

Medidas de divergencia en análisis de datos

Miguel Salicrú Pagés

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tesisenxarxa.net) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tesisenred.net) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tesisenxarxa.net) service has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized neither its spreading and availability from a site foreign to the TDX service. Introducing its content in a window or frame foreign to the TDX service is not authorized (framing). This rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

MEDIDAS DE DIVERGENCIA EN ANALISIS DE DATOS

Memoria presentada para
optar al grado de doctor
en Matemáticas por la
Universidad de Barcelona
por

Miquel Salicrú Pagés

Vº Bº
EL DIRECTOR.

Dr. Carles M^a Cuadras Avellana
Catedrático de Bicestadística
Departamento de Estadística
Universidad de Barcelona

Barcelona, 29 de Enero de 1987

A Ma Teresa, mi esposa

AGRADECIMIENTOS

Quiero expresar mi agradecimiento al Dr. C. Cuadras, director de esta memoria, por sus consejos, orientaciones y constante estímulo. A todos mis compañeros del Dpto. de Estadística por su apoyo constante, destacando de forma muy especial al Dr. A. Arcas y M. Calvo por su inestimable e incansable cooperación, y al Dr. J. Ocaña y al Dr. J.M. Oller por la supervisión de los aspectos de la memoria relacionados con sus líneas de trabajo.

Me siento complacido de poder expresar mi gratitud al Dr. J. Cascante y al Dr. J.E. Martínez por sus orientaciones y por el estímulo que siempre me ha proporcionado. Asimismo quiero agradecer a la Dra. P. García Carrasco, al Dr. D. Nualart y al Dr. C. Alsina su interés y sus orientaciones en los aspectos relacionados con medidas de incertidumbre.

Finalmente, quiero dedicar este trabajo a Ma Teresa y Mónica, por ser ellas quienes han dado verdadero sentido a esta memoria.

INDICE

	<u>Página</u>
PROLOGO	5
I. CONSIDERACIONES GENERALES SOBRE LAS MEDIDAS DE DIVERGENCIA ENTRE POBLACIONES	6
1. Introducción	7
2. Medidas de entropía	8
3. Medidas que provienen de diferencias intrínsecas entre individuos	13
4. Medidas que provienen de funciones de entropía .	16
5. Medidas que provienen de consideraciones empíri- cas	18
6. Interés de las divergencias en análisis de datos	21
II. CONVEXIDAD Y SIMETRIA DE LA J-DIVERGENCIA GENERA- LIZADA	22
1. Introducción	23
2. Definiciones previas	23
3. Propiedades previas	24
4. Condiciones para la simetría de la J-divergencia generalizada	26
5. Convexidad de la J-divergencia generalizada res- pecto del parámetro	31
6. Convexidad de la J-divergencia generalizada res- pecto $p(x)$, $q(x)$	34
III. ANALISIS COMPARATIVO ENTRE DIVERGENCIAS	39
1. Introducción	40

	<u>Página</u>
2. Definiciones previas	40
3. Relaciones de igualdad entre divergencias	42
4. Medidas logarítmicas asociadas a divergencias ..	44
5. Relaciones de desigualdad entre divergencias ...	46
6. La L-divergencia como medida de Csiszar	52
7. Relación de la medida de Csiszar con las entropías	55
8. Convexidad y no negatividad de las divergencias.	57
 IV. METRICAS RIEMANIANAS ASOCIADAS A DIVERGENCIAS PARA FAMILIAS DE DISTRIBUCIONES PARAMETRICAS	 62
1. Introducción	63
2. Definiciones previas	63
3. Condiciones generales para una distancia	64
4. Métrica Riemana asociada a la J-divergencia generalizada	66
5. Métricas Riemana asociadas a la K, L y M divergencias	69
6. Métricas diferenciales invariantes frente a transformaciones no singulares de variables aleatorias	72
7. Invariancia para el funcional ϕ -entropía	77
 V. MEDIDAS DE INCERTIDUMBRE ASOCIADAS A J-DIVERGENCIAS	 78
1. Introducción	79
2. Definiciones previas	79
3. Condiciones para la constante k	80
4. Condiciones para la función ϕ	82

	<u>Página</u>
5. La entropía de Shannon como medida de incertidumbre asociada a J-divergencias	84
6. La entropía de Havrda-Charvat como medida de incertidumbre asociada a J-divergencias	92
7. Entropías adicionales asociadas a J-divergencias	93
8. Los funcionales ϕ -entropía como medidas de incertidumbre asociadas a J-divergencias	95
 VI. CARACTERISTICAS MUESTRALES DEL INDICE DE INCERTIDUMBRE ASOCIADO A LA J-DIVERGENCIA CORRESPONDIENTE A LA FUNCION $\phi(t) = t \log t$	 99
1. Introducción	100
2. Indices a comparar	100
3. Metodología empleada	102
4. Resultados obtenidos	106
5. Nota complementaria	112
ANEXO I: Programa índices σ y v	115
ANEXO II: Programa índices correlación Spearman ...	129
 VII. RESUMEN DE RESULTADOS	 139
 BIBLIOGRAFIA	 144

PROLOGO

Esta memoria es una aportación al estudio de medidas que generalizan los índices de distanciaci3n entre poblaciones com3nmente utilizadas en an3lisis de datos. La idea y posterior desarrollo de la memoria va intimamente ligado a los trabajos realizados por Burbea y Rao (1982) y a las sucesivas visitas del Dr. Burbea a la antigua unidad docente de Bioestadística de la Facultad de Biología.

En la primera parte (cap. 1, 2 y 3), se presentan las distintas divergencias, se estudian inter-relaciones y se analiza la convexidad.

En la segunda parte (cap. 4) se estudian las métricas diferenciales asociadas a divergencias invariantes frente a cambios no singulares de parámetros y variables aleatorias.

En la tercera parte (cap. 5 y 6) se analizan las relaciones entre la J-divergencia y las entropías com3nmente utilizadas.

Finalmente, en los anexos I y II se presentan los programas utilizados en el cap. 6.

Enero, 87

CAPITULO I

CONSIDERACIONES GENERALES SOBRE LAS MEDIDAS DE DIVERGENCIA ENTRE POBLACIONES

RESUMEN

En este capítulo se presentan distintas formas de construcción de medidas de distanciaci3n entre poblaciones, y se observa la necesidad del estudio de medidas que generalizen las distancias com3nmente utilizadas en an3lisis de datos.

Planteamos en primer lugar las distancias entre poblaciones que provienen de distancias entre los individuos, determinamos en segundo lugar las que provienen de funciones de entropía, y finalmente, presentamos distintas formas de distanciaci3n que provienen de consideraciones empíricas.

1. INTRODUCCION

Un problema interesante que se plantea en análisis de datos es la elección de medidas adecuadas que permitan cuantificar las analogías y diferencias entre poblaciones o entre individuos de una misma población. En este sentido gran variedad de índices de similaridad o disimilaridad se han planteado en los numerosos trabajos relacionados con la representación o clasificación de poblaciones.

De forma general, podemos decir que los índices de similaridad entre poblaciones pueden basarse en: 1) las diferencias intrínsecas entre los individuos de cada población, tal como se observa en los trabajos de Simpson (1949), Nei (1978) o Agresti (1978); 2) consideraciones relacionadas con funciones de entropía, siendo interesante entre otros los trabajos de Shannon (1948), Renyi (1961), Havrda y Charvat (1967), Mathay y Rathie (1974) o Burbea y Rao (1982); 3) consideraciones empíricas alejadas de la naturaleza del problema, pudiéndose relacionar entre otros los trabajos de Rao (1945), Wald (1950), Brattacharyya (1946), Jeffreys (1948), Kullback y Leibler (1951) o Matusita (1964).

En la mayoría de los casos, los índices que provienen de las diferencias intrínsecas entre los individuos, suelen ajustarse bien a las condiciones experimentales, pero al carecer, por lo general, de buenas propiedades matemáticas, resulta difícil la representación geométrica de las poblaciones a través de la distancia obtenida. Por el contrario, los índices

que provienen de funciones de entropía tienen atractivas propiedades teóricas de difícil interpretación. Finalmente, los índices que provienen de consideraciones empíricas no suelen ajustarse excesivamente a las condiciones experimentales, pues por lo general, son ajenos a la naturaleza del problema, si bien al haber sido introducidos directamente como índices de distanciaci3n, suelen reflejar con mayor claridad las condiciones experimentales que los índices asociados a entropías.

En este capítulo introductorio de la memoria presentaremos las medidas de dispersi3n y distanciaci3n más usadas en análisis de datos y expondremos las propiedades más generales de cada medida.

2. MEDIDAS DE ENTROPIA

En muchos trabajos experimentales en que las poblaciones han sido determinadas por características cualitativas, resulta de gran interés la obtenci3n de medidas cuantitativas que reflejen las diferencias. Por lo general, se han considerado categorías disjuntas y se han obtenido las frecuencias relativas de aparici3n, de forma que el problema se ha reducido a determinar medidas en el conjunto de distribuciones de probabilidad multinomiales. En este sentido, las medidas de entropía que analizaremos seguidamente proporcionan un índice de la informaci3n contenida en una distribuci3n.

2.1. DEFINICIONES PREVIAS

Dado el conjunto $S^n = \{(p_1, \dots, p_n) \in \mathbb{R}^n \text{ t.g. } \sum_{i=1}^n p_i = 1, p_i \geq 0 \forall_i\}$, se dice que una función H de S^n a valores reales es una función de entropía si cumple las condiciones:

E.1: $H(p) = 0$ si y solo si todas las componentes de p , $p = (p_1, \dots, p_n)$ son todas cero, salvo una de ellas.

E.2: $H(p)$ es una función cóncava de S^n

En trabajos experimentales, las medidas de entropía más usadas son las que relacionamos a continuación:

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

$$H_C = \frac{1}{\alpha-1} \left(1 - \sum_{i=1}^n p_i^\alpha \right) \quad \alpha > 0, \alpha \neq 1 \quad (\text{Havrda-Charvat})$$

$$H_R = \frac{1}{1-\alpha} \log \sum_{i=1}^n p_i^\alpha \quad \alpha > 0, \alpha \neq 1 \quad (\text{Renyi})$$

$$H_\gamma = \frac{1}{1-2^{\gamma-1}} \left(1 - \left(\sum_{i=1}^n p_i^{1/\gamma} \right)^\gamma \right) \quad \gamma > 0, \gamma \neq 1 \quad (\gamma\text{-entropía})$$

$$H_p = - \sum_{i=1}^n p_i \log p_i - \sum_{i=1}^n (1-p_i) \log (1-p_i) \quad (\text{Shannon modificada})$$

$$H_L = 1 - \sum_{i=1}^n p_i^2 - \sum_{i=1}^n p_i^2 (1-p_i)^2 \quad (\text{Latter})$$

A título ilustrativo de la utilidad de estas medidas, Patil y Taille (1972) y Pielou (1975) interpretan estos indi-

ces desde una perspectiva ecológica, Lewontin (1972) usa la entropía de Shannon en estudios de biología, Agresti y Agresti (1978) usan el índice de Havrda-Charvat de grado 2 en estudios sociológicos, al igual que Lieberson (1969) en estudios de economía o Greenberg (1956) en lingüística.

2.2. PROPIEDADES E INTERRELACIONES

H_C , H_R y H_Y no están definidas para $\alpha=1$ ó $\gamma=1$, pero aproximando por límites se comprueba

Proposición 2.2.1.

$$(1) \lim_{\alpha \rightarrow 1} H_C = \lim_{\alpha \rightarrow 1} H_R = H_S \qquad (2) \lim_{\gamma \rightarrow 1} H_Y = \frac{1}{\log 2} H_S$$

Hallando máximos y mínimos condicionados, se obtiene para todas las medidas el siguiente resultado.

Proposición 2.2.2.

Los seis índices son simétricos, no negativos, tienen el máximo para la distribución en la que todos los sucesos son equiprobables y tienen mínimos en las distribuciones degeneradas.

También, sin excesivas dificultades, puede observarse:

Proposición 2.2.3.

Fijada una distribución $p = (p_1, \dots, p_n)$, H_C y H_R son funciones decrecientes en α y H_γ es creciente en γ .

De considerar la desigualdad $(1-x)(x-1-\log x) \geq 0$, se tiene de forma inmediata

Proposición 2.2.4.

$$(\alpha-1) H_S \geq (\alpha-1) H_R \geq (\alpha-1) H_C \text{ para todo } \alpha \in \mathbb{R}^+$$

De observar que H_R , H_C y H_γ son todas funciones de $\sum_{i=1}^n p_i^\alpha$, y de manipulaciones calculísticas sencillas se obtiene

Proposición 2.2.5.

Fijado un α y fijadas dos distribuciones de probabilidad, $p = (p_1, \dots, p_n)$ y $q = (q_1, \dots, q_n)$, se tiene

$$H_R(p) > H_R(q) \iff H_C(p) > H_C(q) \iff H_\gamma(p) > H_\gamma(q)$$

Finalmente, vale la pena mencionar que los seis índices presentados son Schur-cóncavos en S^k , ya que son simétricos y cóncavos.

2.3. ENTROPIA CUADRÁTICA

Rao (1982) introduce una medida de entropía que tiene en cuenta la distancia entre los sucesos disjuntos que determinan la distribución multinomial. En este sentido, dados A_1, \dots, A_n sucesos mutuamente excluyentes de probabilidades p_1, \dots, p_n , y dada la matriz $D = (d_{ij})$, en la que d_{ij} representa la distancia entre la categoría i y la categoría j , se define la entropía cuadrática como

$$H_D(p) = \sum_{i=1}^n \sum_{j=1}^n d_{ij} p_i p_j = p \cdot D \cdot p'$$

Rao (1982) demuestra que $H_D(p)$ cumple la condición E.1 si y solo si $d_{11} = \dots = d_{nn} = 0$, y cumple la condición E.2 si la matriz

$$D^* = (d_{ij}^*) = (d_{in} + d_{jn} - d_{ij} - d_{nn})$$

es semidefinida positiva.

Además, si se toma la matriz $D = (d_{ij})$ de forma que $d_{ij} = 1 - \delta_{ij}$ siendo δ_{ij} la δ de Kronecker entonces

$$H_D(p) = 1 - \sum_{i=1}^n p_i^2 \quad (H_C \text{ para } \alpha = 2)$$

También, combinando en la matriz D los valores 0, 1 y 2, se obtiene H_L .

Una caracterización de la entropía cuadrática a partir de entropías completamente cóncavas, la podemos encontrar en Ka-Sing-Lau (1985).

3. MEDIDAS QUE PROVIENEN DE DIFERENCIAS INTRINSECAS ENTRE INDIVIDUOS

Una forma de distanciar poblaciones muy usada en trabajos aplicados, consiste en determinar la distancia media entre los individuos de cada población y aplicar la diferencia de Jensen a la distancia media entre individuos pertenecientes a poblaciones distintas. Es decir, se parte de una medida de distanciaci3n entre individuos $d(x,y)$, que es simétrica, no negativa y depende de la naturaleza práctica del problema a estudiar y se calcula el valor de los índices:

$$H_i = \int d(x,y) d F_i(x) d F_i(y)$$

$$H_j = \int d(x,y) d F_j(x) d F_j(y)$$

$$H_{ij} = \int d(x,y) d F_i(x) d F_j(x)$$

donde $F_k(x)$ es la distribución correspondiente a la población k , con $k \in \{i,j\}$.

Finalmente, se calcula la distancia entre la población π_i y π_j a partir de la diferencia de Jensen:

$$D_{ij} = H_{ij} - \frac{1}{2} (H_i + H_j)$$

Resulta inmediato comprobar que $D_{ij} \geq 0$ si y solo si H es c3n cava.

3.1. CASOS PARTICULARES MAS SIGNIFICATIVOS

- a) Cuando las poblaciones π_i vienen determinadas por vectores aleatorios n-dimensionales $X \sim (\mu, \Sigma)$, y se toma por distancia entre individuos la distancia

$$d(x,y) = (x-y)' \Sigma^{-1} (x-y)$$

entonces, la distancia D_{ij} se reduce a la distancia Mahalanobis

$$D_{ij} = (\mu_i - \mu_j)' \Sigma^{-1} (\mu_i - \mu_j)$$

- b) Cuando las poblaciones π_i vienen determinadas por vectores aleatorios n-dimensionales $X = (X_1, \dots, X_r)$, de forma que cada variable X_r sólo puede tomar un número finito de valores con probabilidad no nula, determinados por $P(x_{rs}) = P_{irs}$ con $s \in \{1, \dots, k_r\}$, y la distancia entre individuos se ha definido como

$$d(x,y) = m - \sum_{r=1}^m \delta_r \text{ con } \delta_r = \begin{cases} 1 & \text{si } X_r(x) = X_r(y) \\ 0 & \text{si } X_r(x) \neq X_r(y) \end{cases}$$

entonces la distancia D_{ij} se reduce a la distancia de Nei

$$D_{ij} = \frac{1}{2} \sum_{r=1}^m \sum_{s=1}^{k_r} (P_{irs} - P_{jrs})^2$$

además, cuando X es unidimensional, H_i se reduce a la entropía de Havrda-Charvat de grado dos (índice de Gini-Simpson)

$$H_i = 1 - \sum_{s=1}^k p_{is}^2$$

y la distancia D_{ij} se reduce a la distancia euclídea módulo dos.

3.2. DESCOMPOSICION DE LA DIVERGENCIA

Rao (1962, 71, 82) propone la descomposición de la distancia inicial entre individuos en una suma de distancias. En este sentido toma

$$d(x,y) = \hat{d}_1(x,y) + \dots + \hat{d}_s(x,y)$$

y obtiene de forma inmediata por la linealidad de la integral

$$1) H_i = H_i^1 + \dots + H_i^s \qquad 2) H_{ij} = H_{ij}^1 + \dots + H_{ij}^s$$

$$3) D_{ij} = D_{ij}^1 + \dots + D_{ij}^s$$

El interés de esta descomposición se observa al aplicarlo en los casos particulares anteriores, pues en ellos se obtiene:

a) Cuando cada población π_i viene determinada por

$\chi \sim (\mu, \Sigma)$, los valores propios de Σ son $\theta_1 \geq \dots \geq \theta_m$ y

L_1, \dots, L_m son sus vectores propios normalizados, entonces

la distancia D_{ij} definida a partir de la expresión

$\hat{d}(x,y) = (x-y) \Sigma^{-1} (x-y)$ se reduce a

$$D_{ij} = \sum_{k=1}^m \frac{1}{\theta_k} (L'_k \begin{pmatrix} \mu_i \\ \mu_j \end{pmatrix})^2$$

y cuando se toman m vectores arbitrarios $\{v_1, \dots, v_m\}$ cumpliendo:

$$1) v_i' \Sigma v_j = \delta_{ij} \quad 2) v_1 = (\sigma' \Sigma^{-1} \sigma)^{-1/2} \Sigma^{-1} \sigma$$

donde σ es el vector de desviaciones standard, entonces

$$D_{ij} = (v_1' \begin{pmatrix} \mu_i \\ \mu_j \end{pmatrix})^2 + \dots + (v_m' \begin{pmatrix} \mu_i \\ \mu_j \end{pmatrix})^2 = D_{st}^2 + D_{sf}^2$$

donde $D_{st}^2 = (v_1' \begin{pmatrix} \mu_i \\ \mu_j \end{pmatrix})^2$ representa la componente de la distancia Mahalanobis debida al tamaño y D_{sf}^2 representa la componente debida a la forma.

b) Si en lugar de tomar la distancia Mahalanobis, tomamos la distancia euclídea, entonces de $d(x,y) = (x,y)'(x-y)$ resulta

$$1) H_i = 2 \sum_{k=1}^m \theta_k \quad 2) D_{ij} = \sum_{k=1}^m L'_k \begin{pmatrix} \mu_i \\ \mu_j \end{pmatrix}$$

obteniéndose en este caso H_i como la variabilidad total en términos de las componentes principales.

4. MEDIDAS QUE PROVIENEN DE FUNCIONES DE ENTROPIA

En esta sección, plantearemos medidas de distanciamiento concebidas directamente sobre el espacio de las funciones de distribución de probabilidad. Estas medidas resultarán de apli

car las consideraciones de la sección 3, a las funciones de la sección 2. Es decir, consideraremos las distancias

$$D_{ij} = H_{ij} - \lambda_1 H_i - \lambda_2 H_j \quad \text{con} \quad \lambda_2 = 1 - \lambda_1 \quad \text{y} \quad \lambda_1, \lambda_2 \in (0, 1)$$

donde H_k representará la entropía de la población π_k y H_{ij} la entropía de la población $\lambda_1 \pi_i + \lambda_2 \pi_j$.

Al igual que en el caso anterior, se comprueba sin dificultad que D_{ij} es positiva o cero cuando H_k es cóncava.

El estudio general de estas medidas lo abordaremos en sucesivos capítulos al analizar las J-divergencias generalizadas, pero pese a que el estudio lo realizaremos más adelante, nos parece conveniente plantear dos casos particulares de gran importancia en trabajos teóricos y aplicados.

a) Cuando tomamos la entropía de Shannon, entonces

$$D_{ij} = \lambda_1 \sum_{r=1}^n P_{ir} \log \frac{P_{ir}}{\lambda_1 P_{ir} + \lambda_2 P_{jr}} + \\ + \lambda_2 \sum_{r=1}^n P_{jr} \log \frac{P_{jr}}{\lambda_1 P_{ir} + \lambda_2 P_{jr}}$$

que es el radio de información de Jardine-Sibson (1971).

b) Al tomar el índice de Havrda-Charvat de grado 2, entonces

$$D_{ij} = 2 \lambda_1 \lambda_2 \sum_{r=1}^n (P_{ir} - P_{jr})^2$$

que es la distancia euclídea módulo $2 \lambda_1 \lambda_2$.

5. MEDIDAS QUE PROVIENEN DE CONSIDERACIONES EMPIRICAS

En esta sección presentaremos algunas medidas de distancia entre poblaciones, que resultan de consideraciones empíricas sobre las poblaciones. Obviamente, no pretendemos analizar todas las medidas que provienen de consideraciones empíricas, tan solo queremos presentar algunas que utilizaremos en la memoria.

5.1. DISTANCIA GEODESICA (DISTANCIA DE RAO)

Cuando se tienen una familia paramétrica de funciones de densidad de probabilidad $p(x, \theta)$ con $x \in \chi$ espacio medible y $\theta = (\theta_1, \dots, \theta_n) \in S$ subconjunto de R^n , puede definirse la métrica diferencial asociada a la matriz de información de Fisher por el procedimiento siguiente

- 1) Se define el tensor covariante de 2º orden $g_{ij}(\theta)$ como

$$g_{ij}(\theta) = \int_{\chi} \frac{1}{p} \frac{\partial p}{\partial \theta_i} \frac{\partial p}{\partial \theta_j} dx$$

- 2) Se considera el elemento de línea como

$$ds^2 = \sum_{i=1}^n \sum_{j=1}^n g_{ij}(\theta) d\theta_i d\theta_j$$

- 3) Se calcula la distancia entre las poblaciones determinadas por $p(x, \theta_1)$ y $p(x, \theta_2)$ a partir de la expresión

$$S(\theta_1, \theta_2) = \left| \int_{t_1}^{t_2} \left[\sum_{i=1}^n \sum_{j=1}^n g_{ij}(\theta) \frac{d\theta_i}{dt} \frac{d\theta_j}{dt} \right]^{\frac{1}{2}} dt \right|$$

siendo la curva que une los puntos θ_1 y θ_2 la que resulta de las ecuaciones de Euler-Lagrange.

Este método propuesto por Rao (1945) y estudiado posteriormente por Atkinson y Mitchell (1981), Amari (1982), y Cuadras y Oller (1985), es aconsejable en estudios de carácter evolutivo donde los cambios se producen con el paso del estado θ_1 al estado θ_2 .

5.2. DISTANCIA DE WALD

Una alternativa no paramétrica para distanciar dos funciones de densidad de probabilidad, definidas en un espacio medible X , fue propuesta por Wald (1950) a partir de la expresión

$$D(p_i, p_j) = \max_{RCX} \left| \int_R (p_i(x) - p_j(x)) dx \right|$$

Esta distancia que fue introducida por razonamientos de teoría de la decisión, puede expresarse en los siguientes términos:

$$\begin{aligned} D(p_i, p_j) &= \int_{R_1} (p_i(x) - p_j(x)) dx = \frac{1}{2} \int_X |p_i(x) - p_j(x)| dx = \\ &= \int_{R_1} p_i(x) dx + \int_{R_2} p_j(x) dx - 1 \end{aligned}$$

donde $R_1 = \{x \in X \text{ t.q. } p_i(x) \geq p_j(x)\}$ y $R_2 = X - R_1$

Esta medida de distanciaci3n que varía entre 0 y 1, alcanza el m3nimo cuando $p_i(x) = p_j(x)$ y el m3ximo cuando $\text{sop } p_i \cap \text{sop } p_j = \emptyset$. Propiedades y relaciones de esta medida con otras distancias utilizadas en an3lisis de datos, pueden encontrarse en Jeffreys (1948) o Rao (1977).

5.3. INVARIANTES DE JEFFREYS

Para distanciar dos poblaciones determinadas por $p_i(x)$ y $p_j(x)$, Jeffreys (1948) propone las medidas

$$I_n(p_i, p_j) = \int_X \left| \sqrt[n]{p_i(x)} - \sqrt[n]{p_j(x)} \right|^n dx \text{ para } n > 0$$

$$I_0(p_i, p_j) = \int_X (p_i(x) - p_j(x)) \cdot (\log p_i(x) - \log p_j(x)) dx$$

Casos particulares de estas medidas han sido muy utilizadas en an3lisis de datos, siendo muy destacables los 3ndices que provienen de $n=0$, $n=1$ y $n=2$. En este sentido se tiene:

- a) Para $n=0$, I_0 es la suma de los n3meros de informaci3n de Kullback-Leibler correspondientes a $I(p,q)$ e $I(q,p)$
- b) Para $n=1$, I_1 es el doble de la distancia de Wald.
- c) Para $n=2$, I_2 es la expresi3n de la distancia Matusita muy utilizada en problemas de inferencia.

6. INTERES DE LAS DIVERGENCIAS EN ANALISIS DE DATOS

Con los pocos datos presentados en este capítulo, ya se observa la complejidad en la elección de la medida de distancia a utilizar cuando se tienen datos experimentales. Por ello, un problema interesante en análisis de datos es el estudio de medidas que generalicen las presentadas en los apartados anteriores.

En este sentido, Burbea y Rao (1982) presentan las J-divergencias

$$J_{\phi}^{\lambda}(p, q) = H_{\phi}(\lambda p + (1-\lambda)q) - \lambda H_{\phi}(p) - (1-\lambda) H_{\phi}(q)$$

con $H_{\phi}(p) = -\int_{\mathcal{X}} \phi(p) d\mu$, como una generalización de la mayoría de medidas que provienen de consideraciones intrínsecas entre los individuos y de las que provienen de funciones de entropía.

También presentan las K, L y M divergencias

$$K_{\phi}(p, q) = \int_{\mathcal{X}} (p-q) \left(\frac{\phi(p)}{p} - \frac{\phi(q)}{q} \right) d\mu$$

$$L_{\phi}(p, q) = \int_{\mathcal{X}} \left(p\phi\left(\frac{q}{p}\right) + q\phi\left(\frac{p}{q}\right) \right) d\mu$$

$$M_{\phi}^2(p, q) = \int_{\mathcal{X}} \left(\sqrt{\phi(p)} - \sqrt{\phi(q)} \right)^2 d\mu$$

como una generalización de medidas que provienen de consideraciones empíricas.

CAPITULO II

CONVEXIDAD Y SIMETRIA DE LA J-DIVERGENCIA GENERALIZADA

RESUMEN

En este capítulo caracterizamos la convexidad y simetría de la J-divergencia generalizada en los siguientes términos:

- 1) Determinamos la simetría de la J-divergencia generalizada atendiendo a los valores del parámetro y al tipo de función $\phi(t)$ que la determina.
- 2) Analizamos la convexidad de la J-divergencia generalizada respecto del parámetro una vez fijadas las funciones ϕ -integrables que intervienen y relacionamos la derivada de la J-divergencia generalizada respecto del parámetro con la medida de información de Kullback-Leibler.
- 3) Caracterizamos la convexidad de la J-divergencia generalizada respecto de las funciones ϕ -integrables en términos de concavidad y convexidad de las funciones $\phi(t)$ y $(\phi''(t))^{-1}$.

