UNIVERSIDAD DE GRANADA

Departamento de
Didáctica de la Matemática

ANÁLISIS EPISTEMOLÓGICO DE LA VARIABLE ALEATORIA Y COMPRENSIÓN DE OBJETOS MATEMÁTICOS RELACIONADOS POR ESTUDIANTES UNIVERSITARIOS

Tesis doctoral

Blanca Rosa Ruiz Hernández

PROGRAMA DE DOCTORADO EN DIDÁCTICA DE LA MATEMÁTICA

Granada, España, septiembre 2013

Editor: Editorial de la Universidad de Granada
Autor: Blanca Rosa Ruiz Hernández
D.L.: GR 866-2014
ISBN: 978-84-9028-933-4

ANÁLISIS EPISTEMOLÓGICO DE LA VARIABLE ALEATORIA Y COMPRENSIÓN DE OBJETOS MATEMÁTICOS RELACIONADOS POR ESTUDIANTES UNIVERSITARIOS

T e s i s   d o c t o r a l

M E M O R I A

realizada bajo la dirección de la

Dra. Carmen Batanero Bernabeu

que, para optar al grado de Doctor, presenta

Dña. Blanca Rosa Ruiz Hernández

Fdo: Blanca Rosa Ruiz Hernández

Vo. Bo.

Dra. Carmen Batanero Bernabeu

Trabajo realizado en el marco del Proyecto EDU2010.14947 (MCIN) y
Grupo PAI FQM126 (Junta de Andalucía).

La doctoranda Blanca Rosa Ruiz Hernández y la directora de la tesis Carmen Batanero Bernabeu garantizamos, al firmar esta tesis doctoral, que el trabajo ha sido realizado por el doctorando bajo la dirección de los directores de la tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

Directora de la Tesis	Doctoranda
Fdo.: Carmen Batanero Bernabeu	Fdo.: Blanca Rosa Ruiz Hernández

Granada, 20 de Septiembre, 2013

Índice

Resumen

Análisis epistemológico de la variable aleatoria y comprensión de objetos matemáticos relacionados por estudiantes universitarios

La temática abordada en esta tesis es didáctica de la variable aleatoria, una de las ideas fundamentales propuestas por Heitele (1975) en la enseñanza de la probabilidad, que, como tal, proporciona modelos explicativos en cada etapa del desarrollo del concepto a lo largo de la educación del individuo. Su comprensión se apoya en muchos conceptos, como los de variable, distribución, centro, dispersión, forma o probabilidad, función y función inversa. Otras razones que justifican el tema son las dificultades en su aprendizaje y su interés como herramienta de modelación.

Se presentan cuatro estudios diferentes alrededor de la variable aleatoria desde la perspectiva de la ingeniería didáctica, haciendo uso de algunos elementos del enfoque onto-semiótico, y de los antecedentes de investigación:

- *Estudio 1.* Estudio epistemológico de la variable aleatoria desde la disciplina, relacionándolo con la asignación de probabilidad y la actividad de modelación.
- *Estudio 2:* Análisis del desarrollo histórico de la variable aleatoria que permitirá proponer algunas posibles acciones didácticas e identificar algunas dificultades ligadas a su desarrollo.
- *Estudio 3.* Se analiza una exploración cognitiva en profundidad de las concepciones y comprensión de una pareja de estudiantes sobre la variable aleatoria cuando se enfrentan a la resolución de un problema que requiere el uso del objeto «variable aleatoria». La metodología empleada es una entrevista clínica guiada.
- *Estudio 4:* Se analizan las soluciones escritas de una muestra de 101 estudiantes a un proyecto estadístico donde trabajan con tres pares de variables estadísticas y las variables aleatorias subyacentes.

Los dos primeros estudios proporcionan información detallada sobre el objeto matemático «variable aleatoria» desde el punto de vista matemático e histórico, mientras en los dos últimos se analizan exhaustivamente los razonamientos de los alumnos participantes cuando se enfrentan a las tareas propuestas, proporcionando resultados nuevos sobre sus concepciones y dificultades cuando se enfrentan a problemas vinculados con la variable aleatoria y otros objetos matemáticos relacionados con ella. Estos resultados proporcionan información valiosa en un terreno en que la investigación previa es casi inexistente y se han recogido en una serie de publicaciones que se describen a lo largo de la memoria.

# ABSTRACT

Epistemological analyses of random variable and university students' understanding of related mathematical objects

The topic addressed in this dissertation is the random variable, one of the fundamental ideas proposed by Heitele (1975) in the teaching of probability. As such, it provides explanatory models at each developmental stage of the concept along the education of a subject. Its understanding is based on many concepts, such as those of variable, distribution, center, spread, probability, function and inverse function. Other reasons supporting the interest of the topic are the learning difficulties and its interest as a tool for modelling.

We present four different studies around the random variable from the perspective of didactic engineering, using also some elements from the onto-semiotic approach, and the research background:

- *Study 1.* Epistemological study of the random variable from the discipline, relating it to the assignment of probability and the modelling activity.
- *Study 2:* Analysis of the historical development of random variable that is used to propose some possible teaching activities and to identify some difficulties linked to their development that may appear in the students.
- *Study 3.* We analyze in-depth a cognitive exploration of a couple of students' conceptions and understanding of random variable when solving a problem that requires the use of the object "random variable." The methodology is a guided clinical interview.
- *Study 4:* The written solutions from a sample of 101 students to a statistical project where they work with three pairs of statistical variables and the underlying random variables are analyzed.

The first two studies provide detailed information about the random variable from the mathematical and historical points of view, while the last two analyze in depth the reasoning of the participants when faced with the tasks proposed, providing new results on their conceptions and difficulties when faced with problems related to random variable and related mathematical objects. These results provide valuable information in an area where research is almost nonexistent and have been reflected in a series of publications that are described throughout the dissertation.

INTRODUCCIÓN

Actualmente existe un rezago en las instituciones educativas, con respecto a la demanda que las sociedades exigen del uso, cada vez más intenso y extenso, de la estadística a sus ciudadanos y profesionistas. Para disminuir este rezago, se ha hecho necesario abordar la enseñanza de la probabilidad y la estadística desde una perspectiva científica, para trazar líneas que permitan ocuparse de este tema en nuestras instituciones educativas.

Uno de los primeros en explotar esta perspectiva fue Heitele (1975), quien propuso una lista de diez ideas fundamentales en la enseñanza de la estadística basándose en un enfoque epistemológico-pragmático. Para él, una idea fundamental proporciona modelos explicativos en cada etapa del desarrollo del concepto a lo largo de la educación del individuo. Entre esas diez ideas fundamentales que los estudiantes desarrollan desde niveles educativos básicos y que culminan en niveles superiores, el autor incluyó la variable aleatoria. Sin embargo, es una idea que no ha recibido la atención que amerita y en la que se centra esta Tesis.

La pertinencia del desarrollo de una investigación alrededor de la didáctica de la variable aleatoria también se basa en otras razones, como las dificultades en su aprendizaje y las características particulares del desarrollo del concepto en la probabilidad y la estadística como ciencias. Por un lado, la variable aleatoria constituye un pilar para el desarrollo de las habilidades de modelación en los estudiantes, puesto que es una herramienta indispensable para vincular la realidad con la herramienta matemática, pero también es básica para el desarrollo de la teoría sobre las distribuciones de probabilidad, el teorema del límite central y el estudio de la inferencia. En la variable aleatoria convergen muchos conceptos matemáticos, estadísticos, y probabilísticos provenientes de los niveles educativos básico y medio superior, como los de variable, distribución, centro, dispersión, forma o probabilidad, a la vez que está relacionada con conceptos abstractos propios de niveles superiores, entre otros los de función y función inversa.

En esta memoria reportaremos el trabajo realizado en cuatro estudios diferentes alrededor de la variable aleatoria desde la perspectiva de la ingeniería didáctica y también haciendo uso de algunos elementos del enfoque ontosemiótico, para profundizar sobre las vertientes epistemológicas y cognitivas de la enseñanza y su aprendizaje. El trabajo se ha organizado en los siguientes capítulos:

- *Capítulo 1.* Se presentan los fundamentos del trabajo, mediante la descripción precisa el problema, con énfasis en su relevancia para la didáctica, una exposición de los objetivos y un resumen tanto del marco teórico como de la metodología de la investigación.
- *Capítulo 2.* Está dedicado a los antecedentes del trabajo, que comienzan con los estudios de Piaget sobre el desarrollo evolutivo de la idea de distribución. Aunque son pocas las investigaciones específicas sobre la variable aleatoria, describimos otros temas relacionados, que nos sirven de apoyo, como las investigaciones sobre el concepto de distribución, la distribución normal y la aleatoriedad.
- *Capítulo 3.* En éste se presenta el primer estudio epistemológico del trabajo (*Estudio 1*). Tiene como objetivo analizar el objeto «variable aleatoria» desde la disciplina con la intención de profundizar en él y fundamentar los estudios posteriores. Se analiza el objeto matemático, relacionándolo con la asignación de probabilidad y la actividad de modelación y delimitando. Finalmente, se deducen los objetos de análisis que usaremos en los Estudios 3 y 4.
- *Capítulo 4:* Se describe el *Estudio 2* donde se realiza un análisis del desarrollo histórico de la variable aleatoria que complementa el análisis epistemológico desde la disciplina, en el sentido de que permite conocer más profundamente el objeto de estudio a lo largo de su devenir histórico. Este estudio también permitirá proponer algunas posibles acciones didácticas e identificar algunas dificultades ligadas a su desarrollo que podrían presentarse en los estudiantes.
- *Capítulo 5.* Se analiza una exploración cognitiva en profundidad (*Estudio 3*) de las concepciones y comprensión de una pareja de estudiantes sobre la variable aleatoria cuando se enfrentan a la resolución de un problema que requiere el uso del objeto «variable aleatoria». La metodología empleada es una entrevista clínica guiada. Se trata de conocer cómo dos estudiantes resuelven «en voz alta» y en consenso un problema propuesto.
- *Capítulo 5:* El *Estudio 4* trata de complementar el anterior en un doble sentido. En primer lugar amplía el tamaño de la muestra, de modo que los resultados puedan tener mayor generalidad, por otro, se emplea una situación en la que los estudiantes han de trabajar con tres pares de variables estadísticas asociadas a las variables aleatorias subyacentes. Además se añade una dificultad más, el estudiante se enfrenta con la necesidad de

organizar los datos proporcionados, de modo que, en teoría, el primer paso en el análisis de la situación es la construcción de la distribución por parte de los estudiantes. De este modo accederemos a la comprensión que manifiestan sobre la variable estadística y la relación que establecen entre variable estadística y variable aleatoria, así como a las prácticas que realizan al resolver una actividad centrada alrededor de la variable estadística.

La tesis finaliza con las conclusiones, referencias y algunos anexos incluidos en un CD-ROM en donde se muestran, principalmente, las actividades que se usaron para observar el comportamiento de los estudiantes en los estudios cognitivos.

Los dos primeros estudios descritos proporcionan información detallada sobre el objeto matemático «variable aleatoria» desde el punto de vista matemático e histórico. En los siguientes se analizan exhaustivamente los razonamientos de los alumnos participantes cuando se enfrentan a las tareas propuestas, proporcionando resultados nuevos sobre sus concepciones y dificultades cuando se enfrentan a problemas vinculados con la variable aleatoria y otros objetos matemáticos relacionados con ella.

Aunque el concepto de variable aleatoria es muy amplio y los resultados podrían ser diferentes en otras tareas o tipo de alumnos, nuestros estudios aportan una información inicial en un terreno en que la investigación previa es casi inexistente. Todos estos resultados se han recogido en una serie de publicaciones que se describen a lo largo de la memoria.

Capítulo 1.

Fundamentación de la investigación

ÍNDICE DE CAPÍTULO

1. Introducción

En este primer capítulo se presentan el problema de investigación resaltando su importancia, el marco teórico en que se apoya, los objetivos planteados y un resumen de la metodología, que se detallará más ampliamente en los capítulos posteriores.

En primer lugar se describe la justificación del problema de investigación, principiando con una perspectiva general de la problemática de la didáctica de la estadística. El segundo apartado se concluye destacando la particularidad del tema seleccionado. El tercer apartado está dedicado a la descripción del marco teórico. Sustentamos la investigación en la *teoría de situaciones didácticas* (Brousseau, 1981) complementada en algunos puntos con el análisis del sistema de prácticas matemáticas, a través de la perspectiva *semiótico-ontológica de la cognición e instrucción matemática* que sostiene Godino (2002). En ese apartado nos ocuparemos brevemente de ellas. Los elementos teóricos seleccionados conducen, desde el punto de vista metodológico, a la *Ingeniería Didáctica* (Artigue, 1995a y 1995b) y a algunas metodologías auxiliares que le sirven de apoyo.

El capítulo finaliza con la descripción de los objetos de la investigación y una breve exposición de su metodología, que se ampliará en cada uno de los estudios en que se ha organizado la Memoria.

2. El problema de investigación

La educación en probabilidad y estadística es una demanda cada vez más urgente de nuestras sociedades modernas. En respuesta a ello, tanto profesores como investigadores y las mismas oficinas responsables de la producción de la estadística han comenzado a abordarla esta educación de manera científica. Este interés está justificado puesto que es necesario tanto que los especialistas en el área produzcan y analicen estadísticas, como que profesionistas y ciudadanos sepan interpretarlas y tomar decisiones basándose en ellas. Pero también está vinculado con el rápido desarrollo de la estadística como ciencia, las dificultades intrínsecas de la asignatura en sí misma y su vinculación tan cercana, y en ocasiones tan específica, con otras profesiones.

Así mismo dentro del sistema educativo vivimos cambios, no sólo en los temas tratados sino también en la incorporación de nuevas estrategias de aprendizaje y usos de la tecnología, que requieren una revisión cuidadosa de las modificaciones de los contenidos que con ellos se ocasiona. Las directrices de diseños curriculares, como los del NCTM (2000), de la SEP (2006) en México y del MEC (2006) en España, introducen la estadística desde los primeros niveles escolares aunada a un cambio en la metodología de enseñanza.

Se pretende que los niños sean capaces de comprender la idea de distribución de datos, describir su forma y usar las características estadísticas, como el rango y las medidas de tendencia central, para comparar conjuntos de datos. Deben considerar que los datos son muestras recogidas de poblaciones mayores y llevar a cabo investigaciones y proyectos, considerando el ciclo: formular preguntas, recoger datos y representarlos. Los niños deberían, además, usar programas de ordenadores que les ayuden a representar gráficos, por ejemplo, la hoja de cálculo electrónica. Sugerencias similares, pero a nivel más avanzado, se presentan en los currículos de la enseñanza universitaria y las directrices son muy similares en muchos otros países. Se pretende desarrollar el razonamiento estadístico en los estudiantes e implícitamente se reconoce, como lo indica Scheaffer (2006) y Gattuso y Ottaviani (2011), que este tipo de razonamiento es diferente del matemático, pero ambos complementarios en el currículo global de matemáticas y esenciales en la sociedad moderna.

Aunque la didáctica de la estadística está emergiendo con mucha fuerza en los últimos años como área de investigación, la mayor parte de los trabajos realizados se centran en la enseñanza primaria o secundaria, o bien abordan el tema sólo desde su perspectiva psicológica, por demás necesaria, pero no suficiente. Muchas de las dificultades de aprendizaje de varios conceptos básicos no han sido todavía exploradas y el profesor se encuentra sin recursos para hacerles frente cuando se presentan en sus clases con sus estudiantes.

Dentro de este contexto, en este apartado situamos nuestra investigación acerca del aprendizaje del concepto de variable aleatoria como una idea fundamental dentro de la enseñanza de la estadística sobre el que resulta pertinente detenernos a reflexionar. Precisamos el problema específico que se aborda en este trabajo, continuación de otros trabajos previos de didáctica de la probabilidad a nivel universitario realizados por nuestros compañeros (Serrano, 1996; Tauber, 2001; Alvarado, 2007; Olivo, 2008; Arteaga, 2011; Contreras, 2011).

2.1. Demanda social y desarrollo de la educación en estadística

El acelerado desarrollo de la tecnología, de la ciencia y de los medios de comunicación en las sociedades actuales ha exigido que se incorpore a la educación formal una de las áreas matemáticas de más actualidad: la estadística. Los gobiernos y empresas exigen de la estadística cifras más claras y fiables que permitan tomar decisiones más acertadas en el orden de lo económico, social y político. Para ello es necesaria la cultura estadística básica del ciudadano que ha de colaborar en la recolección de datos (Batanero, 2002), pero que también será usuario de esos datos para establecer criterios al participar en las decisiones que toma su sociedad. Como resultado, los institutos y asociaciones de estadística se están implicando cada vez más en la preparación de materiales y organización de acciones que ayuden a mejorar la *cultura estadística* de los ciudadanos. Cabe aclarar que desde nuestra perspectiva, la cultura estadística es la capacidad de comprender y evaluar críticamente los resultados estadísticos que impregnan nuestras vidas a diario, unida a la valoración de la contribución que el razonamiento estadístico puede hacer en la vida privada y pública y en la toma de decisiones (Gal, 2002).

Batanero (2002) indica que la UNESCO y otros organismos internacionales acordaron que es prioritaria la promoción de la formación estadística como clave para el crecimiento. El Instituto Internacional de Estadística (ISI) fundado en 1885 es otro organismo que impulsa la educación a través de las Conferencias Internacionales sobre Enseñanza de la Estadística (International Conferences on Teaching of Statistics, ICOTS), iniciadas en 1982, de las cuáles la última edición se celebró en Eslovenia en 2008. La Asociación Internacional de Educación Estadística (International Association for Statistical Education, IASE) fundada en 1991 se encarga específicamente de impulsar las acciones educativas en estadística del ISI, como los ICOTS y las Mesas Redondas (Round Table Conferences), congresos donde se discuten temas específicos alrededor de la enseñanza de la estadística.

Por otro lado, la educación estadística también se ha desarrollado dentro de la educación matemática, aunque con menor intensidad. El Comité Internacional para la Instrucción en Matemática (International Committee on Mathematical Education, ICMI) inició los Congresos Internacionales sobre Educación Matemática (International Congresses on Matemátics Education, ICME) a partir de 1973, donde, entre otros temas, se organizan grupos de trabajo en educación estadística, en su última edición, que tuvo lugar en 2008 en Monterrey, México, se organizaron dos grupos diferentes. En este mismo espacio, el ICMI también dedicó uno de sus ICMI Studies al tema particular de la estadística, con el título «Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education». Este estudio en particular se realizó de manera conjunta con una Mesa Redonda de la IASE y se celebró en el Instituto Tecnológico de Estudios Superiores de Monterrey.

En los últimos años, tanto en los congresos celebrados por el Grupo Internacional para la PME (Psychology of Mathematics Education), como en el CERME (Congreso Europeo de Educación Matemática), CIBEM (Congreso Iberoamericano de Educación Matemática) y RELME (Reunión Latinoamericana en Matemática Educativa) se están incluyendo grupos temáticos de educación estadística. Todas estas instancias dan cuenta de la importancia de la educación estadística y de la preocupación de la sociedad por mejorarla.

También en psicología ha habido una gran cantidad de investigación sobre desarrollo cognitivo, a partir del trabajo pionero de Piaget e Inhelder (1951), quienes analizaron exhaustivamente el desarrollo de las ideas de aleatoriedad y probabilidad, del razonamiento combinatorio, de la intuición de la frecuencia relativa, distribución y convergencia, así como de la capacidad de cuantificación de probabilidades en los niños desde sus primeros años a la adolescencia, determinando las diferentes etapas en el desarrollo del razonamiento probabilístico. Este trabajo abrió un campo de investigación muy fuerte en psicología que aún se continúa.

Finalmente, la psicología también ha influido fuertemente en la investigación en educación estadística. Desde la perspectiva de Pérez Echeverría (1990), durante los años 80 hubo un cambio de enfoque en la investigación sobre el razonamiento humano. La postura de un hombre pensante únicamente en términos de reglas de una lógica mental fue cambiada por el de un ser humano que quebranta constantemente esas reglas gracias al trabajo de Kahneman, Slovic y Tversky (1982). Un concepto fundamental de estos autores, es el de *heurística* o *estrategia inconsciente* que reduce la complejidad de un problema probabilístico, puesto que no se realiza un análisis exhaustivo de la información. Aunque las heurísticas ayudan en muchos casos a obtener una solución aproximada al problema, en otros, los mismos investigadores producen sesgos en las conclusiones obtenidas, con las consiguientes implicaciones en las decisiones tomadas. Este enfoque de investigación ha aportado resultados importantes, entre otros puntos, sobre el razonamiento correlacional, la inferencia, la probabilidad condicional y la regla de Bayes. Todas estas investigaciones, para el caso de la probabilidad, se resumen en trabajos como los de Borovcnik, Bentz y Kapadia (1991 a y b), Shaughnessy (1992, 2007) y Jones, Langrall y Mooney (2007).

2.2. Rezago de las instituciones educativas

Hasta ahora, la demanda social de la educación estadística ha sido mayor que la capacidad de las instituciones educativas para responder a la misma. Muchas veces, enseñar estadística en la escuela sólo consiste en realizar cálculos o demostrar teoremas matemáticos y se pone poco empeño en diseñar experimentos, analizar datos o conectar la estadística con el proceso general de indagación. Como consecuencia, los estudiantes finalizan el bachillerato con escasa comprensión de los principios básicos que subyacen en el análisis de datos, lo que explica muchos de sus problemas posteriores en el uso de la estadística en su vida cotidiana o profesional o en los cursos de estadística en la universidad.

Particularmente en México, los estudiantes llegan al nivel universitario con serias deficiencias en el ámbito de probabilidad y estadística y de desarrollo del pensamiento estocástico, como lo manifiestan las academias de estadística del sistema ITESM (Instituto Tecnológico de Estudios Superiores de Monterrey), tomado en cuenta el desempeño escolar, los índices de reprobación y el rápido olvido de los contenidos estudiados en sus alumnos. La situación es similar en España, donde los profesores no siempre encuentran apoyo en los libros de texto, pues las aplicaciones mostradas en los mismos se limitan a juegos de azar o no se basan en datos tomados de aplicaciones reales; encontrándose incluso definiciones incorrectas o incompletas (Serradó, Cardeñoso y Azcárate, 2005).

Sin embargo este problema no es propio de un nivel educativo ni tampoco de algún país aislado, como se puso de manifiesto en los trabajos del Joint ICMI/IASE Study (Batanero, Burrill, Reading y Rossman, 2008; Batanero, Burrill y Reading, 2011) y la incorporación de la estadística desde la educación básica no es todavía un hecho. Esto se puede observar en México en el nivel básico, en donde, a pesar de que en los programas oficiales están especificados contenidos de probabilidad y estadística, suele ser la última unidad que, por falta de tiempo e incluso por dificultades del profesor en el tema, no son estudiados o lo son superficialmente (Alatorre, 1998 y Alquicira, 1998). En el nivel medio superior actualmente no hay una política educativa uniforme para todo el país respecto a la enseñanza de la Probabilidad y Estadística (SEP, 2006), a pesar de que la misma Secretaria de Educación Pública (SEP) de México lo maneja como un tópico obligatorio desde 1982. De esta forma, en muchos planes de estudio de bachillerato no es incluida como asignatura, en otros es una asignatura optativa y en otros más es incluida sólo como tópico en alguna otra asignatura.

En el nivel profesional, hay también una necesidad patente de revisión de los contenidos en la materia de probabilidad y estadística, para satisfacer la demanda de las distintas ingenierías y licenciaturas, quienes exigen una orientación hacia sus áreas de trabajo. Actualmente en el Sistema ITESM la asignatura está diversificada en Estadística para Ciencias Sociales, Probabilidad y Estadística para Ingeniería, Estadística para Economistas, Estadística para Ciencias Biológicas y hay intentos encaminados a un programa de estudio sobre Estadística para Ingeniería Industrial.

En resumen, la adecuación de los contenidos de Probabilidad y Estadística a los distintos contextos y necesidades profesionales, la necesidad de actualización permanente de los profesores, así como las dificultades a las que tanto profesores como estudiantes universitarios se enfrentan al tratar de cubrir contenidos propios del nivel y subsanar deficiencias de niveles anteriores (Batanero, 2002), dificultan que las instituciones educativas respondan adecuadamente a las demandas sociales de formación en Probabilidad y Estadística.

2.3. Problemas de la enseñanza y el aprendizaje de la estadística

Existen también dificultades propias de la enseñanza y el aprendizaje de la estadística, que contradicen la falsa creencia de que el discurso escolar es *transparente*. Estas dificultades se han analizado en conceptos estadísticos aparentemente muy sencillos, tales como los de gráficas estadísticas, medidas de posición central, variabilidad y distribución, sobre los que se apoya la construcción de la idea de variable aleatoria (ver, por ejemplo, Shaughnessy, 2007, las actas de los ocho congresos ICOTS, y las revistas *Journal of Statistics Education* y *Statistics Education Research Journal*). Esto sugiere el interés de analizar la comprensión de los estudiantes sobre la variable aleatoria, donde los problemas en los contenidos previos seguramente incidirán.

Algunos de estos problemas también pudieran ser producto de una enseñanza de la estadística no adecuada o heredados de la enseñanza formalista tradicional de la matemática determinística. Así, por ejemplo, tanto en México como en España la enseñanza de la estadística tiene una fuerte tendencia a prescindir de los fenómenos aleatorios y de «algebraizar» los resultados de la estadística, con la consecuencia, entre otras, que se fomenta muy poco el desarrollo del pensamiento estocástico.

Esto se recrudece en las clases de estadística universitarias debido a que se tiende a pensar que las nociones estadísticas se deben enseñar con un mayor grado de formalidad. La situación ha comenzado a cambiar poco a poco, observándose el interés de algunos profesores por realizar proyectos de innovación docente en estadística, que presentan en los congresos internacionales o bien en las reuniones de las sociedades de estadística mexicana o española. Pero de hecho son pocos los profesores de matemáticas que están bien preparados para enseñar estadística. En los niveles de primaria o secundaria ocurre que, aunque los profesores hayan estudiado estadística, ha sido desde el punto de vista teórico y no conocen bien las aplicaciones ni tienen experiencia en análisis de datos reales, como se reconoce en los trabajos del Joint ICMI/IASE Study. En el nivel superior, los profesores generalmente no tienen una buena instrucción pedagógica que les permita adentrarse en el pensamiento de los estudiantes y diseñar estrategias de enseñanza efectivas.

El problema no sólo radica en la forma en cómo se le presente el contenido a los estudiantes, pues algunas de las dificultades se repiten con estudiantes de diversa edad y en contextos educativos muy diferentes. Gattuso y Ottaviani (2011) resaltan la fuerte especificidad de la educación estadística, que se refleja en las cuestiones filosóficas, éticas, procedimentales e incluso políticas que todavía son objeto de debate en la estadística y sus aplicaciones. Pero, por otro lado, la estadística está mucho más relacionada que las matemáticas con otras ciencias (como psicología, lingüística, geografía, física, ingeniería o economía) que han contribuido al desarrollarlo de muchos métodos estadísticos. Por este motivo, es también más sencillo establecer conexiones con otras áreas curriculares de la escuela en estadística que en matemáticas (Scheaffer, 2006).

2.4. El objeto de estudio: La variable aleatoria

La variable aleatoria es un pilar para el desarrollo de habilidades de modelación en los estudiantes, así como una base para el desarrollo de la teoría relativa a distribuciones de probabilidad, el Teorema central del límite y el estudio de la inferencia entre otros temas fundamentales para la Estadística. No es un objeto sencillo, pues en ella convergen muchos conceptos matemáticos, estadísticos y probabilísticos provenientes de los niveles educativos básico y medio superior, a la vez que está relacionada con conceptos abstractos propios de niveles superiores, entre otros lo de función y función inversa. Su interés curricular se manifiesta desde el nivel medio superior, puesto que esta noción se puede considerar un indicador del aprendizaje de elementos revisados por el estudiante en cursos previos al universitario.

Como analizamos con detalle en Ruiz y Sánchez (2007), la variable aleatoria como idea central en la enseñanza estocástica fue propuesta por Heitele (1975), quien se basa en la concepción de 'Idea fundamental' de Bruner (1960). Para Bruner las ideas fundamentales son útiles desde la perspectiva curricular porque permiten definir cuáles son los tópicos que se estudian en cada grado escolar y con qué profundidad, además de vincularlos entre ellos y con los de otras ramas de las matemáticas. De este modo se genera una red de conocimientos que no sólo propicia el aprendizaje formal de los conceptos sino que también educa la intuición y desarrolla conexiones significantes con la realidad.

Para Heitele una idea fundamental se puede enseñar de manera provechosa en cualquier nivel escolar con niveles de aprendizaje bien definidos. Su lógica está de acuerdo con un currículo en espiral en el que en cada etapa se desarrollan modelos explicativos sobre una idea fundamental, que difieren en los niveles cognitivos, en su forma lingüística y en sus niveles de profundización, pero no en su estructura. El principio básico de currículum en espiral es que un niño es capaz de aprender seriamente cualquier tema en un cierto nivel y que por lo tanto todos los tópicos pueden ser introducidos en una edad temprana, pero que no pueden ser exhaustivos a ninguna edad, sino que sólo pueden ser retomados para incrementar su profundidad.

Heitele pone como ejemplo de este proceso curricular el desarrollo cognitivo de la idea de variable aleatoria. La primera aproximación es un modelo explicativo burdo en donde el niño observa repeticiones del fenómeno e interpreta qué es lo que ocurrirá más frecuentemente, a través de actividades de juego en donde no se involucra una instrucción formal ni analítica. En este primer estadio, Heitele propone la noción de esperanza vinculada con una variable de interés que el niño debe detectar, la percepción de algunos valores que puede tomar y el criterio que se establecería para tomar una decisión con respecto a por cuál valor es por el que más le conviene apostar. Aunque en este nivel el modelo es meramente intuitivo, para Heitele se pre-establecen conocimientos analíticos posteriores, como los de espacio muestral, probabilidad frecuencial y esperanza matemática, además de *pre-figuraciones* del concepto de variable aleatoria a través de la variable de interés.

En una etapa intermedia se pondría énfasis en un modelo más cuantitativo en donde no sólo se detecten sino también se enumeren los eventos posibles y se identifiquen los favorables. Así mismo, se retoma a la variable estadística como una *pre-figuración* de la variable aleatoria y se pre-establece la distribución de probabilidad y la esperanza matemática, pero ya se maneja conocimiento analítico como el concepto de probabilidad clásica (laplaciana). Un modelo más elaborado daría cuenta de la identificación e interpretación de la variable aleatoria, el análisis de la situación completa y la observación de parámetros que definen el comportamiento del fenómeno. El conocimiento aquí explorado sería analítico.

Heitele manifiesta que la variable aleatoria y todo el inventario conceptual relacionado con ella, juegan un papel básico en la matematización de la probabilidad, pero también en las aplicaciones de la estocástica a la vida diaria. Como idea fundamental también tiene un sostén psicológico, porque considera que la intuición de magnitudes en las que participa el azar surge en una etapa más temprana que la de experimento aleatorio.

Desde la perspectiva del modelo explicativo, Heitele considera que la variable aleatoria tiene un papel fundamental en tres aspectos: su distribución, su esperanza y la composición entre variables aleatorias. En la distribución juegan un papel importante la equiprobabilidad, la distribución normal y el teorema central del límite que podrían proporcionar modelos explicativos tanto en niveles primarios como universitarios. La distribución de una variable aleatoria está innegablemente vinculada con sus dos características más importantes, su esperanza y su dispersión, pero Heitele considera a la esperanza como un capítulo especial en las teorías de aprendizaje por su valor explicativo y por lo tanto, en la configuración de las etapas cognitivas de sus modelos explicativos. Las operaciones entre variables aleatorias permitirían construir nuevos modelos matemáticos complejos a partir de uno inicial, lo que daría lugar a un modelo explicativo más complejo.

Miller (1998) también considera el concepto de variable aleatoria como fundamental en la Estadística. Debido a la importancia que este concepto tiene en posteriores nociones estadísticas, como las distribuciones de probabilidad, el modelo de regresión o la obtención de estimadores, indica que la confusión que se puede generar en este tema, tendría repercusiones posteriores importantes en los estudiantes. Por su parte Batanero (2001) afirma que la variable aleatoria es la responsable del paso del estudio de los sucesos aislados al estudio de las distribuciones de probabilidad, además de que, junto con las funciones de distribución, son una herramienta muy potente que permiten hacer uso del análisis matemático en los fenómenos aleatorios.

Los argumentos cognitivos y epistemológicos expuestos anteriormente para afirmar que la variable aleatoria como idea fundamental resaltan la importancia de realizar investigaciones que confirmen y sustenten de manera más detallada el desarrollo de ese concepto a lo largo del currículo escolar. Nuestro trabajo se orientará a realizar análisis epistemológicos y cognitivos sobre este objeto matemático. Indirectamente, la metodología usada en los estudios cognitivos permitirá también analizar posible situaciones a-didácticas que se podrían refinar posteriormente y serían útiles para evaluar y desarrollar *pre-figuraciones* intuitivas y posteriormente la profundización y abstracción del concepto.

A continuación presentamos el marco teórico y la metodología de nuestro trabajo. El capítulo concluye con la clarificación de los objetivos de esta investigación desde la perspectiva de nuestro marco teórico y la metodología utilizada.

3. Marco teórico

En esta sección primero expondremos el marco teórico que respalda nuestra investigación para abordar y plantear la problemática tratada en los incisos anteriores, y su metodología. Nos apoyamos fundamentalmente en la *teoría de situaciones didácticas* planteada por Brousseau (1981, 1986 y 1997) que nos conduce, desde el punto de vista metodológico, a la Ingeniería Didáctica (Artigue, 1995a y 1995b), aunque nos limitaremos a la fase del análisis a priori. Aunado a este enfoque, nos parece útil usar algunos elementos de la *perspectiva semiótico-ontológica de la cognición e instrucción matemática* que propone Godino (2002).

Todo ello nos servirá de apoyo en la descripción y análisis del objeto variable aleatoria, tanto desde la cognición matemática como desde su epistemología.

3.1. Teoría de situaciones didácticas

La teoría de situaciones didácticas toma como objeto de estudio al sistema didáctico y, más ampliamente, al sistema de enseñanza, que se sustenta en el triángulo epistemológico: docente-saber matemático-alumno y sus relaciones. A partir de esta perspectiva, la ingeniería didáctica propone un análisis preliminar de los conceptos desde tres dimensiones: la didáctica, la epistemológica y la cognitiva. Los resultados aportados por las tres componentes dan lugar al diseño de una secuencia didáctica que permitirá el estudio de la didáctica del objeto matemático en el aula a través de variables de observación relevantes también fundamentadas en el análisis preliminar. Este análisis requiere, por tanto, de una investigación seria que profundice en la naturaleza del objeto matemático desde cada uno de sus componentes y sus interrelaciones.

Siguiendo esta teoría, en nuestro trabajo se aborda el análisis preliminar del objeto «variable aleatoria». Los resultados de los análisis servirán como orientación para el diseño de situaciones didácticas apropiadas a los tipos de estudiantes que participaron en la investigación y proporcionará fundamentos para futuras investigaciones basadas o no en esta teoría.

Brousseau (1997) vislumbró por primera vez la necesidad de utilizar un modelo propio de la actividad matemática en la investigación en didáctica de la matemática a partir de la problematización y cuestionamiento del conocimiento matemático enseñado. A ese modelo le llamó *Teoría de situaciones didácticas*. Brousseau está de acuerdo con la *Teoría de equilibración mayorante* de Piaget (1991). El reto está en construir *situaciones didácticas* con las que el estudiante se involucre con *situaciones a-didácticas* que permitan que un conocimiento matemático particular emerja de la actividad del estudiante. En los párrafos siguientes describiremos brevemente algunos de los aspectos más relevantes de esta teoría.

3.1.1. Teoría de la equilibración mayorante de Piaget

Una de las ideas centrales de la *Teoría de la Equilibración mayorante* de Piaget (1991) es que tanto la naturaleza como la validez de los conocimientos dependen de su modo de formación. La postura de Piaget es que el conocimiento se adquiere mediante la interacción entre el sujeto y el objeto. De aquí que la acción sea para él un principio fundamental para el aprendizaje. La naturaleza dialéctica de la teoría de Piaget es un proceso complejo de estructuraciones sucesivas a través de una jerarquía de niveles bien definidos. El conocimiento es, desde un punto de vista metodológico, un proceso continuo cuya iniciación o finalización no pueden alcanzarse nunca y por lo tanto debe enfocarse siempre como un relativo a un estado anterior y de menor conocimiento y a su vez susceptible de ser construido a partir de un estado anterior. No se trata, afirma Piaget, de cortes arbitrarios en el seno de un proceso continuo o puramente aditivo, las estructuras adquiridas en un nivel dan lugar a una reconstrucción antes de que estas estructuras reconstruidas puedan ser integradas en las nuevas estructuras elaboradas sobre los niveles anteriores.

Cada uno de los niveles constituye un estado de *equilibrio dinámico* a la manera de los estados de equilibrio estático de un sistema de equilibrio (situaciones «estacionarias») de un sistema termodinámico. Piaget lo llama *equilibración*. En la medida que el desarrollo del conocimiento es concebido como un juego, se entra en mecanismos de *desequilibración* del anterior nivel y de *reequilibración* en los nuevos niveles que se van alcanzando. Esto tiene repercusiones importantes en la didáctica porque no hay acumulación progresiva de saberes, sino una reorganización permanente de conocimientos: los nuevos saberes son integrados al saber anterior, aun a veces modificando o confrontando este último. Es necesario, por tanto que se establezca un *conflicto cognitivo* que provoque un estado de desequilibración-reequilibración para que surja un nuevo conocimiento.

La concepción básica más original de la teoría epistemológica de Piaget consiste en afirmar que la *acción* es constitutiva a todo conocimiento. El conocimiento es dependiente de la acción y la acción es productora de conocimiento. La actividad propia del sujeto no se ejerce forzosamente mediante la manipulación de objetos materiales, sino de una acción finalizada y problematizada que supone una dialéctica pensamiento-acción muy diferente a una simple manipulación guiada, orientada con frecuencia a una simple tarea de constatación.

En la acción elemental todavía no puede hablarse, en sentido estricto, ni de un sujeto ni de un objeto. Poner en el punto de partida la acción es, por un lado, sustituir las opciones clásicas (primacía del sujeto en el idealismo o del objeto en el empirismo) con un nuevo enfoque: la primacía es la del vínculo práctico, de la interacción efectiva, de la acción objetiva. Pero, por otro lado, es adoptar una perspectiva constructivista que dé cuenta de la constitución del sujeto en tanto sujeto cognoscente y del objeto en tanto objeto de conocimiento. Este papel de la acción rompe la vieja dicotomía entre pensamiento y acción. Tal como lo señala Piaget: «todas las teorías no-genéticas conciben al *pensamiento* como anterior a la *acción* y a ésta como una aplicación de aquél» (Piaget, 1991, p. 16). De este modo, la acción en la didáctica de las matemáticas juega un papel esencial, propone una forma de apropiación del conocimiento por parte del alumno.

En esta investigación se tendrá en cuenta esta teoría en los dos estudios cognitivos, en los que se pretenderá que el objeto «variable aleatoria» se construya a partir de situaciones que requieren la implicación activa (acción) del estudiante.

3.1.2. Situaciones a-didácticas

Brousseau (1997) sostiene que la concepción moderna de la enseñanza es pedir al maestro que provoque en el alumno las adaptaciones deseadas, con una elección acertada de los «problemas» que le propone. Para que el alumno pueda aceptar estos problemas, deben hacerle actuar, hablar, reflexionar y evolucionar por sí mismo. Entre el momento en que el alumno acepta el problema como suyo y aquél en el que produce su respuesta, el maestro rehúsa intervenir, es decir, espera que el estudiante descubra por sí solo los conocimientos que quiere ver aparecer. El alumno sabe bien que el problema ha sido elegido para hacerle adquirir un conocimiento nuevo, pero también que este conocimiento está enteramente justificado por la lógica interna de la situación y que *puede* construirlo sin atender a razones didácticas a través de la interacción con el problema. No sólo puede, sino que también debe, pues sólo habrá adquirido verdaderamente este conocimiento cuando él mismo sea capaz de ponerlo en acción, en situaciones que encontrará fuera de todo contexto de enseñanza, y en ausencia de cualquier indicación intencional. Tal situación es llamada *a-didáctica* por Brousseau.

Cada conocimiento puede caracterizarse por una o más situaciones a-didácticas que preservan su sentido y que llamaremos *situaciones fundamentales*. Pero el alumno no puede resolver de golpe cualquier situación a-didáctica, por eso el maestro le propone, entre las situaciones a-didácticas, aquéllas que están a su alcance. Estas situaciones a-didácticas, ajustadas a fines didácticos, determinan el conocimiento enseñado en un momento dado y el sentido particular que este conocimiento va a tomar. Para Brousseau la definición del conocimiento matemático enseñado se establece a través de una *situación fundamental*, que, para un conocimiento matemático *C* en particular, definió como «un conjunto mínimo de situaciones *a-didácticas* (específicas de *C*) que permiten engendrar, por manipulación de los valores que toman sus *variables didácticas*, un campo de problemas suficientemente extenso como para proporcionar una buena representación de *C* en relación a cómo ha sido construido *C* en la institución didáctica en cuestión» (Chevallard, Bosch y Gascón, 1997, p 216).

En nuestro estudio buscamos dos situaciones complementarias[1] de las que surge la idea de variable aleatoria y en las que el estudiante se interesa personalmente:

- La primera (*Estudio 3*) es una situación de sorteo, en la que el estudiante debe tomar una decisión. Se asemeja a problemas de juego de los que surge la idea de variable aleatoria y esperanza matemática y su análisis requiere el manejo de la concepción clásica de probabilidad.
- La segunda (*Estudio 4*) es una investigación que precisa recoger y analizar datos para poner a prueba sus propias intuiciones. En este caso el problema propuesto es similar a otros de los que surge la necesidad de realizar estudios estadísticos y de los que surge tanto la variable estadística como la aleatoria, además de que el tipo de probabilidad empleado es frecuencial.

3.1.3. Situación didáctica

La situación o el problema elegido por el profesor es una parte esencial de la siguiente situación más amplia: el maestro busca *delegar* al alumno una situación a-didáctica que provoque en él una interacción lo más independiente y lo más fecunda posible. Para ello, comunica o se abstiene de comunicar, según el caso, informaciones, preguntas, métodos de aprendizaje, heurísticas, etc. En consecuencia, el enseñante está implicado en un juego con el sistema de interacciones del alumno y con los problemas que él le ha planteado. Esta situación más amplia es la *situación didáctica* (Brousseau, 1986).

Para construir buenas situaciones didácticas se consideran los siguientes aspectos, que se han tenido en cuenta en cada uno de los dos estudios cognitivos realizados:

- La actividad propuesta como punto de partida debe presentar un verdadero problema para los alumnos, pero, a la vez, ser comprendido por ellos. Es

[1] Que potencialmente, mejoradas podrían llegar a convertirse en situaciones a-didácticas.

decir, los estudiantes deben poder pensar y planear la respuesta del problema.

- ❖ Debe permitir al alumno utilizar conocimientos anteriores (para que pueda introducirse en el problema).
- ❖ Y debe ofrecer una resistencia suficiente para llevar al alumno a hacer evolucionar sus conocimientos anteriores, a cuestionarlos, a elaborar nuevos conocimientos.
- ❖ Debe contener, en lo posible, su propia validación, es decir, el alumno debe poder por sí mismo – o confrontado con los otros alumnos – controlar su solución, decidir su validez de respuesta.
- ❖ El conocimiento previsto debe ser el más apropiado para resolver el problema.

La situación didáctica implica una interacción dialéctica del estudiante con situaciones problemáticas, donde el sujeto anticipa y finaliza sus acciones y compromete sus conocimientos anteriores, los somete a revisión, los modifica, los complementa o los rechaza para formar concepciones nuevas. El objeto principal de la didáctica desde la perspectiva de la teoría de situaciones didácticas es estudiar las condiciones que deben cumplir las situaciones planteadas al alumno para favorecer la aparición, funcionamiento o rechazo de esas concepciones.

Para Brousseau (1986) el principio fundamental de la teoría de situaciones didácticas es el principio metodológico «*definir todo 'objeto matemático' mediante una situación*». Esto es, la «planeación documentada» de situaciones didácticas. Desde una perspectiva metodológica se habla de la búsqueda de fundamentos para diseñar secuencias didácticas que planteen problemas de los que surja de manera natural el objeto matemático al interactuar con ellos, del planteamiento de posibles conductas del estudiante al abordar esos problemas y del estudio de las interacciones del profesor con el estudiante al resolver el problema y de la verificación de tales hipótesis en el aula. Estas ideas se ponen de manifiesto en la *ingeniería didáctica*, que ha sido utilizada como la metodología de la teoría de situaciones didácticas. En este sentido, y aunque somos conscientes de la necesidad de revisión y mejora, las situaciones utilizadas en nuestros estudios cognitivos pueden ser punto de partida para unas futuras situaciones didácticas y a-didácticas que formen parte de una serie de situaciones fundamentales para la enseñanza de la variable aleatoria.

3.1.4. Concepciones y errores

Como ya se ha mencionado, para Piaget el desarrollo del conocimiento se realiza a través de mecanismos de equilibración y desequilibración con la consecuente reorganización de conocimientos en el sujeto. En la teoría de situaciones didácticas, los nuevos estadios están vinculados con la superación de *obstáculos* por los que el estudiante necesariamente debe pasar para provocar un estado de desequilibración-equilibración y por lo tanto surja un nuevo conocimiento.

Así, el estudio de las acciones y producciones de los estudiantes al enfrentarse a un problema también debe ser un punto a tomar en cuenta en el análisis preliminar. Es de esperar que de este análisis surjan hipótesis que sustenten nuevos análisis a-priori y el diseño de la secuencia didáctica. De este modo, la didáctica de las matemáticas también se dedica a estudiar las *concepciones* del sujeto que aprende y de las condiciones en las cuales se construye el conocimiento.

El término «concepción» ha sido utilizado de una forma muy amplia en didáctica, (Artigue, 1990) y como sinónimo de otros, como «concepto», «sentido», «noción», «comprensión» etc. Vergnaud (1990) propuso que la piedra angular de la cognición era la conceptualización. Sin embargo se enfocó a la faceta epistémica de las nociones «concepto» y «campo conceptual», sin mencionar una perspectiva cognitiva (aunque es posible que las distinguiera implícitamente). Para Vergnaud, el «concepto» se describe a través de la tripleta formada por el conjunto de *situaciones* que dan sentido al concepto, el conjunto de *invariantes operatorios* que constituyen el concepto (objetos, propiedades y relaciones) y el conjunto de *representaciones simbólicas* usadas para presentar esos invariantes y las situaciones y procedimientos. También introduce la idea de «campo conceptual» como un conjunto de situaciones que da sentido a un concepto e incluye sus invariantes y significantes.

Artigue (1990) usa el término «concepción» para referirse a distintos puntos de vista o forma de comprender un objeto matemático (diferenciándolo del término «concepto»). La autora también indica que en este término se pueden considerar dos puntos de vista complementarios: uno epistémico (objetos matemáticos y su funcionamiento) y otro cognitivo (los conocimientos del sujeto en relación a un objeto matemático particular). Artigue discute dos usos del término concepción por distintos autores:

- ❖ Como estado cognitivo global, tiene en cuenta la totalidad de los conocimientos y creencias (en general la estructura cognitiva) del sujeto respecto a un cierto objeto matemático. En este sentido «concepción» está muy próximo al «concepto» de Vergnaud, entendido como una tripleta (situaciones, invariantes y significantes).

- Como un objeto local, estrechamente asociado al saber puesto en juego y a los diferentes problemas en cuya resolución intervienen. Sería la parte de la concepción que se pone en juego en unas circunstancias dadas.

Aunque en nuestro trabajo, en principio estamos interesados por el punto de vista global de la concepción sobre la variable aleatoria de los sujetos participantes en los estudios, reconocemos que en cada uno de ellos posiblemente observaremos una parte local de estas concepciones; más específicamente la que se pone en juego para resolver las tareas propuestas.

Sfard (1991) por su parte diferencia entre concepto y concepción. Usa «concepto» para referirse al conocimiento matemático establecido, y «concepción» para hacer referencia «al aglomerado completo de representaciones internas y asociaciones evocadas por el concepto -la contrapartida del concepto en el universo interno o subjetivo del conocimiento humano» (p. 3). Tanto para los conceptos como para las concepciones Sfard diferencia los puntos de vista operacional (proceso que se usa en unas circunstancias y conduce a una serie de acciones) y estructural (estructura estática del concepto o concepción en un momento dado), que son complementarios. En este punto de vista la parte estructural de la concepción sobre variable aleatoria implicaría el conjunto de propiedades y relaciones que un sujeto es capaz de establecer en un momento dado y la parte operacional la forma en que la variable aleatoria se pone en juego al resolver problemas y la serie de acciones que implica (tales como formar su distribución o comparar sus promedios).

Es frecuente que el análisis fino de los errores producidos en tareas propuestas ayude al investigador a forjar hipótesis sobre las concepciones de los alumnos y en ese sentido interpretamos el *error* en nuestro trabajo. Lo diferenciamos de la ignorancia, que es un no-conocimiento, en que el error se ejerce sobre un material ya dado, es decir, conlleva conocimiento, pero en su manipulación conceptual lo trastorna y lo ensambla incorrectamente, de modo que no ajusta entre sí. El error aparece, por tanto, como algo inherente al sistema de conocimientos (Muñoz y Valverde, 2000)[2].

[2] También hay errores que no son debidos a una mala manipulación conceptual, sino testimonio de un conocimiento que tuvo su ámbito de validez en otro contexto o nivel de conocimiento, a este tipo de errores, Brousseau (1981) los llamó *obstáculos epistemológicos*. No los estudiaremos en este trabajo.

3.2. Elementos del enfoque onto-semiótico

Como hemos indicado, utilizaremos algunos elementos del enfoque ontológico-semiótico de la cognición e instrucción matemática, desarrollado principalmente por Godino y colaboradores (Godino y Batanero, 1994, 1998; Godino, 2002; Godino, Batanero y Font, 2007), que pretende ser un marco unificado para el estudio de las formas del conocimiento matemático y sus interacciones dentro de los sistemas didácticos. Godino se centra en la problematización de la naturaleza del objeto matemático como base de estudio de los procesos de enseñanza y aprendizaje. Retoma diversas aproximaciones teóricas de la didáctica de la matemática y establece un enfoque útil para caracterizar los significados elementales y sistémicos de los objetos matemáticos puestos en juego en los procesos de estudio matemático. Se apoya fuertemente en las perspectivas ontológica y semiótica de la actividad matemática y reconoce que sus planteamientos surgen de una perspectiva antropológica con elementos de la didáctica francesa (Godino, Font, Contreras y Wilhelmi, 2006).

Los procesos de enseñanza y aprendizaje de los contenidos matemáticos se sitúan en el seno de una institución educativa y unos alumnos tratando de apropiarse de ellos con la ayuda de un profesor en circunstancias específicas. Esto es, el proceso es sistémico y dinámico, de modo que Godino (2002) propone el desarrollo de un modelo teórico que articule las facetas *epistemológica* (centrado en la naturaleza de los distintos componentes del conocimiento matemático a estudiar en una institución en particular), *ontológica-cognitiva* (centrado en los procesos de interpretación de significados institucionales por parte de los sujetos interpretantes) e *instruccional* (centrado en la interacción entre las funciones docentes, discentes y los distintos componentes epistémicos y cognitivos) a partir del análisis del objeto matemático. En este trabajo nos centraremos tanto en la faceta epistemológica (Estudios 1 y 2), como en la cognitiva (Estudios 3 y 4).

3.2.1. Significado de un objeto matemático

También en concordancia con las ideas piagetianas anteriormente expuestas, este marco teórico se sostiene en tres aspectos importantes del conocimiento matemático:

- ❖ Las matemáticas son un quehacer humano que surgen como respuesta o solución a *problemas* externos o internos. Estos problemas y sus soluciones son compartidos por un grupo de personas implicadas en el estudio de una cierta clase de problemas. De modo que los objetos matemáticos son entidades culturales socialmente compartidas que emergen y evolucionan gracias a las acciones de las personas.

- Las matemáticas crean un *lenguaje simbólico* propio en el que se expresan las situaciones problemas y sus soluciones, que cumple con una función comunicativa, pero también instrumental.
- Uno de los propósitos de la actividad matemática es la construcción de un *sistema conceptual lógicamente organizado*, que se ve modificado, en sus relaciones internas y en su tamaño, cuando un nuevo conocimiento se le añade (Godino y Batanero, 1998).

Para formular una ontología del objeto matemático de acuerdo con estos aspectos, Godino (2002) retoma de Chevallard (1991) la idea de objeto matemático como un emergente de un sistema de prácticas para relacionar el *significado sistémico de un objeto matemático* con las prácticas que realiza un individuo o institución cuando enfrenta una situación problema. De Chevallard (1992), también, retoma la relatividad del conocimiento con respecto a las instituciones y añade una faceta personal, para proponer los constructos de *significado institucional* y *significado personal*. El sentido de 'las prácticas' en el contexto de esta teoría se refieren concretamente a *prácticas significativas* que incluyen tanto componentes operatorios como discursivos, esto es, cuando la práctica desempeña una función para la consecución del objetivo en los procesos de resolución de un problema o bien para comunicar la solución, para validarla o para generalizarla a otros contextos o problemas.

A su vez, el significado se concibe como el contenido asignado a una expresión que designa un objeto (función semiótica) es decir, es aquello a lo cual se refiere un sujeto en un momento y circunstancias dadas. Así, las prácticas significativas se relacionan con un campo o tipo de problemas y se conciben en contexto institucional, una persona (o institución) y los instrumentos semióticos que matizan la acción. También se propone una investigación en didáctica que centre su atención principalmente en el estudio de las relaciones de los significados institucionales de los objetos matemáticos y los significados personales construidos por los sujetos y por lo tanto se introduce la necesidad de un análisis más profundo de los componentes del significado del objeto matemático.

3.2.2. Objetos que intervienen en los sistemas de prácticas

Godino se basa en diferentes autores principalmente para desglosar el significado del objeto matemático en sus componentes. Toma la tripleta de la noción de «concepto» de

Vergnaud (1990) y añade explícitamente las acciones de los sujetos en su afán por resolver un problema, sus argumentos y las propiedades de los objetos puestas en juego. La forma en que queda constituido el significado de un objeto matemático en esta teoría y que tendremos en cuenta en nuestro trabajo, es a través de los siguientes elementos:

- *Las situaciones-problema*: Son los problemas y campos de problemas, que incluyen tanto problemas de situaciones simples y complejas, como problemas propiamente matemáticos, de donde emerge un objeto matemático y que inducen la actividad matemática.
- *El lenguaje matemático*: Para resolver los problemas matemáticos, generalizar su solución o describirlos a otra persona se hace uso de expresiones, notaciones, gráficos, etc. Aquí también se incluye la notación simbólica, que puede ser usada para representar objetos abstractos, situaciones concretas y disposiciones tabulares. Estos sistemas de signos no sólo tienen la función comunicativa sino un papel instrumental que modifica al propio sujeto que los utiliza como mediadores.
- *Las acciones del sujeto*. Son los modos de actuar del sujeto (o institución) ante las situaciones-problema o tareas. Con este término asignan las diversas operaciones, algoritmos, técnicas de cálculo, procedimientos y estrategias que llegan a automatizarse.
- *Los conceptos-definición*. Son los conceptos o nociones matemáticas que previamente se conocen y en los que se apoya la resolución del problema mediante sus definiciones o descripciones.
- *Las propiedades o atributos*. Propiedades específicas del objeto matemático. Suelen darse mediante proposiciones o descripciones y se relacionan con las diferentes definiciones que se le atribuya al objeto.
- *Argumentos*. Son razonamientos o argumentaciones que se usan para comprobar la validez de las soluciones o de los procesos que se ponen en juego. Pueden ser deductivas, con contraejemplos, generalización, análisis y síntesis, simulaciones con computadora, demostraciones informales, etc.

Dentro de esta teoría, una *persona comprende* un objeto matemático cuando se apropia del significado institucional del objeto, es decir, se apropia de los distintos elementos que componen los significados institucionales correspondientes al objeto matemático, pero además exige que el sujeto identifique en el objeto un para qué y una intencionalidad. Godino (2002) no diferencia la *competencia*, entendida como un ‘saber hacer’, y la *comprensión*, que implica qué hacer y por qué, tradicionales en la didáctica. En su teoría, los seis elementos primarios del significado matemático están ligados tanto a la competencia como a la comprensión puesto que en cada uno de ellos se puede diferenciar un saber hacer y un saber por qué se hace. En este sentido utilizaremos el término «comprensión» en nuestro trabajo.

3.2.3. Facetas duales del conocimiento matemático

Para el análisis de la actividad matemática también es necesario considerar una serie de procesos interpretativos que involucran a los objetos matemáticos y sus elementos de significado. Con esta finalidad, el modelo ontológico propuesto se complementa con la consideración de las facetas tanto referidas a los objetos y sus significados como un todo, como para cada elemento de significado. Las facetas son duales y dialécticas, se presentan en parejas que se complementan y se pueden aplicar tanto a entidades primarias como secundarias (Godino, Batanero y Font, 2007).

En este trabajo sólo nos ocupamos de la faceta personal-institucional, en cuanto analizamos el significado personal que los alumnos asignan a la variable aleatoria y lo comparamos con el significado institucional fijado en la actividad implementada.

- *Facetas personal-institucional*. Dependiendo de las circunstancias contextuales del estudio, el significado del objeto matemático puede situarse en una persona o en una institución. Si se trata de la manifestación de un individuo, como respuesta a una prueba de evaluación, la realización de una tarea escolar por un estudiante, se habla de objetos personales. Si se trata de documentos curriculares, libros de texto, explicaciones de un profesor ante su clase, se consideran objetos institucionales. Se trata de distinguir «posiciones» en el sistema de didáctico que se refleja también en los propios objetos de enseñanza. Dentro de esta faceta se pueden considerar también diferentes tipos de significados tanto personales (referencial, pretendido, implementado y evaluado) como institucionales (global, declarado y logrado).
- *Facetas elemental y sistémica.* Los componentes del objeto matemático se pueden ver como entidades compuestas, sistémicas, con una cierta organización y estructura, o bien, por el contrario como entidades que se ponen en juego de manera transparente, como si se tratara de entidades unitarias o elementales.
- *Facetas ostensiva y no ostensiva.* Cualquier objeto matemático tiene una faceta ostensiva, esto es perceptible, y una no ostensiva dependiendo del contexto en el que se maneje. Las entidades lingüísticas se muestran directamente ostensivas, pero el resto de las entidades no son directamente

perceptibles como tales. El lenguaje es el medio por el cual se expresan los no ostensivos (es la fase ostensiva de los objetos matemáticos), aunque también tiene una faceta no ostensiva, puesto que se puede referir a una palabra a través de un símbolo o de una representación.

- ❖ *Facetas ejemplar y tipo.* Esta faceta está relacionada con lo que en matemáticas comúnmente se distingue como abstracto y concreto sólo que en este marco teórico se puede aplicar a cualquiera de los componentes del significado no sólo a los conceptos. Un objeto concreto (tipo) es algo que se pone en juego *por sí mismo*, en cambio será un ejemplar cuando es el representante de dichos objetos, como ejemplar de un cierto tipo o componente de un sistema. Esta distinción surge para satisfacer la necesidad del estudio de las matemáticas de generalizar los problemas, las soluciones y el discurso en el que se suscriben y organizan. Una misma entidad puede ser ejemplar o tipo dependiendo del juego de lenguaje en el que se encuentre.
- ❖ *Facetas expresión y contenido.* En esta faceta se encarna la característica relacional de los objetos matemáticos, puesto que no se pueden concebir como entidades aisladas. La expresión y el contenido están relacionados con el significante y significado que relaciona cada uno de los distintos objetos matemáticos, se trata, pues, de las correspondencias entre un antecedente (significado, expresión) con un consecuente (contenido o significado) establecidas por una persona (o institución) de acuerdo a un cierto criterio. Las expresiones pueden ser de tipo instrumental (cuando un objeto usa a otro como instrumento) o referencial (cuando un objeto se pone en lugar de otro).

Nosotros nos centramos en la faceta personal- institucional, porque partiendo del análisis epistemológico disciplinar que define el significado de referencia en la institución, analizamos las prácticas personales de los estudiantes en la exploración cognitiva, que definen los significados personales de los estudiantes.

La faceta elemental-sistémica se tiene siempre en cuenta en forma implícita, pues al objeto «variable aleatoria» le asociamos un sistema de objetos y prácticas.

3.2.4. Idoneidad Didáctica

Desde la perspectiva del enfoque onto-semiótico de la cognición matemática, la idoneidad global de una configuración didáctica se evalúa a través de cinco criterios que articulan las configuraciones docente y discente: idoneidad epistémica, cognitiva, interaccional, mediacional y emocional (Godino, Wilhelmi y Bencomo, 2005; Godino, Contreras y Font, 2006). Estos cinco criterios se describen como:

- *Idoneidad epistémica*: Representatividad de los significados institucionales implementados (o pretendidos), respecto de un significado de referencia. Es decir, se busca identificar si los significados de los objetos presentes en un recurso son adecuados desde el punto de vista matemático.
- *Idoneidad cognitiva:* Grado en que los significados pretendidos/ implementados son asequibles a los alumnos, así como si los significados personales logrados por los alumnos son los significados pretendidos por el profesor (en nuestro caso por el investigador).
- *Idoneidad interaccional*: Grado en que la organización de la enseñanza (investigación) permite identificar errores, dificultades y conflictos y si es posible resolverlos a través del intercambio de opiniones, argumentos y significados.
- *Idoneidad mediacional:* Disponibilidad y adecuación de los recursos necesarios para el desarrollo del proceso de enseñanza-aprendizaje.
- *Idoneidad emocional:* Interés y motivación del alumnado en el proceso de estudio (investigación).

Estos criterios de idoneidad tienen la finalidad de orientar el diseño, la implementación y evaluación de los procesos de instrucción matemática. En nuestro caso, hacemos uso de esta herramienta para evaluar un proceso cuyo objetivo principal está en la investigación de los procesos de aprendizaje. Nuestra evaluación dará prioridad a la profundización sobre los argumentos y razonamientos de los estudiantes, más que a un proceso de enseñanza en sí. Pero para los estudiantes que participan en una investigación, generalmente el proceso sí involucra una situación de enseñanza-aprendizaje por lo que dichos criterios sí son pertinentes para valorar si los resultados serán de utilidad para concretar futuras situaciones de enseñanza.

4. Objetivos de investigación

El objetivo general del presente trabajo es realizar algunos estudios preliminares sobre la didáctica de la variable aleatoria que permitan posteriormente el diseño fundamentado de situaciones didácticas. De acuerdo a las consideraciones hechas en el capítulo y al marco teórico seleccionado, se puede descomponer en dos objetivos fundamentales:

1. *Llevar a cabo un análisis a priori sobre la variable aleatoria desde el punto de vista epistemológico, que aborde tanto los aspectos disciplinares como los históricos.*

   Esta exploración se lleva a cabo en los Estudios 1 y 2. El primero de ellos permite fijar el significado de referencia de la variable aleatoria para nuestra investigación. Dicho significado se utiliza para seleccionar situaciones (que en un futuro podrían derivar en situaciones a-didácticas y didácticas) y los instrumentos de recogida de datos en los Estudios 3 y 4, establecer hipótesis sobre los resultados de dichos estudios e interpretar algunas de las dificultades y errores de comprensión que manifiestan los estudiantes que participan en los mismos.

   El Estudio 2 (histórico) proporciona apoyo para explicar algunas concepciones parciales o incorrectas y errores observados en las respuestas de los estudiantes, por su paralelismo con algunas de las observadas en periodos históricos anteriores al desarrollo de la variable aleatoria. También provee criterios para elegir las situaciones propuestas a los estudiantes en los Estudios 3 y 4; una de ellas próxima a los campos de problemas en los que se desarrolla el significado clásico y la otra a aquellos en los que se desarrolla el significado frecuencial de la probabilidad y la inferencia informal.

2. *Llevar a cabo una exploración cognitiva de la comprensión intuitiva que muestran algunos estudiantes universitarios al trabajar con situaciones en las que se manifiesta la variable aleatoria y otros objetos matemáticos relacionados con ella, y que implican resolución de problemas y proyectos con una orientación hacia la modelación.*

   Esta exploración se lleva a cabo en los Estudios 3 y 4, donde se toman dos situaciones (potencialmente situaciones a-didácticas para la variable aleatoria) y se recogen los datos de estudiantes cuando interactúan con ellas. Las situaciones elegidas son complementarias bajo los criterios ya mencionados de concepción de probabilidad subyacente, método de recogida de datos y tamaño de muestra. Los resultados de esta exploración proporcionarán también elementos que complementan el análisis a priori de la variable aleatoria desde el punto de vista cognitivo.

5. Metodología de investigación

La metodología de esta investigación, tiene diversas fases y características y, acorde con el marco teórico en el que se sustenta, sigue las directrices de la ingeniería didáctica. Específicamente la centramos en los estudios preliminares que se suelen llevar a cabo en esta metodología y dentro de ellos, a las componentes cognitiva y epistemológica. En este último aspecto nos enfocamos al análisis del concepto desde la disciplina y desde la historia. La estructura general de la memoria se presenta en la Figura 1.1.

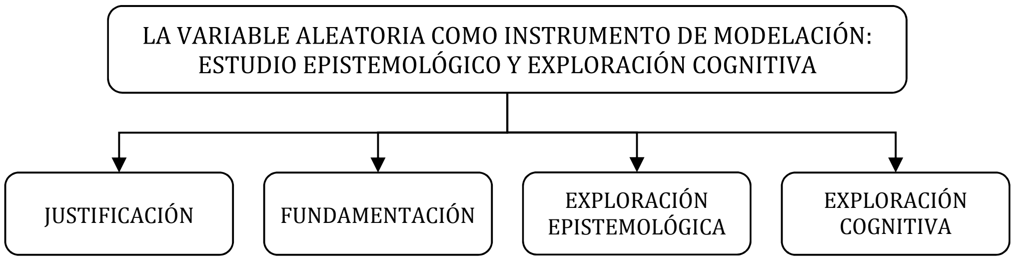


Figura 1.1. Estructura general de la investigación

En la justificación y fundamentación de la investigación (Figura 1.2), puntualizamos su pertinencia, consistencia y coherencia. La primera a través de su justificación social, didáctica, epistemológica y cognitiva. El marco teórico y la metodología proporcionan la fundamentación que dan coherencia y consistencia a la investigación. Con esa base, se proponen dos tipos de acercamientos al estudio de la variable aleatoria, uno de corte epistemológico y el otro de corte cognitivo.

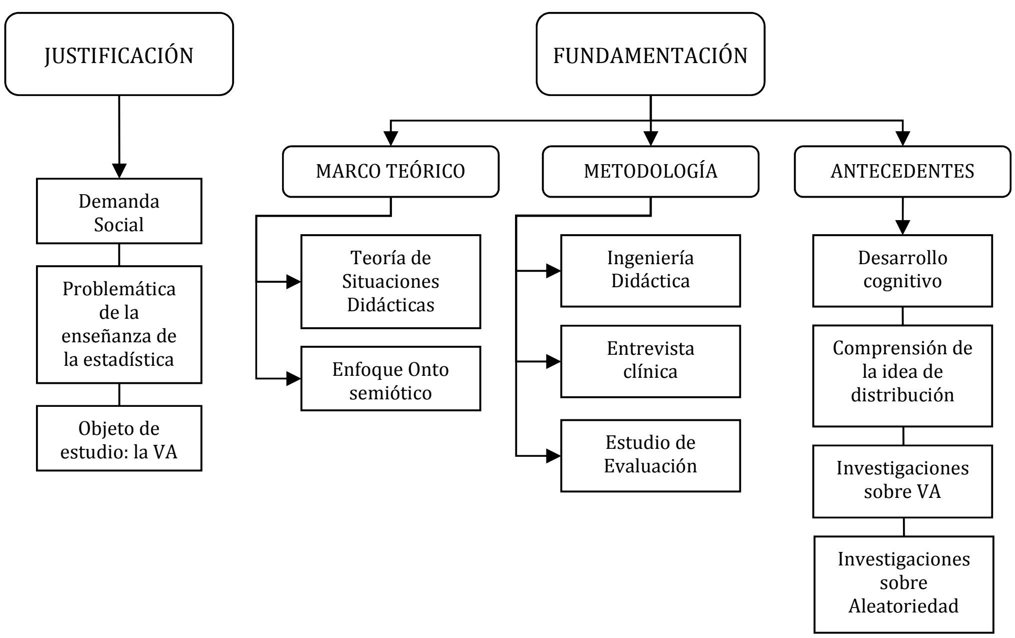


Figura 1.2. Puntos analizados en la justificación y fundamentación de la investigación

La pertinencia de este trabajo se justifica con base en la demanda social de una mejor enseñanza de la estadística en la escuela, los problemas particulares de la enseñanza de la estadística y la importancia de la variable aleatoria como una idea fundamental dentro del currículo escolar. La investigación se fundamenta en el marco teórico descrito y en los antecedentes que analizaremos en el Capítulo 2 organizados en los apartados que se muestran en la Figura 1.2.

Como ya se ha indicado, nuestra investigación consiste en la profundización del estudio de la didáctica de la variable aleatoria bajo las vertientes epistemológica y cognitiva. Los dos análisis realizados en cada una de ellas producen los cuatro estudios (Figura 1.3) que se reportan en esta memoria y que se describen a continuación.

- ❖ *Estudio 1.* Es de tipo epistemológico. Se analiza la variable aleatoria desde la disciplina, con la intención de indagar sobre su naturaleza e interrelaciones. Se presenta un análisis del objeto matemático, relacionándolo con la asignación de probabilidad y la actividad de modelación y delimitando, finalmente, los objetos relacionados con la variable aleatoria que serán tratados en los Estudios 3 y 4.
- ❖ *Estudio 2*. Es de tipo epistemológico. Se presenta un análisis del desarrollo histórico de la variable aleatoria con fines didácticos. Complementa el análisis epistemológico desde la disciplina y permitirá identificar algunas dificultades y obstáculos ligados a su desarrollo y que podrían presentarse en los estudiantes.
- ❖ *Estudio 3*. Consiste en una exploración cognitiva en profundidad de las concepciones y comprensión de una pareja de estudiantes sobre la variable aleatoria cuando se enfrentan a una situación de apuesta y en cuya resolución se pone en juego a la variable aleatoria, desde la concepción clásica de probabilidad. La metodología empleada es una entrevista clínica guiada.
- ❖ *Estudio 4*. Es de tipo cognitivo y trata de complementar el anterior en un doble sentido. En primer lugar ampliando el tamaño de la muestra, de modo que los resultados puedan tener mayor generalidad. Por otro lado, los estudiantes tienen el reto de organizar una investigación estadística completa (experimento, recogida de datos, análisis y producción de informes) para responder a una pregunta sobre sus propias intuiciones sobre el azar. Se pone en juego la variable aleatoria desde la concepción frecuencial. El primer paso de los estudiantes en vías del análisis ha de ser la construcción de la distribución. De este modo accederemos a la comprensión que manifiestan y las prácticas que realizan sobre la variable estadística y la relación que

establecen entre variable estadística y variable aleatoria. Se trata de un estudio de evaluación en un grupo, usando un proyecto abierto, que se analiza, en primer lugar con métodos cualitativos, para posteriormente elaborar tablas de frecuencias asociadas a diversas variables que analizaremos.

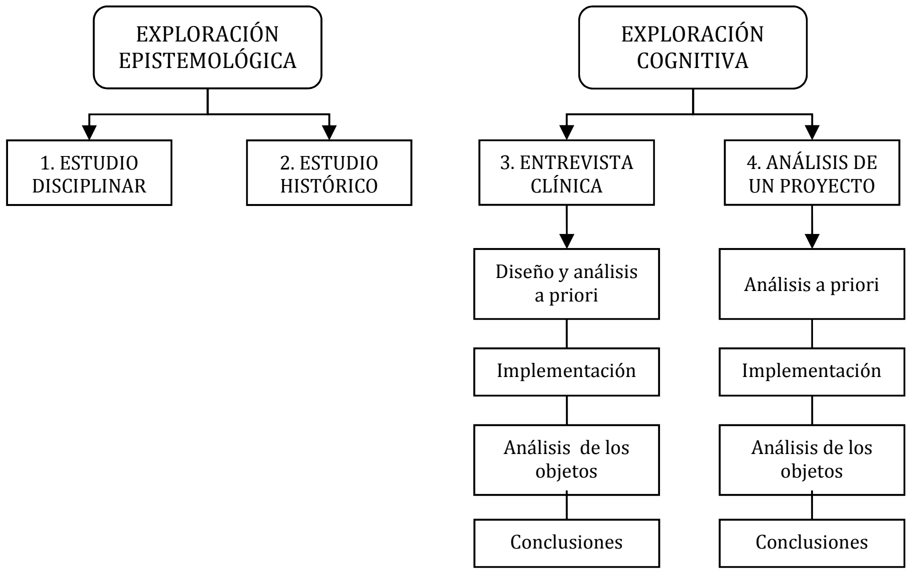


Figura 1.3. Estructura de los Estudios

Como señala Artigue (1995a), la profundización sobre el análisis preliminar del objeto matemático dentro de la ingeniería didáctica acepta el uso de metodologías externas que ayuden en la búsqueda de elementos tendientes a proporcionar información al análisis a priori y al diseño. Sin embargo, el objetivo explícito de las metodologías externas estará supeditado a proporcionar elementos que sean útiles a la ingeniería didáctica. La concordancia entre los sustentos teóricos de los elementos metodológicos usados y la teoría de situaciones didácticas es un aspecto que cuidamos y fundamentamos. En este trabajo nos auxiliamos de dos vertientes externas para el análisis de la dimensión cognitiva: por un lado en la *entrevista clínica* (Confrey, 1980 y Clement, 2000) y por otro el *análisis de contenido* de producciones escritas de los estudiantes.

En el siguiente apartado se describen los rasgos generales de la ingeniería didáctica y posteriormente se puntualiza cada una de las metodologías usadas en los siguientes apartados.

5.1. Ingeniería didáctica

Numerosos trabajos de investigación en didáctica –llamados *ingenierías didácticas*– se centran en la tarea de producir situaciones de aprendizaje destinadas a asegurar de manera controlada la emergencia de conceptos matemáticos en el contexto escolar. Tales situaciones abarcan varias fases descritas por Brousseau: fase de investigación o acción, fase de formulación, fase validación y fase de institucionalización.

La ingeniería didáctica ha sido propuesta como la metodología de investigación de la teoría de situaciones didácticas principalmente por Artigue (1995a), se aplica en *situación escolar,* su análisis es *cualitativo* y recurre a lo que llama *validación interna*. Sus formas de validación son básicamente cualitativas, aunque también usa métodos estadísticos cuando es conveniente, se basan en estudios de casos o de muestras y consiste, en esencia, en la confrontación entre el análisis *a priori* y *a posteriori*. La ingeniería didáctica, con respecto a su metodología experimental, está integrada por tres fases:

- *Análisis preliminar*. Es una investigación previa al planteamiento de una secuencia didáctica alrededor de un objeto matemático. Su objetivo es conocer más de cerca la naturaleza de este objeto desde la perspectiva didáctica, epistemológica y cognitiva con el propósito de identificar hipótesis sobre el proceso de construcción del objeto matemático por parte de los estudiantes en situación escolar, así como aportar elementos para el diseño de la secuencia didáctica.
- *Diseño de la secuencia y análisis a priori*. El diseño de la secuencia supone una vinculación de los aspectos epistemológicos, cognitivos y didácticos, así como una planeación de la forma en la que el estudiante construirá el conocimiento a partir de ella. La secuencia didáctica pretende desarrollar el proceso de construcción del objeto matemático por parte de los estudiantes. Los objetivos del análisis a priori son validar de la secuencia didáctica, poner a prueba las hipótesis que emergieron del análisis preliminar y aportar elementos de observación de la secuencia didáctica que de otro modo podrían pasar desapercibidos (obstáculos epistemológicos, dificultades vinculados con la problemática específica, etc.).
- *Análisis a posteriori y validación interna*. A la puesta en práctica de las secuencias didácticas le sigue un análisis a posteriori que se basa en el conjunto de datos recogidos a lo largo de la experimentación, a saber, las observaciones realizadas de las secuencias. La validación interna es la confrontación de este análisis con el a priori.

Aunque nuestra investigación se centrará únicamente en el análisis preliminar de esta metodología, recorremos los mismos puntos propuestos por Artigue para la ingeniería didáctica en los Estudios 3 y 4 (Figura 1.3). En cada uno de ellos se propone a los estudiantes una situación que permitirá la recolección de datos sobre su forma de abordar las situaciones propuestas. Se comienza con el análisis a priori de dichas situaciones y el establecimiento de hipótesis previas, sustentadas en la revisión bibliográfica. Se sigue con la experimentación de la situación y recogida de datos para finalmente confrontar con las hipótesis previas. En cada caso se finaliza con el análisis de la idoneidad de la situación, tanto desde el punto de vista de evaluación, como del punto de vista de aprendizaje. No obstante, las situaciones propuestas están aún lejos de constituir verdaderas situaciones a-didácticas probadas, consideramos que el estudio realizado con ellas permite sentar las bases de su mejora para la futura construcción de situaciones a-didácticas y didácticas sobre la variable aleatoria.

Análisis preliminar

Para Chevallard (1985) el objeto de estudio de la didáctica no puede ser visto como un fenómeno aislado, sino como un fenómeno sistémico. La didáctica se interesa en el juego que se realiza entre un *docente*, los *alumnos* y un *saber matemático*. Esta relación ternaria forma una relación didáctica, a la que Chevallard denomina *sistema didáctico* y que se sustenta en el triángulo epistemológico: *docente-saber matemático-alumno* y sus relaciones. Así el objeto de estudio de la didáctica es el sistema didáctico, y más ampliamente, el *sistema de enseñanza*, que reúne los sistemas didácticos y se ocupa de todo lo que la sociedad organiza a través de instituciones para el funcionamiento de dichos sistemas didácticos.

El análisis preliminar dentro de situaciones didácticas está sustentado en el sistema didáctico. En él se analizan y se detectan aquellas restricciones o vínculos del conocimiento matemático a partir de distinguir tres dimensiones que surgen del sistema didáctico:

- *Componente epistemológica*. Ésta se refiere a los aspectos relacionados con el conocimiento mismo. Pueden considerarse su emergencia histórica, estado actual o su referencia teórica. En esta investigación, nos concentraremos en el estudio histórico y el estudio disciplinar correspondiente a una introducción elemental a la variable aleatoria en los cursos universitarios de especialidades donde la estadística es sólo una herramienta instrumental[3]. Se

[3] Es decir, no tratamos el estudio disciplinar en especialidades avanzadas como matemáticas o estadística.

trataría de determinar el significado de referencia del objeto matemático para nuestra investigación.

- *Componente cognitiva*. Hace referencia a las concepciones de los estudiantes, a las dificultades y obstáculos, sus prácticas personales en la resolución de problemas, que determinan la evolución de su aprendizaje. Llevamos a cabo una primera aproximación a estas componentes utilizando dos situaciones diferentes: un problema, dirigido en el contexto de una entrevista clínica y un proyecto abierto planteado a un grupo de estudiantes universitarios dentro de una situación normal de clase.
- *Componente didáctica.* Se consideran los elementos didácticos tales como profesor, libros de texto, currículo, método de enseñanza, etc. En nuestro estudio, no consideramos este componente.

5.2. Entrevista clínica

Esta técnica es considerada una herramienta metodológica que proporciona información sobre los procesos cognitivos de los estudiantes a través del cuestionamiento sobre sus concepciones por parte de un profesor (o investigador). Se enfoca al estudio de los procesos de razonamiento y de la formación de estructuras del conocimiento. El primer estudio cognitivo (Estudio 3) de nuestra investigación usará esta técnica.

La entrevista clínica ha sido muy utilizada en investigaciones psicológicas y sobre resolución de problemas. Piaget (1991/1950) fue uno de sus pioneros. Tiene una fuerte connotación constructivista porque supone que los procesos de razonamiento y estructuras de conocimiento que siguen las personas no son los mismos que aquellos seguidos por los académicos (Clement, 2000). Es una herramienta exploratoria que propone el descubrimiento de lo que ocurre en la mente de los estudiantes, no de una forma estática, sino también con la intención de modificar la estructura de conocimiento del estudiante. Desde esta perspectiva, la entrevista clínica es una especie de «enseñanza experimento» cuya finalidad no sólo estará en proporcionar al investigador (o al profesor) una idea adecuada de «donde está» el estudiante, sino también una idea adecuada de «la dirección que debe tomar». Esto proveerá un modelo hipotético de la forma del razonamiento del estudiante complementado con otros estudios.

Para Confrey (1980) en las entrevistas clínicas se busca modelar la *expresión* (verbalización) de un estudiante a través de la *perspectiva* de un conocedor mejor informado, es decir, el modelo de las nociones del entrevistado será construido a partir de elementos conceptuales del entrevistador. Pero la forma de indagación provoca que también se modifique la *perspectiva* del investigador sobre el mismo conocimiento. Ambas, la expresión (del estudiante) y la perspectiva (del profesor-investigador), según Confrey «aportan importante contenido epistemológico a la interacción enseñanza-aprendizaje» (1980, p 40). Puesto que no sólo proporciona al investigador un conocimiento sobre la forma en que piensa el estudiante sino también un nuevo conocimiento sobre el objeto matemático o problema sobre el que se está trabajando.

Hay una gran variedad de formas de llevar a cabo una entrevista clínica, desde entrevistas con preguntas abiertas hasta resolución de problemas en voz alta. Clement (2000) analiza la diversidad de entrevistas clínicas guiado por el protocolo de análisis que se plantee. En particular, en el Estudio 3 de nuestra investigación se pretende conocer la forma en que dos estudiantes, resuelven en «voz» alta una actividad propuesta sin ayuda del profesor, pero en presencia de él. En este caso el papel del profesor/investigador únicamente es intervenir para aclarar o profundizar, a sí mismo o a la investigación, las ideas que los estudiantes expresan. Los estudiantes resuelven la actividad en consenso, expresando sus ideas en voz alta y argumentando sus posiciones.

De acuerdo con la clasificación de Clement (2000), nuestra entrevista se sitúa en un estudio *generativo/interpretativo* en donde se obtienen nuevas categorías de observación y nuevos elementos de modelación de acuerdo a las respuestas que se vayan obteniendo por parte de los estudiantes. Sus resultados serán la formulación de variables relevantes de observación y modelos explicativos de procesos cognitivos basándose en un protocolo más o menos abierto y algunas veces en hipótesis generales y/o variables de observación a grandes rasgos que servirán para guiar y dirigir el estudio, sin que ello limite su interpretación. Son apropiados para tópicos poco estudiados (como lo es la didáctica de la variable aleatoria) o para los que se tiene poca información para establecer una teoría previa puesto que se concentran en la viabilidad y relevancia del modelo de acuerdo a las observaciones detectadas por el profesor/investigador y sólo prestan atención a aspectos no formales de fiabilidad. La generalización de los modelos producidos por nuestro estudio será hipotética y sólo será válida en escasos contextos y algunos individuos. En este tipo de entrevista, es necesario reproducir o confirmar las conclusiones así como ampliar la profundización sobre los aspectos a investigar.

5.3. Análisis de contenido en producciones escritas

El análisis de contenido ha sido una técnica muy usada en educación para evaluar las producciones escritas por los estudiantes y de alguna manera también para realizar estudios que requieran el análisis de documentos, como los estudios epistemológicos. En nuestra investigación la usaremos tanto en el análisis de las producciones escritas del Estudio cognitivo 4 como en los estudios epistemológicos.

En el Estudio 4 se realizó un análisis cualitativo de las producciones escritas de una muestra de estudiantes en un Proyecto abierto de análisis de datos. Cada uno de estos estudiantes resolvió el problema propuesto a través de un proyecto y elaboró un reporte en donde se les pidió que plasmaran su método solución y respondieran a una pregunta que debió guiar todo su análisis. La metodología empleada para el análisis de estos reportes se engloba en el *análisis de contenido* que permite evaluar constructos cognitivos de su autor -sus ideas, concepciones, conocimientos- mediante el análisis de su discurso escrito. Se basa en el hecho de que las unidades de análisis en que se divide el texto pueden clasificarse en un número reducido de categorías (Weber, 1985). Sirve para efectuar inferencias sobre el constructo de interés mediante la identificación sistemática y objetiva de las características específicas del texto (Ghiglione y Matalón, 1989).

En nuestro trabajo, los constructos de interés son las prácticas de los estudiantes relacionadas con la variable aleatoria y estadística (como prácticas observables) y el significado personal que los estudiantes asignan a dichas variables, donde podemos englobar sus concepciones, las propiedades que le asignan, argumentos que emplean, representaciones y procedimientos. Aunque dicho significado como constructo es inobservable, podemos inferirlo al menos parcialmente de las prácticas observables mediante el análisis de contenido.

En el estudio 4 se han seguido las técnicas lógico-semánticas del análisis de contenido temático, donde se recurren a la lógica personal, y al conocimiento previo sobre el tema adquirido a través de la revisión bibliográfica y el estudio histórico previo, para resumir el texto escrito de los alumnos, definir variables y categorías y verificar la validez de las mismas (Ghiglione y Matalón, 1989). El análisis permitió definir variables, tales como:

- Nivel del gráfico mediante el cual el alumno representa una distribución de frecuencias (variable ordinal con 4 niveles).
- Cálculo de determinados estadísticos, por ejemplo, la media. Esta variable toma los valores: calcula correctamente, calcula en forma parcialmente correcta, no calcula.
- Interpretación que se hace de los gráficos y estadísticos calculados. Variable que toma los siguientes valores: interpreta correctamente, en forma parcialmente correcta, o no interpreta

- ❖ Conclusión obtenida sobre el problema planteado: concluye correctamente, conclusión parcial o no concluye/concluye incorrectamente.

De estas variables se han analizado cualitativamente las diferentes categorías presentando un ejemplo de uso por parte de los alumnos y adicionalmente se presentan tablas de frecuencias para informar de la proporción de estudiantes en cada categoría.

El marco teórico del enfoque onto-semiótico nos proporciona una categorización de entidades matemáticas que aportan orientación y concreción al análisis de contenido. El objetivo es la caracterización de los significados personales de los estudiantes puestos de manifiesto en el proyecto. Siendo un análisis de datos cualitativo, el proceso se realizó en varias etapas dividiendo el texto en unidades, que fueron clasificadas en categorías y a continuación ensambladas y relacionadas para conseguir un todo coherente. El proceso se dividió en tres etapas: la reducción de datos, el análisis de datos y la obtención y verificación de conclusiones (Huberman y Miles, 1994):

- ❖ La primera operación fue la separación de segmentos o unidades de análisis, en varios niveles. Siguiendo este criterio, una vez recogidos los protocolos, una lectura detallada de los mismos permitió seleccionar los párrafos, estadísticos o gráficos, relacionados con cada variable analizada.
- ❖ Para cada uno de estos fragmentos se compararon los clasificados en un mismo apartado para cada categoría, depurándose la clasificación.
- ❖ Se elaboraron tablas recogiendo los resultados del análisis de cada variable para poder analizar estos resultados para la extracción de conclusiones.
- ❖ Se elaboró el informe de análisis incluyendo ejemplos de las diferentes categorías, elaboradas mediante un proceso interpretativo guiado por el análisis epistemológico de los conceptos intervinientes y por los resultados de las investigaciones previas.

Capítulo 2.
Antecedentes

ÍNDICE DE CAPÍTULO

1. Introducción

En este capítulo se describen los antecedentes sobre los que se basa nuestro trabajo. Aunque la investigación sobre la didáctica de la probabilidad es muy abundante (Borovcnik, Bentz y Kapadia, 1991a y b; Shaughnessy, 1992, 2007; Jones 2005; Jones, Langrall y Mooney, 2007), hay poca literatura de investigaciones centradas específicamente en la didáctica de la variable aleatoria. Sin embargo, son muy abundantes las que tratan sobre cuestiones vinculadas con nuestro tema de interés, como estudios sobre el desarrollo de la idea de distribución, la enseñanza y el aprendizaje de las distribuciones de probabilidad (en particular de la Normal) y la idea de aleatoriedad. De todos estos últimos estudios, revisaremos exclusivamente los que son más necesarios para introducirnos en el estudio de la enseñanza-aprendizaje de la variable aleatoria.

Como punto de inicio, en este apartado describiremos brevemente las investigaciones de Piaget e Inhelder sobre el desarrollo de la idea de distribución en los niños, pues esta investigación contempla los modelos de las variables aleatorias con las distribuciones Normal y de Poisson y proporciona pautas sobre las concepciones de los individuos en diversos estados de desarrollo. Estos autores, además de ser pioneros en la investigación en didáctica de la probabilidad, también hicieron importantes aportaciones que sirvieron de base para el desarrollo de la teoría de situaciones didácticas que hemos incluido en el marco teórico que usamos en esta tesis (Capítulo 1).

A continuación analizaremos investigaciones más recientes que se centran en la comprensión de la idea de distribución, tanto de la variable estadística, como de la variable aleatoria. La revisión sobre los resultados de investigación en didáctica de la variable estadística proporcionará, por un lado, fundamento al segundo estudio cognitivo (Capítulo 6), en donde analizamos la comprensión de los estudiantes sobre algunas variables estadísticas, sus distribuciones y su relación con algunas variables aleatorias. También, por otro lado, es necesario ahondar sobre su estudio puesto que existe una relación conceptual íntima entre las variables estadísticas y aleatorias (como se describirá en el Capítulo 3), que se refleja en la vertiente cognitiva.

Después nos concentraremos en aquellos autores que, consideramos, se han interesado directamente en el estudio de la enseñanza o el aprendizaje de la variable aleatoria o que han obtenido resultados en esta área aunque su investigación no se centre necesariamente en ella. Eso permitirá tener un panorama del estado actual de las investigaciones que aportan resultados sobre la didáctica de la variable aleatoria. Esta parte de la revisión bibliográfica se analiza y puntualiza de acuerdo a tres aspectos que nos parecen de interés: el didáctico, epistemológico y el cognitivo.

Para finalizar resumimos algunos estudios sobre la percepción de la aleatoriedad en sujetos adultos, pues, por un lado, el proyecto cuya solución analizamos en el Capítulo 6 está basado en una situación experimental inspirada en los experimentos realizados en dichas investigaciones, y, por otro lado, los resultados de nuestros dos estudios cognitivos muestran las dificultades de los estudiantes con el concepto de aleatoriedad, por lo que las investigaciones relacionadas proporcionan un marco explicativo para dichas dificultades.

2. Estudios de Piaget sobre el desarrollo evolutivo de la idea de distribución

El primer estudio conocido sobre la variable aleatoria y su distribución fue realizado por Piaget e Inhelder (1951) y es parte de su investigación sobre la comprensión del razonamiento probabilístico en los niños. Acorde con la teoría de desarrollo de razonamiento en estadios de Piaget, ambos autores son de los principales precursores de los estudios sobre el desarrollo evolutivo del razonamiento probabilístico y uno de los más completos en esta materia, pues incluye el estudio de la forma en que el niño percibe la aleatoriedad, la probabilidad simple y compuesta, su capacidad combinatoria y de comparación de probabilidades, así como las ideas de distribución, convergencia y correlación.

No entraremos en el detalle de todos sus estudios, sólo dedicaremos unos párrafos a la concepción de Piaget sobre la aleatoriedad y su comprensión, pues ello da explicaciones para comprender los experimentos relacionados con la evolución de la idea de distribución y su desarrollo.

2.1. La idea de aleatoriedad según Piaget

Piaget e Inhelder consideraron que el desarrollo cognitivo del niño progresa en diferentes etapas cuya transición entre ellas se da a través de procesos de *asimilación* y *acomodación*. El orden en que las etapas se presentan no varía, pero sí la edad en la que cada niño las alcanza y también la edad en que las alcanza para diversos conceptos matemáticos. Para Piaget el desarrollo cognitivo consiste en un proceso de transformación secuencial durante el que los esquemas sensorio-motores elementales son reestructurados en estructuras cognitivas progresivamente más complejas, inicialmente las estructuras pre-operativas y después las operativas, en donde las operaciones lógicas y aritméticas se podrán componer entre sí de modo riguroso y siempre *reversible*. La reversibilidad (siempre se puede volver al estado primitivo mediante una operación inversa, como la resta para el caso de la suma) en un nivel abstracto es lo que hace posible la deducción. Las estructuras cognitivas operativas constituyen el prerrequisito más importante para la emergencia de conceptos lógico-matemáticos porque el niño es capaz de reconstruir las relaciones cuantitativas en un problema únicamente por medio de operaciones lógicas. Piaget e Inhelder también aplicaron estos principios a la construcción del concepto de probabilidad por los niños. En este campo las tres etapas de desarrollo relevantes son: la pre operacional, de operaciones concretas y de operaciones formales.

- ❖ *En el estadio pre operacional* (2-7 años) el niño organiza el espacio situando y desplazando los objetos (dentro/fuera, encima/debajo); descubre y compara propiedades físicas de los objetos que manipula; diferencia colecciones numéricas de pocos elementos; contrasta magnitudes por comparación y estima a partir de una cantidad la longitud, volumen y peso; ordena en el tiempo; engloba aspectos de tipo espacial; se inicia en el conteo; trabaja con una sola cantidad y resuelve problemas de cambio sencillo. Sin embargo en esta etapa no efectúa operaciones reversibles.
- ❖ *El periodo de las operaciones concretas* (7-11) se caracteriza por la aparición de operaciones reversibles con la adquisición de principios de conservación de cantidad, peso y volumen; agrupa objetos en función de propiedades aditivas o multiplicativas; ordena elementos en función de la cualidad que varía; adquiere la noción de sistema de numeración y de operación con números y orden de sucesión de los objetos en el espacio. Las operaciones concretas se realizan sobre lo actual, presente u observable.
- ❖ *El periodo de operaciones formales* (11-15) constituye el último del desarrollo intelectual del hombre. A este periodo se le atribuye la máxima importancia en el desarrollo de los procesos cognitivos y sociales. En él se contempla lo real como parte de lo posible; se depura el pensamiento proposicional; se desarrollan y amplían las posibilidades combinatorias de pensamiento y se da la integración de formas de pensamiento reversible. Las operaciones sobre lo posible o potencial permiten el pensamiento hipotético deductivo (Inhelder y Piaget, 1955).

Piaget e Inhelder (1951) aseguran que la comprensión de la aleatoriedad para el niño es complementaria a la de la relación causa-efecto. De acuerdo con ellos, el azar es el resultado de la interferencia de una serie de causas que actúan independientemente una de otra produciendo resultados inesperados o impredecibles. Por lo tanto, las transformaciones aleatorias *no pueden componerse en forma rigurosa* y son esencialmente *irreversibles*. Así, estos dos autores conciben el azar como complementario a la composición lógica de las operaciones reversibles, por lo que sugieren que el azar no puede ser comprendido hasta que se adquiere un dominio de las operaciones reversibles, y se le compare con ellas.

Piaget e Inhelder concluyeron que no había una intuición innata de la aleatoriedad en el niño. Ya que en el período pre operacional el niño no tiene un pensamiento reversible, no podría entender la aleatoriedad porque no podría diferenciar entre acontecimientos reversibles y aleatorios originados por mezclas de causas irreversibles, además de que no comprende bien la relación causa-efecto. Sin embargo una de las características funcionales que diferencian el pensamiento formal del pensamiento concreto en la teoría piagetiana es la subordinación de lo «real» a lo «posible», por lo tanto una estructura cognitiva combinatoria daría al niño la posibilidad de abstraer hacia lo «posible» y despegarse de su mera experiencia «real». Es decir, se requiere la posesión de un esquema combinatorio para poder concebir las distintas posibilidades existentes en estas situaciones (Batanero, 2001).

En consecuencia, Piaget e Inhelder (1951) supusieron que la idea de aleatoriedad comenzaría a aparecer alrededor de los 7 años de edad en la etapa de las operaciones concretas, en donde ya hay cierta apreciación de los factores que caracterizan a los fenómenos causales. En esa etapa el niño iniciaría la comprensión de la interacción de cadenas causales que conducen a sucesos impredecibles y la irreversibilidad de los fenómenos aleatorios. Sin embargo, ya sea porque comprende la interferencia de las causas sin reconocer su independencia, o porque comprende la independencia y no la interferencia, no llega a completar el proceso. Durante este período, el pensamiento está todavía muy ligado al nivel concreto y carece de un razonamiento combinatorio como para que comprenda más profundamente la aleatoriedad.

En cambio, según Piaget e Inhelder, el adolescente (etapa de operaciones formales) ya es capaz de relacionar los resultados posibles de los fenómenos aleatorios según esquemas operacionales, en particular el combinatorio. Una vez que una situación aleatoria se presenta, el uso de estos esquemas la hace inteligible, y la síntesis entre el azar y lo operacional conduce al adolescente al concepto de probabilidad. Esto es, los casos aislados son imprevisibles pero el conjunto de posibilidades puede determinarse mediante un razonamiento de tipo combinatorio, con lo que la situación se vuelve previsible. Esta es la vía por la que aparece la idea de probabilidad expresada por la razón entre las posibilidades de un caso y del conjunto de posibilidades (Pérez Echeverría, 1990).

2.2. Distribuciones normales

El trabajo de Piaget e Inhelder (1951) indica que la comprensión progresiva de la aleatoriedad conduce a la de la distribución de una serie de resultados en un experimento aleatorio (que para nosotros sería la distribución de una variable aleatoria). En la primera parte de su libro reportan sus estudios sobre la aparición del concepto de azar físico, en donde enfrentaron a los niños con tareas de mezclas irreversibles y sobre distribuciones simétricas y uniformes en las que introducen el azar de modo progresivo en situaciones espacio-temporales. En la segunda parte estudian el azar y la probabilidad desde la perspectiva lógico-matemática e incluso llegan a pedir la cuantificación de las probabilidades. Todas las experiencias aleatorias comportaban esquemas de simetría.

Para estudiar las concepciones de los niños con respecto a la distribución y convergencia, Piaget e Inhelder realizaron diversas experiencias en las que les preguntaron a los niños por las posiciones finales de distintos experimentos aleatorios. Una de ellas, de la que nos ocuparemos en este apartado, consistió en dejar caer bolitas a través de un pequeño orificio en dispositivos muy similares al aparato de Galton (Ver Figura 2.1).

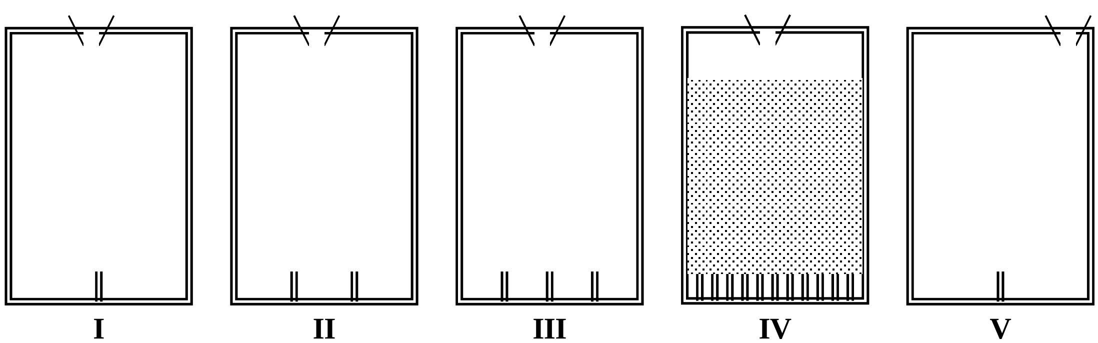


Figura 2.1. Esquema de los aparatos de Galton utilizados en las experiencias de Piaget

El aparato de Galton tiene la forma de un plano inclinado de madera donde se colocan unos clavos regularmente espaciados. Cuando se dejan caer bolitas por el orificio superior, éstas se dispersan al azar y se recogen en unas casillas colocadas al final del plano inclinado. Las casillas centrales deben recibir más bolitas que las casillas situadas en los extremos de manera simétrica. Piaget diseñó cinco aparatos (Figura 2.1), de los cuales cuatro tenían la abertura superior en la parte central y diferente número de casillas: 2, 3, 4 y un número grande. El aparato V, con abertura en el extremo superior derecho, sólo tenía dos casillas.

Haciendo uso de estos aparatos y a partir de las respuestas de los niños en entrevistas clínicas, los investigadores analizaron si éstos llegaban a comprender el mecanismo de convergencia, que hará que después de algún tiempo, la distribución de bolitas tome la forma de la distribución normal (que ellos llamaron *centrada*). Para Piaget e Inhelder esta comprensión implicaba captar la simetría de las trayectorias posibles de las bolas con respecto al eje vertical cuando éstas caían por el orificio. Batanero (2001) y Tauber (2001) explican que esta simetría es debida a que cada bolita tiene la misma probabilidad de orientarse hacia la derecha o hacia la izquierda, hacia adelante o hacia atrás, por lo que la distribución toma una forma regular, cuya representación matemática sería la curva de Gauss.

Durante las entrevistas, en primera instancia se pedía a un niño que hiciera una conjetura sobre la casilla en la que quedaría una de las bolitas si se le dejaba caer por el orificio de una de las cajas. También se le pedía la razón de su respuesta. Después el experimento se realizaba físicamente con unas cuantas bolitas y se le pedía al niño que conjeturara sobre la forma que tomaría la distribución cuando se dejaran caer un número grande de bolitas (alrededor de 60). Al final se dejaban caer las bolitas y se le pedía al niño que interpretara la distribución obtenida.

El primer estadio (hasta 7 años) en la caja I los niños generalmente apostaron a favor de una de las dos casillas, pero no conjeturaron que el número de bolas se igualará en las dos casillas cuando se dejan caer muchas bolitas. En la caja II los niños prevén una distribución igual en las tres casillas o piensan que todas las bolas quedarán en una de las casillas. Con la caja III los niños centran su apuesta en uno de las casillas centrales o por una distribución irregular. En la caja IV, espera una distribución irregular. En resumen, no hay presencia de la idea de distribución en esta edad.

Los niños de entre 7 y 11 años fueron capaces de predecir la desigualdad entre las frecuencias de bolitas que caerán en las casillas laterales y centrales; pero, aunque predijeron la simetría global, no comprendieron la equivalencia de frecuencias esperadas entre las casillas equidistantes del centro. En la caja I predijeron una igualdad aproximada entre las dos casillas, pero sin concebir que esta igualdad se consiguiera progresivamente cuando se incrementa el número de bolitas. En la caja II previeron un máximo central, pero no la igualdad en las casillas laterales. En la caja III llegaron a la conjetura correcta de mayor acumulación en las casillas centrales, pero no previeron equivalencia entre en ellas ni entre las laterales. En la caja IV no previeron una distribución simétrica regular, pero comenzó a haber una generalización de las experiencias anteriores sobre la configuración de conjunto. En suma, el segundo estadio se caracterizó por el reconocimiento de un principio de distribución de conjunto generalizable.

Los niños mayores de 12 años sí previeron una equivalencia entre las partes simétricas correspondientes de la dispersión. Es decir, el tercer estadio (a partir de los 12 años) estuvo marcado por la cuantificación de la distribución de conjunto. En la caja IV, sin embargo, fueron necesarios varios ensayos de graduación hasta descubrir la distribución en forma de campana. El progreso más notable fue la comprensión del papel de los grandes números en la regularidad de la distribución.

2.3. Distribución de Poisson

Las investigaciones de Piaget e Inhelder (1951) sobre la comprensión de lo que ellos llamaron *distribuciones uniformes*, se basaron en la experiencia de los niños en la observación de la distribución de las primeras gotas de lluvia o copos de nieve sobre un piso embaldosado cuando empieza una lluvia o nevada. Piaget e Inhelder usaron una hoja de papel blanco dividida en cuadros de 2 ó 3 cm como representación del embaldosado y sobre ella pidieron a los niños que simularan las gotas de lluvia o copos de nieve con fichas que dejaban caer al azar o con puntos que dibujaban sobre ella. Otra experiencia fue directamente sobre un piso embaldosado, en donde se les pidió a los niños que colocaran piedrecitas sobre el piso en el lugar donde ellos pensaban que caería la siguiente gota o copo cuando empezara a llover o nevar. Piaget e Inhelder comenzaron con unos pocos puntos y aumentaron su número progresivamente para ver la distribución global en el plano (Ver Figura 2.2) que Batanero (2001) identifica como una distribución Poisson en el plano.

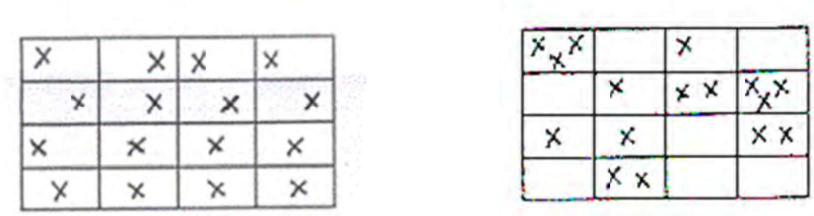


Figura 2.2. Dibujos o disposiciones de puntos en la experiencia del embaldosado

Los niños pequeños (etapa pre operacional) supusieron que las gotas de lluvia se repartirían por toda la superficie y que se distribuirían regularmente en los cuadros. En el experimento, los niños colocaron las piedrecitas o puntos sistemáticamente, rellenando uno a uno todos los cuadros antes de repetir las piedrecitas o puntos en alguno de ellos. Si la retícula tenía un solo cuadro vacío, los niños colocaron la marca o piedra en el cuadro vacío para completar un patrón uniforme. Piaget e Inhelder interpretaron este comportamiento como falta de comprensión de la aleatoriedad en esa etapa.

Los niños con edad situada en el periodo de las operaciones concretas aceptaron la irregularidad de la distribución, aunque si había un solo cuadro sin marca, el cuadrado «seco» se consideraba como el más probable para recibir la siguiente gota o copo. Sin embargo, los puntos se colocaron desordenados dentro de cada cuadrado, mientras en la etapa anterior se colocaban exactamente centrados.

En el periodo de las operaciones formales los niños usaron el mecanismo de la convergencia progresiva. Entre más gotas haya, la diferencia en el número de gotas por baldosa tiende a disminuir más, no en forma absoluta, sino en forma relativa, es decir, en promedio. De acuerdo con Piaget e Inhelder, la ley de los grandes números se comprende por su doble aspecto combinatorio y proporcional. Es decir, en este estadio (12 años o más) también aparece el razonamiento proporcional y por lo tanto, los niños comprenden la ley de los grandes números.

En resumen, las investigaciones de Piaget e Inhelder sobre el desarrollo del razonamiento probabilista en el niño, suponen que la base para la adquisición de este razonamiento es la construcción de relaciones lógicas tales como causa y efecto, porque ello permitirá la comprensión de la diferencia entre los sucesos necesarios (y por tanto predecibles) y los aleatorios (no predecibles). Pero también el desarrollo este tipo de razonamiento requiere construir y representar la totalidad de los sucesos (razonamiento combinatorio) y tener en cuenta múltiples relaciones para calcular proporciones. Piaget e Inhelder también suponen que estas construcciones se adquieren paulatinamente a diferentes edades, culminando en la adolescencia con el dominio de las operaciones concretas. En particular concluyen que en esa etapa se tendrá una buena percepción de la aleatoriedad y de las distribuciones uniformes y normales. Estas conclusiones han sido contradichas, indicando que muchos adultos tienen sesgos respecto a su concepción sobre los experimentos aleatorios y también que los niños tienen intuiciones sobre la aleatoriedad. Sin embargo la obra de Piaget e Inhelder no puede ser pasada por alto, los aportes al razonamiento probabilista han sido de gran utilidad en otras investigaciones con adultos sobre la percepción de la aleatoriedad, además de que las tareas experimentales que usaron son muy ingeniosas y se han incorporado y mejorado en otras muchas investigaciones.

3. Investigaciones sobre la comprensión de la idea de distribución

Al contrario de lo que hemos indicado sobre el reducido número de trabajos centrados explícitamente en la variable aleatoria, el número de trabajos relacionados con la idea de distribución es grande y en constante crecimiento, a tal punto que la conferencia STRL-4 (*Statistical Thinking, Reasoning and Literacy Research Forum*: srtl.stat.auckland.ac.nz/srtl4/) estuvo dedicada al tema «Razonamiento sobre la distribución» y el volumen 4, número 1 de la revista *Statistics Education Research Journal* fue asignado como un número monográfico alrededor de la idea de distribución.

Shaughnessy (2007) indica que uno de los fines básicos de la estadística es desarrollar la capacidad de leer, analizar, criticar y hacer inferencias a partir de distribuciones de datos. Frecuentemente las investigaciones no distinguen entre la distribución de una variable aleatoria o estadística, Shaughnessy, sin embargo, sugiere que el concepto de distribución es múltiple, pues se puede referir a distribución de datos (variable estadística), distribución de probabilidad (variable aleatoria) y distribución muestral (distribución del estadístico en el muestreo, que también es una variable aleatoria). Shaughnessy también menciona que muchas investigaciones que analizan otros conceptos, tienen como móvil subyacente la comprensión de la idea de distribución en los estudiantes, puesto que dicha idea incluye las de posición central y variabilidad o dispersión. Reading y Canada (2011) añaden las ideas de probabilidad, proporcionalidad, forma, frecuencia y simetría. Todas estas ideas están interrelacionadas en una tabla o gráfica puesto que se suele referir a las distribuciones a través de sus representaciones tabular o gráfica (Shaughnessy, 2007).

El razonamiento distribucional también implica relacionar los datos muestrales (variable estadística) con los de la población (variable aleatoria) y transitar entre una y otra, así como comprender la diferencia entre los parámetros (vinculados con la población) y los estadísticos (vinculados con las muestras). Esto es lo que permite hacer inferencias y predicciones y requiere de un tránsito forzoso no sólo entre la distribución de una variable estadística y una aleatoria sino también la relación que este tránsito guarda con las distribuciones muestrales (otro tipo de variable aleatoria).

En su mayoría, las investigaciones educativas centradas en la idea de distribución se refieren a la distribución de una variable estadística. En este punto recogeremos, muy escuetamente, algunos de los principales trabajos y que también se reportan en Shaughnessy (2007) y Reading y Canada (2011).

3.1. Distribución de variables estadísticas

Bakker y Gravemeijer (2004) indican que una característica esencial del análisis estadístico es que trata sobre todo de describir y predecir propiedades de los agregados de datos y no de cada dato aislado. Un problema para conseguirlo es que los estudiantes a veces no ven el dato (por ejemplo altura de un estudiante) como un valor de una variable (altura en el ejemplo), sino como una característica personal de un individuo. Los estudiantes se centran en los datos aislados que usan para calcular, por ejemplo, la media o el rango, pero no siempre ven estos estadísticos (media o rango) como una propiedad del conjunto de datos como un todo (distribución). En su estudio, estos autores analizaron específicamente el razonamiento de los estudiantes para pasar de la idea de dato aislado a la de distribución como un todo cuando se les puso una tarea en la que tuvieron que razonar y comparar conjuntos de datos utilizando un software específico que los ayudó a transitar de los datos a la distribución y viceversa. Su experiencia muestra que los estudiantes pueden razonar sobre algunos aspectos de la distribución y llegar a desarrollar aspectos más avanzados de la misma como frecuencia, clase (intervalo), dispersión, mediana o densidad, si se utiliza un enfoque apropiado.

Shaughnessy y Ciancetta (2002) presentan ejemplos de lo que piensan los estudiantes sobre los promedios, forma y dispersión al razonar sobre las distribuciones de datos. Por ejemplo, algunos de los estudiantes que se centraron demasiado en los valores centrales, son incapaces de pasar de estos valores a la idea de distribución. La dificultad de generar una distribución de datos a partir de un promedio dado, también ha sido descrita en muchas investigaciones sobre promedios, donde al dar un promedio y pedir a los estudiantes que describan una distribución correspondiente al promedio, los estudiantes se limitan a repetir valores muy similares o equidistantes del promedio.

Otra noción importante para comprender las ideas de variable aleatoria y distribución es la idea de variabilidad, que está siempre presente en los datos y tiene múltiples significados en estadística (Reading y Shaughnessy, 2004), entre otros los siguientes: variabilidad de resultados en un experimento aleatorio; variabilidad en los datos; variabilidad en una variable aleatoria; variabilidad en las muestras o la distribución muestral. Es por ello que dos fines importantes de la enseñanza de la estadística es que los estudiantes perciban la variabilidad y manejen modelos que permitan controlarla y predecirla.

Uno de los análisis más frecuentes en los usos de la estadística es la comparación de dos distribuciones, ya sea de la misma variable estadística en dos conjuntos de datos o de dos variables estadísticas en el mismo conjunto de datos. Es por ello que algunos autores han descrito el razonamiento de los estudiantes alrededor de la idea de distribución cuando comparan dos distribuciones de datos e indican que los estudiantes no suelen usar la media para comparar conjuntos de datos, sino que más bien usan parte de los datos, por ejemplo, suelen comparan los valores máximos o las longitudes del rango. Un ejemplo lo tenemos en los resultados obtenidos por Konold, Pollatsek, Well, and Gagnon (1997) quienes trataron de detectar las barreras críticas que encuentran los estudiantes para analizar los datos. A pesar de que formalmente estas comparaciones se llevan a cabo a partir de los promedios y la dispersión de las distribuciones, Konold et al. encontraron que ni siquiera el uso de la media como representante para comparar dos conjuntos de datos es intuitivo para algunos estudiantes, quienes, en su investigación, se centraron en las frecuencias absolutas y no en las relativas. Incluso cuando las muestras eran de tamaño muy diferente, la comparación la realizaron sobre las frecuencias de los valores de la variable que coincidían en ambos grupos. Konold, et al. concluyen que el problema tiene sus orígenes en que los estudiantes no han dado el paso de comparar los valores de la variable como propiedades de los individuos aislados a comparar las propiedades de los conjuntos de datos.

También Batanero, Estepa y Godino (1997) estudiaron las estrategias seguidas por los estudiantes al comparar dos distribuciones como parte de un estudio sobre asociación. Como estrategias correctas encontraron comparar medias o totales en las dos distribuciones o reducir el conjunto de datos a una sola variable restando los valores correspondientes en el caso de muestras relacionadas y luego comparar la media o el total con cero. También encontraron estrategias incorrectas como comparar solo valores aislados en las dos distribuciones o esperar un aumento/disminución similar en todos los casos para las muestras relacionadas.

En una serie de estudios longitudinales, Watson analizó las estrategias utilizadas por estudiantes de primaria y secundaria al comparar conjuntos de datos presentados en forma gráfica (Watson y Moritz, 1999; Watson, 2001). En sus trabajos, clasificó estas estrategias de acuerdo al modelo SOLO[1], que es un modelo jerárquico, en tres niveles cognitivos, que a su vez, se enmarcaron en dos ciclos por la dificultad de la tarea. Los niveles cognitivos denotaron el tipo de razonamiento usado por los estudiantes para resolver las tareas dadas: (a) uso de algún aspecto importante de la tarea (Uniestructural); (b) procesamiento de varios aspectos disjuntos (Multiestructural); y, finalmente, (c) integración de las relaciones entre varios aspectos del dominio de la tarea (Relacional).

[1] Structure of Observed Learning Outcome. La traducción al español es «estructura del resultado del aprendizaje observado».

En el primer ciclo los estudiantes fueron capaces de comparar los conjuntos de igual tamaño calculando el total o usando comparaciones visuales a partir de la gráfica. En el segundo ciclo, los estudiantes pudieron comparar conjuntos de datos de diferente tamaño usando un razonamiento proporcional. En este ciclo se incluyeron estrategias numéricas, generalmente comparación de frecuencias relativas, porcentajes o promedios; y estrategias visuales, a través de la comparación gráfica, y mixtas. En algunos casos usaron las características de las representaciones gráficas para describir variación entre conjuntos. Watson agrupó las estrategias de los estudiantes codificadas bajo el modelo SOLO en visuales, numéricas y mezcla de ambas. Las respuestas codificadas dentro del nivel uniestructural y multiestructural en ambos ciclos, predominantemente usaron estrategias visuales o numéricas, mientras que los relacionales, mezclaron las estrategias numéricas y visuales.

Bakker y Gravemeijer (2004) previeron que el razonamiento con graficas sería la base para el razonamiento sobre distribuciones, y con esa base diseñaron actividades con uso del ordenador en las que los estudiantes crearon sus propias gráficas basándose en datos e hicieron predicciones. El diseño se basó en una trayectoria de aprendizaje que involucraba tres niveles de comprensión de la distribución. En su artículo, los investigadores reportan la forma en que se estimuló el razonamiento de los estudiantes acerca de la idea de distribución. En el primer nivel (*cada dato es representado como una barra*) la actividad haría que los datos se concibieran simplemente como un conjunto de datos aislados, en el que, sin embargo, emergió razonamiento acerca de la distribución en diferentes aspectos: uso de una estimación gráfica de la media (a través de la compensación de las barras); ubicación de la mayoría de los datos (sesgo), probabilidad para datos atípicos o extremos y dispersión (confiabilidad). En el segundo nivel (*los puntos remplazan a las barras*) los datos se representaron a través de puntos que después fueron sustituidos por un histograma. Los estudiantes manifestaron su preferencia por esta representación porque «están mejor organizados» (p 156), así mismo desarrollaron nociones cuantitativas de aspectos más avanzados de la distribución tales como frecuencia, clases, dispersión, cuartiles, mediana y densidad. En el último nivel (*simbolizando los datos como un agregado*), los estudiantes crean sus propias gráficas sin datos, para analizar y describir lo que ellos esperan en varios escenarios. Usan el término «aglomerado» (*bump*) con relación a nociones estadísticas tales como mayoría, datos atípicos y tamaño de la muestra, incluso algunos de ellos llegan a percibir la distribución como una entidad conceptual global. Bakker y Gravemeijer concluyen haciendo explícitas las tres heurísticas para diseños didácticos en la educación estadística que se sustentan en el análisis del comportamiento de los estudiantes en cada nivel y que, a su vez, sustentan esos tres niveles propuestos por ellos.

Prodromou y Pratt (2006) proponen actividades de simulación en donde los estudiantes hacen emerger datos a partir de una distribución de la que puedan manipulan su centralización y su dispersión. De esta manera los estudiantes pueden observar el efecto de esas manipulaciones sobre los datos que surgen a partir de ellas, ofreciéndoles la visión de modelismo[2] de la distribución. Prodromou y Pratt encontraron apoyo en los niveles propuestos por Bakker y Gravemeijer (2004) en donde «las distribuciones pueden ser leídas desde o hacia la colección de datos» (p. 84) pero sugieren un viaje desde dos perspectivas: centrada en los datos y en el modelismo de la distribución y no sólo desde las medidas (de centralización, dispersión, densidad y sesgo) hacia los datos o hacia la distribución, en un solo sentido. Una perspectiva *desde* la distribución daría oportunidad de visualizar el tipo de datos que se generan a partir de ella y, por tanto, de abstraer conceptualmente la idea de distribución. Prodromou y Pratt consideran que lo *conceptual* descansa en la coordinación que se puede dar entre las dos perspectivas.

Ciancetta (2007) refina y construye un marco teórico interpretativo basándose en marcos teóricos propuestos por otros autores y en un análisis con estudiantes universitarios. Su marco tiene la intención de establecer categorías que permitan el estudio de las estrategias de razonamiento (*reasoning strategies*) que ponen en juego los estudiantes cuando se enfrentan a tareas que requieren hacer inferencias informales sobre pares de conjuntos de datos. Él considera cinco niveles de análisis: razonamiento idiosincrático (0), local (1), transicional (2), distribucional inicial (3) y distribucional (4). En el primero (*Nivel cero: idiosincrático*) incluye respuestas que no se centran en un razonamiento de los datos y por lo tanto no proporcionan información sobre razonamiento estadístico. En el *Nivel 1*, *Local*, se refiere a respuestas que implican una vista local de los datos y el uso exclusivo de razonamientos aditivos. El *razonamiento transicional* (*Nivel 2*) manifiesta una visión transicional entre la visión local y la global. Un tercer nivel (*distribucional inicial*) llevaría al uso de razonamiento proporcional y comenzar a reconocer los aspectos globales de los datos. Todos estos aspectos se integrarían en el más alto nivel, el *distribucional (Nivel 4)*.

Como se puede observar, las investigaciones sobre el concepto de distribución de una variable estadística se centran en los conceptos sobre cuya comprensión descansa este concepto y en la definición de niveles de comprensión de la distribución. Con

[2] En inglés los autores usan la palabra *modelling*, refiriéndose a la capacidad de generar modelos. En el sentido cotidiano, modelización se refiere, principalmente, a reproducciones (modelos) a escala. En el contexto que Prodromou y Pratt, lo entendemos como la capacidad de una distribución de modelar diversas situaciones al modificar sus parámetros.

frecuencia se utiliza la comparación de distribuciones para indagar en la comprensión del concepto. En consecuencia, estos resultados serán útiles en el análisis de los datos del Estudio 4 que implicará la comparación de varios pares de distribuciones por parte de los estudiantes.

3.2. Distribuciones de variables aleatorias

Aunque con menor frecuencia, la idea de distribución también ha sido investigada en relación con las variables aleatorias. Truran (1994) estudió la comprensión de los estudiantes sobre los niveles de desviación de la probabilidad empírica (frecuencial) de la probabilidad teórica. Él pidió a los estudiantes que conjeturaran sobre los resultados que esperarían obtener en muestras de nueve elementos tomadas de una población con 1/3 de bolas de un color y el resto de otro. Obtener valores extremos resultó sorprendente para los estudiantes, y casi todos esperaban aproximadamente la mitad de bolas de cada color en cada muestra. Fueron muchas las respuestas de los estudiantes que indicaban una teoría sobre el número de bolas que no guardaba relación con la composición de la urna. La conclusión del autor es que los estudiantes no comprenden la relación entre frecuencia y probabilidad ni la variabilidad de una variable aleatoria.

La mayoría de los estudios sobre el tema se enfocan al estudio de la distribución de una o varias variables estadísticas o aleatorias independientemente, sin embargo, hay estudios que sugieren enlazarlos. Así por ejemplo, Prodromou y Pratt (2006) en su investigación proponen el análisis «desde las distribuciones», es decir, promover que los estudiantes analicen los datos que surgen de manipular los parámetros de dispersión y centralización de una distribución de probabilidad para relacionar el modelo teórico de la variable aleatoria con el modelo empírico de la variable estadística y no sólo a la inversa.

Kazak y Confrey (2007), también sugieren que estos dos conceptos debieran ser relacionados tanto en la enseñanza como en la investigación. Las autoras manifiestan su preocupación por la inexistencia en el currículo escolar básico de una conexión conceptual entre la distribución teórica (usualmente enseñada dentro del tema de «probabilidad») y las distribuciones empíricas de datos (usualmente tratadas dentro del tema de «estadística»). Indican que esta vinculación no se puede posponer hasta el tema de inferencia estadística (en niveles universitarios o más) por la implicación que esto acarrea en la conceptualización de la idea de probabilidad como base para la cuantificación de la incertidumbre estadística. Ellas proponen enfatizar en la noción de *distribución* como una vinculación entre los datos y la probabilidad. En su investigación, ellas trabajan con un grupo pequeño de niños de 9 años a los que les piden realizar una secuencia de tareas enfocadas en las ideas de probabilidad y de ambos tipos de distribuciones y después los entrevistan para analizar sus concepciones.

Para Kazak y Confrey la idea de distribución comprende dos ideas: 1) una visión *estadística*, como agregado de un conjunto de datos (distribución de la variable estadística); y 2) una *visión probabilística* como conjunto de posibles resultados de un experimento aleatorio. En la primera se manejan distribuciones de datos y son descritas por los estadísticos de centralización y dispersión; la segunda se refiere al modelo teórico de una variable aleatoria vinculada con los resultados de un experimento aleatorio (distribución de probabilidad). Las tareas que ellas proponen a los niños incluyen el análisis de distribuciones espaciales de objetos reales (como flores o rebaños de animales), replicaciones de los experimentos de Piaget sobre la comprensión de la distribución normal, lanzamiento de monedas y experimentos que pueden asociarse con la distribución binomial. Concluyen que la comprensión intuitiva de la idea de distribución estadística de los niños puede evolucionar a un modelo más formal de distribución de una variable aleatoria, sobre todo si se realizan tareas de simulación.

En el estudio 4 de nuestra investigación, como Kazak y Confrey, trataremos de coordinar estas dos visiones de la distribución. Propondremos, a estudiantes universitarios, trabajar simultáneamente con variables estadísticas y aleatorias con el objetivo de obtener información sobre las concepciones y errores de los estudiantes al trabajar la vinculación de estos dos objetos matemáticos.

4. Investigaciones sobre la comprensión de la distribución normal

Los estudios descritos con anterioridad examinan la idea de distribución, sin referirse a un modelo particular, tal como la distribución normal o binomial. Sin embargo, gran parte de la utilidad de la variable aleatoria se debe precisamente a que se dispone de un conjunto potente de modelos con un fuerte grado de abstracción que permite que la distribución de la variable aleatoria quede determinada por la estimación de los parámetros del modelo (usualmente la media u otros estadísticos).

Entre los estudios relacionados con la comprensión de los modelos de las distribuciones, los más frecuentes son los relacionados con la distribución normal. En muchos de los cuales se centran en la comprensión de propiedades aisladas. Así, con relación a la tipificación (o estandarización), Huck, Cross y Clark (1986) identificaron dos concepciones erróneas, referentes al rango de variación de las puntuaciones normales tipificadas *Z*. El primer error se debe a que algunos alumnos creen que todas las puntuaciones *Z* han de tomar un valor comprendido entre -3 y +3, mientras que el hecho es que sólo el 99,7% de las observaciones se encuentra entre la media ± 3 desviaciones típicas. Otros estudiantes creen que las puntuaciones *Z* no tienen límite superior ni inferior, pues han aprendido que las colas de la curva normal son asintóticas al eje de abscisas y hacen una generalización incorrecta.

Para Hawkins, Joliffe y Glickman (1992) los errores más comunes en relación a la distribución normal vienen del rol dominante que tiene el teorema central del límite en estadística que llega a inducir la creencia de que todas las variables tendrían una distribución normal asintótica. También aparecen errores cuando los estudiantes usan la distribución normal como una aproximación de la distribución binomial, porque no ven la diferencia entre lo discreto y lo continuo, y en muchos casos aplican la corrección por continuidad de una forma mecánica sin entender su significado.

En su tesis de doctorado, Tauber (Tauber, 2001; Batanero, Tauber y Sánchez, 2001 y 2004; Tauber, Batanero y Sánchez, 2004) estudió el aprendizaje de la distribución normal en un curso de estadística con alumnos universitarios. Su estudio no se centró en la variable aleatoria, aunque es uno de los conceptos que consideró importantes en la descripción del significado de la distribución normal. La autora hizo un estudio sistemático de la enseñanza y aprendizaje de la distribución normal. Describió el significado del tema en los libros de texto universitarios y diseñó una secuencia de enseñanza con uso de ordenadores. Este proceso de enseñanza no siguió la pauta tradicional, sino que formuló problemas más realistas, incorporando los recursos gráficos y de cálculo que posibilita el ordenador, entremezclando actividades tradicionales, tales como exposición del profesor o resolución de problemas con papel y lápiz, con otras basadas exclusivamente en el uso del ordenador. La investigación se realizó en una muestra de 117 alumnos de la Facultad de Ciencias de la Educación y permitió mostrar la diferencia entre el aprendizaje producido a través de esa secuencia diseñada y el observado en una clase tradicional, sobre la que no se tuvo injerencia.

La evaluación del aprendizaje estuvo basada en los reportes escritos de actividades que realizaron los alumnos en sus clases, en un cuestionario final y en una prueba abierta de ensayo para ser resuelta con ayuda del ordenador. Como consecuencia del análisis de estos datos se describen dificultades y errores en un tema en el que la investigación previa era prácticamente inexistente. Asimismo, se describen las diferencias en la evaluación de dos tipos de alumnos: aquellos que tienen conocimientos previos de estadística, adquiridos en la educación secundaria o en cursos anteriores a la universidad y los que no los tienen.

La autora indica que muchos estudiantes que siguieron la secuencia diseñada alcanzaron la idea de distribución normal como modelo de los datos y comprendieron la utilidad de los modelos y la curva normal para resolver problemas prácticos. También fueron bien comprendidas la simetría de la distribución normal, las posiciones relativas de media, mediana y moda en distribuciones simétricas y asimétricas, el significado de los parámetros (media y desviación típica) y el uso del software estadístico. Algunos de los errores que se produjeron fueron la interpretación del área en un histograma de frecuencias, interpretación de probabilidades como áreas bajo la curva normal y la discriminación entre la variable estadística y aleatoria o entre la distribución de frecuencias empírica y la distribución teórica de probabilidad. Estas últimas dificultades pueden ser particularmente importantes en nuestro estudio.

5. Investigaciones específicas sobre la variable aleatoria

La mayoría de los estudios que se consultaron y que tratan específicamente sobre la variable aleatoria (aunque no necesariamente estén centrados en ella) se enfocan al aspecto didáctico del objeto, y dentro de ellos, la mayoría se encuadran en el análisis de libros de texto. Hay escasas investigaciones, sin embargo, en los aspectos epistemológicos y cognitivos. Dentro del primero de ellos, hay pocas referencias al aspecto histórico de la variable aleatoria y una mayor mención de cuestiones referentes a la disciplina.

La exposición en este apartado se divide en estos tres aspectos mencionados: didáctico, epistemológico y cognitivo. Dentro de las investigaciones dedicadas a la didáctica de la variable aleatoria, las enfocadas a libros de textos, por ser las más abundantes, ameritan una sección aparte.

5.1. Investigaciones sobre libros de texto

Diversos autores han estudiado la presentación de la variable aleatoria en los libros de texto, aunque algunos de ellos no se han centrado específicamente en la misma.

Ortiz (2002) hizo un estudio sobre la caracterización del significado de los conceptos probabilísticos que se presentan, como resultado de la transposición didáctica, en los libros de texto del primer curso de bachillerato español. Siguiendo un procedimiento cualitativo, analizó las definiciones, tipos de problemas y representaciones asociadas a los conceptos de aleatoriedad, frecuencia, probabilidad y variable aleatoria, entre otros, en una muestra de libros de texto. A partir de este análisis define los elementos de significado asociados a cada uno de ellos, siguiendo el marco teórico de Godino y Batanero (1998). En particular, los elementos de significado vinculados a la variable aleatoria que obtiene, los presentamos como un resultado de tipo epistemológico importante (sección 5.3), puesto que con ellos, Ortiz definió el significado de referencia de este objeto matemático.

Dentro de los resultados del análisis de los elementos de significado presentes en su muestra de libros de texto, Ortiz reportó que en la mayoría de los textos que analizó no se trata el concepto de variable aleatoria de manera explícita, pero aquellos que lo hacen, lo ubican ya sea en los temas de probabilidad o en los de estadística y que sólo contemplan una mínima parte de sus elementos de significado. Él encontró que el concepto también es tratado de manera implícita, sobre todo cuando se describe la variable estadística, que se obtiene a partir de los valores observados de una variable aleatoria en una muestra. En uno de los textos observó cierta confusión entre los conceptos de variable aleatoria y variable estadística, ya que atribuyen tanto la distribución de probabilidad como la distribución muestral a la variable aleatoria. Recomienda que, dada la importancia que tiene, la variable aleatoria debe ser tratada y relacionada con la probabilidad, aunque de una manera no excesivamente formalizada. Para lograrlo se deberá retomar con diferentes grados de profundización a lo largo de la educación secundaria.

En la caracterización del tipo de actividades (ejercicios, problemas o ejemplos) que se pueden resolver en el contexto de la variable aleatoria, define nueve tipos de actividades, pero en los libros de texto analizados, cuatro de ellas no se presentan. Las actividades presentes más frecuentemente se refieren a la determinación de la distribución de probabilidad de una variable aleatoria y a la representación gráfica de dicha distribución. Así mismo, reporta que las actividades presentes sólo se manejan referidas a espacios muestrales finitos en contexto únicamente de juego, que la asignación de probabilidades predominante es mediante la regla de Laplace y que la forma de presentación más usada es la verbal, en ocasiones con tablas de datos. Ortiz recomendó diversificar las actividades en los libros de texto, sin olvidar que los conceptos deben introducirse acordes al nivel de enseñanza y edades de los alumnos. Él considera que la categorización de la variable aleatoria que realizó de acuerdo a sus elementos de significado (presentados en el apartado 5.3 de este capítulo), sería una buena guía para lograrlo.

Oseguera (1994) hace un análisis del material de apoyo para el alumno que sirve como base para el diseño de la enseñanza y la evaluación del tema de variable aleatoria en la asignatura de probabilidad y estadística en la Unidad Profesional Interdisciplinaria de Ingeniería. Ciencias Sociales y Administrativas (UPIICSA) del Instituto Politécnico Nacional en México. El análisis de este material de apoyo se basa en tres ejes principales. Aquí desglosamos los dos de mayor interés para nuestra investigación:

- Los conceptos básicos de probabilidad relacionados con variable aleatoria. Este criterio fue establecido por el mismo autor. Los conceptos básicos se refirieron principalmente al tipo de distribución que se ponía en juego en cada reactivo (normal, uniforme o exponencial) y al manejo de sus respectivos parámetros, así como a la aproximación de la función de distribución normal a una binomial. Dentro de estas categorías también se incluyeron el área en la que incide el reactivo y si se requiere alguna representación gráfica.
- La propuesta epistemológica de Heitele (1975) que describiremos posteriormente (sección 5.3). Oseguera seleccionó aquellas ideas fundamentales que se relacionan con la variable aleatoria y detectó si estaban presentes en el material analizado.

En su trabajo, él observa que hay una tendencia predominante a interpretar los reactivos meramente como modelos matemáticos, olvidándose del contexto y de la forma en que los números fueron asignados a los valores de las variables aleatorias y cómo estos números pueden relacionarse con las mediciones (datos) y con el modelo matemático. Así mismo observa que hay una ausencia de vinculación entre el tópico de estocásticos y las experiencias intuitivas de los estudiantes dando prioridad a requerimientos puramente matemáticos.

Oseguera concluye asegurando que en el contexto que él analiza no hay aspectos del surgimiento de la idea de variable aleatoria en el proceso de enseñanza-aprendizaje, reduciéndolo al manejo metodológico de reglas y símbolos.

En su breve artículo, Miller (1998) hace una crítica al tratamiento de la variable aleatoria en algunos libros de texto. Él afirma que en los libros de introducción a la estadística casi no desarrollan esta idea y que cuando lo hacen, frecuentemente lo hacen de una forma errónea que acarrea problemas en temas posteriores. Critica, en particular, el tratamiento de la variable aleatoria como asociada al valor de los datos. Este tratamiento que crea una confusión entre el resultado del fenómeno aleatorio propiamente y el valor o dato asignado a la variable aleatoria, como si ese dato no pudiera existir en sí mismo sin la presencia del fenómeno aleatorio.

Miller toma dos ejemplos de un libro de Moore[3], en los que se ilustra el concepto de variable aleatoria con la distribución de las calificaciones de un grupo de estudiantes y con la distribución del tamaño de un grupo de casas. En ambos casos el resultado del fenómeno aleatorio no es la calificación obtenida o la medida de la casa, sino la casa o el estudiante seleccionado aleatoriamente. El valor numérico relacionado con la casa o el estudiante están definidos para cada casa y para cada estudiante, no así el estudiante o la casa seleccionada a través de un fenómeno aleatorio. Los problemas son, afirma Miller, que (1) el valor numérico no determina claramente a la variable aleatoria (el tamaño de la casa no es un «resultado» directo del fenómeno aleatorio) y (2) que en el discurso estadístico sí aparecen variables aleatorias dadas por un valor numérico calculado o determinado (como en el caso de la variable explicativa del modelo de regresión), lo que conlleva, tarde o temprano, a una contradicción en el discurso escolar.

En su tesis de doctorado, Tauber (2001) describe a la variable aleatoria como uno de los objetos matemáticos importantes en el significado de la distribución normal. Entre otros análisis, la autora estudia la forma en que se presenta la distribución normal en una muestra de 11 libros universitarios de estadística aplicada de acuerdo con el enfoque Onto-semiótico (Godino, 1996, 2002) e identifica los siguientes elementos de significado:

- *Campos de problemas:* La distribución normal se utiliza para resolver problemas de ajuste de modelos a datos reales, aproximar distribuciones discretas, obtención de distribuciones muestrales exactas o asintóticas, estimación y contraste de hipótesis
- *Representaciones* verbales, simbólicas, tabulares y gráficas;
- *Procedimientos* para resolver los problemas, tales como estudio descriptivo, tipificación, cálculo de probabilidades y valores críticos a partir de tablas, cálculo de límites del intervalo central y comparación de puntuaciones normales.
- *Definiciones y propiedades,* tales como simetría, asíntota horizontal, área bajo la curva, proporción de casos en intervalos centrales, curtosis, etc.
- *Argumentos:* deductivos, ejemplos- contraejemplos, análisis y síntesis, prueba por contradicción, o simulación.

Tauber concluye que en los libros de texto no siempre se relaciona el estudio de la estadística descriptiva con el de la variable aleatoria y con el de las distribuciones de

[3] 1995. *The basic practice of stadistics*. Freeman: New York

probabilidad. También coincide con Oseguera (1994) en que los libros de texto enfatizan sobre todo en los elementos simbólicos, algebraicos y numéricos, particularmente las tablas de la distribución y se dedica gran espacio al cálculo de probabilidades a partir de las tablas, con menor énfasis en la experimentación, simulación y estudio de propiedades.

En resumen, estas investigaciones indican que algunos autores se enfocaron directa o indirectamente al estudio del objeto de nuestro interés en los libros de texto concluyendo que los libros de texto casi no desarrollan la idea de variable aleatoria de manera explícita y cuando lo hacen sólo contemplan una mínima parte de sus elementos de significado o bien cometen errores en su interpretación (Oseguera, 1994; Miller, 1998; Tauber, 2001 y Ortiz, 2002). También encontraron errores en los libros. Ortiz reportó una confusión entre variable aleatoria y variable estadística en algunos libros de secundaria. Miller (1998) muestra la inconveniencia de asociar a la variable aleatoria con el valor de los datos directamente en los libros de introducción a la estadística universitarios. En este mismo sentido Borovcnik et al (1991, citado por Ortiz, 2002, p. 121) da una definición práctica del concepto: «una variable aleatoria se determina como resultado de un experimento aleatorio» (p. 121) que aunque sería considerada imprecisa por Miller (puesto que relaciona la variable directamente con el resultado de un fenómeno aleatorio), Ortiz la considera apropiada para el nivel secundaria.

5.2. Otras investigaciones sobre didáctica

Aparte del énfasis en el análisis de libros de texto, existen otras investigaciones enfocadas a las diversas formas en que se presenta, o debería presentar, la variable aleatoria en la enseñanza. Heitele (1975), por ejemplo, propone que el éxito en la comprensión de la variable aleatoria en cada nivel escolar estará en ser tratada bajo diferentes niveles cognitivos que desarrollen modelos explicativos que difieran en su forma lingüística y en sus niveles de profundización. Heitele se pregunta qué será más conveniente, si introducir el concepto de variable aleatoria vía el caso especial de equidistribución o introducirlo hasta el final, vía las distribuciones generales. Sin embargo, no encontramos más referencias a cuáles podrían ser esos estadios de la variable aleatoria que propiciarían el crecimiento del concepto de variable aleatoria en cada nivel escolar. Así pues, no se tiene muy claro qué es lo que se espera que los estudiantes aprendan en cada nivel escolar, aunque se manifiesta una necesidad de que la noción sea introducida en niveles escolares primarios.

Nardecchia y Hevia (2003), realizaron una investigación bibliográfica histórica tendiente a encontrar los posibles obstáculos didácticos que el aprendizaje de la variable aleatoria puede implicar. En uno de los resultados preliminares de su investigación, ellos concluyeron que el concepto de variable aleatoria está fuertemente vinculado con el de aleatoriedad, pero que la historia muestra que ha sido difícil aceptar esa relación de modo que una de las principales dificultades con las que el estudiante se podría enfrentar en la construcción de la noción de variable aleatoria sería no visualizar la presencia del azar en el fenómeno aleatorio que se está tratando de modelar a través de la variable aleatoria.

Ellos argumentan que, aunado a esto, los estudiantes tienen predominantemente desarrollado el pensamiento determinístico sobre el probabilístico y que esto puede influir aún más en la presencia de ese obstáculo principalmente en la enseñanza superior. Concluyen, también, que históricamente no ha sido simple la construcción de un modelo adecuado a partir de los datos observados, de modo que esta vinculación entre la realidad y la variable aleatoria (como modelo matemático) puede constituir un obstáculo con el que se podría enfrentar un estudiante. Así mismo, ellos enfatizan en la importancia de realizar estudios que nos indiquen la transposición didáctica que el objeto matemático 'variable aleatoria' ha sufrido para poder ser incorporado a la enseñanza en las instituciones educativas.

Varios de los autores citados en esta tesis concuerdan en que en el material de apoyo de la enseñanza hay un énfasis en el tratamiento de los elementos simbólicos, algebraicos y numéricos de la variable aleatoria y en que hay una ausencia de vinculación entre el tópico y las experiencias empíricas. Oseguera (1994) y Tauber (2001) encontraron que no siempre hay una conexión entre el estudio del modelo probabilístico y el análisis de los datos empíricos en el material de apoyo y textos universitarios que analizaron, olvidándose del contexto y de la forma en que se asignaron los valores a las variables aleatorias y cómo estos números pueden relacionarse con las mediciones (datos) y con el modelo matemático.

Esto significa que hay un rompimiento entre el fenómeno aleatorio y la variable aleatoria o su distribución de probabilidad. Nardecchia y Hevia (2003) ya mencionan la dificultad en aceptar históricamente la relación entre la aleatoriedad y la variable aleatoria. Cosa por demás grave si a esta ausencia añadimos las dificultades que han sido señaladas en numerosas investigaciones sobre la percepción de la aleatoriedad en los estudiantes y las dificultades filosóficas alrededor del concepto de aleatoriedad y azar, que se describen más adelante y encontramos, por ejemplo, en Falk y Konold, (1997) y Batanero y Serrano (1995 y 1999).

5.3. Investigaciones sobre aspectos epistemológicos

Desde el punto de vista conceptual, Heitele (1975) considera que hay tres aspectos que juegan un papel fundamental en la noción de variable aleatoria:

- *Su distribución*, caracterizada en muchos casos por su valor esperado y su varianza o su desviación estándar. El análisis crítico de datos requiere de la caracterización apropiada de una función de distribución, a través de su esperanza y de su desviación estándar.
- *Su esperanza*, que ha mostrado ser una idea muy intuitiva y que tiene un valor explicativo muy grande en un amplio campo de problemas.
- También considera importante la *composición* de variables aleatorias para obtener otras nuevas, como modo de componer modelos complejos a partir de otros más sencillos.

Así mismo, Ortiz (2002) hizo un desglose de los elementos de significado de la variable aleatoria como objeto matemático, haciendo notar su complejidad epistémica, la cantidad de conceptos que es necesario poner en juego para su comprensión y los campos de problemas en los que se encuentra. Los elementos que desglosa se ocupan de la variable aleatoria y de su distribución, así como de los elementos que la definen: su media y su varianza. La base de su análisis fue una definición formal del concepto, aunque su análisis en los libros de texto incluye nociones no dadas explícitamente, aquéllas en las que el concepto se presenta y las relaciones que le son asociadas. Identifica los siguientes elementos, que tendremos en cuenta, entre otros, en el análisis conceptual sobre la variable aleatoria (Ortiz, 2002, p. 121):

- *Definiciones y propiedades de la variable aleatoria:*

VA1: La variable aleatoria toma sus valores dependiendo de los resultados de un experimento aleatorio.

VA2: Es una función del espacio muestral en *R*.

VA3: Queda caracterizada mediante la distribución de probabilidad, que es el conjunto de valores que toma la variable, dando para cada uno su probabilidad.

VA4: Para que una variable definida sobre un espacio muestral sea variable aleatoria, se requiere que, para cada intervalo *I* de *R* el conjunto original de dicho intervalo sea un suceso del espacio muestral.

VA5: Una variable aleatoria define una medida de probabilidad sobre el conjunto de números reales.

VA6: Para cada variable aleatoria podemos definir una función de distribución de la forma siguiente:

$$R :\to [0,1]$$
$$x :\to F(x) = P(\xi \leq x)$$

VA7: La función de distribución de una variable aleatoria es una función real de variable real, monótona no decreciente.

VA8: La función de distribución de una variable aleatoria determina en forma biunívoca la distribución de probabilidad.

VA9: Sea $(x_i, p_i)\, i \in I$ la distribución de probabilidad de una variable aleatoria discreta. Se define la media o esperanza matemática como $E[\xi] = \sum_{i \in I} x_i\, p_i$. Este concepto extiende la idea de media en una variable estadística.

VA10: La moda es el valor más probable de la variable aleatoria y extiende la idea de moda de una variable estadística

VA11: La mediana es el valor de la variable para el cual la función de distribución toma el valor 1/2. Por tanto, la probabilidad de que una variable aleatoria tome un valor menor o igual a la mediana es exactamente 1/2.

- *Las actividades (problemas y ejercicios) característicos de la variable aleatoria, son categorizadas en los siguientes puntos* (Ortiz, 2002, pp 199-202)*:*

SVA1: Determinar la distribución de probabilidad de una variable aleatoria.

SVA2. Determinar la probabilidad de que una variable aleatoria tome un valor determinado.

SVA3. Dada una lista de variables aleatorias, decidir si son cualitativas o cuantitativas y cuáles son discretas y cuáles continuas.

SVA4: Representar gráficamente la distribución de probabilidad de una variable aleatoria.

SVA5: A partir de la representación gráfica, hallar la distribución de probabilidad.

SVA6: Obtener la función de distribución dada la distribución de probabilidad.

SVA7: Obtener alguno de los valores de posición central, media, mediana o moda o de dispersión.

SVA8: Dada una distribución de probabilidad comprobar si verifica los axiomas de probabilidad.

SVA9: Calcular el valor de la variable aleatoria al que corresponde una cierta probabilidad.

Hay que hacer notar que estos problemas corresponden a problemas presentados en libros, son, por tanto, problemas intra-matemáticos, pero no constituyen verdaderas situaciones que den origen a la idea de variable aleatoria. Sobre estos tipos de ejercicios identifica las siguientes variables adicionales, en función de lo encontrado en los libros de texto (pp 50-51):

V1: Tipo de actividad que se pide al alumno, que puede ser: ejemplo introductorio; ejemplo después de la definición; ejercicio introductorio; ejercicio después de la definición y ejercicio después de haber hecho un ejemplo similar.

V2: Concepto al que el ejercicio se refiere explícitamente o implícitamente, entre los citados anteriormente.

V3: Tipos de espacio muestral, considerando espacio muestral infinito, finito, con dos elementos equiprobables; finito, con más de dos elementos equiprobables, finito, con sucesos no equiprobables e impreciso.

V5: Posible asignación de probabilidades a los sucesos; clásica, frecuencial o bayesiana

V6: Contexto del enunciado del ejercicio: Se refiere al campo de aplicación mostrado.

V7: Presentación de la información: Verbal, numérica o gráfica.

El desglose de los elementos de significado de la variable resulta valioso por ser una propuesta de análisis epistémico del concepto para el nivel secundaria.

Tauber (2001) realiza un estudio conceptual detallado de la distribución normal y sus elementos de significado en nivel universitario. En su análisis enfatiza la dificultad de discriminar entre la variable aleatoria (modelos matemáticos) y variable estadística (datos que se quieren modelar). Sin una adecuada discriminación entre ellos se pierde el sentido de la modelación.

Heitele (1975) recalca la diferencia de la modelación en matemáticas y en probabilidad y estadística. Wild y Pfannkuch (1999) consideran que cualquier forma de representar la realidad puede considerarse modelo (un estadístico útil, un texto, un gráfico o una línea de regresión), siempre y cuando se diferencie el modelo de los datos y al mismo tiempo se relacione con ellos. Esta diferenciación se facilita cuando las relaciones entre realidad y matemáticas se ven de acuerdo a la estructura de estratos que propone Heitele. Esto es, los datos numéricos en estadística descriptiva se pueden considerar datos de un modelo, a la vez que desde un nivel más alto (en otro estrato), se les puede considerar parte de la realidad. Lo importante para él es definir cuál es el nivel del estrato en el que se está trabajando y qué es lo que en ese estrato puede fungir como «realidad» y qué como «modelo». Heitele no enfatiza en la variable aleatoria en particular, pero sí menciona la importancia de considerar este problema en la modelación en probabilidad y estadística, puesto que en el contexto estocástico, el «modelo» puede modificar los datos que surgen de la «realidad».

Desde la perspectiva histórica se hace referencia a la importancia de la variable aleatoria en la comprensión de nociones que eran difíciles de entender antes de su surgimiento, tales como el teorema central del límite (Heitele, 1975) y la dificultad en la aceptación histórica de la relación entre variable aleatoria y aleatoriedad. Nardecchia y Hevia (2003) mencionan la posibilidad de que los estudiantes tengan esta misma dificultad cuando se está tratando de modelar a través de la variable aleatoria.

Hemos encontrado algunos estudios epistemológicos históricos del teorema central del límite, cuyo desarrollo también contribuye a la idea de variable aleatoria. En particular Alvarado (2007) comenzó su estudio discutiendo los primeros campos de problemas que parten de situaciones prácticas alrededor del teorema central del límite, por ejemplo, la necesidad de aproximar la distribución de la suma de errores y la búsqueda de aproximaciones para la suma de variables aleatorias. Continuó con el razonamiento de Laplace, quien dio la primera formulación limitada del teorema central del límite, que varios autores trataron de generalizar después, originando campos de problemas cada vez más abstractos y ligados al método deductivo, propio de la matemática.

