QÜESTIIÓ, Vol 11, n° 3, pp. 3-12, 1987

FUNCIONES DE ENTROPÍA ASOCIADAS A MEDIDAS DE CSISZAR

MIQUEL SALICRÚ y CARLES M. CUADRAS
Universitat de Barcelona

En este trabajo se presentan las medidas de entropía que provienen de la distancia, en el sentido de Csiszar, entre una distribución y la distribución en la que todos los sucesos son equiprobables. En segundo lugar, se estudian condiciones para la concavidad y no negatividad de las medidas propuestas. Finalmente, se obtienen los funcionales ϕ-entropía como casos particulares de las medidas estudiadas.

Entropy functions based on Csiszar measures

Keywords: Entropy functions, Csiszar measures.

1. INTRODUCCIÓN

En teoria de la información suelen presentarse las medidas de entropía a partir de funciones, con propiedades matemáticas adecuadas, que permite reflejar la incertidumbre asociada a un experimento. En este sentido, desde perspectivas distintas, la información asociada a un experimento ha sido estudiada por Shannon [13], Havrda–Charvat [16], Renyi [10], y Boekee y Lubbe [1] entre otros.

Una alternativa y a su vez una generalización de los modelos ϕ–entropía podemos obtenerla a partir de las medidas que provienen de la distancia entre una distribución y la distribución en la que todos los sucesos son equiprobables (distancia en el sentido de Csiszar [4]).

En este trabajo presentamos las medidas asociadas a divergencias de Csiszar, estudiamos condiciones para la concavidad, homogeneidad y no negatividad, analizamos la condición de Dalton–Pielou, y finalmente a título ilustrativo presentamos algunas funciones ϕ–entropía como casos particulares de las medidas planteadas.

–Miquel Salicrú i Carles M. Cuadras - Universitat de Barcelona, Dept. d'Estadística - Av. Diagonal, 637 - 08028 Barcelona.
–Article rebut el abril de 1987.

El análisis de las propiedades anteriores se encuadra en el marco general correspondiente al estudio de las medidas de divergencia y entropía en análisis de datos, tal y como se presenta en Rao [9], Burbea y Rao [3], Bourguignon [2] y en Ferentinos y Papaivannou [5].

2. DEFINICIONES PREVIAS

Para la realización de este trabajo, relacionaremos en primer lugar las definiciones básicas que utilizaremos en este estudio. En este sentido, para una función $f : \mathbf{R} \longrightarrow R$ dos veces diferenciable con continuidad y para dos vectores de $R^n_+, x = (x_1, x_2, ...x_n)$ e $y = (y_1, y_2, ...y_n)$, Csiszar [4] define las siguientes medidas.

Definición 2.1.

Una aplicación $I_f : R^n_+ \pm R^n_+ \longrightarrow R$ se dice que es una medida de Csiszar si viene definida a través de la expresión

$$I_f(x, y) = \sum_{i=1}^{n} x_i f\left(y_i x_i^{-1}\right)$$

Asociadas a estas medidas consideramos, para cada constante k y para cada vector de probabilidades p, las funciones de la forma

$$H^*_{f,k,p}(x) = I_f\left((x_1, x_2, ..., x_n), (p_1, p_2, ...p_n)\right) + k$$

En particular, designaremos por $H^*_{f,k}(x)$ la medida correspondiente al vector $p = (n^{-1}, ..., n^{-1})$.

Cuando $H^*_{f,k,p}(x)$ defina una función no negativa y cóncava en x, diremos entonces que es una medida de entropía, y cuando pueda extenderse por continuadad a $R^n_+ \cup \{0\}$, es decir, cuando $\lim_{t \longrightarrow \infty} t^{-1} f(t) = C$, diremos entonces que $H^*_{f,k,p}(x)$ es una medida regular.

