Qüestiió - V. 9, n.º 4 (desembre 1985) pp. 251-255

SOBRE CIERTAS PROPIEDADES DE LA M-DIVERGENCIA EN ANÁLISIS DE DATOS

M. SALICRÚ Y A. ARCAS
UNIVERSIDAD DE BARCELONA

En este trabajo se desarrollan ciertas propiedades relativas a la convexidad de la M-divergencia en términos de la L-divergencia, estudiando como caso particular el cuadrado de la distancia Matusita.

También se relaciona la M-divergencia con la J y la K-divergencia, atendiendo a las condiciones que deben verificar las funciones que definen las divergencias.

Keywords: DIVERGENCE MEASURES, CONVEXITY, DATA ANALYSIS.

1. INTRODUCCION.

Uno de los problemas más interesantes que se plantean en análisis de datos es el de la elección de medidas adecuadas que permitan cuantificar las analogías y diferencias entre poblaciones o entre individuos de una misma población. En este sentido las divergencias pueden entenderse como medidas de las diferencias entre distribuciones o como disimilaridades entre las poblaciones.

Las divergencias pueden basarse en consideraciones empíricas tales como las que presentan Rao /19/, Bhattacharyya /2/, Jeffreys /11/, Wald /24/, Kullback & Leibler /12/, Matusita /16/, en las diferencias intrínsecas de los individuos, pudiendo citarse entre otros los trabajos de Simpson /21/, Nei /18/, Agresti /1/, y por último, pueden deducirse de consideraciones relacionadas con las funciones de entropía, siendo interesantes entre otros los trabajos de Shannon /23/, Renyi /20/, Havrda & Charvat /10/, Mathay & Rathie /13/, Burbea & Rao /4/.

Es evidente que cada uno de estos métodos presentan ventajas e inconvenientes respecto de los demás. Así, los índices que se deducen de los dos primeros métodos suelen ajustarse bastante bien a las condiciones experimentales pero al carecer por lo general de buenas propiedades matemáticas resulta difícil la representación geométrica de los individuos a través de la disimilaridad obtenida. Por el contrario, los índices que provienen de las funciones de entropía tienen atractivas propiedades matemáticas pero serias dificultades de interpretación. Sin embargo, estas últimas medidas de divergencia definidas en términos de funciones de entropía admiten, por lo general, estudios relativos a su convesidad que resultan muy adecuados para abordar aspectos relacionados con la teoría de la información e inferencia estadística tal como podemos ver en Matusita /14/, /15/, /17/ y Shaked /22/. Burbea & Rao /3/, /4/, /5/, para la clasificación de las entropías de grado α.

Las medidas de divergencia han sido utilizadas en una gran variedad de estudios relativos a distintos campos: biología, economía, teoría de la comunicación, cibernética, ... como podemos ver en Bhattacharyya /2/, Csiszar /7/, Havrda & Charvat /10/.

En este trabajo vamos a dedicarnos al estudio particular de la M-divergencia y sus

- M.Salicrú Pagés i A. Arcas Pons - Universitat de Barcelona- Dep. de Bioestadística de la Facultat de Biologia - Avda. Diagonal, 645 - 08028 Barcelona.

- Article rebut el març de 1986.

relasciones con las J,K y L divergencias.

2. DEFINICIONES PREVIAS.

Dada una función ϕ contínua definida en un intervalo I de R, se define la ϕ entropía $H_{n,\phi}(x)$ como

$$H_{n,\phi}(x) = -\sum_{i=1}^{n} \phi(x_i) \tag{1}$$

siendo $x = (x_1,\ldots,x_n)$.

