TRABAJOS DE ESTADISTICA Y
DE INVESTIGACION OPERATIVA
Vol. 32, Núm. 1, 1981, pp. 3 a 42

CANTIDAD DE INFORMACION DE FISHER (1)

Mª Pilar García-Carrasco Aponte
Universidad Complutense de Madrid
Facultad de Ciencias Matemáticas
Sección de Estadística e Investigación Operativa

RESUMEN

This paper is devoted to the Fisher's quantity of information study, it is focused in the same way than that ysed in the Shannon's information and planned under the conditions of regularity that Fourgeaud (5) considered. From this baseline, the corresponding criterion for comparing experiments is defined and studied.

1. Introducción

Este artículo está dedicado al estudio de la cantidad de información de Fisher como tal medida de información, ya que a pesar de su papel fundamental en las teorías de la estimación y tests clásicas, no había sido considerada con detalle en este sentido. Nuestro estudio, enfocado de modo análogo al usual en la información de Shannon y planteado bajo las condiciones de regularidad consideradas por Fourgeaud (5), considera las informaciones conjuntas y condicionadas examinando en particular las propiedades de no negatividad, aditividad, invarianza y suficiencia. Aunque alguno de los resultados allí obtenidos eran ya conocidos, la mayoría de las demostraciones son originales y hemos con-

(1) Este trabajo forma parte de la tesis presentada recientemente por el autor.

siderado conveniente su inclusión, para tras un estudio completo de la cantidad de información de Fisher y sus posibles generalizaciones, obtener el criterio de comparación de experimentos correspondientes, de modo análogo a los de suficiencia, maximización de la utilidad esperada y maximización de la información de Shannon estudiados anteriormente por el autor (6).

2. Definiciones y propiedades principales de la información de Fisher

2.1. Definición y condiciones de regularidad

Definición 1

Dado un espacio Θ, un experimento asociado a Θ consiste en un espacio medible $(\mathfrak{X}, \mathbf{A})$ y una familia de distribuciones de probabilidad sobre él $\{P_\theta, \theta \in \Theta\}$.

Consideremos el experimento $X = \{\mathfrak{X}, \mathbf{A}, P_\theta, \theta \in \Theta\}$. Supongamos que la familia $\{P_\theta\}\ \theta \in \Theta$ está dominada por una medida σ-finita μ y sean $p(x/\theta) = \dfrac{dP_\theta}{d\mu}$ las densidades. Es más, para mayor simplificación supongamos espacios euclideos con la σ-algebra de Borel y que esta medida μ es la de Lebesgue.

Sea Θ, el espacio de estados de la naturaleza, un abierto de $\mathbb{R}$.

Definimos la cantidad de información de Fisher dada por X sobre θ a la cantidad, si existe,

$$I(X, \theta) = E\left\{\left(\frac{\partial}{\partial\theta} \ln p(x/\theta)\right)^2\right\} = \int_{\mathfrak{X}} \left(\frac{\partial}{\partial\theta} \ln p(x/\theta)\right)^2 p(x/\theta)\, dx$$

Condiciones de regularidad

Diremos que X cumple las condiciones de regularidad (C.R.) si verifica:

$$E_\theta = \{x/p(x/\theta) > 0\}$$

es independiente de θ; es decir $E_\theta = E$.

$$E(1) = \int_{\mathfrak{X}} p(x/\theta)\, dx$$

es derivable dos veces bajo el signo integral.

Propiedad

Si X verifica C.R., entonces:

$$E\left(\frac{\partial}{\partial\theta} \ln p(x/\theta)\right) = o \text{ y}$$

$$I[X,\theta] = \mathrm{Var}\left(\frac{\partial}{\partial\theta} \ln p(x/\theta)\right) = -E\left(\frac{\partial^2}{\partial\theta^2} \ln p(x/\theta)\right) \qquad [1]$$

Para su demostración ver Fourgeaud (5), pág. 181.

Ejemplos en los que X no verifica C.R.

