

UNIVERSIDAD COMPLUTENSE DE MADRID

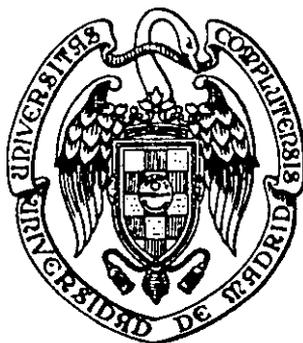
FACULTAD DE MATEMATICAS

Departamento de Estadística e Investigación Operativa

UNIVERSIDAD COMPLUTENSE

**ESTIMADORES DE MINIMA DIVERGENCIA
DE RAO: COMPORTAMIENTO ASINTOTICO
Y APLICACION A CONTRASTES DE
HIPOTESIS**

M^a del Carmen Pardo



21.131

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS MATEMATICAS
Departamento de Estadística e Investigación Operativa



ESTIMADORES DE MINIMA DIVERGENCIA
DE RAO: COMPORTAMIENTO ASINTOTICO
Y APLICACION A CONTRASTES DE
HIPOTESIS

M^a del Carmen Pardo

21.131

Memoria para optar al grado de
Doctor en Ciencias Matemáticas
realizada bajo la dirección de
los Drs. D. Julio Angel Pardo
y D. Leandro Pardo.
Madrid, Marzo 1996.



A mis padres.

INTRODUCCION

El análisis estadístico de datos discretos multinomiales ha despertado en las últimas dos décadas un notable interés como ponen de manifiesto las numerosas publicaciones relacionadas con este tema aparecidas en la literatura estadística. El desarrollo de modelos apropiados, como puede verse en los libros de Cox (1970), Haberman (1974, 1978, 1979), Bishop y otros (1975), Gokhale y Kullback (1978), Upton (1978), Fienberg (1980), Plackett (1981), Agresti (1984), Goodman (1984) y Freeman (1987), ha ocupado un lugar prominente en este campo. En estos libros se lleva a cabo la verificación de los modelos construidos tradicionalmente a través del estadístico X^2 de Pearson o del estadístico del logaritmo del cociente de verosimilitudes. No obstante, estos estadísticos clásicos no siempre dan los mejores resultados como lo demuestran trabajos posteriores.

Otros autores como puede verse en el libro de Read y Cressie (1988) y en las referencias que allí se citan, se han preocupado más que en construir modelos en describir y valorar estadísticos de bondad de ajuste ya existentes por un lado, y por otro en definir nuevos estadísticos que mejoren en algún sentido a los ya existentes. Ambos estudios se encuadran dentro de lo que hoy se conoce como Teoría de la

Información Estadística ya que hacen uso de medidas de divergencia para sus fines. No se debe olvidar que prácticamente todos los estadísticos construidos para abordar el problema de bondad de ajuste se basan en una medida de divergencia.

Además en el supuesto de que la hipótesis nula dependa de algún parámetro desconocido una forma de estimarlo dentro del enfoque de la Teoría de la Información Estadística sería a través de los estimadores de mínima divergencia. La consideración de estos estimadores no es nueva ya que desde un punto de vista clásico la consideración del estadístico X^2 de Pearson lleva asociado la estimación de los parámetros a través del estimador de máxima verosimilitud del modelo discretizado que no es otro que el estimador de mínima divergencia de Kullback.

Esta memoria se encuadra dentro de la línea de actuación iniciada por Read (1982) y que tiene como objetivo el de presentar procedimientos alternativos tanto en el problema de bondad de ajuste como en el de estimación previamente de parámetros en el modelo cuando sea necesario. Ambos problemas se plantean y resuelven a través de las medidas de R_ϕ -divergencia introducidas en la literatura estadística por Burbea y Rao (1982). La utilización de esta familia de divergencias en la resolución de otros problemas estadísticos puede verse en Rao (1982a,1982b), Lau (1985), Rao y Nayak (1985), Nayak (1986), etc. Un hecho importante que puso de manifiesto Read (1982) a través de la divergencia paramétrica que introdujo, divergencia que es un caso particular de la familia de ϕ -divergencias de Csiszár (1967), es que para $\lambda=2/3$ el estadístico resultante es un excelente competidor del estadístico de la X^2 . Hoy en día este es un hecho aceptado por la comunidad científica estadística como así queda recogido en los numerosos trabajos aparecidos en los últimos años en los que siempre que se aborda la posibilidad, para un problema concreto, de encontrar un test competitivo, en algún sentido, con el de la X^2 aparece de forma natural el estadístico introducido por Read para $\lambda=2/3$.

En la unificación de medidas de divergencia que se lleva a cabo en el capítulo I, como luego se verá, aparecen tres familias de divergencias: Divergencias de Csiszár, Divergencias de Burbea-Rao y Divergencias de Bregman. Read encontró, como ya se señaló anteriormente, un test de bondad de ajuste alternativo dentro de la familia de Divergencias de Csiszár. En esta memoria se obtiene un resultado importante cuando se considera en bondad de ajuste para hipótesis nula equiprobable una función $\phi = \phi_\alpha$, ya que el estadístico X^2 de Pearson coincide con el basado en R_{ϕ_2} y en muchos casos aparece como competidor de éste un estadístico alternativo desconocido basado en $R_{\phi_{13/7}}$. Encontrados estadísticos alternativos en las familias de Csiszár y Rao, obviamente, el problema que queda abierto y que debe ser objeto de estudios posteriores es el de encontrar, si esto fuera posible, un estadístico competitivo en la familia de divergencias de Bregman.

En el capítulo I se introduce, como ya se indicó anteriormente, una familia general de distancias que contiene como casos particulares a las tres familias de divergencias más importantes introducidas en la literatura estadística hasta la fecha: Divergencias de Csiszár, Divergencias de Burbea-Rao y Divergencias de Bregman. A la vez que se hace referencia a numerosos trabajos en los que se pueden encontrar diversas propiedades analíticas de las mismas, se obtienen caracterizaciones de alguna de ellas y se establecen condiciones necesarias y suficientes para validar la condición de proceso de datos: "data processing". Es decir, se ha seguido el criterio de citar trabajos donde se pueden encontrar propiedades analíticas y únicamente desarrollar aquellos resultados que son originales y se presentan en esta memoria por primera vez. Este estudio analítico de la familia general de distancias introducida se realiza para el caso discreto ya que en los capítulos siguientes se considera el modelo discretizado asociado a un espacio estadístico general para obtener estimadores puntuales y construir contrastes a partir de las divergencias de

Burbea-Rao.

En el capítulo II se propone un método de estimación basado en la R_ϕ -divergencia (divergencias de Burbea-Rao) para datos agrupados. Este método consiste en estimar el parámetro desconocido de una distribución por aquel valor que minimiza la distancia elegida entre las frecuencias relativas y las probabilidades esperadas de cada una de las clases. Se analizan las propiedades y comportamiento asintótico del estimador propuesto bajo las condiciones de regularidad de Birch (1964). En concreto, se demuestra que el estimador de mínima R_ϕ -divergencia es consistente y asintóticamente normal. Además, se prueba que el método es robusto en el sentido de que a pequeñas desviaciones del modelo le corresponden pequeñas desviaciones de la estimación del parámetro. Por último, se finaliza el capítulo realizando un estudio computacional de una importante familia de R_ϕ -divergencias. Este estudio se realiza en dos direcciones: (1) encontrar el valor del parámetro de dicha familia que nos de un estimador óptimo y (2) comparar los estimadores de mínima R_ϕ -divergencia con otros estimadores conocidos, para poblaciones Normal y Weibull.

A partir de la R_ϕ -divergencia entre la probabilidad observada y esperada de la variable aleatoria multinomial que surge al discretizar los datos de una variable continua en M clases, en el capítulo III, se propone un contraste de bondad de ajuste. La hipótesis nula a contrastar puede ser simple o compuesta. El caso de hipótesis nula simple que se estudia con detalle, debido a su importancia, es cuando se consideran clases equiprobables. Además, en este caso, la distribución del estadístico propuesto basado en la R_ϕ -divergencia es una ji-cuadrado con $M-1$ grados de libertad. En el caso de hipótesis nula compuesta, la probabilidad esperada es una función de parámetros desconocidos que se tendrá que estimar. Dependiendo del método que se utilice para estimar el vector de parámetros y de la función ϕ que se considere, la distribución asintótica del estadístico del contraste es

una combinación lineal de ji-cuadrados o una ji-cuadrado con $M-M_0-1$ grados de libertad, donde M_0 es el número de parámetros estimados. Tanto para una hipótesis como para la otra la función de potencia del contraste converge a 1 cuando $n \rightarrow \infty$ para hipótesis alternativas fijas. Todos estos resultados se han obtenido para M fijo y $n \rightarrow \infty$, sin embargo, obsérvese que parece razonable hacer $M \rightarrow \infty$ cuando $n \rightarrow \infty$ ya que es claro que se está perdiendo información sustancial al hacer $n \rightarrow \infty$ manteniendo fijo el número de clases. En esta situación se demuestra que bajo determinadas condiciones el estadístico del contraste propuesto sigue una Normal. Además, se calcula en este caso la función de potencia para una familia de alternativas y se encuentra el miembro óptimo de la familia de R_ϕ -divergencias utilizada en el estudio computacional realizado en el capítulo anterior en el sentido de máxima potencia.

El último capítulo de este trabajo está dedicado al estudio de la optimalidad para muestras pequeñas de los contrastes de bondad de ajuste propuestos en el capítulo anterior. En capítulos anteriores, se ha obtenido como aproximaciones a la distribución exacta de los estadísticos basados en la R_ϕ -divergencia una χ_{M-1}^2 cuando $n \rightarrow \infty$ y M fijo y una Normal cuando $n \rightarrow \infty$ y $M \rightarrow \infty$ bajo la hipótesis simétrica. En este capítulo, en primer lugar, se proponen dos nuevas aproximaciones cuando $n \rightarrow \infty$ y M fijo. La primera de ellas modifica el estadístico de forma que la esperanza y varianza exactas de éste sean iguales a las asintóticas (las de una ji-cuadrado) más un infinitésimo $o(n^{-1})$. La otra se obtiene mediante el desarrollo asintótico de segundo orden de la distribución de los estadísticos. En segundo lugar, se comparan las cuatro aproximaciones de la distribución exacta de los estadísticos mediante dos criterios diferentes para muestras pequeñas. Afortunadamente, se observa que la aproximación obtenida vía comparación de momentos es tan precisa como la obtenida utilizando los desarrollos Edgeworth en la mayoría de los casos considerados lo que implica un gran ahorro computacional. Finalmente, se calculan las potencias exactas basadas en regiones críticas exactas para muestras

pequeñas.

Finalmente, deseo agradecer a todas aquellas personas que con su ayuda y estímulo han colaborado de alguna manera en la elaboración de esta memoria: al profesor Dr. Igor Vajda por los comentarios e ideas que han contribuido a mejorar el capítulo I; al profesor Dr. Domingo Morales que leyó cuidadosamente el original y siempre estuvo en la mejor disposición cuando discutí con él diversos aspectos de esta memoria; a los directores de la memoria, profesores Dr. Julio Angel Pardo y Dr. Leandro Pardo, mis hermanos, por su inestimable apoyo e indicaciones prestadas sin las cuales no hubiera sido posible su elaboración.

INDICE

Pág.

CAPITULO I: Propiedades analíticas de las medidas de divergencia

1.1.- Introducción.	2
1.2.- Divergencias de Csiszár.	6
1.3.- Divergencias de Burbea-Rao.	14
1.4.- Divergencias de Bregman.	20

CAPITULO II: El estimador de mínima R_ϕ -divergencia: comportamiento asintótico.

2.1.- Introducción.	25
2.2.- El estimador de mínima R_ϕ -divergencia.	28
2.3.- Propiedades y comportamiento asintótico del estimador de mínima R_ϕ -divergencia.	33
2.4.- Robustez de la función de mínima R_ϕ -divergencia.	50
2.5.- Resultados computacionales.	53

CAPITULO III: Contrastes de bondad de ajuste basados en la R_ϕ -divergencia. Comportamiento asintótico.

3.1.- Introducción.	64
3.2.- R_ϕ -divergencia y bondad de ajuste: Hipótesis nula simple. Eficiencia Pitman.	70
3.3.- R_ϕ -divergencia y bondad de ajuste: Hipótesis nula compuesta.	87
3.4.- El problema de bondad de ajuste cuando se incrementa el número de clases.	104

CAPITULO IV: Optimalidad en los contraste de bondad de ajuste basados en la R_ϕ -divergencia para muestras pequeñas.

4.1.- Introducción.	126
4.2.- Desarrollo asintótico de segundo orden de los momentos de los estadísticos R_ϕ -divergencia.	128
4.3.- Desarrollo asintótico de segundo orden de la distribución de los estadísticos R_ϕ -divergencia.	147
4.4.- Comparación de las diferentes aproximaciones de la distribución exacta de los estadísticos R_ϕ -divergencia.	167
4.5.- Potencia exacta basada en regiones críticas exactas.	197

REFERENCIAS	204
--------------------	------------

CAPITULO I

PROPIEDADES ANALITICAS DE LAS MEDIDAS DE DIVERGENCIA

- 1.1.- Introducción.**
- 1.2.- Divergencias de Csiszár.**
- 1.3.- Divergencias de Burbea-Rao.**
- 1.4.- Divergencias de Bregman.**

1.1.- Introducción

Consideremos un espacio medible $(\mathcal{X}, \beta_{\mathcal{X}})$ y las distancias $D(P, Q)$ entre medidas de probabilidad P, Q definidas en este espacio. Estamos interesados en distancias reflexivas, es decir, las que satisfacen la condición

$$D(P, Q) \geq 0, \text{ dándose el } = \text{ si y sólo si } P=Q, \quad (1.1.1)$$

pero no necesariamente simétricas

$$D(P, Q) = D(Q, P) \quad (1.1.2)$$

ni verificando la desigualdad triangular

$$D(P, Q) \leq D(P, \tilde{P}) + D(\tilde{P}, Q). \quad (1.1.3)$$

En otras palabras, las distancias que vamos a considerar no son necesariamente métricas en el espacio \mathcal{P} de todas las distribuciones de probabilidad en $(\mathcal{X}, \beta_{\mathcal{X}})$.

Las distancias o discrepancias consideradas juegan un papel importante en teoría de la probabilidad, teoría de la información y

estadística matemática. Por ejemplo, las distancias de Prohorov, Levy, Kolmogorov, y variación total son métricas muy importantes en teoría de la probabilidad y estadística matemática (Billingsley (1968) y Huber (1981)). La divergencia de Kullback y Leibler no es métrica pero es muy importante en teoría de la información y estadística matemática (Blahut (1987), Cover y Thomas (1991), Kullback (1959), Tchentsov (1972), Amari (1990)), y también tiene aplicaciones interesantes en teoría de la probabilidad (Barron (1985)). Las extensiones de la divergencia de Kullback dadas por Rényi (1961), Csiszár (1963) y Ali y Silvey (1966) contienen otras distancias importantes en teoría de la información, estadística matemática y teoría de la probabilidad (Blahut (1987), Csiszár (1995), Cressie y Read (1988), Clarke y Barron (1990) y Liese y Vajda (1987)).

El teorema denominado "data processing" (Csiszár (1967) ó Cover y Thomas (1991)) es un resultado importante de teoría de la información y estadística matemática. En primer lugar fue enunciado por Kullback y Leibler (1951). Csiszár (1963, 1967) extendió dicho resultado a una clase más amplia de distancias. Pero, probablemente, la versión más general de este teorema aparece en p.17 en Liese y Vajda (1987). Este teorema establece que si se hace una transformación T del espacio de datos $(\mathcal{X}, \beta_{\mathcal{X}})$ en un espacio medible $(\mathcal{Y}, \mathcal{A})$ entonces la distancia $D(PT^{-1}, QT^{-1})$ entre las distribuciones PT^{-1} y QT^{-1} definidas en $(\mathcal{Y}, \mathcal{A})$ no puede superar el valor original de $D(P, Q)$, y debe mantenerse el mismo valor si T preserva la información estadística de los datos, es decir, si T es suficiente para $\{P, Q\}$.

Algunas distancias $D(P, Q)$ satisfacen este teorema y otras no. Por ejemplo, si consideramos un espacio finito \mathcal{X} de puntos $x_1 < x_2 < \dots < x_M$ y una transformación biyectiva $T: \mathcal{X} \rightarrow \mathcal{X}$ que evidentemente es suficiente para cualquier par $\{P, Q\}$, entonces la distancia de Kolmogorov

$$D_K(P, Q) = \max_{1 \leq j \leq M} \left| \sum_{i=1}^j (p_i - q_i) \right| \quad (1.1.4)$$

con $p_i = P(\{x_i\})$ y $q_i = Q(\{x_i\})$ no satisface el teorema. En efecto, es claro que $D_K((1/3, 2/3, 0), (2/3, 0, 1/3)) = 1/3$ mientras que si se considera como transformación T la permutación $(x_1, x_2, x_3) \rightarrow (x_2, x_3, x_1)$ se tiene que $D_K((2/3, 0, 1/3), (0, 1/3, 2/3)) = 2/3$.

Sin embargo, si se considera la transformación anterior en las discrepancias

$$D_a(P, Q) = \sum_{i=1}^M |p_i - q_i|^a, \quad a > 0, \quad (1.1.5)$$

éstas no varían.

Consideremos el espacio finito $\mathcal{X} = \{x_1, x_2, \dots, x_M\}$, las distribuciones $P = (p_1, p_2, \dots, p_M)$, $Q = (q_1, q_2, \dots, q_M)$, y las distancias

$$D(P, Q) = \sum_{i=1}^M \delta(p_i, q_i), \quad (1.1.6)$$

donde $\delta: [0, 1]^2 \rightarrow (-\infty, \infty]$ es finita en $(0, 1]^2$, continua en $[0, 1]^2 - \{(0, 0)\}$, y $\delta(0, 0) = 0$. De esta forma las distancias (1.1.5) están dentro de esta clase mientras que la distancia de Kolmogorov (1.1.4) no. El considerar distribuciones discretas no representará ninguna restricción a lo largo de la presente memoria ya que la teoría que se ha desarrollado en los capítulos siguientes se sustentará en el correspondiente modelo discretizado de un modelo estadístico dado.

Obsérvese que para muchas distancias consideradas en la literatura, la definición de $\delta(p_i, q_i)$ para $p_i = 0$ ó $q_i = 0$ da problemas. Los valores $\delta(p_i, q_i)$ son normalmente finitos y continuos en p_i, q_i para $(p_i, q_i) \in (0, 1]^2$, por lo que es natural extender $\delta(p, q)$ de forma continua a $[0, 1]^2$. La única excepción es el punto $(0, 0) \in [0, 1]^2$ donde se puede suponer que $\delta(0, 0) = 0$, ya que los puntos con $p_i = q_i = 0$ no deberían contribuir a la distancia entre las distribuciones de

probabilidad. De esta forma el $\{(0,0)\}$ se excluye del dominio de la extensión continua y se trata por separado. Además, debido a que los valores de $\delta(p,q)$ no necesitan estar acotados en $(0,1]^2$, la extensión continua debe ser considerada en la topología de la recta real extendida $\bar{R}=[-\infty,\infty]$, y los valores $\delta(0,q)$ y $\delta(p,0)$ del borde $\{(0,q):0<q\leq 1\} \cup \{(p,0):0<p\leq 1\}$ de $[0,1]^2-\{(0,0)\}$ pueden ser infinitos. Finalmente, para evitar en (1.1.6) expresiones del tipo $\infty-\infty$, se excluyen los valores $\delta(p,q)=-\infty$ en $[0,1]^2$. Esto explica las suposiciones hechas en (1.1.6) acerca de $\delta(p,q)$.

Muchos autores han considerado distancias entre distribuciones discretas del tipo (1.1.6). Entre ellos se encuentran Neyman (1949), Kullback y Leibler (1951), Rao (1961), Rényi (1961), Csiszár (1963), Ali y Silvey (1966), Bregman (1967), Robertson (1972), Burbea y Rao (1982), Cressie y Read (1984) y Cohen y otros (1993).

La condición "data processing" para las distancias (1.1.6) se formula de la siguiente forma:

"Consideremos una aplicación T del espacio $\mathcal{X} = \{x_1, \dots, x_M\}$ en $\mathcal{Y} = \{y_1, \dots, y_s\}$ donde $1 \leq s \leq M$ y los conjuntos $A_j = T^{-1}\{y_j\}$, $1 \leq j \leq s$, son distintos del vacío. La clase $\{A_1, \dots, A_s\}$ es una partición de \mathcal{X} y las distribuciones inducidas vienen dadas por

$$\tilde{P} = (\tilde{p}_1, \dots, \tilde{p}_s) = PT^{-1} \quad \text{y} \quad \tilde{Q} = (\tilde{q}_1, \dots, \tilde{q}_s) = QT^{-1}$$

siendo

$$\tilde{p}_j = \sum_{i \in A_j} p_i \quad \text{y} \quad \tilde{q}_j = \sum_{i \in A_j} q_i, \quad j=1, \dots, s. \quad (1.1.7)$$

Por el Teorema de Factorización, la aplicación T es suficiente para $\{P, Q\}$ si existen $h_1 \geq 0, \dots, h_M \geq 0$ tales que

$$p_i = \tilde{p}_j h_i \quad \text{y} \quad q_i = \tilde{q}_j h_i \quad \forall i \text{ tal que } x_i \in A_j \text{ y } 1 \leq j \leq s. \quad (1.1.8)$$

Entonces, la distancia (1.1.6) satisface la condición "data processing" si, para todas las distribuciones $P=(p_1, \dots, p_M)$, $Q=(q_1, \dots, q_M)$ y aplicaciones T consideradas,

$$D(\tilde{P}, \tilde{Q}) \leq D(P, Q) \quad (1.1.9)$$

y se da la igualdad cuando T es suficiente para $\{P, Q\}$."

En el apartado siguiente se encuentran las condiciones que deben verificar las funciones $\delta(u, v)$ para que las correspondientes distancias (1.1.6) verifiquen la condición "data processing" y como consecuencia se obtiene una caracterización de las divergencias de Csiszár. De acuerdo a esta caracterización, únicamente la variación total $D_1(P, Q)$ de entre las distancias consideradas en (1.1.5) satisface la condición "data processing". En los apartados 1.3 y 1.4 se introducen a partir de la expresión (1.1.6) las distancias de Burbea-Rao (Burbea y Rao (1982)) y las distancias de Bregman (Bregman (1967)) utilizadas recientemente en el contexto estadístico por Csiszár (1991, 1994) y se establecen condiciones a través del teorema de caracterización dado en 1.2 para que estas medidas de divergencia verifiquen la condición "data processing".

1.2.- Divergencias de Csiszár

En este apartado se caracteriza las divergencias de Csiszár y se demuestra una condición necesaria y suficiente para que se cumpla la condición "data processing".

Decimos que la distancia (1.1.6) es f-divergencia si existe una función convexa $f:(0, \infty) \rightarrow \mathbb{R}$ tal que

$$\delta(p,q) = qf\left[\frac{p}{q}\right] \quad \forall 0 < p, q \leq 1. \quad (1.2.1)$$

Entonces, por continuidad,

$$\delta(0,q) = q \lim_{t \downarrow 0} f(t) = q f(0) \quad \forall 0 < q \leq 1.$$

y

$$\delta(p,0) = p \lim_{t \rightarrow \infty} \frac{f(t)}{t} = p f(\infty)/\infty \quad \forall 0 < p \leq 1.$$

Esta elección de δ hace que la distancia (1.1.6) coincida con la f-divergencia de Csiszár (1963)

$$D_f(P,Q) = \sum_{i=1}^M q_i f\left[\frac{p_i}{q_i}\right] \quad (1.2.2)$$

con

$$q f\left[\frac{0}{q}\right] = q f(0) \quad \forall 0 < q \leq 1,$$

$$0 f\left[\frac{p}{0}\right] = p f(\infty)/\infty \quad \forall 0 < p \leq 1,$$

y

$$0 f\left[\frac{0}{0}\right] = 0,$$

donde $f(0)$ y $f(\infty)/\infty$ se definen como antes.

Algunas de las propiedades básicas de las f-divergencias probadas en Liese y Vajda (1987) o Vajda (1989) son:

(i) Se verifica que $f(1) \leq f(0) + f(\infty)/\infty$ donde la igualdad se da solo si f es afín en $(0, \infty)$.

(ii) Se verifica que $f(1) \leq D_f(P,Q) \leq f(0) + f(\infty)/\infty$, donde $f(1)=D_f(P,Q)$ si $P = Q$ y $D_f(P,Q) = f(0) + f(\infty)/\infty$ si P es singular con Q , $P \perp Q$.

(iii) Si f no es afín en $(0,\infty)$ entonces $f(1) = D_f(P,Q)$ solo si $P = Q$ y $D_f(P,Q) = f(0) + f(\infty)/\infty < \infty$ solo si $P \perp Q$.

(iv) Dos f -divergencias $D_{f_1}(P,Q)$ y $D_{f_2}(P,Q)$ coinciden para cualquiera de las distribuciones P y Q consideradas si y sólo si existe $c \in \mathbb{R}$ tal que $f_1(t)-f_2(t) = c(t-1)$ para todo $t \in (0,\infty)$.

(v) Toda f -divergencia satisface la condición "data processing".

Ejemplo 1.2.1

Las funciones convexas no negativas con $f_a(1)=0$,

$$f_a(t) = \frac{t^a - at + a - 1}{-a(a-1)} \quad \text{para } a \neq 0, a \neq 1,$$

definen una conocida clase de f -divergencias donde sus extensiones continuas vienen dadas por

$$f_0(t) = -\ln t + t - 1 \quad \text{y} \quad f_1(t) = t \ln t - t + 1$$

(Cressie y Read (1984); Battacharyya (1946) y Rényi (1961) consideraron funciones de estas f -divergencias para $a=1/2$ y $a>0$ respectivamente, y Kullback y Leibler (1951) consideraron el caso $a=0$ y $a=1$). Únicamente para $a=1/2$ la distancia correspondiente es una métrica, además su raíz cuadrada es la distancia de Hellinger.

Otras f -divergencias que también son métricas se pueden obtener con

$$f(t) = \frac{(t-1)^2}{t+1}$$

considerada por Vajda y Kus (1995), con

$$f(t) = |t^a - 1|^{1/a}, \quad 0 < a \leq 1$$

considerada por Matusita (1964) o con

$$f(t) = (t^a + 1)^{1/a} - 2^{(1-a)/a}(t+1), \quad a > 1$$

introducida por Österreicher (1996).

■

Teorema 1.2.1

La distancia (1.1.6) es f-divergencia si y sólo si satisface la condición "data processing". La función convexa f para la que únicamente esto es cierto viene dada por

$$f(t) = \begin{cases} \delta(t, 1) & \text{para } 0 < t \leq 1 \\ t\delta(1, 1/t) & \text{para } t > 1 \end{cases} \quad (1.2.3)$$

excepto por la equivalencia descrita en la propiedad (iv).

Demostración

Por una parte, si la distancia (1.1.6) es f-divergencia entonces se verifica la condición "data processing" por (v) y la relación (1.2.3) entre δ y f. Además, la f-divergencia no puede ser una f^* -divergencia para $f^* \neq f$ excepto en el caso descrito en (iv).

Por otra parte, supongamos que la distancia (1.1.6) satisface la condición "data processing". Sea T una aplicación de $\mathfrak{X} = \{x_1, x_2, \dots, x_M\}$ en $\mathfrak{Y} = \{y_1, y_2, \dots, y_{M-1}\}$ tal que $T(x_1) = T(x_2) = y_1$. Esta aplicación es

suficiente para las P, Q consideradas si y sólo si ó $q_1+q_2=0$ ó $q_1>0$, $q_2>0$ y

$$\frac{p_1}{q_1} = \frac{p_2}{q_2} = t \geq 0.$$

De (1.1.9) para cada $p_i, q_i > 0$ con $p_1+p_2 \leq 1$, $q_1+q_2 \leq 1$

$$\delta(p_1+p_2, q_1+q_2) \leq \delta(p_1, q_1) + \delta(p_2, q_2) \quad (1.2.4)$$

y para cada $t > 0$ y $q_1, q_2 > 0$ con $q_1+q_2 \leq \min\{1, 1/t\}$

$$\delta(t(q_1+q_2), q_1+q_2) = \delta(tq_1, q_1) + \delta(tq_2, q_2). \quad (1.2.5)$$

Demostraremos que (1.2.5) implica la existencia de una función $f: (0, \infty) \rightarrow \mathbb{R}$ satisfaciendo (1.2.1) y que (1.2.4) implica su convexidad.

Sea $t > 0$ arbitrario fijo y consideramos la función continua $\psi(q) = \delta(tq, q)$ de variable $0 < q \leq \min\{1, 1/t\}$. Si $0 < 2\varepsilon < \min\{1/2, 1/2t\}$ entonces (1.2.5) implica para todo $\varepsilon \leq q_1, q_2 \leq \min\{1/2, 1/2t\}$

$$\psi(q_1+q_2) = \psi(q_1) + \psi(q_2).$$

Por Teorema 1 en p.46 de Aczél (1966), esto implica la existencia de $f(t) \in \mathbb{R}$ tal que

$$\delta(tq, q) = f(t) q \quad (1.2.6)$$

para todo $\varepsilon \leq q \leq \min\{1/2, 1/2t\}$. De (1.2.5) tenemos para todo $0 < q < \varepsilon$

$$\delta(t(\varepsilon+q), \varepsilon+q) = \delta(t\varepsilon, \varepsilon) + \delta(tq, q)$$

y, aplicando (1.2.6),

$$f(t)(\varepsilon+q) = f(t)\varepsilon + \delta(tq, q).$$

Por tanto (1.2.6) es válido para todo $0 < q \leq \min\{1/2, 1/2t\}$. Análogamente se puede extender la validez de (1.2.6) a todo $0 < q \leq \min\{1, 1/t\}$, lo que implica (1.2.1) y la relación (1.2.3) entre δ y f .

De este modo únicamente queda probar que f es convexa. Es fácil probar que para cada $0 < \pi < 1$ y $t_1, t_2 > 0$ existe p_i, q_i pertenecientes al dominio considerado en (1.2.4) tal que

$$\frac{q_1}{q_1 + q_2} = \pi \quad \text{y} \quad \frac{p_i}{q_i} = t_i \quad \text{para } i=1,2.$$

Reescribiendo (1.2.4) por medio de (1.2.6) en la forma

$$(q_1 + q_2) f\left(\frac{p_1 + p_2}{q_1 + q_2}\right) \leq q_1 f\left(\frac{p_1}{q_1}\right) + q_2 f\left(\frac{p_2}{q_2}\right)$$

se comprueba después de algunas operaciones algebraicas que (1.2.4) implica la desigualdad de Jensen para f ,

$$f(\pi t_1 + (1-\pi)t_2) \leq \pi f(t_1) + (1-\pi)f(t_2).$$

■

El Teorema 1.2.1 presenta un método directo para verificar si una distancia (1.1.6) es f -divergencia, que consiste en comprobar si la función (1.2.3) es convexa en $(0, \infty)$. En algunos casos esta comprobación no será fácil. El Teorema 1.2.2, basado en los dos lemas siguientes, proporciona un método alternativo.

Lema 1.2.1

Una aplicación $\delta: (0,1]^2 \rightarrow \mathbb{R}$ satisface la condición de homogeneidad

$$\delta(tu, tv) = t \delta(u, v) \quad \forall t > 0 \text{ y } 0 < u, v \leq \min\{1, 1/t\} \quad (1.2.7)$$

si y sólo si existe una función $f: (0, \infty) \rightarrow \mathbb{R}$ verificando (1.2.1).

Demostración

Es inmediato que (1.2.1) implica (1.2.7) por tanto demostraremos que (1.2.7) implica (1.2.1). Para ello definimos f por (1.2.3). De (1.2.7) se sigue que si $0 < u \leq v \leq 1$ entonces para $0 < t = u/v \leq 1$

$$\delta(u,v) = \delta(tv,v) = v \delta(t,1) = v f(t) = v f(u/v),$$

y si $0 < v < u \leq 1$ entonces para $t = u/v > 1$

$$\delta(u,v) = \delta(u,u/t) = u \delta(1,1/t) = v t \delta(1,1/t) = v f(t) = v f(u/v)$$

■

Lema 1.2.2

La aplicación δ en Lema 1.2.1 es continua o convexa si y sólo si la función f considerada en éste es continua o convexa.

Demostración

La continuidad es evidente de la relación entre δ y f en la demostración del Lema 1.2.1. Por tanto probemos la convexidad.

Si $\delta(u,v)$ es convexa entonces $f_v(t) = \delta(vt,v)$ son convexas en los dominios $0 < t \leq 1/v$ para todo $0 < v \leq 1$. Por (1.2.1), $f(t)$ coincide con $f_v(t)/v$ en el dominio $(0, 1/v]$ por tanto $f(t)$ es convexa en $(0, \infty)$. Si, al contrario, $f(t)$ es convexa entonces para cada $0 < \pi < 1$ y $(u_i, v_i) \in (0, 1]^2$, $i=1,2$, una mixtura convexa adecuada con el parámetro

$$\tau = \frac{\pi v_1}{\pi v_1 + (1-\pi)v_2} \in (0, 1)$$

lleva a la desigualdad de Jensen para f ,

$$f\left[\tau\frac{u_1}{v_1} + (1-\tau)\frac{u_2}{v_2}\right] \leq \tau f\left[\frac{u_1}{v_1}\right] + (1-\tau)f\left[\frac{u_2}{v_2}\right].$$

Pero esto es equivalente a

$$(\pi v_1 + (1-\pi)v_2)f\left[\frac{\pi u_1 + (1-\pi)u_2}{\pi v_1 + (1-\pi)v_2}\right] \leq \pi v_1 f\left[\frac{u_1}{v_1}\right] + (1-\pi)v_2 f\left[\frac{u_2}{v_2}\right]$$

lo cual es por (1.2.1) la desigualdad de Jensen

$$\delta(\pi u_1 + (1-\pi)u_2, \pi v_1 + (1-\pi)v_2) \leq \pi \delta(u_1, v_1) + (1-\pi)\delta(u_2, v_2)$$

para $\delta(u, v)$.

■

Teorema 1.2.2

La distancia (1.1.6) es f-divergencia si y sólo si δ es convexa en $(0,1]^2$ y homogénea en el sentido de (1.2.7).

Demostración

Evidente por los Lemas 1.2.1 y 1.2.2.

■

Ejemplo 1.2.2

Para las distancias (1.1.5) la función $\delta_a(u, v) = |u-v|^a$ es convexa en el dominio $u, v \geq 0$ solo si $a \geq 1$. De esta forma las distancias $D_a(P, Q)$ para $0 < a < 1$ no satisfacen la condición "data processing". De las funciones $\delta_a(u, v)$, $a \geq 1$, solo aquellas con $a=1$ satisfacen la condición de homogeneidad

$$\delta_a(tu, tv) = t\delta_a(u, v) \text{ para todo } t, u, v > 0.$$

De esta forma $D_1(P,Q)$ es la única distancia de la clase (1.1.5) que satisface la condición "data processing". Además $D_1(P,Q) = \sum_{i=1}^M |p_i - q_i|$ es la métrica de la variación total, la f-divergencia para $f(t) = |t-1|$.

■

1.3.- Divergencias de Burbea-Rao

A lo largo de esta memoria consideramos una función cóncava continua $\phi:(0,\infty) \rightarrow \mathbb{R}$, donde

$$\phi(0) = \lim_{t \downarrow 0} \phi(t) \in (-\infty, \infty].$$

La función $\delta_\phi:[0,1]^2 \rightarrow (-\infty, \infty]$ definida por

$$\delta_\phi(u,v) = \begin{cases} \phi\left[\frac{u+v}{2}\right] - \frac{\phi(u)+\phi(v)}{2} & \text{si } (u,v) \neq (0,0) \\ 0 & \text{si } (u,v) = (0,0) \end{cases} \quad (1.3.1)$$

satisface todas las suposiciones hechas sobre $\delta(u,v)$ en (1.1.6). La distancia correspondiente

$$R_\phi(P,Q) = \sum_{i=1}^M \delta_\phi(p_i, q_i) \quad (1.3.2)$$

es la divergencia de Burbea y Rao (1982), denominada R_ϕ -divergencia.

Por ser ϕ continua, la condición

$$\delta_\phi(u,v) \geq 0 \quad \forall 0 \leq u, v \leq 1$$

es equivalente a la desigualdad de Jensen

$$\pi\phi(u)+(1-\pi)\phi(v) \leq \phi(\pi u+(1-\pi)v) \quad \forall 0 \leq \pi \leq 1 \text{ y } 0 \leq u, v \leq 1. \quad (1.3.3)$$

Sabemos que (Apéndice A en Liese y Vajda (1987)), (1.3.3) es equivalente a la existencia de una aplicación $\phi^*: (0,1) \rightarrow \mathbb{R}$ tal que

$$\phi(t) \leq \phi(t_0) + \phi^*(t_0)(t-t_0) \quad \forall 0 \leq t \leq 1, 0 < t_0 < 1. \quad (1.3.4)$$

La función (1.3.1) es idénticamente cero si y sólo si ϕ es afín en $(0,1)$. En efecto, ϕ es estrictamente cóncava en $t_0 \in (0,1)$ si la desigualdad en (1.3.4) es estricta para todo $t \in (t_0, 1]$ o todo $t \in [0, t_0)$. Como se prueba en el Apéndice A de Liese y Vajda (1987), esto se da si y sólo si la desigualdad en (1.3.3) es estricta para todo $0 < \pi < 1$ y todo $0 \leq u, v \leq 1$ satisfaciendo la condición

$$\pi u + (1-\pi)v = t_0.$$

De este modo la identidad $\delta_\phi(u,v) \equiv 0$ implica que ϕ no es estrictamente cóncava en $t_0 \in (0,1)$, es decir, para cada $t_0 \in (0,1)$ existen puntos $u_0 \in [0, t_0)$ y $v_0 \in (t_0, 1)$ tales que los valores $\phi(u_0)$, $\phi(t_0)$ y $\phi(v_0)$ pertenecen a la recta

$$\phi(t) = \phi(t_0) + \phi^*(t_0)(t-t_0).$$

Esto junto con (1.3.3) implica que ϕ es afín en $[u_0, v_0]$. Si se considera un conjunto adecuado de valores t_0 , esta afinidad se extiende a $(0,1)$.

Esto nos permitirá caracterizar las distancias de Burbea-Rao.

Lema 1.3.1

La distancia (1.3.2) es simétrica para las funciones ϕ consideradas en el sentido de (1.1.2), no negativa e igual cero para

$P=Q$. Si ϕ es estrictamente cóncava para cada punto del intervalo $(0,1)$ entonces esta distancia es reflexiva en el sentido de (1.1.1).

Demostración

Es evidente, por la concavidad de ϕ que la distancia (1.3.2) es no negativa. Además $\delta_\phi(u,v)$ es siempre simétrica en las variables u y v .

Por otra parte, de lo visto anteriormente, se tiene que si ϕ es estrictamente cóncava para todo $0 < t < 1$ entonces

$$\delta_\phi(u,v) > 0 \quad \forall 0 \leq u, v \leq 1, u \neq v.$$

■

El resultado siguiente es una modificación del Teorema 1 de Burbea y Rao (1982).

Teorema 1.3.1

Sea ϕ estrictamente cóncava y dos veces diferenciable en cada punto del intervalo $(0,1)$. Entonces δ_ϕ es convexa en $[0,1]^2$ (estrictamente convexa para todo $(u,v) \in (0,1)^2$ con $u \neq v$) si y sólo si $1/\phi''(t)$ es convexa (estrictamente convexa) en $(0,1)$.

Demostración

La matriz Hessiana de $\delta_\phi(u,v)$ en $(0,1)^2$ viene dada por

$$H_\phi(u,v) = \frac{1}{2} \begin{pmatrix} \frac{1}{2} \phi''\left[\frac{u+v}{2}\right] - \phi''(u) & \frac{1}{2} \phi''\left[\frac{u+v}{2}\right] \\ \frac{1}{2} \phi''\left[\frac{u+v}{2}\right] & \frac{1}{2} \phi''\left[\frac{u+v}{2}\right] - \phi''(v) \end{pmatrix},$$

donde la concavidad estricta de ϕ implica que todas las derivadas

segundas son negativas. La matriz Hessiana es semidefinida positiva si y sólo si

$$2\phi''(u) > \phi''\left(\frac{u+v}{2}\right)$$

y

$$\phi''(u)\phi''(v) \geq \frac{1}{2} \phi''\left(\frac{u+v}{2}\right)(\phi''(u)+\phi''(v)).$$

La segunda condición es equivalente a

$$\frac{1}{\phi''\left(\frac{u+v}{2}\right)} \leq \frac{1}{2} \left[\frac{1}{\phi''(u)} + \frac{1}{\phi''(v)} \right].$$

La matriz Hessiana es semidefinida positiva si y sólo si se tiene la última desigualdad y es definida positiva si y sólo si esta desigualdad es estricta.

La demostración del teorema se completa teniendo en cuenta que (B.3.d en Chap. 16 de Marshal y Olkin (1979)) una función $\delta(u,v)$ es convexa en el dominio $0 < u, v < 1$ si y sólo si su matriz Hessiana es semidefinida positiva en este dominio, y estrictamente convexa en el dominio $0 < u < v$ y $0 < v < u$ si y sólo si la matriz Hessiana es definida positiva en este recinto. La convexidad en todo $[0,1]^2$ se sigue de la continuidad de $\delta_\phi(u,v)$ en $[0,1]^2 - \{(0,0)\}$ y del hecho de que $\delta_\phi(0,0)=0$.

■

Obsérvese que desde un punto de vista intuitivo es interesante que $\delta_\phi(u,v)$ sea convexa en (u,v) ya que eso significa que la divergencia media entre (u,v) y (z,w) no es menor que entre su combinación lineal $\lambda(u,v) + \mu(z,w)$, donde $\lambda, \mu \geq 0$ y $\lambda + \mu = 1$.

El siguiente resultado establece condiciones necesarias y

suficientes para que las divergencias de Burbea-Rao verifiquen la condición "data processing".

Teorema 1.3.2

Supongamos que ϕ satisface las condiciones del Teorema 1.3.1. Entonces la distancia de Burbea-Rao (1.3.2) satisface la condición "data processing" si y sólo si $1/\phi''(t)$ es convexa en $(0,1)$ y se verifica que

$$\frac{1}{t}\phi\left[t\frac{u+v}{2}\right] - \frac{\phi(tu)+\phi(tv)}{2t} = \phi\left[\frac{u+v}{2}\right] - \frac{\phi(u)+\phi(v)}{2} \quad (1.3.5)$$

para todo $t > 0$ y $0 < u, v \leq \min\{1, 1/t\}$. Si estas condiciones se mantienen $R_{\phi}(P, Q)$ coincide con la f -divergencia de P y Q dada por (1.2.2) para la función convexa

$$f(t) = \phi\left[\frac{t+1}{2}\right] - \frac{\phi(t)+\phi(1)}{2}.$$

Demostación

Evidente por el Teorema 1.2.2 y el Teorema 1.3.1. ■

Ejemplo 1.3.1

Las funciones $\phi_1(t) = -t \ln t$ y $\phi_0(t) = \ln t$ cumplen las condiciones del Teorema 1.3.1 con $\phi_1''(t) = -t^{-1}$ y $\phi_0''(t) = -t^{-2}$. Al ser $1/\phi_1''(t)$ convexa y $1/\phi_0''(t)$ cóncava, por el Teorema 1.3.2,

$$R_{\phi_0}(P, Q) = \sum_{i=1}^M \ln \frac{p_i + q_i}{2(p_i q_i)^{1/2}}$$

no satisface la condición "data processing". Para $\phi_1(t)$ se verifica

que

$$\frac{u+v}{2} \ln\left(t \frac{u+v}{2}\right) - \frac{u \ln(tu) + v \ln(tv)}{2} = \frac{u+v}{2} \ln\left(\frac{u+v}{2}\right) - \frac{u \ln u + v \ln v}{2}$$

por lo que se tiene (1.3.5). Por tanto, por el Teorema 1.3.2,

$$R_{\phi_1}(P, Q) = \frac{1}{2} \sum_{i=1}^M \left[p_i \ln \frac{2p_i}{p_i + q_i} + q_i \ln \frac{2q_i}{p_i + q_i} \right]$$

satisface la condición "data processing" y coincide con la f-divergencia de P y Q para $f(t) = 1/2 \left[t \ln t - (t+1) \ln[(t+1)/2] \right]$ que viene dada por

$$D_f(P, Q) = \frac{1}{2} \sum_{i=1}^M \left[p_i \ln \frac{p_i}{q_i} + (p_i + q_i) \ln \frac{2q_i}{p_i + q_i} \right].$$

La igualdad

$$R_{\phi_1}(P, Q) = \frac{1}{2} \sum_{i=1}^M \left[p_i \ln \frac{p_i}{q_i} + (p_i + q_i) \ln \frac{2q_i}{p_i + q_i} \right]$$

se puede obtener también directamente, sin utilizar el Teorema 1.3.2.

La primera componente de la expresión anterior es la divergencia de Kullback

$$I(P, Q) = \sum_{i=1}^M p_i \ln \frac{p_i}{q_i}$$

pero no está claro a primera vista si la segunda componente

$$D(P, Q) = \sum_{i=1}^M (p_i + q_i) \ln \frac{2q_i}{p_i + q_i}$$

es también f-divergencia. Observemos que esto se puede verificar utilizando el Teorema 1.2.2. De hecho, la función

$$\delta(u,v) = (u+v)\ln\left[\frac{2v}{u+v}\right]$$

es convexa en el dominio $0 < u, v \leq 1$ y homogénea en el sentido de (1.2.7). Por tanto la respuesta es afirmativa y, por (1.2.3), la función convexa correspondiente es $f(t) = (t+1)\ln(2/(t+1))$.

■

1.4.- Divergencias de Bregman

A lo largo de este apartado se considera una función convexa diferenciable $\phi:(0,\infty)\rightarrow\mathbb{R}$, con

$$\phi(0) = \lim_{t \downarrow 0} \phi(t) \in (-\infty, \infty].$$

La función $\delta_\phi:[0,1]^2\rightarrow(-\infty,\infty]$ definida por la extensión continua de

$$\delta_\phi(u,v) = \phi(u) - \phi(v) - \phi'(v)(u-v) \quad \forall 0 < u, v \leq 1 \quad (1.4.1)$$

en $[0,1]^2 - \{(0,0)\}$ y por $\delta_\phi(0,0) = 0$, satisface las condiciones de (1.1.6). La expresión

$$B_\phi(P,Q) = \sum_{i=1}^M \delta_\phi(p_i, q_i) \quad (1.4.2)$$

define la divergencia de Bregman (Bregman (1967) y Csiszár (1991, 1994)). Esta distancia puede no ser simétrica.

Lema 1.4.1

La distancia (1.4.2) es no negativa e igual a cero para $P=Q$. Si ϕ es estrictamente convexa para todo punto del intervalo $(0,1)$ entonces esta distancia es reflexiva en el sentido de (1.1.1).

Demostración

Es evidente, por la convexidad de ϕ que la distancia (1.4.2) es no negativa, es decir, toma valores en $[0,\infty]$.

Si ϕ es estrictamente convexa para $0 < t < 1$ entonces

$$\delta_{\phi}(u,v) > 0 \quad \forall 0 < u, v \leq 1, u \neq v.$$

■

Teorema 1.4.1

Sea ϕ estrictamente convexa y tres veces diferenciable en todo punto del intervalo $(0,1)$. Entonces δ_{ϕ} es convexa en $[0,1]^2$ (estrictamente convexa en $(u,v) \in (0,1)^2$ con $u \neq v$) si y sólo si $1/\phi''(t)$ es cóncava (estrictamente cóncava) en $(0,1)$.

Demostración

La matriz Hessiana de $\delta_{\phi}(u,v)$ viene dada por

$$H_{\phi}(u,v) = \begin{bmatrix} \psi(u) & -\psi(v) \\ -\psi(v) & \psi(v) + \psi'(v)(v-u) \end{bmatrix},$$

donde ψ representa a ϕ'' . Es semidefinida positiva si y sólo si

$$\psi(u) \left[\psi(v) + \psi'(v)(v-u) \right] \geq \psi^2(v)$$

o, equivalentemente,

$$\frac{1}{\psi(\bar{v})} + \frac{\psi'(v)}{\psi^2(v)}(v-u) \geq \frac{1}{\psi(\bar{u})}.$$

es decir

$$\frac{\frac{1}{\psi(\bar{v})} - \frac{1}{\psi(\bar{u})}}{v-u} \geq \left[\frac{1}{\psi(v)} \right]'$$

que es equivalente a la concavidad de $1/\psi(t)$ en $(0,1)$. Análogamente si la matriz Hessiana es definida positiva se tiene que $1/\psi(t)$ es estrictamente cóncava en $(0,1)$. ■

Teorema 1.4.2

Supongamos que ϕ satisface las condiciones del Teorema 1.4.1. Entonces la divergencia de Bregman (1.4.2) satisface la condición "data processing" si y sólo si $1/\phi''(t)$ es cóncava en $(0,1)$ y

$$\frac{\phi(tu) - \phi(tv)}{t} - \phi'(tv)(u-v) = \phi(u) - \phi(v) - \phi'(v)(u-v) \quad (1.4.4)$$

para todo $t > 0$ y $0 < u, v \leq \min\{1, 1/t\}$. Si se verifican estas condiciones entonces $B_\phi(P, Q)$ coincide con la f -divergencia de P y Q dada por (1.2.2) para la función convexa

$$f(t) = \phi(t) - \phi'(1)(t-1) - \phi(1).$$

Demostración

Evidente por el Teorema 1.2.2 y el Teorema 1.4.1. ■

Ejemplo 1.4.1

Las funciones $\phi_0(t) = -\ln t$ y $\phi_1(t) = t \ln t$ satisfacen las condiciones del Teorema 1.4.1. Observemos que

$$1/\phi_0''(t) = t^2 \quad \text{y} \quad 1/\phi_1''(t) = t$$

son convexas. El lado izquierdo de (1.4.4) es para $\phi_0(t)$ igual a

$$\frac{1}{t} \left[\ln \frac{v}{u} + \frac{u}{v} - 1 \right] = \frac{1}{t} \left[\phi_0(u) - \phi_0(v) - \phi_0'(v)(u-v) \right]$$

por tanto (1.4.4) no se verifica. De esta forma por el Teorema 1.4.2 se tiene que que la divergencia de Bregman $B_{\phi_0}(P,Q)$ no satisface la condición "data processing". Para $\phi_1(t)$ el lado izquierdo de (1.4.4) es igual a

$$u \ln \frac{u}{v} - (u-v) = \phi_1(u) - \phi_1(v) - \phi_1'(v)(u-v)$$

por tanto se tiene (1.4.4). De este modo el Teorema 1.4.2 implica que la correspondiente divergencia de Bregman

$$B_{\phi_1}(P,Q) = \sum_{i=1}^M p_i \ln \frac{p_i}{q_i}$$

satisface la condición "data processing". Este resultado fue obtenido por Kullback y Leibler (1951). ■

CAPITULO II

EL ESTIMADOR DE MINIMA R_{ϕ} -DIVERGENCIA: COMPORTAMIENTO ASINTOTICO

- 2.1.- Introducción.**
- 2.2.- El estimador de mínima R_{ϕ} -divergencia.**
- 2.3.- Propiedades y comportamiento asintótico del
estimador de mínima R_{ϕ} -divergencia.**
- 2.4.- Robustez de la función de mínima
 R_{ϕ} -divergencia.**
- 2.5.- Resultados computacionales.**

2.1.- Introducción

En este capítulo se considera una amplia clase de estimadores que se pueden usar cuando los datos son discretos, bien porque la distribución subyacente lo sea bien porque sea continua pero las observaciones se clasifiquen en grupos. Esta clasificación se puede llevar a cabo por razones experimentales o porque el problema de estimación que se desea resolver con los datos no agrupados tiene características no deseables.

Algunos ejemplos sencillos y otros no tan sencillos en los que falla el conocido método de máxima verosimilitud son expuestos por Le Cam (1990). Por ejemplo, supongamos que Y_1, Y_2, \dots, Y_n son variables aleatorias independientes y distribuidas como una mixtura de dos poblaciones normales con función de densidad

$$f_{\theta}(y) = w \frac{1}{(2\pi)^{1/2} \sigma_1} \exp\left[-\frac{1}{2} \left(\frac{y - \mu_1}{\sigma_1}\right)^2\right] + (1-w) \frac{1}{(2\pi)^{1/2} \sigma_2} \exp\left[-\frac{1}{2} \left(\frac{y - \mu_2}{\sigma_2}\right)^2\right]$$

donde $\theta = (\mu_1, \mu_2, \sigma_1, \sigma_2, w)$, $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$ y $w \in (0, 1)$.

La función de verosimilitud para estimar los cinco parámetros de esta distribución viene dada por

$$L(\theta; y_1, \dots, y_n) = \prod_{j=1}^n f_{\theta}(y_j).$$

Si hacemos $\mu_1 = y_i$ para algún i ($i=1, \dots, n$), entonces

$$f_{\theta}(y_i) > w(\sqrt{2\pi} \sigma_1)^{-1}$$

y

$$\frac{f_{\theta}(y_j)}{(1-w)(\sqrt{2\pi} \sigma_2)^{-1} \exp\left[-\frac{1}{2}\left(\frac{y_j - y_i}{\sigma_2}\right)^2\right]} > \quad \text{para } j \neq i.$$

De esta forma

$$L(\theta; y_1, \dots, y_n) > (2\pi)^{-n/2} w(1-w)^{n-1} \sigma_1^{-1} \sigma_2^{-(n-1)} \exp\left[-\frac{1}{2} \sum_{j \neq i} \left(\frac{y_j - y_i}{\sigma_2}\right)^2\right]$$

donde eligiendo σ_1 suficientemente pequeño, podemos hacer L tan grande como queramos. Por tanto no existen valores w , σ_1 , σ_2 , μ_1 y μ_2 que maximizen L . Es decir, no siempre existe el estimador de máxima verosimilitud basado en los datos no agrupados.

El primero en dar solución a este problema fue Pearson (1894) mediante el método de los momentos. No obstante, a pesar de ser muchos los fenómenos aleatorios que siguen esta distribución pasó mucho tiempo hasta que Hassenblad (1966) reabrió el tema. Desde entonces son muchos los autores que han abordado este problema, Cohen (1967) desarrolla un procedimiento iterativo que reduce el esfuerzo computacional requerido para resolver la ecuación de grado nueve que propuso Pearson. Day (1969) y Behboodian (1970) obtienen mediante métodos iterativos los máximos locales de la función de verosimilitud, ya que como se ha visto, es no acotada. Posteriormente, Fryer y Robertson (1972) compararon las estimaciones de los momentos y los de máxima verosimilitud y mínima χ^2 para datos agrupados de los

parámetros de varias mezclas de normales. Estos autores concluyen que las estimaciones para datos agrupados son más precisas que las de los momentos para la mayoría de las distribuciones consideradas. En los últimos años Woodward y otros (1984) y Woodward y otros (1995) han realizado interesantes comparaciones entre el estimador de máxima verosimilitud y los estimadores de mínima distancia basados en la distancia de Cramér-von Mises y en la de Hellinger, respectivamente.

En este capítulo se presenta un procedimiento que entre otros resuelve el problema planteado a través del modelo discretizado. Para discretizar el modelo se divide el recorrido de la variable aleatoria en estudio Y con valores en el espacio estadístico $(X, \beta_X, P_\theta)_{\theta \in \Theta \subseteq \mathbb{R}^{M_0}}$ en M subconjuntos disjuntos A_1, \dots, A_M . Entonces, si se extrae una muestra aleatoria simple de Y de tamaño n se tiene una variable aleatoria (X_1, \dots, X_M) multinomial, donde X_i representa el número de veces que $Y \in A_i$, $i=1, \dots, M$, con parámetros $(n; q_1(\theta), \dots, q_M(\theta))$ y siendo $q_i(\theta) = P_\theta(A_i)$, $i=1, \dots, M$, la probabilidad de la clase A_i . Evidentemente, el problema de estimar θ , por máxima verosimilitud una vez agrupados los datos consiste en maximizar para (x_1, \dots, x_M) fijo

$$P_\theta(X_1=x_1, \dots, X_M=x_M) = \frac{n!}{x_1! \dots x_M!} q_1(\theta)^{x_1} \dots q_M(\theta)^{x_M}$$

o equivalentemente

$$\ln P_\theta(X_1=x_1, \dots, X_M=x_M) = -n D^{\text{KULLBACK}}(\hat{P}, Q(\theta)) + \text{cte}$$

siendo $\hat{P}=(\hat{p}_1, \dots, \hat{p}_M)^t$ con $\hat{p}_i=x_i/n$, $i=1, \dots, M$, $Q(\theta)=(q_1(\theta), \dots, q_M(\theta))^t$ y D^{KULLBACK} la divergencia de Kullback, Kullback (1959). Por tanto estimar θ a través del modelo discretizado mediante máxima verosimilitud es equivalente a minimizar en $\theta \in \Theta \subseteq \mathbb{R}^{M_0}$ la divergencia de Kullback.

