

S/158

**UNIVERSIDAD DE GRANADA
FACULTAD DE CIENCIAS**

**DEPARTAMENTO DE ESTADÍSTICA E INVESTIGACIÓN
OPERATIVA**

UNIVERSIDAD DE GRANADA
Facultad de Ciencias
Fecha 10/04/01
ENTRADA NUM. 1686



UNIVERSIDAD DE GRANADA
2 / ABR. 2001
COMISION DE DOCTORADO

Tests de Hipótesis en los Ensayos Multinomiales 2x2

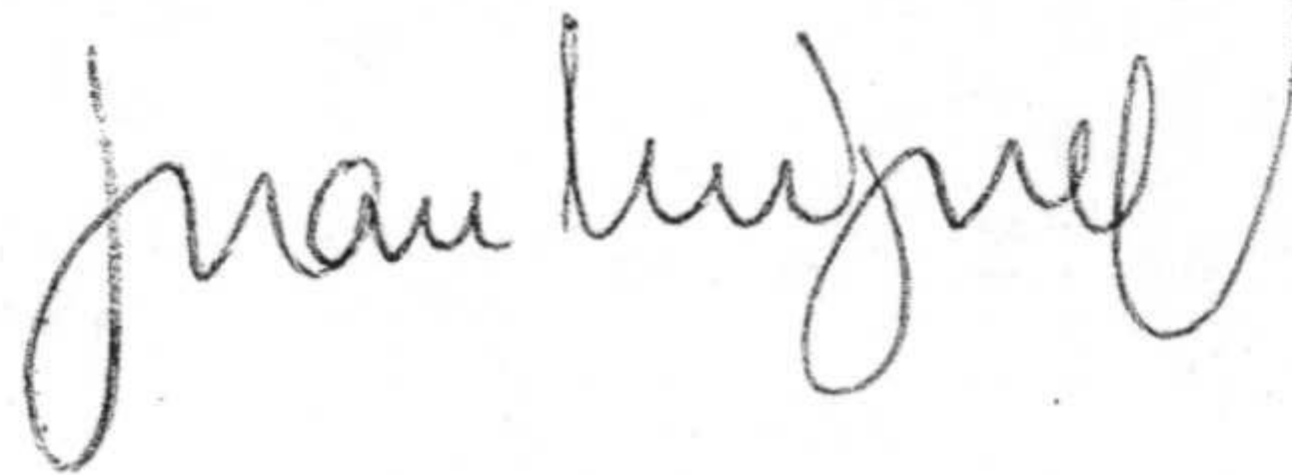
Juan Miguel Tapia García
Tesis doctoral

BIBLIOTECA UNIVERSITARIA
GRANADA
Nº Documento 61338611
Nº Copia 21557495

GRANADA - 2001

TESTS DE HIPÓTESIS EN LOS ENSAYOS
MULTINOMIALES 2×2

Trabajo presentado por Juan Miguel Tapia García ante la Universidad de Granada para optar al grado de Doctor en Ciencias en Granada a 27 de marzo de 2001.



D. ANTONIO MARTÍN ANDRÉS, Catedrático de Universidad del Área de Estadística e Investigación Operativa y perteneciente al Departamento del mismo nombre de la Universidad de Granada

CERTIFICA:

Que **D. Juan Miguel Tapia García**, Licenciado en Ciencias Matemáticas por la Universidad de Granada, ha realizado bajo mi dirección la presente Tesis Doctoral titulada "TESTS DE HIPÓTESIS EN LOS ENSAYOS MULTINOMIALES 2×2 ";

Que dicha Tesis ha sido revisada por mí, estimando que reúne todas las condiciones precisas para ser presentada ante Tribunal y hacer merecedor al autor del título de Doctor.

Granada, 27 de marzo de 2001



Prof. Dr. Antonio Martín Andrés

Validación externa de los resultados obtenidos en esta Tesis Doctoral

Gran parte de los resultados obtenidos en esta Tesis Doctoral han sido publicados en diversos ámbitos científicos. Otros están pendientes de publicación. Así:

1) Los resultados del **Capítulo II** han sido publicados en:

(a) "Sobre la determinación del valor P en los ensayos multinomiales 2×2 ". *VIª Conferencia Española de Biometría*. 21-24 Septiembre 1997. 122-131.

(b) "Sobre el método incondicionado óptimo en el test de independencia para tablas 2×2 ". *XXIVº Congreso Nacional de Estadística e Investigación Operativa*. 20-23 Octubre 1998. 169-170.

(c) 'On determining the P-value in 2×2 multinomial trials'. *Journal of Statistical Planning and Inferences* 69 (1), 33-49(1998). Erratum in 79 (1999) 365.

(d) 'Optimal unconditional test in 2×2 multinomial trials'. *Computational Statistics and Data Analysis* 31 (3), 311-321(1999).

(e) <http://www.ugr.es/~bioest/TMP.EXE>. (1999).

(f) <http://www.ugr.es/~bioest/crbind.htm> (2000).

(g) 'Optimal unconditional critical regions for 2×2 multinomial trials'. *Journal of Applied Statistics* 27(6), 689-695 (2000).

2) Los resultados del **Capítulo III** están en la actualidad sometidos a la Revista *Biometrical Journal* bajo el título: 'Optimal unconditional chi-squared test in 2×2 multinomial trials'.

3) Los resultados del **Capítulo IV** se publicarán en un artículo conjunto con Sánchez Quevedo, M.J. cuando dicha autora obtenga resultados equivalentes para el caso de 2 proporciones. Igual sucede con los resultados alusivos a la comparación entre modelos (aludidos en el **Capítulo II**).

PREÁMBULO

Las tablas 2×2 tienen una apariencia engañosamente simple que empaña su densa y larga historia. Ellas se abordan en cualquier curso introductorio de estadística (por breve que sea), y así los estudiantes y la mayoría de los profesores piensan que las mismas constituyen un problema resuelto. Pero no es así. De hecho esta memoria se dedica exclusivamente a uno de los modos en que puede surgir una tabla 2×2 : el muestreo de dos cualidades dicotómicas (a fin de estudiar la existencia o no de asociación entre ellas).

En el desarrollo de la memoria se ha procurado separar claramente aquello que es material histórico de aquello que constituyen **las aportaciones actuales (cuyos apartados se han marcado con *)**. A veces, por facilitar la exposición, ello no ha sido estrictamente posible, pero en tales casos el texto lo aclarará fehacientemente.

El **Capítulo I** (que no contiene aportaciones especiales, aparte del valor de la propia revisión del tema) enmarca al lector dentro del problema actual; le indica lo que sí y lo que no es objetivo de esta memoria; le señala su importancia, problemática y soluciones que suele admitir; le explica las razones de las discrepancias en los distintos análisis (relacionadas con el modelo asumido y/o los principios de la estadística que se aceptan); y, finalmente, le informa del tipo de soluciones que aquí se defienden y desarrollan.

El **Capítulo II** comienza con una revisión bibliográfica de todo lo conocido acerca de los métodos exactos de analizar las tablas 2×2 que son objetivo de esta memoria. El resto del capítulo consta de diversas aportaciones en las cuales se formaliza el problema; se definen las condiciones a verificar por cualquier región crítica coherente; se definen nuevos métodos para la formación de la región crítica; se evalúa el tamaño del test de un modo más corto (permitiendo así hacer utilizables en la

práctica estos métodos, que requieren de un gran tiempo de cómputo); se obtiene la potencia global de un método para hacerlo comparable con los demás; se seleccionan los tests óptimos y se proporcionan programas y tablas para ellos. El capítulo finaliza con una evaluación de cuánto de conservador es el test exacto de Fisher cuando se usa con datos que provienen del modelo actual.

El **Capítulo III** alude a los métodos asintóticos para analizar el problema, entre otros, el clásico test de la chi-cuadrado. Tras describir lo tratado al respecto por la literatura (que es escaso, pues prácticamente sólo se dedica a evaluar métodos, no a compararlos, y alude en general sólo a métodos sin corrección por continuidad, que teóricamente no son óptimos), se aborda de modo exhaustivo las distintas correcciones por continuidad posibles para cada versión de chi-cuadrado, se propone al P-mid de Fisher como solución aproximada, se indica un procedimiento coherente que permite no sólo evaluar un método, sino también seleccionar el óptimo entre varios, y se determina el óptimo de entre los 53/57 métodos propuestos para tests de 1/2 colas, indicando sus condiciones de validez y su comportamiento liberal o conservador cuando el método no las verifica.

El análisis anterior permite seleccionar el método asintótico óptimo y proporcionar unas condiciones de validez muy genéricas. Pero no justifica el por qué de tales condiciones ni proporciona unas específicas para cada caso. El **Capítulo IV** está dedicado a ello. En él se analiza cómo la validez del test chi-cuadrado está ligada a la magnitud de la asimetría del estadístico base, de modo que afirmar que el test es válido cuando E (la mínima cantidad esperada) es mayor que un determinado número (E^*), es equivalente afirmar que es válido cuando la asimetría es inferior a un número también dado. En el capítulo se comprueba que la cota E^* no es un número fijo, sino una función que depende del tamaño de muestra y del desequilibrio del marginal más raro, al tiempo que se da la forma de la misma. Finalmente se comprueba que afortunadamente E^* alcanza un máximo absoluto, lo que permite dar unas condiciones de validez generales (aunque conservadoras) muy diferentes de la clásica condición de

$E \geq 5$ - $E > 14,38$ (una cola) y $E > 13,03$ (dos colas)- y diferentes también de las del test chi-cuadrado como aproximación del test exacto de Fisher.

ÍNDICE

0. PREÁMBULO	i
ÍNDICE	
I. INTRODUCCIÓN	1
1.1. El problema y su notación	1
1.1.1. Introducción.	1
1.1.2. Modelo I (objeto de esta memoria)	2
1.1.3. Modelo II.	4
1.1.4. Modelo III	5
1.2. Soluciones al problema	6
1.2.1. El test (exacto) de Fisher	6
1.2.2. Los test (exactos) de Barnard	7
1.2.2.1. Introducción	7
1.2.2.2. Caso del Modelo II	8
1.2.2.3. Caso del Modelo I	10
1.2.3. El test (asintótico) de la chi-cuadrado.	11
1.3. Relaciones entre los tres modelos	12
1.3.1. Supuesta cierta H_0	12
1.3.1.1. Modelo I vs. Modelo II	12
1.3.1.2. Modelo II vs. Modelo III	12
1.3.1.3. Modelo I vs. Modelo III	13
1.3.2. En general	13
1.4. ¿Test condicionado o incondicionado?.	16
1.4.1. Introducción.	16
1.4.2. Los principios de suficiencia y de auxiliaridad	17
1.5. La razón del producto cruzado	20
II. TESTS EXACTOS	22
2.1. Estado actual del problema	23
2.1.1. Introducción.	23
2.1.2. Variables base para la formación del espacio muestral	24
2.1.3. Formación de la región crítica: métodos de ordenación	24
2.1.4. Notas acerca del cálculo y lugar del máximo	29
2.1.5. Valores P simplificados	30
2.1.6. Evaluación de los diversos métodos	32
2.1.7. Tablas y programas	35
2.2. Formalización del problema*	36
2.2.1. Introducción.	36

2.2.2. Construcción simplificada del espacio muestral: Equivalencia y Simetría	37
2.2.3. Condiciones a verificar por la región crítica: Convexidad	42
2.2.4. Formación de la región crítica	43
2.2.4.1. Introducción	43
2.2.4.2. Métodos de ordenación "CON memoria"	43
2.2.4.3. Métodos de ordenación "SIN memoria"	44
2.2.4.4. Atractores	46
2.2.5. Equivalencia, Simetría, Convexidad y métodos de ordenación	47
2.2.5.1. Introducción	47
2.2.5.2. En los métodos con memoria	48
2.2.5.3. En los métodos procedentes del test exacto de Fisher	48
2.2.5.4. En los métodos procedentes del estadístico chi-cuadrado	49
2.3. Simplificaciones para el cálculo del máximo*	53
2.3.1. Introducción.	53
2.3.2. Agrupación de cálculos	54
2.3.3. Reducción general del espacio paramétrico	55
2.3.4. Reducción del espacio paramétrico en el caso de α^* sucesivos	59
2.3.5. Reducción del espacio paramétrico en la región crítica inicial (atractores)	68
2.3.5.1. En el test de una cola	68
2.3.5.2. En el test de dos colas	70
2.3.6. Reducción del número de máximos a calcular con el método M.	72
2.3.7. Proceso numérico para la determinación del máximo	75
2.3.8. Atractor teórico	75
2.3.9. Discusión y conclusiones	76
2.4. Potencia y selección del método óptimo	78
2.4.1. Los procedimientos clásicos y sus restricciones	78
2.4.2. Resolución de la primera restricción: potencia a largo plazo	79
2.4.3. Resolución de la segunda restricción: potencia media	82
2.5. Selección del método incondicionado óptimo*	84
2.5.1. Introducción.	84
2.5.2. Datos a fijar de antemano	85
2.5.3. Procedimiento para la obtención de los resultados	86
2.5.4. Comentarios previos a la obtención de los resultados definitivos	87
2.5.4.1. Acerca de la RC inicial	87
2.5.4.2. Acerca de la definición de potencia a considerar	87
2.5.4.3. Acerca de la paridad de n	89
2.5.5. Resultados y su análisis	89
2.5.6. Conclusiones	97
2.6. Sobre el conservadurismo del test exacto de Fisher	98
2.6.1. Introducción.	98
2.6.2. Acotación	98

2.6.3. Sobre la coherencia de la región crítica condicionada*	99
2.6.4. Conservadurismo del condicionamiento y test UMP*	99
2.6.5. Criterio para la comparación*	100
2.6.6. Resultados y conclusiones*	101
2.7. Sobre el conservadurismo de los tests semi-óptimos*	106
2.7.1. Introducción.	106
2.7.2. Resultados y conclusiones.	107
2.8. Programas*	107
2.9. Tablas*	108
2.9.1. Introducción.	108
2.9.2. Tablas de RC por el método de Barnard (M)	109
2.10. Conclusiones*	112
III. TESTS APROXIMADOS	116
3.1. Estado actual del problema	116
3.1.1. Introducción.	116
3.1.2. Convenios	117
3.1.3. Estadísticos sin corrección por continuidad	119
3.1.3.1. Estadísticos del tipo chi-cuadrado clásico	119
3.1.3.2. Otros estadísticos de tipo chi-cuadrado	121
3.1.4. Estadísticos con corrección por continuidad	123
3.1.4.1. Introducción	123
3.1.4.2. Correcciones procedentes del método condicionado.	123
3.1.4.3. Corrección de Pirie and Hamdan.	125
3.1.4.4. Corrección de Conover	126
3.1.5. Evaluación de los diversos métodos asintóticos	127
3.1.5.1. Datos de la literatura	127
3.1.5.2. Comentarios generales*	129
3.1.5.3. Conclusiones.	130
3.1.5.4. Condiciones de validez	131
3.2. Nuevas correcciones por continuidad*	133
3.2.1. Introducción.	133
3.2.2. Amplitud del salto total y estadísticos que provoca.	135
3.2.3. Determinación del número de saltos de \bar{n}	137
3.2.3.1. Introducción	137
3.2.3.2. Número de tablas que alcanzan el mínimo (cero)	138
3.2.3.3. Caso de n impar.	139
3.2.3.4. Caso de n par.	140
3.2.3.5. Número de saltos para los estadísticos U y chi-cuadrado	141
3.2.3.6. Discusión y propuestas	142
3.3. El P-mid de Fisher como método asintótico	145
3.3.1. Introducción.	145
3.3.2. Definiciones clásicas	146

3.3.3. Nueva definición*	147
3.4. Selección del método asintótico óptimo*	149
3.4.1. Objetivo	149
3.4.2. Criterio de selección.	149
3.4.3. Tablas a evaluar para aplicar el criterio	152
3.4.4. Acerca de la globalización en n	153
3.4.5. Acerca de los valores δ a utilizar y primera selección	154
3.4.6. Selección final	163
3.5. Conclusiones*	167
IV. SOBRE LAS CONDICIONES DE VALIDEZ DEL TEST CHI-CUADRADO		
4.1. El valor de E^* en el Modelo III	170
4.1.1. Introducción.	170
4.1.2. Valores experimentales de E^*	170
4.1.3. La condición de asimetría acotada y la condición de validez	174
4.1.4. Valores estimados de E^* : condiciones de validez explícitas.	177
4.1.5. Consecuencias generales	178
4.1.5.1. Introducción	178
4.1.5.2. Sobre el mínimo valor de a_1	178
4.1.5.3. Sobre la forma de la función $E^*(n_1; K)$	179
4.1.5.4. Sobre los valores límites de E^*	180
4.1.5.4. Algunas c.d.v. globales	180
4.1.6. Discusión y conclusiones	181
4.2. El valor de E^* en el Modelo I*	183
4.2.1. Introducción.	183
4.2.2. Valores experimentales de E^*	183
4.2.2.1. Tests a utilizar	183
4.2.2.2. Validez del test aproximado en una tabla dada.	183
4.2.2.3. Generación del espacio muestral.	184
4.2.2.4. Valores experimentales de a_1^* a obtener	186
4.2.2.5. Análisis de los resultados	187
4.2.3. La asimetría del estadístico base para el Modelo I	189
4.2.3.1. Introducción	189
4.2.3.2. Momentos de la multinomial	189
4.2.3.3. Asimetría del estadístico U (bajo H_0)	191
4.2.3.4. Estimación de β_1	193
4.2.4. La condición de asimetría acotada y la condición de validez	194
4.2.5. Valores estimados de E^* : condiciones de validez explícitas.	196
4.2.6. Consecuencias generales	196
4.2.7. Discusión y conclusiones	199
CUADROS	201

ANEXO205
BIBLIOGRAFÍA214

CAPÍTULO I

INTRODUCCIÓN

En este primer capítulo se efectúa una presentación formal del problema objeto de esta memoria, del tipo de soluciones dadas al mismo en la literatura (bien sean exactas o aproximadas), de su naturaleza condicionada o incondicionada y de las razones para éste último conflicto, señalando finalmente qué metodología es la que aquí se asume y el porqué de ello.

1.1. EL PROBLEMA Y SU NOTACIÓN

1.1.1. Introducción

Se entiende por una *Tabla 2×2* a una presentación de datos de frecuencias en dos filas y dos columnas, cada una de ellas aludiendo a una cualidad dicotómica. Pero tras ese término subyacen varios modelos estadísticos que, estando asociados al muestreo que ocasionó los datos, hacen variar el análisis estadístico del problema. Aunque tales diferencias ya fueron puestas de manifiesto por Pearson (1947), lo cierto es que muchos estadísticos no son conscientes de las mismas (y ello por las razones que se verán). Esta memoria está dedicada a sólo uno de los modelos (el menos estudiado en la literatura), pero, de algún modo, los otros dos también intervienen, y así es preciso que se describan. En lo que sigue se alude sólo a los 3 modelos más habituales (y más emparentados entre sí); los otros descritos por Kudô and Tarumi (1978) y relacionados con variables binomiales o multinomiales negativas, son de menor interés práctico y se eluden. Tampoco se contempla el modelo de Tomizawa (1995) que, siendo del tipo actual, alude a otras hipótesis nulas diferentes de las aquí contempladas.

Asímismo, el punto de vista contemplado en esta memoria es el de la *estadística clásica*. Para un punto de vista *bayesiano* ver Altham (1969), Aitchison and Bacon-Shone (1981), Nurminen and Mutanen (1987), Camilli (1995) y Howard (1998). Para

un punto de vista *pseudobayesiano* (a mitad de camino entre los otros dos) ver Ballatori (1987).

Una revisión general de los tres modelos que se detallan en lo que sigue puede verse en Pearson (1947), Kroll (1989), Richardson (1990), Haviland (1990) y Sahai and Khurshid (1995).

1.1.2. Modelo I (objeto de esta memoria)

Si n individuos son tomados al azar de una población y en cada uno de ellos se anota la verificación (A o B) o no (\bar{A} o \bar{B}) de dos características dicotómicas A y B, los resultados de las frecuencias así obtenidas pueden presentarse como en la Tabla 1.1.(a), reseñándose en su parte (b) la probabilidad de obtener una observación en cada una de las casillas. Aquí ninguno de los marginales es conocido de antemano (y se habla del caso de *0 marginales fijados*) y para tener determinada la tabla basta con conocer 3 valores de su interior (por ejemplo x_1, y_1, x_2) cuya distribución será una *multinomial*:

$$P(x_1, y_1, x_2 | n) = \frac{n!}{x_1! y_1! x_2! y_2!} p_{11}^{x_1} p_{12}^{y_1} p_{21}^{x_2} p_{22}^{y_2} \quad (1.1)$$

No se emplea aquí la notación más tradicional de O_{ij} o n_{ij} o x_{ij} para las cantidades observadas x_i e y_j pues para las demostraciones posteriores conviene que la notación sea común a los 3 modelos y, como se verá, la actual será más sugerente.

Tabla 1.1
Notación para el Modelo I

(a) Resultados				(b) Probabilidades			
	A	\bar{A}	Total		A	\bar{A}	Total
B	x_1	y_1	n_1	B	p_{11}	p_{12}	$p_{1\bullet} = q$
\bar{B}	x_2	y_2	n_2	\bar{B}	p_{21}	p_{22}	$p_{2\bullet} = 1 - q$
Total	a_1	a_2	n	Total	$p_{\bullet 1} = p$	$p_{\bullet 2} = 1 - p$	1

El objetivo es decidir si las características A y B son independientes:

$$H_0 \equiv A \text{ y B independientes} \equiv p_{ij} = p_{i\bullet} \times p_{\bullet j} \quad \forall i, j \quad (1.2)$$

contra una alternativa de dos colas ($H_1 \equiv A$ y B son dependientes) o de una cola ($H_1 \equiv$ "A y B están asociadas positivamente" o $H_1 \equiv$ "A y B están asociadas negativamente"). Para efectuar el test es preciso evaluar la (1.1) bajo H_0 :

$$P(x_1, y_1, x_2 | n, H_0) = \frac{n!}{x_1! y_1! x_2! y_2!} p^{a_1} (1-p)^{a_2} q^{n_1} (1-q)^{n_2} \quad (1.3)$$

lo que presenta el problema de la existencia de los dos *parámetros perturbadores* (desconocidos) p y q.

El problema descrito (asociación entre dos cualidades dicotómicas) es frecuente en muchas ciencias experimentales, pero lo es aún más en Medicina y Epidemiología (asociación entre la presencia de un factor de riesgo y la presentación de una enfermedad), lo que ha ocasionado (junto a otras razones que se verán) el que se le hayan dedicado cientos de artículos en los últimos 20 años.

El problema actual, objetivo de esta memoria, ha sido denominado de diversos modos por otros tantos autores: "*Doble dicotomía*" (Barnard, 1947); "*Problema III*" (Pearson, 1947); "*Caso III*" (Kendall and Stuart, 1973); "*Modelo III. Independencia*" (Camilli and Hopkins, 1979); "*Independencia*" (Yates, 1984); "*Origen I*" (Kempthorne, 1979); "*Ensayos Multinomiales 2x2*" (Shuster, 1992); y quizás otros más. El primer nombre probablemente sea el más indicativo desde el punto de vista estadístico; el último desde el punto de vista estadístico-médico. El nombre completo que lo identificaría es el de "*independencia de dos cualidades dicotómicas*", pero es demasiado largo para referirlo brevemente. Kroll (1989) proporciona la lista completa de nombres. Conviene observar la calidad de los estadísticos que participan en el problema, lo que abunda en la frecuencia e importancia del mismo. Desde el punto de vista epidemiológico se le denomina un "*estudio transversal*".

1.1.3. Modelo II

Si n_i individuos ($i=1, 2$) son tomados al azar e independientemente de dos poblaciones y se anota en cada uno de ellos si presentan (A) o no (\bar{A}) una determinada característica dicotómica A, los resultados de las frecuencias así obtenidas pueden presentarse como en la Tabla 1.2.(a), reseñándose en su parte (b) la probabilidad de obtener una observación en cada una de las casillas. Ahora un marginal es conocido de antemano (y se habla del caso de *1 marginal fijado*) y para tener determinada la tabla basta conocer los dos valores independientes x_1 y x_2 , cada uno distribuyéndose como una *binomial*. Con ello la probabilidad de la configuración observada es:

$$P(x_1, x_2 | n_1, n_2) = \binom{n_1}{x_1} \binom{n_2}{x_2} p_1^{x_1} (1-p_1)^{y_1} p_2^{x_2} (1-p_2)^{y_2} \quad (1.4)$$

Tabla 1.2
Notación para el Modelo II

(a) Resultados				(b) Probabilidades			
	A	\bar{A}	Total		A	\bar{A}	Total
Muestra I	x_1	y_1	n_1	Población I	p_1	$1-p_1$	1
Muestra II	x_2	y_2	n_2	Población II	p_2	$1-p_2$	1
Total	a_1	a_2	n				

El objetivo es decidir si las proporciones de presentación de A en ambas poblaciones son la misma o no:

$$H_0 \equiv p_1 = p_2 (=p) \quad (1.5)$$

contra una alternativa de dos colas ($H_1 \equiv p_1 \neq p_2$) o de una cola ($H_1 \equiv p_1 < p_2$ o $H_1 \equiv p_1 > p_2$). Para efectuar el test es preciso evaluar la (1.4) bajo H_0 :

$$P(x_1, x_2 | n_1, n_2, H_0) = \binom{n_1}{x_1} \binom{n_2}{x_2} p^{a_1} (1-p)^{a_2} \quad (1.6)$$

lo que presenta el problema de la existencia del parámetro perturbador (desconocido) p .

El problema actual es tan frecuente o más que el anterior, pero no es nuestro objetivo directo (aunque le servirá de apoyo indirecto). Él ha sido tratado en detalle por Silva (1993). Desde el mismo punto de vista epidemiológico citado en el Modelo I, es conocido como un "*estudio retrospectivo*" (si los n_i aluden a individuos con/sin la enfermedad) o "*prospectivo*" (si los n_i aluden a individuos con/sin el factor de riesgo).

1.1.4. Modelo III

Fisher (1942) presentó su clásico ejemplo de "la señorita y las tazas de té" para ilustrar el modelo. Ahora a una señorita se le ofrecen n_1 tazas con el té echado antes que la leche y n_2 tazas con el té echado después que la leche. Esa información se le proporciona al sujeto experimental, el cual, tras probar el contenido de las tazas, decide de qué tipo son cada una. El investigador tabula los resultados como en la Tabla 1.1.(a). Si la señorita es coherente, ella procurará que $a_1 = n_1$ y $a_2 = n_2$, de modo que ambos marginales están fijados de antemano (aunque para otros problemas no es preciso que sea $a_1 = n_1$) y se habla del caso de *dos marginales fijados*. Para conocer la tabla basta con el valor de x_1 (pues los demás se obtienen por diferencias con n_1 o n_2) cuya distribución es la *hipergeométrica generalizada*:

$$P(x_1 | a_1, n_1, n) = \frac{\binom{n_1}{x_1} \binom{n_2}{a_1 - x_1} \theta^{x_1}}{\sum_{h=r}^s \binom{n_1}{h} \binom{n_2}{a_1 - h} \theta^h} \quad (1.7)$$

con:

$$r = \text{Max}(0; a_1 - n_2) \leq x_1 \leq \text{Min}(a_1; n_1) = s \quad (1.8)$$

y θ la conocida *odds-ratio poblacional* (o razón del producto cruzado).

El objetivo es decidir si la señorita NO distingue las tazas ($H_0 \equiv \theta = 1$) o SÍ las distingue ($H_1 \equiv \theta \neq 1$; $H_1 \equiv \theta > 1$ o $H_1 \equiv \theta < 1$), de modo que, para el test, la distribución de x_1 bajo H_0 es:

$$P(x_1 | a_1, n_1, n, H_0) = \frac{\binom{n_1}{x_1} \binom{n_2}{x_2}}{\binom{n}{a_1}} \quad (1.9)$$

que es la conocida distribución hipergeométrica y que no depende de parámetro perturbador alguno.

1.2. SOLUCIONES AL PROBLEMA

1.2.1. El test (exacto) de Fisher

Fisher (1935), Irwin (1935) y Yates (1934) señalaron que la inferencia en base a los Modelos I y II debe hacerse como en el Modelo III (suponiendo que los marginales están fijados de antemano) pues sólo cabe hacer inferencias basadas en aquello que realmente se ha observado (a_1 para el Modelo II; n_1 y a_1 para el modelo I). Esto lleva a que, bajo H_0 , la probabilidad de una tabla experimental viene dada siempre por la (1.9), dando lugar al conocido *test exacto de Fisher* que se describe a continuación. Para el caso de la (1.7), $\theta = p_{11} p_{22} / p_{12} p_{21}$ en el Modelo I y $\theta = p_1 q_2 / p_2 q_1$ en el Modelo II.

Si x_1 es el valor experimental y se supone que la tabla está reordenada de modo que $\hat{p}_1 = x_1/n_1 > \hat{p}_2 = x_2/n_2$, ó, equivalentemente, que $\hat{\theta} = x_1 y_2 / x_2 y_1 > 1$ (para lo cual basta con permutar convenientemente las filas o columnas), entonces el valor P para $H_1 \equiv \theta > 1$ será:

$$F1(x_1) = \sum_{i=x_1}^s P(i) \quad (1.10)$$

con $P(i)$ dada por la (1.9). Para test de dos colas ($H_1 \equiv \theta \neq 1$) la solución más tradicional y óptima (Martín and Herranz, 1995) consiste en obtener el valor P así:

$$F2(x_1) = \sum_{i=r}^{x'_1} P(i) + \sum_{i=x_1}^s P(i) \quad (1.11)$$

con x'_1 el valor de la otra cola ($x'_1 < E = a_1 n_1 / n$) tal que:

$$P(x'_1) \leq P(x_1) \quad \text{y} \quad P(x'_1 + 1) > P(x_1) \quad (1.12)$$

No se recoge aquí la versión aleatorizada de este test (Tocher, 1950; Lehman, 1959), y que da lugar a un test UMPU, pues, aunque sea teóricamente idóneo, prácticamente ningún autor lo acepta como mecanismo práctico de realizar el test (Plackett, 1964; Mantel and Greenhouse, 1968) por lo irracional de decidir por sorteo la posible significación de la tabla obtenida.

Como regla general para lo que sigue debe tenerse en cuenta que estos tests exactos no son insesgados (salvo el indicado de Tocher) y que ni siquiera se puede hablar de que sean UMP (pues las regiones críticas que producen no están incluidas unas en otras): Upton (1982) y Lloyd (1988).

1.2.2. Los tests (exactos) de Barnard

1.2.2.1. Introducción

Barnard (1945; 1947) asumió directamente los Modelos I y II y propuso eliminar los parámetros perturbadores por maximización del valor del error α en los mismos. Si para Fisher el espacio muestral son todos los posibles valores de x_1 ($r \leq x_1 \leq s$), para Barnard el espacio muestral para el Modelo II son todos los posibles valores de (x_1, x_2) ($0 \leq x_1 \leq n_1; 0 \leq x_2 \leq n_2$) y para el Modelo I todos los posibles valores de (x_1, y_1, x_2) ($0 \leq x_1, y_1, x_2 \leq n; x_1 + y_1 + x_2 \leq n$). Esto implica que los

espacios muestrales consten de un número ascendente de puntos:

$s-r+1$ para el Modelo III (Fisher, 1935)

$(n_1+1)(n_2+1)$ para el Modelo II (Barnard, 1947)

$(n+1)(n+2)(n+3)/6$ para el Modelo I (Pirie and Hamdan, 1972)

lo que a su vez implica menores saltos de probabilidad y favorece la presencia de resultados significativos (Martín and Luna, 1989; Barnard, 1989).

1.2.2.2. Caso del Modelo II

Para determinar el valor P es preciso decidir qué parejas (x_1, x_2) son tan extrañas/extremas o más que la experimentalmente obtenida, lo cual depende del criterio de ordenación elegido para las mismas. Acerca de cuál es el criterio óptimo se ha investigado bastante, pero, no siendo el Modelo II el objetivo de este trabajo, se eludirá tal discusión. El lector interesado puede ver en Martín and Silva (1996) todo lo relevante a estos efectos. Sea cual sea el criterio de ordenación, supongamos que $RC(x_1, x_2)$ es el conjunto de tales parejas. En tal caso, el error $\alpha(p)$ para tal conjunto será:

$$\alpha(p) = \sum_{RC(x_1, x_2)} P(x_1, x_2 | n_1, n_2, H_0) \quad (1.13)$$

con $P(x_1, x_2)$ dada por la (1.6), y el tamaño del test será:

$$\alpha^* = \text{Max}_{0 < p < 1} \alpha(p) = \text{Max}_{0 < p < 1} \sum_{RC(x_1, x_2)} \binom{n_1}{x_1} \binom{n_2}{x_2} p^{a_1} (1-p)^{a_2} \quad (1.14)$$

lo que constituirá el valor P para la pareja (x_1, x_2) dada.

Puede probarse (Martín, 1991) que el valor α^* de la (1.14) es en general más pequeño (para una misma tabla) que el valor obtenido por las (1.10) o (1.11) ó, más generalmente, que el test actual es más potente en general que el de Fisher (Martín and Herranz, 1995), pero, aunque todo ello se supiera desde hace bastante tiempo (Mc Donald et al, 1977), la actual metodología no pudo aplicarse hasta la aparición de los modernos ordenadores, y ello por causa de la complejidad de cálculo que implica la

evaluación de la (1.14). Hoy día tales cálculos son posibles para muestras moderadas, algunos paquetes estadísticos ya los incluyen como una opción (STATXACT 4) y hay programas disponibles en algunas direcciones de Internet:

- <http://www.stat.ncsu.edu/~berger/tables.html>
- <http://www.jiscmail.ac.uk/files/EXACT-STATS/smpexe.uue>
- <http://www.ugr.es/~bioest/SMP.EXE>

El método de ordenación empleado para construir la región crítica (RC en adelante) puede ser cualquiera, pero ha de estar sometido a dos condiciones de coherencia dadas por Barnard (1947):

Condición de Convexidad: Si $(x_1, x_2) \in RC$, con $\hat{p}_1 > \hat{p}_2$, también pertenecerán a ella los puntos $(x_1 + 1, x_2)$ y $(x_1, x_2 - 1)$, pues uno de los \hat{p}_i quedó constante y el otro es más extremo.

Condición de Simetría (para test de dos colas): Si $(x_1, x_2) \in RC$, entonces $(y_1, y_2) \in RC$, pues $p_1 \neq p_2$ es equivalente a $(1-p_1) \neq (1-p_2)$ y se trata de una simple reordenación de la tabla.

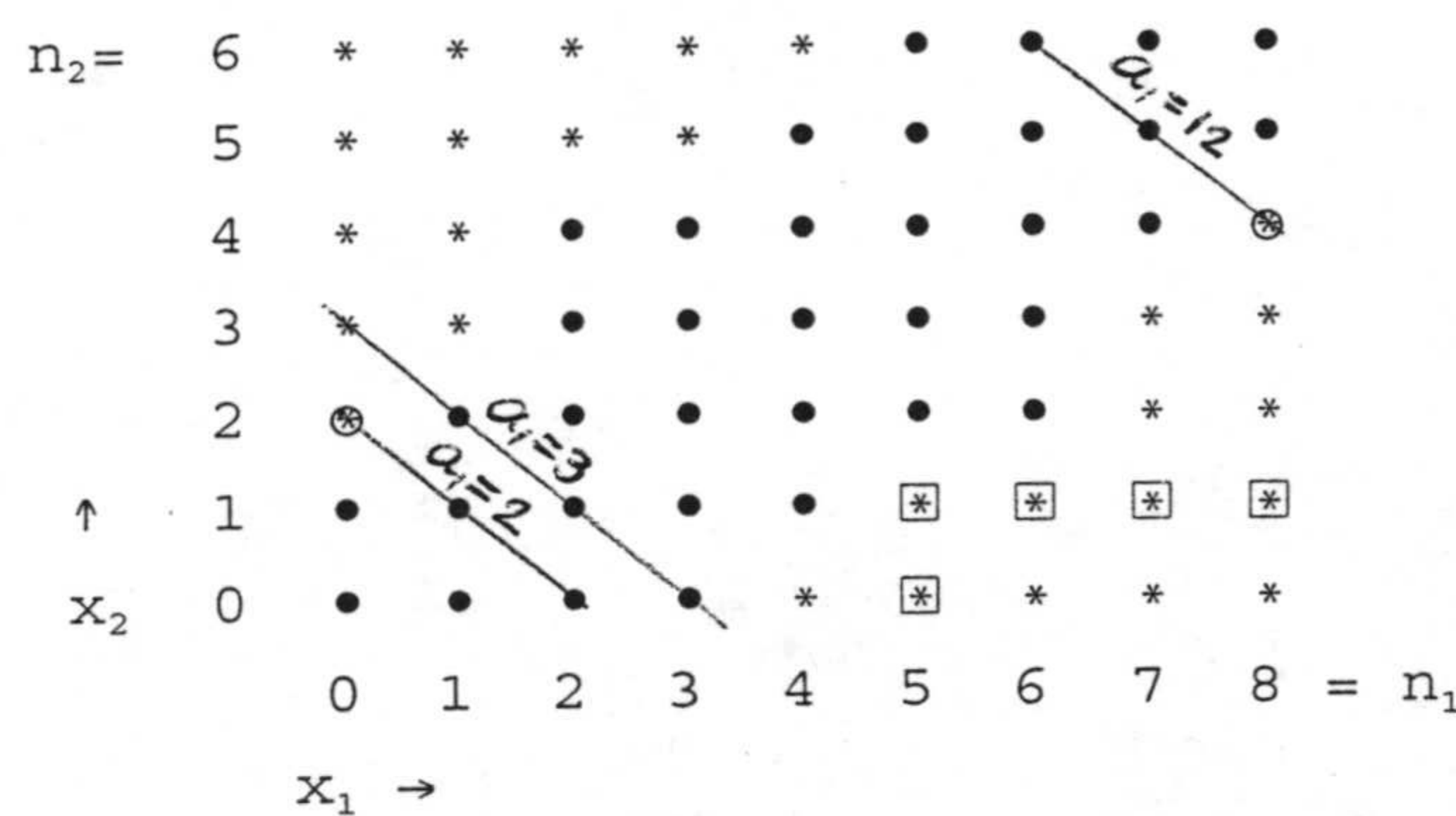


Figura 1.1

Espacio muestral (• y *) y puntos de la RC (*) en un test de dos colas para $H_0 \equiv p_1 = p_2$ con $n_1=8$ y $n_2=6$. Los • y * representan las parejas (x_1, x_2) . Los puntos □ verifican la propiedad de convexidad.

Los puntos ○ verifican la propiedad de simetría.

condiciones que aseguran que la RC no presenta huecos (es convexa) y que es simétrica (si el test es de dos colas). La Figura 1.1. ayuda a entender gráficamente todo lo anterior. En ella, construida para el caso $(n_1, n_2) = (8,6)$, el espacio muestral

viene representado por todos sus puntos; las diagonales secundarias (dibujadas sólo como referencia) se corresponden con el espacio muestral del test exacto de Fisher (pues en ellas a_1 es constante); los puntos marcados con * son los de la RC; los marcados además con \circ son dos puntos simétricos; los marcados con \square aluden a la propiedad de convexidad. Aquí se está prestando atención a estos asuntos de la RC pues más adelante servirán de apoyo y ayuda para el caso del Modelo I.

1.2.2.3. Caso del Modelo I

Para determinar el valor P es preciso decidir qué ternas (x_1, y_1, x_2) son tan extrañas/extremas o más que la experimentalmente obtenida, lo cual depende del criterio de ordenación elegido para las mismas (y sobre los cuales se hablará en el próximo capítulo). Sea cual sea éste, él dará lugar a un conjunto $RC(x_1, y_1, x_2)$ de ternas que proporcionan el error $\alpha(p, q)$ dado por:

$$\alpha(p, q) = \sum_{RC(x_1, y_1, x_2)} P(x_1, y_1, x_2 | n, H_0) \quad (1.15)$$

con $P(x_1, y_1, x_2)$ dado por la (1.3), y el tamaño del test será:

$$\alpha^* = \text{Max}_{0 < p, q < 1} \alpha(p, q) = \text{Max}_{0 < p, q < 1} \sum_{RC(x_1, y_1, x_2)} \frac{n!}{x_1! y_1! x_2! y_2!} p^{a_1} (1-p)^{a_2} q^{n_1} (1-q)^{n_2} \quad (1.16)$$

lo que constituye el valor P para la terna (x_1, y_1, x_2) . Puede probarse (Silva and Martín, 1995) que el valor α^* de la (1.16) es en general más pequeño que el de la (1.14), dando lugar por tanto a un test más potente (en general).

Cabe observar que la (1.16) es aún más compleja de determinar que la (1.14), razón por la cual hay muy poca literatura acerca del Modelo I (en comparación con la mucho más extensa para el Modelo II). De ahí que él sea el objetivo de ésta memoria.

1.2.3. El test (asintótico) de la chi-cuadrado

Los tests exactos (ó no asintóticos) anteriores están basados en las distribuciones exactas y, como se ha dicho, en general dan lugar a valores P distintos (a veces bastante distintos). Sin embargo, sorprendentemente, los tres modelos tienen igual solución asintótica (ó no exacta): el clásico estadístico chi-cuadrado de Pearson en sus dos posibles versiones (Pearson, 1947):

$$\chi_Z^2 = \frac{\{x_1y_2 - x_2y_1\}^2}{a_1a_2n_1n_2} n \quad (1.17)$$

$$\chi_R^2 = \frac{\{x_1y_2 - x_2y_1\}^2}{a_1a_2n_1n_2} (n-1) \quad (1.18)$$

a comparar con una $\chi^2_{1 \text{ g.l.}}$. La raíz cuadrada de la primera expresión da lugar al clásico estadístico Z (z-pooled) generalmente utilizado en el caso del Modelo II:

$$Z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad (1.19)$$

que, con $\hat{p}_i = x_i/n_i$ y $\hat{p} = a_1/n$, ha de compararse con una normal típica.

Sin embargo, las diferencias entre unos modelos y otros aún permanecen si se considera la conveniencia de efectuar en todos los casos una *corrección por continuidad* (c.p.c. en adelante), la cual varía con el modelo asumido. La c.p.c. idónea para el Modelo III fue indicada por Martín et al (1992) y Martín and Herranz (1997), aunque la más conocida es la de Yates (1934). La c.p.c. idónea para el Modelo II es la indicada por Martín and Silva (1996). La c.p.c. idónea para el Modelo I será objeto de esta memoria (Capítulo III).

1.3. RELACIONES ENTRE LOS TRES MODELOS

1.3.1. Supuesta cierta H_0

1.3.1.1. Modelo I vs. Modelo II

Supongamos que se parte del Modelo I en el que se acepta H_0 (la independencia entre A y B). De ser así, como $n_1 \rightarrow B(n; q)$, entonces:

$$P(x_1, y_1, x_2 | n, H_0, p, q) = P(n_1 | n, q) \times P(x_1, x_2 | n_1, n, H_0, p) \quad (1.20)$$

con

$$P(n_1 | n, q) = \binom{n}{n_1} q^{n_1} (1-q)^{n_2} \quad (1.21)$$

lo que fue apuntado (aunque no explícitamente) por Pearson (1947).

Silva and Martín (1995) aprovecharon esta circunstancia para proponer al test del Modelo II como un test válido (aunque conservador) para el Modelo I. Para ello, si $RC(x_1, x_2 | n_1)$ es la RC obtenida para el Modelo II basado en el valor marginal n_1 y el error objetivo α ($\Rightarrow \alpha(p) \leq \alpha$) entonces la $RC(x_1, y_1, x_2 | n) = \bigcup_{n_1} RC(x_1, x_2 | n_1, n)$ puede ser utilizada conservadoramente para el Modelo I, pues en ella:

$$\begin{aligned} \alpha(p, q) &= \sum_{RC(x_1, y_1, x_2 | n)} P(x_1, y_1, x_2) = \sum_{n_1} P(n_1 | n) \sum_{RC(x_1, x_2 | n_1, n)} P(x_1, x_2 | n_1, n) \\ &= \sum_{n_1} P(n_1 | n) \alpha(p) \leq \alpha \sum_{n_1} P(n_1 | n) < \alpha \end{aligned} \quad (1.22)$$

con lo que $\alpha^* < \alpha$ y el test así constituido respeta el error objetivo α .

1.3.1.2. Modelo II vs. Modelo III

Si se parte del Modelo II y se acepta H_0 , entonces $a_1 \rightarrow B(n; p)$ y así (Pearson, 1947):

$$P(x_1, x_2 | n_1, n, H_0, p) = P(a_1 | n, H_0, p) \times P(x_1 | n_1, a_1, n, H_0) \quad (1.23)$$

con $P(x_1)$ dada por la (1.9).

Pearson (1947) aprovechó esta circunstancia para justificar que el test exacto de Fisher (Modelo III) es válido como test para el Modelo II, aunque es conservador. Así, si $RC(x_1 | a_1, n_1, n)$ es la RC de Fisher para los marginales indicados, la RC $(x_1, x_2 | n_1, n) = \bigcup_{a_1} RC(x_1 | a_1, n_1, n)$ es válida para el Modelo II pues:

$$\begin{aligned} \alpha(p) &= \sum_{RC(x_1, x_2 | n_1, n)} P(x_1, x_2) = \sum_{a_1} P(a_1 | n) \times \sum_{RC(x_1 | a_1, n_1, n)} P(x_1 | a_1, n_1, n) \\ &= \sum_{a_1} P(a_1 | n) \alpha(a_1) \leq \alpha \sum_{a_1} P(a_1 | n) < \alpha \end{aligned} \quad (1.24)$$

con lo que $\alpha^* < \alpha$ de nuevo y el test así construido respeta el error objetivo α .

1.3.1.3. Modelo I vs. Modelo III

Pearson (1947) encadenó las expresiones (1.20) y (1.23):

$$P(x_1, y_1, x_2 | n, H_0, p, q) = P(n_1 | n, H_0, q) \times P(a_1 | n, H_0, p) \times P(x_1 | n_1, a_1, n, H_0) \quad (1.25)$$

justificando así que el test exacto de Fisher es válido como test para el Modelo I.

Silva and Martín (1995) fueron algo más allá pues, con similares argumentos, probaron que si los tamaños obtenidos con cada modelo son α^*_I , α^*_{II} y α^*_{III} , entonces:

$$\alpha^*_I < \alpha^*_{II} < \alpha^*_{III} \leq \alpha \quad (1.26)$$

y de ahí que cada modelo, empleado como simplificación del anterior, sea más conservador que éste (dé menos significaciones).

1.3.2. En general

Las factorizaciones anteriores pueden efectuarse también sin necesidad de suponer cierta H_0 , lo que da lugar a relaciones homónimas entre las (1.1), (1.4) y (1.7) en lugar de entre las (1.3), (1.6) y (1.9).

Ahora la (1.21) sigue siendo válida, y $P(x_1, x_2 | n_1, n) = P(x_1, y_1, x_2 | n) \div P(n_1 | n)$, con lo cual:

$$P(x_1, x_2 | n_1, n, p_{ij}/p_{i.}) = \binom{n_1}{x_1} \binom{n_2}{x_2} \left(\frac{p_{11}}{p_{1.}}\right)^{x_1} \left(\frac{p_{12}}{p_{1.}}\right)^{y_1} \left(\frac{p_{21}}{p_{2.}}\right)^{x_2} \left(\frac{p_{22}}{p_{2.}}\right)^{y_2} \quad (1.27)$$

y así la paralela a la (1.20) es:

$$P(x_1, y_1, x_2 | n, p_{ij}) = P(n_1 | n, p_{i.}) \times P(x_1, x_2 | n_1, n, p_{ij}/p_{i.}) \quad (1.28)$$

Ahora sin embargo a_1 no es independiente de n_1 . Como $(x_i | n_i, n) \rightarrow B(n_i; p_{i1}/p_{i.})$ y los valores de $a_1 = x_1 + x_2$ se alcanza en todas las parejas (x_1, x_2) compatibles con que sumen a_1 , será:

$$P(a_1 | n_1, n, p_{ij}/p_{i.}) = \sum_{h=r}^s \binom{n_1}{h} \binom{n_2}{n_1-h} \left(\frac{p_{11}}{p_{1.}}\right)^h \left(\frac{p_{12}}{p_{1.}}\right)^{n_1-h} \left(\frac{p_{21}}{p_{2.}}\right)^{a_1-h} \left(\frac{p_{22}}{p_{2.}}\right)^{n-n_1-a_1+h} \quad (1.29)$$

con lo que la homónima a la (1.23) es ahora:

$$P(x_1, x_2 | n_1, n, p_{ij}/p_{i.}) = P(a_1 | n_1, n, p_{ij}/p_{i.}) \times P(x_1 | n_1, a_1, n, \theta) \quad (1.30)$$

con $P(x_1)$ dada por la (1.7). Finalmente, la paralela a la (1.25) es:

$$P(x_1, y_1, x_2 | n, p_{ij}) = P(n_1 | n, p_{i.}) \times P(a_1 | n, n_1, p_{ij}/p_{i.}) \times P(x_1 | n, n_1, a_1, \theta) \quad (1.31)$$

con valores dados, respectivamente, por las (1.21), (1.29) y (1.7).

Las factorizaciones anteriores es frecuente realizarlas tras una reparametrización adecuada (Lehmann, 1994) a fin de enmarcarlas dentro de las familias de distribuciones clásicas y de los problemas básicos de la inferencia estadística. En el caso actual es habitual definir:

$$\theta = \ln \frac{p_{11}p_{22}}{p_{12}p_{21}} \quad ,, \quad \lambda_1 = \ln \frac{p_{12}}{p_{22}} \quad ,, \quad \lambda_2 = \ln \frac{p_{21}}{p_{22}} \quad (1.32)$$

con lo que la (1.1) pasa a ser:

$$\begin{aligned}
 P(x_1, y_1, x_2 | n, p_{ij}) &= P(x_1, y_1, x_2 | n, \theta, \lambda_1, \lambda_2) = \\
 &= \frac{n!}{x_1! y_1! x_2! y_2!} \cdot \frac{\exp \{x_1 \theta + n_1 \lambda_1 + a_1 \lambda_2\}}{\{\exp(\theta + \lambda_1 + \lambda_2) + \exp(\lambda_1) + \exp(\lambda_2) + 1\}^n} \quad (1.33)
 \end{aligned}$$

lo cual, entendiendo que λ es el vector (λ_1, λ_2) , es de la forma de la familia exponencial:

$$h(x) K(\theta, \lambda) \exp \{x_1 \theta + (n_1, a_1) \lambda\}$$

Sin embargo, a los efectos que se verán, es más cómodo expresar lo anterior omitiendo el "ln" de las definiciones previas:

$$\theta = \frac{P_{11}P_{22}}{P_{12}P_{21}} \quad ,, \quad \lambda_1 = \frac{P_{12}}{P_{22}} \quad ,, \quad \lambda_2 = \frac{P_{21}}{P_{22}} \quad (1.33)$$

y así:

$$P(x_1, y_1, x_2 | n, \theta, \lambda_1, \lambda_2) = \frac{n!}{x_1! y_1! x_2! y_2!} \cdot \frac{\theta^{x_1 + \lambda_1^{n_1} + \lambda_2^{a_1}}}{\theta \cdot \lambda_1 \cdot \lambda_2 + \lambda_1 + \lambda_2 + 1} \quad (1.34)$$

la cual se factoriza así:

$$P(x_1, y_1, x_2 | n, \theta, \lambda) = P(a_1, n_1 | n, \theta, \lambda) \times P(x_1 | n, a_1, n_1, \theta) \quad (1.35)$$

-si de nuevo λ designa al vector (λ_1, λ_2) -, con:

$$P(a_1, n_1 | n, \theta, \lambda) = \frac{\binom{n}{n_1} \lambda_1^{n_1} \lambda_2^{a_1} \sum_{h=r}^s \binom{n_1}{h} \binom{n_2}{a_1 - h} \theta^h}{\{\theta \lambda_1 \lambda_2 + \lambda_1 + \lambda_2 + 1\}^n} \quad (1.36)$$

$$P(x_1 | n, a_1, n_1, \theta) = \frac{\binom{n_1}{x_1} \binom{n_2}{x_2} \theta^{x_1}}{\sum_{h=r}^s \binom{n_1}{h} \binom{n_2}{a_1-h} \theta^h} \quad (1.37)$$

1.4. ¿TEST CONDICIONADO O INCONDICIONADO?

1.4.1. Introducción

La discusión acerca de si debe condicionarse o no supera el ámbito de esta memoria por tratarse de un tema que hunde sus raíces en las propias bases de la estadística. Más aún si, como en el caso actual, la discusión se plantea en presencia de parámetros perturbadores. Sin embargo conviene reseñar aquí lo más relevante, pues es preciso justificar el porqué de la visión que aquí se adopta (la incondicionada).

Aunque la discusión "condicionar/no condicionar" sea algo general de la estadística, y no particular de las tablas 2×2 , la mayoría de los autores que intervienen en ella terminan por ejemplificar sus ideas a través de una tabla 2×2 (generalmente bajo el Modelo II, que es más sencillo).

Una buena muestra de lo complejo de la situación es que Barnard (el creador del método incondicionado en 1947) rápidamente se arrepiente (1949) pasando a defender el método condicionado (1982 y 1989). Entre tanto, un gran número de autores se muestran favorables a condicionar (Yates, 1984; Little, 1989; Havilan, 1990; Cormack and Mantel, 1991; Upton, 1992; Lehmann, 1993; ...) y otros a no hacerlo (Garside and Marck, 1967; Liddell, 1976; McDonald et al, 1977; Upton, 1982; Haber, 1987; D'Agostino et al, 1988; Suissa and Shuster, 1984 y 1985; ...), en donde Upton es otro autor que cambia de opinión. Silva (1993) dedicó un capítulo completo a una revisión crítica del tema, y en él pueden verse con más detalle los argumentos que intervienen en la discusión. En particular, el más relevante es el que se cita a continuación.

1.4.2. Los principios de suficiencia y de auxiliaridad

Aunque los conceptos que siguen tienen una larga historia dentro de la estadística (por ejemplo, Cox, 1958, ya introdujo los conceptos de estadísticos "auxiliar" y "condicionalmente suficiente") en lo que sigue, por homogeneizar conceptos y notaciones, se seguirá el trabajo más reciente y global de Reid (1995).

Si una función de densidad $f(x; \theta)$, con x aludiendo a las observaciones y θ a los parámetros, se puede factorizar como:

$$f(x; \theta) = f_s(s; \theta) \times f_{t|s}(t | s) \quad (1.38)$$

con (s, t) relacionándose con x según una transformación uno a uno, entonces s es un estadístico suficiente para θ y las inferencias sobre θ deben basarse en f_s (la densidad marginal de s) por ser la que contiene toda la información sobre θ . Esto es algo bien aceptado por los estadísticos, pero no hay tanta unanimidad en el caso de que la factorización sea:

$$f(x; \theta) = f_{s|t}(s | t; \theta) \times f_t(t) \quad (1.39)$$

en cuyo caso es la distribución condicionada la que se lleva la información sobre θ : ahora ya no hay tanto acuerdo en que la inferencia sobre θ deba hacerse en base a $f_{s|t}$, con t un estadístico auxiliar para θ (por ejemplo, Suissa and Shuster, 1991, abogan por no condicionar y efectuar la inferencia en base a f), aunque en general está bastante aceptado.

Cuando en el problema intervienen dos parámetros (el de interés θ y el perturbador λ), el acuerdo se complica aún más. Ahora pueden darse varias situaciones, pero la que nos afecta es la siguiente factorización que es una extensión de la (1.39).

$$f(x; \theta, \lambda) = f_{s|t}(s | t; \theta) \times f_t(t; \lambda) \quad (1.40)$$

Ahora la distribución condicionada se lleva la información del parámetro de interés θ (con lo que t es auxiliar para θ), en tanto que la no condicionada se lleva la información del parámetro perturbador λ (con lo que t es suficiente para λ). Esto

indica que, salvo excepciones, la mayoría de los estadísticos estarán de acuerdo en utilizar $f_{s|t}$ para las inferencias sobre θ y f_t para las inferencias sobre λ . En el contexto del Modelo I, y supuesta cierta H_0 , la expresión (1.25) es la homónima de la (1.40), con $(x_1, y_1, x_2) \equiv x$, $x_1 \equiv s$, $(a_1, n_1) \equiv t$, $p = \lambda$ y $\theta = 1$ bajo H_0 . Ello llevó a Gart (1971) y a otros autores a deducir el carácter auxiliar del estadístico (a_1, n_1) y a aconsejar el efectuar la inferencia sobre H_0 en base a la distribución condicional $f_{s|t} \equiv f_{x_1|a_1, n_1}$ (es decir, en base al test exacto de Fisher).

Desgraciadamente la situación actual no es esa, pues la factorización solicitada debe hacerse antes de suponer cierta H_0 . De modo general, una posibilidad alternativa a la (1.39) es que sea:

$$f(x; \theta, \lambda) = f_{s|t}(s|t; \theta) \times f_t(t; \theta, \lambda) \quad (1.41)$$

con lo que t es suficientemente para λ , pero no es auxiliar para θ , y así basar la inferencia sobre θ en $f_{s|t}$ alcanza aún menos acuerdos que en el caso de la (1.40) por causa de que ahora se pierde la información que t posee sobre θ . Para nuestro caso del Modelo I, $f(x; \theta, \lambda)$ viene dado por la (1.34), con $x \equiv (x_1, y_1, x_2)$, $\lambda = (\lambda_1, \lambda_2)$ y la reparametrización de la (1.33), $t = (a_1, n_1)$ y $s = x_1$. Con ello la (1.35), que es la factorización actual, es de la forma de la (1.41), con $f_{s|t}$ y f_t dadas por las (1.37) y (1.36) respectivamente. Así pues, los partidarios del condicionamiento hacen énfasis en que las inferencias sobre θ han de hacerse en base a la (1.37), la cual para $\theta = 1$ (es decir bajo H_0) es la hipergeométrica y da lugar al test exacto de Fisher. Por el contrario, los partidarios del no condicionamiento opinan que la inferencia ha de hacerse a partir de la distribución original (1.34) o (1.1), la cual bajo H_0 da lugar a la ya discutida expresión (1.3) base de toda la inferencia a contemplar en esta memoria.

En realidad, decir que uno es o no partidario del condicionamiento es una afirmación demasiado imprecisa pues, el serlo, puede depender del tipo de densidad a la que nos enfrentemos. Más preciso es definirse partidario o no del principio de suficiencia. De un modo grosero podemos decir que los estadísticos se encuadran en

uno de estos tres tipos:

- 1) Los "*incondicionados totales*" (como Suissa and Shuster) partidarios de efectuar la inferencia incondicionada tanto en el caso de la (1.40) como en el de la (1.41). Ellos no son partidarios ni del principio de suficiencia, ni del de auxiliaridad.
- 2) Los "*incondicionados parciales*" (como el segundo grupo de autores citado en el 3^{er} párrafo de 1.4.1) partidarios de efectuar la inferencia incondicionada en el caso de la (1.41), pero no en el de la (1.40). Ellos son partidarios del principio de suficiencia y del principio de auxiliaridad.
- 3) Los "*condicionados totales*" (como el primer grupo de autores citado en el 3^{er} párrafo de 1.4.1) partidarios de efectuar la inferencia condicionada siempre, tanto en la (1.40) como en el de la (1.41). Ellos son partidarios del principio de suficiencia, pero no del de auxiliaridad.

En realidad muchos de los estadísticos del tercer grupo se muestran partidarios de un principio de auxiliaridad laxo, pues aducen que son favorables al condicionamiento cuando en la (1.41) t posee poca información acerca de θ , lo cual, se comprenderá, es muy relativo. Plackett (1977) vió que los marginales de la tabla contienen en general poca información sobre θ , pero Hinde and Aitkin (1987) indicaron que con pequeñas muestras ésta puede ser importante, y Zhu and Reid (1994) la evaluaron en hasta un 26%. Nuestra posición es la intermedia (la del segundo grupo de autores), lo cual implica que aceptamos la inferencia incondicionada en el caso de la (1.41) -caso de nuestro Modelo I-, pero no en el de la (1.40) -caso de la comparación de dos proporciones a través de muestras apareadas-, y ello por considerar inaceptable la pérdida de información sobre θ que implica trabajar con $f_{s|t}$ en la (1.41) en lugar de con f . Cosa distinta es reconocer los grandes servicios prestados a la estadística por las distribuciones condicionadas (el test exacto de Fisher puede hacerse a mano, el de Barnard no) en épocas de dificultad de cómputo, con la ventaja añadida de que los tests que proporcionan no dan significaciones falsas por ser

conservadores (recordar la (1.26)). Un estudio más detallado de las implicaciones filosófico-estadísticas de los distintos principios (condicionalidad, suficiencia, auxiliaridad y verosimilitud) puede verse en Helland (1995) que, en esencia, apoya estas ideas.

1.5. LA RAZÓN DEL PRODUCTO CRUZADO

Ya se ha visto cómo el parámetro

$$\theta = p_{11}p_{22}/p_{12}p_{21} \quad (1.42)$$

juega un papel fundamental en las tablas 2×2 , y en particular en el Modelo I actual: recordar la reparametrización de la (1.33). Él es conocido por *odds-ratio* o razón de superioridades o razón de ventajas o *razón del producto cruzado*, y es el parámetro básico del modelo logístico. Edwards (1963) probó que θ es la medida de asociación más importante en Tablas 2×2 , recomendando usarla sólo a ella (ó sus funciones) para tal propósito, debiéndose recordarse aquí que el objetivo actual es ver si la fuerza de asociación θ entre las dos cualidades dicotómicas A y B es 1 o $\neq 1$ (pues la (1.2) es equivalente a $H_0 \equiv \theta = 1$). No es el objetivo de esta memoria las inferencias generales sobre θ , pero conviene ver ahora algunas propiedades de θ que se utilizarán más adelante (de hecho algunas ya han sido utilizadas).

Se sabe que $0 \leq \theta < \infty$, con $\theta < 1$ aludiendo a una asociación negativa entre A y B (por ejemplo, la presencia del factor de riesgo favorece la ausencia de la enfermedad: el factor de riesgo es un factor de protección); $\theta > 1$ aludiendo a una asociación positiva entre A y B (y la presencia del factor de riesgo favorece la presencia de la enfermedad); $\theta = 1$ aludiendo a la independencia entre A y B (el factor de riesgo no es tal), y θ y θ^{-1} aludiendo a igual fuerza de asociación, pero de signo contrario.

Supongamos que una tabla dada tiene un valor θ dado. La fuerza de asociación en ella no debe cambiar (y así sucede) por el hecho de permutar las filas entre sí, las columnas entre sí o las filas por columnas, ni tampoco si cada fila y/ó columna de p_{ij}

se multiplica por números dados (con tal de que siga sucediendo que $\sum p_{ij} = 1$), aunque en los dos primeros casos la asociación cambiará de sentido.

El estimador de máxima verosimilitud de θ es:

$$\hat{\theta} = \frac{x_1 y_2}{x_2 y_1} \quad (1.43)$$

el cual no es insesgado (de hecho su esperanza es infinita si $x_2=0$ o $y_1=0$) y su error estándar estimado $\hat{\theta}(\sum 1/x_i + \sum 1/y_i)^{1/2}$ no está definido si alguna frecuencia es nula. Para evitarlo se utiliza la corrección de Haldane-Anscombe

$$\hat{\theta}' = \frac{(x_1+0,5)(y_2+0,5)}{(x_2+0,5)(y_1+0,5)} \quad (1.44)$$

cuyo error estándar estimado es $\hat{\theta}' \{ \sum 1/(x_i+0,5) + \sum 1/(y_i+0,5) \}^{1/2}$, si bien a efectos de inferencia se le utiliza en escala logarítmica:

$$\ln \hat{\theta}' \rightarrow N \left\{ \ln \theta; \left[\sum 1/(x_i+0,5) + \sum 1/(y_i+0,5) \right]^{1/2} \right\} \quad (1.45)$$

que ya sí tiene momentos finitos.

CAPÍTULO II

TESTS EXACTOS

Los métodos incondicionados exactos para el análisis de una tabla 2×2 (es decir, los tests de Barnard) han merecido gran atención en los últimos 20 años, coincidiendo con la implantación y desarrollo de los modernos ordenadores. Sin embargo, de los cientos de artículos (sic) aparecidos sobre el tema, la gran mayoría se refieren al caso del Modelo II (la comparación de dos proporciones), pues éste cuenta con menores problemas de cómputo que el caso del Modelo I (la asociación de dos cualidades dicotómicas). De hecho, hasta el último decenio no se ha podido decir algo de interés sobre el mismo.

Con todo, el método original de Barnard es aún hoy día prácticamente inaplicable, por lo que algunos autores han propuesto diversos métodos abreviados que presentan varias dificultades:

- 1) No existe un método abreviado único que pueda considerarse como el óptimo.
- 2) No se conoce la pérdida de potencia sufrida por utilizar un método abreviado.
- 3) No existe un marco conceptual definido acerca de las reglas que permitan construir un método abreviado lícito.

En este Capítulo se pretende subsanar estos defectos, al tiempo que se proponen nuevos métodos abreviados y se ofrecen diversas propiedades que permitan aminorar el tiempo de cómputo requerido y/o la calidad de las soluciones ofrecidas.

El término "exactos" empleado como título para el capítulo es una herencia del mismo término utilizado para aludir al "test exacto de Fisher". Con ello se quiere indicar que el test se va a efectuar sin utilizar expresiones aproximadas, sino a través de la (1.16), por lo que la frase *Tests Exactos* debe entenderse como equivalente a

Tests No Asintóticos (frase que de hecho es preferida por algunos autores).

2.1. ESTADO ACTUAL DEL PROBLEMA

2.1.1. Introducción

En 1.1.2 se indicó la estructura del Modelo I objeto de esta memoria, y en 1.2.2.1 y 1.2.2.3 se indicó que el test de Barnard para este caso de 0 marginales fijados se basa en las ternas de valores (x_1, y_1, x_2) cuyo espacio muestral viene dado por todos los valores posibles de ella ($0 \leq x_1, y_1, x_2 \leq n; x_1 + y_1 + x_2 \leq n$), por lo que consta de $(n+1)(n+2)(n+3)/6$ ternas posibles o tablas posibles ó, abreviadamente, puntos. También se indicó que, *dado un error objetivo* α , y supuesta ya construida una región crítica $RC(x_1, y_1, x_2)$ para él, el error del test para ella es:

$$\begin{aligned} \alpha(p, q) &= \sum_{RC(x_1, y_1, x_2)} P(x_1, y_1, x_2 | n, H_0) = \\ &= \sum_{RC(x_1, y_1, x_2)} \frac{n!}{x_1! y_1! x_2! y_2!} p^{a_1} (1-p)^{a_2} q^{n_1} (1-q)^{n_2} \end{aligned} \quad (2.1)$$

(con p y q los parámetros perturbadores) y que el tamaño del test será:

$$\alpha^* = \text{Max}_{0 < p, q < 1} \alpha(p, q) \quad (2.2)$$

lo que puede entenderse como el *error real* α^* del test. Los problemas son tres:

- 1) ¿Cómo representar gráficamente el espacio muestral y la RC del modo más simple e intuitivo posible?
- 2) ¿Cómo formar la RC para obtener un test lícito y óptimo?
- 3) ¿Cómo calcular α^* con un tiempo de cómputo mínimo?

los cuales son abordados en lo que sigue.

2.1.2. Variables base para la formación del espacio muestral

Ya se ha visto que cada observación de la distribución multinomial base del Modelo I puede expresarse exhaustivamente mediante una tabla como la Tabla 1.1(a). En ella sólo n estaba fijado de antemano, y así un modo corto de aludirla fue mediante la terna (x_1, y_1, x_2) , que es una observación de la variable multinomial base. Pero es claro que hay otras posibles opciones: cualquier terna que permita completar (por resta y suma) la Tabla 1.1.(a). Como ello va a implicar una representación tridimensional del espacio muestral y de la RC, el asunto no es baladí (a efectos prácticos) si se pretende efectuar una representación lo mas manejable e intuitiva posible.

Barnard (1947) optó por la terna (x_1, y_1, x_2) , lo que daba lugar a un espacio muestral sencillo de visualizar por su forma de tetraedro, sencillo de evaluar en su número de puntos $-N = (n+1)(n+2)(n+3)/6$ según Pirie and Hamdan (1972)-, pero complejo de interpretar en cuanto a la ubicación de la RC. Ver Figura 2.1(a).

Pearson (1947) optó por la terna (x_1, a_1, n_1) que está ligada al formato del test exacto de Fisher: para cada (a_1, n_1) , x_1 toma los valores (1.8) de la distribución hipergeométrica. Ahora la visión de la RC es algo más intuitiva. Ver Figura 2.1(b).

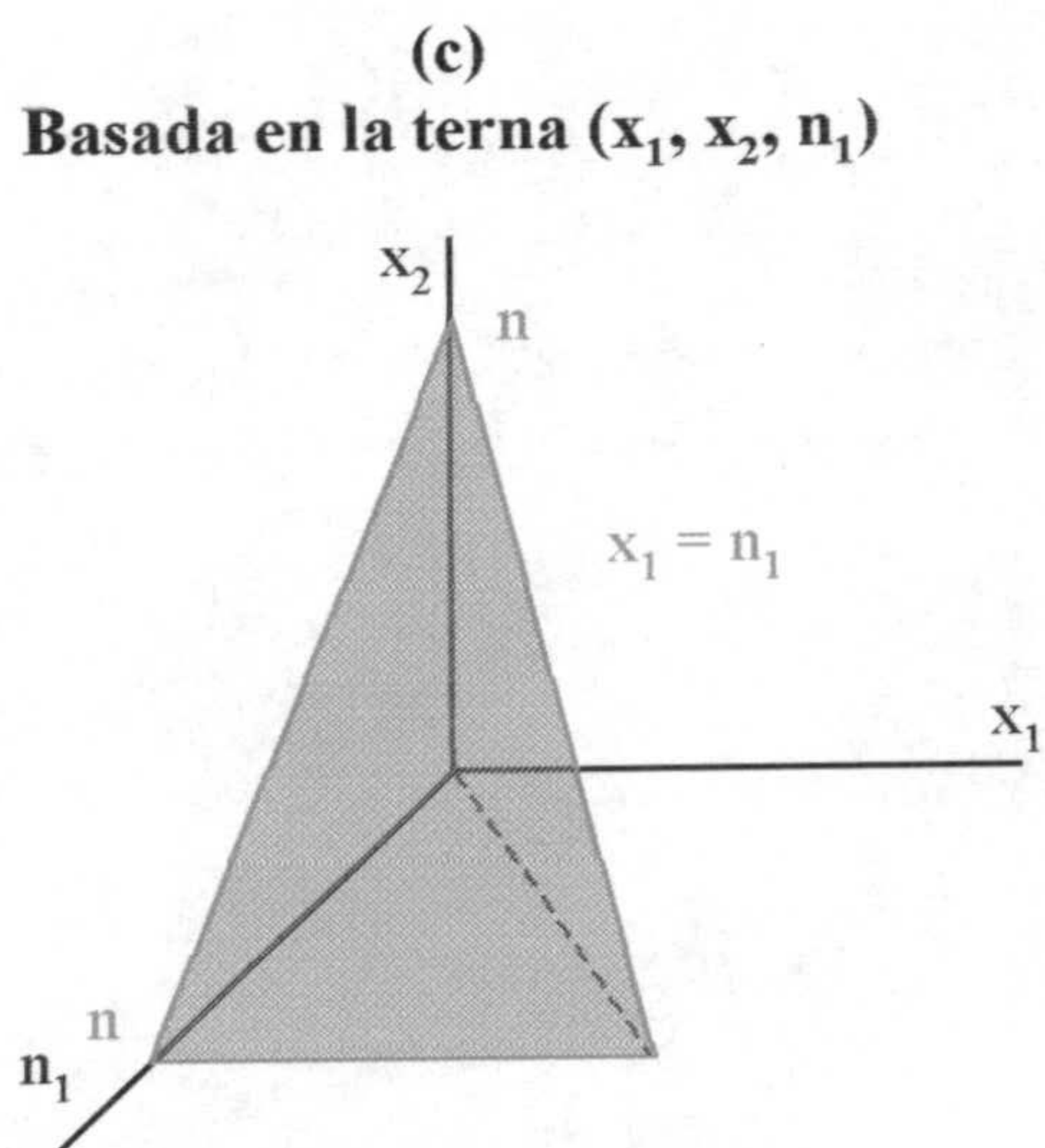
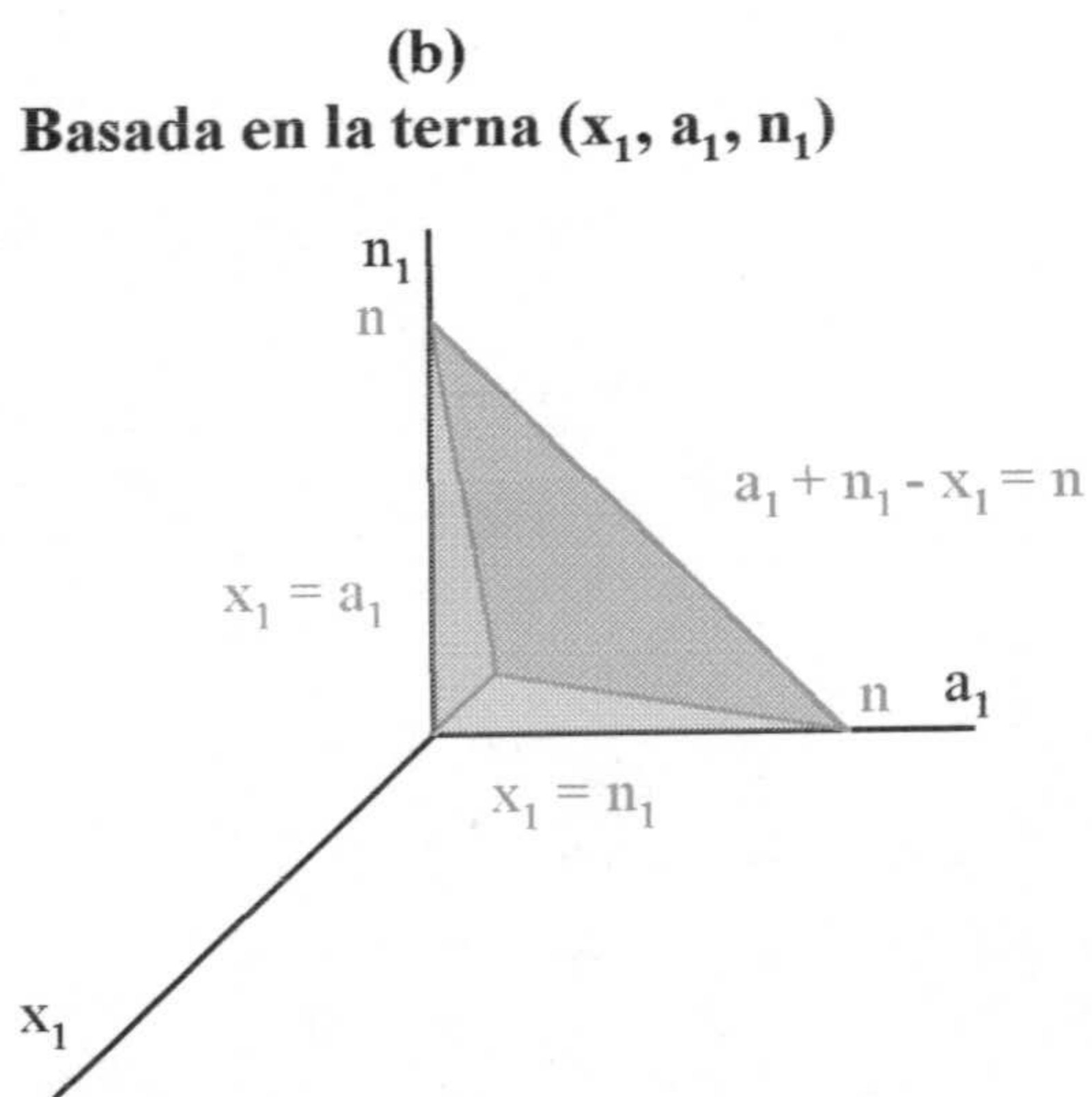
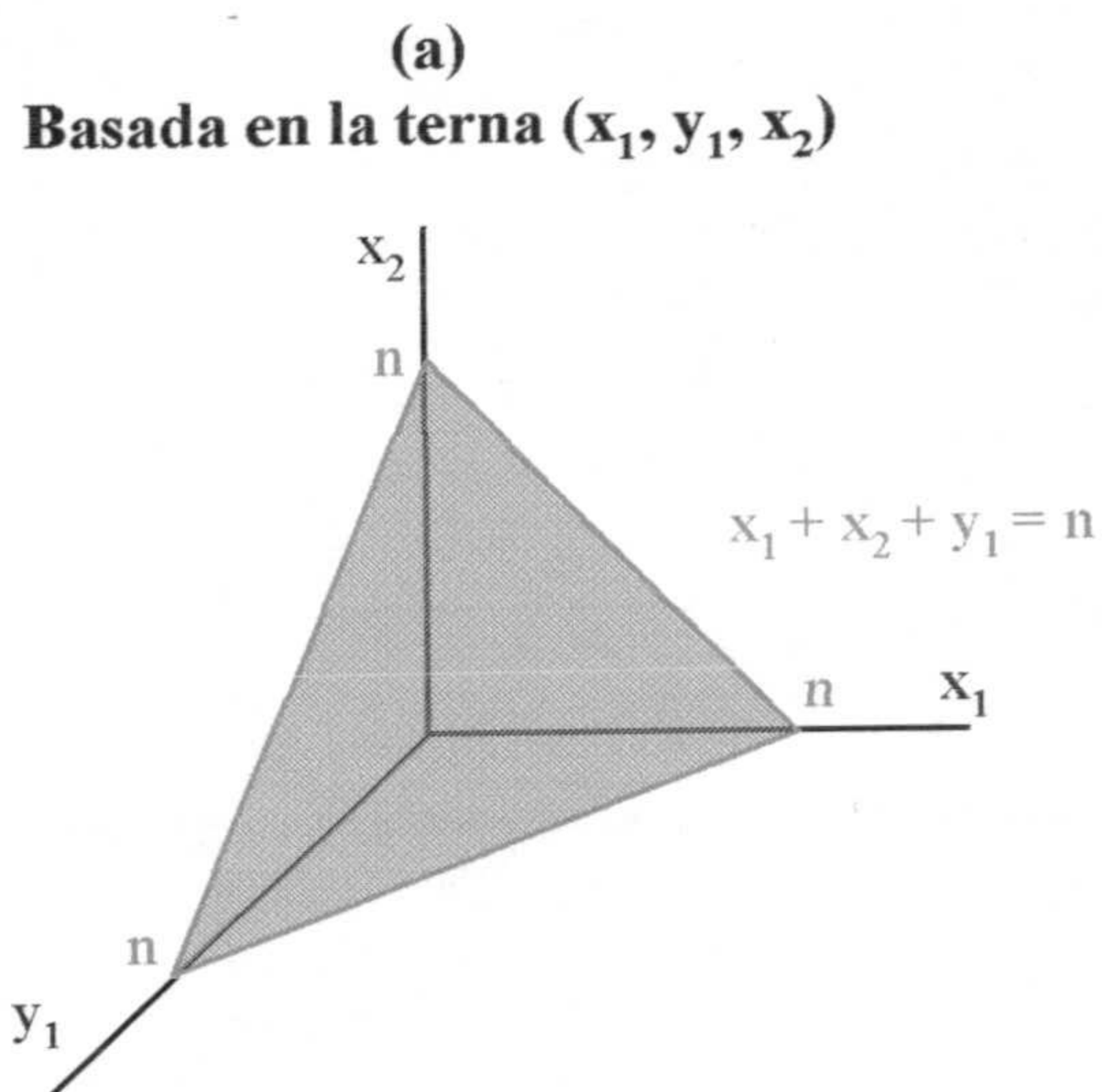
En tiempos de Pearson era impensable trabajar numéricamente con el Modelo II, y así recurrió al Modelo III para visualizar la RC y el espacio muestral. Hoy día el Modelo II está ya bastante bien analizado, y trabajar con su espacio muestral y RC (recordar la Figura 1.1) es bastante sencillo e intuitivo. Ello llevó a Ortiz (1993) a proponer como terna de trabajo a (x_1, x_2, n_1) , lo que implica que cada corte con $n_1 = \text{cte.}$ da el espacio muestral y RC para (x_1, x_2) . Ver Figura 2.1.(c). Como se verá más tarde, esta será nuestra elección para el desarrollo de nuestras aportaciones a ésta memoria.

2.1.3. Formación de la región crítica: métodos de ordenación

Una vez decidida la terna de trabajo y conocido su espacio muestral, ¿cómo se van introduciendo los puntos en la $RC(\alpha)$? Es claro que se precisa *un criterio de ordenación* para los puntos del espacio muestral y que, conocido éste, la formación de la $RC(\alpha)$ es sencilla: ir introduciendo los puntos en la RC según el orden estable-

Figura 2.1

Distintas posibilidades de representar el espacio muestral en el caso del Modelo I



cido, deteniéndose cuando el α^* de la (2.2) supere por primera vez a α , en cuyo caso la $RC(\alpha)$ es la penúltima utilizada.

Barnard (1947), en su creación del método incondicionado, no le presta mayor atención a este asunto, remitiéndose simplemente a una solución paralela a la del Modelo II. Se sobreentiende pues que su criterio, llamado *criterio del mínimo* y al que aludiremos por **M**, es el siguiente:

- 1) Calcular la (2.2) en todas las RC formadas por un único punto del espacio muestral, lo que da tantos α^*_{0i} como puntos posibles (N). El primer punto a introducir en la RC es el que menor valor α^*_{0i} haya ocasionado (de ahí la palabra "mínimo") y tal RC sera la RC_1 .
- 2) Añadir a la RC_1 uno a uno los demás puntos del espacio muestral, obteniendo así diversas RC's ampliadas RC_{1i} (en principio formadas por dos puntos) que proporcionan valores α^*_{1i} . La RC de este segundo paso - RC_2 - es la que menor valor tenga de α^*_{1i} .
- 3) Así se procede para obtener la RC_3 , RC_4 , etc. Si en RC_h es $\alpha^*_h \leq \alpha$ y en RC_{h+1} es $\alpha^*_{h+1} > \alpha$, entonces la RC buscada es $RC(\alpha) = RC_h$.
- 4) Si en algún paso p sucede que dos o más valores α^*_{pi} son iguales (por ejemplo α^*_{p1} y α^*_{p2}) los dos puntos extras añadidos a la RC_p deben estar juntos, es decir $RC_{p1} \cup RC_{p2}$ es la $RC_{p(12)}$ a ensayar. Tales "empates" suceden en la práctica con bastante frecuencia, y así cada RC_p se constituye añadiendo más de un punto a la RC_{p-1} anterior.

Se observará que el método **M** es teóricamente el idóneo, pues al elegir el mínimo α^*_{pi} se garantiza que α^* crece poco de una RC a la siguiente, y así el número de puntos en $RC(\alpha)$ será máximo (alcanzándose una máxima potencia). Pero también se observará que el método **M** es extraordinariamente incómodo por las innumerables veces que ha de determinar la costosa expresión (2.2). De hecho ningún autor lo ha citado ni ejecutado nunca.