1. INTRODUCCION

Un problema interesante que se plantea en análisis de datos es la elección adecuada de medidas que nos permitan cuantificar la diferencia entre poblaciones o entre individuos de una misma población. En este sentido, las J-divergencias generalizadas pueden entenderse como medidas que cuantifican la diferencia entre distribuciones o poblaciones.

Como una primera aproximación a las divergencias, analizaremos en este capítulo las condiciones para la convexidad y simetría de la J-divergencia generalizada, completando y generalizando los trabajos sobre J-divergencias realizadas por Burbea y Rao (1982).

Estudios en este sentido resultan muy adecuados para abordar aspectos relacionados con la teoría de la información e inferencia estadística tal como podemos ver en Matusita (1955, 1957, 1964), Pituran (1979), Wald (1949) o Shaked (1977), y en otro orden, Burbea y Rao (1982, 1984), clasifican las entropías de grado α atendiendo a la convexidad de la J-divergencia.

Las relaciones entre las J, K, L y M divergencias las abordaremos en capítulos sucesivos al analizar la convexidad de las K, L y M divergencias.

2. DEFINICIONES PREVIAS

Dada una medida positiva μ, σ aditiva y σ finita definida en una σ -álgebra de subconjuntos de un espacio medible X , y

una función a valores reales $\phi(t)$, dos veces diferenciable con continuidad y definida en un intervalo T_ϕ , con $[0,1] \subset T_\phi \subset [0,\infty)$, se define el espacio de las funciones ϕ -integrables como

Definición 2.1.

$$L_\phi^1 = \{p; p: X \rightarrow \mathbb{R} \text{ t.q. } 1) \int_X |p(x)| d\mu(x) < +\infty$$

$$2) p(x) \in T_\phi \text{ para todo } x \in X\}$$

Para diferenciar dos funciones ϕ -integrables $p(x)$ y $q(x)$ definimos la J -divergencia generalizada como

Definición 2.2.

$$J_\phi^\lambda(p, q) = \int_X (\lambda\phi(p) + (1-\lambda)\phi(q) - \phi(\lambda p + (1-\lambda)q)) d\mu(x) \text{ para } \lambda \in (0, 1)$$

Donde si $\mu(x)$ es una medida atómica, la J -divergencia generalizada se reduce a

$$J_\phi^\lambda(p, q) = \sum_{i=1}^{\infty} a_i (\lambda\phi(x_i) + (1-\lambda)\phi(y_i) - \phi(\lambda x_i + (1-\lambda)y_i))$$

3. PROPIEDADES PREVIAS

De forma inmediata sustituyendo en la definición 2.2. se tiene

Proposición 3.1.

$$J_{\phi}^{\lambda}(\Psi(p), \Psi(q)) = \lambda(1-\lambda) \int_X |\Psi(p) - \Psi(q)|^2 d\mu(x) \text{ para } \phi(x) = x^2 - x$$

resultando en particular para $\Psi(p) = \sqrt{p}$

$$J_{\phi}^{\lambda}(\sqrt{p}, \sqrt{q}) = 2\lambda(1-\lambda) \left(1 - \int_X \sqrt{p \cdot q} d\mu(x)\right) = 2\lambda(1-\lambda) (1 - \cos(p, q))$$

con $p(x), q(x)$ funciones de densidad y $\langle p, q \rangle = \int_X \sqrt{p(x) \cdot q(x)} d\mu(x)$.

Así, en este caso, $J_{\phi}^{\lambda}(\sqrt{p}, \sqrt{q})$ se expresa en función de la distancia Hellinger y a su vez de la distancia Bhattacharyya.

Proposición 3.2.

$J_{\phi}^{\lambda}(p, q) \geq 0$ para todo $p(x), q(x)$ de L_{ϕ}^1 si y solo si $\phi(x)$ es convexa en T_{ϕ} .

Demostración

Si $\phi(x)$ no fuera convexa en un punto $x_0 \in T_{\phi}$, entonces existiría otro punto $x_1 \in T_{\phi}$ cumpliendo

$$\lambda\phi(x_0) + (1-\lambda)\phi(x_1) < \phi(\lambda x_0 + (1-\lambda)x_1)$$

y tomando para un $K \subset X$ con $0 \neq \mu(K) < +\infty$

$$p(x) = \begin{cases} x_0 & \text{para } x \in K \\ 0 & \text{para } x \notin K \end{cases} \quad q(x) = \begin{cases} x_1 & \text{para } x \in K \\ 0 & \text{para } x \notin K \end{cases}$$

se tendría $J_{\phi}^{\lambda}(p, q) < 0$.

El recíproco resulta evidente al ser la integral de una función positiva.

4. CONDICIONES PARA LA SIMETRIA DE LA J-DIVERGENCIA GENERALIZADA

En esta sección vamos a desarrollar las condiciones que debemos imponer al parámetro λ para que fijada una función $\phi(t)$, la J-divergencia generalizada resulte simétrica. En este sentido, estudiaremos previamente las funciones que cumplen la condición $\phi(\lambda x + (1-\lambda)y) = \phi(\lambda y + (1-\lambda)x)$ para todo $x, y \in I$, $I = [a, b] \subset \mathbb{R}$ y demostraremos en el anexo del capítulo.

Lema 4.1.

Si tomamos λ fijo, $\lambda \in (0, \frac{1}{2})$ y $\phi: I \rightarrow \mathbb{R}$ una función cumpliendo $\phi(\lambda x + (1-\lambda)y) = \phi(\lambda y + (1-\lambda)x)$ para todo $x, y \in I$, entonces se cumple que $\phi(x) = \text{cte}$ para todo $x \in (a, b)$.

La demostración del lema 4.1., la realizaremos en tres partes distribuidas en los lemas A1, A2 y A3.

Lema A.1

Si a las condiciones del lema 4.1., añadimos la condición

$$\min \{x-a, b-y\} > \frac{\lambda(y-x)}{1-2\lambda} > 0 \text{ para } x, y \in (a, b)$$

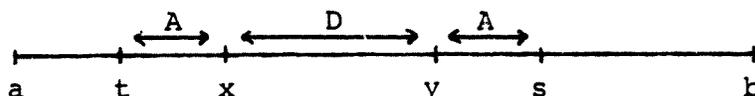
entonces $\phi(t) = \text{cte}$ para $t \in [x, y]$.

Demostración:

De tomar los elementos s y t de I en la forma

$$t = x - \frac{\lambda(y-x)}{1-2\lambda} \quad s = y + \frac{\lambda(y-x)}{1-2\lambda}$$

se tiene el diagrama



$$\text{con } \lambda = \frac{A}{2A+D} \text{ y } 1 - \lambda = \frac{D+A}{2A+D}$$

Así, de la descomposición de x e y en función de s y t en la forma

$$x = \lambda s + (1-\lambda)t \quad y = (1-\lambda)s + \lambda t$$

se obtiene la igualdad $\phi(x) = \phi(y)$ de aplicar la hipótesis del lema.

Además, para todo $t \in (x, y)$ se tiene:

$$\min \{t-a, b-t\} > \frac{\lambda(y-b)}{1-2\lambda}$$

resultando así de la consideración anterior $\phi(y) = \phi(t)$ para todo $t \in (x, y)$.

Lema A.2.

En las condiciones del lema 4.1., $\phi(t) = \text{cte}$ para todo $t \in (a, \frac{1}{2}(a+b)]$

Demostración:

Si y_0 es un valor arbitrario perteneciente al intervalo $(a, \frac{1}{2}(a+b))$ y $\{y_k\}_{k \in \{1, \dots, n\}}$ es la sucesión de valores definida por la expresión:

$$y_k = y_0 + k \frac{1-2\lambda}{1-\lambda} t_0$$

con $y_{n-1} < \frac{1}{2}(a+b) \leq y_n$ y $t_0 = y_0 - a$

entonces se obtiene:

$$\min \{y_j - a, b - y_{j+1}\} > \frac{\lambda}{1-2\lambda} (y_{j+1} - y_j) \text{ para } j \in \{0, \dots, n-1\}$$

de considerar la desigualdad $b-a > 2(y_j - a)$

Así, de aplicar el lema A.1, resulta

$$\phi(t) = \text{cte} \text{ para } t \in (a, y_n]$$

Lema A.3.

En las condiciones del lema 4.1. $\phi(t) = \text{cte}$ para $t \in [\frac{1}{2}(a+b), b)$.

Demstración

De tomar la sucesión $\{x_k\}_{k \in \{1, \dots, n\}}$ de forma simétrica a la considerada en el lema A.2., es decir,

$$x_k = x_0 - k \frac{1-2\lambda}{1-\lambda} t_0 \quad \text{con } t_0 = b - y_0$$

se tiene:

$$\min \{x_{j+1} - a, b - x_j\} < \frac{\lambda}{1-2\lambda} (x_j - x_{j+1})$$

obteniéndose de forma análoga a la anterior

$$\phi(t) = \text{cte} \quad \text{para } t \in [x_n, b)$$

Así, con estos tres lemas hemos demostrado:

Lema 4.1.

Si una función $\phi: I \rightarrow \mathbb{R}$ cumple la condición $\phi(\lambda x + (1-\lambda)y) = \phi(\lambda y + (1-\lambda)x)$ para todo $x, y \in I$ y para un $\lambda \in (0, \frac{1}{2})$, entonces $\phi(x) = \text{cte}$ para todos los valores $x \in (a, b)$.

Proposición 4.1.

- a) $J_\phi^\lambda(p, q)$ es simétrica en p y q para toda $\phi(x)$ de la forma $\phi(x) = ax^2 + bx + c$.
- b) $J_\phi^\lambda(p, q)$ es simétrica en p y q para $\phi(x) = ax^2 + bx + c$ si y solo si $\lambda = \frac{1}{2}$.

Demostración a)

Para $\phi(x) = ax^2 + bx + c$, $J_{\phi}^{\lambda}(p, q) = \int_X a^{\lambda}(1-\lambda)(p^2 + q^2) d\mu(x)$
que es simétrica en p y q .

Demostración b)

De $J_{\phi}^{\lambda}(p, q)$ simétrica para toda $p(x), q(x) \in L_{\phi}^1$ y de tomar

$$p(x) = \begin{cases} s & \text{para } x \in K \\ 0 & \text{para } x \notin K \end{cases} \quad q(x) = \begin{cases} s & \text{para } x \in K \\ 0 & \text{para } x \notin K \end{cases} \quad \text{con } 0 = \mu(K) < +\infty$$

se tendrá para todo $s, t \in T_{\phi}$.

$$\begin{aligned} & (\lambda\phi(s) + (1-\lambda)\phi(t) - \phi(\lambda s + (1-\lambda)t))\mu(K) = \\ & = ((1-\lambda)\phi(s) + \lambda\phi(t) - \phi((1-\lambda)s + \lambda t))\mu(K) \end{aligned}$$

y operando

$$(2\lambda - 1)(\phi(s) - \phi(t)) = \phi(\lambda s + (1-\lambda)t) - \phi((1-\lambda)s + \lambda t) = F(s, t)$$

resultando de derivar respecto a s y t

$$0 = \phi''(\lambda s + (1-\lambda)t)\lambda(1-\lambda) - \phi''((1-\lambda)s + \lambda t)\lambda(1-\lambda) = \frac{\partial^2 F(s, t)}{\partial s \partial t}$$

lo que equivale a

$$\phi''(\lambda s + (1-\lambda)t) = \phi''((1-\lambda)s + \lambda t) \quad \text{para todo } s, t \in T_{\phi}$$

y por el lema 4.1, $\phi''(x) = \text{cte}$ para todo $x \in T_{\phi}$ cuando $\lambda \neq \frac{1}{2}$,
obteniendo así la simetría solo para $\lambda = \frac{1}{2}$.

5. CONVEXIDAD DE LA J-DIVERGENCIA GENERALIZADA RESPECTO DEL PARAMETRO

Prosiguiendo con el estudio de la J-divergencia generalizada atendiendo a los valores que toma el parámetro, analizaremos a continuación las condiciones exigibles a la función $\phi(t)$ para que J-divergencia sea convexa en función del parámetro. Para ello, consideraremos que $J_{\phi}^{\lambda}(p,q)$ es dos veces derivable bajo el signo integral y que el soporte de $\lambda\phi(p)+(1-\lambda)\phi(q) - \phi(\lambda p+(1-\lambda)q)$ no depende del parámetro.

Proposición 5.1.

$J_{\phi}^{\lambda}(p,q)$ es convexa (cóncava) en $(0,1)$ respecto de λ para todas las parejas $p(x), q(x) \in I_{\phi}^1$ fijadas si y solo si $\phi(t)$ es cóncava (convexa) en T_{ϕ} .

Demostración

De $J_{\phi}^{\lambda}(p,q)$ convexa respecto de λ se tiene

$$\int_X \phi''(\lambda p(x)+(1-\lambda)q(x)) (p(x)-q(x))^2 d\mu(x) < 0 \quad (1)$$

Si $\phi''(t)$ fuera positiva en un punto "a" de T_{ϕ} , por la continuidad de $\phi''(t)$ se tendría $\phi''(t) > 0$ para $t \in (a-\delta, a+\delta)$ y tomando

$$p(x) = \begin{cases} \frac{a}{2\lambda} & \text{si } x \in K \\ 0 & \text{si } x \notin K \end{cases} \quad q(x) = \begin{cases} \frac{2a+\delta}{4-4\lambda} & \text{si } x \in K \\ 0 & \text{si } x \notin K \end{cases} \quad \text{con } 0 \neq \mu(K) < +\infty$$

resultaría

$$\int_X \phi''(\lambda p(x) + (1-\lambda)q(x)) (p(x) - q(x))^2 d\mu(x) =$$

$$= \int_X \phi''\left(a + \frac{\delta}{4}\right) \left(\frac{a}{2\lambda} - \frac{2a+\delta}{4-4\lambda}\right)^2 d\mu(x) > 0$$

contradiciendo (1).

El recíproco es evidente, pues de la concavidad de $\phi(t)$ se deduce (1) y a su vez la convexidad de $J_\phi^\lambda(p, q)$.

Proposición 5.2.

$$\left(\frac{\partial}{\partial \lambda} J_\phi^\lambda(p, q)\right)(\lambda) = K(p, \lambda p + (1-\lambda)q) - K(q, \lambda p + (1-\lambda)q)$$

siendo $\phi(t) = t \cdot \log t$, $K(p, q)$ la diferencia de información entre $p(x)$ y $q(x)$ en el sentido de Kullback-Leibler, y de forma que las funciones $p(x)$ y $q(x)$ cumplen la condición $\|\sqrt{p}\| = \|\sqrt{q}\| = 1$ para el producto escalar $\langle p, q \rangle = \int_X \sqrt{p(x) \cdot q(x)} d\mu(x)$

Demostración

$$\left(\frac{\partial}{\partial \lambda} J_\phi^\lambda(p, q)\right)(\lambda) =$$

$$= \int_X [p \cdot \log p - q \cdot \log q - p \cdot \log(\lambda p + (1-\lambda)q) + q \cdot \log(\lambda p + (1-\lambda)q)] d\mu$$

$$= \int_X \left[p \cdot \log \frac{p}{\lambda p + (1-\lambda)q} - q \cdot \log \frac{q}{\lambda p + (1-\lambda)q} \right] d\mu =$$

$$= K(p, \lambda p + (1-\lambda)q) - K(q, \lambda p + (1-\lambda)q)$$

Corolario 5.1.

En las condiciones de la proposición 5.2., se tiene de forma inmediata los resultados siguientes:

$$a) \left(\frac{\partial}{\partial \lambda} J_{\phi}^{\lambda}(p, q) \right) (0) = K(p, q)$$

$$b) \left(\frac{\partial}{\partial \lambda} J_{\phi}^{\lambda}(p, q) \right) (1) = -K(q, p)$$

Corolario 5.2.

En las condiciones de la proposición 5.2., existe un único λ_0 tal que cumple la condición:

$$K(p, \lambda_0 p + (1 - \lambda_0)q) = K(q, \lambda_0 p + (1 - \lambda_0)q)$$

Demostración

Al ser la función $\frac{\partial}{\partial \lambda} J_{\phi}^{\lambda}(p, q)$ una función decreciente en λ , por aplicación de la proposición 5.1. y al tomar distintos signos en los extremos del intervalo $[0, 1]$, resulta inmediata la existencia de un único $\lambda_0 \in (0, 1)$ cumpliendo

$$\left(\frac{\partial}{\partial \lambda} J_{\phi}^{\lambda}(p, q) \right) (\lambda_0) = 0$$

condición que equivale a

$$K(p, \lambda_0 p + (1 - \lambda_0)q) = K(q, \lambda_0 p + (1 - \lambda_0)q)$$

Proposición 5.3.

La simetrización ${}^S J_\phi^\lambda(p, q) = J_\phi^\lambda(p, q) + J_\phi^{1-\lambda}(p, q)$ alcanza su máximo en $\lambda = \frac{1}{2}$ cuando $\phi(t)$ es convexa estricta

Demostración

Al ser la función $\phi(t)$ una función convexa estricta, las funciones $J_\phi^\lambda(p, q)$ y $J_\phi^{1-\lambda}(p, q)$ son cóncavas en $(0, 1)$. Así, la función simetrizada es también cóncava estricta en $(0, 1)$ y el máximo respecto de λ es único.

Por otro lado al cumplirse las condiciones

$$1) \frac{\partial}{\partial \lambda} ({}^S J_\phi^\lambda(p, q)) \left(\frac{1}{2} \right) = 0$$

$$2) \frac{\partial^2}{\partial \lambda^2} ({}^S J_\phi^\lambda(p, q)) \left(\frac{1}{2} \right) < 0$$

el máximo de ${}^S J_\phi^\lambda(p, q)$ se alcanza en $\lambda = \frac{1}{2}$.

6. CONVEXIDAD DE LA J-DIVERGENCIA GENERALIZADA RESPECTO DE $p(x), q(x)$

En esta sección analizaremos las condiciones que debemos imponer a $\phi(t)$ para que $J_\phi^\lambda(p, q)$ sea convexa en función de $p(x)$ y $q(x)$.

Proposición 6.1.

$J_{\phi}^{\lambda}(p, q)$ es convexa (cóncava) en $L_{\phi}^1 \times L_{\phi}^1$ si y solo si $\phi(t)$ es convexa (cóncava) en T_{ϕ} y $(\phi'')^{-1}$ es cóncava (convexa) en T_{ϕ} .

Demostración

Para que $J_{\phi}^{\lambda}(p, q)$ sea convexa en $L_{\phi}^1 \times L_{\phi}^1$, la matriz de la forma cuadrática que define el Hessiano debe ser semidefinida positiva.

En nuestro caso

$$d^2 J_{\phi}((p, q); (f, g), (f, g)) = \int_X (a_1(p, q) f^2 + a_2(p, q) g^2 + 2b(p, q) f \cdot g) d\mu(x)$$

con

$$a_1(p, q) = \lambda \phi''(p) - \lambda^2 \phi''(\lambda p + (1-\lambda)q)$$

$$a_2(p, q) = (1-\lambda) \phi''(q) - (1-\lambda)^2 \phi''(\lambda p + (1-\lambda)q)$$

$$b(p, q) = \lambda(1-\lambda) \phi''(\lambda p + (1-\lambda)q)$$

y

$$M(p, q) = \begin{pmatrix} a_1(p, q) & b(p, q) \\ b(p, q) & a_2(p, q) \end{pmatrix} \text{ semidefinida positiva.}$$

De tomar $p = q$ se tiene

$$0 \leq a_1(p, q) = \lambda \phi''(p) - \lambda^2 \phi''(p) = (\lambda - \lambda^2) \phi''(p(x)) \text{ para todo } p(x) \in L_{\phi}^1$$

resultando así $\phi''(t) \geq 0$ para todo t .

Por otro lado, de $a_1(p,q)a_2(p,q) - (b(p,q))^2 \geq 0$, se obtiene

$$\lambda(1-\lambda)\phi''(p)\phi''(q)\phi''(\lambda p+(1-\lambda)q) \left((\phi'')^{-1}(\lambda p+(1-\lambda)q) - (\phi'')^{-1}(p) - (1-\lambda)(\phi'')^{-1}(q) \right) \geq 0$$

y de la convexidad de $\phi''(t)$ es inmediata la concavidad de $(\phi'')^{-1}$.

Recíprocamente, de la concavidad de $(\phi'')^{-1}$ y la convexidad de ϕ se obtiene de forma inmediata

$$a_1(p,q)a_2(r,q) - (b(p,q))^2 \geq 0$$

y de tomar

$$r_1 = \left(\frac{\lambda}{\phi''(p)}\right)^{\frac{1}{2}}, \quad r_2 = \left(\frac{1-\lambda}{\phi''(q)}\right)^{\frac{1}{2}}, \quad s_1 = (\lambda\phi''(p))^{\frac{1}{2}}, \quad s_2 = 0$$

se tiene

$$r_1^2 + r_2^2 = \frac{\lambda}{\phi''(p)} + \frac{1-\lambda}{\phi''(q)} \leq \frac{1}{\phi''(\lambda p+(1-\lambda)q)}$$

y sustituyendo en la desigualdad de Cauchy

$$\frac{1}{\phi''(\lambda p+(1-\lambda)q)} \cdot \lambda\phi''(p) \geq \lambda^2$$

de donde

$$a_1(p,q) = \lambda\phi''(p) - \lambda^2\phi''(\lambda p+(1-\lambda)q) \geq 0$$

Corolario 6.1.

La divergencia de Jensen definida por

$$J_{\phi}(p, q) = \int_X \left(\frac{1}{2}(\phi(p) + \phi(q)) - \phi\left(\frac{p+q}{2}\right) \right) d\mu$$

es convexa (cóncava) en $L_{\phi}^1 \times L_{\phi}^1$ si y solo si $\phi(t)$ es convexa (cóncava) y $(\phi'')^{-1}$ es cóncava (convexa) en T_{ϕ} .

Corolario 6.2.

$J_{\phi_{\alpha}}^{\lambda}(p, q)$ es convexa en $L_{\phi_{\alpha}}^1 \times L_{\phi_{\alpha}}^1$ si y solo si $\alpha \in [1, 2] \cup \{0\}$ y no es nunca cóncava en $L_{\phi_{\alpha}}^1 \times L_{\phi_{\alpha}}^1$ para

$$\phi_{\alpha}(x) = \begin{cases} (\alpha-1)^{-1} (x^{\alpha}-x) & \text{para } \alpha \neq 1 \\ x \log x & \text{para } \alpha = 1 \end{cases}$$

Proposición 6.2.

Si $\phi(t)$ es cuatro veces diferenciable con continuidad, $J_{\phi}^{\lambda}(p, q)$ es convexa (cóncava) si la matriz

$$A_{\phi}(t) = \begin{pmatrix} \phi''(t) & \sqrt{2} \phi'''(t) \\ \sqrt{2} \phi'''(t) & \phi''''(t) \end{pmatrix}$$

es semidefinida positiva (negativa).

Demostración

$J_{\phi}^{\lambda}(p, q)$ es convexa (cóncava) si y solo si $\phi''(t) \geq 0$ ($\phi''(t) \leq 0$) y

$$((\phi''(t))^{-1})'' = \frac{-\phi(t)\phi''(t) + 2(\phi'''(t))^2}{(\phi''(t))^3} \leq 0 \quad ((\phi''(t))^{-1})'' \geq 0 \quad (2)$$

condición que se cumple cuando la matriz $A_\phi(t)$ es semidefinida positiva (negativa).

De forma inmediata, al integrar la expresión (2) se obtiene el siguiente resultado

Proposición 6.3.

La condición $\det A_\phi(t) = 0$ se cumple para las funciones de la forma $\phi(t) = \frac{1}{a^2} (at+b)\ln(at+b) + ct + d$ y $\phi(t) = et^2 + ht + k$, obteniéndose la convexidad de $J_\phi^\lambda(p, q)$ para $a > 0$ y $b \geq 0$, $\delta \geq 0$.

Observemos que la función $\phi_1(t) = t \log t$ se obtiene para $a=1$, $b=c=d=0$ y la función $\phi_2(t) = t^2 - t$ se obtiene para $e=1$, $h=1$ y $k=0$

CAPITULO III

ANALISIS COMPARATIVO ENTRE DIVERGENCIAS

RESUMEN

En este capítulo realizamos un estudio comparativo de las K, L y M divergencias. En este sentido, el estudio realizado se compone de las siguientes partes:

- 1) Estudio de relaciones de igualdad entre divergencias, atendiendo a las distintas funciones que las determinan.
- 2) Caracterización de relaciones de desigualdad entre la J, K y M divergencia, atendiendo a la convexidad de las funciones que las determinan.
- 3) Extensión de las medidas de divergencia a medidas logarítmicas y obtención de relaciones análogas a las obtenidas entre la entropía de Renyi y Havrda-Charvat.
- 4) Relación entre las medidas de entropía comúnmente utilizadas y las medidas de Csizsar.
- 5) Estudio de la convexidad y no negatividad de las divergencias atendiendo a las funciones que las determinan.

1. INTRODUCCION

Para un estudio global de las medidas de disimilaridad entre poblaciones o grupos de poblaciones, presentamos en este capítulo las K, L y M divergencias que generalizan la mayor parte de las medidas comúnmente utilizadas en estudios relativos a distintos campos: biología, economía, teoría de la comunicación, cibernética,, como podemos ver en Bhattacharyya (1946), Matusita (1957, 64) o Nei (1978).

En este capítulo, partiendo de los trabajos de Burbea y Rao (1982, 84), analizamos, en primer lugar, relaciones de igualdad y desigualdad entre divergencias, atendiendo a las funciones que las determinan. En segundo lugar, extendemos las medidas de divergencia de forma análoga a la realizada por Renyi a partir de la entropía de Havrda-Charvat. En tercer lugar presentamos las medidas de entropía comúnmente utilizadas como medidas de Csiszar entre una distribución y la distribución en la que todos los sucesos son equiprobables. Finalmente, analizamos condiciones para la convexidad y no negatividad de la K, L y M divergencia.

2. DEFINICIONES PREVIAS

Al igual que en el capítulo anterior, consideraremos una medida μ positiva, σ aditiva y σ finita definida en una σ -álgebra de subconjuntos de un espacio medible X , una función a valores reales $\phi(t)$ definida en T_ϕ dos veces diferenciable con

continuidad, y el conjunto de las funciones ϕ -integrables con valores en T_ϕ que denotaremos por L_ϕ^1 .

Para diferenciar a las funciones ϕ -integrables, definiremos en este capítulo tres tipos de divergencia utilizados en numerosos problemas prácticos.

Definición 2.1.

A las expresiones:

$$a) K_\phi(p, q) = \int_X (p-q) \left(\frac{\phi(p)}{p} - \frac{\phi(q)}{q} \right) d\mu(x)$$

$$b) L_\phi(p, q) = \int_X \left(p\phi\left(\frac{q}{p}\right) + q\phi\left(\frac{p}{q}\right) \right) d\mu(x)$$

$$c) M_\phi(p, q) = \left(\int_X (\sqrt{\phi(p)} - \sqrt{\phi(q)})^2 d\mu(x) \right)^{\frac{1}{2}}$$

las llamamos respectivamente las K, L o M divergencia que separa $p(x)$ y $q(x)$.

Es obvio, que al tomar una medida atómica finita, en la que todos los elementos tengan el mismo peso, $p(x)$ y $q(x)$ vendrán representadas por $p = (p_1, \dots, p_n)$ y $q = (q_1, \dots, q_n)$, de forma que las expresiones anteriores se reducirán a

$$a) K_{\phi, n}(p, q) = \sum_{i=1}^n (p_i - q_i) \left(\frac{\phi(p_i)}{p_i} - \frac{\phi(q_i)}{q_i} \right)$$

$$b) L_{\phi, n}(p, q) = \sum_{i=1}^n \left(p_i \phi\left(\frac{q_i}{p_i}\right) + q_i \phi\left(\frac{p_i}{q_i}\right) \right)$$

$$c) M_{\phi, n}(p, q) = \left(\sum_{i=1}^n (\sqrt{\phi(p_i)} - \sqrt{\phi(q_i)})^2 \right)^{\frac{1}{2}}$$

3. RELACIONES DE IGUALDAD ENTRE DIVERGENCIAS

En esta primera sección presentaremos algunos casos particulares, en los que la J, K, L o M divergencia coinciden. En este sentido, las relaciones de igualdad con mayor interés son las que presentamos a continuación.