Ahora bien la divergencia de Kullback, según se vio en el

capítulo I, no es la única medida de divergencia. De esta forma surge el método de estimación basado en la mínima distancia, que consiste en elegir como estimador de θ el valor $\tilde{\theta}$ tal que

$$D(\hat{P}, Q(\tilde{\theta})) = \inf_{\theta \in \Theta \subseteq \mathbb{R}^{M_0}} D(\hat{P}, Q(\theta))$$

siendo D cualquier medida de divergencia.

A lo largo de esta memoria se abordarán los problemas de estimación y contraste basados en las medidas de divergencia de Burbea-Rao; las R_ϕ -divergencias. En este capítulo se estudia el problema de estimación puntual.

2.2.- El estimador de mínima R_ϕ -divergencia.

Sea Y_1, \dots, Y_n una muestra aleatoria simple procedente de una población dependiente de un parámetro desconocido $\theta \in \Theta \subseteq \mathbb{R}^{M_0}$ y supongamos que existe la función $Q(\theta) = (q_1(\theta), \dots, q_M(\theta))^t$ que aplica cada valor de $\theta = (\theta_1, \dots, \theta_{M_0})^t$ en $\Delta_M = \left\{ P = (p_1, \dots, p_M)^t / \sum_{i=1}^M p_i = 1, p_i \geq 0, i=1, \dots, M \right\}$. Cuando θ varia sobre Θ , $Q(\theta)$ varia sobre un subconjunto T de Δ_M . Si el modelo elegido es correcto existirá un valor $\theta^0 \in \Theta$ de tal forma que $Q(\theta^0) = \pi$ donde π es el verdadero valor de la probabilidad de la multinomial, es decir, $\pi \in T$. En caso de que el modelo no sea correcto, en general $\pi \notin T$, es decir, no existe el valor θ^0 en Θ tal que $\pi = Q(\theta^0)$.

Definición 2.2.1

Dada una m.a.s. de tamaño n procedente de una población con espacio estadístico $(\mathcal{X}, \beta_{\mathcal{X}}, P_\theta)_{\theta \in \Theta \subseteq \mathbb{R}^{M_0}}$, el estimador de mínima R_ϕ -divergencia de θ es cualquier $\hat{\theta}_\phi \in \bar{\Theta}$ verificando

$$R_{\phi}(\hat{P}, Q(\hat{\theta}_{\phi})) = \inf_{\theta \in \Theta} R_{\phi}(\hat{P}, Q(\theta)).$$

En lo sucesivo el estimador de mínima R_{ϕ} -divergencia se expresa mediante $\hat{\theta}_{\phi} = \arg \inf_{\theta \in \Theta} R_{\phi}(\hat{P}, Q(\theta))$. Este método elige el punto de T que está más próximo al valor \hat{P} en el sentido de la distancia elegida.

En el Ejemplo 2.2.1 se utiliza este estimador para dar un pronóstico acerca del parámetro de una población que se supone de Poisson.

Ejemplo 2.2.1

Sea Y_1, \dots, Y_n una muestra aleatoria simple de una población que se sospecha es de Poisson de parámetro θ desconocido. A la vista de las observaciones se consideran 3 clases: la clase del 0, la clase del 1 y la clase de los que son mayores o iguales que 2. La variable aleatoria (X_1, X_2, X_3) sigue una distribución trinomial de parámetros $(n; q_1(\theta), q_2(\theta), q_3(\theta))$, donde

$$q_1(\theta) = P_{\theta}(X=0) = e^{-\theta},$$

$$q_2(\theta) = P_{\theta}(X=1) = \theta e^{-\theta}$$

y

$$q_3(\theta) = P_{\theta}(X \geq 2) = 1 - (1+\theta)e^{-\theta}.$$

Si se considera la R -divergencia ($\phi(x) = -x \ln x$), encontrar $\hat{\theta}_1$ basado en la R -divergencia se reduce a encontrar el valor de θ que minimiza la función

$$R(\hat{P}, Q(\theta)) = \frac{\hat{p}_1 \ln \hat{p}_1 + e^{-\theta} \ln e^{-\theta}}{2} + \frac{\hat{p}_2 \ln \hat{p}_2 + \theta e^{-\theta} \ln \theta e^{-\theta}}{2} + \\ + \frac{\hat{p}_3 \ln \hat{p}_3 + (1 - (1+\theta)e^{-\theta}) \ln(1 - (1+\theta)e^{-\theta})}{2} -$$

$$- \left[\frac{\hat{p}_1 + e^{-\theta}}{2} \ln \frac{\hat{p}_1 + e^{-\theta}}{2} + \frac{\hat{p}_2 + \theta e^{-\theta}}{2} \ln \frac{\hat{p}_2 + \theta e^{-\theta}}{2} + \frac{\hat{p}_3 + (1-(1+\theta)e^{-\theta})}{2} \ln \frac{\hat{p}_3 + (1-(1+\theta)e^{-\theta})}{2} \right].$$

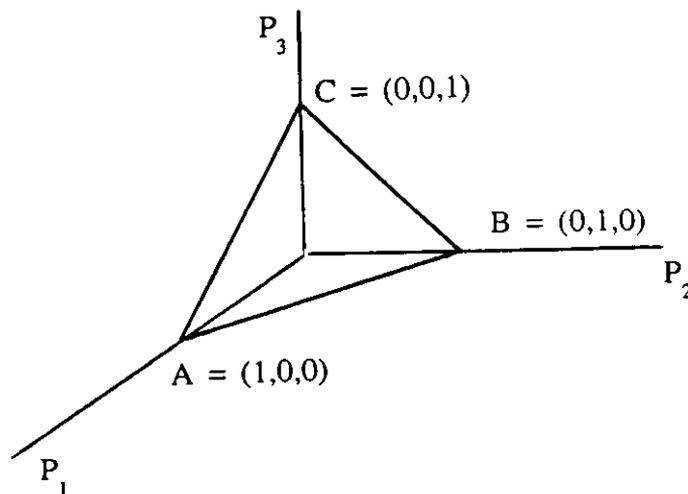
Si suponemos que $\hat{P} = (0.2, 0.3, 0.5)^t$, se obtiene $\hat{\theta}_1 = 1.661$ y por tanto,

$$q_1(\hat{\theta}_1) = 0.19, q_2(\hat{\theta}_1) = 0.31, q_3(\hat{\theta}_1) = 0.5$$

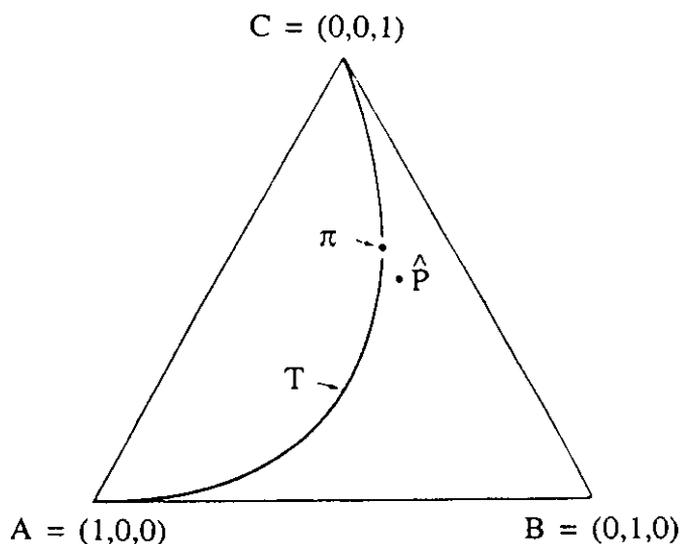
y

$$R(\hat{P}, Q(\hat{\theta}_1)) = 0.1734.$$

Es interesante hacer un análisis geométrico de este ejemplo. Inicialmente Δ_3 es la cara ABC del triedro



que representaremos en el plano mediante el triángulo



Al variar θ en $\mathbb{R}^+ = [0, \infty)$, $Q(\theta) = (e^{-\theta}, \theta e^{-\theta}, 1 - (1+\theta)e^{-\theta})^t$ es una curva sobre Δ_3 que se representa mediante el subconjunto T . Por otro lado el vector de frecuencias observadas, \hat{P} , se podrá representar en Δ_3 mediante un punto. Pues bien, si el modelo que se ha elegido para justificar los datos (en nuestro caso el modelo de Poisson) es correcto, existirá un valor "verdadero" del parámetro, θ^0 , tal que $\pi = Q(\theta^0)$, donde π es el verdadero valor de P , es decir $\pi \in T$.

■

Como se ha puesto de manifiesto en el apartado anterior el problema de estimar los parámetros de una mixtura de normales ha preocupado a muchos autores a lo largo de la historia. En el Ejemplo 2.2.2 no solo se estiman estos parámetros utilizando el estimador de mínima R_ϕ -divergencia para diferentes mixturas sino que se calcula el error cometido por dichas estimaciones.

Ejemplo 2.2.2

En la Tabla 2.2.1 aparecen las estimaciones de mínima

R_ϕ -divergencia con $\phi(x) = x - x^2$ para diversas mixturas de normales estudiadas por Woodward y otros (1984), así como los errores cuadráticos medios cometidos. Las estimaciones se han obtenido generando 500 muestras de tamaño $n=100$ de las mixturas consideradas que se han clasificado en 6 intervalos equiprobables. El procedimiento seguido para el cálculo de los valores que aparecen en la tabla es igual al explicado en el último apartado de este capítulo para poblaciones Normales y Weibull.

Tabla 2.2.1

μ_1 $\hat{\mu}_1$	σ_1 $\hat{\sigma}_1$	μ_2 $\hat{\mu}_2$	σ_2 $\hat{\sigma}_2$	w \hat{w}	ecm
0 .095116	1 .922335	2.32 2.474715	1 .901301	.25 .319910	.107400
0 .177123	1 1.084772	3.6 3.689710	1 .938717	.25 .292838	.086623
0 -.071198	1 .916150	2.56 2.616902	1 .906696	.5 .491556	.072056
0 -.042858	1 .955667	3.76 3.791897	1 .934395	.5 .491940	.059310
0 .136445	$\sqrt{2}$ 1.123112	2.32 2.831053	$\sqrt{2}$ 1.080655	.25 .408027	.271677
0 .137165	$\sqrt{2}$ 1.358219	3.6 3.761637	$\sqrt{2}$ 1.293059	.25 .300400	.178416
0 -.220047	$\sqrt{2}$ 1.168330	2.56 2.769676	$\sqrt{2}$ 1.147163	.5 .495224	.199996
0 -.089573	$\sqrt{2}$ 1.312635	3.76 3.835456	$\sqrt{2}$ 1.289461	.5 .492481	.143369

Evidentemente, el algoritmo utilizado para minimizar la R_ϕ -divergencia debe partir de unos estimadores iniciales. Los utilizados en este ejemplo son los propuestos por los autores citados anteriormente. Es decir, se toman como valores posibles de w únicamente 0.1, 0.2, ..., 0.9. Para cada uno de estos valores se divide la muestra en dos submuestras, Y_1, \dots, Y_{n_1} e Y_{n_1+1}, \dots, Y_n , donde Y_i es el estadístico de orden y y n_1 se obtiene redondeando al entero más próximo nw . Entonces, \hat{w} es aquel valor para el cual $w(1-w)(m_1 - m_2)^2$ es máximo, $\hat{\mu}_1 = m_1$, $\hat{\mu}_2 = m_2$, $\hat{\sigma}_1^2 = ((m_1 - r_1^{(.25)}) / .6745)^2$ y $\hat{\sigma}_2^2 = ((r_2^{(.75)} - m_2) / .6745)^2$, donde m_j es la mediana muestral de la subpoblación j -ésima y $r_j^{(q)}$ es el cuantil q -ésimo de la subpoblación j -ésima.

■

2.3- Propiedades y comportamiento asintótico del estimador de mínima R_ϕ -divergencia.

A lo largo de este apartado se supondrá que el modelo es correcto, $\pi = Q(\theta^0)$, y que $M_0 < M-1$. Además se admitirán las siguientes condiciones de regularidad dadas por Birch (1964):

- 1- El punto θ^0 es un punto interior de Θ .
- 2- $\pi_i = q_i(\theta^0) > 0$ para $i=1, \dots, M$ y por tanto $\pi = (\pi_1, \dots, \pi_M)^t$ es un punto interior de T .
- 3- La aplicación $Q: \Theta \longrightarrow \Delta_M$ es totalmente diferenciable en θ^0 . Por tanto existen las derivadas parciales de q_i con respecto a cada θ_j en θ^0 y $q_i(\theta)$ se puede expresar de la forma:

$$q_i(\theta) = q_i(\theta^0) + \sum_{j=1}^M (\theta_j - \theta_j^0) \frac{\partial q_i(\theta^0)}{\partial \theta_j} + o(\|\theta - \theta^0\|)$$

cuando $\theta \longrightarrow \theta^0$.

4- La matriz

$$J(\theta^0) = \left[\frac{\partial Q(\theta)}{\partial \theta} \right]_{\theta=\theta^0} = \left[\frac{\partial q_i(\theta^0)}{\partial \theta_j} \right]_{\substack{i=1,\dots,M \\ j=1,\dots,M_0}}$$

tiene rango M_0 .

5- La aplicación inversa $Q^{-1}: T \longrightarrow \Theta$ es continua en $Q(\theta^0) = \pi$.

6- La aplicación $Q: \Theta \longrightarrow \Delta_M$ es continua $\forall \theta \in \Theta$.

En los siguientes teoremas se estudiarán las propiedades asintóticas del estimador de mínima R_ϕ -divergencia. Antes se establecerá la siguiente definición:

Definición 2.3.1

Se dice que un estimador, \hat{S} , de $Q(\theta^0) = (q_1(\theta^0), \dots, q_M(\theta^0))'$ es c_n -consistente si verifica

$$c_n \|\hat{S} - Q(\theta^0)\| \leq O_p(1).$$

Dada la sucesión de variables aleatorias $\{Y_n\}_{n \in \mathbb{N}}$ la relación

$$Y_n \leq O_p(1)$$

significa que

$$\lim_{c \rightarrow \infty} \liminf_{n \rightarrow \infty} P(|Y_n| < c) = 1,$$

es decir, o bien $Y_n = O_p(1)$ (la sucesión $\{Y_n\}_{n \in \mathbb{N}}$ está acotada en probabilidad) o bien $Y_n = o_p(1)$ (la sucesión $\{Y_n\}_{n \in \mathbb{N}}$ converge en probabilidad a cero).

Si $c_n \uparrow \infty$, la condición de c_n -consistencia de un estimador es más fuerte que la consistencia. Además si un estimador \hat{S}_1 es c_n^1 -consistente y un estimador \hat{S}_2 es c_n^2 -consistente entonces tanto \hat{S}_1 como \hat{S}_2 son c_n -consistentes para $c_n = \min \{c_n^1, c_n^2\}$.

Dado el vector $P \in \Delta_M$, a la aplicación definida de Δ_M en Θ mediante

$$\theta_\phi(P) = \arg \inf_{\theta \in \Theta} R_\phi(P, \theta)$$

se le denominará función de mínima R_ϕ -divergencia. Obsérvese que dada una m.a.s. y la distribución de frecuencias, \hat{P} , asociada a ella el estimador de θ de mínima R_ϕ -divergencia viene dado por el valor de la función de mínima R_ϕ -divergencia en \hat{P} , es decir, $\hat{\theta}_\phi = \theta_\phi(\hat{P})$.

Teorema 2.3.1

Sea $\phi: (0, \infty) \rightarrow \mathbb{R}$ una función cóncava continua. Supóngase que se verifican las condiciones de regularidad 1-6 de Birch y además que:

i) Θ es un conjunto compacto.

ii) $\arg \inf_{\theta \in \Theta} R_\phi(P, \theta)$ es único en un entorno cerrado de P .

Entonces, se tiene que

$$\hat{\theta}_\phi \xrightarrow{\text{c.s.}} \theta^0.$$

Demostración

En primer lugar probaremos que θ_ϕ es una función continua,

$$\begin{array}{ccc} \theta_\phi : \Delta_M & \longrightarrow & \Theta \\ P & \longrightarrow & \theta_\phi(P) = \arg \inf_{\theta \in \Theta} R_\phi(P, \theta). \end{array}$$

Sea $\{P_n\}$ una sucesión tal que $P_n \xrightarrow{n \rightarrow \infty} P$ y supóngase que

$$\lim_{n \rightarrow \infty} \theta_\phi(P_n) = \lim_{n \rightarrow \infty} \arg \inf_{\theta \in \Theta} R_\phi(P_n, \theta) \neq \theta_\phi(P).$$

Como Θ es un conjunto compacto existe una subsucesión

$$\{\theta_\phi(P_m)\}_m \subset \{\theta_\phi(P_n)\}_n$$

verificando

$$\lim_{m \rightarrow \infty} \theta_\phi(P_m) = \theta_\star^\circ \neq \theta_\phi(P).$$

Sabemos que la función R_ϕ es una función continua, entonces

$$R_\phi(P, \theta_\star^\circ) = \lim_{m \rightarrow \infty} R_\phi(P, \theta_\phi(P_m)).$$

Por otra parte, como

$$\lim_{n \rightarrow \infty} R_\phi(P_n, \theta) = R_\phi(P, \theta) \quad \forall \theta \in \Theta,$$

y Θ es compacto, la convergencia puntual implica la convergencia uniforme, con lo cual

$$\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta} |R_\phi(P_n, \theta) - R_\phi(P, \theta)| = 0, \quad (2.3.1)$$

y por tanto

$$\lim_{n \rightarrow \infty} \left| \inf_{\theta \in \Theta} R_\phi(P_n, \theta) - \inf_{\theta \in \Theta} R_\phi(P, \theta) \right| = 0.$$

Entonces, tenemos que

$$\lim_{n \rightarrow \infty} |R_{\phi}(P_n, \theta_{\phi}(P_n)) - R_{\phi}(P, \theta_{\phi}(P))| = 0,$$

que es lo mismo que

$$\lim_{n \rightarrow \infty} R_{\phi}(P_n, \theta_{\phi}(P_n)) = R_{\phi}(P, \theta_{\phi}(P)).$$

Por otro lado, de (2.3.1) también se sigue que

$$\lim_{n \rightarrow \infty} |R_{\phi}(P_n, \theta_{\phi}(P_n)) - R_{\phi}(P, \theta_{\phi}(P_n))| = 0$$

por lo que se puede concluir que

$$\lim_{n \rightarrow \infty} R_{\phi}(P, \theta_{\phi}(P_n)) = R_{\phi}(P, \theta_{\phi}(P)).$$

Entonces, $R_{\phi}(P, \theta_{\phi}^0) = R_{\phi}(P, \theta_{\phi}(P))$ lo cual está en contradicción con la hipótesis ii) y por tanto θ_{ϕ} es una función continua.

Sabemos que $\hat{P} \xrightarrow[n \rightarrow \infty]{c.s.} \pi = Q(\theta^0)$ siendo \hat{P} el estimador no paramétrico. Entonces por continuidad de θ_{ϕ} se tiene que

$$\hat{\theta}_{\phi} = \theta_{\phi}(\hat{P}) \xrightarrow{c.s.} \theta_{\phi}(\pi) = \theta^0.$$

■

Antes de establecer el siguiente teorema se introducirá notación adicional. Considérese el operador lineal diferencial

$$\frac{d}{d\theta} = \left[\frac{\partial}{\partial \theta_1}, \dots, \frac{\partial}{\partial \theta_{M_0}} \right],$$

y la matriz Jacobiana de dimensión $M \times M_0$, $J(\theta) = (J_{jr}(\theta))$ donde

$$J_{jr}(\theta) = \frac{\partial q_j(\theta)}{\partial \theta_r}$$

Se define

$$q'_j(\theta) = \frac{\partial q_j(\theta)}{\partial \theta} = (J_{j1}(\theta), \dots, J_{jM_0}(\theta))$$

y

$$A(\theta) = \text{diag} \left[\sqrt{-\phi''(q_1(\theta))}, \dots, \sqrt{-\phi''(q_M(\theta))} \right] J(\theta).$$

Teorema 2.3.2

Sea $\phi: (0, \infty) \rightarrow \mathbb{R}$ una función cóncava con derivada segunda continua. Si se verifican las condiciones de regularidad 1-6 así como las hipótesis i) e ii) del Teorema 2.3.1, entonces:

$$\text{a) } \sqrt{n}(\hat{\theta}_\phi - \theta^0) \approx \sqrt{n} B(\theta^0)(\hat{P} - Q(\theta^0)),$$

donde

$$B(\theta) = (A(\theta)^t A(\theta))^{-1} A(\theta)^t \text{diag} \left[\sqrt{-\phi''(q_1(\theta))}, \dots, \sqrt{-\phi''(q_M(\theta))} \right]$$

$$\text{b) } \sqrt{n}(\hat{\theta}_\phi - \theta^0) \approx N(0, \Sigma),$$

donde

$$\Sigma = B(\theta^0) \Sigma_{Q(\theta^0)} B(\theta^0)^t$$

con

$$\Sigma_{Q(\theta^0)} = \text{diag}(Q(\theta^0)) - Q(\theta^0)Q(\theta^0)^t$$

$$c) \sqrt{n} \|Q(\hat{\theta}_\phi) - Q(\theta^0)\| \leq O_p(1) \quad y \quad \sqrt{n} (Q(\hat{\theta}_\phi) - Q(\theta^0)) \approx N(0, \Sigma_Q),$$

donde

$$\Sigma_Q = J(\theta^0)\Sigma J(\theta^0)^t$$

y

$$J(\theta) = \begin{bmatrix} \frac{\partial q_j(\theta)}{\partial \theta_r} \end{bmatrix}_{\substack{j=1,\dots,M \\ r=1,\dots,M_0}}$$

Demostración

a) Al ser

$$R_\phi(P, Q(\theta)) = \sum_{i=1}^M \left\{ \phi \left[\frac{p_i + q_i(\theta)}{2} \right] - \left[\frac{\phi(p_i) + \phi(q_i(\theta))}{2} \right] \right\},$$

la derivada parcial de $R_\phi(P, Q(\theta))$ respecto de θ_r viene dada por

$$\begin{aligned} \frac{\partial R_\phi(P, Q(\theta))}{\partial \theta_r} &= \sum_{i=1}^M \left\{ \frac{1}{2} \frac{\partial q_i(\theta)}{\partial \theta_r} \phi' \left[\frac{p_i + q_i(\theta)}{2} \right] - \frac{1}{2} \frac{\partial q_i(\theta)}{\partial \theta_r} \phi'(q_i(\theta)) \right\} = \\ &= \frac{1}{2} \sum_{i=1}^M \left\{ \phi' \left[\frac{p_i + q_i(\theta)}{2} \right] - \phi'(q_i(\theta)) \right\} \frac{\partial q_i(\theta)}{\partial \theta_r}. \end{aligned} \quad (2.3.2)$$

Si se deriva la expresión anterior respecto a p_i se tiene que

$$\frac{\partial}{\partial p_i} \left[\frac{\partial R_\phi(P, Q(\theta))}{\partial \theta_r} \right] = \frac{1}{4} \phi'' \left[\frac{p_i + q_i(\theta)}{2} \right] \frac{\partial q_i(\theta)}{\partial \theta_r}.$$

Por otro lado el desarrollo de Taylor de $\frac{\partial R_\phi(\hat{P}, Q(\hat{\theta}_\phi))}{\partial \theta_r}$ en torno a $Q(\theta^0)$, viene dado por

$$\begin{aligned} \frac{\partial R_\phi(\hat{P}, Q(\hat{\theta}_\phi))}{\partial \theta_r} &= \frac{\partial R_\phi(Q(\theta^0), Q(\hat{\theta}_\phi))}{\partial \theta_r} + \\ &+ \sum_{i=1}^M \frac{\partial}{\partial p_i} \left[\frac{\partial R_\phi(P, Q(\hat{\theta}_\phi))}{\partial \theta_r} \right]_{P=P^*} (\hat{p}_i - q_i(\theta^0)) \end{aligned}$$

donde

$$\| Q(\theta^0) - P^* \| \leq \| Q(\theta^0) - \hat{P} \|.$$

Así pues

$$\begin{aligned} \sqrt{n} \left[\frac{\partial R_\phi(\hat{P}, Q(\hat{\theta}_\phi))}{\partial \theta_r} - \frac{\partial R_\phi(Q(\theta^0), Q(\hat{\theta}_\phi))}{\partial \theta_r} \right] &= \\ &= \frac{1}{4} \sqrt{n} \sum_{i=1}^M \phi'' \left[\frac{p_i^* + q_i(\hat{\theta}_\phi)}{2} \right] \frac{\partial q_i(\hat{\theta}_\phi)}{\partial \theta_r} (\hat{p}_i - q_i(\theta^0)). \end{aligned}$$

Obsérvese que

$$\hat{\theta}_\phi \xrightarrow[n \rightarrow \infty]{c. s.} \theta^0 \quad \text{y} \quad \hat{P} \xrightarrow[n \rightarrow \infty]{c. s.} Q(\theta^0)$$

implican que

$$Q(\hat{\theta}_\phi) \xrightarrow[n \rightarrow \infty]{c. s.} Q(\theta^0) \quad \text{y} \quad P^* \xrightarrow[n \rightarrow \infty]{c. s.} Q(\theta^0).$$

Por tanto

$$\phi'' \left[\frac{p_i^* + q_i(\hat{\theta}_\phi)}{2} \right] \frac{\partial q_i(\hat{\theta}_\phi)}{\partial \theta_r} \xrightarrow[n \rightarrow \infty]{c. s.} \phi''(q_i(\theta^0)) \frac{\partial q_i(\theta^0)}{\partial \theta_r}.$$

Consecuentemente

$$\begin{aligned} \sqrt{n} \left[\frac{\partial R_\phi(\hat{P}, Q(\hat{\theta}_\phi))}{\partial \theta_r} - \frac{\partial R_\phi(Q(\theta^0), Q(\hat{\theta}_\phi))}{\partial \theta_r} \right]_{r=1, \dots, M_0} &\approx \\ &\approx \sqrt{n} \frac{1}{4} \left[\sum_{i=1}^M \phi''(q_i(\theta^0)) \frac{\partial q_i(\theta^0)}{\partial \theta_r} (\hat{p}_i - q_i(\theta^0)) \right]_{r=1, \dots, M_0} = \\ &= \sqrt{n} \frac{1}{4} J(\theta^0)^t \text{diag}(\phi''(Q(\theta^0))) (\hat{P} - Q(\theta^0)) \end{aligned} \quad (2.3.3)$$

donde

$$\phi''(Q(\theta)) = (\phi''(q_1(\theta)), \dots, \phi''(q_M(\theta)))^t.$$

Al ser $\frac{\partial R_\phi(\hat{P}, Q(\hat{\theta}_\phi))}{\partial \theta_r} = 0$, se sigue de (2.3.2) que

$$\sqrt{n} \left[\frac{\partial R_\phi(\hat{P}, Q(\hat{\theta}_\phi))}{\partial \theta_r} - \frac{\partial R_\phi(Q(\theta^0), Q(\hat{\theta}_\phi))}{\partial \theta_r} \right] =$$

$$= \frac{1}{2} \sqrt{n} \left[\sum_{i=1}^M \left\{ \phi'(q_i(\hat{\theta}_\phi)) - \phi' \left[\frac{q_i(\theta^0) + q_i(\hat{\theta}_\phi)}{2} \right] \right\} \frac{\partial q_i(\hat{\theta}_\phi)}{\partial \theta_r} \right] \quad (2.3.4).$$

Desarrollando $h(\hat{\theta}_\phi) = \phi'(q_i(\hat{\theta}_\phi)) - \phi' \left[\frac{q_i(\theta^0) + q_i(\hat{\theta}_\phi)}{2} \right]$ en serie de Taylor en torno a θ^0 , se obtiene

$$h(\hat{\theta}_\phi) = h(\theta^0) + \sum_{s=1}^{M_0} \frac{\partial h(\theta^*)}{\partial \theta_s} (\hat{\theta}_s - \theta_s^0)$$

donde

$$\| \theta^0 - \theta^* \| \leq \| \theta^0 - \hat{\theta}_\phi \|.$$

Por otra parte,

$$\frac{\partial h(\theta^*)}{\partial \theta_s} = \phi''(q_i(\theta^*)) \frac{\partial q_i(\theta^*)}{\partial \theta_s} - \frac{1}{2} \phi'' \left[\frac{q_i(\theta^0) + q_i(\theta^*)}{2} \right] \frac{\partial q_i(\theta^*)}{\partial \theta_s}$$

y como $\hat{\theta}_\phi \xrightarrow[n \rightarrow \infty]{c. s.} \theta^0$,

$$\frac{\partial h(\theta^*)}{\partial \theta_s} \xrightarrow[n \rightarrow \infty]{c. s.} \left[\frac{1}{2} \phi''(q_i(\theta^0)) \right] \frac{\partial q_i(\theta^0)}{\partial \theta_s}.$$

Como $h(\theta^0) = 0$, de (2.3.4) se tiene

$$\sqrt{n} \left[\frac{\partial R_\phi(\hat{P}, Q(\hat{\theta}_\phi))}{\partial \theta_r} - \frac{\partial R_\phi(Q(\theta^0), Q(\hat{\theta}_\phi))}{\partial \theta_r} \right]_{r=1, \dots, M_0} \approx$$

$$\begin{aligned}
& \approx \sqrt{n} \frac{1}{2} \left\{ \left[\sum_{i=1}^M \left\{ \sum_{s=1}^{M_0} \left[\frac{1}{2} \phi''(q_i(\theta^0)) \frac{\partial q_i(\theta^0)}{\partial \theta_s} (\hat{\theta}_s - \theta_s^0) \right] \right\} \right] \frac{\partial q_i(\theta^0)}{\partial \theta_r} \right\}_{r=1, \dots, M_0} = \\
& = \sqrt{n} \frac{1}{4} \left[\sum_{s=1}^{M_0} \sum_{i=1}^M \phi''(q_i(\theta^0)) \frac{\partial q_i(\theta^0)}{\partial \theta_s} \frac{\partial q_i(\theta^0)}{\partial \theta_r} (\hat{\theta}_s - \theta_s^0) \right]_{r=1, \dots, M_0} = \\
& = \sqrt{n} \frac{1}{4} J(\theta^0)' \text{diag}(\phi''(Q(\theta^0))) J(\theta^0) (\hat{\theta}_\phi - \theta^0). \quad (2.3.5)
\end{aligned}$$

Por tanto de (2.3.3) y (2.3.5) se tiene que

$$\begin{aligned}
\sqrt{n} (\hat{\theta}_\phi - \theta^0) & \approx \\
& \approx \sqrt{n} \left[J(\theta^0)' \text{diag}(\phi''(Q(\theta^0))) J(\theta^0) \right]^{-1} J(\theta^0)' \text{diag}(\phi''(Q(\theta^0))) (\hat{P} - Q(\theta^0)) = \\
& = \sqrt{n} \left[A(\theta^0)' A(\theta^0) \right]^{-1} A(\theta^0)' \text{diag} \left[\sqrt{-\phi''(Q(\theta^0))} \right] (\hat{P} - Q(\theta^0))
\end{aligned}$$

como queríamos demostrar.

b) Aplicando el Teorema Central del Límite, se tiene que

$$\sqrt{n} (\hat{P} - Q(\theta^0)) \xrightarrow[n \rightarrow \infty]{L} N(0, \Sigma_{Q(\theta^0)}),$$

y en consecuencia

$$\sqrt{n} \left[A(\theta^0)' A(\theta^0) \right]^{-1} A(\theta^0)' \text{diag} \left[\sqrt{-\phi''(Q(\theta^0))} \right] (\hat{P} - Q(\theta^0)) \xrightarrow[n \rightarrow \infty]{L} N(0, \Sigma_1)$$

donde

$$\Sigma_1 = \left[A(\theta^0)' A(\theta^0) \right]^{-1} A(\theta^0)' \text{diag} \left[\sqrt{-\phi''(Q(\theta^0))} \right] \left[\text{diag}(Q(\theta^0)) - \right. \\ \left. - Q(\theta^0) Q(\theta^0)' \right] \text{diag} \left[\sqrt{-\phi''(Q(\theta^0))} \right] A(\theta^0) \left[A(\theta^0)' A(\theta^0) \right]^{-1}.$$

Por tanto el resultado enunciado es inmediato a partir del apartado a).

c) Si se considera el desarrollo de Taylor de $q_j(\hat{\theta}_\phi)$ en torno a θ^0

$$q_j(\hat{\theta}_\phi) = q_j(\theta^0) + \sum_{s=1}^{M_0} \frac{\partial q_j(\theta^*)}{\partial \theta_s} (\hat{\theta}_s - \theta_s^0), \quad j=1, \dots, M,$$

donde

$$\| \theta^0 - \theta^* \| \leq \| \theta^0 - \hat{\theta}_\phi \|$$

o lo que es lo mismo

$$Q(\hat{\theta}_\phi) - Q(\theta^0) = \left[\frac{\partial q_j(\theta^*)}{\partial \theta_s} \right]_{\substack{j=1, \dots, M \\ s=1, \dots, M_0}} (\hat{\theta}_\phi - \theta^0).$$

Como $\hat{\theta}_\phi \xrightarrow[n \rightarrow \infty]{c. s.} \theta^0$, se tiene que

$$\sqrt{n} \left[Q(\hat{\theta}_\phi) - Q(\theta^0) \right] \xrightarrow[n \rightarrow \infty]{L} N(0, \Sigma_Q),$$

con

$$\Sigma_Q = J(\theta^0) \Sigma J(\theta^0)'$$

Entonces

$$\sqrt{n} \| Q(\hat{\theta}_\phi) - Q(\theta^0) \| \leq O_p(1).$$

■

En el Teorema 2.3.3 se obtiene un resultado análogo aunque más fuerte al del Teorema 2.3.2 a) cambiando alguna de las condiciones de regularidad. Para obtener este resultado se utiliza el Teorema de la Función Implícita que previamente se enuncia.

"Sea $F=(F_1, \dots, F_{M_0}): \mathbb{R}^{M+M_0} \longrightarrow \mathbb{R}^{M_0}$ una función continuamente diferenciable en un conjunto abierto $U \subset \mathbb{R}^{M+M_0}$ que contiene al punto $(x^*=(x_1^*, \dots, x_M^*)^t, x_0=(x_1^0, \dots, x_{M_0}^0)^t)$ para el cual $F(x^*, x_0) = 0$ y se supone que la matriz

$$\left[\frac{\partial F_i}{\partial x_j} \right]_{\substack{i=1, \dots, M_0 \\ j=M+1, \dots, M+M_0}}$$

es no singular en (x^*, x_0) . Entonces existe un entorno M -dimensional U_0 de x^* en \mathbb{R}^M y una única función continuamente diferenciable $g: U_0 \longrightarrow \mathbb{R}^{M_0}$ de forma que $g(x^*) = x_0$ y $F(x, g(x)) = 0 \forall x \in U_0$ ".

Teorema 2.3.3

Sea $\phi:(0, \infty) \longrightarrow \mathbb{R}$ una función cóncava con derivada segunda continua. Si se verifican las condiciones de regularidad 1-6 de Birch y suponiendo que la aplicación $Q: \Theta \longrightarrow \Delta_M$ tiene derivadas parciales segundas continuas en un entorno de θ^0 , se tiene que

$$\hat{\theta}_\phi = \theta^0 + (A(\theta^0)^t A(\theta^0))^{-1} A(\theta^0)^t \text{diag} \left[\sqrt{-\phi''(Q(\theta^0))} \right] (\hat{P} - Q(\theta^0)) + o(\|\hat{P} - Q(\theta^0)\|)$$

donde $\hat{\theta}_\phi$ es único en un entorno de θ^0 .

Demostración

Sea I^M el interior del cubo unitario M -dimensional con $\Delta_M \subset I^M$ y sea U un entorno de θ^0 en el cual $Q: \Theta \longrightarrow \Delta_M$ tiene derivadas parciales segundas continuas. Considérese la función

$$F=(F_1, \dots, F_{M_0}): I^M \times U \longrightarrow R^{M_0}$$

de tal forma que

$$F_j(p_1, \dots, p_M; \theta_1, \dots, \theta_{M_0}) = \frac{\partial R_\phi(P, Q(\theta))}{\partial \theta_j}, \quad \forall j=1, \dots, M_0.$$

Para $\pi_i = q_i(\theta^0)$, $i=1, \dots, M$, se tiene que

$$F_j(\pi_1, \dots, \pi_M; \theta_1^0, \dots, \theta_{M_0}^0) = 0, \quad \forall j=1, \dots, M_0,$$

ya que

$$\frac{\partial R_\phi(P, Q(\theta))}{\partial \theta_j} = \frac{1}{2} \sum_{i=1}^M \left\{ \phi' \left[\frac{p_i + q_i(\theta)}{2} \right] - \phi'(q_i(\theta)) \right\} \frac{\partial q_i(\theta)}{\partial \theta_j} \quad \forall j=1, \dots, M_0.$$

Seguidamente se comprobará que la matriz

$$\frac{\partial F}{\partial \theta} = \left[\frac{\partial F_j}{\partial \theta_r} \right]_{\substack{j=1, \dots, M_0 \\ r=1, \dots, M_0}}$$

es no singular en θ^0 . En efecto, al ser

$$\frac{\partial}{\partial \theta_r} \left[\frac{\partial R_\phi(P, Q(\theta))}{\partial \theta_j} \right] =$$

$$\begin{aligned}
&= \frac{1}{2} \sum_{i=1}^M \left\{ \frac{\partial q_i(\theta)}{\partial \theta_r} \left[\frac{1}{2} \phi'' \left(\frac{p_i + q_i(\theta)}{2} \right) - \phi''(q_i(\theta)) \right] \frac{\partial q_i(\theta)}{\partial \theta_j} + \right. \\
&\quad \left. + \left[\phi' \left(\frac{p_i + q_i(\theta)}{2} \right) - \phi'(q_i(\theta)) \right] \frac{\partial^2 q_i(\theta)}{\partial \theta_j \partial \theta_r} \right\}
\end{aligned}$$

se tiene que

$$\begin{aligned}
\frac{\partial \underline{F}}{\partial \theta^0} &= \left(\frac{\partial F}{\partial \theta} \right)_{\theta=\theta^0} = \left[\frac{\partial}{\partial \theta_r} \left(\frac{\partial R_\phi(\pi, Q(\theta^0))}{\partial \theta_j} \right) \right]_{\substack{j=1, \dots, M_0 \\ r=1, \dots, M_0}} = \\
&= \left[\frac{1}{2} \sum_{i=1}^M \left\{ \frac{\partial q_i(\theta^0)}{\partial \theta_r} \frac{\partial q_i(\theta^0)}{\partial \theta_j} \left[-\frac{1}{2} \right] \phi''(q_i(\theta^0)) \right\} \right]_{\substack{j=1, \dots, M_0 \\ r=1, \dots, M_0}} = \\
&= \frac{1}{4} A(\theta^0)^t A(\theta^0).
\end{aligned}$$

Teniendo en cuenta que si B es una matriz $p \times q$ y C una matriz no singular de orden q , entonces $\text{rango}(BC) = \text{rango}(B)$, se tiene que $A(\theta^0)^t$ y $A(\theta^0)$ tienen rango M_0 considerando

$$B = \left[\frac{\partial q_i(\theta^0)}{\partial \theta_r} \right]_{\substack{i=1, \dots, M \\ r=1, \dots, M_0}}^t \quad \text{y} \quad C = \text{diag} \left[\sqrt{-\phi''(Q(\theta^0))} \right]_{M \times M}$$

Por otra parte,

$$\text{rango}(A(\theta^0)^t A(\theta^0)) = \text{rango}(A(\theta^0) A(\theta^0)^t) = \text{rango}(A(\theta^0)) = M_0.$$

En definitiva la matriz

$$\left[\frac{\partial F_i}{\partial \theta_j} \right]_{\substack{i=1, \dots, M_0 \\ j=1, \dots, M_0}}$$

es no singular en θ^0 .

Aplicando el Teorema de la Función Implícita existe un entorno M -dimensional U_0 de $\pi = (\pi_1, \dots, \pi_M)^t$ en \mathbb{R}^M y una única función diferenciable con continuidad $\tilde{\theta}: U_0 \longrightarrow \mathbb{R}^{M_0}$ de tal forma que

$$F(P, \tilde{\theta}(P)) = 0 \quad \forall P \in U_0$$

y

$$\tilde{\theta}(\pi) = \theta^0.$$

Aplicando la regla de la cadena se tiene

$$\frac{\partial F(P, \tilde{\theta}(P))}{\partial P} + \frac{\partial F(P, \tilde{\theta}(P))}{\partial \tilde{\theta}(P)} \frac{\partial \tilde{\theta}(P)}{\partial P} = 0$$

y para $P=\pi$

$$\frac{\partial F}{\partial \pi} + \frac{\partial F}{\partial \theta^0} \frac{\partial \theta^0}{\partial \pi} = 0.$$

Ahora bien,

$$\frac{\partial F}{\partial \theta^0} = \frac{1}{4} A(\theta^0)^t A(\theta^0)$$

y

$$\frac{\partial F}{\partial \pi} = \frac{1}{4} J(\theta^0)^t \text{diag}(\phi''(\pi)) = - \frac{1}{4} A(\theta^0)^t \text{diag} \left[\sqrt{-\phi''(Q(\theta^0))} \right]$$

con lo cual

$$\frac{\partial \theta^0}{\partial \pi} = (A(\theta^0)'A(\theta^0))^{-1}A(\theta^0)' \text{diag} \left[\sqrt{-\phi''(Q(\theta^0))} \right].$$

Desarrollando por Taylor $\tilde{\theta}(P)$ en un entorno de π , se tiene

$$\tilde{\theta}(P) = \tilde{\theta}(\pi) + \left[\frac{\partial \tilde{\theta}}{\partial P} \right]_{P=\pi} (P-\pi) + o(\|P-\pi\|)$$

y como $\tilde{\theta}(\pi) = \theta^0$, se llega a

$$\tilde{\theta}(P) = \theta^0 + (A(\theta^0)'A(\theta^0))^{-1}A(\theta^0)' \text{diag} \left[\sqrt{-\phi''(Q(\theta^0))} \right] (P-\pi) + o(\|P-\pi\|).$$

Ahora bien $\hat{P} \xrightarrow[n \rightarrow \infty]{c. s.} \pi$, por lo tanto $\hat{P} \in U_0$ y como consecuencia $\tilde{\theta}(\hat{P})$ es solución única de las ecuaciones

$$\frac{\partial R_\phi(\hat{P}, \tilde{\theta}(\hat{P}))}{\partial \theta_j} = 0, \quad j=1, \dots, M_0,$$

en un entorno de π , luego $\tilde{\theta}(\hat{P})$ es el estimador de mínima R_ϕ -divergencia, $\hat{\theta}_\phi$, que como consecuencia de lo anterior verifica

$$\hat{\theta}_\phi = \theta^0 + (A(\theta^0)'A(\theta^0))^{-1}A(\theta^0)' \text{diag} \left[\sqrt{-\phi''(Q(\theta^0))} \right] (\hat{P}-Q(\theta^0)) + o(\|\hat{P}-Q(\theta^0)\|). \quad \blacksquare$$

2.4.- Robustez de la función de mínima R_ϕ -divergencia.

En los apartados anteriores se ha supuesto que la distribución que rige el modelo discretizado es $Q(\theta) = (q_1(\theta), \dots, q_M(\theta))^t$. En este apartado se consideran desviaciones del modelo dadas por la familia

$$Q_\varepsilon(\theta) = (1-\varepsilon)Q(\theta) + \varepsilon P$$

con $\varepsilon > 0$, $\theta \in \Theta$ y $P \in \Delta_M$.

Sea $\theta_\phi^\varepsilon(P)$ el vector que minimiza la función

$$g_\varepsilon(P, \theta) = \sum_{i=1}^M \phi \left[\frac{p_i + q_i(\theta, \varepsilon)}{2} \right] - \frac{1}{2} \left\{ \sum_{i=1}^M \phi(p_i) + \sum_{i=1}^M \phi(q_i(\theta, \varepsilon)) \right\}$$

donde $q_1(\theta, \varepsilon), \dots, q_M(\theta, \varepsilon)$ son las componentes del vector $Q_\varepsilon(\theta)$.

Para garantizar la robustez de $\theta_\phi(P)$, lo que interesa comprobar es que a pequeñas desviaciones de $Q(\theta)$ le corresponden pequeñas desviaciones de $\theta_\phi^\varepsilon(P)$; o bien, analíticamente, que

$$\lim_{\varepsilon \rightarrow 0} \theta_\phi^\varepsilon(P) = \theta_\phi(P).$$

En el siguiente teorema se dan condiciones que garantizan la robustez de la función de mínima R_ϕ -divergencia.

Teorema 2.4.1

Supóngase que se verifican las condiciones del Teorema 2.3.1, entonces, se tiene que

$$\lim_{\varepsilon \rightarrow 0} \theta_\phi^\varepsilon(P) = \theta_\phi(P).$$

Demostración

Sea $\{\varepsilon_n\}$ una sucesión arbitraria de números positivos verificando $\varepsilon_n \xrightarrow{n \rightarrow \infty} 0$. Por ser ϕ continua y $q_i(\theta, \varepsilon_n) \xrightarrow{\varepsilon_n \rightarrow 0} q_i(\theta)$, $i=1, \dots, M$, se tiene que

$$g_{\varepsilon_n}(P, \theta) \xrightarrow{\varepsilon_n \rightarrow 0} g_0(P, \theta) \quad \forall \theta \in \Theta.$$

Al ser Θ compacto la convergencia puntual implica la convergencia uniforme y en consecuencia

$$\lim_{\varepsilon_n \rightarrow 0} \sup_{\theta \in \Theta} |g_{\varepsilon_n}(P, \theta) - g_0(P, \theta)| = 0$$

lo que implica que

$$\lim_{\varepsilon_n \rightarrow 0} \left| \inf_{\theta \in \Theta} g_{\varepsilon_n}(P, \theta) - \inf_{\theta \in \Theta} g_0(P, \theta) \right| = 0,$$

o lo que es lo mismo

$$\lim_{\varepsilon_n \rightarrow 0} |g_{\varepsilon_n}(P, \theta_{\phi}^{\varepsilon_n}(P)) - g_0(P, \hat{\theta}_{\phi}(P))| = 0.$$

En definitiva se ha demostrado que

$$\lim_{\varepsilon_n \rightarrow 0} g_{\varepsilon_n}(P, \theta_{\phi}^{\varepsilon_n}(P)) = g_0(P, \hat{\theta}_{\phi}(P)). \quad (2.4.1)$$

Si $\lim_{\varepsilon_n \rightarrow 0} \theta_{\phi}^{\varepsilon_n}(P) \neq \hat{\theta}_{\phi}(P)$, resulta que por ser Θ compacto existe una subsucesión

$$\{\theta_{\phi}^{\delta_n}(P)\} \subset \{\theta_{\phi}^{\varepsilon_n}(P)\}$$

verificando

$$\lim_{\substack{\delta \rightarrow 0 \\ n}} \theta_{\phi}^{\delta_n}(P) = \theta_* \neq \hat{\theta}_{\phi}(P).$$

De (2.4.1) se tiene que $g_0(P, \theta_*) = g_0(P, \hat{\theta}_{\phi}(P))$ con $\theta_* \neq \hat{\theta}_{\phi}(P)$, lo cual contradice la unicidad de $\hat{\theta}_{\phi}(P)$.

Finalmente, de la arbitrariedad de la sucesión $\{\varepsilon_n\}$, se deduce el teorema. ■

Otra forma más general de enfocar la robustez es suponer que la verdadera distribución que rige el modelo discretizado, $\pi \in \Delta_M$, cumple

$$\|\pi - Q(\theta)\| < \varepsilon \quad \text{para un } \theta \in \Theta$$

y comprobar que si ε es pequeño, el valor $\theta_{\phi}(\pi)$ es próximo a $\theta_{\phi}(Q(\theta)) = \theta$.

Teorema 2.4.2

Supóngase que se verifican las condiciones del teorema anterior y sea $\pi \in \Delta_M$. Entonces:

$$\lim_{\|\pi - Q(\theta)\| \rightarrow 0} \theta_{\phi}(\pi) = \theta_{\phi}(Q(\theta)) = \theta.$$

Demostración

Inmediata por ser θ_{ϕ} continua. ■

2.5.-Resultados computacionales.

Una familia de R_{ϕ} -divergencias especialmente importante, como pusieron de manifiesto Burbea y Rao (1982) es la que resulta al considerar la función

$$\phi_{\alpha}(x) = \frac{1}{1-\alpha}(x^{\alpha}-x) \quad \alpha > 0, \alpha \neq 1$$

$$\text{y } \phi_1(x) = \lim_{\alpha \rightarrow 1} \phi_{\alpha}(x) = -x \ln x.$$

En este caso es inmediato, a partir del Teorema 2.3.3, que el estimador de mínima $R_{\phi_{\alpha}}$ -divergencia, $\hat{\theta}_{\phi_{\alpha}}$, se puede expresar de la forma

$$\hat{\theta}_{\phi_{\alpha}} = \theta^0 + (A(\theta^0)'A(\theta^0))^{-1} A(\theta^0)' \text{diag} \left[(Q(\theta^0))^{\frac{\alpha}{2}-1} \right] (\hat{P} - Q(\theta^0)) + o(\|\hat{P} - Q(\theta^0)\|)$$

donde

$$A(\theta) = \text{diag} \left[(Q(\theta))^{\frac{\alpha}{2}-1} \right] J(\theta).$$

Obsérvese que si se considera la R-divergencia, R_{ϕ_1} , el estimador de mínima R-divergencia viene dado por

$$\hat{\theta}_{\phi_1} = \theta^0 + (A(\theta^0)'A(\theta^0))^{-1} A(\theta^0)' \text{diag} \left[(Q(\theta^0))^{-\frac{1}{2}} \right] (\hat{P} - Q(\theta^0)) + o(\|\hat{P} - Q(\theta^0)\|)$$

donde

$$A(\theta) = \text{diag} \left[(Q(\theta))^{-\frac{1}{2}} \right] J(\theta)$$

y

$$\sqrt{n}(\hat{\theta}_{\phi_1} - \theta^0) \xrightarrow[n \rightarrow \infty]{L} N(0, I(\theta^0)^{-1}),$$

siendo $I(\theta)$ la matriz de Información de Fisher asociada al modelo discretizado.

Recordemos que un estimador $\hat{\theta}_n$ tal que $\sqrt{n}(\hat{\theta}_n - \theta^0) \xrightarrow[n \rightarrow \infty]{L} N(0, I(\theta^0))$ se denomina asintóticamente eficiente o BAN (Best Asymptotically Normal). Por tanto el estimador $\hat{\theta}_{\phi_1}$ es un estimador BAN.

En este apartado se resuelven los dos siguientes problemas:

(1) Calcular el valor de α , α_{\min} , que minimice el error cuadrático medio que se produce al estimar con la R_{ϕ_α} -divergencia los parámetros de una población Normal y de una Weibull.

(2) Calcular los estimadores de mínima R_{ϕ_α} -divergencia de los parámetros de una población Normal y de una Weibull para diferentes valores de α comparándolos con los estimadores de máxima verosimilitud y de mínima distancia de Kolmogorov D_n .

En primer lugar se aborda el problema (1), pero antes se define lo que se entiende por estimador de mínima D_n y por función de distribución Weibull.

Definición 2.5.1

Se define el estimador de mínima D_n para una familia de distribuciones $\{F_\theta(x), \theta \in \Theta\}$, como el valor $\hat{\theta}_n \in \Theta$ tal que

$$D_n(\hat{\theta}) = \min(D_n(\theta), \theta \in \Theta)$$

siendo

$$D_n(\theta) = \sup_{x \in \mathbb{R}} \{ |F_n^*(x) - F_\theta(x)| \} = \max\{D_n^+(\theta), D_n^-(\theta)\}$$

donde

$$D_n^+(\theta) = \sup_{x \in \mathbb{R}} \{ F_n^*(x) - F_\theta(x) \} = \max\{0, \max_{i=1, \dots, n} \{ \frac{i}{n} - F_\theta(x_{(i)}) \} \},$$

$$D_n^-(\theta) = \sup_{x \in \mathbb{R}} \{ F_\theta(x) - F_n^*(x) \} = \max\{0, \max_{i=1, \dots, n} \{ F_\theta(x_{(i)}) - \frac{i-1}{n} \} \}$$

y $F_n^*(x)$ es la función de distribución empírica de una muestra x_1, \dots, x_n de la población y $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ son los estadísticos de orden.

Definición 2.5.2

Se dice que una variable aleatoria X tiene una distribución Weibull de parámetros (b,c) , $b > 0$, $c > 0$, si la función de distribución de X es

$$F_\theta(x) = 1 - \exp\left\{-\left[\frac{x}{b}\right]^c\right\}, \quad x \geq 0$$

donde $\theta = (b,c)$.

El parámetro b es el parámetro de escala y c es el parámetro de forma.

El esquema general que se seguirá para encontrar el α óptimo en el sentido de (1) es el siguiente:

Paso 1: Se fija:

- (a) tamaño muestral (n),
- (b) número de clases en la partición (M),
- (c) número de muestras simuladas (N).

Se calculan los valores a_0, \dots, a_M , dados por

$$\int_{a_{i-1}}^{a_i} f_{\theta}(x) dx = 1/M, \quad i=1, \dots, M$$

de forma que $A_i = (a_{i-1}, a_i]$, $i=1, \dots, M$, defina una partición $\mathcal{X} \subset \mathbb{R}$.

Paso 2: Se minimiza la función

$$\text{ecm}(\alpha) = \frac{\sum_{i=1}^N (\hat{\theta}_{1\alpha}^i - \theta_1)^2}{2N} + \frac{\sum_{i=1}^N (\hat{\theta}_{2\alpha}^i - \theta_2)^2}{2N}$$

donde $\hat{\theta}_{1\alpha}^i$ es el estimador de mínima $R_{\phi_{\alpha}}$ -divergencia de θ_1 y $\hat{\theta}_{2\alpha}^i$ de θ_2 para la muestra i . Estos valores se calculan en el paso 3.

Paso 3: Dado α fijo, hacer para $i=1$ hasta N

- (a) Generar una muestra aleatoria simple de tamaño n
 - (b) Calcular la frecuencia relativa de las clases del paso anterior
 - (c) Minimizar en θ la función $R_{\phi}(\hat{P}, Q(\theta))$.
- Ir al paso 2.

En la Tabla 2.5.1 y 2.5.2 aparece el valor α , α_{\min} , que minimiza el error cuadrático medio que se comete cuando se utiliza el estimador

de mínima $R_{\phi_{\alpha}}$ -divergencia para estimar los parámetros de una Weibull de parámetros $b=1$ y $c=1$, $We(1,1)$, y $b=1$ y $c=2$, $We(1,2)$ y los parámetros de una Normal $\mu=0$ y $\sigma=1$, $N(0,1)$, y $\mu=0$ y $\sigma=2$, $N(0,2)$, para diferentes tamaños muestrales.

Tabla 2.5.1

	We(1,1)	We(1,2)
n = 20	1 .76875	1 .197607
n = 40	1 .1	0 .613597
n = 60	1 .10158	1 .205

Tabla 2.5.2

	N(0,1)	N(0,2)
n = 20	0 .298348	0.224309
n = 40	0 .280321	0.290615
n = 60	0 .325411	0.298412

Debido a que el algoritmo de minimización que se utiliza para calcular cada uno de los valores de las tablas anteriores puede ser sensible a los puntos iniciales, se ha evaluado cada α_{\min} con diversos puntos iniciales y comprobado que en este caso la diferencia entre los α_{\min} partiendo de un punto inicial u otro es ínfima, por lo que se ha tomado aquél que da el menor error cuadrático medio.