En particular, para cada vector de probabilidades $x, x = (x_1, x_2, ..., x_n)$ con $x_1 \geq 0, x_2 \geq 0, ..., x_n \geq 0$ y $x_1, x_2 + ... + x_n = 1$, la medida $H^*_{f,k}(x)$ representa, salvo constantes, la distancia en el sentido de Csiszar a la distribución en la que todos los sucesos son equiprobables.

Definición 2.2.

Si $\phi : R \longrightarrow R$ es una función dos veces diferenciable con continuidad, se llama ϕ–entropía a la función $H_\phi : R^n \longrightarrow R$ definida a través de la expresión

$$H_\phi(x_1, x_2, ..., x_n) = \sum_{i=1}^{n} \phi(x_i)$$

que sea cóncava y no negativa en x.

Definición 2.3.

Se dice que una función $g : R^n \longrightarrow R$ cumple la condición de Dalton–Pielou si se verifica

$$g(x_1, x_i, ..., x_j, ..., x_n) \leq g(x_1, ..., x_i + \delta, ..., x_j - \delta, ..., x_n)$$

para $x_i < x_i + \delta \leq x_j - \delta < x_j$.

Nótese que esta propiedad, en el caso de vectores de probabilidad, puede expresarse con el siguiente enunciado: "Si se produce una transferencia de probabilidad de un suceso más probable a otro menos probable consevándose el orden, entonces crece el valor de la función".

3. CONCAVIDAD Y CONDICIÓN DE DALTON–PIELOU PARA FUNCIONES $H^*_{f,k,p}(x)$

Al analizar la concavidad de las funciones $H^*_{f,k,p}(x)$ hemos obtenido la siguiente caracterización para cualquier constante k.

Proposición 3.1.

$H^*_{f,k,p}(x)$ es cóncava en R^n_+ si y solo si $f(t)$ es cóncava en R_+

Demostración

La matriz Hessiana $\left(\frac{\partial^2 H^*_{f,k,p}(x)}{\partial x_i \partial x_j}\right)$, que en nuestro caso se reduce a la matriz de elementos.

$$h_{ij}(x_1, x_2, ..., x_n) = \begin{cases} p_i^{-2} x_i^{-3} f''(p_i^{-1} x_i^{-1}) & \text{para } i = j \\ 0 & \text{para } i \neq j \end{cases}$$

es semidefinida negativa en R_+^n si y solo se cumple la condición

$$p_i \cdot t^3 \cdot f''(t) \leq 0 \qquad \forall\, t \in R_+$$

es decir, si $f(t)$ es cóncava en R_+

De considerar restringido el dominio de las medidas $H^*_{f,k,p}(x)$ al conjunto de vectores de probabilidad, se obtiene de forma inmediata el siguiente resultado

Proposición 3.2.

Las medidas $H^*_{f,k,p}(x)$ que cumplen la condición de Dalton–Pielou alcanzan el máximo valor para el vector de propiedades $(n^{-1}, ..., n^{-1})$.

Para vectores de probabilidad $p = (p_1, ...p_n)\,, q = (q_1, ..., q_n)$ y para las medidas $H^*_{f,k,p}(q)$ con $f(t)$ cóncava en R_+, se cumplen

Proposición 3.3.

$H^*_{f,k,p}(q)$ alcanza máximo valor para la distribución p.

Demostración

$$H^*_{f,k,p}(q) = \sum_{i=1}^{n} q_i f\left(p_i . q_i^{-1}\right) \leq f\left(\sum_{i=1}^{n} q_i . p_i . q_i^{-1}\right) = f(1) =$$
$$= \sum_{i=1}^{n} p_i f\left(p_i . p_i^{-1}\right) = H^*_{f,k,p}(p)$$

Corolario 3.1.

La única medida $H^*_{f,k,p}(q)$, con $f(t)$ cóncava en R_+ que puede cumplir la condición de Dalton–Pielou, es la que corresponde al vector de probabilidades $(n^{-1}, ..., n^{-1})$.

Proposición 3.4.

Para funciones $f(t)$ cóncavas en R_+, $H^*_{f,k}(x)$ cumple la condición de Dalton–Pielou.

