

MEDIDAS DE INCERTIDUMBRE ASOCIADAS A J -DIVERGENCIAS

SALICRÚ PAGES, M., y CALVO LLORCA, M.
Departamento de Estadística
Universidad de Barcelona
Avda. Diagonal, 645
08028 BARCELONA

RESUMEN

En este trabajo se presenta la familia de medidas de incertidumbre asociadas a J -divergencias, que resultan de la distancia entre una distribución y la distribución en la que todos los sucesos son equiprobables. Se estudian propiedades teóricas de la familia atendiendo a la pérdida de incertidumbre, a la concavidad y a la condición de medida decisiva. Finalmente se compara a nivel muestral la medida de incertidumbre definida por la función $\phi(t) = -t \log t$ con las medidas de entropía comúnmente usadas.

Palabras clave: Medidas de incertidumbre, J -divergencias, análisis de Montecarlo.

Clasificación A.M.S.: 62H30, 62E25, 62-07.

SUMMARY:

In this paper we show a family of uncertainty measures, related with the J -divergences. They are obtained through a distance between a distribution and the distribution with equiprobable events. We study some theoretical properties of the family, taking into account the loss of uncertainty, the concavity and the condition to be a decisive measure. Finally, we compare some sample characteristics of the uncertainty measures defined by the function $\phi(t) = -t \log(t)$, with the entropy measures most commonly used.

Revisado, mayo 1988

Key words: Uncertainty measures, J -divergences, Montecarlo analysis.

A.M.S. Classification: 62H30, 62E25, 62-07.

Title: Uncertainty measures associated with the J -divergences.

1. INTRODUCCION

Un problema interesante que se plantea en teoría de la información es la elección adecuada de medidas de incertidumbre. En este sentido, las funciones de incertidumbre más usadas provienen de índices de entropía, tal como podemos ver en Lindley (1956), Rényi (1961), De Groot (1970), Havrda-Charvat (1967), Theodorescu (1977) y Boeke (1980).

Una alternativa a estas medidas clásicas de entropía la tenemos en los índices de incertidumbre basados en la distancia a una distribución fijada. En este sentido, analizaremos en este trabajo la J -divergencia entre una distribución y la distribución en la que todos los sucesos son equiprobables.

Hemos dividido el estudio en dos partes: en la primera, hemos analizado la pérdida de incertidumbre y la concavidad de las medidas presentadas, atendiendo a la función que la determina; en la segunda, hemos comparado características muestrales de once índices de entropía con el índice que proviene de la J -divergencia para la función $\phi(t) = -t \log t$.

Hemos elegido la función $\phi(t) = -t \log t$, pues está determinada J -divergencia invariante, a nivel diferencial, frente a transformaciones no singulares de variables aleatorias y parámetros.

2. DEFINICIONES PREVIAS

Dados A_1, A_2, \dots, A_n sucesos mutuamente excluyentes de probabilidades p_1, \dots, p_n , respectivamente, verificando la condición: $p_1 + \dots + p_n = 1$, definimos como *funciones de incertidumbre asociadas a J -divergencias* las funciones del tipo:

$$I_{\phi}(p_1, \dots, p_n) = \sum_{i=1}^n \left[\frac{1}{2} \left(\phi(p_i) + \phi(n^{-1}) \right) - \phi \left(\frac{1}{2}(p_i + n^{-1}) \right) \right] + K \quad (1)$$

donde ϕ es una función de $I = [0, 1]$ en R dos veces diferenciable con continuidad, K es una constante y de forma que $I_\phi(p_1, \dots, p_n)$ sea cóncava y no negativa en I^n .

Además, se dice que una función de incertidumbre $I_\phi(p_1, \dots, p_n)$ es *decisiva* si cumple que

$$I_\phi(1, 0, \dots, 0) = I_\phi(0, 1, 0, \dots, 0) = \dots = I_\phi(0, \dots, 0, 1) = 0.$$

3. CONDICIONES PARA LA CONSTANTE K

En esta sección encontraremos los puntos $P = (p_1, \dots, p_n)$ para los que la función $I_\phi(p_1, \dots, p_n)$ alcanza los valores máximo y mínimo. Así, de forma inmediata, el valor de la constante K quedará determinado por la función $\phi(t)$ y por el número de sucesos n , al imponer la condición decisiva o la no negatividad a la función cóncava $I_\phi(p_1, \dots, p_n)$.

En este sentido, hemos obtenido los siguientes resultados:

Proposición 3.1

Si se produce una transferencia de probabilidad de un suceso más probable a otro menos probable conservándose el orden, la incertidumbre crece, es decir

$$I_\phi(p_1, \dots, p_i, \dots, p_j, \dots, p_n) \leq I_\phi(p_1, \dots, p_i + \delta, \dots, p_j - \delta, \dots, p_n)$$

para $p_i < p_i + \delta < p_j - \delta < p_j$ y con $I_\phi(p_1, \dots, p_n)$ cóncava del tipo (1).

DEMOSTRACIÓN:

Por ser $-I_\phi(x_1, \dots, x_n)$ convexa y simétrica, $-I_\phi(x_1, \dots, x_n)$ es Schur-convexa, es decir, cumple la propiedad (anexo I).

$$(x_i - x_j) \left(\frac{\partial I_\phi}{\partial x_i} - \frac{\partial I_\phi}{\partial x_j} \right) \leq 0 \quad \text{para todo } x_i, x_j \text{ de } I$$

Al ser $I_\phi(x_1, \dots, x_n)$ del tipo (1), la condición anterior se reduce a la desigualdad

$$(x_i - x_j) \left(\left(\phi'(x_i) - \phi'\left(\frac{1}{2}(x_i + n^{-1})\right) \right) - \left(\phi'(x_j) - \phi'\left(\frac{1}{2}(x_j + n^{-1})\right) \right) \right) \leq 0$$

