TRABAJOS DE ESTADISTICA
Vol. 5, núm. 1, 1990, pp. 69 a 79

CONVEXIDAD Y SIMETRIA DE LA *J*-DIVERGENCIA GENERALIZADA

M. SALICRU PAGÉS
Departamento de Estadística
Universidad de Barcelona

RESUMEN

En este trabajo, se caracteriza la simetría de la J-divergencia generalizada en términos del parámetro y de la función que la determina. Se plantea seguidamente la convexidad y simetrización en función del parámetro, atendiendo a la forma de la función $\phi(t)$ que la determina. Finalmente, se revisa la convexidad en función de las variables atendiendo a la concavidad y convexidad, de las funciones $\phi(t)$ y $1/\phi''(t)$.

Palabras clave: J-divergencias, análisis de convexidad, análisis de datos.
Clasificación AMS: 62H30, 62-07.

SUMMARY

In this work the symmetry of the generalized J-divergence is characterized by its parameter and function.

The convexity and symmetrization are studied as a function of the parameter, according to the shape of the function $\phi(t)$ that determines this generalized J-divergence.

Finally, the convexity as a function of the variables, according to the convexity of the functions $\phi(t)$ and $1/\phi''(t)$ is revised.

Key words: J-divergence, convex analysis, data analysis.
AMS Classification: 62H30, 62-07.
Title: Convexity and simmetry of some generalized J-divergence.

Recibido, enero 1989. Revisado, julio 1989.

1. INTRODUCCION

Un problema interesante que se plantea en análisis de datos proviene de la elección adecuada de medidas que nos permitan cuantificar las diferencias entre poblaciones o entre individuos de una misma población. En este sentido, las J-divergencias generalizadas pueden entenderse como medidas que cuantifican analogías y diferencias entre distribuciones o poblaciones.

Como una aproximación a la caracterización de las divergencias, analizamos en este trabajo las condiciones para la convexidad, simetría y simetrización de las J-divergencias generalizadas, completando y extendiendo los aspectos relativos a J-divergencias presentados por Burbea y Rao (1982).

Estudios en este sentido resultan muy adecuados para abordar aspectos relacionados con la teoría de la información e inferencia estadística, siendo destacables, entre otros, los trabajos de Matusita (1955, 1964), Pitman (1979), Wald (1950) y Shaked (1977) y en otro orden, Burbea y Rao (1987) y Salicrú y Calvo (1988) obtienen relaciones con las ϕ-entropías.

2. DEFINICIONES PREVIAS

Dada una medida positiva μ, σ-aditiva y σ finita definida en una σ-álgebra de subconjuntos de un espacio medible X, y una función a valores reales $\phi(t)$, dos veces diferenciable con continuidad y definida en un intervalo T_ϕ, con $[0,1] \subset T_\phi \subset [0,\infty)$, se define el espacio de las funciones ϕ-integrables como el conjunto

$$L_\phi^1 = \left\{ p\,;\, p : X \to R \quad \text{t.q.} \quad \begin{array}{l} 1) \displaystyle\int_X |p(x)|\,d\mu(x) < +\infty \\ 2) \; p(x) \in T_\phi \quad \text{para todo} \quad x \in X \end{array} \right\}$$

Para diferenciar dos funciones ϕ-integrables $p(x)$ y $q(x)$, se define la J-divergencia generalizada a partir de la expresión

$$J_\phi^\lambda(p,q) = \int_X [\lambda\phi(p) + (1-\lambda)\phi(q) - \phi(\lambda p + (1-\lambda)q)]\, d\mu(x) \qquad \text{para} \qquad \lambda \in (0,1)$$

donde si $\mu(x)$ es una medida atómica, la J-divergencia generalizada se reduce a la igualdad

$$J_\phi^\lambda(p, q) = \sum_{i=1}^{\infty} a_i[\lambda\phi(x_i) + (1 - \lambda)\phi(y_i) - \phi(\lambda x_i + (1 - \lambda)y_i)]$$

3. PROPIEDADES PREVIAS

De sencillos cálculos y particularizaciones con la definición de la J-divergencia generalizada, se obtienen de forma inmediata los siguientes resultados.

