TRABAJOS DE ESTADISTICA
Vol. 2. Núm. 2, 1987 pp. 115 a 124

ALGUNAS CONSIDERACIONES SOBRE LA INFORMACION GENERALIZADA

Mª del PILAR ZULUAGA ARIAS

Departamento de Estadística e I.O.
Facultad de Medicina
Universidad Complutense
28040 MADRID

Revisado noviembre 1987

RESUMEN

De Groot ha definido una medida de información que generaliza la información de Kullback como medida de discrepancia entre dos distribuciones.

En el presente artículo definimos la información modal, como caso particular de la información generalizada, la cual servirá para demostrar que las principales propiedades de la información de Kullback no se verifican para la medida dada por De Groot.

Palabras clave: información de Kullback, información generalizada, información modal.

Clasificación A.M.S.: 62B10.

ABSTRACT

This paper deals with De Groot's information measure, which generalizes the Kullback definition of information when discrepance between distributions is to be measured. We define the modal information as a particular generalized information, and it is applied in order to prove that the main properties of Kullback information are not verified by the generalized De Groot measure.

Key word: generalized information, Kullback information, modal information.

A.M.S. Classification: 62B10.

1. INTRODUCCION

Dentro de las medidas de información una de las más usuales es la información de Kullback (Kullback (1968)). Dicha información tiene muy buenas propiedades como medida de la discrepancia entre dos distribuciones de probabilidad q_1 y q_2 definidas ambas sobre el mismo espacio medible, siendo una forma usual de notación $I_k(q \to q_2)$.

Otra de las medidas de información más usuales es la información de Shannon (Lindley (1956)). Dicha medida tiene muy buenas propiedades expresando la información que un experimento X, definido sobre un espacio paramétrico W, da sobre W cuando la distribución a priori es $p(w)$ siendo la forma más usual de notación $I_{sh}(X;p(w))$.

Esta medida de información y la de Kullback están relacionadas por

$$E_x I_k(p(w) \to p(w/x)) = I_{sh}(X;p(w))$$

Una generación de la información de Shannon fue estudiada por De Groot (1962), definiendo la información en función de cualquier función de incertidumbre u como $I_u(X;p(w)) = u(p(w)) - E_x u(p(w/x))$. Estas medidas de información siguen verificando muy buenas propiedades (De Groot (1962), García-Carrasco (1985), (1986a)) a la hora de medir la información que X da sobre W cuando la a priori es $p(w)$.

Una generalización de $I_k(q_1 \to q_2)$ fue dada por De Groot (1984), que notaremos por $I_G(q_1 \to q_2)$. Además se verifica $E_x I_G(p(w) \to p(w/x)) = I_u(X;p(w))$ para ciertas I_u e I_G en cada caso.

El hecho anterior nos sugiere pensar que quizá la medida $I_G(q_1 \to q_2)$ tenga tan buenas propiedades como $I_k(q_1 \to q_2)$ para medir la discrepancia entre dos distribuciones q_1 y q_2. En este artículo se demuestra que esto no es así, para ello se define una información del tipo $I_G(q_1 \to q_2)$ la cual no verifica dichas propiedades.

2. ALGUNAS DEFINICIONES PREVIAS

En este apartado daremos algunas definiciones que serán utilizadas a lo largo de todo el trabajo.

Definicion 2.1 (De Groot (1984))

Dado un problema de decisión (W, A, v) con espacio paramétrico $W = \{w_1, ..., w_m\}$, espacio de acciones A y función de utilidad $v(v(w_i, a)$ con $a \in A)$, y dadas dos distribuciones q_1 y q_2 sobre W, se define la información generalizada que se obtiene al pasar de q_1 a q_2 como:

$$I_G(q_1 \to q_2) = \sup_{a \in A} \sum_{i=1}^{m} v(w_i, a) q_2(w_i) - \\ - \sum_{i=1}^{m} v(w_i, a_1) q_2(w_i)$$

con a_1 tal que

$$\sup_{a \in A} \sum_{i=1}^{m} v(w_i, a) q_1(w_i) = \sum_{i=1}^{m} v(w_i a) q_1(w_i)$$

donde en esta definición y en todo lo que sigue, se supone que a_1 existe y es única.