Así, para todo $\varepsilon_1, \varepsilon_2 \in (0, \delta)$, para $x_i = p_i + \varepsilon_1$ y $x_j = p_j - \varepsilon_2$ se tendrá la condición

$$\begin{aligned} & (\phi'(p_i + \varepsilon_1) - \phi'(\frac{1}{2}(p_i + \varepsilon_1 + n^{-1}))) - \\ & - (\phi'(p_j - \varepsilon_2) - \phi'(\frac{1}{2}(p_j - \varepsilon_2 + n^{-1}))) \geq 0 \end{aligned}$$

de la que se deduce (por el teorema del valor medio)

$$\begin{aligned} & \frac{(\phi(p_i + \delta) - \phi(\frac{1}{2}(p_i + \delta + n^{-1}))) - (\phi(p_i) - \phi(\frac{1}{2}(p_i + n^{-1})))}{\delta} \geq \\ & \geq \frac{(\phi(p_j) - \phi(\frac{1}{2}(p_j + n^{-1}))) - (\phi(p_j - \delta) - \phi(\frac{1}{2}(p_j - \delta + n^{-1})))}{\delta} \end{aligned}$$

y de la reordenación de los términos y su sustitución en la expresión correspondiente, resulta

$$I_\phi(p_1, \dots, p_i + \delta, \dots, p_j - \delta, \dots, p_n) \geq I_\phi(p_1, \dots, p_n)$$

De esta desigualdad se deducen de forma inmediata los siguientes resultados

Corolario 3.1

$I_\phi(p_1, \dots, p_n)$ alcanza el máximo en el punto (n^{-1}, \dots, n^{-1}) y el mínimo en los puntos $(1, 0, \dots, 0), \dots, (0, \dots, 0, 1)$.

Corolario 3.2

Las medidas de incertidumbre asociadas a J -divergencias son decisivas si la constante K toma el valor

$$K_0 = \phi\left(\frac{n+1}{2n}\right) + (n-1)\phi\left(\frac{1}{2n}\right) - \frac{1}{2}[\phi(1) + (n-1)\phi(0)] - \frac{n}{2}\phi(n^{-1})$$

Corolario 3.3

Para $K \geq K_0$, las medidas cóncavas $I_\phi(p_1, \dots, p_n)$ definidas en (1) son siempre medidas de incertidumbre.

4. CONDICIONES PARA LA FUNCION ϕ

En esta sección hemos estudiado las condiciones suficientes para la concavidad de $I_\phi(p_1, \dots, p_n)$ atendiendo a la concavidad y convexidad de $\phi(t)$ y $1/\phi''(t)$. En este sentido, hemos encontrado los siguientes resultados

Proposición 4.1

Si $\phi(t)$ es una función cóncava en $I = [0, 1]$ y $1/\phi''(t)$ es convexa en I , entonces $I_\phi(p_1, \dots, p_n)$ es cóncava en I^n .

DEMOSTRACIÓN 4.1

De la convexidad de $1/\phi''(t)$ se tiene la desigualdad

$$\frac{1}{\phi''(\frac{1}{2}(p_i + n^{-1}))} \leq \frac{1}{2} \frac{1}{\phi''(p_i)} + \frac{1}{2} \frac{1}{\phi''(n^{-1})}$$

desigualdad que por la concavidad de $\phi(t)$ se reduce a

$$\frac{1}{\phi''(\frac{1}{2}(p_i + n^{-1}))} \leq \frac{1}{\phi''(p_i)}$$

y ésta a su vez se escribe en la forma

$$\phi''(p_i) - \frac{1}{2} \phi''(\frac{1}{2}(p_i + n^{-1})) \leq 0$$

Así, se tiene:

$$\frac{\partial^2 I_\phi}{\partial p_i^2} = \frac{1}{2} \phi''(p_i) - \frac{1}{4} \phi''\left(\frac{1}{2}(p_i + n^{-1})\right) \leq 0 \quad \text{para todo } p_i \in I$$

$$\frac{\partial^2 I_\phi}{\partial p_i \partial p_j} = 0 \quad \text{para todo } p_i, p_j \in I \quad \text{con } i \neq j$$

condición equivalente a la concavidad de $I_\phi(p_1, \dots, p_n)$ en I^n .

De forma inmediata obtenemos los siguientes resultados.

Corolario 4.1

Una medida $I_\phi(p_1, \dots, p_n)$ definida por una función $\phi(t)$ cóncava con $1/\phi''(t)$ convexa y por la constante K_o es una medida de incertidumbre decisiva.

Corolario 4.2

Para $\phi(t) = -t \cdot \log t$ y $K_o = \frac{1}{2} \log n + \log 2 - \frac{n+1}{2n} \log(n+1)$ la medida $I_\phi(p_1, \dots, p_n)$ es una medida de incertidumbre decisiva.

5. INTERES FUNCIONAL DE LA FAMILIA PROPUESTA

Además de estas propiedades teóricas, Salicrú (1987) ha comprobado que esta familia contiene los índices de entropía más usados. En este sentido, la entropía de Shannon y la entropía de Havrda-Charvat han resultado como soluciones de las ecuaciones funcionales

$$1/2\phi(x) - \phi(1/2(x + n^{-1})) = -x \log x + ax + b$$

$$1/2\phi(x) - \phi(1/2(x + n^{-1})) = x^\alpha + ax + b$$

De forma inmediata, se han obtenido también la entropía de grado α y la entropía de Shannon modificada a partir de las funciones que determinan la entropía de Havrda-Charvat y la entropía de Shannon.

La entropía de Rényi $H_{n,\alpha}^R(p)$ y la γ -entropía $H_{n,\alpha}^\gamma(p)$ han resultado como función de entropía de Havrda-Charvat a partir de las relaciones

$$H_{n,\alpha}^R(p) = \frac{1}{1-\alpha} \log(1 - (\alpha-1)I_\phi^\alpha(p))$$

$$H_{n,\alpha}^\gamma(p) = \frac{1}{1-2^{(1/\alpha)-1}} (1 - (1 - (\alpha-1)I_\phi^\alpha(p))^{1/\alpha})$$

donde $I_\phi^\alpha(p)$ representa la entropía de Havrda-Charvat de grado α .