Demostración:

Ya que la desigualdad

$$H^*_{f,k,}(x_1, ..., x_i, ..., x_j, ..., x_n) - H^*_{f,k}(x_1, ..., x_i + \delta, ..., x_j - \delta, ..., x_n) \leq 0$$

para $x_i < x_i + \delta \leq x_j - \delta < x_j$ se reduce a

$$\phi(x_i) - \phi(x_i + \delta) + \phi(x_j) - \phi(x_j - \delta) \leq 0$$

con $\phi(t) = t.f\,(n^{-1}.t^{-1})$, y esta a su vez a

$$\frac{\phi(x_i + \delta) - \phi(x_i)}{\delta}(-\delta) + \frac{\phi(x_j) - \phi(x_j - \delta)}{\delta}\delta \leq 0$$

y por el teorema del valor medio a

$$\delta(\phi'(\theta_2) - \phi'(\theta_1)) \leq 0 \qquad \text{con} \quad \theta_1 \leq \theta_2$$

entonces, el resultado es evidente por la concavidad de $\phi(t)$ en R_+.

Particularizando a vectores de probabilidad $p = (p_1, p_2, ..., p_n)$, cumpliendo $p_1 + ... + p_n = 1$ y $p_1 \geq 0, ..., p_n \geq 0$, se tiene.

Corolario 3.2.

a) Las medidas de entropía regulares $H^*_{f,k}(x)$ alcanzan el mínimo valor para los vectores $p_1 = (1, 0, ..., 0), ..., p_n = (0, ..., 1)$.
b) Las medidas de entropía $H^*_{f,k}(x)$ alcanzan el máximo valor para vector $(n^{-1}, ..., n^{-1})$.

Corolario 3.3.

La única medida $H^*_{f,k,p}(x)$ cóncava en R^n_+ y que cumple la condición de Dalton–Pielou es la medida $H^*_{f,k}(x)$.

Corolario 3.4.

Una medida regular $H^*_{f,k}(x)$, definida por una función $f(x)$ cóncava en R_+, es no negativa si la constante k cumple la condición $k \geq -f(n^{-1}) + (1-n)C$ siendo $C = \lim_{t \longrightarrow \infty} t^{-1}.f(t)$.

Demostración:

Ya que el mínimo valor de $H^*_{f,k}(p)$ se alcanza entre otros para el vector de probabilidad $p_1 = (1, 0, ..., 0)$, y en tal caso se cumple

$$H^*_{f,k}(1, 0, ..., 0) = f\left(n^{-1}\right) + (n-1).C + k$$

entonces $H^*_{f,k}(p)$ es no negativo para todo vector de probabilidades, si se cumple la desigualdad $k \geq -f(n^{-1}) + (1-n).C$

4. HOMOGENEIDAD DE LAS FUNCIONES $H^*_{f,k}(x)$

En esta sección hemos determinado las funciones $f(t)$ que caracterizan las funciones $H^*_{f,k}(x)$ homogeneas en R^n_+, y hemos obtenido en este sentido el siguiente resultado

Proposición 4.1.

Las únicas funciones $H^*_{f,k}(x)$ homogeneas de grado λ en R^n_+ son las funciones que se expresan en la forma $H^*_{f,k}(x) = K. \sum\limits_{i=1}^{n} x_i^\lambda$, siendo K una constante arbitraria.

Demostración:

Por el teorema de Euler, $H^*_{f,k}(x)$ es homogenea de grado λ si la función $f(t)$ satisface la ecuación

$$t\left(f\left(n^{-1}t^{-1}\right) - n^{-1}t^{-1}f'\left(n^{-1}t^{-1}\right)\right) = \lambda\left[t.f\left(n^{-1}t^{-1}\right) + k.n^{-1}\right] \; \forall t \in R_+$$

Con el cambio de variable $y = n^{-'}t^{-'}$, la igualdad anterior se reduce a la educación diferencial

$$(1-\lambda)f(y) = y.f'(y) + k.\lambda.y \qquad \forall y \in R_+$$

ecuación que tiene por solución las funciones $f(t) = -k.t + C.t^{1-\lambda}$ para toda C constante real.

