TRABAJOS DE ESTADISTICA
Vol. 6, núm. 1, 1991, pp. 27 a 34

METRICAS RIEMANIANAS ASOCIADAS A M-DIVERGENCIAS

MIQUEL SALICRÚ, MERCÈ ARGEMÍ
Dpto. Estadística
Universidad de Barcelona
Avda. Diagonal, 645
08028 BARCELONA

RESUMEN

Tras plantear las métricas diferenciales asociadas a M-divergencias para funciones de densidad de probabilidad pertenecientes a la misma familia de funciones paramétricas, consideramos para las funciones $\Phi_\alpha(t) = t^\alpha$ la relación entre las matrices que definen las métricas y las matrices α-informativas.

Obtenemos, en segundo lugar, las funciones $\Phi(t)$ que determinan M-divergencias invariantes, a nivel diferencial, frente a cambios no singulares de parámetros y variables aleatorias.

Observamos, finalmente, que las funciones que proporcionan invariancia aseguran un incremento en el valor de la distancia al añadir variables aleatorias estocásticamente independientes.

Palabras clave: M-divergencia, métrica diferencial, invarianza.
Clasificación AMS: 62-07, 62P10.

SUMMARY

After introducing the differential metric associated with M-divergences for probability density functions, which belong to the same family of parametric functions, we consider for the functions $\Phi_\alpha(t) = t^\alpha$ the relation between the matrices which define the metric and the α-information matrices.

Then we obtain the functions $\Phi(t)$ which determine M-divergences invariant with respect to non-singular changes of parameters and random variables.

Finally, we observe that the functions which give invariance ensure an

Recibido: Enero 1990.
Revisado: Marzo 1991.

increase of the value of the distance when we add independent random variables.

Key words: M-divergences, differential metric, invariance.
AMS Classification: 62-07, 62P10.
Title: RIEMANIAN'S METRIC ASSOCIATED WITH M-DIVERGENCES.

1. INTRODUCCION

Al analizar la distancia entre objetos a partir de funciones de densidad de probabilidad, distintos autores han planteado la necesidad de invariancia para la distancia, al variar la forma de tomar las medidas de los objetos. En este sentido Kullback (1959) plantea la invarianza para las métricas diferenciales asociadas a la medida de Kullback-Leibler, a la medida de Jeffreys y a la matriz de información de Fisher, Amari (1985) plantea la invarianza en el estudio de las α-conexiones, y Cuadras y otros (1985) plantean la invarianza para la esperanza del cambio infinitesimal de información.

En este trabajo, planteamos las métricas diferenciales asociadas a M-divergencias para funciones de densidad de probabilidad pertenecientes a la misma familia de funciones paramétricas, relacionamos las matrices que definen las métricas con matrices informativas, y finalmente, determinamos las divergencias que a nivel diferencial resultan invariantes frente a cambios no singulares de parámetros y variables aleatorias.

2. DEFINICIONES PREVIAS

Para una medida positiva μ, σ-aditiva, y σ-finita, definida en una σ-álgebra de subconjuntos de un espacio medible X, para una función real positiva $\Phi(t)$ dos veces diferenciable con continuidad y definida en un intervalo T_Φ, con $[0, 1] \in T_\Phi \subset [0, \infty)$, y para cada subconjunto Ω de $\mathscr{R}^n$.

$$\Omega = \{\Theta; \Theta = (\theta_1, ..., \theta_n) \in \mathscr{R}^n\}$$

se considera la familia paramétrica de funciones de densidad de probabilidad $\mathscr{R}_\Omega$ definida por:

$$\mathscr{R}_\Omega = \left\{ p(x,\Theta);\, p: X \to \mathscr{R} \ tq \ \begin{array}{l} 1. \ \displaystyle\int_X p(x,\Theta)\, d\mu(x) = 1 \\ 2. \ p(x,\Theta) \in T_\Phi \ \forall\, x \in X \\ 3. \ \exists \ \dfrac{\partial^2 p(x,\Theta)}{\partial\theta_i \partial\theta_j} \ \forall\, i, j \in \{1, \dots, n\} \end{array} \right\}$$

y se toma como medidas de distanciación entre poblaciones las M-divergencias definidas por

$$M_\Phi(p,q) = \left[\int_X [\sqrt{\Phi(p)} - \sqrt{\Phi(q)}]^2\, d\mu \right]^{1/2}$$