Lo costoso de la (2.2) hizo que durante muchos años los investigadores

olvidaran el problema o se dedicaran al caso más sencillo del Modelo II. Y así hasta que Haber (1986, 1987) retomó el problema actual.

Haber (1987) especifica que la RC debe ser invariante a una permuta de filas por columnas (pues ya se vió en 1.5 que entonces $\hat{\theta}$ no cambia), lo que le llevó a proponer como criterios de ordenación a todos los del caso del Modelo II (ya entonces bastante estudiado) que verifiquen esto y que se citan más abajo. Todos los métodos que siguen son de una complejidad de cómputo muy inferior a la del método **M** pues todos imprimen a los puntos del espacio muestral un orden que no depende de la (2.2), sino de una expresión más sencilla. La idea que subyace en todos estos métodos "simplificados" es *ordenar los puntos de menor a mayor (ó al revés) valor de un estadístico sencillo de obtener*. Conocido eso, la (2.2) sólo se utiliza una vez en cada ocasión en que se introduce un nuevo punto en la RC. De hecho, si el objetivo es calcular el valor P de una tabla experimental, la ordenación permite construir la RC problema (hasta que el punto experimental entre en ella) y la (2.2) se aplica sólo entonces (una vez): Su α^* es el valor P buscado.

Los métodos de ordenación propuestos por Haber (1987) son los que siguen:

- a) Ordenar de menor a mayor valor del valor P del test exacto de Fisher para cada tabla, idea que es original de Boschloo (1970) para el caso del Modelo II. Aunque Haber no lo explicita, se supone que el valor P al que alude es el de una cola -expresión (1.10)- cuando el test es de una cola, y el de dos colas -expresión (1.11)- cuando el test es de dos colas. Esto da lugar al método de ordenación **F**.
- b) Ordenar de mayor a menor valor de diversos estadísticos utilizados para efectuar el test asintótico en Tablas 2×2 :

$$L = \sum x_i \ln \frac{x_i}{E_{i1}} + \sum y_i \ln \frac{y_i}{E_{i2}} \quad (2.3)$$

$$G = \{ \ln \hat{\theta}' \}^2 / \left\{ \sum \frac{1}{x_i+0,5} + \sum \frac{1}{y_i+0,5} \right\} \quad (2.4)$$

$$Z = \sqrt{\chi_Z^2} \quad (2.5)$$

el primero es el clásico test de la razón de verosimilitudes, con E_{ij} las clásicas cantidades esperadas dadas por $a_i n_j / n$, y da lugar al que llamaremos método **L**; el segundo es el test de Gart (1966) de la odds-ratio, y da lugar al método **G**; el tercero es el clásico test chi-cuadrado (puesto en formato **Z**) de la (1.19) o (1.17), y da lugar al método **Z**.

Años más tarde, Shuster (1992) empíricamente comprobó que la ordenación basada en **Z** se mejoraba sustancialmente si se le incluía la corrección por continuidad (c.p.c. en adelante) clásica de Yates. Así, si

$$\chi_Y^2 = \frac{\left\{ |x_1 y_2 - x_2 y_1| - \frac{n}{2} \right\}^2}{a_1 a_2 n_1 n_2} n \quad (2.6)$$

entonces su estadístico de orden:

$$Z_Y = \begin{cases} +\chi_Y & \text{si } x_1 > E_{11} + 0,5 \\ -\chi_Y & \text{si } x_1 < E_{11} - 0,5 \\ 0 & \text{si } E_{11} - 0,5 \leq x_1 \leq E_{11} + 0,5 \end{cases} \quad (2.7)$$

da lugar al método **Y**. La triple definición anterior persigue evitar valores ficticios en las proximidades de la c.p.c. del 0,5. Kroll (1989) lo solventa de este otro modo:

$$\chi_Y^2 = \frac{n}{a_1 a_2 n_1 n_2} \text{Max} \left\{ 0; |x_1 y_2 - x_2 y_1| - \frac{n}{2} \right\}^2 \quad (2.8)$$

y le asigna a χ_Y el signo de la asociación encontrada. En este caso la ordenación es de mayor a menor Z_Y para $H_1 \equiv$ "asociación positiva"; de menor a mayor Z_Y para $H_1 \equiv$ "asociación negativa"; y de mayor a menor $|Z_Y|$ para $H_1 \equiv$ "asociación cualquiera" (dos colas).

2.1.4. Notas acerca del cálculo y lugar del máximo

Ya se ha indicado varias veces lo complejo de evaluar la (2.2), lo que ha hecho que algunos autores hayan dedicado algún esfuerzo para simplificarla.

Kroll (1989) observa experimentalmente que $\alpha(p,q) = \alpha(q,p)$, por lo que sugiere calcular el máximo no en $0 < p,q < 1$ sino en $0 < p,q \leq 0,5$ por causa de tal simetría. Como única prueba remite al lector el caso del Modelo II (en que eso ya sucedía para el único parámetro p).

Shuster (1992) es el único que especifica algo más y ello para su ordenación Y . Sea Z_{exp} el valor de Z_Y en una tabla experimental. El conjunto de todas las tablas con un valor Z_Y más extremo (en el sentido de arriba) que Z_{exp} formarán las RC_1 y RC_2 de una y dos colas, con errores α de:

$$\alpha_i(Z_{exp}, p, q) = \sum_{RC_i} \frac{n!}{x_1! y_1! x_2! y_2!} p^{a_1} (1-p)^{a_2} q^{n_1} (1-q)^{n_2} \quad (2.9)$$

Si en la tabla se permutan dos filas, $+ Z_Y$ pasa a $-Z_Y$ (y al revés) y $q^{n_1} (1-q)^{n_2}$ pasa a $(1-q)^{n_1} q^{n_2}$, pero lo demás queda inalterado. Un razonamiento similar puede hacerse para la permuta de dos columnas, y ambos llevan a afirmar que:

$$\alpha_1(-Z_{exp}, p, q) = \alpha_1(+Z_{exp}, p, 1-q) = \alpha_1(+Z_{exp}, 1-p, q) \quad (2.10)$$

Para dos colas, y si se supone que $Z_{exp} > 0$, como

$$\alpha_2(Z_{exp}, p, q) = \alpha_1(Z_{exp}, p, q) + \alpha_1(-Z_{exp}, p, q)$$

entonces:

$$\alpha_2(Z_{exp}, p, q) = \alpha_1(Z_{exp}, p, q) + \alpha_1(Z_{exp}, p, 1-q) \quad (2.11)$$

Por otro lado, la permuta de filas por columnas no afecta a la ordenación (como ya apuntó Haber), y así:

$$\alpha_1(Z_{\text{exp}}, p, q) = \alpha_1(Z_{\text{exp}}, q, p) \quad (2.12)$$

de donde deduce que el cálculo del máximo puede restringirse a la región $0 < p \leq q$. Más aún, por la (2.10), si se considera que:

$$\alpha_2^* = \text{Max} \{ \alpha_1(Z_{\text{exp}}, p, q) + \alpha_1(Z_{\text{exp}}, p, 1-q) \} \quad (2.13)$$

bastará restringirse a $0 < q \leq 0,5$. Finalmente añade que $\alpha_2^* = 2 \alpha_1^*$ cuando Z_{exp} es pequeño y que $\alpha_2^* = \alpha_1^*$ si Z_{exp} es grande (lo que es sólo una observación de tipo experimental).

Las observaciones anteriores (justificadas sólo para el método Y) permiten disminuir las parejas posibles de valores (p,q), por lo que el tiempo de cómputo es menor. Pero la evaluación del máximo α^* hay que efectuarla numéricamente, por lo que la pregunta es: ¿con qué valores de p y de q se ensaya? Haber (1987) y Shuster (1992) calculan $\alpha(p,q)$ en valores de p y de q de 1% en 1% y asumen que α^* es el mayor de los $\alpha(p,q)$ así obtenidos. Kroll (1989) lo hace de !10% en 10%! El primero y el tercero evitan los valores $p, q < 5\%$ (?). Todo esto es bastante injustificable y será criticado más tarde.

Finalmente una vez evaluado $\alpha^* = \alpha(p_M, q_M)$, ¿donde suelen estar los (p_M, q_M) del máximo? Kroll (1989) indica que en general es $p_M = q_M = 0,5$. Shuster (1992) indica lo mismo pero sólo para valores $|Z_Y|$ pequeños. Este último añade que, restringiéndose a $p_M = q_M$, se obtienen casi los mismos máximos (pero dice no saber por qué) y que para $Z_Y = 0$ el máximo ocurre en $1/n \leq p, q \leq (n-1)/n$.

2.1.5. Valores P simplificados

En (1.26) ya se vió que, dada una tabla experimental, los valores P de la misma obtenidos por los Modelos I, II o III (α^*_I , α^*_II y α^*_III respectivamente) son cada vez más conservadores, pero tienen la ventaja de ser cada vez más fáciles de computar. Como α^*_I es difícil de calcular, algunos autores proponen sustituirlo por α^*_II o α^*_III como mal menor (puesto que no se obtienen significaciones falsas). Así, Pearson (1947) propone utilizar α^*_III y Barnard (1947) propone α^*_II , añadiendo que no se pierde

gran cosa con ello. Boschloo (1970) opina igual, pero matiza que el condicionamiento que implica el Modelo II (utilizado en lugar del I) puede hacerse en a_1 o en n_1 , por lo que propone tomar el α_{II}^* menor de los dos. Simbólicamente:

$$\alpha_{II}^* = \text{Min} \{ \alpha_{II}^*(a_1); \alpha_{II}^*(n_1) \} \quad (2.14)$$

En realidad α_{II}^* hay muchos (tantos como ordenaciones existen en el Modelo II), y así Boschloo (1970) hace su propuesta para la ordenación F y Silva and Martín (1995) para la M (ambas, naturalmente, dentro del Modelo II). Obsérvese que estos métodos no se proponen como ordenaciones en sí, sino como un modo simplificado de calcular α_{II}^* . Cada uno de ellos (que llamaremos por "**Fisher**" y "**Barnard1**") define una RC condicionada (doble o simplemente) que, de aplicarse la (2.2), podría aún ampliarse. Pero no es ese el objetivo.

Los investigadores del Modelo II conocen desde 1947 que el método "**Fisher**" es bastante conservador, y, por la (1.26), aún más lo será para el caso del Modelo I. Con todo la idea de "**Fisher**" seguía siendo atrayente por los cálculos mínimos (relativos) que requería, y así se desarrollaron diversos esfuerzos por mejorarla (respetando su origen). Al efecto hay dos posibilidades:

- 1) Recurrir al test aleatorizado de Tocher (1950), cuyas fronteras son las de Fisher más algún punto extra que es introducido por un procedimiento de aleatorización. Anteriormente ya se habló del carácter UMPU de tal test y de las circunstancias desfavorables que lo envuelven. Aquí será aludido por método de "**Tocher**" y sólo citado, pero no definido, pues será comentado muy circunstancialmente.
- 2) Recurrir a la versión P-mid del test exacto de Fisher. Definido por Lancaster (1952) como criterio general para variables discretas, Haber (1986) propuso adaptarlo al caso de Fisher. Ahora los puntos frontera de la RC entran en ella con la mitad de su probabilidad, lo que permite que pueda ser un poco más amplia que la de Fisher. El método será aludido por "**P-mid**".

Ambas propuestas permiten hacer α_{III}^* un poco más pequeño, pudiendo llegar a ser más competitivo con α_I^* . Desde este punto de vista, tanto "P-mid" como "Tocher" tienen igual intención que "Fisher" y "Barnard1".

En realidad Haber (1986) propuso el "P-mid" no como un método abreviado, sino como un método exacto justificado por su propia filosofía. Su intención era buscar un camino intermedio entre las filosofías condicionada e incondicionada y de ahí que Barnard (1989), Routledge (1992), Upton (1992), ... lo acogieran con entusiasmo (pues de algún modo conserva sus ideas condicionadas). Sin embargo muchos autores y editores lo contemplan como "basura" (sic), aunque públicamente no se atreven a decir tanto.

Como más adelante el P-mid será propuesto como método de ordenación, conviene recordar aquí su definición. De modo general el P-mid es "la semisuma del valor P de la observación real y de la anterior (más significativa)", lo que lleva a Haber (1986) a definirlo como $\{F1(x_1) + F1(x_1 + 1)\}/2$ en el caso de $H_1 \equiv \theta > 1$, con $F1(x_1)$ dado por la (1.10). Más explícitamente:

$$FM1(x_1) = \sum_{i=x_1+1}^s P(i) + \frac{1}{2} P(x_1) \quad (2.15)$$

en tanto que para el test de dos colas será (Hirji et al., 1991):

$$FM2(x_1) = \sum_{i=r}^{x'_1} P(i) + \sum_{i=x_1+1}^s P(i) + \begin{cases} \frac{1}{2} P(x_1) & \text{si } P(x_1) > P(x'_1) \\ 0 & \text{si } P(x_1) = P(x'_1) \end{cases} \quad (2.16)$$

con $P(i)$ dado por la (1.9), r y s por la (1.8) y x'_1 por la (1.12).

2.1.6. Evaluación de los diversos métodos

Ya se ha visto que en la literatura existen 6 métodos de ordenación (los **M**, **F**, **L**, **G**, **Z** e **Y**) -que en la práctica se reducen a 5 dado que **M** no es tomado en consideración por ningún autor- y otros 4 métodos de análisis abreviado (los "Fisher", "Barnard1", "Tocher" y "P-mid"). No existe ningún trabajo que compare todos estos

procedimientos, sino tan sólo aportaciones globales, aunque el más ambicioso es el de Haber (1987). De otro lado, para la comparación o análisis cada autor emplea una metodología distinta, siendo importante conocerla para saber de la fiabilidad de sus conclusiones. En lo que sigue se hace un breve análisis crítico de cada aportación (ordenadas temporalmente), distinguiendo las que son comparativas de las que no lo son.

Haber (1986) (No comparativa).-

Para tests de 1 o 2 colas, $\alpha = 5\%$ y $10 \leq n \leq 30$, forma la RC condicionada a través del "**P-mid**", calcula α^*_I para ella (no especifica cómo) y lo compara con el error objetivo α , concluyendo que el "**P-mid**" es conservador (especialmente en 1 cola).

En nuestra opinión el trabajo tiene el valor histórico de abordar por primera vez el Modelo I y proponer el P-mid para este ámbito. Aparte de ello sirve para ilustrar cómo el Modelo III "favorecido" por el P-mid da un α^* no lo suficientemente más pequeño que α^*_{III} como para desbordar el error α . El "P-mid", al igual que el Modelo III, sigue siendo conservador (aunque menos).

Haber (1987) (Comparativo).-

Analiza el caso de tests de 2 colas, $\alpha=5\%$ y $n=20, 30, 40$. Para cada situación determina las RC's incondicionadas para las ordenaciones **F**, **L**, **G** y **Z** y las RC's condicionadas para las ordenaciones de los métodos "**Fisher**" y "**Tocher**". Para cada RC calcula α^* a través de la (2.2) y para valores de $(p_{11}, p_{12})=(i/100; j/100)$ de 1% en 1% (pues usa como parámetros perturbadores a p_{11} y p_{12} en lugar de a, p y q). Finalmente, para cada RC, determina la potencia del test para todas las configuraciones $(p_{11}, p_{12}, p_{21}, p_{22})$ con $\sum p_{ij} = 1$ y $\theta = 2, 4, 8, 16, 32$ y 64 (con la condición restrictiva de que sea $p_{ij} \geq 5\%$). Esto le lleva a conocer la potencia del test en un θ dado en función de (p_{11}, p_{12}) , determinando finalmente como mecanismo comparativo:

$$\text{Min}_{P_{11}, P_{12}} \{ \text{Potencia}(\theta) \}$$

Todo ello le permite concluir que:

- 1) **Z** es el mejor método de ordenación en general, aunque **L** puede superarlo en los valores pequeños de θ y **G** en los grandes.
- 2) Todos los métodos (incluso "**Fisher**") son mejores que "**Tocher**".

Kroll (1989) (No comparativo).-

Para $\alpha = 5\%$, $n = 14, 19$ y 30 y test de χ^2 colas? (no lo deja claro su autor) calcula las RC incondicionada por **Z** y la condicionada por "**Fisher**", determinando $\alpha(p, q)$ para $(p, q) = 0,5(0,1)0,9$ y comparándolo con el error objetivo $\alpha = 5\%$, obteniendo que:

- a) **Z** da $\alpha(p, q) \leq 5\%$, acercándose al 5% en $p = q = 0,5$.
- b) "**Fisher**" es muy conservador.

Como se ve el trabajo es muy defectuoso por los escasos valores tomados para p y q , porque la conclusión a) no es tal (dado que, por propia construcción, **Z** ha de respetar α), y porque la conclusión b) es bien conocida ya para el Modelo II (tanto más para el I).

Shuster (1992) (Comparativo).-

El trabajo de Shuster no tenía por fin efectuar una comparación entre métodos, sino proporcionar una tabla de RC. Sin embargo la misma había de basarse en algún criterio de ordenación (que a la vez era el estadístico de test) y por ello hubo de decidir cuál era el más adecuado. Él optó por los clásicos **Z** e **Y**, inclinándose por el segundo pues observó que el **Z** alcanzaba el máximo en valores extremos de p y de q y proporcionaba valores P más elevados que el método **Y** (que será por tanto más potente).

Silva and Martín (1995) (Comparativo).-

Este trabajo no tenía por fin analizar el Modelo I, sino sólo dar tablas de RC para el Modelo II. Sin embargo intenta aprovecharlas para el Modelo I en base a método "**Barnard1**" descrito antes. Sin una prueba exhaustiva, sino tan sólo basándose en unas pocas tablas ejemplificadoras, tales autores comentan que el método "**Barnard1**" da más significaciones (es más potente) que el Y en las tablas con marginales muy desequilibrados (n_1/n_2 o a_1/a_2 muy pequeños o muy grandes).

De todo lo visto (a pesar de los defectos parciales que pueda tener) puede concluirse que:

- I) Los métodos "**P-mid**" y "**Fisher**" son conservadores (especialmente el segundo).
- II) Cualquier método incondicionado es mejor que el de "**Tocher**" (a pesar de ser UMPU), incluso lo es el de "**Fisher**".
- III) De entre los métodos **F**, **Z**, **L** y **G** el mejor es en general el **Z**. De entre **Z** y **Y** es mejor el **Y**, que resulta por tanto el método de ordenación idóneo.
- IV) El método condicionado parcial de "**Barnard1**" es preferible al **Y** en tablas con marginales desequilibrados.

2.1.7. Tablas y Programas

Las únicas tablas existentes, creadas específicamente para éste caso del Modelo I, son las de Shuster (1992). Presenta dos posibilidades:

- 1) Para $\alpha = 5\%$, proporciona los valores críticos del estadístico Z_Y .
- 2) Para valores seleccionados de Z_Y , proporciona los valores P alcanzados en ellos.

en ambos casos para valores $n = 10(1)100$ y test de 1 o 2 colas.

Ya se han citado las tablas de Silva and Martín (1995) para el caso de dos proporciones y que ellos adaptan al caso actual mediante el criterio conservador que

hemos denominado por "**Barnard1**". Ellas son válidas para 1 y 2 colas, errores $\alpha = 1\%$, 5% y 10% y valores $n_1 \leq n_2 \leq 25$, siendo tablas de RC.

En cuanto a los programas, no hay mucho donde elegir. Por un lado Shuster (1992) ofrece a la venta su programa para calcular valores P por el método Y para $n \leq 100$. Por otro el programa para los valores P por "Barnard1" puede encontrarse en las siguientes direcciones electrónicas (Silva and Martín):

<http://www.jiscmail.ac.uk/files/EXACT-STATS/smpexe.uue>

<http://www.ugr.es/~bioest/SMP.EXE>

Finalmente, esta otra dirección (Berger):

<http://www.stat.ncsu.edu/~berger/tables.html>

proporciona los valores P para las ordenaciones **Z** y **F**.

2.2. FORMALIZACIÓN DEL PROBLEMA*

2.2.1. Introducción

Ya se ha indicado que Barnard (1947) prestó mucha atención al caso del Modelo II y prácticamente ninguna al del Modelo I, en particular no comentó nada acerca de si las propiedades de convexidad y simetría del Modelo II serían extensibles, y en qué sentido, al Modelo I. Los pocos autores que desde entonces han abordado el problema tampoco han señalado mucho más, disponiéndose tan sólo de afirmaciones inconexas que ya se han señalado en el apartado anterior.

Ya se ha indicado también que el Modelo II está bastante exhaustivamente tratado, conociéndose cómo formar simplificadaamente el espacio muestral (Barnard, 1947), las condiciones a verificar por la RC (Barnard, 1947), una gama bastante completa de criterios de ordenación y el óptimo de ella (Martín and Silva, 1994) y diversas propiedades para simplificar el cálculo del máximo (Silva and Martín, 1997). El objetivo del resto de este Capítulo es obtener resultados paralelos para el caso del Modelo I.

2.2.2. Contrucción simplificada del espacio muestral: Equivalencia y Simetría.

Ya se han indicado (sección 2.1.2) los diversos modos posibles de aludir al espacio muestral y que, por comodidad y paralelismo con el bien conocido Modelo II, conviene adoptar como variables del problema a la terna (x_1, x_2, n_1) . Esto permitirá además confrontar las propiedades que aquí se obtengan con las ya conocidas del Modelo II. El espacio muestral constará pues de todos los rectángulos de lados $(n_1; n-n_1=n_2)$ con todos los puntos (x_1, x_2) de su interior. Los valores $a_1=x_1+x_2$ son las diagonales de tales rectángulos y se corresponden con el espacio muestral para el test exacto de Fisher. Así:

$$0 \leq n_1 \leq n; 0 \leq x_1 \leq n_1; 0 \leq x_2 \leq n_2 \quad (2.17)$$

Esto es lo que se representó en la Figura 2.1(c) que ahora mejoramos visualmente con la Figura 2.2(a) al incluir la imagen de los "cuadrados de Barnard".

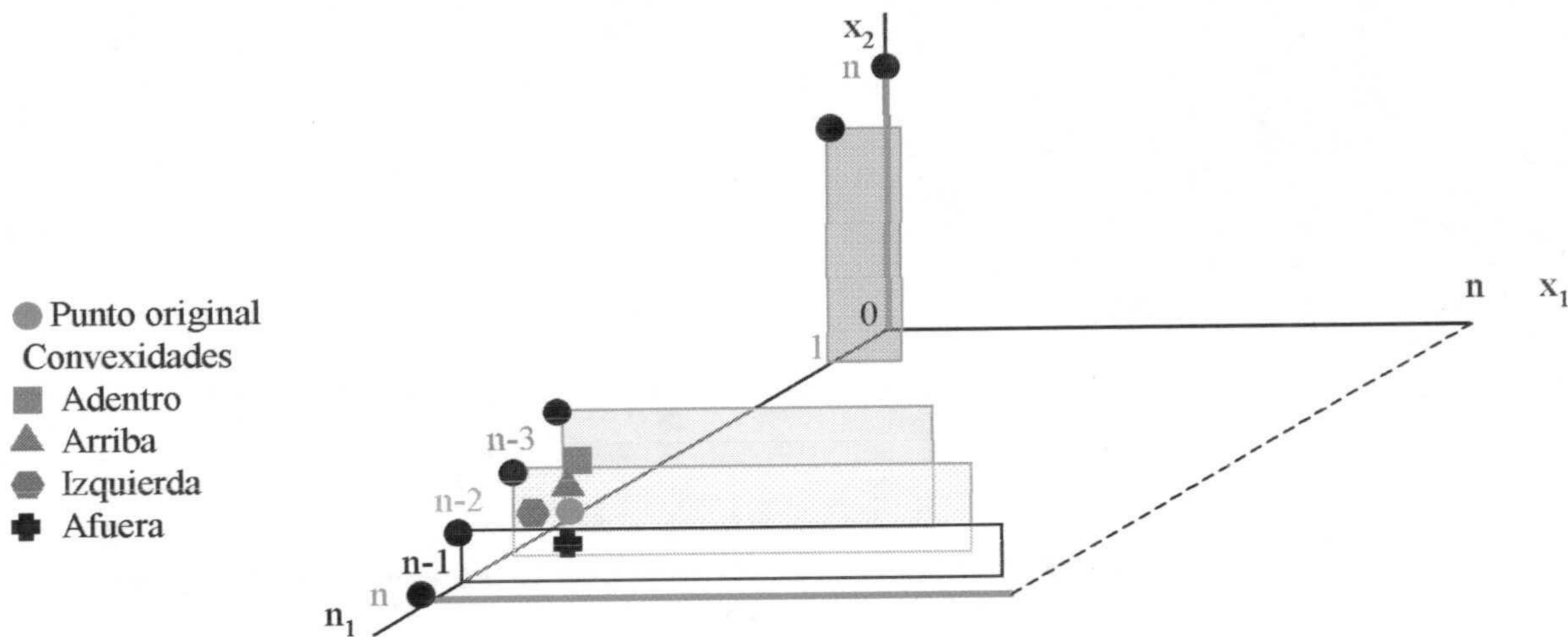
Barnard (1947) dedicó gran atención a las propiedades que debe verificar cualquier RC en el caso de dos proporciones (convexidad y simetría), pero no señaló nada para el caso actual. Otros autores han señalado de pasada algunas propiedades para casos particulares, pero no hay ningún estudio general y exhaustivo al respecto. Como regla general, una RC debe constar de puntos "lógicos" que favorezcan una máxima potencia, pero si con una dimensión (x_1 para el test exacto de Fisher) esto es relativamente sencillo, con más de una dimensión (como en el caso actual) el asunto no es obvio por causa de la ausencia de razón de verosimilitud monótona (Lehmann, 1994). Con tal fin, supongamos que la Tabla E1 de la Tabla 2.1 es la experimental (ella es una versión simplificada de la tabla original 1.1(a)). En lo que sigue toda la argumentación alude a las tablas contempladas en esta Tabla 2.1.

Las Tablas E2 a E4 se obtienen de la E1 permutando los elementos de una o dos de sus diagonales. Las cuatro tablas tienen iguales números en su interior, igual valor de $OR = x_1y_2/x_2y_1$ (la razón del producto cruzado muestral) e iguales marginales, siendo pues *equivalentes*. Si una tabla entra en la RC, todas deben entrar, y así los puntos entran en la RC de 4 en 4, con la excepción de los casos en que $a_1=n_2 \neq n_1$ o $a_1=n_1 \neq n_2$ (en los que sólo hay 2 tablas equivalentes) y del caso $a_1=n_1=n_2$ (en que sólo hay una tabla equivalente). Esto divide el espacio muestral casi por cuatro. La

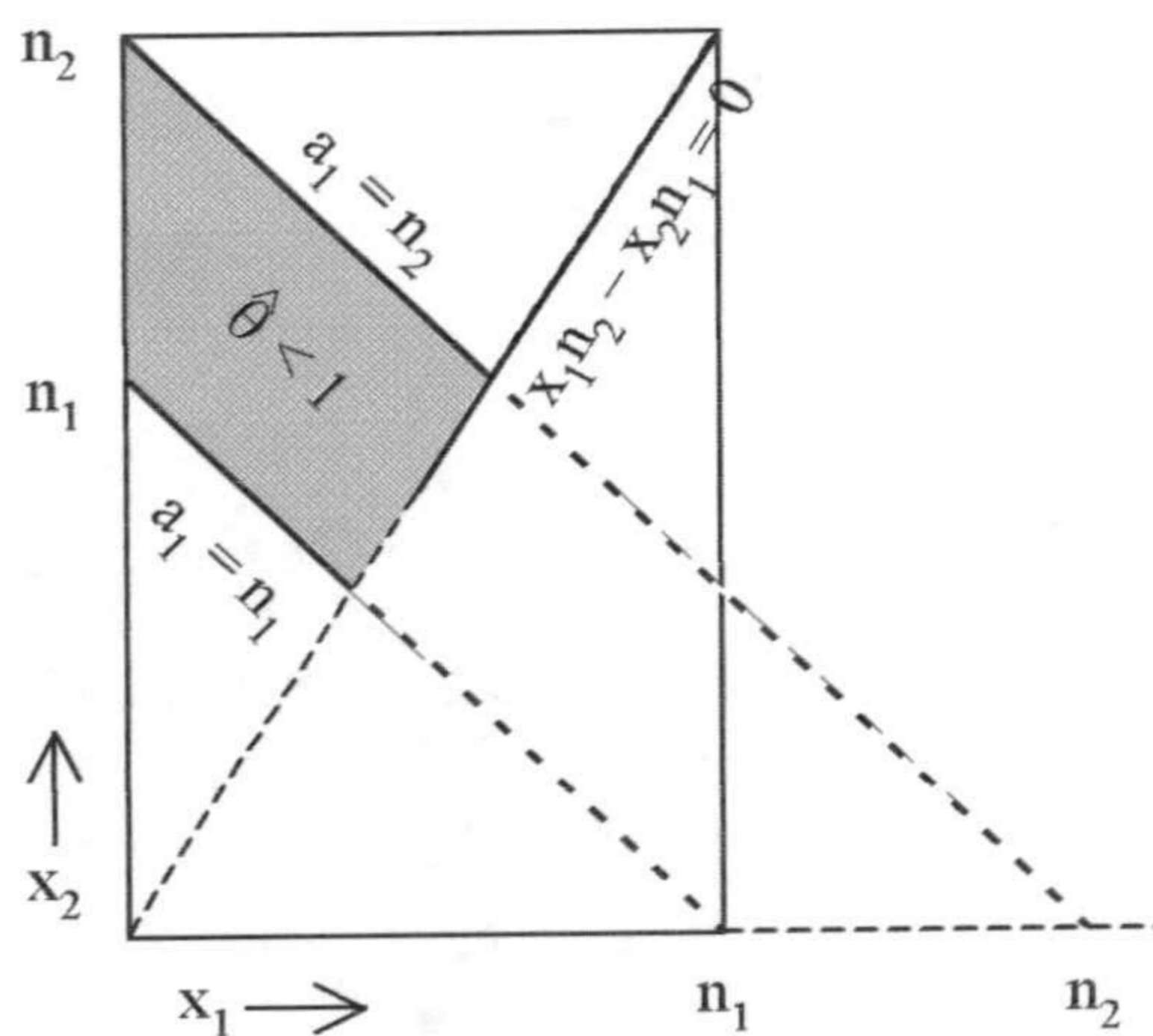
Figura 2.2

Espacios muestrales en base a la terna (x_1, x_2, n_1)

(a) Espacio original (● alude a los atractores)



(b) Una cuadrícula del espacio simplificado (zona sombreada)



(c) Espacio simplificado

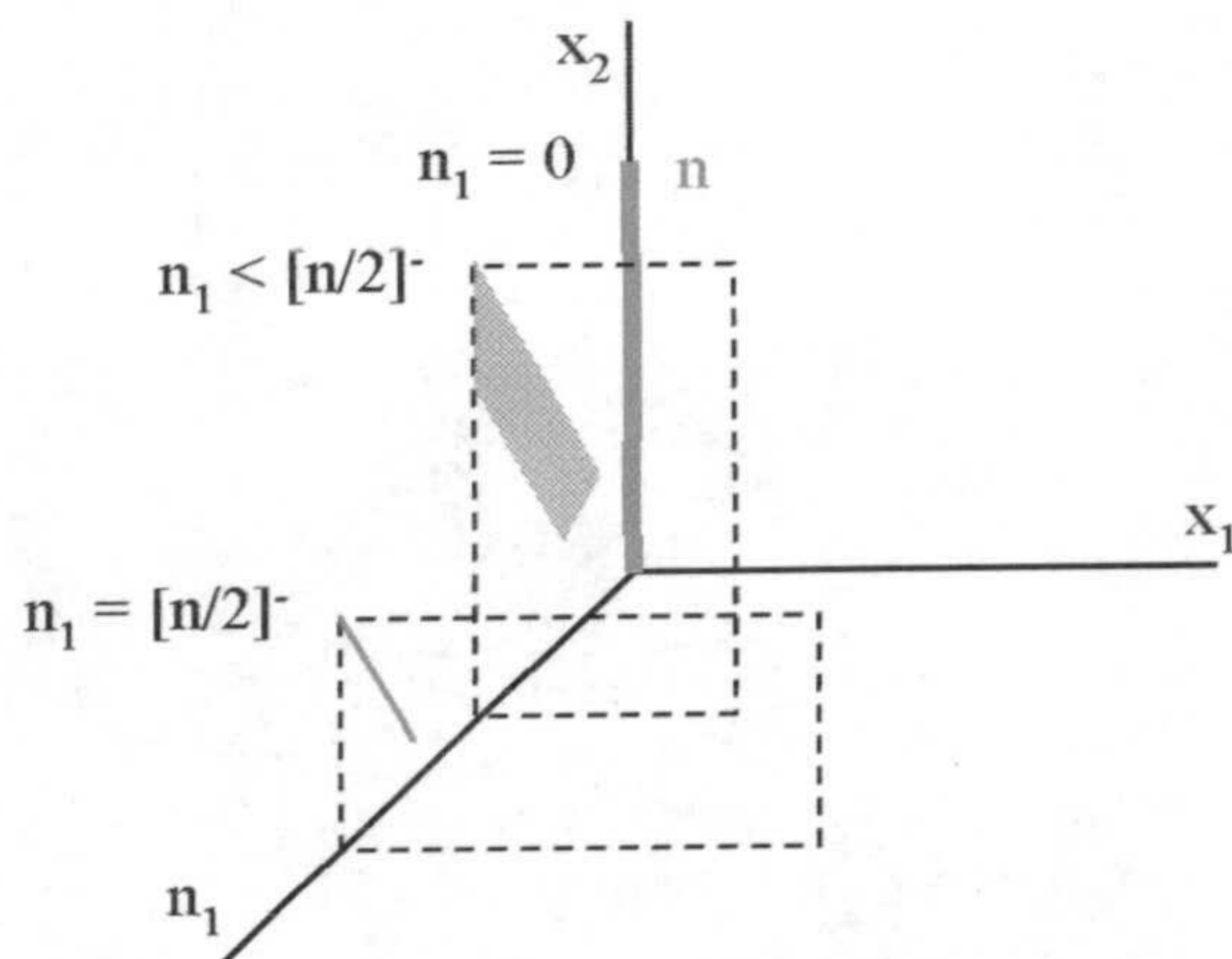


Tabla 2.1

Presentación general de los datos de una tabla 2x2 y propiedades para 0 marginales fijados (Modelo I)
(en filas: presencia o ausencia de B; en columnas: presencia o ausencia de A).

Tablas equivalentes

E1	E2	E3	E4																																				
<table border="1" style="margin: auto;"> <tr><td>x₁</td><td>y₁</td><td>n₁</td></tr> <tr><td>x₂</td><td>y₂</td><td>n₂</td></tr> <tr><td>a₁</td><td>a₂</td><td>n</td></tr> </table>	x ₁	y ₁	n ₁	x ₂	y ₂	n ₂	a ₁	a ₂	n	<table border="1" style="margin: auto;"> <tr><td>y₂</td><td>x₂</td><td>n₂</td></tr> <tr><td>y₁</td><td>x₁</td><td>n₁</td></tr> <tr><td>a₂</td><td>a₁</td><td>n</td></tr> </table>	y ₂	x ₂	n ₂	y ₁	x ₁	n ₁	a ₂	a ₁	n	<table border="1" style="margin: auto;"> <tr><td>x₁</td><td>x₂</td><td>a₁</td></tr> <tr><td>y₁</td><td>y₂</td><td>a₂</td></tr> <tr><td>n₁</td><td>n₂</td><td>n</td></tr> </table>	x ₁	x ₂	a ₁	y ₁	y ₂	a ₂	n ₁	n ₂	n	<table border="1" style="margin: auto;"> <tr><td>y₂</td><td>y₁</td><td>a₂</td></tr> <tr><td>x₂</td><td>x₁</td><td>a₁</td></tr> <tr><td>n₂</td><td>n₁</td><td>n</td></tr> </table>	y ₂	y ₁	a ₂	x ₂	x ₁	a ₁	n ₂	n ₁	n
x ₁	y ₁	n ₁																																					
x ₂	y ₂	n ₂																																					
a ₁	a ₂	n																																					
y ₂	x ₂	n ₂																																					
y ₁	x ₁	n ₁																																					
a ₂	a ₁	n																																					
x ₁	x ₂	a ₁																																					
y ₁	y ₂	a ₂																																					
n ₁	n ₂	n																																					
y ₂	y ₁	a ₂																																					
x ₂	x ₁	a ₁																																					
n ₂	n ₁	n																																					

Tablas simétricas a las de arriba

S1	S2	S3	S4																																				
<table border="1" style="margin: auto;"> <tr><td>y₁</td><td>x₁</td><td>n₁</td></tr> <tr><td>y₂</td><td>x₂</td><td>n₂</td></tr> <tr><td>a₂</td><td>a₁</td><td>n</td></tr> </table>	y ₁	x ₁	n ₁	y ₂	x ₂	n ₂	a ₂	a ₁	n	<table border="1" style="margin: auto;"> <tr><td>x₂</td><td>y₂</td><td>n₂</td></tr> <tr><td>x₁</td><td>y₁</td><td>n₁</td></tr> <tr><td>a₁</td><td>a₂</td><td>n</td></tr> </table>	x ₂	y ₂	n ₂	x ₁	y ₁	n ₁	a ₁	a ₂	n	<table border="1" style="margin: auto;"> <tr><td>x₂</td><td>x₁</td><td>a₁</td></tr> <tr><td>y₂</td><td>y₁</td><td>a₂</td></tr> <tr><td>n₂</td><td>n₁</td><td>n</td></tr> </table>	x ₂	x ₁	a ₁	y ₂	y ₁	a ₂	n ₂	n ₁	n	<table border="1" style="margin: auto;"> <tr><td>y₁</td><td>y₂</td><td>a₂</td></tr> <tr><td>x₁</td><td>x₂</td><td>a₁</td></tr> <tr><td>n₁</td><td>n₂</td><td>n</td></tr> </table>	y ₁	y ₂	a ₂	x ₁	x ₂	a ₁	n ₁	n ₂	n
y ₁	x ₁	n ₁																																					
y ₂	x ₂	n ₂																																					
a ₂	a ₁	n																																					
x ₂	y ₂	n ₂																																					
x ₁	y ₁	n ₁																																					
a ₁	a ₂	n																																					
x ₂	x ₁	a ₁																																					
y ₂	y ₁	a ₂																																					
n ₂	n ₁	n																																					
y ₁	y ₂	a ₂																																					
x ₁	x ₂	a ₁																																					
n ₁	n ₂	n																																					

Convexidad (respecto de E1) a la

<i>Izquierda</i>	<i>Arriba</i>	<i>Adentro</i>	<i>Afuera</i>																																				
CI	CR	CD	CF																																				
<table border="1" style="margin: auto;"> <tr><td>x₁-1</td><td>y₁+1</td><td>n₁</td></tr> <tr><td>x₂</td><td>y₂</td><td>n₂</td></tr> <tr><td>a₁-1</td><td>a₂+1</td><td>n</td></tr> </table>	x ₁ -1	y ₁ +1	n ₁	x ₂	y ₂	n ₂	a ₁ -1	a ₂ +1	n	<table border="1" style="margin: auto;"> <tr><td>x₁</td><td>y₁</td><td>n₁</td></tr> <tr><td>x₂+1</td><td>y₂-1</td><td>n₂</td></tr> <tr><td>a₁+1</td><td>a₂-1</td><td>n</td></tr> </table>	x ₁	y ₁	n ₁	x ₂ +1	y ₂ -1	n ₂	a ₁ +1	a ₂ -1	n	<table border="1" style="margin: auto;"> <tr><td>x₁-1</td><td>y₁</td><td>n₁-1</td></tr> <tr><td>x₂+1</td><td>y₂</td><td>n₂+1</td></tr> <tr><td>a₁</td><td>a₂</td><td>n</td></tr> </table>	x ₁ -1	y ₁	n ₁ -1	x ₂ +1	y ₂	n ₂ +1	a ₁	a ₂	n	<table border="1" style="margin: auto;"> <tr><td>x₁</td><td>y₁+1</td><td>n₁+1</td></tr> <tr><td>x₂</td><td>y₂-1</td><td>n₂-1</td></tr> <tr><td>a₁</td><td>a₂</td><td>n</td></tr> </table>	x ₁	y ₁ +1	n ₁ +1	x ₂	y ₂ -1	n ₂ -1	a ₁	a ₂	n
x ₁ -1	y ₁ +1	n ₁																																					
x ₂	y ₂	n ₂																																					
a ₁ -1	a ₂ +1	n																																					
x ₁	y ₁	n ₁																																					
x ₂ +1	y ₂ -1	n ₂																																					
a ₁ +1	a ₂ -1	n																																					
x ₁ -1	y ₁	n ₁ -1																																					
x ₂ +1	y ₂	n ₂ +1																																					
a ₁	a ₂	n																																					
x ₁	y ₁ +1	n ₁ +1																																					
x ₂	y ₂ -1	n ₂ -1																																					
a ₁	a ₂	n																																					

Tablas que se pueden introducir, tras la E1, en base a la convexidad

\overline{CI}	\overline{CR}	\overline{CD}	\overline{CF}																																				
<table border="1" style="margin: auto;"> <tr><td>x₁+1</td><td>y₁-1</td><td>n₁</td></tr> <tr><td>x₂</td><td>y₂</td><td>n₂</td></tr> <tr><td>a₁+1</td><td>a₂-1</td><td>n</td></tr> </table>	x ₁ +1	y ₁ -1	n ₁	x ₂	y ₂	n ₂	a ₁ +1	a ₂ -1	n	<table border="1" style="margin: auto;"> <tr><td>x₁</td><td>y₁</td><td>n₁</td></tr> <tr><td>x₂-1</td><td>y₂+1</td><td>n₂</td></tr> <tr><td>a₁-1</td><td>a₂+1</td><td>n</td></tr> </table>	x ₁	y ₁	n ₁	x ₂ -1	y ₂ +1	n ₂	a ₁ -1	a ₂ +1	n	<table border="1" style="margin: auto;"> <tr><td>x₁+1</td><td>y₁</td><td>n₁+1</td></tr> <tr><td>x₂-1</td><td>y₂</td><td>n₂-1</td></tr> <tr><td>a₁</td><td>a₂</td><td>n</td></tr> </table>	x ₁ +1	y ₁	n ₁ +1	x ₂ -1	y ₂	n ₂ -1	a ₁	a ₂	n	<table border="1" style="margin: auto;"> <tr><td>x₁</td><td>y₁-1</td><td>n₁-1</td></tr> <tr><td>x₂</td><td>y₂+1</td><td>n₂+1</td></tr> <tr><td>a₁</td><td>a₂</td><td>n</td></tr> </table>	x ₁	y ₁ -1	n ₁ -1	x ₂	y ₂ +1	n ₂ +1	a ₁	a ₂	n
x ₁ +1	y ₁ -1	n ₁																																					
x ₂	y ₂	n ₂																																					
a ₁ +1	a ₂ -1	n																																					
x ₁	y ₁	n ₁																																					
x ₂ -1	y ₂ +1	n ₂																																					
a ₁ -1	a ₂ +1	n																																					
x ₁ +1	y ₁	n ₁ +1																																					
x ₂ -1	y ₂	n ₂ -1																																					
a ₁	a ₂	n																																					
x ₁	y ₁ -1	n ₁ -1																																					
x ₂	y ₂ +1	n ₂ +1																																					
a ₁	a ₂	n																																					

equivalencia de la Tabla E3 fue citada por Haber (1986) y la de la Tabla E2 por Silva and Martín (1997) en el contexto de dos proporciones. Todo esto da lugar a la propiedad E de *equivalencia*. Las excepciones citadas a la regla de que los puntos han

de entrar de "4 en 4" son fáciles de probar:

- i) Diagonal principal igual $\Leftrightarrow x_1 = y_2 \Leftrightarrow x_1 + x_2 = y_2 + x_2 \Leftrightarrow a_1 = n_2$, en cuyo caso sólo hay 2 tablas equivalentes.
- ii) Diagonal secundaria igual $\Leftrightarrow x_2 = y_1 \Leftrightarrow x_2 + x_1 = y_1 + x_1 \Leftrightarrow a_1 = n_1$, en cuyo caso sólo hay 2 tablas equivalentes.
- iii) Ambas diagonales iguales $\Leftrightarrow x_1 = y_2$ y $x_2 = y_1 \Leftrightarrow a_1 = n_2 = n_1 = n/2$, en cuyo caso solo hay 1 tabla equivalente.

Todo ello quiere decir que los puntos entran en la RC en grupos de:

4	si	$a_1 \neq n_1$	y	$a_1 \neq n_2$
2	si	$a_1 \neq n_1$	y	$a_1 = n_2$
2	si	$a_1 = n_1$	y	$a_1 \neq n_2$
1	si	$a_1 = n_1$	y	$a_1 = n_2 (= n/2)$

lo que proporciona la regla simplificada dada arriba.

Para 1 marginal fijado, Barnard (1947) indicó que en un test de dos colas la Tabla E1 tenía una tabla simétrica en la otra cola, la Tabla S1, que habría de entrar en la RC junto con la primera y que se obtiene permutando las dos columnas de ésta. Ahora, por haber 4 tablas equivalentes, hay 4 tablas *simétricas*, las Tablas S1 a la S4 (con las excepciones de antes). Fijadas las primeras, se conocen las segundas, y el espacio muestral queda ahora dividido (de modo extra) por dos. Se observa que las tablas simétricas tienen un valor de odds-ratio inverso ($\hat{\theta}^{-1}$) al de las tablas originales ($\hat{\theta}$). Todo esto da lugar a la propiedad S de *simetría*, que, junto a la anterior, implica que los puntos entran en la RC en grupos de 8, 4 o 2 en las circunstancias descritas arriba. De un modo general (salvo las excepciones comentadas) cada tabla tiene 3 tablas equivalentes y 4 tablas simétricas.

Las condiciones E y S permiten generar el espacio muestral de un modo simplificado. Así, puesto que la tabla E1 es la misma (en esencia) que las E2 a E4,

tomemos un representante canónico de todas ellas eligiendo para tal fin a la que tiene un menor valor del total de la primera fila \Rightarrow

$$0 \leq n_1 \leq [n/2]^- \quad (2.18)$$

con $[x]^-$ aludiendo a la parte entera de x .

Para cada n_1 se tendrá una cuadrícula como la de la Figura 2.2(b). Dentro de ella hay que limitarse a la zona con $\hat{\theta} < 1$ ($\equiv x_1 y_2 / x_2 y_1 < 1 \equiv x_1 y_2 - x_2 y_1 = U < 0 \equiv \hat{p}_1 = x_1/n_1 < \hat{p}_2 = x_2/n_2$) pues la otra zona se obtendrá aplicando la propiedad de simetría $\Rightarrow x_1 n_2 - x_2 n_1 < 0 \Rightarrow x_1 < E = a_1 n_1 / n \Rightarrow$

$$0 \leq x_1 < [a_1 n_1 / n]^- \quad (2.19)$$

y las tablas con $\hat{\theta} = 1$ nunca entrarán en la RC. Finalmente, no todas las diagonales a_1 están permitidas, por causa del convenio que llevó a la (2.18), y así, como n_1 era el menor marginal,

$$\begin{cases} a_1 \geq n_1 \\ a_2 \geq n_1 \Rightarrow n - a_1 \geq n_1 \Rightarrow a_1 \leq n_2 = n - n_1 \end{cases} \Rightarrow n_1 \leq a_1 \leq n - n_1 \quad (2.20)$$

Las tres expresiones (2.18), (2.19) y (2.20) permiten afirmar que el espacio muestral se obtiene limitándose a los valores:

$$0 \leq n_1 \leq [n/2]^- \quad ,, \quad n_1 \leq a_1 \leq n - n_1 \quad ,, \quad 0 \leq x_1 \leq [a_1 n_1 / n]^- \quad (2.21)$$

pues los demás se obtienen por equivalencia o por simetría (si el test es de dos colas), lo que permite generar el espacio muestral de un modo simplificado y sin repeticiones. La Figura 2.2(b) informa del espacio muestral de una cuadrícula verificando las (2.19) y (2.20); la Figura 2.2(c) lo extiende a través de la (2.18). Con respecto a la primera, obsérvese que cada tabla de la zona sombreada da lugar a otras

- i) 3 si el punto no es de las fronteras $a_1 = n_1$ o $a_1 = n_2$
- ii) 1 si el punto es de $a_1 = n_1$ o $a_1 = n_2$.
- iii) 0 si el punto es de $a_1 = n_1 = n_2$ (las dos fronteras coinciden), lo que sólo puede ocurrir cuando n es par.

y que cada tabla así generada tiene otra simétrica con $\hat{\theta}' = \hat{\theta}^{-1} > 1$.

2.2.3. Condiciones a verificar por la región crítica: Convexidad

Supongamos que el test es de una cola con $H_1: \theta < 1$ (asociación negativa). Por definición (para el caso de marginales no nulos), $\hat{\theta}$ es el cociente de odds $(x_1/y_1) \div (x_2/y_2) = O_1/O_2 = \hat{\theta} = (x_1/x_2) \div (y_1/y_2)$, y, si una tabla está en la RC, es porque $\hat{\theta}$ fue pequeña e inferior a 1. Si otra tabla tiene un valor $O'_1/O'_2 = \hat{\theta}' < \hat{\theta}$ con $\{O'_1 < O_1 \text{ y } O'_2 = O_2\}$ o $\{O'_1 = O_1 \text{ y } O'_2 > O_2\}$ y los marginales de los odds son iguales, entonces dicha tabla debe pertenecer también a la RC (pues da un valor de $\hat{\theta}$ más extremo y uno de sus odds, y los totales que los definen, son constantes). Esto obliga a que si la Tabla E1 es de la RC, también tienen que serlo las Tablas CI a la CF, lo que define las cuatro propiedades de *convexidad* aludidas por ellas. Las dos primeras son las señaladas por Barnard (1947), por un razonamiento similar, para el caso de dos proporciones: en el rectángulo (n_1, n_2) , el punto (x_1, x_2) obliga al de su izquierda y al de arriba (de ahí los nombres propuestos). Las otras dos convexidades son propias de este caso y aluden al rectángulo de "adentro" (n_1-1, n_2+1) y al de "afuera" (n_1+1, n_2-1) . Un modo útil de ver las propiedades de convexidad es en sentido inverso: si una tabla como la E1 es de la RC, ella fuerza a que la próxima tabla a introducir en la RC sea una de las \overline{CI} a \overline{CF} (obtenidas por las reglas contrarias a las que definen las Tablas CI a CF) si son compatibles con la convexidad. Como excepción, los puntos con algún marginal nulo no entran nunca en la RC (ó, en todo caso, son los últimos en entrar). Así pues los puntos de la RC deben verificar las condiciones C de convexidad cuya representación gráfica se da en la Fig. 2.2(a).

A fin de enmarcar estos resultados en el problema general de las tablas 2×2 , hagamos algunas observaciones. Con 2 marginales fijados (test exacto de Fisher) sólo hay una variable aleatoria (x_1 , por ejemplo) y una única propiedad de la RC: la convexidad a la izquierda (Davis, 1986). Con 1 marginal fijado hay dos variables aleatorias (x_1 y x_2 , por ejemplo) y existen las propiedades de simetría y de convexidad a la izquierda y arriba (Barnard, 1947). Con 0 marginales fijados las variables son tres (x_1, x_2 y n_1 , por ejemplo) y se ha justificado que existen dos propiedades extras de

convexidad y las nuevas propiedades de equivalencia.

2.2.4. Formación de la región crítica

2.2.4.1. Introducción

Las condiciones E, S y C inducen un orden parcial en los puntos del espacio muestral. Pero para obtener un orden total es preciso definir una regla X extra que efectúe el desempate en aquellos puntos que verifiquen E, S y C, dando lugar así al método de ordenación ESCX (aunque en adelante, por comodidad, será llamado método X a secas). Tales métodos de ordenación serán los ya conocidos de la literatura (y reseñados en 2.1.3) y otros nuevos que aquí se propongan. De aquellos sólo se aludirá a los que se conoce que son competitivos (y se les redefinirá por uniformidad de la exposición).

De un modo general, la introducción de un nuevo punto en la RC puede hacerse en base a un criterio que tenga en cuenta a los puntos que ya pertenecen a ella (*método CON memoria*) o a uno basado exclusivamente en el punto a introducir (*método SIN memoria*). De modo general los métodos CON memoria deben ser más potentes que los sin memoria (por conservar más información) pero también serán más complejos de aplicar.

2.2.4.2. Métodos de ordenación "CON memoria"

Ya se ha citado anteriormente la propuesta de Barnard (1947) del método **M**. De un modo corto éste se definía así: dada una RC y dados los puntos candidatos a ampliar la misma (los que impone la condición C), el siguiente punto a entrar (con sus equivalentes y simétricos, si ha lugar) es aquel que hace mínimo el error $\alpha^* = \max_{p,q} \alpha(p,q)$ de la RC ampliada (criterio **M** del mínimo), lo que dará lugar al test más potente pero obliga a determinar numerosas veces la (2.2) (aumentando desmesuradamente el tiempo de cómputo). Para evitarlo, y por paralelismo a lo propuesto por Martín and Silva (1994), puede definirse el criterio **M'** consistente en cambiar el α^* de la definición de **M** por su estimación $\hat{\alpha}^* = \alpha(\hat{p}, \hat{q})$, con $\hat{p} = a_1/n$ y $\hat{q} = n_1/n$ obtenidos en el punto candidato a ser introducido en la RC: cuando haya

varios, elegir como representante a aquel que ocasione un valor más grande de $\alpha(\hat{p}, \hat{q})$.

2.2.4.3. Métodos de ordenación "SIN memoria"

De un modo general, los métodos de ordenación basados en el clásico estadístico chi-cuadrado suelen ser efectivos. En el caso del Modelo I esto ya se comentó en 2.1.6; en el caso del Modelo II fue notado por Martín and Silva (1994). Sin embargo el test chi-cuadrado admite varias versiones dependiendo de la c.p.c. que se utilice, siendo su forma general:

$$\frac{\{|x_1y_2 - x_2y_1| - c\}^2}{a_1a_2n_1n_2} \quad (2.22)$$

con c aludiendo a la c.p.c. (las demás versiones con factor diferente de n no dan lugar a ordenaciones distintas). Al respecto:

- a) Cuando $c=0$ o $c=n/2$ se obtienen las χ^2_Z sin c.p.c o la χ^2_Y con la c.p.c. de Yates, lo que daba lugar a los métodos **Z** e **Y** de Haber (1987) y Shuster (1992) respectivamente. La precaución de Shuster para el caso de $E-0,5 \leq x_1 \leq E + 0,5$ no es preciso considerarla pues tales puntos se encontrarán cercanos a la diagonal secundaria y nunca formarán parte de una RC construida para los α usuales.
- b) La c.p.c. de $c=n/2$ es típica del Modelo III. Más apropiado parece contemplar la c.p.c. de $c=0,5$ que es la apropiada para el Modelo I o la de $c=n/4$ que es propia del Modelo II más cercano al anterior (Pirie and Hamdan, 1972). Ellas darán lugar a los métodos de ordenación **PH1** y **PH2** respectivamente.
- c) Las dos ordenaciones últimas son nuevas propuestas actuales basadas en c.p.c. conocidas. La ordenación **PH3** que sigue está basada en una c.p.c. nueva cuya prueba puede verse en el próximo Capítulo. Ahora $c = [n^2/8]^- / \bar{n}$ con \bar{n} dado por:

$$\bar{n} = \frac{(n+1)(n+2)(n+3) + (n+1)(3n-33)}{48} + f(n) \quad (2.23)$$

en donde:

$$f(n) = \begin{cases} 1/4 & \text{si } n \text{ es impar y primo} \\ (7n+15)/24 & \text{si } n \text{ es impar y no primo} \\ (15n+3)/16 & \text{si } n \text{ es par y múltiplo de 4} \\ (15n+9)/16 & \text{si } n \text{ es par y no múltiplo de 4} \end{cases}$$

Otros métodos estudiados son los basados en el test exacto de Fisher. Uno de ellos es el indicado de Haber (1987) que propuso lo que entonces se llamó método F basado en las ordenaciones que inducía el valor P de una o de dos colas. Pero ello conviene matizarlo. A tal efecto recuérdese que el orden basta con reseñarlo para el test de una cola (pues la RC para dos colas se deduce de la condición S) y así el método que ordena en base a él- expresión (1.10) -le llamaremos por **F1**, en tanto que al que ordena en base al valor P de dos colas -expresión (1.11)- le llamaremos por **F2**. El método **F1** es válido como ordenación para el test de 1 o 2 colas; el **F2** sólo para el de 2 colas.

De otro lado ya se vió cómo Haber (1986) estudió el método que se llamó "P-mid", pero no lo propuso como método de ordenación. Con respecto a él se pueden hacer las mismas propuestas que para F, obteniendo así los nuevos métodos **FM1** y **FM2** que ordenan los puntos en base a las (2.15) y (2.16) respectivamente. De nuevo, **FM1** es válido para 1 o 2 colas, pero **FM2** lo es sólo para dos colas.

Finalmente, señalar que las condiciones E y S (y, circunstancialmente, el método X, si produce empates) hacen que los puntos entren en la RC no de uno en uno, sino en grupos que pueden llegar a ser de 8 puntos o más. Esto hace que los saltos de α_i^* puedan ser excesivos, dando lugar así a un test conservador. Un modo de paliar algo el problema citado antes (típico de los tests con variables discretas) es a través del P-mid de Lancaster (1952). Ahora el error $\alpha(p,q)$ es el definido anteriormente menos la mitad de la probabilidad del último grupo de puntos que se hayan introducido en la RC. Se ve pues que hay una versión P-mid distinta por cada

versión X definida arriba. Sin embargo esta definición no es frecuente en los métodos incondicionados (como el actual), y así en lo que sigue no se alude más a ella (aunque todas las propiedades de más adelante le son también de aplicación).

2.2.4.4. Atractores

Cualquier método ESCX define perfectamente el orden en que los puntos deben ir entrando en la RC, pero resulta sumamente conveniente conocer de antemano cuál es el primero o primeros a introducir en ella. Para el Modelo III el punto de inicio (ó *atractor* de la RC) es $x_1 = \max \{0; a_1 - n_2\}$ (Fisher, 1935). Para el Modelo II el punto de inicio es $(x_1 = 0; x_2 = n_2)$ propuesto de Barnard (1947) por la lógica del mismo y porque se deduce de su método M. Para el Modelo I, los puntos del caso anterior, para valores $1 \leq n_1 \leq n-1$, proporcionan un conjunto de puntos que en formato reducido son $RC_0 = \{(0, n-1, 1); (0, n-2, 2); \dots; (0, n - [n/2]^-, [n/2]^+)\}$ y que aluden a los puntos del córner izquierdo de cada rectángulo (todos ellos están en la diagonal $a_1 = n_2$, y así sólo tienen un equivalente, salvo el último punto que no tiene ningún equivalente cuando n es par) y que se representan en la Fig. 2.2(a). Las condiciones C obligan ahora a que los puntos de inicio sean al menos dos de los anteriores y en valores consecutivos de n_1 ; sin embargo, dado que todos ellos dan lugar a una asociación negativa máxima ($\hat{\theta} = 0$) con odds $O_1 = 0$ y $O_2 = \infty$, puede opinarse que todos esos puntos deben entrar en la RC a la vez y en el primer paso, constituyendo así lo que será llamada en adelante como RC inicial (esto, en esencia, implica imponer una condición más al método actual, aparte de las E, S y C).

Con todo lo anterior, ya puede resumirse la regla práctica final para formar una RC a un error objetivo de α . De un modo general el proceso consiste en la repetición de los siguientes pasos: 1) Se parte de una RC_i cuyo valor de α^* , obtenido por la (2.2), es α_i^* ; 2) A los puntos de RC_i se les aplican las condiciones \overline{CI} a \overline{CF} para obtener el conjunto de puntos A_i candidatos a entrar en la RC_i ; 3) Se aplica el criterio X a los puntos de A_i , seleccionando así el punto que se introduce en la RC_i para formar la nueva RC_{i+1} . El proceso comienza con los puntos de la RC_0 inicial de los atractores, si se acepta la propiedad de los atractores, o con el conjunto vacío (en cuyo

caso los atractores, uno a uno, son los candidatos a incluir). Para ella (por ejemplo) el conjunto A_0 es $A_0 = \{(1, n_2, n_1); (0, n_2-1, n_1) \mid 1 \leq n_1 \leq n-1\}$. El proceso finaliza seleccionando aquella RC_i en la que $\alpha_i^* \leq \alpha$ pero $\alpha_{i+1}^* > \alpha$.

La Tabla 2.2. presenta un cuadro comparativo de las propiedades/condiciones encontradas en cada Modelo.

Tabla 2.2

Resumen de las condiciones a verificar por la región crítica en cada uno de los modelos de Tablas 2×2

Modelo	Número de:				
	Variables	Equivalencias	Simetrías	Convexidades	Atractores
III	1	0	0	1	1
II	2	0	1	2	1
I	3	4/2/1	1	4	n-1

2.2.5. Equivalencia, Simetría, Convexidad y métodos de ordenación

2.2.5.1. Introducción

Ya se ha indicado que el método de ordenación X es subsidiario de las condiciones E, S y C, aplicándose sólo en aquellos puntos que las verifiquen. Sin embargo la ordenación puede hacerse más rápida si el propio método X las cumple. Aparte de ello, el hecho de que un método verifique E, S y C ya es un dato a favor del mismo, pues indica que actúa coherentemente, en tanto que su no verificación implica que tales condiciones deben ser impuestas de antemano. El objetivo de este apartado es comprobar qué métodos cumplen E, S y C y discutirlo. Queda claro pues que si un método X satisface E, S y C, entonces el método ESCX y el método X son el mismo, y elegir uno u otro es una cuestión de comodidad. Ningún autor se ha preocupado de tal comprobación en cuanto al Modelo I actual. En este sentido, el método de Berger (1994, 1996) no verifica la convexidad del Modelo II (ni, por tanto, la del Modelo I), y por ello no es contemplado en esta memoria.

2.2.5.2. En los métodos con memoria

Los métodos **M** y **M'** son tremendamente costosos en tiempo de cálculo cuando se les impone de antemano **E**, **S** y **C**, y aún lo serían más de no hacerlo. La imposición de **E**, **S** y **C** consigue hacerlos practicables en los pequeños valores de **n**.

El método que se llamó "**Barnard1**" está basado en el método **M** para el Modelo II, el cual, estando sometido a las condiciones de simetría y de convexidad a la "izquierda" y "arriba", lo transmitirá a aquel. De otro lado su definición (2.14) garantiza que las tablas **E1** y **E3** (ó **E2** y **E4**) entran juntas y, finalmente, las tablas **E1** y **E2** también lo harán pues, desde la perspectiva del Modelo II, son la misma (es una permuta de sus filas). Las únicas propiedades que no están garantizadas son las convexidades "adentro" y "afuera", pues ellas son extrañas a su propio fin. De todos modos "**Barnard1**" no fue propuesto como método de ordenación (pues como tal resultaría aún más inaplicable que el propio **M**) sino como un mecanismo conservador de resolver el Modelo I. La posible no verificación de **CD** y **CF** lo descalifica.

2.2.5.3. En los métodos procedentes del test exacto de Fisher

Hajek and Havranek (1978) probaron que el método **F1** verifica la condición **S** y las **CI** y **CR** (pues dichos autores lo estudiaron en el marco del Modelo II). La prueba se basaba en que las probabilidades hipergeométricas de los puntos implicados verifican que:

$$\begin{cases} P(x_1, x_2) = P(y_1, y_2) & (\text{E1 vs S1}) \\ P(x_1, x_2) > P(x_1 - 1, x_2), P(x_1, x_2 + 1) & (\text{E1 vs CI, CR}) \end{cases}$$

y que como ello sucede punto a punto, también sucederá con las sumas de los mismos. Esto se mantiene cuando algún sumando intervenga con un factor multiplicativo constante (como en el P-mid), y así **S**, **CI** y **CR** son verificadas por los métodos **F1** y **FM1**.

Las tablas **E1** a la **E4**, a efectos del test de Fisher, son 4 modos alternativos de presentar un mismo resultado, y así **F1** y **FM1** verifican **E**.

Las probabilidades de las tablas **E1** y **CD** son, respectivamente,

$$P_{E1} = \frac{n_1!n_2!a_1!a_2!}{x_1!x_2!y_1!y_2!n!} \quad ,, \quad P_{CD} = \frac{(n_1-1)!(n_2+1)!a_1!a_2!}{(x_1-1)!(x_2+1)!y_1!y_2!n!}$$

con lo que

$$\begin{aligned} \{P_{E1} - P_{CD}\} &= \text{Cte} \times \left\{ \frac{n_1!n_2!}{x_1!x_2!} - \frac{(n_1-1)!(n_2+1)!}{(x_1-1)!(x_2+1)!} \right\} \\ &= \text{Cte} \times \{n_1(x_2+1) - (n_2+1)x_1\} = \text{Cte} \times \{(n_1x_2 - n_2x_1) + (n_1 - x_1)\} = \\ &= \text{Cte} \times \{-(x_1y_2 - x_2y_1) + y_1\} > 0 \end{aligned}$$

pues nos estamos centrando en las tablas con asociaciones negativas. Con ello $P_{E1} > P_{CD}$, lo cual ocurre con todas las tablas de la línea de $a_1 = \text{cte}$ (que es donde actúa Fisher) a la que pertenecen E1 y CD. Además la línea $a_1 = \text{cte}$ consta de más tablas en el caso de E1 que en el de CD, pues en ésta su total de primera fila (n_1-1) es más pequeño que en aquella (n_1). Esto implica que el valor de **F1** en la tabla E1 es seguro mayor que en la tabla CD, con lo que ésta entrará en la RC antes que aquella. Con ello el método **F1** verifica CD y, por la argumentación de antes, también el **FM1**.

Para la tabla CF la prueba es similar y lleva a que:

$$P_{E1} - P_{CF} = \text{Cte} \times \{-(x_1y_2 - x_2y_1) + x_2\} > 0$$

con similares conclusiones que antes.

Así pues **F1** y **FM1** verifican E, S y C, de modo que sirven como método de ordenación para 1 o 2 colas.

No cabe decir otro tanto de **F2** y **FM2**, de los que se sabe que no verifican la convexidad del Modelo II (y por tanto tampoco la actual), de modo que para que actúen coherentemente es preciso imponérsela de antemano (E y S sí las verifican, puesto que tales métodos se definen como suma de valores de **F1**).

2.2.5.4. En los métodos procedentes del estadístico chi-cuadrado

La expresión (2.22), formato genérico de los métodos **Z**, **Y**, **PH1**, **PH2** y **PH3**, alcanza el mismo valor en todas las tablas E1 a la E4 y S1 a la S4, de modo que todos

ellos verifican E y S.

Las condiciones C son verificadas por **Z**, **PH1**, **PH2** y **PH3**, pero no por **Y**. Por ejemplo, las tablas $(x_1, x_2, n_1) = (1, 10, 10)$ y $(1, 10, 11)$ para $n=100$, que hacen el oficio de E1 y CF, dan valores χ^2_Y de $0,1816 > 0,0877$ respectivamente, con lo que la tabla CF entra en la RC después de la tabla E1 y así no se verifica la convexidad hacia "afuera". Sin embargo tales casos suceden en los puntos con valor P muy alto (puntos que nunca entrarán en la RC) en los que la c.p.c. suele aplicarse incorrectamente. Obsérvese que en las tablas anteriores es $x_1 = 1 \in E \pm 0,5$ (pues $E = 1,1$ y $1,21$ respectivamente), valores en los que Shuster ya indicó que debe considerarse $\chi^2_Y = 0$. Con estas condiciones **Z**, **Y**, **PH1**, **PH2** y **PH3** verifican E, S y C. La prueba general de lo anterior es la siguiente.

Suponiendo que $n_1 = \text{Min} \{a_1, a_2, n_1, n_2\}$ y asociación negativa ($\hat{\theta} < 1$), la expresión (2.22) para la tabla E1 original es:

$$E1 = \frac{(|x_1 y_2 - x_2 y_1| - c)^2}{a_1 a_2 n_1 n_2} n = \frac{(x_2 y_1 - x_1 y_2 - c)^2}{a_1 a_2 n_1 n_2} n = R^2 \frac{n}{a_1 a_2 n_1 n_2} \quad (2.24)$$

donde

$$R = x_2 y_1 - x_1 y_2 - c \quad (2.25)$$

La Tabla CI viene dada por la expresión:

$$CI = \frac{(|(x_1 - 1)y_2 - x_2(y_1 + 1)| - c)^2}{(a_1 - 1)(a_1 + 1)n_1 n_2} n$$

Suponiendo $a_1 > 1$ y efectuando los cambios aplicados antes a la (2.22), se llega a que:

$$CI = \frac{R^2 + 2Rx_2 + 2Ry_2 + x_2^2 + 2x_2 y_2 + y_2^2}{a_1 a_2 - a_2 - 1 + a_1} \frac{n}{n_1 n_2}$$

con lo que:

$$CI - E1 = \left(\frac{R^2 + 2Rx_2 + 2Ry_2 + x_2^2 + 2x_2y_2 + y_2^2}{a_1a_2 - a_2 - 1 + a_1} - \frac{R^2}{a_1a_2} \right) \frac{n}{n_1n_2} \quad (2.26)$$

cuyo signo viene determinado por la expresión dentro del paréntesis. Poniendo en ella común denominador en $(a_1a_2 - a_2 - 1 + a_1) a_1a_2 > 0$, su signo dependerá del signo del numerador, es decir de:

$$\begin{aligned} & (R^2 + 2Rx_2 + 2Ry_2 + x_2^2 + 2x_2y_2 + y_2^2)a_1a_2 - R^2(a_1a_2 - a_2 - 1 + a_1) \\ & = (2Rx_2 + 2Ry_2 + x_2^2 + 2x_2y_2 + y_2^2)a_1a_2 - R^2(-a_2 - 1 + a_1) \\ & \geq 2Rx_2a_1a_2 - R^2(a_1 - a_2 - 1) \geq 2Rx_2a_1a_2 - R^2a_1 = Ra_1(2x_2a_2 - R) \\ & \geq Ra_1(2x_2a_2 - x_2y_1) \geq Ra_1(2x_2y_1 - x_2y_1) = a_1x_2y_1R \end{aligned} \quad (2.27)$$

lo que es mayor o igual que cero cuando $R \geq 0$. Con ello:

$$CI \geq E1 \text{ si } R \geq 0 \quad (2.28)$$

De igual modo, la Tabla CA viene dada por la expresión:

$$CA = \frac{(|x_1(y_2 - 1) - y_1(x_2 + 1)| - c)^2}{(a_1 + 1)(a_2 - 1)n_1n_2} n$$

Suponiendo $a_2 > 1$ y efectuando los cambios aplicados a la (2.22) se obtiene que:

$$CA = \frac{R^2 + 2Rx_1 + 2Ry_1 + y_1^2 + 2x_1y_1 + x_1^2}{a_1a_2 + a_2 - a_1 - 1} \frac{n}{n_1n_2}$$

Calculando CA-E1 como en la (2.26) y realizando una cadena de igualdades y desigualdades similares a la (2.27) -la expresión clave de la 3ª línea de la (2.27) es ahora $2Ry_1a_1a_2 - R^2(a_2 - a_1 - 1)$ - se tiene que:

$$CA \geq E1 \text{ si } R \geq 0 \quad (2.29)$$

De igual modo, la Tabla CF viene dado por la expresión:

$$CF = \frac{(|(x_1(y_2-1) - x_2(y_1+1))| - c)^2}{(n_1+1)(n_2-1)a_1a_2} n$$

Suponiendo que $n_2 > 1$ y efectuando los cambios aplicados a (2.22) se llega a que:

$$CF = \frac{R^2 + 2Rx_1 + 2Rx_2 + x_1^2 + 2x_1x_2 + x_2^2}{n_1n_2 + n_2 - n_1 - 1} \frac{n}{n_1n_2}$$

Calculando CF-E1 como en la (2.26) y realizando una cadena de igualdades y desigualdades paralela a la (2.27) -la expresión clave de la 3ª línea de la (2.27) es ahora $2Rx_2n_1n_2 - R^2(n_2 - n_1 - 1)$ - se tiene que:

$$CF \geq E1 \text{ si } R \geq 0 \quad (2.30)$$

Finalmente, el caso de la Tabla CD no depende de R, ya que al venir dada por la expresión:

$$CD = \frac{(|(x_1-1)y_2 - (x_2+1)y_1| - c)^2}{(n_1-1)(n_2+1)a_1a_2} n$$

suponiendo que $n_1 > 1$ y efectuando los cambios aplicados a la (2.22) se obtiene que:

$$CD = \frac{(R+y_1+y_2)^2}{n_1n_2 + n_1 - n_2 - 1} \frac{n}{a_1a_2}$$

que tiene mayor numerador y menor denominador que (2.24): por tanto $CD > E1$ siempre.

Se ve pues que la verificación de las propiedades de convexidad requiere que sea $R \geq 0$ o, por la (2.25), que sea $U = x_2y_1 - x_1y_2 \geq c$, con c las diversas c.p.c. de los métodos **Z**, **PH3**, **PH1**, **PH2** e **Y**: respectivamente 0, $[n^2/8]/\bar{n}$ con \bar{n} dado por la (2.23), 0,5, $n/4$ y $n/2$. La c.p.c. de **PH3** converge hacia cero cuando $n \rightarrow \infty$ y en $n=5$ ya alcanza un valor menor que 0,5 (exactamente 0,32). Como U salta como mínimo de 1 en 1 y los valores de asociación nula ($U=0$) se están excluyendo, entonces $U > c$ para **Z**, **PH3** y **PH1** siempre. No sucederá así seguro en el caso de los **PH2** e **Y** -

como muestra el ejemplo del inicio de la sección en que $U=10 < n/4=25$ y $n/2=50$ y de ahí que, para protegerse contra ello, Shuster englobara a todas las tablas con $|U| \leq n/2$ en un valor $\chi^2_Y = 0$. Otro tanto puede hacerse con **PH2**: Si $|U| < n/4$ hacer $\chi^2_{PH2} = 0$ (salvo que se controlen a priori las condiciones de convexidad). De cualquier modo todas esas tablas estarán cercanas a la no significación y, a efectos prácticos, no causarán ningún problema.

Finalmente señalar que desde el principio se está asumiendo que $a_1, a_2, n_1, n_2 > 1$, lo que se hace para evitar a las tablas con un valor indeterminado de chi-cuadrado (las cuales tampoco son de interés).

2.3. SIMPLIFICACIONES PARA EL CÁLCULO DEL MÁXIMO*

2.3.1. Introducción

Ya conocemos que la evaluación de la (2.2) es compleja, y así cualquier propiedad que permita disminuir su tiempo de cómputo influirá favorablemente en la difusión y aplicabilidad del método incondicionado para el Modelo I. En el caso del Modelo II, Silva and Martín (1997) ya abordaron este asunto. Aquí se va a efectuar un estudio paralelo que se aprovecha fundamentalmente de la existencia de las propiedades E y S.

La reducción del tiempo de cómputo puede provenir de 3 causas: **i)** Por una agrupación conveniente de cálculos; **ii)** Por una disminución del espacio paramétrico $(0,1) \times (0,1)$ de los parámetros perturbadores (p,q) sobre los que hay que maximizar; y **iii)** Por una reducción del número de veces en que hay que evaluar la (2.2) (en el caso del método **M**).

En lo que sigue será de interés la tabla nemotécnica notada por Tabla 2.3, en la cual se expresan las bases y los exponentes de los términos que dependen de p o de q de las probabilidades de las tablas E1 a E4 y S1 a S4.

Tabla 2.3

Para cada tabla que se indica en el exterior, en el interior se da la base y en la primera línea el exponente al que hay que elevarla, todo ello relativo a la probabilidad de la tabla citada.

Tabla	Exponentes								Tabla
	a ₁	a ₂	n ₁	n ₂	a ₁	a ₂	n ₁	n ₂	
E1	p	1-p	q	1-q	1-p	p	q	1-q	S1
E2	1-p	p	1-q	q	p	1-p	1-q	q	S2
E3	q	1-q	p	1-p	q	1-q	1-p	p	S3
E4	1-q	q	1-p	p	1-q	q	p	1-p	S4
E→	<i>EQUIVALENCIAS</i>				<i>SIMETRÍAS</i>				←S

2.3.2. Agrupación de cálculos

La probabilidad de todas las Tablas E_i consta de igual valor del número combinatorio, y así, agrupándolas, se evita su cálculo repetitivo. Si notamos por:

$$h(p, q, a_1, n_1) = p^{a_1}(1-p)^{a_2}q^{n_1}(1-q)^{n_2} + (1-p)^{a_1}p^{a_2}(1-q)^{n_1}q^{n_2} + q^{a_1}(1-q)^{a_2}p^{n_1}(1-p)^{n_2} + (1-p)^{a_1}q^{a_2}(1-p)^{n_1}p^{n_2} \tag{2.31}$$

y por:

$$g(p, q, a_1, n_1) = \begin{cases} h(p, q, a_1, n_1) & \text{si } a_1 \neq n_1 \text{ y } a_1 \neq n_2 \\ h(p, q, a_1, n_1)/2 & \text{si } a_1 = n_1 \text{ ó } a_1 = n_2 \\ h(p, q, a_1, n_1)/4 & \text{si } a_1 = n_1 = n_2 \end{cases} \tag{2.32}$$

entonces la probabilidad conjunta de todas las Tablas E_i es $\{n!/(x_1!y_1!x_2!y_2!)\} \times g(p, q, a_1, n_1)$.

Los divisores 2 y 4 de la (2.32) son por causa de que en los casos señalados hay sólo 2 o 1 tablas equivalentes (en lugar de las 4 tradicionales) y así tales divisores compensan el hecho de que la (2.31) tiene 2 o 4 términos repetidos inexistentes.

Por otro lado, todos los puntos de la RC que están en igual diagonal a₁ del rectángulo de n₁ tienen igual valor de la función g, y así puede sacarse factor común a ella y calcular la suma de los números combinatorios que le preceden:

$$C(a_1, n_1) = \sum_{\substack{(x_1, x_2, n_1) \in RC \\ x_1 + x_2 = a_1}} \frac{n!}{x_1! y_1! x_2! y_2!} \quad (2.33)$$

con lo que, finalmente, el error α para el test de una cola ($H_1: \theta < 1$) será:

$$\alpha_1(p, q) = \sum_{n_1=1}^{[n/2]^-} \sum_{a_1=n_1}^{n-n_1} C(a_1, n_1) g(p, q, a_1, n_1) \quad (2.34)$$

y para el de dos colas,

$$\alpha(p, q) = \sum_{n_1=1}^{[n/2]^-} \sum_{a_1=n_1}^{n-n_1} C(a_1, n_1) \{g(p, q, a_1, n_1) + g(1-p, q, a_1, n_1)\} \quad (2.35)$$

pues la Tabla Si se obtiene permutando las columnas de la Tabla Ei. Ver también la Tabla 2.3.

Una pequeña ganancia extra en tiempo de cómputo se obtiene haciendo $x = p/(1-p)$ e $y = q/(1-q)$, pues entonces $\{(1-p) \times (1-q)\}^n$ sale como factor común en la (2.31) y así:

$$h(p, q, a_1, n_1) = \{(1-p)(1-q)\}^n \times \left\{ \left(\frac{p}{1-p}\right)^{a_1} \left(\frac{q}{1-q}\right)^{n_1} + \left(\frac{p}{1-p}\right)^{a_2} \left(\frac{q}{1-q}\right)^{n_2} + \left(\frac{p}{1-p}\right)^{n_1} \left(\frac{q}{1-q}\right)^{a_1} + \left(\frac{p}{1-p}\right)^{n_2} \left(\frac{q}{1-q}\right)^{a_2} \right\}$$

Obsérvese que las sumatorias ya se han puesto en el formato canónico de expresar el espacio muestral y que a los errores α del test para una y dos colas se les ha llamado por $\alpha_1(p, q)$ y $\alpha(p, q)$.

2.3.3. Reducción general del espacio paramétrico

Sean R_1 y R las RC para los tests de una cola ($H_1: \theta < 1$) y de dos colas ($H_1: \theta \neq 1$) respectivamente. En cada una de esas RC los errores α toman los valores $\alpha_1(p, q)$ y $\alpha(p, q)$, y la expresión (2.2) para ellos da lugar a los tamaños α_1^* y α^* . Para los errores α usuales, la región $R = R_2 \cup \overline{R_2}$ estará formada por los puntos R_2 (con

$\hat{\theta} < 1$) y $\overline{R^2}$ (sus simétricos, con $\hat{\theta} > 1$), regiones que dan errores $\alpha^2(p, q)$ y $\overline{\alpha^2}(p, q)$ y tamaños α^{2*} y $\overline{\alpha^{2*}}$, con

$$\alpha(p, q) = \alpha^2(p, q) + \overline{\alpha^2}(p, q) \quad (2.36)$$

La probabilidad de la Tabla E1 (ó E3) es la misma que la de la Tabla E2 (ó E4) si se cambia p por $1-p$ y q por $1-q$, y como las cuatro tablas están presentes en $\alpha(p, q)$ entonces ella será igual que $\alpha(1-p, 1-q)$. Igual sucede con las Tablas E1 (ó E2) y la E3 (ó E4) si se cambia p por $q \Rightarrow \alpha(p, q) = \alpha(q, p)$. De ambas, o de las tablas E_i , se deduce que también $\alpha(p, q) = \alpha(1-q, 1-p)$. Con todo ello las siguientes igualdades son ciertas:

$$\alpha(p, q) = \alpha(q, p) = \alpha(1-p, 1-q) = \alpha(1-q, 1-p) \quad (2.37)$$

lo cual es válido para los errores α , α^1 , α^2 , y $\overline{\alpha^2}$. La primera igualdad (aludida por Shuster, 1992, para el método Y) indica que $\alpha(p, q)$ es simétrica respecto de la diagonal secundaria $p=q$ del espacio paramétrico; la segunda indica lo mismo respecto de la diagonal principal $p+q=1$; la tercera, consecuencia de las otras dos, implica la simetría respecto del punto $(0,5; 0,5)$. De ello se deduce que las direcciones $p=q$ y $p=1-q$, y el punto $(p=0,5; q=0,5)$, son lugares privilegiados donde obtener un extremo de la (2.2), aunque no tienen por qué ser los únicos ni dar lugar al máximo: de ahí los comentarios de Shuster (1992) y Kroll (1989) acerca de que la primera y última sean soluciones frecuentes al obtener la (2.2).

De modo similar, la probabilidad de la Tabla E_i es igual a la de la Tabla S_i si se cambia p por $1-p$, con lo cual:

$$\overline{\alpha^2}(p, q) = \alpha^2(1-p, q) \Rightarrow \overline{\alpha^{2*}} = \alpha^{2*} \quad (2.38)$$

y así, por las (2.36) y (2.38):

$$\alpha(p, q) = \alpha^2(p, q) + \alpha^2(1-p, q) = \alpha(1-p, q) \quad (2.39)$$

lo que indica que $\alpha(p, q)$ es simétrica en $p=0,5$ y ella es otra dirección privilegiada para localizar un extremo de ella. Shuster (1992) utilizó las primeras igualdades de las (2.38) y (2.39) para su método Y.