Por otra parte, el esquema general que se sigue para calcular el estimador de mínima $R_{\phi_{\alpha}}$ -divergencia, problema (2), es el siguiente:

Paso 1: Se fija:

- (a) tamaño muestral (n),
- (b) número de clases en la partición (M),
- (c) número de muestras simuladas (N).

Se calculan los valores a_0, \dots, a_M , dados por

$$\int_{a_{i-1}}^{a_i} f_{\theta}(x)dx = 1/M, \quad i=1, \dots, M$$

de forma que $A_i=(a_{i-1}, a_i]$, $i=1, \dots, M$, defina una partición de $\mathcal{X} \subset \mathbb{R}$.

Paso 2: Dado α fijo, hacer para $i=1$ hasta N

- (a) Generar una muestra aleatoria simple de tamaño n
- (b) Calcular la frecuencia relativa de las clases del paso anterior
- (c) Minimizar en θ la función $R_{\phi}(\hat{P}, Q(\theta))$

Paso 3: Se considera $\hat{\theta}_{\phi}$ la media de los valores obtenidos al minimizar la función R_{ϕ} en el Paso 2(c) en todas las muestras y $ecm(\alpha)$ el error cuadrático medio de los parámetros estimados definido anteriormente.

Las Tablas 2.5.3, 2.5.4, 2.5.5 y 2.5.6 contienen los estimadores de máxima verosimilitud (EMV), mínimo D_n (ED_n) y mínima $R_{\phi\alpha}$ -divergencia ($ER_{\phi\alpha}$) para poblaciones Weibull de parámetros $b=1$, $c=1$; $b=1$, $c=2$; para poblaciones Normal de parámetros $\mu=0$, $\sigma=1$; $\mu=0$, $\sigma=2$, respectivamente. Estos valores se han calculado por simulación para 1000 muestras, 6 clases y tamaños muestrales $n=20$, 40 y 60. Se varía el parámetro de forma manteniendo fijo el de escala porque las

estimaciones de c y σ son, en general, peores que las de b y μ . De esta forma, parece más interesante observar el comportamiento de las estimaciones \hat{c} y $\hat{\sigma}$. De hecho, las estimaciones en la Tabla 2.5.4 y 2.5.6 son peores que en la 2.5.3 y 2.5.5 para c y σ , respectivamente. También aparecen en estas tablas la suma de los errores cuadráticos medios de los dos parámetros.

Tabla 2.5.3

We (1,1)		n = 20	n = 40	n = 60
EMV	\hat{b}	0.998783	0.994317	0.994150
	\hat{c}	1.063596	1.029655	1.019258
	ecm	0.055893	0.025969	0.014047
ED _n	\hat{b}	0.984651	1.006958	0.978376
	\hat{c}	1.565195	1.137134	1.185521
	ecm	1.023289	0.121828	0.108134
ER _{ϕ_1}	\hat{b}	1.008978	1.015399	0.983971
	\hat{c}	1.386812	1.068477	1.117729
	ecm	0.745879	0.112524	0.091662
ER _{ϕ_2}	\hat{b}	1.002849	1.006966	0.979121
	\hat{c}	1.414069	1.077251	1.133801
	ecm	0.742312	0.098914	0.096844
ER _{$\phi_{\alpha_{\min}}$}	\hat{b}	1.009632	1.006264	0.978674
	\hat{c}	1.396201	1.051318	1.105390
	ecm	0.734901	0.093519	0.083306

Los programas que calculan los ER _{ϕ_{α}} y los ED_n necesitan partir de una estimación inicial. Para las poblaciones Weibull, esta estimación se realiza por el método dado por Dannenbring (1977), es decir:

$$\hat{b} = x_{((0.6321n)+1)}$$

y

$$\hat{c} = \frac{\ln(\ln 2)}{\ln(x_M/\hat{b})}$$

donde x_M es la mediana muestral.

Tabla 2.5.4

We (1,2)		n = 20	n = 40	n = 60
EMV	\hat{b}	0.992503	0.993879	0.994862
	\hat{c}	2.127185	2.059309	2.038516
	ecm	0.093951	0.039142	0.023382
ED _n	\hat{b}	0.985581	0.997818	0.991716
	\hat{c}	3.030805	2.214321	2.229546
	ecm	3.926376	0.385762	0.280197
ER _{ϕ_1}	\hat{b}	0.992771	0.999766	0.991852
	\hat{c}	2.699849	2.066902	2.135167
	ecm	2.675244	0.271025	0.208801
ER _{ϕ_2}	\hat{b}	0.993954	0.997541	0.989239
	\hat{c}	2.783239	2.145965	2.215762
	ecm	2.700610	0.326649	0.278683
ER _{$\phi_{\alpha_{min}}$}	\hat{b}	0.993631	0.997473	0.988435
	\hat{c}	2.651145	2.019232	2.129087
	ecm	2.560270	0.231978	0.187459

En el caso de poblaciones Normales se han tomado como estimadores iniciales los de máxima verosimilitud.

Puesto que el EMV se basa en valores originales Weibull o Normal, y no en valores discretizados, el error cuadrático medio (ecm) cometido por este procedimiento es más pequeño que el cometido por el estimador de mínima R_{ϕ_α} -divergencia para $n=40$ y 60 y mayor que el cometido por el ER_{ϕ_α} para $n=20$. Sin embargo, el ecm cometido por el estimador de mínimo D_n es más grande que el ER_{ϕ_α} en todos los casos aunque el primero se basa en valores originales y el segundo clasifica los valores originales en clases.

Tabla 2.5.5

N(0,1)		n = 20	n = 40	n = 60
EMV	$\hat{\mu}$	-0.011351	-0.004091	-0.002958
	$\hat{\sigma}$	0.960474	0.978162	0.984269
	ecm	0.038515	0.018526	0.012468
ED_n	$\hat{\mu}$	-0.014687	-0.004947	-0.003634
	$\hat{\sigma}$	0.972221	0.985606	0.987717
	ecm	0.043739	0.021340	0.014099
ER_{ϕ_1}	$\hat{\mu}$	-0.010202	-0.005092	-0.001042
	$\hat{\sigma}$	0.983168	0.989895	0.987750
	ecm	0.041413	0.022099	0.014360
ER_{ϕ_2}	$\hat{\mu}$	-0.009914	-0.004122	-0.002576
	$\hat{\sigma}$	0.978724	0.985323	0.986545
	ecm	0.041710	0.022388	0.014605
$ER_{\phi_\alpha \min}$	$\hat{\mu}$	-0.007035	-0.004988	-0.003618
	$\hat{\sigma}$	0.980223	0.983745	0.988468
	ecm	0.037700	0.020522	0.013907

Entonces, el comportamiento de los ER_{ϕ_α} es bueno cuando las

observaciones se clasifican en clases.

Tabla 2.5.6

N(0,2)		n = 20	n = 40	n = 60
EMV	$\hat{\mu}$	-0.022703	-0.008183	-0.005917
	$\hat{\sigma}$	1.920949	1.956325	1.968539
	ecm	0.154061	0.074107	0.049874
ED _n	$\hat{\mu}$	-0.028261	-0.009410	-0.007361
	$\hat{\sigma}$	1.947619	1.970374	1.975137
	ecm	0.175630	0.084122	0.056079
ER _{ϕ_1}	$\hat{\mu}$	-0.019088	-0.009761	-0.002724
	$\hat{\sigma}$	1.972188	1.975937	1.977961
	ecm	0.169617	0.087569	0.058077
ER _{ϕ_2}	$\hat{\mu}$	-0.017345	-0.006358	-0.005403
	$\hat{\sigma}$	1.955918	1.967490	1.971419
	ecm	0.167734	0.088731	0.057644
ER _{$\phi_{\alpha_{\min}}$}	$\hat{\mu}$	-0.018334	-0.007323	-0.006046
	$\hat{\sigma}$	1.954678	1.968161	1.970706
	ecm	0.149621	0.080399	0.054078

Obsérvese que en general todos los errores cuadráticos medios que se cometen al utilizar los estimadores considerados son bastante más pequeños para las poblaciones Normales que para las Weibull.

CAPITULO III

CONTRASTES DE BONDAD DE AJUSTE BASADOS EN LA R_ϕ -DIVERGENCIA: COMPORTAMIENTO ASINTOTICO

3.1.- Introducción.

3.2.- R_ϕ -divergencia y bondad de ajuste: Hipótesis nula simple. Eficiencia Pitman.

3.3.- R_ϕ -divergencia y bondad de ajuste: Hipótesis nula compuesta.

3.4.- El problema de bondad de ajuste cuando se incrementa el número de clases.

3.1.- Introducción

Una buena parte de los tests de bondad de ajuste se reducen a realizar contrastes sobre el parámetro $\pi = (\pi_1, \dots, \pi_M)$ de una distribución multinomial

$$P_{\pi}(X = x) = \frac{n!}{x_1! \dots x_M!} \pi_1^{x_1} \dots \pi_M^{x_M}$$

donde X toma valores en el conjunto

$$\left\{ x = (x_1, \dots, x_M)^t / \sum_{i=1}^M x_i = n, x_i \geq 0, x_i \in Z^+, i=1, \dots, M \right\}$$

y $\pi \in \Delta_M$.

Esto es posible cuando se discretiza el modelo en estudio y se denota por π_i la probabilidad de la clase A_i , $i=1, \dots, M$. En este caso, el problema de bondad de ajuste se reduce a contrastar una hipótesis H_0 acerca del vector de probabilidad generalmente desconocido $\pi = (\pi_1, \dots, \pi_M)$ que podemos plantear como

$$H_0: \pi = \pi_0 = (\pi_{01}, \dots, \pi_{0M}) \in T \quad (3.1.1)$$

donde $T \subset \Delta_M$ es el conjunto donde π puede tomar valores en la hipótesis

nula. Esta hipótesis puede especificar π completamente en cuyo caso T es únicamente un punto (es decir, hipótesis simple) o puede ser π una función de parámetros desconocidos que deben ser estimados a partir de la muestra (es decir, hipótesis compuesta).

La hipótesis (3.1.1) se valora comparando las frecuencias esperadas en cada clase, dadas por $n\pi_0$, con las frecuencias observadas x . Si hay mucha diferencia entre las frecuencias observadas y las esperadas entonces se rechaza este modelo y se busca otro modelo que se ajuste más a los datos.

El estadístico de bondad de ajuste más utilizado para contrastar (3.1.1) es el X^2 de Pearson (introducido por Pearson, 1900);

$$X^2 = \sum_{i=1}^M \frac{(X_i - n\hat{\pi}_i)^2}{n\hat{\pi}_i}$$

que se distribuye asintóticamente como una ji-cuadrado con $M-1$ grados de libertad en el caso de hipótesis nula simple donde $\hat{\pi}_i = \pi_{0i}$, $i=1, \dots, M$. En el caso de hipótesis compuesta, donde $\hat{\pi}_i$ se elige como el elemento de T más consistente con los datos de la muestra, Pearson (1900) recomendó mantener los $M-1$ grados de libertad del caso anterior pero posteriormente Fisher (1924) obtuvo que los verdaderos grados de libertad son $M-M_0-1$ donde M_0 es el número de parámetros estimados. El método más conocido para elegir $\hat{\pi}$, cuando la hipótesis nula es compuesta, es el de máxima verosimilitud, a partir de los datos discretizados, que como se vio en el capítulo anterior es el caso particular del método de mínima distancia cuando se utiliza la medida de divergencia de Kullback. Es obvio, por tanto, que $\hat{\pi}$ se puede obtener utilizando el método de mínima R_ϕ -divergencia estudiado en el capítulo anterior. El estudio del contraste de bondad de ajuste cuando se utiliza este método de estimación se llevará a cabo en el apartado 3.3.

Cochran (1952) además de presentar una recopilación del desarrollo histórico del estadístico X^2 de Pearson y sus aplicaciones estudió algunos estadísticos alternativos a éste. Entre ellos se encuentra el estadístico del logaritmo del cociente de verosimilitud;

$$G^2 = 2 \sum_{i=1}^M X_i \ln \frac{X_i}{n \pi_i},$$

que es asintóticamente equivalente al estadístico X^2 de Pearson (Neyman (1949)).

Muchos han sido los estudios realizados por diferentes autores sobre las diferencias existentes entre los estadísticos X^2 de Pearson y G^2 , entre otros destacar:

(i) comparaciones para muestras finitas bajo la hipótesis nula (Chapman (1976), Larntz (1978)),

(ii) comparaciones de la potencia asintótica para muestras finitas y varias hipótesis alternativas, incluyendo el efecto de variar los intervalos de las clases (Hoeffding (1965); West y Kempthorne (1972); Goldstein, Wolf y Dillon (1976)),

(iii) cálculo de la distribución asintótica bajo la hipótesis nula y varias alternativas, para M creciendo con n (Holst (1972); Morris (1975); Koehler y Larntz (1980)),

(iv) impacto en los estadísticos de los diferentes métodos de estimación de parámetros (Moore y Spruill (1975)).

Otros estadísticos de bondad de ajuste han sido propuestos a lo largo de los últimos 40 años. Entre estos se incluyen el estadístico de Freeman-Tukey (Freeman y Tukey (1950); Bishop y otros (1975)), el

cual, siguiendo Fienberg (1979) y Moore (1986), se define como

$$F^2 = 4 \sum_{i=1}^M \left[X_i^{1/2} - (n\hat{\pi}_i)^{1/2} \right]^2;$$

el estadístico del logaritmo del cociente de verosimilitud modificado (Kullback (1959, 1985)) que viene dado por

$$GM^2 = 2n \sum_{i=1}^M \hat{\pi}_i \ln \frac{n\hat{\pi}_i}{X_i},$$

y el estadístico X^2 modificado de Neyman (Neyman (1949))

$$NM^2 = \sum_{i=1}^M \frac{(X_i - n\hat{\pi}_i)^2}{X_i}.$$

Todos estos estadísticos comparten la característica común de ser iguales a cero cuando X/n y $\hat{\pi}$ son iguales y la de incrementar su tamaño cuando aumenta la diferencia entre X/n y $\hat{\pi}$.

Además diversos autores han demostrado que bajo la hipótesis nula la distribución asintótica de estos estadísticos es la misma que la de X^2 y G^2 . Así, se pueden encontrar propiedades y comparaciones de los llamados test χ^2 (es decir, aquellos test que, bajo ciertas condiciones, siguen una distribución asintótica χ^2) en Watson (1959), Lancaster (1969), Horn (1977), Fienberg (1979, 1984) y una excelente recopilación en Moore (1986).

Cressie y Read (1984) introducen una familia de estadísticos de bondad de ajuste que contiene a todos los estadísticos clásicos citados anteriormente. Esta familia proporciona por una parte una vía rápida de comparación de los estadísticos clásicos y por otra permite descubrir nuevos estadísticos como alternativa a los clásicos.

Dicha familia se define como

$$2nI^\lambda(X/n, \hat{\pi}) = \frac{2}{\lambda(\lambda+1)} \sum_{i=1}^M X_i \left[\left(\frac{X_i}{n\hat{\pi}_i} \right)^\lambda - 1 \right], \quad -\infty < \lambda < \infty \quad (3.1.2)$$

y mide la divergencia entre X/n y $\hat{\pi}$, donde λ es el parámetro de la familia.

La ecuación (3.1.2) no está definida para $\lambda=-1$ ó $\lambda=0$. Sin embargo, si se definen estos casos por los límites continuos de (3.1.2) cuando $\lambda \rightarrow -1$ y $\lambda \rightarrow 0$, entonces $2nI^\lambda(X/n, \hat{\pi})$ es continua en λ . Además, utilizando el hecho de que $\ln(t) = \lim_{h \rightarrow 0} \frac{t^h - 1}{h}$, se obtiene que

$$2nI^0(X/n, \hat{\pi}) = \lim_{\lambda \rightarrow 0} 2nI^\lambda(X/n, \hat{\pi}) = 2 \sum_{i=1}^M X_i \ln \frac{X_i}{n\hat{\pi}_i} = G^2$$

y

$$2nI^{-1}(X/n, \hat{\pi}) = \lim_{\lambda \rightarrow -1} 2nI^\lambda(X/n, \hat{\pi}) = 2n \sum_{i=1}^M \hat{\pi}_i \ln \frac{n\hat{\pi}_i}{X_i} = GM^2.$$

Además es inmediato que

$$2nI^1(X/n, \hat{\pi}) = X^2,$$

$$2nI^{-1/2}(X/n, \hat{\pi}) = F^2$$

y

$$2nI^{-2}(X/n, \hat{\pi}) = NM^2.$$

Por tanto, todos los estadísticos clásicos considerados

anteriormente se pueden expresar como miembros de la familia (3.1.2).

En el extenso estudio de la familia de estadísticos (3.1.2) realizado por Cressie y Read (1984) se demuestra que bajo ciertas condiciones de regularidad cada miembro de esta familia sigue la misma distribución asintótica (una $\chi_{M-M_0-1}^2$). Además de las comparaciones de los diferentes miembros de la misma aparece el estadístico correspondiente a $\lambda=2/3$ como una excelente alternativa frente a los dos estadísticos más conocidos de bondad de ajuste, $2nI^0$ y $2nI^1$.

Obsérvese que la familia de estadísticos de Cressie y Read y como consecuencia los estadísticos obtenidos como caso particular de ella, se basan en medidas de divergencia entre la distribución teórica de la hipótesis nula y la estimada a partir de la muestra. Es más, no se debe olvidar que las medidas de divergencia de Cressie y Read son un caso particular de la f-divergencia de Csiszár que viene dada, como ya se indicó en el capítulo primero, por

$$D_f(P,Q) = \sum_{i=1}^M q_i f(p_i/q_i),$$

para cualquier función convexa $f:[0,\infty) \rightarrow \mathbb{R} \cup \{\infty\}$ donde $0f(0/0)=0$ y $0f(p/0) = \lim_{u \rightarrow \infty} \frac{f(u)}{u}$.

Morales y otros (1995) establecieron que el estadístico

$$2n \frac{D_f(\hat{\pi}, X/n) - f(1)}{f''(1)} \xrightarrow[n \rightarrow \infty]{L} \chi_{M-M_0-1}^2$$

en el caso de que la hipótesis nula sea compuesta y el parámetro se estime por mínima f-divergencia. Obsérvese que el estimador de mínima f-divergencia para $f(x) = x \ln x$ coincide con el de máxima verosimilitud para el modelo discretizado.

Una vez vista la utilización de las medidas f-divergencia en la construcción de contrastes de bondad de ajuste una pregunta parece obvia ¿los resultados obtenidos en bondad de ajuste a partir de las medidas de divergencia basadas en la f-divergencia de Csiszár se podrán obtener a partir de las R_ϕ -divergencias?. No se debe olvidar que en el capítulo anterior se dio una contestación afirmativa en el caso de estimación puntual.

Así pues en este capítulo se estudia una familia de estadísticos para contrastar bondad de ajuste basada en la divergencia de Burbea-Rao

$$R_\phi(X/n, \pi_0) = \sum_{i=1}^M \phi\left(\frac{X_i/n + \pi_{0i}}{2}\right) - \frac{1}{2} \left\{ \sum_{i=1}^M \phi(X_i/n) + \sum_{i=1}^M \phi(\pi_{0i}) \right\}.$$

3.2.- R_ϕ -divergencia y bondad de ajuste: Hipótesis nula simple.

Eficiencia Pitman

Pardo y otros (1993) demostraron que para una función $\phi:(0,\infty) \rightarrow \mathbb{R}$ cóncava con derivada segunda continua y bajo la hipótesis nula simple

$$8nR_\phi(X/n, \pi_0) \xrightarrow[n \rightarrow \infty]{L} \sum_{i=1}^M \beta_i \chi_1^2$$

donde las χ_1^2 son independientes y los β_i son los autovalores de $D(\pi_0)\Sigma_{\pi_0}$, siendo

$$\Sigma_{\pi_0} = \text{diag}(\pi_0) - \pi_0 \pi_0^t$$

y

$$D(\pi_0) = \text{diag}(-\phi''(\pi_0)).$$

En consecuencia el correspondiente test de bondad de ajuste rechazaría la hipótesis nula a nivel γ si

$$8nR_\phi(X/n, \pi_0) > t_\gamma$$

donde t_γ es tal que $P\left[\sum_{i=1}^M \beta_i \chi_i^2 > t_\gamma\right] = \gamma$.

Uno de los problemas que se presentan al utilizar estos contrastes es el de como elegir las clases cuando se discretiza el modelo, es decir cuál debe ser el tamaño de cada clase. En esta memoria se van a seleccionar clases con igual probabilidad para realizar contrastes de bondad de ajuste, lo que nos permitirá reducir estos contrastes a considerar la hipótesis nula

$$H_0: \pi_i = \pi_{0i} = 1/M, \quad \forall i=1, \dots, M. \quad (3.2.1)$$

Son varias las razones que justifican esta elección. Por una parte, Cohen y Sackrowitz (1975) demostraron que los contrastes que llevan a rechazar la hipótesis (3.2.1) si

$$\sum_{i=1}^M h_i(x_i) > c$$

donde c es una constante positiva, h_i , $i=1, \dots, M$, son funciones convexas y $x_i \geq 0$, $i=1, \dots, M$, son insesgados. Si en nuestro caso se elige ϕ de forma que R_ϕ sea convexa (Teorema 1.3.1), característica deseable que se señaló en el capítulo I, los contrastes propuestos son insesgados cuando las clases son equiprobables. Por otra parte, Bednarski y Ledwina (1978) afirman que $\forall n, \forall h: \Delta_M \times \Delta_M \longrightarrow \mathbb{R}^+$ continua y reflexiva y $\forall 0 < c < \sup\{c / P(h(p, x) \geq c) < 1, p \in \Delta_M\}$ existe $q \in \Delta_M$ tal que el contraste de región crítica $h(q, x) > c$ es sesgado para contrastar $H_0: p=q$. El estadístico en que se basa el contraste propuesto es una función continua en $\Delta_M \times \Delta_M - \{(0,0)\}$ y por ello cuando

las clases no son equiprobables no se puede decir que los contrastes considerados sean insesgados en general. Por último, como se ve en el Teorema 3.2.1, independientemente de la función ϕ elegida la distribución asintótica bajo la hipótesis (3.2.1) es una ji-cuadrado.

Teorema 3.2.1

Sea $\phi: (0, \infty) \rightarrow \mathbb{R}$ una función cóncava con derivada segunda continua y $\phi''(1/M) < 0$. Bajo la hipótesis (3.2.1)

$$S_{\phi}(X/n, \pi_0) = - \frac{M}{\phi''\left(\frac{1}{M}\right)} 8nR_{\phi}(X/n, \pi_0) \xrightarrow[n \rightarrow \infty]{L} \chi_{M-1}^2.$$

Demostración

Si la hipótesis (3.2.1) es cierta, la matriz $D(\pi_0)\Sigma_{\pi_0}$ con

$$\Sigma_{\pi_0} = \text{diag}(\pi_0) - \pi_0 \pi_0^t$$

y

$$D(\pi_0) = \text{diag}(-\phi''(\pi_0))$$

del resultado enunciado anteriormente, queda de la forma

$$-\phi''\left(\frac{1}{M}\right) A/M \tag{3.2.2}$$

donde

$$A = \begin{bmatrix} 1 - \frac{1}{M} & -\frac{1}{M} & \cdots & -\frac{1}{M} \\ -\frac{1}{M} & 1 - \frac{1}{M} & \cdots & -\frac{1}{M} \\ \cdots & \cdots & \cdots & \cdots \\ -\frac{1}{M} & -\frac{1}{M} & \cdots & 1 - \frac{1}{M} \end{bmatrix}.$$

Ahora bien, es inmediato que los autovalores de la matriz A son el 0 con multiplicidad 1 y el 1 con multiplicidad M-1. Por tanto, los autovalores de la matriz (3.2.2) son el 0 con multiplicidad 1 y el $-\phi''(1/M)/M$ con multiplicidad M-1.

Como consecuencia,

$$8nR_{\phi}(X/n, \pi_0) \xrightarrow[n \rightarrow \infty]{L} -\phi''\left(\frac{1}{M}\right) \frac{1}{M} \chi_{M-1}^2$$

y se tiene el resultado enunciado. ■

Por el teorema anterior se tiene que

$$P(S_{\phi}(X/n, \pi_0) > \chi_{M-1, \gamma}^2 | H_0) \xrightarrow[n \rightarrow \infty]{} \gamma$$

donde $\chi_{M-1, \gamma}^2$ es tal que $P(\chi_{M-1}^2 > \chi_{M-1, \gamma}^2) = \gamma$.

Por tanto, para tamaño muestral n grande y número de clases fijo, el correspondiente test de bondad de ajuste llevaría a rechazar la hipótesis nula a nivel γ si

$$S_{\phi}(X/n, \pi_0) > \chi_{M-1, \gamma}^2$$

Obsevación 3.2.1

Tanto en este capítulo como en el siguiente la familia de divergencias $\phi = \phi_{\alpha}$ introducida en el capítulo anterior jugará un papel importante. Obsérvese que bajo la hipótesis nula (3.2.1) el estadístico S_{ϕ_2} coincide con el de la X^2 de Pearson lo que permitirá comparar éste con otros estadísticos no conocidos. El resultado de

estos estudios llevará a la obtención de un excelente competidor de la X^2 .

A continuación se ilustra este contraste con un ejemplo.

Ejemplo 3.2.1

Supóngase que se quiere encontrar un modelo para predecir la probabilidad de ganar en una carrera de galgos en Australia (Read y Cowan (1976)). Se tiene recogida información relativa a la posición ganadora de los 8 galgos participantes en 595 carreras y se supone que los galgos se numeran aleatoriamente al comienzo de cada carrera. El modelo más simple que se puede considerar es aquel con los $8!=40320$ posibles resultados de la carrera igualmente probables, pero en nuestro caso este modelo no es apropiado al tener únicamente 595 observaciones. Por ello, se clasifican las observaciones según la variable aleatoria $X=(X_1, \dots, X_8)$ multinomial donde X_i representa el número de veces que ha ganado el galgo i y se considera el modelo que asigna a todos los galgos la misma probabilidad de ganar. Para comprobar como se adapta este modelo a los datos que tenemos, se realiza el siguiente contraste

$$H_0: \pi_i = 1/8 \quad i=1, \dots, 8$$

donde $\pi = P(\text{galgo número } i \text{ gane})$.

En la Tabla 3.2.1 aparecen las frecuencias observadas y esperadas de que gane cada uno de los galgos.

Para contrastar H_0 se calcula los valores de algunos estadísticos de la familia S_{ϕ_α} con $\phi_\alpha(x) = \frac{1}{1-\alpha}(x^\alpha - x)$ que vienen dados en la Tabla 3.2.2. La elección de los estadísticos considerados quedará totalmente justificada en el capítulo IV dedicado a optimalidad.

Tabla 3.2.1

galgo i	Observada	Esperada
1	0.175	0.125
2	0.16	0.125
3	0.111	0.125
4	0.106	0.125
5	0.104	0.125
6	0.097	0.125
7	0.101	0.125
8	0.146	0.125

Tabla 3.2.2

α	1	13/7	2
S_{ϕ}	29.1768	30.5175	30.788

Utilizando el resultado obtenido en el Teorema 3.2.1, se rechaza H_0 a nivel 0.05 si el valor del estadístico considerado es mayor que el punto crítico $\chi^2_{7,0.05}=14.07$. Ya que todos los valores de los estadísticos calculados son bastante más grandes que este valor, se concluye que es muy poco probable que todos los galgos participantes tengan la misma probabilidad de ganar. ■

Hasta ahora se ha calculado la distribución asintótica de S_{ϕ} bajo la hipótesis (3.2.1). A continuación se calcula la distribución asintótica de S_{ϕ} bajo diferentes hipótesis alternativas con el objetivo de estudiar la eficiencia de estos estadísticos.

La función de potencia de la familia de estadísticos $S_\phi(X/n, \pi_0)$ que representa la probabilidad de rechazar el modelo propuesto en la hipótesis nula H_0 cuando el modelo alternativo es verdadero, es un buen criterio para comparar los estadísticos de la familia, siendo el mejor estadístico el que tiene mayor potencia.

En el caso que las hipótesis alternativas sean fijas la función de potencia de S_ϕ converge a 1 cuando $n \rightarrow \infty$. Sin embargo, es posible que el vector de probabilidad de la alternativa, no sea fijo sino que converja al vector de la hipótesis nula H_0 cuando $n \rightarrow \infty$ de forma que la potencia límite sea menor que 1 (y mayor que el tamaño del test γ). Este límite se denomina eficiencia asintótica del test.

En este sentido, Cochran (1952) propone utilizar las alternativas

$$H_{1,n} : \pi = \pi_0 + n^{-1/2}c, \quad (3.2.3)$$

donde $c = (c_1, \dots, c_M)^t$ satisface $\sum_{i=1}^M c_i = 0$, las cuales convergen a H_0 .

La función de potencia de la familia $S_\phi(X/n, \pi_0)$ cuando se consideran estas alternativas viene dada por

$$\beta_\phi^{(n)}(\pi_0 + n^{-1/2}c) = P(S_\phi(X/n, \pi_0) > \chi_{M-1, \gamma}^2 | H_{1,n}).$$

En este contexto, la eficiencia relativa asintótica Pitman para comparar dos estadísticos S_{ϕ_1} y S_{ϕ_2} se define como el cociente e_{ϕ_1}/e_{ϕ_2} , donde

$$e_\phi = \lim_{n \rightarrow \infty} \beta_\phi^{(n)}(\pi_0 + n^{-1/2}c). \quad (3.2.4)$$

Para calcular e_ϕ , necesitamos conocer la distribución asintótica

de S_ϕ bajo $H_{1,n}$.

Teorema 3.2.2

Sea $\phi:(0,\infty) \rightarrow \mathbb{R}$ una función cóncava con derivada segunda continua y $\phi''(1/M) < 0$. Bajo las hipótesis (3.2.3)

$$S_\phi(X/n, \pi_0) = - \frac{M}{\phi''\left[\frac{1}{M}\right]} \delta n R_\phi(X/n, \pi_0)$$

se distribuye asintóticamente como una ji-cuadrado no centrada con $M-1$ grados de libertad y parámetro de descentralización $\delta = M \sum_{i=1}^M c_i^2$.

Demostración

El desarrollo de Taylor de $R_\phi(X/n, \pi_0)$ en torno al punto $\pi_0 = (1/M, \dots, 1/M)$ viene dado por

$$R_\phi(X/n, \pi_0) = \frac{1}{8} (X/n - \pi_0)' D(\pi_0) (X/n - \pi_0) + R_n,$$

donde

$$D(\pi_0) = \text{diag} \left[-\phi''(1/M) \right].$$

Ya que $nR_n \xrightarrow[n \rightarrow \infty]{P} 0$, se tiene que

$$- \frac{M}{\phi''\left[\frac{1}{M}\right]} \delta n R_\phi(X/n, \pi_0) \quad \text{y} \quad n(X/n - \pi_0)' B(X/n - \pi_0)$$

tienen la misma distribución asintótica con $B = MI$ e I la matriz identidad de dimensión $M \times M$. Ahora bien, bajo las hipótesis (3.2.3)

$$n^{1/2} (X/n - \pi_0) \xrightarrow[n \rightarrow \infty]{L} N(c, \Sigma),$$

donde

$$\Sigma = \frac{1}{M} \left[I - \frac{1}{M} (1)_{ij=1,\dots,M} \right].$$

Para completar la demostración de este teorema se tiene que demostrar que $n(X/n-\pi_0)'B(X/n-\pi_0)$ se distribuye como un ji-cuadrado, para ello basta con comprobar que se verifican las condiciones necesarias y suficientes del Teorema 3.1 de Dik y Gunst (1985),

$$\Sigma B \Sigma B \Sigma = \Sigma B \Sigma \quad (3.2.5)$$

$$c' B \Sigma B c = c' B c \quad (3.2.6)$$

$$\Sigma B \Sigma B c = \Sigma B c \quad (3.2.7)$$

siendo en este caso el número de grados de libertad la traza de $B \Sigma$ y el parámetro de descentralización $c' B c$. Obsérvese que

$$B \Sigma = M I \frac{1}{M} \left[I - \frac{1}{M} (1)_{ij=1,\dots,M} \right] = I - \frac{1}{M} (1)_{ij=1,\dots,M} = \Sigma B,$$

$$\begin{aligned} (\Sigma B)(\Sigma B) &= \left[I - \frac{1}{M} (1)_{ij=1,\dots,M} \right] \left[I - \frac{1}{M} (1)_{ij=1,\dots,M} \right] = \\ &= I - 2 \frac{1}{M} (1)_{ij=1,\dots,M} + \frac{1}{M^2} (M)_{ij=1,\dots,M} = \\ &= \left[I - \frac{1}{M} (1)_{ij=1,\dots,M} \right] = \Sigma B, \end{aligned}$$

$$\begin{aligned} (B \Sigma) B &= \left[I - \frac{1}{M} (1)_{ij=1,\dots,M} \right] M I = M I - (1)_{ij=1,\dots,M} = \\ &= B - (1)_{ij=1,\dots,M} \end{aligned}$$

y

$$c' B \Sigma B c = c' B c - c' (1)_{i,j=1,\dots,M} c$$

con

$$c^t (1)_{i,j=1,\dots,M} c = \sum_{i,j=1}^M c_i c_j = \sum_{i=1}^M c_i \sum_{j=1}^M c_j = 0.$$

Por tanto queda probado que se verifican (3.2.5), (3.2.6) y (3.2.7) y que $\text{traza}(B\Sigma) = M-1$ como se quería demostrar. ■

Este resultado indica que la familia S_ϕ no solo es equivalente bajo la hipótesis nula H_0 sino también bajo las alternativas $H_{1,n}$. Por tanto la ecuación (3.2.4) es independiente de ϕ . Es decir,

$$e_\phi = P(\chi_{M-1}^2(\delta) > \chi_{M-1, \gamma}^2)$$

y la eficiencia relativa asintótica Pitman para dos miembros cualesquiera de la familia S_ϕ es 1. Esto implica que todos los miembros de la familia de estadísticos S_ϕ son equivalentes en el sentido Pitman cuando la hipótesis nula es simétrica.

Observación 3.2.2

Una suposición básica en el desarrollo de los resultados de este apartado ha sido la independencia de las observaciones de la muestra aleatoria utilizada en el cálculo de los estadísticos R_ϕ -divergencia para realizar los contrastes de bondad de ajuste (3.2.1). A continuación se estudian, de forma breve estos estadísticos para realizar contrastes acerca de la distribución estacionaria de cadenas de Markov.

Se considera una cadena de Markov aperiódica irreducible $X=(X_0, X_1, \dots)$ con espacio de estados $\{1, \dots, m\}$. Sea $P=(p_{ij})_{i,j=1}^m$ la matriz de transición de esta cadena y $p=(p_1, \dots, p_m)$ la distribución

estacionaria, e.d., solución de la ecuación $p=pP$.

Los estadísticos que se proponen para contrastar la hipótesis $H_0: p_i = \pi_{0i} = 1/m, i=1, \dots, m$, vienen dados por

$$S_\phi(\hat{p}_n, \pi_0) = - \frac{M}{\phi''(1/m)} 8nR_\phi(\hat{p}_n, \pi_0)$$

donde \hat{p}_{ni} es la frecuencia relativa del estado i en n observaciones de los estados de la cadena (X_1, \dots, X_n) .

Por una parte, si se desarrolla por Taylor $R_\phi(\hat{p}_n, \pi)$ en torno a π_0 , se tiene que

$$R_\phi(\hat{p}_n, \pi_0) = - \frac{\phi''(1/m)}{8} (\hat{p}_n - \pi_0)' (\hat{p}_n - \pi_0)$$

o equivalentemente

$$S_\phi(\hat{p}_n, \pi_0) = n \sum_{i=1}^m \frac{(\hat{p}_{ni} - 1/m)^2}{1/m} + o_p(1).$$

Por otra parte, Tavaré y Altham (1983) demostraron que si

$$\sqrt{n} \left[\frac{\hat{p}_{n1} - \pi_{01}}{\sqrt{\pi_{01}}}, \dots, \frac{\hat{p}_{nm} - \pi_{0m}}{\sqrt{\pi_{0m}}} \right] \xrightarrow{L} N(0, V)$$

entonces

$$X_n^2 = n \sum_{i=1}^m \frac{(\hat{p}_{ni} - 1/m)^2}{1/m} \xrightarrow{L} \sum_{i=1}^m \rho_i Z_i^2$$

donde ρ_i son los autovalores de la matriz V y Z_i son $N(0,1)$ independientes.

Esto nos lleva a conocer la distribución asintótica de S_ϕ bajo la

hipótesis nula simétrica y por tanto, el correspondiente contraste de bondad de ajuste rechazaría la hipótesis nula a nivel γ si

$$S_{\phi}(\hat{p}_n, \pi_0) > t_{\gamma}$$

donde t_{γ} es tal que $P\left[\sum_{i=1}^m \rho_i Z_i^2 > t_{\gamma}\right] = \gamma$.

Además, en el caso de que la cadena sea reversible

$$\rho_i = \frac{1 + \lambda_i}{1 - \lambda_i} \quad \forall i=1, \dots, m-1, \quad \rho_m = 0$$

donde $\lambda_1, \dots, \lambda_{m-1}$ son los autovalores no unitarios de P . En general no se pueden obtener éstos porque la matriz P no es conocida y entonces se tiene que utilizar las frecuencias relativas

$$\hat{P}_{nij} = \frac{\sum_{k=2}^n I_{(i,j)}(X_{k-1}, X_k)}{\sum_{k=2}^n I_{(i)}(X_{k-1})}$$

como estimadores consistentes de los elementos p_{ij} de la matriz P (Billingsley (1961)) para obtener estimadores consistentes $\hat{\lambda}_{ni}$ de λ_i .

Para una cadena de Markov binaria, $m=2$, la clase de matrices de transición posibles vienen dadas por

$$\begin{bmatrix} 1-\beta & \beta \\ \zeta & 1-\zeta \end{bmatrix}, \quad 0 < \beta, \zeta \leq 1 \quad \beta + \zeta < 2$$

y la subclase satisfaciendo la condición de estacionariedad es de la forma

$$\begin{bmatrix} 1-\beta & \beta \\ \theta\beta & 1 - \frac{\theta\beta}{1-\theta} \end{bmatrix}, \quad 0 < \beta \leq \min\left\{1, \frac{1-\theta}{\theta}\right\} \quad \beta < 1.$$

En particular, bajo la hipótesis nula es

$$\begin{bmatrix} 1-\beta & \beta \\ \beta & 1-\beta \end{bmatrix}, 0 < \beta < 1$$

y sus autovalores son 1 y $1 - 2\beta$.

Al ser el estimador de β bajo la hipótesis nula

$$\hat{\beta} = \frac{(1 - \hat{p}_{n11}) + (1 - \hat{p}_{n22})}{2}$$

la región crítica del contraste queda de la forma

$$S_{\phi}(\hat{p}_n, \pi_0) > \frac{\hat{p}_{n11} + \hat{p}_{n22}}{2 - \hat{p}_{n11} - \hat{p}_{n22}} \chi_{1,1-\gamma}^2$$

A continuación, se comparan estos contrastes para $\phi = \phi_{\alpha}$ mediante la función de potencia que viene dada por la probabilidad de la región crítica bajo la hipótesis alternativa $p = (\theta, 1-\theta)$, $0 < \theta \leq 1/(\beta+1)$, para $n=20$ y 50 , $\alpha=.05$ y $\beta \in \{1/4, 1/2, 3/4\}$.

A la vista de las Figuras 3.2.1, 3.2.2, 3.2.3 y 3.2.4 no es clara la elección del mejor estadístico entre los considerados. No obstante, el estadístico S_{ϕ_1} parece ser una buena opción ya que en las figuras en las que no se solapan las gráficas, éste mantiene un buen equilibrio entre no alejarse demasiado del $\alpha=.05$ deseado y tener una potencia alta. Obsérvese que el comportamiento de S_{ϕ_2} (estadístico X^2 de Pearson) y $S_{\phi_{13/7}}$ es parejo.

Figura 3.2.1. Potencia del estadístico S_{ϕ_α} para $\beta=1/2$ y $n=50$.

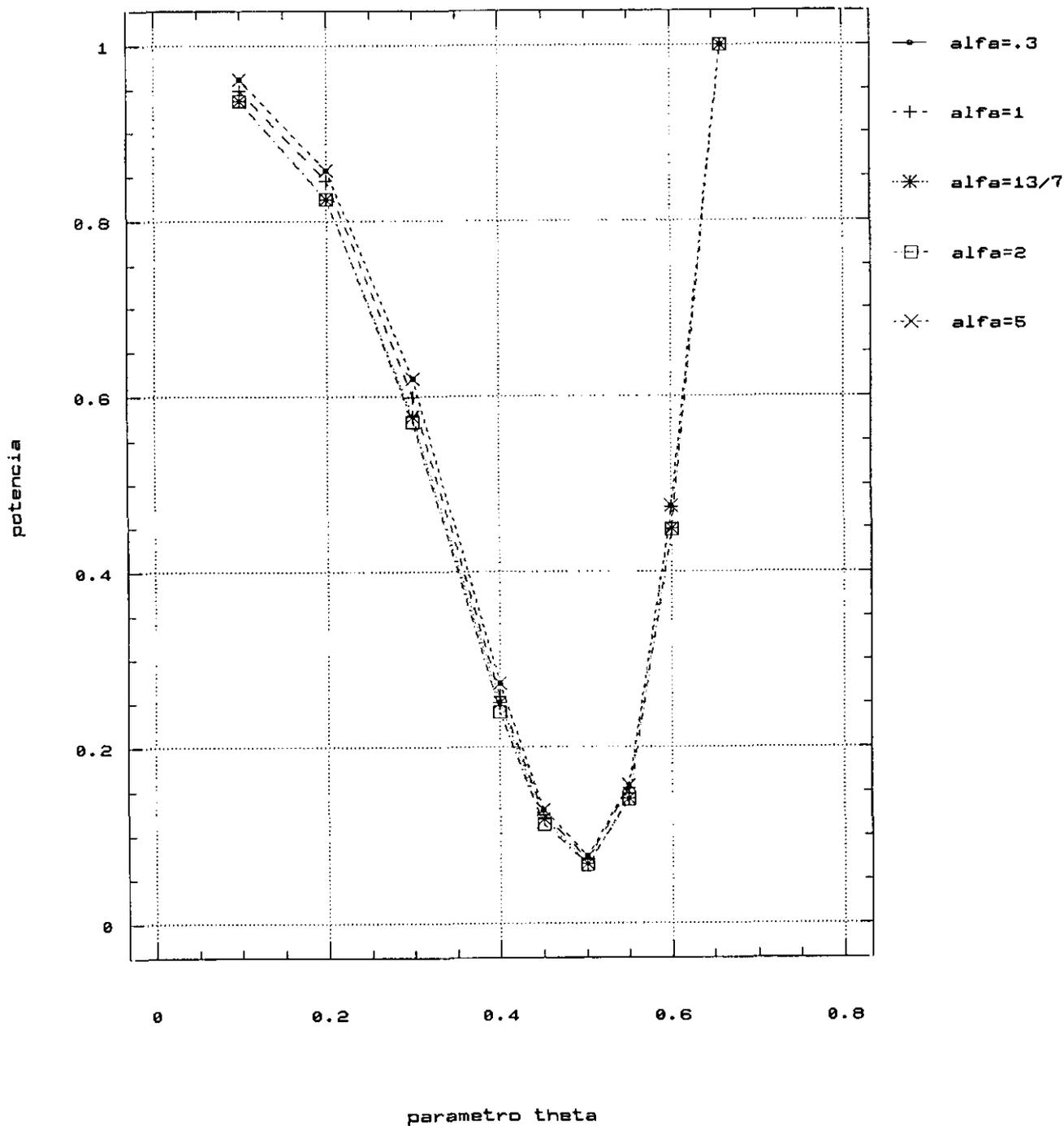


Figura 3.2.2. Potencia del estadístico S_{ϕ_α} para $\beta=1/2$ y $n=20$.

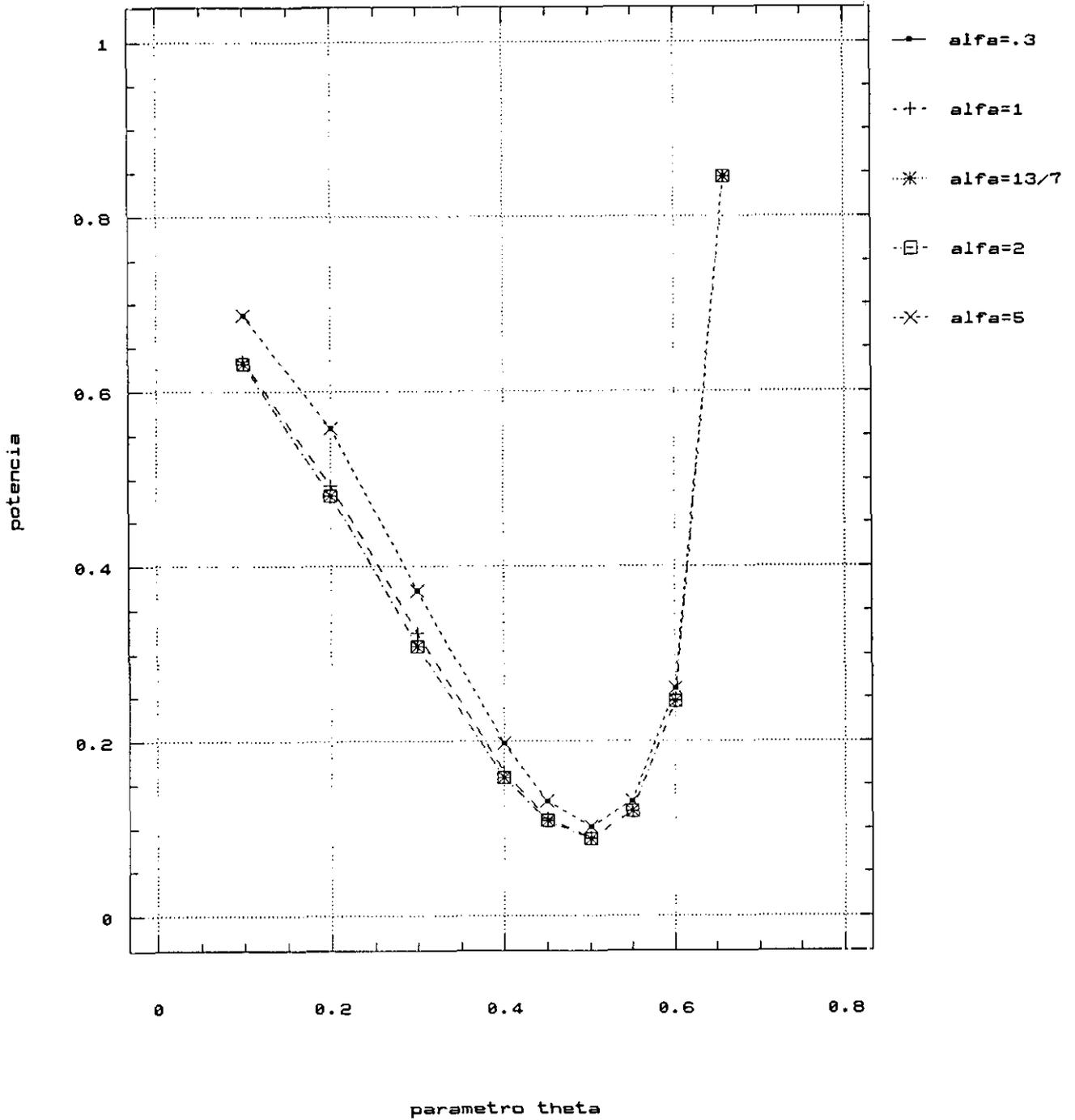


Figura 3.2.3. Potencia del estadístico S_{ϕ_α} para $\beta=3/4$ y $n=20$.

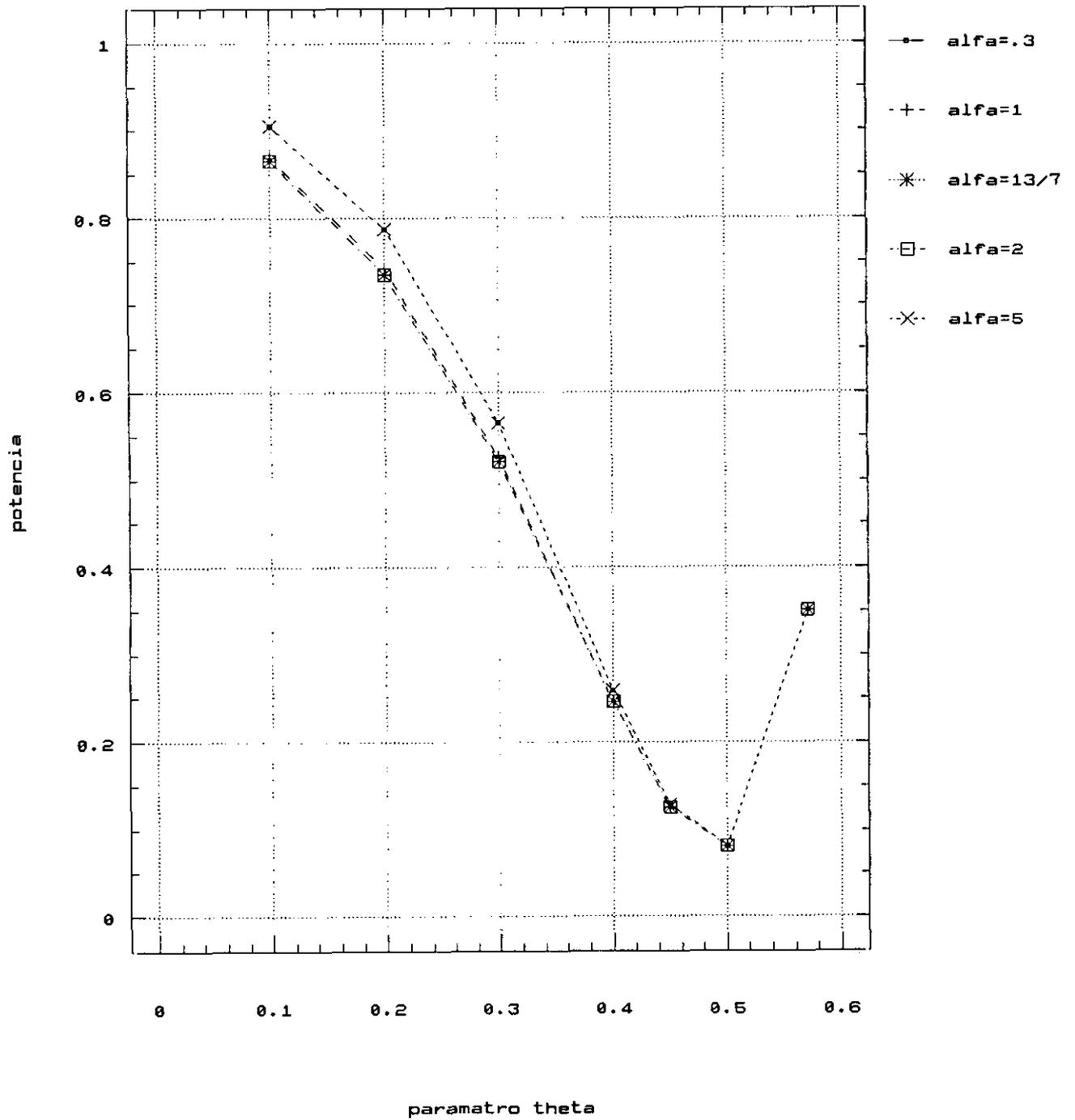
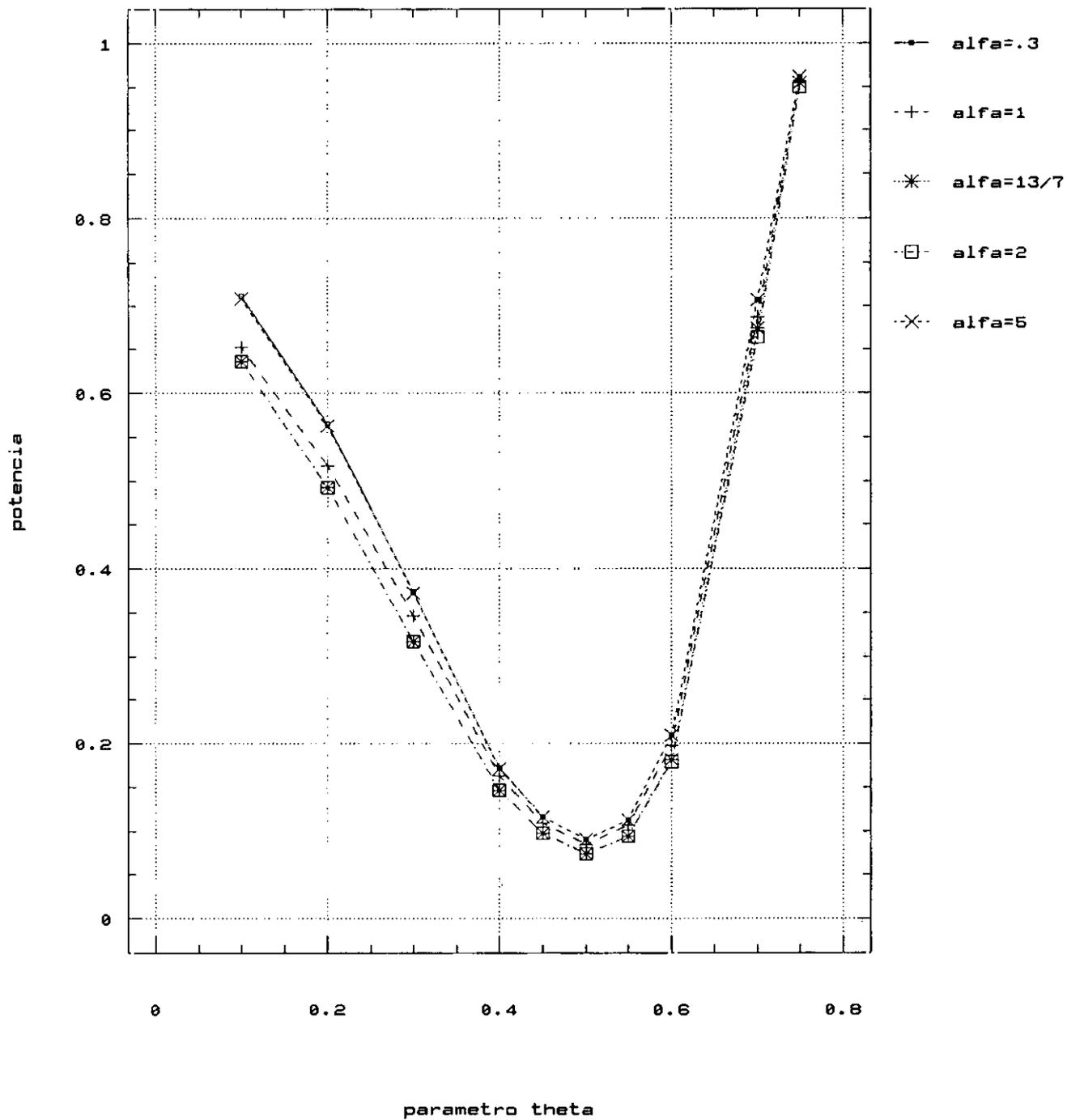


Figura 3.2.4. Potencia del estadístico S_{ϕ_α} para $\beta=1/4$ y $n=50$.



3.3.- R_ϕ -divergencia y bondad de ajuste: Hipótesis nula compuesta

Considérese el caso de hipótesis nula compuesta

$$H_0: \pi = \pi_0 \quad (3.3.1)$$

donde $\pi_0 = Q(\theta) = (q_1(\theta), \dots, q_M(\theta))^t \in T \subset \Delta_M$ y $\theta = (\theta_1, \dots, \theta_{M_0})^t \in \Theta \subseteq R^{M_0}$ el vector de parámetros desconocidos.

