Universidad de Granada
Facultad de Ciencias

VARIABLE BINOMIAL METRICA

JORGE OLLERO HINOJOSA

Granada, Junio 1984

R. 29.526

BIBLIOTECA
FACULTAD DE CIENCIAS
6
3
124

Tesis de Doctorado en Ciencias, Sección de Matemáticas, juzgada el dia 10 de Noviembre de 1984 en la Facultad de Ciencias de la Universidad de Granada y calificada con Sobresaliente Cum Laude por el siguiente Tribunal:

Presidente: Dr.D. Ramón Gutierrez Jaimez, Catedrático de la Universidad de Granada.

Vocales : Dr.D. Antonio Pascual Acosta, Catedrático de la Universidad de Sevilla.

Dr.D. Antonio Martín Andrés, Catedrático de la Universidad de Granada.

Dr.D. Luis Parras Guijosa, Catedrático de la Universidad de Málaga.

Secretario: Dr.D. Elias Moreno Bas, Profesor Adjunto de la Universidad de Granada.

VARIABLE BINOMIAL METRICA

R. 29.526

VARIABLE BINOMIAL METRICA

UNIVERSIDAD DE GRANADA
FACULTAD DE CIENCIAS
SALIDA R.º 2-076

BIBLIOTECA UNIVERSITARIA
GRANADA
Nº Documento 6135768 47
Nº Copia 115587514

Memoria que para optar al grado de Doctor en Ciencias, Sección de Matemáticas, presenta el licenciado Jorge Ollero Hinojosa.

Director de Tesis:
Prof. Dr. D. Ramón Gutierrez Jáimez

Vº Bº

UNIVERSIDAD DE GRANADA. FACULTAD DE CIENCIAS.

Deseo expresar mi gratitud, en primer lugar, al profesor D. Ramón Gutierrez Jáimez en su calidad de Director de la memoria y Director del Departamento de Estadistica Matemática. En segundo lugar, a D. Alfonso Guiraum Martín y D. Ramiro Melendreras Gimeno, también directores del Departamento durante la realización de la memoria. Finalmente, a todos los que han colaborado, de una u otra forma, para llegar a su fin.

Granada, Junio de 1984 .

A todos cuantos me animaron durante estos años y, en especial, a la memoria de los que no pudieron verla terminada.

INDICE

INTRODUCCION GENERAL XVII

CAPITULO I

DISTRIBUCION BINOMIAL

I.1 Introducción 5

I.2 Definición 7

I.3 Pertenencia a Familias Discretas de Probabilidad 9

I.4 Funciones Generatrices 15

I.5 Momentos y Cumulantes. Relaciones de Recurencia . 16

a) Momentos 16

b) Cumulantes 20

I.6 Otros Tipos de Momentos 21

a) Desviación Media 21

b) Momentos Inversos Factoriales 22

c) Momentos Inversos de la Binomial Positiva . . 23

d) Diferencias Medias Generalizadas 26

I.7 Principales Propiedades 28

a) Localización de la Moda 29

b) Teorema de Moivre-Laplace 29

c) Suma de Binomiales 30

d) Diferencia de Binomiales 31
e) Una Distribución Condicionada 32

I.8 Caracterizaciones de la Ley Binomial 33
a) Caracterización de Shanbhag-Basawa 34
b) Caracterización de Janardan 35
c) Caracterización de Lukacs 36

I.9 Estimación de los Parámetros 36
a) Estimación de p 37
b) Estimación de n 47
c) Estimación Conjunta de n y p 50

I.10 Distribuciones Relacionadas 59
a) Distribución Beta 59
b) Distribución F 61
c) Distribución Binomial Negativa 62
d) Distribución de Poisson 63
e) Distribución Hipergeométrica 66

I.11 Aproximaciones a las Probabilidades Binomiales . 70
A) Aproximaciones Normales 81
A.1 Aproximación de Moivre-Laplace 81
A.2 Aproximación de Gram-Charlier 85
A.3 Transformación Raiz Cuadrada 88

A.4 Aproximación de Camp-Paulson 89
A.5 Aproximación de Borges 93
A.6 Aproximación de Peizer-Pratt 97
B) Aproximaciones por la Distribución de Poisson . 102
B.1 Aproximación de Poisson 104
B.2 Aproximación de Gram-Charlier 105
B.3 Aproximación de Bol'shev 106
B.4 Aproximación de Molenaar 108
C) Otros Métodos de Aproximación 108
C.1 Aproximación por la Transformación Arcoseno . 108
C.2 Aproximación de Bahadur 122
C.3 Aproximaciones por Fracciones Continuas . . 129

I.12 Generalización de Poisson de la Distribución Binomial 139
Principales Propiedades y Resultados 145
A) Extremos y Comportamiento Monótono de $f(\bar{p})$. 146
B) Forma de la Distribución 161
C) Localización de la Moda y la Mediana . . . 165
D) Generalización del Teorema de Bernoulli . . 170
E) Convergencia hacia la Distribución de Poisson 172

I.13 Otras Propiedades 180
a) Teorema de Bernoulli 180

b) Relaciones entre las Colas de la Binomial, Normal y Poisson 182

c) Acotaciones basadas en la Aproximación Normal . 187

d) Monotonía de las Probabilidades Binomiales . . 191

CAPITULO II

INTRODUCCION INTUITIVA A LAS VARIABLES BINOMIALES METRICAS

II.1 Introducción 199

II.2 Definiciones Auxiliares 200

a) Fuente de Memoria Nula 200

b) Canal de Información 201

c) Canal Binario Simétrico 202

d) Extensión del Canal Binario Simétrico 203

II.3 Introducción a las Variables Binomiales Métricas 206

a) Distancia de Hamming 207

b) Variable Número de Errores de Transmisión . . 209

c) Variables Binomiales Métricas 210

d) Comentario 214

II.4 Modelos que originan Variables Binomiales Metricas 217

A) Recepción de una Señal Desconocida bajo Ruido . 218

B) Estimación de una Fuente Aleatoria desde una Posición Aleatoria 223

C) Fuente Aleatoria Cifrada por un Desplazamiento Desconocido 228

CAPITULO III

VARIABLES BINOMIALES METRICAS

III.1 Definición 237

III.2 Función de Masa de Probabilidad 243

III.3 Funciones Generatrices 247

III.4 Momentos y Cumulantes 251

a) Momentos 251

b) Cumulantes 260

III.5 Otros Tipos de Momentos 262

a) Desviación Media 263

b) Momentos Inversos Factoriales 269

c) Momentos Negativos de la Variable Binomial Metrica Positiva 279

d) Diferencias Medias Generalizadas 284

III.6 Relaciones entre las Variables Binomiales Metricas 287

III.7 Suma y Diferencia de Variables Binomiales Métricas. Caracterización de la Ley 291

III.8 Principales Propiedades 295

a) Forma y Clasificación de la Distribución . . . 295

b) Localización de la Moda y Mediana 297

c) Relaciones entre las Esperanzas de Funciones de las VBM con valores de λ y n comunes . . . 301

d) Generalizaciones del Teorema de Jogdeo y Samuels 306

e) Convergencia hacia la Ley de Poisson Desplazada 311

III.9 Entropía y Energía Informacional de las Variables Binomiales Generalizadas 315

a) Entropía 316

b) Variable Binomial Generalizada Métrica . . . 326

c) Energía Informacional 327

III.10 Estimación de los Parámetros 341

a) Estimación de p 341

b) Estimación de d 352

c) Estimación de n 355

III.11 Aproximación de las Probabilidades Binomiales Métricas 361

a) Aproximaciones Normales 362

b) Aproximaciones por la Distribución de Poisson . 368

APENDICES

I. Probabilidades Puntuales de la Binomial Métrica . 379

II Estimador de Máxima Verosimilitud de p en B(2,1,p) 389

III Comportamiento de varios Estimadores del parámetro. 393

IV Estimación del Parámetro d en la B(n,d,p) . . . 405

V Comportamiento de varios Estimadores de n . . . 415

VI Comparación entre algunos Métodos Normales de Aproximación de las Probabilidades Binomiales Métricas 429

VII Comparación entre Aproximantes Construidos a partir de la Ley de Poisson para las Probabilidades Binomiales Métricas 437

BIBLIOGRAFIA

Bibliografía Consultada 445

Bibliografía no Consultada 459

INTRODUCCION GENERAL

El objetivo de la memoria es presentar una nueva generalización de la variable binomial de Bernoulli y realizar un estudio detallado de la variable resultante. La variable binomial métrica incorpora un nuevo parámetro que hemos denominado distancia y que puede tomar cualquier valor entero comprendido entre 0 y el número de pruebas n. Esta variable se reduce a la binomial clásica unicamente cuando d coincide con dichos valores o bien la probabilidad de "éxito" p es igual a $\frac{1}{2}$.

Una variable binomial métrica B(n,d,p) puede definirse simplemente como la variable aleatoria correspondiente a la suma de las binomiales independientes B(n-d,p) y B(d,q), lo que significa que nuestra generalización queda englobada en la generalización que realizó Poisson de la

binomial al eliminar la exigencia de que todos los ensayos que intervienen fuesen equiprobables. No obstante, nuestra generalización es fruto de un principio mas general, no solo aplicable a la binomial de Bernoulli, sino a otras distribuciones entre las que podemos incluir a la propia binomial de Poisson. En este sentido, podemos indicar que en la distribución χ^2 no centrada subyace el mismo principio.

La memoria está dividida en tres capítulos y varios apéndices. El primer capítulo está dedicado a estudiar en profundidad la distribución binomial y la generalización de Poisson. Su objetivo es recopilar aquellos resultados poco conocidos o bien aquellos otros que son de reciente publicación y servir de modelo para desarrollar el capítulo tercero.

A excepción de la sección I.12, todas las restantes están dedicadas a la binomial de Bernoulli, destacando las secciones I.9 y I.11 en las que se trata, respectivamente, de la estimación de los parámetros y de la aproximación de las probabilidades. De la primera de estas secciones cabe destacar los recientes artículos sobre la estimación por intervalos de p y sobre la estimación de n, cuando p es conocido o desconocido. En la otra sección se ana-

lizan 13 tipos diferentes de aproximantes para las probabilidades binomiales.

La sección I.12 está dedicada a la binomial generalizada y en ella recogemos practicamente la totalidad de la escasa investigación realizada sobre esta variable aleatoria. Como ya dijimos, la binomial de Poisson surge al suprimir del modelo de Bernoulli la hipótesis de que todas las pruebas sean equiprobables. Al hacer esto, resulta que una variable que siga el nuevo modelo quedará determinada cuando se conozcan las probabilidades de "exito" de cada una de las pruebas que la definen; todas estas probabilidades constituyen el vector de parámetros de la distribución. En el caso particular de la binomial métrica, dicho vector está compuesto por a lo sumo de dos tipos de componentes, siendo las de un tipo las complementarias a la unidad de las otras.

La investigación sobre la binomial generalizada ha versado sobre las siguientes cuestiones: a) Comportamiento de la esperanza matemática de cualquier función de la variable cuando su esperanza se mantiene constante; b) Forma de la distribución; c) Localización de la moda y de la mediana; d) Convergencia hacia la distribución de Poisson; e) Generalización de los teoremas de Bernoulli y de Moivre-La-

place.

Aparte de estas cuestiones, se han considerado algunas aplicaciones practicas de dicha distribución, que no hemos recogido en la presente memoria. A excepción de la última sección para la que ya Poisson generalizó el teorema de Bernoulli, sobre las restantes los resultados han ido apareciendo desde 1956 en adelante.

Respecto a la primera cuestión, W. Hoeffding (1956) encontró sobre qué tipo de vectores paramétricos se localizan los extremos de la esperanza matemática de una función arbitraria de la variable; dichos vectores corresponden esencialmente a variables binomiales de Bernoulli; esto es completamente cierto cuando se trata de la función de distribución y de los momentos absolutos de orden superior a la unidad. Dado que el valor esperado de la binomial generalizada es la suma de las componentes del vector paramétrico, las siguientes aportaciones se produjeron cuando se consideró sobre el espacio paramétrico el orden parcial por mayoración. De este modo, L. Gleser (1975) logró demostrar que si los vectores de parámetros de dos variables eran comparables, entonces se podría saber cuál de las dos esperanzas de un determinado tipo de funciones sería mayor. En ese tipo de fun

ciones se incluye a la función de distribución valuada en cualquier punto, con la excepción de tres situados alrededor de la media.

Los resultados sobre la forma de la distribución se concretan en que ésta es acampanada para cualquier orden y suponen que la distribución se clasifica en la familia de las distribuciones fuertemente unimodales. Dichos resultados se deben a J. Darroch (1964) y a S. Samuels (1965), quienes asimismo consideraron la localización de la moda. Esta se localiza en alguno de los enteros adyacentes a la media y en determinadas condiciones queda completamente determinada por la media o bien por ésta y la mayor o la menor de las probabilidades de "exito". Los resultados relativos a la localización de la mediana son similares y fueron obtenidos por K. Jogdeo y S. Samuels en 1968.

Respecto a la convergencia hacia la ley de Poisson tenemos los resultados de J. Hodges y L. Le Cam (1960), de Le Cam (1960) y de L. Chen (1974). Los dos primeros trabajos contienen teoremas de aproximación, siendo los teoremas del segundo trabajo mucho mas fuertes. En el tercer trabajo se generaliza el teorema límite que G. Simons y N. Johnson dieron para la binomial clásica y que se recoge en la sec-

ción I.10, página 65.

Sobre todas estas cuestiones hemos realizado alguna aportación, ya sea en el caso especial de la binomial métrica ya sea en el caso general de la binomial de Poisson. Aparte de esas aportaciones, hemos añadido una nueva cuestión sobre la binomial de Poisson: la descripción del comportamiento de la entropía y la energía informacional. Al considerar la última medida de aleatoriedad hemos introducido el concepto de variable binomial generalizada métrica. Todos nuestros resultados en estas lineas se encuentran recogidos en las secciones 8 y 9 del capítulo tercero.

Del contenido de las restantes secciones del primer capítulo podemos destacar, además del teorema de Simons y Johnson, el concepto de diferencias medias generalizadas (pag. 26); la caracterización de Shanbhag-Basawa (pag. 34) y el teorema de Jogdeo y Samuels (pag. 192). Ambos teoremas se generalizan en varias formas en el tercer capítulo. Respecto a las diferencias medias generalizadas, se demuestra que son invariantes por extensión métrica y se utilizan para hallar la energía informacional de la variable binomial de Poisson.

El capítulo segundo está dedicado a dar una introducción intuitiva de la noción de variable binomial métrica (VBM). Dicha introducción se realiza desde el punto de vista de la Teoría de la Información, proponiendose algunos modelos y ejemplos ideales en los que intervienen nuestras variables aleatorias. Este capítulo también contiene la génesis de esta nueva noción.

En el capítulo tercero se realiza el estudio a fondo de la VBM siguiendo un desarrollo paralelo al del primer capítulo hasta que se alcanza una profundidad similar a la de cada una de aquellas secciones. De entre nuestros resultados vamos a mencionar los más importantes:

a) La invarianza de las diferencias medias generalizadas respecto al parámetro d (pag. 284). Esta propiedad es extensible a la binomial generalizada métrica y de ella se desprende la invarianza de la energía informacional respecto a la extensión métrica, así como el valor explicito de esta medida de aleatoriedad (Vease la sección III.9).

b) El decrecimiento respecto de d de la función de distribución de la VBM valuada en cualquier punto (pag. 290 y 291) y el no-decrecimiento de la moda y de la mediana

respecto al mismo parámetro (pag. 298).

c) La extensión de la caracterización de Shanbhag y Basawa tanto para la suma como para la diferencia de VBM independientes (pags. 292-294). A su vez, estos resultados son generalizables sin dificultad para la binomial de Poisson.

d) La demostración de la comparabilidad por mayoración entre los vectores de probabilidad de "exito" correspondientes a VBM con el mismo valor esperado (pag. 301). Como consecuencia de ello son aplicables a nuestra distribución los resultados de Hoeffding y Gleser. Por otra parte, en condiciones especiales logramos extender los resultados de Gleser de manera que incluyan completamente a los de Hoeffding (pag. 305).

e) La reformulación del teorema de Jogdeo y Samuels para la binomial de Bernoulli de modo que englobe completamente al resultado similar de Anderson y Samuels (vease I.13.b) (pag. 307) y la extensión del primer teorema a la VBM (pag. 308 y 309).

f) La generalización del teorema de Simons y Johnson, probandose la convergencia de la VBM hacia la distribución de Poisson desplazada (pags. 311-313).

g) La demostración de la Schur-concavidad respecto al vector paramétrico de la entropía de la variable binomial generalizada (pag. 317) y la localización de su máximo absoluto para un número de pruebas de Poisson dado (pag. 318).

h) La demostración para la entropía del análogo del teorema de Jogdeo y Samuels (pag. 323). Como corolario del mismo se obtiene que la entropía de una distribución de Poisson es siempre mayor que la correspondiente a cualquier binomial que tenga el mismo promedio.

i) La definición de la extensión métrica de la variable binomial de Poisson y la demostración de que su energía informacional permanece invariante ante tal extensión (pags. 326-327) .

j) La localización de los extremos de la energía informacional de la binomial generalizada (pag. 331) y la demostración de la Schur-concavidad de la misma sobre un subconjunto del espacio paramétrico (pag. 334).

k) La demostración de que las diferencias medias generalizadas presentan el mismo comportamiento respecto al vector de parámetros que los momentos absolutos de las variables binomiales de Poisson (pag. 337).

l) Una aportación (pag. 340) en la linea de las recientes ideas de R. Hickey (1983) sobre la utilización de la mayorización de vectores para comparar la aleatoriedad de dos distribuciones al margen del uso de cualquier medida (vease pag. 325). A pesar de lo muy limitado de nuestro primer resultado, estas ideas nos han abierto una nueva linea de trabajo en la que hemos obtenido ya algunos resultados importantes, no incluidos en la presente memoria.

m) La demostración de la inexistencia de estimadores insesgados para el parámetro p cuando d es suficientemente grande (pag. 344) y la pérdida de la suficiencia de la suma de las observaciones muestrales al pasar de la binomial de Bernoulli a la binomial métrica (pags. 343 y 344).

n) La finitud del estimador de máxima verosimilitud del parámetro n (pag. 355) y la extensión completa a la VBM de los procedimientos aproximados de estimación de dicho parámetro en la binomial clásica basados en la distribución normal.

ñ) El tratamiento unificado de todos los métodos basados en la distribución normal para la aproximación de las

probabilidades binomiales métricas y la comprobación numérica de que los métodos más importantes alcanzan una precisión similar a la que alcanzan para la binomial clásica.

o) La adecuada construcción de los aproximantes de las probabilidades basados en la ley de Poisson y la comprobación numérica de que presentan un comportamiento uniformemente igual para cualquier valor del parámetro d.

En los apéndices se recoge información numérica de tres tipos diferentes. En primer lugar, en el apéndice I damos una tabla de las probabilidades puntuales para los primeros valores de n. En segundo lugar, los cuatro siguientes apéndices contienen resultados sobre el comportamiento práctico de algunos estimadores de cada uno de los parámetros. Y en tercer lugar, los apéndices VI y VII contienen los errores que se cometen al emplear distintos métodos para aproximar las probabilidades binomiales métricas.

Finalmente vamos a enumerar algunas de las posibles lineas que quedan abiertas para futuros trabajos. Aparte de la basada en las ideas de Hickey, se podría continuar

investigando sobre la estimación de los parámetros y sobre los métodos de aproximación de las probabilidades de la distribución. Estos temas han sido objeto en la memoria de un tratamiento preferentemente empírico, siendo por tanto susceptibles de un enfoque mucho más riguroso. Ademas en el caso de la estimación de los parámetros hemos dejado sin tratar muchas cuestiones.

Así, por ejemplo, se podrían iniciar estudios sobre la aplicación de otros métodos de estimación como son: los métodos bayesianos, los estimadores de mínima distancia, los métodos robustos. También podría considerarse la estimación por intervalos, el contraste de hipótesis y la estimación conjunta de los parámetros, así como el estudio de todas las cuestiones relacionadas.

En el caso de los métodos de aproximación de las probabilidades, se podría intentar fundamentar matemáticamente los resultados empíricos obtenidos. Por otra parte, también se podría ensayar la mejora de los procedimientos normales; en particular, de la extensión de los métodos de Camp-Paulson y de Peizer-Pratt.

Finalmente queda la posibilidad de acometer la extensión métrica de otras distribuciones, tanto univariantes como multivariantes, realizando un estudio semejante al

de esta memoria. En este sentido, recordamos que en la sección III.9 se considera ya la variable binomial de Poisson métrica. Evidentemente, en principio, las distribuciones más apropiadas para llevar a cabo la extensión métrica son las distribuciones relacionadas con el modelo binomial; esto es: la hipergeométrica, la binomial negativa y la multinomial.

CAPITULO I

DISTRIBUCION BINOMIAL

I.1 Introducción

I.2 Definición

I.3 Pertenencia a familias discretas de probabilidad

I.4 Funciones Generatrices

I.5 Momentos y Cumulantes. Relaciones de recurrencia

I.6 Otros tipos de Momentos

I.7 Principales Propiedades

I.8 Caracterizaciones de la Ley Binomial

I.9 Estimación de los Parámetros

I.1o Distribuciones relacionadas

I.11 Aproximaciones a las Probabilidades Binomiales

I.12 Generalización de Poisson de la Ley Binomial

I.13 Otras Propiedades

I.1.- INTRODUCCION

Este capítulo está dedicado a estudiar en profundidad la distribución binomial, haciendose especial hincapié en la estimación de sus parámetros y en la aproximación de sus probabilidades. Asímismo se estudia detenidamente la generalización de la distribución realizada por Poisson. Las siguientes páginas apenas contienen resultados originales, pero si, en cambio, la mayor parte del esfuerzo de documentación llevado a cabo para comprobar la originalidad de nuestra generalización de la binomial, que hemos denominado distribución binomial métrica.

Como parte de la información contenida en el capítulo será conocida por los lectores formados en nuestra disciplina, resulta preferible omitir inicialmente su lectura, realizandola después sólo de aquellas secciones necesarias para seguir el desarrollo del último capítulo de la memoria.

Como todos sabemos, la distribución binomial es una de las distribuciones de probabilidad mas antiguas que se han estudiado. Fue deducida por Jacobo Bernoulli , que la dio a conocer en su obra póstuma "Ars Conjectandi",

publicado en 1713.

Es también una de las distribuciones que más comunmente se conocen, de estructura muy sencilla, y una de las que más importancia han tenido en el desarrollo — del Cálculo de Probabilidades y de la Estadística Matemática. Su estudio en profundidad condujo al descubrimiento de grandes resultados, así como a la creación de importantes unificaciones y generalizaciones dentro de esta rama de las Matemáticas.

En este mismo sentido, dado su importante papel, creemos que cualquier nuevo estudio sobre esta distribución que conlleve la aportación de un nuevo punto de vista o enfoque sobre su naturaleza puede tener alguna importancia al arrojar distinta luz en el campo de las Probabilidades y la Estadística. En esta línea de pensamiento, creemos que podría encuadrarse el contenido de la presente memoria, como tendremos la oportunidad de comprobar — más adelante.

Con objeto de reafirmar el importante papel que desempeña la distribución binomial, recordaremos que está intimamente relacionada con muchas distribuciones

de probabilidad, tanto discretas como contínuas. Destacamos entre las mismas distribuciones discretas a la distribución hipergeométrica, la Poisson y la binomial negativa. Entre las de tipo contínuo destacamos a la distribución Normal, a la beta y a la F de Snedecor. Con la primera está relacionada, como todos conocemos, por el teorema de Moivre-Laplace, cuya generalización dio como fruto la teoría del problema central del límite.

I.2.- DEFINICION

La distribución binomial es una distribución de probabilidad discreta reticular definida en un intervalo de los números enteros no negativos. Depende de dos parámetros, n y p, siendo n un número natural y p un valor del intervalo cerrado $[0,1]$.

Algunos autores prefieren restringir p al intervalo abierto $(0,1)$, pero entonces no se engloban las variables degeneradas correspondientes al caso determinístico, variables que nosotros preferimos incluir.

Una variable aleatoria X, decimos que se distribuye según una ley binomial B(n,p), si todo valor entero r comprendido entre 0 y n tiene asignada una masa de probabilidad dada por

$$P(X=r) = b_r = \binom{n}{r} p^r . (1-p)^{n-r}$$

La denominación de binomial procede del hecho de que la probabilidad de que la variable X tome el valor r, b_r, es igual al término r+1 del desarrollo del binomio

$$(q+p)^n$$

donde q siempre representará 1-p.

Una realización teórica de la distribución binomial se encuentra al estudiar la frecuencia absoluta de un determinado resultado E a lo largo de n pruebas sujetas a las siguientes suposiciones: a) la probabilidad de que el resultado E se presente en cualquier prueba vale siempre p; b) la realización de cada prueba es estadísticamente independiente de las restantes. En tales hipótesis, la distribución de la frecuencia absoluta del resultado E es una binomial B(n,p).

Este modelo se utiliza repetidamente en las situaciones reales que se presentan en la estadística aplicada. Las suposiciones de independencia y probabilidad constante pueden no ser estrictamente correctas en la práctica, pero frecuentemente su adopción conduce a una representación suficientemente precisa de los fenómenos que se están tratando. Incluso cuando no sean válidas, el modelo binomial proporciona un punto de referencia a partir del cual podemos cuantificar las desviaciones existentes entre el modelo y la realidad.

I.3.- PERTENENCIA A FAMILIAS DISCRETAS DE PROBABILIDAD

Vamos a tratar aquí sobre la pertenencia o no de la distribución binomial a las principales familias o clases discretas de distribuciones de probabilidad. Nos centraremos casi exclusivamente sobre las dos grandes familias de distribuciones discretas: la familia de distribuciones de serie de potencias generalizadas y el sistema de Pearson discreto. Nos extenderemos un poco para in

cluir sus definiciones y una interrelación existente entre ambos. De las restantes familias a las que también pertenece la binomial, daremos escuetamente sus definiciones al final de la sección.

El hecho de que las probabilidades puntuales binomiales se expresen como términos del desarrollo en serie de potencias del binomio $(p + q)^n$, permite incluir esta distribución dentro de una amplia familia de distribuciones discretas denominadas clase de distribuciones de serie de potencias generalizadas. Tal clase fue considerada inicialmente por M.C. Tweedie (1947) y difundida extensamente entre los especialistas en distribuciones de probabilidad por A. Noack (1950).

Se dice que una variable aleatoria discreta que tiene valores en un conjunto Ω de números reales sigue una distribución de serie de potencias generalizadas sí y sólo sí

$$f_r = a(r).\theta^r/A(\theta) \qquad \forall r \in \Omega$$

estando $A(\theta)$ definido como $\sum_{r\in\Omega} a(r).\theta^r$, lo cual asegura por normalización que tengamos una auténtica distribución probabilística.

Para mostrar la pertenencia de la binomial a esta clase es necesario sustituir el parámetro p por otro θ definido como p/q, y considerar como función $A(\theta)$ el binomio $(1+\theta)^n$, ya que

$$f_r = a(r).\ \theta^r/A(\theta) = \binom{n}{r}.\ {}^r/(1+\theta)^n =$$
$$= q^n.\binom{n}{r}.\ p^r/q^r = \binom{n}{r}\ p^r.q^{n-r} = b_r$$

Por otra parte, la familia binomial está incluída además en la primera clase importante de distribuciones discretas. Se trata del sistema de Pearson discreto, cuyos miembros verifican cada uno la siguiente ecuación en diferencias finitas:

$$\Delta f_{r-1} = \frac{(a-r).\ f_{r-1}}{b_0+b_1 r+b_2 r.(r-1)}$$

donde a, b_0, b_1, b_2, representan constantes, y Δf_{r-1} a la diferencia $f_r - f_{r-1}$. Tal ecuación fue considerada por primera vez por K. Pearson en 1895.

Según la clasificación expuesta por J.K. Ord (1967), la distribución binomial está incluída en el tipo III y verifica la ecuación en diferencias

$$\Delta b_{r-1} = \frac{(n+1)p-r}{r.q} \cdot b_{r-1}$$

Respecto a estas dos grandes familias de distribuciones podemos señalar que no son completamente independientes. Existe un nexo de unión entre ambas expresado por el siguiente resultado: si una variable aleatoria definida sobre los enteros cumple que para todo r, su masa de probabilidad, f_r, verifica la ecuación reducida en diferencias finitas

$$\Delta f_{r-1} = \frac{(a-r).f_{r-1}}{r.b_1}$$

entonces su función de masa de probabilidad f_r tiene que ser de la forma

$$f_r = a(r,\theta).\theta^r/A(\theta,\varphi)$$

Evidentemente, de forma indirecta, con este resultado comprobamos que la binomial pertenece a la clase de distribuciones de series de potencias generalizadas, sin necesidad de conocer la forma explícita del parámetro θ.

A continuación vamos a dar las definiciones de otras familias de distribuciones a las que también per

tenece la binomial. Estas son las siguientes:

a) Distribuciones unimodales: Se dice que una distribución f_r, discreta definida sobre los enteros, es unimodal si existe al menos un entero s, para el que se verifica:

$$f_{r-1} \leqslant f_r \qquad \text{sí} \quad r \leqslant s$$

y

$$f_{r+1} \leqslant f_r \qquad \text{sí} \quad r \geqslant s$$

b) Distribuciones fuertemente unimodales: Son aquellas f_r tales que su convolución con cualquier otra unimodal siempre es unimodal.

c) Familia exponencial: Una familia uniparamétrica $f_r(\theta)$ de funciones de probabilidad pertenece a la familia exponencial regular, si $f_r(\theta)$ puede expresarse como:

$$f_r(\theta) = \exp\,[Q(\theta).\ T(r) + S(r) + q(\theta)]$$

donde los valores de r no dependen de θ, $Q(\theta)$ es una función contínua no trivial de θ y T(r) es una función no trivial de r.

d) Distribuciones de razón de verosimilitud monótona:

Una familia uniparamétrica $f_\theta(\theta)$, con θ variando en un intervalo, pertenece a esta familia respecto a un cierto estadístico T, $T= t(x_1, x_2,\ldots,x_n)$, si el cociente de verosimilitud

$$\frac{L(x_1,x_2,\ldots,x_n;\theta_1)}{L(x_1,x_2,\ldots,x_n;\theta_2)} \qquad \theta_1 \leq \theta_2$$

es una función monótona de T.

e) Distribución de razón de fallo creciente: Son aquellas distribuciones discretas X para las cuales la razón de fallo, definida como:

$$0 \leq r(j) = \frac{P[X=j]}{P[X\geq y]} = \frac{f_j}{1-F(j-1)} \leq 1$$

es una función no decreciente de j.

Propiedades, resultados y referencia a los originales donde se trata sobre estas clases de distribuciones de probabilidad pueden encontrarse en el libro de Patel, Kapadia y Owen (1976)

I.4.- FUNCIONES GENERATRICES

La función generatriz de probabilidad se establece facilmente en virtud a su pertenencia a la clase de las distribuciones de serie de potencias generalizadas. Cualquier distribución discreta entera positiva de esta familia tiene como función generatriz de probabilidad G(z) a la expresión:

$$\frac{A(\theta z)}{A(\theta)}$$

Por tanto, para la binomial tenemos que

$$G(z) = \frac{(1+\theta z)^n}{(1+\theta)^n} = (q + pz)^n$$

Por supuesto, también resulta inmediata su obtención a partir del cálculo directo.

Las restantes funciones generatrices se obtienen mediante una transformación adecuada sobre la función anterior; de este modo, la función generatriz de momentos y la característica se obtienen, respectivamente, sustituyendo z por e^t ó e^{it}. La función generatriz de cumulantes es simplemente, el logaritmo natural de la función genera

triz de momentos, es decir

$$K(t) = n \ln (q + pe^t)$$

I.5.- MOMENTOS Y CUMULANTES. RELACIONES DE RECURRENCIA

Veamos en primer lugar los momentos de la distribución. Entre estos es de destacar la particular simplicidad de los momentos factoriales, que vienen dados por la fórmula

$$\mu(r) = n_{(r)} \cdot p^r = n.(n-1)...(n-r+1).p^r$$

Los restantes momentos de orden positivo se pueden determinar a partir de estos primeros, empleando las relaciones que ligan unos con otros. Así, teniendo en cuenta que

$$x^r = \sum_{j=0}^{r} \frac{x_{(j)}}{j!} \quad \Delta^j 0^r$$

y tomando esperanzas, se llega a que el r-ésimo momento respecto al origen viene dado por la fórmula simbólica

$$\mu'_r = (1 + p\Delta)^n . \ 0^r$$

De todas los momentos de la distribución binomial, vamos a dar valor de los principales que tienen interés:

μ'_1 = valor medio = n.p

μ_2 = varianza = n.p.q

μ_3 = n.p.q(q-p)

$\mu_4 = 3(n.p.q)^2 + np.q(1-6pq)$

De los tres últimos, deducimos que los coeficientes de asimetria y curtosis valen:

$$\sqrt{\beta_1} = (q-p)/\sqrt{npq}$$

$$\beta_2 = 3 + \frac{1-6pq}{npq}$$

M. Merrel (1933), estudió las relaciones existentes entre β_1 y β_2 . Comprobó que, para valores fijos de p, los puntos (β_1, β_2) cuando el parámetro n varía caen sobre la línea recta

$$\beta_2 - 3 = \left\{1 - \frac{2pq}{(q-p)^2}\right\} \cdot \beta_1$$

Cuando n crece el punto se desplaza a lo largo de la recta acercándose al punto (0,3), teniéndolo como límite.

De la misma manera, se comprueba que, para n fijo, los puntos se hallan sobre la recta

$$\beta_2 - \beta_1 - \frac{3n-2}{n} = 0$$

paralela a la línea teórica que delimita la zona de valores imposibles de (β_1, β_2) en cualquier distribución: $\beta_2-\beta_1 - 1 = 0$. Cuando n tiende a infinito se consigue la línea correspondiente a la distribución de Poisson: $\beta_2-\beta_1 - 3 = 0$. Finalmente estudia la línea que se obtiene cuando el producto np permanece fijo, resultando en esta ocasión que se trata de una hipérbola.

Volviendo otra vez a los momentos, sucede que todos los momentos centrales están relacionados entre sí por una ecuación de recurrencia que, como indicó su descubridor, permite el cálculo de los mismos de una manera más sencilla que a partir de la función generatriz de momentos. Dicha ecuación fue deducida por Romanovsky (1923); su forma es la siguiente:

$$\mu_{r+1} = pq \left[n.r. \ \mu_{r-1} + \frac{d \ \mu_r}{dp} \right]$$

sujeta, indudablemente, a la condición inicial $\mu_0 = 1$, $\mu_1 = 0$.

Una relación similar también se presenta entre los momentos respecto del origen:

$$\mu'_{r+1} = p \cdot q \left[\frac{n}{q} \mu'_r + \frac{d\mu'_r}{dp} \right]$$

Es importante señalar que también esta última ecuación se verifica para los momentos incompletos de la distribución, entendiendo por éstos a los correspondientes a la distribución obtenida al eliminar cualquier número de puntos con masa de probabilidad no nula.

La explicación de la existencia de tales ecuaciones de recurrencia y de su validez para las distribuciones incompletas se conoce ahora claramente a la luz de las propiedades generales de la familia difundida por Tweedie y Noack; este último autor demostró en 1950 que para cualquier miembro de la familia, sus momentos respecto del origen verifican la ecuación de recurrencia:

$$\mu'_{r+1} = \theta \cdot \frac{d\mu'_r}{d\theta} + \mu'_1 \cdot \mu'_r$$

Por otra parte, es evidente que cualquier distribución incompleta obtenida a partir de una perteneciente a la fami

lia, también pertenece a la familia; de donde la extensión de la validez de la relación de recurrencia a las distribuciones incompletas es obvia.

Sin embargo, para el estudio de distribuciones de esta familia resulta mucho más conveniente considerar cumulantes en vez de momentos, por la sencilla razón de que también los cumulantes verifican una ecuación de recurrencia (para otros tipos de momentos y cumulantes existen también ecuaciones similares), que resulta ser mucho más simple. Tal ecuación es

$$K_{r+1} = \theta \frac{dK_r}{d\theta} \qquad r = 1, 2, \ldots$$

Particularizando al caso de la binomial, deducimos que sus cumulantes verificarán:

$$K_{r+1} = p.q \frac{dK_r}{dp}$$

ya que

$$\frac{dK_r}{d\theta} = \frac{dK_r}{dp} \cdot \frac{dp}{d\theta} = \frac{dK_r}{dp} \cdot \frac{1}{\frac{d\theta}{dp}} = q^2 \frac{dK_r}{dp}$$

recordando que θ es $\frac{p}{q}$. Teniendo en cuenta que el primer cumulante es el valor medio, por cálculo sucesi-

vo hemos obtenido los seis primeros, cuyos valores damos a continuación:

$K_1 = n.p$ $\qquad$ $K_6 = n.p.q.\{1-30p.q(1-4pq)\}$

$K_2 = n.p.q.$ $\qquad$ $K_5 = n.p.q.(q-p)[1-12p.q]$

$K_3 = n.p.q.(q-p)$ $\qquad$ $K_4 = n.p.q.(1-6p.q)$

I.6.- OTROS TIPOS DE MOMENTOS

En esta sección vamos a tratar de los siguientes momentos: el primer momento central absoluto, los momentos inversos factoriales, los momentos inversos de la binomial positiva y por último, las diferencias medias generalizadas.

a) El primer momento central absoluto, o corrientemente llamado desviación media, recordamos definido como

$$\nu_1 = E[|X-np|]$$

aparece calculado en el trabajo de N.L. Johnson(1957). Su valor es

$$\nu_1 = 2.n \binom{n-1}{[np]} p^{[np]+1} q^{n-[np]}$$

donde [x] significa la parte entera del número x .
Cuando la media binomial es un número entero la desviación media vale simplemente $2\mu_2 b_m$, siendo b_m la frecuencia del valor modal de la distribución. Este autor seguidamente estudia una expresión aproximada de esta cantidad utilizando la fórmula de Stirling para factoriales. El resultado obtenído es

$$\nu_1 \simeq \sqrt{\frac{2}{\pi} n.p.q} \left[1 + \frac{\delta(np).\delta(nq)}{2npq} - \frac{1-pq}{12npq}\right]$$

donde $\delta(x)$ representa la parte decimal de x, o sea $\delta(x) = x - [x]$. Como era de esperar por la convergencia hacia la normal, el cociente entre ν_1 y la desviación típica se aproxima al valor límite $\sqrt{\frac{2}{\pi}}$ cuando n tiende a infinito.

b) Los momentos inversos factoriales, definidos como

$$E\left[\frac{1}{(x+r)_{(r)}}\right] \text{ ó equivalentemente } E\left[\frac{1}{(x+1)_{[r]}}\right]$$

Se pueden calcular en la distribución binomial por integración sucesiva de la función generatriz de pro-

babilidad entre cero y uno como límites de integración. Su resultado es

$$E\left[\frac{1}{(x+r)_{(r)}}\right] = \frac{1}{(n+r)_{(r)}.p^r}\left\{1 - \sum_{y}^{n+r} p^y.q^{n+r-y}\right\}$$

fórmula que queda simplificada a

$$\frac{1}{(n+r)_{(r)}.p^r} \cdot P\,[y \geqslant r]$$

Definiendo la variable aleatoria y como distribuída según la binomial B(n + r, p)

c) Los momentos negativos respecto del origen, a diferencia de los factoriales negativos, valen siempre infinito o, mejor dicho, no están definidos, puesto que b_0 es siempre positiva. No obstante, podemos considerar estos momentos si excluímos de la variable el valor cero. La variable truncada así obtenida se suele llamar binomial positiva.

Estos momentos tienen importancia porque son utilizados para algunas situaciones prácticas. Para ellos, no existe ninguna fórmula breve que proporcione su

valor. De entre los autores que se han dedicado a su estudio podemos citar a:

- F.F. Stephan (1945), quien calcula estos momentos a partir de un desarrollo en serie de factoriales inversos de $1/X$ (que al tomar esperanza matemática se convierten en los momentos inversos factoriales tratados en el apartado b). El cálculo de estos momentos es aproximado empleando el número de términos que se desee del desarrollo en serie, el cual converge muy rápidamente y sus términos son de fácil iteración.

- E. Grab y J. Savage (1954), quienes tabulan el primer momento basándose en la relación de recurrencia

$$E\left[\frac{1}{X};\ n+1;\ p\right] = \frac{1}{n+1} + q.\ \frac{1-q^{n}}{1-q^{n+1}}\ .\ E\left[\frac{1}{X};\ n;p\right]$$

y además, establecen una aproximación empírica para el mismo

$$E\left[\frac{1}{X}\right] \simeq \frac{1}{np-q}$$

- W. Mendenhall y E.H. Lehman (1960), que ob-

tienen fórmulas aproximadas al ajustar una distribución beta a la binomial relativa positiva utilizando los correspondientes momentos negativos de la beta. Entre otras fórmulas obtienen que

$$E\left[\frac{1}{X}\right] \simeq \frac{1-\frac{2}{n}}{(n-2)p-q}$$

$$V\left[\frac{1}{X}\right] \simeq \frac{(1-\frac{2}{n})\cdot(1-\frac{1}{n})\cdot q}{((n-2)p-q)^2((n-2)p-q-1)}$$

Como puede observarse en el libro de Johnson y Kotz (1969), pág. 74, las fórmulas que allí se encuentran no están bien expresadas.

En el trabajo de Mendenhall y Lehman, se hace una aplicación de estos momentos a problemas de test de vida en los que intervienen la distribución de Weibull.

- Z. Govindarajulu (1963), que generaliza la relación de recurrencia de Grab y Savage para órdenes superiores y estudia además los errores de redondeo acumulado que se cometerían al aplicarlas para la determinación de estos momentos. También es notable la mención que allí

se hace de una ecuación diferencial para el primer momento negativo que obtuvo en 1952 el profesor portugués Tiago de Oliveria.

d) Para cualquier distribución discreta, las diferencias medias generalizadas vienen definidas por la expresión

$$\Delta_r = \sum_i \sum_{i \neq j} |i - j|^r \; f_i . f_j$$

Constituyen una generalización de la diferencia media de Gini, que es una medida de dispersión independiente de cualquier medida de localización. Esta última diferencia se puede calcular también para cualquier distribución discreta entera y positiva a partir de su función generatriz de probabilidad G(z), viniendo dada por la fórmula teórica

$$\Delta_1 = \left[P \frac{d}{dz} \{ G(z) \; G(z^{-1}) \} \right]_{z=1}$$

donde P denota un operador que realiza únicamente la suma sobre los términos en z con el exponente no negativo.

Esta fórmula está mencionada en el libro de

J.K. Ord (1972). Nosotros ahora vamos a dar su generalización, la cual permitirá obtener las diferencias medias factoriales generalizadas (definidas simplemente sustituyendo en la definición dada $|i-j|^r$ por $|i-j|_{(r)}$)

$$\Delta_{(r)} = P\left[\frac{d^r}{dz^r}\left\{G(z).\ G(z^{-1})\right\}\right]_{z=1}$$

A partir de estas diferencias la obtención de las primeras es idéntica a la obtención de los momentos ordinarios respecto de los factoriales. Pero también las diferencias medias de orden par se expresan en función de los momentos centrales por medio de la expresión

$$\Delta_{2r} = 2\,\mu_{2r} + \sum_{j=2}^{2r-2} (-1)^j \binom{2r}{j} \mu_{2r-j}\ \mu_j$$

En cambio, cuando el orden es impar, las diferencias son de difícil cálculo. T. Ramasubban (1959) estudia las diferencias medias generalizadas en el caso particular de la binomial y la poisson. Para la distribución binomial vienen dadas por una expresión complicada en la que interviene la función hipergeométrica con ciertos argumentos. Concretamente para el primer orden resulta

$$\Delta_1 = 2 \cdot \mu_2 \cdot F(-n + 1, \nu 2, 2; 4pq)$$

recordando que la función hipergeométrica está definida por el desarrollo en serie

$$F(\alpha, \beta, \gamma; x) = 1 + \frac{\alpha \cdot \beta}{\gamma} \frac{x}{1!} + \frac{\alpha \cdot (\alpha+1) \beta \cdot (\beta+1)}{\gamma \cdot (\gamma+1)} \cdot \frac{x^2}{2!}$$

donde el argumento γ es siempre positivo; la convergencia de la serie está asegurada siempre que $|x| < 1$.

En la conclusión de su artículo, el autor citado indica la posibilidad de conseguir relaciones de recurrencia entre las diferencias medias generalizadas de esta distribución, añadiendo que él mismo, en aquella fecha, se hallaba muy próximo a su descubrimiento completo.

I.7.- PRINCIPALES PROPIEDADES

Vamos a tratar de las siguientes propiedades: Localización de la moda; teorema de Moivre-Laplace; suma y diferencia de binomiales independientes con igual parámetro p; distribución condicionada de una binomial cuando la suma con otra independiente es conocida. Pasemos segui

damente, a desarrollar estas propiedades.

a) Localización de la moda:

El valor modal se determina inmediatamente a partir de la ecuación en diferencias de la familia

$$\frac{\Delta b_{r-1}}{b_{r-1}} = \frac{(n+1).p - r}{r \cdot q}$$

La última vez que Δb_{r-1} es positivo tiene lugar cuando r vale $[(n+1).p]$, ó de otra forma $[E[X] + p]$.

En el caso particular de que $(n+1).p$ fuese entero, la distribución es bimodal, siendo la segunda moda el punto inmediatamente anterior a la moda dada.

Obsérvese que este resultado visto de otro modo significa que la moda en la distribución binomial queda perfectamente determinada cuando se conoce su media. Posteriormente, veremos que esta afirmación es, hasta cierto punto, también válida para la mediana.

b) Teorema de Moivre-Laplace:

Podemos enunciarlo en los siguientes términos: Cuando el parámetro n tiende a infinito, la variable bino

mial tipificada $X^1 = (X-np)/\sqrt{npq}$ tiende a la distribución Normal N(0,1); lo cual significa que

$$\lim_{n \to \infty} P[\alpha < X < \beta] = \int_{\beta'}^{\alpha'} \frac{1}{\sqrt{2\pi}} e^{-1/2 \cdot x^2} dx$$

siendo

$$\beta' = \frac{\beta - n.p}{\sqrt{npq}} \quad y \quad \alpha' = \frac{\alpha - n.p}{\sqrt{npq}}$$

Este resultado constituye un teorema límite. No obstante, su principal empleo en la práctica es realizar un cálculo aproximado rápido de las probabilidades binomiales. Sobre este punto y los teoremas de aproximación resultantes profundizaremos en la sección 11 y 13 de este capítulo.

c) Suma de binomiales:

La distribución binomial, manteniendo el parámetro p constante, es reproductiva respecto del parámetro n; es decir, dadas dos variables binomiales independientes X e Y de parámetros (n_x,p) y (n_y,p), la variable Z, suma de ambas, se distribuye también binomialmente con igual parámetro p y parámetro n_z igual a n_x+n_y.

En términos de convolución, lo que hemos dicho se puede escribir como

$$B(n_x,p) * B(n_y,p) = B(n_x+n_y,\ p)$$

Como después veremos, este resultado puede invertirse convirtiéndose así en una caracterización de esta distribución.

d) Diferencia de binomiales:

Considerando las mismas variables X e Y del apartado anterior, la distribución de su diferencia $V = X - Y$, por construcción directa, tiene como función de masa de probabilidad

$$P\,[V= K\,] = \sum_{x} \binom{n_x}{x} \cdot \binom{n_y}{x-K} \cdot p^{2x-K} \cdot q^{n_x+n_y-2x-K}$$

donde la suma se realiza entre los límites

$$\max(0,K) \leq x \leq \min\,(N_1,\ N_2+K)$$

En el caso particular en que p y q son iguales, la distribución de la diferencia adopta la forma

$$P[V=k] = \binom{n_x+n_y}{n_y+k} \cdot \left\{\frac{1}{2}\right\}^{n_x+n_y}$$

con k comprendido entre $-n_y$ y n_x, ambos inclusive; es decir, la variable V se distribuye como la binomial $B(n_x+n_y, \frac{1}{2})$ desplazada n_y unidades a la izquierda. Ahora bien, si englobamos en la definición de variable binomial a toda variable que se obtenga a partir de una binomial X, mediante una transformación lineal $a+bX$, con b no nulo, entonces resulta, en este caso particular, que tanto la suma como la diferencia de binomiales independientes siguen también leyes binomiales.

Si aplicamos el teorema de Laplace y la independencia entre X e Y, se deduce la convergencia hacia la normal unitaria de la distribución de la diferencia estandarizada

$$\frac{X-Y-p(n_x-n_y)}{\left\{p \cdot q \cdot (n_x+n_y)\right\}^{1/2}}$$

cuando n_x y n_y tienden a infinito sin ninguna restricción.

e) Una distribución condicionada:

Siguiendo con las variables aleatorias de los

dos puntos anteriores, consideremos la distribución condicionada que tiene la variable X cuando sabemos que su suma con Y toma un valor dado C. Tal distribución condicionada se deduce fácilmente, resultando destacable que ésta es independiente del parámetro p. Su función correspondiente de masa puntual de probabilidad es

$$P\left[X=K \;/\; X+Y=C\right]=\frac{\binom{n_x}{K}\binom{n_y}{C-K}}{\binom{n_x+n_y}{C}}$$

verificándose naturalmente que

$$\max(0, C-n_y) \leqslant K \leqslant \min(n_x, C)$$

Como se distingue claramente, esta variable aleatoria sigue una ley hipergeométrica de parámetros n_x+n_y, n_x y C. Seguidamente veremos en la siguiente sección que este mismo resultado se puede utilizar como caracterización de la binomial.

I.8.- CARACTERIZACIONES DE LA LEY BINOMIAL

Vamos a tratar tres caracterizaciones de la

distribución binomial. Las dos primeras son los reciprocos de las propiedades c) y d) de la sección anterior; en la ultima caracterización se emplean los cumulantes de la distribución.

a) Si una variable Z sigue una ley binomial y es suma de dos variables X e Y, independientes y no negativas, entonces ambas variables siguen leyes binomiales.

Esta propiedad es corolario del siguiente resultado de D.N. Shanbhag y I.V. Basawa (1974), del cual tuvimos noticia por su abstract publicado el año 1971, que caracteriza a la distribución multinomial:

> Teorema. Si X e Y son variables aleatorias k-dimensionales, no negativas e independientes, y si Z=X+Y es una variable multinomial con probabilidades $(p_1, p_2, \ldots, p)$, entonces X e Y son también variables multinomiales con el mismo conjunto de probabilidades.

La demostración es extremadamente simple, basandose en la estructura polinomica de las funciones generatrices de probabilidad de X e Y.

b) Si la probabilidad condicional de una variable aleatoria X cuando se conoce el valor de su suma con otra variable aleatoria Y, ambas independientes entre sí, sigue una ley hipergeométrica $H(n, n_x, C)$, entonces X e Y tienen distribuciones binomiales de parámetros (n_x, p) y $(n-n_x, p)$, respectivamente, siendo p algún valor del intervalo (0,1).

La primera mención que tenemos del presente resultado se encuentra en el abstract de K.G. Janardan aparecido el año 1971, en A.M.S. 42, 3, pág. 1151, que también es una generalización para la multinomial y la multinomial negativa. Puede encontrase la demostración desarrollada, para la binomial, en el libro de A. Kagan, Y.V. Linnik y C.R. Rao (1973), en la página 424. En la misma se obtienen las distribuciones de X e Y en la forma de serie de potencia generalizada de la binomial con parámetro θ indeterminado.

c) Sean $X_1, X_2, \ldots, X_N$ variables aleatorias independientes e identicamente distribuídas con función de distribución F(x) tal que admite como mínimo momentos finitos de tercer orden. Sean k_1, k_2 los dos primeros cumulantes, tales que

$$0 < \frac{K_2}{K_1} = p < 1 \quad y \quad \frac{K_1^2}{(K_1 - K_2)} = n \in Z$$

Sean $L_K = \sum_{i=1}^{n} x_i^k$, para k= 1, 2, 3, 4 , y

$$T = N.L_4 + (N-4)L_3.L_1 + (3-2N)L_2^2$$

$$+ L_2L_1^2 - NL_3 + (N+1)L_2L_1 - L_1^3$$

Entonces, la regresión de T sobre L_1 es cero si y sólo sí la función de distribución F(x) es binomial de parámetros n y p.

Este resultado procede de E. Lukacs (1965); fue presentado en el Simposium sobre distríbuciones discretas clásicas y de contagio, celebrado en Montreal.

I.9.- ESTIMACION DE LOS PARAMETROS

Vamos a comenzar esta sección tratando de la estimación individual de cada parámetro cuando el valor del otro resulta conocido. Estudiemos en primer lugar la

estimación de la probabilidad de éxito. Para ello consideremos una muestra aleatoria simple $x_1, x_2, \ldots, x_N$, de tamaño N, correspondiente a una variable binomial B(n,p), con n conocido.

Como todos sabemos, el estimador de máxima verosimilitud de p es la frecuencia relativa del número total de éxitos, o sea,

$$\hat{p} = \frac{1}{nN} \sum_{j=1}^{n} x_j$$

De hecho $S = \sum_{j=1}^{n} x_j$ es un estadístico suficiente para p, y como su distribución en el muestreo es también binomial, nos podemos limitar a considerar muestras unitarias en relación con la estimación de p. Esta misma afirmación también es válida cuando la muestra procede de varias variables binomiales con distintos números de pruebas.

El estimador $\hat{p}$ es insesgado. Su varianza es pq/nN, que coincide con la cota inferior de Cramer-Rao para los estimadores insesgados; de modo que $\hat{p}$ es el estimador de mínima varianza de p.

Después de este trozo corrientemente conocido por el lector, pasamos a considerar, de manera general, la estimación del parámetro de una familia de distribuciones definidas por una serie de potencias generalizadas. Estos resultados proceden de J.Roy y S. Mitra (1957).

Supongamos una muestra aleatoria simple $x_1, x_2, \ldots, x_N$ de una variable de la familia definida por la función $A(\theta)$. La verosimilitud de la muestra será entonces

$$\prod_{j=1}^{N} \left\{ \frac{a(x_j)\theta^j}{A(\theta)} \right\} = \theta^{\sum_{j=1}^{N} x_j} \cdot A(\theta)^{N} \prod_{j=1}^{N} a(x_j)$$

Derivando respecto a θ, se encuentra que el estimador de máxima verosimilitud $\hat{\theta}$ debe satisfacer la ecuación

$$\hat{\theta}^{-1} \cdot \sum_{j=1}^{N} x_j - N.A'(\hat{\theta})/A(\hat{\theta}) = 0$$

Por tanto $\hat{\theta}$ es una función de S, si empleamos la notación anterior; $\hat{\theta}$ no depende de las observaciones muestrales x_j en ninguna otra forma. Entonces, la distribución condicional de $x_1, x_2, \ldots, x_N$, conocido S, no depende de θ; es decir, S es siempre un estadístico suficiente de θ. Como distribución de probabilidad, S también pertenece a

la familia de las distribuciones de series de potencias generalizadas, estando determinada por la función $\{A(\theta)\}^N$. Por otro lado, también se puede demostrar que S es completo.

Para este tipo de distribuciones siempre existen estimadores insesgados de mínima varianza para θ suponiendo que su conjunto de definición Ω está incluído en los enteros no negativos. La definición de este estimador se realiza por medio de los coeficientes del desarrollo en serie de la función $\{A(\theta)\}^N$, definidora de la distribución de S. Si representamos por c(k) al coeficiente de θ^k de esta función, el estimador deseado viene definido por

$$\hat{\theta} = \begin{cases} c(S-1)/c(S) & \text{si } S>0 \\ 0 & \text{si } S=0 \end{cases}$$

Si aplicamos lo que acabamos de ver a la binomial, vemos que $\{A(\theta)\}^N$ vale

$$\left[(1+\theta)^n\right]^N = (1+\theta)^{n.N}$$

de modo que los coeficientes de su desarrollo correspondientes a S - 1 y S son

$$b(S-1) = \binom{nN}{S-1}, \quad b(S) = \binom{nN}{S}$$

Por tanto, el estimador insesgado de mínima varianza de θ será

$$\hat{\theta} = \frac{\frac{(nN)!}{(S-1)!.(nN-S+1)!}}{\frac{(nN)!}{S!.(nN-S)!}} = \frac{\sum_{j=1}^{N} x_j}{nN + 1 - \sum x_j}$$

Suponiendo que $\hat{\theta}$ viniese dado por el cociente de ciertas estimaciones de p y q, $\tilde{p}$ y $1-\tilde{p}$, despejándola obtendríamos este estimador suficiente para p:

$$\tilde{p} = \frac{\sum_{j=1}^{N} x_j}{nN + 1} = \frac{n \cdot N}{nN + 1} \hat{p}$$

que, como vemos, es distinto del estimador máximo-verosimil y además es sesgado. Si expresamos $\hat{\theta}$ en función de $\hat{p}$, resulta entonces que

$$\hat{\theta} = \frac{nN \cdot \hat{p}}{nN \cdot q + 1}$$

Consideremos ahora la construcción de intervalos de confianza para la probabilidad de éxito. Por tratarse de una distribución discreta, generalmente no es po

sible realizar la estimación por intervalos no aleatorizada del parámetro p con un nivel de confianza exactamente especificado. En la practica se utilizan dos procedimientos que dan niveles de confianza mayores o iguales al especificado.

El mas antiguo consiste en determinar aquellos valores de p que verifican que las colas de la distribución en el muestreo a partir del valor observado tienen una probabilidad exactamente igual a la mitad del nivel de riesgo. Dicho de otra forma, los limites inferior y superior del intervalo buscado, p_I y p_S, son, respectivamente, las soluciones de las ecuaciones:

$$\sum_{j=S}^{nN} \binom{nN}{j} \cdot p_I^j \cdot (1-p_I)^{nN-j} = \frac{1}{2}\alpha$$

y

$$\sum_{j=o}^{S} \binom{nN}{j} \cdot p_S^j (1-p_S)^{nN-j} = \frac{1}{2}\alpha$$

Clopper y Pearson (1934) dieron en forma grafica estos intervalos a los niveles del 95% y 99%. Tablas numéricas de estos intervalos pueden encontrarse, por ejemplo, en Fisher y Yates (1951).

En virtud de la interrelación que existe entre

las funciones de distribución de la binomial y la beta, de la cual trataremos en la sección siguiente, los valores de p_I y p_S pueden obtenerse a partir de :

$$F^{\beta}_{p_I}(S;nN-S+1) = \frac{1}{2}\alpha$$

y

$$F^{\beta}_{p_S}(S+1;nN-S) = 1 - \frac{1}{2}\alpha$$

Donde $F^{\beta}_{p}(\alpha,\beta)$ representa la función de distribución de la ley beta de parámetros α y β, en el punto p. Dicha cantidad se denomina también razón incompleta de la función beta; fue tabulada por K. Pearson para parámetros inferiores o iguales a 50.

Igualmente, estos intervalos también pueden encontrarse expresados en función de los valores críticos de la F de Snedecor con determinados grados de libertad.(cf. J.A. Viedma (1972),pág. 128). Tal hecho se basa en que el valor crítico izquierdo $F_{\alpha}(\nu_1,\nu_2)$, F, satisface la ecuación

$$F^{\beta}_{\frac{\nu_2 F}{\nu_1+\nu_2 F}}(\tfrac{1}{2}\nu_1,\tfrac{1}{2}\nu_2) = \alpha$$

de donde se sigue que

$$p_I = \frac{\nu_1 \cdot F_{\alpha/2}(\nu_1, \nu_2)}{\nu_2 + \nu_1 \cdot F_{\alpha/2}(\nu_1, \nu_2)}$$

siendo $\nu_1 = 2 \cdot S$ y $\nu_2 = 2(nN-S+1)$.

Por el segundo procedimiento se construyen regiones de confianza a partir de los intervalos de aceptación que contienen las observaciones mas probables. No siempre las regiones de confianza obtenidas son intervalos, pero se puede lograr que siempre lo sean exigiendo que los extremos de los intervalos de aceptación sean funciones no decrecientes de p.

Este procedimiento ha sido empleado por E.L. Crow (1956), quién ha tabulado los intervalos correspondientes a los niveles de confianza del 90, 95 y 99%, cuando n varía entre 1 y 30; para valores de n superiores a 30, los métodos aproximados dan practicamente los mismos resultados. Crow ha demostrado que estos intervalos tienen la propiedad de minimizar la suma de las n+1 amplitudes posibles.

C.W. Clunies-Ross ha estudiado algunas propiedades de ambos procedimientos, señalando que este último es mas selectivo, puesto que a igualdad de nivel de confianza da intervalos mas cortos.

Aparte de estos procedimientos, se han publicado tablas de intervalos basandose en otros métodos . Así, en 1960, C.R. Blyth y D.W. Hutchinson publicaron tablas de los intervalos de confianza insesgados mas cortos según el método de Neyman. Estos intervalos fueron obtenidos por M.W. Eudey (1949) para distribuciones discretas, necesitan aleatorización para su construcción y poseen la propiedad de minimizar uniformemente la probabilidad de recubrir valores falsos de p.

Recientemente, Blyth y H.A.Still (1983) han obtenido unos nuevos intervalos de confianza al imponer 4 condiciones razonables a sus extremos. Dichas condiciones son las siguientes:

a) La región de confianza debe ser siempre un intervalo. Esta condición es equivalente al no decrecimiento de los extremos del intervalo de aceptación asociado.

b) Que sea invariante al remplazar S por n-S y p por q; es decir

$$p_I(n-S) = 1 - p_S(S) \quad , \text{ y} \quad p_S(n-S) = 1 - p_I(S)$$

c) Monotonía respecto a n; es decir

$$p_I(S;n+1) \quad p_I(S;n) \quad \text{y} \quad p_S(S;n+1) \quad p_S(S;n)$$

d) Monotonía respecto a S; es decir, para n fijo, el aumento de S debe implicar un aumento de p_I y p_S.

Basandonos en la aproximación normal de la binomial, se obtienen intervalos de confianza aproximados que, como ya dijimos, son suficientemente precisos cuando el parámetro n es superior a 30. En la practica se emplean dos tipos de intervalos aproximados, que han sido estudiados comparativamente por B.K. Ghosh (1979),y Blyth y Still. Ambos intervalos requieren de la corrección por continuidad para asegurar el nivel de confianza nominal.(Para garantizar que el nivel de riesgo real sea como mucho un 5% superior al nominal,sin corrección por continuidad, es necesario que n sea superior a 2000 para los niveles de confianza usuales).

El intervalo de confianza mas preciso es el que obtiene p_I y p_S como las soluciones admisibles de las ecuaciones:

$$\frac{S-0.5 - nN\cdot p_I}{\left\{n\cdot N\cdot p_I\cdot(1-p_I)\right\}^{1/2}} = z_{\alpha/2}$$

y

$$\frac{S+0.5 - nN\cdot p_S}{\left\{n\cdot N\cdot p_S\cdot(1-p_S)\right\}^{1/2}} = -z_{\alpha/2}$$

siendo $z_{\alpha/2}$ el valor de la normal tipificada que deja por encima de él una probabilidad igual a $\alpha/2$.

Blyth y Still corrigen p_I, cuando S vale 0 ó 1, dandole el valor 0 ó $1-(1-\alpha)^{1/n}$, respectivamente. Analogamente, p_S debe corregirse cuando S valga n y n-1, aplicando la condición de invarianza .

El otro intervalo es el mas habitualmente utilizado, obteniendose p_I y p_S de las ecuaciones anteriores al simplificar el denominador por $\{\hat{p}.\hat{q}/nN\}^{1/2}$; es decir

$$\hat{p} \pm z_{\alpha/2} \left\{ \sqrt{\frac{\hat{p}\cdot\hat{q}}{nN}} + \frac{1}{2n} \right\}$$

Cuando $\hat{p}$ es pequeño, Blyth y Still proponen sustituir nN por $nN - z^2_{\alpha/2} - 2\cdot z_{\alpha/2}/\sqrt{nN} - 1/nN$, para alcanzar el nivel de confianza nominal.

Cuando $\hat{p}$ es pequeño, esto es, inferior a 0.1, también pueden construirse intervalos aproximados a partir de la estimación por confianza del parámetro λ de una ley de Poisson que tenga una estimación puntual igual a $n\cdot\hat{p}$.

Finalmente, a estos intervalos aproximados debe

añadirse el que se deduce de la transformación arcoseno.(Dicha transformación y la expresión explicita del intervalo está expuesta en el punto C.1 de la sección 11). En general, para los niveles de confianza usuales este procedimiento es superior a los restantes, y, en el caso particular del nivel de confianza igual al 95%, puede decirse que se trata del procedimiento óptimo entre los de tipo aproximado.

Pasemos a considerar la estimación del parámetro n cuando el valor de p es conocido. Esta situación se presenta raramente, pero pueden encontrarse algunos ejemplo en la bibliografía. En estas condiciones, la estimación natural de n resulta ser

$$\tilde{n} = \frac{S}{Np}$$

Es insesgada y su varianza vale nq/Np , suponiendo que se admiten los valores no enteros de la estimación.