Finalmente, la entropía de Latter $H_L(p)$

$$H_L(p_1, \dots, p_n) = 1 - \sum_{i=1}^n p_i^2 - \sum_{i=1}^n p_i^2(1-p_i)^2$$

se ha obtenido para la función $\phi(t) = a_2 t^2 + a_3 t^3 + a_4 t^4$ siendo

$$a_4 = -\frac{16}{7}, \quad a_3 = \frac{16}{3} - \frac{32}{21} n^{-1} \quad \text{y} \quad a_2 = -8 + \frac{8}{n} - \frac{8}{7} n^{-2}.$$

A título ilustrativo, vale la pena mencionar que la entropía cuadrática se obtiene para la función $\phi(t) = -4t^2$, y la entropía de Havrda-Charvat de grado 3 se obtiene para la función $\phi(t) = -\frac{4}{3}t^3 - \frac{2}{n}t^2$.

6. INDICES A COMPARAR

Para A_1, A_2, \dots, A_n sucesos mutuamente excluyentes con probabilidades p_1, p_2, \dots, p_n , respectivamente y verificando la condición $p_1 + p_2 + \dots + p_n = 1$, consideramos las doce medidas de incertidumbre que relacionamos a continuación

$$H_1 = - \sum_{i=1}^n p_i \log p_i \quad (\text{entropía de Shannon})$$

$$H_2 = 1 - \sum_{i=1}^n p_i^2 \quad (\text{entropía cuadrática})$$

$$H_i = \frac{\alpha}{\alpha - 1} \left(1 - \left(\sum_{i=1}^n p_i^\alpha \right)^{1/\alpha} \right) \quad \begin{cases} \text{con } i = 3 \text{ para } \alpha = 1/2 \\ \text{con } i = 4 \text{ para } \alpha = 2 \\ \text{con } i = 5 \text{ para } \alpha = 3 \end{cases}$$

(γ -entropía de grados 1/2, 2 y 3, respectivamente)

$$H_6 = 1 - \max \{p_1, \dots, p_n\} \quad (\text{medida máxima})$$

$$H_7 = \frac{1}{2} \left(1 - \sum_{i=1}^n p_i^3 \right) \quad (\text{entropía de Havrda-Charvat, de grado 3})$$

$$H_i = \frac{1}{1 - \alpha} \log \sum_{i=1}^n p_i^\alpha \quad \begin{cases} \text{con } i = 8 \text{ para } \alpha = 1/2 \\ \text{con } i = 9 \text{ para } \alpha = 2 \\ \text{con } i = 10 \text{ para } \alpha = 3 \end{cases}$$

(entropía de Rényi de grados 1/2, 2 y 3, respectivamente)

$$H_{11} = (\sqrt{2} - 1)^{-1} \left(\sum_{i=1}^n \sqrt{p_i} - 1 \right)$$

(entropía de Havrda-Charvat modificada de grado 1/2)

$$H_{12} = 1/2 \left(\sum_{i=1}^n \left(\log \left(p_i + \frac{1}{n} \right) - p_i \log p_i \right) \right) + \log n - \frac{1+n}{2n} \log(n+1)$$

(medida de incertidumbre asociada a la J -divergencia determinada por la función $\phi(t) = -t \log t$).

Los estimadores de estas doce medidas los obtenemos sustituyendo las probabilidades teóricas p_i por las frecuencias relativas de las observaciones en las particiones A_i .

7. METODOLOGIA EMPLEADA

Como primera aproximación a la comparación de la bondad de los estimadores, planteamos una simulación por el método de Montecarlo con los estadísticos:

$$\sigma^2(H) \quad \text{y} \quad V = \frac{1}{k} \sum_{i=1}^k \left(\frac{H_i - H_t}{H_{\max}} \right)^2$$

donde $\sigma(H)$ es la desviación típica muestral de H ; H_t es el valor teórico de H ; H_{\max} es el valor máximo teórico que puede alcanzar H ; H_i es el valor de H en la muestra « i »; y k es el número de muestras.

La razón de tomar un estadístico de dispersión estandarizado proviene de considerar que la comparación entre índices sólo es posible cuando se han realizado las medidas en las mismas unidades. En este sentido, distintos autores han planteado la necesidad de normalizar las medidas con distinto recorrido y comparar posteriormente. La razón de considerar también un estadístico de dispersión no estandarizado proviene de comprobar que en diversas situaciones la estandarización no es deseable por el efecto de los sucesos residuales. Así, nos ha parecido conveniente en este trabajo realizar en primer lugar una comparación atendiendo solamente al valor de V , y en segundo lugar, comparar considerando conjuntamente los dos índices.

Los puntos de la variedad paramétrica de las distribuciones multinomiales en los que hemos estudiado las características muestrales son los siguientes:

$$P^{(0)} = M(1/n, \dots, 1/n; N)$$

$$P^{(1)} = M(1, 0, \dots, 0; N)$$

$$P^{(2)} = M(p_1^{(2)}, \dots, p_n^{(2)}; N) \text{ siendo } \begin{cases} p_i^{(2)} = (1/n) + ((n/2) + 1 - i)\varepsilon & \text{para } 2i \leq n \\ p_i^{(2)} = (1/n) + ((n/2) - i)\varepsilon & \text{para } 2i > n \end{cases}$$

con n par y $\varepsilon = 2/n^2(1 - 10^{-3})$.

$$P^{(3)} = M(p_1^{(3)}, \dots, p_n^{(3)}; N) \text{ siendo } \begin{cases} p_i^{(3)} = (1/n) + \delta & \text{para } 2i \leq n \\ p_i^{(3)} = (1/n) - \delta & \text{para } 2i > n \end{cases}$$

con n par y $\delta = 0,8/n$.