La J-divergencia entre $x=(x_1,\ldots,x_n)$ e $y=(y_1,\ldots,y_n)$ se define como

$$J_{n,\phi}(x,y) = \sum_{i=1}^{n} 1/2.(\phi(x_i)+\phi(y_i)) - \phi(\frac{x_i+y_i}{2}) \tag{2}$$

La K-divergencia se define como

$$K_{n,\phi}(x,y) = \sum_{i=1}^{n} (x_i-y_i).(\frac{\phi(x_i)}{x_i} - \frac{\phi(y_i)}{y_i}) \tag{3}$$

la L-divergencia se define como

$$L_{n,\phi}(x,y) = \sum_{i=1}^{n} (x_i.\phi(\frac{x_i}{x_i})+y_i.\phi(\frac{y_i}{y_i})) \tag{4}$$

y la M-divergencia como

$$M_{n,\phi}(x,y) = (\sum_{i=1}^{n} (\sqrt{\phi(x_i)} - \sqrt{\phi(y_i)})^2)^{1/2} \tag{5}$$

3. PROPIEDADES PREVIAS.

PROPIEDAD 3.1

$M_{n,\phi}(x,y)$ cumple las condiciones de distancia en IxI si y sólo si ϕ es una función inyectiva de I en R^+.

La demostración es inmediata.

PROPIEDAD 3.2

a) Para $\phi(x)=x^\alpha$, $M_{n,\phi}(x,y)$ resulta la α-distancia Hellinger multiplicada por $1/2.|\alpha|$.

b) Para $\phi(x)=x$, $M_{n,\phi}(x,y)$ coincide con la distancia Matusita dividida por 2.

4. RELACION ENTRE LAS DIVERGENCIAS

En esta sección nos proponemos relacionar la M-divergencia con la L, J y K divergencias.

Proposición 4.1

Se verifica que

$$M^2_{n,\phi}(x,y) = L_{n,\psi}((\phi(x_1),\ldots,\phi(x_n)),(\phi(y_1),\ldots,\phi(y_n)))$$

siendo $\psi(t) = 1 - \sqrt{t}$.

Demostración:

$$M^2_{n,\phi}(x,y) = \sum_{i=1}^{n} (\sqrt{\phi(x_i)} - \sqrt{\phi(y_i)})^2 =$$

$$= \sum_{i=1}^{n}\left[\phi(x_i).\left[1-\sqrt{\frac{\phi(y_i)}{\phi(x_i)}}\right]+\phi(y_i).\left[1-\sqrt{\frac{\phi(x_i)}{\phi(y_i)}}\right]\right] =$$

$$= L_{n,\psi}((\phi(x_1),\ldots,\phi(x_n)),(\phi(y_1),\ldots,\phi(y_n)))$$

A fin de hallar la relación con la J-divergencia y la K-divergencia, veamos previamente el siguiente lema:

Lema 4.1

$\phi(x)$ es log-convexa de I sobre R^+ si y sólo si $(\phi(x).\phi(y))^{1/2} \geq \phi(\frac{x+y}{2})$ para todo x,y de I.

Demostración:

La desigualdad $(\phi(x).\phi(y))^{1/2} \geq \phi(\frac{x+y}{2})$ equivale a la desigualdad

$$\log(\phi(x))+\log(\phi(y)) \geq 2.\log(\phi(\frac{x+y}{2})),$$

y ésta a su vez a

$$\frac{1}{2}.(\log(\phi(x))+\log(\phi(y))-\log(\phi(\frac{x+y}{2}))\geq 0$$

lo que equivale a decir que $\phi(x)$ es log-convexa.

Proposición 4.2

$\phi(x)$ es log-convexa en I si y sólo si $M^2_{n,\phi}(x,y) \leq 2.J_{n,\phi}(x,y)$ en $I^n x I^n$.

Demostración:

El cuadrado de la M-divergencia $M_{n,\phi}(x,y)$ puede escribirse como

$$M^2_{n,\phi}(x,y)=2.\sum_{i=1}^{n}\left(\frac{1}{2}.(\phi(x_i)+\phi(y_i))-(\phi(x_i).\phi(y_i))^{\frac{1}{2}}\right)$$

Por el lema anterior, de $\phi(x)$ log-convexa se deduce que

$$M^2_{n,\phi}(x,y)\leq 2.\sum_{i=1}^{n}\left(\frac{1}{2}.(\phi(x_i)+\phi(y_i))-\phi\left(\frac{x_i+y_i}{2}\right)\right)$$

Así, cuando $\phi(x)$ es log-convexa,

$M_{n,\phi}(x,y)\leq 2.J_{n,\phi}(x,y)$.