Por otra parte, teniendo en cuenta el caracter entero de n, su estimador de maxima verosimilitud es el que verifica que

$$L(n-1) < L(n) \quad y \quad L(n) > L(n+1) \quad ,$$

donde L(n) representa la verosimilitud de la muestra.

D. Feldman y M. Fox (1968) han probado que este estimador siempre existe y es único.

Si expresamos los números combinatorios en términos de la función Γ, este estimador será uno de los enteros adyacentes al valor n que sea la solución de

$$\sum_{j=o}^{M-1} \frac{f_j}{(\hat{n}-j)} = -N \cdot \ln(1-p)$$

donde f_j es el número de observaciones muestrales que superan a j y M la máxima observación ocurrida.

Recientemente, S. Blumenthal y R.C. Dahiya (1981) calculan la estimación entera de n por un procedimiento mejorado propuesto por Dahiya(1981) para la estimación de parametros enteros por maxima verosimilitud; este método no requiere calcular la derivada de la verosimilitud y, en este caso concreto, n es el entero comprendido entre los valores x-1 y x que verifican

$$L(x-1) = L(x) \quad .$$

Ignorando el caracter entero de $\hat{n}$, para grandes muestras se demuestra que

$$\text{var}(\hat{n}) \simeq \frac{1}{\sqrt{N}\left\{\sum_{j=1}^{n} b_j \cdot \sum_{i=o}^{j-1} \frac{1}{(n-i)^2}\right\}}$$

Feldman y Fox han demostrado la normalidad asintotica de n, así como la consistencia relativa para un exponente $\alpha > \frac{1}{2}$, cuando N permanece fijo. Blumenthal y Dahiya han ampliado estos resultados.

Feldman y Fox también han investigado otros estimadores, basandose en que x_i/q tiene media y varianza iguales y considerando la aproximación normal. No obstante, los estimadores así obtenidos, cuando n es grande, no producen una reducción apreciable de la varianza de ñ, que es un estimador mucho mas sencillo.

Intervalos de confianza para n pueden obtenerse aplicando el método de Clopper-Pearson; en este caso, los extremos del intervalo serian las soluciones de las ecuaciones

$$\sum_{j=o}^{S} \binom{Nn_I}{j} p^j \cdot (1-p)^{Nn_i - j} = 1 - \frac{1}{2}\alpha$$

y

$$\sum_{j=o}^{S} \binom{Nn_S}{j} p^j \cdot (1-p)^{Nn_S - j} = \frac{1}{2}\alpha$$

Desgraciadamente estas ecuaciones no van a tener casi nunca soluciones en valores enteros, de modo que en la practica los intervalos van a tener un nivel de confianza bastante superior al nominal.

Veamos, por último, la estimación conjunta de ambos parámetros. Este problema fue estudiado por Student(1919), R.A. Fisher (1941), J.B. Haldane (1941), P.G. Hoel(1947) y F. Binet (1953). Recientemente ha vuelto a ser estudiada por I. Olkin, J. Petkau y J. Zidek (1981), por S. Blumenthal y R. Dahiya (1981) y por D.F. DeRiggi (1983). Por otro lado, la estimación de n con p conocido o desconocido ha sido tratada desde el punto de vista bayesiano por N. Draper y I. Guttman (1971).

Si se aplica el método de los momentos, las estimaciones obtenidas $\tilde{n}$ y $\tilde{p}$ son las soluciones de

$$\bar{x} = \frac{1}{N} \sum_{j=1}^{N} x_j = \tilde{n}\cdot\tilde{p}$$

y

$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_{j=1}^{N} (x_j-\bar{x})^2 = \tilde{n}\cdot\tilde{p}\cdot\tilde{q}$$

o sea, sus valores son

$$\tilde{p} = 1 - \frac{\hat{\sigma}^2}{\bar{x}}$$

y

$$\tilde{n} = \frac{\bar{x}}{p} = \frac{\bar{x}^2}{\bar{x} - \hat{\sigma}^2}$$

Obsérvese que si $\bar{x}$ es menor que $\hat{\sigma}^2$ tanto la estimación de p como la de n resultan negativas, lo cual indicará que la distribución binomial no debe ser el modelo apropiado para tal situación.

Por otro lado, los estimadores de máxima verosimilitud de ambos, n^* y p^*, verifican las ecuaciones

$$n^* p^* = \bar{x}$$

$$\sum_{j=0}^{M-1} \frac{f_j}{(n^* - j)} = - N \ln \left(1 - \frac{\bar{x}}{n^*}\right)$$

La última ecuacion es exactamente la correspondiente a la estimación máximo verosimil de n cuando p es conocido, en la que éste se ha sustituido por su estimación $\bar{x}/n^*$.

Si el tamaño de la muestra es grande, la va-

rianza de n^* es aproximadamente,

$$\frac{n}{N\left\{\sum_{j=2}^{n}(p/q)\frac{j\,(j-1)!}{j(N-1)_{(j-1)}}\right\}}$$

Fisher demuestra que la eficiencia asintótica de $\tilde{n}$ respecto de n^* vale

$$\left\{1 + 2\sum_{j=1}^{n}(p/q)^{j}\frac{j!}{(j+1)(N-2)_{(j-2)}}\right\}^{-1}$$

(Recordamos que la notación $x_{(j)}$ significa $x(x-1)\ldots(x-j+1)$).

Finalmente, Binet señala que la diferencia entre $\tilde{n}$ y n^* es despreciable si se cumplen algunas de estas desigualdades

a) $n > 21$

b) $n > 2$ y $p < \frac{1}{21}$

c) $n > 2$ y $\frac{q}{p}(N-2) > 20$

Además, puntualiza que tanto $\tilde{n}$ como n^* pueden mejorarse considerando el siguiente hecho: n tiene que ser un número entero al menos igual que el máximo M de las observaciones de la muestra. Sugiere entonces tomar el entero más

próximo a la estimación obtenida siempre que sea superior a M y, en caso contrario, tomar M.

Estos dos estimadores adolecen del grave inconveniente de ser muy sensibles a las variaciones muestrales cuando la varianza muestral está muy proxima a la media muestral; pequeñas cambios en las observaciones muestrales dan lugar a grandes variaciones en las estimaciones. Olkin, Petkau y Zidek han propuesto versiones estabilizadas de dichos estimadores.

Estos autores han establecido, mas o menos arbitrariamente, que una muestra conduce a estimaciones inestables, para tamaños muestrales moderados, cuando la media observada es inferior a $(1+\frac{\sqrt{2}}{2})$ veces la varianza muestral. Dichas condiciones se presentan principalmente cuando n es grande y p inferior a 0.4. Desgraciadamente, este criterio tiene el inconveniente de no depender del tamaño muestral.

La versión estable del estimador obtenido por el método de los momentos es la siguiente

$$\tilde{n}_E = \max\{\hat{\sigma}^2 \cdot \frac{\gamma^2}{(\gamma-1)} , M\}$$

siendo $\gamma = \max\{\frac{M-\bar{x}}{\hat{\sigma}^2} , 1+\sqrt{2}\}$.

Este estimador se ha construido basandose en que el desplazamiento de todas las observaciones en una cantidad fija aumenta la media en esa cantidad, mientras que la varianza permanece inalterada.

La versión estable del estimador de maxima verosimilitud es:

$$n_E^{\star} = M + \frac{N-1}{N}(M - M')$$

siendo M' el valor maximo de la muestra que resulta de eliminar M. Este estimador ha sido construido por la técnica del "Jacknife" en una sola etapa.

Estos autores han comparado las caracteristicas de cada estimador simulando muestras con valores de p, n y N elegidos aleatoriamente sobre los intervalos (0,1), [1,100] y [3,22], obteniendo los siguientes resultados:

a) Los casos inestables se presentan aproximadamente el 30% de la veces.

b) Aproximadamente el 70% de las veces $\tilde{n}_E$ cae mas cerca del verdadero parámetro que $n_E^{\star}$. Los empates representan solo un 8% de los casos.

c) Por construcción el estimador n_E^* es esencialmente sesgado; mediante la técnica del "Jacknife" se logra eliminar un sesgo del orden de $1/N$. Empiricamente se comprueba que los estimadores "Jacknife" de mayor número de etapas están practicamente dominados por n_E^*.

Blumenthal y Dahiya han considerado el estimador de maxima verosimilitud generalizado que se deduce al admitir que el parámetro p sigue una distribución beta, demostrando su consistencia respecto al tamaño muestral cuando el parámetro n permanece fijo y su inconsistencia respecto a n para un tamaño muestral fijo. Estos autores han estudiado también el comportamiento asintotico de este estimador y del obtenido por el método de los momentos.

Por otro lado, DeRiggi ha resuelto afirmativamente el problema pendiente de la unicidad de la estimación conjunta por maxima verosimilitud (n^*,p^*). Esta cuestión fue planteada inicialmente por Anscombe en 1950 para la distribución binomial negativa, extendiendose luego, de manera natural, a la binomial. En el caso particular de la binomial negativa, la respuesta afirmativa también ha sido obtenida recientemente por B.Levin y J.Reeds (1977) y J.Bonitzer (1978).

Por ultimo vamos a completar esta sección tratando brevemente la estimación de n y p en la distribución binomial truncada. Esta distribución se presentaría cuando se modificara el 2º modelo del capitulo siguiente para englobar una situación basada en la Fisica Experimental.

Se entiede por una distribución binomial truncada en l_1 y l_2, aquella que se obtiene normalizando la probabilidad distribuida entre l_1 y l_2, ambos inclusive. En el caso particular de que $l_1=1$ y $l_2=n$, la distribución obtenida se suele denominar binomial positiva. Esta distribución se encuentra frecuentemente en algunos campos, ya que las realizaciones nulas no suelen ser observables.

Cuando n es conocido, D.Finney(1949) han mostrado como calcular el estimador de maxima verosimilitud de p y D.Newell (1965) lo ha aplicado a un ejemplo concreto. S. Shah (1966) ha construido otro estimador basandose en los 3 primeros momentos muestrales respecto al origen. Su estimador es

$$\tilde{p} = \frac{\mu_3' - (l_1+l_2+1)\mu_2' + l_1 \cdot (l_2+1)\mu_1'}{(n-2)\mu_2' + \{(n-1)(n-l_2-l_1) - n(n-2)\}\mu_1' + nl_2(l_1-1)}$$

En el caso particular de $l_1=1$ y $l_2=n-1$, la eficiencia asin-

totica relativa de este estimador respecto al de maxima verosimilitud supera siempre el 90%.

En el caso concreto de la binomial positiva, el estimador de maxima verosimilitud puede expresarse como la solución de la ecuación

$$\frac{n \cdot \hat{p}}{1 - \hat{q}^n} = \bar{x} \quad .$$

N.Mantel (1951) ha propuesto otro estimador para p

$$\tilde{p} = \frac{\bar{x} - f_1}{n - f_1}$$

siendo f_1 la frecuencia relativa de la observación 1 en la muestra. J.Gart (1968) ha estudiado su eficiencia asintotica relativa frente al estimador anterior, encontrando que es generalmente alta. D. Thomas y Gart (1968) han realizado un estudio completo de las propiedades de ambos estimadores en muestras pequeñas, concluyendo que $\tilde{p}$ tiene una precisión comparable a $\hat{p}$, pero menor sesgo que éste. De todas formas, para los usos practicos recomiendan dos correcciones de $\tilde{p}$. Estos estimadores y otros detalles pueden verse en primer lugar en el libro de Johnson y Kotz (1969).

Cuando el parámetro n es desconocido, Blumenthal

y Dahiya (1981) han estudiado la estimación conjunta por máxima verosimilitud de ambos parámetros en el caso de la binomial positiva. Sus propiedades son generalmente iguales a las que tiene el estimador conjunto en la binomial no truncada.

La estimación de n, con p conocido también ha sido estudiada por Blumenthal y Dahiya para la binomial positiva. Estos autores también han considerado el problema de la estimación del tamaño muestral correspondiente a la binomial no truncada a partir de una muestra de la binomial positiva.

I.10.- DISTRIBUCIONES RELACIONADAS

Aparte de la distribución normal, la binomial se halla relacionada con algunas otras distribuciones. Sobre la manera en que tienen lugar tales relaciones y sobre las consecuencias que de ello se deduce vamos a tratar en esta sección.

a) Distribución Beta:

Como ya mencionamos, la manera fundamental en que se relacionan estas dos distribuciones es a traves de sus correspondientes funciones de distribución, lo cual es consecuencia de la siguiente identidad:

$$\sum_{j=k}^{n} \binom{n}{j} \cdot p^j \cdot q^{n-j} = \frac{1}{\beta(k,n-k+1)} \int_0^p t^{k-1}(1-t)^{n-k} dt$$

establecida por integración por partes repetida. Se acostumbra notar el segundo miembro por $I_p(k,n-k+1)$ y recibe

el nombre de razón de la función beta incompleta.

Reescribiendo la identidad en términos de las respectivas funciones de distribución, se obtiene que

$$1 - B\,(K-1;n,p) = F_{\beta}(p;\ k,\ n-K+1)$$

o bien

$$B(K-1\ ) \quad = F_{\beta}(q;\ n-K,\ K+1)$$

de donde queda que la probabilidad puntual binomial vendría dada por

$$b_K = I_p(K,\ n-K+1) - I_p(K+1,\ n-K)$$

La aplicación principal de estas fórmulas es emplearlas para determinar probabilidades binomiales a partir de las tablas disponibles de la beta incompleta I(x).

Por otro lado, la distribución beta se suele utilizar de forma generalizada como función de densidad para la mixtura contínua de distribuciones binomiales cuando todas ellas tienen igual parámetro n. Desde el pun

to de vista bayesiano esto es importante ya que se argumenta que la beta es una distribución a priori natural para p. Sin embargo, no existen razones matemáticas fuertes que soporten esta naturalidad salvo la gran variabilidad de la familia beta y la sencillez con que resultan las probabilidades de los sucesos elementales de la mixtura obtenida así:

$$P[X_m = k] = \binom{n}{k} \frac{\Gamma(k+a)\ \Gamma(n-k+b)}{\Gamma(n+a+b)}$$

para k = 0, 1, ..., n, donde hemos supuesto que la distribución a priori del parámetro p era una $\beta(a,b)$.

b) Distribución F

La relación esta distribución se apoya en la dependencia funcional que existe entre la F y la distribución beta (Véase Johnson y Kotz (1970b)). Ya hablamos de una de las aplicaciones que de ella se desprenden: la construcción de intervalos de confianza para p a partir de los valores significativos de la F de Snedecor.

Otra aplicación, importante en nuestro contexto, es una de las mejores aproximaciones de las probabili

dades binomiales, la aproximación de Camp-Paulson. Esta se basa en la aproximación por la normal dada por E. Paulson (1942) para la distribución F.

c) Distribución binomial negativa:

Una relación en el mismo sentido que la existente con la distribución beta es la que liga a la binomial con esta distribución. A la definición de la binomial negativa le corresponde, en cierto modo, una interpretación opuesta a la que dimos para la binomial. Una realización de una binomial negativa, de parámetros r y p, es el número de pruebas de Bernoulli que no son éxito que ocurren antes del r-ésimo éxito; dicho de otra forma, r más el valor de la variable binomial negativa, da el número de ensayos necesarios para que el suceso éxito alcance una frecuencia absoluta igual a r. (Puede encontrase expresado así en V. Rohatgi (1976)).

Relaciones que existen entre ambas variables aleatorias son las siguientes:

- entre las funciones de masa de probabilidad

$$f_{BN(y;\, r,p)} = \frac{r}{r+y} \cdot b_r(\, r+y,\, p)$$

o bien

$$b_r(n,p) = \frac{n}{n-r} \cdot f_{BN}(r; n-r, q)$$

- entre las funciones de distribución

$$F_{BN}(k; r, p) = 1 - B_{r-1}(r + k, p)$$

La deducción de estas relaciones es sencilla basándose en la interpretación de ambas variables. Su principal utilidad práctica es naturalmente la determinación de las probabilidades binomiales negativas a partir de la binomial o sus aproximaciones.

Seguidamente vamos a tratar de las distribuciones ligadas con la binomial por medio de teoremas límites; concretamente, consideraremos las distribuciones de Poisson e Hipergeométrica. La distribución normal ya fue mencionada en una seccion anterior.

d) Distribución de Poisson:

Es bien sabido que la binomial converge hacia la distribución de Poisson, cuando el parámetro n crece

de manera que el valor medio np_n se mantiene igual a un número fijo λ ; esto precisamente significa que

$$\lim \; b_r\,(n, \frac{\lambda}{n}) = p_r(\lambda)$$

Por p_r designaremos de ahora en adelante las probabilidades de la Poisson, siempre que no haya lugar a confusión. Sobre esta convergencia se fundamentan aproximaciones a las probabilidades binomiales. En la próxima sección profundizaremos en esta aplicación.

Sin embargo, lo que no es tan conocido es el hecho de que tal convergencia es mucho más fuerte de lo que se cree habitualmente. En este sentido podemos señalar que se han producido los siguiente reforzamientos

I. $$\lim \sum_{r=0}^{\infty} | b_r(n, \frac{\lambda}{n}) - p_r(\lambda) | = 0$$

demostrado por L. Le Cam en 1960.

Si consideramos este resultado en su forma de teorema de aproximación y empleamos como medida de discrepancia la distancia de variación total entre ambas distribuciones, resulta que ha sido acotada por las

cantidades:

$$\text{II} \qquad \sum_{r=0}^{\infty} \left| b_r\left(n, \frac{\lambda}{n}\right) - p_r(\lambda) \right| \leq 2\,\frac{\lambda}{n} \qquad \text{si } n \geq 4$$

y

$$\text{III} \qquad \sum_{r=0}^{\infty} \left| b_r\left(n, \frac{\lambda}{n}\right) - p_r(\lambda) \right| \leq \frac{\lambda}{n} \sqrt{\frac{2}{1-\frac{\lambda}{n}}}$$

Estos resultados proceden, respectivamente, de J. Kerstan (1964) y de W. Vervaat (1970).

Finalmente, el mayor reforzamiento de esta convergencia se encuentra en el siguiente teorema:

IV

$$\lim_{n \to \infty} \sum_{r=0}^{\infty} h(r) \cdot \left| b_r\left(n, \frac{\lambda}{n}\right) - p_r(\lambda) \right| = 0$$

si y sólo sí la función h (r) es no negativa para todo r y verifica que

$$\sum_{r=0}^{\infty} h(r) \cdot p_r(\lambda) < \infty$$

Este teorema fue establecido por G. Simons y N. Johnson en el año 1971.

Relacionando este punto con el anterior, queremos hacer notar que estas dos distribuciones, la binomial y la Poisson junto con la binomial negativa, se hallan englobadas en el grupo III del sistema de Pearson discreto, y que son todas también miembros de la familia estudiada por Katz (Véase Johnson y Kotz (1969)). Estos hechos justifican más aún la interrelación entre ellas. Si comparamos las distribuciones entre sí, tenemos que situar a la distribución de Poisson como frontera, o caso límite, de las otras dos.

e) Distribución Hipergeométrica:

Ya mencionamos la propiedad que caracteriza a la binomial en la que interviene la distribución hipergeométrica. Ahora hablaremos de la convergencia de una hacia la otra, de la hipergeométrica hacia la binomial. Cuando los parámetros N y N_a de la hipergeométrica $H(N, N_a, n)$ tienden a infinito de modo que la proporción N_a/N se mantiene igual a p, entonces su función de masa de probabilidad converge hacia la correspondiente de la binomial $B(n,p)$; o sea

$$\lim_{\substack{N,N_a \to \infty \\ N_a/N=p}} \frac{\binom{N_a}{k}\cdot\binom{N-N_a}{n-k}}{\binom{N}{n}} = b_k(n,p)$$

La convergencia no es muy rápida para valores cercanos a 0 ó 1, pero para los restantes valores se puede considerar suficientemente buena para los fines prácticos. Cuando se sustituyen las probabilidades hipergeométricas reales por las binomiales se suele decir que se hace una aproximación a población infinita.

Cuando utilizamos la binomial para este fin, teniendo en cuenta que los papeles de N_a y n son equivalentes, disponemos de 4 posibilidades para aproximar la función de distribución $H_k(N, N_a, n)$:

I $\quad B_k(n, \frac{N_a}{N})$

II $\quad B_k(N_a, \frac{n}{N})$

III $\quad B(N_a - k - 1; N-n, \frac{N_a}{N})$

IV $\quad B(n - k - 1; N-N_a, \frac{n}{N})$

Recomendandose el empleo de aquella función de distribución binomial que corresponda al menor valor del argumento.(Véase

el libro de Patel, Kapadia y Owen (1976)).

Queremos aprovechar la ocasión para disculparnos por no habernos mantenido fieles a la notación establecida para función de distribución binomial; varias veces hemos sustituído $B_k(n, p)$ por $B(k; n, p)$. Lo justificamos porque ésta última resulta mucho más comoda cuando se manejan expresiones largas como argumento. Por esta razón y por otra conveniencia, utilizaremos más adelante preferentemente esta última.

Una aproximación mucho más precisa para $H_k(N, N_a, n)$ basándose en la binomial ha sido construída por Wise (1954). Se emplea la binomial de parámetros n y p_0, definido este último como

$$p_0 = \frac{N - \frac{1}{2} k}{N - N_a + 1/2}$$

Obsérvese que p_0 es función de k, lo que supone que al cambiar su valor es necesario considerar una nueva binomial aproximante. Una mayor aproximación se consigue multiplicando la cantidad obtenida por el factor $\{1 + C_1(k)/N\}$ donde

$$C_1(k) = \frac{k(1-2N_a/N) + n(N_a/N)^2 - (k-nN_a/N)^2}{2\cdot(1 - N_a/N)\cdot N_a/N}$$

Este resultado se debe a J.K. Ord (1968).

Para aproximar las probabilidades puntuales se pueden emplear las fórmulas debidas a I. Burr (1952) y (1973):

$$a)\ h_k(N, N_a, n) = b_k(n, \frac{N_a}{N})\left\{1 + \frac{N}{2N_A(N-N_A)} \cdot \right.$$

$$\left. \cdot\ [k-(k-\frac{nN_A}{N})^2\ ]\right\}$$

$$b)\ h_k(N, N_a, n) = b_k(n, \frac{N_a}{N})\left\{1 + \frac{1}{2N_A} \cdot [k-(k-\frac{nN_A}{N})^2\ ]\right\}$$

Si intercambiamos el papel de N_a y n, la primera fórmula quedará como

$$h_k(N, N_a, n) = b_k(N_a, \frac{n}{N})\cdot\left\{1 + \frac{N}{2n(N-n)}\cdot[k-(k-\frac{nN_a}{N})^2]\right\}$$

Análogamente sucedería con la otra. El error absoluto cometido por las primeras fórmulas es del orden de $1/N_a^2$.

Sin embargo, un buen procedimiento alternativo

y quizás mejor que el de Burr, de aproximación de las probabilidades puntuales mediante la binomial es el propuesto por P. Sandiford (1960). Los parámetros de la nueva binomial aproximante, n^* y p^* , se obtienen igualando la media y varianza de ambas variables, obligando luego que n^* se entero. En el artículo de Sandiford se dan varios ejemplos comparando la aproximación binomial ordinaria con esta nueva; también resulta ser un buen procedimiento para aproximar la función acumulativa de probabilidad. En este sentido, también se dan ejemplos comparándola con el procedimiento de Wise, que es mejor.

I. 11.- APROXIMACIONES A LAS PROBABILIDADES BINOMIALES

En esta sección vamos a revisar los principales métodos que se utilizan para aproximar la distibución binomial. Todas estas aproximaciones se pueden clasificar en tres grupos: a) aquellas basadas en la distribución Normal; b) aquellas otras basadas en la distribución de Poisson; c) las restantes que englobaremos con el socorrido: Otros métodos de aproximación.

Al estudiar cada una de las aproximaciones, nos centraremos especialmente en los principios en que se apoya y el grado de exactitud que proporciona respecto de la verdadera distribución. Enlazando con lo que acabamos de decir al final de la sección anterior, y como a lo largo de toda ésta nos vamos a referir a una distribución binomial genérica, por comodidad prescindiremos, siempre que sea posible, de hacer explícitos los parámetros n y p. De este modo, notaremos la función de distribución en el entero k por B(k), análogamente b(k) denotará la masa puntual de probabilidad en k; la correspondiente probabilidad sobre intervalos semiabiertos (j, k] será notada por B(j,k). Con objeto de incluir los intervalos que contengan el cero es necesario añadir a la binomial el valor -1. (Naturalmente se verificará que b(-1) = B(-1)=0). En correspondencia, las aproximaciones de las cantidades mencionadas se notarán, respectivamente, por $B_i^*(k)$, $b_i^*(k)$ y $B_i^*(j, k)$, en donde el subíndice hace referencia a un determinado método de aproximación.

Antes de analizar las diversas aproximaciones vamos a tratar de cómo podemos medir el grado de ajuste a la distribución aproximada que alcanza un método cualquiera. Con fines prácticos, se han propuesto muchas medidas

pero no se ha conseguido ninguna solución completamente satisfactoria. Así sucede que una cierta medida puede ser apropiada para determinadas condiciones,pero inadecuada para otras. En este orden de ideas, habrá que decir que una sola medida es insuficiente. En otras palabras, que una mejor solución consistirá en el empleo simultáneo de varias medidas que se complementaran. Evidentemente, lo óptimo sería conocer una función de error para todos los valores de la distribución, sin embargo, sería muy poco manejable salvo que se dispusiera de una forma analítica sencilla que la representara completamente. Pasemos seguidamente a repasar algunas de las medidas que se pueden utilizar.

La que nos viene a todos inmediatamente a la memoria es la medida empleada por Kolmogorov y Smirnov en su test de normalidad. Se trata simplemente de la distancia habitual del análisis funcional aplicada concretamente a la función de distribución. Particularizando para la binomial, vale el máximo valor absoluto de la diferencia entre su función de distribución y la de su aproximante sobre todos los puntos en que está definida. Admitiendo que los aproximantes también son variables discretas, tenemos que

$$D_i = \max_k \; |B(k) - B_i^*(k)|$$

cuando k varía desde -1 hasta n.

Explícitamente esta distancia ha sido empleada poco en relación con la distribución que nos ocupa. Entre las personas que la han utilizado podemos mencionar a Uspensky en 1937, y A. Mood y F. Graybill (1976). Sin embargo, se prefiere más otra, también mencionada en el texto de los últimos autores, que, en cierto modo, es generalización de D_i. Se le denomina máximo error y se define como el máximo error absoluto que se puede cometer al aproximar cualquier suma consecutiva de términos binomiales b(k). Precisando más, el error cometido al aproximar la probabilidad del intervalo (j, k] vale

$$E_i(j, k) = |B(j,k) - B_i^*(j,k)|$$

de manera que el máximo error es

$$M_i = \max_{j,k} E_i(j,k)$$

donde j y k han de variar entre -1 y n, ambos inclusive.

Evidentemente M_i es una generalización de D_i ya que si fijásemos j en -1 obtendríamos D_i. Además, son inmediatas las siguientes relaciones:

$$D_i \leq M_i \qquad y \qquad M_i \leq 2D_i$$

Aunque, según la definición, el cálculo de M_i parece exigir la comparación de las probabilidades de $\binom{n+2}{2}$ intervalos distintos, una pequeña reflexión nos muestra que no es necesario, puesto que

$$E_i(j,k) = |B(k) - B_i^*(k) - (B(j) - B_i^*(j))|$$

de modo que su valor máximo se alcanza cuando k y j son tales que alcanzan indistintamente el máximo y mínimo de

$$B(r) - B_i^*(r)$$

En la práctica, esto supone una ventaja computacional importante que reduce sustancialmente el tiempo y/o la memoria necesarias para determinar M_i. En otras palabras, hallar M_i es casi idéntico a determinar D_i.

Este criterio de proximidad ha sido utilizado por M. Raff (1956), cuyo artículo- siguiendo sus propias palabras- parece ser la primera investigación numérica comprensiva de la precisión comparativa de las aproximaciones más útiles de la binomial. Este artículo es un resumen de su Tesis Doctoral. En el año 1969 esta investigación fue ampliada en la misma dirección por F. Gebhardt, quien añadió más casos particulares y consideró nuevos métodos de aproximación.

Este criterio de ajuste ante un procedimiento dado, presenta la cualidad, desde el punto de vista práctico, de permitir la acotación del error absoluto y relativo de las estimaciones de las probabilidades grandes. Sin embargo, las estimaciones de las probabilidades más pequeñas que el máximo error resultan sin valor, puesto que no sabemos si ese procedimiento se ajusta bien, en términos relativos, en esa zona. Desde el punto de vista comparativo, parece razonable (suponiendo que las curvas de error de los distintos procedimientos son más o menos homotéticas respecto de algún punto) que cuanto menor sea el error tanto mejor será un método de aproximación sobre todas las posibles combinaciones de resultados de esa va-

riable binomial concreta. Sin embargo, nada nos asegura que tal suposición sea cierta, ya que será perfectamente posible que un procedimiento con un máximo error superior al de otro, fuese siempre mejor sobre los sucesos de las colas.

No obstante, considerando válida esa suposición dudosa, en lugar de este criterio podría ser más conveniente emplear el máximo error relativo de la función de distribución, puesto que tendría la ventaja de servir para acotar las estimaciones de las probabilidades tanto grandes como pequeñas. La definición habitual del error relativo sobre la cola izquierda es

$$r_i(-1,k) = \frac{E_i(-1,k)}{B(k)}$$

y sobre la cola derecha

$$r_i(k,n) = \frac{E_i(k,n)}{1 - B(k)}$$

Teniendo en cuenta que $E_i(-1,k)$ coincide con $E_i(k,n)$, para un mismo error absoluto tenemos dos errores relativos distintos, siendo mayor el que tiene menor denominador. Es evidente que para la búsqueda del máximo error relativo,

se utilice como definición del mismo la primera cantidad cuando B(k) es menor que 1/2, y la segunda en el otro caso. La función error relativo al variar k ha sido empleada por D. Peizer y J. Pratt (1968), quienes han construido la mejor aproximación normal para la binomial y sus distribuciones relacionadas. Los gráficos que presentan estos autores tienen el propósito de mostrar la gran precisión de su método y, a la vez, servir de corrección o mejora de las probabilidades estimadas. Tales gráficos vienen construídos sobre dos parámetros especiales y no utilizan a k como abscisa.

El error relativo, cuando se usa con fines de acotación de la probabilidad verdadera, suele definirse de manera diferente, empleando como denominador, en vez de ésta, la probabilidad aproximante. Cuando el error relativo es pequeño, ambas definiciones son casi iguales, pero cuando es grande, la diferencia es bastante significativa. El error relativo de los procedimientos de aproximación existentes adolece del grave inconveniente de crecer muchísimo en los extremos de la distribución. (Confírmese por ejemplo, en la tabla 2, pág. 196 del 1er volumen de Feller). Por este motivo, el máximo error relativo resulta

inutilizable. Sin embargo, se podría utilizar esta medida si se exigiera que la menor cola tuviese una probabilidad superior, digamos, a 0.01; de esta manera, podríamos tener un criterio más efectivo que M_i.

Por otro lado, Peizer y Pratt, al objeto de mantener en valores moderados el error relativo en las colas extremas, lo definen sustituyendo las probabilidades por sus logaritmos. Ellos mismos hacen notar que los papeles de $B_i^*(K)$ y $B(K)$ no son simétricos en la definición de esta cantidad (lo cual si sucede en la de M_i); con el fin de corregir este defecto y también para manejar números cercanos a la unidad, utilizan como medida de error el siguiente producto:

$$\frac{B(k)}{B_i^*(k)} \cdot \frac{1 - B_i^*(k)}{1 - B(k)}$$

Al escribirlo así, además de conseguir una cierta simetría entre las funciones de distribución verdadera y aproximante, se logra acentuar la diferencia respecto de la unidad conforme aumenta la gravedad del error. En el trabajo mencionado, bajo este producto, se compara la aproximación por ellos construída con otras cuatro aproximacio-

nes de tipo normal, palpándose claramente la superioridad de su nuevo procedimiento.

Otro criterio alternativo que ha sido emplea do con algún éxito, y que ya fue mencionado en la sección anterior al tratar de la distribución de Poisson, es la distancia de variación total entre las funciones de proba bilidad puntual. Siguiendo la notación de esta sección, su definición se concreta en

$$D_i' = \sum_{k=0}^{n} |b(k) - \overset{*}{b}_i(k)|$$

Obsérvese que esta misma cantidad también puede interpretarse como el valor medio del error relativo de la probabilidad puntual. Evidentemente, aquí hubiera sido perfectamente válido utilizar en vez del error absoluto, por ejemplo, el cuadrado del mismo, consiguiéndose así la definición de otro tipo de distancia de variación total.

Naturalmente existen muchos otros criterios válidos para cuantificar el grado de proximidad o ajuste que proporciona un método de aproximación dado. Para fina lizar esta revisión, vamos a mencionar el utilizado por

Freeman y Tukey (1950), cuyo artículo no hemos podido conseguir, pero aparece bosquejado en el artículo de M. Raff y probablemente sea el mismo que también emplean Peizer y Pratt. Para valorar la exactitud de un procedimiento, en lugar de emplear la función de distribución, se considera el valor de la variable normal tipificada cuya función de distribución es esa cantidad. Comparando, de cualquier forma válida, la abscisa de la probabilidad verdadera con la correspondiente a la aproximación se cuantifica el grado de proximidad alcanzado. Este método es perfectamente natural para los procedimientos normales, y no habría ninguna dificultad para adoptarlo a cualquier otro tipo de procedimiento. Una característica peculiar de este criterio es que la cuantificación del error se reduce para valores próximos al centro de la distribución y se agranda para los de las colas. En cierto sentido puede decirse que es una propiedad desfavorable porque se exagera el error donde de hecho ya es grande.

Pasamos seguidamente a tratar los principales métodos de aproximación de nuestra distribución. Consideremos primeramente los que se basan en la distribución Normal.

A. Aproximaciones normales.

Emplearemos aquí las notaciones habitualmente establecidas para la distribución Normal; es decir

$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-1/2\, x^2}$, para la densidad de probabilidad en el punto x, y

$\Phi(x) = \int_{-\infty}^{x} \varphi(u).du$, para la función de distribución en x.

También para cualquier resultado binomial k, notaremos por z(k), o simplemente z, a la abscisa de la normal unitaria que verifica la igualdad

$$\Phi(z) = B(k)$$

A.1.- Aproximación de Moivre-Laplace. La primera aproximación que vamos a tratar, B_1^*, es consecuencia inmediata del Teorema de Moivre-Laplace. Esta aproximación se puede presentar de varias maneras. En la más habitual B(j,k)

FACULTAD DE CIENCIAS BIBLIOTECA GRANADA

viene estimada por

$$B_1^*(j,k) = \Phi\left(\frac{k+1/2-np}{\sqrt{npq}}\right) - \Phi\left(\frac{j-1/2-np}{\sqrt{npq}}\right)$$

Para su deducción seguiremos a A. Renyi (1976). Empleando la fórmula de Stirling, la probabilidad puntual

$$b(k) = \frac{1}{\sqrt{npq}}\ \varphi(z_o).\left\{1+\frac{z_o^3 - z_o}{6\sqrt{npq}}(q-p)+ O(\frac{1}{n})\right\}$$

donde z_o es igual a $(k-np)/\sqrt{npq}$. La fórmula es válida siempre que z_o esté acotado. Si prescindimos de los dos últimos términos entre llaves, y consideramos la fórmula sumatoria de Euler-MacLaurin, se obtiene la expresión aproximante. La modificación de k y j por la adición y substracción de 1/2, suele denominarse corrección por continuidad, y puede probarse que sin ella las probabilidades resultantes serían globalmente inferiores a las reales. Otra posibilidad de corrección se tendrá en cuenta en el punto A.5.

Si se compara B(j,k), aplicando exactamente la fórmula sumatoria, con $B_1^*(j,k)$, se encuentra que el error tiene la forma

$$E_1(j,k) = \frac{q - p}{6\sqrt{npq}} \cdot \left\{ (1-z_o^2(k))\, e^{-\frac{z_o^2(k)}{2}} - \right.$$

$$\left. - (1-z_o^2(j))\, e^{-\frac{z_o^2(j)}{2}} \right\} + O\left(\frac{1}{n}\right)$$

Esto nos da pie inmediatamente para mejorar B_1 restandole esta cantidad, lo cual puede interpretarse co- una deducción simple del desarrollo de Gram-Charlier de 3 terminos, con lo cual se corrige la inflencia del sesgo (Vease el punto siguiente).

Como puede verse, $E_1(j,k)$ es practicamente proporcional a $(npq)^{-1/2}$. La misma pauta sigue el maximo error de esta aproximación. Mood y Graybill (1976) han acotado teoricamente el valor de M_1. Su resultado es

$$M_1 \leq 0.150 \cdot (npq)^{-1/2}$$

siempre que $npq > 25$. Empiricamente M. Raff ha mostrado que la constante de la cota puede reducirse hasta 0.140 y que la restricción es innecesaria. También ha encontrado la relación empirica $M_1(n,p) < 0.05$ siempre que $n \cdot p^{3/2} > 1.07$.

Del comportamiento de $M_1(n,p)$ se puede destacar lo siguiente: para n constante, $M_1(n,p)$ es simetrico respecto de $p=1/2$; al crecer p desde 0 hasta 1/2, disminuye su valor; para p fijo, M_1 decrece cuando n se incrementa. Si la media $n \cdot p$ se mantiene fija, M_1 crece conforme n se incrementa, alcanzando como limite el maximo error de la aproximación normal para la distribución de Poisson.

En la ultima sección de este Capitulo mencionaremos algunos teoremas recientes que muestran que las colas binomiales son mayores que las normales .

Otra versión de la aproximación normal es la que dio el propio Laplace, citada en el libro de Johnson y Kotz (1969). Su formulación es

$$B_1^*,(k) = \Phi(z_o) + \frac{1}{2\sqrt{npq}}\varphi(z_o)$$

Uspensky ha demostrado que el error de esta formula, medido por la distancia del analisis funcional, está acotado inferiormente, en el supuesto de que σ sea mayor que 5, por la cantidad

$$(1/5 + 1/4|q - p|)\frac{1}{\sigma^2} + e^{-3\sigma/2}$$

Por ultimo, para la estimación de las probabilidades puntuales $b(k)$ disponemos de tres posibilidades : $B_1^*(k-1,k)$, $B_{1'}^*(k-1,k)$, y la que empleamos al deducir B_1^* , o sea

$$b_{1''}^*(k) = \frac{1}{\sigma}\varphi(z_o)$$

Para este ultimo procedimiento, Y. Prohorov ha calculado la distancia de variación total, siendo su resultado que

$$D_1^* = \sum_{k=o}^{n} |b(k) - b_{1''}^*(k)| =$$

$$= \frac{1 + 4e^{-3/2}}{3} \cdot \frac{1}{\sqrt{npq}} \cdot \frac{1}{\sqrt{2\pi}} + O\left(\frac{1}{npq}\right)$$

A.2.- Aproximación de Gram-Charlier . Como ya hemos dicho, el procedimiento anterior puede mejorarse añadiendo un termino de ajuste que elimine la influencia del sesgo de la binomial. De este modo, obtenemos la segunda aproximación :

$$B_2^*(k) = B_1^*(k) - \frac{q - p}{6\cdot\sqrt{npq}} \cdot (z_o^2 - 1) \cdot \varphi(z_o)$$

Este metodo y el anterior se pueden englobar

como casos particulares de un procedimiento general para desarrollar en serie una distribución en función de otra. A este procedimiento se le conoce con el nombre de desarrollo de Gram-Charlier.(Para mas información sobre este y otros desarrollos en serie, puede verse el capitulo II del libro de J.K. Ord(1972)). Para cualquier distribución con todos sus momentos finitos, su densidad de probabilidad se puede expresar en función de la densidad normal, salvo conjuntos de medida nula, por la serie

$$f(x) = \varphi(x)\cdot\left\{1 + (\mu_2-1)\, H_2/2! + \mu_3 H_3/3! + \right.$$

$$\left. + (\mu_4 - 6\mu_2 + 3)\cdot H_4/4! + \ldots \right\}$$

o sea

$$f(x) = \varphi(x) \sum_{j=0}^{\infty} c_j \cdot H_j(x)/j!$$

donde $H_j(x)$ son los polinomios de Hermite, ortogonales bajo la función normal como peso; los coeficientes c_j vienen definidos a partir de la esperanza matematica de dichos polinomios respecto de la distribución desarrollada, como muestra la igualdad

$$c_j \cdot j! = E_f\left[H_j(x)\right]$$

Para aplicar este desarrollo a una distribución discreta definida sobre una red de amplitud Δ, es preciso que la relación de la amplitud a su rango de variación sea suficientemente pequeña. Alrededor de cada punto x_o de la red se aproxima la distribución original con probabilidad puntual p_{x_o} por una densidad continua definida de la siguiente manera :

$$f(x) = p_{x_o}/\Delta \quad \text{cuando} \quad x_o - \sqrt{2}\Delta \leq x \leq x_o + \sqrt{2}\Delta$$

Si aplicamos esto a la binomial tipificada y consideramos el desarrollo de solo tres terminos, obtenemos B_2 .

B_2 adolece del desagradable inconveniente de que para ciertos valores proporciona probabilidades negativas. Afortunadamente su cuantía nunca excede de 0.0001, siempre que la media de la distribución sea superior a 0.2. La pauta de M_2 es similar a la de M_1 , pero con errores mucho mas pequeños. Puede demostrarse que M_2 es inferior a $0.056/\sqrt{npq}$ para todos los valores de los parametros.

A.3.- Transformación raiz cuadrada. Los tres metodos siguientes se basan en la idea de utilizar una transformación sobre la variable original que la convierta en una distribución casi normal. En este metodo de aproximación emplearemos la transformación raiz cuadrada. La aplicación a la binomial se halla fundamentada primeramente en el trabajo de Fisher de 1925 sobre la χ^2 y después en el de Freeman y Tukey (1950) para la distribución beta ;este ultimo fue perfeccionado por Pinkhan (1957), siendo su aplicación a nuestra distribución

$$B^{*}_{3}(k) = \Phi(z_{1/2}(k))$$

con

$$z_{1/2}(k) = \sqrt{(4k-3)}q - \sqrt{(4(n-k)-1)p}$$

El subindice 1/2 hace alusión a la transformación raiz cuadrada. En general, se puede considerar la transformación potencial X^m ; en muchos casos, el resultado es similar al de esta. Por otra parte, el valor normal empleado por Freeman y Tukey es

$$z'_{1/2} = 2\left\{ \sqrt{(k+1)q} - \sqrt{(n-k)p} \right\}$$

Su patron de errores, considerando los graficos de Peizer

y Pratt, es completamente analogo al de A.1. Hemos observado también, que en los casos allí mostrados, generalmente a un error por defecto en la primera le corresponde uno por exceso de aproximadamente la misma cuantía en la segunda, y viceversa .

Para terminar este punto queremos notar que esta aproximación es el caso limite de la transformación arcoseno (estudiada en el apartado C) cuando los angulos pequeños son sustituidos por sus senos. Por esta razón, este metodo, a veces, recibe el nombre de transformación cordal. La aproximación arcoseno se presenta aquí como una transformación estabilizante de la varianza. En este mismo sentido, puede decirse que la transformación raiz cuadrada es precisamente una de este tipo para la distribución de Poisson .

A.4.- Aproximación de Camp-Paulson . Se basa en la transformación mediante la función raiz cubica. Fue construida inicialmente por Wilson y Hilferty (1931) para la distribución χ^2. Segun ellos, la variable transformada $(Y/\nu)^{1/3}$, siendo ν los grados de libertad de Y, se distribuye casi normalmente con media $1-2/9\nu$ y varianza $2/9\nu$. Posteriormente E. Paulson (1942) particularizó estos resulta-

dos para la distribución F de Snedecor. Su resultado se obtiene a partir de las igualdades

$$P\left[F \leqslant F_o\right] = P\left[\frac{Y_1^2/\nu_1}{Y_2^2/\nu_2} \leqslant F_o\right] =$$

$$= P\left[F^{1/3} \leqslant F_o^{1/3}\right] \simeq P\left[\frac{Z_1}{Z_2} \leqslant F_o^{1/3}\right]$$

Es decir, la distribución de la raiz cubica de una variable F es aproximadamente igual a la distribución del cociente de dos normales, cuyas media y varianza son, respectivamente, $1 - 2/9\nu_1$, $2/9\nu_1$, y $1 - 2/9\nu_2$, $2/9\nu_2$. Como desgraciadamente la distribución de tal cociente no es expresable de forma simple, Paulson emplea el resultado de Fieller (1932), segun el cual la función R del cociente $V = Y/X$, definida como

$$R = \frac{V \cdot m_x - m_y}{\sqrt{V^2 \sigma_x^2 + \sigma_y^2}}$$

es aproximadamente normal con media cero y varianza unidad, en el supuesto que la probabilidad de que el denominador X sea negativo es pequeña. Por tanto

$$z_F = \frac{(1 - \frac{2}{9\nu_2}) \cdot F_o^{1/3} - (1 - \frac{2}{9\nu_1})}{\sqrt{\frac{2}{9\nu_2} F_o^{1/3} + \frac{2}{9\nu_1}}}$$

será una abscisa de la normal unitaria cuya función de distribución aproximará a la de la F de Snedecor en el punto F_o.

B. Camp(1951) aplicó los resultados de Paulson a la distribución que nos ocupa. Para ello bastaria tener en cuenta lo expresado en las secciones 9 y 10 . Si queremos hallar aproximadamente B(k;n,p) , cuyo valor designaremos por α, podemos razonar indirectamente considerando el problema de estimación por confianza del parametro p, empleando para ello la formula en función de la F que lo resuelve. De la formula de los intervalos de confiza, tenemos

$$p = \frac{2(k+1)\cdot F_{1-\alpha}(2(k+1),2(n-k))}{2(n-k) + 2(k+1)\cdot F_{1-\alpha}}$$

Despejando $F_{1-\alpha}$, y sustituyendolo como F_o en la formula de z_F , obtenemos el valor normal z, cuya función de distribución vale aproximadamente $1-\alpha$. Por tanto, tenemos como aproximación de Camp-Paulson a

$$B_4^*(k) = 1 - \Phi(z) = \Phi(-z) \quad ,$$

donde z puede escribirse como el cociente $y_1/3\sqrt{y_2}$,

siendo

$$y_1 = c\left\{9 - 1/(n-k)\right\} + 1/(k+1) - 9$$

$$y_2 = c^2/(n-k) + 1/(k+1)$$

en donde c es igual a $\left\{(n-k)p/(k+1)q\right\}^{1/3}$.

Esta aproximación es asombrosamente presisa. Según M. Raff su maximo error está acotado por $0.007/\sqrt{npq}$, y nunca excederá de 0.0122 cualquiera que sean n y p . Raff califica este metodo como un procedimiento avanzado, en contraposición con A.1 y A.3 que los califica como simples. Basa esta distinción en el numero y dificultad de los calculos implicados. La otra aproximación que hemos visto, el metodo de Gram-Charlier, también se encasilla como un metodo avanzado. En ambas se necesita practicamente el mismo esfuerzo de calculo. Comparativamente, la aproximación de Camp-Paulson es globalmente superior a las anteriores. Siguiendo a Gebhardt, el error cometido por esta y la de Gram-Charlier, al determinar una probabilidad, es del orden de $1/n$. En las otras dos, es del orden de $1/\sqrt{n}$. Esto mismo puede confirmarse también en Pratt(1968) .

A.5.- Aproximación de Borges . R. Borges en un trabajo aparecido en 1970 ha investigado una clase general de aproximaciones. Se construyen a partir de cualquier función estrictamente creciente $\Psi(t)$ en el intervalo $[0,1]$, cuyas tres primeras derivadas sean continuas, y de una constante d no negativa. Aplicandola a la binomial, la variable obtenida

$$u_k = \Psi\left(\frac{k + d/2}{n + d}\right)$$

tiene una distribución asintoticamente normal con media $\Psi(p)$ y varianza $\Psi'(p)^2 \cdot pq/n$, como mostró J. Curtiss (1943).

Esta transformación general incluye como casos particulares a la aproximación A.1 , cuando $\Psi(t) = t$ y $d=0$, y la aproximación arcoseno (estudiada en el apartado C) , cuando $\Psi(t) = \text{arseno}\sqrt{t}$ y $d=0$. Borges ha investigado estas transformaciones cuando $d > 0$, buscando, en primer lugar, cuál es su valor optimo para una transformación dada; a continuacion, define como transformación optima a aquella que anula el primer coeficiente del desarrollo generico en serie del maximo error, y la encuentra explicitamente en una cierta transformación beta. A partir de ella

se construye el metodo de aproximación que lleva su nombre.

Borges ha demostrado que el maximo error de una transformación arbitraria es del orden de $1/\sqrt{n}$. Desarrollando $E_i(j,k)$ en serie de potencias de $n^{-1/2}$, se comprueba que su primer coeficiente es función del parametro p y de los valores de j y k; su maximo valor, M_i, queda minimizado, para $\Psi(t)$ dada, cuando se elige $d=1/3$. Considerando distintas transformaciones es posible comparar los errores absolutos que cometen. En este sentido se sabe que el error de la transformación arcoseno es la mitad, y de signo opuesto, que el de A.1 .

Por otra parte, la transformación optima, la unica que anula el coeficiente en $n^{-1/2}$, es la integral beta con exponentes $-1/3$ y $d=1/3$. Explicitamente, el transformado del entero k vale

$$y_k = (pq)^{-1/6}\cdot(n+1/3)\cdot\int_p^{(k+1/6)/(n+1/3)} [s\cdot(1-s)]^{-1/3}\cdot ds$$

Es interesante observar que tanto A.1 como la aproximación arcoseno también pueden expresarse como integrales beta con exponentes 0 y $-1/2$, respectivamente.

En otras palabras, estas aproximaciones podrian tratarse todas juntas, de una manera unificada. A este grupo también pertenece la transformación logaritmo, no incluida en esta memoria, pero que ha sido empleada también con fines aproximativos sobre la distribución binomial. A estas mismas observaciones llegó ,con anterioridad, G. Blom (1954), quien también determinó transformaciones optimas para la binomial desde un punto de vista distinto, y menciona explicitamente la transformación de Borges. Por estos motivos, debe considerarse el trabajo de Blom como un precedente importante de Borges. Seguiremos estudiando este articulo en el punto C.1.

Antes de dedicarnos a la aproximación de Borges, vamos a comentar brevemente el comportamiento del maximo error de A.1,cuando la modificamos tomando el valor optimo de d=1/3. Según Gebhardt, este refinamiento mejora A.1 reduciendo M_1 hasta alrededor de la mitad de su valor primitivo,(La reducción es incluso mayor para valores pequeños de n), excepto cuando p está proximo a 0.5 . En ese caso, se invierte la situación, siendo el valor primitivo de M_1 la mitad del nuevo maximo error. En estas condiciones especiales, resulta que M_1 es del orden de 1/n .

Volviendo a la aproximación de Borges, la de-

finición con fines practicos es la siguiente :

$$B_5^*(k) = \begin{cases} 0 & k = -1 \\ \Phi(y_{k+1/2}) & 0 \leqslant k < n \\ 1 & k = n \end{cases}$$

Naturalmente, la distnción que hemos hecho para $k=-1$ y $k=n$ es completamente valida para cualquier otro procedimiento de aproximación.

Como ya se ha dicho, su maximo error es del orden de $1/n$. Ha sido tabulado en algunos casos por Gebhardt. Generalmente es una aproximación de inferior calidad que la de Camp-Paulson: así, en las cercanias de $p=1/2$, $M_5 \simeq 3 \cdot M_4$; para valores pequeños de p ($np \leqslant 0.15$) , $M_5 \simeq 7 \cdot M_4$. Solamente es superior cuando $n \geqslant 20$ y $n \cdot p$ está en un entorno proximo a 2 ó 3; no obstante, en estas condiciones nunca M_4 sobrepasa del doble de M_5. A diferencia de la aproximación de Camp-Paulson, calificada por Raff como avanzada, Gebhardt considera a la presente como simple, siendo entonces la unica aproximación simple, cuyo error es globalmente del orden de $1/n$. El carácter simple debe entenderse cuando se dispone de tablas de la integral beta implicada. Gebhardt (1968), en un estudio dedicado unicamente a la aproximación de Borges, da

como apendice una de estas tablas en la que el limite superior de la integral varia desde 0.5 hasta 1, de milesima en milesima. Si no se dispusiese de tales tablas habria que calificarla de bastante avanzada, puesto que seria necesario resolver numericamente la integral beta.

A.6. Aproximación de Peizer-Pratt . Finalmente para terminar con las aproximaciones de tipo normal, vamos a ver la mejor aproximación que se ha construido para la binomial; mas exactamente, es la mejor para muchas distribuciones mutuamente relacionadas: binomial, binomial negativa, Pascal, beta, F de Snedecor, t de Student, y sus casos limite, Poisson, Gamma y Chi cuadrado. Se encuentra tratada en dos extensos articulos consecutivos escritos en la J.A.S.A. por D. Peizer y J. Pratt, y este ultimo, en el año 1968 .

Se presenta en dos versiones, siendo la segunda un refinamiento de la primera. Siguiendo la notación de los propios autores, la expresión de valor tipificado por cuya función de distribución vamos a aproximar B(k), o sea, $B^{*}_{6}(k) = \Phi(z_i)$, es

$$z_i = \frac{d_i}{|S - np|} \cdot \left\{ \frac{2}{1 + \sqrt{(6n)}} \cdot \left(S \cdot \ln \frac{S}{np} \right. \right. +$$

$$+ \; T \cdot \ln \frac{T}{n \cdot q} \;) \Bigg\}^{1/2} \quad , \quad i=1,2$$

o alternativamente

$$z_i = d_i \cdot \left\{ \frac{1 + q \cdot g(\frac{S}{np}) + p \cdot g(\frac{T}{n \cdot q})}{(n + \frac{1}{6}) \cdot p \cdot q} \right\}^{1/2}$$

siendo $S = k + 1/2$ y $T = n - S$.

g representa una función especial, que aparece tabulada en el 1er articulo, con la cual se facilita, según dicen, el computo de z_i con calculadora de sobremesa (entiendase ,hoy en dia, aquella que realiza solo las cuatro operaciones aritmeticas). La definición explicita de g es :

$$g(x) = \frac{1 - x^2 + 2x \cdot \ln x}{(1 - x^2)} \qquad x > 0, \; x \neq 1$$

$$g(0) = g(1) = 0$$

Es solo necesario tabularla para x comprendido entre 0 y 1, porque presenta la propiedad de que $g(x) = -g(1/x)$.

Los coeficientes d_i para la distribución binomial valen :

$$d_1 = k + 2/3 - (n + 1/3) \cdot p$$

$$d_2 = d_1 + 0.02 \cdot (\frac{q}{k+1} - \frac{p}{n-k} + \frac{q - 0.5}{n + 1})$$

Las sustituciones necesarias para las otras distribuciones mencionadas pueden encontrarse en el articulo original.

La motivación de la aproximación se discute en la sección 3 del articulo de Pratt. A grandes rasgos, podemos decir que se construye en dos etapas, partiendo de A.1 . En la primera etapa se modifica $z'_o = (k+1/2-np)/\sqrt{npq}$, para conseguir una mayor precisión cerca de la mediana. El resultado así obtenido es

$$z'_1 = \frac{d_1}{\sqrt{(n+1/6)pq}}$$

En la segunda etapa, basandose en el trabajo de Bahadur (1960), del que hablaremos en el apartado C, se corrige el comportamiento asintotico de las colas, multiplicando por un factor que transforma z'_1 en z_1. Cuando σ tiende a infinito, se demuestra que

$$\lim \frac{z}{z_o} = \lim \frac{z}{z'_1} = G^{1/2}$$

Por tanto, si definimos z_1 como $z'_1 \cdot G^{1/2}$, se consigue una aproximación precisa de z, puesto que el limite de z/z_1 sería la unidad.

La segunda versión se obtiene a partir de la siguiente generalización de la primera

$$z^* = (d_1 + c^*) \cdot \left\{ \frac{G}{(n + 1/6) \cdot pq} \right\}^{1/2}$$

determinando c^* de manera que la diferencia $z-z^*$ sea lo mas pequeña posible para las colas no muy extremas. La cantidad c^* es del orden de magnitud de $1/\sigma^2$.

Estos autores han estudiado detenidamente el error cometido por la aproximación, pero sin emplear una medida global. Proporcionan unos graficos para el error relativo de z_2, en función de una variables especiales, los cuales pueden usarse posteriormente para mejorar la estimación obtenida. Además establecen las siguientes relaciones generales sobre el error absoluto y relativo:

$$|B(k) - B_6^*(k)| \quad \begin{cases} 0.01 & \text{siempre} \\ 0.001 & \text{si } k \geq 1 \text{ y } n-k \geq 2 \end{cases}$$

$$\frac{|B(k) - B_6^*(k)|}{B(k)} \quad \begin{cases} 0.01 & \text{si } S,T \geq 2.5 \text{ y } 0.2 \leq \frac{Sq}{Tp} \leq 5 \\ 0.02 & \text{si } S,T \geq 1.5 \text{ y } 0.125 \leq \frac{Sq}{Tp} \leq 8 \\ 0.03 & \text{si } S,T \geq 1 \text{ y } 0.1 \leq \frac{Sq}{Tp} \leq 10 \end{cases}$$

Nosotros, por nuestra parte, hemos estudiado numericamente el comportamiento del maximo error M_6 de ambas versiones para algunos casos particulares, y lo hemos comparado con M_1 y M_4. Estos resultados están expuestos en el apendice II. Naturalmente, podrían considerarse otros casos particulares, simplemente cambiando los datos del programa que los proporciona.

En el trabajo de Pratt se estudia también el comportamiento asintotico de la aproximación bajo ciertas condiciones, cuando la varianza tiende a infinito. Los resultados quedan resumidos en la tabla 1 del citado articulo. En general, cuando z está acotada, todas las medidas de error que cita son del orden de $1/\sigma^3$. Para otras aproximaciones con que compara, se tiene que unicamente para A.4 el error es del mismo orden de magnitud que $1/\sigma^2$; para las restantes, esto es, A.1 y A.3(en general, X^m con cualquier exponente excepto 1/3), es del orden de $1/\sigma$. Excepcionalmente, el orden de magnitud puede ser superior para circunstancias especiales, tales como p valiendo 1/2. (Para mas detalles, consultese la tabla mencionada) .

B.- Aproximaciones por la distribución de Poisson .

Al igual que hicimos en el apartado anterior, empezaremos estableciendo la notación que emplearemos en relación con la distribución de Poisson. $P_k(\lambda)$ representará la función de distribución en el punto k de una Poisson de media λ, y $p_k(\lambda)$, la masa de probabilidad en dicho punto.

Antes de considerar los distintos metodos de aproximación basados en esta distribución, vamos a comentar algunas diferencias de índole practica que existen entre la Poisson y la Normal. Como todos sabemos, cualquiera que sea la distribución normal, todas sus probabilidades quedan determinadas a partir de la normal tipificada, cuyas probabilidades podemos conocer por tablas suficientemente precisas, o bien por calculo directo a partir de dos algoritmos de facil iteración y rapida convergencia. (Tales algoritmos son el desarrollo en serie generado por la integración recurrente de la función de densidad, utilizable para valores moderados, y el desarrollo de la razón de Mill, utilizable cuando los valores son mas extremos.)

Sin embargo, para la distribución de Poisson

sabemos que las probabilidades de una distribución concreta no son expresables a partir de ninguna otra de la familia; en la practica, deben ser obtenidas de las tablas disponibles, o bien por cálculo directo, cuya realización puede requerir un tiempo apreciable. En otras palabras, el empleo practico, cualquiera que sea su fin, de la distribución de Poisson requiere manejar una elevada cantidad de información o el empleo de bastante tiempo de cálculo, en comparación con la utilización de la Normal.

Esta diferencia se hace mas ostensible cuando su aplicación es la aproximacíon de la binomial. Por una parte, cuando se dispone de tablas, corrientemente será necesario realizar interpolaciones, salvo que el valor del parametro coincida casualmente con los valores tabulados. Por otra parte, el cálculo directo de las probabilidades implicadas es algoritmicamente muy semejante al cálculo exacto de éstas en la distribucion binomial. Concretamente, las respectivas funciones de distribución vienen dadas por las series :

$$P_k(\lambda) = \sum_{i=o}^{k} a_i \quad y \quad B_k(n,p) = \sum_{i=o}^{k} a'_i$$

en las que sus terminos generales verifican la ecuación de recurrencia

$$a_i = a_{i-1} \cdot \frac{\lambda}{i} \qquad y \qquad a'_i = a'_{i-1} \cdot \left(\frac{p}{q}\right) \cdot \frac{(n-i+1)}{i}$$

cuyo primer termino viene definido como :

$$a_o = e^{-\lambda} \quad , \quad a'_o = q^n \quad .$$

El algoritmo de la distribución de Poisson es ligeramente mas sencillo que el de la binomial, necesitando este ultimo de mas, añadir en cada iteración la redefinición del factor (n-i+1) y la multiplicación por él; todo esto, como mucho, va a suponer que el algoritmo de la binomial requiera el doble de tiempo .

Tras este inciso, pasemos ya a tratar los distintos metodos de aproximación disponibles. Comencemos por el que todos conocemos.

B.1.- Aproximación de Poisson . En este procedimiento empleamos una distribución con la misma media que la binomial que se desea aproximar, o sea tomamos $\lambda = np$. Por tanto, la aproximación de la función de distribución viene dada por :

$$B^*_7(k) = \begin{cases} 0 & \text{cuando} \quad k=-1 \\ P_k(np) & \text{cuando} \quad 0 \leq k < n \\ 1 & \text{cuando} \quad k = n \end{cases}$$

Su fundamento se encuentra en la convergencia hacia la distribución de Poisson, de la que hemos tratado ya en la sección anterior, donde también dimos acotaciones para la distancia de variación total entre ambas.

Respecto a su precisión, podemos señalar que es erronea la creencia generalizada de que esta aproximación es solo buena cuando p es pequeño y n grande. No obstante, empleando M_7 como criterio de proximidad, resulta que esta cantidad es practicamente independiente del parametro n, para p fijo. Disminuye su valor conforme p decrece. Según Gebhardt M_7 es de tamaño $p/4$.

B.2.- Aproximación de Gram-Charlier . Para mejorar la aproximación anterior se puede particularizar el desarrollo de Gram-Charlier a la ley de Poisson. Los polinomios ortogonales que así surgen se suelen conocer como polinomios de Charlier. La aproximación obtenida al añadir el primer termino corrector es

$$B_8^*(k) = P_k(np) + \frac{1}{2}\, p\cdot(k - np)\cdot p_k(np)$$

Comparandola con otros procedimientos, empleando para ello el maximo error, resulta superior a la aproximacíon de Gram-Charlier para la Normal (A.2) cuando el parame-

tro p es inferior a 0.15 ; cuando es un poco mas pequeño que 0.1, también supera a la aproximación de Camp-Paulson (A.4). En general, B^*_8 es bastante buena incluso para valores de p suficientemente grandes. En el trabajo de Raff se muestra en un diagrama n-p las regiones donde son mejores estas aproximaciones de tipo avanzado; asimismo se presenta otro diagrama similar en donde se comparan aproximaciones de tipo simple, concretamente A.1, B.1 y C.1 . Finalmente señalaremos, siguiendo a Gebhardt, que el maximo error M_8 es inferior a $p^2/4$.

B.3.- Aproximaciones de Bol'shev . Partiendo de la relación que liga la ley de Poisson con la distribución Gamma, similar a la que en el mismo sentido existe entre la binomial y la Beta, L. Bol'shev construye los dos siguientes procedimientos, basandose en la aproximación de las integrales beta por medio de integrales gamma :

$$B^*_9(k) = P_k(\lambda_1)$$

$$B^*_{10}(k) = P_k(\lambda_2)$$

cuyos parametros valen

$$\lambda_1 = \frac{p\cdot(2n - k)}{2 - p}$$

$$\lambda_2 = \frac{\lambda_1}{1 + \left[k(k+2) + k\lambda_1 - 2\lambda_1^2\right]/(6(2n-k)^2)}$$

Para k constante, al tender n a infinito, el termino de error para B_9^* es de la forma $O(1/(2n-k)^2)$ y $O(1/(2n-k)^4)$ para B_{10}^*, uniformemente sobre todos los valores de p. Sin embargo, de este analisis de error no se deduce ningún limite concreto para M_9 y M_{10}. Los resultados numericos muestran que M_9 varía solo lentamente con n, mientras que M_{10} decrece conforme se incrementa n; la proporción del decrecimiento es función de p. Para p menor o igual que 0.2, B_{10}^* es superior a todas las aproximaciones aqui consideradas, salvo las posibles excepciones de la de Peizer-Pratt y la de Molenaar, que veremos a continuación.