Como en los métodos "sin memoria" el orden de entrada de un punto no depende de los que ya estén presentes en la RC, entonces el orden a 1 y 2 colas es el mismo, y así $R_2 = R_1$, $\alpha_2(p, q) = \alpha_1(p, q)$, y, por la (2.39):

$$\text{En los métodos sin memoria: } \alpha(p, q) = \alpha_1(p, q) + \alpha_1(1-p, q) \Rightarrow \alpha^* \leq 2\alpha_1^* \quad (2.40)$$

es decir, el valor P del test de dos colas es menor o igual que el doble del valor P para el test de una cola. Esto no tiene porqué suceder en los métodos "con memoria" (como ya mostraron Silva and Martín, 1997, para el caso del Modelo II con un contraejemplo).

Las simetrías anteriores citadas para $\alpha(p, q)$, y que se dan en la Figura 2.3, indican que la determinación de la (2.2) no es preciso efectuarla en $0 < p, q < 1$, sino que basta determinarla en el triángulo inferior definido por las diagonales del rectángulo $(0, 1) \times (0, 1)$, es decir en $\{0 < p \leq 0,5, 0 < q \leq p\} \cup \{0,5 < p < 1; 0 < q \leq 1-p\}$. Para el caso de test de dos colas, la simetría de la (2.39) indica que el espacio paramétrico puede reducirse a $\{0 < q \leq p \leq 0,5\}$. Este último resultado fue señalado por Kroll (1989) sin demostración, en tanto que Shuster (1992) aludió a la restricción $0 < q \leq p < 1$ (ambos para test de dos colas). Todo lo dicho lleva a la siguiente propiedad:

Propiedad I: El máximo para el test de una cola ha de buscarse en la región:

$$\left\{ \begin{array}{l} 0 < p \leq 0,5 \\ 0 < q \leq p \end{array} \right\} \cup \left\{ \begin{array}{l} 0,5 < p < 1 \\ 0,5 < q \leq 1-p \end{array} \right\} \quad (2.41)$$

en tanto que para el test de dos colas estará en:

$$0 < q \leq p \leq 0,5 \quad (2.42)$$

lo que se representa en la Figura 2.4.

Por otro lado, y ya sin el formato reducido pues por ahora no interesa, las derivadas de $\alpha(p, q)$ son:

Figura 2.3

El error $\alpha(p,q)$ del test de dos colas es simétrico respecto de las zonas en negrita de las figuras de abajo. El error $\alpha_1(p,q)$ del test de una cola sólo lo es en los tres primeros casos. El tercer caso es redundante (se obtiene de los dos primeros).

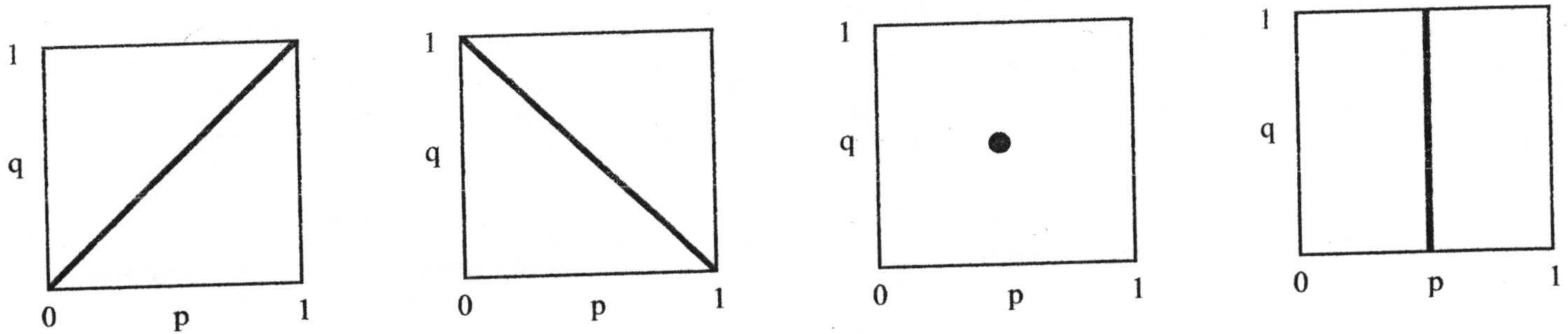
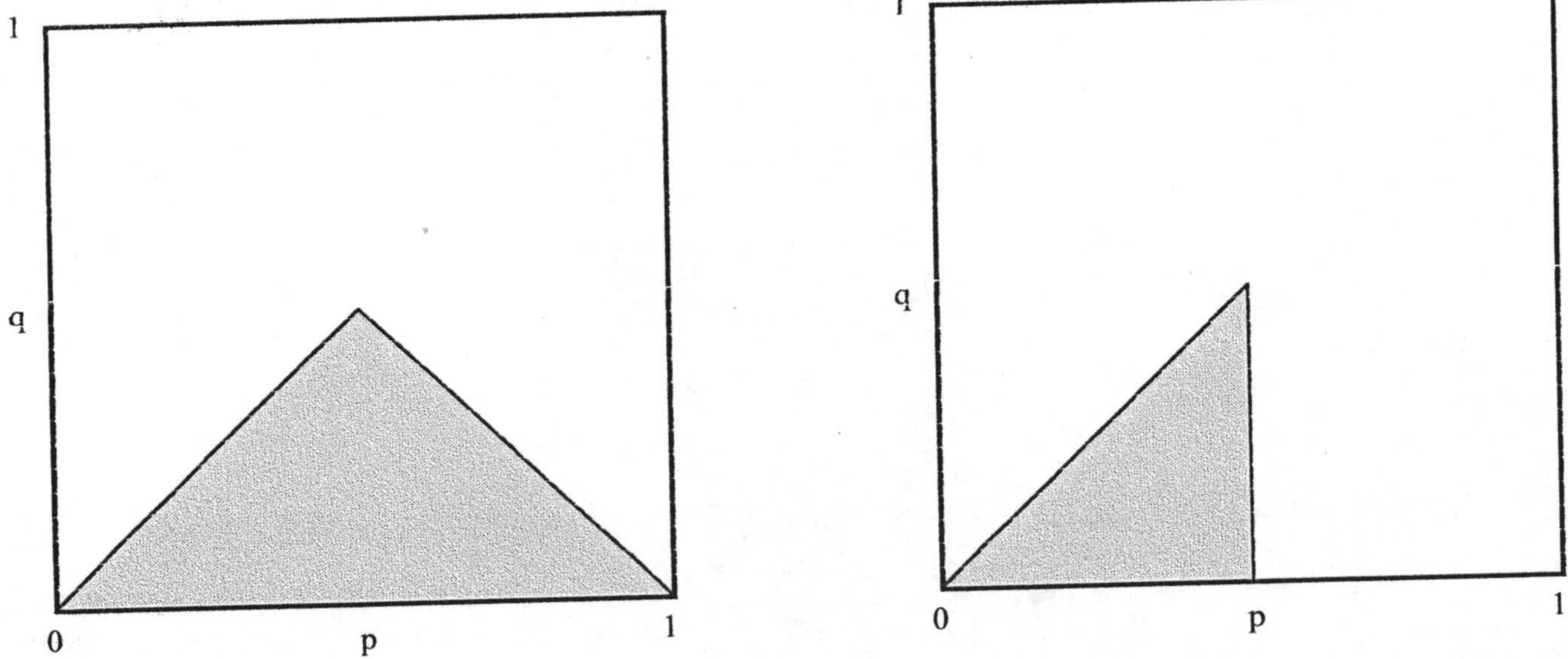


Figura 2.4

Zonas en las que localizar el máximo en el test de una cola (izquierda) o en el de dos colas (derecha).



$$\alpha'_p(p, q) = \sum_{RC} \frac{n!}{x_1! y_1! x_2! y_2!} (a_1 - np) p^{a_1 - 1} (1 - p)^{a_2 - 1} q^{n_1} (1 - q)^{n_2} \quad (2.43)$$

$$\alpha'_q(p, q) = \sum_{RC} \frac{n!}{x_1! y_1! x_2! y_2!} (n_1 - nq) p^{a_1} (1 - p)^{a_2} q^{n_1 - 1} (1 - q)^{n_2 - 1}$$

que serán ambas positivas para p y q menores que todas las fracciones a_1/n y n_1/n , y ambas negativas para p y q mayores que todas esas fracciones. Si en la RC están todos atractores, la mínima fracción es $1/n$ y la máxima $1-1/n$. Con ello $\alpha(p, q)$ siempre crece a la izquierda de $1/n$ y siempre decrece a la derecha de $1-1/n$ (en p , para valores constantes de q ; o en q , para valores constantes de p). Como el máximo de la (2.2) se obtiene allá donde $\alpha'_p = \alpha'_q = 0$, entonces p y q han de estar en el intervalo $(1/n; 1-1/n)$, lo cual, unido a la Propiedad, I indica que:

Propiedad II: El máximo para el test de una cola ha de buscarse en la región:

$$\left\{ \begin{array}{l} 1/n \leq p \leq 0,5 \\ 1/n \leq q \leq p \end{array} \right\} \cup \left\{ \begin{array}{l} 0,5 < p \leq 1-1/n \\ 1/n \leq q \leq 1-p \end{array} \right\} \quad (2.44)$$

en tanto que para el test de dos colas estará en:

$$1/n \leq q \leq p \leq 0,5 \quad (2.45)$$

si se admite que los atractores entran todos desde el principio; en otro caso, el término $1/n$ debe cambiarse por $\text{Min } n_1/n$.

Al respecto, Shuster (1992) indicó, sin demostración, que el máximo para el método Y ha de estar dentro del cuadrado $1/n \leq p, q \leq 1-1/n$. La Figura 2.5 ilustra gráficamente esta propiedad.

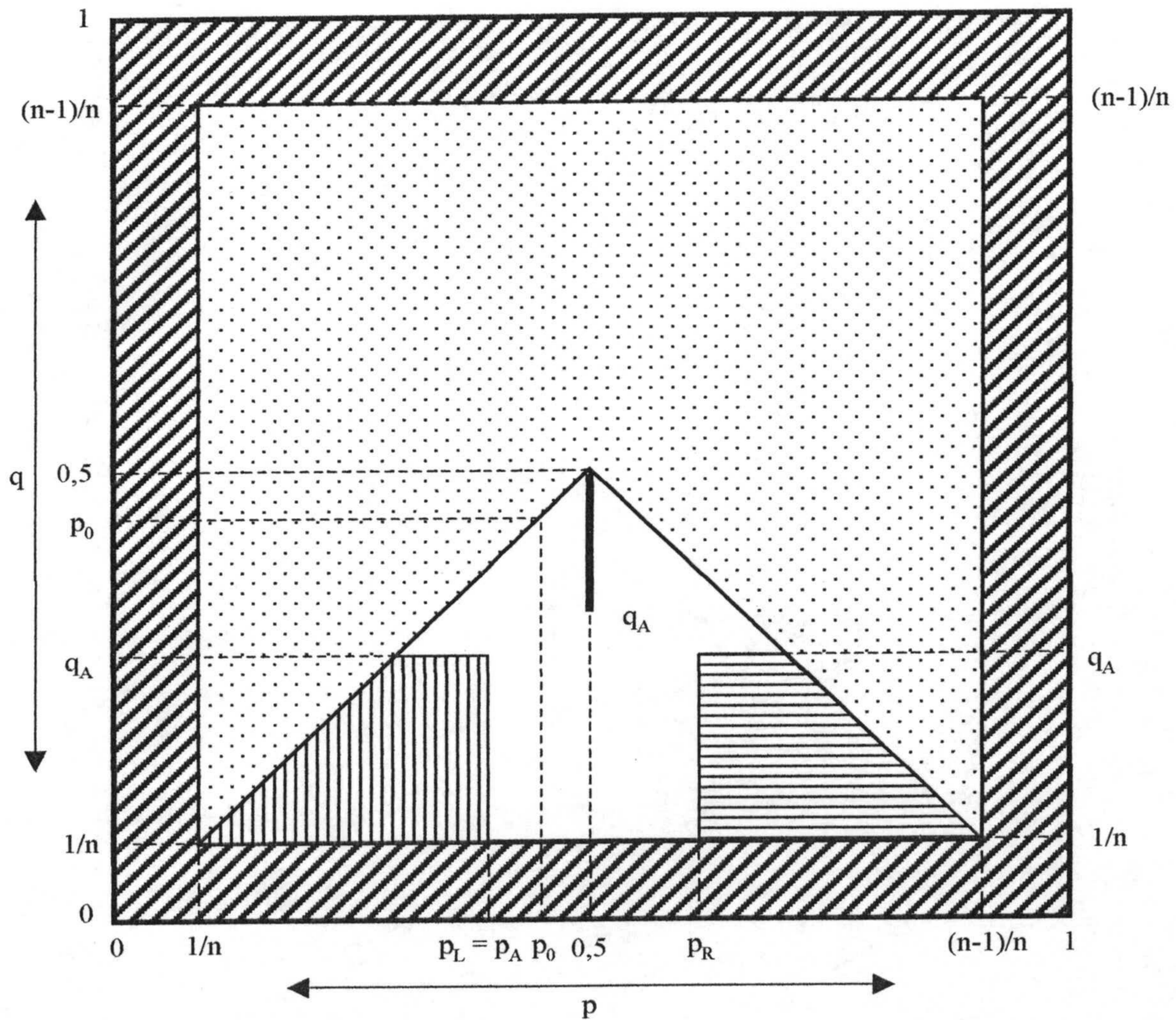
2.3.4. Reducción del espacio paramétrico en el caso de α^* sucesivos

Cuando hay que determinar la RC (por cualquier ordenación), el investigador se ve obligado a calcular α^* en cada una de las RC sucesivas que va obteniendo (con el método M esto ocurre también incluso si sólo se desea el valor P de la tabla

Figura 2.5

Reducción del espacio paramétrico por la aplicación de las propiedades

- 1) La zona rayada oblicuamente se ahorra en base al crecimiento y decrecimiento del error α y la zona punteada se ahorra en base a las simetrías del error α (Propiedad II).
- 2) La zona rayada verticalmente se ahorra en base al conocimiento del lugar del máximo para la RC anterior, siendo ella la aplicación de la Propiedad IV.iii) en lo referente a los dos conjuntos de la izquierda de la (2.53). La zona rayada horizontalmente se ahorra en base a lo anterior, en lo referente a los dos conjuntos de la derecha de la (2.53), y con p_R determinado como en la Propiedad III.iii.a).
- 3) Para el test de dos colas, el espacio paramétrico queda reducido al triángulo rectángulo de la izquierda limitado por $p = 0,5$, siendo de aplicación lo señalado en el primer párrafo de 2) y aludiéndose a la Propiedad V.
- 4) La zona retintada vertical es donde hay que localizar el máximo cuando es $p_0 = p_A = 0,5$.



experimental). Las propiedades que siguen pretenden acortar el espacio paramétrico definido por la Propiedad II aprovechando, en lo posible, la información que proporciona el lugar donde se obtuvo el máximo en el paso anterior. Con tal fin, sean RC_A y RC_N las RC del paso anterior y del nuevo paso, $\alpha_A(p,q)$ y $\alpha_N(p,q)$ sus errores α , (p_A, q_A) y (p_N, q_N) el lugar donde alcanzan el máximo, $RC_o = RC_N - RC_A$ los nuevos puntos incorporados y $\alpha_o(p,q)$ su contribución al error α . Con ello:

$$\alpha_o(p,q) = \sum_{RC_o} \frac{n!}{x_1! y_1! x_2! y_2!} p^{a_1} (1-p)^{a_2} q^{n_1} (1-q)^{n_2} \quad (2.46)$$

$$\alpha_N(p,q) = \alpha_A(p,q) + \alpha_o(p,q) \quad (2.47)$$

Para el test de una cola, el error α alude a la función $\alpha_1(p,q)$. Supongamos que RC_o consta de un único punto y que $a_1/n = p_o$ y $n_1/n = q_o$ en él. La derivada de $\alpha_o(p,q)$ viene dada por la (2.43), omitiendo el símbolo \sum , y de ella se deduce que $\alpha_o(p,q)$: a) alcanza un máximo en (p_o, q_o) ; b) para q fijo, crece a la izquierda de p_o y decrece a la derecha; c) para p fijo, crece a la izquierda de q_o y decrece a la derecha. Con ello:

Propiedad III: La solución (p_A, q_A) divide el espacio paramétrico en 4 cuadrantes:

- i) Si $(p_o, q_o) = (p_A, q_A)$ entonces $(p_N, q_N) = (p_A, q_A)$.
- ii) Si $p_o = p_A$ entonces $(p_N, q_N) = (p_A, q_N)$ en igual semirrecta que q_o respecto de q_A .
- iii) Si $q_o = q_A$ entonces $(p_N, q_N) = (p_N, q_A)$ en igual semirrecta que p_o respecto de p_A, q_A .
- iv) En otro caso, (p_N, q_N) no puede estar en el cuadrante opuesto al que se encuentra (p_o, q_o) .

Prueba: La i) se deduce del hecho de que α_A y α_o se maximizan en igual punto, luego α_N también.

Para probar iv) es preciso ver en qué puntos (p,q) sucede que:

$$\alpha_o(p,q) < \alpha_o(p_A, q_A) \quad (2.48)$$

pues ellos estarán prohibidos. En efecto, como $\alpha_N(p,q) = \alpha_A(p,q) + \alpha_o(p,q) \leq \alpha_A(p_A, q_A) + \alpha_o(p,q)$ (por ser α_A máxima en (p_A, q_A)) $< \alpha_A(p_A, q_A) + \alpha_o(p_A, q_A)$ (por la hipótesis anterior) $= \alpha_N(p_A, q_A)$, entonces tales puntos (p,q) no pueden dar el máximo para α_N . La (2.48) se consigue si se logra probar que:

$$\alpha_o(p,q) < \alpha_o(p_A, q) < \alpha_o(p_A, q_A) \quad (2.49)$$

es decir con dos transiciones de α_o ligadas con el signo $<$, una para q constante y otra para p constante, lo cual, por el crecimiento y decrecimiento de α_o , se consigue en el cuadrante señalado. En efecto, cualquier combinación de las zonas PI y PII para p con las zonas QI y QII para q permite obtener la (2.49) -Figura 2.6- lo que da las zonas prohibidas de la Figura 2.7 que se resumen en la *iv*). Así, para la primera Figura 2.7 es $p_A > p_o$ y $q_A > q_o$, que son las zonas PII y QII de la Figura 2.6, lo que implica que están prohibidos $p > p_A$ y $q > q_A$, es decir la zona rayada de la primera Figura 2.7.

Cuando (p_o, q_o) está en la frontera de dos cuadrantes, será $p_o = p_A$ (por ejemplo), en cuyo caso están prohibidos todos los valores $p \neq p_o$ y el máximo habrá de buscarse en la semirrecta en donde esté q_o : $q \geq q_A$ o $q \leq q_A$, lo que da la *ii*). De igual modo para la *iii*): Ver Figura 2.8. \square

Normalmente la RC_o constará de más de un punto (al menos los 1, 2 o 4 equivalentes), en cuyo caso el crecimiento de α_o está garantizado sólo a la izquierda del mínimo de los p_o (ó de los q_o) y el decrecimiento sólo a la derecha del máximo de los p_o (ó de los q_o) de los puntos implicados. Para cada grupo de tablas equivalentes estos valores son $p_o = \text{Min}(a_1, a_2, n_1, n_2)/n$ y $1-p_o$ (ó sus valores mínimo y máximo, si la RC_o consta de más de un grupo), de modo que sólo fuera del intervalo $(p_o, 1-p_o)$ se puede asegurar el crecimiento (a la izquierda) o el decrecimiento (a la derecha) de α_o

Figura 2.6
Zonas de p o q donde se verifica la desigualdad indicada.

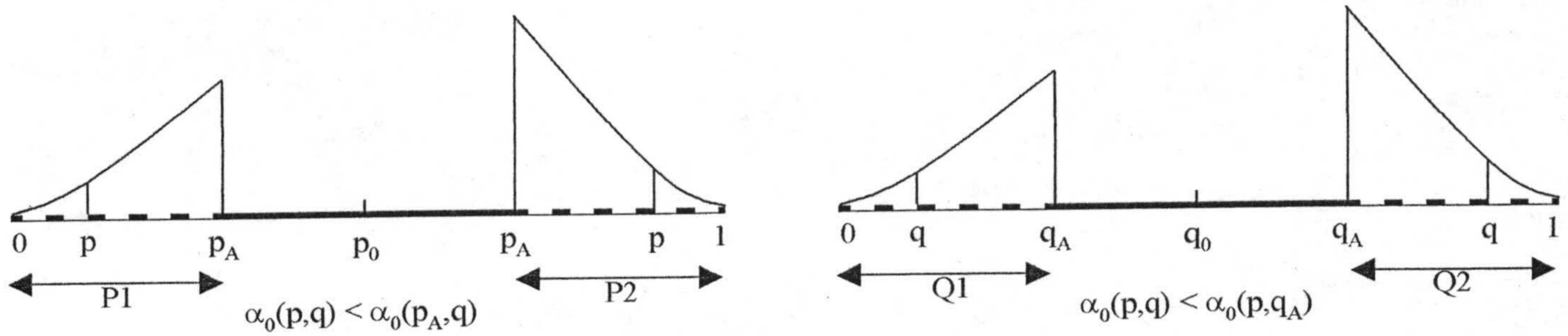
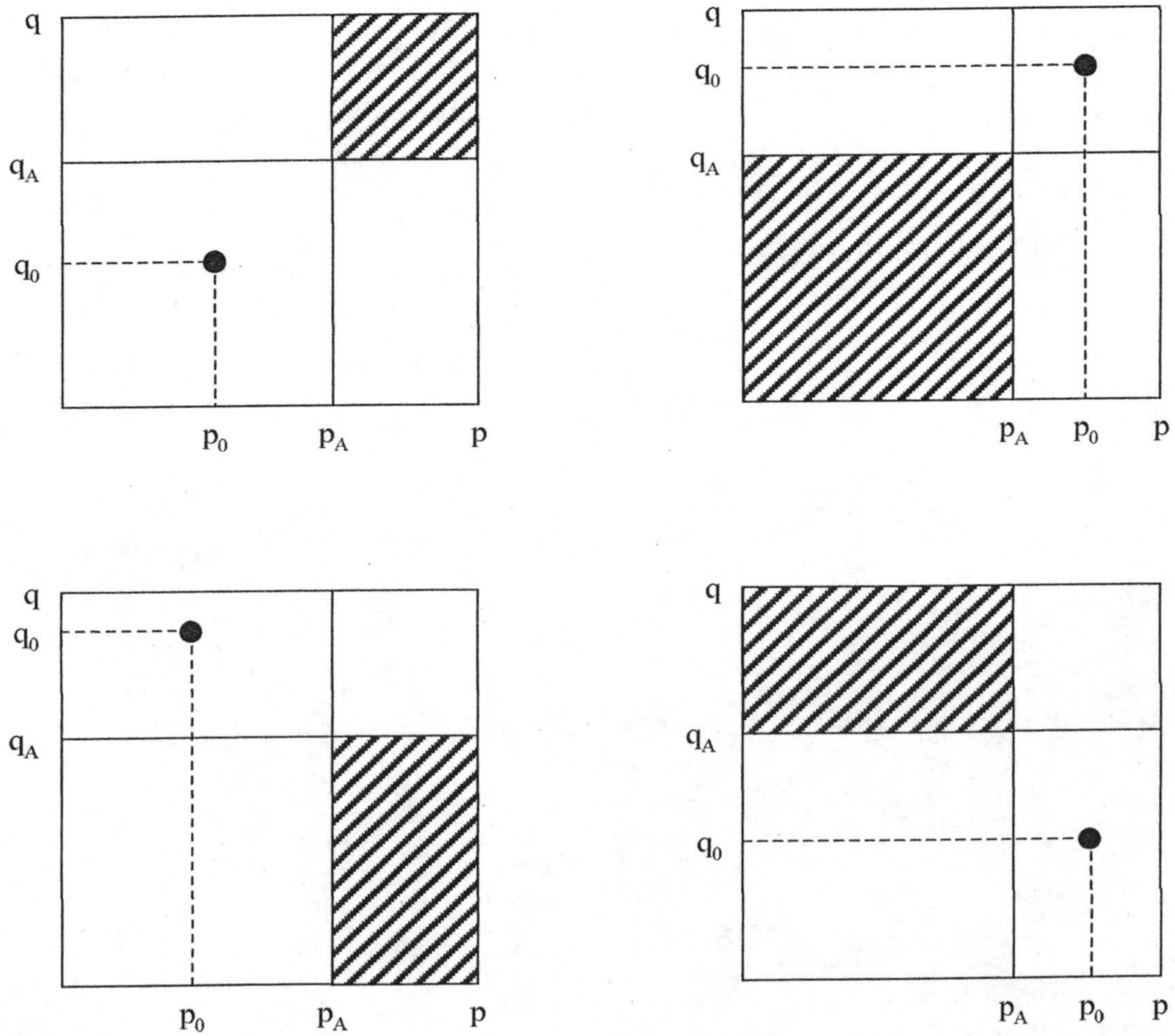


Figura 2.7
Cuadrantes prohibidos para (p_N, q_N) -zona rayada- según el cuadrante donde se encuentre (p_0, q_0) : punto marcado con ●.



en p o en q . Esto quiere decir que si (p_A, q_A) cae dentro de la cuadrícula formada por las cuatro parejas posibles de valores p_o y $1-p_o$, no es posible ahorrar espacio paramétrico, pues los (p_o, q_o) y (p_A, q_A) de la Propiedad III caen en igual cuadrante, pero sí en otro caso: Ver Figura 2.9. La Propiedad III es aplicable entonces con los cambios apropiados, y se la puede completar restringiéndose a la zona a que alude la Propiedad II.

Sin embargo, hasta ahora sólo se ha ahorrado espacio paramétrico (cuando es posible) en la zona izquierda de la (2.44) o en la zona derecha de ella, según donde se encuentre (p_A, q_A) . Veamos cómo ahorrar también por el otro lado. Si (p_A, q_A) está a la izquierda, es decir si $p_A < 0,5$, el objetivo es encontrar una pareja (p'_A, q'_A) a partir de la cual (es decir, para $p > p'_A$ y $q < q'_A$) ocurra la expresión clave (2.48). Con tal fin sea $q'_A = q_A$:

- a) Si sucede que $\alpha_o(p_A, q_A) \geq \alpha_o(1-p_o, q_A)$, entonces para $p > 1-p_o = p'_A$ y $q < q_A = q'_A$ es $\alpha_o(p, q) < \alpha_o(p, q_A)$ (por el crecimiento en q) $< \alpha_o(1-p_o, q_A)$ (por el decrecimiento en p) $\leq \alpha_o(p_A, q_A)$ (por hipótesis), con lo que se verifica la (2.48).
- b) Si no sucede lo anterior, entonces es posible determinar un valor $p = p'_A$ que sea solución de la ecuación $\alpha_o(p, q_A) = \alpha_o(p_A, q_A)$ y así para $p > p'_A$ y $q < q_A$ será $\alpha_o(p, q) < \alpha_o(p, q_A) < \alpha_o(p'_A, q_A) = \alpha_o(p_A, q_A)$, y de nuevo se verifica la (2.48). Con todo ello la pareja (p'_A, q'_A) buscada es:

$$p'_A = \begin{cases} 1-p_o & \text{si } \alpha_o(p_A, q_A) \geq \alpha_o(1-p_o, q_A) \\ \text{la solución en } p (> 1-p_o) \text{ de } \alpha_o(p_A, q_A) = \alpha_o(p, q_A) & \text{(en otro caso)} \end{cases} \quad (2.50)$$

Alternativamente puede fijarse de antemano que sea $p'_A = 1-p_A$, en cuyo caso, y con una prueba similar, sería:

Figura 2.8
 Zonas (retintadas) en donde localizar el máximo según sea $p_0 = p_A$ o $q_0 = q_A$.

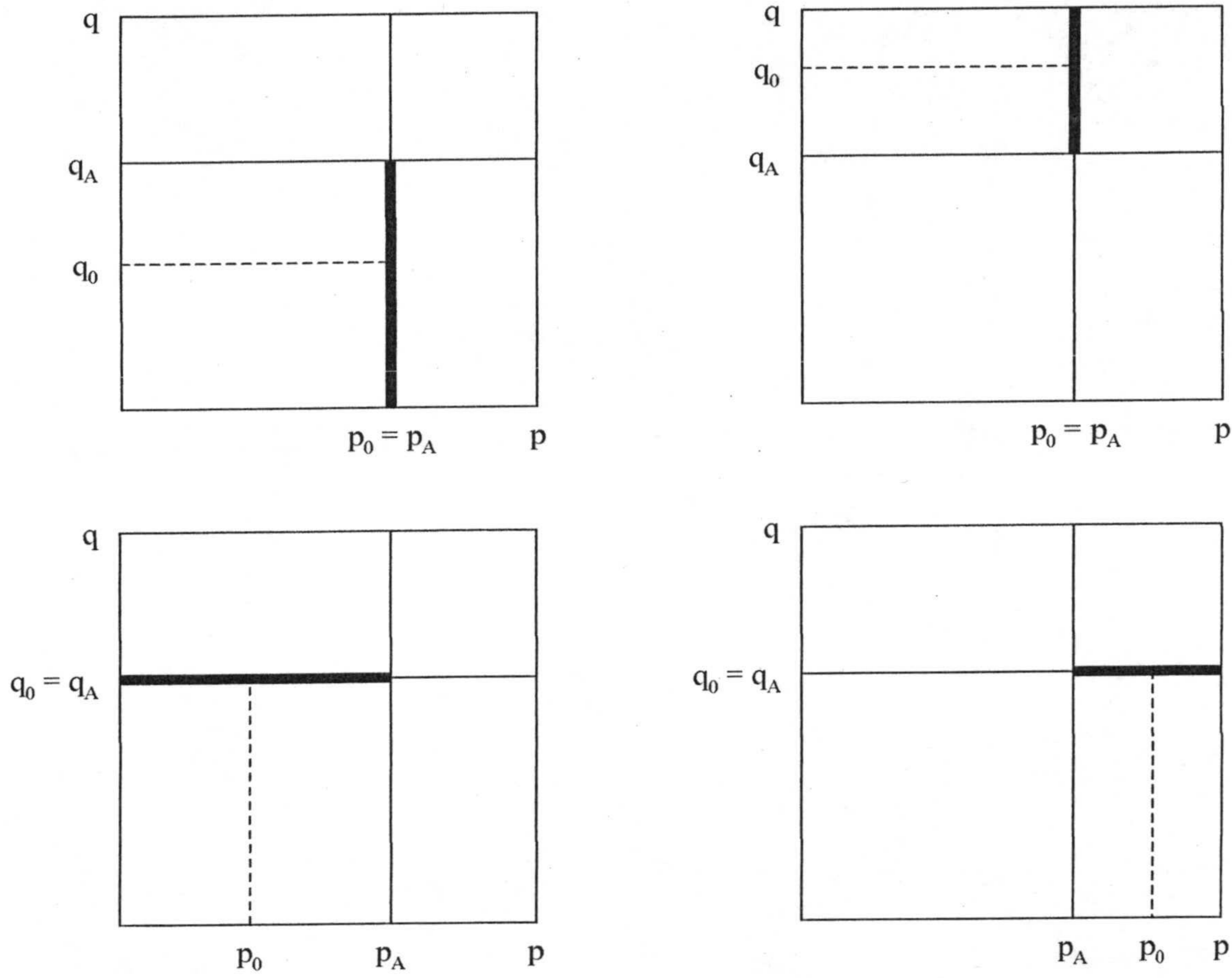
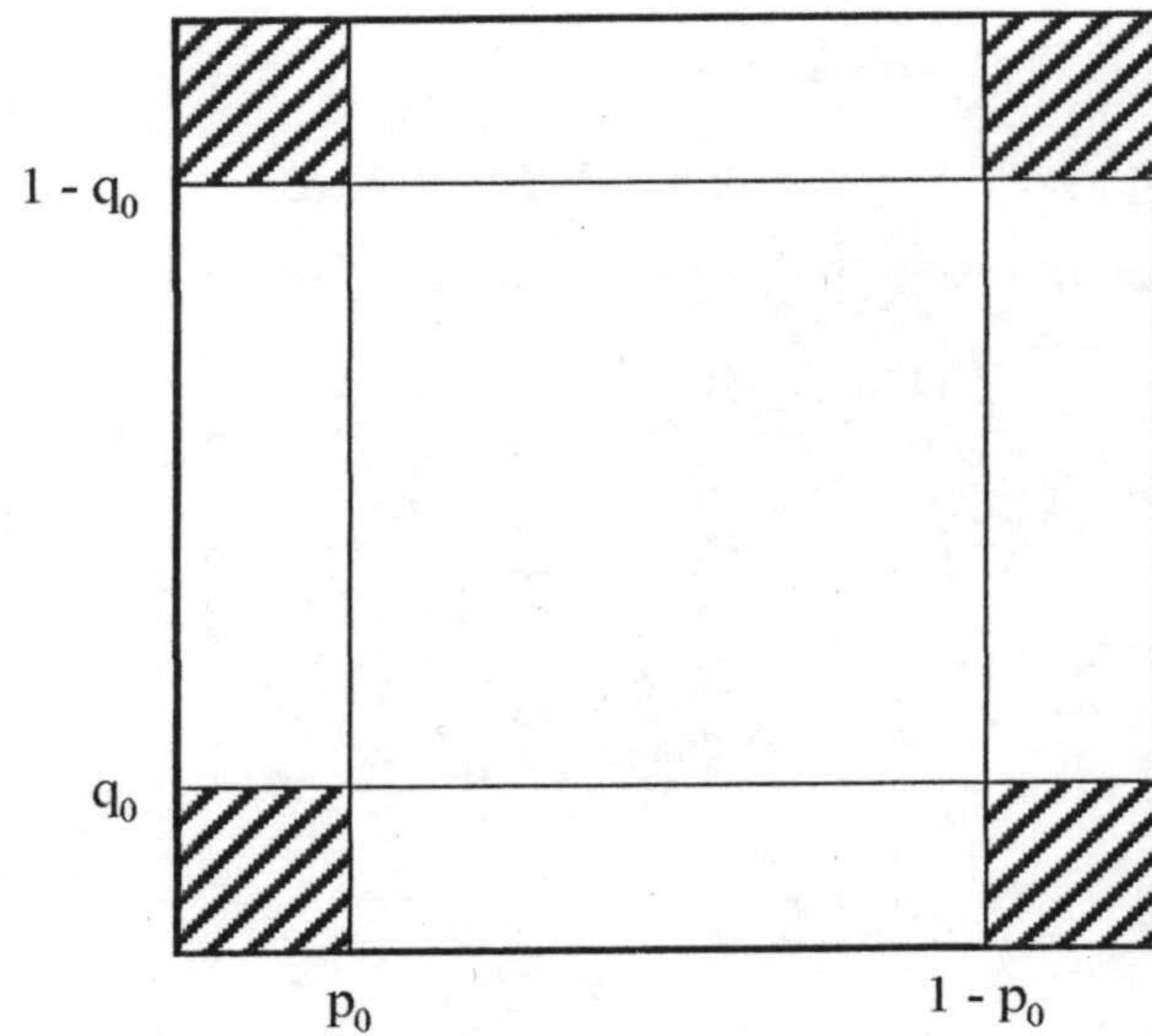


Figura 2.9

Si (p_A, q_A) cae en las zonas rayadas, los puntos equivalentes de RC_0 quedan en el mismo cuadrante y es posible ahorrar espacio paramétrico.



$$\begin{aligned} & \text{Si } p_A < 0,5: \text{ Hacer } p'_A = 1 - p_A \text{ y} \\ q'_A = & \begin{cases} p_o & \text{si } \alpha_o(p_A, q_A) \geq \alpha_o(1 - p_A, p_o) \\ \text{la solución en } q (< q_o) \text{ de } \alpha_o(p_A, q_A) = \alpha_o(1 - p_A, q) & \text{(en otro caso)} \end{cases} \end{aligned} \quad (2.51)$$

Como $\alpha_o(1 - p_o, q_A) = \alpha_o(1 - q_A, p_o)$ -por las simetrías- $< \alpha_o(1 - p_A, p_o)$ -pues $1 - q_A < 1 - p_A$ - se deduce que es más fácil que $\alpha_o(p_A, q_A)$ supere al más pequeño de ambos $-\alpha_o(1 - p_o, q_A)$ - y así interesa más aplicar la (2.50) que la (2.51) pues habrá que resolver en menos ocasiones la igualdad de la primera que la de la segunda.

Finalmente (p_A, q_A) puede estar a la derecha ($p_A > 0,5$) y la solución sería similar a la de antes, con las mismas dos posibilidades de forzar de antemano a que sea $q'_A = q_A$ o $p'_A = 1 - p_A$, aunque de nuevo la ventaja es para la primera opción. Ahora:

$$\begin{aligned} & \text{Si } p_A > 0,5: \text{ Hacer } q'_A = q_A \text{ y} \\ p'_A = & \begin{cases} p_o & \text{si } \alpha_o(p_A, q_A) \geq \alpha_o(p_o, q_A) \\ \text{la solución en } p (< p_o) \text{ de } \alpha_o(p_A, q_A) = \alpha_o(p, q_A) & \text{(en otro caso)} \end{cases} \end{aligned} \quad (2.52)$$

Con todo ello, dado un (p_A, q_A) dentro de las zonas de posible ahorro de espacio paramétrico, hay que determinar su simétrico $(p'_A, q'_A) = (p, q_A)$ de la zona opuesta, con p a obtener según la (2.50) o (2.52), e incorporarlo apropiadamente a las restricciones de la Propiedad III, lo que nos lleva a la siguiente propiedad que viene ilustrada en la Figura 2.5. Antes de ello hay que señalar que toda la argumentación está basada en suponer que RC_o consta de un único grupo de puntos equivalentes. Si, por causa de un empate en el valor α^* , la RC_o consta de 2 o más grupos de puntos equivalentes, como la argumentación anterior está basada en el menor marginal, la misma seguirá siendo válida para $p_o = \text{Min}(a_{1i}, a_{2i}, n_{1i}, n_{2i})/n$, con i aludiendo a los distintos grupos de empates. Con ello:

Propiedad IV: Para test de una cola, sea $p_o = \text{Min}(a_{1i}, a_{2i}, n_{1i}, n_{2i})/n$ para los marginales de todas las tablas de RC_o ; entonces:

- i) Si $p_o = p_A = 0,5$, el máximo está en $p = 0,5$ y $q_A \leq q \leq 0,5$.

- ii) Si $p_0 < p_A < 1-p_0$, el máximo está en la región de la (2.44).
 iii) En otro caso, el máximo está en:

$$\left\{ \begin{array}{l} q_A \leq p \leq p_L \\ q_A \leq q \leq p \end{array} \right\} \cup \left\{ \begin{array}{l} p_L < p \leq 0,5 \\ 1/n \leq q \leq p \end{array} \right\} \cup \left\{ \begin{array}{l} 0,5 < p \leq p_R \\ 1/n \leq q \leq 1-p \end{array} \right\} \cup \left\{ \begin{array}{l} p_R < p \leq 1-q_A \\ q_A \leq q \leq 1-p \end{array} \right\} \quad (2.53)$$

en donde:

- a) Si $p_0 \geq p_A$, serán $p_L = p_A$ y $p_R = 1-p_0$ si $\alpha_0(p_A, q_A) \geq \alpha_0(1-p_0, q_A)$ ó, en otro caso, p_R es la solución en p de la igualdad $\alpha_0(p_A, q_A) = \alpha_0(p, q_A)$ con $1-p_0 < p \leq 1-1/n$.
 b) Si $1-p_0 \leq p_A$, serán $p_R = p_A$ y $p_L = p_0$ si $\alpha_0(p_A, q_A) \geq \alpha_0(p_0, q_A)$ ó, en otro caso, p_L es la solución en p de la igualdad $\alpha_0(p_A, q_A) = \alpha_0(p, q_A)$ con $1/n \leq p < p_0$.

Si el investigador no considera rentable resolver en p las ecuaciones citadas, pues ello no resulta compensado por el ahorro de espacio paramétrico que se produce en el lado opuesto al de (p_A, q_A) , entonces puede aplicar la propiedad anterior tomando las dos primeras zonas de la (2.53) y la segunda de la (2.44), o las dos últimas zonas de la (2.53) y la primera de la (2.44), según el caso. Por otro lado, si para las tablas de la RC se ha adoptado la representación canónica aludida en la expresión (2.21), entonces es $p_0 = \text{Min}(n_{1i})/n$. Finalmente, la función clave $\alpha_0(p, q)$ aludida en varias ocasiones ya sabemos que es:

$$\alpha_0(p, q) = \sum_{RC_0} \frac{n!}{x_1! x_2! y_1! y_2!} p^{a_1} (1-p)^{a_2} q^{n_1} (1-q)^{n_2}$$

pero, en el caso de varios grupos de empates, puede convenir ponerla en el formato reducido que se indicó para $\alpha_1(p, q)$: expresión (2.34).

En el caso de un test de dos colas, el error α alude a la función $\alpha(p, q)$ y todo lo indicado en el Apartado anterior sigue siendo válido, pero restringido a la región que indica la expresión (2.45), lo que hace los cálculos más fáciles. Ahora no hay que forzar ningún "ahorro" por la derecha, y así:

Propiedad V: Para tests de dos colas, sea $p_0 = \text{Min}(a_{1i}, a_{2i}, n_{1i}, n_{2i})/n$ para los marginales de todas las tablas de RC_0 ; entonces:

- i) Si $p_0 = p_A = 0,5$, el máximo está en $p = 0,5$ y $q_A \leq q \leq 0,5$.
- ii) Si $p_0 < p_A$, el máximo está en la región de la (2.45).
- iii) En otro caso, el máximo está en:

$$\left\{ \begin{array}{l} q_A \leq p \leq p_A \\ q_A \leq q \leq p \end{array} \right\} \cup \left\{ \begin{array}{l} p_A < p \leq 0,5 \\ 1/n \leq q \leq p \end{array} \right\} \quad (2.54)$$

Una ilustración gráfica de la Propiedad V se da también en la Figura 2.5.

Obsérvese que tanto en las Propiedades IV como en la V se habla del caso especial en que $p_0 = 0,5$, en cuyo caso es $p_0 = q_0 = 0,5$. Ello sólo puede ocurrir cuando RC_0 consta de puntos equivalentes que verifican $a_1 = a_2 = n_1 = n_2 = n/2$, lo que sólo es factible si n es par. En tal caso, si $p_A = 0,5$ también, entonces el máximo ha de estar en $q_A \leq q \leq 0,5$ y $p = 0,5$ -como indican las propiedades- lo cual alude al caso especial de la línea vertical en $p = 0,5$ que es reflejado también en la Figura 2.5.

2.3.5. Reducción del espacio paramétrico en el caso de la región crítica inicial (atractores)

2.3.5.1. En el test de una cola

La RC inicial descrita en 2.2.4.4 tiene una forma especial que permite evaluar su máximo de un modo bastante abreviado. En ella es

$$\alpha 1(p, q) = \sum_{n_1=1}^{n-1} \binom{n}{n_1} p^{n_2} (1-p)^{n_1} q^{n_1} (1-q)^{n_2} = [q(1-p) + p(1-q)]^n - [p(1-q)]^n - [q(1-p)]^n \quad (2.55)$$

cuyo máximo estará entre los puntos (p, q) que hacen cero sus dos derivadas, y así la suma de ambas también será nula:

$$n(1-p-q) \{ 2[q(1-p) + p(1-q)]^{n-1} - [p(1-q)]^{n-1} - [q(1-p)]^{n-1} \} = 0$$

Como el término entre llaves es siempre positivo, el máximo sólo puede ocurrir en $1-p-q=0$, es decir, en la línea $p+q=1$ (la diagonal secundaria del espacio paramétrico). Haciendo $q=1-p$ en $\alpha 1'_p(p, q)=0$, dividiéndola por $n(1-p)^{2n-1}$, y haciendo $x=p/(1-p)$,

queda que $(1-x)(1+x^2)^{n-1} - 1 + x^{2n-1} = 0$, lo que es divisible por $1-x$ e indica (como ya se sabe por las simetrías) que $x=1$, es decir $p=0,5$, es siempre una posible solución. Dividiendo por $(1-x)$, desarrollando el binomio del primer término y agrupando las potencias pares con la impar anterior, se obtiene:

$$\sum_{h=1}^{n-2} x^{2(h-1)} \left\{ \left[\binom{n-1}{h} - 1 \right] x - 1 \right\} = x^{2n-4}$$

La solución con $x < 1$ (es decir, con $p < 0,5 \Rightarrow$ con $q > 0,5$) no interesa pues cae fuera de la región (2.44). Para $x > 1$ (es decir, $p > 0,5$) todos los términos son positivos, y así todos los términos del primer miembro han de ser menores que el segundo miembro, en particular el último: $x^{2n-6} \{(n-2)x-1\} < x^{2n-4} \Rightarrow x^2 - (n-2)x + 1 > 0$, lo que sólo sucede para $x > \{(n-2) + \sqrt{n(n-4)}\}/2$, es decir para $p > 1 - 2/\{n + \sqrt{n(n-4)}\}$. Con ello hay dos soluciones posibles: $\{p=q=0,5\}$ y $\{1 - 2/\{n + \sqrt{n(n-4)}\} < p \leq 1 - 1/n\}$. Sin embargo la primera da lugar siempre a un mínimo, pues haciendo $q=1-p$ en la (2.55), calculando la derivada segunda y sustituyendo en ella $p=0,5$, se obtiene $1/2^{n-2} - (2n-1)/2^{2n-3}$ que es positivo para $n \geq 4$ (únicos casos, salvo los $n=2$ y 3 , en que puede existir una RC lícita). Como para $n=3$ la solución está en los x que verifican $-x^2 + x - 1 = 0$ y no hay solución real, en tanto que para $n=2$ la solución está en los x que verifican $x=0$ que no es de interés (pues sería $p=0$), entonces:

Propiedad VI: Para $n \geq 4$, el máximo para la RC inicial de una cola a la izquierda (asociación negativa) se obtiene maximizando en p el error $\alpha_1(p, 1-p) = [(1-p)^2 + p^2]^n - p^{2n} - (1-p)^{2n}$ en la región:

$$1 - \frac{2}{n + \sqrt{n(n-4)}} < p \leq 1 - \frac{1}{n} \quad (2.56)$$

y el punto donde se alcanza el máximo es $(p, 1-p)$. Para el caso de $n=2$ o 3 , el máximo se alcanza en $(0,5; 0,5)$. Para la cola derecha (asociación positiva) el máximo es el mismo y se alcanza en $(1-p, 1-p)$.

La última afirmación se deduce a partir de la (2.38) pues en este caso, por ser la

primera RC, es $\alpha_1 = \alpha_2$. Se observa que para $n \rightarrow \infty$ es $p \rightarrow 1 - 1/n$, con lo que el máximo tiende a localizarse en la frontera.

2.3.5.2. En el test de dos colas

Ahora $\alpha(p, q) = \alpha_1(p, q) + \overline{\alpha_1}(p, q) = \alpha_1(p, q) + \alpha_1(1-p, q)$, con lo cual

$$\begin{cases} \alpha'_p(p, q) = \alpha'_1(p, q) - \alpha'_1(1-p, q) \\ \alpha'_q(p, q) = \alpha'_1(p, q) + \alpha'_1(1-p, q) \end{cases}$$

con $\alpha_1(p, q)$ dada por la (2.55). El máximo de $\alpha(p, q)$ estará en los puntos (p, q) que anulan las dos derivadas anteriores, con lo que su diferencia $\alpha'_p - \alpha'_q$ también será nula; dividiéndola por n y agrupando términos comunes, queda que:

$$(p-q) \{ 2 [[q(1-p) + p(1-q)]^{n-1} - [pq + (1-p)(1-q)]^{n-1}] + [(1-p)(1-q)]^{n-1} + [pq]^{n-1} - [p(1-q)]^{n-1} - [q(1-p)]^{n-1} \} = 0$$

Dividiendo todo por $(1-p)^{n-1} (1-q)^{n-1}$, llamando por $x = p/(1-p)$ e $y = q/(1-q)$ y teniendo en cuenta que $p-q = (x-y) \div (1+x)(1+y)$ queda:

$$\frac{x-y}{(1+x)(1+y)} [2 \{ (x+y)^{n-1} - (xy+1)^{n-1} \} + \{ 1 + x^{n-1} y^{n-1} - x^{n-1} - y^{n-1} \} - \{ (xy+1)^{n-1} - 1 - x^{n-1} y^{n-1} \}] = 0$$

Esa expresión es divisible por $x-y$, lo que indica que una posible solución para el máximo es $x=y$, o sea $p=q$. Dividiéndola por $(x-y)$, multiplicándola por $(1+x)(1+y)$ y reordenando términos, queda:

$$\begin{aligned} & [(x+y)^{n-1} - (xy+1)^{n-1}] + [\{ (x+y)^{n-1} - x^{n-1} - y^{n-1} \} - \{ (xy+1)^{n-1} - 1 - x^{n-1} y^{n-1} \}] \\ & - [(1+x)(1+y)(x^{n-2} + x^{n-3}y + x^{n-4}y^2 + \dots + y^{n-2})] = 0 \end{aligned} \tag{2.57}$$

en donde, a efectos de cálculo del máximo, basta restringirse a $y < x < 1$ dado que el máximo ha de encontrarse en el triángulo rectángulo de la Figura 2.4 (en el que $q < p < 0,5$). Veamos lo que sucede con los tres términos entre corchetes de la (2.57):

Primer término: Como $y < 1 = (1-x)/(1-x) \Rightarrow x+y < xy+1 \Rightarrow$ el primer término es negativo.

Tercer término: Todos sus sumandos son positivos y él mismo está precedido del signo negativo \Rightarrow el tercer término es negativo.

Segundo término: Desarrollando el binomio de Newton del primer elemento de cada corchete y operando, queda:

$$\sum_{h=2}^{n-2} x^h y^{n-h} - \sum_{h=2}^{n-2} x^h y^h = \sum_{h=2}^{n-2} x^h (y^{n-h} - y^h)$$

en donde:

- a) Cuando n es par, el término central $x^{n/2} y^{n/2} - x^{n/2} y^{n/2} = 0$
- b) El resto de los términos (que serán todos para n impar) cuentan con un sumando con exponente $h < n/2$ y otro con $(n-h) > n/2$ que es su simétrico.

Agrupándolos:

$$\sum_{h=2}^{[n/2]^-} \{x^h (y^{n-h} - y^h) + x^{n-h} (y^h - y^{n-h})\} =$$

$$- \sum_{h=2}^{[n/2]^-} (y^{n-h} - y^h) (x^{n-h} - x^h) > 0$$

pues al ser $n-h > h \Rightarrow y^{n-h} < y^h$ dado que $y < 1$, y de igual modo con el término en x 's. Con ello los dos paréntesis son negativos y su producto será positivo. Como el segundo término venía precedido de un signo menos \Rightarrow el segundo término es negativo.

Con todo lo anterior, las derivadas α'_p y α'_q no pueden ser cero fuera de $x=y$, es decir de $p=q$. Haciendo $q=p$ en $\alpha'_p(p,q)$ y dividiendo por n , queda que el máximo hay que buscarlo en los valores de p que satisfacen:

$$(1-2p)[2p(1-p)]^{n-1} - (1-2p)[p(1-p)]^{n-1} - (1-2p)[p^2+(1-p)^2]^{n-1} - p^{2n-1} + (1-p)^{2n-1} = 0$$

Dividiendo por $(1-p)^{2n-1}$ y haciendo $x=p/(1-p)$ como antes, entonces $1-2p=(1-x)(1-p)$ y la anterior queda:

$$(1-x)2^{n-1}x^{n-1} - (1-x)x^{n-1} - (1-x)(x^2+1)^{n-1} - x^{2n-1} + 1 = 0$$

la cual, dividiéndola por $1-x$, da:

$$(2^{n-1}-1)x^{n-1} - (x^2+1)^{n-1} + \{1+x+\dots+x^{2n-2}\} = 0 \quad (2.58)$$

El hecho de ser divisible por $1-x$ indica que $x=1$ es una posible solución, es decir $p=0,5$ es siempre una posible solución. Sin embargo la (2.58) no tiene una solución explícita fácil, pero experimentalmente puede comprobarse que suele ser cercana a

1/n.

De todo lo anterior se deduce que en el caso de los atractores para el test de dos colas el máximo ha de buscarse en $p=q$, con lo que la función a maximizar es $\alpha(p,q) = \alpha_1(p, q) + \alpha_1(1-p, p)$, con $\alpha_1(p, q)$ dada por la (2.55). Como:

$$\alpha(p,q) = [q(1-p) + p(1-q)]^n - [p(1-q)]^n - [q(1-p)]^n + [pq + (1-p)(1-q)]^n - [pq]^n - [(1-p)(1-q)]^n$$

entonces haciendo $p=q$ y agrupando términos queda la propiedad final:

Propiedad VII: El máximo para la RC inicial de dos colas se obtiene maximizando en $1/n \leq p \leq 0,5$ el error $\alpha(p,p) = (2^n - 2) [p(1-p)]^n + [p^2 + (1-p)^2]^n - p^{2n} - (1-p)^{2n}$ y el punto donde se alcanza el máximo es (p, p) .

Experimentalmente puede comprobarse que el máximo está en $p=0,5$ en los n pequeños y cerca de $p=1/n$ en los n grandes.

En la Figura 2.10 puede verse la forma de la función $\alpha(p, q)$ para las RC iniciales (las de los atractores) de una y de dos colas, en el caso de $n=10$. Como la inclusión de todos los atractores en la RC inicial no es obligatorio, con fines comparativos la Figura 2.11 representa también la función $\alpha(p, q)$ para $n=10$, pero para una RC que consta de tantos puntos como los atractores ($n-1$ y $2n-2$ respectivamente para una o dos colas), puntos que han sido introducidos por el criterio de Barnard.

2.3.6. Reducción del número de máximos a calcular con el método M

Dada una determinada RC_A y determinadas las tablas candidatas a entrar en ella en el siguiente paso, el método **M** exige formar las RC_i ($i=1, 2, \dots$) (obtenidas incluyendo en RC_A cada una de las tablas candidatas) y obtener sus errores $\alpha_i(p,q)$ y sus máximos α_i^* . Si $\text{Min } \alpha_i^* = \alpha_m^*$, entonces RC_m es la RC siguiente. La propiedad que sigue pretende evitar (cuando se pueda) la determinación de todos esos máximos, seleccionando la RC_m por un procedimiento más corto. La propiedad es aplicable, por

Figura 2.10
Gráfica del error $\alpha(p,q)$ en la RC inicial (atractores) para $n = 10$.

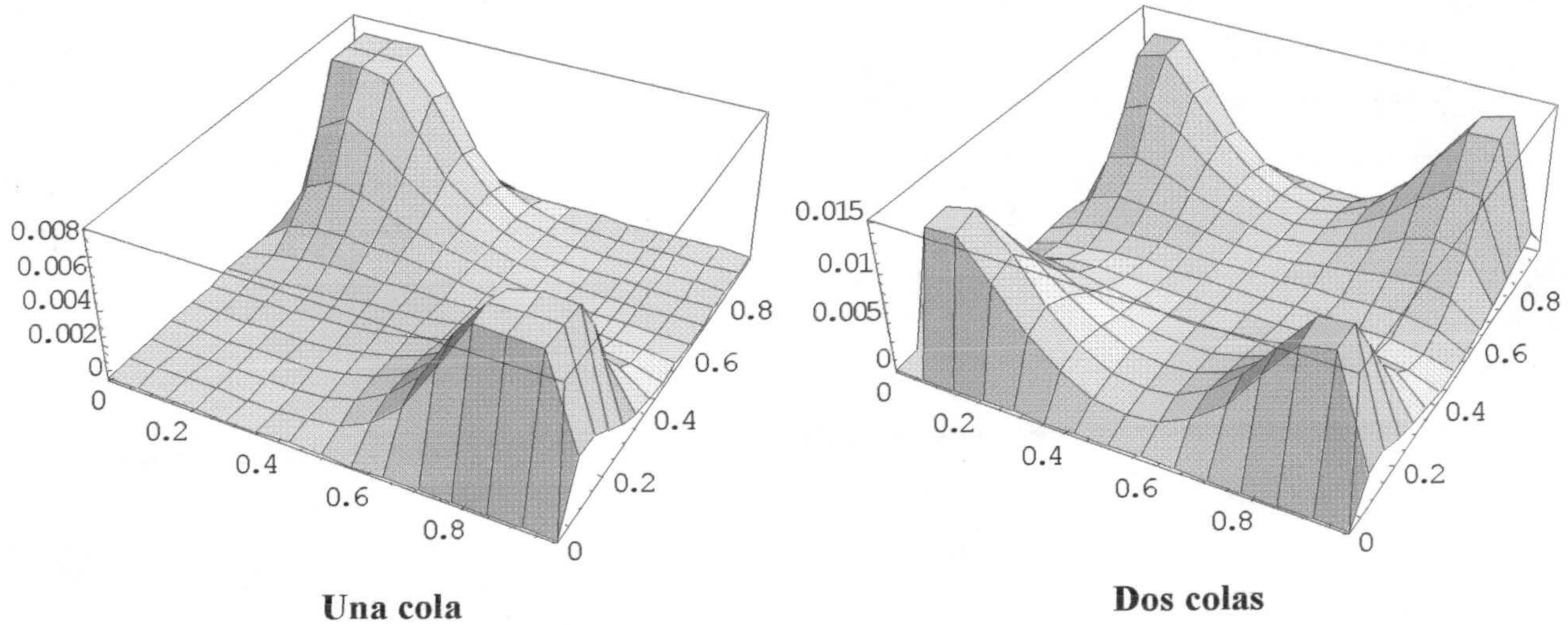
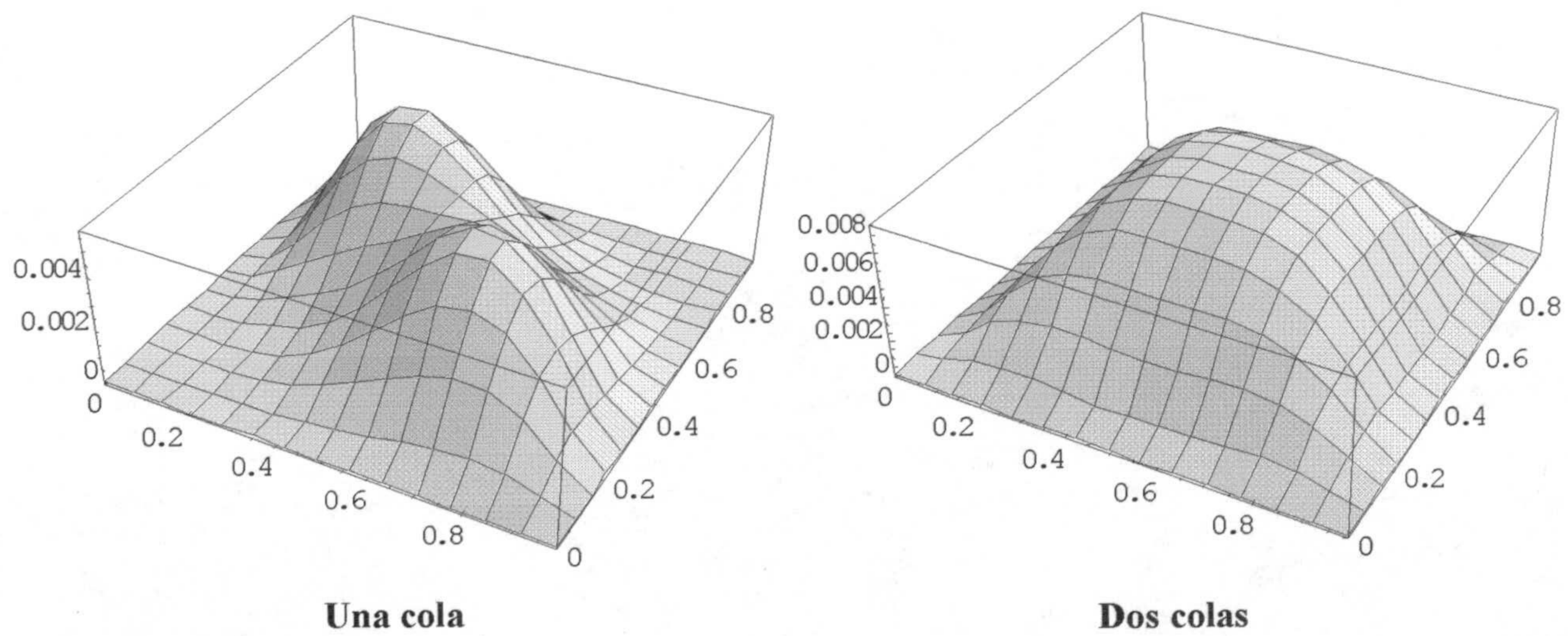


Figura 2.11
Gráfica del error $\alpha(p,q)$ en una RC con tantos puntos como los atractores para $n = 10$ y obtenida por la ordenación M.



tanto, sólo al método **M** (que es el más complejo).

Consideremos por ahora que RC_i-RC_A consta de una única tabla. Para ella, sean $\hat{p}_i = a_i/n, \hat{q}_i = n_i/n$ y sea $\alpha_i(\hat{p}_i, \hat{q}_i)$ el valor que producen al sustituirlos en $\alpha_i(p, q)$. Sea $\alpha_K(\hat{p}_K, \hat{q}_K)$ la menor de todas, y α_K^* el máximo de $\alpha_K(p, q)$. En tal caso, si sucede que $\alpha_i(\hat{p}_i, \hat{q}_i) > \alpha_K^* (\forall i \neq k)$, entonces $\alpha_i^* \geq \alpha_i(\hat{p}_i, \hat{q}_i)$ (por ser su máximo) $> \alpha_K^*$ (por hipótesis) $= \alpha_m^*$ (por definición), y así $RC_m = RC_K$. Con ello, si todos los $\alpha_i(\hat{p}_i, \hat{q}_i)$ son mayores que el α^* del menor de ellos, éste será el máximo de la RC a seleccionar.

Cuando RC_i-RC_A consta de varias tablas equivalentes, la propiedad anterior sigue valiendo si se determinan \hat{p}_i y \hat{q}_i en cualquiera de ellas, como garantiza la expresión (2.37). Si consta de más de un grupo de tablas equivalentes, cada grupo proporciona estimaciones \hat{p}_{ij} y \hat{q}_{ij} ($j=1, 2, \dots, n^\circ$ de grupos) y la propiedad valdrá para $\alpha_i(\hat{p}_i, \hat{q}_i) = \text{Min}_j \alpha_{ij}(\hat{p}_{ij}, \hat{q}_{ij})$, es decir si con $\alpha_i(\hat{p}_i, \hat{q}_i)$ se alude a la más pequeña que proporcionan las tablas empatadas. Para test de dos colas, las tablas simétricas no alteran el razonamiento, pues se verifica la expresión (2.39). Así:

Propiedad VIII: Sean RC_A una RC ya fijada y RC_i las regiones críticas posibles para el paso siguiente, con errores $\alpha_i(p, q)$. Si RC_i-RC_A consta de un único grupo de tablas equivalentes (ó simétricas, si el test es de dos colas), sean $\hat{p}_i = a_i/n$ y $\hat{q}_i = n_i/n$ obtenidos en cualquiera de ellas, sea $\alpha_i(\hat{p}_i, \hat{q}_i)$ y $\alpha_K(\hat{p}_K, \hat{q}_K) = \text{Min}_i \alpha_i(\hat{p}_i, \hat{q}_i)$ cuyo máximo es α_K^* . Entonces, si $\alpha_i(\hat{p}_i, \hat{q}_i) > \alpha_K^* (\forall i \neq k)$, la RC_K es la RC del paso siguiente.

Si alguna RC_i-RC_A consta de más de un grupo de tablas equivalentes (ó simétricas, en su caso) lo anterior sigue valiendo si se considera que $\alpha_i(\hat{p}_i, \hat{q}_i)$ alude al más pequeño valor de esa expresión en cada uno de los grupos citados.

La propiedad sustituye el cálculo de α_i^* por el de $\alpha_i(\hat{p}_i, \hat{q}_i)$, lo que hará disminuir drásticamente el tiempo de cómputo. Según la experiencia del caso de dos proporciones (Martín and Silva, 1994), los métodos **M** y **M'** tienen una potencia similar. De ser así también ahora, ello indica que sus ordenaciones son también

similares y que la condición en la que se basa la propiedad anterior se verificará con frecuencia.

2.3.7. Proceso numérico para la determinación del máximo

La determinación del máximo de $\alpha(p, q)$ no hay otro modo de hacerla que por procedimientos numéricos de ensayo-error. Ya se ha visto que la literatura lo determina evaluando $\alpha(p, q)$ para valores $p, q = 1\%(1\%)99\%$ en el caso más exhaustivo (Haber, 1987; Shuster, 1992), pero las propiedades reseñadas hasta ahora muestran la importancia de los valores de a_i (ó número de diagonales que intervienen) para el crecimiento o decrecimiento de $\alpha(p, q)$. Como ellos van de 1 a $(n-1)$ entonces los cortes claves a efectuar serán $p_i, q_j = 1/n, 2/n, \dots, (n-1)/n$ (complementado con las restricciones del espacio paramétrico que se hayan considerado).

Sea (p_i, q_j) , con $p_i = i/n$ y $q_j = j/n$, la pareja donde se obtuvo el máximo de $\alpha(p, q)$. Para determinarlo con más precisión, se consideran los 4 puntos con valores de p y q dados por:

$$p = p_i \pm \frac{1}{2n} \quad ,, \quad q = q_j \pm \frac{1}{2n}$$

y en ellos se calcula $\alpha(p, q)$ y se selecciona el que produce el máximo: digamos (p_I, q_I) . Ahora se prueba con los 4 puntos:

$$p = p_I \pm \frac{1}{4n} \quad ,, \quad q = q_I \pm \frac{1}{4n}$$

y se selecciona el (p_{II}, q_{II}) que proporciona el máximo. Diversas pruebas realizadas nos han mostrado que con esos dos cortes sucesivos es suficiente, y así se ha considerado que $\alpha^* = \alpha(p_{II}, q_{II})$ lo que concuerda numéricamente con los datos de las tablas de Shuster (1992) y del programa de Berger (1996) ya citado. La Figura 2.12 ilustra lo anterior.

2.3.8. Atractor teórico

Por su forma, y por lo que se verá más tarde, el método más potente es el **M** ¿Cuál es el primer punto, o atractor, que introduce **M** en la RC? La prueba teórica

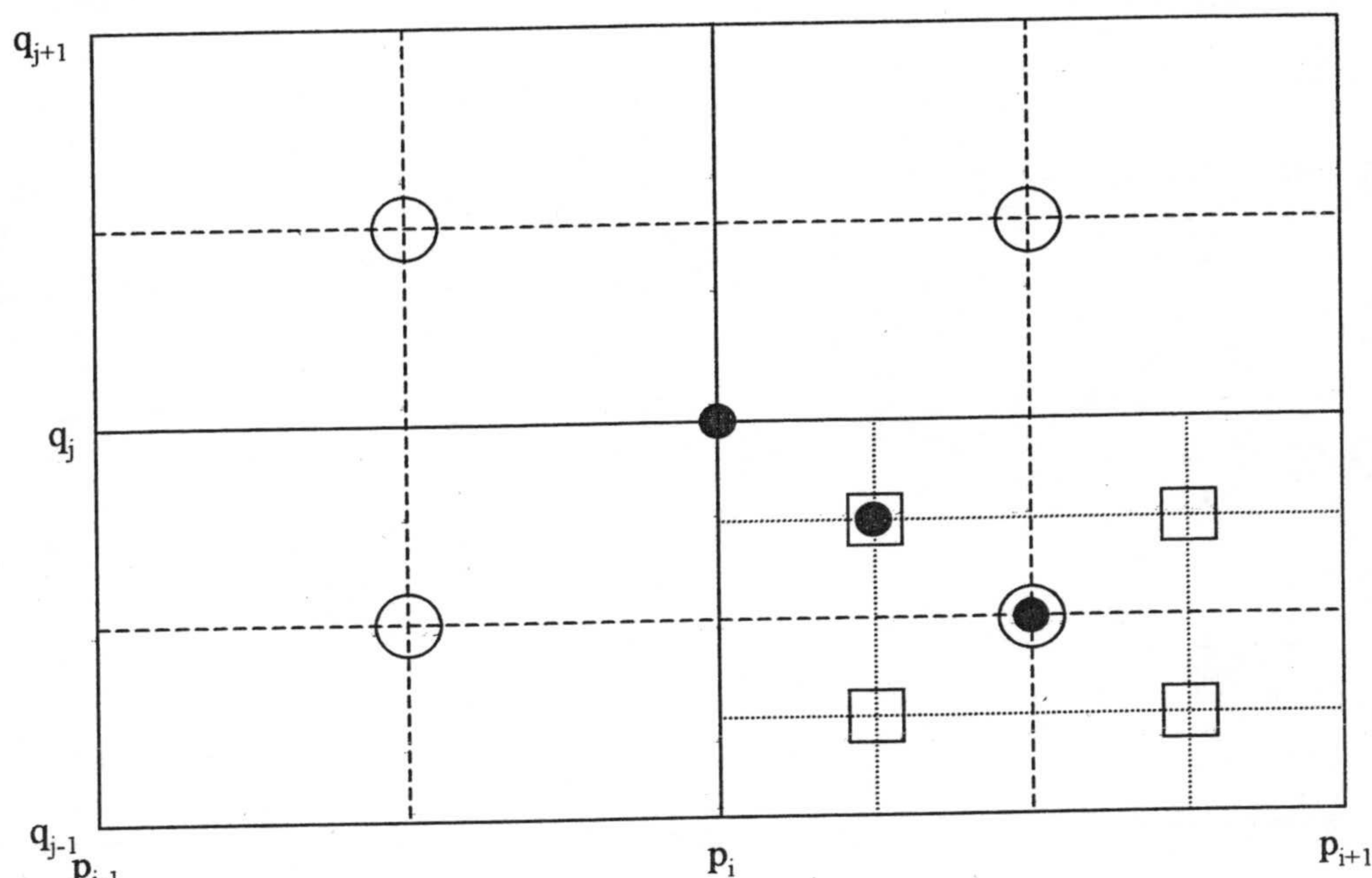


Figura 2.12

Si la pareja (p_i, q_j) , con $p_h = h/n$, es la que da un máximo de $\alpha(p, q)$, ella da el punto \bullet de la gráfica. En el primer corte se prueba con los puntos \circ de la gráfica, siendo el marcado con \odot el que da el máximo. En el segundo corte (y el último) se prueba con los puntos \square y el \square es el del máximo: el valor de $\alpha(p, q)$ en él será considerado como el valor de α^* .

de ello es complicada, pero prácticamente se ha comprobado que en la RC se introduce en primer lugar un único punto (más sus equivalentes y simétricos) en el que $x_1 = y_2 = 0$ y:

- 1) Si $n = \text{par}$, $y_1 = x_2 = n/2$
- 2) Si $n = \text{impar}$, $x_2 = [n/2]$ o $[n/2] - 1$ o $[n/2] - 2$ etc, con el resultado disminuyendo conforme aumenta n : el primero ocurre en $3 \leq n \leq 7$; el segundo en $9 \leq n \leq 25$; el tercero en $27 \leq n \leq 49$ (para una cola); o $n = 3$, $5 \leq n \leq 13$, $15 \leq n \leq 27$ (para dos colas).

Esto quiere decir que la RC de inicio debería dejarse libre para que sea el método utilizado el que la determine. Sin embargo, la idea anterior de atractores (los cuales entran rápidamente en la RC) nos ha permitido hacer algunas observaciones que serán de posterior utilidad en la discusión.

2.3.9. Discusión y conclusiones

Las propiedades dadas anteriormente permitirán acortar drásticamente el tiempo

de cálculo para el método incondicionado actual, haciéndolo así algo más asequible. En particular, la aplicación de la Propiedad II (la más simple de todas) permite disminuir el espacio paramétrico a menos de 1/4 o 1/8 de su tamaño (según que el test sea de una o de dos colas), en tanto que la Propiedad VIII hace practicable al método **M** (el más potente), el cual de otro modo sería prohibitivo en tiempo de cómputo. Las Propiedades III, IV y V son de especial utilidad (aunque complica la programación) cuando con cualquier método haya que determinar una RC o para el método **M**, y con ellas el espacio paramétrico suele reducirse a 1/2 o 2/3 del inicial (ya reducido por la Propiedad II).

Por otro lado, en la introducción se han citado algunas de las escasas aportaciones que la literatura hace al efecto, pero, a la luz de lo visto, conviene efectuar ahora algunos comentarios acerca de las mismas.

Para determinar el valor de α^* , Haber (1987) y Shuster (1992) se limitan a obtener el valor de $\alpha(p,q)$ en valores de p y q de 1% en 1% (!Kroll, 1989, de 10% en 10%!), evitando el primero y el último los valores inferiores a un 5%. En nuestra opinión tal modo de proceder es inapropiado y peligroso. Ya se ha visto la incidencia que tiene el valor de n , y del número de diagonales, en el espacio paramétrico simplificado y en el crecimiento o decrecimiento de $\alpha(p,q)$, de modo que la partición de aquel debería efectuarse primero en múltiplos de $1/n$ y, después, en cortes sucesivos del intervalo seleccionado. Practicando con el caso del Apartado **2.3.5**, puede observarse que pequeñas diferencias en el valor de p (bastante inferiores al 1%) ocasionan grandes diferencias en el valor de $\alpha(p,q)$. Por otro lado, la exclusión de los valores inferiores al 5% no está justificada salvo que sea $n \leq 20$. De hecho, de nuevo en el caso del Apartado **2.3.5**, el máximo está cercano a $1/n$ (precisamente la zona excluida cuando es $n < 20$).

En cuanto al lugar del máximo, Kroll (1989) y Shuster (1992) observan empíricamente que el máximo se alcanza con frecuencia en $p=q$, y en particular en $p=q=0,5$. Ya se ha visto que $p=q$, $p=1-q$ y $p=q=0,5$ son lugares privilegiados en general (por las simetrías), y que para tests dos colas también lo es $p=0,5$. De hecho, las dos primeras igualdades dan el lugar de búsqueda para los casos del Apartado

2.3.5, pero no está garantizado que eso sea siempre así. De modo general, y para las RC con muchos puntos, los nuevos puntos que se incorporan a ella tienden a estar en las diagonales más informativas (aquellas con valores a_1 y n_1 cercanos a $n/2$), lo que produce valores p_0 cercanos a 0,5 y favorece que la localización del máximo sea cercana a $p=q=0,5$ (recordar la Propiedad III). De un modo experimental, y para tablas con $n \leq 40$, los porcentajes experimentales de localizaciones del máximo son (para el método **M**):

<u>Localización</u>	<u>1 cola</u>	<u>2 colas</u>
$p=q = 0,5$	7%	10%
$p=q \neq 0,5$	5%	69%
$p=1-q \neq 0,5$	31%	—
$p=0,5 \neq q$	—	9%

de modo que $p+q=1$ (diagonal secundaria del espacio paramétrico) sucede 1/3 de las veces en los tests de una cola y $p=q \neq 0,5$ (diagonal principal del espacio paramétrico) sucede en 2/3 de las ocasiones en los tests de dos colas, siendo las demás posibilidades bastante más minoritarias. En particular, un 57% de las veces (1 cola) y un 12% de las veces (2 colas) el máximo se produce en valores (p,q) diferentes de los lugares de simetría de $\alpha(p, q)$.

2.4. POTENCIA DEL TEST Y SELECCIÓN DEL MÉTODO ÓPTIMO

2.4.1. Los procedimientos clásicos y sus restricciones

En estadística es tradición comparar los diversos procedimientos para efectuar un test a través de la potencia de los mismos. Para nuestro caso, dada una cierta RC para valores dados de n y de α , la potencia en la alternativa (p_{11}, p_{12}, p_{21}) es:

$$\Theta(p_{11}, p_{12}, p_{21} | \alpha) = \sum_{RC} \frac{n!}{x_1! y_1! x_2! y_2!} p_{11}^{x_1} p_{12}^{y_1} p_{21}^{x_2} p_{22}^{y_2} \quad (2.59)$$

lo que implica dos restricciones (Martín and Silva, 1994, en el contexto del Modelo II).

En primer lugar, una tabla de potencias con tres entradas (p_{11} , p_{12} , p_{21}) se vuelve confusa pues ningún método es uniformemente más potente que otro al no estar siempre la RC de uno contenida en la del otro, lo que obliga a efectuar algunas simplificaciones. Así, ya se dijo que Haber (1987), en el único trabajo en el que se comparan potencias para el caso de asociación, determina la mínima potencia para valores dados de la odds-ratio (que es una buena medida de la lejanía de H_0)

$$\Theta(\theta | \alpha) = \underset{p_{11}, p_{12}}{\text{Min}} \left\{ \Theta(p_{11}, p_{12}, p_{21} | \alpha) \mid p_{11}p_{22}/p_{12}p_{21} = \theta, \sum \sum p_{ij} = 1 \right\} \quad (2.60)$$

procedimiento que aún conserva la variabilidad en el parámetro desconocido θ (lo que le obliga a matizar las conclusiones en función de él) y que pierde la información que implica la minimización en p_{11} y p_{12} . Martín and Silva (1994), en un contexto similar, argumentaron que tal procedimiento es el adecuado para planificar el tamaño de muestra (pudiendo afirmar así que "la potencia cuando θ vale ... es cómo mínimo de ..."), pero no para comparar dos métodos.

La segunda restricción alude a la utilización de un único y mágico error objetivo α . Dichos autores opinan que es injusto decidirse por un método para $\alpha=5\%$ (valor que no será alcanzado, por tratarse de variables discretas) y no tener en cuenta qué sucede para $\alpha=4,5\%$ por ejemplo.

2.4.2. Resolución de la primera restricción: potencia a largo plazo

Martín and Silva (1994) argumentaron a favor de efectuar las comparaciones de potencia en base a la potencia promedio alcanzada en todo el espacio paramétrico, lo que lleva a asignarle al mismo una distribución "a priori". Asumamos que a largo plazo (en todas las ocasiones en que el investigador se enfrente en el futuro a un problema como éste), los parámetros $\{p_{ij}\}$ siguen una distribución uniforme. Alternativa, y equivalentemente, aceptemos que, puesto que las p_{ij} son desconocidas, cualquier valor de ellas es igualmente posible (Rice, 1988, en otro contexto, da una argumentación más detallada), lo que nos lleva de nuevo a la uniforme. Para nuestro caso esto significa que $(p_{11}, p_{12}, p_{21}, p_{22} = 1 - p_{11} - p_{12} - p_{21})$ sigue la distribución de Dirichlet $D_i (v_{11} = v_{12} = v_{21} = v_{22} = 1)$, que es la distribución conjugada natural para los parámetros

de la distribución multinomial, y, en base a ella y a la expresión (2.59), hay que determinar la potencia promedio que llamaremos *potencia a largo plazo*.

Si S alude al simplex $S = \{p_{ij} \mid p_{ij} \geq 0, \sum \sum p_{ij} = 1\}$, como la función de densidad de la distribución de Dirichlet anterior es $3! = 6$, entonces la potencia a largo plazo para el test de dos colas se obtiene integrando la (2.59):

$$\begin{aligned} \Theta(\alpha) &= 6 \int_S \Theta(p_{11}, p_{12}, p_{21} \mid \alpha) dp_{11} dp_{12} dp_{21} \\ &= \sum_{RC} \frac{6 \times n!}{x_1! y_1! x_2! y_2!} \int_S p_{11}^{x_1} p_{12}^{y_1} p_{21}^{x_2} p_{22}^{y_2} dp_{11} dp_{12} dp_{21} \quad (2.61) \\ &= \sum_{RC} \frac{6}{(n+1)(n+2)(n+3)} \end{aligned}$$

pues la última integral es igual a $x_1! y_1! x_2! y_2! / (n+3)!$ por corresponderse con una D_i ($v_{11} = x_1 + 1$; $v_{12} = y_1 + 1$; $v_{21} = x_2 + 1$; $v_{22} = y_2 + 1$). Como, por otro lado, el número de tablas del espacio muestral es $(n+1)(n+2)(n+3)/6$ (Pirie and Hamdan, 1972), entonces: *la potencia a largo plazo para el test de dos colas es:*

$$\Theta(\alpha) = \frac{\text{n}^\circ \text{ de tablas de la RC al error } \alpha}{\text{n}^\circ \text{ de tablas del espacio muestral}} \quad (\text{dos colas}) \quad (2.62)$$

y así, para valores fijados de n y de α , una medida de la potencia de un método la da el número de tablas de que consta su RC. El criterio, de gran fuerza intuitiva, fue utilizado por Upton (1982) sin demostración y coincide con el de Martín and Silva (1994), ambos en el contexto de 1 marginal fijado. Obsérvese que el criterio de Haber dado por la (2.60) hace esto mismo, pero sin ponderar cada punto del espacio paramétrico por la potencia alcanzada en él. Para Haber, el espacio paramétrico es el conjunto de todas las parejas de superficies de valor θ o θ^{-1} (por tratarse de un test de dos colas), y a cada pareja le asigna un peso igual a la mínima potencia alcanzada en ella, perdiendo la información de las potencias individuales alcanzadas en las demás parejas (p_{11} , p_{12}).

Para el test de una cola la demostración es algo más larga. Si la alternativa es $\theta < 1$ (asociación negativa), entonces la primera integral de la (2.61) debe restringirse al simplex $S' = \{p_{ij} \mid p_{ij} \geq 0, \sum \sum p_{ij} = 1, \theta < 1\}$, con lo que la densidad de la uniforme

en él será de $2 \times 3! = 12$. Con ello, la potencia a largo plazo será ahora:

$$\begin{aligned} \Theta(\alpha) &= 12 \int_{S'} \Theta(p_{11}, p_{12}, p_{21} | \alpha) dp_{11} dp_{12} dp_{21} \\ &= \sum_{RC} \frac{12 \times n!}{x_1! y_1! x_2! y_2!} \int_{S'} p_{11}^{x_1} p_{12}^{y_1} p_{21}^{x_2} p_{22}^{y_2} dp_{11} dp_{12} dp_{21} \quad (2.63) \\ &= \sum_{RC} \frac{12}{(n+1)(n+2)(n+3)} P\{\theta < 1\} \end{aligned}$$

pues la segunda integral alude, salvo constante, a $P\{\theta < 1\}$ en la distribución $D_i(x_1+1, y_1+1, x_2+1, y_2+1)$. Altham (1969) obtuvo que:

$$P\{\theta < 1\} = \sum_{s=\text{Max}\{x_2-y_1, 0\}}^{x_2} \binom{a_1+1}{s} \binom{a_2+1}{n_2+1-s} / \binom{n+2}{n_1+1} \quad (2.64)$$

lo cual, reordenando términos, y teniendo en cuenta que $x_2 - y_1 = a_1 - n_1$, da:

$$\sum_{s=\text{Max}\{a_1-n_1, 0\}}^{x_2} \binom{n_2+1}{s} \binom{n_1+1}{a_1+1-s} / \binom{n+2}{a_1+1} \quad (2.65)$$

cantidad que está referida a la casilla de x_2 . A fin de referirnos a la casilla de x_1 (como es habitual) hagamos $r = a_1 + 1 - s$ en la (2.65), con lo cual:

$$\begin{aligned} P\{\theta < 1\} &= \sum_{r=x_1+1}^{\text{Min}\{a_1+1, n_1+1\}} \binom{n_1+1}{r} \binom{n_2+1}{a_1+1-r} / \binom{n+2}{a_1+1} \quad (2.66) \\ &= P_{\bar{F}}(x_1+1, y_1, x_2, y_2+1) \end{aligned}$$

en donde la última igualdad es un modo nemotécnico de aludir al término anterior recordando que él es el valor P del test exacto de Fisher de una cola para la alternativa contraria ($H_1 \equiv \theta > 1$) a la deseada para el test incondicional ($H_1 \equiv \theta < 1$) (y de ahí la negación sobre el símbolo F) para los datos de una tabla como la que se indica entre paréntesis y que se da completa en la Tabla 2.4. Con todo ello:

$$\Theta(\alpha) = \frac{12}{(n+1)(n+2)(n+3)} \sum_{RC} P_{\bar{F}}(x_1+1, y_1, x_2, y_2+1) \text{ (una cola)} \quad (2.67)$$

es decir, la *potencia a largo plazo para el test de una cola* es un promedio de valores

P del test exacto de Fisher, resultado también similar al de Martín and Silva (1994) para el caso del Modelo II. La expresión (2.67) requiere casi la cuarta parte de los cálculos aparentes. La tabla a que alude proviene de la Tabla 1.1(a), y las tablas equivalentes a esta difieren sólo en la permuta de 1 o 2 de sus diagonales, por lo que dan lugar a la misma tabla base para la (2.67) (con iguales permutas), y todas ellas dan igual valor P de Fisher (Martín and Luna, 1990). Como $H_1 \equiv \theta > 1$ es equivalente a $H_1 \equiv \theta^{-1} < 1$, la potencia para el test de una cola a la derecha es la misma que la del caso anterior si se entiende que la (2.67) está aplicada a las tablas de la RC de aquel caso con sus dos columnas permutadas.