Proposición 3.1

$$J_\phi^\lambda(f(p), f(q)) = \lambda(1 - \lambda) \int_X |f(p) - f(q)|^2 \, d\mu(x) \text{ para } \phi(x) = x^2 - x.$$

Obsérvese, que al particularizar a $f(p) = \sqrt{p}$, la expresión anterior se reduce a

$$J_\phi^\lambda(\sqrt{p}, \sqrt{q}) = 2\lambda(1 - \lambda)\left(1 - \int_X \sqrt{pq} \, d\mu(x)\right) = 2\lambda(1 - \lambda)(1 - \cos(p, q))$$

siendo $p(x)$, $q(x)$ funciones de densidad de probabilidad, en las que se ha definido el producto escalar

$$\langle p, q \rangle = \int_X \sqrt{p(x)q(x)} \, d\mu(x)$$

Así, en este caso la J-divergencia se expresa en función de la distancia Hellinger y de la distancia Bhattacharyya.

Proposición 3.2

$J_\phi^\lambda(p, q) \geqslant 0$ para todas las funciones $p(x)$, $q(x)$ de L_ϕ^1 si y sólo si la función $\phi(t)$ es convexa en T_ϕ.

Demostración

Si $\phi(x)$ no fuera convexa en un punto $x_0 \in T_\phi$, entonces existiría otro punto $x_1 \in T_\phi$ que cumpliría

$$\lambda\phi(x_0) + (1 - \lambda)\phi(x_1) < \phi(\lambda x_0 + (1 - \lambda)x_1)$$

para los que al considerar las funciones

$$p(x) = \begin{cases} x_0 & \text{para} \quad x \in K \\ 0 & \text{para} \quad x \notin K \end{cases} \qquad q(x) = \begin{cases} x_1 & \text{para} \quad x \in K \\ 0 & \text{para} \quad x \notin K \end{cases}$$

con K subconjunto de X de medida finita, se tendría que $J_\phi^\lambda(p, q) < 0$.

El recíproco resulta evidente al reducirse la J-divergencia a la integral de una función positiva.

4. CONDICIONES PARA LA SIMETRIA DE LA J-DIVERGENCIA GENERALIZADA

Estudiada la no negatividad de la J-divergencia generalizada, resulta interante el estudio de las condiciones que debemos imponer al parámetro λ para que fijada una función $\phi(t)$, la J-divergencia generalizada resulte simétrica. En este sentido, estudiaremos previamente las funciones que cumplen la condición $f(\lambda x + (1 - \lambda)y) = f(\lambda y + (1 - \lambda)x)$ para todo $x, y \in T_\phi$ siendo $T_\phi = [a, b] \subset R$.

Lema 4.1

Fijado un valor de λ, las únicas funciones continuas $f : T_\phi \to R$ que cumplen la condición, $f(\lambda x + (1 - \lambda)y) = f(\lambda y + (1 - \lambda)x)$ para todo $x, y \in T_\phi$ son las de la forma $f(x) = \text{cte}$ para todo $x \in T_\phi$.

Demostración

Dados $x_0 \in (a, b)$ y λ fijos, $\lambda \in (0, 1/2)$, existen dos puntos de (a, b)

$$\hat{x}_0 = \left[\frac{\lambda(k+1)-k}{2\lambda-1}(x_0-a)\right]+a \quad , \quad \hat{y}_0=\left[\frac{\lambda(k+1)-1}{2\lambda-1}(x_0-a)\right]+a$$

para valores de k cumpliendo

$$k > \frac{x_0 + 2\lambda - 1}{\lambda x_0} - 1 \quad \text{y} \quad \frac{\lambda}{1 - \lambda} < k < 1$$

Para tales valores, al aplicarles la condición

$$f(\lambda \hat{x}_0 + (1 - \lambda)\hat{y}_0) = f(\lambda \hat{y}_0 + (1 - \lambda)\hat{x}_0) \tag{1}$$

se obtiene

$$f(x_0) = f(a + k(x_0 - a))$$

y al aplicar de forma recurrente la expresión anterior

$$f(x_0) = f(a + k^n(x_0 - a))$$

con lo que al tomar límites

$$f(x_0) = \lim_{n \to \infty} f(a + k^n(x_0 - a)) = f(a)$$

Así, las únicas funciones que cumplen la condición (1) son las funciones de la forma $f(x) = \text{cte}$ en T_ϕ.