(1) Sea X : $\mathbf{U}(0, \theta)$, entonces:

$$I(X,\theta) = \frac{1}{\theta^2}, \quad \mathrm{Var}\left(\frac{\partial}{\partial\theta} \ln p(x/\theta)\right) = 0 \quad \text{y}$$

$$-E\left(\frac{\partial^2}{\partial\theta^2} \ln \frac{1}{\theta}\right) = -\frac{1}{\theta^2}$$

(2) Sea X tal que:

$$p(x/\theta) = \begin{cases} \dfrac{1}{2\,\theta} & \text{si } x \in (0,\, \theta) \\[2ex] \dfrac{1}{2\,(1-\theta)} & \text{si } x \in [\theta,\, 1) \end{cases}$$ para $\Theta = (0,\ 1)$; entonces

$$I[X,\, \theta] = \frac{1}{2\,\theta^2} + \frac{1}{2\,(1-\theta)^2},\quad E\left(\frac{\partial}{\partial\theta} \ln p(x/\theta)\right) = -\frac{1}{2\,\theta} + \frac{1}{2(1-\theta)} \neq 0$$

y $$-E\left(\frac{\partial^2}{\partial\theta^2} \ln p(x/\theta)\right) = -\frac{1}{2\,\theta^2} - \frac{1}{2\,(1-\theta)^2}$$

Como vemos en estos ejemplos, aunque X no verifique las C.R., puede definirse la $I(X,\, \theta)$ como se ha hecho anteriormente. Sin embargo, como iremos viendo a lo largo de este estudio, es solo cuando X verifica las C.R. cuando en efecto la definición dada se comporta como una buena medida de información.

2.2. Aditividad y no negatividad

En los resultados de este apartado, a pesar de su gran interés, omitimos las demostraciones por ser en general sencillas de comprobar.

Teorema 1

$I(X,\, \theta) \geqslant 0$ con igualdad si y solo si $p(x/\theta)$ es independiente de θ, excepto a lo sumo en un conjunto de medida nula. (Ver Fourgeaud (5)).

Definición 1

Como caso particular de la definición 1 de 2.1., tenemos:

$$I[(X_1, X_2); \theta] = E_{x_1, x_2/\theta} \left\{ \left(\frac{\partial}{\partial\theta} \ln p(x_1, x_2/\theta) \right)^2 \right\}$$

$$I[X_2(x_1); \theta] = E_{x_2/\theta, x_1} \left\{ \left(\frac{\partial}{\partial\theta} \ln p(x_2/\theta, x_1) \right)^2 \right\}$$

definimos ahora,

$$I[X_2/X_1; \theta] = E_{x_1/\theta}\, I[X_2(x_1); \theta] =$$

$$= \int_{\mathfrak{X}_1} p(x_1/\theta) \int_{\mathfrak{X}_2} \left(\frac{\partial}{\partial\theta} \ln p(x_2/\theta, x_1) \right)^2 p(x_2/\theta, x_1)\, dx_2\, dx_1 =$$

$$= \int_{\mathfrak{X}_1 \times \mathfrak{X}_2} p(x_1, x_2/\theta) \left[\frac{\partial}{\partial\theta} \ln p(x_2/\theta, x_1) \right]^2 dx_1\, dx_2$$

Teorema 2

En las condiciones de regularidad suficientes para que las diversas cantidades de información puedan expresarse de la forma [1], se verifica:

$$I[(X_1, X_2); \theta] = I(X_1; \theta) + I(X_2/X_2; \theta)$$

Generalización del teorema 2

En las condiciones de regularidad que permiten escribir las diversas cantidades de información de la forma [1], $\forall\, n > 1$, se verifica:

$$I[(X_1 \ldots X_n); \theta] = I[X_n; \theta] + I[X_{n-1}/X_n; \theta] + \cdots +$$

$$+ I[X_{n-j}/(X_{n-j+1} \ldots X_n); \theta] + I[(X_1 - X_{n-j-1})/(X_{n-j} - X_n); \theta]$$

$$\forall j \,,,\, 0 \leqslant j \leqslant n-2$$

El resultado se obtiene por inducción, teniendo en cuenta que para $j = 0$ no es mas que un caso particular del teorema 2.

Ejemplos en los que no se verifican las C.R.