Definición 2.2 (De Groot (1984))

Dado un problema de decisión (W, A, v) y un experimento sobre W, $X = \{\mathscr{X}, p(x(w_i), i = 1, ..., m\}$ y se define la información generalizada esperada que X da sobre W cuando la distribución a priori es $p(w)$ como:

$$I_G(X; p(w)) = E_x I_G(p(w) \to p(w/x))$$

Como ya indicamos en el apartado 1 los casos más tratados de estas dos definiciones son, respectivamente, la información de Kullback y la información de Shannon.

Uno de los ejemplos dados en De Groot (1984) es la información modal definida por

$$I_M(q_1 \to q_2) = q_2(w^2) - q_2(w^1)$$

siendo w^i la moda de q_i, $i = 1, 2$, y $A = \{p(\cdot)\}$ conjunto de distribuciones sobre W.

En García-Carrasco (1986 b) se pone de manifiesto el importante papel que tienen en los problemas de decisión las funciones de evaluación modales normalizadas que allí se definen. La información modal puede introducirse definiéndola a partir de dichas funciones de evaluación. Esto hace de la información modal una de las mejores candidatas a la hora de verificar las mismas buenas propiedades que la de Kullback.

3. ALGUNAS PROPIEDADES DE LA INFORMACION GENERALIZADA

La información modal por ser un caso particular de la información generalizada, nos servirá como ejemplo para demostrar que la información generalizada no verifica las mejores propiedades que reflejan el buen comportamiento de la información de Kullback (Kullback (1968)); como en principio se podría esperar.

Notaremos por $I_G(q_1 \to q_2/X, Y)$ a la información generalizada, cuando nos refiramos a las q_i definidas en el espacio producto y por abuso de notación y no dando lugar a error notaremos también por q_1 y q_2 a las marginales con lo que $I_G(q_1 \to q_2/X)$ es la información cuando nos referimos a las marginales sobre X (análogamente para Y).

Proposición 3.1

Sea (W, A, v) un problema de decisión. Sean X e Y definidos sobre W. Sean q_1 y q_2 dos distribuciones sobre $X \times Y$. Entonces no son ciertas las siguientes afirmaciones:

(i) $I_G(q_1 \to q_2/X, Y) \geqslant I_G(q_1 \to q_2/X)$

$I_G(q_1 \to q_2/X, Y) \geqslant I_G(q_1 \to q_2/Y)$

(ii) Si X e Y son independientes respecto a q_1 y q_2:

$I_G(q_1 \to q_2/X, Y) = I_G(q_1 \to q_2/X) + I_G(q_1 \to q_2/Y)$

(iii) $I_G(q_1 \to q_2/X, Y) = I_G(q_1 \to q_2/Y) + I_G(q_1 \to q_2/X/Y)$

donde

$$I_G(q_1 \to q_2/X/Y) = E_y I_G(q_1 \to q_2/X/y)$$

(iv) Denotemos por $X^{(n)}$ a la realización de n pruebas independientes de X:

$$I_G(q_1 \to q_2/X^{(n)}) = n I_G(q_1 \to q_2/X)$$

Demostración

Sean q_1 y q_2 dos distribuciones sobre $X \times Y$, definidas como sigue:

q_1	x_1	x_2	x_3	
y_1	$\frac{1}{81}$	$\frac{2}{81}$	$\frac{6}{81}$	$\frac{1}{9}$
y_2	$\frac{3}{81}$	$\frac{6}{81}$	$\frac{18}{81}$	$\frac{3}{9}$
y_3	$\frac{5}{83}$	$\frac{10}{81}$	$\frac{30}{81}$	$\frac{5}{9}$
	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{6}{9}$	

q_2	x_1	x_2	x_3	
y_1	$\frac{3}{40}$	$\frac{6}{40}$	$\frac{15}{40}$	$\frac{3}{5}$
y_2	$\frac{1}{40}$	$\frac{2}{40}$	$\frac{5}{40}$	$\frac{1}{5}$
y_3	$\frac{1}{40}$	$\frac{2}{40}$	$\frac{5}{40}$	$\frac{1}{5}$
	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{5}{8}$	

con lo cual X e Y son independientes respecto a q_1 y q_2.