De aplicar al lema anterior a las J-divergencias generalizadas, hemos obtenido una caracterización de la simetría que se expresa en los términos siguientes

Teorema 4.1

a) Las J-divergencias $J_\phi^\lambda(p, q)$ son simétricas p y q para toda función de la forma $\phi(x) = ax^2 + bx + c$.

b) Las J-divergencias $J_\phi^\lambda(p, q)$ son simétricas en p y q para las funciones $\phi(x) \neq ax^2 + bx + c$ dos veces diferenciables con continuidad si y sólo si $\lambda = \frac{1}{2}$.

Demostración a)

Para $\phi(x) = ax^2 + bx + c$, $J_\phi^\lambda(p, q) = \int_X a\lambda(1-\lambda)(p-q)^2\, d\mu(x)$ que es simétrica en p y q.

Demostración b)

De $J_\phi^\lambda(p, q)$ simétrica para toda $p(x)$, $q(x) \in L_\phi^1$ y de tomar

$$p(x) = \begin{cases} s \text{ para } x \in K \\ 0 \text{ para } x \notin K \end{cases} \qquad q(x) = \begin{cases} s \text{ para } x \in K \\ 0 \text{ para } x \notin K \end{cases} \qquad \text{con } 0 \neq \mu(K) < +\infty$$

se tendrá para todo s, $t \in T_\phi$.

$$[\lambda\phi(s) + (1 - \lambda)\phi(t) - \phi(\lambda s + (1 - \lambda)t)]\mu(K) =$$
$$= [(1 - \lambda)\phi(s) + \lambda\phi(t) - \phi((1 - \lambda)s + \lambda t)]\mu(K)$$

Simplificando la expresión anterior, se obtiene

$$(2\lambda - 1)[\phi(s) - \phi(t)] = \phi(\lambda s + (1-\lambda)t) - \phi((1-\lambda)s + \lambda t) = F(s,t)$$

y derivando respecto de s y t

$$0 = [\phi''(\lambda s + (1-\lambda)t) - \phi''((1-\lambda)s + \lambda t)](1-\lambda)\lambda = \frac{\partial^2 F(s,t)}{\partial s \partial t}$$

Así, para $\phi''(x)$ se cumple la condición del lema 4.1.

$$\phi''(\lambda s + (1-\lambda)t) = \phi''((1-\lambda)s + \lambda t) \quad \text{para todo} \quad s, t \in T_\phi$$

de la que resulta $\phi''(x) = \text{cte}$ para todo $x \in T_\phi$ cuando $\lambda \neq \frac{1}{2}$, y por ello, se obtiene la simetría para funciones con $\phi''(x) \neq \text{cte}$ sólo para el valor de $\lambda = \dfrac{1}{2}$.

5. CONVEXIDAD DE LA *J*-DIVERGENCIA GENERALIZADA RESPECTO DEL PARAMETRO

Prosiguiendo con el estudio de la J-divergencia generalizada atendiendo a los valores que toma el parametro, analizamos seguidamente las condiciones exigibles a la función $\phi(t)$ para que J-divergencia sea convexa en función del parámetro. En este sentido, con el supuesto que $J_\phi^\lambda(p,q)$ es dos veces derivable bajo el signo integral y que el soporte de $\lambda\phi(p) + (1-\lambda)\phi(q) - \phi(\lambda p + (1-\lambda)q)$ no depende del parámetro, hemos obtenido

Proposición 5.1

$J_\phi^\lambda(p,q)$ es convexa (cóncava) en $(0,1)$ respecto de λ para todas las posibles parejas $p(x), q(x) \in L_\phi^1$ si y sólo si $\phi(t)$ es cóncava (convexa) en T_ϕ.