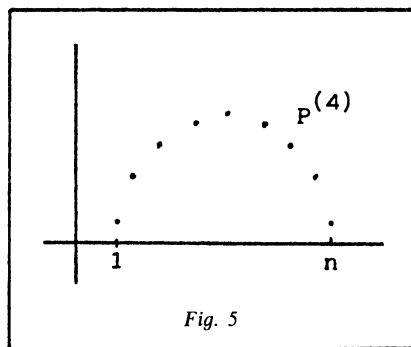
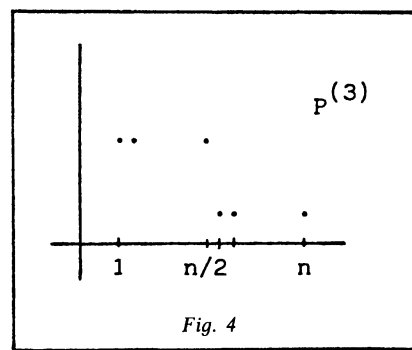
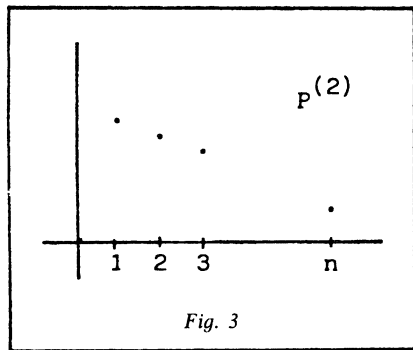
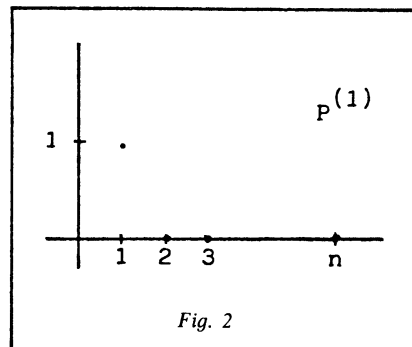
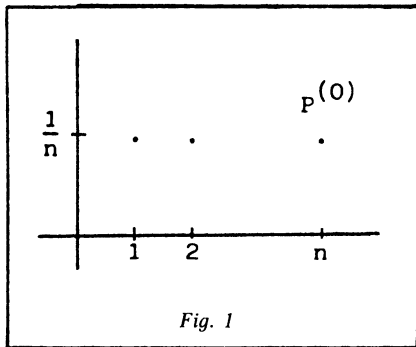
$$P^{(4)} = M(p_1^{(4)}, \dots, p_n^{(4)}; N) \text{ siendo } P_i^{(4)} = \frac{1}{\sqrt{2\pi}} \int_{a_i}^{b_i} e^{-u^2} du,$$

$$\text{y los extremos } a_i = -4 + 8(i - 1)/n \quad \text{y} \quad b_i = -4 + 8i/n$$

Los puntos $P^{(0)}$ y $P^{(1)}$ corresponden, respectivamente, a las distribuciones que determinan máxima y mínima incertidumbre, $P^{(2)}$ corresponde a una distribución con escalonamiento gradual en las probabilidades, $P^{(3)}$ viene asociada a un salto brusco entre dos grupos de particiones, y

$P^{(4)}$ corresponde a realizar n particiones en el intervalo $[-4, 4]$ y asignar a las probabilidades teóricas p_i los valores correspondientes a $P(-4 + 8(i - 1)/n \leq X < -4 + 8i/n)$, siendo X una variable aleatoria que sigue una distribución $N(0, 1)$.

Gráficamente, los puntos descritos corresponden a las distribuciones que representamos en las siguientes figuras



Además de estos cinco modelos de puntos hemos estudiado puntos intermedios calculados a través de las geodésicas del espacio paramétrico de las distribuciones multinomiales tomando como tensor métrico la matriz de información de Fisher. Así, si $P = (p_1, \dots, p_n; N)$ es el punto inicial y $Q = (q_1, \dots, q_n; N)$ el punto final, los puntos intermedios $R(s) = (r_1(s), \dots, r_n(s); N)$ en el arco de la geodésica que enlaza P y Q los hemos obtenido de considerar las relaciones:

$$r_i(s) = \left[\sqrt{p_i} \cos(s/2\sqrt{N}) + \frac{\sqrt{q_i} - \sqrt{q_i} \cos(d/2\sqrt{N})}{\text{sen}(d/2\sqrt{N})} \cdot \text{sen}(s/2\sqrt{N}) \right]^2$$

$$d = 2\sqrt{N} \arccos \left(\sum_{i=1}^n \sqrt{p_i q_i} \right)$$

siendo d la distancia geodésica entre P y Q y $s \in [0, d]$ la distancia entre P y $R^{(s)}$. En nuestro caso, hemos elegido dos puntos intermedios entre $P^{(0)}$ y los restantes $P^{(i)}$, y también otros dos puntos entre $P^{(2)}$ y $P^{(3)}$.

Para cada modelo de punto y para las distribuciones intermedias hemos tomado el número de sucesos n recorriendo el conjunto $\{6, 20, 100, 400\}$ y para cada valor de n hemos generado $N = \min\{20n, 1200\}$ valores aleatorios para determinar el vector de probabilidades y calcular el valor de H_i . Finalmente, hemos repetido el proceso 5.000 veces y hemos calculado el valor de las estadísticas $\sigma^2(H)$ y V .

Hemos realizado también un seguimiento del proceso para N recorriendo el conjunto $\{\min(n, 50), \min(2n, 200), \min(5n, 400), \min(10n, 800)\}$, calculando como en el caso anterior el valor de los estadísticos $\sigma^2(H)$ y V una vez realizadas 5.000 repeticiones.

En la construcción del algoritmo de programación hemos considerado un generador de variables aleatorias con distribución multinomial siguiendo la metodología descrita por Bratley, Fox y Schrage (1983), en el que se construye un vector de probabilidades acumuladas a partir de las frecuencias relativas de los sucesos, obtenidas éstas a través de un generador congruencial de la distribución uniforme en $[0, 1]$. Para acelerar el proceso de asignación de frecuencias, hemos adoptado el método de Ahrens y Korht, descrito también en Bratley, Fox y Schrage (1983). Para suministrar las semillas aleatorias (valor en el cual se inicia

la secuencia de números aleatorios de cada simulación) de forma independiente, hemos superpuesto al generador congruencial un segundo generador (generador de Tausworthe) que proporciona los 5.000 valores necesarios para efectuar las repeticiones.