Para realizar este contraste de bondad de ajuste se deben estimar los parámetros desconocidos, es decir, elegir un valor $Q(\hat{\theta}) \in T$ que sea lo más consistente posible con las frecuencias observadas X/n . El método más conocido para elegir $Q(\hat{\theta})$ consiste en estimar θ por máxima verosimilitud, pero parece también una opción razonable para estimar π_0 elegir el $Q(\hat{\theta}) \in T$ más próximo a X/n con respecto a la medida $R_\phi(\hat{P}, Q(\theta))$. Es decir, considerar $Q(\hat{\theta}_\phi)$ donde $\hat{\theta}_\phi$ es el estimador de mínima R_ϕ -divergencia dado en la Definición 2.2.1 y estudiado en el capítulo anterior.

En segundo lugar, se necesitará conocer la distribución asintótica de $R_\phi(\hat{P}, Q(\hat{\theta}))$ cuando H_0 es verdadera, donde $\hat{P} = (\hat{p}_1, \dots, \hat{p}_M)^t$ es el vector de frecuencias relativas y $Q(\hat{\theta}) = (q_1(\hat{\theta}), \dots, q_M(\hat{\theta}))^t$ siendo $\hat{\theta}$ el estimador de máxima verosimilitud o el estimador de mínima R_ϕ -divergencia.

Antes de calcular la distribución asintótica de $R_\phi(\hat{P}, Q(\hat{\theta}))$, se establece el siguiente resultado más general.

Lema 3.3.1

Sean $\hat{P} = (\hat{p}_1, \dots, \hat{p}_M)^t$ y $\hat{Q} = (\hat{q}_1, \dots, \hat{q}_M)^t$ estimadores c_n -consistentes de la distribución $\pi = Q(\theta^0)$ para algún $c_n \uparrow \infty$. Si

$Q(\theta^0)$ satisface la condición (2) de Birch y ϕ es cóncava y dos veces continuamente diferenciable en $(0, \infty)$

$$c_n^2 R_\phi(\hat{P}, \hat{Q}) \approx \frac{1}{8} c_n^2 (\hat{P} - \hat{Q})^t D(\theta^0) (\hat{P} - \hat{Q})$$

donde

$$D(\theta^0) = \text{diag}(-\phi''(Q(\theta^0))).$$

Demostración

Considérese el vector aleatorio $\hat{W} = (\hat{w}_1, \dots, \hat{w}_{2M})^t = (\hat{P}, \hat{Q})$ y $W = (w_1, \dots, w_{2M})^t = (P, Q)$ donde $P = (p_1, \dots, p_M)^t$ y $Q = (q_1, \dots, q_M)^t$ son distribuciones de probabilidad en \mathcal{X} y se define

$$\psi(W) = R_\phi(P, Q) \quad y \quad W^0 = (Q(\theta^0), Q(\theta^0)).$$

Por el Teorema del Valor Medio

$$\psi(\hat{W}) = \psi(W^0) + (\hat{W} - W^0)^t a(W^0) + \frac{1}{2} (\hat{W} - W^0)^t K(W^*) (\hat{W} - W^0)$$

donde el vector $a(W) = (a_j(W))_{j=1, \dots, 2M}$ viene dado por

$$a_j(W) = \frac{\partial \psi(W)}{\partial w_j},$$

la matriz $K(W) = (k_{jr}(W))_{j,r=1, \dots, 2M}$ viene dada por

$$k_{jr}(W) = \frac{\partial^2 \psi(W)}{\partial w_j \partial w_r}$$

y W^* es un vector que verifica

$$\|W^* - W^0\| \leq \|\hat{W} - W^0\|. \quad (3.3.2)$$

Además, la continuidad de ϕ'' implica que todas las funciones $k_j(W)$ son continuas en W . De esta forma la consistencia de \hat{P} y \hat{Q} junto con (3.3.2) implica que la matriz $K(W^*)$ tiende elemento a elemento a $K(W^0)$ en probabilidad.

Al ser,

$$K(W^0) = \frac{1}{4}K$$

con

$$K = \begin{bmatrix} D(\theta^0) & -D(\theta^0) \\ -D(\theta^0) & D(\theta^0) \end{bmatrix},$$

se tiene

$$\begin{aligned} (\hat{W} - W^0)' K (\hat{W} - W^0) &= (\hat{P} - Q(\theta^0))' D(\theta^0) (\hat{P} - Q(\theta^0)) - 2(\hat{P} - Q(\theta^0))' D(\theta^0) (\hat{Q} - Q(\theta^0)) + \\ &+ (\hat{Q} - Q(\theta^0))' D(\theta^0) (\hat{Q} - Q(\theta^0)) = (\hat{P} - \hat{Q})' D(\theta^0) (\hat{P} - \hat{Q}). \end{aligned}$$

Por último teniendo en cuenta que

$$\psi(W^0) = 0 \quad \text{y} \quad a(W^0) = (0)_{i=1, \dots, M}$$

se obtiene que para cualquier variable aleatoria X y cualquier sucesión c_n

$$\begin{aligned} |c_n^2 R_{\phi}(\hat{P}, \hat{Q}) - X| &\leq \left| \frac{1}{2} c_n^2 (\hat{W} - W^0)' K(W^0) (\hat{W} - W^0) - X \right| + \\ &+ \left| \frac{1}{2} c_n^2 (\hat{W} - W^0)' (K(W^*) - K(W^0)) (\hat{W} - W^0) \right|. \end{aligned}$$

Al ser, el primer sumando igual a

$$|\frac{1}{8}c_n^2 (\hat{P} - \hat{Q})'D(\theta^0)(\hat{P} - \hat{Q}) - X|,$$

bastará probar que el segundo sumando converge en probabilidad a 0.

Ahora bien, el segundo sumando está acotado superiormente por

$$\frac{(c_n \|\hat{P} - Q(\theta^0)\|)^2 + (c_n \|\hat{Q} - Q(\theta^0)\|)^2}{2} \max_{j,r} |k_{jr}(W^*) - k_{jr}(W^0)|$$

donde, por una parte, la c_n -consistencia de \hat{P} y \hat{Q} nos lleva a que

$$\frac{(c_n \|\hat{P} - Q(\theta^0)\|)^2 + (c_n \|\hat{Q} - Q(\theta^0)\|)^2}{2} \leq O_p(1)$$

y por otra parte, como los elementos de $K(W^*)$ convergen a los de $K(W^0)$ en probabilidad, el $\max_{j,r} |k_{jr}(W^*) - k_{jr}(W^0)|$ tiende a 0 en probabilidad.

De esta forma queda probado el resultado enunciado. ■

El siguiente teorema nos da la distribución asintótica de $R_\phi(\hat{P}, Q(\theta))$ bajo H_0 cuando el parámetro θ es estimado por mínima R_ϕ -divergencia.

Teorema 3.3.1

Sea $\phi^* : (0, \infty) \rightarrow \mathbb{R}$ una función cóncava con derivada segunda continua. Sea \hat{P} el vector de frecuencias relativas y $\hat{Q}_\phi = Q(\hat{\theta}_\phi)$, entonces bajo las condiciones de regularidad de Birch y las condiciones i) e ii) del Teorema 2.3.1 se verifica que

$$8n R_\phi^*(\hat{P}, \hat{Q}_\phi) \xrightarrow[n \rightarrow \infty]{L} \sum_{i=1}^M \beta_i \chi_1^2$$

donde las χ_1^2 son independientes y los β_i son los autovalores de la matriz $D(\theta^0)\Sigma_1$, siendo

$$D(\theta^0) = \text{diag}(-\phi''(Q(\theta^0)))$$

y

$$\Sigma_1 = (I - J(\theta^0)B(\theta^0))\Sigma_{Q(\theta^0)}(I - J(\theta^0)B(\theta^0))^t$$

con

$$B(\theta^0) = (A(\theta^0)^t A(\theta^0))^{-1} A(\theta^0)^t \text{diag} \left[\sqrt{-\phi''(Q(\theta^0))} \right]$$

$$J(\theta^0) = \left[\frac{\partial q_j(\theta^0)}{\partial \theta_r} \right]_{\substack{j=1,\dots,M \\ r=1,\dots,M_0}}$$

y

$$\Sigma_{Q(\theta^0)} = \text{diag}(Q(\theta^0)) - Q(\theta^0)Q(\theta^0)^t.$$

Demostración

Por el Lema 3.3.1, al ser \hat{P} y \hat{Q}_ϕ estimadores \sqrt{n} -consistentes, se tiene

$$\sqrt{n} R_\phi^*(\hat{P}, \hat{Q}_\phi) \approx n (\hat{P} - \hat{Q}_\phi)^t D(\theta^0) (\hat{P} - \hat{Q}_\phi).$$

Por el Teorema 2.3.2 a) y c) se sabe que

$$\sqrt{n} \left[\hat{Q}_\phi - Q(\theta^0) \right] \approx \sqrt{n} J(\theta^0) B(\theta^0) \left[\hat{P} - Q(\theta^0) \right],$$

por tanto

$$\begin{aligned}\sqrt{n}(\hat{P} - \hat{Q}_\phi) &= \sqrt{n}(\hat{P} - Q(\theta^0)) + \sqrt{n}(Q(\theta^0) - \hat{Q}_\phi) \approx \\ &\approx \sqrt{n}(I - J(\theta^0)B(\theta^0))(\hat{P} - Q(\theta^0)).\end{aligned}$$

Consecuentemente

$$\sqrt{n}(\hat{P} - \hat{Q}_\phi) \xrightarrow[n \rightarrow \infty]{L} N(0, \Sigma_1),$$

donde

$$\Sigma_1 = (I - J(\theta^0)B(\theta^0))\Sigma_{Q(\theta^0)}(I - J(\theta^0)B(\theta^0))'$$

por lo que $8n R_\phi^*(\hat{P}, \hat{Q}_\phi)$ se distribuye asintóticamente como $\sum_{i=1}^M \beta_i \chi_1^2$ donde las χ_1^2 son independientes y los β_i son los autovalores de la matriz $D(\theta^0)\Sigma_1$. ■

De esta forma para n grande y nivel de significación γ el contraste que se propone llevaría a rechazar la hipótesis nula si

$$8n R_\phi(\hat{P}, Q(\hat{\theta}_\phi)) > t_\gamma \quad (3.3.3)$$

donde t_γ es tal que $P\left[\sum_{i=1}^M \beta_i \chi_1^2 > t_\gamma\right] = \gamma$.

Para calcular la potencia asintótica del contraste (3.3.3) es necesario conocer la distribución asintótica de $R_\phi(\hat{P}, \hat{Q}_\phi)$ cuando el modelo paramétrico propuesto no es correcto.

Teorema 3.3.2

Sea $\phi:(0,\infty) \rightarrow \mathbb{R}$ una función cóncava con derivada continua. Supongamos que se verifican las condiciones de regularidad de Birch y las condiciones i) e ii) del Teorema 2.3.1. Sea \hat{P} el vector de frecuencias relativas y $\hat{Q}_\phi = Q(\hat{\theta}_\phi)$, entonces si $\pi \neq Q(\theta^0)$ se verifica que

$$\sqrt{n} \left[R_\phi(\hat{P}, \hat{Q}_\phi) - R_\phi(\pi, Q_\phi) \right] \xrightarrow[n \rightarrow \infty]{L} N(0, \sigma^2)$$

donde

$$\sigma^2 = T' \Sigma_\pi T, \quad \Sigma_\pi = (\pi_i (\delta_{ij} - \pi_j))_{i,j=1,\dots,M}, \quad T' = (t_1, \dots, t_M)$$

y

$$t_i = \frac{1}{2} \left\{ \phi' \left[\frac{\pi_i + q_i}{2} \right] + \sum_{k=1}^M \left[\phi' \left[\frac{\pi_k + q_k}{2} \right] - \phi'(q_k) \right] \frac{\partial q_k}{\partial \pi_i} - \phi'(\pi_i) \right\}.$$

Demostración

Consideremos el desarrollo de Taylor de $\psi(\hat{P})$ en torno al punto π donde $\psi(\pi) = R_\phi(\pi, Q(\theta_\phi(\pi)))$,

$$\psi(\hat{P}) = \psi(\pi) + (\hat{P} - \pi)' T + R_n$$

donde $T = (t_1, \dots, t_M)'$ con

$$\begin{aligned} t_i &= \frac{\partial \psi(\pi)}{\partial \pi_i} = \\ &= \frac{1}{2} \left\{ \phi' \left[\frac{\pi_i + q_i}{2} \right] + \sum_{k=1}^M \phi' \left[\frac{\pi_k + q_k}{2} \right] \frac{\partial q_k}{\partial \pi_i} - \phi'(\pi_i) - \sum_{k=1}^M \phi'(q_k) \frac{\partial q_k}{\partial \pi_i} \right\}. \end{aligned}$$

Como $\sqrt{n} R_n \xrightarrow[n \rightarrow \infty]{P} 0$, entonces

$$\sqrt{n} \left[R_{\phi}(\hat{P}, \hat{Q}_{\phi}) - R_{\phi}(\pi, Q_{\phi}) \right] \quad \text{y} \quad \sqrt{n} (\hat{P} - \pi)' T$$

tienen asintóticamente la misma distribución.

Por el Teorema Central del Límite

$$\sqrt{n} \left[R_{\phi}(\hat{P}, \hat{Q}_{\phi}) - R_{\phi}(\pi, Q_{\phi}) \right] \xrightarrow[n \rightarrow \infty]{L} N(0, \sigma^2)$$

donde

$$\sigma^2 = T' \Sigma_{\pi} T \quad \text{con} \quad \Sigma_{\pi} = (\pi_i (\delta_{ij} - \pi_j))_{i,j=1,\dots,M}$$

■

Por tanto, utilizando el Teorema 3.3.2 la potencia asintótica del contraste (3.3.3) viene dada por

$$\beta_{\phi}^{(n)}(\pi) = P(8n R_{\phi}(\hat{P}, \hat{Q}_{\phi}) > t_{\gamma}) = 1 - F_N \left[\frac{t_{\gamma} - 8n R_{\phi}(\pi, Q(\theta_{\phi}(\pi)))}{8\sigma n^{1/2}} \right]$$

donde σ^2 es la expresión de la varianza dada en el Teorema 3.3.2 y F_N es la función de distribución normal estándar.

Es inmediato comprobar que

$$\lim_{n \rightarrow \infty} \beta_{\phi}^{(n)}(\pi) = 1$$

es decir, el contraste (3.3.3) es consistente en el sentido de Fraser (1957).

Corolario 3.3.1

Sea \hat{P} el vector de frecuencias relativas y $\hat{Q}_1 = Q(\hat{\theta}_1)$, donde $\hat{\theta}_1$ es el estimador de mínima R-divergencia ($\phi(x) = -x \ln x$), entonces bajo

las condiciones de regularidad de Birch y las condiciones i) e ii) del Teorema 2.3.1 se verifica que

$$8n R(\hat{P}, \hat{Q}_1) \xrightarrow[n \rightarrow \infty]{L} \chi_{M-M_0-1}^2.$$

Demostración

Por el Teorema 3.3.1,

$$8n R(\hat{P}, \hat{Q}_1) \xrightarrow[n \rightarrow \infty]{L} \sum_{i=1}^M \beta_i \chi_1^2$$

donde los β_i son los autovalores de la matriz

$$\begin{aligned} T &= \text{diag} \left[Q(\theta^0)^{-1/2} \right] \left[\begin{array}{cc} \Sigma_{Q(\theta^0)} & -J(\theta^0)B(\theta^0)\Sigma_{Q(\theta^0)} \\ -J(\theta^0)B(\theta^0)\Sigma_{Q(\theta^0)} & -\Sigma_{Q(\theta^0)} \end{array} \right] B(\theta^0)^t J(\theta^0)^t + \\ &+ J(\theta^0)B(\theta^0)\Sigma_{Q(\theta^0)} B(\theta^0)^t J(\theta^0)^t \text{diag} \left[Q(\theta^0)^{-1/2} \right] = \\ &= \text{diag} \left[Q(\theta^0)^{-1/2} \right] \Sigma_{Q(\theta^0)} \text{diag} \left[Q(\theta^0)^{-1/2} \right] - \\ &- \text{diag} \left[Q(\theta^0)^{-1/2} \right] J(\theta^0)B(\theta^0)\Sigma_{Q(\theta^0)} \text{diag} \left[Q(\theta^0)^{-1/2} \right] - \\ &- \text{diag} \left[Q(\theta^0)^{-1/2} \right] \Sigma_{Q(\theta^0)} B(\theta^0)^t J(\theta^0)^t \text{diag} \left[Q(\theta^0)^{-1/2} \right] + \\ &+ \text{diag} \left[Q(\theta^0)^{-1/2} \right] J(\theta^0)B(\theta^0)\Sigma_{Q(\theta^0)} B(\theta^0)^t J(\theta^0)^t \text{diag} \left[Q(\theta^0)^{-1/2} \right] \end{aligned}$$

donde

$$J(\theta^0)B(\theta^0) = \left[\frac{\partial q_i(\theta^0)}{\partial \theta_r} \right]_{i,r} (A(\theta^0)^t A(\theta^0))^{-1} A(\theta^0)^t \text{diag} \left[Q(\theta^0)^{-1/2} \right]$$

siendo

$$A(\theta^0) = \text{diag}\left[Q(\theta^0)^{-1/2}\right]J(\theta^0).$$

Ahora bien,

$$\text{diag}\left[Q(\theta^0)^{-1/2}\right]\Sigma_{Q(\theta^0)}\text{diag}\left[Q(\theta^0)^{-1/2}\right] = I - Q(\theta^0)^{1/2}\left[Q(\theta^0)^{1/2}\right]^t = L(\theta^0)$$

y

$$\begin{aligned} & \text{diag}\left[Q(\theta^0)^{-1/2}\right]J(\theta^0)B(\theta^0)\Sigma_{Q(\theta^0)}\text{diag}\left[Q(\theta^0)^{-1/2}\right] = \\ & = \text{diag}\left[Q(\theta^0)^{-1/2}\right]J(\theta^0)B(\theta^0)\left[\text{diag}\left[Q(\theta^0)\right] - \text{diag}\left[Q(\theta^0)^{1/2}\right]\right. \\ & \quad \left.\times Q(\theta^0)^{1/2}\left[Q(\theta^0)^{1/2}\right]^t\text{diag}\left[Q(\theta^0)^{1/2}\right]\right]\text{diag}\left[Q(\theta^0)^{-1/2}\right] = M(\theta^0)L(\theta^0) \end{aligned}$$

donde

$$M(\theta^0) = \text{diag}\left[Q(\theta^0)^{-1/2}\right]J(\theta^0)B(\theta^0)\text{diag}\left[Q(\theta^0)^{1/2}\right]$$

con lo cual se tiene

$$T = L(\theta^0) - L(\theta^0)M(\theta^0)^t - M(\theta^0)L(\theta^0) + M(\theta^0)L(\theta^0)M(\theta^0)^t.$$

Al ser,

$$M(\theta^0) = A(\theta^0)(A(\theta^0)^tA(\theta^0))^{-1}A(\theta^0)^t \text{ y } \left[Q(\theta^0)^{1/2}\right]^tA(\theta^0)=0,$$

se sigue que,

$$T = I - \left[Q(\theta^0)^{1/2} \right] \left[Q(\theta^0)^{1/2} \right]^t - A(\theta^0)(A(\theta^0)^t A(\theta^0))^{-1} A(\theta^0)^t$$

donde T es idempotente ya que,

$$\begin{aligned} T^2 &= I - \left[Q(\theta^0)^{1/2} \right] \left[Q(\theta^0)^{1/2} \right]^t - A(\theta^0)(A(\theta^0)^t A(\theta^0))^{-1} A(\theta^0)^t - \\ &- \left[Q(\theta^0)^{1/2} \right] \left[Q(\theta^0)^{1/2} \right]^t + \left[Q(\theta^0)^{1/2} \right] \left[Q(\theta^0)^{1/2} \right]^t \left[Q(\theta^0)^{1/2} \right] \left[Q(\theta^0)^{1/2} \right]^t + \\ &+ \left[Q(\theta^0)^{1/2} \right] \left[Q(\theta^0)^{1/2} \right]^t A(\theta^0)(A(\theta^0)^t A(\theta^0))^{-1} A(\theta^0)^t - \\ &- A(\theta^0)(A(\theta^0)^t A(\theta^0))^{-1} A(\theta^0)^t + \\ &+ A(\theta^0)(A(\theta^0)^t A(\theta^0))^{-1} A(\theta^0)^t \left[Q(\theta^0)^{1/2} \right] \left[Q(\theta^0)^{1/2} \right]^t + \\ &+ A(\theta^0)(A(\theta^0)^t A(\theta^0))^{-1} A(\theta^0)^t A(\theta^0)(A(\theta^0)^t A(\theta^0))^{-1} A(\theta^0)^t = \\ &= I - \left[Q(\theta^0)^{1/2} \right] \left[Q(\theta^0)^{1/2} \right]^t - A(\theta^0)(A(\theta^0)^t A(\theta^0))^{-1} A(\theta^0)^t = T. \end{aligned}$$

Al ser la matriz T idempotente tiene únicamente autovalores 0 y 1, siendo el número de autovalores unitarios

$$\begin{aligned} \text{traza}(T) &= \text{traza}(I) - \text{traza} \left[\left[Q(\theta^0)^{1/2} \right] \left[Q(\theta^0)^{1/2} \right]^t \right] - \\ &- \text{traza} \left[A(\theta^0)(A(\theta^0)^t A(\theta^0))^{-1} A(\theta^0)^t \right] = M - 1 - M_0 \end{aligned}$$

con lo cual se tiene el resultado enunciado. ■

El resultado siguiente junto con Teorema 2.3.2 b) y Corolario 3.3.1 proporciona una familia de estadísticos para contrastar la hipótesis de que la distribución de una muestra X_1, \dots, X_n pertenece a $\{P_\theta / \theta \in \Theta\}$ y el parámetro es θ^0 .

Corolario 3.3.2

Bajo las condiciones de regularidad de Birch y las condiciones i) e ii) del Teorema 2.3.1, se tiene que

$$n R(\hat{P}, Q(\hat{\theta}_1)) \quad \text{y} \quad \sqrt{n}(\hat{\theta}_1 - \theta^0)$$

son asintóticamente independientes.

Demostración

Por el Lema 3.3.1, al ser \hat{P} y $Q(\hat{\theta}_1)$ \sqrt{n} -consistentes se tiene

$$\sqrt{n} R_\phi(\hat{P}, Q(\hat{\theta}_1)) \approx n (\hat{P} - Q(\hat{\theta}_1))^t \text{diag} [Q(\theta^0)^{-1}] (\hat{P} - Q(\hat{\theta}_1)),$$

luego bastará con demostrar que

$$n(\hat{P} - Q(\hat{\theta}_1))^t \text{diag} [Q(\theta^0)^{-1}] (\hat{P} - Q(\hat{\theta}_1)) \quad \text{y} \quad \sqrt{n}(\hat{\theta}_1 - \theta^0)$$

son asintóticamente independientes.

$$\text{Sea } Z = \sqrt{n} \text{diag} [Q(\theta^0)^{-1/2}] (\hat{P} - Q(\theta^0)) \approx N[0, I - \sqrt{Q(\theta^0)} \sqrt{Q(\theta^0)}^{-1}],$$

puesto que

$$\sqrt{n}(\hat{\theta}_1 - \theta^0) \approx \sqrt{n} (A(\theta^0)^t A(\theta^0))^{-1} A(\theta^0)^t \text{diag} [Q(\theta^0)^{-1/2}] (\hat{P} - Q(\theta^0))$$

es suficiente demostrar que

$$(A(\theta^0)^t A(\theta^0))^{-1} A(\theta^0)^t Z \quad \text{y} \quad Z^t (I - A(\theta^0)(A(\theta^0)^t A(\theta^0))^{-1} A(\theta^0)^t) Z$$

son asintóticamente independientes.

Aplicando el resultado 3b.4.viii, pag.188 dado por Rao (1973), esto es equivalente a probar que

$$\Sigma B \Sigma C = 0$$

donde

$$\Sigma = \left[I - \sqrt{\pi} \sqrt{\pi}^t \right],$$

$$B = \left[I - A(\theta^0)(A(\theta^0)^t A(\theta^0))^{-1} A(\theta^0)^t \right]$$

y

$$C = \left[A(\theta^0)(A(\theta^0)^t A(\theta^0))^{-1} A(\theta^0)^t \right],$$

lo cual se obtiene después de algunas operaciones algebraicas. ■

En el caso de estimar θ por máxima verosimilitud a partir del modelo discretizado, la distribución asintótica de $R_\phi(\hat{P}, Q(\hat{\theta}_{EMV}))$ bajo H_0 se establece en el siguiente teorema.

Teorema 3.3.3

Sea $\phi: (0, \infty) \rightarrow \mathbb{R}$ una función cóncava con derivada segunda continua. Sea \hat{P} el vector de frecuencias relativas y $\hat{Q}_{EMV} = Q(\hat{\theta}_{EMV})$ donde $\hat{\theta}_{EMV}$ es el estimador de máxima verosimilitud, entonces bajo las condiciones de regularidad de Birch y las condiciones i) e ii) del Teorema 2.3.1 se verifica que

$$8n R_{\phi}(\hat{P}, \hat{Q}_{EMV}) \xrightarrow[n \rightarrow \infty]{L} \sum_{i=1}^M \beta_i \chi_i^2$$

donde las χ_i^2 son independientes y los β_i son los autovalores de la matriz $D(\theta^0)\Sigma_2$, donde

$$D(\theta^0) = \text{diag}[-\phi''(Q(\theta^0))]$$

y

$$\begin{aligned} \Sigma_2 &= \left[I - J(\theta^0)I(\theta^0)^{-1}J(\theta^0)\text{diag}(Q(\theta^0)^{-1}) \right] \Sigma_{Q(\theta^0)} \\ &\quad \times \left[I - J(\theta^0)I(\theta^0)^{-1}J(\theta^0)\text{diag}(Q(\theta^0)^{-1}) \right]^t \end{aligned}$$

siendo, $I(\theta^0)$ la matriz de Información de Fisher del modelo discretizado y $\Sigma_{Q(\theta^0)} = \text{diag}(Q(\theta^0)) - Q(\theta^0)Q(\theta^0)^t$.

Demostración

Por el Lema 3.3.1, al ser \hat{P} y \hat{Q}_{EMV} estimadores \sqrt{n} -consistentes, se tiene que

$$8n R_{\phi}(\hat{P}, \hat{Q}_{EMV}) \approx n (\hat{P} - \hat{Q}_{EMV})^t \text{diag}[-\phi''(Q(\theta^0))] (\hat{P} - \hat{Q}_{EMV}).$$

Además, a partir del Lema 2 de Morales y otros (1995) se tiene que

$$\sqrt{n}(\hat{P} - \hat{Q}_{EMV}) \xrightarrow[n \rightarrow \infty]{L} N(0, \Sigma_2),$$

donde

$$\begin{aligned} \Sigma_2 &= \left[I - J(\theta^0)I(\theta^0)^{-1}J(\theta^0)\text{diag}(Q(\theta^0)^{-1}) \right] \Sigma_{Q(\theta^0)} \times \\ &\quad \times \left[I - J(\theta^0)I(\theta^0)^{-1}J(\theta^0)\text{diag}(Q(\theta^0)^{-1}) \right]^t. \end{aligned}$$

Por tanto, $8n R_{\phi}(\hat{P}, \hat{Q})$ se distribuye asintóticamente como $\sum_{i=1}^M \beta_i \chi_1^2$ donde las χ_1^2 son independientes y los β_i son los autovalores de la matriz $D(\theta^0)\Sigma_2$. ■

Corolario 3.3.3

Bajo las condiciones de regularidad de Birch y las condiciones i) e ii) del Teorema 2.3.1, se tiene que

$$8n R(\hat{P}, \hat{Q}_{EMV}) \xrightarrow[n \rightarrow \infty]{L} \chi_{M-M_0-1}^2.$$

Demostración

Por el Teorema 3.3.3.

$$8n R(\hat{P}, \hat{Q}_{EMV}) \xrightarrow[n \rightarrow \infty]{L} \sum_{i=1}^M \beta_i \chi_1^2$$

donde los β_i son los autovalores de la matriz T del Corolario 3.3.1. ■

En el apartado 3.2 se ilustra el contraste propuesto para hipótesis nula equiprobable con un ejemplo. A continuación se vuelve sobre el mismo ejemplo, ya que el modelo equiprobable allí propuesto no era adecuado, considerando ahora un modelo más complicado.

Ejemplo 3.3.1

El modelo que se plantea en este apartado para el Ejemplo 3.2.1 tiene en cuenta tanto la primera como la segunda posición ganadora.

Sea $\pi_{ij} = P(\text{galgo } i \text{ consiga la primera posición y el galgo } j \text{ la segunda})$, entonces si suponemos que el galgo i gana la primera posición con probabilidad π_i se puede considerar la segunda posición como la posición ganadora de una subcarrera de los 7 galgos restantes. Es decir, se considera el modelo

$$H_0: \pi_{ij} = P(\text{galgo } i \text{ gane})P(\text{galgo } j \text{ gane entre los galgos restantes}) = \pi_i \pi_j / (1 - \pi_i)$$

para $i=1, \dots, 8; j=1, \dots, 8; i \neq j$. Obviamente $\pi_{ii} = 0$ para $i=1, \dots, 8$.

En este ejemplo, $\theta = (\theta_1, \dots, \theta_7) = (\pi_1, \dots, \pi_7) \in \{\theta \in (0, 1)^7 / \theta_1 + \dots + \theta_7 < 1\}$ y $Q(\theta) = (q_{ij}(\theta))$ donde

$$q_{ij}(\theta) = \pi_{ij} = \begin{cases} \frac{\theta_i \theta_j}{1 - \theta_i} & i=1, \dots, 7; j=1, \dots, 7; i \neq j \\ \frac{\theta_i (1 - \theta_1 - \dots - \theta_7)}{1 - \theta_i} & i=1, \dots, 7; j=8 \\ \frac{(1 - \theta_1 - \dots - \theta_7) \theta_j}{\theta_1 + \dots + \theta_7} & i=8; j=1, \dots, 7. \end{cases}$$

La Tabla 3.3.1 contiene las frecuencias observadas y las frecuencias esperadas obtenidas estimando el parámetro θ tanto por máxima verosimilitud como por mínima R-divergencia. El estimador de máxima verosimilitud viene dado por

$$\hat{\theta}_{EMV} = (0.1787, 0.136, 0.1145, 0.1117, 0.1099, 0.1029, 0.1122)$$

y el de mínima R-divergencia por

$$\hat{\theta}_1 = (0.1793, 0.1376, 0.115, 0.1125, 0.1106, 0.101, 0.1123).$$

Tabla 3.3.1

1ª Posición	2ª Posición								Total
	1	2	3	4	5	6	7	8	
1*		14	11	11	17	17	15	19	104
		17.6	14.8	14.5	14.2	13.3	14.5	17.4	106.3
		17.9	14.9	14.6	14.4	13.1	14.6	17.1	106.6
2	22		12	14	15	6	14	12	95
	16.7		10.7	10.5	10.3	9.6	10.5	12.6	80.9
	17.0		10.9	10.7	10.5	9.6	10.7	12.5	81.9
3	13	10		9	9	12	8	5	66
	13.7	10.5		8.6	8.5	7.9	8.6	10.3	68.1
	13.9	10.6		8.7	8.5	7.8	8.7	10.2	68.4
4	10	10	5		13	12	8	5	63
	13.4	10.2	8.6		8.2	7.7	8.4	10.0	66.5
	13.5	10.4	8.7		8.3	7.6	8.5	9.9	66.9
5	12	7	8	9		7	9	10	62
	13.1	10.0	8.4	8.2		7.6	8.2	9.9	65.4
	13.3	10.2	8.5	8.3		7.5	8.3	9.7	65.8
6	10	8	10	5	7		9	9	58
	12.2	9.3	7.8	7.6	7.5		7.7	9.1	61.2
	12.0	9.2	7.7	7.5	7.4		7.5	8.8	60.1
7	8	6	12	8	6	9		11	60
	13.4	10.2	8.6	8.4	8.3	7.7		10.1	66.8
	13.5	10.4	8.7	8.5	8.3	7.6		9.9	66.9
8	27	9	14	16	4	4	13		87
	16.5	12.5	10.6	10.3	10.1	9.5	10.3		79.8
	16.2	12.4	10.4	10.1	10.0	9.1	10.1		78.3
Total	102	64	72	72	71	67	76	71	595
	99.0	80.3	69.5	68.1	67.1	63.4	68.2	79.4	595.0
	99.4	81.1	69.8	68.4	67.4	62.3	68.4	78.1	595.0

* En la primera fila de cada par aparecen las frecuencias observadas y en la segunda y tercera las frecuencias esperadas obtenidas utilizando $\hat{\theta}_{EMV}$ y $\hat{\theta}_1$, respectivamente.

En la Tabla 3.3.2 aparecen los valores del estadístico $8nR(\hat{P}, \hat{Q})$ para los estimadores calculados. Utilizando los resultados de los Corolarios 3.3.1 y 3.3.3 no se debería rechazar la hipótesis nula propuesta a un nivel de significación del 5% puesto que el valor del estadístico estimando el parámetro tanto por máxima verosimilitud como por mínima R-divergencia es menor que $\chi^2_{56-7-1, 0.05} = 64.1$.

Tabla 3.3.2

$8nR(\hat{P}, Q(\hat{\theta}_{EMV}))$	$8nR(\hat{P}, Q(\hat{\theta}_1))$
59.08	59.053

De esta forma, se debería aceptar el modelo propuesto bajo la hipótesis nula.

3.4.- El problema de bondad de ajuste cuando se incrementa el número de clases

Hasta ahora una suposición básica que se ha venido haciendo en los resultados obtenidos ha sido que el número de clases, M , permanecía fijo al incrementarse el tamaño muestral. Sin embargo, parece obvio, que se pierde información sobre el problema original manteniendo fijo el número de clases, y por tanto, es razonable hacer $M \rightarrow \infty$ cuando $n \rightarrow \infty$. En esta situación no se pueden aplicar los resultados utilizados previamente puesto que el número de valores observados en cada celda no tiene por qué crecer con n . Además, evidentemente, la estructura y dimensión del espacio de probabilidad subyacente varían con M .

En este sentido, Hoeffding (1965), Morris (1966,1975) y Holst

(1972) estudiaron las distribuciones asintóticas de los estadísticos χ^2 de Pearson y logaritmo de la razón de verosimilitud cuando $M \rightarrow \infty$ de forma que n/M estuviera acotado. Koehler y Larntz (1980) realizan un estudio de Monte Carlo de la precisión de estas distribuciones bajo la hipótesis nula.

En este apartado, se supondrá que M es una variable independiente y que el tamaño muestral $n = n_M$ crece con M de forma que

$$v_M = \frac{n_M}{M} \rightarrow v \quad 0 < v < \infty. \quad (3.4.1)$$

A partir de ahora, todas las convergencias y expresiones asintóticas se suponen para $M \rightarrow \infty$. Además, las frecuencias observadas en las clases se denotarán por un vector aleatorio $X_M = (X_{M1}, \dots, X_{MM})$ que se distribuye como una multinomial de parámetros n_M y $p_M = (p_{M1}, \dots, p_{MM})$.

Uno de los resultados más importantes que se obtienen en este apartado se basa en un teorema límite propuesto por Holst (1972) que dice lo siguiente:

"Sea $X_M = (X_{M1}, \dots, X_{MM})$ un vector aleatorio con distribución multinomial de parámetros n_M y $p_M = (p_{M1}, \dots, p_{MM})$ y el estadístico

$$S_M = \sum_{i=1}^M h_M(X_{Mi}, i/M),$$

donde $h_M: \{0, 1, 2, \dots\} \times [0, 1] \rightarrow \mathbb{R}$ es una función medible que satisface, para $c_1, c_2 \in \mathbb{R}$ no dependientes de M , la condición

$$|h_M(u, v)| \leq c_1 e^{c_2 u}. \quad (3.4.2)$$

Entonces, si existe $c_0 \in \mathbb{R}$ independiente de M tal que

$$Mp_{Mi} \leq c_0 \quad \forall 1 \leq i \leq M, \quad (3.4.3)$$

y además,

$$0 < \liminf \frac{\sigma_M^2}{n_M} \leq \limsup \frac{\sigma_M^2}{n_M} < \infty, \quad (3.4.4)$$

se verifica que

$$\frac{S_M - \mu_M}{\sigma_M} \xrightarrow{L} N(0,1)$$

donde

$$\mu_M = \sum_{i=1}^M E[h_M(Y_{Mi}, i/M)]$$

y

$$\sigma_M^2 = \sum_{i=1}^M \text{VAR}[h_M(Y_{Mi}, i/M)] - n_M^{-1} \left[\sum_{i=1}^M \text{COV}(Y_{Mi}, h_M(Y_{Mi}, i/M)) \right]^2$$

siendo Y_{M1}, \dots, Y_{MM} v.a.i. de Poisson de parámetros $n_M p_{Mi}$, $Y_{Mi} \approx \text{Poisson}(n_M p_{Mi})$, $i=1, \dots, M$.

A lo largo de este apartado se obtendrán resultados bajo la hipótesis simétrica (3.2.1). En primer lugar se obtiene la distribución asintótica de $S_\phi(X_M/n_M, \pi_0)$ donde $\pi_{0i} = 1/M$, $i=1, \dots, M$.

Teorema 3.4.1

Sea $\phi: (0, \infty) \rightarrow \mathbb{R}$ una función cóncava continua con $\phi''(1/M) < 0$. Si $M \rightarrow \infty$ y $n_M \rightarrow \infty$ de tal forma que $v_M = \frac{n_M}{M} \rightarrow v$ ($0 < v < \infty$), y

$$\left| \frac{-8n_M M}{\phi''(1/M)} \delta_\phi \left(\frac{t}{n_M}, \frac{1}{M} \right) \right| \leq c_1 e^{c_2 t}$$

donde

$$\delta_{\phi}\left(\frac{t}{n_M}, \frac{1}{M}\right) = \phi\left(\frac{t/n_M + 1/M}{2}\right) - \frac{1}{2}\phi\left(\frac{t}{n_M}\right) - \frac{1}{2}\phi\left(\frac{1}{M}\right)$$

se tiene que,

$$\frac{S_{\phi}(X_M/n_M, \pi_0) - \mu_M}{\sigma_M} \xrightarrow{L} N(0,1)$$

siendo

$$S_{\phi}(X_M/n_M, \pi_0) = \frac{-M}{\phi''(1/M)} 8n_M R_{\phi}(X_M/n_M, \pi_0),$$

$$\mu_M = \frac{-8n_M M^2}{\phi''(1/M)} E[\delta_{\phi}(Y/n_M, 1/M)]$$

y

$$\sigma_M^2 = \frac{64n_M^2 M^3}{(\phi''(1/M))^2} \left\{ \text{VAR}[\delta_{\phi}(Y/n_M, 1/M)] - \frac{M}{n_M} \left[\text{COV}[Y, \delta_{\phi}(Y/n_M, 1/M)] \right]^2 \right\}$$

con $Y \approx \text{Poisson}(v_M)$.

Demostación

Tomando

$$\begin{aligned} h_M(X_{M_i}, i/M) = h_M(X_{M_i}) &= \frac{-8n_M M}{\phi''(1/M)} \left[\phi\left(\frac{X_{M_i}/n_M + 1/M}{2}\right) - \frac{1}{2}\phi\left(\frac{X_{M_i}}{n_M}\right) - \right. \\ &\quad \left. - \frac{1}{2}\phi\left(\frac{1}{M}\right) \right] \end{aligned}$$

se tiene que

$$S_{\phi}(X_M/n_M, \pi_0) = \sum_{i=1}^M h_M(X_{M_i}).$$

Por tanto para demostrar el resultado se debe comprobar que se verifica (3.4.3) y (3.4.4).

La condición (3.4.3) se satisface de forma automática ya que $M\pi_{0i}=1 \quad \forall i=1,\dots,M$. Para verificar la condición (3.4.4) se utiliza otro resultado de Holst (1972) que establece lo siguiente

"Si para alguna función de densidad $g(v)$ en $[0,1]$ se tiene que

$$p_k = \int_{(k-1)/M}^{k/M} g(v) \, dv \quad (k=1,\dots,M)$$

y si $h_M(u,v)$ es continua en v para todo u , entonces

$$\frac{\sigma_M^2}{M} = W + o(1)$$

con

$$W = \int_0^1 \text{VAR}[h_M(Z_v, v)] \, dv - \frac{1}{v} \left[\int_0^1 \text{COV}[Z_v, h_M(Z_v, v)] \, dv \right]^2$$

donde Z_v es una v.a. de Poisson de media $vg(v)$, $v \in [0,1]$ y $W = 0$ si y solo si $h_M(u,v) = a_M u + b_M(v)$ casi seguro."

En el caso que nos ocupa basta considerar $g(v) = 1$, $v \in [0,1]$ ya que $p_k = \frac{1}{M} \quad \forall k=1,\dots,M$, entonces

$$\frac{\sigma_M^2}{n_M} = \frac{\sigma_M^2}{M} \frac{M}{n_M} \rightarrow \frac{W}{v}$$

y por tanto la condición (3.4.4) se verifica. La demostración del teorema se tiene sin mas que tener en cuenta el resultado de Holst enunciado antes del Teorema 3.4.1.

■

Por tanto, el correspondiente test de bondad de ajuste rechazaría la hipótesis nula a nivel γ si

$$\frac{S_{\phi}(X_M/n_M, \pi_0) - \mu_M}{\sigma_M} > z_{\gamma}$$

donde z_{γ} es tal que $P(N(0,1) > z_{\gamma}) = \gamma$.

En el siguiente corolario se establece que para la familia $\phi_{\alpha}(x) = \frac{1}{1-\alpha}(x^{\alpha}-x)$ se verifica la acotación del Teorema 3.4.1 y además se obtiene la forma explícita de la distribución asintótica del estadístico X^2 de Pearson que se corresponde con

$$S_{\phi_2}(X_M/n_M, \pi_0) = \sum_{i=1}^M \frac{(X_{Mi} - n_M/M)^2}{n_M/M}$$

Corolario 3.4.1

Si $M \rightarrow \infty$ y $n_M \rightarrow \infty$ de tal forma que $\frac{n_M}{M} \rightarrow v$ ($0 < v < \infty$), entonces

$$i) \quad \left| \frac{-8n_M M}{\phi_{\alpha}''(1/M)} \delta_{\phi_{\alpha}} \left[\frac{t}{n_M}, \frac{1}{M} \right] \right| \leq c_1 e^{c_2 t}$$

$$ii) \quad \frac{S_{\phi_2}(X_M/n_M, \pi_0) - M}{\sqrt{2M}} \xrightarrow{L} N(0,1)$$

Demostración

i) Al ser

$$\left| \frac{-8n_M M}{\phi_{\alpha}''(1/M)} \delta_{\phi_{\alpha}} \left[\frac{t}{n_M}, \frac{1}{M} \right] \right| =$$

$$\begin{aligned}
&= \left| \frac{8n_M}{\alpha(1-\alpha)M} \left[\left(\frac{(tM/n_M)+1}{2} \right)^\alpha - \frac{1}{2} \left(\frac{tM}{n_M} \right)^\alpha - \frac{1}{2} \right] \right| \leq \\
&\leq \frac{8n_M}{\alpha|1-\alpha|M} \left[\left(\frac{(tM/n_M)+1}{2} \right)^\alpha + \frac{1}{2} \left(\frac{tM}{n_M} \right)^\alpha + \frac{1}{2} \right],
\end{aligned}$$

se tendrán que acotar convenientemente cada uno de los sumandos de la expresión anterior para comprobar que se verifica la condición i).

Para el primer término se tiene

$$\begin{aligned}
&\frac{8n_M}{\alpha|1-\alpha|M} \left(\frac{(tM/n_M)+1}{2} \right)^\alpha \leq \\
&\leq \frac{8n_M/M}{\alpha|1-\alpha|} \max \left\{ \left[\frac{(tM/n_M)+1}{2} \right]^{[\alpha]}, \left[\frac{(tM/n_M)+1}{2} \right]^{[\alpha]+1} \right\} \leq \\
&\leq \frac{8n_M/M}{\alpha|1-\alpha|} [\alpha+1]! \max \left\{ \frac{\left[\frac{(tM/n_M)+1}{2} \right]^{[\alpha]}}{[\alpha]!}, \frac{\left[\frac{(tM/n_M)+1}{2} \right]^{[\alpha+1]}}{[\alpha+1]!} \right\} \leq \\
&\leq \frac{8n_M/M}{\alpha|1-\alpha|} [\alpha+1]! \exp \left\{ \frac{(tM/n_M)+1}{2} \right\} = a_M^1 e^{b_M^1 t}
\end{aligned}$$

donde

$$a_M^1 = \frac{8n_M/M}{\alpha|1-\alpha|} [\alpha+1]! e^{1/2}, \quad b_M^1 = \frac{M}{n_M} \frac{1}{2}$$

y por [a] se denota la parte entera de a.

Por tanto

$$\frac{8n_M}{\alpha|1-\alpha|M} \left[\frac{(tM/n_M)+1}{2} \right]^\alpha \leq a_1^* e^{b_1^* t}$$

donde

$$a_1^* = \sup\{a_M^1\} \quad y \quad b_1^* = \sup\{b_M^1\}$$

que sabemos que existen y son finitos porque

$$\lim a_M^1 = \frac{8v}{\alpha|1-\alpha|} [\alpha+1]! e^{1/2}$$

y

$$\lim b_M^1 = \frac{1}{2v}$$

Análogamente, es posible encontrar a_2^* , b_2^* , $a_3^* \in \mathbb{R}$ e independientes de M tales que

$$\frac{4n_M}{\alpha|1-\alpha|M} \left[\frac{tM}{n_M} \right]^\alpha \leq a_2^* e^{b_2^* t}$$

y

$$\frac{4n_M}{\alpha|1-\alpha|M} \leq a_3^*$$

Consecuentemente,

$$\left| \frac{-8n_M M}{\phi_\alpha''(1/M)} \delta_{\phi_\alpha} \left[\frac{t}{n_M}, \frac{1}{M} \right] \right| \leq a_1^* e^{b_1^* t} + a_2^* e^{b_2^* t} + a_3^* \leq c_1 e^{c_2 t}$$

donde

$$c_1 = 3 \max\{a_1^*, a_2^*, a_3^*\}$$

y

$$c_2 = \max\{b_1^*, b_2^*\},$$

como se quería demostrar.

ii) Para $\alpha=2$, se tiene

$$\begin{aligned} E[\delta_{\phi_2}(Y/n_M, 1/M)] &= E\left[-\left(\frac{(Y/n_M)+1/M}{2}\right)^2\right] + \frac{1}{2} E\left[\left(\frac{Y}{n_M}\right)^2\right] + \frac{1}{2M^2} = \\ &= -\frac{1}{M^2} \left\{ E\left[\left(\frac{(YM/n_M)+1}{2}\right)^2\right] - \frac{1}{2} E\left[\left(\frac{YM}{n_M}\right)^2\right] - \frac{1}{2} \right\} = \\ &= -\frac{1}{M^2} \left\{ \frac{1}{4} E\left[\left(\frac{YM}{n_M}\right)^2\right] + \frac{1}{4} + \frac{1}{2} E\left[\left(\frac{YM}{n_M}\right)\right] - \frac{1}{2} E\left[\left(\frac{YM}{n_M}\right)^2\right] - \right. \\ &\quad \left. - \frac{1}{2} \right\} = \frac{1}{4Mn_M}, \end{aligned}$$

$$\begin{aligned} \text{VAR}[\delta_{\phi_2}(Y/n_M, 1/M)] &= \text{VAR}\left[\frac{1}{4}\left(\frac{1}{n_M}\right)^2 Y^2 - \frac{1}{2Mn_M} Y\right] = \\ &= \frac{1}{16}\left(\frac{1}{n_M}\right)^4 \text{VAR}[Y^2] + \frac{1}{4}\left(\frac{1}{Mn_M}\right)^2 \text{VAR}[Y] - \frac{1}{4M}\left(\frac{1}{n_M}\right)^3 \text{COV}[Y^2, Y] = \\ &= \frac{1}{16M}\left(\frac{1}{n_M}\right)^3 + \frac{1}{8}\left(\frac{1}{Mn_M}\right)^2 \end{aligned}$$

y

$$\begin{aligned} \text{COV}[Y, \delta_{\phi_2}(Y/n_M, 1/M)] &= E\left[Y\left[\frac{1}{4}\left(\frac{1}{n_M}\right)^2 Y^2 - \frac{1}{2Mn_M} Y + \frac{1}{4M^2}\right]\right] - \\ &= E[Y]E[\delta_{\phi_2}(Y/n_M, 1/M)] = \frac{1}{4}\left(\frac{1}{n_M}\right)^2 E[Y^3] - \frac{1}{2Mn_M} E[Y^2] + \\ &+ \frac{1}{4M^2} E[Y] - \frac{1}{4M^2} = \frac{1}{4Mn_M} \end{aligned}$$

entonces

$$\mu_M = 4n_M M^2 \left(\frac{1}{4Mn_M}\right) = M$$

y

$$\sigma_M^2 = 16n_M^2 M^3 \left\{ \frac{1}{16M} \left(\frac{1}{n_M}\right)^3 + \frac{1}{8} \left(\frac{1}{Mn_M}\right)^2 - \frac{M}{n_M} \left(\frac{1}{4Mn_M}\right)^2 \right\} = 2M$$

con lo que se tiene el resultado enunciado. ■

Es trivial comprobar que para $\phi_1(x) = -x \ln x$ se verifica la cota del Teorema 3.4.1 y por tanto la obtención de la distribución asintótica del estadístico basado en esta función es inmediata a partir de este teorema.

Corolario 3.4.2

Si $M \rightarrow \infty$ y $n_M \rightarrow \infty$ de tal forma que $\frac{n_M}{M} \rightarrow v$ ($0 < v < \infty$), se tiene que

$$\frac{S_{\phi_1}(X_M/n_M, \pi_0) - \mu_{1M}}{\sigma_{1M}} \xrightarrow{L} N(0,1)$$

siendo

$$S_{\phi_1}(X_M/n_M, \pi_0) = 8nR_{\phi_1}(X_M/n_M, \pi_0),$$

$$\mu_{1M} = 4M \left\{ E \left[Y \ln \frac{Y}{v_M} \right] - E \left[(Y+v_M) \ln \left[\frac{Y}{2v_M} + \frac{1}{2} \right] \right] \right\}$$

y

$$\begin{aligned} \sigma_{1M}^2 = & 16M \left\{ \text{VAR} \left[Y \ln \frac{Y}{v_M} \right] + \text{VAR} \left[(Y+v_M) \ln \left[\frac{Y}{2v_M} + \frac{1}{2} \right] \right] - \right. \\ & - 2\text{COV} \left[Y \ln \frac{Y}{v_M}, (Y+v_M) \ln \left[\frac{Y}{2v_M} + \frac{1}{2} \right] \right] - \\ & \left. - \frac{1}{v_M} \left[\text{COV} \left[Y, Y \ln \frac{Y}{v_M} \right] - \text{COV} \left[Y, (Y+v_M) \ln \left[\frac{Y}{2v_M} + \frac{1}{2} \right] \right] \right]^2 \right\}. \end{aligned}$$

Demostración

Es inmediata con solo considerar que en este caso

$$\delta_{\phi_1} \left(\frac{Y}{n_M}, \frac{1}{M} \right) = \frac{1}{2n_M} \left\{ Y \ln \frac{YM}{n_M} - \left[Y + \frac{n_M}{M} \right] \ln \left[\frac{YM}{2n_M} + \frac{1}{2} \right] \right\}.$$

■

Hasta ahora se ha visto que al contrario que ocurría cuando M era fijo las distribuciones de los estadísticos S_{ϕ} bajo la hipótesis nula cuando el número de clases M crece no son asintóticamente equivalentes. A continuación se calcula la distribución asintótica de S_{ϕ} cuando $M \rightarrow \infty$ bajo las alternativas

$$H_{1,n} : \pi = \pi_0 + \varepsilon \quad (3.4.5)$$

con

$$\pi_{0i} = \frac{1}{M} \quad \text{para } i=1,\dots,M$$

y

$$\varepsilon_i = \int_{(i-1)/M}^{i/M} c(x)/n^{1/m} dx \quad \begin{array}{l} m=1,2,3,\dots \\ i=1,\dots,M \end{array}$$

donde $c(x)$ es una función continua conocida en $[0,1]$ tal que $\int_0^1 c(x)dx=0$. El objetivo es encontrar el estadístico S_ϕ asintóticamente más eficiente, es decir, el estadístico con mayor potencia para estas alternativas.

Teorema 3.4.2

Bajo las hipótesis alternativas (3.4.5) y las condiciones del Teorema 3.4.1, se tiene que

$$\frac{S_\phi(X_M/n_M, \pi) - \mu_{M,m}}{\sigma_{M,m}} \xrightarrow{L} N(0,1)$$

donde

$$\mu_{M,m} = \frac{-8n_M M}{\phi''(1/M)} \sum_{i=1}^M E\left[\delta_\phi\left(\frac{Y_{Mi}}{n_M}, \frac{1}{M}\right)\right]$$

y

$$\sigma_{M,m}^2 = \frac{64n_M^2 M^2}{(\phi''(1/M))^2} \left\{ \sum_{i=1}^M \text{VAR}\left[\delta_\phi\left(\frac{Y_{Mi}}{n_M}, \frac{1}{M}\right)\right] - \frac{1}{n_M} \left[\sum_{i=1}^M \text{COV}\left[Y_{Mi}, \delta_\phi\left(\frac{Y_{Mi}}{n_M}, \frac{1}{M}\right)\right] \right]^2 \right\}$$

siendo Y_{M1}, \dots, Y_{MM} v.a.i. de Poisson de parámetros $n_M \pi_i$,
 $Y_{Mi} \approx \text{Poisson}(n_M \pi_i)$, $i=1, \dots, M$.

Demostración

Este resultado se obtiene a partir del Teorema dado por Holst (1972) de forma análoga al Teorema 3.4.1, es decir, bastará comprobar las condiciones (3.4.3) y (3.4.4).

En relación a la condición (3.4.3),

$$M\pi_i = 1 + \int_{(i-1)/M}^{i/M} M c(x)/n^{1/m} dx,$$

pero $c(x) \leq k$ si $x \in [0,1]$, por tanto

$$M\pi_i \leq 1 + \frac{k}{n^{1/m}} \leq 1 + k \quad \forall i=1, \dots, M$$

entonces la condición se verifica tomando $c_0 = 1+k$.

En cuanto a (3.4.4), como

$$\pi_i = \frac{1}{M} + \int_{(i-1)/M}^{i/M} c(x)/n^{1/m} dx = \int_{(i-1)/M}^{i/M} \left[1 + c(x)/n^{1/m} \right] dx$$

si se define $g(x) = 1 + \frac{c(x)}{n^{1/m}}$ se tiene la condición del resultado de Holst enunciado en la demostración del Teorema 3.4.1 ya que $g(x)$ es continua en $[0,1]$.

Entonces

$$\frac{\sigma_{M,m}^2}{n_M} = \frac{\sigma_{M,m}^2}{M} \frac{M}{n_M} \rightarrow \frac{W}{V}$$

y la condición (3.4.4) queda verificada.

Por tanto, aplicando el Teorema de Holst se tiene el resultado enunciado. ■

Teorema 3.4.3

La potencia asintótica de los estadísticos S_ϕ bajo las hipótesis (3.4.5) es

$$F_N \left[-z_\gamma + e_{\phi, \nu}^{(m)} \right] \quad (3.4.6)$$

donde

$$e_{\phi, \nu}^{(m)} = \lim \frac{\mu_{M, m} - \mu_M}{\sigma_{M, m}}$$

y

$$z_\gamma = F_N^{-1}(1-\gamma)$$

siendo F_N la función de distribución de una Normal de media cero y varianza uno.

Demostración

La función de potencia de la familia S_ϕ cuando se consideran las alternativas (3.4.5) viene dada por

$$\beta_\phi(\pi) = P \left[S_\phi(X_M/n_M, \pi_0) > z_\gamma \sigma_M + \mu_M \mid H_{1,n} \right]$$

donde $z_\gamma = F_N^{-1}(1-\gamma)$.