Proposición 3.1.

$$\text{Para } \phi(p) = p \log p, \quad K_{\phi}(p, q) = L_{\phi}(p, q)$$

es decir, el invariante de Jeffreys lo podemos considerar indistintamente como K o L divergencia.

Proposición 3.2.

$$L_{\psi}(\phi(p), \phi(q)) = M_{\phi}^2(p, q) \quad \text{para } \psi(t) = t - \sqrt{t}$$

Demostración

De las definiciones de L y M divergencia se tienen las siguientes igualdades inmediatas.

$$\begin{aligned} L_{\psi}(\phi(p), \phi(q)) &= \int_X (\phi(p) \left(\frac{\phi(q)}{\phi(p)} - \sqrt{\frac{\phi(q)}{\phi(p)}} \right) + \phi(q) \left(\frac{\phi(p)}{\phi(q)} - \sqrt{\frac{\phi(p)}{\phi(q)}} \right)) d\mu(x) \\ &= \int_X (\sqrt{\phi(p)} - \sqrt{\phi(q)})^2 d\mu(x) = M_{\phi}^2(p, q) \end{aligned}$$

Corolario 3.1.

Para $\psi(t) = t - \sqrt{t}$, $\phi(t) = t$, $\langle p, q \rangle = \int_X \sqrt{p(x) \cdot q(x)} \, d\mu(x)$
y $p(x)$, $q(x)$ funciones de densidad se tiene:

$$L_\psi(p, q) = 4 \operatorname{sen}^2 \frac{\theta}{2} \quad \text{con } \theta = \cos^{-1} \langle p, q \rangle$$

Demostración

Por la proposición 3.2.

$$L_\psi(p, q) = \int_X (\sqrt{p} - \sqrt{q})^2 \, d\mu(x)$$

y operando

$$L_\psi(p, q) = 2 (1 - \int_X \sqrt{p(x) \cdot q(x)} \, d\mu(x)) = 2 (1 - \cos \langle p, q \rangle) = 4 \operatorname{sen}^2 \frac{\theta}{2}$$

Proposición 3.3.

$$K_\eta(\sqrt{\phi(p)}, \sqrt{\phi(q)}) = M_\phi^2(p, q) \quad \text{para } \eta(t) = t^2 - t$$

Demostración

De las definiciones de K y M divergencia se tienen las igualdades

$$\begin{aligned} K_\eta(\sqrt{\phi(p)}, \sqrt{\phi(q)}) &= \int_X (\sqrt{\phi(p)} - \sqrt{\phi(q)}) \left(\frac{\phi(p) - \sqrt{\phi(p)}}{\sqrt{\phi(p)}} - \frac{\phi(q) - \sqrt{\phi(q)}}{\sqrt{\phi(q)}} \right) d\mu(x) \\ &= \int_X (\sqrt{\phi(p)} - \sqrt{\phi(q)})^2 \, d\mu(x) = M_\phi^2(p, q) \end{aligned}$$

Corolario 3.2.

$$L_{\Psi}(p^2, q^2) = \frac{1}{\lambda(1-\lambda)} J_{\phi}^{\lambda}(p, q) = K_{\phi}(p, q) \text{ para}$$

$$\Psi(t) = t - \sqrt{t} \text{ y } \phi(t) = t^2 - t.$$

La demostración es inmediata al considerar las proposiciones 3.2 y 3.3 de este capítulo y la proposición 3.1 del capítulo II.

4. MEDIDAS LOGARITMICAS ASOCIADAS A DIVERGENCIAS

Por analogía a la extensión de la entropía de Havrda-Charvat a la entropía de Rényi podemos considerar para la familia

$\phi_{\alpha}(t) = \frac{1}{\alpha-1} t^{\alpha}$ con $\alpha \in (1, \infty)$ las siguientes medidas:

$$\hat{J}_{\alpha}^{\lambda}(p, q) = \frac{1}{\alpha-1} \log (1+(\alpha-1) J_{\alpha}^{\lambda}(p, q))$$

$$\hat{K}_{\alpha}(p, q) = \frac{1}{\alpha-1} \log (1+(\alpha-1) K_{\alpha}(p, q))$$

$$\hat{L}_{\alpha}(p, q) = \frac{1}{\alpha-1} \log (1+(\alpha-1) L_{\alpha}(p, q))$$

$$\hat{M}_{\alpha}^2(p, q) = \frac{1}{\alpha-1} \log (1+(\alpha-1) M_{\alpha}^2(p, q))$$

Proposición 4.1.

$$a) (\alpha-1) (J_{\alpha}^{\lambda}(p, q) - \hat{J}_{\alpha}^{\lambda}(p, q)) \geq 0$$

$$b) (\alpha-1) (K_{\alpha}(p, q) - \hat{K}_{\alpha}(p, q)) \geq 0$$

$$c) (\alpha-1) (L_{\alpha}(p, q) - \hat{L}_{\alpha}(p, q)) \geq 0$$

$$d) (\alpha-1) (M_{\alpha}^2(p,q) - \hat{M}_{\alpha}^2(p,q)) \geq 0$$

Demostración

Para D_{α} representando J_{α}^{λ} , K_{α} , L_{α} o M_{α}^2 se tienen las igualdades

$$\begin{aligned} (\alpha-1) (D_{\alpha}(p,q) - \hat{D}_{\alpha}(p,q)) &= (\alpha-1) D_{\alpha}(p,q) - \log(1+(\alpha-1)D_{\alpha}(p,q)) = \\ &= x - \log(1+x) \quad \text{para } x = (\alpha-1) D_{\alpha}(p,q) \end{aligned}$$

resultando en todos los casos

$$x = D_{\phi}(p,q) \quad \text{con } \phi(t) = t^{\alpha}$$

y obteniendo en los cuatro casos $x \geq 0$, ya que:

en a), $J_{\phi}^{\lambda}(p,q) \geq 0$ por ser $\phi(t)$ convexa

en b), $K_{\alpha}(p,q) \geq 0$ por ser $\frac{\phi(t)}{t}$ creciente

en c) y d), $L_{\alpha}(p,q) \geq 0$ y $M_{\alpha}^2(p,q) \geq 0$ por ser las integrales funciones positivas.

Así, en los cuatro casos, las expresiones \hat{D}_{α} están bien definidas y las diferencias expresadas por $x - \log(1+x)$ resultan positivas al alcanzar la función $f(t) = t - \log(1+t)$ el mínimo en $t = 0$.

Observemos que si en lugar de tomar la familia $\phi_{\alpha}(t) = \frac{1}{\alpha-1} t^{\alpha}$ tomáramos la familia $\psi_{\alpha}(t) = \frac{1}{\alpha-1} (t^{\alpha} - t)$, entonces se cumplirían las igualdades:

$$J_{\phi_\alpha}^\lambda(p, q) = J_{\psi_\alpha}^\lambda(p, q) \quad \hat{J}_{\phi_\alpha}^\lambda(p, q) = \hat{J}_{\psi_\alpha}^\lambda(p, q)$$

$$K_{\phi_\alpha}(p, q) = K_{\psi_\alpha}(p, q) \quad \hat{K}_{\phi_\alpha}(p, q) = \hat{K}_{\psi_\alpha}(p, q)$$

y al pasar al límite cuando $\alpha \rightarrow 1$, tendríamos la función convexa $\phi(t) = t \log t$, de la que se obtiene de forma inmediata.

Proposición 4.2.

- a) $(\alpha-1) (J_\phi^\lambda(p, q) - \hat{J}_\phi^\lambda(p, q)) \geq 0$
- b) $(\alpha-1) (K_\phi(p, q) - \hat{K}_\phi(p, q)) \geq 0$

Para finalizar con las relaciones de las extensiones, es fácil observar que para la función $\eta_2(t) = -\frac{1}{2} \psi_2(t)$ se tiene la igualdad:

$$K_{\eta_2}(p, q) = \log \cos \theta$$

para $p(x)$ y $q(x)$ funciones de densidad, $\theta = \cos^{-1} \langle p, q \rangle$ y $\langle p, q \rangle = \int_X \sqrt{p(x) \cdot q(x)} \, d\mu(x)$.

5. RELACIONES DE DESIGUALDAD ENTRE DIVERGENCIAS

En esta sección analizaremos algunas relaciones de desigualdad entre la J, K y M divergencia, atendiendo a la convexidad de las funciones que las determinan. En este sentido, hemos obtenido para la función $\psi(t) = \frac{\phi(t)}{t}$ la siguiente igualdad.

Proposición 5.1.

$$\begin{aligned}
& J_{\phi}^{\lambda}(p, q) - \lambda(1-\lambda) K_{\phi}(p, q) = \\
& = \int_X (\lambda p + (1-\lambda)q) (\lambda \Psi(p) + (1-\lambda)\Psi(q) - \Psi(\lambda p + (1-\lambda)q)) d\mu(x)
\end{aligned}$$

Demostración

De las definiciones de J y K divergencia se tienen las siguientes igualdades

$$\begin{aligned}
& J_{\phi}^{\lambda}(p, q) - \int_X (\lambda p + (1-\lambda)q) (\lambda \Psi(p) + (1-\lambda)\Psi(q) - \Psi(\lambda p + (1-\lambda)q)) d\mu(x) = \\
& = \int_X ((\lambda - \lambda^2) \phi(p) + (1-\lambda) \lambda \phi(q) - \lambda(1-\lambda)q \frac{\phi(p)}{p} - \lambda(1-\lambda)p \frac{\phi(q)}{q}) d\mu(x) = \\
& = \lambda(1-\lambda) \int_X (\phi(p) + \phi(q) - p \frac{\phi(q)}{q} - q \frac{\phi(p)}{p}) d\mu(x) = \lambda(1-\lambda) K_{\phi}(p, q).
\end{aligned}$$

De la igualdad obtenida, se deducen de forma inmediata los siguientes resultados

Corolario 5.1.

- a) Cuando $\Psi(t)$ es cóncava, $J_{\phi}^{\lambda}(p, q) - \lambda(1-\lambda) K_{\phi}(p, q) \leq 0$
- b) Cuando $\Psi(t)$ es convexa, $J_{\phi}^{\lambda}(p, q) - \lambda(1-\lambda) K_{\phi}(p, q) \geq 0$

Corolario 5.2.

(α -2) $(J_{\phi}^{\lambda}(p, q) - \lambda(1-\lambda) K_{\phi}(p, q)) \geq 0$ para la familia de funciones

$$\phi_{\alpha}(t) = \begin{cases} \frac{1}{\alpha-1} (t^{\alpha}-t) & \text{con } \alpha \in (0,1) \cup (1,\infty) \\ t \log t & \text{para } \alpha = 1 \end{cases}$$

Demostración

Las funciones $\psi_{\alpha}(t) = \frac{\phi_{\alpha}(t)}{t}$ son convexas para $\alpha - 2 \geq 0$ y cóncavas para $\alpha - 2 \leq 0$. Así

$$\text{para } \alpha - 2 \geq 0, J_{\phi}^{\lambda}(p, q) - \lambda(1-\lambda) K_{\phi}(p, q) \geq 0$$

$$\text{para } \alpha - 2 \leq 0, J_{\phi}^{\lambda}(p, q) - \lambda(1-\lambda) K_{\phi}(p, q) \leq 0$$

de donde resulta

$$(\alpha-2) (J_{\phi}^{\lambda}(p, q) - \lambda(1-\lambda) K_{\phi}(p, q)) \geq 0$$

Corolario 5.3.

$$J_{\phi}^{\lambda}(p, q) = \lambda(1-\lambda) K_{\phi}(p, q) \quad \text{para } \phi(t) = t^2 - t$$

Finalmente, para relacionar la M divergencia con la J divergencia con $\lambda = \frac{1}{2}$, debemos considerar previamente el siguiente resultado

Lema 5.1.

$\phi(t)$ es una función log-convexa (log-cóncava) continua de I en \mathbb{R}^+ si y solo si $(\phi(x) \cdot \phi(y))^{\frac{1}{2}} \geq \phi(\frac{1}{2}(x+y))$ ($(\phi(x) \cdot \phi(y))^{\frac{1}{2}} \leq \phi(\frac{1}{2}(x+y))$) para todo x e y de I .

Demostración

La desigualdad

$$(\phi(x) \cdot \phi(y))^{\frac{1}{2}} \geq \phi\left(\frac{1}{2}(x+y)\right) \text{ para todo } x, y \text{ de } I$$

equivale a

$$\log \phi(x) + \log \phi(y) \geq 2 \log \phi\left(\frac{1}{2}(x+y)\right) \text{ para todo } x, y \text{ de } I$$

y ésta a su vez a la condición de log-convexidad de la función $\phi(t)$ que se expresa como

$$\frac{1}{2} (\log \phi(x) + \log \phi(y)) - \log \phi\left(\frac{1}{2}(x+y)\right) \geq 0 \text{ para todo } x, y \text{ de } I$$

Proposición 5.2.

Si $\phi(t)$ es una función a valores positivos, entonces $\phi(t)$ es log-convexa (log-concava y continua en T_ϕ si y solo si $M_\phi^2(p, q) \leq 2 J_\phi^{\frac{1}{2}}(p, q)$ ($M_\phi^2(p, q) \geq 2 J_\phi^{\frac{1}{2}}(p, q)$) para toda pareja de funciones $p(x)$ y $q(x)$ ϕ -integrables.

Demostración

De la definición del cuadrado de la M-divergencia

$$M_\phi^2(p, q) = 2 \int_X \left(\frac{1}{2}(\phi(p) + \phi(q)) - (\phi(p) \cdot \phi(q))^{\frac{1}{2}} \right) d\mu(x)$$

y de la log-convexidad de la función $\phi(t)$ se tiene

$$M_\phi^2(p, q) \leq 2 \int_X \left(\frac{1}{2}(\phi(p) + \phi(q)) - \phi\left(\frac{1}{2}(p+q)\right) \right) d\mu(x)$$

Así, cuando $\phi(t)$ es log-convexa

$$M_{\phi}^2(p, q) \leq 2 J_{\phi}^{\frac{1}{2}}(p, q)$$

Recíprocamente, de tomar un KCX con $0 \neq \mu(K) < +\infty$, y de considerar las funciones

$$p(x) = \begin{cases} x_0 & \text{cuando } x \in K \\ 0 & \text{cuando } x \notin K \end{cases} \quad q(x) = \begin{cases} x_1 & \text{cuando } x \in K \\ 0 & \text{cuando } x \notin K \end{cases}$$

para todos los valores x_0 y x_1 se tendrá:

$$M_{\phi}^2(p, q) = (\phi(x_0) + \phi(x_1) - 2(\phi(x_0) \cdot \phi(x_1))^{\frac{1}{2}}) d\mu(x)$$

y

$$2J_{\phi}^{\frac{1}{2}}(p, q) = (\phi(x_0) + \phi(x_1) - 2\phi(\frac{1}{2}(x_0 + x_1))) d\mu(x)$$

y de imponer $M_{\phi}^2(p, q) \leq 2J_{\phi}^{\frac{1}{2}}(p, q)$ resultará la condición

$$(\phi(x_0) \cdot \phi(x_1))^{\frac{1}{2}} \geq \phi(\frac{1}{2}(x_0 + x_1))$$

desigualdad que equivale a la log-convexidad de la función $\phi(t)$.

Corolario 5.4.

Las únicas funciones 2-veces diferenciables a valores positivos que cumplen $M_{\phi}^2(p, q) = 2 J_{\phi}^{\frac{1}{2}}(p, q)$ son las funciones de la forma $\phi(t) = e^{ax+b}$.

Demostración

Las únicas funciones que cumplen la condición $(\log \phi(x))'' \geq 0$ son las funciones de la forma $\phi(t) = e^{ax+b}$.

Corolario 5.5.

Si $\phi(t)$ es una función log-convexa a valores positivos y $\psi(t) = \frac{\phi(t)}{t}$ es una función cóncava, entonces

$$K_{\phi}(p,q) \geq 4 J_{\phi}^{\frac{1}{2}}(p,q) \geq 2 M_{\phi}^2(p,q) \geq 0$$

Demostración

De aplicar la proposición 5.2. a la función $\phi(t)$ log-convexa a valores positivos se tiene:

$$4J_{\phi}^{\frac{1}{2}}(p,q) \geq 2 M_{\phi}^2(p,q) \geq 0$$

y de aplicar el corolario 5.1. a $\lambda = \frac{1}{2}$ se tiene:

$$K_{\phi}(p,q) \geq 4 J_{\phi}^{\frac{1}{2}}(p,q)$$

ya que: 1) $\psi(t) = \frac{\phi(t)}{t}$ es una función cóncava

2) $\phi''(t) \geq 0$ por cumplirse $(\log \phi(t))'' \geq 0$ y $\phi(t) \geq 0$

6. LA L-DIVERGENCIA COMO MEDIDA DE CSISZAR

En Csiszár (1963) se definen las medidas de la forma

$$I_f(p, q) = \int_X p \cdot f\left(\frac{q}{p}\right) d\mu(x)$$

para funciones $f(t)$ dos veces diferenciables con continuidad y definidas en \mathbb{R}^+ .

De forma inmediata se tienen los siguientes resultados:

Proposición 6.1.

Si $I_f(p, q)$ es simétrica en p y q , para todas las funciones $p(x)$, $q(x)$ ϕ -integrables, entonces $I_f(p, q) = L_\phi(p, q)$ para $\phi(t) = \frac{1}{2} f(t)$.

Demostración

Si $I_f(p, q)$ es simétrica en p y q , entonces

$$I_f(p, q) = \frac{1}{2} \int_X \left(pf\left(\frac{q}{p}\right) + qf\left(\frac{p}{q}\right) \right) d\mu(x) = L_\phi(p, q)$$

para $\phi(t) = \frac{1}{2} f(t)$

Proposición 6.2.

Toda $L_\phi(p, q)$ se escribe como una medida de Csiszar para $f(t) = \phi(t) + t \phi\left(\frac{1}{t}\right)$.

Demostración

De las igualdades

$$\begin{aligned} L_{\phi}(p,q) &= \int_X (p \phi\left(\frac{q}{p}\right) + q \phi\left(\frac{p}{q}\right)) d\mu(x) = \\ &= \int_X p\left(\phi\left(\frac{q}{p}\right) + \frac{q}{p} \phi\left(\frac{p}{q}\right)\right) d\mu(x) = \int_X p f\left(\frac{q}{p}\right) d\mu(x) \end{aligned}$$

para $f(t) = \phi(t) + t \phi(t^{-1})$, se obtienen las L-divergencias como medidas de Csiszar.

Proposición 6.3.

Si $I_f(p,q)$ es simétrica en p y q , para toda $p(x), q(x)$ de L_{ϕ}^1 , entonces $f(t) = t f(t^{-1})$.

Demostración

Para todo $t \in \mathbb{R}^+$, de considerar las funciones

$$p(x) = \begin{cases} t & \text{cuando } x \in K \\ 0 & \text{cuando } x \notin K \end{cases} \quad q(x) = \begin{cases} 1 & \text{cuando } x \in K \\ 0 & \text{cuando } x \notin K \end{cases}$$

para $K \subset X$ con $0 \neq \mu(K) < +\infty$ se tiene:

$$I_f(p,q) = t \cdot f\left(\frac{1}{t}\right) \cdot \mu(K)$$

y

$$I_f(q,p) = f(t) \mu(K)$$

Así, de la simetría de $I_\phi(p, q)$ se tiene

$$f(t) = t \cdot f(t^{-1})$$

Recíprocamente, si se cumple la condición anterior, entonces se tienen las igualdades

$$I_f(p, q) = \int_X p f\left(\frac{q}{p}\right) d\mu(x) = \int_X q f\left(\frac{p}{q}\right) d\mu(x) = I_f(q, p)$$

Y para finalizar esta sección, presentaremos una serie de medidas de Csizsar muy utilizadas en distintos campos, tales como la genética, psicología o economía.

- 1) Para $f(t) = -\log t$, $I_f(p, q) = \int_X p \log \frac{p}{q} d\mu(x)$ que es la medida de información de Kullback-Leibler.
- 2) Para $f(t) = t \log t$, $I_f(p, q) = \int_X q \log \frac{q}{p} d\mu(x)$ que es la medida de información de Kullback-Leibler modificada.
- 3) Para $f(t) = (t-1) \log t$, $I_f(p, q) = \int_X (q-p) (\log q - \log p) d\mu(x)$ que es el invariante de Jeffreys.
- 4) Para $f(t) = (1 - \sqrt{t})^2$, $I_f(p, q) = \int_X (\sqrt{p} - \sqrt{q})^2 d\mu(x)$, que es la distancia Hellinger de grado 2, comúnmente conocida como distancia Matusita.
- 5) Para $f(t) = t \left|1 - \frac{1}{t}\right|^\alpha$, $I_f(p, q) = \int_X q \left|1 - \frac{p}{q}\right|^\alpha d\mu(x)$, que es la medida de Vajda de grado α , obteniéndose como caso particular para $\alpha = 2$, $I_f(p, q) = \int_X \frac{(q-p)^2}{q} d\mu(x)$, que es la medida de Kagan.

Obsérvese que en realidad esta medida no es una medida de Csiszar, pues $f(t)$ no es 2-veces diferenciable, pero puede entenderse como una medida de Csiszar generalizada.

7. RELACION DE LA MEDIDA DE CSISZAR CON LAS ENTROPIAS

En esta sección, de considerar la media de Csiszar como una medida que cuantifique las diferencias entre las distribuciones multinomiales $p = (p_1, \dots, p_n)$ y $q = (n^{-1}, \dots, n^{-1})$, hemos obtenido.

Proposición 7.1.

Para la familia de funciones

$$f_{\alpha}(x) = \begin{cases} \frac{n^{1-\alpha}}{1-\alpha} (x^{1-\alpha} n^{\alpha-1}) & \text{para } \alpha \neq 1 \\ \log x + \log n & \text{para } \alpha = 1 \end{cases}$$

$I_f(p, q) = H_{n, \alpha}(p)$, siendo $H_{n, \alpha}(p)$ la entropía de Havrda-Charvat de grado α para $\alpha \neq 1$ y $H_{n, 1}(p)$ la entropía de Shannon.

Demostración

En el caso finito, en el que el peso de cada suceso viene determinado por su probabilidad, la medida de Csiszar se reduce a

$$I_f(p, q) = \sum_{i=1}^n p_i f\left(\frac{q_i}{p_i}\right)$$

obteniéndose para la familia propuesta las siguientes igualdades

a) Para $\alpha \neq 1$:

$$\begin{aligned} I_f(p, q) &= \sum_{i=1}^n p_i f\left(\frac{1}{np_i}\right) = \sum_{i=1}^n p_i \frac{n^{1-\alpha}}{1-\alpha} \cdot ((np_i)^{\alpha-1} - n^{\alpha-1}) = \\ &= \frac{1}{\alpha-1} \left(1 - \sum_{i=1}^n p_i^\alpha\right) = H_{n, \alpha}(p) \end{aligned}$$

b) Para $\alpha = 1$:

$$I_f(p, q) = \sum_{i=1}^n p_i f\left(\frac{1}{np_i}\right) = \sum_{i=1}^n (-p_i \log np_i + p_i \log n) = - \sum_{i=1}^n p_i \log p_i = H_{n, 1}(p)$$

Proposición 7.2.

$$I_f(p, q) = - \sum_{i=1}^n p_i \log p_i - \sum_{i=1}^n (1-p_i) \log (1-p_i),$$
 para la función definida como $f(x) = nx \log nx + (1-nx) \log (nx-1)$.

Demostración

De las igualdades

$$\begin{aligned} I_f(p, q) &= \sum_{i=1}^n p_i f\left(\frac{1}{np_i}\right) = \sum_{i=1}^n p_i \left[\frac{1}{p_i} \log \frac{1}{p_i} + \frac{p_i^{-1}}{p_i} \log \frac{1-p_i}{p_i} \right] = \\ &= \sum_{i=1}^n (p_i \log p_i + (1-p_i) \log (1-p_i)) \end{aligned}$$

Es decir, la entropía de Shannon modificada la hemos obtenido a través de las medidas de Csiszar.

Y de forma inmediata, las relaciones de la γ -entropía y la entropía de Renyi con las medidas de Csiszar las tenemos a partir de las siguientes igualdades.

Proposición 7.3.

La entropía de Renyi $H_{n,\alpha}^R(p) = \frac{1}{1-\alpha} \log \sum_{i=1}^n p_i^\alpha$ y la γ -entropía $H_{n,\alpha}^Y(p) = \frac{1}{1-2(1/\alpha)-1} (1 - (\sum_{i=1}^n p_i^\alpha)^{1/\alpha})$ con $\alpha = \frac{1}{\gamma}$ se obtienen a partir de las siguientes igualdades:

$$H_{n,\alpha}^Y(p) = \frac{1}{1-2(1/\alpha)-1} (1 - (1 - (\alpha-1)H_{n,\alpha}^R(p))^{1/\alpha})$$

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

8. CONVEXIDAD Y NO NEGATIVIDAD DE LAS DIVERGENCIAS

En esta sección analizaremos en primer lugar las condiciones para la no negatividad de las divergencias y en segundo lugar estudiaremos su convexidad. En este sentido, se han obtenido los siguientes resultados:

Proposición 8.1.

- a) $K_\phi(p; q) \geq 0$ para toda $p(x), q(x) \in L_\phi^1$ si y solo si $\Psi(t) = \frac{\phi(t)}{t}$ es creciente en T_ϕ
- b) $L_\phi(p, q) \geq 0$ para toda $p(x), q(x) \in L_\phi^1$ si y solo si $f(t) = t \cdot \phi(t^{-1}) + \phi(t)$ es no negativa en R^+ .
- c) $M_\phi(p, q) \geq 0$ para toda $p(x), q(x) \in L_\phi^1$.

Demostración

a) De tomar $K_\phi(p, q) \geq 0$ para las funciones

$$p(x) = \begin{cases} t_0 & \text{cuando } x \in K \\ 0 & \text{cuando } x \notin K \end{cases} \quad q(x) = \begin{cases} t_1 & \text{cuando } x \in K \\ 0 & \text{cuando } x \notin K \end{cases}$$

siendo $K \subset X$, con $0 \neq \mu(K) < +\infty$ y $t_0, t_1 \in T_\phi$, se tiene:

$$(t_0 - t_1) (\Psi(t_0) - \Psi(t_1)) \geq 0 \quad \text{para todo } t_0, t_1 \in T_\phi$$

de donde se deduce el crecimiento de $\Psi(t)$ en T_ϕ .

El recíproco es evidente por ser la integral de una función positiva.

b) De forma análoga a la anterior, tomando para todo $t_0 \in \mathbb{R}^+$ las funciones

$$p(x) = \begin{cases} 1 & \text{para } x \in K \\ 0 & \text{para } x \notin K \end{cases} \quad q(x) = \begin{cases} t_0 & \text{para } x \in K \\ 0 & \text{para } x \notin K \end{cases}$$

se tiene:

$$0 \leq L_\phi(p, q) = f(t_0) \cdot \mu(K)$$

resultando así la no negatividad de la función $f(t)$ en \mathbb{R}^+ . El recíproco es también evidente por ser la integral de una función positiva, al igual que $M_\phi(p, q)$.

Proposición 8.2.

$L_{\phi}^*(p, q) = \int_X \Psi(p) f\left(\frac{\Psi(q)}{\Psi(p)}\right) d\mu(x)$ es convexa (cóncava) en $L_{\phi}^1 \times L_{\phi}^1$ si y solo si la matriz $M = \begin{pmatrix} a(p, q) & b(p, q) \\ b(p, q) & a(q, p) \end{pmatrix}$ es semi-definida positiva (negativa), para

$$a(p, q) = f' \left(\frac{\Psi(p)}{\Psi(q)} \right) \Psi''(p) + f'' \left(\frac{\Psi(p)}{\Psi(q)} \right) \frac{(\Psi'(p))^2}{\Psi(q)}$$

$$b(p, q) = - f'' \left(\frac{\Psi(q)}{\Psi(p)} \right) \frac{\Psi'(p) \Psi'(q) \Psi(q)}{(\Psi(p))^2}$$

Demostración

Para que $L_{\phi}^*(p, q)$ sean convexa (cóncava) en $L_{\phi}^1 \times L_{\phi}^1$, la matriz de la forma cuadrática que define el Hessiano debe ser semidefinida positiva (negativa). En nuestro caso

$$d^2 L_{\phi}^*((p, q); (f, g), (f, g)) = \int_X (a(p, q) f^2 + 2b(p, q) f \cdot g + a(q, p) g^2) d\mu(x)$$

con $a(p, q)$ y $b(p, q)$ definidas en el enunciado.

