QÜESTIIÓ, Vol 16, 1,2,3 pp. 99-116, 1992

CONTRASTES DE HIPÓTESIS BASADOS EN LA (r,s)–DIVERGENCIA: APLICACIÓN A DISTRIBUCIONES MULTINOMIALES Y NORMALES MULTIVARIANTES

MORALES, D., PARDO, L., SALICRÚ, M. y MENÉNDEZ, M.L.

En este trabajo se obtiene la distribución asintótica de la (r,s)–divergencia, introducida por Sharma y Mittal 1975, entre dos densidades $f_{\theta_1}, f_{\theta_2}$ cuando θ_2 es fijo y θ_1 desconocido o bien cuando los dos son desconocidos. Se supone que los parámetros desconocidos se estiman de acuerdo con el principio de máxima verosimilitud. Como caso particular se obtienen las distribuciones asintóticas en el caso de poblaciones multinomiales. Se concluye el trabajo construyendo, sobre la base de los estadísticos citados, contrastes de hipótesis para poblaciones normales multidimensionales.

Test of hypotheses on the (r,s)–divergence: applications to multinomial and multivariate normal distributions.

Keywords: Contrastes de hipótesis, (r,s)–divergencia, distribuciones normales multivariantes.

–MORALES, D. y PARDO, L. Dep. Estadística e I.O. Fac. Matemáticas. Univ. Complutense de Madrid. 28040 MADRID (SPAIN).

–SALICRÚ, M. Dep. Estadística. Univ. Barcelona. Av. Diagonal, 645. 08028 BARCELONA (SPAIN).

–MENÉNDEZ, M.L. Dep. Matemática Aplicada. E.T.S. Arquitectura. Univ. Politécnica Madrid. 20049 MADRID (SPAIN).

–Article rebut el desembre de 1991.

–Acceptat el juny de 1992.

1. INTRODUCCIÓN

Para un espacio estadístico $(\mathfrak{X}, \beta_{\mathfrak{X}}, P_\theta)_{\theta\in\Theta\subset\mathbf{R}^k}$ dominado por una medida μ σ–finita, de forma que $f_\theta(x) = (dP_\theta/d\mu)(x)$, Sharma y Mittal (1977) definen la (r,s)–divergencia entre dos funciones de densidad de probabilidad $f_{\theta_1}(x), f_{\theta_2}(x)$ mediante la expresión

$$(1) \qquad D_r^s(\theta_1,\theta_2) = (s-1)^{-1}\left\{\left(\int_{\mathfrak{X}} f_{\theta_1}(x)^r f_{\theta_2}(x)^{1-r} d\mu(x)\right)^{\frac{s-1}{r-1}} - 1\right\}$$

$$r \neq 1, s \neq 1, r > 0$$

donde $\theta_1 = (\theta_{11},\ldots,\theta_{1k})$ y $\theta_2 = (\theta_{21},\ldots,\theta_{2k})$.

Si s tiende a 1, se obtiene la divergencia de Rényi (1961)

$$\lim_{s\to 1} D_r^s(\theta_1,\theta_2) = (r-1)^{-1}\log\left(\int_{\mathfrak{X}} f_{\theta_1}(x)^r f_{\theta_2}(x)^{1-r} d\mu(x)\right)$$

y al hacer tender r a 1, se obtiene la divergencia

$$D_1^s(\theta_1,\theta_2) = (s-1)^{-1}\left(\exp\{(s-1)D(f_{\theta_1}(x), f_{\theta_2}(x))\} - 1\right) \qquad s \neq 1$$

siendo $D(\theta_1,\theta_2)$ la divergencia dirigida de Kullback-Leibler (1951) definida por

$$D_1^1(\theta_1,\theta_2) = D(\theta_1,\theta_2) = \int_{\mathfrak{X}} f_{\theta_1}(x)\log\frac{f_{\theta_1}(x)}{f_{\theta_2}(x)}\, d\mu(x)$$

Algunas propiedades relativas a la no negatividad, monotonía, convexidad, suficiencia y expresión de la distancia entre distribuciones normales a partir de la (r,s)–divergencia se obtienen en Pardo, L. y otros (1990).

Bajo las siguientes condiciones de regularidad

(*i*) El conjunto $A = \{x \in \mathfrak{X} / f(x,\theta) > 0\}$ no depende de θ.

(*ii*) Para todo $x \in A$ y para todo $\theta = (\theta_1,\ldots,\theta_k) \in \Theta$, existen

$$\frac{\partial}{\partial\theta_i} f(x,\theta), \qquad \frac{\partial^2}{\partial\theta_i\partial\theta_j} f(x,\theta) \quad i,j = 1,\ldots,k$$

(*iii*) Para cada $B \in \beta_{\mathfrak{X}}$ y para cada $i,j = 1,\ldots,k$.

$$\frac{\partial}{\partial\theta_i}\int_B f(x,\theta)d\mu = \int_B \frac{\partial}{\partial\theta_i} f(x,\theta)d\mu$$

y

$$\frac{\partial^2}{\partial\theta_i\partial\theta_j}\int_B f(x,\theta)d\mu = \int_B \frac{\partial^2}{\partial\theta_i\partial\theta_j} f(x,\theta)d\mu$$

y considerando las estimaciones de las funciones de densidad de probabilidad f_{θ_1} y f_{θ_2} por el método de máxima verosimilitud, con muestras de tamaño n y m respectivamente, Kupperman (1957) estableció, bajo la hipótesis $\theta_1 = \theta_2$, que los estadísticos $2nD(\hat{\theta}_1, \theta_2)$ y $\frac{n \cdot m}{n+m} D(\hat{\theta}_1, \hat{\theta}_2)$ se distribuyen asintóticamente según leyes χ^2 con k grados de libertad.