Operando se tiene

$$\begin{aligned} \beta_{\phi}(\pi) &= P \left[\frac{S_{\phi}(X_M/n_M, \pi_0) - \mu_{M,m}}{\sigma_{M,m}} > \frac{z_{\gamma} \sigma_M + \mu_M - \mu_{M,m}}{\sigma_{M,m}} \mid H_{1,n} \right] = \\ &= 1 - P \left[\frac{S_{\phi}(X_M/n_M, \pi_0) - \mu_{M,m}}{\sigma_{M,m}} \leq \frac{z_{\gamma} \sigma_M + \mu_M - \mu_{M,m}}{\sigma_{M,m}} \mid H_{1,n} \right]. \end{aligned}$$

Como

$$\lim \frac{\sigma_M}{\sigma_{M,m}} = 1,$$

la potencia asintótica de S_{ϕ} viene dada por

$$1 - F_N \left[z_{\gamma} - e_{\phi, \nu}^{(m)} \right]$$

como se quería demostrar. ■

Corolario 3.4.3

En el caso $m=2$, Holst (1972) demostró que para estadísticos simétricos, es decir, estadísticos para los cuales $h_M(X_{M_i}, i/M) = h_M(X_{M_i})$, se verifica que $\frac{\mu_{M,m} - \mu_M}{\sigma_{M,m}} \longrightarrow 0$.

Es inmediato que los estadísticos S_{ϕ} son simétricos y por tanto

$$e_{\phi, \nu}^{(2)} = 0 \quad \forall \nu.$$

En el teorema siguiente se establece que el estadístico deducido

a partir de la entropía cuadrática, es decir, el estadístico X^2 , es el de máxima eficiencia relativa asintótica Pitman entre los contrastes basados en R_{ϕ_α} para las alternativas (3.4.5) cuando $m=4$.

Teorema 3.4.4

Si $M \rightarrow \infty$ y $n_M \rightarrow \infty$ de tal forma que $\frac{n_M}{M} \rightarrow v$ ($0 < v < \infty$), entonces para contrastar la hipótesis

$$H_0: \pi_i = \frac{1}{M} \quad i=1, \dots, M,$$

frente a

$$H_{1,n}: \pi_i = \frac{1}{M} + \int_{(i-1)/M}^{i/M} c(x)/n^{1/4} dx, i=1, \dots, M,$$

el contraste basado en la familia S_{ϕ_α} con mayor potencia asintótica es el correspondiente a $\alpha=2$ (estadístico X^2 de Pearson).

Además, la potencia asintótica para S_{ϕ_2} viene dada por

$$F_N \left[-z_\gamma + \sqrt{v/2} \int_0^1 (c(x))^2 dx \right]. \quad (3.4.7)$$

Demostración

Particularizando el resultado de Ivchenko y Medvedev (1978) para S_{ϕ_α} , se tiene

$$e_{\phi_\alpha, v}^{(4)} = \sqrt{v/2} \int_0^1 (c(x))^2 dx \text{ CORR} \left[f_\alpha(Y) - b_f Y, Y^2 - (2v+1)Y \right]$$

donde

$$b_f = v^{-1} \text{COV} \left[f_\alpha(Y), Y \right],$$

$$f_{\alpha}(Y) = \frac{8v}{\alpha(\alpha-1)} \left[\left(\frac{Y+v}{2v} \right)^{\alpha} - \frac{1}{2} \left(\frac{Y}{v} \right)^{\alpha} - \frac{1}{2} \right]$$

e

$$Y \approx \text{Poisson}(v).$$

Como maximizar la función de potencia es equivalente a maximizar $e_{\phi_{\alpha}, v}^{(m)}$ en valor absoluto, la eficiencia de $S_{\phi_{\alpha}}$ para $m=4$ queda determinada por el coeficiente

$$\rho_{\alpha} = |\text{CORR}[(Y+v)^{\alpha-2} Y^{\alpha} - b_{\alpha} Y, Y^2 - (2v+1)Y]|$$

donde

$$b_{\alpha} = v^{-1} \text{COV}[(Y+v)^{\alpha-2} Y^{\alpha}, Y].$$

Al ser

$$b_2 = v^{-1} \text{COV}[-2vY - v^2 + Y^2, Y] = v^{-1} [-2v^2 + v + 2v^2] = 1$$

se tiene

$$\rho_2 = 1$$

por tanto el contraste basado en S_{ϕ_2} tiene máxima potencia. Además, (3.4.7) se obtiene de forma inmediata sustituyendo el valor de ρ_2 en (3.4.6). ■

El teorema anterior indica que el estadístico S_{ϕ_2} es óptimo para

contrastar las alternativas dadas en (3.4.5) con $m=4$. No obstante, a continuación se evalúa ρ_α en general para cuantificar la pérdida de potencia que se produce al utilizar valores de $\alpha \neq 2$. El coeficiente, para $\alpha \neq 1$, viene dado por

$$\rho_\alpha = \left| E \left[(Y^2 - (2v+1)Y + v^2) ((Y+v)^{\alpha-2} Y^\alpha) \right] \times \right. \\ \left. \left\{ 2v^2 E \left[((Y+v)^{\alpha-2} Y^\alpha)^2 \right] - 2v^2 \left[E \left[(Y+v)^{\alpha-2} Y^\alpha \right] \right]^2 - \right. \right. \\ \left. \left. - 2v \left[E \left[(Y-v) ((Y+v)^{\alpha-2} Y^\alpha) \right] \right]^2 \right\}^{-1/2} \right| \quad (3.4.8)$$

y para $\alpha=1$ por

$$\rho_1 = \left| E \left[(Y^2 - (2v+1)Y + v^2) f(Y) \right] \left\{ 2v^2 E \left[(f(Y))^2 \right] - \right. \right. \\ \left. \left. - 2v^2 \left[E \left[f(Y) \right] \right]^2 - 2v \left[E \left[(Y-v) f(Y) \right] \right]^2 \right\}^{-1/2} \right| \quad (3.4.9)$$

donde

$$f(Y) = 4 \left\{ Y \ln \frac{Y}{v} - (Y+v) \ln \left[\frac{Y}{2v} + \frac{1}{2} \right] \right\}.$$

Obsérvese que tanto para $\alpha=1$ o α un valor no entero estas expresiones no se pueden evaluar de forma explícita. En estos casos se aproxima numéricamente teniendo en cuenta que para $Y \approx \text{Poisson}(v)$ se tiene que

$$E[g(Y)] = \sum_{i=0}^{\infty} g(i) \psi_i(v)$$

donde

$$\Psi_i(v) = e^{-v} \frac{v^i}{i!}.$$

En la Tabla 3.4.1 se recogen los valores de ρ_α , para diferentes valores de α y v . Estos valores se han obtenido utilizando los 100 primeros términos de la aproximación numérica de (3.4.8) y (3.4.9) descrita anteriormente. La Tabla 3.4.2 contiene los valores $e_{\phi_{\alpha,v}}^{(4)} \sqrt{\int_0^1 (c(x))^2 dx}$. Si bien la función $c(x)$ no afecta al cociente entre $e_{\phi_{\alpha_1,v}}^{(4)}$ y $e_{\phi_{\alpha_2,v}}^{(4)}$ si que cambia el tamaño absoluto de la potencia de los estadísticos puesto que $c(x)$ mide la distancia entre la hipótesis alternativa y la nula.

Tabla 3.4.1
Valores aproximados de ρ_α

α	v								
	0.1	0.5	1.0	1.5	2.0	3.0	10	20	50
1/3	.9776	.9143	.8439	.7814	.7270	.6444	.8436	.9443	.9786
1/2	.9732	.9165	.8540	.8016	.7590	.7038	.8902	.9550	.9827
2/3	.9732	.9224	.8684	.8262	.7947	.7622	.9189	.9646	.9863
1	.9866	.9452	.9092	.8856	.8722	.8679	.9577	.9803	.9923
1.5	.9972	.9850	.9748	.9699	.9686	.9713	.9903	.9952	.9981
13/7	.9997	.9988	.9980	.9978	.9977	.9980	.9992	.9996	.9998
2	.9999	1.000	1.000	1.000	.9999	1.000	.9999	1.000	.9999
2.5	.9968	.9876	.9825	.9811	.9813	.9832	.9922	.9957	.9981
3	.9856	.9561	.9449	.9428	.9438	.9486	.9727	.9841	.9930
4	.9275	.8573	.8480	.8505	.8555	.8665	.9167	.9463	.9741
5	.8145	.7330	.7362	.7473	.7586	.7783	.8532	.8986	.9468

A partir de estas tablas se observa como, por una parte, para v fijo, el valor de α óptimo es 2 como se vio en el Teorema 3.4.3. Además, en la Tabla 3.4.1, a medida que nos alejamos de $\alpha=2$, ρ_α decrece para valores moderados de v y se mantiene próximo a 1 para valores de v cercanos a 0 o valores grandes. Obsérvese que la opción de considerar $S_{\phi_{13/7}}$ es muy buena puesto que la diferencia con el óptimo es de milésimas.

Por otra parte, para α fijo, de la Tabla 3.4.2 se desprende que $e_{\phi_{\alpha,v}}^{(4)}$ es creciente en v , es decir, cuanto mayor es v mayor es la potencia del estadístico. También se observa que existe una clara tendencia a que $e_{\phi_{\alpha,v}}^{(4)}$ se mantenga cerca de $e_{\phi_{\alpha_2,v}}^{(4)}$ para valores de α próximos a 2.

Tabla 3.4.2
Valores aproximados de $e_{\phi_{\alpha,v}}^{(4)} / \int_0^1 (c(x))^2 dx$

α	v								
	0.1	0.5	1.0	1.5	2.0	3.0	10	20	50
1/3	.2186	.4571	.5967	.6767	.7270	.7893	1.8865	2.9863	4.8934
1/2	.2176	.4582	.6039	.6942	.7590	.8619	1.9905	3.0201	4.9136
2/3	.2176	.4612	.6141	.7155	.7947	.9335	2.0548	3.0504	4.9317
1	.2206	.4726	.6429	.7670	.8722	1.0630	2.1415	3.1002	4.9618
1.5	.2230	.4925	.6893	.8400	.9686	1.1896	2.2144	3.1472	4.9905
13/7	.2235	.4994	.7057	.8641	.9977	1.2223	2.2344	3.1610	4.9992
2	.2236	.5000	.7071	.8660	.9999	1.2247	2.2360	3.1622	4.9999
2.5	.2229	.4938	.6947	.8496	.9813	1.2041	2.2187	3.1488	4.9909
3	.2204	.4780	.6681	.8164	.9438	1.1618	2.1752	3.1121	4.9650
4	.2073	.4286	.5996	.7365	.8555	1.0612	2.0499	2.9926	4.8705
5	.1821	.3665	.5205	.6472	.7586	.9533	1.9080	2.8418	4.7343

Para ilustrar la precisión de estas aproximaciones, se calcula la expresión exacta de ρ_3 , que viene dada por

$$\rho_3 = \frac{-(2v+3)}{(4v^2+15v+9)^{1/2}}$$

obteniéndose, para diferentes valores de v , la Tabla 3.4.3.

Tabla 3.4.3
Valores exactos de ρ_3

v	0.1	0.5	1.0	1.5	2.0	3.0	10	20	50
ρ_3	.9856	.9561	.9449	.9428	.9438	.9486	.9727	.9841	.9930

En este apartado se ha estudiado la eficiencia de los estadísticos S_{ϕ_α} en el caso de hipótesis nula simétrica y se ha llegado al óptimo para contrastar esta hipótesis frente a diversas alternativas. Sin embargo, como ilustraron Ivchenko y Medvedev (1978) en algunos ejemplos, este tipo de conclusiones no es posible para hipótesis nulas no simétricas. Aparte de tener el problema añadido de que para este tipo de hipótesis no es posible realizar un estudio conjunto puesto que el número de parámetros puede incrementar sin límite cuando $M \rightarrow \infty$.

CAPITULO IV

OPTIMALIDAD EN LOS CONTRASTES DE BONDAD DE AJUSTE BASADOS EN LA R_{Φ} -DIVERGENCIA PARA MUESTRAS PEQUEÑAS

- 4.1.- Introducción.
- 4.2.- Desarrollo asintótico de segundo orden de los momentos de los estadísticos R_{Φ} -divergencia.
- 4.3.- Desarrollo asintótico de segundo orden de la distribución de los estadísticos R_{Φ} -divergencia.
- 4.4.- Comparación de las diferentes aproximaciones de la distribución exacta de los estadísticos R_{Φ} -divergencia.
- 4.5.- Potencia exacta basada en regiones críticas exactas.

4.1.- Introducción

Los resultados obtenidos hasta el momento de la familia S_ϕ son asintóticos y por tanto correctos para muestras grandes. En este capítulo se aborda, en primer lugar, el problema de encontrar mejores aproximaciones de la distribución exacta de los estadísticos S_ϕ que la aproximación ji-cuadrado.

Varios autores han utilizado correcciones de momentos para obtener mejores aproximaciones de las distribuciones exactas de X^2 y de G^2 . Lewis y otros (1984) obtuvieron expresiones explícitas para los tres primeros momentos de X^2 en tablas de contingencia múltiple. Lawley (1956) obtuvo una aproximación mejorada para la distribución de G^2 a partir de un estadístico obtenido al multiplicar G^2 por un factor de escala con momentos equivalentes a una ji-cuadrado hasta el término $O(n^{-2})$. Smith y otros (1981) extienden el trabajo anterior obteniendo un estadístico con momentos equivalentes al de la ji-cuadrado hasta el término $O(n^{-3})$. Cressie y Read (1984) hacen lo propio para la familia de estadísticos que lleva su nombre. Menéndez y otros (1996) establecen que dentro de la familia de estadísticos de bondad de ajuste basada en la f-divergencia los óptimos son aquellos que verifican $4f^{III}(1)+3f^{IV}(1)=0$. En este trabajo dicho estudio se realiza en el apartado 4.2 para la familia S_ϕ , donde se propone una

modificación de la familia S_ϕ de forma que la nueva familia tenga esperanza y varianza igual a la de una ji-cuadrado más $o(n^{-1})$.

Otra línea seguida por algunos autores para obtener una aproximación matemáticamente más rigurosa es la que se expone en el apartado 4.3 y que consiste en calcular el término de segundo orden del desarrollo de la distribución de S_ϕ considerado en el capítulo anterior. En este sentido, Hoel (1938) calculó el término de segundo orden para la distribución del estadístico X^2 de Pearson y de este resultado concluyó que el error cometido utilizando la aproximación de primer orden es mejor de lo que se podía esperar en el caso de que la distribución subyacente sea continua. Yarnold (1972) calculó el término de segundo orden correcto para distribuciones multinomiales discretas y lo comparó con otras 4 aproximaciones. Siguiendo los resultados de Yarnold (1972), Siotani y Fujikoshi (1984) calcularon el término de segundo orden para las distribuciones de G^2 y de F^2 y Cressie y Read (1984) generalizaron este resultado para la distribución de la familia de estadísticos que lleva su nombre.

En segundo lugar, todas las aproximaciones de la distribución exacta de S_ϕ conseguidas en este capítulo más la ji-cuadrado y la normal obtenidas en el capítulo anterior son comparadas en el apartado 4.4 para muestras pequeñas mediante dos criterios diferentes. De este estudio, se tiene que la sencilla aproximación propuesta en 4.2 es tan precisa como la complicada de 4.3 en muchos casos.

Uno de los criterios más importantes para comparar tests para muestras pequeñas es el cálculo de la potencia exacta de los contrastes frente a diferentes alternativas sin hacer referencia a resultados asintóticos. Wakimoto y otros (1987) calcularon las potencias exactas de X^2 , G^2 y F^2 y Cressie y Read (1984) de la familia de estadísticos que lleva su nombre frente a varias alternativas. Bajo estas alternativas, en el apartado 4.5 se calcula la potencia exacta de la familia S_ϕ .

4.2.- Desarrollo asintótico de segundo orden de los momentos de los estadísticos R_{Φ} -divergencia

En primer lugar, se comparan los momentos asintóticos de S_{ϕ} con los momentos exactos, con el objetivo de encontrar condiciones sobre los estadísticos S_{ϕ} que permitan mayor proximidad entre los momentos asintóticos y exactos bajo la hipótesis nula equiprobable. Además, utilizando estos resultados, se define un estadístico "modificado" de S_{ϕ} cuya distribución se espera se aproxime más a una χ_{M-1}^2 que la de S_{ϕ} .

Puesto que bajo la hipótesis nula simple (3.2.1), el estadístico S_{ϕ} se distribuye como una χ_{M-1}^2 y los momentos de esta distribución límite son finitos, se tiene que

$$\begin{aligned} E[S_{\phi}(X/n, \pi_0)] &\longrightarrow M-1, \\ E[S_{\phi}(X/n, \pi_0)^2] &\longrightarrow M^2-1, \\ E[S_{\phi}(X/n, \pi_0)^3] &\longrightarrow M^3+3M^2-M-3, \end{aligned}$$

es decir, los tres primeros momentos de $S_{\phi}(X/n, \pi_0)$ son asintóticamente equivalentes a los tres primeros momentos de una χ_{M-1}^2 cuando $n \rightarrow \infty$ y M fijo.

Para estudiar la velocidad de convergencia de los momentos exactos de $S_{\phi}(X/n, \pi_0)$ a los asintóticos, se calcula el desarrollo asintótico de segundo orden de estos momentos lo que proporciona información acerca del error que se comete al utilizar la distribución asintótica en lugar de la exacta. Se supondrá que ϕ es tantas veces derivable como sea necesario en cada caso.

Para ello, se considera el desarrollo en serie de Taylor de $S_{\phi}(X/n, \pi_0)$ en torno al punto $P = (1/M, \dots, 1/M) = \pi_0$, que viene dado por

$$\begin{aligned}
R_\phi(X/n, \pi_0) &= \sum_{j=1}^M \left[\frac{\partial R_\phi(P, \pi_0)}{\partial p_j} \right]_{P=\pi_0} \frac{W_j}{\sqrt{n}} + \frac{1}{2!} \sum_{j=1}^M \left[\frac{\partial^2 R_\phi(P, \pi_0)}{\partial^2 p_j} \right]_{P=\pi_0} \frac{W_j^2}{n} + \\
&+ \frac{1}{3!} \sum_{j=1}^M \left[\frac{\partial^3 R_\phi(P, \pi_0)}{\partial^3 p_j} \right]_{P=\pi_0} \frac{W_j^3}{n\sqrt{n}} + \frac{1}{4!} \sum_{j=1}^M \left[\frac{\partial^4 R_\phi(P, \pi_0)}{\partial^4 p_j} \right]_{P=\pi_0} \frac{W_j^4}{n^2} + \\
&+ O_p(n^{-5/2})
\end{aligned}$$

donde

$$W_j = \frac{1}{\sqrt{n}} \left[X_j - \frac{n}{M} \right].$$

Al ser

$$\left[\frac{\partial R_\phi(P, \pi_0)}{\partial p_j} \right]_{P=\pi_0} = \left[\frac{1}{2} \phi' \left[\frac{p_j + \pi_{0j}}{2} \right] - \frac{1}{2} \phi'(p_j) \right]_{P=\pi_0} = 0,$$

$$\left[\frac{\partial^2 R_\phi(P, \pi_0)}{\partial^2 p_j} \right]_{P=\pi_0} = \left[\frac{1}{4} \phi'' \left[\frac{p_j + \pi_{0j}}{2} \right] - \frac{1}{2} \phi''(p_j) \right]_{P=\pi_0} = -\frac{1}{4} \phi'' \left[\frac{1}{M} \right],$$

$$\left[\frac{\partial^3 R_\phi(P, \pi_0)}{\partial^3 p_j} \right]_{P=\pi_0} = \left[\frac{1}{8} \phi''' \left[\frac{p_j + \pi_{0j}}{2} \right] - \frac{1}{2} \phi'''(p_j) \right]_{P=\pi_0} = -\frac{3}{8} \phi''' \left[\frac{1}{M} \right],$$

$$\left[\frac{\partial^4 R_\phi(P, \pi_0)}{\partial^4 p_j} \right]_{P=\pi_0} = \left[\frac{1}{16} \phi^{IV} \left[\frac{p_j + \pi_{0j}}{2} \right] - \frac{1}{2} \phi^{IV}(p_j) \right]_{P=\pi_0} = -\frac{7}{16} \phi^{IV} \left[\frac{1}{M} \right]$$

se tiene que

$$\begin{aligned}
S_{\phi}(X/n, \pi_0) &= \frac{-8nM}{\phi''(1/M)} R_{\phi}(X/n, \pi_0) = M \sum_{j=1}^M W_j^2 + \frac{M\phi'''(1/M)}{2\sqrt{n}\phi''(1/M)} \sum_{j=1}^M W_j^3 + \\
&+ \frac{7M\phi^{IV}(1/M)}{48n\phi''(1/M)} \sum_{j=1}^M W_j^4 + O_p(n^{-3/2}). \tag{4.2.1}
\end{aligned}$$

Como

$$E[W_j^2] = -\frac{1}{M^2} + \frac{1}{M},$$

$$E[W_j^3] = n^{-1/2} \left[\frac{2}{M^3} - \frac{3}{M^2} + \frac{1}{M} \right],$$

y

$$E[W_j^4] = \left[\frac{3}{M^4} - \frac{6}{M^3} + \frac{3}{M^2} \right] + n^{-1} \left[-\frac{6}{M^4} + \frac{12}{M^3} - \frac{7}{M^2} + \frac{1}{M} \right],$$

el primer momento viene dado por

$$\begin{aligned}
E[S_{\phi}(X/n, \pi_0)] &= \sum_{j=1}^M \left[-\frac{1}{M} + 1 \right] + \frac{\phi'''(1/M)}{2n\phi''(1/M)} \sum_{j=1}^M \left[\frac{2}{M^2} - \frac{3}{M} + 1 \right] + \\
&+ \frac{7\phi^{IV}(1/M)}{48\phi''(1/M)} \left[n^{-1} \sum_{j=1}^M \left[\frac{3}{M^3} - \frac{6}{M^2} + \frac{3}{M} \right] + \right. \\
&+ \left. n^{-2} \sum_{j=1}^M \left[-\frac{6}{M^3} + \frac{12}{M^2} - \frac{7}{M} + 1 \right] \right] + O(n^{-3/2}) = \\
&= M-1 + \frac{1}{n} \left[\frac{\phi'''(1/M)}{2\phi''(1/M)} \left[\frac{2}{M} - 3 + M \right] + \right. \\
&+ \left. \frac{7\phi^{IV}(1/M)}{48\phi''(1/M)} \left[\frac{3}{M^2} - \frac{6}{M} + 3 \right] \right] + O(n^{-3/2}) =
\end{aligned}$$

$$= M-1 + \frac{1}{n} f_{\phi}^1 + O(n^{-3/2})$$

donde

$$f_{\phi}^1 = \frac{\phi'''(1/M)}{2\phi''(1/M)} \left[\frac{2}{M} - 3 + M \right] + \frac{7\phi^{IV}(1/M)}{16\phi''(1/M)} \left[\frac{1}{M^2} - \frac{2}{M} + 1 \right].$$

Elevando al cuadrado los dos miembros de (4.2.1) se tiene

$$\begin{aligned} \left[S_{\phi}(X/n, \pi_0) \right]^2 &= M^2 \sum_{i=1}^M W_i^4 + M^2 \sum_{i \neq j}^M W_i^2 W_j^2 + \frac{M^2 \phi'''(1/M)}{\sqrt{n} \phi''(1/M)} \left[\sum_{i=1}^M W_i^5 + \right. \\ &+ \left. \sum_{i \neq j}^M W_i^2 W_j^3 \right] + \frac{M^2 [\phi'''(1/M)]^2}{4n \phi''(1/M)} \left[\sum_{i=1}^M W_i^6 + \sum_{i \neq j}^M W_i^3 W_j^3 \right] + \\ &+ \frac{7M^2 \phi^{IV}(1/M)}{24n \phi''(1/M)} \left[\sum_{i=1}^M W_i^6 + \sum_{i \neq j}^M W_i^2 W_j^4 \right] + O_p(n^{-3/2}). \quad (4.2.2) \end{aligned}$$

y como para $i \neq j$

$$E \left[W_i^2 W_j^2 \right] = \frac{3}{M^4} - \frac{2}{M^3} + \frac{1}{M^2} + n^{-1} \left[-\frac{6}{M^4} + \frac{4}{M^3} - \frac{1}{M^2} \right],$$

$$E \left[W_i^2 W_j^3 \right] = n^{-1/2} \left[-\frac{20}{M^5} + \frac{20}{M^4} - \frac{7}{M^3} + \frac{1}{M^2} \right],$$

$$\begin{aligned} E \left[W_i^2 W_j^4 \right] &= -\frac{15}{M^6} + \frac{21}{M^5} - \frac{9}{M^4} + \frac{3}{M^3} + n^{-1} \left[\frac{130}{M^6} - \frac{182}{M^5} + \frac{83}{M^4} \right. \\ &\quad \left. - \frac{18}{M^3} + \frac{1}{M^2} \right] + O(n^{-2}), \end{aligned}$$

y

$$E \left[W_i^3 W_j^3 \right] = -\frac{15}{M^6} + \frac{18}{M^5} - \frac{9}{M^4} + n^{-1} \left[\frac{130}{M^6} - \frac{156}{M^5} + \frac{73}{M^4} - \right.$$

$$- \frac{12}{M^3} + \frac{1}{M^2} \Big] + O(n^{-2}),$$

el segundo momento vendrá dado por

$$\begin{aligned} E \left[\left[S_{\phi}(X/n, \pi_0) \right]^2 \right] &= M^2 \left[\frac{3}{M^3} - \frac{6}{M^2} + \frac{3}{M} + n^{-1} \left[-\frac{6}{M^3} + \frac{12}{M^2} - \frac{7}{M} + 1 \right] + \right. \\ &\quad + \frac{3(M-1)}{M^3} - \frac{2(M-1)}{M^2} + \frac{M-1}{M} + n^{-1} \left[-\frac{6(M-1)}{M^3} + \frac{4(M-1)}{M^2} - \right. \\ &\quad \left. \left. - \frac{M-1}{M} \right] \right] + \frac{M^2 \phi'''(1/M)}{\sqrt{n} \phi''(1/M)} \left[n^{-1/2} \left[-\frac{20}{M^4} + \frac{50}{M^3} - \frac{40}{M^2} + \frac{10}{M} - \right. \right. \\ &\quad \left. \left. - \frac{20(M-1)}{M^4} + \frac{20(M-1)}{M^3} - \frac{7(M-1)}{M^2} + \frac{M-1}{M} \right] \right] + \frac{M^2}{4n} \times \\ &\quad \times \left[\frac{\phi'''(1/M)}{\phi''(1/M)} \right]^2 \left[-\frac{15}{M^5} + \frac{45}{M^4} - \frac{45}{M^3} + \frac{15}{M^2} - \frac{15(M-1)}{M^5} + \right. \\ &\quad \left. + \frac{18(M-1)}{M^4} - \frac{9(M-1)}{M^3} \right] + \frac{7M^2 \phi^{IV}(1/M)}{24n \phi''(1/M)} \left[-\frac{15}{M^5} + \frac{45}{M^4} - \frac{45}{M^3} + \right. \\ &\quad \left. + \frac{15}{M^2} - \frac{15(M-1)}{M^5} + \frac{21(M-1)}{M^4} - \frac{9(M-1)}{M^3} + \frac{3(M-1)}{M^2} \right] + O(n^{-3/2}) = \\ &= M^2 - 1 + \frac{1}{n} \left[-2M + 2 + \left[\frac{10}{M} - 13 + 2M + M^2 \right] \frac{\phi'''(1/M)}{\phi''(1/M)} + \right. \\ &\quad + \frac{1}{4} \left[\frac{\phi'''(1/M)}{\phi''(1/M)} \right]^2 \left[\frac{12}{M^2} - \frac{18}{M} + 6 \right] + \frac{7 \phi^{IV}(1/M)}{24 \phi''(1/M)} \left[\frac{9}{M^2} - \frac{15}{M} + \right. \\ &\quad \left. \left. + 3 + 3M \right] \right] + O(n^{-3/2}) = M^2 - 1 + \frac{1}{n} f_{\phi}^2 + O(n^{-3/2}) \end{aligned}$$

donde

$$f_{\phi}^2 = -2M + 2 + \left[\frac{10}{M} - 13 + 2M + M^2 \right] \frac{\phi'''(1/M)}{\phi''(1/M)} + \frac{1}{4} \left[\frac{\phi'''(1/M)}{\phi''(1/M)} \right]^2 \times \\ \times \left[\frac{12}{M^2} - \frac{18}{M} + 6 \right] + \frac{7}{8} \frac{\phi^{IV}(1/M)}{\phi''(1/M)} \left[\frac{3}{M^2} - \frac{5}{M} + 1 + M \right].$$

Finalmente, multiplicando (4.2.2) por (4.2.1) se tiene

$$\left[S_{\phi}(X/n, \pi_0) \right]^3 = M^3 \sum_{i=1}^M W_i^6 + M^3 \sum_{i \neq j}^M W_i^4 W_j^2 + \frac{M^3 \phi'''(1/M)}{2\sqrt{n} \phi''(1/M)} \left[\sum_{i=1}^M W_i^7 + \right. \\ \left. + \sum_{i \neq j}^M W_i^4 W_j^3 \right] + \frac{7M^3 \phi^{IV}(1/M)}{48n \phi''(1/M)} \left[\sum_{i=1}^M W_i^8 + \sum_{i \neq j}^M W_i^4 W_j^4 \right] + \\ + M^3 \left[2 \sum_{i \neq j}^M W_i^4 W_j^2 + \sum_{i \neq j \neq k}^M W_i^2 W_j^2 W_k^2 \right] + \frac{M^3 \phi'''(1/M)}{2\sqrt{n} \phi''(1/M)} \times \\ \times \left[2 \sum_{i \neq j}^M W_i^5 W_j^2 + \sum_{i \neq j \neq k}^M W_i^2 W_j^2 W_k^3 \right] + \frac{7M^3 \phi^{IV}(1/M)}{48n \phi''(1/M)} \times \\ \times \left[2 \sum_{i \neq j}^M W_i^6 W_j^2 + \sum_{i \neq j \neq k}^M W_i^2 W_j^2 W_k^4 \right] + \frac{M^3 \phi'''(1/M)}{\sqrt{n} \phi''(1/M)} \left[\sum_{i=1}^M W_i^7 + \right. \\ \left. + \sum_{i \neq j}^M W_i^5 W_j^2 + \sum_{i \neq j}^M W_i^4 W_j^3 + \sum_{i \neq j}^M W_i^2 W_j^5 + \sum_{i \neq j \neq k}^M W_i^2 W_j^3 W_k^2 \right] + \\ + \frac{M^3}{2n} \left[\frac{\phi'''(1/M)}{\phi''(1/M)} \right]^2 \left[\sum_{i=1}^M W_i^8 + 2 \sum_{i \neq j}^M W_i^5 W_j^3 + \sum_{i \neq j}^M W_i^2 W_j^6 + \right. \\ \left. + \sum_{i \neq j \neq k}^M W_i^2 W_j^3 W_k^3 \right] + \frac{M^3}{4n} \left[\frac{\phi'''(1/M)}{\phi''(1/M)} \right]^2 \left[\sum_{i=1}^M W_i^8 + \sum_{i \neq j}^M W_i^6 W_j^2 + \right. \\ \left. + \sum_{i \neq j \neq k}^M W_i^2 W_j^3 W_k^3 \right]$$

$$\begin{aligned}
& + 2 \sum_{i \neq j}^M W_i^5 W_j^3 + \sum_{i \neq j \neq k}^M W_i^3 W_j^3 W_k^2 \Big] + \frac{7M^3 \phi^{IV}(1/M)}{24 n \phi''(1/M)} \left[\sum_{i=1}^M W_i^8 + \right. \\
& \left. + \sum_{i \neq j}^M W_i^6 W_j^2 + \sum_{i \neq j}^M W_i^4 W_j^4 + \sum_{i \neq j}^M W_i^2 W_j^6 + \sum_{i \neq j \neq k}^M W_i^2 W_j^4 W_k^2 \right] + O_p(n^{-3/2}).
\end{aligned}$$

con lo cual, como para $i \neq j \neq k$

$$E[W_i^4 W_j^3] = n^{-1/2} \left[\frac{210}{M^7} - \frac{315}{M^6} + \frac{183}{M^5} - \frac{45}{M^4} + \frac{3}{M^3} \right] + O(n^{-3/2}),$$

$$E[W_i^4 W_j^4] = \frac{105}{M^8} - \frac{180}{M^7} + \frac{126}{M^6} - \frac{36}{M^5} + \frac{9}{M^4} + O(n^{-1}),$$

$$E[W_i^5 W_j^2] = n^{-1/2} \left[\frac{210}{M^7} - \frac{385}{M^6} + \frac{230}{M^5} - \frac{65}{M^4} + \frac{10}{M^3} \right] + O(n^{-3/2}),$$

$$E[W_i^5 W_j^3] = \frac{105}{M^8} - \frac{195}{M^7} + \frac{135}{M^6} - \frac{45}{M^5} + O(n^{-1}),$$

$$E[W_i^6 W_j^2] = \frac{105}{M^8} - \frac{240}{M^7} + \frac{180}{M^6} - \frac{60}{M^5} + \frac{15}{M^4} + O(n^{-1}),$$

$$\begin{aligned}
E[W_i^2 W_j^2 W_k^2] &= -\frac{15}{M^6} + \frac{9}{M^5} - \frac{3}{M^4} + \frac{1}{M^3} + n^{-1} \left[\frac{130}{M^6} - \frac{78}{M^5} + \right. \\
&\quad \left. + \frac{21}{M^4} - \frac{3}{M^3} \right] + O(n^{-2}),
\end{aligned}$$

$$E[W_i^2 W_j^2 W_k^3] = n^{-1/2} \left[\frac{210}{M^7} - \frac{175}{M^6} + \frac{59}{M^5} - \frac{11}{M^4} + \frac{1}{M^3} \right] + O(n^{-3/2}),$$

$$E[W_i^2 W_j^3 W_k^3] = \frac{105}{M^8} - \frac{105}{M^7} + \frac{45}{M^6} - \frac{9}{M^5} + O(n^{-1}),$$

y

$$E[W_i^2 W_j^2 W_k^4] = \frac{105}{M^8} - \frac{120}{M^7} + \frac{48}{M^6} - \frac{12}{M^5} + \frac{3}{M^4} + O(n^{-1}),$$

se tiene que el tercer momento vendrá dado por

$$\begin{aligned}
E\left[\left[S_{\phi}(X/n, \pi_0)\right]^3\right] &= M^3 \left[-\frac{15}{M^5} + \frac{45}{M^4} - \frac{45}{M^3} + \frac{15}{M^2} + n^{-1} \left(\frac{130}{M^5} - \frac{390}{M^4} + \right. \right. \\
&+ \left. \frac{415}{M^3} - \frac{180}{M^2} + \frac{25}{M} \right) - \frac{45(M-1)}{M^5} + \frac{63(M-1)}{M^4} - \frac{27(M-1)}{M^3} + \\
&+ \frac{9(M-1)}{M^2} + n^{-1} \left(\frac{390(M-1)}{M^5} - \frac{546(M-1)}{M^4} + \frac{249(M-1)}{M^3} - \right. \\
&- \left. \frac{54(M-1)}{M^2} + \frac{3(M-1)}{M} \right) - \frac{15(M-1)(M-2)}{M^5} + \frac{9(M-1)(M-2)}{M^4} - \\
&- \frac{3(M-1)(M-2)}{M^3} + \frac{(M-1)(M-2)}{M^2} + n^{-1} \left(\frac{130(M-1)(M-2)}{M^5} - \right. \\
&- \left. \frac{78(M-1)(M-2)}{M^4} + \frac{21(M-1)(M-2)}{M^3} - \frac{3(M-1)(M-2)}{M^2} \right) \Bigg] + \\
&+ \frac{M^3 \phi'''(1/M)}{2n\phi''(1/M)} \left[\frac{630}{M^6} - \frac{2205}{M^5} + \frac{2835}{M^4} - \frac{1575}{M^3} + \frac{315}{M^2} + \right. \\
&+ \frac{630(M-1)}{M^6} - \frac{945(M-1)}{M^5} + \frac{549(M-1)}{M^4} - \frac{135(M-1)}{M^3} + \\
&+ \frac{9(M-1)}{M^2} + \frac{1260(M-1)}{M^6} - \frac{2310(M-1)}{M^5} + \frac{1380(M-1)}{M^4} - \\
&- \frac{390(M-1)}{M^3} + \frac{60(M-1)}{M^2} + \frac{630(M^2-3M+2)}{M^6} - \frac{525(M^2-3M+2)}{M^5} + \\
&+ \left. \frac{177(M^2-3M+2)}{M^4} - \frac{33(M^2-3M+2)}{M^3} + \frac{3(M^2-3M+2)}{M^2} \right] + \\
&+ \frac{7M^3 \phi^{IV}(1/M)}{48n\phi''(1/M)} \left[\frac{315}{M^7} - \frac{1260}{M^6} + \frac{1890}{M^5} - \frac{1260}{M^4} + \frac{315}{M^3} + \right. \\
&+ \left. \frac{315(M-1)}{M^7} - \frac{540(M-1)}{M^6} + \frac{378(M-1)}{M^5} - \frac{108(M-1)}{M^4} + \right.
\end{aligned}$$

$$\begin{aligned}
& + \frac{27(M-1)}{M^3} + \frac{630(M-1)}{M^7} - \frac{1440(M-1)}{M^6} + \frac{1080(M-1)}{M^5} - \\
& - \frac{360(M-1)}{M^4} + \frac{90(M-1)}{M^3} + \frac{315(M^2-3M+2)}{M^7} - \frac{360(M^2-3M+2)}{M^6} + \\
& + \left. \frac{144(M^2-3M+2)}{M^5} - \frac{36(M^2-3M+2)}{M^4} + \frac{9(M^2-3M+2)}{M^3} \right) + \\
& + \frac{M^3}{4n} \left[\frac{\phi'''(1/M)}{\phi''(1/M)} \right]^2 \left[\frac{315}{M^7} - \frac{1260}{M^6} + \frac{1890}{M^5} - \frac{1260}{M^4} + \frac{315}{M^3} + \right. \\
& + \frac{630(M-1)}{M^7} - \frac{1170(M-1)}{M^6} + \frac{810(M-1)}{M^5} - \frac{270(M-1)}{M^4} + \\
& + \frac{315(M-1)}{M^7} - \frac{720(M-1)}{M^6} + \frac{540(M-1)}{M^5} - \frac{180(M-1)}{M^4} + \\
& + \frac{45(M-1)}{M^3} + \frac{315(M^2-3M+2)}{M^7} - \frac{315(M^2-3M+2)}{M^6} + \\
& + \left. \frac{135(M^2-3M+2)}{M^5} - \frac{27(M^2-3M+2)}{M^4} \right] + O(n^{-3/2}) = \\
& = -3-M+3M^2+M^3 + \frac{1}{n} \left[26-24M-2M^2 + \frac{\phi'''(1/M)}{2\phi''(1/M)} \left(\frac{210}{M} - \right. \right. \\
& - 243 + 3M + 27M^2 + 3M^3 \left. \right) + \frac{7\phi^{IV}(1/M)}{48n\phi''(1/M)} \left(\frac{135}{M^2} - \frac{198}{M} + \right. \\
& + 54M + 9M^2 \left. \right) + \frac{1}{4} \left(\frac{\phi'''(1/M)}{\phi''(1/M)} \right)^2 \left(\frac{180}{M^2} - \frac{234}{M} + 36 + 18M \right) + \\
& + O(n^{-3/2}) = M^3+3M^2-M-3 + \frac{1}{n}f_\phi^3 + O(n^{-3/2})
\end{aligned}$$

donde

$$f_{\phi}^3 = 26 - 24M - 2M^2 + \frac{\phi'''(1/M)}{2\phi''(1/M)} \left(\frac{210}{M} - 243 + 3M + 27M^2 + 3M^3 \right) +$$

$$+ \frac{7\phi^{IV}(1/M)}{16\phi''(1/M)} \left(\frac{45}{M^2} - \frac{66}{M} + 18M + 3M^2 \right) +$$

$$+ \frac{1}{4} \left[\frac{\phi'''(1/M)}{\phi''(1/M)} \right]^2 \left(\frac{180}{M^2} - \frac{234}{M} + 36 + 18M \right).$$

Luego, las funciones ϕ que hacen que se aproximen más los momentos asintóticos a los exactos para M fijo son aquellas para las cuales se verifica que $f_{\phi}^i = 0$, $i=1,2,3$ ya que los desarrollos de segundo orden de los tres primeros momentos de S_{ϕ} son iguales a los momentos de una χ_{M-1}^2 más el factor de corrección de orden $O(n^{-1})$, f_{ϕ}^i , $i=1,2,3$, respectivamente.

En definitiva se puede establecer el siguiente resultado:

Teorema 4.2.1

Sea $\phi:(0,\infty) \rightarrow \mathbb{R}$ cóncava con derivada cuarta continua y $\phi''(1/M) < 0$. El contraste de bondad de ajuste basado en el estadístico $S_{\phi}(X/n, \pi_0)$ es óptimo de acuerdo al criterio de los momentos si y sólo si $f_{\phi}^i = 0$, $i=1,2,3$.

El Teorema 4.2.2 es muy interesante ya que los factores de corrección de orden $O(n^{-1})$ de los tres primeros momentos se anulan para las mismas raíces cuando se considera la familia $\phi_{\alpha}(x) = \frac{1}{1-\alpha}(x^{\alpha} - x)$ si se hace $M \rightarrow \infty$. Una de las raíces es $\alpha = 2$ que se corresponde con el estadístico X^2 de Pearson y la otra es $\alpha = 13/7$ que no se

corresponde con ningún estadístico conocido lo que proporciona una excelente alternativa al estadístico clásico.

Teorema 4.2.2

El contraste de bondad de ajuste basado en el estadístico $S_{\phi_\alpha}(X/n, \pi_0)$ es óptimo de acuerdo al criterio de los momentos cuando $M \rightarrow \infty$ si y sólo si $\alpha = 2$ ó $\alpha = 13/7$.

Demostación

Si se resuelve la primera ecuación del Teorema 4.2.1

$$f_{\phi_\alpha}^1 = \frac{7}{48}(3-6M+3M^2)\alpha^2 + \left[-\frac{81}{48}M^2 + \frac{138}{48}M - \frac{57}{48}\right]\alpha + \left[\frac{13}{8}M^2 - \frac{18}{8}M + \frac{5}{8}\right] = 0$$

y se hace $M \rightarrow \infty$ se tiene que

$$\alpha = \frac{81 \pm \sqrt{81^2 - 6552}}{42},$$

por tanto las soluciones de la ecuación son $\alpha=2$ y $\alpha=13/7$.

Por otra parte, si se resuelve la ecuación

$$f_{\phi_\alpha}^2 = (45-71M+19M^2+7M^3)\alpha^2 + \left[-27M^3-67M^2+215M-121\right]\alpha + \left[26M^3+58M^2-162M+78\right] = 0$$

y se hace $M \rightarrow \infty$ se obtiene

$$\alpha = \frac{27 \pm \sqrt{27^2 - 728}}{14}$$

es decir, $\alpha=2$ y $\alpha=13/7$.

Por último, para aproximar el tercer momento asintótico al exacto debe ser

$$f_{\phi\alpha}^3 = (1035-1398M+144M^2+198M^3+21M^4)\alpha^2 + \left[-81M^4-702M^3-552M^2+4110M-2775 \right] \alpha + \left[78M^4+612M^3+496M^2-3012M+1826 \right] = 0$$

donde si se hace $M \rightarrow \infty$ se obtiene de nuevo que

$$\alpha = \frac{81 \pm \sqrt{81^2 - 6552}}{42}$$

con lo cual $\alpha=2$ y $\alpha=13/7$.

■

El resultado anterior es para M grande, en el caso que M sea pequeño se debería utilizar la Tabla 4.2.1 que contiene las raíces α de las ecuaciones $f_{\phi\alpha}^i = 0$, $i=1,2,3$ para valores de M fijos que crecen hacia ∞ . En esta tabla se observa que para $M > 20$ se puede utilizar el resultado anterior ya que los factores de primer orden de los tres primeros momentos son muy próximos a 0 para $\alpha = 2$ y $\alpha = 13/7$. Para $M \leq 20$, en particular para $M=4, 5$ ó 10 sería razonable elegir un estadístico $S_{\phi\alpha}$ con $\alpha \in [1.5, 2]$.

Tabla 4.2.1

Valores de las raíces ($\alpha_1 > \alpha_2$) de $f_{\phi\alpha}^i = 0$, $i=1,2,3$

M	2	3	4	5	10	20	40	50	100	200	500
$f_{\phi\alpha}^1$ {	α_1	3.0	2.42	2.23	2.14	2.0	2.0	2.0	2.0	2.0	2.0
	α_2	2.0	2.0	2.0	2.0	1.98	1.91	1.88	1.88	1.86	1.86
$f_{\phi\alpha}^2$ {	α_1	3.34	2.52	2.31	2.21	2.07	2.02	2.0	2.0	2.0	2.0
	α_2	1.65	1.68	1.7	1.71	1.76	1.8	1.83	1.83	1.84	1.85
$f_{\phi\alpha}^3$ {	α_1	3.69	2.62	2.37	2.27	2.10	2.04	2.01	2.01	2.0	2.0
	α_2	1.30	1.41	1.47	1.51	1.62	1.72	1.78	1.79	1.82	1.84

Por otra parte, utilizando estos resultados es posible definir un estadístico corregido $S_{\phi}^*(X/n, \pi_0)$ de tal forma que su distribución esté más próxima a una χ_{M-1}^2 . Es decir, si

$$\mu = E\left[\chi_{M-1}^2\right] = M-1$$

y

$$\sigma^2 = V\left[\chi_{M-1}^2\right] = 2(M-1),$$

como se sabe que

$$E[S_{\phi}(X/n, \pi_0)] = \mu + a_{\phi}/n + o(n^{-1})$$

y

$$V[S_{\phi}(X/n, \pi_0)] = \sigma^2 + b_{\phi}/n + o(n^{-1})$$

donde

$$a_{\phi} = \frac{\phi'''(1/M)}{2\phi''(1/M)} \left[\frac{2}{M} - 3 + M \right] + \frac{7\phi^{IV}(1/M)}{16\phi''(1/M)} \left[\frac{1}{M^2} - \frac{2}{M} + 1 \right]$$

y

$$b_{\phi} = -2M + 2 + \frac{\phi'''(1/M)}{\phi''(1/M)} \left[\frac{12}{M} - 18 + 6M \right] + \frac{1}{2} \left[\frac{\phi'''(1/M)}{\phi''(1/M)} \right]^2 \left[\frac{6}{M^2} - \frac{9}{M} + 3 \right] + \frac{7\phi^{IV}(1/M)}{2\phi''(1/M)} \left[\frac{1}{M^2} - \frac{2}{M} + 1 \right]$$

se puede definir

$$S_{\phi}^*(X/n, \pi_0) = \frac{S_{\phi}(X/n, \pi_0) - \gamma_{\phi}}{\sqrt{\delta_{\phi}}}$$

de tal forma que

$$E[S_{\phi}^*(X/n, \pi_0)] = \mu + o(n^{-1})$$

y

$$V[S_{\phi}^*(X/n, \pi_0)] = \sigma^2 + o(n^{-1}).$$

Para ello es necesario considerar

$$\gamma_{\phi} = \mu \left[1 - \sqrt{\delta_{\phi}} \right] + a_{\phi}/n$$

y

$$\delta_{\phi} = 1 + b_{\phi}/n\sigma^2$$

es decir,

$$\delta_{\phi} = 1 - \frac{1}{n} + \frac{1}{n(M-1)} \left[\frac{\phi'''(1/M)}{\phi''(1/M)} \left(\frac{6}{M} - 9 + 3M \right) + \frac{1}{4} \left(\frac{\phi'''(1/M)}{\phi''(1/M)} \right)^2 \left(\frac{6}{M^2} - \frac{9}{M} + 3 \right) + \frac{7\phi^{IV}(1/M)}{4\phi''(1/M)} \left(\frac{1}{M^2} - \frac{2}{M} + 1 \right) \right]$$

y

$$\gamma_{\phi} = (M-1) \left[1 - \sqrt{\delta_{\phi}} \right] + \frac{1}{2n} \left[\frac{\phi'''(1/M)}{\phi''(1/M)} \left(\frac{2}{M} - 3 + M \right) + \frac{7\phi^{IV}(1/M)}{8\phi''(1/M)} \left(\frac{1}{M^2} - \frac{2}{M} + 1 \right) \right]. \quad (4.2.3)$$

A continuación, se calculan los valores de α que hacen que el estadístico modificado $S_{\phi_{\alpha}}^*$ sea igual a $S_{\phi_{\alpha}}$ cuando $M \rightarrow \infty$. Estos estadísticos son iguales cuando $\gamma_{\phi_{\alpha}} = 0$ y $\delta_{\phi_{\alpha}} = 1$ o lo que es lo mismo

$$a_{\phi_{\alpha}} = b_{\phi_{\alpha}} = 0.$$

Por una parte, $a_{\phi\alpha} \approx 0$ si y sólo si $\alpha = 2$ y $\alpha = 13/7$, como ya se ha demostrado.

Por otra parte, $b_{\phi\alpha} \approx 0$ si y sólo si

$$(10M^2 - 23M + 13)\alpha^2 + (-35M^2 + 70M - 35)\alpha + 30M^2 - 52M + 22 = 0$$

donde haciendo $M \rightarrow \infty$ se tiene

$$\alpha = \frac{35 \pm \sqrt{35^2 - 1200}}{20}$$

por tanto las soluciones de la ecuación son $\alpha=2$ y $\alpha=1.5$. De esta forma cuando $M \rightarrow \infty$ ambos términos de corrección se anulan para $\alpha = 2$, es decir, la distribución asintótica ji-cuadrado del estadístico X^2 de Pearson no se puede mejorar con el criterio de los momentos.

Este resultado es para M grande, por ello se calculan las raíces de las ecuaciones $a_{\phi\alpha}$ y $b_{\phi\alpha}$ para valores fijos de M creciendo a ∞ que aparecen en la Tabla 4.2.2. Como era de esperar, para $M \geq 20$, los términos de corrección se aproximan a 0 para $\alpha=2$. Además, para $M=5$ ó 10 los estadísticos S_{ϕ_2} y $S_{\phi_2}^*$ estarían razonablemente próximos.

Tabla 4.2.2
Valores de las raíces ($\alpha_1 > \alpha_2$) de $a_{\phi}=0$ y $b_{\phi}=0$

M	2	3	4	5	10	20	40	50	100	200	500
a_{ϕ}	α_1	3.0	2.42	2.23	2.14	2.0	2.0	2.0	2.0	2.0	2.0
	α_2	2.0	2.0	2.0	2.0	1.98	1.91	1.88	1.88	1.86	1.86
b_{ϕ}	α_1	3.40	2.54	2.33	2.23	2.09	2.04	2.02	2.01	2.0	2.0
	α_2	1.59	1.57	1.55	1.54	1.52	1.51	1.5	1.5	1.5	1.5

Al igual que se ha estudiado la rapidez con que los momentos exactos de S_ϕ se aproximan a los momentos asintóticos bajo la hipótesis (3.2.1), se puede estudiar la rapidez bajo las alternativas (3.2.3) considerando el desarrollo asintótico de segundo orden de estos momentos.

Sea

$$V_i = n^{1/2}(X_i/n - \pi_i)$$

donde $X=(X_1, \dots, X_M)$ es una variable aleatoria multinomial de parámetros n y $\pi=(\pi_1, \dots, \pi_M)$, con $\pi_i = \frac{1}{M} + n^{-1/2}c_i$, entonces se verifica que

$$W_i = V_j + c_i.$$

Además, de (4.2.1) se tiene que

$$E[S_\phi(X/n, \pi_0)] = M \sum_{j=1}^M E[W_j^2] + \frac{M\phi'''(1/M)}{2\sqrt{n}\phi''(1/M)} \sum_{j=1}^M E[W_j^3] + O(n^{-1}). \quad (4.2.4)$$

Como

$$E[W_j^2] = E[(V_j + c_j)^2] = \frac{1}{M} - \frac{1}{M^2} + c_j^2 + n^{-1/2} \left[c_j - \frac{2c_j}{M} \right] + O(n^{-1})$$

y

$$E[W_j^3] = E[(V_j + c_j)^3] = c_j^3 + \frac{3c_j}{M} - \frac{3c_j}{M^2} + O(n^{-1/2})$$

sustituyendo en (4.2.4) tenemos que el primer momento viene dado por

$$\begin{aligned}
 E[S_{\phi}(X/n, \pi_0)] &= M^{-1} + M \sum_{j=1}^M c_j^2 + \frac{1}{\sqrt{n}} \left\{ M \sum_{j=1}^M c_j - 2 \sum_{j=1}^M c_j + \frac{M\phi'''(1/M)}{2\phi''(1/M)} \times \right. \\
 &\quad \times \left. \left[\sum_{j=1}^M c_j^3 + \frac{3}{M} \sum_{j=1}^M c_j - \frac{3}{M^2} \sum_{j=1}^M c_j \right] \right\} + O(n^{-1}) = M^{-1} + \\
 &\quad + M \sum_{j=1}^M c_j^2 + \frac{1}{\sqrt{n}} \left\{ \frac{M\phi'''(1/M)}{2\phi''(1/M)} \sum_{j=1}^M c_j^3 \right\} + O(n^{-1}). \quad (4.2.5)
 \end{aligned}$$

Por otra parte, de (4.2.2) se tiene que

$$\begin{aligned}
 E\left[\left[S_{\phi}(X/n, \pi_0) \right]^2 \right] &= M^2 \sum_{i=1}^M E[W_i^4] + M^2 \sum_{i \neq j}^M E[W_i^2 W_j^2] + \frac{M^2 \phi'''(1/M)}{\sqrt{n} \phi''(1/M)} \times \\
 &\quad + \left[\sum_{i=1}^M E[W_i^5] + \sum_{i \neq j}^M E[W_i^2 W_j^3] \right] + O(n^{-1}).
 \end{aligned}$$

Al ser

$$\begin{aligned}
 E[W_j^4] &= \frac{3}{M^4} - \frac{6}{M^3} + \frac{3}{M^2} + \frac{6c_j^2}{M} - \frac{6c_j^2}{M^2} + c_j^4 + \\
 &\quad + n^{-1/2} \left[\frac{20c_j}{M^3} - \frac{30c_j}{M^2} + \frac{10c_j}{M} + 6c_j^3 - \frac{12c_j^3}{M} \right] \\
 E[W_i^2 W_j^2] &= \frac{3}{M^4} - \frac{2}{M^3} + \frac{1}{M^2} - \frac{4c_i c_j}{M^2} + c_i^2 c_j^2 + \left(\frac{1}{M} - \frac{1}{M^2} \right) c_i^2 + \\
 &\quad + \left(\frac{1}{M} - \frac{1}{M^2} \right) c_j^2 + n^{-1/2} \left[\frac{6c_j}{M^3} + \frac{6c_i}{M^3} - \frac{3c_j}{M^2} - \frac{3c_i}{M^2} + \right.
 \end{aligned}$$

$$+ \frac{c_i}{M} + \frac{c_j}{M} + 2\left(\frac{2}{M^3} - \frac{1}{M^2}\right)c_i + \left[c_j - \frac{2c_j}{M}\right]c_i^2 + 2\left(\frac{2}{M^3} - \frac{1}{M^2}\right)c_j - 4\left(\frac{c_j}{M} + \frac{c_i}{M}\right)c_i c_j + \left[c_i - \frac{2c_i}{M}\right]c_j^2$$

$$E[W_j^5] = 5\left(\frac{3}{M^4} - \frac{6}{M^3} + \frac{3}{M^2}\right)c_j + 10\left(\frac{1}{M} - \frac{1}{M^2}\right)c_j^3 + c_j^5 + O(n^{-1/2}),$$

y

$$E[W_i^2 W_j^3] = 6\left(\frac{1}{M^4} - \frac{1}{M^3}\right)c_i + 3\left(\frac{3}{M^4} - \frac{2}{M^3} + \frac{1}{M^2}\right)c_j + 3\left(\frac{1}{M} - \frac{1}{M^2}\right)c_i^2 c_j - \frac{6c_i c_j^2}{M^2} + \left(\frac{1}{M} - \frac{1}{M^2}\right)c_j^3 + c_i^2 c_j^3 + O(n^{-1/2})$$

se obtiene que el segundo momento viene dado por

$$E\left[\left[S_\phi(X/n, \pi_0)\right]^2\right] = \frac{3}{M} - 6 + 3M + 6M \sum_{j=1}^M c_j^2 - 6 \sum_{j=1}^M c_j^2 + M^2 \sum_{j=1}^M c_j^4 + n^{-1/2} \left[6M^2 \sum_{j=1}^M c_j^3 - 12M \sum_{j=1}^M c_j^3 \right] + \frac{3(M-1)}{M} - 2(M-1) + (M-1)M - 4 \sum_{i \neq j}^M c_i c_j + M^2 \sum_{i \neq j}^M c_i^2 c_j^2 + 2(M-1)^2 \sum_{j=1}^M c_j^2 + n^{-1/2} \left[6 \sum_{i \neq j}^M \frac{c_j}{M} - 6 \sum_{i \neq j}^M \frac{c_i}{M} - 3 \sum_{i \neq j}^M c_j - 3 \sum_{i \neq j}^M c_i + M \sum_{i \neq j}^M c_j + M \sum_{i \neq j}^M c_i + \frac{4}{M} \sum_{i \neq j}^M c_i - 2 \sum_{i \neq j}^M c_i + M^2 \sum_{i \neq j}^M c_i^2 c_j - 2M \sum_{i \neq j}^M c_i^2 c_j + \frac{4}{M} \sum_{i \neq j}^M c_j - 2 \sum_{i \neq j}^M c_j \right]$$

$$\begin{aligned}
& - 4M \sum_{i \neq j}^M c_i c_j^2 - 4M \sum_{i \neq j}^M c_i^2 c_j + M^2 \sum_{i \neq j}^M c_i c_j^2 - 2M \sum_{i \neq j}^M c_i c_j^2 + \\
& + \frac{\phi'''(1/M)}{\sqrt{n} \phi''(1/M)} \left[10(M-1) \sum_{j=1}^M c_j^3 + M^2 \sum_{j=1}^M c_j^5 + 3(M-3) \sum_{i \neq j}^M c_i^2 c_j + \right. \\
& + (M-1) \sum_{i \neq j}^M c_j^3 + M^2 \sum_{i \neq j}^M c_i^2 c_j^3 \left. \right] + O(n^{-1}) = M^2 - 1 + \\
& + 2M(M+1) \sum_{j=1}^M c_j^2 + M^2 \left[\sum_{j=1}^M c_j^2 \right]^2 + n^{-1/2} \left[4M^2 \sum_{j=1}^M c_j^3 + \right. \\
& + \left. \frac{\phi'''(1/M)}{\phi''(1/M)} \left[(M^2 + 5M) \sum_{j=1}^M c_j^3 + M^2 \sum_{j=1}^M c_j^2 \sum_{j=1}^M c_j^3 \right] \right] + O(n^{-1}).
\end{aligned} \tag{4.2.6}$$

Por tanto, los desarrollos de segundo orden de los dos primeros momentos bajo (3.2.3), dados en (4.2.5) y (4.2.6), son iguales a los momentos de una $\chi_{M-1}^2(\delta)$ mas unos términos de corrección $O(n^{-1/2})$. En consecuencia se elegirá ϕ de forma que estos términos se anulen con lo que se conseguirá que sea mayor la velocidad de convergencia de la distribución asintótica a la exacta.