De acuerdo con Alvarado, Poisson dio ejemplos de distribuciones que no pueden ser aproximadas por el teorema central del límite y mejora la demostración para el caso continuo. Dirichlet y Bessel introdujeron el factor de discontinuidad y probaron el teorema de Poisson para un caso general. Cauchy definió la primera cota superior para la diferencia entre la distribución de la suma y la distribución normal. Después los matemáticos rusos consiguieron la primera demostración rigurosa del teorema central del límite; primero Tchebychef y Markov, con el método de los momentos, y posteriormente Liapounoff, con el uso de la condición que lleva su nombre y la función característica. Linderberg finalizó la búsqueda de la condición suficiente y Feller y Lévy hallaron la condición necesaria. Finalmente, fue rigurosamente demostrado por Cramer, quien llegó a la versión moderna del teorema central del límite. Actualmente ya se han obtenido las condiciones necesaria y suficiente para varias variables dependientes, pero los principios básicos de Linderberg, Lévy y Feller aún permanecen (Alvarado y Batanero, 2006). El estudio realizado destaca también algunas dificultades en el desarrollo histórico del teorema central del límite, que podrían ser compartidas por los estudiantes:

- Aceptar que una distribución discreta (por ejemplo la binomial) pueda aproximarse mediante una distribución continua. Esto requiere pasar de considerar valores aislados de la variable a valores continuos y de la idea de función de probabilidad a la de función de densidad de probabilidad. El mismo De Moivre, aunque demostró la aproximación a la binomial a través de la normal, no relacionó la fórmula obtenida con la función de densidad de una variable aleatoria (que sería el límite de la sucesión correspondiente de variables aleatorias binomiales).
- Aceptar las sucesivas generalizaciones del teorema central del límite. El desarrollo histórico de este concepto no ha sido sencillo y ha requerido numerosas etapas. Sin embargo, actualmente el teorema central del límite se enseña con total generalidad, cuando la comprensión de esta generalidad de aplicación podría requerir un tiempo más largo de maduración.
- Pensar que el teorema central del límite se puede aplicar en cualquier condición. La manera tan general en que se enseña, puede llevar a pensar que el teorema central del límite tiene una validez general cuando en realidad está sujeto a condiciones necesarias y suficientes. Cauchy fue el primero en encontrar algunos contraejemplos, pero tomó mucho más tiempo encontrar cuáles eran esas condiciones para que se cumpliera. También es posible que se confundan las condiciones necesarias con las suficientes.

Alvarado indica que es importante que el profesor sea consciente que estos obstáculos pueden estar siendo fomentados por su enseñanza. También en que son necesarias investigaciones que documenten la presencia de estos obstáculos en los procesos de aprendizaje de los distintos niveles educativos.

5.4. Investigaciones cognitivas

En este apartado encontramos algunos aspectos reportados en los que los estudiantes se pueden enfrentar a problemas con el aprendizaje de la variable aleatoria:

- La intuición sobre magnitudes en las que participa el azar surge en etapas más tempranas que la de experimento aleatorio (Heitele, 1975 y Batanero, 2001).
- Los estudiantes tienen predominantemente desarrollado el pensamiento determinístico sobre el probabilístico y eso puede influir en la no visualización del azar en el fenómeno aleatorio cuando se está tratando de modelar haciendo uso de la variable aleatoria (Nardecchia y Hevia, 2003).

En primer lugar analizamos la investigación de Vallecillos (1996 y 1999), en la que hace un análisis didáctico del contraste de hipótesis desde la doble perspectiva, teórica y experimental. Ella hace una investigación bibliográfica profunda sobre los aspectos matemáticos, históricos, filosóficos y didácticos del contraste de hipótesis en donde se pone de manifiesto la complejidad del significado de este concepto. Esta investigación sirve de referencia para plantear una investigación experimental sobre el aprendizaje del contraste estadístico de hipótesis en estudiantes universitarios. Su centro está en la detección de los errores y las concepciones de estos estudiantes. Entre sus principales conclusiones Vallecillos (1999) destaca la confusión entre estadístico (que es un representante de la variable estadística) y parámetro (que es también un representante, pero esta vez, referido a la variable aleatoria), ya que, indica «aunque los alumnos diferencian entre la media de la muestra y la de la población, no perciben la media muestral como una variable aleatoria» (p. 245). Así mismo indica que la comprensión del estadístico como variable aleatoria y su dependencia del parámetro puede ser una de las causas que influyan en la interpretación incorrecta del nivel de significancia en las pruebas de hipótesis. En estos tres puntos se destacan las dificultades del aprendizaje de la variable aleatoria cuando se trata de manera intuitiva y en su papel dentro de la modelación de los fenómenos aleatorios.

En su propuesta de enseñanza, Tauber (2001) introduce la variable aleatoria continua como la generalización de la idea de la distribución de frecuencias de una variable estadística continua, es decir, presenta la función de densidad como el modelo matemático al que tendería el histograma de la variable estadística cuando se aumenta indefinidamente el tamaño de la muestra, haciendo progresivamente cada vez más pequeños los intervalos. Su análisis se centra en las diferencias y concordancias presentadas entre el significado de la distribución normal de los estudiantes que siguieron su propuesta y el significado institucional local pretendido. La propuesta didáctica implementada también está basada en el análisis de los elementos de significado de la distribución normal que se presentan en una muestra de libros de texto universitarios siguiendo su marco teórico.

Basándose en la confusión entre el modelo y los datos reportada por autores como Vallecillos (1999), Tauber, entre unos de sus objetivos de evaluación, se propone observar si los estudiantes discriminan correctamente los conceptos de parámetro y estadístico y la distribución teórica de la empírica. De modo que parte de su análisis desemboca en la comparación entre el manejo de la variable estadística y de la variable aleatoria.

Algunas de las conclusiones que Tauber obtiene, importantes dentro del contexto que nos ocupa, son las siguientes:

- Pudieron observarse ciertas dificultades de los alumnos en la distinción de la distribución teórica y empírica, sobre todo cuando se ven en la necesidad de resolver problemas abiertos.
- Los alumnos llegan a un manejo razonable de la herramienta informática y a realizar correctamente cálculos aislados. Sin embargo, cuando se trata de poner en correspondencias diferentes elementos del significado para tomar una decisión -por ejemplo decidir si la variable estadística podría modelarse adecuadamente mediante la distribución normal- el número de respuestas correctas baja drásticamente.

Observemos que tanto Vallecillos (1999) como Tauber (2001) hacen hincapié en la importancia de una de las concepciones de la media muestral como variable aleatoria. Por otro lado, Miller (1998) afirma que el tratamiento didáctico de las variables aleatorias definidas como el resultado de un experimento aleatorio generará confusión en el estudio de las variables aleatorias dadas por un valor calculado. Ambas cuestiones están sumergidas dentro de la problemática de la composición de variables aleatorias, que ya Heitele (1979) mencionaba como uno de los aspectos epistemológicos fundamentales de este concepto.

Además de estas investigaciones en donde directamente se hace mención a problemas cognitivos vinculados con la variable aleatoria, existen otros que aunque sus autores no los relacionan como vinculados con la variable aleatoria, se puede observar la vinculación directa con este concepto. Así por ejemplo, la creencia de que en el lanzamiento de dos dados, la suma de las caras superiores puede ser 11 ó 12 con la misma probabilidad (Lecoutre, 1992; Bennett, 1998) nos da evidencia, entre otras, de una confusión conceptual: los valores que puede tomar la variable aleatoria se piensan como los eventos posibles del espacio muestral.

De esta revisión podemos concluir que la investigación sobre la variable aleatoria es aún muy somera, incluso cuando este tema forma parte fundamental del discurso escolar de la Probabilidad y Estadística universitaria. Los autores citados concuerdan en que el concepto es básico para introducirse en el estudio de la estadística universitaria y algunos de ellos recomiendan que sea retomado una y otra vez a lo largo de la educación previa a la universitaria con diferentes grados de profundidad (Heitele, 1975; Batanero, 2001; Miller, 1998 y Ortiz, 2002).

6. Investigaciones sobre la percepción de la aleatoriedad

Son diversos los motivos que nos impulsan a detenernos en las investigaciones sobre la percepción de la aleatoriedad. En primera instancia, algunos autores, y en nuestro mismo estudio histórico (Capítulo 4), detectaron relación entre las dificultades de la variable aleatoria con la falta de comprensión de la idea de aleatoriedad. Por otro lado, el análisis conceptual de la variable aleatoria reportado en el Estudio 1 (Capítulo 3) nos hizo tomar a la aleatoriedad como uno de los objetos de análisis en los estudios cognitivos que realizamos tanto en la entrevista clínica (Estudio 3) como en la evaluación del proyecto (Estudio 4). Y, por último, puesto que está tomado de las investigaciones sobre percepción de la aleatoriedad, el contexto del proyecto del Estudio 4 gira alrededor del cuestionamiento sobre la intuición de aleatoriedad de los estudiantes.

En la actualidad, hay abundancia de trabajos reportados sobre las percepciones de la aleatoriedad. A partir de los años ochenta han sido muchos los investigadores, tanto en psicología como en educación matemática, que han tratado de evaluar el significado atribuido a la aleatoriedad por niños y adultos, planteándoles problemas y centrándose en la determinación de las propiedades que se les atribuye a este concepto (e.g., Falk, 1981; Green, 1982; Falk y Konold, 1987; Serrano, 1996; Engel y Sedlmeier, 2005; Chernoff, 2009a y b). Todas estas investigaciones se basan en la respuesta de los sujetos a problemas que se clasifican en dos tipos:

- *Problemas de generación*. En este tipo de tarea se les pide a los sujetos simular una secuencia de resultados aleatorios. Por ejemplo, poner puntos en un folio en una distribución que ellos consideren aleatoria o escribir una sucesión de dígitos que parezca que fue obtenida al azar. Nosotros usaremos este tipo de problema en el segundo estudio cognitivo descrito en el Capítulo 7, donde pediremos a los estudiantes escribir una sucesión de 20 lanzamientos de una moneda equilibrada.
- En los *problemas de reconocimiento* a los participantes se les pone ante ciertas situaciones, secuencias o patrones espaciales y se les pregunta si son o no aleatorios o se les pide elegir entre varias, aquella que parezca obtenida al azar. Este tipo de tarea ha sido considerada por Chernoff (2000a) como *tarea de comparación de verosimilitudes* (CLT, por sus siglas en inglés).

Estas investigaciones muestran que los sujetos no son buenos para generar ni reconocer la aleatoriedad. En general se tiende a encontrar patrones deterministas en las situaciones aleatorias, es decir, tratan de encontrar asociaciones inexistentes, con objeto de reducir la incertidumbre. Simultáneamente, hay una tendencia a inferir aleatoriedad en situaciones en las que no está presente. Otra de las conclusiones de estos estudios es la existencia de sesgos en el razonamiento aleatorio. Uno de los más frecuentes es la tendencia a generar rachas cortas de dos o tres símbolos adyacentes en algún sentido, por ejemplo números consecutivos o letras sucesivas del alfabeto cuando se pide generar secuencias que simulen provenir de procesos aleatorios. Resultados muy similares se encuentran en experimentos con tareas bidimensionales, en las que las personas no consideraron aleatorios los patrones en los que aparecen agrupaciones de puntos en un espacio pequeño. Estos resultados son consistentes con la teoría de Kahneman, Slovic y Tversky (1982) sobre la heurística de la representatividad, según la cual, las personas evalúan la probabilidad de que ocurra un suceso de acuerdo con el grado de similitud del mismo con la población de la que proviene, o al proceso del cual se supone ha sido generado.

La teoría matemática indica que dos sucesiones aleatorias de igual longitud obtenidas a partir del mismo experimento, y cuyos sucesos elementales son equiprobables, también son equiprobables y por tanto son igualmente aleatorias. Sin embargo, la investigación sobre la percepción subjetiva de la aleatoriedad indica que los sujetos consideran más probables unas secuencias que otras (y por lo tanto más o menos factibles de ser aleatorias). Generalmente los sujetos establecen criterios apoyándose en algunas de las características de las secuencias, como el número de rachas o la longitud de la racha más grande (Batanero y Serrano, 1995), y con base en ellos deciden si son o no aleatorias. Por ejemplo, las siguientes secuencias son matemáticamente equiprobables:

O X O X O X O X O O O X X X X O X O X O O

O X O X X X X X O O O O O O X O O O X X O

Sin embargo, muchas personas pensarían que la primera es más probable porque en la segunda secuencia se presentan rachas excesivamente largas. Esta creencia (de que las rachas en una secuencia aleatoria son más cortas) puede estar sustentada en la *falacia del jugador*, según la cual, las cruces serían consideradas más probables que las caras después de una serie de caras consecutivas. En una tarea en donde se le pidiera a un sujeto decidir cuál de las dos series no fue generada a través de un proceso aleatorio, él escogería la segunda secuencia basándose en ese criterio.

Bennett (1998) realizó un estudio histórico-epistemológico sobre la emergencia de la idea de aleatoriedad y los significados que ha recibido a lo largo de la historia, identificando, entre otras, las siguientes concepciones:

- *Aleatoriedad como falta de causalidad*. Fue la primera visión del azar, cuando aún la idea de aleatoriedad no estaba formada y, según Bennett, duró hasta la edad media. De acuerdo con Batanero y Serrano (1999), aún hoy en día aparecen concepciones de azar como falta de causa o bien, azar como la conjunción de una multiplicidad de causas.
- *Aleatoriedad como incertidumbre*; falta de control o imposibilidad de predicción.
- *Aleatoriedad como equiprobabilidad*. Según esta visión, un fenómeno es aleatorio sólo si los posibles resultados son igualmente probables. Esta concepción es errónea porque sí existen los experimentos aleatorios en donde los sucesos no son equiprobables. Esta concepción estuvo asociada con el comienzo del cálculo de probabilidades, cuando se seguía una acepción clásica de la probabilidad.
- Análogamente, *se le asignan diferentes visiones de la aleatoriedad a las concepciones frecuencial y subjetiva de la probabilidad*. Así, la aleatoriedad podría ser vista como la estabilización de la frecuencia relativa en sucesivas repeticiones de un mismo suceso o bien, se puede pensar que depende de la información que un sujeto tiene sobre un suceso. Esto último es muy común y está relacionado con una etapa histórica de la concepción de aleatoriedad.
- *Aleatoriedad como modelo matemático*, es la concepción del estadístico.

Konold et al. (1991a) exploraron las diferencias entre la concepción de experimento aleatorio en personas sin instrucción en probabilidad y expertos y argumentan que, de hecho, el término «aleatoriedad» comprende una familia de conceptos. Resultados similares fueron encontrados por Batanero y Serrano (1999) en un trabajo con alumnos de secundaria, donde describen con más precisión las concepciones anteriores.

6.1. Generación de resultados aleatorios

En la segunda situación experimental de esta investigación, propondremos a los estudiantes que simulen el lanzamiento de una moneda equilibrada. Este tipo de situaciones ha sido utilizado en otras investigaciones. Green y Hunt (1992) pidió a niños de 7 a 11 años escribir una sucesión de caras y cruces que represente lo que ellos creen que se obtendría al lanzar 50 veces una moneda equilibrada. Los resultados mostraron que los niños son muy exactos al reflejar la equiprobabilidad. Además producen sucesiones cuya primera y última mitad son altamente consistentes; quizás demasiado consistentes para reflejar la variabilidad de una sucesión aleatoria.

Green y Hunt también estudiaron la racha de mayor longitud en las secuencias escritas por los niños. Observó que había casos en que la racha de mayor longitud era de longitud 1, pero también había 25 o más signos iguales consecutivos. Asimismo se basó en el número total de rachas en cada una de las sucesiones generadas por los alumnos (relacionado con la variable anterior, pero no es idéntica). Observaron que la longitud media de la racha más larga en su muestra era inferior al valor teórico en niños de todas las capacidades y de diferentes edades. También comparó el número de rachas generadas por los alumnos con las distribuciones teóricas esperadas y encontró que, en general, los alumnos producen demasiadas rachas y que estas son de longitud corta.

Engel y Sedlmeier (2005) propusieron un experimento similar y realizaron el mismo análisis que Green, encontrando resultados muy parecidos. Ellos además, calcularon la probabilidad teórica de alternancia de los resultados de una secuencia aleatoria, que sería de 0,5, y la compararon con el valor medio de la probabilidad de alternancia calculada en cada una de las secuencias producidas por sus estudiantes. Estos valores medios se encontraban entre 0,6 y 0,7 por lo que los autores concluyen, con este nuevo procedimiento, que los sujetos que participaron en su investigación no percibieron adecuadamente la independencia y ese es el motivo por el que producen demasiadas rachas y cortas.

En el proyecto que reportamos en el Capítulo 6 de esta memoria analizamos las mismas variables que analizó Green para observar si los estudiantes perciben la diferencia entre las distribuciones de las mismas en secuencias simuladas y secuencias de resultados aleatorios producidas mediante la realización de experimentos. A diferencia de las investigaciones reportadas en este apartado, en el proyecto planteado en nuestra investigación, a los estudiantes se les propone que sean ellos mismos los que analicen cómo son sus intuiciones sobre la aleatoriedad, siguiendo la línea estipulada en Godino, Batanero, Roa y Wilhelmi (2008) y Arteaga (2011). Así podremos observar, además de la percepción sobre la aleatoriedad de ellos mismos, la forma en que los estudiantes analizan las variables aleatorias y estadísticas involucradas en ese análisis.

6.2. Reconocimiento de resultados aleatorios

Las investigaciones de Green (1982) y Toohey (1995) incluyen también preguntas en las que los niños deben diferenciar entre una sucesión generada por un mecanismo aleatorio y otra que no cumple esta condición. Como resultado, descubren que la mayor parte de los niños elige precisamente la secuencia no aleatoria y que no mejora la apreciación de la aleatoriedad con la edad. Entre las razones aducidas por los niños para justificar su decisión, encontró las siguientes:

- *Razones correctas*: el patrón de la sucesión es demasiado regular para ser aleatorio; hay demasiadas rachas o rachas demasiado cortas.
- *Razones incorrectas*: no se apreciaba suficientemente bien la irregularidad de una sucesión aleatoria; se espera que la frecuencia relativa sea exactamente el 50% de caras y cruces o un valor muy próximo; rachas demasiado largas; no se admite la posibilidad de este tipo de rachas en una sucesión aleatoria.

Toohey quedó sorprendido al ver mejores respuestas en los niños más jóvenes, entre los cuales casi la mitad seleccionó el patrón aleatorio correctamente. En general encontró que los niños considerados más inteligentes seleccionaban con mayor frecuencia que sus compañeros el patrón determinista.

Estos resultados fueron replicados y completados por Batanero y Serrano (1999) en estudiantes universitarios. Ellos sugieren que los alumnos atribuyen diferentes significados a la aleatoriedad y algunos de ellos coinciden con los admitidos en diferentes periodos históricos dentro de la estadística descritos por Bennett (1998), por ejemplo:

- Aleatoriedad como inexistencia de causas o causa desconocida;
- Aleatoriedad como equiprobabilidad;
- Aleatoriedad como estabilidad de las frecuencias relativas;
- Aleatoriedad como impredecibilidad.

Batanero y Serrano concluyen que cada una de estas concepciones recoge propiedades parciales del concepto y por ello pueden ser válidas en unas situaciones e incompletas en otras más complejas. También mencionan que es importante que en la clase el profesor presente a sus estudiantes ejemplos variados de situaciones aleatorias para fomentar en ellos la construcción progresiva del concepto.

7. Conclusiones del capítulo

La mayoría de los estudios descritos en este capítulo se enfocan bien a aspectos relacionados con otros objetos que se relacionan con la variable aleatoria (distribución, aleatoriedad), en modelos particulares de distribuciones o bien al aspecto didáctico de la variable aleatoria, con un mayor acento en el análisis de libros de texto. Hay pocas investigaciones, por lo tanto, centradas en los aspectos cognitivo y epistemológico (a los que dedicamos nuestra investigación), con excepción de los trabajos de Ortiz (2002), Tauber (2001) y Alvarado (2007) quienes no se centran específicamente en la variable aleatoria. Observamos también que algunos de los estudios sobre el aprendizaje o enseñanza del teorema central del límite, de las distribuciones muestrales, de las distribuciones de probabilidad y de frecuencias, o de los contrastes de hipótesis, cuyo interés principal no está en la variable aleatoria, describen dificultades y errores que tienen los estudiantes, los textos o el material de apoyo sobre ese concepto.

Así, aunque hay algunos estudios preliminares sobre variable aleatoria y estudios sobre la distribución, el teorema del límite, las distribuciones muestrales, la distribución normal, los contrastes de hipótesis y algunos otros conceptos básicos asociados con la variable aleatoria, la enseñanza misma de la variable aleatoria no se ha estudiado en forma sistemática a nivel universitario. Sin embargo es un tópico que la investigación en didáctica de otros conceptos estadísticos universitarios forzosamente tiene que abordar. De aquí nuestro interés en centrarnos en este concepto.

Nuestra investigación continúa un estudio sobre la variable aleatoria iniciado en Ruiz (2006) y Arteaga (2011) al abordar la investigación desde las dimensiones cognitiva y epistemológica.

En la dimensión epistemológica profundizaremos en la naturaleza del concepto de variable aleatoria a través del análisis del objeto matemático desde la disciplina que lo formaliza y lo justifica, la estadística (Estudio 1 que se presenta en el Capítulo 3) y del estudio de su emergencia histórica (Estudio 2 que se presenta en el Capítulo 4).

En la componente cognitiva planteamos dos estudios complementarios desde el punto de vista de su metodología y de los objetos matemáticos sobre los que se centran, todos ellos relacionados con la variable aleatoria:

En el Capítulo 5 presentamos una exploración cognitiva por medio del método de entrevista clínica a una pareja de estudiantes (Estudio 3). La situación experimental parte de la resolución de un problema en un contexto familiar a los estudiantes en una toma de decisión. A lo largo de la entrevista el investigador explora sus concepciones sobre las ideas de experimento aleatorio, variable aleatoria asociada a dicho experimento, distribución de probabilidad, así como sus dificultades con la representación gráfica de la variable y la solución del problema. La asignación de probabilidades es a priori y laplaciana, por lo cual no se precisa el trabajo con variables estadísticas. La entrevista trata también de aportar conocimiento sobre el grado de comprensión alcanzado por las estudiantes de la noción formal de variable aleatoria como función y la vinculación entre variable aleatoria y distribución de probabilidad.

En el Capítulo 6 presentamos un estudio de evaluación llevado a cabo sobre una muestra mayor de estudiantes, y a través del análisis del contenido de sus producciones escritas en un proyecto de análisis de datos. Dicho proyecto parte de una situación experimental propuesta en la clase por el profesor, donde se pide a los estudiantes analizar las intuiciones del grupo clase sobre la aleatoriedad. La asignación de probabilidades será frecuencial. Los estudiantes realizan individualmente un experimento similar a los descritos en el apartado 6.1 de este capítulo sobre la generación de resultados aleatorios. A partir de los resultados que cada uno obtiene, la clase genera un conjunto de datos que cada alumno analizará individualmente y reportará en un informe escrito. La evaluación de las producciones de los estudiantes se centrará sobre todo en el uso de las variables y distribuciones estadísticas, así como de sus promedios, representaciones gráficas y dispersión que cada estudiante haga para realizar su proyecto.

Capítulo 3.

Estudio 1.

Análisis Epistemológico desde la Disciplina

ÍNDICE DE CAPÍTULO

1. Introducción

El análisis de la variable aleatoria desde la disciplina indaga sobre la naturaleza e interrelaciones del objeto matemático que se estudia. Se trata de delimitar el *saber sabio,* desde la perspectiva de Chevallard (1985) o el *significado de referencia* desde el marco teórico de Godino (1996; 2002), desde la perspectiva de la ingeniería didáctica. Este análisis permitirá fundamentar el resto de nuestra investigación, a la vez que podrá, por sí mismo, ser utilizado en futuras ingenierías didácticas sobre la variable aleatoria. Lo consideramos una primera aproximación que nos permitirá detectar puntos sobre los que será necesario profundizar en futuros estudios.

Para llevarlo a cabo tomamos libros que consideramos de referencia para nosotros tomando en cuenta que nuestra intención en los cursos universitarios no está enfocada a la enseñanza de una teoría rigurosa de probabilidades. En primera instancia recurrimos a libros de probabilidad, cuyo propósito principal está en la enseñanza de la teoría de probabilidades en niveles universitarios o de postgrado. Cuatro de ellos, Meyer (1989), Feller (1989), Wackerly, Mendenhall y Scheaffer (2002) y Devore (2011), tratan de presentar, junto con el desarrollo de la teoría probabilística, aplicaciones, a la misma probabilidad, a la estadística o a diversas áreas de la ingeniería y a sí mismos se califican de no rigurosos. El resto de los libros: Mood y Graybill (1963), Krickeberg (1973), Cuadras (1999) y Petrov y Mordecki (2002) describen la teoría matemática desde una perspectiva más rigurosa y sin más aplicaciones que las propias de la misma teoría. Uno más, Ríos (1967) presenta una perspectiva diferente. Su enfoque más bien va hacia la estadística y la probabilidad la maneja desde una perspectiva útil para esta última.

Los libros que utilizamos para este análisis tienen en común que se presentan a sí mismos como adecuados para un primer curso de teoría de probabilidades o estadística matemática, no presuponen conocimientos previos dentro de la probabilidad o conocimientos muy profundos de la teoría de la medida, más bien sólo piden como requisito profundización en el análisis matemático.

Recurrimos también a otros libros que añaden otros objetivos diferentes a la propia exposición del tema y que sin embargo proporcionan elementos importantes para concluir nuestro análisis. Nos referimos al libro de Boudot (1979) de corte filosófico y al de Godino, Batanero y Cañizares (1996) que trata de mostrar los temas de azar y probabilidad desde una perspectiva didáctica. De este último tomamos elementos que completan sólo la perspectiva epistemológica.

Son diversos los problemas relacionados con la variable aleatoria, desde hacer una predicción, evaluar una hipótesis, determinar factores que influyan en una cierta magnitud o tomar una buena decisión, hasta describir una cierta característica en una población. A partir de ellos se derivan otros problemas de tipo matemático, tanto en probabilidad como en estadística, como la estimación de parámetros, las pruebas de hipótesis y la obtención de las distribuciones muestrales, o la definición de nuevas variables aleatorias a partir de una o más variables aleatorias, como la definición de suma, multiplicación, máximos y mínimos, o incluso la composición de variables aleatorias.

La solución matemática a estos problemas pasa por definir cuál es la variable aleatoria vinculada a un fenómeno aleatorio. Pero aparecen ligados otros muchos conceptos (parámetro, estadístico, probabilidad, distribución, función), así como un lenguaje específico (verbal, simbólico y gráfico), propiedades, definiciones y argumentos específicos. Consideramos así, la variable aleatoria no sólo como un concepto matemático, sino como una configuración de objetos matemáticos y a la vez un instrumento de resolución de problemas en el que se requiere encontrar la regla de correspondencia que asigne valores numéricos a los resultados de un experimento aleatorio, que cumpla con ciertos criterios matemáticos y que, a su vez, esté vinculada con un contexto real.

2. La variable aleatoria

El tema que nos ocupa parte de una situación aleatoria (experimento), cuyo resultado se puede valorar mediante una medida cuantitativa (usualmente referida a una cierta magnitud). Pero ante la imposibilidad de predecir con exactitud cada resultado en particular, el estudio de un experimento requiere del análisis del espacio muestral que nos brindará la posibilidad de establecer un criterio basado en las probabilidades de todos los resultados posibles.

Esto es, en algunas situaciones «para caracterizar el resultado de un experimento, no basta con decir que se ha producido un determinado suceso, sino que es preciso dar cuenta de las diversas medidas que se hayan efectuado. La variable aleatoria, es decir, la magnitud que ‘depende del azar’, es entonces indispensable» (Boudot, 1979, p 332). La variable aleatoria es la herramienta matemática que permite pasar del estudio de sucesos aislados al estudio de las *distribuciones de probabilidad*, que son funciones reales y por lo tanto, hace posible la aplicación del análisis matemático y de otras herramientas matemáticas a la estadística (Batanero, 2001).

En palabras muy simples, la variable aleatoria es una función que asocia un valor numérico a cada evento del espacio muestral asociado a un cierto experimento aleatorio (Walpole, Myers y Myers, 1999). Una vez en el contexto numérico, y por intermedio de la probabilidad definida en el espacio muestral, es factible trabajar con las probabilidades de los valores que adopta la variable aleatoria en lugar de con las probabilidades de los sucesos. Esto significa que este concepto tan aparentemente sencillo transforma los sucesos a términos numéricos y permite modelar la relación del espacio muestral mediante la distribución de probabilidad en forma funcional. Pero su definición formal no resulta fácil de interpretar por lo que nos detendremos en ella.

Una definición formal de variable aleatoria se reproduce a continuación. A fin de establecer antecedentes y nomenclatura, antes se define *espacio probabilístico*.

Sea Ω, el espacio muestral asociado a un cierto experimento aleatorio, $\mathcal{A}$ un álgebra de sucesos definida en dicho espacio muestral y P una medida de probabilidad definida sobre A, es decir una aplicación:

$$\begin{array}{cccc} P: & \mathcal{A} & \rightarrow & [0,1] \\ & \mathrm{A} & \rightarrow & P(\mathrm{A}) \end{array}$$

tal que se cumplen los tres axiomas:

1. $0 \leq P(A) \leq 1$, para todo suceso A del álgebra de sucesos A
2. $P(\Omega)=1$
3. Si $(A \cap B) = \phi$, entonces $P(A \cup B) = P(A) + P(B)$[1]

En las condiciones anteriores se dice que $(\Omega, \mathcal{A}, P)$ es un *espacio de probabilidad* o espacio probabilístico.

Sea ahora $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y R el cuerpo[2] de los números reales. Se dice que la aplicación:

$$\begin{array}{cccc} \xi: & \Omega & \rightarrow & R \\ & \omega & \rightarrow & \xi(\omega) \in R \end{array}$$

que a cada suceso elemental hace corresponder un número real, es una variable aleatoria[3] si para todo número real x, se verifica la relación:

(1) $$\mathrm{A} = \{\omega \mid \xi(\omega) \leq x\} \in \mathcal{A}$$

es decir, se verifica que A es un suceso. A se indica abreviadamente por $[\xi \leq x]$. (Cuadras, 1999, p 71-72).

[1] El axioma 3 se generaliza para cualquier número de sucesos incompatibles dos a dos.
[2] En algunos libros se le denomina *campo*.
[3] Algunos autores las denominan *magnitudes aleatorias*.

La condición (1) indica, en palabras, que la imagen inversa en la aplicación «variable aleatoria» para todos los intervalos reales acotados superiormente es medible, puesto que, para cada conjunto del álgebra de sucesos está definida la probabilidad. Así, la definición matemática de la variable aleatoria exige que la imagen inversa de todo $\xi^{-1}(\omega)$ sea un elemento del álgebra $\mathcal{A}$, porque una vez definida una medida de probabilidad *P* sobre $(\Omega, \mathcal{A}, P)$, la variable aleatoria puede determinar una medida de probabilidad sobre $(R, \mathcal{B})$, en donde $\mathcal{B}$ es la sigma-álgebra construida por los conjuntos de Borel en *R*. De esta forma, la variable aleatoria induce una medida normada sobre los conjuntos que representan a los sucesos (Godino, Batanero y Cañizares, 1996).

En realidad la variable aleatoria está definida para todo suceso del álgebra de sucesos, y no sólo para los puntos muestrales (elementos del conjunto Ω). Es decir se trata de una aplicación de $\mathcal{A}$ en *R*, lo que garantiza que la imagen inversa de cualquier elemento del conjunto imagen pertenezca a $\mathcal{A}$ y por tanto podamos posteriormente calcular su probabilidad, que estaba definida previamente sobre $\mathcal{A}$.

Existen experimentos aleatorios en los que aparentemente la variable aleatoria está implícita en los puntos muestrales. Por ejemplo, el experimento consiste en observar el tiempo de espera a un autobús, el número de llamadas telefónicas que espera un conmutador o la medida de la altura de una mujer. En tal caso $\omega = \xi(\omega)$ porque en ellos el espacio de descripciones muestrales de cada uno de estos experimentos aleatorios es un conjunto de números reales. Sin embargo, generalmente a partir de un mismo experimento aleatorio se pueden definir diferentes variables aleatorias o bien a partir de operaciones algebraicas o analíticas entre variables (en donde se pueden incluir variables no aleatorias) se generan nuevas variables aleatorias.

Por ejemplo, en el experimento aleatorio de extraer de una urna el nombre de un trabajador de una fábrica (problema tratado en el Estudio 3, Capítulo 5), la variable aleatoria podría ser el «número de miembros de la familia del trabajador ganador» pero también podría ser el «número de años que tiene laborando en la fábrica el trabajador ganador» o cualquier otra característica dependiendo del motivo por el que se está interesado en efectuar esa extracción.

Así mismo, a partir de las repeticiones del experimento aleatorio del lanzamiento de una moneda se puede generar la variable aleatoria «número de caras obtenidas en una serie da lanzamientos» y no solo preguntarnos por la *variable* «si se obtuvo cara o no en el lanzamiento» (este contexto es el usado en el Estudio 4, Capítulo 6). Entonces, la definición de la variable aleatoria está condicionada por dos contextos, el matemático y el dado por el problema. Es en este sentido que para Miller (1998) es tan importante hacer la distinción sobre la forma en que se define a la variable aleatoria en el discurso escolar[4]. Él alerta sobre las dificultades posteriores que puede traer consigo confundir el resultado del experimento con la variable aleatoria o el espacio muestral del experimento con el conjunto de valores que puede tomar la variable. Algunas de las dificultades posteriores a las que se refiere Miller se reportaron en Sánchez, Albert y Ruiz (2011).

En este sentido, una propuesta didáctica interesante es la que hace Parzen[5] (1971). Él plantea introducir la variable aleatoria y «toda la abrumadora cantidad de nociones que hay que introducir simultáneamente» (p. 8) a través de fenómenos aleatorios con resultados numéricos (p. 172). En capítulos posteriores, Parzen define el objeto variable aleatoria y aclara que en los fenómenos aleatorios con resultados numéricos, ésta es la función identidad y pone énfasis en que «hay que aprender a reconocer y formular matemáticamente como *funciones* los objetos descritos verbalmente que sean variables aleatorias» (p. 300). De este modo Parzen enfatiza en esclarecer la forma en que se constituye la regla de correspondencia que define a la variable aleatoria y su papel como una función.

3. La variable aleatoria como función

La condición (1) que se explicita en la definición de variable aleatoria, para el caso discreto se puede traducir como $X^{-1}(\omega) \in \mathcal{A}$ y significa que el conjunto de los elementos de Ω que dan a X un valor determinado deberá pertenecer al álgebra de sucesos $\mathcal{A}$ y por lo tanto que su valor de probabilidad también estará definido, en caso contrario X no podría ser una variable aleatoria sobre $\mathcal{A}$ (Boudot, 1979). Nos auxiliaremos del esquema mostrado en la Figura 3.1 para glosar esta definición.

La variable aleatoria, como toda función matemática, tiene tres componentes: El *espacio muestral* asociado al experimento (dominio); la *regla de correspondencia*, que definirá la forma en que se vinculan el espacio muestral y el número real (perteneciente al conjunto imagen); y el conjunto de *valores numéricos* (imagen), que toma la variable aleatoria. Para cada evento del espacio muestral (dominio), el valor numérico será asignado mediante la regla de correspondencia que define a la variable aleatoria.

[4] La preocupación de Miller (1998) por la definición de variable aleatoria como «el resultado de un experimento aleatorio» se describe de manera más completa en el punto 5.1 del Capítulo 2.
[5] La propuesta de Parzen será tratada más ampliamente en el apartado 6.2 del capítulo 4 de esta misma memoria.

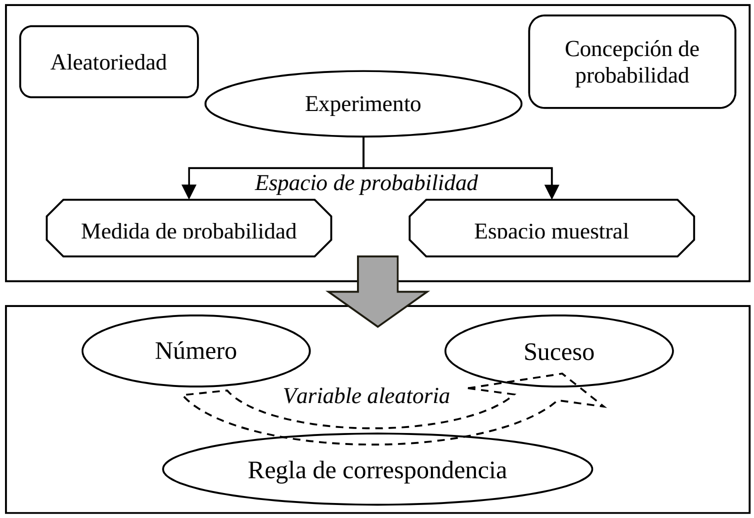


Figura 3.1. Interacciones asociadas a la definición de variable aleatoria

La imagen de la variable aleatoria es un conjunto de números reales relacionados con el espacio muestral, pero que además cumplen con todas las propiedades de campo del conjunto de los reales y que por lo tanto hacen factible establecer una relación funcional (la función de distribución) entre R y el intervalo [0,1]. La imagen de la variable aleatoria, cumple ahora, en la función de distribución, el papel de *variable independiente* y la probabilidad asignada a cada uno de estos valores, el de *variable dependiente* con la que es posible hacer uso de la herramienta del análisis. De modo que los distintos roles que juega la variable aleatoria también podrían generar confusión en los estudiantes.

Así mismo, es notorio que una vez definida la variable aleatoria, la función de distribución también queda definida. En realidad el objetivo final del procedimiento vinculado a la variable aleatoria es la definición de esta función porque ella es el modelo útil para realizar el análisis de la situación problema. Es por ello que la variable aleatoria es inseparable de ella. Los problemas de enseñanza y aprendizaje de la variable aleatoria repercutirán en la comprensión de la función de distribución y viceversa.

4. Función de distribución de una variable aleatoria

La relación (1) hace factible establecer una relación funcional (*función de distribución*)[6] entre el conjunto de números reales y el intervalo [0,1]. Mediante esta función podemos

[6] Los autores de los libros estadísticos consultados difieren en la terminología utilizada. La *función de distribución* de una variable aleatoria se denomina a veces *función de distribución acumulada*.

asignar a cada elemento x de R la probabilidad del subconjunto del espacio muestral; aquél al que la variable aleatoria asigna un valor menor o igual al número x dado. Es decir:

$$F(x) = P(X \le x) = P\big(\omega / X(\omega) \le x\big) \qquad (2)$$

Este proceso se ilustra en la Figura 3.2. El conjunto imagen de la variable aleatoria cumple ahora el papel de conjunto original (o dominio) en la función de distribución, que es, por tanto, una función real de variable real. Se obtiene componiendo la aplicación inversa de la variable aleatoria en el conjunto de valores acotados por x con la probabilidad (que está definida para dicho conjunto). La inversa de la variable aleatoria tiene como conjunto imagen el álgebra de sucesos, que es el conjunto original de la probabilidad (que es una función de conjunto con imagen real). Por ello es posible componer estas dos aplicaciones.

Dicho de otro modo, mientras que la variable *x*, que representa el resultado de aplicar una variable aleatoria $\xi(\omega)$ en un experimento, era la *variable dependiente* en la función de conjunto (variable aleatoria), jugará el papel de *variable independiente* en la función numérica (función de distribución). El valor numérico $F(x)$ de la probabilidad asignada al valor real x imagen de la variable aleatoria, que es un número en el intervalo [0.1], sería la variable dependiente en la función de distribución.

$$\begin{array}{llll}
R: & \xrightarrow{\;X^{-1}\;} \mathcal{A} & \xrightarrow{\;P\;} & [0,1] \\
x: & \longrightarrow \mathbf{A} = \{\omega \mid X(\omega) \le x\} & \longrightarrow & \mathrm{P(A)} \\
x: & \xrightarrow{\qquad\qquad F \qquad\qquad} & & F(x)
\end{array}$$

Figura 3.2. Asignación de probabilidad a los valores de la variable aleatoria

Con la función de distribución es posible trabajar las operaciones usuales en las funciones de variable real. Además, podemos extender la función a todos los números reales (y no sólo a los valores que define la variable aleatoria sobre el espacio muestral), definiendo $F(x)=0$ para x menor que el mínimo de la variable aleatoria y $F(x)=1$ para x mayor que el valor máximo. Intuitivamente esto indica que la probabilidad de que la variable aleatoria tome un valor menor que el mínimo es igual a cero y la probabilidad de que la variable aleatoria tome un valor inferior o igual al máximo es igual a uno. Esto es, la variable aleatoria puede tomar valores que no están estrictamente definidos a través de la variable de interés ni como una aplicación de la función variable aleatoria sobre el espacio muestral del evento aleatorio.

Los tres axiomas principales que deben cumplir las funciones de distribución son:

1. La función debe ser monótona no decreciente. En el sentido que se cumple que $F(a) \leq F(b)$ si $a < b$.
2. Los límites al infinito positivo y negativo de la función deben existir y ser: $\lim_{x \to -\infty} F(x) = 0$ y $\lim_{x \to \infty} F(x) = 1$.
3. Es continua por la derecha. En el sentido de que $\lim_{b \to x^+} F(b) = F(x)$.

Con estas tres condiciones se garantiza que F proporcione un valor de probabilidad para cualquier conjunto de números reales que pueda tomar la variable aleatoria x. A esta propiedad, Parzen (1971) la denomina «probabilizable». Dentro de esta investigación no es nuestra intención caracterizar a los conjuntos reales «probabilizables», solamente diremos que gracias a esa propiedad, la función de distribución queda definida de manera congruente dentro del conjunto de los números reales.

El conjunto imagen de una variable aleatoria puede ser *discreto*, cuando toma un número finito o infinito numerable de elementos o *continuo*, si toma un número infinito no numerable de elementos. Las variables aleatorias definidas sobre espacios muestrales discretos se llaman discretas y las definidas sobre espacios muestrales continuos se llaman continuas. Generalmente el tratamiento del tema es a través de esta división de las variables aleatorias porque ello condiciona la forma en que se obtienen las probabilidades.

4.1. Distribución de probabilidad de la variable aleatoria discreta

En el caso de la variable aleatoria discreta podemos calcular la probabilidad de que la variable tome un valor aislado. La distribución de probabilidad[7] quedaría definida de una manera muy parecida a como se obtiene la función de distribución, pero haciendo uso sólo de la igualdad, puesto que en este caso la función inversa de la variable aleatoria arrojaría como resultado un evento compuesto en el espacio muestral.

$$p(x_i) = P(X = x_i) = P(\omega / X(\omega) = x_i)$$

Es posible, sin embargo, también obtenerla haciendo uso de la función de distribución. Bastaría para ello, hallar la diferencia entre el valor de la función de

[7] Algunos autores la denominan función de probabilidad, pero usaremos el término *distribución de probabilidad* para diferenciarla de la probabilidad definida sobre el espacio muestral que es también una función de probabilidad, pero de conjunto.

distribución para dos valores consecutivos de la variable. Para cada valor x_i, podemos definir una nueva función, nuevamente de variable real, con imagen en el intervalo *[0, 1]* en la forma dada en la siguiente expresión.

$$p(x_i) = F(x_i) - F(x_{i-1}) \tag{3}$$

El valor $p(x_i)$ no sería más que la probabilidad de que la variable tome el valor x_i. La definición formal de la distribución de probabilidad es:

Sean $x_1, x_2, x_3, \ldots x_n$ los distintos valores que puede tomar la variable aleatoria discreta y $p(x_1), p(x_2), \ldots p(x_n)$ la probabilidad con que toma estos valores. El conjunto de pares de valores $(x_j, p(x_j))$ de una variable aleatoria discreta se denomina distribución de probabilidades de la variable aleatoria. La aplicación que a cada valor x en una variable aleatoria discreta le asigna el valor $p(x)$ es una función y debe cumplir con las siguientes propiedades:

1. $0 \leq p(x_j) \leq 1$: p(x) es una probabilidad, y toma valores entre 0 y 1.
2. $\sum_{j=1}^{n} p(x_j) = 1$: la suma de probabilidades repartidas entre todos los valores de la variable debe ser igual a 1.

De la misma forma que se calculan frecuencias acumuladas, podemos sumar probabilidades, obteniendo la *función de distribución de probabilidades*. La función de distribución para estas variables se obtiene en una forma alternativa mediante la expresión:

$$F(x) = \sum_{1}^{j} p(x_j) \tag{4}$$

4.2. Función de densidad de las variables aleatorias continuas

En el caso de variables aleatorias continuas, la probabilidad asociada a un suceso elemental es siempre igual a cero, por lo que no tiene interés considerar directamente la distribución de probabilidad (que en tal caso sería constante). Así, en lugar de trabajar con la probabilidad de valores particulares de la variable, resulta más apropiado trabajar directamente con la función de distribución o bien calcular probabilidades asociadas a intervalos. Para esto último se usa una función que mide «concentración» de probabilidades alrededor de un punto, que se denomina *función de densidad de* *probabilidad* y se denota como *f(x)*. Una función de densidad de probabilidad debe cumplir con las siguientes propiedades:

1. $f(x) \geq 0$ (la función es no negativa para cualquier valor de x, f(x) no es una probabilidad, y puede valer más de 1).
2. $\int_{-\infty}^{\infty} f(x)dx = 1$ (la acumulada para todos los valores de la variable suma 1, el área bajo la curva de la función vale 1).

La función de distribución para una variable aleatoria continua se calcula mediante la relación que liga con la función de densidad, que es la derivada de la función de distribución (para el caso de variables aleatorias continuas), como se muestra en la expresión (5).

$$F(a) = P(X < a) = \int_{-\infty}^{a} f(x)dx \tag{5}$$

La probabilidad de que la variable esté dentro de un intervalo *[a, b]* en una variable aleatoria continua, también se calcula como diferencia de valores de la función de distribución en sus extremos, mediante la expresión (6).

$$P(a < X < b) = F(b) - F(a) \tag{6}$$

Como consecuencia, la probabilidad de que la variable tome un valor particular es nula, como se expresa en (7).

$$P(X=c) = P(c < X < c) = F(c) - F(c) = 0 \tag{7}$$

Esto explica matemáticamente que para el caso de una variable aleatoria continua no tiene sentido trabajar con la probabilidad de un valor particular.

5. Otros conceptos relacionados

A partir de los conceptos anteriores se desprenden otros relacionados con la variable aleatoria, la mayoría de ellos más vinculados con su función de distribución. Entre otros, los momentos de la variable (medidas de posición central y dispersión) y los modelos de funciones que describen las familias de distribuciones de probabilidad (como la Binomial, Poisson, Uniforme o Normal) así como sus respectivos parámetros. Estos conceptos no sólo son importantes desde una perspectiva epistémica, Heitele (1975) propuso que la dispersión, la centralización y la idea de distribución, como las componentes esenciales de la variable aleatoria.

5.1. Promedios

Los promedios evalúan la tendencia central de los datos y se manejan desde dos visiones, muestral y poblacional. La diferencia entre ambas es que en la primera se trabaja con datos o con conglomerados de datos resumidos generalmente en una tabla y no se trabaja con el total de datos de la población. El cálculo puede ser muy simple porque se limita al manejo directo de los datos. La visión poblacional, en cambio, involucra al total de la población, que puede no ser numerable o finita. Bajo esta visión, cuando en el problema de interés se involucra un fenómeno aleatorio, entonces se puede definir un espacio de probabilidad y por lo tanto, también una variable aleatoria y su función de densidad, para el caso continuo, o distribución de probabilidad, para el caso discreto. En las variables aleatorias discretas finitas el cálculo de los promedios puede resultar muy parecido al procedimiento que se sigue en el análisis de una muestra, pero la interpretación es completamente diferente. Es por eso que es importante que se tenga claro la visión desde la que se obtienen los promedios.

Por el objetivo de nuestro estudio, pondremos más énfasis en los promedios de la variable aleatoria, sin embargo no hay que perder de vista la complejidad epistémica que añade a este concepto el que los cálculos de los promedios puedan ser de la población o de la muestra. Algunas de las confusiones que esto genera en los estudiantes están reportadas en De la Cruz (2007).

Los promedios más socorridos son la esperanza matemática, la moda y la mediana. La *esperanza matemática* es un concepto vinculado a la variable aleatoria y que extiende la idea de media muestral. Para las variables aleatorias discretas se define como:

Sea (x_i, p_i) donde $i \in I$, la distribución de probabilidad de una variable aleatoria discreta. Se define la *media o esperanza matemática* como:

$$E(X) = \sum_{i \in I} x_i p_i$$

Para el caso de variables continuas la esperanza matemática está dada por:

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx$$

La esperanza matemática en las variables discretas finitas es una extensión de la media ponderada. El procedimiento de cálculo sería muy similar en el caso que la variable discreta sea finita y que los datos de la muestra estuvieran agrupados en una tabla de frecuencias relativas.

La *moda* es el valor más probable de la variable. Viene dada por los valores de X para los cuales la distribución de probabilidad (variable discreta) o la función de densidad (variable continua) tiene un máximo. En el caso continuo el valor más probable en realidad es un conjunto de valores dados por un intervalo alrededor del máximo de la función de densidad. También es posible que se presente más de una moda. Otra cuestión importante es el hecho de que en un experimento aleatorio la moda sea el valor más probable, no significa que a largo plazo o a nivel masivo se espere que ese sea el valor esperado. Ese papel preponderante lo tiene la esperanza matemática.

La *mediana* es el valor de la variable para el cual la función de distribución toma el valor 1/2. Por tanto, la probabilidad de que una variable aleatoria tome un valor menor o igual a la mediana es exactamente 1/2. Sería la solución de la ecuación:

$$F(a) = P(X < a)=1/2$$

Igualmente se pueden definir las medidas de posición, que son los percentiles, deciles o cuartiles. Por ejemplo el percentil de r% es la solución de la ecuación:

$$F(a) = P(X < a)=r/100 \quad (0 \leq r \leq 100)$$

Esta forma de calcular las medidas de posición requiere la definición de la función de densidad o la distribución de probabilidad. Los cálculos de los percentiles de muestras requieren supuestos y teorías en las que no profundizaremos aquí. La mediana, que además de ser una medida de tendencia central también se puede considerar una medida de posición, no es tan problemática por la facilidad que hay de dividir en dos una muestra o una población discreta finita. El procedimiento para obtener la moda y la mediana puede no diferir demasiado del efectuado para una variable aleatoria discreta finita.

5.2. Medidas de dispersión

La variabilidad siempre se mide con respecto a las medidas de posición central, generalmente con respecto a la media. Al igual que los promedios, las medidas de dispersión se calculan de manera diferente y tienen diferentes connotaciones en una muestra, que en una población. Algunas de las medidas de dispersión más importantes en una variable aleatoria son el rango, que es la diferencia entre el máximo y el mínimo, y el recorrido intercuartílico, que es la diferencia entre el tercer y primer cuartil o lo que es lo mismo entre el percentil del 75 y 25%. Pero las medidas más importantes desde la teoría de probabilidades son la varianza y la desviación estándar. El cálculo para la obtención de las dos primeras es el mismo para una muestra que para una población, pero la varianza se calcula de manera diferente para una población que para una muestra. La varianza de una población en la que se involucra un fenómeno aleatorio es la varianza de una variable aleatoria. La desviación estándar es la raíz cuadrada de la varianza para ambos casos.

La varianza de una variable aleatoria discreta y continua respectivamente está definida por:

$$Var[X] = \sum_{i=1}^{n} (x_i - E(X))^2 p_i$$

$$V[X] = \int_{-\infty}^{\infty} (x - E(x))^2 f(x)dx$$

La varianza puede interpretarse como el momento de inercia de una distribución, mientras que la media sería el centro de gravedad.

Tanto las medidas de posición central como las de dispersión de una variable aleatoria pueden confundirse con la obtención de esas mismas medidas en una muestra si no queda claro si el estudio es sobre una muestra o sobre una población o bien, si no quedan claras las diferencias de cálculo entre unas y otras.

5.3. Momentos

La media y la varianza son casos particulares de los momentos de la distribución de una variable aleatoria. La media es un momento ordinario de orden 1 y la varianza es un momento de orden 2 con respecto a la media.

En general, los momentos ordinarios se definen de la siguiente forma:

$$m_r = \sum_{i=1}^{n} (x_i)^r p_i \quad \textit{para las variables discretas}$$

$$m_r = \int_{-\infty}^{\infty} x^r f(x)dx \quad \textit{para las variables continuas}$$

Los momentos con respecto a la media se definen como:

$$\mu_r = \sum_{i=1}^{n} (x_i - E(X))^r p_i \quad \textit{para las variables discretas}$$

$$\mu_r = \int_{-\infty}^{\infty} (x - E(X))^r f(x)dx \quad \textit{para las variables continuas}$$

Otros momentos importantes son los de orden 3, que sirven para definir los coeficientes de asimetría, y los de orden 4, que sirve para definir la curtosis. Otro motivo por el cual son importantes los momentos es que la función generatriz de una variable aleatoria puede desarrollarse en serie y los coeficientes del desarrollo son precisamente los momentos. De modo que el desarrollo en serie hasta un cierto orden usando los momentos proporciona una aproximación a la función generatriz, que a su vez determina en forma unívoca la función de distribución.

6. El orden en que son presentados los conceptos

En los libros consultados (con excepción del Parzen, 1971) la presentación de los objetos matemáticos más o menos corresponde al orden expuesto en esta memoria, que es coherente con la axiomática de la probabilidad: se define la variable aleatoria, después la función de distribución y a partir de ellas surgen las distribuciones de probabilidad (variable discreta) y la función de densidad (variables continuas). Posteriormente dan paso a la esperanza matemática, varianza y momentos de algunas variables aleatorias.

Sin embargo algunos libros (Krickeberg, 1973; Wackerly, Mendenhall y Scheaffer, 2002; Devore, 2011) inmediatamente después de la definición de la variable aleatoria (o al mismo tiempo) introducen la distribución de probabilidad de una variable aleatoria discreta, luego la definición y cálculo de la esperanza, varianza y momentos de una variable aleatoria discreta y finalmente dan paso a la variable aleatoria continua, su distribución y momentos. A diferencia de Krickerberg, Devore propone problemas más aplicados y su intención no está enfocada a un curso de probabilidad, de modo que el tema que nos ocupa es tratado con una mayor libertad y con menos herramientas teóricas, pero obedeciendo al mismo orden de presentación.

Parzen (1971) propone otro orden diferente. Él define la función de distribución de un fenómeno aleatorio con resultados numéricos, su esperanza, varianza y momentos e introduce algunas funciones de distribución más conocidas (todas las que se trabajan en un curso de probabilidad y estadística para Ciencias Sociales en el nivel universitario y la mayoría de las que se manejan en el mismo curso para ingeniería) y posteriormente define el concepto de variable aleatoria. En Parzen (1971), las funciones de distribución posteriores a este tema involucran varias variables o funciones compuestas.

Ríos (1967), propone un seguimiento muy parecido al de Krickerberg y Devore, pero rescata la asignación de probabilidades frecuentista y la definición de *variable estadística*. Él define la variable aleatoria como una generalización de la variable estadística para después continuar con las distribuciones discretas y continuas. Previamente se ocupó de la definición de variables cuantitativas y de las medidas de centralización y dispersión y momentos. Meyer (1989) propone un orden parecido al que usamos en esta memoria y se ocupa del estudiante con pequeñas notas al pie de página en donde advierte diversas dificultades con las que un estudiante se puede enfrentar al tocar el tema.

Tanto Ríos (1967), como Parzen (1971), Krickerberg (1973), Meyer (1989) y Devore (2011) defienden sus propuestas de orden de los temas tratados basándose en visiones diferentes del mismo tema, enfocados a diferentes objetivos, pero cada uno de ellos propone su propia visión como la mejor opción para introducir a los estudiantes en el tema sin pedirles grandes antecedentes matemáticos de teoría de la medida o análisis matemático. Sin embargo, en ninguno de los casos, sus posturas parten del análisis del conocimiento de los estudiantes, sino más bien un análisis del conocimiento en sí mismo. Así, podemos presumir que cada libro se sustenta en posturas epistémicas diferentes para desarrollar en los estudiantes, aparentemente[8], el mismo tema. Las reflexiones diversas de estos autores, evidencian la complejidad epistémica del tema que nos ocupa.

7. La variable aleatoria y la variable estadística

En los apartados anteriores se introdujo la variable aleatoria directamente a partir de la idea de experimento aleatorio. Otra introducción presentada en algunos textos es a partir de la idea de variable estadística (por ejemplo, en Ríos, 1967). Mientras que la variable aleatoria se refiere a todos los valores posibles (teóricos) de una variable en un experimento aleatorio, la *variable estadística* describe el conjunto de valores obtenidos en los datos al realizar un número *n* concreto de veces el experimento:

> Si consideramos un experimento aleatorio *S* y realizamos un cierto número *n* de pruebas relativas al mismo, obtenemos un conjunto de observaciones, que se llama una *muestra aleatoria de extensión n*. Este conjunto de resultados dará lugar a una tabla estadística en que a unos ciertos valores de la variable corresponden una ciertas frecuencias. A tal *variable, que representa únicamente los n resultados de n realizaciones del experimento aleatorio S la denominaremos variable estadística.* (Ríos, 1967, p. 70. Las cursivas son del texto original.)

La variable aleatoria se concibe como el resultado de la repetición indefinida del experimento:

[8] Nuestra referencia a lo aparente es porque cada uno de ellos lleva distintos objetivos al enunciar la definición de variable aleatoria, como parte del engranaje de la postura conceptual de su libro. Lo cual hace que su forma de abordar la definición, se impregne un poco de su postura aunque en ninguno de ellos deja de ser el mismo concepto matemático.

> Si imaginamos hechas una infinidad de pruebas relativas al experimento *S* la infinidad de resultados posibles da origen a la noción de variable aleatoria asociada al experimento *S*. En este caso la variable aleatoria toma los valores que representan los sucesos elementales posibles de dicho experimento con unas ciertas probabilidades que le corresponden. (Ríos, 1967, p. 70).

El autor explica que los conceptos concretos de muestra, frecuencia y variable estadística, llevan, por un proceso de abstracción, a los de población, probabilidad y variable aleatoria. Habla de la distribución de frecuencias o distribución acumulativa de frecuencias que, al generalizarse, dan lugar a la distribución de probabilidad y función de distribución.

Los momentos, medidas de valor central y dispersión de la variable estadística se denominan *estadísticos*, mientras que en el caso de la variable aleatoria se habla de *parámetros*. Ello es debido a que, en algunos modelos de distribución de las variables aleatorias los momentos (particularmente la media) determinan la expresión algebraica de la función de distribución (actúan como parámetros para una familia de funciones dadas). Los estadísticos se suelen utilizar para *estimar* a los correspondientes parámetros y en estos casos se habla de un *estimador* (por ejemplo, la media de la variable estadística en la muestra se usa para estimar la media de la variable aleatoria en la población).

Sin embargo, hay algunas acotaciones a estas definiciones que merece la pena mencionar. De acuerdo con ellas, la variable estadística se relaciona con una *muestra* de datos, en cambio la variable aleatoria se relaciona con la *población*, aunque en ambos casos se supone la existencia de un «experimento aleatorio *S*» vinculado con la recolección de los datos. Pero esta definición sólo se aplica cuando la asignación de probabilidades es frecuencial, pues en este caso se parte de una variable estadística y es mediante un proceso de inferencia que se llega a la variable aleatoria. Cuando la asignación es clásica, por ejemplo, el espacio muestral puede provenir de un análisis teórico o bien del recuento del total de datos de la población. En tal caso, la definición de variable aleatoria no requeriría del proceso que menciona Ríos. También hay casos en la probabilidad clásica en que sería necesario trabajar con la variable estadística si se plantean problemas de muestreo, como ocurre en el experimento utilizado en el Capítulo 6 en que se parte de una variable aleatoria binomial *B*(20,0.5), número de caras al lanzar 20 veces una moneda, pero se toman muestras de tamaño *n* (en donde *n* es el número de alumnos que realizan el experimento) por lo que aparece la correspondiente variable estadística. También existen variables aleatorias o estadísticas compuestas que se definen a partir de otras variables aleatorias o estadísticas.

Observemos también que Ríos establece la presencia del experimento aleatorio en la conceptualización de ambas variables. Una definición más teórica de variable aleatoria (apartado 2) está de acuerdo en la relación que ella guarda con un experimento aleatorio, lo cual, en algunas ocasiones, no ocurre con la variable estadística. Estas situaciones serían aquellas en las que se obtiene el total de datos de la población (censo) con objetivos descriptivos. De acuerdo con la definición de variable estadística (dada más atrás) habrá un experimento aleatorio que se usa para recolectar datos, pero en el caso de un censo no hay un muestreo aleatorio, puesto que todos los miembros serán seleccionados. En esta situación la definición de la variable como aleatoria o estadística no depende del número de pruebas sino del uso que se le dé a los datos. Si el objetivo es descriptivo, la variable sería estadística, sin la realización del experimento aleatorio en la recolección de los datos, pero si el objetivo fuera conocer la probabilidad de que un miembro tenga cierta característica, sería aleatoria, puesto que el experimento aleatorio involucrado estaría dado por la selección aleatoria de un miembro de esa población y el espacio muestral es conocido (obtenido empíricamente a través de la recolección total de los datos). Es decir, en esta última situación la recolección de datos no estaría dada por la repetición del experimento aleatorio.

Además, en el caso de un censo, y para una variable discreta, la distribución de probabilidad aparentemente sería la misma que la distribución de frecuencias relativas: su gráfica tendría la misma forma y sus valores serían los mismos, pero conceptualmente serían muy diferentes. Más aún, la gráfica y los valores de la distribución de una variable estadística, conceptualmente cambiarían por completo en el momento en que una persona esté interesada por contestarse alguna pregunta probabilística con respecto a esa población, puesto que esa misma gráfica y los mismos números representarían la distribución de una variable aleatoria. Esto podría generar confusiones en las conceptualizaciones de los estudiantes. En el Estudio 3 se presenta una situación problema en donde se podría generar esta confusión. En ella se proporciona a los estudiantes una tabla con los valores de una variable estadística y su frecuencia correspondiente del total de la población de trabajadores de una fábrica (un censo). Los alumnos tienen que darse cuenta que al predecir el resultado de una rifa entre los trabajadores, se involucra un experimento aleatorio y por lo tanto, las representaciones gráfica y tabular originales, conceptualmente son diferentes ya que ahora corresponderían a la distribución de probabilidad de la variable aleatoria.

En resumen y de manera general, la variable estadística está relacionada con la recolección de una muestra de datos y la variable aleatoria con un conocimiento del espacio probabilístico (obtenido teórica o empíricamente a través de un censo). La variable aleatoria siempre está vinculada con un experimento aleatorio, pero la estadística no necesariamente. La aleatoria está vinculada con las diversas asignaciones de probabilidad, ya sea laplaciana, frecuencial, subjetiva o alguna otra, en cambio la estadística lo está con las frecuencias.

8. La variable aleatoria y asignación de probabilidades

Hemos visto que un concepto ligado al de variable aleatoria, e inseparable de ella, es el de *distribución de probabilidad*, pues la variable determina unívocamente su distribución de probabilidad y viceversa. En consecuencia, el trabajo con la variable aleatoria pasará por la determinación de dicha distribución, lo que nos llevará, en un sentido inverso, a un espacio muestral asociado y a su correspondiente probabilidad. Así, aunado y vinculado a las controversias epistémicas ya enunciadas sobre la variable aleatoria, se sitúa la asignación de probabilidades.

A lo largo de su desarrollo histórico, la probabilidad ha recibido diferentes significados, que aún coexisten en la actualidad e influyen en la forma en que se asigna probabilidades a los sucesos del espacio muestral (y por tanto la forma en que se determina la distribución de probabilidad de la variable aleatoria).

Desde un punto de vista matemático, la probabilidad previamente definida en el espacio probabilístico puede ser cualquier función medible que responda al sistema axiomático teórico ya descrito, que Hawkins y Kapadia (1984) denominan *probabilidad formal* y que proporciona el soporte del cálculo de probabilidades.

Sin embargo, esta definición formal, como señala Azcárate (1995), no mantiene relación con los fenómenos naturales, por ello son otras las concepciones de probabilidad las que permitirán asignar valores a la probabilidad en el experimento y, por tanto, determinar la función de distribución de probabilidad.

Una vez determinada la función de distribución, la vinculación de la probabilidad formal, desprovista de relación con la realidad, se reestablecerá a través de la variable aleatoria, puesto que es ella la que mantendrá la relación con el fenómeno aleatorio y con la variable de interés del problema que se esté resolviendo. Pero a su vez, el trabajo de la variable aleatoria exige desproveer de significado a las distintas concepciones de probabilidad porque de otra forma, no se podrá definir la función de distribución. Así, la variable aleatoria en un sentido, vincula la probabilidad formal con la realidad, pero en sentido inverso, es la que permite la abstracción del modelo probabilístico.

La asignación de probabilidades por sí misma, involucra un problema epistemológico (Godino, Batanero y Cañizares, 1996; Batanero, Henry y Parzysz, 2005) y a la vez introduce nuevos objetos matemáticos. En los siguientes apartados describimos la asignación de probabilidades en las concepciones clásica, frecuencial y subjetiva, que son las utilizadas en las aplicaciones prácticas y reflexionaremos sobre la forma en que ellas influyen en la definición de la variable aleatoria y la función de distribución de probabilidad.

En nuestro trabajo se usará la asignación laplaciana de probabilidad en el Estudio 3. En el Estudio 4 se ven involucradas distintas concepciones. La probabilidad laplaciana estará presente en el experimento que realizan los estudiantes, el análisis de datos recogidos de dicho experimento y la respuesta a la pregunta planteada requerirá de la asignación frecuencial. Incluso podemos considerar que de algún modo interviene una asignación subjetiva cuando los estudiantes inventan sus secuencias aleatorias, ya que estas reflejan sus concepciones subjetivas sobre la probabilidad de los sucesos y secuencias de sucesos en el experimento (Chernoff, 2009b).

8.1. Asignación laplaciana de la probabilidad

En 1814, Laplace dio la definición que hoy enseñamos con el nombre de «regla de Laplace» ó «probabilidad clásica» para la probabilidad de un suceso que puede ocurrir solamente en un número finito de modalidades. Él la definió como *una fracción cuyo numerador es el número de casos favorables y cuyo denominador el número de todos los casos posibles*» (Laplace, 2006/1814, p. 28), indicando también la necesidad de reducir los acontecimientos de un cierto tipo a un cierto número de casos igualmente posibles.

Lo característico de esta concepción de la probabilidad es *el análisis a priori del experimento* para determinar todos los casos posibles y favorables, usualmente mediante un razonamiento combinatorio. Una cuestión importante es que todos los sucesos se consideran equiprobables, lo que plantea un problema filosófico, e hizo que esta definición se encontrase inadecuada incluso en la época de Laplace, ya que es circular y restrictiva puesto que la palabra definida (probabilidad) entra en la misma definición (Godino, Batanero y Cañizares, 1996).

Esta definición permite trabajar con la variable aleatoria sin necesidad de vincularla con la variable estadística, es decir, sin la relación empírica con los datos. Sin embargo, no puede aplicarse a los experimentos con un número infinito de posibilidades, por lo que carece de interés en el caso de las variables aleatorias continuas. Tampoco puede aplicarse a aquellos casos en que el espacio muestral es finito, pero en los que no puede aceptarse la condición de simetría, como al lanzar al suelo una chincheta. Como hace notar Bernoulli (2006/1713) «la equiprobabilidad apenas se encuentra fuera del campo de los juegos de azar».

Sin embargo, una vez asignadas algunas probabilidades iniciales (con algún otro método), es muy útil en la definición de algunos modelos de distribuciones de probabilidad muy conocidos. Por ejemplo, la distribución binomial se define mediante probabilidad clásica, una vez asignado el valor del parámetro *p*. Otros ejemplos serían la distribución geométrica o hipergeométrica. De hecho, algunos autores de libros de texto (Wackerly, Mendenhall y Schaeffer, 2002; Devore, 2011) usan este tipo de distribuciones para introducir el concepto de variable aleatoria por la facilidad con la que, a partir de ella, se deriva esta definición y las sucesivas definiciones.

8.2. Asignación frecuencial de probabilidades

En respuesta a las desventajas de la concepción clásica de la probabilidad, Bernoulli sugirió que se podrían asignar probabilidades a los sucesos aleatorios que aparecen en diversos campos a partir de la frecuencia relativa observada en una serie grande de ensayos del experimento (Bernoulli, 2006/1713). Su demostración de la primera Ley de los Grandes Números, fue aceptada en su época como un apoyo al carácter objetivo de la probabilidad (Batanero, 2005). Este teorema indica que la probabilidad de que la frecuencia relativa de un experimento repetido en las mismas condiciones se acerque tanto como queramos a la probabilidad teórica del suceso, puede aproximarse suficientemente a uno, sólo aumentando el número de pruebas (Bernoulli, 2006/1713).

En esta visión se define la probabilidad como el número hipotético hacia el cual tiende la frecuencia relativa al estabilizarse (von Mises, 1952/1928), asumiendo la existencia teórica de dicho límite, del cual la frecuencia relativa observada es solo un valor aproximado. La probabilidad se estimaría, entonces, a partir de la frecuencia de aparición de un valor de la variable en un número elevado de repeticiones del experimento (valores de la *variable estadística*), no siendo posible la determinación de la probabilidad sólo a partir del análisis a priori del experimento, sino que se requiere la recolección de datos. En esta concepción, por tanto, se conecta la estadística con la probabilidad y la variable aleatoria se concebiría como un modelo teórico de la variable estadística.

El enfoque frecuencial tiene diversos problemas prácticos. El primero de ellos es que sólo se obtiene una estimación del valor de la probabilidad (nunca exacto). Además de que es difícil realizar los experimentos exactamente en las mismas condiciones o determinar el número de experimentos que necesarios para obtener una buena estimación de la probabilidad. Más aún, ciertos sucesos (por ejemplo en el campo de la economía o de la historia) son irrepetibles, aunque aleatorios y según esta concepción no podríamos aplicar la teoría de la probabilidad a su estudio (Batanero, 2005).

Sin embargo, el enfoque frecuencial ha tenido un fuerte impulso en las últimas décadas gracias a la proliferación y masificación de ordenadores capaces de realizar simulaciones de fenómenos aleatorios con un número grande de repeticiones, lo que permite encontrar buenas aproximaciones de modelos de funciones de distribuciones de probabilidad a los que no se tenía acceso de manera teórica, incluso para casos de variable continua. Esto ha hecho necesario que en la educación escolarizada se ponga énfasis en esta concepción que había sido muy ignorada en la práctica docente de décadas anteriores.

Una de sus grandes ventajas es que la probabilidad frecuencial es la que vincula de manera más palpable el experimento aleatorio con la probabilidad (Azcárate, 1995). No obstante, los problemas prácticos a los que se enfrenta este enfoque hacen que la abstracción pueda resultar una dificultad epistémica. En esta concepción, no hay un número de ensayos a partir del cual la variable estadística se aproxime lo suficiente a la variable aleatoria, ni la frecuencia a la probabilidad. Es decir, como no es posible obtener una probabilidad a priori, podría haber una confusión de los conceptos variable estadística y variable aleatoria o bien entre los conceptos de frecuencia y probabilidad. La variable aleatoria y la función de distribución tienen los mismos problemas prácticos de construcción que la asignación de probabilidad.

8.3. Asignación subjetiva de probabilidad

En la concepción subjetiva, la probabilidad es un «grado de creencia racional» que asigna un decisor a cada valor de la variable. Ese concepto de probabilidad surgió porque los conceptos de probabilidad frecuentista o laplaciana no resuelven satisfactoriamente los problemas que no pueden considerarse dentro del modelo del espacio probabilístico del experimento aleatorio. Es necesario notar que las magnitudes de las probabilidades no son estimadas por el individuo arbitrariamente, sino que están sujetas a ciertas reglas de consistencia, establecidas como axiomas que el individuo está dispuesto a aceptar. En las últimas décadas, las teorías subjetivistas han adquirido un notorio interés práctico con la aparición de los métodos bayesianos de estimación.

De acuerdo con el método bayesiano, la regla de Bayes permite transformar una probabilidad «a priori» (subjetiva) en probabilidad «a posteriori» en la que se incorpora información (objetiva) que se ha obtenido a partir de datos observados. Esto es, el interés de este enfoque estriba en la integración de información subjetiva con información objetiva para encontrar la probabilidad de unas causas una vez se observan sus consecuencias.

Supongamos que A_i representa el conjunto de posibles sucesos (causas) en el experimento con probabilidades iniciales $P(A_i)$ y se obtiene una muestra aleatoria de datos D (consecuencias), cuya *verosimilitud* se conoce en función de estos datos $P(D/A_i)$. El teorema de Bayes permite calcular las probabilidades finales de los sucesos $P(A_i/D)$, es decir, las probabilidades de las causas A_i dado que se obtuvieron las consecuencias D:

$$P(A_i/D) = \frac{P(A_i)P(D/A_i)}{P(A_1)P(D/A_1)+P(A_2)P(D/A_2)+\cdots+P(A_n)P(D/A_n)} \tag{8}$$

Las probabilidades iniciales en el experimento $P(A_i)$ podrían entonces actualizarse (pasar de probabilidades a priori a probabilidades a posteriori) y pierden de este modo el carácter objetivo que les asigna la concepción frecuencial. Keynes, Ramsey y de Finetti describieron las probabilidades como grados de creencia personal, basadas en el conocimiento y experiencia de la persona que las asigna sobre el suceso dado. Para ellos la probabilidad de un suceso siempre está condicionada por un cierto sistema de conocimientos y puede ser, por tanto, diferente, para distintas personas (Batanero, 2005).

Una dificultad inicial del enfoque de estimación bayesiana es precisamente la asignación subjetiva de la probabilidad al hallar una regla para asignar valores numéricos a las probabilidades a priori, de forma que expresen los grados de creencia personal, pero que estén sujetos a reglas de consistencia. Son diversos las autores que han propuesto sistemas axiomáticos para darle consistencia a la probabilidad subjetiva, entre ellos Ramsey, De Finetti, Koopman, Good, Savage y Anscombe-Aumann. Sin embargo, también se considera que los datos y la aplicación sucesiva del teorema de Bayes restarán importancia a la asignación inicial. Esto es que, al realizar un nuevo experimento, el valor que se asigna a la probabilidad a priori es el valor que en el experimento anterior se obtuvo como probabilidad a posteriori, así, al aumentar los experimentos, los datos harán converger al mismo valor las asignaciones iniciales.

La visión subjetiva amplía también las aplicaciones de la probabilidad, puesto que ya no se tiene el requisito de la repetición de una experiencia en las mismas condiciones. Gradualmente se desarrolla la distinción entre probabilidad frecuencial, empíricamente accesible a través de la frecuencia, y probabilidad epistémica o grado de creencia en la ocurrencia de un suceso en un experimento único (Batanero y Díaz, 2007), mientras se conforman dos escuelas de estadística.

De importancia para nuestro análisis es que, mientras en estadística clásica un parámetro θ de una distribución de probabilidad se considera constante, en inferencia bayesiana un parámetro θ es una variable aleatoria con una distribución inicial de probabilidades $p(\theta)$, de carácter epistémico puesto que indica el conocimiento (o falta de conocimiento) sobre θ antes de tomar los datos. Si se considera $y = (y_1,..., y_n)$ un conjunto de datos, cuya *función de verosimilitud* $p(y/\theta)$ depende del parámetro, entonces la distribución final de θ dados los datos observados y viene dada por el teorema de Bayes (Batanero y Díaz, 2007):

(9)
$$p(\theta / y) = \frac{p(y/\theta)\,p(\theta)}{\int p(y/\theta)\,p(\theta)}$$

A pesar de que en la estadística bayesiana también se habla de variables aleatorias, no tienen el mismo sentido que en la estadística paramétrica. El hecho de que el parámetro también sea aleatorio hace la comprensión de este concepto adquiera mayor realce. Así mismo, la distribución de probabilidades de una variable aleatoria es en realidad una familia de distribuciones de variables aleatorias.

9. Modelación y variable aleatoria

En las situaciones experimentales que propondremos a los estudiantes en la parte empírica de nuestro trabajo (Estudios 3 y 4) no sólo nos interesaremos por las concepciones que ellos presentan de la variable aleatoria o de otros objetos asociados a la misma, sino que también pretendemos analizar el uso que hacen de la variable aleatoria dentro de un proceso completo (aunque sencillo) de modelación.

La variable aleatoria es un modelo matemático para describir situaciones reales y a su vez, uno de los elementos que contribuyen a la modelación de la función de distribución (inseparable de la variable). De este modo la variable aleatoria es un objeto matemático que describe «la realidad» y se convierte en la «realidad matemática» que se modela mediante la función de distribución.

En este apartado nos detendremos en la modelación como un proceso de enseñanza y aprendizaje que en nuestros estudios usaremos para observar las concepciones y procesos que usan para resolver el problema. Nuestra disquisición sobre modelación empieza con el concepto de modelo y su relación con el concepto de teoría desde una perspectiva científica, posteriormente enfocaremos nuestra atención a la modelación con fines de enseñanza en donde nos detendremos en la vinculación entre el razonamiento estadístico y modelación y la modelación en la enseñanza de la probabilidad. Finalizaremos con la modelación del objeto matemático que nos ocupa, en donde se ven involucrados otros objetos, en particular nos detendremos en la relación que se establece entre la variable aleatoria y la variable estadística en la modelación.

9.1. Modelo y teoría

El concepto de modelo es muy amplio y recibe diferentes acepciones tanto en matemáticas como en educación matemática, como se muestra en el ICMI Study *Modelling and applications in Mathematics Education* (Blum, Galbratith, Henn y Niss, 2007). Conviene, por tanto, detenernos a explicitar el sentido que daremos al término en esta sección.

Bunge (1973) menciona que los modelos son medios para comprender lo que la teoría intenta explicar; enlazan lo abstracto con lo concreto. En un contexto escolar, generalmente los estudiantes pasan de lo concreto (la situación problema en la que el profesor los involucra) a lo abstracto (los modelos y las teorías), para, finalmente, volver a valorar su solución en el contexto del problema. En el caso de la probabilidad y estadística, son muchos los modelos que se manejan y se interrelacionan para llegar a constituir una teoría.

Estamos de acuerdo con Yuren (2006) en que es importante la diferenciación entre el término *modelo* del término *teoría*. Un proceso de investigación científica culmina con la elaboración de *teorías*, es decir, con un sistema que relaciona leyes y que ofrece una explicación de las mismas. Los datos tienen la función de concebir nuevas hipótesis que puedan, en su momento, emplearse o sintetizarse en teorías, pero se obtienen a la luz de otras teorías. La observación y la experimentación se realizan no sólo para recoger información y producir hipótesis, sino también para contrastar las consecuencias de la teoría, o bien para saber cuál es su dominio de validez. La función explicativa y de predicción de la ciencia se realiza en el seno de las teorías; la acción misma se basa en las teorías.

Por un lado, el modelo es una representación de la realidad que supone un alejamiento o un distanciamiento de la misma. Es una representación conceptual, simbólica y por lo tanto indirecta, que al ser necesariamente esquemática se convierte en una representación parcial y selectiva de aspectos de esa realidad, forzando la atención en lo que considera importante y despreciando aquello que no lo es y aquello que no aprecia como pertinente a la realidad que considera (Gimeno, 1984). Por otro lado, la teoría incluye modelos, que describen una zona restringida del campo cubierto por la teoría y éstos la representan justamente mostrando la referencia que hace la teoría de la realidad (Yuren, 2006). Una de las características del modelo es que, a la vez que facilita la comprensión de la teoría, muestra sus aspectos más importantes. Esto es, los modelos científicos abarcan tres significaciones: *representan* a la teoría, *muestran* las condiciones ideales en las que se produce el fenómeno al verificarse una ley o teoría y *constituyen una muestra* particular de la explicación general que da la teoría de la realidad. Es en estos dos últimos sentidos que el modelo vincula la teoría con la realidad (Figura 3.3).

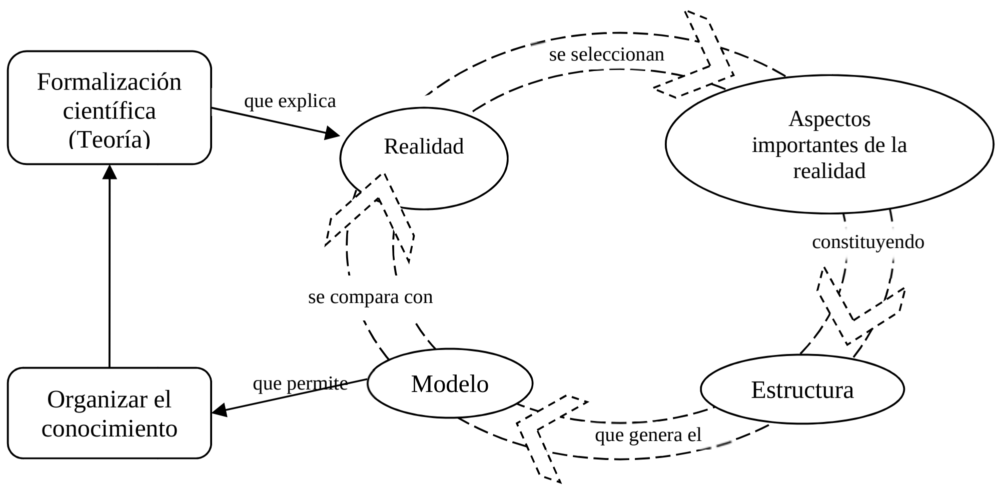


Figura 3.3. Diagrama de la relación entre modelo y teoría

En nuestro campo, los modelos matemáticos son parte de la teoría de probabilidad. La variable aleatoria puede considerarse un modelo porque es el que sustrae la información relevante de la realidad e indica los alcances de la teoría, pero a su vez, es una herramienta para la obtención de un modelo más general que es el que permite la aplicación de la teoría en la situación problemática: la función de distribución. La función de distribución es el objeto que permitirá la solución del problema, pero la variable aleatoria es lo que vincula a la realidad con la teoría.

En este trabajo nos enfocamos a la generación de modelos probabilísticos y estadísticos como un proceso de construcción de conocimiento en la escuela y en la vinculación que, a través de éste proceso, el estudiante puede establecer con la teoría científica ya instaurada. No sólo nos interesamos en cómo un grupo de estudiantes construye conocimiento con ayuda del profesor, sino también por las actividades de la modelación que favorecen que el estudiante se apropie del conocimiento científico. A partir del proceso de construcción y de las teorías adquiridas se pretende que el estudiante genere conocimiento nuevo no sólo para él mismo sino también, a largo plazo, para la sociedad en la que vive. Describir el proceso de modelación que siguen los estudiantes cuando se enfrentan a un problema en el que está en juego el objeto matemático de variable aleatoria y la teoría alrededor de él, es uno de los objetivos de nuestro estudio. Nuestro trabajo, por lo tanto, forzosamente deberá incluir el razonamiento con y a través de la teoría, lo cual conlleva al estudio del razonamiento con modelos probabilísticos y estadísticos, y la modelación como un proceso de enseñanza.

9.2. Modelación y razonamiento estadístico

Para algunos cognoscitivistas, el pensamiento trabaja a través de modelos. Wild y Pfannkuch (1999) describen el razonamiento estadístico como un complejo formado por cuatro componentes, uno de los cuales son las *formas fundamentales* de razonamiento estadístico. Entre ellas incluyen el razonamiento con modelos estadísticos e incluyen a la modelación como uno de los tipos generales de pensamiento comunes en la resolución de problemas matemáticos que también emerge en los procesos de razonamiento estadístico. En estocástica en particular, son de interés los modelos que contemplen componentes aleatorios.

Wild y Pfannkuch proponen a la modelación estadística como una herramienta poderosa para el aprendizaje de nuestro entorno: «El objetivo fundamental de la investigación estadística es el *aprendizaje* en el ámbito del contexto de un problema real» (p. 244). También indican que en el proceso de construcción de un modelo estadístico, el conocimiento estadístico es parte de la acumulación de «comprensión». Wild y Pfannkuch afirman que la modelación es una vía para obtener conocimiento (teoría), incluyendo el conocimiento estadístico. Para ellos, nuestro razonamiento se basa en modelos puesto que el aprendizaje consiste en la construcción de modelos mentales del sistema bajo estudio. Ellos no están interesados en la modelación como un objeto de enseñanza, sino en el aprendizaje que se puede obtener a través de la modelación. En este sentido, la propuesta de Wild y Pfannkuch resulta de gran utilidad para nuestros objetivos de estudio.

Por otro lado, un modelo estadístico para estos autores es un concepto más general que una ecuación e incluiría todos los objetos estadísticos que intervienen en la resolución de un problema, desde la recolección de datos hasta cómo se analizan. Así, los modelos son construcciones cognitivas en el individuo, pero también, objetos abstractos que viven fuera de él y pueden ser menos complejos de lo que en contextos diferentes al educativo se conciben.

El razonamiento estadístico trabaja a partir del conocimiento estadístico, el conocimiento del contexto y la información proporcionada por los datos. En nuestros estudios empíricos asumiremos esta idea, puesto que el trabajo de modelación con la variable aleatoria requiere una síntesis entre el conocimiento del contexto y el conocimiento estadístico en donde la variable aleatoria juega un papel importante: surge del contexto y permite la matematización del problema, como se representa en la Figura 3.4.

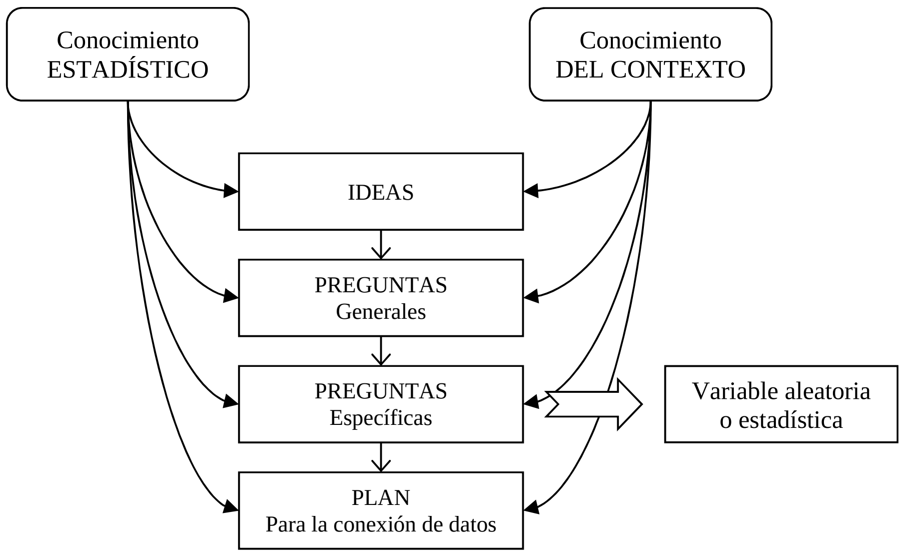


Figura 3.4. Papel de la variable aleatoria en la interacción entre el contexto y la estadística. (A partir de Wild y Pfannkuch, 1999, p. 228)

Como se ha indicado, todos los modelos son simplificaciones de la realidad. Para Wild y Pfannkuch los modelos son formados con datos tomados del contexto real incorporando «conocimiento experto» y en los modelos estadísticos alguna de esa información que se toma de la realidad ya son datos estadísticos (o un modelo extraído de la realidad, como se verá más adelante). Las primeras etapas de una investigación están conducidas casi completamente por el conocimiento del contexto, pero el conocimiento estadístico es el cristalizador del pensamiento en un plan de acción. La formulación de una pregunta específica requiere de teoría estadística para que sea lo suficientemente precisa y que pueda contestarse a través de un análisis de datos. En nuestra investigación el proceso descrito por los autores lleva a construir un modelo de variable aleatoria o estadística porque es en ellas en las que se abstraen las características importantes del fenómeno de estudio tratando de evitar que la pérdida de información no invalide las conclusiones en el contexto del problema. Por tanto, los primeros indicios de trabajo con una variable aleatoria o estadística están en las preguntas específicas que el estudiante se plantea al intentar resolver un problema que darán lugar a la regla de correspondencia que permitirá el trabajo de la variable aleatoria como función.

9.3. La modelación en probabilidad

Nos acogemos en la idea de la modelación como un proceso para la apropiación de conocimiento, no como un objeto de enseñanza. Sin embargo, es necesario tener una idea de las investigaciones que dan cuenta de la enseñanza de esta modalidad. Como referente, tomamos la investigación de Coutinho (2001) sobre la actividad de modelización de situaciones aleatorias elementales. Ella retoma la modelación propuesta por Dantal (1997), quien describe el proceso de modelación deseable en el salón de clases compuesto por los siguientes pasos: (1) Observación de la realidad, (2) Descripción simplificada de la realidad, (3) Construcción de un modelo, (4) Trabajo matemático con el modelo, (5) Interpretación de resultados en la realidad.

La distinción entre este ciclo de modelación con el descrito en la Figura 3.3 es que en éste el objetivo final no es la obtención del modelo para la generación de teoría, sino la solución de un problema haciendo uso de la teoría. No se trata de un proceso rutinario, puesto que el estudiante no estará en posesión de todo el conocimiento matemático necesario para resolver el problema, muy al contrario, en este proceso se espera que el estudiante lo adquiera. La construcción se establece en un ritmo que simula la generación del conocimiento científico.

Generalmente en el salón de clases únicamente se trabaja la construcción y el trabajo matemático con el modelo, puesto que, desde una perspectiva tradicional, son los más sencillos de enseñar. Sin embargo, todas las etapas son igualmente importantes en el aprendizaje y todas tienen sus propios obstáculos. Coutinho (2001) describe las siguientes dificultades en cada uno de los pasos anteriores:

- La observación de la realidad conlleva la dificultad de diferenciar entre situación aleatoria contingente de una que no lo es. Una situación aleatoria es, por lo menos potencialmente, reproducible. La situación se puede reproducir en las mismas condiciones para analizar los diferentes resultados que se producen. En una situación contingente también hay una presencia del azar, pero no es reproducible en las mismas condiciones. En ambas, es posible conocer los resultados posibles, pero no se puede tener la seguridad de cuál será el que ocurrirá en una experiencia particular. Ambas situaciones difieren en cuanto a la forma en que se puede asignar una medida de probabilidad.

En los Estudios 3 y 4 los alumnos habrán de aceptar que los experimentos individuales son reproducibles. En particular, en el estudio 4, habrán de aceptar que las secuencias producidas por los distintos estudiantes son equivalentes.

- ❖ La descripción simplificada de la misma permite pasar de la realidad observada (paso 1) a la construcción del modelo (paso 3), lo que implica una simplificación de la realidad prescindiendo de aspectos poco relevantes para el problema que queremos resolver. Esto no sólo requiere de conocer de la situación sino de tener una idea clara de las condiciones teóricas que harían valida la teoría vinculada con el modelo. En probabilidad, generalmente la simplificación de la realidad se vincula con el planteamiento de un experimento aleatorio que permita tomar información sobre la situación objeto de estudio, pero también puede ser la búsqueda de información relevante y el contexto en que está sumergido el problema los que hagan válida la teoría apropiada. En este paso la definición de la variable aleatoria o estadística juegan un papel muy importante, puesto que es lo que va a concretar qué es lo importante de observar en el experimento aleatorio planteado.
- ❖ Una vez construido el modelo probabilístico para la situación y obtenidos los resultados «matemáticos» a partir del trabajo con el modelo, queda todavía la parte más importante: comparar estas conclusiones con el comportamiento real de la situación analizada y decidir si el modelo elegido nos proporciona una buena descripción de la realidad. Un modelo no es «real», ni tampoco «verdadero»; en el mejor de los casos es consistente y concordante con las observaciones.

Los pasos mencionados no se siguen necesariamente de manera lineal, ni tampoco que se retoman una sola vez. La solución de un problema puede requerir el regreso de un paso hacia cualquier otro anterior o la construcción de varios modelos que, concatenados, proporcionen un modelo teórico más amplio que represente a la realidad. Esto último es particularmente cierto en el caso de la solución de los problemas a los que se enfrentan nuestros estudiantes en los Estudios 3 y 4, puesto que el establecimiento de la función de distribución pasa antes por la definición de la variable aleatoria o estadística, cuestión que por sí misma, constituye ya la formación de un modelo.

9.4. Modelación por estratos

Heitele (1975) asegura que la modelación en contexto estocástico algunas veces los mismos objetos son tomados como «realidad» y otras como «modelo». Así por ejemplo, en muchas situaciones de modelación la simetría o las frecuencias relativas son consideradas parte de la realidad, pero en otras se objeta que el concepto de simetría ya es en sí un modelo y que registrar frecuencias relativas significa cuantificar la realidad, lo cual presupone la generación de un modelo. Heitele resuelve esta aparente inconsistencia estableciendo las relaciones entre realidad y objetos matemáticos en una estructura de estratos. Esto es, sin duda los valores numéricos de los datos se pueden considerar como un modelo (una vez que se ha establecido una variable, por ejemplo), pero desde un estrato más alto, se les puede ver como parte de la realidad. Lo cual significa que se ha pasado de una realidad palpable y visible a una «nueva realidad» en la que los valores de una variable emergen como hechos reales. Este último sería el esquema de acuerdo al cual las frecuencias relativas de una variable son mapeados sobre probabilidades, concebidas en las mentes humanas dentro de un modelo matemático. Es en este sentido que Heitele piensa que un modelo no es una afirmación sobre la realidad más que de manera local y que comúnmente el modelo es algo que reemplaza a la realidad como sustrato operacional en un estrato más alto.

Por otro lado, Henry (1997) afirma que la construcción de modelos y su perfeccionamiento progresivo intervienen en cada fase de la resolución de problemas estadísticos, no sólo en situaciones prácticas, sino también en trabajo de desarrollo teórico. Para él, la modelación no sólo implica un proceso que requiere ser retomado una y otra vez para su mejora sino también que la resolución de un problema estadístico puede dar lugar a varios procesos de modelación sucesivos. En un cierto sentido, Henry está de acuerdo con Heitele en que la modelación puede tener distintos estratos dentro de un mismo proceso de resolución de problemas y afirma que por ello es importante diferenciar el modelo de la realidad.

Tanto Heitele como Henry aportan propuestas útiles para nuestra investigación, ya que en un proceso de resolución de problemas en el que la variable aleatoria se vea involucrada, forzosamente se deberá pasar por lo menos por dos estratos de modelación, o dos procesos de modelación de acuerdo con Henry. El primero es la formulación de la variable (en el sentido mencionado en el apartado 9.2) y el segundo es la formulación de la función de distribución como modelo matemático. En ambos estratos la variable desempeñará un papel dual, en un primer estrato será el modelo obtenido y en el otro, la realidad. Esto ocurre principalmente cuando se considera la probabilidad laplaciana como modelo de asignación de probabilidades al espacio muestral, como se verá con más detalle en el Estudio 3.

9.5. Relación entre las variables estadística y aleatoria en la modelación

Mientras que los textos de estadística matemática o probabilidad, como Meyer (1989), enfocan la variable aleatoria únicamente desde la probabilidad clásica laplaciana para aplicar estos modelos al análisis estadístico, otros autores (como Ríos, 1967) hacen compatibles las concepciones clásica y frecuencial, para posibilitar la aplicación de la inferencia a las situaciones prácticas.

Para Ríos, la teoría de probabilidades es «el modelo matemático de las regularidades que se observan en las series de frecuencias correspondientes a los fenómenos aleatorios» (p 73). Desde esta perspectiva hay dos simplificaciones de la realidad para pasar a un modelo matemático:

- ❖ Se eliminan resultados posibles inconvenientes o poco usuales: que un dado caiga sobre una arista, que la moneda caiga de canto, que exista una carta en blanco en un mazo de póker, etc.
- ❖ Se admite la estabilidad de las frecuencias de los sucesos posibles. Es decir, que las frecuencias tienden a acercarse a un valor teórico de probabilidad.

Ríos presenta la base empírica de la teoría de probabilidades a partir de la observación de la estabilidad de las frecuencias relativas que permite definir la probabilidad frecuencial y por lo tanto, en la búsqueda de su regularidad, hace uso de la variable estadística.

De acuerdo con él, el proceso de construcción del modelo del cálculo de probabilidad se representa en dos esquemas (que se pueden simplificar en uno solo, Figura 3.5). La realidad nos lleva a observar la estabilidad de las frecuencias y sus propiedades. Por un proceso de conceptualización pensamos en la existencia de un límite de las frecuencias relativas (probabilidad) admitiendo como axiomas las propiedades observadas empíricamente en las frecuencias relativas. Una vez admitidos los axiomas básicos, en coherencia con lo visto en el punto 8.2 de este capítulo, es posible deducir todo el cálculo de probabilidades y a partir de él, los modelos que constituyen la estadística matemática de manera inmediata, con lo cual ya es posible trabajar con ellos para dar una solución al problema planteado. La última fase de la modelación sería la interpretación del modelo en la realidad para determinar la solución a las preguntas planteadas.

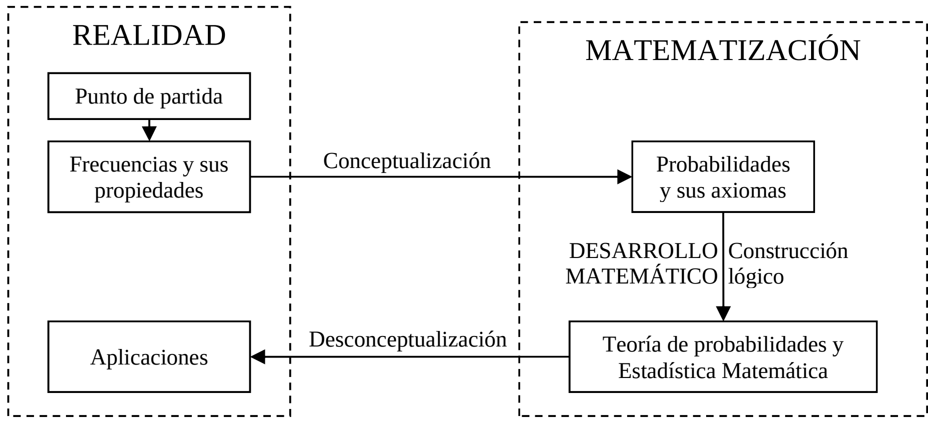


Figura 3.5. Fases de la modelación a partir de la variable estadística según Ríos (1967, p. 73)

Si bien es cierto que el modelo de Ríos es muy acertado, de acuerdo a la modelación por estratos propuesta por Heitele, observamos que su «realidad» es ya un modelo matemático en un estrato más bajo del proceso. Previo a la obtención de las frecuencias relativas y sus propiedades, tuvo que haber un trabajo con los datos en el cual se tuvieron que obtener los valores de la variable y conformar una tabla o una gráfica en donde se reflejara la distribución de las frecuencias relativas de la variable estadística. Actualmente es fácil para nosotros imaginarnos una tabla o gráfica en algún software estadístico que refleje las propiedades de la variable estadística, que se esté actualizando constantemente a medida que se obtienen los datos y que nos permitiera observar la estabilidad de las frecuencias y sus propiedades. La construcción de la variable aleatoria y su distribución sería inmediata a la formulación de las probabilidades y sus axiomas.

Sin embargo el proceso de modelación no resulta tan simple aun desde la modelación estratificada. Después de la estabilización de las frecuencias relativas, la conformación de la distribución de probabilidades sólo puede darse si hay un proceso de descontextualización y los valores de la variable estadística se convierten en sucesos para establecer la relación de conjunto con la probabilidad. En otro estrato del proceso, la variable aleatoria convierte a la variable estadística en su «realidad» para convertirse en el modelo teórico de la variable estadística. En un estrato más, la variable aleatoria se vuelve la realidad de la distribución de probabilidad (o función de densidad), que es el modelo matemático con el que se trabaja para la resolución del problema. A pesar de la aparente simplicidad del proceso de modelación, el desglose respetando todos los procesos y la teoría que debe haber detrás de ellos, se transforma en algo más complicado de lo que Ríos lo conceptualizó (Figura 3.6).

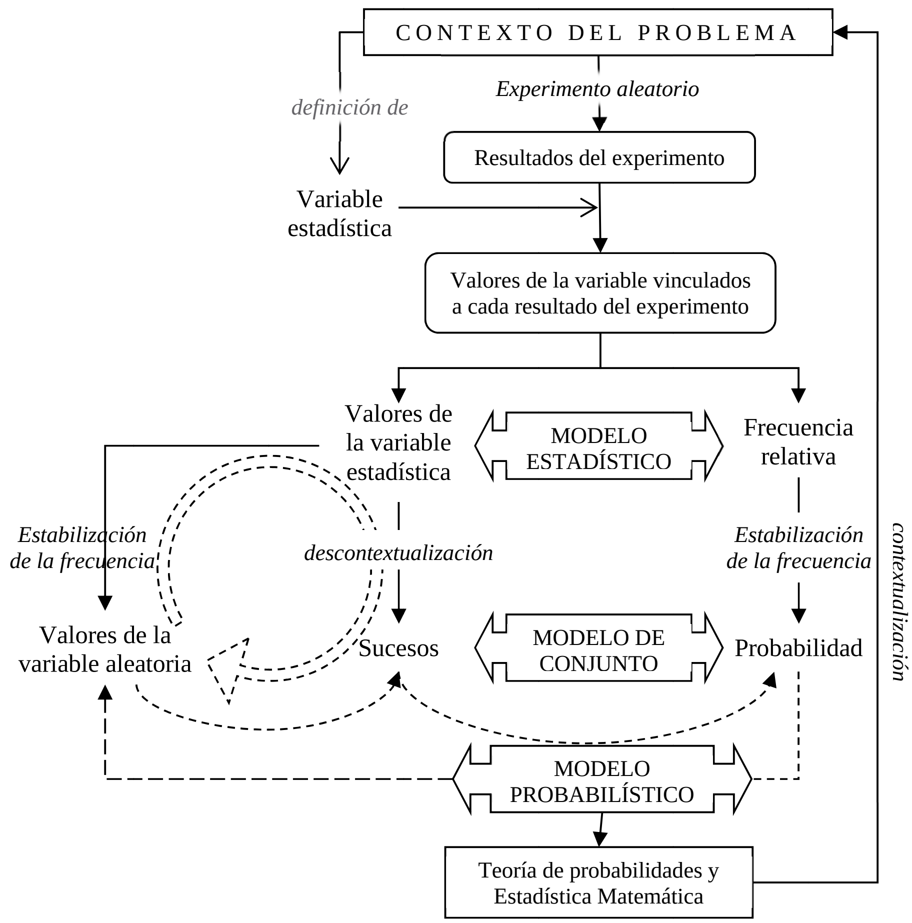


Figura 3.6. Proceso de modelación de la variable aleatoria en una asignación de probabilidad frecuencial

Es notorio que aquí la variable aleatoria no pierde su relación con los sucesos del espacio muestral, sino que se relaciona con ellos a través de la variable estadística. También hay que hacer notar que la existencia del modelo de conjunto es meramente teórico, puesto que el proceso de pasar de la variable aleatoria a la estadística prácticamente es inmediato dando lugar a la distribución de probabilidad.

9.6. La experimentación como puente entre el modelo matemático y la realidad

Con frecuencia los problemas de probabilidad son demasiado complejos para algunos estudiantes. Por ejemplo, en la situación experimental propuesta en el Estudio 4 plantearemos a los estudiantes analizar las variables aleatorias «número de caras», «número de rachas» y «longitud de la racha más larga» de los datos obtenidos al lanzar una moneda equilibrada 20 veces seguidas y compararlas con las mismas variables en secuencias inventadas por los mismos estudiantes. Es claro que el estudio probabilístico de estos problemas está muy lejos de las capacidades de los estudiantes que participan en el estudio.

En primer lugar, los estudiantes no conocen las distribuciones de las variables aleatorias «número de caras», «número de rachas» y «longitud de la racha más larga» en las secuencias inventadas por los estudiantes, pues no sabemos el modelo de secuencia aleatoria que tienen los estudiantes (es decir, no saben si este modelo mental corresponde al modelo matemático de secuencia de ensayos de Bernoulli). Por otro lado, quizás exceptuando el «número de caras en 20 lanzamientos reales» que sigue la distribución binomial *Bin(20, 0.5)*, las otras dos variables aleatorias son demasiado complejas para los estudiantes. Afortunadamente, es posible realizar un estudio intuitivo de estos temas con ayuda de la experimentación o la simulación.

Para Heitele (1970) la simulación en estadística es algo parecido a lo que constituye el isomorfismo en otras ramas de las matemáticas. En la simulación se ponen en correspondencia dos experimentos aleatorios diferentes, uno simulado y otro real, de modo que a cada suceso elemental del primer experimento le corresponda un suceso elemental del segundo y sólo uno, y los sucesos puestos en correspondencia en ambos experimentos son equiprobables. Al trabajar mediante simulación estamos ya modelando, porque debemos no sólo simplificar la realidad, sino fijar los aspectos de la misma que queremos simular y especificar hipótesis matemáticas sobre el fenómeno estudiado (Chaput, Girard y Henry, 2011). De acuerdo con Henry (1997), el papel didáctico de la simulación es inducir implícitamente el modelo teórico a los alumnos, incluso aunque su formulación matemática formalizada no sea posible. En ella ya no se está en la realidad sino en una situación abstracta idealizada, por ejemplo, se imagina que está trabajando con monedas perfectas y prescinde de las condiciones del lanzamiento.

En nuestro trabajo, la situación experimental es lo suficientemente simple como para poder realizarse (simple no significa que no sea complejo) y obtener datos directamente de ella. Se le pedirá a cada estudiante que realice la repetición del experimento físicamente y, por otro lado, que dé los resultados que él cree se podrían obtener si se lanza la moneda el mismo número de veces. Ambos son tipos de experimentos que se pueden realizar en el salón de clases. De este modo, se recogen datos de seis pares de variables estadísticas: número de caras, rachas y racha más larga en cada una de las secuencias reales y simuladas obtenidas por cada uno de los estudiantes. De cada una de estas variables estadísticas obtendremos la correspondiente distribución empírica de frecuencias que podrán compararse por pares para resolver el problema de interés. De este modo, en el Estudio 4, la variable estadística constituye un puente entre el *dominio de la realidad* en que se encuentra la situación que queremos analizar y en la que interviene el azar y el *dominio teórico* donde, con ayuda de la probabilidad se contruye un modelo teórico de variable aleatoria desconocido para el estudiante y, en el caso de sus intuiciones, incluso para nosotros. En el caso de la estadística matemática, existen teorías que ayudan a subsanar la distancia y que generan el modelo apropiado para la variable aleatoria.

En nuestro caso, el estudiante habrá de realizar un proceso de inferencia informal (Rossman, 2008) para obtener conclusiones sobre la variable aleatoria a partir de la estadística. Sin embargo este proceso informal no es posible si el estudiante no tuviera como sustento la posibilidad teórica de la convergencia de la variable aleatoria hacia la estadística. Es decir, puesto que sólo se trabaja con una muestra, el proceso de modelación no pasará por la convergencia de la frecuencia relativa (repetición de experimentos) ni la definición de la función de conjunto, pero sí exige el conocimiento de que, teóricamente, eso ocurrirá. Es en ese sentido que la distribución de frecuencias aporta información sobre la distribución de probabilidades. También es de observar que no se obtiene propiamente al modelo probabilístico (la distribución de la variable aleatoria) sino sólo se infiere de las características de la variable estadística. Es decir, en el proceso de realizar una inferencia informal, se obvian cuestiones teóricas, de las que el profesor debe ser consciente cuando la introduce en el salón de clases.

Esta otra forma de modelar a través del muestreo es diferente a la que implica una asignación de probabilidades laplaciana o frecuencial, aunque comparte elementos con ésta última puesto que la experimentación también funge como puente entre el modelo teórico y la realidad en el primer estrato de modelación.

En la Figura 3.7 se observa que un primer estrato de modelación se finaliza cuando se asignan valores de la variable estadística a cada resultado del experimento aleatorio. Ese momento nos aporta un modelo matemático que será usado como la «realidad» en un siguiente estrato en el que propiamente se establecerá la distribución de frecuencias de la variable estadística. El puente entre el modelo estadístico y el modelo probabilístico (la inferencia informal) está dado por el trabajo matemático con la variable estadística, a través de la cual se efectúa propiamente la inferencia informal, puesto que en una inferencia informal el modelo probabilístico sólo se infiere a través de las medidas de la variable estadística. En este trabajo no se espera que se obtenga propiamente el modelo probabilístico. La modelación por estratos de Heitele (1975) también es apropiada en esta situación. Analizado de acuerdo con la estructura en estratos propuesta, podríamos pensar que en el Estudio 4 hay un doble proceso de modelización: entre la situación real y la variable estadística y entre la variable estadística y la variable aleatoria. Estos procesos los analizaremos con más detalle en el análisis de los resutlados de los estudiantes de acuerdo al Ciclo de Modelación (Punto 7.4 del Capítulo 6).

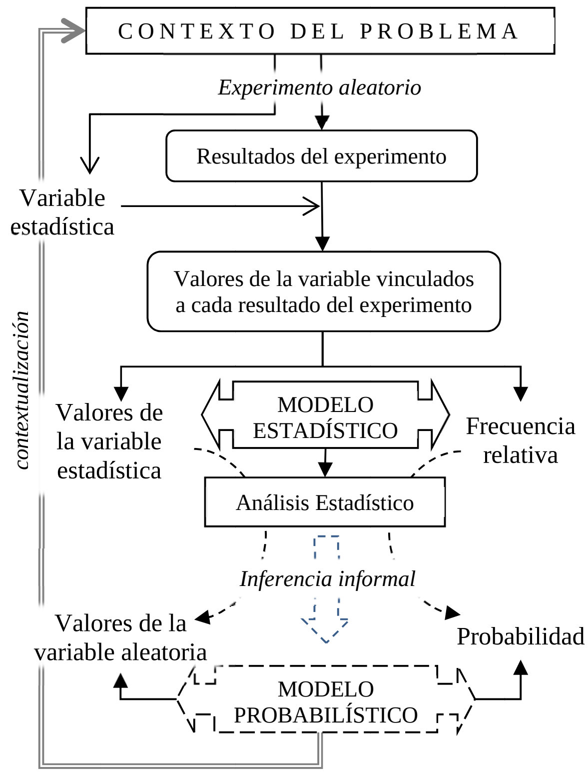


Figura 3.7. Proceso de modelación en una inferencia informal

10. La variable aleatoria y los modelos de distribuciones

Batanero (2001) indica que un ejemplo notable de modelación estadística a partir de un problema práctico son las distribuciones de probabilidad, que permiten describir en forma sintética el comportamiento de las distribuciones empíricas de datos estadísticos y hacer predicciones sobre su comportamiento. Por ejemplo, el modelo de la curva normal de media cero describe con notable precisión la distribución empírica de los errores al realizar medidas repetidas de una misma cantidad (o del tamaño de un cierto producto). Una vez aceptado este modelo como una buena descripción de nuestros datos, basta obtener una estimación de la dispersión en los errores de medida (una estimación de la desviación típica σ en la población de los posibles errores) para tener el modelo completamente determinado. Las aplicaciones prácticas que podemos deducir a partir de esto son muy variadas: previsión del número de defectos (piezas que se saldrán de los límites considerados como «normales»); decisión sobre límites de garantías; coste de la política de garantía; criterios para considerar que el proceso de producción está «controlado», etc.

Krickeberg (1973) sostiene que si estamos interesados en la variable aleatoria *X*, en realidad lo estamos en su función de distribución, puesto que ésta nos proporciona las probabilidades con las cuales el valor $X(\omega)$ está comprendido dentro de los diferentes conjuntos de $\mathcal{A}$ en una observación aleatoria con resultado ω. Sin embargo otros autores (Meyer, 1989; Feller, 1989) sostienen que es en la variable aleatoria en la que radica la importancia de las distribuciones de probabilidad puesto que es ésta en la que se manifiesta el espacio muestral y el experimento y por lo tanto la *realidad*. Lo cierto es que cuando se define la variable aleatoria queda automáticamente definida su función de distribución. Así, el estudio de una implica el estudio de la otra y también la actividad de modelación que vincula la realidad con una representación matemática de ella.

Como regla de correspondencia y desde una perspectiva puramente matemática, la variable aleatoria podría ser arbitraria, siempre y cuando el espacio muestral definido satisfaga las condiciones establecidas en la definición (Mood y Graybill, 1963), pero desde una perspectiva de modelación, la regla de correspondencia debe tener un sentido a partir de una pregunta que se pretenda resolver.

Esto significa que desde una perspectiva exclusivamente teórico-matemática a un espacio muestral se le podrían asignar diferentes reglas de correspondencia (Godino, Batanero y Cañizares, 1996), pero éstas se ven limitadas cuando se condicionan a la pregunta de interés que nos mueve a ocuparnos del fenómeno aleatorio.

Cuando se aplica la teoría de la probabilidad a situaciones reales, no es necesario encontrar una distribución distinta para cada modelo estudiado. A menudo nos encontramos con que muchas situaciones muestran una serie de aspectos comunes, aunque superficialmente parezcan diferentes. En este caso, podemos formular un modelo probabilístico aplicable a estas situaciones. Por este motivo, se han definido varios modelos de variables aleatorias que permiten resolver una gama amplia de situaciones y cuyas distribuciones de probabilidad quedan determinadas, como se ha dicho anteriormente, por uno o varios parámetros. Tradicionalmente los modelos se dividen de acuerdo a la continuidad de las variables aleatorias en continuas y discretas.

Entre dichos modelos citaremos sólo algunos de los más importantes por la frecuencia con la que se presentan en el contexto real.

10.1. Modelos de variables discretas

Distribución uniforme discreta. Se trata de una variable aleatoria que toma todos los valores enteros entre a y b, que son los parámetros de la distribución, teniendo todos ellos la misma probabilidad. La distribución de probabilidad está dada por la siguiente expresión, donde *x* son valores enteros de la variable aleatoria incluidos entre *a* y *b*:

$$p(\mathrm{X}=x)=\frac{1}{b-a}; \qquad x\in(a,b)$$

Distribución binomial B (n, p). Consideremos un experimento aleatorio cualquiera, y en relación a él, estudiemos un suceso *A*, de probabilidad *p* y su contrario $\overline{A}$ de probabilidad *q=1-p.* Diremos que hemos tenido un éxito, si al realizar el experimento obtenemos el suceso *A*, y que hemos obtenido un fracaso en caso contrario. Si, en lugar de realizar únicamente una vez el experimento, efectuamos una serie de repeticiones independientes del mismo, el número total de éxitos obtenido en las *n* realizaciones constituye una variable aleatoria Γ, que puede tomar los valores enteros comprendidos entre 0 y *n.* Esta distribución será la que siga la variable «número de caras» de la situación planteada a los alumnos en el Estudio 4 de este trabajo. La distribución de la variable está dada por la siguiente expresión:

$$P(\Gamma=r)=\binom{n}{r}p^r q^{n-r}$$

Distribución de Poisson. Si en la distribución binomial aumentamos indefinidamente el número de pruebas, manteniendo constante el producto $np=\lambda$, obtenemos una nueva distribución que recibe el nombre de distribución de Poisson. Aunque no la utilizaremos en los estudios empíricos se hace referencia a ella constantemente en el Estudio 2. La variable aleatoria discreta sigue la distribución de Poisson si toma los valores enteros 0,1,2, ... y su distribución de probabilidades es la dada por:

$$p(\mathrm{X}=x)=\frac{\lambda^x e^{-\lambda}}{x!}$$

10.2. Modelos de variables continuas

Distribución continua uniforme. Se trata de una variable que toma cuya densidad de probabilidad es constante en un intervalo (a, b). Sus parámetros son *a* y *b* y su función de densidad está dada por la siguiente expresión:

$$f(x;a,b)=\frac{1}{b-a}; \qquad x\in(a,b)$$

Distribución normal. Esta distribución es la más usada en la estadística clásica. Describe la distribución de probabilidad de las variables cuya probabilidad es mayor conforme los valores se acercan al centro de la distribución y decrece rápidamente hacia ambos extremos. Se caracteriza principalmente por la simetría de la distribución. La función de densidad normal depende de dos parámetros μ y σ, que son su media y su desviación típica respectivamente, y está definida por la siguientes expresión.

$$f(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad ,-\infty<x<\infty$$

11. Conclusiones del capítulo

A lo largo del capítulo se ha visto que, como un objeto matemático complejo, la variable aleatoria se relaciona con otros objetos matemáticos que tienen sus propias complejidades y que, a su vez, éstos influyen en cómo se conforma y las acepciones que de él se tengan. Desde una perspectiva educativa, esto hace que el objeto de nuestro interés sea difícil de analizar si no se tiene una clara definición del alcance que se pretende en los estudiantes. Los libros analizados conforman una muestra de la diversidad de alcances y enfoques con los que se puede enseñar y analizar la teoría en los que se sustenta.

La visión en la escuela tendrá que ser limitada, sobre todo cuando se trata de cursos introductorios en el nivel universitario, pero también completa porque las concepciones tratadas tendrán que mostrarse como la conformación acabada de una idea que les será útil profesionalmente a nuestros estudiantes, sobretodo porque para muchos planes de estudio, los cursos introductorios son los únicos cursos de probabilidad y estadística que se imparten. El concepto explícito de la variable aleatoria como el objeto en el que está sustentado el paso de la matemática de conjuntos al análisis matemático deberá ser pospuesto para cursos más especializados o carreras más específicas, sin embargo no es deseable posponer el rol implícito que juega este concepto en la aplicación de la teoría de probabilidad y estadística matemática en cualquier profesión y en la cotidianeidad mundana. La modelación de situaciones aleatorias y el papel de la variable aleatoria en ella son clave para lograrlo, aunque la modelación tampoco es un aspecto simple. Las acepciones de los conceptos involucrados con nuestro objeto de estudio también modifican la forma en que se desarrolla el proceso al resolver un problema en el que está vinculado la variable aleatoria. Sin embargo, dentro de toda esta diversidad, se condicionan las acepciones y construcciones de la variable aleatoria sobre el significado de la probabilidad.

Como en los primeros apartados de este capítulo se pudo observar, la teoría de la probabilidad no da cuenta de la forma en que se conforma el espacio de probabilidad. Define las condiciones de una probabilidad que llamaremos *axiomática* (o formal, de acuerdo con Hawkins y Kapadia, 1984) y con ello proporciona ciertas condiciones sobre las que descansa el resto de la teoría. Sin embargo, la relación de esta teoría con la realidad está dada a través de la forma en que se asigna el valor de probabilidad (Punto 8 de este capítulo) y por lo tanto, esta asignación también indicará como opera y surge la variable aleatoria. Es decir, los significados que aporta la teoría a la realidad y viceversa provienen de esa asignación. Así, en nuestro trabajo, definiremos las *configuraciones de objetos* vinculados a la variable aleatoria de acuerdo al significado de la probabilidad que esté en juego.

Entendemos por *configuraciones de objetos* las redes de objetos intervinientes y emergentes de los sistemas de prácticas y relaciones que se establecen entre los mismos (Godino, Batanero y Font, 2007). Puesto que la asignación de probabilidades sugiere un distinto sistema de prácticas para la variable aleatoria, podríamos identificar cuatro configuraciones epistémicas diferenciadas para la variable aleatoria: clásica, frecuencial, subjetiva y axiomática, según la concepción subyacente de probabilidad que conlleva la aparición de ciertos objetos matemáticos complementarios. En la Tabla 3.1 se muestran los objetos vinculados a cada configuración de la variable aleatoria. La primera columna contiene elementos comunes a todas las concepciones y que no se repiten en la columna correspondiente. Los objetos vinculados a cada configuración se agrupan de acuerdo a los elementos que conforman el significado de un objeto matemático (Godino, 2002):

- *Situaciones-problemas.* Aquí se incluyen los problemas mencionados a lo largo del capítulo sobre predicción o toma de decisión (para la concepciones clásica, frecuencial y subjetiva); en la concepción frecuencial y subjetiva se añadiría el problema de estimación. Los problemas asociados a la configuración axiomática serían intramatemáticos (por ejemplo, hallar la distribución asintótica de la suma de variables aleatorias).

Tabla 3.1. Algunos objetos matemáticos vinculados a las configuraciones de la variable aleatoria de acuerdo con la asignación de la probabilidad

	Significado de la probabilidad				
	Clásico/Frecuencial /Subjetivo	Clásico	Frecuencial	Subjetivo	Axiomático
Situaciones problemas	Realizar inferencias Toma de decisiones	Juegos de azar	Análisis de datos	Evaluar grado de credibilidad personal	Problemas intra-matemáticos
Lenguaje	Verbal, simbólico, numérico		Tablas, gráficos		Formal
Conceptos	Experimento aleatorio		Experimento estadístico		Experimento aleatorio
	Población		Muestra		Espacio muestral
	Sucesos				Álgebra de sucesos, espacio probabilístico
	Probabilidad		Frecuencia relativa	Probabilidad inicial/final	Función medible, medida
	Variable aleatoria, valor de la variable, rango, máximo, mínimo.		Variable estadística Dato	Dato	Función de conjunto, original, imagen; v. dependiente e independiente
	Distribución de probabilidad		Distribución de frecuencias	Distribución inicial/final, verosimilitud	Función compuesta; inversa
	Parámetros (media, varianza, etc.)	Esperanza Matemática	Estadísticos (media, varianza, etc.)	Distribución del parámetro, inicial y final	Parámetros (media, varianza, etc.)
	Función de distribución		Distribución acumulada de frecuencias		Integral
	Función de densidad		Histograma de frecuencias		Derivada
	Distribución de probabilidad		Polígono de frecuencias		
Propiedades	Dependencia, Independencia	Equi probabilidad	Repetibilidad Convergencia	Carácter subjetivo, parámetros variables	Axiomas
Procedimientos	Análisis Matemático	Combinatoria, proporciones, análisis a priori de la estructura del experimento	Registro de datos estadísticos, simulación, ajuste de curvas	Teorema de Bayes, probabilidad condicional	Conjuntistas
Argumentos	Deductivo	Combinatorios	Inductivos		Deductivo

- ❖ *Conceptos-definiciones.* Además de los relacionados con el experimento aleatorio y probabilidad, encontramos conceptos comunes, como variable, valor, máximo-mínimo, rango, función de distribución, etc. Se amplían estos conceptos con los relacionados con la variable estadística y la inferencia en el significado frecuencial (población, muestra, frecuencia, …) y con los relacionados con la probabilidad condicional en el significado subjetivo de la probabilidad. La configuración vinculada con el significado axiomático utiliza conceptos tomados de la teoría de conjuntos y el análisis matemático.
- ❖ *Lenguaje.* Se refiere a los términos, expresiones, notaciones, gráficos, ... tales como las diversas representaciones gráficas o simbólicas de la variable aleatoria y los objetos relacionados que también varían en cada configuración. Por ejemplo, en la configuración axiomática no se requieren las representaciones gráficas de las funciones de densidad o distribución; en la concepción clásica no se usan los histogramas de frecuencia, etc.
- ❖ *Proposiciones.* Serían las relaciones que ligan los diferentes conceptos-definición entre sí o con otros objetos matemáticos. Por ejemplo la proposición que indica que la función de densidad es la derivada de la función de distribución.
- ❖ *Procedimientos.* Son numerosas las técnicas (algoritmos, operaciones, técnicas de cálculo, ...) asociadas al trabajo con variables aleatorias. Es común el uso del análisis matemático al que se unen los procedimientos propios de cada enfoque: el análisis combinatorio y la estructura a priori en el significado clásico; el registro de datos y ajuste de curvas en el frecuencial; y el uso del teorema de Bayes en el subjetivo.
- ❖ *Argumentos* son los enunciados usados para validar o explicar las proposiciones y procedimientos, deductivos o de otro tipo, que incluirían los argumentos empleados para justificar las proposiciones o las técnicas.

Debido a la gran complejidad del objeto, nuestro acercamiento en los estudios cognitivos será sólo exploratorio y limitado a algunos de los componentes del objeto. Como se ha indicado, en estos estudios se trabajarán dos situaciones diferentes desde el punto de vista metodológico y por el significado de probabilidad desde el que surge el aparato probabilístico y estadístico de la variable aleatoria. En el análisis a priori de cada una de las situaciones se retomará el análisis conceptual para concretar los objetos de análisis de cada situación.

- ❖ En la situación experimental propuesta en el Capítulo 5 (Estudio 3) la asignación de probabilidades es clásica y nos interesamos tanto por objetos matemáticos compartidos en los diferentes significados (aleatoriedad, probabilidad, distribución de probabilidad) como por la comprensión de algunos objetos relacionados con la concepción axiomática (función, componentes de la función, rango, dominio, variables dependientes e independientes).
- ❖ En una segunda situación analizada en el Capítulo 6 (Estudio 4), se combina el uso de las concepciones clásica y frecuencial de la probabilidad y aparecen, además de diferentes variables aleatorias, y los objetos mencionados en el punto anterior las correspondientes variables estadísticas, junto con los objetos matemáticos asociados (frecuencias, distribución de frecuencias, estadísticos) y representaciones gráficas (histogramas, diagramas de barras, etc.).

Además, en ambas situaciones experimentales nos interesamos también por la actividad de modelación de los estudiantes, analizando con detalle las soluciones y conclusiones obtenidas y la interpretación de la solución en el contexto del problema planteado.

Capítulo 4.

Estudio 2.

Análisis Epistemológico Histórico

ÍNDICE DE CAPÍTULO

1. Introducción

La variable aleatoria, como muchos otros objetos de la ciencia, ha surgido progresivamente y ha presentado etapas de mayor o menor desarrollo caracterizadas por sucesos que marcan algún progreso en su conceptualización como objeto matemático. Un aspecto esencial para la comprensión didáctica de la variable aleatoria es conocer su génesis histórico-epistemológica.

En este estudio el análisis epistemológico histórico tendrá la intención no sólo de poner de manifiesto la diversidad de puntos de vista que se han sucedido, algunos de los cuáles en su tiempo se han considerado correctos, pero que después se han rechazado o modificado. Este análisis también tendrá la intensión de buscar elementos epistemológicos constitutivos del significado del objeto y su adaptación a la resolución de distintos problemas (Gómez, 2003). Se tiene también la intención de identificar los obstáculos ligados al desarrollo del objeto matemático.

Nos interesa el análisis del devenir histórico como una herramienta cognitiva puesto que pretendemos encontrar en la historia elementos que ayuden a la comprensión del aprendizaje del concepto de variable aleatoria. Se tiene el propósito de que esta información pueda convertirse en fuente de hipótesis y aporte elementos para el diseño de situaciones didácticas centradas en un objeto matemático y que a su vez, marque etapas de desarrollo en su construcción en los estudiantes (Brousseau, 1997).

La variable aleatoria ha estado presente en casi toda la historia de la probabilidad y la estadística; sin embargo no ha sido sino hasta años relativamente recientes cuando se ha puesto de manifiesto en forma explícita. Es difícil de seguir a lo largo del tiempo debido a la especificidad de su definición, a la trivialización de su importancia, a lo complejo de las herramientas matemáticas necesarias para formalizarlo, a la cercanía temporal de su definición formal y a que la variable aleatoria está vinculada con problemas filosóficos que la teoría de la probabilidad no resuelve (Doob, 1986).

En este análisis nos guiamos por la presencia implícita que percibimos en la historia de la probabilidad y las propiedades distintivas de su formalización denotada en el capítulo anterior.

Iniciamos nuestro análisis basamos en las etapas históricas del desarrollo de la teoría de la probabilidad que propone García (1971) resaltando los momentos importantes que intervinieron en la formulación de la conceptualización de la variable aleatoria como la conocemos actualmente. Destacamos la participación de algunos matemáticos a esta formulación, generalmente agrupados por las fechas en que trabajaron, pero sobre todo por el momento histórico en que se sitúan sus aportaciones al desarrollo del concepto de variable aleatoria. Recurriremos al anacronismo de actualizar términos, e indicaremos el momento en que lo hacemos, para poder observar y comprender el desarrollo del concepto de variable aleatoria desde nuestra perspectiva actual, aunque trataremos de no obviar las diferencias entre las concepciones de entonces y las actuales.

Concluimos el capítulo con una propuesta de ocho etapas históricas de acuerdo al desarrollo histórico de la variable aleatoria en particular, resaltando las características que las distinguen y los principales personajes que contribuyeron a su desarrollo. Así mismo detallamos las dos líneas que conducen el devenir histórico del objeto matemático y las formas en que se interaccionan a lo largo de las ocho etapas. También proponemos los saltos cualitativos conceptuales que se dieron a lo largo de la historia y que contribuyeron al paso de una etapa a otra. Finalmente proponemos algunas posibles recomendaciones didácticas e identificamos algunos obstáculos históricos que pueden ser pertinentes para su conceptualización en las clases de probabilidad y estadística.

2. Primeros indicios. Previsión de la apuesta en juegos de azar. Esperanza matemática

Uno de los primeros campos de problemas en donde se vislumbra la idea de variable aleatoria es el que se vincula con los juegos de azar. Aunque los juegos de azar son tan antiguos como el hombre mismo y se han encontrado indicios de la práctica de estos juegos en las culturas más primitivas, el análisis matemático más formal de los mismos no se inició hasta épocas relativamente recientes (García, 1971). La aparición de ideas vinculadas con la variable aleatoria en estos trabajos es muy poco formal pues no se menciona la existencia de variables o distribuciones de manera general. Sin embargo, sí se definen variables para casos particulares y en algunos casos se trabaja con sus distribuciones. Diversos matemáticos se interesaron por el problema de estimar la apuesta equitativa en un juego de azar, lo que les llevó a considerar implícitamente variables aleatorias y su distribución. En términos modernos, su interés principal era la *esperanza matemática* de la variable.

Así por ejemplo, Bellhouse (2000) analizó el poema del siglo XIII, *De Vetula*, atribuido a Richard de Fournival (1201-1260). Este manuscrito relata principalmente la vida amorosa de Ovidio, pretende dar algunos consejos morales y, en varios párrafos, describe pasatiempos y entretenimientos con mucho detalle. En la parte que nos interesa, alerta a la gente sobre los perjuicios del juego de dados. Se detiene en el análisis de los posibles resultados de un juego en el que se apuesta sobre la suma de los puntos obtenidos en las caras superiores al lanzar tres dados. Fournival busca explicar los resultados obtenidos tras haber observado varias veces repeticiones del experimento aleatorio. Para ello analiza teóricamente las posibles respuestas de un lanzamiento. Según Bellhouse este poema es el escrito más antiguo en el que se establece una relación entre las frecuencias observadas (de los valores de una variable aleatoria al repetir el experimento) y sus posibilidades (espacio muestral del experimento aleatorio): «Se producen dieciséis números. No tienen, sin embargo el mismo valor, puesto que el menor y el mayor ocurren rara vez, mientras los centrales aparecen frecuentemente» (Bellhouse, 2000, p. 134).

En su análisis, Richard de Fournival trata de explicar por qué esos dieciséis números aparecen más o menos veces y se da cuenta que existen diferentes formas en las que puede surgir cada uno de ellos. Enumera las posibles *configuraciones* (es decir, enumera el *espacio muestral* actual) de los tres dados al ser lanzados y conecta cada una de ellas con los dieciséis números que se producen a partir de la suma de las caras de los dados. Finaliza con la determinación de sus posibilidades (probabilidades) y recomienda a los jugadores que apuesten según el beneficio esperado. En ese poema, el autor liga las configuraciones de los dados a un número a través de una *norma* («la suma de los números obtenidos en la cara superior de los tres dados») y después con una probabilidad para tratar de obtener una respuesta a sus inquietudes.

Alrededor de 1620 (400 años después) este mismo juego fue propuesto a Galileo por el Duque de Toscana. La inquietud del Duque era que había notado que las sumas 9 y 12 aparecen menos veces que las sumas 10 y 11, aun cuando para ambas sumas hay el mismo número de combinaciones de números que las producen. Galileo encontró la misma solución que aparece en el poema De Vetula, pero añadió una demostración combinatoria completa de la solución, obteniendo 25 casos teóricos favorables para 9 y 12 y 27 para 10 y 11. En su solución, él tomaba en cuenta el orden de los dados y una definición apropiada del espacio muestral: cada configuración de los sucesos posibles tomando en cuenta el orden de colocación de los dados.

En estas dos soluciones dadas por Fournival y Galileo al juego con tres dados, se observa la necesidad del análisis de la situación completa para contestar una pregunta. El «número obtenido en la cara superior de un dado» es una variable aleatoria, de modo que la «suma de las caras superiores obtenidas en los tres dados» también lo es. Se manifiesta la necesidad de analizar una variable para estudiar el problema y asignar una probabilidad (o posibilidad) a cada valor que toma. Tomando en cuenta que este es uno de los primeros problemas de probabilidad resuelto en la historia (Batanero, Henry y

Parzysz, 2005), podemos decir que la definición de la variable aleatoria está en los orígenes de la probabilidad, aunque de manera implícita.

También hay que observar que la motivación para el análisis teórico de las posibles respuestas (tanto por parte de Galileo como por Fournival) se basa en observaciones empíricas de los distintos posibles resultados y sus frecuencias. La inquietud tanto del Duque de Toscana como de Fournival tuvo que haber sido producto de una recopilación y análisis de muchos datos tomados de la experiencia, puesto que pocas observaciones no habrían sido suficiente para hacer notable la diferencia que se produce entre la aparición de sucesos que tienen 25 ó 27 casos favorables de entre 216 casos posibles. Es decir, se tuvo que haber hecho uso de variables estadísticas. Sin embargo, en ninguno de los dos casos se hizo explícito el uso de la variable estadística ni su vinculación con la variable aleatoria.

Cardano, Pascal, Fermat y Huygens enfocaron su interés por la predicción de los juegos de azar en establecer las normas necesarias para tener un juego equitativo. En su análisis recurrieron intuitivamente al uso de la distribución de las variables aleatorias vinculadas y su valor esperado.

Aunque Cardano no visualizó su libro como un trabajo de índole matemático, su *Liber de Ludo Aleae* (Cardano, 1953/1663) ha sido estudiado por muchos historiadores como el más completo tratamiento del cálculo de probabilidades antes del libro de Huygens. A Cardano se le atribuye la introducción de la idea de asignar una probabilidad *p* entre cero y uno, tomando en cuenta el número total de resultados y el número de casos favorables. Pero su gran aportación, fue sugerir explícitamente el uso del «peso relativo de los sucesos favorables» en los juegos de azar para hacer una apuesta equitativa. Recomendó a los jugadores usar el cálculo combinatorio y les dio una regla general: «…deberemos considerar el circuito completo y el número de lanzamientos que representan como puede ocurrir el resultado favorable en muchas formas y comparar ese número con el resto del circuito, y de acuerdo a esa proporción se deberán hacer las mutuas apuestas, así, se competirá en igualdad de condiciones.» (Cardano 1953/1663, p 202). Por «circuito completo» Cardano se refería al número de resultados elementales posibles.

En su *principio fundamental del juego* definió lo que debería de ser un juego equitativo: «igualdad de condiciones… de dinero, de situación y del dado mismo» (Cardano, 1953/1663, p 189). Cardano sustentó este principio en lo que en lenguaje contemporáneo sería el número esperado de ocurrencias y equiprobabilidad: «…en seis lanzamientos, cada punto debería aparecer una vez, pero ya que alguno se repetirá, a esto seguirá que otro no aparecerá.» (Cardano, 1953/1663, p 191). En lenguaje contemporáneo diríamos que a largo plazo *esperamos* que en seis lanzamientos, cada punto aparezca una vez. Al reconocer que la equiprobabilidad es un ideal y que sólo se puede obtener después de una serie de intentos repetidos, Cardano, no sólo menciona una forma rudimentaria de la ley de los grandes números sino también trabaja implícitamente con la relación tan delicada entre la probabilidad clásica y la frecuencial, al mismo tiempo que maneja una variable aleatoria y una estadística, su vinculación en el contexto del problema y la tendencia de una hacia la otra en tales situaciones.

La correspondencia entre Pascal y Fermat (Pascal y Fermat, 2007/1654) se considera el punto de partida de la teoría de la probabilidad. En ese intercambio de cartas, resuelven varios problemas haciendo uso de la combinatoria y del análisis de las posibilidades que cada jugador tiene de ganar. Entre ellos, el problema de la apuesta interrumpida propuesto a Pascal por el caballero de Meré ha sido de los más relevantes. Dicho problema se cuestionaba sobre la forma en que una apuesta debía ser dividida entre dos jugadores dado que el juego no había sido concluido, pero existían resultados que podían dar cuenta de a quién estaba favoreciendo la suerte. Ambos autores resolvieron el problema por diferentes vías, pero estaban de acuerdo en que había que dividirla proporcionalmente en función de las posibilidades que cada jugador tenía de ganar en el momento en que se interrumpiera el juego. Se dieron cuenta de que «hay que buscar primero en cuántas partidas estará decidido el juego con toda seguridad… Hay que ver, ahora, de cuántas maneras se combinan las partidas… y cuántas hay favorables [a cada jugador]» (Pascal y Fermat, 2007/1654, pp 255-256). Es decir, utilizan una variable aleatoria, aunque no la enunciaron explícitamente. Pascal y Fermat resolvieron el problema para un caso particular de número de partidas, pero lo intentaron resolver para varios casos y con varios jugadores, haciendo más complicada la variable aleatoria y su distribución. Posteriormente, en su libro *Traité du Triangle Arithmétique* (Pascal (1963/1654), demostró algunas propiedades de lo que a partir de entonces se llamó El triángulo de Pascal, y las aplicó para producir una solución sistemática del problema del reparto de las apuestas.

Hald (2003) menciona que uno de los primeros indicios del uso intuitivo de la variable aleatoria fue el realizado por Christian Huygens (1629-94) en su libro *De Ratiociniis in Ludo Aeae* (1714/1657). Huygens, como buen discípulo, retoma la correspondencia de Pascal y Fermat y resuelve varios de los problemas discutidos por ellos, entre ellos el problema de la apuesta interrumpida (Proposición IV).

> Prop. IV: Suponiendo que juego contra otra persona esta condición, que quien gane primero tres juegos se llevará la apuesta; y que ya he ganado dos juegos de los tres y él sólo uno. Quiero saber qué proporción de la apuesta me correspondería, en caso de que

decidiéramos no continuar el juego y dividiéramos la apuesta entre nosotros (Huygens, 1714, p. 4).

En el análisis de su solución, Huygens hace explícita la variable necesaria a analizar: «En primer lugar, debemos considerar el número de juegos que sigue deseando [ganar] cada jugador» (Huygens, 1714/1657, p. 4), para ello, se sitúa en el contexto del problema. También, de manera implícita, relaciona los valores de la variable con el cálculo de la «expectatio», que define en su libro. En la proposición XIV, Huygens plantea un problema donde puede observarse que hace uso de una variable que es una suma y enfatiza: «la suma de dos resultados numéricos del experimento aleatorio de lanzar dos dados»:

> Prop. XIV. Supongamos que otro jugador y yo jugamos por turnos con un par de dados bajo estos términos, yo ganaré si obtengo 7 puntos, o él si obtiene 6 puntos antes, y él tiene la ventaja de lanzar primero. Hallar la proporción de nuestras oportunidades (Huygens, 1714/1657, p 11).

En este primer momento, ya puede observarse el uso de variables que toman valores numéricos (sobre todo en el trabajo de Huygens y de Pascal), que dependen del contexto del problema y que se conciben como una norma que relaciona el espacio muestral del experimento aleatorio con los valores numéricos. La variable aleatoria todavía no se usa explícitamente, pero sí la idea de esperanza matemática y la obtención de distribuciones de probabilidad de algunas variables aleatorias discretas asociadas a problemas particulares, en juegos de dados u otros juegos de azar, aunque sus autores no se plantean generalizar sus resultados.

También puede notarse la observación de los valores de la variable estadística y la frecuencia con la que aparecían, así como que estaban interesados por comparar (o explicar) sus resultados con los obtenidos en análisis teóricos de los fenómenos aleatorios. Desafortunadamente, no se cuenta con alguna constancia de que los jugadores llevaran tal recuento.

3. Elaboración de censos y previsión de datos socio económicos. La variable estadística

Paralelamente al desarrollo de la teoría de la probabilidad a través de la resolución de problemas de juego, en otro ambiente, se gestaba el nacimiento de la estadística a través de la recopilación y descripción de datos sociales o económicos. Desde épocas muy remotas el hombre ha tenido necesidad de realizar conteos y representaciones que podrían considerarse recuentos estadísticos sencillos. La necesidad de conocer y planear, en el sentido de saber qué es lo que se tiene y hacer accesible y manejable esa información para tomar decisiones, ocasionó que poco a poco gobernantes, comerciantes y militares realizaran censos y recuentos cada vez más sofisticados. Siguiendo a Ríos (1994) y a Hald (2003), a continuación se incluye un resumen de estos trabajos.

Se tienen indicios de que las actividades más antiguas de los estadísticos se remontan a los judíos, egipcios, chinos, griegos y romanos, quienes realizaban censos de tierras, riquezas, producción agrícola o de la población con objetivos tributarios, militares o de distribución e incluso ya se mencionan en el Evangelio. Suponemos que en esos recuentos y representaciones hubo un uso implícito de variables estadísticas, pero no se cuenta con evidencias de previsiones con base en recopilaciones y análisis más sistemáticos previos al siglo XVI. Las manifestaciones conocidas sobre el uso implícito más formal de la variable estadística e incluso de la relación entre la variable estadística y la variable aleatoria al tratar de generalizar a casos futuros son posteriores a ese tiempo.

Una de las primeras compilaciones de las que actualmente se tiene conocimiento fue la realizada por Sebastián Münster, quien dedicó gran parte de su vida a la recopilación de datos en los ámbitos geográficos, etnográficos, políticos, militares y sociales que reunió en su libro *Cosmographia*, escrito entre 1544 y 1550 (McLean, 2007). Su mayor influencia fue en cartografía, sin embargo, su recuento de datos, de una vasta utilidad práctica en su época, resaltó la necesidad de las actividades censales.

Alrededor de 1590, en Londres, a propósito de conocer las muertes ocasionadas por la peste, se comenzaron a registrar los bautizos y las defunciones acontecidas junto con sus causas. Graunt, (1968/1662) se dio a la tarea de analizar y sistematizar la información de 30 años para responderse preguntas que supuso de interés: «empecé, entonces, no sólo examinando las Presunciones, Opiniones y Conjeturas, que la vista de unas pocas Tablas me habían sugerido; sino admitiendo también otras nuevas, como encontré razón y ocasión a partir de mis tablas» (p 111). Al plantearse preguntas, definió variables, algunas de ellas, por su naturaleza numérica, fueron variables estadísticas. También fue capaz de detectar, de manera implícita, las variables estadísticas que pudieran ser susceptibles de análisis a partir de los datos con los que contaba.

En su obra *Natural and Political Observations mentioned in a following Index, y made upon the Bills of Mortality*, Graunt (1968/1662) recopiló los datos de múltiples variables estadísticas y se preocupó por describir cómo fueron tomados los datos (por lo tanto su confiabilidad), analizó las proporciones de las causas de las muertes (comparándolas con el total y entre sí) y efectuó predicciones sobre el número de personas que morirían de algunas enfermedades y las proporciones de nacimientos de varones y mujeres que cabría esperar. En estos análisis se puso en evidencia la regularidad de ciertas proporciones y estableció la uniformidad y previsibilidad de muchos fenómenos poblacionales y de salud pública cuando se tienen muchos datos. También dejó planteadas otras preguntas que podrían ser contestadas a partir de datos y que ayudarían a un «buen gobierno» en la planificación, evaluación, control y pronóstico de tendencias sociales. Por lo tanto, Graunt también trabaja implícitamente con variables aleatorias, puesto que se plantea problemas de inferencia a las poblaciones de donde provienen sus datos.

El trabajo de Graunt fue muy reconocido y fomentó el estudio de las estadísticas en Europa, particularmente en Francia e Inglaterra.

Las actividades censales poco a poco se fueron introduciendo en muchos países y en siglos posteriores tuvieron tanta importancia que incluso se crearon asociaciones y sociedades para tal fin. Junto con esta actividad, también se obtuvieron formas de analizar las actividades humanas basándose en datos, lo que trajo consigo usos más complejos de la variable estadística.

Una de las primeras obras en las que la probabilidad y el registro estadístico de datos empíricos fueron aplicados al análisis de fenómenos distintos a los juegos de azar es el ensayo *An argument for divine providence, taken from the constant regularity observed in the births of both sexes*, publicado por John Arbuthnott (1667-1735) en 1710. En él, Arbuthnott, influenciado por los trabajos de Fermat, Pascal y Huygens, aplicó la relación Binomial al estudio de la proporción que debería haber entre los sexos de los pájaros que nacen en una nidada. Supuso que la población de pájaros debería estar constituida por igual número de machos que de hembras (p = ½) y calculó teóricamente el número de machos (M) y hembras (F) posibles procreados en un nido que tiene un número n de huevos mediante la ley Binomial $(M+F)^n$. Sus resultados indicaron que se esperaría que el número de aves machos y hembras que nacieran en el nido fuera el mismo independientemente del número de huevos que hubiera en el nido. Sin embargo, su experiencia como observador de aves le indicaba que había un mayor número de aves machos que hembras y por lo tanto concluyó que no se cumplía la distribución propuesta por Pascal.

Arbuthnott no se conformó con esto, sino que aplicó el mismo análisis teórico a la distribución de sexos en los seres humanos y comparó los resultados obtenidos con los registros de bautismos realizados en Inglaterra entre 1629 y 1710. Los registros indicaron que el número de niños nacidos fue superior al número de niñas en esos 81 años, pero además en una proporción constante. Ante este hecho concluyó que la proporción no igualitaria existente entre el número de niños y de niñas que nacen cada año en Inglaterra no estaba gobernada por el modelo ideal de la probabilidad atribuible al azar, sino que esa proporción entre los dos sexos estaba regida por el misterioso «arte de la Divina Providencia» que escapaba a la comprensión del intelecto humano.

Aunque el escrito de Arbuthnott claramente se inclina por un misticismo religioso, sus estudios pusieron en duda que el modelo clásico de la probabilidad desarrollado por Pascal y Fermat rigiera todos los fenómenos de la naturaleza, pero también para hacer notar la diferencia entre los modelos matemáticos ideales de la probabilidad y la forma en que ocurren esos fenómenos en la realidad. Es decir, dio a conocer que el modelo clásico de la probabilidad desarrollado por Pascal y Fermat puede resultar controvertido para estudios fuera de los juegos de azar.

Observemos que este estudio Arbuthnott siguió un proceso de modelación diferente al seguido por los autores que indagaron sobre los juegos de azar descritos en el momento histórico anterior (Apartado 2 de este capítulo). Él estaba entusiasmado con la teoría de probabilidades y trató de llevarla a los fenómenos naturales, a diferencia de Galileo, Pascal y Fermat, quienes partieron del análisis de la realidad para generar teoría. Pero en este proceso también usó una concepción de aleatoriedad más moderna, puesto que su razonamiento fue que si el fenómeno no seguía los derroteros dados por la entonces naciente teoría de probabilidad, entonces el fenómeno no podía considerarse azaroso y buscó otra explicación para su comportamiento.

De este modo en la mente de los investigadores paulatinamente surgía la necesidad de acoplar cada vez más la teoría de probabilidades al estudio estadístico de los fenómenos aleatorios (Hald, 2003), pero también a la inversa, de aplicar los modelos generados a la realidad en la que vivían.

Una de las ciudades que llevaba el recuento de las defunciones y nacimientos con mucha regularidad y minuciosidad fue Breslau. Alrededor de 1690 el registro había sido llevado por más de un siglo y la cantidad de datos acumulados atrajeron el interés de dos investigadores interesados en resolver distintas cuestiones: Gaspar Neumann y Edmund Halley. Después de revisar miles de partidas de defunción, Neumann probó la falsedad de la antigua creencia popular de que en las edades «climatéricas» (múltiplos de 7) hay más propensión a adquirir enfermedades y morir que en el resto de su vida. Sin embargo demostró que sí había ciertas edades (en la infancia y en la ancianidad) en las que el ser humano estaba más susceptible de morir, puesto que ahí es donde se presentaban mayores defunciones.

Los datos de Breslau fueron conocidos por Halley gracias al trabajo de Neumann y los aplicó en el estudio de la vida humana. Hizo un exhaustivo análisis de las tasas de mortalidad y calculó las «posibilidades»[1] que tiene una persona de morir a una cierta edad para regular el precio de los seguros de vida. Sus cálculos fueron la base de los trabajos de esperanza de vida y los índices de mortalidad que hoy utilizan muchas las compañías de seguros.

Tanto Neumann como Halley analizaron la variable estadística: «edad de las personas que mueren» y la relacionaron con la mortalidad anual (Halley, 1693, p 599). Ambos trabajaron con variables estadísticas y sus distribuciones e incluso trataron de ajustar curvas matemáticas para hacer predicciones. De alguna forma, fueron precursores de la función de densidad, aunque no usaron explícitamente las ideas de variable estadística o aleatoria.

Halley también se dedicó a la astronomía y es de suponerse que conocía el trabajo de Galileo. Su más famoso descubrimiento fue la predicción del año de aparición del cometa que hoy lleva su nombre. Gracias al análisis de viejos datos acumulados descubrió que las orbitas de los cometas aparecidos en 1531, 1607 y 1682, eran muy parecidas y dedujo que correspondían a un mismo cometa. Estimó que el periodo de aparición era de 76 años y predijo que volvería a aparecer en 1758. Su experiencia en estos trabajos hizo que le llamaran la atención los datos de Breslau. Halley empleó los razonamientos estadísticos usados en astronomía para estudiar datos poblacionales y sociales y los hizo progresar hacia nuevas herramientas de análisis. También se percató de la valía de los grandes números, al mencionar que de tener un mayor número de datos, el cálculo sería más confiable:

> …he realizado lo que el corto periodo de observaciones ha permitido con relación a la exactitud… Si su autor, el Dr. Neumann pudiera continuar con ellas de la misma manera por algunos años más, entonces las irregularidades eventuales y la aparentemente discordancia de la tabla podrían ser rectificadas y establecerse con un cierto número de posibilidades (Halley, 1693, p 656).