Sin embargo, enlazando con lo que hemos dicho al comienzo de este apartado, debe notarse que los parametros λ_1 y λ_2 son funciones de k y que tienen que ser calculados nuevamente al cambiar este. Mas concretamente, para aproximar las probabilidades B(j,k) según estos procedimientos, por calculo directo, es necesario evaluar j+k+2 terminos de leyes de Poisson, en vez de los k-j que realmente aparecen en la formula binomial; o sea, estos procedimientos solo pueden tener sentido cuando disponemos de tablas adecuadas de la Poisson.

B.4.- Aproximación de Molenaar . Según J. Ord (1972), un procedimiento mas preciso que B_9^* es la aproximación que construye Molenaar (1970), quien, en esa monografía, estudia de forma unificada aproximaciones a las distribuciones de Poisson, Binomial e Hipergeometrica.

Definiendo esta aproximación de la misma manera que la anterior, tenemos :

$$B_{11}^*(k) = P_k(\lambda_3)$$

con

$$\lambda_3 = \frac{n \cdot p\{(12-2p)n^2 - 7 \cdot k \cdot n\}}{(12 - 8p)n^2 - k(n-1)}$$

No disponemos de ninguna información sobre el comportamiento real del error. Por otro lado, tambien aquí son aplicables las mismas observaciones finales que hemos hecho en el punto anterior.

C.- Otros metodos de aproximación .

C.1.- Aproximación por la transformación arcoseno .

Este procedimiento se fundamenta en el principio de estabilización de la varianza, cuyo objetivo es encontrar transformaciones sobre una familia de varia-

bles aleatorias dependientes de un parametro, de manera que la varianza de la variable transformada sea independiente del mismo. Una ventaja practica que se consigue gracias a la estabilización de la varianza es poder utilizar los modelos del analisis de la varianza para muchas distribuciones a las que les estaría impedido en virtud de fallar la suposición basica de los mismos que exige que las observaciones procedan de poblaciones con igual varianza. No obstante, esta aplicación de la transformación fue criticada por Fisher, quien argumentaba que la interpretación de los resultados puede ser dificil, ya que el modelo se convierte en aditivo respecto de las variables transformadas, en vez de respecto de las primitivas.

Un metodo general para construir tales transformaciones ha sido establecido por Tippett (1934) y Beall (1942). Sea una variable aleatoria X, cuya función de probabilidad f(k) depende de parametros; consideremos una transformación continua sobre ella, $Y = \varphi(X)$. Desarrollando en serie de Taylor alrededor de la media de X, tenemos

$$\varphi(k) = \varphi(m) + (k-m)\cdot\varphi'(m) + \frac{(k-m)^2}{2!}\cdot\varphi''(m) + \ldots\ldots$$

Tomando esperanzas matematicas bajo la suposición de que las derivadas siguientes de la transformación y los correspondientes momentos centrales de X son tales que su producto es pequeño en comparación con los dos primeros terminos dominantes, la varianza de Y tendrá un valor aproximado de

$$V(Y) \simeq Var[X] \cdot \left[\varphi'(m)\right]^2$$

Por tanto, para estabilizar la varianza, la primera derivada de la transformación para el valor medio debe ser inversamente proporcional a la desviación tipica de la variable original. Esta ecuación es la base para la construcción de la mayoria de las transformaciones de estabilización de la varianza, aunque suele ser frecuente que posteriormente se retoque para aumentar el grado de estabilización. Las primeras transformaciones de este tipo datan del año 1936, construyendose inicialmente para la distribución de Poisson, y después para la binomial.

En el caso concreto de la ultima distribución, por el metodo de la estabilización de la varianza, se consigue que practicamente este momento sea independiente de los dos parametros. Según la ecuación que acabamos de dar, la transformación estabilizante debe verificar que :

$$\left.\frac{d\varphi}{dX}\right|_{X=np} = \frac{c}{\sqrt{n\cdot p\cdot(1-p)}} = \frac{c}{\sqrt{X/n}}\cdot\frac{1}{\sqrt{n-X}}$$

cuya integral, si la extendemos a todos los valores de X, puede expresarse como

$$\varphi(X) = c\sqrt{n}\cdot \operatorname{arcsen}\sqrt{\frac{X}{n}}$$

Esta transformación fue introducida por R. A. Fisher en 1922 para situaciones geneticas en las que cierta proporción variaba aleatoriamente de generación a generación. Observó que si las frecuencias muestrales n_{ik}/n se igualan a $\cos^2\psi_{ik}$, estos angulos ψ_i, medidos en radianes, se distribuyen casi normalmente con varianza cercana a $1/4n$. M. Bartlett (1936) propuso su utilización para estabilizar la varianza, cuando datos binomiales eran destinados a un analisis de varianza. La transformación ha sido estudiada en detalle por C. Eisenhart (1947) . El valor en que se estabiliza la varianza tiene la expresión

$$V(y) = \frac{c^2}{4}\cdot\left\{1 + \frac{1}{16npq}\cdot(8pq-3) + O(1/n^2)\right\}$$

Para hacer mas efectiva la transformación, el propio Bartlett propuso una modificación empírica para valores de p

cercanos a 0 o 1. Consiste en incrementar X en media unidad, si es inferior a n/2, y disminuirla en la misma cantidad, en caso contrario.

En esta linea, F. Anscombe (1948) consideró una forma mas general de la transformación :

$$\varphi(X) = c\cdot(n + d)^{1/2}\cdot \operatorname{arsen}\sqrt{\frac{X + a}{n + b}}$$

que nosotros escribiremos, siguiendo a J. Ord (1972), como $c\cdot\varphi(a,b,d)$. Según esta notación la versión de Bartlett se escribiría como $2\cdot\varphi(0,0,0)$. De esta amplia clase de transformaciones, Anscombe concluye que, para valores moderados de n, la mejor es aquella cuyos parametros son

$$a = \frac{b}{2} = \frac{3}{8} \quad , \quad d = \frac{1}{2}$$

Como c es un parametro de escala, fijandolo en 2 la varianza se estabiliza en la unidad con un error de magnitud $O(n^{-2})$.

Veamos otras versiones que se han propuesto de esta transformación. En primer lugar, M. Freeman y J. Tukey (1950), en el mismo articulo citado en el punto A.3, sugieren empiricamente

$$\varphi(0,1,1) \quad + \quad \varphi(1,1,1)$$

La desviación respecto de la varianza unidad, en esta ocasión, es del mismo orden de magnitud que 1/n. Este articulo nació de un estudio numerico asociado al denominado papel binomial de probabilidades, diseñado por F. Mosteller y J. Tukey, del que dan cuenta tratando de sus usos y utilidad en un articulo de J.A.S.A del año 1949. En dicho papel se representan, con fines comparativos, tanto el par de frecuencias esperadas teoricamente (np,nq) como el empirico (X,n-X) a traves de la transformación arcoseno; como normalmente la comparación entre esos pares supone la medición de pequeñas diferencias de angulos, dificilmente medibles con precisión, las sustituyen por sus correspondientes senos, facilmente medibles con una simple regla. A esta modificación se le denomina transformación cordal, y su expresión analitica es precisamente la aproximación por la raiz cuadrada, tratada en el punto A.3 .

Para la versión propuesta por Freeman y Tukey se disponen de tablas construidas por F. Mosteller y C. Youtz (1961) .

Una segunda versión puede obtenerse aplicando

el mismo procedimiento de Anscombe a la versión anterior. Este camino ha sido recorrido por N. L. Laubscher (1961), quien introduce 6 constantes que determina obligando a que la desviación de la varianza real respecto de la unidad sea del orden de $1/n^2$. Como dispone de tres grados de libertad, puede asignar a las tres primeras constantes el valor 0. Así se obtiene

$$\varphi(0,0,0) \quad + \quad \varphi(3/4,3/2,1)$$

En el trabajo de Laubscher puede observarse graficamente el comportamiento de la varianza de estas 4 versiones en función del parametro p cuando n vale 30; en estas condiciones, la mejor versión es siempre la de Anscombe, y la peor la de Bartlett. Según J.K. Ord (1972), la version de Anscombe debe emplearse cuando p está comprendido entre 0.1 y 0.9; en caso contrario, la versión a utilizar es la de Laubscher.

La transformación arcoseno tiene una importancia grande en relación con la distribución binomial. Dicha importancia se hace patente en el tabajo de G. Blom (1954). Su proposito es encontrar la mejor transformación para tres distribuciones bajo ciertos problemas y criterios. Este autor opina que los problemas sobre una distribución pueden

reducirse a dos : Determinar la función de distribución para unos parametros dados, y estimar con un cierto grado de confianza algún parametro cuando se conoce el valor de la función de distribución.

Para hallar las mejores transformaciones sobre la binomial que resuelvan esos problemas, construye dos desarrollos en serie sobre una transformación cualquiera $\psi(p^*)$, donde p^* viene definido como $X/(n+1)$. La elección de esta proporción corregida ya fue empleada,en otro sentido, por Feller (1945); en el presente contexto, se ve sugerida por la interrelación existente entre la binomial y la Z de Fisher (la mitad del logaritmo de la F de Snedecor), que posibilita la obtención de los desarrollos en serie deseados a partir del conocido desarrollo de Cornish -Fisher para la distribución Z (vease Kendall y Stuart, vol. 1, (1969), pag. 380). Actuando de modo analogo sobre una transformación cualquiera,se consigue obtener su desarrollo en serie de potencias, expresado en dos maneras diferentes según sea p^* conocido, o lo sea p :

$$\psi(p) = \psi(p^*) + a_1(n+1)^{-1/2} + a_2(n+1)^{-1} + a_3(n+1)^{-3/2} + \ldots$$

$$\psi(p^*) = \psi(p) + b_1(n+1)^{-1/2} + b_2(n+1)^{-1} + b_3(n+1)^{-3/2} + \ldots$$

En el primer desarrollo, los coeficientes solo dependen de p^* y z (recordemos, aquel valor normal definido como $B(X) = \Phi(z)$), pero no de p. Analogamente, los segundos coeficientes solo dependen de p y z. Evidentemente, el primer desarrollo se utilizará en los problemas de estimación confidencial, y el segundo en las aproximaciones de la función de distribución .

A la vista de estos desarrollos, teniendo en cuenta la rapida convergencia de las sucesivas potencias de $(n+1)^{-1/2}$, sobre todo a partir de n moderado, Blom propone como mejor transformación en un entorno de un determinado z, aquella cuyo coeficiente en $(n+1)$ sea lo mas pequeño posible. Dicho coeficiente, en cada desarrollo, tiene el valor

$$a_2 = \frac{1}{2}zp^*q^*\psi''(p^*) + \left(1 - \frac{z^2+2}{3z^2}\right)\cdot\left(\frac{1}{p^*} - \frac{1}{q^*}\right)\cdot\psi'(p^*)$$

$$b_2 = \frac{1}{2}z\cdot p\cdot q\ \psi''(p) + \frac{z^2+2}{3z^2}\cdot\left(\frac{1}{p} - \frac{1}{q}\right)\cdot\psi'(p)$$

Al igualar a_2 a cero, remplazando p^* por p, se obtiene que la mejor transformación debe verificar la siguiente ecuación diferencial :

$$\frac{\Psi''(p)}{\Psi'(p)} = \alpha_1(z)\cdot\left(\frac{1}{q} - \frac{1}{p}\right)$$

donde $\alpha_1(z)$ representa a $\left(1 - \frac{z^2+2}{3z^2}\right)$.

Razonando igualmente sobre el coeficiente b_2 se llega a la misma ecuación con un parametro $\alpha_2(z)$ definido como

$$\alpha_2(z) = \frac{z^2+2}{3z^2}$$

Observese que $\alpha_1(z) + \alpha_2(z) = 1$. Cuando z varía desde 0 a infinito, $\alpha_1(z)$ crece monotonamente desde $-\infty$ hasta 2/3. Para $z=2$, tenemos justamente que $\alpha_1(z)=\alpha_2(z)=1/2$.

En terminos del parametro α, la solución de la ecuación diferencial, salvo una constante multiplicativa, es

$$\Psi(p;\alpha) = \int_{p_o}^{p} x^{-\alpha}\cdot(1-x)^{-\alpha}\cdot dx$$

o sea, una integral beta de parametros $-\alpha$. Como soluciones particulares para ciertos valores de α, se obtienen las transformaciones arcoseno, logaritmo, y la transformación de Borges. Siguiendo las palabras de Blom, casi todas las transformaciones usadas hasta ahora en la literatura dedicada a la normalización de la binomial y sus distribuciones relacionadas, pueden ser desarrolladas desde un punto de partida

común.

La ultima transformación, de hecho, se encuentra explicitamente mencionada en el articulo de Blom, pero señala que su valor es unicamente de carárter teorico. Por esta razón y por la forma de su planteamiento general, como ya notamos en el punto A.5, el artículo aquí reseñado constituye un precedente destacable del trabajo de Borges.

En cuanto a la transformación arcoseno propiamente dicha, podemos observar que aparece como solución optima cuando α_1 y α_2 valen 1/2, lo que corresponde a un valor normal tipificado de z=2. Por una parte, ésto significa que , por medio de dicha transformación, se puede realizar una estimación optima por intervalos de p a un nivel de confianza cercano al 95%. Como prueba de esta afirmación, puede verse la tabla 1, pag 310 del artículo de Blom, en la que se comparan los verdaderos limites inferiores de intervalos con los calculados por distintos métodos aproximados. Por tanto, a las fórmulas dadas en la sección 1.9 debe añadirse la obtenida por Blom que, incluyendo el termino en $(n+1)^{-1}$ para conseguir una mayor precisión, es

$$2\cdot\operatorname{arsen}\sqrt{p_I} = 2\cdot\operatorname{arsen}\sqrt{p^*} - z\cdot(n+1)^{-1/2} +$$

$$+ \frac{z^2 - 4}{12} \cdot (q^* - p^*) \cdot (p^* \cdot q^*)^{-1/2} \cdot (n + 1)^{-1}$$

Por otra parte, si despejamos del segundo desarrollo la abscisa normal del primer termino, obtenemos

$$z = 2 \cdot (\text{arsen}\sqrt{p^*} - \text{arsen}\sqrt{p}) \cdot (n + 1)^{1/2} +$$

$$+ \frac{z^2 - 4}{12} \cdot (q - p) \cdot (p \cdot q)^{-1/2} \cdot (n + 1)^{-1/2} + \ldots.$$

lo cual demuestra que la transformación arcoseno es un procedimiento de gran precisión para aproximar $B(p^*)$, cuando su valor esté cercano a 0.975. A una conclusión similar habían llegado antes Freeman y Tukey, quienes afirman que es una aproximación extraordinariamente buena entre los puntos del 1% y el 5% de la distribución, y marcadamente buena, en general .

Pasemos ahora a considerar su empleo practico como aproximante de la distribución binomial. De la última identidad se deduce que la variable transformada tiene una distribución asintoticamente normal. En la práctica usaremos, por tanto, la ley normal para calcular de modo aproximado las probabilidades binomiales. Por esta razón, deberíamos haber incluido este procedimiento entre los de tipo nor

mal, maxime cuando su caso limite, la transformación raiz cuadrada, fue incluida en dicho grupo.No obstante, preferimos clasificarla aparte, puesto que consideramos mas importante el principio en que se apoya, la estabilización de la varianza .

Según acabamos de decir, el metodo de aproximación será de la siguiente forma :

$$B^{*}_{12}(k) = \begin{cases} 0 & k = -1 \\ \Phi(z_{12_j}) & 0 \leq k < n \\ 1 & k = n \end{cases}$$

donde z_{12_j} representa el valor tipificado obtenido por cualquiera de las versiones de la transformación, o sea

$$z_{12_j} = 2 \cdot \sqrt{n} \cdot (\varphi_j(X) - \text{arsen}\sqrt{p}) \quad .$$

Como subindice j emplearemos la inicial del autor que propuso la versión utilizada.

Sobre el comportamiento del error de la aproximación, disponemos del siguiente resultado teorico, deducido por Borges : el error absoluto de la versión de Bartlett

es, aproximadamente, la mitad del correspondiente a A.1 , o sea

$$E_{12_B}[-1,k] \simeq -1/2 \cdot E_1[-1,k]$$

Esta misma conclusión puede intuirse de la expresión que Blom da para z. En general, este error será del mismo orden de magnitud que $n^{-1/2}$.

Los estudios numericos sobre esta aproximación han sido realizados principalmente para dos versiones. En el trabajo de Raff, se consideró la versión de Bartlett con corrección por continuidad. El comportamiento del maximo error es similar al de A.1, con la excepción de tender a incrementarse cuando p tiende a cero. Al igual que M_1, M_{12_B} nunca excede de $0.140/\sqrt{npq}$.

La investigación numerica de Gebhardt se concretó sobre la aplicación a la transformación arcoseno de la corrección general propuesta por Borges para el valor optimo de d=1/3 :

$$z_{12_G} = 2 \cdot \sqrt{n+1/3} \cdot \operatorname{arsen} \sqrt{\frac{k + 2/3}{n + 1/3}} - \operatorname{arsen} \sqrt{p}$$

Este refinamiento trae consigo una mejora considerable.

Para valores medios pequeños de la distribución, el maximo error se reduce en una proporción comprendida entre la mitad y la decima parte de los valores de Raff. Cuando $p=0.5$, esta medida de error es del orden de $1/n$, y su valor es casi el doble de M_5, el maximo error de la aproximación de Borges.

C.2.- Aproximación de Bahadur . El punto de partida desde el cual R. Bahadur (1960) construye una serie de aproximaciones a la función de distribución binomial es una representación, aparentemente nueva, de la probabilidad de la cola superior por medio de la función hipergeometrica. Tal representación se consigue a partir de una interrelación entre la binomial y la binomial negativa, semejante a aquella que mencionamos en la sección 1.10 . Ambas surgen de la descripción de un mismo suceso por variables aleatorias distintas.

Consideremos una sucesión ilimitada de pruebas de Bernoulli, cuya probabilidad de exito valga p. Sea X la variable binomial que mide el numero de exitos que pueden ocurrir en las n primeras pruebas, y sea T la variable binomial negativa que denota el número minimo de pruebas necesarias para obtener $n-k+1$ fallos. Es inmediato

comprobar la equivalencia entre los sucesos $X \geq k$ y $T \geq n+1$. De la igualdad entre sus probabilidades obtenemos la representación deseada :

$$P[X \geq k] = B(k-1,n) = P[T \geq n+1] =$$

$$= \binom{n}{k} \cdot p^k \cdot q^{n-k+1} + \binom{n+1}{k+1} \cdot p^{k+1} \cdot q^{n-k+1} + \ldots.$$

$$= \binom{n}{k} \cdot p^k \cdot q^{n-k} \cdot q \cdot \left\{ 1 + \frac{n+1}{k+1} \cdot p + \frac{(n+1)(n+2)}{(k+1)(k+2)} \cdot p^2 + \ldots.. \right\}$$

$$= b(k) \cdot q \cdot F(n+1,1,k+1;p)$$

La interrelación mostrada en 1.lo también puede expresarse en función de la serie hipergeometrica, pero en esta ocasión solo con un número finito de terminos. La variable U que es necesario considerar es la que mide el número minimo de pruebas necesario para alcanzar un total de k exitos ; la descripción del suceso utilizado es ahora $U \leq n$. Si, por comodidad, notamos su probabilidad simplemente por B, entonces tendremos

$$B = b(k) \cdot \left(\frac{k}{n}\right) \cdot \left\{ 1 + \frac{m-1}{n-1} \cdot \frac{1}{q} + \right.$$

$$+\frac{(m-1)(m-2)}{(n-1)(n-2)}\cdot\frac{1}{q^2}+\cdots\cdots+\frac{(m-1)(m-2)\cdot\quad\cdot2\cdot1}{(n-1)(n-2)\cdot\quad\cdot(k+1)\cdot k}\cdot\frac{1}{q^{n-k}}\Big\}=$$

$$=\quad b(k)\cdot\frac{k}{n}\cdot F\left(-(m-1),1,-(n-1);\frac{1}{q}\right)$$

donde m viene definido por n-k+1 . Observese que estas representaciones unicamente son validas cuando $0<p<1$, y ademas la ultima tampoco sirve para $k=0$.

A partir de la representación de Bahadur siempre es posible determinar la función de distribución en cualquier k . Para ello se dispone de dos caminos : hallar la indirectamente a partir de la cola derecha; o bien, hallarla directamente considerando la binomial asociada que mide el número de fallos, intercambiando entonces p por q, y k por n-k .

Dada esta libertad, podemos seguir determinando esas cantidades, si restringimos a k a tomar solo valores a la derecha del valor medio. En dichas condiciones, $np\leq k\leq n$, resulta muy facil manejar la convergencia de la serie hipergeometrica F . Para ello es necesario ademas remplazar el argumento de la serie por otro nuevo definido por

$$x\quad=\quad\frac{(n+1)}{(k+1)}\cdot p$$

cuyo valor hemos asegurado que verifique $0<x<1$.
Ahora F vendrá expresada por la serie $\sum_{s=o}^{\infty} a_s x^s$, cuyos coeficientes tienen que definirse como

$$a_s = \begin{cases} 1 & s=0,1 \\ \prod_{t=1}^{s-1} \dfrac{\left(1+\frac{i}{n+1}\right)}{\left(1+\frac{i}{k+1}\right)} & s=2,3,\ldots. \end{cases}$$

Es inmediato comprobar que $0<a_s\leq 1$. Otra propiedad de estos coeficients que emplearemos mas adelante es la siguiente :

$$\frac{a_s}{a_t} \leq a_{s-t} \qquad s > t$$

La desigualdad siempre es estricta, excepto cuando t valga 0. Según lo dicho, la cola superior queda representada por

$$B = b(k)\cdot q\cdot \sum_{s=o}^{\infty} a_s x^s$$

Es inmediato ahora construir una sucesión creciente de aproximaciones hacia B sin mas que considerar la sucesión de las sumas parciales de F. O sea

$$B^*(j) = b(k)\cdot q\cdot \sum_{s=o}^{j} a_s x^s$$

Es fácil darse cuenta que la convergencia de $B^{*(j)}$ hacia B será tanto mas rapida cuanto mas pequeño sea x . En otras palabras, esta representación debe ser util para la investigación numerica y teórica de la probabílidades de las colas pequeñas .

Vamos a acotar el error relativo que se comete al emplear $B^{*(j)}$ como aproximante de B. Como en las condiciones elegidas B será inferior a 0.5, el error relativo vendrá definido por

$$r^{(j)} = \frac{B - B^{*(j)}}{B} = 1 - \frac{B^{*(j)}}{B}$$

de manera que una cota superior se obtiene mayorando B en función de $B^{*(j)}$. Eso se consigue aplicando la desigualdad entre los coeficientes de la serie y construyendo una serie geometrica :

$$B = b(k).q. \sum_{s=o}^{j} a_s x^s + \sum_{s=j+1}^{\infty} a_s x^s \leq$$

$$< B^{*(j)} + b(k).q \; a_{j+1}x^{j+1} \sum_{s=o}^{j} a_s x^s + a_{2j+2}x^{2j+2} \cdot \sum_{s=o}^{j} a_s x^s + .. <$$

$$< B^{*(j)} . \sum_{r=o} a_{j+1}^r \cdot x^r = \frac{B^{*(j)}}{1 - a_{j+1}x^{j+1}}$$

De donde llegamos a

$$0 < r^{(j)} < a_{j+1} \cdot x^{j+1} < a_j \cdot x^j$$

Es decir, el error relativo es siempre inferior al último sumando usado en el desarrollo incompleto de F. Este método solo es eficiente cuando x es apreciablemente menor que la unidad. Un método alternativo, util incluso para x cercano a 1, se obtiene a partir de la representación en fracción continua finita de la función hipergeometrica; en el punto siguiente trataremos de este procedimiento.

Analogamente, métodos paralelos, con propiedades también similares, pueden construirse basandose en la otra interrelación .

Por otro lado, también es facil obtener una cota superior rapida de B, puesto que la función hipergeometrica está mayorada por la serie geometrica $\sum_{s=0}^{\infty} x^s$. De este modo, obtenemos

$$B < b(k)\frac{q}{1-x} = b(k) \cdot \frac{q \cdot (k+1)}{(k+1)-(n+1) \cdot p}$$

Es a este mayorante a lo que le daremos en este punto el nombre de aproximación de Bahadur. Las formulas explicitas para estimar la función de distribución son las siguientes :

$$B^{*}_{13}(k) = \begin{cases} \dfrac{b(k)\cdot p\cdot(n-k+1)}{(n+1)\cdot p - k} & \text{cuando} \quad k \leqq np \\ 1 - \dfrac{b(k+1)\cdot q\cdot(k+2)}{(k+2)-(n+1)} & \text{cuando} \quad k \geqq np \end{cases}$$

Como ya se ha dicho de la representación de Bahadur, este procedimiento es adecuado para las colas pequeñas, o sea, para valores binomiales tipificados suficientemente lejanos del origen. El grado de exactitud conseguido se desprende del teorema 1 del artículo original, en donde se acota el error relativo superior e inferiormente. Dichas cotas son para las colas superiores :

$$c_i = \frac{(n-k)\cdot(n+1)\cdot p^2}{((k+1)-(n+1)p)((k+1)(k+2)+(n-k)p)}$$

$$c_s = c_i \cdot \frac{(k+1)(k+2)+(n-k)p}{(k+2)\left|(k+1)-(n+1)p\right|+(n-k)p}$$

en el supuesto de que el error relativo viniese definido como

$$r_{13} = \frac{B^{*}_{13} - B}{B}$$

Si nos fijamos unicamente en la cota superior, por tratarse de una aproximación por exceso, puede comprobarse que c_s es siempre inferior a $1/z_o^2$, lo que confirma su validez para las colas pequeñas.

Esta representación ha jugado un papel muy importante en la construcción de la aproximación de Peizer y Pratt , A.6 , utilizandose en la siguiente forma casi equivalente

$$B(k-1,n) = b(k) \cdot \frac{(k+1)}{(k+1)-(n+1)q} \cdot \frac{1}{1 + \theta/z_o'^2}$$

donde z_o' es el valor tipificado de k corregido por continuidad, o sea $z_o' = z_o + 1/2\,\sigma$; θ es un número real comprendido entre 0 y 1 .

C.3.- Aproximaciones por fracciones continuas .
Estos procedimientos se basan también en la representación de las colas binomiales por medio de la función hipergeometrica y además en la expansión de esta última como fracción continua. Dichos procedimientos se encuentran tratados en los artículos de Bahadur (1960), y Peizer y Pratt (1968), donde asimismo se dan las referencias a los originales e información complementaria interesante.

El dominio de aplicación de estos metodos, al igual que el anterior, se reduce casi exclusivamente a las colas extremas para las cuales, los otros procedimientos no proporcionan suficiente precisión . Para las restantes colas, esos otros procedimientos son mucho mas directos, aunque si se precisara una gran exactitud, este metodo la proporcionaria con ayuda de ordenadores.

El fundamento de la expresión como fracción continua de la hipergeometrica descansa sobre las propiedades típicas de esta función, especialmente en la transformación de Euler. La construcción inicial de esta representación se debe a Gauss. Su adaptación para la binomial puede realizarse directamente, empleando representaciones como la de Bahadur, o bien por intermedio de las relaciones existentes entre la hipergeometrica y la función beta incompleta. Al hacerlo de esta ultima manera, las aproximaciones obtenidas a partir de los desarrollos en fracciones continuas no son unicamente validos para la binomial, sino que pueden aplicarse a otras distribuciones, como evidentemente la beta cuando alguno de sus parametros no es un número entero. No obstante, a diferencia de lo que sucede con la binomial,la fracción continua de esos casos está compuesta por un número ilimitado de terminos.

Se dispone de dos representaciones muy parecidas, aunque una ligeramente mejor que la otra, para las colas inferiores de la binomial, $k \leq np$. Naturalmente , considerando la variable binomial asociada a los fallos, se obtendrían las representaciones de las colas superiores. La versión mas eficiente es

$$B(k) = b(k)/(1 + u_1/(v_1 + u_2/(v_2 + \ldots + u_{2k}/v_{2k})))$$

donde $u_1 = -k \cdot q/p$, $v_i = n-k+i$, $i=1,2,\ldots,2k$,

$$\left.\begin{aligned} u_{2j} &= j \cdot (n+j) \cdot q/p \\ u_{2j+1} &= -(n-k+j) \cdot (k-j) \cdot q/p \end{aligned}\right\} \quad j=1,2,\ldots,k$$

Esta representación fue dada por Markov el año 1900. Puede encontrarse también recogida en el libro clasico de J. Uspensky editado el año 1937.

Es inmediato construir una sucesión de aproximaciones truncando la fracción continua en cualquier punto. Llamaremos $B_{14}^{*(j)}(k)$ a la fracción que termina en u_j/v_j . Por simplicidad, prescindiremos de hacer explicito k. De esta manera tenemos

$$B_{14}^{*(1)} = \frac{b(k)}{1+u_1/v_1}$$

$$B_{14}^{*(2)} = b(k)/(1 + u_1/(v_1 + u_2/v_2))$$

$$\vdots \qquad\qquad \vdots$$

tal como están definidos los coeficientes, puede demostrarse que los aproximantes forman dos subsucesiones monotonas convergentes hacia B. Los correspondientes a un indice j de la forma 4m+1 y 4m+2, convergen decrecientemente, mientras que los restantes convergen crecientemente. El decrecimiento y el crecimiento es siempre estricto, acabando la sucesión en el termino j=2k .

El procedimiento ligeramente inferior, escrito de la misma manera, es

$$B(k) = b(k)\cdot p/(1 + u_1'/(v_1' + u_2'/(v_2' + \ldots + u_{2k+1}'/v_{2k+1}')))$$

donde sus coeficientes vienen definidos por

$$u_1' = -(n+1)\cdot q \quad , \quad v_i' = n-k+i \ , \ i=1,2,\ldots,2k+1$$

$$\left.\begin{aligned} u_{2j}' &= j\cdot(k+1-j)\cdot q \\ u_{2j+1} &= -(n-k+j)\cdot(n+1+j)\cdot q \end{aligned}\right\} \quad j=1,2,\ldots,k$$

Aparentemente esta relación fue dada por primera vez por L. Aroian (1941); también se deduce de la representación posterior de Bahadur, quien creia que era nueva .

Del mismo modo que hicimos con la anterior, definimos ahora la sucesión de aproximantes $B_{15}^{*(j)}$, cuyos elementos se subdividen en las mismas sucesiones monotonas. Además puede demostrarse que los aproximantes impares de ambas coinciden. Las dos razones que explican la inferioridad de $B_{15}^{*(j)}$ son las siguientes. En primer lugar, los aproximantes pares son menos precisos; es decir, siempre se verifica

$$|B_{14}^{*(2j)} - B| \leqq |B_{15}^{*(2j)} - B|$$

Esta relación fue conjeturada por Bahadur. En segundo lugar, B_{15}^{*} requiere un termino mas para llegar hasta B; la primera necesita $2k$ terminos y la ultima $2k+1$.

El comportamiento de estos aproximantes para las otras distribuciones, o sea cuando k no es entero, a partir de los terminos $2k+1$ para la primera, y $2k+2$ para la segunda, es siempre monotono convergente, siendo creciente o decreciente según cual sea la paridad de la parte entera de k.

A continuación, vamos a presentar la forma equivalente, dada por Bahadur, de la segunda representación aplicada a la cola superior, siguiendo otra notación habitual en fracciones continuas :

$$B(k-1,n) = b(k)\cdot q\cdot \left\{ \frac{1}{1-}\cdot\frac{d_1}{1+}\cdot\frac{c_1}{1-}\cdot\frac{d_2}{1+}\cdot \quad \ldots\ldots \quad \frac{c_{m-1}}{1-}\cdot d_m \right\}$$

cuyos coeficientes valen

$$d_r = \frac{(n+r)\cdot(k+r-1)}{(k+2r-2)\cdot(k+2r-1)}\cdot p$$

$$c_r = \frac{r\cdot(m-r)}{(k+2r-1)\cdot(k+2r)}\cdot p$$

con $m = n-k+1$, y $np \leq k$.

Ahora la sucesión de aproximantes viene construida así

$$B_{15}^{*(1)}(k-1,n) = b(k)\cdot q\cdot \left\{ \frac{1}{1-}d_1 \right\} = b(k)\cdot q/(1-d_1)$$

$$B_{15}^{*(2)}(k-1,n) = b(k)\cdot q\cdot \left\{ \frac{1}{1-}\frac{d_1}{1+}c_1 \right\} = b(k)\cdot q/(1-d_1/(1+c_1)$$

$$\vdots$$

Las relaciones entre los sucesivos aproximantes son consecuencia de que tanto d_r como c_r están siempre comprendidos entre 0 y 1 .

El calculo de los sucesivos aproximantes de una fracción continua puede realizarse facilmente de manera recurrente, como se describe en algunos textos de analisis numerico, o en libros como Abramowitz,M. y Stegun,I. (1972, pag. 19) y H. Wall (1948, pag 15). Si se expresa la fracción

continua de cada aproximante como un unico cociente, N_j/D_j ; su numerador y denominador pueden calcularse recurrentemente en función de los correspondientes a las dos fracciones anteriores y de los dos nuevos terminos que se añaden al determinar N_j/D_j . Las ecuaciones de recurrencia pueden escribirse en notación matricial de la siguiente manera :

$$\begin{bmatrix} N_j \\ D_j \end{bmatrix} = \begin{bmatrix} N_{j-1} & N_{j-2} \\ D_{j-1} & D_{j-2} \end{bmatrix} \times \begin{bmatrix} v_j \\ u_j \end{bmatrix}$$

donde debe partirse del convenio

$$N_{-1} = N_o = D_o = 1 \qquad y \qquad D_{-1} = 0$$

Pasemos a juzgar el grado de exactitud que se consigue con estos procedimientos. En este sentido, es destacable el hecho de que el primer aproximante coincide con la aproximación de Bahadur. Es decir, se cumple

$$B^*_{13} = B^{*(1)}_{14} = B^{*(1)}_{15}$$

No debe sorprendernos esta identidad, puesto que Bahadur empleó la expansión en fracción continua de la hipergeometrica, redefiniendola como serie, probablemente, de tal manera que al mayorarla por una serie geometrica, se obtuviese el primer aproximante. En este sentido, sería interesan-

te estudiar si es posible encontrar mayoraciones y minoraciones que den lugar a los sucesivos aproximantes por fracción continua. Además las acotaciones que Bahadur construye para el error relativo de su aproximación se basan en este tipo de expansión. Recordemos que dicho error siempre será inferior a $1/z_o^2$, siendo z_o el valor tipificado de k .

Teniendo en cuenta las relaciones entre los sucesivos aproximantes y lo que acabamos de decir, es claro que estos procedimientos son buenos para las colas lejanas. De hecho, en tales condiciones, la sucesión de aproximantes converge rapidamente. Por otra parte, una ventaja adicional de estos procedimientos es que, desde un principio se dispone de una cota superior e inferior precisas de la verdadera probabilidad; un intervalo suficientemente bueno se halla sin ninguna dificultad a partir de los 4 primeros aproximantes con la ayuda de una calculadora ordinaria.

No obstante, es conveniente comparar estos metodos con el cálculo directo de la serie binomial implicada. La determinación exacta de la serie binomial requiere unicamente la suma k terminos de fácil iteración, mientras que estos procedimientos necesitan, respectivamente, $2k$ ó

2k+1 terminos de mas difícil iteración. A grosso modo, podría decirse que en el mismo tiempo en que se calcula un nuevo aproximante por fracción continua, podrían añadirse por lo menos dos terminos a la serie binomial. Desgraciadamente el cálculo directo no proporciona, en principio, cotas superiores de la verdadera probabilidad, de manera que no se puede decidir adecuadamente cuando detener el cálculo, una vez alcanzada una suficiente exactitud. Sin embargo, esta limitación puede superarse combinando este método con la aproximación de Bahadur, de modo que después de cierto número de terminos, puede acotarse superiormente el resto de la serie. Además, este metodo mixto posee una gran precisión, puesto que el error relativo de la cota del resto disminuye conforme avanzamos en la serie, ya que corresponde a colas de valores cada vez mas alejados de la media.

I.12.- GENERALIZACION DE POISSON DE LA DISTRIBUCION BINOMIAL

A partir del modelo genérico del que surge la distribución binomial, se llega de manera inmediata al planteamiento de su generalización. Recordemos que la binomial se define como el número de veces que puede ocurrir un suceso, al que habitualmente denominamos "exito", al realizar n pruebas aleatorias independientes y equivalentes. Como todos sabemos, este tipo de pruebas recibe el nombre de ensayos de Bernoulli.

La generalización que hizo Poisson en 1837 se consigue sin mas que eliminar la exigencia de que todas las pruebas sean equivalentes, posibilitando así que la probabilidad de "exito" pueda variar de una prueba a otra. Por analogia con el termino anterior, este tipo de pruebas, estadisticamente independientes, recibe el nombre genérico de ensayos de Poisson. Con ayuda de este nuevo concepto, la variable binomial generalizada queda definida como el número de "exitos" que pueden ocurrir en n ensayos de Poisson. A menudo, a esta variable se le da también el nombre de distribución binomial de Poisson.

No es difícil encontrar situaciones reales que

FACULTAD DE CIENCIAS
BIBLIOTECA

se ajusten bien a esta distribución; valgan como ejemplos las siguientes :

a.- El número de blancos que consigue un tirador al disparar sobre una diana desde distancia variable. (Ejemplo del libro de Wadsworth y Bryan (1979)) .

b.- El total de individuos que poseen un cierto atributo en una poblacion estratificada al extraer una muestra compuesta por un individuo de cada estrato de la población. (Kendall y Stuart (1969)) .

c.- El número de partidas ganadas por un jugador cualquier noche en un casino.

Pasemos seguidamente a mostrar las principales características de esta distribución. Comencemos determinando la función característica. A partir de ella obtendremos facilmente la función de probabilidad puntual y los dos momentos mas importantes, media y varianza .

Denotemos por p_i , i=1,2,....,n , la probabilidad de que el suceso "exito" ocurra en cada prueba de Poisson. Trivialmente la función caracteristica del número de "exitos" en cada prueba individual vale $(q_i + p_i \cdot e^{it})$. Si notamos por X_g a la variable binomial generalizada, es cla

ro que su función caracteristica vale :

$$\Psi_{X_g}(t) = (q_1+p_1e^{it})\cdot(q_2+p_2e^{it})\cdot\ldots\ldots\cdot(q_n+p_ne^{it})$$

De esta expresión se deduce claramente que cada distribución binomial de Poisson viene completamente especificada por la n-upla $(p_1,p_2,\ldots,p_n)$, que suele denominarse vector de probabilidad de la distribución. No obstante, el lugar ocupado por cada componente en el vector de probabilidad es irrelevante, puesto que Ψ_{X_g} es una función simetrica de dichas cantidades. Igualmente, son funciones simetricas de p_i la función de probabilidad puntual y, en general, la esperanza matematica de una función arbitraria de X_g . Todo eso es consecuencia de que el orden en que se realizan las pruebas no influye de ningún modo sobre X_g. Con fines practicos, esta propiedad nos permitirá reordenar a voluntad las componentes del vector de probabilidad sin alterar esas funciones.

Por otro lado, es inmediato que se obtiene como caso particular la distribución binomial, cuando todas las probabilidades de "exito" son iguales. En estas condiciones bastan los dos parametros habituales para especificar el vector de probabilidad.

La función de probabilidad puntual se determina sin ninguna dificultad a partir de la función característica. Si, para mantener el paralelismo con la binomial, notamos la masa de probabilidad del punto k por $b_g(k)$, puede comprobarse que su valor, el coeficiente del término en $(e^{it})^k$ del desarrollo de $\Psi_{X_g}(t)$, vendrá dado por :

$$P[X_g=k] = b_g(k) = \sum_{\substack{i_j=0,1 \\ \sum_{j=1}^{n} i_j=k}} \prod_{j=1}^{n} p_j^{i_j}.(1-p_j)^{1-i_j}$$

donde cada sumando representa la probabilidad de que los k "exitos" se repartan en aquellas pruebas en que i_j vale la unidad.

Por otro lado, de la propia definición de X_g se deduce directamente que la media y la varianza de la distribución valen :

$$E[X_g] = \sum_{j=1}^{n} p_j = \mu \quad ,$$

$$V[X_g] = \sum_{j=1}^{n} p_j \cdot q_j = \sigma^2$$

Es interesante comparar esta última cantidad con el valor

que tomaría la varianza en otras distribuciones de la familia que con igual número de pruebas tuviesen el mismo valor medio μ. En este sentido, si llamamos p a la probabilidad de "exito" promedio a lo largo de los n ensayos de Poisson, es decir $p=\mu/n$, puede comprobarse que una manera alternativa de expresar $V[X_g]$ es la siguiente:

$$V[X_g] = n \cdot p \cdot q - \sum_{j=1}^{n} (p_j - p)^2 \quad ,$$

lo que significa que la varianza de una distribución binomial generalizada con media dada es tanto menor cuanto mayor sea la dispersión entre las probabilidades de "exito" de las pruebas individuales. Dicho de otro modo, si notamos por $\mathcal{T}_\mu$ o $\mathcal{T}_{np}$ el conjunto de vectores de probabilidad de longitud n que generan distribuciones con valor medio $n \cdot p$, resulta que si $\bar{p}^{(1)}$ y $\bar{p}^{(2)}$ son vectores de $\mathcal{T}_{np}$ tales que $\Sigma p_j^{(1)2} \leqq \Sigma p_j^{(2)2}$, sus distribuciones asociadas verifican

$$V[X_g^{(1)}] \geqq V[X_g^{(2)}] \quad .$$

De aqui concluimos que la binomial clásica es la única distribución binomial generalizada de $\mathcal{T}_{np}$ que maximiza la varianza. Considerando a este momento como medida de dispersión de una distribución probabilistica, resulta a primera

vista sorprendente que la dispersión sea maxima cuando todas las pruebas son homogeneas. Sin embargo, esa sorpresa desaparece cuando suponemos que $n.p$ es entero y consideramos el caso extremo en que todas las probabilidades p_j son solo iguales a cero y a uno, con ambos valores representados, puesto que su variable asociada degenera en un punto, mientras que la correspondiente binomial toma todos los posibles valores con probabilidades no nulas.

Estos dos resultados que hemos obtenido para el caso particular de la esperanza matematica de la función $(X_g - E[X_g])^2$ han sido reproducidos, en cierta medida, para la esperanza matematica de una función arbitraria $g(X_g)$. Así, por ejemplo, tenemos que la propiedad de extremo respecto de la dispersión de la binomial $B(n,p)$ sobre todas las distribuciones generadas por $\mathcal{T}_{n.p}$ puede expresarse de muy variadas formas.

Por otro lado, al igual que la varianza respeta la ordenación que hemos definido sobre $\mathcal{T}_{n.p}$, se puede comprobar que cierta clase de funciones g conservan un orden parcial definido en ese conjunto; es decir, si llamamos

$$f(\vec{p}) = E[g(X_g)] \quad ,$$

veremos que si $\vec{p}_1, \vec{p}_2 \in \mathcal{T}_\mu$ tales que $\vec{p}_1 \preceq \vec{p}_2$, entonces se verificará que

$$f(\vec{p}_1) \leq f(\vec{p}_2)$$

siempre que g(x) sea una función concava. En otras palabras, si consideramos definido sobre $\mathcal{T}_\mu$ el orden $\preceq$, la esperanza matematica de una función concava presenta un comportamiento monótono no decreciente.

A continuación, junto con estas generalizaciones, vamos a dar los principales resultados teoricos que se conocen sobre la distribución binomial generalizada. Sobre trabajos de carácter aplicado en los que se haya empleado dicha distribución pueden encontrarse algunas referencias en el libro de G. Patil y S. Joshi (1968) .

Principales Propiedades y Resultados sobre la Distribución Binomial de Poisson .

Vamos a recoger aquí practicamente toda la investigación realizada sobre esta distribución. Aparte de los estudios relativos a $f(\vec{p})$, trataremos de estos otros temas : Forma de la distribución; Localización de la moda y mediana; Generalización del teorema de Bernoulli; y, Convergencia hacia la distribución de Poisson.

A) Extremos y Comportamiento Monótono de $f(\vec{p})$.

La investigación sobre la esperanza matematica de una función arbitraria de la variable binomial de Poisson ha sido llevada a cabo por W. Hoeffding (1956) y por L. Gleser (1975). El primer trabajo está dedicado a la busqueda de los extremos; el otro al comportamiento monótono de esa función. A continuación, vamos a hacer una revisión del contenido de estos trabajos.

Para la busqueda de aquellos vectores $\vec{p}$ de $\mathcal{P}_\mu$ que proporcionan los maximos y minimos de la esperanza matematica de la función $g(X_g)$, Hoeffding utiliza un procedimiento muy ingenioso desarrollado a partir de la expresión que $f(\vec{p})$ tiene en función de cualesquiera componentes del vector $\vec{p}$ que se quieran separar del mismo. Por ser $f(\vec{p})$ una función simetrica respecto a las permutaciones de las componentes de p, podemos considerar que las componentes suprimidas son las m primeras; el vector de las (n-m) restantes componentes viene notado por $\vec{p}^{1,2,..,m}$. Basandose en que $f(\vec{p})$ es una función de primer grado respecto a cada p_i ,puede comprobarse que $f(\vec{p})$ viene expresado de la siguiente manera:

$$f(\vec{p}) = \sum_{i=o}^{m} C_{m,i}(p_1,p_2,....,p_n) \cdot f_{n-m,i}(\vec{p}^{1,2,...,m})$$

donde $C_{m,i}(p_1,p_2,\ldots,p_n)$ son las sumas simétricas de grado i, lineales en cada p_i, construidas sobre las componentes suprimidas; es decir,

$$C_{m,o}(p_1,p_2,\ldots,p_m) = 1$$

y

$$C_{m,i}(p_1,p_2,\ldots,p_m) = p_1 \cdot p_2 \cdot \ldots \cdot p_i + p_1 \cdot p_2 \cdot \ldots \cdot p_{i-1} \cdot p_{i+1} + \ldots\ldots\ldots + p_{m-i+1} \cdot p_{m-i+2} \cdot \ldots \cdot p_m \quad .$$

Por otro lado, $f_{n-m,i}(\vec{p}^{1,2,..,m})$ son funciones simétricas y lineales de las restantes componentes de $\vec{p}$; no dependen en modo alguno de las m primeras. La forma explícita de dichas funciones se obtiene resolviendo un sistema triangular de ecuaciones lineales. Su solución es

$$f_{n-m,i}(\vec{p}^{1,2,..,m}) = \sum_{h=o}^{i} (-1)^{i-h} \cdot f(0^{m-h}, 1^h, p_{m+1},..,p_n)$$

con la notación $0^{m-h}1^h$ se quiere indicar un vector de probabilidad que tenga m-h componentes nulas y h unitarias.

Al particularizar esta expresión de $f(\vec{p})$ a solo un par de componentes se deducen las condiciones que deben cumplir las restantes para que un vector $\vec{a}$ proporcione el máximo de esa función. Dichas condiciones quedan recogidas

en el siguiente teorema.

Teorema 1.- Si $f(\vec{p})$ alcanza su máximo para el vector $\vec{a}=(a_1,a_2,\ldots,a_n)$, entonces para cualquier par de componentes distintas se verifica

$$f_{n-2,2}(\vec{a}^{i,j}) \leqslant 0 \qquad \text{si} \quad a_i \neq a_j$$

$$f_{n-2,2}(\vec{a}^{i,j}) = 0 \qquad \text{si} \quad 0 < a_i \neq a_j < 1$$

$$f_{n-2,2}(\vec{a}^{i,j}) \geqslant 0 \qquad \text{si} \quad 0 < a_i = a_j < 1$$

Las desigualdades serán estrictas si el máximo no se alcanza sobre los vectores obtenidos al remplazar a_i y a_j por a_i+x y a_j-x con $|x|$ arbitrariamente pequeño .

La clave de la demostración es la comparación de $f(\vec{a})$ con $f(\vec{a}')$, siendo $\vec{a}'$ un vector obtenido por la mencionada sustitución, ya que

$$f(\vec{a}') - f(\vec{a}) = x\cdot(a_j-a_i-x)\cdot f_{n-2,2}(\vec{a}^{ij}) \quad ,$$

de manera que jugando con los posibles valores de x llegamos a las condiciones propuestas.

Para aplicaciones de este teorema es útil la ex-

presión explícita de $f_{n-2,2}$. Si llamamos X_g^{ij} a la variable binomial generalizada que se deduce de X_g al prescindir de las pruebas i y j, tenemos que

$$f_{n-2,2}(\bar{a}^{ij}) = \sum_{k=o}^{n-2} \Delta^2 g(k) \cdot P[X_g^{ij}=k]$$

donde $\Delta^2 g(k)$ denota la segunda diferencia de la función g en el entero k; esto es, $\Delta^2 g(k) = g(k+2)-2\cdot g(k+1)+g(k)$.

Es claro que si la función g(k) fuese estrictamente convexa, es decir $\Delta^2 g(k) > 0$, ya que necesariamente $f_{n-2,2}(\bar{a}^{ij})$ es mayor que cero, se concluye:

Teorema 3.- Si la función g(x) verifica que $\Delta^2 g(k) > 0$ cuando $k=0,1,\ldots,n-2$, entonces su esperanza matemática respecto de cualquier distribución generada por $\mathcal{P}_{np}$ verificará

$$E[g(X_g)] \leqslant \sum_{k=o}^{n} g(k) \cdot b_k(n,p)$$

alcanzandose el signo de igualdad si y solo si $p_1=p_2=p_3=\ldots\ldots=p_n=p$.

En particular, todo momento absoluto $E[|X_g-b|^c]$, respecto de cualquier punto b , cuyo orden c sea superior

a la unidad, alcanza su máximo unicamente si todas las pruebas son equiprobables. En otras palabras, empleando dichos momentos como medidas de dispersión, podemos decir que la binomial es la única distribución de $\mathcal{P}_\mu$ presenta una dispersión maxima.

Si analizamos la dispersión de las distribuciones de $\mathcal{P}_\mu$ desde el punto de vista de la Teoría de la Información, considerando para ello la entropía de Shannon, nosotros hemos comprobado que esta cantidad también resulta maximizada por la binomial, ya que, como veremos en el siguiente punto, el logaritmo de $b_g(X_g)$ es una función concava. Resultados analogos se verifican para otras medidas de informacion; por ejemplo, la binomial minimiza la energia informacional de Onicescu.

Cuando la función $g(x)$ no es convexa, el siguiente teorema da una caracterización parcial de los vectores de probabilidad que corresponden a los extremos de $f(\bar{p})$. Salvo que $g(x)$ sea de una forma particular, no existirá otro tipo de vectores que alcancen dichos extremos.

Teorema 2' .- El máximo y el mínimo de $f(\bar{p})$ se alcanza sobre aquellos vectores cuyas componentes toman, a lo sumo, tres valores diferentes siendo tan solo posible

que uno de ellos sea distinto de 0 o 1. Si r y s son el número de componentes nulas y unitarias de un vector extremo , entonces cualquier otro vector con esas mismas componentes nulas y unitarias será también extremo si y solo si g(x) cumple que

$$g(s+k) = k \cdot g(s+1) - (k-1) \cdot g(s)$$

con k variando desde 2 hasta n-r-s . En tal caso, el valor extremo de $f(\vec{p})$ vale

$$f(\vec{a}) = (1-n \cdot p+s) \cdot g(s) + (n \cdot p-s) \cdot g(s+1) \quad .$$

En otras palabras, este teorema significa que para cualquier g(x) siempre su esperanza respecto a las distribuciones de $\mathcal{P}_\mu$ verificará que

$$\min_{\substack{r=0,1..,[nq] \\ s=0,1..,[np]}} f(\vec{p}(r,s)) \leqslant f(\vec{p}) \leqslant \max_{\substack{r=0,1..,[nq] \\ s=0,1,.,[np]}} f(\vec{p}(r,s))$$

donde $\vec{p}(r,s)$ representa un vector de probabilidad que conste de r componentes nulas, s unitarias y las n-r-s restantes iguales a (np-s)/(n-r-s) .

Según indica S. Samuels (1965), este resultado ya fue obtenido por Tchebichev en 1846 para una función g(x) especial, pero sin hacer uso de esa forma especial .

Aplicando este teorema a determinadas funciones indicador se obtienen resultados importantes sobre las probabilidades $P[X_g < c]$ y $P[b < X_g \leq c]$, que según nuestra notación escribiremos como $B_g(c/\overline{p})$ y $B_g(b,c/\overline{p})$, donde $\overline{p}$ representa el vector de probabilidad de la distribución que se esté considerando. Dichos resultados estan recogidos en los siguientes teoremas :

Teorema 4 .- La función de distribución $B_g(c/\overline{p})$ se encuentra acotada por las siguientes cantidades:

I. $0 \leq B_g(c) \leq B(c)$ si $0 \leq c \leq n \cdot p - 1$

II. $0 < 1 - Q(n-c-1,q) \leq B_g(c) \leq Q(c,p) < 1$
si $np - 1 < c < np$

III. $B(c) \leq B_g(c) \leq 1$ si $np \leq c \leq n$

donde $Q(c,p) = \max_{0 \leq s \leq c} B(c-s;n-s,\alpha)$, con $\alpha = \frac{np-s}{n-s}$.

El valor de s que maximiza $Q(c,p)$ satisface la desigualdad $(c+1-np)\cdot(n-s) < n \cdot q$, salvo que $c=n-1$, en cuyo caso $s=n-1$.

Todas las cotas son alcanzadas. La superior del caso I y la inferior del II solo se alcanzan cuando to

das las pruebas son equiprobables.

Teorema 5.- Sean b y c dos enteros tales que $0 \leqq b \leqq np \leqq c \leqq n$, entonces la probabilidad del intervalo $(b-1,c]$ verifica que

$$B(b-1,c) \leqq B_g(b-1,c) \leqq 1 \quad .$$

Ambas cotas son alcanzadas. La inferior solo se alcanza para la distribución binomial salvo que $b=0$ y $c=n$.

Con este teorema hemos obtenido otra forma de expresar el carácter extremo respecto a la dispersión de la distribución binomial en la familia $\mathcal{T}_\mu$, puesto que es la única que hace corresponder una menor probabilidad a todos los intervalos construidos alrededor de la media.

Por último, Hoeffding da algunas aplicaciones estadisticas de sus resultados. En primer lugar, como consecuencia del teorema 5, los test de hipótesis habituales (de una o dos colas) para el parámetro p de la binomial pueden aplicarse directamente para contrastar el valor de la probabilidad promedio de la distribución binomial de Poisson. Si entendemos como nivel de significación de estos test el máximo de la probabilidad del error de tipo 1 bajo cualquier

condición posible, entonces su valor es exactamente el mismo que tiene el test binomial utilizado.

Esta propiedad conservativa de los test binomiales fue demostrada por inducción sobre el número de pruebas por H. David(1960) , quien en un principio desconocía el trabajo de Hoeffding. J. Walsh (1955) ha considerado qué correcciones aproximadas habría que realizar para obtener las auténticas probabilidades bajo la hipótesis nula.

De manera análoga puede comprobarse que la potencia de dichos test frente a un valor alternativo no muy próximo a la hipótesis nula es al menos igual a la potencia del correspondiente test binomial. Igualmente, los intervalos de confianza basados en la binomial pueden emplearse para la probabilidad promedio cuando el nivel de confianza es suficientemente alto.

Por otra parte, del teorema 3 se desprende la siguiente aplicación a problemas de estimación puntual: supongamos que queremos estimar una función $\theta(p)$ sometidos a la función de perdida $W(p,t)$ cuando $\theta(p)=t$. Si el estimador t es función de X_g y $W(p,t(X_g))$ es una función estrictamente convexa respecto de X_g para cada p, entonces el riesgo

$E[W(p,t(X_g))]$ alcanza su valor maximo unicamente cuando todas las componentes de p son iguales. De aquí resulta que, sin imponer ninguna limitación adicional a la clase de estimadores, el estimador minimax cuando todas la p_j son iguales, también sigue siendolo al no verificarse esa igualdad.

Pasamos ahora a ocuparnos del artículo de L. Gleser, cuyos resultados son generalizaciones mas o menos completas de los tres últimos teoremas de Hoeffding. La idea básica que ha conducido a esos resultados es la aplicación de la noción de mayoración de vectores al conjunto $\mathcal{P}_\mu$.

Dados dos vectores, $\vec{x}$ e $\vec{y}$, de un espacio n-dimensional, se dice que $\vec{x}$ mayora a $\vec{y}$ ($\vec{x} \geqq \vec{y}$) si sus componentes, ordenadas según el valor decreciente de las mismas, verifican las siguientes relaciones :

$$\sum_{i=1}^{k} x_{[i]} \geqq \sum_{i=1}^{k} y_{[i]} \qquad k = 1,2,\ldots,n-1 \quad ,$$

$$\sum_{i=1}^{n} x_{[i]} = \sum_{i=1}^{n} y_{[i]}$$

Dos vectores comparables respecto este orden parcial están caracterizados por el siguiente lema.

Caracterización.- $\vec{x}$ mayora a $\vec{y}$ es equivalente a la exis

tencia de una matriz Π, doblemente estocastica, que transforme $\vec{x}$ en $\vec{y}$. Es decir

$$\vec{x} \overset{\varepsilon}{\leqq} \vec{y} \quad \Leftrightarrow \quad \vec{y} = \vec{x}\cdot\Pi \quad .$$

La relación de mayoración garantiza el comportamiento monotono no decreciente de las funciones $F(\vec{z})$, simétricas respecto a las permutaciones de las componentes del argumento, sobre el dominio donde se verifica la siguiente condición :

$$\forall i,j \quad i\neq j \qquad (z_i - z_j)\cdot\left\{\frac{\partial F}{\partial z_i} - \frac{\partial F}{\partial z_j}\right\} \geqq 0$$

A dicha desigualdad se la conoce con el nombre de condición de Schur.

Al considerar sobre $\mathcal{P}_\mu$ la ordenación por mayoración, este espacio queda dotado de elemento máximo y mínimo. Dichos elementos son, respectivamente, el vector compuesto por el número máximo de componentes iguales a cero y a uno, y aquel cuyas componentes son iguales a la probabilidad promedio; es decir

$$\vec{p}^{\star} = (1,\ldots,1,\mu-[\mu],0,\ldots,0)$$

y

$$\vec{p}^{\triangle} = (p,p,\ldots,p) \quad , \quad \text{con} \quad p = \mu/n \quad .$$

Esta afirmación se desprende sin ninguna dificultad de la propia definición de mayoración. Además, teniendo en cuenta la caracterización matricial, es claro que cualquier vector de probabilidad se transforma en $\vec{p}^{\Delta}$ por intermedio de la matriz cuyos elementos son todos iguales a $1/n$.

Estos dos vectores ocupan un lugar relevante en los teoremas de Hoeffding. En este sentido, tenemos que el máximo y el mínimo de las partes I y III del teorema 4 se alcanzan sobre dichos vectores. En otras palabras, se verifica que $\vec{p}^{\Delta} \leqq \vec{p} \leqq \vec{p}^{\star}$ implica:

$$\text{I} \qquad 0 = B_g(c/\vec{p}^{\star}) \leqq B_g(c/\vec{p}) \leqq B_g(c/\vec{p}^{\Delta}) \qquad c \leqq [\mu - 1]$$

$$\text{III} \qquad B_g(c/\vec{p}^{\Delta}) \leqq B_g(c/\vec{p}) \leqq B_g(c/\vec{p}^{\star}) = 1 \qquad c \geqq [\mu + 1]$$

La principal aportación de Gleser consiste en demostrar que esta implicación sigue siendo valida casi por completo cuando se remplaza $\vec{p}^{\Delta}$ o $\vec{p}^{\star}$ por cualquier otro vector de $\mathcal{P}_{\mu}$. Esas relaciones no se cumplen de manera general cuando c vale $[\mu-1]$ o $[\mu+1]$, como comprueba Gleser con dos contraejemplos; no obstante, imponiendo algunas restricciones a los vectores de probabilidad, se pueden incluir estos valores de c. Veamos el enunciado original de esta aportacion.

Teorema .- Sea $\vec{p}^{(1)} \in \mathcal{P}_\mu$ y supongamos que existe una matriz doblemente estocastica Π para la cual

$$\vec{p}^{(2)} = \vec{p}^{(1)} \cdot \Pi$$

Entonces $\vec{p}^{(2)} \in \mathcal{P}_\mu$ y

$$B_g(c/\vec{p}^{(1)}) \leqq B_g(c/\vec{p}^{(2)}) \quad \text{si} \quad 0 \leqq c \leqq [\mu-2]$$

y

$$B_g(c/\vec{p}^{(1)}) \geqq B_g(c/\vec{p}^{(2)}) \quad \text{si} \quad [\mu+2] \leqq c \leqq n \quad .$$

El procedimiento de demostración consiste en comprobar que $B_g(c)$ verifica sobre todo $\mathcal{P}_\mu$ la condición de Schur. Para ello primeramente calcula el valor del primer miembro de esa condición para $f(\vec{p})$, la esperanza matematica de una función arbitraria $g(X_g)$; después lo particulariza a la función indicador del conjunto $[0,c]$. Recordando que X_g^{ij} representa la variable binomial generalizada resultante de eliminar las pruebas iesima y j-esima, el primer miembro de la condición de Schur vale

$$(p_i-p_j)\cdot\left\{\frac{\partial f(\vec{p})}{\partial p_i} - \frac{\partial f(\vec{p})}{\partial p_j}\right\} = -(p_i-p_j)^2 \cdot \sum_{k=o}^{n-2} \Delta^2 g(k) \cdot P[X_g^{ij}=k]$$

De donde se desprende como corolario inmediato que la esperanza matematica de toda función concava tiene un comportamiento monotono no decreciente respecto de vectores compara

bles. No obstante, esto no es una nueva aportación, sino una redemostración de dicho resultado que ya fue publicado en el trabajo de S. Karlin y A. Novikoff (1963). Si consideramos funciones convexas obtenemos como caso particular el teorema 3 de Hoeffding con la precisión adicional de que el valor mínimo de $f(\overline{p})$ se alcanza sobre $\overline{p}^{\star}$. Esta precisión se puede dar también siguiendo el metodo de Hoeffding. La aplicación que se hizo del teorema 3 a los momentos absolutos puede extenderse a las parejas de vectores de probabilidad comparables.

La aplicación de la última expresión a la función indicador de [0,c] da como resultado

$$-(p_i-p_j)^2 \cdot \left\{ b_g(c/\overline{p}^{ij}) - b_g(c-1/\overline{p}^{ij}) \right\} \quad .$$

Teniendo en cuenta, como veremos en los dos siguientes puntos, que $b_g(c)$ es una función creciente hasta $c=[\lambda]$ y decreciente a partir de $c=[\lambda+1]$, siendo $\lambda = \mu - p_i - p_j$, se puede asegurar que la condición de Schur se cumple respecto de $-B_g(c)$ cuando $c \leq [\mu-2]$, y respecto de $B_g(c)$ cuando $c \geq [\mu+2]$. Por tanto, el teorema queda demostrado.

Para los valores de $c=[\mu-1]$ y $c=[\mu+1]$, Gleser define dos subconjuntos en los que las desigualdades del teo-

rema también quedan garantizadas. Dichos subconjuntos son, respectivamente:

$$\mathcal{S} = \left\{ \overline{p} \in \mathcal{P}_\mu \,/ [\mu - 1] \leqslant \mu - p_i - p_j \quad , \forall i \neq j \right\}$$

$$\mathcal{C} = \left\{ \overline{p} \in \mathcal{P}_\mu / \ [\mu] \leqslant \mu - p_i - p_j \quad , \forall i \neq j \right\} \quad .$$

En este mismo sentido, basandonos en la localización de la moda, nosotros hemos encontrado unos resultados mas fuertes que garantizan el cumplimiento de las desigualdades para esos dos valores críticos de c, cuando μ toma ciertos valores particulares o bien las componentes de $\vec{p}^{(1)}$ verifican ciertas condiciones. Nuestros resultados son los siguientes:

I) Si $\forall i$ se verifica que $\mu + p_i^{(1)} < [\mu + 1]$, entonces el comportamiento monótono está garantizado para $[\mu + 1]$.

II) Si $\forall i$ se verifica que $\mu + p_i^{(1)} \geqq [\mu + 1]$, entonces el comportamiento monótono está garantizado para $[\mu - 1]$.

III) Si el valor esperado verifica las desigualdades

$$[\mu] \leqslant \mu < [\mu] + \frac{1}{[\mu] + 2} \quad ,$$

entonces el teorema es válido también para $[\mu + 1]$.

IV) Si el valor esperado verifica las desigualdades

$$[\mu] + \frac{n - [\mu]}{n - [\mu] + 1} < \mu < [\mu+1]$$

entonces el teorema es válido tambien para $[\mu-1]$.

Finalmente, Gleser obtiene, como corolario inmediato, la siguiente generalización parcial del teorema 5 de Hoeffding.

