UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS MATEMÁTICAS

Departamento de Estadística e Investigación Operativa

LAS MEDIDAS DE DIVERGENCIA EN CONTRASTES DE BONDAD DE AJUSTE CON PONDERACIÓN EN LAS CLASES

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Elena Landaburu Jiménez

Bajo la dirección del doctor

Leandro Pardo Llorente

Madrid, 2002

ISBN: 84-669-1799-3

Indice

Introducción

En las últimas tres décadas el análisis estadístico de datos categóricos ha despertado gran interés entre numerosos estadísticos. El desarrollo de modelos apropiados ha sido tema común de importantes e interesantes libros: Cox (1970), Haberman (1974, 1978, 1979), Bishop y otros (1975), Gokhale y Kullback (1978), Upton (1978), Fienberg (1979, 1980), Plackett (1981), Agresti (1984, 1990), Goodman (1984), Freeman (1987), Leonard (2000), etc...

Generalmente, estos autores han prestado una especial atención a comprobar a través de los estadísticos de la ji-cuadrado y del cociente de verosimilitudes, que los datos se ajustaban correctamente al modelo seleccionado. Frente a este punto de vista de construir modelos y validarlos a través del estadístico de la ji-cuadrado o del cociente de verosimilitudes, otros autores se han preocupado de construir nuevos estadísticos para la validación de modelos que mejoren, en algún sentido, los ya existentes. Cressie y Read (1984) y Read y Cressie (1988) fueron los pioneros a la hora de abordar este problema, utilizando la familia de divergencias que lleva su nombre, para construir una familia de estadísticos de contraste en el caso de hipótesis nula simple como compuesta y establecer, en el caso de hipótesis nula simple, que un elemento de la familia de estadísticos, el correspondiente a $\lambda = 2/3$, tiene un comportamiento mejor que el estadístico del cociente de verosimilitudes y que el de la ji-cuadrado. Posteriormente Zografos y otros (1990) extendieron los resultados de Cressie y Read al considerar la familia de φ-divergencias, en lugar de la familia de divergencias de Cressie y Read. La familia de medidas de φ-divergencias contiene como caso particular a la familia de Cressie y Read, como se verá en el Capítulo I. Posteriormente en Menéndez y otros (1995) se estableció, que en el sentido de Cressie y Read, las mejores φ-divergencias se obtienen como solución de la ecuación

$$4\varphi'''(1) + 3\varphi''''(1) = 0. \tag{1}$$

El valor $\lambda = 2/3$ obtenido por Cressie y Read es una solución de la ecuación anterior cuando se utiliza la familia de divergencias introducida por ellos. Posteriormente en Morales y otros (1995), se utilizaron por primera vez las medidas de φ-divergencias para el problema de bondad de ajuste cuando la hipótesis nula es compuesta. En esta situación al tener que estimar parámetros se propuso el método de estimación de mínima ϕ-divergencia que generaliza, como más tarde se verá en el Capítulo II, el método de máxima verosimilitud. Muchos e importantes han sido los trabajos, aparte de los ya citados, que se han publicado en esta línea. Entre otros cabe destacar los siguientes: Lindsay (1994), Simpson (1987), Basu y Harris (1994), Basu, Harris y Basu (1996), Basu y Lindsay (1994), Basu y Sarkar (1994a,b), Beran (1977), Parr (1981), Pardo, M. C. (1996, 1999), Morales y otros (1997), Menéndez y otros (1995), etc...

Todos estos trabajos tienen en común la utilización de medidas de divergencia para abordar problemas estadísticos de estimación y contraste, encuadrándose pues en lo que hoy día se conoce como Teoría de la Información Estadística: Utilización de las herramientas básicas que proporciona la Teoría de la Información; es decir, entropías y divergencias, para abordar los problemas clásicos de estimación y contraste, tanto en poblaciones multinomiales como generales.

Los trabajos señalados anteriormente consideran las medidas de divergencia como medidas cuantitativas de discriminación entre dos poblaciones, caracterizadas por sus respectivas distribuciones de probabilidad, pero sin tener en cuenta la importancia de los resultados asociados al experimento bajo consideración, respecto a un fin determinado. A la hora de pensar en la adaptación de las medidas de divergencia al contexto en el cual los resultados del experimento bajo consideración no tienen la misma importancia con respecto a un fin determinado, uno puede tomar como referencia la línea establecida inicialmente por Belis y Guiasu (1968) y posteriormente por Gil (1975) en relación a la entropía de Shannon y consistente en ponderar la autoinformación asociada a cada resultado. Es decir, asignar a los resultados del experimento bajo consideración, una ponderación, que represente la importancia del mismo con respecto al fin determinado. Así, parece lógico adaptar las medidas de divergencia mediante la asignación también de una ponderación a los elementos del espacio muestral sobre el que están definidas las distribuciones de probabilidad que intervienen en la medida de divergencia. Para abordar la afirmación hecha repetidamente en los párrafos previos "importancia con respecto a un fin determinado" se presentarán varios ejemplos en el Capítulo I.

El objetivo central de esta memoria es abordar problemas de bondad de ajuste, bajo el supuesto de que los datos están bien o mal clasificados y tanto bajo hipótesis nula simple como compuesta, cuando las clases utilizadas tengan asignada una diferente ponderación. En el caso de hipótesis nula compuesta se requerirá la estimación del parámetro desconocido mediante estimadores de mínima divergencia, que tendrán en cuenta las diferentes ponderaciones. En definitiva, esta memoria se encuadra dentro de la denominada Teoría de la Información Estadística a la que ya se ha hecho referencia anteriormente.

En el Capítulo I tras hacer un análisis de las medidas de divergencia más importantes hasta ahora en la literatura estadística se pone de manifiesto, que tras un estadístico de contraste para bondad de ajuste, siempre subyace una medida de divergencia. Así, en el caso de los dos estadísticos de contraste más conocidos: El estadístico de la ji-cuadrado y el del cociente de verosimilitudes, subyacen la divergencia de Pearson y la de Kullback, que a su vez son un caso particular de la familia de divergencias de Cressie y Read y ésta de la familia de φ-divergencias de Csiszár. Tras este análisis se pasa a definir la familia de divergencias ponderadas que será el objeto de estudio a lo largo de esta memoria: Las (h, φ)-divergencias ponderadas. Una de las propiedades a las que se dedica una especial atención es a la no negatividad y a que toma el valor cero cuando las dos distribuciones de probabilidad coinciden. Esta propiedad será esencial a la hora de dar sentido a los contrastes de bondad de ajuste planteados a lo largo de la memoria.

El Capítulo II está dedicado al estudio de los estimadores de mínima (h, φ)-divergencia ponderada. Tras describir la importancia, en general, de los estimadores de mínima divergencia asociados a las poblaciones multinomiales se pasa a analizar como en realidad los estimadores de mínima divergencia son una extensión natural del estimador de máxima verosimilitud, y se basan en cuantificar la distancia entre el vector de frecuencias observadas, estimador no paramétrico del modelo, y el vector de probabilidades que describe el modelo con una medida de divergencia distinta de la de Kullback. El estimador de máxima verosimilitud no es más que el valor del parámetro que hace que la distancia, cuantificada en términos de la divergencia de Kullback, entre el estimador no paramétrico del modelo y el vector de probabilidades que rige el modelo sea mínima. Esta justificación intuitiva de los estimadores de mínima divergencia permite definir de forma natural el concepto de mínima (h, φ)-divergencia ponderada. Seguidamente, y bajo la suposición de que el modelo bajo consideración verifica las condiciones de

Birch, se pasa al estudio de sus propiedades asintóticas. Se demuestra su normalidad asintótica y consistencia. 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. Se finalizará el capítulo con el estudio del estimador de mínima (h, φ)-divergencia con ponderaciones cuando existen un número determinado de condiciones que restringen el parámetro de nuestro modelo. En esta nueva situación se vuelve a establecer la normalidad asintótica del mismo como otras propiedades de interés.

A partir de la (h, φ)-divergencia ponderada entre el estimador no paramétrico y la distribución de probabilidad teórica del mismo, se propone, en el Capítulo III, un contraste de bondad de ajuste. Se abordará el problema tanto en el caso de hipótesis nula simple como compuesta. En el caso de hipótesis nula compuesta será necesario estimar el parámetro desconocido. Ello se llevará a cabo a través del estimador de mínima (h, φ)-divergencia ponderada introducido y estudiado en el capítulo anterior. La distribución asintótica en ambos casos conduce a una combinación lineal de distribuciones ji-cuadrados. En principio, el que la distribución asintótica del estadístico de contraste sea una combinación lineal de distribuciones ji-cuadrados produce una cierta inquietud. Esto desaparecerá en gran medida si se hace uso de las excelentes aproximaciones que existen en la literatura estadística para las combinaciones de ji-cuadrados. Además, se calcula la potencia del contraste y a partir de ésta se describe un procedimiento para determinar el tamaño muestral necesario que garantice una potencia prefijada. Se presentan otros contrastes de interés y se finaliza este capítulo particularizando los resultados obtenidos al caso binomial. En esta situación la distribución asintótica resultante es una ji-cuadrado, en lugar de una combinación lineal de ji-cuadrados.

El Capítulo IV está dedicado al análisis de la optimalidad en los contrastes de bondad de ajuste introducidos en el capítulo anterior, en el supuesto de que la hipótesis nula sea simple. La familia de estadísticos de contraste utilizada para abordar el problema de bondad de ajuste tenía una distribución asintótica, bajo la hipótesis nula, que no dependía de las distintas funciones que intervienen en la definición del estadístico. Desde un punto de vista práctico se trabajará con muestras finitas y en muchas ocasiones pequeñas, y en estos casos deberá haber diferencias a la hora de la elección de las correspondientes funciones de las familias φ y h que determinan el estadístico de contraste. En este capítulo se analizan dos procedimientos: Uno basado en la función de potencia, introducida y estudiada en el capítulo anterior, del test de bondad de ajuste y el otro consistente en seleccionar las funciones de las familias φ y h de forma que permitan una mayor proximidad entre los momentos exactos y asintóticos del estadístico de contraste bajo la hipótesis nula. Este segundo procedimiento conduce a la resolución de una ecuación diferencial, que en el caso de que las familias φ y h estén compuestas por un único elemento y las ponderaciones sean las mismas coincide con la ecuación planteada en (1) .

Finalmente el Capítulo V está dedicado a abordar el problema de bondad de ajuste, en el caso de hipótesis nula simple, ante la posibilidad de que existan varias observaciones mal clasificadas en las correspondientes clases. Bross (1954) estableció que el estimador no paramétrico en esta situación es sesgado y además, el sesgo es función del número de observaciones mal clasificadas. El efecto de datos mal clasificados sobre el estadístico de la ji-cuadrado, en el problema usual de bondad de ajuste, es el aumento del tamaño del mismo y la disminución de su potencia. Este resultado fue establecido en Mote y Anderson (1965). La forma en la que abordaron el problema Mote y Anderson (1965) es claramente extensible al estadístico de contraste considerado en el capítulo anterior, cuando las clases están ponderadas, y como consecuencia, el estadístico de contraste, para el problema de bondad de ajuste, estudiado en el capítulo anterior tiene el mismo problema. La solución que se da en este capítulo a este problema se basa en sustituir el estimador no paramétrico por otro obtenido a través del muestreo doble introducido y estudiado por Tenenbein (1970, 1971, 1972). Tras describir el método de estimación por muestreo doble, se estudia la distribución asintótica del correspondiente estimador de máxima verosimilitud y se plantea el contraste de bondad de ajuste con datos mal clasificados basado en las (h, φ)-divergencias ponderadas. La distribución asintótica del estadístico de contraste, como era de esperar, resulta ser una combinación lineal de ji-cuadrados. Se finaliza el capítulo haciendo especial hincapié en el caso binomial.

Por último, quiero dedicar unas palabras de agradecimiento, en especial al profesor Leandro Pardo Llorente, por la ayuda, comprensión y apoyo prestado en todo momento, a mi familia por haberme “sufrido” durante este periodo y al Departamento de Estadística e Investigación Operativa de la Facultad de Ciencias Matemáticas por su colaboración.

1

Medidas de (h, ϕ) –Divergencia ponderadas

1.1. Introducción

Sea $X_1, ..., X_n$ una muestra aleatoria simple de una población X con valores en el espacio estadístico $(\mathcal{X}, \beta_{\mathcal{X}}, P)$. El problema de bondad de ajuste consiste en contrastar si P pertenece a la familia paramétrica

$$\mathcal{P} = \left\{ P_\theta : \theta \in \Theta \subset R^{M_0} \right\}$$

de distribuciones sobre $(\mathcal{X}, \beta_{\mathcal{X}})$. Una clase importante de contrastes de hipótesis, para abordar este problema, está basada en distancias, generalmente medidas de divergencia, entre la distribución discreta bajo la hipótesis nula

$$p(\theta) = (p_1(\theta), ..., p_M(\theta))^t,$$

obtenida al considerar una partición fija $\mathcal{A} \equiv \{A_1, ..., A_M\} \subset \beta_{\mathcal{X}}$ de $\mathcal{X}$, con

$$p_j(\theta) = P_\theta(A_j) \qquad 1 \leq j \leq M$$

y el estimador no paramétrico

$$\hat{p} = (\hat{p}_1, ..., \hat{p}_M)^t \tag{1.1}$$

con

$$\hat{p}_j = \frac{1}{n} \sum_{i=1}^{n} I_{A_j}(X_i) \qquad 1 \leq j \leq M.$$

El estadístico $(N_1, .., N_M)$, con $N_j = \sum_{i=1}^{n} I_{A_j}(X_i)$, es suficiente para el modelo estadístico bajo consideración y multinomialmente distribuido con parámetros $(n; p_1(\theta), ..., p_M(\theta))$.

El contraste de bondad de ajuste, en el caso de hipótesis nula simple, se puede entonces formular mediante

$$H_0 : p = p(\theta_0), \tag{1.2}$$

donde θ_0 es un valor fijo y conocido. Los dos estadísticos más comúnmente usados son el estadístico $\mathcal{X}^2$ de Pearson,

$$\mathcal{X}^2 \equiv \sum_{j=1}^{M} \frac{(N_j - np_j(\theta_0))^2}{np_j(\theta_0)} \tag{1.3}$$

y el estadístico del cociente de verosimilitudes, G^2

$$G^2 = 2 \sum_{j=1}^{M} N_j \ln\left(\frac{N_j}{np_j(\theta_0)}\right). \tag{1.4}$$

Es importante observar que (1.3) se puede escribir de la forma

$$\mathcal{X}^2 = n \sum_{j=1}^{M} \frac{(\hat{p}_j - p_j(\theta_0))^2}{p_j(\theta_0)} = 2n\mathcal{X}^2(\hat{p}, p(\theta_0)) \tag{1.5}$$

donde $\mathcal{X}^2(\hat{p}, p(\theta_0))$ es la distancia o medida de divergencia de Pearson. De la misma forma, (1.4) se pude escribir mediante

$$G^2 = 2n \sum_{j=1}^{M} \hat{p}_j \ln\left(\frac{\hat{p}_j}{p_j(\theta_0)}\right) = 2nD_{kullback}(\hat{p}, p(\theta_0)) \tag{1.6}$$

donde $D_{kullback}(\hat{p}, p(\theta_0))$ es la divergencia de Kullback-Leibler (Kullback-Leibler 1951) entre las distribuciones de probabilidad $\hat{p}$ y $p(\theta_0)$.

Si en lugar de considerar hipótesis nula simple se considera la hipótesis nula compuesta,

$$H_0 : p = p(\theta) \tag{1.7}$$

los estadísticos $\mathcal{X}^2$ y G^2 siguen siendo válidos sin más que estimar ahora el parámetro desconocido de forma conveniente. Es bien conocido que la distribución asintótica de $\mathcal{X}^2$ y G^2 es la misma: Ji-cuadrado con $M - 1$ grados de libertad, en el caso de hipótesis nula simple, y Ji-cuadrado con $M - M_0 - 1$ grados de libertad en el caso de hipótesis nula compuesta si el parámetro se estima de forma conveniente. Siendo esto importante, en este capítulo, únicamente nos fijaremos en las relaciones dadas (1.5) y (1.6): Las medidas de divergencias asociadas a los estadísticos $\mathcal{X}^2$ y G^2.

La presente memoria se ocupará de diversas cuestiones relativas al problema de bondad de ajuste cuando los elementos de la partición $\mathcal{A} = \{A_1, ..., A_M\} \subset \beta_{\mathcal{X}}$ de $\mathcal{X}$ tienen una ponderación asociada. Estos problemas se abordarán a través de las denominadas (h, φ)-divergencias ponderadas que se introducirán y estudiarán en esta memoria. Con el fin de centrar convenientemente esta clase de medidas de divergencias, en el siguiente apartado se hará un breve repaso de las diversas familias de divergencias y en el apartado siguiente se introducirá y analizará la familia de (h, φ)-divergencias ponderadas.

1.2. Divergencias entre dos poblaciones

En lo que sigue se denotará por

$$\Delta_{M,\theta} = \left\{p(\theta) = (p_1(\theta), ..., p_M(\theta))^t, \theta \in \Theta \subset R^{M_0}\right\}$$

con $p_j(\theta) = P_\theta(A_j)$, $j = 1, ..., M$ y por

$$\Delta_M = \left\{p = (p_1, ..., p_M)^t, 0 \leq p_i \leq 1, \sum_{i=1}^{M} p_i = 1\right\}$$

el conjunto de todas las distribuciones de probabilidad sobre $\mathcal{A} \equiv \{A_1, ..., A_M\}$. Es claro que $\Delta_{M,\theta} \subset \Delta_M$.

Intuitivamente una medida de divergencia representa, desde un punto de vista estadístico, el "distanciamiento" entre dos poblaciones caracterizadas por sus respectivas distribuciones de probabilidad.

Dados dos vectores de probabilidad $p = (p_1, ..., p_M)^t$ y $q = (q_1, ..., q_M)^t$ pertenecientes a Δ_M, la divergencia de Kullback-Leibler (1951), entre p y q viene dada por

$$D(p, q) = \sum_{j=1}^{M} p_j \log \frac{p_j}{q_j}$$

y admite la siguiente interesante expresión

$$D(p,q) = I(p/q) - H(p)$$

donde

$$I(p/q) = \sum_{j=1}^{M} p_j \log q_j$$

es la Inacuracidad de Kerridge (1961) y

$$H(p) = -\sum_{j=1}^{M} p_j \log p_j \tag{1.8}$$

es la medida de Entropía de Shannon (1948). Un amplio estudio de esta medida de entropía, así como de otras, puede verse en Pardo, L. (1997b).

El primer intento para desarrollar una generalización de la divergencia de Kullback-Leibler fue llevado a cabo por Rényi (1961), que definió la medida que lleva su nombre en los siguientes términos:

$$D_r^1(p,q) = (r(r-1))^{-1} \log \left\{ \sum_{j=1}^{M} p_j^r q_j^{1-r} \right\} \qquad r \neq 1, r \neq 0.$$

Otras generalizaciones fueron dadas por Sharma y Mittal (1975) y se conocen como divergencia de orden 1 y grado s y divergencia de orden r y grado s, respectivamente,

$$D_1^s(p,q) = (s-1)^{-1} \left\{ \exp\left[(s-1) \sum_{j=1}^{M} p_j \log \frac{p_j}{q_j} \right] - 1 \right\}, \qquad s \neq 1$$

$$D_r^s(p,q) = (s-1)^{-1} \left\{ \left[\sum_{j=1}^{M} p_j^r q_j^{1-r} \right]^{\frac{s-1}{r-1}} - 1 \right\}, \quad r \neq 1, s \neq 1.$$

Puede verse que se verifica:

$$\lim_{s \to 1} D_r^s (p, q) = r D_r^1 (p, q)$$

$$\lim_{r \to 1} D_r^s (p, q) = D_1^s (p, q)$$

$$\lim_{\substack{r \to 1 \\ s \to 1}} D_r^s (p, q) = \lim_{s \to 1} D_s^s (p, q) = \lim_{s \to 1} D_1^s (p, q) = D (p, q) .$$

Para más detalles acerca de estas y otras medidas de divergencia consultar Liese y Vajda (1987) y Pardo, L. (1997a,b).

Por otra parte, la divergencia de Kullback-Leibler es un caso particular de las φ-divergencias introducidas y estudiadas, independientemente, por Csiszár (1963) y Ali y Silvey (1966), que se definen mediante la expresión

$$D_\varphi(p, q) = \sum_{j=1}^{M} q_j \varphi \left(\frac{p_j}{q_j} \right) \ , \ \varphi \in \Phi^*$$

donde Φ^* es la clase de todas las funciones convexas φ: $[0, \infty) \to R \cup \{\infty\}$ tales que en $x = 1$, $\varphi(1) = 0$, $\varphi''(1) > 0$ y en $x = 0$, $0\varphi(0/0) = 0$ y $0\varphi(p/0) = p \lim_{u \to \infty} \frac{\varphi(u)}{u}$. Para cada $\varphi \in \Phi^*$ diferenciable, la función

$$\psi(x) = \varphi(x) - \varphi'(1)(x - 1)$$

también pertenece a Φ^*. Es claro que

$$D_\psi (p, q) = D_\varphi (p, q) ,$$

y ψ tiene la propiedad adicional de que $\psi'(1) = 0$.

Al ser las dos medidas de divergencia equivalentes podemos considerar que el conjunto Φ^* es equivalente al conjunto

$$\Phi \equiv \Phi^* \cap \left\{ \varphi : \varphi'(1) = 0 \right\} .$$

Así pues, al hacer referencia a las φ-divergencias se utilizará indistintamente $\varphi \in \Phi$ ó $\varphi \in \Phi^*$.

Seguidamente se presentan algunas medidas importantes de divergencia que se pueden obtener como caso particular de la familia de φ-divergencias de Csiszár.

- Kullback-Leibler

  $\varphi(x) = x \log x$

- Variacional o Estadística

  $\varphi(x) = |x-1|$

- X^2-divergencia o distancia de Pearson

  $\varphi(x) = \frac{1}{2}(1-x)^2$

- Matusita

  $\varphi(x) = |1-x^a|^{\frac{1}{a}}$ con $0 < a \leq 1$

- Balakrishnan y Sanghvi

  $\varphi(x) = \frac{(x-1)^2}{(x+1)}$

- Rathie y Kanappan

  $\varphi_s(x) = (x^s - x)\left(2^{s-1}-1\right)^{-1}, s \neq 1$

- Media armónica

  $\varphi(x) = 1 - 2^{\frac{1}{r}}(1+x^{-r})^{-\frac{1}{r}}$

- Cressie y Read

  $\varphi_\lambda(x) = [\lambda(\lambda+1)]^{-1}\left(x^{\lambda+1}-x\right), \lambda \neq -1, \lambda \neq 0$

- Rukhin

  $\varphi_a(x) = \left\{[a+(1-a)x]^{-1} - 1\right\}, 0 \leq a < 1$

- Lin

  $\varphi_b(x) = bx\log x - (bx+1-b)\log(bx+1-b).$

Desde un punto de vista estadístico la familia más importante es la de Cressie y Read introducida y estudiada por estos autores, ver Cressie y Read (1984) y Read y Cressie (1988).

En las medidas de divergencia listadas anteriormente se ha supuesto que $\varphi \in \Phi^*$. Obsérvese por ejemplo que la familia de medidas de Cressie y Read obtenida mediante $\varphi_\lambda(x) \in \Phi^*$ da lugar a la misma familia de medidas de divergencia que si se considera

$$\psi_\lambda(x) = \varphi_\lambda(x) - (x-1)(\lambda+1)^{-1}, \qquad \lambda \neq -1, \lambda \neq 0$$

con $\psi_\lambda \in \Phi$.

Las indeterminaciones, para algunos valores de los parámetros, de las medidas de divergencia introducidas anteriormente se suelen resolver tomando límites. Así, en el caso de la medida de Rathie y Kanappan

$$\varphi_1(x) = \lim_{s \to 1} \varphi_s(x),$$

y en el caso de la de Cressie y Read

$$\varphi_0(x) = \lim_{\lambda \to 0} \varphi_\lambda(x) \text{ y } \varphi_{-1}(x) = \lim_{\lambda \to -1} \varphi_\lambda(x).$$

Es interesante observar que los estadísticos $\mathcal{X}^2$ y G^2 son un caso particular de la familia de estadísticos asociados a la familia de divergencias de Cressie-Read,

$$2nD_\lambda(\hat{p}, p(\theta_0)) = \frac{2n}{\lambda(\lambda+1)} \sum_{j=1}^{M} \hat{p}_j \left(\left(\frac{\hat{p}_j}{p_j(\theta_0)} \right)^\lambda - 1 \right), -\infty < \lambda < \infty. \quad (1.9)$$

Los estadísticos $2nD_0(\hat{p}, p(\theta_0))$ y $2nD_{-1}(\hat{p}, p(\theta_0))$ se definen como casos límites de $2nD_\lambda\left(\hat{P}, P(\theta_0)\right)$ cuando $\lambda \to 0$ y $\lambda \to -1$, respectivamente. Si en (1.9) se hace $\lambda = 1$, se obtiene el estadístico $\mathcal{X}^2$ y para $\lambda \to 0$, el G^2. Conviene también señalar que para $\lambda = -1/2$ se obtiene el estadístico de Freeman-Tukey y para $\lambda = -2$ el de Neyman modificado. Cressie y Read establecieron que todos los elementos de la familia (1.9) tienen la misma distribución asintótica y ésta es una Ji-cuadrado con $M-1$ grados de libertad. Así mismo, estos autores establecieron que en caso de hipótesis nula compuesta, la familia de estadísticos $2nD_\lambda\left(\hat{p}, p\left(\hat{\theta}\right)\right)$, con $\hat{\theta}$ el estimador de máxima verosimilitud de θ, en el modelo discretizado asociado, se distribuye asintóticamente como una Ji-cuadrado con $M - M_0 - 1$ grados de libertad.

De la misma forma que a partir de la familia de divergencias de Cressie-Read se ha definido una familia de estadísticos para el problema de bondad de ajuste, lo mismo se puede hacer a partir de la familia de Csiszár. Así, Zografos y otros (1990) establecieron que bajo la hipótesis nula simple la familia de estadísticos

$$T_n^\varphi = \frac{2n}{\varphi''(1)} D_\varphi(\hat{p}, p(\theta_0))$$

se distribuye asintóticamente, independientemente de φ, como una Ji-cuadrado con $M-1$ grados de libertad. En el caso de hipótesis nula compuesta Morales y otros (1997) establecieron que la familia de estadísticos

$$T_n^{\varphi,\phi} = \frac{2n}{\varphi''(1)} D_\varphi\left(\hat{p}, p\left(\hat{\theta}_\phi\right)\right),$$

siendo $\hat{\theta}_\phi$ el estimador de mínima divergencia definido mediante

$$\hat{\theta}_\phi = \arg\min_{\theta\in\Theta} D_\phi\left(\hat{p}, p(\theta)\right), \tag{1.10}$$

se distribuye asintóticamente como una Ji-cuadrado con $M - M_0 - 1$ grados de libertad, independientemente de las funciones φ y ϕ consideradas. Ahora no nos detendremos a analizar el estimador de mínima divergencia ya que volveremos a él de forma extensiva en el Capítulo 2. Unicamente señalar que generaliza el concepto de estimador de máxima verosimilitud en el sentido de que para $\phi(x) = x\log x - x + 1$ se obtiene el estimador de máxima verosimilitud.

Puede observarse que tanto la medida de divergencia de Rényi como la de Sharma y Mittal no se pueden obtener como una φ-divergencia.

Por este motivo, Menéndez y otros (1995), definieron un funcional que las incluía, así como a otras muchas. Dicho funcional, llamado (h,φ) −divergencia, tiene la siguiente expresión

$$D_\varphi^h(p,q) = \sum_{a=1}^{\Lambda} \eta_a h_a\left(\sum_{j=1}^{M} q_j \varphi_a\left(\frac{p_j}{q_j}\right)\right) \tag{1.11}$$

donde $h = (h_a)_{a\in\Lambda}$, $\varphi=(\varphi_a)_{a\in\Lambda}$, φ_a satisface las condiciones de la divergencia de Csiszár, h_a es no decreciente y continua en $\left(0, \lim_{u\to\infty} \frac{\varphi_a(u)}{u}\right)$, es decir, sobre el rango de la función $D_{\varphi_a}(p,q)$ con $h_a(0) = 0$ y η_a son pesos positivos.

Es claro que las medidas de Renyi y Sharma-Mittal, que no se podían obtener como un caso particular de las φ-divergencias, se pueden obtener ahora como un caso particular de las (h,φ)-divergencias, con $\Lambda = 1$, $\eta_a = 1$, y las siguientes funciones h y φ:

* Rényi

$h(x) = \frac{1}{r(r-1)}\log\left(1 + r(r-1)x\right); r \neq 0,1$

$\varphi(x) = \frac{x^r - r(x-1) - 1}{r(r-1)}; r \neq 0,1$

* Sharma y Mittal

$h(x) = \frac{1}{s-1}\left\{(1 + r(r-1)x)^{\frac{s-1}{r-1}} - 1\right\}; s, r \neq 1$

$\varphi(x) = \frac{x^r - r(x-1) - 1}{r(r-1)}; r \neq 0, 1.$

En el citado trabajo de Menéndez y otros (1995) se establece que la familia de estadísticos

$$T_n^{\varphi,h} = 2n \frac{D_\varphi^h(\hat{p}, p(\theta_0))}{\sum\limits_{a=1}^{\Lambda} \eta_a h_a'(0)\varphi_a''(1)} \tag{1.12}$$

se distribuye asintóticamente como una Ji-cuadrado con $M-1$ grados de libertad, independientemente de $h = (h_a)_{a\in\Lambda}$, $\varphi = (\varphi_a)_{a\in\Lambda}$ y $(\eta_a)_{a\in\Lambda}$. Mientras que Morales y otros (1995) establecieron que la familia de estadísticos

$$T_n^{\varphi,h,\phi} = 2n \frac{D_\varphi^h\left(\hat{p}, p\left(\hat{\theta}_\phi\right)\right)}{\sum\limits_{a=1}^{\Lambda} \eta_a h_a'(0)\varphi_a''(1)}, \tag{1.13}$$

con $\hat{\theta}_\phi$ definido en (1.10) se distribuye asintóticamente como una Ji-cuadrado con $M - M_0 - 1$ grados de libertad.

Es claro que de (1.11) se obtiene, para $h(x) = x$, $\Lambda = 1$ y $\eta_a = 1$, la familia de divergencias de Csiszár. Por tanto, buena parte de los estadísticos de bondad de ajuste, tanto en el caso de hipótesis nula simple como compuesta, se pueden obtener como un caso particular de $T_n^{\varphi,h}$ o bien de $T_n^{\varphi,h,\phi}$. En consecuencia, haciendo uso de los estadísticos dados en (1.12) y (1.13), se deberá rechazar la hipótesis nula de bondad de ajuste si

$$T_n^{\varphi,h} > \mathcal{X}^2_{M-1,\alpha},$$

en el caso de la hipótesis nula simple dada en (1.2), y

$$T_n^{\varphi,h,\phi} > \mathcal{X}^2_{M-M_0-1,\alpha}$$

en el caso de la hipótesis nula compuesta dada en (1.7).

Otras medidas de divergencia importantes son las R_ϕ- divergencias, las L-, K- y M-divergencias introducidas por Burbea y Rao (1982). Aplicaciones estadísticas interesantes de las R_ϕ- divergencias pueden verse en Pardo, M.C. y

Vajda (1997). Mientras que en Pérez, T. (2000) pueden verse aplicaciones a diversos problemas estadísticos de las L-, K- y M-divergencias.

Todas las divergencias hasta ahora presentadas están diseñadas para el estudio entre dos poblaciones, pero en ciertas aplicaciones como por ejemplo en Biología y Genética es necesario considerar k poblaciones.

Matusita (1967, 1973), propuso por primera vez, una generalización del coeficiente de Battacharya para expresar de forma cuantitativa analogías y diferencias entre k poblaciones, aplicándolo a técnicas analíticas de discriminación. Kaufman y Mathai (1973) ofrecieron una base axiomática en el caso discreto, y algunas propiedades de estas medidas fueron estudiadas por Toussaint (1974). Este último, además presentó una sencilla medida de divergencia, la J-divergencia entre k poblaciones. Gyorfi y Nemetz (1978) introdujeron una clase general de divergencias, llamada f-disimilaridad entre k poblaciones.

Sean $p_1, ..., p_k \in \Delta_M$, con $p_j = (p_{j1}, ..., p_{jM})^t$, entonces la f-disimilaridad entre las k poblaciones $p_1, ..., p_k$ se define mediante

$$D_f\left(p_1, ..., p_k\right) = \sum_{l=1}^{M} f\left(p_{1l}, ..., p_{kl}\right) \tag{1.14}$$

donde f es una función continua, convexa, homogénea y definida sobre

$$S = \left\{(s_1, ..., s_k) : 0 \leq s_i < \infty, i = 1, ..., k\right\}.$$

Si

$$f\left(x_1, ..., x_k\right) = \left(\prod_{j=1}^{k} x_j\right)^{\frac{1}{k}}$$

y

$$f\left(x_1, ..., x_k\right) = \prod_{j=1}^{k} x_j^{a_j}, a_j \geq 0$$

con $\sum\limits_{j=1}^{k} a_j = 1$, entonces la f-disimilaridad es la afinidad de Matusita cambiada de signo y la afinidad introducida por Toussaint, respectivamente.

Si,

$$f(x_1, ..., x_k) = \phi\left(\sum_{i=1}^{k} \lambda_i x_i\right) - \sum_{i=1}^{k} \lambda_i \phi(x_i)$$

se tiene la divergencia "Diferencia de Jensen" para k poblaciones dada por Burbea y Rao (1982). Tomando $\phi(x) = -x \log x$, se tiene el Radio de Información para k poblaciones. En Menéndez y otros (1992) pueden verse tres generalizaciones diferentes del Radio de Información para k poblaciones.

Además la f-disimilaridad también conduce a la φ-divergencia de Csiszár si $f(x_1, x_2) = x_2 \varphi\left(\dfrac{x_1}{x_2}\right)$.

Otras medidas de divergencia entre k poblaciones pueden encontrarse en Kapur (1988), Sahoo y Wong (1988), Rao (1982), Toussaint (1978) y otros.

Zografos (1994) obtuvo la distribución asintótica de la familia de divergencias introducidas por Gyorfi y Nemetz en un contexto multinomial y la utilizó para construir contrastes de homogeneidad. En poblaciones generales el problema fue abordado en Menéndez y otros (1997a).

1.3. Divergencias ponderadas

Tanto la medida de entropía de Shannon dada en (1.8) como las diferentes medidas de divergencias consideradas en la Sección 1.2 fueron introducidas como medidas cuantitativas de información, o lo que es equivalente, como medida de la incertidumbre, en el caso de la entropía de Shannon asociada a una medida de probabilidad $p \in \Delta_M$ y como medidas cuantitativas de discriminación entre dos medidas de probabilidad p y $q \in \Delta_M$ en el caso de las medidas de divergencia.

En un sistema cibernético (biológico o técnico) toda la actividad está dirigida hacia la realización de un fin cualquiera. El sistema debe disponer entonces de un criterio para poder diferenciar los sucesos. El criterio cibernético para una diferenciación cualitativa de los sucesos consiste en la importancia, la significación o la utilidad de la información que reportan respecto al fin. La aparición de un suceso elimina una doble "incertidumbre": una de orden cuantitativo relativa a la probabilidad de aparición y otra de orden cualitativo relativa a su utilidad en relación al fin. Este aspecto de orden cualitativo no queda recogido ni en la entropía de Shannon ni en las familias de divergencias consideradas en la Sección anterior.

La importancia del aspecto cualitativo se verá más claramente a través de los siguientes ajemplos: Así, frente a un viaje por una carretera en la que la probabilidad de sufrir un accidente es de 0.28, la entropía de Shannon ofrece el mismo valor que para el viaje por otra vía en la que la probabilidad de accidentes sea de 0.72. Desde luego, en general, nadie consideraría ambas situaciones como igualmente inciertas. Otro ejemplo: dos tratamientos médicos son tales que uno de ellos lleva a la curación total en el 90% de los casos, y a una mejora visible en el 10% restante, mientras que el segundo proporciona una mejora visible en el 90% y concluye con la curación total del paciente en el 10% restante. Los valores de entropía asociados son idénticos para los dos sistemas de probabilidades pero, evidentemente, nadie los consideraría como equiparables. Para evidenciar aún más estos hechos, se analizarán otras dos nuevas situaciones. La primera se refiere a una leyenda de la mitología griega. Teseo, partiendo para una expedición, ha prometido a su padre Egeo que, si consigue su hazaña, reemplazará a su vuelta la vela negra de su barco por una vela blanca. Podemos hacernos ahora la siguiente pregunta: ¿Cuál es la cantidad de información contenida en las velas? El segundo problema se reduce a la cuestión trivial: ¿Cuál es la cantidad de información que se obtiene cuando se lanza una moneda? Aunque completamente diferentes, estos dos problemas ponen en juego el mismo esquema. Las dos velas como las dos caras de la moneda representan dos sucesos equiprobables y encierran la misma cantidad de información, $\log_2 2 = 1$. Pero para Egeo, entre la información proporcionada por la vela negra y la proporcionada por la vela blanca había una distinción importante; así, al olvidarse Teseo de sustituir la vela negra por la blanca, su padre en la desesperación, al creerlo muerto, se lanzó al mar desde lo alto de una roca. Sin embargo, para la Teoría de la Información esta distinción no es esencial, si no se consideran las denominadas medidas ponderadas, tanto de entropía como de divergencia.

Belis y Guiasu (1968) adaptaron la entropía de Shannon a este esquema. Gil (1975) la redefinió de una forma más razonable. Supóngase que los elementos A_i, $i = 1, ..., M$ de la partición $\mathcal{A}$ son más o menos importantes con respecto al fin que se quiere alcanzar, es decir, tienen ponderaciones diferentes. La ponderación de un elemento de la partición puede ser independiente de su probabilidad. Por ejemplo, un elemento de la partición poco probable puede tener una gran ponderación, mientras que un elemento muy probable, incluso seguro, puede tener una ponderación nula para un fin cualquiera. En los trabajos de Bouchon (1976), Emptoz (1976), Houda y Tuteja (1981), Guiasu (1977, 1986), Longo (1972), Picard (1979) y Sharma, Mitler y Mohan (1978), por ejemplo, puede verse una motivación interesante acerca de la consideración de ponderaciones.

Para diferenciar los elementos de la partición respecto a un fin, se va a asociar a cada elemento A_k de $\mathcal{A}$, $k = 1, ..., M$, un número no negativo u_k, $k = 1, ..., M$, llamado "ponderación del elemento A_k". Los valores u_k, $k = 1, ..., M$, pueden ser cualesquiera números reales no negativos. En esta situación se denomina entropía de Shannon ponderada a la expresión

$$WH(p) = -\sum_{i=1}^{M} \frac{u_i}{E_p(u)} p_i \log p_i$$

donde $E_p(u) = \sum_{i=1}^{M} u_i p_i$.

El siguiente ejemplo analiza una situación real.

Ejemplo 1.1

Una naviera recibe el encargo de transportar petróleo desde un país productor a una cierta refinería, por lo que recibirá un pago de 280 millones de pesetas. La compañía en ese momento solo dispone de un petrolero antiguo cuyo coste de uso para este flete asciende, incluyendo gastos de seguro y amortización a 160 millones en caso de no producirse ninguna avería. Tras una cuidadosa revisión del barco, por parte de los técnicos de la compañía y teniendo en cuenta datos históricos correspondientes a 90 situaciones análogas en la propia compañía, se llega a la conclusión de que la probabilidad de avería es 1/5 y, en este caso las probabilidades de que la avería sea grave, media o ligera son respectivamente del 30%, 50% y 20%. A partir de estos datos se tiene la siguiente estimación de la distribución de probabilidad asociada a la variable aleatoria X (Resultado de la utilización del petrolero antiguo disponible):

	$\widehat{p}_i$
Sin avería	0.8
Con avería grave	0.06
Con avería media	0.1
Con avería ligera	0.04

Es claro que la incertidumbre asociada a esta variable aleatoria, y cuantificada en términos de la entropía de Shannon, viene dada por

$$H(\widehat{p}) = -0.8 \log 0.8 - 0.06 \log 0.06 - 0.1 \log 0.1 - 0.04 \log 0.04 = 0.70633$$

Pero es claro que esta medida no esta recogiendo la importancia que para la empresa tienen los diferentes resultados de la variable aleatoria en estudio. No será lo mismo para ella que el barco tenga avería a que no la tenga. Es más la compañía deberá evaluar las consecuencias asociadas a cada una de estas situaciones. Así por ejemplo despues de un exhaustivo estudio ha llegado a la conclusión de que en cada uno de los diferentes supuestos los gastos de transporte que la naviera deberá afrontar se elevarían respectivamente de los 160 millones iniciales a 360, 220 y 180 millones, respectivamente. Esto lleva a considerar las siguientes ponderaciones (en este caso costes de transporte) sobre cada uno de los resultados de la variable en estudio,

$$u = (160,\ 360,\ 220,\ 180).$$

Si se quiere tener una medida que tenga en cuenta tanto la incertidumbre asociada a las distintas situaciones como la importancia que tiene cada una de ellas para la empresa, sería necesario utilizar la entropía ponderada. Para ello se calculará previamente,

$$E_{\widehat{p}}(u) = 0.8 \times 160 + 0.06 \times 360 + 0.1 \times 220 + 180 \times 0.04 = 178.8$$

para seguidamente calcular

$$\begin{aligned} WH(X) = & \ \tfrac{1}{178.8} (-160 \times 0.8 \log 0.8 - 360 \times 0.06 \log 0.06 \\ & -220 \times 0.1 \log 0.1 - 180 \times 0.04 \log 0.04) \\ = & \ 0.91255 \end{aligned} .$$

Esta sería una medida mas realista para medir tanto la incertidumbre asociada a la aparición de los diferentes resultados como la importancia de los mismos para la empresa.

Un amplio estudio de las propiedades y comportamiento de la entropía de Shannon ponderada puede verse en Gil, Pardo y Gil (1993).

En relación con las medidas de divergencia fue Taneja, H.C. (1985) quien por primera vez introdujo el concepto de medida de divergencia ponderada. De forma más precisa, dadas dos medidas de probabilidad p y $q \in \Delta_M$ y una ponderación

u_i sobre el elemento A_i de la partición $\mathcal{A}$, $i = 1, ..., M$, definió la medida de divergencia de Kullback-Leibler ponderada, en los siguientes y términos,

$$WD(p, q) = \sum_{i=1}^{M} u_i p_i \log \frac{p_i}{q_i}.$$

Además estudió la distribución asintótica del estadístico $WD(\hat{p}, q)$ donde $\hat{p}$ está definida en (1.1).

No hay que profundizar demasiado para advertir que la expresión anterior plantea diversos problemas. En el caso de que $u_i = u \ \forall i = 1, ..., M$, parecería lógico que se obtuviera directamente la divergencia de Kullback-Leibler y no ésta multiplicada por la constante u. Este problema es sencillo de solucionar sin más que multiplicar la expresión de la divergencia Kullback dada anteriormente bien por

$$E_p(u)^{-1} = \left(\sum_{i=1}^{M} u_i p_i \right)^{-1},$$

bien por

$$E_q(u)^{-1} = \left(\sum_{i=1}^{M} u_i q_i \right)^{-1}.$$

No obstante aún siendo importante este problema más lo es el siguiente:

Se consideran las distribuciones de probabilidad

$$p = (p, 1-p)^t \qquad q = (1-p, p)^t$$

y se tiene

$$WD(p, q) = \log\left(\frac{1-p}{p}\right)(u_2(1-p) - u_1 p).$$

Es claro que si

$$\begin{array}{c} 1-p > p \quad \text{y} \quad \dfrac{u_2}{u_1} < \dfrac{p}{1-p} \\ \text{ó} \\ 1-p < p \quad \text{y} \quad \dfrac{u_2}{u_1} > \dfrac{p}{1-p} \end{array}$$

entonces

$$WD\left(p,q\right) < 0.$$

Con el fin de que la medida de divergencia ponderada resultante tenga aplicabilidad en el contraste de hipótesis de bondad de ajuste sería importante el conseguir que $WD\left(p,q\right) \geq 0$ y que la igualdad se diera si $p = q$. Una forma de resolver estos dos problemas es redefinir la divergencia de Kullback-Leibler con ponderaciones en los siguientes términos,

$$WD\left(p,q\right) = \frac{1}{E_p\left(u\right)} \sum_{i=1}^{M} u_i \left(p_i \log \frac{p_i}{q_i} + q_i\right) - 1.$$

En este caso es sencillo comprobar que

$$WD\left(p,q\right) \geq 0$$

y toma el valor cero cuando $p = q$. Esto es obvio sin más que observar que la función $g\left(x\right) = x \log x - x + 1$ es convexa con $g\left(1\right) = 0$ y $g'\left(1\right) = 0$.

En esta memoria se introduce una familia de divergencias basada en las (h, φ)-divergencias y que no presenta los problemas señalados anteriormente: Las (h, φ)-divergencias ponderadas. Esta familia de divergencias ponderadas se define en los siguientes términos

$$WD_{\varphi}^{h}\left(p,q\right) = \sum_{a=1}^{\Lambda} \eta_a h_a \left(\sum_{i=1}^{M} \frac{u_i q_i}{E_p\left(u\right)} \varphi_a \left(\frac{p_i}{q_i}\right)\right) \tag{1.15}$$

donde

- $h = \left(h_a\right)_{a=1,\ldots,\Lambda}$, $\varphi = \left(\varphi_a\right)_{a=1,\ldots,\Lambda}$, h_a y φ_a son funciones de clase 2 para $a = 1, ..., \Lambda$.
- φ_a satisface las condiciones de definición de la divergencia de Csiszár con $\varphi_a\left(1\right) = \varphi_a'\left(1\right) = 0$.
- h_a son funciones no decrecientes y continuas con $h_a\left(0\right) = 0$, para $a = 1, ..., \Lambda$.
- η_a, $a = 1, ..., \Lambda$ son números positivos

- u_i es la ponderación asociada al elemento A_i, $i = 1, ..., M$, de la partición $\mathcal{A}$.

Partiendo del Ejemplo 1.1 se considerará una nueva situación que pone de manifiesto el alcance de la familia de divergencias ponderadas introducidas en (1.15).

Ejemplo 1.2

Continuando con el Ejemplo 1.1, supongamos ahora que la empresa no se fía, por experiencias anteriores, de las probabilidades asignadas por sus técnicos y pide un estudio a una consultoría. Esta consultoría establece, teniendo en cuenta datos históricos correspondientes a 150 situaciones análogas en la propia compañía y en otras compañías, una estimación para la probabilidad de avería de 0.3 *y en este caso sus estimaciones de que la avería sea grave, media o ligera son respectivamente del 28%, 60% y 12%. Entonces se tiene*

	$\widehat{q}_i$
Sin avería	0.7
Con avería grave	$0.3 \times 0.28 = 0.084$
Con avería media	$0.3 \times 0.6 = 0.18$
Con avería ligera	$0.3 \times 0.12 = 0.036$

Ahora se puede plantear el cálculo de la (h, φ)-divergencia ponderada dada en (1.15) *y se pueden hacer consideraciones del tipo contrastar*

$$H_0 : p = q,$$

teniendo en cuenta la diferente importancia que tiene para la naviera los distintos resultados de la variable, utilizando $WD_{\varphi}^{h}(\widehat{p}, \widehat{q})$.

La no negatividad de la familia de divergencias dada en (1.15), así como su valor nulo cuando $p = q$, se sigue ya que φ_a, $a = 1, ..., \Lambda$, se supone que son funciones convexas dos veces diferenciables con $\varphi_a(1) = \varphi_a'(1) = 0$, $a = 1, ..., \Lambda$. Así pues, $WD_{\varphi}^{h}(p, q) \geq 0$ y se da la igualdad para $p = q$, pero no es cierto que si $WD_{\varphi}^{h}(p, q) = 0$, entonces se tenga que verificar que $p = q$, como pone de manifiesto el siguiente ejemplo:

Ejemplo 1.3

Sean $p = (p, 1-p)$, $q = (1-p, p)$, $u_1 = u_2 = \frac{1}{2}$, $\Lambda = 5$, $\varphi_a(x) = \frac{1}{2} + \frac{x}{2} - \sqrt{x}$,

$a = 1, 2, 3, 4$ *y* 5, *y* $h_a(x) = \text{máx}\left\{0, x\left(x^2 + 3a^2\right) - a\left(a^2 + 3x^2\right)\right\}$, $a = 1, 2, 3, 4$ *y* 5.

Es claro que $E_p(u) = \frac{1}{2}$ *y*

$$\begin{aligned} WD_\varphi^h(p,q) &= \sum_{a=1}^{5} \eta_a h_a\left((1-p)\left(\frac{1}{2} + \frac{p}{2(1-p)} - \sqrt{\frac{p}{1-p}}\right)\right. \\ &\left. +p\left(\frac{1}{2} + \frac{1-p}{2p} - \sqrt{\frac{1-p}{p}}\right)\right). \end{aligned}$$

Para $0 < p < 1$, *la función*

$$f(p) = (1-p)\left(\frac{1}{2} + \frac{p}{2(1-p)} - \sqrt{\frac{p}{1-p}}\right) + p\left(\frac{1}{2} + \frac{1-p}{2p} - \sqrt{\frac{1-p}{p}}\right)$$

varía entre 0 y 1, tomando el valor 0 para $p = \frac{1}{2}$.

Por otro lado, las funciones

$$g_a(x) = x\left(x^2 + 3a^2\right) - a\left(a^2 + 3x^2\right)$$

para $a = 1, 2, 3, 4$ *,5 y* $x \in [0, 1]$ *son negativas, por lo tanto*

$$h_a(f(p)) = 0, \ a = 1, 2, 3, 4, 5$$

y como consecuencia

$$WD_\varphi^h(p,q) = \sum_{a=1}^{5} \eta_a h_a(f(p)) = 0.$$

Aunque no será objeto de estudio en esta memoria conviene resaltar que la f-disimilaridad introducida por Gyorfi y Nemetz y definida en (1.14) admite una clara adaptación al caso en el que sobre los elementos A_i de la partición $\mathcal{A}$, $i = 1, ..., M$, existan definidas unas ponderaciones, ya que en este caso se podría definir la f-disimilaridad ponderada en los siguientes términos

$$WD_f(p_1, ..., p_k) = \sum_{l=1}^{M} \frac{u_l}{E_{p_1}(u)^{1/k} ......E_{p_k}(u)^{1/k}} f(p_{1l}, ..., p_{kl}).$$

Si bien esta medida de divergencia entre k poblaciones se define por primera vez en esta memoria, no será objeto de estudio en la misma y constituye un problema abierto para posteriores estudios.

En los siguientes capítulos se centrará nuestro estudio en la familia de las (h, φ)-divergencias ponderadas introducida en (1.15) de acuerdo al esquema señalado en la introducción de esta memoria.

2

Estimadores de Mínima (h, ϕ)-divergencia ponderada

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 o 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 método de máxima verosimilitud son expuestos por Le Cam (1990). Por ejemplo, supongamos que $X_1, X_2, ..., X_n$ son variables aleatorias independientes y distribuidas como una mixtura de dos poblaciones normales con función de densidad

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

donde $\theta = (\mu_1, \mu_2, \sigma_1, \sigma_2, w)$, $\mu_1, \mu_2 \in R$, $\sigma_1, \sigma_2 > 0$ y $w \in (0, 1)$. La función de verosimilitud para estimar los cinco parámetros de esta distribución viene dada por

$$L(\theta; x_1, ...x_n) = \prod_{j=1}^{n} f_\theta(x_j).$$

Si se hace $\mu_1 = \mu_2 = x_i$ para algún i $(i = 1, ..., n)$, entonces

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

y

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

De esta forma

$$L(\theta; x_1, ...x_n) > (2\pi)^{-n/2} w(1-w)^{n-1}\sigma_1^{-1}\sigma_2^{-(n-1)}\exp\left[-\frac{1}{2}\sum_{j\neq i}\left(\frac{x_j - x_i}{\sigma_2}\right)^2\right]$$

donde eligiendo σ_1 suficientemente pequeño, se puede hacer L tan grande como se quiera. Por tanto no existen valores w, σ_1, σ_2, μ_1 y μ_2 que maximicen L. Es decir, no siempre existe estimador de máxima verosimilitud basado en los datos originales.

El primero en dar una 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 no es acotada. Posteriormente, Fryer y Robertson (1972) compararon las estimaciones de los momentos y los de máxima verosimilitud y mínima $\mathcal{X}^2$ para los datos agrupados de los parámetros de varias mixturas 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.

Un procedimiento, que entre otros, resuelve el problema es a través del modelo discretizado. Dado el espacio estadístico $(\mathcal{X}, \beta_\mathcal{X}, P_\theta)_{\theta\in\Theta\subseteq R^{M_0}}$, se considera la partición fija $\mathcal{A} = \{A_1, ..., A_M\} \subset \beta_\mathcal{X}$, de $\mathcal{X}$, con

$$p_j(\theta) = P_\theta(A_j), \qquad j = 1, ..., M, \tag{2.1}$$

y la variable aleatoria M-dimensional $(N_1, ..., N_M)$, basada en una muestra aleatoria simple $X_1, ..., X_n$, con

$$N_j = \sum_{i=1}^{n} I_{A_j}(X_i).$$

Esta variable M-dimensional, $(N_1, ..., N_M)$, es un estadístico suficiente para el modelo bajo consideración y multinomialmente distribuida con parámetros

$$(n; p_1(\theta), ..., p_M(\theta)).$$

El problema de estimar θ, por máxima verosimilitud una vez agrupados los datos consiste en maximizar para $(n_1, ..., n_M)$ fijo

$$P_\theta(N_1 = n_1, ..., N_M = n_M) = \frac{n!}{n_1!...n_M!} p_1(\theta)^{n_1} ... p_M(\theta)^{n_M}$$

o equivalentemente

$$\ln P_\theta(N_1 = n_1, ..., N_M = n_M) = -nD_{Kullback}(\hat{p}, q(\theta)) + cte$$

siendo $\hat{p} = (\hat{p}_1, ..., \hat{p}_M)^t$ con $\hat{p}_i = \frac{n_i}{n}$, $i = 1, ..., M$, $p(\theta) = (p_1(\theta), ..., p_M(\theta))^t$ y $D_{Kullback}$ la divergencia de Kullback-Leibler. Por tanto estimar θ a través del modelo discretizado mediante máxima verosimilitud es equivalente a minimizar en $\theta \in \Theta \subseteq R^{M_0}$ la divergencia de Kullback-Leibler.