Corolario 8.1.

$L_{\phi}(p, q)$ es convexa (cóncava) en $L_{\phi}^1 \times L_{\phi}^1$ si y solo si la función $f(t) = t \cdot \phi(t^{-1}) + \phi(t)$ es convexa (cóncava) en \mathbb{R}^+ .

Demostración

Para $\Psi(t) = t$, la matriz M viene determinada por

$$a(p, q) = f'' \left(\frac{p}{q} \right) \frac{1}{q} \quad \text{y} \quad b(p, q) = - f'' \left(\frac{q}{p} \right) \frac{q}{p^2}$$

resultando en este caso del $M = 0$.

Así, la convexidad de $L_{\phi}(p,q)$ dependerá exclusivamente del signo de $a(p,q)$, y como éste depende del signo de $f''(t)$, la convexidad de $L_{\phi}(p,q)$ solo se tendrá cuando $f(t)$ sea cóncava.

Corolario 8.2.

Para la familia $\phi_{\alpha}(t) = \begin{cases} \frac{1}{\alpha-1} (t^{\alpha}-t) & \text{cuando } \alpha \neq 1 \\ t \log t & \text{cuando } \alpha = 1 \end{cases}$
 $L_{\phi_{\alpha}}(p,q)$ es cóncava en $L_{\phi}^1 \times L_{\phi}^1$ para todo valor $\alpha \in \mathbb{R}^+$.

Demostración

Para todo valor positivo de α , $f''(t) > 0$, ya que

$$f''(t) = \begin{cases} \alpha(t^{-1-\alpha} + t^{\alpha-2}) & \text{para } \alpha \neq 1 \\ t^{-2} + t^{-1} & \text{para } \alpha = 1 \end{cases}$$

Corolario 8.3.

Son cóncavas en $L_{\phi}^1 \times L_{\phi}^1$ las funciones que relacionamos a continuación:

- a) La medida de Kullback-Leibler
- b) La medida de Kullback-Leibler modificada
- c) La medida de Jeffreys
- d) La medida de Vajda
- e) El cuadrado de la distancia Matusita

Demostración

Es inmediato comprobar que las medidas que hemos citado son todas convexas en $L_\phi^1 \times L_\phi^1$, pues son medidas de Csiszar determinadas por funciones convexas.

De considerar finalmente la forma cuadrática que define el Hessiano correspondiente a la K-divergencia se tiene:

Proposición 8.3.

$K_\phi(p,q)$ es convexa (cóncava) en $L_\phi^1 \times L_\phi^1$ si y solo si la matriz

$$M = \begin{pmatrix} a(p,q) & b(p,q) \\ b(p,q) & a(q,b) \end{pmatrix}$$

es semidefinida positiva (negativa) para

$$a(p,q) = \phi''(p) - q \Psi''(p), \quad b(p,q) = -(\Psi'(p) + \Psi'(q)) \quad y$$

$$\Psi(p) = \frac{\phi(p)}{p}$$

Corolario 8.4.

Si $\phi(t)$ es convexa en T_ϕ , $\Psi(t)$ es cóncava en T_ϕ y se cumple la condición $d(x,y) = a(x,y)a(y,x) - b(x,y)^2 \geq 0$, para todo $x,y \in T_\phi$, entonces $K_\phi(p,q)$ es convexa en $L_\phi^1 \times L_\phi^1$.

CAPITULO IV

METRICAS RIEMANIANAS ASOCIADAS A DIVERGENCIAS PARA
FAMILIAS DE DISTRIBUCIONES PARAMETRICAS

RESUMEN:

En este capítulo planteamos las métricas diferenciales asociadas a divergencias definidas en una familia paramétrica de funciones de densidad de probabilidad, y determinamos las funciones que nos proporcionan las divergencias que tienen asociadas métricas diferenciales invariantes frente a transformaciones no singulares de parámetros y de variables aleatorias.

1. INTRODUCCION

Al analizar la distanciación de objetos a partir de funciones de densidad de probabilidad, distintos autores han planteado la necesidad de la invariancia para la distancia al variar la forma de tomar las medidas de los objetos, es decir, la distancia no debe alterarse al efectuar un cambio no singular de variables aleatorias. En este sentido Kullback (1959) plantea la invariancia para las métricas diferenciales asociadas a las medidas de Kullback-Leibler y de Jeffreys, así como para la matriz de información de Fisher, Amari (1985) plantea la invariancia en el estudio de las α -conexiones y Cuadras y otros (1985) plantean la invariancia para la esperanza del cambio infinitesimal de información.

En este capítulo planteamos las métricas diferenciales asociadas a divergencias para funciones de densidad de probabilidad pertenecientes a la misma familia de funciones paramétricas y determinamos las divergencias que a nivel diferencial resultan invariantes frente a cambios no singulares de parámetros y variables aleatorias. Finalmente, presentamos la invariancia para el funcional ϕ -entropía a partir de la métrica diferencial asociada a la J-divergencia generalizada.

2. DEFINICIONES PREVIAS

Al igual que en capítulos precedentes, consideraremos una medida μ positiva, σ aditiva y σ finita definida en una σ -álgebra de subconjuntos de un espacio medible X , una función a valores

reales $\phi(t)$ definida en T_ϕ tres veces diferenciable con continuidad, y el conjunto de las funciones ϕ -integrables con valores en T_ϕ que denotaremos por L_ϕ^1 .

Tenemos también la familia paramétrica de funciones de densidad, de probabilidad $p=p(x, \theta)$ con $x \in X$ y $\theta = (\theta_1, \dots, \theta_n) \in \Omega$, siendo Ω un subconjunto de R^n .

Denotaremos por F_Ω el conjunto:

$$F_\Omega = \{p(x, \theta); p: X \rightarrow R \text{ tq } 1) \int_X |p(x, \theta)| d\mu(x) = 1$$

$$2) p(x, \theta) \in T_\phi \text{ para todo } x \in X\}$$

y finalmente, exigiremos a la función $f(x, \theta)$ que sea dos veces diferenciable respecto de las componentes del parámetro.

3. 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.

Plantaremos en este capítulo la distanciaci3n de objetos que vienen caracterizados por funciones de densidad paramétricas $p(x, \theta)$, pertenecientes a la misma familia, a partir de la métrica diferencial definida por el Hessiano de la segunda derivada en la direcci3n del espacio tangente al espacio paramétrico.

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 parametrizaci3n utilizada, es decir, debe ser invariante frente a transformaciones no singulares de los parámetros, entendiendo como tales las transformaciones biyectivas definidas a través de funciones diferenciables 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 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) Finalmente, la distancia entre los objetos a comparar debe aumentar al agregar variables aleatorias estocásticamente independientes, ya que aumenta la información sobre las diferencias.

4. METRICA RIEMANIANA ASOCIADA A LA J-DIVERGENCIA GENERALIZADA

Para la J-divergencia generalizada

$$J_{\phi}^{\lambda}(p, q) = \int_X (\lambda \phi(p) + (1-\lambda) \phi(q) - \phi(\lambda p + (1-\lambda)q)) d\mu$$

consideraremos a continuación la métrica definida por el Hessiano en la dirección del espacio tangente a F_{Ω} cuando $q \rightarrow p$. En este sentido, el cálculo del elemento diferencial de arco resulta de las siguientes igualdades.

$$\begin{aligned} ds_{\phi}^2 &= \Delta_{\theta} J_{\phi}^{\lambda}(p, p) = d^2 J_{\phi}^{\lambda}(p, p) = \\ &= \frac{d^2}{dt^2} \int_X (\lambda \phi(p) + (1-\lambda) \phi(p+tdp) - \phi(\lambda p + (1-\lambda)(p+tdp))) d\mu \Big|_{t=0} = \\ &= \frac{d}{dt} (1-\lambda) \int_X ((\phi'(p+tdp) - \phi'(\lambda p + (1-\lambda)(p+tdp))) dp) d\mu \Big|_{t=0} = \\ &= (1-\lambda) \int_X ((\phi''(p+tdp) - \phi''(\lambda p + (1-\lambda)(p+tdp))) (1-\lambda)) (dp)^2 d\mu \Big|_{t=0} = \\ &= \lambda(1-\lambda) \int_X \phi''(p) (dp)^2 d\mu. \end{aligned}$$

y puesto que $dp = \sum_{i=1}^n \frac{\partial p}{\partial \theta_i} d\theta_i$, el elemento diferencial de arco se expresará como

$$ds_\phi^2 = \lambda(1-\lambda) \sum_{i,j=1}^n \left(\int_X \phi''(p) \frac{\partial p}{\partial \theta_i} \frac{\partial p}{\partial \theta_j} d\mu \right) d\theta_i d\theta_j$$

expresión que se reduce a

$$ds_\phi^2 = \lambda(1-\lambda) \sum_{i,j=1}^n g_{ij}^\phi d\theta_i d\theta_j$$

donde

$$g_{ij}^\phi = \int_X \phi''(p) \frac{\partial p}{\partial \theta_i} \frac{\partial p}{\partial \theta_j} d\mu \quad \text{y} \quad p = p(x, \theta) \in F_\Omega$$

De forma inmediata, se comprueba el siguiente resultado:

Proposición 4.1.

- a) La matriz (g_{ij}^ϕ) define un tensor covariante de segundo orden
- b) La métrica definida por la matriz (g_{ij}^ϕ) es Riemanniana si y solo si $\phi(t)$ es convexa en T_ϕ .

Demostración

- a) Al efectuar un cambio no singular en los parámetros, $(\theta_1, \dots, \theta_n) \rightarrow (\bar{\theta}_1, \dots, \bar{\theta}_n)$, se tendrá:

$$\begin{aligned} \bar{g}_{ij}^\phi &= \int_X \phi''(\bar{p}) \cdot \frac{\partial \bar{p}}{\partial \bar{\theta}_i} \frac{\partial \bar{p}}{\partial \bar{\theta}_j} d\mu = \\ &= \int_X \phi''(p) \left(\sum_{r=1}^n \frac{\partial p}{\partial \theta_r} \frac{\partial \theta_r}{\partial \bar{\theta}_i} \right) \left(\sum_{s=1}^n \frac{\partial p}{\partial \theta_s} \frac{\partial \theta_s}{\partial \bar{\theta}_j} \right) d\mu \\ &= \sum_{r,s=1}^n \left(\int_X \phi''(p) \frac{\partial p}{\partial \theta_r} \frac{\partial p}{\partial \theta_s} d\mu \right) \frac{\partial \theta_r}{\partial \bar{\theta}_i} \frac{\partial \theta_s}{\partial \bar{\theta}_j} = \sum_{r,s=1}^n g_{rs}^\phi \frac{\partial \theta_r}{\partial \bar{\theta}_i} \frac{\partial \theta_s}{\partial \bar{\theta}_j} \end{aligned}$$

ya que $p=\bar{p}$, pues la función de densidad es invariante en F_{Ω} .

b) Es evidente que el elemento diferencial de arco es positivo para todo $p(x) \in L_{\phi}^1$ si y solo si $\phi''(t) \geq 0$ en T_{ϕ} .

Una alternativa a la expresión del elemento diferencial de arco y al de las componentes del tensor que define la métrica, la obtenemos a partir del concepto de esperanza con las igualdades

$$g_{ij}^{\phi} = E_{\theta} \left(\phi''(p) \cdot p \cdot \frac{\partial \log p}{\partial \theta_i} \frac{\partial \log p}{\partial \theta_j} \right)$$

$$ds_{\phi}^2 = \lambda(1-\lambda) E_{\theta} \left(\phi''(p) \cdot p \cdot (d \log p)^2 \right)$$

Relaciones entre la matriz que define la métrica y las matrices informativas, que comúnmente utilizamos en análisis de datos, las obtenemos a partir de la familia

$$\phi_{\alpha}(t) = \begin{cases} (\alpha-1)^{-1} (t^{\alpha}-t) & \text{para } \alpha \neq 1 \\ t \log t & \text{para } \alpha = 1 \end{cases}$$

En este sentido, la matriz (g_{ij}^{α}) definida por

$$g_{ij}^{\alpha} = \alpha \cdot E_{\theta} \left(p^{\alpha-1} \frac{\partial \log p}{\partial \theta_i} \frac{\partial \log p}{\partial \theta_j} \right)$$

coincide con la matriz de información de orden α módulo α y el elemento de línea

$$ds_{\alpha}^2 = \lambda \cdot (1-\lambda) \cdot \alpha \cdot E_{\theta} \left(p^{\alpha-1} (d \log p)^2 \right)$$

con la distancia geodésica de orden α , módulo $\lambda(1-\lambda)$.

Es interesante señalar que para $\alpha = 1$, la matriz (g_{ij}^1) coincide con la matriz de información de Fisher y que ds_1^2 es el diferencial de la distancia de Rao módulo $\lambda(1-\lambda)$.

5. MÉTRICAS RIEMANNIANAS ASOCIADAS A LA K, L Y M DIVERGENCIAS

De consideraciones análogas a las del apartado anterior, calcularemos la métrica definida por el Hessiano en la dirección del espacio tangente a F_Ω cuando $q \rightarrow p$ para la K, L y M divergencias. En este sentido, para la K-divergencia definida por

$$K_\phi(p, q) = \int_X (p-q) \left(\frac{\phi(p)}{p} - \frac{\phi(q)}{q} \right) d\mu(x)$$

se obtiene el elemento de línea a partir de las siguientes igualdades

$$\begin{aligned} ds_\phi^2 &= d^2 K_\phi(p, p) = \frac{d}{dt^2} \int_X (-tdp) \left(\frac{\phi(p)}{p} - \frac{\phi(p+tdp)}{p+tdp} \right) d\mu \Big|_{t=0} = \\ &= \frac{d}{dt} \int_X \left(\frac{\phi(p+tdp)}{p+tdp} + \frac{\phi'(p+tdp)}{p+tdp} \cdot t \cdot dp - \frac{\phi(p+tdp)}{(p+tdp)^2} \cdot t \cdot dp \right) dp \cdot d\mu \Big|_{t=0} = \\ &= 2 \int_X \left(\frac{\phi'(p)}{p} - \frac{\phi(p)}{p^2} \right) (dp)^2 d\mu(x) = 2 \int_X \left(\frac{\phi(p)}{p} \right)' (dp)^2 d\mu \end{aligned}$$

de forma que la matriz que define la métrica se reduce a

$$g_{ij}^\phi = \int_X \left(\frac{\phi(p)}{p} \right)' \frac{\partial p}{\partial \theta_i} \frac{\partial p}{\partial \theta_j} d\mu$$

y el elemento diferencial de arco se expresa como

$$ds^2 = 2 \sum_{i,j=1}^n g_{ij}^{\phi} d\theta_i d\theta_j$$

para L-divergencia definida por

$$L_{\phi}(p,q) = \int_X \left(p \phi\left(\frac{q}{p}\right) + q \phi\left(\frac{p}{q}\right) \right) d\mu$$

el elemento de línea se obtiene como resultado de los cálculos.

$$\begin{aligned} ds_{\phi}^2 &= d^2 L_{\phi}(p,p) = \frac{d}{dt^2} \int_X \left(p \phi\left(\frac{p+tdp}{p}\right) + (p+tdp) \phi\left(\frac{p}{p+tdp}\right) \right) d\mu \Big|_{t=0} \\ &= \frac{d}{dt} \int_X \left(\phi'\left(\frac{p+tdp}{p}\right) + \phi\left(\frac{p}{p+tdp}\right) - \frac{p}{p+tdp} \phi'\left(\frac{p}{p+tdp}\right) \right) dp d\mu \Big|_{t=0} \\ &= 2 \phi''(1) \int_X \frac{1}{p} (dp)^2 d\mu. \end{aligned}$$

es decir, la matriz que define la métrica se reduce a la matriz de información de Fisher.

$$g_{ij}^{\phi} = \int_X \frac{1}{p} \frac{\partial p}{\partial \theta_i} \frac{\partial p}{\partial \theta_j} d\mu = \int_X p \cdot \frac{\partial \log p}{\partial \theta_i} \frac{\partial \log p}{\partial \theta_j} d\mu$$

y el elemento de línea es el diferencial de la distancia de Rao módulo 2 $\phi''(1)$.

$$ds_{\phi}^2 = 2 \phi''(1) \sum_{i,j=1}^n g_{ij}^{\phi} d\theta_i d\theta_j$$

Para el cuadrado de la M-divergencia definida por

$$M_{\phi}^2(p,q) = \int_X (\sqrt{\theta(p)} - \sqrt{\theta(q)})^2 d\mu$$

el elemento de línea lo obtenemos como resultado de la segunda derivada

$$\begin{aligned}
ds_\phi^2 &= d^2 M_\phi^2(p,p) = \frac{d}{dt^2} \int_X (\sqrt{\phi(p)} - \sqrt{\phi(p+t\dot{p})})^2 d\mu \Big|_{t=0} = \\
&= \frac{d}{dt} \int_X 2 (\sqrt{\phi(p)} - \sqrt{\phi(p+t\dot{p})}) \frac{-\phi'(p+t\dot{p})}{2\sqrt{\phi(p+t\dot{p})}} \cdot \dot{p} \cdot d\mu \Big|_{t=0} = \\
&= 2 \int_X \left(\frac{\phi'(p)}{2\sqrt{\phi(p)}} \right)^2 (\dot{p})^2 d\mu = 2 \int_X ((\sqrt{\phi(p)})')^2 (\dot{p})^2 d\mu
\end{aligned}$$

Así, la matriz que determina la métrica se reduce a

$$g_{ij}^\phi = \int_X ((\sqrt{\phi(p)})')^2 \frac{dp}{d\theta_i} \frac{dp}{d\theta_j} d\mu$$

y el elemento de línea se expresa como

$$ds^2 = 2 \sum_{i,j=1}^n g_{ij}^\phi d\theta_i d\theta_j$$

A la vista de los resultados obtenidos, es inmediato comprobar que para la familia

$$\phi_\alpha(t) = \begin{cases} (\alpha-1)^{-1} (t^\alpha - t) & \text{para } \alpha \neq 1 \\ t \log t & \text{para } \alpha = 1 \end{cases}$$

las métricas que definen la J y la K divergencia coinciden salvo constantes multiplicativas. En este sentido, tenemos la relación

$$d^2 J_\alpha^\lambda(p,p) = \frac{1}{2} \cdot \lambda(1-\lambda)\alpha \cdot d^2 K_\alpha(p,p)$$

Así, la métrica definida por K-divergencia para la familia $\phi_\alpha(t)$ coincide, salvo constantes multiplicativas, con la métrica asociada a la matriz de información de orden α .

Para la misma familia, ya que $\phi_\alpha''(1) = \alpha$ para cualquier valor de α , la métrica que define la L-divergencia coincide con la que define la matriz de información de Fisher módulo 2α .

Finalmente, la métrica que define la matriz de información de orden α se obtiene también para la M-divergencia, salvo la constante multiplicativa $\frac{\alpha^2}{2}$, para la familia $\psi_\alpha(t) = t^\alpha$.

6. METRICAS DIFERENCIALES INVARIANTES FRENTE A TRANSFORMACIONES NO SINGULARES DE VARIABLES ALEATORIAS

En esta sección determinaremos las funciones $\phi(t)$, que a nivel diferencial, para la medida Lebesgue, nos proporcionan distancias invariantes frente a transformaciones no singulares de las variables aleatorias. En este sentido, es interesante demostrar previamente el siguiente resultado:

Lema 6.1.

Si $f(x)$ es una función diferenciable con continuidad a valores reales y cumpliendo la condición $f(x.y).y = f(x)$ para todo $x, y \in \mathbb{R}$, con $y \neq 0$, entonces $f(x) = \frac{a}{x}$.

Demostración

Derivando la condición $f(x.y).y = f(x)$ respecto x e y separadamente, obtenemos las ecuaciones:

$$f'(x \cdot y) \cdot y^2 = f'(x) \quad (1)$$

$$f'(x \cdot y) \cdot x \cdot y + f(x \cdot y) = 0 \quad (2)$$

$$f(x \cdot y) \cdot y = f(x) \quad (3)$$

Ya que de (1) $f'(x \cdot y) = \frac{f'(x)}{y^2}$, y de (3) $f(x \cdot y) = \frac{f(x)}{y}$, la ecuación (2) se transforma en:

$$f'(x) \cdot \frac{x}{y} + \frac{f(x)}{y} = 0$$

condición equivalente a

$$\frac{f'(x)}{f(x)} = -\frac{1}{x}$$

y de integrar la ecuación diferencial resulta

$$f(x) = \frac{a}{x}$$

Cuando realizamos un cambio no singular de variables aleatorias, es decir, cuando transformamos $p = p(x_1, \dots, x_k, \theta_1, \dots, \theta_n)$ en $\bar{p} = \bar{p}(y_1, \dots, y_k, \theta_1, \dots, \theta_n)$, el tensor métrico se transforma para la J, K, L y M divergencia según las siguientes relaciones:

a) Para la J-divergencia generalizada

$$\begin{aligned} g_{ij}^{\phi} &= \int_X \phi''(p) \frac{\partial p}{\partial \theta_i} \frac{\partial p}{\partial \theta_j} dx + \int_X \phi''(\bar{p}) \frac{\partial \bar{p}}{\partial \theta_i} \frac{\partial \bar{p}}{\partial \theta_j} dy = \\ &= \int_X \phi''(p \cdot J) \frac{\partial p}{\partial \theta_i} \frac{\partial p}{\partial \theta_j} J^2 dy = \int_X \phi''(p \cdot J) J \frac{\partial p}{\partial \theta_i} \frac{\partial p}{\partial \theta_j} dx = \bar{g}_{ij}^{\phi} \end{aligned}$$

con $J = \left| \det \left(\frac{\partial x}{\partial y} \right) \right|$.

a) Para la K-divergencia

$$\begin{aligned} g_{ij}^{\phi} &= \int_X \left(\frac{\phi(p)}{p} \right)' \frac{\partial p}{\partial \theta_i} \frac{\partial p}{\partial \theta_j} dx + \int_X \left(\frac{\phi(\bar{p})}{\bar{p}} \right)' \frac{\partial \bar{p}}{\partial \theta_i} \frac{\partial \bar{p}}{\partial \theta_j} dy = \\ &= \int_X \left(\frac{\phi(p, J)}{p \cdot J} \right)' \frac{\partial p}{\partial \theta_i} \frac{\partial p}{\partial \theta_j} J^2 dy = \int_X \left(\frac{\phi(p, J)}{p \cdot J} \right)' J \frac{\partial p}{\partial \theta_i} \frac{\partial p}{\partial \theta_j} dx = \bar{g}_{ij}^{\phi} \end{aligned}$$

c) Para la L-divergencia

$$g_{ij}^{\phi} = \int_X \frac{1}{p} \frac{\partial p}{\partial \theta_i} \frac{\partial p}{\partial \theta_j} dx + \int_X \frac{1}{\bar{p}} \frac{\partial \bar{p}}{\partial \theta_i} \frac{\partial \bar{p}}{\partial \theta_j} dx = \bar{g}_{ij}^{\phi}$$

d) Para la M-divergencia

$$g_{ij}^{\phi} = \int_X \left((\sqrt{\phi(p)})' \right)^2 \frac{\partial p}{\partial \theta_i} \frac{\partial p}{\partial \theta_j} dx + \int_X \left((\sqrt{\phi(p \cdot J)})' \right)^2 \frac{\partial p}{\partial \theta_i} \frac{\partial p}{\partial \theta_j} dx = \bar{g}_{ij}^{\phi}$$

Así, para la obtención de distancias invariantes frente a transformaciones no singulares de variables aleatorias, debe cumplirse para todo $x, y \in \mathbb{R}$, con $x > 0$ e $y \neq 0$ las condiciones:

a) Para la J-divergencia generalizada: $\phi''(x) = \phi''(x \cdot y) \cdot y$

b) Para la K-divergencia: $\left(\frac{\phi(x)}{x} \right)' = \left(\frac{\phi(x \cdot y)}{x \cdot y} \right) y$

c) Para la L-divergencia: $\frac{1}{x} = \frac{1}{x}$

d) Para la M-divergencia: $\left((\sqrt{\phi(x)})' \right)^2 = \left((\sqrt{\phi(x \cdot y)})' \right)^2 \cdot y$

y a partir del lema 6.1., se tiene el siguiente resultado:

Teorema 6.1.

Las funciones $\phi(t)$ que determinan metricas diferenciales invariantes frente a transformaciones no singulares de variables aleatorias se relacionan a continuación:

a) Para la J-divergencia generalizada: $\phi(t) = at \cdot \log t + bt + c$

b) Para la K-divergencia: $\phi(t) = at \cdot \log t + bt$

c) Para la L-divergencia: todas las funciones $\phi(t)$ determinan invariancia

d) Para la M-divergencia: $\sqrt{\phi(t)} = a\sqrt{t} + b$

Demostración

Es inmediato del lema 6.1., que para la J-divergencia generalizada $\phi''(t) = \frac{a}{t}$, para la K-divergencia $(\frac{\phi(t)}{t})' = \frac{a}{t}$ y para la M-divergencia $(\sqrt{\phi(t)})' = \sqrt{\frac{a}{t}}$. Así, integrando una o dos veces obtenemos las funciones de teorema.

Observemos además que las familias de funciones que determinan invariancia a nivel diferencial, cumplen las siguientes condiciones a nivel global:

$$(1) J_{\phi}^{\lambda}(p, q) = a J_{\psi}^{\lambda}(p, q) \quad \text{con } \psi(t) = t \log t$$

$$(2) K_{\phi}(p, q) = a K_{\psi}(p, q) \quad \text{con } \psi(t) = t \log t$$

$$(3) M_{\phi}^2(p, q) = a^2 \int_X (\sqrt{p} - \sqrt{q})^2 d\mu$$

es decir, la métrica diferencial invariante viene determinada por la función $\Psi(t) = t \log t$ para la J-divergencia generalizada y para la K-divergencia y para la M-divergencia, la métrica invariante es la que proviene del cuadrado de la distancia Matusita.

Finalmente, las métricas invariantes vienen determinadas, para todas las divergencias, por el tensor correspondiente a la matriz de información de Fisher módulo una constante,

$$g_{ij} = E \left(\frac{1}{p} \frac{\partial p}{\partial \theta_i} \frac{\partial p}{\partial \theta_j} \right)$$

y en este caso, si x e y son variables aleatorias vectoriales estocásticamente independientes, se verifica

$$p(x, y, \theta) = p_x(x, \theta) p_y(y, \theta)$$

de donde se deduce

$$g_{ij} = E \left(\frac{1}{p_x} \frac{\partial p_x}{\partial \theta_i} \frac{\partial p_x}{\partial \theta_j} \right) + E \left(\frac{1}{p_y} \frac{\partial p_y}{\partial \theta_i} \frac{\partial p_y}{\partial \theta_j} \right) = g_{ij}^x + g_{ij}^y$$

Así, la forma cuadrática diferencial que define el elemento de línea es la suma de las formas cuadráticas correspondientes a las variables x e y separadamente, es decir

$$ds^2 = ds_x^2 + ds_y^2$$

Nótese que esta descomposición asegura el incremento en la distancia estudiada al añadir variables aleatorias independientes.

7. INVARIANCIA PARA EL FUNCIONAL ϕ -ENTROPIA

Finalmente, procediendo de forma análoga a la realizada en las secciones anteriores, se tendrá para el funcional ϕ -entropía definido por

$$H(p) = \int_X \phi(p) d\mu$$

el elemento de línea como resultado de las igualdades

$$ds^2 = d^2 H_\phi(p) = \frac{d^2}{dt^2} \int_X \phi(p+tdp) d\mu \Big|_{t=0} = \int_X \phi''(p) (dp)^2 d\mu$$

Así, la métrica diferencial asociada al funcional ϕ -entropía coincide con la métrica diferencial correspondiente a la J -divergencia generalizada, salvo la constante multiplicativa $\lambda(1-\lambda)$.

CAPITULO V

MEDIDAS DE INCERTIDUMBRE ASOCIADAS A J-DIVERGENCIAS

Resumen

En este capítulo introducimos el concepto de medida de incertidumbre asociada a J-divergencia y analizamos la concavidad y no negatividad de los índices, atendiendo a las funciones y constantes que los determinan.