Trabajando con un problema muy distinto, Galileo Galilei también se ocupó por el recuento de datos y por el análisis en una variable estadística continua. Desde la época de la Revolución Copernicana, los astrónomos organizaban los registros de las observaciones del movimiento de los planetas. Galileo concentró su atención en los diferentes valores que se obtienen al realizar varias veces la misma medida para

[1] Halley (1693) usa la palabra «odds» y en algunas ocasiones la palabra «probability», aparentemente en el mismo sentido. Sin embargo, no debemos confundir su definición con las actuales acepciones de nomios o probabilidad. En Halley traduciremos ambos términos como «posibilidad».

encontrar la distancia de una estrella a la tierra de manera indirecta. Analizó las propiedades de los *errores observados* (aleatorios) y encontró que son más frecuentes los errores pequeños que los grandes, que los errores por exceso son tan frecuentes como los errores por defecto y que la mayoría de las mediciones se agrupan alrededor del valor medio. Además definió una única *distancia verdadera* y *observaciones* sujetas a variaciones (Hald, 1986). Es claro que en estos análisis Galileo manejó implícitamente variables aleatorias y estadísticas y sus distribuciones (aunque nunca las mencionó como tales), pero también con «el cuadrado del error medio», que es nuestra actual *varianza*. Los errores son una variable aleatoria de distribución desconocida y usó a la variable estadística para aproximarse a ella. Actualmente sabemos que esa distribución a la que Galileo se aproximó es la Normal estándar.

Otros grandes matemáticos, como Daniel Bernoulli (1700-1782), se enfocaron al estudio de fenómenos sociales y económicos basándose en datos empíricos. En su obra Bernoulli (1954/1738) afirmó que un comerciante racional siempre debe actuar motivado por obtener la máxima utilidad posible, la cual puede ser calculada mediante procedimientos matemáticos. Además, debe calcular cuál es el riesgo que existe para alcanzar ese premio (la utilidad), y la manera como ese riesgo incide económicamente en la expectativa de ganancia. De este modo Daniel Bernoulli inicia la econometría, es decir, la utilización de datos estadísticos para prever y calcular los riesgos más comunes que fortuitamente pueden afectar las expectativas de ganancia en un negocio, así como la aplicación de la teoría de la probabilidad para adoptar decisiones acertadas frente a los riesgos posibles fortuitos que pueden ocurrir. Estas teorías y modelos matemáticos fueron aprovechados en primer lugar por las empresas aseguradoras y posteriormente en las transacciones comerciales de los mercaderes, prestamistas y burgueses que darían impulso al surgimiento de las primeras fábricas dentro de lo que más tarde sería conocido como la Revolución Industrial.

La obra de Bernoulli sirvió de referente y fundamento para el tratamiento matemático de conceptos tales como la aversión al riesgo, las reservas de capital que se deben hacer para cubrir posibles contingencias, la toma de decisiones en estado de incertidumbre y la teoría de la ruina. Es de suponerse que los primeros casinos que surgieron durante el siglo XVIII comenzaron a tener en cuenta ese conocimiento matemático en la explotación comercial de las apuestas en torno de los juegos de azar (Hald, 1986).

Durante este momento histórico, los registros estadísticos realizados con fines diversos dieron lugar también a la comprensión y mejor análisis de la variable estadística y a vincularla cada vez más con la variable aleatoria, aunque de una forma implícita. Observemos que este avance no hubiera sido posible si no se hubiera percibido la importancia de la variable numérica como una forma de vincularse con los resultados que se estaban obteniendo en probabilidad.

4. Teorema central del límite y modelos generales de distribuciones

A finales del siglo XVII, la introducción de procedimientos y análisis matemáticos en la descripción de fenómenos demográficos, políticos y sociales ayudó a que grandes matemáticos de la época se percataran de la importancia de los datos observados e impulsó la necesidad de abstraer y formalizar el conocimiento estadístico sobre los sucesos aleatorios que hasta entonces se había utilizado en la resolución de problemas específicos.

La noción de variable aleatoria apareció desde los primeros intentos por definir los conceptos generales que había detrás de los casos particulares, pero no había necesidad de definirla como objeto matemático o hacerla explícita. Los primeros pasos en su definición los dio Simón D. Poisson. Posteriormente Pavnutii Chebyshev y algunos de sus discípulos la hicieron patente y evidenciaron sus propiedades. Sin embargo, anteriormente Jacob Bernoulli, Abraham De Moivre y Pierre Simón Laplace no sólo la habían manejado implícitamente en sus teorías y conceptualizaciones, sino también comenzaron a hacerla explícita en casos particulares. Su avance histórico aparece ligado fuertemente al desarrollo progresivo del teorema central del límite y, paralelamente, al estudio de los modelos generales de estas distribuciones.

En este apartado haremos un estudio de este desarrollo. Retomaremos el trabajo de Alvarado (2007), quien no analizó la emergencia de la variable aleatoria en este periodo. Detallaremos las aportaciones de los diferentes autores a esta evolución, desde Jacob Bernoulli hasta Pavnutii Chebyshev y sus discípulos.

4.1. Las aportaciones de Jacob Bernoulli

La obra *El Arte de la Conjetura* (Bernoulli, 2006/1713), fue publicada póstumamente y se estima que fue escrita entre 1685 y 1705 (Hald, 2003 y Sylla, 2006). Es considerada uno de los trabajos más importantes en probabilidad (García, 1971; Hald 2003 y Sylla, 2006). La intención de Bernoulli era fundamentar el cálculo de probabilidades como un instrumento que mejora la habilidad de los ciudadanos para tomar decisiones en ámbitos diversos y situaciones complejas. No logró concluir su obra, pero sí estableció un punto de inflexión entre las matemáticas de los juegos de azar y una teoría matemática del azar, que podía aplicarse a los asuntos morales, cívicos y económicos. De importancia tanto filosófica como matemática, en este trabajo no sólo encontramos las primeras demostraciones probabilísticas, sino también mejores aproximaciones al concepto de probabilidad, cosa que permite, por vez primera, establecerlo en un sentido epistemológico (Sylla, 2006).

La principal aportación de Bernoulli es el enunciado del primer teorema del límite en la teoría de probabilidad, al que él llamó *proposición principal*, pero no es menos importante que lo demostrara de manera rigurosa (Bernoulli, 2006/1713, pp. 330-335). Hald (2003) tradujo a la nomenclatura actual la comprobación de Bernoulli de la ley débil de los grandes números y mostró que está completa y es rigurosa (pp 259-262). Aunado a este rigor matemático, Bernoulli visualiza que debe haber una tolerancia del resultado empírico con respecto al teórico, que permitiría aplicar la probabilidad a ámbitos en los que no se pueden conocer la probabilidad teórica de los resultados posibles de un experimento aleatorio. Es decir, establece un puente entre el mundo empírico con el teórico. Desde nuestro objeto de interés, el establecimiento de una «tolerancia» implica que la variable estadística y la variable aleatoria nunca podrán llegar a ser las mismas.

Bernoulli (2006/1713) también hace un tratamiento sistemático de la combinatoria como una forma de conocer los distintos modos en que se puede dar un acontecimiento y así calcular su probabilidad (3ª parte, pp 193-250). Sin embargo él también fue consciente que ésta teoría sería poco empleada para obtener la probabilidad en situaciones distintas a las de los juegos de azar y propuso obtener la probabilidad a través de aproximaciones sobre la base de frecuencias empíricas:

> si finalmente la certeza moral es alcanzada de esta forma (lo cual demuestro en el siguiente capítulo), podemos encontrar a posteriori el número de casos [favorables y totales] casi con la misma certeza como si los hubiéramos conocido a priori» (p. 328).

Bernoulli concibe la probabilidad como el *grado de certeza* (conocimiento) que los seres humanos podemos tener de los acontecimientos (pp. 315-316) y la deslinda de la asignación, que, afirma, puede hacerse de dos formas: a través del conteo de los casos posibles y totales en el espacio muestral (la actual probabilidad laplaciana) y a través de la observación múltiple de los resultados de pruebas similares (la actual probabilidad frecuencial). Intuitivamente, Bernoulli sabía que las frecuencias relativas tendrían que converger a la auténtica probabilidad y creyó que podría estimarla con el grado de precisión deseado e incluso determinar el número de pruebas necesarias para garantizar esa precisión. Estos intentos fueron los que lo condujeron a la formulación de lo que actualmente conocemos como la ley débil de los grandes números o teorema de Bernoulli.

Bernoulli también se preocupa por hacer comprensible su discurso y usa notación matemática para la definición de sus variables. Es de esperarse que este uso más explícito y metódico de sus variables facilitara la generalización de su teoría. Observemos el uso de las variables en la generalización de la solución de Huygens al problema de las tiradas de dos o más dados en su XI proposición:

> …el número de formas en las cuales *m* puede ser obtenido tirando *n* dados es igual al coeficiente de x^m en el desarrollo de $(x + x^2 + x^3 + x^4 + x^5 + x^6)^n$ como una serie de potencias de *x* (Bernoulli, 2006/1713, p 157).

Aunque en la notación de Bernoulli no distingue una variable de un parámetro, es evidente que en este problema su variable de interés es *m*: «la suma de los números obtenidos en las caras superiores de los dados», a la que llama *resultado*. Es notorio que Bernoulli percibe que, dado el número *n* de dados, *m* será la que varíe durante una partida. También se observa que *n* es un parámetro y *x* sólo es una variable auxiliar que sólo está definida por la forma en que Bernoulli desarrollaba las combinaciones. Es de llamar la atención que la variable auxiliar es representada por una última letra del alfabeto. Es decir, Bernoulli ya comienza a tener una notación diferenciada.

En el Arte de la Conjetura, se retoman y se proporcionan nuevos métodos de solución a los problemas tratados por Huygens (1714/1657). Bernoulli también se interesó por generalizar las soluciones, variando todas las condiciones posibles. Llegó a relacionar algunos de los valores de las variables aleatorias con su probabilidad y en algunos casos, llegó a obtener la distribución de probabilidad completa, como en el caso del problema de la propuesta interrumpida, en donde expresa la distribución de probabilidad en forma tabular o en el caso del problema de la ruina del jugador. En este sentido, su aportación más importante fue la expresión general de la distribución que actualmente conocemos como *Binomial*, y que formula de la siguiente manera:

> …la expresión general para la probabilidad de que [el éxito] ocurra al menos *m* veces en *n* intentos, cuando es conocida la probabilidad de un intento, consistiría de los términos de la expansión de $(b/a + c/a)^n$ desde $(b/a)^n$ hasta el término que involucra $(b/a)^m\ (c/a)^{n-m}$, ambos inclusive, en donde *b/a* y *c/a* es la probabilidad de éxito y falla en un intento, respectivamente (Bernoulli 2006/1713, p. 165).

La función de distribución Binomial usada por Bernoulli tendría su equivalente actual en:

1-B(n-m,n,q) = B(m-1,n,p), donde p=b/a y q=c/a. (Hald, 2003, pp. 226-227)

En esta fórmula, Bernoulli identifica el ensayo que actualmente conocemos como *ensayo Bernoulli*, y se percata de la necesidad de independencia entre uno y otro ensayo. También trabaja implícitamente con dos tipos de variables aleatorias: (1) una variable asociada directamente con el ensayo Bernoulli, a la que se le asocia 1 ó 0 para éxito o fracaso, respectivamente; y (2) otra variable aleatoria corresponde a la suma de éxitos en *n* ensayos, es decir, es la suma de *n* variables aleatorias Bernoulli.

Una tercera variable aleatoria implicada no usada por Bernoulli como tal es la proporción entre el número de éxitos y el tamaño *n* de la muestra extraída. Aparentemente, él comprendió que al realizar empíricamente varios ensayos Bernoulli, la proporción del número de éxitos con respecto al total, variaría, pero no llegó a establecer una teoría de las proporciones muestrales con *n* fija. En cambio, se enfocó a lo que ocurriría si se incrementaba el tamaño *n* de la muestra. Él se propuso demostrar matemáticamente que empíricamente la proporción convergería en el valor numérico de la probabilidad conforme *n* se hiciera cada vez más grande (la ley débil de los grandes números). Bernoulli trabajó con intervalos de proporciones que fluctuaran alrededor de la probabilidad teórica: buscó el número de ensayos tal que la probabilidad de que la proporción de un suceso diferiría de la probabilidad del suceso tuviera una probabilidad muy cercana a uno, a la que llamó certeza moral.

Bernoulli recurrió al ejemplo de extraer bolas de una urna en la que había 3000 bolas blancas y 2000 bolas negras. Esta proporción era desconocida para el observador encargado de extraer con remplazo una bola al azar un número grande de veces. El resultado era anotando cada vez. De esta forma, Bernoulli también trabajó con las variables estadísticas: «1, si la bola extraída era blanca, 0 si la bola extraída era negra» y con «el número de bolas blancas extraídas en *n* ensayos». Desde una perspectiva teórica, se conoce cómo se comportarían estas variables, y por lo tanto, sus correspondientes variables aleatorias son las mencionadas en el párrafo anterior, pero desde una perspectiva empírica (para un observador), su distribución era desconocida. La preocupación de Bernoulli estaba enfocada en estimar la probabilidad cuando no fuera posible deducirla teóricamente, lo que implicaba necesariamente recuento de datos y por lo tanto, el uso de la variable estadística. Esto condujo no sólo a trabajar con ambos tipos de variables sino también se ocupó de su correspondencia y convergencia cuando se quiere estimar la distribución de probabilidad a través de un procedimiento empírico.

Jacob Bernoulli no llegó a aplicar sus resultados a asuntos civiles, morales o económicos, como prometía el título de la cuarta parte de su libro, pero su sobrino, Nicolas Bernoulli sí lo hizo, haciendo más evidente la relación entre la variable aleatoria y la estadística. Nicolas también mejoró el método de aproximación de Jacob a la Binomial con una forma de cálculo más fácil, que posteriormente de Moivre retomaría para introducirse al estudio de la distribución normal (Hald, 2003).

4.2. Las aportaciones de Abraham de Moivre

De Moivre (1756)[2] estableció un cambio con respecto a anteriores libros de probabilidad. Primero, porque el lenguaje evolucionó notablemente después de la publicación del Arte de la Conjetura ya que se comenzó a abandonar el latín y a escribirse o a traducirse contemporáneamente al inglés o al idioma del autor, lo que hizo que un vocabulario especialista se desarrollara más rápidamente al trabajar con un idioma vivo. En segundo lugar porque de Moivre ya mostraba un enfoque conceptual distinto, en donde separó claramente la probabilidad de un suceso de su valor o de su esperanza. En su tercera edición (1756) estableció el paradigma de la probabilidad matemática, desligándose de problemas filosóficos y dando un sustento teórico a todas sus proposiciones (Sylla, 2006). Influenciado por el rigor matemático de la época, se preocupó por establecer definiciones y reglas previas a la aplicación en problemas y por graduar la dificultad de los problemas que resuelve hasta llegar a generalizaciones.

De acuerdo con Pearson (1924), de Moivre escribió el primer tratamiento de la probabilidad integral y la esencia de la curva Normal. Las herramientas matemáticas aportadas al campo de la probabilidad fueron diversas. La regulación, profundización y generalización de los problemas le permitió el estudio del comportamiento de las distribuciones de probabilidad de diversas variables aleatorias. También estudió desde una perspectiva teórica la convergencia de la frecuencia relativa hacia la probabilidad, lo que lo condujo a las sentar bases para caracterizar una distribución de probabilidad y a la expresión de la distribución normal.

Estudió el mismo problema planteado por Bernoulli partiendo de distinta perspectiva. Mientras que Bernoulli estaba interesado en definir el número de ensayos que habría que hacer para conocer la probabilidad de éxito en sucesos en donde no era posible conocerla teóricamente, de Moivre buscaba conocer el grado de precisión de una serie de ensayos en donde se conoce la probabilidad de efectuar un ensayo. La

[2] Aunque De Moivre publicó diversos artículos sobre probabilidad en las *Philosophical Transactions* a partir de 1708, todos esos trabajos se concentraron en las tres ediciones de su obra *La Doctrina del azar* (1718, 1738 y 1756).

respectiva solución de ambos estaba en función de la probabilidad del intervalo del número de éxitos, lo que les obligaba a trabajar con un coeficiente binomial que contenía factoriales de números muy grandes. En palabras de de Moivre, éste era «el problema más difícil que puede ser propuesto en el tema del azar» (de Moivre, 1756, p 242). Jacob y Nicolas Bernoulli propusieron su cálculo a través de aproximaciones suficientes para sustentar la Ley de los grandes números de Jacob Bernoulli. De Moivre buscó una mejor aproximación y encontró una primera expresión de la distribución Normal y la aproximación de la distribución binomial a través de distribución normal, que sería una de las ideas precursoras del teorema central del límite (Alvarado, 2007) o el primer ejemplo de este teorema (Hald, 2003).

De Moivre partió de un problema que, en su forma más general dice:

> *A* y *B* compiten con diferentes posibilidades de ganar un juego, supondré que las posibilidades respectivamente son *a* es a *b*. Después de un cierto número de juegos *n*, la competencia se acaba y el jugador *A* le dará a un espectador *S* tantas piezas como número de juegos ganados tenga por encima de $\frac{a}{a+b}n$ y *B* le dará tantas piezas como juegos ganados tenga por encima de $\frac{b}{a+b}n$. Encontrar la esperanza de *S* (de Moivre, 1756, p. 241).

La esperanza de *S* también podría haber sido interpretada como «la medida de la dispersión de la frecuencia relativa alrededor de la probabilidad *p*» (Hald, 2003, p 471). La solución de este problema condujo a de Moivre a la Ley de los grandes números, formulada a partir de la esperanza de la diferencia entre la probabilidad y la frecuencia relativa. También a preguntarse por las probabilidades en contra de que en una serie de ensayos se presente una desviación del valor esperado muy grande. Este último problema, él lo enuncio como «el grado de asentimiento que le es dado a los experimentos» (de Moivre, 1756, p. 242).

La generalización de ese problema, suponiendo que cada evento tiene las mismas probabilidades de ocurrir, le permitió demostrar que la mayoría de los datos tienden a agruparse alrededor de un valor central (la media), y delimitó el tamaño de la zona donde se deben encontrar la mayoría de los datos:

> …ese evento no aparecerá frecuentemente más de $\frac{1}{2}n + \frac{1}{2}\sqrt{n}$ veces, ni raramente más de $\frac{1}{2}n - \frac{1}{2}\sqrt{n}$ veces, que se puede expresar por la suma doble del valor obtenido en el segundo corolario, esto por 0,682688 (de Moivre, 1757, p. 246).

Con lo cual describió teóricamente, sin percatarse de ello, lo que Galileo ya había notado empíricamente con respecto al comportamiento de la distribución de los errores aleatorios (Normal estándar), pero también descubrió un método para describir las desviaciones con respecto a un valor central, sentando las bases para caracterizar las distribuciones con objetos matemáticos muy usados actualmente, tales como la media, la desviación típica, la regularidad estadística y la simetría.

El método de aproximación de la suma de los términos del binomial, también condujeron a otros aportes importantes. De Moivre encontró la ecuación general: $y = y_o e^{\frac{-l^2}{2^2}}$, halló el valor de $y_o = \frac{1}{\sqrt{2\pi}}$ (con la ayuda de su amigo James Stirling) y mostró que su integral es una buena aproximación para el cálculo de probabilidades de la distribución Binomial cuando n es grande, por lo que muchos autores consideran a de Moivre como el padre de la distribución Normal. Además, con este resultado también determinó una relación importante entre variables aleatorias *discretas* y *continuas* y estableció las bases para el desarrollo posterior del teorema central del límite (Alvarado, 2007). Su trabajo en pro de la Normal también daría lugar a la idea de función de densidad para las variables continuas.

Su afán en la generalización de los problemas que trabajó, también llevaron a de Moivre a deducir el método de cálculo de algunas actuales distribuciones de probabilidad. Así, el problema de despejar el número de ensayos de la distribución Binomial para obtener una cierta probabilidad, lo condujo a deducir la distribución Poisson (de Moivre, 1984/1712, pp 239-243, y Hald, 1986 y 2003) cuando se propuso calcular el número n de ensayos necesarios para que un jugador compitiera de manera equitativa si su apuesta fuera obtener c éxitos en n ensayos y la probabilidad p de un éxito fuera conocida. Así mismo, dedujo la distribución Hipergeométrica (de Moivre, 1984/1712, pp 247-248) al generalizar el problema de obtener 3 fichas blancas al sacar 7 fichas al azar de un recipiente que contiene 4 fichas blancas y 8 negras. Otra distribución más que sobresale en sus trabajos, fue la Binomial negativa al generalizar el problema de la división de una apuesta (de Moivre, 1984/1712, pp 243-244).

De Moivre también era consciente de que el modelo teórico no necesariamente tenía que concordar con lo que se obtenía al realizar experimentos y de que habría que tener cuidado al aplicar la teoría a problemas prácticos:

> Los sucesos en una larga racha podrían caer en una proporción diferente de la real… Por lo tanto, se deberían asignar momios en contra de una variación tan grande de la

igualdad, así la mente estaría mejor dispuesta a las conclusiones derivadas de los experimentos (de Moivre, 1756, p 242).

Sus artículos y libros son totalmente deductivos, por lo que es posible que de Moivre no se haya ocupado de colectar o analizar observaciones. Sin embargo, lo cierto es que profundizó teóricamente en la probabilidad frecuentista y en su relación con un modelo teórico ideal, que, a largo plazo, conduciría a la relación entre la variable estadística y la aleatoria.

Es importante notar que, aunque logra generalizaciones y ecuaciones que posteriormente darían paso a la formulación de diversas distribuciones, él las establece como métodos para el cálculo de probabilidades y no como distribuciones de probabilidad o funciones de densidad, puesto que para ello era necesario el concepto de variable aleatoria que aún no se visualizaba.

Así mismo, se observa que a pesar de sus contribuciones a la aproximación de la Binomial a través de la Normal, de Moivre no trabajó con variables continuas explícitamente, ni tampoco hay indicios de que él pudiera percatarse de la existencia de un tipo de variables que no fueran numéricas. Sin embargo pareciera una necesidad de la época. Si las variables enunciadas por de Moivre no hubiesen sido numéricas, no le hubiese sido posible aplicar el análisis matemático en boga ni tampoco obtener generalizaciones. Esto concordaba enteramente con el pensamiento científico de entonces, en el que se apoyaba la vertiente de las ciencias exactas y la imagen ideal de un universo determinista. Gran parte de este pensamiento se reflejaba en el lenguaje empleado y la estructuración de sus escritos.

En esa época, los científicos usaban la idea de variable ligada al estudio del análisis matemático. Comúnmente la llamaban *cantidad* o *magnitud variable*, lo cual evidenciaba su carácter de cualidad vinculada con la medición, proceso a través del cual, la cualidad podría tomar distintos valores. La economía del lenguaje fue dando paso al uso de una sola palabra y la distinción entre las cantidades que no varían y las que sí, hizo que predominara la sustantivación del adjetivo y fuera nombrada solamente «variable» (Gómez, 2008). Actualmente se conceptualizan distintos tipos de variables, así mismo la idea de variable numérica se concibe de manera distinta en el ámbito determinista que en el probabilista, sin embargo en este momento histórico no se distinguía un tipo de ciencia de la otra y los probabilistas usaban el término bajo una acepción similar a la que los calculistas empleaban para denotar sus variables. No se vislumbraba, aún, la necesidad de distinguir un concepto propio para el ámbito probabilista.

4.3. Las aportaciones de Gauss, Legendre y Laplace

A comienzos del siglo XIX, los matemáticos Johann Karl Friedrich Gauss (1777-1855) Adrien−Marie Legendre (1752-1833) y Pierre−Simon Laplace (1749-1827) retomaron el análisis numérico de los errores de mediciones en física y astronomía, a propósito del problema de cómo determinar el mejor valor leído por un instrumento que entrega diferentes mediciones del mismo fenómeno.

Desde finales del siglo XVI, Galileo ya había detectado algunas características de los errores de las medidas astronómicas (ver apartado 3 de este capítulo) y los había clasificado en dos tipos: los *sistemáticos*, debidos a los métodos y las herramientas de medida; y los *observados* (*aleatorios)*, que varían impredeciblemente de una medida a otra. Sin embargo no había una teoría que sustentara cuál debía ser el valor más confiable de las observaciones empíricas, ni un método que determinara numéricamente el comportamiento aleatorio descrito por Galileo. No fue sino hasta principios del siglo XIX que Legendre, en 1807, y Gauss, en 1809, basándose en el trabajo de de Moivre, propusieron y desarrollaron un método para determinar el valor único *z* de la medición, de manera que una función de los *errores cuadráticos* fuera mínima. La solución dio como resultado el promedio de las mediciones, valor alrededor del cual se agrupan, y el principio del método de mínimos cuadrados para la estimación de parámetros. Gauss, además, demostró que, bajo ciertas condiciones generales, la probabilidad de los errores de medida podía calcularse mediante la forma $\varphi(\Delta) = \frac{h}{\sqrt{\pi}} e^{-h^2\Delta^2}$ y relacionó los errores de las mediciones con la aproximación a la binomial cuando *n* es grande. También reformuló algunos de los corolarios de de Moivre y manejó más implícitamente ideas semejantes a los actuales de varianza y desviación típica para calcular el error de la estimación de probabilidad.

Pierre Simón Laplace también se interesó y aplicó con éxito el cálculo de probabilidades en problemas de las medidas astronómicas, pero, del trabajo de de Moivre, también retomó la aproximación a una Binomial a través de una Normal. Ello dio lugar a una versión del teorema central del límite para una suma de variables aleatorias idénticamente distribuidas (Teorema de Moivre-Laplace), en donde generalizó el resultado para el caso discreto (cualquier *p*) e introdujo algunos casos continuos, aunque no lo demostró para una función arbitraria (Alvarado, 2007). Este resultado extendió y profundizó el uso de las variables aleatorias, con la identificación de las condiciones de *independencia* e *idénticamente distribuidas*, así como con el inició del trabajo con variables continuas. Más tarde, este resultado devendrá en el planteamiento de distribuciones muestrales, lo que también contribuirá a la vinculación entre las variables aleatoria y estadística.

En la recopilación de sus trabajos en *Teoría analítica de las probabilidades*, Laplace (1812) iría más allá del cálculo clásico de probabilidades. Su objetivo fue establecer una teoría que sirviera a la gente para tomar «decisiones razonadas». En él, además del rasgo cualitativo de aplicabilidad, emerge la relevancia de los métodos analíticos específicos de una teoría de probabilidad. Su obra no sólo es importante por la sistematización del conocimiento sino también por el uso de un lenguaje moderno y la incorporación palpable del análisis matemático en probabilidad. Sus antecesores usaban una descripción narrada para indicar la forma del cálculo de probabilidades, Laplace formula ecuaciones y funciones e incluso usa explícitamente los términos *función* y *ecuación*. La más notable herramienta que incorpora del análisis matemático es la integral (incluso en su notación matemática) para el cálculo de las cotas dentro de las cuáles se encontraría la probabilidad de la suma de un número grande de variables aleatorias. Con lo que, de alguna manera, comienza a trabajar con la idea de lo que más adelante sería la *función de densidad*, aunque aún no de manera tan completa, puesto que su interés no era propiamente la definición de una función de la variable aleatoria, sino una función que le proporcionara los valores de las cotas entre las que estaría la probabilidad buscada. Éste uso del lenguaje y del análisis, también le da acceso a un mayor uso y generalización a las variables. En el siguiente ejemplo se puede observar el nivel de formalización de la variable aleatoria:

> Sean *i* cantidades variables y positivas $t, t_1, t_2, \ldots, t_{i-1}$, de las cuáles su suma es *s*, y de las cuales la ley de posibilidad es conocida; se propone encontrar la suma de los productos de cada valor que pueda recibir una función dada $\psi(t, t_1, t_2, \ldots)$ de esas variables, multiplicadas por la probabilidad correspondiente a este valor (Laplace, 1812, p. 262).

En el texto anterior se identifica el uso de los términos *cantidades variables* para referirse a las variables aleatorias *t* así como a su carácter *numérico* y hace referencia a su dominio, que hoy entendemos como los números reales positivos. Por otra parte, Laplace identificaba claramente las *i* variables aleatorias $t, t_1, t_2, \ldots, t_{i-1}$, así como otra variable *s* correspondiente a su suma. Así mismo, realiza operaciones sobre las variables y sus probabilidades. Todos son elementos esenciales de la composición de las variables aleatorias entendidas desde la perspectiva moderna.

Laplace utiliza con frecuencia el término *ley de posibilidad* a lo largo de todo su trabajo y lo distingue de la *probabilidad*, como, a modo de ejemplo, puede verse más claramente en esta otra cita donde acota los valores que puede tener su variable α:

> Ahora, si conociéramos el límite y la ley de posibilidades de los valores de α, nada sería más fácil que resolver exactamente este problema; porque si nombramos q a este límite y si representamos por $\varphi(\alpha)$ a la posibilidad de α, vemos primero que α tiene necesariamente que caer entre 0 y q, la función $\varphi(\alpha)$ debe ser tal que tenemos: $\int d\varphi(\alpha) = 1$. (Laplace, 1778, p. 394).

Aquí usa notación funcional $\varphi(\alpha)$, para la ley de posibilidades de α, y la distingue de su probabilidad. Así, esta última pareciera ser el antecedente de lo que hoy entendemos como función de distribución y la ley de posibilidades sería la función de densidad. Sin embargo, estos conceptos están entremezclados aún, porque $\int d\varphi(\alpha) = 1$, sólo tendría sentido si α fuera continua, pero entonces $\varphi(\alpha)$ no sería la posibilidad, a menos que α fuera discreta (y en tal caso la integral tendría que ser sumatoria). Lo cierto es que ya está presente la distinción entre una función útil para el cálculo de probabilidades y la probabilidad misma.

Desde Bernoulli, las distintas soluciones sobre el problema de acotar la probabilidad de la suma de variables aleatorias, hicieron que se comenzara a trabajar de una forma muy natural con funciones en las que la variable dependiente (los valores obtenidos a partir de aplicar la función sobre un valor de la variable aleatoria) no proporciona la probabilidad, pero que resultaba útil para su cálculo o para aproximarlo (aplicando la integral, por ejemplo). Laplace introduce el término *ley de posibilidad*, que se distingue de la probabilidad, porque se había percatado que la ley de posibilidad era útil para el cálculo de probabilidades, pero en algunos casos no proporciona directamente la probabilidad. Notemos también que el interés explícito de Laplace era establecer las bases de la ciencia de la probabilidad, pero su interés de facto era el cálculo de las probabilidades.

En esta época no se había formalizado completamente la definición de la variable aleatoria ni la de función de densidad. En la obra de Laplace puede observarse que cuando se refiere a variables, lo hace respecto a resultados numéricos de experimentos aleatorios[3]. Sin embargo, cuando Laplace establece operaciones entre

[3] que es la forma en como propone Parzen que se inicie el estudio de las variables aleatorias en estudiantes universitarios (ver apartado 6.2 de este capítulo o apartado 2 del capítulo 3). En forma

variables, hace explícita la regla a través de la cual se calcula una nueva variable. También se nota que, intuitivamente, para él, la nueva variable, que surgía a partir otras variables aleatorias, era de la misma naturaleza aleatoria que sus precedentes, puesto que buscaba acotar su probabilidad.

Hay que recordar que a pesar de la definición de probabilidad que privaba en Laplace y aunque eran manejados de manera informal, en esa época, no había una definición de suceso o espacio muestral, lo que limita el surgimiento de la idea de la variable aleatoria como una función conjunto de valor real, comprensión que sólo llegaría con la axiomatización de la probabilidad por medio de la teoría de conjuntos. Lo cual implica que el establecimiento de la probabilidad como ciencia también está ligado con el desarrollo de la variable aleatoria.

Un aspecto importante, es que en esta época no había una definición del concepto de aleatoriedad. Laplace vio la necesidad de hacer explícitas las bases filosóficas de su teoría y diferenciarlas de su trabajo matemático. En su *Ensayo filosófico sobre las probabilidades* (2006/1814), se sientan los principios filosóficos y resultados generales básicos sobre los que descansa toda su teoría. Algunos de sus principios no son válidos actualmente y también hay contradicciones y redundancias en sus argumentos, pero al hacerlos explícitos, refleja el pensamiento de su época y permite vislumbrar los alcances y limitaciones de su teoría analítica de las probabilidades. En este trabajo se nota, por ejemplo, cual es la definición de aleatoriedad que privaba en el ámbito científico de entonces y que permea toda su obra. Laplace expresa su concepción de aleatoriedad cuando concibe una probabilidad más bien como un mal menor, fruto de nuestra ignorancia:

> Todos los acontecimientos, incluso aquellos que por su insignificancia parecen no atenerse a las grandes leyes de la naturaleza, no son sino una secuencia tan necesaria como las revoluciones del Sol. Al ignorar los lazos que los unen al sistema total del universo, se los ha hecho depender de causas finales o del azar, según que ocurrieran o se sucedieran con regularidad o sin orden aparente, pero estas causas imaginarias han ido siendo descartadas a medida que se han ido ampliando las fronteras de nuestro conocimiento, y desaparecen por completo ante la sana filosofía que no ve en ellas más que la expresión de nuestra ignorancia de las verdaderas causas (Laplace, 2006/1814, p. 24).

Además, en esta cita, se trasluce su apego a la concepción positivista de la ciencia (que influye en el uso que hace de la idea de *cantidad variable*) y que también

resumida, en los fenómenos aleatorios con resultado numérico la función que vincula el espacio muestral con los números reales es la función identidad.

se observa en el desarrollo de su *Teoría analítica de las probabilidades* (1812). Hawking (2006) llamó al *Ensayo filosófico de las probabilidades* la más argumentación más completa y perfecta para una interpretación determinista del mundo.

4.4. Las aportaciones de Simeón Denis Poisson

En su trabajo Recherches *sur la probabilite des jugements en matiere criminelle et en matiere civile,* Poisson (1837) pareciera hacer uso de un concepto intuitivo de la variable aleatoria como función. El problema era sobre la probabilidad de las causas de un envenenamiento observado:

> … llamaremos $C_1, C_2, C_3, ..., C_n, ..., C_m$, a las *m* causas posibles del envenenamiento *E*; siendo $p_1, p_2, p_3, ..., p_n, ..., p_m$, las probabilidades conocidas de sus correspondientes entradas, relativas a esta variedad de posibles causas; de manera que p_n expresa la probabilidad de que *E* tuviera lugar si la causa C_n fuera única, o, lo que es la misma cosa, si ella fuera cierta, excluiría a todas las demás (Poisson, 1837, p 82).

Para resolverlo, Poisson hace la analogía de las *m* causas posibles con un mismo número de urnas $A_1, A_2, A_3, ..., A_n, ..., A_m$. Cada una con una probabilidad de sacar bola blanca de $p_1, p_2, p_3, ..., p_n, ..., p_m$. Luego, expresa esas probabilidades con un común denominador μ :

> Por eso, supongamos que reducimos las fracciones $p_1, p_2, p_3,$ etc., a un mismo denominador, tendríamos:
>
> $$p_1 = \frac{\alpha_1}{\mu},\ \ p_2 = \frac{\alpha_2}{\mu},\ \ p_3 = \frac{\alpha_3}{\mu},\ ...\ p_n = \frac{\alpha_n}{\mu}, ... p_m = \frac{\alpha_m}{\mu};$$
>
> μ y los numeradores $\alpha_1, \alpha_2, \alpha_3$, etc., son unos números enteros. No alteraríamos nada si tomamos al azar una bola blanca de la una A_n, reemplazando las bolas que contiene, para un número α_n de bolas blancas de un número μ de bolas, tanto blancas como negras; y lo mismo para las otras urnas. (Poisson, 1837, pp 82-83).

Se puede observar la presencia de una variable cualitativa: las causas posibles C_n (sucesos) con su probabilidad conocida, pero también Poisson trabaja con una variable numérica α_n asociada, por una parte al suceso C_n, y por otra, a una probabilidad p_n. α pudiera interpretarse como una variable aleatoria discreta. El caso es singular porque, a diferencia de Laplace que estaba interesado en problemas que involucran magnitudes aleatorias, Poisson parte de una variable cualitativa –las posibles causas de envenenamiento- a la cual se le asoció una variable cuantitativa α con probabilidades específicas conocidas. No sólo comienza a distinguir entre valor de la variable y el suceso, sino también se ocupa de que la regla que los vincula tenga relación con el suceso y genere un número real.

Esto se puede observar también en una segunda demostración del teorema central del límite desde un punto de vista más general. Ahí Poisson escribió:

> Supondremos que alguna cosa, a lo que llamaremos *A* para acortar el discurso, es susceptible, por su naturaleza, de tomar todos los valores contenidos entre algunos números, representados por a y b. Sea x, uno de esos valores; si se hacen una serie de experiencias para determinar *A*, la probabilidad de que el valor que saldrá en una de esas observaciones no exceda a x, variará, en general, de una experiencia a otra. Representémosla por $F_n x$ para la n^{th} experiencia. La probabilidad de que este valor sea exactamente x sería infinitamente pequeña, ya que el número de valores posibles es infinito; haciendo $\frac{F_n x}{dx} = f_n x$, sería la expresión $f_n x dx$.
>
> Asignamos X una función dada de x, la cual incrementa sin interrupción desde $x=a$ hasta $x=b$, y representamos a a_1 y b_1, como sus valores extremos. Para más generalidad, buscaremos la probabilidad de que si tomamos la suma de los valores de X, la cual resultará de un número s de sucesivas observaciones, ésta estará contenida entre ciertos límites. (Poisson, 1832, pp. 3-4).

En estos párrafos, son muchos los avances que se observan sobre el desarrollo de la idea de variable aleatoria. Define a X, que es una variable, como una *función creciente*, cosa que sólo podría ser posible si los valores x que toma la variable X son números reales. También veamos que lo que llama «cosa» (chose, en francés) *A* toma los valores de la realización de la experiencia. Es decir, *A* es de naturaleza azarosa, pero la función X es un objeto matemático. Se infiere que X toma todos los valores x («la cual se incrementa sin interrupción»). Estas dos «cosas» (*A* y X) se vinculan a través de los valores x. En *A*, x es el resultado de un fenómeno aleatorio, en X, x son números reales que se pueden ordenar. La definición de X le permite operar con ella (puesto que es una función en los reales) y con x (puesto que son números reales), *A* en cambio le permite vincularse propiamente con la experiencia aleatoria y por lo tanto con el contexto real, incluso se podría pensar que *A* es una noción primitiva de experimento aleatorio, sólo que con resultado numérico. Pero ninguna de las dos es aún el concepto de variable aleatoria actual. Lo notable es que Poisson ya indica el requisito de que X y x tengan características matemáticas, y, al mismo tiempo, estén vinculadas con un experimento aleatorio a través de *A*. También que, aun cuando *A* puede tomar valores numéricos, requiere la definición de *X*, ambas de diferente naturaleza.

Hay que observar que a pesar de que trabaja con valores determinados de *x*, la variable que maneja es continua «ya que el número de valores posibles es infinito» y «se incrementa sin interrupción». Notemos también que hay una insipiente *función de densidad* que está definida a través de algo muy próximo a la *función de distribución* y que ésta última tiene un papel muy importante, puesto que es la que le permite calcular la probabilidad de los valores *x*. La observación de que la probabilidad de un valor determinado *x* tienda a cero, indica que él sabía que la variable *y, dada por* $y = f_n x$, a la que se le llama *ley de probabilidad* (Poisson, 1832, p. 11) es un ente matemático que adquiere sentido en un contexto sólo a través de su integral y, por lo tanto, con su área bajo la curva. Lo curioso es que Poisson no relaciona estas funciones con *X* sino con los valores *x*. Es decir, *x* es la conexión no sólo entre *A* y *X*, sino también del contexto real con la probabilidad, lo cual también se observa en esta cita:

> La curva de la cual la ecuación es $y = f_n x$, representa la ley de probabilidad de los valores de *A* en la n^{th} observación, en el sentido que el elemento *ydx* de su área es la probabilidad del valor de *A* expresado por la abscisa correspondiente a *x*, y la misma área, la probabilidad de que este valor no sea mayor que *x*. (Poisson, 1832, p. 11).

Aquí ya se observa que la ley de probabilidad, $y = f_n x$, es introducida por Poisson de manera muy parecida a la *función de densidad* actual y establece su relación con la probabilidad a partir de la integral. Sin embargo, también hay una confusión, puesto que la relaciona directamente con los valores obtenidos a partir de *A* y no de *X*, por lo que en este término pareciera estar mezclada la función de densidad con la función de probabilidad. Cosa provocada por nombrar *x* tanto a los resultados obtenidos por *A*, como a los obtenidos por *X*. Quizá el motivo por el que Poisson no usó la notación funcional de $y = f(x)$, como ya la usaban los analistas (y muchos probabilistas) en sus trabajos de cálculo, era porque era importante hacer alusión a la *n-ésima* observación obtenida de *A*. Como se nota, el desarrollo de la noción de variable aleatoria está fuertemente vinculado con el de función de densidad y, por lo tanto, también con la introducción del análisis y de la variable continua en probabilidad.

Es indudable que Poisson obtiene resultados notables sobre el desarrollo del objeto matemático variable aleatoria y su distribución, así mismo sobre la variable continua y la idea de función de densidad. Sin embargo, sus trabajos tenían una intención práctica de aplicar el análisis matemático a distintas situaciones físicas y morales, no tenía un propósito de un análisis matemático por sí mismo (Fischer, 2011), así que el concepto de variable aleatoria no tuvo más desarrollo en él, ni tampoco se le dio la importancia que esta aportación tenía; no fue así, en cambio, con su trabajo sobre el teorema central del límite, como lo observa Alvarado (2007).

4.5. Las aportaciones de Pafnuti L. Tchebychef y sus discípulos

Tchebychef (1899 a, b y c), hace un uso más notorio de las variables aleatorias, llevándolas a un terreno matemático, sin nombrarlas aún como tales. Se interesa por varios problemas de probabilidad y de análisis matemático y la integración de los resultados obtenidos en cada uno de ellos es la que hace que dé un fuerte salto hacia una demostración más rígida, aunque incompleta, del teorema central del límite, pero también hacia un uso más explícito de las ideas de variable aleatoria y de función de densidad.

Trabaja con teoremas y demostraciones como el siguiente, actualmente conocido como la desigualdad de Tchebychef:

Teorema. Si se designa por *a, b, c,…* las esperanzas matemáticas de las cantidades *x, y, z, …*, y por $a_1, b_1, c_1, \ldots$ las esperanzas matemáticas de sus cuadrados $x^2, y^2, z^2, \ldots,$ la probabilidad de que la suma $x+y+z+\ldots$ esté dentro de los límites

$$a+b+c\ldots.+\alpha\sqrt{a_1+b_1+c_1+\ldots-a^2-b^2-c^2-\ldots.},$$

$$a+b+c\ldots.-\alpha\sqrt{a_1+b_1+c_1+\ldots-a^2-b^2-c^2-\ldots.},$$

será todavía más grande que $1-\frac{1}{\alpha^2}$, cualquiera que sea α.

Demostración: Sean: $x_1, x_2, x_3, \ldots, x_l$; $y_1, y_2, y_3, \ldots y_m$; $z_1, z_2, z_3, \ldots z_n, \ldots$ todos los valores imaginables de las cantidades *x, y, z, …* y sean $p_1, p_2, p_3, \ldots, p_l$; $q_1, q_2, q_3, \ldots r_m$; $r_1, r_2, r_3, \ldots r_n, \ldots$ las probabilidades respectivas de sus valores, o bien, las probabilidades de los supuestos $x = x_1, x_2, x_3, \ldots, x_l$;; $y = y_1, y_2, y_3, \ldots y_m$; $z = z_1, z_2, z_3, \ldots z_n, \ldots$

Conforme a esta notación, las esperanzas matemáticas de las magnitudes *x, y, z,…*; x^2, y^2, z^2, … pueden ser expresadas así $a = p_1x_1, p_2x_2, p_3x_3, \ldots, p_lx_l$; …

(Tchebychef, 1899b/1867, pp. 687-688).

Tchebychef nombra de dos distintas maneras indistintas a las variables aleatorias, les llama *cantidades* (quantités, en francés) y *magnitudes* (grandeurs) a lo largo de su trabajo. Denomina también *supuestos* (hypothèses) a la forma abreviada en que indica los valores que puede tomar cada una de las variables. Con respecto esta forma abreviada, observemos que primero usa una notación extendida y después introduce su notación abreviada. Cabe mencionar que en trabajos suyos sobre análisis matemático, Tchebychef sí usaba la palabra *variable*. En general Tchebychef denotaba con letras minúsculas a sus variables y distingue a las cantidades de las constantes por la notación con las últimas letras del alfabeto para las primeras y las primeras letras para las últimas, cosa que también hace en el cálculo de probabilidades.

La notación simbólica distinta para los valores de las variables (con subíndice) que para las variables (la misma letra, pero sin subíndice) indican que Tchebychef ya distinguía entre los valores que puede tomar una variable y la variable misma. Gracias a esa distinción, escribió de manera breve las esperanzas matemáticas[4] de las variables que estaba manejando, pero también marca una diferencia no explícita entre los sucesos y el valor de la variable. En algunos de sus trabajos, sin embargo, no hace la distinción en su notación entre valor de una variable y la variable misma, aunque conceptualmente sí se distinguen. Esto se nota en esta otra cita:

> Si las posibilidades del suceso E en μ pruebas consecutivas son p_1, p_2, p_3,…, p_μ y que su suma es S, la probabilidad de que en μ pruebas el número de repeticiones del suceso E suceda menos de m veces y más de n, superará, para m más grande que S+1 y para n más pequeño que S-1, el valor de la expresión:
>
> $$1-\frac{1}{2(m-S)}\sqrt{\frac{m(\mu-m)}{\mu}}\left(\frac{S}{m}\right)^m\left(\frac{\mu-S}{\mu-m}\right)^{\mu-m+1}-\frac{1}{2(S-n)}\sqrt{\frac{n(\mu-n)}{\mu}}\left(\frac{\mu-S}{\mu-n}\right)^{\mu-n}\left(\frac{S}{n}\right)^{n+1}$$
>
> (Tchebychef, 1899a/1846, p. 25).

En la cita anterior, notemos que, además, la letra E de Tchebychef, es distinta a la letra A de Poisson (apartado anterior), la E denota la generalidad de un *suceso* (es decir, cualquier suceso), que usa para vincular su notación matemática con el contexto de algún problema y para definir una variable, que no es E, sino las veces que aparece E. No usa notación matemática para denotar a esta variable, pero sí indica que puede tomar los valores de m y n.

Es posible que Tchebychef haya sido influenciado por la idea de variable aleatoria de Poisson, puesto que conocía su trabajo (lo cita en varias de sus publicaciones), pero no retoma su propuesta de variable como *función*. Sin embargo, su trabajo en análisis matemático, le proporcionan herramientas para una abstracción de ideas en probabilidad. Esto se nota en la demostración de su teorema central del límite:

[4] Fischer (2011) indica que la idea de Poisson de introducir la función X y la «cosa» A, fue tomada de un texto de Carl Friedrich Haubert sobre teoría de valores medios, escrito en 1830. Esto es algo que no pudimos comprobar, pero tenemos la certeza de que la introducción del uso de la variable aleatoria estuvo también fuertemente ligada a la necesidad de distinguir los valores de la variable aleatoria del suceso en el cálculo de las esperanzas matemáticas de las variables, como se observa en Tchebychef.

Para demostrar el teorema de la forma más general, tomaremos -∞ y +∞ como los límites entre los que son comprendidos todos los valores posibles de las cantidades: u_1, u_2, u_3,…. Habiendo designado por $\varphi_1(x)dx$, $\varphi_2(x)dx$, $\varphi_3(x)dx$,… las probabilidades de que los valores de las cantidades u_1, u_2, u_3,…. están comprendidas entre los limites infinitamente vecinos: x_1, $x + dx_1$, señalaremos que:

1) Las funciones $\varphi_1(x)$, $\varphi_2(x)$, $\varphi_3(x)$,… no pueden tomar valores negativos;
2) Las integrales $\int_{+\infty}^{-\infty}\varphi_1(u_1)du_1$, $\int_{+\infty}^{-\infty}\varphi_2(u_2)du_2$, $\int_{+\infty}^{-\infty}\varphi_3(u_3)du_3$,…, que representan las probabilidades de las cantidades u_1, u_2, u_3,…., estando los valores contenidos entre los límites -∞ y +∞, serán iguales a la unidad;
3) Las integrales $\int_{+\infty}^{-\infty}u_1\varphi_1(u_1)du_1$, $\int_{+\infty}^{-\infty}u_2\varphi_2(u_2)du_2$, $\int_{+\infty}^{-\infty}u_3\varphi_3(u_3)du_3$,…, que representan las esperanzas matemáticas de las cantidades u_1, u_2, u_3,…., de acuerdo a nuestro supuesto, serán nulas;
4) En general, todas las cantidades $a_i^{(\mu)}$ definida para la igualdad:

$$a_i^{(\mu)} = \int_{+\infty}^{-\infty}u_i^{\mu}\varphi_i(u_i)du_i,$$

que representan las esperanzas matemáticas de las potencias de las cantidades u_1, u_2, u_3,…., estarán, por hipótesis, contenidas entre límites finitos.

Por otro lado, designaremos por $f(x)dx$ a la probabilidad de la fracción:

$$\frac{u_1 + u_2 + u_3, \dots. + u_n}{\sqrt{n}}$$

que está comprendida entre los límites infinitamente vecinos: x_1, $x + dx_1$…

(Tchebychef, 1899c/1887, pp.483-484).

En esta cita, lo mismo que Poisson, Tchebychef distingue dos variables. La variable *x* tiene un carácter matemático y la cantidad *u*, a la que (se deduce) vincula con los resultados numéricos del experimento aleatorio. Sin embargo, la variable *x* tiene un grado de abstracción mayor que el que Poisson le denotó, puesto que lo usó para vincular *u* con sus resultados sobre los valores límites de las integrales. Esto se observa en la extensión del concepto cuando da lugar a su uso en la suma de las cantidades *u*, pero también en su definición de la función de densidad, $\varphi(x)$ ó $f(x)$, que, al agregar el diferencial, usa para calcular probabilidades comprendidas entre números infinitamente vecinos (limites infiniment voisines, en francés) y que caracteriza con las dos condiciones con las que actualmente la conocemos, es decir, es positiva en todo su dominio y su integral desde infinito hasta más infinito es uno. Tchebychef no usa las variables X y A de Poisson, ni se preocupa por ser explícito en la forma en que se vincula *u* con *x*, y también las usa tanto $\varphi(x)dx$, como $\varphi(u)du$, para denotar la probabilidad y la esperanza, pero ya muestra la necesidad de establecer una variable con carácter abstracto que concreta la definición de funciones de densidad. Observemos, así mismo, que no indica si las variables *x* ó *u* son discretas o continuas, pero por el trato matemático se deduce que son continuas.

Los trabajos de Tchebychef tuvieron algunas imprecisiones por lo que algunos de sus contemporáneos, los consideraron faltos de rigor. Una de las principales objeciones era su uso intuitivo de cantidades infinitamente pequeñas o infinitamente grandes. Frecuentemente pasó de resultados en variable discreta a variable continua sin más explicaciones y en algunas ocasiones manejó un dominio en todos los reales para una variable que había definido acotada entre ciertos valores. Tampoco especificó la condición de independencia de las variables en su demostración del teorema central del límite. Autores anteriores a él hacían explícita esa condición, por lo que se prestó a pensar que Tchebychef asumía la independencia como una característica de las variables aleatorias (Fischer, 2011). Sin embargo, Maistrov (1974) y Fischer (2011) lo consideran el primero en hacer uso de la noción de variable aleatoria con un carácter matemático. Además, la primacía de una versión más general del teorema central del límite, sigue recayendo sobre él, junto con uno de sus más cercanos discípulos, Andrei Markov, quien continuó su trabajo para establecer en 1908 una demostración más rigurosa basándose en su mismo método (Alvarado, 2007; Fischer, 2011).

Markov distinguió entre variables independientes y dependientes y extendió los teoremas límite de probabilidad a la suma de variables aleatorias conectadas en cadena. Por tanto, fue más específico en el concepto ideado de variable ideado por Tchebychef, lo que muestra que en probabilidad se comienzan a desligar del concepto de *cantidades variables* usado en cálculo. En la siguiente cita, se puede observar esta aportación al desarrollo de la variable aleatoria:

> Así, la independencia de las cantidades no constituye una condición necesaria para la existencia de la ley de los grandes números. (Markov, citado por Seneta, 1966, p. 255).

De acuerdo con Fischer (2011), Tchebychef fue el primero en formular un enunciado para el teorema central del límite involucrando una secuencia de variables aleatorias, ampliando el concepto de variable en probabilidad. También menciona que, antes de Tchebychef y sus discípulos, la propuesta laplaciana dio lugar a una época en la que los investigadores se concentraban no en estudiar los teoremas límite sino en encontrar aproximaciones al cálculo de probabilidades. Esto repercutió en que las aproximaciones a la función de densidad (en el caso continuo) o a la distribución de probabilidad (en el caso discreto) y las aproximaciones al cálculo de probabilidades a través de la integral definida fueran consideradas equivalentes. Esta diferenciación está vinculada no sólo con la distinción entre la variable discreta y continua, sino también con la definición de la función de densidad. El siguiente paso hacia la noción de la variable aleatoria fue la generalización de los resultados obtenidos en variable discreta hacia lo continuo y en una definición más precisa de la función de densidad. Cosa a la que contribuiría el otro discípulo notable de Tchebychef, Aleksandr M. Liapounoff (1857-1918).

El trabajo que Liapounoff hizo en pro de la demostración del teorema central del límite, hizo que éste fuera un objeto matemático autónomo, lo que marcó la llegada de una teoría moderna de la probabilidad. No obstante, Fischer (2011) especifica que su demostración tuvo algunas impresiones porque no justifica apropiadamente la valides de sus resultados en un contexto discreto, pero infinito, además de que evade el problema de continuidad de la variable. Esto lo podemos observar en este pasaje, en donde define las variables continuas como variables discretas que toman un número infinito de valores:

> La desigualdad (14) tiene lugar, siempre y cuando se disponga de un número τ positivo de manera que se cumpla que:
>
> $$L_\tau \leq k, \qquad A - \frac{4}{3}L^3\tau > 0$$
>
> Y si, considerando un caso en el que el conjunto de valores posibles contiene una cantidad infinita de números, se le mira como el caso límite del que hemos analizado, esta desigualdad subsistirá aún en el límite, todas las veces que las cantidades α_i, a_i, l_i (i = 1, 2, 3, …) tiendan hacia un límite determinado.
>
> Así pues, el resultado obtenido es totalmente general... (Liapounoff, 1900, p. 379).

Como se puede observar Liapounoff no trabajó propiamente con variables continuas, pero sí hizo posible la generalización[5] de sus resultados para variables que toman un número infinito de valores generadas por medio de un proceso límite de aquellas que toman un número finito de valores, y con ello evadió los problemas de intercambio del orden de integración, pero también el de continuidad. En todos los casos posibles, él usa sumas en lugar de integrales. Se desligó del método de momentos, usados por Tchebychef y Markov, y se apoyó en la definición de una variable auxiliar continua con distribución normal con media cero y varianza que converge cuando n tiende a infinito y que sí es posible integrar. Gracias a ello pudo emitir una demostración más rigurosa del teorema central del límite haciendo uso de condiciones más débiles (Fischer, 2011; Alvarado 2007):

[5] A esto es a lo que se refiere Fischer (2011) cuando habla de la vaguedad de los argumentos de Liapounoff para justificar la generalización de sus resultados.

> Sea x_1, x_2, x_3,… una serie indefinida de variables independientes reales, cuyos valores son debidos al azar. Designando por α_1, α_2, α_3,…, a_1, a_2, a_3,…, d_1, d_2, d_3,…, a las esperanzas matemáticas respectivas de x_1, x_2,…, $(x_1 - \alpha_1)^2$, $(x_2 - \alpha_2)^2$,…, $|x_1 - \alpha_1|^{2+\delta}$, $|x_2 - \alpha_2|^{2+\delta}$,…, donde δ es un número positivo cualquiera, habrá la proposición siguiente:
>
> «*Todas las veces que la existencia de los valores fijos de δ hagan que la siguiente relación:*
>
> $$\frac{(d_1 + d_2 + \ldots + d_n)^2}{(\alpha_1 + \alpha_2 + \ldots + \alpha_n)^{2+\delta}}$$
>
> *tienda a cero cuando n crece indefinidamente*[6]*, la probabilidad de la desigualdad*
>
> $$z_1 < \frac{x_1 - \alpha_1 + x_2 - \alpha_2 + \cdots + x_n - \alpha_n}{\sqrt{2(\alpha_1 + \alpha_2 + \cdots + \alpha_n)}} < z_2$$
>
> *donde z_1 y $z_2 > z_1$, son dos números dados cualquiera, y para n = ∞, tenderá hacia el límite de:*
>
> $$\frac{1}{\sqrt{\pi}} \int_{z_1}^{z_2} e^{-z^2} dz$$
>
> *Y lo hará uniformemente para todos los valores de z_1 y z_2*». (Liapounoff, 1901b, pp. 814-815).

Es posible reconstruir la demostración de Liapounoff desde la perspectiva de funciones características, lo que hubiese atendido el problema de la continuidad (como más tarde lo haría Lévy), pero Liapounoff no lo hizo, sin embargo, sí comenzó a lidiar más fuertemente con el problema de la variable continua en probabilidad y llegó a un resultado que favoreció el estudio de la demostración del teorema como un análisis puramente matemático sin ser un ejemplo de un método o de alguna aplicación. Por otro lado, él es el primero del que se tiene documentado el uso del término *variable* (variable, en francés) y que hace mención a su relación con el azar en su discurso y teoremas:

> El teorema del que aquí se trata, se relaciona con la conocida fórmula de Laplace y Poisson, que sirve para que la probabilidad aproximada de la de la suma de un gran número de variables independientes sometidas al azar, esté contenida entre ciertos límites. Según el teorema en cuestión, esta fórmula dará el límite hacia el que tiende la probabilidad cuando el número de variables aumente indefinidamente. (Liapounoff, 1901a, p. 126).

Aunque las traducciones a otros idiomas generalmente reducen su larga frase a «variables aleatorias independientes», Liapounoff no bautiza a la variable como *aleatoria*. Notemos, que el término «aleatorio» en algunos de sus traductores

[6] A este requisito se le conoce como *Condición de Liapounoff*.

contemporáneos, es un anacronismo. En ese entonces los términos usados para calificar a los fenómenos que nos ocupan eran *del azar* o *fortuitos*. Así, no había una distinción explícita entre los problemas filosóficos y matemáticos vinculados con la definición del azar, por tanto tampoco había una definición de aleatoriedad como un modelo matemático. Es posible que esta haya sido una de las razones por las que otros probabilistas de esa época no hicieran notoria la naturaleza «azarosa» o «fortuita» de la variable con la que estaban trabajando.

Liapounoff, además, también comenzó a trabajar con la función de probabilidad en una forma muy parecida a la actual, tomando como base lo ya definido por Tchebychef:

> Para evaluar esta probabilidad, a la que designaremos Π, debemos estar en condiciones de calcular, para cada una de las variables x_i, la probabilidad de estar contenida entre dos límites dados cualquiera.
> Para este efecto, introduciremos, por ejemplo, las funciones $\varphi_i(x)$, que experimentarán las probabilidades de la forma $x_i<x$, por lo tanto la cantidad $\varphi_i(v) - \varphi_i(u)$, para $v>u$, representa la probabilidad de la desigualdad $u\leq x_i<v$. (Liapounoff, 1900, p. 367).

Como se observa, la noción de función de distribución estaba gestada. Es notable el conocimiento que Liapounoff tenía de variables aleatorias: no sólo las distinguía y las enunciaba sino que usaba sus caracterizaciones más importantes como son su valor esperando y sus varianzas, así como su operatividad, convergencia y composición y manejó la función de distribución, de densidad y la distribución de probabilidad. Además, hizo patente que los problemas vinculados con la demostración del teorema central del límite eran propios de éste (y no aplicaciones de alguna rama de la matemática), impulsando que el teorema central del límite fuera un objeto matemático en sí mismo y que surgieran nuevos problemas propios de una teoría de probabilidades, entre ellos el de la naturaleza de la variable continua y su función de distribución. Sin embargo, la formalidad de la probabilidad y el surgimiento de la variable aleatoria como objeto matemático en sí mismo, aún tendrían que esperar algunos pocos años más para su pleno desarrollo durante el primer tercio del siglo XX con las aportaciones de Kolmogorov.

5. Inferencia estadística. Relación entre las variables estadística y aleatoria

En el siglo XIX y la primera mitad del XX, simultáneamente y en paralelo con el desarrollo teórico probabilístico de la escuela matemática rusa, la estadística sufrió un fuerte desarrollo tanto horizontal, como vertical. Horizontal porque se propagó a través de muchas disciplinas, sufriendo transformaciones internas en el proceso; vertical porque profundizó en la generación de métodos apropiados para el análisis de datos en situaciones bajo incertidumbre. Así, en unos pocos años y a partir de problemas aplicados, en Inglaterra se generaron las técnicas estadísticas inferenciales más comunes en la actualidad. En esos estudios se relacionaron explícitamente las variables estadísticas y las variables aleatorias, haciendo surgir métodos apropiados para estimar los valores de los parámetros que definen las familias de distribuciones de las variables aleatorias a partir de métodos estadísticos basados en el acopio y análisis de datos empíricos. Así mismo surgieron otras útiles herramientas, como los métodos de muestreo, la correlación y el análisis multivariado, que apoyarían la vinculación entre ambas variables.

Dos cosas fueron importantes en el desarrollo de la variable estadística, la primera fue que en su mayoría, la definición de una variable numérica en las ciencias sociales y biológicas surgía de manera natural a partir del contexto de un problema. Lo otro, es que los datos surgidos a partir de esas variables denotaban un comportamiento parecido al descrito por Gauss para la distribución de los errores de las medidas astronómicas. Esto hizo que familiarmente se comenzara a trabajar con la media e incluso con la varianza (como el «cuadrado del error medio») y desviaciones de la media (o «errores») como medidas de centralización y dispersión respectivamente, generando un lenguaje distinto al usado por los probabilistas:

> Más aún, el término Error Probable es absurdo cuando lo aplicamos a los temas ahora tratados, tales como estatura, color de los ojos, facultades artísticas o enfermedades. Por lo tanto, usualmente, hablaré de Desviación Probable. (Galton, 1889, p. 58).

También se fomentó el análisis del comportamiento de los datos a través de la distribución gaussiana. Los científicos de esta época pensaron que ese era el comportamiento «normal» de los datos en ciencias sociales y biológicas, así que, a mediados del siglo XIX, al comportamiento gaussiano de los datos se le comenzó a denominar *distribución normal*, sin que se tenga la certeza de quién o quiénes fueron los que la bautizaron así. En sus trabajos, Galton comenzó a nombraba así de una manera muy natural, a pesar de que en sus publicaciones contemporáneas, los probabilistas la seguían denominando *Ley de probabilidad*:

> … [los valores] no están derivados para nada de observaciones, sino de las bien conocidas tablas de la «Probabilidad Integral» en la forma en que los matemáticos lo hicieron fácilmente comprensible comparando las Tablas 4 a 8, inclusive. No hace falta recordarle al lector que la Ley del Error en la cual estos valores de la Normal están basados, fueron extraídos del uso de los astrónomos y otros quienes están preocupados con extremo cuidado de su medida. (Galton, 1889, p. 54).

En esa época, la distribución binomial ya había sido empleada en algunos estudios biológicos para comparar el resultado obtenido a partir de los datos y el resultado esperado bajo un modelo teórico binomial (ver apartado 3 de este capítulo). Esto ya denotaba un interés en hacer inferencias informales en donde se tenía la intención de relacionar las variables aleatoria y estadística. Así que la distribución normal, que tenía aplicaciones en muchos campos, también fue una herramienta que favoreció las inferencias en las que se pretendía conocer el comportamiento de la población, sólo que de manera un tanto informal.

El problema del análisis del comportamiento de datos bajo incertidumbre en diversas áreas atrajo el interés de diversos científicos, entre ellos, tres se destacan por sus contribuciones: Lambert Adolphe Jacques Quetelet, Francis Galton, Francis Ysidro Edgeworth y Karl Pearson. Ríos (1994) defiende que ellos, y sobre todo éste último, provocaron una revolución probabilista en diversas ramas de las ciencias, similar a la ya introducida por Gauss y Laplace en la astronomía, al establecer las bases para una metodología empírica en ciencias donde la experimentación no era posible de aplicar.

5.1. Las aportaciones de Lambert Adolphe Jacques Quetelet

Quetelet fue de los primeros en hacer intentos por hacer uso de los resultados probabilísticos en el análisis de datos de ciencias sociales. En su libro *Sobre el hombre y el desarrollo de sus facultades*, publicado en 1835, caracterizó al hombre medio a través de las medias de grandes colecciones de datos de diversas variables que supuso se distribuirían como una normal. También notó la regularidad con la que se reproducían ciertos fenómenos sociales y trató de encontrar sus causas a través del análisis estadístico de diversas variables, como el índice de criminalidad, de casamientos y de suicidios, sin conseguirlo. Algunos de sus métodos fueron tachados de simplistas, sin embargo, abrieron paso a un análisis de datos en ciencias sociales y al ajuste de datos empíricos a una distribución de probabilidad teórica, así como el uso de la media y la variación de la media en los análisis de datos.

Quetelet defendía la necesidad de grandes cantidades de datos para acercarse a la verdad (distribución teórica), cosa que de alguna manera es una aplicación de la ley de los grandes números y un intento por conciliar la variable estadística con la aleatoria. Promovió la recolección de datos cuidadosa con fines de análisis y también tuvo gran influencia en el desarrollo de comunidades estadísticas y en la conciliación de métodos y técnicas estadísticas usadas en la recolección y análisis de datos. En 1853, organizó el primer Congreso Internacional de Estadística en Bruselas. Junto con Galton, Pearson, Fisher y otros autores crearon la escuela biométrica, de donde se derivarían posteriormente sociedades tan importantes como el International Statistical Institute, fundado en 1885 durante el segundo congreso internacional de estadística con 81 miembros. La estadística, como un estudio cuantitativo que permite expresar con mayor precisión descripciones verbales, cobró cada vez mayor importancia.

5.2. Las aportaciones de Francis Galton

Fue uno de los pioneros en el desarrollo de la ciencia estadística en Inglaterra. Fundó la revista *Biometrika* en 1901 en colaboración con Pearson y Weldon para promover el estudio de la biometría, que define en la forma siguiente:

> El objeto de la biometría es proporcionar material lo suficientemente preciso para el descubrimiento de cambios incipientes en la evolución, que son demasiado pequeños para ser de otro modo aparentes. La distribución de cualquier atributo dado en cualquier especie, en cualquier momento ha de ser determinada, junto con su relación de influencias externas (Galton, 1901, p. 9).

Las contribuciones de Galton al desarrollo del concepto de variable estadística (y de la estadística en sí misma) fueron abundantes y surgieron principalmente a partir de sus estudios sobre la herencia. Retomó el uso de la estadística que había hecho Quetelet y se enfocó a contar y medir los atributos del ser humano y a caracterizar la forma en cómo se heredaban de una generación a otra. Él denominaba *distribuciones* a la forma en cómo una característica era poseída por los distintos miembros de una población. Se percató que lo importante en la ciencia de la herencia era considerar «grandes poblaciones más que individuos y que a éstas, debemos tratarlas como unidades» (Galton, 1889, p.35). Se preocupó por limitar las poblaciones en las que trabajaba, sin mezclar datos de una etnia con otra, por ejemplo, y en recolectar él mismo sus propios datos. También se dio cuenta que la información sobre un atributo (variable estadística) de cada población debía ser caracterizada por su *distribución* más que sólo por su media, que proporcionaba poca información. Presentó el resumen de los datos en forma gráfica (Figura 4.1), haciendo uso de histogramas (llamados *curva del gráfico del error* por Gauss y los estadísticos de esa época) y sugirió un análisis a través de gráficos muy próximos a las actuales ojivas de frecuencias relativas. En el primer gráfico de la Figura 4.1 está representada la fuerza física de jale de hombres (variable estadística) en una muestra muy grande. La altura de las barras corresponde al valor de la variable y la escala horizontal indica el *grado*[7] correspondiente al valor de la variable. El segundo gráfico representa un actual histograma, sólo que rotado.

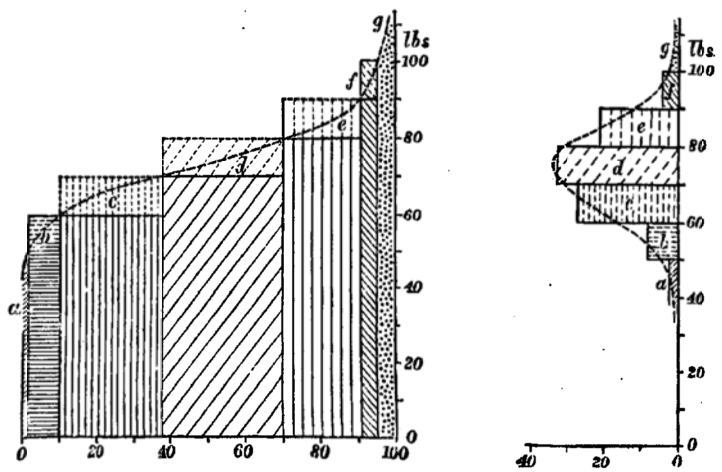


Figura 4.1. Gráficos de la distribución de una variable estadística (fuerza de jale en hombres) propuestos por Galton (Galton, 1889, p. 38)

En el gráfico de frecuencia acumulada Galton ajustaba los extremos de las barras a una curva suavizada a la que llamó *esquema de la distribución* de una característica (variable estadística) de una determinada población. Este gráfico es el precedente de la actual ojiva[8] y fue al que Galton le dio mayor importancia. A través de éste analizaba los diferentes grados, que le servían para caracterizar la distribución de la variable estadística y para calcular la proporción de individuos que se encontraban dentro de un intervalo de valores de la variable estadística. También le indicaba si la distribución era simétrica o asimétrica (haciendo una rotación del gráfico) y por lo tanto si la media y la mediana eran iguales. Galton preferiría la mediana a la media porque le daba más oportunidad de interpretación. Estudió y analizó sus características y la usaba para

[7] Galton empleó la palabra *grados*, refiriéndose a grados centesimales, para aludir a la frecuencia relativa acumulada dada en porcentaje. También empleó la palabra *percentil* tal cuál se usa actualmente: para denotar el valor de la variable correspondiente a un determinado porcentaje acumulado (es decir a un determinado *grado*).

[8] Encontramos algunas referencias de historiadores que atribuyen a Galton el uso de la palabra *ojiva* a este tipo de gráficos, sin embargo nosotros no encontramos la referencia directa de Galton. Lo cual no significa que esa atribución sea falsa, puesto que el número de publicaciones de Galton es muy amplia.

comparar poblaciones a partir de sus datos, junto con los percentiles. En general, los percentiles fueron la forma en que él comparaba una misma variable de diferentes poblaciones. La desviación en una distribución la calculaba a partir del promedio del valor absoluto de las diferencias de diferentes valores con la mediana, pero en sus últimos años (por influencia de Pearson), empleó la desviación estándar y la media[9].

STRENGTH OF PULL.
519 Males aged 23-26.
From measures made at the International Health Exhibition in 1884.

Strength of Pull.	No. of cases observed.	Percentages.	
		No. of cases observed.	Sums from beginning.
Under 50 lbs.	10	2	2
,, 60 ,,	42	8	10
,, 70 ,,	140	27	37
,, 80 ,,	168	33	70
,, 90 ,,	113	21	91
,, 100 ,,	22	4	95
Above 100 ,,	24	5	100
Total	519	100	

Figura 4.2. Tabla de frecuencias acumuladas usadas por Galton (Galton, 1889, p. 199).

Galton también trabajó con tablas de frecuencias (Figura 4.2) en las que definía los intervalos de la variable para poder elaborar esos gráficos porque le proporcionaban facilidad al ajustar sus datos a un esquema (ojiva) y hacer inferencias hacia la población. Sus deducciones sobre el comportamiento de los datos partieron de distribuciones discretas hasta llegar, en 1857, a una distribución continua muy parecida a la normal, cuyos valores, en sus análisis, agrupaba en intervalos de la forma en que se observa en la Figura 4.2. Así mismo sus análisis sobre los valores de las variables continuas los hacía sobre intervalos. Por ejemplo, menciona que la mediana (M) tenía la propiedad de que si se consideraba una medida N como una medida de la población y u una cantidad pequeña, entonces «el número de medidas que caerán entre $(N - \frac{1}{2}u)$ y $(N + \frac{1}{2}u)$ es más grande cuando N=M» (Galton, 1889, p.41).

Galton también propuso una estandarización de los valores de las variables de modo que varias variables correspondientes a la misma población pudieran ser comparables entre sí. Esta estandarización también la ocupó para verificar que los datos de su muestra apoyaban la tesis de que la población se comportaba de acuerdo con la ley

[9] En su artículo: Galton, F. (1907). *Grades and deviates*. Biometrika, 5: 400-408.

de probabilidades del error (la distribución N(0,1)), basándose en la semejanza de los percentiles 5 a 95 de ambas distribuciones. Llegó a la siguiente conclusión:

> Estoy satisfecho de reivindicar que la curva normal es una justa representación promedio de las curvas observadas en el 90% central de sus datos; esto es, para tantos de ellos como caigan entre los grados de 5º y 95º. En particular, estoy de acuerdo en que para la curva de estatura, la curva normal es muy aceptable y constituye un pilar de mi investigación en la ley de la herencia natural. (Galton, 1889, p. 57).

Aunque, también recalcó, sin profundizar demasiado, que la distribución de las cualidades y facultades humanas solo podía ser aproximadamente normal. De modo que, afirmó, sus «resultados también serían sólo aproximaciones a la verdad» (Galton, 1889, p. 58), sin que pudiera cuantificar qué tanto podía aproximarse a esa verdad.

A pesar de sus contribuciones, para Galton la estadística siempre fue una herramienta. Sus estudios sobre la relación entre la altura de hijos y sus padres, lo condujeron al concepto de *regresión,* que posteriormente desarrollaría más ampliamente Karl Pearson y Francis Edgeworth. A partir del registro de datos de muchos grupos familiares, Galton concluyó que los hijos de padres muy altos, heredaban parte de esa altura, pero también revelaban una tendencia a regresar a la mediana[10]. Galton generalizó esta tendencia bajo la *ley de la regresión universal*, que actualmente se conoce como «regresión a la media». En su autobiografía, Galton describe la correlación términos cotidianos como:

> Parecía evidente por observación, y se verificó en ese estudio, que existía un «índice de correlación», o sea, una fracción, ahora llamada simplemente *r*, que relaciona con una buena aproximación cada valor de desviación [de la mediana] por parte del sujeto con el *promedio* de todas las desviaciones asociadas a su pariente. Entonces la aproximación de cualquier parentesco específico puede ser hallada y expresada con un término único. Si un individuo particular se desvía mucho, el *promedio* de las desviaciones de todos sus hermanos será una fracción definida de esa cantidad; del mismo modo que los hijos, padres, primos, etc. Cuando no hay relación alguna, *r* se vuelve igual a 0; cuando es tan cercana que sujeto y pariente poseen idéntico valor, entonces, *r* = 1. Por lo tanto, el valor de *r* en todos los casos toma valores entre los límites extremos de 0 y 1. (Galton, 1908, p. 303).

[10] Como la distribución de las características estudiadas por Galton era normal, la mediana era igual o cercana a la media, pero Galton generalmente hacía ese cálculo basándose en el percentil 50. En sus últimos años, tomó las contribuciones de Pearson y Edgeworth e hizo uso de la media y la desviación estándar.

Así mismo, puesto que se esperaría que las dos generaciones tuvieran un comportamiento estadístico similar, deduce que el promedio de las desviaciones tendría que ser igual en ambos grupos, de lo que se desprende la relación entre la variación entre grupos, la variación dentro de la población y el coeficiente de regresión. Lo que hace a Galton pionero en el estudio de la estimación de las componentes de la varianza atribuibles a causas identificables, y también sentó las bases del estudio de la variable estadística y su distribución a través del estudio de la normal. En sus palabras, esta distribución que fue la que le permitió visualizar un «orden en el aparente caos» e identificar la existencia de una variación «normal» debida al comportamiento natural de los datos. Inventó una maquina llamada *quincunx*[11], que le permitió ilustrar la distribución normal y la diferencia entre la dispersión provocada por causas identificables y la debida a la aleatoriedad. Así mismo en sus publicaciones aparecen las primeras tablas de la normal estándar realizadas por W. F. Sheppard a petición de Galton. También, su trabajo con la normal bivariada (para evaluar el comportamiento de los errores de dos características de una misma población que varían conjuntamente), inició el estudio del análisis multivariante actual. En sus trabajos recalcó la importancia del estudio de la variable estadística a través del análisis de su distribución completa y sentó bases para analizarla. Podemos darnos una idea de la importancia de los estudios de Galton en el desarrollo sucesivo de la variable estadística los primeros renglones de su conclusión a su estudio sobre la herencia:

> La investigación aquí concluida está basada en el hecho de que las características de cualquier población que está en armonía con el ambiente, pueden permanecer estadísticamente idénticas durante generaciones sucesivas. (Galton, 1889, p. 192).

5.3. Las aportaciones de Karl Pearson

En la última década del siglo XVIII, los métodos estadísticos se vieron fuertemente influenciados por los avances que los probabilistas lograron en la formalización de la matemática del azar y uno de los más notables contribuyentes a este crecimiento de la estadística fue Karl Pearson. Sus notables trabajos en ese sentido lo llevaron al establecimiento de los fundamentos filosóficos y matemáticos de la estadística matemática como ciencia (Magnello, 2009). El trabajo de Pearson acrecentó la trascendencia del estudio de los datos en los estudios teóricos y también sustentó la vinculación de los variables estadísticas y aleatorias a través del uso de herramientas

[11] El *quincunx* también es conocido como *aparato de Galton*. Con variaciones, fue usado por Piaget en sus experimentos sobre la normal (ver Capítulo 2, sección 2.2).

probabilísticas y estadísticas, cosa que también definió uno de los derroteros principales de la estadística matemática.

Dentro de los grandes aciertos de Pearson no solo estuvo su formación matemática, que le permitió comprender los modernos resultados que se estaban gestando en probabilidad y matemáticas, sino también que se interesó por aplicarlos a diversas ciencias. A diferencia de algunos probabilistas contemporáneos a él, Pearson basó sus novedosos métodos y propuestas sobre el estudio del azar en el análisis de grandes bases de datos y en preguntas que surgían a partir de ellos. Esto le permitió introducir procedimientos estadísticos para el manejo de los datos, explorar comportamientos de las distribuciones empíricas distintas a la curva normal y proveerse de un conjunto de herramientas estadísticas matemáticas para examinar las gráficas[12] de las distribuciones empíricas basándose en el método de momentos, como la variación (a través de la desviación estándar), el sesgo y la curtosis (Magnello, 2009).

A través de Weldon y por los trabajos de Galton, Pearson se interesó por el desarrollo de métodos matemáticos para el estudio de los procesos de herencia y evolución. Ente 1893 y 1912 escribió diversos artículos con el primer título de *Mathematical Contributions to the Theory of Evolution* en donde describió sus contribuciones al análisis de regresión, el coeficiente de correlación y las pruebas para la determinación de la bondad de ajuste de las curvas de las frecuencias observadas (Chi cuadrada), conceptos que desarrolló ampliamente, contribuyendo así al estudio de la variable estadística. Estudió diversas bases de datos provistas por Galton o Weldon para establecer una relación entre las distribuciones de los datos y las distribuciones teóricas, haciendo explícita la diferencia entre parámetros (de las distribuciones de las variables aleatorias) y los resúmenes estadísticos (de las variables estadísticas obtenidas a partir de los datos). También retomó el método de momentos[13] ideando por Tchebychef y propició la estimación de los momentos poblacionales a partir de los momentos muestrales. Así por ejemplo reconoció la importancia de la media y de la desviación estándar[14] para caracterizar las distribuciones de una variable estadística o aleatoria y las empleó para una mejor caracterización del comportamiento de las variables aleatorias a partir de las estadísticas.

[12] Pearson fue el que le dio el nombre de *histograma* al gráfico de los datos más empleado actualmente.
[13] Pearson fue el que tomó el nombre de *momento* de la física y lo empleó en un sentido estadístico cuando calculó la media de la variable estadística como el centro de gravedad de la curva de las frecuencias observadas y se dio cuenta que podría calcular otros momentos (Pearson, 1894 y 1902).
[14] Fue Karl Pearson también quien definió la *desviación estándar* como la raíz cuadrada del segundo momento, y a partir del concepto de error medio de Gauss, y la simbolizó con la letra σ (Pearson, 1894).También distinguió simbólicamente la desviación estándar muestral con la letra *s*.

Aplicó el método de mínimos cuadrados, ideado por Gauss y retomado por Markov, y encontró mejores estimaciones para los conceptos propuestos por Galton, pero también generalizó y les dio mayor sistematicidad matemática a los métodos estadísticos de la regresión múltiple. Dedujo la fórmula del coeficiente de correlación (que actualmente lleva su nombre) haciendo uso de la media y de la desviación estándar de los datos (a diferencia de Galton que usaba mediana y el promedio del valor absoluto de las desviaciones). Pearson (1902) propuso un método sistemático (basado en el método de momentos) para representar una serie de observaciones a través de una fórmula consistente y accesible, en el que no se limitó al ajuste de los datos a una línea recta. Creía en la correlación como el instrumento adecuado para convertir la psicología, la antropología y la sociología en ciencias tan respetadas como la física y la química (Ríos, 1994).

Sus estudios sobre la forma en qué se comportaban diversos rasgos humanos hereditarios a partir de bases de datos muy grandes, también lo condujeron a la conclusión de que no todos los datos se comportaban de acuerdo a una distribución normal (Pearson, 1893). «Anormal» fue el término que empleo para otro tipo de comportamiento y de este interés surgieron números estudios sobre las distribuciones asimétricas o simétricas que no podían ser consideradas normales (principalmente aquellas vinculadas con la familia Gamma). También abrió los métodos de correlación a datos discretos.

Su mayor contribución a la estadística matemática fueron los resultados surgidos por su preocupación en evaluar la validez de los modelos teóricos encontrados con su metodología a partir de los datos[15]. Propuso el uso de la distribución χ^2 y la probabilidad del cuadrado medio del error *P* para valorar la bondad del ajuste de un modelo teórico sobre los datos, analizando el sistema de errores observados. En su primer trabajo sobre esto, él buscaba un *criterio* que le indicara la certeza del modelo teórico propuesto, pero en sus posteriores incursiones en el tema comenzó a denominarla *prueba*. Karl Pearson ha sido considerado el fundador de la inferencia estadística, ya que desde su primer artículo sobre la prueba χ^2, estaba el germen de lo que posteriormente sería la lógica de las pruebas de hipótesis (Pearson y Kendall, 1970). El desarrollo de esta prueba también indicó que el uso de los métodos estadísticos no dependía de la normalidad de la distribución de los datos para poder hacer válidos sus resultados, lo que también amplió el ámbito de estudio de la estadística a la búsqueda de distribuciones de más diversas

[15] Algunos autores adjudican al geodesta alemán F. R. Helmert haber usado la distribución χ^2 antes que Karl Pearson, en 1876, sin embargo la contribución de Pearson va más allá de un solo trabajo sobre esta distribución. Publicó innumerables artículos sobre la prueba de bondad de ajuste, lo que le permitió mejorarla y aplicarla en diferentes situaciones, así como hacer tablas de χ^2.

variables aleatorias a partir de las estadísticas[16]. La prueba χ^2 posteriormente fue adaptada por el mismo Pearson para probar la asociación de dos variables estadísticas discretas colocadas en un tabla de contingencia.

A partir de los estudios de Pearson, la estadística tomó un rumbo diferente. El uso de modelos probabilísticos denotó un cambio en la forma y la nomenclatura de los artículos sobre estadística. Pero él no sólo estuvo interesado en elevar el prestigio de la estadística como ciencia, sino también como una ciencia útil para el análisis de los datos de cualquier otra ciencia. En palabras de Magnello (2009):

> Su metodología estadística no solo transformó la visión que se tenía sobre la naturaleza de la estadística sino también le dio al mundo un conjunto de herramientas cuantitativas útiles para orientar la investigación, junto con un lenguaje científico universal que estandarizó la escritura científica del siglo XX (p 25).

5.4. Las aportaciones de Gosset, Fisher, Neyman y E. Pearson

En esta época, siguiendo el desarrollo teórico del teorema central del límite y los desarrollos teóricos de Pearson y otros estadísticos, surgió, el concepto de distribución muestral o distribución del estadístico, que es un caso particular de variable aleatoria, y su aplicación en la estimación e inferencia estadísticas. El método de determinar la probabilidad de que la media de la población esté dentro de una distancia de la media suponiendo normalidad y con una desviación estándar igual a $\frac{s}{\sqrt{n}}$, se hizo muy popular. Sin embargo los fundamentos de la estimación y la inferencia estadística no estaban del todo dados. Existían, además, problemas para su aplicación cuando las muestras no eran muy grandes. Así, en contraposición a la investigación exhaustiva (a través de grandes cantidades de datos), se comenzó a defender el muestreo aleatorio y el viejo problema estudiado por Jacob Bernoulli sobre la cantidad necesaria de datos para tener estimaciones confiables volvió a surgir. En 1895, en una reunión del Instituto Internacional de Estadística, fue presentada la idea de *representatividad* de un conjunto de datos tomados de una población. Este ambiente favoreció la emergencia de métodos de análisis de las variables estadísticas en los que se tomó en cuenta el tamaño de la muestra y la forma en la que se recolectaban los datos, así como la generación de modelos de distribuciones útiles para pequeñas muestras.

Bajo el pseudónimo de *Student* (1908), William Gosset, publicó un artículo titulado *El error probable de una media*, en donde propuso la estimación de la

[16] El mérito de concebir a la distribución χ^2 como una familia de funciones relacionadas con la normal, fue de Ronald Fisher.

distribución de variables estadísticas de las que sólo se podían obtener pocos datos a través de una distribución que, en los grandes números, tendía a la N (0,1) y a la que posteriormente se le denominaría *t de Student* y en donde las probabilidades son medidas desde la media de la población en términos de la desviación estándar de la muestra. Gosset fue pionero en los análisis de bases de datos que no fueran muy grandes.

Por otro lado, desde sus primeras publicaciones[17], Ronald Fisher se vio atraído por la búsqueda de un criterio absoluto (la futura *máxima verosimilitud*) para conocer los rasgos de una distribución teórica a partir de un conjunto de datos. Se interesó en los trabajos de Pearson y los usó para sustentar y ampliar la teoría mendeliana sobre la herencia. Esto supuso una vinculación entre las dos líneas teóricas sobre la herencia de principios del siglo XX: la mendeliana y la darwiniana e impulsó nuevos métodos en biometría y estadística. En Fisher (1918), sentó las bases de la genética biométrica e introdujo la descomposición de la varianza en los modelos lineales debida a diferentes fuentes de variabilidad, lo que lo llevó a proponer métodos de análisis mejores que los de correlación.

En 1919 Fisher fue contratado para analizar una serie de datos acumulados en la Estación Experimental de Rothamsted. Ahí tuvo acceso a una serie de estudios agronómicos a partir de los cuáles desarrolló el análisis de la varianza como un nuevo enfoque del diseño de experimentos, al tiempo que, de manera paralela, continuó sus estudios sobre estadística teórica y genética dentro de la biometría. Esto le permitió generar conocimiento estadístico para sus propuestas de diseño de experimentos, pero también se ocupó de teorizar sobre la estimación y la inferencia estadísticas. En 1925 describió, en su artículo *Theory of statistical estimation*[18], una nueva teoría de estimación sustentada en la *máxima verosimilitud* como una forma eficiente de extraer información de los datos basada en los trabajos de Daniel Bernoulli. Y en Fisher (1930), siempre preocupado por mantener una visión frecuencial de la probabilidad, desarrolló un método inferencial basado en la probabilidad «fiducial»[19] con el que propuso determinar la frecuencia con que el valor verdadero de un parámetro toma un valor determinado a partir de los datos observados como una medida de la certeza de la aproximación del estadístico (medida de muestra) al parámetro (medida de población). La inferencia fiducial trata de evitar las probabilidades a priori de las hipótesis (como la

[17] On an absolute criterion for fitting frequency curves, publicado en *Messenger of mathematics*, 41: 155-160 (1912).
[18] Publicado en *Proceedings of the Cambridge Philosophical Society*, 22: 700-725 (1925).
[19] La palabra «fiducial» propuesta por Fisher, que se deriva de fiduciary y cuyas raíces latinas (fiduciaries o fides) significan *en confianza* o *fe*, se refiere a una relación ética de confianza entre dos o más partes. Otra acepción de la palabra está relacionada con un punto de referencia o medida.

estadística frecuencial), pero produce probabilidades a posteriori de las hipótesis, dados los datos (como la inferencia bayesiana). Rouanet (1998) indica que en algunos casos las distribuciones fiduciales de Fisher coinciden con las distribuciones bayesianas a posteriori, de modo que se podría considerar que Fisher fue un estadístico bayesiano adelantado a su tiempo.

Fisher también fue partidario de organizar experimentos y recoger datos para poner a prueba las hipótesis sobre los parámetros de las variables aleatorias (Fisher, 1922). Sus principales metodologías desarrolladas en esta rama las publicó en *Statistical Methods for Research Workers* (1925), *The Arrangement of Field Experiments* (1926) y, más completamente, en *The Design of Experiments* (1935)[20]. En ellos propone métodos para poner a prueba de una forma rigurosa el efecto de diferentes factores y tratamientos, aislados o combinados entre si bajo condiciones de control experimental. Enfatizó en la importancia de aleatorizar los tratamientos o bien tomar elementos aleatoriamente para asegurar la validez de sus pruebas, puesto que en el terreno de la agronomía era difícil asegurar la independencia de las observaciones. En su último trabajo citado, ya había descrito y desarrollado su teoría inferencial en la que sustentaba el análisis y los diseños de sus experimentos. Esto le permitió analizar la distinción entre estadístico (medida de muestra) y parámetro (medida de población), definió los conceptos de hipótesis nula, los niveles de significación, resultado significativo y las ideas fundamentales del test de significación. Es decir, hizo explicita la diferencia entre variable estadística y aleatoria, las puso en relación mediante sus procesos de estimación y se preocupó por evaluar la significancia con la que estaban relacionadas. Para Fischer la prueba de hipótesis era un procedimiento mediante el cual el investigador podría formarse una opinión sobre alguna característica de la población, o parámetro.

Sin embargo, la discusión sobre la teoría de la estimación e inferencia no terminaría ahí. En 1926, Jerzy Neyman (un joven matemático de la universidad de Varsovia) y Egon S. Pearson (hijo de Karl Pearson) comenzaron a ocuparse de los fundamentos matemáticos subyacentes en los trabajos estadísticos de entonces. Su interés se dirigió a encontrar las bases lógicas que permitieran unificar sistemáticamente las pruebas estadísticas propuestas por Gosset y Fisher. La propuesta inicial provino del mismo Gosset en una carta dirigida a Neyman en donde le sugería la conveniencia de inclinarse por rechazar una hipótesis bajo la cual la muestra observada es muy improbable, «si existiera una hipótesis alternativa que explicara la ocurrencia de la muestra con una probabilidad razonable» (Pearson, E.S. citado por Lehmann, 1994, p.

[20] *The Design of Experiments*, 8 editions, 1935 /37 /42 /47 /49 /51 /60 /66, Edinburgh: Oliver & Boyd.

400). En 1928, publicaron sus primeros resultados en un artículo titulado *On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference*. Este artículo ya contenía los principales conceptos de su teoría de hipótesis, tales como los dos tipos de errores, la idea de potencia de una prueba, la distinción entre hipótesis simple y compuesta, pero estaba enteramente basada en la razón de la verosimilitud de las hipótesis, cosa que con lo que Neyman no se sentía satisfecho. En 1930 idearon un nuevo enfoque maximizando la potencia de la prueba suponiendo cierta la hipótesis y con un valor preasignado (el *nivel* de la prueba) para la probabilidad de rechazo. En 1933 publicaron sus resultados en el artículo *On the Problem of the Most Efficient Tests of Statistical Hypotheses*[21], en el que describieron el fundamento de las pruebas de hipótesis más usadas actualmente, conocido como el Lema de Neyman-Pearson. De acuerdo con Lehmann (1994), ellos lograron plantear el problema de encontrar la mejor prueba como un problema matemático lógicamente convincente y claramente formulado, que uno podía, entonces, proceder a resolver (p. 401).

Neyman también abordó el problema de la teoría de estimación. Él era partidario de la estimación de un parámetro a través de un intervalo, no estaba de acuerdo en los enfoques en los que el parámetro era estimado por un único número, pero tampoco congeniaba con la idea bayesiana de que el parámetro era aleatorio (Lehmann, 1994). En 1933 publicó por primera vez una versión de su teoría con el título de *On the two different aspects of the representative method*[22] y con ella inició la teoría moderna de intervalos de confianza. Neyman descompuso el problema de la estimación en una parte práctica y otra teórica. La parte práctica se refiere al cálculo que se necesita realizar con los datos obtenidos durante el muestreo para *estimar* el parámetro. La parte teórica (precisión de dicha estimación) la resuelve en forma probabilística utilizando el teorema central del límite para proponer como solución el intervalo de confianza, como un intervalo de valores aunado a un coeficiente de confianza que mide la probabilidad frecuencial de la proporción de intervalos calculados en muestras sucesivas del mismo tamaño contengan al parámetro (Olivo, 2008). Esta formulación estableció una equivalencia entre intervalos de confianza y familias de pruebas de hipótesis e hizo posible que la prueba de hipótesis formulada en 1933 fuera considerada teoría de estimación (Lehmann, 1994).

[21] Publicado en: *Philosophical Transactions of the Royal Society of London*. Series A, 231, (1933), pp. 289-337.

[22] Aparecido en el volumen 97 del *Journal of the Royal Statistical Society*, en 1934**,** 558-625.

5.5. La variable estadística en la actualidad

Así, a principios del siglo XX, las herramientas estadísticas para el tratamiento de los datos y, por tanto, de la variable estadística, se multiplicaron propiciando que el análisis de datos fuera más rico y tuviera mayores fundamentos. Esto condujo, progresivamente, al desarrollo de las teorías estadísticas con las que convivimos actualmente y a relacionar de manera explícita las variables estadísticas (que se referían a las muestras tomadas de poblaciones) con las variables aleatorias, que definen las poblaciones cuyas características se quiere estimar. Así mismo, el concepto de distribución muestral o distribución del estadístico adquiere una mayor importancia dentro de la estadística matemática y, sobre todo en el caso de la distribución de la media muestral, fomentó el interés por estudiar las distribuciones de las operaciones con variables aleatorias y ofreció un sustento para el crecimiento aún mayor de la inferencia estadística y de la estadística.

Desde el inicio de nuestro análisis histórico hemos denominado a la *variable estadística* como tal, sin embargo, es menester mencionar que, históricamente y a pesar de que en su mayoría no se le confundía con la aleatoria, en sus inicios se le denominaba simplemente con las palabras de *cualidad*, *característica* o *atributo* (el término *magnitud* no era tan socorrido por los contextos en los que comenzó a ser aplicado). A principios del siglo XX se le comenzó a denominar *variable*, sin el calificativo de *estadística*.

La palabra *estadística*, se comenzó a emplear por Godofredo Achenwall en 1760, quien lo tomó del italiano *statista* (hombre de estado) para denominar al análisis de los datos numéricos que en ese entonces comenzaba a surgir. Achenwall creía que éste análisis sería un gran aliado de un gobernante eficaz. A finales del siglo XVIII, John Sinclair, emplearía el término para indicar la generación de información interna que permitiría encontrar falacias y proponer mejoras en el país. A comienzos del siglo XIX, la palabra estadística adoptó un significado más generalizado hacia la recolección y clasificación de datos numéricos de las más variadas situaciones, y posteriormente, a principios del siglo XX, surgieron paulatinamente los fundamentos de la ciencia que llevaría ese nombre. Sin embargo la vinculación entre el adjetivo *estadística* y el sustantivo *variable*, no se dio sino hasta tiempos más recientes[23]. El surgimiento de la estadística como ciencia, la posibilidad de manejar grandes cantidades de datos a través de las computadoras y el fortalecimiento de la teoría formal de la probabilidad dio lugar

[23] La referencia más antigua que encontramos a la definición de variable estadística está en Ríos (1967), aunque ello no implica que no haya sido definida o usada anteriormente. En cambio, encontramos grandes cantidades de libros que la definen, sobre todo de las últimas décadas.

a la necesidad de distinguir palpable y formalmente aquellas variables que estaban definidas teóricamente (las *aleatorias*) de aquellas que se vinculaban directamente con el manejo de los datos (las *estadísticas*).

6. Formalización matemática de la variable aleatoria

La creciente aplicación de los resultados probabilísticos y estadísticos en otras ciencias, como la física, durante el siglo XIX, aunado a definiciones probabilísticas apegadas a la experiencia y en muchas ocasiones, intuitivas, hicieron surgir la necesidad de una mayor precisión conceptual y lógica de la teoría de probabilidad. En particular en el campo que nos interesa, las concepciones de variable aleatoria que surgieron en los trabajos tanto de Poisson como de Tchebychef estaban demasiado apegadas al contexto del problema, como para que la variable aleatoria pudiera ser considerada un objeto matemático.

De acuerdo con Fischer (2011), la consolidación de la teoría de probabilidad fue fuertemente motivada por el interés creciente, principalmente de Feller y Lévy, en la demostración definitiva del teorema fundamental del límite después de la primera demostración dada por Liapounoff, puesto que esto impulsó el apego por el rigor analítico y el esfuerzo por la búsqueda de las condiciones lo más débiles posibles. Cuestiones, que, finalmente, fomentaron la disociación de la teoría de probabilidad de sus aplicaciones y su emancipación a convertirse en una disciplina matemáticamente autónoma. Mientras el problema con la distribución de la suma de variables aleatorias independientes podía ser abordado sobre la base de conceptos elementales de análisis real, la suma de variables aleatorias no independientes, requería el uso de conceptos más sofisticados de teoría de la medida. La formalización tanto de la probabilidad, como de la variable aleatoria, tuvo que esperar hasta el desarrollo de las teorías de conjuntos y de la medida, que surgieron a principios del XX, debidos principalmente a Georg Cantor, Emile Borel y Henri Lebesgue.

Durante el primer tercio del siglo XX, la axiomatización de la teoría de probabilidad tomó forma y fue ahí donde la variable aleatoria adquirió identidad propia y se desenvolvió como una de las bases de la creciente teoría. En este apartado se describe el último escalón de la emergencia de la variable aleatoria como objeto matemático formal, a partir de una necesidad y como una consecuencia de la axiomatización de la teoría de la probabilidad.

6.1. Aportaciones de Kolmogorov y Fréchet

Antes de que Kolmogorov publicara su obra *Fundations of the Theory of Probability*, ya había habido otros intentos por establecer una axiomática de la probabilidad. Richard von Mises en 1919 esbozó una teoría sustentada en la idea de *colectivo* definido a través de dos propiedades de la frecuencia relativa: debe poseer un límite y este límite debe ser constante en todas las sucesiones parciales que pueden seleccionarse de la sucesión original. El mismo Kolmogorov reconocía como el primer sistema axiomático al desarrollado por Sergei Natanovich Bernstein, que se sustentaba en una estructura algebraica de sucesos basándose en la noción de comparación cualitativa de sucesos (aleatorios) de acuerdo a sus (más grandes o más pequeñas) probabilidades (Maistrov, 1974). Andrei Kolmogorov, sin embargo, consolidó la axiomatización de la teoría de la probabilidad vigente hasta nuestros días con la incorporación de las teorías de la medida y la integración.

Los *Fundations of the Theory of Probability* (Kolmogorov, 1993/1956) pueden considerarse la culminación natural del desarrollo experimentado por la teoría de la medida entre 1900 y 1930 a partir de los aportes de muchos matemáticos dedicados a ello, entre los que destaca, desde nuestro enfoque, Maurice Fréchet. A él se le debe el concepto de espacio métrico, pero también la desvinculación de los elementos geométricos (presentes en primer plano en las propuestas de Lebesgue) de las teorías de la medida y la integración. Esta desvinculación fue la condición necesaria para establecer el sustento de la teoría de probabilidad sobre las analogías entre las variables aleatorias independientes y las correspondientes propiedades de las funciones ortogonales. Cosa que le permitió a Fréchet formular la primera vinculación de la herramienta matemática con la noción de variable aleatoria:

> Las variables aleatorias, desde un punto de vista matemático, representan meramente funciones medibles con respecto a $P(A)$, mientras sus esperanzas matemáticas son integrales abstractas de Lebesgue (Esta analogía fue explicada completamente por primera vez en el trabajo de Fréchet) (Kolmogorov, 1956/1993, p. 8).

Kolmogorov define y desarrolla formalmente la definición de variable aleatoria en los capítulos III y IV, sin embargo, discute su significado dos veces durante el capítulo I, a propósito de la inserción de otros distintos conceptos claves en su teoría. La primera vez, cita de arriba, la vincula con el concepto de *independencia*:

> Históricamente, la independencia de los experimentos y la variable aleatoria representan el mismo concepto matemático que ha dado a la teoría de probabilidad su sello

> particular. Los trabajos clásicos de Laplace, Poisson, Tchebychef, Markov, Liapounoff, Mises y Bernstein realmente estuvieron dedicados a la investigación de series de variables aleatorias independientes. Aunque algunos de ellos fallaron en asumir independencia, revelaron la necesidad de introducir condiciones análogas más débiles, para obtener resultados suficientemente significantes (Kolmogorov, 1956/1993, pp. 8-9).

Así, para Kolmogorov, la variable aleatoria independiente es el germen de un campo de problemas en la teoría de probabilidad y por lo tanto era una condición hacerla explícita para poder desarrollar la teoría de probabilidad. No es de extrañarse, entonces, que la segunda vez que menciona a la variable aleatoria durante el Capítulo I sea a propósito de la probabilidad condicional y las cadenas de Markov, cosa que, de alguna forma, coincide con lo ocurrido históricamente con las primeras definiciones de variable aleatoria dadas por Poisson y Tchebychef y el sucesivo desarrollo de las cadenas de Markov (apartado 4.5 de este capítulo):

> Sea $\mathfrak{A}$ una descomposición del conjunto fundamental E: $E = A_1 + A_2 + \ldots + A_n$, y *x* una función real de los sucesos elementales ξ, los cuales para cada conjunto A_q es igual a una correspondiente constante a_q. *x* es entonces llamada *variable aleatoria*, y la suma $E(x) = \sum_q a_q P(A_q)$ es llamada *esperanza matemática* de la variable *x*. La teoría de las variables aleatorias será desarrollada en los capítulos III y IV. No nos limitaremos allí simplemente a aquellas variables aleatorias que pueden asumir solamente un número finito de valores diferentes (Kolmogorov, 1956/1933, p. 12).

Aquí queda claramente establecida la naturaleza funcional de la variable aleatoria como una función conjunto de valor real. Es notorio que esta definición es únicamente para variables aleatorias discretas con un número finito de valores. En su introducción al capítulo I, Kolmogorov aclara que en él partirá de situaciones con un número finito de eventos para enunciar sus principios y que los teoremas ahí enunciados pueden aplicarse también para casos infinitos, pero que «más tarde, cuando éstos sean estudiados, esencialmente nuevos principios serán usados» (p. 1). Con esta aclaración, él denota la necesidad de manipular, en primera instancia, una *teoría elemental de la probabilidad*, en donde enuncia los axiomas y definiciones a partir de los cuales partirá su teoría de probabilidad (*campo de probabilidad* e *independencia*) e introduce una nomenclatura propia de su teoría. Las dos definiciones de variable aleatoria dadas en ese capítulo obedecen ya la necesidad de establecer las relaciones básicas de un sistema axiomático: define, un tanto informalmente, los elementos que requerirá tanto para la justificación de la definición de independencia (p. 9) como de probabilidad condicional (p. 12). La mención a la variable aleatoria se hace necesaria desde la justificación y discusión de la definición de los elementos base de su teoría, sin embargo, también tiene la intención de desvincular lo concreto de sus nociones anteriores, de la abstracción a la que debe obedecer su teoría. Es decir, también hace explícita la necesidad de definir matemáticamente a la variable aleatoria, como parte fundamental de su teoría.

En el capítulo II define el *teorema de continuidad*, los *campos de Borel* y los *campos infinitos de probabilidad*. A partir de este capítulo su teoría es un fluir coherente de elementos basados exclusivamente sobre sus primeros axiomas.

En el capítulo III, Kolmogorov proporciona una definición más formal de la variable aleatoria:

> DEFINICIÓN. Una función de valor real $x(\xi)$, definida sobre el conjunto básico *E*, es llamada una *variable aleatoria* si para cada elección de un número real *a*, el conjunto $\{x < a\}$ de todos los ξ para los cuales la desigualdad $x < a$ es verdadera, pertenece al sistema de conjuntos $\mathfrak{I}$.
>
> Esta función $x(\xi)$ mapea el conjunto básico *E* en el conjunto R^1 de todos los números reales. Esta función determina, como en §1, un campo $\mathfrak{I}^{(x)}$ de un subconjunto de los conjuntos de R^1. Podemos formular nuestra definición de variable aleatoria de esta manera: Una función real $x(\xi)$ es una variable aleatoria si y solo si $\mathfrak{I}^{(x)}$ contiene todos los intervalos de la forma (-∞, *a*).
>
> Ya que $\mathfrak{I}^{(x)}$ es un campo, entonces junto con los intervalos (-∞, *a*), contiene todas las sumas finitas posibles de los intervalos abiertos por la derecha [*a*, *b*). Si nuestro campo de probabilidad es un campo de Borel, entonces $\mathfrak{I}$ y $\mathfrak{I}^{(x)}$ son campos de Borel; por lo tanto, en este caso $\mathfrak{I}^{(x)}$ contiene todos los conjuntos de Borel de R^1 (Kolmogorov, 1956/1933, p. 22).

Esta definición se sostiene en la definición de *campo de probabilidad* y sus axiomas, definidos en el capítulo 1:

> Sea *E* una colección elementos $\xi, \eta, \zeta, \ldots$, los cuales llamaremos *sucesos elementales,* y $\mathfrak{I}$ un conjunto de subconjuntos de *E*; los elementos del conjunto $\mathfrak{I}$ serán llamados *sucesos aleatorios.*
>
> I. $\mathfrak{I}$ *es un* campo *de conjuntos.*
>
> II. $\mathfrak{I}$ *contiene el conjunto E.*
>
> III. Para cada conjunto *A* en $\mathfrak{I}$ es asignado un número real no-negativo P(*A*). Este número P(*A*) es llamado la probabilidad del suceso *A*.

IV. P(E) es igual a 1.

V. Si A y B no tienen elementos en común, entonces $P(A+B)=P(A)+P(B)$.

> Un sistema de conjuntos, $\Im$, junto con una asignación definida de número P(A), que satisfacen los Axiomas I-V, es llamado un *campo de probabilidad* (Kolmogorov, 1956/1933, p. 2).

Esta definición, obedece, por tanto, a la axiomatización de la teoría de la probabilidad propiamente dicha, quedando circunscrita en un *campo de probabilidad*. Formalizó la noción que hasta entonces se había usado con poca precisión y la dejó claramente establecida dentro del sistema axiomático de la teoría de probabilidad.

Llama la atención, también, el desarrollo de la noción de *distribución de probabilidad* dentro de esta teoría. La noción es introducida en el capítulo II como una derivación de la condición que debe cumplir el conjunto contable E de los números reales x, con un campo $\Im$, definido como el conjunto de todos los puntos de Borel de esta línea, para construir un campo de probabilidad:

> Para construir un campo de probabilidad, con el campo dado $\Im$ es suficiente definir una función de conjunto arbitraria no negativa completamente aditiva P(A) sobre $\Im$ tal que satisfaga la condición P(E) = 1. Como es bien sabido, tal función es únicamente determinada por sus valores:
>
> $$P[-\infty; x) = F(x)$$
>
> Para los intervalos especiales:$[-\infty; x)$. La función $F(x)$ es llamada la función de distribución de ξ. (Kolmogorov, 1956/1933, pp. 18-19).

Notemos que esta definición ya requiere que E sea el número real x. En el capítulo III, define la *función de probabilidad* más estrictamente como una función de conjunto $P^{(u)}(A')$ = P$\{u^{-1}(A')\}$ de la función de conjunto $u(\xi)$, definida sobre sobre $\Im^{(u)}$. $u(\xi)$ mapea valores de E sobre E', de modo que cuando $u(\xi)=\xi$, la función de probabilidad $P^{(u)}(A')$ es igual a $P(A)$. La función de probabilidad, entonces, es definida como una función compuesta sobre la pre-imagen de u en E. También es cierto que tanto en la definición dada en el capítulo III, como en el II, Kolmogorov prepara la definición de variable aleatoria que desarrolla inmediatamente después a esta definición en el capítulo III. La función de probabilidad se convierte en función de variable real sólo después de la definición de variable aleatoria:

> En el futuro denotaremos a la función de probabilidad de una variable aleatoria por $P^{(x)}(A')$. Está definida para todos los conjuntos del campo $\Im^{(x)}$, en particular, para el

más importante caso, el campo de Borel de la probabilidad $P^{(x)}$ está definida para todos los conjuntos de Borel en R^1. (Kolmogorov, 1956/1933, p. 23).

La *función de distribución* es definida inmediatamente después de este enunciado tal como la conocemos actualmente y desde un inicio tiene la connotación de ser una *función de una variable aleatoria*. La *función de densidad* de probabilidad es manejada como el caso en que la función de distribución es diferenciable. En el capítulo III también son desarrollados algunos tópicos sobre variables aleatorias como variables aleatorias equivalentes y tipos de convergencia.

Kolmogorov postula como los conceptos base de su teoría en el concepto de *suceso aleatorio* y *su probabilidad*. En otras propuestas, como las de Bernstein y von Mises, estos conceptos eran definidos a través de otros conceptos, lo que ocasionó fallas en sus propuestas. Sin embargo, la principal crítica al sistema propuesto por Kolmogorov es que es incompleto, pues para un conjunto dado *E*, se pueden elegir las probabilidades en $\mathfrak{I}$ de manera diferente. Lo cual, no implica que no sea un sistema razonable, sólo que se debe ser consciente que no indica la forma en que se asignan las probabilidades.

6.2. Aportaciones de Lévy, Petrov y Parzen

Después de la publicación de su teoría, la nomenclatura propuesta por Kolmogorov se extendió por Europa. Así por ejemplo, Paul Lévy comenzó a utilizar el término «variable aleatoria» en enunciados y explicaciones, particularmente alrededor de la suma de variables aleatorias, en trabajos (por ejemplo, Lévy, 1936) en los que citaba a Kolmogorov, por lo que, suponemos, es influencia de éste último. Otras muestras del uso dado por Lévy al concepto de variable aleatoria son sus numerosos artículos en donde amplía sus resultados sobre el tema, estudiando las operaciones y convergencia de las variables aleatorias, por ejemplo, Lévy (1939), sobre la convergencia de series de variables aleatorias, Lévy (1935) sobre la adición de variables y Lévy (1959) sobre el producto de variables aleatorias. Los trabajos sobre variables aleatorias se han ampliado, desde entonces, enriqueciéndolo y haciéndolo más complejo.

También surgen nuevas propuestas del concepto de variable aleatoria, ampliando o modificando la definición dada por Kolmogorov. Así por ejemplo, V. Vladímirovich Petrov, cuya tesis doctoral fue *Some extremal problems in the theory of summation of independent random variables*, y que realizó otros muchos trabajos sobre el tema, como *Sums of independent random variables* publicado en 1972, o *Limit theorems for sums of* *independent random variables* publicado en 1987, definió la variable aleatoria como sigue:

> Consideremos un espacio de probabilidades (Ω, A, P). Llamamos *variable aleatoria* a una función real $X = X(\omega)$ definida en el espacio de sucesos elementales Ω y tal que satisface la condición:
>
> *(3.1)* $$\{\omega \in \Omega : X(\omega) \leq x\} \in A$$
>
> para todo *x* real.
>
> En la terminología del análisis real, una función $X(\omega)$ que cumple la condición (3.1) para todo *x* se denomina *medible.* De esta forma, una variable aleatoria es una función real y medible sobre los sucesos elementales. Se puede verificar que la condición (3.1) para todo *x*, es equivalente a la condición
>
> *(3.2)* $$\{\omega \in \Omega : X(\omega) \in B\} \in A$$
>
> para cualquier conjunto boreliano[24] ***B*** de puntos de la recta real $\Re$. En el caso particular en que ***B*** es el intervalo $(-\infty, x]$, la condición (3.2) se convierte en la condición (3.1) (Petrov y Mordecki, 2002, p. 55).

Esta definición es básicamente la misma dada por Kolmogorov con cambios en la notación y, sobre todo, utilizan los modernos conceptos de σ-álgebra y conjuntos borelianos para referirse a lo que Kolmogorov denotó como campos de Borel, $\Im$, y campos de Borel en R^1, $\Im^{(x)}$, respectivamente. Esta notación es la más usada actualmente.

Emanuel Parzen, en cambio, se aventura en una propuesta más innovadora. En su obra *Modern probability theory and its applications*, que ha tenido un amplio reconocimiento, sobre todo en el contexto de ingeniería (Newton, 2002), propone un tratamiento distinto a la introducción de la variable aleatoria. Él usa el concepto de *fenómenos aleatorios con resultados numéricos* para introducir a la variable aleatoria y las nociones que hay que tratar simultáneamente con ella, en particular, la ley de probabilidades (distribución de probabilidad o función de densidad) y función de distribución, la esperanza, la varianza, la ley de los grandes números y algunas leyes de probabilidad, en particular, la uniforme, normal, Poisson, exponencial y gamma. Inmediatamente después introduce la definición formal de variable aleatoria y la teoría

[24] Nota del autor: La clase de los conjuntos boreleanos en la recta es la mínima σ -álgebra de conjuntos que contiene a todos los intervalos.

alrededor de ella en los capítulos 7 a 10: *Variables aleatorias, Esperanza de una variable aleatoria, sumas de variables aleatorias independientes* y *Sucesiones de variables aleatorias.*

Define fenómeno aleatorio con resultados numéricos en la forma siguiente:

> *Un fenómeno aleatorio con resultados numéricos* es un fenómeno aleatorio cuyo espacio de descripciones muestrales es el conjunto *R* (de los números reales de $-\infty$ a ∞) en cuyos subconjuntos está definida una función $P[\cdot]$, que asigna a cada conjunto boreliano de números reales *E* (que también se llama suceso) un número real no negativo, denotado por $P[E]$, de acuerdo con los axiomas siguientes:
>
> AXIOMA 1. $P[E] \geq 0$ para todo suceso *E*
>
> AXIOMA 2. *P*[*R*] = 1
>
> AXIOMA 3. Para toda sucesión de sucesos $E_1, E_2, \ldots, E_2, \ldots,$ que sea mutuamente exclusiva, tenemos que $P\left[\bigcup_{n=1}^{\infty} E_n\right] = \sum_{n=1}^{\infty} P[E_n]$
>
> (Parzen, 1971, p. 173).

Posteriormente, Parzen extiende el modelo a fenómeno aleatorio con resultados numéricos en *n* dimensiones cuyo espacio muestral es el conjunto $\Re^n$ que consiste en todas las *n*-adas $(x_1, x_2, \ldots, x_n)$ cuyas componentes $x_1, x_2, \ldots, x_n$ son números reales de $-\infty$ a ∞.

Su propuesta pretende partir de situaciones cotidianas (*aplicaciones*) para plantear matemáticamente problemas de probabilidad a fin de resolverlos en forma sistemática, puesto que prioriza el uso de *x* (variable real) en el desarrollo de la teoría. Para Parzen, la conceptualización de los fenómenos aleatorios con resultados numéricos constituye una forma relativamente fácil de adentrarse en el caso de espacios muestrales no necesariamente finitos en donde los eventos son conjuntos borelianos de números reales. Él, igual que Kolmogorov, ve la necesidad de una definición preliminar de la variable aleatoria y de todos los objetos relacionados con ella para analizar los espacios discretos y finitos, pero, también como Kolmogorov, en su primera aproximación, aclara la existencia de otra más rigurosa, en una especie de manifestación de que esta no es la verdadera definición de variable aleatoria:

> Algunos escritores […] llaman variable aleatoria al número *X* determinado por el resultado de un fenómeno aleatorio (como lo es el valor observado de un fenómeno aleatorio con resultados numéricos). En el capítulo 7 damos una definición rigurosa del

> concepto de variable aleatoria en términos del concepto de función, y demostramos que el valor observado *X* de un fenómeno aleatorio con resultados numéricos puede ser considerado como variable aleatoria. Por lo pronto tenemos la siguiente definición:
> *Se dice que una cantidad X es una variable aleatoria (o lo que es lo mismo, se dice que X es un valor observado de un fenómeno aleatorio) si para todo número real x existe una probabilidad de que X sea menor o igual a x.* (Parzen, 1971, p. 195).

Efectivamente, en el Capítulo 7 aparece la definición rigurosa de variable aleatoria. Ahí, aclara que en muchas aplicaciones de teoría de probabilidades, los espacios muestrales, aunque siempre implícitos, no aparecen directamente. Parzen reconoce que en esos casos el método basado en fenómenos aleatorios con resultados numéricos «no es muy satisfactorio» (p. 299). Es en esas aplicaciones donde la variable aleatoria toma un papel primordial: el análisis probabilístico se basa en este concepto, puesto que éste hace innecesario fijar de antemano el número *n* de fenómenos aleatorios. Además de que hace posible generar por medio de operaciones algebraicas y analíticas, nuevos fenómenos aleatorios a partir de fenómenos aleatorios conocidos. Parzen emite su definición rigurosa de esta forma:

> DEFINICIÓN DE VARIABLE ALEATORIA. Decimos que un objeto X es una variable aleatoria i) si es una función de valores reales definida en un espacio de descripciones muestrales, sobre una familia de cuyos subconjuntos hayamos definido una función de probabilidades $P[\cdot]$, y ii) si para todo conjunto boreliano *B* de números reales, el conjunto $\{s : X(s) \text{ pertenece a } B\}$ pertenece al dominio de $P[\cdot]$ (Parzen, 1971, p. 300).

Inmediatamente después, vincula esta definición con la dada anteriormente, en donde resalta la relación bidireccional de ambas:

> Una variable aleatoria es, entonces, una función definida con base en el resultado de un fenómeno aleatorio; consecuentemente, el valor de una variable aleatoria es un fenómeno aleatorio y es, en efecto, un fenómeno aleatorio con resultados numéricos. De manera análoga podemos interpretar todo fenómeno aleatorio con resultados numéricos como el valor de una variable aleatoria *X*; en particular la variable aleatoria definida sobre la línea real, para todo número real, *x* es $X(x) = x$. (Parzen, 1971, p. 300)

A lo largo de su libro, compara los nuevos métodos y conceptos vinculados con la variable aleatoria con lo expresado en los primeros capítulos para los fenómenos aleatorios con resultados numéricos, haciendo a la teoría de probabilidad una extensión del análisis de éstos y ocupa a la variable aleatoria para ello:

> Aunque por definición una variable aleatoria X es una función sobre un espacio de probabilidades, en la teoría de probabilidades nos preocupamos raras veces de la forma funcional de X, puesto que no nos interesa calcular el valor de $X(s)$ que la función X asume de algún elemento individual s del espacio de descripciones muestrales S, sobre el cual se ha definido X. Más bien nos interesa la probabilidad de que un valor observado de la variable aleatoria X pertenezca a un conjunto dado B. Nos interesa una variable aleatoria en la medida en que es un mecanismo que da lugar a un fenómeno aleatorio con resultados numéricos, y las preguntas que hagamos acerca de una variable aleatoria X serán precisamente las mismas que nos sirvieron para estudiar fenómenos aleatorios con resultados numéricos. (Parzen, 1971, p. 302).

A su vez, en esta cita expresa la utilidad del concepto de variable aleatoria en el sentido funcional con la probabilidad desde una perspectiva epistemológica, por la teoría que surge a partir de ella, y desde una visión práctica por las aplicaciones que miran en ella una variable algebraico-funcional con la probabilidad. Sin embargo, Parzen también resalta la definición funcional de la variable aleatoria, aún en las situaciones prácticas:

> Uno de los tropiezos de los estudiantes con el concepto de variable aleatoria consiste en que los objetos que son variables aleatorias no están siempre definidos de tal manera que ello resalte explícitamente. Sin embargo, […] hay que aprender a reconocer y formular matemáticamente como funciones los objetos descritos verbalmente que sean variables aleatorias. (Parzen, 1971, p. 300).

Parzen no rige su discurso por la distinción entre la variable discreta y la variable continua, sino más bien por la naturaleza de los problemas (o aplicaciones) que va tratando. Así define la función de distribución normal como una generalización de la binomial y las distribuciones vinculadas con el proceso Poisson las trata de manera continua, primero la variable discreta y después la exponencial y Gamma. Da una definición de variable discreta y continua inmediatamente después de la definición de función de distribución, pero anteriormente ya había abordado varias distribuciones discretas y continuas. La propuesta de Parzen da prioridad al sentido que puedan tener los objetos matemáticos dentro de las aplicaciones, a la vez que trata de establecer un sistema axiomático. Esto, para el caso de la variable discreta y continua, se puede observar en la anotación que hace antes de definir por primera vez la función de distribución y en donde la definición de fenómeno aleatorio con resultados numéricos le permitió establecer primeramente las funciones de masa y de densidad:

> …para cualquier fenómeno aleatorio con resultados numéricos, existe una función de punto, llamada *función de distribución*, que es suficiente para determinar la función de probabilidades, en el sentido de que ésta puede reconstruirse a partir de la primera. Así pues, la función de distribución constituye una función de punto que contiene toda la información necesaria para describir las propiedades probabilísticas del fenómeno aleatorio. Por consiguiente, para estudiar las propiedades generales con resultados numéricos, sin tener que restringirnos a aquellos cuyas funciones de probabilidades se han especificado, ya sea por una función de densidad o una función de masa de probabilidades, basta con estudiar las propiedades generales de las funciones de distribución. (Parzen, 1971, p. 191).

En su discurso, Parzen también introduce la concepción de función de distribución antes que la de variable aleatoria. Ésta surge como una respuesta a una necesidad impuesta por la definición de la función de distribución.

Así, su propuesta establece un cambio en la definición de variable aleatoria, dando prioridad a las aplicaciones sin perder la formalidad y rigor de la teoría. De alguna manera este interés de Parzen de establecer una «introducción» a la teoría de probabilidad antes de abordar temas más rigurosos, ya se observaba en Kolmogorov, pero Parzen lo lleva más allá al insertar esta introducción a través de aplicaciones. La teoría de Parzen, lo mismo que la de Kolmogorov, requirió establecer definiciones más rigurosas para poder abordar aplicaciones campos más formales de la probabilidad. La definición inicial de Kolmogorov no modifica su definición formal posterior, en cambio la de Parzen sí. A través de la bidireccionalidad entre sus dos definiciones, la misma variable aleatoria formal se convierte también en un fenómeno aleatorio con resultados numéricos. Sin embargo, en ambas teorías la continuidad es la que proporciona la necesidad de una definición formal de variable aleatoria, en el caso de Kolmogorov, más rigurosa, y de Parzen, en la introducida por primera vez, como el resultado de un fenómeno aleatorio con resultados numéricos.

Seguramente en épocas más recientes nuevas propuestas sobre las teorías alrededor de la variable aleatoria están surgiendo, pero, también seguramente, no están al alcance de nuestros propósitos en esta investigación.

7. Conclusiones del Capítulo

La problemática que hizo que el concepto de variable aleatoria deviniera como objeto matemático, no está vinculada a un solo campo de problemas. La variable aleatoria se manejó de manera implícita desde la antigüedad en diferentes contextos culturales, y a partir de múltiples aplicaciones, tanto matemáticas como extra matemáticas, sin que se percibiera la necesidad de definirla hasta que se requirió formalizar la herramienta probabilística. Para construir un sistema axiomático coherente en probabilidad (formular y demostrar teoremas, como el central del límite y la ley de los grandes números) se requería la conceptualización explícita de una variable que difería de la algebraica en diversos sentidos, pero sobre todo en su sentido funcional. Había que definir un objeto matemático que permitiera establecer funciones reales a partir de funciones de conjunto. Esto hizo que los probabilistas estudiaran la teoría de la medida e integración antes de Lebesgue (Doob, 1976).

Sin embargo, antes de la necesidad del surgimiento de la axiomatización de la probabilidad y a lo largo de mucho tiempo, se presentaron manifestaciones sutiles del requerimiento de un objeto matemático de la naturaleza de la variable aleatoria. También, por otro lado, los estadísticos necesitaron fundamentar algunas de sus propuestas de solución a problemas más prácticos en resultados probabilísticos. Esto hizo que de manera paralela y retroalimentándose mutuamente, la variable aleatoria y la estadística fueran trabajadas en ambos campos, el probabilístico y el estadístico, hasta conformarlas teóricamente y diferenciarlas y relacionarlas explícitamente.

Así, a lo largo de este estudio histórico se ha visto que la conceptualización de la variable aleatoria deviene muy lentamente a través de dos caminos diferentes, que se entrecruzaron una y otra vez para complementarse mutuamente:

❖ *El surgido a partir de los análisis probabilísticos* que, inicialmente partieron de los primitivos juegos de azar y el estudio de la previsión de apuestas, que, aunque se practicaron desde tiempos remotos, sólo se estudiaron desde un punto de vista probabilístico a partir del siglo XVII. Este tipo de estudios paulatinamente se fue despegando de sus aplicaciones hasta lograr la conformación de una teoría probabilística en el primer tercio del siglo XX. El principal interés estaba en la formulación de un sustento matemático que diera fundamento a la ciencia del azar basándose en el rigor y formalismo de la ciencia. Los científicos más influyentes estaban instruidos y muy actualizados en las matemáticas y ciencias de su tiempo. Los resultados que se desarrollaban en la creciente matemática del cálculo y de la teoría de conjuntos tuvieron mucha influencia en la teoría de probabilidad, pero también las necesidades de los probabilistas generaron resultados que actualmente pertenecen y son ampliamente usados en la matemática actual, por ejemplo en ecuaciones diferenciales, o a la teoría de la medida, sin que, en ocasiones y desde una perspectiva teórica, guarden vinculación con sus orígenes probabilísticas. Los principales problemas abordados aquí rápidamente se desvincularon de las aplicaciones y más bien estaban vinculados con su fin del desarrollo y fundamentación de una teoría: la generalización de problemas resueltos de manera particular, la obtención de métodos generales, el surgimiento de objetos matemáticos acordes con las generalizaciones, la demostración de teoremas y por último la búsqueda de los fundamentos que dieran lugar a la conformación de una teoría.

- Por otro lado, también había *un interés creciente por el análisis de los datos recopilados por distintas necesidades prácticas*. Los estudios estadísticos, también muy antiguos, se comenzaron a formalizar a partir del siglo XVII con trabajos sobre seguros de vida, estadísticas demográficas, previsión de causas de muerte y estudios poblacionales que hicieron uso de resultados probabilísticos en el análisis de sus datos. A su vez, las rutas del análisis de datos generadas encontraron cada vez mayores aplicaciones y sustento a medida que la teoría de la probabilidad se desarrollaba. Inicialmente los problemas resueltos a través de esta vía estaban vinculados con datos surgidos en diversos contextos prácticos, así que la definición de una variable estadística en ellos era algo relativamente simple. El mismo contexto indicaba la necesidad de resúmenes de datos numéricos en los que se definía una variable numérica. Conforme el número de aplicaciones creció, también lo hizo la necesidad de la conformación de una teoría que fundamentara los métodos y resultados que se obtenían. Se dio paso a la vinculación entre resultados teóricos y empíricos, es decir, al establecimiento de un vínculo formal entre la variable estadística y la variable aleatoria. Uno de los principales derroteros que hizo que surgiera esta necesidad en realidad es una preocupación surgida desde los primeros análisis de datos: la inferencia.

No podemos decir que estas dos vías se hayan desarrollado de manera independiente. Desde sus inicios, la probabilidad tuvo fuertes motivaciones en la observación del comportamiento de la tirada de dos dados: los grandes jugadores (el Duque de Toscana y Fournival) debieron tener registros para percatarse de las diferencias minúsculas entre los resultados producidos por sus análisis y los obtenidos a través de un modelo probabilístico teórico erróneo que no obtenía apropiadamente el espacio muestral del lanzamiento de los dados. Así mismo, desde sus inicios, investigadores preocupados por el comportamiento de los datos, como Arbuthnott, relacionaron modelos obtenidos a partir de los datos con modelos teóricos, en el caso de Arbuthnott, con el desarrollado por Bernoulli de la binomial. Se puede decir que casi desde su surgimiento, los modelos de distribuciones y los teoremas límite se utilizaron con la intención de predecir el comportamiento futuro de fenómenos de tipo estadístico y, a la inversa, los probabilistas hicieron uso de los análisis de datos para obtener y establecer algunos de sus modelos teóricos. Las primitivas aplicaciones de los modelos probabilistas se ampliaron a campos tales como la astronomía o física que se interesaron por magnitudes o cantidades aleatorias[25] y permitieron el desarrollo de algunos modelos de distribuciones de probabilidad. Todos estos puntos de convergencia entre los modelos probabilistas y el análisis de los datos proporcionaron herramientas útiles a los estadísticos y el desarrollo de nuevas teorías a los probabilistas. Lo que, a su vez, abría nuevas preguntas sobre la relación entre las variables aleatorias y las estadísticas.

Sin embargo, también es cierto que durante largo tiempo, el contacto entre estas dos vías sólo fue tangencial. El desarrollo de diversas herramientas matemáticas no hubiera sido posible si las comunidades no se hubieran centrado en sus propios problemas, pero tampoco si no hubiera habido contacto entre ellos.

Así, a lo largo de su historia, estas dos vías dieron lugar a dos corrientes de pensamiento que se conformaron autónomamente nutriéndose de sus mutuos análisis y resultados, generando un proceso dialéctico, pero también con momentos de reflexión propia, mediante los cuales se establecieron formas de razonamiento autónomas que los conformaron como ciencias independientes, cuyo sustento en gran parte pende de la conceptualización de las variables aleatoria y estadística y su vinculación. La historia de las variables aleatoria y estadística es también la historia de cómo se conformaron estos campos como ciencias a través de esas múltiples interacciones en diferentes épocas y etapas de desarrollo. Es en el siglo XX cuando a través de la axiomatización de la probabilidad se llegó a la definición actual de variables aleatorias de notable complejidad. Paralelamente, en Inglaterra y Estados Unidos principalmente, se creó la escuela biométrica que daría paso al estudio formal de la inferencia estadística, en donde se explicita la diferencia entre las variables estadística y aleatoria, estadístico y parámetro y se crean métodos para estimar los parámetros a partir de datos estadísticos.

Aún hoy la estadística conserva una problemática dual, en la cual, sobre todo en el ámbito inferencial, coexisten problemas prácticos y teóricos, y a través de la cual se produce nueva herramienta estadística y surgen nuevas formas de vincular las variables aleatoria y estadística. Así mismo, algunas veces, en las distintas prácticas ingenieriles y

[25] En esa época la variable aleatoria era vista como el resultado de fenómenos aleatorios con resultados numéricos. Acepción influenciada por la forma en que se concebía a la variable en el ámbito determinista.

de ciencias, la variable aleatoria o estadística son utilizadas sin llegar a su formalización.

7.1. Etapas históricas

Un análisis más detallado permite identificar las ocho etapas históricas del desarrollo de la variable aleatoria que a continuación se enumeran, algunas de las cuales se solapan cronológicamente entre sí.

Primera etapa

En el campo de los juegos de azar, se presume que jugadores interesados en conocer la mejor apuesta posible, comenzaron a hacer recuentos de datos obtenidos a partir de su constante participación en esos juegos. No se cuenta con evidencia del análisis de estos datos, ni de los recuentos sin embargo las cartas del Duque de Toscana indican que se tenía un manejo intuitivo de la variable estadística «la suma de las caras superiores de tres dados» y de que se tenía conocimiento de la diferencia de aparición entre algunos valores de la variable cuya probabilidad difería en 0,0093 (0,1157 para los valores de 9 y 12 y 0,125 para las sumas 10 y 11). Esto muestra no solo la amplitud del recuento, sino también la existencia de análisis numéricos intuitivos y rudimentarios de la variable estadística.

Segunda etapa

En el mismo ámbito de los juegos de azar, los jugadores comenzaron a hacer recuentos teóricos de los resultados posibles. Algunos de ellos fueron correctos, como los casos del descrito en el manuscrito «De Vetula» o los planteados por Cardano y Huygens, y algunos otros erróneos e incoherentes con la experiencia aportada por la recopilación de sus datos, como es el caso del establecido por el Duque de Toscana, quien recurrió a Pascal y Fermat a fin de que le proporcionaran modelos adecuados. En todos los casos, se comenzaron a asociar valores numéricos discretos a resultados de experimentos aleatorios. No se usó el término variable, y sólo en algunos de los problemas se hace explícita la variable para el problema en particular. El principal interés es la esperanza matemática de la variable para determinar la apuesta justa en juegos de azar y también la probabilidad de valores aislados para tratar de encontrar estrategias en los juegos. Se identificaron probabilidades asociadas a todos los valores posibles de las variables aleatorias discretas vinculadas a esos problemas e incluso se comenzaron a estudiar algunos modelos concretos de sus distribuciones de probabilidad. Estos resultados se sustentaron principalmente en el análisis teórico del espacio muestral. Los escritos que dan cuenta de la existencia de esta etapa no tienen una inquietud académica, sino más bien, de consejo y recomendaciones de apuesta en los juegos de azar e incluso algunos de ellos caen en consejos morales. El libro de Huygens, sin embargo, ya tiene un manejo más riguroso, establece conceptos y un lenguaje para explicar sus soluciones y algunas generalizaciones.

Tercera etapa

El interés en la recolección de datos estadísticos para fines militares o políticos, conlleva al uso intuitivo de variables estadísticas y a la determinación de su distribución de frecuencias, inicialmente sin gran aparato matemático. Alrededor del siglo XVII, en Inglaterra y posteriormente en otros países, se inicia un estudio más formal en que se recogen datos estadísticos sistemáticamente y se analizan sus regularidades con el objetivo de hacer predicciones, basadas en un uso implícito de la variable aleatoria (Graunt, Arbuthnott, Halley, Galileo). El manejo de la media es importante, pero Galileo Galilei también comenzó a resaltar la variación de la distribución con la introducción del cuadrado del error medio. El proceso de modelación se establecía a partir del análisis de datos empíricos con el uso de algunas herramientas probabilísticas teóricas. Las variables empleadas se definían por el contexto del problema, sin embargo también se comienza a denotar la importancia de las variables numéricas para hacer uso de algunos resultados probabilísticos. Además se comienza el análisis empírico de la variable continua a partir del análisis de los errores en las mediciones (Galileo).

Cuarta etapa

Se incrementa el interés en la formalización y generalización de la herramienta generada a partir de problemas particulares. Las aportaciones teóricas esta etapa son abundantes, pero surgieron principalmente a partir de dos líneas conductoras muy relacionadas entre sí. Por un lado, el interés generado por las múltiples aplicaciones de los resultados teóricos a estudios empíricos, genera cuestionamientos sobre la relación teórica entre la probabilidad frecuencial y la obtenida a través de enfoques teóricos, lo que también conlleva a cuestionamientos sobre la relación entre la variable aleatoria y la estadística. Así, aparece la condición de independencia entre variables aleatorias y la ley de los grandes números (en *Ars Conjectandi* de J. Bernoulli y en la *Doctrina del azar* de de Moivre), así como las primeras versiones del teorema central del límite para el caso particular de la distribución binomial. En otra línea conductora, se inicia la aproximación a las funciones como una forma de cálculo de la probabilidad de los valores de la variable cuando éste les resultaba complicado, lo que conlleva a un estudio más profundo sobre las funciones de densidad con variables continuas (Bernoulli, de Moivre, Gauss, Laplace, Poisson) no siempre de manera consciente. Con ello se definen las primeras operaciones entre variables aleatorias y la convergencia de sucesiones de variables aleatorias, aunque no siempre explícitas. Se comienza el trabajo con variables aleatorias continuas y se define la probabilidad como un diferencial, así mismo se inicia el interés por las funciones de densidad y el cálculo de la probabilidad de intervalos de valores de una variable aleatoria mediante integrales. Se identifican distribuciones de probabilidad para variables aleatorias continuas, particularmente la Normal y la Exponencial. Con ello también se definen variables aleatorias continuas implícitamente y se comienza a denotar más fuertemente la necesidad de conceptualizar una variable numérica vinculada al azar, así como de la función de densidad como una forma de cálculo de la probabilidad, sin distinguirla de la función de probabilidad.

En esta etapa, los problemas reales como los errores de medición, problemas de sobrevivencia o aplicaciones a la teoría matemática de la física molecular aplicada a los gases se describen mediante variables (Gauss, de Moivre, Daniel Bernoulli, Laplace), pero no se usa consistentemente el término *variable*; a veces se le denomina *magnitud* o *cualidad variable* o bien, no se explícita. En todo caso, su uso se asemejaba al dado por los matemáticos dentro del ámbito determinista en ese tiempo. Así mismo, Laplace, principalmente, comienza a diferenciar implícitamente entre la variable y los valores que puede tomar la variable en contextos discretos y continuos.

Quinta etapa

Aparecieron indicios explícitos de variables aleatorias asociadas a contextos y a demostraciones. En esta etapa el principal motor de desarrollo de la variable aleatoria fue la necesidad de establecer una variable de carácter matemático en la función de densidad y el problema de la continuidad, que también estuviera vinculada con el contexto probabilísta. Se observó la necesidad de una *función, X* que vinculara los resultados de un contexto aleatorio con un contexto matemático en donde se le asociaban sus probabilidades (Poisson, Tchebycheff). De manera general se diferenciaron explícitamente los valores de la variable aleatoria y la variable en sí misma y se comenzó a trabajar con el concepto de suceso vinculado con una variable en el contexto matemático, así mismo se trabajó ampliamente con secuencias de variables aleatorias y se profundizó en el problema de la continuidad en el cálculo de probabilidades, cosa que conllevó a diferenciar entre el trabajo con variables continuas y discretas (Poisson, Tchebychev, Markov, Liapounoff). También el cálculo de probabilidades de variables aleatorias dependientes y de sus funciones de distribución surgió de manera explícita (Markov, Liapounff). Todos estos resultados ampliaron el concepto de variable en probabilidad, denotando diferencias importantes con aquellas de cálculo y denotaron la distinción entre la función de densidad o la distribución de probabilidad y la función de probabilidad, lo que marcó fuertemente la naturaleza de la variable continua y de su función de distribución. Así mismo permitió el surgimiento de enunciados más elaborados del teorema central del límite, en particular por parte de Liapounoff y Markov.

Poisson, además, trabajó con la asignación de resultados no-numéricos de experimentos aleatorios a valores numéricos a los que, en un intervalo dado, se le asignan probabilidades. Es decir requirió la definición de una regla que transformara variables cualitativas a cuantitativas, que por tanto no representan una magnitud en el sentido clásico[26].

Al final de esta etapa, las variables aleatorias ya se denotaban como *variables* e incluso algunos autores, el principal, Liapounoff, comenzaron a hacer explícita su condición de «sometidas al azar» (Liapounoff, 1901, p. 126). El apellido de *aleatoria*, aún no es propio de ese tiempo en el que la diferencia entre azar y aleatoriedad aún no es explícita. A finales de esta etapa, la variable aleatoria empieza a perder su carácter físico, es decir, ya no es manejada como el resultado de un fenómeno aleatorio con resultados numéricos, al menos en el ambiente formal de la probabilidad.

Sexta etapa

En la misma época (finales del siglo XIX y principios del XX), en otros ámbitos, las inferencias informales comenzaron a tener auge a partir de la difusión de la distribución normal como un modelo apropiado para muchos contextos en ciencias sociales, biológicas y de la herencia. Estas aplicaciones y formas de análisis atrajeron la mente de muchos científicos con un nivel de conocimientos amplio en su especialidad, pero que no eran matemáticos o estaban actualizados en los últimos resultados de los probabilistas, por lo tanto, sus resultados no tenían la solidez de una teoría en estadística matemática, y, sin embargo, proporcionaron derroteros por el que la estadística pudo desarrollarse como ciencia. La variable estadística jugó un papel primordial porque, a la vez que era fácilmente obtenida a partir de los contextos de interés, también se resaltó su carácter numérico para poder ser analizada y evaluada bajo el modelo de la

[26] Actualmente se ha ampliado el concepto de «medida» y se definen varias escalas de medida, incluyendo las nominales y ordinales, a las que se aplican datos estadísticos. Por tanto las variables aleatoria y estadística podrían modelar nuevos tipos de «magnitudes» que no cumplen la condición matemática de semi módulo.

distribución normal. Así, esta etapa se caracteriza por el interés en inferencias informales, por aportaciones a partir de las aplicaciones a la ciencia y por la ampliación en los campos de aplicación de la estadística. En todo esto, sobresalen los trabajos de Quetelet y Galton.

Inicialmente el interés fundamental era acercarse a la verdadera distribución de la variable estadística, lo que permitiría caracterizar a la población. Se preocuparon por definir lo que era una población en estadística. La recolección de grandes cantidades de datos prevaleció como la mejor forma de conocer como se comportaba una variable estadística en una población. Surgieron las representaciones a través de gráficos y tablas para mostrar los resúmenes de sus datos. Esto propició que las herramientas usadas fueran un tanto ingenuas, pero a su vez, fueran difundidas y comprendidas por muchos científicos a los que les importaba las aplicaciones de la naciente ciencia de los datos. Algunas de las principales herramientas surgidas, con ligeras diferencias de las herramientas actuales, fueron los percentiles, las ojivas, los histogramas, la mediana y la variación alrededor de la mediana. En particular, los percentiles fueron usados para medir no sólo la centralización (la mediana), sino también de dispersión, generando intervalos a partir de los percentiles. Cosa muy útil para comparar poblaciones o datos de una variable que se creía se comportaba como una normal. En esto, las ojivas le fueron muy útiles. Se comenzó a emplear la normal estándar para calcular probabilidades acumuladas o para hacer la prueba de si una variable se comportaba como una normal e incluso se generaron las primeras tablas de la distribución normal estándar para un uso más simple. También se comenzó a diferenciar la variación debida causas identificables y la regresión lineal. Sin embargo, como se expresó al principio, todos estos los avances se hicieron y usaron de manera informal, sin gran aparato matemático que la sustentara.

Séptima etapa

Esta etapa se caracteriza por el surgimiento de la variable aleatoria como objeto matemático a partir de la necesidad de abstracción para la definición de una teoría, pero también por el surgimiento de ampliaciones que nacieron a partir de su primera definición formal. Fréchet y Kolmogorov conceptualizaron a la variable aleatoria generalizada, que puede no tener detrás una magnitud, en el sentido físico. Con base en la teoría de conjuntos y de la medida, se formalizó la definición de variable aleatoria como una función medible de valor real definida en un espacio de probabilidad. Esto contribuyó a que se multiplicaran las aportaciones a la teoría de las variables aleatorias (Levy, Feller, Meyer, Petrov), enriqueciéndola y haciéndola más compleja.

En esta etapa la variable aleatoria adquiere identidad propia y es denotada tal como la conocemos actualmente, no sólo en su simbología sino también en su nomenclatura. También, en épocas más actuales, surge, como producto de un análisis concienzudo de la teoría de Kolmogorov y del uso reciente de la variable aleatoria, la ampliación de Parzen. En la que propone definir la variable aleatoria en dos momentos, comenzando con su definición a partir de un fenómeno aleatorio con resultados numéricos. De esta forma, sugiere un sistema que toma en cuenta el sentido que pueden tener los objetos matemáticos dentro de aplicaciones y que no pierde coherencia, formalidad ni rigor.

Octava etapa

Casi de manera simultánea a la formalización de la probabilidad, también la estadística inicia con tal proceso. A partir del trabajo de Galton, Karl Pearson se preocupó por fundamentar, perfeccionar y ampliar sus métodos haciendo uso de los resultados de los matemáticos y probabilistas, pero también generando objetos matemáticos propios. Las herramientas estadísticas para el análisis de datos se multiplicaron, fundamentándolo y haciéndolo más potente. La inferencia se convierte en uno de los derroteros que hacen que la estadística formule sus propios métodos y procesos. Surge la teoría del contraste de hipótesis e intervalos de confianza, así como el diseño experimental (Pearson, Karl Fisher, Egon Pearson, Neyman). También se ocupan de la teoría del muestreo y la necesidad de controlar el tamaño de muestra. Aparecen distribuciones que explícitamente se ocupaban de la inferencia, en particular la χ^2 y la t de Student. Al mismo tiempo, las aplicaciones de la inferencia en agronomía y biometría se extienden progresivamente a todos los campos científicos. En esta época se vinculan explícita y formalmente las variables estadística y aleatoria mediante los problemas de estimación, así mismo, la necesitad del estudio de las distribuciones muestrales incide en el desarrollo de la teoría de variables aleatorias, particularmente en el teorema central del límite y las variables aleatorias multidimensionales.

7.2. Interacciones entre los análisis estadísticos y probabilísticos.

A lo largo de las ocho etapas históricas, se observan las distintas formas en cómo las dos vías a través de las cuales la variable aleatoria devino como objeto matemático se entrecruzaron y retroalimentaron mutuamente de diferentes formas. A continuación se detallan algunas de las formas en que se da esta interacción y ejemplificamos con algunos de los pasajes más evidentes que identificamos en este estudio.

1. Confrontaciones entre resultados obtenidos de manera empírica que no concuerdan con modelos teóricos. La evidencia empírica estimuló la generación de teoría, lo que propició el desarrollo de la variable aleatoria.

   - El Duque de Toscana planteó un discordancia entre el comportamiento observado en los juegos de azar y un modelo errático de espacio muestral y se lo mostró a Galileo. A partir de este estímulo, se generó un modelo apropiado para el recuento teórico de los posibles resultados.
   - Arbuthnott aplicó la binomial para calcular teóricamente la proporción de pájaros machos y hembras que debían nacer y la comparó con los datos recopilados en muchos años. Esta comparación entre ambos análisis también lo hizo con datos poblacionales de seres humanos. En ninguno de los dos casos concordó el resultado de los tipos de análisis. Esto propició el cuestionamiento sobre la correspondencia entre la probabilidad frecuencial y la clásica.

2. Análisis de datos que buscaban deducir el comportamiento de la variable aleatoria a partir de la estadística. Esto propició nuevas teorías tendientes a la vinculación entre la variable aleatoria y la estadística.

   - Halley estudió grandes bases de datos poblacionales, lo que indujo al análisis de la esperanza de vida y estudios de mortandad.
   - El estudio de Galileo sobre los errores de los datos astronómicos condujeron a una primera aproximación al estudio de la normal estándar.
   - Daniel Bernoulli estudió fenómenos económicos y sociales a partir de datos, lo que lo lleva al análisis de riesgos y a la econometría.

3. Análisis teóricos que justifican resultados surgidos a partir del análisis de datos y que amplían el estudio de la variable aleatoria.

   - De Moivre, Lagrange, Gauss y Laplace ampliaron y demostraron teóricamente los resultados obtenidos por Galileo a partir del análisis de los datos astronómicos sobre la forma en que se agrupan los datos alrededor de la media en la distribución normal estándar. De Moivre, Lagrange, Gauss y Laplace encontraron la fórmula de la distribución normal, el método de mínimos cuadrados y el uso de la media y la desviación estándar en esta distribución.

4. Análisis teóricos que buscaron concordancia entre la variable estadística y la variable aleatoria. No hay un análisis de datos propiamente, sino estudios teóricos a través de los cuales, sin embargo, se genera nueva teoría que favorece la vinculación entre la variable aleatoria y la estadística.

    - ❖ Jacob Bernoulli quería establecer el número de datos que harían que la probabilidad frecuencial convergiera a la probabilidad teórica. Esto genera una primera propuesta de la ley débil de los grandes números.
    - ❖ Gosset estableció la distribución *t* para pequeñas muestras porque se percata que en algunos ambientes los grandes números no eran posibles. Esto conduce a una mejor deducción de la variable aleatoria a partir de la estadística en pequeñas muestras.
    - ❖ De Moivre estudió la diferencia entre la probabilidad y la frecuencia relativa. Se dio cuenta que la igualdad entre ellas no podría ser posible y que más bien se debería medir (aplicar momios) a la variación de la igualdad. Esto lo condujo a la formulación de la ley fuerte de los grandes números.
    - ❖ De Moivre, Laplace, Poisson, Tchebychef, Markov, Liapounoff, Lévy, … contribuyeron a la formulación del teorema central del límite, lo que condujo al planteamiento de las distribuciones muestrales y a la vinculación formal de las variables aleatoria y estadística.

5. Uso de resultados probabilísticos que favorecieron el análisis de datos, lo que fortaleció el análisis de la variable estadística.

    - ❖ Uso de la distribución normal y el uso de diversas herramientas probabilistas en el análisis de datos en diversas áreas científicas. Favoreció no sólo la generación de inferencias informales sino también de análisis propios de la estadística. Destacan los trabajos de Quetelet y Galton, aunque las técnicas se difundieron rápidamente en las últimas décadas del siglo XIX.

6. Teoría probabilística que ayudó a la formalización de ideas que surgieron en análisis de datos en contextos prácticos.