Así, la expresión de $H^*_{f,k}(x)$ se reduce a

$$H^*_{f,k}(x) = K.\sum_{i=1}^{n} x_i^\lambda$$

siendo λ el grado de homogeneidad y K una constante arbitraria. Obsérvese que la concavidad (convexidad) de las funciones $H^*_{f,k}(x)$ se obtiene para

$$K.\lambda(\lambda-1) \leq 0 \qquad (K.\lambda(\lambda-1) \geq 0)$$

y que la condición $H^*_{f,k}(1,...,0) = 0$ solo es posible para la medida constante, si bien las únicas medidas de entropía $H^*_{f,k}(x)$ que provienen de traslaciones de funciones $H^*_{f,k}(x)$, homogeneas en R^n_+, son las de forma

$$H^*_{f,k}(x) = K\left(1 - \sum_{i=1}^{n} x^\lambda\right) \qquad \text{con} \quad K.\lambda(\lambda-1) \geq 0$$

Particularizando a vectores de probabilidad, la medida anterior se reduce a un múltiplo de la medida de entropía de Havrda–Charvat de grado igual al índice de homogeneidad.

5. RELACIÓN DE LAS ϕ–ENTROPÍAS CON MEDIDAS DE CSISZAR

Ya que para vectores de probabilidad, la igualdad entre $H^*_{f,k}(p)$ y $H_{f,k}(p)$ se reduce a la relación funcional

$$t.f\left(n^{-1}t^{-1}\right) = \phi(t) + at + b$$

y esta a su vez a

$$f(x) = n.x.\phi\left(n^{-1}x^{-1}\right) + a + bnx$$

entonces se tiene el siguiente resultado

Proposición 5.1.

Todo funcional ϕ–entropía definido sobre vectores de probabilidas, puede interpretarse como una medida de divergencia (en el sentido de Csiszar) entre una distribución y la distribución en la que todos los sucesos son equiprobables. Además, la función que determina la medida de Csiszar queda univocamente determinada, salvo funciones lineales.

A título ilustrativo, presentamos a continuación una tabla en la que relacionamos algunas medidas clásicas de entropía con medidas de Csiszar.

ENTROPÍA	$H_{\phi}(p) = \sum_{i=1}^{n} \phi(p_i)$	$H^*_{f,k}(p) = \sum_{i=1}^{n} p_i f((np_i)^{-1})$
Shannon	$\phi(t) = -t \log t$	$f(t) = \log nt$
Shannon modificada	$\phi(t) = -t \log t - (1-t) \log (1-t)$	$f(t) = nt \log nt + (1-nt) \log (nt-1)$
Havrda–Charvat	$\phi(t) = (\alpha-1)^{-1}(t-t^{\alpha})$	$f(t) = (\alpha-1)^{-1}(1-(nt)^{1-\alpha})$
Havrda–Charvat modificada	$\phi(t) = (2^{1-\alpha}-1)^{-1}(t^{\alpha}-t)$	$f(t) = (2^{1-\alpha}-1)^{-1}((nt)^{1-\alpha}-1)$
Latter	$\phi(t) = t - 2t^2 + 2t^3 - t^4$	$f(t) = 1 - 2(nt)^{-1} + 2(nt)^{-2} - (nt)^{-3}$

6. NOTAS A LA SIMETRIZACIÓN

Si en lugar de considerar $H^*_{f,k}(x) = I_f\left((x_1,...x_n),(n^{-1},...,n^{-1})\right) + k$, consideramos la simetrización

$$^{s}H^*_{f,k}(x) = I_f\left((x_1,...,x_n),(n^{-1},...,n^{-1})\right) + I_f\left((n^{-1},...n^{-1}),(x_1,...,x_n)\right) + k$$

entonces se obtendria

$$^{s}H^*_{f,k}(x) = \sum_{i=1}^{n} x_i \left[f\left(n^{-1}.x_i^{-1}\right) + n^{-1}.x_i^{-1} f\left(nx_i\right)\right] = H^*_{\phi,k}(x)$$

con $\phi(t) = f(t) + t.f(t^{-1})$, resultando el estudio de las medidas $^{s}H^*_{f,k}(x)$ una consecuencia inmediata de los resultados ya obtenidos.