Si (x^i, y^i) es tal que $\max\limits_{x,y} q_i(x, y) = q_i(x^i, y^i)$ se verifica por la independencia que $\max\limits_{x} q_i(x) = q_i(x^i)$ y $\max\limits_{y} q_i(y) = q_i(y^i)$ para $i = 1, 2$ con lo que se tiene:

$$x^1 = x_3 \quad , \quad y^1 = y_3 \quad , \quad x^2 = x_3 \quad , \quad y^2 = y_1$$

con lo que:

$$I_M(q_1 \rightarrow q_2/X, Y) = q_2(x^2, y^2) - q_2(x^2, y^1) = \frac{15}{40} - \frac{5}{40} = \frac{10}{40}$$

$$I_M(q_1 \rightarrow q_2/X) = q_2(x^2) - q_2(x^1) = \frac{5}{8} - \frac{5}{8} = 0$$

$$I_M(q_1 \rightarrow q_2/Y) = q_2(y^2) - q_2(y^1) = \frac{3}{5} - \frac{1}{5} = \frac{2}{5} = \frac{16}{40}$$

Con lo que

$$I_M(q_1 \rightarrow q_2/Y) > I_M(q_1 \rightarrow q_2/X, Y)$$

con lo que se demuestra que en general no es cierto (i).

Además

$$I_M(q_1 \rightarrow q_2/X, Y) = \frac{10}{40} \neq I_M(q_1 \rightarrow q_2/X) + I_M(q_1 \rightarrow q_2/Y) = 0 + \frac{16}{40}$$

Con lo que se demuestra que en general no es cierto (ii).

Ahora por la independencia y vistos los valores de $I_M(q_1 \rightarrow q_2/X, Y)$ e $I_M(q_1 \rightarrow q_2/Y)$ se ve que en general no es cierto (iii).

Para ver que en general no es cierto (iv) basta observar que

$$I_M(q_1 \rightarrow q_2/X^{(n)}) = q_2(x^2)^n - q_2(x^1)^n$$

que en este caso concreto es

$$I_M(q_1 \rightarrow q_2/Y^{(2)}) = \left(\frac{3}{5}\right)^2 - \left(\frac{1}{5}\right)^2 = \frac{8}{25} \neq \frac{4}{5}$$

Proposición 3.2

Sean q_1 y q_2 dos distribuciones sobre X. Sea T un experimento consistente en una agrupación de X. Entonces, en general, no es cierto

$$I_G(q_1 \rightarrow q_2/T) \leqslant I_G(q_1 \rightarrow q_2/X)$$

Demostración

Basta ver el ejemplo siguiente. Consideramos $X = \{x_1, x_2, x_3, x_4\}$ y $T = \{t_1, t_2\}$ tal que $t_1 = x_1 \cup x_3$, $t_2 = x_2 \cup x_4$ de modo que

X	x_1	x_2	x_3	x_4
q_1	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{8}{20}$	$\frac{9}{20}$
q_2	$\frac{3}{20}$	$\frac{1}{20}$	$\frac{9}{20}$	$\frac{7}{20}$

T	t_1	t_2
q_1	$\frac{9}{20}$	$\frac{11}{20}$
q_2	$\frac{12}{20}$	$\frac{8}{20}$

Observamos que

$$I_M(q_1 \to q_2/X) = q_2(x_3) - q_2(x_4) = \frac{9}{20} - \frac{8}{20} = \frac{1}{20}$$

$$I_M(q_1 \to q_2/T) = q_2(t_1) - q_2(t_2) = \frac{11}{20} - \frac{9}{20} = \frac{2}{20}$$

Con lo que queda demostrado que, en general, no es cierto

$$I_G(q_1 \to q_2/T) \leqslant I_G(q_1 \to q_2/X)$$

Para ver la siguiente proposición daremos previamente la definición de partición suficiente dada en Kullback (1968).