El objetivo en este trabajo es doble, primeramente se calculará la distribución asintótica de $D_r^s(\hat{\theta}_1, \theta_2)$ y $D_r^s(\hat{\theta}_1, \hat{\theta}_2)$ donde $\hat{\theta}_1$ y $\hat{\theta}_2$ denotan los estimadores de máxima verosimilitud de θ_1 y θ_2 respectivamente. Por otro lado utilizando estos resultados y los obtenidos en Pardo, L. y otros (1990), se plantean contrastes de hipótesis en poblaciones normales n-dimensionales acerca del vector de medias, matriz de covarianzas o ambos.

2. COMPORTAMIENTO ASINTÓTICO DE $D_r^s(\theta_1, \theta_2)$

En primer lugar se abordará el problema de encontrar la distribución de $D_r^s(\hat{\theta}_1, \theta_2)$, cuando θ_2 es conocido y $\hat{\theta}_1$ es el estimador de máxima verosimilitud de θ_1. En este sentido hemos encontrado el siguiente resultado

Teorema 1

Si $\theta_1 \neq \theta_2$, entonces

$$n^{1/2}\left(D_r^s(\hat{\theta}_1, \theta_2) - D_r^s(\theta_1, \theta_2)\right) \xrightarrow[n \longrightarrow \infty]{L} \mathrm{N}(0, T^t I_F^{-1}(\theta_1) T)$$

donde $I_F(\theta_1)$ es la matriz de información de Fisher y $T = (t_1, \ldots, t_k)^t$ con

$$(2) \quad t_i = \frac{r}{r-1}\left(\int_{\mathcal{X}} f_{\theta_1}(x)^r f_{\theta_2}(x)^{1-r} d\mu\right)^{\frac{s-r}{r-1}} \cdot \int_{\mathcal{X}} \left(\frac{f_{\theta_1}(x)}{f_{\theta_2}(x)}\right)^{r-1} \frac{\partial f_{\theta_1}(x)}{\partial \theta_{1i}} d\mu$$

cuando $T^t I_F^{-1}(\theta_1) T > 0$.

Demostración

De considerar el desarrollo de Taylor de la función

$$h_r^s(\theta_{11},\ldots,\theta_{1k}) = \frac{1}{s-1}\left[\left(\int f_{\theta_1}(x)^r f_{\theta_2}(x)^{1-r} d\mu\right)^{\frac{s-1}{r-1}} - 1\right]$$

para el punto $\hat{\theta}_1 = (\hat{\theta}_{11},\ldots,\hat{\theta}_{1k})$ en un entorno del punto $\theta_1 = (\theta_{11},\ldots,\theta_{1k})$ se tiene

$$\text{(3)} \qquad \begin{aligned} D_r^s(\hat{\theta}_1,\theta_2) &= D_r^s(\theta_1,\theta_2) + \sum_{i=1}^{k} \frac{\partial D_r^s(\theta_1,\theta_2)}{\partial\theta_{1i}}(\hat{\theta}_{1i}-\theta_{1i}) + \\ &+ \mathcal{E}_n \|\hat{\theta}_1 - \theta_1\| \end{aligned}$$

donde $\mathcal{E}_n \xrightarrow{P} 0$ cuando $\hat{\theta}_1 \longrightarrow \theta_1$.

Ahora bien,

$$\begin{aligned} t_i &= \frac{\partial D_r^s(\theta_1,\theta_2)}{\partial\theta_{1i}} = \\ &= \frac{r}{r-1}\left(\int_{\mathcal{X}} f_{\theta_1}(x)^r f_{\theta_2}(x)^{1-r} d\mu\right)^{\frac{s-r}{r-1}} \cdot \int_{\mathcal{X}} \left(\frac{f_{\theta_1}(x)}{f_{\theta_2}(x)}\right)^{r-1} \frac{\partial f_{\theta_1}(x)}{\partial\theta_{1i}} d\mu \end{aligned}$$

Entonces, de considerar que

$$D_r^s(\hat{\theta}_1,\theta_2) - D_r^s(\theta_1,\theta_2) \quad \text{y} \quad \sum_{i=1}^{k} t_i(\hat{\theta}_{1i}-\theta_{1i})$$

siguen la misma distribución y que

$$\sqrt{n}(\hat{\theta}_{11}-\theta_{11},\ldots,\hat{\theta}_{1k}-\theta_{1k}) \xrightarrow{L} \mathrm{N}(0, I_F^{-1}(\theta_1))$$

se tiene

$$\sqrt{n}\left(D_r^s(\hat{\theta}_1,\theta_2) - D_r^s(\theta_1,\theta_2)\right) \xrightarrow{L} \mathrm{N}(0, T^t I_F^{-1}(\theta_1) T)$$

con

$$T = (t_1,\ldots,t_k)^t$$

■

Cuando $\theta_1 = \theta_2$, los términos t_i del desarrollo de Taylor (3) son nulos, y por eso, es necesario recurrir a los términos de segundo orden. En este supuesto, hemos obtenido

Teorema 2

Si $\theta_1 = \theta_2$, entonces

$$\frac{2nD_r^s(\hat{\theta}_1, \theta_2)}{r} \xrightarrow[n \longrightarrow \infty]{L} \chi_k^2$$

Demostración

De considerar el desarrollo de Taylor de segundo orden para la función $h_r^s(\theta_{11}, \ldots, \theta_{1k})$ cuando $\theta_1 = \theta_2$ se tiene:

$$D_r^s(\hat{\theta}_1, \theta_2) = \frac{1}{2}\sum_{i,j} \frac{\partial^2 h_r^s(\theta_{11}, \ldots, \theta_{1k})}{\partial\theta_{1i}\partial\theta_{1j}} (\hat{\theta}_{1i} - \theta_{1i})(\hat{\theta}_{1j} - \theta_{1j}) + \mathcal{E}_n \|\hat{\theta}_1 - \theta_1\|^2$$

con $\mathcal{E}_n \xrightarrow{P} 0$ cuando $\hat{\theta}_1 \longrightarrow \theta_1$.