    - ❖ Pearson estableció un puente entre las ideas de Galton y resultados probabilísticos y matemáticos, favoreciendo la formalización de la inferencia estadística, así como la vinculación formal entre la variable aleatoria y estadística.

A lo largo de la historia las dos vías históricas también tuvieron largos periodos de reflexión propia, a través de los cuáles la interacción entre ellas fue casi nula. El desarrollo de la teoría de Laplace o la larga discusión que dio lugar a la comprobación del teorema central del límite (que tendría, sin embargo, una fuerte repercusión en la estadística) son momentos en los que la teoría prevalece sobre la práctica. Así mismo, en el análisis de grandes recuentos de datos a través de elementos meramente descriptivos, y sin embargo, con fines inferenciales, permaneció mucho tiempo sin la influencia de herramienta teórica que permitiera análisis más potentes.

7.3. Saltos cualitativos de las etapas históricas

También, dentro de las ocho etapas, se identifican saltos cualitativos en la conceptualización de la variable aleatoria, que marcaron su devenir histórico y que también contribuyeron al desarrollo de la probabilística y estadística como ciencias.

1. *Uso intuitivo de la variable estadística.* Los recuentos de los jugadores (Duque de Toscana) hasta estudios estadísticos en censos, estudio de causas de muerte o seguros de vida y análisis de mediciones astronómicas, impulsaron el desarrollo de análisis detallados de la variable estadística (Graunt, Arbuthnott, Halley, Galileo, Daniel Bernoulli). En muchos de estos estudios los datos condicionaron el análisis. La variable estadística surgía a partir de lo que el investigador podía «visualizar» de los datos, así como su análisis. La naturaleza numérica de la variable se impuso para poder hacer un mejor uso de la información recopilada.
2. *Uso intuitivo de la variable aleatoria.* La probabilidad en un contexto de teoría de juegos hasta problemas de aplicaciones como sobrevivencia, o teoría de errores (Pascal, Fernat, Huygens, de Moivre y Laplace), en donde no se hacía explícita la variable, pero se trabajaba con ella. El desarrollo teórico de la probabilidad es fuerte, sin embargo, a pesar de su uso, la variable aleatoria no tiene importancia explícita.
3. *Uso explícito, sin formalización, de la variable aleatoria.* Progresivamente la variable dentro del ámbito probabilístico adquiere identidad propia. Se percatan de la necesidad de una variable que establezca un puente entre el fenómeno aleatorio y los números reales. Comienza a ser usada como tal, sin que por ello vean la necesidad de definirla o formalizarla. En un inicio es tratada como resultado numérico de un fenómeno aleatorio, pero se llega a reconocer como variable vinculada al azar y se realizan operaciones básicas como objeto matemático (Tchebyshev, Liapunov).

4. *Vinculación informal entre las variables aleatoria y estadística.* Se extiende el uso de la herramienta estadística y probabilística en aplicaciones a problemas biológicos, poblacionales, económicos y sociales. La variable estadística surge directamente de los problemas y se establecen inferencias informales, sin un aparato teórico fuerte que las sustente (Quetelet, Galton).
5. *Axiomatización de la probabilidad.* Se llega a una formalización con base en la teoría de la medida y su reconocimiento como función medible (Borel y Kolmogorov). Se desarrolla la teoría formal de la variable aleatoria: sumas, productos, convergencia de sucesiones y series de variables aleatorias (Lévy, Petrov). Actualmente, la teoría probabilística continúa ampliando y enriqueciendo esta visión del objeto matemático variable aleatoria.
6. *Formalización de la inferencia estadística.* Se establecen las bases formales de la vinculación entre la variable aleatoria y la estadística a través de las distintas visiones sobre inferencia estadística y diseño de experimentos (Karl Pearson, Fisher, Egon Pearson, Neyman) y con ello la fundamentación de la estadística como ciencia. Actualmente la inferencia se ha convertido en una herramienta poderosa tanto para el fortalecimiento y fundamentación científica de diversas ciencias, como para la toma de decisiones prácticas en diversos ámbitos. Sin embargo, existen aún debates filosóficos sobre la mejor forma de vincular ambas variables dentro del ámbito de la estadística matemática.

7.4. Algunos usos posibles en la didáctica

Como Kolmogorov (1956) sostiene, la variable aleatoria fue uno de los dos conceptos[27] que desde el inicio de los tiempos se trabajaron continuamente y que sin embargo no se hicieron explícitos hasta finales del siglo XIX y principios del XX. De lo que se induce el largo tiempo que se requiere para su conceptualización. Una de las primeras recomendaciones didácticas de este estudio histórico debiera ser, entonces, el constante contacto del estudiante con la variable aleatoria desde los primeros niveles escolares hasta los universitarios.

Los niveles de formalización históricos de la variable aleatoria están marcados por los saltos históricos cualitativos, pero la permanencia de estas conceptualizaciones durante las ocho etapas encontradas no ha sido homogénea. Las primeras etapas duraron muchos años (y por lo tanto el uso intuitivo de ambas variables y, en su mayoría, sin contacto entre ellas), con cambios paulatinos, a diferencia de las últimas etapas que

[27] El otro concepto es el de independencia, muy vinculado al desarrollo de la variable aleatoria.

fueron densas por la aparición, no sólo de nuevos objetos matemáticos, sino también de nuevos significados para objetos que habían surgido desde tiempos antiguos, en un tiempo relativamente corto. Así, el primer contacto del estudiante con el concepto debiera partir de aplicaciones concretas y de manera intuitiva. La esperanza matemática podría ser un elemento que contribuyera a su manejo. Una etapa intermedia podría incluir un uso explícito, pero informal, en el que al estudiante se le pusiera en contacto con la relación entre el espacio muestral y la necesidad de una variable numérica. Un uso informal también podría provenir de problemas en donde la variable aleatoria se definiera a partir de fenómenos aleatorios con resultado numérico en los que ya se estableciera la función de densidad.

El largo tiempo que la humanidad ha requerido para la conceptualización de las variables aleatoria y estadística tiene varios motivos. Uno de ellos es el bagaje matemático tan fuerte que ha sido necesario para su formalización, pero también la aparente simplicidad de su manejo. Durante muchos siglos ambas variables fueron tratadas sin la formalidad con las que actualmente las conocemos (su constitución tiene menos de un siglo), sin que ello impidiera que surgiera una solución a los problemas que se pretendían resolver, pero una mejor solución implicaba herramienta matemática que debía ser generada o formalizada. La evolución de la variable aleatoria ha demandado nuevos objetos matemáticos que, en diversas ocasiones, fueron generados a partir del interés en ella, algunos de los cuales actualmente no forman parte del sistema axiomático de la probabilidad o de la estadística, sino de otras ramas de la matemática, como la función generatriz o algunas soluciones a ecuaciones diferenciales, y que, en ocasiones, son necesarios para la comprensión de ciertos niveles de formalización de la variable aleatoria o estadística. En otras situaciones, los elementos matemáticos necesarios surgieron en otras ramas de la matemática, como el cálculo infinitesimal, pero los significados de algunos de ellos tuvieron que ampliarse o transformarse para que pudieran formar parte del sistema axiomático de la probabilidad, como la conceptualización misma de *variable*.

Esto implica que es necesario identificar el tipo de herramienta matemática que el estudiante requiere para la comprensión del concepto en los distintos niveles. Hay objetos matemáticos que hay que esperar a que otros cursos de matemáticas se los proporcionen. Aun así, la historia muestra que la utilidad de la herramienta matemática y los procesos de modelación requieren ser resignificados en un ambiente de incertidumbre puesto que no tienen los mismos significados que en la matemática determinista.

Otro resultado que salta a la vista en este estudio histórico es que el crecimiento de la probabilidad y estadística en gran parte es debido a la interacción que se ha dado entre los datos y la teoría para la conformación de ambas variables en diferentes momentos de la historia. Ya vimos que este contacto entre ellas no se presentó en una sola etapa histórica, sino que se dio en diversos momentos y bajo una diversidad de formas (apartado 7.2). Además muestra que el estudio de la variable aleatoria no es exclusivo de la probabilidad, sino también de la estadística, puesto que las dos vías de surgimiento de este concepto, fueron las que dieron lugar a mayores aplicaciones de él. Así, no es recomendable una enseñanza de acuerdo a un sistema axiomático deductivo sino más bien de un constante contacto entre la teoría y los datos en diversos momentos del currículo, que quizá podrían coincidir con los momentos históricos en los que se dieron, pero mediando los lapsos de tiempo para que la interacción sea más constante, sin relativamente largos periodos de tiempo en los que no se estableció contacto.

7.4.1. Obstáculos posibles

Identificamos, también, algunos posibles obstáculos para el surgimiento y desarrollo del objeto variable aleatoria en la escuela:

- Se requirió de un largo periodo para pasar del estudio de la esperanza matemática de la variable aleatoria y del cálculo de la probabilidad de valores aislados hasta la enumeración de todos sus valores y probabilidades respectivas, esto es, el paso al estudio de la distribución de la variable. Ello podría explicar la dificultad de la idea de distribución para algunos estudiantes, que se centran en el estudio de valores aislados de la variable o sólo en sus valores medios, sin alcanzar una capacidad completa de manejo de distribuciones (Bakker y Gravemeijer, 2004). La solución de algunos problemas históricos, a diferencia del análisis de datos, proporciona un solo número sin la necesidad del análisis completo de todos los valores posibles de la variable.
- A pesar de que el interés radicaba principalmente en el conocimiento del comportamiento poblacional, durante largo tiempo el estudio de las variables estadísticas y su convergencia hacia la aleatoria fue simplemente descriptivo. Ello indica una posible dificultad de los estudiantes para ir más allá del estudio descriptivo de datos cuando trabajen con proyectos estadísticos. Pudiéramos relacionar esta dificultad con la que tienen algunos estudiantes para alcanzar niveles adecuados de complejidad en la lectura de datos, limitándose a la lectura directa sin pasar a la realización de extrapolaciones o inferencias (Shaughnessy, 2007).

- Durante periodos relativamente largos, el estudio de las variables aleatoria y estadística estuvo separado y se realizó desde diferentes comunidades. La variable aleatoria se introdujo dentro de la matemática motivada primeramente por problemas donde la asignación de probabilidad es a priori (clásica) y posteriormente por problemas de tipo matemático. El estudio de la variable estadística provino de problemas prácticos sin el planteamiento de un aparato matemático excesivo, en donde la asignación de probabilidad fue preferentemente frecuencial. La aparición progresiva de la inferencia hizo que se comenzara a relacionar más explícitamente las variables estadísticas y aleatorias. Este desarrollo sugiere que los estudiantes pudieran tener dificultad en conectar, por un lado, las aproximaciones clásica y frecuencial de la probabilidad y por otro, los objetos variable estadística y aleatoria. Posiblemente para hacer esta conexión se requiera plantear un problema de inferencia, al menos de un modo informal, como el que trataremos de implementar en el Estudio 4 para analizar esta conjetura.
- Durante un largo periodo, el énfasis en la preocupación por la fundamentación de la teoría de probabilidad, hizo que los matemáticos se enfocaran a la formalización de la variable aleatoria, olvidando la vinculación de ésta con el fenómeno aleatorio. Es posible que este mismo obstáculo también se presente en los estudiantes cuando, en la enseñanza escolar, se adentra en un concepto más formal de la variable aleatoria. Nardecchia y Hevia (2003) también alertan sobre este posible obstáculo. La aleatoriedad será uno de los elementos de observación en los estudiantes en los Estudios cognitivos 3 y 4.
- La formalización de la variable aleatoria tuvo que esperar a que se conformara la teoría de la medida, las funciones y el análisis matemático. Esta conexión se hizo a través de la definición de la función de distribución y que compone la inversa de la variable aleatoria con la probabilidad. Pensamos que el estudiante puede tener problemas en comprender la definición formal de variable aleatoria, en diferenciar una función de conjunto (variable aleatoria) de una función real (función de distribución) o en comprender la función de distribución como función compuesta. Analizaremos esta conjetura en el Estudio 3.
- Hay una marcada complejidad con sus operaciones y la dificultad para reconocer la existencia de la variable aleatoria al hacer composición entre variables aleatorias (por ejemplo, Bernoulli en la Ley de los grandes números). Históricamente se ha tenido que desarrollar la teoría de conjuntos y de la medida para plantear formalmente la variable aleatoria y abordar más profundamente sus

operaciones como la suma entre variables aleatorias, producto y convergencia de sucesiones y series de variables aleatorias, así como teoremas límite. En esta investigación no se plantean las operaciones con variables aleatorias, pero será necesario entender esta dificultad en el estudio de otros temas por parte de los estudiantes, por ejemplo, en las distribuciones muestrales.

- A lo largo de la historia se hubo una dificultad para comprender la aproximación de la variable continua a través de la discreta (paso de discreto a continuo). En esta investigación tampoco abordamos esta dificultad que fue descrita en el trabajo de Alvarado (2007), pero será importante cuando los alumnos estudien la aproximación de las distribuciones binomial o Poisson y el teorema central del límite.

Capítulo 5.

Estudio 3.

Diseño e implementación de una entrevista clínica

ÍNDICE DE CAPÍTULO

1. Introducción

Uno de los objetivos de nuestro trabajo es llevar a cabo análisis cognitivos de la comprensión intuitiva que muestran algunos estudiantes universitarios sobre la variable aleatoria y otros objetos matemáticos relacionados con ella en la resolución de problemas y proyectos que involucren a la variable aleatoria en situación de modelación. En este capítulo nos concentraremos en la descripción de la fundamentación, diseño, análisis y conclusiones obtenidos a propósito de la entrevista clínica, considerada una herramienta metodológica que proporciona información sobre los procesos cognitivos de los estudiantes. La intención es obtener elementos que permitan discernir sobre la forma en que los estudiantes se aproximan al objeto de estudio, así como elementos que nos acerquen a su análisis por medio de los argumentos, razonamientos, procedimientos y resultados que los estudiantes usen para resolver un problema en donde se ponga en juego a la variable aleatoria.

Este capítulo se inicia con la planeación de la entrevista clínica, que incluye todos los elementos considerados en el diseño del protocolo de investigación. Se comienza con los preliminares de la entrevista clínica (contexto escolar de la investigación y conocimientos esperados por la institución). Formularemos nuestros objetivos y algunas hipótesis generales sobre las posibles dificultades que surgen en estudiantes universitarios de los primeros semestres cuando abordan un problema estocástico relacionado con la idea de variable aleatoria, así como el atisbo de algunas trayectorias que siguen al querer resolver la actividad.

El protocolo del diseño de investigación se concreta en una actividad que, junto con las hipótesis, establecerán las bases del análisis de la entrevista. También se describe y analiza el problema base de la actividad, resaltando la forma en que se manifiesta la variable aleatoria en él. La parte central del capítulo corresponde a la presentación y análisis de los resultados obtenidos que se organiza de acuerdo a los objetos de análisis. Finalmente, en las conclusiones, se confrontan los resultados obtenidos con las hipótesis de investigación y se analiza la idoneidad de la actividad planteada como base del estudio. Esperamos que, del análisis de esos resultados, se obtengan algunas líneas que permitan profundizar en las vertientes epistemológicas, cognitivas y didácticas en futuras investigaciones sobre la variable aleatoria.

2. Preliminares

La entrevista clínica se realizó en el Campus Monterrey del ITESM, México, con estudiantes universitarios de Ciencias sociales recién ingresados a la institución. En este apartado hacemos explícito el contexto escolar de nuestra investigación, así como la profundización pretendida del objeto de estudio basándonos tanto en los análisis epistemológicos realizados anteriormente (Capítulos 3 y 4) como en la experiencia de los profesores participantes dentro de la institución.

2.1. Contexto escolar

A pesar de que Heitele (1975) considera que la intuición sobre las magnitudes relacionadas con el azar aparece en los niños antes que la concepción de experimento aleatorio, consideramos que el planteamiento formal de la variable aleatoria no es sencillo de asimilar espontáneamente ni de comprender en un primer contacto con el tema. Heitele plantea la necesidad de enseñarlo paulatinamente a través de un currículo en espiral a lo largo de la enseñanza formal de los estudiantes. Aunque nuestra investigación se ubica en el nivel universitario, hay una necesidad de conocer el desarrollo previo del concepto en el sistema escolar.

Actualmente los programas de estudio de nivel primaria (6 años) y secundaria (3 años) en México contemplan el estudio la probabilidad, de modo que es factible que durante ese periodo los estudiantes desarrollen ideas sobre las variables vinculadas aleatorias. Sin embargo, en el nivel bachillerato, que es donde podría esperarse que se trabajara este concepto de manera más formal, no está contemplado el tratamiento de la probabilidad y estadística de manera obligatoria, salvo en raras excepciones (como los bachilleratos del IPN y algunos tecnológicos regionales). En algunos otros planes de estudio es una materia optativa o un tópico (en la mayoría de las preparatorias de México afiliadas a la SEP o a la UNAM) y en pocos de estos planes de estudio se contempla el estudio de la variable aleatoria. En el caso de nuestra institución, el ITESM, el estudio de esta materia se restringe a un módulo de estadística descriptiva en el nivel bachillerato.

En la mayoría de las universidades el tema no se introduce hasta el primer curso de probabilidad y estadística, como preámbulo a las distribuciones de probabilidad. De modo que esperamos que los estudiantes recién ingresados a la universidad no muestren condiciones de formalización del concepto. Las prefiguraciones sobre el concepto del estudiante universitario de nuevo ingreso estarán condicionadas por sus experiencias vividas en los niveles básicos y en su vida propia.

2.2. La variable aleatoria en los cursos de estadística del ITESM

Desglosaremos brevemente las ideas y competencias que, sobre la variable aleatoria, se espera que un estudiante del nivel universitario tenga al egresar de una carrera en Ciencias sociales dentro del ITESM, campus Monterrey. Esta será nuestra referencia para diseñar el protocolo de la investigación en lo que se refiere a la profundidad del objeto variable aleatoria y los objetos y variables a analizar.

Uno de los libros de estadística más socorridos en el primer curso de estadística para Ciencias Sociales en el ITESM Campus Monterrey, define a la variable aleatoria como «una variable que toma un valor numérico único para cada uno de los resultados del espacio muestral de un experimento de probabilidad» (Johnson y Kuby, 2008, 270). Es usada como preámbulo para introducir las distribuciones de probabilidad, pero antes de hacerlo, ejemplifican su definición a través de una serie de problemas en donde los estudiantes tienen que identificar cuál es la variable aleatoria vinculada con el experimento y si es continua o discreta. A través de ellos, los autores pretenden esclarecer que la regla que vincula los valores numéricos con el espacio muestral está dada por el contexto del problema.

Observemos que el énfasis de esta definición está en su papel como «variable», en el sentido de que, como Borovcnik, Benz (1991a) y Johnson, R. y Kuby, P (2008) mencionan, es necesario conocer el conjunto de todos los posibles resultados y sus probabilidades asociadas para modelar un fenómeno estocástico. Esto es coherente con los objetivos del ITESM en donde se busca que los estudiantes hagan uso de la herramienta matemática en la modelación que, en el caso de nuestro curso, sería en la modelación del fenómeno aleatorio.

El papel de la variable aleatoria como «función» no está excluido, pero no es mencionado explícitamente. Se entiende, por tanto, que no se está interesado en que el estudiante profundice en la formalización del concepto ni en su definición matemática. Como se analizó en el Capítulo 3, el experimento aleatorio, el espacio muestral y los eventos asociados a él, están definidos en un espacio de probabilidad y la distribución de probabilidad vincula el aparato probabilístico con herramientas de análisis. Este aparato matemático es un tanto complicado por lo que no se trata en los cursos. Sin embargo en esta investigación lo tendremos presente pues nos permite pasar del trabajo matemático con conjuntos al trabajo matemático con números reales.

La aparente sencillez de la definición dada por Johnson y Kuby (2008) se desvanece cuando profundizamos un poco en ella y notamos que están involucrados muchos de los conceptos que están vinculados con la definición formal de la variable aleatoria y que se analizaron en el Capítulo 3. En la Figura 5.1 se muestra, de manera simplificada, la noción de variable aleatoria que, previo a esta investigación, esperamos de los estudiantes, y su relación con los conceptos con los que se vincula más estrechamente.

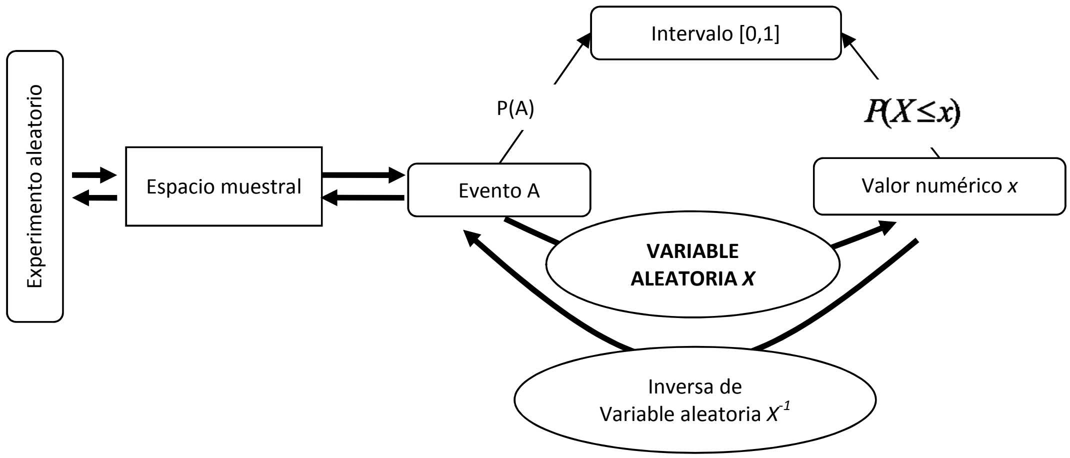


Figura 5.1. Esquema simplificado la noción de variable aleatoria esperada por los estudiantes que egresan del nivel universitario.

La variable aleatoria asigna un valor numérico a cada suceso (evento) del espacio muestral, lo que hace posible que se defina una función de distribución, $F(x)=P(X \leq x)$. Ésta caracterizará a la variable aleatoria vinculada a un experimento aleatorio en particular y concretará una relación funcional dentro de los números reales. A su vez, cada evento está definido en un espacio muestral de un experimento aleatorio. La variable aleatoria vincula al espacio muestral con un valor numérico definido en los reales, al mismo tiempo que la inversa de la variable aleatoria vincula el valor numérico con el experimento aleatorio (o con cada evento elemental). Es decir, la distribución se vincula con el experimento a través de operar con la variable aleatoria en sentido inverso. Se espera que los estudiantes puedan transitar de manera directa e inversa por todas estas relaciones que se establecen entre los conceptos vinculados, es decir, que pueda transitar fluidamente del contexto del problema al contexto matemático y viceversa.

3. Objetivos del estudio

Desde la perspectiva cognitiva, nos interesamos en conocer cómo emergen, a partir de una situación intuitiva para ellos, los elementos del significado de la variable aleatoria en estudiantes recién egresados del bachillerato durante la resolución de un problema en el que surge la necesidad de hacer uso del objeto matemático de estudio. Definimos los objetivos de esta entrevista clínica como sigue:

En primer lugar deseamos describir, en los estudiantes universitarios que inician un curso de estadística en ciencias sociales, los procesos de solución de la actividad propuesta cuando trabajan sin la ayuda directa del profesor. Queremos identificar las dificultades por las que atraviesan, y qué conocimientos, previos a la idea de variable aleatoria utilizan. También nos interesa explorar aquellos que los estudiantes desarrollan sobre nuestro objeto de investigación al resolver una situación problema en donde se pone en juego la variable aleatoria.

Nos preguntamos también por la idoneidad de la situación planteada, en cuanto permita que a partir de ella los alumnos desarrollen la idea de variable aleatoria y sea una situación motivadora.

4. Hipótesis del estudio

De acuerdo con las aportaciones de los análisis epistemológicos y con los preliminares se definieron los objetos matemáticos que queremos analizar en la entrevista y que determinarían el diseño del protocolo. Ellas también permitieron desglosar y concretar, a grandes rasgos, las hipótesis de investigación en el protocolo. Así mismo facultaron al investigador y al profesor para centrar preguntas, así como profundizar en las nociones que las estudiantes utilizaban o mencionaban.

La hipótesis que planeamos son generales en la medida que no tenemos suficiente información previa que nos permita definir hipótesis más específicas. El sustento teórico del método de investigación empleado (entrevista clínica) nos lo posibilita. Así, nos acercamos a los estudiantes con los siguientes juicios con respecto a su comportamiento ante la situación problema:

- *Falta de percepción de la aleatoriedad del proceso.*

  Concordamos, por el análisis histórico reportado en el Capítulo 4, con Nardecchia y Hevia (2003) acerca de la falta de percepción de la aleatoriedad durante el desarrollo histórico del concepto de variable aleatoria, que también podría presentarse en los estudiantes. Así mismo, otros autores reportan dificultades en los estudiantes en la percepción de aleatoriedad (Sección 6 del Capítulo 6). Nosotros esperamos que los estudiantes no se percaten de la vinculación entre la variable aleatoria y la aleatoriedad. Desde la perspectiva formal, esta vinculación se realizaría a través del espacio probabilístico del experimento aleatorio. Desde la perspectiva intuitiva sería la relación que se establece entre una variable numérica (la variable aleatoria) y la aleatoriedad del proceso que genera sus valores. En ese sentido, en los fenómenos aleatorios no se puede asegurar que se obtendrá el mismo resultado, no obstante se repita muchas veces el mismo experimento bajo las mismas condiciones.

- *Tendencia a «algebraizar» y descontextualizar los procedimientos relacionados con la noción de variable aleatoria.*

Esperamos que la componente aleatoria que se presenta en el problema que se trata de matematizar tienda a diluirse para dársele prioridad a los procedimientos matemáticos. La variable aleatoria permite la definición de las funciones de distribución y la distribución de probabilidad y por lo tanto, también el análisis de las situaciones probabilísticas a través de la introducción del análisis matemático. Sin embargo el uso de esta herramienta puede repercutir en que se descontextualice el aparato matemático de las situaciones que hacen surgir a la variable aleatoria y que en cambio se le dé mayor importancia a los procedimientos matemáticos. Es decir, el proceso mencionado por Borovcnik, Benz y Kapadia (1991a) que enfatiza en el papel de la variable aleatoria en la modelación, se reduciría al trabajo con el modelo matemático, olvidando la parte de generación del modelo y de la interpretación de sus resultados. Oseguera (1994) y Tauber (2001) reportan esta ausencia en su análisis de material de apoyo y de los textos universitarios.

- *Extrañeza de trabajar funciones en un contexto probabilístico.*

El estudio de las funciones normalmente se asocia con el estudio del cálculo y por lo tanto con la matemática determinística. Es de esperarse que los estudiantes asocien las gráficas y las ecuaciones de las distribuciones de probabilidad con esa parte de la matemática en la que han profundizado en los últimos semestres de su escolaridad. Por ello esperamos que les sea extraño trabajar con una función tan poco usual como lo es la distribución probabilidad, en el sentido de que no relaciona magnitudes físicas.

- *Dificultades con la noción formal de variable aleatoria.*

Heitele (1975) sugiere a la variable aleatoria como idea fundamental porque proporciona modelos explicativos en cada etapa de desarrollo del individuo. Otros autores (como Batanero, 2001; Miller, 1998 y Ortiz, 2002) se unen a su propuesta de enseñarla desde los primeros niveles escolares, sin embargo no especifican los grados de dificultad con los que se tiene que enseñar en los distintos niveles escolares. Podemos esperar que en nuestra investigación los estudiantes tengan un manejo apropiado de la variable aleatoria de manera intuitiva y que en cambio tengan problemas al formalizarla al nivel que se pretende en el nivel escolar que se encuentran (ver punto 2.1 de este capítulo).

- *Dificultades en la comprensión de los conceptos que intervienen en la noción formal de variable aleatoria.*

De acuerdo a la forma en que se pretende introducir el objeto de estudio, esperamos dificultades en nociones tales como experimento aleatorio, probabilidad, la variable aleatoria y su distribución de probabilidad.

5. El protocolo de investigación

En este apartado describiremos y analizaremos el protocolo de la investigación como guía que orienta y dirige la entrevista. En esta investigación se diseñó en forma de cuestionario que girara alrededor de un problema, de manera tal que la discusión se suscitara a partir de la resolución del problema. Las preguntas, planeadas y definidas por los objetos de análisis y las hipótesis descritas, profundizan en las concepciones de los estudiantes acerca de los objetos matemáticos que emplean y en las estrategias usadas para resolver el problema.

En las siguientes secciones, primero describiremos la estructura del protocolo y el análisis del problema seleccionado para facilitar la comprensión de los objetos de análisis y la forma en que se tomaron en cuenta en el protocolo.

Para realizar la exploración cognitiva se consideró apropiado tomar como base el problema que Castillo y Gómez (1998) proponen para introducir el tema de variable aleatoria, porque permitiría la introducción natural de los objetos matemáticos tenidos en cuenta en el de diseño. Alrededor del problema se diseñó un cuestionario con preguntas abiertas, que, a su vez, guía la solución del problema y conduce a los entrevistadores en la exploración de los conocimientos de los estudiantes. La actividad así diseñada quedó conformada por tres secciones, cada una con un objetivo específico:

PARTE I: Introducción al problema. En esta parte se introduce el problema, se cuestiona a los estudiantes acerca de su comprensión de la probabilidad teórica clásica (dada por la regla de Laplace); se cuestiona la relación entre probabilidad y los eventos y se pide la definición de las variables que intervienen.

PARTE II: El problema. Se les pide a los estudiantes que tomen una decisión, y que sustenten su posición. Para ello deberán exponer sus ideas acerca de riesgo y de probabilidad acumulada. También son cuestionadas sus concepciones acerca de espacio muestral y experimento aleatorio.

PARTE III: Cambio de registro de tabular a gráfico. También se explora su concepción de variable aleatoria. Al final se pide una comparación entre situaciones estocásticas y deterministas.

A continuación describimos el problema utilizado alrededor del cual se establece el protocolo y se genera la actividad, su solución y cómo vive la variable aleatoria en él. El protocolo completo se incluye en el Anexo 1.

5.1. El problema

El problema trabajado en la entrevista es el siguiente (Castillo y Gómez, 1998, p.120):

A raíz de los festejos del día del niño, el departamento de relaciones públicas de una fábrica desea efectuar una rifa que beneficie a los hijos de los trabajadores. Al obrero ganador se le premiará con boletos para que vaya al teatro con su familia completa. La intención es que la rifa beneficie a los hijos de los trabajadores.

Así que encomiendan a la trabajadora social de la empresa que decida cuántos boletos tiene que comprar para que se efectúe la rifa de manera exitosa, pero la empresa pierda lo menos posible. La distribución del número del número de hijos por trabajador se proporciona en la Tabla 5.1.

Tabla 5.1. Frecuencia del número de hijos de cada trabajador

Número de hijos	0	1	2	3	4	5	6	7	8	9
Frecuencia de trabajadores con un número dado de hijos	16	22	33	45	31	20	12	9	7	5

El análisis que realice el alumno debe estar enmarcado en el contexto de esa fábrica y debe tomar en cuenta que es un festejo infantil y que la empresa no desea despilfarrar recursos. Aunque no es una condición explícita, esperamos los estudiantes consideren que al momento de la rifa se tengan los boletos para ser entregados al trabajador ganador.

La dificultad de este problema estriba en que en el momento de comprar los boletos, no se sabe cuántos hijos tiene la familia del obrero que saldrá ganador, puesto que no se sabe quién será. La solución oscila entre dos criterios de decisión:

- Por un lado la empresa debe tener boletos suficientes para que el trabajador ganador tenga posibilidades de llevar a toda su familia a la función de teatro.
- Por otro lado, la empresa desea optimizar sus recursos. No quiere que sobren boletos.

Podemos suponer que todos los trabajadores de la fábrica son casados, así que el número de miembros de la familia del trabajador estará determinado por el número de hijos que éste tenga. Así el número mínimo de miembros de una familia de esa fábrica será 2 (la pareja) y el número máximo de miembros será 11 (9 es el número máximo de hijos de un trabajador de esa fábrica). Los recursos gastados por la empresa estarán dados por el número de boletos que se compren, puesto que no nos proporcionan el precio de los boletos. Para facilitar el razonamiento del problema supondremos que tanto los boletos de los niños como los de los adultos cuestan lo mismo.

La solución puede irse a uno o a otro extremo entre los dos criterios de decisión, pero el objetivo es tomar en cuenta los dos criterios y crear un ambiente de incertidumbre. Si se deja de tomar en cuenta uno de los dos criterios, las soluciones triviales y determinísticas son las siguientes:

- *Se comprarían 11 boletos* si se quiere asegurar que el trabajador ganador lleve a toda su familia: dos para los papás y la cantidad máxima de hijos posible para un trabajador de esa fábrica (9).
- *Se compran 2 boletos* para optimizar al máximo los recursos de la empresa, puesto que todas las familias constan de al menos dos miembros: la pareja.

Pero la empresa quiere realizar esa rifa optimizando el dinero que invertirá y al mismo tiempo quiere quedar bien con sus trabajadores. El estar entre los dos criterios de decisión introduce la incertidumbre y por lo tanto la necesidad de hacer uso de la variable aleatoria y de la herramienta matemática descrita en la Figura 5.4. La aleatoriedad se introduce en la decisión cuando relacionamos los boletos a comprar con el número de hijos del trabajador extraído de la urna y no con el máximo o mínimo número de hijos que tienen los trabajadores de la fábrica porque éstos últimos resultados son fijos (deterministas). Al ubicar la decisión en un contexto de incertidumbre es factible analizar el problema desde la teoría de la probabilidad.

Este problema forma parte de un campo de problemas en donde surge la necesidad de tomar una decisión en una situación aleatoria con un cierto criterio. Por ejemplo en un restaurante o en una compañía de autobuses o aviones, en donde la decisión de cuánta comida hacer o cuántos viajes programar, el criterio es brindar un buen servicio, optimizando los recursos. Lo aleatorio está en no saber el número de clientes que solicitarán el servicio y en tener que planear con anticipación.

5.2. Análisis del problema como situación de exploración

La solución del problema requiere del análisis de la distribución de probabilidad del número de hijos del trabajador premiado y de su función de distribución acumulada. Ambas se proporcionan en la Tabla 5.2.

Tabla 5.2. Distribución de probabilidad y función de distribución del número del número de hijos del trabajador premiado

Número de hijos (x)	Frecuencia	Distribución de Probabilidad	Función de distribución
		Probabilidad de que el trabajador premiado tenga x hijos $p(x) = P(X = x)$	Probabilidad de que el trabajador tenga x hijos o menos $F(x) = P(X \leq x)$
0	16	0,080	0,080
1	22	0,110	0,190
2	33	0,165	0,355
3	45	0,225	0,580
4	31	0,155	0,735
5	20	0,100	0,835
6	12	0,060	0,895
7	9	0,045	0,940
8	7	0,035	0,975
9	5	0,025	1,000

Hay que resaltar que el problema no tiene solución única. De acuerdo con la tabla, la probabilidad de que el trabajador seleccionado tenga tres hijos o menos es mayor que ½, es decir, mayor que la probabilidad de tener 4 ó más hijos. Una posible decisión es que la trabajadora social compre 5 boletos (2 para la pareja y 3 para los hijos), de esa manera la directiva de la empresa tendrá una probabilidad de 0,58 de que los boletos le alcancen al trabajador premiado y de que sobren los menos posibles. Si se quisiera aumentar la probabilidad, por ejemplo, a 70% de tener suficientes boletos, habría que tomar un boleto más. Admitiremos como correctas cualquiera de las dos respuestas.

Además, hay otros objetos matemáticos que aparecen en el problema y que interesan en exploración cognitiva, como el experimento aleatorio, el espacio muestral, la magnitud de interés, los posibles valores de la magnitud en el experimento, la partición que define en el espacio muestral y la variable aleatoria con sus tres componentes (dominio, regla de correspondencia y rango). Respecto a estos objetos, aparte de la comprensión que los estudiantes puedan tener, nos interesamos por el modo en que los definan dentro del contexto del problema (ver Tabla 5.3).

Tabla 5.3. Objetos matemáticos de interés presentes en el problema y su significado en la situación problema

Objeto matemático	Significado en la situación
Experimento aleatorio	Mezclar en una tómbola papeles con el nombre del trabajador y sacar al azar el nombre de uno de ellos, que será el ganador. O bien: Rifar boletos de entrada al teatro para la familia de un trabajador entre los trabajadores de una cierta fábrica de interés.
Espacio muestral	Nombres de los trabajadores que laboran en esa fábrica. Hay 200 sucesos elementales equiprobables.
Magnitud de interés	Número de hijos de los trabajadores.
Variable aleatoria	*Dominio*: Sucesos elementales en el experimento, trabajadores factibles de que reciban el premio.
	Regla de correspondencia: Número de hijos que tiene cada trabajador de la fábrica.
	Rango (valores que toma): números 0, 1, 2, 3, 4, 5, 6, 7, 8, 9.
Partición del espacio muestral	Conjunto de trabajadores que tienen el mismo número de hijos.

La forma en que estas nociones se vinculan con la variable aleatoria en el problema se muestra en la Figura 5.2, en donde se representan dos planos que se retroalimentan y definen la recomendación en el contexto del problema con argumentos matemáticos. En el superior se analiza el problema desde el contexto del sorteo en una fábrica. En la parte inferior se analiza la perspectiva matemática del problema.

A continuación describimos la forma en cómo se refleja este proceso en la Figura 5.2. El modelo matemático está condicionado por la magnitud de interés del problema. Para responder la pregunta planteada es necesario hacer un razonamiento regresivo, es decir, comenzando por lo que queremos alcanzar (Polya, 1965). Supongamos que ya compramos los boletos y que ya efectuamos la rifa, cuyo resultado sería el nombre del trabajador ganador, nos preguntaríamos si le alcanzarán los boletos que ya compramos. La característica de ese trabajador que hace que se responda la pregunta es cuántos hijos tiene. Esa característica permitirá agrupar el espacio muestral en eventos compuestos de interés y asignar probabilidades a esos eventos. Al mismo tiempo, define la variable aleatoria que asignará un valor numérico a cada evento elemental del espacio muestral. Notemos que los eventos simples en el experimento aleatorio son los trabajadores de la fábrica.

Observemos la cercanía entre la característica de interés y la variable aleatoria. No obstante hacemos énfasis en la existencia de las dos porque la característica de interés podría ser otra, por ejemplo, si el trabajador premiado es hombre o mujer o el medio de transporte que usa para llegar a su trabajo. Sin embargo, en un primer acercamiento, nos limitamos a variables cuantitativas que transforman el resultado del experimento aleatorio en un número. La característica de interés (que en este caso es directamente una magnitud), es la que define la partición del espacio muestral. Por supuesto, la partición del espacio muestral y la variable aleatoria tienen que estar vinculadas.

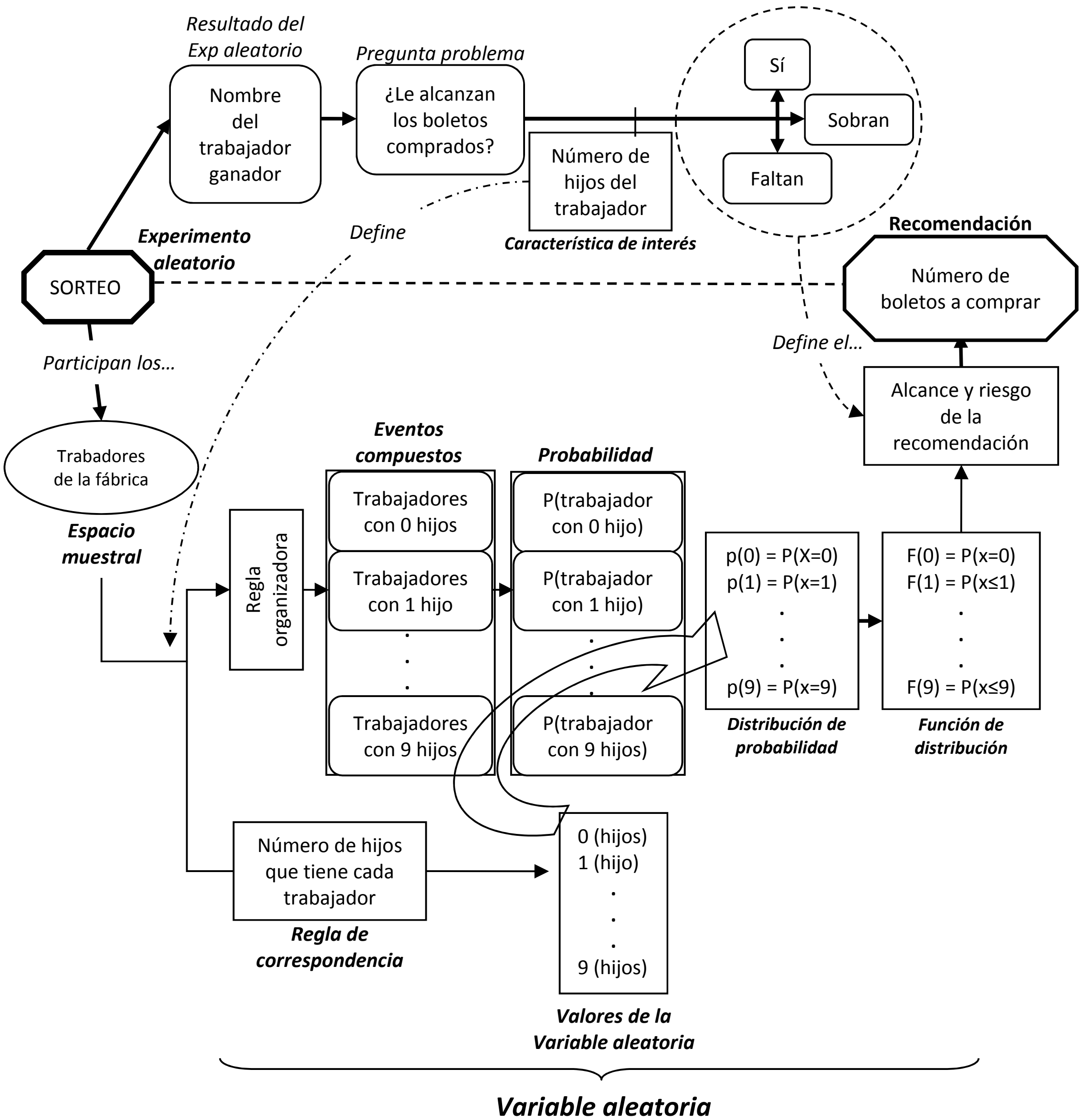


Figura 5.2. Representación de la relación entre los elementos teóricos vinculados con la variable aleatoria y la solución del problema

Una vez definida la variable aleatoria y sus valores, la distribución de probabilidad se obtiene al aplicar, para cada posible valor de la variable aleatoria, la función inversa de la variable aleatoria y asignarle la probabilidad del subconjunto del espacio muestral que toma dicho valor. Esto sólo es viable porque existe una probabilidad definida en el espacio muestral; es decir porque tenemos definido un espacio probabilístico. Los pasos siguientes en la solución del problema serán hacer uso de las distribuciones de probabilidad y de la función de distribución para proporcionar una solución de acuerdo al riesgo que decida asumir el alumno.

En el diagrama (Figura 5.2) se observa como la variable aleatoria está condicionada por el contexto del problema, pero hace posible la utilización de herramienta matemática para cuantificar el riesgo de la solución que se proponga. Una vez comprendido y delimitado el problema, la regla de correspondencia de la variable aleatoria surge de manera natural. El que la variable aleatoria vincule el espacio muestral con valores numéricos, hará posible que ese modelo matemático sea la distribución de probabilidad y/o la función de distribución, aunque la variable aleatoria en sí misma sea un modelo matemático en otro estrato de la estructura de la modelación que propone Heitele (1975).

5.3. Respuestas y soluciones esperadas en el protocolo

En este apartado proporcionaremos las respuestas esperadas a lo largo de la actividad propuesta a los estudiantes. El protocolo completo se incluye en el Anexo 1.

Parte I del Protocolo

a) *Si tomamos un trabajador al azar, ¿cuál es la probabilidad de que tenga 4 hijos? ¿Y 6 hijos? ¿Y 7? ¿Cuál es la probabilidad de que no tenga ningún hijo?*

Hay tres posibles respuestas correctas, todas ellas siguiendo el razonamiento de la probabilidad clásica (o la regla de Laplace). La diferencia está en el uso de fracciones, porcentajes o números decimales.

- De 4 hijos: $\frac{31}{200}$; de 6 hijos: $\frac{12}{200}$; de 7 hijos: $\frac{9}{200}$; de ningún hijo: $\frac{16}{200}$.
- De 4 hijos 0,155; de 6 hijos: 0,06; de 7 hijos: 0,045; de ningún hijo: 0,08.
- De 4 hijos 15.5%; de 6 hijos: 6%; de 7 hijos: 4.5%; de ningún hijo: 8%.

b) *En la distribución de probabilidad, ¿Qué depende de qué? ¿El número de hijos depende de la probabilidad o al revés? ¿Cuál de estas dos variables sería la variable dependiente y cuál la variable independiente?*

Esperamos que el estudiante responda que la probabilidad depende del número de hijos, por lo tanto el número de hijos es la variable independiente y la probabilidad es la variable dependiente en la distribución de probabilidad.

Parte II del Protocolo

En primer lugar, se pide al estudiante el cálculo de la probabilidad de que el trabajador premiado tenga el número de hijos dado. Esperamos que produzca una tabla similar a la Tabla 5.4, obteniendo la distribución de probabilidad y función de distribución de la variable, para auxiliarse en sus cálculos. Aunque también podría simplemente limitarse a calcular valores aislados de estas funciones. Esta misma tabla se puede llenar con las probabilidades en forma de porcentaje. La probabilidad acumulada, no se pide explícitamente en la actividad, sin embargo, la solución se facilita si se calcula.

Tabla 5.4. Probabilidad de que un trabajador tenga un número dado de hijos (distribución de probabilidad y función de distribución)

Número de hijos	0	1	2	3	4	5	6	7	8	9
Frecuencia de trabajadores	16	22	33	45	31	20	12	9	7	5
Probabilidad	$\frac{16}{200}$	$\frac{22}{200}$	$\frac{33}{200}$	$\frac{45}{200}$	$\frac{31}{200}$	$\frac{20}{200}$	$\frac{12}{200}$	$\frac{9}{200}$	$\frac{7}{200}$	$\frac{5}{200}$
	0,080	0,110	0,165	0,225	0,155	0,100	0,06	0,045	0,035	0,025
Probabilidad acumulada	0,080	0,190	0,355	0,580	0,735	0,0835	0,895	0,940	0,975	1

a) *¿Qué recomendación darías a la trabajadora social de cuántos boletos comprar para que la empresa pierda lo menos posible? Explica detalladamente porqué le darías esa recomendación.*

Esperamos que el estudiante deduzca que lo más probable es que el trabajador seleccionado tenga tres 3 hijos o menos, así que es necesario que se compren tres boletos para los niños más los dos boletos del trabajador y su esposa es decir, 5 boletos en total, de ese modo, la mayoría de los trabajadores (el 58%) no tendría problemas para llevar a sus hijos al teatro si salen premiados, al mismo tiempo que la empresa no perdería demasiado si sobran boletos. Hay otras soluciones posibles, como elegir un valor de la variable mayor, pero esperamos que el alumno elija esta, ya que en este problema la mediana es muy próxima a la moda, aunque la media da un valor algo mayor (3,4) se acerca más al valor 3 que al 4. Por tanto todos los promedios coinciden aproximadamente en este contexto.

b) De acuerdo a tu recomendación ¿con qué probabilidad el trabajador premiado ocuparía exactamente el número de boletos que compró la trabajadora social?

Suponiendo que la respuesta óptima es comprar 5 boletos, la respuesta esperada en este apartado es que con una probabilidad de 0,225, es decir, en realidad es poco probable que se ocupen exactamente los 5 boletos. Si su respuesta óptima es otra solución habría que ajustar la respuesta esperada.

c) ¿Qué tanto puedes asegurar que se ocuparan exactamente los boletos que se compraron?

De nuevo esta respuesta está condicionada por la dada en el inciso a. En el supuesto de que el número de boletos a comprar sea 5, es poco probable, es decir es más probable que no se ocupen exactamente puesto que sólo 45 de los trabajadores (de los 200) tienen oportunidad de ocupar exactamente el número de boletos recomendado. Es más alta la posibilidad de que sobren boletos (0,355) o bien de que falten (0,42).

d) De acuerdo a tu recomendación, ¿con qué probabilidad al trabajador premiado le alcanzarán los boletos y hasta le sobrarán?

Si la respuesta dada es la que hemos supuesto como óptima, la probabilidad sería de 0,58, es decir, es más probable que alcancen y hasta sobren los boletos a que hagan falta boletos. Esto significa que de 100 trabajadores, 58 tendrían posibilidades de no tener problemas con sus hijos en caso de que salgan premiados.

e) De acuerdo a tu recomendación, ¿con qué probabilidad al trabajador premiado no le alcanzarán los boletos para poder llevar a todos sus hijos al teatro?

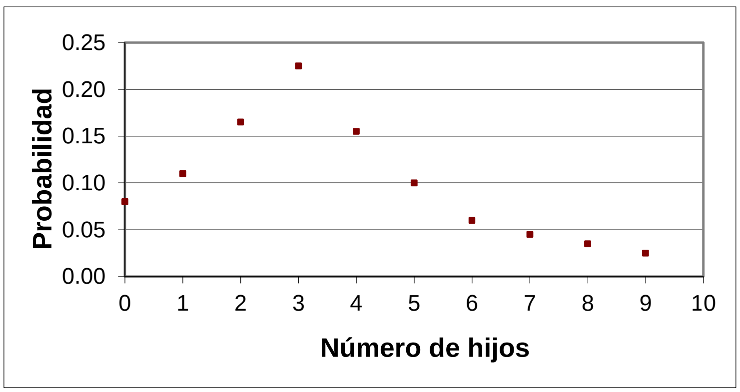
Bajo el mismo supuesto de que la respuesta son 5 boletos, la probabilidad es de 0,42. El 0,42 puede provenir de dos operaciones: la suma de 0,155, 0,1, 0,06, 0,045, 0,035 y 0,025 ó bien de la resta de 1 menos 0,58.

Parte III del Protocolo

En primer lugar se pide al alumno producir una gráfica que represente la probabilidad de que un trabajador con un cierto número de hijos salga premiado. Hay dos posibles gráficas correctas, que se presentan en la Figura 5.3. Inmediatamente después se pide contestar los siguientes incisos.

a) Representa con una letra la variable dependiente y con otra la variable independiente y utiliza estas letras para representar en notación funcional el resultado obtenido en el inciso b) de la parte II.

El resultado del inciso b) de la parte dos se refiere a con qué probabilidad el trabajador premiado ocuparía exactamente el número de boletos que compre la trabajadora social, por lo tanto el número de hijos será de 3, *x = 3*, *y = 0,225*, en notación funcional: *f(3) = 0,225*. Esperamos que los alumnos sugieran una notación aproximada como la siguiente: *x*: variable independiente (número de hijos), también podría ser *n* o cualquier otra letra; *y*: variable dependiente (probabilidad) o *P* o cualquier otra letra. En notación funcional: *y* = *f*(*x*) (utilizando las letras que escogieron para denotar a las variables).



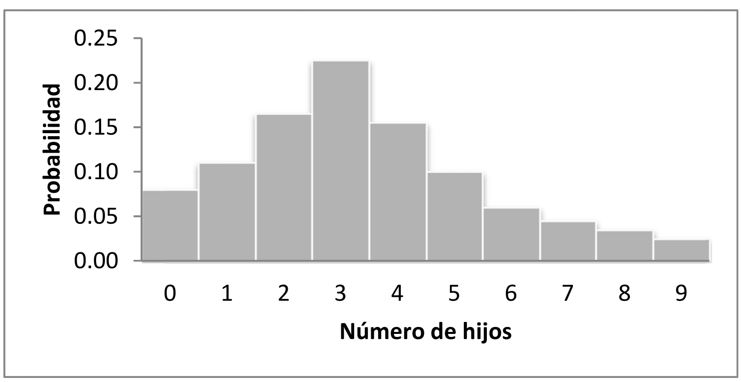


Figura 5.3. Graficas posibles de la distribución de probabilidad del número de hijos de los trabajadores

b) Describe, en el contexto del problema, con tus propias palabras la información que proporciona la variable dependiente de la variable independiente (en la función que acabas de representar).

La variable dependiente indica la probabilidad de que el trabajador que salga premiado tenga ‘*x*’ número de hijos. También se puede expresar como las posibilidades que hay de que un trabajador que tenga un número ‘*x*’ de hijos sea el premiado. Nótese que lo que se espera es que regresen al contexto del problema.

c) *Describe el dominio de la variable independiente y el rango de la variable dependiente.*

En esta pregunta se suscita un conflicto con la teoría. Desde la perspectiva estricta de la teoría de probabilidades, la distribución de probabilidad toma valores de [0,1] en su rango y valores enteros comprendidos entre $[0,\infty)$ en su dominio. Esto es porque la probabilidad definida en la distribución de probabilidad es una extensión de la probabilidad definida en el espacio muestral (en realidad no son iguales). En este caso no esperamos que los estudiantes respondan de ese modo, puesto que no han tomado un curso de teoría de probabilidad sino que esperamos que interpreten en el contexto de una variable discreta, tomando en cuenta solamente los valores de la probabilidad que le asignaron a los valores de la variable aleatoria que estuvieron trabajando. Esperamos descripciones semejantes a las siguientes:

- Dominio D = $\{x \mid x \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}\}$
- Rango R = $\{y \mid y \in \{0{,}025, 0{,}035, 0{,}045, 0{,}06, 0{,}08, 0{,}1, 0{,}11, 0{,}155, 0{,}165, 0{,}225\}\}$

d) *Describe el método a partir del cual puedes obtener el valor de la variable dependiente, a partir del valor de la variable independiente.*

Las respuestas pueden ser diversas, dos que consideramos correctas serían:

- El valor de la probabilidad (variable dependiente) se obtiene la frecuencia (casos favorables) correspondiente al número de hijos de interés (variable independiente) entre 200 (casos posibles). Para conocer la frecuencia correspondiente se recurre a la tabla dada.
- También se puede obtener directamente a partir de la tabla o de la gráfica.

e) *La relación que estableciste ¿es una función matemática? ¿por qué?*

Esperamos que el estudiante responda que esta relación sí es una función matemática porque a cada valor de la variable independiente (número de hijos) le corresponde un solo valor de la variable dependiente (probabilidad).

f) *¿Qué diferencias encuentras entre la variable independiente de esta distribución de probabilidad y las variables independientes de las otras matemáticas que tú conoces, por ejemplo, de álgebra o cálculo?*

Esperamos que el estudiante se percate que en la función (distribución de probabilidad) está vinculada a un proceso aleatorio. En particular, que el orden en que aparecen los valores de la variable independiente en el eje de las abscisas de su gráfica no está vinculado con el resultado del experimento aleatorio. Los valores de la variable independiente (número de hijos) no necesariamente se presentarían uno tras otro si repetimos una y otra vez la rifa en las mismas condiciones. Más aún, algunos valores es posible que no se presenten después de varios intentos. En este tipo de experimentos no se puede predecir el valor que se obtendrá ni siquiera de manera aproximada, sólo podemos saber qué es lo más probable que ocurra, mientras que en los problemas tradicionales de álgebra y cálculo hay un orden de aparición de los valores de las variables.

g) *¿Qué nombre le darías a este tipo de variable?*

Esperamos que el estudiante vincule el nombre de la variable con el fenómeno aleatorio. Podría darle el nombre de ‘aleatoria’ o variable estocástica o variable azarosa u otro nombre similar en el que se haga énfasis en que la aleatoriedad o la incertidumbre están vinculados con ella.

6. Objetos de análisis en la entrevista

Basándonos en los objetivos y el diseño del protocolo de investigación que acabamos de analizar, en este punto se describirán los objetos matemáticos analizados en la entrevista, indicando en cada uno de ellos los procedimientos y la comprensión del estudiante que deseamos explorar. Estos objetos son aleatoriedad, probabilidad, variable aleatoria y solución del problema.

6.1. Aleatoriedad

En relación a este objeto incluiremos tanto la incertidumbre, en el sentido coloquial de la imposibilidad de predecir con exactitud por la intervención del azar, como la aleatoriedad propiamente, como un atributo matemático establecido por la sucesión de resultados de un experimento impredecible realizado repetida e independientemente, que posee, además, la factibilidad de poder ser analizado a través del cálculo de probabilidades (Batanero, 2001).

La aleatoriedad quedará definida por el mismo contexto del problema, de modo que esperamos no haya cuestionamientos a este respecto. Es de interés la percepción del alumno sobre la aleatoriedad en la medida que esperamos que influya sobre su comprensión de la variable aleatoria. Sin embargo no será para nosotros tan importante la formalización que sea capaz de hacer de la aleatoriedad, como su noción intuitiva, en el sentido de la intervención del azar, y la relación que establezca entre la incertidumbre y su solución con la variable aleatoria. De esta forma, dentro de esta categoría nos interesamos por dos puntos:

- ❖ *La aleatoriedad en la situación problema*. Atendiendo a la percepción de la aleatoriedad, interesa la forma en que el estudiante la tiene en cuenta y cómo justifica con un razonamiento estadístico su recomendación. También es de provecho observar las dificultades en este proceso.
- ❖ *Relación entre la incertidumbre y la probabilidad*. Refiriéndonos a cómo el estudiante relaciona estos conceptos, identifica y redefine la incertidumbre con la distribución de probabilidad e interpreta una representación matemática de un problema en el que interviene la incertidumbre.

En la Tabla 5.5 presentamos los objetos matemáticos ligados con el objeto aleatoriedad en la primera columna y en la segunda los elementos que queremos evaluar.

Tabla 5.5. Evaluación de la comprensión de la aleatoriedad

Objeto matemático	Desglose	Evaluación en la actividad
Aleatoriedad en la situación problema	Concepción de experimento aleatorio La intervención de la incertidumbre dentro de una toma de decisión	Presentación de un fenómeno aleatorio Interpretación de la ocurrencia de eventos
Incertidumbre y probabilidad	Incertidumbre y probabilidad: Alcance de la decisión Medida del riesgo Incertidumbre y la distribución de probabilidad involucrada la de solución del problema	Definición de la distribución de probabilidad representativa de la situación Resolución de un problema ligado a una toma de decisión: ¿Cómo se fundamenta la decisión? ¿Qué tanto se puede asegurar que la decisión será certera?

6.2. Probabilidad

Dentro de este rubro la probabilidad será importante en la medida que el alumno la relacione con la variable aleatoria, de manera que llegue a establecer y cuestionar los componentes que definen a la distribución de probabilidad como función. Unimos a esto una exploración de cómo intervienen los significados que asocia a la probabilidad en la apropiación del concepto de la variable aleatoria por parte del estudiante.

- *La probabilidad en el espacio muestral.* Se busca esclarecer la forma en que el estudiante de manera intuitiva asigna la probabilidad.

- *La distribución de probabilidad como función.* Nos interesa cómo el estudiante asigna la probabilidad a cada valor de la variable aleatoria y establece la distribución de probabilidad. También, una vez establecida la representación matemática de la situación, trataremos de analizar la interpretación que el estudiante da a la distribución de probabilidad para contextualizar el problema.

Tabla 5.6. Evaluación de la comprensión de la probabilidad

Objeto matemático	Desglose	Evaluación en la actividad
La probabilidad definida en el espacio muestral	Interpretación de la situación problema en términos del modelo probabilístico: Definición del espacio muestral y de los eventos posibles Interpretación de las probabilidades de los eventos posibles Uso de la regla de Laplace Uso de los axiomas de probabilidad.	Cálculo de la probabilidad de eventos aislados en el experimento aleatorio Interpretación de la probabilidad de un evento
La distribución de probabilidad como función	Distribución de probabilidad como función asociada a una variable aleatoria Valores posibles de la variable aleatoria Asignación de un valor de probabilidad a los valores de la variable aleatoria La distribución de probabilidad como una función y su interpretación en un contexto matemático: su notación funcional, su dominio y su rango y regla de correspondencia Relación entre los contextos gráfico, numérico y algebraico La distribución de probabilidad como función compuesta, uno de cuyos términos es la variable aleatoria Noción de función: Qué es función Representación gráfica y algebraica	Cálculo de la probabilidad puntual y acumulada asociada a los valores de la variable aleatoria Organización de la información en una tabla Interpretación de la distribución de probabilidad y de la función de distribución acumulada Identificación de las variables que intervienen en la modelización del fenómeno aleatorio: ¿Cuál de ellas es la variable independiente en cada una de las funciones? ¿Cuál la dependiente? Las variables que intervienen en el problema: ¿Estamos ante una función real? La gráfica de la función Definición de los elementos de la función: nomenclatura, dominio, y rango

El significado de probabilidad seleccionada en la actividad será la vinculada a la acepción clásica porque esperamos que sea la noción con la que los estudiantes de nuevo ingreso a la universidad estén más familiarizados. Es de nuestro interés que la actividad explore la interpretación y el manejo de la probabilidad por los estudiantes así como la forma de cálculo de esta probabilidad en tres niveles principales: a un nivel intuitivo en eventos aislados (dentro de la situación); a un nivel intuitivo en la situación conjunta y su vinculación con la distribución de probabilidad; y a un nivel más formal en la distribución de probabilidad. Los procedimientos y concepciones a evaluar en este rubro se presentan en la Tabla 5.6.

6.3. Variable aleatoria

Es de esperar que los estudiantes no manejen la definición formal de variable aleatoria, de manera que nuestra intención es centrar la actividad en la exploración de su noción intuitiva y una posible apropiación de la noción con un nivel de profundidad descrito en el Apartado 2.2, que es el esperado en los cursos de estadística para Ciencias Sociales en el ITESM.

Es de nuestro interés la exploración de la comprensión intuitiva de la variable aleatoria en la solución el problema, así como en el nivel de dificultad que involucra su tratamiento formal tanto en relación con la distribución de probabilidad, como con respecto a ella misma definida como la función que relaciona el espacio muestral con un valor numérico. También nos interesa conocer la nomenclatura y el lenguaje que emplean para referirse tanto de la variable aleatoria como de la distribución de probabilidad. El diseño contempla los siguientes cuatro puntos y que se desglosan detalladamente en la Tabla 5.7:

- ❖ *La variable aleatoria en la situación problema.* Estudiamos el vínculo que establece el estudiante entre los valores de la variable aleatoria y la probabilidad definida en el experimento asociado en el contexto del problema. Este vínculo estará relacionado con la forma en que el estudiante maneje la probabilidad y la variable aleatoria para encontrar la solución del problema.
- ❖ *La variable aleatoria como función.* Nos interesa profundizar sobre la forma en que el estudiante maneja la variable aleatoria como función con sus tres componentes: la regla de correspondencia, sucesos del espacio muestral y los valores numéricos vinculados a cada suceso.

Tabla 5.7. Evaluación de la comprensión de la variable aleatoria

Objeto matemático	Desglose	Evaluación en la actividad
La variable aleatoria en la solución del problema	Manejo de la variable aleatoria con miras a la toma de decisión recomendada: Uso de la regla de correspondencia Uso de la variable aleatoria y su distribución (o de su función de distribución acumulada) para resolver el problema	Definición del espacio muestral y de los eventos posibles asociados a la característica de interés dentro del problema Cálculo e interpretación de la probabilidad puntual y acumulada asociada al problema.
Variable aleatoria como función sobre el espacio muestral	Vinculación de la pregunta de interés con la variable aleatoria Partición del espacio muestral en términos de la característica de interés: Representación de los eventos a través de un valor numérico Vinculación entre variable aleatoria y aleatoriedad Familiaridad con la que se maneja un número en lugar de un contexto Identificación de la vinculación de la característica de interés con los eventos en los que se subdivide el espacio muestral y con los valores numéricos	Interpretación de la partición del espacio muestral en términos de la característica de interés ¿Cuál es la diferencia entre la variable aleatoria, considerada como función, con otras funciones matemáticas? ¿Qué nombre le darías a la variable independiente en la variable aleatoria cuando la consideras como función?
La variable aleatoria y su distribución de probabilidad	La variable aleatoria como término en la función compuesta distribución de probabilidad La relación de dependencia entre la variable aleatoria y la probabilidad en la distribución de probabilidad. Características de la variable independiente en la distribución de probabilidad Vinculación entre la variable independiente y la aleatoriedad de la situación	Identificación de relación entre la variable aleatoria como variable independiente en la distribución de probabilidad: ¿Cuál es la variable dependiente en la distribución de probabilidad? En la distribución de probabilidad: Describe del método a partir del cual se obtiene el valor de la variable dependiente (probabilidad) a partir de la independiente (valor de la variable aleatoria) Definición de los elementos de la variable aleatoria: nomenclatura, dominio y características del dominio
Lenguaje	Nomenclatura (oral y escrita) de: la variable aleatoria la función de distribución	Representa con una letra la variable dependiente y con otra la variable independiente y utiliza estas letras para representar en notación funcional la solución del problema. ¿Qué nombre le darías a la variable independiente en la distribución de probabilidad?

- *La variable aleatoria y su distribución de probabilidad.* Buscamos observar la interpretación del estudiante de la variable aleatoria como término en la función compuesta distribución de probabilidad.
- *Lenguaje utilizado.* Dificultades, ambigüedades y nomenclatura al referirse a la variable aleatoria, las relaciones entre la variable aleatoria y la probabilidad, y las relaciones entre la variable aleatoria y la aleatoriedad.

6.4. Solución del problema

Alrededor del problema girará el protocolo de la entrevista, de manera que a la vez que constituya un reto accesible para los estudiantes, también debe permitir dar elementos para realizar la investigación. Por lo tanto, el problema seleccionado será de una dificultad media para permitir la concentración en la exploración cognitiva, pero lo suficientemente rico como para que permita la exploración de la comprensión de los estudiantes. Las características que deberá cumplir son las siguientes características:

- Manejar una situación en donde se involucre la aleatoriedad de manera natural.
- El significado de probabilidad vinculado al problema será la clásica.
- Su solución estará vinculada a la recomendación de una toma de decisión.
- Su solución requerirá el análisis de la situación completa y por lo tanto del manejo de una distribución de probabilidad.
- El riesgo en la toma de decisión podrá ser cuantificable.

Se espera que la solución del problema pase por las etapas de solución representadas en la Figura 5.4. En el problema se pretende vincular los objetos matemáticos por los que nos interesamos en esta investigación, pero también que se apliquen algunas estrategias sencillas de resolución de problemas. Así por ejemplo, esperamos que la verificación de la propuesta a partir del cuestionamiento del riesgo y de la visualización del alcance de la propuesta dada por los estudiantes, despierte cuestionamientos a su modelo planteado. Esto dará pie a redefinir o afianzarse en su propuesta basándose en argumentos matemáticos, proporcionados por las herramientas matemáticas que ponen en juego en esas etapas: la aleatoriedad, la distribución de probabilidad y la variable aleatoria.

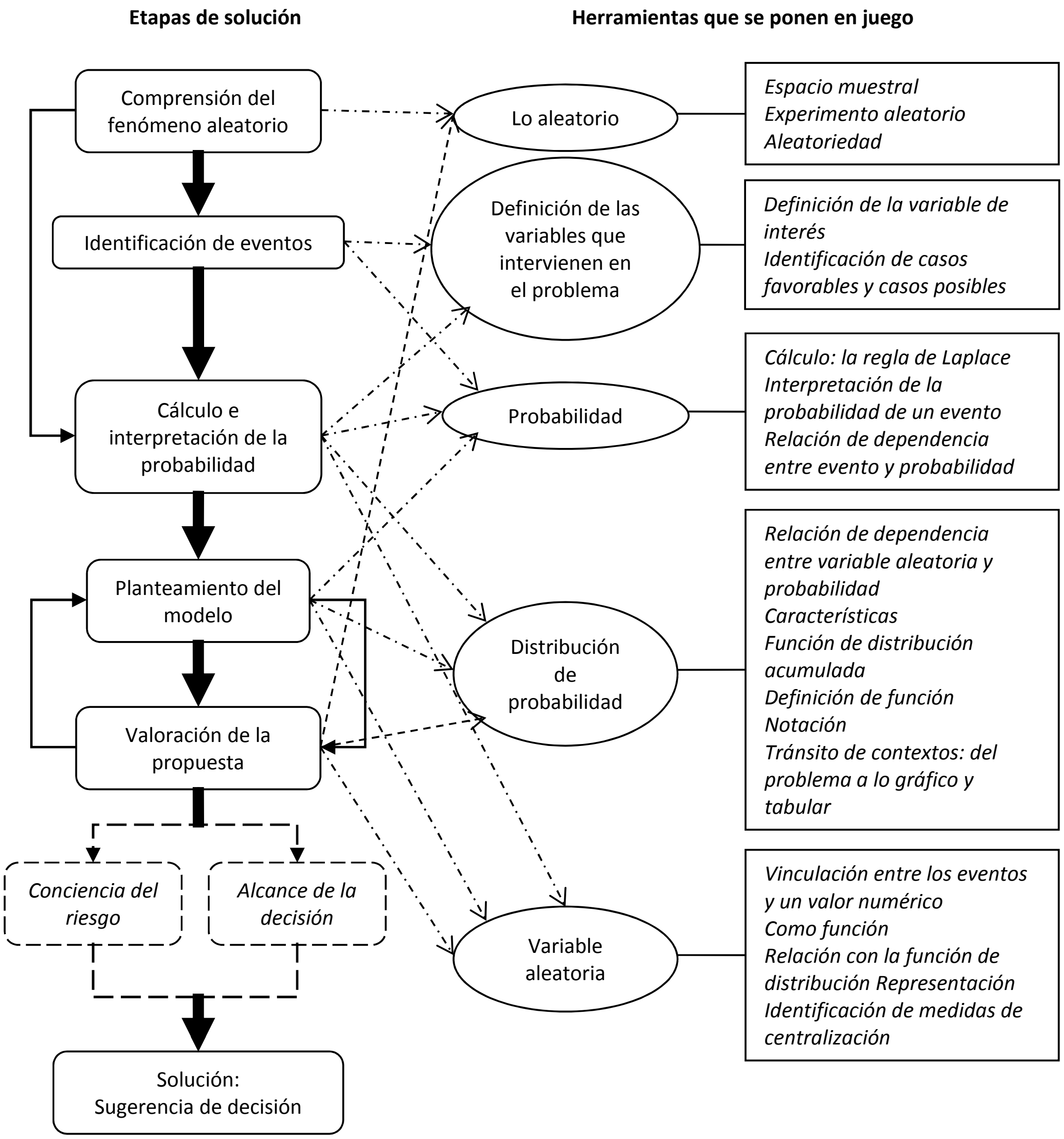


Figura 5.4. Etapas y herramientas esperadas en la solución del problema

7. Implementación

Se entrevistaron a dos estudiantes, Brenda y Mónica, que accedieron de manera voluntaria y que cursaban el primer semestre de universidad en el ITESM Campus Monterrey. Ninguna de las dos había llevado cursos de probabilidad o estadística en ese nivel. Una de ellas, Brenda, había llevado un curso de probabilidad y estadística en el bachillerato y la otra, no. En el curso que había tomado, Brenda llevó estadística descriptiva y reglas de probabilidad. No conocía el término de «variable aleatoria». En el momento de la entrevista, ambas cursaban matemáticas I para ciencias sociales (Cálculo diferencial) y eran consideradas buenas estudiantes por su profesor de matemáticas de ese semestre.

La entrevista se dividió en las tres partes de que consta la actividad y tuvo una duración aproximada de tres horas. A las chicas se les entregó el problema por escrito, una sola copia para ambas, y se les pidió que trabajaran directamente sobre el pizarrón para que sus respuestas fueran accesibles a los entrevistadores (su profesor de matemáticas y la profesora investigadora). Se les dio un tiempo en el que discutían y llegaban a una respuesta conjunta a cada inciso. Una vez que ellas consideraban concluida su respuesta, los entrevistadores dialogaban con ellas para profundizar o esclarecer las ideas que exponían. Al concluir cada sección de la actividad, ellas borraban y pasaban sus resultados a una hoja de papel, Su reporte estuvo constituido únicamente por sus conclusiones. Algunas respuestas las contestaron sólo verbalmente.

Se grabaron (tanto en video como en casete) todas las partes de la experiencia, tanto las discusiones entre ellas para llegar a su respuesta, como los diálogos establecidos con los profesores. La trascripción de las partes principales del diálogo sobre el desarrollo de la actividad se encuentra en el Anexo 3 de esta memoria.

8. Análisis de los resultados

En lo que sigue presentamos el análisis realizado y los resultados obtenidos, organizados en torno a los objetos de estudio. Se incluyen los pasajes de la entrevista relacionados con cada uno de los puntos principales, analizando los objetos matemáticos puestos en juego por el alumno, la comprensión que manifiesta el alumno y sus dificultades al resolver la tarea o al cuestionar sobre el objeto matemático puesto en juego. Así, los pasajes que se presentan a continuación no están descritos cronológicamente.

Nuestra discusión se centra principalmente en la interpretación que hacemos de las respuestas que las estudiantes dan a medida que van resolviendo el problema. Hacemos notar el conocimiento estadístico empleado, su conocimiento del contexto en el que se desenvuelve el problema y sus interpretaciones y validaciones de los resultados que obtienen. Consideramos de especial interés la forma en cómo transitan del dominio matemático al extra-matemático y viceversa. De acuerdo con Wild y Pfannkuch (1999), enfatizamos en el pensamiento estadístico (segunda dimensión de su marco) que integra el conocimiento estadístico y del contexto.

Los resultados de este análisis fueron publicados parcialmente en Ruiz, Batanero, y Albert (2006).

8.1. Objeto de análisis: Aleatoriedad

Este apartado se centra en el estudio de la comprensión del estudiante de la idea de aleatoriedad y su percepción de la incertidumbre, tanto en el experimento como en su recomendación. También interesa cómo vinculan la distribución de probabilidad con la incertidumbre. Se busca conocer el proceso que guía al estudiante a la utilización de términos cuantitativos en la solución de un problema en el que interviene el azar y la forma en que redefine la situación de incertidumbre una vez que ésta está representada a través de objetos matemáticos como la distribución de probabilidad (Figura 5.5).

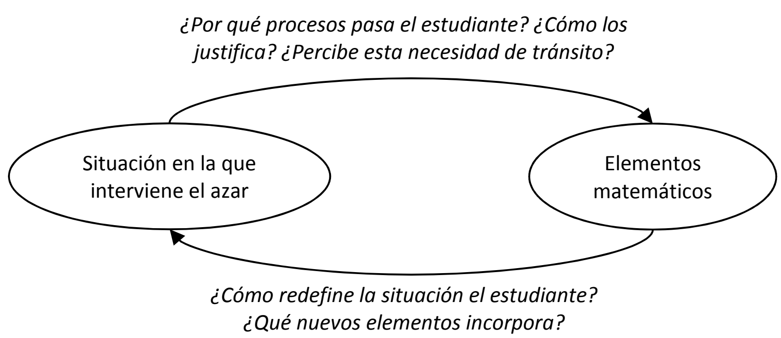


Figura 5.5. Descripción del objeto de análisis 'Aleatoriedad'

Preferimos tomar el aspecto informal de la aleatoriedad porque desde la perspectiva de la modelación de Heitele (1975) en este estrato, el experimento aleatorio es la «realidad», en cambio la distribución de probabilidad (o la función de distribución) es el «modelo matemático» que se está empleando para resolver el problema. Así mismo nos interesa la distribución de probabilidad porque en ella la probabilidad se vincula con la variable aleatoria.

De acuerdo con la Tabla 5.5 se analizará la comprensión del estudiante sobre la aleatoriedad en la situación problema y la relación entre la incertidumbre y la probabilidad.

8.1.1. Aleatoriedad en la situación problema.

En un primer enfrentamiento con el problema, hay una tendencia a dar una recomendación determinista (pasaje 8), principalmente por parte de Mónica que deja ver su tendencia a dar soluciones certeras, en las que no se involucre el azar. Hay dificultad en vincular el sorteo con la recomendación de cuantos boletos comprar. Mónica lee la pregunta: *«¿Qué recomendación le darías a la trabajadora social de cuántos boletos comprar?»* y la resuelve diciendo que compraría dos (para la pareja) y que después, cuando se sepa cuántos hijos tiene el trabajador que ganó, compraría los demás. No aceptan la aleatoriedad en la recomendación ni hace uso de los datos, sólo quiere resolver el problema de la forma más exacta posible.

> Mónica: Pues yo para no perderle y para no errarle, compraría dos nada más porque sí hay familias que tienen 0 hijos, de hecho hay 16 trabajadores que tienen 0 hijos. Entonces nada mas compraría dos para que vayan él y la esposa, y ya después si no sale, por lo menos ya tengo asegurados dos. Por eso preguntaba, si ya tengo asegurado dos... porque dice que pierda lo menos posible. En caso de que compre 3, y si por ejemplo sale una familia de estas, ya perdí un boleto, ya estoy perdiendo un boleto ¿qué voy a hacer con ese boleto? Así pienso yo. Yo compraría dos.

La aceptación a dar una solución no segura es impuesta por las condiciones del problema: después ya no se podrán comprar boletos. En el pasaje 11 Mónica (nuevamente) argumenta que hay una decisión que depende de la ética, que no tiene que ver nada con matemáticas y se quejan de que el problema esté planteado así. Ella se desespera de tener que jugar con lo incierto al no poder dar una solución más precisa y, desde su perspectiva, más justa.

> Mónica: Yo le pondría un 50% [refiriéndose más bien al 58% de probabilidad de que el trabajador premiado tenga 3 hijos] o si acaso un 70% [refiriéndose al 73% de probabilidad de que el trabajador premiado tenga 4 hijos], porque luego para que me salgan con que perdieron... aunque claro... maestro también está mal el problema porque no es cuestión de decirles oigan les vamos a dar boletos y si les toca pues que bueno porque es como decirles a los que tengan más [hijos] llevan las de perder porque no vamos a comprar para todos y luego que... Lo ideal sería que si la empresa está haciendo la campaña, pero eso ya no tiene que ver con matemáticas, sino que es más ético, si la empresa está ofreciendo pues que lo ofrezca bien.

En este mismo pasaje, ellas aún se resisten a aceptar que no puedan dar una solución con mayor seguridad, en cierto modo aceptan que no pueden controlar la situación con una manifiesta resignación.

> Mónica: Yo le pondría un 50% [refiriéndose más bien al 58% de probabilidad de que el trabajador premiado tenga 3 hijos], o si acaso un 70% [refiriéndose al 73% de probabilidad de que el trabajador premiado tenga 4 hijos], porque luego para que me salgan con que perdieron...

Brenda: En todo caso si se está buscando que no vaya a gastar de más y si alcanzan bueno y si no, no, pues yo estaría por 4 hijos.

Mónica: Ajá, yo también y si no alcanzan pues ya ni modo.

Posteriormente a lo largo del problema, ambas llegan a aceptar la incertidumbre de la decisión. Esto lo manifiestan en mayor medida en los pasajes 12 y 20:

Profesor: Pero existirá alguna manera, ¿es ignorancia nuestra el no saber qué boleto va a salir?

Brenda: No es algo que tú puedas controlar, es la suerte.

Así mismo en el pasaje 19 Brenda plantea una concepción de aleatoriedad relacionándola con una falta de patrón:

Brenda: ...son así como... aleatorios. No llevan un patrón así seguido.

8.1.2. Relación establecida entre aleatoriedad y probabilidad

En el pasaje 9 ambas alumnas piensan que la mejor solución a la pregunta sobre cuántos boletos comprar sería 5 boletos, basándose en que esto es «*lo más probable*», por tanto hacen uso explícito de la idea de probabilidad. Están aplicando una concepción laplaciana, puesto que se basan en la consideración de casos favorables y posibles y lo vinculan con los datos de la tabla. En su recomendación usan el número de hijos que tuviera la mayor frecuencia (la moda), apareciendo por tanto la idea de promedio y una estimación de las diferentes probabilidades. Se introduce la condicionante de que no pueden comprar boletos después de la rifa, entonces Mónica está de acuerdo con Brenda en comprar 5 boletos. Seleccionan «*lo más probable*» porque es lo que más seguridad les da. De alguna manera interpretan la probabilidad como una forma de medir la incertidumbre.

Brenda: Yo compraría 5, porque de acuerdo a la tabla, hay más probabilidad de que sea una persona con tres hijos y luego sumándole el obrero y la esposa serían 5.

Mónica: ...¡Ah! [*sin escuchar a Brenda*] entonces compraría 5 boletos porque la mayor incidencia de familias es de 3 hijos, o sea, el mayor número de trabajadores que hay por hijo es 3, o sea, las familias con 3 hijos son las que más abundan en la empresa, entonces como 3 hijos más el papá y la mamá pues ya son 5.

En el pasaje 10 aparecen de nuevo estas ideas. Mónica relaciona la intervención del azar con una situación en la que no puedes saber qué ocurrirá (impredecibilidad de lo aleatorio). Usa la idea de más probable (moda), relacionándola con lo que menos falla. Mónica ahora es la que trata de convencer a Brenda y argumenta a favor de comprar 5 boletos diciendo que es donde hay más probabilidades.

Mónica: ...hay más probabilidades de que yo agarre un trabajador de estos 45 [*se refiere a los trabajadores que tienen 3 hijos*], o sea, en el círculo que dibujamos [*señala la tabla*] éste [*señala el 45*] ocupa más, hay más probabilidades, es como que la que menos falla, es como que azar.

En el pasaje 12 Mónica de alguna manera continúa buscando el determinismo en la situación, pretendiendo afianzarse a algo seguro, y relaciona el porcentaje de trabajadores en la fábrica con un número menor o igual de hijos al de boletos comprados con la probabilidad de acertar en la decisión. Aparece aquí la idea implícita de probabilidad acumulada (por tanto distribución de distribución) y la idea de que la mediana induce una partición de la población en dos grupos con igual número de efectivos, como si con ello estuvieran siendo justas con todos los trabajadores (los que tienen más o menos hijos).

Mónica: Aquí lo que le aseguramos es el 50%, el otro 50% definitivamente va a tener que comprar uno, dos, tres, cuatro boletos por su cuenta, pero lo que usted quiere también es congraciarse o llevarse bien con los trabajadores, entonces pues con más del 50% ya tiene probabilidad de que lo saque, es decir, tiene asegurada la mitad.

8.1.3. Discusión de los resultados obtenidos en el objeto aleatoriedad

En resumen, del análisis anterior se puede observar cómo las estudiantes aceptan que no es posible dar una solución certera al problema. En un principio proporcionan una de las soluciones triviales. Hay una aceptación de la aleatoriedad en la situación puesto que manifiestan no saber cuántos hijos tendrá el trabajador que será premiado pero no aceptan dar una recomendación en la que no aseguren que todos los miembros del trabajador ganador puedan asistir a la función de teatro. Fischbein, Nello y Marino (1991) previenen en contra de estas actitudes que pueden surgir a partir de la enseñanza de la probabilidad y que significan una incomprensión de la aleatoriedad. De acuerdo con Konold (1989) esta actitud está vinculada con la heurística del «enfoque en el resultado aislado» (*outcome aproach*) porque la necesidad por dar una respuesta «exacta» imposibilita al estudiante a emitir un juicio basándose en un razonamiento probabilístico.

En este caso, sin embargo, la reticencia a dar una respuesta de la que no están seguras es sustituida por la aceptación de la aleatoriedad en la solución del problema. Al mismo tiempo la relacionan con la probabilidad e interpretan a la probabilidad como una medida de lo incierto, lo que les permite obtener soluciones y transitar por diversas nociones, como la moda y la mediana de la distribución de probabilidad de la variable aleatoria.

Esta vinculación entre la probabilidad y la aleatoriedad es lo que, de alguna manera, les permite aceptar lo aleatorio de la solución del problema y proporcionar una solución que les satisface. Poder manipular la probabilidad al decidir cuántos boletos comprar es lo que les hace ver cómo pueden ser equitativas o favorecer a la parte que les parece éticamente más apropiada. La aceptación de lo incierto está vinculada con una seguridad de que su solución es más equitativa ya que no pueden tener certeza, buscan equidad.

8.2. Objeto de análisis: Probabilidad

En este apartado es deseable conocer la forma en que la concepción de probabilidad del estudiante influye en su apropiación del concepto de variable aleatoria. En el problema la probabilidad se maneja en dos contextos (Figura 5.5), la vinculada a los eventos y la vinculada a la variable aleatoria, que nos analizar la forma en cómo el estudiante se relaciona con el problema para definir la probabilidad como función sobre el espacio muestral. Y, segundo, profundizar sobre la idea que se genera en el estudiante de la probabilidad y cómo la vincula en el contexto de incertidumbre.

Así, de acuerdo con la Tabla 5.6, la comprensión de la probabilidad será explorada a través de los procedimientos y concepciones que el estudiante haga en estos dos componentes del objeto probabilidad: la probabilidad en el espacio muestral y la distribución de probabilidad como función.

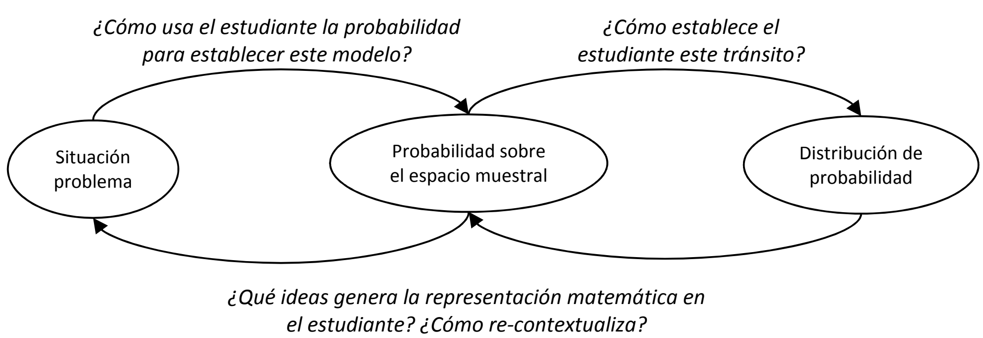


Figura 5.6. Descripción del objeto de análisis ‘ probabilidad’

8.2.1. La probabilidad en el espacio muestral

En el pasaje 13 a ambas estudiantes, en la medida que interpretan el problema, les queda claro el espacio muestral del experimento establecido por el sorteo, que sería el conjunto de los trabajadores de la fábrica (200):

Profesor: ¿es posible que pueda extenderse el número 200 a más valores? ¿por qué?
Mónica: No porque sólo son 200 trabajadores

Así mismo en el pasaje 21 se observa cómo están conscientes de que este espacio muestral está determinado por ciertas circunstancias: una fábrica, en un país, etc., que condicionan un ambiente y que el cambio de esas condiciones afectaría no sólo el espacio muestral sino también la distribución del número de hijos que tendrían los trabajadores.

Mónica: Ajá, porque [*los datos*] pueden estar determinados por diferentes factores por ejemplo, esa población puede haber tenido tres hijos porque se localizaba en México o en Estados Unidos en donde la mayoría de las personas tienen tres hijos, pero en poblaciones de otros lugares donde el número de hijos crece por persona, la gráfica ya no iba a ser igual…

En general se observa una buena interpretación y uso del concepto de probabilidad clásica. Hay evidencia de un buen manejo de dos axiomas de la probabilidad y tienen una interpretación correcta en el contexto del problema, pero tienen dificultades para despegarse de esa interpretación y vincular la probabilidad con un solo número. Así, en el pasaje 4, la probabilidad la expresan a través de la relación entre dos cantidades, «*31 de 200*». El número está escrito como una fracción y lo interpretan como esa relación, pero cuando hablan de ella separan las dos cantidades (pasaje 4):

Profesor: ¿Si tomamos un trabajador al azar cuál es la probabilidad de que tenga 4 hijos?
Ambas: 31 de 200.
Profesor: ¿Por qué no pusieron por ejemplo 4 entre doscientos u otro número porque 31?
Brenda: Porque bueno, en la tabla el número de hijos que son 4 tienen un número de trabajadores de 31 y el universo de trabajadores son 200. Entonces de esos 200 quiere decir que 31 tienen 4 hijos.

Esto repercute en que a pesar de que las alumnas interpretan correctamente el cociente (entre los casos favorables y los casos posibles) se les dificulta ver que ese cociente es un solo número, la probabilidad. Les resulta difícil dejar de situarse en el problema para encontrarse en el dominio matemático. Señalan y hablan del número de trabajadores (casos favorables), pero lo llaman «probabilidad», a pesar de que cuando se les pregunta mencionan que no hay que olvidar el denominador, porque es la forma de relacionar al número de trabajadores con un cierto número de hijos con el total (pasaje 4):

Profesor: Es decir, la probabilidad de tener cuatro hijos...
Brenda: O sea, esta es la probabilidad de 31,... [le arrebata la palabra, señala la respuesta al inciso a), 31/200, escrita a un lado de la tabla].
Brenda: ...entonces estas son todas las probabilidades que hay [*señala la columna de número de trabajadores*]...
Mónica: A pues sí.
Brenda: ...y si las sumamos sí nos van a dar 200. Haz de cuenta que ésta [*señala el 31 en la columna de número de hijos*] es la probabilidad de tener 4 hijos. Está diciendo el profe que se refiere a estas probabilidades [*señala la columna de número de trabajadores*] y al sumarlas nos darían 200.

No se dan cuenta de la contradicción entre el que digan que la probabilidad tiene que ser menor que uno, con el hecho de decir: «*la probabilidad* [*de tener 4 hijos*] *es 31*». La confusión entre el número de trabajadores y la probabilidad (es decir, confusión entre casos favorables, o cardinalidad de los eventos compuestos, y probabilidad) está más presente en Brenda, pero muestra mayor seguridad que Mónica. Ésta es convencida por Brenda de usar el número de trabajadores como probabilidad. Pareciera que fuera más palpable para ellas la variable 'número de trabajadores (con un cierto número de hijos)' que el de 'probabilidad'. A pesar de eso, reconocen que la probabilidad tiene que ser menor que uno y positiva (pasaje 4):

Profesor: Sí, por ejemplo el 31 sobre 200 es un número negativo, positivo...
Mónica: Positivo. Fracción.
Profesor: Fracción, mayor que uno, menor que uno,...
Ambas: Menor que uno.
Profesor: Menor que uno... ¿Y las probabilidades son mayores que uno?
Mónica: Son menores que uno. Todas son menores que uno.

Y que la suma de todas las probabilidades de los eventos es igual a 1 (pasaje 4):

Mónica: Yo digo que aún sumando todas estas [*señala la columna de número de hijos*] no nos da mayor que 1 porque para que fuera mayor que 1 la fracción tendría que ser 201 o más entre 200. O sea 201 sobre 200 y estos [*refiriéndose a la columna de número de hijos*] no van a dar más de 200.

El Profesor intenta utilizar el argumento de que una suma de probabilidades tendría que ser menor o a lo más igual a uno (porque es otra probabilidad) para hacerles ver que 200 no puede ser una probabilidad, esto las ayuda a darse cuenta que omiten el denominador (pasaje 4)

Profesor: Sí, ustedes me acaban de decir hace ratito que la probabilidad de tener 4 hijos es 31 sobre 200 y que esa era una fracción menor que uno ¿no?

Ambas: Ajá.

Profesor: Y luego me dijeron que todas las probabilidades son fracciones, fracciones menores que uno. Entonces si son 10 probabilidades a lo más que podríamos aspirar es a que su suma fuera menor que 10. Cuando yo les pregunto cuál es la suma de probabilidades, ustedes me dicen que 200 así que ¿Cómo le hacen para que llegue a 200?
[*Hacen exclamaciones de obviedad*]

Ambas: ¡Ay maestro!

Brenda: Lo que pasa es que esta fracción es con respecto a 200.

Mónica: Ajá. El común denominador es 200.

Sin embargo en el pasaje 6 nuevamente aparece la confusión entre los casos favorables y la probabilidad. Toman a la probabilidad como si fuera el número de trabajadores, a pesar de que se dan cuenta de que la probabilidad debe ser un número relativo a un total. Sin embargo tampoco son conscientes de su confusión.

Es posible que esa confusión empiece en el primer pasaje, cuando se les pide identificar una variable dependiente y otra independiente entre la probabilidad y el número de hijos. Al principio ellas establecen esta relación correctamente, pero cuando Mónica hace la afirmación: «*Yo digo que la probabilidad depende del número de hijos*», en Brenda se despierta una interrogante relacionada con los trabajadores. En esa fase, Brenda queda conforme hasta que logra establecer una frase en donde están incluidos: «*qué tan probable es que los trabajadores tengan ese número de hijos*». Como si en la relación establecida tuvieran que estar presentes los trabajadores porque son un elemento muy importante de la situación (es entre quienes se organizará la rifa) o como si les costara trabajo identificar la relación entre el número de hijos y los trabajadores.

Mónica: ...Yo digo que la probabilidad depende del número de hijos (recurre a voltear a ver a los profesores para que le confirmen o le refuten esta última afirmación).

Brenda: Pero más bien... el número...el...bueno... no, sí la probabilidad depende del número de hijos porque... [*dudando*]

Mónica: Es que la probabilidad...

Brenda: No, sí la probabilidad... [*muy segura*]

Mónica: Sí, porque según yo...
[*se arrebatan la palabra una a la otra*]

Brenda: depende del número de hijos porque...

Mónica: Ajá, porque qué tan probable es...

Brenda: porque qué tan probable es que los trabajadores tengan ese número de hijos [*en tono triunfalista*]

En el pasaje 5 se observa que interpretan correctamente la probabilidad expresada en forma de un número decimal, es decir llevan los doscientos trabajadores totales como si fuera una unidad.

Profesor: Nos podrían ayudar ustedes a encontrar un significado a ese decimal, ¿qué podría significar 0,155? ¿qué es eso?

Brenda: Yo digo que los doscientos trabajadores se están tomando como una unidad, entonces se supone que ese 0,155 es un pedacito de eso, a eso se refiere, entonces si sumamos cada fracción que da aquí, nos va a dar el entero que sería igual a uno. Bueno, yo entendí eso.

Otras concepciones de probabilidad se relacionan con grado de creencia o confianza en lo que puede pasar. En el pasaje 10 se relaciona con lo que menos falla:

Mónica: ... Ocupa más, hay más probabilidades, es como que la que menos falla, es como que azar.

En el pasaje 11 con una medida del riesgo:

Mónica: Yo le pondría un 50%, o si acaso un 70%, porque luego para que me salgan con que perdieron...

Y en el pasaje 12 se vincula con el grado de confianza:

Brenda: Bueno, es que no se puede asegurar, pero de lo que se habla aquí es de probabilidad, lo que puede pasar.

8.2.2. La distribución de probabilidad como función

En el pasaje 1, ambas estudiantes cambian las variables sobre las que se les pide que trabajen en la distribución de probabilidad, para ellas sería más normal tomar a la variable «número de trabajadores» como la variable dependiente. En un principio ambas dudan de las variables, pero después Mónica responde acertadamente en la distribución de probabilidad. Para Brenda no es tan natural, como si el hecho de que la pregunta pida explícitamente definir a la probabilidad como variable dependiente o independiente choca un poco con sus concepciones.

Brenda: Bueno... primero ¿qué depende de qué? pues el número de hijos... no...el número

Ambas: de trabajadores...

Brenda: ...depende del número de hijos, ¿no?

Mónica: No, no es que dice: el número de hijos depende de la probabilidad, no del...

Brenda: no.. pero la primera no es...

Mónica: ¿qué depende de qué? ...pues...

Ambas: el número de hijos depende de la probabilidad ¿o al revés?...

Mónica: No, la probabilidad depende del número de hijos o sea que qué tan probable... o sea... depende... la probabi...Yo digo que la probabilidad depende del número de hijos [*recurre a voltear a ver a los profesores para que le confirmen o le refuten esta última afirmación*].

Brenda: Pero más bien... el número...el...bueno... no, sí la probabilidad depende del número de hijos porque... [*dudando*]

Mónica: Es que la probabilidad...

Al final de este pasaje definen el valor de la probabilidad como variable dependiente (dentro de la distribución de probabilidad), pero en realidad no están convencidas. El motivo por el que aceptan esa respuesta es porque el texto de la pregunta les condiciona las variables con las que tienen que trabajar, dejando a un lado lo que ellas consideran la verdadera variable dependiente que sería el número de trabajadores (casos favorables o frecuencia). Mónica es la más convencida, pero a lo único que recurre para convencer a Brenda es a leer una y otra vez la pregunta recalcando las variables que se les pide manejar.

En el pasaje 3 hay reincidencia, esta vez no se percatan que la relación en la que les piden trabajar es la misma que la de la pregunta anterior, así que se inclinan porque la variable dependiente sea la frecuencia (o casos favorables).

Mónica: El número de hijos depende de la probabilidad o al revés y ya la pusimos. Ahora dice ¿Cuál de estas dos variables sería la variable dependiente y cuál la independiente?

Brenda: Pues yo digo que el número de trabajadores depende del número de hijos.

Esta confusión se ve más en Brenda, pero eso confunde a Mónica.

En el pasaje 6 se refieren a la probabilidad cuando mencionan el número de trabajadores (frecuencia). Expresan explícitamente su confusión en los términos, aunque también es posible que trabajen con los casos favorables porque a partir de ellos pueden pasar directamente a la probabilidad sin que eso implique un problema para ellas.

Brenda: No, porque, si estamos diciendo que la probabilidad depende del número de hijos, quiere decir que la probabilidad es la dependiente, el número de trabajadores que es la probabilidad es la dependiente [*señala sus respuestas en el pizarrón*].

Blanca: Pero ¿el número de trabajadores es la probabilidad?

Brenda: Si, son estas las probabilidades [señala la columna de número de trabajadores en el pizarrón].

En el pasaje 7 ellas se percatan de su confusión entre los casos favorables y la probabilidad de tener un número dado de hijos. Pero ello no ocurre hasta que pueden ver que para la asignación de probabilidades a la variable aleatoria (número de hijos) se formulan dos relaciones de dependencia, que el número total de trabajadores es una variable (un parámetro, es decir podría cambiar en otro contexto) y que intervienen otras tres variables en ese proceso. Las cuatro variables son:

- ❖ El número de trabajadores (la frecuencia o casos posibles): cardinalidad del evento compuesto en el espacio muestral del experimento (*¿cuántas personas?*),
- ❖ El número total de trabajadores (casos totales): cardinalidad del espacio muestral (*¿de cuántas*?),
- ❖ Número de hijos: variable aleatoria vinculada a la característica de interés que agrupó el espacio muestral en eventos compuestos, (*¿Cuántos hijos*?) y
- ❖ La probabilidad de obtener un cierto número de hijos *(¿Cuántas personas de cuántas?*).

Tratar el número total de trabajadores (o casos totales) como otra variable más las ayuda a descubrir que lo omitían por considerarlo una constante y que no era necesario mencionar a pesar de que, anteriormente (pasaje 4), lo manejaron correctamente («*el común denominador es 200*»). En este pasaje Mónica dice «*tres variables*» pero en realidad está mencionando las cuatro:

Mónica: Es que aquí están las tres variables [*señala la pizarra*],... tres porque de 0 hijos hay 16 personas que tienen 0 hijos de las 200, o sea, es que hay tres factores: ¿cuántos hijos?, ¿cuántas personas? ¿y de cuántas?, pero también, ¿cuántas personas de cuántas?; como que estas dos [*se refiere al número de trabajadores y al total de trabajadores*] van relacionadas muy directamente porque son 16 de 200, 9 de 200, pero esas dos se relacionan con esta [*con el número de hijos*], porque esta es la variable de cuántos hijos son.

Por lo tanto, se hacen evidentes las dos relaciones que las muchachas establecen para asignar al número de hijos, la variable aleatoria, una probabilidad:

Mónica: Según lo que yo entendí, es la probabilidad. Esta [*señala la columna de número de trabajadores en el pizarrón*].

Blanca: Pero ¿esa es la probabilidad? ¿el 12 es probabilidad?

Brenda: No, de aquí se sacaría la probabilidad de 12 sobre doscie... [*señala el pizarrón*].

Mónica: ¡Ah!... Es que hay dos relaciones.

Blanca: ¿Cuáles son las dos relaciones?

Mónica: Esta [señala el número de trabajadores] con esta [señala el número de hijos], y esta [señala el número de trabajadores] con el total [se refiere al total de trabajadores].

Las dos relaciones a las que ellas se refieren, son las siguientes (Figura 5.7):

- ❖ La primera relación se establece cuando se vincula el número de hijos (la variable aleatoria) con el número de trabajadores (cardinalidad de los eventos compuestos o frecuencia). En realidad la relación es más compleja, se está vinculando a la variable aleatoria con la característica de interés (que es algo menos abstracto) y posteriormente con los eventos simples (un trabajador de la fábrica) y con los compuestos (trabajadores con un cierto número de hijos).
- ❖ La segunda es el cálculo de probabilidades. Se establece entre el conjunto de trabajadores que tiene un cierto número de hijos (evento compuesto) y el valor de la probabilidad de tener un número dado de hijos, esto es $P(X=i)=N_i/N$, siendo N_i el número de trabajadores (casos favorables) con i hijos y N el total de los trabajadores (casos totales). En esta relación se define la probabilidad como función en el espacio muestral.

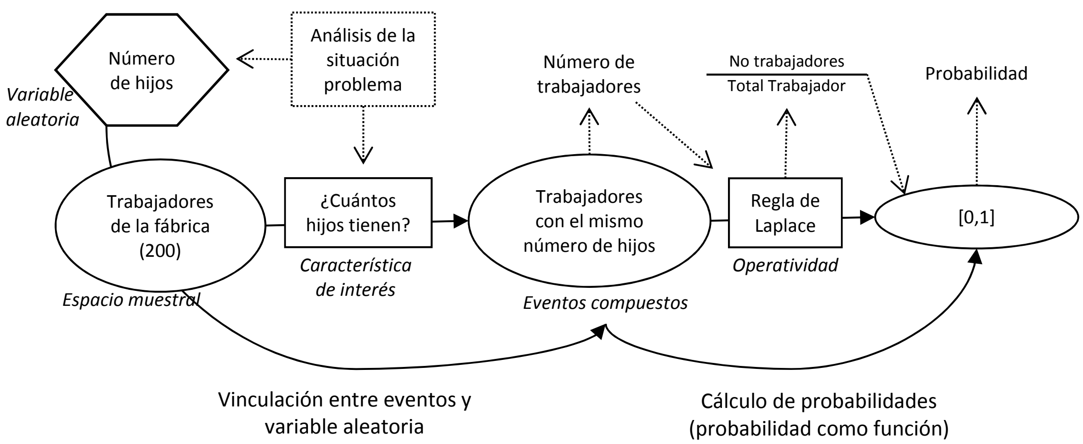


Figura 5.7. Proceso para vincular el número de hijos con una probabilidad

Lo cual indica que para poder asignar un valor de probabilidad a la variable aleatoria, las chicas tuvieron que contextualizar en el problema el significado de la variable aleatoria, es decir a relacionarla con el evento compuesto al que estaba ligada para asignarle la probabilidad de ese evento. Para ellas no representaba ninguna dificultad el cálculo de probabilidades, y es posible que eso es lo que les haya hecho obviarlo, en cambio la vinculación entre la variable aleatoria y el evento compuesto sí representaba un obstáculo[1]. Ellas representan el conjunto de los trabajadores que tienen el mismo número de hijos con su cardinalidad, pero en realidad están pensando en los trabajadores (como conjuntos de personas), no en su número. Lo que trabó su razonamiento es prácticamente aislar al conjunto de trabajadores de su relación entre el número de hijos y la probabilidad.

Observemos también que la cardinalidad de los eventos compuestos (trabajadores con un mismo número de hijos) juega un doble papel en esta descomposición en dos relaciones. Por un lado, al vincular el número de hijos con el subconjunto de trabajadores con un mismo número de hijos, es el resultado de esa aplicación. En cambio en el cálculo de probabilidades, es el argumento sobre el cual se aplica la regla de Laplace para poder obtener la probabilidad. Ambas relaciones son de conjunto, pero finalmente al vincular la variable aleatoria con la probabilidad se establece una relación entre números reales.

En el pasaje 14, al elaborar la gráfica, nuevamente se ven tentadas a usar el número de trabajadores como variable dependiente, pero ellas mismas se corrigieron y usaron la probabilidad como variable dependiente y lo resaltan «*¡Ah! Ya podemos saber que en vez de poner número de trabajadores, vamos a poner directamente la probabilidad*». En posteriores pasajes no vuelven a nombrarlos para referirse a la probabilidad.

En cuanto a la acepción de función, a lo largo de toda la entrevista ellas manejan diferentes concepciones. Una de las primeras de presenta en el pasaje 2, en donde Mónica momentáneamente cree que la distribución de probabilidad no es una función matemática porque ella cree que una función siempre crece. Cuando ella expresa la palabra *relación* se está refiriendo a una relación funcional.

> Mónica: Es que ve estás dos no pueden ser de que ¡ah! dependiente o independiente [*señala las columnas de la tabla*]... así, no. Yo no puedo entrar... de lo que el número de hijos y el número de trabajadores ... no puede ser, pon tú dependiente e independiente o independiente y dependiente [*señala las columnas de la tabla*] porque no, no hay relación... por ejemplo aquí [*señala la columna de número de trabajadores*] empieza un número 16, y luego aquí mayor, mayor, incrementa y luego decrementa y luego decrementa, así como que no hay rela...o sea no hay mucha relación, igual de cero hijos pudo haber un trabajador y de cuatro a cinco mil y así... como que no da resul...pero dice...

En el pasaje 3 tiene una mejor idea de lo que es una función matemática y se detiene a explicarle a Brenda que es una relación de dependencia entre dos variables:

[1] recordemos que en el pasaje 1, Brenda tuvo que recalcar «*qué tan probable es que los trabajadores tengan ese número de hijos*» para poder vincular a la probabilidad con el número de hijos.

Brenda: Sí, pero si te fijas esta es la que... ésta es la que... [*señala la columna de número de hijos*]

Mónica: Va determinando...

Brenda: Ajá, va determinando al número de trabajadores. El número de trabajadores no determina al número de hijos.

Mónica: Ah, ya entiendo.

Brenda: Sino al revés, el número de hijos es lo que está determinando el número de trabajadores. Entonces está sería...

En ese mismo pasaje observamos como diferencian «determinar» de «dependencia». Pareciera que al término «determinar» lo vinculan con una operación. Aunque esto no se observa muy bien en este pasaje, en posteriores recalcan la importancia de que «*una función esté determinada por una ecuación*».

Brenda: ... el número de hijos es lo que está determinando el número de trabajadores. Entonces está sería...

Mónica: Pero eso no significa que depende de los hijos va a ver tantos trabajadores sino que esta [*señala la columna de número de hijos*] rige a esta [*señala la columna de número de trabadores*].

Brenda: Entonces en este caso, el número de hijos es la variable independiente y el número de trabajadores es la variable dependiente.

En el pasaje 15, Mónica expresa que una función es una relación en las que una variable se corresponde con otra y manifiesta un buen manejo de la notación funcional. Ella expresa rápidamente el resultado pedido en notación funcional y eso le da pie a definir que su interpretación de función como una correspondencia entre dos números.

Profesor: ¿Sería posible eso [*se refiere a la gráfica de la distribución de probabilidad que está en el pizarrón*] escribirlo en notación de funciones? ¿qué es una función? Esa gráfica, suponiendo que fuera continua, ¿sería una función?

Mónica: Sí.

Profesor: ¿Por qué?

Mónica: Pues como todas [*se refiere a los valores involucrados*] se corresponden, por ejemplo f de 3 que corresponde a 0,225, ¿no? [*Escribe en el pizarrón: F(3) = 0,225*]

En ese mismo pasaje se observa como relacionan la *F* de la notación funcional con la palabra función, de modo que piensan que esa letra se usa para expresar que la relación establecida sí es una función, de modo que cambian el nombre de la variable dependiente (anteriormente se llamaba ‘*y*’) por esa *F*.

Mónica: Pues… ‘*x*’ sería *3* y ‘*y*’ sería 0,225. La función de ‘*x*’ es ‘*y*’, es decir el resultado.

Brenda: Pero en todo caso… no lo sabes. Lo que yo digo es utilizar…

Mónica: ¿En vez de *F*, 'y', aquí [*señala la F en F(x)*]?

Brenda: No, es que de cualquier forma haría falta especificar que es una función de '*y*'. Lo que podríamos utilizar es, para tomar exactamente las mismas letras, ponerle allá *F* de '*x*' [*se refiere al nombre del eje de las ordenadas*]. Es decir, cambiar aquella, ponerle *F(x)* en lugar de '*y*'.

En el pasaje 17 Mónica expresa una definición incompleta, pero Brenda la completa. No hay discusión alrededor de ambas definiciones, aparentemente hay una aceptación inmediata de las dos por la definición de Brenda. Con esta definición se acepta que la distribución de probabilidades sí es una función.

Profesor: ¿La relación que establecieron es una función matemática? o sea ¿es función? ¿Recuerdan las propiedades de la función? ¿qué es una función?

Mónica: Que hubiera dos variables.

Brenda: Que para el valor de '*x*' haya un valor en '*y*'.

Profesor: ¿Y puede haber dos '*x*' que vayan a dar a un mismo valor de '*y*'?

Ambas: Sí

Profesor: ¿Puede haber una '*x*' que pueda dar dos valores de *y*? Es decir, ¿un número de hijos que de dos probabilidades diferentes?

Ambas: No.

Profesor: Entonces ¿es función o no?

Brenda: Sí, si es función porque para cada número de hijos hay una sola probabilidad.

También hay una comprensión de que la gráfica y la tabla son una forma de expresar una función. Lo mencionan en los pasajes 15 y 17. Relacionan la gráfica de una función con la continuidad del trazo (pasaje 14).

Brenda: Aquí [*señala el eje de las abscisas*] sería, uno, dos,… así hasta nueve.
[*Mónica pone la escala en el eje de las abscisas*].

Brenda: Luego para arriba, ¿de qué la empezamos? ¿de 0,1, no?

Brenda: Bueno, sería de cero hasta uno y luego ya sería irle poniendo las probabilidades. A ver, divídelo…

Mónica: A ver, primero mitad y mitad. Punto cinco…
[*Brenda dicta los valores a Mónica y ella los va colocando en la gráfica*].

Mónica: ¿Hacemos la gráfica?

Brenda: Sí, hay que hacer la gráfica.
[*Brenda une los puntos*]

En el pasaje 16 vuelven a hacer énfasis en que no conciben una gráfica en la que no haya trazos, es decir, que haya sólo puntos.

Brenda: Bueno es que sí es cierto, se supone que nada más serían los que te están dando. Las coordenadas que te está dando la gráfica únicamente son estos [*señala los puntos de la gráfica*] no son mitades ni nada de eso, sólo los enteros.

Mónica: Entonces para qué es la gráfica si vamos a dejar los puntos así aislados.

Brenda: Sí, pero es que realmente no puedes sacar la probabilidad de 0,5.

Mónica: Sí, no puedes tener 8,5 hijos, es lo que yo digo.

Brenda: Pero por ejemplo, o sea, o sea, la gráfica sí está abarcando todo, sí está abarcando todo este espacio.
[*Ambas se quedan calladas un rato*].

Blanca: ¿Entonces están de acuerdo en que ese es el dominio? [*refiriéndose al [0,9], que ellas habían escrito en el pizarrón*]

Ambas: Sí.

En el mismo pasaje 16, ellas mencionan que su grafica sería igual si la variable fuera continua o fuera discreta, la diferencia sería que en una continua los decimales sí tienen sentido y en una discreta no, pero no les importa demasiado que la gráfica represente de una manera más fiel la situación problema.

Blanca: ¿Por qué lo encerraron en un corchete? ¿qué significa eso?

Mónica: Que va de… incluye del 0 todos los números, todos los números, hasta el 9. Cerrado.

Blanca: ¿Todos los números?

Mónica: Entre el cero y el nueve, sí.

Blanca: ¿Todos los números que hay entre cero y nueve?

Mónica: Sí.

Blanca: Cuatro punto cinco está cero y nueve.

Mónica: Pues también es parte porque... No está en la gráfica, pero está en el dominio. Igual, esto [señala el 4.5 en el eje de las abscisas y traza con el lápiz su proyección hacia la gráfica y después hacia el eje de las ordenadas] no tiene un valor escrito pero sí existe.

Lo mismo ocurre con el dominio, se dan cuenta que hay una incoherencia cuando quieren interpretar la gráfica (porque unieron los puntos con líneas) en el contexto del problema puesto que no pueden calcular el valor de la probabilidad (variable dependiente) para valores fraccionarios. Sin embargo consideran necesario tanto la unión de los puntos como la inclusión de números decimales en el dominio con la aclaración de que no se tomarían en cuenta puesto que no pueden obtener las coordenadas de esos puntos. La continuidad en la gráfica y en el dominio la justifican a través de su experiencia en sus clases de matemáticas. Se dan cuenta de que la variable independiente no puede tomar valores decimales, pero eso no las decide a cambiar su gráfica.

Por otro lado, se dan cuenta que la representación algebraica de esta función no es fácil de obtener, pero el argumento para no obtenerla a través de una regresión apunta a que están pensando en una regresión lineal y a que su dominio está limitado (no va más allá de nueve hijos). Además vinculan esta dificultad con el hecho de que los datos están determinados por factores ajenos a las matemáticas. En el pasaje 17 dicen que podrían obtener la expresión algebraica de la tabla si aumentara el número de hijos conforme aumenta la probabilidad, porque usarían regresión, pero como eso no ocurre, entonces no pueden obtenerla.

Mónica: Sin embargo estaba pensando ahorita ¿Cómo le podemos hacer? Lo primero que pensé fue regresión, hacer una regresión y encontrar una ecuación, pero después dije no, porque no tiene nada que ver el número de hijos con las personas, no es de que entre más hijos más personas haya sino de que pueda haber… si va disminuyendo, no va a llegar un momento en que se va a acabar, porque son hijos si fueran otros datos como dinero o alguna otra variable, puede que sí, pero como hijos no, pero como en, no sé, 15 ya se va a acabar, a partir de 3 va a ir disminuyendo.

En el pasaje 21 dicen que los datos que se manejan aquí no están determinados por una ecuación sino por las circunstancias, «*por el azar, yo creo*». Creen que la distribución de probabilidades es diferente a la función algebraica manejada por ellas en cursos de matemáticas porque el valor de ‘*x*’ no controla a la ‘*y*’, en el sentido de que no tiene un comportamiento que se pueda predecir a través de una ecuación. Se extrañan que la función no sea en cierta medida conocida o controlable. Y creen que esto es lo que la hace diferente de otros contextos manejados por ellas en matemáticas. Es decir, vinculan las funciones «*normales*» (determinísticas) con una ecuación.

Mónica: ... en otro tipo de gráficas normales a cada valor en ‘*x*’ corresponde un valor en ‘*y*’, es decir, tiene un comportamiento esperado, dependiendo de la gráfica, no nada más gráfica recta, que puede ser logarítmica o exponencial o polinómica, cualquier tipo de gráfica, pero para cualquier valor en ‘*x*’ se espera uno en ‘*y*’. Aquí no, para un valor en *x*, no sabes cuál va a ser el valor en ‘*y*’, puede ser más grande, más chico, es algo que no está controlado. La ‘*x*’ no controla a la ‘*y*’.

En el mismo pasaje 21, más adelante, lo vuelven a explicar.

Mónica: Cualquier expresión, cualquier ecuación se esperaría que tuviera cierto comportamiento, aquí no, por ejemplo, si por ‘*x*’ o ‘*y*’ razón no hubiera familias con 3 hijos, ahí iría para abajo otra vez la gráfica y no por eso estaría mal la gráfica.

8.2.3. Discusión de los resultados obtenidos en el objeto probabilidad

En la primera parte de los pasajes anteriores se puede observar un buen manejo del cálculo e interpretación de la probabilidad clásica y de su expresión tanto decimal como fraccionaria y de dos axiomas de la probabilidad. Así mismo tienen un correcto manejo de la probabilidad como el grado de certeza de la recomendación que darían. Sin embargo usan repetidamente el término ‘número de hijos’ para referirse a la probabilidad.

Konold, Pollatsek, Well y Gagnon (1997) reportan que los estudiantes generalmente se centran en la comparación de frecuencias absolutas en lugar de en las relativas al comparar dos conjuntos de datos. En nuestro caso es posible que no sea tan evidente que la probabilidad es variable dependiente en la distribución de probabilidad. Pero esa confusión también puede ser ocasionada por la dificultad de asignar un valor de probabilidad a la variable aleatoria, es decir, por la construcción de la distribución de probabilidad. Detengámonos a analizar esta postura.

En la actividad propuesta, pedimos a las estudiantes, que, una vez realizado el cálculo de la probabilidad de los eventos, trabajaran con la distribución de probabilidad[2]. El planteamiento formal de la pregunta constituyó un problema. Ellas no podían pasar directamente del valor de la variable aleatoria (un número de hijos) a la probabilidad, pues en ello interviene una composición de funciones que ellas inconscientemente quisieron obviar. Necesitaban regresar a preguntarse por el significado de la variable aleatoria en el espacio muestral. Es claro que las estudiantes no pensaron que la probabilidad y la frecuencia (casos favorables) son lo mismo, sino que a partir del número de trabajadores pueden calcular la probabilidad. La dificultad se traduciría en que ellas no podían abstraer (descontextualizar, desde la perspectiva de Chevallard, 1985) en dos sentidos:

- No podían desvincular la probabilidad de su cálculo a través de la regla de Laplace, y por lo tanto olvidar su interpretación como la razón entre los favorables y los posibles (en sus palabras: «*¿cuántos de cuántos*?»).
- Inconscientemente no podían desligar el valor de la variable aleatoria (número de hijos) de los eventos compuestos (trabajadores con ese número de hijos). Esto repercutió en que no representaban el conjunto de los trabajadores con un número (el valor de la variable aleatoria correspondiente).

Así, la asignación de probabilidades a los valores de la variable aleatoria y, por lo tanto, el establecimiento de la distribución de probabilidad (como objeto matemático), pasa por dos etapas que requieren, la *contextualización* y *descontextualización* del valor de la variable aleatoria en el espacio muestral y la

[2] Esto se pidió en las preguntas «*¿Qué depende de qué: el número de hijos depende de la probabilidad o al revés? ¿Cuál de estas dos variables sería la variable dependiente y cuál la variable independiente?*».

descontextualización de la probabilidad de los eventos del espacio muestral (Figura 5.8).

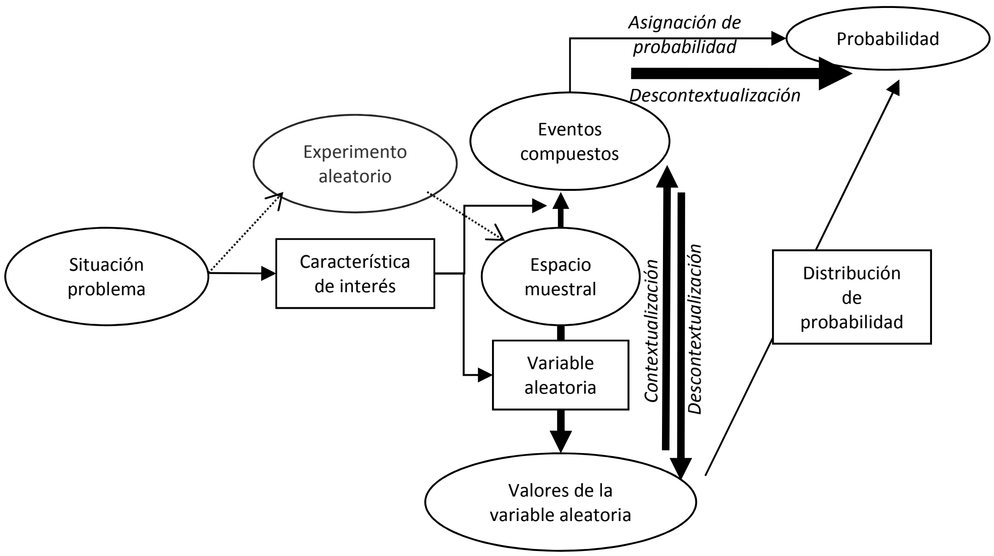


Figura 5.8. Proceso de contextualización y descontextualización de la asignación de probabilidad a la variable aleatoria y el establecimiento de la distribución de probabilidad.

En este proceso las estudiantes podían contextualizar bien el objeto matemático, valor de la variable aleatoria, pero tenían dificultades para descontextualizar. Pudieron pasar de un valor de la variable aleatoria (número de hijos) al contexto, conjunto de trabajadores que tenían ese número de hijos. Pero no fue tan sencillo pasar del contexto a un objeto matemático, es decir, representar a los trabajadores como un valor de la variable aleatoria (número de hijos que tenían)[3].

Así mismo, pudieron interpretarla la probabilidad en el espacio muestral (*¿cuántos de cuántos?*), pero no la formulaban como un solo número y por lo tanto no la desligaban del espacio muestral ni la asignan a la variable aleatoria, es decir, no podían descontextualizarla. Es posible que su correcta interpretación de la probabilidad laplaciana las haga ver a la probabilidad como dos números y no como uno (*30 de 200*). Es decir, este conocimiento que podría llamarse exitoso desde la perspectiva de la probabilidad laplaciana, sería un obstáculo a superar en la construcción de la distribución de probabilidad y la función de distribución. A este respecto, Sierpinska (1992) basándose en diversos obstáculos epistemológicos, reporta la necesidad de generalización y síntesis de la noción de *número* como una de las categorías a considerar en la comprensión del concepto de función en un contexto determinístico.

[3] Desde esta perspectiva, la aplicación de la inversa de la variable aleatoria sobre los valores que toma, es una contextualización de esos valores en el espacio muestral.

Así mismo, Fischbein y Gazit (1984) nos alertan sobre las divergencias entre el pensamiento probabilístico y el razonamiento proporcional, a tal grado que sugieren que ambos están basados en distintos esquemas mentales.

Otra explicación de la dificultad es que la distribución de probabilidad es algo abstracto, en cambio la variable frecuencia de trabajadores con un cierto número de hijos es más concreta para ellas. Esta dificultad para descontextualizar también se presenta cuando ven el dominio limitado a los datos dados en el problema, sin considerar que la distribución de probabilidad daría valor cero a la probabilidad del número de hijos mayor a 9. Esto último está de acuerdo con su idea de no incluir el cero como parte de los valores que puede tomar la probabilidad.

Una situación contraria se observa cuando se niegan a aceptar que la matemática sirve para representar la realidad, como si las representaciones matemáticas no tuvieran que estar vinculadas con un problema sobre todo en este contexto discreto. Incluso no encuentran contradicción en considerar que el número de hijos pueda tomar valores decimales, que no tienen sentido en el contexto del problema, pero que está de acuerdo con el contrato didáctico establecido en lo referido a las gráficas funcionales de sus cursos de matemáticas. Llegan a mencionar que la gráfica de una función sería la misma para variables discretas o continuas. Al obtener el dominio a partir de la gráfica de la función, donde unieron con líneas los valores puntuales, también cometen un error aunque saben que no pueden calcular la probabilidad para un número fraccionario de hijos.

A este respecto, García (2007), quien trabajó en un contexto determinista, reportó que todos los profesores de su estudio consideraron que los modelos matemáticos (en particular, la función lineal) son continuos por definición y que sólo a partir del estudio de la validez del modelo matemático, establecido previamente, en el contexto donde se va a utilizar pueden surgir variables discretas.

En sus estudios a lo largo de 10 años sobre el «sesgo de equiprobabilidad», Lecoutre (1992) concluyó que ninguno de los factores estudiados, incluyendo la formación de los estudiantes en teoría de probabilidad, tiene un efecto fuerte sobre ese sesgo. En cambio la caracterización de los modelos cognitivos usados por los estudiantes, indican que este error está vinculado con la incapacidad de los estudiantes para asociar razonamiento combinatorio y lógico con situaciones en donde interviene el azar[4]. Desde la perspectiva de Lecoutre, la principal dificultad de los estudiantes radica

[4]Para probar su hipótesis, realizó una investigación en donde los estudiantes se enfrentaron a dos problemas que requerían el empleo de la misma herramienta matemática, sólo que en uno de los problemas se enmascaró el aspecto «azaroso» y el otro era un «típico problema de azar». La proporción de respuestas correctas en el primer problema fue mucho mayor (75% contra 45%).

en la vinculación de las situaciones aleatorias con la equiprobabilidad, que viene de la aplicación de la regla de Laplace, pero no es aplicable en todos los casos y cuando lo es, sólo lo es en *sucesos simples*. Algunas veces, en cambio, los estudiantes lo aplican a los valores de la variable aleatoria o a sucesos compuestos. Esto es podría ser un indicio de la dificultad para *contextualizar* los valores de la variable aleatoria o los sucesos compuestos en el espacio muestral.

Ello es un indicador de la complejidad de los procesos de contextualización y descontextualización que tienen lugar en las situaciones aleatorias para vincular el modelo matemático con la realidad y explica la ruptura entre el mundo matemático y el real en situaciones en las que interviene el azar que hicieron las estudiantes de nuestro estudio.

8.3. Objeto de análisis: Variable aleatoria

Aquí trataremos de explorar los tres momentos más importantes del papel que juega la variable aleatoria en la construcción del modelo de distribución de probabilidad: (Figura 5.9). Observemos que en la resolución del problema, la variable aleatoria será el producto de un esfuerzo de síntesis de la integración del conocimiento estadístico y del contexto, de acuerdo con la segunda dimensión del marco de pensamiento estadístico de Wild y Pfannkuch (1999). La variable aleatoria en el problema es, por tanto, la intermediaria entre las constantes conexiones entre el contexto y los resultados de los análisis para acceder a los significados de tal análisis.

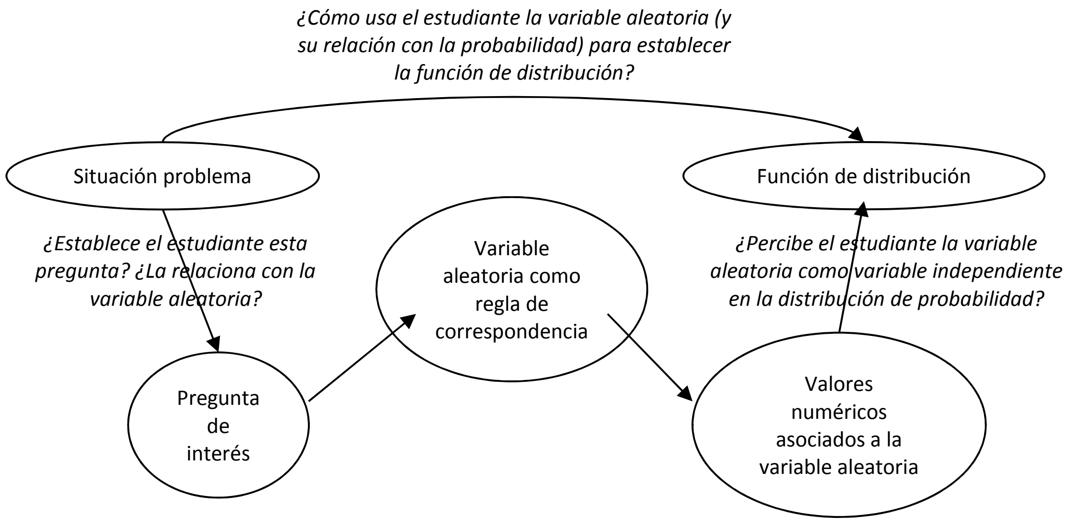


Figura 5.9. Descripción del objeto de análisis 'La variable aleatoria'

Las componentes, de acuerdo a la Tabla 5.7, que nos permitirán analizar los procedimientos y concepciones de los estudiantes en el objeto variable aleatoria son: su concepción propiamente como función, el valor de la variable aleatoria como número vinculado a una probabilidad y como variable independiente dentro de la distribución de probabilidad. Además, dentro de este objeto de análisis incluiremos el lenguaje matemático usado por los estudiantes.

8.3.1. La variable aleatoria en el problema

En el pasaje 9, ya comentado, ambas alumnas piensan que la mejor solución a la pregunta de cuántos boletos comprar sería 5 boletos, basándose en que esto es «*lo más probable*». Por tanto hacen uso explícito de la idea de que el número de hijos es una variable, al mismo tiempo que su respuesta indica que son consientes que el valor de esta variable está vinculado a la aleatoriedad.

En el pasaje 12, Brenda interpreta la distribución para explicar por qué recomiendan el valor que tiene la mayor probabilidad, de qué manera este valor soluciona el problema y como ésta se representa en la tabla de datos dada. Interpretar el contexto se le facilita más que manejar las frecuencias al señalar la acumulación de la mayor parte de los trabajadores.

Brenda: Como además, se encuentran más concentrados los trabajadores en esa parte [*señala los renglones de la tabla correspondientes a trabajadores con 3 hijos o menos*] hay más posibilidades de que no vaya a perder tanto porque más del 50% [*de los trabajadores*] se juntan en ese espacio. No podemos asegurarlo pero es más factible, puede que suceda más, que salga más una de estas personas [*señala los renglones correspondientes a los trabajadores con más de 3 hijos*], es más fácil que del resto porque son más, son mayoría.

En el pasaje 10 Brenda se comenzó a fijar en otra característica de la variable aleatoria: la probabilidad acumulada del complemento. Ella comienza a dudar porque se percata de a cuántos trabajadores podrían no alcanzarles los boletos.

Brenda: Bueno, yo diría eso, comprar los 5, pero también existe la probabilidad de que no vaya a completar después.

En el pasaje 10 Brenda relaciona las probabilidades acumuladas con la probabilidad que les interesa analizar en el problema. Ella introduce las probabilidades acumuladas (aunque usa también la frecuencia acumulada: «*el número de trabajadores hasta aquí*») para argumentar una solución diferente a los 5 boletos. No efectúa las operaciones, sólo hace aproximaciones.

Brenda: Aunque bueno, también por decir un margen más grande sería... bueno más seguro sería comprar los 9, porque si te fijas, de aquí hasta acá [*señala en la tabla la columna de probabilidades las correspondientes de 0 a 6 hijos*] es donde se ocupa la mayor parte, de 0 a 6 es donde se está ocupando la mayor parte inclusive [*se podría tomar*] un poco más, ya en la de 7 hijos que es donde llega a 9 [*es decir, añadiría los trabajadores que tienen 7 hijos*], o sea, es donde está la mayor probabilidad, ésta es muy poquita [*señala la probabilidad de que se tengan 8 ó 9 hijos*]. Estos dos tienen muy poquita probabilidad y serían 7 más los... que serían... Si te fijas realmente ya nada mas tendrías un margen de 13 o sea muy pocos.

Como ya notamos en el punto 8.2.3 y en los pasajes arriba citados, el uso de la variable aleatoria requería que las estudiantes regresaran constantemente a los eventos compuestos. Posteriormente, a partir del pasaje 14, cuando trabajan la distribución de probabilidad y la función de distribución no tienen mayores conflictos en usar los valores de la variable aleatoria. Sin embargo, a lo largo de la entrevista recurrieron constantemente al manejo de la cardinalidad del conjunto de trabajadores (frecuencia) para interpretar sus resultados, en lugar del valor de la variable aleatoria que tomaría ese conjunto de trabajadores.

8.3.2. La variable aleatoria como función

La actividad no permitió que las estudiantes plantearan la regla de correspondencia asociada a la variable aleatoria puesto que se les proporcionaba estaba en el enunciado del problema. Como ya dijimos, la interpretación de la variable aleatoria fue exitosa en la medida que ellas pudieron visualizar bien lo que significaba en el contexto del problema, pero no facilitó la descontextualización de los objetos matemáticos, ni tampoco que aceptaran que las matemáticas podrían responder a un contexto.

Al preguntarles por el rango y dominio en la distribución de probabilidad, en el pasaje 16 asocian las cotas superiores e inferiores del rango y el dominio con los valores más grandes y más pequeños que pueden tomar las variables al interpretar la grafica.

Profesor: Bueno, la que sigue. Describe el dominio de la variable independiente y el rango de la variable dependiente.

Mónica: Pues de 0 a 9 es el dominio de la independiente y el rango de la dependiente sería de 0,025 a 0,225 [*escribe en el pizarrón [0,9] y [0,025, 0,225]*].

El dominio de la distribución correspondería a la imagen de la función variable aleatoria, de modo que no tendría sentido incluir valores decimales y sin embargo, como ya lo hemos mencionado, las estudiantes incluyen valores decimales en el dominio.

No creen en una necesidad del orden en la variable independiente de la distribución. Más bien ven a los números como etiquetas. Los valores de la variable aleatoria aparecen en un orden ascendente en su tabla y en su gráfica, pero lo hacen porque así lo han hecho siempre en sus cursos de matemáticas. Argumentan que podrían estar ordenados de otra forma, por ejemplo de menor a mayor probabilidad o al revés o que incluso los valores de la variable aleatoria podrían haberse ordenado de manera descendente. Explican que los valores de la variable ‘número de hijos’ se presentarán aleatoriamente en caso de efectuar varias veces el experimento volviendo a introducir el papel en la tómbola (pasaje 20). Se observa, también, un buen manejo de la noción de aleatoriedad y un salto directo de la variable aleatoria al experimento sin pasar por los conjuntos de trabajadores. Esto es, vinculan el resultado del experimento aleatorio con el número de hijos directamente, lo que facilita que vinculen directamente la probabilidad con el número de hijos.

Brenda: … bueno en este caso sí va así seguido [*se refiere a los valores de la variable aleatoria*], pero a lo mejor hay casos en que no había nadie con 6 hijos entonces la variable ya no tendría que ser toda seguida.

Mónica: Pues obviamente sí se pueden ordenar siempre las variables [*se refiere a los valores que puede tomar la variable*], así de menor a mayor y de mayor a menor también [*señala los valores de la probabilidad*], pero eso no significa que de mayor a menor van a ir creciendo o van a ir disminuyendo igual [*si se repite muchas veces el experimento aleatorio*], o sea, el que aquí sea de menor a mayor [*se refiere a los valores de la variable aleatoria en la gráfica*], el que estén ordenadas, no significa que también estén creciendo uniformemente [*se refiere al efectuar varias veces el experimento*].

En el pasaje 21 hay un momento en el que Mónica describe las relaciones que se establecen entre las variables de este problema para determinar la variable aleatoria. Menciona la regla de correspondencia y el proceso que se lleva a cabo para obtenerla de manera correcta y detallada y se muestra que conoce la naturaleza del objeto matemático que está manejando. Ella utiliza un razonamiento regresivo (Polya, 1965).

Brenda: No, porque adentro de la tómbola están los trabajadores. Es que se supone que del número de hijos… por decir…¿cómo lo explicaré?

Mónica: O sea no está el número de hijos, de que uno, dos, sino que está un trabajador y tú, a ver cuántos hijos tienes. No pues que dos. Está implícito adentro del trabajador.

Brenda: Sí, ándale. O sea no hay papelitos que digan uno, dos, tres, …

Mónica: No, sino que sale, Juan Ramírez y tú, a ver, cuántos hijos tienes, no pues que tres, a corresponde a… [*señala el eje de las ordenadas en la gráfica*] y ya.

Brenda: O sea que dentro de los papelitos que están en la tómbola, están los [*valores*] posibles o sea estos [*señala la escala del eje de las abscisas*].

Se dan cuenta que el fenómeno aleatorio está relacionado con la elección de un trabajador y que la situación problema es la que condiciona qué pregunta se le haría al trabajador y la partición del espacio muestral (la característica de interés). Ellas logran este aprendizaje para este problema, pero no se puede garantizar a otras situaciones.

8.3.3. La variable aleatoria y su distribución de probabilidad

Las alumnas indican que el dominio de la variable aleatoria es el conjunto de números entre 0 a 9 y el rango de la probabilidad sería de 0,025 a 0,225 (escriben en el pizarrón [0,9] y [0,025, 0,225]). A pesar de que anteriormente habían mencionado la probabilidad cero, a la hora de definir el dominio y el rango no lo toman en cuenta, ni tampoco mencionan la posibilidad de que haya 10 o más hijos. Para ellas la distribución incluye el intervalo entre 0 y 9 (pasaje 16), no lo hacen con fines de extender la probabilidad de la distribución de probabilidad, sino porque piensan que en el ámbito matemático todas las gráficas de las funciones son continuas. Esto se observa en el pasaje 16.

Brenda: Bueno es que sí es cierto, se supone que nada más serían los que te están dando. Las coordenadas que te está dando la gráfica únicamente son estos [*señala los puntos de la gráfica*] no son mitades ni nada de eso, sólo los enteros.

Mónica: Entonces para qué es la gráfica si vamos a dejar los puntos así aislados.

Brenda: Sí, pero es que realmente no puedes sacar la probabilidad de 0,5.

Mónica: Sí, no puedes tener 8,5 hijos, es lo que yo digo.

Brenda: Pero por ejemplo, o sea, la gráfica sí está abarcando todo, sí está abarcando todo este espacio.

Blanca: ¿Entonces están de acuerdo en que ese es el dominio?

Ambas: Sí.

En el pasaje 18 las diferencias inmediatas que ambas estudiantes encuentran entre las variables manejadas en la distribución de probabilidad y otras variables independientes con las que ellas han trabajado radican principalmente en la rareza de los datos discretos y en que es muy restringido el intervalo de valores que puede tomar.

Profesor: ¿Qué diferencia encuentras entre la variable independiente de esta distribución de probabilidad y las variables independientes de las otras matemáticas que tú conoces, por ejemplo del álgebra y del cálculo?

Mónica: Qué es limitada.

Profesor: Supongamos que tiene un sentido «más hijos» indefinida, aun así ¿Hay más características? ¿Es la única característica que ustedes encuentran entre esta función y las otras variables independientes de álgebra y cálculo?

Mónica: Lo de los decimales.

También se observa que vinculan la función con un registro algebraico, que ven como otra diferencia entre las funciones que han trabajado en anteriores cursos de matemáticas y el contexto de esta actividad. En el pasaje 19 mencionan que otra diferencia es que no pueden obtener la ecuación a través de la cual se relacione la variable dependiente con la independiente. Esto lo vinculan con la *«exactitud»* de la variable dependiente, es decir, a partir de la variable independiente no pueden calcular la dependiente de manera exacta (porque sólo puede ser obtenida a partir de la tabla o de la gráfica).

Profesor: Qué más, piensen ¿Habrá alguna característica más que distinga esta variable de las otras que trabajan en matemáticas?

Brenda: Normalmente en álgebra manejamos ecuaciones. Todas las gráficas se manejan por ecuaciones, todos los valores de ’*y*’ se van a dar por la ecuación. Aquí los valores de *y* también se dan según una ‘*x*’, pero no tienen una ecuación, son así como… aleatorios. No llevan un patrón así seguido.

En los pasajes 21 y 22 también se observa esta tendencia a querer interpretar una función a través de una ecuación. Aquí interpretan la palabra «predecir» como obtener el valor de la probabilidad de otro número de hijos, es decir, otros puntos de la gráfica que les indiquen cuál sería su comportamiento después del 9.

Mónica: Las otras son constantes, continúas, se espera que sigan y que vayan, en cambio esta no sabes, no puedes predecir algo.

Expresan que aunque sí podrían sacar una ecuación, no tendría sentido en el contexto del problema (pasaje 19). Observemos que en el diálogo citado arriba:

Brenda: Normalmente en álgebra manejamos ecuaciones. Todas las gráficas se manejan por ecuaciones, todos los valores de ’*y*’ se van a dar por la ecuación. Aquí los valores de *y* también se dan según una ‘*x*’, pero no tienen una ecuación, son así como… aleatorios. No llevan un patrón así seguido.

Piensan que si no tienen un «*patrón*» no pueden obtener la ecuación. Relacionan esa carencia de patrón con un comportamiento aleatorio, pero atañen ese comportamiento aleatorio a la variable dependiente (es decir a la probabilidad) porque es ella la que no pueden calcular.

Brenda: Bueno más que la variable independiente diferente, como que es la variable dependiente la diferente, porque no podemos establecer cuál será la dependiente [*se refieren al valor que tomaría esa variable*]. A diferencia del álgebra, con la independiente podías sacar la dependiente exactamente y en este caso no, son datos que te están dando y tú no puedes... es muy aleatorio, es más difícil que con álgebra. Con álgebra tú podías sacar '*y*' con la ecuación y aquí no.

Ellas vinculan la incertidumbre con la probabilidad más que con la variable aleatoria. Interpretan esa «aleatoriedad» con el no tener una ecuación para poder calcular la '*y*' (probabilidad). Los valores de la variable aleatoria para ellas son números que se debería poder sustituir en una ecuación para obtener el valor de la probabilidad. En el pasaje 21 vuelven a retomar esto.

Mónica: Yo también estoy de acuerdo con ella en que la que varia es la variable dependiente. Varía de forma inconstante o impredecible.

Blanca: Pero ¿por qué impredecible?

Brenda: Sí, porque este valor no es algo que esté dado por una…

Mónica: Función.

Brenda: Sí, por una función, por una ecuación, sino que porque son datos que así se dan o sea que, que pueden ser más o que pueden ser menos, que nos están determinados por algo sino por las circunstancias, por el azar yo creo.

En el pasaje 20, ellas mencionaron que los valores de la variable aleatoria no tienen que presentarse continuamente en caso de realizar varias veces el experimento, ya que no pueden saber en realidad en qué orden se presentará.

Profesor: Entonces ¿habrá orden de aparición en el número que salga de la tómbola?

Brenda: O sea de que primero saque uno y salga uno, después vuelva a sacar y salga dos y así, pues no, no lo hay.

Sin embargo en el pasaje 21 no vinculan, de manera explícita, el número de hijos con la aleatoriedad. Ni tampoco se cuestionan por qué en el eje de las abscisas ellas colocaron los valores que puede tomar la variable aleatoria en forma ordenada. Cuando ellas mencionan las palabras: «aleatorio» y «azaroso» se refieren al valor que tomará la probabilidad:

Mónica: Ajá, porque [*los datos*] pueden estar determinados por diferentes factores por ejemplo, esa población puede haber tenido tres hijos porque se localizaba en México o en Estados Unidos en donde la mayoría de las personas tienen tres hijos, pero en poblaciones de otros lugares donde el número de hijos crece por persona, la gráfica ya no iba a ser igual. En cambio si hablamos de dinero, bueno, no tanto de dinero, sino de otro tipo de cosas como habíamos visto antes [*se refería al curso de matemáticas que acaba de llevar con el profesor*], sí se espera algo, pero aquí el número de hijos… En este tipo de gráficas en donde están midiendo el azar, no podemos saber cuándo va a bajar y cuándo va a subir, no tiene un comportamiento predestinado.

Blanca: ¿La variable dependiente? ¿la probabilidad?

Ambas: Sí.

En otro momento de ese mismo pasaje aseguran que los valores de la variable aleatoria serían no predecibles sólo si la probabilidad de que tome ese valor fuera cero:

Mónica: ... a lo mejor hay casos en que no habría nadie con 6 hijos, entonces la variable ya no tendría que ser toda seguida [*se refiere a la numeración en el eje de las abscisas*].

Como si en tal situación el valor de «6» se pudiera quitar del eje de las abscisas, puesto que no hay nadie que tenga ese número de hijos. Esta idea se expresa más claramente en el pasaje 22 cuando intentan dar un nombre general a las variables independientes de este tipo y Mónica sugiere llamarla «impredecible».

Brenda: No, la impredecible es *y*.

Mónica: No, pero también es *x*.

Blanca: ¿Por qué esta también es impredecible?

Mónica: No, inexacta… ¡Porque! Por… porque a lo mejor podemos quitar el dos, a lo mejor no hay familias con dos hijos y si no hay, no tiene caso…. Incontrolable, no.

En estos casos con probabilidad cero, piensan que la numeración en el eje de las abscisas (los valores de la variable aleatoria) podría ser no seguida. Vinculan lo impredecible con la variable aleatoria, pero no por su relación con el sorteo sino porque podrían saltarse algunos «valores». Esto es coherente con su creencia de que lo aleatorio es la probabilidad.

Relacionar la aleatoriedad con una incapacidad de cálculo las induce a utilizar una noción errónea de aleatoriedad porque aunque ellas pudieran calcular todos los valores de probabilidad a través de una ecuación, eso no haría que supieran cuál es el trabajador que saldría seleccionado. Esta idea la argumentan varias veces a lo largo del pasaje 21.

Relacionan los datos «humanos» (es decir, los obtenidos a través de la experiencia) con la «inexactitud» de la función que están manejando. Piensan que no es posible modelar los datos que tienen a través de una ecuación porque son datos obtenidos a través de un censo y no de una ecuación. Están manifestando una creencia de que las ecuaciones no se obtienen a partir de «datos reales» y precisamente por eso creen no posible obtener una a partir de los datos que están manejando.

Mónica: Aunque sí se observa [*señala la gráfica de la distribución de probabilidad*] que primero crece y luego ya disminuye, disminuye y como son hijos y como somos humanos son funciones más o menos, o sea va a ir disminuye y disminuye hasta que llegue a 12 y ya. En ese caso pero por ejemplo más datos humanos como… número de personas divorciadas o cuántos se han casado, ahí no sería viable sacar una ecuación porque no te iba a decir nada.

Brenda: Bueno en realidad sí puedes sacar una ecuación, pero eso no te daría… no sería muy exacto. Esa es la diferencia, que no te pueden dar datos exactos.

Esta interpretación indica que las estudiantes no están pensando únicamente en la situación planteada, sino en una generalización para cualquier situación. Ellas piensan que una ecuación no debe servir para un caso sino para muchos (en este contexto para muchas fábricas), de manera que les causa conflicto pensar que no pueden determinar cuántos trabajadores tendrían un cierto número de hijos (y por lo tanto la probabilidad) en todas las fábricas con una sola ecuación. En el pasaje 21 esto se observa con más claridad, aunque a partir del pasaje 20 se repite en varios pasajes.

Mónica: Yo también estoy de acuerdo con ella en que la que varia es la variable dependiente. Varía de forma inconstante o impredecible.

Blanca: Pero ¿por qué impredecible?

Brenda: Sí, porque este valor no es algo que esté dado por una…

Mónica: Función.

Brenda: Sí, por una función, por una ecuación, sino que porque son datos que así se dan o sea que, que pueden ser más o que pueden ser menos, que nos están determinados por algo sino por las circunstancias, por el azar yo creo.

Mónica: Ajá, porque pueden estar determinados por diferentes factores por ejemplo, esa población puede haber tenido tres hijos porque se localizaba en México o en Estados Unidos en donde la mayoría de las personas tienen tres hijos, pero en poblaciones de otros lugares donde el número de hijos crece por persona, la gráfica ya no iba a ser igual. En cambio si hablamos de dinero, bueno, no tanto de dinero sino de otro tipo de cosas como habíamos visto anteriormente [*se refiere al curso de cálculo que acaba de llevar con el profesor*], sí se espera algo, pero aquí el número de hijos… En este tipo de gráficas en donde están midiendo el azar, no podemos saber cuándo va a bajar y cuando va a subir, no tiene un comportamiento predestinado.

Además, se observa una vinculación entre una ecuación y la predicción. Por eso para ellas la incapacidad de cálculo está vinculada con la incertidumbre (no saben qué pasará en otras circunstancias). En otras fábricas el valor del número de hijos (1, 2, 3,... 9) sería el mismo a menos que la probabilidad de tener un cierto número de hijos fuera cero, en cambio los valores exactos de la probabilidad cambiarían. Vinculan las ecuaciones con poder predeterminar resultados sin necesidad de adentrarse a la situación.

8.3.4. Lenguaje utilizado

En el pasaje 15 la notación de las variables que proponen son las utilizadas comúnmente en los cursos de matemáticas, '*x*' y '*y*'. Inicialmente, al mencionar notación funcional, Brenda relaciona la respuesta con una ecuación o con una expresión algebraica. Se pregunta si una función sólo está definida a través de una ecuación. Los profesores aprovechan para explorar sus concepciones de función, pero ellas no se muestran muy seguras de sus conocimientos al respecto. En un momento de ese diálogo, Brenda se da cuenta que no necesitan saber la ecuación para dar la respuesta.

Profesor: Representa con una letra la variable dependiente y con otra la variable independiente.

Mónica: 'x' y 'y' [las escribe en la gráfica que está en el pizarrón]. Ya.

....

Brenda [*dirigiéndose a Mónica*]: ¿A qué se refiere con notación funcional? ¿una ecuación?

[*Dudan*]

Mónica: ¡Funciones!, *F* de *x*, *f* de…

Profesor: ¿Sería posible eso [*se refiere a la gráfica que está en el pizarrón*] escribirlo en notación de funciones? ¿qué es una función? Esa gráfica, suponiendo que fuera continua, ¿sería una función?

Mónica: Sí.

Profesor: ¿Por qué?

Mónica: Pues como todas [*se refiere a los valores involucrados*] se corresponden, por ejemplo f de 3 que corresponde a 0,225, ¿no? Escribe en el pizarrón [*F(3) = 0,225*]

Profesor: ¿Qué dice Brenda?

Brenda: Pero lo que pide es poner… ¿cómo es la ecuación?

Blanca: ¿Una función nada más se expresa en forma de ecuación?

Brenda: No.

Blanca: ¿De qué otra forma?

Mónica: Por ejemplo. No, una función no necesariamente tiene que ser una ecuación, por ejemplo está podría ser una función [*escribe F(x) = 3x*], siendo que no es una ecuación.

Profesor: Pero esa [*se refiere a una ecuación*] sería una expresión algebraica, ¿puede algo que no sea una expresión algebraica, una gráfica o una tabla, ser función?, ¿tiene que ser una fórmula?

Mónica: La tabla o la gráfica es la representación de una función.

Brenda está especialmente preocupada por contestar las preguntas de acuerdo a las instrucciones planteadas, así, cuando se le pide escribir en notación funcional, no estará de acuerdo con Mónica porque no está utilizando la notación dada anteriormente. Después se dan cuenta que están utilizando diferentes nombres para la misma variable, *F(x)* y '*y*', optan por dejarle sólo *F(x)* porque consideran que con esa letra (*F*) harían notar que es una función. La notación funcional la utilizan bien, pero mecánicamente. Recurren al lenguaje hablado para después usar el lenguaje escrito. Mónica deja ver que su concepción de función está vinculada con esta notación.