Si se particulariza la expresión (4.2.5) del primer momento para la familia $\phi_\alpha(x)$ y se iguala a cero el término $O(n^{-1/2})$ se tiene

$$-\frac{M^2}{2}(\alpha-2) \sum_{j=1}^M c_j^3 = 0,$$

es decir, $\alpha=2$.

En cuanto al segundo momento, se obtiene

$$M(\alpha-2) = \frac{-4M^2 \sum_{i=1}^M c_i^3}{(M^2+5M) \sum_{i=1}^M c_i^3 + M^2 \sum_{i=1}^M c_i^2 \sum_{i=1}^M c_i^3}$$

de donde

$$\alpha = 2 - \frac{4}{M+5+M \sum_{i=1}^M c_i^2}$$

por tanto $1 < \alpha \leq 2$ y cuando $M \rightarrow \infty$, $\alpha \rightarrow 2$.

De esta forma eligiendo α próximo a 2, es decir, seleccionado el estadístico X^2 de Pearson, se asegura la proximidad a cero de los factores de corrección de los dos primeros momentos bajo (3.2.3). Por tanto, el estadístico X^2 de Pearson ($\alpha=2$) dentro de los estadísticos S_{ϕ_α} es el único que no se puede corregir con el criterio de los momentos para mejorar la distribución asintótica ji-cuadrado tanto bajo la hipótesis nula simétrica como las alternativas consideradas.

4.3.- Desarrollo asintótico de segundo orden de la distribución de los estadísticos R_ϕ -divergencia

En el apartado anterior se utiliza que

$$P(S_\phi(X/n, \pi_0) < c) = P(\chi_{M-1}^2 < c) + o(1) \quad (4.3.1)$$

para comparar los desarrollos asintóticos de segundo orden de los tres primeros momentos de $S_\phi(X/n, \pi_0)$ con los tres primeros momentos de una χ_{M-1}^2 .

En este apartado se va a extraer la componente de segundo orden del término $o(1)$ de la distribución (4.3.1) para obtener una mejor aproximación de la distribución exacta de $S_\phi(X/n, \pi_0)$.

Sea $W_j = \sqrt{n}(X_j/n - \pi_{0j})$, con $\pi_{0j} = 1/M$, $j=1, \dots, M$ y considérese $W = (W_1, \dots, W_r)^t$ donde $r=M-1$. Entonces, W es un vector aleatorio que toma valores en el conjunto

$$L = \{w = (w_1, \dots, w_r)^t / w = \sqrt{n}(x/n - \pi_0^*) \text{ y } x \in K\}, \quad (4.3.2)$$

donde

$$\pi_0^* = (\pi_{01}, \dots, \pi_{0r})^t$$

y

$$K = \{x = (x_1, \dots, x_r)^t / x_j \geq 0 \text{ entero, } j=1, \dots, r; \sum_{j=1}^r x_j \leq n\}.$$

La distribución de probabilidad del vector W (Siotani y Fujikoshi (1984)), se puede expresar en los siguientes términos

$$P(W=w) = n^{-r/2} \varphi(w) \{ 1 + n^{-1/2} h_1(w) + n^{-1} h_2(w) + O(n^{-3/2}) \} \quad (4.3.3)$$

donde

$$\varphi(w) = (2\pi)^{-r/2} |\Omega|^{-1/2} \exp\left[-\frac{1}{2} w^t \Omega^{-1} w\right]$$

es la función de densidad de la normal multivariante, y

$$h_1(w) = -\frac{1}{2} \sum_{j=1}^M \frac{w_j}{\pi_{0j}} + \frac{1}{6} \sum_{j=1}^M \frac{w_j^3}{\pi_{0j}^2}, \quad (4.3.4)$$

$$h_2(w) = \frac{1}{2} \left[h_1(w) \right]^2 + \frac{1}{12} \left[1 - \sum_{j=1}^M \frac{1}{\pi_{0j}} \right] + \frac{1}{4} \sum_{j=1}^M \frac{w_j^2}{\pi_{0j}^2} - \frac{1}{12} \sum_{j=1}^M \frac{w_j^4}{\pi_{0j}^3}$$

con

$$w_M = - \sum_{j=1}^r w_j, \quad \Omega = \text{diag}(\pi_0^*) - \pi_0^* \pi_0^{*t}.$$

Este resultado proporciona el desarrollo de Edgeworth de la probabilidad de W en un punto $w \in L$. Si la función de distribución de W es continua se tiene que

$$P(W \in B) = \int_B \dots \int \varphi(w) \{1 + n^{-1/2} h_1(w) + n^{-1} h_2(w)\} dw + O(n^{-3/2}).$$

Sin embargo, Yarnold (1972) demostró que este desarrollo no es válido si la distribución de W es discreta. Rao (1961) calcula la expresión de la $P(W \in B)$ en el caso que W no sea continua cuando B es un conjunto de Borel, como una integral de Stieltjes cuyo cálculo resulta complicado. Yarnold obtuvo, en el caso de que B sea un conjunto convexo extendido, es decir, si B se puede representar como

$$B = \{ w = (w_1, \dots, w_r)^t / \gamma_s(w) < w_s < \theta_s(w^*), \quad w^* = (w_1, \dots, w_{s-1}, w_{s+1}, \dots, w_r)^t \in B_s \} \quad (4.3.5)$$

donde $B_s \subset R^{r-1}$ y γ_s, θ_s son funciones continuas en R^{r-1} , $s=1, \dots, r$, el siguiente desarrollo alternativo

$$P(W \in B) = J_1 + J_2 + J_3 + O(n^{-3/2})$$

donde

$$J_1 = \int_B \dots \int \varphi(w) \{1 + n^{-1/2} h_1(w) + n^{-1} h_2(w)\} dw,$$

$$J_2 = -n^{-1/2} \sum_{s=1}^r n^{-(r-s)/2} \sum_{w_{s+1} \in L_{s+1}} \dots \sum_{w_r \in L_r} \int_{B_s} \dots \int$$

$$\left[S_1(\sqrt{n} w_s + n\pi_{0s}) \varphi(w) \right]_{\gamma_s(w^*)}^{\theta_s(w^*)} dw_1 \dots dw_{s-1},$$

$$J_3 = O(n^{-1}),$$

con h_1 y h_2 como en (4.3.4),

$$L_j = \{ w_j / w_j = \sqrt{n}(x_j/n - \pi_{0j}) \text{ y } x_j \text{ entero} \},$$

$$S_1(t) = t - [t] - \frac{1}{2},$$

$$\theta_s(w^*) \text{ y } \gamma_s(w^*) \text{ como en (4.3.5)}$$

y

$$\begin{aligned} h(w) \frac{\theta_s(w^*)}{\gamma_s(w^*)} &= h(w_1, \dots, w_{s-1}, \theta_s(w^*), w_{s+1}, \dots, w_r) - \\ &- h(w_1, \dots, w_{s-1}, \gamma_s(w^*), w_{s+1}, \dots, w_r). \end{aligned}$$

La distribución de $S_\phi(X/n, \pi_0)$ bajo la hipótesis nula simétrica, se puede expresar como

$$P(S_\phi(X/n, \pi_0) < c) = P(W \in B_\phi(c))$$

donde

$$B_\phi(c) = \{ w = (w_1, \dots, w_r)' / S_\phi((x/n, x_M/n), \pi_0) < c \}$$

siendo

$$w_M = - \sum_{j=1}^r w_j, \quad x = \sqrt{n} w + n\pi_0^* \text{ y } x_M = \sqrt{n} w_M + n/M.$$

Es inmediato comprobar que, $B_\phi(c)$ es un conjunto convexo extendido donde $\gamma_s(w^*)$ y $\theta_s(w^*)$ se eligen de forma que si $w_s = \gamma_s(w^*)$ ó $w_s = \theta_s(w^*)$, $s=1, \dots, r$, entonces $S_\phi((x/n, x_M/n), \pi_0) = c$. En consecuencia, utilizando el resultado de Yarnold (1972) con $B = B_\phi(c)$ se obtiene el desarrollo de segundo orden para la distribución de S_ϕ en el siguiente teorema.

Teorema 4.3.1

Sea $\phi:(0, \infty) \rightarrow \mathbb{R}$ una función cóncava con derivada segunda continua y $\phi''(1/M) < 0$. La distribución del estadístico $S_\phi(X/n, \pi_0)$ se puede expresar como

$$P(S_\phi(X/n, \pi_0) < c) = J_1^\phi + J_2^\phi + J_3^\phi + O(n^{-3/2})$$

donde J_1^ϕ , J_2^ϕ y J_3^ϕ se obtienen a partir de J_1 , J_2 y J_3 respectivamente del resultado de Yarnold (1972) haciendo $B = B_\phi(c)$.

Además

$$\begin{aligned} J_1^\phi = & P(\chi_r^2 < c) + \frac{(M-1)}{96n} \left\{ P(\chi_r^2 < c) \left[-8(M+1) \right] + P(\chi_{r+2}^2 < c) \left[-\frac{21}{M^2} \times \right. \right. \\ & \times \frac{\phi^{IV}(1/M)}{\phi''(1/M)} (M-1) + \frac{18 \left[\phi'''(1/M) \right]^2}{M^2 \phi''(1/M)} (M-2) + 24M \Big] + \\ & + P(\chi_{r+4}^2 < c) \left[-\frac{24 \phi'''(1/M)}{M \phi''(1/M)} (M-2) + \frac{21 \phi^{IV}(1/M)}{M^2 \phi''(1/M)} (M-1) - \right. \\ & - \frac{36 \left[\phi'''(1/M) \right]^2}{M^2 \phi''(1/M)} (M-2) - 24(M-1) \Big] + P(\chi_{r+6}^2 < c) \left[2(M- \right. \\ & \left. - 2) \left[\frac{12 \phi'''(1/M)}{M \phi''(1/M)} + \frac{9 \left[\phi'''(1/M) \right]^2}{M^2 \phi''(1/M)} + 4 \right] \right\} + O(n^{-3/2}) \end{aligned}$$

y J_2^ϕ se puede aproximar por

$$\hat{J}_2^\phi = (N_\phi(c) - n^{r/2} V_\phi(c)) e^{-c/2} / \left[(2\pi n)^{r/2} M^{M/2} \right],$$

donde

$$N_\phi(c) = \text{número de puntos } w \in L / w \in B_\phi(c)$$

y

$$V_\phi(c) = \text{volumen de } B_\phi(c) =$$

$$= \frac{(\pi c)^{r/2}}{\Gamma(1+r/2)} \left[\frac{1}{M} \right]^{M/2} \left\{ 1 + \frac{c(M-1)}{32M^2(M+1)n} \left[\frac{[\phi'''(1/M)]^2}{[\phi''(1/M)]} (6(M-2) - \frac{7\phi^{IV}(1/M)}{\phi''(1/M)}(M-1)) \right] \right\} + O(n^{-3/2}).$$

Demostración

En primer lugar, para obtener la expresión de J_1^ϕ , se considera la transformación

$$z^t = w^t H = w^t (I_r, -1) D^{-1/2} A \quad (4.3.6)$$

donde

I_r es la matriz identidad de orden $r = M-1$,

$\mathbf{1} = (1, \dots, 1)^t$ es un vector de dimensión $1 \times r$,

$D = \text{diag}(\pi_0)$,

$A^t = (a_1, \dots, a_M)$ es una matriz $r \times M$ ($A, \sqrt{\pi_0}$) es ortogonal y

$$\sqrt{\pi_0} = \left[\sqrt{1/M}, \dots, \sqrt{1/M} \right]^t.$$

Por una parte, al ser la matriz $(A, \sqrt{\pi_0})$ ortogonal se tiene que $A^t A = I_r$ y $A^t \sqrt{\pi_0} = 0$. Por tanto, como $z^t = w^t (I_r, -1) D^{-1/2} A$, se tiene $Az = D^{-1/2} (I_r, -1)^t w = (w_1 \sqrt{M}, \dots, w_M \sqrt{M})^t$. En consecuencia $w_j = \sqrt{1/M} a_j^t z$. Por otra parte,

$$H^t \Omega H = A^t A - A^t \sqrt{\pi_0} \sqrt{\pi_0^t} A$$

y aplicando que la matriz $(A, \sqrt{\pi_0})$ es ortogonal se llega a que $H^t \Omega H = I_r$. Con lo cual (4.3.3) se puede expresar como

$$P(W = w) = n^{-r/2} |\Omega|^{-1/2} \left\{ f(z) + O(n^{-3/2}) \right\}$$

donde

$$f(z) = (2\pi)^{-r/2} \exp \left[-\frac{1}{2} z^t z \right] \left(1 + n^{-1/2} g_1(z) + n^{-1} g_2(z) \right) \quad (4.3.7)$$

con

$$g_1(z) = -T_1/2 + T_3/6,$$

$$g_2(z) = g_1^2(z)/2 + (1-M^2)/12 + T_2/4 - T_4/12$$

y

$$T_1 = \sum_{j=1}^M (a_j^t z) \sqrt{M},$$

$$T_2 = M \sum_{j=1}^M (a_j^t z)^2,$$

$$T_3 = \sum_{j=1}^M (a_j^t z)^3 \sqrt{M},$$

$$T_4 = M \sum_{j=1}^M (a_j^t z)^4.$$

A partir del resultado de Yarnold (1972) y de (4.3.7) se tiene que

$$J_1^\phi = \int_{B_\phi^*(c)} \dots \int f(z) dz$$

donde

$$B_\phi^*(c) = \{z / z^t = w^t H \text{ y } w \in B_\phi(c)\}. \quad (4.3.8)$$

Interpretando $f(z)$ como una función de densidad continua de una variable aleatoria Z , es posible interpretar J_1^ϕ como la función de distribución de $S_\phi((z^t H^t / \sqrt{n})^t + \pi_0, \pi_0)$ que se denota por $S_\phi(z^t H^t)$ y cuya función característica viene dada por

$$c(t) = \int_{R^r} \dots \int \exp[it S_\phi(z^t H^t)] f(z) dz.$$

Utilizando el desarrollo de Taylor de $S_\phi(X/n, \pi_0)$ dado en (4.2.1) y la transformación (4.3.6) se tiene que

$$S_\phi(z^t H^t) = z^t z + \frac{n^{-1/2} \phi'''(1/M)}{2M\phi''(1/M)} T_3 + n^{-1} \frac{7 \phi^{IV}(1/M)}{48M^2 \phi''(1/M)} T_4 + \\ + O(n^{-3/2}). \quad (4.3.9)$$

Además, al ser

$$\exp(\alpha + n^{-1/2}\beta + n^{-1}\gamma) = e^\alpha (1 + n^{-1/2}\beta + n^{-1}(\gamma + \beta^2/2)) + O(n^{-3/2}),$$

se tiene que

$$\begin{aligned}
\exp\left[itS_{\phi}(z^t H^1)\right]f(z) &= (2\pi)^{-r/2} \exp\left[itz^t z - \frac{1}{2}z^t z + \right. \\
&+ \frac{n^{-1/2} \phi'''(1/M)}{2M\phi''(1/M)} T_3 it + n^{-1} \frac{7 \phi^{IV}(1/M)}{48M^2 \phi''(1/M)} T_4 it + \\
&+ O(n^{-3/2})it\left.\right\} (1 + n^{-1/2}g_1(z) + n^{-1}g_2(z)) = \\
&= (2\pi)^{-r/2} \exp((2it-1)z^t z/2) \left[1 + n^{-1/2}v_1(z) + \right. \\
&+ n^{-1}v_2(z)\left.\right] (1 + n^{-1/2}g_1(z) + n^{-1}g_2(z)) + O(n^{-3/2})
\end{aligned}$$

donde

$$v_1(z) = \frac{\phi'''(1/M)}{2M\phi''(1/M)} T_3 it$$

y

$$v_2(z) = \frac{7 \phi^{IV}(1/M)}{48M^2 \phi''(1/M)} T_4 it - \frac{1}{8M^2} \left[\frac{\phi'''(1/M)}{\phi''(1/M)} \right]^2 T_3^2 it^2.$$

Por tanto,

$$c(t) = (2\pi)^{-r/2} \int_{R^r} \dots \int \exp(-z^t z/2\sigma^2) b(z) dz + O(n^{-3/2}) \quad (4.3.10)$$

donde

$$\sigma^2 = (-2it+1)^{-1}$$

y

$$\begin{aligned}
b(z) = & 1 + n^{-1/2} \left[-T_1/2 + T_3/6 + it \frac{\phi'''(1/M)}{2M\phi''(1/M)} T_3 \right] + \\
& + \frac{n^{-1}}{12M\phi''(1/M)} \left(T_3^2 - 3T_1T_3 \right) + \frac{7}{48M^2\phi''(1/M)} \phi^{IV}(1/M) T_4 it - \\
& - \frac{1}{8M^2} \left[\frac{\phi'''(1/M)}{\phi''(1/M)} \right]^2 T_3^2 t^2 + (-T_1/2 + T_3/6)^2/2 + (1-M^2)/12 + \\
& + T_2/4 - T_4/12 \Big].
\end{aligned}$$

Es decir, se tiene que

$$c(t) = \sigma^t E[b(Z)] + O(n^{-3/2})$$

donde

$$Z \approx N(0, \sigma^2 I).$$

Al ser

$$AZ \approx N(0, \sigma^2 AA^t)$$

con

$$AA^t = \begin{bmatrix} 1 - \frac{1}{M} & \cdots & -\frac{1}{M} \\ \vdots & \ddots & \vdots \\ -\frac{1}{M} & \cdots & 1 - \frac{1}{M} \end{bmatrix},$$

la variable $(a_k Z^k, a_j Z^j)$ es normal bidimensional con vector de medias $(0,0)^t$ y matriz de varianzas-covarianzas $\sigma^2 \begin{bmatrix} 1-1/M & -1/M \\ -1/M & 1-1/M \end{bmatrix}$. Por tanto, la v.a. $a_k Z^k$ condicionada a que $a_j Z^j$ tome el valor t es normal de

media $-t/(M-1)$ y varianza $(1-1/M)(1-(1/(M-1))^2)$. Teniendo en cuenta que si X es normal de media μ y desviación típica σ , entonces

$$E[(X-\mu)^r] = \begin{cases} 0 & r \text{ impar} \\ \frac{r! \sigma^r}{(r/2)! 2^{r/2}} & r \text{ par} \end{cases}$$

se tiene que

$$E[(a_j^t Z)] = E[(a_j^t Z)^3] = 0,$$

$$E[(a_k^t Z)(a_j^t Z)] = \begin{cases} -\frac{\sigma^2}{M} & k \neq j \\ \sigma^2 \left(1 - \frac{1}{M}\right) & k = j \end{cases}$$

ya que, para $k \neq j$

$$\begin{aligned} E[(a_k^t Z)(a_j^t Z)] &= E\left[E[(a_k^t Z)(a_j^t Z) | a_k^t Z = t]\right] = E\left[(a_k^t Z)E[a_j^t Z | a_k^t Z = t]\right] = \\ &= E[(a_k^t Z)(-1)(a_k^t Z)/(M-1)] = -E[(a_k^t Z)^2]/(M-1) = -\frac{\sigma^2}{M} \end{aligned}$$

y si $k=j$, es claro que la esperanza viene dada por

$$E[(a_j^t Z)^2] = \sigma^2 \left(1 - \frac{1}{M}\right).$$

Análogamente,

$$E[(a_k^t Z)^3(a_j^t Z)] = \begin{cases} -\frac{3\sigma^4}{M} \left(1 - \frac{1}{M}\right) & k \neq j \\ 3\sigma^4 \left(1 - \frac{1}{M}\right)^2 & k = j \end{cases}$$

$$E[(a_k^t Z)^2(a_j^t Z)^2] = \begin{cases} \sigma^4 \left(1 - \frac{2}{M} + \frac{3}{M^2}\right) & k \neq j \\ 3\sigma^4 \left(1 - \frac{1}{M}\right)^2 & k = j \end{cases}$$

$$E[(a_k^t Z)^3 (a_j^t Z)^3] = \begin{cases} -\sigma^6 \left[9 \left(1 - \frac{1}{M}\right)^2 \frac{1}{M} + \frac{6}{M^3} \right] & k \neq j \\ 15\sigma^6 \left(1 - \frac{1}{M}\right)^3 & k = j \end{cases}.$$

Luego,

$$E[T_1] = E[T_3] = 0,$$

$$E[T_1^2] = M \sum_{k=1}^M \sum_{j=1}^M E[(a_k^t Z)(a_j^t Z)] = M \left\{ -\frac{\sigma^2}{M} (M-1) + \sigma^2 M \left(1 - \frac{1}{M}\right) \right\} = 0,$$

$$E[T_2] = \sigma^2 M(M-1),$$

$$E[T_1 T_3] = 0,$$

$$E[T_4] = 3\sigma^4 (M-1)^2,$$

y

$$E[T_3^2] = 3\sigma^6 (2M^2 - 6M + 4).$$

de donde,

$$\begin{aligned} c(t) &= \sigma^r E[b(Z)] + O(n^{-3/2}) = \sigma^r + \frac{\sigma^2}{n} \left\{ \left[\frac{1-\sigma^{-2}}{2} \right] \frac{\phi''''(1/M)}{12M\phi''(1/M)} \times \right. \\ &\quad \times \left[3\sigma^6 (2M^2 - 6M + 4) \right] + \frac{7}{48M^2} \frac{\phi^{IV}(1/M)}{\phi''(1/M)} \left[\frac{1-\sigma^{-2}}{2} \right] 3\sigma^4 (M-1)^2 + \\ &\quad + \frac{1}{8M^2} \left[\frac{\phi''''(1/M)}{\phi''(1/M)} \right]^2 \left[\frac{1 + \sigma^{-4} - 2\sigma^{-2}}{4} \right] 3\sigma^6 (2M^2 - 6M + 4) + \frac{1}{12} \sigma^6 (M^2 - \\ &\quad - 3M + 2) + \frac{(1-M^2)}{12} + \left. \frac{\sigma^2 M(M-1)}{4} - \frac{\sigma^4 (M-1)^2}{4} \right\} + O(n^{-3/2}) = \end{aligned}$$

$$\begin{aligned}
&= \sigma^r + \frac{\sigma^r}{96n} \left\{ 8(1-M^2) + \sigma^2 \left[-\frac{21\phi^{IV}(1/M)}{M^2\phi''(1/M)} (M-1)^2 + (2M^2-6M + \right. \right. \\
&+4) \frac{9}{M^2} \left[\frac{\phi'''(1/M)}{\phi''(1/M)} \right]^2 + 24M(M-1) \left. \right] + \sigma^4 \left[-\frac{12\phi'''(1/M)}{M\phi''(1/M)} (2M^2- \right. \\
&-6M+4) + \frac{21\phi^{IV}(1/M)}{M^2\phi''(1/M)} (M-1)^2 - \frac{36[\phi'''(1/M)]^2}{M^2[\phi''(1/M)]^2} (M^2-3M+2) - \\
&-24(M-1)^2 \left. \right] + \sigma^6 \left[(2M^2-6M+4) \left[\frac{12\phi'''(1/M)}{M\phi''(1/M)} + \frac{9[\phi'''(1/M)]^2}{M^2[\phi''(1/M)]^2} \right. \right. \\
&\left. \left. +4 \right] \right] \left. \right\} + O(n^{-3/2}).
\end{aligned}$$

Teniendo en cuenta que σ^r es la función característica de una χ_r^2 y que $c(t)$ es la función característica de la distribución J_1^ϕ , se tiene que

$$\begin{aligned}
J_1^\phi &= P(\chi_r^2 < c) + \frac{1}{96n} \left\{ P(\chi_r^2 < c) 8(1-M^2) + P(\chi_{r+2}^2 < c) \left[-\frac{21\phi^{IV}(1/M)}{M^2\phi''(1/M)} \right. \right. \\
&\times (M-1)^2 + \frac{9}{M^2} \left[\frac{\phi'''(1/M)}{\phi''(1/M)} \right]^2 (2M^2-6M+4) + 24M(M-1) \left. \right] + \\
&+ P(\chi_{r+4}^2 < c) \left[-\frac{24\phi'''(1/M)}{M\phi''(1/M)} (M^2-3M+2) + \frac{21\phi^{IV}(1/M)}{M^2\phi''(1/M)} (M-1)^2 - \right. \\
&- \frac{36[\phi'''(1/M)]^2}{M^2[\phi''(1/M)]^2} (M^2-3M+2) - 24(M-1)^2 \left. \right] + P(\chi_{r+6}^2 < c) \left[(2M^2- \right. \\
&-6M+4) \left[\frac{12\phi'''(1/M)}{M\phi''(1/M)} + \frac{9[\phi'''(1/M)]^2}{M^2[\phi''(1/M)]^2} +4 \right] \left. \right] \left. \right\} + O(n^{-3/2})
\end{aligned}$$

como se quería demostrar.

Por último, se calculará \hat{J}_2^ϕ .

Por una parte se sabe que si $w_s = \gamma_s(w^*)$ ó $w_s = \theta_s(w^*)$, $s=1, \dots, r$, entonces $S_\phi((x/n, x_M/n), \pi_0) = c$ y por otra parte que

$$S_\phi((x/n, x_M/n), \pi_0) = w^t \Omega^{-1} w + o(1),$$

por tanto

$$\varphi(w) = (2\pi)^{-r/2} |\Omega|^{-1/2} \exp(-c/2) + o(1)$$

cuando $w_s = \gamma_s(w^*)$ ó $w_s = \theta_s(w^*)$, $s=1, \dots, r$.

En consecuencia,

$$\left[S_1(\sqrt{n} w_s + n\pi_{0s}) \varphi(w) \right]_{\gamma_s(w^*)}^{\theta_s(w^*)}$$

se puede expresar como

$$(2\pi)^{-r/2} |\Omega|^{-1/2} \exp(-c/2) \left[S_1(\sqrt{n} w_s + n\pi_{0s}) \right]_{\gamma_s(w^*)}^{\theta_s(w^*)} + o(1)$$

de donde a partir de la demostración del Teorema 4 de Yarnold (1972, p.1572) se obtiene

$$J_2^\phi = (N_\phi(c) - n^{r/2} V_\phi(c)) e^{-c/2} / \left[(2\pi n)^r |\Omega| \right]^{1/2} + o(1)$$

siendo $N_\phi(c)$ el número de puntos de L que pertenecen a $B_\phi(c)$ y $V_\phi(c)$ el volumen de $B_\phi(c)$. Es decir,

$$V_\phi(c) = \int_{B_\phi(c)} \dots \int dw = |\Omega|^{1/2} \int_{B_\phi^*(c)} \dots \int dz$$

donde z está definida en (4.3.6) y $B_{\phi}^*(c)$ en (4.3.8).

Considérese ahora la transformación que a cada valor z le hace corresponder un valor u tal que

$$u^t u = S_{\phi}(z^t H^{-1}),$$

e.d., de forma que

$$\begin{aligned} u^t u = & z^t z + n^{-1/2} \frac{\phi'''(1/M)}{2M\phi''(1/M)} T_3 + n^{-1} \frac{7\phi^{IV}(1/M)}{48M^2\phi''(1/M)} T_4 + \\ & + O(n^{-3/2}) \end{aligned} \quad (4.3.12)$$

sin mas que considerar (4.3.9), donde si

$$z = d_1(u) + n^{-1/2}d_2(u) + n^{-1}d_3(u) + O(n^{-3/2}),$$

(4.3.12) se puede escribir como

$$\begin{aligned} u^t u = & d_1^t(u)d_1(u) + n^{-1/2} \left[2d_1^t(u)d_2(u) + \frac{\phi'''(1/M)}{2M^{1/2}\phi''(1/M)} \sum_{j=1}^M (a_j^t d_1(u))^3 \right] + \\ & + n^{-1} \left[2d_1^t(u)d_3(u) + d_2^t(u)d_2(u) + \frac{3\phi'''(1/M)}{2M^{1/2}\phi''(1/M)} \sum_{j=1}^M (a_j^t d_1(u))^2 \times \right. \\ & \left. \times (a_j^t d_2(u)) + \frac{7\phi^{IV}(1/M)}{48M\phi''(1/M)} \sum_{j=1}^M (a_j^t d_1(u))^4 \right] + O(n^{-3/2}) \end{aligned}$$

donde $d_1(u)$, $d_2(u)$ y $d_3(u)$ son tales que han de verificar

$$d_1^t(u)d_1(u) = u^t u$$

$$2d_1^t(u)d_2(u) + \frac{\phi'''(1/M)}{2M^{1/2}\phi''(1/M)} \sum_{j=1}^M (a_j^t d_1(u))^3 = 0$$

$$2d_1^t(u)d_3(u) + d_2^t(u)d_2(u) + \frac{3\phi'''(1/M)}{2M^{1/2}\phi''(1/M)} \sum_{j=1}^M (a_j^t d_1(u))^2 (a_j^t d_2(u)) +$$

$$+ \frac{7\phi^{IV}(1/M)}{48M\phi''(1/M)} \sum_{j=1}^M (a_j^t d_1(u))^4 = 0,$$

obteniéndose tras diversos cálculos que

$$d_1(u) = u,$$

$$d_2(u) = - \frac{\phi'''(1/M)}{4M^{1/2}\phi''(1/M)} \sum_{j=1}^M (a_j^t u)^2 a_j$$

y

$$d_3(u) = \frac{1}{96M} \left\{ \left[\frac{\phi'''(1/M)}{\phi''(1/M)} \right]^2 \frac{15}{M} \left[\sum_{j=1}^M M(a_j^t u)^3 a_j - (u^t u)u \right] - \right.$$

$$\left. - \frac{7\phi^{IV}(1/M)}{\phi''(1/M)} \sum_{j=1}^M (a_j^t u)^3 a_j \right\}.$$

La matriz Jacobiana de esta transformación viene dada por

$$(\partial z / \partial u) = I_r + n^{-1/2} \left[- \frac{\phi'''(1/M)}{2M^{1/2}\phi''(1/M)} \right] P_1 + n^{-1} \frac{1}{96M} \left\{ \left[\frac{\phi'''(1/M)}{\phi''(1/M)} \right]^2 \times \right.$$

$$\left. \times \frac{15}{M} \left[3MP_2 - 2u^t u - uu^t I_r \right] - 21 \frac{\phi^{IV}(1/M)}{\phi''(1/M)} P_2 \right\} + O(n^{-3/2})$$

donde

$$P_1 = \sum_{j=1}^M (a_j^t u) a_j a_j^t \quad y \quad P_2 = \sum_{j=1}^M (a_j^t u)^2 a_j a_j^t.$$

Para calcular el determinante Jacobiano se utilizará el siguiente resultado general:

$$|I_r + n^{-1/2}B + n^{-1}C| = 1 + n^{-1/2} \sum_{i=1}^r b_{ii} + n^{-1} \left[\sum_{i=1}^r c_{ii} + \frac{1}{2} \sum_{i,j}^r (b_{ii} b_{jj} - b_{ij} b_{ji}) \right] + O(n^{-3/2})$$

donde B y C son matrices cuadradas $r \times r$.

De este modo se obtiene que

$$|\partial z / \partial u| = 1 + n^{-1/2} \left[- \frac{\phi'''(1/M)}{2M^{1/2} \phi''(1/M)} \right] Q_1 + \frac{n^{-1}}{32M} \left[(15MQ_2 - 5(2+r)u^t u + 4MQ_1^2 - \frac{4MQ_{12}}{M} \left[\frac{\phi'''(1/M)}{\phi''(1/M)} \right]^2 - 7 \frac{\phi^{IV}(1/M)}{\phi''(1/M)} Q_2) \right] + O(n^{-3/2})$$

donde

$$Q_1 = \sum_{j=1}^M (a_j^t u) a_j^t a_j,$$

$$Q_2 = \sum_{j=1}^M (a_j^t u)^2 a_j^t a_j$$

y

$$Q_{12} = \sum_{k,j}^M (a_k^t u) (a_j^t u) (a_k^t a_j)^2.$$

Teniendo en cuenta que

$$Q_1 = \frac{T_1}{M^{1/2}},$$

$$Q_2 = \frac{T_2 - u^t u}{M},$$

y

$$Q_{12} = \frac{T_2 - 2u^t u}{M},$$

donde se ha reemplazado z por u en las correspondientes expresiones de T_1 y T_2 , se obtiene que

$$\begin{aligned} |\partial z / \partial u| = & 1 + n^{-1/2} \left[- \frac{\phi'''(1/M)}{2M\phi''(1/M)} \right] T_1 + \frac{n^{-1}}{32M^2} \left[(15T_2 - 5(5+r)u^t u + \right. \\ & \left. + 4T_1^2 - 4T_2 + 8u^t u) \frac{1}{M} \left[\frac{\phi''''(1/M)}{\phi''(1/M)} \right]^2 - 7 \frac{\phi^{IV}(1/M)}{\phi''(1/M)} (T_2 - u^t u) \right] + O(n^{-3/2}). \end{aligned} \quad (4.3.13)$$

Sustituyendo u por z en $V_\phi(c)$ se tiene que

$$V_\phi(c) = |\Omega|^{1/2} \int_{u^t u < c} \dots \int |\partial z / \partial u| \, du$$

es decir

$$\begin{aligned} V_\phi(c) = & |\Omega|^{1/2} \left\{ M_r + n^{-1/2} \left[- \frac{\phi'''(1/M)}{2M\phi''(1/M)} \right] N_1 + \frac{n^{-1}}{32M^2} \left\{ \left[-(15+5r)N_2 + \right. \right. \right. \\ & \left. \left. + 11N_3 + 4N_4 \right] \frac{1}{M} \left[\frac{\phi''''(1/M)}{\phi''(1/M)} \right]^2 - 7 \frac{\phi^{IV}(1/M)}{\phi''(1/M)} (N_3 - N_2) \right\} + O(n^{-3/2}) \end{aligned}$$

donde

$$M_r = \int_{u_t}^c \dots \int du = (\pi c)^{r/2} / \Gamma(1+r/2),$$

$$N_1 = \sum_{j=1}^M \sqrt{M} \left[\sum_{k=1}^r a_{jk} I_k \right],$$

$$N_2 = \sum_{k=1}^r I_{kk},$$

$$N_3 = \sum_{j=1}^M M \left[\sum_{k,m}^r a_{jk} a_{jm} I_{km} \right],$$

$$N_4 = \sum_{i,j}^M M \left[\sum_{k,m}^r a_{ik} a_{jm} I_{km} \right],$$

e

$$I_k = \int_{u_t}^c \dots \int u_k du,$$

$$I_{km} = \int_{u_t}^c \dots \int u_k u_m du.$$

Además, a partir de la demostración del Teorema 2.1.8 de Read (1982), se sabe que

$$I_k = 0, I_{km} = 0 \text{ para } k \neq m \text{ e } I_{kk} = \frac{M c}{r+2}, k=1, \dots, r.$$

Es decir,

$$N_1 = 0,$$

$$N_2 = (M-1) \frac{M c}{M+1},$$

$$N_3 = (M-1)M \frac{M_r c}{M+1},$$

y

$$N_4 = N_3 - (M-1)M \frac{M_r c}{M+1} = 0,$$

por tanto

$$V_{\phi}(c) = \left[\frac{1}{M} \right]^{M/2} M_r \left\{ 1 + \frac{c(M-1)}{32M^2(M+1)n} \left[\frac{(\phi'''(1/M))^2}{\phi''(1/M)} \right] (6(M-2)) - \frac{7\phi^{IV}(1/M)}{\phi''(1/M)} (M-1) \right\} + O(n^{-3/2}).$$

Este resultado es el que se quería demostrar y con él termina la demostración del teorema. ■

La aproximación obtenida en este teorema se acerca más a la distribución exacta de la familia $S_{\phi}(X/n, \pi_0)$ que la aproximación χ^2 . Sin embargo, la diferencia en los cálculos que hay que realizar para obtener una u otra es significativa.

Observación 4.3.1

De lo anterior se sabe que $J_3^{\phi} = O(n^{-1})$. No obstante, debido a la equivalencia asintótica de la familia $S_{\phi} n(J_3^{\phi_1} - J_3^{\phi_2}) \rightarrow 0$ cuando $n \rightarrow \infty$ por tanto cualquier término ϕ -dependiente en J_3^{ϕ} será $O(n^{-3/2})$. De esta forma como en el desarrollo de la distribución S_{ϕ} en el Teorema 4.3.1 sólo aparecen términos mayores que $O(n^{-3/2})$, J_3^{ϕ} se puede considerar independiente de ϕ . Por esta razón, este término únicamente causa un ajuste constante independiente de ϕ en la distribución.

4.4.- Comparación de las diferentes aproximaciones de la distribución exacta de los estadísticos R_ϕ -divergencia

En este apartado se comparan los errores cometidos al utilizar las aproximaciones de la distribución exacta de S_ϕ para muestras pequeñas mediante dos criterios diferentes. El estudio se realiza en el caso de hipótesis nula simétrica ya que además de las razones dadas en los capítulos anteriores, el número de particiones x que tienen que ser consideradas para las comparaciones que se llevan a cabo se reducen de forma notable cuando la probabilidad de todas las clases es la misma, debido a la invarianza del estadístico S_ϕ frente a las permutaciones en las frecuencias observadas.

A lo largo del estudio realizado se utiliza de forma sistemática la distribución exacta de $S_\phi(X/n, \pi_0)$ que se denotará por

$$T_E(c) \equiv P(S_\phi(X/n, \pi_0) < c)$$

y cuyo procedimiento de cálculo es como sigue:

- 1) Elegir n y M y calcular todas las posibles particiones x de n en M clases. Para cada partición x se calcula la probabilidad multinomial asociada y el valor del estadístico $S_\phi(x/n, \pi_0)$.
- 2) Ordenar las particiones de acuerdo al valor del estadístico de menor a mayor.
- 3) Sumar las probabilidades de las particiones ordenadas hasta llegar a aquella cuyo valor del estadístico asociado es igual o supera a c .

Algunos autores (Tate y Hyer (1973); Kotze y Gokhale (1980)) propusieron un cálculo de la distribución exacta diferente a éste en el paso 2). Estos autores proponían que las particiones se ordenaran

de acuerdo a sus probabilidades multinomiales. Este método es considerado incorrecto por Radlow y Alf (1975) puesto que no necesariamente las particiones con probabilidades pequeñas son las que más se alejan de H_0 .

Las aproximaciones para T_E consideradas a lo largo de este trabajo y con las cuales se lleva a cabo diferentes comparaciones son las siguientes:

a) En el apartado 3.2 se obtuvo la primera aproximación para T_E que ahora se denota por

$$T_{\chi}(c) \equiv P(\chi_{M-1}^2 < c)$$

donde χ_{M-1}^2 representa a una variable aleatoria χ^2 con $M-1$ grados de libertad.

b) A partir del estadístico corregido propuesto en el apartado 4.2 es posible definir una aproximación mejor dada por

$$T_M(c) \equiv P\left[\chi_{M-1}^2 < \frac{c - \gamma_{\phi}}{\sqrt{\delta_{\phi}}}\right]$$

donde γ_{ϕ} y δ_{ϕ} vienen dados en (4.2.3).

c) En el apartado 4.3 se obtuvo la aproximación

$$T_D(c) \equiv J_1^{\phi} + \hat{J}_2^{\phi}$$

donde J_1^{ϕ} y \hat{J}_2^{ϕ} vienen dados en el Teorema 4.3.1.

d) Por último, en el apartado 3.4 se demuestra un resultado asintótico para el caso en que $M \rightarrow \infty$ cuando $n \rightarrow \infty$ de tal forma que $n/M \rightarrow v$, donde $0 < v < \infty$, que se denota por

$$T_N(c) \equiv P\left[N(0,1) < \frac{c - \mu_M}{\sigma_M}\right]$$

con μ_M , σ_M definidos en el Teorema 3.4.1 y $N(0,1)$ representando una variable aleatoria Normal estándar.

Para realizar una comparación de los errores de aproximación cometidos al aplicar los resultados asintóticos cuando los tamaños muestrales son pequeños se siguen dos procedimientos diferentes a los que hemos llamado Criterio 1 y Criterio 2. Evidentemente para llevar a cabo este estudio hay que fijarse en una familia ϕ . Esta será $\phi = \phi_\alpha$ que se ha utilizado en muchas ocasiones ya que tiene el atractivo de contener al estadístico más conocido en bondad de ajuste, el estadístico X^2 de Pearson.

Criterio 1

Se evalúa el máximo error cometido al utilizar cada una de las cuatro aproximaciones consideradas de la distribución exacta T_E en vez de ésta. Es decir, se calcula

$$\max_X |T_E(S_{\phi_\alpha}(x/n, \pi_0)) - T_i(S_{\phi_\alpha}(x/n, \pi_0))|$$

para $i = \chi, M, D$ y N .

En las Figuras 4.5.1 a la 4.5.8 se representan estos errores máximos de aproximación pero conservando el signo de la diferencia para diferentes valores de n y M y valores de α en $(0,3]$. Las aproximaciones T_χ , T_M , T_D y T_N se denotan en las gráficas por Apr1, Apr2, Apr3 y Apr4, respectivamente.

En todos los casos se observa que la aproximación T_D es la mejor

puesto que es la que más se aproxima al eje de las x , es decir, el error máximo de aproximación que se comete frente a la exacta es el que más se aproxima a 0. Las aproximaciones T_χ y T_M son similares aunque siempre es algo mejor esta última como era de esperar. Por último, con T_N se comete un error similar al de T_χ y T_M para $n=10$ pero de diferente signo. Además, para esta aproximación cuando n crece el error no disminuye como ocurre para las otras aproximaciones sino que decrece cuando M crece.

En cuanto a los miembros de la familia S_{ϕ_α} preferidos según este criterio para todas las aproximaciones son los correspondientes a $\alpha \in [1.5, 2]$.

Figura 4.5.1: Máximos errores de aproximación. $n=10$, $M=3$.

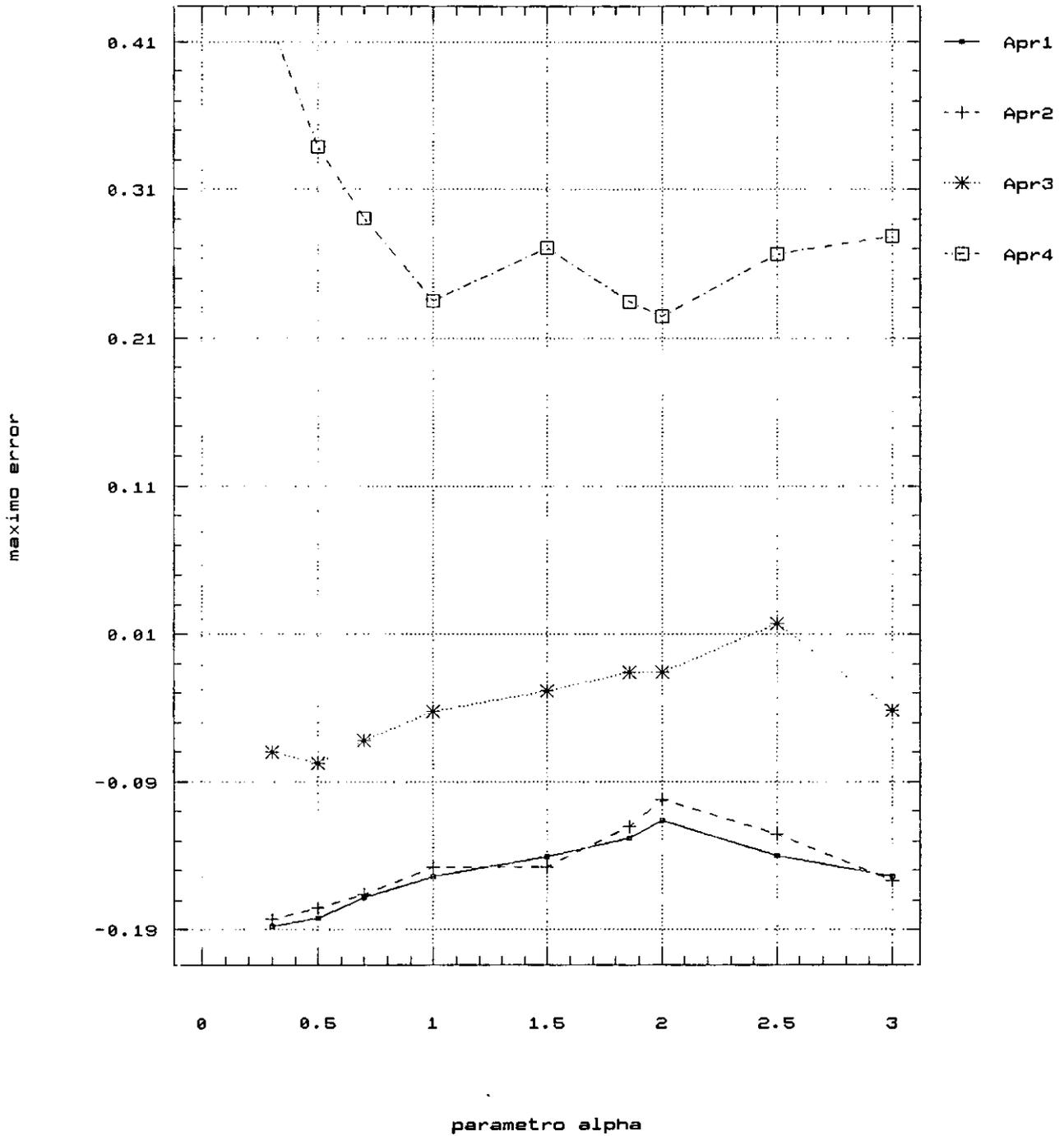


Figura 4.5.2: Máximos errores de aproximación. $n=20$, $M=3$.

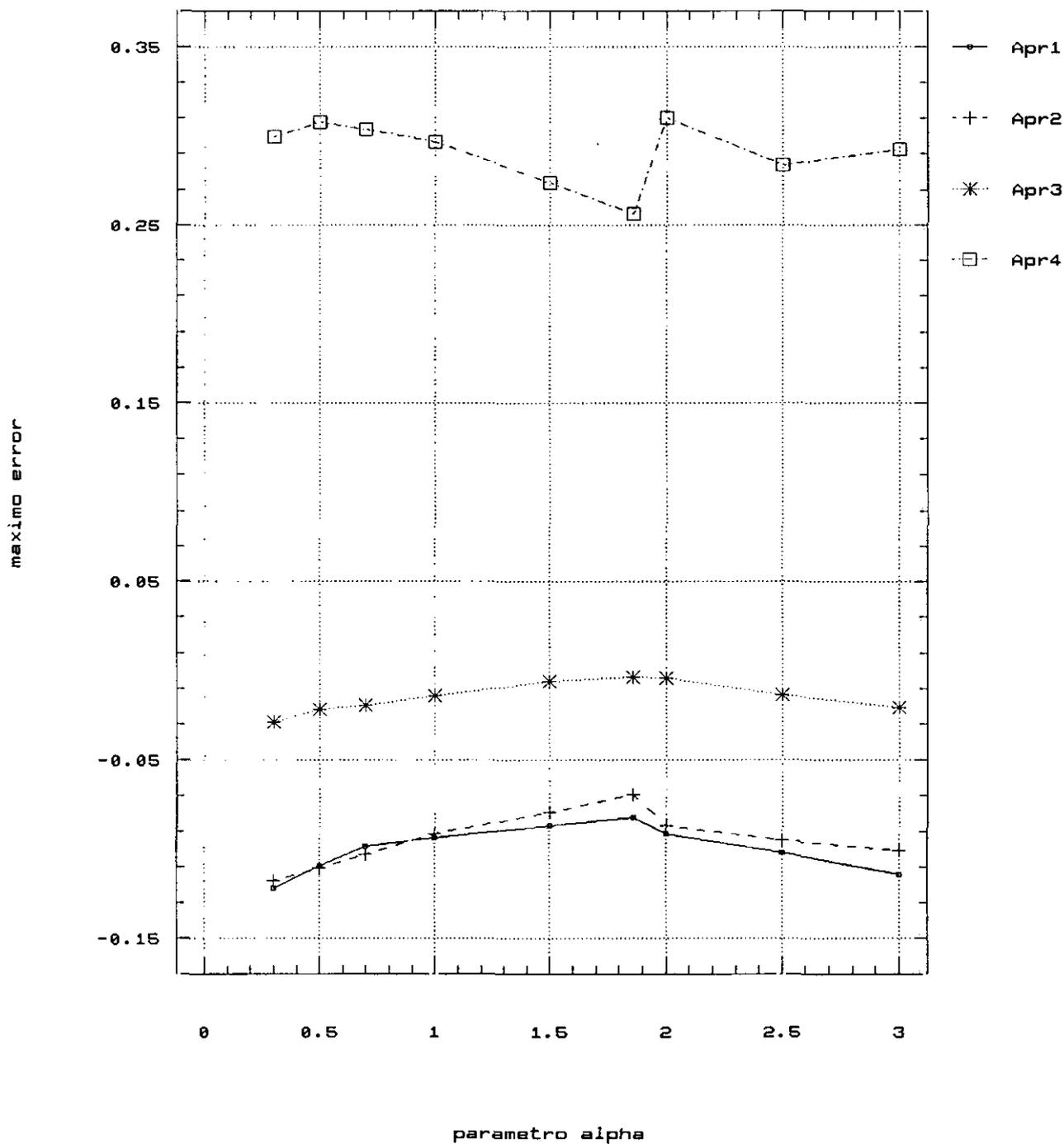


Figura 4.5.3: Máximos errores de aproximación. $n=10$, $M=4$.

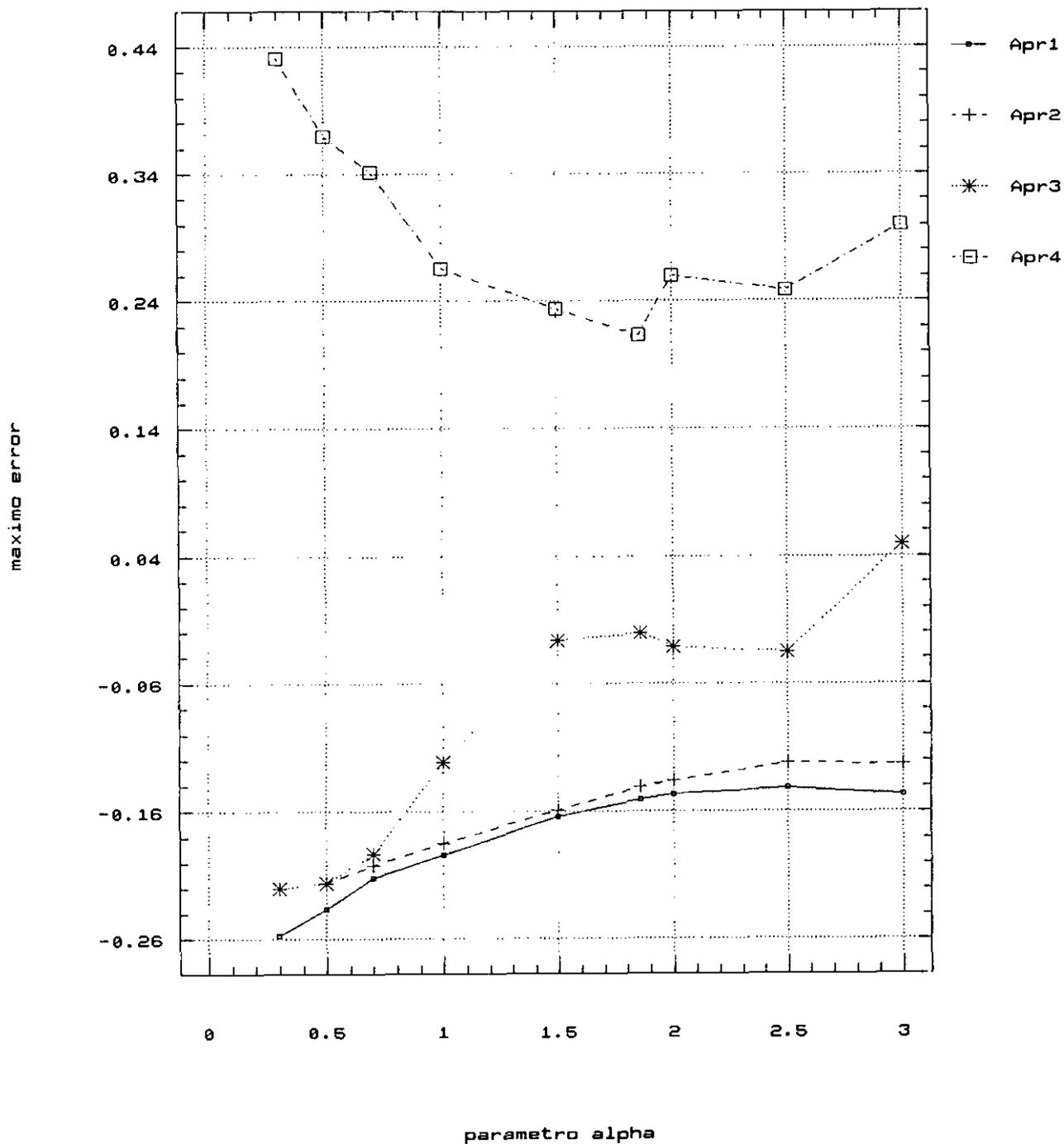


Figura 4.5.4: Máximos errores de aproximación. $n=20$, $M=4$.