Planteamos las entropías más comunes (Shannon, cuadrática, Havrda-Charvat) como casos particulares de las medidas anteriormente definidas y relacionamos la entropía de Renyi y la γ -entropía con las medidas anteriores.

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), Renyi (1961), de Grood (1970), Havrda-Charvat (1967), Theodorescu (1977) y Boekee (1980).

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

Hemos dividido el capítulo en dos partes: en la primera analizamos la pérdida de incertidumbre y la concavidad de la medida, atendiendo a la función que la determina; en la segunda planteamos la entropía de Shannon, la entropía de Havrda-Charvat y la entropía de Shannon modificada como casos particulares de medidas de incertidumbre asociadas a J-divergencias, y además relacionamos la entropía de Renyi y la γ -entropía con las medidas anteriores.

2. DEFINICIONES PREVIAS

Dados A_1, A_2, \dots, A_n sucesos mutuamente excluyentes de probabilidades p_1, \dots, p_n respectivamente y verificando la condi-

ción: $p_1 + \dots + p_n = 1$, definimos como funciones de incertidumbre asociadas a J-divergencias a las funciones del tipo

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

donde ϕ es una función de $I = [0, 1]$ en \mathbb{R} dos veces diferenciable con continuidad, $c = \frac{1}{n}$, 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 $-I_{\phi}(x_1, \dots, x_n)$ cóncava y simétrica, $-I_{\phi}(x_1 \dots x_n)$ es Shur-convexa, es decir, cumple la propiedad

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

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

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

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

$$(\phi'(p_i + \varepsilon_1) - \phi'(\frac{1}{2}(p_i + \varepsilon_1 + c))) - (\phi'(p_j - \varepsilon_2) - \phi'(\frac{1}{2}(p_j - \varepsilon_2 + c))) \leq 0$$

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

$$\frac{(\phi(p_i + \delta) - \phi(\frac{1}{2}(p_i + \delta + c))) - (\phi(p_i) - \phi(\frac{1}{2}(p_i + c)))}{\delta} \geq$$

$$\frac{(\phi(p_j) - \phi(\frac{1}{2}(p_j + c))) - (\phi(p_j - \delta) - \phi(\frac{1}{2}(p_j - \delta + c)))}{\delta}$$

y de la reordenación de los factores resulta la condición deseada

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

De esta desigualdad se deduce de forma inmediata los siguientes resultados

Corolario 3.1.

$I_{\phi}(p_1, \dots, p_n)$ alcanza el máximo en el punto (c, \dots, c) 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

$$k_0 = \phi\left(\frac{1}{2}(1+c)\right) + (n-1)\phi\left(\frac{1}{2}c\right) - \frac{1}{2} \cdot n\phi(c) - \frac{1}{2}(\phi(1) + (n-1)\phi(0))$$

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 encontramos condiciones suficientes para la concavidad de $I_{\phi}(p_1, \dots, p_n)$ atendiendo a la concavidad y convexidad de $\phi(t)$ y $(\phi'')^{-1}(t)$.

En este sentido hemos encontrado el siguiente resultado:

Proposición 4.1.

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

Demostración

De la convexidad de $(\phi'')^{-1}(t)$ en I , se tiene la siguiente desigualdad

$$\frac{1}{\phi''(\frac{1}{2}(p_i+c))} \leq \frac{1}{2} \frac{1}{\phi''(p_i)} + \frac{1}{2} \frac{1}{\phi''(c)}$$

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

$$\frac{1}{\phi''(\frac{1}{2}(p_i+c))} \leq \frac{1}{2} \frac{1}{\phi''(p_i)}$$

y ésta a su vez se escribe en la forma

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

Así, de la condición (2), se deduce de forma inmediata la relación

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

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

deduciendo así 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 $(\phi'')^{-1}(t)$ convexa y por la constante k_0 es una medida de incertidumbre decisiva.

Corolario 4.2.

Para $\phi(t) = t \log t$ y $k_0 = \log(n) - \frac{1}{2}(1+c) \log(n+1)$, la medida $I_\phi(p_1, \dots, p_n)$ es una medida de incertidumbre decisiva

5. LA ENTROPIA DE SHANNON COMO MEDIDA DE INCERTIDUMBRE ASOCIADA

A J-DIVERGENCIAS

Analizadas previamente algunas propiedades teóricas, es interesante plantear las medidas de entropía comunmente usadas como medidas de incertidumbre asociadas a J-divergencias. En este sentido, determinaremos en esta sección una función $\phi(x)$ que proporciona la entropía de Shannon.

Obviamente, una función $\phi(x)$ que determine la entropía de Shannon, debe cumplir la ecuación funcional

$$\frac{1}{2} \phi(x) - \phi\left(\frac{1}{2}(x+c)\right) = -x \log x + ax + b$$

para cualquier valor de a y b , y para todo $x \in [0, 1]$.

En esta sección, hemos determinado la función $\phi(x)$ en un entorno de " c " a partir del desarrollo de Taylor en el punto " c " de la función $f(x) = -x \log x + ax + b$ y hemos extendido la expresión de $\phi(x)$ al intervalo $[0, 1]$ a través de la relación

$$\phi(x) = 2 \phi\left(\frac{1}{2}(x+c)\right) - 2x \log x + 2ax + 2b$$

En nuestro caso, la expresión encontrada por este método se expresa en la forma

$$\phi(x) = -2 \sum_{i=0}^{k_0+1} (x+(2^i-1)c) \log(x+(2^i-1)c) - 2^{k_0+4} \cdot c \cdot P(x)$$

$$\text{con } P(x) = \sum_{i=0}^{\infty} (-1)^i \frac{2^i}{(2^{i+1}-1)(i+1)(i+2)} \left(\frac{x-c}{2^{k_0+2}c}\right)^{i+2}$$

donde k_0-1 es la parte entera de $\log_2(n-1)-1$.

Este método que hemos utilizado para la entropía de Shannon es general para todas las funciones $f(x)$ analíticas en el punto " c ".

Para comprobar que efectivamente la solución hallada es una solución de la familia de ecuaciones planteada, demostraremos en primer lugar, que la serie de potencias que caracteriza la función $\phi(t)$ está bien definida, y en segundo lugar, comprobaremos que es una solución de la ecuación

$$\frac{1}{2}(\phi(x) + \phi(c)) - \phi\left(\frac{1}{2}(x+c)\right) = -x \log x + (1+\log c)x + k$$

para una constante k que determinaremos posteriormente.

Finalmente, en la proposición 8.2, se comprobará que las soluciones de la familia de ecuaciones propuesta se obtienen añadiendo una función lineal a la solución particular encontrada.

Proposición 5.1.

La serie de potencias $\sum_{i=0}^{\infty} (-1)^i \frac{2^i}{(2^{i+1}-1)(i+1)(i+2)} \left(\frac{x-c}{2^{k_0+2}c}\right)^{i+2}$ converge absolutamente en el intervalo $((-2^{k_0+2}+1)c, (2^{k_0+2}+1)c)$.

Demostración

El radio de convergencia de la serie viene determinado por

$$\frac{1}{\rho} = \lim_{i \rightarrow \infty} \left| \frac{a_{i+1}}{a_i} \right| = \lim_{i \rightarrow \infty} \frac{(2^{i+2}-2)(i+1)}{(2^{i+2}-1)(i+3)} = 1$$

Así, la serie será absolutamente convergente para los valores que cumplen la condición

$$\left| \frac{x-c}{2^{k_0+2}c} \right| < 1$$

desigualdad que resulta equivalente a

$$x \in ((-2^{k_0+2}+1)c, (2^{k_0+2}+1)c)$$

Proposición 5.2.

La función $f(x) = \sum_{i=0}^{\infty} (-1)^i \frac{2^i}{(2^{i+1}-1)(i+1)(i+2)} \left(\frac{x-c}{2^{k_0+2}c}\right)^{i+2}$

está bien definida para los valores $x \in [0, 1]$.

Demostración

Basta demostrar

a) $1 < (2^{k_0+2}+1)c$ b) $(-2^{k_0+2}+1)c < 0$

En este sentido, son evidentes las siguientes desigualdades:

a) $(2^{k_0+2}+1)c = (2^{\lfloor \log_2(n-1) \rfloor + 3} + 1)c \geq 2^{1+\log_2(n-1)} > 1$

b) $(-2^{k_0+2}+1)c < 0$ ya que $k_0 \geq 0$

Proposición 5.3.

$$t \log t = t-1 + \sum_{i=0}^{\infty} (-1)^i \frac{(t-1)^{i+2}}{(i+1)(i+2)} \text{ para } t \in [1, 2]$$

Demostración

Para comprobar la condición propuesta demostraremos que el desarrollo de Taylor de la función $f(t) = t \log t$ es con-

vergente en $[0,2]$ y que el resto de Lagrange se aproxima a cero cuando el desarrollo se hace infinito.

El desarrollo de Taylor de la función $f(t) = t \log t$ en el punto $t=1$ viene dado por la expresión

$$t-1 + \sum_{i=0}^{\infty} (-1)^i \frac{(t-1)^{i+2}}{(i+1)(i+2)}$$

Es evidente que la serie de potencias es absolutamente convergente en el intervalo $[0,2]$, ya que la convergencia en $|t-1| < 1$ es consecuencia inmediata del valor del radio de convergencia

$$\frac{1}{\rho} = \lim_{i \rightarrow \infty} \left| \frac{a_{i+1}}{a_i} \right| = \lim_{i \rightarrow \infty} \frac{i+1}{i+3} = 1$$

y la convergencia en $t=0$ y $t=2$ resulta de la comparación con la serie hipergeométrica de grado 2.

$$\sum_{i=0}^{\infty} \left| (-1)^i \frac{(t-1)^{i+2}}{(i+1)(i+2)} \right| = \sum_{i=0}^{\infty} \frac{1}{i^2+3i+2} \sim \sum_{i=0}^{\infty} \frac{1}{i^2} \text{ para } t=0,2$$

Finalmente, vamos a demostrar que el resto i -ésimo de Lagrange, que viene dado por la expresión:

$$R_i(t) = \frac{(-1)^{i+1}}{k^{i+2}(i+1)i} (t-1)^{i+1} \text{ con } k \in (1, t) \text{ ó } k \in (t, 1)$$

cumple la condición

$$\lim_{i \rightarrow \infty} R_i(t) = 0 \text{ para } t \in [\frac{1}{2}, 2] \quad (4)$$

En este sentido, es fácil comprobar:

a) $|\frac{t-1}{k}| < 1$ para $t \in [\frac{1}{2}, 2]$

b) Que la serie $\sum_{i=1}^{\infty} (-1)^{i+1} \frac{1}{k^{i+2} (i+1)^i} (t-1)^{i+1}$ es absolutamente convergente para $|\frac{t-1}{k}| < 1$.

Así, al ser $R_i(t)$ para $t \in [\frac{1}{2}, 2]$ el valor del término general de una serie absolutamente convergente, queda probada la condición (4).

Proposición 5.4.

$$\frac{x + (2^{k_0+2} - 1)c}{2^{k_0+2}c} \in [\frac{1}{2}, 2] \text{ para } x \in [0, 1] \quad (5)$$

Demostración

Para $x \in [0, 1]$ se tiene la desigualdad

$$\frac{(2^{k_0+2} - 1)c}{2^{k_0+2}c} \leq \frac{x + (2^{k_0+2} - 1)c}{2^{k_0+2}c} \leq \frac{1 + (2^{k_0+2} - 1)c}{2^{k_0+2}c}$$

y por las desigualdades

a) $\frac{(2^{k_0+2} - 1)c}{2^{k_0+2}c} = 1 - \frac{1}{2^{k_0+2}} \geq \frac{1}{2}$, ya que $k_0 \geq 0$

b) $\frac{1 + (2^{k_0+2} - 1)c}{2^{k_0+2}c} = 1 + \frac{1-c}{2^{k_0+2}c} < 2$, ya que $(2^{k_0+2} + 1)c > 1$ tal

y como hemos visto en la proposición 4.2. a), se tiene la condición (5).

Corolario 5.1.

$$\sum_{i=0}^{\infty} (-1)^i \frac{1}{(i+1)(i+2)} \left(\frac{x-c}{2^{k_0+2}c}\right)^{i+2} = \frac{c-x}{2^{k_0+2}c} + \frac{x+(2^{k_0+2}-1)c}{2^{k_0+2}c} \log \frac{x+(2^{k_0+2}-1)c}{2^{k_0+2}c}$$

para $x \in [0, 1]$

Demostración

Basta aplicar la proposición 4.3. a $t = \frac{x+(2^{k_0+2}-1)c}{2^{k_0+2}c} \in [\frac{1}{2}, 2]$

Proposición 5.5.

$$\begin{aligned} \frac{1}{2}(\phi(x) + \phi(c)) - \phi\left(\frac{1}{2}(x+c)\right) &= -x \log x + x(1 + \log c) + k \text{ para} \\ \text{la función } \phi(x) &= -2^{k_0+1} \sum_{i=0}^{\infty} (x+(2^i-1)c) \log (x+(2^i-1)c) - \\ &- 2^{k_0+4} c \sum_{i=0}^{\infty} (-1)^i \frac{2^i}{(2^{i+1}-1)(i+1)(i+2)} \left(\frac{x-c}{2^{k_0+2}c}\right)^{i+2} \end{aligned}$$

Demostración

En primer lugar, calcularemos separadamente los valores de las expresiones: a) $\frac{1}{2}(\phi(x) + \phi(c))$ y b) $\phi\left(\frac{1}{2}(x+c)\right)$

$$\begin{aligned} \text{a) } \frac{1}{2}(\phi(x) + \phi(c)) &= -2^{k_0+1} \sum_{i=0}^{\infty} (x+(2^i-1)c) \log (x+(2^i-1)c) - 2^{k_0+1} \sum_{i=0}^{\infty} 2^i c \log 2^i c - \\ &- 2^{k_0+3} c \sum_{i=0}^{\infty} (-1)^i \frac{2^i}{(2^{i+1}-1)(i+1)(i+2)} \left(\frac{x-c}{2^{k_0+2}c}\right)^{i+2} \end{aligned}$$

$$\begin{aligned}
b) \phi\left(\frac{1}{2}(x+c)\right) &= - \sum_{i=0}^{k_0+1} (x+(2^{i+1}-1)c) \log \frac{1}{2} (x+(2^{i+1}-1)c) - \\
&- 2^{k_0+4} c \sum_{i=0}^{\infty} (-1)^i \frac{2^i}{(2^{i+1}-1)(i+1)(i+2)} \left(\frac{x-c}{2^{k_0+2}c}\right)^{i+2} \frac{1}{2^{i+2}} = \\
&= - \sum_{i=1}^{k_0+2} (x+(2^i-1)c) \log (x+(2^i-1)c) + (\log 2) \sum_{i=1}^{k_0+2} (x+(2^i-1)c) - \\
&- 2^{k_0+3} c \sum_{i=0}^{\infty} (-1)^i \frac{1}{(2^{i+1}-1)(i+1)(i+2)} \frac{1}{2} \left(\frac{x-c}{2^{k_0+2}c}\right)^{i+2}
\end{aligned}$$

Así, el valor de la expresión deseada resulta de la forma

$$\frac{1}{2} [\phi(x) + \phi(c)] - \phi\left(\frac{1}{2}(x-c)\right) = -x \log x + (x+(2^{k_0+2}-1)c) \log (x+(2^{k_0+2}-1)c)$$

$$+ k_1 - ((k_0+2) \log 2) x - 2^{k_0+3} c \cdot P(x)$$

$$\text{con } P(x) = \left[\sum_{i=0}^{\infty} (-1)^i \frac{1}{(2^{i+1}-1)(i+1)(i+2)} \left(\frac{x-c}{2^{k_0+2}c}\right)^{i+2} \right] \left(\frac{2^{i+1}-1}{2}\right)$$

$$= -x \log x + (x+(2^{k_0+2}-1)c) \log (x+(2^{k_0+2}-1)c) - ((k_0+2) \log 2) x - c + x -$$

$$- (x+(2^{k_0+2}-1)c) \log (x+(2^{k_0+2}-1)c) + (x+(2^{k_0+2}-1)c) \log 2^{k_0+2} c + k_1 =$$

$$= -x \log x + x + x \log c + k, \text{ donde } k_1 \text{ y } k \text{ son constantes.}$$

Así, de lo expuesto anteriormente se deduce el siguiente resultado

Teorema 5.1.

$I(p_1, \dots, p_n) = - \sum_{i=1}^n p_i \log p_i$ para la función $\phi(x)$ definida en la proposición 5.5. y $k = k_0$ del corolario 3.2.

6. LA ENTROPIA DE HAVRDA-CHARVAT COMO MEDIDA DE INCERTIDUMBRE ASOCIADA A J-DIVERGENCIAS

Por consideraciones análogas a las realizadas para la entropía de Shannon, hemos obtenido los siguientes resultados:

Proposición 6.1.

La serie de potencias $\sum_{i=1}^{\infty} \frac{\alpha(\alpha-1)\dots(\alpha-i+1)}{i!} \frac{2^{i-2}}{2^{i-1}-1} \left(\frac{x-c}{2^{k_0+2}}\right)^i$ converge absolutamente en el intervalo $((-2^{k_0+2}+1)c, (2^{k_0+2}+1)c)$

Proposición 6.2.

La función $f(x) = \sum_{i=1}^{\infty} \frac{(\alpha-1)\dots(\alpha-i+1)}{i!} \frac{2^{i-2}}{2^{i-1}-1} \left(\frac{x-c}{2^{k_0+2}}\right)^i$ está bien definida para los valores de $x \in [0, 1]$.

Proposición 6.3.

$$\sum_{i=1}^{\infty} \frac{\alpha(\alpha-1)\dots(\alpha-i+1)}{i!} \left(\frac{x-c}{2^{k_0+2}}\right)^i = \left(\frac{x+(2^{k_0+2}-1)c}{2^{k_0+2}}\right)^{\alpha} - \frac{x-c}{2^{k_0+2}}$$

para $x \in [0, 1]$.

Proposición 6.4.

$$\frac{1}{2}(\phi(x)+\phi(c))-\phi\left(\frac{1}{2}(x+c)\right)=x^{\alpha}-\frac{1}{2^{k_0+2}}(x-c)+k \text{ para la función}$$

$$\phi(x) = \sum_{i=0}^{k_0+1} 2^{i+1} \left(\frac{x+(2^i-1)c}{2^i}\right)^{\alpha} + 2^{k_0+4} \cdot c \cdot P(x)$$

$$\text{siendo } P(x) = \sum_{i=1}^{\infty} \frac{\alpha(\alpha-1)\dots(\alpha-i+1)}{i!} \frac{2^{i-1}}{2^{i-1}-1} \left(\frac{x-c}{2^{k_0+2}}\right)^i$$

Así, para la función $\Psi(x) = \frac{1}{1-\alpha} \phi(x)$ y $k=k_0$ se tendrá la entropía de Havrda-Charvat expresada de la forma siguiente.

Teorema 6.1.

$$I_{\Psi}(p_1, \dots, p_n) = \frac{1}{\alpha-1} \left(1 - \sum_{i=0}^n p_i^{\alpha} \right)$$

Analogamente, para la función $\phi(x) = (2^{1-\alpha}-1)^{-1} \psi(x)$ y $k=k_0$, se tendrá la entropía de grado α en la forma que sigue

Teorema 6.2.

$$I_{\phi}(p_1, \dots, p_n) = (2^{1-\alpha}-1)^{-1} \left(\sum_{i=0}^n p_i^{\alpha} - 1 \right)$$

Nota: Cuando α es un número natural, a veces, resultará interesante no utilizar la función $\phi(x)$ propuesta anteriormente para los casos generales, ya que funciones de expresión más sencilla nos darán el mismo resultado. En este sentido, y a título ilustrativo, podemos comprobar que para la función $\phi(x) = -4x^2$, obtenemos la entropía cuadrática (entropía de Havrda-Charvat de grado 2), y para la función $\phi(x) = -\frac{4}{3}t^3 - \frac{2}{n}t^2$, obtenemos la entropía de Havrda-Charvat de grado 3.

7. ENTROPIAS ADICIONALES ASOCIADAS A J-DIVERGENCIAS

Para finalizar este capítulo, presentaremos a continuación la entropía de Shannon modificada como medida de incertidumbre asociada a J-divergencias y relacionaremos la entropía

de Renyi y la γ -entropía con la medida encontrada en la sección anterior. En este sentido, los resultados obtenidos son los siguientes:

Proposición 7.1.

La entropía de Shannon modificada $-\sum_{i=1}^n p_i \log p_i - \sum_{i=1}^n (1-p_i) \log (1-p_i)$ resulta, una medida de incertidumbre asociado a J-divergencia al tomar $\eta(t) = \phi(t) + \phi(1-t)$, siendo $\phi(t)$ la función definida en la proposición 5.5.

Demostración

De la igualdad

$$\begin{aligned} & \frac{1}{2} (\eta(x) + \eta(c)) - \eta\left(\frac{1}{2}(x+c)\right) = \\ & = \frac{1}{2} (\phi(x) + \phi(c)) - \phi\left(\frac{1}{2}(x+c)\right) + \frac{1}{2} (\phi(1-x) + \phi(1-c)) - \phi\left(\frac{1}{2}(1-x) + (1-c)\right) \end{aligned}$$

se tiene:

$$I_{\eta}(p_1, \dots, p_n) = -\sum_{i=1}^n p_i \log p_i - \sum_{i=1}^n (1-p_i) \log(1-p_i) + k$$

y tomando como k la constante definida en el corolario 3.2., la medida de incertidumbre $I_{\eta}(p_1, \dots, p_n)$ coincide con la entropía de Shannon modificada.

Proposición 7.2.

La entropía de Renyi $H_{n,\alpha}^R(p) = \frac{1}{1-\alpha} \log \sum_{i=1}^n p_i^\alpha$ y la γ -entropía $H_{n,\alpha}^\gamma(p) = \frac{1}{1-2(1/\alpha)-1} (1 - (\sum_{i=1}^n p_i^\alpha)^{1/\alpha})$ con $\alpha = \frac{1}{\gamma}$ se obtienen a partir de las igualdades

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

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

siendo $I_\phi(p)$ la medida de incertidumbre definida en la sección 6 a través de la función presentada en la proposición 6.2.

Proposición 7.3.

Para la función $\phi(t) = a_2 t^2 + a_3 t^3 + a_4 t^4$ con $a_2 = -8+8c - \frac{8}{7} c^2$, $a_3 = \frac{16}{3} + \frac{32}{21} c$ y $a_4 = \frac{16}{7}$ y para la constante $k = k_0$ definida en el corolario 3.2., se tiene

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

es decir, la entropía de Latter también resulta un caso particular de medida de incertidumbre asociada a J-divergencia.

8. LOS FUNCIONALES ϕ -ENTROPIA COMO MEDIDAS DE INCERTIDUMBRE ASOCIADOS A J-DIVERGENCIAS

Analizadas algunas entropías particulares, es interesante plantear los funcionales ϕ -entropía como medidas asociadas a

J-divergencias. En este sentido, los funcionales ϕ -entropía resultarán como medidas de incertidumbre asociadas a funciones $\phi(t)$ cumpliendo la condición

$$\frac{1}{2} \phi(x) - \phi\left(\frac{1}{2}(x+c)\right) = f(x) + ax + b = g(x) \quad (6)$$

para a y b constantes reales y $x \in I = [\alpha, \beta]$ intervalo cerrado de \mathbb{R} . En la resolución de la ecuación (6) hemos encontrado.

Proposición 8.1.

La ecuación (6) para $a = -f'(c)$ y $b = -f(c) - ac$, tiene por solución $\phi(x) = (x-c) \phi'(c) + \sum_{i=0}^{\infty} 2^{i+1} g\left(c + \frac{x-c}{2^i}\right)$ cuando $f(x)$ es una función 2 veces diferenciable en I .

Demostración

De forma iterativa,

$$\phi(x) = 2\phi\left(\frac{x+c}{2}\right) + 2g(x) = \dots = 2^{n+1} \phi\left(c + \frac{x-c}{2^{n+1}}\right) + \sum_{i=0}^n 2^{i+1} g\left(c + \frac{x-c}{2^i}\right)$$

y pasando al límite,

$$\phi(x) = (x-c) \phi'(c) + \sum_{i=0}^{\infty} 2^{i+1} g\left(c + \frac{x-c}{2^i}\right)$$

ya que por $\phi(c) = 0$

$$\begin{aligned} \lim_{n \rightarrow \infty} 2^{n+1} \phi\left(c + \frac{x-c}{2^{n+1}}\right) &= \lim_{n \rightarrow \infty} \frac{\phi\left(c + \frac{x-c}{2^{n+1}}\right) - \phi(c)}{\frac{x-c}{2^{n+1}}} \cdot (x-c) = \\ &= \lim_{n \rightarrow \infty} \phi'(\theta_n) (x-c) = (x-c) \phi'(c) \end{aligned}$$

$$\text{con } \theta_n \in (c, c + \frac{x-c}{2^{n+1}}) \quad \text{o} \quad \theta_n \in (c + \frac{x-c}{2^{n+1}}, c)$$

Nótese que la función $\phi(x)$ está bien definida, pues del desarrollo de Taylor de grado 2 de la función $g(x)$.

$$g(c + \frac{x-c}{2^i}) = g(c) + g'(c) (\frac{x-c}{2^i}) + g''(\theta) (\frac{x-c}{2^i})^2$$

$$\text{con } \theta \in (c, c + \frac{x-c}{2^i}) \quad \text{o} \quad \theta \in (c + \frac{x-c}{2^i}, c)$$

se tiene que la serie

$$\sum_{i=0}^{\infty} |2^{i+1} g(c + \frac{x-c}{2^i})|$$

está dominada por la serie absolutamente convergente

$$M \sum_{i=0}^{\infty} \frac{1}{2^{i-1}} \quad \text{con} \quad M = (\beta - \alpha)^2 \sup_{x \in I} |g''(x)|$$

Proposición 8.2

Si $\phi_1(x)$ y $\phi_2(x)$ son funciones diferenciables con continuidad y son soluciones de la familia de ecuaciones $\frac{1}{2} \phi(x) - \phi(\frac{1}{2}(x+c)) = f(x) + ax + b$, con $f(x)$ diferenciable en el punto c , entonces $\phi_1(x)$ y $\phi_2(x)$ difieren en una aplicación lineal.

Demostración

De derivar y sustituir en el punto "c" las expresiones

$$\frac{1}{2} \phi_1(x) - \phi_1\left(\frac{1}{2}(x+c)\right) = f(x) + a_1x + b_1$$

$$\frac{1}{2} \phi_2(x) - \phi_2\left(\frac{1}{2}(x+c)\right) = f(x) + a_2x + b_2$$

se tiene $a_1 = a_2 = -f'(c)$.

Así, de derivar la diferencia de las expresiones anteriores se tiene de forma inmediata

$$\eta(x) - \eta\left(\frac{1}{2}(x+c)\right) = 0 \quad \text{para todo } x \in I$$

siendo $\eta(x) = \phi_1'(x) - \phi_2'(x)$.

De forma iterativa,

$$\eta(x) = \eta\left(c + \frac{x-c}{2}\right) = \dots = \eta\left(c + \frac{x-c}{2^n}\right)$$

y pasando al límite

$$\eta(x) = \lim_{n \rightarrow \infty} \eta\left(c + \frac{x-c}{2^n}\right) = \eta(c) = k_1$$

de donde se deduce

$$\phi_1(x) = \phi_2(x) + k_1x + k_2$$

CAPITULO VI

CARACTERISTICAS MUESTRALES DEL INDICE DE INCERTIDUMBRE
ASOCIADO A LA J-DIVERGENCIA CORRESPONDIENTE A LA
FUNCION $\phi(t) = t \log t$ RESUMEN

En este capítulo se han comparado las características muestrales del índice de incertidumbre asociado a la J-divergencia para $\phi(t) = t \log t$, atendiendo en primer lugar a dos estadísticas de dispersión (uno estandarizado y otro no), y en segundo lugar al coeficiente de correlación de Spearman.

Hemos obtenido para el índice de incertidumbre asociado a la J-divergencia propiedades de estimación comparables a los índices comúnmente usados, condición que unida a las buenas propiedades teóricas analizadas en el capítulo anterior, hacen pensar en la necesidad de un análisis más profundo de los índices correspondientes a la familia de las J-divergencias.