3. METRICAS DIFERENCIALES ASOCIADAS A LAS M-DIVERGENCIAS

Distintos autores han planteado la distancia entre individuos, poblaciones o grupos de poblaciones, a partir de métricas definidas a través del elemento diferencial de arco. En este sentido, para familias de distribuciones dependientes de parámetros, Bhattacharya (1946), Kullback (1959), Kagan (1963), Vajda (1973), Boekee (1979) y Ferentines y Papaioannou (1981) consideran, para medidas de disimilaridad D, las métricas definidas a partir de

$$ds^2 = \lim_{t \to \infty} \left[\frac{D(p, p + t\, dp) - D(p,p)}{t} \right]^2 \tag{1}$$

En este trabajo, hemos desarrollado la métrica diferencial correspondiente a las M-divergencias, obteniendo, al desarrollar el numerador por Taylor en el punto 0, que el elemento diferencial de arco se expresa en la forma

$$ds_\Phi^2 = \int_X [(\sqrt{\Phi(p)})']^2\, [dp]^2\, d\mu \tag{2}$$

Teniendo en cuenta que

$$dp = \sum_{i=1}^{n} \frac{\partial p}{\partial \theta_i} d\theta_i \tag{3}$$

la expresión (2) se reduce a

$$ds_\Phi^2 = \sum_{i,j=1}^{n} \left[\int_X [(\sqrt{\Phi(p)})']^2 \frac{\partial p}{\partial \theta_i} \frac{\partial p}{\partial \theta_j} du \right] d\theta_i \, d\theta_j \tag{4}$$

es decir,

$$ds_\Phi^2 = \sum_{i,j=1}^{n} g_{ij}^\Phi \, d\theta_i \, d\theta_j \quad \text{con} \quad g_{ij}^\Phi = \int_X [(\sqrt{\Phi(p)})']^2 \frac{\partial p}{\partial \theta_i} \frac{\partial p}{\partial \theta_j} d\mu \tag{5}$$

Nota 1:

Obsérvese que la matriz (g_{ij}^Φ) define un tensor covariante de segundo orden, ya que, al efectuar un cambio no singular de parámetros $(\theta_1, \theta_2, ..., \theta_n) \to (\bar{\theta}_1, \bar{\theta}_2, ..., \bar{\theta}_n)$, el elemento g_{ij}^Φ se transforma en $\bar{g}_{ij}^\Phi$

$$\bar{g}_{ij}^\Phi = \int_X [(\sqrt{\Phi(\bar{p})})']^2 \frac{\partial \bar{p}}{\partial \bar{\theta}_i} \frac{\partial \bar{p}}{\partial \bar{\theta}_j} d\mu = \sum_{r,s=1}^{n} g_{rs}^\Phi \frac{\partial \theta_r}{\partial \bar{\theta}_i} \frac{\partial \theta_s}{\partial \bar{\theta}_j}$$

de forma que, la matriz (g_{ij}^Φ) y el elemento ds_Φ^2 obtenidos en (5) definen una métrica Riemaniana.

Nota 2:

La matriz (g_{ij}^Φ) que define la métrica y el elemento diferencial de arco ds_Φ^2 pueden expresarse en la forma

$$g_{ij}^\Phi = E\left([(\sqrt{\Phi(p)})']^2 p \frac{\partial \log p}{\partial \theta_i} \frac{\partial \log p}{\partial \theta_j} \right)$$

$$ds_\Phi^2 = E([(\sqrt{\Phi(p)})']^2 p [d \log p]^2)$$

Casos particulares

Para la familia de funciones $\Phi_\alpha(t) = t^\alpha$ definidas en $\mathscr{R}^+$, se tiene:

$$g_{ij}^\alpha = \frac{\alpha^2}{4} E(p^{\alpha-1} \frac{\partial \log p}{\partial \theta_i} \frac{\partial \log p}{\partial \theta_j})$$

$$ds_\alpha^2 = \frac{\alpha^2}{4} E(p^{\alpha-1}[d \log p]^2)$$

Así, (g_{ij}^α) resulta la matriz de orden α módulo $\alpha^2/4$ y ds_α^2 la distancia geodésica de orden α módulo $\alpha^2/4$.

Obsérvese que para $\alpha = 1$, la matriz (g_{ij}^1) se reduce a la matriz de información de Fisher módulo $1/4$ y ds_1^2 al elemento de línea correspondiente a la distancia de Rao, salvo la constante $1/4$.