Teorema.- Supongamos que $0 \leq b \leq [\mu-1]$ y $[\mu+2] \leq c \leq n$. Sea $p^{(1)} \in \mathcal{P}_\mu$ y sea $p^{(2)} = p^{(1)} \cdot \Pi$, donde Π es una matriz $n \times n$, doblemente estocastica. Entonces

$$B_g(b-1, c/p^{(2)}) \leq B_g(b-1, c/p^{(1)})$$

Naturalmente, si imponemos además alguna de las condiciones que hemos añadido al teorema principal de Gleser, se garantiza el cumplimiento de la relación entre las funciones de distribución para $b=[\mu]$, o bien $c=[\mu+1]$, según sea la condición elegida.

B) Forma de la distribución .-

El estudio del modo en que la probabilidad puntual $b_g(k)$ varía con respecto a k ha sido llevado a cabo

por J.N. Darroch (1964) y S.M. Samuels (1965). En dichos artículos también se resuelve el problema de la localización de la moda.

Los resultados de Darroch pueden resumirse en la siguiente afirmación: $b_g(k)$ tiene forma acampanada . Dicho de otro modo, $b_g(k)$ pasa sucesivamente por las siguientes etapas: crecimiento convexo, crecimiento concavo, decrecimiento concavo, y, finalmente, decrecimiento convexo.

Tanto el crecimiento como el decrecimiento son siempre estrictos, salvo que existan, a lo sumo, tres puntos alineados. El orden en que se suceden estas etapas es siempre el mismo, aunque es posible que algunas etapas no se presenten.

Este enunciado es fruto de dos teoremas, cuyas demostraciones se establecen por inducción :

Teorema 1 .-

$$\Delta b_g(k) \geqq 0 \implies \Delta b_g(k-1) > 0 \ , \quad k=1,2,\ldots,n-1 \quad .$$

Teorema 2 .-

$$\Delta^2 b_g(k-1) \geqq 0 \quad y \quad \Delta b_g(k-1) > 0 \implies$$

$$\Delta^2 b_g(k-2) > 0 \quad \text{cuando} \quad k=1,2,\ldots,n$$

En este mismo sentido, Samuels obtiene resultados similares para las diferencias de cualquier orden de las probabilidades puntuales, basandose en una desigualdad atribuida a Newton, la cual establece que los coeficientes $b_0, b_1, \ldots, b_n$, definidos a partir de las constantes reales no nulas $a_1, a_2, \ldots, a_n$ mediante la igualdad

$$\sum_{k=0}^{n} \binom{n}{k} \cdot b_k \cdot z^k = \prod_{i=1}^{n} (1 + a_i \cdot z) \quad ,$$

verifican que $b_k^2 > b_{k-1} \cdot b_{k+1}$ para $k=1,2,\ldots,n-1$, salvo que todas las a_i fuesen iguales, en cuyo caso siempre se verifica la igualdad.

Al aplicar dicha desigualdad a la función generatriz de probabilidad de X_g, se obtiene que

$$\left| \frac{b_g(k)}{\binom{n}{k}} \right|^2 > \frac{b_g(k-1)}{\binom{n}{k-1}} \cdot \frac{b_g(k+1)}{\binom{n}{k+1}} \quad .$$

Resultado que ya fue deducido por P. Levy en 1937, y del que se desprende otra relación ligeramente mas debil:

$$b_g^2(k) > b_g(k-1) \cdot b_g(k+1) \quad , \text{ para } k=1,2,\ldots,n-1 \, .$$

Es decir, $b_g(k)$ es una función logaritmico-concava.(En ello nos apoyamos para dar la propiedad de la entropía que vimos en el punto anterior.)

De ello también se deduce que X_g sigue una distribución unimodal, cuya moda es única o bien compartida entre dos enteros adyacentes. Mas aun, la variable binomial generalizada sigue una distribución fuertemente unimodal, (Vease I.3) como asegura el siguiente teorema de J. Keilson y H. Gerber (1971) .

Teorema.- Una condición necesaria y suficiente para que una distribución $\{p_k\}$ sea fuertemente unimodal es que p_k sea log-concava para cualquier k .

Las relaciones que ligan a las sucesivas diferencias de $b_g(k)$ se deducen teniendo en cuenta que

$$\sum_{k=o}^{n} \Delta^r b_g(k) \cdot z^{k+r} = (1-z)^r \cdot \prod_{i=1}^{n} (q_i + p_i z) \quad , \forall r \in \mathcal{N}$$

de manera que se verifica que

$$\left| \frac{\Delta^r b_g(k)}{\binom{n+r}{k+r}} \right|^2 > \frac{\Delta^r b_g(k-1)}{\binom{n+r}{k+r-1}} \cdot \frac{\Delta^r b_g(k+1)}{\binom{n+r}{k+r+1}}$$

para $r=0,1,\ldots\ldots$ y $k=1,2,\ldots,n+r-1$

De la versión débil obtenida al eliminar los factoriales, se deduce que $\Delta^{r+1}b_g$ puede tener, a lo sumo, un cambio estricto de signo en el intervalo de valores de k en que $\Delta^r b_g$ tiene signo constante. En otras palabras, $\Delta^r b_g$ tiene a lo mas r cambios de signo. Todos estos resultados pueden resumirse bajo el siguiente enunciado: La distribución binomial generalizada tiene forma acampanada de orden r, cualquiera que sea r. El enunciado de Darroch queda englobado en este, ya que corresponde solamente a los ordenes primero y segundo.

C) Localización de la moda y la mediana .

Como antes señalamos, la localización de la moda ha sido tratada por Darroch y Samuels. La localización de la mediana ha sido abordada por K. Jogdeo y el propio Samuels en el año 1968. Veamos los resultados que se han conseguido en relación con estas dos características de la distribución.

Para centrarnos en lo primero, demos de manera diferente el criterio de localización de la moda de la distribución binomial: Si el valor medio μ verifica que $k \leqslant \mu < k+1$,

entonces la moda se localiza en :

$$\begin{cases} k & , \text{ si } \quad \mu+p < k+1 \\ k+1 & , \text{ si } \quad \mu+p > k+1 \\ k \text{ y } k+1 & , \text{ si } \quad \mu+p = k+1 \end{cases} .$$

Dicho resultado ha sido generalizado por Samuels. Sean $p_{[1]}$ y $p_{[n]}$, respectivamente, la mayor y la menor componente del vector de probabilidad de una distribución binomial generalizada. Si $k \leq \mu < k+1$, entonces la moda será :

$$\begin{cases} k & , \text{ si } \quad \mu + p_{[1]} < k+1 \\ k+1 & , \text{ si } \quad \mu + p_{[n]} > k+1 \\ k, k+1 \text{ o ambos} & , \text{ en cualquier otro caso .} \end{cases}$$

Esta generalización se obtiene como corolario inmediato del siguiente teorema.

Teorema .- Si $k \leq \mu \leq k+1$, entonces

(i) $\quad b_g(X^{[n]}=k) > b_g(X^{[n]}=k-1)$

(ii) $\quad b_g(X^{[1]}=k) > b_g(X^{[1]}=k+1)$

salvo que $p_{[1]} = p_{[n]} = \frac{k}{n}$ ó $\frac{k+1}{n}$, en cuyos casos se verifica, respectivamente, la igualdad de (i) o (ii).

Por otro lado, Darroch consigue una caracterización completa de la moda en términos de la media, recogida en el siguiente enunciado.

Teorema.- La moda de una variable binomial de Poisson es :

I) μ , si $\delta(\mu) = 0$

II) $[\mu]$, si $0 < \delta(\mu) < \frac{1}{[\mu]+2}$

III) $[\mu+1]$, si $\frac{n-[\mu]}{n-[\mu]+1} < \delta(\mu) < 1$

IV) $[\mu]$, $[\mu+1]$, o ambos, si $\frac{1}{[\mu]+2} \leq \delta(\mu) \leq \frac{n-[\mu]}{n-[\mu]+1}$

siendo $\delta(\mu)$ la parte fraccionaria de μ; o sea, $\mu - [\mu]$.

Queremos hacer notar, siguiendo a Samuels, que no existe ninguna otra función racional de las probabilidades de "exito", aparte de $\Delta b_g(k)$, que determine de modo único la moda. Incluso es así cuando intervienen solo dos pruebas de Poisson.

Para la demostración de su teorema, Darroch introduce dos números enteros dependientes de μ :

$$k^{\Delta}(\mu) = \max\{k/ \min_{p\in\mathcal{P}_\mu} \{b_g(k) - b_g(k-1)\} > 0 \} ,$$

y

$$k^{\star}(\mu) = \min\{k/ \max_{p\in\mathcal{P}_\mu} \{b_g(k+1) - b_g(k)\} < 0 \} .$$

Es decir, tales valores son, respectivamente, hasta donde todas las distribuciones engendradas por $\mathcal{P}_\mu$ crecen estric<u>t</u>amente,y a partir de donde todas ellas decrecen estrictamente.

Su determinación se realiza con la ayuda del 3er teorema de Hoeffding, inspeccionando, por ello, unicamente las distribuciones correspondientes a los vectores de probabilidad que consten,a lo sumo, de tres valores diferentes, solo uno de los cuales puede ser distinto de 0 y 1. Dichas distribuciones son esencialmente binomiales desplazadas, de modo que su valor modal está perfectamente determinado, y, por tanto, el signo de la diferencia de las probabilidades puntuales se comprueba facilmente. Razonando de esa manera se obtiene que

$$k^{\triangle}(\mu) = \begin{cases} \mu & \text{, si es entero} \\ \mu + \frac{\mu - [\mu]}{n - [\mu]} - 1 & \text{, si } \mu \text{ no es entero, pero si esa cantidad.} \\ \left[\mu + \frac{\mu - [\mu]}{n - [\mu]}\right] & \text{, en el caso restante} \end{cases}$$

La determinación de $k^{\star}(\mu)$ se realiza a partir de $k^{\triangle}(n-\mu)$, correspondiente a la distribución del número de "fracasos", resultando que

$$k^{\star}(\mu) = \begin{cases} \mu & , \text{ si es entero} \\ \left[\mu - \frac{[\mu]+1-\mu}{[\mu]+1}\right] + 1 & , \text{ en el otro caso} \end{cases} .$$

Cuando k^{Δ} y $k^{\star}$ coinciden, la moda es ese mismo valor y queda univocamente definida conocido el valor medio μ. De igualar estos valores se deducen las condiciones que debe cumplir $\delta(\mu)$ para garantizar moda única.

Con respecto a la mediana, los resultados de Jogdeo y Samuels (1968) son paralelos a los que acabamos de dar, y vienen expresados por los siguientes enunciados.

Teorema.- La mediana del número de "exitos" en n pruebas de Poisson solo puede ser uno de los enteros adyacentes a la media. Si la media fuese un valor entero, la mediana coincide con ella.

Teorema.- Si $k \leq \mu < k+1$, entonces la mediana será unica y valdrá :

(i) k , cuando $\mu < \lambda^{\Delta}$

(ii) k+1 , cuando $\mu > \lambda^{\star}$

siendo λ^{Δ} y $\lambda^{\star}$, respectivamente, las soluciones de

$$\frac{1}{2} = \min_{r=0,1,\ldots,n-k-1} B_k\left(n-r, \frac{\lambda^{\Delta}}{n-r}\right)$$

y

$$\frac{1}{2} = \max_{s=0,1,\ldots,k} B_{n-k}(n-s, \frac{\lambda^{\star}}{n-s}) \quad .$$

El primer teorema está basado en propiedades de convergencia monotona de las probabilidades binomiales. El segundo teorema es consecuencia del anterior y del 4º teorema de Hoeffding. Desgraciadamente, a diferencia de lo que sucede con la moda, no es posible dar λ^{Δ} y $\lambda^{\star}$ como funciones racionales de n y k .

D) Generalización del teorema de Bernoulli.

La versión generalizada de dicho teorema fue demostrada por Poisson. Su enunciado es el siguiente :

> Teorema de Poisson .- La probabilidad de que la frecuencia relativa de "exitos" X_g/n difiera del promedio de las probabilidades p_i en una cantidad mayor o igual a ϵ, tiende a cero cuando $n \to \infty$, por pequeño que se elija $\epsilon > 0$.(Cramer(1970), pag 237)

La demostración del teorema se realiza, actualmente, de forma inmediata, aplicando la desigualdad de Tchebychev

$$P\left[\left|\frac{X_g}{n} - p\right| \geq \epsilon\right] \leq \frac{\sum_{i=1}^{n} p_i \cdot q_i}{n^2 \cdot \epsilon^2} \leq \frac{1}{4n \cdot \epsilon^2}$$

La interpretación frecuencial de este teorema es completamente analoga a la del teorema de Bernoulli.(vease I.13)

Por otro lado, también es posible generalizar el teorema de Moivre-Laplace. En este sentido, tenemos el siguiente teorema,debido a Cramer.

Teorema .- Una condición necesaria y suficiente para que la distribución limite de

$$\frac{X_g - \mu}{\sqrt{\sum_{i=1}^{n} p_i \cdot q_i}}$$

sea normal con media cero y varianza unidad,es que la serie $\sum_{i=1}^{n} p_i \cdot q_i$ diverja cuando $n \to \infty$.

Evidentemente, este resultado es simplemente un caso particular de otros teoremas mas fuertes, como , por ejemplo, el teorema de Liapounov. A continuación, pasamos a exponer los principales resultados relativos a la convergencia de la variable binomial generalizada hacia la otra ley limite.

E) Convergencia hacia la distribución de Poisson .

La convergencia de la distribución binomial generalizada hacia la de Poisson puede expresarse de dos formas diferentes: a traves de teoremas limites, o bien, mediante teoremas de aproximación. Los primeros se refieren a sucesiones de variables binomiales generalizadas; los segundos , a una sola variable.

Los teoremas del primer tipo tienen un valor eminentemente matemático. Los del segundo tipo tienen además un valor práctico: justifican el hecho de que observaciones tales como número de accidentes, suicidios, llamadas telefónicas, etc., se ajusten bien a la ley de Poisson. Por una parte, la variable binomial generalizada es un modelo mucho mas realista para esas situaciones que la binomial. Por otra parte, los teoremas de aproximación aseguran que, para tamaños finitos de población, la distribución de Poisson proporcione buenos resultados. Es decir, siguiendo a Hodges y Le Cam, este tipo de teoremas es la mejor manera de explicar el fenómeno empirico de la ley de los pequeños números.

Pasemos a enumerar los distintos resultados empezando por los teoremas limites. Para ello consideremos una sucesión de variables binomiales generalizadas $\{X_{g_n}\}$,

con $X_{g_n} = \xi_{1_n} + \xi_{2_n} + \ldots + \xi_{n_n}$, donde ξ_{i_n} representa el número de "exitos" de cada prueba de Poisson.

En 1950 B.O. Koopman demuestra que es condición necesaria y suficiente para que $b_{g_n}(k)$ converja uniformemente hacia $p_k(\lambda)$ el cumplimiento de las siguientes igualdades:

$$(I) \qquad \lim_{n-\infty} E[X_{g_n}] = \lim_{n-\infty} \sum_{i_n=1}^{n} p_{i_n} = \lambda$$

$$(II) \qquad \lim_{n-\infty} \{ \max_{i_n=1,2,\ldots,n} p_{i_n} \} = \lim_{n-\infty} p_{[1]_n} = 0$$

Con anterioridad, R. von Mises había demostrado en 1921 que $\{X_{g_n}\}$ tiene en el límite una distribución de Poisson de parametro λ, en el supuesto de que p_{i_n} varía con n de tal forma que $E[X_{g_n}]$ se mantiene igual a λ y $p_{[1]_n}$ tiende a 0.

Del propio von Mises procede nuestra observación anterior de que el modelo binomial generalizado se ajusta mejor a la realidad que el modelo binomial. No obstante, para algunas situaciones este modelo aun resulta insatisfactorio al exigir la independencia entre las pruebas. Si consideramos un modelo,un poco mas realista,que admita cierta dependencia entre las pruebas, también es posible asegurar su convergencia hacia la ley de Poisson bajo las siguientes hi

pótesis:

(I') Cada prueba es independiente de al menos n-m-1 de las pruebas restantes y se verifica que

$$\lim_{n\to\infty}\left\{(m+1)\cdot \max P\left[\xi_{i_n}=1/\xi_{1_n},\xi_{2_n},\ldots,\xi_{(i-1)_n}\right]\right\} = 0$$

El máximo se toma sobre todos los posibles valores de $\xi_{1_n}, \xi_{2_n}, \ldots, \xi_{(i-1)_n}$, con $1_n \leq i \leq n$.

(II') La suma de las probabilidades incondicionales de "exito" en cada prueba tiene limite finito,

$$\lim_{n\to\infty} \sum_{i_n=1}^{n} P\left[\xi_{i_n}=1\right] = \lambda < \infty$$

Este teorema ha sido publicado por J.E. Walsh en 1955 ; observese que sus hipótesis se reducen a las de Koopman, cuando existe independencia entre las pruebas.

Por otro lado, la condición suficiente del teorema de Koopman es un caso particular del siguiente teorema, demostrado en el libro de Rényi.

Teorema.- Sean $\xi'_{n1}, \xi'_{n2}, \ldots, \xi'_{nn}$, variables aleatorias independientes definidas sobre los enteros no negativos. Su suma

$$\eta_n = \xi'_{n1} + \xi'_{n2} + \ldots + \xi'_{nn}$$

converge, cuando $n \to \infty$, hacia la función de distribución de la ley de Poisson de parametro λ, si se verifican estas tres condiciones:

$$\text{(i)} \quad \lim_{n\to\infty} \sum_{i=1}^{n} P[\xi'_{ni} = 1] = \lambda$$

$$\text{(ii)} \quad \lim_{n\to\infty} \max P[\xi'_{ni} \geq 1] = 0$$

$$\text{(iii)} \quad \lim_{n\to\infty} \sum_{i=1}^{n} P[\xi'_{ni} \geq 2] = 0$$

Terminamos este tipo de resultados con un teorema debido a L. Chen(1974), con el cual se prueba que la convergencia de esta distribución es tan fuerte como la convergencia de la binomial. En otras palabras, se trata del resultado de Simons y Johnson (vease I.10) aplicado a la distribución binomial generalizada.

Teorema.- Sea $\{X_{g_n}\}$ una sucesión de variables binomiales generalizadas con media fija λ tal que $p_{[1]_n} \to 0$ y sea h(i) una función no negativa. Entonces

$$\lim_{n\to\infty} \sum_{i=0}^{\infty} h(i) \cdot |b_{g_n}(i) - p_i(\lambda)| = 0 \iff \sum_{i=0}^{\infty} h(i) \cdot p_i(\lambda) < \infty$$

La base de su demostración consiste en probar que el cociente $b_{g_n}(i)/p_i(\lambda)$ a partir de un cierto i es

decreciente. Para ello Chen hace uso de la identidad

$$E[X_g \cdot f(X_g)] = \sum_{i=1}^{n} p_i \cdot E[f(X_g^i + 1)] \quad , \forall f(X_g) \text{ acotada},$$

particularizada a la función indicador del punto i+1. La validez de la identidad, que Chen generaliza en su Tesis Doctoral, es consecuencia de que las ξ_i son independientes entre sí y solo toman los valores 0 y 1. La forma particular de la identidad fue empleada por Samuels(1965) en relación con la localización de la moda, y, en este caso, su validez también se deduce por derivación de la función generatriz de probabilidad e identificación de terminos.

A continuación veamos en orden de importancia creciente algunos teoremas de aproximación. En todos ellos se acota alguna medida de desviación entre una determinada distribución binomial generalizada y la ley de Poisson que tenga igual promedio. Como medidas de desviación se consideran la maxima diferencia entre las funciones de distribución y la variación total entre las funciones de masa de probabilidad.(Observese que esta última medida siempre es mayor o igual que la primera.)

Como consecuencia de un teorema general de Kolmogorov (1956) sobre suma de variables aleatorias surge el

primer resultado. Bajo las condiciones tan amplias del teorema general, sólo puede afirmarse que

$$\underset{k=0,\ldots,n}{\text{Max}} \; |B_g(k) - P(k;\lambda)| \leq c_1 \cdot \sqrt[5]{p_{[1]}} \quad ,$$

pero con la importante propiedad de que c_1 es una constante independiente de n y de las p_i.

Restringiendonos a las condiciones concretas de nuestra distribución, la acotación se mejora notablemente. De este modo, J. Hodges y L. Le Cam (1960), empleando argumentos simples, obtuvieron los siguientes resultados :

$$\underset{k=0,\ldots,n}{\text{Max}} \; |B_g(k) - P(k;\lambda)| \leq 2 \cdot \nu$$

y

$$\underset{k=0,\ldots,n}{\text{Max}} \; |B_g(k) - P(k;\lambda)| \leq 3 \cdot \sqrt[3]{p_{[1]}}$$

donde ν representa a $\sum_{i=1}^{n} p_i^2$.

Aunque den mas información que los teoremas limites, estos resultados son de escaso valor practico, ya que si suponemos, por ejemplo, que $p_{[1]}=0.01$, entonces $\sqrt[5]{p_{[1]}} \cong 0.4$, valor que supera cualquier probabilidad puntual de la Poisson salvo que $\lambda \leq 1$, en cuyo caso el ajuste es bastante bueno.

No obstante, utilizando herramientas mas potentes se consiguen resultados verdaderamente importantes. Así, Le Cam en 1960, con ayuda de la función caracteristica y la teoria de la medida como operador sobre grupos abelianos medibles, ha obtenido dos resultados importantes. Estos, expresados en terminos de la distancia de variación total, son

$$\sum_{k=o}^{\infty} |b_g(k) - p(k;\lambda)| \leq c_2 \cdot p_{[1]} \quad , \quad \forall p_{[1]}$$

y

$$\sum_{k=o}^{\infty} |b_g(k) - p(k;\lambda)| \leq c_3 \cdot \frac{\nu}{\lambda} \quad , \quad \text{si } p_{[1]} \leq 1/4$$

donde las constantes c_2 y c_3 toman valores inferiores a 9 y 16, respectivamente. Cuando $p_{[1]}$ y ν son muy pequeños, las constantes se pueden reducir sensiblemente, aunque c_3 no puede ser mucho mas pequeña que 1/2, puesto que un teorema de Prohorov (1953) proporciona un valor de c_3 proximo a 1/2 en el caso particular de la distribución binomial. El segundo resultado de Le Cam debe considerarse como generalización del teorema de Prohorov.

Naturalmente, de todos estos resultados se desprenden de manera inmediata teoremas limites al considerar sucesiones de variables. Así, por ejemplo, si $p_{[1]} \to 0$ cuando $n \to \infty$, entonces

$$\sum_{k=o}^{\infty} |b_g(k) - p(k;\lambda)| \to 0 \quad ,$$

lo cual es simplemente el teorema de Chen particularizado a $h(k) \equiv 1$. Nosotros al mencionar en la sección I.10 el teorema de Le Cam lo hicimos como teorema limite de la binomial. Sin embargo, los resultados de Kerstan y Vervaat suponen simplemente la reducción de la constante c_2 en ese caso particular hasta los valores de 2 y, aproximadamente, $\sqrt{2}$.

Finalmente, es evidente que una aplicación de esta convergencia es el empleo de la distribución de Poisson como aproximante de la binomial generalizada. De lo que aquí hemos visto se deduce que este procedimiento dará buenos resultados siempre que $p_{[1]}$ sea pequeño, independientemente del valor de n; el maximo error cometido nunca será superior a 9 veces esa probabilidad. Además, si se desea mayor precisión pueden emplearse las siguientes fórmulas de Kolmogorov :

$$b_g^{\star}(0) = p_o(\lambda) \cdot \left\{ 1 - \frac{\nu}{2} \right\}$$

$$b_g^{\star}(1) = p_1(\lambda) \cdot \left\{ 1 - \frac{\nu(1-2/\lambda)}{2} \right\}$$

$$k \geq 2 \qquad b_g^{\star}(k) = p_k(\lambda) - \frac{\nu}{2} \cdot p_{k-2}(\lambda) \cdot \left\{ \frac{\lambda^2}{(k-1)k} - \frac{2\lambda}{k-1} + 1 \right\}$$

Estas fórmulas son simplemente el segundo aproximante del

desarrollo Gram-Charlier de tipo B . Este tipo de aproximante adolece del inconveniente de proporcionar valores negativos cuando las probabilidades reales son pequeñas. Eso puede comprobarse en un sencillo ejemplo que Smirnov y Dunin-Barkowskij dan en su libro, donde también se pone de manifiesto la superioridad de estas fórmulas frente al empleo de solo las probabilidades de la distribución de Poisson.

I.13.- OTRAS PROPIEDADES

En esta sección recogemos algunos resultados notables sobre la distribución binomial que no han podido incluirse en las secciones anteriores. Concretamente vamos a tratar de los siguientes puntos: a) Teorema de Bernoulli ; b) Comparación entre las colas binomiales y las de sus distribuciones limite ; c) Acotación de sus probabilidades basada en una modificación de la aproximación normal propuesta por Feller ; d) Propiedades de monotonía de las probabilidades binomiales.

a) Teorema de Bernoulli.

Dada la importancia de este teorema debimos haberlo incluido en la sección I.7 . Como todos sabemos, su

enunciado es el siguiente:

> La probabilidad de que la frecuencia X/n difiera, en valor absoluto, de la probabilidad p en una cantidad mayor o igual a ϵ, tiende a 0 cuando $n \to \infty$, por pequeña que elijamos la cantidad $\epsilon > 0$.

La demostración original de Bernoulli consistió en probar que el cociente entre la probabilidad del suceso contrario y la de este puede hacerse tan grande como se quiera con tal de tomar n suficientemente grande.

La demostración mas habitual consiste en la aplicación inmediata de la desigualdad de Tchebychev:

$$P\left[\left|\frac{X}{n} - p\right| \geq \epsilon\right] < \frac{p \cdot q}{n \cdot \epsilon^2} \leq \frac{1}{4 \cdot n \cdot \epsilon^2}$$

De esta manera se consigue también, para cada n, acotar superiormente la probabilidad de ese suceso. Esta cota ha sido reducida sucesivamente destacando, por ejemplo, la obtenida por P. Levy (1954)

$$P\left[\left|\frac{X}{n} - p\right| \geq \epsilon\right] < \frac{2}{\epsilon\sqrt{n}} \cdot e^{-2n\epsilon^2} \qquad \text{si} \quad p,q > \max\left(\frac{4}{n}, 2\epsilon\right) ,$$

que ha sido ligeramente mejorada por Kambo y Kotz (1966) que

dando como

$$P\left[\left|\frac{X}{n} - p\right| < \epsilon\right] < \sqrt{\frac{2}{n}} \cdot \frac{1}{\epsilon} \cdot e^{-2n\epsilon^2\left(1 + \frac{4}{3}\epsilon^2\right)}$$

en el supuesto adicional de que $n > 2$. Pueden encontrarse otras cotas similares en el libro de Johnson y Kotz (1969).

Una consecuencia relevante de este teorema es su interpretación frecuencial que viene a decir que si para un ϵ dado se toma un n_o fijo suficientemente grande, de modo que $P\left[\left|\frac{X}{n} - p\right| < \epsilon\right]$ sea despreciable desde el punto de vista practico, entonces es practicamente seguro que en un único experimento binomial de n_o pruebas la desviación de $\frac{X}{n}$ respecto a p será inferior a ϵ. Esta interpretación supone la confirmación teorica de la noción intuitiva de la probabilidad a traves de la denominada ley empirica de la estabilidad de las frecuencias. Para una ilustración geometrica o gráfica del teorema de Bernoulli vease el libro de don Sixto .

b) Relaciones entre las colas de la binomial, normal y Poisson.

Las relaciones que veremos a continuación muestran que tanto la aproximación normal (A.1) como la de Poisson (B.1) presentan una desviación sistemática respecto a las probabilidades reales, que en el caso particular de la

cola derecha se traduce en que la distribución normal generalmente subestima las probabilidades reales mientras que la distribución de Poisson siempre las sobreestima.

Las desigualdades existentes entre las probabilidades binomiales y las de la ley de Poisson fueron establecidas por T. Anderson y S.Samuels(1965), quienes las obtuvieron a partir de propiedades de monotonía de las probabilidades binomiales.(De estas propiedades trataremos en el último apartado de la sección). Las desigualdades en cuestión están contenidas en los siguientes resultados:

Teorema.-

$$P(k;n\cdot p) > B(k;n,p) \qquad \text{si} \quad k \leqslant n\cdot p - \frac{n\,p}{n+1}$$

$$P(k;n\cdot p) < B(k;n,p) \qquad \text{si} \quad k \geqslant n\cdot p$$

En otras palabras, con la posible excepción del valor central, la distribución de Poisson siempre sobreestima las colas binomiales.

Teorema.- Sean $k_\lambda^{\triangle} = \lambda + \frac{1}{2} - \sqrt{\lambda+\frac{1}{4}}$ y $k_\lambda^{\star} = \lambda + \frac{1}{2} + \sqrt{\lambda+\frac{1}{4}}$. Si $0 \leqslant k \leqslant k_\lambda^{\triangle}$ ó $k_\lambda^{\star} \leqslant k \leqslant n$, entonces se verifica que

$$p(k;\lambda) > b(k;n,\frac{\lambda}{n})$$

Como corolario se desprende el siguiente resul

tado sobre el error de la aproximación de Poisson:

$$\max_{0\leq k\leq k^{\wedge}_{\lambda}} \left\{ P(k;\lambda) - B(k;n,\frac{\lambda}{n}) \right\}$$

se alcanza cuando $k = k^{\wedge}_{\lambda}$, y

$$\min_{k^{*}_{\lambda}\leq k\leq n} \left\{ P(k;\lambda) - B(k;n,\frac{\lambda}{n}) \right\}$$

se alcanza para el menor valor de k. Por tanto, el maximo error absoluto se alcanza sobre dos enteros que verifican que $k^{\wedge}_{\lambda}\leq k'\leq\lambda\leq k''\leq k^{*}_{\lambda}$.

Como aplicaciones de estas desigualdades, Anderson y Samuels obtienen algunos resultados sobre la utilización de la ley de Poisson para simplificar los tests de hipotesis y los intervalos de confianza del parametro p. Los tests de hipótesis basados en la Poisson verifican, en casi todos los casos, que su nivel de significación real es menor que el nivel nominal dado por la Poisson, de modo que la probabilidad de error de tipo 1 se sobrestima. Analogamente, el nivel de confianza real de los intervalos aproximados es mayor que su nivel nominal.(Para el enunciado preciso de estos resultados vease la sección 4 del trabajo original.) Si tenemos en cuenta además las relaciones entre la binomial y la binomial generalizada, podemos obtener los mismos resultados para la probabilidad promedio de pruebas de Poisson.

Las desigualdades entre las probabilidades de la binomial y la normal han sido probadas por E. Slud en 1977; fueron sugeridas por R.M.Dundley, quien había probado, en comunicación privada, su validez para la aproximación de Peizer-Pratt cuando $p \leqslant \frac{1}{4}$. Slud ha ampliado el campo de validez de la desigualdad basandose en un teorema limite debido a Chernoff(1952), que le permite conocer el sentido de la desigualdad cuando n es suficientemente grande.

Los resultados de Slud son menos generales que los anteriores y además de demostración mucho mas laboriosa. Dichos resultados están presentados considerando que p está comprendido entre 0 y 1/2. Para su enunciado resulta útil volver a emplear la notación $z_o(k)$ para designar el valor tipificado de k, es decir $\frac{k-np}{\sigma}$. Veamos el resultado mas importante.

Teorema.- Si $np \leqslant k \leqslant nq$, entonces

$$B(k-1;n,p) \leqslant \Phi(z_o(k))$$

Además si $p \leqslant \frac{1}{4}$, dicha relación sigue verificandose cuando $k > nq$.

En otras palabras, este teorema afirma que la dis

tribución normal, bajo esas condiciones, subestima las probabilidades de las colas superiores de la binomial. De este resultado Slud también extrae aplicaciones estadisticas similares a las de Anderson y Samuels . Seguidamente veremos otros dos resultados de Slud que muestran que las probabilidades binomiales puntuales son mayores que las deducidas por la fórmula habitual no corregida basada en la normal. El segundo resultado es mas importante, pues da una cota inferior de esa diferencia.

Teorema.- Si $np \leq k \leq nq$ y $k \geq 2$, entonces

$$b(k;n,p) \geq \Phi(z_0(k+1)) - \Phi(z_0(k))$$

Teorema.- Si $p \leq \frac{1}{4}$, entonces

$$b(k;n,p) \geq \Phi(z_0(k+1)) - \Phi(z_0(k)) + \frac{\gamma_i}{\sigma} \cdot \varphi(z_0(k)) \quad , i=1,2$$

siendo $\gamma_1 = \frac{1}{2} \cdot \left\{z_0(k)\right\}^2 \cdot \left\{\frac{1}{k} - \frac{1}{3\sigma}\right\}$ cuando $k \geq \frac{n}{3}$ y $n \geq 27$, y

$\gamma_2 = \max(\gamma_1, 0.16)$ cuando $k > nq$.

Aunque con un objetivo diferente, en el siguiente punto veremos también desigualdades entre las probabilidades binomiales y normales.

c) Acotaciones basadas en la aproximación normal.

Como vimos en la sección 11, la aproximación normal adolece del grave inconveniente de presentar un error relativo que crece rapidamente conforme la variable binomial se aleja de su valor central. Esto trae como resultado que a partir de ciertos valores las probabilidades aproximadas pierdan todo su valor como estimación de las verdaderas. Por tanto, el conocimiento de la calidad de una aproximación concreta requiere la estimación precisa de ese error; pero, desgraciadamente, el estudio de dicho error presenta un tratamiento matemático dificil.

Vamos a ver aquí un artículo de Feller del año 1945 en el que se dan soluciones satisfactorias para superar esta dificultad. Estas soluciones surgen de la adopción de nuevos puntos de vista como son: intentar determinar nuevos limites normales que den lugar al menor error relativo o,al menos, a una acotación precisa de la probabilidad real; o bien,modificar la aproximación normal de manera que sea tratable la estimación del error relativo. Observese que ambos puntos de vista no se excluyen mutuamente entre si.

Para determinar los nuevos límites nos conforma

remos con buscarlos a partir de expresiones manejables tales como polinomios de 2º grado en los límites normales habitua les o ecuaciones del mismo grado en los límites buscados. Se gún Feller, parece que el único autor que ha adoptado este punto de vista es S. Bernstein, quien tras un gran esfuerzo de cálculo ha llegado finalmente al siguiente resultado.

Teorema.- Sean α_x y β_x las soluciones de las ecuaciones de 2º grado

$$x - \frac{3}{2} - np = \alpha_x \sqrt{npq} + \frac{q-p}{6}\cdot\alpha_x^2$$

$$x + \frac{1}{2} - np = \beta_x \sqrt{npq} + \frac{q-p}{6}\cdot\beta_x^2 \quad .$$

Si $\sigma^2 \geqq 62{,}5$, $\alpha \geqq 0$ y $\beta \leqq \sqrt[6]{8}\,\sigma$, entonces

$$\Phi(\beta_k) - \Phi(\beta_j) \leqq B(j-1,k) \leqq \Phi(\alpha_k) - \Phi(\alpha_j)$$

Las condiciones sobre α y β son practicamente equivalentes a que $j \geqq np + \frac{3}{2}$ y $k \leqq np + 2^{1/2}\cdot\sigma^{1/3}$. Una ca- racteristica destacable de este resultado es que el error cometido es mantiene como $O(\frac{1}{\sigma})$ aunque j y k se incremen- ten con σ.

Por otra parte, Feller considera a la vez ambos puntos de vista, consiguiendo así mejorar el teorema de Bern

stein. Su modificación de la aproximación normal consiste en remplazar el valor $z_0(k)$ por

$$x_k = \frac{k+1/2-(n+1)p}{\{(n+1)pq\}^{1/2}} \quad ,$$

al cual aplicaremos la corrección por continuidad cuando se use como argumento de la función de distribución normal. Con esta modificación se consigue desarrollar en serie de potencias de x_k el logaritmo de la probabilidad b_k. Para llegar a ello es necesario aplicar la fórmula de Stirling en dos versiones, una para $n!$ y otra para $k!$ y $(n-k)!$. Notando por $\sigma_\star$ a $\sqrt{(n+1)pq}$ y suponiendo que k y (n-k) son mayores que 3, tenemos que b_k se puede expresar de la siguiente manera

$$b_k = \frac{1}{\sigma_\star\sqrt{2\pi}} \cdot \exp\left\{ - \sum_2^\infty \left\{ \frac{p^{\nu-1} - (-q)^{\nu-1}}{\nu(\nu-1)} \right\} \cdot \frac{x_k^\nu}{\sigma_\star^{\nu-2}} \right. +$$

$$\left. + \frac{1}{24\sigma_\star^2} \sum_3^\infty \left\{ p^{\nu-1} - (-q)^{\nu-1} \right\} \left(\frac{x_k}{\sigma_\star}\right)^{\nu-2} + \frac{1+2pq}{24\sigma_\star} - \rho \right\}$$

donde el término residual ρ está comprendido entre 0 y $\frac{1}{300.\sigma_\star^6}$.

Este desarrollo junto con la fórmula sumatoria de Euler-MacLaurin nos permiten aproximar B(j-1,k) por

$$\Phi(x_k + \tfrac{1}{2\sigma_\star}) - \Phi(x_j + \tfrac{1}{2\sigma_\star}) \quad .$$

Feller deja para el lector la tarea de estimar el error rela

tivo y, en su lugar, pasa a demostrar un teorema similar al de Bernstein, utilizando como límites polinomios de 2º grado en x_k. Su resultado es el siguiente:

Teorema.- Supongamos que $\sigma_\star \geq 3$, $j \geq (n+1)p$ y $x_k \leq \frac{2}{3}\sigma_\star$. Si definimos η_i y ζ_i como

$$\eta_i = \frac{k-(n+1)p}{\sigma_\star} + \frac{p-q}{6\sigma_\star}\left\{\frac{k-(n+1)p}{\sigma_\star}\right\}^2 + \frac{p-q}{3\sigma_\star} - \frac{1}{2\sigma_\star^2} ,$$

$$\zeta_i = \eta_i + \frac{7\cdot M + 6}{42\sigma_\star} + \frac{1}{2\sigma_\star^2}$$

con M igual a $\frac{x_k^3}{\sigma_\star}$, entonces se verifica

$$\Phi(\zeta_{k+1}) - \Phi(\zeta_j) \leq \frac{B(j-1,k)}{A} \leq \Phi(\eta_{k+1}) - \Phi(\eta_j)$$

donde A vale $\exp\left(\frac{5(1-pq)}{36\sigma_\star^2}\right)$.

Este teorema tiene un campo de validez superior al del teorema de Bernstein y además proporciona límites con un menor margen de separación entre ellos. Por otra parte, intercambiando los papeles de p y q , dichos teoremas se verifican para valores de j y k inferiores a la media.

d) Monotonía de las probabilidades binomiales.

Vamos a ver aquí que tanto la función de distribución binomial como sus probabilidades puntuales presentan un comportamiento monótono respecto al parametro n cuando p cumple ciertas restricciones. De todos los resultados que veremos se han desprendido aplicaciones importantes, algunas de las cuales ya han sido tratadas.

Las propiedades sobre la función de distribución han sido publicadas por Anderson y Samuels (1965), K. Jogdeo (1967) y Jogdeo y Samuels (1968). La demostración de Jogdeo es directa, resultando algo engorrosa; en cambio, las otras demostraciones resultan rápidas haciendo uso de la distribución binomial generalizada y el 4º teorema de Hoeffding. Los resultados, por orden cronologico, son los siguientes:

Teorema.- Si la media de la binomial se mantiene fija, entonces al aumentar n, B(k) crece estrictamente siempre que $k \leq np-1$ y decrece estrictamente cuando $k \geq np$.

Teorema.- Sea S_n una variable binomial $B(2n+1, \frac{n+1}{2n+1})$, entonces la sucesión

$$\left\{P[S_{2n+1} \leq n]\right\}, \quad n=1,2,.....$$

converge crecientemente hacia $\frac{1}{2}$.

Teorema.- Si al aumentar el parámetro n, el parámetro p varía de tal modo que los productos np y nq sean valores enteros no decrecientes, entonces $B(np+r;n,p)$ crece estrictamente cuando r es un entero negativo, y decrece estrictamente cuando es no negativo.

Observese que este teorema engloba siempre al anterior; en cambio, el teorema de Anderson y Samuels solo queda incluido cuando np sea entero.

Sobre la monotonía de las probabilidades puntuales solo tenemos noticia de este resultado de Anderson y Samuels :

Teorema.- Sean $k_\lambda^\Delta = \lambda+\frac{1}{2} - \sqrt{\lambda+\frac{1}{4}}$ y $k_\lambda^\star = \lambda + \frac{1}{2} - \sqrt{\lambda+\frac{1}{4}}$. Si $0 \leqq k \leqq k_\lambda^\Delta$ o $k_\lambda^\star \leqq k \leqq n$, entonces $b_k(n,\frac{\lambda}{n})$ es una función estrictamente creciente de n. Si $k_\lambda^\Delta < k < k_\lambda^\star$ y n es suficientemente grande, entonces $b_k(n,\frac{\lambda}{n})$ decrece respecto a n.

Finalmente vamos a enumerar las aplicaciones que se han extraido de estos resultados. Las aplicaciones ya recogidas son las relaciones entre binomial y Poisson (vistas en b)) y la localización de la mediana de la binomial generalizada. El teorema de Jogdeo se ha empleado para generali

zar un ejemplo de Lehmann que mostraba que el estimador de maxima verosimilitud basado en una única prueba de Bernoulli, cuando se sabe, a priori, que p está comprendido entre $\frac{1}{3}$ y $\frac{2}{3}$, resulta uniformemente peor que determinados estimadores. Jogdeo demuestra que esto sigue siendo cierto cuando se considera un número impar de pruebas de Bernoulli.

FACULTAD DE CIENCIAS
BIBLIOTECA

CAPITULO II

INTRODUCCION INTUITIVA A LAS VARIABLES BINOMIALES METRICAS

II.1 Introducción

II.2 Definiciones Auxiliares

II.3 Introducción a las variables binomiales metricas

II.4 Modelos que originan variables binomiales metricas

II.1.- INTRODUCCION

Vamos a dedicar este capítulo a mostrar cómo surgen y se definen las variables aleatorias que hemos convenido denominar variables binomiales metricas. También expondremos tres modelos que se adaptan a situaciones perfectamente posibles en la práctica y en los que se hallan involucradas estas distribuciones de probabilidad.

El procedimiento que vamos a seguir para la introducción de estas variables discretas va a estar basado en un modelo perteneciente al campo de la Teoría de la información. La razón de esta elección es, por una parte, de motivación, puesto que a partir de situaciones similares a las que describiremos aquí llegamos a intuir la idea motriz de la presente memoria; por otra parte, porque supone una forma sencilla y clara de comprender su fundamento matemático. No obstante, es perfectamente posible establecer dicha introducción a partir de determinados experimentos estadisticos, en virtud de la especial dualidad existente entre la Teoría de la Información y la Estadistica Matematica.

Antes de todo eso, daremos primeramente la definición de algunos conceptos de Teoría de la Información que utilizaremos en la formulación de los modelos. Para ello seguiremos el libro de Abramson (1974); definiciones ligeramente diferentes pueden encontrarse en el libro de P. Gil Alvarez (1981) .

II.2.- DEFINICIONES AUXILIARES

Las definiciones que vamos a dar aquí son las siguientes: Fuente de memoria nula, Canal de información, Canal binario simetrico y Extensión del canal binario simetrico.

a.- Fuente de memoria nula

Se entiende por fuente de memoria nula a toda fuente de información tal que el resultado de toda nueva emisión no depende en absoluto de las emisiones realizadas anteriormente. Es decir, estadisticamente hablando, el concepto de memoria nula es esencialmente equivalente al concepto de independencia de sucesos aleatorios.

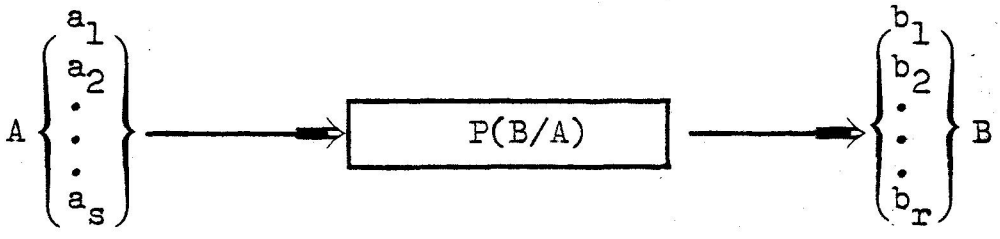
Utilizaremos fuentes binarias de memoria nula en la descripción del segundo y tercer modelo .

b.- Canal de Información

Emplearemos este concepto para la introducción intuitiva de la familia de variables binomiales metricas,y en la descripción del primer modelo.

Veamos, entonces, que es lo que se entiende por canal de información. Un canal de información, con alfabeto de entrada $A=(a_1,a_2,.....,a_s)$ y alfabeto de salida $B=(b_1,b_2,.....,b_r)$, es un medio de transmisión de información tal que al suministrarle un simbolo de entrada a_i produce como salida un simbolo b_j con una cierta probabilidad condicional a priori $p(b_j/a_i)$.

Esquematicamente lo podemos representar así :



Según la definición dada, cada canal de información lleva asociada una matriz de probabilidad P(B/A) con

la cual queda totalmente determinado salvo la manera de elegir los simbolos de los alfabetos de entrada y salida, lo cual es del todo irrelevante. Esta matriz es una matriz estocastica,o también llamada de Markov, ya que la suma de los elementos de cada fila vale siempre la unidad.

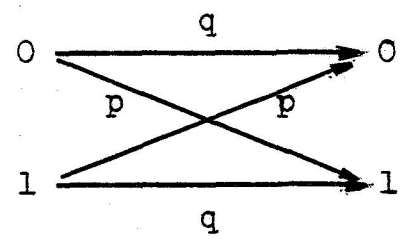
c.- Canal Binario Simetrico

Un caso teorico importante de canal de información es el denominado canal binario simetrico, que abreviaremos por CBS . En este canal existen unicamente dos simbolos de entrada y salida, que notaremos para ambos lados por 0 y 1 , cuya transmisión se realiza de tal manera que las probabilidades condicionales a priori verifican la siguiente igualdad :

$$P(1/0) = P(0/1) = p$$

El calificativo de simetrico procede de esta condición, que significa que la probabilidad de no recibir el simbolo transmitido es independiente del mismo. De esta propiedad se deduce evidentemente que la matriz del canal es simétrica verificandose, por tanto, que la suma de los elementos de las columnas también vale la unidad, es decir la matriz del canal es doblemente estocastica.

Esquematicamente representamos a este canal de la siguiente manera :



siendo $q=1-p$, y suponiendo que p es menor o igual que 1/2 . La matriz del canal CBS es

$$P = \begin{bmatrix} q & p \\ p & q \end{bmatrix}$$

suponiendo las entradas y salidas ordenadas crecientemente.

Este canal representa un caso típico de transmisión de información bajo ruido, puesto que no es posible saber con certeza cúal fue el simbolo realmente emitido por la fuente, una vez conocido el simbolo recibido. Por esta razón, la probabilidad p puede interpretarse como la probabilidad de recibir un simbolo erroneamente; es decir ,

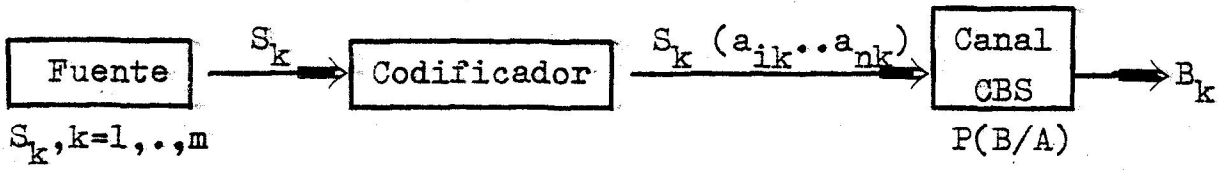
$P(0/1) = P(1/0)$ = Probabilidad de transmisión erronea .

d.- Extensión del canal binario simetrico

En principio, mediante un canal CBS sólo pueden transmitirse dos informaciones distintas; no obstante, por medio de sucesiones de simbolos binarios, es posible utili-

zarlo para la transmisión de cualquier número de mensajes fuente. Concretamente esto significa, en primer lugar, codificar cada uno de los mensajes fuente mediante un codigo binario adecuado y , a contunuación, proceder a la transmisión a través del canal de cada uno de los símbolos binarios que integran la palabra codigo correspondiente al mensaje deseado.

El proceso queda representado, esquematicamente, de la siguiente manera :

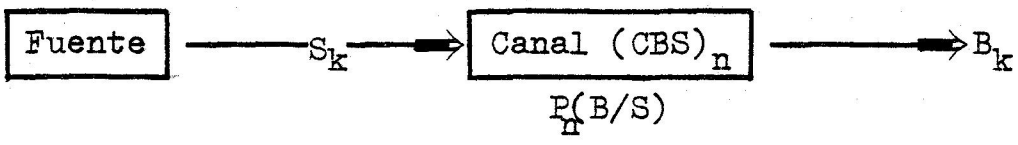


Como procedimiento de codificación nosotros eligiremos cualquier codigo univoco, con tal de que cada mensaje fuente venga representado por palabras de una misma longitud n. Por este procedimiento, sabemos que podrían transmitirse hasta un total de 2^n mensajes fuente diferentes por medio de un canal binario simétrico.

Según el proceso de transmisión que estamos siguiendo, el canal CBS se emplea n veces para cada mensaje fuente. No obstante, podemos englobar la codificación de los mensajes fuente en secuencias de longitud n junto con

su transmisión a través del CBS en un único nuevo canal. A este canal se le denomina extensión de orden n del canal CBS, y lo notaremos por $(CBS)_n$.

De forma grafica, el esquema anterior queda simplificado en el siguiente :



Como es natural, el canal $(CBS)_n$ está identificado por una matriz estocastica de orden 2^n, puesto que es ese valor justamente el número maximo de entradas y salidas posibles del canal. Dicha matriz puede calcularse directamente, analizando ordenadamente todos los posibles pares de entradas y salidas; o bien, a partir de la matriz del canal CBS mediante el cálculo de la denominada potencia tensorial de orden n de la matriz P. Por ejemplo, la matriz del canal $(CBS)_2$ es el cuadrado tensorial, o también llamado de Kroneker, de la matriz P , cuyo valor viene dado por la siguiente relación:

$$P_2 = P^{(2)} = \begin{bmatrix} qP & pP \\ pP & qP \end{bmatrix} = \begin{bmatrix} qq & qp & pq & pp \\ qp & qq & pp & pq \\ pq & pp & qq & qp \\ pp & pq & qp & qq \end{bmatrix}$$

Esta matriz corresponde a las entradas y salidas pertenecientes al producto cartesiano $\{0,1\}\times\{0,1\}$, ordenados sus elementos en el cuadro resultante primero por filas, y luego en cada fila,de izquierda a derecha; en el caso particular de P_2 , sus filas corresponden a las entradas 00, 01, 10 y 11 , respectivamente. Observese que este procedimiento de ordenación es completamente equivalente a la ordenación numérica creciente de los números naturales representados por cada secuencia .

En general, ordenando las entradas y salidas de este modo, la matriz de la extensión enesima de un canal CBS viene dada recurrentemente a partir de la matriz de la extensión anterior por el siguiente producto:

$$P_n = \begin{bmatrix} qP_{n-1} & pP_{n-1} \\ pP_{n-1} & qP_{n-1} \end{bmatrix}$$

II.3.- INTRODUCCION A LAS VARIABLES BINOMIALES METRICAS

Para esta introducción informal, vamos a detenernos por algún tiempo en el canal $(CBS)_2$. Consideremos

una cualquiera de las posibles entradas, por ejemplo, la 01 . Las posibles salidas correspondientes a esa entrada son, naturalmente, todas las respuestas del canal, esto es: 00, 01, 10, 11. Sus respectivas probabilidades a priori respecto de la referida entrada están escritas en la segunda fila de la matriz del canal.

Comparemos cada una de las salidas con la entrada 01 , fijandonos en la fidelidad de la transmisión. Así, por ejemplo, se observa que al recibir 00 se ha cometido un error de transmisión, ya que el segundo 0 debería haber sido un 1 ; analogamente, al recibir 01, 10, 11 se han cometido respectivamente ningún, dos y un errores de transmisión.

a.- Distancia de Hamming .

De la idea de contabilizar el número de errores que contiene una determinada secuencia de salida respecto a la verdadera entrada a un canal ruidoso, se desprende inmediatamente la noción de distancia de Hamming. Esta distancia fue introducida en el año 1950, cuando se inició el estudio de los codigos correctores y detectores de errores ; nosotros la emplearemos para dar la definición rigurosa de la familia de variables binomiales metricas.

La definición de esta distancia es la siguiente: Sean A y B dos secuencias de símbolos binarios de la misma longitud n , la distancia de Hamming D(A,B) entre ambas secuencias es el numero de diferencias existente entre ambas, componente a componente. Matematicamente puede escribirse como

$$D(A,B) = \sum_{i=1}^{i=n} (a_i \oplus_2 b_i) = \sum_{i=1}^{i=n} D_d(a_i,b_i)$$

donde $\oplus_2$ representa la suma módulo 2 y D_d la distancia discreta aplicada al conjunto de las posibles salidas de cada componente de la palabra código. La distancia discreta, llamada así porque genera la topologia discreta, viene definida como :

$$D_d(x,y) = \begin{cases} 0 \text{ , si } x=y \\ 1 \text{ , si } x \neq y \end{cases}$$

Empleando este ultimo concepto, se obtiene una definición de la distancia de Hamming válida para cualquier canal cuyo alfabeto de codificación no sea binario.

Facilmente se puede comprobar que la distancia de Hamming es una verdadera distancia en el sentido matematico, resultado que también se deduce inmediatamente de la ultima definición, por ser la distancia discreta una

verdadera distancia, y la de Hamming, definida de esa manera, una de las posibles metricas definibles sobre un producto cartesiano de espacios.

b.- Variable numero de errores de transmisión .

De lo que hemos visto hasta ahora, resulta completamente natural definir sobre el espacio de las salidas de un canal una variable aleatoria que mida el número de errores cometidos en una transmisión respecto de la verdadera entrada; de esta manera, en el caso particular de la entrada 01 , definiriamos la variable V_{01} , cuya imagen será $D(s,01)$, para cualquier elemento s del conjunto de salida S_2 ; V_{01} tomará unicamente los valores 0,1,2 con probabilidades q^2 , $2pq$, p^2 , respectivamente. Por supuesto, en el caso de los canales $(CBS)_n$, las variables así definidas a partir de cada entrada son todas iguales, y siguen claramente una ley binomial $B(n,p)$, cuyos parametros son el orden de la extensión del canal binario simétrico y la probabilidad de recibir erroneamente cualquier símbolo.

Ahora bien, quizás podría ser importante conocer la familia de variables aleatorias obtenidas por este procedimiento a partir de cualquier canal, binario o no binario. Como resultado de esta memoria, la obtención de ta-

les distribuciones de probabilidad no presentará ninguna dificultad. No obstante, nosotros estamos interesados en estudiar una familia mas amplia, que en el caso de los canales $(CBS)_n$ se concreta en lo que hemos denominado variables binomiales metricas.

c.- Variables Binomiales Metricas .

Para llegar a la generalización de la variable que acabamos de definir, vamos a asumir el papel de receptor del canal de información. Nosotros, desde este punto de vista, generalmente, no vamos a poder conocer con certeza el mensaje fuente emitido, sino solo una distribución de probabilidad sobre el conjunto de entradas al canal, que será el resultado de aunar la información que contenga la matriz del canal con la que dispongamos sobre la fuente. En estas condiciones, no va a ser posible definir la variable anterior. Sin embargo, podemos obviar esta dificultad reemplazando la secuencia emitida por cualquier entrada que haga de referencia,respecto de la cual midamos el número de errores que contiene la salida producida. La elección de esta entrada de referencia se puede realizar por cualquier procedimiento; por ejemplo, considerar la moda de la distribucíon de las entradas; naturalmente, la entrada elegida puede coincidir, por casualidad, con la ver-

daderamente emitida.

De esta forma, en la definición de la nueva variable intervienen tres secuencias diferentes: la realmente emitida, la entrada de referencia,(que podemos tratar como la supuestamente emitida), y la realmente recibida. La primera condiciona la distribución de probabilidad de las salidas y la segunda sirve para medir el número de errores. Para concretar la definición, sigamos con el ejemplo de la emisión de la secuencia 01 por el canal $(CBS)_2$.

Consideremos como entrada de referencia a la secuencia 00 . La variable aleatoria que así definimos vendrá notada por V_{01}^{00} , y su imagen es el número de diferencias que existen entre la salida recibida y la entrada de referencia 00, cuando la secuencia realmente emitida es la 01 ; es decir

$$V_{01}^{00} : S_2 \longrightarrow \mathcal{R}^+ \quad / \quad V_{01}^{00}(s) = D(s,00) \quad ,$$

estando las probabilidades de cada imagen determinadas a partir de la fila de la matriz del canal correspondiente a la entrada 01. De este modo, tenemos que las probabilidad de cada imagen elemental valen :

$$P[V_{01}^{00} = 0] = qp \quad ,$$

ya que este suceso ocurre unicamente cuando se produce la salida 00, cuya probabilidad respecto de 01 es qp (coincidencia en el primer dígito y diferencia en el segundo).

$$P[V_{01}^{00} = 1] = q^2 + p^2 \quad ,$$

el primer sumando procede de la salida 01 y el segundo de 10 .

$$P[V_{01}^{00} = 2] = pq \quad ,$$

correspondiente a la secuencia 11 .

Si consideramos las restantes entradas como mensajes supuestamente emitidos, obtendriamos las variables V_{01}^{01} , V_{01}^{10} y V_{01}^{11} . Observese que la primera de estas coincide con V_{01} y que la última tiene igual distribución que V_{01}^{00} , puesto que estas entradas son completamente equivalentes respecto de la realmente emitida. En el capítulo siguiente demostraremos que este resultado se verifica para todas las variables definidas por entradas cuya distancia sea la misma respecto de la verdadera entrada.

Si hubiesemos razonado utilizando otra entrada

como señal realmente emitida, la situación se habría repetido exactamente, obteniendose unicamente las mismas distribuciones. En virtud de la simetría de la transmisión, manteniendo siempre fija la entrada al canal y considerando unicamente los distintos valores que puede tomar la distancia de Hamming respecto a dicho punto, se llega a la definición de la familia de variables aleatorias que deseamos estudiar. Para la extensión de segundo orden de un canal binario y simetrico, la familia consta de tres variables aleatorias X_0 , X_1 y X_2 , de las que diremos que siguen, respectivamente, una distribución binomial metrica de parametros (2,0,p) , (2,1,p) y (2,2,p).

En el caso general de un canal $(CBS)_n$ cuya probabilidad de transmisión erronea valga p , tenemos un conjunto de 2^n variables aleatorias definidas sobre el espacio muestral de las 2^n posibles salidas del canal, que se clasifican en una familia de n+1 variables binomiales diferentes. Si fijamos una secuencia R como única verdadera entrada al canal, con lo que quedan determinadas todas las probabilidades del espacio de salidas, diremos que una distribución sigue una ley binomial metrica B(n,d,p) si coincide con la distribución de V_R^A, engendrada por la señal A cuya distancia a R, D(A,R), vale d; la variable V_R^A , defi-

nida sobre el espacio de salidas, mide la distancia de Hamming entre la salida recibida S y la señal A, D(A,S/R).

d.- Comentario .

Añadimos el presente comentario para analizar la razón por la cual creemos que la definición de las variables binomiales metricas no ha surgido de manera inmediata en el seno de la Teoría de la Información .

Hasta antes de su definición hemos mantenido la misma linea que se sigue en esta Ciencia para exponer metodologicamente la materia abarcada por el segundo teorema de Shannon. Mas concretamente, nosotros nos hemos guiado por el libro de Abramson .

Sin embargo, llegados a ese punto nos separamos de esa linea adoptando un enfoque diferente del que allí se mantiene. En la Teoría de la Información, uno (particularmente el ingeniero de comunicaciones) está interesado en la maxima transmisión fiable de la información que circule a traves de canales no fiables; para lograrlo, se dispone de gran libertad al poder influir tanto sobre el emisor como sobre el receptor. En esta situación privilegiada, respaldados por el segundo teorema de Shannon, actua sobre la

fuente emisora construyendo unos codigos especiales que le aseguran el objetivo buscado.

Esta visión global de la transmisión no puede ser compartida por nosotros, puesto que no somos la fuente emisora; esta es externa y nos viene impuesta. En tal circustancia, solo podemos adoptar el punto de vista del receptor, siendo nuestro objetivo obtener el maximo conocimiento sobre el canal ruidoso y sobre la verdadera señal emitida. Nuestra situación es semejante a las que se tratan en le Teoria del filtrado de procesos estocasticos. Este enfoque constituye el punto de partida original de la presente memoria. Lo consideramos útil puesto que el ser humano no controla todas las fuentes de información, y gracias a nuestro planteamiento se cubren situaciones teoricas nuevas, no contempladas por el planteamiento corriente de la Teoría de la Información; de hecho ya existen algunas situaciones reales que se ajustan a todo este esquema, y con seguridad aparecerán otros nuevas .

En la siguiente sección vamos a exponer unos modelos junto con algunas realizaciones mas o menos reales de los mismos en las que se hallan envueltas las distribuciones binomiales metricas .

II.4.- MODELOS QUE ORIGINAN VARIABLES BINOMIALES METRICAS

En esta sección vamos a plantear tres modelos diferentes. En el primero utilizaremos canales con ruido con el objeto de plasmar en un modelo lo que hemos expuesto en la sección anterior. Para la descripción de los dos modelos restantes en lugar de emplear canales ruidosos, supondremos que la transmisión es fiable y consideraremos fuentes de información de memoria nula .

En todos los modelos va a ser común la consideración de secuencias binarias de longitud n , de las cuales solo vamos a conocer su distancia de Hamming respecto a una cierta secuencia de referencia R. Desde el punto de vista de la Estadistica Matematica, estamos manejando condensaciones muestrales producidas por estadisticos relacionados con la distancia de Hamming .

Después de leer los modelos se podría argumentar que la mayor parte de la dificultad que estos reflejan procede de estas condensaciones de información, lo cual sería cierto unicamente si los estadisticos utilizados no fuesen suficientes. En ese caso, se nos podría criticar de estar complicando innecesariamente las cosas, al desperdiciar to

da la información contenida en la sucesión de ceros y unos que integran las secuencias observadas.

Sin embargo, queremos resaltar que estas condensaciones pueden ser inevitables, ya sea por venir impuestas a priori por la Naturaleza, o bien ser consecuencia del gran volumen de información que se tendría que manejar de otra manera. Por estas razones y enlazando con el comentario anterior, creemos que la creación de las distribuciones binomiales metricas no son una complicación innecesaria ni una sofisticación de los modelos ya existentes, los cuales no dan cuenta ni resuelven, por ejemplo, los problemas que se derivan de los modelos que ahora expondremos .

Nuestro objetivo en todos ellos se centrará en la determinación de dos cantidades: una probabilidad, la de transmisión erronea o la de emisión del simbolo 1; y una distancia, la existente entre la secuencia R y un cierto punto del espacio muestral.

A.- Recepción de una señal desconocida bajo ruido .

Ante la transmisión de una señal binaria desconocida por un canal con ruido simetrico, nuestro principal ob

jetivo se centrará en el descubrimiento de dicha señal. Si suponemos que la señal puede observarse indefinidamente, un sencillo procedimiento para lograr esto consistiría en ir anotando las sucecivas señales recibidas, o simplemente ir recontando la frecuencia de los distintos tipos de símbolos recibidos en cada posición de la secuencia, hasta que el símbolo mayoritario en cada compomente exceda en una cuantia estadisticamente significativa al minoritario . La estimación de la verdadera señal emitida estaría compuesta por estos símbolos mayoritarios. En el caso general, desconoceremos la probabilidad p de transmisión erronea, de modo que la decisión estadistica podemos adoptarla para la situación de maxima incertidumbre, que se presenta cuando p está muy proximo a su valor maximo, $p_o=1/2$. Observese que valores de p superiores a 1/2 tienen un efecto equivalente a la emisión del símbolo contrario con probabilidad de transmisión erronea inferior a 1/2 ; cuando p toma exactamente dicho valor, resulta imposible la estimación de la señal fuente puesto que a posteriori todas las posibles emisiones serían equiprobables.