Atendiendo a que

$$\frac{\partial^2 h_r^s(\theta_{11}, \ldots, \theta_{1k})}{\partial\theta_{1i}\partial\theta_{1j}} = r I_{ij}^F(\theta_1)$$

obtenemos que

$$(4) \qquad \frac{2nD_r^s(\hat{\theta}_1, \theta_2)}{r} \approx \left[\sqrt{n}(\hat{\theta}_1 - \theta_1)\right]^t I^F(\theta_1) \left[\sqrt{n}(\hat{\theta}_1 - \theta_1)\right]$$

y por Serfling (1980, pág. 149), la forma cuadrática (4) sigue una distribución χ_k^2. Así, el Teorema 2 queda probado.

■

De consideraciones análogas a las realizadas en los Teoremas 1 y 2, cuando θ_1 y θ_2 son desconocidas y han sido obtenidas por el método de máxima verosimilitud a partir de muestras aleatorias independientes de tamaño n y m respectivamente, se tiene:

Teorema 3

Si $\theta_1 \neq \theta_2$, entonces

$$(\mathrm{n}\cdot\mathrm{m})^{1/2}\left(\mathrm{D}_\mathrm{r}^\mathrm{s}(\hat{\theta}_1,\hat{\theta}_2)-\mathrm{D}_\mathrm{r}^\mathrm{s}(\theta_1,\theta_2)\right) \xrightarrow[n,m\to\infty]{L} \mathrm{N}(0,\mathrm{m}\mathrm{T}^t\mathrm{I}_\mathrm{F}^{-1}(\theta_1)\mathrm{T}+\mathrm{n}\mathrm{S}^t\mathrm{I}_\mathrm{F}^{-1}(\theta_2)\mathrm{S})$$

donde T es el vector definido en el teorema 1 y S es el vector de componentes

$$s_i = -\left(\int_{\mathfrak{X}} f_{\theta_1}(x)^r f_{\theta_2}(x)^{1-r} d\mu\right)^{\frac{s-r}{r-1}} \int_{\mathfrak{X}} \left(\frac{f_{\theta_1}(x)}{f_{\theta_2}(x)}\right)^r \frac{\partial f_{\theta_2}(x)}{\partial \theta_{2i}} d\mu \tag{5}$$

cuando $mT^tI_F^{-1}(\theta_1)T + nS^tI_F^{-1}(\theta_2)S \neq 0$.

Demostración

De considerar el desarrollo de Taylor de la función

$$g_r^s(\theta) = g_r^s(\theta_{11},\ldots,\theta_{1k},\theta_{21},\ldots,\theta_{2k}) = D_r^s(\theta_1,\theta_2) \tag{6}$$

se tiene

$$D_r^s(\hat{\theta}_1,\hat{\theta}_2) = D_r^s(\theta_1,\theta_2) + \sum_{j=1}^{2}\sum_{i=1}^{k} \frac{\partial D_r^s(\theta_1,\theta_2)}{\partial \theta_{ji}}(\hat{\theta}_{ji}-\theta_{ji}) + \mathcal{E}_n\|\hat{\theta}-\theta\| \tag{7}$$

con $\mathcal{E}_n \longrightarrow 0$ cuando $\hat{\theta} \longrightarrow 0$.

Entonces

$$(n\cdot m)^{1/2}\left(D_r^s(\hat{\theta}_1,\hat{\theta}_2)-D_r^s(\theta_1,\theta_2)\right) \approx$$
$$m^{1/2}\sum_{i=1}^{k} t_i n^{1/2}(\hat{\theta}_{1i}-\theta_{1i}) + n^{1/2}\sum_{i=1}^{k} s_i m^{1/2}(\hat{\theta}_{2i}-\theta_{2i})$$

con t_i definido en (2) y s_i definido en (5).

Así, el resultado queda probado atendiendo a que

$$T^t n^{1/2}(\hat{\theta}_1-\theta_1) \xrightarrow[n\longrightarrow\infty]{L} \mathrm{N}(0,T^tI_F^{-1}(\theta_1)T)$$

y

$$S^t m^{1/2}(\hat{\theta}_2 - \theta_2) \xrightarrow[m \longrightarrow \infty]{L} \mathrm{N}(0, S^t I_F^{-1}(\theta_2) S)$$

■

Cuando $\theta_1 = \theta_2$ el término de $1^{\underline{\text{er}}}$ orden del desarrollo de Taylor (6) es nulo, y en este caso, con el término de $2^{\underline{\text{do}}}$ orden tenemos:

Teorema 4

Cuando $\theta_1 = \theta_2$ se tiene

$$\frac{2nm}{n+m} \frac{D_r^s(\hat{\theta}_1, \hat{\theta}_2)}{r} \xrightarrow[n,m \longrightarrow \infty]{L} \chi_k^2$$

Demostración

Utilizando el desarrollo de Taylor de la función $g_r^s(\theta)$ se tiene que cuando $\theta_1 = \theta_2$

$$D_r^s(\hat{\theta}_1, \hat{\theta}_2) \approx \frac{r}{2} \left(\sum_{i,j=1}^{k} \int \frac{1}{f_{\theta_1}(x)} \frac{\partial f_{\theta_1}(x)}{\partial \theta_{1i}} \frac{\partial f_{\theta_2}(x)}{\partial \theta_{2i}} d\mu \right) (\hat{\theta}_{1i} - \theta_{1i})$$