8. RESULTADOS OBTENIDOS

Realizadas las simulaciones para $N = \min\{20n, 1200\}$ y analizados conjuntamente los resultados obtenidos para los estadísticos $\sigma^2(H)$ y V hemos obtenido:

- 1) En los puntos $P^{(0)}, P^{(2)}, P^{(3)}, P^{(4)}$ y los intermedios que los enlazan los índices H_7 y H_2 son los que tienen mejor estimación, teniendo en cuenta sobre todo que el valor del estadístico V es muy inferior al de los restantes índices.
- 2) En los puntos mencionados, el estimador del índice H_3 es notablemente inferior a los restantes, así como los de H_9 y H_{10} cuando el número de parámetros es inferior o igual a 100.
- 3) El resto de los índices presentan resultados similares alternándose la mejor eficacia de los estimadores al variar los puntos.
- 4) En la geodésica que enlaza $P^{(0)}$ y $P^{(1)}$ no se observan diferencias significativas entre los estimadores de los índices.

Estos resultados provienen de considerar para cada valor de n los distintos valores que toman los estadísticos $\sigma^2(H)$ y V a lo largo de las geodésicas que enlazan dos o más puntos. A título ilustrativo presentamos a continuación varias gráficas en las que podemos observar la eficacia de los estadísticos al variar el punto. (Figuras 6 a 13.)

En las gráficas, el punto 1^+ indica el salto segundo en la geodésica que une los puntos $P^{(0)}$ y $P^{(1)}$ y el punto 2^+ indica el primer salto en la geodésica que une los puntos $P^{(2)}$ y $P^{(3)}$.

Por último, al realizar el seguimiento del proceso anterior para valores de N variando en el conjunto $\{\min(n, 50), \min(2n, 100), \min(5n, 400), \min(10n, 800)\}$, hemos observado que el orden en los valores de los estadísticos tiende a estacionarse para $N = \min(5n, 400)$.