(1) Sea X_1: $u(0, \theta)$ y $X_2(x_1)$: $u(x_1, 2\theta)$; entonces,

$$I[(X_1, X_2); \theta] = \frac{3 + \ln 16}{\theta^2} > \frac{3}{\theta^2} = I[X_1; \theta] + I[X_2/X_1; \theta]$$

$$\forall\, \theta \in \Theta = (0, +\infty)$$

(2) Sea X_1: $u(0, \theta)$ y $X_2(x_1)$ con $p(x_2/\theta, x_1) = e^{\theta - x_2}$ $x_2 \in (\theta, +\infty)$; entonces,

$$I[(X_1, X_2); \theta] = 1 + \frac{1}{\theta^2} - \frac{2}{\theta} < 1 + \frac{1}{\theta^2} = I[X_1; \theta] + I[X_2/X_1; \theta]$$

$$\forall\, \theta \in \Theta = (0, +\infty)$$

Luego en el caso en que E_θ depende de θ puede darse la desigualdad entre:

$$I[(X_1, X_2); \theta] \quad \text{y} \quad I[X_1; \theta] + I[X_2/X_1; \theta]$$

en uno cualquiera de los dos sentidos. Esta igualdad, es una de las propiedades que debe pedirse a una medida de información y que cuando no se verifican las C.R. deja con esta definición de ser cierta.

Teorema 3

Siempre se verifica $I[X_2/X_1; \theta] \geqslant 0$ con igualdad si y solo si $p(x_2/\theta, x_1)$ es independiente de θ casi seguro.

Corolario

En las condiciones de regularidad exigidas en el teorema 2, $I[(X_1, X_2); \theta] \geqslant I[X_1; \theta]$ con igualdad si y solo si $p(x_2/\theta, x_1)$ es independiente de θ casi seguro.

El resultado es inmediato de los teoremas 2 y 3.

Ejemplos en los que no se verifican las condiciones de regularidad C.R.

Sean:

$$X_1\colon\ p(x_1/\theta) = \frac{1}{\theta} \quad x_1 \in (0, \theta)$$

$$X_2\colon\ p(x_2/\theta) = e^{\theta - x_2} \quad x_2 \in (\theta, +\infty)$$

v.a. independientes; entonces:

$$I[X_1; \theta] = \frac{1}{\theta^2}, \quad I[X_2; \theta] = 1 \quad \text{y} \quad I[(X_1, X_2); \theta] = 1 + \frac{1}{\theta^2} - \frac{2}{\theta}$$

v.a. independientes. Luego:

$$I[(X_1, X_2); \theta] \geqslant I[X_2; \theta] \quad \Leftrightarrow \quad 1 + \frac{1}{\theta^2} - \frac{2}{\theta} \geqslant 1 \quad \Leftrightarrow \quad \frac{1}{2} \geqslant \theta$$

$$I[(X_1, X_2); \theta] \geqslant I[X_1; \theta] \quad \Leftrightarrow \quad 1 + \frac{1}{\theta^2} - \frac{2}{\theta} \geqslant \frac{1}{\theta^2} \quad \Leftrightarrow \quad \theta \geqslant 2$$

Luego $\forall \theta \in \Theta = (0, +\infty)$ $I[(X_1, X_2); \theta]$ no es comparable a $I[X_1; \theta]$ ni a $I[X_2; \theta]$; para $\theta \in \Theta = (2, +\infty)$ por ejemplo tendríamos:

$$I[(X_1, X_2); \theta] > I[X_1; \theta] \quad \text{y} \quad I[(X_1, X_2); \theta] < I(X_2, \theta)$$

Vemos así que en el caso de no verificarse las C.R. no se puede asegurar nada sobre la relación de $I[(X_1, X_2); \theta]$ y $I(X_1; \theta)$ ya que pueden no ser comparables $\forall\, \theta \in \Theta$ y en caso de serlo alcanzar la desigualdad en uno cualquiera de los sentidos.

Teorema 4

Si X_1 y X_2 son independientes, $I[X_2/X_1; \theta] = I[X_2; \theta]$.

Observación:

En particular, si $X_1 \equiv X_2$ el teorema nos dirá que una repetición independiente del mismo experimento es igual de informativa en media que la original.

Corolario

Si X_1 y X_2 son independientes y se verifican las condiciones de regularidad exigidas en el teorema 2,

$$I[(X_1, X_2); \theta] = I(X_1; \theta) + I(X_2; \theta)$$

El resultado es inmediato de los teoremas 2 y 4.

Ejemplos en los que no se verifican las C.R.