Tabla 2.4

Tabla 2×2 a la que se alude con $P_{\bar{F}}$ en la notación de la Tabla 1.1(a)

$x_1 + 1$	y_1	$n_1 + 1$
x_2	$y_2 + 1$	$n_2 + 1$
$a_1 + 1$	$a_2 + 1$	$n + 2$

2.4.3. Resolución de la segunda restricción: potencia media

Martín and Silva (1994) observaron que la potencia a largo plazo $\bar{\Theta}(\alpha)$ aún conserva una fuerte dependencia del error objetivo α utilizado y que, con frecuencia, ello da lugar a empates entre métodos, cosa que no sucede si se amplía la gama de valores de α . Asimismo indicaron que un investigador deseará el test óptimo para una determinada tabla experimental, pero la primera significación de ella se alcanzará en un valor de α que, siendo su valor P, no será uno de los tradicionales. Todo ello les llevó a definir la *potencia media* de un test en el intervalo de interés ($\alpha_I \leq \alpha \leq \alpha_S$) a través de la asignación de la distribución uniforme a α (en el sentido de que cualquier valor de α entre α_I y α_S es igualmente razonable), con densidad $(\alpha_S - \alpha_I)^{-1}$, obteniendo que ella es:

$$\bar{\Theta}(\alpha_I, \alpha_S) = \frac{\alpha_S A(\alpha_S) - \alpha_I A(\alpha_I)}{\alpha_S - \alpha_I} \quad (2.68)$$

con

$$A(\alpha) = \Theta(\alpha) - \frac{1}{\alpha} \sum_{i=1}^r \left\{ \Theta(\alpha_i^*) - \Theta(\alpha_{i-1}^*) \right\} \alpha_i^* \quad (2.69)$$

y α_i^* aludiendo a los tamaños, obtenidos por la (2.2), de cada una de las RC intermedias y final que se obtienen hasta llegar a la RC objetivo al error α . La asignación de la distribución uniforme a α se ve bastante razonable si se recuerda que los valores P (los α_i^* de aquí) siguen la distribución uniforme.

La observación citada de Martín and Silva (1994) para el caso de 1 marginal fijado ha sido comprobada por los autores para el caso actual de 0 marginales fijados (aunque ahora los empates no son tan frecuentes), por lo que su argumento es de aplicación al caso actual. Los datos se dan más adelante. Pero, a fin de particularizar la (2.68) al caso de asociación en tablas 2×2 , es preciso describir el proceso de obtención de los datos. Dados el error objetivo α y el método de ordenación elegido, dicho método irá introduciendo tablas en la RC dando lugar a sucesivas $RC_0 = \phi$, RC_1 , RC_2 , ..., $RC_r = RC(\alpha)$ cuyos tamaños, obtenidos por la (2.2), serán $\alpha_0^* = 0 < \alpha_1^* < \alpha_2^* < \dots < \alpha_r^* \leq \alpha$, valores a los que alude la (2.69). Cada RC_i constará de N_i tablas ($N_0 = 0$, N_1 , N_2 , ..., $N_r = N(\alpha)$) y diferirá de la anterior en el conjunto de tablas $D_i = RC_i - RC_{i-1}$ cuyo número es $\Delta_i = N_i - N_{i-1}$. Con ello, y según la (2.62), los numeradores de $\Theta(\alpha_i^*)$, $\Theta(\alpha_i^*) - \Theta(\alpha_{i-1}^*)$ y $\Theta(\alpha)$ son, respectivamente, N_i , $N_i - N_{i-1} = \Delta_i$ y $N(\alpha)$, con lo cual la (2.69) será:

$$A(\alpha) = 6 \frac{\alpha N(\alpha) - \sum \Delta_i \alpha_i^*}{\alpha(n+1)(n+2)(n+3)} \quad (2.70)$$

y $\bar{\Theta}(\alpha_I, \alpha_S)$ se obtiene sustituyendo la (2.70) en la (2.68):

$$\bar{\Theta}(\alpha_I, \alpha_S) = 6 \frac{\alpha_S N(\alpha_S) - \alpha_I N(\alpha_I) - \sum_{i=I+1}^{r_S} \Delta_i \alpha_i^*}{(\alpha_S - \alpha_I)(n+1)(n+2)(n+3)} \quad (\text{dos colas}) \quad (2.71)$$

con $r = r_I(r_S)$ para $\alpha = \alpha_I(\alpha_S)$, lo que da la *potencia media para el test de dos colas*.

Para el test de una cola, y con la notación de antes, ahora hay que especificar el valor de la (2.69) en base a la (2.67). Con tal fin hay que tener en cuenta que,

salvo constante, el numerador de $\Theta(\alpha_i^*) - \Theta(\alpha_{i-1}^*)$ es la suma de los valores de $P_{\bar{F}}$ en el conjunto D_i , es decir:

$$P_{\bar{F}}(D_i) = \sum_{(x_1, y_1, x_2, y_2) \in D_i} P_{\bar{F}}(x_1+1, y_1, x_2, y_2+1) \quad (2.72)$$

en tanto que el numerador de $\Theta(\alpha)$ será $\sum_i P_{\bar{F}}(D_i)$. Con ello:

$$A(\alpha) = \frac{12}{\alpha(n+1)(n+2)(n+3)} \sum_{i=1}^r (\alpha - \alpha_i^*) P_{\bar{F}}(D_i) \quad (2.73)$$

y así $\Theta(\alpha_I, \alpha_S)$ se obtiene sustituyendo la (2.73) en la (2.68):

$$\bar{\Theta}(\alpha_I, \alpha_S) = 12 \frac{(\alpha_S - \alpha_I) \sum_{i=1}^{r_I} P_{\bar{F}}(D_i) + \sum_{i=r_I+1}^{r_S} (\alpha_S - \alpha_i^*) P_{\bar{F}}(D_i)}{(\alpha_S - \alpha_I)(n+1)(n+2)(n+3)} \quad (\text{una cola}) \quad (2.74)$$

lo que da la *potencia media para el test de una cola*. Todos estos resultados son bastante similares a los de Martín and Silva (1994), de los que difieren sólo en las constantes. Ahora, como antes, la expresión (2.72) requiere de casi la cuarta parte de los cálculos aparentes: Si D_i consta de un único conjunto de tablas equivalentes, basta calcular $P_{\bar{F}}$ para una de ellas y multiplicarlo por 1, 2 o 4 según el número de tablas equivalentes.

2.5. SELECCIÓN DEL MÉTODO INCONDICIONADO ÓPTIMO*

2.5.1. Introducción

En 2.2.4.3 se indicaron los diferentes métodos de ordenación propuestos (unos nuevos y otros no), siendo el objetivo actual el proceder a su comparación a fin de seleccionar el óptimo, comparación que se efectuará a través de las potencias a largo plazo $\Theta(\alpha | n)$ y/o la potencia media $\bar{\Theta}(\alpha_I, \alpha_S)$ lo reseñado en la sección 2.4. El método óptimo, en principio, será el que mayor potencia alcance. Pero también conviene evaluar la pérdida de potencia que se sufre por no usar el óptimo, en relación a la ganancia de tiempo de cómputo que se obtiene.

Como ya se indicó, los métodos en competición son los M , M' , $F1$, $F2$, $FM1$, $FM2$, Z , $PH1$, $PH2$, $PH3$ e Y para el caso del test de dos colas, en tanto que para el de una cola serán esos mismos menos los $F2$ y $FM2$ (que en tal situación no tienen sentido).

2.5.2. Datos a fijar de antemano

Dado un método X cualquiera, para determinar su valor de Θ o de $\bar{\Theta}$ es preciso fijar las condiciones iniciales:

- 1) Nº de colas del test, especificando si es de una o de dos colas.
- 2) Tamaño n de muestra: Para que las comparaciones a efectuar sean válidas es preciso realizarlas para varios valores de n , de modo que las posibles tendencias con el aumento de n se pongan de manifiesto. Desgraciadamente, los problemas de cómputo (especialmente para el método M) obliga a recortar bastante el máximo n a alcanzar (aunque, como se verá, no hay ningún indicio de que ello vaya a afectar a las principales conclusiones). Por todo ello en lo que sigue se eligen valores de

$$n = 13, 20, 25 \text{ y } 30 \quad (2.75)$$

salvo las modificaciones justificadas que se indiquen más adelante. Estos valores de n cubren las cuatro situaciones que requiere el método $PH3$.

- 3) Errores de interés: La comparación en base a Θ se efectuará a un error α u otro, error que habrá que especificar. Aquí se acogen los clásicos

$$\alpha = 1\%, 5\%, 10\% \quad (2.76)$$

para las altas significaciones, las significaciones usuales y los indicios de significación respectivamente. Cuando las comparaciones se efectúan en base a $\bar{\Theta}$, los intervalos para α serán los procedentes de los cortes anteriores:

$$(\alpha_I, \alpha_S) = (0\%; 1\%), (1\%; 5\%), (5\%; 10\%) \quad (2.77)$$

2.5.3. Procedimiento para la obtención de los resultados

Una vez fijado el método de ordenación X a evaluar, el n° de colas del test, el valor n del tamaño de muestra y el error objetivo α , se procede como sigue. En todo el proceso se alude a los conceptos de las secciones 2.2, 2.3 y 2.4.

0º Paso: Partir de una $RC_0 = \phi$ con $N_0 = 0$ puntos y máximo $\alpha_0^* = 0$.

1º Paso: Seleccionar el conjunto de puntos D_1 que según el método X deben introducirse primero en la RC, formando así la RC_1 con N_1 de puntos. Calcular su máximo α_1^* , anotando la pareja (p_1, q_1) donde se obtiene el mismo y el valor de $P_{\bar{F}}$ para tales puntos.

2º Paso: Seleccionar el conjunto de puntos D_2 que, según el método de ordenación X , deben añadirse a la RC_1 para formar la RC_2 . Calcular su número N_2 , el valor $P_{\bar{F}}$ para ellos, el máximo α_2^* y la pareja (p_2, q_2) donde se le obtiene.

3º Paso: Fijada la RC_2 , proceder como en el 2º Paso para obtener así $D_3, RC_3, N_3, P_{\bar{F}}, \alpha_3^*$ y (p_3, q_3) . Etc para los demás pasos.

Fin del proceso: En cada paso comparar el α_i^* obtenido con el error objetivo α .

Entonces:

- 1) Si $\alpha_i^* < \alpha$: Continuar con el paso siguiente
- 2) Si $\alpha_i^* = \alpha$: Final del proceso. La RC buscada es la actual RC_i .
- 3) Si $\alpha_i^* > \alpha$: Final del proceso. La RC buscada es la RC_{i-1} del paso anterior.

Paso Final: Detenido el proceso, ya se disponen de todos los datos para calcular $\Theta(\alpha | n)$ en base a las (2.62) o (2.67).

Nota 1: Si el objetivo es evaluar $\bar{\Theta}(\alpha_1, \alpha_s)$, el proceso anterior se efectuará primero para $\alpha = \alpha_1$, continuándolo luego hasta $\alpha = \alpha_s$. La aplicación de las (2.71) o (2.74) dan la solución. En realidad se ahorra tiempo si desde el principio se plantea un $\alpha = 10\%$ y se llevan controles individuales para los $\alpha = 1\%$ y $\alpha = 5\%$ intermedios, pues así se evita la repetición de cálculos comunes.

Nota 2: En todo lo anterior se ha propuesto conservar los valores (p_i, q_i) pues ellos son los que hacen el oficio de (p_A, q_A) o de (p_N, q_N) en las Propiedades III, IV y V.

Nota 3: La RC_1 puede ser la que indique el método **X** (y será distinta de unos métodos a otros) o la de los atractores, si acaso se opina que debe ser esta. De ser esto último, el valor α_1^* se determinará simplifícadamente a través de las Propiedades VI y VII.

2.5.4. Comentarios previos a la obtención de los resultados definitivos

2.5.4.1. Acerca de la RC inicial

Ya se ha indicado que la RC inicial puede ser la que especifique el método **X** o la de los atractores señalada en 2.2.4.4. (que es común a todos los métodos). Decidirse por una u otra es algo opinable (ver la justificación de entonces), aunque desde el punto de vista de la potencia es preferible la primera opción (que constará de menos puntos). Desde el pnto de vista teórico también hay razones para ello: el marginal $a_1 = [n/2]^+$ es más informativo que el marginal $a_1 = 1$, y así ambas tablas no tienen porqué entrar juntas. De todos modos, en la práctica hemos comprobado que los atractores entran rápidamente en la RC y que, salvo para valores de n muy pequeños, decidirse o no por ellos no afecta a la potencia final para los errores α con que aquí se trabaja (como se comprueba más tarde).

El mantenimiento del concepto de atractores prestó su utilidad a la discusión de 2.3.8.

2.5.4.2. Acerca de la definición de potencia a considerar

Ya se ha señalado anteriormente la conveniencia de elegir el método óptimo en base a los valores $\bar{\Theta}$ de potencia media más que en base a los valores Θ de la potencia a largo plazo. Pero conviene ratificar tal cosa de modo numérico. Con tal fin la Tabla 2.5 presenta los valores de $\Theta(\alpha)$ en los métodos estudiados (considerando los atractores). Comparándolos con los de la posterior Tabla 2.7, se observa que el valor

TABLA 2.5: Potencia a largo plazo (con atractores)
 Para cada número de colas del test (columnas izquierda y derecha), cada valor de n (cabecera de cada subtabla), cada método (primera columna) y cada error α (primera fila) en el interior de cada subtabla se da la potencia a largo plazo $\Theta(\alpha)$ en %.

UNA COLA					DOS COLAS				
n=13	α	0,01	0,05	0,1	n=13	α	0,01	0,05	0,1
M		0,0	31,4	42,1	M		0,0	25,7	32,1
M'		0,0	31,4	42,9	M'		0,0	25,7	34,3
FM1		0,0	27,9	40,0	FM2		0,0	24,3	33,6
F1		0,0	27,9	38,6	FM1		0,0	22,1	27,9
Y		0,0	27,9	38,6	Y		0,0	22,1	27,9
PH2		0,0	30,7	40,0	F1		0,0	22,1	27,9
PH1		0,0	29,3	38,6	F2		0,0	22,1	30,0
PH3		0,0	29,3	38,6	PH2		0,0	23,6	30,7
Z		0,0	27,9	38,6	PH1		0,0	24,3	32,1
					PH3		0,0	24,3	32,1
					Z		0,0	24,3	32,1

n=20	α	0,01	0,05	0,1	n=20	α	0,01	0,05	0,1
M		23,6	40,4	52,2	M		20,7	34,9	43,6
M'		23,8	39,5	49,5	M'		20,9	34,4	41,8
FM1		22,9	38,6	49,5	FM2		19,9	32,0	41,1
F1		22,5	37,3	46,3	FM1		17,3	32,0	38,6
Y		22,9	37,3	46,3	Y		19,1	30,8	37,3
PH2		21,1	39,1	49,9	F1		16,8	30,4	37,3
PH1		13,8	34,7	49,7	F2		19,1	31,1	36,8
PH3		13,3	33,3	49,7	PH2		19,5	32,2	39,1
Z		12,6	33,1	49,7	PH1		13,8	33,3	40,7
					PH3		13,8	33,3	41,1
					Z		12,6	33,1	41,1

n=25	α	0,01	0,05	0,1	n=25	α	0,01	0,05	0,1
M		29,2	45,8	56,2	M		26,0	39,7	47,6
M'		29,2	43,8	52,7	M'		25,8	39,1	47,1
FM1		28,3	43,5	53,4	FM2		24,4	38,2	43,8
F1		28,2	42,9	52,1	FM1		23,2	35,9	43,5
Y		27,8	42,9	52,0	Y		23,6	35,9	42,9
PH2		22,7	42,2	54,3	F1		23,0	35,9	42,9
PH1		12,5	33,2	53,4	F2		22,7	35,0	42,1
PH3		12,2	33,0	52,9	PH2		22,7	38,2	44,3
Z		11,7	32,7	52,9	PH1		12,4	33,2	46,3
					PH3		12,2	33,0	46,8
					Z		11,7	32,7	46,8

n=30	α	0,01	0,05	0,1	n=30	α	0,01	0,05	0,1
M		33,7	49,8	59,5	M		30,1	43,5	51,4
M'		33,0	47,8	54,8	M'		30,2	42,7	49,5
FM1		32,7	48,5	57,5	FM2		29,5	41,4	49,7
F1		32,5	47,3	55,8	FM1		27,8	40,7	48,6
Y		32,0	47,6	55,8	Y		27,9	40,9	47,8
PH2		21,9	46,6	57,8	F1		27,2	40,3	47,4
PH1		11,5	39,8	53,7	F2		27,1	40,1	46,5
PH3		11,3	39,8	53,7	PH2		21,9	42,2	49,0
Z		11,2	39,8	53,7	PH1		11,5	39,8	50,4
					PH3		11,3	39,8	50,4
					Z		11,2	39,8	50,4

de $\bar{\Theta}$ respeta el orden de selección que hace Θ y, eventualmente, lo matiza, permitiendo efectuar desempates. Por todo ello, en lo que sigue los datos se ofrecen para la potencia media $\bar{\Theta}$.

2.5.4.3. Acerca de la paridad de n

En la sección 2.2 ya se observó que el hecho de que n sea par o impar influye en el número de tablas equivalentes. Esto quiere decir que la potencia puede variar sustancialmente al pasar del caso $n = n_0$ al caso $n = n_0 + 1$, especialmente para valores pequeños de n_0 (pues su espacio muestral es más reducido). De ser así, el hecho de que n sea par o impar habría de contemplarse en la comparación de métodos. La Tabla 2.6 (que también considera a los atractores) muestra que la paridad de n prácticamente no influye, y de ahí que en lo que sigue se contemplen exactamente los valores de dados por la (2.75).

2.5.5. Resultados y su análisis

Los valores de $\bar{\Theta}(\alpha_1, \alpha_S)$ para valores de $n = 13, 20, 25, 30$; valores de (α_1, α_S) de $(0\%; 1\%)$, $(1\%, 5\%)$ y $(5\%, 10\%)$; y para los 9 métodos a comparar para el test de una cola o los 11 métodos para el test de dos colas, se ofrecen en la Tabla 2.7 (con atractores) y en la Tabla 2.8 (sin atractores). Se observa la ventaja en potencia que se obtiene por no considerar los atractores (como se hará en adelante).

Para hacer más fácil la lectura comparativa de los resultados, la Figura 2.13 representa gráficamente los datos de la Tabla 2.8, omitiendo los métodos menos competitivos en general (los Z, PH3, PH2 y PH1, todos ellos de tipo chi-cuadrado). El análisis de estos datos se efectúa a continuación.

De un modo general se observan que las diferencias entre métodos se hacen más grandes conforme aumenta n (al principio solo) o decrece el error α o el número de colas del test, y que con el aumento de n los métodos mantienen en general sus posiciones relativas (lo que hace que la selección del óptimo sea bastante estable). Se observa asimismo que, salvo en el test de dos colas a errores α grandes, los métodos

Z, **PH3** y **PH1** son los peores. De hecho se ve que cuanto más grande es la corrección por continuidad mejor es el método (**Z**, **PH3**, **PH1**, **PH2** e **Y** es el orden de peor a mejor), al menos para los errores usuales del 0% al 5%, y que el método del P-mid mejora la actuación de la ordenación de Fisher (los métodos de tipo **FM** son mejores que los métodos de tipo **F**). Los métodos **PH1** y **PH3**, por tener una c.p.c. muy suave, suelen ser parecidos.

Observando la potencia media, se ve que el método **M** es el más potente, especialmente en los grandes valores de n o de α . El lector puede comprobar que, de modo simbólico (en cuanto a la potencia del método que se escriba) los métodos se ordenan en general así:

Test de una cola: $M > > M' > FM1 > F1 \approx Y > PH2 > > PH1 \approx PH3 > Z$

Test de dos colas: $M > > M' > FM2 > FM1 \approx Y > F1 \approx F2 > PH2 > PH1 \approx PH3 > Z$

pudiéndolos agrupar, de un modo general y de mayor a menor potencia, en los siguientes grupos:

$$\{M, M', FM1, FM2\} \quad \{F1, F2, Y\} \quad \{PH1, PH2, PH3, Z\}$$

Ya se ha dicho que el único trabajo comparativo existente para el caso de 0 marginales fijados (Modelo I) es el de Haber (1987), el cual evalúa la potencia en base a la (2.60). Sus resultados aluden sólo al test de dos colas, al error $\alpha=5\%$ y a los métodos **Z** y **F2** (aunque, para éste último, no parece tener en cuenta la condición C). Su conclusión de que **Z** es mejor que **F2** es compatible con lo que sucede en la Tabla 2.8 para $\alpha = 5\%-10\%$, pero para $\alpha = 0\%-5\%$ ocurre al contrario.

Los resultados anteriores son globalmente compatibles con los obtenidos por Martín and Silva (1994) y Martín, Sánchez and Silva (1998) para el caso de 1 marginal fijado (Modelo II), lo que indica que la asignación de la distribución uniforme a 3 (como se hace aquí) o a 2 (como hacen los autores anteriores) de los parámetros relacionados con los de la Tabla 1.1(b) no afecta a las conclusiones.

TABLA 2.6: Potencia media vs paridad n (con atractores)

Para cada número de colas del test (columnas izquierda y derecha), cada intervalo de valores de α (cabecera de la subtabla), cada método (primera columna) y cada valor de n (primera fila) en el interior de cada subtabla se da la potencia media $\bar{\Theta}(\alpha_1, \alpha_2)$ en %.

UNA COLA

$\alpha = 0\% - 1\%$

Métodos	ln	13	14	19	20	24	25
M		0,0	0,0	7,7	8,5	14,4	15,3
M'		0,0	0,0	6,6	7,6	11,2	12,2
FM1		0,0	0,0	6,4	7,4	11,1	11,9
F1		0,0	0,0	6,4	7,3	10,9	11,8
Y		0,0	0,0	6,2	7,0	10,5	11,5
PH2		0,0	0,0	5,9	6,6	9,1	9,8
PH1		0,0	0,0	4,7	5,0	5,8	6,0
PH3		0,0	0,0	4,5	4,9	5,7	5,9
Z		0,0	0,0	4,4	4,7	5,6	5,7

$\alpha = 1\% - 5\%$

Métodos	ln	13	14	19	20	24	25
M		23,7	25,9	32,8	34,0	37,7	38,6
M'		23,3	25,2	32,0	32,9	37,0	37,7
FM1		22,0	24,0	31,3	32,4	36,6	37,4
F1		22,0	23,7	30,8	31,8	36,0	36,9
Y		21,8	23,9	31,0	31,9	36,2	36,9
PH2		21,1	23,0	28,7	29,3	32,3	33,4
PH1		17,4	18,1	21,5	21,8	23,8	24,2
PH3		17,4	17,9	20,8	21,4	23,3	24,0
Z		16,7	17,8	20,2	20,9	23,2	23,7

$\alpha = 5\% - 10\%$

Métodos	ln	13	14	19	20	24	25
M		36,8	38,3	45,0	46,3	50,5	51,4
M'		36,9	37,9	44,4	45,1	48,6	48,8
FM1		35,5	36,5	43,5	45,0	48,8	49,8
F1		34,8	35,6	42,5	43,7	47,6	48,5
Y		34,8	35,5	42,5	43,7	47,6	48,4
PH2		35,6	37,4	44,2	45,2	48,8	49,6
PH1		33,9	35,2	40,2	41,0	43,6	44,4
PH3		33,9	34,6	40,0	40,5	43,2	43,9
Z		32,8	34,5	39,4	40,4	43,1	43,6

DOS COLAS

$\alpha = 0\% - 1\%$

Métodos	ln	13	14	20	21	25	26	30	31
M		0,0	0,0	6,7	7,3	10,6	11,9	15,0	15,7
M'		0,0	0,0	6,8	7,7	11,0	11,8	15,0	15,8
FM2		0,0	0,0	6,6	7,5	10,7	11,4	14,6	15,1
FM1		0,0	0,0	6,1	7,3	10,3	10,7	14,1	14,7
Y		0,0	0,0	6,1	7,2	10,3	10,8	14,1	14,6
F1		0,0	0,0	6,0	7,3	10,2	10,7	14,0	14,6
F2		0,0	0,0	6,2	7,4	10,1	10,7	14,0	14,4
PH2		0,0	0,0	6,2	7,1	9,8	10,3	11,7	12,1
PH1		0,0	0,0	5,0	5,2	6,0	6,2	6,4	6,6
PH3		0,0	0,0	4,9	5,1	5,9	6,0	6,4	6,4
Z		0,0	0,0	4,7	5,0	5,7	5,9	6,3	6,3

$\alpha = 1\% - 5\%$

Métodos	ln	13	14	20	21	25	26	30	31
M		19,2	21,0	29,1	30,2	34,1	34,9	38,1	38,8
M'		19,1	20,8	28,7	29,9	33,5	34,4	37,6	38,3
FM2		18,3	20,0	27,8	28,9	32,3	33,6	36,8	37,4
FM1		17,5	18,0	25,8	27,2	31,0	32,0	35,2	36,2
Y		17,9	18,4	26,2	27,3	30,9	32,2	35,1	36,1
F1		17,5	17,9	25,7	27,0	30,7	31,7	34,8	35,7
F2		17,5	18,3	25,7	26,8	30,3	31,4	34,4	35,3
PH2		18,3	19,3	26,4	27,9	30,9	31,7	34,2	35,1
PH1		16,7	17,1	21,6	22,4	24,2	24,8	26,4	27,0
PH3		16,8	17,1	21,4	22,1	24,0	24,3	26,1	26,7
Z		16,4	17,1	20,9	21,8	24,0	24,0	25,9	26,4

$\alpha = 5\% - 10\%$

Métodos	ln	13	14	20	21	25	26	30	31
M		29,3	31,2	39,1	40,2	44,0	44,8	47,8	48,5
M'		29,6	31,5	38,7	39,7	43,3	44,0	46,7	47,4
FM2		27,4	29,8	38,1	38,6	41,9	43,0	46,5	46,8
FM1		24,7	28,0	35,6	36,5	40,5	41,6	44,7	45,4
Y		24,6	27,7	35,1	36,2	40,1	40,9	44,1	44,8
F1		24,6	27,6	34,9	35,7	39,8	40,7	43,9	44,6
F2		25,5	27,9	34,2	35,2	39,2	40,2	43,2	43,9
PH2		27,0	29,3	37,0	37,8	41,7	42,5	45,4	46,4
PH1		27,7	30,6	37,9	38,4	42,0	43,3	46,2	46,8
PH3		27,7	30,6	37,8	38,6	42,1	43,3	47,8	46,7
Z		28,2	30,5	37,7	38,4	42,1	43,2	45,9	46,6

TABLA 2.7: Potencia media (con atractores)

Para cada número de colas del test (columnas izquierda y derecha), cada intervalo de valores de α (cabecera cada subtabla), cada método (primera columna) y cada valor de n (primera fila) en el interior de cada subtabla se da la potencia media $\bar{\Theta}(\alpha_1, \alpha_2)$ en %.

UNA COLA

DOS COLAS

$\alpha = 0\% - 1\%$

$\alpha = 0\% - 1\%$

Métodos \n	13	20	25	30
M	0,0	8,5	15,3	21,2
M'	0,0	7,6	12,2	16,4
FM1	0,0	7,4	11,9	16,2
F1	0,0	7,3	11,8	16,0
Y	0,0	7,0	11,5	15,4
PH2	0,0	6,6	9,8	11,7
PH1	0,0	5,0	6,0	6,4
PH3	0,0	4,9	5,9	6,4
Z	0,0	4,7	5,7	6,3

Métodos \n	13	20	25	30
M	0,0	6,7	10,6	15,0
M'	0,0	6,8	11,0	15,0
FM2	0,0	6,6	10,7	14,6
FM1	0,0	6,1	10,3	14,1
Y	0,0	6,1	10,3	14,1
F1	0,0	6,0	10,2	14,0
F2	0,0	6,2	10,1	14,0
PH2	0,0	6,2	9,8	11,7
PH1	0,0	5,0	6,0	6,4
PH3	0,0	4,9	5,9	6,4
Z	0,0	4,7	5,7	6,3

$\alpha = 1\% - 5\%$

$\alpha = 1\% - 5\%$

Métodos \n	13	20	25	30
M	23,7	34,0	38,6	42,8
M'	23,3	32,9	37,7	41,7
FM1	22,0	32,4	37,4	41,7
F1	22,0	31,8	36,9	41,0
Y	21,8	31,9	36,9	41,0
PH2	21,1	29,3	33,4	35,9
PH1	17,4	21,8	24,2	26,4
PH3	17,4	21,4	24,0	26,1
Z	16,7	20,9	23,7	25,9

Métodos \n	13	20	25	30
M	19,2	29,1	34,1	38,1
M'	19,1	28,7	33,5	37,6
FM2	18,3	27,8	32,3	36,8
FM1	17,5	25,8	31,0	35,2
Y	17,9	26,2	30,9	35,1
F1	17,5	25,7	30,7	34,8
F2	17,5	25,7	30,3	34,4
PH2	18,3	26,4	30,9	34,2
PH1	16,7	21,6	24,2	26,4
PH3	16,8	21,4	24,0	26,1
Z	16,4	20,9	24,0	25,9

$\alpha = 5\% - 10\%$

$\alpha = 5\% - 10\%$

Métodos \n	13	20	25	30
M	36,8	46,3	51,4	55,1
M'	36,9	45,1	48,8	52,1
FM1	35,5	45,0	49,8	53,4
F1	34,8	43,7	48,5	52,1
Y	34,8	43,7	48,4	52,1
PH2	35,6	45,2	49,6	53,4
PH1	33,9	41,0	44,4	48,1
PH3	33,9	40,5	43,9	46,0
Z	32,8	40,4	43,6	47,7

Métodos \n	13	20	25	30
M	29,3	39,1	44,0	47,8
M'	29,6	38,7	43,3	46,7
FM2	27,4	38,1	41,9	46,5
FM1	24,7	35,6	40,5	44,7
Y	24,6	35,1	40,1	44,1
F1	24,6	34,9	39,8	43,9
F2	25,5	34,2	39,2	43,2
PH2	27,0	37,0	41,7	45,4
PH1	27,7	37,9	42,0	46,2
PH3	27,7	37,8	42,1	47,8
Z	28,2	37,7	42,1	45,9

TABLA 2.8: Potencia media (sin atractores)

Para cada número de colas del test (columnas izquierda y derecha), cada intervalo de valores de α (cabecera de cada subtabla), cada método (primera columna) y cada valor de n (primera fila) en el interior de cada subtabla se da la potencia media en $\bar{\Theta}(\alpha_I, \alpha_S)$ %.

UNA COLA
 $\alpha = 0\% - 1\%$

Métodos \n	13	20	25	30
M	8,8	17,8	23,0	27,3
M'	9,6	17,9	22,5	26,7
FM1	9,5	17,2	22,1	26,2
F1	9,5	17,1	21,9	25,9
Y	9,7	16,6	20,3	23,5
PH2	6,5	10,0	12,6	14,1
PH1	3,4	4,8	5,9	6,4
PH3	3,4	4,7	5,9	6,4
Z	0,0	4,7	5,7	6,3

DOS COLAS
 $\alpha = 0\% - 1\%$

Métodos \n	13	20	25	30
M	7,8	15,9	20,7	24,7
M'	8,0	15,5	20,1	23,9
FM2	7,8	15,2	19,5	23,6
FM1	7,4	13,7	18,5	22,5
Y	7,5	14,4	18,8	22,2
F1	7,4	13,6	18,3	22,4
F2	7,3	13,7	18,3	22,1
PH2	6,2	9,7	12,6	14,1
PH1	3,4	4,8	5,9	6,4
PH3	3,4	4,7	5,9	6,4
Z	0,0	4,7	5,7	6,3

$\alpha = 1\% - 5\%$

Métodos \n	13	20	25	30
M	23,4	33,9	38,5	42,7
M'	23,7	32,9	37,6	41,7
FM1	22,4	32,3	37,4	41,7
F1	22,3	31,8	36,9	41,0
Y	22,4	31,9	36,9	41,0
PH2	21,3	29,3	33,4	36,0
PH1	17,3	21,8	24,2	26,4
PH3	17,3	21,4	24,0	26,1
Z	16,7	20,9	23,7	25,9

$\alpha = 1\% - 5\%$

Métodos \n	13	20	25	30
M	19,4	29,2	33,6	38,2
M'	19,5	28,4	32,8	36,4
FM2	18,4	27,8	32,3	36,8
FM1	17,4	25,6	30,9	35,2
Y	17,9	26,2	30,9	35,1
F1	17,4	25,5	30,6	34,7
F2	17,3	25,3	30,2	34,3
PH2	18,4	26,4	30,9	34,2
PH1	16,7	21,7	24,2	26,5
PH3	16,8	21,4	24,0	26,1
Z	16,4	20,9	23,7	25,9

$\alpha = 5\% - 10\%$

Métodos \n	13	20	25	30
M	35,3	46,1	51,4	55,1
M'	36,9	45,1	49,2	52,1
FM1	35,5	45,0	49,8	53,4
F1	34,8	43,7	48,5	52,1
Y	34,8	43,7	48,4	52,1
PH2	35,6	45,2	49,6	53,4
PH1	33,9	41,0	44,4	48,1
PH3	33,9	40,5	43,9	46,0
Z	32,8	40,5	43,7	47,7

$\alpha = 5\% - 10\%$

Métodos \n	13	20	25	30
M	29,4	39,3	44,0	47,9
M'	29,6	38,1	42,9	46,0
FM2	27,4	38,1	41,9	46,5
FM1	24,7	35,6	40,5	44,7
Y	24,6	35,1	40,1	44,1
F1	24,6	34,9	39,8	43,9
F2	25,5	33,9	39,2	43,1
PH2	27,0	37,0	41,7	45,4
PH1	27,7	37,9	42,0	46,3
PH3	27,7	37,8	42,1	47,8
Z	28,2	37,8	42,1	46,0

Figura 2.13: Potencia media (datos de la Tabla 2.8)

Para cada intervalo de valores de α (cabecera de cada gráfica) y cada número de colas del test (partes izquierda y derecha de cada página), se representa la potencia media $\bar{\Theta}(\alpha_I, \alpha_S)$, en %, en función del valor de n y para cada uno de los métodos indicados (su orden de escritura es el de aparición, por la derecha, en la gráfica).

(a) 0% - 1%

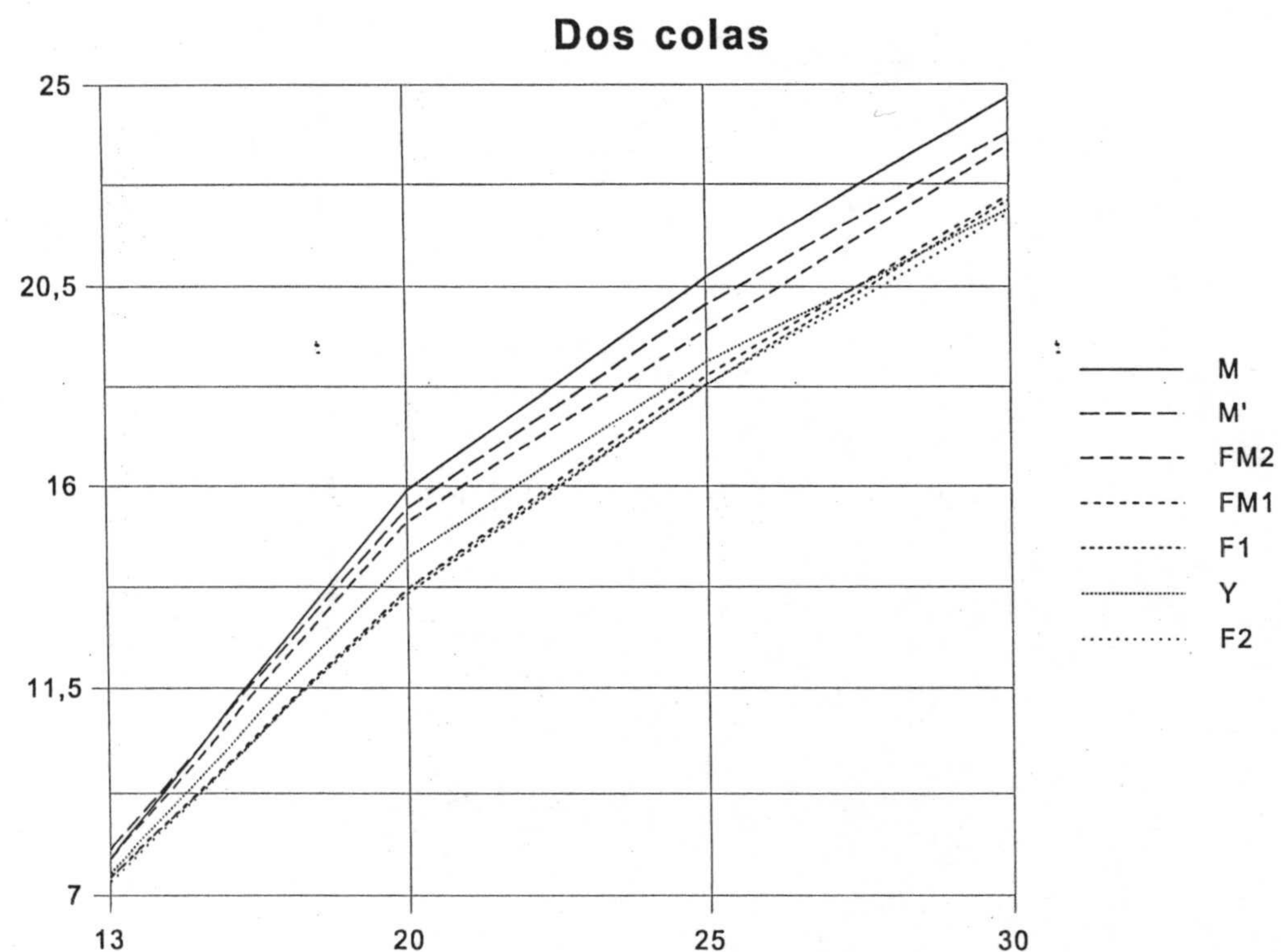
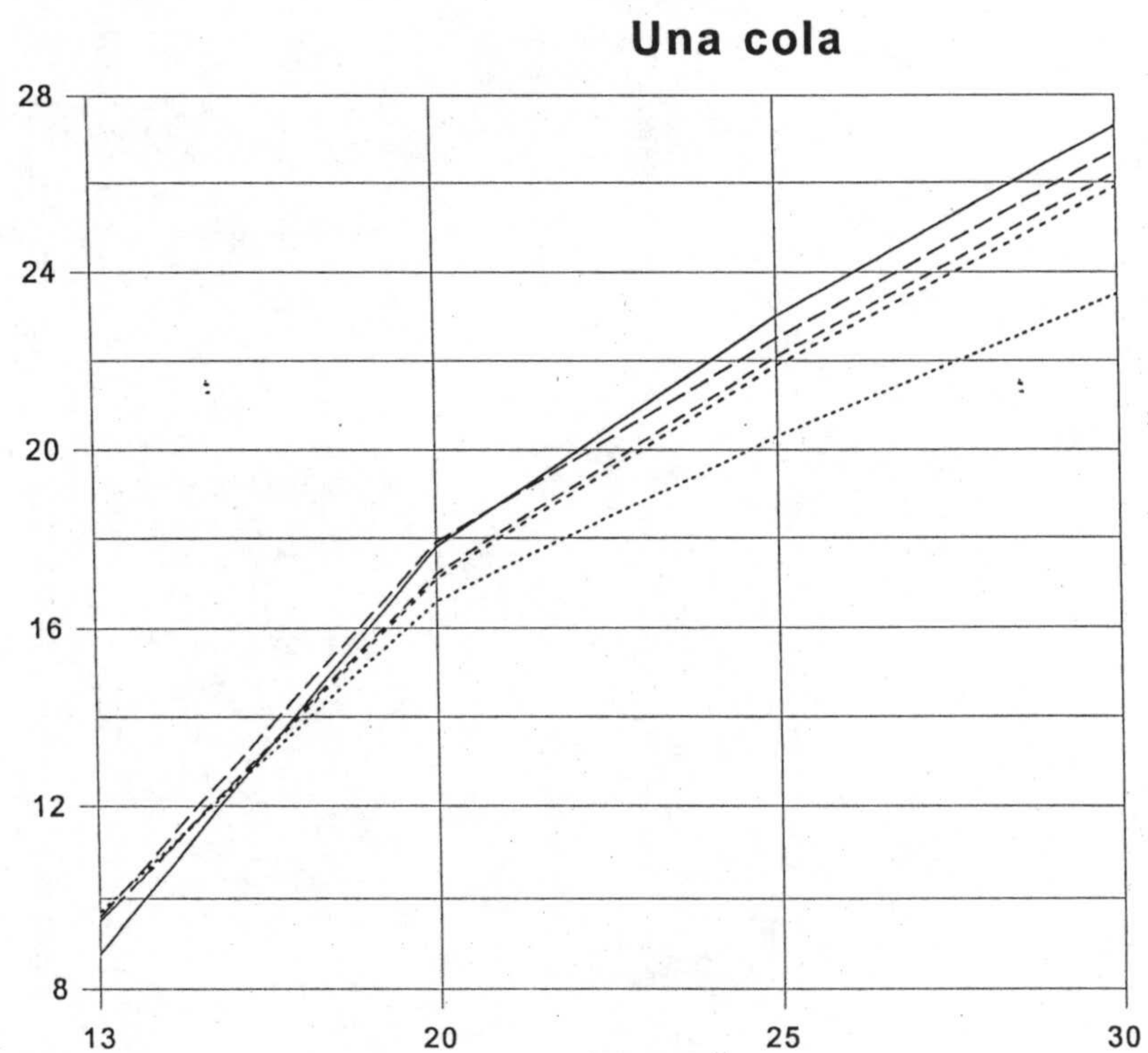
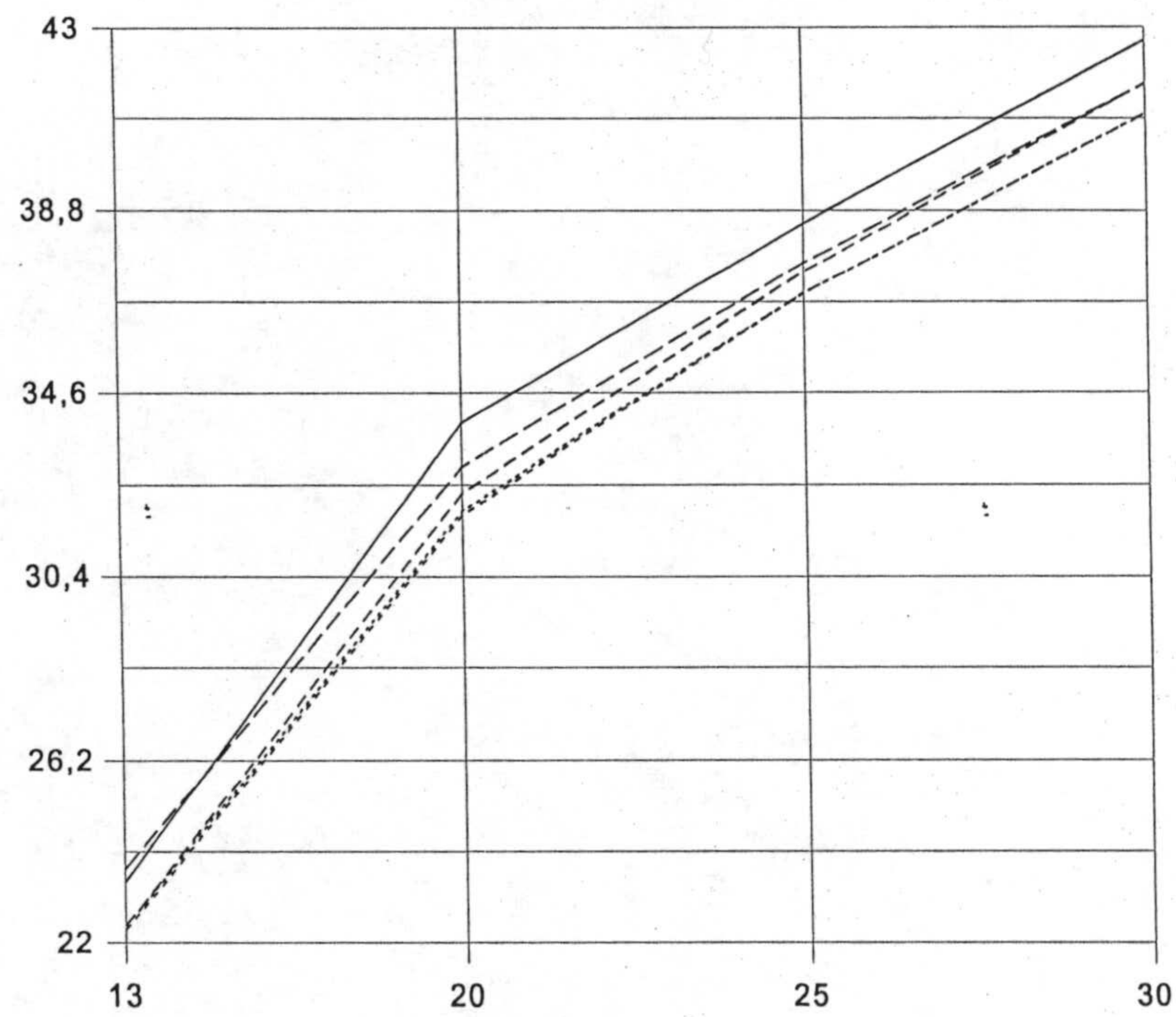


Figura 2.13

(b) 1% - 5%

Una cola



Dos colas

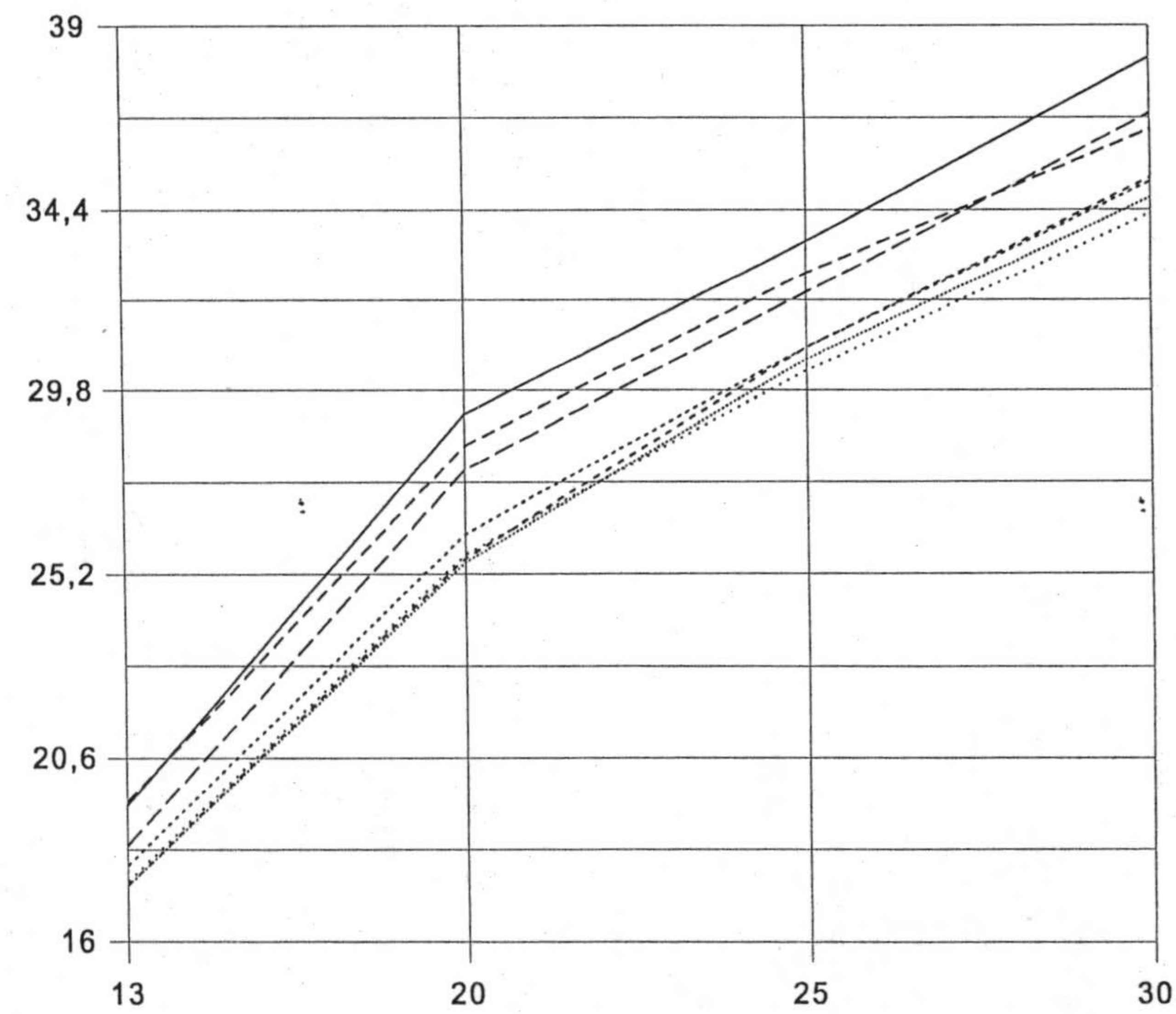
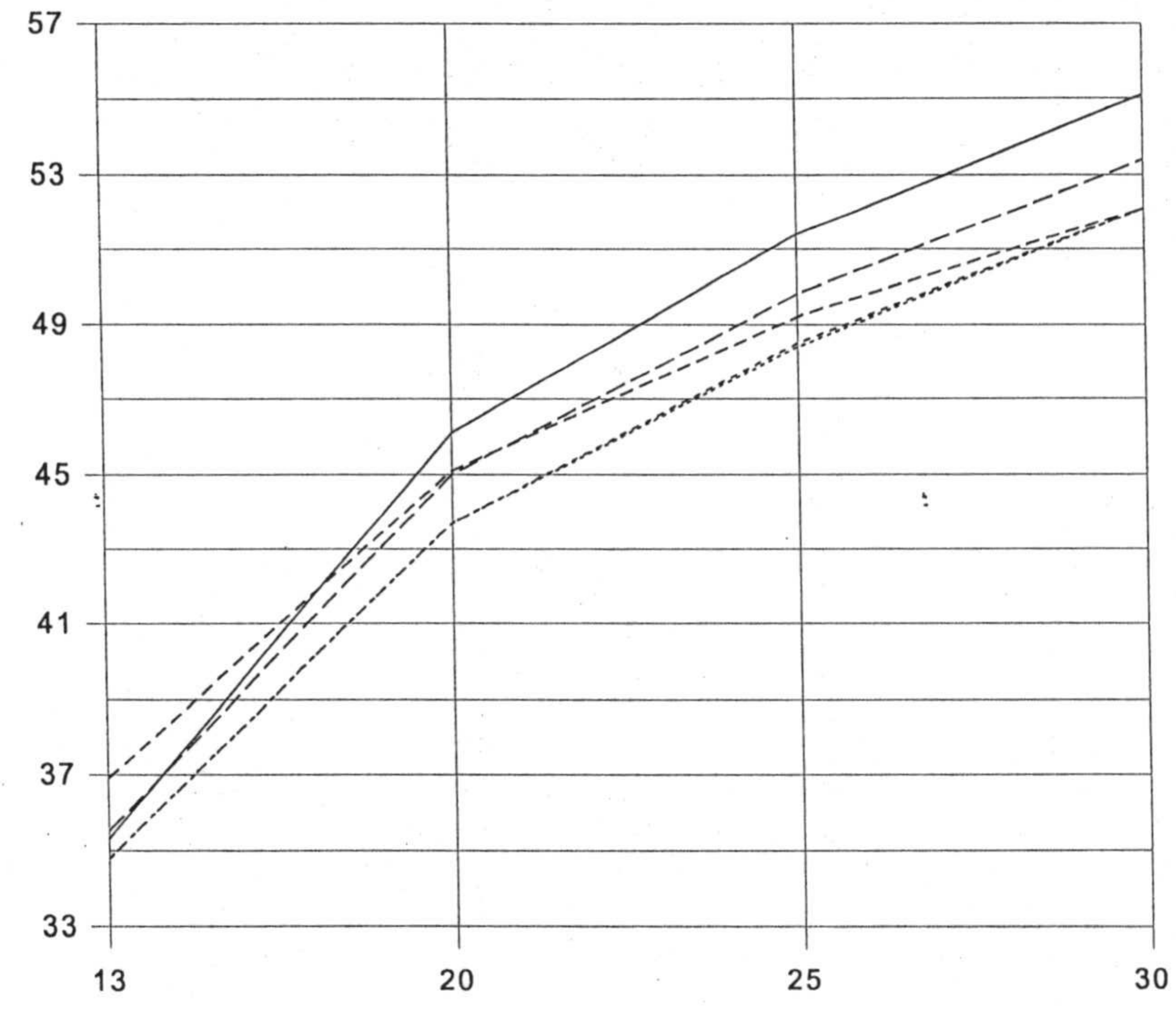


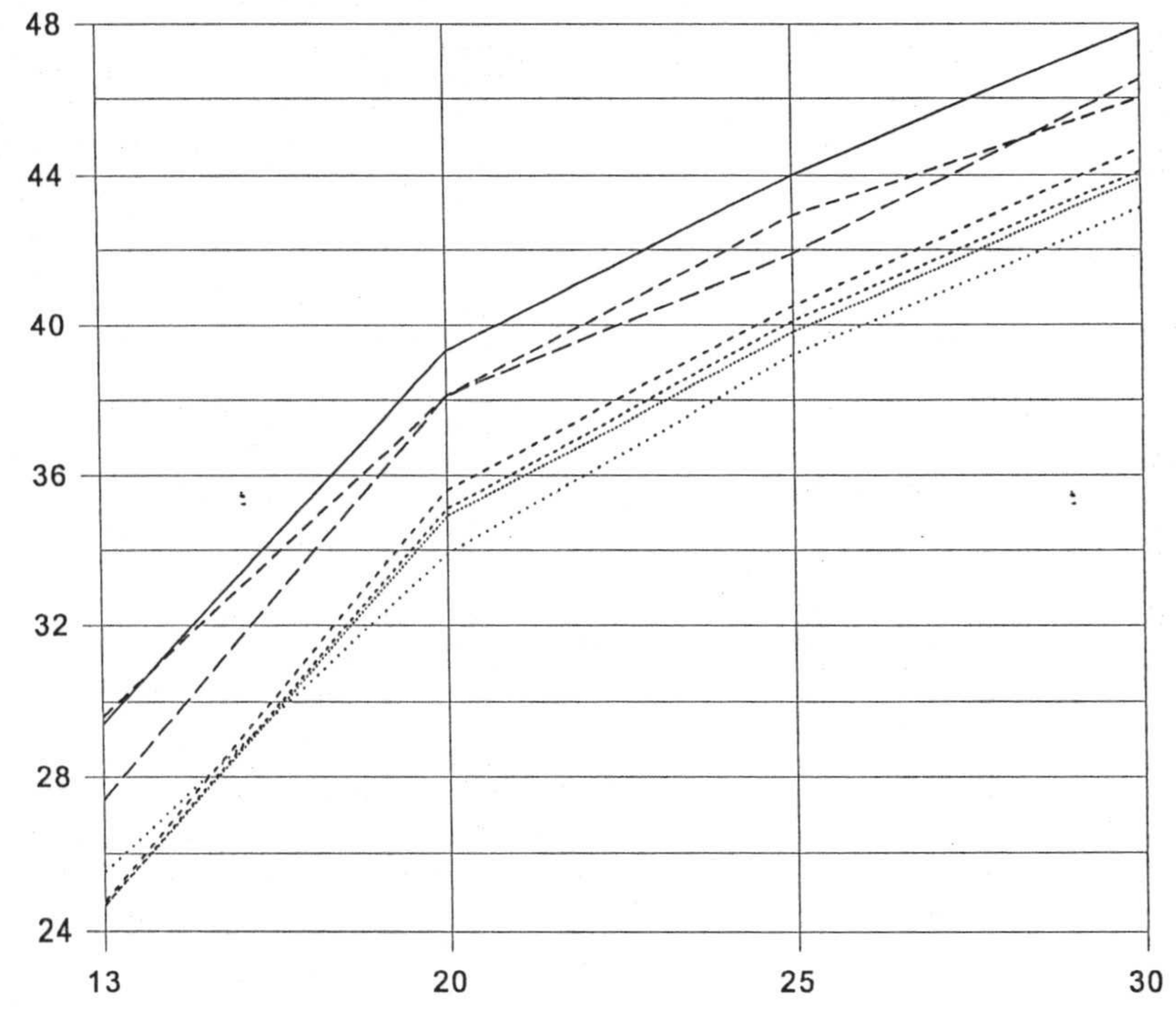
Figura 2.13

(c) 5% - 10%

Una cola



Dos colas



- M
- - FM1
- · M'
- · F1
- · · Y

- M
- - FM2
- · M'
- · FM1
- · · Y
- · · F1
- · · F2

Ante el éxito del método **FM**, y dado que él consiste en disminuir el valor P del método **F** mediante la penalización de $1/2$ en la probabilidad del último punto introducido en la RC, se ha intentado aplicar otras penalizaciones intermedias (fijas, como $1/3$ o $1/4$, o variables) sin éxito apreciable, por lo que en lo que sigue no se les citará más.

2.5.6. Conclusiones

Existen varios métodos para realizar el test incondicionado en una tabla 2×2 con ningún marginal fijado de antemano (test de independencia en tablas 2×2), pero hasta ahora no se conoce cuál de ellos es el más potente. Esta sección ha abordado el problema desde el punto de vista de la potencia media, obteniéndose que el más potente es el método **M**, seguido del método **M'** y de los métodos **FM** (**FM1** para una cola; **FM2** para dos colas) basados en el valor P -mid de Fisher; finalmente, el método **Y** es el mejor de los de tipo chi-cuadrado para los errores α usuales.

Sin embargo, a la hora de seleccionar un método incondicionado no basta con considerar la potencia del mismo, sino que hay que contemplar la magnitud del tiempo de cómputo requerido para realizarlo, pues la resolución de la expresión (2.2) es enormemente costosa. Según nuestra experiencia, para una tabla con $n=20$ y un valor P de aproximadamente el 5%, el método **M** (que ha de calcular α^* numerosas veces) tarda del orden de 2,4 veces más tiempo que el método **M'** (que ha de calcular $\hat{\alpha}^*$ numerosas veces), el cual es a su vez 95 veces más costoso que los demás métodos (que calculan α^* una sola vez), y todos ellos son bastante lentos. Se ve pues que, salvo para valores de n muy pequeños, los métodos **M** y **M'** pueden llegar a ser prohibitivos, y así los métodos más rápidos (aunque algo menos potentes) son buenos competidores de los **M** y **M'**. Para que el lector tenga idea de los tiempos absolutos requeridos, la tabla con $(x_1=2; y_1=4; x_2=17)$, para $n=30$, que tiene un valor P de dos colas cercano al 5%, requiere de un tiempo de cómputo de 1.080 segundos con el método **M**, de 520 segundos con el método **M'**, y de solo 1,5 con los demás métodos.

Así pues el método óptimo (que guarda un buen equilibrio entre su potencia y

el tiempo de cómputo que precisa para ser aplicado) es el **FM1** para test de una cola y el **FM2** para test de dos colas. De hecho, el método **FM1/2** difiere poco del **M'** y puede llegar a superarlo cuando n y α son grandes. Finalmente, de entre los métodos de tipo chi-cuadrado, el método **Y** suele ser el mejor para $\alpha \leq 5\%$, siéndolo **PH2(PH3)** para $\alpha > 5\%$ y tests de una cola (dos colas).

2.6. SOBRE EL CONSERVADURISMO DEL TEST EXACTO DE FISHER

2.6.1. Introducción

En el Capítulo I ya se puso de manifiesto que el test exacto de Fisher, desde el punto de vista incondicionado, es un test válido, pero conservador. La magnitud de su conservadurismo ha sido resaltada históricamente frente al Modelo II (ver referencias de entonces), pero poco o nada se ha dicho para el caso de que se le utilice como sustituto del test de Barnard para el Modelo I, que es el caso actual. El objetivo de este apartado es evaluar la pérdida de potencia que se produce por tal actuación. Pero antes de ello es preciso ver con más detalle el modo en que se produce la acotación indicada en la (1.26) y que tiene como base la (1.25).

2.6.2. Acotación

El espacio muestral completo de una muestra de tamaño n está formado por los puntos de los rectángulos $(n_1; n-n_1)$ con $0 \leq n_1 \leq n$. Cada rectángulo está formado por los puntos (x_1, x_2) de las diagonales de valor a_1 ($x_1 + x_2 = a_1$), con $0 \leq a_1 \leq n$. Cada diagonal viene identificada por tanto por el par (n_1, a_1) y sus puntos son el espacio muestral del test exacto de Fisher. Fijado un error α , el test exacto de Fisher proporcionará una región crítica que llamaremos por $RC(n_1, a_1)$ y que tiene un error real de $\alpha_{III}^*(n_1, a_1) \leq \alpha$. Sea $RC(n_1) = \bigcup_{a_1} RC(n_1, a_1)$ y $\alpha_{III}^*(n_1) = \text{Max}_{a_1} \alpha_{III}^*(n_1, a_1)$. Sea $RC = \bigcup_{n_1} RC(n_1)$ y $\alpha_{III}^* = \text{Max}_{n_1} \alpha_{III}^*(n_1)$. Si RC es una región crítica válida para el Modelo I, su tamaño nominal habrá de ser $\alpha_{III}^* \leq \alpha$. Para verlo, el error $\alpha(p, q)$ de ella será, por la (1.25),

$$\begin{aligned}
\alpha(p, q) &= \sum_{RC} P(x_1, y_1, x_2) = \sum_{RC} P(n_1) P(a_1) P(x_1 | n_1, a_1) = \\
&= \sum_{n_1 \in RC} P(n_1) \sum_{a_1 \in RC(n_1)} P(a_1) \sum_{x_1 \in RC(n_1, a_1)} P(x_1 | n_1, a_1) = \\
&= \sum_{n_1 \in RC} P(n_1) \sum_{a_1 \in RC(n_1)} P(a_1) \alpha_{III}^*(n_1, a_1) \leq \sum_{n_1 \in RC} P(n_1) \alpha_{III}^*(n_1) \sum_{a_1 \in RC(n_1)} P(a_1) \\
&\leq \sum_{n_1 \in RC} P(n_1) \alpha_{III}^*(n_1) \leq \alpha_{III}^* \sum_{n_1 \in RC} P(n_1) \leq \alpha_{III}^* \leq \alpha
\end{aligned} \tag{2.78}$$

con lo cual el tamaño α_I^* de tal RC bajo el Modelo I será:

$$\alpha_I^* \leq \alpha_{III}^* \leq \alpha \tag{2.79}$$

lo que implica que es una región crítica válida (pues $\alpha_I^* \leq \alpha$) y que el test exacto de Fisher es conservador (pues $\alpha_I^* \leq \alpha_{III}^*$).

2.6.3. Sobre la coherencia de la región crítica condicionada*

Una cosa es que la RC así construida respete el error α (con lo que es válida) y otra muy distinta es que ella sea coherente, es decir que respete las propiedades de equivalencia, simetría y convexidad definidas para el Modelo I. En 2.2.5.3 se probó que las ordenaciones F1 y F2 verifican las propiedades de equivalencia y de simetría, pero que sólo F1 verificaba la propiedad de convexidad. En consecuencia, el test exacto de Fisher, que consiste en aplicar F1/F2 con valores nominales, es coherente sólo cuando se aplica como test de una cola. En lo que sigue, FC aludirá al test exacto de Fisher (F1 o F2) aplicado con valores nominales (no como ordenación). Sin embargo FC a dos colas sí verifica una mínima propiedad de coherencia que podemos llamar de *Diagonalidad*: Si el punto $(x_1 | n_1, a_1)$ es de la RC, también lo es el punto $(x_1 - 1 | n_1, a_1)$, pues en la hipergeométrica $P(x_1 - 1 | n_1, a_1) \leq P(x_1 | n_1, a_1)$.

2.6.4. Conservadurismo del condicionamiento y test UMP*

Con independencia de lo anterior, hay un asunto acerca del condicionamiento que no ha sido bien matizado por la literatura. Las (2.78) y (2.79) han de entenderse como sigue: Si para formar una RC para el Modelo I se utiliza el test exacto de Fisher

en sus valores nominales (el **FC**), ello da una $RC(FC)$ válida pues su tamaño $\alpha_1^* \leq \alpha$. De otro lado, si el mismo test es utilizado como método de ordenación para el Modelo I (método **F**), es seguro que su $RC(F)$ contiene a la $RC(FC)$ anterior, pues ésta es válida, y probablemente a algún punto más: $RC(FC) \subset RC(F)$. De ello se concluye que el método **F** es uniformemente más potente (UMP) que el método **FC** (para ello es preciso que F_2 se aplique tal cual, sin imponerle la convexidad). En consecuencia puede afirmarse que el test incondicionado **F** es UMP respecto del condicionado **FC**. Sin embargo, si se usa otro criterio de ordenación (como el **M** óptimo), las cosas ya no son iguales, pues el orden de entrada de puntos que induce **M** no es el que induce **F** y así $RC(FC)$ y $RC(M)$ no están sometidas a una relación de inclusión. Es por ello que para estos casos sólo puede afirmarse que **M** es "en general" UMP respecto de **FC** (pues **M** es mejor que **F** y éste es mejor que **FC**). Dado que **M** es el óptimo, en lo que sigue se evaluará la pérdida de potencia producida por usar **FC** (el condicionado) en lugar de **M**.

Arriba se ha hablado de UMP, no de UMPU, pues Suissa and Shuster (1984) ya probaron que un test incondicionado puede ser sesgado.

2.6.5. Criterios para la comparación*

Ya se ha dicho que las comparaciones de la potencia $\Theta(p_{11}, p_{12}, p_{21})$ alcanzada por dos métodos son complejas de evaluar por su dependencia de las p_{ij} , y que la potencia a largo plazo $\Theta(\alpha)$ o la potencia media $\bar{\Theta}(\alpha_1; \alpha_s)$ constituían mejores y más globales procedimientos. A fin de poder dar un consejo específico a un investigador acerca de cuánta potencia pierde por usar el método **FC**, en lugar del óptimo **M**, cuando tiene en mente un error específico α (por ejemplo, el 5% habitual), en lo que sigue se van a considerar las potencias a largo plazo $\Theta(FC)$ y $\Theta(M)$ obtenidas para las regiones críticas $RC(FC)$ y $RC(M)$ que proporcionan cada procedimiento. La pérdida de potencia será por tanto:

$$\Delta\Theta(M, FC) = \Theta(M) - \Theta(FC) \quad (2.80)$$

la cual será evaluada para:

<i>Nº colas:</i>	1	2			
<i>Error α:</i>	1%	5%	10%		
<i>Tamaño n:</i>	10	20	30	40	50

(2.81)

a fin de estudiar su evolución con el tamaño de muestra.

El valor de $\Theta(M)$ es el indicado por las (2.62) y (2.67), cuyas RC aluden a $RC(M)$. El valor $\Delta\Theta$ de la (2.80) puede obtenerse simplificadaamente así: para dos colas es

$$\Delta\Theta = 6 \frac{\text{Card } RC(M) - \text{Card } RC(FC)}{(n+1)(n+2)(n+3)} \quad (2.82)$$

en tanto que para una cola será:

$$\Delta\Theta = \frac{12}{(n+1)(n+2)(n+3)} \left\{ \sum_{RC(M)-RC(FC)} P_{\bar{F}}(x_1+1, y_1, x_2, y_2+1) - \sum_{RC(FC)-RC(M)} P_{\bar{F}}(x_1+1, y_1, x_2, y_2+1) \right\} \quad (2.83)$$

con $P_{\bar{F}}$ como en la (2.66).

2.6.6. Resultados y Conclusiones*

La Tabla 2.9 presenta los valores de $\Theta(M)$ e $\Delta\Theta(M, FC)$, en tanto que la Figura 2.14 (c) los representa gráficamente para su mejor comprensión. Puede observarse que los valores del incremento de potencia:

- 1) Son más marcados a una que a dos colas, conservándose las diferencias con el aumento de n .
- 2) Son más elevados cuanto mayor es el error α , suavizándose los mismos con el aumento de n .
- 3) Son más pequeños cuanto mayor es n , alcanzándose el mínimo para dos colas

Tabla 2.9

Potencia a largo plazo del método M (1ª entrada) e incremento absoluto de potencia del mismo respecto de los métodos M' (2ª entrada), FM (3ª entrada) y FC (4ª entrada) para cada combinación del número de colas, error α y tamaño de muestra n.

(a) UNA COLA

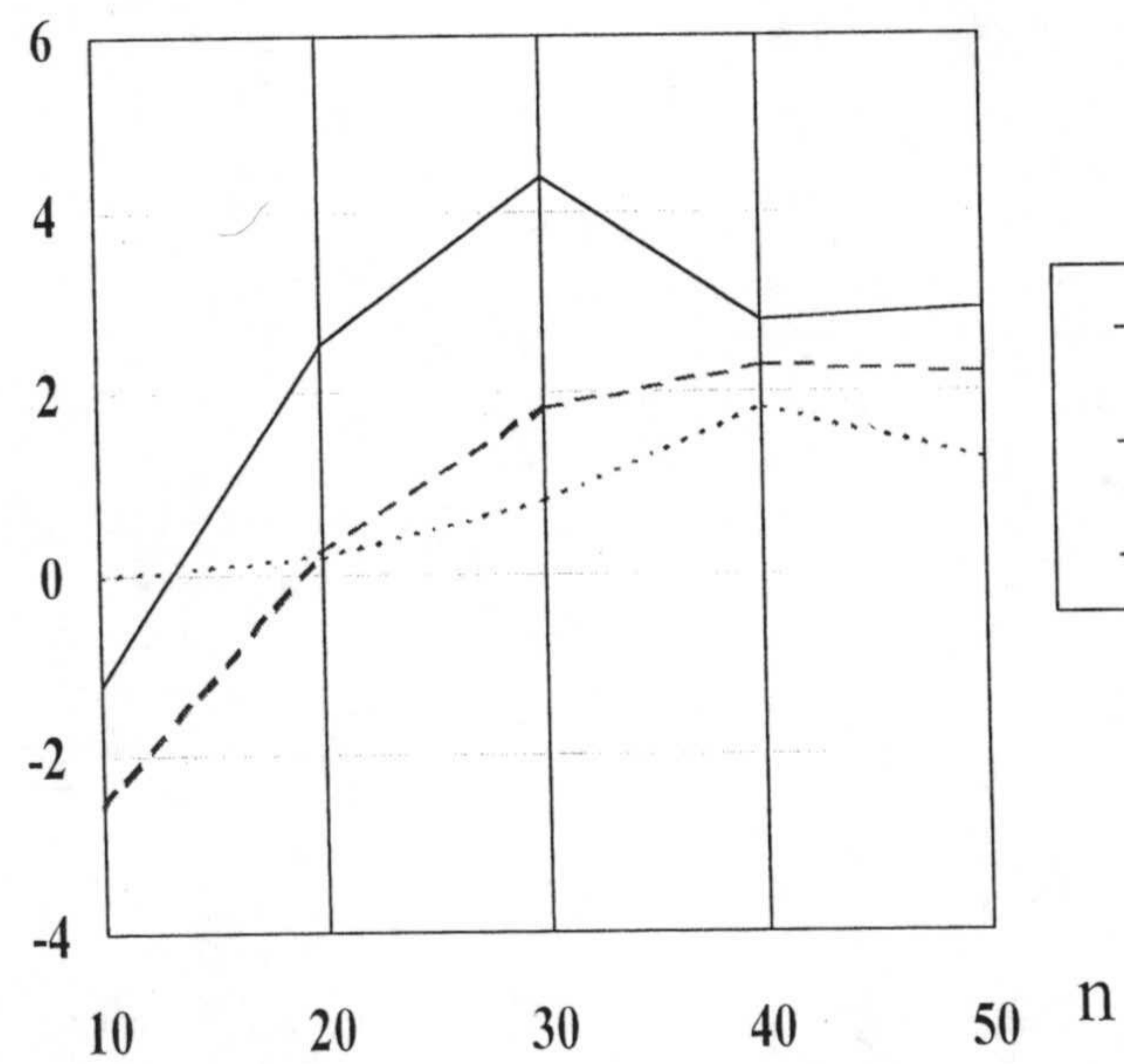
$\alpha \backslash n$	10	20	30	40	50
1%	10,43	23,97	33,76	40,49	45,32
	0	0,22	0,8	1,84	1,26
	0	1,33	1,23	1,06	0,98
	6,94	8,4	7,75	6,8	6,4
5%	21,9	39,21	49,16	55,17	59,14
	-2,61	0,24	1,84	2,32	2,22
	-1,4	1,04	1,17	1,28	1,34
	11,47	11	9,62	8,37	7,24
10%	33,62	50,67	58,21	63,33	66,66
	-1,24	2,52	4,4	2,78	2,91
	1,2	2,4	1,81	1,13	0,71
	14,99	14,23	10,7	8,98	7,53

(b) DOS COLAS

$\alpha \backslash n$	10	20	30	40	50
1%	7,69	20,67	30,32	36,71	41,79
	0	0,68	0,99	1,66	1,6
	0	1,13	2,39	2,26	1,92
	4,2	5,99	5,46	4,52	4,06
5%	16,8	34,9	43,73	49,59	53,87
	-2,08	1,36	1,83	1,52	2,47
	0,02	2,49	2,27	2,61	1,18
	5,59	8,81	6,89	5,79	4,45
10%	23,78	43,59	51,61	56,75	60,62
	-5,59	1,81	2,12	1,78	2,56
	-2,79	1,58	1,76	2,01	1,25
	7,69	10,05	8,1	5,95	4,71

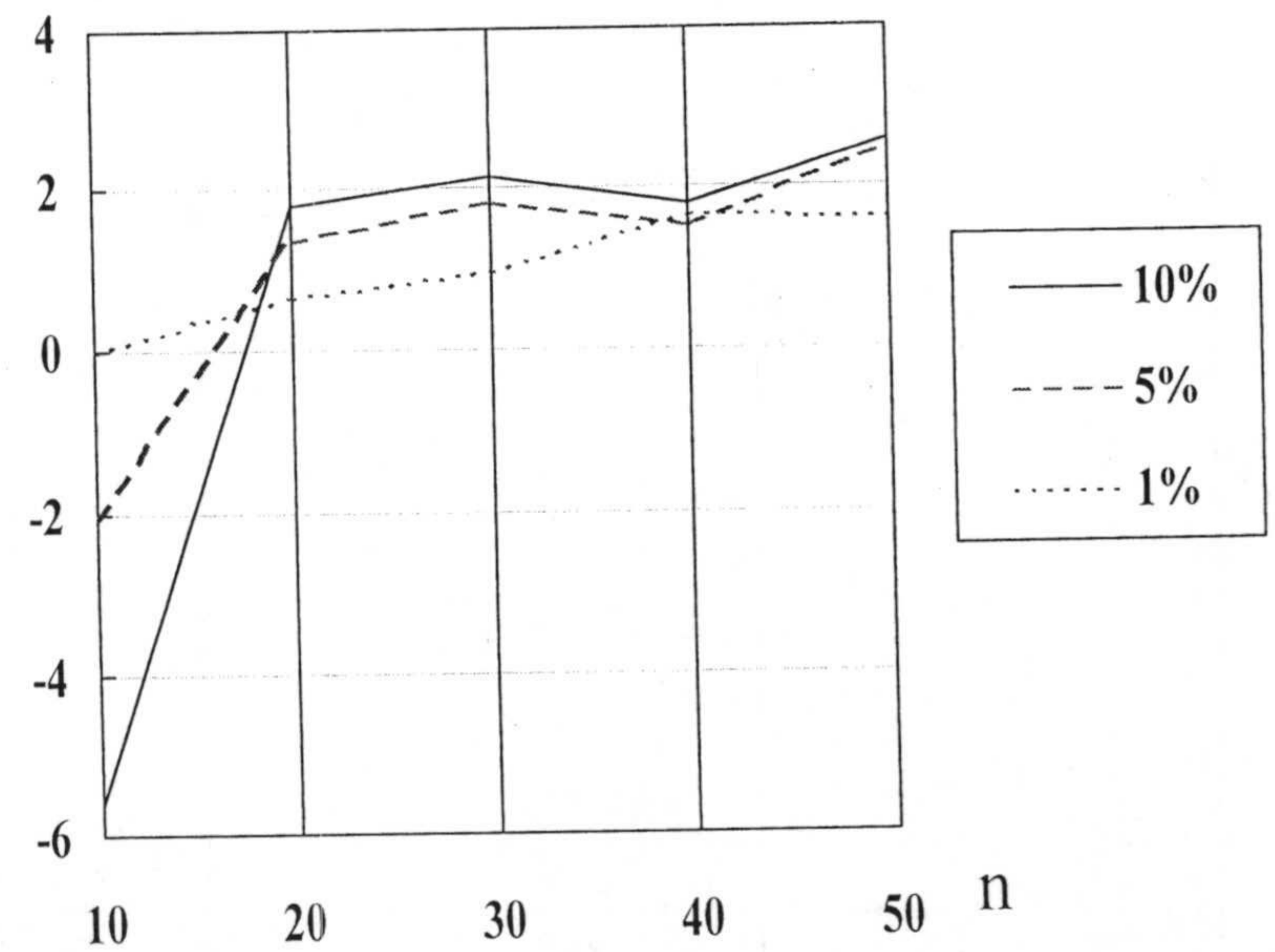
Figura 2.14 (a)
 Incremento absoluto de potencia a largo plazo por usar el método óptimo M en lugar del método semi-óptimo M', en función del tamaño de muestra n y para cada valor del error α
 (su orden de escritura es el de aparición de su curva por la izquierda).

Δ Potencia x 100



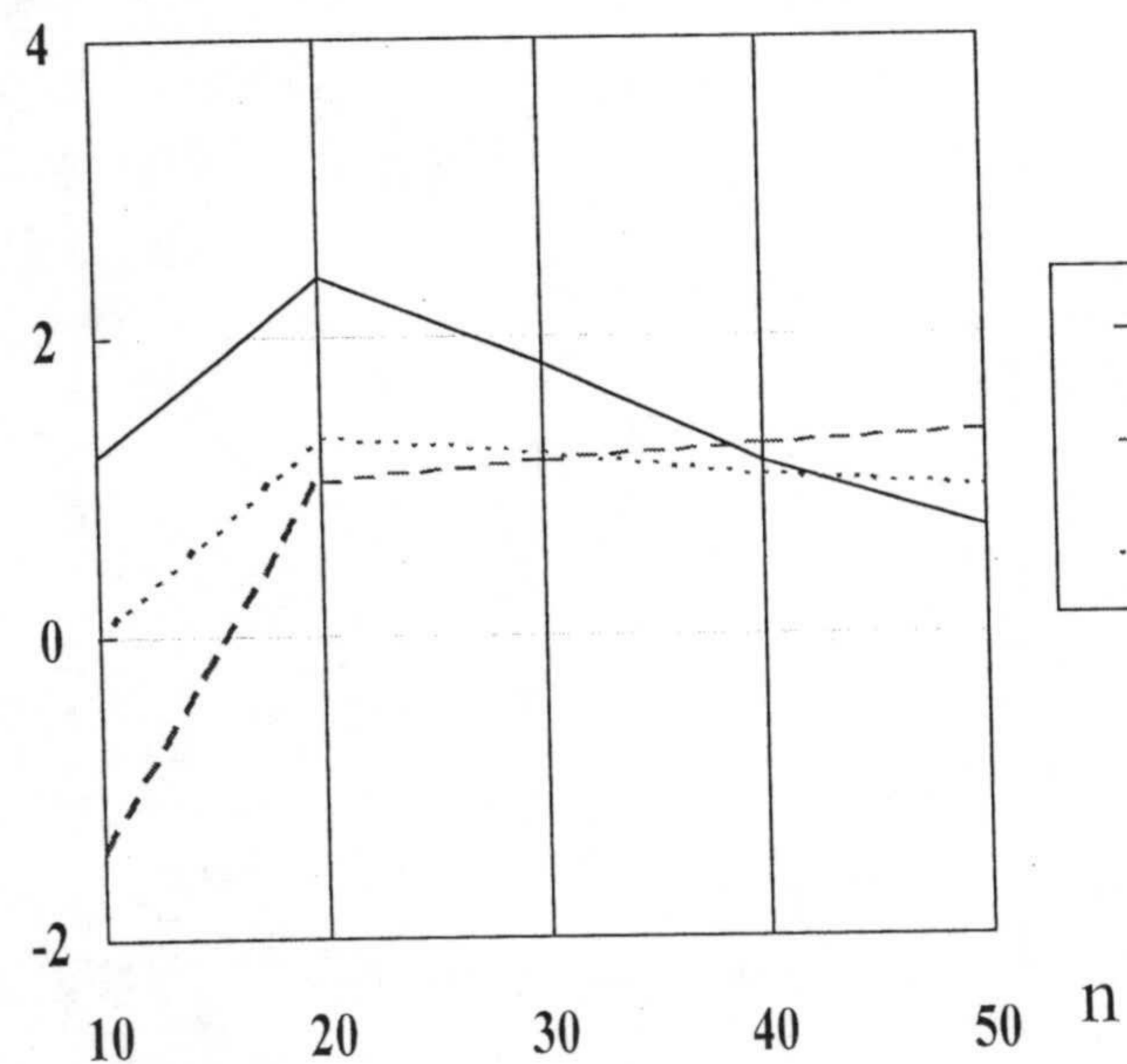
UNA COLA

Δ Potencia x 100

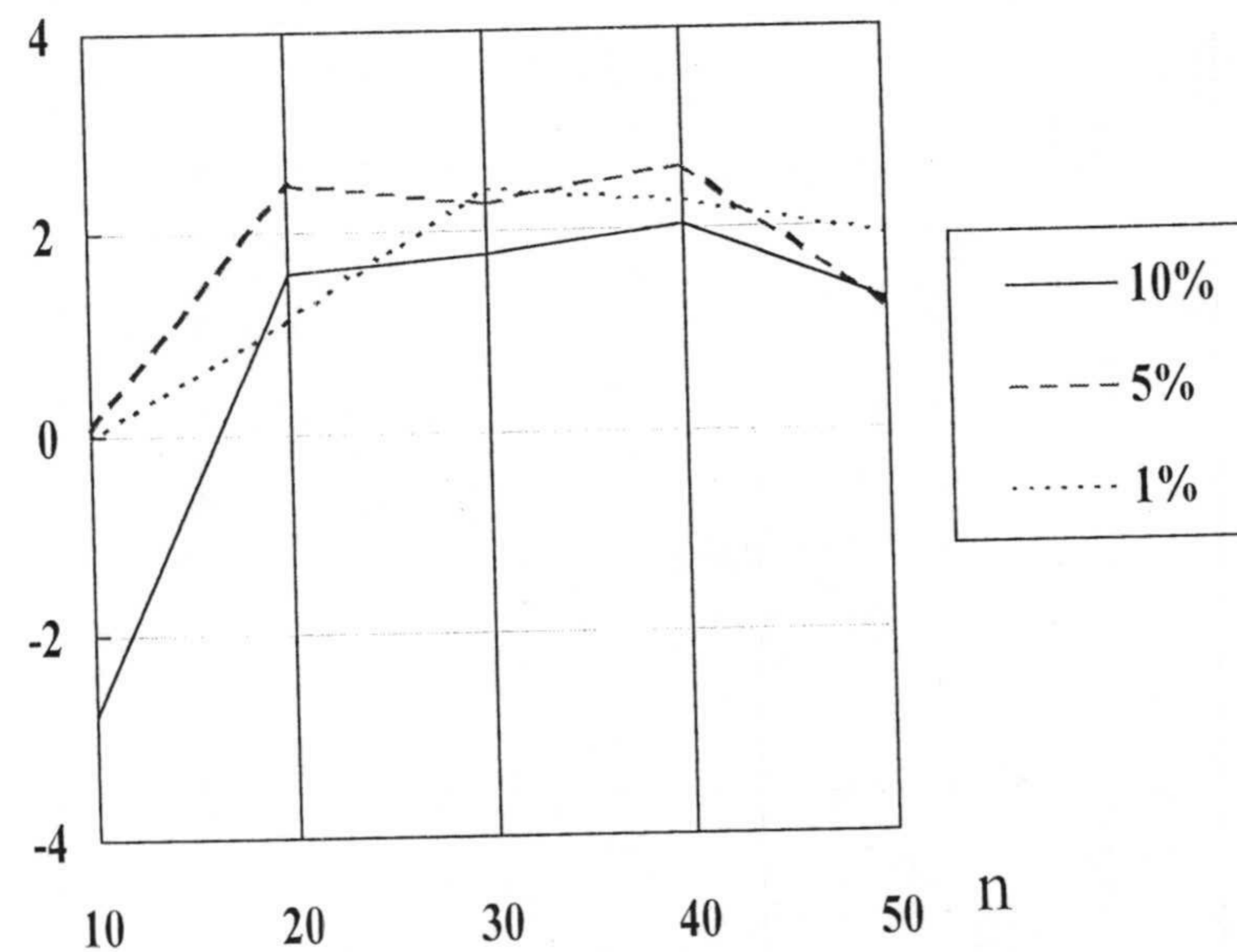


DOS COLAS

Figura 2.14 (b)
Incremento absoluto de potencia a largo plazo por usar el método óptimo M en lugar del método semió-ptimo FM, en función del tamaño de muestra n y para cada valor del error α (su orden de escritura es el de aparición de su curva por la izquierda).