Ahora bien la divergencia de Kullback-Leibler, según se vio en el Capítulo 1, 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

$$D\left(\hat{p}, p\left(\tilde{\theta}\right)\right) = \inf_{\theta \in \Theta \subseteq R^{M_0}} D(\hat{p}, p(\theta))$$

siendo D cualquier medida de divergencia. Para más detalles acerca de este procedimiento de estimación, ver Pardo, L. (1997b).

A lo largo de este capítulo se abordará el problema de estimación basándonos en las (h, ϕ)-divergencias ponderadas ya que se supondrá como se hizo ver en el apartado 1.3 del Capítulo 1 que los elementos A_j, $j = 1, ..., M$, asociados a la partición $\mathcal{A}$ están ponderados.

2.2. El estimador de mínima WD_{ϕ}^{h}-divergencia

Sea $X_1, ..., X_n$ una muestra aleatoria simple procedente de una población dependiente de un parámetro desconocido $\theta \in \Theta \subseteq R^{M_0}$. Supóngase que existe la función $p(\theta) = (p_1(\theta), ..., p_M(\theta))^t$ siendo $p_j(\theta)$ el definido en (2.1) que aplica cada valor de $\theta = (\theta_1, ..., \theta_{M_0})^t$ en Δ_M.

Cuando θ varía sobre Θ, $p(\theta)$ varía sobre un subconjunto $\Delta_{M,\theta}$ de Δ_M. Si el modelo elegido es correcto existirá un valor $\theta_0 \in \Theta$ de tal forma que $p(\theta_0) = \pi$ donde π es el verdadero valor de la probabilidad de la multinomial, es decir, $\pi \in \Delta_{M,\theta}$. En caso de que el modelo no sea correcto, en general $\pi \notin \Delta_{M,\theta}$, es decir, no existirá el valor $\theta_0 \in \Theta$ tal que $p(\theta_0) = \pi$.

Definición 2.1

Dada una muestra aleatoria simple de tamaño n, $X_1, ..., X_n$, procedente de una población con espacio estadístico $(\mathcal{X}, \beta_{\mathcal{X}}, P_\theta)_{\theta \in \Theta \subseteq R^{M_0}}$, el estimador de mínima (h, ϕ)-divergencia ponderada de θ, basado en la partición $\mathcal{A} = \{A_1, ..., A_M\} \subset \beta_{\mathcal{X}}$ de $\mathcal{X}$, es cualquier $\hat{\theta}_{\phi,h,w} \in \Theta$ que verifique

$$WD_{\phi}^{h}\left(\hat{p}, p\left(\hat{\theta}_{h,\phi,w}\right)\right) = \inf_{\theta \in \Theta \subseteq R^{M_0}} WD_{\phi}^{h}(\hat{p}, p(\theta)),$$

siendo $\hat{p} = (\hat{p}_1, ..., \hat{p}_M)^t$ y $\hat{p}_j$, $j = 1, ..., M$, está definido en (1.1).

En lo sucesivo el estimador de mínima (h, ϕ)-divergencia ponderada se denominará estimador de mínima WD_{ϕ}^{h}-divergencia y se expresará mediante

$$\hat{\theta}_{h,\phi,w} = \arg \inf_{\theta \in \Theta} WD_{\phi}^{h}(\hat{p}, p(\theta))$$

siendo

$$WD_{\phi}^{h}(\hat{p}, p(\theta)) = \sum_{a=1}^{\Lambda} \eta_a h_a \left(\sum_{i=1}^{M} \frac{u_i}{E_{\hat{p}}(u)} p_i(\theta) \phi_a \left(\frac{\hat{p}_i}{p_i(\theta)} \right) \right)$$

con u_i la ponderación del elemento A_i de la partición $\mathcal{A}$ y suponiendo que las funciones h_a, ϕ_a $a = 1, ..., \Lambda$ verifican las condiciones señaladas en (1.15).

Este método elige el punto de $\Delta_{M,\theta}$ que está más próximo al valor $\hat{p}$ en el sentido de la distancia elegida.

Veamos un ejemplo.

Ejemplo 2.1

Sea X una variable aleatoria que toma los valores x_1, x_2 y x_3 con probabilidades respectivas $\frac{1}{3}-\theta$, $\frac{2}{3}-\theta$ y 2θ siendo $0<\theta<\frac{1}{3}$. Supuesto que los resultados de x_1, x_2 y x_3 están ponderados por u_1, u_2 y u_3 respectivamente, se obtendrá el estimador de mínima WD_ϕ^h-divergencia, basado en una muestra de tamaño n, y supuesto que $\Lambda=1$, $\eta_1=1$, $h(x)=x$ y

$$\phi(x)=\frac{1}{2}\left(\frac{1}{x}+x-2\right).$$

Al ser,

$$\begin{aligned} WD_\phi^h(\hat{p},q(\theta)) &= \frac{1}{2}\sum_{i=1}^{3}\frac{u_i}{E_{\hat{p}}(u)}p_i(\theta)\left(\frac{p_i(\theta)}{\hat{p}_i}+\frac{\hat{p}_i}{p_i(\theta)}-2\right)\\ &= \frac{1}{2}\sum_{i=1}^{3}\frac{u_i}{E_{\hat{p}}(u)}\frac{n}{n_i}\left(p_i(\theta)-\hat{p}_i\right)^2, \end{aligned}$$

el cálculo del estimador de mínima WD_ϕ^h-divergencia se obtendrá al minimizar en θ, la función

$$g(\theta)=\frac{1}{2E_{\hat{p}}(u)}\left[\frac{u_1 n}{n_1}\left(\frac{1}{3}-\theta-\frac{n_1}{n}\right)^2+\frac{u_2 n}{n_2}\left(\frac{2}{3}-\theta-\frac{n_2}{n}\right)^2+\frac{u_3 n}{n_3}\left(2\theta-\frac{n_3}{n}\right)^2\right].$$

Derivando respecto de θ e igualando a cero se tiene,

$$g'(\theta)=\frac{u_1}{n_1}\left(\frac{1}{3}-\theta-\frac{n_1}{n}\right)(-1)+\frac{u_2}{n_2}\left(\frac{2}{3}-\theta-\frac{n_2}{n}\right)(-1)+\frac{u_3}{n_3}\left(2\theta-\frac{n_3}{n}\right)2=0.$$

Luego,

$$-\frac{u_1}{3n_1}+\frac{n_1}{n}-\frac{2u_2}{3n_2}+\frac{n_2}{n}-\frac{2u_3}{n}=\theta\left(-\frac{u_1}{n_1}-\frac{u_2}{n_2}-\frac{4u_3}{n_3}\right)$$

y en consecuencia

$$\hat{\theta}_{h,\phi,w}(n_1,n_2,n_3,u_1,u_2,u_3)=\frac{\dfrac{u_1}{3n_1}+\dfrac{2u_2}{3n_2}+\dfrac{1}{n}(2u_3-u_1-u_2)}{\dfrac{u_1}{n_1}+\dfrac{u_2}{n_2}+\dfrac{4u_3}{n_3}}.$$

Obsérvese que para $u_1 = u_2 = u_3$ se obtiene

$$\hat{\theta}_{\phi}(n_1, n_2, n_3) = \frac{\dfrac{1}{3n_1} + \dfrac{2}{3n_2}}{\dfrac{1}{n_1} + \dfrac{1}{n_2} + \dfrac{4}{n_3}} \tag{2.2}$$

y además $\phi(x) = \phi_{(-2)}(x)$ con $\phi_{(-2)}(x)$ la función asociada a la divergencia de Cressie y Read con $\lambda = -2$. Así, el estimador obtenido en (2.2) se corresponde con el estimador de mínima Ji-cuadrado modificado o Neyman modificado, obtenido al minimizar en θ la expresión

$$h(\theta) = \frac{1}{2}\sum_{j=1}^{3}\frac{1}{\hat{p}_j}\left(p_j(\theta) - \hat{p}_j\right)^2.$$

En lo que sigue se supondrá que el modelo es correcto, $p(\theta_0) = \pi$ 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. $\pi_i = p_i(\theta_0) > 0$, $i = 1, ..., M$, por tanto $\pi = (\pi_1, ..., \pi_M)^t$ es un punto interior de $\Delta_{M,\theta}$.

3. La aplicación $p : \Theta \to \Delta_M$ es totalmente diferenciable en θ_0. Por tanto existen las derivadas parciales de p_i con respecto a cada θ_j en θ_0 y $p_i(\theta)$ se puede expresar de la forma:

$$p_i(\theta) = p_i(\theta_0) + \sum_{i=1}^{M}(\theta_j - \theta_0)\frac{\partial p_i(\theta_0)}{\partial \theta_j} + o(\|\theta - \theta_0\|) \text{ cuando } \theta \to \theta_0,$$

   donde $\lim\limits_{\theta \to \theta_0} \dfrac{o(\|\theta - \theta_0\|)}{\theta - \theta_0} = 0.$

4. La matriz

$$J(\theta_0) = \left(\frac{\partial p(\theta_0)}{\partial \theta}\right)_{\theta = \theta_0} = \left(\frac{\partial p_i(\theta_0)}{\partial \theta_j}\right)_{\substack{i=1,...,M \\ j=1,...,M_0}} \tag{2.3}$$

   tiene rango M_0.

5. La aplicación inversa $p^{-1} : \Delta_{M,\theta} \to \Theta$ es continua y $p(\theta_0) = \pi$.

6. La aplicación $p : \Theta \to \Delta_M$ es continua para todo $\theta \in \Theta$.

2.3. Propiedades asintóticas del estimador de mínima WD_ϕ^h-divergencia

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

$$\theta_{h,\phi,w}(p) = \arg \inf_{\theta \in \Theta} WD_\phi^h(p, p(\theta))$$

se le denominará en lo que sigue función de mínima WD_ϕ^h-divergencia. Obsérvese que dada una muestra aleatoria simple de tamaño n y la distribución de frecuencias, $\hat{p}$, asociada a ella, el estimador de θ de mínima WD_ϕ^h-divergencia viene dado por el valor de la función de mínima WD_ϕ^h-divergencia en $\hat{p}$, es decir,

$$\hat{\theta}_{h,\phi,w} = \theta_{h,\phi,w}(\hat{p}).$$

Los siguientes teoremas establecen propiedades asintóticas del estimador de mínima WD_ϕ^h-divergencia. El primero de ellos establece la convergencia casi seguro del estimador de mínima WD_ϕ^h-divergencia.

Teorema 2.1

Sea Θ compacto y las funciones h_a, ϕ_a, $a \in \Lambda$, verifican las condiciones requeridas en (1.15). *Suponiendo que se verifican las condiciones de regularidad 1-6 de Birch y* $\arg \inf_{\theta \in \Theta} WD_\phi^h(p, p(\theta))$ *es único en un entorno cerrado de π, se tiene*

$$\hat{\theta}_{h,\phi,w} \underset{n \to \infty}{\overset{c.s.}{\to}} \theta_0.$$

Demostración

Se comenzará probando que $\theta_{h,\phi,w}$ definida mediante

$$\begin{array}{llll} \theta_{h,\phi,w}: & \Delta_M & \to & \Theta \\ & p & \to & \theta_{h,\phi,w}(p) = \arg \inf_{\theta \in \Theta} WD_\phi^h(p, p(\theta)). \end{array}$$

es una función continua.

Se establecerá que dada una sucesión $\{p_n\}_{n \in N}$ con $p_n \in \Delta_M$, $\forall n$, tal que $\lim_{n \to \infty} p_n = \pi$ y supuesto que se hace la hipótesis

$$\lim_{n \to \infty} \theta_{h,\phi,w}(p_n) = \lim_{n \to \infty} \arg \inf_{\theta \in \Theta} WD_\phi^h(p_n, p(\theta)) \neq \theta_{h,\phi,w}(\pi)$$

entonces se está en contradicción con las hipótesis formuladas en el enunciado del teorema. Esto conducirá a la continuidad de la función $\theta_{h,\phi,w}$.

Como por hipótesis Θ es un conjunto compacto existe una subsucesión

$$\left\{\theta_{h,\phi,w}\left(p_m\right)\right\}_{m\in N} \subset \left\{\theta_{h,\phi,w}\left(p_n\right)\right\}_{n\in N}$$

verificando

$$\lim_{m\to\infty} \theta_{h,\phi,w}\left(p_m\right) = \theta_0^* \neq \theta_{h,\phi,w}\left(\pi\right).$$

Por hipótesis $WD_\phi^h\left(p, p\left(\theta\right)\right)$ es una función continua de θ $\forall p \in \Delta_M$, luego

$$WD_\phi^h\left(p, p\left(\theta_0^*\right)\right) = \lim_{m\to\infty} WD_\phi^h\left(p, p\left(\theta_{h,\phi,w}\left(p_m\right)\right)\right).$$

Por otra parte, como

$$\lim_{n\to\infty} WD_\phi^h\left(p_n, p\left(\theta\right)\right) = WD_\phi^h\left(\pi, p\left(\theta\right)\right) \qquad \forall\theta \in \Theta$$

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

$$\lim_{n\to\infty} \sup_{\theta\in\Theta} \left| WD_\phi^h\left(p_n, p\left(\theta\right)\right) - WD_\phi^h\left(\pi, p\left(\theta\right)\right)\right| = 0, \tag{2.4}$$

y por tanto

$$\lim_{n\to\infty} \left| \inf_{\theta\in\Theta} WD_\phi^h\left(p_n, p\left(\theta\right)\right) - \inf_{\theta\in\Theta} WD_\phi^h\left(\pi, p\left(\theta\right)\right)\right| = 0. \tag{2.5}$$

Al ser

$$\inf_{\theta\in\Theta} WD_\phi^h\left(p_n, p\left(\theta\right)\right) = WD_\phi^h\left(p_n, p\left(\theta_{h,\phi,w}\left(p_n\right)\right)\right)$$

e

$$\inf_{\theta\in\Theta} WD_\phi^h\left(\pi, p\left(\theta\right)\right) = WD_\phi^h\left(\pi, p\left(\theta_{h,\phi,w}\left(\pi\right)\right)\right)$$

se tiene de (2.5) que

$$\lim_{n\to\infty} WD_\phi^h\left(p_n, p\left(\theta_{h,\phi,w}\left(p_n\right)\right)\right) = WD_\phi^h\left(\pi, p\left(\theta_{h,\phi,w}\left(\pi\right)\right)\right). \tag{2.6}$$

De (2.4) se deduce que

$$\lim_{n\to\infty} \left| WD_\phi^h\left(p_n, p\left(\theta_{h,\phi,w}\left(p_n\right)\right)\right) - WD_\phi^h\left(\pi, p\left(\theta_{h,\phi,w}\left(p_n\right)\right)\right)\right| = 0 \tag{2.7}$$

por lo que estamos en condiciones de concluir de, (2.6) y (2.7), que

$$\lim_{n\to\infty} WD_\phi^h\left(\pi, p\left(\theta_{h,\phi,w}\left(p_n\right)\right)\right) = WD_\phi^h\left(\pi, p\left(\theta_{h,\phi,w}\left(\pi\right)\right)\right).$$

Entonces,

$$WD_\phi^h\left(\pi, p\left(\theta_0^*\right)\right) = WD_\phi^h\left(\pi, p\left(\theta_{h,\phi,w}\left(\pi\right)\right)\right)$$

lo cual está en contradicción con la hipótesis de que $\arg\inf_{\theta\in\Theta} WD_\phi^h\left(p, p\left(\theta\right)\right)$ es único en un entorno cerrado de π y por tanto $\theta_{h,\phi,w}$ es una función continua.

Al ser

$$\hat{p} \underset{n\to\infty}{\overset{c.s.}{\to}} \pi = p\left(\theta_0\right),$$

con $\hat{p}$ definido en (1.1), se tiene por la continuidad de $\theta_{h,\phi,w}$ que

$$\hat{\theta}_{h,\phi,w} = \theta_{h,\phi,w}\left(\hat{p}\right) \overset{c.s.}{\to} \theta_{h,\phi,w}\left(p\left(\theta_0\right)\right) = \theta_0.$$

$\blacksquare$

El siguiente teorema da una descomposición para el estimador de mínima WD_ϕ^h-divergencia.

Teorema 2.2

Bajo las hipótesis requeridas a las funciones ϕ_a y h_a, $a = 1, ..., \Lambda$, en (1.15), supuesto que se verifican las condiciones de regularidad 1-6 de Birch y suponiendo que la aplicación $p : \Theta \to \Delta_M$ tiene segundas derivadas parciales continuas en un entorno de θ_0, se tiene

$$\hat{\theta}_{h,\phi,w} = \theta_0 + B^*\left(\theta_0\right) diag\left(\mathcal{U}_*^{1/2}\right)\left(\hat{p} - p\left(\theta_0\right)\right) + o\left(\left\|\hat{p} - p\left(\theta_0\right)\right\|\right)$$

donde $\hat{\theta}_{h,\phi,w}$ es único en un entorno de θ_0,

$$B^*\left(\theta_0\right) = \left(A^*\left(\theta_0\right)^t A^*\left(\theta_0\right)\right)^{-1} A^*\left(\theta_0\right)^t$$

$$A^*\left(\theta_0\right)^t = diag\left(\mathcal{U}_*^{1/2}\right)_{M\times M}\left(\left(\frac{\partial p_i\left(\theta_0\right)}{\partial\theta_r}\right)\right)^t_{\substack{i=1,\dots,M\\ r=1,\dots,M_0}}$$

y $\mathcal{U}_*$ es el vector definido mediante,

$$\mathcal{U}_* = \left(\frac{u_1}{p_1\left(\theta_0\right)}, ..., \frac{u_M}{p_M\left(\theta_0\right)}\right)^t.$$

Demostración

Sea l^M el interior del cubo unitario M-dimensional con $\Delta_M \subset l^M$ y sea U un entorno de θ_0 en el cual $p : \Theta \to \Delta_M$ tiene derivadas segundas continuas. Considérese la función

$$F = (F_1, ..., F_{M_0}) : l^M \times U \to R^{M_0}$$

de tal forma que

$$F_j (p_1, ..., p_M; \theta_1, ..., \theta_{M_0}) = \frac{\partial W D_\phi^h (p, p(\theta))}{\partial \theta_j}, \quad j = 1, ..., M_0$$

siendo

$$W D_\phi^h (p, p(\theta)) = \sum_{a=1}^{\Lambda} \eta_a h_a \left(\sum_{i=1}^{M} \frac{u_i}{E_p(u)} p_i(\theta) \phi_a \left(\frac{p_i}{p_i(\theta)} \right) \right).$$

Para $\pi_i = p_i(\theta_0)$, $i = 1, ..., M$, se tiene que

$$F_j (\pi_1, ..., \pi_M; \theta_{01}, ..., \theta_{0M_0}) = 0 \quad \forall j = 1, ..., M_0.$$

En efecto, llamando

$$A = \frac{\partial W D_\phi^h (p, p(\theta))}{\partial \theta_j}$$

se tiene $\forall j = 1, ..., M_0$

$$\begin{aligned} A &= \sum_{a=1}^{\Lambda} \eta_a h_a' \left(\sum_{i=1}^{M} \frac{u_i}{\sum\limits_{i=1}^{M} u_i p_i} p_i(\theta) \phi_a \left(\frac{p_i}{q_i(\theta)} \right) \right) \\ &\quad \times \frac{1}{\sum\limits_{i=1}^{M} u_i p_i} \left[\sum_{i=1}^{M} u_i \frac{\partial p_i(\theta)}{\partial \theta_j} \left(\phi_a \left(\frac{p_i}{p_i(\theta)} \right) - \phi_a' \left(\frac{p_i}{p_i(\theta)} \right) \frac{p_i}{p_i(\theta)} \right) \right] \end{aligned}$$

y entonces

$$\begin{aligned} F_j (\pi_1, ..., \pi_M; \theta_{01}, ..., \theta_{0M_0}) &= \sum_{a=1}^{\Lambda} \eta_a h_a'(0) \frac{1}{\sum\limits_{i=1}^{M} u_i \pi_i} \sum_{i=1}^{M} u_i \frac{\partial p_i(\theta_0)}{\partial \theta_j} \big(\phi_a(1) - \phi_a'(1) \big) \\ &= 0 \quad \forall j = 1, ..., M_0. \end{aligned}$$

ya que por hipótesis $\phi_a(1) = \phi'_a(1) = 0$.

Seguidamente se comprueba que la matriz

$$\frac{\partial F}{\partial \theta} = \left(\frac{\partial F_i}{\partial \theta_j} \right)_{\substack{i=1,...,M \\ j=1,...,M_0}}$$

es no singular en $(\pi_1, ..., \pi_M; \theta_{01}, ..., \theta_{0M_0})$.

Al ser,

$$\begin{aligned}
\frac{\partial}{\partial \theta_r}(A) &= \sum_{a=1}^{\Lambda} \eta_a h''_a \left(\sum_{i=1}^{M} \frac{u_i}{E_p(u)} p_i(\theta) \phi_a \left(\frac{p_i}{p_i(\theta)} \right) \right) \frac{1}{(E_p(u))^2} \left[\sum_{i=1}^{M} u_i \frac{\partial p_i(\theta)}{\partial \theta_j} \left(\phi_a \left(\frac{p_i}{p_i(\theta)} \right) \right. \right. \\
&\quad \left. \left. - \phi'_a \left(\frac{p_i}{p_i(\theta)} \right) \frac{p_i}{p_i(\theta)} \right) \right] \left[\sum_{i=1}^{M} u_i \frac{\partial p_i(\theta)}{\partial \theta_r} \left(\phi_a \left(\frac{p_i}{p_i(\theta)} \right) - \phi'_a \left(\frac{p_i}{p_i(\theta)} \right) \frac{p_i}{p_i(\theta)} \right) \right] \\
&\quad + \sum_{a=1}^{\Lambda} \eta_a h'_a \left(\sum_{i=1}^{M} \frac{u_i}{\sum_{i=1}^{M} u_i p_i} p_i(\theta) \phi_a \left(\frac{p_i}{p_i(\theta)} \right) \right) \left[\sum_{i=1}^{M} \frac{u_i}{\sum_{i=1}^{M} u_i p_i} \frac{\partial^2 p_i(\theta)}{\partial \theta_j \partial \theta_r} \phi_a \left(\frac{p_i}{p_i(\theta)} \right) \right. \\
&\quad - \sum_{i=1}^{M} \frac{u_i}{\sum_{i=1}^{M} u_i p_i} \frac{\partial p_i(\theta)}{\partial \theta_j} \phi'_a \left(\frac{p_i}{p_i(\theta)} \right) \frac{p_i}{p_i(\theta)^2} \frac{\partial p_i(\theta)}{\partial \theta_r} \\
&\quad - \sum_{i=1}^{M} \frac{u_i}{\sum_{i=1}^{M} u_i p_i} \frac{\partial p_i(\theta)}{\partial \theta_r} \phi'_a \left(\frac{p_i}{p_i(\theta)} \right) \frac{p_i}{p_i(\theta)^2} \frac{\partial p_i(\theta)}{\partial \theta_j} \\
&\quad + \sum_{i=1}^{M} \frac{u_i p_i(\theta)}{\sum_{i=1}^{M} u_i p_i} \phi''_a \left(\frac{p_i}{p_i(\theta)} \right) \frac{p_i}{p_i(\theta)^2} \frac{\partial p_i(\theta)}{\partial \theta_r} \frac{p_i}{p_i(\theta)^2} \frac{\partial p_i(\theta)}{\partial \theta_j} \\
&\quad + \sum_{i=1}^{M} \frac{u_i p_i(\theta)}{\sum_{i=1}^{M} u_i p_i} \phi'_a \left(\frac{p_i}{p_i(\theta)} \right) \frac{2 p_i p_i(\theta)}{p_i(\theta)^4} \frac{\partial p_i(\theta)}{\partial \theta_r} \frac{\partial p_i(\theta)}{\partial \theta_j} \\
&\quad \left. - \sum_{i=1}^{M} \frac{u_i p_i(\theta)}{\sum_{i=1}^{M} u_i p_i} \phi'_a \left(\frac{p_i}{p_i(\theta)} \right) \frac{p_i}{p_i(\theta)^2} \frac{\partial^2 p_i(\theta)}{\partial \theta_r \partial \theta_j} \right],
\end{aligned}$$

se tiene,

$$\begin{aligned}
\frac{\partial F}{\partial \theta_0} &= \left(\frac{\partial F}{\partial \theta}\right)_{\left(\pi_1,\ldots,\pi_M;\theta_{01},\ldots,\theta_{0M_0}\right)} \\
&= \sum_{a=1}^{\Lambda} \eta_a h_a'(0)\,\phi_a''(1)\left[\frac{1}{\sum\limits_{i=1}^{M} u_i p_i(\theta_0)} \sum_{i=1}^{M} \frac{u_i}{p_i(\theta_0)} \frac{\partial p_i(\theta_0)}{\partial \theta_r} \frac{\partial p_i(\theta_0)}{\partial \theta_j}\right]_{\substack{r=1,\ldots,M_0 \\ j=1,\ldots,M_0}} \\
&= \frac{\sum\limits_{a=1}^{\Lambda} \eta_a h_a'(0)\,\phi_a''(1)}{E_{p(\theta_0)}(u)} A^*(\theta_0)^t A^*(\theta_0).
\end{aligned}$$

Donde

$$A^*(\theta_0)^t = \left(\left(\frac{\partial p_i(\theta_0)}{\partial \theta_r}\right)_{\substack{i=1,\ldots,M \\ r=1,\ldots,M_0}}\right)^t diag\left(\mathcal{U}_*^{1/2}\right)_{M\times M}.$$

Se sigue que $rango\left(A^*(\theta_0)^t\right) = M_0$ ya que si C es una matriz $u \times v$ y B es una matriz no singular de orden v, entonces $rango\,(CB) = rango\,(C)$. Tomando

$$C = \left(\left(\frac{\partial p_i(\theta_0)}{\partial \theta_r}\right)_{\substack{i=1,\ldots,M \\ r=1,\ldots,M_0}}\right)^t$$

y

$$B = diag\left(\mathcal{U}_*^{1/2}\right)_{M\times M}$$

se tiene que $A^*(\theta_0)^t$ tiene rango M_0. Por otra parte si dado $B_{M\times N}$, $(M \leq N)$ y $rango\,(B) = M$, se tiene que $rango\left(B^t B\right) = M$. En consecuencia como rango de $A^*(\theta_0)^t = M_0$ se tiene que

$$rango\left(A^*(\theta_0)^t A^*(\theta_0)\right) = M_0.$$

En definitiva la matriz

$$\left(\frac{\partial F}{\partial \theta}\right) = \left(\frac{\partial F_i}{\partial \theta_j}\right)_{\substack{i=1,\ldots,M \\ j=1,\ldots,M_0}}$$

es no singular en $(\pi_1, ..., \pi_M; \theta_{01}, ..., \theta_{0M_0})$.

Aplicando El Teorema de la Función Implícita existe un entorno M-dimensional U_0 de $\pi = (\pi_1, ..., \pi_M)^t$ en R^M y una única función diferenciable con continuidad

$\tilde{\theta}: U_0 \to R^{M_0}$ de tal forma que $F\left(p, \tilde{\theta}(p)\right) = 0\ \forall p \in U_0$ y $\tilde{\theta}(\pi) = \theta_0$. Aplicando la regla de la cadena se tiene

$$\frac{\partial F\left(p, \tilde{\theta}(p)\right)}{\partial p} + \frac{\partial F\left(p, \tilde{\theta}(p)\right)}{\partial \tilde{\theta}(p)} \frac{\partial \tilde{\theta}(p)}{\partial p} = 0$$

y para $p = \pi$,

$$\frac{\partial F}{\partial \pi} + \frac{\partial F}{\partial \theta_0} \frac{\partial \theta_0}{\partial \pi} = 0.$$

Ahora

$$\begin{aligned}
\frac{\partial F}{\partial p_i} &= \frac{\partial}{\partial p_i}\left(\frac{\partial WD_{\phi}^{h}(p, p(\theta))}{\partial \theta_j}\right) \\
&= \sum_{a=1}^{\Lambda} \eta_a h_a'' \left(\sum_{i=1}^{M} \frac{u_i p_i(\theta)}{\sum\limits_{i=1}^{M} u_i p_i} \phi_a\left(\frac{p_i}{p_i(\theta)}\right)\right) \\
&\quad \times \frac{1}{\sum\limits_{i=1}^{M} u_i p_i} \left(\sum_{i=1}^{M} u_i \frac{\partial p_i(\theta)}{\partial \theta_j} \phi_a\left(\frac{p_i}{p_i(\theta)}\right) - \sum_{i=1}^{M} u_i p_i(\theta)\, \phi_a'\left(\frac{p_i}{p_i(\theta)}\right) \frac{p_i}{p_i(\theta)^2} \frac{\partial p_i(\theta)}{\partial \theta_j}\right) \\
&\quad \times \frac{1}{\left(\sum\limits_{i=1}^{M} u_i p_i\right)^2} \left(u_i p_i(\theta)\, \phi_a'\left(\frac{p_i}{p_i(\theta)}\right) \frac{1}{p_i(\theta)} \sum_{i=1}^{M} u_i p_i - \sum_{i=1}^{M} u_i p_i(\theta)\, \phi_a\left(\frac{p_i}{p_i(\theta)}\right) u_i\right) \\
&\quad + \sum_{a=1}^{\Lambda} \eta_a h_a' \left(\sum_{i=1}^{M} \frac{u_i p_i(\theta)}{\sum\limits_{i=1}^{M} u_i p_i} \phi_a\left(\frac{p_i}{p_i(\theta)}\right)\right) \frac{1}{\left(\sum\limits_{i=1}^{M} u_i p_i\right)^2} \\
&\quad \times \left(u_i \frac{\partial p_i(\theta)}{\partial \theta_j} \phi_a'\left(\frac{p_i}{p_i(\theta)}\right) \frac{1}{p_i(\theta)} \sum_{i=1}^{M} u_i p_i\right. \\
&\quad - u_i p_i(\theta)\, \phi_a''\left(\frac{p_i}{p_i(\theta)}\right) \frac{1}{p_i(\theta)} \frac{p_i}{p_i(\theta)^2} \frac{\partial p_i(\theta)}{\partial \theta_j} \sum_{i=1}^{M} u_i p_i \\
&\quad - u_i p_i(\theta)\, \phi_a'\left(\frac{p_i}{p_i(\theta)}\right) \frac{1}{p_i(\theta)^2} \frac{\partial p_i(\theta)}{\partial \theta_j} \sum_{i=1}^{M} u_i p_i
\end{aligned}$$

$$- u_i \left(\sum_{i=1}^{M} u_i \frac{\partial p_i(\theta)}{\partial \theta_j} \phi_a \left(\frac{p_i}{p_i(\theta)} \right) - \sum_{i=1}^{M} u_i p_i(\theta) \phi_a' \left(\frac{p_i}{p_i(\theta)} \right) \frac{p_i}{p_i(\theta)^2} \frac{\partial p_i(\theta)}{\partial \theta_j} \right) \Bigg) ,$$

y para $(\pi_1, ..., \pi_M; \theta_{01}, ..., \theta_{0M})$ se tiene

$$\frac{\partial}{\partial p_i} \left(\frac{\partial WD_\phi^h(p, p(\theta_0))}{\partial \theta_j} \right) = -\sum_{a=1}^{\Lambda} \eta_a h_a'(0) \phi_a''(1) \frac{\partial p_i(\theta_0)}{\partial \theta_j} \frac{u_i}{p_i(\theta_0)} \frac{1}{\sum_{i=1}^{M} u_i p_i(\theta_0)} .$$

Entonces, hemos llegado que por una parte

$$\frac{\partial F}{\partial \theta_0} = \frac{\sum_{a=1}^{\Lambda} \eta_a h_a'(0) \phi_a''(1)}{E_{p(\theta_0)}(u)} A^*(\theta_0)^t A^*(\theta_0)$$

y por otra

$$\frac{\partial F}{\partial \pi} = -\frac{\sum_{a=1}^{\Lambda} \eta_a h_a'(0) \phi_a''(1)}{E_{p(\theta_0)}(u)} A^*(\theta_0)^t \, diag\left(\mathcal{U}_*^{1/2}\right)$$

con lo cual

$$\frac{\partial \theta_0}{\partial \pi} = \left(A^*(\theta_0)^t A^*(\theta_0)\right)^{-1} A^*(\theta_0)^t \, diag\left(\mathcal{U}_*^{1/2}\right).$$

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

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

y como $\tilde{\theta}(\pi) = \theta_0$, se llega a

$$\tilde{\theta}(p) = \theta_0 + \left(A^*(\theta_0)^t A^*(\theta_0)\right)^{-1} A^*(\theta_0)^t \, diag\left(\mathcal{U}_*^{1/2}\right)(p - \pi) + o(\|p - \pi\|) .$$

Ahora bien $\hat{p} \underset{n \to \infty}{\overset{c.s.}{\to}} \pi$, por lo tanto $\hat{p} \in U_0$ y como consecuencia $\tilde{\theta}(p)$ es solución única de las ecuaciones

$$\frac{\partial WD_\phi^h\left(\hat{p}, \tilde{\theta}(\hat{p})\right)}{\partial \theta_j} = 0, \quad j = 1, ..., M_0$$

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

$$\begin{aligned}\hat{\theta}_{h,\phi,w} &= \theta_0 + \left(A^*(\theta_0)^t A^*(\theta_0)\right)^{-1} A^*(\theta_0)^t \, diag\left(\mathcal{U}_*^{1/2}\right)(\hat{p}-\pi) \\ &\quad + o\left(\|\hat{p}-p(\theta_0)\|\right).\end{aligned}$$

■

Teorema 2.3

Bajo las condiciones del teorema anterior se tiene

$$\sqrt{n}\left(\hat{\theta}_{h,\phi,w}-\theta_0\right) \underset{n\to\infty}{\overset{L}{\to}} N\left(0,\Sigma^*\right)$$

con

$$\Sigma^* = B^*(\theta_0)\, diag\left(\mathcal{U}_*^{1/2}\right)\Sigma_{p(\theta_0)} diag\left(\mathcal{U}_*^{1/2}\right) B^*(\theta_0)^t$$

donde la matriz $B^*(\theta_0)$ está definida en el teorema anterior.

Demostración

Aplicando el Teorema Central del Límite se tiene que

$$\sqrt{n}\left(\hat{p}-p(\theta_0)\right) \underset{n\to\infty}{\overset{L}{\to}} N\left(0,\Sigma_{p(\theta_0)}\right)$$

siendo

$$\Sigma_{p(\theta_0)} = diag\left(p(\theta_0)\right) - p(\theta_0)\, p(\theta_0)^t.$$

Por el teorema anterior

$$\begin{aligned}\hat{\theta}_{h,\phi,w}-\theta_0 &= \left(A^*(\theta_0)^t A^*(\theta_0)\right)^{-1} A^*(\theta_0)^t \, diag\left(\mathcal{U}_*^{1/2}\right)(\hat{p}-\pi) \\ &\quad + o\left(\|\hat{p}-p(\theta_0)\|\right) \\ &= B^*(\theta_0)\, diag\left(\mathcal{U}_*^{1/2}\right)(\hat{p}-\pi) + o\left(\|\hat{p}-p(\theta_0)\|\right)\end{aligned}$$

con

$$B^*(\theta_0) = \left(A^*(\theta_0)^t A^*(\theta_0)\right)^{-1} A^*(\theta_0)^t.$$

Como consecuencia

$$\sqrt{n}\left(\hat{\theta}_{h,\phi,w}-\theta_0\right) \underset{n\to\infty}{\overset{L}{\to}} N\left(0,\Sigma^*\right)$$

con

$$\Sigma^* = B^*(\theta_0)\, diag\left(\mathcal{U}_*^{1/2}\right)\ \Sigma_{p(\theta_0)} diag\left(\mathcal{U}_*^{1/2}\right) B^*(\theta_0)^t$$

■

Ejemplo 2.2

Continuando con el Ejemplo 2.1, ahora se obtendrá la distribución asintótica del estimador de mínima WD_ϕ^h-divergencia.

En primer lugar se calcularán las matrices $\Sigma_{p(\theta_0)}$, $A^*(\theta_0)$ y $B^*(\theta_0)\, diag\left(\mathcal{U}_*^{1/2}\right)$.

Se tiene:

- $$\Sigma_{p(\theta_0)} = diag\left(p(\theta_0)\right) - p(\theta_0)\, p(\theta_0)^t$$

$$= \begin{pmatrix} \frac{1}{3}-\theta & 0 & 0 \\ 0 & \frac{2}{3}-\theta & 0 \\ 0 & 0 & 2\theta \end{pmatrix} - \begin{pmatrix} \frac{1}{3}-\theta \\ \frac{2}{3}-\theta \\ 2\theta \end{pmatrix} \begin{pmatrix} \frac{1}{3}-\theta, & \frac{2}{3}-\theta, & 2\theta \end{pmatrix}$$

$$= \begin{pmatrix} \left(\frac{1}{3}-\theta\right)\left(\frac{2}{3}+\theta\right) & -\left(\frac{1}{3}-\theta\right)\left(\frac{2}{3}-\theta\right) & -\left(\frac{1}{3}-\theta\right)2\theta \\ -\left(\frac{2}{3}-\theta\right)\left(\frac{1}{3}-\theta\right) & \left(\frac{2}{3}-\theta\right)\left(\frac{1}{3}+\theta\right) & -\left(\frac{2}{3}-\theta\right)2\theta \\ -\left(\frac{1}{3}-\theta\right)2\theta & -\left(\frac{2}{3}-\theta\right)2\theta & (1-2\theta)\,2\theta \end{pmatrix}.$$

- $$A^*(\theta_0) = diag\left(\mathcal{U}_*^{1/2}\right)_{M\times M} \left(\left(\frac{\partial p_i(\theta_0)}{\partial \theta_r}\right)_{\substack{i=1,\dots,M \\ r=1,\dots,M_0}}\right)$$

$$= \begin{pmatrix} \left(\dfrac{u_1}{\frac{1}{3}-\theta}\right)^{1/2} & 0 & 0 \\ 0 & \left(\dfrac{u_2}{\frac{2}{3}-\theta}\right)^{1/2} & 0 \\ 0 & 0 & \left(\dfrac{u_3}{2\theta}\right)^{1/2} \end{pmatrix} \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}$$

$$= \begin{pmatrix} -\left(\dfrac{u_1}{\frac{1}{3}-\theta}\right)^{1/2} \\ \\ -\left(\dfrac{u_2}{\frac{2}{3}-\theta}\right)^{1/2} \\ \\ 2\left(\dfrac{u_3}{2\theta}\right)^{1/2} \end{pmatrix}.$$

- $$B^{*}(\theta_0)\, diag\left(\mathcal{U}_{*}^{1/2}\right) = \left(A^{*}(\theta_0)^{t} A^{*}(\theta_0)\right)^{-1} A^{*}(\theta_0)^{t}\, diag\left(\mathcal{U}_{*}^{1/2}\right)$$

$$= c\left(-\left(\frac{u_1}{\frac{1}{3}-\theta}\right)^{1/2},\ -\left(\frac{u_2}{\frac{2}{3}-\theta}\right)^{1/2},\ 2\left(\frac{u_3}{2\theta}\right)^{1/2} \right)$$

$$\times \begin{pmatrix} \left(\dfrac{u_1}{\frac{1}{3}-\theta}\right)^{1/2} & 0 & 0 \\ 0 & \left(\dfrac{u_2}{\frac{2}{3}-\theta}\right)^{1/2} & 0 \\ 0 & 0 & \left(\dfrac{u_3}{2\theta}\right)^{1/2} \end{pmatrix}$$

$$= c\left(-\frac{u_1}{\frac{1}{3}-\theta},\ -\frac{u_2}{\frac{2}{3}-\theta},\ \frac{u_3}{\theta} \right),$$

donde

$$c = \left(A^*(\theta_0)^t A^*(\theta_0)\right)^{-1} = \left(\frac{u_1}{\frac{1}{3}-\theta} + \frac{u_2}{\frac{2}{3}-\theta} + \frac{2u_3}{\theta}\right)^{-1}.$$

Luego,

$$B^*(\theta_0)\, diag\left(\mathcal{U}_*^{1/2}\right) = \left(\frac{-u_1 c}{\left(\frac{1}{3}-\theta\right)}, \frac{-u_2 c}{\left(\frac{2}{3}-\theta\right)}, \frac{u_3 c}{\theta}\right).$$

Finalmente,

$$\begin{aligned}
\Sigma^* &= B^*(\theta_0)\, diag\left(\mathcal{U}_*^{1/2}\right) \Sigma_{p(\theta_0)} diag\left(\mathcal{U}_*^{1/2}\right) B^*(\theta_0)^t \\
&= -\frac{(-1+3\theta)(-2+3\theta)^2\theta^2 u_1(3u_1\theta - 6u_3\theta + 3u_2\theta + 2u_3 + 2u_1 - u_2)}{(-6u_1\theta + 9u_1\theta^2 - 3u_2\theta + 9u_2\theta^2 - 4u_3 + 18u_3\theta - 18u_3\theta^2)^2} \\
&- \frac{(-1+3\theta)^2(-2+3\theta)\theta^2 u_2(3u_1\theta - 6u_3\theta + 3u_2\theta + 4u_3 - 2u_1 + u_2)}{(-6u_1\theta + 9u_1\theta^2 - 3u_2\theta + 9u_2\theta^2 - 4u_3 + 18u_3\theta - 18u_3\theta^2)^2} \\
&+ 2\frac{(-1+3\theta)^2(-2+3\theta)^2\theta u_3(u_1\theta - 2u_3\theta + u_2\theta + u_3)}{(-6u_1\theta + 9u_1\theta^2 - 3u_2\theta + 9u_2\theta^2 - 4u_3 + 18u_3\theta - 18u_3\theta^2)^2}.
\end{aligned} \tag{2.8}$$

Es decir,

$$\sqrt{n}\left(\hat{\theta}_{h,\phi,w} - \theta_0\right) \underset{n\to\infty}{\overset{L}{\longrightarrow}} N(0, \Sigma^*)$$

con Σ^* *la calculada anteriormente. Teniendo en cuenta la expresión de* $\hat{\theta}_{h,\phi,w}$*, se tiene*

$$\sqrt{n}\left(\frac{\frac{u_1}{3n_1} + \frac{2u_2}{3n_2} + \frac{1}{n}(2u_3 - u_1 - u_2)}{\frac{u_1}{n_1} + \frac{u_2}{n_2} + \frac{4u_3}{n_3}} - \theta_0\right)$$

converge en ley a una normal de media cero y varianza dada en (2.8).

Si se toma $u_1 = u_2 = u_3$*, se tiene entonces que*

$$\begin{aligned}
\Sigma^* &= B^*(\theta_0)\, diag\left(\mathcal{U}_*^{1/2}\right) \Sigma_{p(\theta_0)} diag\left(\mathcal{U}_*^{1/2}\right) B^*(\theta_0)^t \\
&= \frac{(-1+3\theta)(-2+3\theta)(4-9\theta)}{81}.
\end{aligned}$$

Es decir,

$$\sqrt{n}\left(\frac{\dfrac{1}{3n_1}+\dfrac{2}{3n_2}}{\dfrac{1}{n_1}+\dfrac{1}{n_2}+\dfrac{4}{n_3}}-\theta_0\right)$$

converge en ley a una normal con media cero y varianza

$$\frac{(-1+3\theta)(-2+3\theta)(4-9\theta)}{81}.$$

Teorema 2.4

Bajo las condiciones del teorema anterior se tiene,

$$\sqrt{n}\left(p\left(\hat{\theta}_{h,\phi,w}\right)-p(\theta_0)\right)\xrightarrow{L}N\left(0,\Sigma^{**}\right)$$

siendo

$$\Sigma^{**}=J(\theta_0)\,C(\theta_0)\,\Sigma_{p(\theta_0)}C(\theta_0)^t\,J(\theta_0)^t\,,$$

$$C(\theta_0)=B^*(\theta_0)\,diag\left(\mathcal{U}_*^{1/2}\right)$$

y

$$J(\theta_0)=\left(\left(\frac{\partial p_i(\theta_0)}{\partial\theta_r}\right)_{\substack{i=1,\ldots,M\\ r=1,\ldots,M_0}}\right).$$

Demostración

Por el teorema anterior se tiene,

$$\sqrt{n}\left(\hat{\theta}_{h,\phi,w}-\theta_0\right)\underset{n\to\infty}{\xrightarrow{L}}N\left(0,\Sigma^*\right)$$

con

$$\Sigma^*=C(\theta_0)\,\Sigma_{p(\theta_0)}C(\theta_0)^t\,.$$

Desarrollando en serie de Taylor $p\left(\hat{\theta}\right)$ en un entorno de θ_0, se tiene

$$p\left(\hat{\theta}_{h,\phi,w}\right)=p(\theta_0)+\left(\frac{\partial p_i(\theta_0)}{\partial\theta_r}\right)_{\substack{i=1,\ldots,M\\ r=1,\ldots,M_0}}\left(\hat{\theta}_{h,\phi,w}-\theta_0\right)+o\left(\left\|\hat{\theta}_{h,\phi,w}-\theta_0\right\|\right).$$

Con lo cual

$$\sqrt{n}\left(p\left(\hat{\theta}_{h,\phi,w}\right)-p(\theta_0)\right)\xrightarrow{L}N\left(0,\Sigma^{**}\right)$$

siendo

$$\begin{aligned}\Sigma^{**} &= J(\theta_0)\, C(\theta_0)\, \Sigma_{p(\theta_0)} C(\theta_0)^t\, J(\theta_0)^t \\ &= J(\theta_0)\, \Sigma^* J(\theta_0)^t\,.\end{aligned}$$

■

En los teoremas anteriores se ha supuesto que la distribución que rige el modelo discretizado es $p(\theta) = (p_1(\theta), ..., p_M(\theta))^t$. En el siguiente teorema se consideran desviaciones del modelo dadas por la familia

$$p_\varepsilon(\theta) = (1-\varepsilon)\, p(\theta) + \varepsilon p$$

con $\varepsilon > 0$, $\theta \in \Theta$ y $p \in \Delta_M$. Los elementos del vector $p_\varepsilon(\theta)$ se denotarán mediante $p_i(\theta,\varepsilon)$. Es decir,

$$p_i(\theta,\varepsilon) = (1-\varepsilon)\, p_i(\theta) + \varepsilon p_i$$

$i = 1, ..., M$.

Denotemos por $\theta^\varepsilon_{h,\phi,w}(p)$ el vector que minimiza la función

$$g_\varepsilon(p,\theta) = \sum_{a=1}^{\Lambda} \eta_a h_a \left(\sum_{i=1}^{M} \frac{u_i}{E_p(u)} p_i(\theta,\varepsilon)\, \phi_a\left(\frac{p_i}{p_i(\theta,\varepsilon)}\right)\right).$$

Para garantizar la robustez de $\theta^\varepsilon_{h,\phi,w}(p)$, lo que hay que comprobar es que a pequeñas desviaciones de $p(\theta)$ le corresponden pequeñas desviaciones de $\theta^\varepsilon_{h,\phi,w}(p)$; o bien que

$$\lim_{\varepsilon\to 0} \theta^\varepsilon_{h,\phi,w}(p) = \theta_{h,\phi,w}(p)\,.$$

Las condiciones que garantizan la robustez de la función de mínima WD^h_ϕ-divergencia vienen dadas en el siguiente teorema:

Teorema 2.5

Sea Θ compacto y las funciones h_a, ϕ_a, $a \in \Lambda$, verifican las condiciones requeridas en (1.15). *Suponiendo que se verifican las condiciones 1-6 de Birch y $\hat{\theta}_{h,\phi,w}$ es único en un entorno cerrado de $\pi = p(\theta_0)$, se tiene*

$$\lim_{\varepsilon\to 0} \theta^\varepsilon_{h,\phi,w}(p) = \theta_{h,\phi,w}(p)\,.$$

Demostración

Sea $\{\varepsilon_n\}$ una sucesión arbitraria de números positivos verificando $\varepsilon_n \underset{n\to\infty}{\to} 0$. Por ser h_a, ϕ_a, $a \in \Lambda$, continuas y

$$p_i(\theta, \varepsilon_n) \underset{\varepsilon_n \to 0}{\to} p_i(\theta) \qquad i = 1, ..., M,$$

se tiene que

$$g_{\varepsilon_n}(p, \theta) \underset{\varepsilon_n \to 0}{\to} g_0(p, \theta) \qquad \forall \theta \in \Theta.$$

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

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

lo que implica que

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

o lo que es lo mismo

$$\lim_{\varepsilon_n \to 0} \left| g_{\varepsilon_n}\left(p, \theta_{h,\phi,w}^{\varepsilon_n}\right) - g_0\left(p, \hat{\theta}_{h,\phi,w}\right) \right| = 0.$$

En definitiva se ha demostrado que

$$\lim_{\varepsilon_n \to 0} g_{\varepsilon_n}\left(p, \theta_{h,\phi,w}^{\varepsilon_n}\right) = g_0\left(p, \hat{\theta}_{h,\phi,w}\right).$$

Si

$$\lim_{\varepsilon_n \to 0} \theta_{h,\phi,w}^{\varepsilon_n}(p) \neq \hat{\theta}_{h,\phi,w}(p),$$

resulta que por ser Θ compacto existe una subsucesión

$$\left\{\theta_{h,\phi,w}^{\delta_n}(p)\right\} \subset \left\{\theta_{h,\phi,w}^{\varepsilon_n}(p)\right\}$$

verificando

$$\lim_{\delta_n \to 0} \theta_{h,\phi,w}^{\delta_n}(p) = \theta^* \neq \hat{\theta}_{h,\phi,w}(p).$$

De,

$$\lim_{\varepsilon_n \to 0} g_{\varepsilon_n}\left(p, \theta_{h,\phi,w}^{\varepsilon_n}\right) = g_0\left(p, \hat{\theta}_{h,\phi,w}\right)$$

se sigue que

$$g_0(p, \theta^*) = g_0\left(p, \hat{\theta}_{h,\phi,w}\right)$$

con $\theta^* \neq \hat{\theta}_{h,\phi,w}(p)$ lo cual contradice la unicidad de $\hat{\theta}_{h,\phi,w}(p)$.

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, $\pi = p(\theta_0) \in \Delta_M$, cumple

$$\|\pi - p(\theta)\| < \varepsilon$$

para un $\theta \in \Theta$ y comprobar que si ε es pequeño, el valor $\theta_{h,\phi,w}(\pi)$ es próximo a $\theta_{h,\phi,w}(p(\theta)) = \theta$.

Teorema 2.6
Bajo las condiciones del teorema anterior supuesto que $\pi \in \Delta_M$, se tiene

$$\lim_{\|\pi - p(\theta)\| \to 0} \theta_{h,\phi,w}(\pi) = \theta_{h,\phi,w}(p(\theta)) = \theta.$$

Demostración

Inmediata por ser $\theta_{h,\phi,w}$ continua. ∎

2.4. El estimador de mínima WD$_\phi^h$-divergencia con restricciones

Un nuevo problema de estimación se plantea si se tienen ν $(\nu < M_0)$ funciones $f_1(\theta), f_2(\theta) ..., f_\nu(\theta)$, definidas de Θ en R, que restringen el parámetro $\theta \in \Theta \subset R^{M_0}$, $f_m(\theta) = 0$, $m = 1, .., \nu$. El problema de estimación que aparece recibe el nombre de estimación con restricciones. En caso de poblaciones generales Atchison y Silvey (1958) definieron y estudiaron por primera vez el problema del estimador de máxima verosimilitud con restricciones, mientras que Diamond, Milra y Roy (1960) lo hicieron en caso de poblaciones multinomiales. En Pardo, J.A. y otros (2001) se introdujo el estimador de mínima ϕ-divergencia bajo restricciones y se estudiaron sus propiedades. El estimador de mínima ϕ-divergencia con restricciones de θ es el valor $\hat{\theta}_{\phi}^{(r)} \in \Theta$ que satisface la condición

$$D_{\phi}\left(\hat{p}, p\left(\hat{\theta}_{\phi}^{(r)}\right)\right) = \inf_{\left\{\theta \in \Theta \subset R^{M_0}: f_m(\theta)=0, m=1,..,\nu\right\}} D_{\phi}\left(\hat{p}, p\left(\theta\right)\right).$$

En este apartado se introduce y se estudia el problema de estimación con restricciones, supuesto que los elementos A_j, $j = 1, ..., M$, asociados a la partición $\mathcal{A}$ están ponderados, basándonos en las (h, ϕ)-divergencias ponderadas. A lo largo de esta Sección se supondrán las dos condiciones siguientes:

1. Cada función $f_m(\theta)$ tiene derivadas parciales segundas continuas,

2. La matriz
$$D(\theta) = \left(\frac{\partial f_m(\theta)}{\partial \theta_k}\right)_{\substack{m=1,...,\nu \\ k=1,...,M_0}}$$
tiene rango ν.

Supuestas estas dos condiciones. El estimador de mínima (h, ϕ)-divergencia ponderada con restricciones de θ, es cualquier $\hat{\theta}_{h,\phi,w}^{(r)} \in \Theta$ que verifique

$$WD_{\phi}^{h}\left(\hat{p}, p\left(\hat{\theta}_{h,\phi,w}^{(r)}\right)\right) = \inf_{\left\{\theta \in \Theta \subset R^{M_0}: f_m(\theta)=0, m=1,..,\nu\right\}} WD_{\phi}^{h}\left(\hat{p}, p\left(\theta\right)\right).$$

En lo sucesivo el estimador de mínima (h, ϕ)-divergencia ponderada con restricciones se denominará estimador de mínima WD_{ϕ}^{h}-divergencia con restricciones y se expresará mediante

$$\hat{\theta}_{h,\phi,w}^{(r)} = \arg \inf_{\theta \in \Theta} WD_{\phi}^{h}\left(\hat{p}, p\left(\theta\right)\right).$$

De la misma forma que en la Sección 2.3 se estudiaron las propiedades asintóticas del estimador de mínima WD_{ϕ}^{h}-divergencia se pueden estudiar para el estimador de mínima WD_{ϕ}^{h}-divergencia con restricciones. Dada la similitud de las demostraciones únicamente se señalarán los resultados que se obtienen, pero sin dar su demostración.