4. CONDICIONES GENERALES PARA UNA DISTANCIA

En análisis de datos, el concepto de distancia se utiliza para expresar de forma cuantitativa las semejanzas y diferencias entre objetos, siendo de especial interés las caracterizaciones de los objetos a partir de términos matemáticos susceptibles de ser tratados. En este sentido, con frecuencia se determinan los objetos a partir de un conjunto de medidas que pueden ser consideradas como una muestra aleatoria de una población estadística, de forma que la caracterización puede realizarse a través de la función de densidad.

Exigiremos que las distancias consideradas en las familias de funciones paramétricas verifiquen las siguientes condiciones:

1. La distancia entre los objetos debe depender de las funciones de densidad.
2. La distancia entre dos funciones de densidad debe ser independiente de la parametrización utilizada, es decir, debe ser invariable frente a transformaciones no singulares de los parámetros, entendiendo como tales las transformaciones biyectivas definidas a través de funciones diferenciales con continuidad. Ello es debido a que los parámetros juegan en nuestro caso el papel de las coordenadas en la variedad de las funciones de densidad de probabilidad.
3. La distancia entre dos funciones de densidad de probabilidad debe ser invariante frente a transformaciones no singulares de las variables aleatorias ya que dichos cambios no afectan a los objetos a distanciar, sino a la forma de tomar las medidas.
4. La distancia entre los objetos a comparar debe aumentar al agregar variables aleatorias estocásticamente independientes, ya que aumenta la información sobre sus diferencias.

5. INVARIANZA FRENTE A TRANSFORMACIONES NO SINGULARES DE VARIABLES ALEATORIAS

Hemos determinado las funciones $\Phi(t)$ que a nivel diferencial, para la medida de Lebesgue, nos proporcionan M-divergencias invariantes frente a transformaciones no singulares de variables aleatorias. En este sentido, cuando transformamos

$$p = p(x_1, ..., x_k, \theta_1, ..., \theta_n) \text{ en } \bar{p} = \bar{p}(y_1, ..., y_k, \theta_1, ..., \theta_n),$$

el tensor métrico se transforma siguiendo la relación

$$g_{ij}^{\Phi} = \int_X [(\sqrt{\Phi(p)})']^2 \frac{\partial p}{\partial \theta_i} \frac{\partial p}{\partial \theta_j} dx \to \bar{g}_{ij}^{\Phi} = \int_X [(\sqrt{\Phi(\bar{p})})']^2 \frac{\partial \bar{p}}{\partial \theta_i} \frac{\partial \bar{p}}{\partial \theta_j} dy =$$

$$= \int_X [(\sqrt{\Phi(pJ)})']^2 \frac{\partial pJ}{\partial \theta_i} \frac{\partial pJ}{\partial \theta_i} \frac{dx}{J} = \int_X [(\sqrt{\Phi(pJ)})']^2 \frac{\partial p}{\partial \theta_i} \frac{\partial p}{\partial \theta_i} J\, dx$$

siendo J el valor absoluto del jacobiano del cambio.

Particularizando a la familia de funciones de densidad de probabilidad

$$p(x_1, ..., x_k, \theta_1, ..., \theta_n) = \theta_1 + c$$

donde c es una constante, la transformación anterior se reduce a

$$g_{ij}^{\Phi} = \begin{cases} \int_X [(\sqrt{\Phi(\theta_1 + c)})']^2\, dx & \text{para } i = j = 1 \\ 0 & \text{para } i \neq j \text{ o } i = j \neq 1 \end{cases} \to$$

$$\to \bar{g}_{ij}^{\Phi} = \begin{cases} \int_X [(\sqrt{\Phi((\theta_1 + c)J)})']^2\, J\, dx & \text{para } i = j = 1 \\ 0 & \text{para } i \neq j \text{ o } i = j \neq 1 \end{cases}$$

Así, por la arbitrariedad del jacobiano del cambio y de la constante $\theta_1 + c$, la invariancia frente a transformaciones no singulares de variables aleatorias en la métrica diferencial asociada a M-divergencias, se verifica cuando

$$[(\sqrt{\Phi(x)})']^2 = [(\sqrt{\Phi(x \cdot y)})']^2 y \quad \text{con } x, y > 0$$

condición que se cumple únicamente para funciones de la forma

$$\Phi(t) = [a\sqrt{t} + b]^2 \quad \text{con } a \text{ y } b \text{ constantes} \tag{6}$$

Es decir, la invarianza se obtiene para las funciones que a nivel de divergencias definen la distancia o múltiplos de la distancia Matusita.