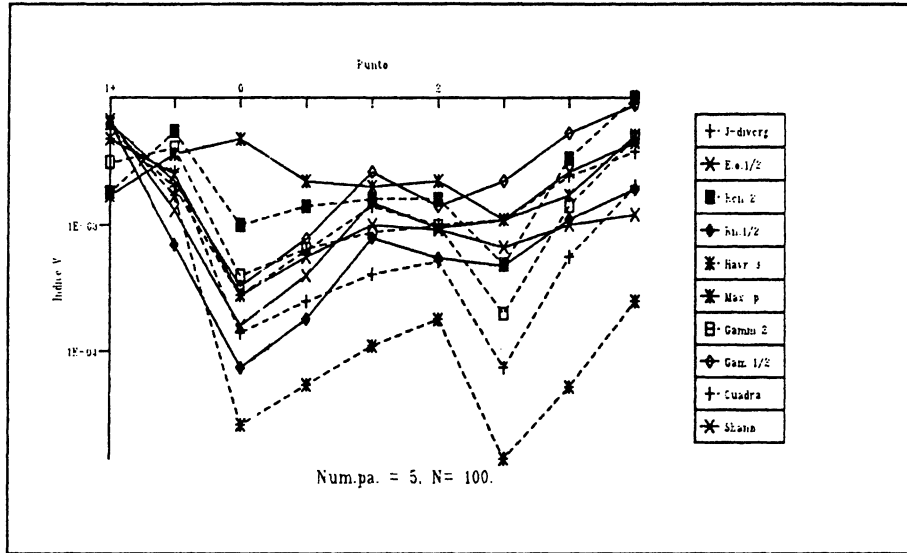


Fig. 6

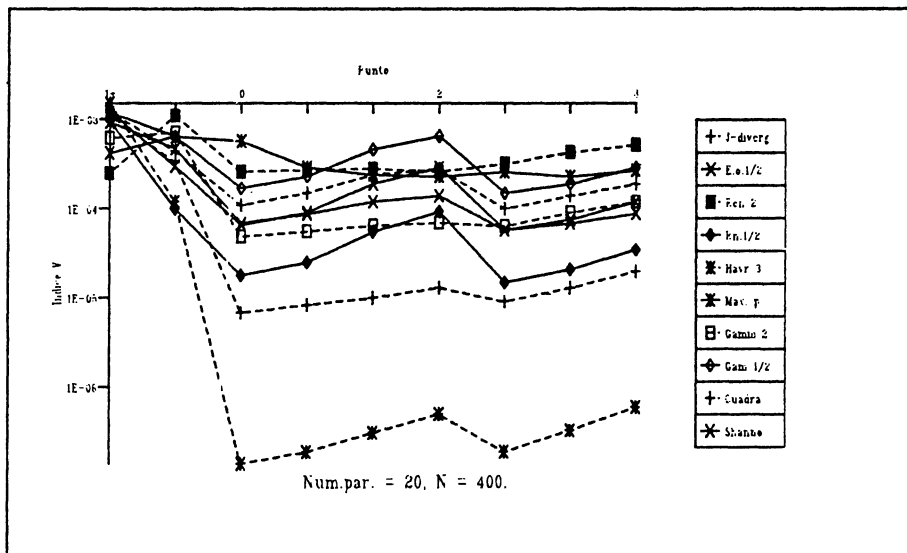


Fig. 7

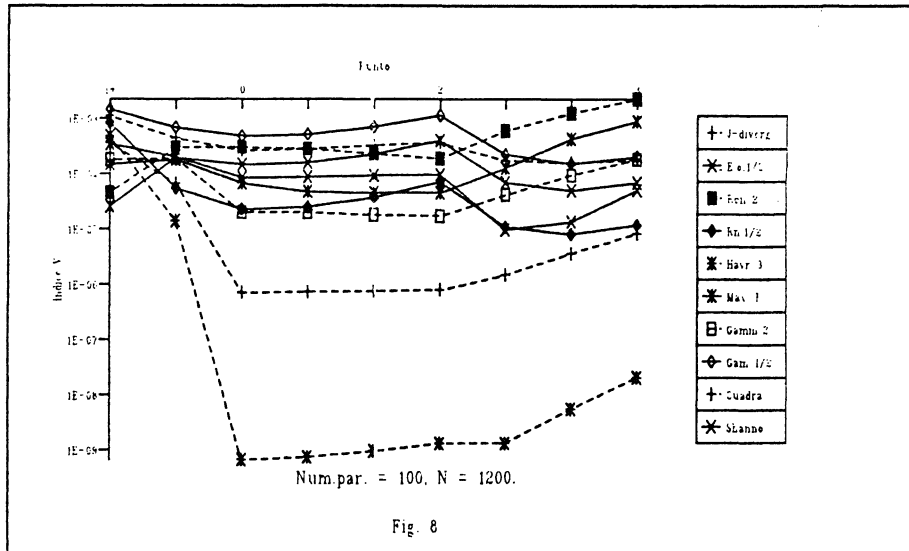


Fig. 8

Fig. 8

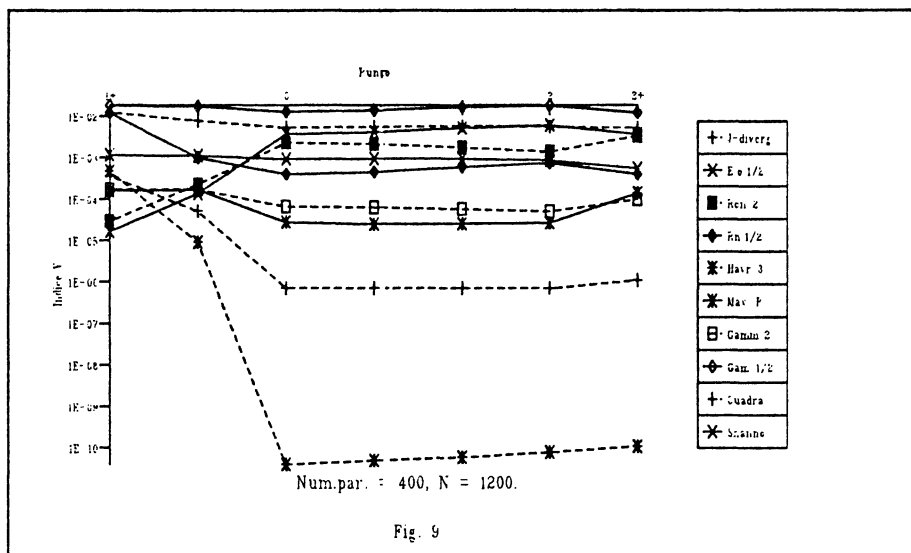


Fig. 9

Fig. 9

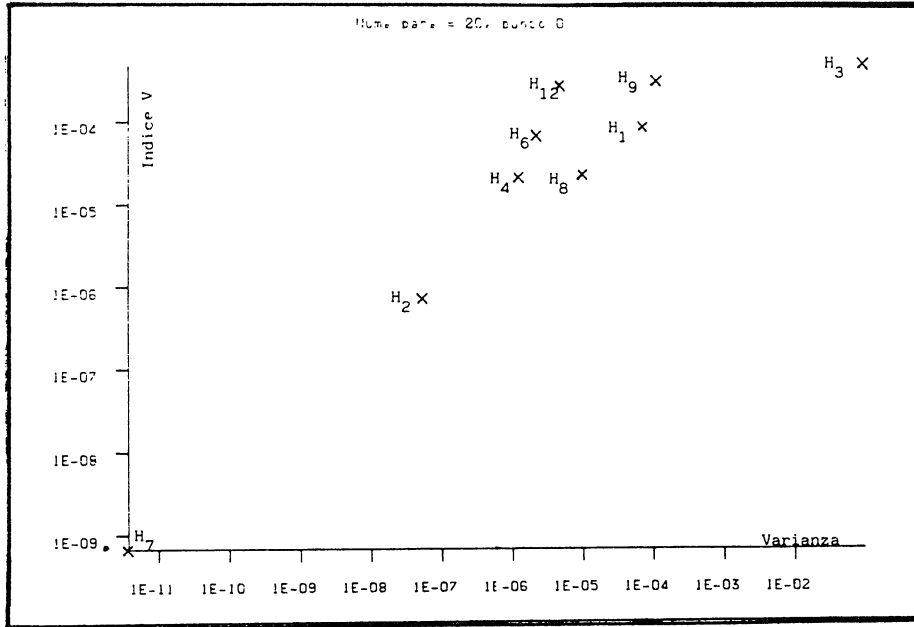


Fig. 10

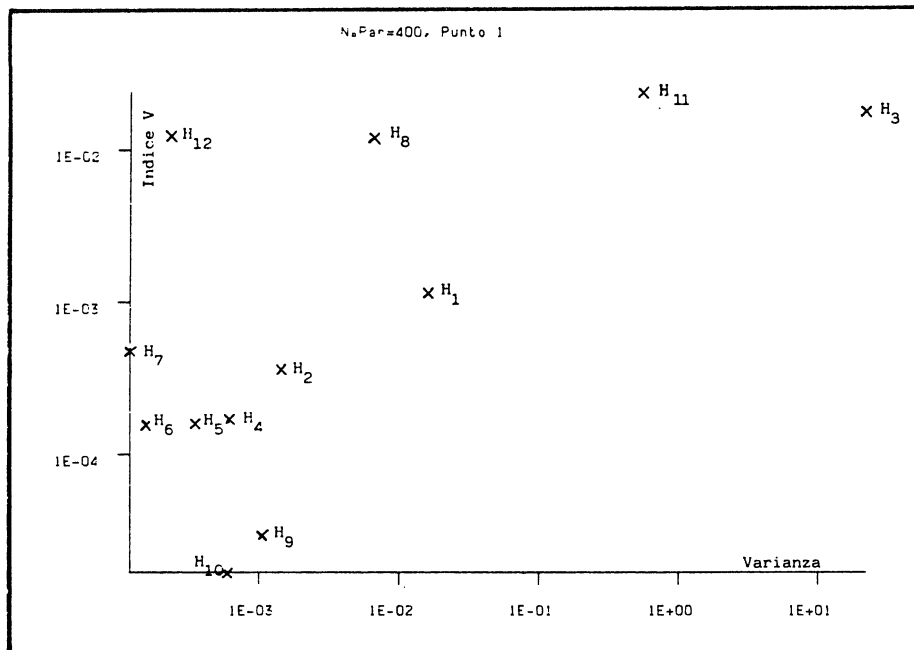


Fig. 11

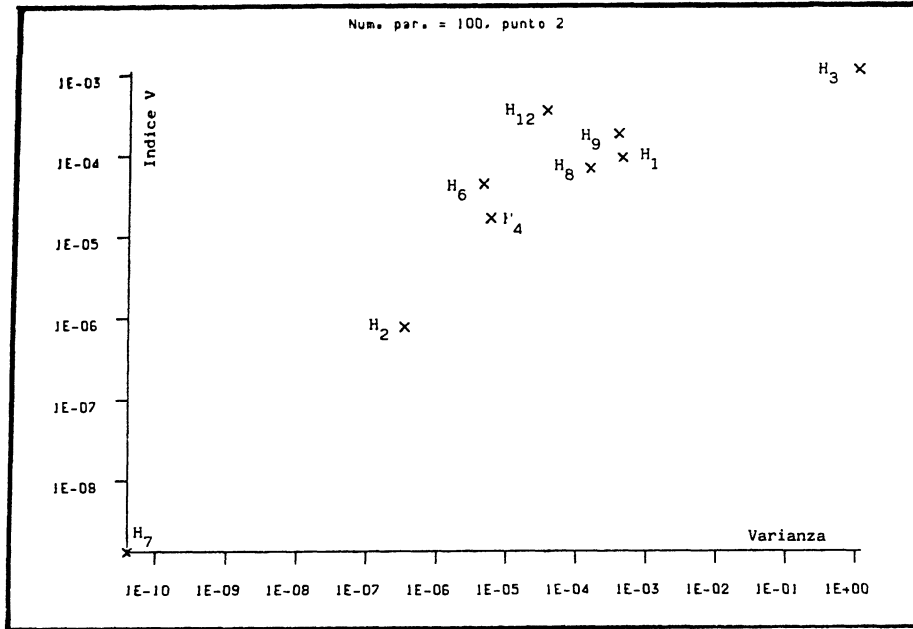


Fig. 12

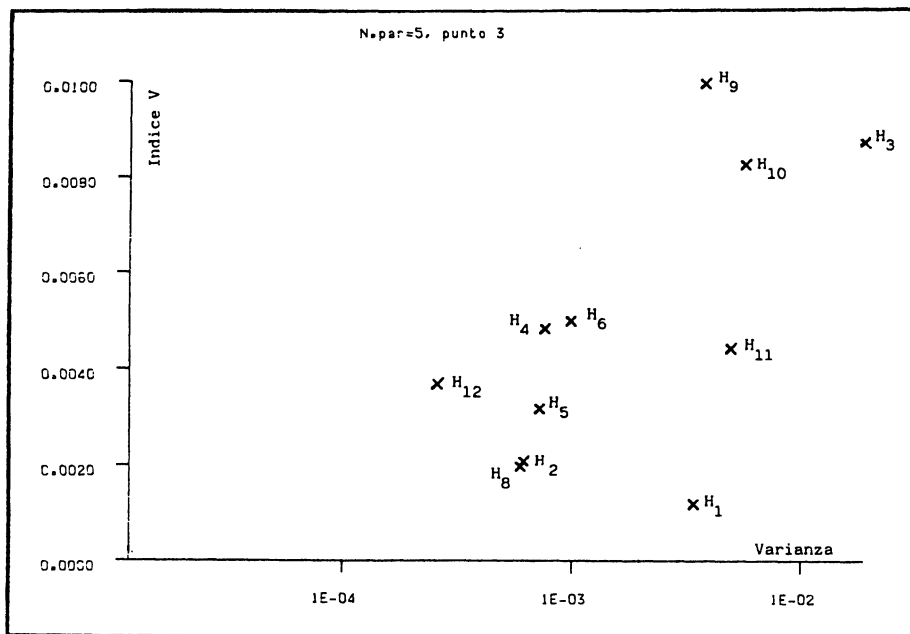


Fig. 13

9. COMPARACION POR EL COEFICIENTE DE CORRELACION DE SPEARMAN

Para complementar el estudio realizado con los estadísticos $\sigma^2(H)$ y V , nos ha parecido conveniente realizar una comparación con los doce índices expuestos a partir de un índice de asociación por rangos. En este sentido hemos calculado el valor del coeficiente de correlación de Spearman entre los valores teóricos y los valores que resultan por simulación, obteniendo por este proceso la tabla 1 que presentaremos a continuación.

La tabla se ha realizado escogiendo al azar 100 y 200 puntos dentro de la familia de las distribuciones multinomiales y considerando para cada uno de ellos un vector acumulativo procedente de 100 extracciones aleatorias a partir de un generador congruencial de la distribución uniforme en $[0, 1]$. Para cada vector acumulativo se ha hallado el valor de la incertidumbre y se ha repetido el proceso 4.000 veces, siendo el valor medio de la incertidumbre de estas 4.000 repeticiones el que hemos correlacionado con el valor teórico. El proceso se ha repetido para valores del parámetro recorriendo el conjunto $\{5, 10, 20\}$.