Bajo las hipótesis reseñadas en el Teorema 2.2 y supuesto que se verifiquen las condiciones 1 y 2 señaladas anteriormente, se tiene

$$\hat{\theta}_{h,\phi,w}^{(r)} = \theta_0 + H^*(\theta_0) B^*(\theta_0) \, diag\left(\mathcal{U}_*^{1/2}\right)(\hat{p} - p(\theta_0)) + o\left(\|\hat{p} - p(\theta_0)\|\right) \tag{2.9}$$

donde $\hat{\theta}^{(r)}_{h,\phi,w}$ es único en un entorno de θ_0, $B^*(\theta_0)$ y $\mathcal{U}_*$ están definidas en el Teorema 2.2 y la matriz $H^*(\theta_0)$ está definida mediante,

$$\begin{aligned} H^*(\theta_0) &= I_{M_0\times M0} - \left(A^*(\theta_0)^t A^*(\theta_0)\right)^{-1} D(\theta_0)^t \\ &\quad \times \left(D(\theta_0)\left(A^*(\theta_0)^t A^*(\theta_0)\right)^{-1} D(\theta_0)^t\right)^{-1} D(\theta_0). \end{aligned}$$

Es interesante observar que si $u_1 = u_2 = ... = u_M = u$, se tiene

$$\begin{aligned} \hat{\theta}^{(r)}_{h,\phi} &= \theta_0 + H(\theta_0) I_F(\theta_0) A(\theta_0)^t diag\left(p(\theta_0)^{1/2}\right)(\hat{p} - p(\theta_0)) \\ &\quad + o(\|\hat{p} - p(\theta_0)\|) \end{aligned}$$

donde $\hat{\theta}^{(r)}_{h,\phi}$ es único en un entorno de θ_0 y las matrices $I_F(\theta_0)^{-1}$ y $H(\theta_0)$ se definen mediante,

$$\begin{aligned} I_F(\theta_0)^{-1} &= A(\theta_0)^t A(\theta_0) \\ A(\theta_0) &= diag\left(p(\theta_0)^{1/2}\right) J(\theta_0) \end{aligned}$$

y

$$H(\theta_0) = I_{M_0\times M0} - I_F(\theta_0)^{-1} D(\theta_0)^t \left(D(\theta_0) I_F(\theta_0)^{-1} D(\theta_0)^t\right)^{-1} D(\theta_0).$$

La matriz $J(\theta_0)$ está definida en (2.3). Es decir, se llega al resultado obtenido en Pardo, J.A. (2001).

A partir de la descomposición dada en (2.9) se tiene por un lado que

$$\sqrt{n}\left(\hat{\theta}^{(r)}_{h,\phi,w} - \theta_0\right) \underset{n\to\infty}{\overset{L}{\to}} N(0, \Sigma_1^*)$$

con

$$\Sigma_1^* = H^*(\theta_0) B^*(\theta_0) diag\left(\mathcal{U}_*^{1/2}\right) \Sigma_{p(\theta_0)} diag\left(\mathcal{U}_*^{1/2}\right) B^*(\theta_0)^t H^*(\theta_0)^t,$$

y por otro

$$\sqrt{n}\left(p\left(\hat{\theta}^{(r)}_{h,\phi,w}\right) - p(\theta_0)\right) \underset{n\to\infty}{\overset{L}{\to}} N(0, \Sigma_2^*)$$

con

$$\Sigma_2^* = J(\theta_0) \Sigma_1^* J(\theta_0)^t.$$

3

Distribución asintótica de las (h, φ)-divergencias ponderadas: Aplicaciones estadísticas

3.1. Introducción

En este capítulo se estudiará el problema de bondad de ajuste (hipótesis nula simple y compuesta), en el supuesto de que se tengan ponderaciones sobre los elementos A_i, $i = 1, ..., M$, de la partición $\mathcal{A}$ de $\mathcal{X}$, haciendo uso del estimador analógico de las (h, ϕ)-divergencias ponderadas, $WD_{\varphi}^{h}(\hat{p}, p_0)$, en el caso de hipótesis nula simple y de $WD_{\varphi}^{h}\left(\hat{p}, p\left(\hat{\theta}_{h,\phi,w}\right)\right)$ en el caso de hipótesis nula compuesta.

Si bien los antecedentes de la utilización de las medidas de divergencia en el análisis de datos categorizables hay que situarlos en Pearson (1900), ya que el estadístico Ji-cuadrado no es más que una (h, ϕ)-divergencia ponderada con $h(x) = x$, $\phi(x) = \frac{1}{2}(1-x)^2$ y $u_i = u \; \forall i = 1, ..., M$, es en los estudios de Cressie y Read (1984), Zografos (1993), Zografos y otros (1990), Pardo, L y otros (1993a), Morales y otros (1994a,b), Pardo, M.C. (1994a,b), etc... en donde de una forma explícita se hace referencia a la importancia de las medidas de divergencia en el análisis de datos categorizables.

Desde un punto de vista intuitivo si se desea contrastar $H_0 : p = p_0$ frente a $H_1 : p \neq p_0$ supuestas ponderaciones sobre los elementos de la partición realizada sobre $\mathcal{X}$ (o sobre las clases), parece lógico rechazar la hipótesis nula si $WD_{\phi}^{h}(\hat{p}, p_0)$ es superior a una constante c (p_0 conocida).

La cuestión radica en encontrar c de tal forma que el test resultante tenga un nivel de significación prefijado α. La obtención de c se llevará a cabo mediante la obtención de la distribución, bajo la hipótesis nula, de la familia de estadísticos $WD_{\phi}^{h}(\hat{p}, p_0)$. La distribución exacta resulta imposible, en la mayoría de los casos, de obtener por lo que se hará uso de la distribución asintótica. Así, el contraste se realizará en base a la distribución asintótica y ésta se obtendrá, por separado, en primer lugar para hipótesis nula simple (p_0 conocida) y en segundo lugar para hipótesis nula compuesta ($p = p(\theta)$, θ desconocido).

En el caso de que $\theta \in \Theta \subset R^{M_0}$, sea desconocido, será necesario recurrir previamente a la estimación de θ. A lo largo del capítulo se procederá a estimar θ mediante el estimador de mínima WD$_{\phi}^{h}$-divergencia introducido y estudiadas sus propiedades en el capítulo anterior.

El esquema que se seguirá en este capítulo será el siguiente: En primer lugar será necesario calcular la distribución de $WD_{\varphi}^{h}(\hat{p}, p_0)$ con p_0 conocida para posteriormente obtener la distribución de $WD_{\varphi}^{h}\left(\hat{p}, p\left(\hat{\theta}_{h,\phi,w}\right)\right)$. Obsérvese que para evitar problemas de notación, en lo que sigue φ se utilizará para definir la correspondiente familia de estadísticos asociada a las (h, ϕ)-divergencias ponderadas y ϕ se dejará para definir el correspondiente estimador de mínima (h, ϕ)-divergencia ponderada. Así, cuando se escriba $WD_{\varphi}^{h}\left(\hat{p}, p\left(\hat{\theta}_{h,\phi,w}\right)\right)$ se estará indicando que se utiliza la familia φ para construir la familia de estadísticos de contraste y la familia ϕ para construir la correspondiente familia de estimadores de mínima divergencia ponderada.

Antes de establecer la distribución asintótica del estadístico $WD_{\varphi}^{h}(\hat{p}, p_0)$ se dará una observación, cuya demostración puede verse en Dik y Gunst (1985) que será de gran utilidad a lo largo de la presente memoria.

Observación 3.1

Sea X una variable aleatoria normal d-dimensional con vector de medias 0 y matriz de varianzas covarianzas Σ. Sea A una matriz real simétrica de orden d, $r = rango(\Sigma A \Sigma)$, $r \geq 1$ y $\lambda_1, ..., \lambda_r$ los autovalores no nulos de $A\Sigma$, entonces la distribución de la variable aleatoria $X^t A X$ coincide con la de la variable aleatoria $\sum_{i=1}^{r} \lambda_i Z_i^2$ donde $Z_1, ..., Z_r$ son variables aleatorias normales e independientes de media cero y varianza uno.

En el supuesto que se verifique

$$\Sigma A \Sigma A \Sigma = \Sigma A \Sigma$$

se tiene que

$$X^t A X \underset{n\to\infty}{\overset{L}{\to}} \mathcal{X}_s^2$$

siendo $s = traza\,(A\Sigma)$.

Un resultado más general se tiene en el supuesto de que la variable aleatoria d-dimensional X *tenga como vector de medias* μ, *en lugar de cero, ya que en este caso la forma cuadrática* $X^t A X$ *sigue una distribución Ji-cuadrado no centrada si y sólo si*

$$\begin{aligned} \Sigma A \Sigma A \Sigma &= \Sigma A \Sigma \\ \mu^t A \Sigma A \mu &= \mu^t A \mu \\ \Sigma A \Sigma A \Sigma \mu &= \Sigma A \mu \end{aligned}$$

siendo el número de grados de libertad $traza\,(A\Sigma)$ *y el parámetro de no centralidad* $\mu^t A \mu$.

3.2. Distribución asintótica de $WD_{\varphi}^{h}(\hat{p}, p_0)$

Seguidamente se obtendrá la distribución asintótica del estadístico $WD_{\varphi}^{h}(\hat{p}, p_0)$ bajo la hipótesis de que las observaciones proceden de una población $p = (p_1, ..., p_M)^t$ con $p \neq p_0$. Posteriormente se obtendrá la distribución asintótica bajo la hipótesis de que $p = p_0$ y finalmente bajo hipótesis contiguas alternativas, es decir $p = p_0 + n^{-1/2} d$ donde $d = (d_1, ..., d_M)^t$ con $\sum_{i=1}^{M} d_i = 0$.

3.2.1. Distribución asintótica con $p \neq p_0$

Se considera el desarrollo en serie de Taylor de la función $WD_{\varphi}^{h}(\hat{p}, p_0)$ en el punto $\hat{p} = (\hat{p}_1, ..., \hat{p}_M)^t$ y en un entorno de $p = (p_1, ..., p_M)^t$. Se tiene entonces,

$$WD_{\varphi}^{h}(\hat{p}, p_0) = WD_{\varphi}^{h}(p, p_0) + \sum_{i=1}^{M} \left(\frac{\partial WD_{\varphi}^{h}(p, p_0)}{\partial p_i} \right)_{(p_1, ..., p_M)} (\hat{p}_i - p_i) + o\left(\|\hat{p} - p\|\right)$$

donde $\forall i=1,....,M$, la expresión de

$$\left(\frac{\partial WD_{\varphi}^{h}\left(p,p_{0}\right)}{\partial p_{i}}\right)_{p=(p_{1},...,p_{M})}$$

viene dada por

$$\sum_{a=1}^{\Lambda}\left\{\eta_{a}h_{a}^{\prime}\left(\frac{\sum\limits_{i=1}^{M}u_{i}p_{i0}}{E_{p}\left(u\right)}\varphi_{a}\left(\frac{p_{i}}{p_{i0}}\right)\right)\times\left[\frac{u_{i}}{\left(E_{p}\left(u\right)\right)^{2}}\left(\varphi_{a}^{\prime}\left(\frac{p_{i}}{p_{i0}}\right)\sum_{i=1}^{M}u_{i}p_{i}-\sum_{i=1}^{M}u_{i}p_{i0}\varphi_{a}\left(\frac{p_{i}}{p_{i0}}\right)\right)\right]\right\}.$$

Por el Teorema Central del Límite se tiene que

$$n^{\frac{1}{2}}\left(\hat{p}-p\right)\underset{n\rightarrow\infty}{\overset{L}{\rightarrow}}N\left(0,\Sigma_{p}\right)$$

con

$$\Sigma_{p}=diag\left(p\right)-pp^{t}$$

entonces

$$n^{1/2}\left(\hat{p}-p\right)=O_{p}\left(1\right)$$

o lo que es lo mismo

$$\begin{aligned}n^{1/2}o\left(\left\Vert \hat{p}-p\right\Vert \right) &= n^{1/2}o\left(O_{p}\left(n^{-1/2}\right)\right)\\ &= n^{1/2}o_{p}\left(n^{-1/2}\right)\\ &= o_{p}\left(1\right).\end{aligned}$$

Por tanto las variables aleatorias

$$n^{\frac{1}{2}}\left(WD_{\varphi}^{h}\left(\hat{p},p_{0}\right)-WD_{\varphi}^{h}\left(p,p_{0}\right)\right)$$

y

$$n^{\frac{1}{2}}\left(\sum_{i=1}^{M}t_{i}\left(\hat{p}_{i}-p_{i}\right)\right)=n^{\frac{1}{2}}T^{t}\left(\hat{p}-p\right)$$

tienen la misma distribución asintótica, siendo $T=\left(t_{1},....,t_{M}\right)^{t}$ y

$$t_{i}=\left(\frac{\partial WD_{\varphi}^{h}\left(p,p_{0}\right)}{\partial p_{i}}\right)_{p=(p_{1},...,p_{M})},\ i=1,....,M.$$

Ahora bien

$$n^{\frac{1}{2}} T^t (\hat{p} - p) \underset{n \to \infty}{\overset{L}{\to}} N\left(0, \sigma_p^2\right)$$

siendo

$$\begin{aligned} \sigma_p^2 &= T^t \Sigma_p T \\ \Sigma_p &= diag(p) - pp^t \\ T &= (t_1, ..., t_M)^t \end{aligned}$$

Con los resultados anteriores se puede establecer el siguiente teorema:

Teorema 3.1

Considérese el estimador analógico, $WD_{\varphi}^{h}(\hat{p}, p_0)$, obtenido de reemplazar los p_i por sus frecuencias relativas observadas $\hat{p}_i$, $i = 1, ..., M$, basadas en una muestra aleatoria simple de tamaño n. Si las funciones h_a y φ_a, $a = 1, ..., \Lambda$, verifican las condiciones requeridas en (1.15), *entonces*

$$n^{\frac{1}{2}} \left[WD_{\varphi}^{h}(\hat{p}, p_0) - WD_{\varphi}^{h}(p, p_0) \right] \underset{n \uparrow \infty}{\overset{L}{\to}} N\left(0, \sigma_p^2\right)$$

siempre y cuando $\sigma_p^2 > 0$, donde

$$\sigma_p^2 = T^t \Sigma_p T = \sum_{i=1}^{M} t_i^2 p_i - \left(\sum_{i=1}^{M} t_i p_i \right)^2$$

con

$$\Sigma_p = (p_i (\delta_{ij} - p_j))_{i,j=1,...,M} = diag(p) - pp^t$$

y

$$T = (t_1, ..., t_M)^t$$

siendo $\forall i = 1, ..., M$

$$t_i = \sum_{a=1}^{\Lambda} \left\{ \eta_a h_a' \left(\frac{\sum\limits_{i=1}^{M} u_i p_{i0}}{E_p(u)} \varphi_a \left(\frac{p_i}{p_{i0}} \right) \right) \left[\frac{u_i}{(E_p(u))^2} \left(\varphi_a' \left(\frac{p_i}{p_{i0}} \right) \sum_{i=1}^{M} u_i p_i - \sum_{i=1}^{M} u_i p_{i0} \varphi_a \left(\frac{p_i}{p_{i0}} \right) \right) \right] \right\}.$$

Corolario 3.1

En el caso de las $(h, \varphi)-$divergencias, es decir, si $u_i = u$ $\forall i = 1, ..., M$, se tiene

$$t_i = \sum_{a=1}^{\Lambda} \left\{ \eta_a h_a' \left(\sum_{i=1}^{M} p_{i0} \varphi_a \left(\frac{p_i}{p_{i0}} \right) \right) \times \left(\varphi_a' \left(\frac{p_i}{p_{i0}} \right) - \sum_{i=1}^{M} p_{i0} \varphi_a \left(\frac{p_i}{p_{i0}} \right) \right) \right\}$$

y

$$
\begin{aligned}
\sigma_p^2 &= \sum_{i=1}^{M}\left[\sum_{a=1}^{\Lambda}\left\{\eta_a h_a'\left(\sum_{i=1}^{M} p_{i0}\varphi_a\left(\frac{p_i}{p_{i0}}\right)\right)\varphi_a'\left(\frac{p_i}{p_{i0}}\right)\right\}\right]^2 p_i \\
&\quad -\left[\sum_{i=1}^{M}\left[\sum_{a=1}^{\Lambda}\left\{\eta_a h_a'\left(\sum_{i=1}^{M} p_{i0}\varphi_a\left(\frac{p_i}{p_{i0}}\right)\right)\varphi_a'\left(\frac{p_i}{p_{i0}}\right)\right\}\right] p_i\right]^2.
\end{aligned}
$$

Demostración

Para hallar t_i basta reemplazar en la fórmula general u_i por u para todo $i = 1, ..., M$.

Sean

$$A_a = \eta_a h_a'\left(\sum_{i=1}^{M} p_{i0}\varphi_a\left(\frac{p_i}{p_{i0}}\right)\right)$$

$$B_a^i = \varphi_a'\left(\frac{p_i}{p_{i0}}\right)$$

$$C_a = \sum_{i=1}^{M} p_{i0}\varphi_a\left(\frac{p_i}{p_{i0}}\right)$$

entonces, se tiene

$$t_i = \sum_{a=1}^{\Lambda} A_a\left(B_a^i - C_a\right).$$

Se sabe que

$$\sigma_p^2 = \sum_{i=1}^{M} t_i^2 p_i - \left(\sum_{i=1}^{M} t_i p_i\right)^2,$$

luego

$$
\begin{aligned}
\sigma_p^2 &= \sum_{i=1}^{M}\left(\sum_{a=1}^{\Lambda} A_a\left(B_a^i - C_a\right)\right)^2 p_i - \left(\sum_{i=1}^{M}\sum_{a=1}^{\Lambda} A_a\left(B_a^i - C_a\right) p_i\right)^2 \\
&= \sum_{i=1}^{M}\left[\left(\sum_{a=1}^{\Lambda} A_a B_a^i\right)^2 + \left(\sum_{a=1}^{\Lambda} A_a C_a\right)^2 - 2\left(\sum_{a=1}^{\Lambda} A_a B_a^i\right)\left(\sum_{a=1}^{\Lambda} A_a C_a\right)\right] p_i
\end{aligned}
$$

$$
\begin{aligned}
& -\left(\sum_{i=1}^{M}\left(\sum_{a=1}^{\Lambda}A_aB_a^i-\sum_{a=1}^{\Lambda}A_aC_a\right)p_i\right)^2 \\
= \; & \sum_{i=1}^{M}\left(\sum_{a=1}^{\Lambda}A_aB_a^i\right)^2 p_i+\left(\sum_{a=1}^{\Lambda}A_aC_a\right)^2-2\left(\sum_{a=1}^{\Lambda}A_aC_a\right)\sum_{i=1}^{M}\left(\sum_{a=1}^{\Lambda}A_aB_a^i\right)p_i \\
& -\left(\sum_{i=1}^{M}\left(\sum_{a=1}^{\Lambda}A_aB_a^i\right)p_i-\sum_{a=1}^{\Lambda}A_aC_a\sum_{i=1}^{M}p_i\right)^2,
\end{aligned}
$$

con lo cual

$$\sigma_p^2=\sum_{i=1}^{M}\left(\sum_{a=1}^{\Lambda}A_aB_a^i\right)^2 p_i-\left(\sum_{i=1}^{M}\sum_{a=1}^{\Lambda}A_aB_a^ip_i\right)^2.$$

Es decir,

$$
\begin{aligned}
\sigma_p^2 \; = \; & \sum_{i=1}^{M}\left[\sum_{a=1}^{\Lambda}\left\{\eta_a h_a'\left(\sum_{i=1}^{M}p_{i0}\varphi_a\left(\frac{p_i}{p_{i0}}\right)\right)\varphi_a'\left(\frac{p_i}{p_{i0}}\right)\right\}\right]^2 p_i \\
& -\left[\sum_{i=1}^{M}\left[\sum_{a=1}^{\Lambda}\left\{\eta_a h_a'\left(\sum_{i=1}^{M}p_{i0}\varphi_a\left(\frac{p_i}{p_{i0}}\right)\right)\varphi_a'\left(\frac{p_i}{p_{i0}}\right)\right\}\right]p_i\right]^2
\end{aligned}
$$

que es lo que se quería demostrar.

Corolario 3.2

En el caso de las φ-divergencias ponderadas, $WD_{\varphi}(\hat{p},p_0)$. Es decir, si $\Lambda=1$, $\eta_1=1$, $h_1(x)=x$ y $\varphi_1(x)=\varphi(x)$, se tiene

$$t_i=\frac{u_i}{\left(\sum\limits_{i=1}^{M}u_ip_i\right)^2}\left(\varphi'\left(\frac{p_i}{p_{i0}}\right)\sum_{i=1}^{M}u_ip_i-\sum_{i=1}^{M}u_ip_{i0}\varphi\left(\frac{p_i}{p_{i0}}\right)\right)$$

y

$$\sigma_p^2=\frac{1}{\left(\sum\limits_{i=1}^{M}u_ip_i\right)^2}\left[\sum_{i=1}^{M}\left(u_i\varphi'\left(\frac{p_i}{p_{i0}}\right)\right)^2p_i-\left(\sum_{i=1}^{M}u_i\varphi'\left(\frac{p_i}{p_{i0}}\right)p_i\right)^2\right].$$

Corolario 3.3

Si además $u_i = u$ $\forall i = 1, ..., M,$ es decir en el caso de las φ-divergencias, entonces

$$t_i = \varphi'\left(\frac{p_i}{p_{i0}}\right) - \sum_{i=1}^{M} p_{i0}\varphi\left(\frac{p_i}{p_{i0}}\right)$$

con lo cual

$$\sigma_p^2 = \sum_{i=1}^{M}\left(\varphi'\left(\frac{p_i}{p_{i0}}\right)\right)^2 p_i - \left(\sum_{i=1}^{M}\varphi'\left(\frac{p_i}{p_{i0}}\right)p_i\right)^2.$$

Este resultado se obtuvo por primera vez en Zografos y otros (1990).

En el caso de la divergencia de Kullback, $\varphi(x) = x\log x - x + 1$, se tiene

$$\sigma_p^2 = \sum_{i=1}^{M} p_i\left(\log\frac{p_i}{p_{i0}}\right)^2 - \left(\sum_{i=1}^{M} p_i\log\frac{p_i}{p_{i0}}\right)^2.$$

Si el primer término del desarrollo en serie de la función $WD_{\varphi}^{h}(\hat{p}, p_0)$ se anula, entonces hay que acudir al segundo término de dicho desarrollo.

A continuación se verá cuando ocurre ésto.

Proposición 3.1

Sea $S_n = \sqrt{n}\sum\limits_{i=1}^{M} t_i(\hat{p}_i - p_i)$. Entonces

$$S_n = 0\ \forall n \Leftrightarrow \sigma_p^2 = 0.$$

Demostración

Si se cumple que $S_n = 0\ \forall n$ entonces, $Va(S_n) = 0\ \forall n$, con lo cual $\lim\limits_{n\to\infty} Va(S_n) = \sigma_p^2 = 0$ (donde $Va(S_n)$ denota la varianza de S_n).

Sea ahora $\sigma_p^2 = 0$, es decir, $T^t\Sigma_p T = 0$. Al ser

$$T^t\Sigma_p T = \sigma_p^2 = \sum_{i=1}^{M} t_i^2 p_i - \left(\sum_{i=1}^{M} t_i p_i\right)^2$$

la varianza de una variable aleatoria que toma los valores $t_1,....,t_M$ con probabilidades $p_1,....,p_M$ se deberá verificar que $t_1 = ... = t_M$ con lo cual $S_n = \sqrt{n}\sum_{i=1}^{M} t_i(\hat{p}_i - p_i) = 0$. ■

Si $p = p_0$ se tiene que $t_i = 0\ \forall i = 1,....,M$, luego $t_1 = ... = t_M$ con lo cual $\sigma_p^2 = 0$.

3.2.2. Distribución asintótica con $p = p_0$

Antes de abordar el caso $p = p_0$ se analizará el caso más general que se obtiene cuando $\sigma_p^2 = 0$. Así, se supone inicialmente $\sigma_p^2 = 0$.

Se considera el segundo término del desarrollo en serie de Taylor de $WD_{\varphi}^{h}(\hat{p},p_0)$.

$$\begin{aligned} WD_{\varphi}^{h}(\hat{p},p_0) &= WD_{\varphi}^{h}(p,p_0) + \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M}\left(\frac{\partial^2 WD_{\varphi}^{h}(p,p_0)}{\partial p_i \partial p_j}\right)(\hat{p}_i - p_i)(\hat{p}_j - p_j) \\ &\quad + o\left(\|\hat{p}-p\|^2\right) \end{aligned}$$

entonces

$$\begin{aligned} \|\hat{p}-p\|^2 &= O_p\left(n^{-1}\right) \\ no\left(\|\hat{p}-p\|^2\right) &= no\left(O_p\left(n^{-1}\right)\right) \\ &= o_p(1). \end{aligned}$$

Por consiguiente se puede afirmar que las variables aleatorias

$$2n\left[WD_{\varphi}^{h}(\hat{p},p_0) - WD_{\varphi}^{h}(p,p_0)\right]$$

y

$$n(\hat{p}-p)^t A (\hat{p}-p)$$

tienen la misma distribución asintótica, siendo

$$A = \left(\frac{\partial^2 WD_{\varphi}^{h}(p,p_0)}{\partial p_i \partial p_j}\right)_{i,j=1,....,M}.$$

Por otra parte, como $n^{1/2}(\hat{p}-p) \xrightarrow{L} N(0,\Sigma_p)$ con $\Sigma_p = diag(p) - pp^t$, se tiene que

$$n(\hat{p}-p)^t A(\hat{p}-p) \underset{n\to\infty}{\overset{L}{\to}} \sum_{i=1}^{r} \beta_i Z_i^2$$

donde las Z_i son variables aleatorias normales independientes de media cero y varianza uno, los β_i son los autovalores no nulos de $A\Sigma_p$ y $r = rango(\Sigma_p A \Sigma_p)$.

Así, con estos resultados se puede escribir el siguiente teorema:

Teorema 3.2

Sea $WD_\varphi^h(p,p_0)$ y su estimador analógico $WD_\varphi^h(\hat{p},p_0)$ obtenido al reemplazar p_i por sus frecuencias relativas observadas $\hat{p}_i$, $i=1,....,M$, basadas en una muestra aleatoria simple de tamaño n y supóngase que se verifican las hipótesis requeridas a las funciones h_a y φ_a, $a=1,....,\Lambda$, en (1.15). Si $\sigma_p^2 = 0$, entonces

$$2n\left[WD_\varphi^h(\hat{p},p_0) - WD_\varphi^h(p,p_0)\right] \underset{n\to\infty}{\overset{L}{\to}} \sum_{i=1}^{r} \beta_i Z_i^2$$

donde las Z_i son variables aleatorias normales independientes de media cero y varianza uno, los β_i son los autovalores no nulos de $A\Sigma_p$ y $r = rango(\Sigma_p A \Sigma_p)$. Siendo

$$\Sigma_p = diag(p) - pp^t$$

y

$$A = \left(\frac{\partial^2 WD_\varphi^h(p,p_0)}{\partial p_i \partial p_j}\right)_{i,j=1,....,M} = (a_{ij})_{i,j=1,...,M}$$

con

$$\begin{aligned} a_{ij} \quad &= \quad \sum_{a=1}^{\Lambda}\left\{\eta_a h_a''\left(\frac{\sum\limits_{i=1}^{M} u_i p_{i0}}{\sum\limits_{i=1}^{M} u_i p_i}\varphi_a\left(\frac{p_i}{p_{i0}}\right)\right) \times \frac{u_i u_j}{\left(\sum\limits_{i=1}^{M} u_i p_i\right)^4}\right. \\ &\times \left(\varphi_a'\left(\frac{p_i}{p_{i0}}\right)\sum_{i=1}^{M} u_i p_i - \sum_{i=1}^{M} u_i p_{i0}\varphi_a\left(\frac{p_i}{p_{i0}}\right)\right) \\ &\times \left.\left(\varphi_a'\left(\frac{p_j}{p_{j0}}\right)\sum_{i=1}^{M} u_i p_i - \sum_{i=1}^{M} u_i p_{i0}\varphi_a\left(\frac{p_i}{p_{i0}}\right)\right)\right\} \end{aligned}$$

$$-\sum_{a=1}^{\Lambda}\left\{\eta_a h_a'\left(\frac{\sum\limits_{i=1}^{M}u_ip_{i0}}{\sum\limits_{i=1}^{M}u_ip_i}\varphi_a\left(\frac{p_i}{p_{i0}}\right)\right)\right.$$

$$\left.\times\frac{u_iu_j}{\left(\sum\limits_{i=1}^{M}u_ip_i\right)^2}\left(\varphi_a'\left(\frac{p_i}{p_{i0}}\right)+\varphi_a'\left(\frac{p_j}{p_{j0}}\right)-2\frac{\sum\limits_{i=1}^{M}u_ip_{i0}\varphi_a\left(\frac{p_i}{p_{i0}}\right)}{\sum\limits_{i=1}^{M}u_ip_i}\right)\right\}$$

y

$$\begin{aligned}
a_{ii} \;=\; & \sum_{a=1}^{\Lambda}\left\{\eta_a h_a''\left(\frac{\sum\limits_{i=1}^{M}u_ip_{i0}}{\sum\limits_{i=1}^{M}u_ip_i}\varphi_a\left(\frac{p_i}{p_{i0}}\right)\right)\times\frac{u_i^2}{\left(\sum\limits_{i=1}^{M}u_ip_i\right)^4}\right.\\
&\left.\times\left(\varphi_a'\left(\frac{p_i}{p_{i0}}\right)\sum_{i=1}^{M}u_ip_i-\sum_{i=1}^{M}u_ip_{i0}\varphi_a\left(\frac{p_i}{p_{i0}}\right)\right)^2\right\}\\
&+\sum_{a=1}^{\Lambda}\left\{\eta_a h_a'\left(\frac{\sum\limits_{i=1}^{M}u_ip_{i0}}{\sum\limits_{i=1}^{M}u_ip_i}\varphi_a\left(\frac{p_i}{p_{i0}}\right)\right)\right.\\
&\left.\times\left[\frac{u_i\varphi_a''\left(\frac{p_i}{p_{i0}}\right)}{p_{i0}\sum\limits_{i=1}^{M}u_ip_i}-\frac{2u_i^2\varphi_a'\left(\frac{p_i}{p_{i0}}\right)}{\left(\sum\limits_{i=1}^{M}u_ip_i\right)^2}+\frac{2u_i^2\sum\limits_{i=1}^{M}u_ip_{i0}\varphi_a\left(\frac{p_i}{p_{i0}}\right)}{\left(\sum\limits_{i=1}^{M}u_ip_i\right)^3}\right]\right\}.
\end{aligned}$$

Corolario 3.4

Si $p=p_0$, se tiene que $WD_{\varphi}^{h}(p,p_0)=0$ y $A=(a_{ij})_{i,j=1,\dots,M}$ con

$$a_{ij}=\begin{cases}0 & si \quad i\neq j\\ \sum\limits_{a=1}^{\Lambda}\eta_a h_a'(0)\,\varphi_a''(1)\,\dfrac{u_i}{p_{i0}\sum\limits_{i=1}^{M}u_ip_{i0}} & si \quad i=j\end{cases}.$$

De esta forma resulta

$$2nWD_{\varphi}^{h}\left(\hat{p},p_0\right) \underset{n\to\infty}{\overset{L}{\to}} \sum_{i=1}^{r}\beta_i Z_i^2$$

donde las Z_i son variables aleatorias normales independientes de media cero y varianza uno, los β_i son los autovalores no nulos de $A\Sigma_p$ y $r = rango\left(\Sigma_p A\Sigma_p\right)$.

Corolario 3.5

Si $p = p_0$, en el caso de las (h,φ)-divergencias, es decir, si $u_i = u$ $\forall i = 1,...,M$, se tiene

$$a_{ij} = \begin{cases} 0 & si \quad i \neq j \\ \sum\limits_{a=1}^{\Lambda} \eta_a h_a'\left(0\right)\varphi_a''\left(1\right)\dfrac{1}{p_{i0}} & si \quad i = j \end{cases},$$

con lo cual

$$A = \sum_{a=1}^{\Lambda}\eta_a h_a'\left(0\right)\varphi_a''\left(1\right)C$$

donde $C = diag\left(p_0^{-1}\right)$.

De esta forma $C\Sigma_{p_0} = \left(\delta_{ij} - p_{j0}\right)_{i,j=1,\dots,M}$ y $C\Sigma_{p_0}C\Sigma_{p_0} = \left(c_{ij}^*\right)_{ij}$ siendo

$$\begin{aligned} c_{ii}^* &= \left(1-p_{i0}\right)^2 + p_{i0}\sum_{j\neq i}^{M}p_{j0} = \left(1-p_{i0}\right)^2 + p_{i0}\left(1-p_{i0}\right) \\ &= \left(1-p_{i0}\right)\left(1-p_{i0}+p_{i0}\right) = 1-p_{i0} \\ c_{ij}^* &= -\left(1-p_{i0}\right)p_{j0} - p_{j0}\left(1-p_{j0}\right) + p_{j0}\sum_{k\neq i,j}^{M}p_{k0} \\ &= p_{j0}\left(-1+p_{i0}-1+p_{j0}+1-p_{i0}-p_{j0}\right) = -p_{j0}. \end{aligned}$$

Es decir, $C\Sigma_{p_0}C\Sigma_{p_0} = C\Sigma_{p_0}$ y aplicando la Observación 3.1, se tiene,

$$\frac{2nWD_{\varphi}^{h}\left(\hat{p},p_0\right)}{\sum\limits_{a=1}^{\Lambda}\eta_a h_a'\left(0\right)\varphi_a''\left(1\right)} \underset{n\to\infty}{\overset{L}{\to}} \mathcal{X}^2_{traza\left(C\Sigma_{p_0}\right)}$$

donde

$$Traza\left(C\Sigma_{p_0}\right) = \sum_{i=1}^{M}\left(1-p_{i0}\right) = M-1.$$

En definitiva, si $\sum_{a=1}^{\Lambda} \eta_a h_a'(0) \varphi_a''(1) \neq 0$, *se obtiene*

$$\frac{2nWD_{\varphi}^{h}(\hat{p}, p_0)}{\sum_{a=1}^{\Lambda} \eta_a h_a'(0) \varphi_a''(1)} \underset{n\to\infty}{\overset{L}{\longrightarrow}} \mathcal{X}_{M-1}^2.$$

Resultado conocido para las (h, φ)*-divergencias. Para más detalles ver Menéndez y otros (1995).*

Corolario 3.6

Si $p = p_0$, *en el caso de* $WD\varphi(\hat{p}, p_0)$, *es decir,* $\Lambda = 1$, $\eta_1 = 1$, $h_1(x) = x$ *y* $\varphi_1(x) = \varphi(x)$, *entonces*

$$a_{ij} = \begin{cases} 0 & si\ i = j \\ \dfrac{u_i \varphi''(1)}{p_{i0} \sum_{i=1}^{M} u_i p_{i0}} & si\ i \neq j \end{cases}$$

y

$$2nWD\varphi(\hat{p}, p_0) \underset{n\to\infty}{\overset{L}{\longrightarrow}} \sum_{i=1}^{M} \beta_i Z_i^2$$

donde los β_i *son los autovalores de la matriz* $A\Sigma_p$ *y* Z_i *son variables aleatorias normales independientes de media cero y varianza uno.*

Corolario 3.7

Si en el caso anterior además $u_i = u\ \forall i = 1, ..., M$, *es decir, en el caso de las* φ*-divergencias, se tiene*

$$a_{ij} = \begin{cases} 0 & si\ i \neq j \\ \varphi''(1)\, p_{i0}^{-1} & si\ i = j \end{cases}$$

por lo que $A = \varphi''(1)\, C$, *donde* $C = diag\left(p_0^{-1}\right)$.

Además ya se ha visto anteriormente que la matriz $C\Sigma_{p_0}$ *es idempotente. En conclusión, si* $\varphi''(1) \neq 0$ *entonces*

$$\frac{2nD\varphi(\hat{p}, p_0)}{\varphi''(1)} \underset{n\to\infty}{\overset{L}{\longrightarrow}} \mathcal{X}_{M-1}^2.$$

3.2.3. Distribución asintótica con hipótesis contiguas alternativas

En este apartado se analizará la distribución asintótica de la familia de los estadísticos

$$2nWD_{\varphi}^{h}(\hat{p}, p_0) \tag{3.1}$$

bajo las hipótesis alternativas contiguas

$$H_{1,n} : p^{(n)} = p_0 + n^{-1/2}d \tag{3.2}$$

donde $d = (d_1, ..., d_M)^t$ con $\sum\limits_{i=1}^{M} d_i = 0$. Es claro que estas hipótesis convergen a la hipótesis nula $H_0 : p = p_0$.

El siguiente teorema presenta la distribución asintótica de la familia de estadísticos dadas en (3.1) bajo las hipótesis contiguas dadas en (3.2).

Antes de establecer el teorema se dará una observación, cuya demostración puede verse en Dik y Gunst (1985).

Observación 3.2

Sea X una variable aleatoria normal d-dimensional con vector de medias μ y matriz de varianzas covarianzas Σ. Sea A una matriz real simétrica de orden d, $r = rango(\Sigma A \Sigma)$, $r \geq 1$ y sean $\lambda_1, ..., \lambda_r$ los autovalores no nulos de $A\Sigma$. Entonces las variables aleatorias

$$X^t AX \;\; y \;\; \sum_{i=1}^{r} \lambda_i (Z_i + \omega_i)^2 + \xi$$

están idénticamente distribuidas donde $Z_1, ..., Z_r$ son variables aleatorias normales e independientes de media cero y varianza uno. Los valores de $\omega = (\omega_1, ..., \omega_r)^t$ y ξ vienen dados mediante

$$\begin{aligned} \omega &= \Lambda^{-1} R^t S^t A \mu \\ \xi &= \mu^t A \mu - \omega^t \Lambda \omega \end{aligned}$$

siendo $\Lambda = diag(\lambda_1, ..., \lambda_r)$, S^t es una raíz arbitraria de Σ y R es la correspondiente matriz de autovectores de $S^t AS$.

Teorema 3.3

Bajo las hipótesis alternativas

$$H_{1,n} : p^{(n)} = p_0 + n^{-1/2}d$$

con $d = (d_1, ..., d_M)^t$, *y* $\sum_{i=1}^{M} d_i = 0$, *y supuesto que las funciones* φ_a, h_a, $a = 1, ..., \Lambda$, *satisfacen las condiciones requeridas en* (1.15), *se tiene*

$$2nWD_{\varphi}^{h}(\hat{p}, p_0) - \xi \underset{n\to\infty}{\overset{L}{\to}} \sum_{i=1}^{r} \beta_i (Z_i + \omega_i)^2$$

siendo $r = rango(\Sigma_{p_0} A \Sigma_{p_0})$, $\beta_1, ..., \beta_r$ *los autovalores positivos de* $A\Sigma_{p_0}$, *A la matriz dada en el Teorema 3.2,* Z_i, $i = 1, ..., r$, *variables aleatorias independientes normales de media cero y varianza uno,* $\omega = \Lambda^{-1} R^t S^t A d$, $\xi = d^t A d - \omega^t \Lambda \omega$ *donde* $\Lambda = diag(\beta_1, ..., \beta_r)$, S^t *es una raíz arbitraria de* Σ_{p_0} *y R es la correspondiente matriz de autovectores de* $S^t A S$.

Demostración

Se tiene,

$$\begin{aligned} \sqrt{n}(\hat{p} - p_0) &= \sqrt{n}\left(\hat{p} - p^{(n)}\right) + \sqrt{n}\left(p^{(n)} - p_0\right) \\ &= \sqrt{n}\left(\hat{p} - p^{(n)}\right) + d. \end{aligned}$$

Entonces al ser, bajo $H_{1,n}$,

$$\sqrt{n}\left(\hat{p} - p^{(n)}\right) \underset{n\to\infty}{\overset{L}{\to}} N(0, \Sigma_{p_0})$$

se llega a

$$\sqrt{n}(\hat{p} - p_0) \underset{n\to\infty}{\overset{L}{\to}} N(d, \Sigma_{p_0}).$$

Por otro lado la distribución asintótica de la familia de estadísticos $2nWD_{\varphi}^{h}(\hat{p}, p_0)$ y de la forma cuadrática $\sqrt{n}(\hat{p} - p_0)^t A \sqrt{n}(\hat{p} - p_0)$ es la misma según se vio en el Teorema 3.2. El resultado ahora se sigue sin más que hacer uso de la observación anterior.

■

3.3. Distribución asintótica de $WD_{\varphi}^{h}(p_0, \hat{p})$

En este apartado se va a hallar la distribución asintótica de $WD_{\varphi}^{h}(p_0, \hat{p})$, es decir, si se permutan los argumentos del apartado anterior.

Se tiene p_0 conocida y de la misma forma que en el caso anterior p se estima a través del vector de frecuencias relativas basado en una muestra aleatoria simple de tamaño n.

Sea

$$WD_{\varphi}^{h}(p_0, p) = \sum_{a=1}^{\Lambda} \eta_a h_a \left(\frac{\sum\limits_{i=1}^{M} u_i p_i}{\sum\limits_{i=1}^{M} u_i p_{i0}} \varphi_a \left(\frac{p_{i0}}{p_i} \right) \right)$$

De manera análoga al apartado 3.2, desarrollando en serie de Taylor la función $WD_{\varphi}^{h}(p_0, \hat{p})$ en el punto $(\hat{p}_1, ..., \hat{p}_M)$ y en un entorno de $(p_1, ..., p_M)$, se obtiene que las variables aleatorias

$$n^{\frac{1}{2}} \left(WD_{\varphi}^{h}(p_0, \hat{p}) - WD_{\varphi}^{h}(p_0, p) \right)$$

y

$$n^{\frac{1}{2}} \sum_{i=1}^{M} s_i (\hat{p}_i - p_i) = n^{\frac{1}{2}} S^t (\hat{p} - p)$$

siguen la misma distribución asintótica, siendo $S = (s_1, ..., s_M)^t$ y

$$s_i = \left(\frac{\partial WD_{\varphi}^{h}(p_0, p)}{\partial p_i} \right)_{p=(p_1,....,p_M)}$$

con $i = 1, ..., M$.

Ahora bien

$$n^{\frac{1}{2}} S^t (\hat{p} - p) \underset{n \to \infty}{\overset{L}{\longrightarrow}} N\left(0, \sigma_p^2\right)$$

siendo

$$\begin{array}{ll} \sigma_p^2 & = S^t \Sigma_p S \\ \Sigma_p & = diag(p) - pp^t \\ p & = (p_1, ..., p_M)^t . \end{array}$$

Con lo que se puede establecer el siguiente teorema:

Teorema 3.4

Sea $WD_{\varphi}^{h}(p_0, p)$ y su estimador analógico, $WD_{\varphi}^{h}(p_0, \hat{p})$, obtenido al reemplazar p_i por sus frecuencias relativas observadas $\hat{p}_i$, $i = 1, ..., M$, basadas en una *muestra aleatoria simple de tamaño n. Bajo las hipótesis de que las funciones* h_a *y* φ_a, $a = 1, ..., \Lambda$, *verifican las hipótesis requeridas en* (1.15), *se tiene*

$$n^{\frac{1}{2}}\left(WD_{\varphi}^{h}(p_0, \hat{p}) - WD_{\varphi}^{h}(p_0, p)\right) \underset{n\to\infty}{\overset{L}{\to}} N\left(0, \sigma_p^2\right)$$

siempre y cuando $\sigma_p^2 > 0$, *donde*

$$\sigma_p^2 = S^t \Sigma_p S = \sum_{i=1}^{M} s_i^2 p_i - \left(\sum_{i=1}^{M} s_i p_i\right)^2$$

con

$$\Sigma_p = \left(p_i\left(\delta_{ij} - p_i\right)\right)_{i,j=1,....,M} = diag\,(p) - pp^t$$

y

$$p = (p_1, ..., p_M)^t$$

siendo, $\forall i = 1, ..., M$,

$$s_i = \sum_{a=1}^{\Lambda}\left\{\eta_a h_a'\left(\frac{\sum\limits_{i=1}^{M} u_i p_i}{\sum\limits_{i=1}^{M} u_i p_{i0}}\varphi_a\left(\frac{p_{i0}}{p_i}\right)\right)\left[\frac{u_i}{\sum\limits_{i=1}^{M} u_i p_{i0}}\left(\varphi_a\left(\frac{p_{i0}}{p_i}\right) - \varphi_a'\left(\frac{p_{i0}}{p_i}\right)\frac{p_{i0}}{p_i}\right)\right]\right\}.$$

Corolario 3.8

En el caso de las (h, φ)*-divergencias, es decir, si* $u_i = u$ $\forall i = 1, ..., M$, *se tiene*

$$s_i = \sum_{a=1}^{\Lambda}\left\{\eta_a h_a'\left(\sum_{i=1}^{M} p_i \varphi_a\left(\frac{p_{i0}}{p_i}\right)\right)\left(\varphi_a\left(\frac{p_{i0}}{p_i}\right) - \varphi_a'\left(\frac{p_{i0}}{p_i}\right)\frac{p_{i0}}{p_i}\right)\right\}$$

y

$$\begin{aligned}\sigma_p^2 &= \sum_{i=1}^{M}\left(\sum_{a=1}^{\Lambda}\eta_a h_a'\left(\sum_{i=1}^{M} p_i\varphi_a\left(\frac{p_{i0}}{p_i}\right)\right)\left(\varphi_a\left(\frac{p_{i0}}{p_i}\right) - \varphi_a'\left(\frac{p_{i0}}{p_i}\right)\frac{p_{i0}}{p_i}\right)\right)^2 p_i \\ &\quad - \left(\sum_{i=1}^{M}\left(\sum_{a=1}^{\Lambda}\eta_a h_a'\left(\sum_{i=1}^{M} p_i\varphi_a\left(\frac{p_{i0}}{p_i}\right)\right)\left(\varphi_a\left(\frac{p_{i0}}{p_i}\right) - \varphi_a'\left(\frac{p_{i0}}{p_i}\right)\frac{p_{i0}}{p_i}\right)\right)p_i\right)^2.\end{aligned}$$

Corolario 3.9

En el caso de las φ-divergencias ponderadas, $WD_{\varphi}(p_0, \hat{p})$. Es decir, si $\Lambda = 1$, $\eta_1 = 1$, $h_1(x) = x$ y $\varphi_1(x) = \varphi(x)$, se tiene

$$s_i = \frac{u_i}{\sum\limits_{i=1}^{M} u_i p_{i0}} \left(\varphi\left(\frac{p_{i0}}{p_i}\right) - \varphi'\left(\frac{p_{i0}}{p_i}\right)\frac{p_{i0}}{p_i} \right)$$

y

$$\begin{aligned} \sigma_p^2 &= \sum_{i=1}^{M} \left(\frac{u_i}{\sum\limits_{i=1}^{M} u_i p_{i0}} \left(\varphi\left(\frac{p_{i0}}{p_i}\right) - \varphi'\left(\frac{p_{i0}}{p_i}\right)\frac{p_{i0}}{p_i} \right)^2 p_i \right) \\ &\quad - \left(\sum_{i=1}^{M} \left(\frac{u_i}{\sum\limits_{i=1}^{M} u_i p_{i0}} \left(\varphi\left(\frac{p_{i0}}{p_i}\right) - \varphi'\left(\frac{p_{i0}}{p_i}\right)\frac{p_{i0}}{p_i} \right) \right) p_i \right)^2 . \end{aligned}$$

Corolario 3.10

En el caso de las φ-divergencias, es decir, si $u_i = u$ $\forall i = 1, ..., M$, $\Lambda = 1$, $\eta_1 = 1$, $h_1(x) = x$ y $\varphi_1(x) = \varphi(x)$, se tiene

$$s_i = \varphi\left(\frac{p_{i0}}{p_i}\right) - \varphi'\left(\frac{p_{i0}}{p_i}\right)\frac{p_{i0}}{p_i}$$

y

$$\sigma_p^2 = \sum_{i=1}^{M} \left(\varphi\left(\frac{p_{i0}}{p_i}\right) - \varphi'\left(\frac{p_{i0}}{p_i}\right)\frac{p_{i0}}{p_i} \right)^2 p_i - \left(\sum_{i=1}^{M} \left(\varphi\left(\frac{p_{i0}}{p_i}\right) - \varphi'\left(\frac{p_{i0}}{p_i}\right)\frac{p_{i0}}{p_i} \right) p_i \right)^2 .$$

En el caso de la divergencia de Kullback, $\varphi(x) = x \log x - x + 1$, se tiene

$$\sigma_p^2 = \sum_{i=1}^{M} \frac{p_{i0}^2}{p_i} - 1.$$

Si el primer término del desarrollo de Taylor de la función $WD_{\varphi}^{h}(p_0, \hat{p})$ es cero, entonces para obtener su distribución asintótica será preciso recurrir al término de segundo orden de dicho desarrollo.

Proposición 3.2

Sea $S_n = \sqrt{n}\sum\limits_{i=1}^{M} s_i\left(\hat{p}_i - p_i\right)$. *Entonces*

$$S_n = 0 \ \forall n \Leftrightarrow \sigma_p^2 = 0$$

Demostración

Análoga a la demostración de la Proposición 3.1 ■

Si $p_0 = p$ se tiene que $\sigma_p^2 = 0$ y por consiguiente el primer término del desarrollo en serie de Taylor de $WD_{\varphi}^{h}(p_0,\hat{p})$ se anula. Antes de abordar esta situación, en el siguiente teorema se presenta la distribución asintótica de $WD_{\varphi}^{h}(p_0,\hat{p})$ cuando $\sigma_p^2 = 0$.

Teorema 3.5

Sea $WD_{\varphi}^{h}(p_0,p)$ *y su estimador analógico* $WD_{\varphi}^{h}(p_0,\hat{p})$ *obtenido al reemplazar* p_i *por sus frecuencias relativas observadas* $\hat{p}_i$, $i = 1, ..., M$, *basadas en una muestra aleatoria simple de tamaño* n. *Bajo las hipótesis de que las funciones* h_a *y* φ_a, $a = 1, ..., \Lambda$, *verifican las hipótesis requeridas en* (1.15). *Si* $\sigma_p^2 = 0$, *entonces*

$$2n\left[WD_{\varphi}^{h}(p_0,\hat{p}) - WD_{\varphi}^{h}(p_0,p)\right] \underset{n\to\infty}{\overset{L}{\to}} \sum_{i=1}^{r}\beta_i Z_i^2$$

donde las Z_i *son variables aleatorias normales independientes de media cero y varianza uno, los* β_i *son los autovalores no nulos de* $B\Sigma_p$ *y* $r = rango\left(\Sigma_p B \Sigma_p\right)$. *Siendo*

$$\Sigma_p = diag\left(p\right) - pp^t$$

y

$$B = \left(\frac{\partial^2 WD_{\varphi}^{h}(p_0,p)}{\partial p_i \partial p_j}\right)_{i,j=1,....,M} = \left(b_{ij}\right)_{i,j=1,...,M}$$

con

$$b_{ij} = \left\{\sum_{a=1}^{\Lambda}\eta_a h_a''\left(\frac{\sum\limits_{i=1}^{M}u_i p_i}{\sum\limits_{i=1}^{M}u_i p_{i0}}\varphi_a\left(\frac{p_{i0}}{p_i}\right)\right)\frac{u_i u_j}{\left(\sum\limits_{i=1}^{M}u_i p_{i0}\right)^2}\left(\varphi_a\left(\frac{p_{i0}}{p_i}\right) - \varphi_a'\left(\frac{p_{i0}}{p_i}\right)\frac{p_{i0}}{p_i}\right)\right.$$

$$\left.\times\left(\varphi_a\left(\frac{p_{j0}}{p_j}\right) - \varphi_a'\left(\frac{p_{j0}}{p_j}\right)\frac{p_{j0}}{p_j}\right)\right\}$$

y

$$b_{ii}=\sum_{a=1}^{\Lambda}\left\{\eta_a h_a''\left(\frac{\sum\limits_{i=1}^{M}u_ip_i}{\sum\limits_{i=1}^{M}u_ip_{i0}}\varphi_a\left(\frac{p_{i0}}{p_i}\right)\right)\frac{u_i^2}{\left(\sum\limits_{i=1}^{M}u_ip_{i0}\right)^2}\left(\varphi_a\left(\frac{p_{i0}}{p_i}\right)-\varphi_a'\left(\frac{p_{i0}}{p_i}\right)\frac{p_{i0}}{p_i}\right)^2\right\}$$

$$+\sum_{a=1}^{\Lambda}\left\{\eta_a h_a'\left(\frac{\sum\limits_{i=1}^{M}u_ip_i}{\sum\limits_{i=1}^{M}u_ip_{i0}}\varphi_a\left(\frac{p_{i0}}{p_i}\right)\right)\frac{u_ip_{i0}^2}{p_i^3\sum\limits_{i=1}^{M}u_ip_{i0}}\varphi_a''\left(\frac{p_{i0}}{p_i}\right)\right\}.$$

Demostración

Se siguen los mismos pasos que en la demostración del Teorema 3.2. ∎

Observación 3.3
Si $p=p_0$ se tiene que

$$b_{ij}=\begin{cases}0 & si\ i\neq j\\ \sum\limits_{a=1}^{\Lambda}\eta_a h_a'(0)\,\varphi_a''(1)\,\dfrac{u_i}{p_{i0}\sum\limits_{i=1}^{M}u_ip_{i0}} & si\ i=j\end{cases}.$$

Corolario 3.11
Si $p=p_0$, en el caso de las (h,φ)-divergencias, es decir, si $u_i=u\ \forall i=1,...,M$, se tiene

$$b_{ij}=\begin{cases}0 & si\quad i\neq j\\ \sum\limits_{a=1}^{\Lambda}\eta_a h_a'(0)\,\varphi_a''(1)\,\dfrac{1}{p_{i0}} & si\quad i=j\end{cases}.$$

Con lo cual

$$B=\sum_{a=1}^{\Lambda}\eta_a h_a'(0)\,\varphi_a''(1)\,diag\left(p_0^{-1}\right)$$

y si $C=diag\left(p_0^{-1}\right)$, se tiene que $C\Sigma_{p_0}$ es una matriz idempotente.