Demostración

De la convexidad de $J_\phi^\lambda(p,q)$ respecto del parámetro λ resulta para toda $p(x)$, $q(x) \in L_\phi^1$

$$\int_X \phi''[\lambda p(x) + (1-\lambda)q(x)][p(x) - q(x)]^2 \, d\mu(x) < 0 \quad \text{para} \quad \lambda \in (0,1) \tag{2}$$

Por otro lado, si $\phi''(t)$ fuera positiva en un punto *"a" de* T_ϕ, existiría un entorno de "a" en T_ϕ cumpliendo $\phi''(t) > 0$ para $t \in (a - \delta, a + \delta)$, y al tomar

$$p(x) = \begin{cases} a & \text{si} \quad x \in K \\ 0 & \text{si} \quad x \notin K \end{cases} \qquad q(x) = \begin{cases} a + \dfrac{\delta}{2} & \text{si} \quad x \in K \\ 0 & \text{si} \quad x \notin K \end{cases}$$

con $0 \neq \mu(K) < +\infty$, resultaría

$$\int_X \phi''[\tfrac{1}{2}p(x) + \tfrac{1}{2}q(x)][p(x) - q(x)]^2 \, d\mu(x) = \int_X \phi''\left(a + \frac{\delta}{4}\right)\frac{\delta^2}{4} \, d\mu(x) > 0$$

que contradice (2) en el valor $\lambda = \frac{1}{2}$.

El recíproco es evidente, pues de la concavidad de $\phi(t)$ se deduce (2) y a su vez la convexidad de $J_\phi^\lambda(p, q)$.

Proposición 5.2

$\left(\dfrac{\partial}{\partial\lambda} J_\phi^\lambda(p, q)\right)(\lambda) = K(p, \lambda p + (1 - \lambda)q) - K(q, \lambda p + (1 - \lambda)q)$ siendo $\phi(t) = t \log t, K(p, q)$ la diferencia de información entre $p(x)$ en el sentido de Kullback-Leibler y las funciones $p(x)$ y $q(x)$ cumpliendo la condición $\|p\| = \|q\| = 1$ para el producto escalar $\langle p, q \rangle$, $\langle p, q \rangle = \displaystyle\int_X \sqrt{p(x) \cdot q(x)} \, d\mu(x)$.

Demostración

$$\left(\frac{\partial}{\partial\lambda} J_\phi^\lambda(p, q)\right)(\lambda) =$$

$$= \int_X [p \log p - q \log q - p \log(\lambda p + (1 - \lambda)q + q \log(\lambda p + (1 - \lambda)q)] \, d\mu =$$

$$= K(p, \lambda p + (1 - \lambda)q) - K(q, \lambda p + (1 - \lambda)q)$$

Corolario 5.1

En las condiciones de la proposición 5.2, se tiene de forma inmediata los resultados siguientes:

a) $\left(\dfrac{\partial}{\partial\lambda} J_\phi^\lambda(p, q)\right)(0) = K(p, q)$

b) $\left(\frac{\partial}{\partial\lambda} J_\phi^\lambda(p,q)\right)(1) = -K(q,p)$

Corolario 5.2

En las condiciones de la proposición 5.2, existe un único λ_0 tal que cumple la condición:

$$K(p, \lambda_0 p + (1-\lambda_0)q) = K(q, \lambda_0 p + (1-\lambda_0)q)$$

Demostración

Al ser la función $\frac{\partial}{\partial\lambda} J_\phi^\lambda(p,q)$ una función decreciente en λ y tomar distintos signos en los extremos del intervalo $[0, 1]$, resulta inmediata la existencia de un único $\lambda_0 \in (0, 1)$ cumpliendo

$$\left(\frac{\partial}{\partial\lambda} J_\phi^\lambda(p,q)\right)(\lambda_0) = 0$$

condición que equivale a

$$K(p, \lambda_0 p + (1-\lambda_0)q) = K(q, \lambda_0 p + (1-\lambda_0)q)$$

Proposición 5.3

La simetrización ${}^sJ_\phi^\lambda(p,q) = J_\phi^\lambda(p,q) + J_\phi^{1-\lambda}(p,q)$ alcanza el máximo en $\lambda = \frac{1}{2}$ cuando $\phi(t)$ es una función estrictamente convexa.