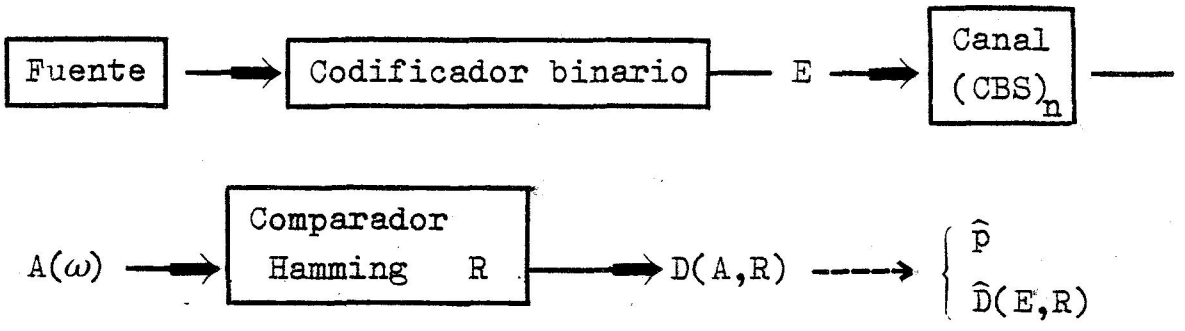
Nosotros vamos a presentar un modelo más restrictivo en el que la señal recibida no se va a poder conocer componente a componente, sino que nos vamos a limitar a un

tratamiento global de la misma, por el cual se resume toda la secuencia en un número que medirá su distancia a otra secuencia elegida por nosotros. Evidentemente en este planteamiento especial no es aplicable el procedimiento descrito, siendo por tanto necesario crear uno nuevo que se adapte a este modelo. En principio, creemos que no va a ser posible identificar completamente la señal realmente emitida, sino que solamente podremos estimar la distancia a nuestra secuencia de referencia. Después, una vez conocida la distancia se podría ir modificando la señal de referencia acercandonos sucesivamente a la señal fuente.

Ante este modelo nuestro proposito se centrará en la estimación conjunta de la probabilidad de error del canal y de la distancia existente entre la señal verdaderamente emitida y la de referencia.(Suponemos que normalmente se conocerá la longitud de las secuencias implicadas; no obstante, podriamos considerarla desconocida ampliando así nuestros objetivos con la estimación de esta cantidad) . Como ya vimos en la sección anterior, el resultado de cada recepción es una realización de una cierta distribución binomial metrica, precisamente aquella cuyos parámetros son las cantidades que deseamos estimar. Queda claro que el estudio de nuestra familia es la base para la

resolución de los problemas planteados por este modelo.

Esquematicamente, todo esto puede representarse de la siguiente manera :



Muchas situaciones reales podrian modificarse o simplificarse con objeto de sirvirnos de ilustración de este modelo; cualquier magnitud cuantificable puede transformarse en una magnitud binaria estableciendo un umbral separador. Sin embargo, en ninguna de las que nos vienen a la memoria encaja la restricción de que la observación deba limitarse a la medición de la distancia de Hamming, puesto que no hallamos nada que impida conocerla componente a componente.

A pesar de ello, obviando estas dificultades, vamos a presentar la siguiente situación,perfectamente posible en la realidad. Consideremos que el espacio exterior que rodea a un astro frente a unas hipoteticas emisiones del mismo se comporta como un canal binario simetrico, cuyo ruido puede imputarse a perturbaciones energéticas que

traen como consecuencia la permutación de los símbolos de un tipo en los del otro.

Como ya dijimos antes, las restricciones que este modelo impone sobre las observaciones de este ejemplo, tienen como efecto que no se pueda determinar directamente la señal emitida, sino solo verificar si coincide o no con la fijada como referencia. En caso de no coincidencia, sabemos que se obtendría una medida de la discrepancia que existe entre ambas, de cuyo valor concreto depende que dispongamos de mayor o menor información sobre dicha señal.

Por tanto, resulta claro que la elección adecuada de la señal de referencia juega un papel importante en este modelo. Generalmente dicha elección se producirá como consecuencia del conocimiento parcial, directo o indirecto, que dispongamos sobre el fenomeno de la emisión.

Normalmente sucederá que las emisiones no son un hecho aislado del astro, sino que existirán otras manifestaciones del mismo mutuamente relacionadas (en las que podemos incluir la observación imprecisa de esas emisiones por otros metodos distintos) , que junto con las teorías

cientificas vigentes permitirán prever la señal emitible; de modo que puede considerarse que la señal de referencia hace las veces de la que teoricamente se espera que emita el astro. Con el presente modelo, entonces, se puede confirmar o refutar la validez de las teorías cientificas empleadas.

Si desgraciadamente las emisiones del ejemplo fuesen un hecho aislado y/o nuestros conocimientos sobre el astro nulos, siempre podriamos fijar al azar la señal de referencia.

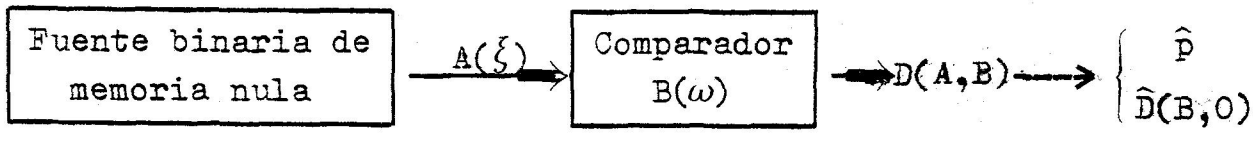
B.- Estimación de una fuente aleatoria desde una posición desconocida .

En este modelo y en el siguiente vamos a hacer uso de fuentes de información binarias de memoria nula. Recordando que su definición fue dada al comienzo del capítulo, consideremos una fuente de este tipo que emita los símbolos 1 y 0 con probabilidades fijas desconocidas p y q, respectivamente. Evidentemente, nuestra labor será la estimación de dichas probabilidades. Para ello vamos a observar cierto número de emisiones, con la particularidad de que solo podremos tratarlas en lotes de longitud n , los cuales se comparan con una cierta secuencia fija y des

conocida de la misma longitud, para dar, por supuesto, como único resultado para cada lote de observaciones su distancia de Hamming a la secuencia desconocida.

La presencia de esa secuencia desconocida trae consigo que tenga que generalizarse el proceso de estimación, que sin dicha presencia se reduciría simplemente a la estimación del parametro p de una distribución binomial. Sin embargo, al realizar esta nueva estimación puede conseguirse alguna información sobre la secuencia desconocida. Esa información se concreta en la distancia de Hamming a uno cualquiera de los dos lotes distinguidos, esto es, aquellas compuestos unicamente de ceros o unos.(Tales secuencias corresponden a los elementos de maxima y minima probabilidad del espacio probabilistico subyacente; cualquier secuencia situada a la misma distancia de ellas resulta indistinguible bajo este modelo).

En esquema el presente modelo queda expresado de la siguiente manera :



Vamos a describir una situación real de la física de partículas de alta energia, que debidamente simplificada y modificada, se puede considerar como ilustración de este modelo. Se trata de la investigación de la denominada radiación cósmica de maxima energia. Su estudio se realiza de modo indirecto por dos motivos: Primero, los detectores de partículas son muy voluminosos encontrandose solo a nivel del suelo y no en el espacio exterior de donde procede dicha radiación. En segundo lugar, la radiación interacciona fuertemente con la materia; al llegar a la atmosfera, estas partículas, denominadas ahora radiación cósmica primaria, sufren colisiones que producen la formación de nuevas partículas con suficiente energia para que este fenomeno se repita varias veces.(Como sabemos, desde el punto de vista del Cálculo de Probabilidades, este fenómeno estaría descrito por un proceso estocastico de ramificación, y como tal ya fue tratado este fenómeno en los años treinta). Como consecuencia de ello, la partícula primaria pierde su energia en el proceso de frenado y queda enmascarada dentro del conjunto de todas esas partículas formadas, que recibe el nombre colectivo de chubasco aereo de radiación secundaria.

Solo una parte del chubasco secundario se detecta a nivel del suelo por una red de contadores de partícu-

las. El problema físico consiste en, a partir del número de cuentas producido en cada detector y de la geometría de la red de detectores, obtener información sobre la partícula primaria responsable del chubasco aereo. Esta información se concreta en la determinación de la energia y la dirección de la partícula primaria, así como la altura sobre el nivel del suelo de la colisión desencadenante del chubasco. Para más detalles sobre este estudio, puede consultarse el número de Septiembre de 1978 de la revista divulgativa "Investigación y Ciencia" .

Vamos a mostrar las semejanzas y diferencia que existen entre el modelo propuesto y la situación que acabamos de describir; como podremos apreciar el grado de concordancia es suficientemente satisfactorio. Comencemos viendo las semejanzas.

Nuestro modelo puede considerarse como una versión n-dimensional de la situación fisica: el espacio muestral $\{0,1\}^n$ hace las veces de la superficie sobre la que se detecta el chubasco de radiación, de manera que cada secuencia de n emisiones puede interpretarse como las coordenadas del lugar del suelo donde se manifiesta una partícula secundaria. En el espacio muestral podemos fijar el sistema de referencia en el punto donde se manifestaría la partícu

la primaria si no sufriese ninguna colisión, de modo que la secuencia desconocida indica la posición relativa del sistema detector; dicho de otra forma, la secuencia desconocida juega un papel similar a la dirección del chubasco. Por otro lado, la probabilidad p de emisión puede considerarse equivalente a la altura, puesto que cuanto mayor son ambos, mayor dispersión existe entre las secuencias y las partículas del chubasco. El papel de la energia lo ocupa el tamaño del muestreo, ya que cuanto mayor sea la energia mayor número de partículas se crean y detectan.

En cuanto a las diferencias, podemos decir que se centran solamente en el proceso de detección. En la situación física se emplean varios detectores que son disparados por parte del chubasco. En cambio, en nuestro planteamiento solo se emplea un único detector especial cuyo alcance es ilimitado, de manera que todas las secuencias son comparadas.

Estas diferencias pueden eliminarse complicando nuestro modelo con la introducción de varios comparadores que detectaran unicamente aquellas secuencias que cayesen en sus proximidades .

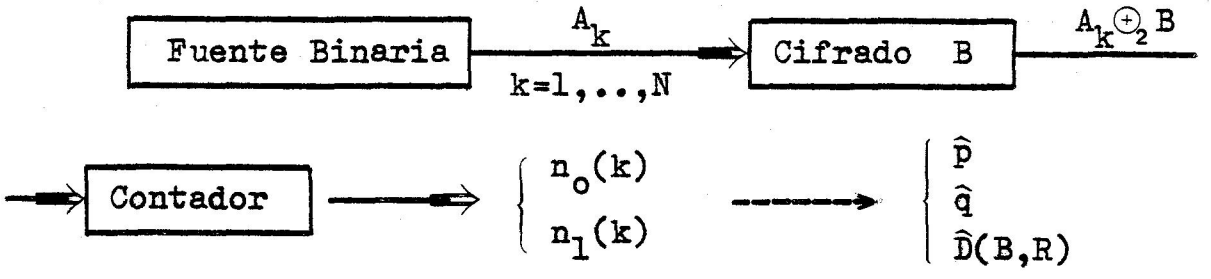
C.- Fuente aleatoria cifrada por un desplazamiento desconocido .

Finalmente, para terminar con estas situaciones modelo, consideremos una fuente binaria con memoria nula de orden n que, con el fin de enmascarar su verdadera aleatoriedad, somete sus emisiones a un proceso de cifrado por medio de una secuencia fija B, lo que trae como consecuencia que la frecuencia natural de los símbolos 0 y 1 sea alterada en cierta medida. El procedimiento de cifrado consiste simplemente en transmitir la secuencia $A \oplus_2 B$ en vez de la secuencia A, la realmente producida. A este tipo de cifrado se le suele denominar clave cesarea.

En estas condiciones, nuestro proposito será descubrir las verdaderas probabilidades de emisión y el mensaje B empleado en la clave cesarea. Como siempre,el proceso de muestreo que utilizaremos consistirá en recontar unicamente el número de símbolos de cada tipo que contienen cierto número de secuencias cifradas por la fuente. Como ya sabemos , la estimación exacta de B no es posible, puesto que, en virtud de la independencia de cada componente, toda reordenación de los simbolos de B da lugar a una secuencia cuyo efecto es estadisticamente equivalente a la propia B .

Esquematicamente, todo lo expuesto aquí puede re

sumirse en



Como se habrá observado éste y el modelo anterior son practicamente idénticos, diferenciandose unicamente en el pequeño detalle de en qué lugar, receptor o emisor, se encuentra englobada la secuencia desconocida. En el modelo anterior, dicha secuencia formaba parte del comparador; esto es, del receptor. En cambio, ahora el emisor fija la secuencia como clave cesarea. Es evidente que, a pesar de la gran similitud existente, habrá situaciones que se ajustarán mejor a un modelo que al otro.

Como ejemplo ilustrativo del presente modelo vamos a considerar una situación ficticia basada en la realidad biológica. Imaginemonos que un organismo pluricelurar, cuya estructura genetica se desconoce, ha estado sometido durante cierto tiempo a la acción de un agente inductor de mutaciones.

Este suceso podría utilizarse para medir el poder mutagénico de ese agente. Con el fin de ajustarnos al modelo, debe suponerse que el proceso de mutación verifica las siguiente condiciones: i) cada gen admite una única posible mutación; ii) la presencia de una mutación en un determinado gen es independiente de la presencia o ausencia de mutación en los restantes genes; iii) la probabilidad de que se presente mutación es una cantidad desconocida p constante para todos los genes.(Esta probabilidad se puede utilizar como medida del poder mutagénico).

Bajo estas condiciones, después de la exposición al agente, la estructura genética de cada célula puede identificarse con una señal binaria de una fuente de memoria nula que emita el simbolo 1 con probabilidad p. Supongamos que existiera una técnica rápida y poco costosa que determinara el número de genes en que se diferencian dos fragmentos de acidos nucleicos. Esta técnica, usada sistematicamente, podría utilizarse para estimar la probabilidad p.

Elijamos una célula cualquiera del organismo y extraigamos de ella un determinado fragmento cromosómico B, del cual fabricaremos un gran número de copias. Apliquemos la supuesta técnica para determinar el número de diferencias

que existen entre B y la secuencia genética A_k correspondiente a ese mismo fragmento cromosómico perteneciente a varias células del organismo. Evidentemente al proceder de esta manera, B hace las veces de clave cesarea y nuestra observación es el número de unos de la secuencia $A_k \oplus_2 B$.

Este procedimiento podría ser mas favorable, en terminos de tiempo de trabajo y costos, que la técnica alternativa que supone determinar la estructura genética del mismo fragmento cromosómico en varias células para deducir la secuencia genética original y estimar p. Observese que si solo interesa conocer la probabilidad p, el descubrimiento de la verdadera secuencia de genes resulta inutil. No obstante, si interesase conocer esta secuencia, podria ser mas ventajoso aplicar repetidamente nuestra técnica hasta localizar una secuencia de referencia que coincida con la original y proceder una sola vez a la determinación de su estructura gen por gen.

Finalmente queremos observar que estos modelos se han expuesto centrado la atención en la estimación conjunta de los parámetros p y d. Este será el problema de estimación que mas frecuentemente se presentará en esta nueva familia de distribuciones. Ocasionalmente, junto con ellos

habrá que estimar también el parámetro n. A pesar de que estos sean los problemas mas importantes, es evidente que los que se derivan del desconocimiento de cualquier otra combinación de parámetros tienen interes estadistico propio. El estudio de la estimación de los parámetros de la distribución binomial métrica será objeto de la sección 9 del capítulo siguiente.

CAPITULO III

VARIABLES BINOMIALES METRICAS

FACULTAD DE CIENCIAS
BIBLIOTECA
GRANADA

III.1 Definición

III.2 Función de Masa de Probabilidad

III.3 Funciones Generatrices

III.4 Momentos y Cumulantes

III.5 Otros Tipos de Momentos

III.6 Relaciones entre las Variables Binomiales Métricas

III.7 Suma y Diferencia de Variables Binomiales Métricas. Caracterización de la ley

III.8 Principales Propiedades

III.9 Entropía y Energía Informacional de las Variables Binomiales Generalizadas

III.10 Estimación de los Parámetros

III.11 Aproximación de las Probabilidades Binomiales Métricas

III.1.- DEFINICION

Como ya hemos visto en el tratamiento intuitivo de estas distribuciones, se trata de una familia dependiente de tres parametros cuyos significados ya son conocidos. Dichos parametros son: n, un número entero positivo cualquiera; p, un número real del intervalo $[0,1]$; d, un entero comprendido entre 0 y n, ambos inclusive.

Para dar la definición rigurosa de la familia, consideremos todos los elementos teoricos precisos para definir una variable aleatoria; esto es, establezcamos el espacio probabilistico basico $(\Omega,\mathcal{A},\mathcal{P})$, soporte de nuestra familia de variables aleatorias.

El espacio muestral Ω es el conjunto discreto de todas las n-uplas cuyas componentes sean solo ceros o unos. Concretamente Ω es la enesima potencia cartesiana del conjunto probabilistico mínimo no trivial, que podemos notar por conveniencia como $U=\{0,1\}$; o sea $\Omega=U \overset{n \text{ veces}}{\cdots\cdots} U$. El conjunto Ω se halla enriquecido por ciertas estructuras matematicas especialmente interesantes para nuestros propositos.

En primer lugar, es un espacio metrico inducido

por la distancia de Hamming, cuyo valor entre dos puntos $\vec{a}$ y $\vec{b}$, recordamos, está dado por

$$D(\vec{a},\vec{b}) = \sum_{i=1}^{n} (a_i \oplus_2 b_i)$$

Esta cantidad oscilará entre 0 y n.

La topología del espacio métrico está constituida por todos los conjuntos de Ω ; es decir, se trata de la topología discreta, que notaremos $\mathcal{C}(\Omega)$, puesto que cualquier conjunto puntual es una bola abierta. Evidentemente $\mathcal{C}(\Omega)$ coincide con la clase de conjuntos medibles $\mathcal{A}$, ya que como sabemos en un espacio discreto cualquier otra clase menor sería insuficiente puesto que trataría al menos un par de puntos distintos de Ω como si fuesen uno solo.

En segundo lugar, si consideramos el cuerpo binario de Boole, $(\mathcal{U}, +, \cdot)$, $(\Omega, \oplus, \cdot)$ tiene estructura de espacio vectorial de dimensión n sobre $\mathcal{U}$. Las definiciones de suma y producto por escalar en el espacio Ω vienen dadas, trivialmente, por :

Suma $\qquad \vec{a} \oplus \vec{b} = (a_i \oplus_2 b_i)$

Producto

$$0 \cdot \vec{a} = (0,.....,0) = \vec{0}$$
$$1 \cdot \vec{a} = \vec{a}$$

En virtud de esto último, el espacio Ω no solo es un espacio metrico, sino que también es un espacio vectorial topológico; mas aun, es un espacio normado, ya que es fácil comprobar que la metrica de Hamming sobre Ω es invariante por translación y dilatación; es decir, se verifica que $\forall \vec{a},\vec{b},\vec{c} \in \Omega$ y $\lambda \in \mathcal{U}$

$$\text{I} \quad D(\vec{a},\vec{b}) = D(\vec{a} \oplus \vec{c}, \vec{b} \oplus \vec{c})$$
$$\text{II} \quad D(\lambda\vec{a},\lambda\vec{b}) = |\lambda| D(\vec{a},\vec{b})$$

De esta invarianza sabemos que se deduce que la estructura de entornos alrededor de cada punto del espacio coincide exactamente con la que tiene el origen o cero de Ω. Esta importante propiedad será de utilidad puesto que nos permitirá establecer la función de masa de probabilidad de cada miembro de la familia siguiendo el mismo razonamiento empleado en la determinación de dicha función en la binomial clásica.

El conjunto de medidas de probabilidad $\mathcal{P}$ está compuesto por las medidas producto inducidas al considerar sobre cada componente de Ω la misma probabilidad elemental

dada por:

$$P[1] = p \qquad P[0] = 1-p = q$$

siendo p cualquier valor del intervalo [0,1] . Sin embargo, en algunas ocasiones será conveniente restringir p al intervalo $[0,1/2]$. Mas adelante explicaremos por qué razón nos hemos limitado a este intervalo en vez de seguir utilizando el habitual de la binomial clasica.

Bajo estas medidas de probabilidad, cualquier número de compomentes de Ω es estadisticamente independiente de las restantes; como cada componente individual tiene, por construcción, el mismo espacio probabilistico, la probabilidad de cualquier punto $\bar{a}=(a_1,a_2,\ldots,a_n)$ viene dada por

$$P[\bar{a}] = \prod_{i=1}^{n} p^{a_i}.q^{1-a_i}$$

o bien usando la distancia de Hamming

$$P[\bar{a}] = p^{D(\bar{a},0)}.\ q^{n-D(\bar{a},0)} = p^{\|\bar{a}\|}.\ q^{n-\|\bar{a}\|}$$

Esto significa que la probabilidad de cada punto depende unicamente de la distancia de este punto al origen, o, lo que es lo mismo, de su norma.

A continuación, exactamente como hicimos en el capítulo anterior, vamos a asociarle a cada punto de Ω una variable aleatoria definida de la siguiente manera:

$$Y_{\vec{a}}: \Omega \longrightarrow \mathcal{R}^{+}$$
$$Y_{\vec{a}}(x) = D(\vec{x},\vec{a})$$

De esta forma obtenemos un total de 2^n variables aleatorias. No obstante, no todas ellas tienen diferentes leyes de probabilidad, sino que existen unicamente $n+1$ leyes distintas, tantas como valores posibles puede tomar la distancia de Hamming en el espacio muestral. Demostraremos esta afirmación a partir del hecho de que todos los puntos del espacio situados a igual distancia del origen tienen asociadas variables aleatorias con igual distribución.

Para ello, consideremos dos puntos $\vec{a}$ y $\vec{b}$ cuya norma valga d

$$\|\vec{a}\| = \|\vec{b}\| = d$$

Consideremos el giro G que transforma el punto $\vec{a}$ en el $\vec{b}$, y tengamos en cuenta los siguientes hechos: que la probabilidad de cada punto depende unicamente de su distancia al origen; y que un giro es una isometria, o dicho de otra forma,

la distancia entre transformados es siempre igual a la existente entre sus respectivos originales.

Sea $\vec{x}$ un punto cualquiera de Ω tal que su norma sea r y la distancia respecto de $\vec{a}$ valga k, entonces su punto transformado $G(\vec{x})$ verificará que

$$D(G(\vec{x}),G(0)) = D(G(\vec{x}),0) = r$$

y

$$D(G(\vec{x}),G(\vec{a})) = D(G(\vec{x}),\vec{b}) = k$$

Por consiguiente, para cada punto que aporte una cierta probabilidad al valor k de la variable $Y_{\vec{a}}$, existe otro que aporta esa misma cantidad al valor k de la variable $Y_{\vec{b}}$, cualquiera que sea $\vec{b}$ siempre que tenga la misma norma que $\vec{a}$.

De aquí concluimos que $Y_{\vec{a}}$ e $Y_{\vec{b}}$ tienen igual distribución siempre que sus puntos definidores disten la misma distancia desde el origen.

A la ley común que tienen todas estas $Y_{\vec{a}}$ la denominaremos binomial metrica de parámetros n, p y d, siendo n la dimensión del espacio Ω soporte de las variables aleatorias; p el valor de una probabilidad que induce la medida de probabilidad del espacio; y d, el nuevo parámetro

que añadimos a la binomial clasica, la norma de cualquier punto generador de la variable que notaremos por X_d .

Alternaremos esta notación con otra mas coherente con la que estamos siguiendo para la binomial clasica; concretamente emplearemos $B(n,d,p)$ o cualquiera de sus formas al suprimir los parametros perfectamente explicitos en el texto. Por otra parte, esta ultima notación con subindice, o con este englobado como primer argumento, hará referencia a la función de distribución de las variables binomiales metrica (VBM); cuando la empleemos en minuscula $b_k(n,d,p)$, indicaremos la masa de probabilidad de $X_d=k$.

III.2.- FUNCION DE MASA DE PROBABILIDAD

Nos basaremos para su obtención, como ya hemos dicho, en la equivalencia de la estrutura de entornos de todos los puntos de Ω. Esto significa que a distancia k de cualquier punto existen $\binom{n}{k}$ puntos. Por otra parte, todos esos puntos estarán situados a distintas distancias respecto del origen. El número de ellos que se encuentran a distancia r del origen depende unicamente del parametro d

correspondiente a la VBM que se esté considerando. Es decir, para cada valor de d se produce una partición de $\binom{n}{k}$ de la siguiente forma :

$$\binom{n}{k} = \pi_d(0;k) + \pi_d(1;k) + \ldots\ldots + \pi_d(n;k)$$

de modo que a partir de ella, la probabilidad puntual de $X_d=k$ se puede expresar como

$$b_k(n,d,p) = \sum_{r=o}^{n} \pi_d(r;k) \cdot p^r \cdot q^{n-r}$$

Pasemos,pues, a determinar los valores de $\pi_d(r;k)$. Consideremos un punto cualquiera $\vec{a}$ que nos defina X_d . Este punto evidentemente estará formado por d componentes unitarias y n-d componentes nulas dispuesta en un cierto orden. Para pasar desde el punto $\vec{a}$, a uno cualquiera del conjunto $X_d=k$ es preciso realizar naturalmente sobre $\vec{a}$ k permutaciones de sus componentes que transformarán ceros en unos y viceversa. El número maximo de permutaciones sobre unos está limitado por el número de estos existentes en el punto $\vec{a}$, o sea d ; cada permutación de estas reduce en una unidad la distancia del punto resultante respecto del origen. Analogamente, sucede que el número maximo de permutaciones sobre ceros es n-d y que cada permutación de es

te tipo incrementa en una unidad la distancia del transformado.

Llamemos i el número de permutaciones sobre unos que realizamos para alcanzar un punto de $X_d=k$. El número de permutaciones realizadas sobre ceros será $k-i$, siendo su valor maximo $n-d$. Según lo que acabamos de escribir, i siempre va a estar comprendido entre

$$\max(k-n+d,0) \leqq i \leqq \min(k,d)$$

El número de formas posibles de obtener i permutaciones sobre unos es $\binom{d}{i}$. Analogamente refiriendonos a los ceros es $\binom{n-d}{k-i}$. De modo que el número total de formas de conseguir k permutaciones sobre $\vec{a}$ que contengan i permutaciones sobre unos es entonces

$$\binom{d}{i} \cdot \binom{n-d}{k-i}$$

Mediante estas permutaciones hemos obtenido todos los puntos de Ω situados a distancia k de $\vec{a}$, y que siempre distan una distancia r igual a

$$d + 1\cdot(k-i) - 1\cdot i = d + k - 2\cdot i$$

respecto de origen; es decir

$$\pi_d(d+k-2i;k) = \binom{d}{i} \cdot \binom{n-d}{k-i}$$

Observese que el valor de π_d consecutivo a uno no nulo siempre es nulo, o sea si

$$\pi_d(r;k) \neq 0 \implies \pi_d(r\ 1;k) = 0$$

El razonamiento seguido incluye todos los posibles π_d no nulos porque, como es sabido, se cumple la siguiente igualdad entre números combinatorios :

$$\sum_{i=o}^{d} \binom{d}{i} \cdot \binom{n-d}{k-i} = \binom{n}{k}$$

Tal igualdad se utiliza para demostrar que la hipergeometrica es una verdadera distribución de probabilidad. Precisamente i , el número de permutaciones realizadas sobre unos, sigue una distribución hipergeometrica; concretamente, sigue H(n,d,k) .

Por simplicidad hemos eliminado las restricciones sobre i , pero debe entenderse que siguen en vigor y, por lo tanto, aquellos números combinatorios en que el ter

mino inferior es negativo o rebasa al superior deben considerarse nulos.

Según lo anterior, tenemos que la función de masa de probabilidad viene dada por

$$P[X_d=k] = \sum_{i=0}^{d} \binom{d}{i}\cdot\binom{n-d}{k-i}\cdot p^{d+k-2i}\cdot q^{n+2i-d-k} =$$

$$= \left(\frac{p}{q}\right)^{d+k}\cdot q^{n}\cdot \sum_{i=0}^{d} \binom{d}{i}\cdot\binom{n-d}{k-i}\cdot \left(\frac{q}{p}\right)^{2i}$$

Esta última forma solamente valida cuando ni p ni q son nulos.

Examinaremos mas adelante la forma de esta función de probabilidad, así como el lugar donde se localizan la moda y la mediana de la distribución. Para ello es necesario conocer previamente las funciones generatrices de las VBM, de lo cual trataremos a continuación.

III.3.- FUNCIONES GENERATRICES

Calcularemos unicamente la función generatriz de probabilidad; las restantes se obtienen a partir de es

ta por cambio de variable o por una sencilla transformación. Por definición, la función generatriz buscada es

$$G(z;d) = \sum_{k=o}^{n} b_k(d)\cdot z^k =$$

$$= \sum_{k=o}^{n} \sum_{i=o}^{d} \binom{d}{i}\cdot\binom{n-d}{k-i}\cdot p^{d+k-2i}\cdot q^{n-d-(k-2i)}\cdot z^k$$

Cambiando el orden de las sumatorias, sacando fuera lo que no depende de k , teniendo en cuenta los números combinatorios nulos y multiplicando y dividiendo por z^i ,pasamos a

$$\sum_{i=o}^{d} \sum_{k=i}^{n+i-d} \binom{d}{i}\cdot\binom{n-d}{k-i}\cdot p^{d-i+k-i}\cdot q^{i+(n-d)-(k-i)}\cdot z^k =$$

$$= \sum_{i=o}^{d} \binom{d}{i}\cdot p^{d-i}\cdot q^i\cdot z^i \cdot \sum_{k=i}^{n-d+i} \binom{n-d}{k-i}\cdot p^{k-i}\cdot q^{n-d-(k-i)}\cdot z^{k-i}$$

Sustituyendo el índice k por el índice j definido por $j=k-i$, nos damos cuenta que

$$\sum_{j=o}^{n-d} \binom{n-d}{j}\cdot p^j\cdot q^{n-d-j}\cdot z^j$$

es el desarrollo del binomio $(q + pz)^{n-d}$. De modo que sustituyendolo en la expresión anterior, sacandolo factor

común, y razonando de manera similar sobre la otra sumatoria, llegamos finalmente a obtener que

$$G(z;d) = (q+pz)^{n-d} \cdot (p+qz)^{d}$$

Como bien puede observarse, se trata del producto de las funciones generatrices de probabilidad de dos binomiales: una de parametros $n-d$ y p ; y otra de parametros d , q. Por tanto, la DBM de parametros (n,d,p) es la distribución de la suma de estas dos binomiales independientes; o sea

$$B(n,d,p) = B(n-d,p) * B(d,q)$$

Veamos cómo podemos razonar sobre el espacio muestral para expresar la distribución $B(n,d,p)$ efectivamente como la suma de $B(n-d,p)$ y $B(d,q)$. La variable X_d está asociada a cualquier punto que tenga d componentes unitarias y $n-d$ nulas. Por definición, la variable $Y_{\bar{a}}$ mide el número de permutaciones o diferencias que existen entre cada original del espacio muestral y el punto $\bar{a}$. Ahora bien, respecto de $\bar{a}$ estas diferencias pueden tener lugar sobre ceros o bien sobre unos, de modo que

$$Y_{\bar{a}} = Y_{\bar{a}}(0) + Y_{\bar{a}}(1)$$

Como por construcción cada componente es independiente de las restantes, resulta que la probabilidad de que se produzca una diferencia sobre una componente nula es exactamente igual a la probabilidad de que en esa componente haya un uno; es decir p. Analogamente, la probabilidad de una diferencia sobre un uno vale q . Por tanto, $Y_{\bar{a}}(0)$ sigue una ley binomial B(n-d,p), y $Y_{\bar{a}}(1)$ una binomial B(d,q). Ademas ambas son independientes entre si. De donde concluimos que

$$X_d = Y_{\bar{a}} = B(n-d,p) * B(d,q)$$

Como consecuencia de esto tenemos que la DBM B(n,d,p) es un caso particular de la distribución binomial generalizada por Poisson; concretamente aquella que viene generada por el vector de probabilidades $(\underbrace{p,\ldots\ldots,p}_{n-d},\underbrace{q,\ldots\ldots,q}_{d})$.

No obstante, consideramos que existe una diferencia sustancial entre ambos tipos de generalización, hasta el punto de que la binomial de Poisson es suceptible de ser generalizada por via metrica. Naturalmente al realizar esto no obtenemos un nuevo tipo de distribución, de modo que podemos decir que la generalización de Poisson de la binomial es cerrada respecto de la extensión metrica. Por todo lo cual, pensamos que se tratan de tipos completamen

te diferentes de generalización de una misma distribución. Mas aun, la generalización metrica se puede aplicar a cualquier espacio probabilistico; en cambio, la de Poisson solo es aplicable a la binomial.

Continuando con las funciones generatrices, realizadas las transformaciones necesarias, obtenemos que la generatriz de momentos, de cumulantes y la función característica son, respectivamente :

$$\Phi(t) = (q + p.e^{t})^{n-d}.(p + q.e^{t})^{d}$$

$$\Psi(t) = (n-d).\ln(q+pe^{t}) + d.\ln(p+qe^{t})$$

$$\varphi(t) = (q + pe^{it})^{n-d}.(p + qe^{it})^{d}$$

III.4.- MOMENTOS Y CUMULANTES

Calculemos en primer lugar la esperanza matematica y la varianza de la VBM. Por tratarse de suma de dos variables aleatorias independientes, sabemos que estos momentos son simplemente la suma de los correspondientes a cada binomial, así que

$$E[X_d] = E[X(0)] + E[X(1)] =$$

$$= (n-d)\cdot p + d\cdot q = n\cdot p + d\cdot(q-p) =$$

o bien

$$= n\cdot p + d\cdot(1-2p) = (n-2d)\cdot p + d = (np-d) + 2dq \quad .$$

Damos todas estas expresiones equivalentes porque serán utilizadas, cuando sea conveniente, a lo largo de las restantes páginas de este capítulo. Para referirnos a un valor concreto de la media emplearemos la letra λ ó μ; tendremos oportunidad de hacer uso de esta notación cuando nos dediquemos a la comparación de distintas DBM que tengan el mismo valor medio.

Por otra parte, la varianza vale

$$V(X_d) = (n-d)\cdot p\cdot q + d\cdot p\cdot q = n\cdot p\cdot q$$

Un valor concreto de la misma será notado por σ^2.

De estos dos momentos podemos sacar las siguientes observaciones :

a.- La varianza de la familia para p y n fijos, es independiente del parametro d .

b.- La media es función lineal de d. Dicho de otra

manera, cuando d varía los valores de la media están en progresión aritmetica de diferencia (q-p) .

c.- Si n es par, n=2m , y d vale precisamente m , la media es independiente de p, teniendo siempre un valor de m .

d.- Por ultimo, también como consecuencia de la observación b, la media es independiente de d si p=½ . Mas precisamente, en este caso todas las variables binomiales metricas son iguales.

Siguiendo con los restantes momentos, podemos obtenerlos por derivación sucesivas de las funciones generatrices: los momentos respecto del origen y los centrales a partir de la función generatriz de momentos; los momentos factoriales a partir de la generatriz de probabilidad. Ahora bien, como la DBM es suma de dos distribuciones binomiales independientes, $X_d = X(0) + X(1)$, sus momentos pueden expresarse en función de los de estos sumandos. Consideremos la fórmula de la derivada r-esima de un producto de funciones :

$$\frac{d^r}{dt^r} \cdot (f(t) \cdot g(t)) = (f^{(} + g^{(})^r = \sum_{i=o}^{r} \binom{r}{i} f^{(i} \cdot g^{(r-i}$$

siendo $f^{(i}$ la derivada iesima de f(t). Eligiendo convenien

temente las funciones f y g , y particularizando para t=1 o t=0, obtenemos que los momentos factoriales y los ordinarios de X_d vienen dados por

$$\mu_{(r)} = \sum_{i=o}^{r} \binom{r}{i} \cdot \mu_{(i)}(0) \cdot \mu_{(r-i)}(1)$$

$$\mu'_r = \sum_{i=o}^{r} \binom{r}{i} \cdot \mu'_i(0) \cdot \mu'_{r-i}(1)$$

donde $\mu_{(i)}(0)$ y $\mu_{(i)}(1)$ indican los momentos factoriales de orden i de las binomiales X(0) y X(1) . Analogamente $\mu'_j(0)$ y $\mu'_j(1)$ representan los momentos ordinarios. Teniendo en cuenta la sencillez de los momentos factoriales de la binomial, resulta que

$$\mu_{(r)} = \sum_{i=o}^{r} \binom{r}{i} \cdot (n-d)_{(i)} \cdot p^i \cdot d_{(r-i)} \cdot q^{r-i}$$

que simbolicamente queda reducida a

$$\mu_{(r)} = ((n-d)_{()} \cdot p + d_{()} \cdot q)^r$$

Por otro lado, la expresión que da los momentos ordinarios es valida para cualquier punto que se tome como origen. En particular, los momentos centrales verifican la misma expresión, con la salvedad de que los terminos en que aparezca el momento de primer orden de cualquiera de los su

mandos de X_d son siempre nulos; esto es

$$\mu_r = \mu_r(1) + \binom{r}{2}\mu_{r-2}(1)\cdot\mu_2(0) + \ldots\ldots + \binom{r}{r-2}\mu_2(1)\cdot\mu_{r-2}(0) + \mu_r(0)$$

Así, aplicando esto al tercer y cuarto momento, obtenemos que sus valores son :

$$\mu_3 = \mu_3(0) + \mu_3(1) = (n-d)\cdot p\cdot q\cdot(q-p) + d\cdot p\cdot q\cdot(p-q) = (n-2d)\cdot p\cdot q\cdot(q-p) \quad .$$

$$\mu_4 = \mu_4(0) + 6\cdot\mu_2(0)\cdot\mu_2(1) + \mu_4(1) =$$

$$= 3\{(n-d)\cdot p\cdot q\}^2 + (n-d)\cdot p\cdot q\cdot(1-6pq) + 6(n-d)pq\cdot dpq +$$

$$+ 3\{d\cdot p\cdot q\}^2 + d\cdot q\cdot p\cdot(1-6qp) =$$

$$= 3\{npq\}^2 + npq\cdot(1-6pq) \quad .$$

De aquí podemos extraer las siguientes observaciones :

a.- El momento de tercer orden, al igual que la media, también es función lineal de d. Esta es tal que distribuciones con valores de d complementarios a n tienen momentos de asimetria opuestos.

b.- Como consecuencia de la observación anterior, siempre que n sea par, n=2m, las distribuciones

de la familia correspondientes a d=m poseen un momento μ_3 nulo. Mas exactamente, en este caso la distribución es simétrica respecto de la media, como vamos a comprobar a continuación. Se trata de mostrar que

$$P[X_m=m-a] = P[X_m=m+a] \qquad \forall a \ ,$$

o equivalentemente que

$$b_k(n,m,p) = b_{n-k}(n,m,p) \qquad \forall k \leq m$$

Las probabilidades de cada miembro valen :

$$b_k(m) = \sum_{i=o}^{k} \binom{m}{i} \cdot \binom{m}{k-i} \cdot p^{m+k-2i} \cdot q^{m-(k-2i)}$$

y

$$b_{n-k}(m) = \sum_{j=m-k}^{m} \binom{m}{j} \binom{m}{2m-k-j} \cdot p^{3m-k-2j} \cdot q^{k+2j-m}$$

La igualdad entre ambas probabilidades se consigue haciendo en $b_{n-k}(m)$ la sustitución del indice j por el i definido como

$$i = k + j - m \ ,$$

sustitución que es el resultado de otras dos mas naturales

$$j' = m-j \quad e \quad i = k-j' \ ;$$

y posteriormente cambiando el número combinatorio

$\binom{m}{m-i}$ por su equivalente $\binom{m}{i}$.

c.- Al igual que la varianza, el momento de cuarto orden también es independiente del parametro d.

Relaciones semejantes pueden seguir estableciendose para los sucesivos momentos; por ejemplo se verifica que: el momento de 5º orden también es función lineal del parametro distancia; y el de 6º orden es función cuadratica de ella; o mas concretamente, es de la forma cte + cte'. $.(n-2d)^2$, de modo que alcanza el valor mínimo cuando el parametro d está lo mas proximo a $\frac{n}{2}$, y toma valores iguales para valores complementarios a n (o simetricos respecto de $\frac{n}{2}$) .

Procediendo de la misma manera que M. Merrell, vamos a considerar las relaciones que existen entre los coeficientes β_1 y β_2 cuando individualmente varia cada parametro mientras que los restantes permanecen fijos.

Por definición los coeficientes β_1 y β_2 son :

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(n-2d)^2 \cdot p^2 \cdot q^2 \cdot (p-q)^2}{n^3 \cdot p^3 \cdot q^3} =$$

$$= \frac{(n-2d)^2}{n^2} \cdot \frac{1 - 4pq}{p \cdot q}$$

y

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{1 + 3 \cdot p \cdot q \cdot (n-2)}{n \cdot p \cdot q}$$

Evidentemente este coeficiente es independiente del parametro d.

Cuando consideramos al parametro p como variable, si despejamos el producto $p \cdot q$ de la ecuación de β_2 y lo sustituimos en β_1 , obtenemos la relación deseada

$$\beta_1 = \frac{(n-2d)^2}{n^3} \cdot \left\{ 2 + n \cdot (\beta_2 - 3) \right\}$$

Como vemos, en este caso, igual que ocurre en la binomial clasica, para las distribuciones de la familia estos coeficientes se hallan sobre una linea recta. Cuando n tiende a infinito y aunque d varie, siempre que sea como $o(n)$, obtenemos como recta limite a

$$\beta_1 = \beta_2 - 3$$

Ahora bien, como β_2 tiende a 3, la recta degenera también en el punto $(0,3)$.

Cuando sea n el parametro variable, la linea

sobre la que caen estos coeficientes de forma tiene por ecuación a

$$\beta_1 = \frac{\{1-2pq(d(\beta_2-3)+3)\}^2}{(1-6pq)^3} \cdot (1-4pq) \cdot (\beta_2-3)$$

Esta linea es una recta unicamente cuando d vale 0 ó n; aquí se pierde el paralelismo existente con la binomial clásica.

En el último caso, cuando d varia, como β_2 es independiente de este parametro, la linea obtenida es

$$\beta_2 = \frac{1 + 3pq \cdot (n-2)}{npq}$$

es decir, una linea horizontal . Conocido el valor de β_1, los valores de d que proporcionan este coeficiente son

$$\frac{n}{2} \cdot \left(1 \pm \sqrt{\frac{\beta_1 \cdot n \cdot p \cdot q}{1 - 4pq}} \right)$$

Dejamos ya los momentos de las distribuciones y pasamos a tratar los cumulantes, los cuales resultan mas convenientes de calcular en este tipo de distribuciones. Esto viene justificado por dos razones. En primer lugar, porque los cumulantes de una suma de variables aleatorias independientes son la suma de los correspondientes a cada

sumando. En segundo lugar, por la sencilla relación de recurrencia que liga a los semiinvariantes de la distribución binomial.(Vease I.5). Dicho de otra forma,porque:

$$\text{I} \qquad \kappa_r = \kappa_r(0) + \kappa_r(1)$$

$$\text{II} \begin{cases} \text{a} & \kappa_r(0) = p\cdot q\cdot \dfrac{d\kappa_{r-1}(0)}{dp} \\ \text{b} & \kappa_r(1) = q\cdot p\cdot \dfrac{d\kappa_{r-1}(1)}{dq} = -p\cdot q\cdot \dfrac{d\kappa_{r-1}(1)}{dp} \end{cases}$$

sujetas respectivamente a

$$\kappa_1(0) = (n-d)\cdot p$$

$$\kappa_1(1) = d\cdot q = d - d\cdot p$$

De este modo, aplicando repetidamente las ecuaciones de rerrencia y sumando, obtenemos el valor de los 4 primeros cumulantes :

$$\kappa_1 = (n-d)\cdot p + d - d\cdot p = (n-2d)\cdot p + d$$

$$\kappa_2 = (n-d)\cdot p\cdot q + d\cdot p\cdot q = n\cdot p\cdot q$$

$$\kappa_3 = (n-d)pq\cdot(q-p) + dqp\cdot(p-q) = (n-2d)\cdot p\cdot q\cdot(q-p)$$

$$\kappa_4 = (n-d)pq(1-6pq) + dpq(1-6pq) = n\cdot p\cdot q\cdot(1-6pq)$$

Afortunadamente también podemos obtener los cumulantes, al

margen de su función generatriz, a partir de una fórmula muy similar a la ecuación de recurrencia de la binomial. Para hallarla emplearemos esa última ecuación para el caso particular de una única prueba de Bernoulli cuya probabilidad de exito valga indistintamente p o q.

Como X(0) surge de la suma de (n-d) pruebas de Bernoulli con probabilidad de exito p, sus cumulantes serán (n-d) veces el valor de los correspondientes a esa variable de Bernoulli; si llamamos $\kappa^{\star}_{r}$ a los cumulantes de esta última variable, tenemos

$$\kappa_r(0) = (n-d)\cdot\kappa^{\star}_{r} = (n-d)\cdot p\cdot q\cdot \frac{d\kappa^{\star}_{r-1}}{dp}$$

Razonando de manera analoga sobre X(1) , llegamos directamente a la fórmula buscada :

$$\kappa_r = (n-d)\cdot p\cdot q\cdot \frac{d\kappa^{\star}_{r-1}}{dp} + d\cdot q\cdot p\cdot \frac{d\kappa^{\star}_{r-1}}{dq} =$$

$$= \{n + d\cdot((-1)^r-1)\}\cdot p\cdot q\cdot \frac{d\kappa^{\star}_{r-1}}{dp}$$

sujeta naturalmente a $\kappa^{\star}_{1} = p$.

Dicha fórmula puede escribirse como ecuación de recurrencia para los cumulantes de una VBM de la siguiente

manera :

$$\kappa_r = \left\{ \frac{n}{n-2d} \right\}^{(-1)^r} \cdot p \cdot q \frac{d\kappa_{r-1}}{dp}$$

Observando la fórmula en su versión anterior, podemos sacar las siguientes conclusiones sobre estos semiinvariantes :

a.- Todos los cumulantes de orden par coinciden con los correspondientes a la binomial que genera la familia. Dicho de otro modo, son invariantes para toda la familia; son independientes del parametro distancia.

b.- Todos los cumulantes de orden impar están en progresión aritmetica respecto del parámetro d ; la diferencia de la progresión vale $-2\kappa_r^{\star}$.

III.5.- OTROS TIPOS DE MOMENTOS

Desarrollaremos esta sección en el mismo orden que seguimos en la sección 6 del primer capítulo; así pues, comenzaremos tratando la desviación media de la familia de variables binomiales metricas.

a.- La desviación media de una VBM B(n,d,p) viene definida por

$$\nu_1 = E[|X_d-\mu|] = \sum_{k=o}^{n} |k-\mu|\cdot b_k(n,d,p)$$

Ahora bien, como el primer momento central vale siempre cero, el primer momento central absoluto puede determinarse por

$$\nu_1 = 2 \sum_{k>\mu} (k-\mu)\cdot b_k(n,d,p)$$

Para realizar su cálculo vamos a emplear dos identidades. La primera relaciona el producto de un número combinatorio por su índice inferior con otro producto similar. La segunda ha sido utilizada en la teoría matematica de riesgos en Seguros; fue demostrada por Gruder en 1930 y está mencionada por N. L. Johnson (1957), quien partiendo de esta propiedad calculó de nuevo la desviación media de la distribución binomial. Tales identidades son las siguientes :

$$\text{I)} \qquad m\cdot\binom{n}{m} = n\cdot\binom{n-1}{m-1}$$

Nosotros la emplearemos para sumar la expresión

$$\sum_{i=o}^{d} i\cdot\binom{d}{i}\cdot\binom{n-d}{k-i}\cdot p^{d+k-2i}\cdot q^{n-d-k+2i}$$

quedando su resultado en términos de la masa de

probabilidad de otra DBM :

$$d \cdot q \cdot b_{k-1}(n-1, d-1, p)$$

También se empleará en la misma expresión, pero cuando se sustituya el primer i por k-i.

II) $$\sum_{r=m}^{n} \binom{n}{r} \cdot p^r \cdot q^{n-r} \cdot \left| r - np \right| = m \cdot \binom{n}{m} \cdot p^m \cdot q^{n-m+1}$$

que también puede escribirse, aplicando I, como

$$n \cdot \binom{n-1}{m-1} \cdot p^m \cdot q^{n-m+1}$$

Veamos ahora como hemos determinado v_1 . Como ya hemos visto, tenemos que

$$v_1 = 2 \cdot \sum_{k>\mu} (k-(n-d)\cdot p - d \cdot q) \cdot \sum_{i=0}^{d} \binom{d}{i} \cdot \binom{n-d}{k-i} \cdot p^{d+k-2i} \cdot q^{n-d-k+2i}$$

Su cálculo lo hemos realizado en los siguientes pasos:

- En primer lugar, restando y sumando convenientemente i en el primer paréntesis, descomponemos $k-\mu$ en tres sumandos; a saber :

$$k-i-(n-d)\cdot p \quad , \quad i \quad , \quad -d \cdot q \quad ,$$

a los cuales se les puede determinar su esperanza

matematica utilizando las identidades anteriormente indicadas.

- La esperanza del primer sumando se halla, en primer lugar, permutando el orden de las sumatorias y aplicando la identidad II, lo que da el resultado intermedio

$$2 \cdot \sum_{i=0}^{d} \binom{d}{i} \cdot \left\{ [\mu+1] - i \right\} \cdot \binom{n-d}{[\mu+1]-i} \cdot p^{d+[\mu+1]-2i} \cdot$$

$$\cdot q^{n-d-[\mu+1]+2i}$$

que aplicando I da finalmente

$$2 \cdot (n-d) \cdot p \cdot q \cdot b_{[\mu]}(n-1,d)$$

- La esperanza correspondiente a i se determina aplicando la identidad I y considerando la función de distribución de la BM resultante. Su valor es

$$2 \cdot d \cdot q \cdot \left\{ 1 - B(\mu-1;n-1,d-1) \right\}$$

- La última esperanza se halla directamente interpretandola como función de distribución. Su valor es

$$2 \cdot d \cdot q \cdot \left\{ 1 - B(\mu;n,d) \right\}$$

Ahora bien, este termino se puede comparar con el anterior si expresamos B(n,d) como B(n-1,d-1)*B(1,1), ya que podemos sustituir $B(\mu;n,d)$ por

$$B(\mu-1;n-1,d-1) + p\cdot b_{[\mu]}(n-1,d-1) \quad .$$

Por tanto, sumando los tres terminos, obtenemos la siguiente expresión para la desviación media:

$$\nu_1 = 2\cdot p\cdot q\left\{d\cdot b_{[\mu]}(n-1,d-1) + (n-d)\cdot b_{[\mu]}(n-1,d)\right\}$$

Como vemos, el cálculo de ν_1 requiere el conocimiento de la probabilidad del punto $[\mu]$ respecto a dos distribuciones. Si $n\geq 3$ y $d\geq 1$, podemos expresar ν_1 en función de una única distribución. Para ello, del mismo modo que hicimos con B(n,d), expresamos B(n-1,d) y B(n-1,d-1) a partir de B(n-2, d-1). Actuando de esta manera, llegamos a la siguiente fórmula:

$$\nu_1 = 2\cdot p\cdot q\left\{\mu\cdot b_{[\mu]}(n-2,d-1) + (n-\mu)\cdot b_{[\mu-1]}(n-2,d-1)\right\}$$

Hemos tabulado este momento con el fín de encontrar algunas regularidades en su comportamiento respecto a cada parametro cuando los otros permanecen fijos. En los casos tabulados, hemos observado que ν_1 nunca decrece al aumentar n. Los casos en que no crece corresponden a distribuciones cuyo valor medio es entero. De estas afirmaciones empí-

ricas solo hemos podido demostrar la última cuyo enunciado preciso es el siguiente:

Teorema.- Sea $B(n,d,p)$ una distribución binomial métrica con valor medio μ entero. Entonces se verifica que

$$\nu_1(n,d) = \nu_1(n-1,d) = \nu_1(n-1,d-1)$$

Para demostrar esta propiedad vamos a sustituir, directamente en la definición de ν_1, $b_k(n,d)$ y μ por sus valores en función de cada una de las otras dos distribuciones. En estas condiciones, si llamamos μ_d a $E[X(n-1,d)]$, su valor es $\mu-p$. Analogamente tenemos que $\mu_{d-1} = \mu-q$. De este modo tenemos:

$$\nu_1(n,d) = 2\cdot \sum_{k=\mu+1}^{n} (k-\mu_d-p)\cdot\left\{p\cdot b_{k-1}(n-1,d)+q\cdot b_k(n-1,d)\right\}$$

$$= 2\cdot p \sum_{k=\mu+1}^{n} (k-\mu_d)\cdot b_{k-1}(n-1,d) + 2\cdot q \sum_{k=\mu+1}^{n} (k-\mu_d)\cdot b_k(n-1,d) -$$

$$- 2 \sum_{k=\mu+1}^{n} p^2\cdot b_{k-1}(n-1,d) + 2pq\cdot \sum_{k=\mu+1}^{n} b_k(n-1,d)$$

Cambiando el índice de la 1ª y 3ª sumatoria, descomponiendo la 1ª en dos sumatorias, añadiendo $2pqb_\mu$ a las otras dos sumatorias y teniendo en cuenta que $b_n(n-1,d)$ vale 0, obtenemos :

$$2\cdot p \sum_{k=\mu}^{n-1}(k-\mu_d)\cdot b_k(n-1,d) \;+\; 2\cdot p \sum_{k=\mu}^{n-1} b_k(n-1,d) \;+$$

$$+\; 2\cdot q \sum_{k=\mu}^{n-1}(k-\mu_d)\cdot b_k(n-1,d) \;-\; 2\cdot p \sum_{k=\mu}^{n-1}(p+q)\cdot b_k(n-1,d)$$

Como el 2º y 4º término se cancelan y los otros dos suman $\nu_1(n-1,d)$, ya hemos demostrado la 1ª parte del teorema. De manera similar se demostraría la 2ª parte, pero en este caso sustituyendo $b_k(n,d)$ por $q\cdot b_{k-1}(n-1,d-1) + p\cdot b_k(n-1,d-1)$.

Respecto a los otros dos parámetros, ν_1 es simétrico, es decir, vale igual para distribuciones cuyos valores de p ó d sean complementarios a 1 ó n, respectivamente. Esto se desprende facilmente de las relaciones que veremos en la proxima sección. Limitandonos, entonces, a los intervalos $[0,\frac{1}{2}]$ y $[0,\frac{n}{2}]$, resulta que la tendencia general de ν_1 es crecer respecto a p y decrecer respecto a d. No obstante, esta tendencia está salpicada, al menos cuando n es moderado, de mínimos relativos, debidos a que μ es entero o proximo a entero. Estos mínimos no existen cuando n es par y d vale $\frac{n}{2}$, como aseguran los resultados de Gleser (Vease I.12).

Por otra parte, hemos observado que cuando n es par, ν_1 siempre alcanza su mínimo respecto a d para $d=\frac{n}{2}$; esto se debe probablemente a la simetría de esa distribución.

b.- Momentos inversos factoriales :

En primer lugar, señalaremos que de manera general los momentos de este tipo se pueden determinar para cualquier distribución discreta no negativa a partir de su correspondiente función generatriz de probabilidad $G(z)$; concretamente, estos momentos vienen dados por la integración sucesiva de $G(z)$ como lo muestra la fórmula

$$E\left[\frac{1}{(X+r)_{(r)}}\right] = \int_0^1 dz_r \int_0^{z_r} dz_{r-1} \cdots\cdots \int_0^{z_2} G(z_1)\cdot dz_1$$

Aplicandola a nuestra distribución obtendremos una fórmula general para calcular dichos momentos factoriales. Por comodidad, los notaremos para esta familia por $m_r(d)$ ó $m_r(n,d)$, empleando la última cuando sea necesario especificar claramente el parámetro n.

Según lo que acabamos de ver, el primer momento viene dado entonces por

$$m_1(d) = \int_0^1 G(z;d)\cdot dz = \int_0^1 (p+qz)^d \cdot (q+pz)^{n-d}\, dz$$

que integrando por partes, al tomar como dv a $(p+qz)^d \cdot dz$, nos da como resultado

$$\frac{1}{(d+1)\cdot q}\cdot(p+qz)^{d+1}\cdot(q+pz)^{n-d}\Big|_{o}^{1} -$$

$$-\frac{(n-d)\cdot p}{(d+1)\cdot q}\int_{o}^{1}(p+qz)^{d+1}\cdot(q+pz)^{n-d-1}\cdot dz$$

Ahora bien, la nueva integral que aparece es simplemente $m_1(d+1)$, de donde podemos interpretar la identidad obtenida como la siguiente ecuación de recurrencia que liga los momentos de primer orden

$$(d+1)\cdot q\cdot m_1(d) + (n-d)\cdot p\cdot m_1(d+1) = 1 - b_o(n+1,d+1)$$

o sea

$$= 1 - p^{d+1}\cdot q^{n-d}$$

la cual permite, conocido este momento para B(n,d,p), calcularlo para B(n,d+1,p). Así, por ejemplo, utilizando la fórmula dada en la sección I.6 para este momento en la binomial, hallamos que

$$m_1(1) = \frac{1}{n\cdot p}\cdot\left\{1 - p\cdot q^n - \frac{q}{(n+1)p}\cdot[1 - q^{n+1}]\right\}$$

Por otro lado, si repetimos de la misma manera la integración por partes hasta agotar el exponente de (q+ +pz), se obtiene el desarrollo

$$m_1(d) = \frac{1}{(d+1)q}(1-p^{d+1}\cdot q^{n-d}) - \frac{(n-d)\cdot p}{(d+1)(d+2)q^2}(1-p^{d+2}q^{n-d})$$

$$+ \frac{(n-d)(n-d-1)\cdot p^2}{(d+1)(d+2)(d+3)\cdot q^3}(1 - p^{d+3}\cdot q^{n-d-2}) - \ldots\ldots$$

$$+ (-1)^{n-d}\frac{(n-d)!\ p^{n-d}}{(n+1)_{(n-d+1)}\cdot q^{n-d+1}}\cdot\left\{1 - p^{n+1}\right\}$$

o lo que es lo mismo, escrito en forma abreviada

$$m_1(d) = \frac{1}{(d+1)\cdot q}\sum_{j=o}^{n-d}\frac{(-(n-d))_{[j]}}{(d+2)_{[j]}}\left(\frac{p}{q}\right)^j\cdot\left\{1 - p^{d+j+1}\cdot q^{n-d-j}\right\}$$

Aprovechamos esta ocasión para recordar que las notaciones $x_{(j)}$ y $x_{[j]}$ significan, respectivamente, $x\cdot(x-1)\cdot\ldots\cdot(x-j+1)$ y $x\cdot(x+1)\cdot\ldots\cdot(x+j-1)$.

Teniendo en cuenta la definición de la función hipergeometrica, (dada en I.6.d), el primer momento queda finalmente expresado en la fórmula

$$\frac{1}{(d+1)\cdot q}\left\{F(1,-(n-d),d+2;\frac{p}{q}) - p^{d+1}\cdot q^{n-d}\cdot\right.$$

$$\left.\cdot\ F(1,-(n-d),d+2;\frac{p^2}{q^2})\right\}$$

Obviamente cuando se emplee esta fórmula para fines de cál-

culo, es conveniente deducirla de manera que incluya el menor número posible de términos; así, cuando d sea menor que (n-d) , deben intercambiarse sus papeles simultaneamente con la permutación de p por q. En cualquiera de las funciones hipergeometricas que así aparecen, aunque $\frac{p}{q}$ sea mayor que la unidad, está asegurada su convergencia porque constan siempre de un número finito de términos.

Pasemos ahora a considerar el segundo momento inverso factorial. En primer lugar establezcamos una ecuación de recurrencia para el mismo. Para ello, partamos de la ecuación análoga que verifica el primer momento, pero escrita en forma de integral indefinida; o sea

$$(d+1).q.\int_0^{z_2}(p+qz_1)^d(q+pz_1)^{n-d}.dz_1 + (n-d).p.\int_0^{z_2}(p+qz_1)^{d+1}\cdot$$

$$\cdot(q+pz_1)^{n-d-1}.dz_1 = (p+qz_2)^{d+1}.(q+pz_2)^{n-d} - p^{d+1}.q^{n-d}$$

Si integramos nuevamente entre 0 y 1, el primer miembro contiene los momentos buscados correspondientes a las variables X_d y X_{d+1} :

$$(d+1).q.m_2(d) + (n-d).p.m_2(d+1) \quad =$$

$$= \int_o^1 \left\{ (p+qz_2)^{d+1} \cdot (q+pz_2)^{n-d} - p^{d+1} \cdot q^{n-d} \right\} dz_2$$

Ahora bien, interpretando el primer integrando como momento, el segundo miembro vale

$$m_1(n+1,d+1) \; - \; p^{d+1} \cdot q^{n-d} \cdot z \Big|_o^1$$

con lo cual obtenemos la ecuación de recurrencia deseada; naturalmente, por medio de ella podemos determinar el momento de segundo orden para una distribución de distancia d+1 a partir de $m_2(d)$ y el momento de primer orden de la variable B(n+1,d+1,p), o inversamente, cuando $d > n/2$, expresar $m_2(d)$ en función de $m_2(d+1)$ y $m_1(n+1,d+1)$.

Ahora bien, como la ecuación de recurrencia es formalmente identica a la que verifican los momentos de primer orden, utilizandola repetidamente en la dirección de aumento del parametro d, hallamos una expresión para calcular $m_2(d)$

$$m_2(d) = \frac{1}{(d+1) \cdot q} \cdot \sum_{j=o}^{n-d} (-1)^j \frac{(n-d)_{(j)}}{(d+2)_{[j]}} \cdot \left(\frac{p}{q}\right)^j \cdot$$

$$\left\{ m_1(n+1,d+j+1) - p^{d+j+1} \cdot q^{n-(d+j)} \right\}$$

La parte del desarrollo correspondiente al sustraendo de la

expresión entre llaves es la misma que aparece en el cálculo de $m_1(d)$; por consiguiente, su valor es

$$-p^{d+1}\cdot q^{n-d}\cdot F(1,-(n-d),d+2;\ \frac{p^2}{q^2})$$

Por otra parte, si sustituimos los momentos de primer orden por sus desarrollos, obtenemos

$$\frac{1}{(d+1)\cdot q}\cdot \sum_{j=o}^{n-d}(-1)^j\cdot\frac{(n-d)_{(j)}}{(d+2)_{[j]}}\left\{\frac{p}{q}\right\}^j\cdot\frac{1}{(d+2+j)\cdot q}\cdot\sum_{k=o}^{n-d-j}(-1)^k\frac{(n-d-j)_{(k)}}{(d+3+j)_{[k]}}$$

$$\cdot\left\{\frac{p}{q}\right\}^k\cdot\left\{1-p^{d+1+j}\cdot q^{n-d-j}\cdot\left\{\frac{p}{q}\right\}^k\right\}$$

Realizando el producto de las dos series, obtenemos dos series iguales, pero con distintos argumentos: la primera en $\frac{p}{q}$, y la segunda en $\frac{p^2}{q^2}$. Consideremos un producto cualquiera de las series de partida que dé lugar a un termino de exponente i de la primera serie resultante; es decir, elijamos j y k tales que sumen exactamente i, y veamos cuanto vale el producto de los terminos correspondientes

$$\frac{1}{(d+1)\cdot q}\cdot(-1)^j\frac{(n-d)_{(j)}}{(d+2)_{[j]}}\cdot\left\{\frac{p}{q}\right\}^j\cdot\frac{1}{(d+2+j)\cdot q}\cdot\frac{(-1)^k(n-d-j)_{(k)}}{(d+3+j)_{[k]}}\cdot\left\{\frac{p}{q}\right\}^k=$$

$$=\frac{1}{(d+1)(d+2)q^2}\cdot(-1)^{j+k}\cdot\frac{(n-d)_{(j+k)}}{(d+3)_{[j+k]}}\cdot\left\{\frac{p}{q}\right\}^{j+k}$$

Claramente se observa que su valor depende unicamente de los indices j y k a traves de su suma; como la ecuación $i=j+k$ tiene $i+1$ soluciones sobre el conjunto de los enteros no negativos, tenemos que el desarrollo de la 1ª serie será entonces

$$\frac{1}{(d+1)\cdot(d+2)\cdot q^2} \sum_{i=o}^{n-d} (-1)^i \frac{(i+1)\cdot(n-d)_{(i)}}{(d+3)_{[i]}} \cdot \left\{\frac{p}{q}\right\}^i$$

que corresponde simplemente a la expresión desarrollada de la función hipergeometrica $F(2,-(n-d),d+3;\ \frac{p}{q}\ )$. Razonando analogamente sobre el sustraendo, concluimos que $m_2(d)$ viene dado por :

$$\frac{1}{(d+1)\cdot q}\cdot\left\{\frac{1}{(d+2)\cdot q}\cdot\left\{F(2,-(n-d),(d+3);\ \frac{p}{q}\ ) - \right.\right.$$

$$\left. - p^{d+2}\cdot q^{n-d}\cdot F(2,-(n-d),(d+3);\ \frac{p^2}{q^2}\ )\right\} -$$

$$\left. - p^{d+1}\cdot q^{n-d}\cdot F(1,-(n-d),(d+2);\ \frac{p^2}{q^2}\ )\right\}$$

Si en la definición inicial de $m_2(d)$ reemplazamos el limite superior unidad por la indeterminada z, obtenemos una función que notaremos por $m_2(z,d)$; se puede comprobar que dicha función viene dada por la siguiente expresión:

$$\frac{1}{(d+1)\cdot q}\cdot\left\{\frac{(p+qz)^{d+2}\cdot(q+pz)^{n-d}}{(d+2)\cdot q}\left\{F(2,-(n-d),(d+3);\frac{p(p+qz)}{q(q+pz)}\right\}\Bigg|_{o}^{z}-\right.$$

$$\left.-\,p^{d+1}\cdot q^{n-d}\cdot F(1,-(n-d),(d+3);\frac{p^2}{q^2})\cdot z\right|$$

Sigamos adelante generalizando estos resultados a los momentos inversos factoriales de cualquier orden. En este sentido, tenemos lo siguiente:

1) Todos los momentos se hallan relacionados entre si por la ecuación de recurrencia:

$$(d+1)\cdot q\cdot m_r(d) + (n-d)\cdot p\cdot m_r(d+1) =$$

$$= m_{r-1}(n+1,d+1) - \frac{1}{(r-1)!}\cdot p^{d+1}\cdot q^{n-d}$$

Su validez se establece sin ninguna dificultad por inducción matematica.