Entonces, si $\sum_{a=1}^{\Lambda} \eta_a h'_a(0)\, \varphi''_a(1) \neq 0$

$$\frac{2nWD_{\varphi}^{h}(p_0, \hat{p})}{\sum_{a=1}^{\Lambda} \eta_a h'_a(0)\, \varphi''_a(1)} \underset{n\to\infty}{\overset{L}{\to}} \mathcal{X}^2_{M-1}.$$

Corolario 3.12

Si $p = p_0$*, en el caso de las* φ*-divergencias ponderadas,* $WD_{\varphi}(p_0, \hat{p})$*, se tiene*

$$b_{ij} = \begin{cases} 0 & si \quad i \neq j \\ \varphi''(1) \dfrac{u_i}{p_{i0} \sum_{i=1}^{M} u_i p_{i0}} & si \quad i = j \end{cases}$$

y

$$2nWD_{\varphi}(p_0, \hat{p}) \underset{n\to\infty}{\overset{L}{\to}} \sum_{i=1}^{M} \beta_i Z_i^2$$

donde los β_i *son los autovalores de la matriz* $B\Sigma_p$ *y* Z_i *son variables aleatorias normales independientes de media cero y varianza uno.*

Si en el caso anterior además $u_i = u\ \forall i = 1, ..., M$*, es decir, en el caso de las* φ*-divergencias, se tiene*

$$b_{ij} = \begin{cases} 0 & si\ i \neq j \\ \varphi''(1)\, p_{i0}^{-1} & si\ i = j \end{cases}$$

por lo que $B = \varphi''(1)\, C$*, donde* $C = diag\left(p_0^{-1}\right)$*.*

Además como ya se ha visto la matriz $C\Sigma_{p_0}$ *es idempotente. En conclusión, si* $\varphi''(1) \neq 0$ *entonces*

$$\frac{2nD\varphi(p_0, \hat{p})}{\varphi''(1)} \underset{n\to\infty}{\overset{L}{\to}} \mathcal{X}^2_{M-1}.$$

3.4. Distribución asintótica de $WD_{\varphi}^{h_1}\left(\hat{p}, p\left(\hat{\theta}_{h_2,\phi,w}\right)\right)$

Para realizar el contraste de bondad de ajuste

$$H_0 : \pi = \pi_0 \tag{3.3}$$

donde $\pi_0 = p(\theta) = (p_1(\theta), ..., p_M(\theta))^t \in \Delta_{M,\theta} \subset \Delta_M$ y $\theta \in \Theta \subseteq R^{M_0}$ es desconocido, se deberá estimar previamente el parámetro. Es decir, hay que elegir un valor $p\left(\hat{\theta}\right) \in \Delta_{M,\theta}$ que sea lo más consistente posible con las frecuencias observadas $\hat{p}$. El método más conocido para seleccionar $p(\theta)$ consiste en estimar θ por máxima verosilimitud en el modelo multinomial asociado, pero también es una opción razonable elegir el valor de $p(\theta) \in \Delta_{M,\theta}$ más próximo a $\hat{p}$ con respecto a la medida $WD_\varphi^h(\hat{p}, p(\theta))$. En definitiva, se puede considerar $p\left(\hat{\theta}_{h,\varphi,w}\right)$, donde $\hat{\theta}_{h,\varphi,w}$ es el estimador de mínima WD_φ^h-divergencia introducido y estudiado en el Capítulo 2.

En el siguiente teorema se presenta la distribución asintótica de $WD_\phi^h(\hat{p}, p(\theta))$ bajo la hipótesis dada en (3.3).

Teorema 3.6

Supóngase que se verifican las condiciones de regularidad señaladas en el Teorema 2.2, entonces

$$\frac{2nWD_\varphi^{h_1}\left(\hat{p}, p\left(\hat{\theta}_{h_2,\phi,w}\right)\right)}{\sum\limits_{a=1}^{\Lambda} \eta_a h'_{1_a}(0)\, \varphi''_a(1)} \underset{n\to\infty}{\overset{L}{\to}} \sum_{i=1}^{M} \beta_i Z_i^2$$

donde β_i son los autovalores de la matriz

$$T = D\left(\Sigma_{p(\theta_0)} - L\Sigma_{p(\theta_0)} - \Sigma_{p(\theta_0)}L^t + L\Sigma_{p(\theta_0)}L^t\right)$$

con

$$D = diag\left(\mathcal{U}_*\right)\frac{1}{E_{p(\theta_0)}(u)}$$

$$\Sigma_{p(\theta_0)} = diag\left(p(\theta_0)\right) - p(\theta_0)\, p(\theta_0)^t$$

$$L = \left(\left(\frac{\partial p_i(\theta)}{\partial \theta_r}\right)_{\substack{i=1,...,M\\ r=1,...,M_0}}\right)\left(A^*(\theta_0)^t A^*(\theta_0)\right)^{-1} A^*(\theta_0)^t\, diag\left(\mathcal{U}_*^{1/2}\right)$$

$$\mathcal{U}_* = \left(\frac{u_1}{p_1(\theta_0)}, ..., \frac{u_M}{p_M(\theta_0)}\right)^t$$

y las Z_i son variables aleatorias normales e independientes de media cero y varianza uno.

Demostración

Se sabe que

$$\begin{aligned} \hat{\theta}_{h_2,\phi,w} - \theta_0 &= \left(A^*(\theta_0)^t A^*(\theta_0)\right)^{-1} A^*(\theta_0)^t \, diag\left(\mathcal{U}_*^{1/2}\right)(\hat{p} - p(\theta_0)) + o_p\left(n^{-1/2}\right) \\ &= C(\theta_0)(\hat{p} - p(\theta_0)) + o_p\left(n^{-1/2}\right) \end{aligned}$$

y que

$$p\left(\hat{\theta}_{h_2,\phi,w}\right) - p(\theta_0) = \left(\left(\frac{\partial p_i(\theta)}{\partial \theta_r}\right)_{\substack{i=1,\dots,M \\ r=1,\dots,M_0}}\right)\left(\hat{\theta}_{h_2,\phi,w} - \theta_0\right) + o_p\left(n^{-1/2}\right).$$

Por tanto

$$p\left(\hat{\theta}_{h_2,\phi,w}\right) - p(\theta_0) = \left(\left(\frac{\partial p_i(\theta)}{\partial \theta_r}\right)_{\substack{i=1,\dots,M \\ r=1,\dots,M_0}}\right) C(\theta_0)(\hat{p} - p(\theta_0)) + o_p\left(n^{-1/2}\right)$$

con $C(\theta_0)$ definido de la misma forma que en el Teorema 2.4.

En lo que sigue se denotará

$$\begin{aligned} L &= \left(\left(\frac{\partial p_i(\theta)}{\partial \theta_r}\right)_{\substack{i=1,\dots,M \\ r=1,\dots,M_0}}\right)\left(A^*(\theta_0)^t A^*(\theta_0)\right)^{-1} A^*(\theta_0)^t \, diag\left(\mathcal{U}_*^{1/2}\right) \\ &= \left(\left(\frac{\partial p_i(\theta)}{\partial \theta_r}\right)_{\substack{i=1,\dots,M \\ r=1,\dots,M_0}}\right) C(\theta_0). \end{aligned}$$

Al ser

$$\begin{pmatrix} \hat{p} - p(\theta_0) \\ p\left(\hat{\theta}_{h_2,\phi,w}\right) - p(\theta_0) \end{pmatrix}_{2M\times 1} = \begin{pmatrix} I \\ L \end{pmatrix}_{2M\times M} (\hat{p} - p(\theta_0))_{M\times 1} + o_p\left(n^{-1/2}\right)$$

se tiene que

$$\sqrt{n}\begin{pmatrix} \hat{p} - p(\theta_0) \\ p\left(\hat{\theta}_{h_2,\phi,w}\right) - p(\theta_0) \end{pmatrix}_{2M\times 1} \underset{n\to\infty}{\overset{L}{\longrightarrow}} N\left(0, \begin{pmatrix} I \\ L \end{pmatrix} \Sigma_{p(\theta_0)}\left(I, L^t\right)\right)$$

ya que

$$\sqrt{n}(\hat{p} - p(\theta_0)) \underset{n\to\infty}{\overset{L}{\longrightarrow}} N\left(0, \Sigma_{p(\theta_0)}\right).$$

Es decir,

$$\sqrt{n}\begin{pmatrix} \hat{p}-p\left(\theta_0\right) \\ p\left(\hat{\theta}_{h_2,\phi,w}\right)-p\left(\theta_0\right)\end{pmatrix}_{2M\times 1} \underset{n\to\infty}{\overset{L}{\longrightarrow}} N\left(0,\begin{pmatrix} \Sigma_{p(\theta_0)} & \Sigma_{p(\theta_0)}L^t \\ L\Sigma_{p(\theta_0)} & L\Sigma_{p(\theta_0)}L^t\end{pmatrix}\right)$$

y en consecuencia

$$\sqrt{n}\left(\hat{p}-p\left(\hat{\theta}_{h_2,\phi,w}\right)\right) \underset{n\to\infty}{\overset{L}{\longrightarrow}} N\left(0,\Sigma_{p(\theta_0)}-\Sigma_{p(\theta_0)}L^t-L\Sigma_{p(\theta_0)}+L\Sigma_{p(\theta_0)}L^t\right).$$

Por otro lado, llamando

$$A=WD_{\varphi}^{h_1}\left(\hat{p},p\left(\hat{\theta}_{h_2,\phi,w}\right)\right),$$

se tiene

$$\begin{aligned} A &= \frac{1}{2}\sum_{i,j=1}^{M}\left(\frac{\partial^2 WD_{\varphi}^{h_1}\left(p,p\left(\theta_0\right)\right)}{\partial p_i\partial p_j}\right)_{p=p(\theta_0)}\left(\hat{p}_i-p_i\left(\theta_0\right)\right)\left(\hat{p}_j-p_j\left(\theta_0\right)\right) \\ &\quad +\frac{1}{2}2\sum_{i,j=1}^{M}\left(\frac{\partial^2 WD_{\varphi}^{h_1}\left(p,p\left(\theta_0\right)\right)}{\partial p_i\partial p_j\left(\theta_0\right)}\right)_{p=p(\theta_0)}\left(\hat{p}_i-p_i\left(\theta_0\right)\right)\left(p_j\left(\hat{\theta}_{h_2,\phi_2,w}\right)-p_j\left(\theta_0\right)\right) \\ &\quad +\frac{1}{2}\sum_{i,j=1}^{M}\left(\frac{\partial^2 WD_{\varphi}^{h_1}\left(p,p\left(\theta_0\right)\right)}{\partial p_i\left(\theta_0\right)\partial p_j\left(\theta_0\right)}\right)_{p=p(\theta_0)}\left(p_i\left(\hat{\theta}_{h_2,\phi,w}\right)-p_i\left(\theta_0\right)\right)\left(p_j\left(\hat{\theta}_{h_2,\phi,w}\right)-p_j\left(\theta_0\right)\right) \\ &\quad +o_p\left(n^{-1}\right). \end{aligned}$$

Ahora

$$\left(\frac{\partial^2 WD_{\varphi}^{h_1}\left(p,p\left(\theta_0\right)\right)}{\partial p_i\partial p_j}\right)_{p=p(\theta_0)} = \begin{cases} 0 & si \quad i\neq j \\ \sum\limits_{a=1}^{\Lambda}\eta_a h_{1_a}'\left(0\right)\varphi_a''\left(1\right)\dfrac{u_i}{p_i\left(\theta_0\right)E_{p(\theta_0)}\left(u\right)} & si \quad i=j \end{cases}$$

$$\left(\frac{\partial^2 WD_{\varphi}^{h_1}\left(p,p\left(\theta_0\right)\right)}{\partial p_i\partial p_j\left(\theta_0\right)}\right)_{p=p(\theta_0)} = \begin{cases} 0 & si \quad i\neq j \\ -\sum\limits_{a=1}^{\Lambda}\eta_a h_{1_a}'\left(0\right)\varphi_a''\left(1\right)\dfrac{u_i}{p_i\left(\theta_0\right)E_{p(\theta_0)}\left(u\right)} & si \quad i=j \end{cases}$$

$$\left(\frac{\partial^2 WD_{\varphi}^{h_1}\left(p, p\left(\theta_0\right)\right)}{\partial p_i\left(\theta_0\right)\partial p_j\left(\theta_0\right)}\right)_{p=p(\theta_0)} = \begin{cases} 0 & si \quad i \neq j \\ \sum\limits_{a=1}^{\Lambda} \eta_a h_{1_a}'\left(0\right)\varphi_a''\left(1\right)\dfrac{u_i}{p_i\left(\theta_0\right)E_{p(\theta_0)}\left(u\right)} & si \quad i = j \end{cases}$$

en consecuencia

$$\begin{aligned} A &= \frac{1}{2}\sum_{i=1}^{M}\sum_{a=1}^{\Lambda}\eta_a h_{1_a}'\left(0\right)\varphi_a''\left(1\right)\frac{u_i}{p_i\left(\theta_0\right)E_{p(\theta_0)}\left(u\right)}\left(\hat{p}_i - p_i\left(\theta_0\right)\right)^2 \\ &\quad -\sum_{i=1}^{M}\sum_{a=1}^{\Lambda}\eta_a h_{1_a}'\left(0\right)\varphi_a''\left(1\right)\frac{u_i}{p_i\left(\theta_0\right)E_{p(\theta_0)}\left(u\right)}\left(\hat{p}_i - p_i\left(\theta_0\right)\right)\left(p_i\left(\hat{\theta}_{h_2,\phi,w}\right) - p_i\left(\theta_0\right)\right) \\ &\quad +\frac{1}{2}\sum_{i=1}^{M}\sum_{a=1}^{\Lambda}\eta_a h_{1_a}'\left(0\right)\varphi_a''\left(1\right)\frac{u_i}{p_i\left(\theta_0\right)E_{p(\theta_0)}\left(u\right)}\left(p_i\left(\hat{\theta}_{h_2,\phi,w}\right) - p_i\left(\theta_0\right)\right)^2 \\ &\quad +o_p\left(n^{-1}\right) \\ &= \frac{1}{2}\sum_{i=1}^{M}\sum_{a=1}^{\Lambda}\eta_a h_{1_a}'\left(0\right)\varphi_a''\left(1\right)\frac{u_i}{p_i\left(\theta_0\right)E_{p(\theta_0)}\left(u\right)}\left[\left(\hat{p}_i - p_i\left(\theta_0\right)\right)^2\right. \\ &\quad \left. -2\left(\hat{p}_i - p_i\left(\theta_0\right)\right)\left(p_i\left(\hat{\theta}_{h_2,\phi,w}\right) - p_i\left(\theta_0\right)\right) + \left(p_i\left(\hat{\theta}_{h_2,\phi,w}\right) - p_i\left(\theta_0\right)\right)^2\right] \\ &\quad +o_p\left(n^{-1}\right) \\ &= \frac{1}{2}\sum_{i=1}^{M}\sum_{a=1}^{\Lambda}\eta_a h_{1_a}'\left(0\right)\varphi_a''\left(1\right)\frac{u_i}{p_i\left(\theta_0\right)E_{p(\theta_0)}\left(u\right)}\left(\hat{p}_i - p_i\left(\hat{\theta}_{h_2,\phi,w}\right)\right)^2 + o_p\left(n^{-1}\right). \end{aligned}$$

Así pues,

$$\frac{2WD_{\varphi}^{h_1}\left(\hat{p}, p\left(\hat{\theta}_{h_2,\phi,w}\right)\right)}{\sum\limits_{a=1}^{\Lambda}\eta_a h_{1_a}'\left(0\right)\varphi_a''\left(1\right)} = \sum_{i=1}^{M}\frac{u_i}{p_i\left(\theta_0\right)E_{p(\theta_0)}\left(u\right)}\left(\hat{p}_i - p_i\left(\hat{\theta}_{h_2,\phi,w}\right)\right)^2 + o_p\left(n^{-1}\right)$$

o lo que es lo mismo

$$\frac{2WD_{\varphi}^{h_1}\left(\hat{p}, p\left(\hat{\theta}_{h_2,\phi,w}\right)\right)}{\sum\limits_{a=1}^{\Lambda}\eta_a h_{1_a}'\left(0\right)\varphi_a''\left(1\right)} = \left(\hat{p} - p\left(\hat{\theta}_{h_2,\phi,w}\right)\right)^t D\left(\hat{p} - p\left(\hat{\theta}_{h_2,\phi,w}\right)\right) + o_p\left(n^{-1}\right)$$

con

$$D = diag\left(\mathcal{U}_*\right) \frac{1}{E_{p(\theta_0)}\left(u\right)}.$$

Resumiendo, por un lado se tiene

$$\sqrt{n}\left(\hat{p} - p\left(\hat{\theta}_{h_2,\phi,w}\right)\right) \underset{n\to\infty}{\overset{L}{\to}} N\left(0, \Sigma_{p(\theta_0)} - \Sigma_{p(\theta_0)}L^t - L\Sigma_{p(\theta_0)} + L\Sigma_{p(\theta_0)}L^t\right)$$

y por otro lado

$$\frac{2nWD_{\varphi}^{h_1}\left(\hat{p}, p\left(\hat{\theta}_{h_2,\phi,w}\right)\right)}{\sum\limits_{a=1}^{\Lambda} \eta_a h'_{1_a}\left(0\right)\varphi''_a\left(1\right)} = n\left(\hat{p} - p\left(\hat{\theta}_{h_2,\phi,w}\right)\right)^t D\left(\hat{p} - p\left(\hat{\theta}_{h_2,\phi,w}\right)\right) + o_p\left(1\right)$$

con lo cual

$$\frac{2nWD_{\varphi}^{h_1}\left(\hat{p}, p\left(\hat{\theta}_{h_2,\phi,w}\right)\right)}{\sum\limits_{a=1}^{\Lambda} \eta_a h'_{1_a}\left(0\right)\varphi''_a\left(1\right)} \underset{n\to\infty}{\overset{L}{\to}} \sum_{i=1}^{M} \beta_i Z_i^2$$

donde β_i son los autovalores de la matriz

$$T = D\left(\Sigma_{p(\theta_0)} - L\Sigma_{p(\theta_0)} - \Sigma_{p(\theta_0)}L^t + L\Sigma_{p(\theta_0)}L^t\right)$$

y Z_i son variables aleatorias normales e independientes de media cero y varianza uno.

■

Corolario 3.13

En el caso de las (h, φ)-divergencias, es decir, si $u_i = u$ $\forall i = 1, ..., M$, entonces se tiene

$$\frac{2nD_{\varphi}^{h_1}\left(\hat{p}, p\left(\hat{\theta}_{h_2,\phi,w}\right)\right)}{\sum\limits_{a=1}^{\Lambda} \eta_a h'_{1_a}\left(0\right)\varphi''_a\left(1\right)} \underset{n\to\infty}{\overset{L}{\to}} \mathcal{X}^2_{M-M_0-1}.$$

Demostración

En este caso se tiene que

$$\frac{2nD_{\varphi}^{h_1}\left(\hat{p}, p\left(\hat{\theta}_{h_2,\phi,w}\right)\right)}{\sum\limits_{a=1}^{\Lambda}\eta_a h'_{1_a}(0)\,\varphi''_a(1)} \underset{n\to\infty}{\overset{L}{\longrightarrow}} \sum_{i=1}^{M}\beta_i Z_i^2$$

donde β_i son los autovalores de la matriz

$$T = D\left(\Sigma_{p(\theta_0)} - L\Sigma_{p(\theta_0)} - \Sigma_{p(\theta_0)}L^t + L\Sigma_{p(\theta_0)}L^t\right)$$

con

$$D = diag\left(p(\theta_0)^{-1}\right)$$

$$\Sigma_{p(\theta_0)} = diag\left(p(\theta_0)\right) - p(\theta_0)\,p(\theta_0)^t$$

$$L = \left(\left(\frac{\partial p_i(\theta)}{\partial\theta_r}\right)_{\substack{i=1,\dots,M\\ r=1,\dots,M_0}}\right)\left(A(\theta_0)^t A(\theta_0)\right)^{-1} A(\theta_0)^t\, diag\left(p(\theta_0)^{-1}\right)^{1/2}$$

siendo

$$A(\theta_0) = diag\left(p(\theta_0)^{-1/2}\right)\left(\left(\frac{\partial p_i(\theta)}{\partial\theta_r}\right)_{\substack{i=1,\dots,M\\ r=1,\dots,M_0}}\right).$$

Es inmediato comprobar que los autovalores de la matriz

$$diag\left(p^{-1}\right)M$$

coinciden con los autovalores de la matriz

$$diag\left(p^{-1/2}\right)M diag\left(p^{-1/2}\right)$$

ya que

$$\left|diag\left(p^{-1}\right)M - \lambda I\right| = 0 \Longleftrightarrow \left|diag\left(p^{-1/2}\right)M diag\left(p^{-1/2}\right) - \lambda I\right| = 0.$$

Por tanto los autovalores de la matriz T son los mismos que los de la matriz T^*, con

$$\begin{aligned}
T^* &= diag\left(p(\theta_0)^{-1/2}\right)\left(\Sigma_{p(\theta_0)} - L\Sigma_{p(\theta_0)} - \Sigma_{p(\theta_0)}L^t + L\Sigma_{p(\theta_0)}L^t\right) diag\left(p(\theta_0)^{-1/2}\right) \\
&= diag\left(p(\theta_0)^{-1/2}\right)\Sigma_{p(\theta_0)} diag\left(p(\theta_0)^{-1/2}\right) - diag\left(p(\theta_0)^{-1/2}\right)L\Sigma_{p(\theta_0)} diag\left(p(\theta_0)^{-1/2}\right) \\
&\quad + diag\left(p(\theta_0)^{-1/2}\right)L\Sigma_{p(\theta_0)}L^t diag\left(p(\theta_0)^{-1/2}\right).
\end{aligned}$$

Ahora bien, llamando

$$S = diag\left(p\left(\theta_0\right)^{-1/2}\right)\Sigma_{p(\theta_0)} diag\left(p\left(\theta_0\right)^{-1/2}\right)$$

se tiene

$$\begin{aligned} S &= diag\left(p\left(\theta_0\right)^{-1/2}\right)\left(diag\left(p\left(\theta_0\right)\right) - p\left(\theta_0\right)p\left(\theta_0\right)^t\right) diag\left(p\left(\theta_0\right)^{-1/2}\right) \\ &= I - diag\left(p\left(\theta_0\right)^{-1/2}\right) p\left(\theta_0\right)p\left(\theta_0\right)^t diag\left(p\left(\theta_0\right)^{-1/2}\right) \\ &= I - p\left(\theta_0\right)^{1/2}\left(p\left(\theta_0\right)^{1/2}\right)^t. \end{aligned}$$

A continuación, se da una forma más sencilla de expresar la matriz

$$B = diag\left(p\left(\theta_0\right)^{-1/2}\right) L\Sigma_{p(\theta_0)} diag\left(p\left(\theta_0\right)^{-1/2}\right).$$

Se tiene

$$\begin{aligned} B &= diag\left(p\left(\theta_0\right)^{-1/2}\right) L\left(diag\left(p\left(\theta_0\right)\right) - p\left(\theta_0\right)p\left(\theta_0\right)^t\right) diag\left(p\left(\theta_0\right)^{-1/2}\right) \\ &= diag\left(p\left(\theta_0\right)^{-1/2}\right) Ldiag\left(p\left(\theta_0\right)^{1/2}\right) \\ &\quad - diag\left(p\left(\theta_0\right)^{-1/2}\right) Lp\left(\theta_0\right)p\left(\theta_0\right)^t diag\left(p\left(\theta_0\right)^{-1/2}\right) \end{aligned}$$

y teniendo en cuenta que

$$p\left(\theta_0\right)^t = \sqrt{p\left(\theta_0\right)^t} diag\left(p\left(\theta_0\right)^{1/2}\right)$$

$$\begin{aligned} B &= diag\left(p\left(\theta_0\right)^{-1/2}\right) Ldiag\left(p\left(\theta_0\right)^{1/2}\right) \\ &\quad - diag\left(p\left(\theta_0\right)^{-1/2}\right) Ldiag\left(p\left(\theta_0\right)^{1/2}\right)\sqrt{p\left(\theta_0\right)}\sqrt{p\left(\theta_0\right)^t} \\ &= diag\left(p\left(\theta_0\right)^{-1/2}\right) Ldiag\left(p\left(\theta_0\right)^{1/2}\right)\left(I - \sqrt{p\left(\theta_0\right)}\sqrt{p\left(\theta_0\right)^t}\right) \\ &= KS \end{aligned}$$

siendo

$$K = diag\left(p\left(\theta_0\right)^{-1/2}\right) Ldiag\left(p\left(\theta_0\right)^{1/2}\right).$$

Con lo que se llega a

$$T^* = S - KS - SK^t + KSK^t$$

siendo

$$\begin{aligned}
K &= diag\left(p\left(\theta_0\right)^{-1/2}\right) Ldiag\left(p\left(\theta_0\right)^{1/2}\right) \\
&= diag\left(p\left(\theta_0\right)^{-1/2}\right) J\left(\theta_0\right)\left(A\left(\theta_0\right)^t A\left(\theta_0\right)\right)^{-1} A\left(\theta_0\right)^t diag\left(p\left(\theta_0\right)^{-1}\right)^{1/2} diag\left(p\left(\theta_0\right)^{1/2}\right)
\end{aligned}$$

con

$$J\left(\theta_0\right) = \left(\left(\frac{\partial p_i\left(\theta\right)}{\partial \theta_r}\right)_{\substack{i=1,\ldots,M \\ r=1,\ldots,M_0}}\right).$$

Ahora bien,

$$\begin{aligned}
A\left(\theta_0\right) &= diag\left(p\left(\theta_0\right)^{-1/2}\right)\left(\left(\frac{\partial p_i\left(\theta\right)}{\partial \theta_r}\right)_{\substack{i=1,\ldots,M \\ r=1,\ldots,M_0}}\right) \\
&= diag\left(p\left(\theta_0\right)^{-1/2}\right) J\left(\theta_0\right)
\end{aligned}$$

con lo cual se tiene

$$K = A\left(\theta_0\right)\left(A\left(\theta_0\right)^t A\left(\theta_0\right)\right)^{-1} A\left(\theta_0\right)^t.$$

Como consecuencia

$$\begin{aligned}
T^* &= S - A\left(\theta_0\right)\left(A\left(\theta_0\right)^t A\left(\theta_0\right)\right)^{-1} A\left(\theta_0\right)^t S - SA\left(\theta_0\right)\left(A\left(\theta_0\right)^t A\left(\theta_0\right)\right)^{-1} A\left(\theta_0\right)^t \\
&\quad + A\left(\theta_0\right)\left(A\left(\theta_0\right)^t A\left(\theta_0\right)\right)^{-1} A\left(\theta_0\right)^t SA\left(\theta_0\right)\left(A\left(\theta_0\right)^t A\left(\theta_0\right)\right)^{-1} A\left(\theta_0\right)^t \\
&= I - \sqrt{p\left(\theta_0\right)}\sqrt{p\left(\theta_0\right)^t} - A\left(\theta_0\right)\left(A\left(\theta_0\right)^t A\left(\theta_0\right)\right)^{-1} A\left(\theta_0\right)^t\left(I - \sqrt{p\left(\theta_0\right)}\sqrt{p\left(\theta_0\right)^t}\right) \\
&\quad - \left(I - \sqrt{Q\left(\theta_0\right)}\sqrt{Q\left(\theta_0\right)^t}\right) A\left(\theta_0\right)\left(A\left(\theta_0\right)^t A\left(\theta_0\right)\right)^{-1} A\left(\theta_0\right)^t \\
&\quad + A\left(\theta_0\right)\left(A\left(\theta_0\right)^t A\left(\theta_0\right)\right)^{-1} A\left(\theta_0\right)^t\left(I - \sqrt{p\left(\theta_0\right)}\sqrt{p\left(\theta_0\right)^t}\right) \\
&\quad \times A\left(\theta_0\right)\left(A\left(\theta_0\right)^t A\left(\theta_0\right)\right)^{-1} A\left(\theta_0\right)^t.
\end{aligned}$$

Además

$$\sqrt{p\left(\theta_0\right)^t} A\left(\theta_0\right) = 0$$

entonces

$$T^* = I - \sqrt{p\left(\theta_0\right)}\sqrt{p\left(\theta_0\right)^t} - A\left(\theta_0\right)\left(A\left(\theta_0\right)^t A\left(\theta_0\right)\right)^{-1} A\left(\theta_0\right)^t.$$

Veamos que es idempotente

$$\begin{aligned}
\left(T^*\right)^2 &= I-\sqrt{p\left(\theta_0\right)}\sqrt{p\left(\theta_0\right)^t}-A\left(\theta_0\right)\left(A\left(\theta_0\right)^t A\left(\theta_0\right)\right)^{-1}A\left(\theta_0\right)^t \\
&\quad -\sqrt{p\left(\theta_0\right)}\sqrt{p\left(\theta_0\right)^t}+\sqrt{p\left(\theta_0\right)}\sqrt{p\left(\theta_0\right)^t}\sqrt{p\left(\theta_0\right)}\sqrt{p\left(\theta_0\right)^t} \\
&\quad +\sqrt{p\left(\theta_0\right)}\sqrt{p\left(\theta_0\right)^t}A\left(\theta_0\right)\left(A\left(\theta_0\right)^t A\left(\theta_0\right)\right)^{-1}A\left(\theta_0\right)^t \\
&\quad -A\left(\theta_0\right)\left(A\left(\theta_0\right)^t A\left(\theta_0\right)\right)^{-1}A\left(\theta_0\right)^t \\
&\quad +A\left(\theta_0\right)\left(A\left(\theta_0\right)^t A\left(\theta_0\right)\right)^{-1}A\left(\theta_0\right)^t\sqrt{p\left(\theta_0\right)}\sqrt{p\left(\theta_0\right)^t} \\
&\quad +A\left(\theta_0\right)\left(A\left(\theta_0\right)^t A\left(\theta_0\right)\right)^{-1}A\left(\theta_0\right)^t A\left(\theta_0\right)\left(A\left(\theta_0\right)^t A\left(\theta_0\right)\right)^{-1}A\left(\theta_0\right)^t
\end{aligned}$$

como,

$$A\left(\theta_0\right)\left(A\left(\theta_0\right)^t A\left(\theta_0\right)\right)^{-1}A\left(\theta_0\right)^t\sqrt{p\left(\theta_0\right)}\sqrt{p\left(\theta_0\right)^t}=0$$

y

$$\sqrt{p\left(\theta_0\right)^t}\sqrt{p\left(\theta_0\right)}=1$$

se tiene

$$\begin{aligned}
\left(T^*\right)^2 &= I-\sqrt{p\left(\theta_0\right)}\sqrt{p\left(\theta_0\right)^t}-A\left(\theta_0\right)\left(A\left(\theta_0\right)^t A\left(\theta_0\right)\right)^{-1}A\left(\theta_0\right)^t \\
&= T^*.
\end{aligned}$$

Al ser idempotente la matriz T^*, tiene únicamente los autovalores "0" y "1". Veamos el número de autovalores unitarios

$$\begin{aligned}
traza\left(T\right) &= traza\left(I\right)-traza\left(\sqrt{p\left(\theta_0\right)}\sqrt{p\left(\theta_0\right)^t}\right) \\
&\quad -traza\left(A\left(\theta_0\right)\left(A\left(\theta_0\right)^t A\left(\theta_0\right)\right)^{-1}A\left(\theta_0\right)^t\right) \\
&= M-1-traza\left(\left(A\left(\theta_0\right)^t A\left(\theta_0\right)\right)^{-1}A\left(\theta_0\right)^t A\left(\theta_0\right)\right) \\
&= M-1-M_0
\end{aligned}$$

ya que

$$\left(A\left(\theta_0\right)^t A\left(\theta_0\right)\right)^{-1}A\left(\theta_0\right)^t A\left(\theta_0\right)=I_{M_0\times M_0}.$$

■

Corolario 3.14

En el caso de las $WD_\varphi\left(\hat{p},p\left(\hat{\theta}_\phi\right)\right)$, *es decir, si* $\Lambda=1$, $\eta_1=1$, $h_{1_1}\left(x\right)=$

$h_{1_2}(x) = x$, $\varphi_1(x) = \varphi(x)$ y $\phi_1(x) = \phi(x)$, entonces se tiene

$$\frac{2nWD_\varphi\left(\hat{p}, p\left(\hat{\theta}_\phi\right)\right)}{\varphi''(1)} \underset{n\to\infty}{\overset{L}{\to}} \sum_{i=1}^{M} \beta_i Z_i^2$$

donde β_i son los autovalores de la matriz

$$T = D\left(\Sigma_{p(\theta_0)} - L\Sigma_{p(\theta_0)} - \Sigma_{p(\theta_0)}L^t + L\Sigma_{p(\theta_0)}L^t\right)$$

con

$$D = diag\left(\mathcal{U}_*\right)\frac{1}{E_{p(\theta_0)}(u)}$$

$$\Sigma_{p(\theta_0)} = diag\left(p(\theta_0)\right) - p(\theta_0)\, p(\theta_0)^t$$

$$L = \left(\left(\frac{\partial p_i(\theta)}{\partial \theta_r}\right)_{\substack{i=1,\dots,M\\ r=1,\dots,M_0}}\right)\left(A^*(\theta_0)^t A^*(\theta_0)\right)^{-1} A^*(\theta_0)^t\, diag\left(\mathcal{U}_*^{1/2}\right)$$

y Z_i son variables aleatorias normales e independientes de media cero y varianza uno.

Corolario 3.15

Si además $u_i = u$ $\forall i = 1, ..., M$, es decir en el caso de las φ-divergencias, entonces

$$\frac{2nD_\varphi\left(\hat{p}, p\left(\hat{\theta}_\phi\right)\right)}{\varphi''(1)} \underset{n\to\infty}{\overset{L}{\to}} \mathcal{X}^2_{M-M_0-1}.$$

Demostración

Análoga al Corolario 3.13. ■

3.5. Contraste de bondad de ajuste

Sea $X_1, ..., X_n$ una muestra aleatoria que se sospecha que procede de una población X con valores en el espacio $(\mathcal{X}, \beta_\chi, P)$. Sea $\mathcal{A} = \{A_1, ..., A_M\} \subset \beta_\chi$ una partición de $\mathcal{X}$, con

$$p_{j0} \equiv p_j(\theta_0) = P_{\theta_0}(A_j) \qquad 1 \leq j \leq M.$$

Es decir, la distribución de X se supone completamente conocida. Supuesto que sobre cada elemento A_j, $j = 1, ..., M$, de la partición $\mathcal{A}$ existe una ponderación u_j, $j = 1, ..., M$, se trata de contrastar

$$H_0 : p = p_0$$

con $p_0 \equiv p(\theta_0)$.

Así, el test basado en $WD_\varphi^h(\hat{p}, p_0)$ viene dado por

$$\phi_1(\hat{p}_1, ..., \hat{p}_M) = \begin{cases} 1 & si\ T_1 > t_\alpha \\ 0 & en\ otro\ caso \end{cases} \tag{3.4}$$

siendo

$$T_1 = 2nWD_\varphi^h(\hat{p}, p_0)$$

y t_α un número real tal que

$$P\left(\sum_{i=1}^{r} \beta_i Z_i^2 > t_\alpha\right) = \alpha$$

donde β_i y Z_i son los definidos en el Corolario 3.4.

Si se desea contrastar la hipótesis nula compuesta de que $H_0 : p = p(\theta)$, $\theta \in \Theta \subset R^{M_0}$, el test basado en $WD_\varphi^{h_1}\left(\hat{p}, p\left(\hat{\theta}_{h_2,\phi,w}\right)\right)$ viene dado por

$$\phi_2(\hat{p}_1, ..., \hat{p}_M) = \begin{cases} 1 & si\ T_2 > t_\alpha^* \\ 0 & en\ otro\ caso \end{cases} \tag{3.5}$$

siendo

$$T_2 = \frac{2nWD_\varphi^{h_1}\left(\hat{p}, p\left(\hat{\theta}_{h_2,\phi,w}\right)\right)}{\sum\limits_{a=1}^{\Lambda} \eta_a h'_{1_a}(0)\, \varphi''_a(1)}$$

con $\hat{\theta}_{h_2,\phi,w}$ el estimador de θ de mínima (h_2, ϕ)-divergencia ponderada y t_α^* un número real tal que

$$P\left(\sum_{i=1}^{r} \beta_i Z_i^2 > t_\alpha^*\right) = \alpha$$

donde β_i y Z_i están definidas en el Teorema 3.5.

Observación 3.4

Al realizar el procedimiento de contraste anterior uno se puede sentir preocupado al haber obtenido un procedimiento de contraste basado en una combinación lineal de Ji-cuadrados. Pero si uno lee cuidadosamente el trabajo de Rao y Scott (1981) esta preocupación desaparece ya que en él se presentan diversos, interesantes y precisos procedimientos para aproximar combinaciones lineales de Ji-cuadrados mediante una distribución Ji-cuadrado. Además como señalan Jensen y Solomon (1972) no debe uno de extrañarse al encontrarse un procedimiento inferencial que de lugar a una combinación lineal de Ji-cuadrados.

Ahora se analizarán diversas aproximaciones de combinaciones lineales de Ji-cuadrados mediante una Ji-cuadrado.

a) Si se considera el estadístico

$$(\beta^*)^{-1}\, 2nWD_\varphi^h\,(\hat{p}, p_0)$$

con $\beta^* = \max\{\beta_1, ..., \beta_r\}$, *se tiene*

$$(\beta^*)^{-1}\, 2nWD_\varphi^h\,(\hat{p}, p_0) \leq \sum_{i=1}^{r} Z_i^2.$$

Como la distribución de $\sum\limits_{i=1}^{r} Z_i^2$ *es una Ji-cuadrado con* r *grados de libertad, si rechazamos la hipótesis nula* $H_0 : p = p_0$ *si*

$$(\beta^*)^{-1}\, 2nWD_\varphi^h\,(\hat{p}, p_0) \geq \mathcal{X}_r^2$$

se tendrá un test conservativo. En esta aproximación únicamente se necesitará el mayor autovalor. En general β^* *será desconocido, pero no será difícil obtener un estimador consistente* $\hat{\beta}^*$ *de* β^*.

b) Si se considera el estadístico

$$\left(\bar{\beta}\right)^{-1} 2nWD_\varphi^h\,(\hat{p}, p_0)$$

con

$$\bar{\beta} = \frac{\sum\limits_{i=1}^{r} \beta_i}{r}$$

se tiene

$$\left(\bar{\beta}\right)^{-1} 2nWD_\varphi^h\,(\hat{p}, p_0) \leq \sum_{i=1}^{r} Z_i^2,$$

y por tanto se puede suponer que aproximadamente la distribución asintótica del estadístico $\left(\bar{\beta}\right)^{-1} 2nWD_{\varphi}^{h}\left(\hat{p},p_0\right)$ *es un Ji-cuadrado con* r *grados de libertad.*

Podemos observar lo siguiente:

$$\begin{aligned} E\left[\left(\bar{\beta}\right)^{-1} 2nWD_{\varphi}^{h}\left(\hat{p},p_0\right)\right] &= \left(\bar{\beta}\right)^{-1} \sum_{i=1}^{r} \beta_i E\left[Z_i^2\right] = r = E\left[\mathcal{X}_r^2\right] \\ Va\left[\left(\bar{\beta}\right)^{-1} 2nWD_{\varphi}^{h}\left(\hat{p},p_0\right)\right] &= \frac{2\sum\limits_{i=1}^{r}\beta_i}{\bar{\beta}^2} = 2r + \sum_{i=1}^{r} \frac{\left(\beta_i - \bar{\beta}\right)^2}{\bar{\beta}^2} 2r\left(1 + CV\left(\beta\right)^2\right) > Va\left[\mathcal{X}_r^2\right], \end{aligned}$$

donde $CV\left(\beta\right)$ *es el coeficiente de variación de las* β_i.

Luego con esta aproximación se consigue un estadístico que tiene igual esperanza que la Ji-cuadrado correspondiente, pero mayor varianza.

Si se denota por $\Lambda = diag\left(\beta_1, ..., \beta_r\right)$ *se tiene*

$$E\left[\sum_{i=1}^{r} \beta_i Z_i^2\right] = \sum_{i=1}^{r} \beta_i = traza\left(\Lambda\right) = traza\left(A\Sigma_p\right)$$

siendo A *la matriz diagonal obtenida en el Corolario 3.4 y que viene dada por*

$$A = \left(a_{ij}\right)_{i,j=1,\dots,M} = \begin{cases} 0 & si \quad i \neq j \\ \sum\limits_{a=1}^{\Lambda} \eta_a h_a'\left(0\right) \varphi_a''\left(1\right) \dfrac{u_i}{p_{i0} \sum\limits_{i=1}^{M} u_i p_{i0}} & si \quad i = j \end{cases}.$$

Por tanto,

$$\begin{aligned} \sum_{i=1}^{r} \beta_i &= \left(\sum_{a=1}^{\Lambda} \eta_a h_a'\left(0\right) \varphi_a''\left(1\right)\right) \sum_{i=1}^{r} p_{i0}\left(1 - p_{i0}\right) \frac{u_i}{p_{i0} \sum\limits_{i=1}^{M} u_i p_{i0}} \\ &= \left(\sum_{a=1}^{\Lambda} \eta_a h_a'\left(0\right) \varphi_a''\left(1\right)\right) \left(\frac{\sum\limits_{i=1}^{r} u_i\left(1 - p_{i0}\right)}{\sum\limits_{i=1}^{M} u_i p_{i0}}\right). \end{aligned}$$

En consecuencia se podrá calcular $\bar{\beta}$, mediante

$$\bar{\beta} = \frac{\sum_{a=1}^{\Lambda} \eta_a h'_a(0)\, \varphi''_a(1)}{r} \left(\frac{\sum_{i=1}^{r} u_i (1-p_{i0})}{\sum_{i=1}^{M} u_i p_{i0}} \right).$$

En el procedimiento de aproximación anterior se consideraba un estadístico con la misma esperanza que la Ji-cuadrado a la que se le aproximaba. En el siguiente procedimiento no solamente el estadístico tendrá la misma esperanza sino que también tendrá la misma varianza.

c) Si se considera el estadístico

$$\left(\bar{\beta}\right)^{-1} \left(1+\lambda^2\right)^{-1} 2nWD_{\varphi}^{h}(\hat{p}, p_0)$$

y se le aproxima por una $\mathcal{X}_v^2$ *siendo* $v = \frac{r}{1+\lambda^2}$ *y* $\lambda^2 = \sum_{i=1}^{r} \frac{\left(\beta_i - \bar{\beta}\right)^2}{r\bar{\beta}^2}$ *se tiene que*

$$E\left[\left(\bar{\beta}\right)^{-1} \left(1+\lambda^2\right)^{-1} 2nWD_{\varphi}^{h}(\hat{p}, p_0)\right] = \frac{1}{1+\lambda^2} \frac{r}{\sum_{i=1}^{r} \beta_i} \sum_{i=1}^{r} \beta_i = \frac{r}{1+\lambda^2} = E\left[\mathcal{X}_v^2\right]$$

$$Va\left[\left(\bar{\beta}\right)^{-1} \left(1+\lambda^2\right)^{-1} 2nWD_{\varphi}^{h}(\hat{p}, p_0)\right] = \frac{1}{\bar{\beta}^2} \frac{1}{\left(1+\lambda^2\right)^2} 2\sum_{i=1}^{r} \beta_i^2 = \frac{2r}{1+\lambda^2} = Va\left[\mathcal{X}_v^2\right].$$

Al igual que en las situaciones anteriores será necesario calcular $\bar{\beta}$ y λ^2. La forma de calcular $\bar{\beta}$ ya se estableció. En cuanto a la forma de estimar λ^2 será suficiente con ser capaces de calcular $\sum_{i=1}^{r} \beta_i^2$. Al ser ,

$$\begin{aligned}
\sum_{i=1}^{r} \beta_i^2 &= traza\left(\Lambda^2\right) = traza\left(\left(A\Sigma_p\right)^2\right) \\
&= \frac{\left(\sum_{a=1}^{\Lambda} \eta_a h'_a(0)\, \varphi''_a(1)\right)^2}{\left(\sum_{i=1}^{M} u_i p_{i0}\right)^2} \left(\sum_{i=1}^{r} u_i^2 (1-2p_{i0})^2 + \left(\sum_{i=1}^{r} u_i p_{i0} \right)^2 \right)
\end{aligned}$$

se tiene que $\sum_{i=1}^{r} \beta_i^2$ *viene dado por*

$$\frac{\left(\sum_{a=1}^{\Lambda} \eta_a h_a'(0)\, \varphi_a''(1)\right)^2}{\left(\sum_{i=1}^{M} u_i p_{i0}\right)^2} \left(\sum_{i=1}^{r} u_i^2 (1-2p_{i0})^2 + \left(\sum_{i=1}^{r} u_i p_{i0}\right)^2\right).$$

d) Para valores pequeños de k está tabulada la función de distribución de la variable aleatoria

$$\sum_{i=1}^{k} a_i Z_i^2$$

Estas tabulaciones se pueden encontrar en Solomon (1960), Johnson y Kotz (1968), Jensen y Solomon (1972), Eckler (1969) y Gupta (1963).

Un estudio similar se puede llevar a cabo cuando se considera una hipótesis nula compuesta.

3.5.1. Potencia del contraste

Sea $p^* = (p_1^*, ..., p_M^*)^t \neq p_0 = (p_{10}, ..., p_{M0})^t$ entonces la potencia del test en $p^* = (p_1^*, ..., p_M^*)$ dado en (3.4) viene dada por

$$\beta_n(p_1^*, ..., p_M^*) = P_{F^*}(T_1 > t_\alpha) = P_{F^*}\left(2nWD_\varphi^h(\hat{p}, p_0) > t_\alpha\right).$$

Ahora bien, como

$$n^{\frac{1}{2}}\left(WD_\varphi^h(\hat{p}, p_0) - WD_\varphi^h(p^*, p_0)\right) \underset{n\to\infty}{\overset{L}{\to}} N\left(0, \sigma_p^2\right)$$

se tiene

$$\begin{aligned}
\beta_n(p_1^*, ..., p_M^*) &= P\left(n^{1/2} WD_\varphi^h(\hat{p}, p_0) > \frac{t_\alpha}{2n^{1/2}}\right) \\
&= P\left(\frac{n^{\frac{1}{2}}\left(WD_\varphi^h(\hat{p}, p_0) - WD_\varphi^h(p^*, p_0)\right)}{\sigma_{p^*}} > \frac{\frac{t_\alpha}{2n^{1/2}} - n^{1/2} WD_\varphi^h(p^*, p_0)}{\sigma_{p^*}}\right).
\end{aligned}$$

Es decir,

$$\beta_n(p_1^*, ..., p_M^*) = 1 - \Phi_n\left(\frac{t_\alpha - 2nWD_\varphi^h(p^*, p_0)}{2n^{1/2}\sigma_{p^*}}\right) \tag{3.6}$$

donde $\Phi_n(x)$ es una sucesión de funciones de distribución que tiende uniformemente a la función de distribución, Φ, de una normal de media cero y varianza uno y σ_{p^*} es la expresión definida en el Teorema 3.1 para $p^* = (p_1^*, ..., p_M^*)^t$.

Además

$$\lim_{n\to\infty} \beta_n(p_1^*, ..., p_M^*) = 1$$

por lo que el test obtenido es asintóticamente consistente en el sentido de Fraser.

Ahora se calculará una aproximación a la función de potencia para el contraste de hipótesis nula compuesta dado en (3.5). Sea $q \notin \{p(\theta) : \theta \in \Theta\}$. Supuesto que la hipótesis alternativa q es cierta, se tiene

$$\hat{p} \underset{n\to\infty}{\overset{P}{\to}} q.$$

Supongamos que existe

$$\theta_a = \arg\inf_{\theta\in\Theta} WD_\phi^h(q, p(\theta))$$

verificando que

$$p\left(\hat{\theta}_{h,\phi,w}\right) \underset{n\to\infty}{\overset{c.s.}{\to}} p(\theta_a)$$

y además

$$\sqrt{n}\left(\hat{p} - p\left(\hat{\theta}_{h,\phi,w}\right)\right) \underset{n\to\infty}{\overset{L}{\to}} N(0, \Sigma)$$

donde

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \ \Sigma_{11} = diag(q) - q^t q, \ \Sigma_{12} = \Sigma_{21}.$$

En estas condiciones se tiene el siguiente resultado.

Teorema 3.7

Bajo las hipótesis dadas en el Teorema 3.6, se verifica que

$$\sqrt{n}\left(WD_\varphi^h\left(\hat{p}, p\left(\hat{\theta}_{h,\phi,w}\right)\right) - WD_\varphi^h(q, p(\theta_a))\right)$$

converge en ley a una normal de media cero y varianza σ^2 *dada por*

$$\sigma^2 = T\Sigma_{11}T^t + T\Sigma_{12}S^t + S\Sigma_{21}T^t + S\Sigma_{22}S^t \tag{3.7}$$

siendo

$$T = (t_1, ..., t_M)^t \ \ y \ S = (s_1, ..., s_M)^t$$

con

$$t_i = \left(\frac{\partial}{\partial q_i} WD_\varphi^h (q, p) \right)_{p=p(\theta_a)}, \ i = 1, ..., M$$

y

$$s_i = \left(\frac{\partial}{\partial p_i} WD_\varphi^h (q, p) \right)_{p=p(\theta_a)}, \ i = 1, ..., M.$$

Demostración

El desarrollo en serie de Taylor de primer orden da

$$\begin{aligned} WD_\varphi^h \left(\hat{p}, p\left(\hat{\theta}_{h,\phi,w}\right)\right) &= WD_\varphi^h (q, p(\theta_a)) + T(\hat{p} - q) + S\left(p\left(\hat{\theta}_{h,\phi,w}\right) - p(\theta_a)\right)^t \\ &\quad + o\left(\|\hat{p} - q\| + \left\| \hat{p}\left(\hat{\theta}_{h,\phi,w}\right) - p(\theta_a)\right\| \right). \end{aligned}$$

Ahora el resultado se sigue de forma inmediata de las hipótesis del teorema. ■

Haciendo uso de este teorema se tiene que la potencia del test (3.5) viene dada por

$$\beta_n (q_1, ..., q_M) = 1 - \Phi_n \left(\frac{t_\alpha^* - 2nWD_\varphi^h (q, p(\theta_a))}{2n^{1/2}\sigma} \right) \tag{3.8}$$

donde σ viene dada en (3.7) y $\Phi_n(x)$ es una sucesión de funciones de distribución que tiende a la función de distribución, Φ, de una normal de media cero y varianza uno.

Observación 3.5

Las aproximaciones obtenidas de las funciones de potencia permiten dar un procedimiento para la determinación del tamaño muestral. Para fijar ideas se considerará la función de potencia dada en (3.6) *para el test estadístico considerado en* (3.4)*. La cuestión que se plantea ahora es ¿Cómo utilizar la función de potencia para la determinación el tamaño muestral? Sea* $p^* = (p_1^*, ..., p_M^*)^t$ *un punto de la alternativa de forma que en él se desea tener una potencia igual a* β^**. El tamaño muestral,* n^**, necesario para alcanzar en* p^* *una potencia* β^**, se obtendrá como solución de la ecuación*

$$\beta^* = 1 - \Phi_{N(0,1)} \left(\frac{t_\alpha - 2nWD_\varphi^h (p^*, p_0)}{2n^{1/2}\sigma_{p^*}} \right).$$

Luego el n^* deberá ser solución de la ecuación

$$4n^2 WD_\varphi^h(p^*,p_0) - 4n\left(t_\alpha WD_\varphi^h(p^*,p_0) + \sigma_{p^*}^2 \Phi_{N(0,1)}^{-1}(1-\beta^*)^2\right) + t_\alpha^2 = 0$$

y esta viene dada por

$$\begin{aligned} n^* &= \left[\frac{4\left(t_\alpha WD_\varphi^h(p^*,p_0) + \sigma_{p^*}^2 \Phi_{N(0,1)}^{-1}(1-\beta^*)^2\right)}{8WD_\varphi^h(p^*,p_0)} \right. \\ &\left. + \frac{\sqrt{\left(t_\alpha WD_\varphi^h(p^*,p_0) + \sigma_{p^*}^2 \Phi_{N(0,1)}^{-1}(1-\beta^*)^2\right)^2 - 16WD_\varphi^h(p^*,p_0)\, t_\alpha^2}}{8WD_\varphi^h(p^*,p_0)} \right] + 1 \end{aligned}$$

donde $[x]$ denota la parte entera de x.

Un estudio análogo se podría realizar con la función de potencia dada en (3.8) para el estadístico considerado en (3.5).

Una aproximación, distinta a la dada en (3.6), se puede obtener para la función de potencia del test estadístico definido en (3.4) si se considera el resultado obtenido en el Teorema 3.3. Así dado $p^{(n)} = p_0 + n^{-1/2} d$ con $d = (d_1, ..., d_M)^t$ y $\sum_{i=1}^{M} d_i = 0$, se tiene

$$\beta\left(p^{(n)}\right) = P\left(T_1 > t_\alpha / H_{1,n}\right)$$

donde $H_{1,n}$ está definido en (3.2). Luego

$$\beta\left(p^{(n)}\right) = 1 - G_n(t_\alpha)$$

para una sucesión de funciones de distribución $G_n(x)$ que tienden uniformemente a la función de distribución de la variable aleatoria

$$\sum_{i=1}^{r} \beta_i (Z_i + w_i)^2 .$$

De la misma forma se puede proceder para el test estadístico definido en (3.5).

3.6. Otros contrastes de interés

A continuación se muestran otros contrastes de hipótesis que se pueden realizar basándose en los resultados asintóticos obtenidos previamente para las (h, φ)-divergencias ponderadas.

3.6.1. La $WD_{\varphi}^{h}(p, p_0)$ es un valor prefijado

Se trata de contrastar que la (h, φ)-divergencia ponderada entre las distribuciones p y p_0 es un valor prefijado W_0, supuesto que p_0 es conocida, es decir

$$\begin{cases} H_0 : & WD_{\varphi}^{h}(p, p_0) = W_0 \\ H_1 : & WD_{\varphi}^{h}(p, p_0) \neq W_0 \end{cases}$$

con $W_0 \neq 0$.