Así, atendiendo a que

$$\left(\frac{n \cdot m}{n+m} \right)^{1/2} (\hat{\theta}_1 - \hat{\theta}_2) \xrightarrow[n,m \longrightarrow \infty]{L} \mathrm{N}(0, I_F^{-1}(\theta_1))$$

se tiene que

$$\frac{n \cdot m}{n+m} (\hat{\theta}_1 - \hat{\theta}_2)' I_F(\theta_1)(\hat{\theta}_1 - \hat{\theta}_2) \xrightarrow[n,m \longrightarrow \infty]{L} \chi_k^2$$

y por tanto

$$\frac{2nm}{m+n} \frac{D_r^s(\hat{\theta}_1, \hat{\theta}_2)}{r} \xrightarrow[n,m \longrightarrow \infty]{L} \chi_k^2$$

■

Nota 1

Los valores de t_i y s_i en el caso de $r = 1$ se obtienen de las expresiones (2) y (5) por continuidad.

3. APLICACIÓN A CONTRASTES DE HIPÓTESIS

Los resultados dados en los teoremas anteriores se pueden utilizar para realizar contrastes de bondad de ajuste y homogeneidad. Así, sobre la base del Teorema 2, es posible contrastar $H_0 : \theta = \theta_1$ frente $H_1 : \theta \lessgtr \theta_1$, sin más que utilizar el estadístico

$$T = \frac{2nD_r^s(\hat{\theta}_1, \theta_1)}{r} \tag{8}$$

y rechazar la hipótesis nula si $T > \chi^2_{k(\alpha)}$. La Potencia asintótica del test se obtiene a partir del Teorema 1 y la consistencia a partir de Fraser (1957).

El teorema 4 permite contrastar si dos muestras aleatorias simples e independientes obtenidas de las poblaciones $f_{\theta_1}(x)$ y $f_{\theta_2}(x)$, son tales que $\theta_1 = \theta_2$. En este caso, se rechazará la hipótesis $H_0 : \theta_1 = \theta_2$ si

$$T = \frac{2mn}{m+n} \frac{D_r^s(\hat{\theta}_1, \hat{\theta}_2)}{r} > \chi^2_{k(\alpha)} \tag{9}$$

De forma análoga al caso anterior, se obtiene la potencia asintótica y la consistencia del test.

Nota 2

Al categorizar variables no necesariamente categóricas, es posible resolver los siguientes contrastes relativos a una o varias muestras.

(1) Test de bondad de ajuste, $H_0 : p = p_0$ (p_0 conocida).

(2) Test de ajuste de m poblaciones, $H_0 : p^{(1)} = \cdots = p^{(m)} = p_0$ (p_0 conocida).

(3) Test de homogeneidad de dos poblaciones, $H_0 : p = q$.

(4) Test de homogeneidad de m poblaciones, $H_0 : p^{(1)} = \cdots = p^{(m)}$.

4. CASOS PARTICULARES

Al particularizar los resultados anteriores a funciones de densidad de probabilidad $f_{\theta_1}(x)$ y $f_{\theta_2}(x)$ siguiendo funciones de densidad de probabilidad multinomiales $M(p_1, \ldots, p_k)$ y $M(q_1, \ldots, q_k)$ se recuperan los resultados obtenidos por Menéndez y otros (1991) en los que

$$\begin{aligned} D_r^s(\theta_1, \theta_2) &= (s-1)^{-1}\left[\left(\sum_{i=1}^{k} p_i^r q_i^{1-r}\right)^{\frac{s-1}{r-1}} - 1\right] \\ t_i &= \frac{r}{r-1}\left(\sum_{i=1}^{k} p_i^r q_i^{1-r}\right)^{\frac{s-r}{r-1}} p_i^{r-1} \cdot q_i^{1-r} \end{aligned}$$

y

$$s_i = -\left(\sum_{i=1}^{k} p_i^r q_i^{1-r}\right)^{\frac{s-r}{r-1}} p_i^r \cdot q_i^{-r}$$

Así, la varianza en el teorema 1 es

$$\sigma^2 = \frac{r^2}{(r-1)^2}\left(\sum_{i=1}^{k} p_i^r q_i^{1-r}\right)^{2\frac{s-r}{r-1}} \left\{\sum_{i=1}^{k} p_i^{2r-1} q_i^{2(1-r)} - \left(\sum_{i=1}^{k} p_i^r q_i^{1-r}\right)^2\right\}$$

y la varianza en el teorema 3 es

$$\begin{aligned} \sigma^2 = \frac{1}{(r-1)^2}\left(\sum_{i=1}^{k} p_i^r q_i^{1-r}\right)^{2\frac{s-r}{r-1}} &\left\{r^2 \sum_{i=1}^{k} p_i^{2r-1} q_i^{2(1-r)} + (1-r)^2 \sum_{i=1}^{k} p_i^{2r} q_i^{1-2r} - \right. \\ &\left. - r^2\left(\sum_{i=1}^{k} p_i^r q_i^{1-r}\right)^2 - (1-r)^2\left(\sum_{i=1}^{k} p_i^r q_i^{1-r}\right)^2\right\} \end{aligned}$$