1. INTRODUCCION

Analizadas algunas propiedades teóricas de los índices de incertidumbre asociados a J-divergencias, nos ha parecido interesante realizar un estudio comparativo de los índices comúnmente usados con un índice de incertidumbre asociado a J-divergencia. En este sentido hemos tomado el índice correspondiente a la función $\phi(t) = -t \log t$, pues para esta función, la J-divergencia resulta invariante a nivel diferencial frente a transformaciones no singulares de las variables aleatorias y de los parámetros.

La comparación se ha realizado en dos partes: en la primera se han analizado conjunta y separadamente dos estadísticas de dispersión, una estandarizada y otra sin estandarizar; en la segunda parte se ha comparado la incertidumbre muestral con la teórica a partir del coeficiente de correlación de Spearman.

Para realizar las simulaciones de los puntos, se ha construido un vector de probabilidades acumuladas a partir de las frecuencias relativas de los sucesos, siguiendo la metodología descrita por Bratley, Fox y Schrage (1983).

2. 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:

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

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

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

(γ -entropía de grados $\frac{1}{2}$, 2 y 3 respectivamente)

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

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

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

(entropía de Renyi de grados $\frac{1}{2}$, 2 y 3 respectivamente)

$$H_{11} = (\sqrt{2}-1)^{-1} \left(\sum_{i=1}^n \sqrt{p_i} - 1 \right) \quad (\text{entropía de Havrda-Charvat modificada de grado } \frac{1}{2})$$

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

(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 .

3. 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:

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

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:

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

$$P^{(2)} = M(p_1^{(2)}, \dots, p_n^{(2)}; N) \text{ siendo:}$$

$$\begin{cases} p_i^{(2)} = (1/n) + ((n/2)+1-i)\epsilon & \text{para } 2i \leq n \\ p_i^{(2)} = (1/n) + ((n/2)-i)\epsilon & \text{para } 2i > n \end{cases}$$

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

$$P^{(3)} = M(p_1^{(3)}, \dots, p_n^{(3)}; N) \text{ siendo:}$$

$$\begin{cases} p_i^{(3)} = (1/n) + \delta & \text{para } 2i \leq n \\ p_i^{(3)} = (1/n) - \delta & \text{para } 2i > n \end{cases}$$

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

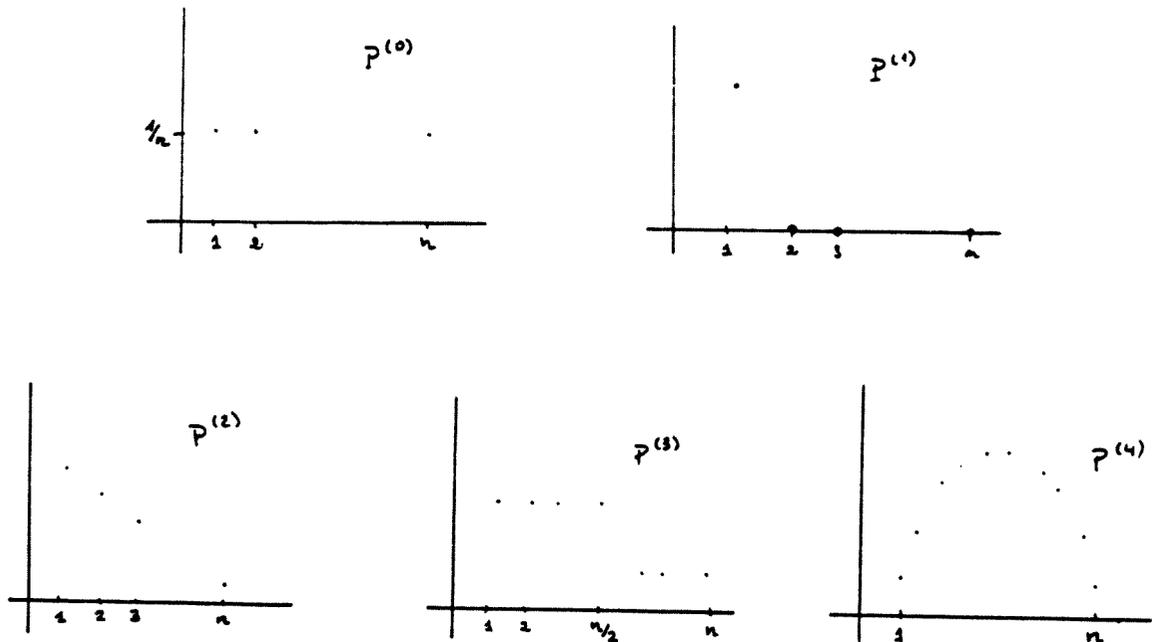
$$P^{(4)} = M(p_1^{(4)}, \dots, p_n^{(4)}; N) \text{ siendo } p_i^{(4)} = \frac{1}{\sqrt{2\pi}} \int_{a_i}^{b_i} e^{-u^2} du,$$

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 $P(-4+8(i-1)/n \leq X \leq -4+8i/n)$, siendo X una variable aleatoria que sigue una distribución $N(0, 1)$.

Gráficamente, los puntos descritos corresponden a las distribuciones que representamos a continuación.



Además de estos cinco modelos de puntos hemos estudiado puntos intermedios calculados a través de las geodésicas del espacio paramétrico de las distribuciones multinomiales. 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:

$$r_i(s) = (\sqrt{p_i} \cos(s/2\sqrt{N}) + \frac{\sqrt{q_i} - \sqrt{p_i} \cos(d/2\sqrt{N})}{\sin(d/2\sqrt{N})}) \cdot \sin(s/2\sqrt{N})^2$$

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

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 5000 veces y hemos calculado el valor de las estadísticas $\sigma^2(H)$ y V .

Hemos realizado también un seguimiento del proceso anterior 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 5000 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), y hemos añadido al generador congruencial un

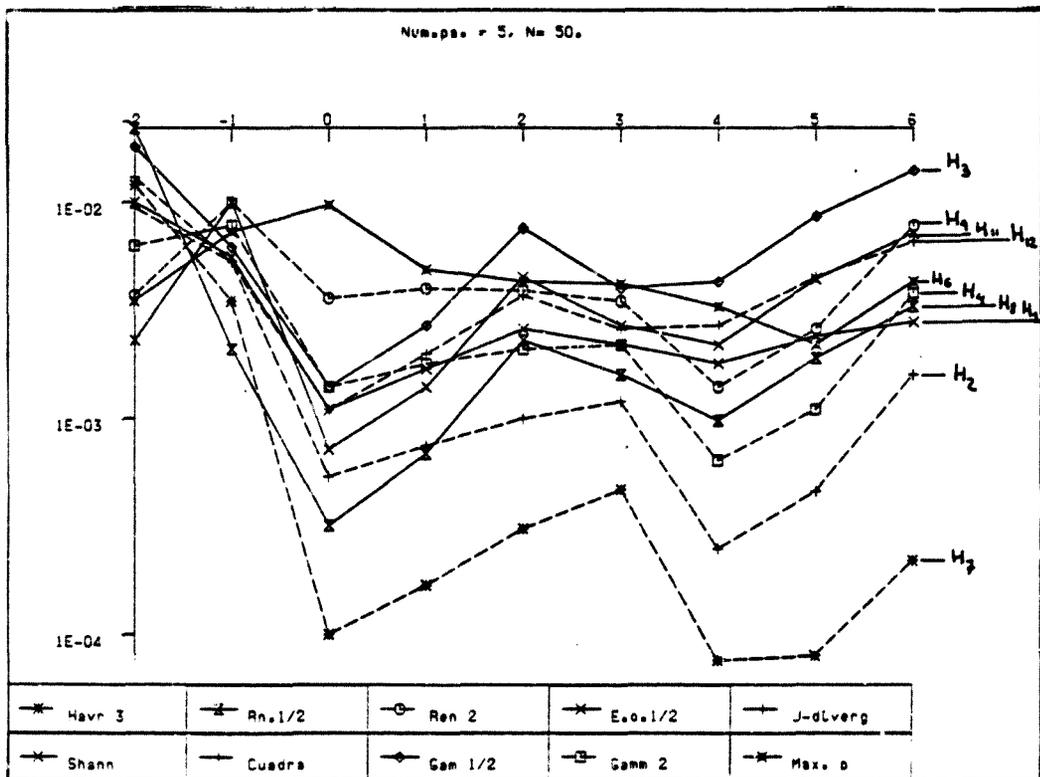
segundo generador (generador de Transworthe) para suministrar las 5000 semillas aleatorias independientes.

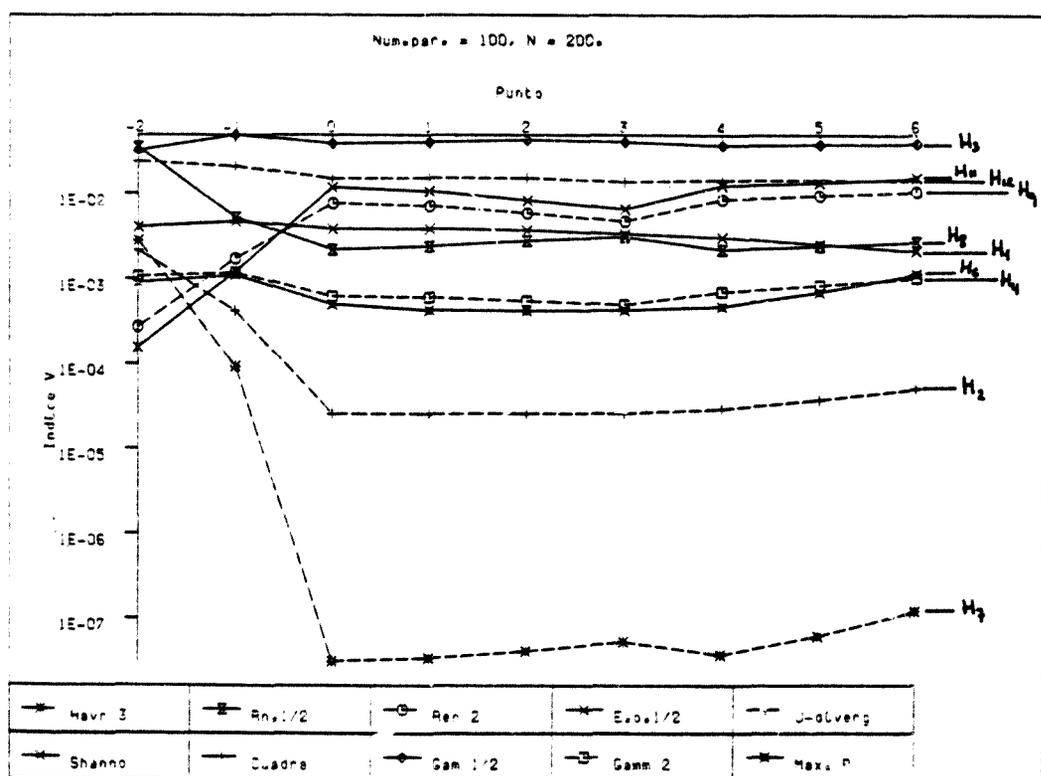
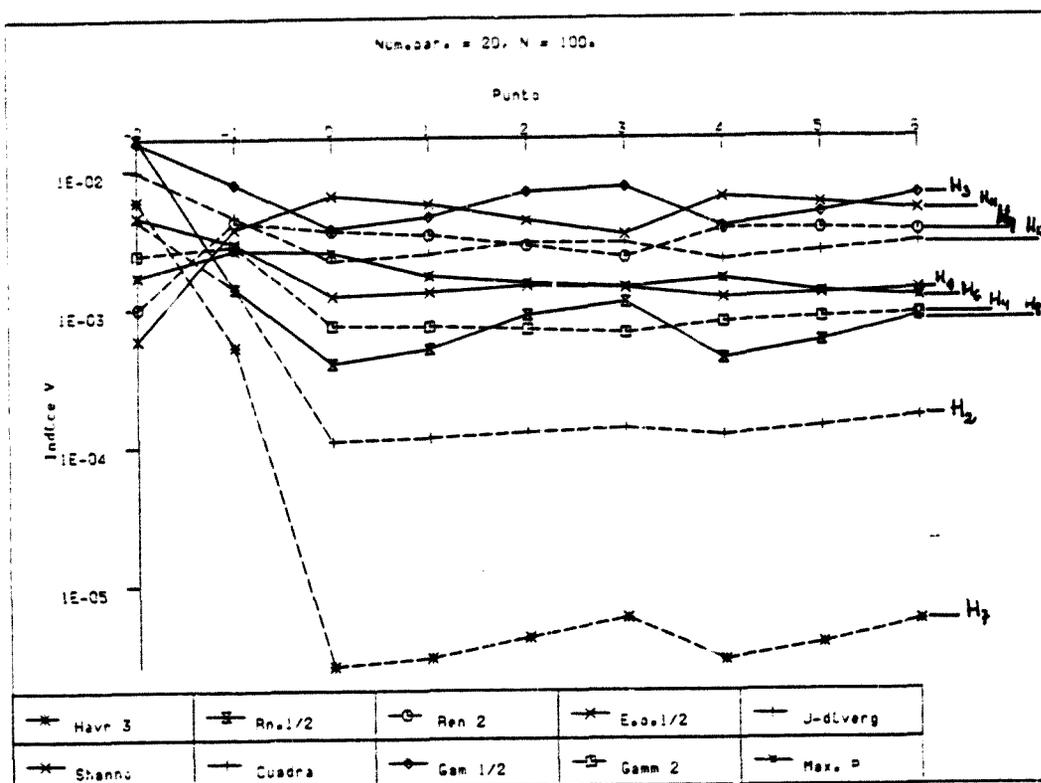
4. RESULTADOS OBTENIDOS

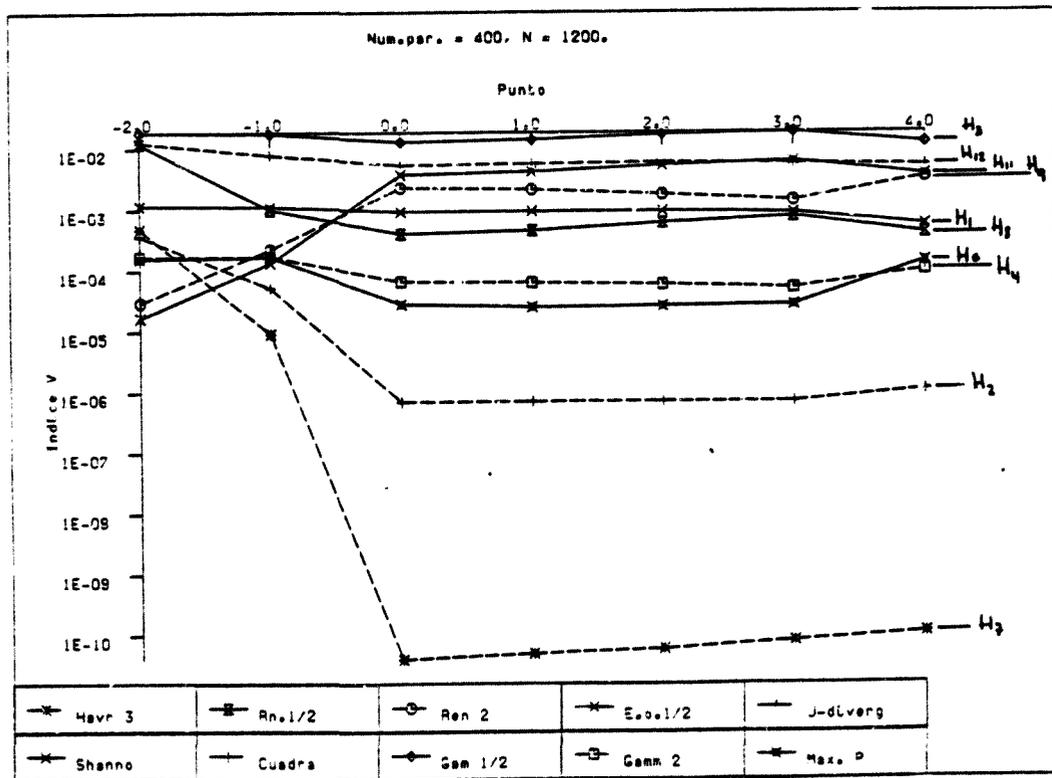
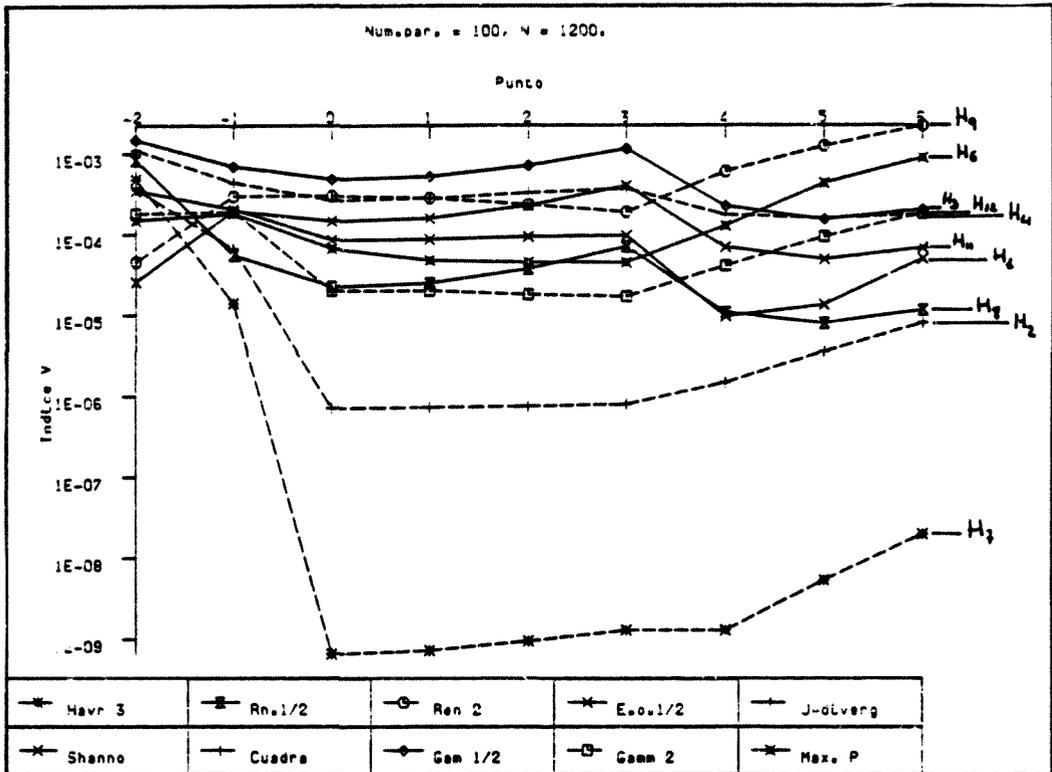
Realizadas las simulaciones para $N = \min \{20n, 1200\}$ y analizadas conjuntamente los resultados obtenidos para los estadísticos $\sigma^2(H)$ y V , hemos obtenido:

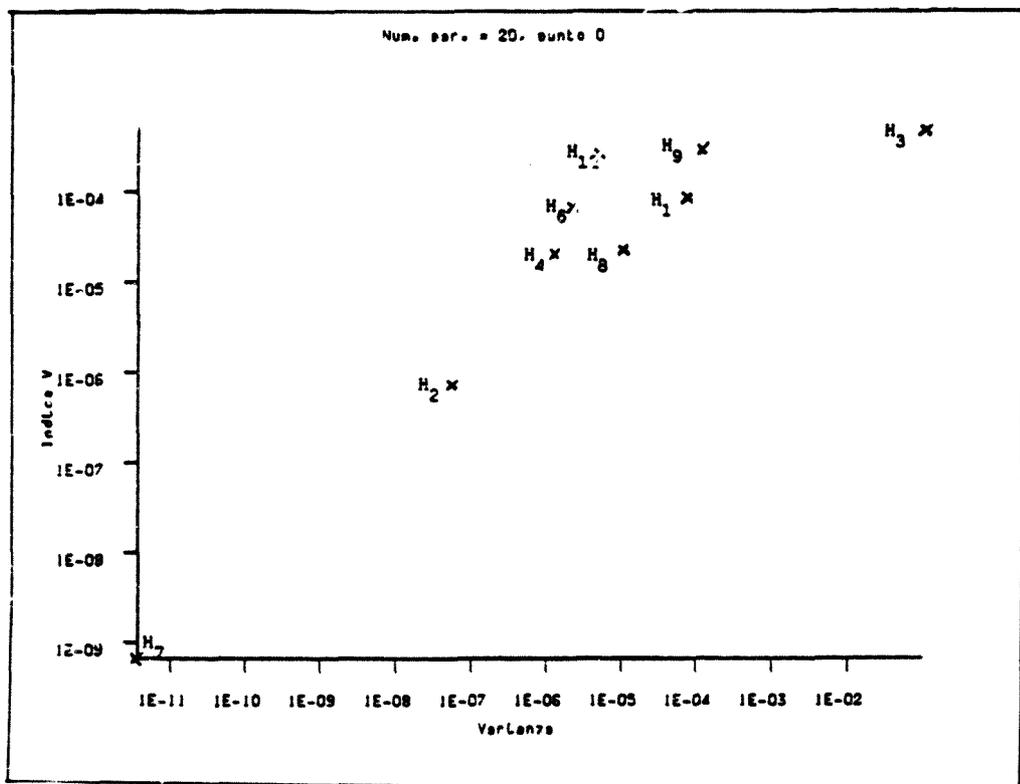
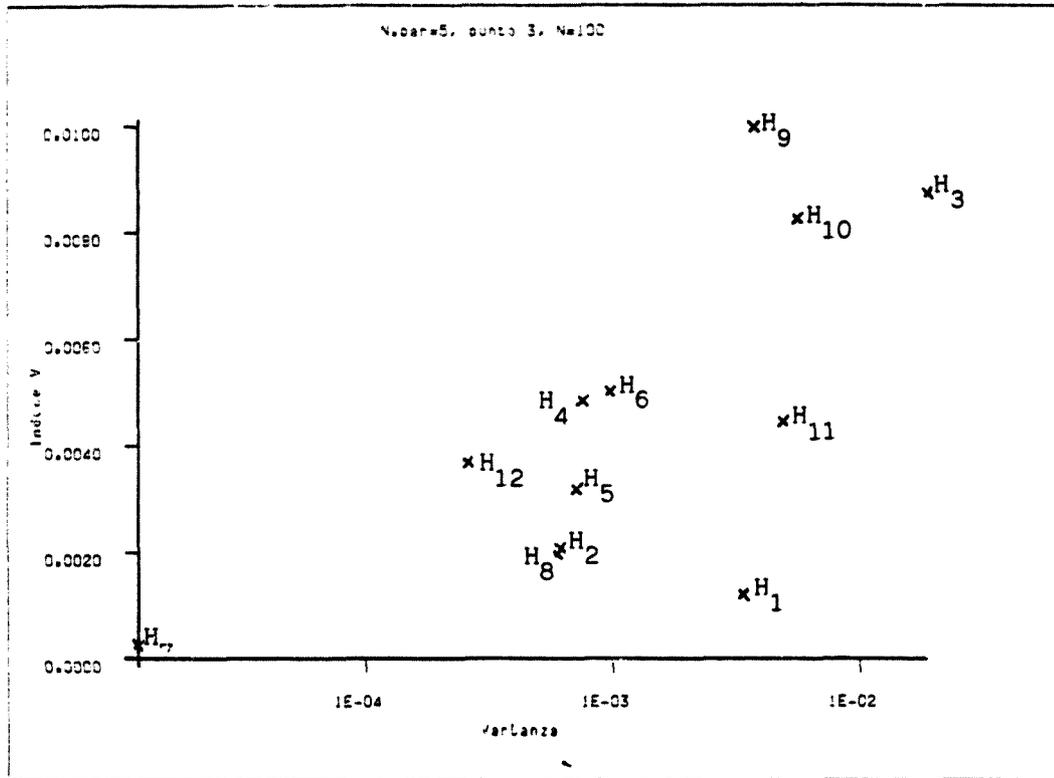
- 1) En los puntos $P^{(0)}$, $P^{(2)}$, $P^{(3)}$, $P^{(4)}$ y los intermedios que los enlazan, los índices H_7 y H_2 son los que tienen mejor estimación, teniendo en cuenta sobre todo que el valor del estadístico V es muy inferior al de los restantes índices.
- 2) En los puntos mencionados, el estimador 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.

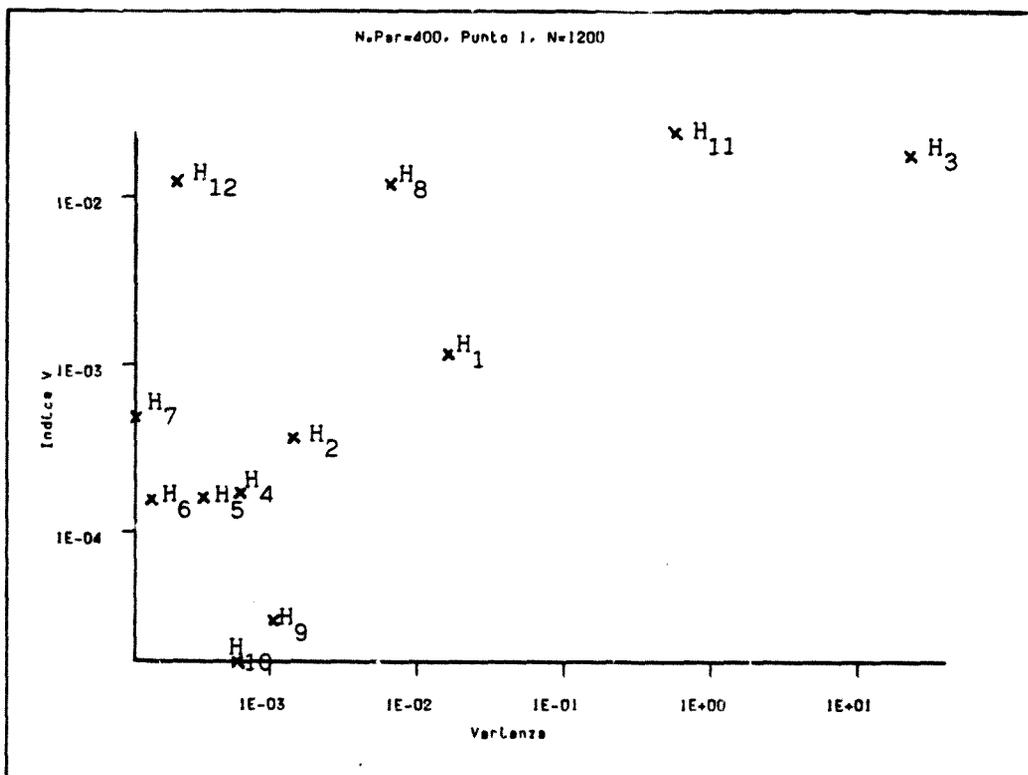
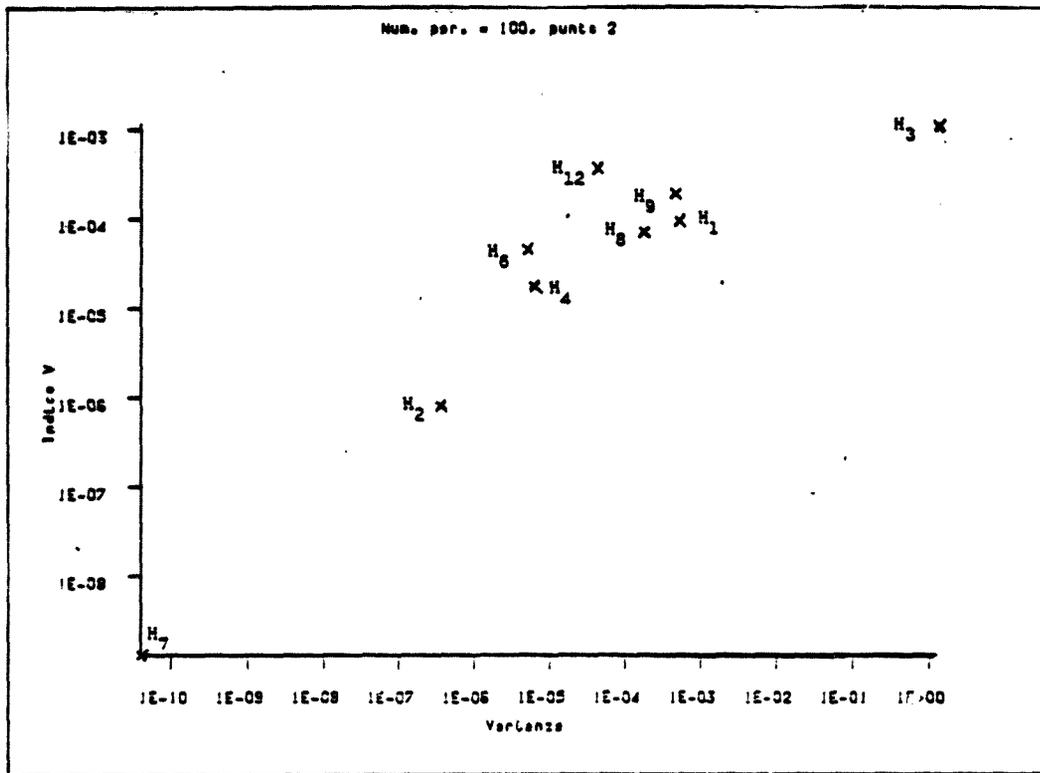
Estos resultados provienen de considerar para cada valor de n los distintos valores que toman los estadísticos $\sigma^2(H)$ y V a lo largo de las geodésicas que enlazan dos o más puntos. A título ilustrativo presentamos a continuación varias gráficas en las que podemos observar la variación de los estadísticos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











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 $N=\min(5n,400)$.