El test viene definido por

$$\phi_3(\hat{p}_1, ..., \hat{p}_M) = \begin{cases} 1 & si \;\; |T_3| > z_{\alpha/2} \\ 0 & en\ otro\ caso \end{cases}$$

donde

$$T_3 = \frac{n^{1/2}\left[WD_{\varphi}^{h}(\hat{p}, p_0) - W_0\right]}{\hat{\sigma}_p}$$

siendo $\hat{\sigma}_p$ el valor de σ_p dado en el Teorema 3.1 para $\hat{p}$ y z_α es un número real tal que

$$P(N(0,1) > z_\alpha) = \alpha.$$

Obsérvese que por el Teorema 3.1,

$$\frac{n^{1/2}\left[WD_{\varphi}^{h}(\hat{p}, p_0) - W_0\right]}{\sigma_p} \underset{n\to\infty}{\overset{L}{\to}} N(0,1)$$

y por otro lado

$$\hat{\sigma}_p \underset{n\to\infty}{\overset{P}{\to}} \sigma_p$$

luego aplicando el teorema de Slutsky (ver Ferguson 1996, Capítulo 6), se tiene que

$$T_3 \underset{n\to\infty}{\overset{L}{\to}} N(0,1).$$

En este contexto, un intervalo de confianza a nivel $1-\alpha$ para $WD_{\varphi}^{h}(p,p_0)$ viene dado por

$$WD_{\varphi}^{h}(p,p_0) \pm z_{\alpha/2}\frac{\hat{\sigma}_p}{n^{1/2}}.$$

3.6.2. Las $WD_{\varphi}^{h}(p,p_0)$ entre r pares de distribuciones coinciden con un valor prefijado

Sean $p_j,\ p_{j0} \in \Delta_M$ para $j=1,....,r$. Se trata de contrastar que las (h,φ)-divergencias ponderadas entre r pares (p_j,p_{j0}); $j=1,....,r$, son iguales a un valor W_0 supuesto que las p_{j0} son conocidas, es decir

$$H_0 : WD_{\varphi}^{h}(p_1,p_{10}) = ... = WD_{\varphi}^{h}(p_r,p_{r0}) = W_0$$

con W_0 conocido, frente a la alternativa $H_1 : \exists j,k \in \{1,....,r\}$ de tal forma que $WD_{\varphi}^{h}(p_j,p_{j0}) \neq WD_{\varphi}^{h}(p_k,p_{k0})$. Se consideran r muestras independientes y a partir de ellas $\hat{p}_j,\ j=1,...,r$.

Se sabe que

$$\frac{\sqrt{n_j}\left[WD_{\varphi}^{h}(\hat{p}_j,p_{j0}) - W_0\right]}{\hat{\sigma}_{p_j}} \underset{n\to\infty}{\overset{L}{\to}} N(0,1)\ \ \forall j \in \{1,....,r\}.$$

Por tanto

$$\frac{n_j\left[WD_{\varphi}^{h}(\hat{p}_j,p_{j0}) - W_0\right]^2}{\hat{\sigma}_{p_j}^2} \underset{n\to\infty}{\overset{L}{\to}} \mathcal{X}_1^2$$

y por consiguiente

$$\sum_{j=1}^{r}\frac{n_j\left[WD_{\varphi}^{h}(\hat{p}_j,p_{j0}) - W_0\right]^2}{\hat{\sigma}_{p_j}^2} \underset{n\to\infty}{\overset{L}{\to}} \mathcal{X}_r^2.$$

Así, si se quiere contrastar

$$H_0 : WD_{\varphi}^{h}(p_1,p_{1j0}) = ... = WD_{\varphi}^{h}(p_r,p_{r0}) = W_0$$

el test vendrá dado por

$$\phi_4\left(\hat{p}_1^{(1)},....,\hat{p}_M^{(1)},....,\hat{p}_1^{(r)},....,\hat{p}_M^{(r)}\right) = \begin{cases} 1 & si\ T_4 > \mathcal{X}_{r,\alpha}^2 \\ 0 & en\ otro\ caso \end{cases}$$

siendo

$$T_4 = \sum_{j=1}^{r} \frac{n_j \left[WD_\varphi^h\left(\hat{p}_j, p_{j0}\right) - W_0\right]^2}{\hat{\sigma}_{p_j}^2}$$

donde $\hat{\sigma}_{p_j}^2$ son las expresiones de la varianza σ_p^2 dadas en el Teorema 3.1 cuando p_i se sustituye por $\hat{p}_i^{(j)}$; $i = 1, ..., M$, $j = 1, ..., r$ y $\mathcal{X}_{r,\alpha}^2$ es un número real positivo verificando $P\left(\mathcal{X}_r^2 > \mathcal{X}_{r,\alpha}^2\right) = \alpha$.

3.6.3. Las $WD_\varphi^h\left(p_j, p_{j0}\right)$ entre r pares de distribuciones coinciden

Si se considera que las p_{j0} son conocidas; $j = 1, ..., r$, se trata, en este caso, de realizar el siguiente contraste

$$H_0 : WD_\varphi^h\left(p_1, p_{10}\right) = ... = WD_\varphi^h\left(p_r, p_{r0}\right)$$

a partir de r muestras independientes.

El test, entonces vendrá dado por

$$\phi_5\left(\hat{p}_1^{(1)}, ..., \hat{p}_M^{(1)}, ..., \hat{p}_1^{(r)}, ..., \hat{p}_M^{(r)}\right) = \begin{cases} 1 & si\ T_5 > \mathcal{X}_{r-1,\alpha}^2 \\ 0 & en\ otro\ caso \end{cases}$$

siendo

$$T_5 = \sum_{j=1}^{r} \frac{n_j \left[WD_\varphi^h\left(\hat{p}_j, p_{j0}\right) - \bar{W}\right]^2}{\hat{\sigma}_{p_j}^2}$$

donde

$$\bar{W} = \left(\sum_{j=1}^{r} \frac{n_j}{\hat{\sigma}_{p_j}^2}\right)^{-1} \sum_{j=1}^{r} \frac{n_j WD_\varphi^h\left(\hat{p}_j, p_{j0}\right)}{\hat{\sigma}_{p_j}^2}$$

y las $\hat{\sigma}_{p_j}^2$ se obtienen como el contraste anterior.

A continuación se verá que la distribución asintótica del estadístico T_5 es una Ji-cuadrado con r-1 grados de libertad.

$$\begin{aligned}
\sum_{j=1}^{r} \frac{n_j \left[WD_\varphi^h\left(\hat{p}_j, p_{j0}\right) - W_0\right]^2}{\hat{\sigma}_{p_j}^2} &= \sum_{j=1}^{r} \frac{n_j \left[WD_\varphi^h\left(\hat{p}_j, p_{j0}\right) - \bar{W} + \bar{W} - W_0\right]^2}{\hat{\sigma}_{p_j}^2} \\
&= \sum_{j=1}^{r} \frac{n_j \left[WD_\varphi^h\left(\hat{p}_j, p_{j0}\right) - \bar{W}\right]^2}{\hat{\sigma}_{p_j}^2} + \sum_{j=1}^{r} \frac{n_j}{\hat{\sigma}_{p_j}^2} \left(\bar{W} - W_0\right)^2
\end{aligned}$$

ya que el doble producto se anula sin más que considerar la definición de $\bar{W}$.

Por un lado se tiene que el primer término de la igualdad se distribuye asintóticamente como una $\mathcal{X}_r^2$.

Por otro, se verá que el segundo término del lado derecho de la igualdad sigue una distribución $\mathcal{X}_1^2$.

Se sabe que

$$\frac{\sqrt{n_j}\left[WD_{\varphi}^h(\hat{p}_j, p_{j0}) - W_0\right]}{\hat{\sigma}_{p_j}} \underset{n\to\infty}{\overset{L}{\to}} N(0,1) \;\; \forall j \in \{1, ..., r\}$$

por tanto

$$\begin{array}{ccc}
\dfrac{\sqrt{n}WD_{\phi}^h\left(\hat{P}_j, Q_j\right)}{\hat{\sigma}_{p_j}} & \approx & N\left(\dfrac{\sqrt{n_j}W_0}{\hat{\sigma}_{p_j}}, 1\right) \\
\dfrac{\sqrt{n_j}}{\hat{\sigma}_{p_j}}\dfrac{\sqrt{n_j}WD_{\varphi}^h(\hat{p}_j, p_{j0})}{\hat{\sigma}_{p_j}} & \approx & N\left(\dfrac{n_jW_0}{\hat{\sigma}_{p_j}^2}, \dfrac{n_j}{\hat{\sigma}_{p_j}^2}\right) \\
\sum\limits_{j=1}^{r}\dfrac{n_jWD_{\varphi}^h(\hat{p}_j, p_{j0})}{\hat{\sigma}_{p_j}^2} & \approx & N\left(W_0\sum\limits_{j=1}^{r}\dfrac{n_j}{\hat{\sigma}_{p_j}^2}, \sum\limits_{j=1}^{r}\dfrac{n_j}{\hat{\sigma}_{p_j}^2}\right) \\
\sum\limits_{j=1}^{r}\dfrac{n_j}{\hat{\sigma}_{p_j}^2}\bar{W} = \sum\limits_{j=1}^{r}\dfrac{n_jWD_{\varphi}^h(\hat{p}_j, p_{j0})}{\hat{\sigma}_{p_j}^2} & \approx & N\left(W_0\sum\limits_{j=1}^{r}\dfrac{n_j}{\hat{\sigma}_{p_j}^2}, \sum\limits_{j=1}^{r}\dfrac{n_j}{\hat{\sigma}_{p_j}^2}\right) \\
\sqrt{\sum\limits_{j=1}^{r}\dfrac{n_j}{\hat{\sigma}_{p_j}^2}}\bar{W} & \approx & N\left(W_0\sqrt{\sum\limits_{j=1}^{r}\dfrac{n_j}{\hat{\sigma}_{p_j}^2}}, 1\right).
\end{array}$$

Entonces,

$$\sqrt{\sum_{j=1}^{r}\frac{n_j}{\hat{\sigma}_{p_j}^2}}\left(\bar{W} - W_0\right) \underset{n\to\infty}{\overset{L}{\to}} N(0,1)$$

y

$$\sum_{j=1}^{r}\frac{n_j}{\hat{\sigma}_{p_j}^2}\left(\bar{W} - W_0\right)^2 \underset{n\to\infty}{\overset{L}{\to}} \mathcal{X}_1^2.$$

Es decir, se tiene

$$\underbrace{\sum_{j=1}^{r}\frac{n_j\left[WD_{\varphi}^h(\hat{p}_j, p_{j0}) - W_0\right]^2}{\hat{\sigma}_{p_j}^2}}_{\mathcal{X}_r^2} = \sum_{j=1}^{r}\frac{n_j\left[WD_{\varphi}^h(\hat{p}_j, p_{j0}) - \bar{W}\right]^2}{\hat{\sigma}_{p_j}^2} + \underbrace{\sum_{j=1}^{r}\frac{n_j}{\hat{\sigma}_{p_j}^2}\left(\bar{W} - W_0\right)^2}_{\mathcal{X}_1^2}.$$

Si las r muestras son independientes, se puede aplicar el siguiente resultado del Teorema de Cochran; (Rao 1973, pág. 187. Resultado iii)):

Sea $Y^tY = Y^tAY + Y^tBY$ con Y normal r-dimensional con vector de medias cero y matriz de varianzas covarianzas la identidad e Y^tAY distribuida como una Ji-cuadrado con a grados de libertad, entonces Y^tBY se distribuye como una Ji-cuadrado con r-a grados de libertad.

Se aplicará este resultado para establecer que

$$T_5 = \sum_{j=1}^{r} \frac{n_j \left[WD_\varphi^h\left(\hat{p}_j, p_{j0}\right) - \bar{W}\right]^2}{\hat{\sigma}_{p_j}^2}$$

sigue una distribución Ji-cuadrado con r-1 grados de libertad.

En efecto, si

$$Y_j = \frac{\sqrt{n_j}\left[WD_\varphi^h\left(\hat{p}_j, p_{j0}\right) - W_0\right]}{\hat{\sigma}_{p_j}}$$

e $Y = (Y_1, ..., Y_r)^t$. Basta comprobar que

$$Y^tAY = \sum_{j=1}^{r} \frac{n_j}{\hat{\sigma}_{p_j}^2}\left(\bar{W} - W_0\right)^2$$

con $rango\,(A) = 1$ y que

$$Y^tBY = \sum_{j=1}^{r} \frac{n_j \left[WD_\varphi^h\left(\hat{p}_j, p_{j0}\right) - \bar{W}\right]^2}{\hat{\sigma}_{p_j}^2}.$$

Como

$$\bar{W} = \frac{1}{a}\sum_{j=1}^{r} \frac{n_j WD_\varphi^h\left(\hat{p}_j, p_{j0}\right)}{\hat{\sigma}_{p_j}^2}$$

con

$$a = \sum_{j=1}^{r} \frac{n_j}{\hat{\sigma}_{p_j}^2}$$

se tiene que

$$\sum_{j=1}^{r} \frac{n_j}{\hat{\sigma}_{p_j}^2}\left(\bar{W} - W_0\right)^2 \quad = \quad a\left(\bar{W} - W_0\right)^2$$

$$\begin{aligned}
&= a\left(\frac{1}{a}\sum_{j=1}^{r}\frac{n_j WD_\varphi^h\left(\hat{p}_j, p_{j0}\right)}{\hat{\sigma}_{p_j}^2} - W_0\right)^2 \\
&= \frac{1}{a}\left(\sum_{j=1}^{r}\frac{n_j WD_\varphi^h\left(\hat{p}_j, p_{j0}\right)}{\hat{\sigma}_{p_j}^2} - \sum_{j=1}^{r}\frac{n_j}{\hat{\sigma}_{p_j}^2}W_0\right)^2 \\
&= \frac{1}{a}\left(\sum_{j=1}^{r}\frac{n_j WD_\varphi^h\left(\hat{p}_j, p_{j0}\right) - W_0}{\hat{\sigma}_{p_j}^2}\right)^2 \\
&= \frac{1}{a}\left(\sum_{j=1}^{r}\frac{\sqrt{n_j}}{\hat{\sigma}_{p_j}}Y_j\right)^2 \\
&= \frac{1}{a}\sum_{j=1}^{r}\sum_{i=1}^{r}\frac{\sqrt{n_j}}{\hat{\sigma}_{p_j}}\frac{\sqrt{n_i}}{\hat{\sigma}_{pi}}Y_jY_i = Y^tAY
\end{aligned}$$

donde $A = (a_{ij})_{i,j=1,....,r}$ y

$$a_{ij} = \left(\frac{\sqrt{n_j}}{\hat{\sigma}_{p_j}}\frac{\sqrt{n_i}}{\hat{\sigma}_{pi}}\right)\left(\sum_{j=1}^{r}\frac{n_j}{\hat{\sigma}_{p_j}^2}\right)^{-1}.$$

Además $\forall i, j, k, l = 1, ..., r$

$$\begin{aligned}
\begin{vmatrix} a_{ij} & a_{ik} \\ a_{lj} & a_{lk} \end{vmatrix} &= a_{ij}a_{lk} - a_{ik}a_{lj} \\
&= \left(\frac{\sqrt{n_j}}{\hat{\sigma}_{p_j}}\frac{\sqrt{n_i}}{\hat{\sigma}_{pi}}\frac{\sqrt{n_l}}{\hat{\sigma}_{pl}}\frac{\sqrt{n_k}}{\hat{\sigma}_{pk}} - \frac{\sqrt{n_k}}{\hat{\sigma}_{pk}}\frac{\sqrt{n_i}}{\hat{\sigma}_{pi}}\frac{\sqrt{n_j}}{\hat{\sigma}_{p_j}}\frac{\sqrt{n_l}}{\hat{\sigma}_{pl}}\right)a^{-2} = 0
\end{aligned}$$

lo que significa que el $rango\,(A) = 1$.

Finalmente, $B = I - A$ y aplicando el corolario anteriormente citado se tiene en definitiva que

$$Y^tBY \approx \mathcal{X}_{r-1}^2.$$

En conclusión, se ha visto lo que se quería

$$T_5 = \sum_{j=1}^{r}\frac{n_j\left[WD_\varphi^h\left(\hat{p}_j, p_{j0}\right) - \bar{W}\right]^2}{\hat{\sigma}_{p_j}^2} \underset{n\to\infty}{\overset{L}{\to}} \mathcal{X}_{r-1}^2.$$

3.7. Contrastes para datos binomiales

Un caso especialmente importante se presenta cuando $M = 2$, es decir si se tienen únicamente dos clases. En esta situación, entonces se tiene el siguiente resultado.

Teorema 3.8

Sea $M = 2$ y supóngase que se verifican las hipótesis requeridas a las funciones h_a y φ_a, $a = 1, ..., \Lambda$ en (1.15). Bajo la hipótesis que $p = p_0$, se verifica

$$\frac{2nWD_{\varphi}^{h}\left(\hat{p}, p_0\right)}{\left(\sum\limits_{a=1}^{\Lambda} \eta_a h_a'\left(0\right) \varphi_a''\left(1\right)\right)\left(u_1\left(1-p\right)+u_2 p\right)\left(u_1 p + u_2\left(1-p\right)\right)^{-1}} \underset{n\to\infty}{\overset{L}{\to}} \mathcal{X}_1^2.$$

Demostración

Si $p = p_0$ y $M = 2$ se tiene a partir del Corolario 3.4 que

$$a_{ij} = \begin{cases} 0 & si \quad i \neq j \\ \sum\limits_{a=1}^{\Lambda} \eta_a h_a'\left(0\right) \varphi_a''\left(1\right) \dfrac{u_i}{p_i \sum\limits_{i=1}^{2} u_i p_i} & si \quad i = j \end{cases}, \ i,j = 1,2$$

es decir,

$$A = \begin{pmatrix} \sum\limits_{a=1}^{\Lambda} \eta_a h_a'\left(0\right) \varphi_a''\left(1\right) \dfrac{u_1}{p_{10} \sum\limits_{i=1}^{2} u_i p_{i0}} & 0 \\ 0 & \sum\limits_{a=1}^{\Lambda} \eta_a h_a'\left(0\right) \varphi_a''\left(1\right) \dfrac{u_2}{p_{20} \sum\limits_{i=1}^{2} u_i p_{i0}} \end{pmatrix}$$

$$= \sum_{a=1}^{\Lambda} \eta_a h_a'\left(0\right) \varphi_a''\left(1\right) \begin{pmatrix} \dfrac{u_1}{p_0\left(u_1 p_0 + u_2\left(1-p_0\right)\right)} & 0 \\ 0 & \dfrac{u_2}{\left(1-p_0\right)\left(u_1 p_0 + u_2\left(1-p_0\right)\right)} \end{pmatrix},$$

habiéndose denotado $p_{10} = p_0$ y $p_{20} = 1 - p_0$.

Se sabe por el Corolario 3.4 que,

$$2nWD_{\varphi}^{h}\left(\hat{p}, p_0\right) \underset{n\to\infty}{\overset{L}{\to}} \sum_{i=1}^{r} \beta_i Z_i^2$$

donde las Z_i son variables aleatorias normales independientes de media cero y varianza uno, los β_i son los autovalores no nulos de $A\Sigma_{p_0}$ y $r = rango\,(\Sigma_{p_0}A\Sigma_{p_0})$.

En este caso se tiene,

$$\Sigma_{p_0} = \begin{pmatrix} p_0\,(1-p_0) & -p_0\,(1-p_0) \\ -p_0\,(1-p_0) & p_0\,(1-p_0) \end{pmatrix},$$

con lo cual

$$\begin{aligned} A\Sigma_{p_0} &= L \begin{pmatrix} \dfrac{u_1}{p_0\,(u_1p_0 + u_2\,(1-p_0))} & 0 \\ 0 & \dfrac{u_2}{(1-p_0)\,(u_1p_0 + u_2\,(1-p_0))} \end{pmatrix} \Sigma_{p_0} \\ &= L \begin{pmatrix} \dfrac{u_1\,(1-p_0)}{u_1p_0 + u_2\,(1-p_0)} & \dfrac{-u_1\,(1-p_0)}{u_1p_0 + u_2\,(1-p_0)} \\ \dfrac{-u_2p_0}{u_1p_0 + u_2\,(1-p_0)} & \dfrac{u_2p_0}{u_1p_0 + u_2\,(1-p_0)} \end{pmatrix} \end{aligned}$$

con $L = \sum_{a=1}^{\Lambda} \eta_a h'_a\,(0)\,\varphi''_a\,(1)$.

Los autovalores β_i de $A\Sigma_{p_0}$ son los mismos que los autovalores λ_i de $C\Sigma_{p_0}$ multiplicados por $\sum_{a=1}^{\Lambda} \eta_a h'_a\,(0)\,\varphi''_a\,(1)$, donde

$$C\Sigma_{p_0} = \begin{pmatrix} \dfrac{u_1\,(1-p_0)}{u_1p_0 + u_2\,(1-p_0)} & \dfrac{-u_1\,(1-p_0)}{u_1p_0 + u_2\,(1-p_0)} \\ \dfrac{-u_2p_0}{u_1p_0 + u_2\,(1-p_0)} & \dfrac{u_2p_0}{u_1p_0 + u_2\,(1-p_0)} \end{pmatrix}.$$

Entonces si se calculan los autovalores λ_i se tiene

$$\begin{aligned} |C\Sigma_{p_0} - \lambda I| &= \begin{vmatrix} \dfrac{u_1\,(1-p_0)}{u_1p_0 + u_2\,(1-p_0)} - \lambda & \dfrac{-u_1\,(1-p_0)}{u_1p_0 + u_2\,(1-p_0)} \\ \dfrac{-u_2p_0}{u_1p_0 + u_2\,(1-p_0)} & \dfrac{u_2p}{u_1p_0 + u_2\,(1-p_0)} - \lambda \end{vmatrix} \\ &= \frac{\lambda^2\,(u_1p_0 + u_2\,(1-p_0))^2 - \lambda\,(u_1p_0 + u_2\,(1-p_0))\,u_1\,(1-p_0)}{(u_1p_0 + u_2\,(1-p_0))^2} \end{aligned}$$

$$= \frac{-\lambda\left(u_1p_0+u_2\left(1-p_0\right)\right)u_2p_0}{\left(u_1p_0+u_2\left(1-p_0\right)\right)^2}.$$

Por lo que $|C\Sigma_{p_0}-\lambda I|=0$, si y sólo si

$$\lambda\left(u_1p_0+u_2\left(1-p_0\right)\right)\left[\lambda\left(u_1p_0+u_2\left(1-p_0\right)\right)-u_1\left(1-p_0\right)-u_2p_0\right]=0$$

con lo cual los autovalores que resultan son

$$\begin{aligned}\lambda_1 &= 0\\ \lambda_2 &= \frac{u_1\left(1-p_0\right)+u_2p_0}{u_1p_0+u_2\left(1-p_0\right)}.\end{aligned}$$

Así, el único autovalor no nulo β de $A\Sigma_{p_0}$ es

$$\frac{\left(u_1\left(1-p_0\right)+u_2p_0\right)\left(\sum\limits_{a=1}^{\Lambda}\eta_a h_a'\left(0\right)\varphi_a''\left(1\right)\right)}{u_1p_0+u_2\left(1-p_0\right)}$$

con lo que la distribución asintótica en este caso es de la forma

$$\frac{2nWD_{\varphi}^{h}\left(\hat{p},p_0\right)}{\left(\sum\limits_{a=1}^{\Lambda}\eta_a h_a'\left(0\right)\varphi_a''\left(1\right)\right)\left(u_1\left(1-p_0\right)+u_2p_0\right)\left(u_1p_0+u_2\left(1-p_0\right)\right)^{-1}} \underset{n\to\infty}{\overset{L}{\longrightarrow}} \mathcal{X}_1^2.$$

■

Teorema 3.9

Supóngase que se verifican las condiciones del teorema anterior. Bajo las hipótesis alternativas

$$H_{1,n}: p^{(n)} = p_0 + n^{-1/2}d$$

donde $p_0=\left(p_{10},p_{20}\right)^t$ *y* $d=\left(d_1,d_2\right)^t$ *con* $d_1+d_2=0$*, se tiene*

$$\frac{2nWD_{\varphi}^{h}\left(\hat{p},p_0\right)}{\left(\sum\limits_{a=1}^{\Lambda}\eta_a h_a'\left(0\right)\varphi_a''\left(1\right)\right)\left(u_1\left(1-p_0\right)+u_2p_0\right)\left(u_1p_0+u_2\left(1-p_0\right)\right)^{-1}} \underset{n\to\infty}{\overset{L}{\longrightarrow}} \mathcal{X}_1^2\left(\delta\right)$$

siendo $\mathcal{X}_1^2\left(\delta\right)$ *una Ji-cuadrado no central con 1 grado de libertad y parámetro de no centralidad*

$$\delta = \frac{d^2}{p_0\left(1-p_0\right)}.$$

Demostración

Se tiene,

$$\begin{aligned}\sqrt{n}\left(\hat{p}-p_0\right) &= \sqrt{n}\left(\hat{p}-p^{(n)}\right)+\sqrt{n}\left(p^{(n)}-p_0\right)\\ &= \sqrt{n}\left(\hat{p}-p^{(n)}\right)+d.\end{aligned}$$

Entonces al ser, bajo $H_{1,n}$,

$$\sqrt{n}\left(\hat{p}-p^{(n)}\right)\underset{n\to\infty}{\overset{L}{\to}} N\left(0,\Sigma_{p_0}\right)$$

se llega a

$$\sqrt{n}\left(\hat{p}-p_0\right)\underset{n\to\infty}{\overset{L}{\to}} N\left(d,\Sigma_{p_0}\right).$$

Por otro lado la distribución asintótica de la familia de estadísticos $2nWD_{\varphi}^{h}\left(\hat{p},p_0\right)$ y de la forma cuadrática $\sqrt{n}\left(\hat{p}-p_0\right)^t A\sqrt{n}\left(\hat{p}-p_0\right)$ con

$$A=\frac{\sum\limits_{a=1}^{\Lambda}\eta_a h_a'\left(0\right)\varphi_a''\left(1\right)}{u_1p_0+u_2\left(1-p_0\right)}\begin{pmatrix} u_1p_0^{-1} & 0\\ 0 & u_2\left(1-p_0\right)^{-1}\end{pmatrix}$$

y llamando $p_{10}=p_0$ y $p_{20}=1-p_0$, es la misma según se vio en el Teorema 3.2.

Entonces,

$$\frac{2nWD_{\varphi}^{h}\left(\hat{p},p_0\right)}{\left(\sum\limits_{a=1}^{\Lambda}\eta_a h_a'\left(0\right)\varphi_a''\left(1\right)\right)\left(u_1\left(1-p_0\right)+u_2p_0\right)\left(u_1p_0+u_2\left(1-p_0\right)\right)^{-1}}$$

y la forma cuadrática $\sqrt{n}\left(\hat{p}-p_0\right)^t B\sqrt{n}\left(\hat{p}-p_0\right)$ con

$$B=\left(u_1\left(1-p_0\right)+u_2p_0\right)^{-1}\begin{pmatrix} u_1p_0^{-1} & 0\\ 0 & u_2\left(1-p_0\right)^{-1}\end{pmatrix}$$

tienen la misma distribución asintótica.

Ahora,

$$\sqrt{n}B^{1/2}\left(\hat{p}-p_0\right)\underset{n\to\infty}{\overset{L}{\to}} N\left(B^{1/2}d,B^{1/2}\Sigma_{p_0}\left(B^{1/2}\right)^t\right)$$

siendo

$$\begin{aligned}
\Sigma^* &= B^{1/2}\Sigma_{p_0}\left(B^{1/2}\right)^t \\
&= \left(u_1\left(1-p_0\right)+u_2p_0\right)^{-1}\begin{pmatrix} u_1\left(1-p_0\right) & -\sqrt{u_1u_2p_0\left(1-p_0\right)} \\ -\sqrt{u_1u_2p_0\left(1-p_0\right)} & u_2p_0 \end{pmatrix}.
\end{aligned}$$

Como $\Sigma^*\Sigma^* = \Sigma^*$ y rango$(\Sigma^*) = 1$, se tiene que Σ^* es una proyección de rango 1. Así, aplicando el Lema de p. 63 en Ferguson (1996) se obtiene

$$\frac{2nWD_{\varphi}^{h}\left(\hat{p},p_0\right)}{\left(\sum\limits_{a=1}^{\Lambda}\eta_a h_a'\left(0\right)\varphi_a''\left(1\right)\right)\left(u_1\left(1-p_0\right)+u_2p_0\right)\left(u_1p_0+u_2\left(1-p_0\right)\right)^{-1}} \underset{n\to\infty}{\overset{L}{\longrightarrow}} \mathcal{X}_1^2\left(\delta\right)$$

con

$$\delta = \left(dB^{1/2}\right)^t\left(dB^{1/2}\right) = \frac{d^2}{p_0\left(1-p_0\right)}$$

y $d_1 = d$, $d_2 = -d$. ∎

3.8. Otros resultados asintóticos de interés. Aplicaciones estadísticas

Al hacer el contraste de bondad de ajuste, con hipótesis nula simple, se obtenía la distribución asintótica del estadístico $WD_{\varphi}^{h}\left(\hat{p},p_0\right)$ con p_0 conocido y determinado a través de la hipótesis nula. En este caso únicamente hay una población involucrada. En muchas situaciones será necesario estimar los dos argumentos de las (h,φ)-divergencias ponderadas, $WD_{\varphi}^{h}\left(\hat{p},\hat{q}\right)$, es decir, estarán involucradas dos poblaciones cuyas distribuciones de probabilidad asociadas, en relación a la partición $\mathcal{A} = \{A_1, ..., A_M\}$, serán p y q, respectivamente, y desconocidas. Los estimadores no paramétricos de p y q se obtendrán a través de dos muestras aleatorias independientes de tamaños n y m respectivamente. La distribución asintótica de $WD_{\varphi}^{h}\left(\hat{p},\hat{q}\right)$ nos permitirá hacer el contraste de homogeneidad $H_0 : p = q$, así como obtener una aproximación de la función de potencia del mismo.

3.8.1. Distribución asintótica de $WD_{\varphi}^{h}\left(\hat{p},\hat{q}\right)$

Sean $p = (p_1, ..., p_M)^t$ y $q = (q_1, ..., q_M)^t$ las distribuciones de probabilidad sobre $\mathcal{A} = \{A_1, ..., A_M\}$, es decir, $p, q \in \Delta_M$ y sean $\hat{p} = (\hat{p}_1, ..., \hat{p}_M)^t$ y $\hat{q} =$

$(\hat{q}_1, ..., \hat{q}_M)^t$ los estimadores no paramétricos de p y q basados en dos muestras aleatorias e independientes entre ellas de tamaños n y m respectivamente

Seguidamente se obtendrá la distribución asintótica del estadístico $WD_\varphi^h(\hat{p}, \hat{q})$.

Recuérdese que

$$WD_\varphi^h(p,q) = \sum_{a=1}^{\Lambda} \eta_a h_a \left(\frac{\sum\limits_{i=1}^{M} u_i q_i}{\sum\limits_{i=1}^{M} u_i p_i} \varphi_a \left(\frac{p_i}{q_i} \right) \right)$$

y considérese el desarrollo en serie de Taylor de la función $WD_\varphi^h(p,q)$ en el punto $(\hat{p}_1, ..., \hat{p}_M, \hat{q}_1, ..., \hat{q}_M)^t$ y en un entorno de $(p_1, ..., p_M, q_1, ..., q_M)^t$. Se tiene,

$$\begin{aligned} WD_\varphi^h(\hat{p}, \hat{q}) &= WD_\varphi^h(p,q) \\ &+ \sum_{i=1}^{M} \left(\frac{\partial WD_\varphi^h(p,q)}{\partial p_i} \right)_{(p_1,....,p_M,q_1,....,q_M)} (\hat{p}_i - p_i) \\ &+ \sum_{i=1}^{M} \left(\frac{\partial WD_\varphi^h(p,q)}{\partial q_i} \right)_{(p_1,....,p_M,q_1,....,q_M)} (\hat{q}_i - q_i) \\ &+ o\left(\|\hat{p} - p\| + \|\hat{q} - q\| \right) \end{aligned}$$

donde $\forall i = 1, ..., M$

$$\begin{aligned} t_i &= \frac{\partial WD_\varphi^h(p,q)}{\partial p_i} = \sum_{a=1}^{\Lambda} \left\{ A_{a,i} \times \left[\frac{u_i}{(E_p(u))^2} \left(\varphi_a' \left(\frac{p_i}{q_i} \right) \sum_{i=1}^{M} u_i p_i - \sum_{i=1}^{M} u_i q_i \varphi_a \left(\frac{p_i}{q_i} \right) \right) \right] \right\} \\ s_i &= \frac{\partial WD_\varphi^h(p,q)}{\partial q_i} = \sum_{a=1}^{\Lambda} \left\{ A_{a,i} \times \left[\frac{u_i}{(E_p(u))} \left(\varphi_a \left(\frac{p_i}{q_i} \right) - \varphi_a' \left(\frac{p_i}{q_i} \right) \frac{p_i}{q_i} \right) \right] \right\} \end{aligned}$$

siendo

$$A_{a,i} = \eta_a h_a' \left(\frac{\sum\limits_{i=1}^{M} u_i q_i}{\sum\limits_{i=1}^{M} u_i p_i} \varphi_a \left(\frac{p_i}{q_i} \right) \right).$$

Suponiendo que

$$\frac{m}{n+m} \underset{n,m\to\infty}{\longrightarrow} \lambda$$

se tiene que

$$\sqrt{\frac{nm}{n+m}}o\left(\|\hat{p}-p\|+\|\hat{q}-q\|\right)=o_p\left(1\right).$$

Por tanto las variables aleatorias

$$\left(\frac{nm}{n+m}\right)^{\frac{1}{2}}\left(WD_{\varphi}^{h}\left(\hat{p},\hat{q}\right)-WD_{\varphi}^{h}\left(p,q\right)\right)$$

y

$$\left(\frac{nm}{n+m}\right)^{\frac{1}{2}}\left(\sum_{i=1}^{M}t_i\left(\hat{p}_i-p_i\right)+\sum_{i=1}^{M}s_i\left(\hat{q}_i-q_i\right)\right)$$

tienen la misma distribución asintótica.

Ahora bien

$$\left(\frac{nm}{n+m}\right)^{\frac{1}{2}}\sum_{i=1}^{M}t_i\left(\hat{p}_i-p_i\right)=\left(\frac{m}{n+m}\right)^{\frac{1}{2}}\left(\sqrt{n}T^t\left(\hat{p}-p\right)\right)$$

converge en ley a una distribución normal unidimensional con media cero y varianza $\lambda\sigma_p^2=\lambda T^t\Sigma_pT$, siendo $T=\left(t_1,....,t_M\right)^t$ y $\Sigma_p=diag\left(p\right)-pp^t$, mientras que

$$\left(\frac{nm}{n+m}\right)^{\frac{1}{2}}\sum_{i=1}^{M}s_i\left(\hat{q}_i-q_i\right)=\left(\frac{n}{n+m}\right)^{\frac{1}{2}}\left(\sqrt{m}S^t\left(\hat{q}-q\right)\right)$$

converge en ley a una distribución normal unidimensional con media cero y varianza $\left(1-\lambda\right)\sigma_q^2=\left(1-\lambda\right)S^t\Sigma_qS$, siendo $S=\left(s_1,....,s_M\right)^t$ y $\Sigma_q=diag\left(q\right)-qq^t$.

En definitiva, se tiene

$$\left(\frac{nm}{n+m}\right)^{\frac{1}{2}}\left(WD_{\varphi}^{h}\left(\hat{p},\hat{q}\right)-WD_{\varphi}^{h}\left(p,q\right)\right)\underset{n,m\to\infty}{\overset{L}{\longrightarrow}}N\left(0,\sigma^2\right)$$

siendo

$$\sigma^2=\lambda\sigma_p^2+\left(1-\lambda\right)\sigma_q^2.$$

Con los resultados anteriores se puede establecer el siguiente teorema

Teorema 3.10

Considérese el estimador analógico $WD_{\varphi}^{h}\left(\hat{p},\hat{q}\right)$, obtenido al reemplazar las probabilidades p_i y q_i por sus frecuencias relativas observadas $\hat{p}_i$ y $\hat{q}_i$ $\left(i=1,....,M\right)$ respectivamente, basadas en sendas muestras aleatorias simples de tamaños n y

m, y supóngase que se verifican las hipótesis requeridas a las funciones h_a y φ_a, $a=1,....,\Lambda$, en (1.15), entonces si $\frac{m}{n+m} \underset{n,m\to\infty}{\longrightarrow} \lambda$ con $\lambda \in (0,1)$, se tiene

$$\left(\frac{nm}{n+m}\right)^{\frac{1}{2}} \left(WD_{\varphi}^{h}\left(\hat{p},\hat{q}\right) - WD_{\varphi}^{h}\left(p,q\right)\right) \underset{n,m\to\infty}{\overset{L}{\longrightarrow}} N\left(0,\sigma^2\right)$$

siempre y cuando $\sigma^2 > 0$, donde

$$\sigma^2 = \lambda\sigma_p^2 + (1-\lambda)\,\sigma_q^2$$

con

$$\begin{aligned} \sigma_p^2 &= T^t\Sigma_p T = \sum_{i=1}^{M} t_i^2 p_i - \left(\sum_{i=1}^{M} t_i p_i\right)^2 \\ \sigma_q^2 &= S^t\Sigma_q S = \sum_{i=1}^{M} s_i^2 q_i - \left(\sum_{i=1}^{M} s_i q_i\right)^2 \end{aligned}$$

$$\begin{aligned} \Sigma_p &= diag\,(p) - pp^t \\ \Sigma_q &= diag\,(q) - qq^t \end{aligned}$$

y $T = (t_1,....,t_M)^t$ siendo $\forall i = 1,....,M$

$$t_i = \sum_{a=1}^{\Lambda}\left\{A_{a,i} \times \left[\frac{u_i}{\left(\sum\limits_{i=1}^{M} u_i p_i\right)^2}\left(\varphi_a'\left(\frac{p_i}{q_i}\right)\sum_{i=1}^{M} u_i p_i - \sum_{i=1}^{M} u_i q_i \varphi_a\left(\frac{p_i}{q_i}\right)\right)\right]\right\},$$

$S = (s_1,....,s_M)^t$ siendo $\forall i = 1,....,M$

$$s_i = \sum_{a=1}^{\Lambda}\left\{A_{a,i} \times \left[\frac{u_i}{\left(\sum\limits_{i=1}^{M} u_i p_i\right)}\left(\varphi_a\left(\frac{p_i}{q_i}\right) - \varphi_a'\left(\frac{p_i}{q_i}\right)\frac{p_i}{q_i}\right)\right]\right\}$$

donde

$$A_{a,i} = \eta_a h_a'\left(\frac{\sum\limits_{i=1}^{M} u_i q_i}{\sum\limits_{i=1}^{M} u_i p_i}\varphi_a\left(\frac{p_i}{q_i}\right)\right).$$

Corolario 3.16

En el caso de las (h,φ)-divergencias, es decir, si $u_i = u$ $\forall i = 1, ..., M$, entonces se tiene

$$
\begin{aligned}
t_i &= \sum_{a=1}^{\Lambda} \left\{ \eta_a h'_a \left(\sum_{i=1}^{M} q_i \varphi_a \left(\frac{p_i}{q_i} \right) \right) \times \left(\varphi'_a \left(\frac{p_i}{q_i} \right) - \sum_{i=1}^{M} q_i \varphi_a \left(\frac{p_i}{q_i} \right) \right) \right\} \\
s_i &= \sum_{a=1}^{\Lambda} \left\{ \eta_a h'_a \left(\sum_{i=1}^{M} q_i \varphi_a \left(\frac{p_i}{q_i} \right) \right) \left(\varphi_a \left(\frac{p_i}{q_i} \right) - \varphi'_a \left(\frac{p_i}{q_i} \right) \frac{p_i}{q_i} \right) \right\}
\end{aligned}
$$

y

$$
\begin{aligned}
\sigma_p^2 &= \sum_{i=1}^{M} \left[\sum_{a=1}^{\Lambda} \left\{ \eta_a h'_a \left(\sum_{i=1}^{M} q_i \varphi_a \left(\frac{p_i}{q_i} \right) \right) \varphi'_a \left(\frac{p_i}{q_i} \right) \right\} \right]^2 p_i \\
&\quad - \left[\sum_{i=1}^{M} \left[\sum_{a=1}^{\Lambda} \left\{ \eta_a h'_a \left(\sum_{i=1}^{M} q_i \varphi_a \left(\frac{p_i}{q_i} \right) \right) \varphi'_a \left(\frac{p_i}{q_i} \right) \right\} \right] p_i \right]^2 \\
\sigma_q^2 &= \sum_{i=1}^{M} \left[\sum_{a=1}^{\Lambda} \left\{ \eta_a h'_a \left(\sum_{i=1}^{M} q_i \varphi_a \left(\frac{p_i}{q_i} \right) \right) \left(\varphi_a \left(\frac{p_i}{q_i} \right) - \varphi'_a \left(\frac{p_i}{q_i} \right) \frac{p_i}{q_i} \right) \right\} \right]^2 q_i \\
&\quad - \left[\sum_{i=1}^{M} \left[\sum_{a=1}^{\Lambda} \left\{ \eta_a h'_a \left(\sum_{i=1}^{M} q_i \varphi_a \left(\frac{p_i}{q_i} \right) \right) \left(\varphi_a \left(\frac{p_i}{q_i} \right) - \varphi'_a \left(\frac{p_i}{q_i} \right) \frac{p_i}{q_i} \right) \right\} \right] q_i \right]^2 .
\end{aligned}
$$

Corolario 3.17

En el caso de las $WD_\varphi(p,q)$, es decir, si $\Lambda = 1$, $\eta_1 = 1$, $h_1(x) = x$ y $\varphi_1(x) = \varphi(x)$, se tiene

$$
\begin{aligned}
t_i &= \frac{u_i}{\left(\sum\limits_{i=1}^{M} u_i p_i \right)^2} \left(\varphi' \left(\frac{p_i}{q_i} \right) \sum_{i=1}^{M} u_i p_i - \sum_{i=1}^{M} u_i q_i \varphi \left(\frac{p_i}{q_i} \right) \right) \\
s_i &= \frac{u_i}{\left(\sum\limits_{i=1}^{M} u_i p_i \right)} \left(\varphi \left(\frac{p_i}{q_i} \right) - \varphi' \left(\frac{p_i}{q_i} \right) \frac{p_i}{q_i} \right)
\end{aligned}
$$

y

$$\begin{aligned}
\sigma_p^2 &= \frac{1}{\left(\sum\limits_{i=1}^{M} u_i p_i\right)^2}\left[\sum_{i=1}^{M}\left(u_i\varphi'\left(\frac{p_i}{q_i}\right)\right)^2 p_i - \left(\sum_{i=1}^{M} u_i\varphi'\left(\frac{p_i}{q_i}\right)p_i\right)^2\right] \\
\sigma_q^2 &= \frac{1}{\left(\sum\limits_{i=1}^{M} u_i p_i\right)^2}\sum_{i=1}^{M}\left[\left(\varphi\left(\frac{p_i}{q_i}\right)-\varphi'\left(\frac{p_i}{q_i}\right)\frac{p_i}{q_i}\right)^2 u_i^2 q_i\right. \\
&\quad \left. - \left(\sum_{i=1}^{M}\left(\varphi\left(\frac{p_i}{q_i}\right)-\varphi'\left(\frac{p_i}{q_i}\right)\frac{p_i}{q_i}\right)u_i q_i\right)^2\right].
\end{aligned}$$

Corolario 3.18

Si además $u_i = u$ $\forall i = 1, ..., M$, es decir en el caso de las φ-divergencias, entonces

$$\begin{aligned}
t_i &= \varphi'\left(\frac{p_i}{q_i}\right) - \sum_{i=1}^{M} q_i\varphi\left(\frac{p_i}{q_i}\right) \\
s_i &= \varphi\left(\frac{p_i}{q_i}\right) - \varphi'\left(\frac{p_i}{q_i}\right)\frac{p_i}{q_i}
\end{aligned}$$

y

$$\begin{aligned}
\sigma_p^2 &= \sum_{i=1}^{M}\left(\varphi'\left(\frac{p_i}{q_i}\right)\right)^2 p_i - \left(\sum_{i=1}^{M}\varphi'\left(\frac{p_i}{q_i}\right)p_i\right)^2 \\
\sigma_q^2 &= \sum_{i=1}^{M}\left(\varphi\left(\frac{p_i}{q_i}\right)-\varphi'\left(\frac{p_i}{q_i}\right)\frac{p_i}{q_i}\right)^2 q_i - \left(\sum_{i=1}^{M}\left(\varphi\left(\frac{p_i}{q_i}\right)-\varphi'\left(\frac{p_i}{q_i}\right)\frac{p_i}{q_i}\right)q_i\right)^2.
\end{aligned}$$

Observación 3.6

En el caso de la divergencia de Kullback, se tiene

$$\begin{aligned}
\sigma_p^2 &= \sum_{i=1}^{M} p_i\left(\log\frac{p_i}{q_i}\right)^2 - \left(\sum_{i=1}^{M} p_i\log\frac{p_i}{q_i}\right)^2 \\
\sigma_q^2 &= \sum_{i=1}^{M}\frac{p_i}{q_i}^2 - 1.
\end{aligned}$$

Si el primer término del desarrollo en serie de Taylor es cero, entonces es necesario recurrir al segundo término de dicho desarrollo para poder encontrar la distribución asintótica de $WD_{\varphi}^{h}\left(\hat{p},\hat{q}\right)$.

Proposición 3.3

Sea $S_{n,m}=\left(\frac{nm}{n+m}\right)^{1/2}\left(\sum\limits_{i=1}^{M}t_i\left(\hat{p}_i-p_i\right)+\sum\limits_{i=1}^{M}s_i\left(\hat{q}_i-q_i\right)\right)$. *Entonces*

$$S_{n,m}=0\ \forall n,m\Leftrightarrow\sigma^2=0$$

Demostración

Se sabe que

$$\begin{aligned}Va\left(\sqrt{n}\sum_{i=1}^{M}t_i\left(\hat{p}_i-p_i\right)\right)&=\sigma_p^2=\sum_{i=1}^{M}t_i^2p_i-\left(\sum_{i=1}^{M}t_ip_i\right)^2\\Va\left(\sqrt{m}\sum_{i=1}^{M}s_i\left(\hat{q}_i-q_i\right)\right)&=\sigma_q^2=\sum_{i=1}^{M}s_i^2q_i-\left(\sum_{i=1}^{M}s_iq_i\right)^2.\end{aligned}$$

Así pues,

$$Va\left(S_{n,m}\right)=\frac{m}{n+m}\sigma_p^2+\frac{n}{n+m}\sigma_q^2$$

y

$$\lim_{n,m\to\infty}Va\left(S_{n,m}\right)=\lambda\sigma_p^2+\left(1-\lambda\right)\sigma_q^2=\sigma^2.$$

Si se cumple que $S_{n,m}=0\ \forall n,m$ entonces, $Va\left(S_{n,m}\right)=0\ \forall n,m$ con lo cual $\lim\limits_{n,m\to\infty}Va\left(S_{n,m}\right)=\sigma^2=0$.

Sea ahora $\sigma^2=0$, es decir $0=\lambda\sigma_p^2+\left(1-\lambda\right)\sigma_q^2=\sigma^2$ con lo que $\sigma_p^2=0$ y $\sigma_q^2=0$. Al ser $\sigma_p^2=\sum\limits_{i=1}^{M}t_i^2p_i-\left(\sum\limits_{i=1}^{M}t_ip_i\right)^2$ la varianza de una variable aleatoria que toma los valores $t_1,...,t_M$ con probabilidades $p_1,....,p_M$ se deberá verificar que $t_1=....=t_M$ y en consecuencia $\sum\limits_{i=1}^{M}t_i\left(\hat{p}_i-p_i\right)=0$. Análogamente se obtiene que $\sum\limits_{i=1}^{M}s_i\left(\hat{q}_i-q_i\right)=0$, con lo que $S_{n,m}=0\ \forall n,m$. ■

Si $p = q$ se tiene que $t_i = 0$ y $s_i = 0$ $\forall i = 1, ..., M$ con lo cual $\sigma^2 = 0$. En consecuencia, si $p = q$ el primer término del desarrollo de Taylor de $WD_{\varphi}^{h}(\hat{p}, \hat{q})$ es cero.

Se supone $p = q$. Por consiguiente, se considera el segundo término del desarrollo en serie de Taylor de $WD_{\varphi}^{h}(\hat{p}, \hat{q})$

$$WD_{\varphi}^{h}(\hat{p}, \hat{q}) = \frac{1}{2}\sum_{i=1}^{2M}\sum_{j=1}^{2M}\frac{\partial^2 WD_{\phi}^{h}(W)}{\partial w_i \partial w_j}(\hat{w}_i - w_i)(\hat{w}_j - w_j) + o\left(\|\hat{p} - p\|^2 + \|\hat{q} - q\|^2\right)$$

donde

$$w_i = \begin{cases} p_i & si \quad i = 1, ..., M \\ q_{i-M} & si \quad i = M+1, ..., 2M \end{cases}$$

y

$$\hat{w}_i = \begin{cases} \hat{p}_i & si \quad i = 1, ..., M \\ \hat{q}_{i-M} & si \quad i = M+1, ..., 2M \end{cases}$$

además, $W = (p, q)$, $\hat{W} = (\hat{p}, \hat{q})$.

Las derivadas parciales segundas son de la forma

$$\begin{aligned}
\left(\frac{\partial^2 WD_{\varphi}^{h}(W)}{\partial p_i^2}\right) &= \sum_{a=1}^{\Lambda}\eta_a h_a'(0)\,\varphi_a''(1)\,\frac{u_i}{p_i\sum_{i=1}^{M}u_i p_i} \\
\left(\frac{\partial^2 WD_{\varphi}^{h}(W)}{\partial p_i \partial q_j}\right) &= 0 \\
\left(\frac{\partial^2 WD_{\varphi}^{h}(W)}{\partial q_i^2}\right) &= \sum_{a=1}^{\Lambda}\eta_a h_a'(0)\,\varphi_a''(1)\,\frac{u_i}{p_i\sum_{i=1}^{M}u_i p_i} \\
\left(\frac{\partial^2 WD_{\varphi}^{h}(W)}{\partial p_i \partial p_j}\right) &= 0 \\
\left(\frac{\partial^2 WD_{\varphi}^{h}(W)}{\partial p_i \partial q_i}\right) &= -\sum_{a=1}^{\Lambda}\eta_a h_a'(0)\,\varphi_a''(1)\,\frac{u_i}{p_i\sum_{i=1}^{M}u_i p_i} \\
\left(\frac{\partial^2 WD_{\varphi}^{h}(W)}{\partial q_i \partial q_j}\right) &= 0.
\end{aligned}$$

Si denotamos por C la matriz

$$C = \left(\begin{array}{ccc|ccc} \frac{u_1}{p_1 \sum_{i=1}^{M} u_i p_i} & & & -\frac{u_1}{p_1 \sum_{i=1}^{M} u_i p_i} & & \\ & \ddots & & & \ddots & \\ & & \frac{u_M}{p_M \sum_{i=1}^{M} u_i p_i} & & & -\frac{u_M}{p_M \sum_{i=1}^{M} u_i p_i} \\ \hline -\frac{u_1}{p_1 \sum_{i=1}^{M} u_i p_i} & & & \frac{u_1}{p_1 \sum_{i=1}^{M} u_i p_i} & & \\ & \ddots & & & \ddots & \\ & & -\frac{u_M}{p_M \sum_{i=1}^{M} u_i p_i} & & & \frac{u_M}{p_M \sum_{i=1}^{M} u_i p_i} \end{array} \right)$$

se tiene

$$\frac{\partial^2 WD_\varphi^h(W)}{\partial w_i \partial w_j} = \sum_{a=1}^{\Lambda} \eta_a h_a'(0)\, \varphi_a''(1)\, C.$$

Al ser,

$$\frac{nm}{n+m} o\left(\|\hat{p} - p\|^2 + \|\hat{q} - q\|^2 \right) = o_p(1)$$

las variables

$$\frac{2mn}{n+m} \frac{WD_\varphi^h(\hat{p}, \hat{q})}{\sum_{a=1}^{\Lambda} \eta_a h_a'(0)\, \varphi_a''(1)}$$

y

$$\left(\frac{nm}{n+m} \right)^{\frac{1}{2}} \left(\hat{W} - W \right)^t C \left(\frac{nm}{n+m} \right)^{\frac{1}{2}} \left(\hat{W} - W \right)$$

tienen la misma distribución asintótica.

Ahora, el vector $\left(\frac{nm}{n+m} \right)^{\frac{1}{2}} \left(\hat{W} - W \right)$ sigue una distribución normal de media cero y matriz de varianzas covarianzas

$$\Sigma^* = \begin{pmatrix} \lambda \Sigma_p & 0 \\ 0 & (1 - \lambda)\, \Sigma_q \end{pmatrix}$$

donde Σ_p y Σ_q son los mismos que se definieron en el Teorema 3.1.

Por tanto

$$\frac{2mn}{n+m}\frac{WD_{\varphi}^{h}(\hat{p},\hat{q})}{\sum\limits_{a=1}^{\Lambda}\eta_a h_a'(0)\varphi_a''(1)} \underset{n,m\to\infty}{\overset{L}{\longrightarrow}} \sum_{i=1}^{r}\beta_i Z_i^2$$

donde $r = rango\,(\Sigma^* C \Sigma^*)$, las Z_i son variables aleatorias normales de media cero y varianza uno y los β_i son los autovalores de la matriz $C\Sigma^*$.