En el pasaje 14 colocan de manera correcta las variables dependiente e independiente, probabilidad y número de hijos en la gráfica de la distribución, en el eje cartesiano aunque espontáneamente hay un titubeo en tomar el número de hijos como la variable dependiente. Utilizan los valores decimales de la probabilidad, no los fraccionarios. No prevén los valores posibles de su variable dependiente para hacer su gráfica más visual, así que les queda amontonada en la parte de abajo. Unen los puntos por líneas.

[*Mónica hace unos ejes cartesianos en el pizarrón. En su sistema coordenado sólo incluyen el primer cuadrante*].

Brenda: Aquí sería el número de hijos [*señala el eje de las abscisas*].

[*Mónica etiqueta los ejes: número de hijos y número de trabajadores*].

Brenda: No, pero entonces ahí sería la probabilidad [*se refiere al nombre de la variable dependiente*].

Brenda: Aquí [*señala el eje de las abscisas*] sería, uno, dos,… así hasta nueve.

[*Mónica pone la escala en el eje de las abscisas*].

Brenda: Luego para arriba, ¿de qué la empezamos? ¿de 0,1, no?

Brenda: Bueno, sería de cero hasta uno y luego ya sería irle poniendo las probabilidades. A ver, divídelo…

Así mismo se presentan otros problemas vinculados con el lenguaje. Por ejemplo, no ven la necesidad de expresar en la gráfica, más fielmente, el problema. Muy probablemente se debe a que en el momento que hicieron el análisis ellas sabían perfectamente la vinculación entre el problema y la matemática. De modo que la gráfica no resulta ser algo que les sirva para representar la situación problema (pasaje 16).

Otro problema ya mencionado es que prefieren llamar probabilidad al número de trabajadores. Esto se observa en diversos pasajes, pero principalmente en el 4.

8.3.5. Discusión de los resultados obtenidos en el objeto variable aleatoria

El análisis de la forma en que las alumnas transitan de un contexto a otro al tratar de profundizar en la herramienta matemática empleada en la solución del problema, nos lleva a explicar algunas de sus dificultades a partir de la modelación desde la perspectiva de Heitele (1975), quien considera que en el momento que se genera un modelo matemático, éste reemplaza a la «realidad» como sustrato operacional y por ello propone representar las relaciones entre realidad y herramienta matemática en una estructura de estratos. También tomamos en cuenta la propuesta de Wild y Pfannkuch (1999), quienes indican que los elementos matemáticos pueden llegar a interiorizarse que llegan a ser parte integral del pensamiento.

Nuestro modelo tiene dos estratos de modelación en los que los valores de la variable aleatoria juegan dos papeles diferentes:

- *Primer estrato:* Vinculación de los eventos elementales con los valores de la variable aleatoria (aplicación de la regla de correspondencia que la define). En este estrato, la variable aleatoria es propiamente el «modelo matemático». Sus valores numéricos son el resultado de aplicar la regla de correspondencia sobre el espacio muestral, que constituiría la «realidad». Hay un proceso de descontextualización.

  En realidad este proceso involucraría un estrato de modelación anterior en el cual el resultado sería la generación de un espacio de probabilidad, en el que se definirían tanto el espacio muestral como la probabilidad relacionada. El espacio muestral es parte del «modelo» resultante y la «realidad» del primer estrato de modelación.

- *Segundo estrato:* La asignación de una probabilidad a cada valor de la variable aleatoria (la construcción de la distribución de probabilidad). Aquí los valores de la variable aleatoria vinculan al modelo matemático (la distribución de probabilidad) con la realidad. El proceso de construcción de la distribución de probabilidad seguido es complejo (ver sección 8.2.3 de este capítulo y Figura 5.8).

Desde una perspectiva matemática el proceso al que nos estamos refiriendo en el segundo estrato implica una composición de funciones no expresadas a través de ecuaciones, sino a través de varios registros incluyendo la descripción verbal. Según Engelke, Oehrtma y Carlson (2005) la comprensión de la composición de funciones requiere que los estudiantes tengan un concepto de función completo, estático como objeto y dinámico como proceso, para coordinar el uso de una función como variable de otra (pasar de un estrato a otro dentro del proceso de modelación). También mencionan que el aprendizaje se dificulta cuando la composición de funciones exige cambios de registro y que uno de los que presenta mayores dificultades es la *descripción verbal*. Esto puede explicar las dificultades con la variable aleatoria para lograr verla como un objeto sobre el cual se puede operar.

En el momento en que las estudiantes logran vincular los valores de la variable aleatoria con la distribución de probabilidad, también afirman que los valores de la variable aleatoria no tendrían por qué estar ordenados en forma ascendente. Los interpretan como los resultados de repetir varias veces un experimento aleatorio, en lugar de verlos como números reales. A este respecto, Miller (1998) crítica la definición escolar de variable aleatoria como un valor numérico resultante del experimento aleatorio, indicando que el valor numérico no siempre es un «resultado» del fenómeno aleatorio. Sin embargo en esta investigación observamos que interpretar los valores de la variable aleatoria como el resultado del experimento aleatorio, que es el *primer estrato de modelación*, les ayudó a las estudiantes a descontextualizar los valores de la variable aleatoria del conjunto de trabajadores con un cierto número de hijos. Cuando argumentan sobre la aleatoriedad del proceso, interpretan los valores de la variable como su «realidad» en la distribución de probabilidad (*segundo estrato de modelación*). Este resultado también apoya la propuesta de Parzen (1971) sobre la forma de introducir la variable aleatoria (sección 5.3 del Capítulo 4) como el resultado de un experimento aleatorio.

Sin embargo en nuestro estudio, este también favoreció que el doble papel que juegan los valores de la variable aleatoria (como resultado de la aplicación de la regla de correspondencia y como argumento de la distribución de probabilidad) confundiera a las estudiantes. Puesto que cuando toman los valores de la variable aleatoria como el resultado del experimento aleatorio, nuevamente están pensando en eventos, no en números (reales) y por lo tanto la relación que establecen entre la probabilidad y los valores de la variable aleatoria sigue siendo una función de conjunto (entre la probabilidad y sus eventos) y no una función vinculada con una variable numérica. De acuerdo con los resultados de las investigaciones de las dificultades sobre la composición de funciones de Ayers, Davis, Dubinsky y Lewin (1988), se requeriría que las estudiantes lograran un grado de abstracción de la variable aleatoria como objeto estático para hacer inmediato el camino a través de la inversa de la variable aleatoria y llegar a la distribución de probabilidad. Pero la abstracción reflexiva en ellas no se concretó suficientemente como lograrlo.

Por otro lado, las estudiantes mostraron una comprensión deficiente de la distribución de probabilidad al adjudicarle un carácter aleatorio, pues aunque la probabilidad se aplica a sucesos aleatorios, una vez definida matemáticamente, tiene un valor fijo. Asumimos que el razonamiento erróneo mostrado en las estudiantes es consecuencia de conocimientos y razonamientos matemáticos exitosos en contextos determinísticos. Se presenta una fuerte convicción del poder de las operaciones formales con expresiones algebraicas (Sierpinska, 1992) que proviene de dos creencias primarias: (1) las ecuaciones no están vinculadas con datos extraídos de la realidad y (2) una ecuación debe poder generalizarse para servir en varios contextos (Figura 5.10). Alrededor de esta idea surgen otras dos.

Por un lado, como se observa en la Figura 5.10, las estudiantes argumentan la inutilidad de la distribución porque no serviría para «*predecir*» o «*predeterminar*» (palabras usadas por ellas) el valor de la probabilidad de otras fábricas. Sin embargo, es posible que si la situación problema planteada correspondiera a una distribución conocida (como la binomial), ellas hubieran determinado su fórmula sin preguntarse si serviría en otros contextos. Probablemente no vieron la utilidad de descontextualizar la herramienta matemática que emplearon para resolver el problema, porque su objetivo estuvo únicamente en la solución del mismo y no fue necesario.

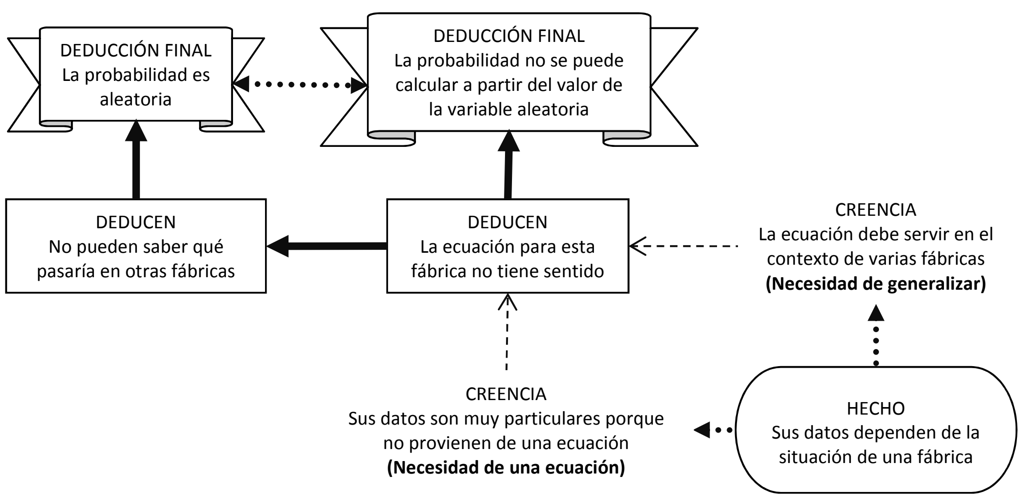


Figura 5.10. Razonamiento seguido por las estudiantes para vincular la aleatoriedad con la incapacidad de cálculo de la probabilidad

Por otro lado, cuando adjudican aleatoriedad a la probabilidad están pensando en un nivel más amplio: encontrar una ecuación que sirva para muchas fábricas. El experimento aleatorio sería la selección de una fábrica al azar; la característica de interés, observar en ella la distribución de las frecuencias de los trabajadores que tienen un mismo número de hijos; el espacio muestral serían las fábricas de una ciudad o un país. En esta nueva situación nuestra distribución de probabilidad sería una distribución marginal y pierde su relación con el problema que nos interesa, que se restringe a una fábrica en particular. Su intento infructuoso por tratar de descontextualizar la distribución de probabilidad finalmente, desemboca en relacionar la aleatoriedad con una incapacidad de cálculo.

Históricamente, el concepto de función es protagónico a partir de que se le concibe como una fórmula, es decir, cuando se logra la interacción de los dominios de representación geométrico y algebraico. Desde la perspectiva tradicional, en la enseñanza de éste concepto se tiende a sobrevalorar los aspectos analíticos y los procedimientos algorítmicos dejando de lado los argumentos visuales o los enfoques numéricos, por no considerarlos, entre otras causas, procesos plenamente matemáticos (Cantoral y Farfán, 1998). Por lo que la necesidad de las estudiantes de buscar una ecuación que representara sus datos podría provenir de la forma en cómo han aprendido el tema de funciones.

Por último, creemos que las dificultades vinculadas al lenguaje son aplicables al lenguaje matemático en general y a sus creencias sobre las matemáticas. Así, de alguna manera, ven la gráfica como un objeto que no está relacionado con el contexto del problema, algo matemático que vive en sí mismo y no con un fin práctico.

8.4. Objeto de análisis: Solución del problema

En este apartado se analizarán las estrategias y rutas de solución para resolver el problema, las principales dificultades, la simbología y el lenguaje utilizados y los objetos matemáticos utilizados. Puesto que algunas estrategias exigen razonamientos que se discuten en pasajes muy largos de la entrevista, no incluimos todos pasajes en nuestro análisis, sino remitimos al Anexo 3 para su lectura.

8.4.1. Estrategias y recursos en la solución del problema

En el pasaje 8 se observa que la solución inmediata es determinista, pues las estudiantes desvinculan la decisión de comprar el número de boletos con el sorteo y quieren que su recomendación sea exacta. Se les dificulta un poco dar una recomendación en términos de lo más probable. En su solución inicial no usan los datos de la tabla, sólo expresan que comprarían dos y esperarían el resultado de la rifa para comprar los que hagan falta. Esta postura es defendida principalmente por Mónica.

En el pasaje 9 Brenda sugiere comprar 5 boletos y argumenta que es el valor más probable (puesto que 3 es la moda de la distribución de probabilidad y añade dos boletos para los padres). Mónica insiste en comprar sólo 2 para no hacer perder a la empresa, sin embargo cambia de opinión con un supuesto que antes no habían utilizado: los boletos tienen que comprarse por anticipado. El profesor tiene que imponer ese supuesto, pero es Mónica la que se da cuenta que necesita conocer ese dato del que no se habla explícitamente en el problema.

Mónica: Si, hay mucha probabilidad. Pero y si te salen por ejemplo estas [*señala los trabajadores con menos de tres hijos*]. Ah profe, es que eso es lo que le preguntaba; lo que yo le estoy preguntando es muy importante porque, en caso de que yo compre dos me van a faltar, pero si tengo que comprar los exactos que lo menos pierda o puedo comprar ya después boletos.

Profesor: Ya no, después ya no.

Mónica acepta que se deben comprar 5 boletos porque es la solución más probable. Al parecer hay un acuerdo entre ambas. Pero en el pasaje 10 hay un cambio de opinión de Brenda, quien duda porque comienza a hacer uso de la distribución de probabilidad (en la tabla que tiene escrita en el pizarrón). Lo primero que expresa es su preocupación por que el trabajador ganador «*no vaya a completar después*». Ella comienza a ver otras características a partir de la interpretación de la distribución, como el complemento de la probabilidad acumulada.

En este pasaje se observa como ambas estudiantes se posicionan en el problema y se produce la devolución. Se entabla una discusión en la que Brenda, para argumentar la compra de 5 boletos, usa de manera intuitiva tanto la probabilidad como la frecuencia acumulada, «*el número de trabajadores hasta aquí*». Mónica en un principio usa la moda (como el valor más probable, «*el que menos falla*») y después recurre al argumento de que se debe gastar lo menos posible, que no se había mencionado, pero que, probablemente, es lo que haya impulsado a Mónica a su solución inicial (comprar sólo dos boletos).

En contraste con el argumento de gastar menos, surge otra solución obvia posible que se descarta de inmediato: comprar 11 boletos para no introducir riesgo alguno («si *yo quiero que nadie se quede sin boletos, pues entonces de una vez compran los... 11...de una vez para ya no errarle*»).

Mónica presta menos atención al argumento apoyado en la probabilidad (o frecuencia) acumulada que a la preocupación de Brenda porque la familia premiada tenga más miembros que los boletos que recibirán. Hay un pequeño debate entre todos los argumentos generados hasta ese momento. No se llega a un acuerdo porque no usan la misma herramienta matemática (Mónica usa la moda, Brenda la probabilidad acumulada y la mediana). El profesor les pide que escriban el cálculo de la probabilidad acumulada (los cálculos que Brenda sólo había hecho eran solo mentales y aproximados) y se rompe la discusión, porque interpretan la intervención del profesor como una guía velada sobre lo que espera de ellas. No vuelven a utilizar la moda como argumento. De manera natural ellas no deciden calcular la probabilidad acumulada ni la del complemento, aunque las usan, sólo las estiman. En los argumentos utilizados por Brenda en esta parte también hay una referencia a la mediana de la distribución, pero, de la misma forma, no se detienen en ella.

En el pasaje 11, Mónica también usa la probabilidad acumulada para argumentar cuántos boletos comprarían «*si tienes cubierto todo esto*». Dudan y no se convencen mutuamente. Considerando la misma herramienta matemática, se observa claramente la posición de cada una. Mónica, argumenta que se debe ahorrar lo más posible (viendo la empresa) y Brenda intenta que el número de boletos alcance (tratando de ponerse del lado del trabajador). No pueden desligarse de su posición dentro del contexto. Argumentan que hay una decisión que depende de la ética, que no tiene nada que ver con matemáticas y se quejan con el profesor del planteamiento del problema.

Después, aparentemente hay un acuerdo entre ambas, que implica también una aceptación de la aleatoriedad de la situación y un riesgo de ambas partes, empresa y trabajadores.

Brenda: En todo caso si se está buscando que no vaya a gastar de más y si alcanzan bueno y si no, no, pues yo estaría por 4 hijos.

Mónica: Ajá, yo también y si no alcanzan pues ya ni modo.

Ante la respuesta de comprar 6 boletos (4 para los hijos y dos para los padres), nuevamente el profesor interviene. Les indica que la recomendación tiene que estar en lo mínimo que se pueda gastar. De esa forma ambas llegan a un consenso: comprarían 5 boletos con una probabilidad de 0,58 de que los boletos les alcancen exactamente o sobren. Sin embargo esta respuesta está condicionada por la intervención del profesor.

8.4.2. Discusión de resultados obtenidos en el objeto solución del problema

Las estudiantes proporcionan diferentes soluciones utilizando diferentes herramientas y bajo distintos argumentos, también van incorporando supuestos a medida que descartan soluciones y se apropian del problema (Tabla 5.8).

Tabla 5.8. Síntesis de las soluciones proporcionadas por las alumnas

Solución	Argumento	Herramientas	Motivo que las hace descartar la solución
2 boletos	La empresa debe perder lo menos posible	No usan los datos de la tabla de probabilidades Solución determinista	Cuestionan al profesor sobre si es posible comprar boletos después de la rifa. El indica que no
5 boletos	Es más probable que el trabajador premiado tenga 3 hijos	Usan la moda de la distribución de probabilidad	Se preocupan por los trabajadores que tienen muchos hijos
11 boletos	Es seguro que le alcanzarían los boletos al trabador premiado	Usan la probabilidad acumulada y la noción de evento seguro	La empresa correría el riesgo de perder demasiado. Es descartada de inmediato
6 boletos	Se busca favorecer a los trabajadores pero también que la empresa no gastarte de más	Hay una interpretación de la función de distribución a través del uso de la probabilidad acumulada y de su complemento Aparece el uso de la mediana de la distribución de probabilidad Hay una aceptación de la aleatoriedad como riesgo	El profesor interviene recordando que se debe gastar lo menos posible
5 boletos	Se busca favorecer a los trabajadores pero que la empresa se arriesgue lo menos posible.	Hay una interpretación de la función de distribución a través de la probabilidad acumulada y de su complemento No queda claro si recurren a la moda o a la mediana Hay una aceptación de la aleatoriedad como riesgo	Ésta es la solución que ellas finalmente proporcionan.

Observemos que su segunda solución (5 hijos) no es igual a su última solución (también, 5 hijos) pues en esta última cambian los argumentos y usan herramientas estocásticas. Necesitaron de ese tránsito entre varias soluciones para hacer uso de esas herramientas y ser conscientes del riesgo de su decisión.

Así mismo hay una intervención del profesor que hace que ellas se sientan seguras de su solución y no la cuestionen más. Si el profesor no hubiera intervenido, es posible que la solución final hubiera sido comprar 6 boletos, que fue su conclusión después de su pequeña discusión. De alguna manera fue la que les hizo sentirse bien con los trabajadores pero que no cumple con las condiciones del problema (porque en el problema se especifica una mínima inversión). Notemos que las dos soluciones finales no son muy distintas en cuanto a la argumentación matemática que utilizan, la diferencia está en los riesgos que se aceptan para la empresa y los trabajadores. Aunque el mínimo riesgo para la empresa sería 2 boletos, habría poca posibilidad de favorecer a los trabajadores. El problema es abierto, por tanto, no hay una única solución, pero la elección de 5 boletos es preferible porque la probabilidad de que el trabajador quede favorecido es mayor a 0,5 y a la vez, la empresa no se arriesga excesivamente.

La discusión de las estudiantes puso en claro los momentos en que se introducen los criterios de decisión y cómo se están poniendo en juego. Si el único criterio hubiera sido favorecer al trabajador, la decisión hubiera sido determinista. La respuesta de comprar 6 boletos demuestra que ellas preferían que la empresa gastara más a dejar a una familia con boletos insuficientes. Así, ellas no usaban la función de distribución acumulada para medir la probabilidad de que la empresa gastara de más, sino la probabilidad de que al trabajador ganador le alcancen los boletos.

Sin embargo, en la discusión de las alumnas también mencionaron las posibilidades de que la empresa gaste de más o al trabajador ganador no le alcancen los boletos cuando se referían al complemento de la distribución (Tabla 5.9). Esta función sería más importante si cambiamos el orden de los criterios de decisión y se quisiera favorecer a la empresa antes que al trabajador. En tal caso se comprarían 4 boletos (Figura 5.11).

Tabla 5.9. La distribución de probabilidad del número de hijos, función de distribución y el complemento de la función de distribución

Número de hijos (x)	Distribución de Probabilidad	Función de distribución $F(x)=P(X \le x)$	Complemento $1-F(x)$
	Probabilidad de que el trabajador premiado tenga x hijos $p(x)=P(X=x)$	Probabilidad de que al trabajador premiado le alcancen los boletos si se compran x boletos	Probabilidad de que la empresa no gaste de más. (Probabilidad de que al trabajador premiado no le alcancen los boletos)
0	0,080	0,080	0,920
1	0,110	0,190	0,810
2	0,165	0,355	0,645
3	0,225	0,580	0,420
4	0,155	0,735	0,265
5	0,100	0,835	0,165
6	0,060	0,895	0,105
7	0,045	0,940	0,060
8	0,035	0,975	0,025
9	0,025	1,000	0,000

Es factible cambiar el orden de los criterios de decisión porque no podemos encontrar un número de boletos que proporcione exactamente el 50% (eso garantizaría ecuanimidad), de modo que se tiene que favorecer a una de las dos partes. Desde la perspectiva ética, el problema se trabajó pensando en la recomendación sobre cuál es la situación que favorecería más a los trabajadores y en la que se gastaría menos. A partir de esa recomendación, los directivos serían los encargados de decidir cuántos comprarán.

Observemos que la entrevista obligó a las estudiantes a entrar constantemente en el ciclo interrogativo (Wild y Pfannkuch, 1999). Esto hizo que ellas obtuvieran una solución del problema satisfactoria, pero no podemos afirmar que se haya tenido el mismo éxito en la apropiación de la herramienta probabilística que usaron.

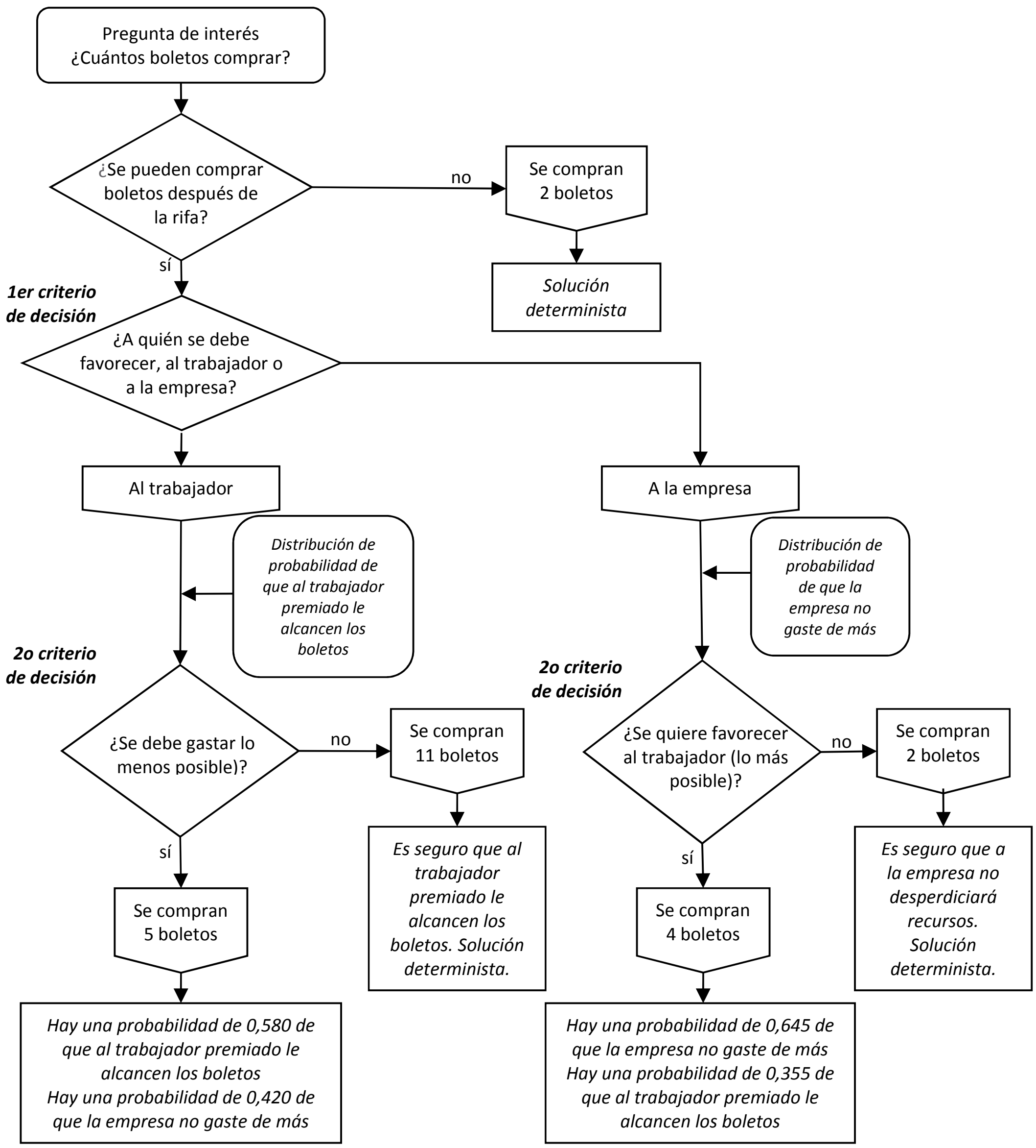


Figura 5.11. Ruta de solución tomando en cuenta los criterios de decisión posibles

9. Conclusiones del capítulo

El objetivo en la entrevista clínica fue analizar el proceso cognitivo de dos estudiantes al enfrentarse a un problema en el que se ponía en juego la variable aleatoria. No perseguimos documentar una apropiación formal del objeto, pero sí explorar si en la actividad es posible llegar a un nivel de comprensión tal que dé acceso a los estudiantes no sólo a resolver problemas como el que se les plantea en esta experiencia, sino también a algunas herramientas estadísticas y probabilísticas sencillas. Nuestra búsqueda se enfoca a hacer accesible a los estudiantes la estructura teórica necesaria para darle sentido a la experiencia empírica descrita en el problema. La experiencia es sólo un primer paso para iniciar un estudio posterior sobre la variable aleatoria que les permita penetración y discernimiento en otros ámbitos (Wild y Pfannkuch, 1999), pero también dentro de la misma estadística.

En las conclusiones primeramente confrontamos las hipótesis planeadas con los resultados obtenidos y después analizaremos la idoneidad del protocolo de la entrevista clínica, el problema y la forma en que fue planteado para cumplir el objetivo planteado.

9.1. Confrontación de los resultados con las hipótesis

En el Punto 4 de este capítulo se plantearon las hipótesis sobre lo que esperábamos encontrar en la entrevista. A continuación discutimos las conclusiones que pueden deducirse del estudio de los resultados empíricos.

Falta de percepción de la aleatoriedad del proceso

Esta hipótesis no se llega a confirmar, pues las estudiantes manejaron un buen concepto de aleatoriedad, al contrario de lo reportado en otras investigaciones, como la de Serrano (1996). Manifestaron no poder predecir cuál sería el número de hijos que tendría el trabajador premiado (impredecibilidad del resultado aislado) y que si repitieran el sorteo muchas veces, no sabrían cuál es la secuencia que seguirían los resultados (falta de patrón o de orden).

La aleatoriedad del proceso aleatorio (el sorteo) no fue cuestionada. Sin embargo en las primeras etapas del trabajo, hubo reticencia a dar una recomendación que no sabían si sería certera, es decir, que su solución «heredara» la aleatoriedad del sorteo. Una vez aceptada la incertidumbre, buscaron argumentos de equidad (entre los trabajadores y la directiva de la empresa) para proporcionar una solución con la que pudieran sentirse más confiadas. Este argumento de equidad se obtuvo a partir del uso que hicieron de la probabilidad acumulada.

Sin embargo en un intento por formular la distribución de probabilidad, vincularon la aleatoriedad con una incapacidad de cálculo (no tenían una ecuación que vinculara la probabilidad con la variable aleatoria y por ello no podían «calcular» la probabilidad de un número dado de hijos), de modo que más que al número de hijos, asignaron la aleatoriedad a la probabilidad. Es decir, no llegan a relacionar apropiadamente las expresiones matemáticas (fórmulas de las distribuciones) con el estudio de los fenómenos aleatorios.

Tendencia a «algebraizar» y descontextualizar los procedimientos relacionados con la noción de variable aleatoria

Esta hipótesis se confirma parcialmente puesto que se encontraron dificultades en ambos sentidos. Sin embargo, de manera general, al contrario de lo esperado, para las alumnas fue más difícil pasar del contexto del problema al contexto matemático. Varias fueron las razones que nos llevan a concluir que el proceso de contextualización es más simple para ellas que el inverso:

- Hay una tendencia a centrarse preferentemente en la frecuencia (número de trabajadores) de un valor de la variable, en lugar de en su probabilidad cuando trataron de descontextualizar la distribución de probabilidad (al inicio de la actividad) y al argumentar alguna propuesta de solución del problema.
- A pesar de que hay una interpretación apropiada de la probabilidad clásica, la perciben como un cociente de dos números en lugar de un solo número.
- Constantemente estuvieron recordando que el valor de la variable aleatoria ‘Número de hijos’ está vinculada con su imagen inversa (conjunto de trabajadores con ese número de hijos).
- En la gráfica, no incluyen valores con probabilidad nula a pesar de que habían mencionado, por ejemplo, que la probabilidad de que el trabajador premiado tuviera diez hijos sería cero.
- Manifiestan que el eje de las ordenadas (en la gráfica de la distribución de probabilidad) podría no estar ordenado siguiendo el orden numérico natural debido a la naturaleza aleatoria del experimento.
- Piensan que no es útil encontrar la ecuación de la distribución de probabilidad para el problema planteado porque se trabaja con datos demasiado «*reales*».

- ❖ No sintieron la necesidad de ir más allá del problema en la herramienta matemática empleada en su resolución. Su esfuerzo se concentró en obtener la solución al problema.

En otros pasajes desvinculan las matemáticas del contexto del problema, mostrando una concepción de las mismas como algo que no tiene porqué tener una relación con una situación del contexto real, como vemos en los siguientes argumentos:

- ❖ Piensan que una ecuación debe servir para hacer predicciones en distintas situaciones y no para una en particular.
- ❖ Creen que los datos de una función deben provenir de un contexto matemático y no de datos estadísticos (tomados de la «realidad»).
- ❖ Piensan que una gráfica debe ser continua, por lo tanto, el dominio y el rango de una función también. No importa que el contexto del problema sea discreto.
- ❖ Piensan que una gráfica siempre es infinita.

Estas evidencias de descontextualización, refuerzan la desvinculación entre los conceptos matemáticos y el contexto de un problema. Sin embargo en nuestro caso, fue más fácil para las estudiantes contextualizar que descontextualizar.

Esto coincide con el desarrollo histórico de la variable aleatoria, puesto que, a pesar del manejo de diversos recuentos de datos desde épocas remotas, tuvieron que transcurrir muchos años para que los análisis de los datos se vincularan de manera más formal con las herramientas probabilísticas y matemáticas.

Extrañeza de trabajar funciones en un contexto probabilístico

Esto también se confirma parcialmente. Las alumnas no encontraron extraño asociar el término función a la distribución de probabilidad. Esta facilidad de aceptación fue debida a que la pudieron escribir muy naturalmente en notación funcional y porque recurrieron a la definición y observaron que concordaba con ella. También fue inmediato el uso de la representación tabular para las frecuencias y probabilidades. Sin embargo sí encontraron extraño que hubiera funciones con las características encontradas para la distribución de probabilidad, por ejemplo:

- ❖ Encontraron raro trabajar con una función discontinua (una función puntual), donde la variable independiente es discreta.
- ❖ Hallan restringido el intervalo de valores de la variable independiente porque sólo toma valores entre 0 y 9.

- Encuentran excepcional no poder encontrar los valores de la variable dependiente a partir de la variable independiente a través de una ecuación.
- Se les hace extraño que la función esté definida a partir de datos estadísticos (ellas los denominan «*reales*»).
- Piensan que es novedoso que la ecuación de la función, en caso de que pudiera obtenerse, solo serviría para la probabilidad de la fábrica del problema.

La idea de que una ecuación es útil sólo si sirve para conocer la probabilidad de muchas fábricas, las lleva a la idea errónea de que la aleatoriedad está vinculada con una incapacidad de cálculo de la probabilidad. Este razonamiento es el que conlleva a la conclusión de que el valor de la probabilidad es aleatorio.

Dificultades con la noción formal de variable aleatoria

No pudimos observar la construcción de la regla de correspondencia de la variable aleatoria porque esta información se les proporcionó en el problema, pero sí se tienen conclusiones sobre el manejo y la interpretación de la variable aleatoria de las estudiantes. Ellas comprendieron que la regla de correspondencia vinculaba el espacio muestral con los valores de la variable aleatoria. Pudieron percatarse, también, que esa regla de correspondencia provenía de una característica de interés (extraída del planteamiento del problema) que agrupaba a los trabajadores en eventos compuestos. Así mismo interpretaron correctamente que la probabilidad se obtenía a partir de los eventos compuestos y que ello permitiría asignar una probabilidad a cada valor de la variable aleatoria. Así, concluimos que se tuvo una buena interpretación y manejo de los valores de variable aleatoria.

Se presentaron dificultades, sin embargo, al construir la distribución de probabilidad porque eso requería no sólo la formulación de la probabilidad como la relación entre dos conjuntos de personas (las favorables y las posibles) sino también el planteamiento de los valores de la variable aleatoria como un valor numérico y no como la característica de interés que agrupó a los trabajadores en los eventos compuestos. Esto provocó que las alumnas prefirieran relacionar los valores de la variable aleatoria con la frecuencia, más que con la probabilidad.

Pudimos observar que en el proceso de construcción de la distribución de probabilidad a partir de la situación problema, se generaron dos estratos de modelación desde la perspectiva de Heitele (1975). Desde una perspectiva cognitiva, para poder pasar al segundo estrato de modelación, es necesario que el modelo matemático del primer estrato sea abstraído por el estudiante para poder usarlo como entrada del segundo estrato. Las estudiantes no lograron una descontextualización completa del primer estrato de modelación, como se observa cuando piensan que los valores de la variable aleatoria en el eje numérico no tendrían por qué ir ordenados.

Observamos que en la tentativa por descontextualizar la distribución de probabilidad, las estudiantes pasan por el proceso de transformar el resultado del experimento aleatorio. Piensan que del sorteo se obtendrá el nombre de un trabajador sino un número de hijos. Esto hace que el experimento se transforme en un experimento con resultado numérico, proceso que se observa también históricamente (Parzen, 1971).

Por último, en su intento por descontextualizar la distribución de probabilidad, ellas desvinculan la distribución de probabilidad del experimento aleatorio y la transforman en la distribución de frecuencias. Lo que conduce al planteamiento de una función de distribución doble, en donde nuestra fábrica proporcionaría sólo una distribución marginal y los valores de sus frecuencias (probabilidades para ellas) serían aleatorios (en la distribución doble).

Dificultades respecto a los conceptos que intervienen en la definición formal de variable aleatoria

No se consiguió una definición formal de la variable aleatoria (al nivel que esperábamos) sin embargo sí hubo una buena aproximación a ella en la que intervinieron diversos conceptos probabilísticos y estocásticos.

Las alumnas usan la idea de probabilidad en forma natural, interpretándola como una medida de lo incierto, aunque la consideran una razón y no un único valor numérico. Esto contribuye a que no vinculen directamente a la probabilidad con el valor de la variable aleatoria, más que a través de los eventos compuestos, lo que representa un obstáculo de descontextualización.

La interpretación de los eventos compuestos (trabajadores con un mismo número de hijos) juega un papel muy importante en las etapas de modelación para formular la distribución de probabilidad. En la experiencia realizada su interpretación permitió la vinculación entre el experimento aleatorio y la probabilidad.

Así mismo, al visualizar los valores de la variable aleatoria como el resultado de un experimento aleatorio, se define un nuevo espacio muestral numérico que facilita el primer estrato del proceso de modelación, lo que contribuyó a la conceptualización de la variable aleatoria y la distribución de probabilidad. Sin embargo, es conceptualización equívoca de la variable aleatoria como objeto matemático y también acarreó una concepción errónea: los valores de la variable se vincularon con el proceso aleatorio, lo que dificultó su total abstracción como números reales con un orden definido.

También se observó que, como ocurre epistemológicamente, los procesos cognitivos para la construcción de la distribución de probabilidad y la variable aleatoria están intrínsecamente vinculados y que se condicionan el uno al otro.

9.2. Idoneidad de la actividad propuesta

El protocolo de la entrevista se diseñó para que girara alrededor de un problema. Esperábamos que fuera atractivo a las estudiantes para que se involucraran en su solución, pero que también las obligara a transitar por los objetos matemáticos de estudio. La profundidad de las herramientas matemáticas estuvo condicionada por los conocimientos previos de las alumnas, tomado en cuenta que no habían llevado cursos de estadística universitaria, y también por el manejo del objeto esperado en un curso de estadística en una carrera de ciencias sociales en el ITESM.

A continuación se analiza la idoneidad didáctica del problema junto con el protocolo de la entrevista. Esta idoneidad incluye la articulación de los cinco componentes que definen Godino, Wilhelmi y Bencomo (2005) y Godino, Contreras y Font (2006): idoneidad epistémica, cognitiva, interaccional, mediacional y emocional.

- *Idoneidad epistémica*. La actividad permitió que las alumnas avanzaran hacia la conceptualización de la variable aleatoria y trabajaran con los objetos relacionados con ella, así mismo que se cuestionaran sobre sus concepciones que intervienen para construir el concepto de variable aleatoria. No obstante algunos objetos matemáticos vinculados con la variable aleatoria, como sus momentos o la variable estadística, no fueron considerados.
- *Idoneidad cognitiva*. La situación movilizó el aprendizaje de las alumnas. Encontramos ideas espontáneas correctas sobre aleatoriedad y probabilidad, no tuvieron dificultad en identificar correctamente el espacio muestral ni cómo depende del experimento. Hubo una buena interpretación y uso de la probabilidad, tanto desde un punto de vista laplaciano, como desde un punto de vista subjetivo, como grado de creencia, aunque tuvieron tropiezos con la distribución de probabilidad. También reconocieron espontáneamente dos de los axiomas de la probabilidad. Mostraron correcta comprensión de los números fraccionarios y de la probabilidad como fracción y como decimal. Encontraron natural considerar aleatorio el número de hijos, manejaron bien

el conjunto de valores que toma, aunque restringieron su dominio. Usaron correcta y espontáneamente las ideas de moda, probabilidad acumulada, mediana, complemento de una probabilidad y equidad (valor esperado). También aceptaron que la probabilidad proporciona una solución al problema y la representación de su numerador en la tabla de datos por medio de frecuencias.

- *Idoneidad interaccional*. La situación problema conjugó la complicación y la simplicidad del manejo de una variable aleatoria definida de manera contextual. Esto, aunado a que era necesario desvincular los eventos y los valores de la variable aleatoria para poder descontextualizarla, hicieron que las estudiantes tuvieran acceso al problema, pero el proceso era lo suficientemente complicado como para representarles un reto que pudieron enfrentar. A su vez, la situación, junto con la entrevista, provocó el debate entre las estudiantes, quienes en todo momento estuvieron interesadas en resolver las tareas y se sintieron cómodas con ellas. El debate sirvió para explicitar sus concepciones, dificultades y estrategias. Lo que les permitió explorar sus propias concepciones, dificultades y estrategias, así como poner en juego sus concepciones matemáticas en diversos aspectos, tanto en el contexto de la probabilidad y estadística como en el de la matemática determinística. Esto también provocó que progresaran en la solución del problema.
- Por otro lado, se observaron algunos puntos para mejorar el protocolo. Al ser la moda igual la mediana en la distribución de probabilidad, fue difícil saber si las argumentaciones de las alumnas se basaban en el análisis de la moda o de la mediana, aunque esto también favoreció la discusión de los dos criterios de decisión que se ponían en juego. También nos dimos cuenta de la necesidad de hacer explícita la condición de que no se podían comprar boletos después de la rifa, necesaria para explicitar la aleatoriedad de la situación. Por el contrario no fue necesario explicitar el criterio de que había que favorecer a las familias de los trabajadores lo más posible. Al proporcionar los datos agrupados en la tabla, se impidió la observación de la variable aleatoria como regla de correspondencia de una función. Es posible que si el problema hubiera sido más abierto, se hubiera promovido una mejor exploración de la generación de la característica de interés y de la regla de correspondencia.

- *Idoneidad mediacional*. La puesta a punto de la situación requirió pocos medios, únicamente papel y lápiz y el pizarrón, además de los dispositivos para grabar la entrevista (que no serían necesarios si se trabaja esta situación sólo con fines de docencia). Deducimos que la situación tiene idoneidad mediacional, aunque por supuesto se puede ampliar dando cabida a calculadoras u ordenadores, lo que también permitiría complicar la situación planteada.
- *Idoneidad emocional*: La situación fue interesante para las estudiantes, quienes trabajaron en ella hasta completarla, a pesar del tiempo que les tomó. Pensamos que se consigue la devolución del problema a las estudiantes, quienes incluso llegaron a pensar que no se trata de un problema matemático y a tratarla como una situación a-didáctica. Ello indica una alta idoneidad emocional del protocolo.

En resumen, los puntos anteriores indican una idoneidad razonable de la situación y del protocolo de investigación. También constatamos que la situación, a pesar de que debe ser mejorada, contiene algunos elementos interesantes como base para el diseño de una futura ingeniería didáctica que recoja esta y otras situaciones que sirvan para ayudar al alumno a desarrollar la idea de variable aleatoria.

Capítulo 6.

Estudio 4.

Evaluación a través de un proyecto abierto

ÍNDICE DEL CAPÍTULO 6

1. Introducción

Esta segunda exploración de las concepciones y dificultades de los estudiantes alrededor del concepto de variable aleatoria, se realizó a través del análisis de la forma en cómo abordan el objeto dentro de un proyecto abierto. Este estudio complementa al descrito en el Capítulo 5 en el análisis cognitivo y difiere de él en el método utilizado, así como en la configuración epistémica de los objetos matemáticos asociados a la variable aleatoria en los que centramos nuestra atención. La intención es evaluar los conocimientos y procesos implicados en la resolución de una tarea asignada por el profesor en situación escolar con un grupo relativamente amplio de estudiantes. Metodológicamente nos interesa la recolección de datos por escrito, en los que podamos observar el análisis y los razonamientos que surgen en los estudiantes al tratar de comparar variables aleatorias por medio de la comparación de variables estadísticas. Es decir, de realizar un proceso de inferencia informal.

Los datos se obtuvieron a través de una prueba de ensayo abierta, que fue previamente utilizada en investigaciones dentro de nuestro grupo de investigación (Godino, Batanero, Roa y Wilhelmi, 2008; Arteaga, 2011). En este trabajo, tanto esta prueba como los datos obtenidos a través de ella se analizan desde una perspectiva diferente, nos centramos en la variable aleatoria y los objetos relacionados con ella para complementar los trabajos anteriores con nuevos resultados, algunos de los cuales han sido publicados en Ruiz, Arteaga y Batanero (2009 a y b) y Ruiz, Batanero y Arteaga (2011).

La tarea consistió en hacer un informe del análisis de un conjunto de datos, obtenidos por los mismos estudiantes, para tratar de responder a una pregunta propuesta por el profesor sobre las intuiciones de los estudiantes acerca de la aleatoriedad. El proyecto fue abierto porque los estudiantes participantes tuvieron la libertad de analizar los datos en la forma que considerasen mejor, incluyendo o no gráficos, tablas o resúmenes estadísticos.

La actividad involucra a la variable aleatoria, tanto en la fase de recogida de datos, como en el trabajo de análisis de los estudiantes y en la elaboración de sus conclusiones. Además, hace surgir la variable estadística y a diferentes objetos matemáticos asociados a ella. También es una situación potencialmente rica para analizar el uso que el estudiante hace de la variable aleatoria y estadística en una situación de modelación.

Al tratarse del análisis de una producción escrita, nos permite acceder a un mayor número de alumnos que en el caso de la entrevista clínica, por lo que, presuponemos, los resultados que se obtengan pueden ser más generalizables. Como contraste, puesto que el análisis de los razonamientos de los estudiantes es indirecto, a través de sus producciones escritas, nos será más difícil acceder al nivel profundo del razonamiento del estudiante. Dicho de otro modo, los resultados de este capítulo, en comparación con los presentados en el Capítulo 5 tendrán mayor fiabilidad pero menor validez.

La estructura de este capítulo es similar a la del anterior. Se inicia con los preliminares del estudio y la planeación y el diseño del protocolo de observación. Éste se concreta en un proyecto que será entregado y resuelto por los estudiantes y que, junto con las hipótesis, establecerán las bases del análisis de las producciones de los estudiantes. Se presenta la solución del proyecto propuesto, en donde es menester analizar los datos obtenidos por los estudiantes durante la realización de la actividad en clase, y se destaca la relación que se establece entre la variable aleatoria y la estadística. Los objetos de análisis del estudio se describen antes de presentar los resultados y su análisis. Estos últimos, se organizan y exponen a un tiempo y de acuerdo a los objetos de análisis. Finalmente, en las conclusiones, se confrontan los resultados obtenidos con las hipótesis de investigación y se analiza la idoneidad del proyecto planteado como base del estudio.

2. Preliminares

En este apartado situaremos el estudio en el entorno en el que fue realizado. Hacemos explícito el contexto escolar en el que se desenvolvían los estudiantes que participaron en la investigación, las características de los estudiantes que participaron en el estudio y el contexto así como la descripción de las condiciones en que se llevó a cabo la situación de enseñanza.

El proyecto abierto se aplicó a estudiantes que se encontraban cursando el segundo año del antiguo plan de estudio de la Diplomatura de Magisterio en la especialidad de Educación Primaria de la Facultad de Ciencias de la Educación de la Universidad de Granada, España.

2.1. Contexto escolar

El estudio se realizó con estudiantes de la asignatura «Currículo de Matemáticas en la Educación Primaria» en la Universidad de Granada, en el curso 2007-2008. Esta asignatura era obligatoria en el anterior plan de estudios y tenía un carácter eminentemente práctico (dos tercios de las horas docentes se dedicaban a actividades prácticas) y su contenido exclusivo era la Didáctica de la Matemática enfocada a la matemática incluida en el currículo de educación primaria en España. Abarcaba los cuatro bloques temáticos de la educación primaria: *Números y operaciones*, *La medida: estimación y cálculo de magnitudes*, *Geometría y Tratamiento de la información, azar y probabilidad* (MEC, 2006). Debido a la adaptación de los planes de estudio por el Tratado de Bolonia, esta asignatura ya no se imparte como tal, aunque su contenido se conserva en la asignatura de Enseñanza y Aprendizaje de las Matemáticas y tiene una extensión mayor (6 créditos) a partir de la generación 2011-2012.

Los estudiantes participantes en el estudio tenían en su haber pocas experiencias con asignaturas de probabilidad y estadística. La mayoría procedían del bachillerato especializado en humanidades, en el que los contenidos matemáticos son menores y las matemáticas son opcionales en sus últimos cursos. Algunos estudiantes provenían del bachillerato de ciencias sociales, y en este caso, tuvieron un curso de estadística en el cuatrimestre previo a su ingreso a la Facultad de Educación. Este curso incluye el concepto de variable aleatoria y una introducción a la inferencia estadística. No es habitual que un estudiante de la asignatura en la que se realizó esta parte de la investigación haya cursado el bachillerato con especialidad en ciencias, pero en tal caso, aunque hubieran cursado matemáticas durante el año anterior a su ingreso a la Facultad, no se incluye la estadística dentro de su currículo. No hay que obviar que, independientemente del bachillerato que hayan cursado, se espera que durante la enseñanza primaria (6-12 años) y enseñanza secundaria obligatoria (12-16) los estudiantes hayan recibido una enseñanza mínima de estadística que incluye los mismos conceptos trabajados en el proyecto.

En el primer año de la Diplomatura de Magisterio especialidad de Educación Primaria, los estudiantes cursaron la asignatura obligatoria de Matemáticas y su Didáctica (asignatura de 9 créditos, cada uno de los cuales equivale a 25 horas de trabajo del estudiante, incluido el trabajo dentro y fuera del aula). En ese curso se dedicaban 2 créditos (50 horas del estudiante, 10 de ellas dentro del aula) al bloque de *Tratamiento de la información, azar y probabilidad.* Los estudiantes revisaban y ampliaban su información sobre datos, variables estadística, distribuciones de frecuencias, gráficos sencillos (diagramas de barras, polígonos de frecuencias, histogramas, gráficos de sectores), medidas de posición central (media, mediana y moda) y medidas de dispersión (rango y desviación típica). Igualmente se revisaban y ampliaban las ideas de aleatoriedad, probabilidad y la asignación de probabilidades mediante regla de Laplace o probabilidad frecuencial. En el curso en el que aplicamos la investigación, los estudiantes previamente habían trabajado con un proyecto sencillo de análisis de datos, aunque sólo una vez y los datos no fueron recogidos por ellos, sino dados por el profesor. No se dedicó tiempo al estudio formal de la variable aleatoria.

En resumen, los estudiantes participantes tenían ideas elementales de estadística descriptiva y probabilidad, por lo que podemos considerar que la actividad propuesta en el proyecto fue el primer encuentro con la idea de variable aleatoria para la mayoría.

2.2. Descripción de la muestra.

Como ya se mencionó, los estudiantes participantes en esta parte del estudio se preparaban para ser maestros de niños de primaria entre 6 y 12 años. En general, sus edades oscilaban entre los 18 y 20 años y en su mayoría tenían 19 años, aunque también había estudiantes de mayor edad, que cursaban la asignatura excepcionalmente. Como maestros generalistas, no estarán especializados en matemáticas, sino que en su vida profesional tendrán que impartir las diferentes materias (matemáticas, lengua, ciencias naturales, ciencias sociales, etc.) dentro de su aula. El estudio se aplicó a 111 estudiantes del curso 2007-2008.

En dicho curso de la asignatura del Currículo de Matemáticas en Educación Primaria hubo un total de 318 alumnos matriculados y 50 de ellos (15,7%) eran repetidores. De entre ellos se eligieron tres grupos naturales que formaron la muestra de nuestro estudio. Estos grupos tenían características similares al total de alumnos matriculados.

Respecto a la competencia matemática de los futuros maestros, tanto en nuestra experiencia como en varias investigaciones hemos constatado sus dificultades. Por ejemplo, Hernández, Noda, Palarea y Socas (2003) describen un estudio de evaluación inicial sobre conocimientos y destrezas matemáticas básicas que aplicaron a una muestra de 883 estudiantes de siete universidades. En algunos ítems, los porcentajes de acierto sobre «números y operaciones» y «medida», sólo llegan al 42% y 48%. Los alumnos presentan grandes dificultades en proporcionalidad, fracciones y decimales; confunden los conceptos de longitud, perímetro, área y volumen.

En relación a la estadística, Estrada, Batanero y Fortuny (2004) en un estudio con una muestra de 367 estudiantes de magisterio indican que pocos apreciaron el efecto de un valor atípico en el cálculo de la media o son capaces de discernir cuándo un valor es atípico para un contexto dado. Un 40% tuvo dificultad en comprender una estimación frecuencial de la probabilidad. Otros errores con alto porcentaje fueron la confusión entre correlación y causalidad, no apreciar el tamaño de la muestra en su relación con el muestreo, no apreciar el sesgo en el muestreo y no saber cuál medida de posición central elegir, en caso de distribuciones asimétricas.

Estos ejemplos, así como otros descritos en las investigaciones de Moreno, Hernández y Socas (2010), indican las dificultades matemáticas de los futuros profesores españoles que, presuponemos, son compartidas por los participantes en nuestro estudio.

2.3. Descripción de la situación de enseñanza.

La actividad que se analizará en este capítulo, fue parte de una práctica en la asignatura mencionada, a la que se dedicaron dos sesiones de clase con una duración de dos horas cada una y se relaciona con el cuarto bloque del currículo de matemáticas en la educación primaria: *Tratamiento de la información, azar y probabilidad.* Los estudiantes realizaron la actividad individualmente y produjeron un informe escrito fuera de su horario lectivo. La actividad fue coherente con la metodología de enseñanza universitaria implantada en aquél curso en la Facultad de Ciencias de la Educación para adecuar los planes de estudio al Espacio Europeo de Educación Superior (EEES). El enunciado del proyecto completo propuesto y las instrucciones entregadas a los estudiantes se incluye en el Anexo 2.

El proyecto fue puesto en la cuarta sesión práctica del curso. Se le denominó «Evaluación de experiencias de enseñanza. Estadística en primaria» y constaba de dos componentes, de los cuales sólo el primero es de interés para nuestra investigación:

- ❖ La realización del proyecto de análisis de datos «Comprueba tus intuiciones sobre el azar», en una versión adaptada de Batanero (2001) y Godino, Batanero, Roa y Wilhelmi (2008), donde se plantea una investigación para comprobar si los alumnos tienen una buena intuición sobre la aleatoriedad. La actividad contempló dos fases: primero se trabajó en clase y después se les pidió a los estudiantes que realizaran en casa un informe de su análisis y lo entregaran en la siguiente sesión. Los informes de los alumnos son el instrumento de análisis de este estudio centrado en la variable aleatoria, mientras que el realizado por Arteaga (2011), con otros estudiantes trabajando sobre el mismo proyecto, se centra en la elaboración e interpretación de los gráficos por parte de los estudiantes.
- ❖ Una semana después, una vez recogidos los informes individuales, en la clase se organizó una actividad de análisis didáctico. En ella, los futuros profesores analizaron el proyecto, las actividades realizadas en el mismo, las posibles soluciones correctas e incorrectas y produjeron un informe sobre la idoneidad del proyecto como instrumento de enseñanza de estadística en la secundaria o en cursos de formación de profesores. Este informe no es objeto

de estudio en esta Memoria, pero se analiza con detalle en la tesis de Arteaga (2011).

Así, la práctica descrita contempla dos componentes de la formación de profesores: matemático y didáctico. En nuestro estudio de evaluación sólo nos ocupamos del componente matemático y por tanto de la primera parte de la práctica consistente en la realización del proyecto «Comprueba tus intuiciones sobre el azar».

3. Objetivos del estudio de evaluación a partir de un proyecto

Nuestra investigación se enfoca a evaluar la comprensión de la variable aleatoria y la forma en que los estudiantes la vinculan con otros objetos matemáticos en una actividad de modelación en la que están implicada la realización de inferencias estadísticas informales (Rossman, 2008). Éstas se producen al analizar las diferencias entre dos variables aleatorias que se deducen a partir de la comparación de sus correspondientes variables estadísticas. Este objetivo general se desglosa en los siguientes objetivos particulares:

- ❖ En primera instancia se pretenden evaluar las intuiciones de los estudiantes acerca de la aleatoriedad, haciendo uso de los resultados e informes obtenidos en el proyecto. Los resultados de algunos estudios (por ejemplo, Falk, 1981; Green, 1989; Konold, et al., 1991; Toohey, 1995; Serrano, 1996; Batanero y Serrano, 1999; Nickerson, 2002) indican que parte de las intuiciones de los seres humanos sobre los fenómenos aleatorios son correctas pero otra parte importante es incorrecta. El proyecto tiene por objetivo que el alumno evalúe sus propias intuiciones sobre el comportamiento de los fenómenos aleatorios, en particular sobre el lanzamiento de una moneda. Al mismo tiempo que el análisis de los datos arrojados por el grupo al querer simular un experimento reflejarán sus creencias sobre la aleatoriedad, el análisis que hagan los estudiantes sobre esos datos, también será una fuente de información para nuestro análisis.
- ❖ El segundo objetivo del estudio es evaluar las ideas de los estudiantes sobre las variables estadísticas y sus elementos, complementando, de este modo, el estudio hecho sobre las variables aleatorias en el Capítulo 5. Para responder las preguntas propuestas en el proyecto, el estudiante debe realizar un experimento aleatorio. Al realizar el proyecto se pretende que el estudiante reflexione acerca de las características de la aleatoriedad, que se percibe a la vez como ausencia de modelos y multiplicidad de modelos. Entre otros posibles modelos, es deseable que el estudiante llegue a reflexionar sobre

variables aleatorias como «número de caras en n lanzamientos», «número de rachas en n lanzamientos» y «longitud de la racha más larga» para evaluar la aleatoriedad de un experimento. Puesto que el estudiante no tiene los conocimientos matemáticos necesarios para estudiarlas directamente, debe recurrir a las correspondientes variables estadísticas asociadas al tomar datos de las variables anteriores.

- La respuesta a la pregunta sobre las intuiciones requerirá que el estudiante comparare los datos obtenidos en los experimentos aleatorios y simulados organizados en la clase. Se pretende que los alumnos agrupen los datos de forma natural, formando las distribuciones de las variables estadísticas estudiadas y usen de forma natural los gráficos y tablas estadísticas que ya conocen, así como las medidas de posición central y dispersión para comparar las variables citadas. Desde el punto de vista de la investigación *evaluaremos si el estudiante llega o no de una forma natural a la idea de distribución, sus medidas de posición central y dispersión* y sus tablas y representaciones gráficas, en qué modo compara dos distribuciones y si usa estos instrumentos para resolver el problema propuesto.
- En su tesis, Arteaga (2011) indica que para resolver el proyecto planteado, el estudiante se enfrenta a una situación de modelación y que falla en los pasos iniciales de dicho proceso, pero no profundiza en esta idea. Retomaremos esta conclusión de Arteaga, centrándonos específicamente en la *forma en que el estudiante enfrenta una situación de modelación donde debe partir de preguntas sobre la realidad para formular hipótesis, definir variables y a partir de ellas obtener un modelo matemático*. Evaluaremos hasta qué punto el alumno construye modelos adecuados para las variables estadística y aleatoria, si trabaja con sus modelos y si completa el ciclo de modelación interpretando los resultados para responder las preguntas planteadas.
- Desde otra perspectiva, complementamos el estudio sobre la modelación con el análisis de cómo los estudiantes se enfrentan a un proyecto donde es necesario recorrer las diferentes fases del ciclo de investigación PPDAC[1] propuesto por Wild y Pfannkuch (1999): formulación de un problema, refinamiento de las preguntas de investigación, recolección y análisis de los datos, interpretación de los resultados y obtención de conclusiones.

[1] PPDAC es el acrónimo que resume las etapas del ciclo: Problema-Preguntas y Plan-Datos-Análisis-Conclusiones

4. Descripción y análisis del proyecto

En este apartado se hace el análisis a priori del proyecto con fines de investigación. Esto permitirá una mejor comprensión de la actividad que se pide que haga el estudiante. Primero se hace una descripción breve del proyecto, después se especifica la forma en cómo se trabajó el proyecto en el salón de clases, se detallan los objetivos de aprendizaje de cada parte y se delimita el tipo de análisis que se espera que el estudiante haga, junto con un análisis a profundidad de las variables implicadas en el proyecto y su distribución. Posteriormente se resuelven las actividades planteadas (acorde con los datos obtenidos en la clase) y se enumeran algunas posibles soluciones por parte de los estudiantes. Esta sección se finaliza con un ejemplo del tipo de conclusiones esperadas por los estudiantes.

4.1. El proyecto: «Comprueba tus intuiciones sobre el azar»

El proyecto que usamos en esta investigación está retomado de Batanero (2001) y ha sido usado en otros trabajos de investigación (Godino, Batanero, Roa y Wilhelmi, 2008; Arteaga, 2011). En él se les propone a los estudiantes una de las tareas fundamentales en el análisis de datos que es la realización de inferencias informales (Rossman, 2008), lo que implica el uso coordinado de las variables estadística (distribución de datos) y aleatoria (distribución de probabilidad). Mientras que la variable estadística es un primer modelo matemático que representa los datos obtenidos en una muestra, la variable aleatoria supone un segundo nivel de modelación, al imaginar que la toma de datos se extiende al total de la población de donde se tomó la muestra.

La comprensión de la relación entre la distribución de frecuencias y la distribución de probabilidad permitirá al estudiante hacer inferencias informales que, finalmente, han de interpretarse en el contexto donde se tomaron los datos. Esta comprensión y coordinación del carácter dual de la distribución involucra las ideas de aleatoriedad, independencia, tendencia, valor esperado y variabilidad (Shaughnessy, 2007). Reading y Canada (2011) añaden las ideas de forma, densidad, asimetría, frecuencia relativa, probabilidad y proporcionalidad.

Varios autores sugieren la necesidad de desarrollar en los estudiantes la capacidad de realizar inferencias informales para preparar un estudio posterior más formal del tema en la Universidad. Se comenzaría con la discriminación entre posición central y variabilidad en las distribuciones de datos (Rubin, Hammerman y Konold, 2006). Después hay que deducir las características de una población a partir del análisis de los datos o bien, inferir diferencias o similitudes entre dos poblaciones a partir de las diferencias observadas en dos muestras de datos (Zieffler, Garfield, Delmas y Reading, 2008). Rossman (2008) sugiere usar la simulación para conectar la aleatoriedad que los estudiantes perciben en el proceso de recogida de datos con la inferencia del modo siguiente: a) Comenzar con una hipótesis sobre los datos; b) usar experimentación o simulación para reflexionar si los datos observados son plausibles cuando la hipótesis es cierta; y c) rechazar (o confirmar) la hipótesis inicial basándose en los resultados.

Acordes con los autores anteriormente citados, en el proyecto se les propone a los estudiantes una investigación para comprobar si los alumnos de la clase en su conjunto tienen una buena intuición sobre la aleatoriedad. En el transcurso de la misma los estudiantes recogen datos individuales y colectivos sobre un experimento de percepción de la aleatoriedad basado en las investigaciones sobre este tema de Green (1989), Green y Hunt (1992), Toohey (1995) y Serrano (1996). En la primera sesión de clase dedicada a la práctica (2 horas), el profesor plantea el proyecto, siguiendo la metodología y actividades descritas en la Sección 4.2 de este capítulo. Para poder dar respuestas a las preguntas planteadas es necesario que cada uno de los estudiantes realice un experimento que pone a prueba sus intuiciones. Se recogen los datos de toda la clase en una hoja preparada para ello. A cada alumno se le entrega una copia de la hoja y se le deja como tarea extra-clase analizar los datos obtenidos para responder las preguntas planteadas. Antes de cerrar la sesión se discute brevemente qué es la aleatoriedad, cómo se podrían evaluar las intuiciones sobre ella y qué variables sería pertinente utilizar para ello. En el Anexo 2 se encuentra la actividad completa y las instrucciones dadas a los estudiantes.

4.2. La secuencia de actividades

En esta sección se describirá con detalle una secuencia de actividades sugerida por primera vez por Batanero (2001), el tipo de análisis que se les propuso y algunas de las reacciones que se espera por parte de los estudiantes. Este análisis es una ampliación del realizado por Arteaga (2011) profundizando en particular sobre las variables aleatorias y estadísticas que intervienen.

4.2.1. Consigna inicial y discusión colectiva

El trabajo comienza motivando a los estudiantes sobre el tema de la aleatoriedad y de sus intuiciones al respecto. Se pide a los estudiantes ejemplos sobre experimentos aleatorios y no aleatorios, con los que tengan experiencia propia y se incita a la reflexión con preguntas similares a las siguientes, que se discuten colectivamente en la clase.

> **1. Consigna inicial y discusión colectiva:**
>
> *¿Tienes una buena intuición sobre el azar? ¿Cómo piensas que deberían ser los resultados de lanzar una moneda 20 veces seguidas? ¿Serías capaz de escribir 20 resultados de lanzar una moneda (sin lanzarla realmente, sino como tú pienses que debieran salir) de forma que otras personas piensen que has lanzado la moneda en realidad?. O, ¿podría otra persona adivinar que estás haciendo trampa?*

Las conjeturas iniciales de los alumnos pueden ser variadas e incluir el caso de estudiantes que crean que pueden adivinar los resultados de un experimento aleatorio o bien los pueden controlar de algún modo. Estas serían manifestaciones de la llamada *ilusión de control*[2], así como del *enfoque en el resultado aislado (Outcome approach)*[3].

Con el profesor dirigiendo, la clase discute que un estudiante podría acertar su previsión sobre el resultado de un experimento aleatorio, sin embargo, hay que resaltar, que eso sólo sería por mero azar, pues, como bien sabemos, el resultado de un único experimento aleatorio no puede preverse con seguridad. También hay que subrayar que el objetivo de la estadística es prever el comportamiento de una serie de ensayos o de una población, es decir, una distribución de valores, aunque no se pueda predecir cada valor particular. Se ponen ejemplos del uso de la estadística para realizar predicciones, por ejemplo, en las encuestas de intención de voto y explica cómo se recogen los datos estadísticos para analizarlos.

Con este planteamiento, el profesor propone la necesidad de realizar una prueba con toda la clase y se organiza un experimento del cual se extraerán datos para que cada estudiante los analice en la forma que considere mejor. También de este modo se completará el ciclo completo de investigación de Wild y Pfannkuch (1999), pues una vez planteada la pregunta (¿Son buenas las intuiciones respecto al azar?) surge la necesidad de tomar datos para responderla. Justamente reconocer la necesidad de datos en una investigación es otro de los modos de razonamiento estadístico fundamental en el modelo de Wild y Pfannkuch.

4.2.2. Realización del experimento individual

La recogida de datos en la clase se organiza en la forma que se describe a continuación. Se proporciona una pauta como la mostrada abajo para que cada estudiante escriba los

[2] La *ilusión de control* fue descrita originalmente por Langer (1975). Él observó que, en juegos de azar, como la lotería o de cartas, se podía generar en las personas la creencia ilusoria de que poseían la capacidad de poder influir en el resultado.

[3] Éste fue descrito por Konold (Konold, 1989; Konold, et al, 1991) como la actitud de las personas de enfocarse en saber cuál será exactamente el resultado del siguiente experimento, sin observar el comportamiento poblacional. Se describirá más detalladamente en la Sección 7.1.1.

datos de dos experimentos: En el primero, el estudiante inventa la secuencia de caras y cruces como él piense que podría ocurrir. Se espera que la secuencia refleje su concepción de una secuencia de resultados aleatorios (secuencia simulada). La segunda parte del experimento consiste en lanzar una moneda realmente y anotar los resultados (secuencia real).

2. Realización individual del experimento de lanzar una moneda y anotar los resultados en una pauta cuadriculada.

Vamos a comprobar qué tal son tus intuiciones respecto a los resultados aleatorios. Abajo tienes dos cuadrículas. En la primera de ellas escribe 20 resultados sin realizar realmente el experimento. En la segunda mitad lanza la moneda 20 veces y escribe los resultados obtenidos. Pon C para cara y + para cruz.

Lanzamiento simulado

Lanzamiento real

4.2.3. Las primeras variables

Cuando todos los alumnos han concluido su propio experimento se plantearán preguntas similares a la que reproducimos a continuación. Confiamos en que los estudiantes producirán ideas sobre las características de las secuencias obtenidas y que de ahí surgirán variables con las que se pueda evaluar la aleatoriedad de las secuencias inventadas por los alumnos.

3. Gestión de la clase, nuevas cuestiones y actividades:

¿Cómo podremos distinguir una secuencia realmente aleatoria de otra que hemos inventado?

Normalmente al lanzar una moneda equilibrada la primera variable visible con la que trabajamos es una variable aleatoria con distribución de Bernoulli (o distribución dicotómica) que toma valor 1 cuando ocurre, por ejemplo cara (éxito) y 0 cuando ocurre cruz (fracaso) y que cumple la condición de que los ensayos son independientes unos de otros. Dicha distribución está caracterizada por su probabilidad de éxito p y probabilidad de fracaso $q=1-p$. En este proyecto cada estudiante trabaja implícitamente con dos variables aleatorias diferentes respecto a las secuencias producidas (real y simulada):

- ❖ La variable aleatoria de la secuencia real ζ_r, que correspondería al lanzamiento de la moneda equilibrada (la variable estaría vinculada a la pregunta ¿el resultado del experimento es cara o cruz?), por tanto sus probabilidades son $p=q=1/2$. Es una variable Bernoulli porque los resultados de los lanzamientos son independientes unos de otros. Su media, sería igual a p y su varianza igual a pq. Sobre esta variable podemos analizar si el estudiante la percibe y si tiene una intuición de su valor medio.

- ❖ La variable aleatoria, ζ_s, es producto de la intuición del estudiante y estaría vinculada con la pregunta ¿qué caería ahora, cara o cruz? Toma valor 1 cuando ocurre, por ejemplo, cara (éxito) y 0 cuando ocurre cruz (fracaso) en la secuencia simulada. No sabemos si es o no Bernoulli, pues no sabemos si el estudiante produce ensayos independientes. Tampoco sabemos el valor de p, pues no sabemos si el estudiante tiene alguna preferencia por producir caras o cruces, ni en qué proporción lo hace.

Además, puesto que el estudiante repite el experimento 20 veces tenemos una serie de 20 ensayos de cada una de las variables aleatorias anteriores, y por lo tanto, cada estudiante obtiene dos variables estadísticas:

- ❖ La variable estadística X_r surge como consecuencia de obtener una muestra de 20 valores de la variable aleatoria ζ_r (el recuento de los resultados de los ensayos que realizó al lanzar la moneda).
- ❖ La variable estadística X_s surge como consecuencia de obtener una muestra de 20 valores de la variable aleatoria $\zeta_{s.}$ (el recuento de los resultados al simular haber lanzado la moneda).

Se espera que el estudiante pueda sugerir de manera natural trabajar con las anteriores variables, no en el sentido específico de que sea consciente de la existencia y diferencias entre las variables aleatorias y estadísticas, pero sí que sea capaz de enunciarlas y analizarlas informalmente como variables útiles en la evaluación de la aleatoriedad de las secuencias. En particular, suponemos que, mediante un proceso de inferencia informal, será capaz de comparar las características de las distribuciones de las variables X_r y X_s para conjeturar sobre las semejanzas o diferencias entre ζ_r y ζ_s y de ahí deducir sobre las intuiciones respecto al azar.

Seguramente algún alumno sugerirá contar el número de caras y cruces y deducir que deben ser aproximadamente iguales, ya que hay las mismas posibilidades para obtener cara que cruz. Esto mostrará una concepción correcta sobre la variable aleatoria ζ_r y la correspondiente variable estadística X_r cuyo valor medio ha de coincidir con p, puesto que la media de una muestra es un estimador insesgado de la media en una población (Ríos, 1967). También esperamos que denote un cierto grado de capacidad de modelación, pues es de suponer que el alumno trate de estudiar la veracidad de sus intuiciones analizando la similitud de las variables aleatorias ζ_r y ζ_s y estas mediante las correspondientes variables estadísticas X_r y X_s. Es decir, usa un modelo matemático para responder a una pregunta sobre la realidad.

4.2.4. Variables vinculadas al número de caras

Sin embargo, el análisis de las variables anteriores involucraría analizar un solo caso (el propio estudiante) en donde, para las variables estadísticas, los 20 lanzamientos son repeticiones del experimento. Esto sería una pequeña muestra (por lo tanto poco fiable) que no arrojaría mucha luz sobre el problema de las intuiciones de todo el grupo ya que aun cuando el estudiante obtuviese exactamente el mismo número de caras y cruces en las dos secuencias no obtendría ninguna conclusión, como tampoco la obtendría con números diferentes, pues es natural la variabilidad de resultados en este tipo de experimentos. Incluso cuando se piense en unir todos los datos para hacer una muestra más confiable de la proporción del número de caras y de cruces obtenidas por todo el grupo, no arrojarían mucha luz sobre el problema de las intuiciones del grupo, puesto que se dejaría de lado el el orden en que aparecieron las caras y las cruces, puesto que un mismo número de caras puede aparecer de diferentes de formas y es de esperar que algunas parezcan más «aleatorias» que otras.

Si analizan sus propios datos, algunos estudiantes pueden manifestar ideas erróneas como esperar que se obtengan exactamente 10 caras y 10 cruces e incluso, la misma secuencia de resultados en los dos experimentos, lo cual pudiera indicar la ilusión de control descrita por Langer (1975). Otros pensarán que tienen buenas intuiciones si tienen muchas coincidencias en las dos secuencias, manifestando el *enfoque en el resultado aislado* (Konold, 1989). En esta fase del proyecto, el profesor tratará de llevar a los estudiantes del estudio de su caso particular (es decir, si sus intuiciones fueron o no correctas) al estudio del colectivo clase (si las intuiciones de la clase como grupo son correctas). El profesor organizará una discusión con preguntas como las siguientes:

4. Cuestionamiento sobre los resultados individuales obtenidos:

Pero, ¿hemos de obtener exactamente 10 caras y 10 cruces? ¿Qué pasa si obtenemos 11 caras y 9 cruces? ¿Y si obtenemos 18 y 2?

5. Comparación entre lo simulado y lo obtenido:

¿Qué os parece si comparamos el número de caras en las secuencias real y simulada de todos los alumnos de la clase? Cada alumno puede contar el número de caras en sus secuencias simulada y real ¿Qué valores habéis obtenido? ¿Cuáles son el valor mínimo y máximo obtenido? ¿Cuál es el valor más frecuente? ¿Cómo representar los datos de modo que sepamos cuántas veces aparece cada valor? ¿Cómo podríamos organizar y resumir estos datos?

Se espera que, con sus preguntas, el profesor guíe a los alumnos a ir más allá de sus propios datos y a poner atención a los de sus compañeros: propone trabajar con el número de caras que obtuvo cada alumno en su experimento particular (es decir, el número de caras en una secuencia de 20 lanzamientos) y analizar esta variable en el total de los alumnos. Se comienza por identificar el valor máximo y mínimo del número de caras que apareció en la clase, apreciar que los diferentes alumnos obtienen diversos valores e interesarse por los más frecuentes. Se trata, aunque no explícitamente, de inducir en los estudiantes la idea de organizar un recuento y una tabla de frecuencias para resumir la información tanto en las secuencias reales como en las simuladas para finalmente comparar las dos distribuciones y analizar si existen diferencias importantes que indiquen que su intuición los engaña (o no) con respecto a la aleatoriedad.

Así, el siguiente paso es organizar la recolección de los datos de todos los alumnos de la clase, tanto del número de caras en las secuencias simuladas como en las reales. Con esta actividad, aparecen dos nuevas variables aleatorias y sus correspondientes variables estadísticas:

- ❖ La variable η_r que correspondería al número de caras en 20 lanzamientos de la moneda equilibrada, por tanto sería una variable aleatoria Binomial con parámetros $n=20$ $p=q=1/2$. Su media, sería igual a np y su varianza igual a npq (Ríos, 1967). Con respecto a esta variable podemos analizar si el estudiante percibe su variabilidad y si tiene una intuición de su valor medio y de los valores más probables.
- ❖ La variable aleatoria η_s que correspondería al número de caras en una secuencia de longitud 20 inventada por los estudiantes y que usamos como modelo matemático para reflejar las intuiciones colectivas sobre los experimentos aleatorios. Esperamos que no siga la distribución Binomial, pues la investigación didáctica muestra que los ensayos producidos en este tipo de experimentos no son independientes, aunque la probabilidad de cara sigue siendo p=1/2 (Green, 1982, 1989; Toohey, 1995; Serrano, 1996).

Si m es el número de estudiantes de la muestra, entonces, puesto que cada estudiante repite una vez el experimento, habrá una serie de m ensayos de cada una de las variables aleatorias anteriores. Por lo tanto obtenemos dos variables estadísticas:

- La variable estadística Y_r o resultados de una muestra de m valores de la variable aleatoria η_r. Se espera que el estudiante use esta variable estadística para sacar algunas conclusiones sobre la variable aleatoria η_r, puesto que sus conocimientos matemáticos no le permiten trabajar directamente con η_r.
- La variable estadística Y_s o resultados de una muestra de m valores de la variable aleatoria η_s. El estudiante ha de usar esta variable estadística para sacar algunas conclusiones sobre la variable aleatoria η_s.

Una vez recogidos los datos del grupo completo, la finalidad es que el estudiante los resuma formando la distribución de frecuencias de las correspondientes variables estadísticas Y_r y Y_s para proceder, primeramente al análisis de cada una de estas dos variables y luego a la comparación de las principales diferencias en su distribución. De ello esperamos que extrapole a las diferencias entre las correspondientes variables aleatorias η_r y η_s y de ahí obtenga conclusiones respecto a sus intuiciones a través de un proceso de inferencia informal. Esperamos que los estudiantes se den cuenta que las variables tienen la misma media, pero distinta dispersión, es decir, que las intuiciones sobre el número de caras son buenas con respecto al promedio, pero la secuencia de los alumnos tiene menos variabilidad que las secuencias reales.

4.2.5. Variables vinculadas al número de rachas

El número de caras es sólo una de las variables que podemos analizar en una secuencia de resultados aleatorios. En realidad aparecen muchos otros modelos probabilísticos (Batanero y Serrano, 1995; Batanero, Serrano y Green, 1998) y en este estudio también se quiere que los estudiantes tengan contacto con alguno de ellos, en particular con variables vinculadas con las rachas.

Con base en las investigaciones sobre aleatoriedad subjetiva (Green, 1982; 1989; Engel y Sedlmeier, 2005) se espera que las rachas simuladas por los estudiantes sean cortas. Detrás de este razonamiento está la *falacia del jugador*, según la cual se cree que después de una corta racha de, por ejemplo, caras, la probabilidad de que aparezca una cruz tiene que aumentar, por lo que, inconscientemente, se tiende a que el resultado de un lanzamiento simulado dependa de lo que se puso en el lanzamiento anterior (Green, 1982; 1989; Shaughnessy, 1992; Engel y Sedlmeier, 2005). Para evaluar esta dependencia de ensayos, en este proyecto proponemos analizar dos nuevas variables en las secuencias producidas por los alumnos: el número de rachas y la longitud de la racha más larga.

La introducción de estas variables requiere una intervención más dirigida del profesor, puesto que son menos conocidas para el estudiante. Él deberá explicar a los alumnos que una *racha* es una secuencia de resultados iguales, de modo que, si después de una cara aparece una cruz (o viceversa) la racha tiene longitud 1. A modo de ejemplo, se usaron las secuencias obtenidas por un alumno, mostradas en la Tabla 6.1, donde C, denota *cara* y + denota *cruz* para motivar las variables propuestas. En esa tabla se pueden observar las rachas. Así, en la secuencia simulada, la racha más larga es de longitud 3 (3 caras) y el número de rachas es 12, mientras que en secuencia real hay una racha de 5 cruces y el número de rachas es 11.

Tabla 6.1. Ejemplo de las secuencias obtenidas por un estudiante

Lanzamiento simulado:

C	C	+	C	+	+	+	C	C	+	C	+	C	+	+	C	C	C	+	+

Lanzamiento real:

+	C	+	C	+	+	C	C	+	C	C	C	+	+	+	+	+	C	+	+

A partir del ejemplo, el profesor podría preguntar si la racha donde aparecen 5 cruces seguidas, parece razonable. Es posible que algún alumno sugiera que la moneda utilizada no estaba bien construida o que se hizo trampa al realizar los lanzamientos. Este tipo de respuestas ayudarán al profesor a conducir a los estudiantes hacia un cuestionamiento sobre las características de las rachas de una secuencia aleatoria y al análisis de las rachas en las secuencias generadas por el grupo. En esta parte del estudio aparecerán cuatro nuevas variables aleatorias y sus correspondientes variables estadísticas:

- ❖ La variable θ_r correspondería al número de rachas en 20 lanzamientos de la moneda equilibrada. Con respecto a esta variable podemos analizar si el estudiante percibe su variabilidad y si tiene una intuición de su valor medio y cuáles son los valores más probables. En una secuencia de 20 lanzamientos, el valor esperado del número de rachas según Engel y Sedlmeier (2005) es 10,5.
- ❖ La variable aleatoria θ_s correspondería al número de rachas en la secuencia de longitud 20 inventada por los estudiantes y que usamos como modelo matemático para reflejar las intuiciones colectivas sobre los experimentos aleatorios. No conocemos su distribución, pero las investigaciones de Green

(1982, 1989), Toohey (1995), Falk y Konold (1997) y Serrano (1996) indican que hay una tendencia a un exceso de alternancia entre caras y cruces y que, mientras la probabilidad esperada de cambio de resultado en una secuencia aleatoria es 0,5, las personas aumentan esa probabilidad hasta 0,6 ó 0,7.

Puesto que cada estudiante realiza el experimento, se tiene una serie de m ensayos de cada una de las variables aleatorias anteriores y por lo tanto, también las dos variables estadísticas correspondientes:

- La variable estadística T_r o resultados de una muestra de m valores de la variable aleatoria θ_r.
- La variable estadística T_s o resultados de una muestra de m valores de la variable aleatoria θ_s.

Lo mismo que en las variables anteriores, suponemos que los estudiantes, en un proceso de inferencia informal, compararán las características de las distribuciones de las variables T_s *y* T_r para conjeturar sobre las semejanzas o diferencias entre θ_s y θ_r y de ahí deducir el comportamiento de sus intuiciones con respecto al azar. En particular, se espera que el estudiante observe que el valor medio del número de rachas es menor en las secuencias reales que en las simuladas, pero la dispersión es parecida.

4.2.6. Variables vinculadas la racha más larga

Finalmente se analizaría la longitud de la racha más larga, donde aparecen dos nuevas variables aleatorias:

- La variable aleatoria λ_r que correspondería a la longitud de la racha mayor en 20 lanzamientos de una moneda equilibrada. De acuerdo a Engel y Sedlmeier (2005), el valor medio de la racha más larga en una secuencia de longitud n está dada por el logaritmo con base 2 de n, que para $n=20$ sería aproximadamente igual a 4,322. Respecto a esta variable podemos analizar si el estudiante percibe su variabilidad y cuáles son los valores más probables.
- La variable aleatoria λ_s que correspondería a la longitud de la racha mayor en secuencia de longitud 20 inventada por los estudiantes y que usamos como modelo matemático para reflejar las intuiciones colectivas sobre los experimentos aleatorios.

Igual que en las variables anteriores, aparecen dos variables estadísticas L_r *y* L_s que el estudiante ha de analizar para conjeturar sobre las semejanzas o diferencias entre λ_r y λ_s y de ahí deducir sobre las intuiciones respecto al azar. Esperamos que el estudiante observe que el valor medio de las rachas de las secuencias reales es mayor que el de las de las secuencias simuladas.

La sesión finaliza con algunas instrucciones muy generales sobre el proyecto que deberán de completar en casa. Se les da la hoja de los datos recogidos y se les pide que los analicen y hagan un informe escrito que entregarán la siguiente semana. Se aclara que en su informe deberán incluir todos los análisis y gráficos producidos y escribir sus conclusiones (razonadas y sustentadas en el análisis realizado) sobre las intuiciones de la clase con respecto a los experimentos aleatorios.

4.3. Respuestas y soluciones del proyecto: análisis de los datos

Para obtener una respuesta a la pregunta principal planteada en el proyecto: *¿Tienes una buena intuición sobre el azar?*, es necesario analizar los datos generados por los estudiantes durante su trabajo en clase. Los alumnos que participaron en la experiencia estuvieron divididos en tres grupos. Cada grupo tenía datos diferentes, puesto que los datos fueron recopilados en cada clase como producto del experimento que realizó cada uno de los alumnos de la clase. Los datos recogidos en uno de los grupos se muestran en la Tabla 6.2. Mostraremos sólo el análisis de los datos de este grupo, puesto que los resultados son muy similares en los otros dos y en los realizados en cursos anteriores.

Se espera que los estudiantes analicen estos datos a través de cada una de las variables mencionadas en el apartado anterior (4.2) de manera independiente, aunque también es deseable que el análisis de cada una de ellas les permita llegar a una conclusión conjunta acerca de las intuiciones sobre el azar. Suponemos que los estudiantes harán uso de herramientas estadísticas que les faciliten la comparación entre las secuencias reales y las simuladas de cada una de las variables. No se puede exigir el uso de pruebas de hipótesis o intervalos de confianza, puesto que los estudiantes no han tenido contacto con esos temas, sin embargo sí se espera que tengan la capacidad de un manejo apropiado de las medidas de tendencia central, de dispersión y de gráficas y tablas y que su correcta interpretación les permitan hacer una comparación con estos elementos. Es decir, confiamos en que estén en capacidad de hacer un adecuado análisis exploratorio de datos que los conduzca a realizar inferencias informales (Rossman, 2008).

En este apartado se expondrá el análisis de los datos obtenidos por los alumnos mostrados en la Tabla 6.2. En primera instancia mostraremos el análisis que se esperaría que los estudiantes hicieran y conjeturaremos sobre las dificultades que podrían presentarse al realizar ese trabajo. Para completar el análisis de Arteaga (2011), utilizaremos pruebas de hipótesis para dar una solución matemática completa a la actividad y justificar las afirmaciones hechas sobre la percepción de la aleatoriedad en el apartado anterior. Puesto que el análisis de la comparación de las distintas variables es similar, el análisis del número de caras será más detallado que el del resto de las variables.

Tabla 6.2. Hoja de datos de una clase

Alumno No.	Secuencia simulada			Secuencia Real		
	Número de Caras	Número de Rachas	Racha Mayor	Número de Caras	Número de Rachas	Racha Mayor
1	10	13	3	12	9	4
2	11	13	3	9	10	4
3	10	13	3	11	13	3
4	10	11	3	15	9	6
5	11	13	2	11	11	4
6	10	14	3	11	16	2
7	10	14	2	10	9	6
8	9	12	3	7	7	8
9	12	15	3	12	10	4
10	9	14	2	9	12	4
11	11	11	3	12	12	3
12	10	8	4	11	11	4
13	11	13	3	9	12	4
14	11	14	3	8	12	6
15	10	12	3	10	11	5
16	9	12	3	11	11	3
17	10	14	3	7	14	3
18	10	12	3	8	11	4
19	9	10	3	10	12	4
20	11	12	3	10	10	4
21	10	12	3	14	9	7
22	10	13	3	10	10	3
23	13	11	3	13	8	9
24	10	10	3	13	8	6
25	11	12	3	8	10	3
26	11	11	3	11	10	5
27	9	12	3	11	10	5
28	10	14	3	14	8	5
29	10	12	3	8	10	4
30	9	15	3	11	11	4

4.3.1. Comparación del número de caras en las secuencias reales y simuladas

El primer paso en esta comparación es resumir la información en una tabla de frecuencias (Tabla 6.3) que permita analizar los datos más fácilmente. Para preparar la tabla, los estudiantes necesitan contar la frecuencia de valores de cada una de las dos variables estadísticas, con ello formularán la distribución de las mismas. Los estudiantes podrían construir una tabla como la mostrada o bien dos tablas separadas para cada una de las variables. Podrían también completar la tabla con frecuencias relativas, frecuencias acumuladas y porcentajes. Todos estos conocimientos les son accesibles por sus estudios de secundaria y del primer año en la Facultad.

Tabla 6.3. Distribución de frecuencias del número de caras.

Número de caras	Secuencia	
	Simulada	Real
6	0	0
7	0	2
8	0	4
9	6	3
10	14	5
11	8	8
12	1	3
13	1	2
14	0	2
15	0	1
16	0	0
Total	30	30

Al realizar tablas o representaciones gráficas de la distribución de frecuencias, los estudiantes hacen uso de uno de los razonamientos estadísticos fundamentales, según el modelo de Wild y Pfannkuch (1999), el de transnumeración, que consiste en realizar un cambio de representación a los datos para, a partir del mismo, conseguir un conocimiento no accesible en el listado de datos. Lo más adecuado sería que las dos distribuciones se representen conjuntamente, por ello que, en lo que sigue, nosotros presentaremos todas las tablas y los gráficos por pares de variables relacionadas.

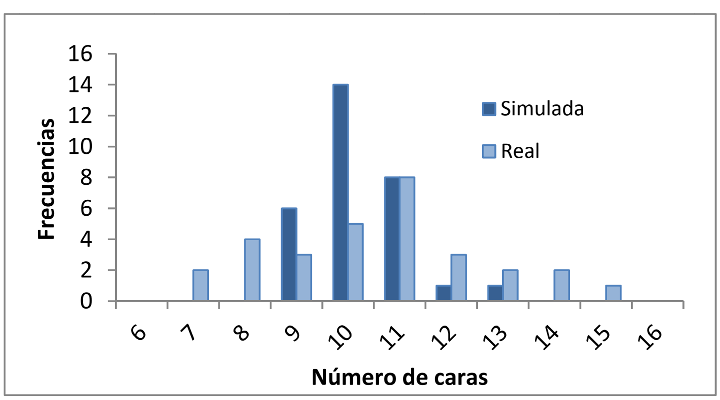


Figura 6.1. Diagramas de Barras del número de caras en secuencias reales y simuladas

Nuestros estudiantes están familiarizados con los diagramas de barras y polígonos de frecuencia, que podrían construir a mano o bien, como en la Figura 6.1 y Figura 6.2, utilizando el paquete Excel, que ellos ya han empleado en la asignatura de Matemáticas y su didáctica el curso anterior. Ambas representaciones gráficas permiten a los estudiantes observar las diferencias y semejanzas de la distribución del número de caras en las secuencias real y simulada, especialmente la cercanía de las modas y diferencia de rangos. También se observa claramente que la mayor parte de los estudiantes produjo exactamente 10 caras, que es el valor esperado de la variable aleatoria binomial *B(20, 0,5)* que modela el número de caras al lanzar 20 veces una moneda. De ahí deberán concluir que los estudiantes tienen una buena intuición del valor esperado de dicha variable aleatoria.

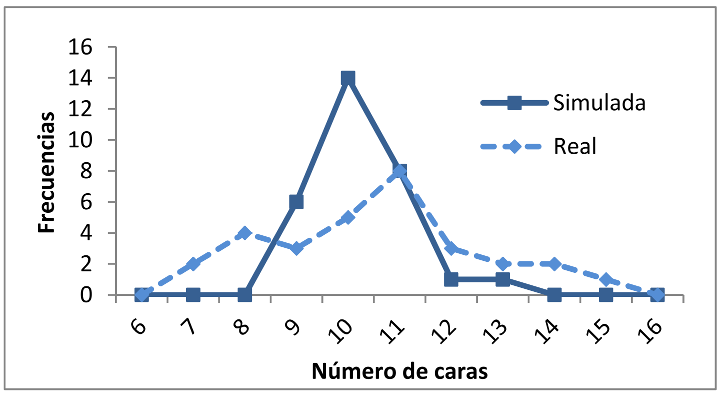


Figura 6.2. Polígono de frecuencia del número de caras en las secuencias reales y simuladas

En caso de que los estudiantes los conozcan, también podrían realizar algunos gráficos de caja (Figura 6.3) de las dos secuencias, donde se muestran más claramente la coincidencia de las medianas y al mismo tiempo podemos comparar visualmente la media y los cuartiles. Es muy evidente la menor dispersión, que indica que al tratar de simular un fenómeno aleatorio somos menos variables de lo que ocurre en la realidad; no percibimos completamente la variabilidad aleatoria, tenemos una tendencia a reducirla (Engel y Sedlmeier, 2005). También permite identificar los alumnos que produjeron resultados atípicos.

En caso de que los estudiantes produzcan algunos de los gráficos anteriores en forma separada, sería importante usar las mismas escalas y si construyen histogramas los mismos intervalos y tomar valores centrados en números enteros, al ser la variable discreta. Con el fin de poder comparar los gráficos y de obtener gráficos fáciles de interpretar.

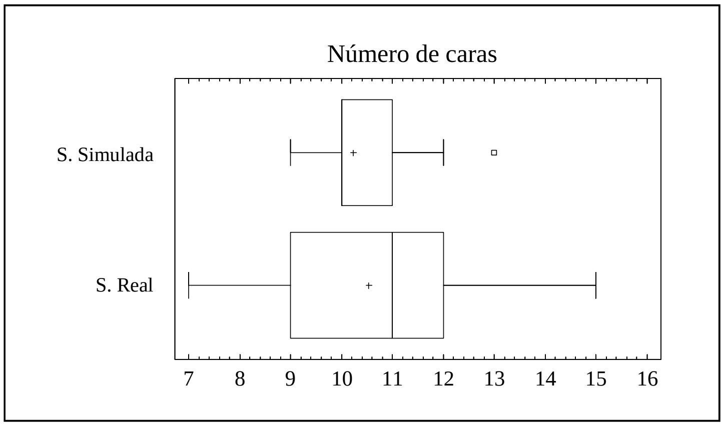


Figura 6.3. Número de caras en secuencias reales y simuladas

En estudios previos (Arteaga, Batanero y Ruiz, 2009; Ruiz, Arteaga y Batanero, 2009 a y b; Ruiz, Batanero y Arteaga, 2011) hemos reportado algunas posibles dificultades que aparecen en la realización de tablas y gráficos. Esperamos que algunos de ellos se reproduzcan también aquí, como que los estudiantes no llegasen a formar la distribución de la variable estadística, analizando solamente su caso particular (los datos de su experimento) o bien estudiasen los valores obtenidos uno a uno, sin agruparlos, trabajando con las frecuencias de la variable cara o cruz. Estos errores indicarían que los estudiantes no usan intuitivamente la idea de distribución para comparar dos distribuciones.

Una característica esencial del análisis estadístico es que trata de describir y predecir propiedades de los conjuntos de datos (y no de cada dato aislado). Sin embargo, Konold, Pollatsek, Well y Gagnon (1997) han observado que a algunos estudiantes se les dificulta dejar de pensar en los valores de la variable como propiedad de los individuos aislados, en lugar de comparar propiedades de conjuntos de datos. También Batanero, Estepa y Godino (1997) estudian estrategias de estudiantes de magisterio al comparar dos distribuciones, y encuentran que algunos estudiantes solo usan valores aislados en la comparación de dos distribuciones. Esta conducta podría presentarse en nuestro estudio.

Otra herramienta indispensable para el análisis de los datos son los estadísticos. Es necesario que la confrontación de las dos distribuciones se realice basándose en la comparación de los promedios (media, mediana y moda) y dispersión (varianza, desviación típica y rango). Esperamos, por tanto, que los estudiantes calculen algunos de los estadísticos mostrados en la Tabla 6.4, incluso algunos de ellos podrían no requerir de las tablas y los gráficos anteriores, sino únicamente hacer la comparación directamente a partir de estos estadísticos.

Tabla 6.4. Estadísticos del número de caras

	Secuencia	
	Simulada	Real
Media	10.2333	10.5333
Moda	10	11
Mediana	10	11
Máximo	13	15
Mínimo	9	7
Rango	4	8
Desviación Típica	0.9353	2.0634
Varianza	0.8747	4.2575

Numéricamente se pueden comprobar las conjeturas gráficas hechas anteriormente a través de estos resúmenes estadísticos: la media, mediana y moda coinciden o son muy próximas, mientras la varianza, desviación típica y recorrido intercuartílico son menores en la distribución de caras simuladas. Esto, nuevamente, indica una buena intuición del valor medio (esperanza) del número de caras en la secuencia aleatoria, pero no de la dispersión de los datos. Podemos decir, que la clase produce secuencias con menor dispersión del número de caras que las obtenidas al realizar el experimento, que es una forma de no aceptar la aleatoriedad de las mismas.

También esperamos que en el cálculo de los estadísticos, los estudiantes tengan algunas de las dificultades ya descritas en las investigaciones de varios autores, en particular de Cobo (2003) y Mayén (2009):

- Respecto a la media, la principal dificultad descrita en estos trabajos es el error en el cálculo de medias ponderadas; aunque también se han descrito otros errores tales como calcular la media de las frecuencias (confundiendo frecuencia y dato).
- Respecto a la mediana, el principal problema que se presenta es no ordenar los datos en el cálculo de la mediana o no tener en cuenta el caso de indeterminación.
- En muchos casos, los estudiantes se limitan a calcular los estadísticos sin interpretarlos, pues tienen problemas al elegir el mejor promedio para representar un conjunto de datos o bien en interpretar el significado de la varianza o desviación estándar.

Finalmente presentamos en la Tabla 6.5 y en la Tabla 6.6 los contrastes de hipótesis de la diferencia de medias (mediante el test *t* de Student de muestras relacionadas) y de diferencias de varianzas (mediante la prueba *F*). Los estudiantes no tienen acceso a este conocimiento, por lo que no se espera que ellos hagan este tipo de estudios; sin embargo a nosotros como investigadores, nos dará la referencia de la solución matemática del problema.

El contraste de diferencias de medias (Tabla 6.5), permite ver en primer lugar, el solape de los intervalos de confianza del 95% y que el intervalo de confianza para la diferencia de medias contiene el origen de coordenadas. Por otro lado, el valor de la *t* obtenido para la diferencia de medias es muy cercano al origen, por lo que la probabilidad de obtener la diferencia dada en caso de igualdad en las medias de las poblaciones es 0,47, es decir, muy alta. Todos estos indicadores permiten afirmar que la igualdad de medias del número de caras en secuencias reales y simuladas se podría generalizar a las poblaciones de alumnos de donde se obtuvieron las muestras.

Tabla 6.5. Resultado del contraste de diferencia de medias

Intervalo de confianza de 95%	
Media del número de caras en las secuencias simuladas	(9,88 - 10,58)
Media del número de caras en las secuencias reales	(9,76 - 11.3)
Diferencia de medias en las dos secuencias	(-1,12 - 0,52)
Valor *t*	*t* = -0,72
Valor *p*	0,47

Por el contrario, no hay solape en los intervalos de confianza de las desviaciones típicas del número de caras en las dos distribuciones y la razón entre las dos varianzas (mayor de 5) es muy poco probable en caso de igualdad de varianzas en las poblaciones, por lo tanto, la diferencia de varianzas (y de dispersión) es estadísticamente significativa. El valor *p* es muy pequeño. Todos estos indicadores nos permiten aceptar la diferencia de dispersión en las dos distribuciones (Tabla 6.6).

Tabla 6.6. Resultado del contraste de diferencia de varianzas

Intervalo de confianza de 95%	
desviación típica del número de caras en las secuencias simuladas	(0,74 - 1,25)
desviación típica del número de caras en las secuencias reales	(1,6 - 2,77)
Valor *F*	*F* = 5,2
Valor *p*	0,00005

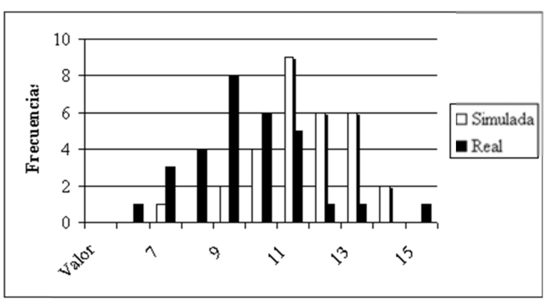
4.3.2. Comparación del número de rachas en las secuencias reales y simuladas

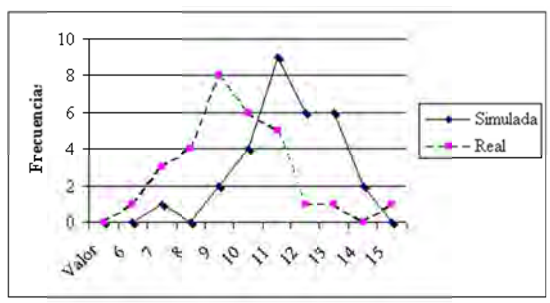
El análisis matemático para comparar las secuencias a través del número de rachas es parecido al análisis de la variable anterior, por lo que será más escuetamente tratado en este apartado. Tanto en la tabla de frecuencias (Tabla 6.7), como en los distintos gráficos (Figura 6.4), se observa un desplazamiento de la distribución del número de rachas en las secuencias reales respecto a las simuladas, lo que muestra notoriamente un mayor valor medio del número de rachas en las secuencias simuladas. También se puede notar con claridad una dispersión semejante o incluso menor. La conclusión que se espera que los estudiantes obtengan de esta observación es que los valores medios son ahora diferentes, pero con la misma dispersión. Dicho de otra manera, cuando simulamos secuencias aleatorias producimos más rachas que las que se esperarían en una secuencia realmente aleatoria.

Tabla 6.7. Distribución de frecuencias y estadísticos del número de rachas

Valor	Secuencia	
	Simulada	Real
7	0	1
8	1	3
9	0	4
10	2	8
11	4	6
12	9	5
13	6	1
14	6	1
15	2	0
16	0	1
Total	30	30

Estadístico	Secuencia	
	Simulada	Real
Media	12,4	10,5
Moda	12	10
Mediana	12	10
Máximo	15	16
Mínimo	8	7
Rango	7	9
Desv. Típ.	1,6	1,9
Varianza	2,4	3,6





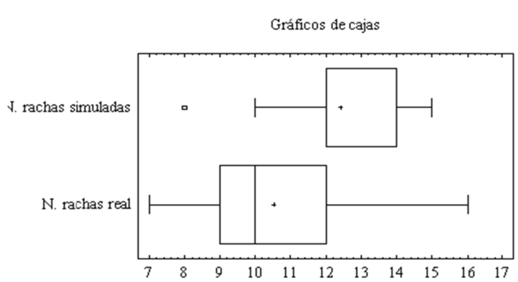


Figura 6.4. Algunos gráficos del número de rachas en secuencias reales y simuladas

Las dificultades que podrían tener los estudiantes en la producción e interpretación de tablas, gráficos y resúmenes estadísticos son similares a las descritas en el apartado anterior. Sin embargo, de acuerdo con Arteaga (2011), la variable «número de rachas» será menos utilizada porque al realizar este tipo de experimentos, la habían consideran menos que la variable «número de caras», que les resulta más familiar.

En la comparación entre varianzas (Tabla 6.9), los intervalos de confianza del 95% se solapan, y obtenemos un valor *F* (razón de varianzas) cercano a la unidad con una probabilidad (valor *p*) igual a 0,31, valor que ocurre una de cada tres veces, por tanto es relativamente probable. Como consecuencia no podemos asumir que las distribuciones tengan diferente dispersión en este caso.

Tabla 6.8. Resultado del contraste de diferencia de medias

Intervalo de confianza de 95%	
Media del número de rachas en las secuencias simuladas	(11,81 - 12,98)
Media del número de rachas en las secuencias reales	(9,82 - 11.23)
Diferencia de medias en las dos secuencias	(0,96 - 2,86)
Valor *t*	*t* = -0,72
Valor *p*	0,47

Tabla 6.9. Resultado del contraste de diferencia de varianzas

Intervalo de confianza de 95%	
desviación típica del número de rachas en las secuencias simuladas	(1,24 - 2,10)
desviación típica del número de rachas en las secuencias reales	(1,5 - 2,53)
Valor *F*	*F* = 0,67
Valor *p*	0,31

4.3.3. Comparación de la longitud de la racha mayor en las secuencias reales y simuladas

A continuación se repiten los análisis para la variable «longitud de la racha más grande». En este caso también las diferencias son evidentes. En la tabla de frecuencias y en los estadísticos (Tabla 6.10) se observa que la longitud mayor de la racha de las secuencias producidas por los estudiantes se concentra entre 2 y 4 mientras que en las reales varía entre 2 y 9.

Estas diferencias son más claras en las diferentes gráficas (Figura 6.5), tanto en lo que concierne a la diferencia de valores centrales (moda en los diagramas de barra, polígonos de frecuencia e histogramas y mediana en los gráficos de caja) como diferencia de dispersión. Las rachas generadas por los estudiantes son muy cortas con respecto a las que se presentan en un proceso aleatorio. Además, hay una gran coincidencia en los resultados pues hay mucho menos dispersión que cuando se repiten varios lanzamientos aleatorios de 20 monedas.

Tabla 6.10. Distribución de frecuencias y estadísticos de la longitud de la racha mayor

Valor	Secuencia	
	Simulada	Real
2	3	1
3	26	6
4	1	12
5	0	4
6	0	4
7	0	1
8	0	1
9	0	1
Total	30	30

	Secuencia	
	Simulada	Real
Media	2,9	4,6
Moda	3	4
Mediana	3	4
Máximo	4	2
Mínimo	2	9
Rango	2	7
Desv. Típ.	0,36	1,56
Varianza	0,13	2,46

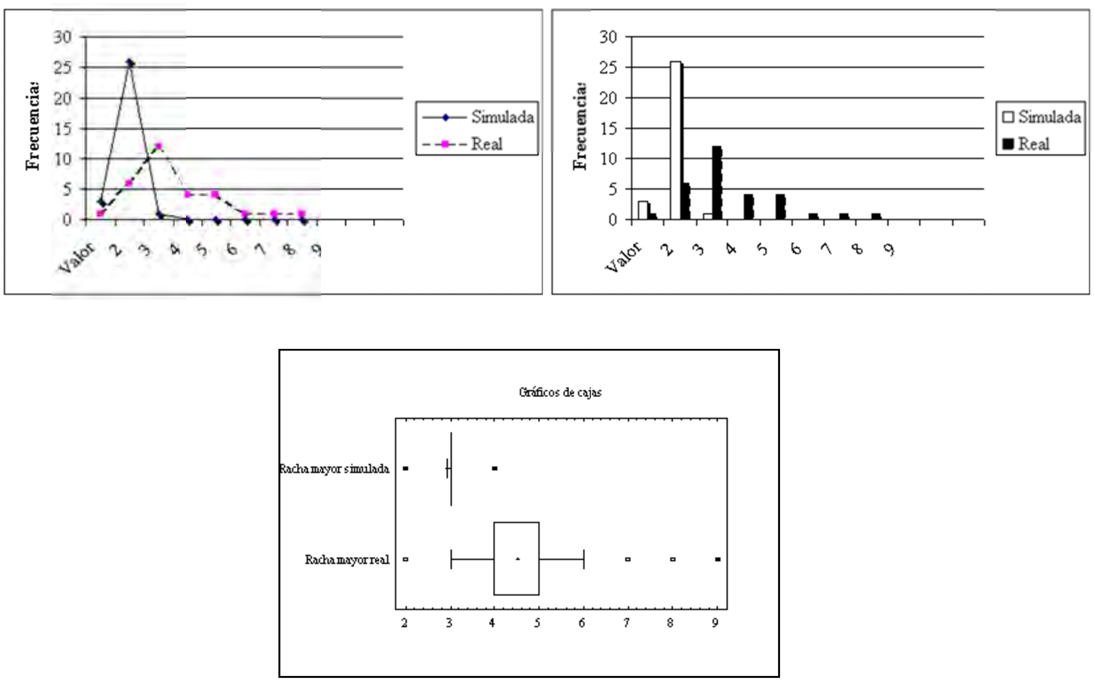


Figura 6.5. Algunos gráficos de la longitud de racha mayor en secuencias reales y simuladas

Como en el caso de las variables anteriores realizaremos contrastes de hipótesis para obtener la solución matemática (Tabla 6.11 y Tabla 6.12). En esta variable los valores medios de la distribución simulada son menores y lo mismo ocurre con la dispersión. Tanto las diferencias de medias como las de varianza son estadísticamente significativas. En ninguno de los dos casos se solapan los intervalos de confianza del

95% de la media o desviación típica en cada distribución. Las pruebas *t* y *F* dan como resultado un valor con una probabilidad extremadamente pequeña, por lo que la conclusión es que los estudiantes producen rachas más cortas que lo esperado en una secuencia aleatoria. Asimismo la dispersión de la racha más larga en las secuencias reales es menor que la correspondiente a las simuladas por lo que la concordancia entre los resultados obtenidos por los alumnos es mayor que la esperada en una situación aleatoria real.

Tabla 6.11. Resultado del contraste de diferencia de medias

Intervalo de confianza de 95%	
Media de la longitud de la racha mayor en las secuencias simuladas	(2,79 - 3,06)
Media de la longitud de la racha mayor en las secuencias reales	(3,94 - 5,11)
Diferencia de medias en las dos secuencias	(-2,18 - 1,01)
Valor *t*	t = -5,43
Valor *p*	0,000001

Tabla 6.12. Resultado del contraste de diferencia de varianzas

Intervalo de confianza de 95%	
desviación típica de la longitud de la racha mayor en las secuencias simuladas	(0,29 - 0,49)
desviación típica de la longitud de la racha mayor en las secuencias reales	(1,25 - 2,11)
Valor *F*	F = 0,05
Valor *p*	0,000000

4.3.4. La conclusión: Una respuesta a la pregunta del proyecto.

Suponemos que una parte de los estudiantes integrarán las conclusiones obtenidas en los análisis individuales de cada una de las variables de los incisos anteriores y que serán capaces de emitir una conclusión global con respecto al grupo completo que podrían aplicar con respecto a sí mismo (como parte del grupo), para deducir si tiene buenas intuiciones sobre el azar.

Se espera que concreten una conclusión en donde manifieste que la aleatoriedad en la secuencia simulada está condicionada, principalmente, por las variables número de caras y longitud de la racha más grande. Al igual que en la investigación de Arteaga (2011) esperamos que algunos de ellos generen argumentos parecidos al siguiente o a algunas partes del mismo:

> Los resultados indican que los estudiantes tienen una buena intuición del número de caras más probable al lanzar una moneda. Sin embargo, están tan preocupados por lograr que la media del número de caras sea 10, que tienden a hacer muchas rachas y por lo tanto la longitud de la racha más larga que logran inventar es muy corta. La gran variabilidad del número de rachas indica que los estudiantes no estaban

> preocupados por el número de rachas que generaban, en cambio la poca variabilidad en la distribución de la longitud de la racha más larga, señala que la gran mayoría hizo lo posible porque sus rachas fueran cortas. Podemos concluir que los estudiantes del grupo tienen intuiciones parcialmente correctas (el número de caras) y parcialmente incorrectas (tratan de que no haya repetición consecutiva muy larga del mismo resultado) porque hay demasiada preocupación por el equilibrio entre caras y cruces.

Aunque no lo expresen explícitamente, esperamos que sus respuestas reflejen algunos indicadores explícitos de comprensión de que la aleatoriedad no significa necesariamente ausencia de modelos (Batanero y Serrano, 1999), sino que también puede «observarse» a través de modelos que se reflejan en los parámetros o estadísticos y distribuciones de las variables trabajadas, con algún razonamiento similar al siguiente en el que también hay evidencias de uso de las variables estadísticas y aleatorias:

> Nos dimos cuenta que la aleatoriedad puede ser observada y evaluada por medio de las distribuciones de las tres variables que manipulamos en el proyecto. Así mismo, que es importante considerar las medidas de tendencia central y de dispersión en su conjunto.

5. Objetos de análisis en el proyecto

Para completar el estudio de Arteaga (2011) en este apartado analizaremos el proyecto más profundamente como un instrumento de evaluación de la comprensión de los modelos matemáticos relacionados con la variable aleatoria y que están implicados en la realización de inferencias informales por parte de los estudiantes participantes en la investigación.

Este análisis lo presentamos en cuatro aparados de acuerdo a los objetos matemáticos que el estudiante tendrá que usar implícita o explícitamente en la resolución del proyecto: (1) la aleatoriedad, (2) la variable estadística y su distribución, (3) la variable aleatoria y su relación con la variable estadística y (4) las conclusiones sobre la pregunta planteada y la forma en que el estudiante completa el *ciclo de modelación*.

En cada apartado se describe y analiza la forma en que el objeto se hace presente en el estudio, cómo esperamos que el estudiante lo ponga en juego al desarrollar el proyecto y la manera en que esperamos evaluar su comprensión de acuerdo con los objetivos de nuestra investigación, así como algunas de las posibles respuestas de los estudiantes. Para nuestro análisis nos basaremos en las investigaciones que han descrito dificultades de los estudiantes con cada uno de los objetos analizados. Todo ello nos servirá para plantear nuestras hipótesis de este estudio.

5.1. Aleatoriedad

La aleatoriedad no sólo es un objeto matemático con el que se relaciona epistémicamente la variable aleatoria, también es el contexto en el que se desarrolla el proyecto, puesto que proponemos un experimento clásico sobre percepción de la aleatoriedad. Es particularmente importante indagar, entonces, sobre el conocimiento que los estudiantes tienen acerca de ella y la forma en que ese conocimiento influye sobre su razonamiento estadístico y viceversa (Makar, Bakker y Ben-Zvi, 2011).

Así, la aleatoriedad se presenta en dos niveles en la investigación. Por un lado, las concepciones de los estudiantes se deberán ver reflejadas en los datos recogidos en el experimento propuesto en el proyecto. Por otro, puesto que se pide a los estudiantes que usen los resultados de un experimento aleatorio para analizar sus propias concepciones sobre «el azar», las conclusiones obtenidas por los estudiantes y sus argumentos como parte de su análisis nos permitirán también acceder a su visión sobre la aleatoriedad. Así, en el proyecto se evaluará la concepción de aleatoriedad desde dos perspectivas, a) en una observaremos globalmente las percepciones sobre el azar del grupo completo en el análisis de los datos de sus secuencias inventadas; b) en la otra a través de las razones individuales de los estudiantes para justificar su conclusión ante la pregunta sobre las intuiciones del azar del grupo completo. Por el tipo de información y el diseño de la recopilación de la información, la primera permite un análisis más cuantitativo, la segunda, en cambio, será cualitativo.

En las preguntas planeadas, hemos evitado el uso de términos más sofisticados que el estudiante podría desconocer; por eso utilizamos el término «azar» como expresión coloquial para referirnos a la aleatoriedad. La idea de azar es más amplia e involucra concepciones filosóficas y epistemológicas sobre las que no profundizaremos en este trabajo, sin embargo estaremos atentos a ellas porque implícitamente determinan comportamientos y creencias de los estudiantes. Dentro de las múltiples concepciones históricas sobre aleatoriedad (ver resúmenes en Batanero y Serrano, 1995; Serrano, 1996; Bennett, 1998; Batanero Henry y Parzysz, 2005), en este proyecto adoptamos los dos aspectos constitutivos de este objeto que devinieron como importantes a partir de los desarrollos teóricos de la inferencia estadística y que fueron diferenciados por Zabell (1992):

- ❖ El *proceso de generación* de resultados aleatorios, que es lo que matemáticamente se conoce como *experimento aleatorio*. Detrás de este aspecto, está el supuesto de que el carácter aleatorio del proceso asegura que los resultados obtenidos a partir de él, sean aleatorios. En nuestra investigación consideramos la evaluación del experimento más simple posible: el lanzamiento de una moneda equilibrada, que consta de dos resultados posibles, equiprobables e independientes (cara y cruz). Esperamos que los estudiantes comprendan este experimento, su carácter aleatorio y la equiprobabilidad. Sin embargo, puesto que repiten 20 veces este experimento y, para el objetivo del proyecto, lo estadísticamente analizable es la secuencia de resultados a partir de esos 20 resultados, esperamos que manejen intuitivamente la extensión del experimento y la recursividad de la aleatoriedad de este nuevo experimento. Los estudiantes de nuestra muestra no han llevado cursos de probabilidad, así que no esperamos la conceptualización formal de un experimento aleatorio compuesto o de su espacio muestral. Sin embargo, sí se espera que el mismo contexto los ayude a manejar implícitamente el experimento aleatorio compuesto. Es deseable la concientización de que el objetivo del proyecto es el que dirige este proceso de extensión del experimento aleatorio simple, puesto que el lanzamiento de una moneda permitiría sólo una evaluación parcial de la aleatoriedad, ya que sólo sería posible analizarla a través de la variable «¿cara o cruz?».

- Las *características de la secuencia aleatoria* producida como consecuencia del experimento. La *aleatoriedad* también se puede observar en las características de la secuencia producida, puesto que a pesar de que la aleatoriedad se vincula con la imposibilidad de encontrar reglas o patrones, también lo es que existen formas de definir y estudiar la aleatoriedad de las secuencias. En nuestra investigación la aleatoriedad de las secuencias reales está garantizada por el proceso seguido, por lo tanto, el análisis de sus características le servirá para cuestionar la aleatoriedad de las que no se obtienen a través de tal proceso. Se espera que el estudiante encuentre patrones en las secuencias reales obtenidas y que los caracterice a través de las variables propuestas: número de caras, el número de rachas y la longitud de la racha más grande.

Como se ha indicado, además en este proyecto, se estudia la percepción subjetiva de la aleatoriedad, puesto que en la secuencia simulada se ponen a prueba las intuiciones de los estudiantes. Se pretende que el estudiante (y nosotros) estudien la *concepción* de los estudiantes de la clase sobre la aleatoriedad, como se ha hecho en muchos otros experimentos de estudio de la percepción de la aleatoriedad en sujetos adultos ya mencionados.

Como podemos observar, en la actividad se pone de manifiesto el hecho de que los dos componentes de la aleatoriedad descritos por Zabell (1992) se pueden separar, puesto que es posible, al menos teóricamente, que se pudiese simular una secuencia de modo que apareciese como aleatoria a los ojos de las personas que no supiesen como se ha generado, lo mismo que es posible hoy día que algoritmos deterministas (como un programa de ordenador) produzca secuencias de resultados «aleatorios».

Recalcamos que en este proyecto la aleatoriedad es recursiva, pues la secuencia de 20 resultados de lanzar una moneda por cada estudiante podría a su vez considerarse un nuevo experimento (proceso) aleatorio compuesto, cuyos sucesos serían los posibles ordenamientos con repetición de caras y cruces de longitud 20. La repetición del experimento por cada uno de los estudiantes, se podría considerar como una secuencia de resultados aleatorios de dicho experimento compuesto, de modo que teóricamente es posible estudiar los patrones de dichas «secuencias».

Esperamos también que, en el trabajo con el proyecto, los alumnos manifiesten algunas ideas sobre la aleatoriedad y parte de ellos aprecien las características esenciales de los fenómenos aleatorios descritas por Batanero y Serrano (1995) y Serrano (1996):

- En un experimento considerado aleatorio hay *más de un resultado posible* de acuerdo a las condiciones fijadas de antemano. Esperamos que el estudiante sea capaz de identificar los sucesos del espacio muestral correspondiente a los diferentes experimentos que aparecen a lo largo del proyecto.
- Con los conocimientos que posee el sujeto que emite el juicio, el resultado concreto que ocurrirá es *impredecible*. La *impredecibilidad* es una diferencia esencial entre experimentos aleatorios y deterministas. Sin embargo, algunos estudiantes podrían suponer que es posible predecir el resultado o al menos predecir una parte de los resultados de la secuencia aleatoria, mostrando la ilusión de control. Los estudiantes también pueden tener la creencia de que al emitir una conjetura sobre un experimento aleatorio, tienen que adivinar el resultado, mostrando el enfoque en el resultado aislado.
- Hay posibilidad -al menos teórica- de *repetir indefinidamente la observación* o producción del fenómeno. Esta es una característica de la concepción de aleatoriedad en la escuela de estadística frecuencial, que aparece en este estudio, puesto que una parte del proyecto consiste precisamente en repetir el experimento de lanzar la moneda en las mismas condiciones[4].
- Las secuencias de resultados obtenidas *carecen de un patrón* que el sujeto

[4] La asignación de probabilidades para la variable aleatoria «número de caras» podría ser clásica, sin embargo, por su nivel de estudios, no esperamos que los estudiantes la usen. En el estudio descrito en el Capítulo 5 la asignación de probabilidades era clásica o a priori.

pueda controlar o predecir. Esta visión está relacionada con la idea de impredecibilidad, y a la vez entronca con la concepción de aleatoriedad de Von Mises (1952/1928)[5], quien define que en cualquier secuencia aleatoria se pueden seleccionar subsecuencias que conservarán la misma frecuencia relativa sin importar el método de selección utilizado para armar las subsecuencias. Esta es la base de los contrastes de hipótesis a través de los cuales se prueban las tablas de números aleatorios (Batanero y Serrano, 1995). En esta investigación, no usaremos esta definición para probar aleatoriedad de las secuencias, sin embargo sí esperamos que los estudiantes manifiesten una idea de la relación entre la aleatoriedad y «carecer de patrón» al comparar entre sí sus secuencias y con las secuencias de sus compañeros.

❖ En el aparente desorden que caracteriza a la aleatoriedad, y también en aparente contradicción, pueden descubrirse *una multitud de regularidades globales*, comenzando por la estabilización de las frecuencias relativas de cada uno de los resultados posibles. Esta regularidad global es el fundamento que permite el estudio de estos fenómenos aleatorios mediante el cálculo de probabilidades (Serrano, 1996) y la que nos permitirá trabajar con los estudiantes en este proyecto. En el proyecto se analizarán algunos de los modelos vinculados al análisis de las secuencias de los lanzamientos de la moneda. En particular, el número de caras, número de rachas y la longitud de la racha más larga; definidos, precisamente, a partir del establecimiento de una variable aleatoria y su correspondiente variable estadística.

Estos diferentes componentes del objeto aleatoriedad están conectados con el conocimiento del estudiante, pues, en nuestro proyecto, para responder la pregunta si se tienen buenas intuiciones sobre la aleatoriedad, el estudiante debe movilizar su conocimiento matemático y cotidiano del término. Algunas de las concepciones (correctas o parcialmente correctas) sobre la idea de aleatoriedad que podrían aparecer en nuestra investigación son encontradas por Serrano (1996) y Batanero y Serrano (1999) también se deducen de nuestro análisis anterior:

❖ Aleatoriedad como *impredecibilidad o carencia de control*. Se manifestaría en los estudiantes que indiquen que no se puede saber o controlar el

[5] Von Mises, basándose en la idea de *colectivo*, definió la aleatoriedad como una serie de acontecimientos o procesos que «satisfacen la condición de ser reacios a toda ley o arbitrarios» (p. 45) y que también cumplen con las propiedades de que «la frecuencia relativa de sus atributos debe poseer valor límite… (y que) éstos valores límite no deben alterarse en las sucesiones parciales que podamos seleccionar de la original» (p. 46).

resultado de los experimentos aleatorios, o bien que no puede predecirse la secuencia particular que aparecerá. Esta concepción sería correcta si se refieren a un resultado en particular, pero deberá conjugarse la percepción de que algunos sucesos son más probables que otros y que, aunque no se puede predecir cada caso particular, sí se puede prever el comportamiento del colectivo. La creencia inversa es incorrecta; si los estudiantes piensan que se tendría una buena intuición cuando se acierta lo que va a ocurrir en un número grande de las repeticiones del experimento o bien cuando esperan que las secuencias simuladas y reales sean idénticas. También cuando piensen que podría haber un control sobre el experimento, de manera que es posible predecir la secuencia que se obtendrá.

Aunque la impredecibilidad y la carencia de control están vinculadas, pueden presentarse por separado. La impredecibilidad está vinculada con una incapacidad de predicción y la carencia de control con la imposibilidad de manipular o influir en el resultado en un experimento aleatorio.

- Aleatoriedad como *equiprobabilidad*, concepción que sería próxima a la existente durante los primeros pasos en el desarrollo del cálculo de probabilidades (Bennet, 1998), pero se espera que sea superada rápidamente. Se manifestaría en los estudiantes que mencionen que han de obtenerse 10 caras porque las caras y cruces son equiprobables. Nuevamente esto tiene parte de verdad, porque aunque 10 es el valor más probable, no tiene necesariamente que ocurrir.
- Aleatoriedad como *conjunto de modelos matemáticos* en la secuencia de resultados es la concepción actual; esperamos que esta concepción no sea espontánea en los estudiantes, pero el profesor la impondrá al sugerirles que analicen diversas propiedades de las secuencias. En el proyecto analizaremos su aceptación o reticencia a esta concepción al usar o no las propiedades que el profesor les pedirá que analicen.

El estudiante movilizará su conocimiento y creencias sobre la situación experimental; por ejemplo, su experiencia le dirá que no se puede conocer con seguridad el resultado de lanzar la moneda; por tanto esperamos que admita que nos encontramos ante un experimento aleatorio. Además sabe que no hay razón para preferir caras o cruces; es decir, aceptará la moneda como equiprobable. También suponemos que sabrá que el resultado de cada lanzamiento no depende de los anteriores. Finalmente, ellos esperarán un número aproximado de caras y cruces y cierta irregularidad pero no excesiva.

Esperamos que la unión del conocimiento estadístico y del contexto les lleve, con ayuda del profesor, a transformar la pregunta inicial, que es demasiado vaga, a unas preguntas que se puedan contestar, ¿Cuáles son las características que poseen las secuencias aleatorias? ¿Cuál será el número esperado de caras y cruces? ¿Cuántas rachas se podrían esperar en una secuencia aleatoria?, ¿qué tan largas? ¿Cuáles son las diferencias, en términos de las variables analizadas, entre una secuencia real y una simulada? ¿Qué medidas auxilian a este análisis y cuáles no? ¿Las mismas medidas tienen la misma importancia en todas las variables? Para finalmente recoger, resumir y analizar la información que, al final, les permitirá concluir en el contexto del problema.

Tabla 6.13. Evaluación de la comprensión de la aleatoriedad

Objeto matemático	Desglose	Evaluación en el proyecto
Aleatoriedad	Concepciones en ejemplos individuales Concepciones del grupo (globalmente)	Diferenciación entre resultado y proceso Referencia a impredecibilidad de resultados en el experimento Referencia a repetibilidad del experimento en sus argumentos Conexión entre aleatoriedad y causalidad Referencia a aleatoriedad como equiprobabilidad Referencia a aleatoriedad como variabilidad Referencia a aleatoriedad como ausencia o multiplicidad de modelos
		Características de las secuencias que produce el estudiante en la simulación Si las secuencias reflejan la equiprobabilidad Si reflejan la variabilidad aleatoria
Independencia de ensayos	Concepciones del grupo (globalmente)	Evaluación a partir de la independencia en las secuencias producidas por los estudiantes (rachas)

También es de interés conocer si los estudiantes ponen en práctica la concepción de independencia entre ensayos cuando inventan las secuencias, lo que observaremos en los datos generados por el grupo completo. En la Tabla 6.13 mostramos un esquema de estas concepciones y procedimientos que queremos evaluar en el proyecto y los indicadores que usaremos para llevar a cabo dicha evaluación.

5.2. Variable estadística y su distribución

La mayor cantidad de información recogida en los protocolos de los estudiantes se refiere a la concepción y uso de variables estadísticas y a los objetos relacionados con ellas: distribución, promedios y dispersión. En esta parte del análisis nos acercaremos a su comprensión desde los objetos matemáticos que es capaz de construir: los procedimientos usados, los conceptos puestos en juego y el lenguaje que utiliza a través de los indicios que nos muestre su informe escrito.

Nos concentraremos en las variables estadísticas ya descritas anteriormente en el Apartado 4.2 de este capítulo. El número de caras, rachas y la racha más larga varían de un estudiante a otro, aunque algunos valores pudieron haber coincidido. La tabla de datos recopilada en la clase mostraría como variable independiente a la repetición del experimento (o lo que es lo mismo, el número de orden en que se recogieron los datos de los estudiantes) y el resultado obtenido por cada uno de ellos para cada variable (por ejemplo, el número de caras) sería la variable dependiente correspondiente. Es necesario un primer análisis para encontrar la distribución de frecuencias, en donde la variable independiente sería el valor de la variable y la dependiente la frecuencia con la que se obtiene dicho valor.

Las concepciones de los estudiantes sobre la variable estadística se pondrán de manifiesto en el trabajo matemático que realicen los estudiantes para comparar las distribuciones de las secuencias simuladas y las reales: uso de tablas estadísticas y gráficos, momentos, interpretaciones y dificultades que se plantean. Analizaremos también si hay coincidencia en algunos de los errores aparecidos en la entrevista clínica analizada en el Capítulo 5, en particular, si el estudiante confunde las variables dependiente e independiente en la distribución de frecuencias.

La necesidad de trabajar con la variable estadística aparece cuando los estudiantes se dan cuenta de la necesidad de comparar las características de las secuencias simuladas y real en el grupo completo. A partir del conjunto de datos en bruto es difícil obtener alguna conclusión sobre la igualdad o diferencia de alguna de las variables que están trabajando porque los datos sin agrupar no permiten visualizar tendencias ni variabilidad. Esperamos que los estudiantes sean capaces de formar la distribución de frecuencias de cada una de las variables estadísticas y las presenten en una o varias tablas o gráficos similares a los analizados en el Apartado 4.3.

Se requiere un proceso de transnumeración (Wild y Pfannkuch, 1999) para, mediante el cambio de representación de los datos a una tabla de frecuencia o un gráfico, visualizar aspectos que nos permitan llegar a una mejor comprensión del sistema. Algunos estudiantes podrían no llegar a la transnumeración ni a formar la distribución de las variables. En vez de ello, podrían optar por estudiar los datos individuales, tal como han sido producidos y basar su juicio de la diferencia entre las secuencias reales y simuladas sólo en la comparación entre valores aislados. Esta conducta apareció en la investigación sobre asociación estadística en Batanero, Estepa y Godino (1997) en la comparación de dos muestras relacionadas o dos muestras independientes. Los autores denominaron *concepción local sobre la asociación* a la tendencia de algunos estudiantes a deducir una asociación entre variables o una diferencia de distribuciones basándose tan sólo en la comparación de algunos valores de los datos.

La elaboración de la tabla o gráfico supone una serie de pasos: clasificación de valores similares, recuento y cálculo de frecuencias absolutas, ordenación de los valores de la variable en orden creciente o decreciente y asignación de una frecuencia a cada valor de la variable (estableciendo entonces una función real de los valores de la variable en *R*). Algunos estudiantes podrían completar la tabla o gráfico con frecuencias relativas o porcentajes aunque en este proyecto, no es tan necesario porque para comparar las poblaciones se tomaron muestras de un mismo tamaño. A este respecto, podrían aparecer dificultades en la elaboración de las tablas o gráficos, como las descritas en Arteaga (2011): confusión de los ejes del gráfico, representar variables no relacionadas en el mismo gráfico, confusión con las escalas o la representación de los números reales en las mismas, o no completar el polígono de frecuencias.

También estamos interesadas en la idea de promedio y dispersión: si aparecen en forma intuitiva en los estudiantes, si calculan estos estimadores y cuáles, si hacen referencia a su definición o propiedades y si tienen errores de cálculo, en particular en el cálculo de medias ponderadas o no ordenar los datos en el cálculo de la mediana (Cobo, 2003; Mayén, 2009). Estas medidas resultan importantes porque, como señala Heitele (1975), el valor central y dispersión son dos características fundamentales de la variable aleatoria (por tanto lo son de la estadística) junto con su distribución.

Tabla 6.14. Evaluación de la comprensión de la variable estadística

Objeto matemático	Desglose	Evaluación en el proyecto
Variable estadística	Variable estadística como la variable de interés que toma diferentes valores en un conjunto de datos Diferenciación de conjunto original e imagen	Referencia a las variables como forma de evaluar aleatoriedad Referencia al rango de variación Referencia a que cada estudiante podría obtener un valor diferente Referencia a los distintos valores que toma la variable
Distribución de la variable estadística	Comprensión de la idea de distribución Representación de la distribución	Agrupación de valores similares, ordenación y cálculo de frecuencias Paso de datos desordenados a una distribución de frecuencias en tabla o gráfico Nivel de comprensión de la distribución
Medidas de posición central	Medidas de posición que se usan Comprensión de su significado	Medidas que se calculan Referencia que se hace a sus propiedades, definición o significado
Medidas de dispersión	Medidas de dispersión que se usan Comprensión de su significado	Medidas que se calculan Referencia que se hace a sus propiedades, definición o significado

Por otro lado, acordes con Konold, Pollatsek, Well y Gagnon (1997), estimamos que los estudiantes se apoyarán en las medidas de tendencia central y de dispersión para realizar las comparaciones entre las distribuciones. Sin embargo estos autores también indican que para algunos estudiantes no es intuitivo el uso de la media como representante para comparar dos conjuntos de datos. En su lugar, los estudiantes usan parte de los datos, por ejemplo comparan los máximos o el rango. También se podría presentar que los estudiantes usasen las medias pero las malinterpretasen o llegasen a una conclusión errónea en la comparación.

En la Tabla 6.14 indicamos, a modo de resumen, el desglose de los elementos que se evaluarán en las producciones de los estudiantes en relación a la variable estadística, entre los que incluimos el nivel de comprensión mostrado sobre la distribución, que deduciremos del tipo de gráfico o tabla construido.

5.3. Variable aleatoria y su relación con la variable estadística

El razonamiento distribucional implica establecer una conexión entre las muestras (variable estadística) y la población (variable aleatoria), así como pasar de una a otra y comprender la diferencia entre los parámetros (valores en la población) y los estadísticos (valores en las muestras) (Reading y Canada, 2011). Es decir, realizar un proceso informal de inferencia (Rossman, 2008).

En el proyecto surgen las variables aleatorias descritas en la Sección 4.2. Cuando cada uno de los estudiantes realiza un experimento y se juntan los datos de la clase, se tienen datos de las variables estadísticas correspondientes, que están relacionados con la distribución de variables aleatorias que subyacen en el experimento. Por tanto, nuestro acceso a las concepciones de los estudiantes sobre las variables aleatorias pasará, en primer lugar por las que manifiesten sobre las variables estadísticas correspondientes; y, en segundo lugar, por las relaciones que establecen entre ellas y las correspondientes variables aleatorias.

Para estudiar estas concepciones, analizaremos si los estudiantes hacen referencia al carácter aleatorio de las variables número de caras, número de rachas y longitud de la racha más larga en las secuencias simuladas y reales de sus compañeros, como variables cuyos resultados dependen de un experimento aleatorio. Es decir si explicitan las ideas de variable, valores e impredecibilidad, debida a la existencia de un experimento aleatorio y si hacen mención de su distribución y de sus valores esperados, así como de su dispersión. Analizaremos también si utilizan todas las variables del proyecto o bien sólo alguna de ellas, y si hacen referencia al rango o dominio de la variable. También nos interesan los argumentos que nos indiquen si de los estudiantes diferencian las variables aleatoria y estadística, es decir, la variable teórica vinculada con los resultados de sus experimentos, que puede tomar otros valores distintos a los que tomó cuando ellos hicieron el experimento y que está propiamente relacionada con la probabilidad, de la variable estadística que se refiere al experimento en concreto y está vinculada con la frecuencia.

No esperamos que los alumnos se refieran explícitamente al concepto de variable aleatoria, puesto que no han estudiado formalmente el concepto. Sin embargo, de modo implícito se tienen que utilizar cuando pedimos una conclusión sobre las características de las secuencias que serían capaces de producir los estudiantes y las características de las secuencias aleatorias en general, es decir, no restringidas a las producidas en el experimento. Es en este momento cuando, al hacer una inferencia, hay un uso implícito de las variables aleatorias.

No podemos observar la definición de las variables como una necesidad vinculada a la solución del problema que les fue propuesto en el informe escrito del proyecto, puesto que éstas surgieron en la interacción del profesor con el grupo en la sesión de clase (es decir, no se puede observar la construcción de la regla de correspondencia que vincula el espacio muestral con los números reales deducida a partir de la necesidad de resolver un problema). Pero sí podemos observar la comprensión de las variables propuestas a través del uso que el estudiante les da para resolver el problema propuesto.

La concepción de distribución de probabilidad de la variable aleatoria no se puede estudiar más que indirectamente a partir de los argumentos que los estudiantes proporcionan al comparar sus creencias (de equiprobabilidad, de variabilidad, de la forma de la distribución, etc.) con de lo obtenido. Algunos estudiantes podrían hacer referencia a los valores esperados de alguna de las variables que intervienen en el proyecto, por ejemplo hacer referencia a que en una secuencia de 20 lanzamientos se esperarían 10 caras (valor medio) o que lo más probable sería obtener 10 caras (moda de la distribución).

Con todo lo descrito anteriormente, se podría observar la forma en que los estudiantes relacionan la variable estadística con la aleatoria, tanto en la generalización de una variable estadística a una variable aleatoria (considerada cuando se repite indefinidamente el experimento) como viceversa, es decir, cuando pasa de la variable aleatoria teórica considerada en el experimento a la variable estadística como conjunto de valores obtenidos en una muestra de resultados de la variable aleatoria. En la Tabla 6.15 describimos los criterios mediante los cuales nos acercamos a las concepciones de los estudiantes sobre la variable aleatoria.

Tabla 6.15. Evaluación de la comprensión de la variable aleatoria en el proyecto

Objeto matemático	Desglose	Evaluación en el proyecto
Variable aleatoria	Percepción de la variable aleatoria	Referencia a variable cuyos valores dependen de los resultados de un experimento aleatorio Variables que usan (entre las discutidas en el proyecto)
Distribución de probabilidad de una variable aleatoria	Definición de los valores posibles de la variable aleatoria y estimación de su probabilidad	Referencia al rango en la variable aleatoria Estimación de probabilidades de algunos valores a partir de experimentación Generalización de la distribución de frecuencias obtenida a la distribución de probabilidad
Momentos de la variable aleatoria	Comprensión de ideas de posición central o dispersión	Referencia a tendencias o dispersión de las variables Estimación de valores esperados

5.4. Ciclo de modelación

La conceptualización de las variables aleatorias y estadísticas también está relacionada con la capacidad del estudiante de vincular la herramienta matemática con el objetivo de la pregunta en cuestión en las diversas fases del ciclo de modelación (Chaput, Girard y Henry, 2011), tal como se describe en el Punto 9.6 del Capítulo 3. En la situación problema de nuestra investigación, se observaría por la forma en que el estudiante es capaz de vincular el trabajo estadístico con las variables que surgieron a partir de la discusión en clase orientada por su profesor con la respuesta a la pregunta sobre las intuiciones del grupo con respecto al azar del grupo. La forma en cómo concebimos este objeto de análisis está sintetizado en la Tabla 6.16.

Tabla 6.16. Evaluación de competencia en el ciclo de modelización

Desglose	Evaluación en el proyecto
Formulación de preguntas	Si el alumno plantea preguntas nuevas o reformula las dadas por el profesor
Construcción de modelos matemáticos	Gráficos que construye Modelos que utiliza
Trabajo con modelos matemáticos	Corrección de los gráficos construidos Corrección de las tablas y resúmenes estadísticos Corrección de las expresiones matemáticas
Interpretación de resultados matemáticos	Interpretación que hace de los resultados
Interpretación de resultados en relación al contexto	Conclusiones respecto al problema planteado

Adoptamos la percepción de Wild y Pfannkuch (1999) en cuanto a su definición de *modelo estadístico*: «se refiere a todas nuestras concepciones estadísticas del problema que influyen en cómo recolectamos datos acerca del sistema y cómo lo analizamos» (p. 230). De esta forma, consideramos que el estudiante trabaja con modelos matemáticos en cuanto produce un gráfico o calcula un estadístico, puesto que éstos se convierten en partes de la forma en que se evalúa, representa o visualiza un conjunto de datos. Además, en estos gráficos, tablas o resúmenes el estudiante ha abstraído de la realidad, pasando de ella (estudiantes, sus intuiciones) a un objeto matemático (tabla, gráfico, frecuencias, variables) con el que trabaja.

La inferencia informal que se pretende que el estudiante haga, plantea el ciclo de modelación completo que describe Dantal (1997), puesto que para que el estudiante pueda contestar la pregunta de investigación del proyecto, es necesario que visualice a la variable aleatoria como una generalización de la variable estadística y a la probabilidad como el modelo teórico de la frecuencia relativa para poder formular la distribución del modelo teórico último de la situación. A partir de ese modelo, es necesario retornar a la realidad y concluir en ella. El proceso, por tanto se describe a través de una modelación por estratos, tal como se describe en la Figura 3.6 del Capítulo 3.

El análisis de la variable estadística, así como la forma en que se establece la relación entre la variable aleatoria y la variable estadística ya se trataron en los puntos anteriores (5.2 y 5.3). Para no ser repetitivos, en objeto de análisis nos enfocaremos de manera más intensa en la corrección del uso de dichos modelos matemáticos y, en particular, a las fases iniciales y finales del proceso, describiendo someramente las fases iniciales. En la última fase queremos ver si el estudiante interpreta correctamente el resultado de su trabajo matemático, tanto desde una perspectiva matemática, como al contexto de la investigación planteada. Es decir, tanto si realizan una inferencia informal, como si completan el ciclo de investigación de Wild y Pfannkuch (1999).

Como indica Dantal (1997) los profesores en las clases de matemáticas insisten más en el trabajo con el modelo (cálculos matemáticos) y no en las fases iniciales (planteamiento de las preguntas y construcción del modelo) y finales (interpretación de resultados para sacar unas conclusiones) del ciclo de modelización. Es por esto que esperamos que los estudiantes de la muestra tengan problemas en llegar a las conclusiones finales sobre las intuiciones de los estudiantes de la clase.

6. Hipótesis del estudio

Tomamos como base el análisis a priori del proyecto y de la solución esperada por los estudiantes, así como la solución matemática del mismo (a través de inferencia) para elaborar nuestras hipótesis sobre los resultados que podemos esperar en el análisis de las producciones de los estudiantes.

- *En nuestro trabajo se encontrarán algunas de las concepciones (correctas o parcialmente correctas) sobre la idea de aleatoriedad descritas por Serrano (1996).* Esta hipótesis se basa en los resultados de las investigaciones sobre la percepción subjetiva de la aleatoriedad realizadas con personas adultas, entre ellas, las de Serrano (1996) con una muestra de futuros profesores de educación primaria españoles, similares a los participantes en nuestro estudio. En uno de los ítems de su estudio se mostraban varias secuencias de resultados de lanzamientos de una moneda equilibrada y se pedía decidir si cada una de ellas era aleatoria y argumentar su respuesta. En otro ítem, pidió a los estudiantes que escribieran una secuencia de resultados de lanzar una moneda 50 veces sin hacerlo realmente (una secuencia simulada). Aunque la tarea de nuestro estudio es diferente, tiene partes comunes con las de Serrano, como el pedir a los participantes escribir una secuencia simulada y justificar sus razonamientos sobre secuencias aleatorias. Esperamos, por lo tanto, que en nuestro trabajo se reproduzcan parte de los resultados de dicho autor.
- *Una proporción importante de estudiantes harán uso de las variables estadísticas y de la idea de distribución en forma intuitiva y al menos parcialmente correcta.* Esta hipótesis se deduce de los resultados de las investigaciones que se han realizado utilizando este mismo proyecto (Godino, Batanero, Roa y Wilhelmi, 2008; Arteaga, 2008; 2011). Aunque ninguno de estos trabajos se centró específicamente en la variable aleatoria, sus resultados muestran que en muchos casos los estudiantes construyeron la distribución de algunas de las variables estadísticas, como el número de caras y las compararon intuitivamente. Por otro lado, pudieran aparecer las dificultades ya citadas y descritas por Pollatsek, Well y Gagnon (1997) o Batanero, Estepa y Godino (1997) en estudiantes de magisterio al comparar dos distribuciones.
- *Aparecen dificultades en la representación gráfica de las distribuciones.* Esperamos encontrar las mismas dificultades encontradas en el trabajo realizado por Arteaga (2011) en donde hace un análisis gráfico del mismo proyecto en una muestra de estudiantes más amplia.
- *El uso de los promedios es mucho más intuitivo que el uso de la dispersión.* La comprensión y coordinación del carácter dual de la distribución involucra las ideas de aleatoriedad, independencia, tendencia, valor esperado y variabilidad (Shaughnessy, 2007). Sin embargo, los resultados de

investigaciones sobre comparación de distribuciones por parte de estudiantes indican que muchos se basan únicamente en la tendencia central, sin tener en cuenta la distribución de los datos. Shaughnessy y Ciancetta (2002) presentan ejemplos de cómo piensan los estudiantes al razonar sobre las distribuciones de datos e indican que algunos estudiantes que se centran demasiado en los valores centrales, son incapaces de pasar de estos valores a la idea de distribución y tienen dificultades con la idea de variabilidad y dispersión.

- *Los alumnos usan intuitivamente algunas características de la variable aleatoria, en particular tienen una buena concepción del valor esperado.*En muchos de los estudios que hemos citado en este capítulo se estudia la distribución de una o varias variables estadísticas o aleatorias independientemente. Kazak y Confrey (2007) sugieren que estos dos conceptos debieran ser relacionados en la enseñanza y en la investigación. El proyecto planteado tiene esa intención y esperamos que los alumnos puedan relacionar la variable aleatoria y estadística intuitivamente. Por otro lado, las investigaciones sobre percepción de la aleatoriedad (como, por ejemplo, Falk y Konold, 1997; Batanero y Serrano, 1995 y 1999) indican que los sujetos tienen una buena percepción del valor esperado en la distribución binomial, que será una de las distribuciones utilizadas en nuestro estudio. Por lo tanto esperamos que este concepto sea ampliamente empleado en sus inferencias hacia la variable aleatoria.
- *Los alumnos no completan el ciclo de modelación, limitándose al trabajo con los modelos matemáticos, sin llegar a interpretar los resultados en el contexto del problema.* Proponemos una situación de modelación para acceder a la comprensión de la variable aleatoria, puesto que la relación entre las variables aleatoria y estadística está fuertemente involucrada en el proceso de modelación (Apartado 9.5 del Capítulo 4). Los alumnos de nuestra investigación sólo habían trabajado anteriormente con un proyecto estadístico en la asignatura que nos ocupa. En consecuencia, esperamos que tengan dificultades en completar la última fase del ciclo de modelización (iinterpretación de resultados en la realidad) del ciclo propuesto por Dantal (1997). El autor indica que en las clases de matemáticas no se suele dar énfasis a este paso. Además, Arteaga (2011), trabajando con el mismo proyecto, encontró que muchos estudiantes realizaron gráficos, pero no los leyeron ni obtuvieron conclusiones de ellos. Esta dificultad en obtener

conclusiones de los gráficos podría extenderse al resto del análisis hecho por los estudiantes, quienes tampoco obtendrían conclusiones del cálculo de estadísticos o de las tablas o emitirán una conclusión general sobre sus intuiciones sobre el azar en donde se involucren todas las variables.

- *Se reproducen algunas de las dificultades encontradas en la entrevista clínica, en particular la confusión de algunos estudiantes entre variable dependiente e independiente.* Esta hipótesis se basa en los resultados de la entrevista (Capítulo 5) y también en algunos resultados de Arteaga (2011).

7. Análisis de las producciones de los estudiantes

En esta sección presentaremos a un tiempo mismo los resultados y discusión del estudio de evaluación. Cada estudiante entregó un reporte por escrito de su solución en forma individual. En el Anexo 4 se muestran algunos reportes entregados que consideramos representativos del trabajo realizado.

Las producciones de los estudiantes variaron, tanto en extensión (entre 1 y 12 páginas), como en los tipos de análisis estadísticos realizados (gráficos, tablas, resúmenes estadísticos). Dos tercios de los estudiantes entregaron su informe en lápiz y papel; y una tercera parte hizo uso un procesador de texto. Estos últimos también utilizaron la hoja Excel, aunque, desde el análisis que expondremos, esto no siempre supuso una ventaja.

El análisis de contenido de los informes se realizó siguiendo la metodología de análisis de producciones escritas, descrita en el Apartado 5.3 del Capítulo 1. Fue de tipo cualitativo y cíclico, esto es, se efectuaron varias revisiones del reporte de cada estudiante, comparándolo con el de sus compañeros y con el análisis a priori descrito en las secciones 4 y 5 de este capítulo. A partir de este análisis, logramos identificar inductivamente algunas categorías de respuestas de cada uno de los apartados en que dividiremos esta sección y que corresponden a los diferentes objetos matemáticos de interés en el estudio. En este apartado analizamos, la forma en que dichos objetos se ponen de manifiesto en las producciones escritas de los estudiantes, su corrección matemática y la comprensión que se deduce de sus respuestas. También tomaremos en cuenta las investigaciones que han descrito dificultades de los estudiantes con cada uno de los objetos analizados.

7.1. Objeto de análisis: Aleatoriedad

El estudio de la comprensión de los estudiantes alrededor de la aleatoriedad es importante porque éste es el contexto en el cual se desarrolla el proyecto. El objetivo detonador del proyecto es hacer que el estudiante se cuestione acerca de sus mismas concepciones sobre la aleatoriedad, lo que requerirá hacer uso crítico de la herramienta estadística y probabilística. Por otro lado, la aleatoriedad, como objeto matemático, tiene un papel crucial dentro de la red de objetos emergentes vinculados a la variable aleatoria, por lo que su comprensión influirá en la forma en que trabaje y comprenda este objeto.

Como hemos analizado en la Sección 5.1, dentro del proyecto, la aleatoriedad se manifiesta en el lanzamiento de una moneda equilibrada, experimento aleatorio que se utiliza para recoger los datos que se analizan en el proyecto y como referencia para que los estudiantes evalúen las concepciones sobre aleatoriedad del grupo completo, haciendo uso, también, de sus propias concepciones sobre la aleatoriedad. De acuerdo a la Tabla 6.13, en esta sección analizaremos de dos maneras las producciones escritas para visualizar las concepciones sobre la aleatoriedad de los estudiantes desde dos enfoques:

- ❖ *Individualmente*, del análisis de los argumentos que los estudiantes emplean a lo largo del proyecto para obtener sus propias conclusiones o realizar su propio análisis. Esto se puede observar principalmente cuando se refieran al experimento que ellos realizaron y que consiste en lanzar 20 veces la moneda y sus posibles resultados y en las conclusiones de su análisis, cuando evalúen la intuición del grupo de alumnos de la clase sobre aleatoriedad, pues en esta evaluación sin duda influyen sus propias concepciones del objeto.
- ❖ *Globalmente*, al analizar las características de las secuencias aleatorias simuladas por los estudiantes, pues la tarea es similar a las propuestas en estudios sicológicos sobre la percepción de la aleatoriedad (Falk y Konold, 1997; Batanero y Serrano, 1995 y 1999). El análisis de los datos del total de los estudiantes será similar al presentado en la Sección 5.1. Los resultados se compararan con otras investigaciones sobre la percepción subjetiva de la aleatoriedad.

Esperamos que estos dos enfoques de análisis de las concepciones de los estudiantes sobre la aleatoriedad nos proporcionen un panorama más amplio sobre la forma en que visualizan y usan las variables estadísticas y aleatorias involucradas. Avances de estos resultados se han presentado en Arteaga, Batanero y Ruiz (2010) y Batanero, Arteaga, Ruiz y Roa (2010).