5. NOTA COMPLEMENTARIA

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 4000 veces, siendo el valor medio de la incertidumbre de estas 4000 repeticiones el que hemos correlacionado con el valor teórico. El proceso se ha repetido para valores del parámetro recorriendo el conjunto $\{5,10,20\}$.

		VALOR DEL PARAMETRO			
		5	10	20	
S E C I D I N I	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	0,8141	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	

Tabla 1

A la vista de los resultados para el coeficiente de correlación, podemos dividir los índices en cuatro grupos:

$$G.1 = \{H_2, H_4, H_9\} \quad G_2 = \{H_1, H_{12}, H_5, H_7, H_{10}\}$$

$$G.3 = \{H_3, H_8, H_{11}\} \quad G_4 = \{H_6\}$$

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 asociados 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.

ANEXO I

PROGRAMA INDICES σ Y V

```

C -----
SUBROUTINE OUTP(C,W)

C   Escritura en el fichero de salida de la matriz W con el titulo 'C'

COMMON /E/ ETIQU(18)
REAL*8 W(18,5)
CHARACTER*17 C
CHARACTER*10 ETIQU

WRITE (6,1) C

DO 100 I = 1,18
  WRITE (6,2) ETIQU(I),(W(I,J),J=1,5),ETIQU(I)
100 CONTINUE

1   FORMAT(/, ' - ',A17,/)
2   FORMAT(1X,A10,5(2X,G20.10),2X,A10)

RETURN
END

```

```

C -----
BLOCK DATA INITI

C   Inicializaciones de variables common

COMMON /E/ ETIQU(18)
COMMON /CHISQ/ CHMEAN(5),CHMAX(5)
REAL*4 CHMEAN,CHMAX
CHARACTER*10 ETIQU

DATA ETIQU /'Shannon :','Cuadrati.:', 'Gamma 1/2:', 'Gamma 2 :',
1          'Gamma 3 :','Max. p(i):',
2          'Havrda 2 :','Havrda 3 :',
3          'Renyi 1/2:', 'Renyi 2 :', 'Renyi 3 :',
4          'E.ord.1/2:', 'E.orden 2:', 'E.orden 3:',
5          'J-diverg.:', 'J-diver 2:', 'J-diver 3:', 'J-div. -1:/'

DATA CHMAX /5*0./,CHMEAN /5*0./

END

```

```

C -----
SUBROUTINE ITAUSW(LL,IX)

C   Generador de Tausworthe modificado de numeros aleatorios.
C   Entrada:
C     - LL = entero empleado como semilla
C   Salida :
C     - LL = IX = entero pseudo-aleatorio (ver main utilidad ix)
C   Referencia: Bratley et alt. "A Guide to Simulation" - Springer-Verlag

INTEGER*4 I, ICOMP, IX, J, JCOMP, LL
LOGICAL A, B, ACOMP, BCOMP

```

```
EQUIVALENCE(I,A),(J,B),(ACOMP,ICOMP),(BCOMP,JCOMP);
PARAMETER (N = 262144,M = 8192,K = 2147483647)
```

```
I = IL
J = I / M
ICOMP = K - I
JCOMP = K - J
```

```
B = A .AND. BCOMP .OR. ACOMP .AND. B
```

```
I = J
J = J * N
ICOMP = K - I
JCOMP = K - J
```

```
A = A .AND. BCOMP .OR. ACOMP .AND. B
```

```
IX = I
LL = I
```

```
RETURN
END
```

C

```
FUNCTION UNIF(IX)
```

```
C   Generador de numeros aleatorios, utiliza la formula recursiva:
C       IX = 16807 * IX MOD (2**31 -1)
C   Entrada:
C       - IX   = semilla del generador,  0 < IX <= 2147483647
C   Salida:
C       - UNIF = real pseudo-aleatorio entre 0 y 1.
C   Referencia: Bratley et alt. "A Guide to Simulation" - Springer-Verlag
```

```
INTEGER*4 IX, K1
```

```
K1 = IX / 127773
IX = 16807 * (IX - K1 * 127773) - K1 * 2836
IF (IX .LT. 0) IX = IX + 2147483647
```

```
UNIF = IX * 4.656612875E-10
```

```
RETURN
END
```

C

```
FUNCTION IVDISC(U)
```

```
C   Invierte una funcion de distribucion discreta utilizando un
C   vector de indices.
```

```
C   Entrada :
C       - Common /INTERN/
C       - U   = real tal que:  0 < U <= 1
```

```
C   Salida :
C       - IVDISC = F inversa de U.
C   Referencia: Bratley et alt. "A Guide to Simulation" - Springer-Verlag
```

```

COMMON /INTERN/ N,M,L(1500),CDF(500)
REAL*4 CDF,U
INTEGER L,M,N,NDX,JVDISC

IF (U .GE. CDF(N)) THEN
  IVDISC = N
  RETURN
END IF

NDX = U * M + 1.
JVDISC = L(NDX)

IF (JVDISC .GT. 0) THEN
  IVDISC = JVDISC
  RETURN
END IF

JVDISC = - JVDISC

100 IF (CDF(JVDISC) .LT. U) THEN
  JVDISC = JVDISC + 1
  GOTO 100
END IF

IVDISC = JVDISC
RETURN

END

```

C -----

SUBROUTINE VDSETP

C Inicializa el vector de probabilidades acumuladas CDF y el vector
C de pointers L para la inversion de la funcion de de distribucion
C discreta CDF, realizada mediante la funcion IVDISC.

C Entrada: mediante el common /PROBAB/

C - N2 = numero de sucesos (tambien tamaño util de P)

C - P = vector de probabilidades

C Salida: ccnstruccion del common /INTERN/

C - CDF = vector de probatibilidades acumuladas

C - L = vector de pointers.

C - M = tamaño util del vector L

C - N = tamaño util de CDF = N2

C Referencia: Bratley et alt. "A Guide to Simulation" - Springer-Verlag

```

COMMON /INTERN/ N,M,L(1500),CDF(500)
COMMON /GLOBAL/ N2,IX,RIN,D,S,P(500),A(500),B(500),F(500)
REAL*4 CDF,D,P,S,RIN,A,B,F
INTEGER I,IX,J,L,N,N2,NDX,ND1,NOLD

```

N = N2

M = 2 * N

CDF(1) = P(1)

DO 100 I = 2, N

```

      CDF(I) = CDF(I-1) + P(I)
100  CONTINUE

      DO 200 I = 1, M
        L(I) = 1
200  CONTINUE
      IF (N .EQ. 1) RETURN

      NOLD = M
      I = N

      DO 400 J = 2, N
        I = I - 1

        NDX = CDF(I)*M + 1
        L(NDX) = - I
        ND1 = NDX + 1

        IF (ND1 .LE. NOLD) THEN
          DO 300 K = ND1, NOLD
            L(K) = I + 1
300        CONTINUE
          END IF
          NOLD = NDX - 1
400  CONTINUE

      RETURN
      END

```

C -----

SUBROUTINE ENTROP(N,Y,H)

C Calculo J-divergencia y entropias.

C Entrada:

C - Y = vector de probabilidades (o frecuencias relativas)

C - N = tamaño util del vector Y (N <= 500)

C Salida:

C - H = vector que contiene las entropias y la J-divergencia

REAL*4 Y(500)

REAL*8 P,Q,AUX,H(18),SQR2,C1D3,C71,C73,LO2,DN,DRIN

PARAMETER(C1D3 = 1.D0/3.D0,C73 = -4.D0/3.D0,

1 SQR2 = 1.41421356237309515,C71 = 2.41421356237309448,

2 LO2 = 0.301029995663981212)

C ---- SQR2 = sqrt(2), C71 = 1/(sqrt(2)-1), LO2 = log10(2)

INTEGER I,N

DN = DFLOAT(N)

DRIN = 1 .D0 / DFLOAT(N)

DO 100 I = 1,18

 H(I) = 0.D0

100 CONTINUE

C ----- Calculo en acumuladores

```

DO 200 I = 1,N
  P = DBLE(Y(I))
  Q = P + DRIN
  IF (P .NE. 0.) THEN
    AUX = (P * DLOG10(P) - Q * DLOG10(Q)) / LO2
    H(1) = H(1) - P * DLOG10(P) / LO2
    H(2) = H(2) + P**2
    H(3) = H(3) + DSQRT(P)
    H(5) = H(5) + P**3
    H(6) = DMAX1(H(6),P)
    H(15) = H(15) + AUX
    H(16) = H(16) + AUX**2
    H(17) = H(17) + AUX**3
  ELSE
    AUX = Q * DLOG10(Q) / LO2
    H(15) = H(15) - AUX
    H(16) = H(16) + AUX**2
    H(17) = H(17) - AUX**3
  END IF

```

200 CONTINUE

C ----- calculo definitivo coeficientes

```

H(4) = H(2)
H(2) = 2.D0 - 2.D0 * H(2)
H(6) = 1.D0 - H(6)

H(18) = 1.D0/(H(15) + DLOG10(5.*DRIN) / LO2)
H(15) = -.5D0*H(15) + (DLOG10(DN)-(.5+.5*DRIN)*DLOG10(DN+1))/ LO2

H(7) = 1.D0 - H(4)
H(8) = .5D0*(1.D0 - H(5))
H(9) = 2.D0 * DLOG(H(3))
H(10) = -LOG(H(4))
H(11) = -.5D0*LOG(H(5))
H(12) = C71*(H(3)-1.D0)
H(13) = -2.D0*(H(4)-1.D0)
H(14) = C73*(H(5)-1.D0)
H(3) = H(3)**2 - 1.D0
H(4) = 2.D0*(1.D0 - DSQRT(H(4)))
H(5) = 1.5D0*(1.D0 - H(5)**C1D3)

RETURN
END

```

C -----

FUNCTION CHISQU(P,X,N)

C Efectua un test ji-cuadrado asumiendo distribucion multinomial
C de parametros (p(1),...,p(N)).

C Entrada:

C - P = vector de probabilidades
C - X = vector de frecuencias relativas observadas

```

C      - N = tamaño util de los vectores P y X (N <= 500)
C      Salida :
C      - CHISQU = valor del estadístico ji-cuadrado

```

```

REAL*4 P(500),X(500),CH
INTEGER I,N

```

```

CH = 0.

```

```

DO 100 I = 1,N
  CH = CH + (X(I) - P(I))**2/P(I)
100 CONTINUE

```

```

CHISQU = CH

```

```

RETURN
END

```

```

C -----

```

SUBROUTINE MULTIN

```

C      Calcula la suma y la suma al cuadrado de entropías y J-divergencia
C      correspondientes a una muestra de tamaño TM de N variables aleato-
C      rias con distribución conjunta multinomial de parámetros (p(1),p(2),
C      ...,p(N),TM).

```

```

C      La muestra se simula mediante extracción de números aleatorios.
C      TM se hace variar desde POPSIZ(1) hasta POPSIZ(5), evaluando en cada
C      caso las entropías. Si se desea un test ji-cuadrado de la muestra
C      para verificar su distribución, quitar las marcas de comentario
C      C*** en las líneas 382 a 384.

```

```

C      Funciones y subrutinas utilizadas:

```

```

C      - UNIF, IVDISC (CHISQU si se quita C****)

```

```

C      Entrada: mediante los common /GLOBAL/ y /F/

```

```

C      - N = tamaño del vector X

```

```

C      - IX semilla del generador UNIF

```

```

C      - POPSIZ = vector con el rango muestral

```

```

C      Salida : mediante el common /F/

```

```

C      - MEANH = matriz con las sumas de los 18 estadísticos en las 5
C      extracciones.

```

```

C      - VARH = idem con la suma al cuadrado.

```

```

C      mediante el common /CHISQ/

```

```

C      - CHMEAN = vector acumulado de estadísticos ji-cuadrado para
C      cada tamaño muestral.

```

```

C      - CHMAX = vector con el valor máximo hallado en TODAS las re-
C      plicas (no solo en esta llamada) para cada TM.

```

```

COMMON /GLOBAL/ N,IX,RIN,D,S,P(500),A(500),B(500),F(500)
COMMON /F/ POPSIZ(6),HT(18),RANK(18),MEANH(18,5), VARH(18,5),
1      VINDEXT(18,5)

```

```

COMMON /CHISQ/ CHMEAN(5),CHMAX(5)

```

```

REAL*4 U,RIN,P,S,A,B,F,CHMEAN,CHMAX,RTM,X(500)
REAL*8 HT,RANK,MEANH,VARH,VINDEXT,H(18)
INTEGER COUNT(500),DF,I,N,IX,POPSIZ,START

```

```

START = 1

DO 100 I = 1,N
  COUNT(I) = 0
100 CONTINUE

DO 500 I = 1,5

  DO 200 J = START,POPSIZ(I)
    U = UNIF(IX)
    LOCATE = IVDISC(U)
    COUNT(LOCATE) = COUNT(LOCATE) + 1
200 CONTINUE
C -----> el bucle 200 simula una muestra de tamaño POPSIZ(I)

  RTM = REAL(POPSIZ(I))
  DO 300 J = 1,N
    X(J) = REAL(COUNT(J)) / RTM
C -----> paso a frecuencias relativas
300 CONTINUE

  CALL ENTROP(N,X,H)
C -----> calculo entropias y J-Divergencia muestrales
  DO 400 J = 1,18
    MEANH(J,I) = MEANH(J,I) + H(J)
    VARH(J,I) = VARH(J,I) + H(J)**2
    VINDEXT(J,I) = VINDEXT(J,I) + ((H(J) - HT(J))/RANK(J))**2
C -----> sumar al acumulado
400 CONTINUE

  START = POPSIZ(I) + 1
C -----> siguiente punto de arranque

C*** CH = CHISQU(P,X,N)
C*** CHMEAN(I) = CHMEAN(I) + CH
C*** CHMAX(I) = MAX(CHMAX(I),CH)

500 CONTINUE

RETURN
END

```

C -----

SUBROUTINE FCALCU(KIND)

C Construcción del punto F, punto final de la geodesica

```

REAL*4 D,P,RIN,S,A,B,F,DELTA
REAL*8 HT,RANK,MEANH,VARH,VINDEXT
INTEGER I,KIND,N,M,POPSIZ

```

```

COMMON /GLOBAL/ N,IX,RIN,D,S,P(500),A(500),B(500),F(500)
COMMON /F/ POPSIZ(6),HT(18),RANK(18),MEANH(18,5), VARH(18,5),
1 VINDEXT(18,5)

```

IF (2*(N/2) .LT. N) N = N + 1

```

C      ----> las particiones han de ser pares

      M = N/2
      RIN = 1. / REAL(N)

      DELTA = 2. * RIN**2
C      -----> delta maximo posible en los casos 2 y 3
      DELTA = DELTA - DELTA/1000.

      DO 100 I = 1,N
        P(I) = RIN
100    CONTINUE
C      -----> se asigna inicialmente el punto 0 = mult. equiprobable
      CALL ENTROP(N,P,RANK)
C      -----> punto de maxima entropia.

      F(1) = 1.
      DO 200 I = 2,N
        F(I) = 0.
200    CONTINUE
      CALL ENTROP(N,F,HT)
C      -----> punto de minima entropia, tambien valido como punto final
C      en el caso 1.

      DO 300 I = 1,18
        RANK(I) = RANK(I) - HT(I)
300    CONTINUE

      IF (KIND .EQ. 2) THEN

        DO 400 I = 1,M
          F(I) = RIN + (M+1-I)*DELTA
          F(I+M) = RIN - I*DELTA
400    CONTINUE

        ELSE IF (KIND .EQ. 3) THEN

          DO 500 I = 1,M
            F(I) = 1.8 * RIN
            F(I+M) = .2 * RIN
            P(I) = RIN + (M+1-I)*DELTA
            P(I+M) = RIN - I*DELTA
500    CONTINUE

        ELSE IF (KIND .EQ. 4) THEN
          CALL NORMAL
        END IF

      RETURN
      END

C -----
      SUBROUTINE NORMAL

C      Tipo de problema 4: particion de la normal (0,1) en N intervalos
C      de probabilidad f(i)

```

```

REAL*4 D,F,RIN,S,A,B,F
REAL*16 ISIZE,STEP,AUX,CAUX,SQR2
INTEGER I,IX,N

```

```

COMMON /GLOBAL/ N,IX,RIN,D,S,P(500),A(500),B(500),F(500)

```

```

AUX = 0.00
STEP = 0.00
SQR2 = QSQR(2.00)
ISIZE = 4.00 / QFLOAT(N/2)

```

```

DO 100 I = 1,N,2

```

```

    STEP = STEP + ISIZE
    CAUX = QERF(STEP/SQR2)
    F(I) = SNGLQ(5.0-1*(CAUX-AUX))
    F(I+1) = F(I)
    AUX = CAUX

```

```

100 CONTINUE

```

```

    RETURN
    END

```

C-----

```

SUBROUTINE DISTAN(KIND)

```

```

C   Calcula la distancia entre el punto O = (p(1),...,p(n)) y el punto
C   F en el espacio parametrico de las distr. multinomiales, asi como
C   las constantes de integracion A y B.

```

```

COMMON /GLOBAL/ N,IX,RIN,D,S,P(500),A(500),B(500),F(500)

```

```

REAL*4 RIN,D,P,S,A,B,F,AUX,BUX
INTEGER I,KIND

```

```

AUX = 0.
D = 0.

```

```

DO 100 I = 1,N
    AUX = AUX + SQRT(P(I)*F(I))

```

```

100 CONTINUE

```

```

S = ACOS(AUX)

```

```

C   -----> s = distancia(O-F) / (2 * SQRT(Tamaño muestral))

```

```

BUX = SQRT(1. - AUX**2)

```

```

DO 200 I = 1,N
    A(I) = SQRT(P(I))
    B(I) = (SQRT(F(I)) - SQRT(P(I))*AUX) / BUX

```

```

200 CONTINUE

```

```

    RETURN
    END

```

C-----

SUBROUTINE GEODES

C Desplaza el punto $P = (p(1), \dots, p(n), N)$ a lo largo de la curva G
 C una distancia s. La curva G es la geodesica que une en la variedad
 C parametrica los puntos O y F.

C Entrada:

C - S = incremento de arco en la direccion O - F en la geodesica G
 C y mediante el common /GLOBAL/
 C - P = vector de probabilidades (punto P)
 C - N = n, numero de variables aleatorias observadas
 C - D = Jis^2 variancia O - C / sqrt (tamaño muestral)

C Salida:

C - P = vector P en D + S
 C - D = D + S

COMMON /GLOBAL/ N,IX,RIN,D,S,P(500),A(500),B(500),F(500)
 REAL*4 RIN,D,P,S,A,B,F,AUX,BUX
 INTEGER IX,N

D = D + S
 AUX = SQRT(RIN)*COS(D)
 BUX = SIN(D)

DO 100 I = 1,N
 P(I) = (AUX + B(I)*BUX)**2
 100 CONTINUE

RETURN
 END

C -----

PROGRAM MAIN

C
 C Estudio, mediante simulacion, de algunas propiedades muestrales
 C de ciertas entropias (ver BLOCK DATA INITI) en una serie de puntos
 C , elegidos por el usuario, de la variedad parametrica asociada
 C a las funciones de densidad multinomiales.

C Entrada:

C - N = numero de sucesos
 C - NR = numero de replicas
 C - ISG = numero de pasos intermedios entre los puntos O y F
 C - KIND= tipo de problema.

C Salida : para cada punto

C - HT = entropias y J-divergencia teoricas
 C - MEANH = media de las NR entropias y J-divergencia muestrales
 C - VARH = varianza de las idem
 C - SIGTIP = desviacion tipica de las idem
 C - SESGO = sesgo tipificado de las idem

COMMON /GLOBAL/ N,IX,RIN,D,S,P(500),A(500),B(500),F(500)
 COMMON /F/ POPSIZ(6),HT(18),RANK(18),MEANH(18,5), VARH(18,5),
 1 VINDEX(18,5)
 COMMON /CHISQ/ CHMEAN(5),CHMAX(5)

```

COMMON /E/ ETIQU(18)

REAL*4  A,E,F,D,P,RIN,S,CHMEAN,CHMAX
REAL*8  MEANH,VARH,H(18),ET,RNR,RANK,SIGTIP(18,5),SESGO(18,5),
1      VINDEK
INTEGER I,LL,N,NR,POPSIZ,STP
CHARACTER*17 C
CHARACTER*10 ETIQU

WRITE(5,1)
1  FORMAT (' Entra tamaño del vector. ')
   READ(5,*) N

   IF ((N .GT. 500) .OR. (N .LE. 1)) THEN
       WRITE(5,2)
2      FORMAT (' Numero de parametros incorrecto: 1 < N <= 500')
       STOP
   END IF

   WRITE(5,3)
3      FORMAT (' Entra numero de replicas. ')
   READ(5,*) NR
   WRITE(5,4)
4      FORMAT (' Entra tipo de problema escogido. ')
   READ(5,*) KIND

   IF ((KIND .LT. 0) .OR. (KIND .GT. 4)) THEN
       WRITE(5,5)
5      FORMAT (' Tipo de problema incorrecto (solo 0, 1, 2, 3, 4)')
       STOP
   END IF

   IF (KIND .GT. 0) THEN
       WRITE(5,6)
6      FORMAT (' Entra numero de saltos en la geodesica. ')
       READ(5,*) ISG
       IF ((KIND .EQ. 1) .AND. (ISG .EQ. 1)) STOP
   ELSE
       ISG = 1
   END IF

   WRITE(6,7) KIND,NR,N

   RNR = DBLE(NR)

   POPSIZ(1) = MIN0(N,100)
   POPSIZ(2) = MIN0(2*N,200)
   POPSIZ(3) = MIN0(5*N,400)
   POPSIZ(4) = MIN0(10*N,800)
   POPSIZ(5) = MIN0(20*N,1200)

   CALL FCALCU(KIND)
C      -----> asignaciones iniciales puntos inicial (O) y final (F).
   CALL DISTAN(KIND)
C      -----> calcula distancia (O-F), y la ctes. A y B
   IF (ISG .GT. 0) S = S / REAL(ISG)
C      -----> s = distancia(O-F) / isg / (2 * SQRT(TM))

```

```

IF (KIND .EQ. 1) ISG = ISG - 1
C      -----> no tiene sentido el caso (1,0,....,0)

DO 800 STP = 1,ISG

IF (KIND .GT. 0) CALL GEODES
C      -----> Desplazar el punto actual P hacia F 's' unidades

LL = 524287
C      -----> valor inicial para el generador de Tausworthe
CALL VDSETP
C      -----> inicializar tabla de indices
CALL ENTROP(N,P,HT)
C      -----> calculo entropias y J-divergencia teoricas

DO 200 I = 1,18
DO 200 J = 1,5
    MEANH(I,J) = 0.
    VARH(I,J) = 0.
    VINDEK(I,J) = 0.
200 CONTINUE
C      -----> inicializacion media y varianza

DO 300 J = 1,NR
CALL ITAUSW(LL,IX)
C      -----> LL es empleada como semilla inicial de cada replica
CALL MULTIN
C      -----> simulacion muestra 'J' y calculo entropias
300 CONTINUE

IF (KIND .NE. 0) THEN
    WRITE(6,8) D,STP
ELSE
    WRITE(6,*)
END IF

DO 400 I = 1,18
    WRITE(6,9) ETIQUE(I),HT(I)
400 CONTINUE
WRITE (6,10) (POPSIZ(I),I=1,5)

DO 600 J = 1,5

DO 500 I = 1,18
    MEANH(I,J) = MEANH(I,J) / RNR
    VARH(I,J) = VARH(I,J) / RNR - MEANH(I,J)**2
    VINDEK(I,J) = VINDEK(I,J) / RNR
    IF (VARH(I,J) .LT. 0.DO) VARH(I,J) = 0 DO
        SESGO(I,J) = (HT(I) - MEANH(I,J))/ HT(I) * 100.DO
        SIGTIP(I,J) = DSQRT(VARH(I,J)) / HT(I) * 100.DO
500 CONTINUE

600 CONTINUE

C = ' Media (m) de:'
CALL OUTP(C,MEANH)
C = ' Varianza de:'
CALL OUTP(C,VARH)

```

```
C = ' Sigma tipi. de:'  
CALL OUTP(C,SIGTIP)  
C = ' (E-m) / E de:'  
CALL OUTP(C,SESGO)  
C = ' Nuevo indice:'  
CALL OUTP(C,VINDEX)  
  
800 CONTINUE  
  
7   FORMAT(/,30X,' Tipo de problema ---->',G6,  
1   /,30X,' Numero de simulaciones:',G6,  
2   /,30X,' Numero de parametros  :',G6)  
8   FORMAT (/,30X,' Distancia:',F6.3,' ----> salto num. ', G3,  
1   /,' Valor teorico de'/)  
9   FORMAT(1X,A10,5(2X,G20.10))  
10  FORMAT(// ' Tamaño poblacional :',/,10X,5(10X,G4,8X))  
  
STOP  
END
```

ANEXO II

PROGRAMA INDICE CORRELACION SPEARMAN

C -----

SUBROUTINE ITAUSW(LL,IX)

C Generador de Tausworthe modificado de numeros aleatorios.
 C Entrada:
 C - LL = entero empleado como semilla
 C Salida :
 C - LL = IX = entero pseudo-aleatorio (ver main utilidad ix)
 C Referencia: Bratley et alt. "A Guide to Simulation" - Springer-Verlag

```
INTEGER*4 I,ICOMP,IX,J,JCOMP,LL
LOGICAL A,B,ACOMP,BCOMP
EQUIVALENCE(I,A),(J,B),(ACOMP,ICOMP),(BCOMP,JCOMP)
PARAMETER (N = 262144,M = 8192,K = 2147483647)
```

```
I = LL
J = I / M
ICOMP = K - I
JCOMP = K - J
```

```
B = A .AND. BCOMP .OR. ACOMP .AND. B
```

```
I = J
J = J * N
ICOMP = K - I
JCOMP = K - J
```

```
A = A .AND. BCOMP .OR. ACOMP .AND. B
```

```
IX = I
LL = I
```

```
RETURN
END
```

C -----

FUNCTION UNIF(IX)

C Generador de numeros aleatorios, utiliza la formula recursiva:
 C $IX = 16807 * IX \text{ MOD } (2^{**}31 - 1)$
 C Entrada:
 C - IX = semilla del generador, $0 < IX \leq 2147483647$
 C Salida:
 C - UNIF = real pseudo-aleatorio entre 0 y 1.
 C Referencia: Bratley et alt. "A Guide to Simulation" - Springer-Verlag

```
INTEGER*4 IX, K1
```

```
K1 = IX / 127773
IX = 16807 * (IX - K1 * 127773) - K1 * 2836
IF (IX .LT. 0) IX = IX + 2147483647
```

```
UNIF = IX * 4.656612875E-10
```

```
RETURN
END
```

C -----

```
      FUNCTION IVDISC(U)
C      Invierte una funcion de distribucion discreta utilizando un
C      vector de indices.

C      Entrada :
C      - Common /INTERN/
C      - U = real tal que: 0 < U <= 1
C      Salida :
C      - IVDISC = F inversa de U.
C      Referencia: Bratley et al. "A Guide to Simulation" - Springer-Verlag
```

```
      COMMON /INTERN/ N,M,L(1500),CDF(500)
      REAL*4 CDF,U
      INTEGER L,M,N,NDX,JVDISC

      IF (U .GE. CDF(N)) THEN
          IVDISC = N
          RETURN
      END IF

      NDX = U * M + 1.
      JVDISC = L(NDX)

      IF (JVDISC .GT. 0) THEN
          IVDISC = JVDISC
          RETURN
      END IF

      JVDISC = - JVDISC

100  IF (CDF(JVDISC) .LT. U) THEN
          JVDISC = JVDISC + 1
          GOTO 100
      END IF

      IVDISC = JVDISC
      RETURN

      END
```

C -----

```
      SUBROUTINE VDSETP

C      Inicializa el vector de probabilidades acumuladas CDF y el vector
C      de pointers L para la inversion de la funcion de de distribucion
C      discreta CDF, realizada mediante la funcion IVDISC.

C      Entrada: mediante el common /PROBAB/
C      - N2 = numero de sucesos (tambien tamaño util de P)
C      - P = vector de probabilidades
C      Salida: construccion del common /INTERN/
C      - CDF = vector de probabilidades acumuladas
C      - L = vector de pointers.
C      - M = tamaño util del vector L
```

C - N = tamaño util de CDF = N2
 C Referencia: Bratley et alt. "A Guide to Simulation" - Springer-Verlag

```

COMMON /INTERN/ N,M,L(1500),CDF(500)
COMMON /GLOBAL/ N2,IX,P(500)
REAL*4 CDF,P
INTEGER I,IX,J,L,M,N,N2,NDX,ND1,NOLD

N = N2
M = 2 * N
CDF(1) = P(1)

DO 100 I = 2, N
  CDF(I) = CDF(I-1) + P(I)
100 CONTINUE

DO 200 I = 1, M
  L(I) = 1
200 CONTINUE
IF (N .EQ. 1) RETURN

NOLD = M
I = N

DO 400 J = 2,N
  I = I - 1

  NDX = CDF(I)*M + 1
  L(NDX) = - I
  ND1 = NDX + 1

  IF (ND1 .LE. NOLD) THEN
    DO 300 K = ND1,NOLD
      L(K) = I + 1
300 CONTINUE
    END IF
    NOLD = NDX - 1
400 CONTINUE

RETURN
END

```

C -----

SUBROUTINE ENTROP(N,Y,H)

C Calculo J-divergencia y entropias.