Con lo que se puede establecer

Teorema 3.11

Considérese el estimador analógico $WD_{\varphi}^{h}(\hat{p},\hat{q})$, obtenido al reemplazar las probabilidades p_i y q_i por sus frecuencias relativas observadas $\hat{p}_i$ y $\hat{q}_i$ $(i = 1, ..., M)$ respectivamente, basadas en sendas muestras aleatorias simples independientes de tamaños n y m, y supóngase que se verifican las hipótesis requeridas a las funciones h_a y φ_a, $a = 1, ..., \Lambda$, en (1.15), entonces si $p = q$, se tiene

$$\frac{2mn}{n+m}\frac{WD_{\varphi}^{h}(\hat{p},\hat{q})}{\sum\limits_{a=1}^{\Lambda}\eta_a h_a'(0)\varphi_a''(1)} \underset{n,m\to\infty}{\overset{L}{\longrightarrow}} \sum_{i=1}^{r}\beta_i Z_i^2$$

donde $r = rango\,(\Sigma^* C \Sigma^*)$, las Z_i son variables aleatorias normales e independientes de media cero y varianza uno y los β_i son los autovalores no nulos de la matriz $C\Sigma^*$, siendo

$$\Sigma^* = \begin{pmatrix} \lambda\Sigma_p & 0 \\ 0 & (1-\lambda)\Sigma_q \end{pmatrix}$$

y C la matriz

$$C = \left(\begin{array}{ccc|ccc} \frac{u_1}{p_1\sum\limits_{i=1}^{M}u_ip_i} & & & -\frac{u_1}{p_1\sum\limits_{i=1}^{M}u_ip_i} & & \\ & \cdot & & & \cdot & \\ & & \cdot & & & \cdot \\ & & \frac{u_M}{p_M\sum\limits_{i=1}^{M}u_ip_i} & & & -\frac{u_M}{p_M\sum\limits_{i=1}^{M}u_ip_i} \\ \hline -\frac{u_1}{p_1\sum\limits_{i=1}^{M}u_ip_i} & & & \frac{u_1}{p_1\sum\limits_{i=1}^{M}u_ip_i} & & \\ & \cdot & & & \cdot & \\ & & \cdot & & & \cdot \\ & & -\frac{u_M}{p_M\sum\limits_{i=1}^{M}u_ip_i} & & & \frac{u_M}{p_M\sum\limits_{i=1}^{M}u_ip_i} \end{array}\right).$$

Corolario 3.19

Si $p = q$, en el caso de las (h,φ)-divergencias, es decir, si $u_i = u$ $\forall i = 1, ..., M$, entonces se tiene

$$\frac{2mn}{n+m}\frac{WD_{\varphi}^{h}(\hat{p},\hat{q})}{\sum\limits_{a=1}^{\Lambda}\eta_a h'_a(0)\,\varphi''_a(1)} \underset{n,m\to\infty}{\overset{L}{\to}} \mathcal{X}^2_{M-1}.$$

Demostración

Si $u_i = u$ $\forall i = 1, ..., M$, se tiene

$$C = \left(\begin{array}{cccc|cccc}
p_1^{-1} & & & & -p_1^{-1} & & & \\
 & \cdot & & & & \cdot & & \\
 & & \cdot & & & & \cdot & \\
 & & & p_M^{-1} & & & & -p_M^{-1} \\
\hline
-p_1^{-1} & & & & p_1^{-1} & & & \\
 & \cdot & & & & \cdot & & \\
 & & \cdot & & & & \cdot & \\
 & & & -p_M^{-1} & & & & p_M^{-1}
\end{array}\right).$$

Pero en este caso $\Sigma^* C \Sigma^* C \Sigma^* = \Sigma^* C \Sigma^*$ y por tanto la variable anterior seguirá una distribución Ji-cuadrado con la traza de $C\Sigma^*$ como grados de libertad.

En efecto, para comprobar que $\Sigma^* C \Sigma^* C \Sigma^* = \Sigma^* C \Sigma^*$ basta ver que $C\Sigma^* C\Sigma^* = C\Sigma^*$.

Primero se calcula la matriz $C\Sigma^*$

$$C\Sigma^* = \left(\begin{array}{c|c}
\lambda(\delta_{ij}-p_j) & -(1-\lambda)(\delta_{ij}-p_j) \\
\hline
-\lambda(\delta_{ij}-p_j) & (1-\lambda)(\delta_{ij}-p_j)
\end{array}\right).$$

Sean $C\Sigma^* = (c_{ij})_{i,j}$ y $C\Sigma^* C\Sigma^* = \left(c^*_{ij}\right)_{i,j}$ entonces se tendrán que estudiar los siguientes casos

(i) $1 \leq i, j \leq M$

(ii) $1 \leq i \leq M < j \leq 2M$

(iii) $1 \leq j \leq M < i \leq 2M$

(iv) $M < i, j \leq 2M.$

Se comprobará el primer caso

$$\begin{aligned} c^*_{ii} &= \lambda^2 (1-p_i)^2 + \lambda^2 p_i \sum_{j \neq i}^{M} p_j + \lambda (1-\lambda) (1-p_i)^2 + \lambda (1-\lambda) p_i \sum_{j \neq i}^{M} p_j \\ &= \lambda (1-p_i)^2 + \lambda p_i \sum_{j \neq i}^{M} p_j \\ &= \lambda \left[(1-p_i)^2 + p_i (1-p_i) \right] \\ &= \lambda (1-p_i) . \end{aligned}$$

$$\begin{aligned} c^*_{ij} &= -\lambda^2 (1-p_i) p_j - \lambda^2 pj (1-p_j) + \lambda^2 p_j \sum_{k \neq i,j}^{M} p_k - \lambda (1-\lambda) (1-p_i) p_j \\ &\quad -\lambda (1-\lambda) p_j (1-p_j) + \lambda (1-\lambda) p_j \sum_{k \neq i,j}^{M} p_k \\ &= \lambda^2 p_j (p_i - 1 + p_j - 1 + 1 - p_i + 1 - p_j) + \lambda p_j \left(\sum_{k \neq i,j}^{M} p_k + p_i - 1 + p_j - 1 \right) \\ &= -\lambda p_j . \end{aligned}$$

Por tanto

$$c^*_{ij} = c_{ij} \ si \ 1 \leq i, j \leq M.$$

De forma análoga se comprueba para los otros casos.

Por otra parte

$$Traza\,(C\Sigma^*) = \lambda \sum_{i=1}^{M} (1-p_i) + (1-\lambda) \sum_{i=1}^{M} (1-p_i) = M - 1$$

con lo cual, utilizando la Observación 3.1,

$$\frac{2mn}{n+m}\frac{WD_{\varphi}^{h}\left(\hat{p},\hat{q}\right)}{\sum\limits_{a=1}^{\Lambda}\eta_a h_a'\left(0\right)\varphi_a''\left(1\right)}\underset{n,m\rightarrow\infty}{\overset{L}{\rightarrow}}\mathcal{X}_{M-1}^{2}.$$

■

Corolario 3.20

Si $p=q$, en el caso de $WD_{\varphi}\left(\hat{p},\hat{q}\right)$es decir, $\Lambda=1$, $\eta_1=1$, $h_1\left(x\right)=x$ y $\varphi_1\left(x\right)=\varphi\left(x\right)$, entonces

$$\frac{2mn}{n+m}\frac{WD_{\varphi}\left(\hat{p},\hat{q}\right)}{\varphi''\left(1\right)}\underset{n,m\rightarrow\infty}{\overset{L}{\rightarrow}}\sum_{i=1}^{r}\beta_i Z_i^2,$$

donde $r=rango\left(\Sigma^{*}C\Sigma^{*}\right)$, las Z_i son variables aleatorias normales e independientes de media cero y varianza uno y los β_i son los autovalores no nulos de la matriz $C\Sigma^{*}$.

Demostración

No hay más que ver que en este caso se tiene

$$\sum_{a=1}^{\Lambda}\eta_a h_a'\left(0\right)\varphi_a''\left(1\right)=\varphi''\left(1\right)$$

■

Corolario 3.21

Si en el caso anterior además, $u_i=u\ \forall i=1,....,M$, es decir, en el caso de las φ-divergencias, se tiene

$$\frac{2mn}{n+m}\frac{D_{\varphi}\left(\hat{p},\hat{q}\right)}{\varphi''\left(1\right)}\underset{n,m\rightarrow\infty}{\overset{L}{\rightarrow}}\mathcal{X}_{M-1}^{2}$$

Demostración

Si $u_i = u \ \forall i = 1, ..., M$ se tiene

$$C = \left(\begin{array}{ccc|ccc} p_1^{-1} & & & -p_1^{-1} & & \\ & \cdot & & & \cdot & \\ & & \cdot & & & \cdot \\ & & p_M^{-1} & & & -p_M^{-1} \\ \hline -p_1^{-1} & & & p_1^{-1} & & \\ & \cdot & & & \cdot & \\ & & \cdot & & & \cdot \\ & & -p_M^{-1} & & & p_M^{-1} \end{array} \right).$$

Como se ha visto anteriormente la matriz $C\Sigma^*$ es idempotente y $traza\,(C\Sigma^*) = M - 1$ por lo que

$$\frac{2mn}{n+m} \frac{D_\varphi(\hat{p}, \hat{q})}{\varphi''(1)} \underset{n,m\to\infty}{\overset{L}{\longrightarrow}} \mathcal{X}^2_{M-1}.$$

■

3.8.2. Aplicaciones estadísticas

Se van a plantear dos tipos de contrastes de homogeneidad, el primero homogeneidad con una distribución conocida y el segundo homogeneidad entre dos muestras. En ambos casos los contrastes se basarán en las (h, φ)-divergencias ponderadas ya que supondremos que los elementos A_j, $j = 1, ..., M$, asociados a la partición $\mathcal{A}$ están ponderados .

i) Homogeneidad con una distribución conocida

Sean $X^{(1)} = \left(X_1^{(1)}, ..., X_{n_1}^{(1)}\right), ..., X^{(m)} = \left(X_1^{(m)}, ..., X_{n_m}^{(m)}\right)$, m muestras independientes de tamaños $n_1, ..., n_m$, respectivamente. Se desea contrastar H_0 : Las m muestras proceden de la misma población cuya distribución de probabilidad perfectamente conocida se denotará por Q. Se considera el espacio estadístico $(\mathcal{X}, \beta_\mathcal{X}, Q)$ y la partición $\mathcal{A} = \{A_1, ..., A_M\} \subset \beta_\mathcal{X}$, con $P\left(X^{(j)} \in A_i\right) = p_{ij}$, $i = 1, ..., M$, $j = 1, ..., m$.

Entonces, el contraste será

$$H_0 : p_1 = ... = p_M = q$$

o equivalentemente

$$H_0 : p_{i1} = p_{i2} = ... = p_{im} = q_i = Q\left(A_i\right), i = 1, ..., M.$$

Sean $n_j\hat{p}_1^{(j)}, ..., n_j\hat{p}_M^{(j)}$ el número respectivo de observaciones $X^{(j)}$, $j = 1, ..., m$, sobre los elementos de la partición $A_1, ..., A_M$. Los vectores

$$\left(n_j\hat{p}_1^{(j)}, ..., n_j\hat{p}_M^{(j)}\right),\ j = 1, ..., m$$

se distribuyen multinomialmente con parámetros $(n_j; p_{1j}, ..., p_{Mj})$, $j = 1, ..., m$.

Si H_0 es cierta, entonces intuitivamente se espera que $WD_\varphi^h\left(\hat{p}_j, q\right)$ sea pequeña, el alejamiento del cero indica la poca compatibilidad de los datos con la hipótesis nula.

Así, para tamaños muestrales grandes, el test de tamaño α basado en $WD_\varphi^h\left(\hat{p}, q\right)$, viene dado por

$$\phi\left(\hat{p}_1^{(1)}, ..., \hat{p}_M^{(1)}, ..., \hat{p}_1^{(m)}, ..., \hat{p}_M^{(m)}\right) = \begin{cases} 1 & si\ T_6 > t_\alpha \\ 0 & en\ otro\ caso \end{cases}$$

siendo

$$T_6 = \frac{2}{\sum\limits_{a=1}^{\Lambda} \eta_a h_a'\left(0\right)\varphi_a''\left(1\right)} \sum_{j=1}^{m} n_j WD_\varphi^h\left(\hat{p}_j, q\right)$$

donde $\hat{p}_j = \left(\hat{p}_1^{(j)}, ..., \hat{p}_M^{(j)}\right)^t$ y $q = (q_1, ..., q_M)^t$ y t_α un número real tal que

$$P\left(\sum_{j=1}^{m}\sum_{i=1}^{r} \lambda_{ij} Z_{ij}^2 > t_\alpha\right) = \alpha.$$

Obsérvese que por el Corolario 3.4 se tiene para cada j, con $j = 1, ..., m$

$$2n_j WD_\varphi^h\left(\hat{p}_j, q\right) \underset{n\to\infty}{\overset{L}{\to}} \sum_{i=1}^{r} \beta_{ij} Z_{ij}^2$$

donde Z_{ij} son variables aleatorias normales independientes de media cero y varianza uno, los β_{ij} son los autovalores no nulos de $A\Sigma_p$ y $r = rango\left(\Sigma_p A \Sigma_p\right)$, donde la matriz A viene dada por

$$a_{ij} = \begin{cases} 0 & si \quad i \neq j \\ \sum\limits_{a=1}^{\Lambda} \eta_a h_a'\left(0\right)\varphi_a''\left(1\right) \frac{u_i}{p_i \sum\limits_{i=1}^{M} u_i p_i} & si \quad i = j \end{cases}.$$

Al ser

$$A=\sum_{a=1}^{\Lambda}\eta_a h_a'(0)\,\varphi_a''(1)\,C$$

con $C=(c_{ij})_{i,j=1,....,M}$

$$c_{ij}=\begin{cases}0 & si \quad i\neq j\\ \frac{u_i}{p_i\sum\limits_{i=1}^{M}u_ip_i} & si \quad i=j\end{cases}.$$

Entonces se tiene para cada $j=1,....,m$

$$\frac{2n_jWD_{\varphi}^{h}(\hat{p}_j,q)}{\sum\limits_{a=1}^{\Lambda}\eta_a h_a'(0)\,\varphi_a''(1)}\underset{n\to\infty}{\overset{L}{\to}}\sum_{i=1}^{r}\lambda_{ij}Z_{ij}^2$$

donde Z_{ij} son variables aleatorias normales independientes de media cero y varianza uno, los λ_{ij} son los autovalores no nulos de $C\Sigma_p$ y $r=rango\,(\Sigma_pA\Sigma_p)=rango\,(\Sigma_pC\Sigma_p)$.

ii) Homogeneidad

Supóngase ahora que se tienen dos muestras aleatorias independientes $X=(X_1,....,X_n)$ e $Y=(Y_1,....,Y_m)$ de tamaños n y m respectivamente. Para contrastar que las dos muestras proceden de la misma población, se considera la partición $\mathcal{A}=\{A_1,...,A_M\}$ y sean $P\,(X\in A_i)=p_i$ y $P\,(Y\in A_i)=q_i$, $i=1,....,M$. La hipótesis que se desea contrastar es

$$H_0:p_1=q_1,....,p_M=q_M.$$

Sean $\hat{p}=(\hat{p}_1,....,\hat{p}_M)^t$ y $\hat{q}=(\hat{q}_1,....,\hat{q}_M)^t$ los vectores de frecuencias relativas basados en X e Y respectivamente.

El test, entonces vendrá dado por

$$\phi\,(p_1,....,p_M,q_1,....,q_M)=\begin{cases}1 & si\ T_7>t_\alpha\\ 0 & en\ otro\ caso\end{cases}$$

donde

$$T_7=\frac{2mn}{n+m}\frac{WD_{\varphi}^{h}(\hat{p},\hat{q})}{\sum\limits_{a=1}^{\Lambda}\eta_a h_a'(0)\,\varphi_a''(1)}$$

y t_α un número real tal que

$$P\left(\sum_{i=1}^{r}\beta_i Z_i^2 > t_\alpha\right) = \alpha$$

donde β_i y Z_i son los definidos en el Teorema 3.11.

Observación 3.7

El Teorema 3.11 se puede utilizar para calcular la potencia asintótica del test anterior en $p = (p_1, ..., p_M)^t$ *y* $q = (q_1, ..., q_M)^t$*, con* $p \neq q$*. Esta viene dada por*

$$\beta_{n,m}(p,q) = P_{p,q}(T_7 > t_\alpha) = P_{p,q}\left(\frac{2mn}{n+m}\frac{WD_\varphi^h(\hat{p},\hat{q})}{\sum\limits_{a=1}^{\Lambda}\eta_a h'_a(0)\varphi''_a(1)} > t_\alpha\right).$$

Ahora bien, como

$$\left(\frac{nm}{n+m}\right)^{\frac{1}{2}}\left(WD_\varphi^h(\hat{p},\hat{q}) - WD_\varphi^h(p,q)\right) \underset{n,m\to\infty}{\xrightarrow{L}} N\left(0,\sigma^2\right)$$

se tiene

$$\begin{aligned}\beta_{n,m}(p,q) &= P\left(\left(\frac{mn}{n+m}\right)^{1/2} WD_\varphi^h(\hat{p},\hat{q}) > \frac{t_\alpha\sum\limits_{a=1}^{\Lambda}\eta_a h'_a(0)\varphi''_a(1)}{2\left(\frac{mn}{n+m}\right)^{1/2}}\right) = \\ &= P\left(\frac{\left(\frac{nm}{n+m}\right)^{\frac{1}{2}}\left(WD_\varphi^h(\hat{p},\hat{q}) - WD_\varphi^h(p,q)\right)}{\sigma(p,q)} > \frac{t_\alpha L - 2\left(\frac{nm}{n+m}\right)WD_\varphi^h(p,q)}{2\left(\frac{mn}{n+m}\right)^{1/2}\sigma(p,q)}\right)\end{aligned}$$

con $L = \sum\limits_{a=1}^{\Lambda}\eta_a h'_a(0)\varphi''_a(1)$*, es decir,*

$$\beta_{n,m}(p,q) = 1 - \Phi_{n,m}\left(\frac{t_\alpha\sum\limits_{a=1}^{\Lambda}\eta_a h'_a(0)\varphi''_a(1) - 2\left(\frac{nm}{n+m}\right)WD_\varphi^h(p,q)}{2\left(\frac{mn}{n+m}\right)^{1/2}\sigma(p,q)}\right)$$

donde $\sigma^2(p,q)$ *es la expresión de la varianza del Teorema 3.11 y* $\Phi_{n,m}(x)$ *es una sucesión de funciones de distribución que tiende uniformemente a la función de distribución,* Φ*, de una normal de media cero y varianza uno.*

Además,

$$\lim_{n,m\to\infty} \beta_{n,m}(p,q) = 1,$$

es decir, el test es consistente en el sentido de Fraser.

4

Optimalidad en los contrastes de bondad de ajuste

4.1. Introducción

En el Capítulo 3 se obtuvo que la distribución asintótica de la familia de estadísticos

$$T_1 = 2nWD_{\varphi}^{h}\left(\hat{p}, p_0\right),$$

bajo la hipótesis de que $p = p_0 = (p_{01}, ..., p_{0M})^t$, coincide con la de la variable

$$\sum_{i=1}^{r} \beta_i Z_i^2$$

donde las Z_i son variables aleatorias normales e independientes de media cero y varianza uno, los β_i son los autovalores no nulos de $A\Sigma_{p_0}$, A viene dada en el Teorema 3.2, $\Sigma_{p_0} = diag\left(p_0\right) - p_0 p_0^t$ y $r = rango\left(\Sigma_{p_0} A \Sigma_{p_0}\right)$. Es decir desde un punto de vista asintótico el comportamiento de la familia de las (h, φ)-divergencias ponderadas es el mismo, es decir, la distribución asintótica es independiente de $h = (h_a)_{a \in \Lambda}$, $\varphi = (\varphi_a)_{a \in \Lambda}$ y $(\eta_a)_{a \in \Lambda}$. La cuestión es que desde un punto de vista práctico se trabajará con muestras finitas y en muchas ocasiones pequeñas, y en estos casos si existirá diferencia significativa a la hora de la elección de las correspondientes funciones h y φ, y en consecuencia resultará interesante el disponer de algún criterio que permita la selección de tales funciones. En general el estudio no se podrá llevar a cabo para todas las (h, φ)-divergencias y se tendrá que considerar una familia concreta de medidas de divergencia dependientes de un parámetro y dar algún criterio que permita seleccionar de forma conveniente el valor óptimo del parámetro. En este capítulo se analizarán dos procedimientos. Uno basado en la función de potencia, introducida y estudiada en el capítulo anterior, del test de bondad de ajuste y el otro consistente en seleccionar las funciones h y φ, asociadas al estadístico T_1, que permitan una mejor proximidad entre los momentos exactos y asintóticos bajo la hipótesis nula.

4.2. Función de potencia

Dado un punto $p^* = (p_1^*, ..., p_M^*)^t$, la función de potencia asociada a la familia de estadísticos T_1 cuantifica la probabilidad de rechazar la hipótesis nula $H_0 : p = p_0$ cuando el valor de p es p^*. Los mejores estadísticos de la familia son áquellos cuya función de potencia se aproxime a 1. Dada la distribución de probabilidad p^* distinta de p_0 una aproximación de la función de potencia del test basado en el estadístico, T_1, viene dada, según se vio en el Capítulo 3 mediante la expresión

$$\beta_n^1 (p_1^*, ..., p_M^*) = 1 - \Phi_{N(0,1)} \left(\frac{t_\alpha - 2nWD_\varphi^h (p^*, p_0)}{2\sqrt{n}\sigma_{p^*}} \right)$$

donde t_α y σ_{p^*} aparecen explicitados en la Sección 3.5.1 y $p^* = (p_1^*, ..., p_M^*)^t$.

Así dado un tamaño muestral n y una familia de estadísticos, por ejemplo la basada en la divergencia de Renyi, se puede elegir el valor r del parámetro que proporciona una mayor potencia para diversas alternativas dadas.

Se puede obtener otra aproximación a la función de potencia si el vector de probabilidades de la alternativa no es fijo sino que converge al vector p_0, de la hipótesis nula cuando $n \to \infty$. La función de potencia del test basándose en el estadístico T_1 cuando se consideran las alternativas, que convergen a H_0,

$$H_{1,n} : p^{(n)} = p_0 + n^{-1/2} d$$

donde $d = (d_1, ..., d_M)^t$ es tal que $\sum_{i=1}^{M} d_i = 0$, viene dada por

$$\beta_n^2 \left(p_1^{(n)}, ..., p_M^{(n)} \right) = P \left(T_1 > t_\alpha / H_{1,n} \right).$$

Haciendo uso del resultado del Teorema 3.3 se tiene que la función de potencia se podrá aproximar mediante la expresión

$$\beta_n^2\left(p_1^{(n)},...,p_M^{(n)}\right)=1-G\left(t_\alpha\right)$$

donde G es la función de distribución de la variable aleatoria $\sum\limits_{i=1}^{r}\beta_i\left(Z_i+\omega_i\right)^2+\xi$.

Si en lugar de considerar el estadístico T_1 se consideran los estadísticos

a) $(\beta^*)^{-1}\,2nWD_\psi^h\left(\hat{p},p_0\right)$ con $\beta^*=\max\left\{\beta_1,...,\beta_r\right\}$,

b) $(\beta^*)^{-1}\,2nWD_\psi^h\left(\hat{p},p_0\right)$ con $\beta^*=\frac{1}{r}\sum\limits_{i=1}^{r}\beta_i$,

c) $(\beta^*)^{-1}\,2n\left(1+\lambda^2\right)^{-1}WD_\psi^h\left(\hat{p},p_0\right)$ con $\beta^*=\frac{1}{r}\sum\limits_{i=1}^{r}\beta_i$ y $\lambda^2=\sum\limits_{i=1}^{r}\frac{\left(\beta_i-\bar{\beta}\right)^2}{r\bar{\beta}^2}$,

que sabemos siguen una distribución Ji-Cuadrado con r grados de libertad el primero y el segundo, y $v=r/\left(1+\lambda^2\right)$ el tercero, se tendrá que

$$\beta_n^2\left(p_1^{(n)},...,p_M^{(n)}\right)=1-F\left(\mathcal{X}_{s,\alpha}^2-\xi\right)$$

donde F es la función de distribución de una Ji-Cuadrado con s grados de libertad ($s=r$ para los dos primeros y $s=v$ para el tercer estadístico) y parámetro de no centralidad $\delta=\sum\limits_{i=1}^{r}\omega_i^2$. Este resultado se sigue por el hecho de que la variable aleatoria $\sum\limits_{i=1}^{r}\left(Z_i+\omega_i\right)^2$ es una variable aleatoria Ji-Cuadrado con s grados de libertad y parámetro de no centralidad $\delta=\sum\limits_{i=1}^{r}\omega_i^2$.

4.3. Momentos asintóticos y exactos: Comparación

En este apartado se buscan condiciones sobre las funciones φ y h asociadas al estadístico T_1 que permitan mayor proximidad entre los momentos exactos y asintóticos bajo la hipótesis nula.

Se comenzará dando un teorema que establece la esperanza y varianza asintótica del estadístico T_1 bajo la hipótesis nula $H_0: p=p_0$.

Teorema 4.1
Bajo la hipótesis nula $H_0: p=p_0$ se tiene que la esperanza y la varianza

asintótica del estadístico T_1 vienen dadas por

$$E\left[\sum_{i=1}^{r}\beta_i Z_i^2\right]=\frac{1}{E_{p_0}(u)}\sum_{i=1}^{M}u_i\left(1-p_{i0}\right)\left(\sum_{a=1}^{\Lambda}\eta_a h_a'(0)\,\varphi_a''(1)\right)$$

y

$$Va\left[\sum_{i=1}^{r}\beta_i Z_i^2\right]=\frac{2}{E_{p_0}(u)}\left\{\sum_{i=1}^{M}u_i^2\left(1-2p_{i0}\right)+\left(\sum_{i=1}^{M}u_i p_{i0}\right)^2\right\}\left(\sum_{a=1}^{\Lambda}\eta_a h_a'(0)\,\varphi_a''(1)\right)^2,$$

respectivamente.

Demostración

Se tiene, denotando por $\Delta=diag\left(\beta_1,...,\beta_r\right)$, que

$$\begin{aligned}E\left[\sum_{i=1}^{r}\beta_i Z_i^2\right] &= \sum_{i=1}^{r}\beta_i E\left[Z_i^2\right]=\sum_{i=1}^{r}\beta_i=traza\left(\Delta\right)=traza\left(A\Sigma_{p_0}\right)\\ &= \frac{1}{E_{p_0}(u)}\sum_{i=1}^{M}u_i\left(1-p_{i0}\right)\left(\sum_{a=1}^{\Lambda}\eta_a h_a'(0)\,\varphi_a''(1)\right)\end{aligned}$$

ya que es inmediato comprobar que la diagonal de la matriz $A\Sigma_{p_0}$, con A definida en el Corolario 3.4, viene dada por

$$l_{1121;1}\left(u_1\left(1-p_{10}\right),...,u_M\left(1-p_{1M}\right)\right),$$

siendo $l_{1121;1}=\frac{1}{E_{p_0}(u)}\left(\sum\limits_{a=1}^{\Lambda}\eta_a h_a'(0)\,\varphi_a''(1)\right)$.

En lo sucesivo se denotará por

$$l_{ijkl;r}=\frac{1}{E_{p_0}(u)^r}\left(\sum_{a=1}^{\Lambda}\eta_a h_a^{(i}(0)^j\,\varphi_a^{(k}(1)^l\right). \tag{4.1}$$

En relación a la varianza, se tiene

$$\begin{aligned}Va\left[\sum_{i=1}^{r}\beta_i Z_i^2\right] &= \sum_{i=1}^{r}\beta_i^2 Va\left[Z_i^2\right]=2traza\left(\Delta^2\right)=2traza\left(\left(A\Sigma_{p_0}\right)^2\right)\\ &= \frac{2}{E_{p_0}(u)^2}\left\{\sum_{i=1}^{M}u_i^2\left(1-2p_{i0}\right)+\left(\sum_{i=1}^{M}u_i p_{i0}\right)^2\right\}\left(\sum_{a=1}^{\Lambda}\eta_a h_a'(0)\,\varphi_a''(1)\right)^2\\ &= 2l_{1121;1}^2\left\{\sum_{i=1}^{M}u_i^2\left(1-2p_{i0}\right)+\left(\sum_{i=1}^{M}u_i p_{i0}\right)^2\right\}\end{aligned}$$

es sencillo comprobar que la diagonal de la matriz $(A\Sigma_{p_0})^2$ viene dada por

$$l_{1121;1}^2\left(u_1^2\left(1-p_{10}\right)^2+u_1p_{10}\sum_{j=2}^{M}u_jp_{j0},...,u_M^2\left(1-p_{M0}\right)+u_Mp_{M0}\sum_{j=1}^{M-1}u_jp_{j0}\right).$$

■

Observación 4.1

Dado que luego se necesitará la expresión asintótica del momento de orden 2 respecto al origen, se procederá a su cálculo,

$$\begin{aligned} E\left[\left(\sum_{i=1}^{r}\beta_iZ_i^2\right)^2\right] &= Va\left[\sum_{i=1}^{r}\beta_iZ_i^2\right]+E\left[\sum_{i=1}^{r}\beta_iZ_i^2\right]^2 \\ &= l_{1121;1}^2\left\{2\sum_{i=1}^{M}u_i^2\left(1-2p_{i0}\right)+\left(\sum_{i=1}^{M}u_i\left(1-p_{i0}\right)\right)^2\right. \\ &\quad \left.+2\left(\sum_{i=1}^{M}u_ip_{i0}\right)^2\right\}. \end{aligned}$$

Para los posteriores estudios de selección de los mejores h y φ vía momentos resulta más conveniente escribir la expresión anterior del momento de orden 2, en los siguientes términos

$$\begin{aligned} E\left[\left(\sum_{i=1}^{r}\beta_iZ_i^2\right)^2\right] &= l_{1121;1}^2\left\{2\sum_{i=1}^{M}u_i^2-4\sum_{i=1}^{M}u_i^2p_{i0}+\left(\sum_{i=1}^{M}u_i\right)^2\right. \\ &\quad \left.-2\left(\sum_{i=1}^{M}u_i\right)\left(\sum_{i=1}^{M}u_ip_{i0}\right)+3E_{p_0}\left(u\right)^2\right\}. \end{aligned}$$

Puesto que bajo la hipótesis nula simple

$$H_0 : p = p_0$$

el estadístico

$$T_1=\frac{2nWD_{\varphi}^{h}\left(\hat{p},p_0\right)}{\sum\limits_{a=1}^{\Lambda}\eta_a h_a'\left(0\right)\varphi_a''\left(1\right)}$$

se distribuye asintóticamente como una combinación lineal de Ji-cuadrados y los momentos de esta distribución límite son finitos y fueron hallados en el Teorema 4.1, se tiene que

$$\begin{array}{lcl} E\left[T_1\right] & \rightarrow & \dfrac{1}{E_{p_0}\left(u\right)} \sum\limits_{i=1}^{M} u_i\left(1-p_{i0}\right) \\ E\left[\left(T_1\right)^2\right] & \rightarrow & \dfrac{1}{E_{p_0}\left(u\right)^2} \left\{ 2\sum\limits_{i=1}^{M} u_i^2 - 4\sum\limits_{i=1}^{M} u_i^2 p_{i0} + \left(\sum\limits_{i=1}^{M} u_i\right)^2 \right. \\ & & \left. -2\left(\sum\limits_{i=1}^{M} u_i\right)\left(\sum\limits_{i=1}^{M} u_i p_{i0}\right) + 3E_{p_0}\left(u\right)^2 \right\} \end{array}$$

Obsérvese que para $u_i = u$, $i = 1, ..., M$, se llega a

$$\begin{array}{lcl} E\left[T_1\right] & \rightarrow & M-1 \\ E\left[\left(T_1\right)^2\right] & \rightarrow & M^2-1 \end{array}$$

es decir, los dos primeros momentos de T_1 son asintóticamente equivalentes a los dos primeros momentos de una Ji-cuadrado cuando $n \rightarrow \infty$.

Para estudiar la velocidad de convergencia de los momentos exactos

$$\mu_\beta\left(T_1\right) = E\left[\left(T_1\right)^\beta\right], \qquad \beta = 1, 2$$

del estadístico T_1 a los asintóticos, se cálcula el desarrollo asintótico de segundo orden de estos momentos y se comprobará que se pueden escribir de la forma

$$\mu_\beta\left(T_1\right) = m_{\beta,0}\left(\varphi, h\right) + \frac{m_{\beta,1}\left(\varphi, h\right)}{N} + o\left(n^{-1}\right), \qquad \beta = 1, 2$$

siendo además

$$m_{\beta,0}\left(\varphi, h\right) = \mu_\beta\left(\sum_{i=1}^{r} \beta_i Z_i^2\right).$$

Luego, habrá que elegir las φ y las h que verifiquen

$$m_{\beta,1}\left(\varphi, h\right) = 0, \quad \beta = 1, 2 \qquad .$$

La proposición siguiente da el desarrollo en serie de Taylor del estadístico T_1 para $\hat{p}$ en un entorno de p_0.

Proposición 4.1

Denotando por $W_i = \sqrt{n}\left(\hat{p}_i - p_{i0}\right)$, $i = 1, ..., M$, *se tiene*

$$\begin{aligned}
T_1 \;\; = \;\; & \frac{1}{E_{p_0}(u)} \sum_{i=1}^{M} \frac{u_i}{p_{i0}} W_i^2 + \frac{1}{3} \left\{ \frac{l_{1131;1}}{l_{1121;0}} \sum_{i=1}^{M} \frac{u_i}{p_{i0}^2} \frac{W_i^3}{\sqrt{n}} - \frac{3}{E_{p_0}(u)} \sum_{i=1}^{M} \sum_{j=1}^{M} \frac{u_i u_j}{p_{i0}} \frac{W_i^2 W_j}{\sqrt{n}} \right\} \\
& + \frac{1}{12} \left\{ \frac{l_{1141;1}}{l_{1121;0}} \sum_{i=1}^{M} \frac{u_i}{p_{i0}^3} \frac{W_i^4}{n} + 3 \frac{l_{2122;2}}{l_{1121;0}} \sum_{i=1}^{M} \sum_{j=1}^{M} \frac{u_i u_j}{p_{i0} p_{j0}} \frac{W_i^2 W_j^2}{n} \right. \\
& - 4 \frac{l_{1131;2}}{l_{1121;0}} \sum_{i=1}^{M} \sum_{j=1}^{M} \frac{u_i u_j}{p_{i0}^2} \frac{W_i^3 W_j}{n} + \frac{12}{E_{p_0}(u)^3} \left[\sum_{i=1}^{M} \frac{u_i^3}{p_{i0}} \frac{W_i^4}{n} \right. \\
& + 2 \sum_{\substack{i=1 \\ i \neq j}}^{M} \sum_{j=1}^{M} \frac{u_i^2 u_j}{p_{i0}} \frac{W_i^3 W_j}{n} + \sum_{\substack{i=1 \\ i \neq j}}^{M} \sum_{j=1}^{M} \frac{u_i u_j^2}{p_{i0}} \frac{W_i^2 W_j^2}{n} \\
& \left. \left. + \sum_{\substack{i=1 \\ i \neq j \neq k}}^{M} \sum_{j=1}^{M} \sum_{i=1}^{M} \frac{u_i u_j u_k}{p_{i0}} \frac{W_i^2 W_j W_k}{n} \right] \right\} + O_p\left(n^{-3/2}\right)
\end{aligned}$$

con $l_{ijkl;r}$ *definido en* (4.1).

Demostración

Se comenzará dando el desarrollo de Taylor de $WD_\varphi^h(p, p_0)$ en torno al punto p_0. Este viene dado en los siguientes términos:

$$\begin{aligned}
WD_\varphi^h(p, p_0) \;\; = \;\; & \sum_{i=1}^{M} \left(\frac{\partial WD_\varphi^h(p, p_0)}{\partial p_i} \right)_{p=p_0} (p_i - p_{i0}) \\
& \frac{1}{2!} \left\{ \sum_{i=1}^{M} \left(\frac{\partial^2 WD_\varphi^h(p, p_0)}{\partial p_i^2} \right)_{p=p_0} (p_i - p_{i0})^2 \right. \\
& \left. + 2 \sum_{i=1}^{M} \sum_{\substack{j=1 \\ i \neq j}}^{M} \left(\frac{\partial^2 WD_\varphi^h(p, p_0)}{\partial p_i \partial p_j} \right)_{p=p_0} (p_i - p_{i0})(p_j - p_{j0}) \right\} \\
& + \frac{1}{3!} \left\{ \sum_{i=1}^{M} \left(\frac{\partial^3 WD_\varphi^h(p, p_0)}{\partial p_i^3} \right)_{p=p_0} (p_i - p_{i0})^3 \right.
\end{aligned}$$

$$+3\sum_{i=1}^{M}\sum_{\substack{j=1\\i\neq j}}^{M}\left(\frac{\partial^3 WD_\varphi^h(p,p_0)}{\partial p_i^2\partial p_j}\right)_{p=p_0}(p_i-p_{i0})^2(p_j-p_{j0})$$

$$\left.+\sum_{i=1}^{M}\sum_{\substack{j=1\\i\neq j\neq k}}^{M}\sum_{k=1}^{M}\left(\frac{\partial^3 WD_\varphi^h(p,p_0)}{\partial p_i\partial p_j\partial p_k}\right)_{p=p_0}(p_i-p_{i0})(p_j-p_{j0})(p_k-p_{k0})\right\}$$

$$+\frac{1}{4!}\left\{\sum_{i=1}^{M}\left(\frac{\partial^4 WD_\varphi^h(p,p_0)}{\partial p_i^4}\right)_{p=p_0}(p_i-p_{i0})^4\right.$$

$$+4\sum_{i=1}^{M}\sum_{\substack{j=1\\i\neq j}}^{M}\left(\frac{\partial^4 WD_\varphi^h(p,p_0)}{\partial p_i^3\partial p_j}\right)_{p=p_0}(p_i-p_{i0})^3(p_j-p_{j0})$$

$$+3\sum_{i=1}^{M}\sum_{\substack{j=1\\i\neq j}}^{M}\left(\frac{\partial^4 WD_\varphi^h(p,p_0)}{\partial p_i^2\partial p_j^2}\right)_{p=p_0}(p_i-p_{i0})^2(p_j-p_{j0})^2$$

$$+6\sum_{i=1}^{M}\sum_{\substack{j=1\\i\neq j\neq k}}^{M}\sum_{k=1}^{M}\left(\frac{\partial^4 WD_\varphi^h(p,p_0)}{\partial p_i^2\partial p_j\partial p_k}\right)_{p=p_0}(p_i-p_{i0})^2(p_j-p_{j0})(p_k-p_{k0})$$

$$+\sum_{i=1}^{M}\sum_{\substack{j=1\\i\neq j\neq k\neq l}}^{M}\sum_{k=1}^{M}\sum_{k=1}^{M}\left(\frac{\partial^4 WD_\varphi^h(p,p_0)}{\partial p_i\partial p_j\partial p_k\partial p_l}\right)_{p=p_0}(p_i-p_{i0})(p_j-p_{j0})$$

$$\left.\times(p_k-p_{k0})(p_l-p_{l0})\right\}+o\left(\|p-p_0\|\right).$$

Utilizando la notación dada en (4.1), se tiene

$$\left(\frac{\partial^2 WD_\varphi^h(p,p_0)}{\partial p_i^2}\right)_{p=p_0} = l_{1121;1}\frac{u_i}{p_{i0}}$$

$$\left(\frac{\partial^2 WD_\varphi^h(p,p_0)}{\partial p_i\partial p_j}\right)_{p=p_0} = 0$$

$$\left(\frac{\partial^3 WD_\varphi^h(p,p_0)}{\partial p_i^3}\right)_{p=p_0} = -3l_{1121;2}\frac{u_i^2}{p_{i0}} + l_{1131;1}\frac{u_i}{p_{i0}^2}$$

$$\left(\frac{\partial^3 WD_\varphi^h(p,p_0)}{\partial p_i^2 \partial p_j}\right)_{p=p_0} = -l_{1121;2}\frac{u_i u_j}{p_{i0}}$$

$$\left(\frac{\partial^3 WD_\varphi^h(p,p_0)}{\partial p_i \partial p_j \partial p_k}\right)_{p=p_0} = 0.$$

Por lo que el término correspondiente a las derivadas terceras se puede escribir de la forma

$$\begin{aligned}
&\sum_{i=1}^{M}\left(\frac{\partial^3 WD_\varphi^h(p,p_0)}{\partial p_i^3}\right)_{p=p_0}(p_i-p_{i0})^3 \\
&+3\sum_{i=1}^{M}\sum_{\substack{j=1\\ i\neq j}}^{M}\left(\frac{\partial^3 WD_\varphi^h(p,p_0)}{\partial p_i^2\partial p_j}\right)_{p=p_0}(p_i-p_{i0})^2(p_j-p_{j0}) \\
=\ & l_{1131;1}\sum_{i=1}^{M}\frac{u_i}{p_{i0}^2}(p_i-p_{i0})^3 - 3l_{1121;2}\sum_{i=1}^{M}\sum_{j=1}^{M}\frac{u_i u_j}{p_{i0}}(p_i-p_{i0})^2(p_j-p_{j0}).
\end{aligned}$$

Veamos las expresiones correspondientes a las derivadas cuartas,

$$\left(\frac{\partial^4 WD_\varphi^h(p,p_0)}{\partial p_i^4}\right)_{p=p_0} = 3l_{2122;2}\frac{u_i^2}{p_{i0}^2} + 12l_{1121;3}\frac{u_i^3}{p_{i0}} - 4l_{1131;2}\frac{u_i^2}{p_{i0}^2} + l_{1141;1}\frac{u_i}{p_{i0}^3}$$

$$\left(\frac{\partial^4 WD_\varphi^h(p,p_0)}{\partial p_i^3 \partial p_j}\right)_{p=p_0} = 6l_{1121;3}\frac{u_i^2 u_j}{p_{i0}} - l_{1131;2}\frac{u_i u_j}{p_{i0}^2}$$

$$\left(\frac{\partial^4 WD_\varphi^h(p,p_0)}{\partial p_i^2 \partial p_j^2}\right)_{p=p_0} = l_{2122;2}\frac{u_i u_j}{p_{i0}p_{j0}} + 2l_{1121;3}\frac{u_i^2 u_j}{p_{j0}} + 2l_{1121;3}\frac{u_i u_j^2}{p_{i0}}$$

$$\left(\frac{\partial^4 WD_\varphi^h(p,p_0)}{\partial p_i^2 \partial p_j \partial p_k}\right)_{p=p_0} = 2l_{1121;3}\frac{u_i u_j u_k}{p_{i0}}$$

$$\left(\frac{\partial^4 WD_\varphi^h(p,p_0)}{\partial p_i \partial p_j \partial p_k \partial p_l}\right)_{p=p_0} = 0,$$

con lo cual se tiene,

$$\begin{aligned}
&\sum_{i=1}^{M}\left(\frac{\partial^4 WD_\varphi^h(p,p_0)}{\partial p_i^4}\right)_{p=p_0}(p_i-p_{i0})^4+\\
&+4\sum_{i=1}^{M}\sum_{\substack{j=1\\ i\neq j}}^{M}\left(\frac{\partial^4 WD_\varphi^h(p,p_0)}{\partial p_i^3 \partial p_j}\right)_{p=p_0}(p_i-p_{i0})^3(p_j-p_{j0})\\
&+3\sum_{i=1}^{M}\sum_{\substack{j=1\\ i\neq j}}^{M}\left(\frac{\partial^4 WD_\varphi^h(p,p_0)}{\partial p_i^2 \partial p_j^2}\right)_{p=p_0}(p_i-p_{i0})^2(p_j-p_{j0})^2\\
&+6\sum_{i=1}^{M}\sum_{\substack{j=1\\ i\neq j\neq k}}^{M}\sum_{k=1}^{M}\left(\frac{\partial^4 WD_\varphi^h(p,p_0)}{\partial p_i^2 \partial p_j \partial p_k}\right)_{p=p_0}(p_i-p_{i0})^2(p_j-p_{j0})(p_k-p_{k0})\\
=\;& l_{1141;1}\sum_{i=1}^{M}\frac{u_i}{p_{i0}^3}(p_i-p_{i0})^4+3l_{2122;2}\sum_{i=1}^{M}\sum_{j=1}^{M}\frac{u_i u_j}{p_{i0}p_{jo}}(p_i-p_{i0})^2(p_j-p_{j0})^2\\
&-4l_{1131;2}\sum_{i=1}^{M}\sum_{j=1}^{M}\frac{u_i u_j}{p_{i0}^2}(p_i-p_{i0})^3(p_j-p_{j0})+12l_{1121;3}\left\{\sum_{i=1}^{M}\frac{u_i^3}{p_{i0}}(p_i-p_{i0})^4\right.\\
&+2\sum_{i=1}^{M}\sum_{\substack{j=1\\ i\neq j}}^{M}\frac{u_i u_j^2}{p_{i0}}(p_i-p_{i0})^3(p_j-p_{j0})+\sum_{i=1}^{M}\sum_{\substack{j=1\\ i\neq j}}^{M}\frac{u_i u_j^2}{p_{i0}}(p_i-p_{i0})^2(p_j-p_{j0})^2\\
&\left.+\sum_{i=1}^{M}\sum_{\substack{j=1\\ i\neq j\neq k}}^{M}\sum_{k=1}^{M}\frac{u_i u_j u_k}{p_{i0}}(p_i-p_{i0})^2(p_j-p_{j0})(p_k-p_{k0})\right\}.
\end{aligned}$$

En definitiva se llega a,

$$
\begin{aligned}
WD_{\varphi}^{h}\left(p,p_{0}\right) &= \frac{1}{2}l_{1121;1}\sum_{i=1}^{M}\frac{u_i}{p_{i0}} \\
&+\frac{1}{3!}\left\{l_{1131;1}\sum_{i=1}^{M}\frac{u_i}{p_{i0}^2}\left(p_i-p_{i0}\right)^3-3l_{1121;2}\sum_{i=1}^{M}\sum_{j=1}^{M}\frac{u_iu_j}{p_{i0}}\left(p_i-p_{i0}\right)^2\left(p_j-p_{j0}\right)\right\} \\
&+\frac{1}{4!}\left\{l_{1141;1}\sum_{i=1}^{M}\frac{u_i}{p_{i0}^3}\left(p_i-p_{i0}\right)^4+3l_{2122;2}\sum_{i=1}^{M}\sum_{j=1}^{M}\frac{u_iu_j}{p_{i0}p_{jo}}\left(p_i-p_{i0}\right)^2\left(p_j-p_{j0}\right)^2\right. \\
&-4l_{1131;2}\sum_{i=1}^{M}\sum_{j=1}^{M}\frac{u_iu_j}{p_{i0}^2}\left(p_i-p_{i0}\right)^3\left(p_j-p_{j0}\right)+12l_{1121;3}\left[\sum_{i=1}^{M}\frac{u_i^3}{p_{i0}}\left(p_i-p_{i0}\right)^4\right. \\
&+2\sum_{i=1}^{M}\sum_{\substack{j=1\\ i\neq j}}^{M}\frac{u_iu_j^2}{p_{i0}}\left(p_i-p_{i0}\right)^3\left(p_j-p_{j0}\right)+\sum_{i=1}^{M}\sum_{\substack{j=1\\ i\neq j}}^{M}\frac{u_iu_j^2}{p_{i0}}\left(p_i-p_{i0}\right)^2\left(p_j-p_{j0}\right)^2 \\
&\left.\left.+\sum_{i=1}^{M}\sum_{\substack{j=1\\ i\neq j\neq k}}^{M}\sum_{k=1}^{M}\frac{u_iu_ju_k}{p_{i0}}\left(p_i-p_{i0}\right)^2\left(p_j-p_{j0}\right)\left(p_k-p_{k0}\right)\right]\right\}+o\left(\left\|p-p_0\right\|\right).
\end{aligned}
$$

Haciendo $p=\hat{p}$, $W_i=\sqrt{n}\left(\hat{p}_i-p_{i0}\right)$ y teniendo en cuenta que

$$T_1=\frac{2nWD_{\varphi}^{h}\left(\hat{p},p_0\right)}{l_{1121;0}}$$

se tiene el resultado enunciado. ∎

Teorema 4.2

El momento de orden 1, $\beta=1$, $\mu_1\left(T_1\right)=E\left[T_1\right]$ satisface, para $p_0=\left(\frac{1}{M},...,\frac{1}{M}\right)^t$, la relación asintótica

$$E\left[T_1\right]=M-1+\frac{1}{n}f_{h,\varphi,u}^{1}+O\left(n^{-3/2}\right)$$

siendo

$$f_{h,\varphi,u}^{1}=\frac{1}{3}a_1+\frac{1}{4}a_2+a_3$$

con

$$
\begin{aligned}
a_1 &= \frac{l_{1131}}{l_{1121}}\left(2-3M+M^2\right)-6+3M+\left(6M-3M^2\right)\frac{\sum\limits_{i=1}^{M}u_i^2}{\left(\sum\limits_{i=1}^{M}u_i\right)^2}\\
a_2 &= \frac{l_{1141}}{l_{1121}}\left(1-2M+M^2\right)+\frac{l_{2122}}{l_{1121}}\left[(3-2M+M_2)+\left(2M^2-4M\right)\frac{\sum\limits_{i=1}^{M}u_i^2}{\left(\sum\limits_{i=1}^{M}u_i\right)^2}\right]\\
a_3 &= -\frac{l_{1131}}{l_{1121}}\left[(1-M)+\left(M^2-M\right)\frac{\sum\limits_{i=1}^{M}u_i^2}{\left(\sum\limits_{i=1}^{M}u_i\right)^2}\right]\\
&+ 3-M+\left(3M^2-5M\right)\frac{\sum\limits_{i=1}^{M}u_i^3}{\left(\sum\limits_{i=1}^{M}u_i\right)^3}-\frac{5M\sum\limits_{\substack{i=1\\i\neq j}}^{M}\sum\limits_{j=1}^{M}u_i^2u_j}{\left(\sum\limits_{i=1}^{M}u_i\right)^3}-\frac{2M\sum\limits_{\substack{i=1\\i\neq j}}^{M}\sum\limits_{j=1}^{M}u_iu_j^2}{\left(\sum\limits_{i=1}^{M}u_i\right)^3}\\
&+ \frac{M^2\sum\limits_{\substack{i=1\\i\neq j}}^{M}\sum\limits_{j=1}^{M}u_iu_j^2}{\left(\sum\limits_{i=1}^{M}u_i\right)^3}.
\end{aligned}
$$

Demostración

Llamando,

$$
\begin{aligned}
A_1 &= \frac{1}{E_{p_0}(u)}\sum_{i=1}^{M}\frac{u_i}{p_{i0}}W_i^2\\
A_2 &= \frac{l_{1131;1}}{l_{1121;0}}\sum_{i=1}^{M}\frac{u_i}{p_{i0}^2}\frac{W_i^3}{\sqrt{n}}-\frac{3}{E_{p_0}(u)}\sum_{i=1}^{M}\sum_{j=1}^{M}\frac{u_iu_j}{p_{i0}}\frac{W_i^2W_j}{\sqrt{n}}\\
A_3 &= \frac{l_{1141;1}}{l_{1121;0}}\sum_{i=1}^{M}\frac{u_i}{p_{i0}^3}\frac{W_i^4}{n}
\end{aligned}
$$

$$A_4 = 3\frac{l_{2122;2}}{l_{1121;0}}\sum_{i=1}^{M}\sum_{j=1}^{M}\frac{u_i u_j}{p_{i0}p_{j0}}\frac{W_i^2 W_j^2}{n}$$

$$A_5 = 4\frac{l_{1131;1}}{l_{1121;0}}\sum_{i=1}^{M}\sum_{j=1}^{M}\frac{u_i u_j}{p_{i0}^2}\frac{W_i^3 W_j}{n}$$

$$A_6 = \frac{12}{E_{p_0}(u)^3}\left[\sum_{i=1}^{M}\frac{u_i^3}{p_{i0}}\frac{W_i^4}{n} + 2\sum_{\substack{i=1\\ i\neq j}}^{M}\sum_{j=1}^{M}\frac{u_i^2 u_j}{p_{i0}}\frac{W_i^3 W_j}{n} + \sum_{\substack{i=1\\ i\neq j}}^{M}\sum_{j=1}^{M}\frac{u_i u_j^2}{p_{i0}}\frac{W_i^2 W_j^2}{n} + \right.$$
$$\left. + \sum_{\substack{i=1\\ i\neq j\neq k}}^{M}\sum_{j=1}^{M}\sum_{i=1}^{M}\frac{u_i u_j u_k}{p_{i0}}\frac{W_i^2 W_j W_k}{n}\right]$$

se tiene

$$T_1 = A_1 + \frac{1}{3}A_2 + \frac{1}{12}\left(A_3 + A_4 - A_5 + A_6\right) + O_p\left(n^{-3/2}\right).$$

Calculemos ahora la esperanza de A_i, $i = 1, ..., 6$.