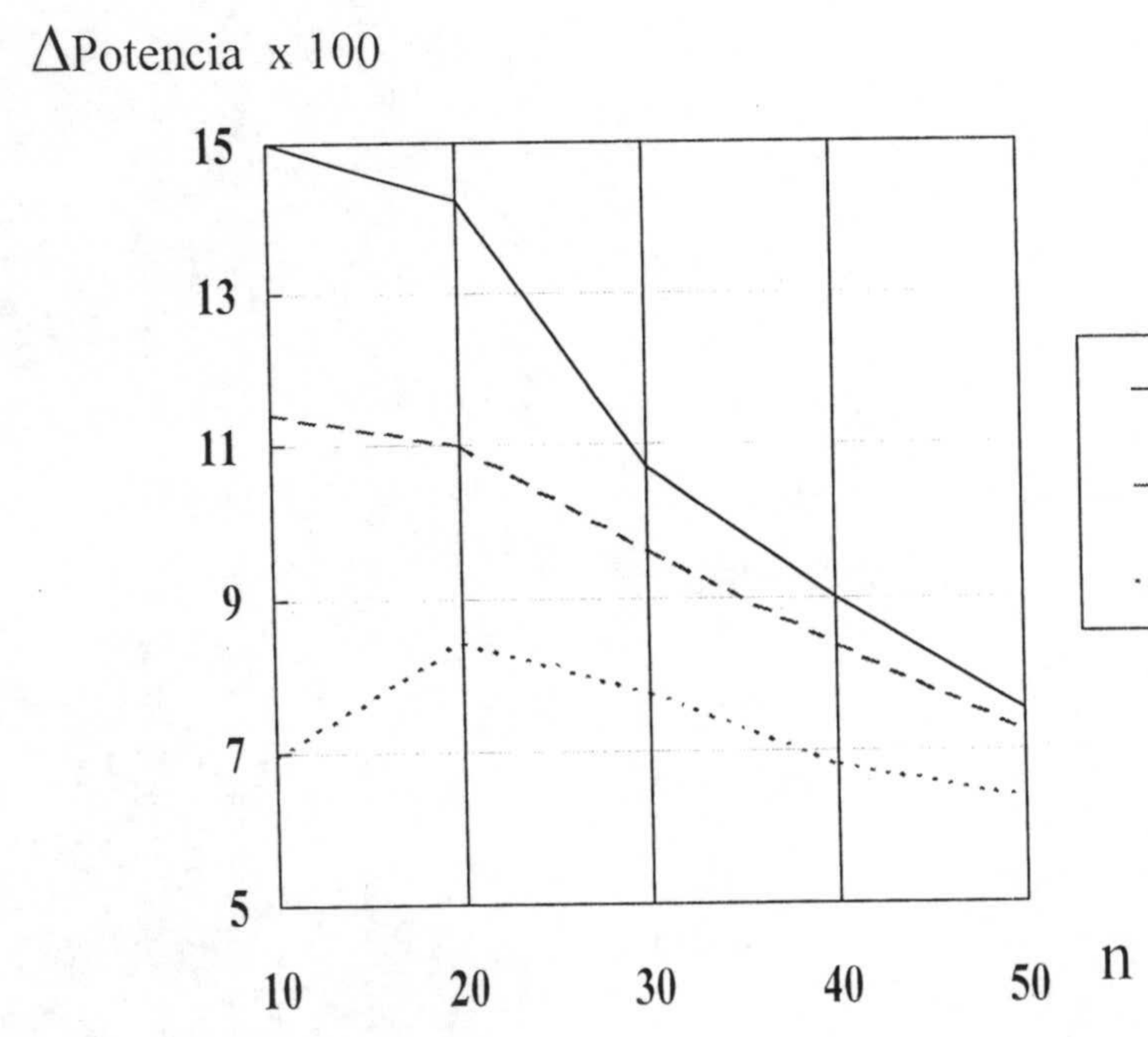
 Δ Potencia x 100

UNA COLA

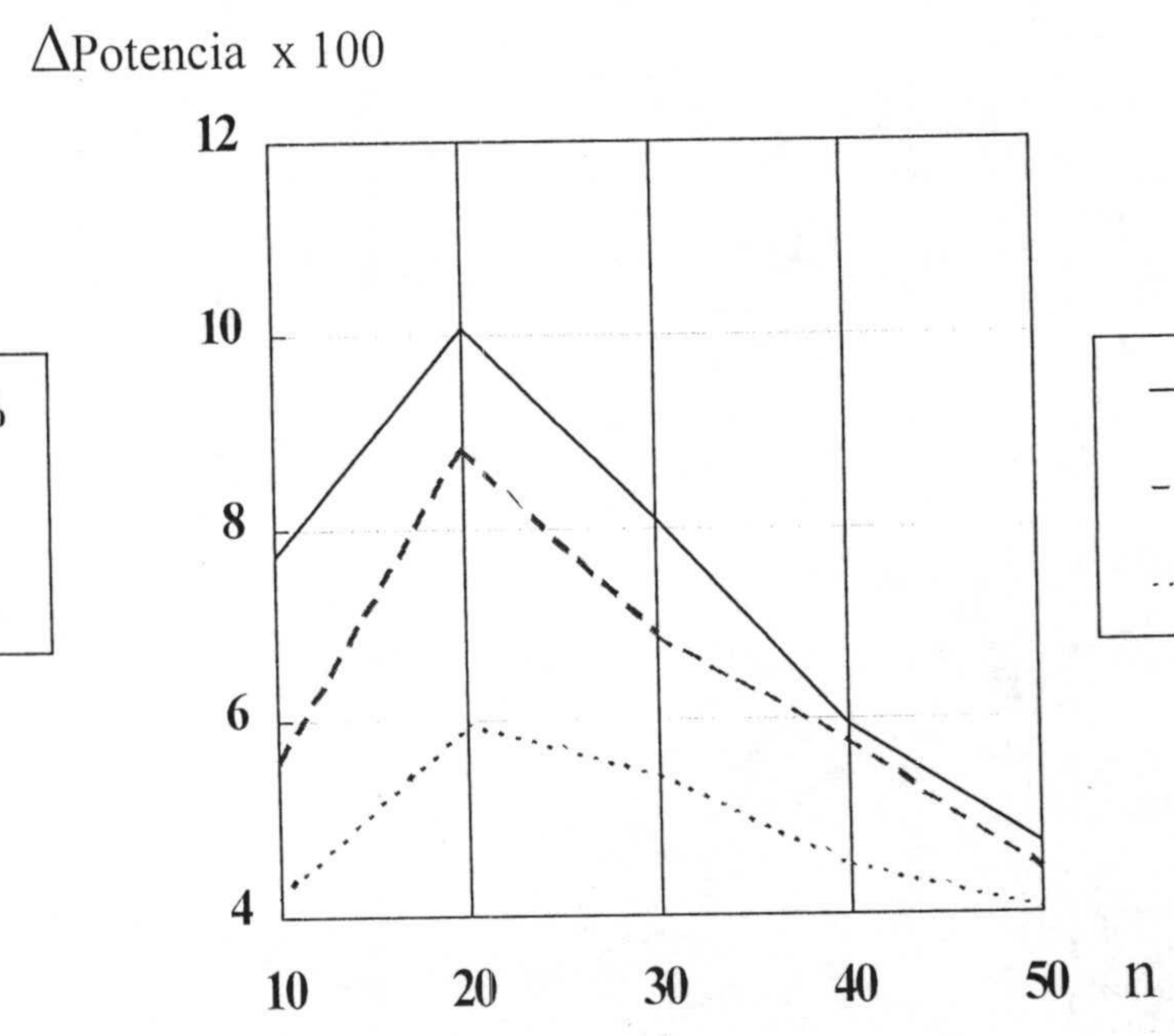
 Δ Potencia x 100

DOS COLAS

Figura 2.14 (c)
 Incremento absoluto de potencia a largo plazo por usar el método óptimo M en lugar del método condicionado FC, en función del tamaño de muestra n y para cada valor del error α (su orden de escritura es el de aparición de su curva por la izquierda).



UNA COLA



DOS COLAS

y $\alpha = 1\%$.

4) En todos los casos $\Delta\Theta > 0$, por lo que FC es siempre menos potente que M. de donde se deduce que el test incondicionado es siempre más ventajoso (en potencia) que el condicionado, tanto más cuanto menor es el número de colas, mayor el error α o menor el tamaño n de muestra.

Por otro lado se observa que para $n=50$ y $\alpha=5\%$, el $\Delta\Theta$ es el del orden del 7,2% (una cola) o 4,5% (dos colas) en términos absolutos y de un 12,2% (una cola) o 8,3% (dos colas) en términos relativos, lo que evidentemente son aún cantidades importantes y muestran el gran conservadurismo del test condicionado en los pequeños tamaños de muestra. Si se entiende que un valor $\Delta\Theta=1\%$ es ya poco relevante, una extrapolación de los resultados (que es lo único factible con la capacidad de cómputo actual) permite deducir que desde $n=110$ (una cola) o $n=80$ (dos colas) ambas metodologías difieren poco. De ahí la importancia de contar con tablas y/ó programas del test incondicionado para valores de n inferiores a los señalados (todo ello aparte de la inconsistencia, ya reseñada, del test de Fisher).

2.7. SOBRE EL CONSERVADURISMO DE LOS TESTS SEMI-ÓPTIMOS*

2.7.1. Introducción

En la sección 2.5 se concluyó que el método más potente era el **M**, seguido del **M'** y, un poco más distante, del **FM1** (una cola) o **FM2** (dos colas), a lo que se llamará abreviadamente por método **FM** en lo que sigue. Sin embargo **M'** y **FM** son métodos bastante más rápidos que **M**, y así el objetivo actual es estudiar a partir de qué valor de n ellos pueden sustituir al **M** con una pérdida mínima de potencia. A tal efecto se va a seguir igual metodología y argumentación que en la sección anterior, y así las fórmulas (2.80) a (2.83) de entonces siguen siendo válidas ahora si se cambia FC por **M'** o por **FM** (según el caso).

2.7.2. Resultados y Conclusiones

Los datos de $\Delta\theta(M, M')$ y $\Delta\theta(M, FM)$ se incluyen también en la Tabla 2.9, en tanto que las Figuras 2.14 (a) y (b) efectúan la representación gráfica de los mismos. Puede observarse que:

- 1) Para M' , la pérdida de potencia $\Delta\theta$ tiende a estabilizarse en los alrededores del 2% para $n \geq 40$ (algo mayor que el 2% en los α grandes; algo menor en los α pequeños).
- 2) Para FM la pérdida de potencia $\Delta\theta$ tiende a estabilizarse en los alrededores del 1% para $n \geq 40$ (una cola) o $n \geq 50$ (dos colas y $\alpha \geq 5\%$) o en el 2% para $n \geq 30$ (dos colas y $\alpha = 1\%$).

Esto quiere decir que, en contra de lo afirmado en la sección 2.5, el método FM es más potente que el M' cuando el tamaño de muestra es moderadamente grande ($n \geq 40$ o 50). En realidad la contradicción no es tal pues ahora se están contemplando valores de n más elevados que entonces, y es justo en ellos cuando empieza a cambiar la situación.

Dado que FM es un método más rápido que el M' , la conclusión final es que si se admite una pérdida de potencia $\Delta\theta$ de alrededor de un 1% entonces el método óptimo es el FM para $n \geq 40$ (una cola) o $n \geq 50$ (dos colas).

Alternativamente, los $\Delta\theta$ de la Tabla 2.9 también indican que el método M' se acerca tanto o más que el FM (y por tanto que el resto) al M cuando $n \leq 30, 20, 10$ (1 cola) o $n \leq 50, 40, 10$ (dos colas) en los $\alpha = 1\%, 5\%, 10\%$ respectivamente. Como en el primer caso M es practicable, se deduce que M' sólo será conveniente en las dos primeras circunstancias del segundo caso. Sin embargo M' es más potente que M para $n \leq 10$ (en todo caso).

2.8. PROGRAMAS*

En 2.1.7 ya se señaló la escasez de programas fiables para el Modelo I. Pero además cuentan con serios defectos:

- 1) El programa de Shuster (1992) sólo es válido para el método Y que ya se ha

visto no es del grupo de los mejores (aunque sí lo es de entre los tipo chi-cuadrado) ni en potencia ni en el equilibrio potencia/tiempo de cálculo. Por otro lado, como test de dos colas no verifica las propiedades de convexidad (lo que es algo no deseable).

- 2) El programa de Berger (1996) sólo es válido para los métodos **Z** y **F1**. El primero es el peor de todos; el segundo es de calidad intermedia (es similar al **Y**).
- 3) El programa de Silva and Martín (1997) ya se vio que se proporcionaba como una solución aproximada a la que se llamó "**Barnard1**". Aparte no hay garantías de que el método verifique convexidad, ni tan siquiera de que sea válido (de que respete el error α).

De todo ello se desprende la conveniencia de proporcionar un programa que contemple a todos los métodos o, alternativamente, a los más potentes, que puede encontrarse en la dirección electrónica <http://www.ugr.es/~bioest/TMP.EXE>.

Todos estos programas permiten obtener el valor P de una tabla experimental, lo que implica que, salvo en el caso del método **M**, el máximo α^* sólo habrá que calcularse una vez y al final: cuando la tabla experimental entre en la RC. De ahí que, para tales programas (salvo, de nuevo, el de **M**), las Propiedades **III** a la **V** no sean de utilidad. No debe olvidarse que la utilización de los mismos ha de ser prudente: los **M** y **M'** son desesperantes para $n > 30$; los demás para $n > 60$ (todo ello de modo aproximado).

De otro lado, el lector puede estar interesado en el programa que genera las tablas de potencia $\Theta(\alpha)$ y $\bar{\Theta}(\alpha_1, \alpha_s)$; el mismo puede solicitarse a los autores.

2.9. TABLAS*

2.9.1. Introducción

En 2.1.7 se citó que las únicas tablas existentes son las tablas de valores críticos para el método **Y** de Shuster (1992), lo que tiene la misma crítica que la realizada para

su programa.

La obtención de unas tablas de RC para los métodos de ordenación óptimos del Modelo I (M y M' , que lo era para valores pequeños de n) presentan la dificultad (aparte del tiempo de cómputo preciso para su construcción) de tener que especificar uno a uno los valores de las fronteras de su RC (los puntos * frontera de la Figura 1.1) lo que implica que precisan de un gran espacio. Como ello es común a M y M' , lo aconsejable es obtener una tabla para M (que es el óptimo).

Las tablas de RC para los demás métodos son más rápidas de obtener y requieren de menos espacio para expresarlas, pues, por depender de un estadístico de orden, basta con especificar el valor crítico de éste para cada α seleccionado y para cada tipo de test (de una o dos colas). Así, puesto que **FM1** y **FM2** eran las ordenaciones óptimas, ellos son los métodos aconsejables a incluir en unas tablas (en lugar del **Y** de Shuster). Desgraciadamente el método **FM2** no verifica las condiciones de convexidad, por lo que la solución no es factible.

2.9.2. Tablas de RC por el método de Barnard (M)

Las tablas se han obtenido para valores de n entre 3 y 50; valores de α del 1%, 5% o 10%; y tests de una o de dos colas. A fin de disminuir su tamaño aproximadamente a una cuarta parte (en los tests de una cola) o a una octava parte (en los tests de dos colas), los datos se generan y obtienen en el formato canónico utilizado en esta memoria, lo que implica que una tabla experimental debe reformatearse para que verifique que:

$$n_1 = \text{Min}(n_1, n_2, a_1, a_2) \quad ,, \quad x_1 y_2 / x_2 y_1 < 1 \quad (2.84)$$

ó, equivalentemente,:

$$n_1 = \text{Min}(n_1, n_2, a_1, a_2) \quad ,, \quad \hat{p}_1 = \frac{x_1}{n_1} < \hat{p}_2 = \frac{x_2}{n_2} \quad (2.84 \text{ Bis})$$

Para un valor dado de n , α y número de colas, la RC obtenida contará con algún punto en los diversos valores de n_1 (no tiene porqué haberlo en todos), como por ejemplo el caso de la Figura 1.1 (aunque en ella es $n_1 > n_2$). Para explicitar tal RC (que es de dos colas) basta con dar las fronteras de una de las colas (en nuestro caso

la de arriba, que es la que da $\hat{\theta} < 1$) pues la otra se obtiene por simetría. Las fronteras se indicarán señalando cada uno de los valores de x_1 en los que haya puntos de la RC y el valor de $x_2 = C$ a partir del cual están los puntos de la RC ($x_2 \geq C$). Para el caso de la Figura 1.1 los datos serían:

$\frac{n_1}{}$	$\frac{x_1}{}$	$\frac{x_2}{}$
8	0	2
	1	3
	2	5
	3	5
	4	6

con la última columna aludiendo a los valores C señalados antes. La constante C es pues un valor que depende de los valores n , n_1 y x_1 de la tabla experimental $-C(n, n_1, x_1)-$ y la significación se alcanza si ocurre que en nuestra tabla experimental es

$$x_2 \geq C(n, n_1, x_1) \quad (2.85)$$

Por el modo de generar los datos, en la tabla buscada sólo aparecerán valores de n_1 y x_1 compatibles con la (2.21). Si algún n_1 o x_1 no aparece, es porque para él no hay puntos de la RC. Igual deben entenderse los casos en que como valor de x_2 aparece un "-": ellos han sido incluidos porque para otro error α u otro n° de colas del test sí hay puntos en la RC.

Las tablas de RC para el método **M** óptimo son las del **ANEXO** (construidas sin introducir simultáneamente a todos los atractores). Se observará que, a pesar de llegar sólo hasta el valor $n = 50$, han sido precisas 8 hojas para especificarlas (y eso a pesar de todas las precauciones tomadas para llevarlas al mínimo de su tamaño). Los dos ejemplos que siguen sirven, por un lado, para ilustrar su uso y, por otro, para comprobar que con ellas se obtienen resultados más significativos que con las de Shuster (1992).

Ejemplo de test de 1 cola:

Los datos experimentales son los de la Tabla 2.10(a), para los que se

desea contrastar la alternativa de asociación negativa ($H_1 \equiv \theta < 1$). Como en ella $\hat{\theta} = 1 \times 0 / 1 \times 11 < 1$, es posible que haya significación a los errores α ordinarios (de haber sucedido $\hat{\theta} \geq 1$, la conclusión sería directamente por H_0).

Tabla 2.10

(a)	(b)	(c)																											
<table style="width: 100%; border-collapse: collapse;"> <tr><td style="border-right: 1px solid black; padding: 2px 10px;">1</td><td style="padding: 2px 10px;">11</td><td style="padding: 2px 10px;">12</td></tr> <tr><td style="border-right: 1px solid black; padding: 2px 10px;">1</td><td style="padding: 2px 10px;">0</td><td style="padding: 2px 10px;">1</td></tr> <tr style="border-top: 1px solid black;"><td style="border-right: 1px solid black; padding: 2px 10px;">2</td><td style="padding: 2px 10px;">11</td><td style="padding: 2px 10px;">13</td></tr> </table>	1	11	12	1	0	1	2	11	13	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="border-right: 1px solid black; padding: 2px 10px;">1</td><td style="padding: 2px 10px;">0</td><td style="padding: 2px 10px;">1</td></tr> <tr><td style="border-right: 1px solid black; padding: 2px 10px;">1</td><td style="padding: 2px 10px;">11</td><td style="padding: 2px 10px;">12</td></tr> <tr style="border-top: 1px solid black;"><td style="border-right: 1px solid black; padding: 2px 10px;">2</td><td style="padding: 2px 10px;">11</td><td style="padding: 2px 10px;">13</td></tr> </table>	1	0	1	1	11	12	2	11	13	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="border-right: 1px solid black; padding: 2px 10px;">0</td><td style="padding: 2px 10px;">1</td><td style="padding: 2px 10px;">1</td></tr> <tr><td style="border-right: 1px solid black; padding: 2px 10px;">11</td><td style="padding: 2px 10px;">1</td><td style="padding: 2px 10px;">12</td></tr> <tr style="border-top: 1px solid black;"><td style="border-right: 1px solid black; padding: 2px 10px;">11</td><td style="padding: 2px 10px;">2</td><td style="padding: 2px 10px;">13</td></tr> </table>	0	1	1	11	1	12	11	2	13
1	11	12																											
1	0	1																											
2	11	13																											
1	0	1																											
1	11	12																											
2	11	13																											
0	1	1																											
11	1	12																											
11	2	13																											

Como n_1 ha de ser el más pequeño marginal (es decir 1), se permutan las dos filas de la tabla para obtener la Tabla 2.10(b). Como en ella $\hat{\theta} = 1 \times 11 / 1 \times 0 > 1$, y el ANEXO está preparado para tablas con $\hat{\theta} < 1$, hay que permutar sus dos columnas para obtener así la Tabla 2.10(c) que es la que finalmente hay que localizar en el ANEXO. Los valores para ella son:

$$n = 13 \text{ ,, } n_1 = 1 \text{ ,, } x_1 = 0 \text{ ,, } x_2 = 11$$

La terna $(n, n_1, x_1) = (13, 1, 0)$ se localiza en el ANEXO y en su línea, columna de "UNA COLA", aparecen los valores límites $C(n, n_1, x_1) = 10, 11$ y - para los errores del 10%, 5% y 1% respectivamente. Como el valor experimental $x_2 = 11$ es mayor o igual que los dos primeros, pero no que el tercero (para el que no existe límite), la tabla es significativa para errores α del 10% y del 5%, pero no para el 1%, con lo que su valor P es $1\% < P \leq 5\%$ (el igual se debe a que $x_2 = 11$ coincide con la frontera al 5%), probablemente cercano al 5% (el valor real es del 4,84%).

Sin embargo, la misma tabla evaluada por las tablas de Shuster (1992) da un $Z_Y = 0,999$ con un valor P de $5,37\% < P < 11,47\%$ superior al obtenido por las tablas del ANEXO (el valor exacto es 5,41%). Puede observarse aquí la ventaja de no condicionar. El valor $Z_Y = 0,999$ clásico del método condicionado da un valor P (en el supuesto de que fuera válido) $> 10\%$ (muy superior a los anteriores).

Tabla 2.11

2	3	5
0= x_1	7= x_2	7
2= n_1	10	12= n

Ejemplo de test de 2 colas:

Los datos experimentales son los de la Tabla 2.11, cuya asociación es positiva ($\hat{\theta} = 7 \times 2 / 0 \times 3 > 1$ o $\hat{\theta}' = 7,5 \times 2,5 / 0,5 \times 3,5 > 1$) de modo que la significación, de ocurrir, es por causa de una asociación positiva.

Para adaptar la tabla al formato del ANEXO, en realidad no es preciso ir rotándola sucesivas veces como se hizo con la Tabla 2.10, sino que basta con asignar del modo adecuado a cada número el valor de n , n_1 , x_1 y x_2 (en ese orden). Así, observando la Tabla 2.11:

- 1) El valor n es sin duda $n=12$.
- 2) El valor n_1 es $n_1=2$ (pues ha de ser el mínimo marginal).
- 3) El valor x_1 ha de ser 2 o 0, el que consiga que $\hat{p}_1 < \hat{p}_2$. Como $2/2 > 3/10$, entonces $x_1=0$ (pues $0/2 < 7/10$).
- 4) Finalmente, $x_2=7$ (su pareja de la otra columna).

Con todo ello $(n, n_1, x_1) = (12, 2, 0)$ que en el ANEXO (columna "DOS COLAS") da las fronteras $C=6, 7$ y 9 para $\alpha = 10\%$, 5% y 1% respectivamente. Como $x_2=7$ es mayor o igual que los dos primeros, pero no que el tercero, entonces $1\% < P \leq 5\%$ (el valor real es $4,56\%$).

Sin embargo la misma tabla evaluada por las tablas de Shuster (1992) da un $Z_Y = 1,047$ con una significación $9,95\% < P < 15,12\%$ netamente inferior a la de arriba (el valor exacto es $11,15\%$).

2.10. CONCLUSIONES*

A lo largo del Capítulo se ha visto cómo la literatura acerca del Modelo I es escasa, dispersa, con análisis parciales del problema y basada en métodos, programas y tablas no óptimos, lo que se ha intentado corregir aquí dado lo frecuente de tales

tablas 2×2 en las Ciencias Experimentales (especialmente en Psicología y Ciencias de la Salud). Con tal fin:

- 1) Se ha formalizado el problema estableciendo un formato simplificado (ó canónico) para el espacio muestral (para lo que se precisó definir las condiciones de Equivalencia y de Simetría) y las condiciones a verificar por cualquier región crítica coherente (las condiciones de Convexidad).
- 2) Se han definido nuevos métodos de ordenación (a añadir a los 5 ya existentes).
- 3) Se han proporcionado diversas propiedades que faciliten la determinación del máximo de $\alpha(p, q)$, bien por agrupación de cálculos, bien por reducción del espacio paramétrico, bien por reducción del número de máximos a calcular (en el caso del método original de Barnard).
- 4) Se ha particularizado al Modelo I la modificación del concepto tradicional de potencia realizada por Martín and Silva (1994) para el Modelo II, obteniendo así los conceptos de "potencia a largo plazo" y de "potencia media", conceptos que, por ser una globalización de la potencia ordinaria a todo el espacio paramétrico a un conjunto de errores α , permiten evaluar comparativamente, de un modo más sencillo y apropiado, los diversos métodos de ordenación estudiados.
- 5) Se han estudiado comparativamente todos los métodos de ordenación existentes y los nuevos aquí propuestos, obteniendo que:
 - a) El nuevo método **M** es el más potente, pero es 2,4 veces más costoso en tiempo de cómputo que el **M'**.
 - b) El nuevo método **M'** es el siguiente en potencia para los n pequeños (y muy cercano al **M** en el caso del test de dos colas), pero es 95 veces más costoso en tiempo de cálculo que los que le siguen. De hecho es el más potente para $n \leq 10$.
 - c) Los nuevos métodos **FM** (FM1 para 1 cola; FM2 para dos colas) son los siguientes en potencia (incluso mejores que **M'** en los n más altos), pero resultan ser los óptimos en el equilibrio potencia/tiempo de cómputo.

- d) De entre los métodos de tipo chi-cuadrado (que son los peores) los nuevos métodos **PH2/PH3** son los mejores para los errores $\alpha > 5\%$ en tests de 1 cola/2 colas, en tanto que el clásico **Y** lo es para $\alpha \leq 5\%$, aunque de modo global es preferible el **Y**.
- 6) Se ha evaluado la pérdida de potencia por usar el test condicionado en lugar del actual, obteniendo que siempre es menos potente y que:
- a) El test incondicionado es especialmente conveniente a una cola, a errores α altos y con tamaño de muestra pequeño.
- b) El test condicionado puede entenderse casi igual que el incondicionado (un 1% de pérdida de potencia) desde $n=110$ (una cola) o $n=80$ (dos colas), si bien tales resultados han sido obtenidos por extrapolación.
- 7) Se ha evaluado la pérdida de potencia por usar un método incondicionado no óptimo, concluyéndose que el FM puede entenderse casi igual que el M (un 1% de pérdida de potencia) desde $n \geq 40$ (1 cola) o $n \geq 50$ (2 colas).
- 8) De todo lo anterior se desprende que la regla general de actuación es la que se indica en la Tabla 2.12.

Tabla 2.12

Método exacto incondicionado óptimo a seleccionar en el Modelo I según el tamaño n de muestra (en el interior de la tabla)

(a)

En base a la potencia

Método	M'	M
1 o 2 COLAS	≤ 15	≥ 16

(b)

En base a su accesibilidad por tablas o tiempo de cómputo y, subsidiariamente, a la potencia (en los métodos no óptimos puede perderse alrededor de un 1% de potencia respecto del óptimo)

Método	M'	M	FM	FC
1 COLA	≤ 15	16-50	51-109	≥ 110
2 COLAS			51-79	≥ 80

- 9) Se han ofrecido programas en lenguaje C para calcular el valor P de una tabla por cualquiera de los métodos descritos, así como la dirección electrónica donde encontrarlos.
- 10) Se han ofrecido tablas de regiones críticas para el método óptimo M.

CAPÍTULO III

TESTS APROXIMADOS

En el Capítulo anterior ya se ha citado la gran complejidad de cómputo del test exacto de Barnard, el cual es impracticable para tamaños de muestra incluso pequeños (digamos, para $n > 40$). Es por ello que diversos autores emprendieron soluciones exactas menos potentes que la anterior pero con menor complejidad de cómputo. Con todo, tales tests (por ejemplo el basado en la ordenación P-mid de Fisher) son también impracticables para tamaños de muestra incluso moderados (digamos, para $n > 70$). Es por ello que se hace patente la necesidad de buscar *métodos aproximados* que permitan resolver el problema con ciertas garantías. Otros autores, en lugar de la dicotomía "métodos exactos"/"métodos aproximados", prefieren hablar de "métodos no asintóticos"/"métodos asintóticos".

Como ya se indicó anteriormente, tanto el método condicionado como el incondicionado (es decir, el basado en los tres Modelos I, II y III) tienen como método asintótico el que proporciona el clásico estadístico chi-cuadrado -en sus versiones (1.17) o (1.18)-, radicando la diferencia entre unos y otros en la *corrección por continuidad* (*c.p.c* en adelante) a efectuar. Sin embargo es claro que un método asintótico no es válido en cualquier circunstancia, sino "asintóticamente", con lo que aún resta el problema complementario de decidir bajo qué *condiciones de validez* (*c.d.v.* en adelante) es válido el mismo. Todo esto es el objetivo del actual Capítulo.

3.1. ESTADO ACTUAL DEL PROBLEMA

3.1.1. Introducción

La propuesta y evaluación de métodos asintóticos para el Modelo I actual está

bastante menos desarrollada que la del caso del Modelo II, y ello, una vez más, por los problemas de cómputo del test exacto al que se pretende aproximar. Para el Modelo II se conoce bastante bien las diversas posibilidades (Martín et al., 1992), las c.p.c. a efectuar (Conover, 1974; Haber, 1980 y 1982; Martín and Silva, 1996) y la opción óptima con sus c.d.v. (Martín and Silva, 1996). Pero para el Modelo I se está actualmente en la fase de propuestas de estadísticos sin c.p.c. o en las primeras evaluaciones de algunas c.p.c. De hecho, la literatura para las versiones asintóticas del Modelo II es muy abundante (pudiendo hablarse de cientos de artículos) en tanto que la del Modelo I es bastante más restringida (pudiendo hablarse sólo de una par de decenas de artículos). En esta sección se describe en orden conceptual (no en orden histórico) el estado actual de la cuestión.

3.1.2. Convenios

La práctica totalidad de los métodos asintóticos para el Modelo I son de tipo chi-cuadrado, y así al símbolo χ^2 del estadístico que los ocasiona habrá que ponerle un subíndice X para distinguir unos métodos de otros: χ^2_X . Tales estadísticos son sin c.p.c. y, cuando se la usa, suelen dar lugar a varios otros estadísticos según cuál sea el tipo de c.p.c. impuesta. Para distinguir unas de otras habremos de colocar uno o más subíndices extras Y (tras el X anterior): χ^2_{XY} . Con frecuencia, en el texto, si no hay lugar a confusión, se citará abreviadamente al método a través de subíndice (X por ejemplo) en lugar del símbolo completo (χ^2_X).

Por otro lado, dado un método asintótico X cualquiera cuyo estadístico de contraste χ^2_X se distribuye asintóticamente como una $\chi^2_{1\text{ g.l.}}$, el valor P para el test de una cola en una tabla específica será:

$$P_X = \frac{P\{\chi^2_{1\text{ g.l.}} \geq \chi^2_X\}}{2} = \frac{P(\chi^2_X)}{2} \quad (\text{una cola}) \quad (3.1)$$

si la alternativa está en la dirección apropiada, en tanto que para dos colas será:

$$P_X = P\{\chi^2_{1\text{ g.l.}} \geq \chi^2_X\} = P(\chi^2_X) \quad (\text{dos colas}) \quad (3.2)$$

es decir, el doble del anterior. Esto de por sí ya presenta un primer problema en cuanto a la actuación del test asintótico, pues, según lo anterior, el valor P del test de dos colas es el doble del valor P del test de una cola, cosa que no sucedía en los tests exactos que se pretenden aproximar (ver 2.3.3).

Finalmente, en esta sección se nos va a permitir manejar una doble notación para los datos de la tabla experimental, y así la notación tradicional de las Tabla 3.1 (a) y (c) se verá complementada por la notación eventual de la Tabla 3.1(b) (que es más clásica). La razón de tal proceder es que determinadas expresiones son más cómodas de escribir y demostrar con la nueva notación, pero la vieja notación nos será más útil cuando haya que aludir al espacio muestral (pues así se mantienen las imágenes gráficas reseñadas en el capítulo anterior). La última tabla -la 3.1(d)- se obtiene del modo clásico:

Tabla 3.1

Notaciones para el Modelo I

<p>(a)</p> <p>Resultados: 1ª opción (actual)</p> <table border="1" style="margin-left: auto; margin-right: auto; border-collapse: collapse; text-align: center;"> <thead> <tr> <th></th> <th>A</th> <th>\bar{A}</th> <th>Total</th> </tr> </thead> <tbody> <tr> <th>B</th> <td>x_1</td> <td>y_1</td> <td>n_1</td> </tr> <tr> <th>\bar{B}</th> <td>x_2</td> <td>y_2</td> <td>n_2</td> </tr> <tr> <th>Total</th> <td>a_1</td> <td>a_2</td> <td>n</td> </tr> </tbody> </table>		A	\bar{A}	Total	B	x_1	y_1	n_1	\bar{B}	x_2	y_2	n_2	Total	a_1	a_2	n	<p>(b)</p> <p>Resultados: 2ª opción (clásica)</p> <table border="1" style="margin-left: auto; margin-right: auto; border-collapse: collapse; text-align: center;"> <thead> <tr> <th></th> <th>A</th> <th>\bar{A}</th> <th>Total</th> </tr> </thead> <tbody> <tr> <th>B</th> <td>n_{11}</td> <td>n_{12}</td> <td>$n_{1\cdot}$</td> </tr> <tr> <th>\bar{B}</th> <td>n_{21}</td> <td>n_{22}</td> <td>$n_{2\cdot}$</td> </tr> <tr> <th>Total</th> <td>$n_{\cdot 1}$</td> <td>$n_{\cdot 2}$</td> <td>$n_{\cdot\cdot} = n$</td> </tr> </tbody> </table>		A	\bar{A}	Total	B	n_{11}	n_{12}	$n_{1\cdot}$	\bar{B}	n_{21}	n_{22}	$n_{2\cdot}$	Total	$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot\cdot} = n$
	A	\bar{A}	Total																														
B	x_1	y_1	n_1																														
\bar{B}	x_2	y_2	n_2																														
Total	a_1	a_2	n																														
	A	\bar{A}	Total																														
B	n_{11}	n_{12}	$n_{1\cdot}$																														
\bar{B}	n_{21}	n_{22}	$n_{2\cdot}$																														
Total	$n_{\cdot 1}$	$n_{\cdot 2}$	$n_{\cdot\cdot} = n$																														
<p>(c)</p> <p>Probabilidades</p> <table border="1" style="margin-left: auto; margin-right: auto; border-collapse: collapse; text-align: center;"> <thead> <tr> <th></th> <th>A</th> <th>\bar{A}</th> <th>Total</th> </tr> </thead> <tbody> <tr> <th>B</th> <td>p_{11}</td> <td>p_{12}</td> <td>$p_{1\cdot} = q$</td> </tr> <tr> <th>\bar{B}</th> <td>p_{21}</td> <td>p_{22}</td> <td>$p_{2\cdot} = 1 - q$</td> </tr> <tr> <th>Total</th> <td>$p_{\cdot 1} = p$</td> <td>$p_{\cdot 2} = 1 - p$</td> <td>$p_{\cdot\cdot} = 1$</td> </tr> </tbody> </table>		A	\bar{A}	Total	B	p_{11}	p_{12}	$p_{1\cdot} = q$	\bar{B}	p_{21}	p_{22}	$p_{2\cdot} = 1 - q$	Total	$p_{\cdot 1} = p$	$p_{\cdot 2} = 1 - p$	$p_{\cdot\cdot} = 1$	<p>(d)</p> <p>Esperadas</p> <table border="1" style="margin-left: auto; margin-right: auto; border-collapse: collapse; text-align: center;"> <thead> <tr> <th></th> <th>A</th> <th>\bar{A}</th> <th>Total</th> </tr> </thead> <tbody> <tr> <th>B</th> <td>E_{11}</td> <td>E_{12}</td> <td>$n_{1\cdot}$</td> </tr> <tr> <th>\bar{B}</th> <td>E_{21}</td> <td>E_{22}</td> <td>$n_{2\cdot}$</td> </tr> <tr> <th>Total</th> <td>$n_{\cdot 1}$</td> <td>$n_{\cdot 2}$</td> <td>n</td> </tr> </tbody> </table>		A	\bar{A}	Total	B	E_{11}	E_{12}	$n_{1\cdot}$	\bar{B}	E_{21}	E_{22}	$n_{2\cdot}$	Total	$n_{\cdot 1}$	$n_{\cdot 2}$	n
	A	\bar{A}	Total																														
B	p_{11}	p_{12}	$p_{1\cdot} = q$																														
\bar{B}	p_{21}	p_{22}	$p_{2\cdot} = 1 - q$																														
Total	$p_{\cdot 1} = p$	$p_{\cdot 2} = 1 - p$	$p_{\cdot\cdot} = 1$																														
	A	\bar{A}	Total																														
B	E_{11}	E_{12}	$n_{1\cdot}$																														
\bar{B}	E_{21}	E_{22}	$n_{2\cdot}$																														
Total	$n_{\cdot 1}$	$n_{\cdot 2}$	n																														

$$E_{ij} = \frac{n_{i.} \times n_{.j}}{n_{..}} \quad (3.3)$$

reservando la letra E, sin subíndices, para aludir a la *mínima cantidad esperada*:

$$E = \text{Min}_{i,j} E_{ij} \quad (3.4)$$

cantidad que es la relevante para especificar las c.d.v.

3.1.3. Estadísticos sin corrección por continuidad

3.1.3.1. Estadísticos de tipo chi-cuadrado clásico

Desde Pearson (1900) ya se conoce el estadístico $\chi^2_U = \sum (n_{ij} - E_{ij})^2 / E_{ij}$, que es el que figura en casi todos los libros especializados o no. Realizando operaciones, es conocido que tal expresión puede simplificarse en esta otra que ya fue aludida con anterioridad:

$$\chi^2_Z = \frac{\{x_1 y_2 - x_2 y_1\}^2}{a_1 a_2 n_1 n_2} n \quad (3.5)$$

Menos conocida es la siguiente opción debida a Pirie and Hamdan (1972). Ellos indican que como $H_0 \equiv \theta = 1$ es equivalente a $H_0 \equiv p_{11}p_{22} = p_{12}p_{21}$ o también a $H_0 \equiv p_{11}p_{22} - p_{12}p_{21} = \delta = 0$, y como el estimador insesgado de δ es $(n_{11}n_{22} - n_{12}n_{21}) / \{n(n-1)\}$, entonces un estadístico apropiado para el test es $U = n_{11}n_{22} - n_{12}n_{21}$. Su media es (bajo H_0):

$$E(U) = E(n_{11}n_{22}) - E(n_{12}n_{21}) = n(n-1)\delta = 0$$

y de ahí la insesgadez de antes. Su varianza, bajo H_0 , es:

$$V(U) = E(U^2) = E(n_{11}n_{22})^2 + E(n_{12}n_{21})^2 - 2 E(n_{11}n_{12}n_{21}n_{22}) = n^2(n-1) p_{11}p_{22} = \sigma^2$$

La prueba de lo anterior puede verse en detalle en las pgs 40 a 42 de Ortiz (1993) por ejemplo, pero en esencia todas las claves están en la siguiente demostración que alude al tercer sumando de la expresión anterior:

$$\begin{aligned}
E\left(\prod_{i,j} n_{ij}\right) &= \sum \frac{n!}{\prod_{i,j} n_{ij}!} \prod_{i,j} p_{ij}^{n_{ij}} \left(\prod_{i,j} n_{ij}\right) = \\
&= n(n-1)(n-2)(n-3) \prod_{i,j} p_{ij} \left\{ \sum \frac{(n-4)!}{\prod_{i,j} (n_{ij}-1)!} \prod_{i,j} p_{ij}^{n_{ij}-1} \right\} \\
&= n(n-1)(n-2)(n-3) p_{11} p_{12} p_{21} p_{22}
\end{aligned}$$

Como los estimadores de máxima verosimilitud de $p_{i\cdot}$ y $p_{\cdot i}$ son $n_{i\cdot}/n$ y $n_{\cdot i}/n$ y, por el Teorema de Zehna, los estimadores de máxima verosimilitud de p_{ii} serán $\hat{p}_{ii} = n_{i\cdot} \times n_{\cdot i} / n^2$, entonces, finalmente, $\hat{\sigma}^2 = (n-1)n_{1\cdot}n_{2\cdot}n_{\cdot 1}n_{\cdot 2} / n^2$. Con ello, asintóticamente, $U/\hat{\sigma}$ se distribuye como una normal típica (bajo H_0), con lo que el estadístico final, en formato chi-cuadrado, será:

$$\chi_P^2 = \frac{\{x_1 y_2 - x_2 y_1\}^2}{a_1 a_2 n_1 n_2} \cdot \frac{n^2}{n-1} \quad (3.6)$$

sí bien ellos mismos proponen regresar a la clásica χ_Z^2 pues para n grande será $n/(n-1) \approx 1$.

Años más tarde, Haber (1986), sin justificación teórica sino sólo por paralelismo con el caso del Modelo II, propone el estadístico:

$$\chi_R^2 = \frac{\{x_1 y_2 - x_2 y_1\}^2}{a_1 a_2 n_1 n_2} (n-1) \quad (3.7)$$

que Richardson (1990) justifica del modo siguiente: si se hace:

$p_{11} p_{22} = (p_{1\cdot} p_{\cdot 1})(p_{2\cdot} p_{\cdot 2}) = \{p_{1\cdot}(1-p_{1\cdot})\}\{p_{\cdot 1}(1-p_{\cdot 1})\} = q(1-q) p(1-p)$, como los estimadores de máxima verosimilitud de $p(1-p)$ y de $q(1-q)$ son $\hat{p}(1-\hat{p})n/(n-1)$ y $\hat{q}(1-\hat{q})n/(n-1)$, es decir, $n_{\cdot 1}n_{\cdot 2}/\{n(n-1)\}$ y $n_{1\cdot}n_{2\cdot}/\{n(n-1)\}$, entonces el estimador de máxima verosimilitud de $p_{11} p_{22}$ es el producto de ellos y el de σ^2 será:

$$\hat{\sigma}^2 = n^2(n-1) \frac{n_{1\cdot}n_{2\cdot}}{n(n-1)} \frac{n_{\cdot 1}n_{\cdot 2}}{n(n-1)} = \frac{n_{1\cdot}n_{2\cdot}n_{\cdot 1}n_{\cdot 2}}{n-1}$$

con lo que $U/\hat{\sigma} \rightarrow N(0,1)$, asintóticamente, y su cuadrado da la (3.7). Puesto que

Richardson es quien justifica la forma del estadístico para el Modelo I, se ha puesto R como subíndice de él. Es claro que:

$$\chi_P^2 = \chi_Z^2 \frac{n}{n-1} \quad ,, \quad \chi_R^2 = \chi_Z^2 \frac{n-1}{n} \quad (3.8)$$

y también que:

$$\chi_R^2 < \chi_Z^2 < \chi_P^2 \quad (3.9)$$

con lo que el método P dará más significaciones que el método Z, que a su vez las dará con más frecuencia que el R.

3.1.3.2. Otros estadísticos de tipo chi-cuadrado

Los estadísticos que siguen, todos ellos en formato chi-cuadrado, son menos tradicionales y de menos importancia a los efectos de esta Memoria. De ahí que a continuación nos limitemos a citarlos en su forma y origen.

El clásico χ_Z^2 fue modificado por Neyman (1949) en el siguiente sentido:

$$\chi_N^2 = \sum \frac{(n_{ij} - E_{ij})^2}{n_{ij}} = \sum \frac{E_{ij}^2}{n_{ij}} - n \quad (3.10)$$

Awad and Sarie (1987) alude a diversos estadísticos en el ámbito más amplio de las tablas $r \times c$, en particular es reseñable el clásico estadístico de Freeman-Tukey:

$$\chi_F^2 = 4 \sum \left\{ \sqrt{n_{ij}} - \sqrt{E_{ij}} \right\}^2 \quad (3.11)$$

Sin embargo, los propios Award and Sarie prueban que tales estadísticos lo más que hacen es comportarse tan bien como el clásico y, además, los mismos no han recibido ninguna atención por ningún autor en el ámbito restringido de las Tablas 2×2 . De ahí que no se les cite en lo que sigue.

Más tradicional es el estadístico procedente del test de razón de verosimilitudes,

bien en su formato usual

$$X_L^2 = 2 \sum n_{ij} \ln \frac{n_{ij}}{E_{ij}} \quad (3.12)$$

bien con la modificación de Kullback (1959):

$$X_K^2 = 2 \sum E_{ij} \ln \frac{E_{ij}}{n_{ij}} \quad (3.13)$$

Los estadísticos (3.5) y (3.10) a (3.13) pueden ponerse en este formato único:

$$X_\lambda^2 = \frac{2}{\lambda(\lambda+1)} \sum n_{ij} \left\{ \left(\frac{n_{ij}}{E_{ij}} \right)^\lambda - 1 \right\} \quad (3.14)$$

pues haciendo $\lambda = 1, -2, -1/2, 0$ y -1 se obtiene cada uno de los citados (los casos $\lambda=0$ y $\lambda=-1$ se definen en el límite). Cressie and Read (1989) probaron que los valores $\lambda=0$ (X_L^2) y $\lambda=1$ (X_Z^2) daban lugar a estadísticos que verificaban determinadas propiedades deseables, proponiendo acoger las ventajas de ambos a través del valor intermedio $\lambda=2/3$.

$$X_{CR}^2 = \frac{9}{5} \sum n_{ij} \left\{ \left(\frac{n_{ij}}{E_{ij}} \right)^{2/3} - 1 \right\} \quad (3.15)$$

Finalmente, es también tradicional el estadístico de Gart (1966) cuyo uso suele restringirse a la obtención de intervalos de confianza para las diversas medidas de asociación, más que para el test de hipótesis objeto de esta memoria:

$$X_G^2 = \frac{\{\ln \hat{\theta}'\}^2}{\sum 1/(n_{ij}+0,5)} \quad (3.16)$$

con θ aludiendo a la odds-ratio, como siempre, y $\hat{\theta}'$ a su estimación ya citada de $(n_{11}+0,5)(n_{22}+0,5)/(n_{12}+0,5)(n_{21}+0,5)$: recordar la (1.45).

Todos los estadísticos citados son asintóticamente equivalentes al χ_Z^2 . El caso

del χ^2_L puede verse en las pgs. 48 y 49 de Ortiz (1993).

3.1.4. Estadísticos con corrección por continuidad

3.1.4.1. Introducción

Para compensar el hecho de que una variable aleatoria discreta -el estadístico chi-cuadrado, que proviene de la multinomial original- se está aproximando a través de una variable aleatoria continua -la χ^2 con 1 g.l.- es tradicional y conveniente efectuar una c.p.c., lo que fue justificado gráficamente por Pearson (1947) y analíticamente por Cox (1970).

De un modo general (Kendall and Stuart, 1973) una c.p.c. consiste en tomar como "valor observado" la media entre el valor realmente observado y su inmediato siguiente menos extremo, ó, equivalentemente, en sumar o restar al primero de ellos la mitad del salto entre ambos. Esto lleva a que, en el caso actual, la c.p.c. sea una u otra dependiendo de quién se considere que es la variable base del problema: el estadístico χ^2 , el estadístico χ o el estadístico $U = x_1y_2 - x_2y_1$.

3.1.4.2. Correcciones procedentes del método condicionado

La c.p.c. más tradicional es la clásica ya citada de Yates (1934), pero ella, teóricamente, sólo es válida para el caso condicionado del Modelo III, no para el Modelo I incondicionado actual. Sin embargo muchos autores la han contemplado (por sí misma, o con sus variantes) para criticar su actuación incondicionada y mostrar que era preciso buscar otras alternativas. Es por ello que aquí se la va a definir, indicando a la vez el criterio que le da origen.

Razonar sobre los saltos de U o de χ es equivalente en el Modelo III, pues las a_i y n_i son siempre las mismas. La variable U salta de n en n (cuando x_1 se incrementa o decrementa en 1), y así el clásico estadístico de Yates (1934) es:

$$\chi^2_Y = \frac{\left\{ |x_1y_2 - x_2y_1| - \frac{n}{2} \right\}^2}{a_1a_2n_1n_2} n \quad (3.17)$$

Sin embargo tal estadístico es apropiado sólo para test de una cola, pues, de aplicarse como test de dos colas, se asume que el valor P de dos colas es el doble del de una cola, y tal cosa no sucede en general en el test exacto de Fisher al que se está intentando aproximar. Mantel (1974) y Conover (1974) ya señalaron este hecho (ante las críticas que recibía el test χ^2_Y en su actuación en dos colas), indicando que el error consiste en no tener en cuenta que el valor observado x_1 da lugar a un valor χ^2_Y , pero que, por la otra cola, el valor x_1' (tan extremo o más que x_1) da lugar a otro valor $\chi^2'_Y$ que, en general, es distinto. Estos son sus razonamientos:

Mantel (1974): Si x_1 es el valor elemental y se supone, sin pérdida de generalidad, que da una asociación negativa ($\hat{\theta} < 1$), el valor x_1' de la otra cola (asociación positiva) que da un χ^2_Z mayor o igual que el de x_1 es

$$x_1' = \left[\frac{2a_1n_1}{n} - x_1 \right]^+ \quad (3.18)$$

con $[x]^+$ aludiendo al primer entero mayor o igual que x . La tabla de x_1 da una chi-cuadrado de Yates de χ^2_Y , en tanto que la de x_1' da $\chi^2'_Y$, ambas a partir de la (3.17); con ello, el valor P de Mantel para el test de dos colas es:

$$P_M = \frac{P(\chi^2_Y) + P(\chi^2'_Y)}{2} = P(\chi^2_M) \quad (3.19)$$

con χ^2_M aludiendo al valor de la $\chi^2_{1 \text{ g.l.}}$ cuyo percentil es $(1-P_M)$. Mantel, por tanto, promedia los valores P de cada cola.

Conover (1974): Conover razonó de un modo algo diferente, pues pensaba que no hay que trabajar con las dos colas por separado, sino que hay que aplicar el criterio de Kendall and Stuart literalmente. Para ello consideró que, así como x_1 da lugar a un valor χ^2_1 de χ^2_Z , los valores x_1-1 , x_1' y $x_1'+1$ darán valores de χ^2_Z de χ^2_2 , $\chi^2_{1'}$ y $\chi^2_{2'}$, en donde χ^2_2 es el inmediato χ^2_Z más pequeño que χ^2_1 , y $\chi^2_{2'}$ es lo propio con respecto a $\chi^2_{1'}$ (que, recuérdese, era $\geq \chi^2_1$). Con ello, χ^2_1 debe promediarse con su inmediato más pequeño que será χ^2_2 o $\chi^2_{2'}$, y al que llamaremos $\chi^2_{22'}$, y así el valor P para el test de dos colas según Conover es:

$$P_C = P\left\{\frac{\chi_1^2 + \chi_{22'}^2}{2}\right\} = P(\chi_C^2) \quad (3.20)$$

Se observa pues que para Mantel hay dos colas (puesto que las distingue), pero para Conover no (puesto que las mezcla). Aún existen otras posibilidades de c.p.c. condicionada (tanto a una como a dos colas), pero la literatura incondicionada no las contempla y, así, tampoco es citada aquí. Ver Haber (1980) y Martín et al (1992).

La notación para estos métodos (Y, M y C) no es la prometida (debería de anteponérsele la Z del estadístico de que provienen) pues ellos no ocupan el papel de métodos incondicionados.

3.1.4.3. Corrección de Pirie and Hamdan

Pirie and Hamdan (1972) fueron los primeros autores en observar que la c.p.c para el test chi-cuadrado debería variar con el modelo asumido y en dar argumentos teóricos para cada caso.

Ellos indican que para contrastar $H_0 \equiv \delta = 0$ vs $H_1 \equiv \delta \neq 0$, con δ un parámetro de la función de probabilidad P_D (discreta), si U es un estimador insesgado y suficiente de δ (ó de un múltiplo de él) que salta de h en h , entonces la RC es de la forma $|U| \geq r$, con r el máximo valor tal que $P_D\{|U| \geq r\} \leq \alpha$. Si U tiene por distribución asintótica a la normal P_N , Cox (1970) ya que indicó que:

$$\alpha \geq P_D\{|U| \geq r\} \approx P_N\left\{|U| \geq r - \frac{h}{2}\right\}$$

con $h/2$ aludiendo a "la mitad del salto". Supongamos que $\hat{\sigma}^2$ es un estimador de máxima versosimilitud y consistente de $\sigma^2 = V(U)$, entonces:

$$\alpha \geq P_D\left\{\frac{|U|}{\hat{\sigma}} \geq \frac{r}{\hat{\sigma}}\right\} \approx P_N\left\{\frac{|U|}{\hat{\sigma}} \geq \frac{r-h/2}{\hat{\sigma}}\right\}$$

pues $U/\hat{\sigma}$ será asintóticamente normal bajo H_0 . Con ello el estadístico con c.p.c. a comparar con la normal típica (ó su cuadrado con una $\chi^2_{1.g.l.}$) será:

$$\frac{|U| - h/2}{\hat{\sigma}} \quad (3.21)$$

y de modo similar para el caso del test de una cola.

La (3.21) puede aplicarse a cualquiera de los estadísticos χ^2_Z , χ^2_P y χ^2_R dados por las (3.5), (3.7) y (3.8) respectivamente, pues la notación usada aquí es la misma de allí, con lo que los estadísticos con c.p.c. serán:

$$\chi^2_{ZP} = \frac{\{|x_1y_2 - x_2y_1| - 0,5\}^2}{a_1a_2 n_1n_2} n \quad (3.22)$$

$$\chi^2_{PP} = \frac{\{|x_1y_2 - x_2y_1| - 0,5\}^2}{a_1a_2 n_1n_2} \frac{n^2}{n-1} \quad (3.23)$$

$$\chi^2_{RP} = \frac{\{|x_1y_2 - x_2y_1| - 0,5\}^2}{a_1a_2 n_1n_2} (n-1) \quad (3.24)$$

aunque Pirie and Hamdan sólo proponen los dos primeros y, además, se decantan por el primero por la razón señalada en 3.1.3.1. La letra P añadida al final de Z, P o R alude a que se trata de la c.p.c. de Pirie and Hamdan. La c.p.c de 0,5 proviene de la observación de Pirie and Hamdan de que la variable $U = x_1y_2 - x_2y_1$ salta casi siempre (en el Modelo I) de 1 en 1, siendo difícil localizar los casos en que esto no es así.

3.1.4.4. Corrección de Conover

Conover (1974) apuesta por promediar los valores de χ^2_Z , más que por promediar U como Pirie and Hamdan, con lo que:

$$\frac{\chi^2_{Z^+} + \text{su próximo más pequeño } (\chi'^2_Z)}{2}$$

Sin embargo él mismo reconoce que el " χ^2_z " próximo más pequeño que el experimental χ^2_z es imposible de obtener de un modo sencillo (como se supone que debe ser un método asintótico), abandonando cualquier discusión o estudio de tal método.

3.1.5. Evaluación de los diversos métodos asintóticos

3.1.5.1. Datos de la literatura

En lo que sigue se hace una breve exposición temporal de los diversos artículos que abordan el análisis asintótico de una tabla 2×2 emergente del Modelo I, bien desde el punto de vista de evaluar uno o más métodos, bien con la intención de compararlos y obtener el óptimo. En la exposición se alude no sólo a las conclusiones más relevantes, sino también al modo en que éstas han sido obtenidas (pues ello puede afectar a su fiabilidad). Un modo frecuente de proceder (y que se cita aquí para evitar repetirlo en varias ocasiones) consiste en decidir el valor n , el error α (5% por ejemplo) y el número de colas del test (2 por ejemplo), determinando a continuación todas las tablas con un valor $\chi^2_x \leq \chi^2_{\text{nominal}}$ ($1,96^2$ para el ejemplo asumido) y determinando, para la RC así obtenida, el error real α^* a partir de la (2.2). La comparación del α^* real con el α nominal (5%) permite extraer las conclusiones.

Grizzle (1967): Contempla sólo el caso de 2 colas, $\alpha=1\%$ y 5% y $n=20(20)120$.

Para α y n fijados, asume H_0 y fija también 5 parejas variadas de valores de (p,q) . Para cada una de ellas genera por Montecarlo 500 tablas 2×2 y determina el porcentaje experimental $\hat{\alpha}$ de rechazos (indebidos) de H_0 por los criterios χ^2_z y χ^2_y . Haciendo la gráfica de $\hat{\alpha}$ vs n , estudia la evolución de la poligonal obtenida frente a la línea de $\alpha=1\%$ o 5% . Sus conclusiones son:

- 1) χ^2_y siempre es conservador.
- 2) χ^2_z va bien en promedio, pero es conservador en los pequeños tamaños de muestra.

Camilli and Hopkins (1978 y 1979): Mantiene el esquema de Grizzle, con las modificaciones de añadir $\alpha=10\%$, restringir n a 20, 50 y 100, observar 9 parejas de valores de (p,q) y generar 10.000 tablas. Sus conclusiones son

similares:

- 1) χ^2_Y es conservador.
- 2) χ^2_Z va bien en promedio.

Haber (1986): Para test de 1 o 2 colas, $\alpha=5\%$ y $10 \leq n \leq 30$, evalúa α^* y lo compara con α para el caso de χ^2_R , concluyendo que es un test liberal (aunque muy débilmente a dos colas). Obsérvese que es la primera vez que se evalúa α^* (que es lo correcto).

Kroll (1989): Para tests de 2 colas (aunque no lo especifica claramente), $\alpha=5\%$ y $n=14, 19$ y 30 , genera las RC para los métodos $\chi^2_Z, \chi^2_R, \chi^2_Y, \chi^2_M$ y χ^2_G y determina $\alpha(p,q)$ en $p,q=0,5(0,1)0,9$ (lo que es un prelude del α^*) comparándolos con α . El procedimiento es pues intermedio entre los dos anteriores. Sus conclusiones son:

- 1) χ^2_Z y χ^2_R son ligeramente conservadores (del orden de un 6% frente a un 5%).
- 2) χ^2_Y, χ^2_M y χ^2_G son muy conservadores.

Richardson (1990): Con igual planteamiento que el anterior, emplea los valores $n=20$ y 100 y sólo 6 parejas de valores de (p,q) , evaluando los métodos $\chi^2_Z, \chi^2_R, \chi^2_Y$ y χ^2_C . Sus conclusiones son:

- 1) χ^2_R y χ^2_Z son aceptables.
- 2) χ^2_Y y χ^2_C son conservadores.

lo cual ratifica generando 10.000 tablas por Montecarlo y actuando como Camilli and Hopkins.

Para el caso de $n=20$, obtiene la potencia de tales test para valores $\theta=2$ y 10 , concluyendo que:

- a) χ^2_Y y χ^2_C son poco potentes.
- b) χ^2_R es poco potente para p muy diferente de q .
- c) χ^2_Z va bien.

Lee and Shen (1994): Para tests de 2 colas (aunque no lo especifica claramente), $\alpha=1\%$ y 5% , $n=20(10)70$ y 4 parejas de valores de $(p_{12}/p_{11}, p_{21}/p_{11})$ determina por simulación $\hat{\alpha}^*$ y lo compara con α bajo la idea de Cochran (1942) de que

un α^* del 4% al 6% es adecuado para el objetivo de $\alpha = 5\%$, y que uno del 0,7% al 1,5% lo es para el $\alpha = 1\%$. Los métodos analizados son los que se aluden en estas conclusiones:

- 1) χ^2_Z y χ^2_{CR} son los mejores.
- 2) χ^2_L no se comporta tan bien.
- 3) χ^2_K , χ^2_N y χ^2_F no son satisfactorios.

continuando luego con un análisis de potencia que no tiene mayor interés.

3.1.5.2. Comentarios generales*

Las limitaciones y/o defectos de los 7 estudios anteriores son de diverso tipo:

- 1) Casi todos se han realizado para el test de dos colas, siendo Haber el único que aborda también el caso de una cola.
- 2) La evaluación del error real se efectúa por tres procedimientos que, de peor a mejor, son: determinar $\hat{\alpha}(p,q)$ por simulación, determinar $\alpha(p,q)$ en algunas parejas y determinar α^* . El único fiable (determinar α^*) lo efectúa sólo Haber. Los demás no son de fiar por su dependencia de unas pocas parejas de (p,q) .
- 3) Todas las alusiones a la potencia del método no tienen ningún interés, pues cada método da lugar a un α^* distinto y lógicamente, cuanto más grande es α^* menos potente será el método (por eso los métodos conservadores son poco potentes). El objetivo es ver qué método aproximado respeta más el error α nominal.
- 4) Todos los trabajos se limitan a un máximo de 3 errores α nominales y muchos de ellos sólo a uno (el 5%). Aunque ello es una buena guía, hoy día los investigadores toman sus decisiones en función del valor P (que puede tomar cualquier valor) y no está garantizado que lo que suceda para $\alpha = 5\%$ sea válido también para $\alpha = 3\%$ o $\alpha = 8\%$ por ejemplo.
- 5) Los valores contemplados para n son bastante bajos (especialmente en el caso de Haber), aunque ello es comprensible dada la dificultad de cómputo del método exacto. Los trabajos más fiables serán aquellos en que las afirmaciones realizadas se basen en valores de n variados y amplios.

- 6) Ningún autor introduce en sus comparaciones a los métodos con c.p.c. (ZP, PP, RP) ni al método P. Lo último no presenta problema pues se supone que P se comportará similarmente a los Z y R. Lo primero sí, pues ellos deberían ser el objetivo principal de estudio (por ser las c.p.c. apropiadas para el Modelo I).

Algunos de los defectos citados constituyen errores conceptuales (como el caso de la potencia), y así las conclusiones basadas en ellos no tienen ningún interés. Otros defectos (como la determinación aproximada de α^*) pueden ser comprensibles para los años 60 o 70 (por los problemas de cómputo), pero no después. Obsérvese que en general el trabajo más completo y fiable es el de Haber, pero desgraciadamente alude a sólo un método (el R). Con todo, los trabajos defectuosos por causa del α^* aún pueden ser útiles: si un método es conservador según ellos, también lo será en la realidad, pues $\alpha(p,q) \leq \alpha^*$ es decir, el método es aún más conservador de lo que los autores creen.

3.1.5.3. Conclusiones

En base a todo lo anterior, en lo que sigue se analizan globalmente las conclusiones de la literatura en referencia a los diversos métodos agrupados por familias coherentes:

Métodos condicionados (Y, M y C):

En realidad los autores los incluyen con el fin de probar a los lectores que deben erradicarse de la práctica cuando se analiza el Modelo I. Aunque los análisis están realizados para el test de dos colas, es de suponer que para una cola sucederá igual (pues hay trabajos con α variados e Y vale también para una cola). Todos los autores están de acuerdo que tales métodos son conservadores y deben dejar de utilizarse en este contexto.

Métodos no clásicos (L, F, K, N, CR y G):

El método G es considerado muy conservador y los demás son insatisfactorios (Lee and Shen) salvo el CR que es asimilado en su comportamiento al del clásico Z (por lo que será preferible el clásico, que es más conocido y fácil de calcular).

Métodos clásicos sin c.p.c. (Z, R y P):

Ellos son los más estudiados por su carácter de clásicos y porque diversos autores han mantenido que son los adecuados en el caso del Modelo I. Desgraciadamente las conclusiones son diversas, por lo que resulta difícil saber exactamente a qué atenerse.

En el caso de **R**, Richardson lo califica de adecuado, Haber de liberal y Kroll de ligeramente conservador. Dado que el trabajo más fiable es el de Haber, se deduce que el método P requiere de una pequeña c.p.c. que le quite su carácter de liberal (especialmente para n pequeño).

En el caso de **Z** las discrepancias son menores, pues todos los autores lo califican de ligeramente conservador (Kroll) o de ir bien en general (otros cuatro autores). Aunque ninguno de esos estudios es fiable (pues no evalúan α^*), sus coincidencias dan alguna garantía. Sin embargo, como $\chi^2_Z > \chi^2_R$ y el segundo el ligeramente liberal, el primero debería serlo aún más, lo que no es reflejado por la literatura.

En el caso de **P** no hay nada dicho, pero por la (3.9), el comentario último del párrafo anterior le es de aplicación.

Como resumen final de todo, puede concluirse que:

- 1) Los únicos métodos de interés son los clásicos Z, R y P (muy parecidos entre sí para valores moderados o grandes de n), todos ellos ligeramente liberales (Z es el que menos).
- 2) Los métodos con c.p.c. (PP, RP y ZP) no han sido evaluados pero, por hacer más pequeño el valor del estadístico del que provienen (P, R y Z respectivamente), es previsible que corrijan su pequeña liberalidad. Ellos son pues unos firmes candidatos para su análisis futuro.
- 3) Los métodos procedentes del test condicionado y con c.p.c. (Y, M y C) deben abandonarse por su gran conservadurismo.

3.1.5.4. Condiciones de validez

Ya se ha indicado que un método asintótico no suele ir bien en todo caso, sino

que será fiable bajo determinadas condiciones de validez que habrá que determinar. Desgraciadamente, si ya de por sí los estudios de comportamiento del test asintótico incondicionado son escasos y poco fiables, los de sus c.d.v. lo son aún menos. De hecho no hay ningún trabajo específico ni moderadamente exhaustivo al respecto, sino tan sólo algunas referencias aisladas y basadas en la "impresión" de los autores (algunas de ellas realizada en el marco de tablas más amplias que las 2×2 o en otros modelos). Estas son las diversas afirmaciones:

- 1) Cochran (1954) indica que para $\alpha = 5\%$ debe ser $E > 5$, si bien esto lo hace en el contexto del Modelo III.
- 2) Grizzle (1967) indica que el test va bien para $E \geq 5$, siendo conservador en otro caso.
- 3) Larntz (1978), Koehler and Larntz (1980) y Richardson (1990) indican que el test va bien incluso con valores de E pequeños.
- 4) Camilli and Hopkins (1978 y 1979) indican que para $n \geq 8$ basta con que sea $E = 1$ o 2 .
- 5) Haber (1986), sin especificar E , dice que el test va bien en toda la gama $10 \leq n \leq 30$.
- 6) Martín (1996) estima que el método debe ir bien si algún marginal está equilibrado ($a_1 = a_2$ o $n_1 = n_2$) o si $E > 1,5$.
- 7) Loh (1989) hizo ver que es preciso exigir que E sea mayor que algo (preferiblemente grande) pues de lo contrario el error α puede llegar a hacerse igual a 1.

Hay pues muchos indicios de que las c.d.v. para el Modelo I no van a ser muy exigentes. Por un lado las anteriores afirmaciones de la literatura. Por otro el hecho de que las c.d.v. se suavizan cuando se pasa del Modelo III (Martín and Herranz, 1997) al Modelo II (Martín and Silva, 1996), siendo previsible una nueva suavización de las mismas al pasar al Modelo I actual (que goza de más aleatoriedad en sus marginales), aunque ello puede venir condicionado por el cambio de los criterios de validez por otros más liberales (ver la sección 3.4). Pero todo esto es algo que está pendiente de estudiar.

3.2. NUEVAS CORRECCIONES POR CONTINUIDAD

3.2.1. Introducción

Ya se ha visto anteriormente que los estadísticos competitivos son los Z, P y R dados por las (3.5), (3.6) y (3.7) respectivamente. También se ha visto que, para corregir su ligero liberalismo, convenía realizarles alguna c.p.c., la cual, basada en el clásico criterio de "sumar o restar la mitad del salto", podía efectuarse sobre el estadístico U (Pirie and Hamdan, 1972), sobre el estadístico χ^2 (Conover, 1974) o sobre el estadístico χ (Haber, 1980), lo que **añadiría los subíndices P, C o H** que recuerdan a sus autores. Desgraciadamente los saltos consecutivos de tales estadísticos no son fáciles de localizar, lo que llevó a Pirie and Hamdan a simplificar su método asumiendo que el salto de U en general es de la unidad, definiendo así los estadísticos ZP, PP y RP dados por las (3.22), (3.23) y (3.24).

En el contexto del Modelo II, Haber (1982) observa que, puesto que los saltos de χ no son fáciles de evaluar ni tampoco constantes, conviene cambiar la definición de c.p.c. de "sumar o restar la mitad del salto" a "**sumar o restar la mitad del salto promedio**", lo que Martín and Silva (1996) aplicaron también al estadístico χ^2 y no hay inconveniente en aplicarlo asimismo al estadístico U. Esta segunda definición de c.p.c. era fácilmente evaluable y efectiva en el caso del Modelo II (Martín and Silva, 1996), constituyendo una razonable aproximación a la definición primera (que es la correcta), por lo que también será ensayada aquí con el Modelo I y será recordada **añadiendo un tercer subíndice H** en honor al autor que la propuso. El modo de proceder para ello será el siguiente:

- 1) Para el estadístico X de interés (U, χ^2 o χ), calcular sus valores mínimo (m) y máximo (M) en el rango de todo el espacio muestral. Con ello $m \leq X \leq M$ y $A = M - m$ es el salto total del estadístico o amplitud del salto.
- 2) Calcular el número $(\bar{n} + 1)$ de valores distintos que puede tomar X.
- 3) El salto promedio será entonces $(M - m) / \bar{n}$ y la mitad del mismo:

$$c = \frac{M-m}{2\bar{n}} = \frac{A}{2\bar{n}} \quad (3.25)$$

será la c.p.c. que hay que efectuar a U, χ^2 o χ . Esto da los siguientes formatos simbólicos:

$$|U| - c \quad \text{para el método P} \quad (3.26)$$

$$\chi^2 - c \quad \text{para el método C} \quad (3.27)$$

$$(\chi - c)^2 \quad \text{para el método H} \quad (3.28)$$

En lo que sigue se desarrolla tal concepto para los estadísticos base Z, P y R, pero, antes de continuar, conviene recordar la notación que se asumirá que, aún siendo algo engorrosa, es indicativa del origen del método:

- 1^{er} Subíndice:** Alude al método chi-cuadrado que le da origen: los Z, P o R de las (3.5) a (3.7). Si no hay más subíndices es que se trata de esos propios métodos (3 métodos).
- 2^o Subíndice:** Su presencia alude a que se está efectuando una c.p.c. sobre los estadísticos U, χ^2 o χ , delatando al tipo de estadístico por las letras P, C y H respectivamente. Si no hay más subíndices es que la c.p.c. es la clásica de "la mitad del salto", y de ella, sólo existen los métodos ZP, PP y RP (3 métodos).
- 3^{er} Subíndice:** Será siempre H, y su presencia alude a que se está efectuando una c.p.c. del tipo "la mitad del salto promedio". Ella puede hacerse en tres métodos chi-cuadrado (Z, P o R) sobre tres estadísticos (U, χ^2 y χ aludidos por las letras P, C y H) (9 métodos).
- 4^o Subíndice:** Hay diversas posibilidades de evaluar simplificadaamente el valor de \bar{n} , lo que lleva a que los 9 métodos anteriores se bifurquen en varios. Esto será indicado con los números 0, 1 y 2 o con la letra

A. Si no hay 4° subíndice, ello es señal de que \bar{n} fue determinado del modo más complejo.

De un modo abreviado:

1^{er} Subíndice: Z, P o R según el método chi-cuadrado.

2^o Subíndice: P, C o H según que la c.p.c. se haga sobre U, χ^2 o χ .

3^{er} Subíndice: H aludiendo a que es la c.p.c. "promedio".

4^o Subíndice: 0, 1, 2 o A aludiendo a las diversas opciones del valor \bar{n} . Su ausencia indica que se está tomando el \bar{n} teórico.

Así pues, los estadísticos Z, P, R, ZP, PP y RP ya han sido definidos con anterioridad (pues son conocidos por la literatura), pero los demás son nuevos y se definen explícitamente en lo que sigue. Para facilidad del lector (dada la abundancia de métodos), el Cuadro II los contiene a todos de modo resumido.

3.2.2. Amplitud del salto total y estadísticos que provoca

Es común afirmar que $0 \leq \chi^2_Z \leq n$, con lo que para χ^2_Z sería $M=n$, $m=0$ y $A=n$. El valor $M=n$ se alcanza en $(x_1=n_1, x_2=n_2)$ pero el valor $m=0$ no siempre tiene por qué alcanzarse (como se verá más tarde) salvo que se entienda que una tabla con $x_1=x_2=0$ da una $\chi^2_Z=0$. En principio χ^2_Z en $x_1=x_2=0$ da un valor indeterminado, pero si se hace $x_1=x_2=\Delta$ se obtiene:

$$\chi^2_Z = \frac{n(n_2-n_1)^2}{2n_1n_2} \frac{1}{\frac{n}{\Delta}-2} \rightarrow 0 \text{ si } \Delta \rightarrow 0$$

y así la asunción anterior tiene sentido. Con ello $0 \leq \chi_Z \leq \sqrt{n}$, y la amplitud para χ_Z es $A=\sqrt{n}$. Finalmente, por las (3.25), (3.27) y (3.28), las c.p.c. oportunas son:

$$\chi^2_{ZCH} = \chi^2_Z - \frac{n}{2\bar{n}} \quad ,, \quad \chi^2_{ZHH} = \left(\chi_Z - \frac{\sqrt{n}}{2\bar{n}} \right)^2 \quad (3.29)$$

con \bar{n} por determinar (ver siguiente apartado). Antes, como en lo que sigue, se ha contemplado χ_Z con valor positivo pues sus valores idénticos pero negativos provienen

de los puntos simétricos y \bar{n} se va a determinar de un único modo: el n° de valores distintos del estadístico en puntos con asociación positiva o nula.

Para el caso de χ^2_P o χ^2_R , la (3.8) indica que $A = n^2/(n-1)$ para el primero y que $A = (n-1)$ para el segundo. Con ello:

$$\chi_{PCH}^2 = \chi_P^2 - \frac{n^2}{2(n-1)\bar{n}} \quad \gg \quad \chi_{PHH}^2 = \left(\chi_P - \frac{n}{2\bar{n}\sqrt{n-1}} \right)^2 \quad (3.30)$$

$$\chi_{RCH}^2 = \chi_R^2 - \frac{n-1}{2\bar{n}} \quad \gg \quad \chi_{RHH}^2 = \left(\chi_R - \frac{\sqrt{n-1}}{2\bar{n}} \right)^2 \quad (3.31)$$

Cuando el estadístico es $U = x_1y_2 - x_2y_1$ el asunto varía algo. Ahora $x_1 = x_2 = 0$ da $U=0$ también, y así $m=0$ como antes. Sin embargo el valor máximo de U se alcanzará cuando todas las observaciones estén en una diagonal de la tabla (por ejemplo, cuando $y_1 = x_2 = 0$) y ellas sean lo más iguales posibles ($x_1 \approx y_2$) para que así su producto sea máximo. Cuando n es par esto sucede en $x_1 = y_2 = n/2$, con lo que $M = n^2/4$, compatible con la afirmación de Pirie and Hamdan de que $0 \leq |U| \leq n^2/4$. Sin embargo, cuando n es impar lo más que puede hacerse es $x_1 = [n/2]^- = (n-1)/2$ e $y_2 = [n/2]^+ = (n+1)/2$, con lo que $M = (n-1)(n+1)/4$ y así:

$$\text{Max } U = [n^2/4]^- = \begin{cases} n^2/4 & \text{si } n \text{ es par} \\ (n-1)(n+1)/4 & \text{si } n \text{ es impar} \end{cases} \quad (3.32)$$

con $[x]^-$ aludiendo a la parte entera de x . Con ello por la (3.25) y (3.26):

$$\chi_{ZPH}^2 = \frac{\left\{ |x_1y_2 - x_2y_1| - \frac{[n^2/8]^-}{\bar{n}} \right\}^2}{a_1 a_2 n_1 n_2} \quad (3.33)$$

$$\chi^2_{\text{PPH}} = \frac{\left\{ |x_1 y_2 - x_2 y_1| - \frac{[n^2/8]^-}{\bar{n}} \right\}^2}{a_1 a_2 n_1 n_2} \frac{n^2}{n-1} \quad (3.34)$$

$$\chi^2_{\text{RPH}} = \frac{\left\{ |x_1 y_2 - x_2 y_1| - \frac{[n^2/8]^-}{\bar{n}} \right\}^2}{a_1 a_2 n_1 n_2} (n-1) \quad (3.35)$$

3.2.3. Determinación del número de saltos \bar{n}

3.2.3.1. Introducción

Ya se ha dicho que las c.p.c. anteriores exigen determinar el número $\bar{n} + 1$ de valores posibles del estadístico de trabajo, con \bar{n} el número de saltos del mismo. También se ha dicho que no se van a distinguir las colas, con lo que el número de saltos para χ^2 es el mismo que para χ , y así sólo nos preocuparemos del primero. Como los estadísticos Z, P y R difieren sólo en el factor final n , $n^2/(n-1)$ o $(n-1)$, su valor de \bar{n} será el mismo, y así sólo trabajaremos con χ^2_Z . Por último, ya Haber (1982) y Martín and Silva (1996) indicaron que el estudio teórico de los valores de χ^2_Z es difícil, acudiendo al método reducido de estudiar cuándo alcanza los valores mínimos y considerar para el resto sólo las simetrías (y, en nuestro caso, las equivalencias). Esto da lugar a un valor probablemente falso de \bar{n} , pero el procedimiento fue efectivo en el Modelo II y así será ensayado también en el Modelo I. Esto implica que trabajar con χ^2_Z es lo mismo que trabajar con U (salvo una pequeña diferencia que se advertirá al final), y así se hará en lo que sigue.

En lo que sigue el valor $\bar{n} + 1$ se va a determinar en pasos sucesivos. En cada paso se detectará un valor distinto para \bar{n} , y su último valor será el \bar{n} final. En cada uno de estos pasos se parte de un número N_i de tablas por analizar, siendo las de partida igual al número total de puntos del espacio muestral:

$$N = (n+1)(n+2)(n+3)/6 \quad (3.36)$$

A los efectos de la prueba trabajaremos con el espacio muestral completo (no con el simplificado), el cual se muestra abreviadamente en la Figura 3.1

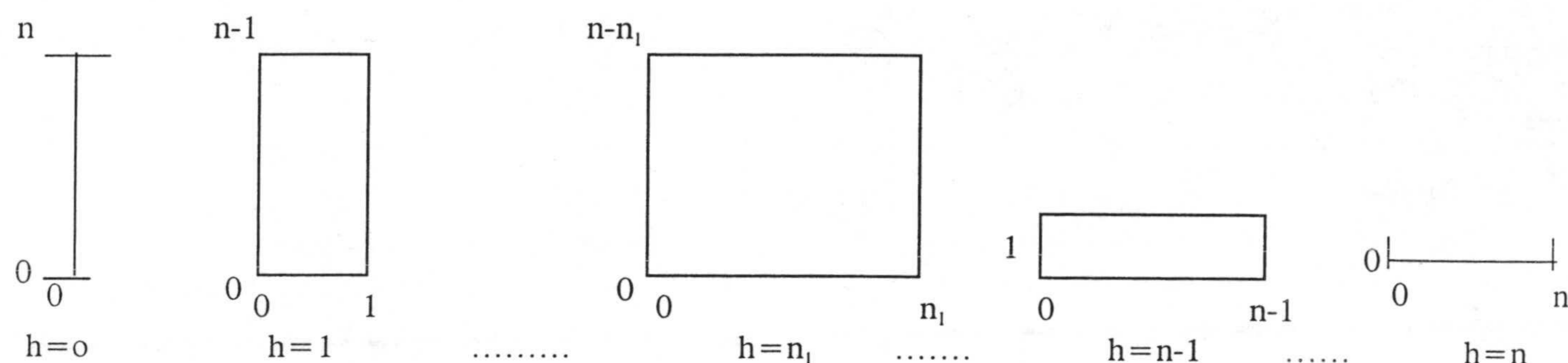


Figura 3.1

Espacio muestral para el Modelo I de tamaño n . Los valores de h son los posibles valores de n_1 , y el rectángulo para cada uno de ellos alude al espacio muestral (tipo Modelo II) que produce.