Nota 3

Al tratarse de una métrica Riemaniana, los cambios no singulares de parámetros no alteran el valor de la distancia.

Nota 4

Para X, Y vectores aleatorios estocásticamente independientes y para las funciones $\Phi(t)$ definidas en (6), se verifican las siguientes igualdades

$$\begin{aligned}
g_{ij}^{X,Y} &= E\left\{\frac{a^2}{4} p^{-2}(x,y) \frac{\partial p(x,y)}{\partial \theta_i} \frac{\partial p(x,y)}{\partial \theta_j}\right\} = \\
&= E\left\{\frac{a^2}{4} p^{-2}(x) p^{-2}(y) \frac{\partial p(x)p(y)}{\partial \theta_i} \frac{\partial p(x)p(y)}{\partial \theta_j}\right\} = \\
&= E\left\{\frac{a^2}{4} p^{-2}(x) \frac{\partial p(x)}{\partial \theta_i} \frac{\partial p(x)}{\partial \theta_j}\right\} + E\left\{\frac{a^2}{4} p^{-2}(y) \frac{\partial p(x)}{\partial \theta_i} \frac{\partial p(y)}{\partial \theta_j}\right\} + \\
&+ E\left\{\frac{a^2}{4} p^{-1}(x) p^{-1}(y) \frac{\partial p(x)}{\partial \theta_i} \frac{\partial p(y)}{\partial \theta_j}\right\} + \\
&+ E\left\{\frac{a^2}{4} p^{-1}(x) p^{-1}(y) \frac{\partial p(x)}{\partial \theta_j} \frac{\partial p(y)}{\partial \theta_i}\right\} = g_{ij}^X + g_{ij}^Y
\end{aligned}$$

ya que

$$\begin{aligned}
E\left\{\frac{a^2}{4} p^{-1}(x) p^{-1}(y) \frac{\partial p(x)}{\partial \theta_r} \frac{\partial p(y)}{\partial \theta_s}\right\} &= \int\int \frac{a^2}{4} \frac{\partial p(x)}{\partial \theta_r} \frac{\partial p(y)}{\partial \theta_s}\, dy\, dx = \\
&= \frac{a^2}{4} \int \frac{\partial p(x)}{\partial \theta_r}\, dx \int \frac{\partial p(y)}{\partial \theta_s}\, dy = 0 \quad \text{por} \quad \int p(x)\, dx = 1
\end{aligned}$$

Así, la forma cuadrática diferencial que define el elemento de línea es suma de las formas cuadráticas correspondientes a las variables X e Y separadamente. Es decir

$$ds^2_{(X,\,Y)} = ds^2_X + ds^2_Y \tag{7}$$

con lo que, la descomposición (7) asegura el incremento en el valor de la distancia al añadir variables aleatorias independientes.

Nota 5

Hemos estudiado el caso absolutamente continuo, ya que en el caso discreto todas las funciones $\Phi(t)$ proporcionana invariancia.

BIBLIOGRAFIA

- AMARI, S. (1985): *Lecture Notes in Statistics*, Springer-Verlag, Nueva York.
- BHATTACHARYA, A. (1946): «A measure of divergence between two multinomial populations», Sankhya, 7, 401-406.
- BOEKEE, D. E. (1979): *The D information of order s*. Transactions of the 8th. Prague Conference on Information Theory, Statistical Decision Functions and Random Processes 1978, p. 55-66.
- BURBEA, J., RAO, C. R. (1985): «Differential metrics in probability spaces», *Probability and Mathematical Statistics 3*, 2, p. 241-258.
- CUADRAS, C. M.; OLLER, J. M.; ARCAS, A., y RIOS, M. (1985): «Métodos geométricos de la Estadística», *Qüestió 9*, 4, 219-250.
- FERENTINOS, K., y PAPAIVANNOU, T. (1981): «New parametric measures of information», *Information and Control*, 51, p. 193-208.
- KULLBACK, S. (1959): *Information theory and Statistics*, Jhon W iley & Sons, Nueva York.
- RAO, C. R. (1945): «Information and accurateness in the estimation of parameters», *Bull. Calcuta Mathematics Society*, 37, p. 81-91.
- SALICRU, M., y ARCAS, A. (1985): «Sobre ciertas propiedades de la M-divergencia en análisis de datos», *Qüestió 9*, 4, p. 251-255.