La expresión de la $(r, s)-$divergencia entre distribuciones normales $n-$dimensionales con parámetros (μ_1, Σ_1) y (μ_2, Σ_2) ha sido obtenida por Pardo y otros (1990). En este sentido,

$$D_r^s\left((\mu_1,\Sigma_1),(\mu_2,\Sigma_2)\right) =$$
$$= (s-1)^{-1}\left(\exp\left\{\frac{r(s-1)}{2}\left((\mu_1,\mu_2)^T(r\Sigma_2+(1-r)\Sigma_1)^{-1}(\mu_1,\mu_2)\right)\right\}\cdot\right.$$
$$\left.\cdot\,\frac{|r\Sigma_2+(1-r)\Sigma_1|^{\frac{1-s}{2(r-1)}}}{|\Sigma_1|^{\frac{s-1}{2}}|\Sigma_2|^{\frac{(1-s)r}{2(r-1)}}}-1\right) \qquad r\pm 1,\ s\pm 1$$

y por continuidad

$$D_1^s\left((\mu_1,\Sigma_1),(\mu_2,\Sigma_2)\right) =$$
$$= (s-1)^{-1}\left(\exp\left\{\frac{(r-1)}{2}\left((\mu_1,\mu_2)^T\Sigma_2^{-1}(\mu_1,\mu_2)+\operatorname{tr}(\Sigma_2^{-1}\Sigma_1-I)\right)\right\}\cdot\right.$$
$$\left.\cdot\,\frac{|\Sigma_2|^{\frac{s-1}{2}}}{|\Sigma_1|^{\frac{s-1}{2}}}-1\right)$$

$$D_r^1\left((\mu_1,\Sigma_1),(\mu_2,\Sigma_2)\right) = \frac{r}{2}\left((\mu_1,\mu_2)^T(r\Sigma_2+(1-r)\Sigma_1)^{-1}(\mu_1,\mu_2)\right)+$$
$$+\ \frac{1}{2(1-r)}\log\frac{|r\Sigma_2+(1-r)\Sigma_1|}{|\Sigma_1|^{1-r}|\Sigma_2|^r}$$

$$D\left((\mu_1,\Sigma_1),(\mu_2,\Sigma_2)\right) =$$
$$\frac{1}{2}\left((\mu_1,\mu_2)^T\Sigma_2^{-1}(\mu_1,\mu_2)+\operatorname{tr}(\Sigma_2^{-1}\Sigma_1-I)+\log\frac{|\Sigma_2|}{|\Sigma_1|}\right)$$

Así, sobre la base de los resultados obtenidos anteriormente para una distribución normal $\mathrm{N}(\mu,\Sigma)$ es posible contrastar: (1) $\mathrm{H}_0 : \mu=\mu_0$ con Σ conocida, (2) $\mathrm{H}_0 : \Sigma=\Sigma_0$ con μ conocida y (3) $\mathrm{H}_0 : (\mu,\Sigma)=(\mu_0,\Sigma_0)$, y para dos poblaciones normales $\mathrm{N}(\mu_1,\Sigma_1)$ y $\mathrm{N}(\mu_2,\Sigma_2)$ es posible contrastar: (1) $\mathrm{H}_0 : \mu_1=\mu_2$ con Σ_1 y Σ_2 conocidas, (2) $\mathrm{H}_0 : \mu_1=\mu_2$ con Σ_1 y Σ_2 desconocidas pero iguales, y (3) $\mathrm{H}_0 : (\mu_1,\Sigma_1)=(\mu_2,\Sigma_2)$.

Teniendo en cuenta que:

a) Para $\theta_1=(\mu_{11},\ldots,\mu_{1k})$,

$$t_i=\left(D_r^s(\theta_1,\theta_2)+\frac{1}{s-1}\right)\frac{r(s-1)}{2}\left\{\sum_{j=1}^{k}(r\Sigma_2+(1-r)\Sigma_1)_{ij}^{-1}(\mu_{1j}-\mu_{2j})+\right.$$
$$\left.+\sum_{i=1}^{k}(r\Sigma_2+(1-r)\Sigma_1)_{ij}^{-1}(\mu_{1i}-\mu_{2i})\right\}$$

siendo

$$(r\Sigma_2 + (1-r)\Sigma_1)^{-1}_{ij}$$

el elemento ij de la matriz

$$(r\Sigma_2 + (1-r)\Sigma_1)^{-1}$$

b) Para $\theta = \left({}^1\sigma_{11}, {}^1\sigma_{12}, \ldots, {}^1\sigma_{1k}, \ldots, {}^1\sigma_{k1}, \ldots, {}^1\sigma_{kk}\right)$

$$t_{(i-1)k+j} = \left(D_r^s(\theta_1, \theta_2) + \frac{1}{s-1} \right) \left((\mu_1 - \mu_2)^t (l^{\cdot i}\, l^{j\cdot})(\mu_1 - \mu_2) \right) +$$

$$+ (s-1)^{-1} \cdot \exp\left\{ \frac{r-1}{2} \left((\mu_1 - \mu_2)^t (r\Sigma_2 + (1-r)\Sigma_1)^{-1} (\mu_1 - \mu_2) \right) \right\} \cdot$$

$$\cdot |\Sigma_2|^{\frac{(s-1)r}{2(r-1)}} \cdot \left\{ \frac{\frac{1-s}{2(r-1)} |r\Sigma_2 + (1-r)\Sigma_1|^{\frac{1-s}{2(r-1)}} A_{ij} |\Sigma_1|^{\frac{s-1}{2}}}{|\Sigma_1|^{s-1}} - \right.$$

$$\left. - \frac{\frac{s-1}{2} |\Sigma_1|^{\frac{s-1}{2}-1} A^*_{ij} |r\Sigma_2 + (1-r)\Sigma_1|^{\frac{1-s}{2(r-1)}}}{|\Sigma_1|^{s-1}} \right\}$$

donde:

A_{ij} es el adjunto del elemento (i,j) en la matriz $r\Sigma_2 + (1-r)\Sigma_1$
A^*_{ij} es el adjunto del elemento (i,j) en la matriz Σ_1
$l^{\cdot i}$ es la i–ésima columna de la matriz $(r\Sigma_2 + (1-r)\Sigma_1)^{-1}$
$l^{j\cdot}$ es la j–ésima fila de la matriz $(r\Sigma_2 + (1-r)\Sigma_1)^{-1}$
$i = 1, \ldots, k, \quad j = 1, \ldots, k.$

5. EJEMPLO

Como una aplicación de los resultados para distribuciones normales multivariantes, procederemos seguidamente al análisis de los datos presentados en Mardia, Kent y Bibby (1979, p. 121) en base a la distancia $D_r^s\left((\mu, \Sigma), (\mu^*, \Sigma^*)\right)$ con $s = 2$ y $r = 0.5$. En este sentido, los datos siguientes indican la longitud de la cabeza en el primer y segundo hijo correspondiente a 25 familias.