Se tiene,

$$E\left[A_1\right] = \frac{1}{E_{p_0}(u)}\sum_{i=1}^{M}u_i\left(1 - p_{i0}\right)$$

$$E\left[A_2\right] = \frac{1}{n}\left\{\frac{l_{1131;1}}{l_{1121;0}}\sum_{i=1}^{M}u_i\left(2p_{i0} - 3 + \frac{1}{p_{i0}}\right) - \frac{3}{E_{p_0}(u)^2}\left[\sum_{i=1}^{M}u_i^2\left(2p_{i0}^2 - 3p_{i0} + 1\right)\right.\right.$$
$$\left.\left. + \sum_{\substack{i=1\\ i\neq j}}^{M}\sum_{j=1}^{M}u_i u_j\left(2p_{i0}p_{j0} - p_{j0}\right)\right]\right\}$$

$$E\left[A_3\right] = \frac{1}{n}\frac{l_{1141;1}}{l_{1121;0}}\sum_{i=1}^{M}u_i\left(3p_{i0} - 6 + \frac{3}{p_{i0}}\right)$$

$$E\left[A_4\right] = \frac{3}{n}\frac{l_{2122;2}}{l_{1121;0}}\left\{\sum_{i=1}^{M}u_i^2\left(3p_{i0}^2 - 6p_{i0} + 3\right) + \sum_{\substack{i=1\\ i\neq j}}^{M}\sum_{j=1}^{M}u_i u_j\left(3p_{i0}p_{j0} - p_{i0} + 1 - p_{j0}\right)\right\}$$

$$+O\left(n^{-3/2}\right)$$

$$E\left[A_5\right] = \frac{4}{n}\frac{l_{1131;2}}{l_{1121;0}}\left\{\sum_{i=1}^{M} u_i^2\left(3p_{i0}^2-6p_{i0}+3\right)+\sum_{\substack{i=1\\ i\neq j}}^{M}\sum_{j=1}^{M} u_i u_j\left(3p_{i0}p_{j0}-3p_{j0}\right)\right\}$$

$$+O\left(n^{-3/2}\right)$$

$$E\left[A_6\right] = \frac{12}{nE_{p_0}\left(u\right)^3}\left\{\sum_{i=1}^{M} u_i^3\left(3p_{i0}^3-6p_{i0}^2+3p_{i0}\right)+2\sum_{\substack{i=1\\ i\neq j}}^{M}\sum_{j=1}^{M} u_i^2 u_j\left(3p_{i0}^2p_{j0}-3p_{i0}p_{j0}\right)\right.$$

$$+\sum_{\substack{i=1\\ i\neq j}}^{M}\sum_{j=1}^{M} u_i u_j^2\left(3p_{i0}^2p_{j0}-p_{i0}p_{j0}+p_{j0}-p_{j0}^2\right)$$

$$\left.+\sum_{i=1}^{M}\sum_{\substack{j=1\\ i\neq j\neq k}}^{M}\sum_{k=1}^{M} u_i u_j u_k\left(3p_{i0}p_{j0}p_{k0}-p_{j0}p_{k0}\right)\right\}+O\left(n^{-3/2}\right)$$

Para $p_0=\left(\frac{1}{M},...,\frac{1}{M}\right)^t$, se llega a

$$E\left[A_1\right] = M-1$$

$$E\left[A_2\right] = \frac{1}{n}\left\{\frac{l_{1131}}{l_{1121}}\left(2-3M+M^2\right)-6+3M+\left(6M-3M^2\right)\frac{\sum_{i=1}^{M}u_i^2}{\left(\sum_{i=1}^{M}u_i\right)^2}\right\}$$

$$E\left[A_3\right] = \frac{1}{n}\left\{\frac{l_{1141}}{l_{1121}}\left(3-6M+3M^2\right)\right\}$$

$$E\left[A_4\right] = \frac{3}{n}\left\{\frac{l_{2122}}{l_{1121}}\left[\left(3-2M+M^2\right)+\left(2M^2-4M\right)\frac{\sum_{i=1}^{M}u_i^2}{\left(\sum_{i=1}^{M}u_i\right)^2}\right]\right\}+O\left(n^{-3/2}\right)$$

$$E[A_5] = \frac{4}{n}\left\{\frac{l_{1131}}{l_{1121}}\left[(3-3M)+\left(3M^2-3M\right)\frac{\sum\limits_{i=1}^{M}u_i^2}{\left(\sum\limits_{i=1}^{M}u_i\right)^2}\right]\right\}+O\left(n^{-3/2}\right)$$

$$E[A_6] = \frac{12}{n}\left\{3-M+\left(3M^2-5M\right)\frac{\sum\limits_{i=1}^{M}u_i^3}{\left(\sum\limits_{i=1}^{M}u_i\right)^3}-\frac{5M\sum\limits_{\substack{i=1\\i\neq j}}^{M}\sum\limits_{j=1}^{M}u_i^2u_j}{\left(\sum\limits_{i=1}^{M}u_i\right)^3}\right.$$

$$\left.-\frac{2M\sum\limits_{\substack{i=1\\i\neq j}}^{M}\sum\limits_{j=1}^{M}u_iu_j^2}{\left(\sum\limits_{i=1}^{M}u_i\right)^3}+\frac{M^2\sum\limits_{\substack{i=1\\i\neq j}}^{M}\sum\limits_{j=1}^{M}u_iu_j^2}{\left(\sum\limits_{i=1}^{M}u_i\right)^3}\right\}+O\left(n^{-3/2}\right).$$

Con lo cual se tiene el resultado enunciado. ■

Corolario 4.1
Para $u_i = u,\ i = 1,..,M$ *y* $p_0 = \left(\frac{1}{M},...,\frac{1}{M}\right)^t$ *se tiene*

$$E[T_1] = M-1+\frac{1}{n}f_{h,\varphi}^1+O\left(n^{-3/2}\right)$$

siendo

$$f_{h,\varphi}^1 = \frac{1}{3}\left\{\frac{l_{1131}}{l_{1121}}\left(2-3M+M^2\right)\right\}+\frac{1}{12}\left\{\frac{l_{1141}}{l_{1121}}\left(3-6M+3M^2\right)+3\frac{l_{2122}}{l_{1121}}\left(M^2-1\right)\right\}.$$

Demostración

El resultado se sigue del teorema anterior haciendo $u_i = u$, $i = 1,..,M$, en la expresión de $f_{h,\varphi,u}^1$. ■

Corolario 4.2
Para $u_i = u,\ i = 1,..,M,\ p_0 = \left(\frac{1}{M},...,\frac{1}{M}\right)^t$ *y* $h_a(x) = x$ *con* $\Lambda = 1$ *y* $\eta_a = 1$, *se tiene*

$$E[T_1] = M-1+\frac{1}{n}f_{\varphi}^1+O\left(n^{-3/2}\right)$$

siendo

$$f_{\varphi}^1 = \frac{\varphi'''(1)}{3\varphi''(1)}\left(2-3M+M^2\right)+\frac{\varphi''''(1)}{4\varphi''(1)}\left(1-2M+M^2\right).$$

Demostración

El resultado se sigue del corolario anterior haciendo $h_a(x) = x$ con $\Lambda = 1$ y $\eta_a = 1$, en la expresión de $f^1_{h,\varphi}$. ■

Teorema 4.3

Para $M \to \infty$, la condición que deben verificar las φ y las h para que los términos $f^1_{h,\varphi,u}$, $f^1_{h,\varphi}$ y f^1_{φ} se anulen vienen dadas mediante las ecuaciones:

$$\begin{aligned} 0 \;=\;& \frac{1}{3}\frac{l_{1131}}{l_{1121}} - \frac{\sum\limits_{i=1}^{M} u_i^2}{\left(\sum\limits_{i=1}^{M} u_i\right)^2} + \frac{1}{4}\frac{l_{1141}}{l_{1121}} + \frac{1}{4}\frac{l_{2122}}{l_{1121}}\left(1 + \frac{2\sum\limits_{i=1}^{M} u_i^2}{\left(\sum\limits_{i=1}^{M} u_i\right)^2}\right) - \frac{l_{1131}}{l_{1121}}\frac{\sum\limits_{i=1}^{M} u_i^2}{\left(\sum\limits_{i=1}^{M} u_i\right)^2} \\ & + \frac{3\sum\limits_{i=1}^{M} u_i^3}{\left(\sum\limits_{i=1}^{M} u_i\right)^3} + \frac{\sum\limits_{\substack{i=1\\ i\neq j}}^{M}\sum\limits_{j=1}^{M} u_i u_j^2}{\left(\sum\limits_{i=1}^{M} u_i\right)^3} \end{aligned}$$

para $f^1_{h,\varphi,u} = 0$.

$$4l_{1131} + 3l_{1141} + 3l_{2122} = 0$$

para $f^1_{h,\varphi} = 0$, supuesto que $l_{1121} \neq 0$.

Finalmente,

$$4\varphi'''(1) + 3\varphi''''(1) = 0$$

para $f^1_{\varphi} = 0$, supuesto que $\varphi''(1) \neq 0$

Demostración

El resultado se sigue del Teorema 4.2 y de los Corolarios 4.1 y 4.2 haciendo tender M a infinito. ■

Teorema 4.4

El momento de orden 2, $\beta = 2$, $\mu_2\left((T_1)^2\right) = E\left[(T_1)^2\right]$ satisface, para $p_0 = \left(\frac{1}{M}, ..., \frac{1}{M}\right)^t$, la relación asintótica

$$E\left[(T_1)^2\right] = E\left[\left(\sum_{i=1}^{r} \beta_i Z_i^2\right)^2\right] + \frac{1}{n} f^2_{h,\varphi,u} + O\left(n^{-3/2}\right)$$

siendo

$$f^2_{h,\varphi,u} = b_1 + \frac{1}{3}b_2 + \frac{2}{3}b_3 + \frac{1}{9}b_4 - 2b_5 + \frac{1}{2}b_6 + \frac{1}{2}b_7 + 2b_8$$

con

$$b_1 = \left(-6 + 4M - M^2\right) + \left(8M - 6M^2 + M^3\right) \frac{\sum\limits_{i=1}^{M} u_i^2}{\left(\sum\limits_{i=1}^{M} u_i\right)^2}$$

$$b_2 = \left(\frac{l_{1131}}{l_{1121}}\right)^2 \left\{ \left(-5 + 6M - 3M^2\right) + \left(9M - 12M^2 + 5M^3\right) \frac{\sum\limits_{i=1}^{M} u_i^2}{\left(\sum\limits_{i=1}^{M} u_i\right)^2} \right\}$$

$$b_3 = \frac{l_{1131}}{l_{1121}} \left\{ \left(10 - 4M - M^2 + M^3\right) + \left(30M - 33M^2 + 9M^3\right) \frac{\sum\limits_{i=1}^{M} u_i^2}{\left(\sum\limits_{i=1}^{M} u_i\right)^2} \right.$$

$$\left. - 2\left(33M - 42M^2 + 15M^3\right) \frac{\sum\limits_{i=1}^{M} u_i^3}{\left(\sum\limits_{i=1}^{M} u_i\right)^3} - 2\left(33M - 24M^2 + 3M^3\right) \frac{\sum\limits_{\substack{i=1 \\ i\neq j}}^{M} \sum\limits_{j=1}^{M} u_i^2 u_j}{\left(\sum\limits_{i=1}^{M} u_i\right)^3} \right\}$$

$$b_4 = \left(-15 + 6M - M^2\right) + \left(39M - 44M^2 + 15M^3\right) \frac{\sum\limits_{i=1}^{M} u_i^4}{\left(\sum\limits_{i=1}^{M} u_i\right)^4} + \left(39M - 26M^2 + 3M^3\right)$$

$$\times \frac{\sum_{\substack{i=1 \\ i\neq j}}^{M} \sum_{j=1}^{M} u_i^2 u_j^2}{\left(\sum_{i=1}^{M} u_i\right)^4} + \left(78M - 44M^2 + 6M^3\right) \frac{\sum_{\substack{i=1 \\ i\neq j}}^{M} \sum_{j=1}^{M} u_i^3 u_j}{\left(\sum_{i=1}^{M} u_i\right)^4}$$

$$+ \left(51M - 12M^2 + M^3\right) \frac{\sum_{i=1}^{M} \sum_{\substack{j=1 \\ i\neq j\neq k}}^{M} \sum_{k=1}^{M} u_i^2 u_j u_k}{\left(\sum_{i=1}^{M} u_i\right)^4}$$

$$b_5 = \left(-20 + 10M - 2M^2\right) + \left(40M - 38M^2 + 10M^3\right) \frac{\sum_{i=1}^{M} u_i^3}{\left(\sum_{i=1}^{M} u_i\right)^3}$$

$$+ \left(40M - 18M^2 + 2M^3\right) \frac{\sum_{\substack{i=1 \\ i\neq j}}^{M} \sum_{j=1}^{M} u_i^2 u_j}{\left(\sum_{i=1}^{M} u_i\right)^3}$$

$$b_6 = \frac{l_{1141}}{l_{1121}} \left\{ \left(-5 + 7M - 3M^2 + M^3\right) + \left(8M - 12M^2 + 4M^3\right) \frac{\sum_{i=1}^{M} u_i^2}{\left(\sum_{i=1}^{M} u_i\right)^2} \right\}$$

$$b_7 = \frac{l_{2122}}{l_{1121}} \left\{ \left(-15 + 9M - 3M^2 + M^3\right) + \left(36M - 42M^2 + 14M^3\right) \frac{\sum_{i=1}^{M} u_i^3}{\left(\sum_{i=1}^{M} u_i\right)^3} \right.$$

$$\left. + \left(36M - 18M^2 + 6M^3\right) \frac{\sum\limits_{\substack{i=1 \\ i\neq j}}^{M} \sum\limits_{j=1}^{M} u_i^2 u_j}{\left(\sum\limits_{i=1}^{M} u_i\right)^3} \right\}$$

$$\begin{aligned} b_8 &= \left(-15 + 6M - M^2\right) + \left(39M - 44M^2 + 15M^3\right) \frac{\sum\limits_{i=1}^{M} u_i^4}{\left(\sum\limits_{i=1}^{M} u_i\right)^4} \\ &\quad + \left(78M - 44M^2 + 6M^3\right) \frac{\sum\limits_{\substack{i=1 \\ i\neq j}}^{M} \sum\limits_{j=1}^{M} u_i^3 u_j}{\left(\sum\limits_{i=1}^{M} u_i\right)^4} + \left(39M - 24M^2 + 3M^3\right) \frac{\sum\limits_{\substack{i=1 \\ i\neq j}}^{M} \sum\limits_{j=1}^{M} u_i^2 u_j^2}{\left(\sum\limits_{i=1}^{M} u_i\right)^4} \\ &\quad + \left(39M - 12M^2 + M^3\right) \frac{\sum\limits_{i=1}^{M} \sum\limits_{\substack{j=1 \\ i\neq j\neq k}}^{M} \sum\limits_{k=1}^{M} u_i^2 u_j u_k}{\left(\sum\limits_{i=1}^{M} u_i\right)^4}. \end{aligned}$$

Demostración

Se siguen los mismos pasos que en el Teorema 4.2 ■

Corolario 4.3
Para $u_i = u$, $i = 1, ..., M$, *y* $p_0 = \left(\frac{1}{M}, ..., \frac{1}{M}\right)^t$ *se tiene*

$$E\left[(T_1)^2\right] = M^2 - 1 + \frac{1}{n} f_{h,\varphi}^2 + O\left(n^{-3/2}\right)$$

siendo

$$\begin{aligned} f_{h,\varphi}^{2} &= (2-2M)+\tfrac{1}{3}\left(\frac{l_{1131}}{l_{1121}}\right)^{2}\left(4-6M+2M^{2}\right) \\ &+ \tfrac{1}{2}\frac{l_{1141}}{l_{1121}}\left(3-5M-2M^{2}+(M+3)\,M^{2}\right)+\tfrac{1}{2}\frac{l_{2122}}{l_{1121}}\left(M^{3}+9M^{2}-M-3\right) \\ &+ \tfrac{2}{3}\frac{l_{1131}}{l_{1121}}\left(10-13M-6M^{2}+(M+8)\,M^{2}\right). \end{aligned}$$

Demostración

El resultado se sigue del teorema anterior haciendo $u_i = u$, $i = 1, ..., M$, en la expresión de $f^2_{h,\varphi,u}$. ∎

Corolario 4.4

Para $u_i = u$, $i = 1, ..., M$, $p_0 = \left(\frac{1}{M}, ..., \frac{1}{M}\right)^t$ *y* $h_a(x) = x$ *con* $\Lambda = 1$ *y* $\eta_a = 1$, *se tiene*

$$E\left[(T_1)^2\right] = M^2 - 1 + \frac{1}{n} f^2_\varphi + o\left(n^{-3/2}\right)$$

siendo

$$\begin{aligned} f_{\varphi}^{2} &= (2-2M)+\tfrac{1}{3}\left(\frac{\varphi'''(1)}{\varphi''(1)}\right)^{2}\left(4-6M+2M^{2}\right) \\ &+ \tfrac{1}{2}\frac{\varphi''''(1)}{\varphi''(1)}\left(3-5M-2M^{2}+(M+3)\,M^{2}\right) \\ &+ \tfrac{2}{3}\frac{\varphi'''(1)}{\varphi''(1)}\left(10-13M-6M^{2}+(M+8)\,M^{2}\right). \end{aligned}$$

Demostración

El resultado se sigue del corolario anterior haciendo $h_a(x) = x$ con $\Lambda = 1$ y $\eta_a = 1$, en la expresión de $f^2_{h,\varphi}$. ∎

Teorema 4.5

Para $M \to \infty$ la condición que deben verificar las φ y las h para que los términos $f^2_{h,\varphi,u}$, $f^2_{h,\varphi}$ y f^2_{φ} se anulen viene dada mediante las ecuaciones

$$
\begin{aligned}
0 \;=\; & \frac{\sum_{i=1}^{M} u_i^2}{\left(\sum_{i=1}^{M} u_i\right)^2} + \frac{5}{3}\left(\frac{l_{1131}}{l_{1121}}\right)^2 \frac{\sum_{i=1}^{M} u_i^2}{\left(\sum_{i=1}^{M} u_i\right)^2} + \frac{2}{3}\frac{l_{1131}}{l_{1121}}\left\{1 - 30\frac{\sum_{i=1}^{M} u_i^3}{\left(\sum_{i=1}^{M} u_i\right)^3} - 6\frac{\sum_{\substack{i=1\\ i\neq j}}^{M}\sum_{j=1}^{M} u_i^2 u_j}{\left(\sum_{i=1}^{M} u_i\right)^3}\right\} \\
& + \frac{5}{3}\frac{\sum_{i=1}^{M} u_i^4}{\left(\sum_{i=1}^{M} u_i\right)^4} + \frac{1}{3}\frac{\sum_{\substack{i=1\\ i\neq j}}^{M}\sum_{j=1}^{M} u_i^2 u_j^2}{\left(\sum_{i=1}^{M} u_i\right)^4} + \frac{2}{3}\frac{\sum_{\substack{i=1\\ i\neq j}}^{M}\sum_{j=1}^{M} u_i^3 u_j}{\left(\sum_{i=1}^{M} u_i\right)^4} + \frac{1}{9}\frac{\sum_{i=1}^{M}\sum_{\substack{j=1\\ i\neq j\neq k}}^{M}\sum_{k=1}^{M} u_i^2 u_j u_k}{\left(\sum_{i=1}^{M} u_i\right)^4} \\
& - 20\frac{\sum_{i=1}^{M} u_i^3}{\left(\sum_{i=1}^{M} u_i\right)^3} - 4\frac{\sum_{\substack{i=1\\ i\neq j}}^{M}\sum_{j=1}^{M} u_i^2 u_j}{\left(\sum_{i=1}^{M} u_i\right)^3} + \frac{1}{2}\frac{l_{1141}}{l_{1121}}\left\{1 + 4\frac{\sum_{i=1}^{M} u_i^2}{\left(\sum_{i=1}^{M} u_i\right)^2}\right\} \\
& + \frac{1}{2}\frac{l_{2122}}{l_{1121}}\left\{1 + 14\frac{\sum_{i=1}^{M} u_i^3}{\left(\sum_{i=1}^{M} u_i\right)^3} + 6\frac{\sum_{\substack{i=1\\ i\neq j}}^{M}\sum_{j=1}^{M} u_i^2 u_j}{\left(\sum_{i=1}^{M} u_i\right)^3}\right\} \\
& + 30\frac{\sum_{i=1}^{M} u_i^4}{\left(\sum_{i=1}^{M} u_i\right)^4} + 12\frac{\sum_{\substack{i=1\\ i\neq j}}^{M}\sum_{j=1}^{M} u_i^3 u_j}{\left(\sum_{i=1}^{M} u_i\right)^4} + 6\frac{\sum_{\substack{i=1\\ i\neq j}}^{M}\sum_{j=1}^{M} u_i^2 u_j^2}{\left(\sum_{i=1}^{M} u_i\right)^4} + 2\frac{\sum_{i=1}^{M}\sum_{\substack{j=1\\ i\neq j\neq k}}^{M}\sum_{k=1}^{M} u_i^2 u_j u_k}{\left(\sum_{i=1}^{M} u_i\right)^4}
\end{aligned}
$$

para $f^2_{h,\varphi,u} = 0$.

$$0 \;=\; 4l_{1131} + 3l_{1141} + 3l_{2122}$$

para $f^2_{h,\varphi} = 0.$

Finalmente

$$0 = 4\varphi'''(1) + 3\varphi''''(1)$$

para $f^2_{\varphi} = 0.$

Demostración

El resultado sigue del Teorema 4.4 y Corolarios 4.3 y 4.4 haciendo tender M a infinito.

■

En definitiva la mejor elección de las funciones h y φ vienen dadas por la resolución del sistema

$$\begin{cases} f^1_{h,\varphi,u} = 0 \\ f^2_{h,\varphi,u} = 0 \end{cases}$$

donde $f^1_{h,\varphi,u}$ y $f^2_{h,\varphi,u}$ vienen dadas, respectivamente en los Teoremas 4.2 y 4.4. En el caso general obsérvese que habrá que recurrir a procedimientos númericos para su resolución. Sin embargo para $u_i = u$, $i = 1, ..., M$, se llega a que ambas ecuaciones son la misma y viene dada por

$$0 = 4l_{1131} + 3l_{1141} + 3l_{2122}.$$

Así por ejemplo si se considera la familia de divergencias de Renyi que viene caracterizada por

$$\varphi_r(x) = \frac{x^r - r(x-1) - 1}{r(r-1)} \qquad r \neq 0, 1$$

y

$$h_r(x) = \frac{1}{r(r-1)} \log(r(r-1)x + 1) \qquad r \neq 0, 1$$

se tiene que

$$\begin{aligned} l_{1131} &= r - 2 \\ l_{1141} &= (r-2)(r-3) \\ l_{2122} &= -r(r-1) \end{aligned}$$

con lo cual se llega a que el r óptimo es $5/4$.

La misma sencillez se presenta en el caso $u_i = u$, $i = 1, ..., M$, y $h_a(x) = x$ con $\Lambda = 1$ y $\eta_a = 1$ ya que para este caso las divergencias óptimas se obtendrán como solución de la ecuación

$$0 = 4\varphi'''(1) + 3\varphi''''(1).$$

Así por ejemplo en el caso de la familia de divergencias de Rukhin, dada por

$$\varphi_a(x) = \left(\frac{1}{(a + (1-a)x)} - 1 \right)$$

se tiene que el valor de a óptimo viene dado por $a = 2/3$.

5

Bondad de ajuste con datos mal clasificados basados en $(h, \varphi)-$ Divergencias ponderadas

5.1. Introducción

En el Capítulo 3 se analizó ampliamente el problema de bondad de ajuste en el caso de hipótesis nula simple $H_0 : p = p_0$ en el supuesto de que se tengan ponderaciones sobre los elementos A_i, $i = 1, ..., M$ de la partición $\mathcal{A}$ de $\mathcal{X}$, haciendo uso de la familia de estadísticos de contraste

$$T_1 = 2nWD_{\varphi}^{h}\left(\hat{p}, p_0\right).$$

Una de las dificultades, que se suele encontrar en la práctica, es la posibilidad de una falsa o mala clasificación de uno o más individuos en las respectivas clases o categorías. Este problema fue abordado por primera vez por Bross (1954) para el caso de dos clases. Bross estableció que la proporción muestral es un estimador sesgado y el sesgo es una función de la cantidad de observaciones mal clasificadas. Mote y Anderson (1965) estudiaron el efecto de una clasificación errónea sobre el estadístico $\mathcal{X}^2$ de Pearson y llegaron a la conclusión de que si se ignoran los errores de clasificación, el tamaño del test aumentará y la potencia asintótica se reducirá. Un razonamiento similar con la familia de tests estadísticos $T_1 = 2nWD_{\varphi}^{h}(\hat{p}, p_0)$ conduciría a los mismos resultados. Esto nos lleva a que en el supuesto de que haya datos mal clasificados, la familia de tests estadísticos $T_1 = 2nWD_{\varphi}^{h}(\hat{p}, p_0)$ se debe modificar para paliar los problemas señalados previamente.

En este capítulo se introduce una familia de tests basados en las $(h,\varphi)-$ divergencias ponderadas, para el contraste de bondad de ajuste cuando el procedimiento o clasificador que se utiliza para asignar las observaciones $X_1, ..., X_n$ a los elementos A_i, $i = 1, ..., M$ de la partición $\mathcal{A}$ de $\mathcal{X}$, puede conducir a la clasificación errónea de algunas observaciones y existen ponderaciones en las clases. En esencia el nuevo procedimiento se basará en sustituir el estimador de máxima verosimilitud, $\hat{p}$, de p por uno obtenido por muestreo doble.

5.2. Método de estimación por muestreo doble

El método de muestreo doble fue introducido por Tenenbein (1970, 1971, 1972) en el contexto de la siguiente situación experimental. Supongamos que tenemos dos procedimientos para asignar las observaciones $X_1, ..., X_n$ a los elementos A_i, $i = 1, ..., M$ de la partición $\mathcal{A}$ de $\mathcal{X}$. Uno de ellos resulta muy costoso y está prácticamente libre de error y el otro es muy barato, pero resulta fiable hasta un cierto punto. Generalmente el aspecto costo es importante y se supondrá que los recursos están limitados. Un ejemplo sencillo sería el siguiente: Supongamos que las observaciones hay que clasificarlas como buenas o defectuosas. La única forma de saber si se debe clasificar como buena conduce a su destrucción (procedimiento caro, pero seguro) mientras que por otro lado mediante una inspección visual se puede clasificar como buena o defectuosa (procedimiento barato, pero no seguro). Diamond y Lilienfield (1962) analizan una interesante situación experimental relativa a una determinada enfermedad. Esta se puede diagnosticar mediante una serie de pruebas de bajo coste, pero no totalmente fiables o mediante una serie de pruebas muy costosas, pero prácticamente libres de error.

El método de muestreo doble de Tenenbein da un procedimiento alternativo que trata de dar a un coste razonable unos resultados suficientemente precisos. El esquema consiste en considerar n unidades experimentales en una primera etapa y clasificarlas en los correspondientes elementos de la partición A_i, $i = 1, ..., M$, utilizando los dos procedimientos de clasificación y en una segunda etapa tomar

$N-n$ unidades experimentales adicionales y clasificarlas en los elementos A_i, $i=1,...,M$, de la partición utilizando el procedimiento de clasificación sujeto a errores. Hochberg (1977) extendió el uso del muestreo doble dado por Tenenbein al caso de tablas de contingencia múltiples y Cheng y otros (1998) utilizaron el muestreo doble para dar contrastes alternativos a los de la ji-cuadrado y cociente de verosimilitudes en el problema de bondad de ajuste.

Veamos de una forma más precisa la forma de proceder con el muestreo doble de Tenenbein:

Se considera una muestra, $X_1,...,X_n$, de forma que cada unidad se pueda clasificar en una de las M diferentes clases, A_i, $i=1,...,M$ de la partición $\mathcal{A}$ de $\mathcal{X}$. Para cada unidad muestral se definen las variables aleatorias Y e Y^0 en la forma siguiente:

$$\begin{aligned} Y &= i \text{ si la unidad muestral verdaderamente pertenece a la categoría} \\ &\quad\; A_i \text{ con } i=1,...,M \\ Y^0 &= j \text{ si la unidad muestral es clasificada, por el procedimiento de} \\ &\quad\; \text{clasificación que puede dar lugar a una clasificación errónea,} \\ &\quad\; \text{en la categoría } A_j \text{ con } j=1,...,M. \end{aligned}$$

Las distribuciones marginales de Y e Y^0 son respectivamente

$$p_i = P\left(Y=i\right),\ \pi_j = P\left(Y^0=j\right),\ i,j=1,...,M,$$

con $\sum\limits_{i=1}^{M} p_i = \sum\limits_{j=1}^{M} \pi_j = 1.$

Para describir la clasificación errónea se definen las siguientes probabilidades:

$$\begin{aligned} \theta_{ij} &= \text{Probabilidad de que una unidad muestral perteneciendo a la} \\ &\quad\; \text{categoría } A_i \text{ se clasifique en la categoría } A_j. \end{aligned}$$

Es decir,

$$\theta_{ij} = P\left(Y^0=j/Y=i\right),\ i,j=1,...,M.$$

Es claro que:

$$\sum_{j=1}^{M} \theta_{ij} = 1$$

y

$$\pi_j = \sum_{i=1}^{M} p_i \theta_{ij},\ i, j = 1, ..., M.$$

Con esta notación el procedimiento de muestreo doble se puede describir mediante las dos siguientes etapas:

1. Se obtiene una muestra de n observaciones y se clasifican según los dos procedimientos: el que asigna las observaciones a las clases sin ningún género de duda, $Y_1, ..., Y_n$, en lo sucesivo se denominará clasificador correcto, y el que las puede clasificar de forma errónea, $Y_1^0, ..., Y_n^0$, en lo sucesivo se denominará clasificador erróneo. La utilización de ambos procedimientos da lugar a la siguiente tabla de contingencia $M \times M$

Clasificador erróneo

Clasificador correcto

	A_1	A_2	.	.	A_j	.	.	A_M	
A_1	n_{11}	n_{12}	.	.	n_{1j}	.	.	n_{1M}	n_{1*}
A_2	n_{21}	n_{22}	.	.	n_{2j}	.	.	n_{2M}	n_{2*}
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
A_i	n_{i1}	n_{i2}	.	.	n_{ij}	.	.	n_{iM}	n_{i*}
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
A_M	n_{M1}	n_{M2}	.	.	n_{Mj}	.	.	n_{MM}	n_{M*}
	n_{*1}	n_{*2}	.	.	n_{*j}	.	.	n_{*M}	n

   donde n_{ij} es el número de unidades de la muestra cuya verdadera categoría es A_i y son clasificadas en la categoría A_j, y

$$n_{i*} = \sum_{j=1}^{M} n_{ij}, \qquad n_{*j} = \sum_{i=1}^{M} n_{ij}, \qquad n = \sum_{i=1}^{M} \sum_{j=1}^{M} n_{ij}.$$

2. Posteriormente, se obtiene una muestra de $N - n$ unidades adicionales y se clasifica por el procedimiento que puede producir errores, $Y_{n+1}^0, ..., Y_N^0$. Denotamos por

$$m_j = \sum_{k=n+1}^{N} I_{\left(Y_k^0 = j\right)}, \qquad j = 1, ..., M$$

el número de unidades que según este clasificador se han asignado a la categoría A_j, $j = 1, ..., M$, entre las $N - n$ observadas y por $(m_1, ..., m_M)^t$ el vector de frecuencias asociado a $Y^0_{n+1}, ..., Y^0_N$.

En la primera etapa del muestreo doble, se tiene que de las n observaciones tomadas hay n_{ij} unidades, cuya verdadera categoría es A_i y una vez clasificadas han ido a parar a la categoría A_j. Consideremos n_{ij} para i fijo. La tabla de contingencia $M \times M$ dada anteriormente se puede ahora considerar como una tabla $2 \times M$, dependiendo si la verdadera categoría de una unidad es A_i o no A_i, es decir

$$
\begin{array}{ccccccc}
 & & \multicolumn{4}{c}{\overbrace{\hspace{12em}}^{Y^0}} & \\
 & & A_1 & A_2 & \cdot\ \cdot\ A_j\ \cdot\ \cdot & A_M & \\
\cline{3-6}
Y\,\{ & A_i & \multicolumn{1}{|c|}{n_{i1}} & \multicolumn{1}{c|}{n_{i2}} & \multicolumn{1}{c|}{\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot} & \multicolumn{1}{c|}{n_{iM}} & n_{i*} \\
\cline{3-6}
 & \text{no } A_i & \multicolumn{1}{|c|}{n_{*1} - n_{i1}} & \multicolumn{1}{c|}{n_{*2} - n_{i2}} & \multicolumn{1}{c|}{\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot} & \multicolumn{1}{c|}{n_{*M} - n_{iM}} & n - n_{i*} \\
\cline{3-6}
 & & n_{*1} & n_{*2} & \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot & n_{*M} & n
\end{array}
$$

La distribución de la variable aleatoria $2M$-dimensional

$$(n_{i1}, ..., n_{iM}, n_{*1} - n_{i1}, ..., n_{*M} - n_{iM})$$

es una multinomial con parámetros

$$(n; \theta_{i1} p_i, ..., \theta_{iM} p_i, \pi_1 - \theta_{i1} p_i, ..., \pi_M - \theta_{iM} p_i)$$

ya que

$$
\begin{aligned}
P\left(Y = i, Y^0 = j\right) &= P\left(Y^0 = j / Y = i\right) P\left(Y = i\right) = \theta_{ij} p_i \\
P\left(Y \neq i, Y^0 = j\right) &= P\left(Y^0 = j\right) - P\left(Y = i, Y^0 = j\right) = \pi_j - \theta_{ij} p_i.
\end{aligned}
$$

En la segunda etapa, utilizando el clasificador erróneo, se observan m_j unidades (de entre las $N - n$) en la clase A_j. Por tanto $(m_1, ..., m_M)$ se distribuye según una multinomial con parámetros

$$(N - n; \pi_1, ..., \pi_M)$$

ya que

$$P\left(Y^0 = j\right) = \pi_j, \qquad j = 1, ..., M.$$

Como la segunda muestra es independiente de la primera, la función de verosimilitud conjunta es proporcional a

$$L\left(p_i,\theta\right)\approx\prod_{j=1}^{M}\left(\theta_{ij}p_i\right)^{n_{ij}}\left(\pi_j-\theta_{ij}p_i\right)^{n_{*j}-n_{ij}}\prod_{j=1}^{M}\left(\pi_j\right)^{m_j}$$

donde $\theta=\left(\theta_{ij}\right)_{i,j=1,\ldots,M}$.

Sea $\lambda_{ij}=\dfrac{\theta_{ij}p_i}{\pi_j}$, entonces se tiene

$$\begin{aligned}
L &= L\left(\lambda_{i1},...,\lambda_{iM},\pi_1,..,\pi_M\right)\\
&= \prod_{j=1}^{M}\left(\lambda_{ij}\right)^{n_{ij}}\left(\pi_j\right)^{n_{ij}}\left(\pi_j-\lambda_{ij}\pi_j\right)^{n_{*j}-n_{ij}}\prod_{j=1}^{M}\left(\pi_j\right)^{m_j}\\
&= \prod_{j=1}^{M}\left(\lambda_{ij}\right)^{n_{ij}}\left(\pi_j\right)^{n_{*j}}\left(1-\lambda_{ij}\right)^{n_{*j}-n_{ij}}\prod_{j=1}^{M}\left(\pi_j\right)^{m_j}\\
&= \prod_{j=1}^{M}\left(\lambda_{ij}\right)^{n_{ij}}\left(1-\lambda_{ij}\right)^{n_{*j}-n_{ij}}\prod_{j=1}^{M}\left(\pi_j\right)^{m_j+n_{*j}}.
\end{aligned}$$

Tomando logarítmos se tiene,

$$\begin{aligned}
\log L &= \sum_{j=1}^{M}n_{ij}\log\lambda_{ij}+\sum_{j=1}^{M}\left(n_{*j}-n_{ij}\right)\log\left(1-\lambda_{ij}\right)+\sum_{j=1}^{M}\left(m_j+n_{*j}\right)\log\pi_j\\
&= \sum_{j=1}^{M}n_{ij}\log\lambda_{ij}+\sum_{j=1}^{M}\left(n_{*j}-n_{ij}\right)\log\left(1-\lambda_{ij}\right)+\sum_{j=1}^{M-1}\left(m_j+n_{*j}\right)\log\pi_j\\
&\quad+\left(m_M+n_{*M}\right)\log\left(1-\sum_{j=1}^{M}\pi_j\right).
\end{aligned}$$

Derivando respecto a λ_{ij} e igualando a cero se llega a,

$$\frac{\partial\log L}{\partial\lambda_{ij}}=\frac{n_{ij}}{\lambda_{ij}}-\frac{n_{*j}-n_{ij}}{1-\lambda_{ij}}=0,$$

luego

$$\hat{\lambda}_{ij}=\frac{n_{ij}}{n_{*j}}.$$

Derivando respecto a π_j e igualando a cero se tiene,

$$\frac{\partial \log L}{\partial \pi_j} = \frac{n_{ij}}{\lambda_{ij}} \frac{\partial \lambda_{ij}}{\partial \pi_j} + \frac{n_{*j} - n_{ij}}{1 - \lambda_{ij}} \frac{\partial (1 - \lambda_{ij})}{\partial \pi_j} + \frac{m_j + n_{*j}}{\pi_j} - \frac{m_M + n_{*M}}{\pi_M}.$$

Al ser,

$$\lambda_{ij} = \frac{\theta_{ij} p_i}{\pi_j}$$

se tiene,

$$\frac{\partial \lambda_{ij}}{\partial \pi_j} = -\frac{\theta_{ij} p_i}{\pi_j^2} = -\frac{\lambda_{ij}}{\pi_j}$$

y como

$$\hat{\lambda}_{ij} = \frac{n_{ij}}{n_{*j}}$$

se llega a

$$\begin{aligned} \frac{\partial \log L}{\partial \pi_j} &= -\frac{n_{ij}}{\lambda_{ij}} \frac{n_{ij}}{n_{*j}} \frac{1}{\pi_j} + \frac{n_{*j} - n_{ij}}{1 - \lambda_{ij}} \frac{n_{ij}}{n_{*j}} \frac{1}{\pi_j} + \frac{m_j + n_{*j}}{\pi_j} - \frac{m_M + n_{*M}}{\pi_M} \\ &= -\frac{n_{ij} n_{*j}}{n_{ij}} \frac{n_{ij}}{n_{*j}} \frac{1}{\pi_j} + \frac{(n_{*j} - n_{ij})\, n_{*j}}{n_{*j} - n_{ij}} \frac{n_{ij}}{n_{*j}} \frac{1}{\pi_j} + \frac{m_j + n_{*j}}{\pi_j} - \frac{m_M + n_{*M}}{\pi_M} = 0. \end{aligned}$$

De la anterior igualdad se tiene,

$$\frac{m_j + n_{*j}}{\pi_j} = \frac{m_M + n_{*M}}{\pi_M}$$

luego

$$\pi_M (m_j + n_{*j}) = \pi_j (m_M + n_{*M})$$

y sumando en j se llega a

$$\frac{1}{N} = \frac{\pi_M}{m_M + n_{*M}}$$

con lo cual se tiene que

$$\hat{\pi}_j = \frac{m_j + n_{*j}}{N}, \qquad j = 1, ..., M.$$

Como $\lambda_{ij} = \frac{\theta_{ij} p_i}{\pi_j}$ y $p_i = \sum_{j=1}^{M} \lambda_{ij} \pi_j$ la expresión de los estimadores $\hat{\theta}_{ij}$ y $\hat{p}_i$ viene dada por

$$\begin{aligned} \hat{\theta}_{ij} &= \frac{(m_j + n_{*j})\, n_{ij}}{N n_{*j} \hat{p}_i}, \qquad i, j = 1, ..., M \\ \hat{p}_i &= \sum_{j=1}^{M} \frac{(m_j + n_{*j})\, n_{ij}}{N n_{*j}}, \qquad i = 1, ..., M. \end{aligned}$$

Observación 5.1

Ahora se calculará la matriz de información de Fisher para i fijo. Los parámetros del modelo son

$$\left(\lambda_{i1},...,\lambda_{iM},\pi_1,...,\pi_{M-1}\right).$$

En lo sucesivo se utilizará la siguiente notación

$$\alpha_1=\lambda_{i1},\alpha_2=\lambda_{i2},...,\alpha_M=\lambda_{iM},\alpha_{M+1}=\pi_1,...,\alpha_{2M-1}=\pi_{M-1}.$$

Así el elemento (u,v) de la matriz de información de Fisher, $I_{uv}\left(\alpha_1,...,\alpha_{2M-1}\right)$ viene dado por

$$I_{uv}\left(\alpha_1,...,\alpha_{2M-1}\right)=E\left(-\frac{\partial^2\log L\left(p,\theta\right)}{\partial\alpha_u\partial\alpha_v}\right),\qquad u,v=1,...,2M-1.$$

Ahora se calculará esta expresión según los diferentes valores de los índices u y v.

- *Si $u\neq v$, $u,v=1,...,M$, se tiene*

$$\begin{aligned}\frac{\partial\log L\left(p,\theta\right)}{\partial\lambda_{iu}}&=\frac{n_{iu}}{\lambda_{iu}}-\frac{n_{*u}-n_{iu}}{1-\lambda_{iu}}\\ \frac{\partial^2\log L\left(p,\theta\right)}{\partial\lambda_{iu}\partial\lambda_{iv}}&=0\end{aligned}$$

  y

$$E\left(-\frac{\partial^2\log L\left(p,\theta\right)}{\partial\alpha_u\partial\alpha_v}\right)=E\left(-\frac{\partial^2\log L\left(p,\theta\right)}{\partial\lambda_{iu}\partial\lambda_{iv}}\right)=0$$

- *Si $u=v$, $u=1,...,M$, se tiene*

$$\frac{\partial^2\log L\left(p,\theta\right)}{\partial\lambda_{iu}^2}=-\frac{n_{iu}}{\lambda_{iu}^2}-\frac{n_{*u}-n_{iu}}{\left(1-\lambda_{iu}\right)^2}.$$

Al ser, como ya se vio anteriormente, la distribución conjunta de

$$\left(n_{i1},...,n_{iM},n_{*1}-n_{i1},...,n_{*M}-n_{iM}\right)$$

una multinomial con parámetros

$$\left(n;\theta_{i1}p_i,...,\theta_{iM}p_i,\pi_1-\theta_{i1}p_i,...,\pi_M-\theta_{iM}p_i\right)$$

se tiene que n_{ij}, a i fijo, es una binomial de parámetros $(n; \theta_{ij} p_i)$, con lo cual

$$E(n_{ij}) = n\theta_{ij} p_i = n\lambda_{ij}\pi_j,$$

y $n_{*j} - n_{ij}$, a i fijo, es también una binomial, pero ahora los parámetros son $(n; \pi_j - \theta_{ij} p_i)$, con lo cual

$$E(n_{*j} - n_{ij}) = n(\pi_j - \theta_{ij} p_i) = n\pi_j (1 - \lambda_{ij}).$$

Así, se tiene

$$\begin{aligned} I_{uu}(\alpha_1, ..., \alpha_{2M-1}) &= E\left(-\frac{\partial^2 \log L(p,\theta)}{\partial \alpha_u^2}\right) = E\left(-\frac{\partial^2 \log L(p,\theta)}{\partial \lambda_{iu}^2}\right) \\ &= \frac{n\lambda_{iu}\pi_u}{\lambda_{iu}^2} + \frac{n\pi_u (1-\lambda_{iu})}{(1-\lambda_{iu})^2} \\ &= n\pi_u \left(\frac{1}{\lambda_{iu}} + \frac{1}{1-\lambda_{iu}}\right) \\ &= \frac{n\pi_u}{\lambda_{iu}(1-\lambda_{iu})}. \end{aligned}$$

De forma análoga se obtiene,

- *Si $u = 1, ..., M$, $v = M+1, ..., 2M-1$, entonces*

$$I_{uv}(\alpha_1, ..., \alpha_{2M-1}) = E\left(-\frac{\partial^2 \log L(p,\theta)}{\partial \lambda_{iu} \partial \pi_v}\right) = 0.$$

- *Si $u = M+1, ..., 2M-1$, $v = 1, ..., M$, entonces*

$$I_{uv}(\alpha_1, ..., \alpha_{2M-1}) = E\left(-\frac{\partial^2 \log L(p,\theta)}{\partial \pi_u \partial \lambda_{iv}}\right) = 0.$$

- *Si $u \neq v$, $u, v = M+1, ..., 2M-1$, entonces*

$$I_{uv}(\alpha_1, ..., \alpha_{2M-1}) = E\left(-\frac{\partial^2 \log L(p,\theta)}{\partial \pi_u \partial \pi_v}\right) = \frac{N}{\pi_M}.$$

- *Si $u = v$, $u, v = M+1, ..., 2M-1$, entonces*

$$I_{uv}(\alpha_1, ..., \alpha_{2M-1}) = E\left(-\frac{\partial^2 \log L(p,\theta)}{\partial \pi_u \partial \pi_v}\right) = \frac{N}{\pi_u} + \frac{N}{\pi_M}.$$

En definitiva la matriz de información de Fisher, viene dada mediante

$$I_{uv}(\alpha_1,...,\alpha_{2M-1}) = \begin{pmatrix} ndiag(T) & 0 \\ 0 & NA \end{pmatrix}$$

siendo $diag(T)$ *una matriz* $M \times M$ *cuyos elementos en la diagonal principal son*

$$\frac{\pi_u}{\lambda_{iu}(1-\lambda_{iu})} \qquad u = 1,...,M$$

y $A = (a_{uv})_{u,v=M+1,...,2M-1}$ *cuyos elementos* a_{uv} *vienen dados mediante*

$$a_{uv} = \begin{cases} \dfrac{1}{\pi_u} + \dfrac{1}{\pi_M} & si \quad u = v \\ \\ \dfrac{1}{\pi_M} & si \quad u \neq v \end{cases}.$$

Es claro a partir de este resultado que la variable aleatoria

$$\sqrt{n}\left(\left(\hat{\lambda}_{i1},...,\hat{\lambda}_{iM},\hat{\pi}_1,...,\hat{\pi}_M\right) - (\lambda_{i1},...,\lambda_{iM},\pi_1,...,\pi_M)\right)$$

converge en Ley a una normal $2M$-dimensional con vector de medias cero y matriz de varianzas covarianzas dada mediante

$$\begin{pmatrix} diag(T)^{-1} & 0 \\ 0 & \frac{n}{N}\Sigma_\pi \end{pmatrix}$$

siendo

$$\Sigma_\pi = diag(\pi) - \pi^t\pi.$$

El siguiente resultado que se establecerá es importante ya que permite obtener los resultados asintóticos más importantes que se obtienen en este capítulo.

Teorema 5.1

Sea

$$\hat{p}_i = \sum_{j=1}^{M} \frac{(m_j + n_{*j})\, n_{ij}}{N n_{*j}}, \qquad i = 1,...,M$$

el estimador de máxima verosimilitud de p_i*,* $i = 1,...,M$*, bajo la hipótesis del muestreo doble. Supuesto que* n/N *tiende a* $f > 0$*, cuando* N *tiende a infinito, entonces*

$$\sqrt{N}\,(\hat{p}_1 - p_1,...,\hat{p}_M - p_M) \underset{N\to\infty}{\overset{L}{\to}} N(0,\Sigma^*)$$

siendo $\Sigma^* = (\sigma_{ij})_{i,j=1,...,M}$ *con*

$$\sigma_{ij} = \begin{cases} \dfrac{p_i q_i}{f}\left(1-(1-f)\,K_i\right) & si \quad i=j \\ \\ \left(1-\dfrac{1}{f}\right)\sum\limits_{k=1}^{M}\lambda_{ik}\lambda_{jk}\pi_k - p_i p_j & si \quad i\neq j \end{cases}.$$

Además,

$$\begin{aligned} q_i &= 1-p_i \\ K_i &= \left(Corr\left(I_{(Y=i)}, E\left(I_{(Y=i)}/Y^0\right)\right)\right)^2 \\ &= \frac{p_i}{q_i}\left(\sum_{k=1}^{M}\frac{\theta_{ik}^2}{\pi_k}-1\right) \\ \lambda_{ij} &= \frac{p_i\theta_{ij}}{\pi_j} = E\left(I_{(Y=i)}/Y^0=j\right). \end{aligned}$$

Demostración

En primer lugar se calculará la varianza de las $\hat{p}_i$. Es decir σ_{ii}. Al ser,

$$\hat{p}_i = \sum_{j=1}^{M}\hat{\lambda}_{ij}\hat{\pi}_j, \qquad i=1,...,M$$

se tiene que el desarrollo de Taylor de primer orden de $\hat{p}_i$ en p_i viene dado por

$$\hat{p}_i - p_i = \sum_{u=1}^{M}\pi_u\left(\lambda_{iu}-\hat{\lambda}_{iu}\right) + \sum_{u=1}^{M}\lambda_{iu}\left(\pi_u-\hat{\pi}_u\right) + o\left(\left\|\lambda-\hat{\lambda}\right\| + \|\pi-\hat{\pi}\|\right)$$

con $\lambda = (\lambda_{i1},...,\lambda_{iM})^t$ y $\pi = (\pi_1,...,\pi_M)^t$.

Por tanto la varianza asintótica de $\hat{p}_i$, teniendo en cuenta las relaciones existentes entre las diferentes covarianzas, viene dada por

$$Va\left(\hat{p}_i\right) = \sum_{u=1}^{M}\pi_u^2 Va\left(\hat{\lambda}_{iu}\right) + \sum_{u=1}^{M}\lambda_{iu}^2 Va\left(\hat{\pi}_u\right) + \sum_{\substack{u=1 \\ u\neq v}}^{M}\sum_{v=1}^{M} Cova\left(\hat{\pi}_u,\hat{\pi}_v\right)\lambda_{iu}\lambda_{iv}.$$

Ahora bien, al ser

$$Va\left(\hat{\lambda}_{iu}\right) = \frac{\lambda_{iu}\left(1-\lambda_{iu}\right)}{n\pi_u}$$

$$\begin{aligned} Va\left(\hat{\pi}_u\right) &= \frac{1}{N}\pi_u\left(1-\pi_u\right) \qquad u=1,...,M \\ Cova\left(\hat{\pi}_u,\hat{\pi}_v\right) &= -\frac{1}{N}\pi_u\pi_v \end{aligned}$$

y teniendo en cuenta que $\lambda_{iu} = p_i\theta_{iu}/\pi_u$, se tiene

$$Va\left(\hat{p}_i\right) = \frac{p_i}{n} - \frac{p_i^2}{n}\sum_{u=1}^{M}\frac{\theta_{iu}^2}{\pi_u} + \frac{p_i^2}{N}\sum_{u=1}^{M}\frac{\theta_{iu}^2}{\pi_u} - \frac{p_i^2}{N}\sum_{u=1}^{M}\theta_{iu}^2 - \frac{p_i^2}{N}\sum_{\substack{u=1\\u\neq v}}^{M}\sum_{v=1}^{M}\theta_{iu}\theta_{iv}.$$

Ahora bien, llamando

$$C_1 = \frac{p_i}{n} - \frac{p_i^2}{n}\sum_{u=1}^{M}\frac{\theta_{iu}^2}{\pi_u}$$

y

$$C_2 = \frac{p_i^2}{N}\left(\sum_{u=1}^{M}\frac{\theta_{iu}^2}{\pi_u} - \sum_{u=1}^{M}\theta_{iu}^2 - \sum_{\substack{u=1\\u\neq v}}^{M}\sum_{v=1}^{M}\theta_{iu}\theta_{iv}\right)$$

se tiene,

$$\begin{aligned} C_1 &= \frac{1}{n}\left(\left(p_iq_i + p_i^2\right) - p_i^2\sum_{u=1}^{M}\frac{\theta_{iu}^2}{\pi_u}\right) \\ &= \frac{1}{n}\left(p_iq_i\left(1 - \frac{p_i^2}{p_iq_i}\sum_{u=1}^{M}\frac{\theta_{iu}^2}{\pi_u} + \frac{p_i^2}{p_iq_i}\right)\right) \\ &= \frac{p_iq_i}{n}\left(1 - \frac{p_i}{q_i}\left(\sum_{u=1}^{M}\frac{\theta_{iu}^2}{\pi_u} - 1\right)\right) \\ &= \frac{p_iq_i}{n}\left(1-K_i\right) \end{aligned}$$

y

$$\begin{aligned} C_2 &= \frac{p_i^2}{N}\left(\sum_{u=1}^{M}\frac{\theta_{iu}^2}{\pi_u} - 1\right) \\ &= \frac{p_iq_i}{N}\left(\frac{p_i}{q_i}\left(\sum_{u=1}^{M}\frac{\theta_{iu}^2}{\pi_u} - 1\right)\right) \\ &= \frac{p_iq_i}{N}K_i. \end{aligned}$$

Luego,

$$\begin{aligned} Va\left(\hat{p}_i\right) &= \frac{p_i q_i}{n}\left(1-K_i\right)+\frac{p_i q_i}{N}K_i=\frac{p_i q_i}{n}\left(1-K_i+\frac{n}{N}K_i\right) \\ &= \frac{p_i q_i}{n}\left(1-K_i\left(1-\frac{n}{N}\right)\right)=\frac{p_i q_i}{n}\left(1-K_i\left(1-f\right)\right). \end{aligned}$$

En definitiva

$$Va\left(\sqrt{N}\hat{p}_i\right)=\frac{p_i q_i}{f}\left(1-K_i\left(1-f\right)\right)=\sigma_{ii}.$$

Para comprobar que

$$Cova\left(\sqrt{N}\hat{p}_i,\sqrt{N}\hat{p}_j\right)=\sigma_{ij}$$

hay que tener en cuenta que

$$Cova\left(\hat{\lambda}_{is}-\lambda_{is},\hat{\lambda}_{jm}-\lambda_{jm}\right)=\begin{cases} 0 & si \quad s\neq m \\ -\dfrac{\lambda_{is}\lambda_{js}}{\pi_s n} & si \quad s=m \end{cases}$$

y la relación existente entre las diferentes covarianzas.