A la vista de los resultados (ver tabla 1) para el coeficiente de correlación podemos dividir los índices en cuatro grupos:

$$\begin{aligned} G \cdot 1 &= \{H_2, H_4, H_9\} & G \cdot 2 &= \{H_1, H_{12}, H_5, H_7, H_{10}\} \\ G \cdot 3 &= \{H_3, H_8, H_{11}\} & G \cdot 4 &= \{H_6\} \end{aligned}$$

es decir, los índices afectados por potencias de grado dos, agrupados en el bloque primero, tienen mayor coeficiente de correlación, los índices afectados por funciones logarítmicas o potencias de grado 3 agrupados en el bloque segundo, se hallan en una zona intermedia, y los índices afectados por raíces cuadradas, agrupados en el tercer bloque, son los que presentan menor correlación. Finalmente, el índice H_6 se observa totalmente separado con muy baja correlación.

A la vista de los resultados, nos parece interesante profundizar en el estudio de índices de incertidumbre asociado a J -divergencias, ya que, por un lado, la familia de las J -divergencias contiene los índices más usados, y por otro, la familia aporta nuevos índices con propiedades de convexidad y de estimación comparables a los comúnmente utilizados.

TABLA 1

		VALOR DEL PARAMETRO			
		5	10	20	
I N D I C E S	H ₁	p = 100	0,7104	0,7451	0,2252
		p = 200	0,7810	0,5558	0,3216
	H ₂	p = 100	0,9364	0,8273	0,5775
		p = 200	0,8328	0,6675	0,3602
	H ₃	p = 100	0,6357	0,2599	0,1480
		p = 200	0,5243	0,2767	0,0583
	H ₄	p = 100	0,9853	0,8207	0,5473
		p = 200	0,8525	0,6571	0,3787
	H ₅	p = 100	0,8260	0,4048	0,4928
		p = 200	0,6392	0,3687	0,2585
	H ₆	p = 100	0,4529	0,2087	0,0425
		p = 200	0,1838	0,0282	0,1420
H ₇	p = 100	0,8375	0,5696	0,6055	
	p = 200	0,7910	0,5024	0,3075	
H ₈	p = 100	0,5851	0,3044	0,1708	
	p = 200	0,5245	0,3080	0,0683	
H ₉	p = 100	0,9566	0,9141	0,6549	
	p = 200	0,8059	0,6692	0,3991	
H ₁₀	p = 100	0,7318	0,2875	0,4009	
	p = 200	0,5094	0,2636	0,1939	
H ₁₁	p = 100	0,5873	0,2961	0,1715	
	p = 200	0,5182	0,2826	0,0831	
H ₁₂	p = 100	0,8906	0,7371	0,2745	
	p = 200	0,7655	0,3037	0,1950	

BIBLIOGRAFIA

ACZEL, J., y DARACZY, Z. (1975): «On Measures of Information and their characterizations», *Academic Press*, New York.

BOEKKEE, D. E., y VAN DER LUBBE (1980): «The R-Norm Information Measure», *Information and Control*, V.45,136-155.

- BOURGUIGNON, F. (1979): «Descomposable income inequality measures», *Econometrika*, V.47, 4,901-920.
- BRATLEY, P.; FOX, B. L.; SCHRAGE, L. E. (1983): *A Guide to Simulation*, Springer-Verlag, New York.
- DE GROOT, M. H. (1970): «Optimal Statistical Decisions», New York, McGraw.
- HAVDRA, J., CHARVAT, F. (1967): «Quantification method of classification processes. Concept of structural α -entropy», *Cybernetica*, 3,20-25.
- KRUSKAL, W. H. (1956): «Ordinal measures of association», *Journal of the Statistical Association*, 53,814-861.
- LIEBERSON, S. (1969): «Measuring population diversity», *Amer. Soc. Rev.*, 34,840-862.
- LINDLEY, D. V. (1956): «On a Measure of the Information provided by an experiment», *Annals of Mathematical Statistics*, 27,986-1005.
- MARSHALL, A., y OLKIN I. (1979): *Inequalities: Theory of majorization and its applications*, Academic Press, New York.
- RENYI, A. (1961): «On Measures of Entropy and Information», Proc. 4th Berkeley Symp. Math. Statistics and Probability, 1,547-561.
- ROCKAFELLAR, R. T. (1970): *Convex Analysis*, Princeton, New Jersey.
- SALICRU M. (1987): *Medidas de divergencia en análisis de datos*, Tesis Doctoral.
- THEODORESCU, A. (1977): «Energie Informationelle et notions apparentees», *Trabajos de estadística e I.O.*, 28, n.º 2 y 3.
- WAYNE, A., y VARBERG, D. E. (1973): *Convex functions*, Academic Press, New York.

ANEXO I: FUNCIONES SCHUR-CONVEXAS

Debido al uso poco frecuente de las funciones Schur-convexas, presentamos seguidamente algunos resultados de intereses para la realización de este trabajo.

Definición A.1

Para un intervalo I de R , se dice que una función $f: I^n \rightarrow R$ ($n > 1$) es Schur-convexa si se cumple la condición $f(Sx) \leq f(x)$ para todo elemento x de I^n y para cada matriz estocástica S de orden n .

De consideraciones sobre matrices estocásticas, se obtienen los siguientes resultados.

Proposición A.1

Una función convexa $f: I^n \rightarrow R$ es Schur-convexa si y sólo si es simétrica en los argumentos.

Proposición A.2

Si $f(x) = f(x_1, \dots, x_n)$ es una función simétrica de los argumentos y tiene derivadas parciales continuas en I^n , entonces la función $f: I^n \rightarrow R$ es Schur-convexa si y sólo si se cumple la condición

$$(x_i - x_j) \left(\frac{\partial f}{\partial x_i} - \frac{\partial f}{\partial x_j} \right) \geq 0 \quad \text{en } I^n$$

Para mayor información sobre el tema puede verse Wayne A. y Dale E. (1973) o Marshall A. y Olkin I. (1979).