3.2.3.2. Número de tablas que alcanzan el mínimo (cero)

El valor $U=0$ se alcanza cuando $x_1y_2 - x_2y_1 = x_1n_2 - x_2n_1 = 0$, ecuación diofántica que tiene $\{\text{mcd}(n_1, n_2) + 1\}$ soluciones (Martín and Silva, 1996). Con ello el número N_0 de tablas con valor $U=0$ es:

$$N_0 = \sum_{h=1}^{n-1} \{\text{mcd}(h; n-h) + 1\} + 2(n+1) = \sum_{h=1}^{n-1} \text{mcd}(h; n-h) + (3n+1) = N'_0 + (3n+1) \quad (3.37)$$

en donde el término $2(n+1)$ alude a las tablas con $h=0$ o $h=n$ de la Figura 3.1 (pues en todas ellas $U=0$). El cálculo de N'_0 no se presta a una fórmula fácil. Si r es un divisor de n ($n=r\dot{r}$), la pareja $(r; n-r)$ da $\text{mcd}(r; n-r)=r$ y, a partir de ahí, $(2r; n-2r)$, $(3r; n-3r)$, tienen de mcd al menos r . Con ello N'_0 contiene a la suma de todos los divisores de n menos el último, más todos los divisores de n menos el

penúltimo etc, lo que imposibilita su evaluación por una fórmula sencilla (que es el objetivo). Veamos una alternativa simplificadora:

- * Cuando n es primo $\Rightarrow \text{mcd}(h;n-h)=1$ siempre (por lo indicado arriba) $\Rightarrow N'_0 = n-1 \Rightarrow N_0 = 4n$.
- * Cuando n es par, la tabla $(n/2; n/2)$ da un $\text{mcd} = n/2$. De las $n-2$ restantes tablas, la mitad $(n-2)/2$ son de la forma $(2r; n-2r)$ con $\text{mcd} = 2$ como mínimo \Rightarrow contribuyen a N'_0 con $(n-2)/2$. Con ello $N'_0 \geq n/2 + (n-2) + (n-2)/2 = 2n-3$.
- * Cuando n es de otro tipo (un múltiplo distinto de 2, por ejemplo r) ya sabemos que cada r tablas (menos la de $h=n$) contribuye con un mcd de al menos r y las demás con al menos 1. Lo peor ocurre cuando $n=3$, en cuyo caso $2n/3$ de tablas contribuyen con al menos 1 y $(n/3-1)$ contribuyen con al menos 3, lo que da $N'_0 \geq \{5n-9\}/3$.

Sumando tales valores de N'_0 a la (3.37) se obtienen los valores mínimos para N_0 :

$$N_0 = \begin{cases} 7n+1 & \text{si } n \text{ es primo} \\ 5n-2 & \text{si } n \text{ es par} \\ \frac{14n-6}{3} & \text{en otro caso} \end{cases} \quad (3.38)$$

Con todo ello, actualmente es:

$$\begin{cases} \bar{n} = 0 \\ N_1 = N - N_0 \end{cases} \quad (3.39)$$

pues ya se ha detectado un valor χ^2_Z (el 0), con lo que $\bar{n} + 1 = 1$, y N_1 siendo el número de tablas pendientes de analizar: N y N_0 dadas por las (3.37) y (3.38).

3.2.3.3. Caso de n impar

Si n es primo, las tablas son $N_1 = N - 7n - 1$. De ellas:

- 1) No hay ninguna que tenga una sola tabla equivalente, pues para ello tendría que ocurrir que fuera $n_1 = n_2$ (lo que no es posible por ser n impar).
- 2) Las tablas con $a_1 = n_1$ o $a_1 = n_2$ tienen otra equivalente y dos simétricas y proporcionan igual valor de χ^2_Z . La mitad de ellas están en los casos con

$n_1 < n_2$. Con $a_1 = n_1$ hay $n_1 + 1$ tablas y con $a_1 = n_2$ hay otras $n_1 + 1$ tablas, lo que da $2(n_1 + 1)$ tablas o $4(n_1 + 1)$ considerando que eran la mitad. Como n_1 va de 1 a $(n-1)/2$, entonces el número total de tablas en esas condiciones son:

$$4 \sum_{h=1}^{\frac{n-1}{2}} (h+1) = \frac{(n-1)(n+5)}{2}$$

Esto indica que:

(a) $\bar{n} = (n-1)(n+5)/8$ saltos por ahora

(b) $N_2 = N - 7n - 1 - (n-1)(n+5)/2$ son las tablas pendientes.

- 3) De las N_2 tablas restantes, cada una tiene 3 equivalentes y 4 simétricas $\Rightarrow N_2/8$ es el número de valores distintos de $\chi^2_Z \Rightarrow \bar{n}$ serán los de antes $-(n-1)(n+5)/8$ más los actuales $N_2/8$, y así:

$$\text{Si } n \text{ es impar y primo : } \bar{n} = \frac{(n+1)(n+2)(n+3) + (n+1)(3n-33)}{48} + \frac{1}{4} \quad (3.40)$$

Finalmente, si n no es primo, el número de tablas de partida son $N_1 = N - (14n-6)/3$. Como la prueba anterior sigue valiendo, entonces $N_2 = N_1 - (n-1)(n+5)/2$ y el \bar{n} definitivo será $N_2/8 + (n-1)(n+5)/8$, es decir:

$$\text{Si } n \text{ es impar y no primo : } \bar{n} = \frac{(n+1)(n+2)(n+3) + (n+1)(3n-19)}{48} + \frac{1}{3} \quad (3.41)$$

3.2.3.4. Caso de n par

Por las (3.39) y (3.38), las tablas de partida son ahora $N_1 = N - 5n + 2$. De ellas, y en modo paralelo al apartado anterior:

- 1) La tabla de la Figura 3.1 con $n_1 = n/2$ ($= n_2$) tiene una diagonal $a_1 = n_1$ ($= a_2 = n_2$) en la que todos sus puntos -que son $(n+2)/2$ - no tienen ningún equivalente. Si $n \neq 4$, cada punto tiene un simétrico \Rightarrow las $(n+2)/2$ tablas dan $(n+2)/4$ valores distintos de χ^2_Z . Si $n=4$, hay que descontar la tabla con $x_1 = x_2 = n/4$ que ya fue considerada por dar $\chi^2_Z = 0$, con lo que hay $n/2$ tablas y $n/4$ valores distintos. Con todo ello:

(a) $\bar{n} = n/4$ si $n=4$ o $(n+2)/4$ si $n \neq 4$ son los saltos

(b) $N_2 = N - n/2$ si $n=4$ o $N - (n+2)/2$ si $n \neq 4$ son las tablas pendientes de analizar.

- 2) Este paso es idéntico al del apartado anterior pero para las tablas de la Figura 3.1 que van desde $h=1$ hasta $h=n/2-1=(n-2)/2$, pues el caso $h=n/2$ ya ha sido considerado en 1). Con ello el número total de tablas es ahora:

$$4 \sum_{h=1}^{\frac{n-2}{2}} (h+1) = \frac{(n-2)(n+4)}{2}$$

cuyo valor debe restarse del N_2 del paso anterior y su cuarta parte sumarse al \bar{n} del paso anterior. Con ello:

- (a) El número total de saltos actual es:

$$\begin{cases} \text{Si } n = 4 & : \bar{n} = (n^2 + 4n - 8)/8 \\ \text{Si } n \neq 4 & : \bar{n} = (n^2 + 4n - 4)/8 \end{cases}$$

- (b) El número de tablas pendientes de analizar son:

$$\begin{cases} \text{Si } n = 4 & : N_3 = N - (n^2 + 3n - 8)/2 \\ \text{Si } n \neq 4 & : N_3 = N - (n^2 + 3n - 6)/2 \end{cases}$$

- 3) De las N_3 tablas que restan, cada una tiene 3 equivalentes y 4 simétricas $\Rightarrow N_3/8$ es el número de valores distintos de $\chi^2_Z \Rightarrow$ el \bar{n} de arriba se ve incrementado en $N_3/8$ y así:

$$\begin{cases} \text{Si } n = 4 & : \bar{n} = \frac{(n+1)(n+2)(n+3) + (n+1)(3n+12)}{48} - \frac{3}{4} \\ \text{Si } n \neq 4 & : \bar{n} = \frac{(n+1)(n+2)(n+3) + (n+1)(3n+12)}{48} - \frac{3}{8} \end{cases} \quad (3.42)$$

3.2.3.5. Número de saltos para los estadísticos U y χ^2 -cuadrado

Hasta ahora las deducciones se han efectuado sin distinguir si el estadístico va

a ser U o χ^2 . En realidad, como ya se avisó, los dos casos son exactamente el mismo (y los valores de \bar{n} también) salvo una ligera diferencia. La tabla $x_1=y_1=x_2=0$, de la que hay otra equivalente y dos simétricas (4 en total), tiene un valor $U=0$, pero el valor de χ^2_Z es indeterminado. Un modo de deducir qué valor razonable podría asignársele es hacer $x_1=y_1=x_2=\Delta$, con lo que:

$$\chi^2_Z = \frac{n}{4} \left(\frac{n-4\Delta}{n-2\Delta} \right)^2 \rightarrow \frac{n}{4} \quad \text{si } \Delta \rightarrow 0$$

de aceptarse eso, entonces esas 4 tablas tienen la característica de dar $U=0$ pero $\chi^2_Z \neq 0$, con lo que los valores N_0 de la (3.38) están incrementados en 4 y los valores N_1 de la (3.39), que sirvió de base para toda la prueba, están decrementados en 4. Esas 4 tablas tienen un mismo valor de χ^2_Z y se incorporan al caso 2) de los dos apartados anteriores, lo que ocasiona que el \bar{n} de ellos se incremente en 1 y que los \bar{n} finales también se vean incrementados en la unidad. Con esto, las expresiones (3.40) a (3.42) son apropiadas para U , en tanto que ellas mismas, más la unidad, son apropiadas para χ^2 o χ . Con todo ello y adoptando como fórmula base la (3.40), que es la de menor valor, los valores de \bar{n} para todos los casos vienen sintetizados en la Tabla 3.2, de modo que los valores \bar{n} para U de la misma son los que han de ponerse en las expresiones (3.33) a (3.35), y los \bar{n} para χ^2 o χ son los que han de ponerse en las (3.29) a (3.31).

3.2.3.6. *Discusión y propuestas*

Los valores deducidos para \bar{n} son aproximados. Por un lado serán superiores a los reales pues al calcular N_0 no se consideraron todos los mcd que podían darse, con lo que el N_1 anotado en la (3.39) fue excesivo. Por otro serán inferiores a los reales pues en alguna ocasión el valor de χ^2 o χ o U de una tabla coincidirá con los de otra (que no sea simétrica ni equivalente la primera), y así habrá más empates de los previstos y un menor \bar{n} del previsto. Desde ese punto de vista, los actuales \bar{n} pueden ser un buen punto de equilibrio.

Por otro lado, en los valores \bar{n} de la Tabla 3.2. se ha sido explícito, pues ellos

Tabla 3.2

Valores \bar{n} del número de saltos en función de n y del estadístico utilizado

n		Estadístico	
		$U = x_1 y_2 - x_2 y_1$	χ^2 o χ
Impar	Primo	$\bar{n}_0 + \frac{1}{4}$	$\bar{n}_0 + \frac{5}{4}$
	No primo	$\bar{n}_0 + \frac{7n+15}{24}$	$\bar{n}_0 + \frac{7n+39}{24}$
Par	$=4$	$\bar{n}_0 + \frac{15n+3}{16}$	$\bar{n}_0 + \frac{15n+19}{16}$
	$\neq 4$	$\bar{n}_0 + \frac{15n+9}{16}$	$\bar{n}_0 + \frac{15n+25}{16}$
$\bar{n}_0 = \frac{(n+1)(n+2)(n+3) + (n+1)(3n-33)}{48}$		(3.43)	

son los proporcionados por la teoría, pero probablemente no sea preciso serlo tanto. Para ver esto, consideremos en primer lugar que, puestos a "redondear" el valor de \bar{n} , es preferible poner uno que sea más pequeño que lo real pues, por estar dividiendo a la c.p.c., la c.p.c. será más grande de lo real, el estadístico χ^2 será más pequeño de lo real y el test será más conservador (que es preferible a que sea liberal, para que así respete α). El valor \bar{n} más pequeño (y común a todos) de la Tabla 3.2 es \bar{n}_0 , lo que constituye la primera propuesta simplificadora. La máxima diferencia de \bar{n} con \bar{n}_0 ocurre en el caso de n par, no múltiplo de 4 y con el estadístico χ , en el cual es $\bar{n} = \bar{n}_0 + (15n+25)/16$. ¿Qué ocurre con la c.p.c. de $\sqrt{n}/2\bar{n}$ de la (3.29), por ejemplo, de utilizar el falso \bar{n}_0 al real \bar{n} ? Puede comprobarse que para $n \geq 10$ la c.p.c. basada en \bar{n} difiere de la c.p.c. basada en \bar{n}_0 en menos de 0,01, lo que es irrelevante para las significaciones ordinarias: que χ^2_{ZHH} sea $1,96^2$ o $1,95^2$ es prácticamente igual. Igual sucede si en lugar del citado \bar{n}_0 se coloca este otro valor más simple de recordar

$$\bar{n}_1 = \frac{(n+1)(n-2)(n+10)}{48} \quad (3.44)$$

que es más pequeño aún que \bar{n}_0 y difiere de él en la cantidad $5(n+1)/48$. De ser cierto lo anterior, la propia c.p.c. de $\sqrt{n/2\bar{n}_1}$ es despreciable (es inferior a 0,005) para $n \geq 27$, de modo que de ahí en adelante no sería preciso efectuar corrección por continuidad en el Modelo I. Como para $n < 27$ muy probablemente emplearemos el test exacto (mediante las tablas del Anexo) ello indicaría que la c.p.c. no es preciso efectuarla. Una opción más directa y elemental para \bar{n} , que conviene probar, es:

$$\bar{n}_2 = \frac{(n+1)(n+2)(n+3)}{48} \quad (3.45)$$

consistente en dividir por 8 (el número genérico de equivalencias y simetrías por tabla) el número total de puntos del espacio muestral. Con todo ello, en lo que sigue se estudiará la posible efectividad de las c.p.c. dadas por \bar{n} , \bar{n}_0 , \bar{n}_1 y \bar{n}_2 -Tabla 3.2, (3.43), (3.44) y (3.45) respectivamente- para los **9 nuevos estadísticos** (3.29) a (3.31) y (3.33) a (3.35).

Finalmente cabe observar que las c.p.c. para los métodos XCHx, XHHx y XPHx son, respectivamente, de $A/2\bar{n}$, $A^{1/2}/2\bar{n}$ y $[n^2/8]^{-}/\bar{n}$. Como sus denominadores son de $O(n^3)$ y sus numeradores son, respectivamente, de $O(n)$, $O(n^{1/2})$ y $O(n^2)$, entonces las c.p.c. son de $O(n^{-2})$, $O(n^{-5/2})$ y $O(n^{-1})$. Las dos primeras pueden considerarse por tanto prácticamente cero, pero la última puede aceptarse. Como $[n^2/8]^{-}/\bar{n} \simeq 48O(n^2)/8O(n^3) = 6/O(n) \simeq 6/n$, la propuesta es considerar que $6/n$ es una c.p.c. **promedio asintótica** y así tres nuevos estadísticos serán:

$$\chi^2_{ZPHA} = \frac{\left\{ |x_1 y_2 - x_2 y_1| - \frac{6}{n} \right\}^2}{a_1 a_2 n_1 n_2} n \quad (3.46)$$

$$\chi^2_{\text{PPHA}} = \frac{\left\{ |x_1 y_2 - x_2 y_1| - \frac{6}{n} \right\}^2}{a_1 a_2 n_1 n_2} \frac{n^2}{n-1} \quad (3.47)$$

$$\chi^2_{\text{RPHA}} = \frac{\left\{ |x_1 y_2 - x_2 y_1| - \frac{6}{n} \right\}^2}{a_1 a_2 n_1 n_2} (n-1) \quad (3.48)$$

3.3. EL P-MID DE FISHER COMO MÉTODO ASINTÓTICO

3.3.1. Introducción

Hasta ahora, todos los métodos asintóticos propuestos lo son de origen (aún cuando en el capítulo anterior algunos de ellos se utilizaran tentativamente como ordenación para el exacto), con lo que su estudio actual está plenamente justificado. Sin embargo proponer el P-mid de Fisher como método asintótico es en principio "antinatural", por proceder de un test condicionado y exacto, con lo que su inclusión aquí debe justificarse.

Antes que nada señalar que, a efectos prácticos, la idea no es absurda. Es cierto que la aplicación del P-mid de Fisher requiere del uso de un ordenador (al menos para grandes muestras) pero, mientras que en tales casos el test de Barnard es imposible de aplicar, el de Fisher es accesible con muchos paquetes (STATXACT por ejemplo) y sin ningún problema de tiempo de cómputo. Con todo se mantiene la diferencia de que los tests de tipo chi-cuadrado pueden hacerse a mano, pero el actual requiere de un ordenador.

El P-mid para tablas 2×2 fue propuesto por Plackett, en la discusión de Yates (1984), como medio de quitarle su conservadurismo al test exacto de Fisher cuando se le aplica al Modelo II. La idea fue acogida favorablemente por Barnard (1989), Routledge (1992) y Upton (1992), pero es Haber (1986) el primero que lo propone en firme y lo estudia, si bien hasta ese momento se le acepta como exacto. Hirji et al (1991) hacen un diagnóstico más certero al calificarlo de método "quasi-exacto" y

contemplanlo como una aproximación al de Barnard. Ambos autores y Davis (1993) están de acuerdo con que es conservador, pero que eventualmente puede ser liberal. Haber (1992) entiende que el P-mid de Fisher puede poner paz en la discusión condicionado/incondicionado.

Aquí entendemos que el P-mid de Fisher podría aceptarse como test aproximado para el Modelo I si acaso va bien, y, puesto que en el Modelo II funciona bastante bien, conviene estudiarlo en el contexto actual. Con todo, es esperable que en el Modelo I actúe de modo conservador, pues en el Modelo II actúa correctamente y los valores P de éste son superiores a los de aquel. Con todo, su actuación como ordenación exacta ha sido excelente, lo que de por sí constituye una buena razón para proponerlo también ahora.

3.3.2. Definiciones clásicas

El valor del P-mid de Fisher para una y dos colas ya fue definido en las (2.15) y (2.16). Entonces se les denominó FM1 y FM2 pues constituían dos métodos de ordenación distintos para el test exacto. Ahora ambos son un único método (en sus dos versiones, según el número de colas del test) al que llamaremos por **FM**. Para este método no cabe hablar de c.p.c. pues proviene de cálculos exactos (si bien el 1/2 que incorpora podría interpretarse como tal).

Sin embargo, para el caso de dos colas hay otras posibilidades:

Criterio de Armitage: Hirji et al (1991) contemplan la posibilidad de Armitage de definir el valor P de dos colas como el doble del de una cola;

$$FMA(x_1) = 2 \times FM1(x_1) \quad (3.49)$$

lo que da lugar al método **FMA** que se propone como aproximado.

Criterio de Mantel: Ya se ha citado cómo Mantel (1974), en el contexto del test chi-cuadrado condicionado, defiende que el valor P de dos colas ha de obtenerse en cada cola separadamente. Si esto se hace para el P-mid, él deberá aplicarse para cada cola por separado y luego sumar, con lo que:

$$\text{FMM}(x_1) = \text{FM1}(x_1) + \text{FM1}(x_1') = \sum_{i=r}^{x_1'-1} P(i) + \sum_{i=x_1+1}^s P(i) + \frac{1}{2} \{P(x_1') + P(x_1)\} \quad (3.50)$$

con igual notación que en las (2.15) y (2.16). Esto da lugar al método **FMM**.

3.3.3. Nueva definición*

Ya se ha señalado que, puesto que el P-mid anterior funciona bien en el Modelo II, no es lógico que él mismo (sin ningún cambio) funcione bien también en el Modelo I. ¿Qué cambio convendría efectuar en él? Para verlo, conviene dar antes una justificación teórica del por qué el P-mid funciona bien en el Modelo II (para así luego intentar extenderla al Modelo I).

Omitiendo el condicionamiento en n (que está fijo en todo el razonamiento), una tabla del Modelo II vendrá especificada por $(x_1, x_2 | n_1)$, en tanto que si es del Modelo III lo está por $(x_1 | n_1, a_1)$. La misma tabla, bajo cada modelo, tiene un valor P de una cola de P_{II} y P_{III} , con $P_{II} < P_{III}$ como ya conocemos. Esto implica que el valor de P_{III} hay que hacerlo más pequeño para que sea una buena aproximación de P_{II} . Bajo el Modelo III, la tabla con un P_{III} inmediatamente más pequeño que el de la tabla original $(x_1 | n_1, a_1)$ es la tabla anterior $(x_1 - 1 | n_1, a_1)$ (si nos centramos en el caso de asociaciones negativas). Parece razonable obtener la disminución solicitada de P_{III} mediante el promedio de ambos valores P ; así:

$$P_{II}(x_1, x_2 | n_1) \approx \frac{P_{III}(x_1 | n_1, a_1) + P_{III}(x_1 - 1 | n_1, a_1)}{2} = \text{FM1}(x_1 | n_1, a_1) \quad (3.51)$$

y de ahí que el P-mid vaya bien en el Modelo II.

Para repetir el razonamiento en el caso del Modelo I, hay que hacerlo en dos fases: pasando del Modelo I al II y, luego, del II al III en base a la (3.51). Ahora una tabla viene dada por la terna (x_1, y_1, x_2) , cuyo valor P será $P_I < P_{II}$. Si se conociera la tabla $(x_1^F, x_2^F | n_1)$ que entró antes que la $(x_1, x_2 | n_1)$ en la RC del Modelo II, entonces:

$$P_I(x_1, y_1, x_2) \approx \frac{P_{II}(x_1, x_2 | n_1) + P_{II}(x_1^F, x_2^F | n_1)}{2}$$

Sin embargo, al pasar del Modelo I al II se puede condicionar en filas (en n_1) como antes (de ahí la letra F) o en columnas (en a_1), por lo que si la tabla que entra antes que la $(x_1, y_1 | a_1)$ es la $(x_1^C, y_1^C | a_1)$ entonces también:

$$P_I(x_1, y_1, x_2) \approx \frac{P_{II}(x_1, y_1 | a_1) + P_{II}(x_1^C, y_2^C | a_1)}{2}$$

de modo que por la (3.51):

$$P_I(x_1, y_1, x_2) \approx \frac{FM1(x_1 | n_1, a_1) + FM1(x_1^F | n_1, a_1^F)}{2} \approx \frac{FM1(x_1 | a_1, n_1) + FM1(x_1^C | n_1^C, a_1)}{2} \quad (3.52)$$

Las tablas aludidas por F y C no pueden obtenerse literalmente pues ello implicaría aplicar el test de Barnard para el Modelo II en todos los puntos de ambos espacios muestrales, y el objetivo es llegar al Modelo III. Admitiendo el orden FM1 para el Modelo I, las tablas $(x_1^F, x_2^F | n_1)$ y $(x_1^C, y_1^C | a_1)$ serán las que proporcionen valores $FM1(x_1^F | n_1, a_1^F)$ y $FM1(x_1^C | a_1, n_1^C)$, con $a_1^F, n_1^C = 0, 1, \dots, n$, lo más cercanos posibles, pero inferiores, al $FM1(x_1 | n_1, a_1)$ experimental. Por la (3.52), de ambos es útil el más grande de ellos, y así:

$$P_I(x_1, y_1, x_2) \approx \frac{FM1(x_1 | n_1, a_1) + \text{Max}\{FM1(x_1^F | n_1, a_1^F); FM1(x_1^C | n_1^C, a_1)\}}{2} \quad (3.53 a)$$

ó simbólicamente:

$$P_I(x_1, y_1, x_2) \approx FMN1(x_1 | n_1, a_1) = \frac{FM1(\text{original}) + FM1(\text{más cercano para } n_1 \text{ ó } a_1 \text{ fijos})}{2} \quad (3.53 b)$$

lo que da el nuevo método FMN1 para una cola.

Para dos colas las definiciones serán como en el caso del FM, pero basándose en las (3.53): métodos FMN2, FMNA y FMNM. Los FMN1 y FMN2 dan lugar al único método FMN; FMN1 y los otros dos dan lugar a los métodos FMNA y FMNM.

3.4. SELECCIÓN DEL MÉTODO ASINTÓTICO ÓPTIMO*

3.4.1. Objetivo

A la largo de éste capítulo se han señalado diversos métodos aproximados. A saber:

- 1) Los 3 métodos sin c.p.c.: **Z**, **P** y **R**. Todos ellos son conocidos.
- 2) Los 3 métodos con c.p.c. basada en el salto de **U**: **ZP**, **PP** y **RP**. Todos ellos son también conocidos.
- 3) Los 9 métodos con c.p.c. basados en el salto de **U**, χ^2 o χ : **(Z o P o R)(P o C o H)H** con una terminación extra en $x = 0, 1, 2$ o **A** según el valor \bar{n} asumido. Esto da $9 \times 5 = 45$ métodos nuevos.
- 4) Los nuevos métodos **FM** y **FMN** -con sus versiones extras terminadas en **A** o en **M** para dos colas- basados todos ellos en el **P**-mid. Esto añade 2 métodos nuevos más para una cola y 6 más para dos colas.

En total hay pues 47/51 métodos nuevos para 1/2 colas y 53/57 métodos en total. Para no confundir al lector, los Cuadros II y III los resume. El primero es sólo nemotécnico (pues presenta el origen y el nombre del método); el segundo es funcional (pues presenta la fórmula del método).

Es claro que los métodos nuevos no han sido aún evaluados, en tanto que los viejos o no han sido evaluados o lo fueron aislada y defectuosamente (recordar **3.1.5**), aunque los datos de la literatura nos han permitido descartar otros 10 métodos (los **N**, **F**, **L**, **K**, λ , **CR**, **G**, **Y**, **M** y **C**). Finalmente, el objetivo de esta sección será determinar el método asintótico óptimo (bajo el criterio que se verá).

3.4.2. Criterio de selección

En **3.1.5** ya se vió las distintas estrategias de la literatura tendentes a evaluar cada método aproximado. En esencia todas consistían en lo mismo, aunque sólo haber la efectuaba correctamente: comparar los errores α^* reales de la RC construida por el método asintótico con los errores α objetivo (ver **3.1.5.1**). Esto presenta dos desventajas:

- 1) El investigador se limita a unos pocos valores de α (1% y 5% generalmente)

sin evaluar lo que sucede en los demás.

- 2) El procedimiento permite evaluar lo que le sucede a un método, pero no se proporciona un criterio para efectuar una selección entre métodos.

De hecho, hoy día es frecuente acudir al valor P de un resultado más que a su significación o no a un error α , y el procedimiento de la bibliografía no permite saber cómo de fiable es el valor P aproximado obtenido. Al respecto Martín et al (1992), para el Modelo III, argumentaron que lo relevante es comparar los valores P obtenidos por el método aproximado X (P_X) con los del método exacto (P_B). Lo idóneo es que $P_X \approx P_B$, por lo que la cuestión es ¿cuánta diferencia $|P_X - P_B|$ se permite para admitir que el método X es una buena aproximación del método exacto? Aprovechando el criterio -que lo plantearon para el caso de Fisher- de exigir que $|P_X - P_B| \leq \delta P_B$ con δ del orden del 20% o del 50% para valores de P_B del 5% o del 1% respectivamente (lo que es bastante razonable), tales autores propusieron extenderlo para $1\% \leq P_B < 1\%$ (manteniendo el $\delta = 50\%$) y para $5\% < P_B \leq 10\%$ (manteniendo el $\delta = 20\%$), interpolando los valores de δ para los $1\% < P_B < 5\%$, lo que les permitió afirmar que el método X va bien cuando:

$$|P_X - P_B| \leq \delta P_B \quad (3.54)$$

con:

$$\delta = \begin{cases} 0,5 & \text{si } 1\% \leq P_B \leq 1\% \\ 0,575 - 7,5 P_B & \text{si } 1\% < P_B < 5\% \\ 0,2 & \text{si } 5\% \leq P_B \leq 10\% \end{cases} \quad (3.55)$$

En la regla se omiten las tablas con $P_B < 1\%$ (por ser significaciones demasiado altas) y $P_B > 10\%$ (por ser significaciones demasiado bajas). Cuando sucede la (3.54) diremos que "el método aproximado no falla" y, en otro caso, que "el método aproximado falla". La (3.54) es equivalente a estas otras:

$$-\delta P_B \leq P_X - P_B \leq +\delta P_B \quad ,, \quad (1-\delta)P_B \leq P_X \leq (1+\delta)P_B \quad ,, \quad -\delta \leq \frac{P_X - P_B}{P_B} \leq +\delta \quad (3.56)$$

la última de ellas indicando que son admisibles imprecisiones relativas en el intervalo $(-\delta; +\delta)$.

De todo lo dicho se desprenden dos cosas:

- El método óptimo será el que menos fallos tenga, es decir, el que más veces verifique la (3.54).
- Para evaluar un método interesa también saber cómo actúa cuando falla. Diremos que los "*fallos son conservadores*" cuando P_X supere el límite derecho de la (3.56), en tanto que los "*fallos serán liberales*" cuando P_X sea inferior al límite izquierdo de tal expresión.

Por otro lado Martín and Silva (1996) comprobaron experimentalmente que el criterio para los δ de la (3.55) era demasiado estricto en el caso del Modelo II, pues la chi-cuadrado incondicionada era más imprecisa que la condicionada, viéndose obligados a modificar al alza los valores de δ , aunque siendo menos permisivos cuando el método X era liberal (a fin de no dar demasiadas significaciones falsas). Esto les llevó a proponer dos valores de δ (δ^- y δ^+) en función del tipo de fallo, con lo que la (3.56) pasa a ser:

$$-\delta^- P_B \leq P_X - P_B \leq +\delta^+ P_B \quad ,, \quad (1-\delta^-)P_B \leq P_X \leq (1+\delta^+)P_B \quad ,, \quad -\delta^- \leq \frac{P_X - P_B}{P_B} \leq +\delta^+ \quad (3.57)$$

-la última de ellas indicando que son admisibles imprecisiones relativas en el intervalo $(-\delta^-; +\delta^+)$ - todo lo cual justificaron con apoyos bibliográficos. Las expresiones para los δ^- y δ^+ , similares a la (3.55), son ahora:

$$\delta^-(\delta^+) = \begin{cases} 0,5 & (1) & \text{si } 1^0/00 \leq P_B \leq 1^0/0 \\ 0,55 - 5P_B & (1,15 - 15P_B) & \text{si } 1^0/0 < P_B < 5^0/0 \\ 0,2 + 2P_B & (0,1 + 6P_B) & \text{si } 5^0/0 \leq P_B \leq 10^0/0 \end{cases} \quad (3.58)$$

cuyo significado preciso para los valores P_B de tránsito se da en la Tabla 3.3. De

sucedir aquí igual, ésta sería la segunda opción. Evidentemente, ahora el test exacto es el de Barnard (de ahí el subíndice B que se viene utilizando).

Tabla 3.3

Valores de δ^- y δ^+ para ciertos valores clave de P_B y valores P_X que se consideran una buena aproximación de P_B

P_B	1‰	1%	5%	10%
δ^-	0,5	0,5	0,3	0,4
δ^+	1	1	0,4	0,7
P_X	0,5‰-2‰	5‰-2%	3,5%-7%	6%-17%

Sin embargo es claro que un método aproximado unas veces fallará y otras no —unas veces no verificará la (3.54) y otras sí— y que lo relevante es saber cuándo no falla, lo que dará sus *condiciones de validez* o c.d.v. Determinadas éstas para cada método, entonces el método óptimo será el que tenga unas c.d.v. menos exigentes, pues él será el que puede aplicarse con fiabilidad con mayor frecuencia. Esto define el criterio de óptimo prometido al final del apartado anterior.

3.4.3. Tablas a evaluar para aplicar el criterio

No es esperable que un método aproximado funcione bien en cualquier circunstancia. La actuación del mismo puede variar (en principio) en función de:

- A) El **número de colas** del test (1 o 2), pues recuérdese que en el test asintótico chi-cuadrado el valor P de dos colas es el doble del de una cola, cosa que no sucedía con el test exacto.
- B) El valor P_B , pues la aproximación chi-cuadrado asume la asimetría de la distribución variable base y, no siendo esto cierto en general, es esperable que en los valores bajos de P_B se obtengan mayores discrepancias entre P_X y P_B .
- C) El valor n del tamaño de muestra, pues ya se ha visto la teórica influencia de que éste sea par (múltiplo de 4 o no) o impar (primo o no). También pudiera ser que la propia magnitud de n influyera.
- D) El valor E de la mínima cantidad esperada, por ser éste un parámetro clásico del test chi-cuadrado.

con lo que la selección del método óptimo debe hacerse para cada combinación de los parámetros anteriores, ó, para que el estudio sea factible, para cada combinación de intervalos de ellos. Es por esto que los intervalos/grupos contemplados en este trabajo para cada parámetro son:

$$\begin{array}{ll}
 P_B: & 1\% \leq P_B \leq 1\% \quad ,, \quad 1\% < P_B \leq 10\% \\
 E: & \leq 1,5 \quad ,, \quad 1,5(1)4,5(1,5)9 \quad ,, \quad >9 \\
 Colas: & 1 \quad ,, \quad 2 \\
 n: & \text{Depende de la finalidad}
 \end{array} \tag{3.59}$$

Fijado esto, el modo práctico de proceder será el siguiente:

- 1) Elegir un método aproximado X, el n° de colas del test y un criterio para las δ : el (3.51) o (3.52).
- 2) Elegir un grupo de valores de n y seleccionar todas las tablas con $1\% \leq P_B \leq 10\%$, determinando para ellas el valor P_X y si fallan o no.
- 3) Elegir un intervalo para P_B y para E, determinando el número N de tablas que verifican ambas condiciones.
- 4) De las N tablas, N^0 fallarán por la regla de los δ 's (N^+ por razones conservadoras; N^- por razones liberales), con $N^0 = N^+ + N^-$.
- 5) Determinar los porcentajes de fallos globales (H^0), conservadores (H^+) o liberales (H^-) con $H^0 = H^+ + H^-$ y $H^? = (N^?/N) \times 100$.

Los valores H^0 (para cada gama de valores de E) servirán para evaluar el test y seleccionar el óptimo: aquel en que $H^0 = 0$ con mayor frecuencia. Los valores H^+ y H^- servirán para evaluar en qué sentido se producen tales fallos (conservador o liberal).

3.4.4. Acerca de la globalización en n

Por los trabajos ya citados relativos a los métodos asintóticos en los Modelos III y II, se conoce que la actuación de un método aproximado viene fuertemente afectada por los valores de P_B , n° de colas y E, pero no por el valor de n (cuya

influencia es despreciable), siendo esperable que aquí suceda otro tanto. Ello es además deseable pues sería excesivamente prolijo aconsejar a los investigadores el uso de un método u otro en función, no sólo del número de colas y de P_B , sino también de n . La influencia de n puede darse en dos sentidos (por su magnitud o por su tipo), por lo que en éste apartado se consideran dos agrupaciones posibles de valores de n :

Por su tipo: $n = \{16,32,40,48\}, \{18,34,42,50\}, \{17,31,41,47\}, \{15,33,45,49\}$

Por su magnitud: $n = \{15,16,17,18\}, \{31,32,33,34\}, \{40,41,42,45\}, \{47,48,49,50\}$

la primera de las cuales agrupa a los valores de n que son múltiplos de 4, pares no múltiplos de 4, impares primos o impares no primos (en correspondencia con las posibles situaciones descritas en la Tabla 3.2). Para estudiar lo anterior se hace preciso elegir algún método aproximado de entre los muchos disponibles. El más representativo de todos es el método Z (por ser el más tradicional) con la c.p.c. basada en los saltos "promedio" de U (que fue lo efectivo para el Modelo II), lo que nos lleva a elegir el método ZPH (es decir, el basado en los \bar{n} no simplificados de la primera columna de la Tabla 3.2). Finalmente, a efectos de lo actual es indiferente elegir uno u otro criterio para los δ 's, de modo que en ésta primera aproximación se trabajará con los δ 's clásicos del Modelo III dados por la (3.56).

Las Tablas 3.4 y 3.5 presentan los valores de H^0 para ZPH en las circunstancias en ellas indicadas: la primera con los n agrupados según su tipo; la segunda con los n agrupados según su tamaño. Se observa que, comparando las H^0 de una misma tabla, igual número de colas e iguales gamas de P_B y E según el grupo de n considerado (es decir, H^0 's por columnas), las mismas no varían apreciable y sistemáticamente, de modo que en adelante no habrá inconveniente en globalizar todos los valores de n utilizados.

3.4.5. Acerca de los valores de δ a utilizar y primera selección

Ya se ha reseñado que el criterio de la (3.56) se utilizó para evaluar el Modelo III, pero que con el Modelo II Martín and Silva (1996) se vieron obligados a relajar

Tabla 3.4
Porcentaje de fallos H^0 del método ZPH en función del tipo de n y de la magnitud de E
(δ 's del Modelo III)
(N alude al número de tablas en que se basa H^0)

$n = \{ 16, 32, 40, 48 \} = \text{Múltiplo de 4}$

E	Una cola				Dos colas			
	N	0,001 - 0,01	N	0,01 - 0,1	N	0,001 - 0,01	N	0,01 - 0,1
$\leq 1,5$	115	75	288	60	113	58	213	59
1,5 - 2,5	102	47	214	34	79	37	175	32
2,5 - 3,5	139	57	178	18	98	39	157	22
3,5 - 4,5	88	23	169	14	95	29	144	20
4,5 - 6,0	132	8	184	9	117	10	162	19
6,0 - 7,5	88	8	126	14	83	13	105	19
7,5 - 9,0	48	6	69	16	44	14	63	17
$>9,0$	36	17	50	28	33	9	45	29

$n = \{ 18, 34, 42, 50 \} = \text{Par no múltiplo de 4}$

E	Una cola				Dos colas			
	N	0,001 - 0,01	N	0,01 - 0,1	N	0,001 - 0,01	N	0,01 - 0,1
$\leq 1,5$	124	76	302	58	125	59	229	61
1,5 - 2,5	110	46	244	39	86	40	189	23
2,5 - 3,5	154	56	190	20	106	41	168	26
3,5 - 4,5	85	27	196	10	106	35	165	25
4,5 - 6,0	156	8	207	14	136	12	185	23
6,0 - 7,5	104	4	145	12	93	5	124	31
7,5 - 9,0	67	4	93	15	60	10	86	31
$>9,0$	50	14	68	19	46	11	59	22

$n = \{ 17, 31, 41, 47 \} = \text{Primo}$

E	Una cola				Dos colas			
	N	0,001 - 0,01	N	0,01 - 0,1	N	0,001 - 0,01	N	0,01 - 0,1
$\leq 1,5$	113	76	281	59	113	57	212	61
1,5 - 2,5	103	49	213	36	77	43	169	30
2,5 - 3,5	134	59	175	19	96	38	153	20
3,5 - 4,5	94	14	168	15	90	28	149	19
4,5 - 6,0	127	5	175	8	118	11	154	26
6,0 - 7,5	91	3	132	11	85	5	112	29
7,5 - 9,0	51	2	70	21	46	9	63	24
$>9,0$	27	7	39	23	26	23	34	29

$n = \{ 15, 33, 45, 49 \} = \text{Impar no primo}$

E	Una cola				Dos colas			
	N	0,001 - 0,01	N	0,01 - 0,1	N	0,001 - 0,01	N	0,01 - 0,1
$\leq 1,5$	119	78	302	55	122	61	225	62
1,5 - 2,5	110	40	240	42	886	38	189	25
2,5 - 3,5	156	60	190	10	105	32	173	23
3,5 - 4,5	98	13	191	12	102	35	159	20
4,5 - 6,0	157	3	217	8	144	10	192	27
6,0 - 7,5	104	4	147	15	93	13	129	28
7,5 - 9,0	75	8	104	10	69	3	92	24
$>9,0$	46	7	64	16	43	7	56	29

Tabla 3.5
Porcentaje de fallos H^0 del método ZPH en función de las magnitudes de n y de E
(δ 's del Modelo III)
(N alude al número de tablas en que se basa H^0)

$n = \{ 15, 16, 17, 18 \}$

E	Una cola				Dos colas			
	N	0,001 - 0,01	N	0,01 - 0,1	N	0,001 - 0,01	N	0,01 - 0,1
$\leq 1,5$	26	50	96	54	27	48	70	66
1,5 - 2,5	35	34	42	40	19	47	40	45
2,5 - 3,5	14	21	26	58	16	19	21	57
3,5 - 4,5	3	0	9	22	5	40	7	43
4,5 - 6,0	0	--	0	--	0	--	0	--
6,0 - 7,5	0	--	0	--	0	--	0	--
7,5 - 9,0	0	--	0	--	0	--	0	--
$>9,0$	0	--	0	--	0	--	0	--

$n = \{ 31, 32, 33, 34 \}$

E	Una cola				Dos colas			
	N	0,001 - 0,01	N	0,01 - 0,1	N	0,001 - 0,01	N	0,01 - 0,1
$\leq 1,5$	100	73	257	53	99	59	190	59
1,5 - 2,5	91	38	187	37	62	50	157	37
2,5 - 3,5	116	53	150	27	90	40	130	28
3,5 - 4,5	70	3	135	15	73	27	112	32
4,5 - 6,0	95	4	130	21	87	14	116	30
6,0 - 7,5	45	9	64	30	41	7	54	39
7,5 - 9,0	8	13	10	20	6	17	11	18
$>9,0$	0	--	0	--	0	--	0	--

$n = \{ 40, 41, 42, 45 \}$

E	Una cola				Dos colas			
	N	0,001 - 0,01	N	0,01 - 0,1	N	0,001 - 0,01	N	0,01 - 0,1
$\leq 1,5$	150	77	371	61	153	58	278	60
1,5 - 2,5	135	47	305	39	105	35	241	23
2,5 - 3,5	194	61	238	16	133	37	212	21
3,5 - 4,5	130	22	248	15	133	37	213	24
4,5 - 6,0	196	7	270	12	176	7	236	21
6,0 - 7,5	136	4	190	15	125	14	166	23
7,5 - 9,0	79	9	113	22	75	11	102	29
$>9,0$	33	12	46	4	30	7	39	21

$n = \{ 47, 48, 49, 50 \}$

E	Una cola				Dos colas			
	N	0,001 - 0,01	N	0,01 - 0,1	N	0,001 - 0,01	N	0,01 - 0,1
$\leq 1,5$	195	81	449	59	194	60	341	62
1,5 - 2,5	164	50	377	36	142	37	284	23
2,5 - 3,5	259	60	319	9	166	38	288	19
3,5 - 4,5	181	22	332	9	182	33	285	14
4,5 - 6,0	281	5	383	5	252	12	341	24
6,0 - 7,5	206	4	296	8	188	6	250	27
7,5 - 9,0	154	3	213	11	138	13	191	23
$>9,0$	126	11	175	19	118	13	155	28

las δ 's y a cambiar al criterio de la (3.58) pues, de otro modo, los valores de H^0 en muy pocas circunstancias se anulaban. Esto era un reflejo de la actuación más errática del test chi-cuadrado como incondicionado, y ello mismo vuelve a suceder ahora: las Tablas 3.4 y 3.5 casi nunca contienen un valor $H^0=0$.

Si ahora se repite la Tabla 3.5. (por ejemplo) para los nuevos δ 's de la (3.58), se obtienen los datos de la Tabla 3.6. Ahora se sigue observando que los valores de H^0 no tienen una tendencia sistemática a evolucionar con la magnitud de n (con lo que sigue siendo cierto que se les puede globalizar en n). También se observa que las H^0 han disminuido de modo apreciable con respecto a las de la Tabla 3.5 (como era de esperar), pero, sin embargo, en muy pocas ocasiones se logra que $H^0=0$. De hecho, sorprendentemente, ¡las H^0 son más grandes en $E>9$ que en los intervalos de E inmediatamente anteriores! Como esto puede deberse a que el método seleccionado (el ZPH) actúe peor de lo esperado, conviene ratificarlo para los demás métodos, pero globalizando ya en todos los valores disponibles de n :

$$n = \{15, 16, 17, 18, 31, 32, 33, 34, 40, 41, 42, 45, 47, 48, 49, 50\} \quad (3.60)$$

para lo cual se ha construido la Tabla 3.7. De ella puede concluirse lo siguiente:

- 1) Los métodos basados en el P-mid funcionan en general peor que los demás, por lo que pueden descartarse en lo que sigue. En la tabla no se dan los datos del método FMN pues, a pesar de ser bastante más complejo que el FM, sólo logra mejorarlo ligeramente (pero no competitivamente respecto de los métodos chi-cuadrado); por ejemplo, sus valores H^0 para la Tabla 3.7 (a) son de 39, 27, 29, 6, 1, 2, 1, 3.
- 2) Los métodos de tipo chi-cuadrado con c.p.c. complicada (todos los XXX, XXX0, XXX1 y XXX2) no mejoran la actuación de los que no tienen c.p.c. (los X) o la tienen muy sencilla (los XX y los XXXA), por lo que pueden ser descartados también en lo que sigue.
- 3) Los métodos basados en el poco habitual estadístico P (los P, PP y PPHA) son siempre peores que el resto, y así también serán descartados.

Tabla 3.6
Porcentaje de fallos H^0 del método ZPH en función de las magnitudes de n y de E
(δ 's del Modelo II)
(N alude al número de tablas en que se basa H^0)

$n = \{ 15, 16, 17, 18 \}$

E	Una cola				Dos colas			
	N	0,001 - 0,01	N	0,01 - 0,1	N	0,001 - 0,01	N	0,01 - 0,1
$\leq 1,5$	26	50	96	40	27	30	70	36
1,5 - 2,5	35	14	42	17	19	21	40	10
2,5 - 3,5	14	14	26	31	16	13	21	38
3,5 - 4,5	3	0	9	11	5	40	7	0
4,5 - 6,0	0	--	0	--	0	--	0	--
6,0 - 7,5	0	--	0	--	0	--	0	--
7,5 - 9,0	0	--	0	--	0	--	0	--
$>9,0$	0	--	0	--	0	--	0	--

$n = \{ 31, 32, 33, 34 \}$

E	Una cola				Dos colas			
	N	0,001 - 0,01	N	0,01 - 0,1	N	0,001 - 0,01	N	0,01 - 0,1
$\leq 1,5$	100	73	257	42	99	46	190	39
1,5 - 2,5	91	24	187	10	62	29	157	10
2,5 - 3,5	116	29	150	9	90	4	130	7
3,5 - 4,5	70	1	135	4	73	8	112	13
4,5 - 6,0	95	0	130	4	87	2	116	6
6,0 - 7,5	45	0	64	8	41	2	54	17
7,5 - 9,0	8	0	10	20	6	0	11	9
$>9,0$	0	--	0	--	0	--	0	--

$n = \{ 40, 41, 42, 45 \}$

E	Una cola				Dos colas			
	N	0,001 - 0,01	N	0,01 - 0,1	N	0,001 - 0,01	N	0,01 - 0,1
$\leq 1,5$	150	77	371	46	153	51	278	44
1,5 - 2,5	135	34	305	11	105	24	241	3
2,5 - 3,5	194	38	238	7	133	12	212	3
3,5 - 4,5	130	18	248	4	133	17	213	8
4,5 - 6,0	196	4	270	2	176	3	236	6
6,0 - 7,5	136	1	190	3	125	2	166	7
7,5 - 9,0	79	1	113	5	75	1	102	8
$>9,0$	33	9	46	9	30	3	39	8

$n = \{ 47, 48, 49, 50 \}$

E	Una cola				Dos colas			
	N	0,001 - 0,01	N	0,01 - 0,1	N	0,001 - 0,01	N	0,01 - 0,1
$\leq 1,5$	195	81	449	47	194	55	341	40
1,5 - 2,5	164	42	377	10	142	24	284	5
2,5 - 3,5	259	41	319	2	166	14	288	5
3,5 - 4,5	181	18	332	2	182	13	285	5
4,5 - 6,0	281	2	383	1	252	5	341	3
6,0 - 7,5	206	1	296	1	188	3	250	5
7,5 - 9,0	154	1	213	0	138	0	191	4
$>9,0$	126	3	175	3	118	3	155	8

Tabla 3.7 (a)
 Porcentaje de fallos H^0 de los métodos asintóticos que se aluden
 (δ 's del Modelo II) (n = todos los citados)

UNA COLA

0,001-0,01

E	N	Z	P	R	ZP	PP	RP	ZPHA	PPHA	RPHA	FM
<1,5	471	76	80	73	76	78	71	76	80	72	46
1,5-2,5	425	33	40	28	32	38	28	33	39	29	34
2,5-3,5	583	37	33	42	39	34	44	37	33	43	34
3,5-4,5	384	15	16	14	15	15	14	15	16	14	8
4,5-6	572	2	4	2	2	4	1	2	4	2	2
6-7,5	387	1	2	1	1	2	1	1	2	1	2
7,5-9	241	1	1	0	1	1	0	1	1	0	2
>9	159	4	6	2	4	6	2	4	6	2	3

E	N	ZPH	ZCH	ZHH	PPH	PCH	PHH	RPH	RCH	RHH
<1,5	471	76	76	76	80	80	80	73	73	73
1,5-2,5	425	33	33	33	40	40	40	29	28	29
2,5-3,5	583	37	37	37	33	33	33	43	43	42
3,5-4,5	384	15	15	15	16	16	16	14	14	14
4,5-6	572	2	2	2	4	4	4	2	2	2
6-7,5	387	1	1	1	2	2	2	1	1	1
7,5-9	241	1	1	1	1	1	1	0	0	0
>9	159	4	4	4	6	6	6	2	2	2

E	N	ZPH0	ZCH0	ZHH0	PPH0	PCH0	PHH0	RPH0	RCH0	RHH0
<1,5	471	76	76	76	80	80	80	73	73	73
1,5-2,5	425	33	33	33	40	40	40	29	28	29
2,5-3,5	583	37	37	37	33	33	33	43	43	42
3,5-4,5	384	15	15	15	16	16	16	14	14	14
4,5-6	572	2	2	2	4	4	4	2	2	2
6-7,5	387	1	1	1	2	2	2	1	1	1
7,5-9	241	1	1	1	1	1	1	0	0	0
>9	159	4	4	4	6	6	6	2	2	2

E	N	ZPH1	ZCH1	ZHH1	PPH1	PCH1	PHH1	RPH1	RCH1	RHH1
<1,5	471	76	76	76	80	80	80	73	73	73
1,5-2,5	425	33	33	33	40	40	40	29	28	29
2,5-3,5	583	37	37	37	33	33	33	43	43	42
3,5-4,5	384	15	15	15	16	16	16	14	14	14
4,5-6	572	2	2	2	4	4	4	2	2	2
6-7,5	387	1	1	1	2	2	2	1	1	1
7,5-9	241	1	1	1	1	1	1	0	0	0
>9	159	4	4	4	6	6	6	2	2	2

E	N	ZPH2	ZCH2	ZHH2	PPH2	PCH2	PHH2	RPH2	RCH2	RHH2
<1,5	471	76	76	76	80	80	80	72	73	73
1,5-2,5	425	33	33	33	40	40	40	29	29	29
2,5-3,5	583	37	37	37	33	33	33	43	43	43
3,5-4,5	384	15	15	15	16	16	16	14	14	14
4,5-6	572	2	2	2	4	4	4	2	2	2
6-7,5	387	1	1	1	2	2	2	1	1	1
7,5-9	241	1	1	1	1	1	1	0	0	0
>9	159	4	4	4	6	6	6	2	2	2

Tabla 3.7 (b)
 Porcentaje de fallos H^0 de los métodos asintóticos que se aluden
 (δ 's del Modelo II) (n = todos los citados)

UNA COLA

0,01-0,1

E	N	Z	P	R	ZP	PP	RP	ZPHA	PPHA	RPHA	FM
<1,5	1173	46	48	42	43	47	39	45	48	40	40
1,5-2,5	911	11	15	11	10	13	12	10	14	11	16
2,5-3,5	733	6	10	4	5	8	2	6	9	4	2
3,5-4,5	724	3	5	2	3	4	2	3	4	2	2
4,5-6	783	2	3	1	1	2	1	2	3	1	2
6-7,5	550	2	4	1	2	3	1	2	3	1	2
7,5-9	336	3	3	2	2	3	2	2	3	2	4
>9	221	5	7	5	5	5	5	5	6	5	4

E	N	ZPH	ZCH	ZHH	PPH	PCH	PHH	RPH	RCH	RHH
<1,5	1173	45	45	46	48	48	48	41	41	41
1,5-2,5	911	11	11	11	14	14	14	11	11	11
2,5-3,5	733	6	6	6	9	9	9	4	4	4
3,5-4,5	724	3	3	3	4	4	4	2	2	2
4,5-6	783	2	2	2	3	3	3	1	1	1
6-7,5	550	2	2	2	4	3	4	1	1	1
7,5-9	336	2	2	2	3	3	3	2	2	2
>9	221	5	5	5	6	6	6	5	5	5

E	N	ZPH0	ZCH0	ZHH0	PPH0	PCH0	PHH0	RPH0	RCH0	RHH0
<1,5	1173	45	45	46	48	48	48	41	41	41
1,5-2,5	911	11	11	11	14	14	14	11	11	11
2,5-3,5	733	6	6	6	9	9	9	4	4	4
3,5-4,5	724	3	3	3	4	4	4	2	2	2
4,5-6	783	2	2	2	3	3	3	1	1	1
6-7,5	550	2	2	2	4	3	4	1	1	1
7,5-9	336	2	2	2	3	3	3	2	2	2
>9	221	5	5	5	6	6	6	5	5	5

E	N	ZPH1	ZCH1	ZHH1	PPH1	PCH1	PHH1	RPH1	RCH1	RHH1
<1,5	1173	45	45	46	48	48	48	41	41	41
1,5-2,5	911	11	11	11	14	14	14	11	11	11
2,5-3,5	733	6	6	6	9	9	9	4	4	4
3,5-4,5	724	3	3	3	4	4	4	2	2	2
4,5-6	783	2	2	2	3	3	3	1	1	1
6-7,5	550	2	2	2	4	3	4	1	1	1
7,5-9	336	2	2	2	3	3	3	2	2	2
>9	221	5	5	5	6	6	6	5	5	5

E	N	ZPH2	ZCH2	ZHH2	PPH2	PCH2	PHH2	RPH2	RCH2	RHH2
<1,5	1173	45	45	46	48	48	48	41	41	41
1,5-2,5	911	11	11	11	14	14	14	11	11	11
2,5-3,5	733	6	6	6	9	9	9	4	4	4
3,5-4,5	724	3	3	3	4	4	4	2	2	2
4,5-6	783	2	2	2	3	3	3	1	1	1
6-7,5	550	2	2	2	4	3	4	1	1	1
7,5-9	336	2	2	2	3	3	3	2	2	2
>9	221	5	5	5	6	6	6	5	5	5

Tabla 3.7 (c)
 Porcentaje de fallos H^0 de los métodos asintóticos que se aluden
 (δ 's del Modelo II) (n = todos los citados)

DOS COLAS 0,001-0,01

E	N	Z	P	R	ZP	PP	RP	ZPHA	PPHA	RPHA	FM	FMA	FMM
<1,5	473	50	55	50	50	53	50	51	55	50	100	96	100
1,5-2,5	328	24	23	27	25	24	29	25	23	28	100	89	99
2,5-3,5	405	11	10	17	12	10	19	11	10	18	60	41	51
3,5-4,5	393	13	9	17	14	9	18	14	9	17	17	13	10
4,5-6	515	3	4	3	3	4	3	3	4	3	7	4	5
6-7,5	354	3	3	3	3	3	3	3	3	3	6	6	4
7,5-9	219	0	3	3	0	2	3	0	2	3	6	7	3
>9	148	3	9	3	3	7	3	3	9	3	11	6	9

E	N	ZPH	ZCH	ZHH	PPH	PCH	PHH	RPH	RCH	RHH
<1,5	473	50	50	50	55	55	55	50	50	50
1,5-2,5	328	25	25	25	23	23	23	28	28	28
2,5-3,5	405	11	11	11	10	10	10	18	18	18
3,5-4,5	393	14	14	14	9	9	9	16	16	16
4,5-6	515	3	3	3	4	4	4	3	3	3
6-7,5	354	3	3	3	3	3	3	3	3	3
7,5-9	219	0	0	0	2	2	2	3	3	3
>9	148	3	3	3	9	9	9	3	3	3

E	N	ZPH0	ZCH0	ZHH0	PPH0	PCH0	PHH0	RPH0	RCH0	RHH0
<1,5	473	50	50	50	55	55	55	50	50	50
1,5-2,5	328	25	25	25	23	23	23	28	28	28
2,5-3,5	405	11	11	11	10	10	10	18	18	18
3,5-4,5	393	14	14	14	9	9	9	16	16	16
4,5-6	515	3	3	3	4	4	4	3	3	3
6-7,5	354	3	3	3	3	3	3	3	3	3
7,5-9	219	0	0	0	2	2	2	3	3	3
>9	148	3	3	3	9	9	9	3	3	3

E	N	ZPH1	ZCH1	ZHH1	PPH1	PCH1	PHH1	RPH1	RCH1	RHH1
<1,5	473	50	50	50	55	55	55	50	50	50
1,5-2,5	328	25	25	25	23	23	23	28	28	28
2,5-3,5	405	11	11	11	10	10	10	18	18	18
3,5-4,5	393	14	14	14	9	9	9	16	16	16
4,5-6	515	3	3	3	4	4	4	3	3	3
6-7,5	354	3	3	3	3	3	3	3	3	3
7,5-9	219	0	0	0	2	2	2	3	3	3
>9	148	3	3	3	9	9	9	3	3	3

E	N	ZPH2	ZCH2	ZHH2	PPH2	PCH2	PHH2	RPH2	RCH2	RHH2
<1,5	473	50	50	50	55	55	55	50	50	50
1,5-2,5	328	25	25	25	23	23	23	28	28	28
2,5-3,5	405	11	11	11	10	10	10	18	18	18
3,5-4,5	393	14	14	14	9	9	9	17	17	16
4,5-6	515	3	3	3	4	4	4	3	3	3
6-7,5	354	3	3	3	3	3	3	3	3	3
7,5-9	219	0	0	0	2	2	2	3	1	3
>9	148	3	3	3	9	9	9	3	3	3

Tabla 3.7 (d)
 Porcentaje de fallos H^0 de los métodos asintóticos que se aluden
 (δ 's del Modelo II) (n = todos los citados)

DOS COLAS 0,01-0,1

E	N	Z	P	R	ZP	PP	RP	ZPHA	PPHA	RPHA	FM	FMA	FMM
<1,5	879	40	39	43	42	40	44	41	39	44	99	97	98
1,5-2,5	722	5	8	6	6	7	6	6	7	6	52	37	41
2,5-3,5	651	6	9	6	6	7	5	6	8	6	17	11	16
3,5-4,5	617	7	11	7	7	10	5	7	11	6	15	8	17
4,5-6	693	4	8	3	4	6	3	4	6	3	12	7	12
6-7,5	470	7	11	6	7	9	5	7	10	5	11	6	17
7,5-9	304	6	8	6	6	6	6	6	7	6	12	10	12
>9	194	9	10	8	7	10	8	8	10	8	12	10	18

E	N	ZPH	ZCH	ZHH	PPH	PCH	PHH	RPH	RCH	RHH
<1,5	879	41	41	41	39	39	39	43	43	43
1,5-2,5	722	5	5	5	8	8	7	6	6	6
2,5-3,5	651	6	6	6	8	8	8	6	6	6
3,5-4,5	617	7	7	7	11	11	11	6	6	6
4,5-6	693	4	4	4	7	6	7	3	3	3
6-7,5	470	7	8	7	10	10	10	5	5	6
7,5-9	304	6	6	6	7	7	7	6	6	6
>9	194	8	8	8	10	10	10	8	8	8

E	N	ZPH0	ZCH0	ZHH0	PPH0	PCH0	PHH0	RPH0	RCH0	RHH0
<1,5	879	41	41	41	39	39	39	43	44	43
1,5-2,5	722	5	5	5	8	8	7	6	6	6
2,5-3,5	651	6	6	6	8	8	8	6	6	6
3,5-4,5	617	7	7	7	11	11	11	6	6	6
4,5-6	693	4	4	4	7	6	7	3	3	3
6-7,5	470	7	8	7	10	10	10	5	5	6
7,5-9	304	6	6	6	7	7	7	6	6	6
>9	194	8	8	8	10	10	10	8	8	8

E	N	ZPH1	ZCH1	ZHH1	PPH1	PCH1	PHH1	RPH1	RCH1	RHH1
<1,5	879	41	41	41	39	39	39	43	44	43
1,5-2,5	722	5	5	5	8	8	7	6	6	6
2,5-3,5	651	6	6	6	8	8	8	6	6	6
3,5-4,5	617	7	7	7	11	11	11	6	6	6
4,5-6	693	4	4	4	7	6	7	3	3	3
6-7,5	470	7	8	7	10	10	10	5	5	6
7,5-9	304	6	6	6	7	7	7	6	6	6
>9	194	8	8	8	10	10	10	8	8	8

E	N	ZPH2	ZCH2	ZHH2	PPH2	PCH2	PHH2	RPH2	RCH2	RHH2
<1,5	879	41	41	41	39	39	39	43	44	43
1,5-2,5	722	5	5	5	8	8	7	6	6	6
2,5-3,5	651	6	6	6	8	8	8	6	6	6
3,5-4,5	617	7	7	7	11	11	11	6	6	6
4,5-6	693	4	4	4	7	6	7	3	3	3
6-7,5	470	7	8	7	10	10	10	5	5	5
7,5-9	304	6	6	6	7	7	7	6	6	6
>9	194	8	8	8	10	10	10	8	8	8

- 4) De entre los métodos restantes, los sin c.p.c. (Z y R) son iguales o algo peores que sus homónimos con una c.p.c. sencilla de 0,5 o 6/n (los XP y XPHA), de donde se deduce la conveniencia de efectuar una c.p.c. (aunque sea tan ligera como las indicadas) en la chi-cuadrado para el Modelo I.
- 5) Los métodos restantes son prácticamente iguales, con alguna ligera ventaja para los métodos de tipo R (RP para una cola; RPHA para dos colas).

Como resumen del apartado, en él nos hemos visto obligados a acoger los valores δ del Modelo II (más laxos que los del Modelo III) y a rechazar todos los métodos del tipo P-mid, los de χ^2 basados en c.p.c. complicadas y los de χ^2 no habituales. Los métodos óptimos son por tanto los ZP, ZPHA, RP y RPHA, aunque mantendremos como seleccionados los Z y R por ser habituales en la literatura.

3.4.6. Selección final

Desgraciadamente, en los métodos seleccionados sigue produciéndose la circunstancia desafortunada de que prácticamente nunca ocurre que $H^0=0$, así como que, en los grandes valores de E, cuando crece E también crece H^0 (lo que es contrario a lo que habitualmente se piensa: que el test chi-cuadrado funciona mejor cuanto más grande es E). Para ver de qué tipo son tales fallos, la Tabla 3.8 proporciona los valores H^+/H^- para los 6 métodos seleccionados. Se observa que, para cada situación, los fallos son de los dos tipos, no habiendo valores de H^+/H^- sistemáticamente nulos, y de ahí que el criterio no sirva para intentar decantarse por un método u otro.

Por otro lado, y por si acaso ocurre como en el Modelo II en que la magnitud de H^0 varía dependiendo del desequilibrio marginal, definamos por :

$$K = n_2/n_1 \geq 1 \quad \text{con} \quad a_1 = \text{Min}\{a_1, a_2, n_1, n_2\} \quad \text{y} \quad n_1 = \text{Min}\{n_1, n_2\} \quad (3.61)$$

de modo que K puede entenderse como una medida del "mínimo desequilibrio marginal"; en efecto, si para una tabla dada los desequilibrios marginales son $K_F = \text{Max} n_i / \text{Min} n_i$ (filas) y $K_C = \text{Max} a_j / \text{Min} a_j$ (columnas), entonces $K = \text{Min}\{K_F; K_C\}$. Cada tabla tiene un valor dado de K y así, al igual que con E, conviene agruparlas en intervalos del mismo como por ejemplo los siguientes:

$$K=1 \quad 1 < K \leq 1,5 \quad 1,5 < K \leq 3 \quad K > 3 \quad (3.62)$$

Tabla 3.8
 Porcentaje de fallos liberales (H^-) y conservadores (H^+), de los métodos asintóticos
 que se aluden en función de la magnitud de E
 (δ 's del Modelo II) (n = todos los citados)

Métodos		UNA COLA				(a) 0,001-0,01							
		Z		R		ZP		RP		ZPHA		RPHA	
E	N	H^-	H^+	H^-	H^+	H^-	H^+	H^-	H^+	H^-	H^+	H^-	H^+
<1,5	471	76	0	73	0	76	0	71	0	76	0	72	0
1,5-2,5	425	30	3	23	5	28	4	22	6	30	3	23	6
2,5-3,5	583	6	30	3	38	6	33	3	41	6	31	3	39
3,5-4,5	384	6	9	4	10	5	9	4	10	6	9	4	10
4,5-6	572	2	0	1	0	2	0	1	0	2	0	1	0
6-7,5	387	1	0	0	1	1	0	0	1	1	0	0	1
7,5-9	241	1	0	0	0	1	0	0	0	1	0	0	0
>9	159	4	0	2	0	4	0	1	1	4	0	2	0

Métodos		UNA COLA				(b) 0,01-0,1							
		Z		R		ZP		RP		ZPHA		RPHA	
E	N	H^-	H^+	H^-	H^+	H^-	H^+	H^-	H^+	H^-	H^+	H^-	H^+
<1,5	1173	45	1	41	1	42	1	37	2	44	1	39	1
1,5-2,5	911	10	2	7	4	8	3	6	5	9	2	7	5
2,5-3,5	733	6	0	4	0	5	0	3	0	6	0	4	0
3,5-4,5	724	3	0	2	0	2	0	2	1	3	0	2	0
4,5-6	783	2	0	1	0	1	0	1	0	1	0	1	0
6-7,5	550	2	0	1	0	2	0	1	0	2	0	1	0
7,5-9	336	1	1	1	1	1	1	1	1	1	1	1	1
>9	221	4	0	4	1	4	0	4	2	4	0	4	1

Métodos		DOS COLAS				(c) 0,001-0,01							
		Z		R		ZP		RP		ZPHA		RPHA	
E	N	H^-	H^+	H^-	H^+	H^-	H^+	H^-	H^+	H^-	H^+	H^-	H^+
<1,5	473	49	1	46	4	47	3	44	6	49	2	45	5
1,5-2,5	328	6	18	4	23	5	20	4	25	6	20	4	24
2,5-3,5	405	2	9	1	16	1	10	0	19	1	10	0	16
3,5-4,5	393	2	11	1	16	2	12	1	17	2	12	1	16
4,5-6	515	2	1	1	2	2	1	1	2	2	1	1	2
6-7,5	354	2	1	1	2	2	1	1	2	2	1	1	2
7,5-9	219	0	0	0	3	0	0	0	3	0	0	0	3
>9	148	3	0	2	1	3	0	2	1	3	0	2	1

Métodos		DOS COLAS				(d) 0,01-0,1							
		Z		R		ZP		RP		ZPHA		RPHA	
E	N	H^-	H^+	H^-	H^+	H^-	H^+	H^-	H^+	H^-	H^+	H^-	H^+
<1,5	879	20	21	18	25	18	24	16	28	20	21	17	26
1,5-2,5	722	2	3	1	5	20	4	0	6	2	3	1	5
2,5-3,5	651	4	1	4	2	4	2	3	2	4	1	3	2
3,5-4,5	617	6	1	4	1	6	1	3	2	6	1	4	2
4,5-6	693	4	1	2	2	3	1	1	2	4	1	2	2
6-7,5	470	7	1	4	1	6	1	3	2	7	1	4	1
7,5-9	304	4	2	2	4	3	3	2	4	4	2	2	4
>9	194	6	2	5	4	5	2	4	4	6	2	4	4

es decir, tablas con desequilibrio nulo, ligero, moderado o grande. Con ello, si las N tablas de cada casilla de la Tabla 3.7 se descomponen en 4 grupos según su valor de K y la (3.62), se obtienen los datos de la Tabla 3.9 (que, por razones de espacio, sólo incorpora a los métodos testigo de Z , R , ZP o ZR). Desgraciadamente tal dependencia de K no se observa ahora, y así el recurrir a tal parámetro no mejora las cosas.

Otra posibilidad de "ver" mejor los resultados es incrementar los valores de δ utilizados para que así surjan valores $H^0=0$ (decisión bastante difícil pues los δ 's actuales del Modelo II ya son de por sí elevados). Sin embargo la Tabla 3.10 elimina la idea. En ella se anotan los valores máximos (equivalentes a $+\delta^+$) y mínimos (equivalentes a $-\delta^-$) para $d=(P_X-P_B)\div P_B$. El mínimo d es $d=-1$, lo que implica permitir que P_X sea cero en cualquier P_B . El máximo es $d=5,2$, lo que implica permitir que $P_X=6,2 P_B$. Ambas cosas son disparatadas.

Como consecuencia de todo lo anterior, hemos de limitarnos a los datos de la Tabla 3.7 para efectuar la selección final. Con tal fin, y dado que raramente es $H^0=0$, adoptemos el criterio previo de aceptar como válido un método cuando su valor de H^0 en los intervalos seleccionados de E comienza a ser pequeño (del orden del 5%) o cuando a partir de un cierto E el H^0 global lo es. La Tabla 3.11 contiene los datos relevantes para los métodos seleccionados y ha sido obtenida de la Tabla 3.7. De ella pueden deducirse las siguientes conclusiones generales:

- 1) Los 6 métodos son bastantes similares, actuando peor en los P_B bajos que en los altos (como era de esperar) y en tests de una cola que en los de dos colas (para las significaciones ordinarias).
- 2) Los métodos de tipo Z son algo peores que sus homónimos de tipo R .
- 3) Los métodos X son peores o iguales que los $XPHA$, que a su vez son peores o iguales que los XP , con lo que los métodos sin c.p.c. son peores que los con c.p.c. y la c.p.c. es mejor cuanto más fuerte es.
- 4) El método óptimo es el RP , con unas c.d.v. genéricas de $E > 4,5$ para los P_B bajos o $E > 2,5$ (1,5) para los P_B altos y tests de una cola (dos colas).
- 5) De entre los métodos sin c.p.c., es mejor el R , con unas c.d.v. genéricas iguales a las de arriba.

Tabla 3.9
Porcentaje de fallos H^0 , de los métodos asintóticos que se aluden, en función
de la magnitud de E y de K (el mínimo desequilibrio marginal)
(δ 's del Modelo II) (n = todos los citados)

0,001-0,01						(a) UNA COLA					0,01-0,1				
K=1	N	Z	R	ZP	RP	N	Z	R	ZP	RP	N	Z	R	ZP	RP
E<1,5	0	-	-	-	-	16	0	0	0	0	16	0	0	0	0
1,5-2,5	10	0	10	10	10	14	0	0	0	0	14	0	0	0	0
2,5-3,5	10	20	40	20	40	17	18	12	18	0	17	18	12	18	0
3,5-4,5	11	0	0	0	0	18	6	11	6	6	18	6	11	6	6
4,5-6	19	0	0	0	0	30	3	0	0	0	30	3	0	0	0
6-7,5	20	0	0	0	0	27	4	4	4	4	27	4	4	4	4
7,5-9	18	0	0	0	6	23	9	9	9	9	23	9	9	9	9
>9	24	13	8	13	4	32	16	16	16	16	32	16	16	16	16
1<K≤1,5	N	Z	R	ZP	RP	N	Z	R	ZP	RP	N	Z	R	ZP	RP
E<1,5	0	-	-	-	-	185	0	0	0	2	185	0	0	0	2
1,5-2,5	135	1	4	1	5	193	4	0	2	1	193	4	0	2	1
2,5-3,5	175	26	41	31	46	269	5	18	4	3	269	5	18	4	3
3,5-4,5	164	5	6	5	5	312	1	6	1	1	312	1	6	1	1
4,5-6	346	1	1	1	1	465	2	0	1	1	465	2	0	1	1
6-7,5	307	1	0	1	1	435	2	4	2	1	435	2	4	2	1
7,5-9	221	1	0	1	0	309	2	9	2	2	309	2	9	2	2
>9	135	3	1	3	1	189	3	16	3	4	189	3	16	3	4
1,5<K≤3	N	Z	R	ZP	RP	N	Z	R	ZP	RP	N	Z	R	ZP	RP
E<1,5	141	27	18	26	14	388	32	24	28	20	388	32	24	28	20
1,5-2,5	205	41	33	39	33	525	10	11	10	12	525	10	11	10	12
2,5-3,5	384	42	42	43	44	432	7	5	6	4	432	7	5	6	4
3,5-4,5	209	23	21	22	21	394	4	3	4	3	394	4	3	4	3
4,5-6	207	5	3	5	2	288	2	1	2	1	288	2	1	2	1
6-7,5	60	2	3	2	5	88	2	1	2	0	88	2	1	2	0
7,5-9	2	0	0	0	0	4	0	0	0	0	4	0	0	0	0
>9	0	-	-	-	-	0	-	-	-	-	0	-	-	-	-
K>3	N	Z	R	ZP	RP	N	Z	R	ZP	RP	N	Z	R	ZP	RP
E<1,5	330	97	96	97	95	584	71	68	69	64	584	71	68	69	64
1,5-2,5	75	72	63	72	60	179	26	22	25	25	179	26	22	25	25
2,5-3,5	14	36	36	36	36	15	0	0	0	0	15	0	0	0	0
3,5-4,5	0	-	-	-	-	0	-	-	-	-	0	-	-	-	-
4,5-6	0	-	-	-	-	0	-	-	-	-	0	-	-	-	-
6-7,5	0	-	-	-	-	0	-	-	-	-	0	-	-	-	-
7,5-9	0	-	-	-	-	0	-	-	-	-	0	-	-	-	-
>9	0	-	-	-	-	0	-	-	-	-	0	-	-	-	-

0,001-0,01						(b) DOS COLAS					0,01-0,1				
K=1	N	Z	R	ZP	RP	N	Z	R	ZP	RP	N	Z	R	ZP	RP
E<1,5	0	-	-	-	-	8	0	0	0	0	8	0	0	0	0
1,5-2,5	2	0	50	50	50	14	7	14	7	7	14	7	14	7	7
2,5-3,5	15	33	47	27	47	13	23	23	23	15	13	23	23	23	15
3,5-4,5	5	40	20	40	20	18	11	6	6	6	18	11	6	6	6
4,5-6	18	0	0	0	0	20	10	0	10	0	20	10	0	10	0
6-7,5	18	6	0	6	0	24	13	4	13	0	24	13	4	13	0
7,5-9	14	0	7	0	7	25	8	8	8	8	25	8	8	8	8
>9	23	4	4	4	4	25	12	16	12	16	25	12	16	12	16
1<K≤1,5	N	Z	R	ZP	RP	N	Z	R	ZP	RP	N	Z	R	ZP	RP
E<1,5	0	-	-	-	-	105	42	57	53	61	105	42	57	53	61
1,5-2,5	82	56	66	60	70	187	5	4	5	4	187	5	4	5	4
2,5-3,5	149	13	20	14	21	216	5	3	4	2	216	5	3	4	2
3,5-4,5	159	5	8	5	9	275	4	4	4	4	275	4	4	4	4
4,5-6	314	1	2	1	2	417	4	3	3	3	417	4	3	3	3
6-7,5	279	2	3	2	3	373	7	6	7	5	373	7	6	7	5
7,5-9	203	0	2	0	2	276	5	5	5	6	276	5	5	5	6
>9	125	3	2	3	3	169	8	7	7	7	169	8	7	7	7
1,5<K≤3	N	Z	R	ZP	RP	N	Z	R	ZP	RP	N	Z	R	ZP	RP
E<1,5	140	5	13	8	19	252	29	35	33	39	252	29	35	33	39
1,5-2,5	168	11	14	11	15	406	4	4	4	5	406	4	4	4	5
2,5-3,5	236	9	13	10	16	405	6	6	5	6	405	6	6	5	6
3,5-4,5	229	18	22	19	23	324	10	7	10	6	324	10	7	10	6
4,5-6	183	8	6	8	5	256	5	4	4	4	256	5	4	4	4
6-7,5	57	5	5	5	5	73	7	5	5	5	73	7	5	5	5
7,5-9	2	0	0	0	0	3	0	0	0	0	3	0	0	0	0
>9	0	-	-	-	-	0	-	-	-	-	0	-	-	-	-
K>3	N	Z	R	ZP	RP	N	Z	R	ZP	RP	N	Z	R	ZP	RP
E<1,5	333	69	66	68	63	514	44	45	45	44	514	44	45	45	44
1,5-2,5	76	21	14	18	13	115	12	13	12	15	115	12	13	12	15
2,5-3,5	5	0	0	0	0	17	18	24	24	24	17	18	24	24	24
3,5-4,5	0	-	-	-	-	0	-	-	-	-	0	-	-	-	-
4,5-6	0	-	-	-	-	0	-	-	-	-	0	-	-	-	-
6-7,5	0	-	-	-	-	0	-	-	-	-	0	-	-	-	-
7,5-9	0	-	-	-	-	0	-	-	-	-	0	-	-	-	-
>9	0	-	-	-	-	0	-	-	-	-	0	-	-	-	-

Tabla 3.10
Valores mínimo y máximo de $d = (P_X - P_B) \div P_B$ por el método RP

	UNA COLA		DOS COLAS	
	1‰-1%	1%-10%	1‰-1%	1%-10%
Mínimo	-1	-1	-1	-1
Máximo	+3,2	+1,3	+5,2	+2,3

Tabla 3.11
Mínimo valor de E a partir del cual el método aludido (en las circunstancias que se señalan) comete sólo el % de fallos indicados entre paréntesis (δ 's del Modelo II)

Colas	P_B	Métodos					
		Z	R	ZP	RP	ZPHA	RPHA
1	1‰-1%	4,5 (1,8%)	4,5 (1,4%)	4,5 (1,8%)	4,5 (0,9%)	4,5 (1,8%)	4,5 (1,4%)
	1%-10%	3,5 (2,7%)	2,5 (2,2%)	2,5 (2,8%)	2,5 (1,8%)	3,5 (2,5%)	2,5 (2,2%)
2	1‰-1%	4,5 (2,5%)	4,5 (2,5%)	4,5 (2,5%)	4,5 (3,0%)	4,5 (2,5%)	4,5 (3,0%)
	1%-10%	1,5 (5,9%)	1,5 (6,0%)	1,5 (6,0%)	1,5 (5,1%)	1,5 (6,0%)	1,5 (5,4%)

3.5. CONCLUSIONES*

Dada la gran complejidad de cómputo del test exacto de Barnard, en este capítulo se ha estudiado la posibilidad de utilizar un test aproximado que permita obtener un valor P (P_X) que no difiera excesivamente del valor P exacto (P_B): $|P_X - P_B| \leq \delta P_B$, con δ las del Modelo III o las del Modelo II (menos exigentes que las anteriores).

Los tests aproximados más habituales son de tipo chi-cuadrado, pero una revisión bibliográfica permite eliminar del estudio a todos aquellos que tienen un formato no habitual (los del subapartado 3.1.3.2.), quedándonos así con los habituales Z (con factor n) y R (con factor n-1), y con el menos habitual P (con factor $n^2/(n-1)$): expresiones (3.5), (3.7) y (3.6). Cada uno de ellos puede ser utilizado con o sin c.p.c., y así en el capítulo se han propuesto hasta 48 métodos distintos con c.p.c. (según que se contemplen los saltos de χ^2 , de χ o del numerador de χ). De modo complementario, se ha acogido también la propuesta de utilizar el P-mid como método aproximado.

En una primera fase se ha comprobado que, para el análisis de la actuación de los diferentes métodos, puede globalizarse en varios valores de n y que conviene manejar los más permisivos valores δ del Modelo II (para que las comparaciones entre métodos sean más claras). Bajo tales premisas se ha comprobado que los métodos

óptimos son los de tipo X, XP y XPHA, con $X=Z$ o R, pues los demás son peores o iguales que estos y más costosos de cálculo.

En la fase final del capítulo se ha deducido que los métodos óptimos son el R (de entre los sin c.p.c.) y el RP (de entre todos), ambos con unas c.d.v. genéricas de E (mínima cantidad esperada) $>4,5$ para $1‰ \leq P_E < 1\%$ y $E > 2,5$ (ó 1,5) para $1\% \leq P_E \leq 10\%$ y tests de una (ó de dos colas). Sin embargo, al contrario que sucede con el Modelo II, tales c.d.v. no garantizan que el método no falle, sino sólo que lo hace menos del 5% (2%) aproximadamente de las veces en tests de dos (una) cola. La razón de ello es que ahora el test aproximado para el actual Modelo I funciona más imprecisamente que con el Modelo II (al igual que con el Modelo II funcionaba más imprecisamente que con el Modelo III, y de ahí que se tuvieron que modificar sus valores δ). En consecuencia, para lograr que los % de fallos sean cero haría falta modificar los δ 's al alza (lo que es improcedente pues los actuales del Modelo II ya están de por sí bastante elevados), aumentar el tamaño de n para que así surjan valores más elevados de E (lo que es imposible por la capacidad actual de cómputo), o limitarse a tablas con valores de P_B más altos del 10% (lo que no es conveniente pues tales significaciones no suelen ser de interés).

La conclusión final obtenida es que el mejor método es el RP, con las c.d.v. señaladas arriba pero que el mismo puede fallar entre un 1% y un 5% de las veces (por razones generalmente liberales en el caso del test de una cola).

CAPÍTULO IV

SOBRE LAS CONDICIONES DE VALIDEZ DEL TEST CHI-CUADRADO

El test chi-cuadrado es el método asintótico más conocido para contrastar la independencia en una Tabla 2×2 , y ello con independencia de cuál sea el Modelo estadístico asumido. En el caso actual del Modelo I, el Capítulo anterior se ha dedicado a estudiar cuál es la c.p.c. óptima, entendiendo por tal aquella que con más frecuencia ocasiona que el valor P aproximado (P_A) de una tabla caiga dentro de un entorno fijado del valor P exacto (P_E). Sin embargo el test chi-cuadrado no es universalmente válido, sino que ha de estar sometido a unas c.d.v. que son las que garantizan que los valores de P_A y P_E sean aproximadamente iguales. Hasta ahora las c.d.v. (basadas exclusivamente en la mínima cantidad esperada E) han sido tratadas de un modo muy general, pues el fin primordial era seleccionar la c.p.c. óptima, pero en el Capítulo actual se pretende estudiarlas con algo más de detalle.