7.1.1. Evaluación individual de las concepciones sobre la aleatoriedad

La información de las concepciones individuales de los estudiantes dejó vislumbrar ideas que se aproximan a los diferentes significados históricos sobre aleatoriedad, similares a las descritas por Batanero y Serrano (1995) y detectadas en los estudiantes por Batanero y Serrano (1999). No todos los estudiantes explicitaron por escrito estas concepciones, pues no se les pidió explícitamente una definición de aleatoriedad ni se les preguntó sobre las características que atribuyen a las secuencias aleatorias. Sin embargo algunos de ellos hicieron referencia a la aleatoriedad en general o al carácter aleatorio de las secuencias al tratar de concluir sobre las intuiciones del conjunto de la clase con respecto a los experimentos aleatorios.

A continuación presentamos las categorías en que clasificamos los argumentos de los estudiantes.

Aleatoriedad como impredecibilidad

Bennett (1998) indica que durante mucho tiempo, se asoció la aleatoriedad con la incapacidad de predecir un resultado. Hoy en día, esta es una concepción parcialmente correcta, puesto que si ciertamente el resultado individual de un experimento aleatorio no se puede vaticinar, el conocimiento de la distribución de probabilidades del fenómeno permite explicar y predecir (Laplace, 2006/1814). La acepción de la incapacidad de predecir un resultado se volvería incorrecta si el estudiante no apreciase el papel del cálculo de probabilidades en el estudio de las distribuciones del conjunto de resultados.

En nuestro estudio, algunos estudiantes que manifestaron que la aleatoriedad era impredecible también mencionaron que cualquier cosa podría presentarse en la distribución de los resultados de las secuencias, puesto que el proceso había sido aleatorio. Es decir, no apreciaron la estabilidad de las frecuencias debida a la convergencia, ni la mayor probabilidad de unos resultados frente a otros. Por ejemplo, el testimonio de AA que niega cualquier tipo de predicción:

> Sobre todo destacamos que es imposible hacer una previsión de resultados ya que en este tipo de experimentos aleatorios todo resultado es imprevisible [*Alumno AA*].

Konold (1989) describió esta actitud mostrada por los dos estudiantes, como el *enfoque en el resultado aislado*. Se caracteriza porque los estudiantes no interpretan las preguntas de una forma probabilística, sino que enfocan la pregunta hacia saber cuál será exactamente el siguiente resultado, lo que les impide tomar una decisión. Fischbein (1975) y Fischbein, Nello y Marino (1991) reportan que esta incapacidad de emitir una predicción sustentada es debida a la falta de habilidad para considerar la estructura racional del riesgo, es decir, de sintetizar lo necesario (en el sentido filosófico de causa y dependencia) y lo aleatorio en el concepto de probabilidad. Ellos también indican que este comportamiento está fuertemente enraizado en los estudiantes.

El estudiante MN también hizo referencia a la imposibilidad de predicción total, pero a diferencia de SG y AA, deja vislumbrar algún tipo de posibilidad de predicción (inferimos que piensa en el cálculo de probabilidades al hacer referencia a sucesos más y menos probables). En este caso la concepción sería correcta:

> Mi valoración final es que en las experiencias relacionadas con el azar existen sucesos más o menos probables pero resulta imposible predecir con exactitud el resultado completo [*Alumno MN*].

Aleatoriedad como ausencia de control

Otros estudiantes han mostrado una concepción de aleatoriedad bajo la cual los acontecimientos aleatorios escapan al control humano. De acuerdo con Batanero, Henry y Parzysz (2005) esta es una de las acepciones presentes en una primera fase histórica exploratoria del desarrollo de la idea de aleatoriedad que se extiende desde la antigüedad hasta el comienzo de la Edad Media. En esta etapa los dispositivos aleatorios, como loterías, dados o huesos de astrágalo, se usaban para predecir el futuro y tomar decisiones, bajo el supuesto de que la voluntad de los dioses gobernaba tanto el futuro como los dispositivos aleatorios y, por lo tanto, era de esperar que los fenómenos aleatorios dieran indicio de sus designios para el futuro (Hacking, 1975; Gutiérrez, 1992; Bennett, 1998). En esta etapa también se usaban estos dispositivos en los juegos para impedir que alguna de las partes interesadas tuviera ventaja, pues se suponía que lo aleatorio no podía ser controlado humanamente. Ya sea que se le atribuyera a fuerzas sobrenaturales o no, el azar suprimía la posibilidad de que la voluntad, inteligencia o conocimiento humano influenciara de alguna forma este tipo de fenómenos (Batanero y Serrano, 1995; Batanero, Serrano y Green, 1998).

Al igual que en el caso anterior, hoy día esta concepción es parcialmente correcta, pues aunque es cierto que no se puede controlar totalmente un experimento aleatorio, a veces se pueden determinar variables que afectan a la variabilidad y de este modo reducirla. Es decir, nuevamente el estudio de las distribuciones puede tener un papel importante en el control parcial de la aleatoriedad en los grandes números. Algunas respuestas que explicitan la imposibilidad de control en nuestro estudio son las siguientes:

> Este resultado ha aparecido en la secuencia real a pesar de no tener nosotros ningún control en el azar, el resultado ha sido igual entre el número de caras y de cruces [*Alumno AG*].

> A pesar de nuestra incapacidad para controlar el azar, obtuvimos igual número de caras y cruces [*Alumna SF*].

También se observó el caso contrario, es decir, estudiantes que piensan que los fenómenos aleatorios pueden ser controlados. Fischbein, Nello y Marino (1991) reportan una investigación en la que los estudiantes pensaron que el lanzamiento de un dado puede llegar a ser controlado por un jugador experto a tal extremo que los resultados serían los que él desea, presentándose la ilusión de control (Langer, 1975).

Un caso extremo lo encontramos en el estudiante LG, quien clasificó los resultados de todos sus compañeros de grupo de acuerdo a su capacidad para predecir los resultados, suponiendo que una buena capacidad para predecir implicaría una concepción correcta de la aleatoriedad. Obtuvo las frecuencias de los alumnos con «*buenas y malas intuiciones*» y encontró la distribución de los alumnos de acuerdo al «error cometido». En la Figura 6.6, reproducimos el gráfico construido por él, en el que muestra el porcentaje de estudiantes que «*acierta exactamente*», «*aciertan con un margen de error de* ± *1*» o «*tienen un margen de error de* ± *2*»:

> Tan sólo el 21,73% de la muestra han acertado el nº de caras (y consecuentemente el nº de cruces) del experimento. El 39,13% de la muestra se acercaron a la previsión del juego de azar con un margen de (± 1); El resto, 18, tuvo una mala aproximación (± 2) [*Alumno LG*].

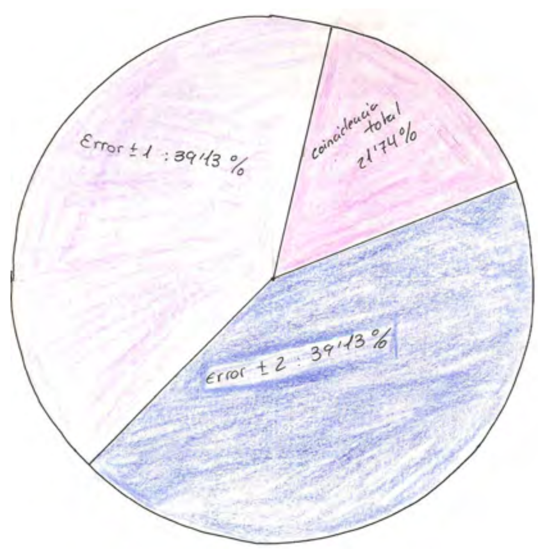


Figura 6.6. Distribución de los estudiantes en el grupo según sus «aciertos en el experimento» de acuerdo con los criterios establecidos por el estudiante LG

Aleatoriedad y causalidad

A través de la historia ha prevalecido la concepción de la aleatoriedad como lo opuesto a las causas conocidas, pero también se interpreta como si la aleatoriedad fuera una causa en sí misma. Poincare (1979/1908) menciona que los filósofos clásicos distinguían dos clases de fenómenos: aquellos que parecían obedecer a leyes armónicas, establecidas de una vez y para siempre, y aquellos que atribuían al azar y que no podían preverse porque se revelan a toda ley. El autor indica que cuatrocientos años antes de Cristo, Demócrito postulaba que todo cuanto existía en el mundo era fruto de la razón y la necesidad. Esta postura estaba vinculada con la doctrina atomista en la que todos los fenómenos respondían a una causa, con excepción de los átomos, que eran eternos y móviles. De acuerdo con Bennett (1998) los fenómenos aleatorios no eran vistos como algo «sin causa», sino que más bien, el azar era una especie de «causa oculta» (en el sentido de desconocida). Durante los siglos XVII, XVIII y XIX, esta postura se robusteció con el fortalecimiento y surgimiento de nuevas herramientas probabilísticas. Laplace (2006/1814) atribuía el azar con el que se explicaban algunos acontecimientos, como *causas imaginarias* «que se descartarían a medida que se ampliaran las fronteras del conocimiento y tenderían a desaparecen por completo ante la sana filosofía que no ve en ellas más que la expresión de nuestra ignorancia de las verdaderas causas» (p. 361). De modo que la concepción de aleatoriedad como causa puede estar vinculada con la controversia filosófica entre el *azar* y la *necesidad*, que no ve en el azar sino la medida de nuestra ignorancia.

Poincare (1979/1908) comenzó a diferenciar el concepto matemático del filosófico y propone la existencia de *eventos fortuitos* cuya principal característica es que siguen las leyes del azar (vistas desde una perspectiva probabilística). Poincare continuaba creyendo que todos los efectos son producidos por causas, pero aceptó que las causas pueden ser complejas y numerosas o que un cambio pequeño (o difícil de medir) en las causas produce un efecto grande, lo que dificultaría conocer o aplicar la ley que los rige (en el sentido causal). Para Poincare el azar no es producto de nuestra ignorancia, puesto que las herramientas probabilísticas proporcionan conocimiento sobre los fenómenos aleatorios y aun cuando se pudieran descubrir las leyes causales que los rigen, su modelo probabilístico seguiría funcionando.

La relación entre el azar y la causalidad sigue siendo tema de controversias aún en nuestros días. Actualmente el paradigma determinista prevalece en muchos campos científicos y desde este paradigma la ocurrencia de un acontecimiento con probabilidad muy pequeña (*inverosimilitud de los hechos extraordinarios*, desde la perspectiva de

Laplace) está vinculada con una causa que, según Newton, «no podía ser fortuita, sino dotada de grandes dosis de Geometría y Mecánica» (Hawking, 2006, p 355).

Actualmente, los matemáticos han desligado el objeto aleatoriedad de la idea de azar para sustentar el desarrollo del cálculo de probabilidades, independientemente de los debates filosóficos y epistemológicos alrededor de la idea de azar (Batanero, Henry y Parzysz, 2005). No es raro que nuestros estudiantes defiendan la postura del azar como causa para argumentar aleatoriedad, puesto que en su grado de estudios el concepto matemático de aleatoriedad, no es muy tratado. Los modelos matemáticos involucrados en su definición son complicados aún para un estudiante universitario, pero también influye que el límite entre ambos conceptos aún es confuso y ha generado controversias.

La respuesta del alumno AA denota este tipo de creencia:

> Creemos que los datos obtenidos tanto en la secuencia simulada como en la secuencia real son fruto del azar [*Alumno AA]*.

Suceso aleatorio vinculado a un proceso temporal

De acuerdo con Batanero, Henry y Parzysz (2005), históricamente las primeras aproximaciones al concepto de probabilidad (Pascal y Fermat, 2007/1654; de Moivre, 1756; Bernoulli, 2006/1713) relacionaron la aleatoriedad de los resultados posibles de una situación azarosa con la incertidumbre de eventos futuros, es decir estaban preocupados con la predicción. Sin embargo, Laplace (2006/1814) ya maneja ejemplos en los que se pregunta por la probabilidad de eventos ocurridos en el pasado, como es el problema en el que analiza la probabilidad de que un testigo mienta en su testimonio (p. 368).

En el ámbito educativo, esta concepción de temporalidad la describió Falk (1986) quien encontró que algunos estudiantes asociaron la probabilidad con un eje temporal, generalmente futuro. Esto produce que estos estudiantes tengan dificultad para asignar probabilidades a un suceso que ocurrió en el pasado.

Esta misma concepción aparece más explícitamente en el siguiente caso. La definición del estudiante es, además, redundante, pues utiliza el concepto que quiere definir (aleatorio) en su definición:

> Proceso estocástico es un concepto matemático que sirve para caracterizar y estudiar todo tipo de fenómenos aleatorios (estocásticos) que evolucionan, generalmente, con el tiempo [*Alumno GS*].

Aleatoriedad como equiprobabilidad

Bennett (1998) indica que la equiprobabilidad no podía ser un concepto con el que nuestros ancestros estuvieran familiarizados, debido, principalmente, a la falta de utensilios «simétricos» para realizar experimentos aleatorios. Los astrágalos guardaban distintas proporciones geométricas en sus lados además de que tal proporción variaba dependiendo de la clase de animal de la que se extrajera el hueso. Sin embargo, hay evidencia de que la irregularidad fue superada rápidamente. Hacking (1975) sostiene que los dados tallados en marfil y otros instrumentos regulares existieron desde hace mucho tiempo. En los siglos XII y XVI, en los primeros trabajos sobre teoría de juegos de De Fournival (Bellhouse, 2000) y de Cardano (1953/1663) ya se observa la descripción de juegos de azar con instrumentos equilibrados[6].

De acuerdo con Batanero y Serrano (1995), la relación de estos primeros desarrollos teóricos con los juegos de azar, en los que el número de posibilidades era finito y equiprobables, ocasionaron que al surgir el cálculo de probabilidades, la aleatoriedad se relacionara con la equiprobabilidad. Laplace (2006/1814), por ejemplo, describe que «la teoría del azar consiste en reducir todos los acontecimientos del mismo tipo a un cierto número de casos igualmente posibles» (p. 363). Sin embargo, sí había un conocimiento de la diferencia entre los experimentos en donde los resultados eran igualmente posibles y en los que no. Así, a mediados del siglo XVIII, de Moivre (1718) en su libro *The doctrine of chances* se preocupó por explicar a sus lectores la existencia de posibilidades no uniformes.

Batanero, Henry y Parzysz (2005) explican que estas controversias en el desarrollo y filosofía de la probabilidad han tenido influencia en la enseñanza de la probabilidad. En particular, la enseñanza de la visión clásica de la probabilidad ha favorecido de una manera simplista la visión de equiprobabilidad como una característica de aleatoriedad. Lecoutre (1992) describió el «*sesgo de equiprobabilidad*» como la creencia de que todos los sucesos asociados a cualquier experimento aleatorio tienen la misma probabilidad, incluso aquellos experimentos en los que no es aplicable el principio de indiferencia o donde no hay una simetría física.

En nuestro experimento hay *equiprobabilidad* en los resultados de la variable aleatoria ζ_r, Bernoulli, referida al lanzamiento de una moneda. Sin embargo, la variable aleatoria η_r, número de caras en las secuencias de 20 lanzamientos, sigue una distribución Binomial, por ello sería mucho más probable que nuestros estudiantes obtuvieran los valores centrales de la distribución (como 9, 10, 11 caras) que los

[6] En el mismo sentido, a principios del siglo XVII Galileo (citado por Bennett, 1998, p. 47) detalla las características de un dado justo: «tiene 6 caras y cuando se tira pueden caer por igual cualquiera de ellas».

extremos. En la siguiente justificación del estudiante EB se visualiza, sin embargo, la visión de aleatoriedad como equiprobabilidad y la confusión entre los sucesos compuestos o los valores de la variable aleatoria con los sucesos simples:

> La probabilidad de cara y cruz es la misma, de modo que en 20 lanzamientos hay la misma posibilidad de obtener 20 caras, 20 cruces o cualquier posible combinación de caras y cruces [*Alumno EB*].

Este alumno cae en el *sesgo de equiprobabilidad* (Lecoutre, 1992). Lecoutre indica que esta creencia no es debida a la falta de razonamiento combinatorio, sino a que los modelos combinatorios no se asocian fácilmente con las situaciones aleatorias. Este sesgo supondría una extensión errónea de la regla de Laplace, en donde manipulan los sucesos compuestos o los valores de la variable aleatoria como si fueran sucesos simples.

En la siguiente respuesta se detecta, además, la confusión entre el resultado del experimento aleatorio y los valores de la variable aleatoria (0, 1, 2, 3, 4,…20) y por eso a cada uno de ellos les asigna igual probabilidad. No se percata que detrás de los valores de la variable aleatoria hay sucesos compuestos, que se podrían desglosar en sucesos simples y que éstos son los resultados del experimento aleatorio. Esta dificultad también se observó en la entrevista clínica y se describió en la Sección 7.2.3 del Capítulo 5.

> Cualquier resultado puedo ocurrir; ya que se trata de un experimento aleatorio, hay igual probabilidad para cada resultado [*Alumno CG*].

Aleatoriedad como variabilidad

Algunos estudiantes hacen referencia a la variabilidad de los resultados como una característica de los procesos aleatorios o justifican que el proceso sea aleatorio precisamente porque los resultados «varían». En términos de Wild y Pfannkuch (1999) la percepción de la variabilidad es uno de los nodos fundamentales del razonamiento estadístico porque la variabilidad es una característica esencial de los fenómenos estadísticos. Estos alumnos, por lo tanto, muestran en su respuesta una concepción correcta que se sustenta en un razonamiento estadístico fundamental.

Las respuestas siguientes aluden a esta variabilidad; en la primera, también se refiere al azar como un tipo de causa y además se hace referencia a la variedad de valores, por lo que implícitamente se está aludiendo a la idea de variable aleatoria. En la segunda, el estudiante IE deduce que los experimentos aleatorios han de presentar mayor variabilidad que los producidos por las personas. Cosa que, de acuerdo con el análisis de los datos obtenidos en el proyecto (Sección 4.3), es cierta, por lo que muestra no sólo una buena intuición sino también que ha sido capaz de obtener una conclusión del comportamiento del grupo en la actividad.

> Hay más variedad de valores que aparecen al azar. Son lógicos estos resultados ya que son de forma aleatoria y en mano del azar [*Alumno NC*].

> En el número de caras obtenidas se observa una diferencia... ya que al inventar nosotros las simuladas los datos son más igualados y los reales son más desiguales al haberlos obtenidos al azar [*Alumno IE*].

Aleatoriedad como ausencia de modelos

Una idea muy extendida es que la aleatoriedad está asociada con la desorganización, irregularidad y ausencia de patrones o modelos, incluso se ha contrapuesto a la idea de «conocimiento científico» (Falk, 1981).

Serrano (1996) considera que este pensamiento coincide con la idea de von Mises (1952/1928), quien sostuvo que en una secuencia aleatoria no es posible obtener una algoritmo que la reproduzca a ella ni a ninguna de sus posibles sub-sucesiones. También puede estar vinculado con la definición basada en las ideas de Kolmogorov: una secuencia aleatoria es aquella con una máxima complejidad; es decir, una ley simple no puede describir la secuencia (Bennett, 1998, p. 163). Las dos visiones coinciden en vincular la aleatoriedad con el «desorden» e «irregularidad» de las secuencias de resultados aleatorios, pero ambas tienen dificultades prácticas en los procedimientos a seguir para distinguir una secuencia aleatoria de otra que no lo es. Además de que la primera sólo podría asegurar la aleatoriedad en secuencias infinitas y la segunda establece grados de aleatoriedad para secuencias finitas, pero difícilmente se podría validar la aleatoriedad perfecta. Serrano (1996) señala que a finales del siglo XIX, con el desarrollo de la inferencia estadística, se impulsó el estudio de modelos útiles para probar aleatoriedad y evaluar las tablas de números aleatorios o los generadores aleatorios de calculadoras y ordenadores. Desde la concepción moderna, la aleatoriedad no se visualiza como una cualidad del proceso, sino como una colección de modelos matemáticos que podemos aplicar al resultado (Serrano y Batanero, 1995; Serrano, Batanero y Cañizares, 1999). Sin embargo, la controversia sigue en pie: buscando la manera de caracterizar el aparente desorden de las secuencias aleatorias, han surgido conceptos como *aleatoriedad local* y pruebas específicas que se basan en la utilidad práctica de la aleatoriedad sustentada en modelos particulares viables para cierto tipo de secuencias. También continúa la discusión de la relación entre *necesidad y azar*, porque muy recientemente han encontrado procesos aleatorios que exhiben naturalmente una reminiscencia de *orden* (Beltrami, 1999).

Como humanidad hemos vivido la dificultad de caracterizar o definir la aleatoriedad. No es sorpresivo, entonces, que nuestros estudiantes interpreten la irregularidad de las secuencias aleatorias como *ausencia de modelos*, puesto que esta acepción se ha utilizado para definir lo que sería una secuencia aleatoria. Así por ejemplo, los estudiantes de nuestro estudio que citamos abajo, usan el término «azar» como sinónimo de «sin pensar» y de «desigual» para describir un mayor desorden de los datos en la gráfica de distribución de las secuencias reales:

> En el número de rachas tanto en las simuladas como las reales los datos son muy parecidos porque en las simuladas no tuvimos en cuenta las rachas y las pusimos sin pensar, casi al azar [*Alumno IE*].

> …se observa una diferencia entre los resultados simulados y los reales, ya que al inventarlos nosotros, las simuladas son más iguales y los reales son más desiguales al haberlos obtenido al azar [*Alumno IEA*].

Estos resultados coinciden con los trabajos de Serrano (1996) y Serrano, Batanero y Cañizares (1999), quienes encontraron que una de las razones más frecuentes de los estudiantes para apoyar la no aleatoriedad de una serie de puntos distribuidos en un plano fue la regularidad de su patrón. Igualmente cuando a los estudiantes se les pidió generar distribuciones aleatorias, la mayoría de ellos buscaron erróneamente la regularidad, especialmente en lo concerniente a no dejar espacios vacíos. En Serrano, Batanero, Ortiz y Cañizares (2001) en la tarea de identificación de secuencias aleatorias binarias, también se asoció el patrón regular con la no aleatoriedad en una alta proporción de los estudiantes, aunque no fue el más frecuente.

En nuestro estudio, estos otros alumnos al ver una secuencia aleatoria demasiado ordenada pensaron que era inventada porque no admiten *orden* en los procesos aleatorios.

> Observando estos datos yo diría que algunas de las secuencias que han dado los alumnos son falsas y algunas secuencias están muy ordenadas para haberlas hecho de forma aleatoria [*Alumno EA*].

> No puedes encontrar un patrón, pues la secuencia se realizó al azar [*Alumno BS*].

> No puede ser aleatorio, es demasiado ordenado [*Alumno LZ*].

Sin embargo al evaluar la aleatoriedad, siempre se compara con un modelo. Por ejemplo, en el estudio de Serrano, Batanero y Cañizares (1999) sería aplicable la distribución de Poisson para analizar el patrón de la distribución de cruces en el plano. Green (1982) reporta 5 modelos diferentes relacionados con la aleatoriedad de una secuencia aleatoria de lanzamientos de monedas, entre ellos, a través de la variable «número de caras en *n* lanzamientos», con una distribución Binomial. Falk (1981) trabaja con la variable aleatoria «razón de alternancia de las secuencias», que se comporta con una distribución Normal. Green y Hunt (1992) trabajan con el «número de rachas» que relaciona con la *recencia* como una forma de medir el grado de alternancia de rachas. En la Sección 4.2, describimos las 4 variables aleatorias y sus modelos que usaremos en la presente investigación.

7.1.2. Evaluación global de las concepciones sobre aleatoriedad

Mientras el análisis anterior (Sección 7.1.1) era de corte cualitativo, puesto que sólo se enfocó a los estudiantes que explicitaron concepciones en su reporte. En esta sección el análisis será de corte cuantitativo y podrá caracterizar globalmente las concepciones de todo el grupo de participantes en nuestro estudio. Nos basaremos en el análisis de los datos proporcionados cuando se les solicitó que inventaran una secuencia que simulara los lanzamientos de una moneda para describir sus tendencias generales y variabilidad. Este procedimiento ha sido seguido en diversos estudios sobre percepción de la aleatoriedad, más concretamente en Green (1982, 1989) y Serrano (Serrano, 1996; Serrano, Batanero, Ortiz y Cañizares, 2001).

A diferencia del análisis realizado en la Sección 4.3, en este espacio incluiremos a todo el grupo de estudio, formado por 3 clases (111 estudiantes). Además, no sólo nos basaremos en la comparación entre variables estadísticas sino que también haremos uso de los modelos de las variables aleatorias, puesto que conocemos la distribución teórica de las variables, cosa que no podíamos esperar de los alumnos del estudio. El análisis se organizará de acuerdo a los tres estimadores con los que caracterizamos las concepciones globales del grupo: el valor esperado y la dispersión de cada una de las variables y la independencia entre los ensayos en cada secuencia.

Percepción del valor esperado

En primer lugar analizamos si hay una percepción intuitiva del valor esperado de cada una de las variables analizadas, comparando los valores medios obtenidos en las secuencias reales y simuladas. En la Tabla 6.17 se muestran las medias y sus intervalos de confianza con un 95% de confianza para cada una de las variables del proyecto en el total de los alumnos. Las características globales de las concepciones del grupo de estudiantes que se pueden inferir de estos datos sobre el valor esperado no se diferencian mucho de las descritas en otras investigaciones sobre percepción subjetiva de la aleatoriedad.

Tabla 6.17. Medias e intervalo de confianza de las variables consideradas en el análisis de las secuencias simuladas y reales (n=111)

	Media	Intervalo de confianza al 95% de la media
Número de caras en las secuencias simuladas	10,29	[10,18;10,73]
Número de caras en las secuencias reales	10,45	[10,12;10,45]
Número de rachas en las secuencias simuladas	10,7814	[10,40;11,16]
Número de rachas en las secuencias reales	10,1	[9,70;10,49]
Racha mayor en las secuencias simuladas	3,32	[3,17;3,47]
Racha mayor en las secuencias reales	4,35	[4,14;4,57]

Al realizar el contraste de hipótesis para igualdad de medias entre las secuencias reales y simuladas con la variable número de caras, obtenemos un valor de: t = -1,0008, con un p-valor = 0,31. Por tanto, el resultado no es estadísticamente significativo y no podemos rechazar la hipótesis nula de que las medias de las dos distribuciones sean iguales. En los intervalos de confianza (Tabla 6.17) observamos que los intervalos para las secuencias reales y simuladas se traslapan. Así mismo, el intervalo de confianza para la diferencia entre las medias contiene el valor 0,0, puesto que se extiende desde -0,16 hasta 0,48. Estos resultados confirman que no existe diferencia estadísticamente significativa entre las medias de las dos muestras.

Interpretamos que los estudiantes son muy exactos al reflejar el número esperado de caras en la secuencia de 20 lanzamientos de la distribución Binomial B(20, 0,5), incluso la moda de la distribución obtenida es igual a la del valor esperado de la Binomial np=10 y la frecuencia del valor 10 supera en mucho a la de otros valores, hasta a los adyacentes a él. Esto se observa claramente en el gráfico de caja (Figura 6.7), donde la mediana coincide con el primer cuartil debido al alto número de estudiantes que obtienen exactamente 10 caras en la secuencia inventada. Estos resultados coinciden con los de Serrano (1996), quien trabajó con una secuencia de 50 lanzamientos y en donde el número de caras simuladas (25,85), resultó ser muy próxima al valor teórico esperado (25).

Con respecto a la variable número de rachas, el test t para la igualdad de medias entre las secuencias simuladas y reales, mostró un valor t = 2,48 y un p-valor = 0,013. Ya que el p-valor es inferior a 0,05, el resultado es estadísticamente significativo y rechazamos la hipótesis nula, que suponía la igualdad de medias para las dos secuencias, a favor de la hipótesis alternativa. El resultado se confirma con el intervalo de confianza a un nivel de confianza del 95% para la diferencia de las medias entre el número de rachas de la secuencia simulada y de la secuencia real, que fue [0,14; 1,23] y no contiene al cero.

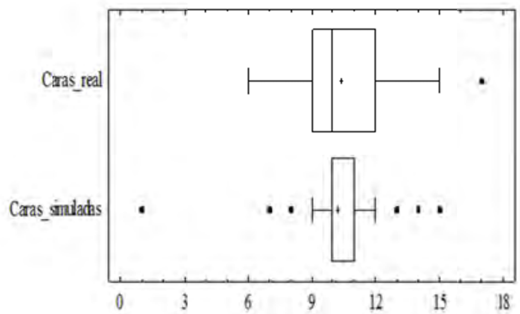


Figura 6.7. Número de caras en las secuencias reales y simuladas de los estudiantes

En consecuencia, los estudiantes no tienen tan buena intuición para los valores medios del número de rachas. En la Figura 6.8 se observa que los estudiantes tienden a generar muchas rachas, puesto que en la gráfica de caja las secuencias simuladas están sesgadas a la izquierda, además el número medio de rachas en la secuencia simulada es mayor que en la secuencia real (Tabla 6.17).

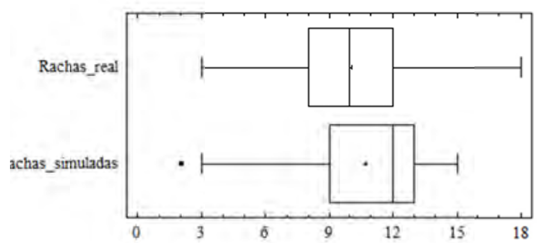


Figura 6.8. Número de rachas en las secuencias reales y simuladas de los estudiantes

En concordancia con los resultados obtenidos en otras investigaciones (Green, 1989; Green y Hunt, 1992; Serrano, 1996), los estudiantes presentan recencia negativa, es decir, una tendencia a alternar en exceso los resultados del lanzamiento de la moneda. No obstante, nuestros resultados son menos marcados que los de tales investigaciones posiblemente debido a que el tamaño de la muestra que ellos trabajaron era más grande que la nuestra (50). Por ejemplo, Serrano obtiene un promedio de número de rachas de 30,015 y 29,238 para estudiantes de 18 y 14 años respectivamente, que difieren ampliamente del valor teórico de una secuencia de 50 lanzamientos (25,5 rachas). La distribución teórica del número de rachas para una secuencia de 20 lanzamientos es N(10,5, 2,18) (Engel y Sedlmeier, 2005), de modo que la media teórica y el promedio obtenido por nuestros estudiantes (10,78) no difieren tan visiblemente. Los valores teóricos también se pueden aproximar a través de la simulación en un ordenador de 100.000 secuencias de 20 monedas, en donde la media obtenida fue de 10,4999 y la desviación típica de 2,186 (Figura 6.9).

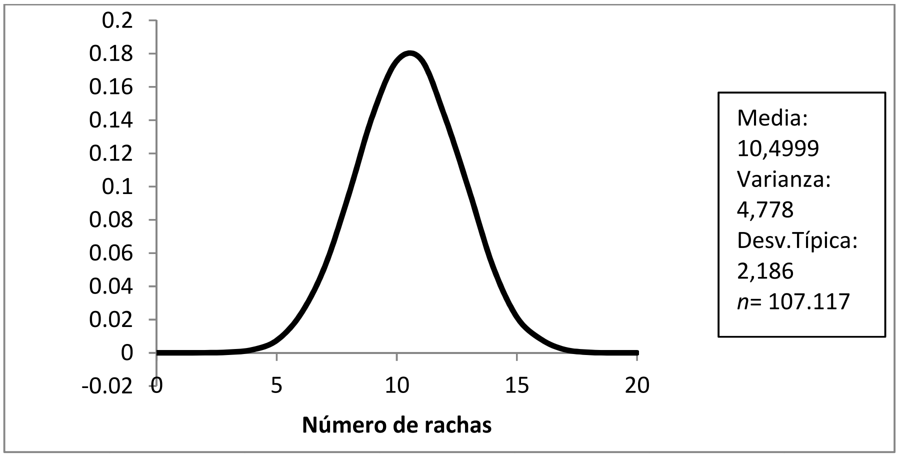


Figura 6.9. Distribución del número de rachas obtenida a partir de la simulación de 107.117 secuencias de 20 lanzamientos de una moneda

Esta diferencia fue observada por algunos de los estudiantes, quienes al analizar los datos de su clase, fueron conscientes de que alternaron excesivamente los resultados.

Finalmente, al estudiar la racha más larga en las secuencias inventadas por los alumnos (Figura 6.10) observamos que produjeron rachas excesivamente pequeñas (en forma excepcional 4 y 5, pero la moda estuvo en 3). La media de la racha más larga producida es solo 3,32 frente a la media en las secuencias reales (4,35). Eso mismo lo podemos observar en el gráfico de caja, en donde se nota que la distribución de las secuencias inventadas por los alumnos está totalmente sesgada hacia la derecha, más allá de lo que está la distribución del número de rachas de los lanzamientos reales de la moneda.

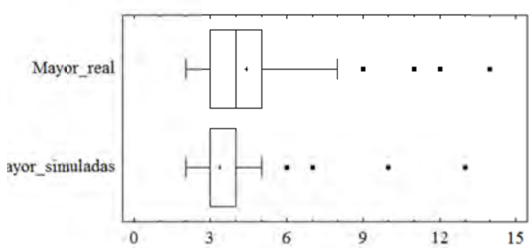


Figura 6.10. Racha más larga en las secuencias reales y simuladas de los estudiantes

Como en las variables anteriores, realizamos un contrate de hipótesis *t* para la diferencia de medias de las variables longitud de la racha más larga en las dos secuencias, simulada y real, en el que se han obtenido los valores $t = -7{,}761$ y un *p*-valor = 0,000, con lo que, puesto que el *p*-valor es menor que 0,05, el resultado es estadísticamente significativo y rechazamos la hipótesis nula que supone la igualdad de medias frente a la hipótesis alternativa. Además, en la Tabla 6.17 se aprecia que los intervalos de confianza para las medias en ambas secuencias son totalmente disjuntos, por lo que hay una notable diferencia entre los valores medios en ambas secuencias. Por otro lado el intervalo de confianza al 95% para la diferencia de medias, [-1,29; -0,77], no contiene el valor 0,0.

El efectuar la simulación de un gran número de experimentos, obtuvimos valores aproximados de los parámetros de la distribución de la variable longitud de la racha más larga (Figura 6.11). Su media es de 4,67, que es aún mayor que la obtenida en la muestra del lanzamiento de la moneda obtenida por los estudiantes, con lo cual, las distribuciones simuladas por los estudiantes se alejan aún más de lo esperado en una secuencia aleatoria.

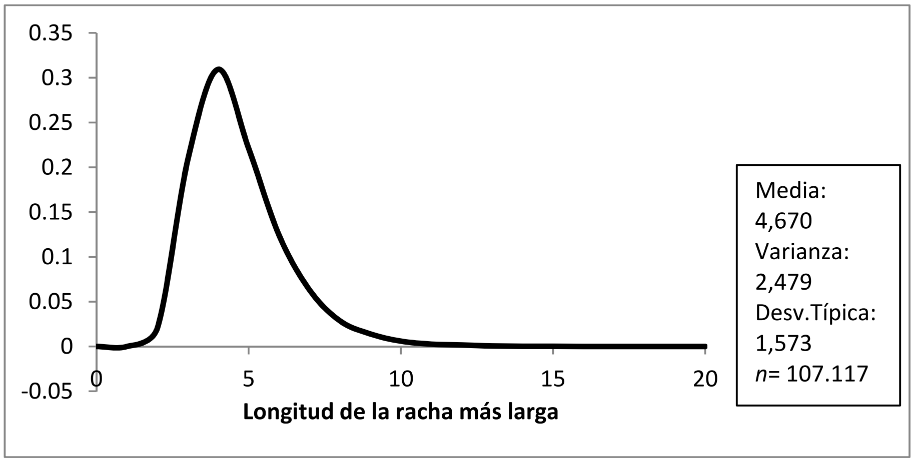


Figura 6.11. Distribución teórica de la longitud de la racha más larga obtenida a partir de la simulación de 107.117 secuencias de 20 lanzamientos de una moneda

Estos resultados indican un fallo en la percepción de la independencia de ensayos en una secuencia aleatoria por parte de los estudiantes lo que concuerda con los trabajos de Green y Serrano (Green y Hunt, 1992; Serrano 1996). Este hecho también fue destacado por algunos estudiantes que dieron argumentaciones similares a la que sigue:

> Veremos como en las secuencias reales las rachas alcanzan tamaños mayores que en las secuencias simuladas, lo que obedece a la falsa creencia de que en los fenómenos aleatorios es bastante improbable obtener rachas grandes [*Alumno EE*].
>
> …es más bien difícil que haya una racha muy grande al azar [*Alumno UE*]

Percepción de la dispersión

En la Tabla 6.18 se muestran tanto la desviación típica, como el intervalo de confianza del 95% de las desviaciones típicas para de una de las variables en estudio en el proyecto. En los párrafos sucesivos interpretaremos los datos de estos resultados para cada una de las variables de interés.

Tabla 6.18. Desviación típica e intervalo de confianza de las variables consideradas en el análisis de las secuencias simuladas y reales (n=111)

	Desv. Típ.	Intervalo de confianza al 95% de la Desv. Típ.
Número de caras en las secuencias simuladas	1,25	[1,11;1,34]
Número de caras en las secuencias reales	2,05	[1,87;2,26]
Número de rachas en las secuencias simuladas	2,85	[2,56;3,10]
Número de rachas en las secuencias reales	2,95	[2,65;3,21]
Racha mayor en las secuencias simuladas	1,12	[1,02;1,23]
Racha mayor en las secuencias reales	1,65	[1,46;1,77]

El contraste F para comparar la igualdad de varianzas en el número de caras da un valor F = 2,82, lo que señala que la variabilidad entre grupos es casi tres veces la variabilidad dentro de cada grupo. El p-valor=0,0, así que el resultado es estadísticamente significativo con lo que se rechaza la hipótesis de igualdad de varianzas. Observamos también que los intervalos de confianza de la desviación típica del número de caras en las secuencias simuladas y reales (Tabla 6.18) no se solapan. La conclusión se corrobora con el intervalo de confianza para la razón de varianzas, que se extiende desde 2,16 hasta 3,69 sin incluir al 1. Así pues, deducimos que hay una diferencia estadísticamente significativa entre las dos varianzas. Esto también se aprecian visualmente en la Figura 6.7.

El grupo de estudiantes no refleja suficientemente bien la variabilidad en la distribución teórica esperada del número de caras (Binomial), pues el valor teórico esperado de la varianza es npq = 5, con una desviación típica de 2,236, mientras que la obtenida en las secuencias producidas por los estudiantes es sólo igual a 1,25. Esta baja desviación típica concuerda con el agrupamiento notorio de los datos en el gráfico de caja (Figura 6.7). Asimismo, es notorio que el rango de las secuencias reales es mayor y el recorrido intercuartílico de las secuencias reales es más amplio. Todos estos resultados nuevamente coinciden con los obtenidos por Serrano (1996), cuyos estudiantes de 18 años producen una desviación típica de 2,99, menor que la esperada teóricamente que sería de 3.54, sin embargo, nuestros resultados están notoriamente más distantes del esperado.

Para comparar las desviaciones típicas del número de rachas en la secuencia simulada y en la real, se ha realizado un contraste F de igualdad de varianzas. Se obtuvo una F = 0,93 y un p-valor = 0,609, menor que 0,05, por lo tanto, no se obtiene evidencia estadística para rechazar la hipótesis de desviaciones típicas iguales en ambas secuencias. Además, en la Tabla 6.18 se observa que los intervalos de confianza del 95% para las desviaciones típicas se solapan y el intervalo [0,72; 1,22] de nivel 95%

para el cociente de varianzas contiene el 1, lo que muestra que no hay diferencias estadísticamente significativas entre las desviaciones típicas de las dos secuencias. De modo que aunque la percepción de los estudiantes acerca del valor medio del número de rachas en las secuencias simuladas fue pobre, la percepción de la variabilidad, ha sido buena.

En la Figura 6.8 podemos observar que, el rango de las secuencias simuladas y reales no es muy diferente. El valor de la desviación típica en la simulación de un gran número de ensayos (Figura 6.9) es 2,186, que no es muy distinto al obtenido en las secuencias simuladas por los estudiantes (2,85). Serrano (1996), en cambio, obtuvo una desviación típica mayor que la teórica (8,442 contra 3,4) en las secuencias simuladas por los estudiantes. Serrano no hizo prueba de hipótesis para las desviaciones típicas puesto que no comparó con una muestra de lanzamientos reales, sin embargo sus resultados fueron similares a los de Green y Hunt (1992) ya que las secuencias simuladas reflejan una mayor variabilidad que el valor teórico esperado.

Por último realizamos un contraste F de igualdad de varianzas para la longitud de la racha más larga en las secuencias simuladas y reales, obteniéndose una $F = 0{,}48$ y un p-valor = 1,74E-7, por lo que se rechazaría la hipótesis nula de igualdad de varianzas. Por otro lado, los intervalos de confianza de las varianzas al 95% de confianza (Tabla 6.18) no se solapan y el intervalo de confianza del 95% para el cociente de las varianzas no contiene el 1, puesto que se extiende desde 0,37 hasta 0,63. Se concluye, por lo tanto, que la diferencia entre las desviaciones típicas es estadísticamente significativa.

En la Figura 6.10 se puede observar estos resultados gráficamente en la mayor variabilidad de la longitud de la racha mayor en la secuencia real que en la simulada. En el gráfico de caja, vemos que el 50% de los datos se concentraron en esos dos valores. Más aún, como la mediana es 3, podemos decir que el 25% o más de los estudiantes tuvieron como racha mayor a 3. Lo cual explica la poca variabilidad de los datos. En comparación con el valor obtenido por la simulación de un número grande de ensayos (1,573), observamos que la dispersión obtenida por los estudiantes es pequeña (1,12). Todos estos resultados muestran la pobre percepción de la dispersión por parte de los alumnos para la longitud de la racha mayor en la secuencia simulada.

Esta conclusión concuerda con los valores obtenidos por Serrano, quién observó una concentración de datos alrededor de los valores de 4 y 5, pero un importante número de estudiantes no inventaron rachas mayores a 2 y 3 para una longitud de secuencias simuladas de 50.

Percepción de la independencia de ensayos

Un último punto de análisis es la percepción de la independencia de ensayos por parte de los estudiantes. Podemos evaluar la concepción del estudiante sobre la independencia de los lanzamientos sucesivos de la moneda a partir de la evaluación de la organización de las secuencias simuladas por ellos. Una secuencia cuyos resultados son independientes unos de otros tienen una distribución tanto del número de rachas como de la longitud de la racha más larga, conforme a la que aparece en un proceso aleatorio. Si el número de rachas es estadísticamente mayor o menor, ello indica que los sucesos no son independientes.

De acuerdo al análisis de la percepción del valor esperado y de la variabilidad de las variables número de rachas y longitud de la racha más grande (Tabla 6.17 y Tabla 6.18), la percepción de la independencia fue pobre en los futuros profesores ya que en promedio produjeron rachas de longitud más corta y mayor número de rachas de lo que se esperaría en un proceso aleatorio. Además, ideas erróneas sobre la independencia también fueron observadas en las conclusiones obtenidas por los estudiantes acerca de las intuiciones del conjunto de la clase sobre el azar. Estos alumnos argumentan que determinadas secuencias no pueden ser aleatorias debido a que algunas de sus rachas son demasiado largas:

> Algunos estudiantes hacen trampas e inventan sus secuencias reales ya que hay demasiadas caras sucesivas o cruces para ser aleatoria [*Alumna EA*]

De ello deducimos la creencia de los estudiantes de que en la aleatoriedad ocurre un cambio frecuente en el tipo de suceso subsiguiente. Pareciera ser que interpretan una racha larga como una manifestación de un *orden local* y por lo tanto, la rechazan como característica de aleatoriedad. La negación de un orden local en la aleatoriedad también implicaría que las rachas simuladas no fueran demasiado cortas (de tamaño 1) lo cual se observa en el número de rachas obtenidas por los estudiantes de nuestro estudio, puesto que el número mayor de rachas fue de 15 y nadie obtuvo 20 rachas. Aunque, indudablemente, esta hipótesis no se puede comprobar del todo con los análisis de nuestro estudio, sino que queda para analizar en un posible estudio posterior.

Estos resultados implican una manifestación de la *falacia del jugador* en algunos estudiantes, que también se explica mediante la teoría de heurísticas de Kahneman, Solvic y Tversky (1982), que sostiene que se juzga un proceso como más aleatorio que otro con base en la incorrecta creencia de que las pequeñas muestras deberían parecerse a la población de la que proceden. Por ello se juzga menos aleatorio un proceso que produce rachas largas: no representa una distribución equitativa de caras y cruces de la población.

Según Green y Hunt (1992) esta actitud es consistente, es decir, no importa que tan larga sea la secuencia, los estudiantes pretenderán «equilibrar» los resultados en pequeñas secuencias, alternando sus resultados frecuentemente y buscando la representatividad del número de caras y cruces esperado. Sin embargo, en su análisis, Serrano (1996) concluye que el estudiante pretende mantener la frecuencia esperada en el total de la secuencia, ya que en su análisis no encontró relación entre el número de caras de dos trozos de secuencia producida por un mismo alumno. Por otro lado, Chernoff (2009b) sugiere que los estudiantes hacen una partición subjetiva del espacio muestral, en la cual consideran sucesos compuestos a las longitudes de las rachas (y no los sucesos compuestos formados en la distribución binomial). Una longitud de racha corta es en realidad más probable que una larga (las rachas largas aparecen menos en la secuencia), por lo que ellos podrían concentrarse en procurar producir secuencias con rachas de la longitud más probable.

Lo indudable es que, de acuerdo con los resultados de estos autores y los nuestros, los estudiantes buscan representatividad en muestras más pequeñas que aquellas en las que se esperaría encontrar.

7.1.3. Discusión de los resultados obtenidos en el objeto aleatoriedad

De manera general, las concepciones de la aleatoriedad indican que nuestros resultados son coherentes con lo encontrado en otras investigaciones anteriores.

El análisis de las intuiciones globales sobre la aleatoriedad, indica que los alumnos se dejan llevar por sus creencias en el tema al tratar de simular lanzamientos aleatorios. Desconocen el valor esperado del número de rachas o de la longitud de la racha más larga, pero saben que el valor esperado del número de caras es 10 y es lo que condiciona su secuencia. Desconocen la variabilidad del número de caras, así que la ignoran, y sólo se concentran en su media. Sin embargo, también creen que no debe haber regularidad en la secuencia, así que tratan de alternar, lo más irregularmente que pueden, las caras y las cruces provocando un exceso de alternancia y condicionando su próxima cara o cruz a los resultados inmediatamente anteriores. Ese exceso de alternancia produce rachas demasiado cortas y resultados individuales dependientes. Las distribuciones de las variables vinculadas a las rachas quedan condicionadas por el valor esperado del número de caras y por la creencia de que la alternancia entre caras y cruces debe ser continua, pero irregular. De modo que, mientras ellos están preocupados porque *su* secuencia parezca aleatoria, en realidad están siguiendo el modelo matemático que les indica que debe haber el mismo número de caras que de cruces y que la longitud de las rachas no puede ser muy larga.

Los resultados de las concepciones individuales muestran que la comprensión del objeto aleatoriedad es complejo y sutil y posiblemente requiera de mucho tiempo para una total adquisición. En los argumentos de los estudiantes encontramos entremezcladas distintas concepciones que se relacionan con las aparecidas históricamente, pero que también siguen siendo vigentes en el contexto cotidiano. Muchas son correctas parcialmente, lo que denota una necesidad de complementarlas con otras propiedades para poseer una comprensión más completa de la aleatoriedad. Entre ellas, destacamos las siguientes:

- Observamos una tensión entre *impredecibilidad* local de cada suceso particular, y la *predecibilidad* de la distribución de cada variable aleatoria que aparece en el experimento. Algunos estudiantes se enfocan demasiado en la impredecibilidad local, por lo que no comprenden la utilidad de la estadística para estudiar matemáticamente la distribución global y dar predicciones de la misma. Esto también se observó en el análisis de las concepciones globales de los estudiantes puesto que rechazan la posibilidad de que en la aleatoriedad las rachas sean muy largas. Esta tensión están vinculada con una ausencia de modelos porque también niega la existencia de modelos vinculados a las variables aleatorias.
- Los alumnos alternan su visión de *ausencia o presencia de control* porque aún en la aleatoriedad existe la posibilidad de control sobre los resultados del experimento, pero al mismo tiempo, no puede llegar a ser total. Los estudiantes, en cambio, visualizan el control de manera absoluta: hay o no hay. Sería necesario comprender que la estadística proporciona herramientas mediante la que es posible explicar, reducir y por tanto controlar parte de la variabilidad (Wild y Pfannkuch, 1999). Nuevamente esta visión de los estudiantes está relacionada con la aleatoriedad como algo en lo que no hay modelos.
- En *la relación establecida entre azar y causalidad* hay una contradicción, puesto que al aceptar el azar como causa, aceptan que «no hay azar», sólo causas, o bien, lo aceptan como «algo ininteligible» a lo que se puede recurrir cuando no se conocen las causas. En nuestra investigación hemos soslayado discutir directamente con los estudiantes esta problemática filosófica, aunque un profesor bien informado, podría conocerla y llegar a tratarla en su salón de clases con el fin de relacionar la aleatoriedad (aunque le llame azar) con modelos matemáticos.
- Actualmente se *diferencia entre proceso y resultado aleatorio* como dos

aspectos de la aleatoriedad. En nuestro estudio nadie cuestionó la aleatoriedad en el proceso de lanzar una moneda. Al contrario, de manera muy natural, ligaron los dos aspectos al usar los resultados del proceso que sabían aleatorio (los lanzamientos de las monedas) para evaluar resultados que sabían que no habían sido generados por un proceso aleatorio (su intuición de la aleatoriedad). Sin embargo, en ocasiones se presentó una confusión entre el «proceso» y «resultado», y también vincularon el proceso con la temporalidad.

- ❖ La acepción de aleatoriedad vinculada a la *equiprobabilidad* es un arma de doble filo. Es necesaria para el cálculo de probabilidades desde la perspectiva clásica, pero limita la percepción del estudiante en la asignación de probabilidad, olvidándose de aquellas en las cuales no es aplicable el principio de inferencia o donde no hay simetría física. Además, favorece la confusión entre los sucesos de un experimento aleatorio equiprobable con los valores que adopta la variable aleatoria.
- ❖ La *variabilidad* actualmente es vista como una característica fundamental de los fenómenos aleatorios. Fue apreciada parcialmente por los estudiantes en sus argumentos. Sin embargo no fueron capaces de reproducirla en las variables que aparecen en el experimento ni, como vernos más adelante, tampoco de observar diferencias de variabilidad al comparar dos distribuciones. Para los estudiantes tuvo más sentido el valor esperado del número de caras.
- ❖ La idea de aleatoriedad como *ausencia de modelos* es muy fácil de aceptar por los estudiantes puesto que es acorde con la idea de azar como desorden y desconocimiento. Sin embargo, actualmente la aleatoriedad de un fenómeno es analizada a través de la modelación del fenómeno mediante el cálculo de probabilidades. Esta concepción de los estudiantes también se observó en el análisis de las concepciones globales, puesto que sólo percibieron parcialmente la presencia de un modelo vinculado con el número de caras.

Como se puede apreciar, casi todas estas percepciones responden a una idea cotidiana o histórica de la aleatoriedad y la mayoría de ellos son incompletas. Así por ejemplo, la aleatoriedad vinculada a la impredecibilidad es percibida como contraria a la modelación, pero la modelación probabilística (que es la que se maneja en los fenómenos aleatorios) no permite conocer ni predecir el fenómeno en el sentido determinista. El error, entonces, está en la no aceptación de un modelo matemático en un sentido probabilista. Así, hay contradicciones en la percepción de la aleatoriedad del estudiante que están vinculadas con un sentido determinista y absoluto del conocimiento. Una excepción a esta afirmación es la visualización del proceso y resultado aleatorio como dos aspectos distintos de la aleatoriedad que, posiblemente, es debida el actual uso de los números pseudoaleatorios, así como los modernos juegos electrónicos de azar. Creemos que la aleatoriedad vista como equiprobabilidad también es contraria a un pensamiento determinista, puesto que encierra una visión parcial de la asignación de probabilidades, sin embargo, Lecoutre (1992) indica que es debida a una falta de vinculación entre los fenómenos aleatorios y los modelos combinatorios.

Entonces, el estudio de los fenómenos aleatorios implica la aceptación de un tipo de razonamiento distinto al que comúnmente están acostumbrados los estudiantes. Es importante, por tanto, ayudar los, sobre todo a aquellos que serán profesores, a mejorar sus creencias sobre azar y aleatoriedad. La escuela o universidad debe constituirse en ese espacio privilegiado en donde se puedan discutir este tipo de cuestiones y con ello contribuir a distinguir entre las acepciones y creencias cotidianas no fundamentadas y aquellas con un sustento teórico y razonable.

Finalmente indicamos que es posible que estas creencias influyan en la forma en que los estudiantes analicen y usen la herramienta estadística para hacer inferencias informales y concluir la tarea expuesta en el proyecto (Makar, Bakker y Ben-Zvi, 2011).

7.2. Objeto de análisis: Variable estadística

En esta sección analizaremos el uso y concepciones de los estudiantes de diversos objetos que componen la variable estadística de acuerdo a lo expuesto en la Sección 5.2 de este capítulo. De acuerdo con Heitele (1975), analizaremos a la variable estadística a través de la formulación de la distribución y de las medidas de tendencia central y dispersión que usa el estudiante. Algunos resultados parciales de este análisis han sido publicados en Arteaga, Batanero y Ruiz (2009).

7.2.1. Distribución de la variable estadística

Un punto importante es si los estudiantes logran formular la distribución de frecuencias de las variables estadísticas de interés, es decir, si proporcionan un listado de todos los valores que la variable ha tomado en el conjunto de alumnos y para cada uno de ellos ha calculado su frecuencia, presentando la distribución de frecuencias, mediante una tabla o un gráfico. Para llegar a formar esta distribución, el estudiante tendrá que establecer una correspondencia entre cada valor de la variable y su frecuencia, con lo que usa la idea de función, variable independiente y dependiente. Al realizar la tabla o gráfico, tendrá que explicitar el rango de variación de dicha función así como su dominio. Este proceso es un paso hacia la conceptualización posterior de la distribución de probabilidad de las variables aleatorias correspondientes, como generalización de las variables estadísticas estudiadas, mediante un proceso de inferencia informal (Rossman, 2008).

La formulación de la distribución de la variable estadística de la forma descrita anteriormente, indicará que los estudiantes son capaces de abstraer la información relevante que les permitirá el análisis del fenómeno a través de los datos del grupo. Por lo tanto, considerará sus propios datos y los de cada uno de sus compañeros de grupo como una repetición del mismo experimento (y no como un fenómeno único), a la vez que acepta la variabilidad del fenómeno al obtener diferentes resultados de un mismo experimento y su posibilidad de análisis a través de las variables especificadas (Bakker y Gravemeijer, 2004). Por tanto, permitirá la observación de distintos grados de abstracción, la complejidad de los procedimientos, errores y deducir la proximidad del trabajo del estudiante con un razonamiento estadístico.

En este análisis tendremos que considerar forzosamente las variables estadísticas descritas en la Sección 4.2: número de caras obtenidas en cada una de las secuencias simulada (Y_s) y real (Y_r); número de rachas en cada una de las secuencias reales (T_r) y simuladas (T_s) y longitud de la racha más grande en cada una de las secuencias reales (L_r) y simuladas (L_s).

Tabla 6.19. Clasificación de estudiantes según la construcción de tabla y gráfico

Presentación de la distribución	No.
No presenta	7
Solo tabla	16
Solo gráfico	42
Tabla y gráfico	46
Total	111

En primer lugar, observamos que la mayoría de los estudiantes producen una representación de las distribuciones de las variables estadísticas en forma de tabla o gráfico y aproximadamente la mitad de ellos, hacen ambas (Tabla 6.19). Los estudiantes que realizaron tabla o gráficos para una de las variables repitieron el procedimiento para todas, casi sin excepción. Así que, aparentemente, los estudiantes comprenden y utilizan intuitivamente la idea de variable estadística y su distribución, así mismo, visualizan la necesidad de ese ella, pues en las instrucciones del proyecto no se les pedía explícitamente realizar tablas o gráficos.

Tablas de frecuencias de las distribuciones de la variable estadística

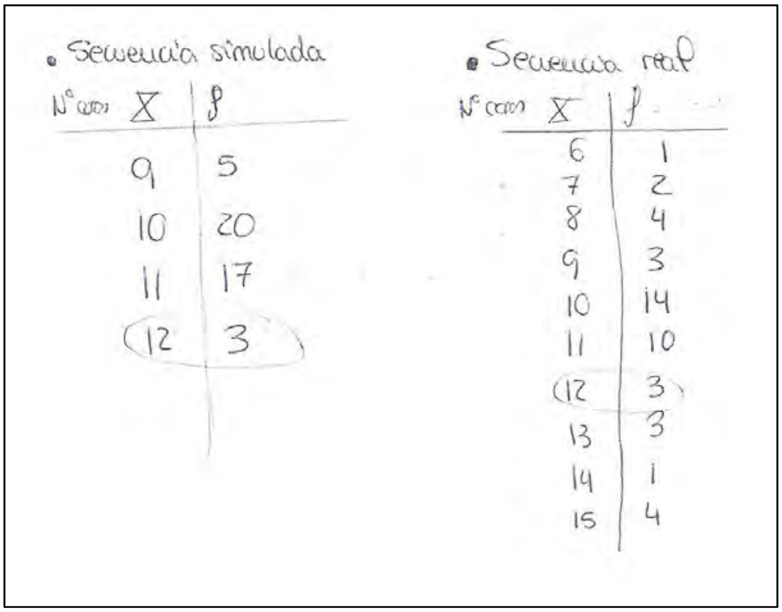
Un total de 62 estudiantes (55,8%) han elaborado tablas, bien conjuntamente o por separado de las variables mencionadas. Es importante notar que todos los estudiantes que hicieron tabla, la hicieron para las tres variables trabajadas en el estudio, aunque no todos hicieron uso de la información vertida en la tabla. En su mayoría y de manera general, las tablas son correctas, aunque básicamente se limitan a presentar sólo las frecuencias absolutas y no las frecuencias relativas.

A este respecto, Watson y Moritz (1999) y Watson (2001) sugieren que en un primer nivel de comprensión de la distribución, los estudiantes son capaces de comparar conjuntos de igual tamaño y sólo en un segundo nivel se comparan conjuntos de datos de diferente tamaño usando un razonamiento proporcional. También Batanero, Estepa y Godino (1997) encontraron que algunos estudiantes en su estudio usan frecuencias absolutas y no relativas al comparar distribuciones. Sin embargo en el proyecto planteado por nosotros, las muestras relacionadas tienen el mismo tamaño y la comparación de las distribuciones podría hacerse únicamente con base en las frecuencias absolutas, motivo por el que no es necesario recurrir a las relativas. Seguramente algunos alumnos se percataron de ello y por eso limitaron sus tablas a las frecuencias absolutas, aunque otros podrían no haberse preocupado por el tamaño de la muestra, quedándose en el primer nivel de comprensión de la distribución descrito en los trabajos de Watson. Nuestro estudio no permitirá distinguir entre estos dos grupos de estudiantes.

Es importante hacer notar, que a diferencia del problema planeado en el Estudio 3, en donde el estudiante se limita a interpretar la distribución de la variable, pues viene dada en el enunciado, en la tarea planeada en este estudio, el estudiante debe construirla, aspecto que sí podremos evaluar en este estudio.

Generalmente la interpretación de los datos se realizó a partir de los gráficos, o de los estadísticos calculados, sólo excepcionalmente los estudiantes comentaron directamente la tabla de frecuencias. Esto nos da indicios de que para ellos, la tabla es sólo un paso previo a la construcción del gráfico. Por otro lado, una cantidad importante de estudiantes (42) presentan sólo el gráfico, sin la tabla. Ellos tuvieron que hacer el conteo y relacionar el valor de la variable con la frecuencia, pero no vieron la necesidad de construir una tabla. También pudieron haber apuntado los datos por ahí, sin que para ellos fuera necesario añadir la tabla en su reporte. Puesto que no se les dio libertad también en el reporte, no solo en el tipo de análisis, la construcción de tablas no formó parte del contrato didáctico, y muchos no vieron la necesidad de incluirlas.

A continuación mostraremos los diferentes tipos de tablas que se presentaron encontraron más frecuentemente en las producciones de los estudiantes y que no siempre corresponden con las estándares de los libros de texto.



- Secuencia simulada

Nº caras X	f
9	5
10	20
11	17
12	3

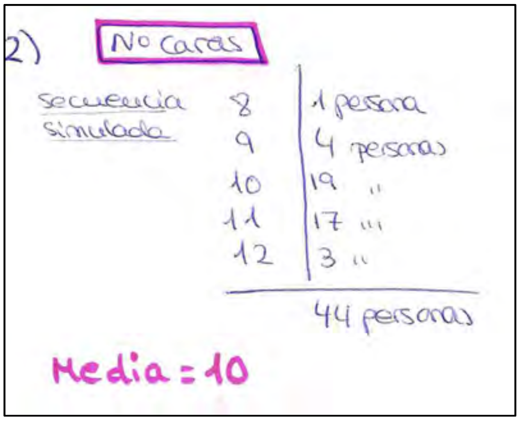
- Secuencia real

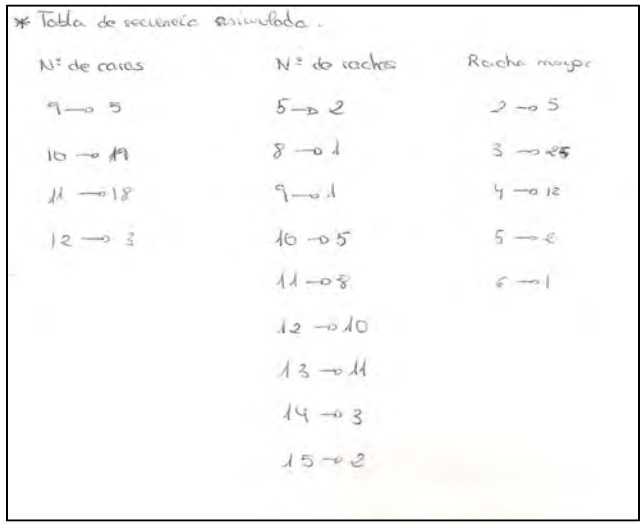
Nº caras X	f
6	1
7	2
8	4
9	3
10	14
11	10
12	3
13	3
14	1
15	4

Figura 6.12. Ejemplo de tablas de frecuencia separadas con una notación formal

La tabla más frecuentemente presentada se muestra en Figura 6.12, donde el estudiante sólo calcula las frecuencias absolutas y presenta dos tablas separadas para las secuencias reales y simuladas. Para hacerla, en primer lugar, el estudiante comprendió la diferencia entre las dos muestras para la variable estadística representada (en este caso, número de caras). Después clasificó los datos según el valor de la variable y efectuó el recuento. Usa las ideas de variable, valor de la variable y frecuencia. Además usó símbolos adecuados. Lo que muestra una aparente comprensión de la idea de distribución puesto que maneja las dos variables, la independiente (*X:* número de caras) y la dependiente (*f:* frecuencia de cada valor) y lo está haciendo para el conjunto de los datos y no individuo a individuo.

En este caso el estudiante ha usado la notación simbólica *X* para referirse a la variable estadística y *f* para referirse a las frecuencias. Aunque no lo explicitó, es posible que haya usado esta notación por su semejanza con la notación de funciones; por otro lado, los libros de texto suelen designar la frecuencia absoluta con la letra *f*. Cualquiera de estas dos opciones denota una abstracción del estudiante, aunque la primera sería mucho más abstracta. También el uso de una *X* mayúscula es notorio porque está más vinculado con la notación de la variable aleatoria en probabilidad y estadística, aunque el estudiante también tiene la necesidad de recurrir al contexto al escribir, de manera abreviada y muy pequeño, que se trata del número de caras.





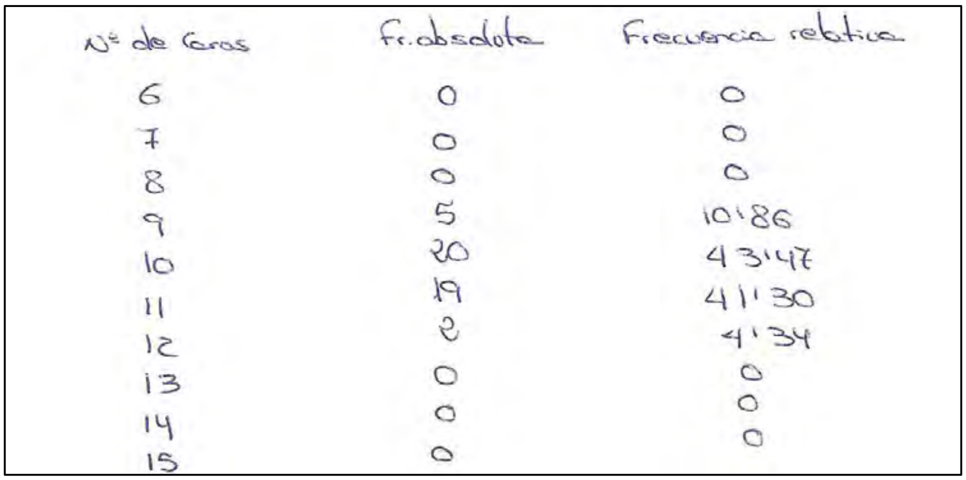
A) B)

Figura 6.13. Ejemplo de tablas separadas con distinta notación

En otros estudiantes con tablas separadas (ver ejemplo en Figura 6.13A) no se observa esta formalización y el estudiante se refiere a «número de caras» y «número de personas», o, en otros ejemplos, se denomina a las frecuencias absolutas «veces repetidas», sin usar ni los símbolos matemáticos ni explicitar los conceptos de variable y frecuencia. Esto hace más difícil valorar si el estudiante usa o no la idea de función, pues no se deduce claramente de la notación. Hay una predilección por el uso del contexto del problema lo que puede ser generado por una falta de abstracción de la situación, como para usar la palabra 'frecuencia'.

En la Figura 6.13B se observa el reporte en donde simplemente no se pone etiqueta a las columnas de la tabla, pero usa una notación de aplicación en donde, aparentemente, se establece una relación funcional y se trasluce la distinción entre la variable dependiente y la independiente. En esta figura también se observa que el alumno ha sido capaz de formar la distribución de las tres variables en estudio para la secuencia simulada.

En la Figura 6.14 mostramos otra tabla de frecuencias en donde el estudiante ha tenido en cuenta los valores de la variable obtenidos en las dos secuencias real y simulada. Es por ello que algunos valores de la variable en la tabla aparecen con frecuencia cero (son los valores de la variable obtenidos en la secuencia real que no aparecen en la simulada). Aunque esta estrategia es ventajosa en la comparación de las dos secuencias, posteriormente surge el problema de que este estudiante calcula mal el rango de las variables, pues considera como mínimo el valor 6 y como máximo el 15 en ambas secuencias, puesto que en las tablas, ambas secuencias aparentan el mismo rango. Es decir el alumno no comprende bien el concepto de rango de la variable. Este mismo problema se detectó en otros estudiantes; incluso algunos construyeron la tabla de frecuencias y luego los gráficos desde el valor 0, aunque el mínimo que aparece en las dos secuencias es de 6. Esto fue particularmente grave porque el rango fue la medida que más se usó para evaluar la variación en la distribución de las variables, así que esto no permitió ver la diferencia de variación en la distribución de las variables en las dos secuencias.

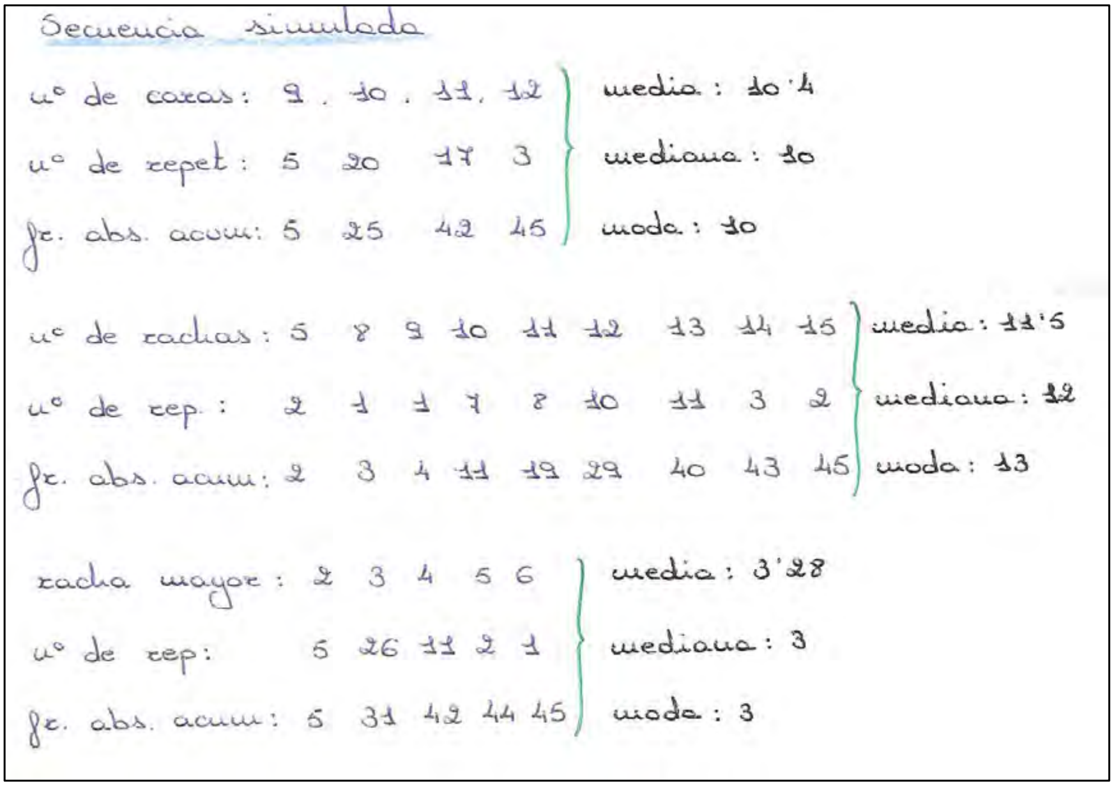


Nº de Giros	Fr. absoluta	Frecuencia relativa
6	0	0
7	0	0
8	0	0
9	5	10'86
10	20	43'47
11	19	41'30
12	2	4'34
13	0	0
14	0	0
15	0	0

Figura 6.14. Ejemplo de tablas de frecuencia con frecuencias nulas

Sin embargo, también se observa una referencia implícita al uso de la variable aleatoria (pues se tienen en cuenta los valores posibles y no sólo los resultados) y a la puesta en relación de variable estadística y aleatoria, es decir, la realización de una inferencia informal (Zieffler, Garfield, Delmas y Reading, 2008).

La asignación de la frecuencia absoluta y de la relativa a cada valor de la variable, implica una composición de funciones, donde la frecuencia absoluta (variable dependiente en la primera función) se convierte en independiente en el cálculo de la frecuencia relativa. Pero también donde a cada valor de la variable independiente le corresponden dos valores de la dependiente (la frecuencia y la frecuencia relativa), que, a su vez, se relacionan entre sí. El proceso implica una modelación estratificada en la que se requiere la descontextualización de los casos favorables (frecuencia absoluta) para poder abstraer la distribución de la frecuencia relativa de la variable. Es una situación muy parecida a la que describimos en el Estudio 3 para el caso del número de hijos (sección 8.2.3 del Capítulo 5), sólo que en aquel caso, los casos favorables eran la cardinalidad de los eventos compuestos puesto que se hace uso de la probabilidad clásica y directamente de la variable aleatoria.


Secuencia simulada

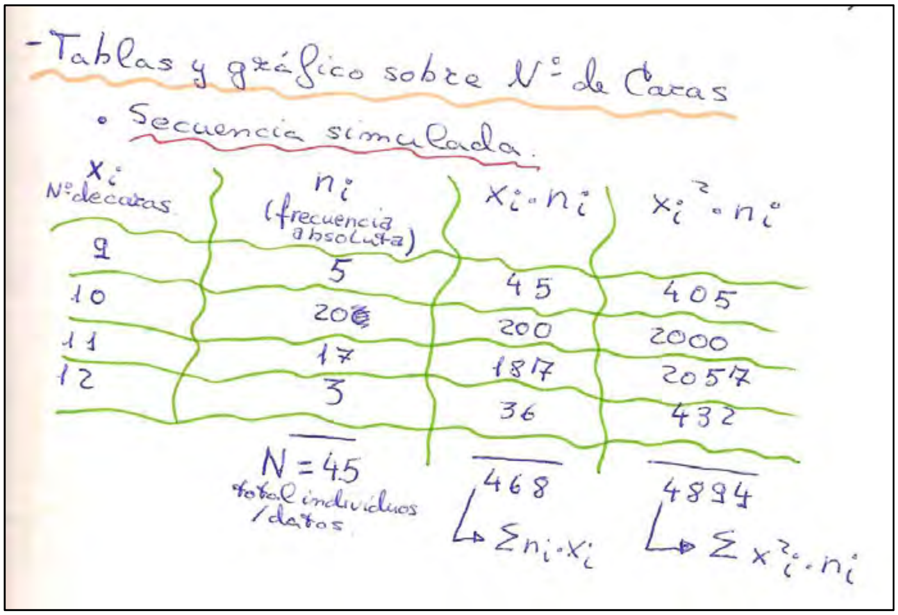
nº de caras:	9,	10,	11,	12						media: 10'4
nº de repet:	5	20	17	3						mediana: 10
fr. abs. acum:	5	25	42	45						moda: 10

nº de rachas:	5	8	9	10	11	12	13	14	15	media: 11'5
nº de rep:	2	1	1	7	8	10	11	3	2	mediana: 12
fr. abs. acum:	2	3	4	11	19	29	40	43	45	moda: 13

racha mayor:	2	3	4	5	6	media: 3'28
nº de rep:	5	26	11	2	1	mediana: 3
fr. abs. acum:	5	31	42	44	45	moda: 3

Figura 6.15. Ejemplo de tabla con frecuencias acumuladas

En el ejemplo de la Figura 6.15 el alumno calcula correctamente las frecuencias absolutas (que denomina simplemente repeticiones) y acumuladas, aunque luego no hace uso de dichas frecuencias acumuladas, sino simplemente representa los datos con diagramas de barras correctos de frecuencias absolutas. Nuevamente hay un efecto del contrato didáctico donde el estudiante asume que ha de hacer todos los cálculos estadísticos que pueda (o sepa), aunque no le sean necesarios para resolver el problema o no sepa realmente su utilidad. En la misma tabla el estudiante presenta la media, mediana y moda de cada variable, mostrando comprensión y capacidad de cálculo de estos estadísticos. Como en el caso anterior, aquí el estudiante también tiene que manejar una función compuesta para encontrar la frecuencia acumulada. De modo que además de manejar una función compuesta o bien con doble valor, el alumno realiza operaciones algebraicas con los datos para encontrar estos estadísticos.



- Tablas y gráfico sobre Nº de Caras

• Secuencia simulada.

x_i Nº de caras	n_i (frecuencia absoluta)	$x_i \cdot n_i$	$x_i^2 \cdot n_i$
9	5	45	405
10	20	200	2000
11	17	187	2057
12	3	36	432
	$N = 45$ total individuos / datos	468 $\to \Sigma n_i \cdot x_i$	4894 $\to \Sigma x_i^2 \cdot n_i$

Figura 6.16. Ejemplo de tabla con disposición para cálculo abreviado

En el ejemplo de la Figura 6.16, además de calcularse las frecuencias absolutas, se añaden columnas para el cálculo simplificado de la media y varianza. Aun cuando el profesor ha permitido que los estudiantes usen ordenador o calculadora, este estudiante ha preferido recurrir al medio aprendido de cálculo manual que, además de ser innecesario, seguramente realizó con auxilio de una calculadora. No obstante, estos cálculos nos hacen notar que el estudiante muestra un buen uso de los símbolos matemáticos y tiene un alto nivel de formalización de la variable estadística, diferenciando entre valores y frecuencias y expresándolos simbólicamente, así como sus operaciones. La disposición tabular le sirve también para facilitarle los cálculos; además de que los símbolos tienen una función instrumental y no sólo representacional. Por otro lado, la tabla también hace referencia a varias funciones diferentes, donde la variable independiente es el valor de la variable estadística, pero también se podrían considerar funciones compuestas, como en las tablas anteriores.

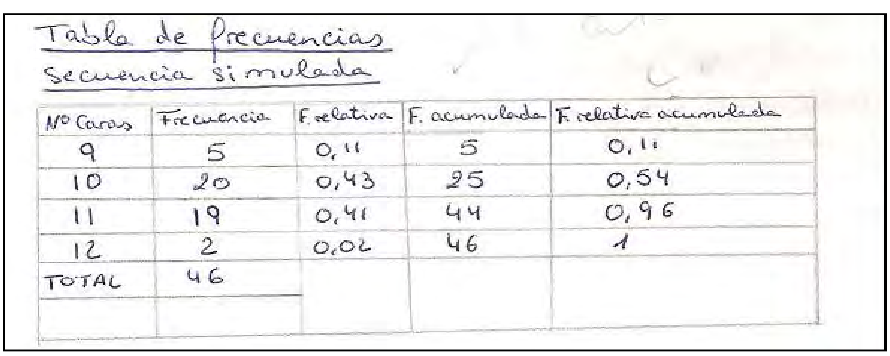


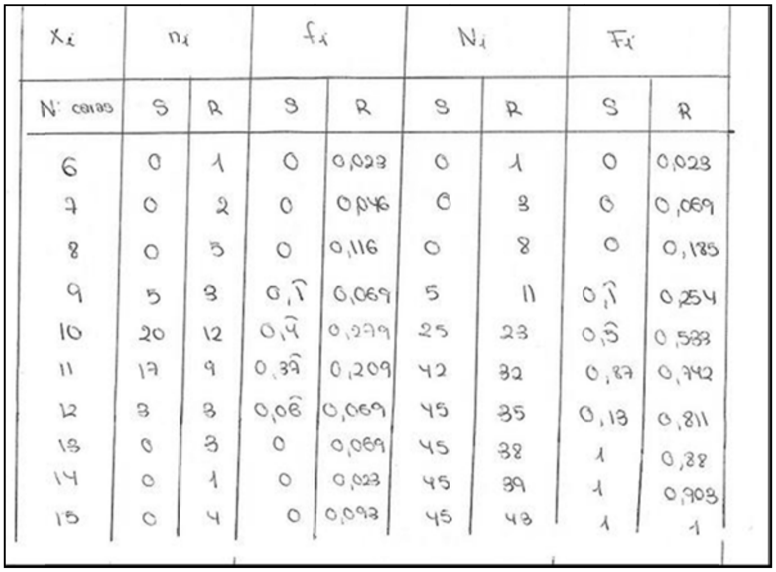
Tabla de frecuencias

Secuencia simulada

Nº Caras	Frecuencia	F. relativa	F. acumulada	F. relativa acumulada
9	5	0,11	5	0,11
10	20	0,43	25	0,54
11	19	0,41	44	0,96
12	2	0,02	46	1
TOTAL	46			

Figura 6.17. Ejemplo de tabla de frecuencias completa producida por un estudiante

En la Figura 6.17 presentamos un ejemplo de una tabla de frecuencias con las columnas de valor de la variable, frecuencia absoluta, relativa, acumulada y relativa acumulada, todas ellas calculadas correctamente. En algunos otros casos los estudiantes han añadido la columna de porcentaje. Se nota la complejidad de la tabla de frecuencias porque se representan cuatro funciones diferentes con la misma variable independiente (número de caras).



x_i	n_i		f_i		N_i		F_i	
N: caras	S	R	S	R	S	R	S	R
6	0	1	0	0,023	0	1	0	0,023
7	0	2	0	0,046	0	3	0	0,069
8	0	5	0	0,116	0	8	0	0,185
9	5	3	$0,\widehat{1}$	0,069	5	11	$0,\widehat{1}$	0,254
10	20	12	$0,\widehat{4}$	0,279	25	23	$0,\widehat{5}$	0,533
11	17	9	$0,3\widehat{7}$	0,209	42	32	0,87	0,742
12	3	3	$0,0\widehat{6}$	0,069	45	35	0,13	0,811
13	0	3	0	0,069	45	38	1	0,88
14	0	1	0	0,023	45	39	1	0,903
15	0	4	0	0,093	45	43	1	1

Figura 6.18. Tabla de frecuencias conjunta y completa

Se presentó un caso que implica más complejidad (Figura 6.18), en donde el estudiante muestra una tabla de frecuencias conjunta para ambas secuencias (*S*, simulada; *R*, real) y usa correctamente la notación simbólica, incluyendo las mayúsculas para las frecuencias acumuladas. Además el estudiante coloca las columnas de las secuencias simulada y real en paralelo lo cual facilita la comparación entre ambas variables. También se observa una comprensión y un uso correcto de la nomenclatura de los números decimales periódicos. Como en un caso anterior (Figura 6.14), el estudiante también incluye frecuencia cero para los valores de la variable en la secuencia simulada, puesto que, aunque en ella no se presenten casos, sí los hay para la secuencia real.

Cinco estudiantes presentan tablas similares a la de la Figura 6.19. Se observa que, aunque realizan una tabla conjunta, ésta no les facilita el análisis posterior, porque usan una disposición tabular no estándar. Aquí no se ve tan claramente que al mismo valor de la variable le corresponden dos valores de la frecuencia en las dos distribuciones, además que no se obtuvieron otros cálculos como la frecuencia relativa o la acumulada y por supuesto, tampoco se tiene la ventaja de establecer un paralelismo entre ambas secuencias.

Simulada					real									
8	9	10	11	12	7	8	9	10	11	12	13	14	15	16
3	6	14	16	2	2	6	8	7	4	7	2	3	2	1

Figura 6.19. Tabla de frecuencias con una disposición no estándar

Tampoco el uso del ordenador ayuda siempre al estudiante. En el siguiente ejemplo (Figura 6.20), el alumno introdujo los datos en una hoja Excel y calculó las frecuencias relativas (en realidad los porcentajes), pero las instrucciones dadas al ordenador fueron incorrectas. La frecuencia calculada es el resultado de dividir el valor de la variable entre la suma de todos los valores de la variable de la muestra (430) multiplicado por 100. Es decir confundió el valor de la variable con la frecuencia absoluta.

Es claro que el alumno no tenía correctamente definido el experimento aleatorio como el experimento compuesto de lanzar 20 monedas, ni tampoco supo relacionar la variable estadística con los resultados del experimento, puesto que no logró identificar cuáles son los valores que toma la variable estadística, ni cuáles son los casos favorables. Eso acarrea como consecuencia que no sepa cómo realizar el análisis de datos, ya que no agrupa los datos alrededor de los valores de la variable estadística. Al contrario, en su tabla prevalece el orden arbitrario en que fueron numerados los estudiantes, lo cual indica que no se dio cuenta que el número de alumno no implica otra cosa que la repetición del mismo experimento aleatorio.

Alumnos	Nº de caras	FREC. REL.(Fi)	Nº de rachas	Racha mayor	Nº de cruces	
1	10	2,320185615	10	2	10	3,5952381
2	10	2,320185615	11	4	10	3,5952381
3	11	2,552204176	13	3	9	3,5952381
4	9	2,088167053	10	3	11	3,5952381
5	8	1,856148492	13	3	12	3,5952381
6	11	2,552204176	12	4	9	3,5952381
7	10	2,320185615	8	3	10	3,5952381
8	11	2,552204176	10	3	9	3,5952381
9	10	2,320185615	12	3	10	3,5952381
10	10	2,320185615	13	3	10	3,5952381
11	11	2,552204176	14	3	9	3,5952381
12	11	2,552204176	11	4	9	3,5952381
13	11	2,552204176	12	4	9	3,5952381
14	12	2,784222738	12	4	8	3,5952381
15	8	1,856148492	10	2	12	3,5952381
16	13	3,016241299	14	5	7	3,5952381
17	10	2,320185615	12	3	10	3,5952381
18	9	2,088167053	11	4	11	3,5952381

Figura 6.20. Tabla de frecuencias en ordenador

Tabla 6.20. Clasificación de tablas, según datos representados

Categoría del gráfico construido	Frecuencia	Porcentaje
Listado de datos individuales	2	3,23
Tabla separada para cada variable	54	87,10
Tabla conjunta para las dos variables	6	9,68
Total	62	100

En conclusión, como vemos en la Tabla 6.20, 62 estudiantes (56% de la muestra) construye la tabla de las distribuciones de las variables de las secuencias, aunque, posteriormente, en su mayoría no las usan en el análisis de la comparación. Dos de estas tablas son simplemente listado de los datos sin organizar (es decir, no llegan a formar la distribución) y seis son tablas donde se representa conjuntamente las dos variables; el resto son tablas separadas para cada una de las variables a comparar.

Las tablas generalmente incluyen solamente las frecuencias absolutas. La mayor parte de los alumnos que hace tabla las elabora correctamente, aunque se presentan diferentes grados de complejidad en su elaboración, lo que también indica una comprensión de la función de la distribución de frecuencias en la resolución del proyecto y la definición de la variable estadística en ellos. También encontramos confusiones entre el valor de la variable y la frecuencia, frecuencia absoluta y relativa, y entre frecuencia relativa y porcentaje. Aunque algunos alumnos usan la hoja Excel, eso no les facilita la tarea, puesto que la hoja Excel no tabula ni agrupa los datos automáticamente y aunque así lo hiciera, se requiere una comprensión del objetivo del problema y del fenómeno aleatorio y su repetitividad para poder dar las instrucciones apropiadas al ordenador.

Representación gráfica de la distribución de la variable estadística

En este apartado hacemos un análisis resumido de las representaciones gráficas de las variables estadísticas producidas por los estudiantes en el proyecto. Un análisis más exhaustivo se encuentra en la tesis de Arteaga (2011), quien estudió los gráficos producidos por los estudiantes en el mismo proyecto que analizamos nosotros de acuerdo a su nivel de complejidad semiótica sustentándose en el análisis semiótico propuesto por Godino, Batanero y Font (2007). También relacionó esta complejidad semiótica con la corrección y el nivel de comprensión lectora de los gráficos y con las conclusiones obtenidas en el proyecto. Nosotros hemos colaborado con este autor en algunas publicaciones conjuntas (Arteaga, Batanero y Ruiz, 2009; 2010), ahora, en lugar de repetir su análisis, haremos una interpretación de los gráficos producidos en términos del nivel de comprensión de la variable estadística y su distribución, siguiendo las jerarquías definidas por diferentes autores.

En lo que sigue analizaremos muy brevemente estos niveles en los gráficos que produjeron los estudiantes. Mostramos algunos ejemplos de gráficos en cada uno de dichos niveles y destacamos el uso que se hace de las variables aleatoria y estadística en cada uno de los niveles. En algunos casos, el análisis de los gráficos también nos permitirá observar cómo los estudiantes reflejan sus concepciones de aleatoriedad en ellos y su relación con el manejo de las variables. Tendremos en cuenta los siguientes niveles:

0. No se producen gráficos.
1. Se producen gráficos que sólo representan los resultados del experimento individual del estudiante que lo elaboró.
2. Gráficos que representan o listan los datos de toda la clase, sin conformar la distribución de frecuencias.
3. Gráficos en que se forma la distribución de frecuencia separada para cada una de las variables que se comparan.
4. Gráficos conjunto de las dos distribuciones de las variables que se pueden comparar.

0. No se producen gráficos.

En nuestro estudio, el 20% de los estudiantes no realizaron ningún tipo de gráfico (Tabla 6.19). Su análisis se basó en tablas o medidas de tendencia central. Cuando lo realizan, lo mismo que Arteaga (2011), encontramos que en su mayoría los estudiantes se concentran en elaborar la gráfica de la variable «número de caras».

Observamos, entonces, que la mayoría de los estudiantes ha realizado un proceso de transnumeración (Wild y Pfannkuch, 1999) gráfico, aunque no todos ellos llegan a formar la distribución. Tampoco el que no hayan hecho gráficos significa que no tengan un cierto nivel de comprensión de la distribución. En nuestro estudio, cuatro de los estudiantes que no reportan gráfica calcularon medidas de tendencia central. Puesto que estas medidas son un resumen de la distribución de datos, asumimos que, aunque es posible que sea como efecto del contrato didáctico, los estudiantes están considerando la distribución de la variable bajo estudio, sin embargo no podemos saber que tan conscientes son de ello. Los otros tres estudiantes no usaron explícitamente este objeto matemático.

1. *Gráficos que sólo representan los resultados del experimento individual del estudiante que lo elaboró*

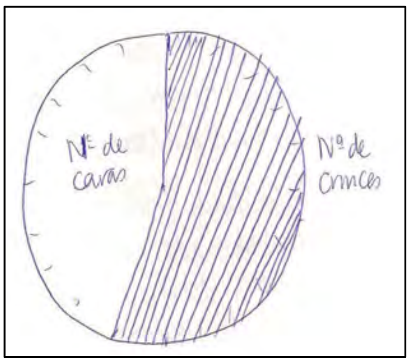
Dentro de este nivel se consideraron aquellos gráficos que produjeron los alumnos para representar únicamente sus propios datos, sin considerar los de sus compañeros. No se encontraron tablas equivalentes a este nivel posiblemente.

Los estudiantes clasificados en esta categoría no usaron explícitamente la idea de distribución. En general, trataron de resolver la pregunta para su caso particular (si él mismo tiene una buena intuición) comparando lo obtenido en las dos secuencias que él mismo había generado, por lo que deducimos que no comprendieron el propósito del proyecto de realizar un análisis estadístico, o bien, no fueron capaces de realizar un análisis global de los datos. En cualquier caso, no vislumbraron que el análisis estadístico está orientado a describir y predecir las propiedades de las distribuciones y no de datos aislados (Bakker y Gravemeijer, 2004) y por lo tanto no presentan un buen razonamiento estadístico (Wild y Pfannkuch, 1999). En los ejemplos presentados en la Figura 6.21, los gráficos representan las frecuencias de caras y cruces en 20 lanzamientos, usando diferentes tipos de gráficos.

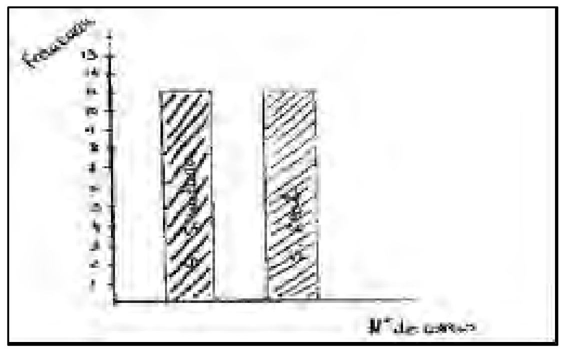
A través del manejo de su gráfico se deduce que estos estudiantes manifiestan una intuición errónea sobre la aleatoriedad porque suponen que una buena intuición implicaría que su secuencia simulada fue idéntica en alguna característica a su secuencia real. Es común que traten de ver si sus resultados particulares son iguales en las dos secuencias. De acuerdo con el análisis de la sección anterior (7.1), presentan la falacia de ilusión de control (Langer, 1975) o el enfoque en el resultado aislado (Falk, 1989).

También es evidente que este grupo de estudiantes no se percató de que no se trabajaba el experimento aleatorio simple (el lanzamiento de una moneda) sino sobre el compuesto por 20 lanzamientos de una moneda. Esto repercutió en que, de manera general, la variable «número de caras en 20 lanzamientos» no tuviera sentido para ellos. Sin embargo, esto no explica su falta de análisis global de los datos, puesto que aun cuando se considera solamente el experimento aleatorio simple, y por lo tanto la variable «cara o cruz en un lanzamiento», hubiera sido posible que tomaran en cuenta los datos de todo el grupo.

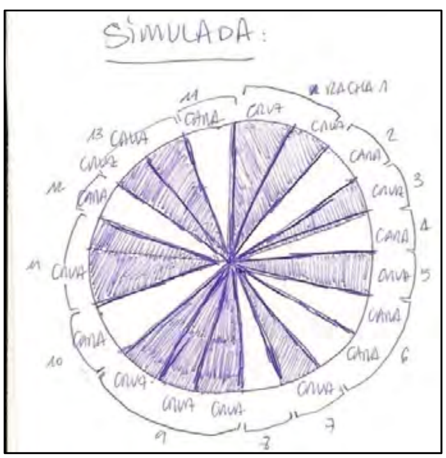
En el ejemplo A de la Figura 6.21, el alumno representa la frecuencia de los resultados de sus propias secuencias. Para él, los valores que toma la variable son «cara» y «cruz», lo cual manifiesta una idea intuitiva de variable aleatoria y estadística y su distribución. Sin embargo considera sólo la variable «resultado del lanzamiento de la moneda en sus 20 ensayos», es decir, considera la variable aleatoria Bernoulli ζ_r, que correspondería al lanzamiento de la moneda equilibrada, con probabilidades $p=q=1/2$ y que toma solo esos dos valores. También estudian la variable aleatoria ζ_s que toma valor 1 cuando ocurre, por ejemplo, cara (éxito) y 0 cuando ocurre cruz (fracaso) en la secuencia simulada, así como sus correspondientes variables estadísticas (tanto las variables aleatorias como las estadísticas se describieron en la Sección 4.2.3). Como no toma en cuenta los resultados de toda la clase, no puede considerar las variables Binomiales η_r y η_s, número de caras en las secuencias de cada alumno, que tomaría valores numéricos, ni sus correspondientes variables estadísticas definidas en las secciones 4.2.4, 4.2.5 y 4.2.6.



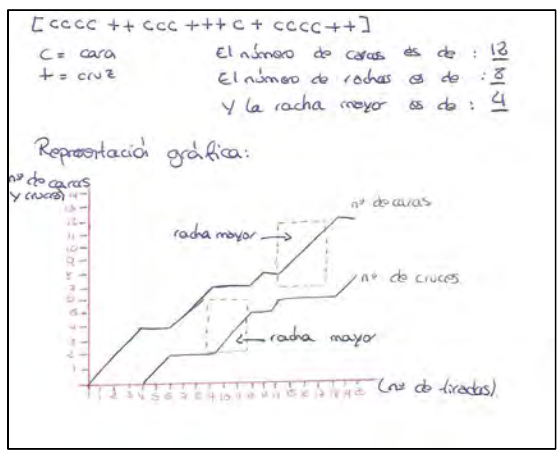
A)



B)



C)



D)

Figura 6.21. Ejemplos de gráficos que sólo representan sus datos

El ejemplo B de la Figura 6.21 el estudiante pretendió hacer un diagrama de barras, pero aunque anota que la variable independiente es el «número de caras», en realidad se refiere al «tipo de secuencia» y los valores que toma son: «real» y «simulada». Nuevamente se observa una confusión entre las variables «cara o cruz en un lanzamiento» y «número de caras en 20 lanzamientos de una moneda» puesto que el eje vertical lo rotula como «frecuencia», y lo es, sólo que de la primera variable estadística X, que, en un lanzamiento real, correspondería a una variable aleatoria Bernoulli, ζ_r. Este estudiante, sin embargo, sí representa los datos de ambas secuencias en una sola gráfica, lo que hace más fácil la comparación de las variables. Observemos que los valores de la frecuencia son los mismas (11), lo que, desde la concepción del estudiante (ilusión de control o enfoque en el resultado aislado), se interpretaría como una buena percepción de aleatoriedad.

Algunos alumnos representaron la sucesión de resultados en su experimento como mostramos en los ejemplos C y D de la Figura 6.21. En ellos, los estudiantes están haciendo una descripción de su propia experiencia al lanzar 20 veces una moneda, representando de diferente manera si el resultado era «cara» o «cruz» indicando, nuevamente, que ellos no lograron visualizar su experimento como uno compuesto, sino como repeticiones del experimento sencillo (lanzar una moneda). En el gráfico circular (inciso C) el alumno indica la obtención de «cara» con un sector circular en blanco, y la obtención de una «cruz» con un segmento coloreado. En este gráfico se visualizan rápidamente las rachas, así como la racha más grande, así mismo, se observa que el estudiante usó su gráfico para contar el número de rachas en su secuencia, aunque no hay evidencia de que haya hecho lo mismo con la racha más larga.

En el gráfico D, cada resultado (cara o cruz) es representado en una línea diferente, donde, en cada repetición, el estudiante acumuló la frecuencia del resultado del número de caras o cruces dependiendo de cuál fue el resultado. Observemos que la variable independiente es «número de tiradas» y que cuando la línea del número de caras crece, la línea del número de cruces permanece constante porque en una tirada se obtiene sólo uno de los dos resultados. El crecimiento de la línea indica la existencia de una racha, lo que permite representar, además del número de caras (o cruces), las rachas e incluso identificar la racha mayor. De este modo representa las tres variables de su secuencia en un solo gráfico, que además resulta ingenioso. Para su propósito, las variables están bien definidas y el gráfico es correcto dentro de este nivel de complejidad, aunque resulta inútil representar el número de caras y el número de cruces de una misma secuencia en su gráfico. Este ejemplo también está relacionado con la percepción de la aleatoriedad como un proceso temporal (7.1.1), puesto que a través del gráfico el estudiante está representando cómo se incrementa el número de caras (o cruces) a medida que transcurre el número de tiradas (tiempo).

A diferencia de los otros dos ejemplos, en estos (C y D) claramente se observa que los estudiantes hicieron uso de su gráfico para realizar su análisis, aunque no logren la comprensión que se esperaba en el proyecto.

En este nivel el estudiante no puede (ni tiene la intensión de) realizar una inferencia estadística informal basándose en sus gráficos, por lo tanto realmente no está vinculando una variable estadística con una aleatoria a través de esta herramienta. De acuerdo con el análisis semiótico realizado por Arteaga (2011) los conceptos, proposiciones y procedimientos puestos en juego por estos estudiantes son de un nivel de complejidad bajo, lo que potencialmente sólo permitiría un nivel de *lectura de datos* (Friel, Curcio y Bright, 2001).

2. *Gráficos que representan los datos de toda la clase, sin conformar la distribución de frecuencias*

En estos gráficos, los estudiantes no agruparon los valores similares de las variables (número de caras, número de rachas o longitud de la racha mayor) de las secuencias reales o simuladas de los alumnos de la clase, para formar la distribución de frecuencias. En lugar de ello, representaron el valor (o valores) obtenidos para cada alumno dentro del gráfico en el orden en que los datos fueron recopilados, que es un orden arbitrario. Es decir, sí toman en cuenta el total de los datos, pero no grafican la distribución de las variables.

En este tipo de gráficos hay un uso implícito de las variables estadísticas del problema, pero no llegan a formar su distribución porque no hacen el recuento de las frecuencias asociadas a cada valor de las variables y por tanto no emplean la idea de distribución de las variables en sus gráficos. En cambio, relacionan correctamente un elemento del eje de las abscisas (número de alumno) con el de las ordenadas (número de caras, por ejemplo) formando gráficos de barras o de líneas en donde sólo se visualiza la dispersión a través de medidas como el rango, el mínimo y el máximo. También se observa el orden en que se llevó a cabo el proceso de recolección de los resultados de cada estudiante, dándole una mayor importancia de la que tiene en la distribución, donde es irrelevante.

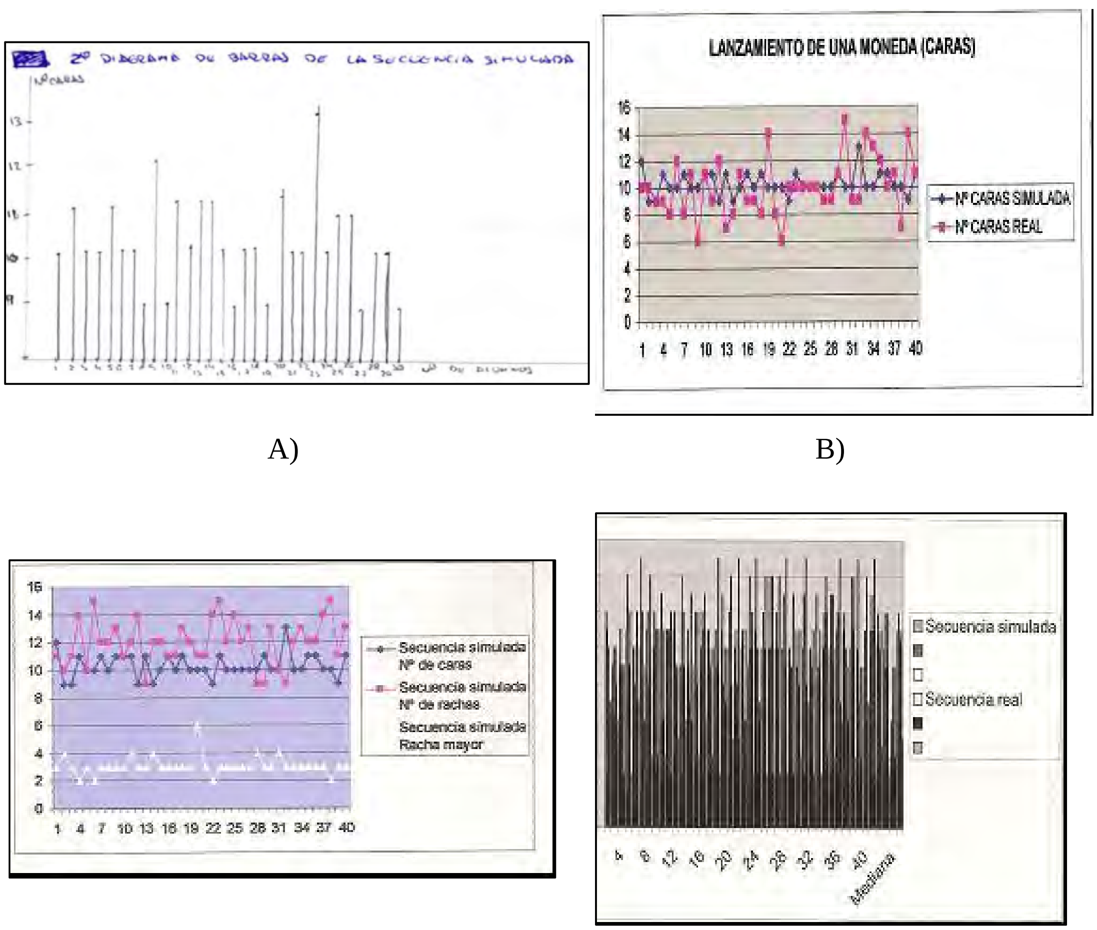
Este nivel de gráfico puede promover que el estudiante interprete los datos provenientes de una variable como repeticiones de un mismo experimento aleatorio (el lanzamiento de 20 monedas), es decir, como provenientes de una misma población, o bien puede favorecer la comparación de los datos uno a uno como si no fueran repeticiones de un mismo experimento, sino datos aislados. En el primer caso, aun cuando no formaron la distribución, los estudiantes podrían buscar una visión general del comportamiento de la variabilidad de los valores de la variable en la muestra, sin embargo sería limitada, puesto que no se propicia la comparación entre medidas de tendencia central. Aun así, el estudiante estaría obteniendo una caracterización parcial de una de las variables de interés en nuestro problema (número de caras, número de rachas o longitud de la racha más larga en 20 lanzamientos). En el segundo caso, las variables que se pondrían en juego, serían las relacionadas con el experimento simple X_r y X_s (4.2.3). Es decir, harían un análisis de todos los estudiantes del grupo, pero sin considerarlos elementos de una misma población. Esto favorecería que los estudiantes se centraran en observar el grado de coincidencia entre los datos provenientes de la secuencia simulada y de la real, favoreciendo el enfoque en el resultado aislado (Konold, 1989) o la ilusión de control (Langer, 1975). También podría fomentar la relación de la aleatoriedad con un enfoque temporal, puesto que el eje horizontal (número de alumno) se puede interpretar como el orden en el que se recopilaron los datos, cosa que es irrelevante en el proyecto.

En general, aunque en algunos casos este nivel de graficación es propicio para la comparación, no se favorecería una buena interpretación ni la realización de inferencias informales apropiadas. De acuerdo con Arteaga (2011), los gráficos de este nivel permitirían responder preguntas hasta el nivel de *leer entre datos* pero no de *leer más allá de los datos* en la terminología de Friel, Curcio y Bright (2001).

Entre los estudiantes que hicieron este tipo de gráfico, uno de ellos realizó la tabla de frecuencias para cada variable de manera correcta, por lo tanto sí conformó la distribución de frecuencias para las variables estadísticas, aunque no lo hizo gráficamente. En su mayoría, el resto de los estudiantes clasificados en este grupo estaría en un primer nivel de comprensión de la distribución, según Bakker y Gravemeijer (2004), donde la distribución se concibe como un conjunto de datos aislados. Algunos gráficos situados en esta categoría se ilustran en la Figura 6.22.

En el primero de estos ejemplos (A), el estudiante usa un diagrama de barras para representar el número de caras en la secuencia real; el diagrama permite visualizar la variabilidad de la variable, así como sus valores máximo y mínimo (por tanto el rango). Este estudiante hizo gráficas separadas para las secuencias real y simulada, sin embargo, la comparación entre ambas no resulta fácil puesto que, además de que en las gráficas no se observan claramente las medidas de tendencia central, usó distinta escala en el eje correspondiente al número de caras, por lo tanto tampoco se aprecia a simple vista la diferencia en la variabilidad de ambas secuencias.

El ejemplo B es un diagrama de líneas donde el alumno representa el número de caras en las secuencias real y simulada para cada estudiante en los mismos ejes. Este gráfico permite mostrar la coincidencia de valores de un mismo estudiante que puede reforzar las concepciones sobre aleatoriedad basadas en el enfoque en el resultado aislado (Konold, 1989) o la ilusión de control (Langer, 1975). Sin embargo, el gráfico sería correcto dentro de su nivel puesto que permite comparar la variabilidad de las variables estadísticas relacionadas con el número de caras (es decir las variables Y_r y Y_s, descritas en 4.2.4). Además, si el estudiante lo hubiese percibido, también podría haber representado la media de cada secuencia mediante una línea recta horizontal, lo que hubiera permitido una mejor comparación de las secuencias. Muestra una comprensión media del objetivo del problema al representar en un mismo eje ambas secuencias y ofrece una potencial comparación de la variabilidad por parte del estudiante. Aunque, por sí mismo, no indica claramente la variable que el estudiante utilizaría en su interpretación, puesto que puede favorecer que observe la variabilidad del número de caras en 20 lanzamientos en el total del grupo (variables Y_r y Y_s,) o las variables relacionadas con el experimento simple X_r y X_s (4.2.3) si se concentra en la coincidencia de valores entre la secuencia simulada y la real.



A) B)

C) D)

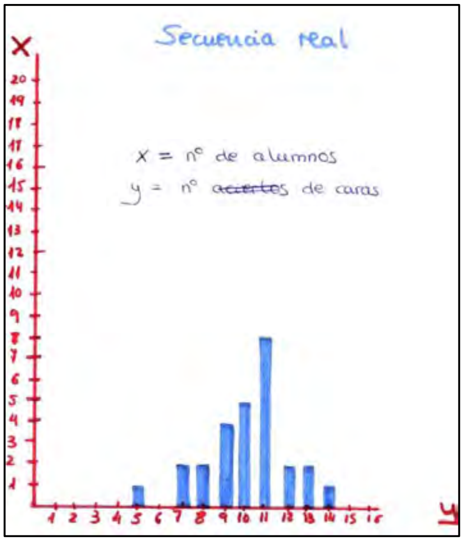
Figura 6.22. Ejemplos de gráficos que no representan la distribución de la variable

En el tercer ejemplo (C), se representan las tres variables (número de caras, número de rachas y longitud de la racha más larga) de una de las secuencias (secuencia simulada) en un mismo eje. Las variables, sin embargo, no son comparables, de modo que la gráfica no tiene sentido. Este mismo error se produce en el gráfico de barras adosado del ejemplo D, en donde, además, hay un mal uso del ordenador que se deduce de la inclusión de la mediana en el rango a graficar y del tipo de gráfico seleccionado en el que no se llega a visualizar ninguna de las características de las variables. Estos dos gráficos (C y D) muestran la incomprensión de estos estudiantes del objetivo del problema y la desvinculación de las variables con la evaluación de las concepciones de aleatoriedad de los estudiantes, puesto que a pesar de que colocan en un solo gráfico varias variables, parece que no les queda claro cuáles son las variables que tienen que comparar para realizar la evaluación. Estos gráficos (C y D) parecen indicar que sus autores están interesados en realizar un análisis por alumno, centrándose en la comparación de los valores de las variables que obtuvo cada uno de ellos (sin percatarse que no son comparables), o bien, no hay una intención de uso del gráfico.

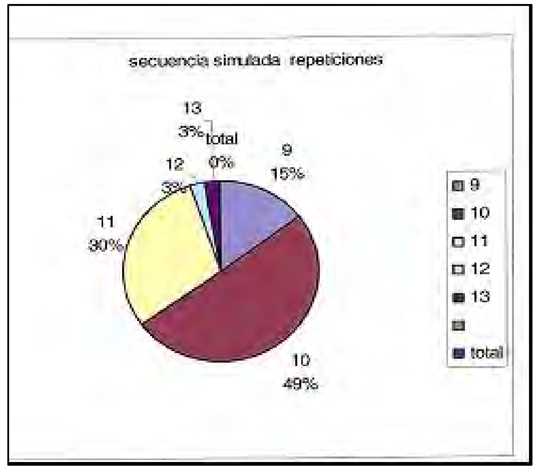
3. *Gráficos con una sola distribución*

En esta categoría los estudiantes calcularon las frecuencias de cada una de las variables y elaboraron dos o más gráficos en donde relacionaron cada una de las variables por separado. Este proceso no siempre supuso la elaboración previa de la tabla, pero a través de la gráfica se detecta que el alumno pasa del conjunto de datos a la variable estadística y su distribución de frecuencias y que potencialmente podría efectuar una inferencia estadística informal de manera correcta. La característica principal de los gráficos de este nivel es que los elaboran separados para cada una de las secuencias y a veces esto dificulta la comparación de las variables, sobre todo cuando no usan la misma escala de representación en los dos gráficos. Presentamos algunos ejemplos en la Figura 6.23.

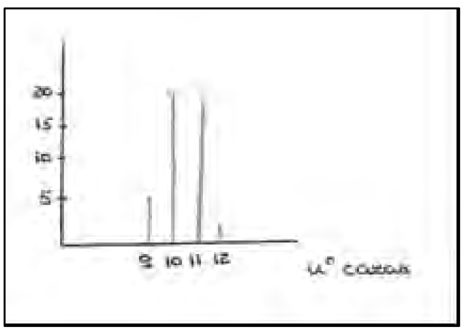
En el primer ejemplo (A), el estudiante produce un diagrama de barras correcto, tanto la escala, como las barras centradas en el valor entero. El estudiante ha tenido que agrupar los datos, calcular las frecuencias y representarlas. Muestra una comprensión correcta de la idea de distribución. Estarían en el nivel superior de comprensión de Bakker y Gravemeijer (2004), donde la distribución se concibe en relación a un agregado y no a datos aislaros. Ha elegido un gráfico adecuado para observar estadísticos útiles en la solución del problema, como la moda y el recorrido de la variable. Además, este estudiante amplía el eje vertical hasta el 20 porque eso le permite la comparación del número de caras de esta secuencia con la secuencia simulada (cuya máxima frecuencia fue 19). Lo mismo podemos decir del ejemplo C.



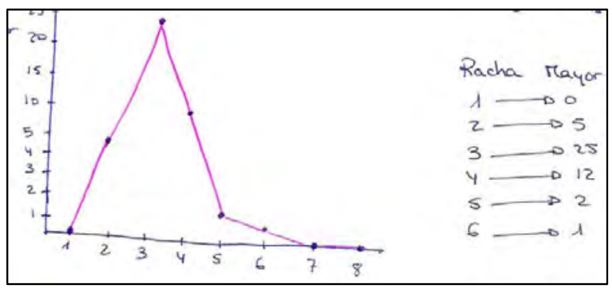
A)



B)



C)



D)

Figura 6.23. Ejemplos de gráficos que representan las variables por separado

En el segundo ejemplo (Figura 6.23B), la distribución de frecuencias está bien formada, como los casos anteriores, pero el alumno elige un gráfico poco adecuado para efectuar la comparación con la secuencia real. El diagrama de sectores permite ver las proporciones de cada valor dentro del total, pero no se visualiza tan bien el rango y la moda como en el gráfico de barras. El ejemplo D básicamente está correcto, pero la escala del eje no es homogénea, mostrando dificultades en el razonamiento proporcional.

Aparentemente este grupo de estudiantes se percatara de que a partir de un experimento aleatorio es posible definir diversas variables. Esto significaría que, al menos implícitamente, sí se logró la comprensión de la recursividad de la aleatoriedad y la extensión del experimento aleatorio simple de lanzar una moneda al experimento aleatorio compuesto por el lanzamiento de veinte monedas, lo que contribuyó al manejo matemático de las variables estadísticas y aleatorias correctamente.

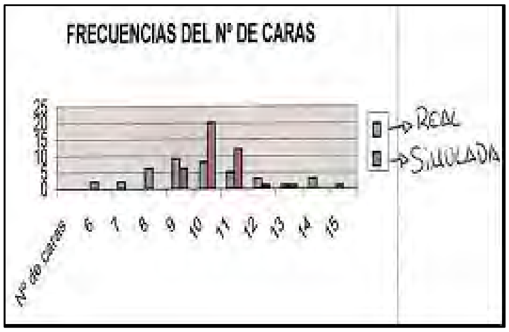
De acuerdo con Arteaga (2011) los gráficos clasificados en este nivel permiten responder preguntas hasta un nivel de lectura *más allá de los datos* en la terminología de Friel, Curcio y Bright (2001), puesto que una lectura directa de los gráficos permite identificar rápidamente la moda y la variabilidad de la variable.

4. *Gráficos conjuntos de las dos distribuciones.*

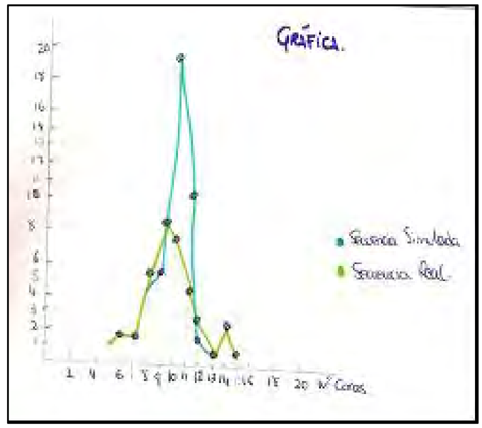
En este nivel quedaron agrupados los gráficos en los que el alumno formuló las distribuciones de cada par de variables de las dos secuencias y las representó conjuntamente en el mismo gráfico. Estos gráficos muestran una comprensión aún mayor de la variable estadística y su distribución puesto que presenta dos variables estadísticas que sí son comparables y relaciona una frecuencia con el valor de cada variable. Esto, potencialmente, facilita la comparación de las mismas para resolver la pregunta planteada en el proyecto. De acuerdo a Ciancetta (2007), los alumnos clasificados en los dos últimos niveles han superado la visión local de los datos y comienzan a reconocer los aspectos globales de los datos.

En la Figura 6.24 presentamos algunos ejemplos de los gráficos de este nivel. En el ejemplo A se muestra un diagrama de barras adosado de la distribución del número de caras en las secuencias real y simulada. El gráfico es correcto, tiene una escala apropiada para ambas secuencias, el rótulo del gráfico y la etiqueta de los ejes son apropiados. El alumno tuvo que haber relacionado las frecuencias de las secuencias real y simulada y luego elegir un gráfico en la hoja Excel que le pareciera adecuado para representar conjuntamente las dos secuencias y saber definirlas apropiadamente en el software.

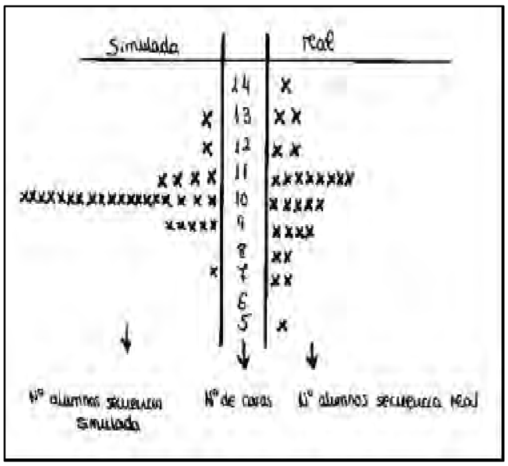
En el ejemplo B, el estudiante optó por dos polígonos de frecuencia conjuntos. Básicamente están bien representados, salvo que el estudiante no cuidó una limpia y buena presentación (líneas torcidas, se salta valores en la escala del eje-*y*, distinto tamaño en las unidades del eje-*x*) y eso no facilita una lectura inmediata. Aunque sí se preocupó por distinguir los gráficos de las dos secuencias con colores diferentes, lo que ayuda a una comparación de las dos secuencias porque se observa a simple vista el recorrido y la moda.



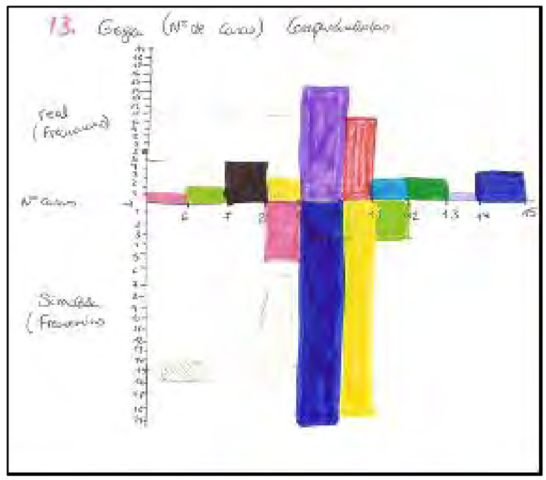
A)



B)



C)



D)

Figura 6.24. Ejemplos de gráficos que representan las variables conjuntamente

Igual que en el ejemplo A, el alumno B tuvo que haber relacionado las frecuencias de las secuencias real y simulada para realizar el gráfico y después seleccionar el tipo de gráfico que le pareció más adecuado. Se observa (por el trazo continuo de la línea) que los valores máximo y mínimo en los ejes los seleccionó previamente, sin embargo, pareciera que los números y las líneas de división los fue colocando sobre los ejes a medida que colocaba los puntos en el gráfico, por lo tanto, no hay proporcionalidad entre las distancias a las que están colocados los valores sobre los ejes.

El ejemplo D (Figura 6.24), también es básicamente correcto, salvo que las barras no están centradas. Aquí el alumno hace un uso creativo de la escala, porque aunque ellos están acostumbrados a que la parte de abajo del eje-*y* corresponde a números negativos, en este caso representa la frecuencia (con valores positivos) de la secuencia simulada. Los rótulos y las etiquetas del eje-*y* indican que el estudiante era consciente que lo que estaba dibujando en la parte de abajo era la frecuencia. El diagrama permite comparar muy bien ambas distribuciones de las secuencias, tanto en la variación, como la tendencia central de los datos. El gráfico C muestra un tratamiento parecido, con un manejo apropiado de los ejes, rótulos y escala. En este caso, es el lado izquierdo del eje-*x* no representa valores negativos sino la frecuencia (números positivos) de cada valor de la variable en la secuencia simulada.

Este nivel de graficación potencia la realización de inferencias informales pues posibilita responder preguntas hasta un nivel de lectura *más allá de los datos* en la terminología de Friel, Curcio y Bright (2001).

Niveles de comprensión de la distribución

Observamos que todos los gráficos mostrados en los niveles 2, 3 y 4 representan a una o más variables estadísticas, sin embargo, cada nivel potencia una mayor comprensión de las variables estadísticas y distribuciones que el anterior y su ejecución involucra el uso de diferentes objetos matemáticos que fueron analizados por Arteaga (2011).

Así por ejemplo, el orden de los valores que toma la variable en el eje-*x* sólo coincide con el orden numérico de algunos gráficos de los niveles más altos, 3 y 4. Las frecuencias absolutas únicamente aparecen representadas en los gráficos de barras y líneas (eje-*y*), aunque implícitamente también se consideran en el gráfico de puntos (ejemplo C del nivel 4, Figura 6.24). La idea de proporcionalidad aparece en los gráficos de puntos, barras o líneas de los niveles 3 y 4, pero no en el gráfico de puntos.

Mientras en los gráficos de puntos, barras o líneas de los niveles 3 y 4, los valores individuales se han clasificado y organizado de acuerdo a los valores que toma la variable para formar la distribución de frecuencias, en los gráficos de nivel 2, los datos individuales no presentan esa organización. Además, en los niveles 3 y 4 se representa una función de variables reales, en donde a cada valor de la variable, se le asigna su frecuencia (nivel 3) o una frecuencia en cada variable (nivel 4). Esta función no llega a construirse en los niveles anteriores, aunque en el nivel 2 sí se presenta una función no real que relaciona el valor de una variable estadística a cada estudiante o número de estudiante.

Es importante destacar que nuestra investigación indica que los diversos niveles reflejan el manejo de las variables estadísticas y aleatorias por parte del estudiante. De manera general, en el nivel 1 sólo hay un manejo de las variables relacionadas con un resultado del lanzamiento de una moneda, ζ_r, ζ_s y X_r, X_s (Sección 4.2.3), aunque no hay un manejo real de la distribución; los niveles 3 y 4 implican ya un manejo del resto de las variables, las relacionadas con el número de caras, η_r, η_s y Y_r, Y_s (Sección 4.2.4), número de rachas, θ_r, θ_s y T_r, T_s (Sección 4.2.5), y racha más larga, λ_r, λ_s y L_r, L_s (Sección 4.2.6). Sin embargo, en el nivel 3 la capacidad de vincular entre sí las variables estadísticas y con sus respectivas aleatorias, dependerá de las habilidades con otros recursos matemáticos a los que acuda el estudiante, mientras que en el nivel 4, la vinculación puede resultar más directa. El 2 implica el manejo de las variables relacionadas con el resultado del lanzamiento de la moneda (ζ_r, ζ_s y X_r, X_s) con una complejidad mayor, puesto que se toman en cuenta los datos de todo el grupo y también se potencia el manejo de las otras variables, pero no se llega a conformar ni a usar la distribución de las mismas. Así observamos que la clasificación de gráficos realizada por Arteaga (2011) también es útil para dar indicios de cómo los estudiantes podrían usar los gráficos como herramienta para realizar una inferencia estadística informal, aunque posteriormente veremos que no todos llegan a ella, pues finalmente dependerá de la capacidad interpretativa de cada estudiante (Makar, Bakker y Ben-Zvi, 2011).

Tabla 6.21. Clasificación de gráficos, según el nivel de comprensión de la distribución

Nivel	Categoría del gráfico construido	Frecuencia	Porcentaje
1	Representa sólo sus datos	4	4,54
2	Representa resultados individuales	13	14,77
3	Gráficos separados para cada distribución	46	52,27
4	Gráficos conjuntos de dos distribuciones	25	28,41
	Total	88	100

En la Tabla 6.21 presentamos la distribución de los gráficos elaborados por los 88 (79,3% del total de los alumnos) produjeron algún tipo de gráfico para analizar sus datos, incluso cuando las instrucciones de la tarea no lo pedían explícitamente, en función de los niveles descritos anteriormente. La mayoría de ellos (52,2%), produjeron gráficos separados para cada variable (nivel 3), el 28,4% de los estudiantes trabajan al nivel 4 y producen un solo gráfico de las dos variables. Son pocos los estudiantes que analizan sólo sus propios datos (nivel 1) y sólo el 14,7% estudian los valores obtenidos por cada estudiante caso a caso sin llegar a formar la distribución (nivel 2). Por lo que en su mayoría los gráficos realizados potencian el buen uso de las variables estadísticas y aleatorias por parte de los estudiantes.

Como hemos indicado, algunos estudiantes que no realizaron gráfico, sí construyeron una tabla de la distribución de una o varias variables. En la Tabla 6.22 clasificamos el conjunto de estudiantes de nuestra investigación en cuatro categorías de acuerdo a la representación gráfica o tabular de las variables del proyecto y sus distribuciones.

Tabla 6.22. Clasificación de alumnos de acuerdo a cómo representan los datos (en tabla o gráfico)

Forma de representar los datos	Frecuencia	Porcentaje
No representan	7	6,31
Muestran valores aislados	12	10,81
Representan distribuciones por separado	67	60,36
Representan distribuciones conjuntamente	25	22,52
Total	111	100,00

- *No representan tablas ni gráficos*; tampoco calcularon resúmenes estadísticos. Sólo encontramos 7 alumnos en este grupo que representaron gráficamente sus propios datos individuales (4 casos), o bien, hicieron referencia a ellos por escrito, sin representarlos. En general, estos alumnos no llegaron a usar las variables sugeridas por el profesor (número de caras, rachas y racha mayor en 20 lanzamientos), sino que manejan únicamente la variable «Cara o cruz» como producto de su experimento aleatorio simple.
- *Realizan un gráfico o un listado (o ambos) de los valores* de alguna de las variables citadas, sin llegar a formar la distribución (11%).
- *Forman las distribuciones de las variables y las representan en tablas o en gráficos (o ambos) por separado* (60%); usan tanto las variables sugeridas por el profesor, como sus distribuciones; pero no siempre llegaron a ser capaces de compararlas completamente.
- *Representan conjuntamente en tabla o gráfico (o ambos) las dos distribuciones de las variables de ambas secuencias*. Estos son los estudiantes que, hasta este momento, han hecho un mejor uso de las variables estadísticas y sus distribuciones y que, potencialmente, tienen mayor facilidad de interpretación y conclusión del proyecto.

7.2.2. Momentos de la variable estadística

Dentro de los elementos importantes que caracterizan a la variable aleatoria, Heitele (1975) señala su esperanza y variabilidad. La variable estadística que nos interesa analizar en el trabajo de los estudiantes es aquella vinculada con el fenómeno aleatorio y cuya generalización da lugar a la variable aleatoria, por lo tanto, las medidas de tendencia central y dispersión serán también fundamentales en la variable estadística. La generalización de estas medidas da lugar a la comprensión de las medidas de tendencia central y dispersión en las correspondientes variables aleatorias. Parte de los resultados incluidos en esta sección se han publicado en Ruiz, Arteaga y Batanero (2009a y b).

Tabla 6.23. Estadísticos calculados por los estudiantes (n=111)

	Correcto	Incorrecto	Total
Media	79	11	90
Mediana	55	17	72
Moda	77	3	80
Rango	54	6	60
Otra medida dispersión	41	0	41

En la Tabla 6.23 presentamos los estadísticos reportados en los productos escritos de los alumnos. Lo primero que nos llama la atención es que generalmente los estudiantes calculan más de uno, pero ninguno de los estadísticos fue calculado por el total de los alumnos. El estadístico más frecuente fue la media, seguida por la moda y al final la mediana. En general los estudiantes se detuvieron poco en medir la variabilidad de las variables, a pesar de ser un requisito para la comprensión de las ideas de variable aleatoria y distribución (Reading y Shaughnessy, 2004). El estadístico de variabilidad más frecuente fue el rango. En la tabla resumen que presentamos, sólo reflejamos si se calcula o no una medida y si se realiza correctamente, pues el comportamiento de un estudiante con respecto a estas medidas fue el mismo con todas las variables que analizaron.

Medidas de posición central

De un total de 111 estudiantes, 91 calculan algún promedio y la mayoría calculan más de uno. Esto indica una buena aproximación a la idea de distribución en los estudiantes, pues pasan del dato aislado a un resumen estadístico del conjunto de datos realizando una transnumeración (Wild y Pfannkuch, 1999). Los 20 estudiantes restantes, no usaron medidas de posición central al comparar dos distribuciones de datos, conducta que no es infrecuente, de acuerdo con Konold, Pollatsek, Well y Gagnon (1977).

El estadístico calculado con preferencia fue la media (sólo uno de los estudiantes que reportaron promedios, no la calculó) o la moda (80); la mediana fue calculada por 72 estudiantes (ver Tabla 6.23). Esto indica una dificultad algo mayor de comprensión de la mediana respecto a la moda o la media, coincidiendo con lo expuesto en la investigación de Mayén (2009).

En general el cálculo de los promedios fue correcto, lo que indica un buen dominio de los algoritmos de cálculo, aunque aparecieron algunos errores, como no ordenar los datos en la mediana (17 estudiantes), falta de ponderación en el cálculo de la media ponderada o error en la aplicación del algoritmo del cálculo de la media (11 estudiantes) y no tener en cuenta el caso de bimodalidad para la moda (3 estudiantes). Estos errores coinciden con los ya señalados en estudiantes de secundaria en investigaciones previas como la de Cobo (2003) y Mayén (2009). También encontramos dos ejemplos de confusión entre la media y la moda, uno de los cuáles se reproduce en la Figura 6.25, en que el estudiante hace referencia a la media, pero calcula la moda.

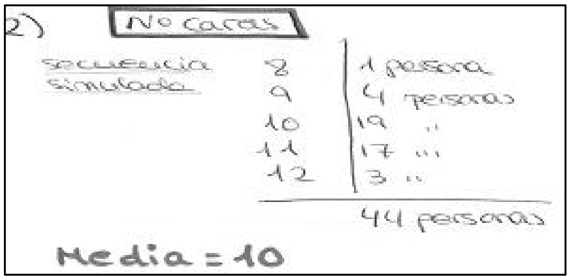


Figura 6.25. Confusión entre media y moda

También encontramos algunas concepciones erróneas sobre los promedios, explicitadas en la solución del proyecto por los estudiantes. Por ejemplo, el siguiente estudiante, confunde la variable de observación (número de caras en las diferentes secuencias producidas por cada estudiante) con los sucesos del experimento (cara y cruz), al calcular la moda da la siguiente respuesta:

> Moda = Son las caras porque su número máximo es 12, mientras que el número máximo de cruces es de 11 [*Alumno CE*].

En lugar de buscar el número de caras que aparezca con más frecuencia, compara el máximo del número de caras con el del número de cruces de entre todos los resultados del grupo, como las caras aparecen más veces, reporta que la moda es el suceso «caras». La confusión de este estudiante está claramente vinculada con una incomprensión de los valores que adopta la variable aleatoria, pero también se presenta una incomprensión del experimento aleatorio compuesto por los 20 lanzamientos de la moneda. Esta afirmación se refuerza cuando analizamos otro error que este mismo alumno comete en el cálculo de la mediana. Confunde el total de la muestra (46 secuencias) con el número total de experimentos simples en el conjunto de todas las secuencias de 20 elementos de los estudiantes (es decir, 46 por 20), como vemos en su respuesta:

> Mediana: Buscamos el valor que corresponde a N/2=(920/2-1), porque es par; 920=46*20 y veríamos si se corresponde con una cara o una cruz, pero no lo sabemos porque no podemos deducir la secuencia de caras y cruces del alumno nº 24 [*Alumno CE*].

El alumno trata de calcular la mediana considerando que cada repetición del lanzamiento de una moneda como un dato, es decir, no logra pasar al experimento aleatorio compuesto por el lanzamiento de 20 monedas (de ello supone 920 observaciones, en lugar de 46). También confunde el valor de la mediana (que debe ser un valor numérico) con los sucesos del experimento. Supone que la mediana se calcula localizando al alumno que queda al centro (él dice que es el alumno 24) cuando tiene ordenadas todas las secuencias de los 46 estudiantes de su grupo, es decir tiene una concepción correcta de mediana como centro. Pero como no conoce cómo aparecieron las caras y cruces del alumno 24, no puede deducir cuál es la mediana. El alumno *CE* pretende trabajar con las variables relacionadas con los resultados del experimento aleatorio simple (ζ_r, ζ_s y X_r, X_s, Sección 4.2.3) en la que los resultados son caras y cruces, lo que hace que su comprensión del objetivo del proyecto sea muy pobre, puesto que se restringe a trabajar con la variable «cara o cruz» y no lo hace correctamente.

En otros casos, cuando calculan la mediana, los estudiantes no tienen en cuenta la frecuencia, sino que le asignan el punto medio de los diferentes valores que toma la variable, lo que es equivalente a considerar la mediana como el centro geométrico del rango. Este error también fue encontrado por Mayén (2009) en su estudio sobre comprensión de los promedios. La ejemplificamos en la siguiente respuesta del alumno *CF* que, además de dar la definición de mediana a partir de su algoritmo, indicando que hay que ordenar los datos, confunde valor de la variable con los valores de las frecuencias y las ordena (que varían de 2 a 15) para obtener la mediana. Además, tampoco hace referencia a la mediana como centro de la distribución, sino sólo como una especie de rango.

La mediana: ordenar de mayor a menor 2-15 [*Alumno CF*].

De manera general, el cálculo de los promedios fue, en su mayoría, correcto. El mayor número de errores se cometieron entre los alumnos que calcularon la mediana (23.6%), puesto que es la medida de tendencia central menos intuitiva, en el sentido de que el promedio es muy usado y la moda es la más simple de localizar. El menor porcentaje de errores se tuvo en la obtención de la moda (3.75%).

Medidas de dispersión

Fueron pocos los estudiantes que calcularon las medidas de dispersión, lo que sugiere que los alumnos no sienten la necesidad de calcular estas medidas para comparar distribuciones. Sin embargo, la comprensión de la idea de distribución entraña conjugar a la vez las ideas de promedio y dispersión (Bakker y Gravemeijer, 2004), que muchos estudiantes en nuestro estudio no supieron relacionar.

Sólo el 54% del grupo calculó el rango y 37% calcularon otra medida de dispersión (ver Tabla 6.23), a pesar de que, como indica Shaughnessy (2007), la dispersión es una característica importante de la distribución de los datos. Casi la mitad de los participantes en la muestra no considera esa necesidad; es posible que por falta de comprensión de su significado o falta de familiaridad con el trabajo con las distribuciones.

La mayoría de estudiantes que usaron alguna medida de dispersión se inclinaron por el rango y 54 (90% de los que lo calcularon) lo hicieron correctamente. En su mayoría los que cometieron errores, tomaron incorrectamente los mismos valores del máximo, mínimo y rango tanto en la secuencia real, como en la simulada, lo que, como indicamos anteriormente, implica una incorrecta comprensión de la distribución de la variable estadística, puesto que unen el dominio de dos variables, lo cual no es válido aunque esas dos variables sean comparables. En algunos casos, los estudiantes incluyeron su definición correcta, por lo que, sospechamos, el error en el cálculo se debió a la necesidad de comparar dos variables relacionadas (es decir, esta comparación implica un nivel de comprensión más alto):

> El rango: es la medida de dispersión más elemental, la diferencia entre las dos observaciones extremas, o se la diferencia entre los valores mayor y menor del conjunto de datos. *[Alumno JC]*.

Los que prefirieron otra medida de dispersión calcularon la desviación típica (25), la varianza (15) o el coeficiente de variación (1) y en algunos casos dan una descripción de su significado con sus propias palabras:

> La desviación típica: Es la medida de dispersión más conocida y usada en combinación con la media como medida central. La idea clave es comprender bien la desviación con respecto a la media. Si se toma a ésta como referencia, las distancias de cada uno de los datos a ella, darán una buena idea de la dispersión total del conjunto de datos. *[Alumno JC]*.

Todas estas medidas de dispersión fueron obtenidas correctamente, lo cual indica que únicamente se inclinaron por estas medidas aquellos que sabían calcularlas bien. Algunos de ellos indicaron expresamente su necesidad:

> Medidas de centralización (media, mediana y moda), debe ir acompañadas necesariamente de las medidas de dispersión (rango y desviación típica). De esta manera obtenemos información más precisa para realizar el análisis de los datos de las tablas. *[Alumno ER]*.

Es evidente que las medidas de dispersión son menos intuitivas para los futuros profesores, por el menor uso que se hace de las mismas, como también se confirma en la investigación de Arteaga (2011). En ambas muestras fueron muy pocos los que usaron estas medidas y menos aun los que las interpretaron correctamente. Sin embargo, aquellos estudiantes que las usan cometen menos errores que en el caso de las medidas de tendencia central y cuando dan una definición intuitiva esta es correcta.

7.2.3. Discusión de resultados sobre el objeto de análisis variable estadística

Los resultados presentados en esta sección indican que los estudiantes participantes han comenzado a construir el concepto de variable estadística, valor y frecuencia de la variable y su distribución en un porcentaje apreciable, pues, a pesar de no habérseles indicado, tratan de formar la distribución de las variables para comparar las variables por parejas.

En nuestro estudio, de manera general, los estudiantes trataron de representar la distribución de las variables estadísticas y presentarlas ya sea en tabla o en gráfico, realizando un proceso de transnumeración (Wild y Pfannkuch, 1999). En la Tabla 6.19, se observa que sólo 7 alumnos (6,30%) usan sus datos aislados y 12 (10,8%) se limitan a listar todos los valores de la variable; es decir, los primeros no manejan las variables estadísticas y los segundos aunque sí las reportaron no usaron sus distribuciones. El resto, que es la amplia mayoría, produce una tabla o gráfico (o ambas cosas) de variables estadísticas, mostrando una comprensión intuitiva de la distribución.

Comparando la producción de tablas con la de gráficos, se nota una mayor construcción de gráficos que de tablas. El 79% de los estudiantes hicieron gráficos y el 56% hicieron tablas. Un 38% sólo reportaron gráficos sin tablas, mientras que un 14% presentó sólo tablas. Esto muestra una necesidad de visualización de las datos a través de gráficos o bien una mayor vinculación de las distribuciones con sus gráficos. Estos resultados coinciden con los de Arteaga (2011).

La notación usada en la tabla por diversos estudiantes denotó diversos grados de abstracción, siendo menos los que usaron un lenguaje más funcional o probabilístico, sin embargo, la abstracción que implica elaborar la distribución de frecuencias a partir de los datos, sí se logró en todos los estudiantes que elaboraron tablas. En su mayoría, la presentación de estas distribuciones en forma tabular en forma de listado vertical, pero pocos elaboraron una gráfica que facilitara la comparación rápida de las distribuciones entre las secuencias simulada y la real. En la representación gráfica de la distribución tampoco prevaleció la comparación como guía para la elaboración y selección del gráfico.

Algunos de los errores vinculados con la comprensión de la variable que afectan en la concreción de la distribución de la variable son:

- ❖ *Extienden el rango de la variable usando, implícitamente, frecuencia cero para valores de la variable que no aparecieron en los datos.* Agregan valores mayores o menores a los que aparecen en la muestra cuando elaboran el gráfico o la tabla, es decir, extienden el rango hacia arriba, hacia abajo o ambos, con valores de la variable estadística que no se obtuvieron en la secuencia analizada, pero sí en la otra secuencia y que resultarían hipotéticos en la variable aleatoria. A algunos estudiantes, esta maniobra les facilitó la comparación entre las distribuciones de las secuencias simuladas y las reales o la elaboración de sus gráficos. Sin embargo, en otros, ocasionó un cálculo incorrecto del rango que provocó que interpretaran que no había diferencia entre la variación de las variables vinculadas con la secuencia simulada y la real.

Por otro lado, para generalizar la variable estadística hacia la variable aleatoria, se requiere aceptar que la variable podría tomar valores diferentes a los que se presentan en una muestra. Por ello el error en la concepción del rango de la variable estadística, podría implicar confusión entre los valores de la variable aleatoria y los de la estadística y por lo tanto confusión entre estas dos variables. Esta confusión apareció en la entrevista clínica (Capítulo 5) de manera inversa y ocasionó que las estudiantes pensaran que lo aleatorio tendría que ser la probabilidad con el argumento de que en cualquier caso podrían conocer los valores que podría tomar la variable aleatoria, no así, la probabilidad. Esto, a su vez, transformó en la mente de las estudiantes, la variable aleatoria en una variable estadística, puesto que la distribución se transformó en la distribución de una muestra y no de la población bajo estudio.

- *Confusión entre el valor de la variable con el resultado del suceso aleatorio simple*. Algunos estudiantes daban el valor de «cara» o «cruz» al valor de la moda, media o mediana de las variables o bien graficaban estos dos resultados en lugar de las variables propuestas. Este es un error que encontramos en tablas, gráficos y cálculo de medidas de tendencia central y de dispersión. Para estos estudiantes el experimento aleatorio es el lanzamiento de una moneda, es decir no pudieron manejar implícitamente el experimento aleatorio compuesto de 20 lanzamientos de una moneda. Esto es muy natural puesto que a pesar de que el análisis requiere la definición de un experimento aleatorio compuesto, la evaluación de la aleatoriedad es sobre el lanzamiento de una moneda, además de que no se podía esperar la conceptualización formal de un espacio aleatorio compuesto de los estudiantes de la muestra. Se esperaba que el mismo contexto los ayudara a trabajar intuitivamente con el experimento aleatorio compuesto.

- *Confusión entre la convergencia de la frecuencia relativa y la frecuencia absoluta*. Algunos estudiantes sostuvieron la convergencia de la frecuencia absoluta, en lugar de la relativa. Esta confusión la consideramos, producto de la dificultad de la modelación estratificada en una inferencia informal discutida en las Secciones 9.5 y 9.6 del Capítulo 3. Ahí hicimos notar que la frecuencia relativa no es parte de la realidad sino que ya es un modelo matemático, puesto que hay que establecer la frecuencia como el cociente de las frecuencias para un valor entre el tamaño de la muestra. Este modelo estadístico no resulta simple de establecer y genera un obstáculo para la generación de inferencias estadísticas informales. Esta dificultad también se presentó en el Estudio 3, en la entrevista clínica, en donde las estudiantes no abstraían el cociente y preferían el manejo directo del número de trabajadores. Lo cual indica que esta dificultad se puede presentar tanto en el manejo de una inferencia informal, como en la probabilidad laplaciana (el Estudio 3 hacia uso de esta asignación de probabilidades).
- *Confusión entre el valor de la variable y la frecuencia absoluta*. Esto ocasiona que inviertan el gráfico o que reporten de manera errónea las medidas de tendencia central o de dispersión, esto último ha sido reportado en otras investigaciones anteriores, como la de Cobo (2003). Así mismo, ocasiona una confusión entre la variable dependiente e independiente en la distribución de frecuencias, tal como apareció en el estudio de la entrevista clínica (Capítulo 5).
- *Percibieron el lanzamiento de 20 monedas como la repetición del experimento aleatorio «lanzar una moneda»*. Esto provoca que la única variable con la que puedan trabajar sea con la de «cara o cruz» y por lo tanto no lograron trabajar con el resto de las variables aleatorias ni establecer sus distribuciones. Estos estudiantes manifestaron este error de diversas formas: graficación, cálculo de medidas y en el reporte de su tabla (sin agrupar). Algunos de ellos lograron trabajar con la variable «el número de caras en el lanzamiento de 20 monedas», pero en general, no lograron darle sentido a las variables vinculadas con las rachas.

Arteaga reporta que el porcentaje de estudiantes que grafican la distribución del número de caras (87.44%) es mucho mayor que el de aquellos que se detienen a graficar las distribuciones de las otras dos variables (70,53% para número de rachas y 61,84% para longitud de la racha mayor). Sin embargo, en nuestro estudio, los alumnos que presentan tablas lo hacen para todas las variables que se analizan en la Sección 4.2. Esto marca una diferencia importante con respecto a la elaboración de gráficos. Es posible que la elaboración de tablas represente una tarea más sencilla que la elaboración de gráficos; pero también es posible que estuviera relacionado con que, en su mayoría los estudiantes basaron su trabajo en la interpretación de los gráficos y estadísticos, más que en la de las tablas, y que no vieran la necesidad de analizar todas las variables. Esto significaría que los estudiantes no entendieron el sentido del análisis conjunto de las tres variables para evaluar la aleatoriedad, conformándose con un solo modelo, aunque los distintos estudios sobre la percepción de aleatoriedad subjetiva, muestran la necesidad de realizar el análisis haciendo uso de varios modelos (Green, 1982).

Como ya lo vimos en la sección anterior (7.1.2) al analizar sus concepciones sobre aleatoriedad, los estudiantes tienden a dejarse influir por sus conocimientos sobre la distribución de la variable «número de caras» para simular los resultados de una secuencia de 20 lanzamientos de la moneda. Más aún, observamos que los estudiantes condicionan involuntariamente el comportamiento de las otras dos variables (número de rachas y longitud de la racha más larga) a sus creencias con respecto a la primera variable. También hay que tomar en cuenta que en el momento de realizar el proyecto en clase, la única variable de análisis que emergió de manera natural por parte de los estudiantes fue el número de caras y las otras dos tuvieron que ser propuestas por el profesor que dirigió la sesión. Ello indica que las variables relacionadas con las rachas fueron más difíciles de comprender para los estudiantes, lo cual explicaría que sean menos estudiantes los que las grafiquen y analicen.

7.3. Variable aleatoria y su relación con la variable estadística

En esta sección evaluaremos las concepciones de los estudiantes sobre la variable aleatoria y su relación con la variable estadística. Puesto que no han estudiado dentro de sus cursos anteriores el concepto de variable aleatoria, sus concepciones respecto a la misma serán intuitivas. Por otro lado, no se les preguntó explícitamente sobre ella, de modo que hemos de inferir a partir de los argumentos en que hacen referencia a las variables consideradas en el proyecto y que se analizaron en la Sección 4.2. A continuación presentamos y analizamos estos argumentos.

7.3.1. Referencia a la variable

Son muchos los estudiantes que hacen referencia implícita a las variables aleatorias cuando se refieren a las estadísticas. En nuestro primer ejemplo, el alumno compara la variable número de caras en las dos secuencias señalando la variabilidad de valores que toma y diferenciando claramente entre el valor de la variable y la posición que ocupa en el conjunto de datos.

> En cuanto a la hoja de datos tengo que decir que el nº de caras de las secuencia simulada está en torno de 7 a 14 caras; en cuanto el nº se rachas simultáneas están entre 4 y 9 rachas y el tamaño mayor de estas rachas es de 7 que pertenece al alumno 21 que tiene 7 rachas [*Alumno CB*].

Aquí el alumno menciona explícitamente el rango de las variables estadísticas y hace una deducción hacia la variable aleatoria. Se observa a un estudiante consciente de que el comportamiento descrito es de una muestra.

En el segundo ejemplo, el alumno describe la tabla de datos y hace referencia a las seis variables analizadas y al valor obtenido para cada una por cada estudiante en su grupo:

> La tabla de datos está dividida en 3 columnas; en la primera encontramos el número de alumnos que ha realizado el experimento; en la segunda encontramos la secuencia simulada y vemos como a su vez se divide en 3 columnas independientes donde en la subcolumnna 1 aparece el número de caras que se han obtenido en total en la secuencia, en la subcolumna 2 encontramos el número de rachas obtenidas y en la 3 la longitud de la racha más larga [*Alumno SG*].

La siguiente alumna utiliza explícitamente el término variable, para referirse a la longitud de las rachas. Su respuesta refleja, sin embargo, una concepción incorrecta de la independencia de ensayos:

> Otra de las variables a tener en cuenta es la longitud de las rachas, valorando que a medida que el número de caras o de cruces aumenta, la posibilidad de que salga la contraria es mayor [*Alumna MU*].

En el siguiente ejemplo se comparan los diferentes valores obtenidos en dos de las variables de los dos tipos de secuencias. Aunque se usa incorrectamente la palabra «posibilidades», en vez de «valores», el uso de esa palabra indica que se está pensando en términos de azar y, en consecuencia, de variable aleatoria.

> En las gráficas que se han obtenido al representar las rachas se pueden destacar varios resultados. Por ejemplo, en la secuencia simulada encontramos cinco posibilidades y en la real son siete. *[Alumna HC]*.

Cuando los estudiantes hacen referencia al azar o al experimento aleatorio, relacionan la variable con la aleatoriedad del experimento y la dificultad de predecir los valores, es decir, ven el valor de la variable como resultado de un experimento aleatorio. La siguiente alumna indica que, debido al carácter aleatorio será difícil que los valores de las variables sean exactamente los mismos al tomar nuevos datos de la variable:

> El azar nos demuestra que nunca, o es muy difícil, que varias veces seguidas salga la misma muestra de valores *[Alumna AM]*.

En ocasiones sus concepciones correctas sobre la convergencia de la frecuencia relativa (de que a la larga se equilibra el número de caras y cruces) les llevan a una percepción incorrecta sobre la independencia de ensayos, y por lo tanto, también sobre los valores que tomaría la variable aleatoria. Esperan una convergencia rápida en pequeñas muestras y recurren a la heurística de representatividad (Kahnmenam, Slovic y Tversky, 1982):

> De ahí que la probabilidad de que nos salga 4 o 5 veces seguidas el mismo resultado es casi imposible, de ahí que al hacer una tirada simulada haya puesto como racha más larga la de 3, ya que casi seguro que se van a ir alternando [*Alumna AM*].

7.3.2. Referencia a las probabilidades o distribución

En algunos casos los estudiantes aluden a la probabilidad de valores de las variables analizadas, que estiman a partir de las frecuencias obtenidas en los experimentos, es decir, aplicando una inferencia informal para obtener la probabilidad. De este modo, al hacer referencia a la probabilidad de obtener un cierto valor de algunas de las variables, hacen referencia implícita a la variable aleatoria que corresponde a la variable estadística dada y a su distribución de probabilidad o al menos a la probabilidad de algunos de sus valores aislados. En otros casos, los estudiantes recurren a analizar el experimento aleatorio simple («cara o cruz») haciendo uso de la concepción laplaciana y hacen algunas deducciones sobre los sucesos del espacio muestral, aunque, ello no les permite obtener probabilidades de las variables aleatorias ni conformar la distribución de probabilidad.