Definición 3.1

Sea $X = \{x_{11}, ..., x_{1m_1}, ..., x_{k1}, ..., x_{hm_h}\}$ con distribución q_1 o q_2. Sea T una partición de X tal que $t_i = T(x_{ij})$ son $j = 1, ..., m_i$ para $i \in \{1, ..., h\}$; se dice que T es una partición suficiente para X si

$$\frac{q_1(x_{ij})}{q_1(t_i)} = \frac{q_2(x_{ij})}{q_2(t_i)} \text{ para todo } j, i$$

Proposición 3.3

Con la notación anterior, en general, no es cierto $I_G(q_1 \to q_2/X) = I_G(q_1 \to q_2/T)$ siendo T una partición suficiente de X.

Demostración

Utilizaremos también aquí un ejemplo basado en la información modal. Consideramos

X	x_{11}	x_{12}	x_{21}	x_{31}	x_{32}	x_{33}	x_{34}	x_{35}	x_{36}
q_1	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{4}{18}$	$\frac{2}{18}$	$\frac{2}{18}$	$\frac{2}{18}$	$\frac{2}{18}$	$\frac{2}{18}$	$\frac{2}{18}$
q_2	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{54}$	$\frac{1}{54}$	$\frac{1}{54}$	$\frac{1}{54}$	$\frac{1}{54}$	$\frac{1}{54}$

sea T tal que $t_i = T(x_{ij})$ $\forall j$, de donde se obtienen las probabilidades inducidas

T	t_1	t_2	t_3
q_1	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{6}{9}$
q_2	$\frac{1}{9}$	$\frac{7}{9}$	$\frac{1}{9}$

Observamos que por construcción

$$\frac{q_1(x_{ij})}{q_1(t_i)} = \frac{q_2(x_{ij})}{q_2(t_i)}$$

con lo que T es partición suficiente y además

$$I_M(q_1 \to q_2/X) = q_2(x_{21}) - q_2(x_{21}) = 0$$

$$I_M(q_1 \to q_2/T) = q_2(t_1) - q_2(t_3) = \frac{6}{9}$$

4. DISCUSION

Las proposiciones del apartado 3 son todas negación de una serie de propiedades, dichas propiedades son las principales y más intuitivas verificadas por la información de Kullback.

Concluimos, por lo tanto, que la información generalizada propuesta por De Groot (Definición 2.1) no tiene, en general, tan buenas propiedades cmo podría esperarse. Quizás la generalización de la información de Kullback, como medida de discrepancia, deba ser en la línea de las f-divergencias propuesto por Csiszar (1977).

Hay que destacar, sin embargo, el buen comportamiento de la información generalizada esperada de De Groot (Definición 2.2) según se pone de manifiesto en los artículos citados.

5. REFERENCIAS

CSISZAR, I. (1977): «Information Measures: A Critical Survey», *Transactions of 7th Prague Conference on Information Theory, Stat*, Decision Function, Random Processes 1974, Boston: Reidel Publications.

DE GROOT, M. H. (1962): «Uncertainty, information and sequential experiments», *Ann. Math. Statist.*, 33, 404-419.

DE GROOT, M. H. (1984): «Changes in utility as information», *Theory and Decision*, 17, 287-303.

GARCIA-CARRASCO, M. P. (1985): «Criterio de comparación de experimentos basado en la información generalizada», *Trabajos de Estadística* (pendiente de publicación).

GARCIA-CARRASCO, M. P. (1986 a): «Distribuciones mínimo informativas, caso de espacio paramétrico discreto y finito», *Qüestió*, vol. 10, n.° 1, 7-12.

GARCIA-CARRASCO, M. P. (1986 b): «Funciones de evaluación modales», *Trabajos de Estadística* (pendiente de aceptación).

KULLBACK, S. (1968): «Information Theory and Statistics», *Dover Publications*, New York.

LINDLEY, D. V. (1956): «On a measure of the information provided by an experiment», Ann. Math. Statist, 27, 986-1005.