Ya se ha visto (ver 3.1.5.4) cómo todos los autores están de acuerdo en que la c.d.v. apropiada es del tipo $E > E^*$, con E^* un número por determinar (que tradicionalmente se han entendido que es de 5 (Fisher, 1925), lo que ha sido admitido sin más en todo lo anterior. Pero, ¿ E^* es un número o una función?, es decir ¿es constante o es variable? De ser variable, ¿de qué depende? En cualquier caso, ¿hay algún argumento teórico que justifique su forma o su valor? De existir éste, ¿puede aprovecharse para, a través de un estudio teórico-experimental, estimar E^* para que así las c.d.v. puedan extrapolarse con ciertas garantías más allá de los pequeños valores de n con que ha sido factible trabajar?

A estos efectos la literatura siempre ha considerado que E^* es una constante, pero Martín and Herranz (1999), con referencia al Modelo III, comprueban que esto no es así, sino que E^* depende de los desequilibrios marginales y del tamaño total. Dado que este

es el único trabajo que aborda tal problema y que muchos de sus conceptos y demostraciones pueden ser aprovechados para el actual Modelo I, la primera sección del actual Capítulo resume dicho trabajo. Todo ello en la esperanza de que lo que funciona para el Modelo III también funcione en el Modelo I (como es esperable por la íntima relación existente entre los tres modelos de Tablas 2×2), lo que será puesto de manifiesto en la segunda sección de este Capítulo.

4.1. EL VALOR DE E^* EN EL MODELO III

4.1.1. Introducción

Martín, Herranz and Luna (1992) y Martín and Herranz (1997) ya observaron que en el caso del Modelo III (el condicionado clásico, que es el más frecuente en la literatura) la clásica condición de $E > 5$ o $E \geq 5$ podía ser demasiado liberal o demasiado estricta según el caso, lo que da indicios de que el valor E^* no es un único número, sino una función por determinar. El trabajo de Martín and Herranz (1999) es el único existente acerca de este asunto, y todo lo que sigue de esta sección alude a él (al cual queda remitido el lector para más detalles).

4.1.2. Valores experimentales de E^*

A fin de no incurrir en repeticiones, y dado que la mínima cantidad esperada E es la clave, conviene asegurarnos de que E se corresponde con la cantidad esperada de la casilla (1,1) con tal fin, cualquier tabla del Modelo III se reordenará de modo que:

$$a_1 = \text{Min} (a_1, a_2, n_1, n_2) \text{ y } n_1 = \text{Min} (n_1, n_2) \quad (4.1)$$

con lo cual $E = a_1 n_1 / n$ y el espacio muestral vendrá dado por:

$$0 \leq x_1 \leq a_1 \quad (4.2)$$

pues $\text{Max} \{0; a_1 - n_2\} = 0$ y $\text{Min} (a_1, n_1) = a_1$. Cada terna (a_1, n_1, n) determina el espacio muestral (4.2) con probabilidades $P(x_1)$ dadas por la (1.9) y correspondientes a la distribución hipergeométrica.

Fijados los valores de una pareja (n_1, n) , el valor clave para determinar la c.d.v. de $E > E^*$ es a_1 (por ser el menor), de modo que será $E = a_1 n_1 / n > E^*$ cuando sea $a_1 >$

$E^*n/n_1 = a_1^*$ o, si hacemos $K = n_2/n_1$, cuando sea $a_1 > a_1^* = E^*(K+1)$. Esto implica que para $(n_1, n) \equiv (n_1, K)$ dados, el objetivo es determinar un entero a_1^* a partir del cual ($a_1 > a_1^*$) el test chi-cuadrado sea válido. Determinado a_1^* , el valor de E^* es inmediato: $E^* = a_1^*/(K+1)$. Los valores (n_1, K) para los que Martín and Herranz determinaron a_1^* fueron:

$$\begin{aligned} n_1 &= 40, 60, 100 (50) 250 \\ K &= 1 (0,1) 2 (0,5) 4, 5, 6 \end{aligned} \tag{4.3}$$

y ellos se dan en la Tabla 4.1 (aunque para el test de dos colas añadieron algunos valores extras de K por conveniencia experimental). Pero para su obtención es preciso indicar qué se entiende al afirmar que “el test chi-cuadrado es válido para $a_1 > a_1^*$ ”.

Dada una tabla específica $(x_1 | a_1, n_1, n)$, en ella el test exacto de Fisher dará un valor P de P_E , en tanto que el test aproximado de la chi-cuadrado que se utilice proporcionará un valor P de P_A . Diremos que el test chi-cuadrado “falla” para esa tabla si:

$$|P_E - P_A| > \delta P_E \tag{4.4}$$

con:

$$\delta = \begin{cases} 0,5 & \text{si } P_E \leq 1\% \\ 0,575 - 7,5 P_E & \text{si } 1\% < P_E < 5\% \\ 0,2 & \text{si } P_E \geq 5\% \end{cases} \tag{4.5}$$

lo que es una extensión del criterio de Cochran (1954) de aceptar imprecisiones relativas del 20% o del 50% para valores $P_E = 5\%$ o 1% respectivamente (y que en el Capítulo III fueron aludidos como δ 's del Modelo III). Cuando el test chi-cuadrado “no falla” para esa tabla es porque $P_A \approx P_E$, es decir, los valores P exactos y aproximados difieren en una cantidad pequeña (menor o igual que δP_E). Finalmente, diremos que el test chi-cuadrado “es válido” para un espacio muestral (a_1, n_1, n) si “no falla” en ninguna de las tablas del mismo –las dadas por la (4.2)– que tengan un P_A entre el 1% y el 10%. La limitación a tales valores de P_A se debe a que las c.d.v. varían con el error α manejado (lo que es conocido de antiguo; de hecho la condición $E > 5$ suele decirse que es válida para $\alpha = 5\%$). Tal limitación se basa en P_A , más que en P_E , a fin de que el investigador pueda saber cuándo la c.d.v. le es de aplicación a su tabla particular. Pero para aplicar lo

Tabla 4.1

Valores mínimos reales (a_1^* y E^*) del marginal más pequeño (a_1) y de la mínima cantidad esperada (E) a partir de los cuales el test chi-cuadrado es válido (como aproximación del test exacto de Fisher) y predicciones conservadoras \hat{a}_1^* del primero. Se asume que $K=n_2/n_1 \geq 1$.

n_1	K	Una cola			Dos colas			n_1	K	Una cola			Dos colas		
		a_1^*	\hat{a}_1^*	E^*	a_1^*	\hat{a}_1^*	E^*			a_1^*	\hat{a}_1^*	E^*	a_1^*	\hat{a}_1^*	E^*
40	1	3	8	1.5	7	8	3.5	150	1	3	10	1.5	8	10	4.0
40	1.1	3	8	1.4	7	8	3.3	150	1.1	3	10	1.4	8	10	3.8
40	1.2	3	9	1.4	8	8	3.6	150	1.2	8	11	3.6	8	10	3.6
40	1.3	3	9	1.3	3	9	1.3	150	1.3	3	12	1.3	7	11	3.0
40	1.4	3	10	1.2	9	9	3.8	150	1.4	9	13	3.8	10	11	4.2
40	1.5	9	11	3.6	9	9	3.6	150	1.5	10	15	4.0	9	12	3.6
40	1.6	10	12	3.8	3	10	1.2	150	1.6	10	16	3.8	11	12	4.2
40	1.7	10	13	3.7	11	11	4.1	150	1.7	15	18	5.6	11	13	4.1
40	1.8	11	15	3.9	10	11	3.6	150	1.8	16	20	5.7	10	14	3.6
40	1.9	16	16	5.5	12	12	4.1	150	1.9	17	22	5.9	12	15	4.1
40	2	12	17	4.0	3	12	1.0	150	2	18	24	6.0	14	16	4.7
40	2.5	20	23	5.7	16	16	4.6	150	2.5	32	34	9.1	16	20	4.6
40	2.6	--	--	--	2	16	0.6	150	2.6	--	--	--	2	21	0.6
40	2.7	--	--	--	17	17	4.6	150	2.7	--	--	--	17	22	4.6
40	2.8	--	--	--	2	18	0.5	150	2.8	--	--	--	2	23	0.5
40	2.9	--	--	--	18	18	4.6	150	2.9	--	--	--	20	24	5.1
40	3	24	29	6.0	2	18	0.5	150	3	43	45	10.8	2	23	0.5
40	3.5	33	35	7.3	2	20	0.4	150	3.5	49	56	10.9	23	26	5.1
40	4	38	40	7.6	23	23	4.6	150	4	62	67	12.4	4	29	0.8
40	5	38	40	6.3	19	27	3.2	150	5	84	89	14.0	9	36	1.5
40	6	40	40	5.7	32	32	4.6	150	6	108	111	15.4	39	42	5.6
60	1	3	8	1.5	8	9	4.0	200	1	3	10	1.5	8	10	4.0
60	1.1	3	9	1.4	8	9	3.8	200	1.1	3	10	1.4	8	10	3.8
60	1.2	3	9	1.4	8	9	3.6	200	1.2	8	11	3.6	10	10	4.5
60	1.3	3	10	1.3	3	9	1.3	200	1.3	3	12	1.3	10	11	4.3
60	1.4	9	11	3.8	9	10	3.8	200	1.4	9	14	3.8	10	11	4.2
60	1.5	9	13	3.6	9	10	3.6	200	1.5	10	15	4.0	9	12	3.6
60	1.6	10	14	3.8	11	11	4.2	200	1.6	15	17	5.8	11	13	4.2
60	1.7	10	15	3.7	11	12	4.1	200	1.7	15	19	5.6	11	14	4.1
60	1.8	11	17	3.9	10	12	3.6	200	1.8	16	21	5.7	10	14	3.6
60	1.9	16	18	5.5	12	13	4.1	200	1.9	17	23	5.9	12	15	4.1
60	2	17	20	5.7	3	14	1.0	200	2	18	25	6.0	14	16	4.7
60	2.5	26	27	7.4	16	17	4.6	200	2.5	32	36	9.1	16	21	4.6
60	2.6	--	--	--	2	18	0.6	200	2.6	--	--	--	2	22	0.6
60	2.7	--	--	--	17	19	4.6	200	2.7	--	--	--	19	23	5.1
60	2.8	--	--	--	2	20	0.5	200	2.8	--	--	--	2	24	0.5
60	2.9	--	--	--	18	20	4.6	200	2.9	--	--	--	20	25	5.1
60	3	30	34	7.5	2	20	0.5	200	3	43	48	10.8	2	24	0.5
60	3.5	41	42	9.1	23	23	5.1	200	3.5	56	60	12.4	23	27	5.1
60	4	46	49	9.2	23	25	4.6	200	4	63	72	12.6	4	30	0.8
60	5	56	60	9.3	9	30	1.5	200	5	93	96	15.5	9	37	1.5
60	6	57	60	8.1	32	36	4.6	200	6	109	119	15.6	39	43	5.6
100	1	3	9	1.5	8	9	4.0	250	1	3	10	1.5	8	10	4.0
100	1.1	3	10	1.4	8	9	3.8	250	1.1	3	10	1.4	8	10	3.8
100	1.2	8	10	3.6	8	10	3.6	250	1.2	8	11	3.6	10	11	4.5
100	1.3	3	11	1.3	3	10	1.3	250	1.3	8	12	3.5	10	11	4.3
100	1.4	9	12	3.8	10	11	4.2	250	1.4	9	14	3.8	10	12	4.2
100	1.5	9	14	3.6	9	11	3.6	250	1.5	10	15	4.0	9	12	3.6
100	1.6	10	15	3.8	11	12	4.2	250	1.6	15	17	5.8	11	13	4.2
100	1.7	15	17	5.6	11	13	4.1	250	1.7	15	19	5.6	11	14	4.1
100	1.8	16	19	5.7	10	13	3.6	250	1.8	21	21	7.5	13	15	4.6
100	1.9	17	20	5.9	12	14	4.1	250	1.9	17	23	5.9	12	15	4.1
100	2	17	22	5.7	14	15	4.7	250	2	18	26	6.0	14	16	4.7
100	2.5	32	32	9.1	16	19	4.6	250	2.5	32	37	9.1	16	21	4.6
100	2.6	--	--	--	2	20	0.6	250	2.6	--	--	--	2	22	0.6
100	2.7	--	--	--	16	21	4.3	250	2.7	--	--	--	19	23	5.1
100	2.8	--	--	--	2	22	0.5	250	2.8	--	--	--	2	24	0.5
100	2.9	--	--	--	18	23	4.6	250	2.9	--	--	--	20	25	5.1
100	3	37	41	9.2	2	22	0.5	250	3	44	50	11.0	2	25	0.5
100	3.5	48	50	10.7	23	25	5.1	250	3.5	56	62	12.4	23	28	5.1
100	4	54	60	10.8	23	28	4.6	250	4	70	75	14.0	4	31	0.8
100	5	75	78	12.5	9	34	1.5	250	5	93	100	15.5	9	37	1.5
100	6	97	97	13.9	39	40	5.6	250	6	119	125	17.0	25	44	3.6

anterior es preciso indicar cómo se obtienen P_E y P_A .

El valor de P_E para una tabla dada se obtiene mediante el test exacto de Fisher ya indicado en 1.2.1. Los valores P de una y dos colas son por tanto los de (1.10) y (1.11) para una tabla con $x_1 > E$.

Para el valor P_A es preciso elegir una de las muchas versiones de test chi-cuadrado existentes para el Modelo III. Tales autores optaron por los test óptimos para una o dos colas de su artículo de 1997. Ellos se basan en los estadísticos:

$$\chi^2_Y = \frac{\left\{ |x_1 y_2 - x_2 y_1| - \frac{n}{2} \right\}^2}{a_1 a_2 n_1 n_2} n \tag{4.6}$$

$$\chi^2_C = \chi^2_Y + \frac{n^3}{4a_1 a_2 n_1 n_2} \tag{4.7}$$

de modo que si:

$$P(\chi^2_X) = P\{\chi^2_{1 \text{ g.l.}} \geq \chi^2_X\} \tag{4.8}$$

entonces el valor P_A es:

$$P_A = \begin{cases} P(\chi^2_C)/2 & \text{para una cola} \\ \{P(\chi^2_Y) + P(\chi^2_{Y'})\}/2 & \text{para dos colas} \end{cases} \tag{4.9}$$

con χ^2_Y aludiendo al valor de la (4.6) en la tabla con $x_1 = x_1' = [2E - x_1]^-$, valor sobre el que, por comparabilidad de resultados, se aplica también la (1.11).

Con todo ello, hay espacios muestrales (a_1, n_1, n) en los que el test chi-cuadrado SÍ es válido y espacios en que NO es válido, indicando la Tabla 4.1 los valores de a_1 ($a_1 > a_1^*$), para unos (n_1, K) fijados, a partir del cual el test SÍ es válido. Se observa que:

- 1) Para n_1 constante, los valores de a_1^* aumentan en general con K , aunque como por construcción es $a_1 \leq n_1$ el crecimiento no puede mantenerse indefinidamente. Para dos colas el crecimiento es más lento e irregular, observándose caídas bruscas en varias ocasiones (de ahí los valores extra de K incluidos en este caso). Algo similar ocurre para K constante: a_1^* aumenta con n_1 en general.
- 2) Para $K = 1$ es $a_1^* = 3$ en una cola, y 7 u 8 en dos colas, lo que, junto con lo

anterior, indica que las c.d.v. son más exigentes en dos colas que en una cola en los valores pequeños de K , pero menos exigentes en los moderados o grandes.

- 3) Los valores de E^* son bastante diferentes en unos casos que en otros y, especialmente en una cola, con frecuencia son bastante mayores que 5 (el ≤ 5 se da en los pequeños valores de K).

En conclusión, los datos experimentales dejan bastante claro que E^* no es una constante, sino que depende de que el test sea de una o dos colas y de los valores de n_1 y de K , de modo que, fijado lo primero, E^* es una función $E^*(n_1, K)$ o, alternativamente, a_1^* es una función $a_1^*(n_1, K)$. ¿Qué función es esta?

4.1.3. La condición de asimetría acotada y la condición de validez

Martín and Herranz argumentaron que, puesto que el test chi-cuadrado actúa aproximando la distribución hipergeométrica a través de la distribución normal, tal aproximación debe ir mal cuando la asimetría de aquella (el cuadrado, β_1 , de su coeficiente de asimetría) está demasiado alejada de la de esta (que es de 0), es decir cuando $\beta_1 > B$, con $B > 0$ una constante por determinar. Como según Johnson and Kotz (1969, Cap. 6) es:

$$\beta_1 = \frac{(n_2 - n_1)^2 (a_2 - a_1)^2 (n - 1)}{a_1 a_2 n_1 n_2 (n - 2)^2} \quad (4.10)$$

entonces, definiendo por:

$$f(x) = \frac{(n - 2x)^2}{x(n - x)}, \quad C = f(1) = \frac{(n - 2)^2}{n - 1} \quad (4.11)$$

a una función que mide el desequilibrio del marginal de valores $(x, n - x)$ —en $x = 1$ el desequilibrio es máximo; en $x = n/2$ el equilibrio es total—, la aproximación irá mal cuando

$$\beta_1 = \frac{f(a_1) f(n_1)}{C} \geq B \quad (4.12)$$

Esta suposición es concordante con el análisis de Boos and Hughes-Oliver (2000) acerca del papel de la asimetría en los intervalos de confianza aproximados para una media. Sin

embargo, en $K = 1$ es $f(n_1) = 0 \Rightarrow \beta_1 = 0 < B$ seguro, lo que es contrario a los datos experimentales de que en $K = 1$ el test no siempre es válido. Para hacer compatible ambas cosas basta incrementar $f(n_1)$ en una constante A (por determinar), con lo que el test aproximado irá mal si

$$\frac{\{f(n_1) + A\} f(a_1)}{C} \geq B \tag{4.13}$$

o, equivalentemente, si:

$$y \leq a + bx \tag{4.14}$$

con:

$$y = C/f(x) \quad ,, \quad x = f(n_1) = (K - 1)^2/K \quad ,, \quad a = A/B \quad ,, \quad b = 1/B \tag{4.15}$$

Como x depende sólo de K , la igualdad de (4.14) se alcanzará, para valores fijados de K , en un valor

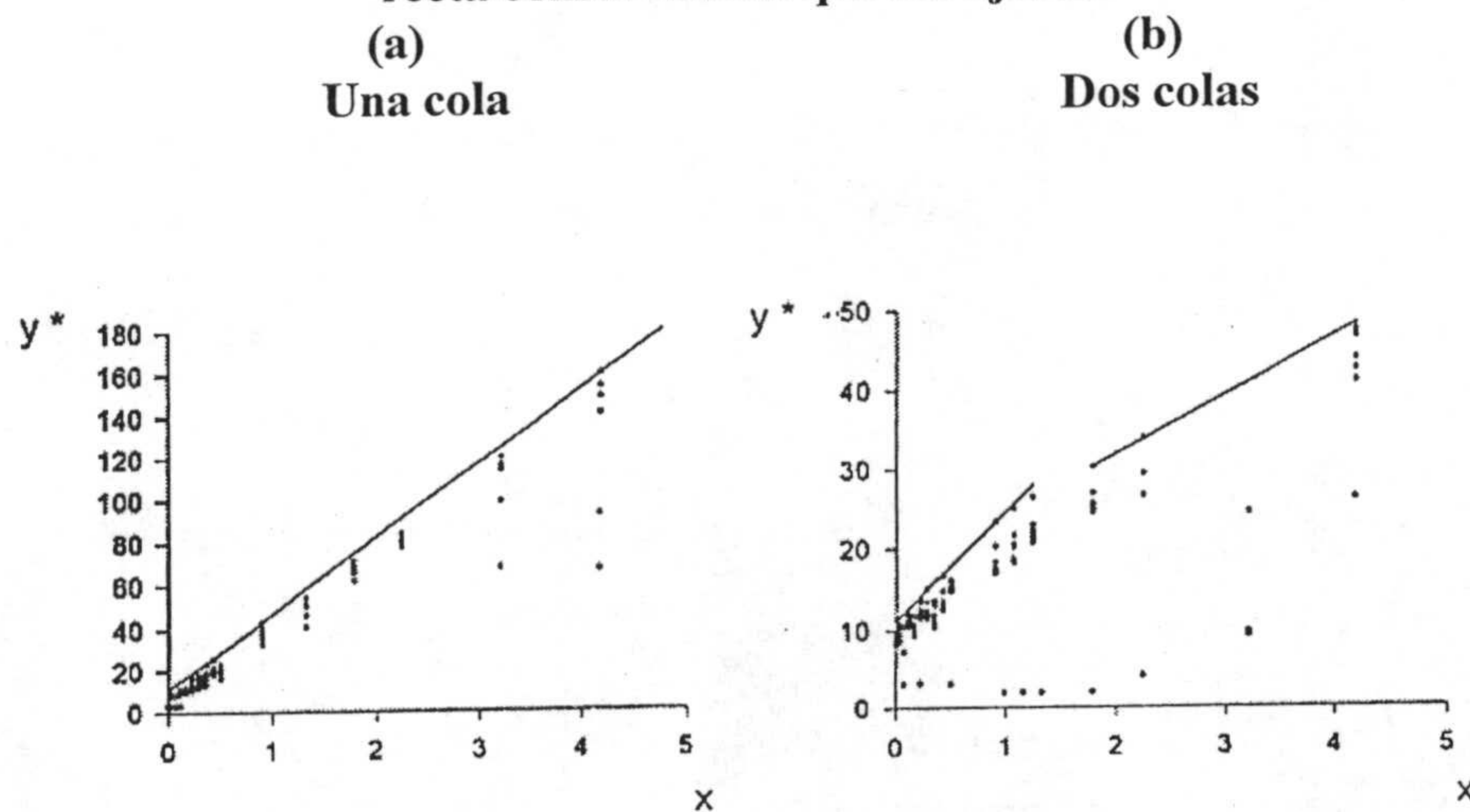
$$y^* = \ddot{a} + bx \tag{4.16}$$

que experimentalmente viene dado por:

$$y^* = C/f(a_1^*) \tag{4.17}$$

de modo que el test aproximado será válido cuando $y > y^*$ (pues en a_1^* es la primera vez en que el test chi-cuadrado no es válido).

Figura 4.1
Nube de puntos para los valores de $(x; y^*)$ del Modelo III –Tabla 4.1– y recta conservadora que los ajusta.



En conformidad con todo lo anterior, lo que habrá que hacer es ajustar la recta de regresión (4.16) a los puntos experimentales (x, y^*) de la Tabla 4.1 para así estimar

a y b (y, en consecuencia, A y B). La Figura 4.1 presenta las dos nubes de puntos (x, y*) para los tests de una y de dos colas. Para el caso de una cola, la recta de regresión mínimo cuadrática da una razón de correlación entre a_1^* y \hat{a}_1^* (sus predicciones) de 0,99, lo que avala la importancia de la asimetría a los efectos actuales (Martín and Herranz justifican lo que sucede con los puntos de peor ajuste).

Sin embargo, en la práctica el investigador no desea unas c.d.v. que vayan bien en promedio, sino unas c.d.v. que garanticen que puede aplicarse el test chi-cuadrado a esa tabla en particular con cierta seguridad, lo que implica ajustar una recta conservadora como la/s marcada/s en la misma Figura 4.1. Tales rectas son:

$$\hat{y}^* = 11,194594 + 35,5282 x \quad (\text{una cola}) \quad (4.18)$$

$$\hat{y}^* = \begin{cases} 11,307150 + 13,3303 x & \text{si } K < 3 \\ 17,064775 + 7,4455 x & \text{si } K \geq 3 \end{cases} \quad (\text{dos colas}) \quad (4.19)$$

y garantizan que $\hat{a}_1^* \geq a_1^*$ (al menos para los datos observados y para muchas otras pruebas extras de tales autores), lo que da cierta seguridad en las c.d.v. que proporcionan. Se observa pues que, como es habitual, el test de dos colas actúa más irregularmente que el de una cola; de hecho requiere de dos rectas para su ajuste, rectas que se cortan en $K \approx 2,6$ y de ahí los puntos extras incluidos para este caso.

Tabla 4.2

Valores de A y B en el Modelo III de asimetría controlada

Valor	Una cola	Dos colas	
		K < 3	K ≥ 3
A	0,31509	0,8482	2,2920
B	0,028147	0,075017	0,13431

Los valores de a y b en (4.18) y (4.19) permiten calcular los de A y B a través de la (4.15), obteniendo los indicados en la Tabla 4.2. Se observa que en el test de dos colas se permite más asimetría inicial (B), pero que ella es disminuida más rápidamente (A) con el marginal más asimétrico. Esto es así porque la (4.13) original equivale a :

$$\beta_1 = \frac{f(n_1)f(a_1)}{C} \geq B - Af(a_1)/f(1) = B - Ay^{-1} \quad (4.20)$$

con lo que la máxima asimetría permitida no es una constante, sino que depende de y^{-1}

(el desequilibrio relativo del marginal más desequilibrado). También se observa que en dos colas se exigen unas c.d.v. iniciales más exigentes que en una cola, pero que su crecimiento es más lento. El momento en que las c.d.v. a una y dos colas coinciden es en $K \approx 1,1$.

4.1.4. Valores estimados de E*: condiciones de validez explícitas

Para comprobar si una tabla dada (a_1, n_1, n) es analizable por chi-cuadrado, el investigador tiene varias opciones prácticas. En primer lugar puede comprobar si no se verifica la (4.13) : si sucede lo contrario

$$\frac{\{(n_2 - n_1)^2 + An_1n_2\}(a_2 - a_1)^2}{a_1a_2n_1n_2BC} < 1 \tag{4.21}$$

el test chi-cuadrado será válido (A y B son las de la Tabla 4.2). Pero para propósitos más amplios es preferible el planteamiento que sigue.

El test chi-cuadrado es válido si $y > y^*$ como se dijo anteriormente, con y^* dado por las (4.18) o (4.19) y "x" e "y" dado por la (4.15). Despejando a_1 en tal desigualdad, y teniendo en cuenta que a_1 es un entero, se deduce que el test chi-cuadrado es válido si:

$$a_1 > \left\lceil \frac{n}{2} \left\{ 1 - \left(1 + \frac{4y^*}{C} \right)^{-0,5} \right\} \right\rceil = [a_1^{**}] = a_1^* (\leq n_1) \tag{4.22}$$

con a_1^{**} aludiendo a la cantidad dentro del primer corchete y a_1^* aludiendo a la parte entera de a_1^{**} (en realidad debería ser \hat{a}_1^* , por tratarse de una predicción, pero en lo que sigue, salvo que haya lugar a error, no se distinguirá). Como $C = (n - 2)^2 / (n - 1) = f(1) \approx (n - 3) = (K + 1)n_1 - 3$, a lo cual converge bastante rápido, entonces :

$$a_1^{**} \approx \frac{(K + 1)n_1}{2} \left\{ 1 - \sqrt{\frac{K(K + 1)n_1 - 3K}{K(K + 1)n_1 + (4a - 3)K + 4b(K - 1)^2}} \right\} \tag{4.23}$$

La condición de validez de $a_1 > a_1^*$ de la (4.22) permite obtener la c.d.v. del test chi-cuadrado en función de cualquier otro parámetro de control. Por ejemplo, $E > E^* = a_1^* / (K + 1)$ o $p = a_1 / n > p^* = a_1^* / n$ o $K' = a_2 / a_1 < K'^* = (n - a_1^*) / a_1^*$. Martín and Herranz estudiaron lo que sucede asintóticamente con las cotas a_1^* , E^* , p^* y K'^* de los parámetros de control a_1 , E , p y K' , comprobando que cuando $n_1 \rightarrow \infty$ o $K \rightarrow \infty$ o

ambos, el único que mantiene una cota distinta de cero y de infinito es E , por lo que éste es el parámetro de control idóneo (lo que coincide con lo tradicional). Como $E^* = a_1^*/(K + 1)$, la (4.22) indica que el test chi-cuadrado es válido si

$$E > E^* = \frac{\left[\frac{n}{2} \left\{ 1 - \left(1 + \frac{4\{aK + b(K - 1)^2\}}{KC} \right)^{-0,5} \right\} \right]}{K + 1} \quad (4.24)$$

que es la c.d.v. idónea y práctica buscada y que ¡no es una constante! Obsérvese que E^* (al igual que a_1^*) es una función de n_1 y K , por lo que debe entenderse que $E^* = E^*(n_1; K)$.

Las c.d.v. de (4.21), (4.22) o (4.24) provienen de un ajuste conservador de los datos, por lo que la afirmación “el test chi-cuadrado es válido” es segura, en tanto que la afirmación contraria no lo es (aunque no se verifiquen las c.d.v., el test aproximado puede ir bien). Pero eso es lo habitual de cualquier c.d.v. : pretenden garantizar la validez de la aproximación propuesta. La Tabla 4.1 también contiene los valores \hat{a}_1^* y \hat{E}^* predichos a través de las (4.22) y (4.24); como era esperable, en todos los casos $\hat{a}_1^* > a_1^*$ y $\hat{E}^* > E^*$ (la c.d.v. predicha es más fuerte que la real).

4.1.5 Consecuencias generales

4.1.5.1 Introducción

En este apartado se detallan las consecuencias que se obtienen del modelo de asimetría controlada. Muchas de ellas son independientes de los valores específicos de “a” y de “b”, y otras prácticamente independientes (válidas para valores razonables), detallándose finalmente algunas conclusiones específicas del ajuste obtenido.

4.1.5.2 Sobre el mínimo valor de a_1

La (4.20) indica que se exige más simetría cuanto más pequeño es el marginal más raro (cuanto más grande es $f(a_1)$). Ella implica dos casos extremos:

- 1) Cuando todos los marginales están equilibrados ($a_i = n_i = n/2$, lo que sólo puede

ocurrir si n es par) es $f(a_1) = 0$ y $\beta_1 = 0 \Rightarrow$ se verifica la (4.20). En consecuencia, en la tablas totalmente equilibradas el test chi-cuadrado siempre es válido.

- 2) Por la (4.20), el test chi-cuadrado nunca será válido cuando $B - Af(a_1)/C \leq 0$, lo que ocurre aproximadamente cuando:

$$a_1 \leq \frac{n}{2} \left\{ 1 - \sqrt{\frac{n-3}{4a+n-3}} \right\} \quad (4.25)$$

cantidad que converge hacia "a" cuando $n \rightarrow \infty$. Esto indica que, para grandes muestras, las tablas no equilibradas con $a_1 \leq [a]^-$ nunca pueden analizarse por chi-cuadrado.

Las dos consecuencias generales anteriores, unificadas, permiten extraer consecuencias específicas para los datos obtenidos. Así, la (4.25) indica que, en el test de una cola, el test chi-cuadrado no puede aplicarse en tablas no equilibradas y con $a_1 \leq 6$ (si $n = 40$) o $a_1 \leq 8$ (si $n = 100$) o $a_1 \leq 10$ (si $n = 1000$) o $a_1 \leq 11$ (si $n \geq 1725$). De un modo general y conservador, la expresión $a_1 \leq [a]^-$ indica que "en tablas grandes no equilibradas y con $a_1 \leq 11$ (una cola o dos colas y $K < 3$) o $a_1 \leq 17$ (dos colas y $K \geq 3$) el test chi-cuadrado no puede aplicarse".

4.1.5.3. Sobre la forma de la función $E^*(n_1; K)$

Derivando en K o en n_1 la función E^* dada por la (4.24), Martín and Herranz obtuvieron que para los valores a y b de las (4.18) y (4.19):

- 1) En K constante, E^* crece con n_1 , y tanto más rápido cuanto más grande es K .
- 2) En n_1 constante, E^* decrece con K hasta las proximidades de $K = 1,1$ (1,3 en dos colas), luego crece hasta aproximarse a un máximo y, finalmente, decrece hasta cero (aunque todo ello de un modo muy general).
- 3) Las anteriores descripciones son de tipo macroscópico pues los valores de a_1^* son discretos. Así, la afirmación de que E^* "crece" con n_1 debe entenderse que microscópicamente significa que "no decrece" (que es lo que pasa cuando se

transita de a_1^* a $a_1^* + 1$).

4.1.5.4. Sobre los valores límites de E^*

Con independencia de cuánto valgan "a" y "b", la (4.24) verifica que:

$$E^* \rightarrow \frac{\left[a + b \frac{(K - 1)^2}{K} \right]}{K + 1} \quad \text{si } n_1 \rightarrow \infty \quad (4.26)$$

$$E^* \rightarrow \frac{n_1}{K(n_1) + 1} \leq \frac{n_1}{2} \left\{ 1 - \sqrt{\frac{n_1}{n_1 + 4b}} \right\} \quad \text{si } K \rightarrow K(n_1) \quad (4.27)$$

$$E^* \rightarrow 0 \quad \text{si } K(n_1) < K \rightarrow \infty$$

$$E^* \rightarrow b \quad \text{si } n_1, K \rightarrow \infty \quad (4.28)$$

en donde $K(n_1)$ es la solución en K de la igualdad:

$$n_1 = \frac{3K^2 + aK(K - 1)^2 + b(K - 1)^4}{K^2(K + 1)} \quad (4.29)$$

y se verifica que:

$$K(n_1) \geq 1 + \frac{n_1 + \sqrt{n_1(n_1 + 4b)}}{2b} \simeq \frac{n_1 + b}{b} \quad (4.30)$$

4.1.5.5. Algunas c.d.v. globales

La conjunción de los dos apartados anteriores permite extraer algunas conclusiones:

- 1) Puesto que E^* crece con n_1 (para K fijo), su máximo lo da la (4.26) que, por ser creciente en K , tendrá tal máximo en el límite de (4.28). En consecuencia el máximo absoluto de E^* es b , y así una c.d.v. universal y conservadora es $E > b$, lo que garantiza que E^* tiene una cota finita y, por tanto, no crece indefinidamente.

Para nuestro caso la c.d.v. general es $E > 35,5282$ (una cola) o $E > 7,4455$ (dos colas). La primera de ellas, como se ve, es tremendamente lejana del clásico valor de 5.

- 2) Un caso particular de lo anterior es $E^*(n_1 = \infty; K = 1) = [a]^-/2$. Como $E^*(n_1; K = 1) \leq E^*(n_1 = \infty; K = 1)$, entonces $E > [a]^-/2$ es la c.d.v. para cualquier tabla con $K = 1$.

Para nuestro caso, $E > 5,5$ es la c.d.v. para cualquier tabla con $K=1$ (tanto en una como en dos colas), valor sorprendentemente cercano al clásico de 5.

- 3) De igual modo, $E^*(n_1 = \infty; K)$ sirve como una cota conservadora para cada valor de K en cualquier n_1 .

Por ejemplo, para una cola, las cotas conservadoras en $K = 1; 1,5; 2; 5$ y 10 son $E > 5,5; 6,8; 9,53; 20,67$ y $27,09$ respectivamente.

- 4) Ya se ha indicado que E^* , para n_1 fijo, crece con K de modo general hasta alcanzar el máximo que indica la (4.27), para luego decrecer hasta cero. Tal máximo se alcanza en $K = K(n_1)$ dado por la solución de la (4.29), y algunos de sus valores son $K(n_1) = 4,74; 6,17$ y $32,6$ para $n_1 = 60; 100$ y 1000 , respectivamente, en el caso de una cola. Para dos colas son bastante mayores. En consecuencia el máximo se alcanza bastante pronto en los n_1 pequeños, pero es muy tardío en los n_1 grandes.

4.1.6 Discusión y conclusiones

En esta sección se ha visto cómo la validez del test chi-cuadrado para tablas 2×2 emergentes del Modelo III, como una aproximación del test exacto de Fisher, está muy ligada a la asimetría de la distribución hipergeométrica base, de modo que el test aproximado es válido si la asimetría es menor que un número dado (que depende del tamaño de muestra y del desequilibrio del marginal más raro).

A efectos prácticos, un modo sencillo de comprobar la validez del test chi-cuadrado es chequeando si se verifica la (4.21) o la (4.22) o la (4.24), todas ellas equivalentes, expresiones que están sujetas a la presentación de datos que indica la (4.1) y en las que a y b son las de (4.18) o (4.19), A y B las de la Tabla 4.2, y K alude a n_2/n_1 .

Se ha visto también que la condición de asimetría acotada ($\beta_1 < \beta_1^*$) es equivalente a la condición clásica de $E > E^*$, pero que la primera permite deducir la

forma de la segunda. Asimismo se ha visto que E es el parámetro más adecuado para controlar la validez del test chi-cuadrado, y ello por sus convenientes propiedades asintóticas.

En particular se ha comprobado cómo E^* es una función de n_1 y K , función que en general crece con n_1 y con K , alcanzando un máximo absoluto de 35,53 (una cola) y de 7,75 (dos colas) —muy distintos del valor comúnmente admitido de 5— que pueden servir como c.d.v. generales y conservadoras. Una alternativa de c.d.v., intermedia entre las muy conservadoras anteriores y la demasiado compleja de la (4.24), es $E > E^*$ ($n_1 = \infty$; K) dada por la (4.26). Un caso particular que no es preciso evaluar es cuando todos los marginales son iguales: el test chi-cuadrado es entonces siempre válido.

Un comentario general ha sido que las c.d.v. son menos estrictas en los test de dos colas que en los de una cola (salvo para tablas con un marginal muy equilibrado), pero también son más irregulares en su relación con la realidad. A estos efectos (lo que no ha sido citado anteriormente), Martín and Herranz comprobaron que la curtosis de la distribución hipergeométrica juega un papel complementario a la asimetría (especialmente en tests de dos colas), ayudando a explicar algunos desajustes de la segunda.

Para obtener los resultados anteriores se ha determinado experimentalmente la ecuación de una recta $y = a + bx$, en donde x e y dependen sólo de los marginales de la tabla y las constantes a y b tienen varios significados físicos. Por un lado $[a]^-$ es el valor mínimo del marginal más pequeño para que el test chi-cuadrado pueda ser válido en una tabla no equilibrada, en tanto que $[a]^-/2$ es el valor mínimo de E para la validez del test cuando uno de los marginales está equilibrado. Por otro lado, b es el máximo valor posible de E^* . Ambos, a su vez, están relacionados con la máxima asimetría permitida, pues $A = a/b$ y $B = 1/b$, con B la máxima asimetría que se le permite a una tabla para que el test chi-cuadrado sea válido y A una constante reductora de aquella (que actúa proporcionalmente sobre la asimetría relativa del marginal más asimétrico). Los valores de B son del orden de 0,03 en una cola y de 0,10 en dos colas.

4.2. EL VALOR DE E^* EN EL MODELO I*

4.2.1. Introducción

El objetivo de esta sección es efectuar para el Modelo I un desarrollo similar al realizado para el Modelo III en la sección anterior. Muchas de las motivaciones, demostraciones y fórmulas para el caso actual serán las del anterior, de ahí que sean omitidas o reflejadas más sucintamente, haciendo hincapié sólo en lo que de distinto suceda (lo que lleva a subdividir los apartados de la sección 4.1 cuando sea preciso).

4.2.2. Valores experimentales de E^*

4.2.2.1. Tests a utilizar

Para el actual Modelo I, el test exacto a utilizar para obtener el P_E de una tabla dada debería ser el óptimo de Barnard que se ha aludido por M. Sin embargo, las limitaciones de cómputo del mismo no permitirían acoger valores grandes de n (que es cuando el test asintótico puede funcionar bien), y de ahí que en lo que sigue se utilice como exacto al método Y: él fue el óptimo de entre los de bajo tiempo de cómputo y su orden está basado en el test asintótico.

Para obtener el valor P_A aproximado se utilizará el test chi-cuadrado que en el Capítulo III ha sido considerado óptimo de modo global para los errores α del 1% al 10% que son de interés actual: el estadístico RP dado por la (3.24).

4.2.2.2. Validez del test aproximado en una tabla dada

Para una tabla experimental, diremos que el test chi-cuadrado RP “falla” en esa tabla cuando P_A sea excesivamente distinta de P_E , es decir, cuando ocurre la (4.4) para los δ 's de la (4.5). Obsérvese que tales δ 's son los que se utilizaron para el Modelo III y no los usados en el Capítulo III para el actual Modelo I. La razón es que aquellos son más razonables a los propósitos perseguidos, habiendo sido utilizados los otros en el Capítulo III por los bajos valores de n ensayados y para que las comparaciones entre métodos sean más claras. Cuando no ocurra la (4.4) diremos que el método RP “no falla” para esa tabla.

4.2.2.3. Generación del espacio muestral

Para un valor dado de n , el espacio muestral en el Modelo I viene determinado por todas las ternas posibles (x_1, y_1, x_2) , si bien en 2.2.2. se optó por la terna más intuitiva de (x_1, x_2, n_1) , con lo que al espacio muestral se genera mediante la (2.17):

$$0 \leq n_1 \leq n; \quad 0 \leq x_1 \leq n_1; \quad 0 \leq x_2 \leq n_2 \quad (4.31)$$

Sin embargo, ya entonces se observaron las propiedades de equivalencia y de simetría, propiedades que implicaban que las tablas formaban parte de la RC de una cola en grupos de 4 tablas (en algunos casos 2 o 1), todas ellas con igual valor P , y que cada una de ellas tenía una tabla simétrica (de asociación opuesta) que entraba en la RC de dos colas junto con la otra y con valor P también idéntico. Ello llevó a generar el espacio muestral de un modo simplificado (canónico) que se basaba en dos principios reorganizadores de la tabla (parte izquierda de la expresión que sigue) y que daban lugar a las tres expresiones generadoras del espacio muestral (parte derecha de la expresión que sigue) que son las de (2.21):

$$\left\{ \begin{array}{l} n_1 = \text{Min} (a_1, a_2, n_1, n_2) \\ x_1 y_2 / x_2 y_1 \leq 1 \\ \text{(Asociación negativa o nula)} \end{array} \right. \Rightarrow \left\{ \begin{array}{l} 0 \leq n_1 \leq [n/2]^- \\ n_1 \leq a_1 \leq n - n_1 \\ 0 \leq x_1 \leq \left[\frac{a_1 n_1}{n} \right]^- \end{array} \right. \quad (4.32)$$

La representación gráfica de tal espacio muestral simplificado se dió en la Figura 2.2(c), pero, a efectos comparativos con lo que sigue, conviene insistir ahora en ello. Para cada valor de n_1 de los que autoriza la (4.32), el espacio muestral simplificado lo definen las dos últimas expresiones de la misma (4.32) y viene representado en la Figura 4.2(a).

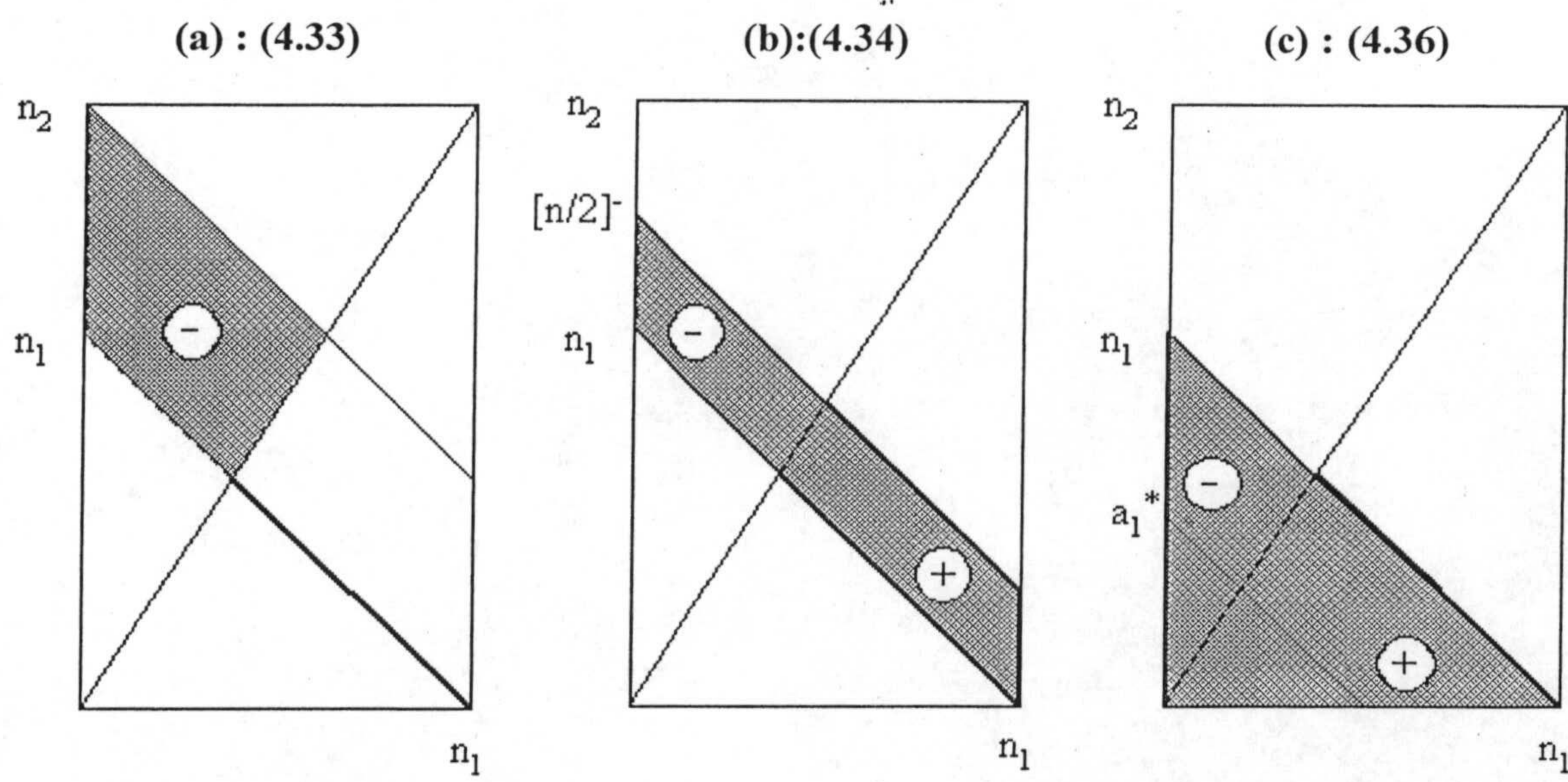
Sin embargo, en paralelo a lo indicado en 4.1.2, ahora estamos interesados en asegurar que el valor de x_1 sea el de la casilla con la menor cantidad esperada, para lo cual hay que obligar a que sea $a_1 = \text{Mín} (a_1, a_2)$. Esto implica que las (4.32) cambian a estas otras:

$$\begin{cases} n_1 = \text{Min}(a_1, a_2, n_1, n_2) \\ a_1 = \text{Min}(a_1, a_2) \end{cases} \Rightarrow \begin{cases} 0 \leq n_1 \leq [n/2]^- \\ n_1 \leq a_1 \leq [n/2]^- \\ 0 \leq x_1 \leq n_1 \end{cases} \quad (4.33)$$

cuya representación gráfica se da en la Figura 4.2(b). Ambos sistemas de generar datos son equivalentes. La diferencia es que mientras que en la (4.32) se tomaba como representante canónico a una de las tablas E_i de asociación negativa (recordar la Tabla 2.1), en la (4.33) se toma bien una tabla E_i , bien una tabla S_i (asociación negativa), lo que es compensado por el hecho de que a_1 disminuye su variabilidad a la mitad. Como excepción, la (4.33) contienen algunas tablas más que la (4.32). En efecto, cuando $n = \text{par}$, es decir cuando $n/2$ es un entero, las tablas con $a_1 = n/2$ sólo deberían intervenir en la (4.33) si son de asociación negativa, pues las de asociación positiva son simétricas de ellas y ya están representadas, lo que implicaría que la línea frontera superior de la Figura 4.2(b) sólo intervendría en su mitad izquierda. Sin embargo, a efectos de lo que sigue, la matización es irrelevante pues las tablas repetidas son del mismo grupo y tienen igual valor P de una o de dos colas.

Figura 4.2

Espacio muestral simplificado (zona sombreada) ilustrado para una de las cuadrículas (n_1, n_2) y basado en las expresiones que se indican. Los símbolos \ominus y \oplus aluden a que en esa zona las tablas son de asociación negativa o positiva respectivamente. En el caso (c), las zonas que contienen a tales símbolos aluden a la zona de validez del test aproximado (pues en ella es $a_1 > a_1^*$).



Finalmente, aún es conveniente una modificación más en el modo de generar las tablas, y ello porque todos los razonamientos y demostraciones (que pretendemos aprovechar) de la sección 4.1 están basados en el convenio de la (4.1), el cual, respecto el de la (4.33), tiene permutados los papeles de a_1 y n_1 . Haciéndolo así ahora, la (4.33) pasa a ser:

$$\begin{cases} a_1 = \text{Min}(a_1, a_2, n_1, n_2) \\ n_1 = \text{Min}(n_1, n_2) \end{cases} \Rightarrow \begin{cases} 0 \leq a_1 \leq [n/2]^- \\ a_1 \leq n_1 \leq [n/2]^- \\ 0 \leq x_1 \leq n_1 \end{cases} \quad (4.34)$$

Un pequeño cambio extra es conveniente aún en la (4.34) pues pretendemos fijar el espacio muestral para cada valor dado de n_1 . Como $a_1 \leq n_1$, entonces la (4.34) es equivalente a esta otra:

$$\begin{cases} a_1 = \text{Min}(a_1, a_2, n_1, n_2) \\ n_1 = \text{Min}(n_1, n_2) \end{cases} \Rightarrow \begin{cases} 0 \leq n_1 \leq [n/2]^- \\ 0 \leq a_1 \leq n_1 \\ 0 \leq x_1 \leq a_1 \end{cases} \quad (4.35)$$

cuya representación gráfica es la indicada en la Figura 4.2(c). Aquí los comentarios son los de antes, y el caso citado de $n = \text{par}$ implica que toda la mitad izquierda del espacio muestral del rectángulo $n/2 \times n/2$ es una repetición de la mitad derecha: la primera contiene tablas E_i que son del mismo grupo de tablas equivalentes-simétricas que la segunda (que contiene a los representantes S_i). Para este formato canónico será:

$$E = \frac{a_1 n_1}{n} ; \quad \hat{p} = \frac{a_1}{n} \leq \hat{q} = \frac{n_1}{n} \leq 0,5 \quad (4.36)$$

4.2.2.4. Valores experimentales de a_1^* a obtener

Cada pareja (n, n_1) proporciona un espacio muestral como el de (4.36) y la Figura 4.2(c) en el que hay varias diagonales a_1 : $0 \leq a_1 \leq n_1$. Como se indicó en 4.1.2 el objetivo es determinar para él un valor a_1^* como el de tal figura que verifique que para $a_1 > a_1^*$ el test chi-cuadrado sea válido (es decir, en ninguna de sus tablas el test falla). Para cada n_1 habrá un a_1^* en general distinto, y así ahora las c.d.v. del test indicarán que el test chi-cuadrado es válido en ciertas partes del espacio muestral y no lo es en otras. La forma de

obtener el valor de a_1^* es en principio clara:

(4.37) “Para un a_1 dado, generar todas las tablas con $0 \leq x_1 \leq a_1$, retener sólo aquellas en que $1\% \leq P_A \leq 10\%$ y comprobar si alguna verifica la (4.4), es decir, si alguna falla. Si ninguna falla, disminuir a_1 en una unidad y repetir lo anterior. Si alguna falla $a_1^* = a_1$. El proceso comienza con $a_1 = n_1$ y finaliza cuando se llega a algún fallo o se llega a $a_1 = 0$, en cuyo caso será $a_1^* = 0$ (pues tal valor de a_1 nunca da tablas de la RC)”.

Cada a_1^* está asociado a una pareja $(n, n_1) \equiv (n_1, K)$ cuyos valores hay que proporcionar. La propuesta de (4.3) no es asumible ahora por la gran dificultad de cómputo del Modelo I, por lo que habrá que rebajarla a estos otros valores:

$$n_1 = 10(5)30$$

$$K = 1(0,1)2; 2,4(0,2)3; 4; 5$$

$$n \leq 60$$

los cuales hay que entenderlos de modo genérico; así, el valor $K = 1,1$ no es posible cuando $n_1 = \text{impar}$ (por lo que se toman otros para compensar), y los grandes valores de K tampoco son posibles en la gama alta de valores de n_1 . La Tabla 4.3 presenta los valores realmente utilizados para (n_1, K) , así como los valores a_1^* y E^* obtenidos por el procedimiento descrito en (4.37).

4.2.2.5. Análisis de los resultados

Observando los datos de la Tabla 4.3 se ve que su comportamiento es similar a los de la Tabla 4.1 (salvando las diferencias que ocasionan los muy diferentes valores de n_1), aunque hay diferencias importantes de matices que obligan a comentar los nuevos datos en paralelo a lo realizado al final del apartado 4.1.2 para los datos de la Tabla 4.1:

- 1) Para n_1 constante, los valores de a_1^* aumentan en general con K y a un ritmo similar en una que en dos colas (mientras que en el Modelo III el ritmo era más lento a dos colas). Tal crecimiento tiene ligeras irregularidades para una cola (en el Modelo III había grandes irregularidades a dos colas).
- 2) Para $K = 1$, los valores de a_1^* también crecen con n_1 , siendo similares a una y a dos colas (en el Modelo III eran constantes y diferentes).

3) Los valores de E^* son similares a una y a dos colas, aumentando con K y con n_1 (en el Modelo III eran diferentes). Ahora nunca es $E^* > 5$, pero ello puede ser consecuencia de los bajos valores observados para n_1 .

de donde se deduce que a_1^* (y E^*) crecen con n_1 y con K y de un modo similar a una que a dos colas.

Tabla 4.3
Valores mínimos reales (a_1^* y E^*) del marginal más pequeño (a_1) y de la mínima cantidad esperada (E) a partir de los cuales el test chi-cuadrado RP es válido (como aproximación del test exacto Y) y predicciones conservadoras \hat{a}_1^* del primero. Se asume que $K=n_2/n_1 \geq 1$.

n_1	K	Una cola			Dos colas		
		a_1^*	\hat{a}_1^*	E^*	a_1^*	\hat{a}_1^*	E^*
10	1	3	4	1.5	3	5	1.5
10	1.1	3	4	1.4	3	5	1.4
10	1.2	3	4	1.4	3	5	1.4
10	1.3	4	5	1.7	4	5	1.7
10	1.4	3	5	1.3	3	5	1.3
10	1.5	4	5	1.6	4	6	1.6
10	1.7	4	6	1.5	4	6	1.5
10	1.8	4	6	1.4	4	6	1.4
10	1.9	5	6	1.7	5	7	1.7
10	2	5	7	1.7	6	7	2
10	2.4	7	8	2.1	7	8	2.1
10	3	9	10	2.3	9	10	2.3
10	4	10	10	2	10	10	2
10	5	10	10	1.7	10	10	1.7
15	1	3	5	1.5	4	6	2
15	1.2	6	6	2.7	4	7	1.8
15	1.4	5	6	2.1	5	7	2.1
15	1.6	6	7	2.3	6	8	2.3
15	1.8	6	8	2.1	6	9	2.1
15	2	8	9	2.7	8	9	2.7
15	2.2	9	9	2.8	9	10	2.8
15	2.4	10	10	2.9	9	11	2.6
15	2.6	10	11	2.8	12	12	3.3
15	2.8	9	12	2.4	10	12	2.6
15	3	13	13	3.3	11	13	2.8
20	1	5	6	2.5	4	8	2
20	1.1	6	7	2.9	7	8	3.3
20	1.2	6	7	2.7	8	8	3.6
20	1.3	5	7	2.2	6	8	2.6
20	1.4	7	8	2.9	8	9	3.3
20	1.5	8	8	3.2	8	9	3.2
20	1.7	7	9	2.6	10	10	3.7
20	1.8	9	9	3.2	10	10	3.6
20	1.9	9	10	3.1	10	11	3.4
20	2	8	10	2.7	11	11	3.7
25	1	6	7	3	7	9	3.5
25	1.2	8	8	3.6	8	9	3.6
25	1.4	9	9	3.8	8	10	3.3
30	1	6	8	3	10	10	5

4.2.3. La simetría del estadístico base para el Modelo I

4.2.3.1. Introducción

La variable base en el Modelo III era x_1 , cuya distribución, bajo H_0 , es la hipergeométrica, distribución de la cual se conoce perfectamente el cuadrado β_1 del coeficiente de asimetría. En el caso del Modelo I ya se justificó en 3.1.3.1. que Pirie and Hamdan (1972) propusieron como variable base al estadístico $U = x_1y_2 - x_2y_1$, cuya media y varianza se dieron entonces, pero de la que no se conoce el valor de β_1 . Es pues claro que, para efectuar el desarrollo paralelo al Modelo III que se persigue, es preciso determinar el valor de β_1 para la variable aleatoria U . Pero con tal fin, y para facilitar la notación de las demostraciones, hemos de cambiar momentáneamente los convenios notacionales empleados hasta ahora, cambiando los clásicos de la Tabla 1.1 por los de la Tabla 4.4. Con esta nueva notación, la probabilidad de una tabla dada es:

Tabla 4.4

Notación transitoria para el Modelo I

x_1	x_2	n_1	p_1	p_2	q
x_3	x_4	n_2	p_3	p_4	$1 - q$
a_1	a_2	n	p	$1 - p$	1

$$P(x_1, x_2, x_3, x_4) = \frac{n!}{\prod x_i!} \prod p_i^{x_i} \tag{4.38}$$

La variable base es:

$$U = x_1x_4 - x_2x_3 \tag{4.39}$$

y su media y varianza bajo la hipótesis de independencia $H_0 \equiv p_1p_4 = p_2p_3$ son (ver subapartado 3.1.3.1.):

$$\begin{cases} E(U) = n(n - 1)(p_1p_4 - p_2p_3) \Rightarrow E(U|H_0) = 0 \\ V(U|H_0) = E(U^2|H_0) = n^2(n - 1)p_1p_4 = \sigma^2 \end{cases} \tag{4.40}$$

4.2.3.2. Momentos de la multinomial

La función generatriz de momentos de la multinomial es:

$$\varphi(t_1, t_2, t_3, t_4) = \left(\sum p_i e^{t_i} \right)^n = \psi^n \tag{4.41}$$

por lo que cualquier momento no central puede obtenerse así:

$$\mu'_{r_1, r_2, r_3, r_4} = E(\prod x_i^{r_i}) = \left[\frac{\partial^{\sum r_i} \psi^n}{\partial t_1^{r_1} \partial t_2^{r_2} \partial t_3^{r_3} \partial t_4^{r_4}} \right]_{t_1=t_2=t_3=t_4=0} \tag{4.42}$$

Pirie and Hamdan, en su demostración, evaluaron los siguientes momentos:

$$\mu'_{1,0,0,1} = E(x_1 x_4) = n(n-1)p_1 p_4 \tag{4.43}$$

$$\mu'_{0,1,1,0} = E(x_2 x_3) = n(n-1)p_2 p_3 \tag{4.44}$$

$$\mu'_{1,1,1,1} = E(x_1 x_2 x_3 x_4) = n(n-1)(n-2)(n-3)p_1 p_2 p_3 p_4 \tag{4.45}$$

$$\begin{aligned} \mu'_{2,0,0,2} = E(x_1^2 x_4^2) &= n(n-1)(n-2)(n-3)p_1^2 p_4^2 + n(n-1)(n-2)p_1 p_4 (p_1 + p_4) \\ &+ n(n-1)p_1 p_4 \end{aligned} \tag{4.46}$$

$$\begin{aligned} \mu'_{0,2,2,0} = E(x_2^2 x_3^2) &= n(n-1)(n-2)(n-3)p_2^2 p_3^2 + n(n-1)(n-2)p_2 p_3 (p_2 + p_3)n \\ &(n-1)p_2 p_3 \end{aligned} \tag{4.47}$$

aunque lo hicieron por un mecanismo excesivamente largo. Más corto es aplicar la (4.42). Por ejemplo:

$$\begin{aligned} \mu'_{1,0,0,1} &= \left[\frac{\partial^2 \psi^n}{\partial t_1 \partial t_4} \right]_{t_1=t_2=t_3=t_4=0} = \left[\frac{\partial}{\partial t_4} \left\{ n \psi^{n-1} p_1 e^{t_1} \right\} \right]_{t_1=t_2=t_3=t_4=0} \\ &= \left[n p_1 e^{t_1} \left\{ (n-1) \psi^{n-2} p_4 e^{t_4} \right\} \right]_{t_1=t_2=t_3=t_4=0} = n(n-1)p_1 p_4 \end{aligned}$$

pues en $t_1 = t_2 = t_3 = t_4 = 0$ es $\psi = 1$ por ser $\sum p_i = 1$.

Para lo que sigue es preciso determinar también otros momentos. Con una prueba basada en la (4.42), y que no tiene más dificultad que el gran número de operaciones, se llega a que:

$$\begin{aligned} \mu'_{2,1,1,2} &= n(n-1)(n-2)(n-3)(n-4)(n-5)p_1^3 p_4^3 + \\ &n(n-1)(n-2)(n-3)(n-4)p_1^2 p_4^2 (p_1 + p_4) + \\ &n(n-1)(n-2)(n-3)p_1^2 p_4^2 \end{aligned} \tag{4.48}$$

y, por simetría,

$$\mu'_{1,2,2,1} = n(n-1)(n-2)(n-3)(n-4)(n-5)p_2^3 p_3^3 +$$

$$\begin{aligned} & n(n-1)(n-2)(n-3)(n-4)p_2^2 p_3^2 (p_2 + p_3) + \\ & n(n-1)(n-2)(n-3)p_2^2 p_3^2 \end{aligned} \quad (4.49)$$

De igual modo puede probarse que:

$$\begin{aligned} \mu'_{3,0,0,3} = & n(n-1)(n-2)(n-3)(n-4)(n-5)p_1^3 p_4^3 + \\ & 3n(n-1)(n-2)(n-3)(n-4)p_1^2 p_4^2 (p_1 + p_4) + \\ & n(n-1)(n-2)(n-3)p_1 p_4 (p_1^2 + p_4^2 + 9p_1 p_4) + \\ & 3n(n-1)(n-2) p_1 p_4 (p_1 + p_4) + n(n-1)p_1 p_4 \end{aligned} \quad (4.50)$$

y, por simetría,

$$\begin{aligned} \mu'_{0,3,3,0} = & n(n-1)(n-2)(n-3)(n-4)(n-5)p_2^3 p_3^3 + \\ & 3n(n-1)(n-2)(n-3)(n-4)p_2^2 p_3^2 (p_2 + p_3) + \\ & n(n-1)(n-2)(n-3)p_2 p_3 (p_2^2 + p_3^2 + 9p_2 p_3) + \\ & 3n(n-1)(n-2) p_2 p_3 (p_2 + p_3) + n(n-1)p_2 p_3 \end{aligned} \quad (4.51)$$

4.2.3.3. Asimetría del estadístico U (bajo H_0)

Si μ_r y μ'_r aluden a los momentos r del estadístico U , centrales y no centrales respectivamente, entonces, por definición, el cuadrado de la asimetría de U es $\beta_1 = \mu_3^2 / \mu_2^3$. Como tal asimetría se va a precisar bajo H_0 , en cuyo caso $E(U) = \mu_1' = 0$ por la (4.40), entonces:

$$\beta_1 = \frac{\mu_3'^2}{\mu_2'^3} = \frac{E^2(U^3|H_0)}{E^3(U^2|H_0)} = \frac{E^2(U^3|H_0)}{\{n^2(n-1)p_1 p_4\}^3} \quad (4.52)$$

también por la (4.40), y así todo lo que queda por conocer es el valor de $E(U^3|H_0)$.

Para ello, como:

$$\begin{aligned} E(U^3) &= E\{(x_1 x_4 - x_2 x_3)^3\} = E\{(x_1^2 x_4^2 + x_2^2 x_3^2 - 2x_1 x_2 x_3 x_4)(x_1 x_4 - x_2 x_3)\} \\ &= E\{x_1^3 x_4^3 - x_1^2 x_2 x_3 x_4^2 + x_1 x_2^2 x_3^2 x_4 - x_2^3 x_3^3 - 2x_1^2 x_2 x_3 x_4^2 + 2x_1 x_2^2 x_3^2 x_4\} \\ &= \mu'_{3,0,0,3} - \mu'_{2,1,1,2} + \mu'_{1,2,2,1} - \mu'_{0,3,3,0} - 2\mu'_{2,1,1,2} + 2\mu'_{1,2,2,1} \\ &= (\mu'_{3,0,0,3} - \mu'_{0,3,3,0}) + 3(\mu'_{1,2,2,1} - \mu'_{2,1,1,2}) \end{aligned}$$

entonces el primer paréntesis es la diferencia entre las expresiones (4.50) y (4.51) y el segundo la diferencia entre las (4.49) y (4.48). Como tales diferencias hay que evaluarlas bajo H_0 , entonces, teniendo en cuenta que $p_1 p_4 = p_2 p_3$, serán:

$$\begin{aligned} \mu'_{3,0,0,3} - \mu'_{0,3,3,0} &= 3n(n-1)(n-2)(n-3)(n-4)p_1^2 p_4^2 (p_1 + p_4 - p_2 - p_3) + \\ &\quad n(n-1)(n-2)(n-3)p_1 p_4 (p_1^2 + p_4^2 - p_2^2 - p_3^2) + \\ &\quad 3n(n-1)(n-2)p_1 p_4 (p_1 + p_4 - p_2 - p_3) \end{aligned}$$

$$3(\mu'_{1,2,2,1} - \mu'_{2,1,1,2}) = 3n(n-1)(n-2)(n-3)(n-4)p_1^2 p_4^2 (p_2 + p_3 - p_1 - p_4)$$

lo que sustituido en $E(U^3)$ da:

$$\begin{aligned} E(U^3|H_0) &= n(n-1)(n-2)(n-3)p_1 p_4 (p_1^2 + p_4^2 - p_2^2 - p_3^2) + \\ &\quad 3n(n-1)(n-2)p_1 p_4 (p_1 + p_4 - p_2 - p_3) \end{aligned} \quad (4.53)$$

Por otro lado, como $\sum p_i = 1$, la igualdad $p_1 p_4 = p_2 p_3$ puede ponerse como $(1 - p_2 - p_3 - p_4)(1 - p_1 - p_2 - p_3) = (1 - p_1 - p_3 - p_4)(1 - p_1 - p_2 - p_4)$. Realizando operaciones y eliminando términos comunes queda que:

$$p_1^2 + p_4^2 - p_2^2 - p_3^2 = p_1 + p_4 - p_2 - p_3$$

lo que sustituido en la (4.53) indica que:

$$E(U^3|H_0) = n^2(n-1)(n-2)p_1 p_4 (p_1 + p_4 - p_2 - p_3)$$

o, para que todo esté en función de p_1 y p_4 , haciendo $p_2 + p_3 = 1 - p_1 - p_4$:

$$E(U^3|H_0) = n^2(n-1)(n-2)p_1 p_4 (2p_1 + 2p_4 - 1)$$

Finalmente, elevando al cuadrado y sustituyendo en la (4.53), el valor buscado del cuadrado de la asimetría del estadístico U (bajo H_0) será:

$$\beta_1 = \frac{(n-2)^2 (2p_1 + 2p_4 - 1)^2}{n^2(n-1)p_1 p_4} \quad (4.54)$$

A fin de escribir la (4.54) en un modo más tradicional, téngase en cuenta que la hipótesis nula de independencia implica que —ver notación de la Tabla 4.4— $p_1 = pq$ y $p_4 = (1-p)(1-q)$, lo que puede sustituirse en el denominador de (4.54). Para el numerador, como:

$$2p_1 + 2p_4 - 1 = 2pq + 2(1-p)(1-q) - 1 = 4pq + 1 - 2p - 2q$$

$$(1-2p)(1-2q) = 1 - 2q - 2p + 4pq$$

entonces $2p_1 + 2p_4 - 1 = (1-2p)(1-2q)$ y así la (4.54) es:

$$\beta_1 = \frac{(n-2)^2 (1-2p)^2 (1-2q)^2}{n^2(n-1)pq(1-p)(1-q)} \quad (4.55)$$

expresión que también está en la notación tradicional de la Tabla 1.1 y que depende

sólo de n y de los parámetros perturbadores p y q . Es de interés observar que como los marginales a_1 y n_1 siguen distribuciones binomiales $B(n; p)$ y $B(n; q)$ respectivamente, y como el cuadrado de la asimetría de una $B(n; p)$ es $\beta_1(p) = (1 - 2p)^2 / \{np(1 - p)\}$, entonces la (4.55) será:

$$\beta_1 = \frac{(n - 2)^2}{n - 1} \beta_1(p) \times \beta_1(q) \quad (4.56)$$

con lo que el cuadrado de la asimetría del U , bajo H_0 , es proporcional al producto de los cuadrados de las asimetrías marginales.

4.2.3.4. Estimación de β_1

A efectos de reproducir el modelo de asimetría acotada, la expresión (4.55) no es de utilidad por depender de los parámetros desconocidos p y q , y así habrá que estimarlos. Para tal fin conviene observar que el denominador de la (4.55) es $\sigma^2 = V(U|H_0)$ según la (4.40), con:

$$\sigma^2 = n^2(n - 1)p(1 - p)q(1 - q) \quad (4.57)$$

de modo que:

$$\beta_1 = \frac{(n - 2)^2 (1 - 2p)^2 (1 - 2q)^2}{\sigma^2}$$

Una estimación plausible de β_1 implica estimar su numerador y su denominador por separado. Para el numerador no hay ambigüedad: basta tomar los tradicionales $\hat{p} = a_1/n$ y $\hat{q} = n_1/n$. Como $1 - 2\hat{p} = (n - 2a_1)/n = (a_2 - a_1)/n$ y $1 - 2\hat{q} = (n_2 - n_1)/n$, entonces

$$\hat{\beta}_1 = \frac{(n - 2)^2}{n^4} \times \frac{(a_2 - a_1)^2 (n_2 - n_1)^2}{\hat{\sigma}^2} \quad (4.58)$$

Pero, como se vió en 3.1.3.1., la estimación de (4.57) puede hacerse de dos modos. Si se asumen los mismos estimadores individuales anteriores, entonces $\hat{\sigma}_P^2 = (n - 1)a_1a_2n_1n_2/n^2$, y así:

$$\hat{\beta}_{1P} = \frac{(n - 2)^2}{n^2(n - 1)} \times \frac{(a_2 - a_1)^2 (n_2 - n_1)^2}{a_1a_2n_1n_2} \quad (4.59)$$

expresión que será la apropiada cuando se utilice el estadístico asintótico χ^2_P de Pirie

and Hamdan de la (3.6) –pues el mismo se obtuvo a través de $\hat{\sigma}_P^2$ – y de ahí el subíndice P que se le ha puesto a β_1 . Si en lugar de estimar p y q individualmente, se estiman $p(1 - p)$ y $q(1 - q)$, entonces ya se vio que $\hat{\sigma}_R^2 = a_1 a_2 n_1 n_2 / (n - 1)$, y así:

$$\hat{\beta}_{1R} = \frac{(n - 2)^2 (n - 1)}{n^4} \times \frac{(a_2 - a_1)^2 (n_2 - n_1)^2}{a_1 a_2 n_1 n_2} \quad (4.60)$$

expresión que será la apropiada cuando se utilice el estadístico asintótico χ^2_R de Richardson de la (3.7) –pues el mismo se obtuvo a través de $\hat{\sigma}_R^2$ – y de ahí el subíndice R.

A fin de contar con una expresión de $\hat{\beta}_1$ apropiada para el estadístico χ^2_Z de la (3.5), y dado que este se obtenía mediante una prueba diferente a la de los anteriores, conviene observar que para que $\chi^2_Z = U^2 / \hat{\sigma}_Z^2$ ocasione la (3.5), es preciso que $\hat{\sigma}_Z^2 = a_1 a_2 n_1 n_2 / n$. Haciendo esto mismo en la (4.58), se obtiene que el valor de $\hat{\beta}_1$ apropiado para el estadístico χ^2_Z es:

$$\hat{\beta}_{1Z} = \frac{(n - 2)^2}{n^3} \times \frac{(a_2 - a_1)^2 (n_2 - n_1)^2}{a_1 a_2 n_1 n_2} \quad (4.61)$$

Los tres valores obtenidos de $\hat{\beta}_1$ para el Modelo I son iguales –salvo constante de proporcionalidad– al valor β_1 de la (4.10) para el Modelo III. Para grandes muestras, los cuatro son proporcionales a $1/n$.

Finalmente, las expresiones de $\hat{\beta}_1$ anteriores pueden ponerse en términos de la función $f(x)$ de la (4.11). Por ejemplo:

$$\hat{\beta}_{1R} = \frac{(n - 2)^2 (n - 1)}{n^4} f(a_1) f(n_1) \quad (4.62)$$

que es similar a la (4.12) que dio lugar a todos los razonamientos de entonces. Dado que el estudio se va a efectuar para el estadístico RP (que es el óptimo), la (4.62) es la expresión de la asimetría para el mismo.

4.2.4. La condición de asimetría acotada y la condición de validez

En paralelo con el apartado 4.1.3., si se asume que la $f(x)$ de (4.11) es válida para $x = a_1$ y $x = n_2$, pero que ahora:

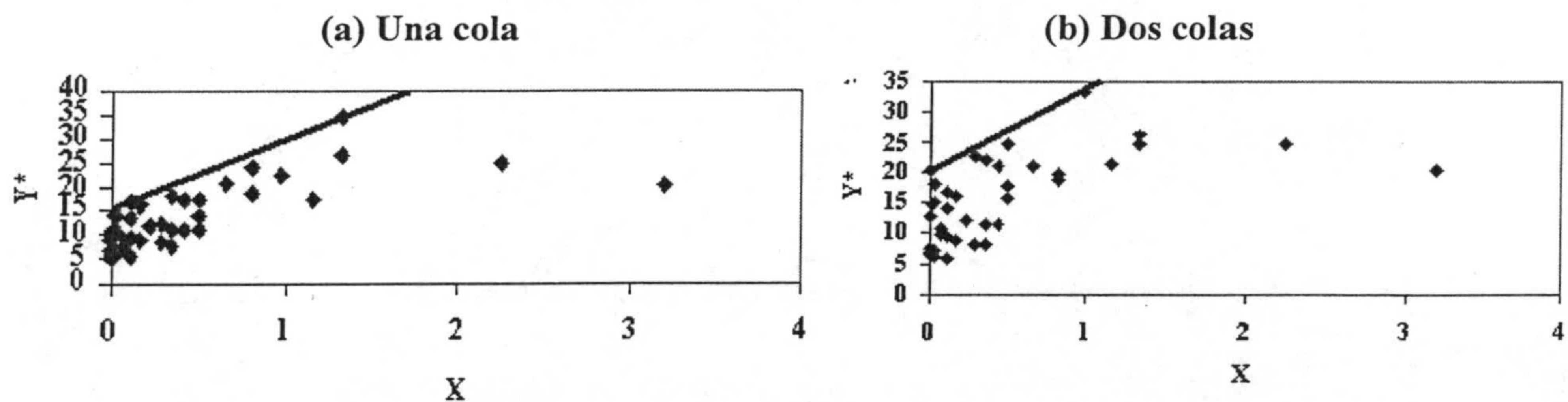
$$C = \frac{n^4}{(n - 2)^2(n - 1)} \tag{4.63}$$

en lugar de $(n - 2)^2/(n - 1)$ como allí, entonces las (4.12), (4.13), (4.14), (4.15), (4.16), (4.17) y (4.20) siguen siendo válidas (puesto que en $K = 1$ sigue siendo verdad que el test no siempre es válido).

Ahora la Figura 4.3 presenta las dos nubes de puntos de los $(x; y^*)$ de la Tabla 4.3 (una y dos colas). Se observa que ambas nubes de puntos son muy similares y no presentan la anomalía de la Figura 4.1 en el caso del test de dos colas (lo que puede deberse a los bajos valores de n_1). Los dos puntos “anómalos” de la Figura 4.3 no son tal: ellos se corresponden con los dos casos de $K = 4$ y $K = 5$ ($n_1 = 10$) en los que ya fue $a_1^* = 10$ (lo máximo posible) y así a_1^* no pudo seguir creciendo (como prevé la tendencia observada). En la misma figura se han representado las dos nuevas rectas conservadoras que serán útiles para las predicciones. Ellas han sido obtenidas a partir de los dos puntos más extremos de la nube de puntos y sus ecuaciones:

Figura 4.3

Nube de puntos para los valores de $(x; y^*)$ del Modelo I –Tabla 4.3– y recta conservadora que los ajusta.



$$\hat{y}^* = 15,34799472 + 14,37361494x = a + bx \text{ (una cola)} \tag{4.64}$$

$$\hat{y}^* = 20,4054893 + 13,02102090x = a + bx \text{ (dos colas)}$$

garantizan que $\hat{a}_1^* \geq a_1^*$ como entonces. Sus valores de a y b determinan los de A y B a través de la (4.15), lo que da:

$$\begin{array}{lll} A = 1,0678 & B = 0,06957. & \text{(una cola)} \\ A = 1,5671 & B = 0,07680 & \text{(dos colas)} \end{array} \tag{4.65}$$

Ahora se observa que la máxima asimetría inicial permitida (B) es ligeramente mayor en los tests de dos colas, pero que ella es disminuida algo más rápidamente (A) con el marginal más asimétrico (recordar la (4.20)). Es por ello que los valores de a_1^* son generalmente algo mayores en tales tests (salvo en algunos valores pequeños de K).

4.2.5. Valores estimados de E*: condiciones de validez explícita

La verificación de si una tabla dada (a_1, n_1, n) es analizable por el test aproximado RP puede hacerse como se indicó en 4.1.4., con los cambios oportunos. Ahora las expresiones (4.21), (4.22) y (4.24) siguen siendo válidas, al igual que los comentarios de entonces, para los nuevos valores de A, B, C, a y b. Ahora $C \approx (n + 5)$, de modo que la paralela a la (4.23) es:

$$a_1^{**} \approx \frac{(K + 1)n_1}{2} \left\{ 1 - \sqrt{\frac{K(K + 1)n_1 + 5K}{K(K + 1)n_1 + (4a + 5)K + 4b(K - 1)^2}} \right\} \tag{4.66}$$

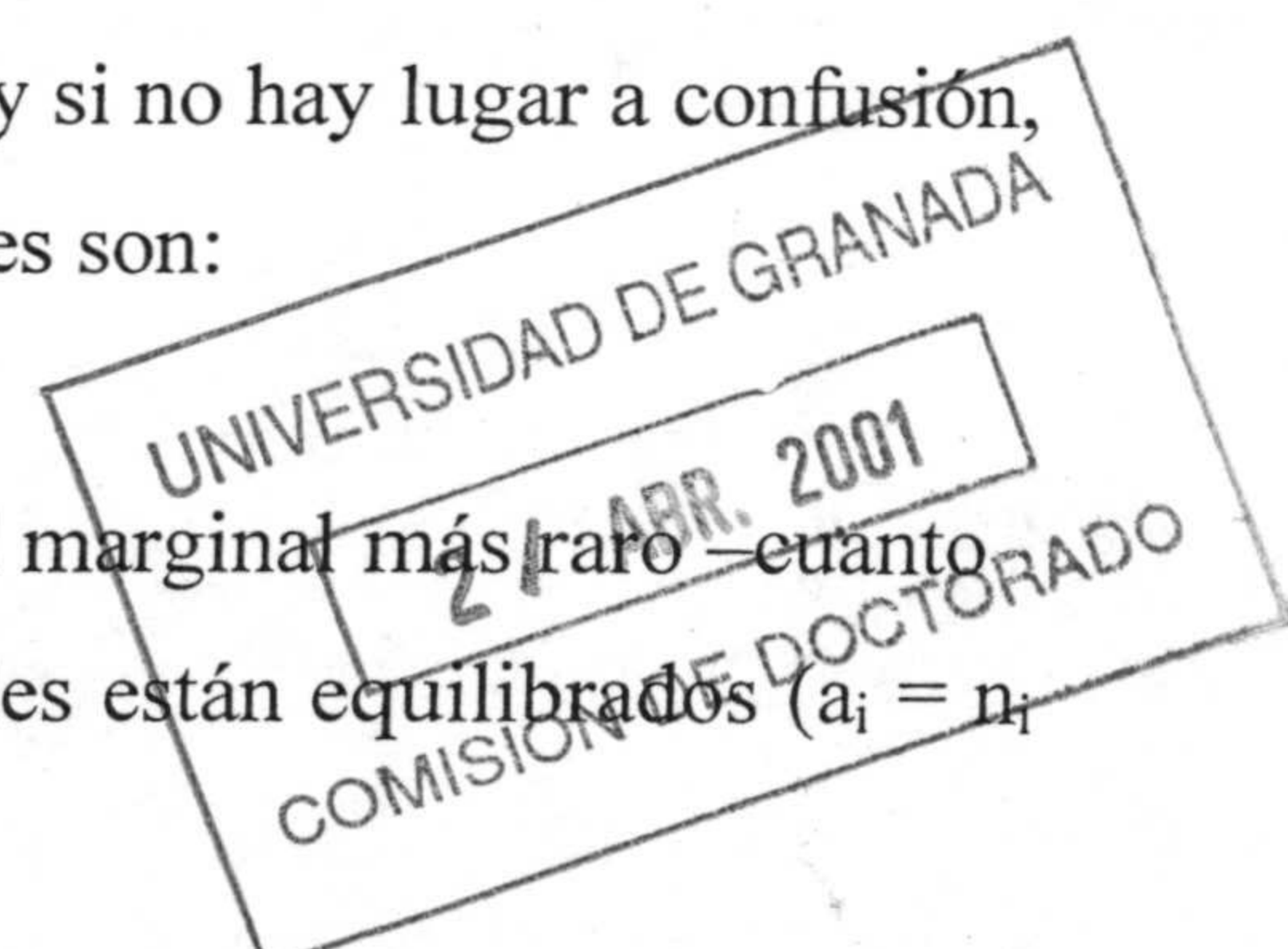
4.2.6. Consecuencias generales

En lo que sigue se reiteran las consecuencias obtenidas en 4.1.5. deteniéndose sólo en lo que de diferente haya. También a partir de ahora y si no hay lugar a confusión, se omite el símbolo “^” encima de a y de E. Las conclusiones son:

- 1ª) El modelo exige más simetría cuanto más pequeño es el marginal más raro —cuanto más grande es $f(a_1)$ — y así, cuando todos los marginales están equilibrados ($a_i = n_i = n/2$), el test RP siempre es válido.
- 2ª) El test RP nunca será válido cuando $B - Af(a_1)/C$ sea nulo o negativo, lo que ocurre aproximadamente cuando:

$$a_1 \leq \frac{n}{2} \left\{ 1 - \sqrt{\frac{n + 5}{4a + n + 5}} \right\} \rightarrow a \text{ (si } n \rightarrow \infty) \tag{4.67}$$

Esto indica que, para grandes muestras, las tablas no equilibradas con $a_1 \leq [a]$ nunca pueden analizarse por RP, es decir, las tablas con $a_1 \leq 15$ (una cola) o $a_1 \leq 20$ (dos colas). Para pequeñas muestras, la cota depende del valor de n. Para n =

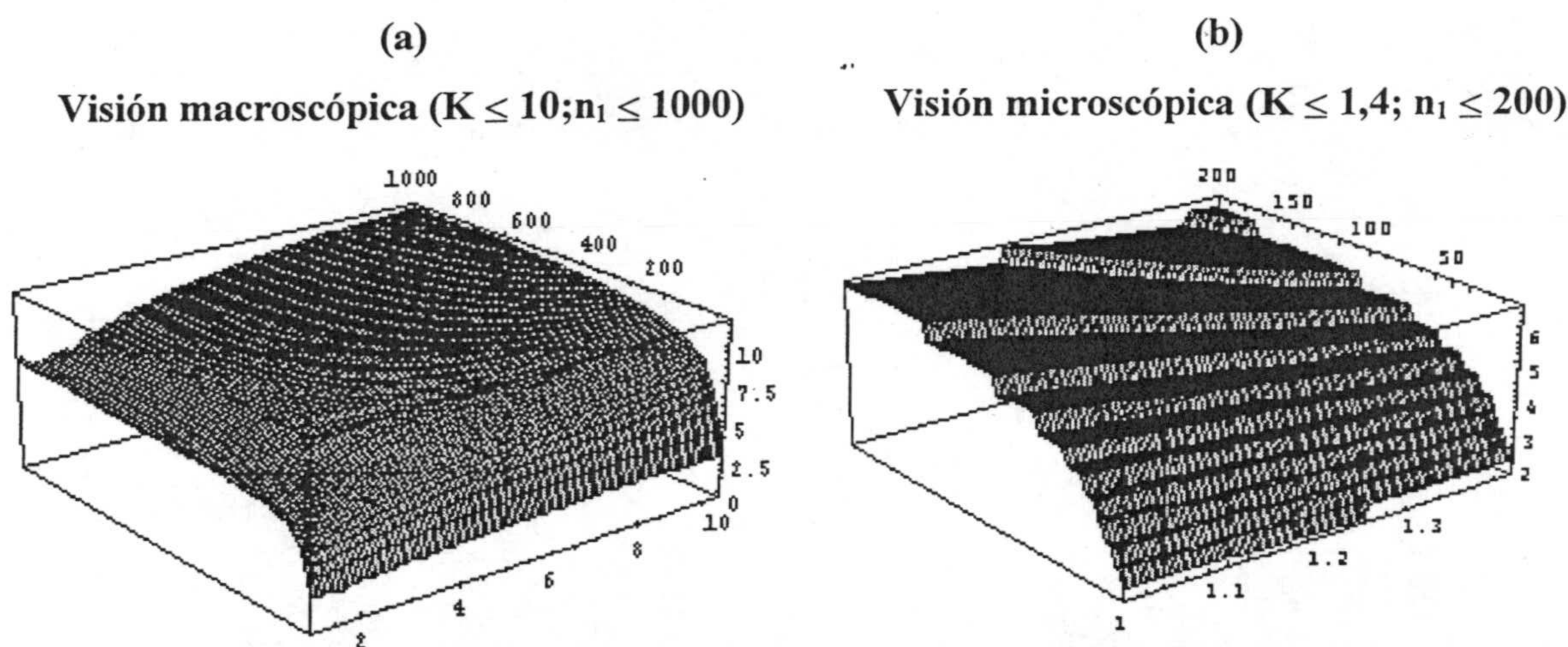


40, 100, 1000 las cotas son $a_1 \leq 6, 10, 14$ (una cola) y 8, 12, 19 (dos colas).