2) El momento $m_r(d)$ viene dado por la siguiente fórmula:

$$m_r(d) = \frac{1}{(d+1)_{[r]}\cdot q^r}\cdot F(r,-(n-d),(d+r+1);\frac{p}{q}) -$$

$$-\,p^d\cdot q^{n-d}\cdot\sum_{j=1}^{r}\frac{p^j}{(r-j)!\cdot(d+1)_{[j]}\cdot q^j}\cdot F(j,-(n-d),d+j+1;\frac{p^2}{q^2})$$

Resultado que también es demostrable por inducción . Para ello, teniendo en cuenta que

$$m_r(d) = \int_0^1 m_{r-1}(z,d)\cdot dz \quad ,$$

es suficiente comprobar que

$$\int_0^z (p+qz)^{d+r}.(q+pz)^{n-d}.\ F\left(r,-(n-d),d+r+1;\frac{p(p+qz)}{q(q+pz)}\right) dz$$

es igual a

$$\frac{(p+qz)^{d+r+1}}{(d+r+1).q}\cdot (q+pz)^{n-d}.F\left(r+1,-(n-d),d+r+2;\frac{p(p+qz)}{q(q+pz)}\right)\Bigg|_0^z$$

Para lo cual sustituyamos la función hipergeometrica por su desarrollo e integremos término a termino

$$\int_0^z (p+qz)^{d+r}(q+pz)^{n-d}\sum_{j=0}^{n-d}\frac{(-1)^j.(n-d)_{(j)}}{(d+r+1)_{[j]}}.\frac{r_j}{j!}.\left|\frac{p(p+qz)}{q(q+pz)}\right|^j.\ dz$$

$$=\sum_{j=0}^{n-d}\frac{(-1)^j(n-d)_{(j)}}{(d+r+1)_{[j]}}.\frac{r_j}{j!}.\frac{p}{q}^j.\int_0^z (p+qz)^{d+r+j}.(q+pz)^{n-d-j}\ dz$$

donde las nuevas integrales que aparecen representan $m_1(z;n+r,d+r+j)$; por consiguiente, sustituyendolas por su fórmula, tenemos que

$$= \sum_{j=o}^{n-d} \frac{(-1)^j \cdot (n-d)_{(j)} \cdot r_{[j]}}{(d+r+1)_{[j]} \; j!} \left\{\frac{p}{q}\right\}^j \cdot \frac{(p+qz)^{d+r+j+1} \cdot (q+pz)^{n-d-j}}{(d+r+j+1) \cdot q} \cdot$$

$$\cdot \sum_{k=o}^{n-d-j} \frac{(-1)^k \cdot (n-d-j)_{(k)}}{(d+r+j+2)_{[k]}} \cdot \left\{\frac{p}{q} \cdot \frac{p+qz}{q+pz}\right\}^k$$

sacando fuera de los signos de sumatoria el factor

$$\frac{(p+qz)^{d+r+1}}{(d+r+1) \cdot q} \cdot (q+pz)^{n-d} \quad ,$$

realizando el producto de las series y reescribiendolo en orden creciente de las potencias de los argumentos, llegamos a

$$\frac{(p+qz)^{d+r+1} \cdot (q+pz)^{n-d}}{(d+r+1) \cdot q} \cdot \sum_{i=o}^{n-d} \sum_{j=o}^{i} \left\{\frac{r}{j!}\right\} \cdot$$

$$\cdot \frac{(-1)^i \cdot (n-d)_{(i)}}{(d+r+2)_i} \cdot \left[\frac{p \cdot (p+qz)}{q \cdot (q+pz)}\right]^i$$

con lo que terminamos la demostración puesto que fácilmente se puede probar que para cualquier r la suma entre llaves vale

$$\frac{(r+1)_{[i]}}{i!}$$

A partir de la fórmula que hemos obtenido, el cálculo de los sucesivos momentos presenta la ventaja de basarse en el de los anteriores, teniendose que añadir para el cálculo del siguiente unicamente una nueva función hipergeometrica valuada en $\frac{p}{q}$ y en su cuadrado; concretamente esta función es la que corresponde a los parametros $\alpha=r$ y $\gamma=d+r+1$; los restantes terminos ya han sido previamente calculados y se emplean siempre valuados en $\frac{p^2}{q^2}$; solo necesitan ser divididos por $r-\alpha$, siendo α el primer parámetro de la hipergeométrica considerada. Por otro lado, también aquí sigue siendo válida la observación que hicimos en relación con el primer momento cuando el valor del parámetro d era inferior a la diferencia n-d , cuyo objetivo es reducir el número de cálculos implicados.

c) Momentos negativos de la binomial métrica positiva.

El momento negativo de orden r de esta variable truncada vale, por definición :

$$E\left[X_d^{-r}\right] = \frac{1}{1-b_0(d)} \cdot \sum_{k=1}^{n} b_k(d) \cdot \left\{\frac{1}{k}\right\}^r$$

Por comodidad, en las siguientes fórmulas vamos a omitir el factor $\{1-b_0(d)\}^{-1}$, el cual generalmente tendrá valores muy cercanos a la unidad.

Nosotros vamos a centrarnos principalmente en el primero de estos momentos. Como vimos en la sección I.6 no se conocen fórmulas reducidas para este momento respecto a la distribución binomial; por esta razón solo nos limitaremos a considerar los métodos aproximados de cálculo, adaptando en primer lugar los utilizados sobre la binomial.

El procedimiento de Stephan consiste en expresar $E[\frac{1}{X_d}]$ en función de los momentos inversos factoriales. Para ello se parte del siguiente desarrollo :

$$\frac{1}{x} = \frac{0!}{x+1} + \frac{1!}{(x+1)_{[2]}} + \dots + \frac{(i-1)!}{(x+1)_{[i]}} + R_i(x)$$

donde el resto $R_i(x)$ vale exactamente $i!/x_{[i+1]}$. La validez del desarrolo se comprueba facilmente por inducción.

Al tomar la esperanza matemática respecto a la variable positiva, nos falta el primer término de todos los momentos factoriales. Por tanto tenemos que

$$E\left[\frac{1}{X_d}\right] = \sum_{i=1}^{t}(i-1)!\cdot\left\{m_i(d) - \frac{1}{i!}\cdot b_o(d)\right\} + \sum_{k=1}^{n} R_t(k)\cdot b_k(d)$$

Observese que la expresión entre llaves es equivalente a la fórmula de $m_i(d)$, pero haciendo comenzar el índice por 0 en vez de por 1.

En teoría, tomando t suficientemente grande, se hace despreciable la esperanza del resto, de modo que solo intervienen los momentos inversos factoriales. En la practica este procedimiento funciona bien cuando el valor medio λ de la VBM es grande. Por ejemplo, si $\lambda>20$, se obtienen al menos 5 cifras significativas exactas con menos de 10 términos. Cuando λ es pequeño, se requieren muchos términos siendo preferible el cálculo directo. Además este método tiene el inconveniente importante de tener que manejar simultaneamente números muy grandes y muy pequeños.

Otro procedimiento mucho mas rápido y que proporciona suficiente exactitud es el empleado por Mendenhall y Lehman, consistente en ajustar la distribución discreta por una beta que iguale los dos primeros momentos y, de esta manera, utilizar los momentos negativos de la beta para aproximar a los de la VBM.

Los parámetros de la beta ajustada se pueden escribir formalmente de la misma manera que los correspondientes a la binomial clásica, es decir

$$a=(n^{\star}-1)\cdot p^{\Delta} \quad y \quad b=(n^{\star}-1)\cdot q^{\Delta} \quad ,$$

definiendo $p^{\Delta}=\frac{\lambda}{n}$, $q^{\Delta}=1-p$ y $n^{\star}=n\cdot\frac{p^{\Delta}\cdot q^{\Delta}}{p\cdot q}$

Teniendo en cuenta que los momentos negativos de la beta solo existen cuando su orden es menor que el paráme tro a y que su valor es

$$E[y^{-k}] = \frac{B(a-k,b)}{B(a,b)} = \frac{(a+b-1)_{(k)}}{(a-1)_{(k)}} \quad ,$$

podemos aproximar los momentos negativos de la binomial por dicha cantidad dividida por n^k, siempre que $k<a$. En particular tenemos

$$E\left[\frac{1}{X_d}\right] = \frac{1}{n}\cdot\frac{(n^*-2)}{(a-1)} = \frac{n^*-2}{(n^*-1)\cdot\lambda - n}$$

y

$$V\left[\frac{1}{X_d}\right] = \frac{(n^*-1)\cdot(n^*-2)\cdot q^\Delta}{((n^*-1)\lambda-n)^2\cdot((n^*-1)p^\Delta-2)}$$

Esta aproximación se hace mas precisa conforme a y b aumentan. Observese, entonces, que si limitamos p al intervalo $[0,\frac{1}{2}]$ la precisión se hace mayor cuando d aumenta, ya que $n^*\geq n$ y $p^\Delta\geq p$. Como índice del grado de precisión sirven estas apreciaciones empíricas: si $\lambda>5$ están aseguradas 2 cifras significativas de precisión; si $\lambda>10$, 3 cifras y si $\lambda>25$, 4 cifras.

A continuación vamos a ver un nuevo procedimiento basado en la convergencia de la VBM a la distribución Normal. Para ello aplicaremos el resultado de Fieller, recogido en I.11.d, según el cual la función R del cociente de dos

normales Y/X, definida como

$$R = \frac{\frac{Y}{X}\cdot m_x - m_y}{\sqrt{\left\{\frac{Y}{X}\right\}^2\cdot\sigma_x^2 + \sigma_y^2}}$$

tiene una distribución aproximadamente N(0,1), siempre que P[X<0] sea muy pequeña. Si consideramos que Y sigue la distribución degenerada N(1,0) y sustituimos X por X_d, resultará que

$$\frac{1}{X_d} \simeq \frac{1}{\lambda - R\sigma}$$

Ahora bien, suponiendo que $P\left[\frac{\sigma R}{\lambda} \geq 1\right]$ es suficientemente pequeña, obtenemos el siguiente desarrollo

$$\frac{1}{X_d} \simeq \frac{1}{\lambda}\cdot \sum_{i=0}^{\infty} \left\{\frac{\sigma}{\lambda}\cdot R\right\}^i \quad ,$$

que al tomar esperanzas y tener en cuenta que los momentos de la normal estandar valen 0 ó $(2k)!/2^k\cdot k!$, nos lleva a que

$$E\left[\frac{1}{X_d}\right] \simeq \frac{1}{\lambda}\cdot\left\{1 + \left\{\frac{\sigma}{\lambda}\right\}^2 + 3\left\{\frac{\sigma}{\lambda}\right\}^4 + 15\left\{\frac{\sigma}{\lambda}\right\}^6 + \ldots\ldots\right\}$$

Al considerar los dos primeros términos, nos queda la fórmula

$$E\left[\frac{1}{X_d}\right] \simeq \frac{\lambda^2 + \sigma^2}{\lambda^3} = \frac{E[X^2]}{E[X]^3}$$

Comparandola con la fórmula empírica sin justificación que

Grab y Savage mencionan para la binomial clásica, observamos que tiene la misma precisión, es decir, dos cifras significativas cuando $\lambda>10$, con la ventaja de poseer cierto fundamento matemático.

d) Diferencias medias generalizadas.

La diferencia media de orden r de una VBM X_d se define como

$$\Delta_r(d) = \sum_{i=o}^{n} \sum_{i\neq j} |i-j|^r \cdot b_i(d) \cdot b_j(d)$$

En la sección I.6 vimos un procedimiento teórico para el cálculo de estas cantidades a partir de la función generatriz probabilidad. Por ese procedimiento se obtienen en primer lugar las diferencias medias factoriales y a partir de ellas $\Delta_r(d)$ mediante la fórmula

$$\Delta_r(d) = \sum_{j=o}^{r} \frac{\Delta^j 0^r}{j!} \Delta_{(j)}(d)$$

donde $\Delta^j 0^r$ significa la diferencia finita de orden j del polinomio x^r en el punto 0 .

Teniendo en cuenta ese procedimiento se llega al siguiente resultado importante :

Teorema.- Las diferencias medias generalizadas de las

VBM son independientes del parámetro d. Por tanto, sus valores numéricos coinciden con los de la binomial clásica.

Demostración. Teniendo en cuenta el procedimiento teórico de su cálculo, es suficiente comprobar que

$$G_d(z)\cdot G_d(z^{-1}) = G_o(z)\cdot G_o(z^{-1}) \quad ,$$

lo cual resulta inmediato.

Como consecuencia de este teorema, disponemos de un sencillo contraejemplo de que las diferencias medias generalizadas, a diferencia de los que sucede con los momentos, no pueden determinar de modo único una distribución.

Finalmente, diremos que el presente teorema también se verifica para la distribución binomial generalizada métrica, entendiendo por esta distribución a toda variable binomial de Poisson que se obtenga a partir de un vector de probabilidad dado al cambiar cualquier número de sus componentes por las probabilidades de "fracaso". El correspondiente teorema afirmaría que todas esas variables aleatorias resultan indistinguibles entre si respecto a estas medidas de dispersión.

III.6.- RELACIONES ENTRE LAS VARIABLES BINOMIALES METRICAS

Vamos a recoger aquí las principales relaciones existentes entre las VBM cuando cada uno de sus parámetros varía individualmente. Comenzaremos viendo las correspondientes a la variación del parámetro probabilidad.

Como todos sabemos, cada situación binomial concreta admite dos descripciones diferentes: una en términos de "exitos" y otra en términos de "fracasos". Esto trae consigo que las probabilidades de la B(n,q) queden completamente determinadas en función de la B(n,p), ya que $b_k(n,q)$ coincide con $b_{n-k}(n,p)$.

Nosotros vamos a probar que al introducir el parametro d se sigue manteniendo esa misma relación entre las probabilidades de las VBM. Concretamente veremos que siempre se verifica que

$$b_k(n,d,q) = b_{n-k}(n,d,p)$$

Para probarlo, observemos primeramente que el modelo binomial métrico también admite una doble descripción; en esta ocasión, en términos de diferencias, o bien en términos de coincidencias entre las componentes de los puntos del espacio muestral

y el punto de referencia.

Siguiendo la segunda descripción, veamos que la variable $Y'_{\vec{a}}$ que mide el número de componentes coincidentes respecto al punto $\vec{a}$, situado a una distancia d del origen, se distribuye segun una ley B(n,d,q). Para ello consideremos el punto $\vec{a}^{\star}$, obtenido intercambiando las componentes de $\vec{a}$; es decir, $\vec{a}^{\star} = \vec{1} + \vec{a}$. Respecto a este punto, las coincidencias respecto a $\vec{a}$ se convierten en diferencias, y recíprocamente. Como por construcción $\vec{a}^{\star}$ solo tiene d componentes nulas, resulta que $Y'_{\vec{a}}$ sigue una ley B(n,n-d,p). Ahora bien, como los parámetros p y q son intercambiables, por ejemplo, en la función característica, podemos expresar la ley de $Y'_{\vec{a}}$ en la forma deseada. El intercambio de p y q puede interpretarse sobre Ω como una traslación del origen a punto $\vec{1}$; simultaneamente a este intercambio, la distancia de cada punto al origen pasa a valer n-d.

De este razonamiento se desprenden dos observaciones importantes. En primer lugar, cada VBM viene determinada por dos puntos diferentes del espacio paramétrico. Esta peculiaridad trae consigo que el proceso de estimación de los parámetros se vea afectado en cierta medida, ya que, por ejemplo, en la estimación simultanea de p y d siempre son es

timaciones equivalentes p y d, y,q y n-d. Para paliar esta dificultad es conveniente reducir el espacio paramétrico de modo que haya una correspondencia unívoca entre distribuciones y puntos paramétricos. Esta reducción puede realizarse de dos maneras: limitando d a variar entre 0 y $\frac{n}{2}$, o bien limitando p al intervalo $[0,\frac{1}{2}]$. Particularmente preferimos esta última forma, maxime cuando hemos interpretado p como probabilidad de transmisión erronea. A pesar de esta reducción no se consigue una correspondencia completamente unívoca, ya que cuando $p = \frac{1}{2}$ todas las distribuciones resultantes son identicas.

La segunda observación es de índole practica. Al igual que ocurre con la binomial clásica, no es necesario realizar una tabulación completa de las probabilidades binomiales métricas. En este caso, es suficiente realizarla para valores de p comprendidos entre 0 y $\sqrt{2}$, y valores de d comprendidos entre 0 y $\frac{n}{2}$.

Veamos ahora las relaciones de recurrencia que ligan a las VBM cuyo parámetro n difiere en una unidad. Dado que las componentes de los puntos de Ω son estadisticamente independientes, toda variable B(n,d,p) puede expresarse como la suma de las variables B(n-1,d,p) y B(1,0,p).De aquí se

obtienen inmediatamente las siguientes relaciones entre las probabilidades puntuales y acumuladas:

$$b_k(n,d,p) = p \cdot b_{k-1}(n-1,d,p) + q \cdot b_k(n-1,d,p)$$

y

$$B_k(n,d,p) = B_{k-1}(n-1,d,p) + q \cdot b_k(n-1,d,p)$$

Estas relaciones fueron empleada en la sección anterior para el cálculo de la desviación media de la distribución, donde también vimos otra forma de expresar B(n,d,p),en función de B(n-1,d-1,p),que da lugar a las igualdades:

$$b_k(n,d) = q \cdot b_{k-1}(n-1,d-1) + p \cdot b_k(n-1,d-1)$$

y

$$B_k(n,d) = B_{k-1}(n-1,d-1) + p \cdot b_k(n-1,d-1)$$

Veamos, por último, las relaciones que corresponden a la variación de solo el parámetro d. Para ello apliquemos las igualdades que acabamos de ver para expresar las probabilidades de B(n,d) y B(n,d-1) en función de B(n-1,d-1). Haciendo esto, obtenemos :

$$b_k(n,d) = b_k(n,d-1) - (q-p) \cdot \Delta b_{k-1}(n-1,d-1)$$

y

$$B_k(n,d) = B_k(n,d-1) - (q-p) \cdot b_k(n-1,d-1) \quad .$$

Esta última igualdad puede ser expresada en forma de teorema, bajo el siguiente enunciado :

Teorema.- Para valores de n y p fijos, con $p \leq \frac{1}{2}$, la función de distribución de las VBM, $B_k(d)$, es una función decreciente del parámetro d, salvo que $k = n$, en cuyo caso vale la unidad para todas las variables de la familia.

III.7.- SUMA Y DIFERENCIA DE VBM. CARACTERIZACION DE LA LEY

Vamos a mostrar en esta sección que tanto la suma como la diferencia de VBM independientes con parámetro p comun siguen también distribuciones binomiales métricas, con la particularidad de que la distribución de la diferencia está desplazada a la izquierda. Igualmente probaremos que estos resultados pueden invertirse, en el sentido de que si la suma o diferencia de variables independientes sigue una DBM, necesariamente ambas variables serán binomiales métricas. Veamos en primer lugar los resultados relativos a su suma.

Teorema de adición.- Sean X y Y VBM independientes de parámetros (n_x, d_x, p) y (n_y, d_y, p), entonces su suma Z se distribuye segun la ley $B(n_x+n_y, d_x+d_y, p)$.

La demostración es inmediata realizando el pro-

ducto de las funciones características de los sumandos.

Teorema de Caracterización.- Sea Z una variable aleatoria obtenida como suma de dos variables aleatorias no negativas independientes y que se distribuye según una ley B(n,d,p). Entonces cada sumando sigue DBM cuyos parámetros han de verificar: $n_x+n_y=n$, $d_x+d_y=d$, y $p_x=p_y=p$.

La demostración se basa en la estructura polinómica de $G_z(t)$, la función generatriz de probabilidad de Z. Al exigir que los sumandos X e Y sean no negativos, se deduce que ambos han de tomar el valor 0 y de lo cual se deduce también que sus restantes valores solo podrán ser enteros. Por tanto existen sus respectivas funciones generatrices de probabilidad, que verificarán

$$G_z(t) = G_x(t) \cdot G_y(t)$$

Ahora bien, como los factores $(q+pt)$ y $(p+qt)$ de $G_z(t)$ son polinomios irreducibles, son también los únicos factores que pueden formar parte de G_x y G_y. Además las condiciones que han de verificar los parámetros son consecuencia del número de factores de cada tipo que contiene $G_z(t)$.

Pasemos a considerar ahora la diferencia $X-Y$, que notaremos por V. Esta variable está definida sobre los enteros comprendidos entre $-n_y$ y n_x. Para identificar su distribución conviene desplazarla n_y unidades, de modo que obtengamos la variable no negativa

$$V + n_y = X - Y + n_y = X + (n_y - Y)$$

Como vimos en la sección anterior, $(n_y - Y)$ tambien sigue una DBM; concretamente la de parámetros $(n_y, n_y - d_y, p)$. Por tanto, aplicando el teorema de adición, $V+n_y$ sigue la ley BM $B(n_x+n_y, d_x+n_y-d_y, p)$. En otras palabras, tenemos el siguiente resultado:

> Teorema de Substracción.- La diferencia de VBM independientes también sigue una ley BM, pero desplazada a la izquierda.

Como corolario tenemos que la diferencia de binomiales clásicas, tratada en I.7.d, sigue una $B(n_x+n_y, n_y, p)$. Casualmente, cuando $p=\frac{1}{2}$, se identificó como binomial clásica desplazada porque en ese caso particular todas las VBM son iguales.

Veamos ahora el teorema inverso que sirve tambien

como caracterización de la ley.

Teorema de Caracterización.- Sea V, $V = X - Y$, la diferencia de dos variables independientes no negativas que asignan probabilidades positivas al punto 0. Si existe un entero n tal que $(V+n)$ se distribuye según una $B(n_v, d_v, p)$, entonces las variables X e Y son también VBM cuyos parámetros verifican: $n = n_y$, $n_x = n_v - n$ y $d_x - d_y = d_v - n$.

La demostración es completamente análoga a la del otro teorema de caracterización. La hipótesis de que X e Y tomen el valor cero con probabilidad no nula es necesaria para asegurar que sean VBM no desplazadas.

Todos los resultados de esta sección se pueden trasladar directamente a las variables binomiales generalizadas, de manera que se verifica que tanto la suma como la diferencia de esas variables pertenece y caracteriza a su familia.

III.8.- PRINCIPALES PROPIEDADES

Las propiedades que vamos a exponer en esta sección van a tratar de las siguientes cuestiones: a) Forma y clasificación de la distribución; b) Localización de la moda y la mediana; c) Relaciones entre las esperanzas de variables con valores comunes de la media y del número de pruebas; d) Generalización del Teorema de Jogdeo y Samuels ; e) Convergencia hacia la distribución de Poisson desplazada.

En la siguiente sección veremos otras propiedades importantes de las VBM, pero todas ellas serán corolarios de resultados mas generales sobre variables binomiales de Poisson. En esta sección emplearemos también la pertenencia de las VBM a las binomiales generalizadas para demostrar las propiedades de los 4 primeros apartados; para ello aplicaremos los principales resultados sobre estas distribuciones, que hemos recogido en la sección I.12.

a) Forma y clasificación de la distribución.-

Como demostraron Darroch y Samuels la densidad de probabilidad discreta de las variables binomiales generalizadas tiene una forma acampanada de orden r, cualquiera

que sea r. Esto supone que las probabilidades puntuales de las VBM respecto a k pasan, aunque en casos concretos no se presenten todas, sucesivamente por las siguientes etapas: crecimiento convexo, crecimiento concavo, decrecimiento concavo y decrecimiento convexo.

Esta forma de la densidad de probabilidad clasifica a la distribución como miembro de la familia unimodal. Ahora bien, como consecuencia de que el logaritmo de la densidad puntual es una función concava y del teorema de Keilson y Gerber, las distribución binomial métrica pertenece también a la familia de distribuciones fuertemente unimodales. Esto significa que su convolución con cualquier otra distribución discreta unimodal será siempre unimodal.

Con respecto a otras familias de distribuciones que incluyen a la binomial clásica, vamos a comprobar en primer lugar, mediante contraejemplos, que la BM no pertenece ni al sistema de Pearson discreto ni a la familia de distribuciones de series generalizadas de potencias. Para la primera familia tenemos que la ecuación en diferencias finitas cuando el parámetro d vale 1 es la siguiente

$$\frac{\Delta b_{k-1}(1)}{b_{k-1}(1)} = \frac{(n+1)p - k + \frac{p(q-p)\cdot\{n-(k-1)\}}{np+(k-1)\cdot(q-p)}}{k\cdot q}$$

La no-pertenencia a la familia de distribuciones de series generalizadas de potencias se demuestra por la imposibilidad de que exista un parámetro θ que verifique la relación de recurrencia entre los cumulantes para valores del parámetro d distintos de 0 ó n; tal parámetro θ debería valer simultaneamente

$$\left\{\frac{p}{q}\right\}^{\frac{n}{n-2d}} \qquad y \qquad \left\{\frac{p}{q}\right\}^{\frac{n-2d}{n}} .$$

Debido a la estructura de la función de masa de probabilidad se puede comprobar que las VBM tampoco pertenecen a la familia exponencial respecto al parámetro p. Al estudiar en la siguiente sección la estimación de este parámetro se verán otras razones que impiden también dicha pertenencia. Por otro lado, hemos dejado sin resolver si nuestra familia pertenece o no a la familia de razón de verosimilitud monótona y a la de razón de fallo creciente.

b) Localización de la moda y la mediana.-

Los resultados generales sobre la localización de estas medidas en la binomial de Poisson aplicados a las VBM afirman que: si su media λ es número entero, entonces la mediana y la moda coinciden con ella; si no es un valor ente-

ro, entonces ambas solo se pueden localizar en los enteros adyacentes.

En el caso concreto de la moda se puede afinar mas, ya que disponemos de los resultados de Darroch y Samuels (Vease I.12.C). El primero proporciona unos intervalos racionales sencillos en los que el valor de λ localiza univocamente la moda. El segundo da condiciones suficientes que localizan la moda en función de λ y la componente máxima o mínima del vector de probabilidad de la distribución. En el caso particular de la binomial métrica estas condiciones son:

$$\lambda + q < [\lambda+1] \qquad Mo = [\lambda]$$

$$\lambda + p > [\lambda+1] \qquad Mo = [\lambda+1]$$

Aparte de la particularización de los resultados anteriores, se obtiene un importante resultado de otro tipo teniendo en cuenta las peculiaridades de las VBM:

Teorema.- La mediana y la moda de la binomial métrica no decrecen al aumentar el parámetro d, cuando los otros parámetros permanecen fijos.

La demostración del teorema es equivalente a pro

bar las siguientes implicaciones:

$$Me(d-1) = k+1 \quad \Longrightarrow \quad Me(d) = k+1$$

y

$$Mo(d-1) = k+1 \quad \Longrightarrow \quad Mo(d) = k+1$$

para las variables B(d-1) y B(d) cuyos valores medios están comprendidos entre los enteros k y k+1 .

La demostración de la implicación sobre la mediana es consecuencia del teorema que dimos en la sección III.6 sobre el decrecimiento respecto al parámetro d de la función de distribución, ya que

$$B(k;d) < B(k;d-1) < 0.5$$

Para probar la otra implicación se necesita aplicar las condiciones de Samuels y las relaciones entre las probabilidades puntuales de ambas distribuciones. Empleando estas relaciones, tenemos

$$b(k+1;d) = b(k+1;d-1) - (q-p)\Delta b(k;n-1,d-1)$$

y

$$b(k;d) = b(k;d-1) - (q-p)\Delta b(k-1;n-1,d-1)$$

de modo que su diferencia vale

$$\Delta b(k;d) = \Delta b(k;d-1) - (q-p)\,\Delta^2 b(k-1;n-1,d-1)$$

Ahora bien, la segunda diferencia finita es negativa como consecuencia de que las condiciones de Samuels aseguran que la moda de $B(n-1,d-1)$ se localiza en k, ya que su correspondiente valor medio verifica

$$\lambda(n-1,d-1) \geq k \text{ , porque } \lambda(n-1,d-1) + 1 = \lambda(n,d-1) + q \geq k + 1 \text{ , y}$$

$$\lambda(n-1,d-1) + q = \lambda(n,d) < k + 1$$

Observese que de esta parte de la demostración no solo se deduce que la diferencia $\Delta b_k(d)$ es positiva, sino que además es mayor que $\Delta b_k(d-1)$. Esta afirmación es muy probable que siga siendo cierta cuando ambas distribuciones tengan igual moda, pero las partes enteras de sus medias sean diferentes; sobre todo cuando las varianzas sean mayores que 1. El motivo de esto se encuentra en la forma concava que suele presentar la distribución en las proximidades de la moda. Desgraciadamente, por el momento, no hemos encontrado un procedimiento de demostración.

c) Relaciones entre las esperanzas de las VBM con valores de λ y n comunes.-

Vamos a probar que las VBM como variables binomiales de Poisson verifican los resultados de Gleser. En otras palabras, vamos a ver que los vectores de probabilidad de VBM que pertenezcan a un cierto $\mathcal{T}_\lambda$ siempre están ordenados por mayorización.

Teniendo en cuenta la definición de mayoración entre vectores, dada en el punto I.12.A, y que hemos limitado al parametro p a variar en el intervalo $[0, \frac{1}{2}]$, obtenemos el siguiente

Teorema.- Sean $B(n,d_1,p_1)$ y $B(n,d_2,p_2)$ dos VBM con el mismo valor medio. Notemos por $\vec{p}^{[1]}$ y $\vec{p}^{[2]}$ a sus respectivos vectores de probabilidad. Entonces se verifica que $p_1 \leq p_2$ es equivalente a $\vec{p}^{[1]} \geq \vec{p}^{[2]}$.

Demostración. Los vectores de probabilidad, escritas sus componentes en orden decreciente, son:

$$\vec{p}^{(1)} = (\underbrace{q_1,\ldots,q_1}_{d_1},p_1,\ldots,p_1)$$

$$\vec{p}^{(2)} = (\underbrace{q_2,\ldots,q_2}_{d_2},p_2,\ldots,p_2)$$

con $p_1 \leq p_2$ y $q_1 \geq q_2$

La implicación hacia la derecha vamos a probarla por reducción al absurdo. Supongamos que existiese un j tal que

$$\sum_{i=1}^{j} p^{(1)}_{[i]} < \sum_{i=1}^{j} p^{(2)}_{[i]}$$

Como la suma de todas las componentes de cada vector es la misma y como p_1 es inferior a p_2, para conseguir dicha igualdad no tenemos mas remedio que añadir a la primera sumatoria componentes iguales a q_1. Esto significa que

$$\sum_{i=1}^{j} p^{(1)}_{[i]} = j \cdot q_1 \quad ,$$

que siempre es superior a la otra sumatoria cualquiera que sean sus componentes.

Veamos ahora la implicación inversa. Distinguiremos dos casos: d_2 inferior a n y d_2 igual a n. En el primer caso tenemos que

$$\sum_{i=1}^{n-1} p_{[i]} \geq \sum_{i=1}^{n-1} p_{[i]} \implies p_1 < p_2$$

En el otro caso, debe verificarse $q_1 \geq q_2$, por tanto $p_1 \leq p_2$.

Como consecuencia de este teorema, aplicando directamente los resultados de Gleser, tenemos las siguientes

propiedades:

i) Si g(X) es una función convexa, entonces

$$E[g(X_1)] \leq E[g(X_2)] \quad .$$

En particular, esta desigualdad se verifica para todo momento absoluto de orden superior o igual a la unidad.

ii) Entre sus correspondientes funciones de distribución se cumplen las desigualdades:

$$B_k(p_1) \leq B_k(p_2) \quad \text{cuando} \quad 0 \leq k \leq [\lambda - 2]$$

y

$$B_k(p_1) \geq B_k(p_2) \quad \text{cuando} \quad [\lambda + 2] \leq k \leq n \quad .$$

A estos resultados generales, podemos añadir el que se desprende del caso particular en que X_1 localiza unívocamente su moda por las condiciones de Darroch o Samuels.

Teorema.- Sean $B(n,d_1,p_1)$ y $B(n,d_2,p_2)$ dos VBM con el mismo valor medio. Si X_1 localiza univocamente su moda y $p_1 < p_2$, entonces X_2 presenta la misma moda y se verifica la siguiente desigualdad :

$$B_{[\lambda-1]}(p_1) \leq B_{[\lambda-1]}(p_2) \quad \text{,si la moda es } [\lambda+1]$$

$$B_{[\lambda+1]}(p_1) \geq B_{[\lambda+1]}(p_2) \quad \text{,si la moda es } [\lambda] \quad .$$

Demostración.- Es inmediato comprobar que X_1 y X_2 tienen la misma moda: cuando se verifican las condiciones de Darroch esto siempre es cierto; cuando X_1 verifica las condiciones de Samuels, X_2 también las verifica porque la mayorización asegura que la componente mínima del vector mayorado no disminuye y la máxima no aumenta.

Para probar la validez de las relaciones entre las funciones de distribución, sigamos el razonamiento del teorema de Gleser. Consideremos el conjunto de todas las variables binomiales de Poisson X_g con igual valor medio que X_1 y tales que las componentes del vector de probabilidad verifiquen

$$p_1 \leqslant \min p_1^g, p_2^g, .., p_n^g \leqslant \max p_1^g, p_2^g, .., p_n^g \leqslant q_1$$

Veamos que, según donde localice X_1 la moda, la variable X_g^{ij} jamás podrá tener determinada moda. Cuando la moda de X_1 es $[\lambda]$, las condiciones de Darroch y Samuels impiden que X_g^{ij} localice su moda en $[\lambda+1]$. Analogamente, en el otro caso, $[\lambda-2]$ nunca puede ser la moda de X_g^{ij}.

Observese que, como el proceso de demostración no utiliza ninguna propiedad especifica de las VBM, este

teorema será valido para las variables binomiales de Poisson si existe mayoración entre sus vectores de probabilidad. Seguidamente, vamos a dar otro teorema sobre estas variables que aplicaremos en la siguiente sección.

Teorema.- Sean X_1 y X_2 variables binomiales generalizadas definidas por los vectores de probabilidad $\vec{p}^{(1)}$ y $\vec{p}^{(2)}$. Si $\vec{p}^{(1)} \trianglelefteq \vec{p}^{(2)}$ y se verifica

$$\sum_{i=1}^{n} p_i^{(1)} = \lambda \in Z \quad y \quad \frac{1}{3} < p_{\underline{n}}^{(1)} \leq p_{\underline{1}}^{(1)} < \frac{2}{3} \quad ,$$

siendo $p_{\underline{1}}^{(1)}$ y $p_{\underline{n}}^{(1)}$, respectivamente, la componente máxima y mínima de $\vec{p}^{(1)}$, entonces se tiene

$$B_g(k/\vec{p}^{(1)}) \leq B_g(k/\vec{p}^{(2)}) \quad , \quad \text{si} \quad k < \lambda$$

y

$$B_g(k/\vec{p}^{(1)}) \geq B_g(k/\vec{p}^{(2)}) \quad , \quad \text{si} \quad k \geq \lambda \quad .$$

Demostración. Basta comprobar que cualquiera que sea X_g, con media λ y probabilidades de "exito" comprendidas entre $\frac{1}{3}$ y $\frac{2}{3}$, siempre dará lugar a variables X_g^{ij} con la moda localizada en $\lambda-1$. Esto es cierto porque la media ν de dicha variable verifica las condiciones de Samuels, ya que

$$\lambda - \frac{4}{3} < \lambda - p_i - p_j < \lambda - \frac{2}{3} \quad .$$

Naturalmente, de este teorema se desprende el correspondiente corolario para las VBM. Por otro lado, como bajo estas condiciones tan restrictivas hemos logrado una generalización completa del 4º teorema de Hoeffding, también se obtiene un importante corolario sobre las probabilidades de los intervalos alrededor del valor medio.

Corolario.- Bajo las hipótesis del teorema anterior, se verifica que

$$P\left[j \leq X_1 \leq k\right] \geq P\left[j \leq X_2 \leq k\right]$$

siempre que $0 \leq j \leq \lambda \leq k \leq n$.

Emplearemos este corolario en la siguiente sección al tratar de la mayoración entre distribuciones.

d) Generalizaciones del teorema de Jogdeo y Samuels.

Vamos a dar dos generalizaciones de dicho teorema que hemos obtenido al ampliar sus hipótesis; nuestros resultados, aunque ligeramente menos generales, incluyen todas las posibilidades reales. La primera generalización se refiere unicamente a la binomial clásica; la segunda a la binomial métrica.

Como vimos en la sección I.13, en las hipótesis del teorema de Jogdeo y Samuels se exige que los valores medios de una sucesión de variables binomiales sean números enteros; bajo esta condición no se logra englobar completamente al teorema de Anderson y Samuels, salvo en el caso particular de que el valor medio considerado sea entero. Hemos observado que en la demostración del teorema de Jogdeo y Samuels lo realmente esencial es que las diferencias entre los valores medios sean enteros. Con esta ampliación de las hipótesis, se reduce la validez de la tesis en un solo caso.

Teorema.- Sea $\{X_m\}$ una sucesión de variables binomiales que verifica:

i) La sucesión $\{n_m\}$ es creciente; las sucesiones $\{n_m p_m\}$ y $\{n_m q_m\}$ son no decrecientes.

ii) Para cualquier m, la diferencia $\{n_m p_m - n_l p_l\}$ es siempre un número entero.

Entonces, la sucesión $\{B(n_m p_m + r; n_m, p_m)\}$ es estrictamente creciente o decreciente, según sea r un entero negativo o positivo. En el caso particular de que los promédios fuesen enteros, dicha sucesión tambien será decreciente para $r = 0$.

FACULTAD DE CIENCIAS
BIBLIOTECA
GRANADA

La demostración se basa en el 4º teorema de Hoeffding (Vease I.12.A). Consideremos dos variables de la sucesión X_m y $X_{m'}$, con $m<m'$. Veamos que X_m se puede transformar en una variable binomial de Poisson de igual media y número de pruebas que $X_{m'}$; consideremos la variable ${}_gX_m$ generada por el vector de probabilidad obtenido al añadir al vector de X_m $(n_{m'}-n_m)$ componentes, de las cuales $(n_{m'}p_{m'}-n_mp_m)$ sean unitarias y las restantes nula. Por definición esta nueva variable verifica

$$P[{}_gX_m \leq n_{m'}p_{m'} + r] = B(n_mp_m+r;n_m,p_m)$$

de modo que basta aplicar el teorema de Hoeffding para terminar la demostración.

La generalización que vamos a dar para las VBM se basa en el teorema de Gleser, de manera que solo tendremos que comprobar que nuestras hipótesis garantizan la ordenación por mayoración de sus vectores de probabilidad. Desgraciadamente, los resultados no son tan fuertes, ya que no se asegura la monotonía cuando $|r|\leq 1$.

Teorema.- Sea $\{B(n_m,d_m,p_m)\}$ una sucesión de VBM que verifica las siguientes condiciones:

i) La sucesión $\{n_m\}$ es creciente.

ii) La sucesión $\{p_m\}$ es no decreciente, bajo el supuesto de que ese parámetro se elige de manera que no supere a $\frac{1}{2}$.

iii) Si $\{\lambda_m\}$ es la sucesión de los valores medios, dicha sucesión y $\{n_m-\lambda_m\}$ son sucesiones no decrecientes, y además la diferencia $\lambda_m-\lambda_1$ es siempre un número entero.

Entonces, la sucesión $\{B(\lambda_m+r)\}$, con $|r|\geq 2$, es no decreciente o no creciente, según sea r negativo o positivo.

Demostración: Vamos a comprobar que si $m<m'$, el vector de probabilidad de la variable ${}_gX_m$, definida en la prueba del teorema anterior, siempre dominará al vector de probabilidad de la VBM $B(n_m,d_m,p_m)$. Dichos vectores de probabilidad son

$$\overrightarrow{p}^{(m)} = (1,\ldots,1,q_m,..,q_m,p_m,\ldots,p_m,0,..,0)$$

y

$$\overrightarrow{p}^{(m)} = (q_{m'},\ldots\ldots,q_{m'},p_{m'},\ldots\ldots,p_{m'}) \quad ,$$

con $\quad 0\leq p_m\leq p_{m'}\leq \frac{1}{2}\leq q_{m'}\leq q_m\leq 1$.

Razonemos por reducción al absurdo. Supongamos que existiera un índice j tal que

$$\sum_{i=1}^{j} p^{(m)}_{[i]} > \sum_{i=1}^{j'} p^{(m)}_{[i]} + \lambda_{m'} - \lambda_{m} \quad ,$$

entonces, como $p_m < p_{m'}$ y la suma de todas las componentes es la misma, deben existir todavía componentes del vector $\vec{p}^{(m)}$ iguales a q_m, pero eso es contradictorio porque en tal caso la 1ª sumatoria no podría superar jamás a la 2ª.

Queremos observar que este teorema desafortunadamente no engloba totalmente al teorema de Jogdeo y Samuels. Esto se debe a que la hipótesis del no decrecimiento de p excluye a muchas sucesiones que lo verificarían. No obstante, dicha hipótesis no es superflua como se comprueba considerando a las VBM B(7,0,0.4) y B(8,1,0.3). Ambas distribuciones tienen como valor medio 2.8, de modo que si pudiesemos prescindir de la hipótesis sobre p, debería verificarse que $b_o(7,0,0.4)$ es menor que $b_o(8,1,0.3)$; en la realidad se verifica lo contrario.(En este caso ambas variables tienen también igual varianza. Creemos que los teoremas de este punto y el anterior serán ciertos, aun cuando no habiendo ordenación por mayoración, haya desigualdad de varianzas.)

e) Convergencia hacia la ley de Poisson desplazada.

Consideremos una sucesión de variables binomiales clásicas que verifica las condiciones para garantizar la convergencia hacia una distribución de Poisson de parametro λ. Veamos hacia que distribución converge cada una de las sucesiones de VBM que se generan para cada valor del parametro d.

Es facil comprobar que la correspondiente sucesión de funciones generatrices verifica:

$$\lim_{\substack{n\to\infty \\ n\cdot p\to\lambda}} G(z;n,d) = z^d \cdot e^{\lambda(z-1)}$$

Por tanto, la distribución límite es una Poisson desplazada d unidades a la derecha.

Veamos que dicha convergencia es tan fuerte como la que presenta la binomial clásica; en otras palabras, vamos a generalizar el teorema de Simons y Johnson (Vease el punto I.10.d).

Teorema.- Sea h(k) cualquier función positiva cuya esperanza respecto a la ley de Poisson de parámetro λ exis

te, entonces se verifica que

$$\lim_{n\to\infty} \sum_{k=0}^{\infty} h(r)\,|b(k+d;n,d,\frac{\lambda}{n}) - p(k;\lambda)| = 0$$

La clave de la demostración del teorema primitivo es probar que a partir de un cierto k se verifica que el cociente

$$\frac{b(k;d)}{p(k-d)} : \frac{b(k-1;d)}{p(k-1-d)} < 1$$

Vamos a probar por inducción matematica que esa desigualdad sigue cumpliendose para cualquier d finito. Para ello vamos a aplicar las relaciones de la sección 6 que muestran que

$$B(k;n,d) = p\cdot b(k;n-1,d-1) + q\cdot b(k-1;n-1,d-1)$$

Sustituyendo en el cociente, tenemos

$$\frac{p(k-d)}{p(k-d+1)} \cdot \frac{p\cdot b(k;n-1,d-1) + q\cdot b(k-1;n-1,d-1)}{p\cdot b(k-1;n-1,d-1) + q\cdot b(k-2;n-1,d-1)} =$$

$$= \frac{k-d}{\lambda} \cdot \frac{b(k-1;n-1,d-1)}{b(k-2;n-1,d-1)} \cdot \frac{p\cdot b(k)/b(k-1) + q}{p\cdot b(k-1)/b(k-2) + q} < 1$$

ya que las dos primeras fracciones son menores que 1 por la hipótesis de inducción y la otra lo es por la

log-concavidad de b(k).

Debido a que esta razón es menor que 1, el máximo del cociente b(k;d)/p(k-d) es finito, resultando de ello que se verifica que

$$h(k)\,|b(k+d)-p(k)| = h(k)\cdot p(k)\left|\frac{b(k+d)}{p(k)} - 1\right| < h(k)(S+1)\cdot p(k)$$

es decir, que la serie de la tesis está dominada por una serie convergente. Esto significa que a partir de un k suficientemente grande la cola de la serie puede hacerse tan pequeña como queramos. La suma de los primeros términos se hace tan pequeña como queramos avanzando en la sucesión de VBM.

Como veremos en la última sección, la distribución de Poisson desplazada no es un buen aproximante de la binomial métrica, ya que presenta un error creciente respecto al parametro d. Allí daremos un aproximante cuyo error será practicamente independiente del valor de dicho parámetro.

III.9.- ENTROPIA Y ENERGIA INFORMACIONAL DE LAS VARIABLES BINOMIALES GENERALIZADAS

En esta sección vamos a estudiar el comportamiento que tienen sobre la familia de las variables binomiales de Poisson la entropía de Shannon y la energía informacional de Onicescu. Estas dos magnitudes son dos medidas opuestas asociadas al concepto de aleatoriedad o incertidumbre; generalmente, a un incremento de aleatoriedad le corresponde un aumento de entropía y una disminución de energía informacional.

Teniendo en cuenta que el concepto de aleatoriedad es muy semejante al de dispersión, pensamos que sería de esperar que la entropía y el opuesto de la energía informacional se comportaran de igual modo que los momentos absolutos; esto es, que verificaran los resultados de Hoeffding y Gleser. Como veremos a continuación, la entropía los verifica completamente,mientras que la energía informacional solo de modo parcial. Veamos primero las definición de ambas medidas.

La entropía y la energía informacional de una variable aleatoria se definen, respectivamente, como los valo

res esperados del logaritmo del inverso de la función de densidad y de la propia densidad de probabilidad. En el caso concreto de las variables binomiales generalizadas valen

$$H(X_g) = -\sum_{k=1}^{n} b_g(k) \cdot \ln b_g(k) \quad y \quad \mathcal{E}(X_g) = \sum_{k=1}^{n} b_g^2(k) \quad .$$

Pasemos a estudiar el comportamiento de la entropía. Comencemos con el resultado analogo al corolario de Gleser. Aunque la probabilidad puntual $b_g(k)$ es una función log-concava respecto de k, nuestro resultado no se deduce directamente de dicho corolario, porque b_g no es independiente del vector de probabilidad que engendra cada distribución. Para demostrarlo se requiere además la aplicación del lema de Gibbs, cuya demostración puede encontrarse en el libro de Gil, pag. 22. Su enunciado es el siguiente:

Lema de Gibbs.- Dados dos sistemas de números $r_1, r_2, \ldots, r_n$ y $s_1, s_2, \ldots, s_n$ no negativos y tales que

$$\sum_{k=1}^{n} r_k = \sum_{k=1}^{n} s_k \quad ,$$

siempre se verifica

$$-\sum_{k=1}^{n} r_k \cdot \ln r_k \leq -\sum_{k=1}^{n} r_k \cdot \ln s_k$$

cumpliendose la igualdad unicamente se $r_k = s_k \quad \forall k$.

Veamos nuestro resultado y dos corolarios que de él se desprenden :

Teorema.- Sean X_1 y X_2 dos variables binomiales generalizadas pertenecientes a $\mathcal{P}_\mu$; sean $\vec{p}^{(1)}$ y $\vec{p}^{(2)}$ sus correspondientes vectores de probabilidad. Si $\vec{p}^{(1)}$ domina estrictamente a $\vec{p}^{(2)}$, $\vec{p}^{(1)} \succ \vec{p}^{(2)}$, se verifica entonces que $H(X_1) < H(X_2)$.

Corolario.- Sean X_1 y X_2 dos VBM con iguales valores del parametro n y de la media. Entonces $p_1 < p_2$ equivale a que $H(X_1) < H(X_2)$.

Corolario.- La única distribución de $\mathcal{P}_\mu$ que maximiza la entropía de Shannon es la distribución binomial clásica.

Demostración del teorema.- Apliquemos el corolario de Gleser para la función convexa

$$g(X) = -\ln b_g(X/\vec{p}^{(2)}) \quad ,$$

obteniendose así

$$E[g(X)/\vec{p}^{(1)}] \leq E[g(X)/\vec{p}^{(2)}] \quad ,$$

es decir

$$-\sum_{k=1}^{n} b_g(k/\bar{p}^{(1)}) \cdot \ln b_g(k/\bar{p}^{(2)}) \leq H(X_2) \quad .$$

El eslabón restante se obtiene al aplicar el lema de Gibbs, tomando como r_k a $b_g(k/\bar{p}^{(1)})$ y como s_k a $b_g(k/\bar{p}^{(2)})$.

Seguidamente vamos a considerar en que condiciones se alcanza el máximo absoluto de la entropía correspondiente a un número dado de pruebas de Poisson. Vamos a probar, como era de esperar, que dicho máximo se alcanza cuando las probabilidades de "exito" y "fracaso" de cada prueba coinciden.

Teorema.- El máximo absoluto de la entropía del número de "exitos" de un número dado de pruebas de Poisson se alcanza cuando la probabilidad de "exito" de cada una de las pruebas vale $\frac{1}{2}$.

Demostración.- Por nuestro último corolario, sabemos que el máximo solo tiene que buscarse sobre pruebas de Bernoulli. Ahora bièn, como en estas condiciones el número de "exitos" y "fracasos", salvo el orden, definen la misma distribución, resulta que $H(n,p)$ es una función simétrica respecto a $p=\frac{1}{2}$, con lo cual está asegurado que se anula la derivada de la entropía en ese punto.

Comprobemos que corresponde a un máximo y que no existe ningún otro posible extremo. Lo haremos viendo que la segunda derivada de la entropía es siempre es negativa. Su primera derivada vale :

$$\frac{dH}{dp} = -\sum_{k=1}^{n} \frac{db_k}{dp} \cdot \left\{ \ln b_k + 1 \right\}$$

Expresando las derivadas de b_k en términos de las probabilidades puntuales de la binomial con una prueba menos, se llega a

$$\frac{dH}{dp} = -n \cdot \sum_{k=o}^{n-1} b_k(n-1) \cdot \ln\left\{ \frac{p}{q} \cdot \frac{n-k}{k+1} \right\}$$

Volviendo a derivar y haciendo la misma sustitución en las derivadas de las probabilidades puntuales, puede comprobarse que

$$\frac{d^2H}{dp^2} = -n \cdot \left\{ (n-1) \sum_{k=1}^{n-2} (b_{k-1}(n-2) - b_k(n-2)) \cdot \ln \frac{p}{q} \cdot \frac{n-k}{k+1} + \frac{1}{pq} \right\} =$$

$$= -n \cdot \left\{ \frac{1}{pq} + (n-1) \sum_{k=o}^{n-2} b_k(n-2) \cdot \ln\left\{ \frac{k+1}{k+2} \cdot \frac{n-k-1}{n-k} \right\} \right\} =$$

$$= -n \cdot \left\{ \frac{1}{pq} + (n-1) \cdot E\left[\ln \frac{k+1}{k+2} + \ln \frac{n-k-1}{n-k} \; ; n-2, p \right] \right\}$$

Vamos a demostrar que la expresión entre llaves es siempre positiva. Para ello vamos a acotar inferiormente cada sumando de la esperanza. Partamos del conocido desa-

rrollo del logaritmo de (1-x)

$$\ln(1-x) = -(x+\frac{x^2}{2}+\frac{x^3}{3}+\frac{x^4}{4}+\cdots) \quad |x|<1$$

y sustituyamos los denominadores por la unidad, con lo que obtendremos la progresión geométrica cuya suma vale $x/(1-x)$. Aplicandolo a nuestro caso, tenemos

$$\ln\frac{k+1}{k+2} = \ln(1-\frac{1}{k+2}) > -\frac{\frac{1}{k+2}}{1-\frac{1}{k+2}} > -\frac{1}{k+1}$$

y analogamente

$$\ln\frac{n-k-1}{n-k} > -\frac{1}{n-k-1} \quad ,$$

de modo que

$$E\left[\ln\frac{k+1}{k+2}\cdot\frac{n-k-1}{n-k}\right] > -\left\{E\left[\frac{1}{k+1};n-2,p\right] + E\left[\frac{1}{n-k-1};n-2,p\right]\right\}$$

Ahora bien, la primera esperanza representa al primer momento inverso factorial de la binomial B(n-2,p). Analogamente sucede con la otra esperanza, pero respecto a la distribución de los "fracasos". Teniendo en cuenta, vease I.6.b, que sus valores son respectivamente

$$\frac{1}{(n-1)p}\cdot(1-q^{n-1}) \quad y \quad \frac{1}{(n-1)q}\cdot(1-p^{n-1}) \quad ,$$

llegamos finalmente a que

$$\frac{d^2H}{dp^2} < -n\left\{\frac{1}{pq} - \frac{1-q^{n-1}}{p} - \frac{1-p^{n-1}}{q}\right\} < -\frac{n(q^n+p^n)}{pq} < 0$$

Como resultado del procedimiento de demostración que hemos empleado, se obtiene la generalización para cualquier número de pruebas de Bernoulli de la propiedad que tiene la entropía de una única prueba de ser una función concava y simétrica respecto a $p = \frac{1}{2}$.

Corolario.- Cualquiera que sea el número de pruebas de una variable binomial su entropía siempre es una función concava y simétrica respecto a $p = \frac{1}{2}$.

Pasemos ahora a describir el comportamiento de esta medida respecto al número de pruebas de Poisson. Como parece razonable, comprobaremos que es siempre no decreciente. Esta propiedad se deduce facilmente a partir de una versión simplificada de la desigualdad de Jensen para funciones concavas.

Lema: Sea $y=f(x)$ una función cóncava en el intervalo $[a,b]$. Si x_1 y x_2 son valores interiores al intervalo y si p y q son dos números arbitrarios positivos cuya suma es la unidad, se tiene

$$pf(x_1) + qf(x_2) < f(px_1+qx_2)$$

Su demostración puede encontrarse en el libro de

A.Yaglon y I.Yaglon (1969), pagina 286.

Teorema.- Sea X_n una variable binomial generalizada definida por n pruebas. Sea X_{n+1} la variable obtenida al añadirle a X_n una nueva prueba de Poisson con probabilidad de "exito" p no nula, entonces

$$H(X_{n+1}) > H(X_n)$$

Demostración: Como la función $-x\ln x$ es cóncava y nula para $x=0$, podemos considerar que X_n toma además los valores -1 y $n+1$ con probabilidades nulas, con lo que resulta que todas las probabilidades elementales de X_{n+1} se expresan como

$$b_g(k;n+1) = p\cdot b_g(k-1;n) + q\cdot b_g(k;n) \quad .$$

Aplicando el lema, tenemos que

$$-b_g(k;n+1)\cdot \ln b_g(k;n+1) > -p\cdot b_g(k-1;n)\cdot \ln b_g(k-1;n) - q\cdot b_g(k;n)\cdot \ln b_g(k;n)$$

de modo que sumando desde 0 hasta n+1 se obtiene la desigualdad buscada, ya que

$$H(X_{n+1}) > -p\sum_{k=0}^{n+1} b_g(k-1;n)\cdot \ln b_g(k-1;n) - q\sum_{k=0}^{n+1} b_g(k;n)\cdot \ln b_g(k;n)$$

$$-(p+q)\sum_{k=o}^{n} b_g(k;n)\cdot \ln b_g(k;n) \geq H(X_n)$$

Observese que cuando la prueba añadida es deterministica la entropía no aumenta, lo cual permite obtener importantes resultados basados en las hipótesis de nuestra generalización del teorema de Jogdeo y Samuels. Su demostración se reduce a la aplicación del primer teorema de esta sección.

Teorema.- Sea $\{X_m\}$ una sucesión de variables binomiales que verifica:

i) La sucesión $\{n_m\}$ es creciente y las sucesiones $\{n_m p_m\}$ y $\{n_m q_m\}$ son no decrecientes.

ii) La diferencia entre los valores esperados de cualquier par de variables de la sucesión siempre es número entero.

Bajo estas condiciones, se cumple que la sucesión de las entropías $H(X_m)$ es estrictamente creciente.

Corolario.- La distribución de Poisson tiene siempre mayor entropía que cualquier distribución binomial con igual promedio.

Teorema.- Sea $B(n_m, d_m, p_m)$ una sucesión de VBM que cumple las siguientes condiciones:

i) La sucesión $\{n_m\}$ es creciente.

ii) La sucesión $\{p_m\}$ es no decreciente, bajo el supuesto de que dicho parámetro se define en el intervalo $[0, \frac{1}{2}]$.

iii) La sucesión $\{\lambda_m\}$ de los valores esperados y la sucesión $\{n_m - \lambda_m\}$ son no decrecientes y todo incremento de las misma es entero.

Entonces la sucesión de sus correspondientes entropías es estrictamente creciente.

Volviendo a nuestro resultado del no decrecimiento de la entropía al añadir una nueva prueba, debemos señalar que, si incluimos las pruebas deterministicas, nuestro resultado sería consecuencia del segundo apartado del ejercicio 1.5, pag. 212, del libro de P.Gil. Ahora bien, a la luz de un reciente articulo de R.Hickey (1983), ambos resultos son consecuencia de un teorema mucho mas fuerte.

Hickey emplea la noción de mayoración para comparar la aleatoriedad de distribuciones discretas. Considerando las probabilidades puntuales de las distribuciones como

componentes de un vector, da la siguiente definición:

> Definición.- Si el vector de masa de probabilidad de una distribución S domina al correspondiente vector de otra distribución T, diremos que T es al menos tan aleatoria como S.

La ventaja de esta definición es que si dos distribuciones son comparables por mayoración siempre lo serán en el mismo sentido respecto a cualquier medida de aleatoriedad. De ahi la importancia del siguiente teorema de Hickey.

> Teorema.- Sean X e Y dos variables discretas independientes. La distribución de $X+Y$ es mas aleatoria que las de X o Y, salvo que alguna de ellas sea degenerada, en cuyo caso es tan aleatoria como la otra distribución.

Como consecuencia de esto, tenemos que la energía informacional siempre disminuye al añadir pruebas de Poisson con probabilidades de "exito" no nulas. Pero antes de seguir con la energía informacional, queremos hacer aun un comentario sobre la entropía. Podría pensarse que, al igual que la entropía de n pruebas se hace máxima cuando todas las probabilidades de "exito" valen $\frac{1}{2}$, que el máximo in

cremento de la misma al añadir una nueva prueba tendría lugar en esas mismas condiciones. Desgraciadamente, esta conjetura no es cierta, como lo prueba el hecho de que una única prueba con p=0.4 presenta el incremento maximo de entropía para un valor de p cercano a 0.5285.

Pasemos a considerar ahora el comportamiento de la energía informacional. Nuestro interes hacia esta medida surgió de manera casual gracias a un error de programación, que dió lugar a que calculasemos la energía informacional de las VBM y nos diesemos cuenta de que su valor no depende del parámetro d. Encontrada la explicación de esta propiedad en la invarianza, respecto a dicho parámetro, de la diferencia media generalizada de orden cero (Vease III.5.d), nos disponemos a extenderla a cualquier distribución binomial de Poisson. Para ello veamos lo que se entiende por variable binomial métrica generalizada.

Definición.- Dada una variable binomial de Poisson X definida por el vector de probabilidad de "exito" $\vec{p}$, se dice que Y es una generalización métrica de X si es una variable binomial de Poisson cuyo vector de probabilidad de "exito" se deduce de $\vec{p}$ sustituyendo cualquier número de componentes por su complementario a 1.

Veamos que todas estas distribuciones tienen la misma energía informacional y demos una expresión sencilla de su valor.

Teorema.- Sea X una variable binomial generalizada definida por el vector de probabilidad $\vec{p}$. Todas las variables métricas que engendra tienen la misma energía informacional, coincidiendo su valor con la probabilidad central de la distribución binomial generalizada definida por el vector de probabilidad $\vec{\pi}=(\vec{p},\vec{q})$, donde $\vec{q}$ representa al vector de las probabilidades de "fracaso" de todas las pruebas. En otras palabras, se verifica que

$$\mathcal{E}(X) = b_g(n/(\vec{p},\vec{q}))$$

Demostración.- Recordando que la diferencia media generalizada de orden cero vale

$$\Delta_o = \sum_{i=1}^{n} \sum_{j=1}^{n} b_g(i) \cdot b_g(j) \cdot |i-j|^o$$

y que se obtiene a partir de la función generatriz de probabilidad por la fórmula

$$\Delta_o = P\left\{G(z) \cdot G(z^{-1})\right\}\Big|_{z=1} \quad ,$$

donde P representa al operador que extrae solo los tér-

minos que tienen en z grado no negativo, es claro que la energía informacional corresponde unicamente a los sumandos de Δ_o en que i=j y, por tanto, se puede deducir a partir de la función generatriz de probabilidad empleando una fórmula analoga a la anterior

$$\mathcal{E}(X) = O\left(G(z)\cdot G(z^{-1})\right)$$

donde el operador O actua extrayendo el término de grado cero en z del operando.

Aplicando esta fórmula general al presente caso, teniendo en cuenta que

$$G(z;\vec{p}) = \sum_{i=1}^{n}(q_i+p_i z) \quad ,$$

resulta que el producto $G(z;\vec{p})\cdot G(z^{-1};\vec{p})$ vale

$$\frac{1}{z^n}\sum_{i=1}^{n}(q_i+p_i z)\cdot(p_i+q_i z)$$

Ahora bien, reescribiendo el último producto como

$$\sum_{j=1}^{2n}(\rho_i + \pi_i z)$$

representa la función generatriz de una variable binomial de Poisson Y compuesta por 2n pruebas y definida por el vector de probabilidad de "exito" $\vec{\pi}=(\vec{p,q})$, de modo que el término independiente de $z^{-n}\cdot G(z,\vec{\pi})$ coinci-

de con la probabilidad del suceso Y=n . Ahora bien, en este punto se localiza la moda de la variable Y, ya que su promedio es ese mismo valor entero.

Consideremos ahora la localización de los extremos absolutos de esta cantidad. Veamos que $\mathcal{E}(X)$ alcanza sus extremos en las mismas distribuciones que la entropía, pero correspondiendose los máximos de una a los mínimos de la otra, y viceversa. En otras palabras, la energía informacional es máxima cuando todas las pruebas son deterministicas, mientras que es mínima cuando los resultados de cada prueba son equiprobables.

La demostración de este resultado se simplifica debido a que los extremos de $\mathcal{E}(X)$ coinciden también con los extremos de la probabilidad central de la variable Y sin imponer ninguna restricción a sus probabilidades de "exito". Por poderse expresar esta probabilidad como el valor esperado de la función indicador de la moda de Y, la busqueda de sus extremos se reduce drasticamente gracias al 2º teorema de Hoeffding, que caracteriza a los vectores de probabilidad de las distribuciones donde es posible se presente un extremo. Recordamos que las componentes de dichos vectores solo pueden tomar tres valores: 0 ó 1 ó el mismo para las restan

tes componentes.

Nuestra demostración de la existencia de un mínimo de $\mathcal{E}(X)$ en las condiciones mencionadas se reducirá entonces a comprobar que $b_n(2n, \frac{1}{2})$ es inferior a la probabilidad central de las variables binomiales de Poisson definidas por los vectores de probabilidad cuyas 2n componentes suman n y verifican la caracterización de Hoeffding. Para realizar esta comprobación necesitamos hacer uso de dos lemas.

Lema.- La sucesión de probabilidades $\{b_n(m, \frac{n}{m})\}$ es decreciente en m, cualquiera que sea n.