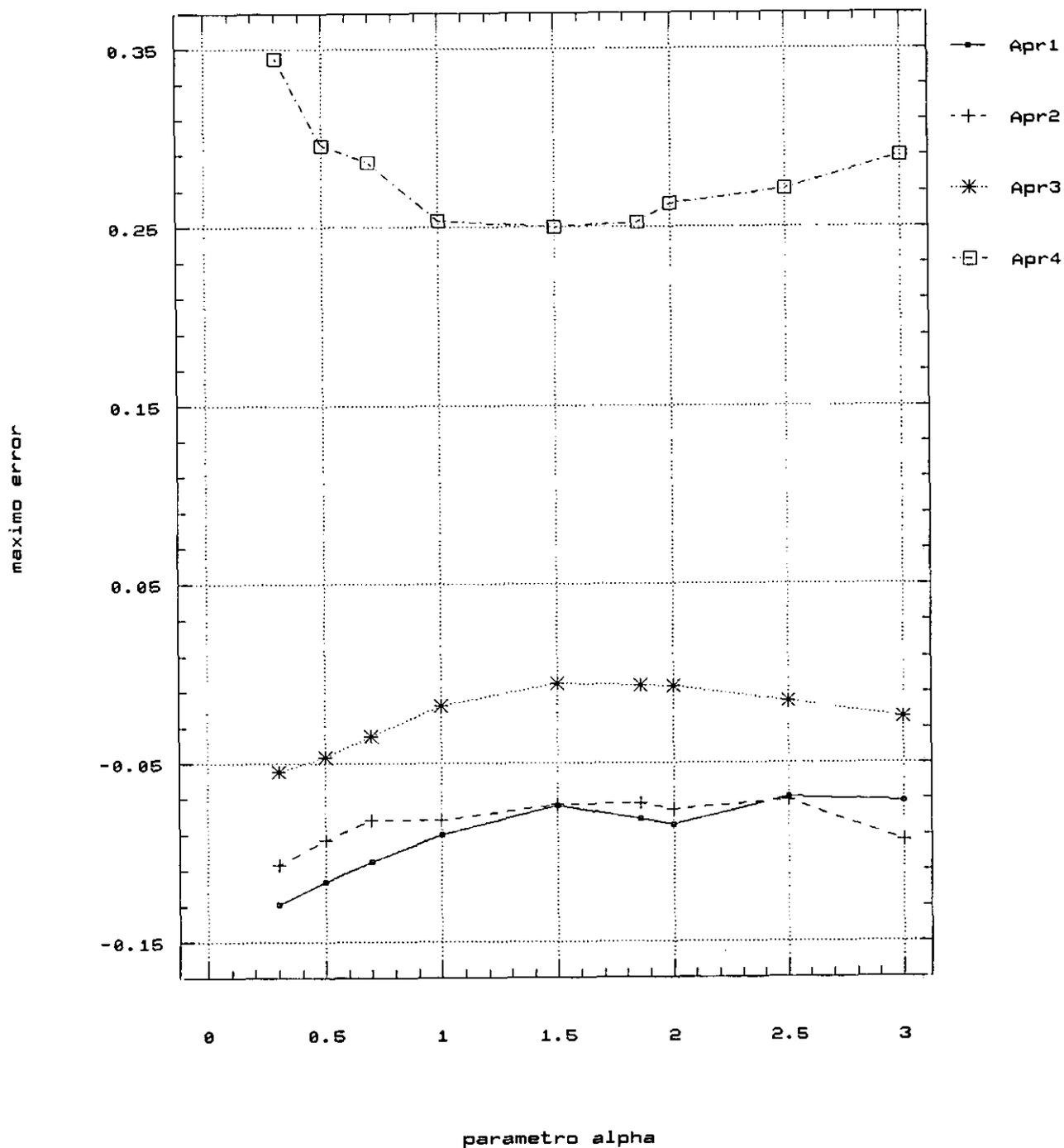


Figura 4.5.5: Máximos errores de aproximación. $n=10$, $M=5$.

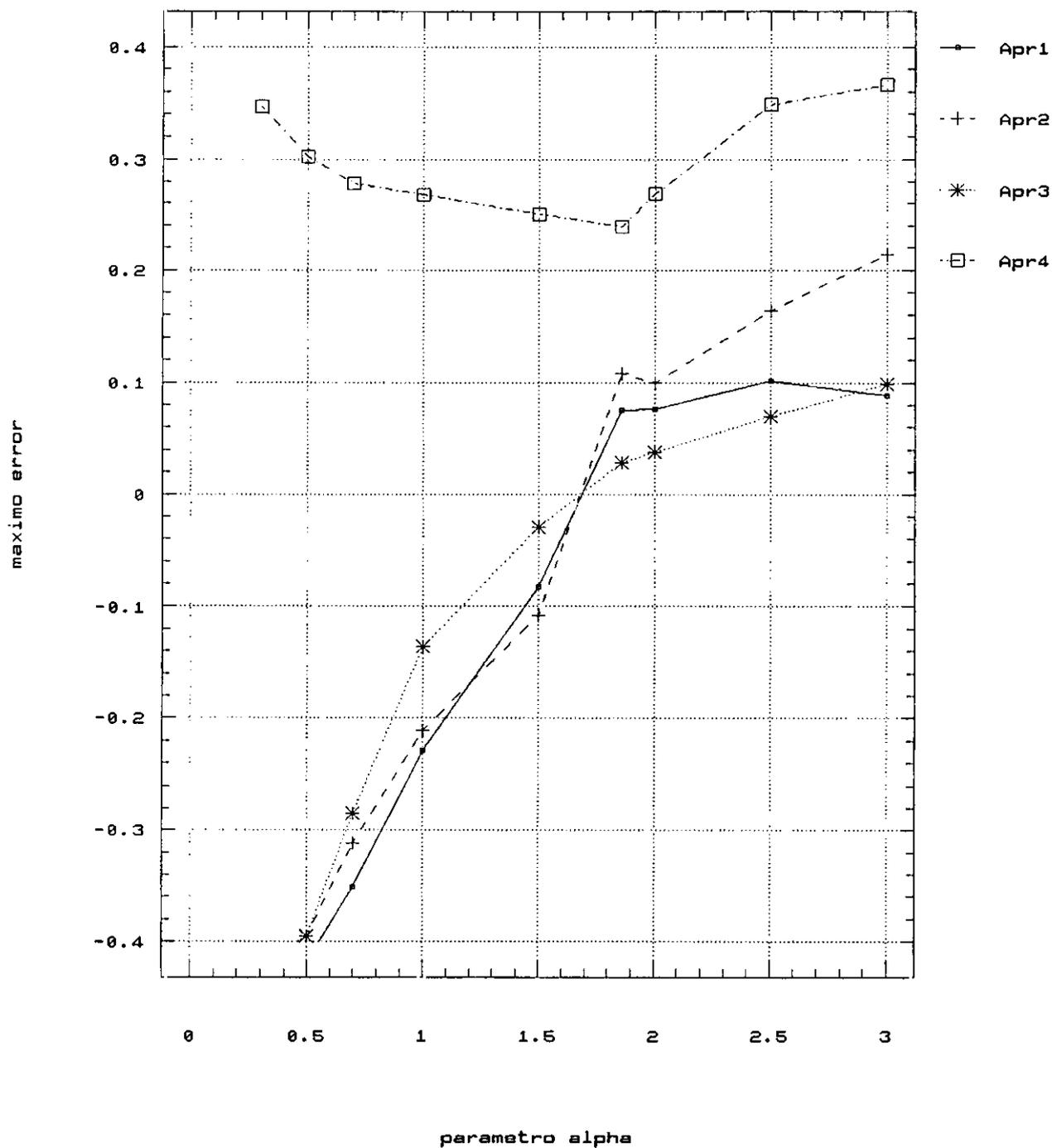


Figura 4.5.6: Máximos errores de aproximación. $n=20$, $M=5$.

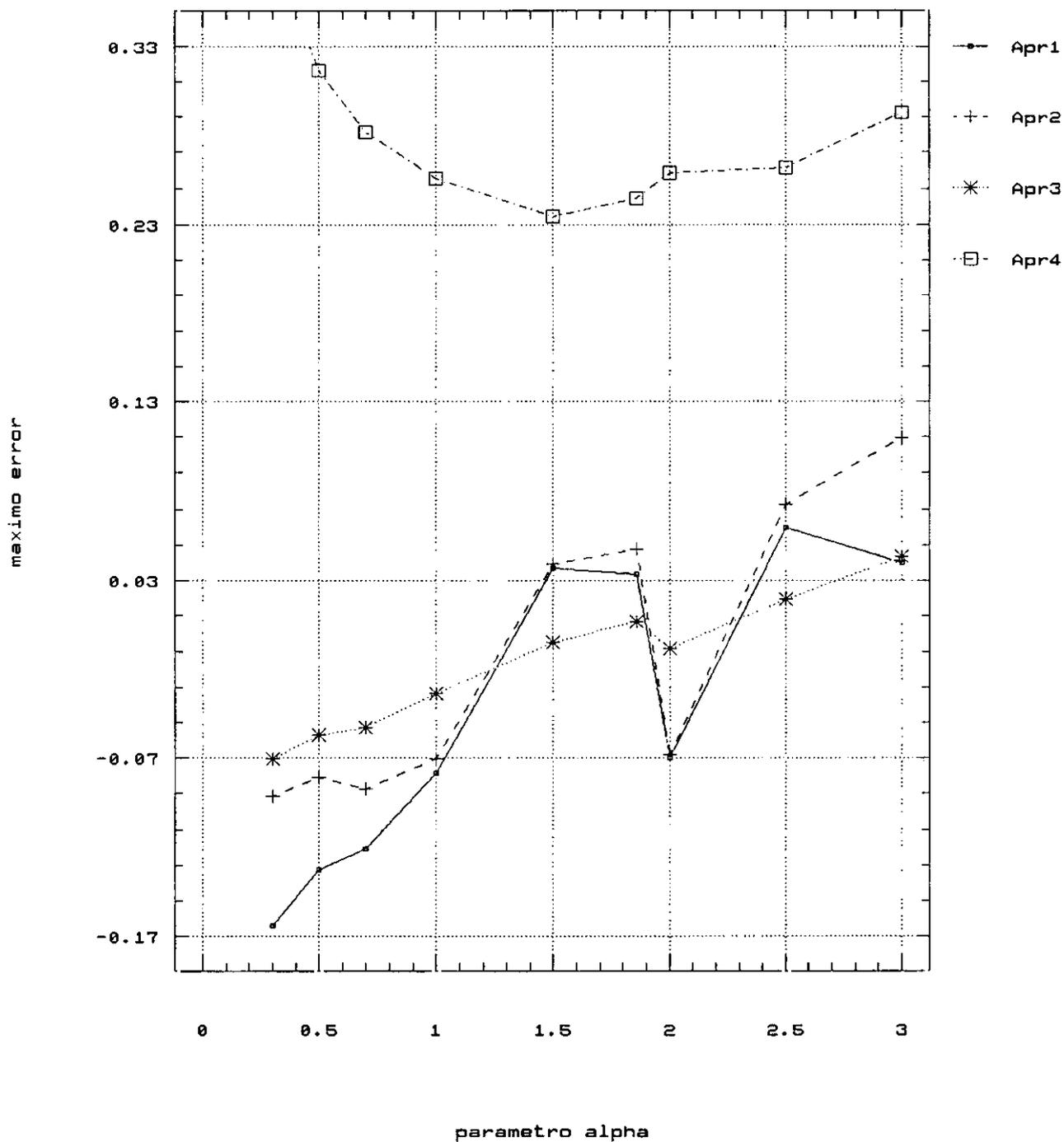


Figura 4.5.7: Máximos errores de aproximación. $n=10$, $M=6$.

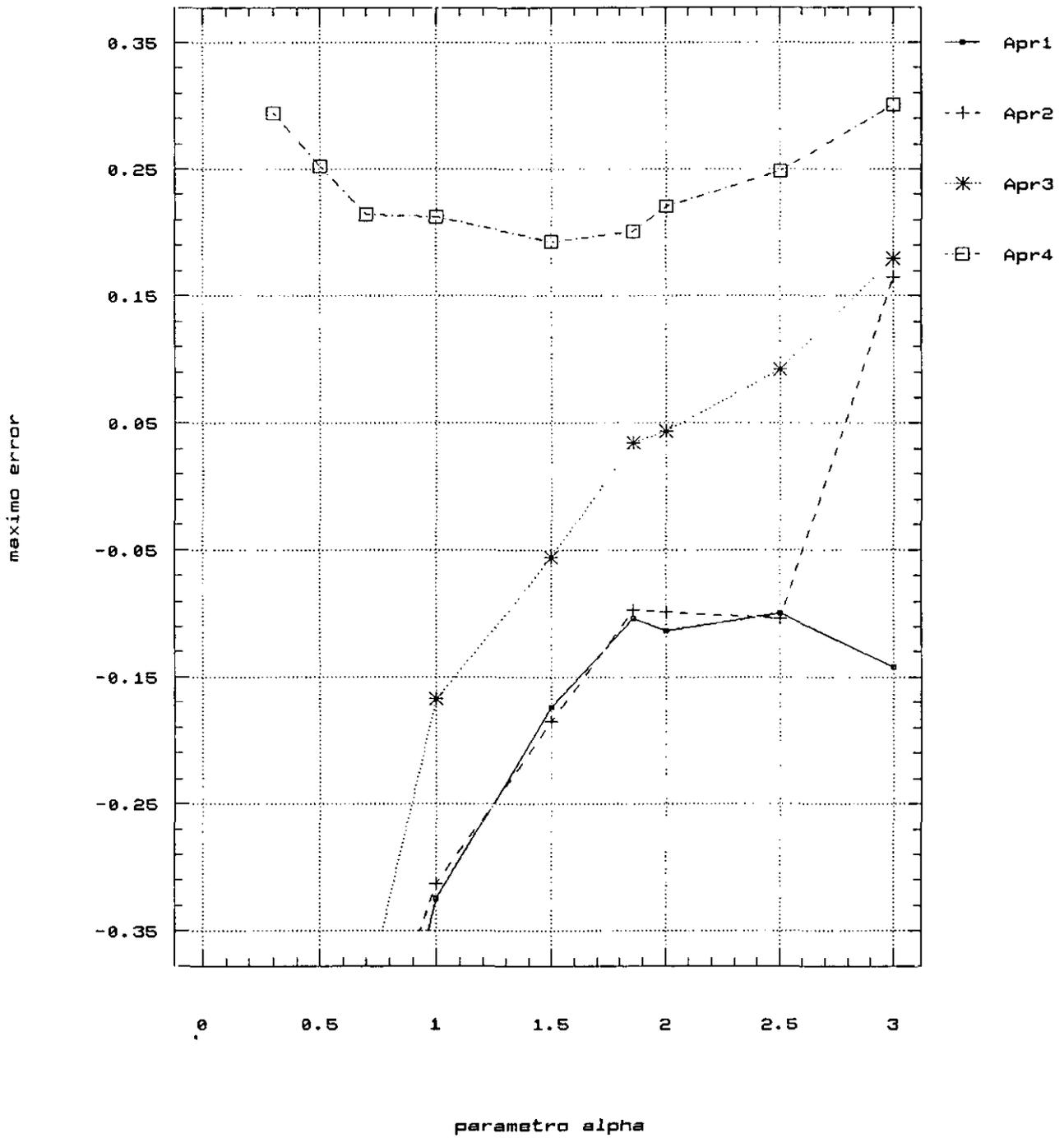
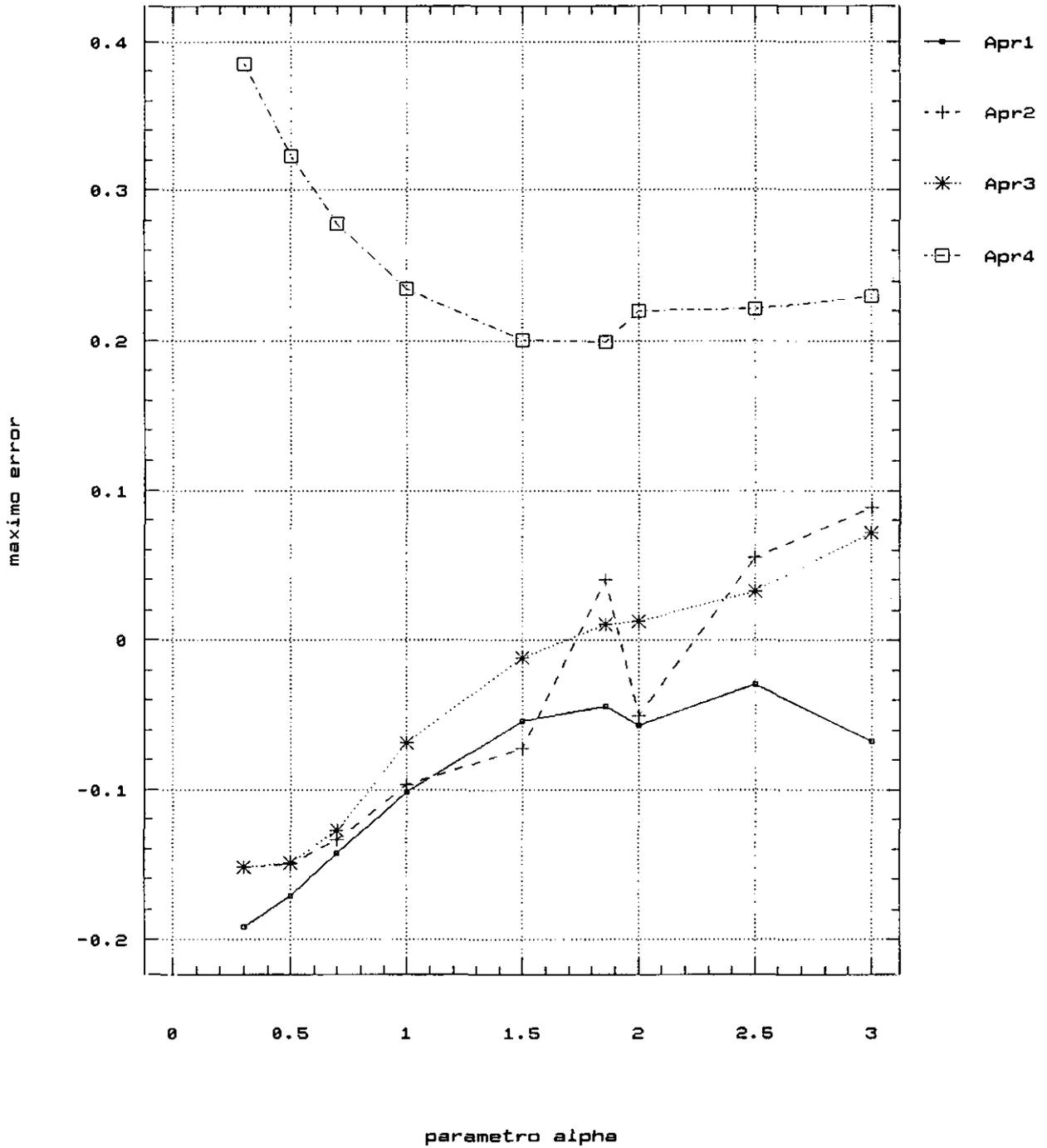


Figura 4.5.8: Máximos errores de aproximación. $n=20$, $M=6$.



Criterio 2

Se valora la precisión de las aproximaciones de T_E calculando el tamaño del contraste con cada una de ellas. Se utiliza la aproximación T_χ para obtener un contraste aproximado de tamaño γ , es decir el contraste con región crítica $[\chi_{M-1, \gamma}^2, \infty)$. Entonces se calcula el nivel de significación exacto de este contraste y se compara con los niveles de significación obtenidos con el resto de las aproximaciones, es decir, se calcula

$$1 - T_i(\chi_{M-1, \gamma}^2) \quad \text{para } i=E, M, D, N.$$

Existen dos razones por las que hemos tomado como contraste de referencia el de la aproximación χ^2 . Por una parte es la aproximación de uso más frecuente y por otra la región crítica es independiente de α .

En las Figuras 4.5.9 a 4.5.16 se representan los niveles de significación exacto y de las cuatro aproximaciones para diferentes valores de n y M y $\gamma=0.1$. En las Figuras 4.5.17 a 4.5.24 se ilustran los niveles de significación para $\gamma=0.01$.

En las gráficas correspondientes al nivel $\gamma=0.1$, la aproximación T_D no destaca como la mejor como ocurría con el criterio anterior sino que tiene un comportamiento en todos los casos parecido a la T_M . Por tanto es preferida la T_M por su sencillez de cálculo con respecto a la complicada fórmula que tenemos que resolver para obtener T_D . Tanto una como otra aproximación son mejores para $n=20$.

La aproximación T_N es peor que estas dos y como era de esperar no mejora cuando aumenta n sino cuando aumenta M . Además, la mejoría es bastante acusada.

El nivel de significación de la aproximación T_{χ} no varía con α , es para todos los casos 0.1. Para esta aproximación hay que tener especial cuidado con el miembro de la familia $S_{\phi_{\alpha}}$ que se elige ya que si está fuera del intervalo $[1.5, 2.5]$ el error cometido puede ser bastante grande.

Para el nivel $\gamma=0.01$, no solo el error cometido por la aproximación T_{χ} frente a la exacta T_E , cuando se eligen miembros de la familia $S_{\phi_{\alpha}}$ fuera del intervalo $[1.5, 2]$, incrementa considerablemente sino que también ocurre con el resto de las aproximaciones. Aunque sigue siendo con la T_{χ} con la que se comete un mayor error al salirse de ese intervalo. Otra diferencia es que la aproximación T_N mejora de forma clara aunque siguen siendo la T_M y la T_D preferidas en la mayoría de los casos.

Como conclusión de los dos criterios anteriores parece conveniente en el caso que se utilice la aproximación T_{χ} elegir un valor de α en $[1.5, 2]$ cuando $M \leq 6$. Por otra parte, cuando se estudia que valores de α hacían que los tres primeros momentos de la $S_{\phi_{\alpha}}$ se aproximaban más a los de una χ_{M-1}^2 se obtuvo los valores $\alpha=13/7$ ó $\alpha=2$, por tanto también se recomienda utilizar un $\alpha \in [1.5, 2]$ para $M > 6$. Además, por el Teorema 1.3.1 es fácil comprobar que las $R_{\phi_{\alpha}}$ son convexas para $\alpha \in [1, 2]$ que supone otra propiedad deseable para los estadísticos con parámetro α en este intervalo.

En el caso que se desee utilizar un α fuera del intervalo $[1.5, 2]$ la aproximación T_M aparece como una buena alternativa frente a la T_D al ser más sencilla de calcular.

Figura 4.5.9: Niveles de significación exacto y aproximados.
 $\gamma=0.1, n=10, M=3.$

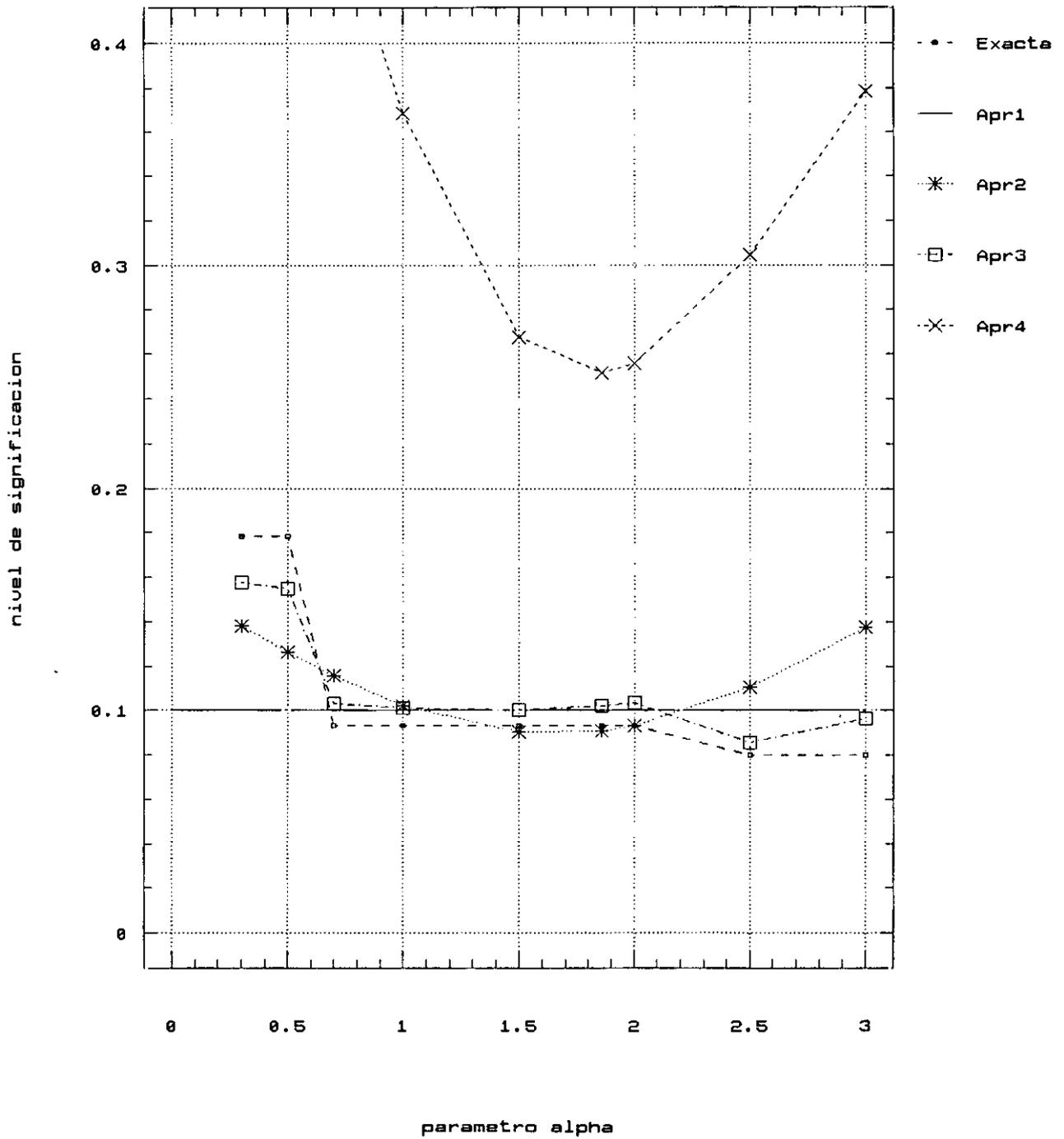


Figura 4.5.10: Niveles de significación exacto y aproximados.
 $\gamma=0.1, n=20, M=3.$

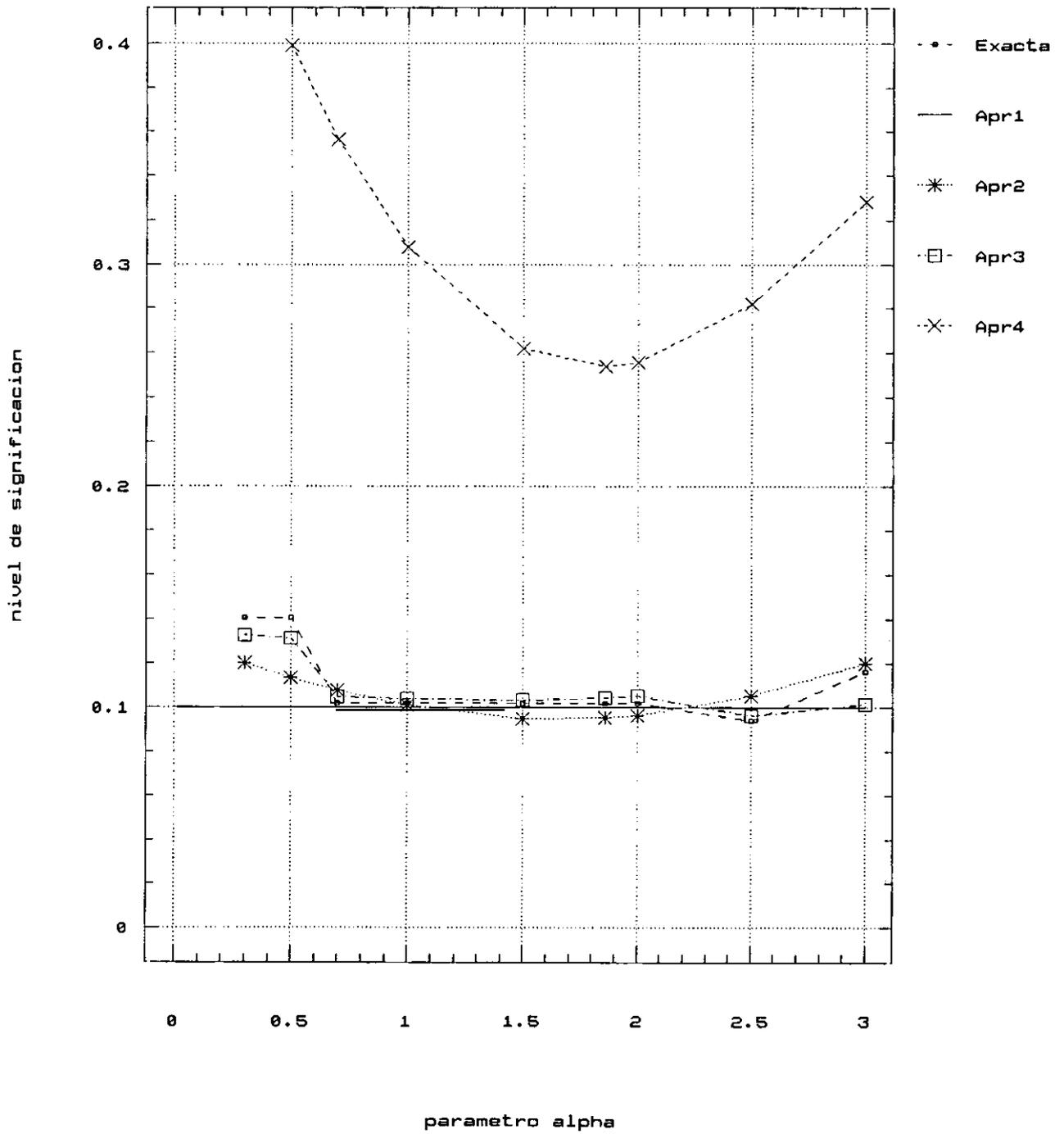


Figura 4.5.11: Niveles de significación exacto y aproximados.

$\gamma=0.1, n=10, M=4.$

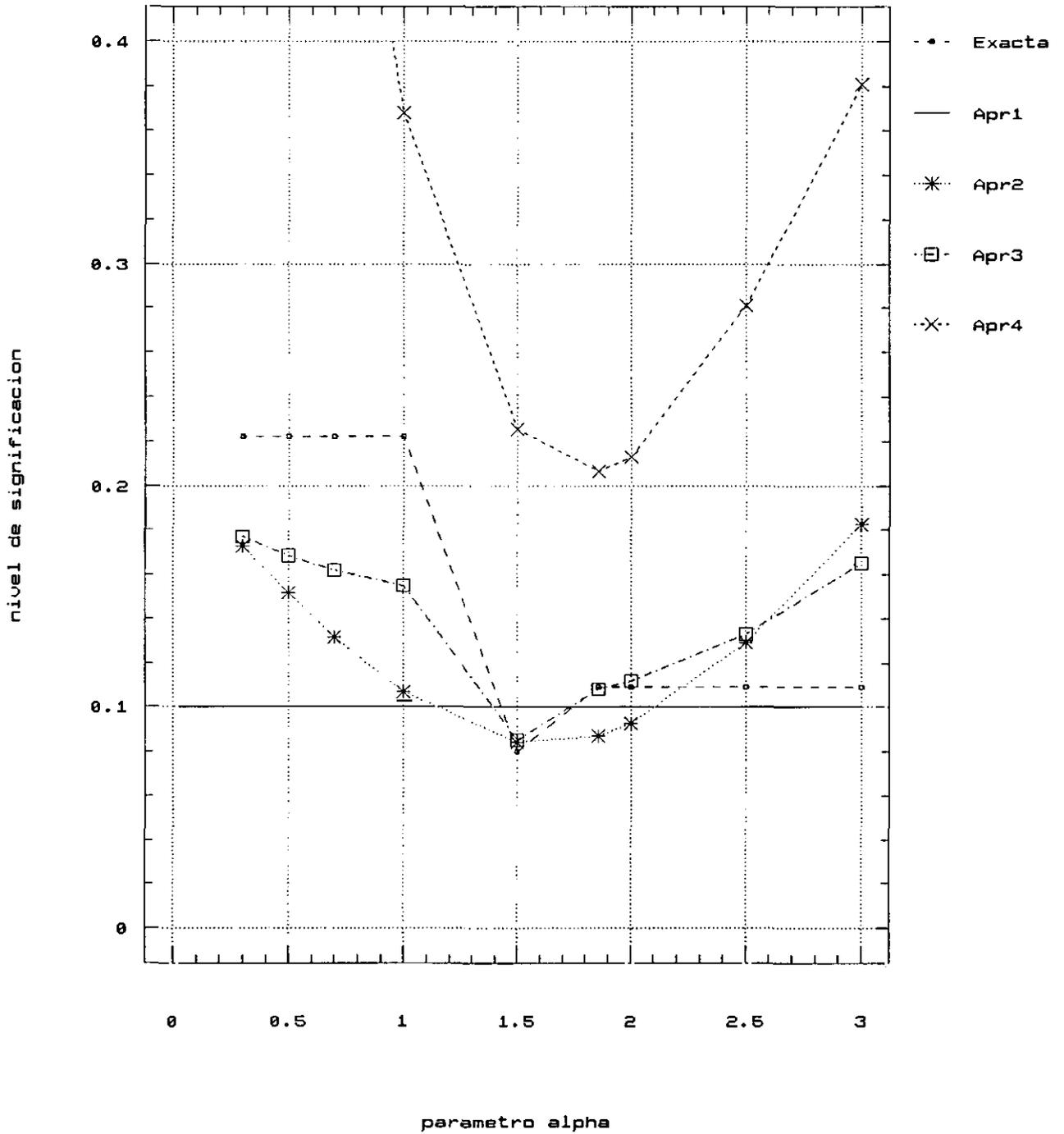


Figura 4.5.12: Niveles de significación exacto y aproximados.
 $\gamma=0.1$, $n=20$, $M=4$.

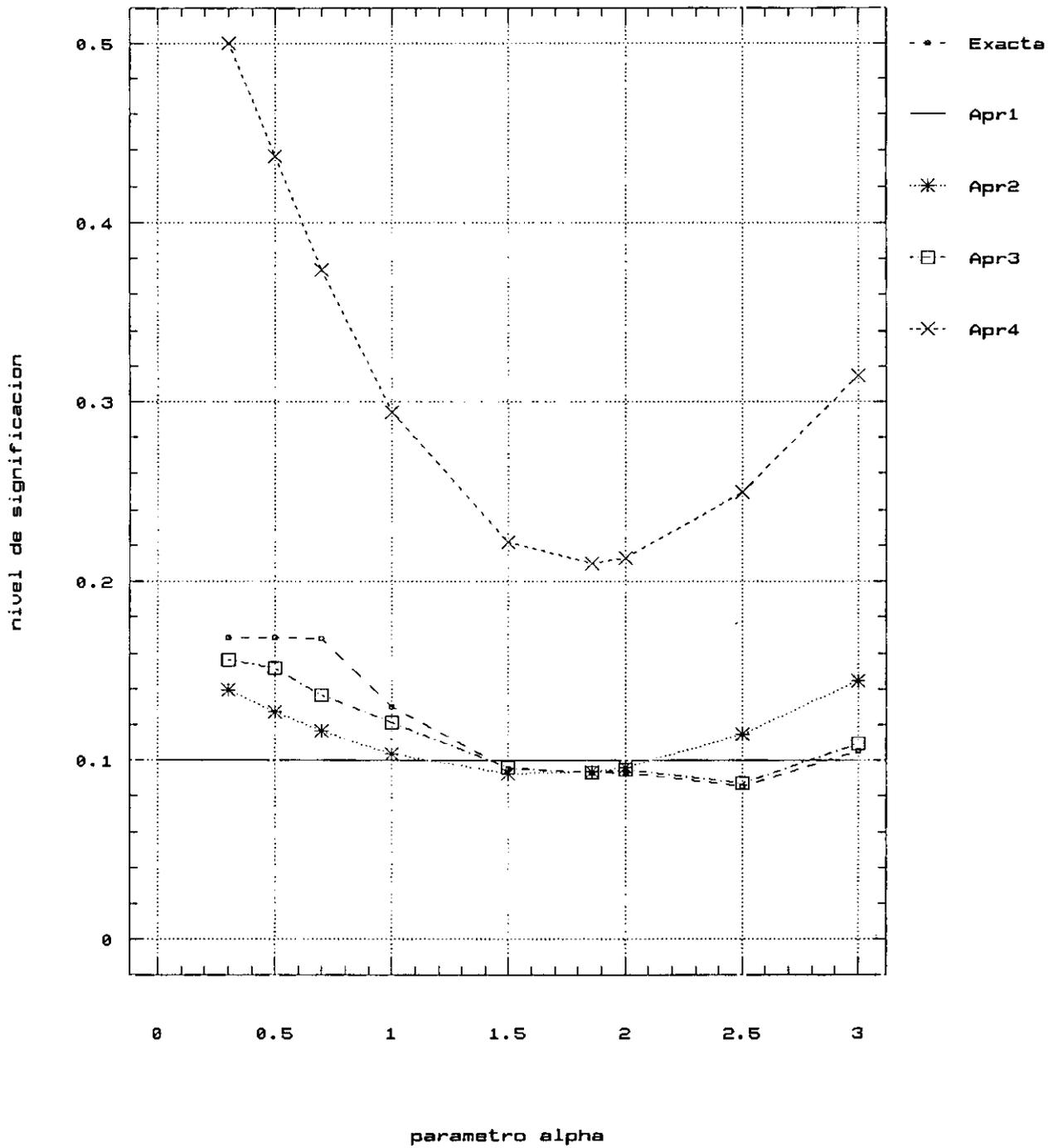


Figura 4.5.13: Niveles de significación exacto y aproximados.
 $\gamma=0.1, n=10, M=5.$

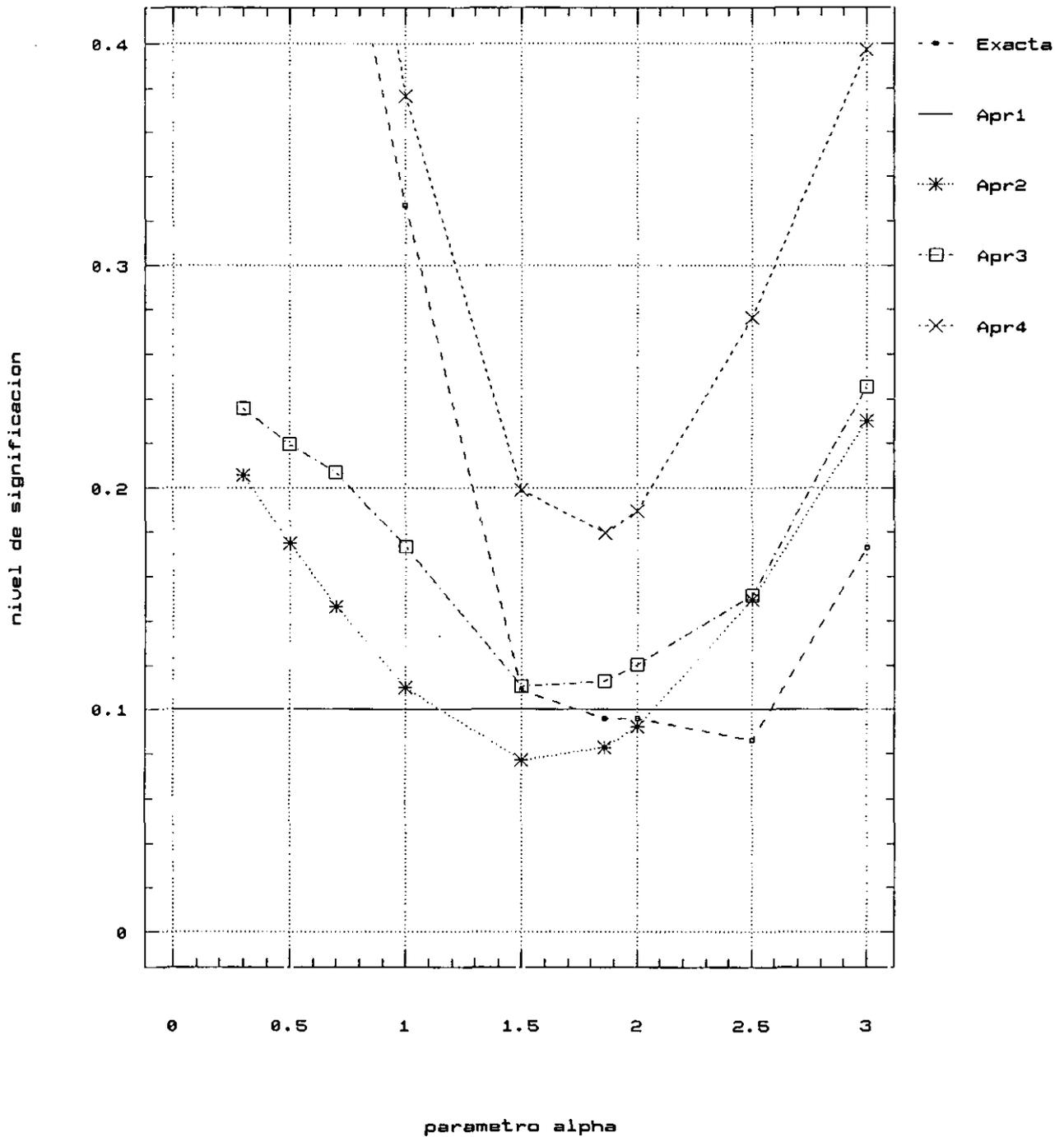


Figura 4.5.14: Niveles de significación exacto y aproximados.

$\gamma=0.1, n=20, M=5.$

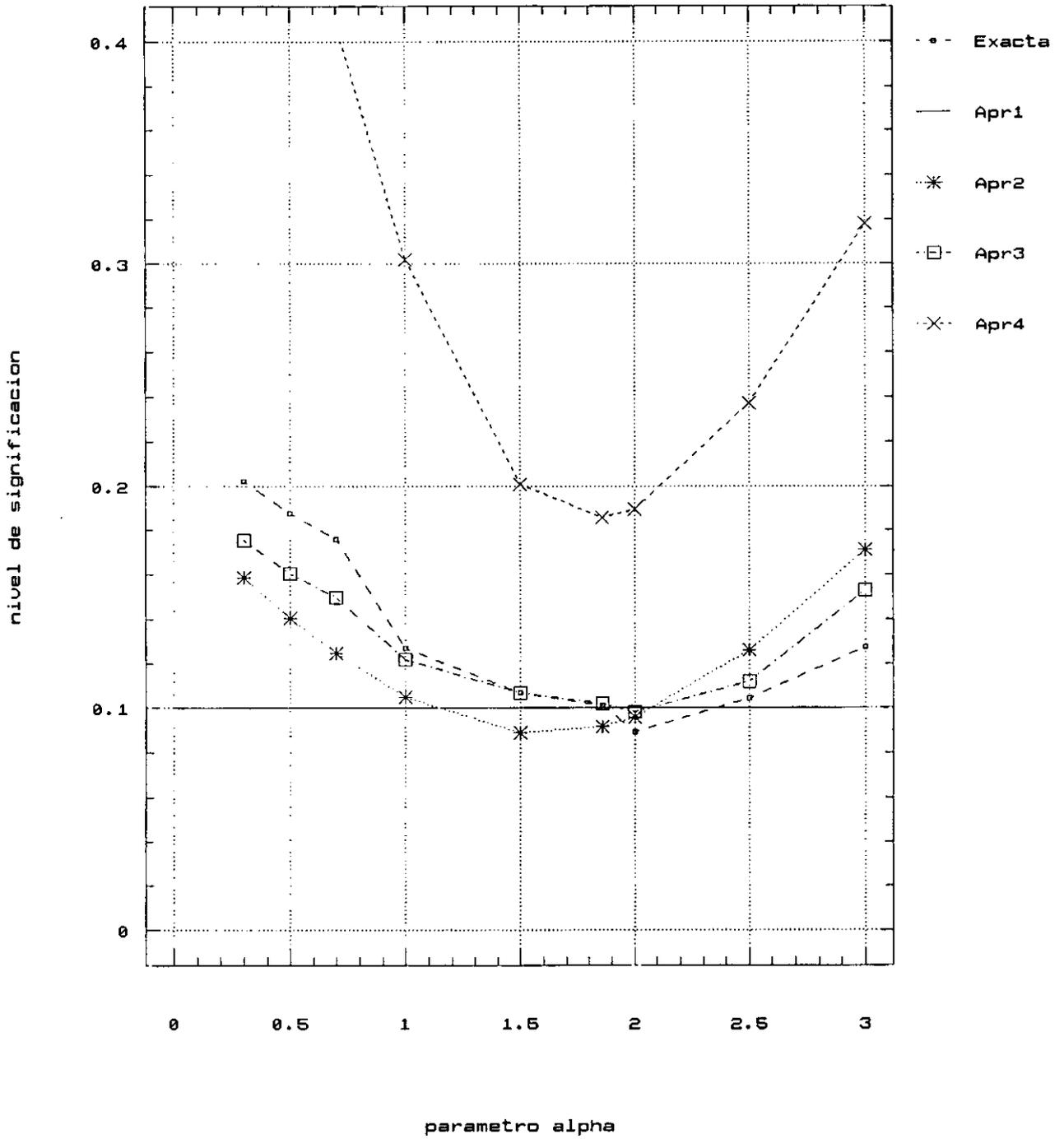


Figura 4.5.15: Niveles de significación exacto y aproximados.
 $\gamma=0.1, n=10, M=6.$

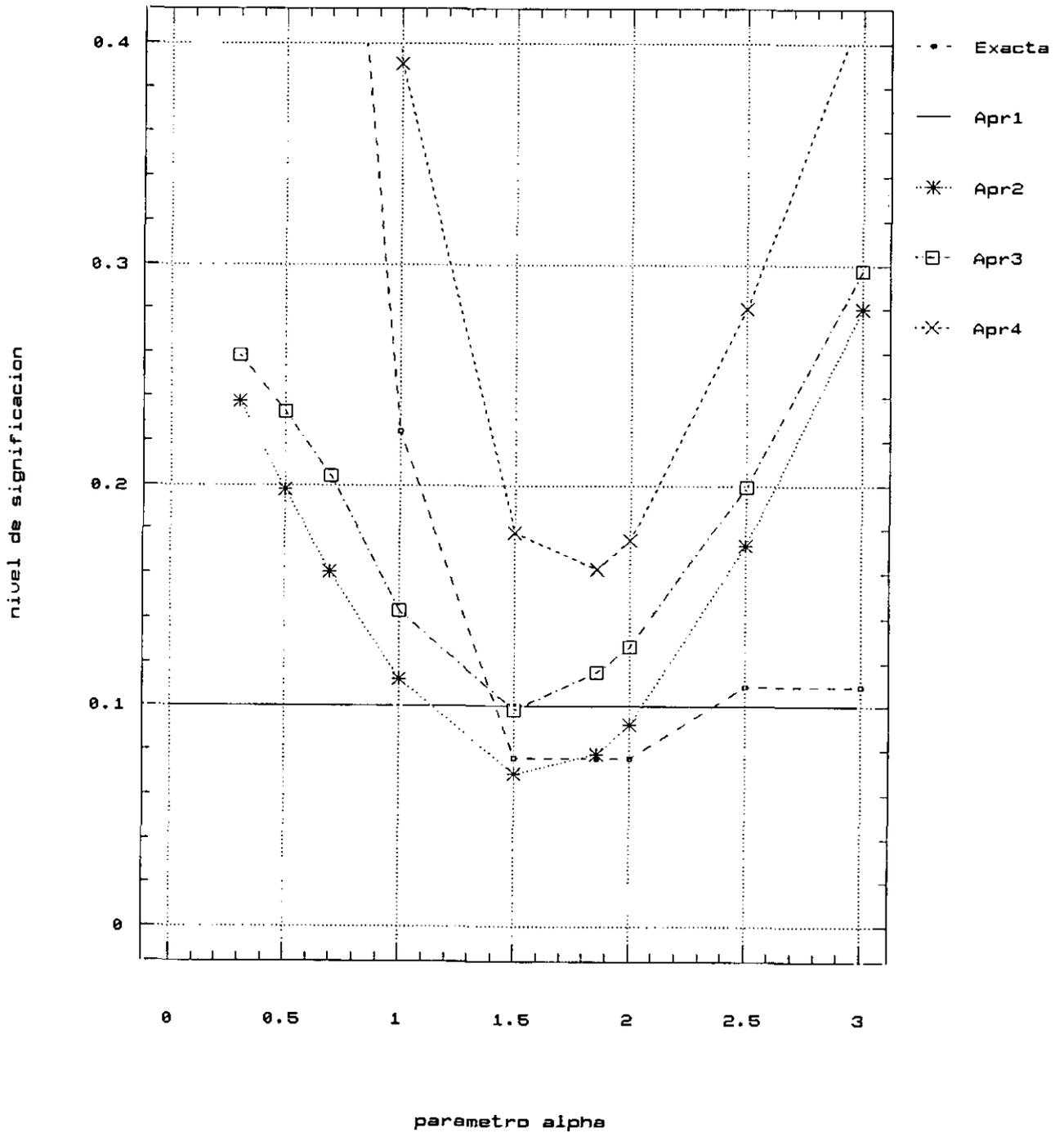


Figura 4.5.16: Niveles de significación exacto y aproximados.

$\gamma=0.1, n=20, M=6.$

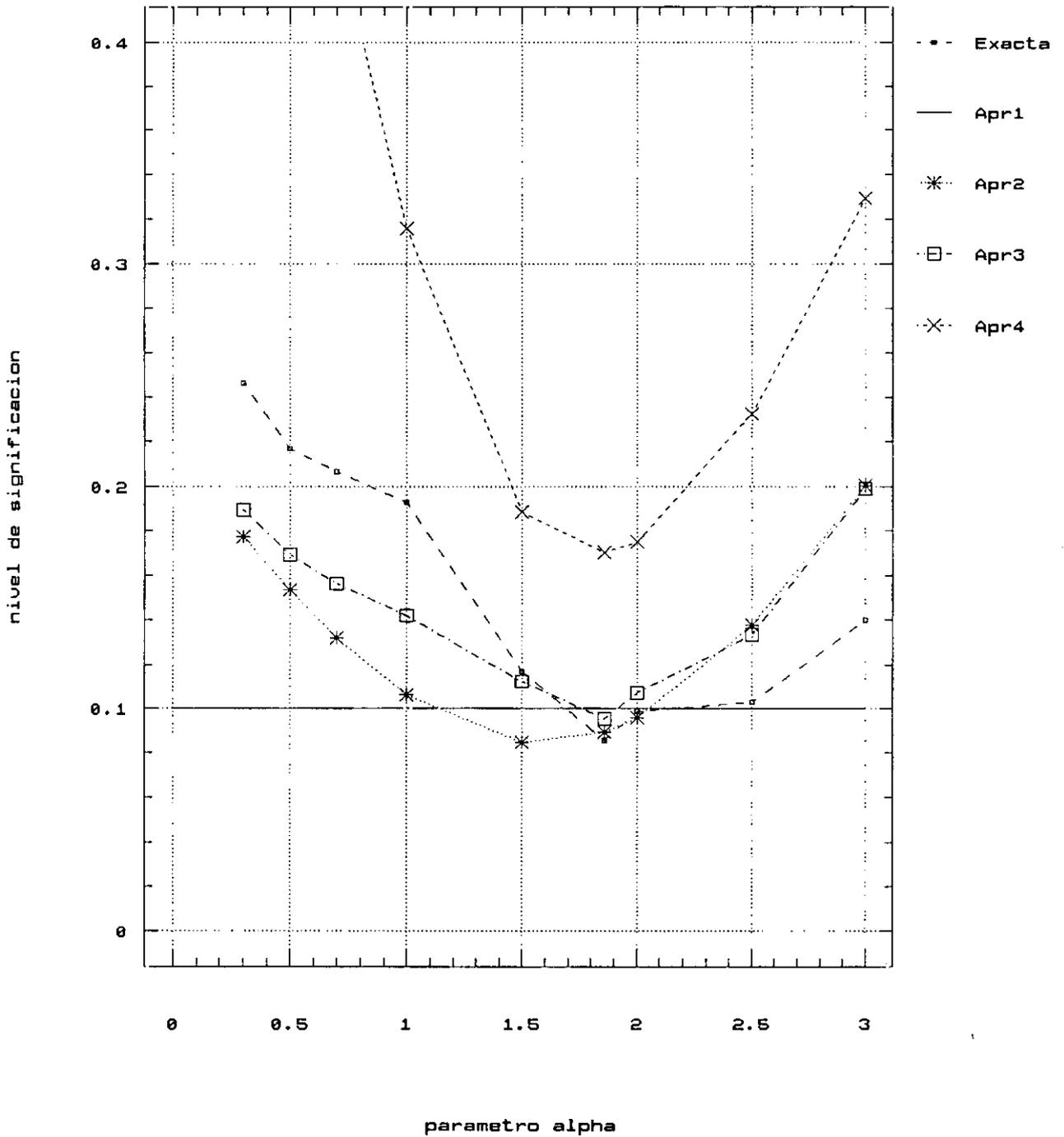


Figura 4.5.17: Niveles de significación exacto y aproximados.
 $\gamma=0.01, n=10, M=3$.