En nuestro primer ejemplo, el alumno indica que algunos valores son más probables que otros basándose en los datos empíricos y, asimismo, hace una estimación de la moda. Por tanto, implícitamente usa las ideas de variable aleatoria, sus valores y las probabilidades de sus valores y su valor esperado:

> En la secuencia real nos damos cuenta de que la mayoría de las veces en las que aparecen las caras se encuentran en torno a 10… Como observamos en el fenómeno real, estamos en torno a diez, distribuidos principalmente entre 8 y 12. [*Alumna LAE*].

Son muchos los estudiantes que hacen referencia a la equiprobabilidad de resultados en el experimento simple y sacan la consecuencia sobre el valor esperado de veces que se obtendrá cada suceso a lo largo de los 20 ensayos:

> Al lanzar una moneda 20 veces seguidas, los resultados que debemos obtener, sería una igualización, sin ser exacta, entre el número de caras y de cruces, puesto que al lanzarla, tenemos un 50% de posibilidades de obtener tanto cara como cruz [*Alumna MU*].

> Se trata de un juego de azar en que la probabilidad juega un papel muy importante, por tanto cabe la misma posibilidad de que salga cara o cruz *[Alumna AG]*

> Con todo esto se llega a la conclusión de que hay un 50% de posibilidades de que salga cara y otro 50% de que salga cruz. Esto se debe a que sólo pueden obtenerse dos resultados diferentes, y por ello los 40 alumnos hemos obtenido resultados diferentes *[Alumna CG]*

Algunos alumnos, como AR, incluso cuantifican la probabilidad de obtención de cara y cruz y llegan a estimar el valor esperado de la variable Binomial, número de caras al lanzar las 20 monedas, sin dejar de reconocer la variabilidad de dicha variable en los diferentes experimentos.

> En el caso de las caras y las cruces suponemos que hay un 50% de probabilidades de que salga una u otra, por tanto las simulaciones son todas en torno a 10 caras. Sin embargo en el ejercicio real el número de caras y de cruces varía mucho de una experiencia a otra [*Alumno AR*].

El alumno EB hace referencia a la probabilidad de los sucesos aleatorios en el experimento simple correctamente:

> Existe la misma probabilidad de que salga cara o cruz y en veinte tiradas la probabilidad de que salga todo cara es la misma de que salga todo cruz y de esta forma con cada una de las combinaciones que se pueda formar [*Alumno EB*].

No queda claro si la afirmación de EB sobre las combinaciones que se pueden formar en el experimento aleatorio compuesto se refieren a los valores de la variable (es decir, «*todo cara*» sería 20 caras y «*todo cruz*», cero caras); o a los sucesos del espacio muestral (es decir, las «*combinaciones que se pueden formar*» serían los sucesos del espacio muestral teniendo en cuenta el orden de las caras y las cruces). El primer razonamiento sería erróneo y mostraría el sesgo de equiprobabilidad con respecto a los valores que toma la variable aleatoria; estaría confundiendo los valores de la variable aleatoria con los sucesos del espacio muestral. No obstante, estaría haciendo referencia a la variable aleatoria binomial y su distribución, aunque la estimación que hace de sus probabilidades sería incorrecta. Si estuviéramos en el segundo caso, la afirmación sobre las probabilidades de cada combinación sería correcta.

También encontramos algunos casos en que de forma intuitiva se alude a la distribución de frecuencias de la variable estadística, como el caso del siguiente estudiante, quien explícitamente usa el término «variable estadística» y, en forma correcta, la propiedad cuantitativa de la variable. El alumno describe la forma de construir la tabla de frecuencias que luego elabora correctamente.

> Para efectuar una interpretación global del conjunto de datos, hacemos una tabla de frecuencias. La más sencilla consta de dos columnas; una para las modalidades distintas de la variable estadística ordenadas de, al ser la variable cuantitativa, de menor a mayor y otra con las frecuencias [*Alumno LG*].

En algún caso se relaciona la probabilidad a priori de los dos sucesos del experimento aleatorio simple (que es de 0,5) con la frecuencia observada de caras y cruces que es aproximadamente la misma. Es decir, se explica la frecuencia de resultados de la variable estadística por las probabilidades asociadas a los sucesos en el experimento aleatorio simple, estableciendo una relación entre frecuencia y probabilidad, es decir, entre variable aleatoria y estadística:

> Pero si nos podemos dar cuenta en su observación de que la representación de la secuencia es demasiado lineal y parecida, tanto en número de caras como de cruces. Lo cual es debido a que hemos utilizado el 50% de que salga cara y el 50% de que salga cruz [*Alumno AF*].

También otros son conscientes de que la convergencia se produce en la distribución de la variable aleatoria y esta se puede predecir, aunque no el resultado aislado o el orden en que aparecen los diferentes resultados del experimento aleatorio simple (en un suceso):

> En cuanto a la tabla que hemos realizado tengo que decir que en mi caso han coincidido bastante los resultados, no en el mismo lugar, pero sí en el número; tanto en el lanzamiento real como en el simulado tengo 11 caras y 9 cruces; en cambio no he coincidido en el número de rachas y la racha más larga *[Alumna AG]*.

7.3.3. Referencia a los parámetros

Nos interesamos también por ver si los estudiantes tienen concepciones intuitivas sobre los parámetros de posición central y dispersión de las variables aleatorias, elementos fundamentales de ésta (Heitele, 1975), y los relacionan con los correspondientes resúmenes estadísticos de las variables estadísticas.

En nuestro primer ejemplo, hay referencias a la moda de la variable aleatoria binomial «número de caras al lanzar 20 monedas» en algunos estudiantes, que hacen mención explícita al valor más probable de esta variable:

> Que el número que es más probable que salga (nº de caras) es de 10 [*Alumno CE*].

Mucho menos intuitivos han sido los resúmenes estadísticos de las otras variables. Algunos alumnos muestran explícitamente sus concepciones erróneas de la independencia, que implícitamente quedaron plasmadas en la distribución del número de rachas de las secuencias simuladas. En la afirmación de la alumna AM se observa este error en el valor esperado de la longitud de la racha más larga:

> Sabemos de sobra que en una secuencia de tiradas de una moneda el azar nos demuestra que nunca o es muy difícil que varias veces seguida salga la misma muestra, es decir, la probabilidad de que nos salga 4 o 5 veces seguida el mismo resultado es casi imposible, de ahí que al hacer una tirada simulada hay puesto como racha más larga la de 3 ya que casi seguro se va a ir alternando el resultado [*Alumna AM*].

El alumno EJG, en cambio, analiza correctamente las creencias de sus compañeros sobre la independencia que hacen que la longitud de la racha más larga sea diferente en la distribución de la secuencia simulada y la real.

> Esta diferencia se debe a que si a un alumno/a se le ofrece predeterminar un lanzamiento de moneda, éste no suele poner varias veces un mismo resultado, es decir, no suele decir: cara, cara, cara, cara, cruz *[Alumno EJG]*.

En algunos estudiantes hay confusión entre las ideas de variable y valor de la variable, como el siguiente ejemplo, en donde el alumno trata de explicar la mayor dispersión de valores en las secuencias reales. Sin embargo, el estudiante hace uso explícito de la idea de dispersión, que expresa con el término variedad:

> Se observa que en las secuencias reales hay más variedad de variables que en las simuladas. Por lo tanto, los alumnos se arriesgan poco [*Alumno ER*].

Con otras expresiones se hace referencia a la variabilidad. En general los estudiantes tienen poca competencia en comunicación verbal y un uso poco preciso del lenguaje matemático. En estos dos ejemplos, las estudiantes ni siquiera hacen referencia a las variables sobre las que están concluyendo, aunque por el desarrollo del reporte de su proyecto, se intuye que se refieren al número de caras:

> Como conclusión, a raíz del análisis comprobamos que con las gráficas se puede observar como en la secuencia simulada hay más constancia, es decir, que se tiende a los mismos valores y en la secuencia real hay mucha más variedad, dispersión, incluso en ocasiones cambios bruscos *[Alumna AC]*

> Llegamos a la conclusión de que la experiencia simulada tiene un número de valores más parecidos entre ellos que los de la tabla de reales *[Alumna JM]*.

7.3.4. Discusión de los resultados obtenidos en la relación entre las variables aleatoria y estadística

Aunque los estudiantes no habían estudiado formalmente el concepto de variable aleatoria, hubo varios indicios que indican una transición de la variable estadística a la variable aleatoria. Dentro de las más importantes están, por ejemplo, el uso constante de la palabra «probabilidad» o «posibilidad» en lugar de «frecuencia», vinculado con algún valor de la variable puesto que en estos casos hay una intuición de la distribución de probabilidad de alguna de las variables y no solo de la probabilidad de un suceso del experimento aleatorio. También encontramos referencias a la convergencia de la frecuencia relativa a la probabilidad y algunas muestras de distinción entre muestra y población. Además, varios estudiantes hicieron uso de la palabra variable y distinguieron los valores de la variable de la variable en sí misma.

En lo que sigue resumiremos algunas acciones que en nuestro estudio contribuyeron de manera favorable a facilitar la realización de una inferencia estadística y, por lo tanto, a la generalización de la variable estadística.

- ❖ En ocasiones la elaboración gráficos o tablas explícitamente para comparar las distribuciones de las variables de las secuencias reales con las simuladas, ayudaron a la transición de la variable estadística hacia la aleatoria, pues una correcta comparación requeriría suponer la existencia de valores teóricos posibles para la variable aleatoria que no se obtuvieron en la variable estadística y cuya frecuencia sería cero. Sin embargo, como ya notamos en la Sección anterior (7.2.3), esto también generó errores en el cálculo del rango. La mayoría de los estudiantes recurrieron al rango y una sobrevaloración de la dispersión de las distribuciones.
- ❖ El manejo apropiado de los resúmenes estadísticos como las medidas de tendencia central y dispersión también favorecen la realización de inferencias, puesto que se encontraron estudiantes que hacían uso de ellos y mostraban ideas apropiadas sobre el uso implícito de la variable aleatoria, sus valores y sus probabilidades.
- ❖ Clasificar a la variable estadística como cuantitativa es un paso hacia la abstracción de la variable aleatoria. Esto no lo pudieron hacer aquellos estudiantes que consideraron sólo el experimento aleatorio del lanzamiento de una moneda, puesto que para ellos, los únicos valores que podía tomar la variable era cara o cruz.
- ❖ La aceptación de la convergencia de la frecuencia relativa, hacia la probabilidad daría lugar al cumplimiento de una de las condiciones más

importantes para la modelación de la variable aleatoria al realizar una inferencia informal (Figura 3.6, Capítulo 3).

- ❖ La relación directa que establecen algunos estudiantes entre el resultado del experimento con el valor que toma la variable aleatoria[7] también fue una señal de la vinculación entre las variables aleatoria y estadística. Esto favoreció la modelación de la distribución de probabilidades en un primer estrato de modelación en el Estudio 3 (Punto 8.3.5, Capítulo 5). Sin embargo, Miller (1998) alerta sobre los inconvenientes de establecer esta relación, puesto que favorece ver los valores que toma la variable aleatoria como sucesos.

Al contrario, acciones que dificultaron el establecimiento de la relación ente la variable aleatoria y la estadística y que se presentaron en nuestro estudio, fueron:

- ❖ La confusión entre los valores de la variable aleatoria con los sucesos del espacio muestral llevaría al estudiante no sólo al problema de equiprobabilidad sino a confundir la probabilidad (función de conjunto), con la distribución de probabilidad (función real).
- ❖ El enfoque en el resultado aislado (Falk, 1986), hace que el estudiante no acepte la predicción estadística sustentada en la distribución de probabilidades. Esta creencia provoca que no pueda ir más allá de sus propios datos o de una predicción determinista.
- ❖ El trabajo únicamente con la distribución Bernoulli merma el trabajo con las variables relacionadas con las rachas y provoca errores con la variable «número de caras». Es posible que una de las causas sea que los estudiantes no visualizaron el experimento aleatorio compuesto por 20 lanzamientos sino solamente el experimento de lanzar una moneda.

7.4. Objeto de análisis: Ciclo de modelación

En la Sección 9 del Capítulo 3, vimos que gran parte de la actividad estadística puede ser descrita como un proceso de modelación e hicimos hincapié en un proceso de inferencia informal en el que existen dos estratos de modelación (Figura 3.7) en los que la variable estadística juega un doble papel, como «modelo» en el primer estrato y como «realidad» en el segundo[8]. A su vez, el proyecto planteado a los estudiantes está

[7] En su mayoría coincide con la acepción histórica de Parzen (1971), discutido en la Sección 6.2 del Capítulo 4, en el que visualizan el resultado del experimento aleatorio como el valor que toma la variable aleatoria.
[8] Para Henry (1997), el primer estrato sería la construcción de un modelo pseudoconcreto.

diseñado para que, potencialmente, los estudiantes recorran todos los pasos de la modelación; por este motivo también nos interesamos por el grado en que los estudiantes de nuestra muestra lograron completar los pasos del ciclo de modelación señalados por Dantal (1997).

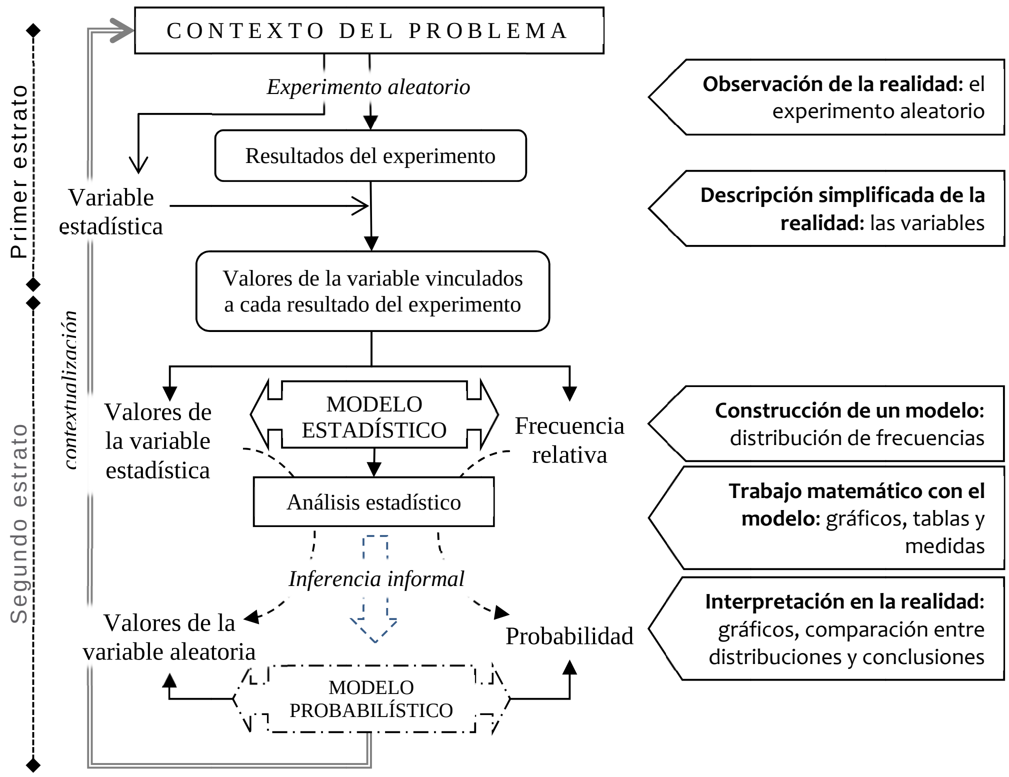


Figura 6.26. Modelación en estratos descrita a través de los pasos de modelación de Dantal, 1997.

En la Figura 6.26 relacionamos la modelación por estratos discutida en la sección 9.6 del Capítulo 3 con los pasos de modelación de Dantal para visualizar la forma en que analizaremos los resultados obtenidos por los estudiantes en esta sección. La descripción simplificada de la realidad (Dantal, 1997) implica por sí misma una modelación, por lo tanto, la definimos como un estratos de modelación (Heitele, 1975). La inferencia informal la hacemos ver como una interpretación de la realidad, puesto en las conclusiones los estudiantes requieren vincular las variables definidas con el contexto del problema[9].

[9] El modelo probabilístico sólo se infiere a través de sus medidas, lo cual, en nuestro proyecto, se observa en la comparación entre las distribuciones para obtener una conclusión.

7.4.1. Observación de la realidad

El primer paso en la actividad de modelación es observar la realidad. Nuestro proyecto gira alrededor del cuestionamiento sobre las intuiciones de los estudiantes, es decir, sobre el carácter aleatorio de las secuencias simuladas por ellos. La realidad, por tanto, estará suscrita por el análisis de qué es aleatoriedad y la forma de evaluarla. Para poder hacer el estudio es indispensable la aceptación de la aleatoriedad del lanzamiento de una moneda en las secuencias reales, lo que dará un punto de referencia del comportamiento de la aleatoriedad en los lanzamientos de una moneda. También es necesario que los estudiantes se percaten que se requiere el planteamiento de un experimento aleatorio compuesto que auxilie en la evaluación de las secuencias de resultados del lanzamiento de una moneda. No esperamos que enuncien el espacio muestral ni que obtengan la probabilidad vinculada a cada evento, sin embargo sí se espera un manejo intuitivo que les permita vincular las variables enunciadas con la evaluación de la aleatoriedad.

Dantal (1997) señala la dificultad de diferenciar, dentro de los fenómenos en los que el azar está presente, entre situación aleatoria y contingente. Las situaciones aleatorias son reproductibles en las mismas condiciones y cuyos resultados pueden analizase estadísticamente. En una situación contingente también hay incertidumbre, pero no es reproductible en las mismas condiciones. A efectos del análisis estadístico, en nuestro estudio los estudiantes tienen que aceptar la equivalencia de los experimentos de cada alumno en particular. Es decir, asumir no sólo que las condiciones de los lanzamientos reales de la moneda son, al menos teóricamente, iguales (que no hay sesgos personales de lanzamiento, ni monedas cargadas, etc.); sino también que las simulaciones de los estudiantes se realizan en las mismas condiciones, es decir, que las secuencias simuladas son parte de una misma población; sólo así se podrá caracterizar la aleatoriedad del experimento real como un modelo para el análisis de las intuiciones de los estudiantes reflejadas en los experimentos simulados.

Como lo hicimos notar en la Sección 7.1, de manera general este objetivo se cumplió. Los estudiantes aceptaron el carácter aleatorio de los lanzamientos reales puesto que se avinieron a analizar conjuntamente los datos de toda la clase sin cuestionar si se podían o no mezclar los datos de los estudiantes. Sin embargo, pudiera ser que aquellos alumnos que se limitaron a analizar solo los datos de su experimento, no llegaran a visualizar cada experimento individual como un ensayo que forma parte de un experimento colectivo. Una explicación alternativa es que consideraran sus propias intuiciones diferentes y las de sus compañeros, y por lo tanto, sin carácter aleatorio.

Cabe aclarar que el experimento consistente en la simulación de 20 lanzamientos de una moneda por parte de los estudiantes también tiene carácter aleatorio y que éste también fue aceptado por la mayoría de los estudiantes, puesto que los valores de las variables deducidas a partir del resultado de sus simulaciones y las de sus compañeros fueron consideradas repeticiones de un mismo experimento. Es claro que sin este supuesto, no tendría sentido el análisis estadístico de este experimento ni su comparación con los resultados obtenidos en el experimento real. En nuestro estudio, buscamos que el estudiante cuestione el comportamiento de las secuencias simuladas como una imitación de los lanzamientos reales, no su carácter aleatorio.

Es posible que los que analizaron únicamente sus datos sólo tomaran en cuenta el experimento simple (el lanzamiento de una moneda) y por lo tanto, una repetición del experimento compuesto sería vista como 20 repeticiones del experimento aleatorio simple. Implícitamente, los estudiantes tenían que visualizar los 20 lanzamientos de una moneda como el experimento aleatorio que tenían que analizar y del que surgen las variables con las que se evalúa la aleatoriedad de las secuencias. Esto, como observamos en la sección 7.1.1, no fue aceptado por todos los estudiantes y trajo como consecuencia que ellos no pudiesen abstraer suficientemente para trabajar con todas las variables, sobre todo aquellas vinculadas con las rachas.

7.4.2. Descripción simplificada de la realidad

La simplificación de la realidad permite pasar de la realidad observada (paso 1) a la construcción del modelo (paso 3), puesto que no es posible trabajar con todos los aspectos que intervienen en la realidad. Implica ser conscientes de la dimensión del problema a tratar y reducirlo a una realidad en la que sí podamos emitir una respuesta.

En el desarrollo del proyecto, la realidad se ha simplificado al pasar del problema general de estudiar las intuiciones sobre aleatoriedad al estudio de un experimento aleatorio concreto. Podríamos haber sugerido el estudio de muchos otros experimentos aleatorios, por ejemplo, escribir dígitos al azar, dibujar puntos al azar en un plano, etc., pero para tratar de responder a la pregunta de si los estudiantes tienen buena intuición sobre la aleatoriedad se planteó el experimento de lanzar 20 veces una moneda. Este esclarecimiento de los límites también nos obliga a dar la respuesta a la pregunta inicial con base en este experimento.

Ríos (1967) afirma que otra simplificación de la realidad al trabajar con variables estadísticas consiste en la aceptación de la estabilidad de las frecuencias de los valores de la variable. Como hicimos notar en la Figura 3.7 (Capítulo 3), esta aceptación también es importante en nuestro proyecto puesto que la inferencia hacia la variable aleatoria sólo es posible si se acepta la convergencia de las frecuencias a la probabilidad. En la sección 7.3.2, consideramos una evidencia de vinculación entre la variable aleatoria y la estadística la referencia por parte de los estudiantes a la probabilidad y dimos algunos testimonios de los estudiantes sobre este punto. Los estudiantes visualizaron esta convergencia principalmente en el experimento simple (el lanzamiento de una moneda), cuya probabilidad para cada uno de sus resultados sería de 0,5. En el resto de las variables, los estudiantes no tenían conocimiento suficiente para saber cuál sería el valor al que convergería la probabilidad de cada valor de la variable, sin embargo sí encontramos comentarios sobre el número esperado de caras, que sería 10 en los 20 lanzamientos.

La simplificación también involucra una abstracción de la realidad[10]. Heitele (1975) propone una modelación en varios niveles de realidad, en donde en cada estrato de modelación se toman los aspectos más importantes de ésta para generar un modelo que funcione como «realidad» en otro estrato. En nuestro caso, en el proceso de modelación estratificado (Figura 6.26) observamos la importancia de la variable estadística (que es un primer modelo matemático) como vínculo entre la realidad y el modelo de la variable aleatoria.

El análisis de los resultados del experimento aleatorio (secuencias de caras y cruces) requiere la definición de variables útiles (número de caras, rachas y longitud más de la racha más larga) para dar una respuesta al proyecto. Esto implicará una abstracción de la realidad, puesto que tomaremos los valores que estas variables adopten en cada secuencia en lugar de trabajar con la secuencia de caras y cruces en sí misma. El hecho de que en nuestro proyecto sea necesaria la definición de varias variables, favorece que el estudiante pueda ver la aleatoriedad como multiplicidad de modelos. Al mismo tiempo, las variables estadísticas analizadas constituyen un estrato de la modelación porque al aplicarlas como regla de correspondencia (función) a la secuencia de cada estudiante, da como resultado los valores numéricos que adopta la variable estadística (por ejemplo, el valor del número de caras).

En resumen se parte de una «realidad» no matemática (intuiciones de los estudiantes) y de ella al conjunto de secuencias producidas por ellos para construir unos modelos matemáticos (variables estadísticas definidas por el profesor), que a su vez funcionarán como «realidad» en la construcción de las distribuciones de dichas variables estadísticas. Los datos dejan de ser el conjunto de las secuencias obtenidas por

[10] De acuerdo con Heitele (1975) diversos problemas en la comprensión de la modelación en probabilidad y estadística son debidos a que se considera «realidad» algo que en sí ya es un modelo. En el estudio disciplinar de esta memoria de tesis, analizamos estos puntos más ampliamente (Secciones 9.4 y 9.5, Capítulo 3).

cada estudiante, para convertirse en el conjunto de valores de cada variable que se obtiene de la secuencia obtenida por cada estudiante (tabla de datos). La definición de la regla de correspondencia que permite definir a la variable estadística (función que asigna un número a cada secuencia) tiene las mismas dos aristas que ya hemos discutido en el Capítulo 3 sobre la variable aleatoria: es necesario que sea útil para resolver el problema tratado y también tiene que garantizar una matematización de la situación.

En nuestra investigación, no se esperaba la definición de las variables por parte de los estudiantes, aunque algunos sugirieron espontáneamente analizar el número de caras o de cruces en las secuencias, como lo mostramos en el análisis de sus concepciones sobre la aleatoriedad (Secciones 7.1.1 y 7.1.2). Los estudiantes no solo son conscientes de que una característica importante en una secuencia de lanzamientos de una moneda es el número de caras que se obtienen, sino además saben que la forma en que éstas están ordenadas, también lo es. Esto facilitó al profesor proponerles el análisis de las variables vinculadas con las rachas sustentándose en investigación sobre el tema de aleatoriedad, lo que además, permite comparar los resultados con esos estudios.

Esto explica que estas variables fueran menos interpretadas y analizadas por los estudiantes. En unos pocos casos los alumnos analizaron también el número de cruces en las secuencias producidas, sin embargo esta variable no era necesaria, puesto que puede ser descrita rápidamente a través del número de caras.

7.4.3. Construcción de un modelo

Wild y Pfannkuch (1999) tienen una visión muy amplia de modelo estadístico y consideran el trabajo con modelos una parte esencial del razonamiento estadístico. Para ellos un diagrama de barras o gráfico de sectores es ya un modelo que refleja la distribución de los datos, pues los datos ya se han abstraído y agrupado. En este sentido, los estudiantes han mostrado una buena capacidad para construir espontáneamente modelos matemáticos, que han sido analizados con detalle en las secciones anteriores, y son:

- Tablas de la distribución de una o varias variables estadísticas.
- Gráficos de la distribución de una o varias variables estadísticas. Los estudiantes mostraron una gran variedad de gráficos e incluso recurrieron a gráficos no convencionales y un estudiante inventó su propio gráfico no estándar.
- Medidas de posición central y dispersión: media, moda, mediana, rango y en algunos casos varianza y desviación típica.

No todos los modelos resultaron apropiados ni correctos, puesto que gran parte de ellos no sólo reflejaron los conocimientos de los estudiantes sobre estadística sino también su conocimiento sobre el contexto del problema (Makar, Bakker, y Ben-Zvi, 2011).

Dicho conocimiento del contexto se observa cuando los estudiantes hicieron uso de sus creencias sobre aleatoriedad, que analizamos en el punto 7.1 de este capítulo. Estas concepciones son parcialmente correctas y es muy sutil el paso de una concepción correcta a una que no lo es. Así mismo, en la simulación del lanzamiento de 20 monedas, desconocían el valor esperado del número de rachas o la longitud más larga, pero sabían que el número de caras medio es 10, lo que condicionó la secuencia simulada del grupo, ignorando la variación. Otro fuerte condicionante fue su creencia de un azar vinculado a la irregularidad: las secuencias no fueron muy largas. Los estudiantes se mostraron preocupados porque su secuencia simulara el lanzamiento aleatorio de una moneda, cuando en realidad estaban siguiendo un modelo matemático subjetivo generando ensayos dependientes.

Así, los estudiantes construyeron y tuvieron oportunidad de analizar modelos matemáticos que dependen de sus creencias sobre la forma en que se comporta una sucesión de resultados aleatorios al lanzar 20 veces una moneda. Ese *modelo interno* (conocimiento sobre el contexto) influye, como veremos más adelante, en la forma en que los estudiantes analizan e interpretan sus resultados estadísticos.

7.4.4. Trabajo matemático con el modelo

El trabajo matemático con el modelo ha sido variable en los estudiantes, no sólo por la diversidad de modelos que han obtenido, sino porque matemáticamente el trabajo desarrollado no siempre fue correcto. En lo que sigue veremos que la mayoría de los estudiantes han trabajado correctamente al menos con algunas medidas de posición central, también fue mayoría el número de estudiantes que construyó tablas y gráficos de la distribución al menos parcialmente correctos. También la notación empleada en el cálculo de los estadísticos marca una diferencia en el grado de abstracción de los estudiantes. En este apartado analizaremos estos dos puntos.

Errores en las tablas y gráficos producidos

Identificar los errores cometidos en tablas y gráficos, así como tratar de rastrear el razonamiento detrás de los que son correctos es de nuestro interés porque la construcción e interpretación de tablas y gráficos estadísticos es parte importante de la cultura estadística definida por Gal (2002). Además, Shaughnessy (2007) sugiere que muchos de los errores descritos en las investigaciones sobre comprensión de gráficos enmascaran errores en la comprensión de la idea de distribución.

Los resultados de los errores en los gráficos encontrados en nuestro estudio fueron publicados parcialmente en Batanero, Arteaga, y Ruiz (2009; 2010).

En su tesis doctoral Arteaga (2011) hizo un estudio exhaustivo de estos errores. Sin llegar al nivel de detalle del autor, usamos a continuación su clasificación general en tres tipos de gráficos y lo ampliamos a las tablas:

- ❖ *Gráficos o tablas básicamente correctos*: En esta categoría agrupamos los gráficos que no tienen ningún error, aunque algún detalle los pudiera distanciar de los gráficos correctos. Se incluyen los gráficos o tablas no habituales, que suponen maneras originales de representar los datos por parte de los estudiantes. También, gráficos convencionales a los que los estudiantes han creído apropiado añadir líneas de unión innecesarias. Esto se puede deber a la confusión de los gráficos estadísticos (principalmente del polígono de frecuencias) con los gráficos usados en la matemática determinista en los que se unen las líneas de los puntos dibujados.
- ❖ *Gráficos o tablas parcialmente correctos*. El estudiante tiene errores en los ejes, título confuso, escala, sistema de coordenadas elegido o bien las etiquetas de la escala o de los ejes. Un ejemplo es usar escalas no proporcionales, representación errónea de los números naturales en la recta real o barras no centradas.
- ❖ *Gráficos o tablas incorrectos*. En esta categoría se incluyen gráficos con errores que están vinculados con el manejo de las variables estadísticas, por ejemplo que no son apropiados para el tipo de variable representada, presentan variables no relacionadas dentro de un mismo gráfico o tabla o que intercambian el valor de las variables con la frecuencia.

Arteaga (2011) no encontró relación entre los errores cometidos y la variable estadística representada, pero la mayoría de los errores que reporta están relacionados con el manejo de las variables estadísticas, tales como: confusión entre el valor de la variable y la frecuencia, confusión entre la variable dependiente e independiente en la distribución de frecuencias o asociar la idea de distribución a un conjunto de variables y no a una sola, elección adecuada de un gráfico al tipo de variable que se esté representando y los valores que tome, así como de la naturaleza de las variables representadas. En nuestro estudio también encontramos estos errores.

En la Tabla 6.24 anotamos el número de estudiantes que hicieron tablas y gráficos en cada una de las categorías anteriores. A primera vista salta que el porcentaje de errores cometidos en las tablas fue mucho menor que en los gráficos (6,5% contra

27,3%), lo cual es indicador de la mayor dificultad que acarrea hacer estos últimos de manera correcta. En general los errores vinculados con el uso de tablas se refieren a confusión entre frecuencia relativa y porcentaje o a la confusión entre el valor de la variable y la frecuencia absoluta. Ejemplos de estas tablas las analizamos en la sección 7.2.1.

Tabla 6.24. Clasificación de tablas y gráficos por su corrección

	Total	Correcto	Parcialmente correcto	Incorrecto
Construyen gráficos	88	40	24	24
Construyen tabla	62	40	18	4

Algunos de los errores no relacionados con la variable aleatoria o estadística que se presentaron en nuestro estudio fueron barras del diagrama no centradas en los valores del eje-*x*, o la escala del eje-*y* no homogénea, falta del título del gráfico o no se indica la variable representada en el eje-*x*.

También se presentaron errores por la conceptualización de la variable, por ejemplo, cuando el gráfico no es apropiado para la variable representada o bien grafican juntas dos que no se pueden comparar. Esto último indica que no comprenden el propósito de un gráfico conjunto ni discriminan las situaciones en que dos variables estadísticas son o no comparables. Un ejemplo lo tenemos en la Figura 6.27 en donde el estudiante representa conjuntamente el número de caras, número de rachas y longitud de la racha más larga en tres polígonos de frecuencias sobre el mismo eje cartesiano. Los polígonos están incompletos al no unir los extremos con el eje de ordenadas.

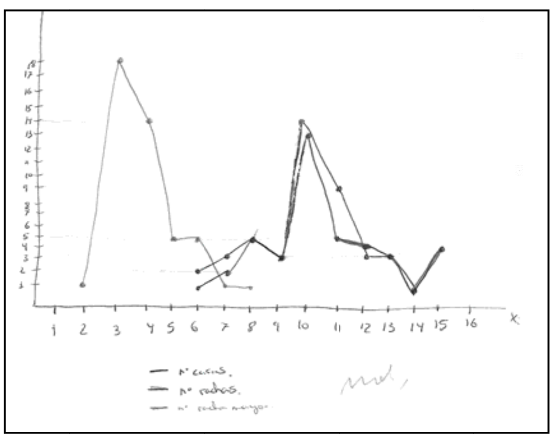


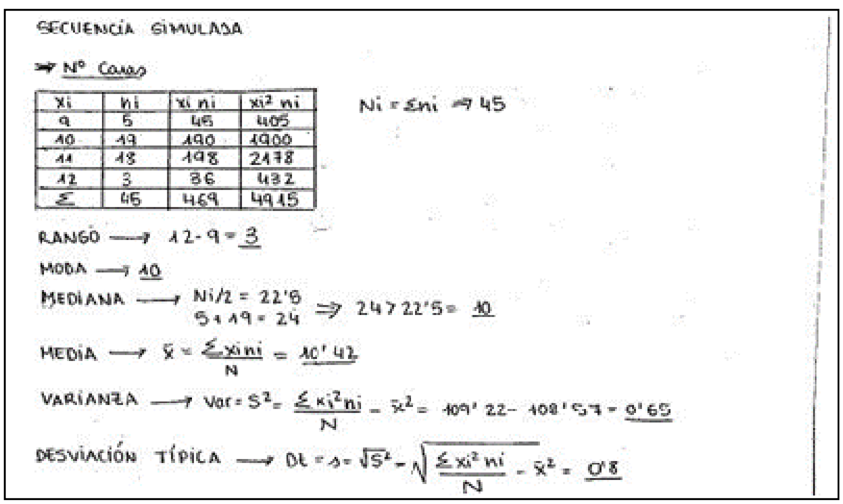
Figura 6.27. Ejemplo de estudiante que representa tres variables no comparables

Unos pocos estudiantes intercambiaron los ejes en los gráficos, representando las frecuencias en el eje-*x* y los valores de la variable en el eje-*y*. Estos estudiantes confunden la variable dependiente e independiente en la distribución de frecuencias de la variable estadística mostrando un error similar al que ya analizamos en la entrevista clínica (Capítulo 5) donde las dos chicas entrevistadas confunden la variable dependiente con la independiente en la distribución de probabilidad y ya descrito en Ruiz (2006).

Notación utilizada y cálculo de estadísticos

Fueron muchos los alumnos (ver ejemplos en la Figura 6.28) que muestran un buen conocimiento y uso de las fórmulas, e incluso de las fórmulas de cálculo abreviado y la disposición auxiliar para realizar los cálculos. Sin embargo, el uso de los símbolos matemáticos varía en formalización. Mientras que el alumno de la Figura 6.28A, emplea con propiedad el símbolo de sumatoria y los subíndices, así como los de media y otros estadísticos, el segundo (Figura 6.28B) emplea sólo símbolos numéricos. En este último, hay, además, un uso incorrecto de la igualdad en el reporte del cálculo de la mediana, pues lo que se está calculando ahí es la mitad de la suma de frecuencias y no el valor en sí mismo de la mediana.

A)



SECUENCIA SIMULADA

→ Nº Casos

xi	ni	xi ni	xi² ni
9	5	45	405
10	19	190	1900
11	18	198	2178
12	3	36	432
Σ	45	469	4915

$Ni = \Sigma ni \rightarrow 45$

RANGO → $12 - 9 = 3$

MODA → 10

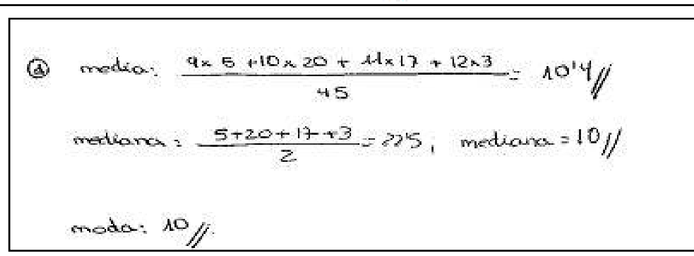
MEDIANA → $Ni/2 = 22'5$; $5 + 19 = 24$ ⇒ $24 > 22'5 = 10$

MEDIA → $\bar{x} = \frac{\Sigma xini}{N} = 10'42$

VARIANZA → $Var = S^2 = \frac{\Sigma xi^2 ni}{N} - \bar{x}^2 = 109'22 - 108'57 = 0'65$

DESVIACIÓN TÍPICA → $DT = s = \sqrt{S^2} = \sqrt{\frac{\Sigma xi^2 ni}{N} - \bar{x}^2} = 0'8$

B)



media: $\frac{9 \times 5 + 10 \times 20 + 11 \times 17 + 12 \times 3}{45} = 10'4$

mediana: $\frac{5+20+17+3}{2} = 22'5$, mediana $= 10$

moda: 10

Figura 6.28. Ejemplo de cálculo de estadísticos con uso de notación matemática

En algunos otros casos, como el ejemplo que reproducimos en la Figura 6.29A, no hay uso de símbolos, por lo que no podemos deducir si el alumno conoce las fórmulas de cálculo. En el ejemplo B es evidente el uso de hoja de cálculo, pero en el caso A no podemos saber cómo ha obtenido los estadísticos. Tampoco nos queda claro si el alumno comprende las fórmulas en el caso de calcularlas mediante ordenador, pues en Excel las funciones dan los estadísticos automáticamente.

A)

nº de caras:	6	7	8	9	10	11	12	13	14	15	media: 10'5
nº de repet:	1	2	5	3	11	9	3	3	1	4	mediana: 10
fr. abs. acum:	1	3	8	11	22	31	37	40	41	45	moda: 10

B)

	MEDIA	MODA	MEDIANA	VARIANZA
Nº CARAS	10,6190476	9	10	5,31475029
Nº RACHAS	11	12	11	3,56097561
RACHA MAYOR	4,35714286	4	4	2,23519164
Nº CRUCES	9,38095238	11	10	5,31475029

Figura 6.29. Cálculo de estadísticos que no muestran procedimiento

Por el contrario, en otros ejemplos (Figura 6.30) los estudiantes muestran comprensión de las ideas de variable, frecuencia y del cálculo de medias ponderadas, al explicitar estas fórmulas en la solución aportada. Notamos en este ejemplo una confusión terminológica entre los conceptos de frecuencia absoluta y acumulada, aunque el uso de las frecuencias absolutas (que se denomina "veces repetidas") al ponderar en el cálculo de la media es correcto. Deducimos que la frecuencia acumulada, únicamente fue usada para calcular el total en el cálculo de la media.

nº de caras →	9	10	11	12
veces repetidas →	5	20	17	3
frecuen. absoluta →	5	25	42	45

Media $\dfrac{9\times5+10\times20+11\times17+12\times3}{45} = \boxed{10'4}$

Figura 6.30. Ejemplo de cálculo de estadísticos

7.4.5. Interpretación de resultados en la realidad

Además de su trabajo matemático con las tablas, gráficos y promedios correspondientes, los estudiantes tenían que trasladar sus resultados al contexto del problema (traducir estos resultados a lo que indican respecto de las intuiciones de los estudiantes) porque, como sugieren Wild y Pfannkuch (1999), «…el objetivo fundamental de la investigación estadística es el aprendizaje en el ámbito del contexto de un problema real» (p. 244). De ahí la importancia de la modelación.

Para contestar la pregunta planteada en el proyecto, los estudiantes forzosamente tenían que efectuar una comparación apropiada entre las distribuciones de las variables, que implicaba una inferencia estadística informal. Es decir, comparar las distribuciones obtenidas mirando la distribución de probabilidades generalizando la distribución de frecuencias y los estadísticos como una aproximación a los parámetros de la variable aleatoria. La interpretación de las comparaciones de las tres variables se tenía que integrar para emitir una conclusión sobre las intuiciones del grupo. Así, dividimos este trabajo de interpretación de resultados por parte de los estudiantes en tres apartados, la interpretación que hacen de los gráficos construidos, la comparación entre las variables haciendo uso de los estadísticos calculados y la emisión de una conclusión que dé respuesta a la pregunta sobre si la clase en conjunto tiene o no buenas intuiciones sobre la aleatoriedad.

Interpretación de gráficos

Como se ha indicado, 88 estudiantes produjeron algún tipo de gráfico. En la Tabla 6.25 clasificamos su trabajo de acuerdo a la interpretación que hacen de su gráfico para vincularlo con la «realidad», es decir, con el problema planeado. Algunos resultados previos fueron publicados en Arteaga, Batanero y Ruiz (2009; 2010) y Arteaga (2011) los analizó con mayor detalle.

Tabla 6.25. Clasificación de estudiantes según interpretación de los gráficos

Interpretación del gráfico	
Correcta	29
Parcialmente correcta	37
Incorrecta o no interpreta	22
Total	88

Destacamos, en primer lugar que una cuarta parte de los futuros profesores se limita a producir el gráfico sin incluir ninguna conclusión de él o bien da una interpretación errónea no siendo capaz de vincularlo con el contexto del problema.

Es posible que una parte de los estudiantes que no leyeron su gráfico, lo hayan incluido en su reporte sólo como un efecto del contrato didáctico, pues están acostumbrados a que en la clase de estadística se presenten gráfico o hagan cálculos, pero no han comprendido la importancia que éstos tienen. Es decir, no hacen un esfuerzo para que la elaboración de la gráfica les aporte alguna comprensión sobre el tema. Aparentemente estos estudiantes no tienen los niveles mínimos de lectura de gráficos, ni puedan completar el ciclo de modelación (Dantal, 1997).

Las interpretaciones incorrectas, en general, se deben a errores conceptuales. Un ejemplo de interpretación incorrecta se presenta a continuación:

> En el caso de las gráficas he optado por realizar un diagrama de barras en el cual se puede observar que una secuencia simulada la frecuencia absoluta que más se repite se encuentra en la tabla de racha mayor, siendo el valor 3 y la frecuencia 26 [*Alumno SL*].

El estudiante SL usa un lenguaje poco preciso al hablar de la «frecuencia absoluta que más se repite» para referirse a la mayor frecuencia e indirectamente a la moda. Sin embargo, su error más grave es que compara la distribución de las tres variables vinculadas a la secuencia simulada al mismo tiempo, incluso cuando había representado cada una de las variables por separado en un diagrama de barras. Es decir, aparece un conflicto en la idea de moda, pues calcula la moda (valor de mayor frecuencia), como si todos los datos provinieran de una misma variable y no para cada variable por separado. Así mismo, se observa una incomprensión del objetivo del proyecto y de la función de las variables que está utilizando.

En el siguiente ejemplo, la interpretación errónea se debe a errores conceptuales, relacionados con la dependencia funcional lineal. No queda claro qué es dependencia lineal para este estudiante ni la importancia de que pase de «– a +». Además de que también hay una incomprensión del objetivo del problema puesto que establece una comparación entre todos los datos, en lugar de comparar sólo los de aquellas variables que sí son comparables.

> Todos los gráficos presentan una dependencia funcional lineal de – a + o viceversa, según las circunstancias [*Alumno PF*].

Un 42% de los estudiantes hacen una interpretación parcialmente correcta de los gráficos. La mayor proporción de ellos se limitan a representar los datos uno a uno, sin llegar a formar la distribución de la variable estadística, lo cual resulta desventajoso porque aunque pueden observar la variabilidad de las variables, sólo lo hacen de forma parcial o con inexactitud.

Algunos estudiantes muestran una gran imprecisión del lenguaje, confundiendo conceptos. Por ejemplo, en el caso siguiente, el estudiante ha representado los datos de sus compañeros individualmente en dos diagramas de líneas sobre un mismo gráfico, como se muestra en la Figura 6.22, y concluye lo siguiente sobre su gráfico:

> Cuando comparamos los resultados obtenidos en las secuencias simuladas y en las secuencias reales, observamos que las secuencias simuladas reflejan líneas de tendencia menos pronunciadas que las de las secuencias reales. En el caso de las caras y las cruces suponemos que hay un 50% de probabilidades de que salga una u otra, por tanto las simulaciones son todas en torno a 10 caras. Sin embargo en el ejercicio real el

> número de caras y de cruces varía mucho de una experiencia a otra. Como el número de experimentos es relativamente bajo, creemos que si se aumenta, el número de caras y de cruces se irá haciendo cada vez más equitativo [*Alumno AR*].

El estudiante compara las dos variables estadísticas referidas al «número de caras» y detecta una mayor variabilidad en las secuencias reales. Usa implícitamente la idea de variable aleatoria al mencionar las probabilidades y también ha comprendido el fenómeno de convergencia, es decir, relaciona la asignación frecuencial y clásica de probabilidad. Sin embargo, la convergencia a que él alude se dará en la frecuencia relativa de número de caras y no en la absoluta. Observemos que en su argumentación alude poco a la frecuencia relativa y cuando lo hace, de inmediato concreta en la frecuencia absoluta, por lo que, más que un error de lenguaje, podría estar confundiendo estos conceptos. Una confusión parecida, entre la frecuencia relativa y la absoluta, también la observamos en la entrevista clínica (Capítulo 5). Lo cual nos indica que la frecuencia relativa no es un concepto simple para los estudiantes.

El alumno AG también hace referencia a las dos variables estadísticas, pero se limita a la interpretación de las medidas de tendencia central:

> El sistema de representación de datos utilizado es el diagrama de líneas ya que se pueden ver las diferencias entre la secuencia simulada y la real; si la línea asciende, la frecuencia aumenta. El punto más alto del diagrama corresponde al valor de la variable que tiene mayor frecuencia, como ocurre cuando… [*continúa con un ejemplo; Alumno AG*].

Se observan puntos de interpretación acertada del gráfico, puesto que relaciona el crecimiento y decrecimiento en la gráfica con el aumento o disminución en la frecuencia, así mismo detecta correctamente a la moda y justifica la elaboración de su gráfico. Sin embargo compara las modas y porcentajes de valores pero no se ocupa de la dispersión.

Encontramos también interpretaciones correctas de los gráficos, algunas de ellas en estudiantes que sólo han trabajado representando sus propios datos o listados de datos sin formar la distribución. Por ejemplo, el alumno CG une los datos del total de los alumnos para el número de caras y número de cruces y lo representa en un sector circular, después de haber representado los datos del número de caras alumno a alumno en un diagrama de barras semejante al mostrado en la Figura 6.22A, realiza el siguiente comentario:

> Si observamos en el sector circular, nos daremos cuenta que los alumnos han puesto al azar más caras que cruces y que en la secuencia real dan unos resultados casi iguales en el porcentaje de caras y cruces [*Alumno CG*].

Aunque la interpretación del gráfico de la variable «cara o cruz» en sí misma es correcta, el alumno no entendió la finalidad del proyecto, puesto que no consideró el análisis del resto de las variables. Para CG la aleatoriedad de las secuencias se evalúa al comparar el equilibrio entre el número de caras y el número de cruces simuladas con el valor esperado en el experimento Bernoulli. De manera que aunque analiza correctamente esa variable, sus conclusiones sobre las intuiciones de su grupo estarán incompletas. Este estudiante no visualizó la importancia de evaluar la aleatoriedad del lanzamiento de una moneda a través del experimento compuesto por 20 lanzamientos, por lo tanto no podrá interpretar la aleatoriedad como una multitud de modelos.

Una tercera parte de los estudiantes conjuga las ideas de promedio y dispersión detectando las tendencias y analizando la estructura de los datos, visualizándolo a partir de la gráfica y no sólo a partir de los números. Esto es mayor en la proporción de los alumnos que representaron gráficamente la distribución de alguna de las variables, ya sea en gráficos separados o conjuntos. El siguiente es un ejemplo de este tipo de argumentación:

> Para comentar las diferencias, el gráfico aporta más posibilidades, pues visualmente podemos percibir las diferencias. Las tablas necesitan de una interpretación de los datos. Centrándome en el gráfico observamos que los valores de la frecuencia simulada se concentran en torno a unos mismos intervalos 10 y 11, mientras que en la frecuencia real, los valores están más disgregados, aunque también vemos que se concentran en un núcleo 8-15, fluctuando al alza y a la baja. De esta forma en la frecuencia real comprobamos que 10 es la secuencia con mayor número de veces saliendo cara, tanto en la simulada (20) como en la real (14).... la horquilla de la secuencia real es mucho más amplia que la de la simulada [*Alumno FL*].

En primer lugar, el alumno reconoce que le es más sencillo interpretar el gráfico que la tabla. Compara correctamente el comportamiento de la distribución de las dos variables estadísticas a través de los intervalos modales y diferencia la dispersión entre ambas. Aunque en toda su argumentación usa un lenguaje impreciso.

En este apartado encontramos que no es suficiente que los estudiantes hayan construido sus gráficos para que los interpreten o lo hagan correctamente. Aquellos estudiantes que lograron establecer la distribución de las variables gráficamente tienen un mejor desempeño en su interpretación, pero no todos estos estudiantes tuvieron una correcta interpretación. Tampoco fue cierto que todos los estudiantes que graficaron los datos uno a uno tuvieron una interpretación incorrecta de su gráfico, aunque fueron los menos los que lo hicieron correctamente. Esto significa que la correcta elaboración del gráfico auxilia en la interpretación de los datos, pero eso no significa que el estudiante comprenda cuál es la forma en que puede emplearlo para resolver el problema.

Coincidimos, entonces, con Arteaga (2011) en que la interpretación de gráficos es difícil para los estudiantes, a pesar de es una herramienta indispensable en su vida profesional.

Comparación entre las distribuciones: uso de los estadísticos

La comparación entre las distribuciones de las variables de la secuencia real y la simulada se realizó principalmente haciendo uso de los estadísticos, sin embargo, una parte de los estudiantes la realizaron haciendo uso de otras herramientas, como valores aislados, la forma de la gráfica, o no hicieron tal comparación. En la Tabla 6.26 mostramos que, aunque los estudiantes calculen los estadísticos, no siempre los emplean, limitándose en muchos casos a presentarlos sin ningún comentario respecto a su significado o las diferencias entre las dos secuencias. A este respecto Konold, Pollatsek, Well y Gagnon (1997) indican que para los estudiantes no es siempre intuitivo el uso de los promedios y dispersión para comparar dos distribuciones, por lo que es posible que muchos considerasen, por efecto del contrato didáctico, que la tarea estaba completa una vez calculados los estadísticos.

En primera instancia nos enfocamos a las medidas de tendencia central. 36 (32,4%) estudiantes comparan las medias (32 de ellos, correctamente), 18 (16,2%) las modas (12 correctamente) y doce (10,8%) las medianas (dos incorrectamente). Como notamos, en su mayoría los estudiantes tienen un mejor manejo de la media que del resto de las medidas de tendencia central y también es a la que más recurren, pero de manera general, los estudiantes que hicieron uso de las medidas de posición central para comparar las distribuciones, lo hicieron correctamente.

Tabla 6.26. Interpretación de estadísticos en la comparación de las distribuciones (n=111)

Tipo de comparación	Frecuencia
Comparan media	32
Comparan moda	18
Comparan mediana	12
Comparan dispersión	30
Comparan valores aislados	10
Comparan forma gráfica	12
Comparan solo sus datos	11
Indican lo que esperan	12
No comparan	23

En los siguientes ejemplos haremos notar la forma en que realizaron tales comparaciones, en el primero se hace uso de las modas, en el segundo, de las medias y en el tercero, de las medianas:

> Pasamos al gráfico del número de rachas, en la secuencia real el 10 es el que más se repite; en cambio en la secuencia simulada el valor más repetido es el 12 seguido del 13 [*Alumno MC*].

> Con respecto al número de caras en la secuencia simulada la media está en 10,41, es decir, que de 20 veces que hemos tirado la moneda 10 veces para redondear ha salido cara. Muy igualado está el caso de la secuencia real en que la media del número de caras está en 10,76, 11 en números redondos [*Alumno CG*].

> La mediana del número de caras y del número de rachas de la secuencia simulada y la real también coinciden [*Alumno CL*].

Algunos estudiantes presentaron los estadísticos pero no los comparan (23, es decir, 20,7%), o indican lo que esperan sin comparar (12, el 10,8%), siguiendo sus creencias previas. Por ejemplo, en el siguiente caso, el alumno calculó la media, mediana y moda, pero no las usa al efectuar la comparación. Él se limita a exponer sus creencias sobre lo que debe ocurrir: piensa que no puede haber coincidencia, pues muestra una concepción de la aleatoriedad errónea como falta de control (Langer, 1975):

> Respecto a la conclusión podemos decir que en un primer momento los alumnos intentan que salga el número más parecido de caras y cruces, cosa que es algo que no lo puedes lograr [*Alumno IA*].

En los dos siguientes ejemplos, aunque se comparan los valores de las variables, la idea de aleatoriedad como imposibilidad de predicción, hace que el alumno desista de su comparación. Estos estudiantes no han comprendido la utilidad de la estadística para analizar datos experimentales ni que ello les proporcionaría información sobre la situación, porque no visualizan la utilidad de una predicción estadística a través de una distribución de valores y no de valores aislados.

> En un principio es factible que el nº de caras y de cruces a priori debería ser el mismo, pero debido a que esto es un proceso aleatorio, las secuencias de caras y cruces en realidad es imprevisible [*Alumno ES*].

> Al comparar los gráficos no se parecen mucho, ya que en la simulada lo que se hacía era inventarse los datos y no sabes cómo pueden salir los datos en realidad [*Alumno AG*].

Otros alumnos comparan sólo sus propios datos (11, 9,9%), tratando de evaluar únicamente sus propias intuiciones. Se limitan a comparar el valor de las diferentes variables en las dos secuencias, sin hacer referencia a los estadísticos. Generalmente, manifiestan que una buena intuición supone poder acertar el experimento, es decir, tienen una concepción errónea sobre la aleatoriedad, cayendo en la ilusión de control:

> Secuencia simulada: Longitud de la racha más larga=5, número de rachas: 0; número de caras=10. Secuencia real: Longitud de la racha más larga=3; número de rachas: 15; número de caras 11; he acertado 13 veces [*Alumno EL*].

> Los datos obtenidos son parecidos; el número de caras en el lanzamiento simulado es de doce mientras que en el real es de nueve; sólo se llevan de diferencia tres caras. Lo mismo ocurre con la longitud de la racha más larga que sólo se diferencia en uno y el número de rachas que es de dos la diferencia [*Alumno SL*].

Unos pocos estudiantes compararon directamente la forma de la gráfica (12), como el siguiente caso, en el que se ha calculado la media, mediana y moda, pero sólo compara directamente la gráfica de la distribución. La comparación resulta certera, al observar, de manera general, la diferencia en la variación de las distribuciones. En el segundo caso, la observación del gráfico no le auxilia a ver la distribución como un todo, sino que compara valor a valor:

> Al comparar los gráficos, podemos observar que el número de caras de la simulación real está mucho más repartido que en la simulación, cuyos datos están más centrados en los números 10 y 11 [*Alumno MC*].

> En los gráficos que se han obtenido al representar las rachas que resultaron de la actividad realizada por los alumnos se pueden destacar distintos resultados. Por ejemplo, en la primera gráfica, de la secuencia simulada encontramos cinco posibilidades y en la gráfica de la secuencia real las posibilidades son siete [*Alumno HC*].

Otro ejemplo es el de este estudiante, que usa sus gráficas para comparar entre sí la variación del número de caras y del número de rachas de la secuencia simulada a través del rango. Aunque su percepción de la distribución es como un todo es correcta, se equivoca al comparar los estadísticos de las variables de una misma secuencia:

> Se puede observar que en la secuencia simulada al calcular el rango (r), el valor más disperso se encuentra en el número de caras con respecto a la media (x) también hay más variación en el número de rachas, será porque nos encontramos con valores más diversos [*Alumno SL*].

Algunos (10 alumnos, 9%) comparan valores aislados de la variable o porcentajes de valores aislados (10 estudiantes, 9%); cuatro de ellos comparan únicamente los máximos y mínimos en cada distribución. Esta estrategia de comparación de valores aislados al estudiar la diferencia de dos distribuciones ya apareció en la investigación de Estepa y Batanero (1994), quienes la explican por la existencia de una concepción local de la asociación estadística, consistente en juzgar la asociación entre dos variables considerando tan sólo una parte de los datos y no el conjunto completo.

> El máximo que he encontrado en las dos secuencias es 15 en el número total de caras, 16 el número más alto en rachas y 8 el máximo en la racha real [*Alumno EG*].

Sólo 30 estudiantes (27%) utilizan explícitamente las medidas de dispersión (27 de ellos el rango y seis la desviación típica o la varianza). Un ejemplo se reproduce a continuación, en que el alumno es capaz de apreciar la diferencia de dispersión en el número de caras y su semejanza en el número de rachas, aunque su redacción no es muy buena:

> En primer lugar compararé los datos del número de caras en ambas secuencias, donde destacaré las diferencias que se aprecian principalmente entre el rango, la varianza y, consecuentemente la desviación típica, de hasta seis veces más de diferencia la secuencia simulada con respecto a la real [*quiere decir lo contrario, pues la dispersión es mayor en la secuencia real*]. En segundo lugar destacaré los datos del número de rachas de las secuencias simuladas y la secuencia real. He de decir que son sorprendentes las similitudes del rango e incluso la varianza y la desviación típica que tan sólo llegan a diferenciarse en décimas [*Alumno LC*].

Otros alumnos tienen imprecisión en el lenguaje al confundir variable y valor de la variable, como en el caso siguiente, que también compara la dispersión; de hecho la desviación típica calculada es positiva, aunque el estudiante indica lo contrario:

> Se observa que en las secuencias reales hay más variedad de variables que en las simuladas. Aunque la desviación típica es más pequeña en la secuencia simulada, siendo inferior a cero [*Alumno JC*].

En el caso siguiente se confunde frecuencia y valor de la variable, error también encontrado por Cobo (2003) en el cálculo de la media. Como vemos en el ejemplo, el alumno se refiere a la frecuencia, pues al ser sólo 20 los lanzamientos, la longitud de la racha no puede ser igual a 25.

> En la secuencia simulada la racha mayor y en la que más ha coincidido la clase es 25 [*Alumno TF*].

En resumen, un porcentaje importante de estudiantes de la muestra ha calculado algún estadístico, principalmente estadísticos de posición central (media, mediana y moda) y en menor medida de dispersión. En general, aunque el cálculo es correcto, con pocas excepciones (ver Punto 7.4.4), y el de los promedios muy abundante, la mayoría de los estudiantes se limita a calcularlos, pero no los interpreta ni los usa para la comparación, aunque en caso de compararlos, la comparación, en general, es correcta. Menos aún usan la idea de dispersión, aunque el cálculo del rango lo hacen correctamente, parece que, aunque se comprende el procedimiento de cálculo, los alumnos no llegan a captar el significado de la dispersión ni su utilidad en la comparación de dos distribuciones.

Así mismo observamos que encontramos estudiantes que sólo se dejan llevar por lo que creen que va a pasar sin hacer uso de los datos ni de la estadística.

Concluimos que la idea de distribución no es comprendida de manera completa por los participantes en el estudio, pues aunque la mayoría llega a formar la distribución, a representarla, bien mediante una gráfica o mediante una tabla o ambas y aunque la mayoría calcula los promedios, sólo unos pocos basan la comparación de las distribuciones en los promedios y casi ninguno tiene en cuenta la dispersión, que, de acuerdo con Reading y Shaughnessy (2004) es fundamental para comparar dos distribuciones.

También notamos 35 estudiantes que no efectúan tal comparación (indican lo que esperan o no comparan), por lo tanto, un número relativamente alto de estudiantes (76, el 68,5%) hacen (o intentan hacer) la comparación entre las distribuciones simuladas y reales aunque algunos la realizan mal. Esto nos indica que hubo un porcentaje importante de estudiantes con una comprensión parcial del objetivo del proyecto, puesto que era necesario realizar una comparación para obtener una conclusión sobre las intuiciones del azar del grupo.

Obtención de una conclusión, respuesta a la pregunta del proyecto

Aunque los alumnos completaron con mayor o menor éxito los otros pasos de la actividad de modelación, pocos han sido capaces de completar el último paso emitiendo una conclusión en el contexto del problema. Este punto fue estudiado por Arteaga (2011) para los alumnos que elaboraron gráficos, pero él lo relaciona con la complejidad semiótica de los datos y no con la actividad de modelación como lo haremos nosotros.

Interpretar los resultados del análisis de datos en términos de la pregunta planteada (¿son buenas las intuiciones de los estudiantes?) fue el más complicado para los estudiantes; posiblemente por su falta de familiaridad con proyectos estadísticos y actividades de modelación. Sin embargo, algunos estudiantes, parecen haber comprendido la necesidad de este último paso en el proceso de modelación pues explícitamente indican la utilidad de la estadística para analizar los datos del experimento y obtener una conclusión:

> La estadística nos permite ver si nos acercamos o no a los resultados, si intuimos y prevemos bien [*Alumna HC*].

En la Tabla 6.27 clasificamos las conclusiones obtenidas por los estudiantes. En ella observamos que la conclusión es la tarea más difícil para los estudiantes, puesto que sólo un poco más de la cuarta parte (27%) obtienen una conclusión por lo menos parcialmente correcta. Al igual que en los resultados de la entrevista, pareciera que a los estudiantes les resulta más sencillo descontextualizar (pasar de la realidad al modelo matemático) que la operación inversa (contextualizar, pasar del modelo a la realidad).

Tabla 6.27. Clasificación de estudiantes, según su conclusión final

Conclusión	Frecuencia
Correcta	4
Parcialmente correcta	26
Incorrecta o no concluye	81
Total	111

Estudiantes que concluyen incorrectamente o que no lo hacen

Fueron muchos los estudiantes que no conectaron los resultados del trabajo matemático con el modelo en la situación (73%), es decir, no vieron las aplicaciones de lo obtenido en el análisis estadístico sobre las intuiciones de los estudiantes.

Algunas de las conclusiones más alejadas, simplemente comentaban cuestiones que no tenían que ver con la pregunta que debían contestar ni hacían referencia al análisis estadístico que habían efectuado. Las conclusiones de SG y JOCH son ejemplos que muestran que para algunos estudiantes emitir una conclusión fue sólo una forma de cumplir con el contrato didáctico con el profesor. Este tipo de comentarios, podrían ser valiosos si hubieran sido acompañados de una respuesta a la pregunta del proyecto o de un análisis más profundo de sus resultados, pero se presentaron aislados. Comentarios como el de JOCH, denotan una revisión bibliográfica que no citaron y que se vincula con el propósito del proyecto solo de manera tangencial:

> El diseño de experimentos aleatorios en el aula, no solo permite su realización de manera sencilla y rápida sino que hace necesaria la comprensión previa del problema, de las herramientas disponibles, de la manera de hacerlos y el interpretar los resultados obtenidos. [*Alumno SG*].

> La comprensión de estas nociones no es inmediata y tiene cierta dificultad para los niños, incluso para los adolescentes y adultos… Los niños se suelen dejar guiar por sus creencias y experiencias subjetivas, considerando, por ejemplo, que un suceso es imposible porque a él nunca le ha ocurrido [*Alumno JOCH]*.

Dentro de este grupo, son muchos los estudiantes (23) de la muestra que anteponen sus ideas sobre el contexto al análisis de datos. Así por ejemplo, el alumno EL, no comprendió que un modelo matemático puede aplicarse a una situación aleatoria. Eso hizo que negara la evidencia de los datos y se quedaran con sus ideas prestablecidas sobre el contexto. Esta concepción errónea también la detectamos en el estudio de la entrevista (Estudio 3, Capítulo 5):

> Comparando los datos me he dado cuenta que son muchos los resultados entre los compañeros los que coinciden, pero aun así, sigo pensando que es mera casualidad, porque en la simulada hemos puesto lo que hemos querido [*Alumno EL*].

Observemos que aun cuando logró conectar el modelo que obtuvo con el contexto del problema, falla en la obtención de conclusiones debido a que supone que una buena intuición ha de corresponder a obtener los mismos resultados en las secuencias reales y simuladas (concepción errónea de aleatoriedad). Desde distintas perspectivas Wild y Pfannkuch (1999), Makar, Bakker y Ben-Zvi (2011) y Chaput, Girard, y Henry (2011), han mostrado la importancia que los estudiantes dan al contexto al trabajar con un modelo. Makar, Bakker y Ben-Zvi se han preocupado por analizar más detenidamente el conocimiento del contexto y la forma en que éste influirá en la forma en que se analicen los datos. Chaput, Girard, y Henry (2011) han remarcado las dificultades que tienen los estudiantes al trabajar con un modelo pseudo-concreto en lugar de con la «realidad». En estos otros ejemplos (alumnos GEE, LG y AG) es notorio como mezclan el análisis estadístico (la evidencia de los datos) con sus creencias (conocimiento del contexto) para emitir una conclusión.

> En los primeros alumnos, aparecen unos datos imposibles. En general, podemos decir que los alumnos han tenido una visión de aleatoriedad más elevada en la situación simulada que en la real [*Alumno GEE*].

> Partiendo del estudio de los gráficos, se podría decir que en valores absolutos, la previsión del grupo no ha sido demasiado desafortunada. Un juego de azar es imposible de prever con total exactitud, pero las aproximaciones sumadas a los aciertos son mayores a las previsiones muy alejadas del resultado real [*Alumno LG*].

> ...al comparar las gráficas no se parecen mucho, ya que en la simulada lo que se hacía era inventarse los datos y no sabe cómo pueden salir en la realidad [*Alumna AG*].

El alumno GEE antepone totalmente sus creencias a los datos, concluyendo que es más aleatoria la secuencia simulada que la real. Los estudiantes LG y AG muestran una concepción correcta del azar (no se puede prever) y otra incorrecta (evaluar el número de coincidencias o la diferencia de valores obtenidos en cada estudiante), en lugar de comparar directamente las distribuciones de las variables o el comportamiento general del gráfico. Es decir, comparan caso a caso sin utilizar la distribución o sus estadísticos al hacer la comparación.

Como analizamos en la Sección 7.1, aunque los estudiantes de la muestra están familiarizados con el experimento «lanzar una moneda» no lo están con el experimento «lanzar 20 monedas», que sería parte del contexto, ni son conscientes de sus propias intuiciones incorrectas (que sería la otra parte del contexto del proyecto planteado). Su falta de conocimiento sobre el contexto podría influir en que el análisis estadístico se vea mermado.

En nuestro proyecto, esta relación entre el contexto del problema y el análisis estadístico resultó conflictiva, puesto que al mismo tiempo que se les pretendía mostrar a los estudiantes las ventajas del análisis estadístico[11], también se les quería concientizar sobre sus intuiciones incorrectas sobre la aleatoriedad. Sin embargo, al no tener una buena perspectiva sobre la utilidad de la estadística, desechan los resultados matemáticos y se guían por sus intuiciones, sin que el proyecto les permita darse cuenta de que éstas son erróneas ni les ayude a tener una idea más global de la estadística.

En otros estudiantes faltó capacidad de trabajo con el modelo matemático para detectar las diferencias entre las secuencias inventadas y reales. Esto, como se observa en el siguiente ejemplo, los llevó a concluir que los resultados en las dos secuencias eran similares:

> Como conclusión podemos ver que los resultados son prácticamente los mismos tanto en la real como en la simulada, de lo que podemos deducir que lo real y lo simulado es muy parecido así que lo que te inventes puede ser prácticamente igual a lo real [*Alumno AC*].

En este ejemplo, el alumno tampoco llega a relacionar su análisis con la intuición de los estudiantes, lo cual significa que no vinculó las variables trabajadas con la respuesta a la pregunta planteada. En su comentario ni siquiera menciona a qué variable se refiere cuando habla de «los resultados».

Algunos estudiantes se plantearon nuevas preguntas para tratar de responder la pregunta planteada en el proyecto. Según Wild y Pfannkuch (1999), esto también es parte de la interacción del contexto con la estadística y, en consecuencia, del razonamiento estadístico. Esto, de hecho, era necesario en nuestro proyecto, puesto que la pregunta sobre las intuiciones sobre el azar, sólo podía ser respondida en términos de las variables que se manejaron. La pregunta original no podía ser respondida con un sí o un no, de modo que los estudiantes debieron haberse preguntado sobre el comportamiento de la distribución de cada variable. En el ejemplo del estudiante BG, se observa cómo transforma la pregunta sobre las intuiciones a una evaluación sobre la variabilidad de los resultados entre sus compañeros, lo que sería equivalente a tratar de analizar la variabilidad de la variable. Sin embargo, él no hace uso de las medidas de

[11] Un análisis estadístico tiene el objetivo de *conocer* en un sentido probabilístico (Laplace, 2006/1814), de modo que un proyecto planteado en situación escolar, también debería enfatizar en esa faceta de la estadística.

dispersión (que, aparentemente, no comprende) sino que se limita a hacer una observación somera de sus gráficos, en donde había graficado los datos de todo el grupo, sin formar la distribución:

> Al comparar las secuencias (simuladas y reales) observamos que no existen muchas diferencias entre unas y otras. Casi toda la clase hemos tenido resultados muy parecidos. Yo creo que esto se debe al azar. Lo mismo que depende del azar que salga cara o cruz también depende del azar que 2 personas pongan los mismos resultados o que yo tenga los mismos resultados simulados que reales [*Alumno BG*].

BG también manifiesta una concepción errónea de la aleatoriedad pues el hecho de que la coincidencia entre estudiantes sea muy grande es precisamente fruto de una intuición global errónea (como hemos analizado en la Sección 7.1.2), ya que la probabilidad de obtener estos resultados en secuencias aleatorias es extremadamente pequeña (de hecho viene dada por el nivel de significación que se calculó en dicha sección al realizar los contrastes de hipótesis correspondientes). Este estudiante, por tanto, también antepuso sus creencias a lo que el análisis estadístico pudiera decirle.

Estudiantes con conclusiones parcialmente correctas

Un segundo grupo de estudiantes (26, el 23,4%) obtiene una conclusión parcial. A continuación mostraremos algunos ejemplos que nos darán cuenta de la diversidad de causas por la que su trabajo con las conclusiones fue considerado dentro de esta categoría.

El comportamiento más llamativo de este grupo es que fueron muchos (14) los que limitaron su conclusión a las intuiciones de los estudiantes sobre las medidas de tendencia central del número de caras. Estos estudiantes muestran una buena comprensión de la idea de esperanza matemática (media de la variable aleatoria) y la discriminan de la media de la correspondiente variable estadística. También han sido capaces de usar los promedios al comparar dos distribuciones, por lo que estarían en el nivel más alto de comprensión de la distribución según Bakker y Gravemeijer (2004), sin embargo no han sido capaces de tomar en cuenta las otras dos variables, ni la variabilidad de esta variable, como se observa en los siguientes ejemplos:

> La intuición de mis compañeros observando la tabla del nº de caras es buena, ya que los valores más repetidos en la secuencia simulada coinciden con los valores de la secuencia real: 10 y 11 son las más repetidas. La media de las dos secuencias es 10, por lo tanto creo que la intuición es buena [*Alumno TG*].

> Según los datos obtenidos y tras representaciones gráficas, podemos observar como la mayoría de la clase tiene buena intuición ya que si comparamos la moda del número

de caras, en la situación simulada de la real, vemos como se diferencia solo en una [*Alumna AJF*].

Algunos de estos estudiantes realizaron un trabajo matemático con el resto de las variables y con las medidas de dispersión, lo cual indica que no es el trabajo estadístico lo que les limitó el uso de la dispersión y el trabajo el resto de las variables, sino una incomprensión de la vinculación de las otras dos variables con la aleatoriedad. Suponemos que creyeron que responder la pregunta planteada, cuando la restringían a las intuiciones sobre el número esperado de veces con que ocurre un suceso. Como hicimos notar en la sección anterior, era necesario que los alumnos refinaran la pregunta general del profesor, demasiado amplia, en preguntas más pequeñas referidas al comportamiento de la distribución de las variables, para luego integrarlo en una conclusión general. No es el objetivo reducir el problema a otro más simple, que más estudiantes son capaces de responder, pero que no es exactamente el planteado.

Otro punto importante es que algunos de estas respuestas son acordes con las intuiciones incorrectas sobre la aleatoriedad de los estudiantes. Así, estos estudiantes analizaron los datos siguiendo el modelo subjetivo que nos fue revelado en la Sección 7.1.2 de este capítulo en el que esperan el mismo número de caras que de cruces, rachas cortas y con ensayos dependientes.

Parcialmente correctas también fueron las conclusiones de estudiantes que siguieron un análisis correcto, pero por alguna razón no finalizan la conclusión. Como los siguientes ejemplos, en donde el estudiante HC tiene una concepción incorrecta de la aleatoriedad y eso hace que, aunque el análisis de datos le muestre una concepción incorrecta del grupo, concluya que la intuición es correcta pues coincide con la suya. En el segundo ejemplo, el alumno MAB hace una interpretación correcta de la variación y el valor más grande de la racha más larga (en cierta medida el rango), pero eso no le hace cambiar su idea de buenas intuiciones sobre la aleatoriedad:

> Respecto a las gráficas construidas sobre el número de caras he de comentar que se parece observar algunos cambios en la secuencia simulada donde encontramos cuatro posibilidades y en la real hay diez posibilidades. Con esto podríamos decir que los alumnos no han sido intuitivos, pero no creo que esto sea así, ya que si nos detenemos atentamente en las gráficas podemos ver que los valores de 9, 10, 11 y 12 en ambas han sido dados por un mayor número de alumnos que en los demás casos [*Alumno HC*].

> … en la secuencia real hay mayor rango de variación de valores, ya que estas discurren desde las 14 a 5 caras, mientras que la simulada de las 12 a las 7 caras. Por todo ello, creo que nuestras intuiciones sobre el número de caras que se obtienen al lanzar 20 veces una moneda, son totalmente correctas… Además, en la real, la racha

> más larga es de 9 y en la simulada de 4, ambas muy distantes. A pesar de ello y en su conjunto, aunque un poco peor que anteriormente, comentar que nuestra intuición con respecto a la aleatoriedad, no nos engaña en absoluto [*Alumno MAB*].

Es de llamar la atención, puesto que los alumnos visualizan, en un caso tanto la diferencia de dispersión como los valores centrales similares, en el otro sólo la dispersión, pero no llegan a concluir correctamente sobre la intuición de los estudiantes porque sus resultados matemáticos no concuerdan con sus creencias. El alumno MAB, además, se limita al análisis de la variabilidad.

En el mismo tenor que los ejemplos anteriores, aunque esta alumna es capaz de ver las semejanzas y diferencias en los promedios y dispersión de las variables, concluye que la intuición es buena. Como en los casos anteriores, pareciera que no puede reconocer que su intuición es incorrecta.

> En la secuencia se puede ver como más o menos son parecidos el número de caras, sólo se da un margen del 8 al 11 con una media de 10,26. En cambio en la secuencia real existe una variedad más extensa del número de caras del 7 al 16, con una media del número de casa de 10,61. Quedando así visible que la intuición y lo real es parecido, con lo cual no se diferencia mucho la realidad de la moneda que nosotros imaginariamente tirábamos [*Alumno AV*].

También se presentaron estudiantes que únicamente se enfocaron a la variabilidad, sin tomar en cuenta las medidas de tendencia central. En su mayoría, estos estudiantes hicieron un gráfico que representa el listado de resultados individuales sin formar la distribución. Como el alumno MM, que reconoce la diferencia de dispersión en las distribuciones, pero interpreta la pregunta del problema como la búsqueda de diferencias entre las intuiciones de los estudiantes y eso hace que no pueda concretar una conclusión sobre las intuiciones del grupo. Su siguiente conclusión se basa en el análisis de su gráfico que representa el listado de resultados individuales:

> Realmente las intuiciones de los alumnos han sido más o menos muy parecidas. No hay muchas irregularidades. Pero cuando las comparamos con sus secuencias reales, se puede ver a simple vista en el gráfico que existen grandes irregularidades. En la secuencia real el número máximo de caras es 16 y el mínimo 7, sin embargo en la simulada el máximo es 13 y el mínimo 8; su recorrido es más pequeño que el real [*Alumno MM*].

Concluye que las intuiciones son similares en distintos estudiantes, con lo cual se observa que aprecia la poca variabilidad existente en el número de caras. También nota la diferencia entre la distribución del número de caras de las secuencias simuladas y reales. Es decir, es capaz de percibir la diferencia de variabilidad y entre máximos y mínimos en su gráfica, pero no llega a relacionar estas diferencias de los resultados de los experimentos con las intuiciones globales de los estudiantes sobre las características del fenómeno aleatorio. Para el tipo de gráfico que este estudiante hizo, su interpretación es correcta, sin embargo, no asoció la diferencia de dispersión en las distribuciones real y simulada con una pobre intuición sobre la variabilidad aleatoria en los estudiantes. Es decir, no pudo descontextualizar la matemática en términos del contexto del problema.

También se presentó que a pesar de comparar correctamente el promedio y la dispersión para las tres variables, se produce un error de comprensión del significado de la medida de dispersión que ocasiona que no se llegue a una conclusión correcta. En el caso que sigue, aun cuando el estudiante aprecia la igualdad de medias y diferencia de rangos, le atribuye la diferencia de rangos a la representatividad de la media en lugar de a la mayor variabilidad de los datos. Por otro lado, supone que los rangos son iguales, a pesar de la diferencia que nota en las desviaciones típicas.

> Las intuiciones de los estudiantes han sido bastante acertadas. Se ha llegado a esta conclusión analizando las 3 variables en las secuencias real y simulada. Analizando el nº de caras observamos que los resultados obtenidos de la moda, la mediana y la media coinciden prácticamente. El rango de la secuencia simulada permite apreciar más uniformidad en la media que en la secuencia real donde el rango es 9. A pesar de ello en ambas desviaciones típicas se ve una gran representatividad de los valores de las medias ya que están muy próximas a cero [*Alumna ER; quien comenta en forma parecida las otras dos variables*].

Como se ha comentado, los estudiantes tienen poca capacidad de comunicación verbal y a veces hacen uso no adecuado de las expresiones matemáticas. Nuestro último ejemplo en la emisión de conclusiones parcialmente correctas, muestra imprecisión del lenguaje, lo que no le ayuda a obtener una conclusión certera:

> Hemos encontrado que tras realizar las gráficas correspondiente al experimento, llegamos a la conclusión de que la tabla de experiencia simulada tiene un número de valores más parecidos entre ellos que los de las tablas reales, donde los números son más diversos. Consideramos entonces que nuestra imaginación tiene limitaciones, ya que nos imaginamos cosas que no son esperadas [*Alumno AR*].

Este estudiante hace referencia a las limitaciones de las intuiciones (las llama «imaginación») con lo que relaciona el análisis de datos con las intuiciones de la clase. Por lo tanto no sólo fue capaz de detectar la diferencia en variabilidad de las dos distribuciones que compara sino también de contextualizar e interpretar estos resultados en términos de la pregunta pedida. Su respuesta es sólo parcialmente correcta porque no hace referencia a los promedios, únicamente a la dispersión y con un lenguaje muy impreciso.

Estudiantes que concluyen correctamente

Solo cuatro estudiantes concluyen sobre la diferencia en dispersión en las dos distribuciones del número de caras, número de rachas y longitud de la racha más larga, lo cual coincide con los resultados de Arteaga (2011). A continuación reproducimos la respuesta de uno de ellos:

> En el estudio estadístico comparativo que se hizo entre las secuencias se demostró que sí hay varias diferencias entre una secuencia simulada y una inventada. En cuanto a los promedios, el número de caras simulados fueron muy parecidos a los reales. Esto indica que en esto la intuición humana sobre el lanzamiento de moneda es muy similar a la realidad. Sin embargo el número de rachas en la secuencia inventada es más alto. Finalmente, el valor de la racha más larga fue el más distante. En cuanto la variación típica, los valores de la secuencia inventada son considerablemente menores en los tres casos (# de caras, # de rachas y racha más larga) en relación a la secuencia real. [*Alumno RB]*.

Así mismo, dos de los estudiantes interpretaron correctamente cada una de las variables y añadieron una explicación del porqué de sus resultados. Este ejemplo, aunque con un lenguaje poco preciso, evidencia un cambio en la concepción del azar:

> Las diferencias entre la secuencia real y simulada [*encontradas en el análisis estadístico*] han advertido sobre los errores que en la secuencia simulada han llevado a cabo los alumnos. Esto debido a que priorizamos lo que creemos que puede pasar, es decir, … los alumnos han pensado que algo no se podía repetir muchas veces [*se refiere a la longitud de las rachas*] pero en el azar se evidencia todo lo contrario [*Alumna LAE*].

Esta estudiante manifiesta no sólo un buen manejo estadístico sino una mejor comprensión del contexto del problema en el que lo aplicó. Es decir, el análisis que realizó le sirvió para *conocer* (en un sentido probabilístico) las intuiciones que tienen los seres humanos sobre el azar.

Aunque son sólo estos cuatro estudiantes los que concluyeron en forma correcta todo el ciclo de modelación en el proyecto, hemos visto que otros 26 lo concluyen en forma parcial; bien porque se conforman con analizar la intuición respecto a la tendencia central del número de caras (caso más frecuente); porque no apreciaron la diferencia de dispersión o de promedio; o porque apreciándola, no quieren reconocer que sus intuiciones son incorrectas. En todo caso, estos estudiantes mostraron un buen razonamiento estadístico a lo largo de la actividad, aunque esto sólo se observó poco más de la cuarta parte de la muestra (27%). Lo que indica que hay una necesidad fuerte de preparar a los estudiantes para el trabajo con variables estadísticas y aleatorias en actividades de modelación a través de la aplicación de proyectos.

7.4.6. Discusión de los resultados obtenidos en el objeto Ciclo de Modelación

De manera general, podemos decir que fueron la minoría de los estudiantes de nuestro estudio los que lograron completar el ciclo de modelación, puesto aunque todos realizaron los primeros pasos y algún tipo de trabajo matemático con las distribuciones, pocos lograron llegar al último punto. Es decir, no lograron interpretar en el contexto del problema y tuvieron un desempeño pobre en su conclusión sobre las intuiciones de la aleatoriedad. Estos resultados confirman los de Arteaga (2011).

Llama la atención que un número importante de estudiantes que intentaron comparar las distribuciones de las distintas variables (aunque algunos lo hicieron mal y sólo algunos para las tres variables), pocos emitieron una conclusión al menos parcialmente correcta sobre las intuiciones del azar de los estudiantes del grupo. Esto nos indica que hubo una comprensión parcial del objetivo del proyecto pero pocos completaron las conclusiones que podrían extraerse del objetivo del análisis estadístico.

Para que un estudiante pudiera emitir la conclusión esperada sobre las intuiciones de la aleatoriedad, no bastaba con que comprendiera que se tenía que construir un modelo matemático o aceptara los propuestos por el profesor, ni que hiciera un trabajo estadístico con el modelo o realizara la comparación entre las variables de las secuencias reales y las simuladas. También era necesario comprender la vinculación entre las variables estadísticas y aleatorias con el experimento aleatorio en juego y cómo permiten responder la pregunta planteada sobre las intuiciones. Esto es, se debe vincular las variables aleatorias con el objetivo del proyecto.

Esta desconexión en la emisión de las conclusiones, puede deberse a la incomprensión en la definición de las variables. Es decir, el último paso de la modelación (*interpretación de la realidad*) está condicionado por el segundo paso (*simplificación de la realidad*), en donde, elegido un experimento para recolectar los datos, hay que definir las variables aleatorias que permitirán extraer la información valiosa del resultado del experimento (una secuencia) para evaluar la aleatoriedad de la secuencia.

En nuestro estudio, como observamos en el Punto 7.4.2, en un inicio, los estudiantes sólo proponen el número de caras para analizar los resultados del experimento aleatorio y las otras variables les fueron sugeridas por el profesor. Éstas surgieron en interacción con la clase, bajo la motivación del número de secuencias que se podían obtener con un mismo número de caras pero en distinto orden y resaltando que el orden también era importante. Además se recurrió a trabajar con otros ejemplos semejantes, como el nacimiento de varones en una familia y la posibilidad de tener dos o tres o cuatro seguidos.

Sin embargo, la variable con la que los estudiantes privilegiaron trabajar fue con el número de caras (resultado en el que concordamos con Arteaga, 2011). Así mismo, algunos estudiantes propusieron la variable número de cruces, a pesar de que realmente no es necesaria para el análisis, pero que también es una variable con la que el estudiante tiene más familiaridad que las propuestas por el profesor. Esto, en su conjunto, nos hace deducir que la variable con la que prefieren trabajar es con aquella que les resulta más familiar y por lo tanto, muchos de ellos restringieron su análisis a ella, ignorando el resto de las variables y la vinculación de todas ellas con la pregunta del proyecto.

Podemos inferir que, vinculando con los resultados obtenidos en la emisión de conclusiones, no se comprendió por qué se definieron esas dos variables, cuál era su importancia en la evaluación de aleatoriedad de una secuencia y por qué tenían que ser tres variables y no sólo una (*el número de caras*), a pesar de que el profesor trató de hacer palpable su relación. Esto influyó en que los estudiantes finalmente no pudieran concluir haciendo uso de su trabajo previo, puesto que no pudieran contextualizar su trabajo matemático, lo cual se observó precisamente en las variables vinculadas con las rachas, que fueron las que menos compararon.

En la simplificación de la realidad (*primer estrato de modelación*, Figura 6.26), los estudiantes necesitan comprender que las variables permiten descontextualizar los resultados del experimento, pero también de qué manera responden a la pregunta (o problema) planteada. El hecho de que, en su mayoría, los estudiantes hayan trabajado con los valores de la variable y hayan podido conformar la distribución de frecuencias no es indicador de que los estudiantes hayan comprendido el porqué de la definición de las variables.

La aparente «falta de cuestionamiento» en la obtención de los valores de la variable estadística a partir de los resultados del experimento aleatorio durante ese primer estrato de modelación puede ser otro efecto del contrato didáctico establecido con su profesor o a que el estudiante simplemente efectuó el proceso mecánicamente sin detenerse a pensar en porqué eran esas variables y no otras. Este problema coincide con los resultados encontrados en el Estudio 3 (Capítulo 5) de nuestra investigación en donde también se encontraron dificultades en retornar a la «realidad» en el primer estrato de modelación.

Wild y Pfannkuch, (1999) afirman que las primeras etapas de la modelación están conducidas por el conocimiento del contexto y en nuestra investigación notamos que éste condiciona también otras etapas de la modelación. Si el estudiante no es capaz de comprender plenamente la simplificación de la realidad a través de las variables, eso mermará sus conclusiones y por lo tanto, el sentido del análisis estadístico. Sostenemos entonces, que el paso 2 amerita mayor importancia para que nuestros estudiantes puedan completar el ciclo de modelación y lograr un razonamiento estadístico.

También observamos que, de manera general, los estudiantes se dejan influir por su conocimiento del contexto (que, en este caso, se reduce sus creencias sobre la aleatoriedad descritas en el apartado 7.1) más que por su análisis estadístico para interpretar sus resultados. En este sentido, se comportan de manera diferente al efectuar la comparación de las variables vinculadas con las secuencias reales y simuladas, que muchos lo hicieron de forma correcta, y al emitir sus conclusiones, que pocos finalizaron y donde se dejaron llevar más por sus creencias individuales (Punto 7.1.1). Así, algunos estudiantes mostraron sus creencias sobre la forma en que se comporta una sucesión de resultados aleatorios al lanzar 20 veces una moneda, tanto al inventar secuencias como en la interpretación de sus resultados en la comparación de los estadísticos de las secuencias simuladas y reales.

Aunque de manera general los estudiantes presentan diversos cálculos estadísticos, ellos prefieren hacer uso únicamente de las medidas de tendencia central para la comparación de distribuciones, cosa que, en general hacen bien. En cambio, el rango, que es la medida de variación más calculada, es poco socorrida para esa comparación, y, en general, el desempeño de los estudiantes es medio en ella. Esto concuerda con el modelo obtenido en las variables vinculadas a la secuencia simulada (punto 7.1.2), que indica una familiaridad de los estudiantes con la esperanza del número de caras, descuidando la variabilidad. Otro número importante de estudiantes, en cambio, simplemente ignoraron el análisis estadístico para emitir sus conclusiones de acuerdo a sus creencias individuales (Punto 7.1.1).

El contrato didáctico también jugó un papel importante en que los estudiantes no completen el ciclo de modelación, puesto que muchos gráficos, tablas y estadísticos e incluso algunas interpretaciones y conclusiones se añaden al reporte sólo como una manera de cumplir con la tarea asignada sin percatarse de la importancia de ellos en su análisis o sin comprender cuál es su papel dentro de la respuesta a la pregunta del proyecto.

Todos estos puntos nos ponen en una alerta sobre lo que hay que observar en la realización del proyecto, puesto que aunque alrededor de la cuarta parte de los alumnos participantes en el estudio pudieron transitar por todas las etapas de la modelación y una gran mayoría sí realizaron las comparaciones, se detectaron diversos errores y muy pocos llegaron a conclusiones certeras.

8. Conclusiones del capítulo

El objetivo del proyecto planteado fue evaluar la comprensión de la variable aleatoria y la forma en que los estudiantes la vinculan con otros objetos matemáticos en una actividad de modelación en la que está implicada la realización de inferencias estadísticas informales (Rossman, 2008). Como perspectiva complementaria a la anterior, también estamos interesados en recopilar información sobre la forma en cómo los estudiantes recorren las diferentes fases del ciclo de modelación propuesto por Dantal (1997) y haciendo consideración al modelo de investigación propuesto por Wild y Pfannkuch (1999). En las conclusiones primeramente confrontamos las hipótesis planeadas con los resultados obtenidos y después concluiremos analizamos la idoneidad del proyecto para cumplir el objetivo planteado.

8.1. Confrontación de los resultados con las hipótesis

En el punto 6 de este capítulo se plantearon las hipótesis sobre lo que esperábamos encontrar en las producciones escritas de los estudiantes sobre el proyecto planteado, que se discuten a continuación.

Comprensión e intuiciones sobre la idea de aleatoriedad

La evaluación de las intuiciones sobre el azar se realizó a través del análisis de los argumentos que algunos estudiantes, espontáneamente incluyeron en sus informes y de un análisis estadístico completo de los datos de los experimentos realizados por los estudiantes. Como resultado, se obtuvieron ejemplos de las siguientes concepciones, muchas parcialmente correctas. La mayoría también aparecieron en el trabajo de Serrano (1999) o en momentos históricos de desarrollo del concepto descritos en Bennet (1998):

- *Aleatoriedad como impredecibilidad*: es aleatorio aquello que no se puede predecir. Aunque esta idea es correcta, a veces se lleva al límite de considerar que tampoco se puede predecir la distribución de una variable aleatoria. Un fenómeno relacionado con esta creencia es el enfoque en el resultado aislado, descrito por Konold (1989).
- *Aleatoriedad como falta de control*: es aleatorio aquello que no se puede controlar. Por el contrario, algunos estudiantes piensan que sería posible, bien por inteligencia, conocimiento de los factores o suerte controlar el fenómeno aleatorio, mostrando la ilusión de control (Langer, 1975).

- *Aleatoriedad como ausencia de causas*, como un tipo especial de causa o como causa desconocida (ignorancia).
- *Diferencia entre proceso y resultado aleatorio*. Aunque un proceso se puede considerar aleatorio a priori es sólo mediante el análisis estadístico (a posteriori) de la secuencia de resultados que se puede confirmar la aleatoriedad.
- *Aleatoriedad como equiprobabilidad*: un suceso es aleatorio si es equiprobable. Esta creencia lleva en ocasiones al sesgo de equiprobabilidad (Lecoutre, 1992).
- *Aleatoriedad como variabilidad*: aleatorio es equivalente a que varía. Aunque de acuerdo a Wild y Pfannkuch (1999) la percepción de la variabilidad es parte del razonamiento estadístico, esta creencia no es totalmente cierta pues hay variabilidad de tipo determinista (relacionada con la razón de cambio en una función). La variabilidad aleatoria a veces también se puede reducir por medio de la determinación de los factores que la afectan.
- *Aleatoriedad como falta de modelo*. Aunque esta concepción está en la base de la definición de aleatoriedad de algunos matemáticos (como Von Mises, 1756, y Kolmogorov, 1956), muchos estudiosos de la aleatoriedad subjetiva (Green, 1982; Konold 1989; Batanero y Serrano, 1995) que argumentan que es necesario considerarla como una multiplicidad de modelos. Existen también, diversos estudios probabilísticos que analizan la aleatoriedad a través de diversos modelos (Beltrami, 1999).

Esta variabilidad y riqueza de ideas sobre la aleatoriedad podría ser explotada en la clase por parte del profesor para hacer evolucionar a los estudiantes hacia una concepción más completa de lo que es aleatoriedad como modelo matemático del cálculo de probabilidades (Batanero y Serrano, 1999). Esta posibilidad se hace indispensable cuando observamos lo sutil que es la diferencia entre una buena y mala concepción de aleatoriedad en cada una de esas concepciones. El proyecto propuesto a los alumnos y analizado en este capítulo puede ser una buena forma de introducir esta necesaria discusión.

El foco principal de nuestra investigación no se centró en el concepto de aleatoriedad, sin embargo éste es el contexto del análisis de los estudiantes en el proyecto y uno de los objetivos de aprendizaje del mismo proyecto. Observamos la forma en cómo el estudiante concibe y evalúa la aleatoriedad y también cómo ambas perspectivas se retroalimentan. El análisis de las concepciones individuales y globales del grupo sobre aleatoriedad proporcionó información que pudimos vincular con la forma en que los estudiantes analizaron las variables estadísticas, así como la forma en que a partir de éstas realizan inferencias sobre las variables aleatorias. Así, coincidimos con Makar, Bakker y Ben-Zvi (2011) sobre la importancia del conocimiento del contexto y que es necesario darle mayor énfasis a éste en las clases de estadística.

Uso de variables estadísticas y su distribución

La mayoría de los estudiantes espontáneamente llegó a formar la distribución de las variables estadísticas bajo estudio; sólo 7 estudiantes trabajan con sus propios datos, y 15 más usan las variables sin llegar a formar la distribución, trabajando con un listado de los valores correspondientes a las mismas.

De manera general, los alumnos lograron establecer la distribución de las variables ya sea por medio de tablas (55%) y gráficos (70%) o ambos (80%), predominando los gráficos de distintos tipos. El uso de la variable «número de caras» fue mucho más intuitivo que las variables relacionadas con las rachas; sin embargo todos los que construyen tablas las construyen para las tres variables en el estudio; todos los que construyen gráficos los hacen para la variable «número de caras» y el 70% de ellos para alguna o las otras dos variables.

También la mayoría de los estudiantes calculó algún estadístico, principalmente estadísticos de posición central (media, mediana y moda) y en menor medida de dispersión. Esto es un indicador de la comprensión alcanzada de la distribución de acuerdo con Shaughnessy (2007). Por otro lado, la mayoría de los estudiantes trabajó con frecuencias absolutas (y no con relativa) en sus tablas y al comparar las distribuciones. La tarea no requería el uso de frecuencias relativas, por lo que no se ha podido observar si alcanzan el segundo de los niveles de comprensión de distribución definido por Watson (2001).

Representación de las distribuciones

Aunque sólo seis alumnos construyen una tabla conjunta para dos de las variables del estudio, 25 alumnos son capaces de construir un gráfico conjunto para las dos distribuciones; lo que implica una mejor comprensión del gráfico que de la tabla por parte de los alumnos.

En los gráficos construidos, la mayoría llega a formar la distribución, aunque aproximadamente la mitad de los gráficos tienen algún error. Respecto a la interpretación de gráficos, la cuarta parte de los estudiantes que realiza gráfico se limita a presentarlos sin hacer comentarios sobre los mismos, pensamos que por efecto del contrato didáctico. Un 42% de los que construyen gráfico los interpretan pero cometen algún error o bien hacen una interpretación incompleta que no permite llegar a una conclusión completa sobre el problema y el resto hacen una buena interpretación del gráfico. Estos resultados coinciden con los encontrados por Arteaga (2011).

Uso de promedios y dispersión

El cálculo de estadísticos es en general correcto, con pocas excepciones, aunque también aparecen algunos de los errores descritos por Cobo (2003) o Mayén (2009); por ejemplo, no ponderar en el cálculo de la media ponderada o no ordenar los datos para calcular la mediana pero con mucha menor frecuencia que en aquel estudio. Si los estadísticos de dispersión, son calculados, lo son correctamente. Apenas aparecen errores en el cálculo de estos estadísticos.

Aunque todos los estudiantes presentaron medidas de centralización, la mayoría de ellos se limitó a calcularlas, pero no las interpretó ni las usó para la comparación de las variables en el estudio, por lo que parece que no alcanzan los niveles superiores de comprensión de la distribución en la clasificación de Bakker y Gravemeijer (2004). Sin embargo, en caso de compararlas, la comparación es correcta, lo que implica una buena comprensión del significado de la media y moda (apenas se usa la mediana). Pensamos, entonces, que el hecho de presentar los estadísticos sin interpretarlos es un efecto del contrato didáctico, puesto que en realidad no comprenden para qué necesitan calcularlos. En gran parte esto es debido a que están poco acostumbrados al trabajo con proyectos abiertos en la clase de estadística.

La idea de dispersión es menos expuesta que la de centralización. Aunque el cálculo del rango, en general, lo hacen correctamente, los alumnos no llegan a captar el significado de la dispersión ni su utilidad en la comparación de dos distribuciones. La desviación estándar o varianza sólo fue usada por unos cuantos alumnos y no fue interpretada.

Uso de la variable aleatoria

Aunque no se pidió explícitamente, algunos alumnos hacen referencia explícita a las variables aleatorias en juego en el estudio. Otros relacionan la variable con la aleatoriedad del experimento y la dificultad de predecir sus valores cuando hacen referencia al azar o al resultado del experimento, es decir ven el valor de la variable como el resultado de un experimento aleatorio, lo cual coincide con la propuesta de Parzen (1971).

Esto se observó como una ventaja en la entrevista clínica puesto que favoreció que los estudiantes vieran el resultado de aplicar la variable aleatoria o estadística, como un resultado del primer estrato de modelación y por tanto como la «realidad» del segundo estrato. Sin embargo, en ambos estudios también se percibió una dificultad posterior, al tratar de interpretar los resultados estadísticos en el contexto del problema. Ante esta dificultad, los estudiantes recurren a su conocimiento del contexto (en este caso a sus creencias sobre aleatoriedad) para tratar de solventar este hueco en el tránsito hacia la realidad para contestar la pregunta del proyecto. Dos de las dificultades más importantes puestas en juego en este sentido son las creencias de la aleatoriedad como enfoque aislado, falta de control o ausencia de modelos, que hacen que los estudiantes tiendan a una visión determinista, invalidando los resultados obtenidos en el análisis estadístico como conocimiento en sí mismo. Bajo estos enfoques, los estudiantes no pueden ir más allá de sus datos, que en ocasiones se reducen a los suyos propios.

La dificultad de contextualizar está vinculada con la aplicación inversa de las variables, es decir, con el proceso inverso en el primer estrato de modelación. En este estudio, se vinculó con la incomprensión del papel de las variables para responder la pregunta del proyecto. En este sentido, coincidimos con Chaput, B., Girard, J. C. y Henry, M. (2011) quienes resaltan la importancia de que el estudiante comprenda el resultado del primer estrato de modelación como un modelo pseudoconcreto[12], puesto que esto evitaría dificultades en la interpretación de los resultados en el contexto (realidad, desde su propuesta).

Otras veces la referencia al uso de la variable aleatoria es implícita, al hablar de las probabilidades o posibilidades de algunos valores, o de los valores posibles de la variable y no sólo de los resultados en el experimento. En estos razonamientos se observa una puesta en relación de variable estadística y aleatoria, es decir, la realización de una inferencia informal (Zieffler, Garfield, Delmas y Reading, 2008).

En algunos casos los estudiantes aluden a la probabilidad de algunos valores de las variables analizadas, que estiman a partir de las frecuencias obtenidas en los experimentos, aplicando una concepción frecuencial de la probabilidad. Ligan también en sus razonamientos la variable aleatoria que corresponde a una variable estadística dada y su distribución de probabilidad o la probabilidad de algunos de sus valores aislados con la distribución de frecuencia obtenida, mediante una inferencia de tipo informal.

Son bastantes los que explicitan el valor esperado de la distribución binomial (10 caras en 20 experimentos) apoyándose en la equiprobabilidad de los resultados y el número de ensayos (que serían los parámetros de la distribución binomial). En casos aislados hay una referencia a la convergencia (de la distribución de frecuencias a la

[12] El modelo de Henry (1997), retomado por Chaput, Girad y Henry (2011), está descrito en la sección 9.4 del Capítulo 3. Lo que para Henry es un modelo pseudo-concreto para nosotros (acordes con Wild y Pfannkuch, 1999) es un modelo y adopta el papel de «realidad» en el siguiente estrato de modelación.

distribución de probabilidad). También en pocos casos se hizo patente que la variable era cuantitativa.

Ciclo de modelación

En el proyecto se pretendió que los estudiantes completaran un ciclo de modelación (Dantal, 1997). Sumergimos al estudiante en una situación aleatoria de la que él mismo formaba parte al proponerles evaluar sus propias concepciones del azar y llevar a cabo un experimento aleatorio real para confrontarse. Durante la discusión inicial con el profesor, en la que los estudiantes participaron activamente expresando sus ideas sobre cómo se manifiesta el azar y sus concepciones de la aleatoriedad, hubo una aceptación general de la pregunta planteada y del experimento aleatorio. Sin embargo, el análisis completo de los reportes de los estudiantes dejó entrever que algunos estudiantes no manejaron intuitivamente el experimento aleatorio compuesto por 20 lanzamientos de la moneda. Esto trajo como consecuencia que su análisis se restringiera a la variable «cara o cruz» o bien a cálculos erróneos en las variables vinculadas con las rachas. Eso resulta comprensible, puesto que lo que se está evaluando es la aleatoriedad del lanzamiento de una moneda, pero para poder hacerlo se recurre a analizar secuencias de 20 resultados de ella. El planteamiento de un experimento aleatorio compuesto es, por tanto, un recurso teórico difícil de comprender.

El segundo paso del ciclo, simplificar la realidad, surgió a partir de una discusión grupal en la cual una de las variables es espontáneamente sugerida por algunos estudiantes, estudiar el número de caras o cruces. Ellos también propusieron la equiprobabilidad de caras y cruces en el experimento simple. Otra parte, en cambio, el estudio de las rachas, es básicamente sugerida por el profesor, lo cual, como es sabido, no garantiza la comprensión de su propósito en el análisis.

En su totalidad, los estudiantes hicieron uso de las variables, cuyos valores sustituyeron el resultado del experimento aleatorio, e incluso los llegaron a tratar como el resultado de su experimento aleatorio. Sin embargo, una gran parte de los estudiantes no comprendieron la necesidad del estudio de las tres variables en su conjunto, ni su relación con el proyecto en su totalidad, puesto que en su mayoría, aunque hicieron un análisis estadístico, no llegaron a sacar conclusiones de las mismas ni a hacer uso de ellas para responder la pregunta del proyecto. Por lo tanto, creemos que gran parte de la aplicación de las variables al resultado de su experimento, fue solamente la aplicación de reglas (contar el número de caras o de rachas u observar la longitud de la racha más larga) sin comprender su sentido en el contexto del problema.

En su mayoría, los estudiantes ignoraron el análisis de las variables vinculadas con las rachas, por lo que también es posible que sus intuiciones incorrectas sobre las rachas (en general, esperaban rachas cortas) de estos estudiantes, les llevara a ignorar los datos del análisis que contradecían dichas intuiciones.

A partir de la definición de las variables, los estudiantes continúan el ciclo de modelación por sí mismos, construyendo sus tablas o gráficos y realizando un proceso de transnumeración (Wild y Pfannkuch, 1999). En su totalidad, fueron capaces de construir la distribución. Definieron y trabajaron con modelos matemáticos propuestos por ellos, lo cual nos permitió observar diversos tipos de tablas, gráficos y estadísticos. El cálculo de estadísticos, en general, fue correcto, con pocas excepciones. En los gráficos, sin embargo, se presentó una mayor dificultad, en cambio, la elaboración de tablas fue mucho más sencilla para ellos.

Respecto al lenguaje matemático, muchos alumnos muestran un buen conocimiento y uso de las fórmulas, e incluso de las fórmulas de cálculo abreviado y la disposición auxiliar para realizar los cálculos. Sin embargo, el uso de los símbolos matemáticos varía en formalización y su lenguaje verbal es muy impreciso. Falta competencia en comunicación verbal en los estudiantes.

El paso último en el proceso de modelación (interpretación del modelo matemático con la realidad) fue la más difícil, pues para los estudiantes fue más sencillo descontextualizar (pasar del problema a la matemática) que el proceso inverso (contextualizar de nuevo), hecho que también se observó en la entrevista y está vinculado con que la modelación sigue un proceso estratificado. Muchos estudiantes no vieron las implicaciones del análisis estadístico sobre las intuiciones de los estudiantes, es decir, no comprendieron cómo se relacionaban las variables con la evaluación de la aleatoriedad de las secuencias por lo que les fue difícil regresar al contexto. En otros casos sus propias intuiciones incorrectas (y el no querer aceptarlas como incorrectas) les llevaron a pensar que habían obtenido resultados en los que parecía que su modelo estadístico no era adecuado, puesto que habían realizado un experimento aleatorio, «por azar», en el que cualquier cosa podía ocurrir. En otros, su falta capacidad de análisis (capacidad de trabajo con el modelo matemático) para detectar las diferencias, los llevó a concluir que los resultados en las dos secuencias son similares. Esto último principalmente debido a la falta de interpretación de las medidas de dispersión.

Otras dificultades de los estudiantes

Además de los puntos señalados, se han observado dificultades particulares que también aparecieron en la entrevista clínica, de lo que se deduce que podrían aparecer en otras situaciones de trabajo con la variable aleatoria:

- *Confusión entre frecuencias absolutas y relativas*, que apareció en la entrevista clínica y también en la tesis de Cobo (2003), Mayen (2009) y Contreras (2011), donde algunos estudiantes confunden la probabilidad con los casos favorables. Desde nuestra perspectiva de modelación, esto es producto de la dificultad de visualizar la frecuencia relativa como un modelo matemático a partir de la realidad.
- *Confusión entre valor de la variable y frecuencia*, que también apareció en la entrevista clínica y en las investigaciones de Cobo (2003), Mayén (2009) y Arteaga (2011). Ello implica falta de diferenciación entre una variable dependiente e independiente en la distribución de frecuencias. Este problema también se vincula con el primer estrato de modelación de la situación estudiada desde nuestra perspectiva de modelación.
- La necesidad de comparar las distribuciones de las variables estadísticas de las secuencias simuladas y reales provoca que algunos estudiantes *extiendan el rango de las variables* en sus gráficos. Este proceso, que facilita la comparación visual, también hizo que algunos estudiantes se dejaran llevar por la apariencia errónea de que los rangos de ambas variables son iguales. También hubo personas que, al calcular el rango, usaron el mínimo y máximo del conjunto de valores de las dos variables estadísticas, es decir, no diferenciaron las distintas poblaciones que están interviniendo en el estudio y asociaron un solo rango a un par de distribuciones (y no a una sola). Ambos procesos son diferentes porque en el segundo, el cálculo lo hacen a partir de la tabla, en cambio en el primero hacen uso de los gráficos. El rango podría ser mayor en el primero si se percataron de todos los valores posibles que podía adoptar la variable y los añadieron a su gráfico. Estos errores dificultarán, posteriormente, la comparación de la dispersión de las dos distribuciones, que tendrían, bajo estos razonamientos, la misma variabilidad.
- Al tratar de relacionar la variable estadística y aleatoria mediante una inferencia informal, los estudiantes a veces *confunden los valores posibles de la variable aleatoria con los datos* (valores que han resultado en la variable estadística). Es decir, no diferencian los datos empíricos del modelo

teórico.

- Los estudiantes tienen muchos *errores en la construcción de gráficos*, repitiéndose los descritos en Arteaga (2011); en particular son muy frecuentes los errores en las escalas, pero también la elección de un gráfico inadecuado o la representación de variables no comparables en los gráficos.

8.2. Idoneidad del proyecto

El proyecto fue tomado de investigaciones anteriores, porque pensamos que era especialmente adecuado para nuestro trabajo.

A continuación se analiza la idoneidad didáctica del proyecto como articulación de los cinco componentes que definen Godino, Wilhelmi y Bencomo (2005) y Godino, Contreras y Font (2006): idoneidad epistémica, cognitiva, interaccional, mediacional y emocional.

- *Idoneidad epistémica*. El proyecto permitió el trabajo de los estudiantes con el concepto de aleatoriedad, un experimento aleatorio particular y la definición de tres pares de variables aleatorias. Puesto que no estaban preparados suficientemente en cálculo de probabilidades, se estudiaron dichas variables a través de las correspondientes variables estadísticas. Ello hizo que los estudiantes trabajaran con distribuciones de frecuencias, sus representaciones tabulares y gráficas, sus medidas de tendencia central y su dispersión. Así mismo se involucraron en un ciclo completo de modelación, incluyendo la realización de inferencias informales para poder obtener conclusiones sobre las variables aleatorias a partir del trabajo con las variables estadísticas.
- *Idoneidad cognitiva*. La situación moviliza el aprendizaje de los estudiantes, quienes expresan y reflexionan sobre sus ideas respecto a la aleatoriedad y eventualmente las mejoran. Los estudiantes también recogieron datos de un experimento y los analizaron utilizando tablas, gráficos y resúmenes estadísticos. Pensaron en términos de probabilidades y esperanza y se involucraron en la realización de un proyecto estadístico. También vieron la utilidad de la estadística para analizar situaciones aleatorias. Tratándose de una tarea abierta, el aprendizaje fue variable, dependiendo del estudiante, sin embargo el trabajo recopilado nos indica que el proyecto estuvo dentro de sus capacidades.
- *Idoneidad interaccional*. La situación problema se conjugó con la complicación de un proceso completo de investigación estadística que los

estudiantes pudieron abordar con ayuda del profesor. Las variables aleatorias y estadísticas surgieron de una forma contextualizada, aunque no se pudieron comprender del todo bien. El interés por comprobar cómo eran sus intuiciones provocó la devolución del problema a los alumnos, a la vez, el debate. Su argumentación al responder las preguntas planteadas sirve para explicitar sus concepciones, dificultades y estrategias. Lo que permitió explorar sus concepciones, dificultades y estrategias, así como poner en juego sus concepciones matemáticas en diversos aspectos. Sin embargo no se logró del todo que aceptasen los resultados del análisis de datos, lo que hizo que los estudiantes le restaran importancia a su propio trabajo matemático y estadístico y prevalecieran sus creencias sobre la aleatoriedad en sus conclusiones (Makar, Bakker, Ben-Zvi, 2011).

❖ *Idoneidad mediacional*. La puesta a punto de la situación requirió pocos medios, únicamente papel y lápiz y el pizarrón, y una moneda para cada estudiante, así como hojas para registrar los datos. Aunque algunos estudiantes utilizaron la hoja Excel para realizar sus análisis, no siempre fue bueno, pues algunos que la utilizaron añadieron nuevos errores provocados por el software.

❖ *Idoneidad emocional*: La situación fue interesante para los estudiantes, quienes estuvieron interesados por saber cómo eran sus intuiciones y trabajaron con gran interés en producir sus informes. Pensamos que se consigue la devolución del problema a las estudiantes, quienes incluso llegan a pensar en un momento de que no se trata de un problema matemático y la llegaron a tratar como una situación a-didáctica. Ello indica una alta idoneidad emocional del proyecto.

En resumen, los puntos anteriores indican una idoneidad razonable del proyecto, tanto como instrumento de evaluación (como lo hemos usado en el trabajo), como con fines didácticos.

7. Conclusiones

ÍNDICE DE CONCLUSIONES

1. Introducción

En esta Tesis hemos realizado un acercamiento hacia la variable aleatoria mediante dos estudios de tipo epistemológico (descrito en los Capítulos 3 y 4) y otros dos de carácter cognitivo (Capítulos 5 y 6). Previamente se ha justificado el interés de centrarnos en este objeto matemático, sustentándonos en tres consideraciones principales: (a) su carácter de idea fundamental en estadística, sugerido por Heitele (1975), quien considera que puede ser enseñada con diferentes grados de complejidad a lo largo de todo el currículo escolar; (b) las dificultades previsibles de los estudiantes, puesto que el estudio de la variable aleatoria y su distribución está ligado a la comprensión de otros objetos matemáticos (Reading y Canada, 2011); (c) la escasez de investigaciones previas, pues este objeto matemático apenas ha recibido atención desde la investigación didáctica.

La investigación se ha apoyado como marco teórico principal de la Teoría de Situaciones (Brousseau, 1986, 1997) y, como consecuencia, metodológicamente, hacemos uso de la ingeniería didáctica (Artigue, 1995a). Más concretamente, se pretende producir información dentro del análisis preliminar previo al diseño de situaciones didácticas en los componentes epistemológicos y cognitivos de dicho análisis preliminar. La recogida de datos en los estudios empíricos ha requerido una experimentación exploratoria de dos situaciones adidácticas complementarias en las que surge implícitamente la idea de variable aleatoria, una de ellas basada en el significado clásico de la probabilidad y la otra en los significado clásico y frecuencial y la realización de una inferencia informal, que podrían, en el futuro, ser base en el diseño de situaciones didácticas.

Por otro lado, también se han incorporado algunos elementos del Enfoque Onto Semiótico desarrollado por Godino y sus colaboradores. La clasificación que proponen los autores de objetos matemáticos primarios se usa para diferenciar varios significados de la variable aleatoria, que dependen de la forma de asignar probabilidades y para determinar el significado de referencia de nuestro trabajo. También nos hemos apoyado en la idea de idoneidad didáctica para analizar las situaciones utilizadas en los Estudios 3 y 4.

A continuación, se resumirán las principales conclusiones de cada uno de estos dos tipos de estudios desde los objetivos generales de nuestra investigación. Los objetivos e hipótesis de cada estudio ya fueron discutidos con más detalles en los diferentes capítulos y en sus conclusiones. Después se presentan las aportaciones más relevantes de nuestro trabajo. Se finaliza indicando las implicaciones de los resultados para la enseñanza de la estadística en la educación universitaria y su alcance para futuras líneas de investigación.

2. Sobre los estudios epistemológicos

El primer objetivo de nuestro trabajo (Sección 4 del Capítulo 1) se formuló en los siguientes términos:

Objetivo 1: *Llevar a cabo un análisis a priori preliminar sobre la variable aleatoria desde el punto de vista epistemológico, que aborde tanto los aspectos disciplinares como los históricos*.

Para responder a este objetivo se han llevado a cabo dos estudios epistemológicos que se describen en los Capítulos 3 y 4. Cada uno de ellos ha producido resultados originales y otros que confirman los de otras investigaciones que nos precedieron. Estos resultados, junto con el estudio de los antecedentes que se reporta en el Capítulo 2, nos han permitido posteriormente establecer de una manera fundamentada, las hipótesis de investigación en los estudios cognitivos. Del mismo modo, han servido para discutir nuestros hallazgos en los estudios cognitivos y explicar algunas de las dificultades que muestran los estudiantes en los mismos.

2.1. Desde la disciplina

En el *Estudio 1* (Análisis Epistemológico desde la Disciplina) se realizó un análisis matemático detallado del objeto «variable aleatoria» y sus relaciones con otros objetos matemáticos (función de distribución y densidad, parámetros, medidas de tendencia central y dispersión, variable estadística, probabilidad, espacio probabilístico, función y función medible, componentes de una función).

Se analizan con detalle las diferentes funciones que giran en torno a la variable aleatoria, comenzando por la misma variable, como función medible en un espacio probabilístico y se destaca la importancia de la función de distribución (función compuesta de la inversa de la variable aleatoria y la probabilidad) que permite introducir el análisis matemático en el trabajo con las variables aleatorias, y en consecuencia, con la estadística. También se resalta la naturaleza dual de la variable aleatoria, puesto que su definición debe permitir el manejo del espacio muestral o del resultado del experimento aleatorio en el espacio de los números reales, pero también es el vínculo con el contexto y por lo tanto, su definición debe hacerse de modo que permita dar respuesta a la pregunta que motiva el análisis de la situación.

En este sentido Parzen (1971) enfatiza en la importancia de la definición de la variable aleatoria como función, al aprender a reconocer y formular matemáticamente como *función* objetos descritos verbalmente a partir del contexto del problema. Es decir, es importante percibir como una *aplicación* (en el sentido matemático) el texto que traduce los sucesos del espacio muestral en números reales. Es precisamente esa dualidad la que hace importante a la variable aleatoria dentro del ciclo de modelación, puesto que en una primera fase de la modelación describe la «realidad» y, en un estrato más alto, reemplaza a la realidad como sustrato operacional. Así la variable aleatoria es una *realidad matemática* que se modela mediante la función de distribución y permite introducir el análisis matemático en el estudio de los fenómenos aleatorios. La descripción de los diferentes modelos de distribuciones (como la Binomial o Normal) constituyen, según Batanero (2003), un ejemplo de la potencia de la modelación en estadística.

Este doble proceso de modelación, que será de especial interés en el Estudio 3, aparece desde el contexto del problema a la variable aleatoria y de esta a su función de distribución. En nuestro trabajo, re-interpretamos la modelación estocástica dentro del marco descrito por Wild y Pfannkuch (1999), haciendo uso de la propuesta de Heitele (1975), para definir una modelación estratificada.

También se contempla la relación entre variable estadística y aleatoria desde esta perspectiva (Ríos, 1967) observando un doble proceso de modelación en la construcción de la variable aleatoria: desde los datos a la variable estadística; y desde este a la variable aleatoria, siguiendo una inferencia informal (contemplado en el Estudio 4). Asimismo, señalamos algunas posibles dificultades del aprendizaje de la modelación en este campo, puesto que existe una mayor la distancia entre realidad y modelo matemático que en otras ramas de las matemáticas (Coutinho, 2001), que sin embargo, ofrece una gran oportunidad para enseñar la modelación en matemáticas, por las muchas situaciones cotidianas en las que es posible modelar fenómenos aleatorios (Chaput, Girard y Henry, 2011).

Un punto que no se suele discutir en los textos de estadística es la influencia que la forma en que se asigna probabilidad tiene sobre el estudio de la variable aleatoria, puesto que añade nuevos objetos matemáticos (problemas, conceptos, propiedades y procedimientos) al estudio de la misma y que difieren dependiendo del tipo de asignación de probabilidades (Batanero, 2005). La descripción de algunos de los significados que ha recibido la probabilidad a lo largo de la historia, de su campo de aplicación y problemática filosófica en la asignación de probabilidades nos ha llevado a interpretar diferentes significados de la variable aleatoria dependiendo de esta asignación.

La conclusión del estudio es la complejidad del objeto variable aleatoria, cuya comprensión ha de ser gradual y requiere un largo tiempo de dedicación. El estudio nos permite analizar con detalle los componentes que habría que incluir en dicha enseñanza y proporciona la base para la construcción de futuras ingenierías didácticas sobre la variable aleatoria.

2.2. Desde la historia

En el *Estudio 2* (Análisis Epistemológico histórico) se resume el desarrollo histórico de la idea de variable aleatoria, desde múltiples aplicaciones en todas las áreas de la actividad humana, y que se puede clasificar en dos grandes corrientes de pensamiento surgidos a partir de los análisis probabilísticos y del análisis de datos (estudios estadísticos), que no se suceden sino que interactúan en diversos momentos de la historia.

A partir de los análisis probabilísticos.

Inicialmente, a partir de los primitivos juegos de azar y del estudio de la previsión de apuestas surgió directa, aunque lentamente, la idea de variable aleatoria discreta como variable numérica vinculada a los sucesos de un espacio muestral. Los primeros problemas estaban vinculados con la determinación de la esperanza matemática de ciertas variables aleatorias (la determinación de la apuesta; el estudio de la equitatividad de un juego, etc.). En estos primitivos problemas, el espacio muestral del experimento es, al menos potencialmente, conocido a priori y la asignación de probabilidades es Laplaciana (García, 1971; Hald, 2003; Bellhouse, 2000).

La difusión de estos conocimientos, generó un interés en su aplicación a diversos campos, que condujo a cuestionamientos teóricos sobre la relación entre la variable aleatoria y la variable estadística (Halley, 1693; Bernoulli, 1954/1738; Hald, 1986). Estas inquietudes ocasionaron el desarrollo de un gran número de problemas intra-matemáticos que desembocaron en el trabajo con las primeras operaciones de las variables aleatorias y la convergencia de sucesiones de variables aleatorias. Se introdujeron las primeras versiones de los teoremas del límite, la ley de los grandes números y la del límite central, así como el trabajo con las variables continuas y con funciones de densidad, particularmente con la Normal y la Exponencial (de Moivre, 1756; Pearson, 1924; Hald, 2003; Sylla, 2006; Bernoulli, 2006/1713; Alvarado, 2007).

Con ello, se detectó la necesidad de la definición de una variable numérica vinculada a los resultados de experimentos aleatorios, pero aún se trabaja sólo implícitamente con ella y la asociaban más bien al azar porque aún no había una distinción entra azar y aleatoriedad. Algunas veces se le denominaba *variable*, y otras *magnitud* o *cualidad variable* (Bernoulli, 1954/1738; Laplace, 1812;). Una etapa de trabajo matemático más fuerte fue la necesidad de establecer la función de densidad y de atacar el problema de continuidad a mediados del siglo XIX, lo que conllevó a diferenciar el trabajo con la variable aleatoria discreta y continua, así como a la definición de *función de densidad*. En este momento, dentro del ámbito formal de la probabilidad, la variable aleatoria comenzó a perder su carácter físico, puesto que ya no se manejaba como el resultado de un fenómeno aleatorio con resultados numéricos sino se comenzó a vincular con la asignación de valores numéricos al espacio muestral (Poisson, 1832; Tchebycheff, 1899b/1867; Liapounoff, 1900; Seneta, 1966; Fischer, 2011). Una última etapa, con el advenimiento del trabajo de Fréchet y Kolmogorov (Kolmogorov, 1993/1956), se caracteriza por el surgimiento de la variable aleatoria como objeto matemático a partir de la necesidad de abstracción de una teoría de probabilidad, tal como la conocemos actualmente, no sólo en su definición sino también en su nomenclatura.

Esta corriente de pensamiento estaba más influida por la formulación de un sustento matemático que diera fundamento a la, en ciernes, teoría de probabilidad. Los principales problemas abordados, rápidamente se desvincularon de sus aplicaciones, para ocuparse de generalizaciones de las soluciones de problemas particulares, obtención de métodos generales, definición de conceptos, demostración de teoremas y finalmente, la conformación de una teoría. La variable aleatoria fue despojada de sus significados intuitivos y se admitió la posibilidad de variables aleatorias abstractas (por ejemplo, variables multidimensionales detrás de las cuáles no encontramos una magnitud, como usualmente la entendemos). Estos resultados y trabajo estaban sustentados en la creciente matemática del cálculo, la teoría de la medida y la teoría de conjuntos, pero, a medida que la teoría de probabilidad se desarrollaba y resolvía sus propios problemas, también tuvo una fuerte influencia sobre aquellas, a tal grado que la probabilidad, quedó ligada a ellas.

Los estudios estadísticos

Los estudios estadísticos son conocidos desde mucho tiempo atrás, aparecen en los censos romanos y en todas las civilizaciones primitivas para fines militares y políticos. Científicos y políticos e incluso algunos jugadores interesados en conocer la mejor apuesta posible dirigieron sus esfuerzos a la recopilación de datos y su análisis. Alrededor del siglo XVII, ya se estudiaban las regularidades de los datos con el objetivo de hacer predicciones, por ejemplo, en el estudio de las tablas de mortandad y la esperanza de vida. En ese entonces, las variables se definían por el contexto del problema, pero se comenzó a denotar la importancia de variables numéricas para hacer uso de algunos resultados y herramientas probabilísticos de la época. En particular, el trabajo de Galileo durante el siglo XVII, da un empuje fuerte al estudio de la variación y del análisis empírico de la variable continua a partir del análisis de los errores de las mediciones astronómicas (Hald, 1986).

A finales del siglo XIX y principios del XX, las inferencias informales tuvieron un fuerte auge gracias a la difusión de la distribución normal como modelo apropiado para muchos contextos en ciencias sociales, biológica y de la herencia. Surgieron las representaciones gráficas y tabulares para mostrar los resúmenes de datos, la caracterización de la población de estudio, las mediciones de la variación alrededor de la mediana, así como las primeras pruebas de normalidad de los datos. Todo esto dio énfasis a la variable estadística y su distribución, puesto que se resaltó su carácter numérico para poder ser analizada, a la vez que su definición respondía a un contexto de interés, y se destacó su vinculación con la aleatoria. Sin embargo, muchos de estos estudios hicieron uso de inferencias informales, sin hacer uso de las herramientas que se gestaban en el campo probabilístico. Casi de manera simultánea a la formalización de la probabilidad, en el ámbito estadístico, Karl Pearson se preocupó por fundamentar, perfeccionar y ampliar los métodos usados principalmente por Quetelet y Galton, haciendo uso de los resultados generados en el área de probabilidad, pero también generando una matemátización propia. Esto comenzó a hacer explícita y formal la vinculación entre la variable aleatoria y la estadística. Otros autores como Gosset, Fisher, Neyman y Ergon Pearson generan herramienta que sustenta y consolida a la inferencia como una de las principales ramas de la estadística.

Así, desde problemas externos a la matemática se desarrolló lentamente la idea de variable estadística y su distribución, para poco a poco considerar experimentos aleatorios potenciales, donde los datos recogidos se ampliasen indefinidamente a una cierta población. La variable aleatoria sería, desde esta perspectiva, la generalización de una variable estadística cuando el número de datos es infinito. Además, no hay posibilidad de examinar a priori el espacio muestral del experimento, por lo que los parámetros y distribución de las variables aleatorias han de estimarse a partir de los correspondientes en las variables estadísticas. En consecuencia, la asignación de probabilidades es frecuencial (inferencia frecuencial) y se justifica mediante un proceso de inferencia, cuyos métodos se apoyan en las propiedades de las distribuciones de las variables subyacentes.

Independientemente, Bayes, con la formalización de su teorema y la aplicación de éste a la determinación de las probabilidades de las causas y estimación de proporciones, sienta las bases de lo que hoy en día es la inferencia Bayesiana.

A lo largo de la historia, se observan las distintas formas en cómo las dos vías a través de las cuales la variable aleatoria devino como objeto matemático se entrecruzaron y retroalimentaron mutuamente de diferentes formas. Por ejemplo, cuando el Duque de Toscana planteó un discordancia entre el comportamiento observado en los juegos de azar y un modelo errático de espacio muestral y se lo mostró a Galileo. A partir de este estímulo, se generó un modelo apropiado para el recuento teórico de los posibles resultados. O cuando Halley estudió grandes bases de datos poblacionales, lo que indujo al análisis de la esperanza de vida y estudios de mortandad. Otro ejemplo es cuando Jacob Bernoulli quería establecer el número de datos que harían que la probabilidad frecuencial convergiera a la probabilidad teórica, lo que genera una primera propuesta de la ley débil de los grandes números. Y, desde nuestro estudio, las vinculaciones más importantes se establecen cuando el uso de la distribución normal y diversas herramientas probabilistas en el análisis de datos en diversas áreas científicas, favoreció no sólo la generación de inferencias informales sino también de análisis propios de la estadística. Así mismo, cuando Pearson estableció un puente entre las ideas de Galton y resultados probabilísticos y matemáticos, favoreciendo la formalización de la inferencia estadística, así como la vinculación formal entre la variable aleatoria y estadística.

Sin embargo, estas dos corrientes de pensamiento también tuvieron largos periodos de reflexión propia, a través de los cuáles la interacción entre ellas fue casi nula. El desarrollo de la teoría de Laplace o la larga discusión que dio lugar a la comprobación del teorema central del límite (que tendría, sin embargo, una fuerte repercusión en la estadística) son momentos en los que la teoría prevalece sobre la práctica. Así mismo, en el análisis de grandes recuentos de datos a través de elementos meramente descriptivos, y sin embargo, con fines inferenciales, permaneció mucho tiempo sin la influencia de herramienta teórica que permitiera análisis más potentes.

Este desarrollo histórico podría constituir una hipótesis de estudio tendiente al análisis de la forma en cómo un estudiante puede apropiarse de la idea de la variable aleatoria y estadística, su vinculación entre ellas y, finalmente, la formalización de la relación entre ambas variables.

3. Sobre los estudios cognitivos

El segundo objetivo de nuestro trabajo (Sección 4 del Capítulo 1) se formuló en los siguientes términos:

Objetivo 2. *Llevar a cabo una exploración cognitiva de la comprensión intuitiva que muestran algunos estudiantes universitarios al trabajar con situaciones en las que se manifiesta la variable aleatoria y otros objetos matemáticos relacionados con ella, y que implican resolución de problemas y proyectos con una orientación hacia la modelación*.

Para cumplir con este objetivo, se realizaron dos tipos de estudios, cada uno de los cuáles fue diseñado y analizado sustentándonos en un análisis preliminar que se describe en cada uno de los capítulos 5 y 6. El análisis se sustentó en el análisis epistemológico disciplinar (Estudio 1) y en las conclusiones se confrontaron los resultados con las hipótesis previamente planteadas, siguiendo la metodología de ingeniería didáctica. También se analizó la idoneidad didáctica de las dos situaciones, que resultó bastante alta, aunque con algunos puntos de mejora. Nuestro objetivo general está vinculado con la búsqueda de indicios que sirvan de sustento en un futuro diseño de situaciones didácticas que auxilien en la enseñanza de la variable aleatoria, y todo el bagaje de conceptos unidos a ella, en contexto escolar.

3.1. La entrevista clínica

En el primero de dichos estudios (Estudio 3), se realizó una entrevista en profundidad a dos alumnas mientras resolvían un problema que involucraba a una variable aleatoria. El problema planteaba la toma de decisión en una situación en la que asignación de probabilidades era laplaciana y en la que se realiza un sorteo. Por tanto, este contexto se próxima a uno de los campos de problemas de los que surge históricamente la variable aleatoria: los juegos de azar, aunque involucra una decisión que tienta éticamente a las estudiantes. El análisis de las entrevistas se organizó alrededor de cuatro puntos principales:

- *Percepción de la aleatoriedad en la situación y en la solución del problema*. Aunque las alumnas tuvieron una buena comprensión intuitiva de la aleatoriedad en la situación planteada y explicitaron propiedades tales como impredecibilidad, convergencia y equiprobabilidad, en un inicio mostraron reticencia a aceptar una solución incierta, tratando, finalmente, de justificarla con argumentos, tales como la equidad. Las estudiantes usaron la probabilidad acumulada como un argumento útil para ser justas.

  Al tratar de encontrar una fórmula que expresara la distribución de probabilidad, sin embargo, asignaron la propiedad de aleatoria a la probabilidad en lugar de a la variable aleatoria, puesto que los valores que ésta podría tomar, podían ser conocidos previamente, en cambio el valor de la probabilidad resultaba ser incierto en el mismo contexto de una fábrica, pero en otra ciudad o país. Así, observamos, como al introducir ellas mismas la necesidad de una fórmula, se olvidan del contexto y al, abstraer, equivocan el concepto de aleatoriedad, vinculándolo con una incapacidad de cálculo.

- *Comprensión de la probabilidad definida en el espacio muestral y de la distribución de probabilidad*. Las alumnas usan la idea de probabilidad en forma natural, interpretándola como una medida de lo incierto, pero al usar el significado clásico de probabilidad, la consideran una razón y no un único valor numérico. Esto contribuye a que no vinculen directamente la probabilidad con el valor de la variable aleatoria, más que a través de los eventos compuestos lo que representa un obstáculo de descontextualización, relacionado con los sucesos compuestos.

- *La variable aleatoria en la solución del problema*, su carácter de función en el espacio muestral y la relación con la distribución de probabilidad. No pudimos observar la construcción de la regla de correspondencia de la variable aleatoria por parte de los estudiantes porque fue proporcionada en el problema, pero las alumnas pudieron reconocer que la regla de correspondencia vinculaba el espacio muestral con los valores de la variable aleatoria. Pudieron percatarse, también, que esa regla de correspondencia provenía de una característica de interés (extraída del planteamiento del problema) que agrupaba a los trabajadores en eventos compuestos.

  Encontramos algunas dificultades de descontextualización de la variable aleatoria en un primer estrato de modelación que son superadas cuando las estudiantes ven el valor de la variable aleatoria como el resultado del experimento aleatorio, conceptualización que Parzen (1971) sugiere para un curso introductorio de probabilidad. Sin embargo, esto también provoca que piensen que los valores de la variable aleatoria puedan ser colocados en un orden cualquiera en el eje-*x* de la gráfica porque son *aleatorios*. Al operar con los valores como otra «realidad», dejaron de ser entes abstractos y perdieron sus propiedades numéricas. Sin embargo las estudiantes interpretaron correctamente que la probabilidad se obtenía a partir de los eventos compuestos y que esto era lo que permitiría asignarle un valor de probabilidad a cada valor de la variable aleatoria. Así, concluimos que se tuvo una buena interpretación y manejo de los valores de variable aleatoria y aunque se pasaron por algunos obstáculos, lograron ser superados.

Las alumnas encontraron también natural asociar el término función a la distribución de probabilidad, debido a que la pudieron escribir muy naturalmente en notación funcional y porque que concordaba con la definición de función. También fue inmediato el uso de la representación tabular para las frecuencias y probabilidades. Sin embargo sí encontraron extraño que hubiera funciones con las características encontradas para la distribución de probabilidad del problema, lo que ocasiona que le adjudiquen el carácter aleatorio a la probabilidad en lugar de a la variable aleatoria. Sostenemos que esta afirmación proviene de la creencia de que las funciones forzosamente deben poder expresarse como una ecuación y que una función (expresión matemática) debe ser útil en varios contextos y no en uno solo.

- ❖ También se tuvieron en cuenta *los procesos de modelación* vinculados a la solución del problema. Las alumnas manifestaron un buen manejo de las herramientas matemáticas necesarias para resolver el problema, pero, al contrario de lo esperado, para ellas fue más difícil pasar del contexto del problema al contexto matemático que el proceso inverso. A pesar de eso, también en algunos pasajes desvinculan las matemáticas del contexto del problema, mostrando una concepción de las mismas como algo que no tiene por qué tener una relación con una situación del contexto real. Así mismo pudimos detectar que gran parte de esta problemática es engendrada por la modelación en estratos para pasar del contexto del problema hasta la obtención de la función de distribución.

También pudimos observar las distintas soluciones por las que transitan las estudiantes para llegar a una conclusión que ellas puedan argumentar satisfactoriamente y los motivos por los que buscan una nueva solución. Las estudiantes entran varias veces en el ciclo interrogativo de Wild y Pfannkuch (1999), lo que también hace que hagan uso de herramientas matemáticas que inicialmente no contemplaron. Así mismo, es notorio cómo el contexto del problema condiciona la discusión y las soluciones de las estudiantes.

3.2. La evaluación de un proyecto abierto

En el segundo de dichos estudios (Estudio 4), se examinan las producciones escritas de 111 estudiantes que trabajaron individualmente en un proyecto abierto de análisis de datos que involucraba tres pares de variables aleatorias y sus correspondientes variables estadísticas. Se trata de que los estudiantes recorran el ciclo completo de modelación (Dantal, 1997) para responder a una pregunta sobre las intuiciones sobre el azar de los estudiantes. Para ello se organiza un experimento en el que los mismos estudiantes son partícipes en la recolección de los datos y, parcialmente, en el planteamiento de las variables. El análisis estadístico necesario será decidido por los mismos estudiantes, esperando que obtengan conclusiones sobre las intuiciones sobre el azar del grupo. El análisis de los informes escritos elaborados por los estudiantes se organizó alrededor de los puntos siguientes:

- *Comprensión de la aleatoriedad y su percepción en el experimento.* Se realizó una doble evaluación: globalmente fue posible analizar los datos de las secuencias simuladas para inferir indirectamente las características de estas concepciones; individualmente fue posible observar las concepciones de los estudiantes manifestadas en sus argumentos al analizar las intuiciones sobre el azar de su grupo completo.

  En el primer análisis, se observó que los estudiantes en su mayoría tienen una buena percepción del valor esperado en la distribución binomial, pero poca percepción del valor esperado en variables asociadas con las rachas. Esto concuerda con los resultados de investigaciones previas, como la de Serrano (1999) y Toohey (1995). Los estudiantes tienen problemas para percibir la independencia de ensayos sucesivos produciendo rachas muy cortas. Individualmente, se encontraron ejemplos de concepciones parcialmente correctas que corresponden a las sostenidas en diferentes periodos históricos sobre la aleatoriedad (aleatoriedad como contrapuesto a causa; aleatoriedad como variabilidad, ausencia de control, impredecibilidad o equiprobabilidad) y algunos presentan la ilusión de control (Langer, 1975).

De manera general, observamos que el conocimiento que los estudiantes tienen sobre el contexto (en este caso, sus creencias), condicionó la generación de las secuencias que simulaban el lanzamiento de una moneda. También el análisis que hicieron de los datos generados por ellos mismos, motivó que ellos hicieran patentes sus creencias sobre la aleatoriedad, puesto que hicieron uso de ellas para argumentar los resultados encontrados en sus datos.

- *Comprensión de la variable estadística y su distribución:* Se ha estudiado la forma en que los estudiantes trabajan con las variables estadísticas, sus distribuciones y momentos. Más de las dos terceras partes de los estudiantes formaron la distribución de las variables estadísticas que intervenían en el proyecto y la representaron mediante una tabla, gráfica o ambas, a pesar de que no se les pedía explícitamente. De ello se deduce un primer nivel intuitivo en la comprensión de la idea de la distribución. Sin embargo, fueron muy pocos los estudiantes que elaboraron gráficos o sus tablas conjuntas o con una escala apropiada para hacer una comparación rápida de las distribuciones de las secuencias simuladas con las de las reales.

  Poco menos de la tercera parte no llegaron a formar la distribución tratando de comparar los resultados individuales alumno a alumno y analizaron sólo sus datos, ya sea porque no comprendieron el objetivo del proyecto o por falta de conocimiento estadístico. Recordemos que cada estudiante trabajó de manera individual, lo que no permitió la interacción con sus compañeros, ni con el profesor.

  Así mismo, encontramos algunos errores vinculados con la comprensión de la variable que afectan la concreción de su distribución, entre ellos destaca la dificultad de las variables vinculadas con las rachas, que no eran conocidas para los estudiantes. Este obstáculo se agranda por su falta de familiaridad con experimentos aleatorios compuestos, puesto que la definición de las tres principales variables de análisis no estaba vinculada con el lanzamiento de una moneda sino con el lanzamiento de 20 monedas. Esto explicaría porque algunos estudiantes se restringieron al análisis de la variable «cara o cruz» o incluso a sus propios datos. En cualquier caso, también es cierto que los estudiantes se concentran más en la graficación del número de caras que en las otras dos, en cambio, en general, los que elaboraron tablas, lo hicieron para las tres variables.

La comprensión y uso de medidas de tendencia central y su interpretación fue mucho mayor que las de dispersión que, en caso de ser calculadas, no se interpretan, salvo algunas excepciones. La comprensión de la idea de dispersión no parece intuitiva para los estudiantes, quienes también repiten los errores en sus representaciones gráficas que han sido descritos en investigaciones previas como las de Bruno y Espinel (2005), Espinel (2007) y Arteaga (2011).

- ❖ *Percepción y referencia a las variables aleatorias y relación con la variable estadística.* Se ha analizado los argumentos en que los estudiantes ligan las variables a los resultados de un experimento aleatorio o hacen referencia a probabilidades de valores de la variable. Uno de los más frecuentes es la referencia espontánea al valor esperado de la variable aleatoria binomial, «número de caras al lanzar 20 monedas», además de que casi todos estiman correctamente su valor, resultado que concuerda con el hecho de que la esperanza matemática es el primer desarrollo en la historia de la variable aleatoria. También con frecuencia ligan la variable estadística y aleatoria, explicando los resultados en la variable estadística por medio de la probabilidad teóricas en el lanzamiento de la moneda. Por el contrario, el resto de variables aleatorias en el proyecto (número y longitud de rachas) no parecen intuitivas para los estudiantes, puesto que son poco mencionadas. También en casos aislados hay referencia a la convergencia de la distribución de frecuencias a la distribución de probabilidad, es decir, explícitamente se realiza un proceso de inferencia informal (Rossman, 2008) que es implícito en el resto de los estudiantes. Concluimos que la actividad es potencialmente útil para introducir las variables aleatorias, como generalizaciones de las variables estadísticas.
- ❖ *Ciclo de modelación y razonamiento estadístico.* Se han analizado los diferentes pasos en el ciclo de modelación de acuerdo con Dantal (1999) y el razonamiento estadístico según el modelo de Wild y Pfannkuch (1999), así como el modelo por estratos establecido en el análisis disciplinar. Los dos primeros pasos del ciclo son sugeridos por el profesor, aunque algunos estudiantes espontáneamente sugieren el número de caras en los 20

lanzamientos como variable que con la que se pueden estudiar los datos para resolver la cuestión planteada.

En los pasos 3 y 4, los estudiantes muestran buena capacidad para construir y trabajar con diferentes modelos matemáticos sencillos (tablas, gráficos y estadísticos resúmenes), aunque en algunos de ellos aparecen errores matemáticos o de simbolización. Sin embargo, encontramos una gran dificultad de los estudiantes en el último paso del ciclo de modelización, pues no tradujeron los resultados matemáticos al contexto del problema. Una tercera parte de los estudiantes sí llegaron a comparar las variables de las secuencias reales y simuladas, pero pocos llegaron a concluir sobre las intuiciones de la aleatoriedad de los estudiantes. Esto nos dio indicios de que no les fue claro cuál era la pregunta de investigación que tenían que contestar en el proyecto o el papel que jugaban las variables aleatorias (y su distribución) en la respuesta a esa pregunta. Por tanto, la definición de las variables (en el paso 2 de la modelación) juega un papel importante en la realización del proyecto, puesto que su incomprensión hace que, aun cuando hayan realizado un trabajo matemático bueno, el ciclo de modelación se quede trunco.

Hay que tomar en cuenta que los estudiantes de nuestra muestra tenían poco trabajo previo con proyectos, sólo una parte de ellos habían realizado uno en el curso anterior. Además de que no se fomenta este tipo de trabajo en otras ramas de las matemáticas. Así que esta falta de experiencia con proyectos estadísticos también influye en la incomprensión de la utilidad del análisis estadístico para resolver problemas reales y su capacidad para vincular el trabajo matemático con las variables y el contexto del problema. Ello concuerda con lo señalado por Dantal (1997) y Chaput, Girard y Henry (2011) que sugieren que es necesario recorrer con los estudiantes todas las fases del ciclo de modelización en la clase de matemáticas y no sólo centrarse en el trabajo matemático.

Así mismo, observamos que muchos estudiantes se dejan llevar por su conocimiento sobre el contexto también en el análisis estadístico. Es decir, dan preferencia a sus creencias por sobre los resultados de su análisis. Esto también se ha observado en otras investigaciones, por ejemplo Estepa y Batanero (1994) donde los estudiantes también siguen sus creencias al resolver algunos problemas de asociación en tablas de contingencia.

Por último, también encontramos argumentos sin sentido o gráficos y cálculos que los estudiantes no usan ni leen. Lo que nos hace pensar que, ante la incomprensión del objetivo del proyecto, los estudiantes recurren a contestar lo que creen que el profesor espera de ellos.

En este sentido, Makar, Bakker y Ben Zvi (2011) proponen fomentar el razonamiento inferencial informal a través de la inserción de experiencias en donde se favorezca el aprendizaje basado en la indagación colaborativa, inserción de conocimiento estadístico dentro de un contexto y en la profundización de los conflictos que surgen entre la respuesta intuitiva del estudiante y su interpretación del análisis de datos.

Como se puedo observar, a pesar de la diferencia de perspectiva de ambos, los estudios cognitivos presentaron puntos en común, referidos principalmente al proceso de modelación involucrado en la solución de los problemas presentados tanto en la entrevista como en el proyecto.

4. Aportaciones y limitaciones de la investigación

Consideramos que nuestro trabajo realiza aportaciones de interés al campo de la educación estadística, donde la investigación sobre la variable aleatoria es prácticamente inexistente.

Una primera aportación viene dada por los estudios epistemológicos descritos. Pues, por un lado, hemos realizado un análisis detallado del objeto matemático «variable aleatoria» y sus relaciones con otros objetos, además de constituir una base para el análisis a priori en la metodología de ingeniería didáctica, sienta las bases para describir el significado de referencia del objeto, dentro del enfoque onto-semiótico. Asimismo, el estudio epistemológico-histórico nos permitió identificar algunos campos de problemas de donde surge la variable aleatoria y proporciona hipótesis sobre posibles obstáculos y dificultades de los estudiantes. También identificamos saltos cualitativos entre las etapas históricas, que pueden contribuir a un diseño para el desarrollo cognitivo de la variable aleatoria.

La exploración cognitiva, realizada mediante dos estudios de diferente metodología, ha proporcionado datos muy detallados de las dificultades de los estudiantes, su comprensión de diversos objetos matemáticos ligados con la variable aleatoria, sus estrategias y uso de dichos objetos en la resolución de problemas y su capacidad de modelación. Todos estos datos son fundamentales en la labor del profesor, que ha de planificar la enseñanza teniendo en cuenta el aprendizaje de los estudiantes. Del mismo modo proporcionan una base para una posterior construcción de la ingeniería didáctica para continuar esta investigación.

Dentro del campo de la modelación, retomamos la modelación por estratos de Heitele (1975) y basándonos en el modelo de Wild y Pfannkuch (1999) propusimos un modelo de los procesos que se siguen para concretar una distribución de probabilidad bajo distintas asignaciones de probabilidad (laplaciana y frecuencial) y en una inferencia informal. En este modelo se resalta el papel que juega la variable aleatoria y se usa en ambos estudios cognitivos para explicar algunas dificultades que los estudiantes tuvieron al contextualizar los resultados matemáticos y, por lo tanto, completar el ciclo de modelación.

La conjugación de estudios cognitivos y epistemológicos permitió mostrar que el desarrollo del pensamiento estadístico supone dos procesos aparentemente contradictorios. Por un lado se requiere el dominio de herramienta matemática abstracta y por otro es necesaria una ruptura con el pensamiento determinista que permita aceptar otro tipo de conocimiento que se caracteriza por aportar resultados con un grado de incertidumbre. Estos dos procesos se combinan fuertemente en la variable aleatoria.

El instrumento diseñado (entrevista), así como la metodología de análisis descrita en cada uno de los capítulos 4 y 5 constituyen también aportaciones para la investigación futura, pues podrían aplicarse en otras investigaciones. Finalmente, tanto el estado de la cuestión, como la bibliografía incluida en la memoria apoyarán la investigación futura sobre la variable aleatoria.

Como en cualquier investigación limitada cronológica y geográficamente, queremos reconocer también algunos puntos en que podría ser mejorada.

Una de las principales limitaciones del estudio de exploración cognitiva descrito en el Capítulo 5 es el tamaño de muestra, que se reduce a dos estudiantes. Esta limitación se asumió conscientemente, pues se prefirió realizar un trabajo de gran profundidad de análisis a expensas del tamaño muestra. La decisión pensamos fue correcta, por la cantidad y variedad de conclusiones e hipótesis surgidas del estudio, que, por este motivo tiene una alta validez. No podemos, por el contrario, proporcionar información de si las dificultades descritas se presentarán en otros estudiantes o con qué frecuencia. Para analizar la fiabilidad de este estudio, se tendría que replicar la entrevista en otros estudiantes, o bien diseñar un cuestionario escrito, basado en los resultados de la entrevista y que sirviese para confirmar la existencia de estas dificultades en otros estudiantes.

En el segundo estudio cognitivo se amplía el tamaño de muestra, a expensas de pasar a un método de obtención de datos menos potente que la entrevista. Aun así, el análisis de los protocolos recogidos de los estudiantes ha confirmado la existencia de algunas de las dificultades descritas en la entrevista en otros estudiantes, incluso en un contexto educativo y geográfico diferente. Asimismo proporciona descripción de otras dificultades relacionadas con la variable estadística, sus parámetros y el proceso de modelización. También en este segundo caso, pensamos que sería necesario un mayor tamaño de muestra para la confirmación de las conclusiones, como un estudio a profundidad (entrevista) para conocer con mayor detalle el porqué de los errores cometidos por algunos estudiantes. En particular, la indagación de porque algunos estudiantes se quedan en el análisis de sus propios datos o no conforman la distribución y cómo superarlo sería de gran ayuda para la elaboración de proyectos en clase tomando en cuenta estos errores. Otro estudio a profundidad necesario es la indagación del rompimiento en el ciclo de modelación por la no conclusión en el contexto del problema.

5. Algunas ideas para investigaciones futuras

En esta Memoria hemos iniciado la fase preliminar de la ingeniería didáctica para el objeto matemático «variable aleatoria». Aunque se han aportado resultados importantes, tanto de los estudios epistemológicos como de los estudios cognitivos realizados, es necesario continuar la investigación, para que sus implicaciones puedan llevarse a cabo al aula.

En primer lugar, sería necesario diseñar una secuencia de enseñanza del tema, identificando las situaciones fundamentales de las cuáles surge la variable aleatoria y la comprensión de sus propiedades y relaciones con otros objetos. En dichas situaciones fundamentales sería también necesario prestar atención a las competencias procedimental y argumentativa de los estudiantes, para atender a los diferentes elementos del significado de la variable aleatoria.