Primer hijo:	191 195 181 183 176 208 189 197 188 192 179 183 174 190 188 163 195 186 181 175 192 174 176 197 190.
Segundo hijo:	179 201 185 188 171 192 190 189 197 187 186 174 185 195 187 161 183 173 182 165 185 178 176 200 187.

En primer lugar se supondrá que la matriz de varianzas covarianzas es conocida y se desea hacer un contraste de igualdad de medias:

1) **Contrastar $\boldsymbol{H}_0$: $(\mu_1, \mu_2) = (182, 182)$. Supuesto que**

$$\boldsymbol{\Sigma} = \begin{pmatrix} 100 & 0 \\ 0 & 100 \end{pmatrix},$$

Puesto que $\hat{\mu}_1 = \overline{x}_1 = 185.72$ y $\hat{\mu}_2 = \overline{x}_2 = 183.84$, el estadístico (8) toma el valor

$$T = \frac{2n(s-1)\left(\exp\left\{\frac{r(s-1)}{2}\left((\hat{\mu}-\mu_0)^t(\Sigma)^{-1}(\hat{\mu}-\mu_0)\right)\right\}-1\right)}{r} =$$

$$= 100\left(\exp\left\{0.25\left((3.74, 1.84)\begin{pmatrix} 100 & 0 \\ 0 & 100 \end{pmatrix}^{-1}\begin{pmatrix} 3.72 \\ 1.84 \end{pmatrix}\right)\right\}-1\right) = 4.39$$

Por tanto al ser $\chi^2_{2,0.05} = 5.99$ se acepta la hipótesis nula con una confianza del 95%.

Ahora se supondrá que (μ_1, μ_2) es conocido y se efectuará un contraste de igualdad de matrices de varianzas covarianzas.

2) **Contrastar, $\boldsymbol{H}_0 : \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0 = \begin{pmatrix} 100 & 0 \\ 0 & 100 \end{pmatrix}$. Supuesto que $\mu = (182, 182)$.**

En este caso al ser μ conocido, $\hat{\Sigma} = S + dd^t$ donde $d = \overline{x} - \mu_0$ y S es la matriz de varianzas covarianzas muestral. Por tanto se tiene

$$\hat{\Sigma} = \begin{pmatrix} 93.481 & 66.875 \\ 66.875 & 96.775 \end{pmatrix} + \begin{pmatrix} 13.83 & 6.84 \\ 6.84 & 3.38 \end{pmatrix} = \begin{pmatrix} 107.31 & 73.715 \\ 73.71 & 100.110 \end{pmatrix}$$

En este caso el estadístico (8) toma el valor

$$\begin{aligned} T &= 2n\frac{(s-1)^{-1}}{r}\left(\frac{\left|r\Sigma_0 + (1-r)\hat{\Sigma}\right|^{\frac{1-s}{2(r-1)}}}{|\Sigma_0|^{\frac{s-1}{2}}\left|\hat{\Sigma}\right|^{\frac{(1-s)r}{2(r-1)}}} - 1\right) = \\ &= 100\left(\frac{\left|0.5\hat{\Sigma} + 0.5\Sigma_0\right|}{|\Sigma_0|^{1/2}\left|\hat{\Sigma}\right|^{1/2}} - 1\right) \end{aligned}$$

Ahora bien, por un lado

$$0.5\hat{\Sigma} + 0.5\Sigma_0 = \begin{pmatrix} 103.65 & 36.85 \\ 36.85 & 100.05 \end{pmatrix}$$

y por otro, $\left|\hat{\Sigma}\right| = 5308.967$, $|\Sigma_0| = 10000$ y $\left|0.5\hat{\Sigma} + 0.5\Sigma_0\right| = 9012.725$. Con lo cual se tiene que $T = 23.69$. Por otro lado $\chi^2_{3,0.05} = 7.82$ y en consecuencia se rechaza la hipótesis nula.

3) Finalmente, vamos a contrastar

$\boldsymbol{H}_0$: $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0 = \begin{pmatrix} \mathbf{100} & \mathbf{0} \\ \mathbf{0} & \mathbf{100} \end{pmatrix}$, $\boldsymbol{\mu} = (\mathbf{182}, \mathbf{182})$ **frente**

$\boldsymbol{H}_1$: $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0 \neq \begin{pmatrix} \mathbf{100} & \mathbf{0} \\ \mathbf{0} & \mathbf{100} \end{pmatrix}$, $\boldsymbol{\mu} \neq (\mathbf{182}, \mathbf{182})$