3ª) Para K constante, E^* crece (no decrece) con n_1 , y tanto más rápidamente cuanto más grande es K . Para n_1 constante, E^* decrece (no crece) con K hasta las proximidades de 1,5 (una cola) o de 2 (dos colas), pero luego crece (no decrece) hasta alcanzar un máximo y finalmente decrece (no crece) hasta cero. La Figura 4.4 da una visión macroscópica y microscópica de la forma de E^* para el test de una cola (para el de dos colas es similar). El escalonamiento que se observa en la visión microscópica es debido a lo discreto de la variable a_1^* (con lo que E^* también es discreta); de ahí las aclaraciones anteriores acerca de “no crece” o “no decrece”.

Figura 4.4

Representación gráfica de la mínima cantidad esperada $E^*(n_1, K)$ para que el test chi-cuadrado RP de una cola sea válido. En el eje vertical E^* ; en el plano horizontal K y n_1 .



4ª) Los valores límites de las (4.26), (4.27) y (4.28) siguen siendo ciertos, aunque ahora el valor $K(n_1)$ a que alude la (4.27) es la solución en K de la igualdad de la (4.66) a n_1 (pues se trata de determinar para qué valor de $K - K(n_1) - a_1^{**}$ alcanza el valor máximo posible: n_1), es decir la solución de:

$$n_1 = \frac{-5K^2 + aK(K - 1)^2 + b(K - 1)^4}{K^2(K + 1)} \tag{4.68}$$

que es la paralela a la (4.29). Por ejemplo para dos colas y $n_1 = 100$ es $K(n_1) = 10,62$, lo que indica que el tope no es alcanzado en la mayoría de los casos prácticos (pues los K 's tan grandes no son habituales) para n_1 moderadamente

grande. La (4.30) sigue siendo válida (pues en ella sólo intervienen n_1 y b).

5ª) Una c.d.v. universal y conservadora es $E > b$; para nuestro caso $E > 14,38$ (una cola) y $E > 13,03$ (dos colas). Lo primero es más débil que en el Modelo III, pero lo segundo es más fuerte. Ambas son muy discrepantes con el clásico $E \geq 5$.

6ª) Para cualquier tabla con $K = 1$, $E > [a]^-/2$ es una c.d.v. válida y conservadora. Para nuestro caso es $E > 7,5$ (una cola) y $E > 10$ (dos colas), lo cual es bastante discrepante con la cota de 5,5 del Modelo III.

7ª) $E^*(n_1 = \infty; K)$ sirve también ahora como una cota conservadora para cada valor de K en cualquier n_1 . Por ejemplo, para una cola en $K = 1; 1,5; 2; 5$ y 10 son –por la (4.26)– $E > 7,5; 6,8; 7,3; 10,2$ y $11,9$ respectivamente, lo cual, salvo para $K < 1,5$, es favorable al Modelo I (por dar valores más pequeños que los del Modelo III). Para dos colas los valores son $E > 10; 8,8; 8,7; 10,3$ y $11,4$.

Aún pueden hacerse algunos comentarios extras:

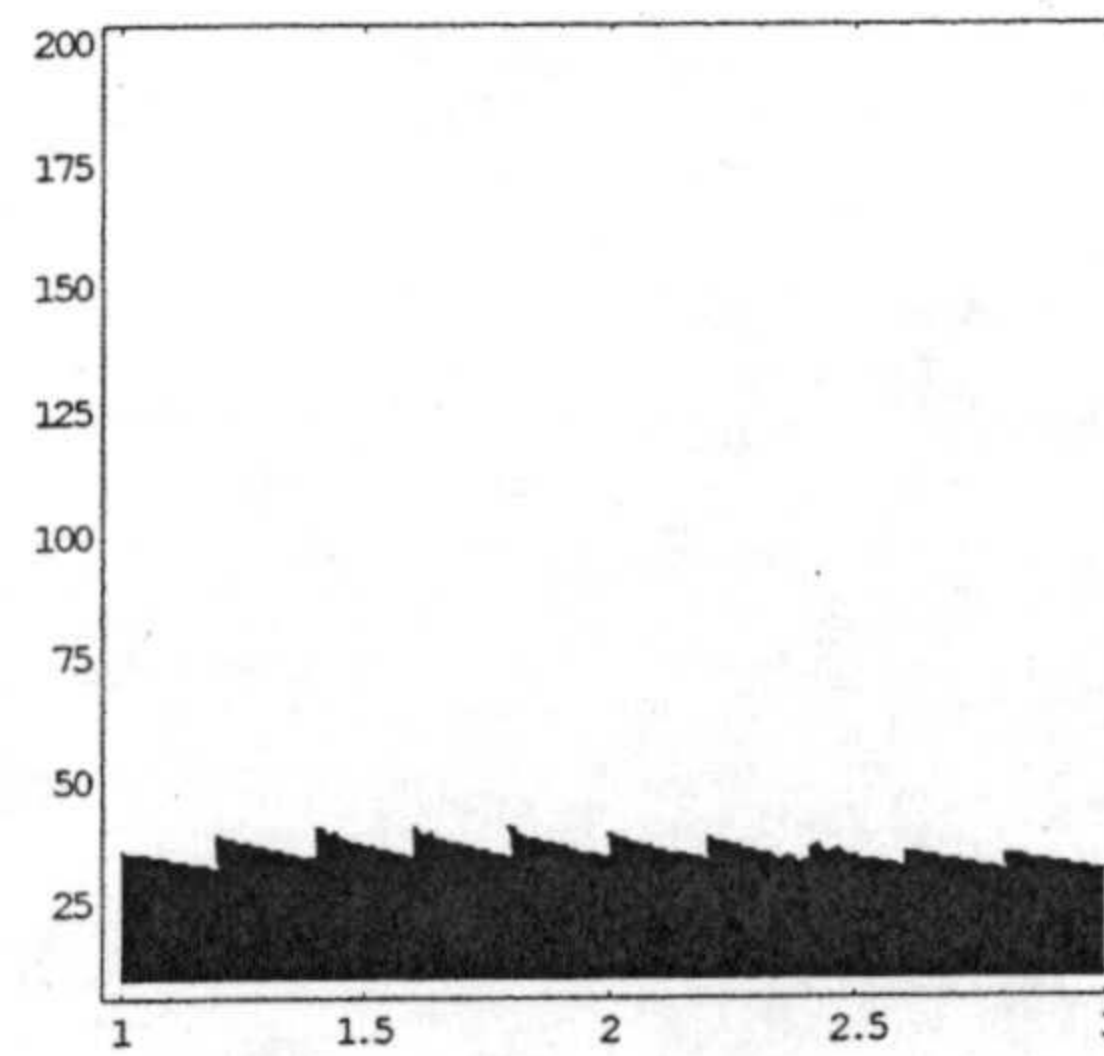
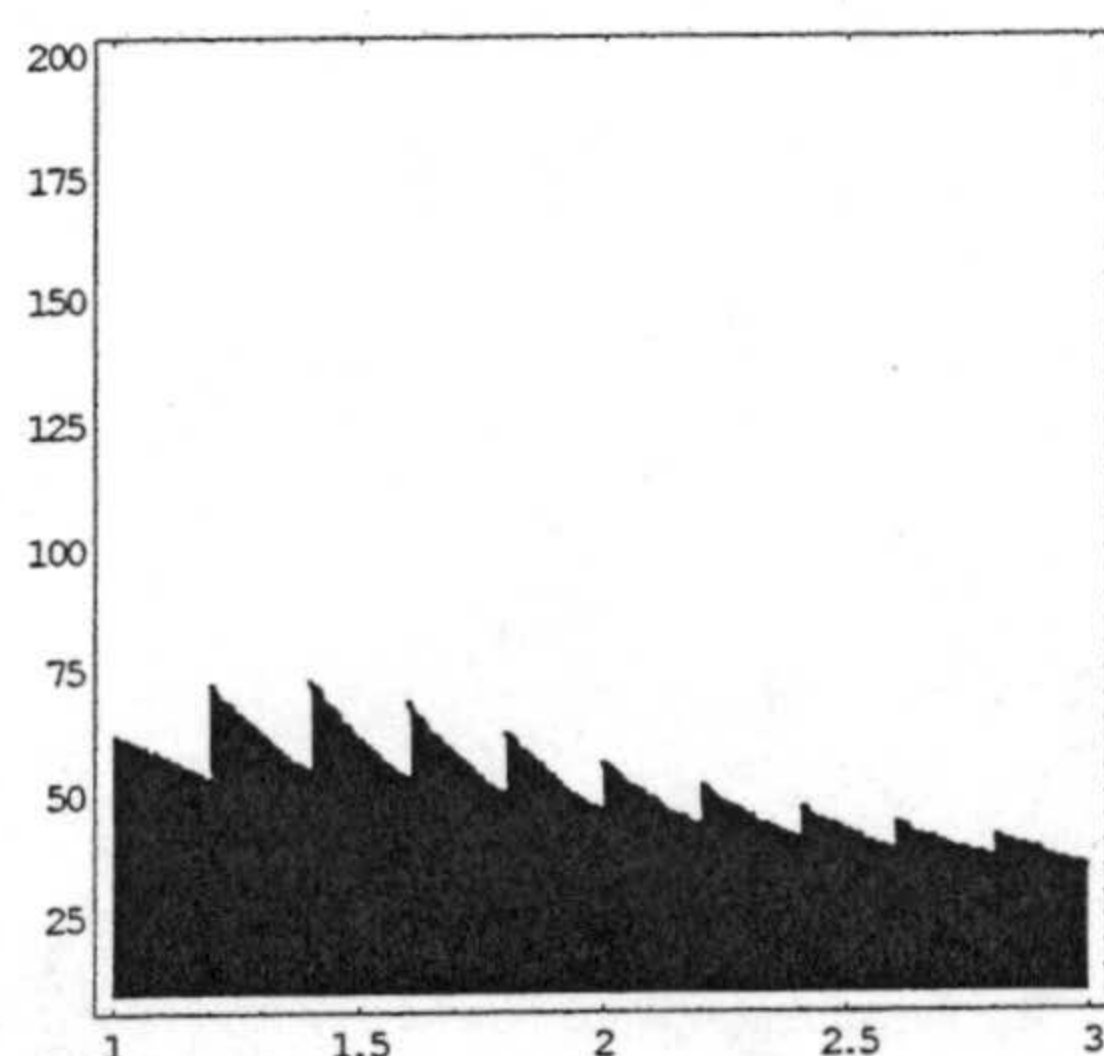
8ª) Para evaluar cómo de peligroso es asumir para el Modelo I la clásica c.d.v. de $E \geq 5$, se da la Figura 4.5. Obsérvese cómo para $n_1 \geq 75$ (50) la regla falla siempre en los tests de una (dos) colas. De hecho por la (4.27), $E \geq 5$ es una c.d.v. válida (cualquiera que sea K) sólo para $n_1 \leq 20$ (21) en los tests de una (dos) colas.

Figura 4.5

Las zonas oscuras se corresponden con los valores de K (eje horizontal) y de n_1 (eje vertical) en donde es válida para el Modelo I la clásica condición de validez de $E \geq 5$

(a) Una cola

(b) Dos colas



9ª) En igualdad de condiciones, la c.d.v. es más fuerte cuanto mayor es el valor de \hat{y}^* para la misma. Observando la (4.64) se ve que en $x = 0$ (es decir en $K = 1$) la c.d.v. es más fuerte en dos colas que en una cola (pues $20,4 > 15,3$), lo cual se mantiene

hasta un determinado valor de K (pues la pendiente de la recta para dos colas es inferior a la pendiente para una cola), a partir del cual los papeles se invierten. El valor de K en que ello sucede está asociado al valor de $x = (K - 1)^2/K$ en el que las dos rectas se cortan, lo que da $x = 3,74$ y $K = 5,56$. Con ello, las c.d.v. para el test de dos colas son más exigentes que en el de una cola para $K < 5,56$, pero menos exigentes para $K > 5,56$ (lo que concuerda con lo observado en 7)).

4.2.7. Discusión y conclusiones

En esta sección se ha visto cómo la validez del test chi-cuadrado RP para tablas 2×2 emergentes del Modelo I, como aproximación del test exacto Y, está muy ligada a la asimetría del estadístico en que se basa (el estadístico $U = x_1y_2 - x_2y_1$), de modo que el test RP es válido si la asimetría es menor que un número dado (que depende del tamaño de muestra y del desequilibrio del marginal más raro).

A efectos prácticos, un modo sencillo de comprobar la validez del test RP es chequeando si se verifica la (4.21) o la (4.22) o la (4.24), todas ellas equivalentes, expresiones que están sujetas a la presentación de datos que indica la primera parte de (4.35) y en las que a y b son las de (4.64), A y B las de (4.65), K alude a n_2/n_1 y C viene dado por la (4.63).

La forma más tradicional de presentar la c.d.v. es como en (4.24) $-E > E^*$ con E^* una función de n_1 y de K que generalmente crece con ellas pero que alcanza un máximo absoluto de 14,38 (una cola) o 13,03 (dos colas) que puede servir como c.d.v. generales y conservadoras (y que son muy diferentes del tradicional valor de 5). Una c.d.v. intermedia a la muy conservadora anterior y a la muy compleja de (4.24), es $E > E^*$ ($n_1 = \infty$; K) dada por la (4.26). Un caso particular que no es preciso evaluar es cuando todos los marginales son iguales: el test RP es entonces siempre válido. Otros valores particulares han sido que para $K = 1$, la c.d.v. general y conservadora es $E > 7,5$ (una cola) y $E > 10$ (dos colas).

De modo general, la c.d.v. para el test de dos colas es más débil que en el de una cola sólo para valores muy desequilibrados de las n_i ($K > 5,6$). Efectuando similar análisis que entonces para evaluar comparativamente las c.d.v. de los Modelos I y III —es

decir comparando las (4.64) y (4.18) o (4.19)— se deduce que las c.d.v. del Modelo I son siempre más exigentes que las del Modelo III en el caso del test de dos colas; para una cola son más exigentes sólo cuando $K > 1,55$.

Para obtener los resultados anteriores se ha determinado experimentalmente la ecuación de una recta $y = a + bx$, en donde x e y dependen sólo de los marginales de la tabla y las constantes a y b tienen varios significados físicos. Por un lado $[a]^-$ es el valor mínimo del marginal más pequeño para que el test chi-cuadrado pueda ser válido en una tabla no equilibrada, en tanto que $[a]^-/2$ es el valor mínimo de E para la validez del test cuando uno de los marginales está equilibrado. Por otro lado, b es el máximo valor posible de E^* . Ambos, a su vez, están relacionados con la máxima asimetría permitida, pues $A = a/b$ y $B = 1/b$, con B la máxima asimetría que se le permite a una tabla para que el test chi-cuadrado sea válido y A una constante reductora de aquella (que actúa proporcionalmente sobre la asimetría relativa del marginal más asimétrico). Los valores de B son del orden de 0,07.

CUADROS

Cuadro I

ORDENACIONES EN LOS TESTS EXACTOS

Definición conceptual y real de los métodos de ordenación utilizados en ésta memoria

Método	Concepto	Estadístico de orden
Z	χ^2 clásica sin c.p.c.	$-(x_1y_2-x_2y_1)^2 n/a_1a_2n_1n_2$
Y	χ^2 con la c.p.c. de Yates	$-\{ x_1y_2-x_2y_1 - n/2 \}^2 n/a_1a_2n_1n_2$
PH1	χ^2 de Pirie-Hamdan (Modelo I)	$-\{ x_1y_2-x_2y_1 - 0,5 \}^2 n/a_1a_2n_1n_2$
PH2	χ^2 de Pirie-Hamdan (Modelo II)	$-\{ x_1y_2-x_2y_1 - n/4 \}^2 n/a_1a_2n_1n_2$
PH3	χ^2 con la c.p.c. nueva (Modelo I)	$-\{ x_1y_2-x_2y_1 - [n^2/8]^{-1/\bar{n}} \}^2 n/a_1a_2n_1n_2$
F1	Valor P de Fisher (1 cola)	$F1(x_1) = P(x_1) + \dots + P(s)$
F2	Valor P de Fisher (2 colas)	$F2(x_1) = P(r) + \dots + P(x'_1) + F1(x_1)$
F	El conjunto F1/F2	F1 para una cola; F2 para dos colas.
FC	F con valores nominales	A usar como test condicionado.
FM1	P-mid Fisher (1 cola)	$FM1(x_1) = F1(x_1) - P(x_1)/2$
FM2	P-mid Fisher (2 colas)	$FM2(x_1) = F2(x_1) - P(x_1)/2$ si $P(x_1) > P(x'_1)$ $= F2(x_1) - P(x_1)$ si $P(x_1) = P(x'_1)$
FM	El conjunto FM1/FM2	FM1 para una cola; FM2 para dos colas.
M	Barnard exacto	α^* : máximo $\alpha(p,q)$ en $0 < p, q < 1$ en la RC alcanzada
M'	Barnard estimado	$\hat{\alpha}^*$: máximo $\alpha(\hat{p}, \hat{q})$ en la RC alcanzada; \hat{p} y \hat{q} en el último punto.
NOTAS		
<p>(1) La notación para la tabla observada es la tradicional de la Tabla 1.1.(a). Se entiende que tal tabla está reordenada de modo que $n_1 \leq n_2$ y $x_1y_2 - x_2y_1 > 0$. En todos los casos el orden de entrada de los puntos en la RC es de menor a mayor valor del estadístico de orden aquí definido.</p> <p>(2) \bar{n} es el indicado en la primera columna de la Tabla 3.2.</p> <p>(3) $r = \text{Máx} \{0; a_1 - n_2\}$,, $s = \text{Mín} \{a_1, n_1\}$,, $\hat{p} = a_1/n$; $\hat{q} = n_1/n$</p> <p>(4) x'_1 es el primer entero de la otra cola que es tan improbable ó más que x_1.</p>		

Cuadro II
MÉTODOS APROXIMADOS
 Origen conceptual y nombre de los 53/57 métodos asintóticos estudiados

Tipo	Estadístico de origen	Sin c.p.c	Con c.p.c. del tipo: Salto...				
			Individual	Promedio			Asintótico
			P	PH	CH	HH	PHA
Chi-Cuadrado	Z	Z	ZP	ZPHx	ZCHx	ZHHx	ZPHA
	P	P	PP	PPHx	PCHx	PHHx	PPHA
	R	R	RP	RPHx	RCHx	RHHx	RPHA
							Versiones x = blanco, 0, 1 ó 2
		Tradicional	Armitage (A)	Mantel (M)			
P-mid	FM	FM	FMA	FMM			
	FMN	FMN	FMNA	FMNM			
							Válidos sólo para dos colas
Claves de las columnas							
P	≡ Salto individual de $U = x_1y_2 - x_2y_1$.						
PH	≡ Salto promedio de U.						
CH	≡ Salto promedio de χ^2 .						
HH	≡ Salto promedio de χ .						
PHA	≡ Salto promedio asintótico de U.						
Armitage	≡ Doblar el P de una cola.						
Mantel	≡ El P de cada cola se calcula por separado.						
Versiones x	≡ Valor estimado del número de saltos y aproximaciones al mismo.						
Claves de las filas							
Z	≡ Chi-cuadrado clásica: factor n.						
P	≡ Chi-cuadrado clásica: factor $n^2/(n-1)$.						
R	≡ Chi-cuadrado clásica: factor (n-1).						
FM	≡ P-mid del test exacto de Fisher (tradicional).						
FMN	≡ P-mid del test exacto de Fisher (nuevo).						

Cuadro III
MÉTODOS APROXIMADOS
 Definición de los métodos asintóticos estudiados en ésta memoria

(a)
Métodos chi-cuadrado

Fórmulas básicas			
$\chi_{XCHx}^2 = \chi_x^2 - \frac{A}{2\bar{n}} \quad ,, \quad \chi_{XP}^2 = \frac{\{ x_1y_2 - x_2y_1 - 0,5\}^2}{a_1a_2n_1n_2} A \quad ,, \quad \chi_{XPHA}^2 = \frac{\left\{ x_1y_2 - x_2y_1 - \frac{6}{n} \right\}^2}{a_1a_2n_1n_2} A$			
$\chi_x^2 = \frac{(x_1y_2 - x_2y_1)^2}{a_1a_2n_1n_2} A \quad ,, \quad \chi_{XHHx}^2 = \left(\chi_x - \frac{\sqrt{A}}{2\bar{n}} \right)^2 \quad ,, \quad \chi_{XPHx}^2 = \frac{\{ x_1y_2 - x_2y_1 - [n^2/8]^{-1/\bar{n}}\}^2}{a_1a_2n_1n_2} A$			
X	A	ñ	x
Z	n	Como en Tabla 3.2: 1º Columna para PH; 2º para CH ó HH	Blanco
P	n ² /(n-1)		0
R	n-1		1
			2

(b)
Métodos P-mid

Fórmulas básicas			
Si $x_1 > E = a_1n_1/n$, $r = \text{Máx}\{0; a_1 - n_2\}$, $s = \text{Mín}(a_1, n_1)$ y $x'_1 < E$ es tal que $P(x'_1) \leq P(x_1)$ pero $P(x'_1 + 1) > P(x_1)$, entonces:			
$\text{FM1}(x_1) = P(s) + P(s-1) + \dots + P(x_1 + 1) + P(x_1)/2$			
$\text{FM1}(x'_1) = P(r) + P(r+1) + \dots + P(x'_1 - 1) + P(x'_1)/2$			
$\text{FM2}(x_1) = \begin{cases} \{P(r) + \dots + P(x'_1)\} + \{P(x_1 + 1) + \dots + P(s)\} + P(x_1)/2 & \text{si } P(x_1) > P(x'_1) \\ \{P(r) + \dots + P(x'_1)\} + \{P(x_1 + 1) + \dots + P(s)\} + 0 & \text{si } P(x_1) = P(x'_1) \end{cases}$			
con P(i) en la hipergeométrica.			
Método			
Tradicional		Armitage (A)	Mantel (M)
1 cola	2 colas	2 colas	2 colas
FM		FMA	FMM
FM1(x ₁)	FM2(x ₁)	2FM1(x ₁)	FM1(x ₁) + FM1(x' ₁)
FMN			
Se definen como los FM, pero basándose en que ahora:			
$\text{FMN1/2} = \frac{\text{FM1/2}(x_1) + \text{FM1/2}(\text{más cercano para } n_1 \text{ ó } a_1 \text{ fijos})}{2}$			

ANEXO

REGIONES CRÍTICAS PARA EL TEST DE INDEPENDENCIA EN TABLAS 2×2

(obtenidas por el test original de Barnard)

Presentación de los datos

A ^B	SÍ	NO	Total	Condiciones
SÍ	x_1	y_1	n_1	$N \leq 50$ $n_1 = \text{Min}(n_1, n_2, a_1, a_2)$ $\hat{p}_1 = x_1/n_1 < x_2/n_2 = \hat{p}_2$ $(\equiv x_1 y_2 < x_2 y_1)$
NO	x_2	y_2	n_2	
Total	a_1	a_2	N	

Uso de las tablas

La significación se alcanza si es:

$$x_2 \geq C(N, n_1, x_1)$$

con $C(N, n_1, x_1)$ el número que se indica en las tablas para la combinación de valores de N, n_1, x_1 , número de colas del test y error α (en ese orden). El símbolo "--" (o la no presencia del valor C) indica ausencia de límite.

N = 3						
UNA COLA			DOS COLAS			
n1	x1	10%	5%	1%	10%	5%
1	0	2	--	--	--	--
2	0	2	--	--	--	--
3	0	2	--	--	--	--
4	0	2	--	--	--	--
5	0	2	--	--	--	--
6	0	2	--	--	--	--
7	0	2	--	--	--	--
8	0	2	--	--	--	--
9	0	2	--	--	--	--
10	0	2	--	--	--	--
11	0	2	--	--	--	--
12	0	2	--	--	--	--
13	0	2	--	--	--	--
14	0	2	--	--	--	--
15	0	2	--	--	--	--
16	0	2	--	--	--	--
17	0	2	--	--	--	--
18	0	2	--	--	--	--
19	0	2	--	--	--	--
20	0	2	--	--	--	--
21	0	2	--	--	--	--
22	0	2	--	--	--	--
23	0	2	--	--	--	--
24	0	2	--	--	--	--
25	0	2	--	--	--	--
26	0	2	--	--	--	--
27	0	2	--	--	--	--
28	0	2	--	--	--	--
29	0	2	--	--	--	--
30	0	2	--	--	--	--

N = 14						
UNA COLA			DOS COLAS			
n1	x1	10%	5%	1%	10%	5%
4	0	4	5	7	6	6
5	0	5	5	7	5	5
6	0	6	6	6	6	6
7	0	7	7	7	7	7
8	0	8	8	8	8	8
9	0	9	9	9	9	9
10	0	10	10	10	10	10
11	0	11	11	11	11	11
12	0	12	12	12	12	12
13	0	13	13	13	13	13
14	0	14	14	14	14	14
15	0	15	15	15	15	15
16	0	16	16	16	16	16
17	0	17	17	17	17	17
18	0	18	18	18	18	18
19	0	19	19	19	19	19
20	0	20	20	20	20	20
21	0	21	21	21	21	21
22	0	22	22	22	22	22
23	0	23	23	23	23	23
24	0	24	24	24	24	24
25	0	25	25	25	25	25
26	0	26	26	26	26	26
27	0	27	27	27	27	27
28	0	28	28	28	28	28
29	0	29	29	29	29	29
30	0	30	30	30	30	30

N = 18						
UNA COLA			DOS COLAS			
n1	x1	10%	5%	1%	10%	5%
8	2	6	7	--	7	8
9	0	9	9	9	9	9
10	1	9	9	9	9	9
11	2	7	7	7	7	7
12	3	6	--	--	--	--
13	4	5	--	--	--	--
14	5	4	--	--	--	--
15	6	3	--	--	--	--
16	7	2	--	--	--	--
17	8	1	--	--	--	--
18	9	0	--	--	--	--
19	10	0	--	--	--	--
20	11	0	--	--	--	--
21	12	0	--	--	--	--
22	13	0	--	--	--	--
23	14	0	--	--	--	--
24	15	0	--	--	--	--
25	16	0	--	--	--	--
26	17	0	--	--	--	--
27	18	0	--	--	--	--
28	19	0	--	--	--	--
29	20	0	--	--	--	--
30	21	0	--	--	--	--

N = 21						
UNA COLA			DOS COLAS			
n1	x1	10%	5%	1%	10%	5%
10	1	9	9	9	9	9
11	2	8	8	8	8	8
12	3	7	8	--	8	--
13	4	6	--	--	--	--
14	5	5	--	--	--	--
15	6	4	--	--	--	--
16	7	3	--	--	--	--
17	8	2	--	--	--	--
18	9	1	--	--	--	--
19	10	0	--	--	--	--
20	11	0	--	--	--	--
21	12	0	--	--	--	--
22	13	0	--	--	--	--
23	14	0	--	--	--	--
24	15	0	--	--	--	--
25	16	0	--	--	--	--
26	17	0	--	--	--	--
27	18	0	--	--	--	--
28	19	0	--	--	--	--
29	20	0	--	--	--	--
30	21	0	--	--	--	--

N = 24						
UNA COLA			DOS COLAS			
n1	x1	10%	5%	1%	10%	5%
6	3	15	15	--	15	--
7	0	7	7	8	7	7
8	1	9	10	12	9	10
9	2	10	11	14	11	13
10	3	13	14	--	14	14
11	4	12	--	--	--	--
12	5	11	--	--	--	--
13	6	10	--	--	--	--
14	7	9	--	--	--	--
15	8	8	--	--	--	--
16	9	7	--	--	--	--
17	10	6	--	--	--	--
18	11	5	--	--	--	--
19	12	4	--	--	--	--
20	13	3	--	--	--	--
21	14	2	--	--	--	--
22	15	1	--	--	--	--
23	16	0	--	--	--	--
24	17	0	--	--	--	--
25	18	0	--	--	--	--
26	19	0	--	--	--	--
27	20	0	--	--	--	--
28	21	0	--	--	--	--
29	22	0	--	--	--	--
30	23	0	--	--	--	--

N = 4						
UNA COLA			DOS COLAS			
n1	x1	10%	5%	1%	10%	5%
1	0	3	--	--	--	--
2	0	2	2	--	2	2
3	0	2	2	--	2	2
4	0	2	2	--	2	2

N = 5						
UNA COLA			DOS COLAS			
n1	x1	10%	5%	1%	10%	5%
1	0	4	4	--	4	4
2	0	2	3	--	3	--
3	0	2	3	--	3	--
4	0	2	3	--	3	--
5	0	2	3	--	3	--

N = 19						
UNA COLA			DOS COLAS			
n1	x1	10%	5%	1%	10%	5%
1	0	13	17	18	13	17
2	0	9	11	16	11	12
3	0	6	8	12	9	10
4	0	6	7	10	7	9
5	0	5	6	9	6	8
6	0	4	5	8	5	7
7	0	3	4	7	4	6
8	0	2	3	6	3	5
9	0	1	2	5	2	4
10	0	0	1	4	1	3
11	0	0	0	3	0	2
12	0	0	0	2	0	1
13	0	0	0	1	0	0
14	0	0	0	0	0	0
15	0	0	0	0	0	0
16	0	0	0	0	0	0
17	0	0	0	0	0	0
18	0	0	0	0	0	0
19	0	0	0	0	0	0

N = 22						
UNA COLA			DOS COLAS			
n1	x1	10%	5%	1%	10%	5%
1	0	16	20	21	16	20
2	0	11	13	18	12	13
3	0	7	11	14	11	12
4	0	7	8	12	8	11
5	0	5	8	9	8	8
6	0	4	7	8	7	7
7	0	3	6	7	6	6
8	0	2	5	6	5	5
9	0	1	4	5	4	4
10	0	0	3	4	3	3
11	0	0	2	3	2	2
12	0	0	1	2	1	1
13	0	0	0	1	0	0
14	0	0	0	0	0	0
15	0	0	0	0	0	0
16	0	0	0	0	0	0
17	0	0	0	0	0	0
18	0	0	0	0	0	0
19	0	0	0	0	0	0
20	0	0	0	0	0	0
21	0	0	0	0	0	0
22	0	0	0	0	0	0

N = 25						
UNA COLA			DOS COLAS			
n1	x1	10%	5%	1%	10%	5%
1	0	18	22	24	18	22
2	0	11	15	20	14	15
3	0	9	11	16	13	14
4	0	7	10	13	9	12
5	0	5	7	11	8	11
6	0	4	6	10	7	10
7	0	3	5	9	6	9
8	0	2	4	8	5	8
9	0	1	3	7	4	7
10	0	0	2	6	3	6
11	0	0	1	5	2	5
12	0	0	0	4	1	4
13	0	0	0	3	0	3
14	0	0	0	2	0	2
15	0	0	0	1	0	1
16	0	0	0	0	0	0
17	0	0	0	0	0	0
18	0	0	0	0	0	0
19	0	0	0	0	0	0
20	0	0	0	0	0	0
21	0	0	0	0	0	0
22	0	0	0	0	0	0
23	0	0	0	0	0	0
24	0	0	0	0	0	0
25	0	0	0	0	0	0

N = 6						
UNA COLA			DOS COLAS			
n1	x1	10%	5%	1%	10%	5%
1	0	5	5	--	5	5
2	0	3	4	--	4	4
3	0	3	3	3	3	3
4	0	3	3	3	3	3
5	0	3	3	3	3	3
6	0	3	3	3	3	3

N = 15						
UNA COLA			DOS COLAS			
n1	x1	10%	5%	1%	10%	5%
1	0	11	13	--	12	13
2	0	6	9	12	8	9
3	0	6	6	10	7	8
4	0	4	6	7	6	7
5	0	4	5	6	5	6
6	0	3	4	5	4	5
7	0	2	3	4	3	4
8	0	1	2	3	2	3
9	0	0	1	2	1	2
10	0	0	0	1	0	1
11	0	0	0	0	0	0
12	0	0	0	0	0	0
13	0	0	0	0	0	0
14	0	0	0	0	0	0
15	0	0	0	0		

N = 26

		UNA COLA			DOS COLAS		
n1	x1	10%	5%	1%	10%	5%	1%
8	0	8	8	8	8	8	10
	1	7	10	12	9	10	12
	2	10	11	14	11	13	15
	3	12	14	15	14	15	—
	4	14	—	—	—	—	—
9	0	9	9	9	9	9	9
	1	8	8	11	8	10	11
	2	9	10	13	10	11	14
	3	10	12	14	12	13	14
	4	13	13	—	13	—	—
10	0	10	10	10	10	10	10
	1	9	9	9	9	9	11
	2	8	8	11	10	10	12
	3	10	11	13	10	12	13
	4	12	12	—	12	12	—
11	0	11	11	11	11	11	11
	1	10	10	10	10	10	10
	2	9	9	10	9	9	12
	3	8	10	12	10	11	12
	4	10	11	—	11	11	—
12	0	12	12	12	12	12	12
	1	11	11	11	11	11	11
	2	10	10	10	10	10	10
	3	9	9	11	9	10	11
	4	9	10	—	10	—	—
13	0	13	13	13	13	13	13
	1	12	12	12	12	12	12
	2	11	11	11	11	11	11
	3	10	10	10	10	10	10
	4	9	9	—	9	9	—
	5	8	—	—	—	—	—

N = 28

		UNA COLA			DOS COLAS		
n1	x1	10%	5%	1%	10%	5%	1%
5	3	20	—	—	20	—	—
6	0	6	7	10	9	10	13
	1	11	12	16	13	14	16
	2	14	17	19	16	18	20
	3	18	19	—	19	19	—
7	0	7	7	10	7	10	12
	1	9	12	15	11	12	15
	2	13	14	18	14	16	18
	3	16	17	—	17	17	—
	4	17	—	—	—	—	—
8	0	8	8	8	8	8	11
	1	9	11	13	10	11	14
	2	11	12	15	12	14	16
	3	14	16	17	15	16	17
	4	16	16	—	16	16	—
9	0	9	9	9	9	9	9
	1	8	8	12	8	10	12
	2	9	11	14	11	12	15
	3	11	13	16	13	15	16
	4	14	15	—	15	—	—
10	0	10	10	10	10	10	10
	1	9	9	11	9	9	12
	2	8	11	12	10	11	13
	3	11	13	14	12	13	15
	4	13	13	—	14	14	—
	5	13	—	—	—	—	—
11	0	11	11	11	11	11	11
	1	10	10	10	10	10	10
	2	9	9	11	9	11	12
	3	10	11	13	11	13	14
	4	11	12	—	12	13	—
	5	12	—	—	—	—	—
12	0	12	12	12	12	12	12
	1	11	11	11	11	11	11
	2	10	10	10	10	10	10
	3	9	9	12	9	11	13
	4	10	12	—	12	12	—
	5	11	—	—	—	—	—
13	0	13	13	13	13	13	13
	1	12	12	12	12	12	12
	2	11	11	11	11	11	11
	3	10	10	10	10	10	10
	4	9	9	—	9	9	—
	5	8	—	—	—	—	—

N = 29

		UNA COLA			DOS COLAS		
n1	x1	10%	5%	1%	10%	5%	1%
12	5	12	—	—	12	—	—
13	0	13	13	13	13	13	13
	1	12	12	12	12	12	12
	2	11	11	11	11	11	11
	3	10	10	12	10	10	12
	4	9	10	12	11	11	—
	5	11	—	—	11	—	—
14	0	14	14	14	14	14	14
	1	13	13	13	13	13	13
	2	12	12	12	12	12	12
	3	11	11	11	11	11	11
	4	10	10	11	10	10	—
	5	9	10	—	10	—	—
.....							
N = 30							
**	**	***	**	**	***	**	**
1	0	22	26	29	21	25	29
2	0	14	18	24	16	18	23
	1	26	27	—	26	27	—
3	0	11	13	19	15	16	19
	1	20	22	26	20	22	26
	2	25	—	—	25	—	—
4	0	8	11	17	13	15	17
	1	17	19	23	18	20	23
	2	22	24	—	22	24	—
5	0	7	9	13	9	12	16
	1	14	16	20	16	17	19
	2	18	20	23	19	20	23
	3	22	—	—	22	—	—
6	0	6	8	12	9	10	14
	1	12	15	17	15	16	18
	2	16	18	21	17	18	21
	3	20	21	—	20	21	—
7	0	7	7	10	9	10	12
	1	10	12	16	11	14	17
	2	13	16	19	15	16	19
	3	16	19	20	18	20	—
	4	19	—	—	—	—	—
8	0	8	8	10	8	8	11
	1	10	11	14	11	12	15
	2	12	13	17	14	15	17
	3	15	15	19	16	18	19
	4	17	18	—	18	—	—
9	0	9	9	9	9	9	10
	1	8	10	13	10	11	14
	2	10	12	15	12	13	16
	3	13	15	17	14	16	18
	4	16	17	—	17	17	—
10	0	10	10	10	10	10	10
	1	9	9	11	9	9	12
	2	9	11	14	11	12	14
	3	12	13	16	13	14	17
	4	14	15	—	15	16	—
	5	15	—	—	—	—	—

N = 31

		UNA COLA			DOS COLAS		
n1	x1	10%	5%	1%	10%	5%	1%
6	0	6	8	12	9	11	15
	1	12	14	18	15	17	18
	2	17	19	22	17	19	22
	3	20	22	—	21	21	—
7	0	7	7	11	9	10	12
	1	11	13	16	11	14	17
	2	14	15	19	16	17	20
	3	17	19	21	19	21	21
	4	20	—	—	20	—	—
8	0	8	8	10	8	8	12
	1	10	11	15	11	13	16
	2	12	14	17	15	16	18
	3	15	18	20	17	18	20
	4	18	18	—	19	—	—
9	0	9	9	9	9	9	11
	1	8	10	13	10	11	14
	2	10	13	16	12	13	16
	3	14	15	18	15	16	19
	4	17	18	—	18	18	—
	5	17	—	—	—	—	—
10	0	10	10	10	10	10	10
	1	9	9	12	9	11	13
	2	10	11	14	11	13	15
	3	12	14	16	13	15	17
	4	14	16	—	16	17	—
	5	16	—	—	—	—	—
11	0	11	11	11	11	11	11
	1	10	10	10	10	10	10
	2	9	10	13	11	12	14
	3	11	12	15	12	14	16
	4	13	14	16	14	15	16
	5	15	15	—	15	—	—
12	0	12	12	12	12	12	12
	1	11	11	11	11	11	11
	2	10	10	12	10	10	13
	3	9	11	13	11	12	15
	4	12	13	15	13	14	15
	5	13	14	—	14	14	—
13	0	13	13	13	13	13	13
	1	12	12	12	12	12	12
	2	11	11	11	11	11	11
	3	10	10	13	10	11	14
	4	13	14	16	14	15	16
	5	15	15	—	15	—	—
14	0	14	14	14	14	14	14
	1	13	13	13	13	13	13
	2	12	12	12	12	12	12
	3	11	11	11	11	11	11
	4	10	10	10	10	10	10
	5	11	12	14	12	13	14
	6	12	—	—	—	—	—
15	0	15	15	15	15	15	15
	1	14	14	14	14	14	14
	2	13	13	13	13	13	13
	3	12	12	12	12	12	12
	4	11	11	11	11	11	11
	5	10	10	10	10	10	10
	6	10	10	—	10	10	—
	7	11	11	11	11	11	11
	8	12	12	12	12	12	12
	9	13	13	13	13	13	13
	10	14	14	14	14	14	14
	11	15	15	15	15	15	15
	12	16	16	16	16	16	16
	13	17	17	17	17	17	17
	14	18	18	18	18	18	18
	15	19	19	19	19	19	19
	16	20	20	20	20	20	20
	17	21	21	21	21	21	21
	18	22	22	22	22	22	22
	19	23	23	23	23	23	23
	20	24	24	24	24	24	24
	21	25	25	25	25	25	25
	22	26	26	26	26	26	26
	23	27	27	27	27	27	27
	24	28	28	28	28	28	28
	25	29	29	29	29	29	29
	26	30	30	30	30	30	30
	27	31	31	31	31	31	31
	28	32	32	32	32	32	32
	29	33	33	33	33	33	33

N = 38

		UNA COLA				DOS COLAS				
n1	x1	10%	5%	1%	10%	5%	1%	10%	5%	1%
17	6	13	14	15	14	15	—			
	7	14	14	—	14	—	—			
18	0	18	18	18	18	18	18			
	1	17	17	17	17	17	17			
	2	16	16	16	16	16	16			
	3	15	15	15	15	15	15			
	4	14	14	14	14	14	14			
	5	13	13	14	13	13	15			
	6	12	12	—	12	14	—			
	7	13	—	—	—	—	—			
19	0	19	19	19	19	19	19			
	1	18	18	18	18	18	18			
	2	17	17	17	17	17	17			
	3	16	16	16	16	16	16			
	4	15	15	15	15	15	15			
	5	14	14	14	14	14	14			
	6	13	13	13	13	13	—			
	7	12	12	—	12	—	—			

N = 39

		UNA COLA				DOS COLAS				
n1	x1	10%	5%	1%	10%	5%	1%	10%	5%	1%
16	0	16	16	16	16	16	16			
	1	15	15	15	15	15	15			
	2	14	14	14	14	14	14			
	3	13	13	13	13	13	15			
	4	12	12	15	12	14	16			
	5	13	14	17	14	15	17			
	6	14	16	17	15	17	—			
	7	15	16	—	16	—	—			
17	0	17	17	17	17	17	17			
	1	16	16	16	16	16	16			
	2	15	15	15	15	15	15			
	3	14	14	14	14	14	14			
	4	13	13	13	13	13	15			
	5	12	12	16	12	14	16			
	6	13	14	16	15	15	—			
	7	14	15	—	15	—	—			
18	0	18	18	18	18	18	18			
	1	17	17	17	17	17	17			
	2	16	16	16	16	16	16			
	3	15	15	15	15	15	15			
	4	14	14	14	14	14	14			
	5	13	13	15	13	13	16			
	6	12	14	15	14	15	—			
	7	14	14	—	14	—	—			
19	0	19	19	19	19	19	19			
	1	18	18	18	18	18	18			
	2	17	17	17	17	17	17			
	3	16	16	16	16	16	16			
	4	15	15	15	15	15	15			
	5	14	14	14	14	14	14			
	6	13	13	14	13	13	—			
	7	12	13	—	13	—	—			

N = 40

		UNA COLA				DOS COLAS				
n1	x1	10%	5%	1%	10%	5%	1%	10%	5%	1%
14	2	12	12	14	12	12	15			
	3	11	13	16	14	15	17			
	4	14	15	18	15	16	19			
	5	15	17	20	17	18	21			
	6	18	19	—	19	20	—			
	7	19	—	—	—	—	—			
15	0	15	15	15	15	15	15			
	1	14	14	14	14	14	14			
	2	13	13	13	13	13	13			
	3	12	12	15	12	14	16			
	4	12	14	17	14	15	18			
	5	14	15	19	16	17	20			
	6	15	18	19	18	19	—			
	7	18	18	—	18	—	—			
16	0	16	16	16	16	16	16			
	1	15	15	15	15	15	15			
	2	14	14	14	14	14	14			
	3	13	13	13	13	13	15			
	4	12	12	15	12	14	17			
	5	13	14	17	14	16	18			
	6	14	16	18	16	17	—			
	7	16	17	—	17	—	—			
17	0	17	17	17	17	17	17			
	1	16	16	16	16	16	16			
	2	15	15	15	15	15	15			
	3	14	14	14	14	14	14			
	4	13	13	15	13	13	16			
	5	12	14	16	13	15	17			
	6	14	15	17	15	16	17			
	7	15	16	—	16	16	—			
	8	15	—	—	—	—	—			
18	0	18	18	18	18	18	18			
	1	17	17	17	17	17	17			
	2	16	16	16	16	16	16			
	3	15	15	15	15	15	15			
	4	14	14	14	14	14	14			
	5	13	13	15	13	13	16			
	6	12	14	16	14	16	16			
	7	14	14	—	14	—	—			
19	0	19	19	19	19	19	19			
	1	18	18	18	18	18	18			
	2	17	17	17	17	17	17			
	3	16	16	16	16	16	16			
	4	15	15	15	15	15	15			
	5	14	14	14	14	14	14			
	6	13	13	14	13	13	—			
	7	12	13	—	13	—	—			

N = 41

		UNA COLA				DOS COLAS				
n1	x1	10%	5%	1%	10%	5%	1%	10%	5%	1%
11	0	11	11	11	11	11	11			
	1	10	12	15	12	13	16			
	2	12	14	18	14	15	20			
	3	15	17	21	18	19	22			
	4	19	20	24	20	21	25			
	5	21	23	25	22	24	25			
	6	24	24	—	24	—	—			
12	0	12	12	12	12	12	12			
	1	11	11	14	11	13	15			
	2	10	13	17	14	15	18			
	3	14	16	20	16	18	20			
	4	17	19	22	19	19	23			
	5	19	21	24	20	23	24			
	6	21	23	—	23	23	—			
13	0	13	13	13	13	13	13			
	1	12	12	12	12	12	14			
	2	11	11	15	13	14	16			
	3	13	15	18	15	16	19			
	4	15	17	20	17	18	21			
	5	17	19	22	18	20	23			
	6	19	21	—	21	22	—			
	7	21	—	—	—	—	—			
14	0	14	14	14	14	14	14			
	1	13	13	13	13	13	13			
	2	12	12	15	12	12	16			
	3	13	14	16	14	15	17			
	4	14	15	19	15	17	20			
	5	16	18	21	17	19	22			
	6	18	20	21	20	21	—			
	7	20	20	—	—	—	—			
15	0	15	15	15	15	15	15			
	1	14	14	14	14	14	14			
	2	13	13	13	13	13	13			
	3	12	12	15	12	12	16			
	4	14	15	18	14	14	14			
	5	16	18	20	16	16	16			
	6	18	19	—	18	—	—			
	7	19	—	—	—	—	—			
16	0	16	16	16	16	16	16			
	1	15	15	15	15	15	15			
	2	14	14	14	14	14	14			
	3	13	13	15	13	13	16			
	4	14	15	18	14	14	14			
	5	16	18	20	16	16	16			
	6	18	20	21	20	21	—			
	7	20	20	—	—	—	—			
17	0	17	17	17	17	17	17			
	1	16	16	16	16	16	16			
	2	15	15	15	15	15	15			
	3	14	14	14	14	14	14			
	4	13	13	15	13	13	16			
	5	14	15	18	15	16	19			
	6	15	16	19	16	18	19			
	7	17	18	—	18	—	—			
	8	18	18	—	18	—	—			
18	0	18	18	18	18	18	18			
	1	17	17	17	17	17	17			
	2	16	16	16	16	16	16			
	3	15	15	15	15	15	15			
	4	14	14	14	14	14	14			
	5	13	13	15	13	13	16			
	6	12	14	16	14	16	16			
	7	15	15	—	15	—	—			
	8	15	—	—	—	—	—			
19	0	19	19	19	19	19	19			
	1	18	18	18	18	18	18			
	2	17	17	17	17	17	17			
	3	16	16	16	16	16	16			
	4	15	15	15	15	15	15			
	5	14	14	14	14	14	14			
	6	13	13	15	13	13	—			
	7	12	14	—	14	—	—			

N = 46										N = 46										N = 47										N = 48										N = 48									
UNA COLA					DOS COLAS					UNA COLA					DOS COLAS					UNA COLA					DOS COLAS					UNA COLA					DOS COLAS														
n1	x1	10%	5%	1%	n1	x1	10%	5%	1%	n1	x1	10%	5%	1%	n1	x1	10%	5%	1%	n1	x1	10%	5%	1%	n1	x1	10%	5%	1%																				
11	0	11	11	11	22	2	20	20	20	15	6	20	22	25	6	3	33	36	39	19	4	15	15	17																									
	1	10	13	17		3	19	19	19		7	22	24	25		4	38	38	38		5	14	15	18																									
	2	13	16	21		4	18	18	18		8	24	24	24		7	0	9	11		6	16	16	20																									
	3	18	21	24		5	17	17	17		16	0	16	16		1	17	21	27		7	17	18	21																									
	4	21	24	27		6	16	16	16		1	15	15	15		2	23	26	32		8	18	20	21																									
	5	24	26	30		7	15	15	17		2	14	14	14		3	28	32	36		9	20	20	20																									
	6	27	29	29		8	14	16	15		3	13	14	17		4	34	36	33		20	0	20	20																									
12	0	12	12	12		9	15	15	15		4	14	16	19		8	0	8	10		1	19	19	19																									
	1	11	11	16	23	0	23	23	23		5	16	18	21		1	16	18	23		2	18	18	18																									
	2	13	15	19		1	22	22	22		6	18	20	23		2	21	23	29		3	17	17	17																									
	3	17	19	22		2	21	21	21		7	20	22	24		3	26	28	33		4	16	16	16																									
	4	19	21	25		3	20	20	20		8	22	23	23		4	30	32	36		5	15	15	17																									
	5	22	24	28		4	19	19	19		17	0	17	17		5	34	35	35		6	14	16	19																									
	6	25	27	28		5	18	18	18		1	16	16	16		9	0	9	9		7	15	17	20																									
	7	27	29	29		6	17	17	17		2	15	15	15		1	13	16	22		8	17	18	19																									
13	0	13	13	13		7	16	16	16		3	14	14	16		2	18	21	26		9	19	19	19																									
	1	12	12	15		8	15	15	15		4	13	15	18		3	23	26	30		21	0	21	21																									
	2	13	14	18		9	14	14	14		5	15	17	20		4	27	29	33		1	20	20	20																									
	3	15	17	21			6	17	18	22		5	31	33	33		2	19	19	19																									
	4	18	20	23			7	19	20	23		10	0	10	10		3	18	18	18																									
	5	20	22	26			8	21	22	22		1	13	15	20		4	17	17	17																									
	6	22	24	27			9	21	21	21		2	16	19	24		5	16	16	16																									
	7	25	26	26			18	0	18	18		3	21	23	28		6	15	15	18																									
14	0	14	14	14			1	17	17	17		4	25	27	30		7	14	17	19																									
	1	13	13	13			2	16	16	16		5	27	30	33		8	16	18	19																									
	2	12	12	16			3	15	15	15		6	31	32	32		9	17	18	18																									
	3	14	16	19			4	14	14	17		11	0	11	11		22	0	22	22																									
	4	16	18	22			5	14	16	19		1	10	14	18		1	21	21	21																									
	5	19	20	24			6	16	17	20		2	15	17	22		2	20	20	20																									
	6	21	23	26			7	17	19	22		3	19	21	25		3	19	19	19																									
	7	23	25	25			8	19	21	21		4	22	25	29		4	18	18	18																									
	8	24	25	25			9	20	21	21		5	26	28	32		5	17	17	17																									
15	0	15	15	15			19	0	19	19		6	28	31	31		6	16	16	16																									
	1	14	14	14			1	18	18	18		7	30	30	30		7	15	15	18																									
	2	13	13	15			2	17	17	17		12	0	12	12		8	14	17	18																									
	3	12	15	18			3	16	16	16		1	11	13	17		9	16	17	17																									
	4	15	16	20			4	15	15	15		2	14	16	20		10	16	16	16																									
	5	17	19	23			5	14	14	18		3	17	20	24		23	0	23	23																									
	6	19	21	24			6	15	16	19		4	20	22	26		1	22	22	22																									
	7	21	23	24			7	16	18	21		5	23	25	29		2	21	21	21																									
	8	23	24	24			8	18	20	20		6	26	28	28		3	20	20	20																									
16	0	16	16	16			9	19	19	19		7	29	29	29		4	19	19	19																									
	1	15	15	15			20	0	20	20		13	0	13	13		5	18	18	18																									
	2	14	14	14			1	19	19	19		1	12	12	16		6	17	17	17																									
	3	13	13	17			2	18	18	18		2	13	14	19		7	16	16	16																									
	4	14	15	19			3	17	17	17		3	16	18	22		8	15	15	15																									
	5	16	18	21			4	16	16	16		4	19	21	24		9	16	16	16																									
	6	18	19	23			5	15	15	15		5	21	23	27		24	0	24	24																									
	7	20	22	23			6	14	16	18		6	24	26	29		1	23	23	23																									
	8	22	22	23			7	15	17	19		7	27	28	28		2	22	22	22																									
17	0	17	17	17			8	16	18	21		14	0	14	14		3	21	21	21																									
	1	16	16	16			9	18	18	18		1	13	13	13		4	20	20	20																									
	2	15	15	15			21	0	21	21		2	12	13	17		5	19	19	19																									
	3	14	14	16			1	20	20	20		3	15	16	20		6	18	18	18																									
	4	13	15	18			2	19	19	19		4	17	19	23		7	17	17	17																									
	5	15	16	19			3	18	18	18		5	20	22	25		8	16	16	16																									
	6	16	18	21			4	17	17	17		6	22	24	27		9	15	15	15																									
	7	18	20	22			5	16	16	16		7	24	26	27																										
	8	20	21	21			6	15	15	17		8	26	26	26		N = 49																									
18	0	18	18	18			7	14	16	18		15	0	15	15		**	**	**	**																									
	1	17	17	17			8	16	17	17		1	14	14	14		1	0	36	41																									
	2	16	16	16			9	17	17	17		2	13	13	16		2	0	23	29																									
	3	15	15	15			22	0	22	22		3	12	15	19		1	42	45	45																									
	4	14	14	17			1	21	21	21		4	16	18	21		3	0	17	22																									
	5	13	16	18			2	20	20	20		5	18	20	24		1	33	37	43																									
	6	15	17	19			3	19	19	19		6	21	22	26		2	44	44	44																									
	7	17	18	21			4	18	18	18		7	22	24	25		4	0	14	18																									
	8	19	20	20			5	17	17	17		8	25	25	25		1	27	32	38																									
	9	19	19	19			6	16	16	16		16	0	16	16		2	38	39	43																									
19	0	19	19	19			7	15	15	17		1	15	15	15		5	0	11	16																									
	1	18	18	18			8	14	16	17		2	14	14	16		1	24	27	33																									
	2	17	17	17			9	16	16	16		3	13	15	18		2	31	35	39																									
	3	16	16	16			23</																																						

N = 49
UNA COLA DOS COLAS

n1	x1	10%	5%	1%	10%	5%	1%
10	5	28	31	34	30	32	34
	6	32	33	---	33	---	---
11	0	11	11	12	11	11	15
	1	12	14	18	14	15	20
	2	14	18	23	18	20	24
	3	20	22	26	22	24	28
	4	23	25	29	25	27	30
	5	26	29	32	28	29	33
	6	29	31	---	31	32	---
	7	31	---	---	---	---	---
12	0	12	12	12	12	12	14
	1	11	13	17	13	15	18
	2	14	16	21	16	18	23
	3	18	20	24	20	23	25
	4	21	23	27	23	24	28
	5	24	26	30	26	27	31
	6	26	29	31	29	30	---
	7	30	30	---	30	---	---
13	0	13	13	13	13	13	13
	1	12	12	16	12	14	17
	2	13	14	19	15	17	21
	3	16	19	23	17	21	24
	4	19	22	25	22	23	26
	5	22	24	28	24	25	29
	6	25	26	30	26	28	30
	7	27	29	---	29	---	---
14	0	14	14	14	14	14	14
	1	13	13	15	13	13	16
	2	12	14	18	15	16	19
	3	15	17	21	16	18	23
	4	18	19	23	21	22	24
	5	20	22	26	22	24	27
	6	23	24	28	24	26	29
	7	25	27	---	27	28	---
	8	27	---	---	---	---	---
15	0	15	15	15	15	15	15
	1	14	14	14	14	14	16
	2	13	13	16	13	15	18
	3	14	16	19	16	17	21
	4	16	18	22	18	20	23
	5	19	21	24	21	22	25
	6	21	23	26	22	24	27
	7	23	25	27	25	27	---
	8	26	26	---	26	---	---
16	0	16	16	16	16	16	16
	1	15	15	15	15	15	15
	2	14	14	16	14	14	17
	3	13	15	18	15	17	19
	4	16	17	21	17	18	22
	5	16	19	23	19	21	24
	6	20	21	25	21	22	26
	7	21	24	26	24	25	---
	8	23	25	---	25	---	---
17	0	17	17	17	17	17	17
	1	16	16	16	16	16	16
	2	15	15	15	15	15	17
	3	14	14	17	14	16	18
	4	13	16	19	16	17	20
	5	16	18	21	18	19	22
	6	18	20	23	19	21	24
	7	20	21	25	21	23	25
	8	22	24	---	24	24	---
	9	23	---	---	---	---	---
18	0	18	18	18	18	18	18
	1	17	17	17	17	17	17
	2	16	16	16	16	16	16
	3	15	15	16	15	15	17
	4	14	14	18	14	17	19
	5	16	17	20	17	18	21
	6	17	18	21	18	20	23
	7	18	20	24	20	22	24
	8	20	22	---	22	23	---
	9	22	---	---	---	---	---
19	0	19	19	19	19	19	19
	1	18	18	18	18	18	18
	2	17	17	17	17	17	17
	3	16	16	16	16	16	16
	4	15	15	17	15	15	18
	5	14	16	19	16	17	20
	6	16	17	20	17	19	21
	7	17	19	22	19	20	23
	8	19	21	---	21	22	---
	9	21	21	---	21	---	---
20	0	20	20	20	20	20	20
	1	19	19	19	19	19	19
	2	18	18	18	18	18	18
	3	17	17	17	17	17	17
	4	16	16	16	16	16	16
	5	15	15	18	15	15	19
	6	14	16	19	17	18	20
	7	16	18	21	18	19	22
	8	18	20	21	19	21	---
	9	19	20	---	20	---	---
21	0	21	21	21	21	21	21
	1	20	20	20	20	20	20
	2	19	19	19	19	19	19
	3	18	18	18	18	18	18
	4	17	17	17	17	17	17
	5	16	16	16	16	16	16
	6	15	15	18	15	15	19
	7	14	16	19	17	18	20
	8	16	18	21	18	19	22
	9	18	20	21	19	21	---
	10	19	20	---	20	---	---
	11	21	21	21	21	21	21
	12	21	21	21	21	21	21
	13	21	21	21	21	21	21
	14	21	21	21	21	21	21
	15	21	21	21	21	21	21
	16	21	21	21	21	21	21
	17	21	21	21	21	21	21
	18	21	21	21	21	21	21
	19	21	21	21	21	21	21
	20	21	21	21	21	21	21
	21	21	21	21	21	21	21

N = 49
UNA COLA DOS COLAS

n1	x1	10%	5%	1%	10%	5%	1%
21	8	17	18	20	18	19	---
	9	18	19	---	19	---	---
	10	18	---	---	---	---	---
22	0	22	22	22	22	22	22
	1	21	21	21	21	21	21
	2	20	20	20	20	20	20
	3	19	19	19	19	19	19
	4	18	18	18	18	18	18
	5	17	17	17	17	17	17
	6	16	16	16	16	16	18
	7	15	15	18	15	17	19
	8	16	17	19	17	18	---
	9	17	18	---	18	18	---
	10	17	---	---	---	---	---
23	0	23	23	23	23	23	23
	1	22	22	22	22	22	22
	2	21	21	21	21	21	21
	3	20	20	20	20	20	20
	4	19	19	19	19	19	19
	5	18	18	18	18	18	18
	6	17	17	17	17	17	17
	7	16	16	18	16	16	18
	8	15	15	18	15	17	---
	9	16	17	---	17	17	---
	10	16	---	---	---	---	---
24	0	24	24	24	24	24	24
	1	23	23	23	23	23	23
	2	22	22	22	22	22	22
	3	21	21	21	21	21	21
	4	20	20	20	20	20	20
	5	19	19	19	19	19	19
	6	18	18	18	18	18	18
	7	17	17	17	17	17	17
	8	16	16	17	16	16	---
	9	15	15	---	16	---	---
	10	15	---	---	---	---	---

N = 50

**	**	***	**	**	***	**	**
1	0	36	42	48	37	42	48
2	0	24	29	40	27	29	39
	1	43	46	---	42	46	---
3	0	18	23	32	25	27	32
	1	34	38	44	33	38	44
	2	45	---	---	45	---	---
4	0	14	18	27	20	24	28
	1	28	33	39	31	33	39
	2	38	41	44	38	40	44
5	0	12	16	24	18	20	27
	1	24	28	35	27	30	35
	2	32	35	40	34	35	40
	3	38	41	---	38	42	---
6	0	10	14	20	16	18	24
	1	20	24	31	25	27	31
	2	27	31	36	31	33	36
	3	34	36	41	35	36	41
	4	40	---	---	39	---	---
7	0	9	11	19	12	17	21
	1	19	22	28	20	25	29
	2	25	27	33	27	30	33
	3	31	33	37	32	33	37
	4	35	38	---	35	37	---
8	0	8	11	16	12	14	20
	1	16	19	25	20	21	26
	2	21	24	30	24	27	31
	3	26	30	34	29	31	33
	4	31	34	38	33	35	38
	5	36	37	---	37	37	---
9	0	9	9	14	12	13	18
	1	14	17	22	18	20	25
	2	19	22	27	22	24	29
	3	25	27	31	27	29	32
	4	29	31	35	30	32	35
	5	32	35	---	34	36	---
	6	35	---	---	---	---	---
10	0	10	10	13	10	13	16
	1	13	15	21	15	18	22
	2	17	20	25	20	22	27
	3	21	25	29	25	27	30
	4	25	28	32	28	30	33
	5	29	31	35	31	32	35
	6	33	34	---	34	---	---
	7	32	---	---	---	---	---
11	0	11	11	13	11	11	15
	1	12	14	19	14	16	21
	2	15	18	23	19	21	25
	3	20	22	26	22	24	28
	4	24	26	30	26	27	31
	5	27	29	33	28	31	33
	6	29	32	---	32	32	---
	7	32	---	---	---	---	---
12	0	12	12	12	12	12	14
	1	11	13	18	14	15	19
	2	15	16	21	16	19	23
	3	19	20	25	21	23	26
	4	21	24	28	24	25	29
	5	25	26	31	26	27	32
	6	27	30	32	29	32	---
	7	31	31	---	31	---	---
13	0	13	13	13	13	13	13
	1	12	12	16	12	14	17
	2	13	15	19	15	17	21

REFERENCIAS

REFERENCIAS

- AITCHISON, J. and BACON-SHONE, J. (1981). 'Bayesian risk ratio analysis'. *The American Statistician* 35 (4), 254-257.
- ALTHAM, P.M.E. (1969). 'Exact Bayesian Analysis of a 2×2 Contingency Table, and Fisher's "Exact" Significance Test'. *Journal of the Royal Statistical Society B* 31, 261-269.
- AWAD, A.M. and SARIE, T. (1987). 'A comparative study of tests of independence'. *Commun. Statist. - Simula. and Comp.* 16 (3), 645-671.
- BALLATORI, E. (1987). 'Generalizing Fisher's exact test'. *Metron* 45 (3-4), 90-114.
- BARNARD, G.A. (1945). 'A new test for 2×2 tables'. *Nature*, 156, 1777 and 783-784.
- BARNARD, G.A. (1947). 'Significance tests for 2×2 tables'. *Biometrika* 34, 123-138.
- BARNARD, G.A. (1949). 'Statistical Inference'. *Journal of the Royal Statistical Society B* 11, 115-139.
- BARNARD, G.A. (1982). 'Conditionality versus similarity in the analysis of 2×2 tables'. *Statistics and probability: Essays in Honor of C.R. Rao*. North Holland Publishing Company. 59-65.
- BARNARD, G.A. (1989). 'On alleged gains in power from lower P-values'. *Statistics in Medicine* 8 (12), 1469-1477.
- BERGER, R.L. (1994). 'Power comparison of exact unconditional tests for comparing two binomial proportions'. *Institute of Statistics Mimeo Series* No. 2266.
- BERGER, R.L. (1996). 'More powerful tests from confidence interval p values'. *The American Statistician* 50 (4), 314-8.
- BERGER, R.L. (1996). <http://www.stat.ncsu.edu/~berger/tables.htm>.
- BOOS, D.D. and HUGHES-OLIVER, J.M. (2000). 'How Large Does n Have to be for Z and t Intervals?' *The American Statistician* 54(2), 121-128.
- BOSCHLOO, R.D. (1970). 'Raised conditional level of significance for the 2×2 Table when testing the equality of two probabilities'. *Statistica Neerlandica* 24 (1), 1-35.
- CAMILLI, G. (1995). 'The relationship between Fisher's exact test and Pearson's chi-square test: A Bayesian perspective'. *Psychometrika* 60 (2), 305-312.
- CAMILLI, G. and HOPKINS, K.D. (1978). 'Appicability of chi-square to 2×2 contingency tables with small expected cell frequencies'. *Psychological Bulletin* 85, 163-167.
- CAMILLI, G. and HOPKINS, K.D. (1979). 'Testing for association in 2×2 contingency tables with very small samples sizes'. *Psychological Bulletin* 86 (5), 1011-1014.
- COCHRAN, W.G. (1942). 'The 2×2 correction for continuity'. *Iowa State College Journal of Science*, 16, 421-436.
- COCHRAN, W.G. (1954). 'Some methods for strengthening the common χ^2 tests'. *Biometrics* 10, 417-451.
- CONOVER, W.J. (1974). 'Some reasons for not using the Yates' continuity corrections on 2×2 contingency tables'. *Journal of the American Stat. Assoc.* 69, 374-376.
- CORMACK, R.S. and MANTEL, N. (1991). 'Fisher's exact test: the marginal totals as seen from two different angles'. *The Statistician* 40, 27-34.
- COX, D.R. (1958). 'Some problems corrected with statistical inference'. *Ann. Mathe. Stat.* 29, 357-72.
- COX, D.R. (1970). 'The continuity correction'. *Biometrika* 57, 217-219.

- CRESSIE, N. and READ, T.R.C. (1989). 'Pearson's χ^2 and log likelihood ratio statistics G^2 : A comparative review'. *International Stat. Review* 57 (1), 19-43.
- D'AGOSTINO, R.B., CHASE, W. and BELANGER, A. (1988). 'The appropriateness of some common procedures for testing the equality of two independent binomial populations'. *The American Statistician* 42 (3), 198-202.
- DAVIS, A.B. (1993). 'Power of testing proportions in small two-sample studies when sample sizes are equal'. *Statistics in Medicine* 12, 777-787.
- DAVIS, L.J. (1986). 'Exact tests for 2×2 contingency tables'. *The American Statistician* 40 (2), 139-141.
- EDWARDS, W.F. (1963). 'The measure of association in a 2×2 table'. *Journal of the Royal Statistical Society A* 126, 109-114.
- FISHER, R.A. (1925). *Statistical Methods for Research Workers*, 5th ed., Edinburgh: Oliver and Boyd.
- FISHER, R.A. (1935). 'The logic of inductive inference'. *Journal of the Royal Statistical Society A* 98, 39-54.
- FISHER, R.A. (1942). *The Design of Experiment*. Chap. 2. 3rd ed. Edinburgh: Oliver and Boyd.
- GARSDIE, G.R. and MACK, C. (1967). 'Correct confidence limits for the 2×2 homogeneity contingency table with small frequencies'. *The New Journal of Statistics and Operational Research* 3 (2), 1-25.
- GART, J.J. (1966). 'Alternative analyses of contingency tables'. *Journal of the Royal Statistical Society B* 28, 164-179.
- GART, J.J. (1971). 'The comparison of proportions: A review of significance tests, confidence intervals and adjustments for stratification'. *Review of the Inter. Stat. Inst.* 39 (2), 148-169.
- GRIZZLE, J. E. (1967). 'Continuity correction in the χ^2 test for 2×2 tables'. *The American Statistician* 21 (4), 28-32.
- HABER, M. (1980). 'A comparison of some continuity corrections for the chi-squared test on 2×2 tables'. *Journal of the American Stat. Assoc.* 75, 510-515.
- HABER, M. (1982). 'The continuity correction and statistical testing'. *International Statistical Review* 50, 135-144.
- HABER, M. (1986). 'A modified exact test for 2×2 Contingency Tables'. *Biometrical Journal* 28 (4), 455-463.
- HABER, M. (1986). 'An exact unconditional test for the 2×2 comparative trial'. *Psychological Bulletin* 99 (1), 129-132.
- HABER, M. (1987). 'A comparison of some conditional and unconditional exact tests for 2×2 contingency tables'. *Commun. Statist. - Simula. and Comp.* 16 (4), 999-1013.
- HABER, M. (1992). 'On the expected significance probabilities and Bahadur efficiencies of tests for comparing two binomial proportions'. *Journ. Stat. Comput. and Simula.* 43, 243-251.
- HAJEK, P. and HAVRANEK, T. (1978). *Mechanizing Hypothesis Formation*. Springer, Berlin. Heidelberg, p. 163.
- HAVILAND, M.B. (1990). 'Yates' correction for continuity and the analysis of 2×2 contingency tables'. *Statistics in Medicine* 9 (4), 363-383.

- HELLAND, I.S. (1995). 'Simple counterexamples against the conditionality principle'. *The American Statistician* 49 (4), 351-356.
- HINDE, J. and AITKIN, M. (1987). 'Canonical likelihoods: A new likelihood treatment of nuisance parameters'. *Biometrika* 74 (1), 45-58.
- HIRJI, K.F., TAN, S.J. and ELASHOFF, R.M. (1991). 'A quasi-exact test for comparing two binomial proportions'. *Statistics in Medicine* 10, 1137-1153.
- HOWARD, J.V. (1998). 'The 2x2 Table: A discussion from a Bayesian Viewpoint'. *Statistical Science* 13 (4), 351-367.
- IRWIN, J.D. (1935). 'Test of significance for differences between percentages based on small numbers'. *Metron* 12 (2), 84-94.
- JOHNSON, N. L. and KOTZ, S. (1969). *Discrete distributions*. New York: Wiley.
- KEMPTHORNE, O. (1979). 'In dispraise of the exact test:reactions'. *Jour. Stat. Planning Inferences* 3, 199-213.
- KENDALL, M. G. and STUART, A. (1973). *The Advanced Theory of Statistics*, Vol. 2, 2nd. ed. New York: Hafner Pub. Co.
- KOEHLER, K.J. and LARNTZ, K. (1980). 'An empirical investigation of ggodness-of-fit statistics for sparse multinomials'. *Journal of the American Stat. Assoc.* 75, 336-344.
- KROLL, N.E.A. (1989). 'Testing Independence in 2x2 Contingency Tables'. *Journal of Educational Statistics* 14 (1), 47-79.
- KUDÔ, A. and TARUMI, T. (1978). '2x2 tables emerging out of different chance mechanisms'. *Comm. in Stat.- Theo. and Met.* 7 (10), 977-986.
- KULLBACK, S. (1959). *Information Theory and Statistics*. New York: Wiley.
- LANCASTER, H.O. (1952). 'Statistical control of counting experiment'. *Biometrika* 39, 419-422.
- LARNTZ, K. (1978). 'Small-sample comparisons of exact levels for chi-square goodness-of-fit statistics'. *Journal of the American Stat. Assoc.* 73, 253-263.
- LEE, C.C. and SHEN, S. (1994). 'Convergence rates and powers of six power-divergence statistics for testing independence in 2 by 2 contingency table'. *Commun. Statist. - Theory and Meth.* 23 (7), 2113-2126.
- LEHMANN, E.L. (1959). *Testing for Statistical Hypotheses*. New York: John Wiley & Sons.
- LEHMANN, E.L. (1993). 'The Fisher, Neyman-Pearson theories of testing hypotheses: One theory or two?'. *Journal of the American Stat. Assoc.* 88, 1242-1249.
- LEHMANN, E.L. (1994). *Testing Statistical Hypothesis* (2nd ed.). New York: Chapman and Hall. Cap. 3, sec. 1.
- LIDDELL, D. (1976). 'Practical test of 2x2 tables'. *The Statistician* 25 (4), 295-304.
- LITTLE, R.J.A. (1989). 'Testing the equality of two independent binomial proportions'. *The American Statistician* 43 (4), 283-288.
- LLOYD, C.J. (1988). 'Doubling the one-sided p-value in testing independence in 2x2 tables against a two-sided alternative'. *Statistics in Medicine* 7, 1297-1306.
- LOH, WEI-YIN (1989). 'Bounds on the size of the χ^2 -test of independence in a contingency table'. *The Annals of Statistics* 17 (4), 1709-1722.
- MANTEL, N. (1974). 'Comment and a Suggestion'. *Journal of the American Stat. Assoc.* 69, 378-380.

- MANTEL, N. and GREENHOUSE, S.W. (1968). 'What is the continuity correction?'. *The American Statistician* 22 (5), 27-30.
- MARTÍN ANDRÉS, A. (1991). 'A review of classic non-asymptotic methods for comparing two proportions by means of independent samples'. *Commun. Statist. - Simula. and Comp.* 20 (2&3), 551-583.
- MARTÍN ANDRÉS, A. (1996). 'Tablas 2×2: Pearson, Yates, Fisher y cía (VI): La chi-cuadrado incondicionada'. *JANO* 1192, 66-8.
- MARTÍN ANDRÉS, A. and HERRANZ TEJEDOR, I. (1995). 'Is Fisher's exact test very conservative?'. *Computational Statistics and Data Analysis* 19, 579-591.
- MARTÍN ANDRÉS, A. and HERRANZ TEJEDOR, I. (1997). 'On condition for validity of the approximations to Fisher's exact test'. *Biometrical Journal* 39, 935-54.
- MARTÍN ANDRÉS, A. and HERRANZ TEJEDOR, I. (1999). 'On the minimum expected quantity for the chi-square test in 2×2 tables'. To appear in *Journal of Applied Statistics*.
- MARTÍN ANDRÉS, A. and LUNA DEL CASTILLO, J.D. (1989). 'On the sensitivity of Fisher's exact test to minor perturbations in 2×2 contingency tables'. Letters to the Editors. *Statistics in Medicine* 8 (2), 243-245.
- MARTÍN ANDRÉS, A. and LUNA DEL CASTILLO, J.D. (1990). 'P-value for the optimal version of Fisher's exact test in the comparison of two independent proportions'. *Biometrical Journal* 32 (2), 213-227. Corrections in *B. J.* 32 (8), 1018 (1990).
- MARTÍN ANDRÉS, A. and SILVA MATO, A. (1994). 'Choosing the optimal unconditioned test for comparing two independent proportions'. *Computational Statistics and Data Analysis* 17, 555-574.
- MARTÍN ANDRÉS, A. and SILVA MATO, A. (1996). 'Optimal correction for continuity and conditions for validity in the unconditional chi-squared test'. *Computational Statistics and Data Analysis* 21, 609-626. Erratum in 26, 1997, p. 235.
- MARTÍN ANDRÉS, A., HERRANZ TEJEDOR, I. and LUNA DEL CASTILLO, J.D. (1992). 'Optimal correction for continuity in the chi-squared test in 2×2 tables (conditioned method)'. *Commun. Statist. - Simula. and Comp.* 21 (4), 1077-1101.
- MARTÍN ANDRÉS, A., SANCHEZ QUEVEDO, M. J. and SILVA MATO, A. (1998). 'Fisher's mid-p-value arrangement in 2×2 comparative trials'. *Computational Statistics and Data Analysis* 29 (1), 107-115.
- MARTÍN ANDRÉS, A., SILVA MATO, A. and HERRANZ TEJEDOR, I. (1992). 'A critical review of asymptotic methods for comparing two proportions by means of independent samples'. *Commun. Statist. - Simula. and Comp.* 21 (2), 551-586.
- McDONALD, L.L., DAVIS, B.M. and MILLIKEN, G.A. (1977). 'A non-randomized unconditional test for comparing two proportions in a 2×2 contingency table'. *Technometrics* 19, 145-150.
- NEYMAN, J. (1949). 'Contribution to the theory of the χ^2 test'. *Proc. 1st Berkeley Symp.*, 239-273.

- NURMINEN, M. and MUTANEN, P. (1987). 'Exact Bayesian Analysis of two Proportions'. *Scandinavian Journal of Statistics* 14 (1), 67-77.
- ORTÍZ RODRÍGUEZ, I.M. (1993). *Doble Dicotomía en Tablas 2×2*. Memoria de Licenciatura. Universidad de Almería.
- PEARSON, E.S. (1947). 'The choice of statistical tests illustrated on the interpretation of data classed in a 2×2 table'. *Biometrika* 34, 139-167.
- PEARSON, K. (1900). 'On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can reasonably be supposed to have arisen from random sampling'. *Philos. Mag. Series 5*, 50, 157-172.
- PIRIE, W.R. and HAMDAN, M.A. (1972). 'Some revised continuity corrections for discrete distributions'. *Biometrics* 28, 693-701.
- PLACKETT, R.L. (1964). 'The continuity correction in 2×2 tables'. *Biometrika* 51 (3 and 4), 327-337.
- PLACKETT, R.L. (1977). 'The marginal totals of a 2×2 table'. *Biometrika* 64 (1), 37-42.
- REID, N. (1995). 'The roles of conditioning in inference'. *Statistical Sciences* 2 (10), 138-199.
- RICE, W.R. (1988). 'A New Probability Model for Determining Exact p-Values for 2×2 Contingency Tables when Comparing Binomial Proportions'. *Biometrics* 44, 1-22.
- RICHARDSON, J.T.E. (1990). 'Variants of chi-square for 2×2 contingency tables'. *British Journal of Mathematical and Statistic Psychology* 43, 309-326.
- ROUTLEDGE, R.D. (1992). 'Resolving the conflict over Fisher's exact test'. *The Canadian Journal of Statistics* 20 (2), 201-209.
- SAHAI, H. and KHURSHID, A. (1995). 'Formulas and tables for the determination of sample sizes and power in clinical trials for testing differences in proportions for the two-sample design: a review'. *Statistics in Medicine* 15, 1-21.
- SHUSTER, J.J. (1992). 'Exact unconditional tables for significance testing in the 2×2 multinomial trial'. *Statistics in Medicine* 11 (7), 913-922. Correction in 11, 1619 (1992).
- SILVA MATO, A. (1993). *Comparación de dos proporciones independientes por el método incondicionado*. PhD Tesis. Departamento de Estadística e I.O. Universidad Complutense de Madrid (Spain).
- SILVA MATO, A. and MARTÍN ANDRÉS, A. (1995). 'Optimal unconditional tables for comparing two independent proportions'. *Biometrical Journal* 37 (7), 821-836.
- SILVA MATO, A. and MARTÍN ANDRÉS, A. (1997). 'Simplifying the calculation of the P-value for Barnard's test and its derivatives'. *Statistics and Computing* 7, 137-143.
- SILVA MATO, A. and MARTÍN ANDRÉS, A. (1997 and 1999).
<http://www.jiscmail.ac.uk/files/EXACT-STATS/smpexe.uue;>
[http://www.ugr.es/~bioest/SMP.EXE.](http://www.ugr.es/~bioest/SMP.EXE)
- STATXACT-4 for Windows (1998). *Cytel Software Corporation*, Cambridge, Mass.
- SUISSA, S. and SHUSTER, J.J. (1984). 'Are uniformly most powerful unbiased tests really best?'. *The American Statistician* 38, 204-206.

REFERENCIAS

- SUISSA, S. and SHUSTER, J.J. (1985). 'Exact unconditional samples sizes for the 2×2 binomial trial'. *Journal of the Royal Statistical Society A* 148 (4), 317-327.
- SUISSA, S. and SHUSTER, J.J. (1991). 'The 2×2 matched-pairs trial: exact unconditional design and analysis'. *Biometrics* 47, 361-372.
- TOCHER, K.D. (1950). 'Extension of the Neyman-Pearson theory of tests of discontinuous variables'. *Biometrika* 37, 130-144.
- TOMIZAWA, V.S. (1995). 'Double independence models for a 2×2 contingency table with the same classifications and their application to data'. *Allg. Statistisches Archiv* 79, 140-145.
- UPTON, G.J.G. (1982). 'A comparison of alternative tests for the 2×2 comparative trial'. *Journal of the Royal Statistical Society A* 145 (1), 86-105.
- UPTON, G.J.G. (1992). 'Fisher's exact test'. *Journal of the Royal Statistical Society A* 155 (3), 395-402.
- YATES, F. (1934). 'Contingency tables involving small numbers and the χ^2 test'. *Journal of the Royal Statistical Society Suppl.* 1, 217-235.
- YATES, F. (1984). 'Test of significance for 2×2 contingency tables'. *Journal of the Royal Statistical Society A* 147 (3), 426-463.
- ZHU, Y. and REID, N. (1994). 'Information, and sufficiency in the presence of nuisance parameter'. *The Canadian Journal of Statistics* 22 (1), 111-123.