Analogamente, para las combinaciones convexas

$$^{\lambda}H^*_{f,k}(x) = \lambda I_f\left((x_1,...,x_n),(n^{-1},...,n^{-1})\right) +$$
$$+(1-\lambda)I_f\left((n^{-1},...,n^{-1}),(x_1,...,x_n)\right) + k$$

se tendría

$$^{\lambda}H^{*}_{f,k}(x) = H^{*}_{\psi,k}(x) \qquad \text{con} \qquad \psi(t) = \lambda.f(t) + (1-\lambda).t.f\left(t^{-1}\right)$$

y para el caso particular

$$^{t}H^{*}_{f,k}(x) = I_f\left(\left(n^{-1},...,n^{-1}\right),(x_1,...,x_n)\right)$$

resultaría

$$^{t}H^{*}_{f,k}(x) = H^{*}_{g,k}(x), \qquad \text{siendo} \qquad g(t) = t.f\left(t^{-1}\right)$$

7. COMENTARIOS A LA EXTENSIÓN AL CASO ABSOLUTAMENTE CONTINUO

Para las familias G. de funciones de densidad de probabilidad definidas sobre un mismo soporte compacto, podemos considerar medidas $H^{*}_{f,k}$ a traves de la expresión

$$H^{*}_{f,k}(g) = \int_{\mathrm{Sop}\ g} g.f\left(\frac{u}{g}\right)dx + k$$

siendo u la distribución uniforme en el soporte de las funciones de G. Obviamente, el estudio de tales medidas es paralelo al estudio ya realizado, y además, sus resultados son análogos.

Observese que, si la medida del soporte de las funciones de G no es finito, la extensión no es válida para para la indefinición de la función u, y en tal caso, se deberá pensar en la distancia a una distribución fijada de la familia.

8. BIBLIOGRAFÍA

[1] **Boekee, D.E.** y **Van Der Lubbe**. "The R–norm information measure". Information and Control 45, 135-155 (1980).

[2] **Bourguignon, F.** "Descomposable income inequality measures". Econometrika 47, 901-920 (1979).

[3] **Burbea, J.** y **Rao, C.R.** "On the convexity of some divergence measures based on entropy functions". IEEB Trans. Information Theory 28, 489-495 (1982).

[4] **Csiszar, I** "A class of measures of informativity of observation channels". Periodica Math. Hungarica 2, 192-213 (1972).

[5] **Ferentinos, K.** y **Papaivannou, T.** "New parametric measures of information". Information and Control 51, 193-208 (1981).

[6] **Havrda, J.** y **Charvat, F.** "Quantification method in classification peocesses: Concept of structural α–entropy". Kybernetika 3, 30-35 (1967).

[7] **Kullback, S.** "Information theory and Statistics". Dover (1967).

[8] **Latter, B.D.H.** "Measures of genetic distance between individuals and populations". Genetic Structure of Populations, 27-39. Publicaciones Univ. Hawaii, Honolulu (1973).

[9] **Rao, C.R.** "Diversity: its measurement, descomposition, apportionment and analysis". Sankhya 44, 1-22 (1982).

[10] **Renyi, A.** "On measures of entropy and information". Proceedings fourth Berkeley Symp. 1, 547-561 (1961).

[11] **Rockafellar, R.T.** "Convex analysis". Princeton. New Jersey (1970).

[12] **Salicrú, M** y **Calvo, M.** "Un índice de incertidumbre asociado a J–divergencias". XVI Congreso SEIO (1986).

[13] **Shannon, C.E.** "A mathematical theory of communications". Bell System Tech. Journal 27, 379-423, 623-656 (1948).