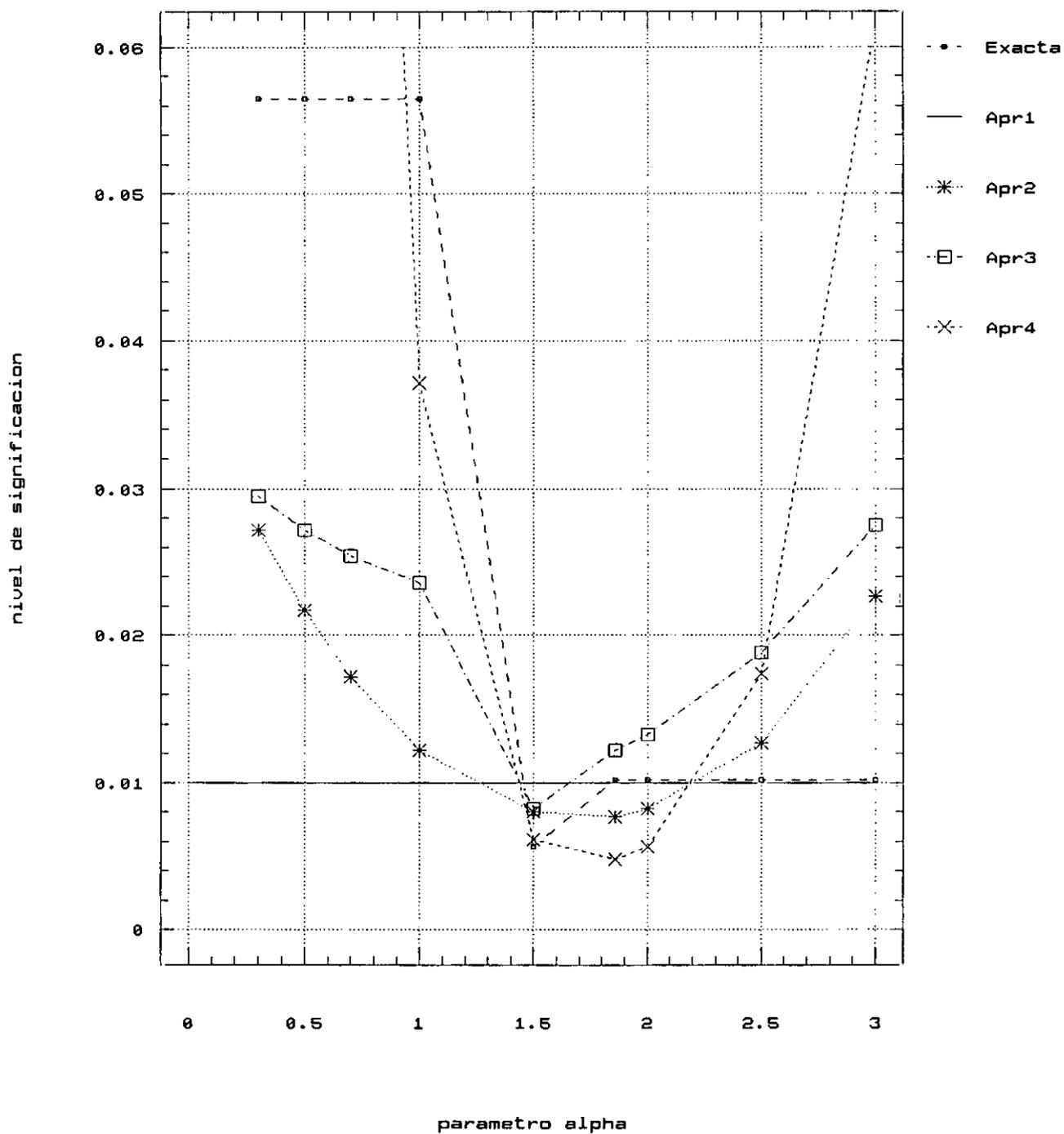


Figura 4.5.18: Niveles de significación exacto y aproximados.
 $\gamma=0.01, n=20, M=3.$

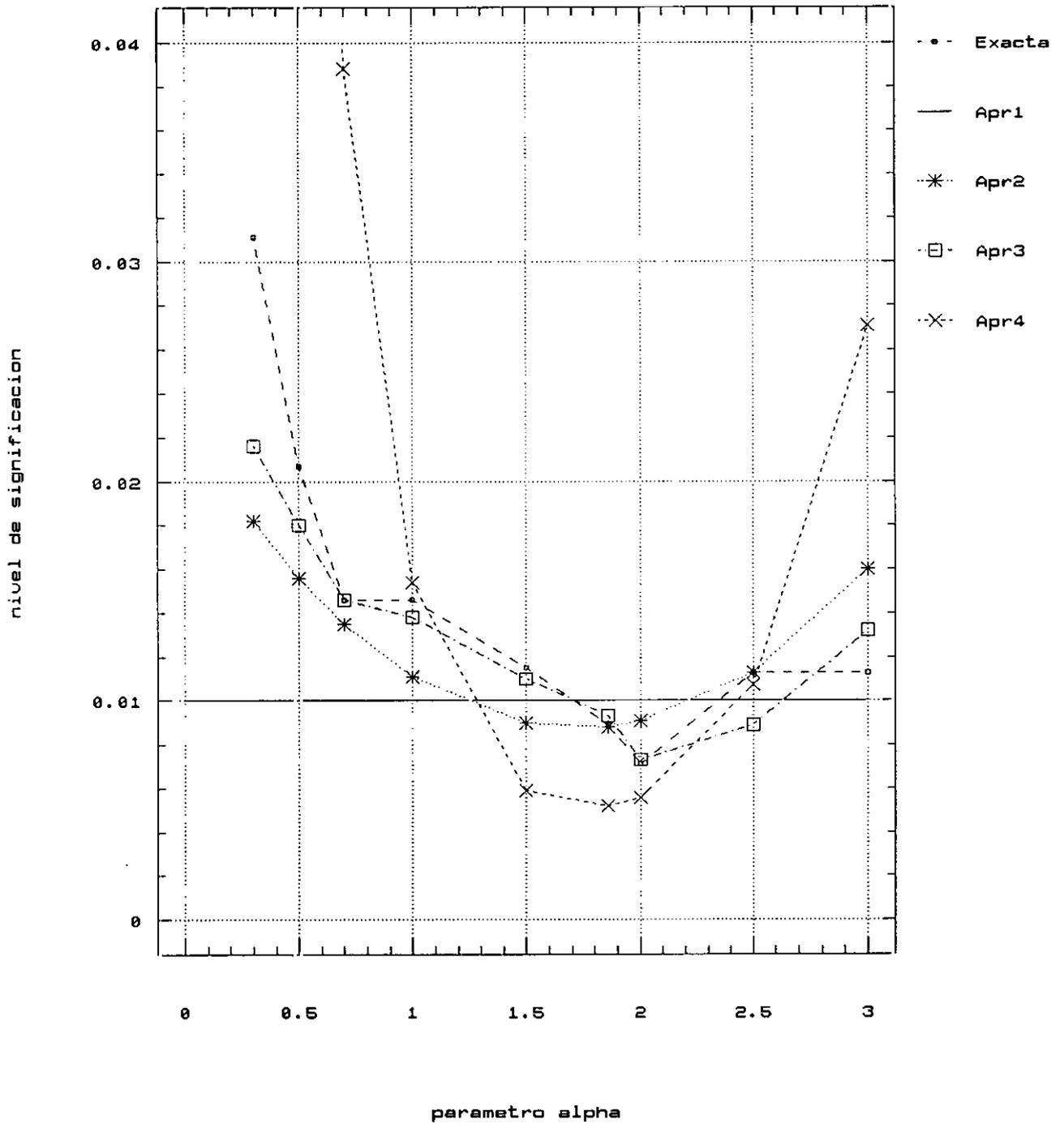


Figura 4.5.19: Niveles de significación exacto y aproximados.
 $\gamma=0.01, n=10, M=4.$

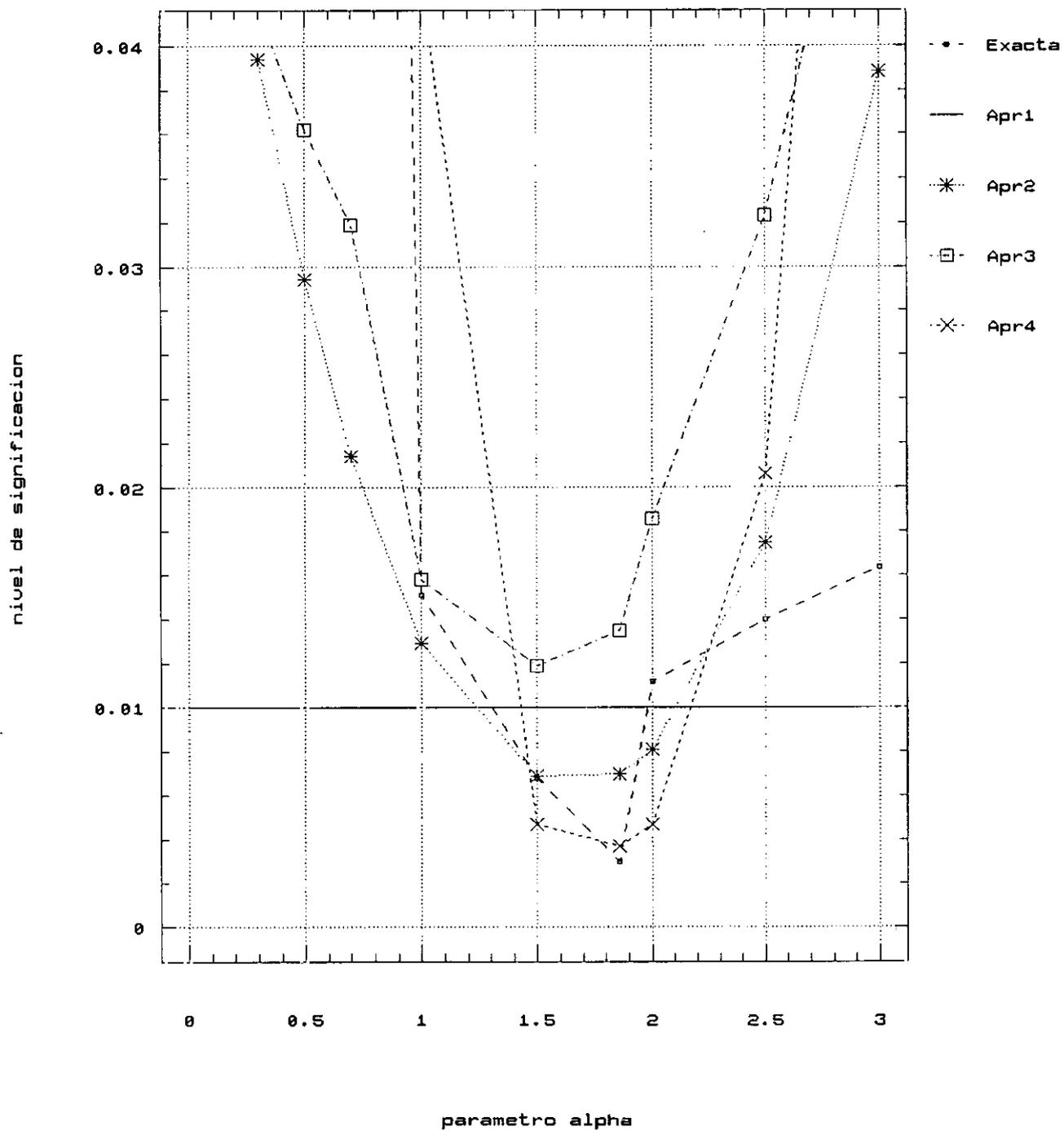


Figura 4.5.20: Niveles de significación exacto y aproximados.
 $\gamma=0.01$, $n=20$, $M=4$.

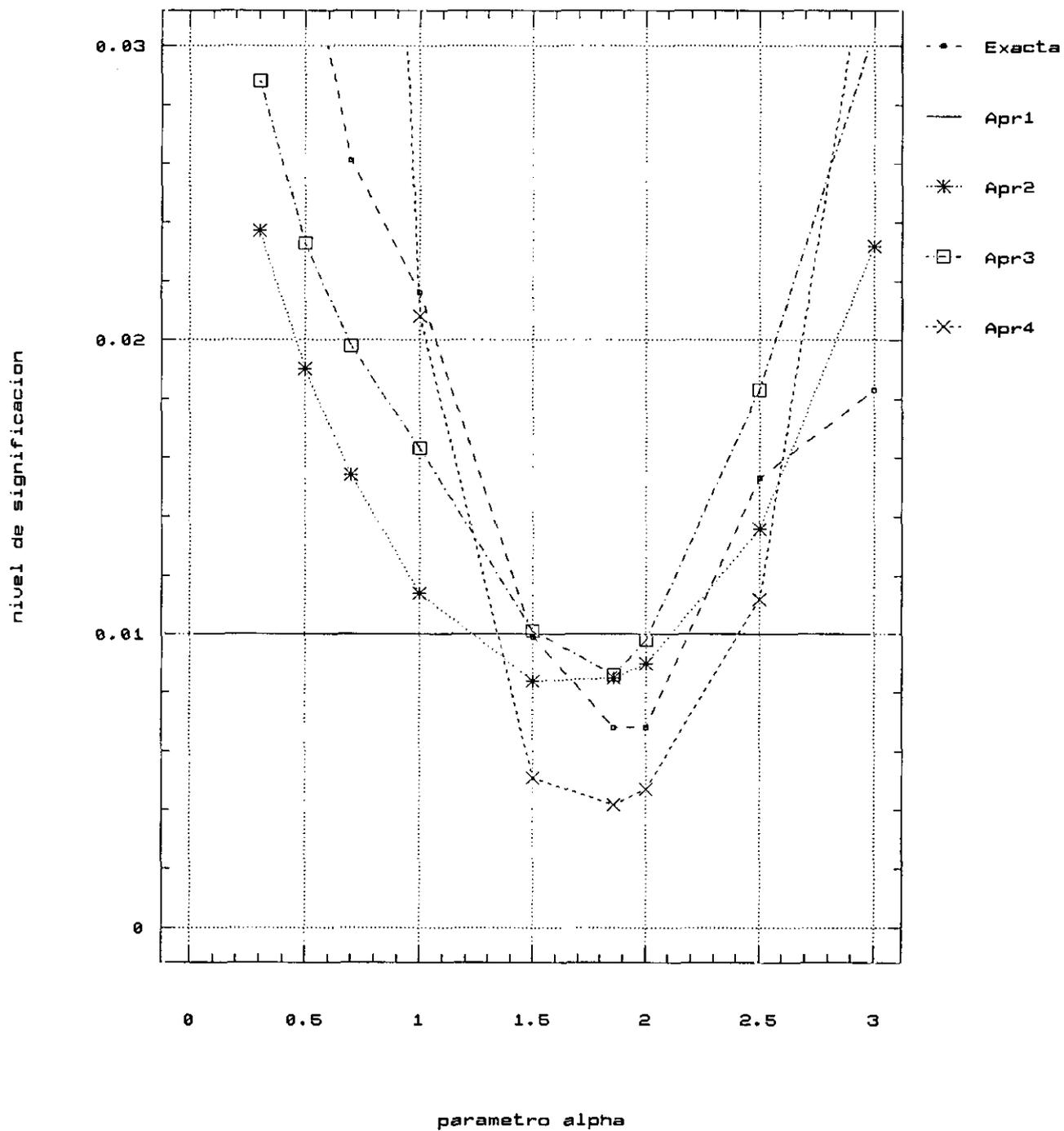


Figura 4.5.21: Niveles de significación exacto y aproximados.
 $\gamma=0.01$, $n=10$, $M=5$.

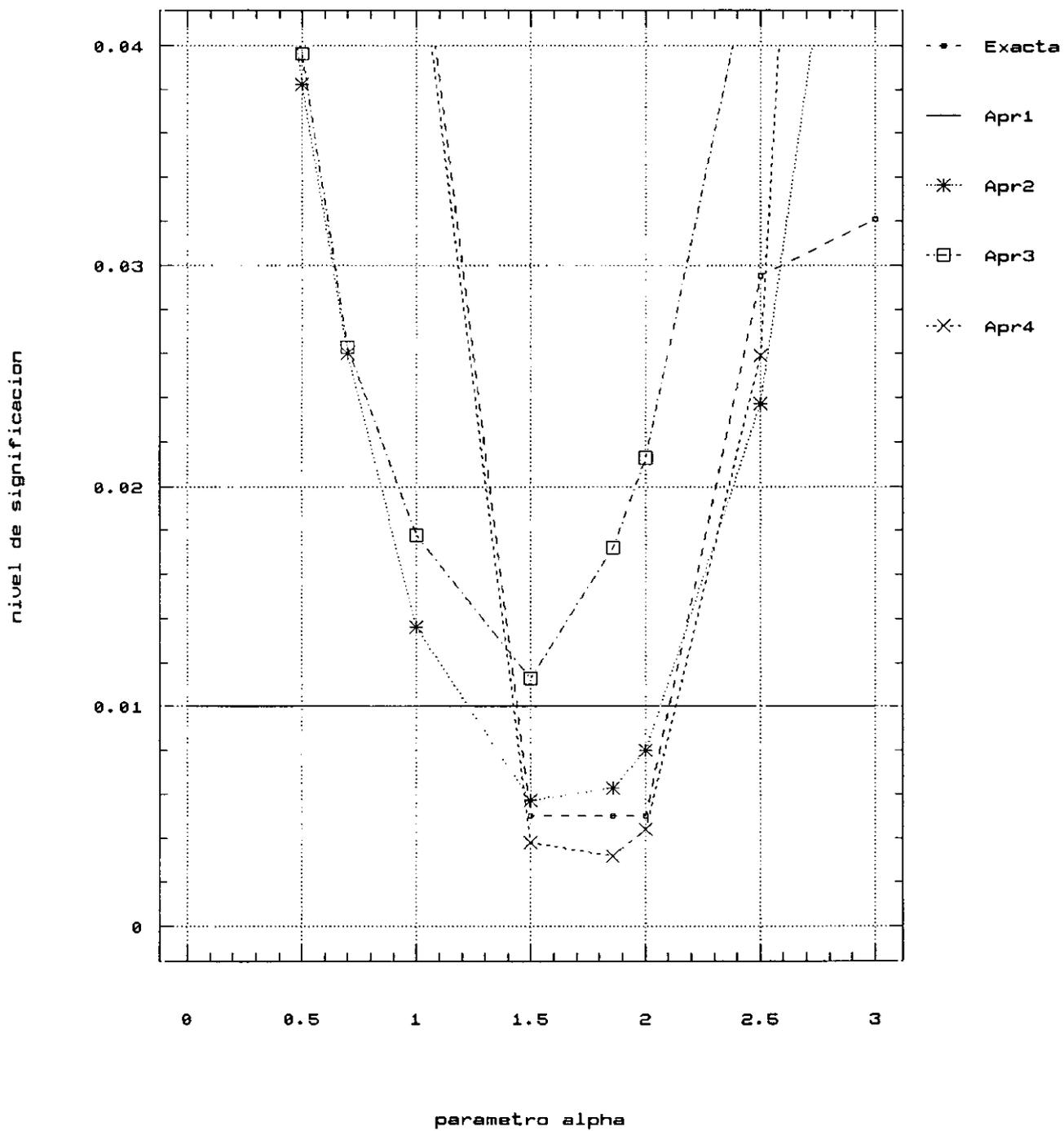


Figura 4.5.22: Niveles de significación exacto y aproximados.
 $\gamma=0.01$, $n=20$, $M=5$.

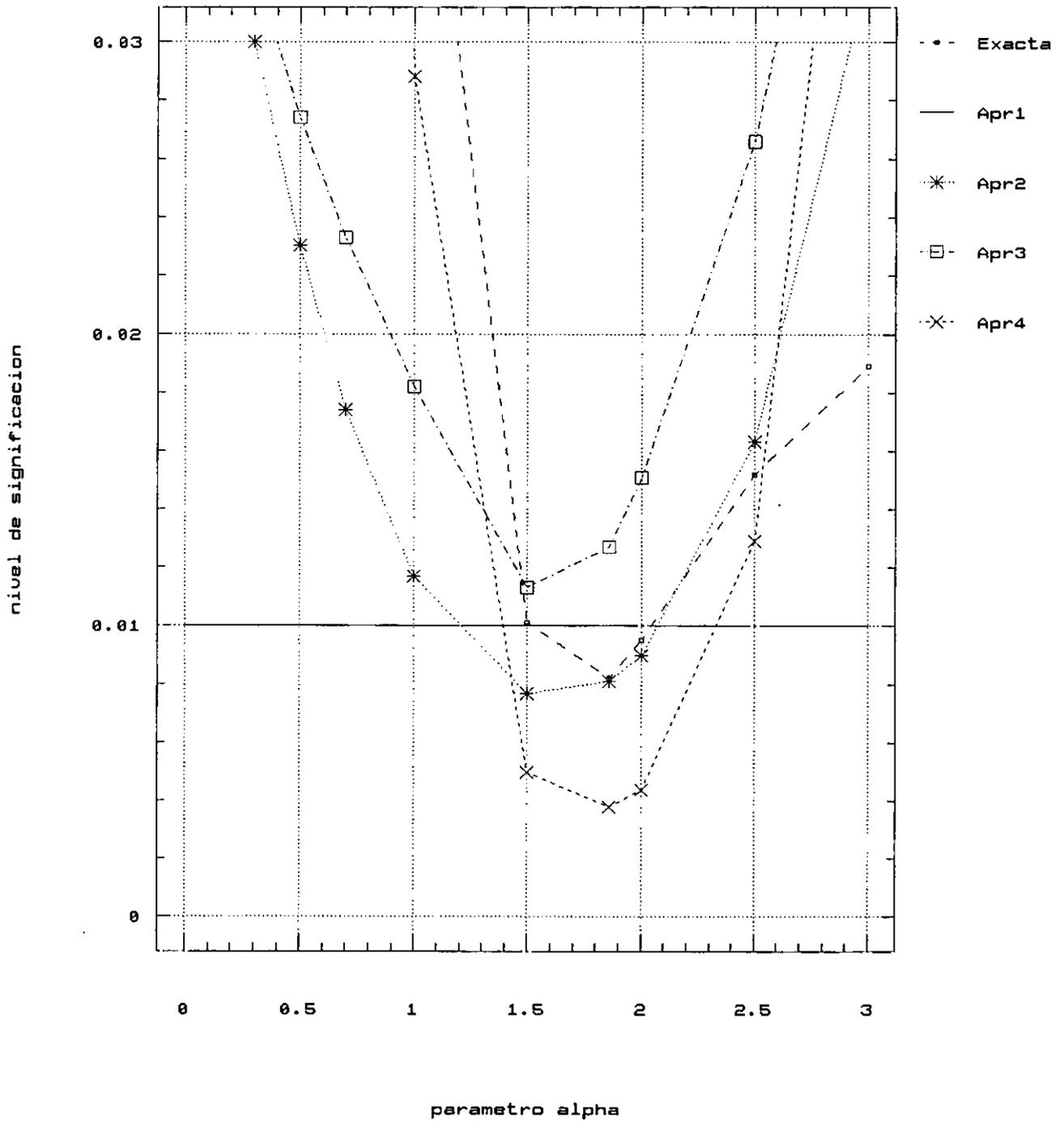


Figura 4.5.23: Niveles de significación exacto y aproximados.
 $\gamma=0.01, n=10, M=6.$

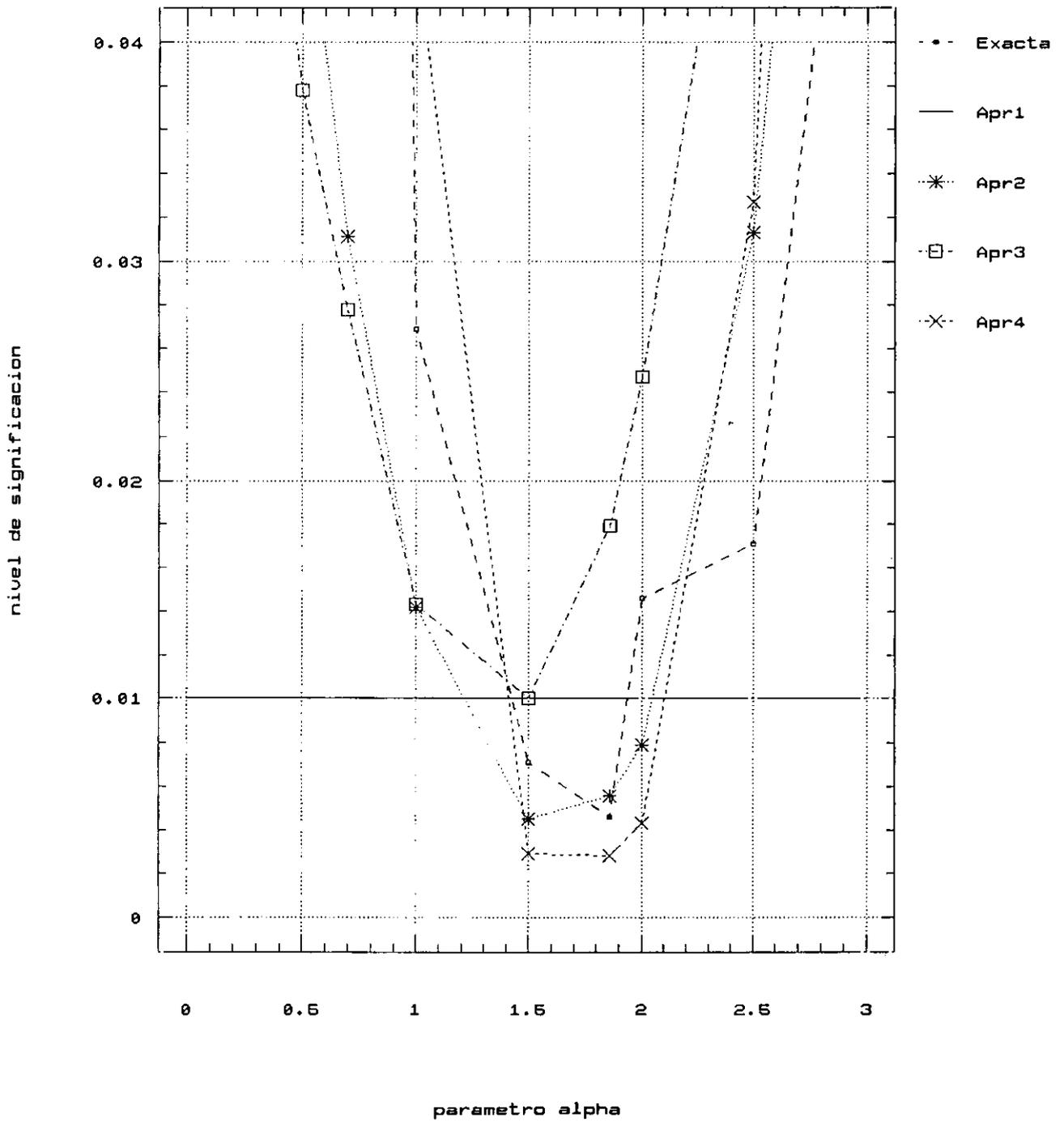
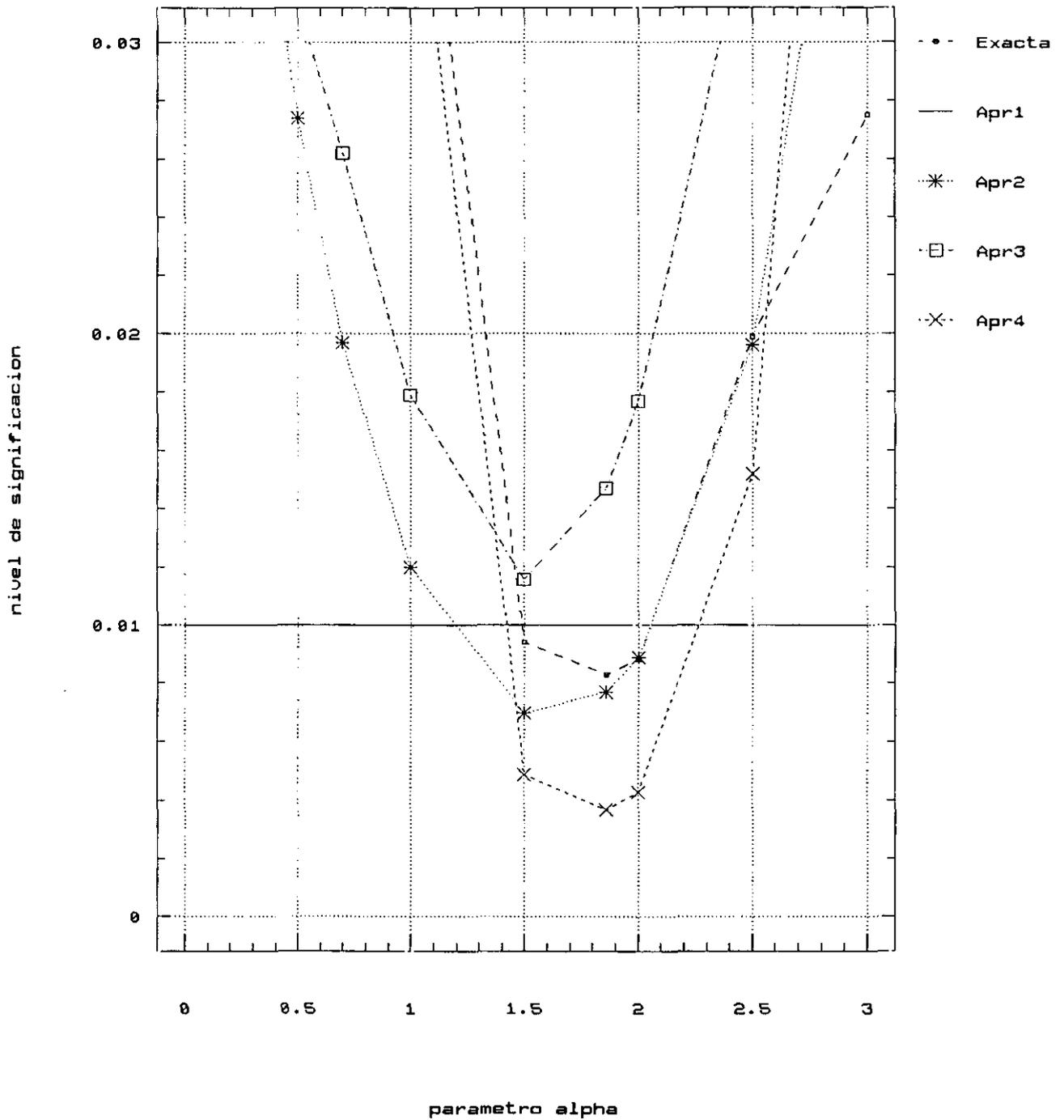


Figura 4.5.24: Niveles de significación exacto y aproximados.
 $\gamma=0.01$, $n=20$, $M=6$.



4.5.- Potencia exacta basada en regiones críticas exactas

Existen diferentes formas de investigar la eficiencia de los miembros de la familia S_{ϕ_α} para muestras pequeñas. Una solución a este problema consiste en encontrar condiciones bajo las cuales los resultados asintóticos obtenidos sean correctos para muestras pequeñas. Sin embargo esto no es fácil, Frosini (1976) demostró que son muy restrictivas las condiciones para que los resultados obtenidos en 3.2 sean apropiados con el estadístico X^2 de Pearson al considerar muestras pequeñas.

Otra solución más directa consiste en calcular la potencia exacta para cada estadístico S_{ϕ_α} al considerar diferentes alternativas sin utilizar ningún resultado asintótico. Este método es parecido al utilizado por West y Kempthorne (1972) para comparar los contrastes de la X^2 de Pearson y del cociente del logaritmo de verosimilitud para alternativas compuestas.

Por las razones dadas anteriormente se restringe el estudio al caso de hipótesis nula, H_0 , simétrica y se consideran alternativas donde la probabilidad de una de las clases se perturba y el resto se ajustan de forma que sumen 1 para la familia de estadísticos S_{ϕ_α} . En particular, se trabaja con las alternativas:

$$H_1: \pi_i = \begin{cases} \frac{M-1-\delta}{M(M-1)} & \text{si } i=1, \dots, M-1 \\ \frac{1+\delta}{M} & \text{si } i=M, \end{cases} \quad (4.5.1)$$

donde $-1 \leq \delta \leq M-1$ es fijo.

Para calcular la potencia exacta de cada estadístico S_{ϕ_α} es necesario fijar un tamaño del test γ y calcular la región crítica

asociada. La forma más sencilla de calcular ésta sería utilizando una de las aproximaciones estudiadas. Sin embargo debido a que el error de aproximación dependería de α , ésto influiría en las comparaciones de las potencias que también dependen de α . Por tanto se calcula la región crítica exacta utilizando T_E . Además, debido a que es bastante improbable que un test de tamaño γ exacto no aleatorizado exista, se utiliza un test aleatorizado basado en S_{ϕ_α} de tamaño γ que se obtiene de la siguiente forma:

Sea $c_\alpha(\gamma)$ un valor posible de $S_{\phi_\alpha}(X/n, \pi_0)$ tal que

$$P\left[S_{\phi_\alpha}(X/n, \pi_0) > c_\alpha(\gamma) \mid H_0\right] = \gamma_{1,\alpha}$$

y

(4.5.2)

$$P\left[S_{\phi_\alpha}(X/n, \pi_0) \geq c_\alpha(\gamma) \mid H_0\right] = \gamma_{2,\alpha}$$

siendo $\gamma_{1,\alpha} < \gamma \leq \gamma_{2,\alpha}$. Entonces dado un vector x , el test aleatorizado de tamaño γ rechaza H_0 con probabilidad

$$\begin{cases} 1, & \text{si } S_{\phi_\alpha}(X/n, \pi_0) > c_\alpha(\gamma) \\ \frac{\gamma - \gamma_{1,\alpha}}{\gamma_{2,\alpha} - \gamma_{1,\alpha}}, & \text{si } S_{\phi_\alpha}(X/n, \pi_0) = c_\alpha(\gamma). \\ 0, & \text{si } S_{\phi_\alpha}(X/n, \pi_0) < c_\alpha(\gamma) \end{cases}$$

De (4.5.2) se tiene que el tamaño del test es

$$1\gamma_{1,\alpha} + \frac{\gamma - \gamma_{1,\alpha}}{\gamma_{2,\alpha} - \gamma_{1,\alpha}}(\gamma_{2,\alpha} - \gamma_{1,\alpha}) = \gamma$$

como se quería.

Además, si se define

$$\beta_{1,\alpha} = P\left[S_{\phi_\alpha}(X/n, \pi_0) > c_\alpha(\gamma) \mid H_1\right]$$

y

$$\beta_{2,\alpha} = P\left[S_{\phi_\alpha}(X/n, \pi_0) \geq c_\alpha(\gamma) \mid H_1\right]$$

se tiene que la potencia del test aleatorizado de tamaño γ viene dada por

$$\beta_\alpha = \beta_{1,\alpha} + \frac{\gamma - \gamma_{1,\alpha}}{\gamma_{2,\alpha} - \gamma_{1,\alpha}} \left[\beta_{2,\alpha} - \beta_{1,\alpha} \right].$$

Las potencias exactas para el test aleatorizado (4.5.2) de tamaño 0.05 frente a las alternativas (4.5.1) aparecen en las Tablas de la 4.5.1 a la 4.5.8 para diferentes valores de δ y de α . Los valores de n y M son los mismos que los utilizados en el apartado anterior.

Para las alternativas $\delta < 0$ la potencia decrece cuando α crece y para $\delta > 0$, al contrario, la potencia crece cuando α crece, para $n=20$. Un comportamiento similar tienen las potencias para $n=10$ pero con dos salvedades: una de ellas obvia, los valores de las potencias son menores y la otra es que para $\alpha > 2$ y $\delta > 0$ la potencia deja de crecer o crece poco.

Por tanto las recomendaciones son que para las alternativas del tipo $\delta < 0$ se elija el α lo menor posible para obtener la mejor potencia. Mientras que para las alternativas $\delta > 0$ se debe elegir el α tan grande como sea posible, no obstante si n es pequeño con respecto a M es aconsejable que $\alpha \leq 2$.

Si se estuviera interesado en encontrar un test con potencia razonable frente a este tipo de alternativas, δ cualquier valor, se observa de forma casi general que hay un decrecimiento notable de la potencia cuando α aumenta de 2.5 a 5 y un crecimiento notable cuando α disminuye de 1 a 13/7. Esto sugiere que en este caso se debería elegir un $\alpha \in [13/7, 2.5]$.

Tabla 4.5.1: Potencia exacta para el test aleatorizado (4.5.2) de tamaño .05 frente a las alternativas (4.5.1). $n=10$, $M=3$.

α	$\delta = -0.9$	$\delta = -0.5$	$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$
0.3	0.7137	0.1701	0.1090	0.3037	0.6761
0.5	0.7137	0.1701	0.1090	0.3037	0.6761
0.7	0.7137	0.1701	0.1090	0.3037	0.6761
1.	0.6326	0.1518	0.1125	0.3056	0.6763
13/7	0.2917	0.1365	0.1141	0.4160	0.8445
2.	0.2917	0.1365	0.1141	0.4160	0.8445
2.5	0.2824	0.1168	0.1771	0.5596	0.9302
5	0.2082	0.1159	0.1481	0.4945	0.8915

Tabla 4.5.2: Potencia exacta para el test aleatorizado (4.5.2) de tamaño .05 frente a las alternativas (4.5.1). $n=20$, $M=3$.

α	$\delta = -0.9$	$\delta = -0.5$	$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$
0.3	0.9514	0.2986	0.1902	0.6301	0.9774
0.5	0.9523	0.3106	0.2230	0.7284	0.9916
0.7	0.9523	0.3106	0.2230	0.7284	0.9916
1.	0.9523	0.3106	0.2230	0.7284	0.9916
13/7	0.9165	0.2811	0.2668	0.8148	0.9972
2.	0.9165	0.2811	0.2668	0.8148	0.9972
2.5	0.9165	0.2811	0.2668	0.8148	0.9972
5	0.7792	0.2205	0.2847	0.8274	0.9975

Tabla 4.5.3: Potencia exacta para el test aleatorizado (4.5.2) de tamaño .05 frente a las alternativas (4.5.1). $n=10$, $M=4$.

α	$\delta = -0.9$	$\delta = -0.5$	$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$
0.3	0.2001	0.0909	0.0967	0.2469	0.4970
0.5	0.2039	0.0977	0.0884	0.2353	0.4890
0.7	0.2472	0.1027	0.0929	0.2418	0.4933
1.	0.2001	0.0909	0.0967	0.2469	0.4970
13/7	0.1825	0.0875	0.1039	0.2903	0.5829
2.	0.1825	0.0875	0.1039	0.2903	0.5829
2.5	0.1990	0.0931	0.0975	0.2751	0.5646
5	0.1471	0.0871	0.0852	0.2459	0.5305

Tabla 4.5.4: Potencia exacta para el test aleatorizado (4.5.2) de tamaño .05 frente a las alternativas (4.5.1). $n=20$, $M=4$.

α	$\delta = -0.9$	$\delta = -0.5$	$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$
0.3	0.7509	0.1825	0.1280	0.3810	0.7351
0.5	0.7382	0.1767	0.1332	0.3992	0.7512
0.7	0.7338	0.1749	0.1390	0.4300	0.7960
1.	0.7334	0.1786	0.1425	0.4401	0.8018
13/7	0.6367	0.1641	0.1572	0.5058	0.8646
2.	0.5766	0.1587	0.1617	0.5156	0.8697
2.5	0.3354	0.1280	0.1742	0.5595	0.8978
5	0.2172	0.1121	0.1762	0.5693	0.9045

Tabla 4.5.5: Potencia exacta para el test aleatorizado (4.5.2) de tamaño .05 frente a las alternativas (4.5.1). $n=10$, $M=5$.

α	$\delta = -0.9$	$\delta = -0.5$	$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$
0.3	0.1619	0.0797	0.0721	0.1414	0.2618
0.5	0.1521	0.0760	0.0733	0.1423	0.2622
0.7	0.1538	0.0765	0.0741	0.1440	0.2639
1.	0.1582	0.0791	0.0741	0.1552	0.3039
13/7	0.1305	0.0758	0.0846	0.2078	0.4186
2.	0.1305	0.0756	0.0872	0.2144	0.4276
2.5	0.1246	0.0735	0.0851	0.2076	0.4177
5	0.1199	0.0759	0.0816	0.2027	0.4131

Tabla 4.5.6: Potencia exacta para el test aleatorizado (4.5.2) de tamaño .05 frente a las alternativas (4.5.1). $n=20$, $M=5$.

α	$\delta = -0.9$	$\delta = -0.5$	$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$
0.3	0.5952	0.1289	0.0783	0.1470	0.2573
0.5	0.5952	0.1289	0.0783	0.1470	0.2577
0.7	0.5689	0.1244	0.0797	0.1566	0.3063
1.	0.5627	0.1281	0.0886	0.2204	0.4705
13/7	0.2839	0.1081	0.1218	0.3677	0.6950
2.	0.2684	0.1063	0.1229	0.3725	0.7007
2.5	0.1958	0.0965	0.1255	0.3897	0.7236
5	0.1432	0.0860	0.1294	0.4091	0.7484

Tabla 4.5.7: Potencia exacta para el test aleatorizado (4.5.2) de tamaño .05 frente a las alternativas (4.5.1). $n=10$, $M=6$.

α	$\delta = -0.9$	$\delta = -0.5$	$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$
0.3	0.1093	0.0690	0.0703	0.1396	0.2627
0.5	0.1113	0.0694	0.0705	0.1400	0.2631
0.7	0.1061	0.0675	0.0716	0.1418	0.2650
1.	0.1054	0.0670	0.0723	0.1430	0.2665
13/7	0.1067	0.0671	0.0743	0.1522	0.2890
2.	0.1036	0.0676	0.0726	0.1495	0.2864
2.5	0.1037	0.0683	0.0708	0.1497	0.2928
5	0.1031	0.0682	0.0700	0.1473	0.2889

Tabla 4.5.8: Potencia exacta para el test aleatorizado (4.5.2) de tamaño .05 frente a las alternativas (4.5.1). $n=20$, $M=6$.

α	$\delta = -0.9$	$\delta = -0.5$	$\delta = 0.5$	$\delta = 1$	$\delta = 1.5$
0.3	0.2680	0.0961	0.0792	0.1626	0.2967
0.5	0.2611	0.0945	0.0799	0.1642	0.2986
0.7	0.2672	0.0960	0.0805	0.1692	0.3125
1.	0.2595	0.0950	0.0845	0.1923	0.3780
13/7	0.1904	0.0889	0.1003	0.2789	0.5530
2.	0.1791	0.0872	0.1019	0.2847	0.5610
2.5	0.1651	0.0853	0.1006	0.2849	0.5637
5	0.1223	0.0773	0.1033	0.2985	0.5855

REFERENCIAS

- Ali M.S. y Silvey, S.D. (1966): A general class of coefficients of divergence of on distribution from another. *Journal of the Royal Statistical Society, Ser. B*, 28, 131-140.
- Aczél, J. (1966): *Lectures on Functional Equations and their Applications*. Academic Press, New York.
- Agresti, A. (1984): *Analysis of Ordinal Categorical Data*. New York, John Wiley.
- Amari, S. (1990): *Diferential Geometric Methods in Statistics, 2nd ed.* Springer, New York.
- Barron, A. R. (1985): The strong ergodic theorem for densities: generalized Shannon-McMillan-Breiman theorem. *Annals of Probability*, 13, 1292-1303.
- Bednarski, T. y Ledwina, T. (1978): A note on a biasedness of tests of fit. *Mathematische Operationsforschung und Statistik, Series Statistics* 9, 191-193.
- Behboodian, J. (1970): On a mixture of normal distributions. *Biometrika*, 57, 215-217.
- Bhattacharyya, A. (1946): On a measure of divergence between two statistical populations defined by their probability distributions. *Bull. Calcutta. Math. Soc.*, 35, 99-109.
- Billingsley, P. (1961): Statistical methods in Markov chains. *Annals of Mathematical Statistics*, 32, 12-40.
- Billingsley, P. (1968): *Convergence of Probability Measures*. John Wiley, New York.
- Birch, M.W. (1964): A new proof of the Pearson-Fisher theorem. *Annals of Mathematical Statistics*, 35, 817-824.
- Bishop, Y.M.M., Fienberg, S.E. y Holland, P.W.(1975): *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, Mass.: MIT Press.
- Blahut, R.E. (1987): *Principles and Practice of Information Theory*. Adisson-Wesley, Reading, MA.
- Bregman, L.M. (1967) : The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *U.S.S.R. Comput. Math. and Math.*

- Phys.*, 7, 200-217.
- Burbea, J. y Rao, C.R. (1982): On the convexity of some divergence measures based on Entropy functions. *IEEE Transactions on Information Theory*, 28, 489-495.
- Chapman, J.W. (1976): A comparison of the χ^2 , $-2\log R$, and the multinomial probability criteria for significance testing when expected frequencies are small. *Journal of the American Statistical Association*, 71, 854-863.
- Clarke, B.S. y Barron, A.R. (1990): Information-theoretic asymptotics of Bayes methods. *IEEE Transaction on Information Theory*, 36, 453-471.
- Cochran, W.G. (1952): The χ^2 test of goodness of fit. *Annals of Mathematical Statistics*, 23, 315-345.
- Cohen, A.C. (1967): Estimation in mixtures of two normal distributions. *Technometrics*, 9, 15-28.
- Cohen, A. y Sackrowitz, H. B. (1975): Unbiasedness of the chi-square, likelihood ratio, and other goodness of fit tests for the equal cell case. *Annals of Statistics*, 3, 959-964.
- Cohen, J.E.; Derriennic, Y; Zbaganu, Gh. (1993): Majorization, monotonicity of relative entropy and stochastic matrices. *Contemporary Mathematics*, 149, 251-259.
- Cover, T.M. y Thomas, J.B. (1991): *Elements of Information Theory*. Wiley, New York.
- Cox, D.R. (1970): *The Analysis of Binary Data*. London, Methuen.
- Cressie, N. y Read, T.R.C. (1984): Multinomial goodness of fit test. *Journal of the Royal Statistical Society, Ser. B*, 46, 440-464.
- Cressie, N. y Read, T.R.C. (1988): *Goodness of Fit Statistics for Discrete Multivariate Data*. Springer, New York.
- Csiszár, I. (1963): Eine Informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markhoffschen Ketten. *Publ. Math. Inst. Hungar. Acad. Sci. Ser. A*, 8, 85-108.
- Csiszár, I. (1967): Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar*, 2, 299-318.

- Csiszár, I. (1991): Why least squares and maximum entropy?. An axiomatic approach to inference for linear inverse problems. *Annals of Statistics*, 19, 2031-2066.
- Csiszár, I. (1994): Maximum entropy and related methods. *Trans. 12th Prague Conf. on Information Theory* (Eds. J.A. Višek and P. Lachout), 58-62. Czech Acad. Sci., Prague.
- Csiszár, I. (1995): Generalized cutoff rates and Rényi's information measures. *IEEE Transactions on Information Theory*, 41, 26-34.
- Dannenbring, D.G. (1977): Procedures for estimating optimal solution values for large combinatorial problems. *Man. Sci.*, 23, 1273-1283.
- Day, N. E. (1969): Estimating the components of a mixture of normal distributions. *Biometrika*, 56(3), 463-474.
- Dik, J.J. y Gunst, M.C.M. (1985): The distribution of general quadratic forms in normal variables, *Statistica Neerlandica*, 39, 14-26.
- Fienberg, S.E. (1979): The use of Chi-squared statistics for categorical data problems. *Journal of the Royal Statistical Society*, B, 41, 54-64.
- Fienberg, S.E. (1980): *The Analysis of Cross-Classified Categorical Data* (2nd edition). Cambridge, MA, the MIT Press.
- Fienberg, S.E. (1984): The contributions of William Cochran to categorical data analysis. In *W.G. Cochran's Impact on Statistics* (editors P.S.R.S. Rao y J. Sedransk), 103-118. New York, John Wiley.
- Fisher, R.A. (1924): the conditions under which χ^2 measures the discrepancy between observation and hypothesis. *Journal of the Royal Statistical Society*, 87, 442-450.
- Fraser, D.A.S. (1957): *Nonparametric Methods in Statistics*, Wiley, New York.
- Freeman, D.H. (1987): *Applied Categorical Data Analysis*. New York, Marcel Dekker.
- Freeman, M.F. y Tukey, J.W. (1950): Transformations related to the angular and the square root. *Annals of Mathematical Statistics*,

- 21, 607-611.
- Frosini, B.V. (1976): On the power function of the χ^2 test. *Metron*, 34, 3-6.
- Fryer, J.G. y Robertson, C.A. (1972): A comparison of some methods for estimating mixed normal distributions. *Biometrika*, 59(3), 639-648.
- Gokhale, D.V. y Kullback, S. (1978): *The Information in Contingency Tables*. New York, Marcel Dekker.
- Goldstein, M., Wolf, E. y Dillon, W. (1976): On a test of independence for contingency tables. *Communications in Statistics. Theory and Methods*, 2, 159-169.
- Goodman, L.A. (1984): *Analysis of Cross-Classified Data Having Ordered Categories*. Cambridge, MA, Harvard University Press.
- Haberman, S.J. (1974): *The Analysis of Frequency Data*. Chicago, University of Chicago Press.
- Haberman, S.J. (1978): *Analysis of Qualitative Data*, Vol.1. New York, Academic Press.
- Haberman, S.J. (1979): *Analysis of Qualitative Data*, Vol.2. New York, Academic Press.
- Hassenblad, V. (1966): Estimation of parameters for a mixture of normal distributions. *Technometrics*, 8, 431-434.
- Hoeffding, W. (1965): Asymptotically optimal tests for multinomial distributions. *Annals of Mathematical Statistics*, 36, 369-408.
- Hoel, P.G. (1938): On the chi-square distribution for small samples. *Annals of Mathematical Statistics*, 9, 158-165.
- Holst, L. (1972): Asymptotic normality and efficiency for certain goodness-of-fit tests. *Biometrika*, 59, 137-145.
- Horn, S.D. (1977): Goodness-of-fit tests for discrete data: A review and an application to a health impairment scale. *Biometrics*, 33, 237-248.
- Huber, P.J. (1981): *Robust Statistics*. J. Wiley, New York.
- Ivchenko, G.I. y Medvedev, Y.I. (1978): Separable statistics and hypothesis testing. The case of small samples. *Theory of Probability and Its Applications*, 23, 764-775.

- Koehler, K.J. y Larntz, K. (1980): An empirical investigation of goodness-of-fit statistics for sparse multinomials. *J. Amer. Statist. Ass.*, 75, 336-344.
- Kotze, T.J.v.W. y Gokhale, D.V. (1980): A comparison of the Pearson- X^2 and the log-likelihood-ratio statistics for small samples by means of probability ordering. *Journal of Statistical Computation and Simulation*, 12, 1-13.
- Kullback, S. y Leibler, R. (1951): On information and sufficiency. *Annals of Mathematical Statistics*, 22, 79-86.
- Kullback, S. (1959): *Information Theory and Statistics*. J. Wiley, New York.
- Kullback, S. (1985): Minimum discrimination information (MDI) -estimation. In *Encyclopedia of Statistical Sciences*, (editores S. Kotz y N.L. Johnson), New York, John Wiley, 5, 527-529.
- Lancaster, H.O. (1969): *The Chi-squared Distribution*. New York. Wiley.
- Larntz, K. (1978): Small sample comparisons of exact levels for Chi-squared goodness of fit statistics. *Journal of the American Statistical Association*, 73, 253-263.
- Lau, K. (1985): Characterization of Rao's quadratic entropies. *Sankhya Series A* 47, 295-309.
- Lawley, D.N. (1956): A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika*, 43, 295-303.
- Le Cam, L. (1990): Maximum Likelihood: An Introduction. *International Statistical Review*, 58, 2, 153-171.
- Lewis, T., Saunders, I.W. y Westcott, M. (1984): The moments of the Pearson chi-squared statistic and the minimum expected value in two-way tables. *Biometrika*, 71, 515-522.
- Liese F. y Vajda, I. (1987): *Convex Statistical Distances*. Teubner, Leipzig.
- Marshall, A.W. y Olkin, I. (1979): *Inequalities: Theory of Majorization and its Applications*. Academic Press, New York.
- Matusita, K. (1964): Distances and decision rules, *Ann. Inst. Statist. Math.*, 16, 305-320.

- Menéndez, M.L., Morales, D., Pardo, L. y Vajda, I. (1996): About divergence-based goodness-of-fit tests in the Dirichlet-Multinomial model. Aparecerá en *Communications in Statistics (Theory and Methods)*.
- Moore, D.S. y Spruill, M.C. (1975): Unified large-sample theory of general chi-squared statistics for tests of fit. *Annals of Statistics*, 3, 599-616
- Moore, D.S. (1986): Tests of chi-squared type. In *Goodness-of-Fit Techniques* (editores R.B. D'Agostino y M.A. Stephens), 63-95. New York, Marcel Dekker.
- Morales, D. Pardo, L y Vajda, I. (1995): Asymptotic divergence of estimates of discrete distributions. *Journal of Statistical Planning and Inference* 48, 347-369.
- Morris, C. (1966): Admissible Bayes procedures and classes of epsilon Bayes procedures for testing hypotheses in a multinomial distribution. *Technical Report 55*, Department of Statistics, Stanford University, Stanford, CA.
- Morris, C. (1975): Central limit theorems for multinomial sums. *Annals of Statistics*, 3, 165-188.
- Nayak, T.K. (1986): Sampling distributions in analysis of diversity. *Sankhya Series B*, 48, 1-9.
- Neyman, J. (1949): Contribution to the theory of the χ^2 -test. *Proc. 1st Berkeley Symp. on Math. Statist. and Probab.*, 239-273. Univ. of Berkeley Press, Berkeley.
- Osterreicher, F. (1996): On a class of perimeter-type distances of probability distributions. Aparecerá en *Kybernetika*
- Pardo, L., Morales, D., Salicrú, M. y Menéndez, M.L. (1993): R_{ϕ}^h -divergence statistics in applied categorical data analysis with stratified sampling. *Utilitas Mathematica*, 44, 145-164.
- Pearson, K. (1894): Contributions to the mathematical theory of evolution. *Phil. Trans. R. Soc. Ser.A*, 185, 71-110.
- 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 be reasonably supposed to have

- arisen from random sampling. *Philosophy Magazine*, 50, 157-172.
- Plackett, R.L. (1981): *The Analysis of Categorical Data* (2nd edition). High Wycombe, Griffin.
- Radlow, R. y Alf, E.F. (1975): An alternate multinomial assessment of the accuracy of the χ^2 test of goodness of fit. *Journal of the American Statistical Association* 70, 811-813.
- Rao, C.R. (1961): Asymptotic efficiency and limiting information. *Proc. 4th Berkeley Symp. on Math. Statist. Probab.*, 1, 531-546. Univ. of California Press, Berkeley.
- Rao, C.R. (1973): *Linear Statistical Inference and Its Applications* (2nd edition). New York, John Wiley.
- Rao, C.R. (1982a): Diversity and dissimilarity coefficients: a unified approach. *Theoretical Population Biology*, 21, 24-43.
- Rao, C.R. (1982b): Diversity: its measurement, descomposition, apportionment and analysis. *Sankhya Series A*, 44, 1-22.
- Rao, C.R. y Nayak, T.K. (1985): Cross entropy, dissimilarity measures, and characterizations of quadratic entropy. *IEEE Transactions on Information Theory*, 31, 589-593.
- Read, T.R.C. y Cressie, N.A.C. (1988): *Goodness of fit Statistics for Discrete Multivariate Data*. Springer-Verlag.
- Read, T.R.C. (1982): Choosing a goodness-of-fit test. *Ph. D. Dissertation, School of Mathematical Sciences*. The Flinders University of South Australia, Adelaide, South Australia.
- Read, T.R.C. y Cowan, R. (1976): Probabilistic modelling and hypothesis testing applied to permutation data. Private correspondence.
- Rényi, A. (1961): On measures of entropy and information. *Proc. 4th Berkeley Symp. on Math. Statist. Probab.*, 1, 547-561. Univ. of California Press, Berkeley.
- Robertson, C.A. (1972): On minimum discrepancy estimators. *Sankhya*, Ser. A, 133-144.
- Siotani, M. y Fujikoshi, Y. (1984): Asymptotic approximations for the distributions of multinomial goodness-of-fit statistics. *Hiroshima Mathematics Journal* 14, 115-124.

- Smith, P.J., Rae, D.S. Manderscheid, R.W. y Silbergeld, S. (1981): Approximating the moments and distribution of the likelihood ratio statistic for multinomial goodness of fit. *Journal of the American Statistical Association*, 76, 737-740.
- Tate, M.W. y Hyer, L.A. (1973): Inaccuracy of the X^2 test of goodness of fit when expected frequencies are small. *Journal of the American Statistical Association* 68, 836-841.
- Tavaré, S. y Altham, P.M.E. (1983): Serial dependence of observations leading to contingency tables, and corrections to chi-squared statistics. *Biometrika*, 70, 139-144.
- Tchentsov, N.N. (1972): Statistical Decision Rules and Optimal Inference (in Russian). *Nauka, Moskva* (see also Translation of Math. Monographs, 53, *Amer. Math. Soc.*).
- Upton, G.J.G. (1978): *The Analysis of Cross-Tabulated Data*. New York, John Wiley.
- Vajda, I. (1989): *Theory of Statistical Inference and Information*. Kluwer, Boston.
- Vajda, I. y Kus, V. (1995): Relation between Divergences, Total Variation and Euclidean distances, *Technical Report*, 1853. Prague: Inst. of Inform. Theory.
- Wakimoto, K., Odaka, Y. y Kang, L. (1987): Testing the goodness of fit of the multinomial distribution based on graphical representation. *Computational Statistics and Data Analysis*, 5, 137-147.
- Watson, G.S. (1959): Some recent results in chi-square goodness-of-fit tests. *Biometrics*, 15, 440-468.
- West, E.N. y Kempthorne, O. (1972): A comparison of the Chi^2 and likelihood ratio tests for composite alternatives. *J. Statist. Computation and Simulation*, 1, 1-33.
- Woodward, W.A., Parr, W.C., Schucany, W.R. y Lindsay, H. (1984): A comparison of minimum distance and maximum likelihood estimation of a mixture proportion. *Journal of American Statistical Association*, 79, 590-598.
- Woodward, W.A., Whitney, P. y Eslinger, P.W. (1995): Minimum Hellinger

distance estimation of mixture proportions. *Journal of Statistical Planning and Inference*, 48, 303-319.

Yarnold, J.K. (1972): Asymptotic approximations for the probability that a sum of lattice random vectors lies in a convex set. *Annals of Mathematical Statistics*, 43, 1566-1580.