En este caso el estadístico, viene dado por

$$\begin{aligned} T_3 &= \frac{2n}{(1-s)r}\left(\exp\left\{\frac{r(s-1)}{2}\left((\hat{\mu}-\mu_0)^t\left(r\Sigma_0+(1-r)S\right)^{-1}(\hat{\mu}-\mu_0)\right)\right\}\cdot\right.\\ &\quad\left.\cdot\ \frac{\left|rS+(1-r)\hat{\Sigma}\right|^{\frac{1-s}{2(r-1)}}}{|S|^{\frac{s-1}{2}}\left|\hat{\Sigma}\right|^{\frac{(1-s)r}{2(r-1)}}}-1\right) = \\ &= 100\left(\exp\left\{0.25\left[(3.74, 1.84)\left(\begin{pmatrix} 93.481 & 66.875 \\ 66.875 & 96.775 \end{pmatrix}0.5+\right.\right.\right.\right.\\ &+\left.\left.\left.0.5\begin{pmatrix} 100 & 0 \\ 0 & 100 \end{pmatrix}\right)^{-1}\begin{pmatrix} 3.72 \\ 1.84 \end{pmatrix}\right]\right\}\cdot\\ &\quad\left.\cdot\ \frac{\left|\begin{pmatrix} 93.481 & 66.875 \\ 66.875 & 96.775 \end{pmatrix}0.5+0.5\begin{pmatrix} 100 & 0 \\ 0 & 100 \end{pmatrix}\right|}{\left|\begin{pmatrix} 100 & 0 \\ 0 & 100 \end{pmatrix}\right|^{1/2}\left|\begin{pmatrix} 93.481 & 66.875 \\ 66.875 & 96.775 \end{pmatrix}\right|^{1/2}}-1\right) = 28.95 \end{aligned}$$

Por otro lado, $\chi^2_{5,0.05} = 11.07$ y en consecuencia se rechaza la hipótesis nula.

6. BIBLIOGRAFÍA

[1] **Fraser, D.A.S.** (1957). "Nonparametric Methods in Statistics", John Wiley and Sons. New York.

[2] **Kullback, S.** and **Leibler, A.** (1951). "On the Information and Sufficiency", *Ann. Math. Statist.,* **22**, 79–86.

[3] **Kupperman, M.** (1957). "Further applications of information theory to multivariate analysis and statistical inference", Ph. D. dissertation on file at the library of The George Washington University.

[4] **Mardia, K.V.; Kent, J.T.** and **Bibby, J.T.** (1979). *Multivariate Analysis.* Academic Press, New York.

[5] **Menéndez, M.L.; Morales, D.; Pardo, L.** y **Salicrú, M.** (1991). "Some statistical application of (r,s)−directed divergences". To appear in *Utilitas Mathematica.*

[6] **Morales, D.; Pardo, L.; Salicrú, M.** y **Menéndez, M.L.** (1991). "A test of independence based on the (r,s)−directed divergence". To appear in *TamKang Journal of Mathematics.*

[7] **Pardo, L., Pardo, J.A.** y **Menéndez, M.L.** (1990). "The convexity with respect to gaussian distribution of (r,s)−directed divergence". Communicated.

[8] **Salicrú, M.; Menéndez, M.L.; Morales, D.** y **Pardo, L.** (1991). "Asymptotic properties of unified (r,s)−directed divergence in a stratified sampling". Communicated.

[9] **Serfling, R.J.** (1980). "Approximation theorems of Mathematical Statistics". Wiley, New York.

[10] **Sharma, B.D.** y **Mittal, D.P.** (1977). "New Non-additive Measures of Relative Information". *J. Comb. Inform. and Syst. Sci.,* **2**, 122–133.

ENGLISH SUMMARY:

TEST OF HYPOTHESES ON THE (r,s)–DIVERGENCE: APPLICATIONS TO MULTINOMIAL AND MULTIVARIATE NORMAL DISTRIBUTIONS

Morales, D., Pardo, L., Salicrú, M. and Menéndez, M.L.

Let $(\mathfrak{X}, \beta_{\mathfrak{X}}, P_\theta)_{\theta\in\Theta\subset\mathbb{R}^k}$ be a statistical space dominated by a measure μ σ–finite, such that $f_\theta(x) = \left(\frac{dP_\theta}{d\mu}\right)(x)$. Sharma and Mittal (1977) define the (r,s)–divergence between two probability density functions $f_{\theta_1}(x)$ and $f_{\theta_2}(x)$ as follows

$$D_r^s(\theta_1,\theta_2) = (s-1)^{-1}\left\{\left(\int_{\mathfrak{X}} f_{\theta_1}(x)^r\, f_{\theta_2}(x)^{1-r}\, d\mu(x)\right)^{\frac{s-1}{r-1}} - 1\right\},$$

$$r \neq 1, s \neq 1, r > 0$$

where $\theta_1 = (\theta_{11},\ldots,\theta_{1k})$ and $\theta_2 = (\theta_{21},\ldots,\theta_{2k})$.

When $s \longrightarrow 1$ or $r \longrightarrow 1$ it is possible to define the (r,s)–divergence by continuity. In this sense, when $s \longrightarrow 1$ the divegence given by Renyi (1961) is obtained and when $r \longrightarrow 1$ and $s \longrightarrow 1$ the divergence given by Kullback-Leibler (1951) is obtained. Under the usual regularity assumptions, the asymptotic distributions of $D_r^s(\hat\theta_1,\theta_2)$ and $D_r^s(\hat\theta_1,\hat\theta_2)$ are obtained, where $\hat\theta_1$ and $\hat\theta_2$ are the maximum likelihood estimators of θ_1 and θ_2 based on samples of sizes n and m respectively. By using these results and those one obtained by Pardo L. *et al.* (1990), a widerange of statistical hypothesis about the mean vectors and the variance-covariance matrices of n–variate normal distributions can be tested.

First we find the asymptotic distribution of $D_r^s(\hat\theta_1,\theta_2)$, when θ_2 is known and $\hat\theta_1$ is the maximum likelihood estimator of θ_1. We have obtained the following result.