La investigación realizada aporta algunos datos en este sentido, pues hemos identificado dos tipos principales de situaciones didácticas fundamentales, que, debidamente relacionadas, complementadas y secuenciadas pueden constituir la base de un primer encuentro de los estudiantes con la variable aleatoria:

- *Situaciones de toma de decisión en un ambiente aleatorio.* Un ejemplo de dichas situaciones se ha utilizado en el diseño e implementación de la entrevista y la exploración cognitiva, descrita y analizada en el Capítulo 5.

- *Situaciones de análisis de datos empíricos por medio de modelos matemáticos que permita dar respuesta a un problema de investigación (situaciones de inferencia en un ambiente aleatorio)*: En el Capítulo 6 hemos utilizado una de estas situaciones, dentro de un proyecto abierto de análisis de datos que ha llevado a los estudiantes a comparar tres pares de variables estadísticas e implícitamente a considerar modelos matemáticos (variables aleatorias) subyacentes en un experimento.

Una hipótesis factible sería que los estudiantes se apropian del concepto a través de la constante interacción entre estos dos tipos de situaciones y otras situaciones no consideradas en este estudio (por ejemplo, la revisión de probabilidades a la luz de una nueva información, utilizada en la inferencia Bayesiana). Tal diseño, entonces, deberá contemplar las interacciones entre los contextos estadísticos y probabilísticos que, detectamos, se han dado a lo largo del surgimiento del concepto de variable aleatoria.

Una vez diseñada la secuencia de enseñanza, sería necesario llevarla a cabo en uno o varios grupos naturales de clase, observando su desarrollo, para comparar el significado realmente implementado en la clase, con el significado pretendido en la secuencia instruccional.

Asimismo sería necesario la construcción de instrumentos de evaluación del aprendizaje de la variable aleatoria, que se basen en los ya utilizados en este trabajo, pero que contemple la totalidad del significado de referencia del objeto variable aleatoria. Es decir, que tengan suficiente validez de contenido. Dichos instrumentos, aplicados a muestras de estudiantes después de la enseñanza del tema, permitirían comprobar si las dificultades descritas en esta Memoria están extendidas en otros estudiantes.

Esta investigación también abre nuevas oportunidades de estudio en la comprensión de variables aleatorias en diversas situaciones, por ejemplo en las distribuciones particulares, siguiendo estudios previos como el de Tauber (2001) sobre la distribución normal. O bien, la operatividad de variables aleatorias, en especial, los estadísticos y distribuciones del muestreo de una y varias variables y la inferencia estadística para el caso de los estudiantes universitarios. Algunos estudio base serían los Alvarado (2007) y Olivo (2008).

Otra vía de investigación es desarrollar una propuesta curricular que considere la enseñanza de la variable aleatoria como una idea fundamental, tal como propone Heitele (1979). De acuerdo con nuestros análisis epistemológicos histórico y disciplinar de la variable aleatoria, la complejidad de este concepto matemático hace que las ideas vinculadas con la variable aleatoria requieren de mucho tiempo para ser construidas. Por ello, podría resultar interesante el estudio de su introducción desde las primeras etapas de desarrollo del niño siguiendo la línea de Heitele. El principio básico sería que no se puede ser exhaustivo en ninguno de los niveles, pero sí se pueden enseñar prefiguraciones del concepto en cada nivel escolar de manera provechosa, de modo que proporcionen un significado redondo que puedan ser retomado en niveles escolares superiores para incrementar su profundidad. Estas prefiguraciones podrían estar inspiradas en las etapas históricas marcadas en nuestro estudio.

Nuestro trabajo también hace acentúa la importancia de la vinculación entre las variables aleatoria y estadística en la inferencia informal y la importancia del conocimiento del contexto en la realización de estas inferencias, cuestiones que puede ser retomadas a la luz de muy recientes investigaciones que han surgido alrededor de estos temas.

8. Referencias Bibliográficas

Alatorre, S. (1998). Acerca del tratamiento didáctico de la probabilidad. *Correo del Maestro 26*, 23-35.

Alquicira, Z. M. (1998). *Probabilidad: Docencia y praxis. Hacia una fundamentación epistemológica.* Tesis de Maestría. Cinvestav, México.

Alvarado, H. (2007). *Significados del teorema central del límite en la enseñanza de la estadística en ingeniería.* Tesis Doctoral. Universidad de Granada, España.

Arteaga, P. (2011). *Los gráficos estadísticos en la formación de profesores de educación primaria: evaluación de conocimiento matemático-didáctico en la realización y valoración de un proyecto de análisis de datos.* Tesis Doctoral. Universidad de Granada, España.

Arteaga, P., Batanero, C. y Ruiz, B., (2008). Complejidad semiótica de gráficos producidos por futuros profesores en la comparación de dos muestras. En L. Blanco y cols. (Eds.). *Actas del XII Simposio de las Sociedad Española de Investigación en Educación Matemática. Comunicaciones en los grupos de trabajo.* CD- ROM. Badajoz: SEIEM

Arteaga, P., Batanero, C. y Ruiz, B. (2009). Comparación de distribuciones por futuros profesores. En M. J. González, M. T. González y J. Murillo (Eds.), *Investigación en Educación Matemática XIII* (pp. 129-138). Santander: SEIEM.

Arteaga, P., Batanero, C. y Ruiz, B. (2010). Pre-service primary school teachers' perception of randomness. En M. M. F Pinto y T. F. Kawasaki (Eds), *Proceedings of the 34th Conference of the International Group for the Psychology of Mathematics Education* (Vol 2, pp. 169-176). Belo Horizonte, Brasil: PME group.

Arteaga, P., Ruiz, B. y Batanero, C. (2008). Complejidad semiótica de gráficos estadísticos en la comparación de dos distribuciones por futuros profesores. Trabajo presentado en el *Encuentro Latino Americano de Educación Estadística.* Monterrey, México, Julio, 2008.

Artigue, M. (1990). Epistémologie et didactique. *Recherches en Didactique des Mathématiques*, 10(2-3), 241-286.

Artigue, M. (1995a). Ingeniería didáctica. En P. Gómez (Ed.), *Ingeniería didáctica en educación matemática* (pp. 39-59). México: Iberoamericana.

Artigue, M. (1995b). Une recherche d'ingénierie didactique sur l'enseignement des mathématiques. Trabajo presentado en la *Reunión Recherches en Didactique des mathématiques et de l'Informatique.* Grenoble: IMAG-LSD.

Ayers, T., Davis, G., Dubinsky, E. y Lewin, P. (1988). Computer experiences in learning composition of functions. *Journal for Research in Mathematics Education, 19* (3), 246-259.

Azcárate, P. (1995). *El conocimiento profesional de los profesores sobre las nociones de aleatoriedad y probabilidad. Su estudio en el caso de la educación primaria*. Tesis Doctoral. Universidad de Cádiz, España.

Bakker, A. y Gravemeijer, K. P. E. (2004). Learning to reason about distribution. En J. Garfield y D. Ben-Zvi (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp 147-168). Dordrecht: Kluwer.

Batanero, C. (2001). *Didáctica de la estadística*. Granada: Grupo de investigación en educación estadística.

Batanero, C. (2002). Los retos de la cultura estadística. Conferencia inaugural. *Jornadas Interamericanas de Enseñanza de la Estadística.* Buenos Aires, Argentina. Online: www.ugr.es/~batanero/.

Batanero, C. (2003). La simulación como instrumento de modelización en probabilidad. *Educación y Pedagogía*, *35,* 37-64.

Batanero, C. (2005). Significados de la probabilidad en la educación secundaria. *Relime 8*(3). 247-264.

Batanero, C., Arteaga, P. y Ruiz, B. (2009). Statistical graphs produced by prospective teachers in comparing two distributions. En V. Durand-Guerrier, S. Soury-Lavergne y F. Arzarello (Eds.), *Proceedings of the Sixth Congress of the European Society for Research in Mathematics Education*. Lyon, Francia: ERME. Online: www.inrp.fr/editions/editions-electroniques/cerme6/.

Batanero, C., Arteaga, P. y Ruiz, B. (2010). Análisis de la complejidad semiótica de los gráficos producidos por futuros profesores de educación primaria en una tarea de comparación de dos variables estadísticas. *Enseñanza de las Ciencias, 28*(1), 141-154.

Batanero, C., Arteaga, P., Ruiz, B. y Roa, R. (2010). Assessing pre-service teachers conceptions of randomness through project work. En C, Reading (Ed.), *Proceedings of the Eight International Conference on Teaching* Statistics. Lubjana: International Association for Statistical Education. Online: www.stat.auckland.ac.nz/~iase/publications.

Batanero, C. Arteaga, P., Serrano, L. y Ruiz. B. (2013). Prospective primary school teachers' perception of randomness. En E. J Chernoff, y B. Sriraman, (Eds.), Probabilistic thinking: presenting plural perspectives. Advances in Mathematics Education Series. Springer, en prensa.

Batanero, C. Burrill, G. y Reading, C. (Eds.) (2011). *Teaching statistics in school mathematics- challenges for teaching and teacher education. A Joint ICMI/IASE Study*. New York: Springer.

Batanero, C. Burrill, G., Reading, C. y Rossman, A. (2008). *Proceedings of the Joint ICMI /IASE Study Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education.* Monterrey, México: ICMI e IASE. CD ROM.

Batanero, C. y Díaz, C. (2007). Probabilidad, grado de creencia y proceso de aprendizaje. *Actas de las XIII Jornadas Nacionales de Enseñanza y Aprendizaje de las Matemáticas* [CD-ROM]. Granada: Federación de Sociedades de Profesores de Matemáticas.

Batanero, C., Estepa, A. y Godino, J. D. (1997). Evolution of students' understanding of statistical association in a computer-based teaching environment. En J. B. Garfield y G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics. 1996 IASE Round Table Conference* (pp. 183-198). University of Minnesota: IASE.

Batanero, C., Henry, M. y Parzysz, B. (2005). The nature of chance and probability. En G. Jones (Ed.). *Exploring probability in school: challenges for teaching and learning* (pp. 15-37). Dordrecht: Kluwer.

Batanero, C. y Serrano, L. (1995). La aleatoriedad, sus significados e implicaciones educativas. *Uno*, *5*, 15-28.

Batanero, C. y Serrano, L. (1999). The meaning of randomness for secondary students. *Journal for Research in Mathematics Education*, *30*, 558-567.

Batanero, C., Serrano, L. y Green, D. R. (1998). Randomness, its meanings and implications for teaching probability. *International Journal of Mathematics Education in Science and Technology*, 29 (1), 113-123.

Batanero, C., Tauber, L. y Sánchez V. (2001). Significado y comprensión de la distribución normal en un curso introductoria de análisis de datos. *Quadrante*, *10* (1), 59-92.

Batanero, C., Tauber, L. y Sánchez, V. (2004). Student’s reasoning about the normal distribution. En D. Ben-Zvi y J. B. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 257-276). Dordrecht: Kluwer.

Beltrami, E. (1999). *What is random? Chance and order in mathematics and life*. New York: Copernicus, Springer-Verlag.

Bellhouse, D. R. (2000). De Vetula: a medieval manuscript containing probability calculations. *International Statistical Review*, *68*(2), 123-136.

Bennett, D. J. (1998). *Randomness.* Cambridge: Harvard University Press.

Bernoulli, D. (1954). Exposition of a new theory on the measurement of risk *Econometrica*, 22 (1), 23-36. (Trabajo original publicado en 1738), Bernoulli, J. (2006). The art of conjecturing. En E. D. Sylla (Ed.). *Ars conjectandi together with letter to a friend on sets in court tennis* (pp 127-340). Baltimore: The Johns Hopkins University Press (Trabajo original publicado en 1713).

Blum, W., Galbraith, P. L., Henn, H. W. y Niss, M. (2007). *Modelling and applications in Mathematics Education. The 14th ICMI Study*. New York: Springer.

Boudot, M. (1979). *Lógica inductiva y probabilidad*. Madrid: Paraninfo.

Borovcnik, M., Bentz, H. J. y Kapadia, R. (1991a). A probabilistic perspective. En R. Kapadia y M. Borovcnik (Eds), *Chance encounters: Probability in education* (pp 27-71). Dordrecht, Netherlands: Kluwer.

Borovcnik, M., Bentz, H. J. y Kapadia, R. (1991b). Empirical research in understanding probability. En R. Kapadia y M. Borovcnik (Eds.), *Chance encounters: probability in education* (pp. 73-105). Dordrecht: Kluwer.

Brousseau, G. (1981). *Los obstáculos epistemológicos y los problemas en matemáticas*. México: Cinvestav.

Brousseau, G. (1986). Fondements et méthodes de la didactique des mathématiques. *Recherches en Didactique des Mathématiques¸ 7*(2), 33-115.

Brousseau, G. (1997). *Theory of didactical situations in mathematics. Didactique des mathématiques, 1970-1990*. Dordrecht, Netherlands: Kluwer.

Bruner, J. S. (1960). *The process of education*. Massachusetts: Cambridge.

Bunge, M. (1973). *La investigación científica*. Barcelona: Ariel.

Bruno, A. y Espinel, M. C. (2005). Recta numérica, escalas y gráficas estadísticas: un estudio con estudiantes para profesores. *Formación del Profesorado e Investigación en Educación Matemáticas 7*, 57-85.

Cantoral, R. y Farfán, R. (1998). Pensamiento y lenguaje variacional en la introducción al análisis. *Épsilon*, 42, 353-369.

Cardano, G. (1953). The book on games of chance. En O. Ore (Ed.), *Cardano: the Gambling Scholar* (pp 181-241). Princeton: Princeton University Press. (Trabajo original publicado en 1663).

Castillo, J. y Gómez, J. (1998). *Estadística inferencial básica*. México: Grupo Editorial Iberoamérica.

Chaput, B., Girard, J. C. y Henry, M. (2011). Modeling and simulations in statistics education. En C. Batanero, G. Burrill y C. Reading (Eds.) (2011). *Teaching Statistics in School Mathematics- Challenges for Teaching and Teacher Education. A Joint ICMI/IASE Study* (pp. 85-95). New York: Springer.

Chernoff, E. (2009a). *Subjective probabilities derived from the perceived randomness of sequences of outcomes.* Unpublished Ph. D. Simon Fraser University.

Chernoff, E. (2009b). Sample space partitions: An investigative lens. *Journal of Mathematical Behavior, 28*, 19–29.

Chevallard, Y. (1985). *La transposición didáctica*. Argentina: Aique

Chevallard, Y. (1991) Dimension instrumentale, dimension sémiotique de l'activé mathématique. Trabajo presentado en el *Séminaire de Didactique des Mathématiques et de l'informatique de Grenoble*. Grenoble: Université Joshep-Fourier,

Chevallard, Y. (1992). Concepts fondamentaux de la didactique: perspectives apportées par un apache anthropologique. *Recherches en Didactique des Mathématiques, 12* (1), 73-112.

Chevallard, Y., Bosch, M. y Gascón, J. (1997). Estudiar matemáticas. *El eslabón perdido entre la enseñanza y el aprendizaje*. Barcelona: Horsori.

Ciancetta, M. (2007). *Students' reasoning when comparing distributions of data.* Tesis doctoral, Portland State University, Estados Unidos.

Clement, J. (2000). Analysis of clinical interviews: Foundations and model viability. En A. E. Kelly y R. A. Lesh (Eds), *Handbook of research design in mathematics and science education* (pp 547-589). London: Lawrence Erlbaum.

Cobo, B. (2003). *Significado de las medidas de posición central para los estudiantes de Secundaria*. Tesis Doctoral. Universidad de Granada, España.

Confrey, J. (1980). Clinical interview: Its potential to reveal insights in mathematical education. En R. Karplus (Ed.), *Proceedings of the 4th International Conference for the Psychology of Mathematics Education* (pp. 400-408). Berkeley, C.A: PME-NA.

Contreras, J. M. (2011). *Evaluación de conocimientos y recursos didácticos en la formación de profesores sobre probabilidad condicional.* Tesis Doctoral. Universidad de Granada, España.

Coutinho, C. (2001). *Introduction aux situations aléatoires dès le Collège: de la modélisation à la simulation d'expériences de Bernoulli dans l'environnement informatique Cabri-géomètre-II*. Tesis Doctoral. Universidad de Grenoble.

Cuadras, C. (1999). *Problemas de probabilidades y estadística.* (Vol. 1). Barcelona: EUB.

Dantal, B. (1997). Les enjeux de la modélisation en probabilité. En *Enseigner les probabilités au lycée* (pp. 57-59). Reims: Commission Inter-IREM.

De la Cruz, A. T. (2007). *Un estudio sobre la construcción social de la noción de promedio en un contexto probabilístico*. Tesis de Maestría. Centro de Investigación en Ciencia Aplicada y Tecnología Avanzada. México.

Devore, J. (2011). *Probabilidad y estadística para ingeniería y ciencias*. México: Cengage Learning.

Doob, J. L. (1976). Foundations of probability theory and its influence on the theory of statistics. En D. B. Owen. *On the history of statistics and probability: proceedings of a symposium on the American mathematical heritage* (pp. 196-204). New York: Dekker.

Doob, J. L. (1986). The central limit theorem around 1935: Comment. *Statistical Science*, *1*(1), 93-94.

Engel, J. y Sedlmeier, P. (2005). On middle-school students' comprehension of randomness and chance variability in data. *Zentralblatt fur Didaltik der Mathematik*, *37*(3), 168-177.

Engelke, N., Oehrtman, M. y Carlson, M. (2005). Composition of functions: Precalculus students' understandings. En G. M. Lloyd, M. Wilson, J. L. Wilkins y S. L. Behm (Eds). *Proceedings of the 27th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* [CD_ROM]. Virginia Technological University.

Estepa, A. y Batanero, C. (1994). Concepciones iniciales sobre la asociación estadística. *Enseñanza de las Ciencias*, *13*(2), 155-170.

Espinel, C. (2007). Construcción y razonamiento de gráficos estadísticos en la formación de profesores. *Investigación en Educación Matemática* 11, 99-119.

Estrada, A., Batanero, C. y Fortuny, J. M. (2004). Un estudio de evaluación de conocimientos estadísticos en profesores en formación e implicaciones didácticas. *Educación Matemática*, 16, 89-112.

Falk, R. (1981). The perception of randomness. In *Proceedings of the fifth international conference for the psychology of mathematics education* (Vol. 1, pp. 222–229). Grenoble, France: Laboratoire IMAG.

Falk. R. (1986). Misconceptions of statistical significance. *Journal of Structural Learning*, 9, 83-96.

Falk, R. y Konold, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review*, *104*, 301-318.

Feller, W. (1989). *Introducción a la teoría de probabilidades y sus aplicaciones* (Vol. 1). México: Limusa.

Fischbein, E. (1975). *The intuitive sources of probabilistic thinking in children*. Dordrecht: D. Reidel.

Fischbein, E. y Gazit, A. (1984). Does the teaching of probability improve probabilistic intuitions?. *Educational Studies in Mathematics*, *15*(1), 1-24.

Fischbein, E., Nello, M.S. y Marino, M.S. (1991). Factors afffecting probabilistic judgments in children and adolescents. *Educational Studies in Mathematics*, 22(6), 523-549.

Fischer, H. (2011). A history of the central limit theorem from classical to modern probability theory. New York: Springer.

Fisher, R. A. (1918). The correlation between relatives on the supposition of mendelian inheritance. *Royal Society of Edinburgh from Transactions of the Society*, 52, 399-433.

Fisher, R.A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London*, A, 222, 309-368.

Fisher, R.A. (1930). Inverse probability. *Proceedings of the Cambridge philosophical society*, 26, 528-535.

Friel, S., Curcio, F. y Bright, G. (2001). Making sense of graphs: critical factors influencing comprehension and instructional implications. *Journal for Research in mathematics Education* 32(2), 124-158.

Gal, I. (2002). Adult's statistical literacy. Meanings, components, responsibilities. *International Statistical Review*, *70* (1), 1-25.

Galton, F. (1889). *Natural inheritance.* London: Macmillan.

Galton, F. (1901). Biometry. *Biometrika*, *1*(1), 7-10.

Galton, F. (1908). *Memories of my life*. London: Methuen

García, M. (1971). Desarrollo histórico de la teoría de probabilidad. *Estadística Española*, *71*, 51-86. Online: www.ine.es/revistas/estaespa/estaespa51a60.htm

García, M. (2007). *Resignificando el concepto de función lineal en una experiencia de educación a distancia*. Tesis de Maestría. México: CICATA-IPN, México.

Gattuso, L. y Ottaviani, G. (2011). Complementing mathematical thinking and statistical thinking in school mathematics. En C. Batanero, G. Burrill y C. Reading (Eds.) (2011). *Teaching Statistics in School Mathematics- Challenges for Teaching and Teacher Education. A Joint ICMI/IASE Study* (pp. 121-132). New York: Springer.

Ghiglione, R. y Matalón, B. (1989). *Les enquêtes sociologiques. Théorie et practique.* París: Armand Colin.

Gimeno, J. (1984). *Teoría de la enseñanza y desarrollo del currículo*. Madrid: Paraninfo.

Godino, J. D. (1996). Mathematical concepts, their meaning and understanding. En L. Puig y A. Gutierrez (Eds.), *Proceedings of XX Conference of the International Group for the Psychology of Mathematics Education.* (Vol. 2, pp. 417-425). Valencia, España: PME.

Godino, J. D. (2002). Un enfoque ontológico y semiótico de la cognición matemática. *Recherches en Didactique des Mathématiques*, *22*(2/3), 237-238.

Godino, J. D. y Batanero, C. (1994). Significado institucional y personal de los objetos matemáticos. *Recherches en Didactique des Mathématiques*, *14*(3), 325-355.

Godino, J. D. y Batanero, C. (1998). Clarifying the meaning of mathematical objects as a priority area of research in mathematics education. En A. Sierpinska y J. Kilpatrick (Eds.), *Mathematics Education as a research domain: A search for identity* (pp. 177-195). Dordrecht: Kluwer.

Godino, J. D., Batanero, C. y Cañizares, M. J. (1996). *Azar y probabilidad*. Madrid: Síntesis.

Godino, J. D. Batanero, C. y Font, V. (2007). The onto-semiotic approach to research in mathematics education. *ZDM. The International Journal on Mathematics Education,* 39 (1-2), 127-135.

Godino, J. D., Batanero, C., Roa, R. y Wilhelmi, M. R. (2008). Assessing and developing pedagogical content and statistical knowledge of primary school teachers through project work. En C. Batanero, G. Burrill, C. Reading & A. Rossman (2008). *Proceedings of the Joint ICMI/IASE Study: Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education. Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*. Monterrey: ICMI and IASE. Online: www.stat.auckland.ac.nz/~iase/publications.

Godino, J. D., Contreras A. y Font, V. (2006). Análisis de procesos de instrucción basado en el enfoque ontológico-semiótico de la cognición matemática. *Recherches en Didactiques des Mathématiques*, *26* (1), 39-88.

Godino, J. D., Font, V., Contreras, A. y Wilhelmi, M. R. (2006). Una visión de la didáctica francesa desde el enfoque ontosemiótico de la cognición e instrucción matemática. *Revista Latinoamerica de Investigación en Matemática Educativa, 9*(1), 117-150.

Godino, J., Wilhelmi, M. y Bencomo, D. (2005). Suitability criteria of a mathematical instruction process. A teaching experience of the function notion. *Mediterranean Journal for Research in Mathematics Education*, *4*(2), 1-26.

Gómez, B. (2003). La investigación histórica en Didáctica de las matemáticas. En E. Castro, P. Flores, T. Ortega, L. Rico y A. Vallecillos (Eds.). *Investigación en Educación Matemática. VII Simposio de la Sociedad Española de Investigación en Educación Matemática (SEIEM)* (pp. 79-85). Universidad de Granada.

Gómez, E. (2008). *La construcción de la noción de variable*. Tesis de doctorado. Centro de investigación en ciencia aplicada y tecnología avanzada del IPN, México.

Graunt, J. (1968). Natural and political observations mentioned in a following index and made upon the bills of mortality. En J. Newman (Ed), *Sigma: El mundo de las matemáticas* (Vol. 3, pp 108-121). Barcelona: Grijalbo (Trabajo original publicado en 1662).

Green, D. R. (1982). Testing randomness. *Teaching Mathematics and its Application, 1*(3), 95-100.

Green, D. R. (1989). School pupils' understanding of randomness. En R. Morris (Ed.), *Studies in Mathematics Education*, (Vol. 7, pp. 27-39). París: Unesco.

Green, D. R. y Hunt, D.N. (1992). *Generation and recognition of random sequences*. Mathematical Education Report ME9102, Department of Mathematical Sciences. Loughborough, UK: University of Technology.

Gutiérrez, S. (1992). *Filosofía de la probabilidad*. Valencia, España: Tirant lo Blanc.

Hacking I. (1975). *El surgimiento de la probabilidad. Un estudio filosófico de las ideas tempranas acerca de la probabilidad, la inducción y la inferencia estadística.* Barcelona: Gedisa.

Hald A. (1986). Galileo´s statistical analysis of astronomical observations. *International Statistical Review*, *54*(2), 211-220.

Hald, A. (2003). *A history of probability and statistics and their applications before 1750*. New Jersey: John Wiley.

Halley, E.(1693). An estimate of the degrees of the mortality of mankind, drawn from curious tables of the births and funerals at the city of Breslaw; with an attempt to ascertain the price of annuities upon lives. *Philosophical Transactions of the Royal Society of London*, *17*(1), 596-610.

Hawking, S. (2006). *Dios creó los números. Los descubrimientos matemáticos que cambiaron la historia*. Barcelona: Crítica.

Hawkins, A., Jolliffe, F. y Glickman, L. (1992). *Teaching statistical concepts.* Essex: Longman.

Hawkins, A. y Kapadia, R. (1984). Children's conceptions of probability: A psychological and pedagogical review. *Educational Studies in Mathematics, 15*(4), 349-377.

Heitele, D. (1975). An epistemological view on fundamental stochastic ideas. *Educational Studies in Mathematics*, *6*(3), 187-205.

Henry, M. (1997). Notion de modéle et modélization en l'enseignement. En *Enseigner les probabilités au lycée,* Commission Inter-IREM, Reims, 77-84.

Hernández, J., Noda, A, Palarea, M. y Socas, M. (2003). *Estudio sobre habilidades básicas en matemáticas de alumnos de magisterio*. Universidad de La Laguna. Departamento de Análisis Matemático.

Huberman, A. M. y Miles, M. (1994). Data management and analysis methods. En N. K. Denzin y Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 428 – 444). London: Sage Publications.

Huck, S., Cross, T. y Clark, S. (1986). Overcoming misconceptions about z-scores. *Teaching Statistics*, *8*(2), 38-40.

Huygens, C. (1714). *Libellus de ratiociniis in ludo aleae or the value of all chances in games of fortune.* London: T. Woodward. (Trabajo original publicado en 1657). Online: www.york.ac.uk/depts/maths/histstat/huygens.pdf

Inhelder, B. y Piaget, J. (1955). *De la logique de l´enfant à la logique de l´adolescent*. París: Presses Universitaires de France.

Johnson, R. y Kuby, P. (2008). *Estadística elemental: Lo esencial* (10ª Ed). México: Cengage Learning.

Jones, G. (2005). (Ed), *Exploring probability in school: Challenges for teaching and learning.* New York: Springer.

Jones, G., Langrall, C. y Mooney, E. (2007). Research in probability: responding to classroom realities. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning.* Greenwich, CT: Information Age Publishing y NCTM.

Kahneman, D., Slovic, P. y Tversky, A. (1982). *Judgement under uncertainty: heuristics and biases.* Cambridge: Cambridge University Press.

Kazak, S. y Confrey, J. (2007). Elementary school students' intuitive conceptions of random distribution. *International Electronic Journal of Mathematics Education, 2*(3). Online: www.iejme.com.

Kolmogorov A. (1956). *Foundations of the theory of probability*, 2° edition. New York: Chelsea Publishing Company (trabajo original publicado en 1933).

Konold, C. (1989). Informal conceptions of probability. *Cognition and instruction*, 6(1), 59-98.

Konold, C., Lohmeier, J., Pollatsek, A., Well, A. D., Falk, R. y Lipson, A. (1991). Novice views on randomness. En R. G. Underhill (Ed.), *Proceedings of the Thirteenth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 167-173). Blacksburg: Virginia Polytechnic Institute and State University.

Konold, C., Pollatsek, A., Well, A. y Gagnon, A. (1997). Students analyzing data: Research of critical barriers. En J. B. Garfield & G. Burrill (Eds.), *Research on the role of technology in teaching and learning statistics.* Voorburg, The Netherlands: International Statistical Institute.

Krickeberg, K. (1973). *Teoría de la probabilidad*. Barcelona: Teide.

Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology*, *32*(2), 311-328.

Laplace P. S. (1778). Mémoire sur les probabilités. *Mémoires de l'Académie Royale des Sciences de Paris*, 383-485.

Laplace P. S. (1812). *Théorie analytique des probabilités*. Gauthier-Villars.

Laplace P. S. (2006). Ensayo filosófico sobre las probabilidades. En S. Hawking (Ed), *Dios creó los números. Los descubrimientos matemáticos que cambiaron la historia* (pp. 361-419). Barcelona: Crítica (Trabajo original publicado en 1814).

Lecoutre, M. P. (1992). Cognitive models and problem spaces in "purely random" situations. *Educational Studies in Mathematics*, *23,* 557-568.

Lehmann, E. L. (1994). Jerzy Neyman. En National Research Council (Eds). *Biographical Memoirs* (Vol 63, pp. 394-421). Washington, DC: The National Academies Press.

Lévy P. (1935). Sur la sommabilité des séries aléatoires divergentes. *Bulletin de la Société Mathématique Française, 63*, 1-35.

Lévy, P. (1936). Sur quelques points de la théorie des probabilités dénombrables. *Institute Henry Poincaré, 6*(2), 153-184.

Lévy P. (1939). L'addition des variables aléatoires définies sur une circonférence. *Bulletin de la Société Mathématique Française, 67*, 1-41.

Lévy P. (1959). Esquisse d'une théorie de la multiplication des variables aléatoires. *Annales scientifiques de l'École Normale Supérieure, 3[a] série, 76*(1), 59-82.

Liapounoff, A. M. (1900). Sur une proposition de la théorie des probabilités. *Bulletin de l'Académie Impériale des Sciences de St. Petersburg*, 13(4), 359-386.

Liapounoff, A. M. (1901a). Sur théorème du calcul des probabilités. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences de Paris* 132(3), 126-128.

Liapounoff, A. M. (1901b). Une proposition générale du calcul des probabilités. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences de Paris,* 132(13), 814-815.

Maistrov, L. E. (1974). *Probability theory. A historical sketch.* New York : Academic Press.

Makar, K., Bakker, A. y Ben-Zvi, D. (2011). The reasoning behind informal statistical inference. *Mathematical thinking and learning*, *13*, 152-173.

Mayén (2009). *Significados de las medidas de posición central para estudiantes mexicanos de Bachillerato.* Tesis doctoral. Universidad de Granada.

Magnello, M.E. (2009). Karl Pearson and the establishment of mathematical estatistics. *International Statistical Review*, 77(1), 3-29.

McLean, M. (2007). *The cosmographia of Sebastian Münster: describing the word in the reformation.* Surreu, UK: Ashgate.

Ministerio de Educación y Ciencia/MEC (2006). *Real Decreto 1513/2006, de 7 de diciembre, por el que se establecen las enseñanzas mínimas correspondientes a la Educación primaria.*

Meyer, P. (1989). *Probabilidad y aplicaciones estadísticas*. México: SITESA: Addison Wesley Iberoamericana.

Miller, T. K. (1998). The random variable concept in introductory statistics. En L. Pereira- Mendoza (Ed.). *Proceedings of the Fifth International Conference on Teaching Statistics*. (pp. 1221-1222). Singapur: IASE.

Mises, R. von (1952). *Probabilidad, estadística y verdad*. Madrid: Espasa Calpe (Trabajo original publicado en 1928).

Moivre, A. de (1756). *The doctrine of chances or a method of calculating the probability of events in play,* 3rd edition. London: A. Millar.

Moivre, A. de (1984). On the measurement of chance, or, on the probability of events in games depending upon fortuitous chance. *International Statistical Review, 52*, 237-262. (Trabajo original publicado en 1712).

Mood, A. y Graybill, F. (1963). *Introducción a la teoría de la estadística*. Madrid: Aguilar.

Muñoz, J. y Valverde, J. (2000). *Compendio de epistemología*. Madrid: Trotta.

Moreno, M.D., Hernández, V., y Socas, M.M. (2010). Interpretación y análisis de los resultados obtenidos antes y después de un programa de formación sobre los números decimales. *Formación del Profesorado e Investigación en Educación Matemática*, 10,179-222.

Nardecchia, G. y Hevia, H. (2003). Dificultades en la enseñanza del concepto de variable aleatoria. Trabajo presentado en el *V Simposio de Educación Matemática*. Chivilcoy, Argentina.

National Council on Teaching Mathematics [NCTM]. (2000). *Principles and standards for school mathematics.* Reston, VA: NCTM. Online: standards.nctm.org/.

Newton, J. H. (2002). A conversation with Emanuel Parzen. *Statistical Science*, *17*(3), 357-378.

Nickerson, R. S. (2002). The production and perception of randomness. *Psychological Review*, *109*, 330-357.

Olivo, E. (2008). *Significados del intervalo de confianza en la enseñanza de la ingeniería en México.* Tesis Doctoral. Universidad de Granada.

Ortiz J. J. (2002). *La probabilidad en los libros de texto.* Granada: Grupo de investigación en educación estadística.

Oseguera, F. (1994). *El concepto de variable aleatoria en el contexto del currículo. Análisis y Alternativas*. Tesis de Maestría. Cinvestav, México.

Pascal, B. (1963). Traité du triangle arithmétique. En *Œuvres Complètes* (pp. 43-49). París: Seuil. (Trabajo original publicado en 1654).

Pascal, B. y Fermat, P. (2007). Correspondencia entre Pascal y Fermat durante 1654. En J. Basurto y J. A. Camuñez (Eds.), *La geometría del azar* (pp. 231-278). Madrid: Nivola. (Trabajo original publicado en 1654).

Parzen, E. (1971). *Teoría moderna de probabilidades y sus aplicaciones*. México: Limusa.

Pearson, E. S. y Kendall, M. (1970) (Eds). *Studies in the history of statistics and probability*. Londres: Charles Griffin.

Pearson, K. (1893). Contributions to the mathematical theory of evolution. *Journal of the Royal Statistical Society*, 56(4), 675-679.

Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society. Series A*, 185, 71-110.

Pearson, K. (1902). On the systematic fitting of curves to observations and measurements. *Biometrika*, 1(3), 265-303.

Pearson, K. (1924). Historical note on the origin of the normal curve of errors. *Biometrika, 16*(3), 402-404.

Pérez Echeverría, M. P. (1990). *Psicología del razonamiento probabilístico*. Madrid: ICE de la Universidad Autónoma.

Petrov, V. y Mordecki, E. (2002). *Teoría de probabilidades*. Moscú: Editorial URSS.

Piaget, J. (1991). *Introducción a la epistemología genética,* (Vol I) (Trad. M. T. Cevasco y Fischman). México: Paidós. (Trabajo original publicado en 1950).

Piaget, J. e Inhelder, B. (1951). *La genèse de l'idée de hasard chez l'enfant.* Paris: Presses Universités de France.

Poincare, H. (1979). El azar. En J. Newman (Ed), *Sigma: El mundo de las matemáticas* (Vol. 3, pp. 68-82). Barcelona: Grijalbo (Trabajo original publicado en 1908).

Poisson, S. D. (1832). Suite du mémoire sur la probabilité des résultats moyens des observations, inséré dans la connaissance des temps pour l'an 1827, *Connaissance des temps pour l'an 1832*, 3-22. (Leído a la academia el 20 de abril de 1829).

Poisson, S. D. (1837). *Recherches sur la probabilité des jugements en matière criminelle et en matière civile*. Paris: Bachelier. Online: gallica.bnf.fr/ark:/12148/bpt6k110193z.pagination.

Polya, G. (1965). *Cómo plantear y resolver problemas*. México: Trillas.

Prodromou, T. y Pratt, D. (2006). The role of causality in the co-ordination of two perspectives on distribution within a virtual simulation. *Statistics Education Research Journal, 5*(2), 69–88. Online: http://iase-web.org/Publications.php?p=SERJ_issues.

Reading, C. y Canada, D. (2011). Teachers' knowledge of distribution. En C. Batanero, G. Burrill y C. Reading (Eds.) (2011). *Teaching statistics in school mathematics-challenges for teaching and teacher education. A joint ICMI/IASE Study* (pp. 223-234). New York: Springer.

Reading, C. y Shaughnessy, J. M. (2004). Reasoning about variation. En J. Garfield & D. Ben-Zvi (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 201-226). Dordrecht: Kluwer.

Ríos, S. (1967). *Métodos estadísticos*. Madrid: Del Castillo.

Ríos, S. (1994). La revolución probabilística. *Historia de la Matemática*, 135-147.

Rossman, A. (2008). Reasoning about informal statistical inference: One statistician's view**.** *Statistics Education Research Journal*, 7(2), 40-58.

Rouanet, H. (1998). Significance testing in a Bayesian framework: Assessing direction of effects. *Behavioral and Brain Sciences,* 21, 217-218.

Rubin, A., Hammerman, J. y Konold, C. (2006). Exploring informal inference with interactive visualization software. En A. Rossman y B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics.* Salvador (Bahía): International Association for Statistical Education. CD ROM.

Ruiz, B. (2006). *Un acercamiento cognitivo y epistemológico a la didáctica del concepto de variable aleatoria*. Centro de investigación en ciencia aplicada y tecnología avanzada del IPN, México.

Ruiz, B., Albert, A. y Batanero, C. (2003). Hacia una didáctica de la variable aleatoria. Actas del 27 Congreso Nacional de Estadística e Investigación Operativa (pp. 1013-1030). Lérida: SEIO. CD ROM.

Ruiz, B., Albert, A. y Batanero, C. (2006). An exploratory study of students' difficulties with random variables. En A. Rossman y B. Chance (Eds.), *Proceedings of the Seventh International Conference on Teaching Statistics.* Salvador (Bahía): International Association for Statistical Education. CD ROM.

Ruiz, B., Arteaga, P. y Batanero, C. (2009a). Uso de promedios y dispersión en la comparación de distribuciones por futuros profesores. *Actas del VI Congreso Iberoamericano de Educación Matemática* (pp. 1351-1356). Puerto Mont, Chile: Universidad de Los Lagos.

Ruiz, B., Arteaga, P. y Batanero, C. (2009b). La comparación de distribuciones, ¿una actividad sencilla para los futuros profesores? *Actas del II Encontro da Probabilidade e Estatística na Escola.* CD-ROM. Universidad de Minho, Braga, Portugal.

Ruiz, B., Batanero, C. y Arteaga, P. (2011). Vinculación de la variable aleatoria y estadística en la realización de inferencias informales por parte de futuros profesores. *Bolema 24* (39), 413-429.

Ruiz, B. y Sánchez, T. (2007). La variable aleatoria como una idea fundamental dentro del currículum escolar. *Actas de las XIII Jornadas sobre Enseñanza y Aprendizaje de las Matemáticas*. Granada, España: Federación Española de Profesores de Matemáticas.

Sánchez, T., Albert, J. A. y Ruiz, B. (2011). Elementos cognitivos del estadístico como variable aleatoria en las distribuciones muestrales: el caso de la media. En: J.D. Zacarías. *Memorias del Primer Encuentro Internacional en la Enseñanza de la probabilidad y la estadística*. CD-ROM. Benemérita Universidad Autónoma de Puebla, México.

Scheaffer. R. L. (2006). Statistics and mathematics: On making a happy marriage. En G. Burrill (Ed.), *NCTM 2006 Yearbook: Thinking and reasoning with data and chance* (pp. 309-321). Reston, VA: NCTM.

Secretaria de Educación Pública [SEP]. (2006). *Planes y programas de estudio para la reforma de la educación secundaria.* México: Subsecretaría de Educación Básica de la Secretaría de Educación Pública. Online: www.sep.gob.mx/

Seneta, E. (1966). Markov and the birth of chain dependence theory. *International Statistical Review* 64 (3), 255-263.

Serradó, A., Cardeñoso, J. M. y Azcárate, P. (2005). Los obstáculos en el aprendizaje del conocimiento probabilístico: Su incidencia desde los libros de texto. *Statistics Education Research Journal 4*(2), 59-81.

Serrano, L. (1996). *Significados institucionales y personales de objetos matemáticos ligados a la aproximación frecuencial a la enseñanza de la probabilidad.* Tesis Doctoral. Universidad de Granada, España.

Serrano, L., Batanero, C. y Cañizares, M. J. (1999). Concepciones sobre distribuciones aleatorias planas en alumnos de secundaria. *Epsilon, 43-44*, 149-162.

Serrano, L., Batanero, C., Ortiz, J. J. y Cañizares, M. J. (2001). Concepciones de los alumnos de secundaria sobre modelos probabilísticos en las secuencias de resultados aleatorios. *Suma,* 36, 23-32.

Sfard, A. (1991). On the dual nature of mathematical conceptions: Reflections on processes and objects as different sides of the same coin. *Educational Studies in Mathematics*, 22 (1), 1-36.

Shaughnessy, J. M. (1992). Research in probability and statistic: Reflections and directions. En D. A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 465-494). New York: Macmillan Publishing Company.

Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. En F. Lester (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning* (pp. 957-1010). Greenwich, CT: Information Age Publishing, Inc, y NCTM.

Shaughnessy, J. M., y Ciancetta, M. (2002). Students' understanding of variability in a probability environment. En B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching Statistics.* Cape Town, South Africa: International Statistics Institute [CD-Rom].

Sierpinska, A. (1992). On understanding the notion of function. En E. Dubinsky y G. Harel (Eds.) *The concept of function: Aspects of epistemology and pedagogy*. (Notes, Vol. 25, pp 25-58). Washington: Mathematical Association of America.

Student (1908). The probable error of a mean. *Biometrika*, 6, 1–25.

Sylla, E. D. (2006). *Ars conjectandi together with letter to a friend on sets in court tennis*. Baltimore: The Johns Hopkins University Press.

Tauber, L. (2001). *La construcción del significado de la distribución normal a partir de actividades de análisis de datos.* Tesis Doctoral. Universidad de Sevilla, España.

Tauber, L., Batanero, C. y Sánchez, V. (2005). Diseño, implementación y análisis de enseñanza de la distribución normal en un curso universitario. *EMA, 9* (2), 82-105.

Tchebychef, P. L. (1899a). Démonstration élémentaire d'une proposition générale de la théorie des probabilités. En A. Markoff y N. Sonin (Eds.), *Œuvres de P. L. Tchebychef,* (Vol. 1, pp. 15-26). San Petersburgo: Académie Impériale des sciences. (Trabajo original publicado en 1846).

Tchebychef, P. L. (1899b). Des valeurs moyennes. En A. Markoff y N. Sonin (Eds.), *Œuvres de P. L. Tchebychef,* (Vol. 1, pp. 687-694). San Petersburgo: Académie Impériale des sciences. (Trabajo original publicado en 1867).

Tchebychef, P. L. (1899c). Sur deux théorèmes relatifs aux probabilités. En A. Markoff y N. Sonin (Eds.), *Œuvres de P. L. Tchebychef,* (Vol. 2, pp. 479-491). San Petersburgo: Académie Impériale des sciences. (Trabajo original publicado en 1887).

Toohey, P. G. (1995). *Adolescent perceptions of the concept of randomness.* Tesis de Maestría. University of Adelaide.

Truran, J. (1994). Children's intuitive understanding of variance. En J. Garfield (Ed.), *Research papers from the Fourth International Conference on Teaching Statistics.* Minneapolis, MN: The International Study Group for Research on Learning Probability and Statistics.

Vallecillos, A. (1996). *Inferencia estadística y enseñanza: Un análisis didáctico del contraste de hipótesis estadísticas*. Madrid: Comares.

Vallecillos, A. (1999). Some empirical evidence on learning difficulties about testing hypotheses. *Bulletin of the International Statistical Institute: Proceedings of the Fifty-Second Session of the International Statistical Institute* (Tome LVIII, Book 2, pp. 201-204). Helsinki: International Statistical Institute.

Vergnaud, G. (1990). La théorie des champs conceptuels. *Recherches en Didactique de Mathématiques*, *10*(2-3), 133-170.

Wackerly, D., Mendenhall, W. y Scheaffer L. (2002). *Estadística matemática con aplicaciones*. México: Thompson.

Walpole, R., Myers, R. y Myers, Sh. (1999). *Probabilidad y estadística para ingenieros* (6ª Ed.) (trad. R. Cruz). México: Prentice-Hall Hispanoamericana, S. A. (Trabajo original publicado en 1998).

Watson, J. M. (2001). Longitudinal development of inferential reasoning by school students. *Educational Studies in Mathematics, 47*, 337-372.

Watson, J. M., & Moritz, J. B. (1999). The beginning of statistical inference: Comparing two data sets. *Educational Studies in Mathematics, 37,*145-168.

Weber, R. P. (1985). *Basic content analysis.* Londres: Sage.

Wild, C. y Pfannkuch, M. (1999). Statistical thinking in empirical enquiry (con discussion). *International Statistical Review*, *67* (3), 223-265.

Yuren, M.T. (2006). *Leyes, teorías y modelos* (3ª ed). México: Trillas

Zabell, S. L. (1992). Randomness and statistical applications. En F. Gordon and S. Gordon (Eds,), *Statistics for the XX1 Century*. The Mathematical Association of America.

Zieffler, A., Garfield, J., Delmas, R. y Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, 7(2), 40-58.

Anexos

Anexo 1.

Protocolo.
Estudio 3: La entrevista clínica.

PARTE I

A raíz de los festejos del día del niño, el departamento de relaciones públicas de una fábrica desea conocer el número de hijos que tiene cada uno de los 200 obreros que ahí laboran. La intención es efectuar una rifa que beneficie a los hijos de los trabajadores. Los resultados de la encuesta realizada para tal efecto son:

Número de hijos	Número de trabajadores
0	16
1	22
2	33
3	45
4	31
5	20
6	12
7	9
8	7
9	5

a) Si tomamos un trabajador al azar, ¿cuál es la probabilidad de que tenga 4 hijos? ¿Y 6 hijos? ¿Y 7? ¿Cuál es la probabilidad de que no tenga ningún hijo?
b) ¿Qué depende de qué: el número de hijos depende de la probabilidad o al revés? ¿Cuál de estas dos variables sería la variable dependiente y cuál la variable independiente?

PARTE II

En la fiesta se premiará a la familia de un obrero con boletos para el teatro, pero los boletos de teatro se tienen que reservar con días de anticipación, así que le encomiendan a la trabajadora social de la empresa que decida cuántos boletos tiene que comprar.

En la siguiente tabla, calcula la probabilidad de que el trabajador premiado tenga un hijo, dos hijos, tres hijos, … y nueve hijos.

Número de hijos	0	1	2	3	4	5	6	7	8	9
Número de trabajadores	16	22	33	45	31	20	12	9	7	5
Probabilidad										

Indica las operaciones que efectuaste para llenar la tabla y contesta:

c) ¿Qué recomendación darías a la trabajadora social de cuántos boletos comprar para que la empresa pierda lo menos posible? Explica detalladamente porqué le darías esa recomendación.
d) De acuerdo a tu recomendación ¿con qué probabilidad el trabajador premiado ocuparía exactamente el número de boletos que compró la trabajadora social?
e) De acuerdo a tu recomendación ¿qué tanto puedes asegurar que se ocuparan exactamente los boletos que se compraron?
f) De acuerdo a tu recomendación, ¿con qué probabilidad al trabajador premiado le alcanzarán los boletos y hasta le sobrarán?
g) De acuerdo a tu recomendación, ¿con qué probabilidad el trabajador premiado no le alcanzarán los boletos para poder llevar a todos sus hijos al teatro?

PARTE III

Analicemos matemáticamente la situación…

Utiliza la información que obtuviste en la Parte I acerca de la variable dependiente e independiente y la tabla que hiciste en la segunda parte para hacer una gráfica que represente la probabilidad de que un trabajador con un cierto número de hijos salga premiado.

h) Representa con una letra la variable dependiente y con otra la variable independiente y utiliza estas letras para representar en notación funcional el resultado obtenido en el inciso b) de la parte II.
i) Describe, en el contexto del problema, con tus propias palabras la información que proporciona la variable dependiente de la variable independiente.
j) Describe el dominio de la variable independiente y el rango de la variable dependiente.
k) Describe el método a partir del cuál puedes obtener el valor de la variable dependiente, a partir del valor de la variable independiente.
l) La relación que estableciste ¿es una función matemática? ¿Por qué?
m) ¿Qué diferencias encuentras entre la variable independiente de esta función de probabilidad y las variables independientes de las otras matemáticas que tú conoces, por ejemplo, de álgebra o cálculo?
n) ¿Qué nombre le darías a este tipo de variable?

Anexo 2.

El proyecto.
Estudio 4: Evaluación de un proyecto.

PRÁCTICA 4:
EVALUACIÓN DE EXPERIENCIAS DE ENSEÑANZA: ESTADÍSTICA EN PRIMARIA

Objetivos:
1. Adquirir competencias para consultar bibliografía sobre didáctica de contenidos específicos de matemáticas.
2. Adquirir competencias para analizar la "idoneidad didáctica" de experiencias de enseñanza y aprendizaje de las matemáticas, programadas o implementadas.

Contenidos:

Teóricos:
- Didáctica de la estadística en la educación primaria.
- Idoneidad de una propuesta didáctica

Matemáticos:
- Tratamiento de la información, aleatoriedad, experimentos y sucesos aleatorios, frecuencias, tablas, gráficos, medidas de posición central y dispersión.

Reactivos y consignas:
PARTE A:
1. Realizar el experimento descrito en esta práctica (Anexo 1) y analizar los datos obtenidos para evaluar las propias concepciones sobre la aleatoriedad.
2. Analizar los contenidos sobre tratamiento de la información, estadística y probabilidad que se proponen para la educación primaria en los Decretos de enseñanzas mínimas (MEC y Junta)
3. Estudiar la "Didáctica" de la Estadística en la bibliografía recomendada.

PARTE B:
4. Estudiar la "Pauta de análisis de la idoneidad didáctica de procesos de enseñanza y aprendizaje de la matemática" entregada en clase.
5. Analizar la programación de una didáctica sobre tratamiento de la información propuesta en el aneco para valorar su idoneidad didáctica, conforme a los siguientes apartados:
   a) Idoneidad epistémica- matemática (el contenido estadístico estudiado es adecuado)
   b) Idoneidad cognitiva (adecuación a los conocimientos previos y edad. Los objetivos de aprendizaje se pueden alcanzar)
   c) Idoneidad de medios técnicos y del tiempo empleado
   d) Idoneidad motivacional y afectiva
   e) Idoneidad interaccional (la interacción entre profesores y estudiantes favorece la superación de dificultades y conflictos)
   f) Idoneidad ecológica (grado en que el proceso de estudio se adapta a las orientaciones curriculares y a las condiciones del entorno)
6. Indicar cambios en la experiencia para mejorar su idoneidad didáctica global.

Bibliografía recomendada *(*Disponible desde el directorio de la asignatura).

Godino, J. D., (Director) (2003). *Didáctica de las matemáticas para maestros*. Tema 6: Estocástica (Capítulo 1. Estadística, páginas 411-423). Disponible en Fotocopias y Directorio de la asignatura.

Decretos de enseñanzas mínimas en Educación Primaria. Área de Matemáticas. MEC, 2006 y Junta de Andalucia.

FICHA PARA ACTIVIDADES PRÁCTICAS DE CURRÍCULO DE PRIMARIA

Nombre del estudiante
Práctica Nº 4: Evaluación de experiencias de enseñanza (PARTE A: Desarrollo del proyecto)
RESOLUCIÓN: (Utilizar esta hoja por delante y por detrás, añadiendo otras si fuera necesario) 1. Analizar los datos obtenidos en la clase durante el experimento realizado para evaluar las concepciones sobre la aleatoriedad. 2. Preparar un informe sobre el análisis, incluyendo todas las tablas y gráficos estadísticos elaborados 3. Resumir las conclusiones obtenidas del estudio anterior acerca de las intuiciones de la clase sobre experimentos aleatorios. 4. Elaborar una tabla que resuma en una columna los contenidos sobre tratamiento de datos y estadística que has trabajado durante el proyecto y en otra los que se describen en el Decreto de Enseñanzas Mínimas del MEC (tanto en contenidos, como en criterios de evaluación) 5. Estudiar las páginas 411-423 del Tema 6: Estocástica (Capítulo 1. Estadística) del texto recomendado.

Práctica Nº 4: Evaluación de experiencias de enseñanza (PARTE B)
REFLEXIÓN: (Utilizar esta hoja por delante y por detrás, añadiendo otras si fuera necesario) 1. Estudiar la pauta de evaluación de idoneidad didáctica de un proceso de estudio 2. Aplica la pauta para analizar el experimento de enseñanza sobre estadística para valorar su idoneidad didáctica, conforme a los apartados: a) Idoneidad epistémica o matemática (contenido estadístico estudiado) b) Idoneidad cognitiva (adecuación a los conocimientos previos y edad. Los objetivos de aprendizaje se pueden alcanzar) c) Idoneidad de medios técnicos y del tiempo empleado d) Idoneidad motivacional y afectiva e) Idoneidad interaccional (la interacción entre profesores y estudiantes favorece la superación de dificultades y conflictos) f) Idoneidad ecológica (grado en que el proceso de estudio se adapta a las orientaciones curriculares y a las condiciones del entorno) 3. Indicar cambios en la experiencia para mejorar su idoneidad didáctica global.

Programación de una unidad didáctica sobre el contenido "Tratamiento de la información" (Estadística)

Nivel: Profesores de Educación Primaria

Desarrollo del tema a partir de un proyecto de análisis de datos: "Comprueba tus intuiciones sobre el azar"

1. CONTEXTO CURRICULAR

En las propuestas curriculares para el área de matemáticas en educación primaria, tanto a nivel regional, nacional e internacional se incluye un bloque temático sobre "Tratamiento de la información, azar y probabilidad". Concretamente en el borrador de Decreto del MEC (Octubre 2006) se incluyen contenidos de este bloque temático para los tres ciclos de educación primaria.

La Estadística ha cobrado gran desarrollo en los últimos años, contribuyendo al avance de la ciencia y la técnica y al crecimiento de la economía, por lo que la mayor parte de los países han introducido su enseñanza desde la educación primaria. La estadística es hoy una parte de la educación general deseable para los ciudadanos, quienes precisan adquirir la capacidad de lectura e interpretación de tablas y gráficos estadísticos que con frecuencia aparecen en los medios de comunicación. Las principales razones que fundamentan la enseñanza de la estadística son las siguientes:

- Es útil para la vida posterior a la escuela, ya que en muchas profesiones se precisan unos conocimientos básicos del tema.
- Su estudio ayuda al desarrollo personal, fomentando un razonamiento crítico, basado en la valoración de la evidencia objetiva, apoyada en los datos, frente a criterios subjetivos.
- Ayuda a comprender los restantes temas del currículo, tanto de la educación obligatoria como posterior, donde con frecuencia aparecen gráficos, resúmenes o conceptos estadísticos.

Se trata de superar el carácter exclusivamente determinista del currículo de matemáticas, y mostrar al alumno una imagen más equilibrada de la realidad, en la que hay una fuerte presencia de fenómenos aleatorios.

2. OBJETIVOS

- Hacer reflexionar al alumno y evaluar sus propias intuiciones sobre el comportamiento de un experimento aleatorio sencillo (lanzamiento de una moneda).
- Realizar un experimento aleatorio, recoger los datos obtenidos y aplicar las técnicas y conceptos estadísticos básicos para evaluar las intuiciones probabilísticas.
- Apreciación de la utilidad de los resúmenes gráficos y del cálculo de estadísticos para obtener conclusiones sobre experimentos aleatorios y para analizar datos obtenidos en una investigación.

3. CONTENIDOS CURRICULARES

- Recogida y registro de datos de experimentos aleatorios sencillos, mediante la realización y observación de sus resultados.
- Elaboración de tablas de frecuencias de variables discretas.
- Elaboración de gráficos estadísticos (diagramas de barras, puntos y sectores, ...)
- Concepto de posición central y dispersión. Cálculo de la media, moda, mediana y rango.
- Comparación de dos distribuciones de datos.
- Evaluación de las propias intuiciones sobre las características de los experimentos aleatorios.

4. SECUENCIA DE ACTIVIDADES

4.1. Consigna inicial y discusión colectiva:

1. ¿Tienes una buena intuición sobre el azar? ¿Cómo piensas que deberían ser los resultados de lanzar una moneda 20 veces seguidas? ¿Serías capaz de escribir 20 resultados de lanzar una moneda (sin lanzarla realmente, sino como tú pienses que debieran salir) de forma que otras personas piensen que has lanzado la moneda en realidad. O, ¿podría otra persona adivinar que estás haciendo trampa?

4.2. Realización individual del experimento de lanzar una moneda y anotar los resultados en una pauta cuadriculada.

2. Vamos a comprobar qué tal son tus intuiciones respecto a los resultados aleatorios. Abajo tienes dos cuadrículas. En la primera de ellas escribe 20 resultados sin realizar realmente el experimento. En la segunda mitad lanza la moneda 20 veces y escribe los resultados obtenidos. Pon C para cara y + para cruz.

Lanzamiento simulado:																			
Lanzamiento real:																			

4.3. Gestión de la clase, nuevas cuestiones y actividades

Cuando todos los alumnos han finalizado la realización del experimento se plantearán preguntas similares a las que reproducimos a continuación.

3. ¿Cómo podremos distinguir una secuencia realmente aleatoria de otra que hemos inventado?

Se dejará algún tiempo para pensar y a continuación se organiza una discusión colectiva. Seguramente algún alumno sugerirá contar el número de caras y cruces que debe ser aproximadamente igual, ya que hay las mismas posibilidades para la cara que para la cruz.

4. *Pero, ¿hemos de obtener exactamente 10 caras y 10 cruces? ¿Qué pasa si obtenemos 11 caras y 9 cruces? ¿Y si obtenemos 18 y 2? ¿Qué os parece si comparamos el número de caras en las secuencias real y simulada de todos los alumnos de la clase?*

Para realizar esta comparación se recogen los datos de todos los alumnos de la clase, tanto del número de caras en las secuencias simuladas como en las reales, para proceder, primeramente al análisis de cada una de estas dos variables y luego a la comparación de las principales diferencias en su distribución.

5. Hemos recogido el número de caras en las secuencias simuladas por cada alumno de la clase ¿Cómo podríamos organizar y resumir estos datos? ¿Cuáles son el valor mínimo y máximo obtenido? ¿Cómo representar los datos de modo que sepamos cuántas veces aparece cada valor? ¿Cuál es el valor más frecuente?

El profesor ayudará a los chicos a identificar el valor máximo y mínimo y a organizar un recuento y tabla de frecuencias, haciéndoles ver su utilidad para resumir la información.

De igual modo se realizará el estudio del número de caras en las secuencias reales, para finalmente comparar las dos distribuciones y analizar si existen algunas diferencias importantes que indiquen que nuestra intuición respecto a la aleatoriedad nos engaña.

6. Compara ahora los gráficos del número de caras en las secuencias reales y simuladas. ¿En qué se parecen? ¿En qué se diferencian? ¿Es el valor más frecuente el mismo? ¿Hay el mismo rango de variación de valores? ¿Cuál de las dos variables tiene mayor variabilidad? ¿Piensas que nuestras intuiciones sobre el número de caras que se obtienen al lanzar 20 veces una moneda equilibrada es totalmente correcta? ¿Podrías idear algún gráfico en que se viesen más claramente las diferencias?

El número de caras es sólo una de las variables que podemos analizar en una secuencia de resultados aleatorios, en la que aparecen otros muchos modelos probabilísticos. Uno de ellos es la longitud de las rachas que, intuitivamente esperamos que sean cortas. Es bien conocida la *falacia del jugador* por la que esperamos que, tras una corta racha de, por ejemplo caras, la probabilidad de que aparezca una cruz aumente.

En este proyecto proponemos analizar dos nuevas variables en las secuencias producidas por los alumnos: el número de rachas y la longitud de la racha más larga.

Para aclarar el lenguaje llamaremos *racha* a una secuencia de resultados iguales, de modo que, si después de una cara aparece una cruz (o viceversa) la racha tiene longitud 1. Supongamos que la tabla siguiente contiene los resultados C, + simulados y obtenidos por un alumno. Se muestran las rachas que aparecen. Vemos que en la secuencia simulada, la racha más larga es de longitud 3 (3 caras) y que el número de rachas es 12, mientras que en secuencia real hay una racha de 5 cruces y el número de rachas es 11.

Lanzamiento simulado:																			
C	C	+	C	+	+	+	C	C	+	C	+	C	+	+	C	C	C	+	+
Lanzamiento real:																			
+	C	+	C	+	+	C	C	+	C	C	C	+	+	+	+	+	C	+	+

Para motivar el estudio de estas variables, el profesor preguntará si el resultado obtenido arriba, donde aparecen 5 cruces seguidas, parece razonable. Probablemente algún alumno sugiera que la moneda utilizada no está bien construida y se plantea el estudio de las rachas en las secuencias.

El profesor explicará cómo identificar las rachas y sugerirá a los niños que busquen cuál es la racha más larga en cada una de sus dos secuencias, así como que cuenten el número de rachas, procediendo de nuevo al estudio y comparación de estas variables en las dos secuencias, tal y como se ha hecho con el número de caras y finalizando con una discusión sobre sus diferencias y si nuestras intuiciones sobre las rachas son o no correctas.

El profesor puede usar una hoja de registro como la que reproducimos a continuación donde cada niño anota sus resultados. Luego la hoja se fotocopia y se reparte a los chicos. Si hay poco tiempo, la clase puede dividirse en grupos para que cada uno de ellos se encargue de analizar una de las variables y posteriormente, una vez disponibles los gráficos se realiza la discusión conjunta.

Hoja de recogida de datos de la clase sobre las rachas

	Secuencia simulada			Secuencia real		
Alum. Nº	N. caras	N. rachas	Racha mayor	N. caras	N. rachas	Racha mayor
1	10	14	4	11	9	4
2						
...						

Anexo 3.

Transcripción de la entrevista.
Estudio 3: Entrevista Clínica.

En la transcripción del diálogo de las personas que intervienen en la entrevista, se añaden en cursivas notas explicativas necesarias para la comprensión de la acción durante su lectura, así como tiempos de grabación y notas de los archivos de video y audio. Durante la entrevista participaron: Brenda (*alumna*), Mónica (*alumna*), Profesor (*de matemáticas de Brenda y Mónica de primer semestre*) y Blanca (*investigadora*).

PRIMERA PARTE

Brenda y Mónica transcribieron los datos (en forma de tabla) al pizarrón. La tabla contenía tres columnas correspondientes al número de hijo y al número de trabajadores y a la probabilidad. Contestan el primer inciso directamente en la tabla. La probabilidad la escriben como el cociente de dos números enteros en todos los casos. También escribieron en el pizarrón sus respuestas a cada inciso de la primera parte.

En el pizarrón:

Número de hijos	*Número de trabajadores*	*Probabilidad*
0	16	$\frac{16}{200}$
1	22	$\frac{22}{200}$
2	33	$\frac{33}{200}$
3	45	$\frac{45}{200}$
4	31	$\frac{31}{200}$
5	20	$\frac{20}{200}$
6	12	$\frac{12}{200}$
7	9	$\frac{9}{200}$
8	7	$\frac{7}{200}$
9	5	$\frac{5}{200}$

Primer pasaje: ***Definición de la probabilidad como variable dependiente o independiente.***

ByM01 (video) --- 0:22.8
Mónica tiene la actividad por escrito y se lo lee a Brenda. Brenda, en tanto, está terminando de escribir sus respuestas en el pizarrón.
Mónica: ¿Qué depende de qué: el número de hijos depende de la probabilidad o al revés? Y luego ¿cuál de estas dos variables será la variable independiente e independiente?
Brenda: Bueno... primero ¿qué depende de qué? pues el número de hijos... no...el número
Ambas: de trabajadores... *(muy seguras de su respuesta)*
Brenda: ...depende del número de hijos, ¿no?
Mónica: No, no es que dice: el número de hijos depende de la probabilidad, no del...
Brenda: no.. pero la primera no es...
Mónica: ¿qué depende de qué? ...pues...
Ambas: el número de hijos depende de la probabilidad ¿o al revés?...
Mónica: No, la probabilidad depende del número de hijos o sea que qué tan probable... o sea... depende... la probabi...Yo digo que la probabilidad depende del número de hijos. (*recurre a voltear a ver a los profesores para que le confirmen o le refuten esta última afirmación)*
Brenda: Pero más bien... el número...el...bueno... no, sí la probabilidad depende del número de hijos porque... *(dudando)*
Mónica: Es que la probabilidad...
Brenda: No, sí la probabilidad... *(muy segura)*
Mónica: Sí, porque según yo...
(se arrebatan la palabra una a la otra)
Brenda: depende del número de hijos porque...
Mónica: Ajá, porque qué tan probable es...
Brenda: porque qué tan probable es que los trabajadores tengan ese número de hijos.
Mónica: Ajá.
Brenda: Entonces sí *(escribe la conclusión en el pizarrón).*

En el pizarrón:

b) La probabilidad depende del número de hijos

-- 01:31.1

Segundo pasaje: ***Dan una primera aproximación a la noción de función.***

ByM01(video)-- 1:38.9
Nuevamente Mónica tiene la actividad por escrito y se la lee a Brenda.
Mónica: ¿Cuál de estas dos variables sería la variable dependiente y cual la independiente? Ésta. ¿Hablan de aquí *(se refiere al inciso anterior)* o de... ?
Brenda está terminando de escribir la conclusión en el pizarrón, así que no le hace mucho caso a Mónica.
Mónica: Es que ve estás dos no pueden ser de que ¡ah! dependiente o independiente *(señala las columnas de la tabla)...* así, no. Yo no puedo entrar... de lo que el número de hijos y el número de trabajadores ... no puede ser, pon tú dependiente e independiente o independiente y dependiente *(señala las columnas de la tabla)*

porque no, no hay relación... por ejemplo aquí *(señala la columna de número de trabajadores)* empieza un número 16, y luego aquí mayor, mayor, incrementa y luego decrementa y luego decrementa, así como que no hay rela...o sea no hay mucha relación, igual de cero hijos pudo haber un trabajador y de cuatro a cinco mil y así... como que no da resul...pero dice...
-- 2:22.9

Tercer pasaje: *Definen cuál es la variable independiente y cuál es la variable dependiente en el problema*.

ByM01(video) -- 2:32.0
Nuevamente Mónica tiene la actividad por escrito y se la lee a Brenda.
Mónica: El número de hijos depende de la probabilidad o al revés y ya la pusimos. Ahora dice ¿Cuál de estas dos variables sería la variable dependiente y cuál la independiente?
Brenda: Pues yo digo que el número de trabajadores depende del número de hijos.
Mónica: ¿Por qué? Porque es más probable que sea del tipo 3 ó 4
Brenda: Sí, porque, bueno se supone que... o sea aquí por decir, x... o sea'x'sería la... la variable independiente, o sea los número de hijos son independientes, porque según o sea... como... Sí has de cuenta... ya ves que sería como poner 'x' y 'y'...
Mónica: Ajá. A ver ponlo.
(Brenda escribe una'x'al final de la columna del número de hijos y una 'y' al final de la columna del número de trabajadores)

En el pizarrón:

Número de hijos	*Número de trabajadores*	*Probabilidad*
0	16	$\frac{16}{200}$
1	22	$\frac{22}{200}$
2	33	$\frac{33}{200}$
3	45	$\frac{45}{200}$
4	31	$\frac{31}{200}$
5	20	$\frac{20}{200}$
6	12	$\frac{12}{200}$
7	9	$\frac{9}{200}$
8	7	$\frac{7}{200}$
9	5	$\frac{5}{200}$
x	y	

Brenda: Entonces está sería la ‘x’ *(señala la columna de número de hijos)* porque realmente... o sea si te fijas...
Mónica: Bueno es que aquí sí crece y luego vuelve para abajo otra vez *(señala la columna de número de trabajadores)*.
Brenda: Sí, pero si te fijas esta es la que.. esta es la que... *(señala la columna de número de hijos)*
Mónica: Va determinando...
Brenda: Ajá, va determinando al número de trabajadores. El número de trabajadores no determina al número de hijos.
Mónica: Ah, ya entiendo.
Brenda: Sino al revés, el número de hijos es lo que está determinando el número de trabajadores. Entonces está sería...
Mónica: Pero eso no significa que depende de los hijos va a ver tantos trabajadores sino que esta *(señala la columna de número de hijos)* rige a esta *(señala la columna de número de trabadores)*.
Brenda: Entonces en este caso, el número de hijos es la variable independiente y el número de trabajadores es la variable dependiente.
(Brenda escribe la conclusión en el pizarrón)

En el pizarrón:

Número de hijos = V. independiente Número de trabajadores = V. dependiente

--4:23.8

Cuarto pasaje: *Cuestionamiento sobre los motivos para definir las variables independiente y dependiente*.

ByM01(video)-- 4: 42.9
Profesor: ¿Si tomamos un trabajador al azar cuál es la probabilidad de que tenga 4 hijos?
Ambas: 31 de 200.
Profesor: ¿Por qué no pusieron por ejemplo 4 entre doscientos u otro número porque 31?
Brenda: Porque bueno, en la tabla el número de hijos que son 4 tienen un número de trabajadores de 31 y el universo de trabajadores son 200. Entonces de esos 200 quiere decir que 31 tienen 4 hijos.
Mónica: Es como que al sumar todos estos *(señala la columna de número de trabajadores)* y te dan 200, pero pues ya te lo da de antemano el problema, te dice que son 200 obreros en total. Este es el número de trabajadores obreros *(señala la columna)* que si hacemos la suma te da 200. Entonces por ejemplo de 5, a pues a ver 5 *(busca en la tabla)*, aquí están los 5. Hay 20 trabajadores de 200 que tienen 5 hijos.
Profesor: La cuestión sería ¿Qué pasaría si sumáramos todas las probabilidades? La probabilidad de que tenga un hijo, dos hijos, así..., ¿qué pasaría?
Brenda: ¿Sumar el número de hijos?
Profesor: Las probabilidades. Miren por ejemplo, cuál es la probabilidad de que... haya 4 hijos.
Brenda: 31 sobre 200.
Profesor: Bueno entonces 31 sobre 200 es una probabilidad. ¿Qué pasaría si sumo cada caso?

Brenda: Deben de dar 200, o sea debe de dar todo el universo.
--- 6:27.5
ByM01(video) --- 6:52.1
Profesor: Bueno, entonces la pregunta es si la probabilidad puede valer 200.
Brenda: Sí, bueno, ¿qué la probabilidad dé algo o sea pueda valer 200?
Profesor: Sí, por ejemplo el 31 sobre 200 es un número negativo, positivo...
Mónica: Positivo. Fracción.
Profesor: Fracción, mayor que uno, menor que uno,...
Ambas: Menor que uno.
Profesor: Menor que uno.¿Y las probabilidades son mayores que uno?
Mónica: Son menores que uno. Todas son menores que uno.
Profesor: Entonces si sumamos todas esas probabilidades, ¿qué creen que pueda pasar?
Brenda: Sería igual a 200. O sea te quedaría 200 y sería igual a... o sería la unidad uno. Quedaría 200...
--- 7:27.5
ByM01 (video) -- 8:04.4
Profesor: Es decir, la probabilidad de tener cuatro hijos...
Brenda: O sea, esta es la probabilidad de 31,... *(le arrebata la palabra, señala la respuesta al inciso a), 31/200, escrita a un lado de la tabla).*
Profesor: 31 de 200 *(ninguna de las dos lo escucha, enfrascadas en su discusión)*.
Brenda: ...entonces estas son todas las probabilidades que hay *(señala la columna de número de trabajadores)*...
Mónica: A pues sí.
Brenda: ...y si las sumamos sí nos van a dar 200. Haz de cuenta que ésta *(señala el 31 en la columna de número de hijos)* es la probabilidad de tener 4 hijos. Está diciendo el profe que se refiere a estas probabilidades *(señala la columna de número de trabajadores)* y al sumarlas nos darían 200.
Mónica: Pues sí
Profesor: ¿16 sería una probabilidad?
Ambas: ¿16? *(extrañadas)*
Brenda: ¿así, sola?
Profesor: Sí, así sola.
Brenda: No, así solo, no.
Profesor: ¿Qué le faltaría?
Mónica: 16 de cuántos.
Profesor: ¿De cuántos sería?
Ambas: De 200.
Profesor: Entonces la cuestión es, ¿si sumando todas? ¿qué pasa si sumo probabilidades? ¿qué se esperaría que diera?
Brenda: El universo en general, los 200.
(Mónica asiente con la cabeza)
Profesor: Denme una respuesta entre las dos. Lleguen a un acuerdo.
Mónica: Pero es que.. probabilidad sí yo estoy de acuerdo que sí son estos *(señala la columna de número de trabajadores)* porque es todo.. es un porcentaje... es que tanto, que tan probable es que salgan... el número de hijos. Pues sí, aquí *(señala la columna de número de trabajadores)* de hecho serían 200. Pero lo que estábamos discutiendo de que... ya nos había preguntado de 200... y luego que tan probable de que si sumo ¿éstos *(señala la columna de número de hijos)* nos preguntó ahorita?
Profesor: Las probabilidades.

Ambas: No, las probabilidades son 200.
Mónica: Las probabilidades son 200. Ya.
Brenda: Al sumar las probabilidades te debe dar todo el conjunto que estás.. o sea, todo el universo.
(Las dos asienten, están de acuerdo y convencidas)
Profesor: Nada más la pregunta es si una probabilidad puede valer 200. Es que me dicen que la probabilidad suma 200.
Mónica: Pues ya no sería probabilidad sino un total *(riéndose).*
Brenda: Bueno, lo que podría ser es que el número... es decir, podría ser la probabilidad de que los trabajadores tuvieran de 0 a 9 hijos. Esa sería la única forma en que hubiera una probabilidad de 200.
Mónica: Ajá. Total ya, sin dividir. En conjunto.
Profesor: Sin embargo, en la primera respuesta de cuatro hijos dice 31 sobre 200. Y ustedes me dijeron que todas las probabilidades son fracciones menores que uno. Entonces en total, hay 10 fracciones porque son 10 casos y si todas son menores de uno, ¿cómo le hacen para que llegue a 200?
Desconcertadas Miran las tablas en el pizarrón.
Mónica: ¿Cómo?... o sea ¿cómo maestro?
Profesor: Sí, ustedes me acaban de decir hace ratito que la probabilidad de tener 4 hijos es 31 sobre 200 y que esa era una fracción menor que uno ¿no?
Ambas: Ajá.
Profesor: Y luego me dijeron que todas las probabilidades son fracciones, fracciones menores que uno. Entonces si son 10 probabilidades a lo más que podríamos aspirar es a que su suma fuera menor que 10. Cuando yo les pregunto cuál es la suma de probabilidades, ustedes me dicen que 200 así que ¿Cómo le hacen para que llegue a 200?
Hacen exclamaciones de obviedad
Mónica: ¡Ay maestro!
Brenda: Lo que pasa es que esta fracción es con respecto a 200.
Mónica: Ajá. El común denominador es 200.
-- 11:28.3
Profesor: ¿Tú que piensas Mónica?
Mónica: Yo digo que aun sumando todas estas *(señala la columna de número de hijos)* no nos da mayor que 1 porque para que fuera mayor que 1 la fracción tendría que ser 201 o más entre 200. O sea 201 sobre 200 y estos *(refiriéndose a la columna de número de hijos)* no van a dar más de 200.
Profesor: O sea ¿estás de acuerdo con ella?
(Ninguna de las dos contesta. Se miran confundidas)
Brenda: No, pero la probabilidad se refiere a ésta *(señala la columna de número de trabajadores).*
Brenda voltea a ver al profesor, busca su apoyo. Mónica está confundida. Brenda tiene mayor seguridad en lo que dice.
Profesor: Decidan ente ustedes.
Mónica: No, ¡ya me dejó... ¿cómo la probabilidad va a ser ésta?! *(extrañada, señala la columna de número de trabajadores).*
ByM01 (video)--- 13:04.8
Profesor: ¿Podría tener sentido decir que la probabilidad tiene un tope, un límite, un número al cuál no puede rebasar? ¿o puede tomar cualquier valor positivo? Ustedes me acaban de decir que no tiene sentido que sea negativo, pero por ejemplo, ¿puede ser cero?

Mónica: Sí.
Brenda: Pues sí.
Mónica: Por ejemplo, de 10 hijos es cero, la probabilidad.
Profesor: Entonces, ¿la probabilidad puede tomar cualquier valor? ¿Puede haber probabilidad 2, probabilidad 3... probabilidad 200, puede haber probabilidad 1000 por ejemplo? ¿Tiene sentido?
Mónica: Dependiendo del número total, ¿no? No se puede exceder del 200, en este caso. ¿Sí estás de acuerdo?
Brenda: Sí.
Mónica: Dependiendo del valor total. También ni modo que haya 201 de 200. Pues no. O sea mayor que 200, no.
--- 14:15.7

Quinto pasaje: *Interpretación de la probabilidad expresada con números decimales*.

ByM02 (video) -- 00:10.9
El Profesor solicita a Brenda y Mónica que efectúen la división en cada caso, añaden otra columna a la tabla en donde escriben el número decimal resultado de cada división.

En el pizarrón:

Número de hijos	*Número de trabajadores*	*Probabilidad*	*Probabilidad*
0	16	$\frac{16}{200}$	0.08
1	22	$\frac{22}{200}$	0.11
2	33	$\frac{33}{200}$	0.165
3	45	$\frac{45}{200}$	0.225
4	31	$\frac{31}{200}$	0.155
5	20	$\frac{20}{200}$	0.1
6	12	$\frac{12}{200}$	0.06
7	9	$\frac{9}{200}$	0.045
8	7	$\frac{7}{200}$	0.035
9	5	$\frac{5}{200}$	0.025

Profesor: Nos podrían ayudar ustedes a encontrar un significado a ese decimal, ¿qué podría significar 0.155? ¿qué es eso?

Brenda: Yo digo que los doscientos trabajadores se están tomando como una unidad, entonces se supone que ese 0.155 es un pedacito de eso, a eso se refiere, entonces si sumamos cada fracción que da aquí, nos va a dar el entero que sería igual a uno. Bueno, yo entendí eso.
Mónica: Si yo también, o se puede hacer como le hacemos, o sea, se puede dividir todo en doscientas unidades y ya sumas cada cosita. O sea los 31 representan... es que es como un porcentaje, es un porcentaje.
Brenda: No, bueno yo diría que no, en este caso no sería porcentaje, o sea, haz de cuenta que los doscientos lo estas tomando como un entero, o sea como una sola unidad y esto sería una fracción, porque si lo quisiéramos sacar como porcentaje, necesitaríamos sacar lo proporcional, o sea 200 es a 100 y 31 es a...
Mónica: Ah sí, por eso, no es exactamente el porcentaje pero si es la parte, no es porcentaje porque es de 100, pero es una fracción. Es la palabra porcentaje.
Profesor: Entonces que fue ese 0.155.
Mónica: Una parte del entero.
Brenda: Si, o sea, que el universo, o sea los doscientos trabajadores se están tomando como uno solo, entonces esta probabilidad se refiere a una parte, es una fracción, una parte de ese entero que se está tomando, entonces al sumar todo debe de dar un uno que sería el entero que se está considerando.
Profesor: Muchísimas gracias.
-- 2:07.5

Sexto pasaje: *Surgimiento, nuevamente, de la discusión en la definición de la variable dependiente e independiente*.

ByM02 (video)-- 3:32.7
Profesor: Respecto a esa última pregunta, el inciso b) dice: ¿qué depende de qué? ¿el número de hijos depende de la probabilidad o al revés?
Brenda: Que la probabilidad depende del número de hijos.
Blanca: Si pero... ahí fíjate, ustedes están manejando "la probabilidad depende del número de hijos" y cuando ponen variable independiente ponen "número de hijos", y variable dependiente "número de trabajadores"
Mónica: Ah sí es cierto, esta al revés, nos estamos contradiciendo aquí.
Brenda: No porque, si estamos diciendo que la probabilidad depende del número de hijos, quiere decir que la probabilidad es la dependiente, el número de trabajadores que es la probabilidad es la dependiente *(señala sus respuestas en el pizarrón)*.
Blanca: Pero ¿el número de trabajadores es la probabilidad?
Brenda: Si, son estas las probabilidades *(señala la columna de número de trabajadores en el pizarrón).*
Mónica: ¿no?
Blanca: No sé.
Profesor: Bueno, yo más bien lo que quería decir es que la pregunta esta parece que se está refiriendo nada más a la probabilidad y al número de hijos, porque dice ¿qué depende de qué? ¿el número de hijos depende de la probabilidad o al revés?, y eso ya lo contestaron.
Mónica: Ajá, que es la probabilidad es la que depende, entonces sería la dependiente, los trabajadores, dependen del número de hijos.
Profesor: Y luego donde dice ¿cuál de estas dos variables sería la variable dependiente y cuál la independiente?
Mónica: Aquí. *(señala su respuesta a esta pregunta en el pizarrón)*

Profesor: Entonces el número de hijos sería la independiente, y entiendo que el número de trabajadores es, como de ahí se sacan las probabilidades lo hacen equivalente, es como si estuviéramos hablando de las probabilidades, ¿así lo están entendiendo?
Ambas: Sí.
-- 5:29.6

Séptimo pasaje: *Discusión entre la diferencia entre probabilidad y número de trabajadores. Surgimiento de las dos relaciones importantes: probabilidad-número de trabajadores y número de hijos-número de trabajadores*.

ByM02 (video) --- 6:41.4
Profesor: Podemos estar de acuerdo en que entendemos por probabilidad, ¿simplemente un número suelto?
Brenda: No, es un número que, o sea, de la unidad que se está tomando, no puede ser así...
Profesor: O sea, "es relativo a". Entonces cuando ustedes hablan allá abajo *(se refiere a la respuesta al inciso b) que está escrita en el pizarrón)* del número de trabajadores como variable dependiente, también sería "relativo a" .
Mónica: Relativo al total de trabajadores.
Profesor: ¿O no? ¿cómo ven?
Mónica: ¿éste maestro? *(señala la respuesta a la variable dependiente "número de trabajadores" escrita en el pizarrón).*
Profesor: Sí, sí, como variable dependiente.
Mónica: O sea, es que el número de trabajadores es variable dependiente del número de hijos porque aquí ponemos la 'x' es el número de hijos y dependiendo de, si hay 4 hijos va a haber... el número de trabajadores.
Profesor: Si, si hay una asociación de dependencia, ahora nada más el detalle es sobre probabilidad, es que dice ¿el número de hijos depende de la probabilidad o al revés? Pero de la probabilidad.
Brenda: O sea, de esto *(señala los valores de probabilidad obtenidos en el inciso a) escritos en el pizarrón)*.
Blanca: Es que en la función de dependencia que ustedes establecen donde dice: "la probabilidad depende del número de hijos", ¿cuáles son las variables que manejan ustedes?
Mónica: Según lo que yo entendí, es la probabilidad. Esta *(señala la columna de número de trabajadores en el pizarrón).*
Blanca: Pero ¿esa es la probabilidad? ¿el 12 es probabilidad?
Brenda: No, de aquí se sacaría la probabilidad de 12 sobre 2.. *(señala el pizarrón).*
Mónica: Es que hay dos relaciones.
Blanca: ¿Cuáles son las dos relaciones?
Mónica: Esta *(señala el número de trabajadores)* y esta *(señala el número de hijos)*, y esta *(señala el número de trabajadores)* y el total *(se refiere al total de trabajadores)*
Blanca: Esa, y esa, y esa del total.
Brenda: Bueno, lo que pasa es que por ejemplo, se supone que el número de trabajadores depende del número de hijos, pero por ejemplo, de 31 esto sería la probabilidad, entonces de acuerdo a esta tabla *(señala la tabla)* se está sacando que los hijos, o sea, la probabilidad de que haya esos hijos es el número de trabajadores. A ver ¿cómo te lo puedo explicar?

Mónica: Es que aquí están las tres variables *(señala el pizarrón)*, tres porque de 0 hijos hay 16 personas que tienen 0 hijos de las 200, o sea, es que hay tres factores: ¿cuántos hijos?, ¿cuántas personas? ¿y de cuántas?, o sea, ¿cuántas personas de cuántas?; como que estas dos *(se refiere al número de trabajadores y al total de trabajadores)* van relacionadas muy directamente porque son 16 de 200, 9 de 200, pero esas dos se relacionan con esta *(con el número de hijos)*, porque esta es la variable de cuántos hijos son.
Blanca: Pero esas dos, 16 sobre 200, ¿son dos números o es un número?
Mónica: Es un número.
Blanca: Es un número. Entonces ¿cuáles son las dos variables que ustedes tienen?
Brenda: Lo que pasa es que la probabilidad, este número *(el número de trabajadores)* que te da sobre el universo, está dependiendo del número de hijos que tú le estas poniendo, porque por decir, allá arriba *(se refiere al inciso a) escrito en la parte de arriba del pizarrón)* te está pidiendo: "quiero la probabilidad de 4 hijos", entonces tú estás sacando la probabilidad de acuerdo a esta tabla *(señala la columna del número de trabajadores)*, que te da 31, o sea, 31 de todo el universo y esa probabilidad está dependiendo del número de hijos. Por eso la probabilidad está dependiendo de los hijos. Esa probabilidad es de todo el universo, pero lo estas tomando de acuerdo al número de hijos que tú estás buscando.
Mónica: Esta probabilidad *(señala la palabra probabilidad en la frase "La probabilidad depende del número de hijos" en el pizarrón)* es este número *(señala el 16/200, referida a la respuesta a la probabilidad de que el trabajador seleccionado tenga cero hijos)* que es el mismo que este *(señala el 31/200 respuesta referida a la probabilidad de que tenga cuatro hijos)*, o sea, cerrado en un solo número, y esa es la probabilidad, este *(nuevamente 16/200)* que es igual que este *(nuevamente el 31/200)*.
Blanca: Entonces ¿cuáles son las variables que ustedes manejan ahí?
Mónica: Yo diría que es el número de trabajadores entre total de trabajadores.
Blanca: ¿Es decir?
Mónica: Probabilidad.
Blanca: ¿cuál sería entonces la variable dependiente?
Mónica: La probabilidad.
Brenda: Sí, sería la variable dependiente.
--- 11:46.7

SEGUNDA PARTE

Ellas transcriben sus respuestas a la hoja de su reporte y borran sus respuestas a la primera parte del pizarrón, pero no la tabla de los datos.

Octavo pasaje: ***Primera respuesta a la recomendación que darían a la trabajadora social: Hay uso escaso de los datos de la tabla*.**

ByM03 (video)-- 4:35.4
Mónica lee el problema
Mónica: ¿qué recomendación le darías a la trabajadora social de cuántos boletos comprar para que la empresa pierda lo menos posible? ¡Ah, pero dice que se tiene que comprar por anticipado! ¿no? Pero si faltan se pueden comparar ahí.
Profesor: El caso es, que quieren comprar de tal manera que no falten...
Mónica: Pero tampoco que no pierdan.

Profesor: Ajá, porque ¿qué pasa si compran nueve?. Nueve más dos son... ¿qué pasa si compraran 11 boletos?
Mónica: Puede perder, pero ¿y si sale?
Profesor: Pero resulta que 9... a ver ustedes ¿cómo ven?. Pero ¿será posible que salgan los 9?
Mónica: Pues yo para no perderle y para no errarle, compraría dos nada mas porque sí hay familias que tienen 0 hijos, de hecho hay 16 trabajadores que tienen 0 hijos. Entonces nada más compraría dos para que vayan él y la esposa, y ya después si no sale, por lo menos ya tengo asegurados dos. Por eso preguntaba, si ya tengo asegurado dos... porque dice que pierda lo menos posible. En caso de que compre 3, y si por ejemplo sale una familia de estas, ya perdí un boleto, ya estoy perdiendo un boleto ¿qué voy a hacer con ese boleto?. Así pienso yo, yo compraría dos.
--- 6:03.4

Noveno pasaje: ***Establecimiento del primer supuesto implícito en su decisión de cuántos boletos comprar: los boletos deben comprarse con anticipación*.**

ByM03 (video) -- 6:03.8
Brenda: Yo compraría 5, porque de acuerdo a la tabla, hay más probabilidad de que sea una persona con tres hijos y luego sumándole el obrero y la esposa serían 5.
Mónica: Si, hay mucha probabilidad. Pero y si te salen por ejemplo estas *(señala los trabajadores con menos de tres hijos)*. Ah profe, es que eso es lo que le preguntaba; lo que yo le estoy preguntando es muy importante porque, en caso de que yo compre dos me van a faltar, pero si tengo que comprar los exactos que lo menos pierda o puedo comprar ya después boletos.
Profesor: Ya no, después ya no.
Mónica: Ah pues entonces si tiene razón ella, se tienen que comprar 5.
Profesor: ¿Por qué?
Mónica: Porque yo había dicho que 2 para no fallarle y no perder nada, porque dije: compro dos mínimo y ya, si me sale una familia de 0 hijos, o sea, que nada más iría la pareja pues ya se los doy y ya no perdí yo nada, pero si me salen más pues vuelvo a comprar, no hay problema. Pero no, porque tengo que comprar una vez. ¡Ah! entonces compraría 5 boletos porque la mayor incidencia de familias es de 3 hijos, o sea, el mayor número de trabajadores que hay por hijo es 3, o sea, las familias con 3 hijos son las que más abundan en la empresa, entonces como 3 hijos más el papá y la mamá pues ya son 5.
--- 7:33.9

Décimo pasaje: ***Discuten la decisión sobre cuantos boletos comprar usando más fuertemente los datos que la tabla de probabilidad les proporciona. Surge la necesidad de calcular la frecuencia acumulada*.**

ByM03 (video) -- 7:47.4
Profesor: ¿Tú que dices Brenda de eso?
Brenda: Bueno, yo diría eso, comprar los 5, pero también existe la probabilidad de que no vaya a completar después.
Mónica: Pero va a ser un sorteo, yo siempre me lo imagino como una canasta con papelitos, entonces, hay más probabilidades de que yo agarre un trabajador de estos 45 *(se refiere a los trabajadores que tendrían 3 hijos)*, o sea, en el círculo que dibujamos *(se refiere a la tabla)* éste *(señala el 45)* ocupa más, hay más probabilidades, es como que la que menos falla, es como que azar.

Brenda: Aunque bueno, también por decir un margen más grande sería... bueno más seguro sería comprar los 9, porque si te fijas, de aquí hasta acá *(señala en la tabla la columna de probabilidades las correspondientes de 0 a 6 hijos)* es donde se ocupa la mayor parte, de 0 a 6 es donde se está ocupando la mayor parte inclusive (*se podría tomar)* un poco más, ya en la de 7 hijos que es donde llega a 9 *(es decir, añadiría los trabajadores que tienen 7 hijos)*, o sea, es donde está la mayor probabilidad, ésta es muy poquita *(señala la probabilidad de que se tengan 8 ó 9 hijos)*. Estos dos tienen muy poquita probabilidad y serían 7 más los... que serían... Si te fijas realmente ya nada más tendrías un margen de 13 o sea muy pocos.
Mónica: Pero dice ¿cómo le hace la empresa para gastar lo menos posible? Requieres gastar lo menos posible.
Brenda: Por eso, por decir, sí, este es el que tiene más *(se refiere a los trabajadores que tienen 3 hijos)*, pero si compras nada más 5, resulta que, tienes todavía 31 posibilidades de que salgan con 4, 20 posibilidades de que salgan con 5 y 12 todavía con 6 y así, entonces si te fijas hasta acá *(señala el renglón con 6 hijos)* esto abarca la mayor posibilidad, si tienes cubierto todo esto, ya no vas a tener problemas de que la familia (*premiada)* se vaya a quedar sin boletos. Si te fijas realmente todo esto está cubriendo la mayor probabilidad.
Mónica: Pero es que entonces, yo quiero que nadie se quede sin boletos, pues entonces de una vez compran los... 11...de una vez para ya no errarle.
Brenda: No pero la cuestión es...
Mónica: Ahorrar más es la cuestión, ¿qué no?
Brenda: Sí ahorrar de modo que no se vaya a..., pero (*con mi propuesta)* también va a quedar abierta la posibilidad de que no vayan a alcanzar los boletos. Si en todo caso fuera, ay pues ya compra todos, pero no ¿verdad?, ya no habría problema; pero en todo caso, yo diría que si sería hasta aquí *(trabajadores con 6 hijos)* o hasta aquí *(trabajadores con 7 hijos)*, cualquiera de los dos, porque entonces abarcaría la mayor parte de la probabilidad.
Profesor: Una sugerencia, me gusta la idea que están ustedes desarrollando pero sería bueno como registrarla tantito, ustedes dicen: “porque hasta ahí abarcaría la mayor parte de la probabilidad”, ese fue su último argumento, entonces ¿por qué no ven que tanta probabilidad abarca según el número de boletos?, o sea por ejemplo, si no compran ningún boleto ¿qué probabilidad abarcaron?, si compran un boleto ¿qué probabilidad abarcaron? Y así...
--- 11:23.7

Undécimo pasaje: *Uso de la probabilidad acumulada para argumentar cuántos boletos comprar. Ambas se preocupan en diferentes intereses: en los trabajadores o en la empresa*.

ByM04 (video)--- 0:4.18
En el pizarrón, a la tabla anterior le añadieron la columna de la probabilidad acumulada:

Número de hijos	*Número de trabajadores*	*Probabilidad*	*Probabilidad*	*Probabilidad Acumulada*
0	16	$\frac{16}{200}$	0.08	0.08
1	22	$\frac{22}{200}$	0.11	0.19
2	33	$\frac{33}{200}$	0.165	0355
3	45	$\frac{45}{200}$	0.225	0.58
4	31	$\frac{31}{200}$	0.155	0.735
5	20	$\frac{20}{200}$	0.1	0.835
6	12	$\frac{12}{200}$	0.06	0.895
7	9	$\frac{9}{200}$	0.045	0.94
8	7	$\frac{7}{200}$	0.035	0.975
9	5	$\frac{5}{200}$	0.025	1

Mónica: Es que yo también estoy de acuerdo contigo porque sí es cierto por ejemplo hasta aquí (*señala el renglón correspondiente a 7 hijos*) ya es casi seguro que les toquen, un 94% (*es la probabilidad acumulada hasta 7 hijos*) ya casi seguro que le toque a todos, más casi no estás ahorrando porque imagínate que te toque ya no éste (*señala el renglón de 4 hijos*) sino éste ya estás perdiendo (*señala el renglón correspondiente a 6 hijos*). Un escalón menos ya estás perdiendo, dos más, imagínate que te toqué éste (*refiriéndose a 4 hijos*) o éste (*refiriéndose a 2 hijos*), ya perdiste mucho más. Imagínate que, a lo mejor es una empresa chica y no sé, a lo mejor los boletos están en mil pesos, así que no es lo mismo gastar...
Mónica: Yo le pondría un 50% (*refiriéndose más bien al 58% de probabilidad de que el trabajador premiado tenga 3 hijos*), o si acaso un 70% (*refiriéndose al 73% de probabilidad de que el trabajador premiado tenga 4 hijos*), porque luego para que me salgan con que perdieron... aunque claro, maestro también está mal el problema porque no es cuestión de decirles oigan les vamos a dar boletos y si les toca pues que bueno porque es como decirles a los que tengan más (*hijos*) llevan las de perder porque no vamos a comprar para todos y luego que... Lo ideal sería que si la empresa está haciendo la campaña, pero eso ya no tiene que ver con matemáticas, sino que es más ético, si la empresa está ofreciendo pues que lo ofrezca bien.
Brenda: En todo caso si se está buscando que no vaya a gastar de más y si alcanzan bueno y si no, no, pues yo estaría por 4 hijos.
Mónica: Ajá, yo también y si no alcanzan pues ya ni modo.
Profesor: Y si se pusieran codas, si se pusieran así estrictas, la lana de veras es que... vamos a ofrecer este beneficio pero... trabajadora social de veras trata de hacernos gastar lo menos posible, ¿cuál sería el valor que escogerían?
Brenda: En la codez, en la codez sería tres hijos.

Mónica: Yo también porque es donde hay más trabajadores.
Profesor: ¿Y en términos de probabilidades?
Brenda: Hasta aquí alcanzaría más de la mitad, estaríamos hablando de más de la mitad. Es decir tiene bastante probabilidad.
-- 2:54.6

Décimo segundo pasaje: *Sobre su percepción y aceptación del azar*.

ByM04 (video)--- 9:57.1
Profesor: Supongamos que yo soy el patrón y ustedes son las trabajadoras sociales y me dicen ‘te convienen cinco boletos’ y me van a convencer: ‘cinco boletos es mucho, ¿por qué tanto?’, ustedes ¿qué argumento me darían?
Brenda: Bueno con cinco boletos tendría asegurado que la mayor parte de… o sea más la mitad de la probabilidad de que la familia que resultara seleccionada complete con esos boletos. Abarcaría a la mayor parte de sus trabajadores.
Profesor: Oigan pero eso es posible, a lo mejor eso ni sucede porque es azar…
Mónica: Pues es lo que yo decía…
Profesor: …acuérdense que será una ruleta o una tómbola en la que echaremos los boletos y es azar y pues, quién sabe si...
Mónica: Pero por la misma razón que es azar igual pueden salir estos (*señala en el pizarrón los renglones referidos a los trabajadores que tendrían de tres a nueve hijos*).
Profesor: Claro, como es azar, pueden salir nueve.
Mónica: Bueno, lo que diríamos aquí lo que le aseguramos es el 50%, el otro 50% definitivamente va a tener que comprar uno, dos, tres, cuatro boletos por su cuenta, pero lo que usted quiere también es congraciarse o llevarse bien con los trabajadores, entonces pues con más del 50% ya tiene probabilidad de que lo saque, es decir, tiene asegurada la mitad.
Profesor: Pero yo sigo así como renuente, porque es azaroso, como que es la tómbola, como están tan seguras de esos números, porqué le creen a esos números si es todo azar, ¿es posible que en medio de tanto azar ustedes, que nadie sabe que va a salir, ustedes puedan asegurarme qué va a salir?
Mónica: No, no le podemos asegurar nada. Yo digo que en los azares no se puede asegurar nada.
Brenda: Bueno, es que no se puede asegurar, pero de lo que se habla aquí es de probabilidad, lo que puede pasar. Como además, se encuentran más concentrados los trabajadores en esa parte (*señala los renglones de la tabla correspondientes a trabajadores con 3 hijos o menos*) hay más posibilidades de que no vaya a perder tanto porque más del 50% (*de los trabajadores*) se juntan en ese espacio. No podemos asegurarlo pero es más factible, puede que suceda más, que salga más una de estas personas (*señala los renglones correspondientes a los trabajadores con 3 hijos o menos*), es más fácil, que del resto porque son más, son mayoría.
Profesor: O sea que ustedes me están diciendo que uno puede predecir qué boleto va a salir.
Mónica: Cuál es más probable pero no cuál va a salir. No hay manera.
Profesor: Pero existirá alguna manera, ¿es ignorancia nuestra el no saber qué boleto va a salir?
Brenda: No es algo que tú puedas controlar, es la suerte.
--- 13:50.6

Décimo tercer pasaje: ***Sobre el espacio muestral dentro del contexto del problema.***

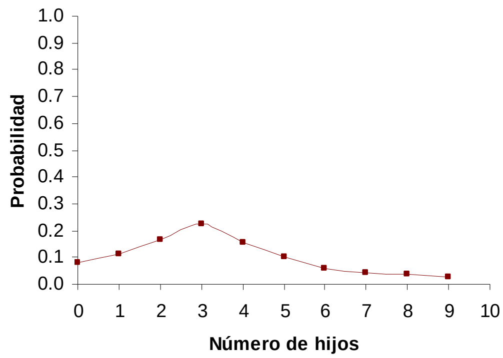
ByM04 (video) --- 15:13.2
Profesor: ¿Cuántos casos posibles pueden haber de hijos? Puede ser que en una de esas salga una familia de 11 hijos.
Mónica: No, 9.
Profesor: ¿y mínimo?
Brenda: ¿Mínimo? Cero hijos.
Profesor: ¿es posible que pueda extenderse el número 200 a más valores? ¿por qué?
Mónica: No porque sólo son 200 trabajadores.
Profesor: ¿o sea que la probabilidad sólo tiene que ver con esos trabajadores o con otras cosas ahí?
Mónica: No, nada más con esos trabajadores.
-- 16:00.5

TERCERA PARTE

Décimo cuarto pasaje: ***En la elaboración de la gráfica*****.**

Mónica tiene el gis, ella hace la gráfica mientras Brenda le va dando indicaciones, de manera que lo que Mónica va haciendo está consensuado por las dos.
Mónica hace unos ejes cartesianos en el pizarrón. El sistema coordenado sólo incluye el primer cuadrante.
ByM04 (video) --- 16:48.0
Brenda: Aquí sería el número de hijos (*señala el eje de las abscisas*).
Mónica etiqueta los ejes: número de hijos y número de trabajadores.
Brenda: No, pero entonces ahí sería la probabilidad (*se refiere al nombre de la variable dependiente*).
Mónica: ¡Ah! Aquí ya podemos saber que en vez de poner número de trabajadores, ya vamos a poner directamente la probabilidad. Vamos a trabajar con la probabilidad directamente.
Brenda: Aquí (*señala el eje de las abscisas*) sería, uno, dos,… así hasta nueve.
Mónica pone la escala en el eje de las abscisas.
Brenda: Luego para arriba, ¿de qué la empezamos? ¿de 0.1, no?
Brenda: Bueno, sería de cero hasta uno y luego ya sería irle poniendo las probabilidades. A ver, divídelo…
Mónica: A ver, primero mitad y mitad. Punto cinco…
Brenda dicta los valores a Mónica y ella los va colocando en la gráfica.
Mónica: ¿Hacemos la gráfica?
Brenda: Sí, hay que hacer la gráfica.
Brenda une los puntos.

En el pizarrón:

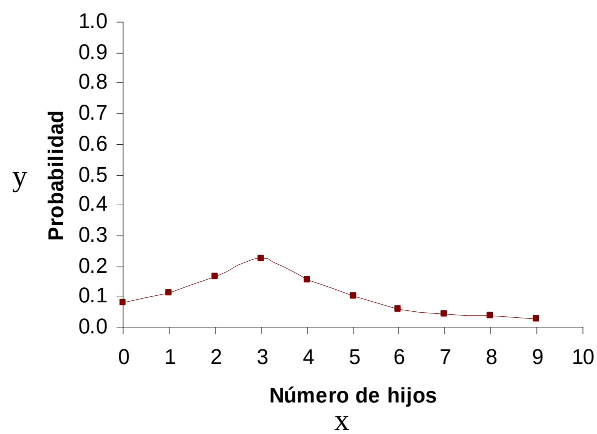


ByM05 (video)--- 05:36.8

Décimo quinto pasaje: *Sobre la noción de función*.

ByM05 (video)--- 05:36.8
<u>Profesor</u>: Representa con una letra la variable dependiente y con otra la variable independiente.
<u>Mónica</u>: ‘x’ y ‘y’ (*las escribe en los ejes de la gráfica que está en el pizarrón*). Ya.

En el pizarrón:



<u>Profesor</u>: Utiliza estas letras para representar en notación funcional el resultado obtenido en el inciso b) de la parte 2. Es decir con que probabilidad el trabajador premiado ocuparía exactamente el número de boletos que compró la trabajadora social.
Dudan, leen una y otra vez la pregunta.
<u>Mónica</u>: Sería esté no. (*Señala el punto máximo en la gráfica* y *escribe en el pizarrón y = 0.225*), ¿No está bien?

En el pizarrón:

$$\boxed{y = 0.225}$$

Brenda: (*dirigiéndose a Mónica*) ¿A qué se refiere con notación funcional? ¿una ecuación?
Dudan..
Mónica: ¡Funciones!, F de x, f de…
Profesor: ¿Sería posible eso (*se refiere a la gráfica que está en el pizarrón*) escribirlo en notación de funciones? ¿qué es una función? Esa gráfica, suponiendo que fuera continua, ¿sería una función?
Mónica: Sí.
Profesor: ¿Por qué?
Mónica: Pues como todas (*se refiere a los valores involucrados*) se corresponden, por ejemplo f de 3 que corresponde a 0.225, ¿no? (*Escribe en el pizarrón F(3) = 0.225*)

En el pizarrón:

y = 0.225 F(3) = 0.225

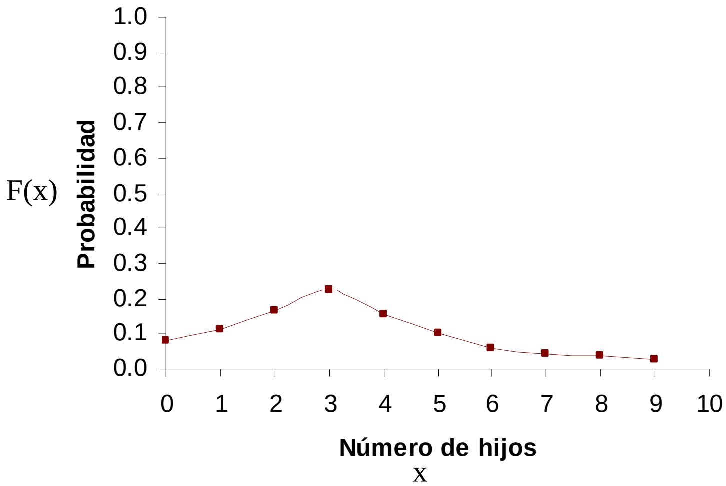
Profesor: ¿Qué dice Brenda?
Brenda: Pero lo que pide es poner… ¿cómo es la ecuación?
Blanca: ¿Una función nada más se expresa en forma de ecuación?
Brenda: No.
Blanca: ¿De qué otra forma?
Mónica: Por ejemplo. No, una función no necesariamente tiene que ser una ecuación, por ejemplo está podría ser una función (*escribe F(x) = 3x*), siendo que no es una ecuación.

En el pizarrón:

y = 0.225 F(3) = 0.225 F(x) = 3x

Profesor: Pero esa sería una expresión algebraica, ¿puede algo que no sea una expresión algebraica, una gráfica o una tabla, ser función? ¿tiene a fuerzas que ser una fórmula?
Mónica: La tabla o la gráfica es la representación de una función.
Profesor: ¿Qué dice Brenda?
Brenda: No, porque pueden variar los datos o que estén dados.
Mónica: De hecho la fórmula se saca de la tabla y la gráfica también.
Profesor: ¿Qué opinas de eso que escribió Mónica de F(3) = 0.225? ¿Eso sería una notación funcional?
Brenda: Bueno pues es que… que aquí dice que nombremos las variables y utilice estas las letras para escribir en notación funcional el resultado obtenido, o sea que tendríamos que utilizar ‘x’ y ‘y’, puesto que esas fueron las letras que escogimos.
Profesor: ¿Quién sería ‘x’?
Brenda: En este caso, pues sería tres.
Profesor: Pregúntale a Mónica, ¿quién sería ‘x’ y quien sería ‘y’?
Mónica: Pues… ‘x’ sería 3 y ‘y’ sería 0.225. La función de ‘x’ es ‘y’, es decir el resultado.
Brenda: Pero en todo caso… no lo sabes. Lo que yo digo es utilizar…
Mónica: ¿En vez de F, ‘y’, aquí (*señala la F en F(x)=0.225*)?

Brenda: No, es que de cualquier forma haría falta especificar que es una función de 'y'. Lo que podríamos utilizar es, para tomar exactamente las mismas letras, ponerle allá F de 'x' (*se refiere al nombre del eje de las ordenadas*). Es decir, cambiar aquella, ponerle F(x) en lugar de 'y'.
Mónica: Pero necesitamos ponerle F porque es lo que dice ella (*se refiere a Blanca*) que hace falta ponerle la función.
Brenda: Por eso te digo, mejor cambiar aquella (*señala la y en el eje de las ordenadas de la gráfica*).
Mónica: (*Borra la y en el eje de las ordenadas escribe F(x)*) Así nos queda una sola variable.



ByM06 (video)--- 02.50.4

Décimo sexto pasaje: *Definición del dominio de la función*.

ByM06 (video)--- 03.03.7
Profesor: Describe en el contexto del problema, con tus propias palabras, la información que proporciona la variable dependiente de la variable independiente.
Mónica: ¿La variable dependiente (*señala en el pizarrón el eje de las ordenadas*) de la… (*Señala en el pizarrón el eje de las abscisas*)?
Brenda: Bueno pues se supone que esta (*señala en el pizarrón el eje de las ordenadas*) te está diciendo que probabilidad hay de que… cuál es la probabilidad de la variable independiente.
Mónica: Lo que dice es que todos estos números (*señala la escala del eje de las abscisas*), ninguno pasa de esto (*señala el punto máximo de la gráfica*). Es decir, ninguno va más para arriba.
Brenda: Pues básicamente la variable dependiente te está diciendo cuál es la probabilidad.
Profesor: ¿Llegan a un consenso?
Ambas: Sí
Profesor: Bueno, la que sigue. Describe el dominio de la variable independiente y el rango de la variable dependiente.
Mónica: Pues de 0 a 9 es el dominio de la independiente y el rango de la dependiente sería de 0.025 a 0.225 (*escribe en el pizarrón [0,9] y [0.025, 0.225]*).

En el pizarrón:

[0,9] [0.025,0.225]

Blanca: ¿Por qué lo encerraron en un corchete? ¿qué significa eso?
Mónica: Que va de… incluye del 0 todos los números, todos los números, hasta el 9. Cerrado.
Blanca: ¿Todos los números?
Mónica: Entre el cero y el nueve, sí.
Blanca: ¿Todos los números que hay entre cero y nueve?
Mónica: Sí
Blanca: Cuatro punto cinco está cero y nueve.
Mónica: Pues también es parte porque... No está en la gráfica, pero está en el dominio. Igual, esto (*señala el 4.5 en el eje de las abscisas y traza con el lápiz su proyección hacia la gráfica y después hacia el eje de las ordenadas*) no tiene un valor escrito pero sí existe.
Blanca: ¿Si podemos calcularlo?
Mónica: Sí.
Brenda: Yo diría que no, se supone que estamos…
Mónica: Pero es que, es que… viéndolo así para no poner cuatro punto cinco hijos… pues es de que... sí son valores enteros, pero es que en las gráficas también te sale, por ejemplo población 1970.3, ni modo que haya 0.3 de humano, pero sí se puede calcular, es como un valor de…
Brenda: Bueno es que sí es cierto, se supone que nada más serían los que te están dando. Las coordenadas que te está dando la gráfica únicamente son estos (*señala los puntos de la gráfica*) no son mitades ni nada de eso, sólo los enteros.
Mónica: Entonces para qué es la gráfica si vamos a dejar los puntos así aislados.
Brenda: Sí, pero es que realmente no puedes sacar la probabilidad de 0.5.
Mónica: Sí, no puedes tener 8.5 hijos, es lo que yo digo.
Brenda: Pero por ejemplo, o sea, o sea, la gráfica sí está abarcando todo, sí está abarcando todo este espacio.
Blanca: ¿Entonces están de acuerdo en que ese es el dominio?
Ambas: Sí.
Profesor: ¿Y el rango? ¿Cómo lo encontraron?
Mónica: Igual, vimos cual es el valor más chiquito y cuál era el más grande y no puede salirse de ahí, ni para abajo ni para arriba.
Profesor: ¿Qué significa el 4, 5, 6…?
Ambas: El número de hijos.
Profesor: Entonces no tienen sentido los decimales.
Ambas: No.
Blanca: Pero a mí me acaban de decir que sí tenían sentido.
Brenda: No, lo que pasa es que no podemos saber cuál sería el punto, la coordenada aquí (*señala un punto que está entre 0 y 1 en la gráfica*) en este punto sería decimal, pero en este caso uno no usa los decimales, no sé la probabilidad de la mitad de un hijo.
Mónica: Si fuera dinero sí se podría saber.
Brenda: Sí podemos usar los decimales, porque entran dentro de la misma gráfica, pero en ciertas cosas no usas decimales.
Profesor: En cambio en el eje vertical ¿si tienen sentido los decimales?

<u>Brenda</u>: Es la probabilidad.
<u>Profesor</u>: Por ejemplo, puede haber 0.255, o sea valores intermedios.
<u>Ambas:</u> Sí.
<u>Profesor</u>: Ahí sí se puede escribir como intervalo cerrado. El que está dando lata es el de arriba (*se refiere al orden en que escribieron ambos intervalos en el pizarrón, el de arriba es el dominio, el de abajo es el rango*), el de abajo como que no hay discusión, el de arriba con la salvedad que dicen ustedes. “bueno, así se suele manejar”. Ese sería el argumento, ¿sí?
<u>Ambas:</u> Sí.
-- 10.57.8

Décimo séptimo pasaje: *Definición de función en la relación establecida.*

ByM06 (video)--- 10:57.0
<u>Profesor</u>: Describe el método para encontrar el valor de la variable dependiente a partir de la variable independiente.
<u>Mónica</u>: ¿La ecuación?
<u>Profesor</u>: La F, es decir la variable dependiente, a partir del valor de la variable independiente.
<u>Mónica</u>: Pues fijarse en la gráfica o sea checar.
<u>Brenda</u>: ¿Con un método más allá del gráfico?
<u>Profesor</u>: No dice un método más allá del gráfico. ¿Qué opinas de la respuesta de Mónica?
<u>Blanca</u>: Sí, por ejemplo si ‘x’ es igual a dos, ¿cuánto vale la variable independiente?
<u>Mónica</u>: Pues se vería aquí, (*se dirige a la gráfica y señala el dos en el eje de las abscisas, la proyección a la gráfica y posteriormente al eje de las ordenadas*) o sea dos y checar el resultado aquí, que es el mismo que de la tabla.
<u>Blanca</u>: ¿Lo sacarías de ahí o de la tabla?
<u>Mónica</u>: Es lo mismo. Bueno, si estuviera esto (*señala el eje de las ordenadas*) perfectamente numerado, no así tan… sería lo mismo.
<u>Brenda</u>: Es decir si se pudiera tener mayor precisión, pues sí, aquí pues está un poco más complicado.
<u>Mónica</u>: Ajá, si tuviera cuadrícula, pues si, la tabla y la gráfica es lo mismo.
Silencio
<u>Mónica</u>: Sin embargo estaba pensando ahorita ¿Cómo le podemos hacer? Lo primero que pensé fue regresión, hacer una regresión y encontrar una ecuación, pero después dije no, porque no tiene nada que ver el número de hijos con las personas, no es de que entre más hijos más personas haya sino de que pueda haber… si va disminuyendo, no va a llegar un momento en que se va a acabar, porque son hijos si fueran otros datos como dinero o alguna otra variable, puede que sí, pero como hijos no, pero como en, no sé, 15 ya se va a acabar, a partir de 3 va a ir disminuyendo.
<u>Profesor</u>: ¿La relación que establecieron es una función matemática? ¿o sea es función? ¿Recuerdan las propiedades de la función? ¿qué es una función?
<u>Mónica</u>: Que hubiera dos variables.
<u>Brenda</u>: Que para el valor de ‘x’ haya un valor en ‘y’.
<u>Profesor</u>: ¿Y puede haber dos ‘x’ que vayan a dar a un mismo valor de ‘y’?
<u>Ambas:</u> Sí
<u>Profesor</u>: ¿Puede haber una ‘x’ que pueda dar dos valores de y? Es decir, ¿un número de hijos que de dos probabilidades diferentes?
<u>Ambas:</u> No.
<u>Profesor</u>: Entonces ¿es función o no?

Brenda: Sí, si es función porque para cada número de hijos hay una sola probabilidad.
-- 15:24.3

Décimo octavo pasaje: *La continuidad como principal diferencia entre la variable independiente usada en el contexto manejado en esta actividad y las variables usadas en sus cursos normales de matemáticas*.

ByM06 (video) -- 15:25.0
Profesor: ¿Qué diferencia encuentras entre la variable independiente de esta función de probabilidad y las variables independientes de las otras matemáticas que tú conoces, por ejemplo del álgebra y del cálculo?
Mónica: Qué es limitada.
Profesor: Supongamos que tiene un sentido "más hijos" indefinida, aun así ¿Hay más características? ¿Es la única característica que ustedes encuentran entre esta función y las otras variables independientes de álgebra y cálculo?
Mónica: Lo de los decimales.
Profesor: Que se suele manejar variable continúa.
Mónica: Enteros.
Profesor: Aquí enteros. ¿Te pareció que esa es una variable discreta? A esa se le llama variable discreta, cuando nada más puede tomar valores enteros.
Mónica: Ah, pues es así, discreta.
-- 16:50.0

Décimo noveno pasaje: *El problema de obtener una ecuación a partir de los datos establecidos como otra diferencia entre las funciones que han trabajado en anteriores cursos de matemáticas y el contexto de esta actividad.*

-- 16:50.0
Profesor: Qué más, piensen ¿Habrá alguna característica más que distinga esta variable de las otras que trabajan en matemáticas?
Brenda: Normalmente en álgebra manejamos ecuaciones. Todas las gráficas se manejan por ecuaciones, todos los valores de'y'se van a dar por la ecuación. Aquí los valores de y también se dan según una 'x', pero no tienen una ecuación, son así como… aleatorios. No llevan un patrón así seguido.
Mónica: Es lo que decía, no podemos sacar una ecuación porque no nos va a dar nada preciso o sea de número de hijos no se puede sacar una ecuación.
Brenda: Bueno más que la variable independiente diferente, como que es la variable dependiente la diferente, porque no podemos establecer cuál será la dependiente. A diferencia del álgebra, con la independiente podías sacar la dependiente exactamente y en este caso no, son datos que te están dando y tú no puedes... es muy aleatorio, es más difícil que con álgebra. Con álgebra tú podías sacar 'y' con la ecuación y aquí no.
ByM06 (video) -- 10:57.0
En audio -- Segundo casete
Mónica: Aunque sí se observa que primero crece y luego ya disminuye, disminuye y como son hijos y como somos humanos son funciones más o menos, o sea va a ir disminuye y disminuye hasta que llegue a 12 y ya. En ese caso pero por ejemplo más datos humanos como… número de personas divorciadas o cuántos se han casado, ahí no sería viable sacar una ecuación porque no te iba a decir nada.

Brenda: Bueno en realidad sí puedes sacar una ecuación, pero eso no te daría… no sería muy exacto. Esa es la diferencia, que no te pueden dar datos exactos.
-- Segundo casete

Vigésimo pasaje: *Sobre el valor esperado del número de hijos al realizar la rifa de los boletos*.

En audio-- Segundo casete
Profesor: Sí tomo la tómbola y le doy vueltas, ¿cuál sería el número de hijos que creen que tanga el trabajador seleccionado?
Brenda: Tres hijos.
Profesor: ¿Por qué?
Mónica: Es el de mayor incidencia.
Profesor: Supongamos que sale un trabajador que tiene dos hijos, si salió dos, la siguiente salida de la tómbola ¿cuántos hijos creen ustedes que saldría?
Mónica: Tres.
Profesor: Supongamos que sale tres otra vez y volvemos a sacar otro papelito, ¿cuántos hijos tendría el trabajador ahora?
Brenda: Pues podría salir dos o cuatro.
Profesor: Vamos a suponer que sale 4, y volvemos a seleccionar un papelito, ¿cuánto creen ustedes que salga?
Mónica: Pues es que nomás está entre esos números. Puede salir cualquier número, pero las mayores probabilidades están entre esos números. Pero ya después de varias veces se van a acabar los 3.
Profesor: No, siempre estamos regresando el papelito que sale.
Mónica: Ah, pues tres siempre, es el que se espera.
Profesor: Pero si saco uno al azar y sale que el trabador tiene un hijo. Lo regreso y vuelvo a revolver, ¿cuál es el valor esperado?
Mónica: Pues tres maestro, ¿es que por qué no?
Profesor: Y bueno, si sacamos uno más, ¿podemos saber cuál es el que le sigue? ¿Alguna forma de conocer cuál va a salir antes de sacarlo?
Ambas: No.
Profesor: ¿Es posible que vuelva a salir el uno?
Brenda: Es poco probable, es más probable que salga entre el 2 y el 4 a que vuelva a salir el uno.
Mónica: Más sí puede volver a salir.
Profesor: Entonces ¿habrá orden de aparición en el número que salga de la tómbola?
Brenda: O sea de que primero saque uno y salga uno, después vuelva a sacar y salga dos y así, pues no, no lo hay.
--- Segundo casete

Vigésimo primero pasaje: *La variable dependiente como la principal diferencia entre las funciones que han trabajado en anteriores cursos de matemáticas y el contexto de esta actividad.*

En audio--- Segundo casete
Blanca: Comparen con los modelos que hayan trabajado en álgebra o cálculo. Por ejemplo, que ustedes van en un vehículo con una velocidad constante y ustedes quisieran tratar de describir esa situación, ¿cuáles serían las dos variables que manejarían?
Brenda: Sería el tiempo y la distancia.

Blanca: ¿Cuál sería la variable dependiente ahí?
Brenda: La dependiente es la distancia y la independiente el tiempo.
Blanca: Y díganme, si nosotros estamos ahorita en el tiempo uno, después ¿en cuál estaría?
Ambas: En el dos.
Blanca: ¿Y después?
Ambas: En el tres.
Blanca: ¿Por qué no pasa lo mismo aquí?
Mónica: Porque es azar.
Brenda. Ajá.
Profesor: La pregunta inicial era qué diferencia había entre esta variable y las variables como las que menciona Blanca.
Mónica: Las otras son constantes, continúas, se espera que sigan y que vayan, en cambio esta no sabes, no puedes predecir algo.
Brenda: Sí, a lo mejor hay personas, como dices, bueno, voy al tiempo uno, dos y tres, voy una hora, una hora, pero a lo mejor… bueno en este caso sí va así seguido, pero a lo mejor hay casos en que no había nadie con 6 hijos entonces la variable ya no tendría que ser toda seguida.
Mónica: Pues obviamente sí se pueden ordenar siempre las variables (*se refiere a los valores que puede tomar la variable*), así de menor a mayor y de mayor a menor también, pero eso no significa que de mayor a menor van a ir creciendo o van a ir disminuyendo igual, o sea, el que sea de menor a mayor, el que estén ordenadas, no significa que también estén creciendo uniformemente.
Blanca: Y entonces ¿está variable independiente es igual a las otras variables que ustedes han manejado?
Mónica: No.
-- En audio
ByM07 (video) -- 00:00.0
Profesor: Perdón, pero, ¿podrían sintetizar alguna idea de lo que están discutiendo hasta ahora?
Brenda: Yo decía que si lo comparamos con el ejemplo del carro, ahí la ‘x’ siempre va a ser continua, tiene que haber uno y luego dos, o sea siempre una hora va a seguir a otra hora y así, y aquí por ejemplo, podría haberse dado el caso de que a lo mejor no había hijos, no había una persona con cierto número de hijos. Como no está que tiene que ser así, como es muy aleatorio todo esto, no tiene porque ser así, creciendo igual.
Mónica: Lo que yo dije es que la variable independiente está ordenada y se pueden ordenar tanto la dependiente, como la independiente, pero en otro tipo de gráficas normales a cada valor en ‘x’ corresponde un valor en y, es decir, tiene un comportamiento esperado, dependiendo de la gráfica, no nada más gráfica recta, que puede ser logarítmica o exponencial o polinómica, cualquier tipo de gráfica, pero para cualquier valor en ‘x’ se espera uno en ‘y’. Aquí no, para un valor en x, no sabes cuál va a ser el valor en ‘y’, puede ser más grande, más chico, es algo que no está controlado. La ‘x’ no controla a la ‘y’.
ByM07 (video) -- 02:34.2
ByM08 (video) -- 00:00.0
Blanca: Bueno, estábamos en si era diferente o no era diferente y si es diferente ¿en qué?
Brenda: Bueno yo digo que básicamente la diferencia, lo mismo que decía ella, pero más en todo caso, más que nada la diferencia radica en la variable dependiente

porque, en las funciones matemáticas o algebraicas que hemos visto, con la ecuación podemos saber el valor de la variable dependiente, pero con esta varía mucho. Básicamente la que varía más es la dependiente.
Mónica: Yo también estoy de acuerdo con ella en que la que varia es la variable dependiente. Varía de forma inconstante o impredecible.
Blanca: Pero ¿por qué impredecible?
Brenda: Sí, porque este valor no es algo que esté dado por una…
Mónica: Función.
Brenda: Sí, por una función, por una ecuación, sino que porque son datos que así se dan o sea que, que pueden ser más o que pueden ser menos, que nos están determinados por algo sino por las circunstancias, por el azar yo creo.
Mónica: Ajá, porque pueden estar determinados por diferentes factores por ejemplo, esa población puede haber tenido tres hijos porque se localizaba en México o en Estados Unidos en donde la mayoría de las personas tienen tres hijos, pero en poblaciones de otros lugares donde el número de hijos crece por persona, la gráfica ya no iba a ser igual. En cambio si hablamos de dinero, bueno, no tanto de dinero sino de otro tipo de cosas como habíamos visto anteriormente, sí se espera algo, pero aquí el número de hijos… En este tipo de gráficas en donde están midiendo el azar, no podemos saber cuándo va a bajar y cuando va a subir, no tiene un comportamiento predestinado.
Blanca: ¿La variable dependiente? ¿la probabilidad?
Ambas: Sí.
-- 02:07.3
ByM09 (video)-- 00:00.0
Blanca: O sea lo que ustedes piensan es que lo que hace diferente a esta función del resto es la variable dependiente y no la independiente.
Brenda: Sí, yo digo que esa sería la gran diferencia. Porque igual por ejemplo, aquí (*señala un punto sobre el eje de las abscisas*) puedes no tener un dato, pero esto (*señala la altura de ese punto sobre el eje de las abscisas*) es lo que siempre va a estar variando. En las otras tienes que seguir cierto orden y a lo mejor aquí (*señala el eje de las abscisas*) te puedes brincar un dato, pero esto no lo puedes tener determinar.
Mónica: Cualquier expresión, cualquier ecuación se esperaría que tuviera cierto comportamiento, por ejemplo, si por ‘x’ o ‘y’ razón no hubiera familias con 3 hijos, ahí iría para abajo otra vez la gráfica y no por eso estaría mal la gráfica.
Brenda: Es como lo que estaba preguntando ahorita el profe, que si podemos saber exactamente qué número va a salir, pues no, no se puede saber. Y toda la situación es de que no puedes saber cuándo va a salir, a diferencia de la matemática que nos decía, sí podemos saber cuánto se va a recorrer en cierto tiempo y en esta no, no puedes decir: “va a salir este número”.
Mónica: Ajá.
Brenda: Como que más bien es una aproximación, no es más… es probabilidad.
-- 02:31.0

Vigésimo segundo pasaje: *Poniéndole nombre a la variable independiente, lo que permite resumir las características importantes de esa variable*.

ByM09 (video)-- 02:31.4
Profesor: Como ven esta variable del eje de las ‘x’ es muy distinta a la que están acostumbradas a trabajar, entonces, ¿qué nombre le darían? Si es distinta, pues vale la pena ponerle un nombre a la variable.

Brenda: ¿o sea a la x?
Profesor: Porque la y no hay problema ¿no? ¿cómo se llama la y?
Mónica: Probabilidad.
Profesor: Entonces ya nada más nos queda la x, ¿no?
Mónica: Estadística. Probabilidad y Estadística.
Risas
Mónica: Pues sí, es que… datos.
Blanca: Datos ¿Variable datos?
-- 03:03.3
ByM010 (video) -- 00:00.0
Mónica: Inconstante.
Blanca: ¿Por qué inconstante?
Mónica: No, impredecible.
Brenda: No, la impredecible es y.
Mónica: No, pero también es x.
Blanca: ¿Por qué esta también es impredecible?
Mónica: No, inexacta… ¡Porque! Por… porque a lo mejor podemos quitar el dos, a lo mejor no hay familias con dos hijos…. Incontrolable, no.
Profesor: ¿De dónde provendría este valor de las x?
Brenda: ¿De la x? En este caso es el número de hijos
Mónica: Pues del número de trabajadores.
Profesor: Pero ya sabemos el número de trabajadores, ya sabemos el número de hijos. Pero se acuerdan que se trataba de hacer un sorteo y el sorteo tenía una tómbola y de la tómbola sale un trabajador seleccionado con cierto número de hijos. Se hacía un experimento. Bueno si eso es así, ¿dónde estaría esa x, en los trabajadores, en la tómbola, en los boletos? ¿dónde quedó la x?
Mónica: En los boletos. No, en lo que hay adentro de la tómbola, en las bolitas.
Brenda: No, porque adentro de la tómbola están los trabajadores. Es que se supone que del número de hijos… por decir…¿cómo lo explicaré?
Mónica: O sea no está el número de hijos, de que uno, dos, sino que está un trabajador y tú, a ver cuántos hijos tienes. No pues que dos. Está implícito adentro del trabajador.
Brenda: Sí, ándale. O sea no hay papelitos que digan uno, dos, tres, …
Mónica: No, sino que sale, Juan Ramírez y tú, a ver, cuántos hijos tienes, no pues que tres, corresponde a… (*señala el eje de las ordenadas en la gráfica*) y ya.
Brenda: O sea que dentro de los papelitos que están en la tómbola, están las posibilidades o sea estos (*señala la escala del eje de las abscisas*).
Mónica: …están los trabajadores y el número de hijos es una característica de ellos.
Profesor: Entonces esa ‘x’ es obtenida de la tómbola, ¿es cierto eso?
Mónica: Pero no directamente ¿verdad? Indirectamente, sí. Hay que juntarlo con un trabajador.
Profesor: Y entonces ahí sale, y si en el primer boleto salió un trabajador, vamos a suponer, con dos hijos, ¿se puede saber en el siguiente boleto el número de hijos que va a salir?
Mónica: No. Ni siquiera podemos saber qué trabajador va a salir.
Profesor: Entonces ¿hay orden de aparición en esos números?
Mónica: No.
Profesor: Entonces ¿por qué los pusieron así seguidos, uno, dos, tres, cuatro,…?
Brenda: Pues porque es un orden y así va la numeración, pero orden pues así como la mayoría...

Mónica: Pudimos haberlos ordenado, primero tres, luego cuatro y así, y luego iba a salir una gráfica así (*dibuja en el aire una gráfica decreciente*) porque van por orden en que salieron. Cómo se esperaría que salieran porque se espera que salgan muchas veces el tres, pero no sale, igual sale más veces otro, pero se espera que salga más veces el tres.
Profesor: ¿Es posible, de veras, que a la mera hora, supongamos que se sacan varios boletos, resulte que salen varios nueves?
Mónica: Sí, es poco probable, pero sí es posible. Es como lo de los boletos de las casas, a lo mejor no tiene nada que ver, pero por ejemplo de que se va a rifar una botella y una sola persona compra muchos boletos y uno dice, pues se espera que gane, pero puede no ganar.
Brenda: Sí.
Blanca: Entonces, ¿cómo llamarían a esa variable?
Mónica: Suerte.
--- 05:43.4
ByM11 (video)--- 00:00.0
Brenda: La implícita.
Mónica: Sí.
Brenda: Porque está dentro de la tómbola. Porque está dentro de la tómbola, lo que estábamos explicando ahorita, no está dentro de los papelitos. O ¿qué otro nombre le ponemos?
Mónica: Implícita determinada. Implícita por lo que ya quedamos y determinada porque está determinada por el trabajador. Está implícita por el trabajador y determinada por el trabajador. ¿Te parece buen nombre?
Brenda: Yo creo que sí.
--- 01:42.6

Anexo 4.

Ejemplos de producciones de los alumnos.
Estudio 4: Evaluación de un proyecto.

Ejemplo 1: MJDG

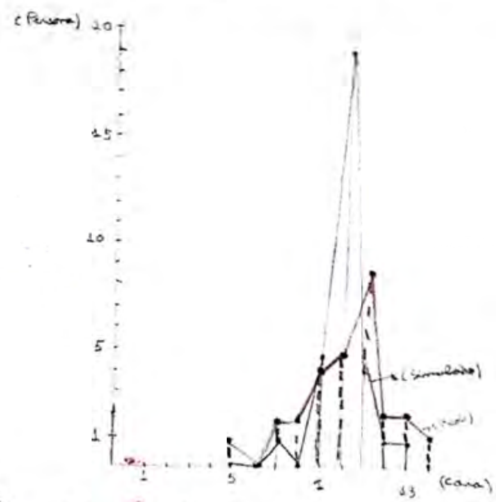
Práctica Nº 4: Evaluación de experiencias de enseñanza (PARTE A: Desarrollo del proyecto)

RESOLUCIÓN: (Utilizar esta hoja por delante y por detrás, añadiendo otras si fuera necesario)

1. Analizar los datos obtenidos en la clase durante el experimento realizado para evaluar las concepciones sobre la aleatoriedad.
2. Preparar un informe sobre el análisis, incluyendo todas las tablas y gráficos estadísticos elaborados
3. Resumir las conclusiones obtenidas del estudio anterior acerca de las intuiciones de la clase sobre experimentos aleatorios.
4. Elaborar una tabla que resuma en una columna los contenidos sobre tratamiento de datos y estadística que has trabajado durante el proyecto y en otra los que se describen en el Decreto de Enseñanzas Mínimas del MEC (tanto en contenidos, como en criterios de evaluación)
5. Estudiar las páginas 411-423 del Tema 6: Estocástica (Capítulo 1. Estadística) del texto recomendado.

1.

Caras	Persona S	Persona R
5	–	1
6	–	–
7	1	2
8	0	2
9	5	4
10	19	5
11	4	8
12	1	2
13	1	2
14	–	1



En esta gráfica hay que comentar que:

- tanto en el lanzamiento simulado como el real tienen unos valores centrales.
- La simulada tiene una tendencia similar a la real.
- En la simulada 19 personas han puesto diez caras y 11 personas han puesto 8 caras en la real. Son los máximos de estos dos tipos de lanzamientos.
- La simulada es más intuitiva que la real. En general, la clase tiene una buena intuición.
- El valor central de los dos lanzamientos es diez.

MJDG, página 2

2.

Lanzamiento Simulado:

C C X C X X C C X C X X C X X C C X C C

$C \to 11$ $X \to 9$ Rachas $\to 13$

Lanzamiento real:

X X C X C C C X C C C C X C C X C X X X

$C \to 11$ $X \to 9$ Rachas $\to 11$

Caras	Persona S	Persona R
5	–	1
6	–	–
7	1	2
8	0	2
9	5	4
10	19	5
11	4	8
12	1	2
13	1	2
14	–	1

Lanzamiento de una moneda

Simulada (Caras)

x_i	f_i	$x_i \cdot f_i$	$(x_i - \bar{x})^2 \cdot f_i$
9	5	45	7'2
10	16	160	0'64
11	7	77	4'48
12	1	12	3'24
13	1	13	7'84
	N=30	Σ=307	Σ=23'4

- Moda → 10
- Media → $\bar{x} = \frac{\Sigma 370}{30} = 10'2$

  $\bar{x} = \frac{\Sigma x_i f_i}{N}$
- Mediana → 10
- Varianza → varianza $= \frac{\Sigma (x_i - \bar{x})^2 \cdot f_i}{N} = \frac{23'4}{30} = 0'78$
- Desviación típica → $\sigma = \sqrt{0'78} = 0'$
- Rango → $13 - 9 = 4$

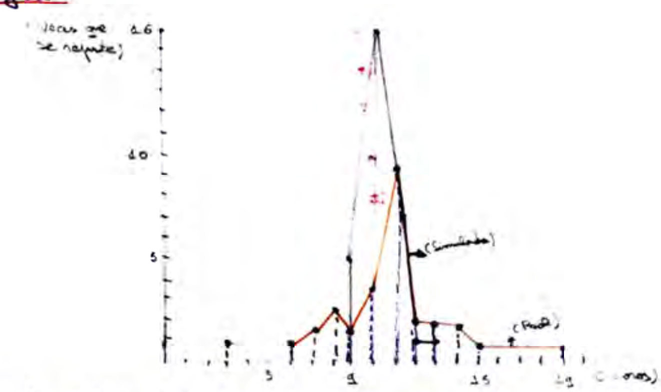
MJDG, página 3

Real (Caras)

x_i	f_i	$x_i \cdot f_i$	$(x_i - \bar{x})^2 \cdot f_i$
3	1	3	56'25
6	1	6	20'25
7	2	14	24'5
8	3	24	18'75
9	2	18	4'5
10	4	40	1
11	9	99	2'25
12	2	24	4'5
13	2	26	12'5
14	2	28	24'5
15	1	15	20'25
19	1	19	72'25
	N=30	Σ=316	Σ=261'5

- Moda → 11
- Media → $\bar{x} = \frac{\Sigma x_i \cdot f_i}{N} = \frac{316}{30} = 10$
- Mediana → 11
- Varianza → varianza $= \frac{\Sigma (x_i - \bar{x})^2 \cdot f_i}{N}$ $= \frac{261'5}{30} = 8'71$
- Desviación típica → $\sigma = \sqrt{8'71} =$
- Rango → 19-3 = 16

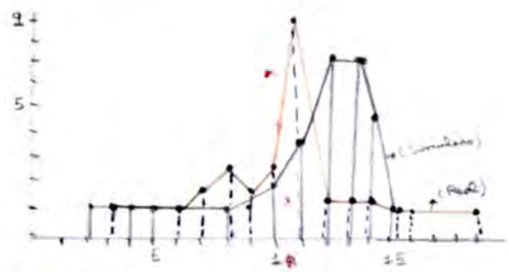
Gráfica



En esta gráfica los valores se concentran en el medio. La media es el diez y se v
que en la simulada el número de caras que se repite es mayor que la real. Ést
tiene más altibajos; sin embargo, la simulada tiene un único pico.

MJDG, página 4

Gráfica (Rachas)



- En esta gráfica la si-
lada está mayormente
buida que la real. Los va
no se concentran tanto
en el anterior caso. Aho
la real destaca sobre la
simulada en cuanto a l
rachas. La media sería
ximadamente once.

Simulada (Rachas)

x_i	f_i	$x_i \cdot f_i$	$(x_i - \bar{x})^2 \cdot f_i$
2	1	2	81
4	1	4	49
5	1	5	36
8	1	8	9
10	2	20	2
11	4	44	0
12	7	84	7
13	7	91	28
14	5	70	245
15	1	15	16
	$\Sigma = 30$ → N	$\Sigma = 343$	$\Sigma = 473$

✱ Moda → 12 y 13

✱ Media → $\bar{x} = \frac{\Sigma x_i \cdot f_i}{N} = \frac{343}{30}$

✱ Mediana → 12

✱ Varianza → varianza $= \frac{\Sigma (x_i - \bar{x})^2 \cdot f_i}{N}$ $= \frac{473}{30} = 15'7$

✱ Desviación típica → $\sigma = \sqrt{15}$ $= 3'96$

✱ Rango → $15 - 2 = 13$

Real (Rachas)

x_i	f_i	$x_i \cdot f_i$	$(x_i - \bar{x})^2 \cdot f_i$
7	1	7	9
8	3	24	12
9	3	27	3
10	8	80	0
11	6	66	6
12	5	60	20
13	1	13	9
14	1	14	16
15	1	15	25
16	1	16	36
	N=30	$\Sigma = 322$	$\Sigma = 136$

✱ Moda → 10

✱ Media → $\bar{x} = \frac{322}{30} = 10'7$

✱ Mediana →

✱ Varianza → varianza $= \frac{136}{30} = 4'5$

✱ Desviación típica → $\sigma = \sqrt{4'5} = 2$

✱ Rango → $16 - 7 = 9$

3.

En general la clase tiene una buena intuición. El lanzamiento real se ajusta bastante a la simulada aunque halla alguna excepción (tanto en cara como rachas).

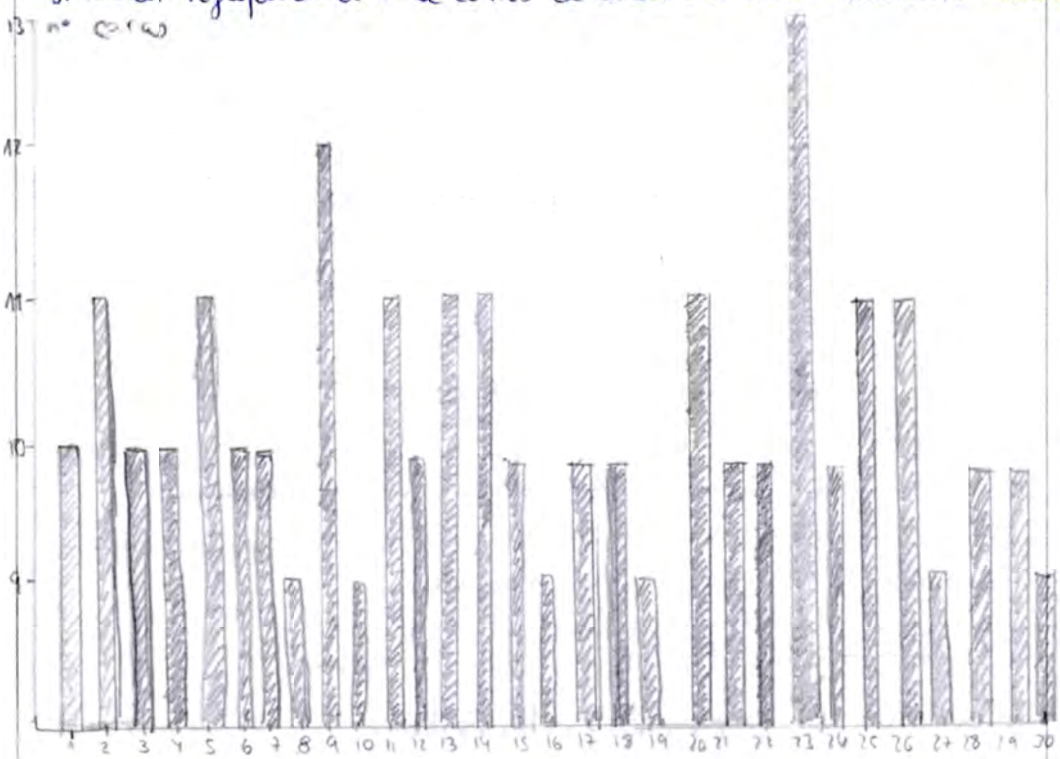
EJEMPLO 2: FJS

RESOLUCIÓN: (Utilizar esta hoja por delante y por detrás, añadiendo otras si fuera necesario)

1. Analizar los datos obtenidos en la clase durante el experimento realizado para evaluar las concepciones sobre la aleatoriedad.
2. Preparar un informe sobre el análisis, incluyendo todas las tablas y gráficos estadísticos elaborados
3. Resumir las conclusiones obtenidas del estudio anterior acerca de las intuiciones de la clase sobre experimentos aleatorios.
4. Elaborar una tabla que resuma en una columna los contenidos sobre tratamiento de datos y estadística que has trabajado durante el proyecto y en otra los que se describen en el Decreto de Enseñanzas Mínimas del MEC (tanto en contenidos, como en criterios de evaluación)
5. Estudiar las páginas 411-423 del Tema 6: Estocástica (Capítulo 1. Estadística) del texto recomendado.

1. Se presenta en la hoja de recogida de datos, en lanzamientos de una moneda (Final de la práctica)

2. ~~Tabla~~ ~~que~~ Gráfico estadístico (Diagrama de barras) de la secuencia simulada reflejando el nº de caras de cada uno de los alumnos. Tabla



FJS, página 2

Gráfica estadística (diagrama de barras) representando el nº de caras de la secuencia real de cada alumno de la clase. Tabla 2.

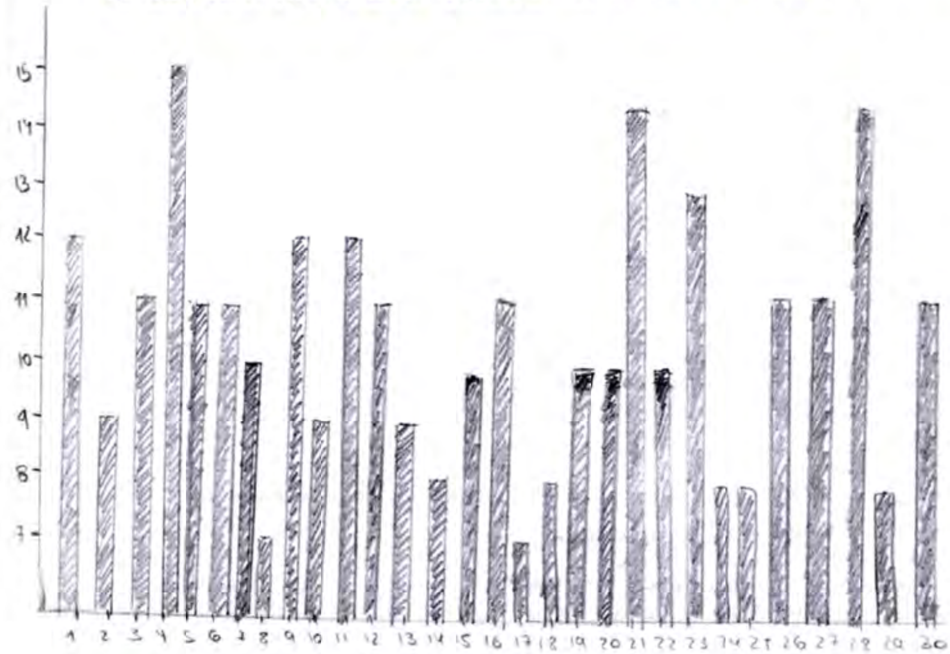


Tabla 1

Media = Sumamos todos los datos que intervienen y dividimos entre el total de los mismos.

$$\frac{307}{30} = \boxed{10'2}$$

Mediana. Dato que ocupa el lugar central en una ordenación, podemos apreciar que el nº 15 es el que ocupa el centro y, pero como el nº de datos es par tomamos como referencia los dos datos que ocupan la posición central y obtenemos la media, por tanto resultaría la mediana:

$\frac{15+16}{2} = \boxed{15'5}$; 15 = 10, 16 = 9 $\frac{10+9}{2} = \boxed{9'5}$

Moda = Es el valor que más veces se repite

Nº veces 9 = 6
Nº veces 10 = 14
Nº veces 11 = 8
Nº veces 12 = 1
Nº veces 13 = 1

Podemos deducir que el 10 es el nº que más se repite

FJS, página 3

Tabla 2

Media = 311 / 30 = $\boxed{10'3}$

Mediana = 15 y 16 → valores centrales

10 11 → Media = $\frac{10+11}{2}$ = $\boxed{10'5}$

Moda = Nº veces 7 = 2
Nº veces 8 = 5
Nº veces 9 = 3
Nº veces 10 = 5
Nº veces 11 = 8
Nº veces 12 = 3
Nº veces 13 = 1
Nº veces 14 = 2
Nº veces 15 = 1

El nº que más se repite $\boxed{11}$

Tabla que recoge el nº rachas de la secuencia simulada y real

Racha Simulada		Racha Real
	16	X
XX	15	
XXXXXX	14	X
XXXXXX	13	X
XXXXX XXXX	12	XXXXX
XXXX	11	XXXXXX
X	10	XXXXXXXX
	9	XXXX
X	8	XXX
	7	X
X	3	

FJS, página 4

Secuencia Simulada → Racha mayor

Para mejor claridad de los datos vamos a calcular en % (porcentajes)

Rachas: 13 → 1 ; $\left.\begin{array}{l} 30 — 100 \\ 1 — x \end{array}\right\} x = \boxed{3'3\%}$ → $\boxed{13}$

4 → 1 ; $\boxed{3'3\%}$ → $\boxed{4}$

3 → 24 $\left.\begin{array}{l} 30 — 100 \\ 24 — x \end{array}\right\} x = \boxed{80\%}$ → $\boxed{3}$

2 → 3 $\left.\begin{array}{l} 30 — 100 \\ 3 — x \end{array}\right\} x = \boxed{6'\widehat{6}\%}$ → $\boxed{2}$

1 → 1 ; $\boxed{3'\widehat{3}\%}$ → $\boxed{1}$

Secuencia ~~Simulada~~ Real → Racha Mayor

Rachas:

9 → 1 ; $\boxed{3'\widehat{3}\%}$ → $\boxed{9}$

8 → 1 ; $\boxed{3'\widehat{3}\%}$ → $\boxed{8}$

7 → 1 ; $\boxed{3'\widehat{3}\%}$ → $\boxed{7}$

6 → 4 ; $\left.\begin{array}{l} 30 — 100 \\ 4 — x \end{array}\right\} x = \boxed{13'3\%}$ → $\boxed{6}$

FJS, página 5

$5 \to 4$, $\boxed{13'3\%} \to \boxed{5}$

$4 \to 12$; $\left.\begin{array}{l} 30 - 100 \\ 12 - x \end{array}\right\}$ $x = \boxed{40\%} \to \boxed{4}$

$3 \to 6$; $\left.\begin{array}{l} 30 - 100 \\ 6 - x \end{array}\right\}$ $x = \boxed{20\%} \to \boxed{3}$

$2 \to 1$; $\boxed{3'3\%} \to \boxed{2}$

3.

Con respecto del nº de caras de la secuencia simulada y la real cabe decir que la media en ambos es 10 nº de caras, con lo que podemos afirmar que en el nº de caras existen buenas intuiciones para cada uno de los alumnos.

En la secuencia simulada el nº de caras que más se repite es 10, sin embargo en la secuencia real son 11.

Finalmente decir que en lo que respecta al nº de caras en una secuencia simulada a una real existe buena intuición por lo que no hay grandes diferencias en este punto de cantidades.

Por otro lado decir que por el contrario el nº de rachas en la secuencia simulada y real no existe intuición alguna se da el caso de un alumno tener una racha de 3 frente a otro de 16.

FJS, página 6

Por último decir que casi todos los alumnos han
coincidido el el nº de racha mayo con una media de 3
rachas mayor pues el 80% es un número elevado, pero
no existe buena intuición con relación del nº de rachas mayores
reales.