Veamos esta última igualdad. Al ser,

$$\begin{aligned} h\left(x,y\right) &= \frac{x}{y}=h\left(x_0,y_0\right)+\left(\frac{\partial h\left(x,y\right)}{\partial x}\right)_{\left(x_0,y_0\right)}\left(x-x_0\right) \\ &\quad +\left(\frac{\partial h\left(x,y\right)}{\partial y}\right)_{\left(x_0,y_0\right)}\left(y-y_0\right)+o\left(\|x-x_0\|+\|y-y_0\|\right) \\ &= \frac{x_0}{y_0}+\frac{1}{y_0}\left(x-x_0\right)-\frac{x_0}{y_0^2}\left(y-y_0\right)+o\left(\|x-x_0\|+\|y-y_0\|\right) \end{aligned}$$

se tiene, tomando

$$x=n_{is} \qquad x_0=E\left[n_{is}\right]=n\lambda_{is}\pi_s$$

$$y=n_{*s} \qquad y_0=E\left[n_{*s}\right]=n\pi_s,$$

que

$$\hat{\lambda}_{is}=\frac{n_{is}}{n_{*s}}=\frac{n\lambda_{is}\pi_s}{n\pi_s}+\frac{1}{n\pi_s}\left(n_{is}-E\left[n_{is}\right]\right)-\frac{n\lambda_{is}\pi_s}{n^2\pi_s^2}\left(n_{*s}-E\left[n_{*s}\right]\right).$$

Sea $i \neq j$, entonces

$$
\begin{aligned}
Cova\left(\hat{\lambda}_{is}-\lambda_{is},\hat{\lambda}_{js}-\lambda_{js}\right) &= \frac{1}{n^2\pi_s^2}Cova\left(n_{is},n_{js}\right)-\frac{\lambda_{js}}{n^2\pi_s^2}Cova\left(n_{is},n_{*s}\right)\\
&\quad -\frac{\lambda_{is}}{n^2\pi_s^2}Cova\left(n_{js},n_{*s}\right)+\frac{\lambda_{js}\lambda_{is}}{n^2\pi_s^2}Va\left(n_{*s}\right)\\
&= \frac{1}{n^2\pi_s^2}\left(-n\lambda_{is}\pi_s\lambda_{js}\pi_s\right)\\
&\quad -\frac{\lambda_{js}}{n^2\pi_s^2}\left(Cova\left(n_{is},n_{is}\right)+\sum_{u\neq i}Cova\left(n_{is},n_{us}\right)\right)\\
&\quad -\frac{\lambda_{is}}{n^2\pi_s^2}Cova\left(n_{js},n_{*s}\right)+\frac{\lambda_{js}\lambda_{is}}{n^2\pi_s^2}n\pi_s\left(1-\pi_s\right)\\
&= -\frac{\lambda_{js}\lambda_{is}}{n}-\frac{\lambda_{js}}{n^2\pi_s^2}\left(n\lambda_{is}\pi_s\left(1-\lambda_{is}-\pi_s\right)\right.\\
&\quad \left.-\sum_{u\neq i}n\lambda_{is}\lambda_{us}\pi_s^2\right)-\frac{\lambda_{is}}{n^2\pi_s^2}Cova\left(n_{js},n_{*s}\right)\\
&\quad +\frac{\lambda_{js}\lambda_{is}}{n\pi_s}-\frac{\lambda_{js}\lambda_{is}}{n}\\
&= -\frac{\lambda_{js}\lambda_{is}}{n}-2\frac{\lambda_{js}}{n\pi_s}\left(\lambda_{is}-\lambda_{is}\pi_s\right)+\frac{\lambda_{js}\lambda_{is}}{n\pi_s}-\frac{\lambda_{js}\lambda_{is}}{n}\\
&= -\frac{\lambda_{js}\lambda_{is}}{n}-2\frac{\lambda_{js}\lambda_{is}}{n\pi_s}+2\frac{\lambda_{js}\lambda_{is}}{n}+\frac{\lambda_{js}\lambda_{is}}{n\pi_s}-\frac{\lambda_{js}\lambda_{is}}{n}\\
&= -\frac{\lambda_{js}\lambda_{is}}{n\pi_s}.
\end{aligned}
$$

Así se tiene que

$$
\begin{aligned}
Cova\left(\sqrt{N}\hat{p}_i,\sqrt{N}\hat{p}_j\right) &= N\sum_{s=1}^{M}Cova\left(\hat{\lambda}_{is}-\lambda_{is},\hat{\lambda}_{js}-\lambda_{js}\right)\pi_s^2\\
&\quad +N\sum_{s=1}^{M}\sum_{m=1}^{M}Cov\left(\hat{\pi}_s,\hat{\pi}_m\right)\lambda_{jm}\lambda_{is},
\end{aligned}
$$

con

$$\begin{aligned}
\sum_{s=1}^{M}\sum_{m=1}^{M} Cova\left(\hat{\pi}_s,\hat{\pi}_m\right)\lambda_{jm}\lambda_{is} &= -\sum_{\substack{s=1\\m\neq s}}^{M}\sum_{m=1}^{M}\frac{1}{N}\pi_s\pi_m\lambda_{jm}\lambda_{is}\\
&+\sum_{s=1}^{M}\frac{\pi_s\left(1-\pi_s\right)}{N}\lambda_{js}\lambda_{is}\\
&= -\sum_{\substack{s=1\\m\neq s}}^{M}\sum_{m=1}^{M}\frac{1}{N}\pi_s\pi_m\frac{p_j\theta_{jm}}{\pi_m}\frac{p_i\theta_{is}}{\pi_s}\\
&+\sum_{s=1}^{M}\frac{\pi_s\left(1-\pi_s\right)}{N}\lambda_{js}\lambda_{is}\\
&= -\frac{1}{N}p_ip_j\sum_{\substack{s=1\\m\neq s}}^{M}\theta_{is}\sum_{m=1}^{M}\theta_{jm}\\
&+\sum_{s=1}^{M}\frac{\pi_s\left(1-\pi_s\right)}{N}\lambda_{js}\lambda_{is}.
\end{aligned}$$

Teniendo en cuenta que

$$\sum_{\substack{s=1\\m\neq s}}^{M}\theta_{is}\sum_{m=1}^{M}\theta_{jm}=1-\sum_{m=1}^{M}\theta_{im}\theta_{jm}$$

se llega a

$$\begin{aligned}
\sum_{s=1}^{M}\sum_{m=1}^{M} Cova\left(\hat{\pi}_s,\hat{\pi}_m\right)\lambda_{jm}\lambda_{is} &= -\frac{1}{N}p_ip_j+\frac{1}{N}\sum_{m=1}^{M}\theta_{im}\theta_{jm}p_ip_j\\
&+\sum_{s=1}^{M}\frac{\pi_s\left(1-\pi_s\right)}{N}\lambda_{js}\lambda_{is}\\
&= -\frac{1}{N}p_ip_j+\frac{1}{N}\sum_{m=1}^{M}\lambda_{im}\pi_m\lambda_{jm}\pi_m\\
&+\frac{1}{N}\sum_{s=1}^{M}\left(\pi_s-\pi_s^2\right)\lambda_{js}\lambda_{is}
\end{aligned}$$

$$= \quad -\frac{1}{N}p_ip_j + \frac{1}{N}\sum_{s=1}^{M}\pi_s\lambda_{js}\lambda_{is}.$$

Así, se tiene

$$Cova\left(\sqrt{N}\hat{p}_i, \sqrt{N}\hat{p}_j\right) = -p_ip_j + \sum_{s=1}^{M}\pi_s\lambda_{js}\lambda_{is} - \frac{N}{n}\sum_{s=1}^{M}\pi_s\lambda_{js}\lambda_{is}.$$

■

Bajo la hipótesis nula

$$H_0 : p = p_0 = (p_{01}, ..., p_{0M})^t$$

Cheng y otros (1998) establecieron que

$$N\sum_{i=1}^{M-1}\sum_{j=1}^{M-1}(\hat{p}_i - p_{0i})\,\hat{\tau}_{ij}\,(\hat{p}_j - p_{0j}) \underset{N\to\infty}{\overset{L}{\longrightarrow}} \mathcal{X}^2_{M-1}$$

siendo $\Sigma^{-1} = (\tau_{ij})_{i,j=1,...,M-1}$, Σ la matriz obtenida al eliminar la última fila y la última columna de la matriz Σ^* con $\hat{\tau}_{ij}$ el estimador de máxima verosimilitud de τ_{ij} y $\hat{p} = (\hat{p}_1, ..., \hat{p}_M)^t$ el estimador de máxima verosimilitud de p con muestreo doble.

En el apartado siguiente se presenta una nueva familia de estadísticos de contraste, basados en las (h,φ)-divergencias ponderadas para abordar aquellas situaciones en las que se tengan ponderaciones sobre los elementos A_i, $i = 1, ..., M$, de la partición $\mathcal{A}$ de $\mathcal{X}$. Es decir, se trata de presentar una familia de estadísticos de contraste que sustituya a la familia

$$T_1 = 2nWD^h_\varphi(\hat{p}, p_0)$$

cuando p se estima a través de máxima verosimilitud, pero utilizando un muestreo doble.

5.3. Bondad de ajuste con datos mal clasificados

Teorema 5.2

Sea $\hat{p} = (\hat{p}_1, ..., \hat{p}_M)^t$ *obtenido por el procedimiento de muestreo doble de*

Tenenbein, siendo

$$\hat{p}_i = \sum_{j=1}^{M} \frac{(m_j + n_{*j})\, n_{ij}}{N n_{*j}}, \qquad i = 1, ..., M$$

con $n/N \underset{N\to\infty}{\longrightarrow} f > 0$. *Bajo la hipótesis nula*

$$H_0 : p = p_0 = (p_{01}, ..., p_{0M})^t$$

y para Σ^* *la matriz de varianzas covarianzas asintótica del vector aleatorio* $\sqrt{N}\,(\hat{p} - p_0)$, *dada en el apartado anterior, entonces*

a) *Si* $q \neq p_0$, *se tiene*

$$\sqrt{N}\left(WD_\varphi^h(\hat{p}, q) - WD_\varphi^h(p_0, q)\right) \underset{N\to\infty}{\overset{L}{\longrightarrow}} N\left(0, \sigma^2\right)$$

donde

$$\sigma^2 = T\Sigma^* T^t \tag{5.1}$$

y T *la definida en el Capítulo 3.*

b) *Si* $q = p_0$, *se tiene*

$$2NWD_\varphi^h(\hat{p}, p_0) \underset{N\to\infty}{\overset{L}{\longrightarrow}} \sum_{i=1}^{r} \beta_i Z_i^2$$

donde las Z_i *son variables aleatorias independientes normales de media cero y varianza uno, los* β_i *son los autovalores no nulos de* $A\Sigma^*$, *siendo* A *la matriz definida en el Corolario 3.4,* $A = (a_{ij})_{i,j=1,\dots,M}$ *con*

$$a_{ij} = \begin{cases} 0 & si \quad i \neq j \\ \sum\limits_{a=1}^{\Lambda} \eta_a h'_a(0)\, \varphi''(1) \dfrac{u_i}{p_i \sum\limits_{i=1}^{M} u_i p_i} & si \quad i = j \end{cases}$$

y $r = rango\,(\Sigma^* A \Sigma^*)$.

Demostración

a) El desarrollo en serie de Taylor de primer orden de la función $WD_\varphi^h(\hat{p},q)$, viene dado por

$$\sqrt{N}\left(WD_\varphi^h(\hat{p},q)-WD_\varphi^h(p_0,q)\right)=\sqrt{N}T(\hat{p}-q)^t+\sqrt{N}o(\|\hat{p}-q\|).$$

Al ser,

$$\sqrt{N}(\hat{p}-q)\underset{N\to\infty}{\overset{L}{\to}}N(0,\Sigma^*),$$

se tiene

$$\sqrt{N}o(\|\hat{p}-q\|)=o_p(1).$$

El resultado, ahora, se sigue de forma inmediata.

b) El desarrollo en serie de Taylor de segundo orden de la función $WD_\varphi^h(\hat{p},q)$, viene dado por

$$WD_\varphi^h(\hat{p},q)-WD_\varphi^h(p_0,q)=\frac{1}{2}(\hat{p}-p_0)^t A(\hat{p}-p_0)+o\left(\|\hat{p}-p_0\|^2\right)$$

con A la matriz dada en el enunciado.

Como $WD_\varphi^h(p_0,q)=0$ si $q=p_0$ y

$$\sqrt{N}(\hat{p}-p_0)\underset{N\to\infty}{\overset{L}{\to}}N(0,\Sigma^*),$$

se tiene

$$No\left(\|\hat{p}-p_0\|^2\right)=o_p(1).$$

Aplicando la Observación 3.1 del Capítulo 3, se sigue que

$$2NWD_\varphi^h(\hat{p},p_0)\underset{N\to\infty}{\overset{L}{\to}}\sum_{i=1}^{r}\beta_i Z_i^2$$

donde las Z_i son variables aleatorias independientes normales de media cero y varianza uno, los β_i son los autovalores no nulos de $A\Sigma^*$ y $r=rango(\Sigma^* A\Sigma^*)$.

■

Observación 5.2

A la vista del apartado b) de este teorema para contrastar

$$H_0: p=p_0$$

cuando $p = (p_1, ..., p_M)^t$ *se ha estimado utilizando el muestreo doble y sobre los elementos* A_i*,* $i = 1, ..., M$*, de la partición* $\mathcal{A}$ *de* $\mathcal{X}$*, existen ponderaciones, se deberá rechazar la hipótesis nula si*

$$T^* = 2NWD_{\varphi}^{h}(\hat{p}, p_0) > t_\alpha$$

siendo t_α *de forma que*

$$P\left(\sum_{i=1}^{r} \beta_i Z_i^2 > t_\alpha\right) = \alpha.$$

Observación 5.3

Dado un punto $q \neq p_0$*, el apartado a) del teorema anterior permite obtener la potencia en* q*. Esta viene dada mediante,*

$$\beta_N(q) = 1 - \Phi_N\left(\frac{t_\alpha - 2NWD_{\varphi}^{h}(p_0, q)}{2N^{1/2}\sigma}\right)$$

donde σ *viene dada en* (5.1) *y* $\Phi_N(x)$ *es una sucesión de funciones de distribución que tiende a la función de distribución,* Φ*, de una normal de media cero y varianza uno.*

Se puede obtener otra aproximación a la función de potencia cuando se consideran alternativas, que convergen a la hipótesis nula H_0, como se vio en el Capítulo 3,

$$H_{1,N} : p^{(N)} = p_0 + N^{-1/2}d$$

con $d = (d_1, ..., d_M)^t$, y $\sum_{i=1}^{M} d_i = 0$.

Teorema 5.3

Bajo las hipótesis alternativas

$$H_{1,N} : p^{(N)} = p_0 + N^{-1/2}d$$

con $d = (d_1, ..., d_M)^t$*, y* $\sum_{i=1}^{M} d_i = 0$*, se tiene*

$$2NWD_{\varphi}^{h}(\hat{p}, p_0) - \xi \underset{N\to\infty}{\overset{L}{\longrightarrow}} \sum_{i=1}^{r} \beta_i (Z_i + \omega_i)^2$$

siendo $r = rango\left(\Sigma^* A \Sigma^*\right)$, $\beta_1, ..., \beta_r$ *los autovalores positivos de* $A\Sigma^*$, A *y* Σ^* *las matrices dadas en el Teorema 5.2,* Z_i, $i = 1, ..., r$, *variables aleatorias independientes normales de media cero y varianza uno,* $\omega = \Lambda^{-1} R^t S^t A d$, $\xi = d^t A d - \omega^t \Lambda \omega$ *donde* $\Lambda = diag\left(\beta_1, ..., \beta_r\right)$, S^t *es una raíz arbitraria de* Σ^* *y* R *es la correspondiente matriz de autovectores de* $S^t A S$.

Demostración

Se tiene,

$$\begin{aligned} \sqrt{N}\left(\hat{p} - p_0\right) &= \sqrt{N}\left(\hat{p} - p^{(N)}\right) + \sqrt{N}\left(p^{(N)} - p_0\right) \\ &= \sqrt{N}\left(\hat{p} - p^{(N)}\right) + d. \end{aligned}$$

Entonces al ser, bajo $H_{1,N}$,

$$\sqrt{N}\left(\hat{p} - p^{(N)}\right) \underset{N\to\infty}{\overset{L}{\to}} N\left(0, \Sigma^*\right)$$

se llega a

$$\sqrt{N}\left(\hat{p} - p_0\right) \underset{N\to\infty}{\overset{L}{\to}} N\left(d, \Sigma^*\right).$$

Por otro lado la distribución asintótica de la familia de estadísticos

$$2NWD_{\varphi}^{h}\left(\hat{p}, p_0\right)$$

y de la forma cuadrática

$$\sqrt{N}\left(\hat{p} - p_0\right)^t A \sqrt{N}\left(\hat{p} - p_0\right)$$

es la misma según se vio en el Teorema 5.2 parte b). El resultado ahora se sigue sin más que hacer uso de la Observación 3.3. ■

Observación 5.4

Usando el resultado del teorema previo se obtiene una aproximación de la función de potencia en $p^{(N)} = p_0 + N^{-1/2} d$ *mediante la expresión*

$$\beta_N\left(p^{(N)}\right) = 1 - G\left(t_\alpha\right)$$

donde G *es la función de distribución de la variable aleatoria* $\sum_{i=1}^{r} \beta_i \left(Z_i + \omega_i\right)^2 + \xi$.

5.4. El caso binomial con errores de clasificación

En esta sección se van a particularizar los resultados obtenidos en la sección anterior, para el caso binomial. En esta situación la hipótesis nula es

$$H_0 : p = p_0$$

donde p es la probabilidad de tener el resultado uno y $q = 1-p$ es la probabilidad de obtener el resultado cero.

En este caso se utilizará la siguiente notación para designar a las probabilidades de clasificación errónea,

$$\begin{aligned} \theta &= P\left(Y^0 = 0/Y = 1\right) = \theta_{10} \\ 1-\theta &= P\left(Y^0 = 1/Y = 1\right) = \theta_{11} \\ \psi &= P\left(Y^0 = 1/Y = 0\right) = \theta_{01} \\ 1-\psi &= P\left(Y^0 = 0/Y = 0\right) = \theta_{00}. \end{aligned}$$

Es inmediato que

$$\begin{aligned} \pi &= P\left(Y^0 = 1\right) = p\left(1-\theta\right) + q\psi \\ 1-\pi &= P\left(Y^0 = 0\right) = p\theta + q\left(1-\psi\right). \end{aligned}$$

Al igual que en el caso general se denotará por n_{ij} el número de unidades cuya verdadera categoría es i y una vez clasificadas han ido a parar a la categoría j, $i, j = 0, 1$, y

$$m_k = \sum_{j=n+1}^{N} I_{\left(Y_j^0 = k\right)}, \qquad k = 0, 1.$$

En este contexto los estimadores de máxima verosimilitud de las probabilidades p, θ y ψ son de la forma

$$\begin{aligned} \hat{p} &= \frac{n_{11}}{n_{*1}}\frac{m_1 + n_{*1}}{N} + \frac{n_{10}}{n_{*0}}\frac{m_0 + n_{*0}}{N} \\ \hat{\theta} &= \frac{n_{10}}{n_{*0}}\frac{m_0 + n_{*0}}{N\hat{p}} \\ \hat{\psi} &= \frac{n_{01}}{n_{*1}}\frac{m_1 + n_{*1}}{N\left(1-\hat{p}\right)}. \end{aligned}$$

En este caso la matriz A, como ya se vio en el Teorema 3.8, viene dada por

$$A = L \begin{pmatrix} \dfrac{u_1}{p_0\left(u_1 p_0 + u_2\left(1-p_0\right)\right)} & 0 \\ 0 & \dfrac{u_2}{\left(1-p_0\right)\left(u_1 p_0 + u_2\left(1-p_0\right)\right)} \end{pmatrix}$$

con

$$L = \sum_{a=1}^{\Lambda} \eta_a h_a'\left(0\right)\varphi_a''\left(1\right).$$

Por otro lado se tiene,

$$\Sigma^* = \left(\sigma_{ij}\right)_{i,j=1,2},$$

con

$$\sigma_{ij} = \begin{cases} \dfrac{p_i q_i}{f}\left(1-\left(1-f\right)K_i\right) & si \quad i=j \\ \\ \left(1-\dfrac{1}{f}\right)\sum\limits_{k=1}^{M}\lambda_{ik}\lambda_{jk}\pi_k - p_i p_j & si \quad i \neq j \end{cases}.$$

Ahora se obtendrán las expresiones de σ_{11} y σ_{12} en el caso binomial. La expresión de σ_{11} viene dada por

$$\sigma_{11} = \frac{p\left(1-p\right)}{f}\left(1-\left(1-f\right)K_1\right)$$

con

$$K_1 = \frac{p}{1-p}\left\{\frac{\theta_{11}^2}{\pi_1} + \frac{\theta_{12}^2}{\pi_2} - 1\right\}.$$

Seguidamente se obtendrá la expresión de K_1 cuando $p = p_0$

$$\begin{aligned} K_1 &= \frac{p_0}{1-p_0}\left\{\frac{\left(1-\theta\right)^2}{\pi} + \frac{\theta^2}{1-\pi} - 1\right\} \\ &= \frac{p_0}{1-p_0}\left\{\frac{\left(1-\theta\right)^2\left(1-\pi\right) + \theta^2\pi - \pi\left(1-\pi\right)}{\pi\left(1-\pi\right)}\right\} \\ &= \frac{p_0}{\left(1-p_0\right)\pi\left(1-\pi\right)}\left\{\left(1-\theta\right)^2\left(1-\pi\right) + \theta^2\pi - p_0 q_0\left(1-\theta-\psi\right)^2\right. \\ &\quad \left. - p_0\left(1-\theta\right)\theta - q_0\left(1-\psi\right)\psi\right\} \end{aligned}$$

$$\begin{aligned}
&= \frac{p_0}{(1-p_0)\,\pi\,(1-\pi)} \left\{(1-\theta)\left((1-\theta)(1-\pi)-p_0\theta\right)+\theta^2\pi\right. \\
&\left. -p_0q_0\,(1-\theta-\psi)^2 - q_0\,(1-\psi)\,\psi\right\} \\
&= \frac{p_0}{(1-p_0)\,\pi\,(1-\pi)} \left\{(1-\theta)\left((1-\theta)\left(p_0\theta+q_0\,(1-\psi)\right)-p_0\theta\right)\right. \\
&\left. +\theta^2\left(p_0\,(1-\theta)+q_0\psi\right) - p_0q_0\,(1-\theta-\psi)^2 - q_0\,(1-\psi)\,\psi\right\} \\
&= \frac{p_0}{(1-p_0)\,\pi\,(1-\pi)} \left\{(1-\theta)\big((1-\theta)\,\theta p_0+(1-\theta)\,q_0\,(1-\psi)-p_0\theta+\theta^2p_0\big)\right. \\
&\left. +\theta^2q_0\psi - p_0q_0\,(1-\theta-\psi)^2 - q_0\,(1-\psi)\,\psi\right\} \\
&= \frac{p_0}{(1-p_0)\,\pi\,(1-\pi)} \left\{(1-\theta)^2\left(\theta p_0+q_0\,(1-\psi)-p_0\theta\right)\right. \\
&\left. +q_0\psi\left(\theta^2-(1-\psi)\right) - p_0q_0\,(1-\theta-\psi)^2\right\} \\
&= \frac{p_0}{(1-p_0)\,\pi\,(1-\pi)} \left\{(1-\theta)^2\,q_0\,(1-\psi)+q_0\psi\left(\theta^2-(1-\psi)\right)\right. \\
&\left. -p_0q_0\,(1-\theta-\psi)^2\right\} \\
&= \frac{p_0q_0}{(1-p_0)\,\pi\,(1-\pi)} \left\{(1-\theta)^2\,(1-\psi)+\psi\left(\theta^2-(1-\psi)\right)-p_0\,(1-\theta-\psi)^2\right\} \\
&= \frac{p_0}{\pi\,(1-\pi)} \left\{(1-\theta-\psi)^2 - p_0\,(1-\theta-\psi)^2\right\} \\
&= \frac{p_0q_0}{\pi\,(1-\pi)} \left\{(1-\theta-\psi)^2\right\}.
\end{aligned}$$

Así, se tiene

$$\sigma_{11} = \frac{p_0q_0}{f}\left[1-(1-f)\,\frac{p_0q_0}{\pi\,(1-\pi)}\,(1-\theta-\psi)^2\right].$$

Ahora se calculará σ_{12}. Se tiene,

$$\sigma_{12} = \left(1 - \frac{1}{f}\right) \sum_{k=1}^{2} \lambda_{1k}\lambda_{2k}\pi_k - p_0 q_0.$$

Al ser,

$$\begin{aligned}
\lambda_{11} &= \frac{p_0}{\pi}(1-\theta) \\
\lambda_{21} &= \frac{q_0}{\pi}\psi \\
\lambda_{12} &= \frac{p_0}{1-\pi}\theta \\
\lambda_{22} &= \frac{q_0}{1-\pi}(1-\psi)
\end{aligned}$$

se tiene,

$$\lambda_{11}\lambda_{21}\pi + \lambda_{12}\lambda_{22}(1-\pi) = p_0 q_0 \left(1 - \frac{p_0 q_0}{\pi(1-\pi)}(1-\theta-\psi)^2\right).$$

Luego

$$\begin{aligned}
\sigma_{12} &= \left(1 - \frac{1}{f}\right) p_0 q_0 \left(1 - \frac{p_0 q_0}{\pi(1-\pi)}(1-\theta-\psi)^2\right) - p_0 q_0 \\
&= \frac{p_0 q_0}{f}\left[(1-f)\frac{p_0 q_0}{\pi(1-\pi)}(1-\theta-\psi)^2 - 1\right] = -\sigma_{11}
\end{aligned}$$

y

$$\begin{aligned}
\sigma_{21} &= \sigma_{12} \\
\sigma_{11} &= \sigma_{22}.
\end{aligned}$$

De esta forma la matriz Σ^*, viene dada por

$$\Sigma^* = \begin{pmatrix} Va(\hat{p}) & -Va(\hat{p}) \\ -Va(\hat{p}) & Va(\hat{p}) \end{pmatrix}$$

con

$$Va(\hat{p}) = \frac{p_0 q_0}{f}\left[1 - (1-f)\frac{p_0 q_0}{\pi(1-\pi)}(1-\theta-\psi)^2\right]$$

y $q_0 = 1 - p_0$.

Por lo tanto se tiene,

$$
\begin{aligned}
A\Sigma^* &= L \begin{pmatrix} \dfrac{u_1}{p_0\left(u_1 p_0 + u_2\left(1-p_0\right)\right)} & 0 \\ \\ 0 & \dfrac{u_2}{\left(1-p_0\right)\left(u_1 p_0 + u_2\left(1-p_0\right)\right)} \end{pmatrix} \Sigma^* \\
\\
&= L \begin{pmatrix} \dfrac{u_1 Va\left(\hat{p}\right)}{p_0\left(u_1 p_0 + u_2\left(1-p_0\right)\right)} & -\dfrac{u_1 Va\left(\hat{p}\right)}{p_0\left(u_1 p_0 + u_2\left(1-p_0\right)\right)} \\ \\ -\dfrac{u_2 Va\left(\hat{p}\right)}{\left(1-p_0\right)\left(u_1 p_0 + u_2\left(1-p_0\right)\right)} & \dfrac{u_2 Va\left(\hat{p}\right)}{\left(1-p_0\right)\left(u_1 p_0 + u_2\left(1-p_0\right)\right)} \end{pmatrix}.
\end{aligned}
$$

Los autovalores β_i de $A\Sigma^*$ son los mismos que los autovalores λ_i de $C\Sigma^*$ multiplicados por $L = \sum_{a=1}^{\Lambda} \eta_a h_a'\left(0\right) \varphi_a''\left(1\right)$, donde

$$
C\Sigma^* = \begin{pmatrix} \dfrac{u_1 Va\left(\hat{p}\right)}{p_0\left(u_1 p_0 + u_2\left(1-p_0\right)\right)} & -\dfrac{u_1 Va\left(\hat{p}\right)}{p_0\left(u_1 p_0 + u_2\left(1-p_0\right)\right)} \\ \\ -\dfrac{u_2 Va\left(\hat{p}\right)}{\left(1-p_0\right)\left(u_1 p_0 + u_2\left(1-p_0\right)\right)} & \dfrac{u_2 Va\left(\hat{p}\right)}{\left(1-p_0\right)\left(u_1 p_0 + u_2\left(1-p_0\right)\right)} \end{pmatrix}.
$$

Al ser,

$$
\begin{aligned}
\left|C\Sigma^* - \lambda I\right| &= \begin{vmatrix} \dfrac{u_1 Va\left(\hat{p}\right)}{p_0\left(u_1 p_0 + u_2\left(1-p_0\right)\right)} - \lambda & -\dfrac{u_1 Va\left(\hat{p}\right)}{p_0\left(u_1 p_0 + u_2\left(1-p_0\right)\right)} \\ \\ -\dfrac{u_2 Va\left(\hat{p}\right)}{\left(1-p_0\right)\left(u_1 p_0 + u_2\left(1-p_0\right)\right)} & \dfrac{u_2 Va\left(\hat{p}\right)}{\left(1-p_0\right)\left(u_1 p_0 + u_2\left(1-p_0\right)\right)} - \lambda \end{vmatrix} \\
\\
&= \frac{\lambda^2\left(u_1 p_0 + u_2 q_0\right) p_0 q_0 - \lambda\left(u_i q_0 Va\left(\hat{p}\right) + u_2 p_0 Va\left(\hat{p}\right)\right)}{p_0 q_0\left(u_1 p_0 + u_2 q_0\right)},
\end{aligned}
$$

se tiene que $\left|C\Sigma^* - \lambda I\right| = 0$, si y sólo si,

$$
\lambda\left[\lambda\left(u_1 p_0 + u_2 q_0\right) p_0 q_0 - Va\left(\hat{p}\right)\left(u_1 q_0 + u_2 p_0\right)\right] = 0
$$

con lo cual los autovalores que resultan son

$$\begin{aligned}\lambda_1 &= 0 \\ \lambda_2 &= \frac{Va\left(\hat{p}\right)\left(u_1 q_0 + u_2 p_0\right)}{\left(u_1 p_0 + u_2 q_0\right) p_0 q_0}.\end{aligned}$$

Así, el único autovalor no nulo β de $A\Sigma^*$ es

$$\begin{aligned}\beta &= L\frac{Va\left(\hat{p}\right)\left(u_1 q_0 + u_2 p_0\right)}{\left(u_1 p_0 + u_2 q_0\right) p_0 q_0} \\ &= \sum_{a=1}^{\Lambda} \eta_a h_a'\left(0\right)\varphi_a''\left(1\right)\frac{u_1 q_0 + u_2 p_0}{u_1 p_0 + u_2 q_0}\frac{1}{f}\left(1-\left(1-f\right)\frac{p_0 q_0}{\pi\left(1-\pi\right)}\left(1-\theta-\psi\right)^2\right).\end{aligned}$$

Entonces, se tiene,

$$\frac{2NWD_{\varphi}^{h}\left(\hat{p},p_0\right)}{\beta} \underset{N\to\infty}{\overset{L}{\longrightarrow}} \mathcal{X}_1^2.$$

Si llamamos

$$\hat{\beta} = \sum_{a=1}^{\Lambda} \eta_a h_a'\left(0\right)\varphi_a''\left(1\right)\frac{u_1 q_0 + u_2 p_0}{u_1 p_0 + u_2 q_0}\frac{1}{f}\left(1-\left(1-f\right)\frac{p_0 q_0}{\pi\left(1-\pi\right)}\left(1-\hat{\theta}-\hat{\psi}\right)^2\right),$$

entonces

$$\frac{2NWD_{\varphi}^{h}\left(\hat{p},p_0\right)}{\hat{\beta}} \underset{N\to\infty}{\overset{L}{\longrightarrow}} \mathcal{X}_1^2.$$

En este caso la potencia asintótica en $p^* \neq p_0$, viene dada por

$$\beta_N\left(p^*\right) = 1 - \Phi_{N(0,1)}\left(\frac{\mathcal{X}_{1,\alpha}^2\hat{\beta} - 2NWD_{\varphi}^{h}\left(p^*,p_0\right)}{2N^{1/2}\sigma_{p^*}}\right)$$

con $p^* = \left(p^*, 1-p^*\right)^t$ y σ_{p^*} la expresión definida en el Teorema 5.2 parte a) para p^*.

Observación 5.5

Si $\Lambda = 1$, $\eta_1 = 1$, $h_1\left(x\right) = x$, $\varphi_1\left(x\right) = \frac{1}{2}\left(x-1\right)^2$ *y* $u_1 = u_2$*; es decir, en el caso de las* φ*-divergencias, se tiene*

$$D_{\varphi}\left(\hat{p},p_0\right) = \frac{\left(\hat{p}-p_0\right)^2}{p_0 q_0}$$

y

$$\begin{aligned}\frac{2NWD_{\varphi}^{h}(\hat{p},p_0)}{\hat{\beta}} &= \frac{2ND_{\varphi}(\hat{p},p_0)}{\varphi''(1)\frac{1}{f}\left(1-(1-f)\frac{p_0q_0}{\pi(1-\pi)}(1-\theta-\psi)^2\right)} \\ &= \frac{n(\hat{p}-p_0)^2}{p_0q_0\left(1-(1-f)\frac{p_0q_0}{\pi(1-\pi)}(1-\theta-\psi)^2\right)}.\end{aligned}$$

Es decir, el estadístico obtenido por Cheng y otros (1998).

Se consideran ahora, las hipótesis alternativas de la forma

$$H_{1,N} : p^{(N)} = p_0 + N^{-1/2}d$$

donde $p_0 = (p_0, q_0)^t$, $q_0 = 1 - p_0$ y $d = (d_1, d_2)^t$ con $d_1 + d_2 = 0$.

Teorema 5.4

Bajo las hipótesis alternativas

$$H_{1,N} : p^{(N)} = p_0 + N^{-1/2}d$$

donde $p_0 = (p_0, q_0)^t$, $q_0 = 1 - p_0$ *y* $d = (d_1, d_2)^t$ *con* $d_1 = d$ *y* $d_2 = -d$, *se tiene*

$$\frac{2NWD_{\varphi}^{h}(\hat{p},p_0)}{\beta} \underset{N\to\infty}{\overset{L}{\longrightarrow}} \mathcal{X}_1^2(\delta)$$

donde $\mathcal{X}_1^2(\delta)$ *es una distribución Ji-cuadrado no central con un grado de libertad y parámetro de no centralidad*

$$\delta = \left(\frac{1}{f}\left(1-(1-f)\frac{p_0q_0}{\pi(1-\pi)}\left(1-\hat{\theta}-\hat{\psi}\right)^2\right)\right)^{-1}\frac{d^2}{p_0(1-p_0)},$$

β *el autovalor no nulo de la matriz* $A\Sigma^*$ *y* A *la matriz definida en el Teorema 5.2 parte b).*

Demostración

Se tiene,

$$\begin{aligned}\sqrt{N}(\hat{p}-p_0) &= \sqrt{N}\left(\hat{p}-p^{(N)}\right)+\sqrt{N}\left(p^{(N)}-p_0\right) \\ &= \sqrt{N}\left(\hat{p}-p^{(N)}\right)+d.\end{aligned}$$

Entonces al ser, bajo $H_{1,N}$,

$$\sqrt{N}\left(\hat{p}-p^{(N)}\right)\underset{N\to\infty}{\overset{L}{\to}}N\left(0,\Sigma^{*}\right)$$

se llega a

$$\sqrt{N}\left(\hat{p}-p_{0}\right)\underset{N\to\infty}{\overset{L}{\to}}N\left(d,\Sigma^{*}\right)$$

con

$$\Sigma^{*}=\begin{pmatrix} Va\left(\hat{p}\right) & -Va\left(\hat{p}\right)\\ -Va\left(\hat{p}\right) & Va\left(\hat{p}\right)\end{pmatrix}$$

y

$$Va\left(\hat{p}\right)=\frac{p_{0}\left(1-p_{0}\right)}{f}\left[1-\left(1-f\right)\frac{p_{0}\left(1-p_{0}\right)}{\pi\left(1-\pi\right)}\left(1-\theta-\psi\right)^{2}\right].$$

Por otro lado la distribución asintótica de la familia de estadísticos $2NWD_{\varphi}^{h}\left(\hat{p},p_{0}\right)$ y de la forma cuadrática $\sqrt{N}\left(\hat{p}-p_{0}\right)^{t}A\sqrt{N}\left(\hat{p}-p_{0}\right)$ con

$$A=\frac{\sum\limits_{a=1}^{\Lambda}\eta_{a}h_{a}'\left(0\right)\varphi_{a}''\left(1\right)}{u_{1}p_{0}+u_{2}\left(1-p_{0}\right)}\begin{pmatrix} u_{1}p_{0}^{-1} & 0\\ 0 & u_{2}\left(1-p_{0}\right)^{-1}\end{pmatrix}$$

y llamando $p_{10}=p_{0}$ y $p_{20}=1-p_{0}$, es la misma según se vio en el Teorema 5.2 parte b).

Entonces,

$$\frac{2NWD_{\varphi}^{h}\left(\hat{p},p_{0}\right)}{\beta}$$

con β el autovalor no nulo de $A\Sigma^{*}$ hallado anteriormente y la forma cuadrática $\sqrt{N}\left(\hat{p}-p_{0}\right)^{t}B\sqrt{N}\left(\hat{p}-p_{0}\right)$ con

$$B=C\left(u_{1}\left(1-p_{0}\right)+u_{2}p_{0}\right)^{-1}\begin{pmatrix} u_{1}p_{0}^{-1} & 0\\ 0 & u_{2}\left(1-p_{0}\right)^{-1}\end{pmatrix}$$

y

$$C=\left(\frac{1}{f}\left(1-\left(1-f\right)\frac{p_{0}q_{0}}{\pi\left(1-\pi\right)}\left(1-\hat{\theta}-\hat{\psi}\right)^{2}\right)\right)^{-1}$$

tienen la misma distribución asintótica.

Ahora,

$$\sqrt{N}B^{1/2}\left(\hat{p}-p_{0}\right)\underset{N\to\infty}{\overset{L}{\to}}N\left(B^{1/2}d,B^{1/2}\Sigma^{*}\left(B^{1/2}\right)^{t}\right)$$

siendo

$$\begin{aligned}
\Sigma^{**} &= B^{1/2}\Sigma^{*}\left(B^{1/2}\right)^{t} \\
&= \left(u_1\left(1-p_0\right)+u_2p_0\right)^{-1}\begin{pmatrix} u_1\left(1-p_0\right) & -\sqrt{u_1u_2p_0\left(1-p_0\right)} \\ -\sqrt{u_1u_2p_0\left(1-p_0\right)} & u_2p_0 \end{pmatrix}.
\end{aligned}$$

Como $\Sigma^{**}\Sigma^{**}=\Sigma^{**}$ y rango$(\Sigma^{**})=1$, se tiene que Σ^{**} es una proyección de rango 1. Así, aplicando el Lema de p. 63 en Ferguson (1996) se obtiene

$$\frac{2NWD_{\varphi}^{h}\left(\hat{p},p_0\right)}{\beta}\underset{N\rightarrow\infty}{\overset{L}{\rightarrow}}\mathcal{X}_1^2\left(\delta\right)$$

con

$$\delta=\left(dB^{1/2}\right)^{t}\left(dB^{1/2}\right)=\left(\frac{1}{f}\left(1-\left(1-f\right)\frac{p_0q_0}{\pi\left(1-\pi\right)}\left(1-\hat{\theta}-\hat{\psi}\right)^{2}\right)\right)^{-1}\frac{d^2}{p_0\left(1-p_0\right)}$$

y $d_1=d$, $d_2=-d$. ■

6

Bibliografía

ALI, S. M. y SILVEY, S. D. : A general class of coefficient of divergence of one distribution from another. *Journal of the Royal Statistical Society 28,* 131-142, 1966.

AGRESTI, A. : *Analysis of Ordinal Categorical Data.* new york, John wiley, 1966.

ATCHISON, J. y SILVEY, S. D. : Maximum likehood estimation of parameters subject to constraints. *Annals of Mathematical Statistics, 29,* 813-828, 1958.

BASU, A. y HARRIS, I. R. : Robust predeictive distributions for exponential families. *Biometrika, 81,* 790-794, 1994.

BASU, A., HARRIS, I. R. y BASU, S. : Tests of hypotheses in discrete models based on the penalized Hellinger distance. *Statist. Prob. Letters, 27,* 367-373, 1996.

BASU, I. R. y LINDSAY, B. G. : Minimum disparity estimation for continuous models. *Annals of the Institute Statistical Mathematics 46,* 683-705, 1994.

BASU, I. R. y SARKAR, B. G. : Minimum disparity estimation in the errors-in-variables model. *Statist. Prob. Letters, 20,* 69-73, 1994a.

BASU, I. R. y SARKAR, B. G. : The trade-off between robutness and efficiency and the efect of model smoothing in minimum disparity inference. *J. Statist. Comput. Simul. 50,* 173-185, 1994b.

BEHBOODIAN, J. : On a mixture of normal distributions. *Biometrika, 57,* 215-217, 1970.

BELIS, M. y GUIASU, S. : A quantitative-qualitative measure of information in cybernetics systems. *IEEE Transactions on Information Theory, IT-4,* 593-594, 1968.

BERAN, R. : Minimum Hellinger distance for parametric model. *Ann. Statist. 5,* 445-463, 1977.

BISHOP, Y. M. M., FIENBERG, S. E. y HOLLAND, P. W. : *Discrete Multivariate Analysis: Theory and Practice.* Cambridge, Mass.: MIT Press, 1975.

BOUCHON, B. : Useful Information and Questionnaires. *Information and Control, 16,* 36-51, 1976.

BROSS, I. : Misclassification in 2x2 tables. *Biometrics 10,* 478-486, 1954.

BURBEA, J. y RAO, C. R. : On the convexity of some divergence measures based on entropy functions. *IEEE Transactions on Information Theory, 28,* 489-495, 1982.

CHENG, K. F., HSUEH, H. M. y CHIEN, T. H. : Goodness of fit tests with misclassified data. *Comm. Statist. Theory Methods, 27,* 1379-1393, 1998.

COHEN, A. C. : Estimation in mixtures of two normal distributions. *Technometrics, 9,* 15-28, 1967.

COX, D. R. : *The analysis of Binary Data.* London, Methuen, 1970.

CRESSIE, N. y READ, T.R.C. : Multinomial goodness of fit test. *Journal of the Royal Statistical Society, Ser. B, 46,* 440-464, 1984.

CSISZAR, I. : Eine Informationstheoretische Ungleichung und ihre Anwendung auf den Bewis der Ergodizität on Markhoffschen Ketten. *Publ. Math. Inst. Hungar. Acad. Sci. 8,* 84-108, 1963.

DAY, N. E. : Estimating the components of a mixture of normal distributions. *Biometrika, 56, 3,* 463-474, 1969.

DIAMOND, E. L., MITRA, S. K. y ROY, S. N. : Asymptotic power and asymptotic independence in the statistical analysis of categorical data. *Bulletin of the International Institute, 37,* 3-23,1960.

DIAMOND, E. L. y LILIENFELD, A. M. : Effects of errors in classification and diagnosis in various types of epidemiologocal studies. *American Journal of Public Health, 10,* 2106-2110, 1962.

DIK, J. J. y GUNST, M. C. M. : The distribution of general quadratic forms in normal variables. *Statistica Neerlandica, 39,* 14-26, 1985.

ECKLER, A. R. : A Survey of Corage Problems Associated with Point and Area Targets. *Technometrics,11,* 561-589, 1969.

EMPTOZ, H. : Information de type β integrant un concept d ´utilite. *C. R. Acad. Sci. Paris, 282A,* 911-914, 1976.

FERGUSON, T. S. : *A Course in Large Sample Theory.* Chapman & Hall, 1996.

FIENBERG, S. E. : The use of Chi-squared statistics for categorical data problems. *Journal of Royal Statistical Society, B,* 41, 54-64, 1979.

FIENBERG, S. E. : *The Analysis of Cross-Clasified Categorical Data (2ª edición).* Cambrige, MA, the MIT Press, 1980.

FRASER, D. A. S. : *Nonparametric Methods in Statistics.* John Wiley and Sons. New York, 1957.

FREEMAN, D. H. : *Applied Categorical Data Analysis.* New York, Marcel Dekker, 1987.

FRYER, J. G. y ROBERSON, C. A. : A comparison of some methods for estimating mixed normal distributions. *Biometrika, 59, 3,* 639-648, 1972.

GIL, P. : Medidas de incertidumbre e información en problemas de decisión estadística. *Revista de la Real Academia. de Ciencias Exactas, Físicas y Naturales de Madrid, LXIX,* 549-610, 1975.

GIL, P., PARDO, L. y GIL, M. A. : *Matemáticas de la incertidumbre y de la información y sus aplicaciones estadísticas.* Publicaciones de la Universidad de Oviedo. 1993.

GOKHOLE, D. V. y KULLBACK, S. : *The Information in Contingency Tables.* New York, Marcel Dekker, 1978.

GOODMAN, L. A. : *Analysis of Cross-Clasified Data Harving Ordered Categories.* Cambridge, MA, Harvard University Press, 1984.

GUIASU, S. : *Information Theory with Applications.* McGraw Hill, New York, 1977.

GUIASU, S. : Grouping data by using the weighted entropy. *Journal of Statistical Planning and Inference, 15,* 63-69, 1986.

GUPTA, S. S. : Bibliography on Multivariate Normal Integrals and Rrlated Topics. *Annals of Mathematical Statistics, 34,* 829-838, 1963.

GYORFI, L. y NEMETZ, K. : $f-$dissimilarity: a generalization of the affinity of several distributions. *Annals of the Institute Statistical Mathematics 30,* 105-113, 1978.

HABERMAN, S. L. : *Analysis of Qualitative Data,* Vol. 1, New York, Academic Press, 1978.

HABERMAN, S. L. : *Analysis of Qualitative Data,* Vol. 2, New York, Academic Press, 1979.

HASSENBLAD, V. : Estimation of parameters for a mixture of normal distributions. *Technometrics, 8,* 431-434, 1966.

HOCHBERG, Y. : On the use of double sampling schemes in analyzing categorical data with misclassification errors. *Journal of American Statistical Association, 72, n 360, part 1,* 914-921, 1977.

HOUDA, D. S. y TUTEJA, R. K. : Two generalized measures of useful Information. *Information Sciences, 23,* 11-24, 1981..

JENSEN, D. R. y SOLOMON, H. : A Gaussian Aproximation to the Distribution of a Definite Quadratic Form. *Journal of American Statistical Association, 67 (340),*898-902, 1972.

JOHNSON, N. L. y KOTZ, S. : Tables of Distributions of Positive Definite Quadratic Forms in Central Normal Variables. *Sankhya, Ser. B,* 303-314, 1968.

KAPUR, J. N. : On Measures of divergence based on Jensen difference. *Nat. Acad. Sci. Lett. 11,* 1, 23-27, 1988.

KAUFMAN, H. y MATHAI, A. M. : An axiomatic foundation for a multivariate measure of affinity among a number of distributions. *Journal of Multivariate Analysis, 3,* 236-242, 1973.

KERRIDGE, D. F. : Inaccuracy and inference. *Journal of the Royal Statistical Society, Ser. B 23,* 184-194, 1961.

KULLBACK, S. y LEIBLER, A. : On information and sufficiency. *Annals of Mathematical Statistics, 22,* 79-86, 1951.

LE CAM, L. : Maximum likelihood: an introduction. *International Statistical Review, 58,* 2, 153-171, 1990.

LEONARD, T. : *A Course in Categorical Data Analysis.* Chapman and Hall, 2000.

LIESE, F. y VAJDA, I. : *Convex Statistical Distances.* Teubner, Leipzig. 1987.

LINDSAY, B. G. : Efficiency versus robustness: The case for minimum Hellinger distance and related methods. *Ann. Statist. 22,* 1081-1114, 1994.

LONGO, G. : *Quantitative-Qualitative Measures of information.* Springer-Verlay, New York, 1972.

LONGO, G. : A noiseless coding theorem for sources having utilities. *SIAM Journal of Applied Mathematics 30,* 739-748, 1976.

MATUSITA, K. : On the notion of affinity of several distributions and some of its applications. *Annals of the Institute Statistical Mathematics 19,* 181-192, 1967.

MATUSITA, K. : Discrimination and the affinity of distributions in Discriminant Analysis and Applications, T. Cacoullus, Ed. New York. Academic Press, 213-223, 1973.

MENENDEZ, M. L., PARDO, L. y TANEJA, I. J. : On $M-$dimensional unified $(r, s)-$Jensen difference divergence measures and their applications. *Kybernetika, 28,* 309-324, 1992.

MENENDEZ, M. L., MORALES, D., PARDO, L. y SALICRU, M. : Asymptotic behaviour and statistical applications of divergence measures in multinomial populations: A study. *Statistical Papers*, 1-29, 1995.

MENENDEZ, M. L., MORALES, D., PARDO, L. y SALICRU, M. : Divergence measures between s populations: Statistical applications in the exponential family. *Communications in Statistics (Theory and Methods), 26,* 5, 1099-1117, 1997a.

MENENDEZ, M. L., MORALES, D., PARDO, L. y VAJDA, I. : Divergence-Based Estimation and Testing of Statistical Models of Classification. *Journal of Multivariate Analysis, 54, 1,* 329-354, 1995.

MENENDEZ, M. L., MORALES, D., PARDO, L. y VAJDA, I. : Asymptotic distributions of $f-$divergences of hypotetical and observed frequencies in sparse testing schemes. *Statistica Neerlandica,* 1997b.

MENENDEZ, M. L. , PARDO, J. A., PARDO, L. y PARDO, M. C. : Asymptotic approximations for the distributions of the $(h, \phi)-$ divergence goodness-of-fit statistics: Applications to Renyi's statistic. *Kybernetes, 26,* 442-452, 1997c.

MORALES, D., PARDO, L. y VAJDA, I. : Asymptotic divergence of estimates of discrete distributions. *Journal Statistical Planning and Infernece*, 48, 347-369,1995.

MORALES, D., PARDO, L. y VAJDA, I. : Some New Statistics for Testing Hypothesis in Parametric Models. *Journal of Multivariate Analysis*, 62, 137-168, 1997.

MOTE, V. L. y ANDERSON, R. L. : An investigation of effect of misclassification on the properties of Chi-Squared tests in the analysis of categorical data. *Biometrika 52,* 95-109, 1965.

PARDO, J. A., PARDO, L. y ZOGRAFOS, K. : Minimum ϕ-divergence stimators with constrains in multinomials populations. *Journal of Statistical Planning and Inference.* En prensa.

PARDO, L. : Generalized divergence measures: Statistical applications. Encyclopedia of microcomputers (Editores, A. Kent, J. G. Willianms y N. L. Johnson). Marcel Dekker, 1997a.

PARDO, L. : *Teoría de la información estadística.* Editorial HESPERIDES. Salamanca, 1997b.

PARDO, M. C. : On testing independence for multidimensional contingency tables with stratified random sampling. *Information Sciences, 78,* 101-118, 1994a.

PARDO, M. C. : Tests of independence for multidimensional contingency tables based on $(h, \phi)-$divergence measures. *Utilitas Mathematica,* 1994b.

PARDO, M. C. : Estimadores de mínima divergencia de Rao: Comportamiento asintótico y Aplicación a contrastes de hipótesis. *Tesis Doctoral.* Universidad Complutense de Madrid. 1996.

PARDO, M. C. : On Burbea-Rao Divergence Based Goodness-of-fit tests for Multinomial Models. *Journal of Multivariate Analysis, 68,* 1999.

PARDO, M. C. y VAJDA, I. : About distances of discrete distributions satisfying the data processing theorem of information theory. *IEEE Transactions on Information Theory, 43,* 1288-1293, 1997.

PARR, W. C. : Minimum distance estimation: a bibliography. *Communictions in Statistics (Theory and Methods), 10,* 1205-1224, 1981.

PEARSON, K. : Contributions to the mathematical theory of evolution. *Phil. Trans. R. Soc. Ser. A, 185,* 71-110, 1894.

PEARSON, K. : 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, 1900.

PEREZ, T. : *Tesis Doctoral.* Universidad Complutense de Madrid, 2000.

PLACKETT, R. L. : *The Analysis of Categorical Data* (2^a edición). High Wycombe, Griffin, 1981.

PICARD, C. F. : *Weighted Probabilistic Information Measures.* J. Com. Inferm. and Sy, 1979.

RAO, C. R. : *Linear Statistical inference and its applications.* John Wiley. New York, 1973.

RAO, C. R. : Diversity and disimilarity coefficients: a unified approach. *Journal Theoretical Population Biology, 21,* 24-43, 1982.

RAO, J. N. K. y SCOTT, A. J. : The Analysis of Categorical Data from Complex Sample Surveys: Chi-squared tests for Goodness-of-Fit and Independence in Two-way Tables. *Journal of American Statistical Association, 76,* 221-230, 1981.

READ, T. R. C. y CRESSIE, N. : Goodness-of-fit statistics for discrete multivariate data. *Springer Series in Statistics.* Springer-Verlag, New York, 1988.

RENYI, A. : On measures of entropy and information. *Proceedings of the fourth Berkeley Symposium on Mathematical Statistics and Probability, 1,* 547-561, 1961.

SHANNON, C. E. : A mathematical theory of communication. *Bell System Tech. J. 27,* 379-423, 1948.

SHAOO, P. K. y WONG, A. K. C. : Generalized Jensen difference based on entropy functions. *Kybernetika, 24,* 4, 241-250, 1988.

SHARMA, B. D. y MITTAL, D. P. : New Non-additive measures of relative information. *Journal of Combinatory Information & Systems Science, 2,* 122-133, 1975.

SIMPSON, D. G. : Minimum Hellinger distance for the analysis of count data. *Journal of the American Statistical Association, 82,* 802-807, 1987.

SOLOMON, H. : Distribution of Quadratic Forms-Tables and Applications, Aplied Mathematics and Statistics Laboratories. *Technical Report, 45, Standfor University,* 1960.

TANEJA, H. C. : On measures of relative "usefull" information. *Kybernetika, 21, n 2,* 148-156, 1985.

TENENBEIN, A. : A Double Sampling Scheme for Estimating from Binomial Data with misclassifications. *Journal of American Statistical Association, 65,* 1351-1361, 1970.

TENENBEIN, A. : A Double Sampling Scheme for Estimating from Binomial Data with misclassifications: Sample Size Determination. *Biometrics 27,* 935-944, 1971.

TENENBEIN, A. : A Double Sampling Scheme for Estimating from Misclassified Multinomial Data with Applications to Sampling Inspection. *Technometrics 14,* 187-202, 1972.

TOUSSAINT, G. T. : Some properties of Matusitas measure of affinity of several distributions. *Annals of the Institute of Statistical Mathematics 26,* 389-394, 1974.

TOUSSAINT, G. T. : Probability of error, expected divergence, and affinity of several distributions. *IEEE Transactions on System, Man and Cybernetics, 8,* 6, 482-485, 1978.

UPTON, G. J. G. : *The Analysis of Cross-Tabulated Data.* New York, John Wiley, 1978.

WOODWARD, W.A., PARR, W.C., SCHUCANY, W.R. y LINDSAY, B. G.: A comparison of minimum distance and maximum likelihood estimation of a mixture proportion. *Journal of American Statistical Association, 79,* 590-598, 1984.

WOODWARD, W.A., WHITNEY, P. y ESLINGER, P.W. : Minimum Hellinger distance of mixture proportions. *Journal of Statistical Planning and Inference, 48,* 303-319, 1995.

ZOGRAFOS, K., FERENTINOS, K. y PAPAIOANNOU, T. : $\varphi-$divergence statistics: sampling properties, multinomial goodness of fit and divergence tests. *Communications in Statistics (Theory and Methods), 19(5),* 1785-1802, 1990.

ZOGRAFOS, K. : Asymptotic distributions of estimated $f-$dissimilarity between populations in stratified sampling. *Statistics & Probability Letters, 21,* 147-151, 1994.

ZOGRAFOS, K. : Asymptotic properties of $\varphi-$divergence statistics and its application in contingency tables. *International Journal of Mathematics and Statistical Science, 2 (1),* 5-22, 1993.