Demostración: Veamos que el cociente entre dos términos consecutivos,

$$C(m) = \frac{b_n(m+1, \frac{n}{m+1})}{b_n(m, \frac{n}{m})} = \left(\frac{m}{m+1}\right)^m \cdot \left(\frac{m+1-n}{m-n}\right)^{m-n} ,$$

es siempre menor que la unidad. Para ello consideremos que m varía de modo continuo y probemos que la función C(m), cuyo limite cuando $m \to \infty$ vale 1, es creciente. Su derivada vale

$$\frac{dC}{dm} = C(m) \cdot \left\{ \frac{1}{m+1} + \ln\left(\frac{m}{m+1}\right) - \frac{1}{m+1-n} - \ln\left(\frac{m-n}{m-n+1}\right) \right\}$$

que, teniendo en cuenta el desarrollo, ya empleado, de $\ln(1-x)$, puede expresarse como

$$\frac{dC}{dm} = C(m)\cdot\left\{\frac{1}{m+1} - \sum_{i=1}^{\infty}\frac{1}{i}\cdot\left(\frac{1}{m+1}\right)^i - \frac{1}{m-n+1} + \sum_{i=1}^{\infty}\frac{1}{i}\cdot\left(\frac{1}{m-n+1}\right)^i\right\} =$$

$$= C(m)\cdot\sum_{i=2}^{\infty}\frac{1}{i}\cdot\left\{\frac{1}{(m-n+1)^i} - \frac{1}{(m+1)^i}\right\} \quad .$$

Claramente se ve que su valor siempre es positivo.

Lema.- La sucesión $\{b_n(2n, \frac{1}{2})\}$ es decreciente.

En este caso la demostración es inmediata ya que el cociente entre dos términos consecutivos vale

$$\frac{b_{n+1}(2n+2, \frac{1}{2})}{b_n(2n, \frac{1}{2})} = \frac{2n+1}{2n+2} \quad .$$

Pasemos a enunciar el teorema sobre los extremos absolutos de la energía informacional.

Teorema.- El máximo y el mínimo absolutos de la energia informacional de la distribución de probabilidad de n pruebas de Poisson son, respectivamente, la unidad y $\binom{2n}{n}.2^{-2n}$. El primero se alcanza cuando la probabilidad de "exito" de cada prueba vale indistintamente 0 ó 1. El segundo solamente cuando todas estas valen $\frac{1}{2}$.

Demostración.- En el caso del máximo resulta trivial, ya que el mayor valor posible de la energía informacional es la unidad y se alcanza sobre distribuciones degeneradas.

En el caso del mínimo, como ya hemos dicho, la demostración consiste en comprobar que el término central $b_g(n/\vec{v})$ de la distribución del número de "exitos" de 2n pruebas de Poisson, cuando el promedio de "exitos" es n, tiene un mínimo igual a $b_n(2n, \frac{1}{2})$. Para ello veamos el valor que tiene $b_g(n/\vec{v})$ sobre todas las distribuciones en que se pueden localizar sus extremos. Tales distribuciones estan definidas por los vectores de probabilidad de la siguiente forma:

$$\vec{v}(r,m,s) = (\overbrace{0,\ldots,0}^{r}, \overbrace{\tfrac{n-s}{m},\ldots,\tfrac{n-s}{m}}^{m}, \overbrace{1,\ldots,1}^{s})$$

Respecto a estas distribuciones, el término central queda reducido a una probabilidad binomial cuyo valor es

$$b_g(n/\vec{v}(r,m,s)) = b_{n-s}(m, \tfrac{n-s}{m})$$

y además se verifica que

$$b_g(n/\vec{v}(r,m,s)) = b_g(n/\vec{v}(s,m,r))$$

con lo que es suficiente considerar su valor sobre aque

llas distribuciones en que $s \leqslant r$, con $0 \leqslant s \leqslant n$.

Para mostrar que $b(n;2n,\frac{1}{2})$ es el mínimo de $b_g(n)$, vamos a hallar primero su menor valor sobre las distribuciones cuyo vector de probabilidad tiene el mismo número de componentes unidad y luego consideraremos el mínimo de dichos valores.

Para s fijo, $b_{n-s}(m,\frac{n-s}{m})$ toma su menor valor posible, aplicando el primer lema, cuando m sea lo mayor posible. Luego debe verificarse que r sea igual a s, correspondiendole un valor mínimo de

$$b_{n-s}(2n-2s,\tfrac{1}{2}) \quad .$$

Al variar s obtenemos la sucesión del segundo lema, presentandose por tanto el mínimo absoluto cuando $s=0$.

Pasemos a considerar ahora el comportamiento de la energía informacional de las distribuciones de $\mathcal{P}_\mu$ que están ordenadas por mayoración. A diferencia de lo que sucede con la entropía, esta cantidad no se comporta globalmente de la misma manera. Así, por ejemplo, dos pruebas de Poisson con promedio $\mu=0.02$ maximizan su valor cuando son equiprobables, en lugar de minimizarla que sería lo esperable.

Seguidamente vamos a mostrar que al menos localmente se puede garantizar que la energía informacional sigue el comportamiento esperado. Recordando que $p_{[1}$ y $p_{[n}$ denotan, respectivamente, la mayor y la menor componente de de un vector de probabilidades de "exito", enunciemos nuestro resultado y un corolario del mismo.

Teorema. Sea $\mathcal{S}_\mu$ el subconjunto de $\mathcal{P}_\mu$ cuyos elementos verifican que $p_{[1} < \frac{2}{3}$ y $p_{[n} > \frac{1}{3}$. Entonces si $\vec{p}^{(1)} \in \mathcal{S}_\mu$ y si $\vec{p}^{(1)} \underset{\leq}{\prec} \vec{p}^{(2)}$, se verifica que $\mathcal{E}(\vec{p}^{(1)}) \geq \mathcal{E}(\vec{p}^{(2)})$.

Corolario.- Sobre $\mathcal{S}_\mu$ la energía informacional presenta un mínimo en la distribución binomial $B(n,\frac{\mu}{n})$.

Demostración del teorema: Consistirá en probar que $\mathcal{E}(X)$ es una función Schur-convexa sobre $\mathcal{S}_\mu$. Esto está asegurado si se verifica que

$$(p_i - p_j)\left(\frac{\partial \mathcal{E}}{\partial p_i} - \frac{\partial \mathcal{E}}{\partial p_j}\right) \geq 0$$

cualesquiera que sean p_i y p_j de los vectores de $\mathcal{S}_\mu$.

Para ello partamos de la expresión que adopta $b_g(k)$ cuando del vector de probabilidad separamos las componentes p_i y p_j :

$$b_g(k) = b_g(k;\vec{p}^{ij}) + (p_i+p_j)\cdot\left\{b_g(k-1;\vec{p}^{ij})-b_g(k;\vec{p}^{ij})\right\} +$$

$$+ p_ip_j\cdot\left\{b_g(k;\vec{p}^{ij})-2b_g(k-1;\vec{p}^{ij})+b_g(k-2;\vec{p}^{ij})\right\} ,$$

que por comodidad notaremos como

$$\beta_k + \gamma_k(p_i+p_j) + \delta_k p_i p_j \quad .$$

Sustituyendola en la definición de $\mathcal{E}(X)$ y derivando, por ejemplo, respecto a p_i obtenemos

$$\frac{\partial \mathcal{E}}{\partial p_i} = 2\sum_{k=o}^{n} b_g(k)\left\{\gamma_k + \beta_k p_j\right\}$$

de modo que

$$(p_i-p_j)\left\{\frac{\partial \mathcal{E}}{\partial p_i} - \frac{\partial \mathcal{E}}{\partial p_j}\right\} = -(p_i-p_j)^2 \sum_{k=o}^{n} b_g(k)\cdot\delta_k$$

donde la sumatoria, interpretando sus productos como probabilidades de una cierta variable binomial de Poisson, puede comprobarse que vale

$$\Delta^2 b_g(n-2/(p,\overline{q}^{ij}))$$

Ahora bien, como las componentes de $(p,\overline{q}^{ij})$ tienen una suma

$$\eta = np + (n-2-np+p_i+p_j) = n-2+p_i+p_j \quad ,$$

resulta que su distribución asociada siempre localiza la moda en n-1, con lo que $\Delta^2 b_g(n-2)$ es negativa.(Vease la

demostración del último teorema del punto III.8.c).

Como consecuencia inmediata de estos dos teoremas, teniendo en cuenta que Δ_0 vale $\frac{1}{2}(1+\delta)$, obtenemos los siguientes corolarios relativos a esta medida de dispersión.

Corolario.- El máximo y el mínimo absolutos de la diferencia medias de orden cero de n pruebas de Poisson se alcanza, respectivamente, cuando todas las pruebas son determinísticas o bien cuando sus resultados son equiprobables. Todas las distribuciones métricas que engendra una variable binomial generalizada tienen el mismo valor de Δ_0.

Corolario.- Sean X_1 y X_2 dos variables binomiales de Poisson cuyos vectores de probabilidad verifican $\vec{p}^{(1)} \geq \vec{p}^{(2)}$. Entonces, si $p_1^{(1)} < \frac{2}{3}$ y $p_n^{(1)} > \frac{1}{3}$, se verificará que

$$\Delta_0(X_1) < \Delta_0(X_2) \quad .$$

Seguidamente vamos a demostrar que las diferencias medias de orden superior se comportan respecto a la relación de mayoración de la misma forma que los momentos absolutos de orden superior a la unidad. El fun-

damento de esta propiedad reside en la identidad entre las diferencias medias de la variable binomial de Poisson definida por el vector $\vec{p}$ y los momentos centrales absolutos de la variable engendrada por $\vec{\pi}=(\vec{p},\vec{q})$. Veamos primero un resultado auxiliar.

Lema.- Si $\vec{p}^{(1)} \succeq \vec{p}^{(2)}$, entonces $\vec{\pi}^{(1)} \succeq \vec{\pi}^{(2)}$.

Demostración: Por hipotesis existe una matriz Φ doblemente estocastica que verifica

$$\vec{p}^{(2)} = \vec{p}^{(1)}\Phi \quad .$$

Entonces la matriz Ψ definida como

$$\Psi = \begin{pmatrix} \Phi & 0 \\ 0 & \Phi \end{pmatrix}$$

es doblemente estocastica y transforma $\vec{\pi}^{(1)}$ en $\vec{\pi}^{(2)}$.

Teorema.- Sean X_1 y X_2 dos variables binomiales generalizadas. Si sus correspondientes vectores de probabilidad verifican que $p^{(1)} \succeq p^{(2)}$, entonces se cumple que

$$\Delta_r(X_1) \geq \Delta_r(X_2) \quad , \quad \forall r \geq 1 \quad .$$

Demostración.- Vamos a cambiar el procedimiento de obtención de las diferencias medias generalizadas para que re

sulten directamente, en vez de por intermedio de las diferencias medias factoriales. Si $\Phi(t)$ es la función generatriz de momentos de una variable discreta, puede comprobarse que Δ_r puede calcularse como

$$\Delta_r = P\frac{d^r}{dt^r}\left\{\Phi(t)\cdot\Phi(-t)\right\}\Big|_{t=o}$$

donde el operador P actua dejando solo los términos con potencias no negativas de e^t.

Teniendo en cuenta que la función generatriz de momentos de una variable binomial de Poisson vale

$$\Phi(t;\vec{p}) = \prod_{i=1}^{n}(q_i + p_i e^t) \quad ,$$

resulta que el producto $\Phi(t)\cdot\Phi(-t)$ puede interpretarse como la función generatriz de momentos centrales de la variable definida por el vector $\vec{\pi}=(\vec{p},\vec{q})$, ya que

$$\Phi(t;\vec{p})\cdot\Phi(-t;\vec{p}) = \prod_{i=1}^{n}(q_i + p_i e^t)\cdot\frac{(q_i e^t + p_i)}{e^t} =$$

$$= e^{-nt}\cdot\prod_{j=1}^{2n}(\rho_j + \pi_j e^t) = \sum_{k=o}^{2n} b_g(k)e^{k-n\,t} \quad .$$

De aquí que la fórmula de Δ_r sea equivalente a

$$\Delta_r = \sum_{k=n}^{2n}(k-n)^r b_g(k;\vec{\pi}) \quad .$$

Ahora bien, como las probabilidades $b_g(k;\vec{\pi})$ son simétricas respecto de n, puesto que el vector de probabilidades de "fracaso" coincide, reordenando sus componentes, con el de las probabilidades de "exito", resulta que Δ_r es igual a la mitad del momento absoluto de orden r de la variable definida por $\vec{\pi}$. Por tanto, teniendo en cuenta el lema previo y el corolario de Gleser queda demostrado el teorema.

Finalmente vamos a hacer un comentario sobre algunas cuestiones que quedan abiertas. Por ejemplo, podría estudiarse el comportamiento de la familia de medidas de aleatoriedad definidas por Renyi:

$$\gamma_\alpha(X_g) = \frac{1}{1-\alpha} \ln(\Sigma b_g^\alpha(k))$$

(Observese que para $\alpha=2$ se obtiene el logaritmo del inverso de $\delta(X_g)$.) En este estudio puede jugar un papel muy importante las ideas de Hickey sobre la mayoración de las distribuciones. En nuestro caso, sería muy interesante establecer bajo qué condiciones se deduciría la mayoración entre variables binomiales de Poisson cuando sus correspondientes vectores de probabilidades de "exito" son comparables. En este sentido, nosotros hemos obtenido un resultado muy particular. Para su demostración se requiere el último corolario del pun

to III.8.c . Nuestro enunciado tiene el siguiente enunciado:

Teorema.- Si se verifica que $\vec{p}^{(1)} \geq \vec{p}^{(2)}$, $p_{\boxed{1}}^{(1)} < \frac{2}{3}$ y $p_{\boxed{n}}^{(1)} > \frac{1}{3}$, entonces la distribución engendrada por el vector $\vec{\pi}^{(1)}=(\vec{p}^{(1)},\vec{q}^{(1)})$ domina a la engendrada por $\vec{\pi}^{(2)}$.

Demostración.- Por construcción las distribuciones son simétricas. Como a partir de la moda las probabilidades puntuales decrecen monotonamente, es suficiente para demostrar la dominación comparar unicamente las sumas parciales de la forma

$$\sum_{k=n}^{j} b_g(k) \qquad \text{con} \quad n < j \leq 2n \ .$$

Ahora bien, aplicando el corolario mencionado, resulta que dichas sumas en la primera distribución son siempre mayores o iguales que en la segunda, con lo que queda probada la mayoración.

III.10.- ESTIMACION DE LOS PARAMETROS

En esta sección vamos a tratar unicamente de la estimación individual de cada uno de los parámetros de la VBM. La estimación conjunta de cualquier combinación de los mismos será objeto de un trabajo posterior, aunque de nuestra investigación previa se desprende que, en general, los estimadores posibles son bastante imprecisos.

En cualquiera de los tres casos posibles, supondremos que partimos de una muestra aleatoria simple de tamaño N, $x_1, x_2, \ldots, x_N$, procedente de una población $B(n,d,p)$ con el correspondiente parámetro desconocido. Notaremos por S la suma de las observaciones muestrales y por $\bar{x}$ la media muestral.

a) Estimación de p.

En este caso, con objeto de que la estimación de p resulte lo mas precisa posible, nos vemos obligados a cambiar la definición del espacio paramétrico de la VBM que hemos mantenido en las paginas anteriores; consideraremos que p varía en el intervalo unidad, mientras que d toma valores comprendidos entre 0 y $\frac{n}{2}$; en el caso especial de que

d tome este último valor, tendremos que limitar a p al intervalo $[0,\frac{1}{2}]$, ya que valores complementarios del mismo definen la misma distribución. Volveremos a considerar esta determinación del espacio paramétrico en la estimación de n , ya que ésta resulta tanto mas precisa cuanto mayor sea el valor de p. La adopción de la otra determinación paramétrica, en el caso particular de la binomial de Bernoulli, supone la estimación del mínimo de p y q. Su estimador natural,

$$\min\left(\frac{S}{n\cdot N}, 1-\frac{S}{n\cdot N}\right) \quad ,$$

ha sido tratado por R. Greenwood y M. Glasgow (1950) y M. Sandeluis (1952); los primeros estudiaron sus momentos; el segundo estudió su función de distribución, construyendo un intervalo de confianza para $\min(p,q)$.

El procedimiento mas sencillo de estimación de p es el método de los momentos. Hay que distinguir dos casos. Si el parámetro d es distinto de $\frac{n}{2}$, tenemos

$$\tilde{p} = \frac{\bar{x} - d}{n - 2\cdot d} \quad .$$

Si d vale $\frac{n}{2}$, como el valor esperado de la BM es independiente de p, hay que recurrir al siguiente momento, obteniendose

$$\tilde{\tilde{p}}\cdot\tilde{\tilde{q}} = \frac{1}{nN}\sum_{i=1}^{N}\left(x_i - \frac{n}{2}\right)^2$$

Consideremos el primer estimador. En principio, cabría esperar que $\tilde{p}$ presentase buenas propiedades, dado que al particularizarlo a $d=0$ se obtiene un estimador eficiente. A primera vista, parece ser un estimador insesgado con una varianza igual a

$$V(\tilde{p}) = \frac{n \cdot p \cdot q}{N(n-2d)}$$

Sin embargo, al fijarnos mas atentamente, se observa que estos resultados no son ciertos cuando d es distinto de 0, debido a que $\tilde{p}$ toma valores fuera del intervalo paramétrico. Así, si $\bar{x} < d$, toma valores negativos, mientras que toma valores superiores a 1 si $\bar{x} > n-d$. Basandose en que las probabilidades de las colas de las VBM son inferiores a las de la binomial de igual promedio y numero de pruebas, y en las cotas asociadas con el teorema de Bernoulli (Vease I.13 .a), puede juzgarse en que medida los valores inadmisibles afectan al estimador. Evidentemente, su influencia será tanto mayor cuanto mas proximo esté d a $\frac{n}{2}$.

Otra propiedad que se pierde al pasar de la binomial a la binomial métrica es la suficiencia de S respecto al parámetro p. En efecto, la distribución de la muestra condicionada a S depende de p; por ejemplo, cuando S vale $N \cdot d$ la probabilidad del suceso $\{x_1=d, x_2=d, \ldots, x_N=d/S=Nd\}$ es i-

gual o distinta de 1, según p sea igual o distinto de 0. La cantidad de información perdida al emplear S aumenta al acercarse el parámetro d a $\frac{n}{2}$.

Naturalmente, a pesar de estos inconvenientes, dada su sencillez, $\tilde{p}$ resulta un estimador bastante útil y cómodo; no obstante, cuando las observaciones muestrales conduzcan a estimaciones inadmisibles, será preciso emplear otros estimadores. A pesar de los valores inadmisibles, este estimador conserva las propiedades de ser consistente y asintoticamente normal, como se desprende de que S siga una distribución B(n·N,d·N,p). De un estudio empírico que hicimos sobre la eficiencia relativa de $\tilde{p}$ respecto a la cota de Frechet-Cramer-Rao, hemos obtenido una consecuencia teórica importante:

Teorema.- El parámetro p no admite ningún estimador insesgado si el parámetro d es mayor que $\frac{n}{2} - \sqrt{n}$.

Demostración: Vamos a hacerla por reducción al absurdo, viendo que si existiese un estimador insesgado, su varianza sería inferior a la que predice la cota. Dicha cota siempre existe, puesto que la VBM verifica las condiciones de regularidad de Cramer:

1) Los valores de las observaciones muestrales no dependen de p.
2) p está definido en un intervalo finito.
3) Las observaciones muestrales son también finitas.
4) La derivada $\frac{d}{dp} b(k;n,d,p)$ siempre está acotada.

Numericamente puede comprobarse con facilidad que, bajo las hipótesis del teorema, la cota supera al producto $p \cdot q$, máxima varianza posible de un estimador insesgado de un parámetro definido sobre el intervalo unidad. Vamos a mostrar esto mismo analiticamente para un entorno de $p=\frac{1}{2}$, basandonos en la siguiente forma de expresar el denominador de la cota

$$E\left[\frac{d \ln b_k}{dp}^2\right] =$$

$$= \frac{1}{p \cdot q}\left\{ n + 4(n-d)\left\{\frac{(n-d)\cdot p}{q}\{\gamma_o(n,d) - 1\} - 1\right\}\right\}$$

donde $\gamma_o(n,d)$ representa a

$$\sum_{k=1}^{n} \frac{b^2(k-1;n-1,d)}{b(k;n,d)} = E\left[\frac{b_k}{qb_{k+1} + pb_k};n-1,d\right] ,$$

fórmula que se obtiene al expresar las probabilidades puntuales de B(n,d) en función de las correspondientes a B(n-1,d), pasando por el resultado intermedio

$$\frac{db_k}{dp} = \frac{1}{p.q}\left\{(d-k-p.n)b_k + 2(n-d)b_{k-1}(n-1,d)\right\} \quad .$$

(Fórmulas semejantes se obtienen intercambiando los papeles de p y d con los de q y n-d, apareciendo entonces la variable B(n-1,d-1).)

Teniendo en cuenta que para d=0 la cota vale $\frac{pq}{n}$, deducimos que

$$\gamma_o(n,0) = \frac{pn + q}{pn} \quad ,$$

de manera que al considerar el caso particular en que $p=\frac{1}{2}$, como las VBM no dependen del parámetro d, resulta que $\gamma_o(n,d)$ tiene el mismo valor, de modo que

$$\text{Cota}(n,d,\tfrac{1}{2}) = \frac{n}{4(n-2d)} \quad ;$$

cantidad que coincide con la varianza del estimador $\tilde{p}$ no corregido y que en las condiciones del teorema supera a $\frac{1}{4}$.

Nosotros creemos que la inexistencia de un estimador insesgado debe verificarse también para cualquier distribución cuyo valor de d sea distinto de 0, aunque por el momento no hemos encontrado ningún procedimiento para demostrarlo.

Consideremos ahora el caso especial en que n sea par y d valga $\frac{n}{2}$. En dicho caso, el denominador de la cota para $p=\frac{1}{2}$ se anula. Analicemos detalladamente lo que ocurre en esta situación. La explicación se encuentra en la siguiente propiedad:

Lema.- $b_k(n, \frac{n}{2})$ es siempre función del producto $\theta=p\cdot q$.

La demostración es inmediata, ya que la función generatriz de probabilidad de $B(n, \frac{n}{2})$ puede expresarse como $(\theta + (1-2\theta)z + \theta z^2)^n$, siendo el trinomio la función generatriz de B(2,1).

Teniendo en cuenta esta propiedad, resulta que

$$\frac{d\ln b_k}{dp} = \frac{d\theta}{dp}\cdot\frac{d\ln b_k}{d\theta} = (q-p)\cdot\frac{d\ln b_k}{d\theta} \quad ,$$

de modo que el denominador de la cota se anula por la presencia del factor $(q - p)$. No obstante, a pesar de ser nulo este denominador, existe la cota ya que simultaneamente la derivada del valor esperado de cualquier estimador, por ser esta también función de θ , admite este mismo factor. Simplificando, resulta

$$\text{cota}(n, \frac{n}{2}, \theta) = \frac{\left\{\frac{d}{d\theta}E[\tilde{p}]\right\}^2}{N\cdot E[\{\frac{d}{d\theta}\ln b_k\}^2]} \quad ,$$

expresión que siempre toma valores finitos ya que $\frac{db_o}{d\theta}$ no se anula para $p \neq 0$.

En este caso especial, la función de verosimilitud también presenta algunas peculiaridades:

1) Es simétrica respecto a $p=\frac{1}{2}$, de modo que si alcanza su maximo en p_o, también lo alcanzará en q_o.
2) $p=\frac{1}{2}$ siempre corresponde a un extremo de la verosimilitud muestral, de modo que si no existiese máximo respecto a θ, $\frac{1}{2}$ será el estimador de maxima verosimilitud.

Debido a que los papeles de p y q son indistinguibles, pensamos que cualquier proceso de estimación de este parámetro debe dar simultaneamente ambos valores. Por esta razón es necesario restringir su valor o bien al intervalo $[0,\frac{1}{2}]$ o bien al intervalo $[\frac{1}{2},1]$; es decir, elegir el mínimo o el máximo de las estimaciones. Nosotros consideraremos el primero, porque es coherente con nuestra determinación anterior del espacio paramétrico. En estas condiciones pasemos a estudiar algunas propiedades del estimador que se deduce del método de los momentos:

$$\tilde{\tilde{p}} = \frac{1}{2} - \sqrt{\frac{1}{4} - \frac{1}{nN}\Sigma(x_i - \frac{n}{2})^2}$$

Este estimador, tomado al pie de la letra, también da valores inadmisibles; éstos se presentan cuando la desviación cuadrática muestral supera a $\frac{n}{4}$. Ahora bien, interpretando el significado de este hecho, es razonable asignarle como estimación el valor $\frac{1}{2}$.

Para n=2 el estimador, así corregido, coincide con el estimador de máxima verosimilitud y además es función de un estadistico suficiente. Si llamamos f_i a la frecuencia absoluta de la observación i, la verosimilitud muestral vale

$$L(x_1,x_2,\ldots,x_N) = \theta^{f_0}(1-2\theta)^{f_1}\theta^{f_2} = \theta^{\Sigma(x_i-1)^2}(1-2\theta)^{N-\Sigma(x_i-1)^2},$$

es decir, queda factoriza al considerar el error cuadrático muestral.

En este caso particular, es facil estudiar la distribución exacta en el muestreo de dicho estadistico para tamaños muestrales elevados. Concretamente, nosotros hemos tabulado su valor esperado, sesgo, varianza y la cota de Cramer-Rao para valores de N comprendidos entre 1 y 10.000. Estos resultados sirven de punto de referencia de lo que debe suceder para otros valores de n cuando el tamaño muestral

es inversamente proporcional.(Vease apendice II)

Por otro lado, para valores grandes de n, este estimador coincide con el de máxima verosimilitua que se deduce al considerar la aproximación normal

$$\hat{b_k}(n,\frac{n}{2}) = \frac{1}{2\sqrt{npq}} \cdot \exp\left(-\frac{(k-\frac{n}{2})^2}{2npq}\right)$$

Además, en estas condiciones, $\Sigma(x_i-\frac{n}{2})^2$ también es un estadistico suficiente de p.

Para otros valores del parámetro d podemos generalizar este estimador de dos formas distintas. En primer lugar, considerar el que se desprende del método de los momentos al igualar la varianza poblacional a la variaza o la cuasivarianza muestral, eligiendo la solución que esté mas acorde con la media muestral; evidentemente si las soluciones son complejas tomaremos como estimación a $\frac{1}{2}$.

La otra posibilidad es considerar el estimador de máxima verosimilitud que se obtiene al aproximar las probabilidades puntuales por la densidad de probabilidad de la distribución Normal $N(d+(n-2d)p;npq)$. El estimador resultante es una de las raices de la siguiente ecuación cúbica:

$$2Nn\hat{p}^3 + \left\{N(n-2d)^2 - 3Nn - 2(n-2d)\,\Sigma(x_i-d)\right\}\hat{p}^2 + \left\{Nn + 2\Sigma(x_i-d)^2\right\}\hat{p} - \Sigma(x_i-d)^2 = 0$$

Este estimador utiliza conjuntamente la información contenida en la media y varianza muestral, las cuales son un estadistico minimal suficiente del parámetro p respecto a la normal aproximante. Otra propiedad de este estimador es su consistencia cuando N tiende a infinito, que se comprueba viendo que la ecuación se verifica para $\hat{p}=p$ cuando se sustituyen las expresiones muestrales por las esperanzas a que convergen.

Hemos estudiado por simulación ambos estimadores, junto con el estimador del método de los momentos, comparando sus valores medios y error cuadrático. Los resultados se recogen en el apendice III, de los que se puede concluir que el estimador obtenido al igualar la varianza es, en general, peor que los otros dos. El estimador basado en la aproximación normal suele ser superior al de los momentos cuando el parámetro d toma valores próximos a $\frac{n}{2}$.

Finalmente, queremos señalar que no hemos considerado el estimador de máxima verosimilitud basado en todas las observaciones o bien en S por desconocer su forma explicita y por ser muy lenta su determinación numérica.

b) Estimación de d.

Para la estimación de este parametro volvemos a considerar la definición inicial del espacio paramétrico de las VBM, con la salvedad de que debemos excluir el caso en que p valga $\frac{1}{2}$, debido a que las distribuciones resultantes no dependen del valor de d.

Por tratarse de un parámetro discreto es relativamente sencillo determinar numericamente el estimador de máxima verosimilitud, del que teoricamente sabemos que debe ser consistente y asintoticamente normal; de su estudio por simulación, se desprende que presenta un sesgo pequeño y que no es necesario considerar la verosimilitud respecto a todos los posibles valores de d, puesto que es practicamente seguro que su máximo se alcanza para uno de los enteros adyacentes al estimador del método de los momentos.

Dicho estimador viene definido como

$$\tilde{d} = \frac{\bar{x} - np}{(q - p)} \quad .$$

Al margen de que raramente tomará valores enteros, este estimador, al igual que el de $\tilde{p}$, toma valores inadmisibles: negativos, cuando $\bar{x} < np$, y superiores a n, cuando $\bar{x} > nq$.

Afortunadamente, esas estimaciones se corrigen de forma natural asignandoles, respectivamente, los valores cero y n. Incluso admitiendo estas deficiencias, $\tilde{d}$ es un estimador asintoticamente normal y consistente, con una varianza dada por

$$V(\tilde{d}) = \frac{npq}{N(q-p)^2} \quad ,$$

cuyo valor crece rapidamente al aproximarse p a q, lo cual significa que en esas condiciones las VBM dependen muy poco del parámetro d. En este sentido, puede decirse que para que $\tilde{d}$ tenga una varianza inferior a la unidad son suficientes tamaños muestrales inferiores a n si $p<\frac{1}{4}$, mientras que se necesitan tamaños superiores a $10 \cdot n$ si $p>0.42$.

Para obtener valores enteros de $\tilde{d}$ hemos considerado tres criterios de elección sobre los enteros adyacentes:

i) El entero mas proximo.

ii) El que maximice la verosimilitud muestral.

iii) El que maximice la verosimilitud basada unicamente en S.

Hemos comparado por simulación estos estimadores con el de máxima verosimilitud, empleando como criterio de comparación el porcentaje de dominación; es decir, el porcen

taje de veces que esa estimación está tan cerca o mas cerca del verdadero valor del parámetro que la estimación de máxima verosimilitud. Los resultados se encuentran recogidos en el apendice IV y de ellos se desprende que el estimador obtenido por el segundo criterio tiene unos porcentajes de dominación muy elevados, que frecuentemente llegan hasta el 100%. Los otros estimadores son muy parecidos. En el estudio numerico es suficiente considerar que el verdadero valor del parámetro está comprendido entre 0 y $\frac{n}{2}$, ya que, teniendo en cuenta la equivalencia entre los sucesos de las variables $B(n,n-d,p)$ y $B(n,d,p)$, toda muestra $x_1,x_2,\ldots,x_N$ de la primera distribución está en correspondencia univoca con la muestra $n-x_1,n-x_2,\ldots,n-x_N$ perteneciente a la segunda distribución.

Finalmente, si para valores grandes de n consideramos la aproximación de la VBM por la distribución normal, resulta que la densidad normal obtenida pertenece a la familia exponencial regular respecto al parámetro d, de modo que admite un estimador insesgado y de varianza uniformemente mínima. Dicho estimador es el de máxima verosimilitud y coincide con el estimador del método de los momentos correspondiente a la verdadera distribución.

c) Estimación de n.

Para este parámetro se pueden extender los estimadores utilizados en la binomial de Bernoulli cuando el parámetro p es conocido. Dichos estimadores han sido estudiados por Feldman y Fox (1968) y Blumenthal y Dahiya (1981); la mayoría de las propiedades que en esos artículos se recogen resultan aplicables también en nuestro caso. Al igual que para los otros dos parámetros, se consideran dos tipos de estimadores: los obtenidos directamente de la distribución exacta y los obtenidos por intermedio de la aproximación normal. Todos ellos tienen la propiedad de ser consistentes y asintoticamente normales respecto al tamaño muestral; a excepción del estimador de maxima verosimilitud, todos los otros dan lugar a estimaciones inadmisibles, puesto que n no puede tomar nunca valores inferiores al parámetro d ni inferiores a M, siendo $M=\max(x_1,x_2,\ldots,x_N)$. Evidentemente, cuando la muestra conduzca a estimaciones inadmisibles, la corrección natural será tomar $n^{\wedge}=\max(d,M)$.

Consideremos en primer lugar los estimadores exactos. El estimador de máxima verosimilitud puede obtenerse numericamente entre los enteros comprendidos entre $n^{\wedge}$ y $n^{\star}$, definido este último valor como $n^{\star}=\max(n^{\wedge}, \frac{M-d(q-p)}{p}+1)$,

puesto que está asegurado el decrecimiento de la verosimilitud muestral a partir de $n^{\star}$, como consecuencia de la unimodalidad de ls DBM y de las relaciones entre las probabilidades puntuales de B(n,d,p) y B(n+1,d,p). Sin embargo, el proceso de determinación podría abreviarse si la verosimilitud muestral de la DBM fuese además unimodal como lo es la verosimilitud de la binomial clasica.

El estimador del método de los momentos, ignorando el carácter entero de n, viene dado por

$$\tilde{n} = \frac{\bar{x} - d(q-p)}{p}$$

Aceptando también las estimaciones inadmisibles, resulta un estimador insesgado con una varianza igual a

$$V(\tilde{n}) = \frac{n.q}{N.p} \quad ,$$

cuyo valor crece ilimitadamente cuando p tiende a 0, lo cual puede atribuirse a que la muestra no aporta practicamente ninguna información sobre el parámetro. En particular, cuando p vale 0, las observaciones muestrales se reducen a d y cualquier valor mayor o igual que d resulta una estimación igualmente válida del parámetro.

Para redondear $\tilde{n}$ a entero podrían considerarse

los mismos procedimientos que empleamos para el parámetro d. Nosotros, en esta ocasión, solo hemos utilizado los dos primeros, midiendo unicamente el porcentaje de dominación correspondiente al redondeo simétrico.

Aparte de estos dos estimadores podriamos haber considerado el estimador que se obtiene al igualar la varianza de la distribución con la cuasivarianza muestral, pero no lo hemos considerado porque, a partir de $n=3$, tiene siempre mayor varianza que $\tilde{n}$, como puede comprobarse al considerar los valores esperados de los momentos muestrales (Vease, por ejemplo, Kendall y Stuart, vol. 1). Sim embargo, este estimador puede ser cuando el parámetro d sea desconocido, ya que tiene la ventaja de ser independiente del valor del mismo.

Pasemos a considerar los estimadores obtenidos a partir de la aproximación normal. Veremos el estimador insesgado de mínima varianza y el estimador de máxima verosimilitud, en el supuesto de que n fuese un parámetro continuo. El estimador del método de los momentos evidentemente coincide con $\tilde{n}$. Dichos estimadores fueron obtenidos para la binomial clásica por Feldman y Fox, para lo cual hubo que dividir las observaciones muestrales por q, consiguiendose a

sí que la media y la varianza de la distribución fuesen iguales. Esta operación se puede extender facilmente a la VBM, sin mas que restarle $d(q-p)$ a cada observación muestral. Notemos por $y_1, y_2, \ldots, y_N$ los valores deducidos de la muestra por la transformación

$$Y = \frac{X - d(q-p)}{q} \quad .$$

Bajo la aproximación, estos valores siguen una distribución $N(\eta, \eta)$, con $\eta = \frac{np}{q}$. En estas condiciones, $\bar{y}$ es un estimador insesgado de η y $T = \Sigma y_i^2$ es un estadístico suficiente y completo del mismo, de modo que, por el teorema de Blackwell-Rao, $E[\bar{y}|T]$ es su estimador insesgado de mínima varianza; multiplicando por $\frac{q}{p}$ obtenemos el estimador de n, que tiene la siguiente expresión

$$\check{n} = \frac{q}{p} \cdot \sqrt{\frac{T}{N}} \cdot \frac{I_{N/2}(\sqrt{NT})}{I_{\frac{N}{2}-1}(\sqrt{NT})}$$

donde I representa la función modificada de Bessel de primer orden; es decir

$$I_\alpha(z) = \frac{\left(\frac{z}{2}\right)^\alpha}{\Gamma(\alpha+\frac{1}{2})\Gamma(\frac{1}{2})} \cdot \int_{-1}^{1} (1-x^2)^{\alpha-\frac{1}{2}} \cdot e^{z \cdot x} \, dx \quad .$$

Teniendo en cuenta la cota de Cramer-Rao, se comprueba que la varianza de n es como mucho

$$\frac{2np}{2np + q}$$

veces menor que la varianza de $\tilde{n}$, de modo que cuando se sepa que n es grande es preferible utilizar el último estimador, puesto que es un estadistico mucho mas simple.

Por otra parte, el estimador de máxima verosimilitud viene dado por

$$\hat{n} = \frac{q}{p}\sqrt{\frac{T}{N} + \frac{1}{4}} - \frac{q}{2p} \quad .$$

Feldman y Fox demuestran que cuando N tiende a infinito, $\hat{n}$ converge uniformemente hacia $\check{n}$ al variar T. Blumenthal y Dahiya describen sus leyes límite cuando n permanece fijo o tiende a infinito; asimismo, lo comparan por simulación con los estimadores exactos, entre los que incluyen al estimador que minimiza al siguiente estadístico tipo χ^2

$$\Sigma(x_i - np)^2/(npq) \quad ,$$

cuyo valor generalizado a la VBM es $\hat{\tilde{n}} = \frac{q}{p}\sqrt{\frac{T}{N}}$.

Por último, nosotros hemos comparado por simulación el estimador exacto de máxima verosimilitud con $\tilde{n}$, en sus dos versiones redondeadas, y $\check{n}$, con redondeo simétrico. Los resultados se hallan recogidos en el apéndice V y de ellos se concluye, al igual que Blumenthal y Dahiya, que el estimador mas preferible es el de máxima verosimilitud.

III.11.- APROXIMACION DE LAS PROBABILIDADES BINOMIALES METRICAS

En esta sección vamos a considerar la extensión a la VBM de algunos de los procedimientos de aproximación de las probabilidades binomiales. Concretamente, de los métodos recopilados en la sección I.11, solamente consideraremos los basados en las distribuciones Normal y de Poisson. De entre los primeros nos reduciremos unicamente a los siguientes: Aproximación de Moivre-Laplace (A.1); aproximación de Camp-Paulson (A.4) y las dos versiones de la aproximación de Peizer-Pratt (A.6). De entre los segundos nos limitaremos a la aproximación de Poisson (B.1) y a la aproximación de Gram-Charlier (B.2).(A este último método también nos hemos referido (sección I.12) con el nombre de modificación de Kolmogorov.

Nuestro objetivo va a ser describir el comportamiento empírico de las aproximaciones obtenidas, así como dar argumentos que puedan explicarlo en cierta manera. Solamente para algunas afirmaciones relativas a la aproximación de Poisson hemos obtenido demostraciones matemáticas. De todas formas, es nuestro proposito demostrar las restantes afirmaciones, si son ciertas, cuando encontremos las herramientas

adecuadas.

Como medida de la precisión que cada procedimiento alcanza emplearemos, al igual que Raff(1954) y Gebhardt(1969) hicieron en sus estudios numéricos, el máximo error absoluto. No obstante, en el caso de los procedimientos del segundo tipo resulta conveniente utilizar otras medidas; en particular, emplearemos también el error de variación total. Si notamos respectivamente por $B_i^*(k,j)$ y $b_i^*(k)$ las aproximaciones por el método i-esimo de las probabilidades del intervalo $[k,j]$ y el punto k, recordamos que la definición de las medidas de precisión es la siguiente:

$$M_i(n,d,p) = \max_{k,j} |B(k,j;n,d,p) - B_i^*(k,j;n,d,p)|$$

y

$$D_i'(n,d,p) = \sum_{k=o}^{n} |b(k;n,d,p) - b_i^*(k;n,d,p)| \quad .$$

a) Métodos Normales.

Como normalmente, cuando se recurre a este tipo de procedimientos, lo que interesa es aproximar la función de distribución de un punto o bien la probabilidad de un intervalo, lo cual, a su vez, se reduce a lo primero, nosotros nos centraremos aquí en la aproximación de la función de dis

tribución de un punto arbitrario k de una VBM B(n,d,p) de media λ y desviación típica σ.

Vamos a realizar un tratamiento unificado de la extensión a la VBM de todos los métodos normales construidos para la binomial de Bernoulli, aunque luego nos restringiremos a los antes citados. La elección de A.1 se debió a que es el método mas inmediato; los otros métodos se eligieron por ser los mas precisos.

El tratamiento unificado se realiza por intermedio de la binomial de Bernoulli que tiene igual promedio y número de pruebas que la VBM aproximada. Sean p^{Δ} y σ^{Δ} su correspondiente probabilidad de "exito" y desviación típica. Notemos por z el valor de la normal tipificada cuya función de distribución coincida con $B_k(n,d,p)$; analogamente notemos por z^{Δ} el valor que verifica la misma propiedad respecto a $B(n,p^{\Delta})$. Llamemos f al factor que existe entre ambos valores; es decir

$$z = f \cdot z^{\Delta} \quad .$$

Evidentemente, este factor será, en principio, función de n, d, p y k. Por el 4º teorema de Hoeffding (I.12.a) y por el teorema de localización de la mediana de Jogdeo y Samuels

(I.12.c), se deduce que f toma siempre valores mayores que la unidad. (En esta afirmación también podemos incluir el caso en que $k = n$, ya que entonces z y z^{Δ} son ambos iguales a $+\infty$.

Consideremos ahora el valor límite al que converge f cuando n tiende a infinito. En dichas condiciones es válida la generalización de Cramer del teorema de Moivre-Laplace (I.12), de modo que se desprende que $B(n,d,p)$ y $B(n,p^{\Delta})$ se distribuyen aproximadamente según normales de igual media y desviaciones típicas respectivamente iguales a σ y σ^{Δ}. Por tanto, para k fijo tenemos que

$$\lim_{n\to\infty} f = \frac{\sigma^{\Delta}}{\sigma} = \sqrt{\frac{p^{\Delta}\cdot q^{\Delta}}{p\cdot q}} = \bar{f} \quad ,$$

que evidentemente resulta independiente del valor de k. Teniendo en cuenta esto y si notamos por $z_i^{\star}$, el valor de la normal tipificada cuya función de distribución aproxime a $B(k;n,p^{\Delta})$ a partir del método normal i-esimo, resulta que, para n suficientemente grande, $B(k;n,d,p)$ quedará aproximado a partir de dicho método por $\Phi(\bar{f}\cdot z_i^{\star})$.

A la vista de este resultado, nosotros proponemos utilizar la misma expresión para extender a cualquier

valor del parámetros n las aproximaciones normales de las probabilidades de las VBM. Observese que esta elección no contradice en absoluto al teorema de Gleser (I.12.a). De todo lo cual, se tiene que los métodos de aproximación de las probabilidades binomiales métricas basados en la distribución Normal quedan de la siguiente forma:

A.1) Aproximación de Moivre-Laplace:

$$B_1^{\star}(k;d) = \Phi\left(\overline{f}\cdot\frac{k+\frac{1}{2}-\lambda}{\sigma^{\Delta}}\right) = \Phi\left(\frac{k+\frac{1}{2}-\lambda}{\sigma}\right) \quad ,$$

es decir, coincide, como era de esperar, con la aproximación a partir de la normal que tenga la misma media y varianza que la variable aproximada.

A.4) Aproximación de Camp-Paulson:

$$B_4^{\star}(k;d) = \Phi\left(\overline{f}\cdot\frac{y_1}{3\sqrt{y_2}}\right) \quad ,$$

siendo

$$y_1 = c\cdot\left(9-\frac{1}{n-k}\right)+\frac{1}{k+1}-9$$

y

$$y_2 = \frac{c^2}{n-k}+\frac{1}{k+1} \quad ,$$

donde c viene dado por $\sqrt[3]{\frac{(n-k)\cdot p^{\Delta}}{(k+1)\cdot q^{\Delta}}}$.

A.6) Aproximación de Peizer-Pratt:

$$B_6^{\star}(k;d) = \Phi\left(\overline{f}\cdot\frac{\delta_j}{|k+\frac{1}{2}-\lambda|}\cdot y_3\right) \quad , \; j=1,2 \; ,$$

siendo

$$y_3 = \frac{2}{1+\frac{1}{6n}}\cdot\left((k+\tfrac{1}{2})\cdot\ln\frac{k+\frac{1}{2}}{\lambda} + (n-k-\tfrac{1}{2})\cdot\ln\frac{n-k-\frac{1}{2}}{n-\lambda}\right)^{\frac{1}{2}}$$

donde el coeficiente δ_j de cada versión tiene el siguiente valor:

$$\delta_1 = k + \frac{2}{3} - (n+\tfrac{1}{3})\cdot p^{\Delta}$$

y

$$\delta_2 = \delta_1 + .02\cdot\left(\frac{q^{\Delta}}{k+1} - \frac{p^{\Delta}}{n-k} + \frac{.5-p^{\Delta}}{n+1}\right) \quad .$$

En el apendice VI aparecen tabulados los máximos errores absolutos de estas aproximaciones para algunos casos particulares. De sus valores numéricos pueden extraerse las siguientes conclusiones empíricas:

i) El máximo error de $B_1^{\star}$ se mantiene siempre inferior a la cota propuesta por Raff para dicha aproximación en la binomial clásica, cuyo valor es $.140/\sigma$.

ii) Para p fijo, este error presenta una tendencia a disminuir conforme la distribución se hace mas simétrica; es decir, conforme d se acerca a $\frac{n}{2}$.

iii) En el peor de los casos, $B_4^{\star}$ y $B_6^{\star}$ tienen errores

muy similares a $B_1^{\star}$, aunque ligeramente superiores. Estos casos se presentan cuando p es inferior a .2 y d proximo a $\frac{n}{2}$. De todas formas, se mantienen acotados por $.140/\sigma$.

iv) Para p y n fijos, el máximo error absoluto de $B_4^{\star}$ y $B_6^{\star}$ crece al aumentar el parámetro d hasta $\frac{n}{2}$. Evidentemente, esto se debe a que conforme d crece la VBM se va diferenciando cada vez mas de una binomial de Bernoulli. En particular, la VBM tiene un coeficiente de asimetría siempre mayor que la binomial de igual promedio; la razón entre ambos coeficientes es justamente $\overline{f}$. Este inconveniente se podría eliminar modificando adecuadamente a $\overline{f}$ para que incluya la asimetría. No obstante, cuando p y n son superiores a .20 y 50, para cualquier valor de d se verifica que el error se mantiene por debajo de la cota que dio Raff para la aproximación de Camp-Paulson, cuyo valor es $.007/\sigma$.

v) En las mismas condiciones que acabamos de citar, $B_6^{\star}$ se muestra uniformemente mejor que $B_4^{\star}$. En los restantes casos, para d pequeños $B_6^{\star}$ sigue siendo mejor que $B_4^{\star}$, pero cuando d está proximo a $\frac{n}{2}$ se cambia el orden.

Finalmente, para terminar este apartado queremos hacer dos observaciones. Por una parte, se puede observar que, a partir de $n \geq 200$, los errores tabulados empiezan a aumentar, en contra de lo que sería esperable por el teorema central del límite. Ello se debe al predominio de los errores de cálculo sobre los verdaderos errores absolutos, ya que para el cálculo de los números combinatorio recurrimos, para esos valores de n, a la fórmula de Stirling.

Por otro lado, hay que señalar que nuestro estudio numérico restringido a la binomial clásica y a la aproximación de Peizer-Pratt, complementa el trabajo de dichos autores, ya que ellos hacen un estudio de su aproximación desde un punto de vista diferente.

b) Metodos basados en la distribución de Poisson.

En este apartado vamos a alterar la notación para estar acorde con la notación habitual en relación con la distribución de Poisson. Llamaremos ahora λ al producto np; es decir, el valor esperado de la binomial de Bernoulli que engendra la familia de VBM, en el supuesto de que tomemos a p inferior a $\frac{1}{2}$.

En principio, en virtud de la convergencia hacia la distribución de Poisson desplazada (III.8.e), parece razonable emplear la siguiente aproximación:

$$B^{\star}(k;d) = P(k-d;\lambda) \quad o \quad b^{\star}(k;d) = p(k-d;\lambda) \quad .$$

Por este procedimiento se asignan las mismas probabilidades, aúnque desplazadas, a todas las VBM y además se deja sin aproximar la cola que queda por debajo de d. La probabilidad de la cola no recubierta, $P[X_d<d]$, puede utilizarse, junto con el máximo error absoluto y error de variación total, como medida de la calidad de la aproximación.

Del estudio numérico de este procedimiento (no recogido en los apéndices) se sacan las siguientes observaciones:

i) Todas las medidas de error crecen al aumentar d.

ii) Como cotas empíricas de estos errores pueden utilizarse d·p para el máximo error absoluto y la probabilidad no recubierta, y 2·d·p para el error de variación total.

iii) Cuando n es grande, el máximo error es casi exactamente la mitad del error de variación total. Esta re

lación es particularmente exacta cuando $\lambda<1$.

De estas observaciones empíricas hemos podido demostrar completamente la relativa a la acotación de la probabilidad de la cola no recubierta y parcialmente, en el caso particular de la binomial de Bernoulli y de que $\lambda\leq\frac{1}{2}$, la relativa a la relación entre el máximo error y el error de variación total. Nuestros resultados son los siguientes:

Teorema.- Para cualesquiera que sean n y p, siempre que $d\geq 1$ se verifica que $P[X_d< d]<d\cdot p$.

Demostración: La haremos por inducción sobre d. Evidentemente para cualquier n y $d=1$, el teorema es cierto, puesto que

$$P[X_1<1] = P[X_1=0] = q^{n-1}\cdot p < 1\cdot p$$

Supongamos el teorema cierto para $d-1$ y consideremos la siguiente relación que liga las funciones de distribución de $B(n,d)$ y $B(n-1,d-1)$ (Vease III.6) :

$$P[X_d\leq k] = P[X_{d-1}\leq k-1] + p\cdot P[X_{d-1}=k] \quad .$$

Entonces, tomando $k=d-1$ y aplicando la hipótesis de inducción, resulta

$$P[X_d < d] = P[X_{d-1} < d-1] + p \cdot P[X_{d-1} = d-1] <$$

$$< (d-1) \cdot p + p \cdot P[X_{d-1} = d-1] < d \cdot p \quad .$$

Teorema.- Si $\lambda \leq \frac{1}{2}$, entonces el error de variación total es el doble del máximo error absoluto de la aproximación de Poisson de la binomial clásica.

Demostración: Nos basaremos en los resultados de Anderson y Samuels, expuestos en I.13.b, sobre las relaciones entre las probabilidades de la binomial y la Poisson de igual promedio.

De las relaciones entre las funciones de distribución, se tiene que

$$P(0) > B(0)$$

y

$$P(k) < B(k) \text{ , para } k \geq 1 \quad .$$

De la relación entre las probabilidades puntuales, se tiene que

$$p(k) > b(k) \text{ , para } k = 0 \quad y \quad k \geq 2 \quad .$$

Ahora bien, como $\Sigma |b(k) - p(k)| = 0$, resulta que el error de variación total vale

$$D' = \sum_{k \neq 1} p(k) - b(k) \quad + |b(1) - p(1)| = 2|b(1) - p(1)|$$

Por otra parte, como el máximo error absoluto es igual a

$$M = \text{Max}\left(B(k) - P(k)\right) - \text{Min}\left(B(j) - P(j)\right) \quad ,$$

resulta que su valor en estas condiciones es

$$M = B(1) - P(1) - \left(B(0) - P(0)\right) \quad ,$$

es decir, $b(1) - p(1)$, como queriamos demostrar.

Este teorema sigue siendo válido hasta $\lambda = 2 - \sqrt{2}$. Además la validez aproximada de la relación entre los errores para otros valores de λ tiene su base también en las relaciones de Anderson y Samuels o, mas precisamente, en la probable relación sistemática entre las probabilidades puntuales de la distribución binomial y de Poisson, tal que las primeras probabilidades son primero menores, luego mayores y nuevamente menores que las segundas probabilidades.

Por otra parte, las cotas de estos errores probablemente puedan deducirse de los resultados de Vervaat, Kerstand, y Hodges y LeCam (Vease I.10.d); los dos primeros se refieren a la variación total, mientras que el tercero a aproximadamente a la mitad del máximo error absoluto.

Si tenemos en cuenta que $B(n,d,p)$ se puede definir como la suma de las binomiales $B(n-d,p)$ y $B(d,q)$, una aplicación superficial del teorema de convergencia nos conduce a una ligera mejora de la aproximación. Si suponemos que p es muy pequeño, resulta que $B(d,q)$ tiene aproximadamente una distribución degenerada en el punto d y, por tanto, la distribución $B(n,d,p)$ debe aproximarse por $P((n-d)p)$ desplazada d unidades.

Con este análisis se consigue que a cada binomial métrica se le asignen distintas probabilidades, con lo cual se disminuyen ligeramente los errores de aproximación, aúnque sigue persistiendo su crecimiento respecto a parámetro d. Por otro lado, la utilización directa de la modificación de Kolmogorov, en el sentido del apartado I.12.E, no conduce a ninguna mejora de la aproximación cuando $d \neq 0$, en comparación con estos dos métodos, sino que, todo lo contrario, se comporta mucho peor que éstas.

Finalmente, un análisis mas detallado nos lleva a métodos de aproximación mucho mas precisos. Observese que si $B(n-d,p)$ está correctamente aproximada por $P((n-d)\cdot p)$, con errores acotados matematicamente por $(n-d)\cdot p^2$ ó $2\cdot p$ (según convenga), $B(d,q)$, puesto que sus probabilidades coinci-

den con las de B(d,p) tomadas en orden opuesto, debería aproximarse por las probabilidades de P(d.p), quedando garantizada una precisión similar. De esta manera se llega a que B(n,d,p) debe aproximarse por las probabilidades de la variable aleatoria

$$Y = P((n-d).p) - P(d.p) + d \quad .$$

Por tanto, tenemos la siguiente generalización de $B_7^{\star}$:

$$B_7^{\star}(k;d) = \begin{cases} \sum_{i=0} p(i;(n-d)p).p(i+d-k;dp) & \text{si } k<d \\ \sum_{i=0} p(k-d+i;(n-d)p).p(i;dp) & \text{si } k\geq d \end{cases}$$

Ahora bien, como cada probabilidad de la distribución de Poisson es el aproximante $B_7^{\star}$ de la correspondiente binomial, podemos extender la modificación de Kolmogorov a las VBM remplazando dichas probabilidades por el aproximante $B_8^{\star}$ de las probabilidades binomiales puntuales implicadas.

En el apéndice VII hemos tabulado el máximo error y el error de variación total de estas dos aproximaciones.(Observese que con estos procedimientos no queda ninguna cola de la distribución sin recubrir). El comportamiento de ambas aproximaciones es francamente bueno como lo sugie-

ren los siguientes hechos:

i) Los errores están uniformemente acotados por las cotas de la binomial. Así resulta que M_7 y M_8 son aproximadamente iguales a $\frac{p}{4}$ y $\frac{p^2}{4}$, respectivamente. Ambos errores practicamente son independientes del parámetro n. (Véase Gebhardt(1969)) .

ii) Ambos errores tienden a disminuir conforme el parámetro d aumenta. Para estas aproximaciones, salvo cuando p es muy pequeño, deja de cumplirse que el error de variación total sea practicamente el doble que el máximo error.

No hemos encontrado ningún procedimiento para demostrar estos hechos, pero una posible explicación es que, al hacer la convolución de los aproximantes,los errores por exceso de una se compensan con los errores por defecto de la otra, y viceversa, de modo que aumentaría la precisión. Esta explicación es coherente con la perdida de la correspondencia entre ambos errores, que es debida a que las probabilidades aproximadas oscilan alrededor de las verdaderas, en vez de seguir un comportamiento sistemático como en la aproximación por la Poisson desplazada.