Demostración

Cuando la función $\phi(t)$ es convexa estricta, las funciones $J_\phi^\lambda(p,q)$ y $J_\phi^{1-\lambda}(p,q)$ son cóncavas en $(0, 1)$. Así, la función simetrizada es también cóncava estricta en $(0, 1)$ y el máximo respecto de λ es único.

Por otro lado, al cumplirse las condiciones

$$\frac{\partial}{\partial\lambda}({}^sJ_\phi^\lambda(p,q))(\tfrac{1}{2}) = 0 \qquad \text{y} \qquad \frac{\partial^2}{\partial\lambda^2}({}^sJ_\phi^\lambda(p,q)(\tfrac{1}{2}) < 0$$

el máximo de ${}^sJ_\phi^\lambda(p,q)$ se alcanza en $\lambda = \frac{1}{2}$.

6. CONVEXIDAD DE LA *J*-DIVERGENCIA GENERALIZADA RESPECTO DE $p(x), q(x)$

De forma paralela al estudio de la convexidad de la diferencia de Jensen realizado por Burbea y Rao (1982), hemos obtenido para la J-divergencia generalizada el siguiente resultado.

Proposición 6.1

$J_\phi^\lambda(p, q)$ es convexa (cóncava en $L_\phi^1 \times L_\phi^1$ si y sólo si $\phi(t)$ es convexa (cóncava) en T_ϕ y $\dfrac{1}{\phi''}(t)$ es cóncava (convexa) en T_ϕ.

Demostración

Para que $J_\phi^\lambda(p, q)$ sea convexa en $L_\phi^1 \times L_\phi^1$, la matriz de la forma cuadrática que define el Hessiano debe ser semidefinida positiva. En nuestro caso, el cálculo se reduce a

$$d^2 J_\phi((p, q); (f, g), (f, g)) =$$
$$= \int_X [a_1(p, q) f^2 + a_2(p, q) g^2 + 2b(p, q) f \cdot g] \, d\mu(x)$$

con

$$a_1(p, q) = \lambda\phi''(p) - \lambda^2\phi''(\lambda p + (1 - \lambda)q)$$
$$a_2(p, q) = (1 - \lambda)\phi''(q) - (1 - \lambda)^2\phi''(\lambda p + (1 - \lambda)q)$$
$$b(p, q) = \lambda(1 - \lambda)\phi''(\lambda p + (1 - \lambda)q)$$

y

$$M(p, q) = \begin{pmatrix} a_1(p, q) & b(p, q) \\ b(p, q) & a_2(p, q) \end{pmatrix} \text{ semidefinida positiva.}$$

Al tomar $p = q$ se tiene

$$0 \leqslant a_1(p, q) = \lambda\phi''(p) - \lambda^2\phi''(p) = (\lambda - \lambda^2)\phi''(p(x)) \text{ para todo } p(x) \in L_\phi^1$$

de donde resulta $\phi''(t) \geqslant 0$ para todo t.

Por otro lado, de $a_1(p, q)a_2(p, q) - (b(p, q))^2 \geqslant 0$, se obtiene

$$\lambda(1 - \lambda)\phi''(p)\phi''(q)\phi''(\lambda p + (1 - \lambda)q)[(\phi'')^{-1}(\lambda p + (1 - \lambda)q) -$$

$$- (\phi'')^{-1}(p) - (1 - \lambda)(\phi'')^{-1}(q)] \geqslant 0$$

y de la convexidad de $\phi''(t)$ es inmediata la concavidad de $\dfrac{1}{\phi''}(t)$.