Recíprocamente, si $M^2_{n,\phi}(x,y)\leq 2.J_{n,\phi}(x,y)$ en $I^n x I^n$, para todo z,t de I, tomando $x=(z,..z)$ $y=(t,..,t)$ se tendrá la desigualdad

$$\frac{1}{2}.M^2_{n,\phi}(x,y)=\sum_{i=1}^{n}\left(\frac{1}{2}.(\phi(z)+\phi(t))-(\phi(z).\phi(t))^{\frac{1}{2}}\right)\leq$$

$$\leq\sum_{i=1}^{n}\left(\frac{1}{2}.(\phi(z)+\phi(t))-\phi\left(\frac{z+t}{2}\right)\right)=J_{n,\phi}(x,y)$$

de la que se deduce

$$-n.(\phi(z).\phi(t))^{\frac{1}{2}}\leq -n.\phi\left(\frac{z+t}{2}\right)$$

para todo z,t de I, resultando a partir del lema 1 que $\phi(x)$ es log-convexa en I.

Corolario 4.1

Si $\psi(x)=\phi(x)/x$ es cóncava en I y $\phi(x)$ es log-convexa en I, entonces

$$K_{n,\phi}(x,y)\geq 4.J_{n,\phi}(x,y)\geq 2.M^2_{n,\phi}(x,y)\geq 0$$

Demostración:

Por ser $\phi(x)$ positiva y log-convexa en I, $\phi''(x)$ debe ser positiva en I, resultando de /4/, de $\phi(x)$ convexa en I y de $\psi(x)$ cóncava en I que

$$K_{n,\phi}(x,y)\geq 4.J_{n,\phi}(x,y)$$

y de considerar la proposición anterior, el corolario es inmediato.

5. SOBRE LA CONVEXIDAD DE LA M-DIVERGENCIA.

En esta última parte vamos a considerar ciertas propiedades relativas a la convexidad de la M-divergencia, convexidad que tendrá su interés en aspectos de inferencia estadística relacionados con el cálculo de la estimación máximo verosímil.

Proposición 5.1

$M^2_{n,\phi}(x,y)$ es convexa en $I^n x\ I^n$ si y sólo si $M^2_{1,\phi}(x,y)$ es convexa en IxI.

Demostración:

Es evidente comprobar que $M^2_{n,\phi}(x,y)=$

$$=\sum_{i=1}^{n} M^2_{1,\phi}(x_i,y_i)$$

Así, si $M^2_{1,\phi}(x,y)$ es convexa en IxI, $M^2_{n,}(x,y)$ será convexa en $I^n xI^n$ por ser la suma de convexas.

Recíprocamente, para todo z,t de I podemos - tomar $x=(z,...,z)$, $y=(t,...,t)$, resultando

$$M^2_{n,\phi}(x,y) = n.M_{1,\phi}(z,t)$$

y de la convexidad de $M^2_{n,\phi}(x,y)$ en $I^n xI^n$ se deduce la convexidad de $M^2_{1,\phi}(z,t)$ en IxI.

Proposición 5.2

$M^2_{n,\phi}(x,y)$ es convexa en términos de $\Phi(x)=(\phi(x_1),\ldots,\phi(x_n))$, es decir, $M^2_{n,\phi}(\lambda(\Phi(x),\Phi(y))+(1-\lambda)(\Phi(z),\Phi(t)))\leq\lambda M^2_{n,\phi}(\Phi(x),\Phi(y))+(1-\lambda)M^2_{n,\phi}(\Phi(z),\Phi(t))$

Demostración : **De la propiedad 4.1**

$$M^2_{n,\phi}(x,y)=L_{n,\psi}((\phi(x_1),\ldots,\phi(x_n)),(\phi(y_1),\ldots,\phi(y_n)))=$$

$$=\sum_{i=1}^{n}\phi(y_i).f\left(\frac{\phi(x_i)}{\phi(y_i)}\right)$$

con $f(t) = (1-\sqrt{t})^2 = \psi^2(t)$

A partir de /4/ y de f(t) convexa en R^+ se deduce que $L_{n,\psi}(\phi(x),\phi(y))$ es convexa en $R^n_+ \times R^n_+$.