C Entrada:

C - Y = vector de probabilidades (o frecuencias relativas)
 C - N = tamaño util del vector Y (N <= 500)

C Salida:

C - H = vector que contiene las entropias y la J-divergencia

REAL*4 Y(500)

REAL*8 P,Q,AUX,H(18),SQR2,C1D3,C71,C73,LO2,DN,DRIN

PARAMETER(C1D3 = 1.D0/3.D0,C73 = -4.D0/3.D0,

1 SQR2 = 1.41421356237309515,C71 = 2.41421356237309448,

```

2          LC2 = 0.301029995663981212)

C      ---- SQR2 = sqrt(2), C71 = 1/(sqrt(2)-1), LO2 = log10(2)

      INTEGER I,N

      DN   = DFLOAT(N)
      DRIN = 1 .DO / DFLOAT(N)

      DO 100 I = 1,18
        H(I) = 0.DO
100    CONTINUE

C      ----- Calculo en acumuladores

      DO 200 I = 1,N
        P = DBLE(Y(I))
        Q = P + DRIN
        IF (P .NE. 0.) THEN
          AUX = (P * DLOG10(P) - Q * DLOG10(Q)) / LO2
          H(1) = H(1) - P * DLOG10(P) / LO2
          H(2) = H(2) + P**2
          H(3) = H(3) + DSQRT(P)
          H(5) = H(5) + P**3
          H(6) = DMAX1(H(6),P)
          H(15) = H(15) + AUX
          H(16) = H(16) + AUX**2
          H(17) = H(17) + AUX**3
        ELSE
          AUX = Q * DLOG10(Q) / LO2
          H(15) = H(15) - AUX
          H(16) = H(16) + AUX**2
          H(17) = H(17) - AUX**3
        END IF
200    CONTINUE

C      ----- calculo definitivo coeficientes

      H(4) = H(2)
      H(2) = 2.DO - 2.DO * H(2)
      H(6) = 1.DO - H(6)

      H(18) = 1.DO/(H(15) + DLOG10(5.*DRIN) / LO2)
      H(15) = -.5D0*H(15) + (DLOG10(DN)-(.5+.5*DRIN)*DLOG10(DN+1))/ LO2

      H(7) = 1.DO - H(4)
      H(8) = .5D0*(1.DO - H(5))
      H(9) = 2.DO * DLOG(H(3))
      H(10) = -LOG(H(4))
      H(11) = -.5D0*LOG(H(5))
      H(12) = C71*(H(3)-1.DO)
      H(13) = -2.DO*(H(4)-1.DO)
      H(14) = C73*(H(5)-1.DO)
      H(3) = H(3)**2 - 1.DO
      H(4) = 2.DO*(1.DO - DSQRT(H(4)))
      H(5) = 1.5D0*(1.DO - H(5)**C1D3)

```

```

RETURN
END

```

C -----

```

SUBROUTINE GENERA(SEMGEN)

```

C Genera un vector aleatorio de N componentes con suma = 1.

```

COMMON /GLOBAL/ N,IX,P(500)

```

```

REAL*4 P
INTEGER SEMGEN,I

```

```

CDF = 0.
CALL ITAUSW(SEMGEN,IX)

```

```

DO 100 I = 1,N
  P(I) = UNIF(IX)
  CDF = CDF + P(I)
100 CONTINUE

```

```

DO 200 I = 1,N
  P(I) = P(I) / CDF
200 CONTINUE

```

```

RETURN
END

```

C -----

```

SUBROUTINE SORT(NPOINT,E,IE)

```

C Ordena el vector de indices IE de acuerdo a la ordenacion
C ascendente del vector E.

C Entrada:

C - NPOINT = numero de objetos a ordenar, NPOINT <= 500
C - E = vector de objetos a ordenar
C - IE = vector de indices (inicializado a 1,2,...,NPOINT)

C Salida :

C - IE = posicion de menor a mayor valor de los elementos de E

```

REAL*8 E (500,18)
INTEGER*4 NPOINT,IE(500,18)

```

```

DO 100 NH = 1,18
  DO 100 I = 1,NPOINT-1
    DO 100 J = 1,NPOINT-I
      IF (E(IE(J,NH),NH) .GT. E(IE(J+1,NH),NH)) THEN
        IAUX = IE(J,NH)
        IE(J,NH) = IE(J+1,NH)
        IE(J+1,NH) = IAUX
      END IF
    CONTINUE
  CONTINUE
100 CONTINUE

```

```

RETURN
END

```

```

C -----
SUBROUTINE SPEARM(NPOINT,I1,I2,S)

C   Calcula la correlacion de Spearman entre los indices I1 e I2.
C   Entrada:
C     - NPOINT = numero de elementos de I1 e I2
C     - I1 = vector de indices 1.
C     - I2 = vector de indices 2.
C   Salida :
C     - S = correlacion de Spearman

REAL*8    S(18),RAUX
INTEGER*4  I1(500,18),I2(500,18),NPOINT

RAUX = DFLOAT(NPOINT**3-NPOINT)

DO 100 I = 1,18
    S(I) = 0.D0
100 CONTINUE

DO 200 I = 1,18
    DO 200 J = 1,NPOINT
        S(I) = S(I) + DFLOAT(I1(J,I) - I2(J,I))**2
200 CONTINUE

DO 300 I = 1,18
    S(I) = 1.D0 - (6.D0*S(I))/RAUX
300 CONTINUE

RETURN
END

```

```

C -----
SUBROUTINE MULTIN

C   Calcula la suma de entropias y J-divergencia correspondientes a
C   una muestra de tamaño POPSIZ de N variables aleatorias con distri-
C   bucion conjunta multinomial de parametros (p(1),p(2),...,p(N),POP).
C   La muestra se simula mediante extraccion de numeros aleatorios.

C   Funciones y subrutinas utilizadas:
C     - UNIF, IVDISC
C   Entrada: mediante los common /GLOBAL/ y /F/
C     - N = tamaño del vector X
C     - IX = semilla del generador UNIF
C   Salida : mediante el common /F/
C     - ACUMU = matriz con las sumas de los 18 estadisticos

COMMON /GLOBAL/ N,IX,P(500)
COMMON /F/ POPSIZ,II,ACUMU(18)

REAL*4  U,P,RTM,X(500)
REAL*8  ACUMU,H(18)
INTEGER COUNT(500),I,N,IX,POPSIZ

```

```

DO 100 I = 1,N
  COUNT(I) = 0
100 CONTINUE

DO 200 J = 1,POPSIZ
  U      = UNIF(IX)
  LOCATE = IVDISC(U)
  COUNT(LOCATE) = COUNT(LOCATE) + 1
200 CONTINUE
C      -----> el bucle 200 simula una muestra de tamaño POPSIZ

RTM = REAL(POPSIZ)
DO 300 J = 1,N
  X(J) = REAL(COUNT(J)) / RTM
C      -----> paso a frecuencias relativas
300 CONTINUE

CALL ENTROP(N,X,H)
C      -----> calculo entropias y J-Divergencia muestrales
DO 400 J = 1,18
  ACUMU(J) = ACUMU(J) + H(J)
C      -----> sumar al acumulado
400 CONTINUE

RETURN
END

C -----
PROGRAM MAIN

C
C Estudio, mediante simulacion, de la correlacion de Spearman entre
C las esperanzas muestral y teorica de una coleccion de entropias
C (ver DATA ETIQUE), calculadas en una serie de puntos elegidos
C al azar en la variedad parametrica de las funciones de densidad
C multinomiales.

C Entrada:
C - N      = numero de sucesos
C - NR     = numero de replicas (en cada punto).
C - POPSIZ = tamaño poblacional (numero de extracciones)
C - NPOINT = numero de puntos extraidos al azar en el espacio
C           de las multinomiales.
C Salida :
C - S      = vector con los coef. de correlacion de Spearman

COMMON /GLOBAL/ N,IX,P(500)
COMMON /F/ POPSIZ,II,ACUMU(18)

REAL*4 P
REAL*8 ACUMU,HT(18),RNR,ET(500,18),EM(500,18),S(18)
INTEGER I,LL,N,NR,NPOINT,POPSIZ,SEMGEM,IM(500,18),IT(500,18),
1 IS(18)
CHARACTER*10 ETIQUE(18)

DATA ETIQUE /'Shannon  :','Cuadrati.:','Gamma 1/2:','Gamma 2  :',
1 'Gamma 3  :','Max. p(i):'

```

```

2          'Havrda 2 :','Havrda 3 :',
3          'Renyi 1/2:','Renyi 2  :','Renyi 3  :',
4          'E.ord.1/2:','E.orden 2:','E.orden 3:',
5          'J-diverg.:','J-diver 2:','J-diver 3:','J-div. -1:/'

WRITE(5,1)
1  FORMAT (' Entra tamaño del vector.')
   READ(5,*) N

   IF ((N .GT. 500) .OR. (N .LE. 1)) THEN
     WRITE(5,2)
2    FORMAT (' Numero de parametros incorrecto: 1 < N <= 500')
     STOP
   END IF

   WRITE(5,3)
3  FORMAT (' Entra numero de replicas.')
   READ(5,*) NR
   WRITE(5,4)
4  FORMAT (' Entra tamaño poblacional.')
   READ(5,*) POPSIZ

   IF (POPSIZ .LE. 0) THEN
     WRITE(5,5)
5    FORMAT (' Tamaño poblacional incorrecto')
     STOP
   END IF

   WRITE(5,6)
6  FORMAT (' Entra numero de puntos.')
   READ(5,*) NPOINT

   IF (POPSIZ .LE. 0) THEN
     WRITE(5,7)
7    FORMAT (' Numero de puntos incorrecto')
     STOP
   END IF

WRITE(6,8) NR,N,POPSIZ,NPOINT

RNR    = DBLE(NR)
SEMGEN = 572586

DO 800 NPO = 1,NPOINT

   CALL GENERA(SEMGEN)
C      -----> genera un vector aleatorio con suma de componentes = 1

   LL = 524287
C      -----> valor inicial para el generador de Tausworthe
   CALL VDSETP
C      -----> inicializar tabla de indices
   CALL ENTROP(N,P,HT)
C      -----> calculo entropias y J-divergencia teoricas-

DO 200 I = 1,18
   ET(NPO,I) = HT(I)

```

```
        IT(NPO,I) = NPO
        ACUMU(I) = 0.D0
200    CONTINUE

        DO 300 J = 1,NR
            CALL ITAUSW(LL,IX)
C            -----> LL es empleada como semilla inicial de cada replica
            CALL MULTIN
C            -----> simulacion muestra 'J' y calculo entropias
300    CONTINUE

        DO 400 I = 1,18
            EM(NPO,I) = ACUMU(I) / RNR
            IM(NPO,I) = NPO
400    CONTINUE

800    CONTINUE

        CALL SORT(NPOINT,ET,IT)
        CALL SORT(NPOINT,EM,IM)

        CALL SPEARM(NPOINT,IT,IM,S)

        DO 900 I = 1,18
            IS(I) = I
900    CONTINUE

        DO 1000 I = 1,17
            DO 1000 J = 1,18-I
                IF (S(IS(J)) .GT. S(IS(J+1))) THEN
                    IAUX = IS(J)
                    IS(J) = IS(J+1)
                    IS(J+1) = IAUX
                END IF
1000    CONTINUE

        WRITE(6,*)
        WRITE(6,9) (ETIQUE(IS(I)),S(IS(I)),I=1,18)

8        FORMAT(/,30X,' Numero de simulaciones:',G6,
1          /,30X,' Numero de parametros  :',G6,
2          /,30X,' Tamaño poblacional   :',G6,
3          /,30X,' Numero de puntos    :',G6)
9        FORMAT(1X,A10,2X,G20.10)

        STOP
        END
```

VII. RESUMEN DE RESULTADOS

En la realización de esta memoria hemos obtenido los siguientes resultados:

1) Hemos caracterizado la simetría de la J-divergencia generalizada en el siguiente sentido:

a) $J_{\phi}^{\lambda}(p, q)$ es simétrica en p y q para toda función $\phi(t)$ de la forma $\phi(t) = at^2 + bt + c$

b) $J_{\phi}^{\lambda}(p, q)$ es simétrica en p y q para toda función $\phi(t) = at^2 + bt + c$ si y solo si $\lambda = \frac{1}{2}$

2) Hemos caracterizado la convexidad de la J-divergencia generalizada respecto del parámetro λ en el sentido:

$J_{\phi}^{\lambda}(p, q)$ es convexa (cóncava) en $(0, 1)$ respecto de λ para todas las parejas de funciones (p, q) si y solo si $\phi(t)$ es cóncava (convexa) en T_{ϕ} .

3) Hemos relacionado la variación de la J-divergencia generalizada respecto del parámetro para la función $\phi(t) = t \log t$ con la información de Kullbach-Leibler en el sentido

$$\left(\frac{\partial}{\partial \lambda} J_{\phi}^{\lambda}(p, q)\right)(\lambda) = k(p, \lambda p + (1-\lambda)q) - k(q, \lambda p + (1-\lambda)q)$$

$$\text{con } \|\sqrt{p}\| = \|\sqrt{q}\| = 1$$

4) Hemos observado que la simetrización $S J_{\phi}^{\lambda}(p, q) = J_{\phi}^{\lambda}(p, q) + J_{\phi}^{1-\lambda}(p, q)$ alcanza su máximo (mínimo) para $\lambda = \frac{1}{2}$ cuando $\phi(t)$ es convexa (cóncava) estricta.

- 5) Hemos obtenido relaciones de igualdad y desigualdad entre divergencias para distintas funciones $\phi(t)$, y en particular, atendiendo a propiedades de concavidad hemos obtenido

$$k_{\phi}(p, q) \geq 4 J_{\phi}^{\frac{1}{2}}(p, q) \geq 2 M_{\phi}^2(p, q) \geq 0$$

cuando $\phi(t)$ es una función log-convexa a valores positivos y $\phi(t) = \frac{\phi(t)}{t}$ es una función cóncava.

- 6) Hemos extendido las medidas de divergencia a medidas logarítmicas, obteniendo relaciones análogas a las que a las que se obtienen entre la entropía de Renyi y la entropía de Havrda-Charvat.
- 7) Hemos hallado una relación entre la medida de Csiszar y las medidas de entropía comunmente usadas, en el sentido:

$$I_f(p, q) = \frac{1}{\alpha-1} \left(1 - \sum_{i=1}^n p_i^{\alpha} \right) \quad \text{para } \alpha \neq 1$$

$$I_f(p, 1) = - \sum_{i=1}^n p_i \log p_i \quad \text{para } \alpha \neq 1$$

$$I_g(p, q) = - \sum_{i=1}^n p_i \log p_i - \sum_{i=1}^n (1-p_i) \log(1-p_i)$$

siendo:

$$f_{\alpha}(x) = \begin{cases} \frac{n^{1-\alpha}}{1-\alpha} (x^{1-\alpha} - n^{\alpha-1}) & \text{para } \alpha \neq 1 \\ \log x + \log n & \text{para } \alpha = 1 \end{cases}$$

$$g(x) = n x \log n x + (1-nx) \log (nx-1)$$

$$p = (p_1 \dots p_n) \quad \text{y} \quad q = (n^{-1}, \dots, n^{-1})$$

- 8) Tras plantear las métricas diferenciales asociadas a divergencias que vienen definidas en familias paramétricas de funciones de densidad de probabilidad, hemos estudiado la invariancia a nivel infinitesimal respecto de los cambios no singulares de parámetros y variables aleatorias. En este sentido, la invariancia la hemos obtenido para las siguientes funciones:
- a) Para la J-divergencia generalizada: $\phi(t) = at \log t + bt + c$
 - b) Para la K-divergencia: $\phi(t) = at \log t + bt$
 - c) Para la L-divergencia: Todas las funciones dan invariancia.
 - d) Para la M-divergencia: $\sqrt{\phi(t)} = a \sqrt{t} + b$
- 9) Hemos planteado las medidas de incertidumbre asociadas a J-divergencias y hemos analizado la convexidad y no negatividad de tales medidas, atendiendo a las funciones y constantes que la determinan. En este sentido, hemos observado que las medidas introducidas cumplen la condición de Dalton y hemos determinado el valor de la constante que proporciona incertidumbres decisivas.

- 10) Hemos obtenido como medidas de incertidumbre asociadas a J-divergencias las medidas siguientes: la entropía de Shannon, la entropía de Havrda-Charvat, la entropía de grado α , la entropía de Shannon modificada y la entropía de Latter. También hemos relacionado la entropía de Renyi y la γ -entropía con las medidas anteriores.

- 11) Hemos comparado, por métodos de Monte Carlo, características muestrales de los índices de incertidumbre comúnmente utilizados, con el índice que proviene de la J-divergencia para $\phi(t) = t \log t$. Hemos observado que el índice propuesto tiene propiedades similares a los ya estudiados.

BIBLIOGRAFIA

- ACZEL, J. y DAROCZY, Z. (1975), "On Measures of information and their characterizations". Academic Press. New York.
- AGRESTI, A. y AGRESTI, B.F. (1978). "Statistical analysis of qualitative variation", *Social methodology*, 204-237.
- AMARI, S. (1982). "Differential geometry of curved exponential families—curvatures and information loss". *The Annals of Statistics*, vol. 10, nº 2, 357-385.
- ATKINSON, C. y MITCHELL, A.F. (1981). "Rao's distance measure" *Sankhya*, 345-365.
- BERAU, R. (1977). "Minimum Hellinger distance estimates for parametric models". *The Annals of Statistics*, 5, 445-463.
- BOEKEE, D.E. y VAN DER LUBBE (1980). "The R-norm information measure". *Information and Control* 45, 136-155.
- BOURGUIGNON, F. (1979). "Descomposable income inequality measures". *Econometrika* 47, 4, 901-920.
- BRATLEY, P., FOX, B.L., SCHRODE, L.E. (1983). "A guide to simulation". Springer-Verlag, New York.
- BRATTACHARYA, A. (1946). "A measure of divergence between two multinomial populations". *Sankhya* 7, 401-406.
- BURBEA, J. y RAO, C.R. (1982). "Entropy differential metric, distance and divergence measures in probability spaces: a unified approach". *T. Multivariate Anal*, 12, 575-596.
- BURBEA, J. y RAO, C.R. (1982). "On the convexity of some divergence measures based on entropy functions". *IEEE Trans. Information Theory*, 28, 489-495.

- BURBEA, J. y RAO, C.R. (1984). "Differential metrics in probability spaces". *Probability and Mathematical Statistics* 3, 241-258.
- BURBEA, J. (1984). "The convexity with respect to Gaussian distributions of divergences of order α ". Technical Report 84-26 Univ. of Pittsburgh.
- BURBEA, J. (1984). "Informative geometry of probability spaces" Technical Report 84-52. Univ. of Pittsburgh.
- BURBEA, J. (1986). "Informative geometry of probability spaces" *Expositiones Mathematicae* 4, 347-378.
- CSISZAR, I. (1972). "A class of measures of informativity observation channels". *Periodica Math. Hungarica* 2, 191-213.
- CSISZAR, I. (1975). "I divergence geometry of probability distributions and minimization problems". *The Annals of Probability* 3, 146-158.
- CUADRAS, C.M. (1981). "Métodos de análisis multivariante". eunibar.
- CUADRAS, C.M., OLLER, J.M., ARCAS, A., RIOS, M. (1985). "Metodos geométricos de la estadística". *Qüestió* 9, 219-250.
- DAROCZY, Z. (1970). "Generalized information functions". *Information and Control*, 16, 83-88.
- DE GROOD, M.H. (1970). "Optimal Statistical Decisions". McGraw, New York.
- DIEUDONNE, J. (1977). "Elementos de Análisis", tomo II, Reverté.
- FERENTINOS, K., PAPAIVANNOU, T. (1981). "New parametric measures of information". *Information and Control*, 51, 193-208.

- GOOD, I.J. (1963). "Maximum entropy for hypothesis formulation especially for multidimensional contingency tables". *Annals Mathematical Statistics*, 34, 911-933.
- GREENBERG, J.H. (1956). "The measurement of linguistic diversity". *Language* 32, 109-115.
- HAVRDA, J. y CHARVAT, F. (1967). "Quantification method in classification processes: Concept of structural α -entropy". *Kybernetika*, 3, 30-35.
- HICKS, N.J. (1974). "Notas sobre geometría diferencial". *Hispano-europea*.
- JARDINE, N. y SIBSON, R. (1971). "Mathematical Taxonomy". Wiley, New York.
- JEFFREYS, H. (1948). "Theory of Probability", 2ª edición, Univ. Oxford, Londres.
- KRUSTAL, W. (1958). "Ordinal measures of association". *Journal of the Statistical Association*, 814-861.
- KUCZMA, M. (1968). "Functional equations in a single variable". Polish Scientific Publishers, T. 46.
- KULLBACK, S. y LEIBLER, R.A. (1951). "On information and sufficiency". *Annals Mathematical Statistics*, 22, 79-86.
- KULLBACK, S. (1959). "Information theory and statistics", John Wiley and Sons. New York.
- KUMAR, T. (1985). "On diversity measures based on entropy functions". *Commun. Statist. Theor. Meth.* 14, 203-215.
- LAU, K.S. (1985). "Characterization of Rao's quadratic entropies". *Sankhya* 47, 3, 295-309.

- LATTER, B.D.H. (1973). "Measures of genetic distance between individuals and populations", Genetic Structure of Populations, 27-39, Publicaciones Univ. Hawaii, Honolulu.
- LE CAM, L. (1970). "On the assumptions used to prove asymptotic normality of maximum likelihood estimates". Annals of Mathematical Statistics, 41, 802-828.
- LEWONTIN, R.C. (1972). "The apportionment of human diversity" Evolutionary Biol. 6, 381-398.
- LIEBERSON, S. (1969). "Measuring population diversity". Amer. Soc. Rev. 34, 850-862.
- LINDLEY, D.V. (1956). "On a measure of the information provided by and experiment". Annals of Mathematical Statistics, 27, 986-1005,
- LINDMAN, H. y CAELLI, T. (1975). "Constant curvature riemannian scaling". Journal of Math. Psychol. 17, 89-109.
- LOH, W.Y. (1985). "A note on the geometry of Kullback-Leibler information numbers". Commun. Statist. Theor. Meth. 14, 4, 895-904.
- MATHAI, A. y RATHIE, P.N. (1974). "Basic concepts in information theory and statistics". Wiley, New York.
- MATUSITA, K. (1955). "Decision rules, based on the distance, for problems of fit, two samples, and estimation". Annals Math. Statistics 26, 631-640.
- MATUSITA, K. (1957). "Decision rule based on the distance for the classification problem". Annals of the Inst. Statistica Math. 8, 67-77.
- MATUSITA, K. (1964). "Distance and decision rules". Annals of the Inst. Statistica Math. 14, 305-315.

- MATUSITA, K. (1966). "Classification based on distance in multivariate Gaussian cases". Proceed. of the fifth Berkeley Symposium on Mathematical Statistics and Probability.
- NEI, M. (1978). "The theory of genetic distance and evolution of human races". Japan J. Human Genetics 23, 341-369.
- OLLER, J.M. y CUADRAS, C.M. (1982). "Defined distances for some probability distributions". Proceed. II, Word Conf. Math. Serv. Man.
- PATIL, G.P. y TAILLE, C. (1979). "An overview of diversity" Ecological Diversity in Theory and Practice" 1, 3-28.
- PIELON, E.C. (1975). "Ecological diversity". Wiley, New York.
- PITMAN, E.J.G. (1979). "Some basic theory for statistical inference". John Wiley and Sons. New York.
- RAO, C.R. (1945). "Information and accuracu attainable in the estimation of parameters". Bull. Calcutta Math. Soc. 37, 81-91.
- RAO, C.R. (1948). "The utilization of multiple measurements in problems of biological classification". J. Royal Statistic Soc. 10, 159-193.
- RAO, C.R. (1954). "On the use and interpretation of distance functions in statistics". Bull. International Statist. Inst. 34, 90.
- RAO, C.R. (1981). "Gini-Simpson index of diversity: a characterization, generalization and applications". Technical Report 81-22. Univ. of Pittsburgh.
- RAO, C.R. (1982). "Analysis of diversity: a unified approach" Statistical Decision Theory and Related Topics 2, 233-250.

- RAO, C.R. (1982). "Diversity and dissimilarity coefficients: a unified approach". *Theoretical Population Biology* 21, 24-42.
- RAO, C.R. (1982). "Diversity: its measurement, descomposition, apportionment and analysis". *Sankhya* 44, 1-22.
- RENYI, A. (1961). "On measures of entropy and information". *Proceedings fourth Berkeley Symp.* 1, 547-561.
- RENYI, A. (1976). "Cálculo de Probabilidades". Reverté, S.A.
- ROCKAFELLAR, R.T. (1970). "Convex analysis". Princeton. New Jersey.
- RIOS, M. y CUADRAS, C.M. (1986). "Distancia entre modelos lineales normales". *Qüestió* 10, 83-92.
- RUDIN, W. (1979). "Análisis real y complejo". Alhambra.
- SALICRU, M. y ARCAS, A. (1985). "Sobre ciertas propiedades de la M-divergencia en análisis de datos". *Qüestió*, 9, 4, 251-255.
- SHAKED, M. (1977). "Statistical inference for a class of life distributions" 13, 1323-1339.
- SHANNON, C.E. (1948). "A mathematical theory of communications" *Bell System Tech. J.* 27, 379-423, 623-656.
- SHORROCKS, A.F. (1980). "The class of additively descomposable inequality measures". *Econometrica* 48, 613-625.
- SIMPSON, E.H. (1949). "Measurement of diversity". *Nature* 163, 688.

THEODORESCU, A. (1977). "Energie informationelle et notions apparentoes". Trabajos de estadística e I.O. 28, nº 2 y 3.

WALD, A. (1950). "Statistical decision functions". Wiley. New York.

WAYNE, A. y VARBERG, D.E. (1973). "Convex functions". Academic Press. New York.

WILLIAMSON, J.H. (1973). "Integración Lebesgue". Tecnos.