Recíprocamente, de la concavidad de $\dfrac{1}{\phi''}(t)$ y la convexidad de ϕ se obtiene de forma inmediata

$$a_1(p, q)a_2(p, q) - (b(p, q))^2 \geqslant 0$$

Por otro lado, al considerar los valores,

$$r_1 = \left(\frac{\lambda}{\phi''(p)}\right)^{1/2} \quad , \quad r_2 = \left(\frac{1-\lambda}{\phi''(q)}\right)^{1/2} \quad , \quad s_1 = (\lambda\phi''(p))^{1/2} \quad , \quad s_2 = 0$$

se tiene

$$r_1^2 + r_2^2 = \frac{\lambda}{\phi''(p)} + \frac{1 - \lambda}{\phi''(q)} \leqslant \frac{1}{\phi''(\lambda p + (1 - \lambda)q)}$$

y sustituyendo en la desigualdad de Cauchy

$$\frac{1}{\phi''(\lambda p + (1 - \lambda)q)} \cdot \lambda\phi''(p) \geqslant \lambda^2$$

de donde

$$a_1(p, q) = \lambda\phi''(p) - \lambda^2\phi''(\lambda p + (1 - \lambda)q) \geqslant 0$$

Corolario 6.1

La diferencia de Jensen definida por

$$J_\phi(p, q) = \int_X \left(\frac{1}{2}[\phi(p) + \phi(q)] - \phi\left(\frac{p + q}{2}\right)\right) d\mu$$

es convexa (cóncava $L_\phi^1 \times L_\phi^1$ si y sólo si $\phi(t)$ es convexa (cóncava) y $\dfrac{1}{\phi''}(t)$ es cóncava (convexa) en T_ϕ.

Corolario 6.2

$J^{\lambda}_{\phi_\alpha}(p, q)$ es convexa en $L^1_{\phi_\alpha} \times L^1_{\phi_\alpha}$ si y sólo si $\alpha \in [1, 2] \cup \{0\}$ y no es nunca cóncava en $L^1_{\phi_\alpha} \times L^1_{\phi_\alpha}$ para

$$\phi_\alpha(\mathrm{x}) = \begin{cases} (\alpha - 1)^{-1}(x^\alpha - x) & \text{para} \quad \alpha \neq 1 \\ x \log x & \text{para} \quad \alpha = 1 \end{cases}$$

AGRADECIMIENTOS

Agradezco a los referees de este trabajo las valiosas sugerencias que han servido para la mejora en los contenidos de este escrito.

BIBLIOGRAFIA

BURBEA, J., y RAO, C. R. (1982): «Entropy differential metric, distance and divergence measures in probability spaces: a unified approach», *J. Multivariate Anal.*, 12, 757-596.

BURBEA, J., y RAO, C. R. (1984): «Differential metrics in probability spaces», *Probability and Mathematical Statistics*, 3, 241-258.

MATUSITA, K. (1955): «Decision rules, based on the distance, for problems of fit, two samples, and estimation», *Annals Math. Stat.*, 26, 631-640.

MATUSITA, K. (1964): «Distance and decision rules», *Annals of the Inst. Statistica Math.*, 14, 305-315.

PITMAN, E. J. G. (1979): «Some basic theory for statistical inference», *John Wiley and Sons.*, New York.

SALICRU, M., y CUADRAS, C. M. (1987): «Funciones de entropía asociadas a medidas de Csiszar», *Qüestió*, 11(3), 3-12.

SALICRU, M., y CALVO, M. (1988): «Medidas de incertidumbre asociadas a *J*-divergencias», *Trabajos de Estadística* (pendiente de publicación).

SHAKES, M. (1977): «Statistical inference for a class of life distributions», *Com. Stat. Theo. Math.*, A 6(13), 1323-1339.

WALD, A. (1950): *Statistical decision functions*, Wiley, New York.