APENDICES

APENDICE I

PROBABILIDADES PUNTUALES DE LA BINOMIAL METRICA

Se tabula $b_k(n,d,p)$, definida como

$$b_k = \sum_{i=i_1}^{i_2} \binom{d}{i} \cdot \binom{n-d}{k-i} \cdot p^{d+k-2i} \cdot q^{n+2i-d-k} \quad ,$$

siendo $i_1 = \max(k-n+d,0)$ y $i_2 = \min(k,d)$, en los siguientes casos particulares:

$$n = 2,3,\ldots,10$$
$$d = 0,1,\ldots,[\tfrac{n}{2}]$$
$$p = .05,(.05),.5 \quad .$$

En caso de que p fuese mayor que 0.5 ó que d fuese mayor que $\frac{n}{2}$, $b_k(n,d,p)$ se obtiene por alguna de las siguientes identidades:

$$b_k(n,d,p) = b_k(n,n-d,1-p)$$
$$b_k(n,d,p) = b_{n-k}(n,d,1-p)$$
$$b_k(n,d,p) = b_{n-k}(n,n-d,p) \quad .$$

N\ D\ K\ P:	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
2										
0										
0	.902500	.810000	.722500	.640000	.562500	.490000	.422500	.360000	.302500	.250000
1	.095000	.180000	.255000	.320000	.375000	.420000	.455000	.480000	.495000	.500000
2	.002500	.010000	.022500	.040000	.062500	.090000	.122500	.160000	.202500	.250000
1										
0	.047500	.090000	.127500	.160000	.187500	.210000	.227500	.240000	.247500	.250000
1	.905000	.820000	.745000	.680000	.625000	.580000	.545000	.520000	.505000	.500000
2	.047500	.090000	.127500	.160000	.187500	.210000	.227500	.240000	.247500	.250000
3										
0										
0	.857375	.729000	.614125	.512000	.421875	.343000	.274625	.216000	.166375	.125000
1	.135375	.243000	.325125	.384000	.421875	.441000	.443625	.432000	.408375	.375000
2	.007125	.027000	.057375	.096000	.140625	.189000	.238875	.288000	.334125	.375000
3	.000125	.001000	.003375	.008000	.015625	.027000	.042875	.064000	.091125	.125000
1										
0	.045125	.081000	.108375	.128000	.140625	.147000	.147875	.144000	.136125	.125000
1	.862125	.747000	.652375	.576000	.515625	.469000	.433875	.408000	.389125	.375000
2	.090375	.163000	.220125	.264000	.296875	.321000	.338625	.352000	.363375	.375000
3	.002375	.009000	.019125	.032000	.046875	.063000	.079625	.096000	.111375	.125000
4										
0										
0	.814506	.656100	.522006	.409600	.316406	.240100	.178506	.129600	.091506	.062500
1	.171475	.291600	.368475	.409600	.421875	.411600	.384475	.345600	.299475	.250000
2	.013537	.048600	.097537	.153600	.210938	.264600	.310537	.345600	.367537	.375000
3	.000475	.003600	.011475	.025600	.046875	.075600	.111475	.153600	.200475	.250000
4	.000006	.000100	.000506	.001600	.003906	.008100	.015006	.025600	.041006	.062500
1										
0	.042869	.072900	.092119	.102400	.105469	.102900	.096119	.086400	.074869	.062500
1	.821275	.680400	.570775	.486400	.421875	.372400	.333775	.302400	.275275	.250000
2	.128962	.221400	.284962	.326400	.351563	.365400	.371962	.374400	.374962	.375000
3	.006775	.024400	.049275	.078400	.109375	.140400	.170275	.198400	.224775	.250000
4	.000119	.000900	.002869	.006400	.011719	.018900	.027869	.038400	.050119	.062500
2										
0	.002256	.008100	.016256	.025600	.035156	.044100	.051756	.057600	.061256	.062500
1	.085975	.147600	.189975	.217600	.234375	.243600	.247975	.249600	.249975	.250000
2	.823537	.688600	.587537	.513600	.460938	.424600	.400537	.385600	.377537	.375000
3	.085975	.147600	.189975	.217600	.234375	.243600	.247975	.249600	.249975	.250000
4	.002256	.008100	.016256	.025600	.035156	.044100	.051756	.057600	.061256	.062500
5										
0										
0	.773781	.590490	.443705	.327680	.237305	.168070	.116029	.077760	.050328	.031250
1	.203627	.328050	.391504	.409600	.395508	.360150	.312386	.259200	.205889	.156250
2	.021434	.072900	.138178	.204800	.263672	.308700	.336416	.345600	.336909	.312500
3	.001128	.008100	.024384	.051200	.087891	.132300	.181147	.230400	.275653	.312500
4	.000030	.000450	.002152	.006400	.014648	.028350	.048770	.076800	.112767	.156250
5	.000000	.000010	.000076	.000320	.000977	.002430	.005252	.010240	.018453	.031250

N\ K\	P: .05	.10	.15	.20	.25	.30	.35	.40	.45	.50
5										
1										
0	.040725	.065610	.078301	.081920	.079102	.072030	.062477	.051840	.041178	.031250
1	.782355	.619650	.498976	.409600	.342773	.291550	.250595	.216000	.185092	.156250
2	.163578	.267300	.327834	.358400	.369141	.367500	.358597	.345600	.330103	.312500
3	.012884	.044100	.084628	.128000	.169922	.207900	.240866	.268800	.292359	.312500
4	.000452	.003250	.009830	.020800	.036133	.055350	.077711	.102400	.128714	.156250
5	.000006	.000090	.000430	.001280	.002930	.005670	.009754	.015360	.022553	.031250
2										
0	.002143	.007290	.013818	.020480	.026367	.030870	.033642	.034560	.033691	.031250
1	.081789	.133650	.163917	.179200	.184570	.183750	.179298	.172800	.165052	.156250
2	.786659	.634500	.527903	.454400	.404297	.370300	.347141	.331200	.320134	.312500
3	.122853	.201700	.249609	.276800	.291016	.297900	.301372	.304000	.307378	.312500
4	.006442	.022050	.042314	.064000	.084961	.103950	.120433	.134400	.146180	.156250
5	.000113	.000810	.002438	.005120	.008789	.013230	.018115	.023040	.027565	.031250
6										
0										
0	.735092	.531441	.377149	.262144	.177979	.117649	.075419	.046656	.027681	.015625
1	.232134	.354294	.399334	.393216	.355957	.302526	.243661	.186624	.135887	.093750
2	.030544	.098415	.176177	.245760	.296631	.324135	.328005	.311040	.277950	.234375
3	.002143	.014580	.041453	.081920	.131836	.185220	.235491	.276480	.303218	.312500
4	.000085	.001215	.005486	.015360	.032959	.059535	.095102	.138240	.186066	.234375
5	.000002	.000054	.000357	.001536	.004395	.010206	.020484	.036864	.060894	.093750
6	.000000	.000001	.000011	.000064	.000244	.000729	.001838	.004096	.008304	.015625
1										
0	.038689	.059049	.066556	.065536	.059326	.050421	.040610	.031104	.022648	.015625
1	.745273	.564246	.435875	.344064	.276855	.225694	.184754	.150336	.120331	.093750
2	.194517	.302535	.353505	.368640	.362549	.344715	.320796	.293760	.264848	.234375
3	.020419	.066420	.121109	.174080	.219727	.255780	.282072	.299520	.309344	.312500
4	.001073	.007335	.021049	.042240	.069580	.101115	.134815	.168960	.202354	.234375
5	.000028	.000406	.001840	.005184	.011230	.020574	.033539	.050176	.070326	.093750
6	.000000	.000009	.000065	.000256	.000732	.001701	.003414	.006144	.010149	.015625
2										
0	.002036	.006561	.011745	.016384	.019775	.021609	.021867	.020736	.018530	.015625
1	.077807	.121014	.141402	.147456	.145020	.137886	.128319	.117504	.105939	.093750
2	.751416	.584415	.473305	.399360	.349365	.314335	.288396	.267840	.250347	.234375
3	.156043	.244980	.291353	.312320	.319336	.319620	.317391	.314880	.313118	.312500
4	.012263	.040015	.073408	.106560	.136475	.162135	.183761	.202240	.218719	.234375
5	.000429	.002934	.008420	.016896	.027832	.040446	.053926	.067584	.080942	.093750
6	.000006	.000081	.000366	.001024	.002197	.003969	.006340	.009216	.012404	.015625
3										
0	.000107	.000729	.002073	.004096	.006592	.009261	.011775	.013824	.015161	.015625
1	.006126	.019926	.036333	.052224	.065918	.076734	.084621	.089856	.092803	.093750
2	.117033	.183735	.218515	.234240	.239502	.239715	.238043	.236160	.234839	.234375
3	.753469	.591220	.486159	.418880	.375977	.348580	.331122	.320320	.314394	.312500
4	.117033	.183735	.218515	.234240	.239502	.239715	.238043	.236160	.234839	.234375
5	.006126	.019926	.036333	.052224	.065918	.076734	.084621	.089856	.092803	.093750
6	.000107	.000729	.002073	.004096	.006592	.009261	.011775	.013824	.015161	.015625

N\ D\ K\	P: .05	.10	.15	.20	.25	.30	.35	.40	.45	.50
7										
0										
0	.698337	.478297	.320577	.209715	.133484	.082354	.049022	.027994	.015224	.007813
1	.257282	.372009	.396007	.367002	.311462	.247063	.184776	.130637	.087194	.054688
2	.040623	.124003	.209651	.275251	.311462	.317652	.298485	.261274	.214022	.164063
3	.003563	.022963	.061662	.114688	.173035	.226894	.267871	.290304	.291848	.273438
4	.000188	.002551	.010882	.028672	.057678	.097240	.144238	.193536	.238785	.273438
5	.000006	.000170	.001152	.004301	.011536	.025005	.046600	.077414	.117221	.164063
6	.000000	.000006	.000068	.000358	.001282	.003572	.008364	.017203	.031969	.054688
7	.000000	.000000	.000002	.000013	.000061	.000219	.000643	.001638	.003737	.007813
1										
0	.036755	.053144	.056572	.052429	.044495	.035295	.026397	.018662	.012456	.007813
1	.709944	.513726	.380477	.288358	.222473	.173112	.134304	.102643	.076373	.054688
2	.222055	.328706	.365861	.363725	.341125	.309009	.273181	.236390	.199815	.164063
3	.029124	.090031	.155969	.212992	.255432	.282460	.295625	.297216	.289321	.273438
4	.002040	.013243	.036058	.068608	.107117	.147514	.186355	.221184	.250500	.273438
5	.000080	.001099	.004722	.012595	.025818	.044736	.068986	.097690	.129739	.164063
6	.000002	.000049	.000331	.001242	.003357	.007363	.013958	.023757	.037229	.054688
7	.000000	.000001	.000010	.000051	.000183	.000510	.001195	.002458	.004567	.007813
2										
0	.001934	.005905	.009983	.013107	.014832	.015126	.014214	.012442	.010192	.007813
1	.074018	.109569	.121954	.121242	.115708	.108003	.091060	.078797	.066605	.054688
2	.717735	.538075	.423520	.348979	.296279	.261400	.232369	.207706	.185364	.164063
3	.185812	.278923	.318646	.329728	.326843	.318034	.307243	.296064	.284671	.273438
4	.019452	.060511	.106100	.147712	.182190	.209380	.230532	.247296	.261199	.273438
5	.001021	.006642	.018168	.034829	.054993	.076953	.099368	.121446	.142941	.164063
6	.000027	.000366	.001574	.004198	.008606	.014912	.022995	.032563	.043246	.054688
7	.000000	.000008	.000055	.000205	.000549	.001191	.002219	.003686	.005582	.007813
3										
0	.000102	.000656	.001762	.003277	.004944	.006483	.007653	.008294	.008339	.007813
1	.005825	.018006	.031194	.042598	.051086	.056492	.059125	.059443	.057864	.054688
2	.111487	.167354	.191188	.197837	.196106	.190821	.184346	.177638	.170923	.164063
3	.721647	.550471	.446012	.381952	.341858	.315920	.298544	.286656	.278594	.273438
4	.148854	.224483	.258662	.271168	.273621	.272374	.270621	.269824	.270639	.273438
5	.011671	.036307	.063660	.088627	.109314	.125628	.138319	.148378	.156719	.164063
6	.000408	.002649	.007212	.013722	.021423	.029503	.037271	.044237	.050100	.054688
7	.000005	.000073	.000311	.000819	.001648	.002778	.004121	.005530	.006822	.007813
8										
0										
0	.663420	.430467	.272491	.167772	.100113	.057648	.031864	.016796	.008373	.003906
1	.279335	.382638	.384692	.335544	.266968	.197650	.137262	.089579	.054808	.031250
2	.051456	.148803	.237604	.293601	.311462	.296475	.258687	.209019	.156949	.109375
3	.005416	.033067	.083860	.146801	.207642	.254122	.278586	.278692	.256826	.218750
4	.000356	.004593	.018499	.045875	.086517	.136137	.187510	.232243	.262663	.273438
5	.000015	.000408	.002612	.009175	.023071	.046675	.080773	.123863	.171925	.218750
6	.000000	.000023	.000230	.001147	.003845	.010002	.021747	.041288	.070333	.109375
7	.000000	.000001	.000012	.000082	.000366	.001225	.003346	.007864	.016441	.031250
8	.000000	.000000	.000000	.000003	.000015	.000066	.000225	.000655	.001682	.003906

N\ P:	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
8										
1										
0	.034917	.047830	.048087	.041943	.033371	.024706	.017158	.011197	.006851	.003906
1	.676284	.467668	.331892	.241172	.177979	.131767	.096536	.069051	.047611	.031250
2	.246449	.347208	.368054	.348652	.311462	.268240	.224574	.182891	.144266	.109375
3	.038770	.113899	.187452	.243138	.276855	.290425	.287770	.272886	.249043	.218750
4	.003395	.020922	.054045	.097485	.144196	.187998	.224599	.251597	.267969	.273438
5	.000178	.002313	.009422	.023798	.046143	.075570	.110065	.147087	.184081	.218750
6	.000006	.000154	.000990	.003512	.008972	.018575	.033217	.053330	.078858	.109375
7	.000000	.000006	.000058	.000289	.000977	.002566	.005662	.010977	.019265	.031250
8	.000000	.000000	.000001	.000010	.000046	.000153	.000418	.000983	.002055	.003906
2										
0	.001838	.005314	.008486	.010486	.011124	.010588	.009239	.007465	.005605	.003906
1	.070414	.099202	.105158	.099615	.088989	.076640	.064164	.052255	.041219	.031250
2	.685549	.495224	.378285	.303432	.252136	.213881	.182911	.156142	.131922	.109375
3	.212408	.304839	.334377	.333578	.319702	.301044	.281037	.260721	.240093	.218750
4	.027770	.082353	.137982	.184115	.218353	.241977	.257381	.266803	.271851	.273438
5	.001942	.012029	.031358	.057405	.086792	.116681	.145276	.171786	.196157	.218750
6	.000077	.000994	.004063	.010324	.020203	.033524	.049726	.068116	.088109	.109375
7	.000002	.000044	.000283	.001004	.002563	.005307	.009491	.015237	.022531	.031250
8	.000000	.000001	.000008	.000041	.000137	.000357	.000777	.001475	.002512	.003906
3										
0	.000097	.000590	.001498	.002621	.003708	.004539	.004975	.004977	.004586	.003906
1	.005539	.016271	.026779	.034734	.039551	.041489	.041110	.038984	.035578	.031250
2	.106204	.152419	.167189	.166789	.159851	.150522	.140518	.130360	.120046	.109375
3	.691139	.512160	.407789	.345129	.305420	.278391	.258575	.243049	.230142	.218750
4	.177494	.257082	.286764	.293325	.290680	.285438	.280394	.276557	.274219	.273438
5	.018530	.055125	.092910	.125135	.150391	.169652	.184625	.196956	.207983	.218750
6	.000971	.006015	.015679	.028703	.043396	.058340	.072638	.085893	.098079	.109375
7	.000025	.000330	.001346	.003400	.006592	.010796	.015724	.021012	.026297	.031250
8	.000000	.000007	.000047	.000164	.000412	.000833	.001442	.002212	.003070	.003906
4										
0	.000005	.000066	.000264	.000655	.001236	.001945	.002679	.003318	.003752	.003906
1	.000388	.002391	.006177	.011141	.016479	.021486	.025669	.028754	.030625	.031250
2	.011108	.032941	.055193	.073646	.087341	.096791	.102952	.106721	.108741	.109375
3	.141995	.205666	.229411	.234660	.232544	.228351	.224315	.221245	.219375	.218750
4	.693008	.517873	.417910	.359795	.324799	.302857	.288771	.279923	.275014	.273438
5	.141995	.205666	.229411	.234660	.232544	.228351	.224315	.221245	.219375	.218750
6	.011108	.032941	.055193	.073646	.087341	.096791	.102952	.106721	.108741	.109375
7	.000388	.002391	.006177	.011141	.016479	.021486	.025669	.028754	.030625	.031250
8	.000005	.000066	.000264	.000655	.001236	.001945	.002679	.003318	.003752	.003906

N\ P:	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
K\										
9										
0										
0	.630249	.387420	.231617	.134218	.075085	.040354	.020712	.010078	.004605	.001953
1	.298539	.387420	.367862	.301990	.225254	.155650	.100373	.060466	.033912	.017578
2	.062850	.172187	.259667	.301990	.300339	.266828	.216188	.161243	.110986	.070313
3	.007718	.044641	.106922	.176161	.233597	.266828	.271621	.250823	.211881	.164063
4	.000609	.007440	.028303	.066060	.116798	.171532	.219386	.250823	.260036	.246094
5	.000032	.000827	.004995	.016515	.038933	.073514	.118131	.167215	.212757	.246094
6	.000001	.000061	.000588	.002753	.008652	.021004	.042406	.074318	.116049	.164063
7	.000000	.000003	.000044	.000295	.001236	.003858	.009786	.021234	.040693	.070313
8	.000000	.000000	.000002	.000018	.000103	.000413	.001317	.003539	.008323	.017578
9	.000000	.000000	.000000	.000001	.000004	.000020	.000079	.000262	.000757	.001953
1										
0	.033171	.043047	.040874	.033554	.025028	.017294	.011153	.006718	.003768	.001953
1	.644216	.425684	.289321	.201327	.141827	.099649	.068754	.045909	.029269	.017578
2	.267941	.359254	.362629	.327156	.278091	.227298	.179761	.137355	.100771	.070313
3	.049154	.137230	.214543	.264241	.285507	.283769	.265651	.236888	.201894	.164063
4	.005163	.030220	.074056	.126616	.177361	.218726	.246709	.260112	.259453	.246094
5	.000339	.004174	.016116	.038535	.070656	.109298	.150152	.188891	.221831	.246094
6	.000014	.000370	.002254	.007569	.018265	.035673	.060114	.090833	.126208	.164063
7	.000000	.000020	.000198	.000934	.002975	.007369	.015306	.027918	.046082	.070313
8	.000000	.000001	.000010	.000066	.000278	.000877	.002253	.004981	.009799	.017578
9	.000000	.000000	.000000	.000002	.000011	.000046	.000146	.000393	.000925	.001953
2										
0	.001746	.004783	.007213	.008389	.008343	.007412	.006005	.004479	.003083	.001953
1	.066985	.089814	.090657	.081789	.069523	.056824	.044940	.034339	.025193	.017578
2	.654793	.455622	.337316	.262668	.211349	.172709	.141349	.114587	.091106	.070313
3	.236065	.323877	.340963	.327549	.302811	.274895	.246693	.218889	.191416	.164063
4	.037002	.104601	.167441	.214008	.243690	.259697	.265660	.264370	.257560	.246094
5	.003234	.019061	.047351	.082747	.119682	.154270	.184512	.209793	.230220	.246094
6	.000170	.002097	.008157	.019741	.036850	.058471	.083168	.109584	.136731	.164063
7	.000005	.000139	.000850	.002868	.006973	.013772	.023573	.036389	.052041	.070313
8	.000000	.000005	.000049	.000233	.000744	.001842	.003827	.006980	.011520	.017578
9	.000000	.000000	.000001	.000008	.000034	.000107	.000272	.000590	.001130	.001953
3										
0	.000092	.000531	.001273	.002097	.002781	.003177	.003234	.002986	.002522	.001953
1	.005267	.014703	.022987	.028312	.030590	.030404	.028463	.025381	.021631	.017578
2	.101171	.138805	.146127	.140378	.129776	.117812	.105725	.093810	.082035	.070313
3	.661892	.476186	.371699	.309461	.269028	.240030	.217255	.197973	.180599	.164063
4	.203176	.282590	.304918	.303686	.294365	.283324	.272757	.263154	.254384	.246094
5	.026478	.075320	.121988	.158773	.185463	.204388	.218144	.228796	.237789	.246094
6	.001849	.010926	.027264	.047989	.070145	.091734	.111833	.130318	.147536	.164063
7	.000073	.000899	.003496	.008460	.015793	.025059	.035644	.046965	.058599	.070313
8	.000002	.000040	.000242	.000811	.001957	.003822	.006441	.009732	.013522	.017578
9	.000000	.000001	.000007	.000033	.000103	.000250	.000505	.000885	.001382	.001953

N\ \ P:	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
9										
4										
0	.000005	.000059	.000225	.000524	.000927	.001361	.001741	.001991	.002064	.001953
1	.000369	.002159	.005290	.009044	.012669	.015623	.017622	.018579	.018532	.017578
2	.010572	.029886	.047840	.061145	.069626	.074199	.075903	.075534	.073589	.070313
3	.135451	.188393	.203279	.202457	.196243	.188883	.181838	.175436	.169589	.164063
4	.665457	.486652	.389635	.334768	.301735	.280505	.266211	.256452	.249977	.246094
5	.169546	.236887	.257686	.259687	.255608	.250702	.246875	.244717	.244413	.246094
6	.017652	.050214	.081326	.105849	.123642	.136259	.145429	.152531	.158526	.164063
7	.000924	.005446	.013529	.023642	.034195	.044077	.052718	.059941	.065777	.070313
8	.000024	.000298	.001151	.002753	.005047	.007807	.010725	.013492	.015845	.017578
9	.000000	.000007	.000040	.000131	.000309	.000583	.000938	.001327	.001689	.001953
10										
0										
0	.598737	.348678	.196874	.107374	.056314	.028248	.013463	.006047	.002533	.000977
1	.315125	.387420	.347425	.268435	.187712	.121061	.072492	.040311	.020724	.009766
2	.074635	.193710	.275897	.301990	.281568	.233474	.175653	.120932	.076303	.043945
3	.010475	.057396	.129834	.201327	.250282	.266828	.252220	.214991	.166478	.117188
4	.000965	.011160	.040096	.088080	.145998	.200121	.237668	.250823	.238367	.205078
5	.000061	.001488	.008491	.026424	.058399	.102919	.152570	.200658	.234033	.246094
6	.000003	.000138	.001249	.005505	.016222	.036757	.068910	.111477	.159566	.205078
7	.000000	.000009	.000126	.000786	.003090	.009002	.021203	.042467	.074603	.117188
8	.000000	.000000	.000008	.000074	.000386	.001447	.004281	.010617	.022890	.043945
9	.000000	.000000	.000000	.000004	.000029	.000138	.000512	.001573	.004162	.009766
10	.000000	.000000	.000000	.000000	.000001	.000006	.000028	.000105	.000341	.000977
1										
0	.031512	.038742	.034743	.026844	.018771	.012106	.007249	.004031	.002072	.000977
1	.613664	.387420	.252054	.167772	.112627	.074942	.048593	.030233	.017793	.009766
2	.286755	.365897	.351633	.301990	.244025	.189003	.140908	.100777	.066595	.043945
3	.060094	.159432	.236756	.276824	.283653	.266828	.235590	.197075	.156389	.117188
4	.007363	.040921	.095129	.154141	.204397	.238239	.253339	.250823	.233551	.205078
5	.000580	.006779	.024807	.056151	.097332	.142127	.183947	.217380	.238761	.246094
6	.000031	.000750	.004334	.013763	.031363	.057761	.091627	.130056	.169239	.205078
7	.000001	.000055	.000506	.002261	.006798	.015860	.030989	.053084	.082139	.117188
8	.000000	.000003	.000038	.000240	.000953	.002825	.006822	.014156	.026126	.043945
9	.000000	.000000	.000002	.000015	.000078	.000295	.000884	.002228	.004918	.009766
10	.000000	.000000	.000000	.000000	.000003	.000014	.000051	.000157	.000416	.000977
2										
0	.001659	.004305	.006131	.006711	.006257	.005188	.003903	.002687	.001696	.000977
1	.063723	.081310	.078141	.067109	.054228	.042001	.031313	.022395	.015243	.009766
2	.625402	.419041	.300317	.226492	.175893	.137943	.107606	.082488	.061445	.043945
3	.257002	.337052	.340416	.314573	.279945	.244239	.209823	.177168	.146276	.117188
4	.046955	.126529	.193470	.236716	.258471	.264256	.259022	.246178	.227795	.205078
5	.004922	.027615	.065365	.108999	.150684	.185898	.212914	.231624	.242523	.246094
6	.000323	.003794	.014036	.032342	.057558	.087211	.118639	.149668	.178801	.205078
7	.000014	.000335	.001946	.006242	.014442	.027182	.044431	.065667	.090151	.117188
8	.000000	.000019	.000169	.000760	.002301	.005421	.010738	.018743	.029755	.043945
9	.000000	.000001	.000008	.000053	.000212	.000628	.001516	.003146	.005806	.009766
10	.000000	.000000	.000000	.000002	.000009	.000032	.000095	.000236	.000509	.000977

N\ \ D\ \ K\ P:	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
10										
3										
0	.000087	.000478	.001082	.001678	.002086	.002224	.002102	.001792	.001387	.000977
1	.005008	.013286	.019730	.023069	.023638	.022236	.019632	.016423	.013032	.009766
2	.096376	.126394	.127656	.117965	.104980	.091590	.078683	.066438	.054854	.043945
3	.633856	.442448	.337863	.275644	.234215	.203365	.178220	.156308	.136245	.117188
4	.226112	.301950	.314935	.304841	.288031	.270336	.253331	.237082	.221181	.205078
5	.035313	.096047	.149428	.187756	.212688	.228069	.237258	.242539	.245257	.246094
6	.003081	.017365	.041472	.070146	.098974	.125530	.149042	.169710	.188150	.205078
7	.000162	.001902	.007061	.016366	.029381	.045062	.062310	.080306	.098620	.117188
8	.000005	.000126	.000730	.002341	.005416	.010193	.016662	.024625	.033807	.043945
9	.000000	.000005	.000042	.000188	.000566	.001322	.002582	.004424	.006845	.009766
10	.000000	.000000	.000001	.000007	.000026	.000075	.000177	.000354	.000622	.000977
4										
0	.000005	.000053	.000191	.000419	.000695	.000953	.001132	.001194	.001135	.000977
1	.000351	.001949	.004530	.007340	.009733	.011345	.012064	.011944	.011121	.009766
2	.010062	.027113	.041458	.050725	.055387	.056626	.055505	.052752	.048813	.043945
3	.129207	.172543	.179963	.174195	.164589	.154478	.144761	.135475	.126389	.117188
4	.638957	.456826	.361682	.308306	.275362	.253018	.236681	.224046	.213802	.205078
5	.194341	.261863	.277479	.274703	.267139	.259643	.253642	.249411	.246916	.246094
6	.025247	.068881	.107780	.136616	.156633	.170592	.180935	.189405	.197175	.205078
7	.001760	.009923	.023698	.040083	.056557	.071732	.085167	.096977	.107514	.117188
8	.000069	.000813	.003008	.006930	.012334	.018688	.025423	.032072	.038314	.043945
9	.000001	.000036	.000206	.000655	.001493	.002751	.004363	.006193	.008059	.009766
10	.000000	.000001	.000006	.000026	.000077	.000175	.000328	.000531	.000760	.000977
5										
0	.000000	.000006	.000034	.000105	.000232	.000408	.000609	.000796	.000929	.000977
1	.000023	.000269	.000984	.002228	.003862	.005640	.007299	.008626	.009475	.009766
2	.000879	.004931	.011672	.019464	.026908	.033196	.038020	.041361	.043308	.043945
3	.016816	.045737	.071156	.089407	.101280	.108604	.112980	.115495	.116789	.117188
4	.161951	.218219	.231232	.228919	.222616	.216369	.211369	.207842	.205764	.205078
5	.640661	.461675	.369843	.319752	.290203	.271564	.259443	.251758	.247473	.246094
6	.161951	.218219	.231232	.228919	.222616	.216369	.211369	.207842	.205764	.205078
7	.016816	.045737	.071156	.089407	.101280	.108604	.112980	.115495	.116789	.117188
8	.000879	.004931	.011672	.019464	.026908	.033196	.038020	.041361	.043308	.043945
9	.000023	.000269	.000984	.002228	.003862	.005640	.007299	.008626	.009475	.009766
10	.000000	.000006	.000034	.000105	.000232	.000408	.000609	.000796	.000929	.000977

APENDICE II

ESTIMADOR DE MAXIMA VEROSIMILITUD DE p EN B(2,1,p)

Se tabula el valor esperado, sesgo y varianza del estimador de máxima verosimilitud del parámetro p en la distribución B(2,1,p) para los siguientes tamaños muestrales:

N = 1,2,5,10,25,50,100,500,1000,5000,1000 .

Asimismo se muestra la cota de Cramer-Frechet-Rao, calculada de modo aproximado, para la estimación de dicho parámetro.

Estos resultados sirven de punto de referencia de lo mas que puede esperarse de otros estimadores de p cuando el parámetro d coincide con $\frac{n}{2}$.

Estimador de Maxima Verosimilitud de p en B(2,1)

PR :	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
TM 1										
E.MV.	.047500	.090000	.127500	.160000	.187500	.210000	.227500	.240000	.247500	.250000
SESGO	-.002500	-.010000	-.022500	-.040000	-.062500	-.090000	-.122500	-.160000	-.202500	-.250000
VZA	.021494	.036900	.047494	.054400	.058594	.060900	.061994	.062400	.062494	.062500
COTA	.021493	.036890	.047461	.054386	.058613	.060884	.062039	.062383	.062291	.062521
TM 2										
E.MV.	.090487	.163800	.222487	.268800	.304688	.331800	.351487	.364800	.372487	.375000
SESGO	.040487	.063800	.072487	.068800	.054688	.031800	.001487	-.035200	-.077513	-.125000
VZA	.037056	.055070	.061743	.062147	.059509	.055809	.052200	.049321	.047497	.046875
COTA	.035208	.049606	.052673	.050286	.045804	.040963	.036885	.033737	.031679	.031310
TM 5										
E.MV.	.058099	.117089	.173010	.222811	.264740	.298181	.323287	.340587	.350673	.353983
SESGO	.008099	.017089	.023010	.022811	.014740	-.001818	-.026713	-.059413	-.099327	-.146017
VZA	.007632	.016033	.022811	.026929	.028576	.028534	.027656	.026613	.025836	.025554
COTA	.007448	.015545	.021971	.025619	.026689	.025965	.024597	.023082	.021866	.021791
TM 10										
E.MV.	.053662	.110761	.171401	.231075	.284088	.326822	.358340	.379416	.391357	.395210
SESGO	.003662	.010761	.021401	.031075	.034088	.026822	.008340	-.020584	-.058643	-.104790
VZA	.003288	.008701	.015834	.021780	.024621	.024643	.023184	.021467	.020224	.019781
COTA	.003208	.008102	.014477	.019742	.021844	.020948	.018836	.016561	.014840	.014680
TM 25										
E.MV.	.051259	.103243	.156901	.213065	.269324	.319799	.359372	.386353	.401592	.406474
SESGO	.001259	.003243	.006901	.013065	.019324	.019799	.009372	-.013647	-.048408	-.093526
VZA	.001131	.002586	.004854	.008290	.011762	.013589	.013552	.012618	.011738	.011400
COTA	.001128	.002546	.004663	.007838	.011147	.012665	.012307	.010861	.009426	.009445
TM 50										
E.MV.	.050608	.101517	.153060	.206300	.262728	.319255	.367241	.400602	.419123	.424944
SESGO	.000608	.001517	.003060	.006300	.012728	.019255	.017241	.000602	-.030877	-.075056
VZA	.000547	.001212	.002134	.003783	.006689	.009528	.010414	.009714	.008780	.008401
COTA	.000546	.001203	.002092	.003604	.006265	.008801	.009453	.008195	.006620	.006694
TM 100										
E.MV.	.050299	.100737	.151444	.202759	.255874	.312285	.366306	.406645	.429164	.436131
SESGO	.000299	.000737	.001444	.002759	.005874	.012285	.016306	.006645	-.020836	-.063869
VZA	.000269	.000590	.001010	.001650	.002945	.005322	.007109	.007022	.006233	.005869
COTA	.000269	.000586	.000993	.001611	.002843	.004963	.006693	.006120	.004613	.004811
TM 500										
E.MV.	.050058	.100139	.150271	.200502	.250968	.302073	.355996	.410548	.446064	.456762
SESGO	.000058	.000139	.000271	.000502	.000968	.002074	.005996	.010548	-.003936	-.043238
VZA	.000053	.000116	.000197	.000308	.000483	.000838	.001867	.003144	.002925	.002589
COTA	.000053	.000113	.000183	.000291	.000508	.000819	.001920	.002903	.001787	.002362
TM 1000										
E.MV.	.050028	.100060	.150123	.200236	.250476	.300955	.352677	.408265	.450610	.463563
SESGO	.000028	.000060	.000123	.000236	.000476	.000955	.002677	.008265	.000610	-.036437
VZA	.000027	.000059	.000100	.000156	.000238	.000406	.000828	.002000	.002149	.001825
COTA	.000027	.000055	.000086	.000138	.000264	.000393	.000949	.001853	.001044	.001824
TM 5000										
E.MV.	.049997	.099955	.149947	.199961	.250094	.300013	.350262	.401611	.455056	.475580
SESGO	-.000003	-.000045	-.000053	-.000039	.000094	.000013	.000262	.001611	.005056	-.024420
VZA	.000006	.000017	.000032	.000048	.000047	.000130	.000214	.000465	.001126	.000814
COTA	.000006	.000009	.000009	.000017	.000077	.000076	.000323	.000357	.000040	.001277
TM 10000										
E.MV.	.049985	.099888	.149852	.199846	.250047	.299738	.349815	.400359	.454434	.479457
SESGO	-.000015	-.000112	-.000148	-.000154	.000047	-.000262	-.000185	.000360	.004434	-.020543
VZA	.000004	.000018	.000034	.000051	.000023	.000145	.000216	.000358	.000927	.000575
COTA	.000003	.000003	.000001	.000004	.000057	.000038	.000288	.000165	.000076	.001293

APENDICE III

COMPORTAMIENTO DE VARIOS ESTIMADORES DEL PARAMETRO p

Se comparan el valor medio y el error cuadrático de 1000 estimaciones simuladas, en 9 situaciones diferentes, de los siguientes estimadores:

Estimador a partir de la varianza, es decir, la raiz de la ecuación

$$p \cdot q = \frac{1}{n(N-1)} \quad (x_i - \bar{x})^2$$

que mejor se ajusta a la media muestral.

Estimador del método de los momentos, corregido por el anterior cuando sea inadmisible.

Estimador de máxima verosimilitud a partir de la distribución normal que tenga igual media y varianza que distribución estimada.

Comparacion entre varios estimadores del parametro p
de la distribucion binomial metrica B(n,d,p)

n = 5 T.m. = 10 N.r. =1000

p :		.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
D = 0											
E.VAR	V.M	.052493	.107782	.167398	.223813	.271319	.308935	.337866	.368101	.418545	.485529
	E.C	.001609	.005890	.012383	.018910	.021906	.022463	.023279	.027195	.033124	.038308
E.MOM	V.M	.051035	.100734	.151514	.201033	.249413	.300872	.350918	.401256	.449816	.500103
	E.C	.000947	.001872	.002410	.003082	.003762	.004144	.004455	.004767	.004868	.004884
E.NOR	V.M	.048039	.096337	.146581	.196204	.245046	.297052	.347974	.399085	.448576	.499708
	E.C	.000950	.002029	.002711	.003556	.004251	.004609	.004848	.005115	.005162	.005111
D = 1											
E.VAR	V.M	.052217	.105404	.164471	.219872	.263032	.307951	.352705	.400451	.459149	.520864
	E.C	.001971	.005754	.012841	.019292	.022608	.027797	.034062	.040097	.047539	.052418
E.MOM	V.M	.056795	.101960	.150996	.202078	.250213	.298226	.346026	.394383	.443346	.490844
	E.C	.002012	.004147	.006635	.008698	.010402	.011580	.012134	.012744	.012856	.012709
E.NOR	V.M	.049261	.096450	.146694	.197482	.245342	.293706	.341875	.391521	.442615	.494035
	E.C	.001251	.002720	.004781	.006847	.008682	.010610	.012363	.013883	.014875	.015872
D = 2											
E.VAR	V.M	.054765	.130696	.226402	.301643	.361177	.407475	.441594	.464390	.475743	.490049
	E.C	.004482	.024194	.055077	.068280	.072353	.069089	.064915	.059951	.058885	.065352
E.MOM	V.M	.095150	.159486	.218100	.260892	.301651	.335824	.367532	.390368	.412601	.428198
	E.C	.011761	.022873	.033767	.041369	.049080	.052797	.056587	.061243	.065318	.069684
E.NOR	V.M	.052565	.115633	.190413	.258977	.312041	.354979	.389700	.417793	.436513	.449369
	E.C	.003168	.014431	.034115	.049623	.056443	.058010	.059067	.060303	.063778	.072776

Comparacion entre varios estimadores del parametro p
de la distribucion binomial metrica B(n,d,p)

n = 10 T.m. = 10 N.r. =1000

p :		.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
D = 0											
E.VAR	V.M	.048995	.101355	.157965	.213834	.263734	.298857	.326055	.344569	.395086	.486659
	E.C	.001074	.004273	.010248	.016376	.021469	.023077	.023512	.027879	.036088	.044663
E.MOM	V.M	.048725	.098703	.147394	.197383	.247424	.297123	.346787	.396959	.446658	.496933
	E.C	.000477	.000942	.001286	.001641	.001922	.002173	.002347	.002500	.002547	.002634
E.NOR	V.M	.045963	.095086	.143956	.194341	.244704	.294908	.345099	.395688	.446035	.496836
	E.C	.000517	.001040	.001448	.001794	.002067	.002321	.002463	.002626	.002649	.002721
D = 1											
E.VAR	V.M	.051819	.106057	.161507	.218841	.263340	.300043	.322489	.349694	.399829	.495611
	E.C	.001225	.004821	.011830	.018418	.022481	.023933	.024823	.029173	.038575	.045313
E.MOM	V.M	.050588	.100882	.150157	.200982	.250244	.300880	.350605	.400685	.451092	.500524
	E.C	.000668	.001352	.001820	.002369	.002752	.003131	.003319	.003565	.003615	.003640
E.NOR	V.M	.048084	.097131	.145418	.196899	.246290	.297661	.348108	.398983	.450297	.500489
	E.C	.000592	.001295	.001889	.002458	.002880	.003296	.003460	.003723	.003739	.003777
D = 2											
E.VAR	V.M	.051001	.105706	.162842	.219435	.263081	.302597	.338845	.370217	.432766	.498027
	E.C	.001114	.004835	.011820	.019211	.022148	.024320	.029041	.034371	.041560	.045822
E.MOM	V.M	.052880	.099605	.148826	.198495	.248010	.299562	.348966	.397937	.448911	.498074
	E.C	.001119	.002369	.003429	.004309	.005164	.005743	.006245	.006652	.006793	.006811
E.NOR	V.M	.048607	.096455	.144619	.194286	.243581	.295792	.345832	.395453	.447654	.498326
	E.C	.000707	.001766	.002836	.003902	.004959	.005747	.006447	.007045	.007235	.007360
D = 3											
E.VAR	V.M	.050483	.105979	.165998	.225162	.272419	.317171	.356657	.398336	.441643	.510958
	E.C	.001123	.004575	.012410	.021119	.026002	.030120	.036298	.040994	.044435	.047552
E.MOM	V.M	.059125	.109789	.158326	.206511	.256340	.307424	.356600	.405607	.454080	.502650
	E.C	.001987	.004202	.006701	.008479	.010352	.012229	.013243	.013721	.013769	.014332
E.NOR	V.M	.048857	.099746	.150283	.199599	.250591	.303489	.352755	.404041	.454767	.505685
	E.C	.000830	.002340	.004221	.006301	.008574	.011221	.013230	.014572	.015460	.016603
D = 4											
E.VAR	V.M	.050753	.108017	.180306	.263959	.326716	.374415	.417378	.447199	.476447	.504440
	E.C	.001080	.006578	.023924	.045445	.051535	.052816	.052179	.049585	.047975	.050656
E.MOM	V.M	.079250	.136555	.185252	.230816	.270955	.312195	.354771	.395023	.431114	.469715
	E.C	.006497	.012157	.019490	.025926	.032838	.038559	.045264	.050214	.053305	.053140
E.NOR	V.M	.050127	.103070	.164046	.229244	.282588	.329930	.372527	.412803	.448506	.484989
	E.C	.000972	.004476	.015711	.028940	.034620	.039906	.043033	.045561	.047548	.051030
D = 5											
E.VAR	V.M	.051146	.104602	.161351	.216457	.263849	.295841	.314842	.324509	.319361	.299778
	E.C	.001206	.005115	.011581	.017764	.021110	.022376	.024126	.028342	.039627	.062179
E.MOM	V.M	.055696	.105176	.161294	.218478	.267687	.299661	.318692	.329395	.324828	.300946
	E.C	.002942	.004827	.010808	.017478	.021088	.021625	.023081	.026783	.037427	.060252
E.NOR	V.M	.055696	.105176	.161294	.218478	.267687	.299661	.318692	.329395	.324828	.300946
	E.C	.002942	.004827	.010808	.017478	.021088	.021625	.023081	.026783	.037427	.060252

Comparacion entre varios estimadores del parametro p
de la distribucion binomial metrica B(n,d,p)

n = 20 T.m. = 10 N.r. =1000

p :		.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
D = 0											
E.VAR	V.M	.048330	.101458	.157005	.209179	.255366	.288620	.312776	.329503	.360372	.491619
	E.C	.000868	.004276	.010946	.017132	.021107	.023367	.025122	.028528	.037960	.048077
E.MOM	V.M	.049235	.099188	.149029	.199048	.248933	.298744	.348855	.398986	.449042	.499007
	E.C	.000232	.000427	.000607	.000764	.000881	.001008	.001097	.001173	.001211	.001233
E.NOR	V.M	.046926	.096837	.146805	.197120	.247335	.297480	.347877	.398307	.448688	.498988
	E.C	.000275	.000492	.000673	.000826	.000929	.001042	.001126	.001192	.001230	.001246
D = 2											
E.VAR	V.M	.050884	.104767	.163450	.217737	.261847	.296062	.315910	.336668	.373896	.486984
	E.C	.001003	.004821	.012299	.018917	.022652	.023919	.025074	.029072	.038728	.047276
E.MOM	V.M	.050355	.100170	.150195	.200094	.250219	.299800	.350356	.399911	.449760	.500219
	E.C	.000398	.000754	.001081	.001342	.001565	.001754	.001921	.002036	.002063	.002095
E.NOR	V.M	.048233	.097496	.147579	.197765	.248132	.298104	.348975	.398940	.449237	.500117
	E.C	.000382	.000773	.001132	.001413	.001637	.001829	.001993	.002106	.002116	.002144
D = 4											
E.VAR	V.M	.050709	.103569	.162655	.218703	.263836	.299578	.324846	.354093	.403908	.500596
	E.C	.001019	.004011	.011057	.017520	.021504	.022752	.024150	.028758	.039007	.043650
E.MOM	V.M	.050375	.100232	.150594	.199985	.250719	.300302	.350896	.400361	.450388	.500810
	E.C	.000598	.001185	.001690	.002105	.002433	.002800	.003027	.003234	.003288	.003342
E.NOR	V.M	.048039	.097024	.147208	.196852	.247577	.297825	.348933	.399115	.449719	.500861
	E.C	.000430	.000997	.001549	.002034	.002419	.002804	.003077	.003316	.003374	.003440
D = 6											
E.VAR	V.M	.049754	.100855	.158105	.213402	.253793	.293458	.327292	.369995	.427415	.499192
	E.C	.000857	.003961	.010808	.017426	.020099	.023996	.027766	.036035	.043719	.046781
E.MOM	V.M	.054458	.102125	.150578	.200354	.249707	.299393	.350155	.399362	.449679	.500300
	E.C	.001137	.002601	.003966	.005041	.006146	.006797	.007464	.007821	.008099	.008204
E.NOR	V.M	.048388	.096211	.145376	.195434	.244764	.294980	.346379	.396642	.448385	.500428
	E.C	.000549	.001582	.002792	.004012	.005344	.006376	.007377	.008019	.008475	.008636
D = 8											
E.VAR	V.M	.052216	.110352	.171661	.234478	.295391	.351984	.398322	.437748	.471202	.498302
	E.C	.000937	.005836	.014750	.024880	.035072	.042687	.046451	.046412	.045471	.045126
E.MOM	V.M	.072097	.122058	.170477	.215959	.262544	.310707	.359272	.407584	.455479	.504197
	E.C	.003487	.007382	.011171	.015266	.018914	.022584	.025054	.027484	.029521	.029546
E.NOR	V.M	.051267	.103502	.155299	.208930	.261558	.315240	.366163	.415127	.459861	.504687
	E.C	.000709	.002817	.006288	.010864	.015981	.021732	.026500	.029240	.030997	.031150
D =10											
E.VAR	V.M	.051014	.104605	.161743	.218052	.263010	.299267	.323564	.337764	.344738	.344801
	E.C	.000877	.004303	.010835	.018285	.022209	.023856	.024384	.027077	.033972	.046808
E.MOM	V.M	.051280	.103408	.160437	.217288	.263469	.303670	.331015	.345501	.351998	.352242
	E.C	.000967	.003512	.009558	.016720	.020457	.022780	.023265	.025281	.031430	.043541
E.NOR	V.M	.051280	.103408	.160437	.217288	.263469	.303670	.331015	.345501	.351998	.352242
	E.C	.000967	.003512	.009558	.016720	.020457	.022780	.023265	.025281	.031430	.043541

Comparacion entre varios estimadores del parametro p
de la distribucion binomial metrica B(n,d,p)

n = 40 T.a. = 10 N.r. =1000

p :		.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
D = 0											
E.VAR	V.M	.051748	.105606	.166209	.223423	.270026	.305120	.327820	.344813	.363077	.518053
	E.C	.000834	.003801	.011089	.018499	.021923	.022510	.023055	.025240	.031961	.042245
E.MOM	V.M	.050067	.100261	.150173	.200108	.250316	.300324	.350303	.400238	.450180	.500199
	E.C	.000118	.000218	.000312	.000388	.000454	.000528	.000576	.000620	.000650	.000662
E.NOR	V.M	.049212	.099372	.149446	.199484	.249807	.299962	.350204	.400074	.450092	.500214
	E.C	.000127	.000226	.000317	.000390	.000455	.000530	.000580	.000619	.000654	.000670
D = 4											
E.VAR	V.M	.051866	.106055	.165551	.222699	.265725	.298793	.323758	.340079	.365039	.500644
	E.C	.000818	.004192	.011666	.019207	.022319	.022975	.024544	.027193	.034735	.045286
E.MOM	V.M	.050683	.100835	.150894	.200863	.251257	.301109	.351388	.401110	.451400	.501301
	E.C	.000182	.000338	.000480	.000601	.000701	.000800	.000867	.000927	.000968	.000978
E.NOR	V.M	.049426	.099434	.149609	.199780	.250325	.300467	.350885	.400722	.451202	.501297
	E.C	.000182	.000348	.000490	.000611	.000709	.000815	.000873	.000943	.000982	.000993
D = 8											
E.VAR	V.M	.050268	.103067	.163249	.218232	.263565	.296156	.319059	.335590	.380081	.497700
	E.C	.000785	.003885	.011921	.019356	.023424	.024840	.026006	.029515	.039943	.048471
E.MOM	V.M	.049119	.098994	.148756	.198627	.248518	.298318	.348324	.398076	.448224	.498281
	E.C	.000322	.000602	.000853	.001074	.001227	.001393	.001531	.001629	.001696	.001728
E.NOR	V.M	.047718	.096919	.146761	.196703	.246942	.296992	.347292	.397371	.447911	.498325
	E.C	.000262	.000552	.000822	.001054	.001212	.001389	.001540	.001641	.001708	.001748
D =12											
E.VAR	V.M	.051288	.105556	.163945	.220685	.266052	.298362	.324149	.355997	.408220	.498033
	E.C	.000831	.004252	.011555	.019430	.022948	.023953	.026043	.031895	.041008	.046373
E.MOM	V.M	.050473	.099450	.149645	.199595	.249426	.299651	.349407	.399237	.449232	.499333
	E.C	.000680	.001403	.001986	.002481	.002940	.003289	.003594	.003810	.003912	.003924
E.NOR	V.M	.048746	.097253	.146472	.196388	.246532	.296976	.347186	.397622	.448313	.499125
	E.C	.000401	.001056	.001734	.002326	.002900	.003345	.003725	.003981	.004094	.004091
D =16											
E.VAR	V.M	.052341	.109030	.170012	.224512	.275762	.319490	.365730	.406781	.467188	.512146
	E.C	.000912	.005105	.013197	.020024	.026323	.030914	.036983	.041624	.045631	.045958
E.MOM	V.M	.063664	.112774	.158846	.205795	.254065	.303023	.352359	.402160	.452173	.502088
	E.C	.001991	.004292	.006800	.009660	.011570	.013525	.014983	.015891	.016260	.016680
E.NOR	V.M	.051137	.101839	.151791	.201829	.251960	.302267	.352599	.402431	.452454	.502463
	E.C	.000634	.002043	.004008	.006508	.009005	.011537	.013830	.015618	.016671	.017295
D =20											
E.VAR	V.M	.050348	.103835	.164067	.218340	.262767	.295944	.318468	.333179	.341853	.344163
	E.C	.000784	.003851	.012005	.018925	.022469	.024080	.025096	.028320	.035433	.047919
E.MOM	V.M	.050658	.104080	.164587	.220894	.265908	.300093	.322355	.338491	.347695	.350580
	E.C	.000724	.003459	.010985	.018637	.021908	.023066	.023510	.026367	.032863	.044746
E.NOR	V.M	.050658	.104080	.164587	.220894	.265908	.300093	.322355	.338491	.347695	.350580
	E.C	.000724	.003459	.010985	.018637	.021908	.023066	.023510	.026367	.032863	.044746

Comparacion entre varios estimadores del parametro p
de la distribucion binomial metrica B(n,d,p)

n = 40 T.m. = 10 N.r. =1000

p :		.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
D =15											
E.VAR	V.M	.051315	.106156	.164114	.222210	.271397	.313815	.353283	.398240	.454390	.515438
	E.C	.000806	.004101	.010583	.018324	.023892	.027753	.032047	.038088	.042763	.044623
E.MOM	V.M	.056924	.104398	.153917	.203679	.253613	.303565	.353811	.403893	.454057	.503980
	E.C	.001368	.002992	.004609	.005948	.007073	.008068	.008622	.009144	.009503	.009558
E.NOR	V.M	.049850	.099494	.149453	.199679	.249802	.300622	.351614	.402477	.453760	.504672
	E.C	.000500	.001503	.002818	.004342	.005910	.007442	.008511	.009584	.010184	.010384
D =16											
E.VAR	V.M	.050050	.102005	.161738	.217962	.261440	.309942	.360221	.410909	.459810	.509862
	E.C	.000706	.003453	.010935	.018319	.021767	.028438	.035452	.041449	.044783	.045401
E.MOM	V.M	.060609	.108460	.154606	.203103	.251594	.301608	.350278	.400484	.450521	.500675
	E.C	.001782	.003996	.006387	.008424	.010505	.011868	.013250	.013846	.014501	.014757
E.NOR	V.M	.048872	.097568	.147185	.197025	.246576	.297699	.347803	.399133	.450035	.501643
	E.C	.000503	.001694	.003494	.005539	.008109	.010419	.012684	.014427	.015537	.016059
D =17											
E.VAR	V.M	.050077	.104083	.160819	.223997	.284914	.339844	.386504	.426045	.460822	.511075
	E.C	.000710	.003876	.011120	.022786	.033951	.041059	.044207	.045493	.045955	.046388
E.MOM	V.M	.069407	.120340	.166645	.212954	.259257	.305745	.354961	.402603	.451340	.501769
	E.C	.003021	.006486	.009927	.013815	.017424	.020254	.023346	.025105	.026627	.026963
E.NOR	V.M	.049299	.099823	.149249	.202057	.254971	.305268	.356562	.403608	.450989	.501328
	E.C	.000573	.002191	.004904	.009282	.014747	.019520	.023423	.026031	.027571	.027871
D =18											
E.VAR	V.M	.049850	.104826	.173905	.252339	.323648	.380612	.416730	.461399	.479774	.502886
	E.C	.000753	.005173	.020692	.040051	.053178	.057614	.054038	.053923	.048740	.046800
E.MOM	V.M	.083251	.139300	.186475	.233134	.278222	.322360	.369160	.415338	.457234	.500330
	E.C	.005887	.011822	.017867	.023477	.029507	.035385	.039272	.043932	.047031	.047472
E.NOR	V.M	.049059	.100464	.158410	.222647	.281940	.334471	.378954	.428558	.462878	.500223
	E.C	.000647	.003444	.012470	.025550	.035832	.041741	.044194	.047466	.046796	.046099
D =19											
E.VAR	V.M	.066092	.183762	.288133	.358964	.411254	.442379	.462110	.480558	.500253	.510639
	E.C	.013663	.066448	.097407	.101088	.094828	.080596	.065867	.055292	.048967	.045495
E.MOM	V.M	.135322	.212121	.275201	.329601	.372628	.410288	.437947	.465022	.484890	.505700
	E.C	.023469	.045535	.060664	.074610	.080471	.083659	.083087	.080512	.080186	.079025
E.NOR	V.M	.064710	.175127	.268557	.334086	.385825	.423130	.446125	.472025	.495343	.512444
	E.C	.013048	.062603	.090403	.095885	.093427	.085023	.073646	.066450	.061466	.058217
D =20											
E.VAR	V.M	.050820	.105488	.164167	.217548	.265496	.297250	.319055	.334390	.342883	.345217
	E.C	.000842	.004310	.011974	.018569	.023087	.024156	.024810	.027902	.034739	.047111
E.MOM	V.M	.050582	.104997	.162196	.218253	.266241	.298731	.320866	.335388	.344231	.346864
	E.C	.000786	.004155	.010801	.018473	.022570	.023688	.024353	.027051	.033687	.045873
E.NOR	V.M	.050582	.104997	.162196	.218253	.266241	.298731	.320866	.335388	.344231	.346864
	E.C	.000786	.004155	.010801	.018473	.022570	.023688	.024353	.027051	.033687	.045873

Comparacion entre varios estimadores del parametro p
de la distribucion binomial metrica B(n,d,p)

n = 40 T.m. = 20 N.r. =1000

p :		.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
D =15											
E.VAR	V.M	.051065	.103458	.158558	.218041	.270606	.310941	.342810	.387883	.446137	.512204
	E.C	.000409	.001914	.005599	.011915	.016326	.017951	.018563	.024253	.030264	.032021
E.MOM	V.M	.054099	.102405	.152608	.202802	.252518	.302701	.352973	.403177	.453245	.503391
	E.C	.000778	.001765	.002532	.003283	.003831	.004401	.004675	.005022	.005164	.005234
E.NOR	V.M	.050457	.100476	.150348	.200941	.250798	.301136	.351860	.402574	.453297	.503832
	E.C	.000273	.000823	.001500	.002278	.003018	.003822	.004370	.004928	.005231	.005383
D =16											
E.VAR	V.M	.051419	.104152	.161300	.222531	.275514	.319173	.360442	.405207	.455568	.502910
	E.C	.000359	.001724	.005551	.011979	.016812	.018592	.020907	.025196	.028877	.029761
E.MOM	V.M	.055801	.103935	.150727	.199368	.249344	.299650	.349188	.399157	.449479	.498906
	E.C	.001123	.002475	.003928	.005274	.006229	.007077	.007570	.007971	.008269	.008435
E.NOR	V.M	.050799	.101098	.150958	.201102	.250313	.300515	.350000	.399822	.449944	.499132
	E.C	.000290	.001008	.001975	.003115	.004392	.005653	.006719	.007581	.008206	.008525
D =17											
E.VAR	V.M	.049976	.101552	.155833	.213028	.271607	.324045	.368528	.411917	.449823	.495034
	E.C	.000345	.001627	.004786	.010260	.017264	.022903	.025945	.028840	.030684	.031499
E.MOM	V.M	.060673	.106341	.152796	.198555	.248362	.296564	.346043	.396227	.445148	.495100
	E.C	.001558	.003544	.005781	.007869	.009495	.011060	.012101	.012979	.013570	.013685
E.NOR	V.M	.049508	.099130	.148881	.197978	.248608	.298352	.347828	.397851	.445492	.494699
	E.C	.000299	.001065	.002304	.004064	.006154	.008689	.010739	.012471	.013658	.013864
D =18											
E.VAR	V.M	.049667	.100351	.154327	.221036	.294571	.356962	.402735	.441675	.480206	.510566
	E.C	.000341	.001507	.005166	.016301	.029850	.037175	.037593	.035239	.034527	.032492
E.MOM	V.M	.073449	.125061	.172302	.217649	.264760	.308553	.355642	.404984	.452625	.502956
	E.C	.003376	.007092	.010691	.014531	.018842	.022933	.025509	.028007	.029999	.030706
E.NOR	V.M	.049541	.099618	.150601	.206766	.267487	.320421	.368188	.416416	.461425	.504815
	E.C	.000308	.001153	.003167	.008348	.015930	.020997	.024166	.025933	.027396	.027492
D =19											
E.VAR	V.M	.056122	.129022	.223779	.316254	.388479	.438305	.463889	.479855	.496156	.509258
	E.C	.004205	.021124	.048416	.067890	.071733	.064031	.050788	.039188	.032127	.028744
E.MOM	V.M	.110110	.177384	.235724	.289001	.335110	.375434	.413626	.445816	.482094	.513716
	E.C	.012971	.025092	.036704	.045099	.052798	.058914	.061907	.064031	.065061	.065089
E.NOR	V.M	.055785	.125908	.211005	.291105	.353986	.405794	.438343	.463261	.486635	.510818
	E.C	.004005	.019463	.042229	.057902	.061579	.059234	.051725	.044548	.039487	.037347
D =20											
E.VAR	V.M	.050585	.101953	.157706	.214816	.270114	.314292	.346670	.365324	.375939	.379121
	E.C	.000363	.001577	.005034	.010476	.015837	.017892	.018445	.019036	.022721	.031625
E.MOM	V.M	.050664	.101993	.157538	.214335	.270860	.315888	.348283	.368010	.378688	.382588
	E.C	.000342	.001479	.004728	.009796	.015432	.017374	.017679	.018230	.021625	.030178
E.NOR	V.M	.050664	.101993	.157538	.214335	.270860	.315888	.348283	.368010	.378688	.382588
	E.C	.000342	.001479	.004728	.009796	.015432	.017374	.017679	.018230	.021625	.030178

Comparacion entre varios estimadores del parametro p
de la distribucion binomial metrica B(n,d,p)

n = 40 T.m. = 40 N.r. =1000

p :		.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
D =15											
E.VAR	V.M	.051404	.103359	.157398	.214412	.277093	.330168	.368524	.399936	.441824	.496900
	E.C	.000168	.000690	.001920	.004918	.010874	.013645	.013019	.013116	.016398	.018426
E.MOM	V.M	.050978	.100696	.150743	.200643	.251156	.300762	.350799	.400772	.450677	.500899
	E.C	.000464	.000930	.001296	.001638	.001910	.002151	.002336	.002486	.002590	.002601
E.NOR	V.M	.050916	.101345	.151732	.201590	.251906	.301404	.351281	.401107	.450799	.500803
	E.C	.000135	.000407	.000771	.001143	.001552	.001900	.002175	.002436	.002577	.002605
D =16											
E.VAR	V.M	.049576	.099244	.150676	.204470	.262060	.312921	.353708	.389018	.438183	.505657
	E.C	.000157	.000653	.001792	.004498	.009151	.012367	.013610	.015640	.020020	.023131
E.MOM	V.M	.050796	.099582	.148816	.198816	.248788	.298531	.348937	.398421	.448483	.498606
	E.C	.000628	.001296	.001930	.002416	.002843	.003156	.003466	.003602	.003824	.003815
E.NOR	V.M	.049291	.098377	.147995	.197564	.247534	.297068	.347635	.397457	.448054	.498709
	E.C	.000131	.000438	.000901	.001441	.002031	.002588	.003192	.003546	.003894	.003966
D =17											
E.VAR	V.M	.050907	.102400	.154854	.211621	.270676	.325256	.370737	.418935	.459400	.504828
	E.C	.000176	.000765	.002050	.005481	.010668	.014656	.016513	.019693	.020947	.021294
E.MOM	V.M	.055548	.103481	.151946	.201770	.252018	.301793	.351995	.401789	.452030	.501979
	E.C	.001061	.002463	.003745	.004722	.005549	.006308	.006782	.007186	.007459	.007504
E.NOR	V.M	.050798	.101678	.152257	.202996	.253541	.304043	.354417	.403642	.452816	.501647
	E.C	.000161	.000594	.001297	.002256	.003338	.004654	.005809	.006851	.007511	.007586
D =18											
E.VAR	V.M	.050584	.101898	.154469	.213626	.275831	.339612	.389703	.434360	.466580	.497554
	E.C	.000184	.000768	.002128	.006705	.013588	.020315	.022536	.023100	.021766	.021331
E.MOM	V.M	.061892	.111040	.157628	.204546	.251776	.301583	.350269	.400649	.450703	.500125
	E.C	.001747	.004019	.006264	.008708	.010536	.012577	.014033	.014741	.015212	.015490
E.NOR	V.M	.050356	.101164	.152095	.205642	.257519	.311458	.361752	.410942	.455532	.499739
	E.C	.000169	.000637	.001464	.003561	.006176	.009338	.011751	.013517	.014131	.014357
D =19											
E.VAR	V.M	.050419	.107167	.171422	.245009	.325240	.384950	.433536	.465308	.491893	.507948
	E.C	.000154	.005107	.014247	.026634	.038019	.038190	.034842	.028592	.023456	.020725
E.MOM	V.M	.090057	.148029	.198625	.247855	.294853	.340216	.385829	.426543	.470334	.512979
	E.C	.006263	.012189	.018188	.024555	.030407	.035206	.038962	.042508	.044451	.044625
E.NOR	V.M	.050365	.104275	.167299	.236235	.305164	.360285	.408364	.448229	.482335	.509718
	E.C	.000150	.002927	.011949	.023031	.031058	.032416	.031550	.029263	.026637	.024890
D =20											
E.VAR	V.M	.050632	.101517	.154212	.210350	.269153	.320837	.360415	.385183	.400142	.404289
	E.C	.000166	.000704	.001916	.005146	.009925	.013194	.013271	.012551	.014302	.020684
E.MOM	V.M	.050634	.101529	.154221	.209997	.268942	.321308	.360707	.386141	.400782	.404647
	E.C	.000162	.000681	.001844	.004870	.009687	.013113	.013060	.012390	.014060	.020411
E.NOR	V.M	.050634	.101529	.154221	.209997	.268942	.321308	.360707	.386141	.400782	.404647
	E.C	.000162	.000681	.001844	.004870	.009687	.013113	.013060	.012390	.014060	.020411

Comparacion entre varios estimadores del parametro p
de la distribucion binomial metrica B(n,d,p)

n = 80 T.m. = 10 N.r. =1000

p :		.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
D = 0											
E.VAR	V.M	.050264	.103443	.162032	.216437	.261738	.297270	.320915	.336515	.346102	.497821
	E.C	.000663	.003650	.011040	.017386	.021748	.023808	.024853	.027932	.034544	.046319
E.MOM	V.M	.049894	.099931	.149940	.199955	.249951	.299961	.349901	.399909	.449881	.499885
	E.C	.000061	.000117	.000165	.000205	.000242	.000283	.000317	.000346	.000364	.000380
E.NOR	V.M	.049075	.099277	.149437	.199548	.249613	.299686	.349692	.400100	.449807	.499889
	E.C	.000066	.000122	.000166	.000206	.000243	.000283	.000318	.000354	.000369	.000379
D = 8											
E.VAR	V.M	.048808	.100336	.156307	.208049	.250616	.283312	.308862	.326133	.338314	.496183
	E.C	.000645	.003388	.010470	.016504	.020222	.021969	.024660	.028306	.035980	.049140
E.MOM	V.M	.049913	.099974	.149912	.199912	.249816	.299874	.349889	.399908	.449872	.499813
	E.C	.000094	.000179	.000256	.000319	.000368	.000427	.000477	.000508	.000531	.000549
E.NOR	V.M	.048991	.099062	.149069	.199147	.249157	.299350	.349537	.399703	.449736	.499804
	E.C	.000093	.000180	.000257	.000320	.000370	.000430	.000482	.000514	.000542	.000562
D =16											
E.VAR	V.M	.053186	.109790	.173171	.232374	.280244	.315234	.339079	.354051	.370938	.500331
	E.C	.000765	.004046	.012077	.019611	.022659	.022981	.022774	.024115	.030056	.039745
E.MOM	V.M	.049341	.099098	.148927	.198910	.248662	.298561	.348559	.398393	.448433	.498374
	E.C	.000156	.000300	.000422	.000531	.000618	.000705	.000776	.000828	.000859	.000884
E.NOR	V.M	.048888	.098530	.148401	.198407	.248240	.298301	.348372	.398208	.448339	.498378
	E.C	.000138	.000284	.000409	.000519	.000611	.000707	.000771	.000831	.000862	.000889
D =24											
E.VAR	V.M	.050094	.103407	.161020	.215853	.260020	.292365	.316053	.332811	.375316	.498174
	E.C	.000731	.003887	.011127	.018255	.021980	.023495	.024676	.028082	.038427	.047691
E.MOM	V.M	.049217	.098957	.148801	.198591	.248501	.298174	.348368	.398050	.448085	.498234
	E.C	.000349	.000679	.000968	.001209	.001406	.001593	.001751	.001844	.001906	.001942
E.NOR	V.M	.047775	.096922	.146578	.196587	.246671	.296599	.347117	.397190	.447627	.498190
	E.C	.000253	.000586	.000904	.001174	.001392	.001598	.001775	.001879	.001944	.001982
D =32											
E.VAR	V.M	.050985	.106261	.165040	.221469	.266746	.300991	.336520	.380322	.434345	.507801
	E.C	.000801	.004563	.011768	.019477	.023422	.025687	.029805	.037053	.044000	.047137
E.MOM	V.M	.055596	.101536	.149877	.199060	.248945	.298905	.348570	.398643	.448837	.498787
	E.C	.001092	.002585	.004011	.005224	.006179	.006912	.007492	.007957	.008139	.008288
E.NOR	V.M	.049343	.098229	.147252	.196467	.246331	.296176	.346366	.397188	.448082	.498905
	E.C	.000459	.001405	.002588	.003903	.005151	.006330	.007304	.008057	.008486	.008718
D =40											
E.VAR	V.M	.050313	.103472	.162463	.217638	.263967	.299992	.324925	.341417	.349877	.352215
	E.C	.000620	.003383	.010372	.017185	.021170	.022922	.023635	.025986	.032423	.044058
E.MOM	V.M	.050184	.103070	.161334	.217500	.263787	.300646	.325677	.344247	.353054	.355418
	E.C	.000571	.003066	.009526	.016407	.020127	.021894	.022440	.024772	.030873	.042165
E.NOR	V.M	.050184	.103070	.161334	.217500	.263787	.300646	.325677	.344247	.353054	.355418
	E.C	.000571	.003066	.009526	.016407	.020127	.021894	.022440	.024772	.030873	.042165

Comparacion entre varios estimadores del parametro p
de la distribucion binomial metrica B(n,d,p)

n = 80 T.m. = 10 N.r. =1000

p :		.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
D =35											
E.VAR	V.M	.051665	.106471	.167648	.225609	.277414	.332806	.377394	.414855	.454507	.504879
	E.C	.000747	.003984	.012117	.020106	.026716	.035732	.039512	.041811	.043357	.044683
E.MOM	V.M	.065276	.113932	.159170	.205783	.254625	.301967	.351018	.400643	.450479	.500241
	E.C	.002188	.004828	.007763	.010661	.012886	.015474	.017318	.018472	.019348	.019556
E.NOR	V.M	.050414	.100677	.150840	.201370	.251756	.302079	.352020	.401156	.450424	.499809
	E.C	.000535	.001942	.004166	.007017	.010277	.013857	.016858	.018883	.020282	.020641
D =36											
E.VAR	V.M	.049940	.102154	.161326	.225413	.290402	.341347	.390313	.421087	.462772	.497808
	E.C	.000649	.003252	.010861	.022614	.035042	.041444	.045528	.044766	.046239	.046724
E.MOM	V.M	.070644	.120617	.165974	.209330	.255854	.302389	.349365	.397053	.446737	.495250
	E.C	.003132	.006534	.010408	.014800	.018891	.022249	.025277	.027979	.029298	.030263
E.NOR	V.M	.049129	.098434	.149017	.201270	.254571	.304293	.353398	.400501	.448317	.493501
	E.C	.000515	.002040	.004991	.010149	.016053	.021085	.025286	.028628	.030581	.031650
D =37											
E.VAR	V.M	.050397	.106449	.173610	.251900	.316498	.365754	.407379	.437918	.467418	.489903
	E.C	.000639	.006011	.018485	.037657	.047672	.050335	.049368	.046769	.045120	.044825
E.MOM	V.M	.082486	.136268	.184369	.230159	.275582	.318327	.361836	.405450	.449710	.494591
	E.C	.005599	.011362	.017009	.022943	.028507	.034809	.039797	.044423	.047309	.048166
E.NOR	V.M	.049943	.102554	.159436	.220867	.276402	.325595	.371264	.412806	.452646	.488341
	E.C	.000571	.003933	.011153	.022472	.030881	.036544	.040420	.042596	.043484	.044261
D =38											
E.VAR	V.M	.051796	.130986	.213949	.289193	.355379	.402594	.434372	.460057	.483817	.502596
	E.C	.002300	.025923	.047616	.061803	.069112	.067082	.060480	.053747	.049790	.047865
E.MOM	V.M	.101762	.165060	.219268	.268964	.311369	.347769	.384015	.419583	.452092	.485451
	E.C	.011316	.022095	.032372	.042297	.049087	.055031	.059628	.062009	.061581	.063449
E.NOR	V.M	.050954	.124016	.195151	.259727	.316265	.365288	.404217	.435405	.465668	.495165
	E.C	.002023	.022578	.040127	.052184	.059090	.062283	.061266	.058528	.057056	.056891
D =39											
E.VAR	V.M	.117260	.241375	.326262	.384639	.436294	.457459	.475638	.492303	.497932	.509629
	E.C	.060692	.112951	.125439	.118011	.107895	.086847	.070575	.058508	.049248	.045898
E.MOM	V.M	.161971	.254762	.319064	.370235	.409310	.437971	.455641	.477550	.493583	.507237
	E.C	.040917	.076886	.096283	.102495	.101183	.095277	.089008	.081323	.078069	.075214
E.NOR	V.M	.110676	.228527	.311840	.370684	.416351	.445604	.463174	.479948	.491580	.508150
	E.C	.055874	.107005	.123173	.119891	.109502	.094290	.078936	.067735	.060195	.057461
D =40											
E.VAR	V.M	.051553	.106979	.167111	.225080	.272116	.307687	.330865	.346084	.355112	.358643
	E.C	.000706	.003898	.011133	.018628	.022330	.023341	.023604	.025922	.031740	.042512
E.MOM	V.M	.051203	.105611	.164856	.223521	.270405	.307125	.331884	.348305	.358441	.361718
	E.C	.000642	.003377	.009857	.017449	.020853	.022043	.022468	.024754	.030205	.040713
E.NOR	V.M	.051203	.105611	.164856	.223521	.270405	.307125	.331884	.348305	.358441	.361718
	E.C	.000642	.003377	.009857	.017449	.020853	.022043	.022468	.024754	.030205	.040713

APENDICE IV

ESTIMACION DEL PARAMETRO d EN LA B(n, d, p)

Se tabula el valor medio y la desviación típica de 1000 estimaciones simuladas del parámetro d , en 8 situaciones diferentes, por el método de máxima verosimilitud. Asimismo se tabula el porcentaje de dominación, respecto del estimador anterior, de tres procedimientos de redondeo del estimador del método de los momentos:

- Redondeo simétrico.
- Redondeo por la verosimilitud muestral.
- Redondeo por la verosimilitud de x_i .

Estimacion del parametro d de la B(n,d,p)

Parametro n = 5
Tamaño muestral = 5
Numero de simulaciones = 1000

	PR :	.05	.10	.15	.20	.25	.30	.35	.40	.45
	0									
EMV	V.M.	.01200	.09800	.16200	.23700	.37100	.48900	.67100	.93500	1.52800
	D.T.	.10889	.29731	.38439	.46779	.61103	.73612	1.02115	1.39455	1.94299
	P.D.1	.98800	.95300	.97000	.98200	.96600	.97900	.98300	.99500	1.00000
	P.D.2	1.00000	.99800	.99900	.99900	.99900	.99800	1.00000	1.00000	1.00000
	P.D.3	.98800	.95300	.96800	.98200	.93800	.97900	.94000	.99500	.95400
	1									
EMV	V.M.	1.02800	1.05700	1.02500	1.07900	1.13100	1.12800	1.30400	1.48800	1.95200
	D.T.	.25537	.35743	.55711	.68903	.81354	1.02060	1.27890	1.64069	2.05078
	P.D.1	.99800	.94200	.96400	.92600	.93100	.95000	.97700	1.00000	1.00000
	P.D.2	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
	P.D.3	.99800	.94200	.88000	.92600	.86300	.95000	.93500	1.00000	.94900
	2									
EMV	V.M.	2.00700	2.01800	1.99100	2.05100	2.05300	2.03600	2.10600	2.19000	2.21500
	D.T.	.17015	.41675	.57352	.71722	.89230	1.11835	1.42575	1.81656	2.10732
	P.D.1	.93800	.95900	.87900	.94700	.88900	.97200	.86400	1.00000	.88700
	P.D.2	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
	P.D.3	.93800	.95900	.95300	.94700	.97900	.97200	.97600	1.00000	1.00000