Theorem 1

If $\theta_1 \neq \theta_2$, then

$$n^{1/2}\left(D_r^s(\hat\theta_1,\theta_2) - D_r^s(\theta_1,\theta_2)\right) \xrightarrow[n\to\infty]{L} \mathrm{N}\left(0, T^t I_F^{-1}((\theta_1)T\right),$$

where $I_F(\theta_1)$ is the Fisher information matrix and $T = (t_1, \dots, t_k)^t$ with

$$t_i = \frac{r}{r-1}\left(\int_{\mathcal{X}} f_{\theta_1}(x)^r f_{\theta_2}(x)^{1-r}\, d\mu\right)^{\frac{s-1}{r-1}} \int_{\mathcal{X}} \left(\frac{f_{\theta_1}(x)}{f_{\theta_2}(x)}\right)^{r-1} \frac{\partial f_{\theta_1}(x)}{\partial \theta_{1i}}\, d\mu$$

when $T^t I_F^{-1}(\theta_1) T > 0$.

The result can be proved as follows. Consider the Taylor series expansion of the function

$$h_r^s(\theta_{11}, \dots, \theta_{1k}) = \frac{1}{s-1}\left(\left[\int_{\mathcal{X}} f_{\theta_1}(x)^r f_{\theta_2}(x)^{1-r}\, d\mu\right]^{\frac{s-1}{r-1}} - 1\right)$$

at the point $\hat{\theta}_1 = (\hat{\theta}_{11}, \dots \hat{\theta}_{1k})$ in a neighbourhood of the point $\theta_1 = (\theta_{11}, \dots, \theta_{1k})$. Study the asymptotic distribution of the first order term by using the Central Limit Theorem.

When $\theta_1 = \theta_2$, the first order term equals to zero, so it is necessary to use the second order term and to study its asymptotic distribution. In this last case, we have obtained the following result.

Theorem 2

If $\theta_1 = \theta_2$, then

$$\frac{2n D_r^s(\hat{\theta}_1, \theta_2)}{r} \xrightarrow[n\to\infty]{L} \chi_k^2$$

When θ_1 and θ_2 are both unknown, we estimate them with maximum likelihood estimators based on independent random samples of sizes n and m respectively. Similar arguments to those given in theorems 1 and 2 yields to:

Theorem 3

If $\theta_1 \neq \theta_2$, then

$$(nm)^{1/2}\left(D_r^s(\hat{\theta}_1, \hat{\theta}_2) - D_r^s(\theta_1, \theta_2)\right) \xrightarrow[n,m\to\infty]{L} \mathrm{N}\left(0, mT^t I_F^{-1}(\theta_1) T + nS^t I_F^{-1}(\theta_2) S\right),$$

where T has been defined in theorem 1 and S is a vector whose components are:

$$s_i = -\left(\int_{\mathcal{X}} f_{\theta_1}(x)^r f_{\theta_2}(x)^{1-r}\, d\mu\right)^{\frac{s-r}{r-1}} \int_{\mathcal{X}} \left(\frac{f_{\theta_1}(x)}{f_{\theta_2}(x)}\right)^{r-1} \frac{\partial f_{\theta_1}(x)}{\partial \theta_{2i}} d\mu$$

when $mT^t I_F^{-1}(\theta_1) T + nS^t I_F^{-1}(\theta_2) S \neq 0$.

Theorem 4

If $\theta_1 = \theta_2$, then

$$\frac{2nm}{n+m}\frac{D_s^r(\hat{\theta}_1, \hat{\theta}_2)}{r} \xrightarrow[n,m\to\infty]{L} \chi_k^2$$

The previous theorems can be used to test for goodness of fit and homogeneity. Theorem 2 allows us to test $H_0: \theta = \theta_1$ against $H_1: \theta \lessgtr \theta_1$, by using the statistics

$$T = \frac{2nD_r^s(\hat{\theta}_1, \theta_1)}{r}$$

We reject the null hypothesis if $T > \chi^2_{k(\alpha)}$. The asymptotic power of this test can be obtained from theorem 1 and its consistency from Fraser (1957).

Theorem 4 allows us to test that two independent random samples from $f_{\theta_1}(x)$ and $f_{\theta_2}(x)$ respectively, are such that $\theta_1 = \theta_2$. We reject $H_0: \theta_1 = \theta_2$ if

$$T = \frac{2mn}{m+n}\frac{D_r^s(\hat{\theta}_1, \hat{\theta}_2)}{r} > \chi^2_{k(\alpha)}.$$

Similarly its asymptotic power and consistency can be obtained.

After discretizing in different categories a random variable, the following hypotheses can be tested with one or more samples

(1) Test for goodness of fit $H_0: p = p_0$ (p_0 known)

(2) Test for goodness of fit of m populations $H_0: p^{(1)} = \cdots = p^{(m)} = p_0$ (p_0 known)

(3) Test for homogeneity of two populations, $H_0: p = q$

(4) Test for homogeneity of m populations, $H_0: p^{(1)} = \cdots = p^{(m)}$.

When $f_{\theta_1}(x)$ and $f_{\theta_2}(x)$ are taken to be multinomial distributions $\mathcal{M}(p_1, \ldots, p_k)$ and $\mathcal{M}(q_1, \ldots, q_k)$, the results of Menéndez *et al.* (1991) are obtained as particular cases.

On the other hand, when $f_{\theta_1}(x)$ and $f_{\theta_2}(x)$ are multivariate normal distributions, it is possible to test

(1) $H_0: \mu = \mu_0$ with Σ known,

(2) $H_0: \Sigma = \Sigma_0$ with μ known

(3) $(\mu, \Sigma) = (\mu_0, \Sigma_0)$ with μ_0 and Σ_0 both known.

Furthermore, when dealing with two samples, the following hypotheses can also be tested:

(1) $H_0\colon \mu_1 = \mu_2$ with Σ_1 and Σ_2 known

(2) $H_0\colon (\mu_1, \Sigma_1) = (\mu_2, \Sigma_2)$.