Corolario 5.1

El cuadrado de la distancia Matusita es convexo en $R^n_+ \times R^n_+$

Demostración:

Ya que la distancia Matusita al cuadrado se obtiene con $\phi(x)=x$, $M^2_{n,\phi}(x,y)$ es convexa en x,y.

6. CONCLUSIONES.

En este trabajo se ha estudiado la relación entre la M-divergencia y la J-divergencia, resultando la relación con la K-divergencia una consecuencia inmediata del estudio de Burbea & Rao /4/.

Por otro lado se han estudiado ciertos aspectos de la convexidad de la M-divergencia, -- propiedad de gran interés en estudios de teoría de la información e inferencia estadística. En este sentido, la M-divergencia se ha obtenido como una L-divergencia entre elementos a los que se les ha aplicado una transformación.

7. BIBLIOGRAFIA.

/1/ AGRESTI, A. y AGRESTI, B.F.: "Statistical analysis of qualitative variation". Social Methodology, 204-237 (1978).

/2/ BHATTACHARYA, A.: "A measure of divergence between two multinomial populations" Sankhya 7, 401 (1946).

/3/ BURBEA, J. y RAO, C.R.: "Entropy differential metric, distance and divergence measures in probability spaces: a unified approach". Jornal of multivariate analysis, V.12, nº 4. (1982).

/4/ BURBEA, J. y RAO, C.R.: "On the convexity of some divergence measures based on entropy functions". IEE transactions on information theory" V. 28, 3, 489-495 (1982) (1982).

/5/ BURBEA, J. & RAO, C.R.: "Differential metrics in probability spaces". Prob. Math.Stat., 3, 241-258 (1984)

/6/ BURBEA, J.: "Informative geometry of probability spaces. Technical Report, nº 84-52, Univers. Pittsburgh. (1984).

/7/ CSISZAR, I.: "A class of measures of informativity of observation channels". Perioriodica Math. Hungarica, V.2, 191-213. (1972).

/8/ DAROCZY, Z.: "Generalized information functions". Inform. and Contr., V. 16, 83-88 (1970).

/9/ GOOD, I.J.: "Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables". Anals Math. Statistics, 911-933. (1963).

/10/ HAVRDA, M.E. y CHARVAT, F.: "Quantification method of classification processes: concept of structural α-entropy.Kybernetica, V.3, 30-35. (1967).

/11/ JEFFREYS, H.: "Theory of probability." Oxford Univ. Press. London (1948).

/12/ KULLBACK, S. y LEIBLER, R.A.: "On information and sufficiency". Ann. Mathe Statist., V.22, 79-86 (1951).

/13/ MATTAI, A. y RATHIE, P.N.: "Basic concepts in information theory and statistics". John Wiley, New York.(1974).

/14/ MATUSITA, K.: "Decision rules, based on the distance, for problems of fit, two samples, and estimation". Anals Math. Statistical, V.26, 631-640 (1955).

/15/ MATUSITA, K.: "Decision rule based on the distance for the classification problem". Ann. Inst. Statis. Math., V.8, 66-77. (1957).

/16/ MATUSITA, K.: "Distance and decision rules". Annals of the Ins. Statis. Math., V. 14, 305-315. (1964).

/17/ MATUSITA, K.:"Classification based on distance in multivariate Gaussian cases". Proceed. of the fifth Berkeley Symposium on Mathematical Statistics and probability. (1966).

/18/ NEI, M.: "The theory of genetic distance and evolution of human races". Japan, J.Human Genet. V. 23, 341-369. (1978).

/19/ RAO, C.R.: "The utilization of multiple measurements in problems of biological classification". J.Roy Statist. Soc. V.10, 159-193 (1948).

/20/ RENYI, A.: "On measures of entropy and information". Proceedings fourth Berkeley Symp., V.1, 547-561 (1961).

/21/ SIMPSON, E.H.: "Measurement of diversity". Nature 163-688, (1949).

/22/ SHAKED, M.: "Statistical inference for a class of life distributions". Com. Stat. Theo. Math., A6(13), 1323-1339. (1977).

/23/ SHANNON, C.E.: "A mathematical theory of communications". Bell System. Tech. V. 27, 379-423, 623-656. (1948).

/24/ WALD, A.: "Statistical Decision Functions". John Wiley, New York. (1950).