(1) Sean $X_1 : u(0, \theta)$ y $X_2 : u(0, 2\theta)$ independientes, entonces:

$$I[(X_1, X_2); \theta] = \frac{4}{\theta^2} > \frac{1}{\theta^2} + \frac{1}{\theta^2} = I[X_1; \theta] + I[X_2; \theta]$$

(2) Sean:

$$X_1 \,,, \; p(x_1/\theta) = \frac{1}{\theta}, \quad x_1 \in (0, \theta)$$

$$X_2 \,,, \; p(x_2/\theta) = e^{\theta - x_2}, \quad x_2 \in (\theta, +\infty)$$

independientes como vimos anteriormente (ejemplo 2 del teorema 2).

$$I[(X_1, X_2); \theta] = 1 + \frac{1}{\theta^2} - \frac{2}{\theta} < 1 + \frac{1}{\theta^2} = I[X_1, \theta] + I[X_2, \theta]$$

Luego tampoco en el caso de experimentos independientes, cuando no se verifican las C.R. podemos afirmar nada sobre la relación entre $I[(X_1, X_2); \theta]$ y $I[X_1; \theta] + I[X_2; \theta]$.

2.3. Combinaciones lineales convexas de informaciones

Definición 1

Sean $X_i = \{\mathfrak{X}_i;\ \mathbf{A}_i;\ P_{\theta i};\ \theta \in \Theta\}$ $i = 1{,}2$ dos experimentos sobre Θ de densidades $p_i(x_i/\theta)$. Consideremos el nuevo experimento $X = (\lambda X_1 + (1 - \lambda) X_2) = \{\mathfrak{X}_1 \cup \mathfrak{X}_2; \mathbf{A}_1 \cup \mathbf{A}_2; P_\theta; \theta \in \Theta\}$ donde P_θ viene definida por la densidad

$$p(x/\theta) = \begin{cases} \lambda\, p_1(x/\theta) & \text{si } x \in \mathfrak{X}_1 \\ (1 - \lambda)\, p_2(x/\theta) & \text{si } x \in \mathfrak{X}_2 \end{cases}$$

Propiedad

Siempre se verifica:

$$I[\lambda X_1 + (1 - \lambda) X_2; \theta] = \lambda I[X_1; \theta] + (1 - \lambda) I[X_2; \theta] \quad \forall \lambda \,,, 0 \leqslant \lambda \leqslant 1$$

(fácil de comprobar).

2.4. Información en muestras de tamaño n

Vamos a estudiar a continuación la información de Fisher en muestras de tamaño n. Sea el experimento $X = \{\mathfrak{X}; \mathbf{A}; P_\theta; \theta \in \Theta\}$ de densidad $p(x/\theta)$. Consideramos el nuevo experimento $X^{(n)}$ cuya probabilidad vendrá descrita por la densidad $p(x_1 \ldots x_n/\theta) = \Pi_{i=1}^n p(x_i/\theta)$.

Como caso particular de la definición de información de Fisher:

$$I[X^{(n)}; \theta] = E\left(\frac{\partial}{\partial\theta} \ln p(x_1 \ldots x_n/\theta)\right)^2 =$$

$$= \int_{\mathfrak{X}^n} \prod_{i=1}^n p(x_i/\theta) \left[\frac{\partial}{\partial\theta} \ln \prod_{i=1}^n p(x_i/\theta)\right]^2 dx_1 \ldots dx_n$$

Propiedad 1

$$I[X^{(n)}; \theta] = n\, I(X; \theta) + 2\binom{n}{2} \left[E\left(\frac{\partial}{\partial\theta} \ln p(x/\theta)\right)\right]^2$$

(fácil de comprobar).

Propiedad 2

Si X verifica las C.R., entonces $I[X^{(n)}; \theta] = n\, I[X; \theta]$.

En efecto, en las C.R. se vio en 2.1 que:

$$E\left(\frac{\partial}{\partial\theta} \ln p(x/\theta)\right) = 0$$

y por la propiedad 1, $I(X^{(n)}; \theta) = n\, I(X; \theta)$.

Observación: Este resultado puede obtenerse también como caso particular del corolario del teorema 4 de 2.2.

Propiedad 3

$I[X^{(n)}; \theta]$ es una función no decreciente de n (se obtiene fácilmente de la propiedad 1).

Propiedad 4

En las C.R. del teorema 2 de 2.2 se verifica $\forall\lambda,, 0 \leqslant \lambda \leqslant 1$, $I[\lambda\ X^{(k)} + (1-\lambda)\ X^{(m)};\ \theta] = I[X^{(n)};\ \theta]$ siendo $n = \lambda\ K + (1-\lambda)\ m$. En efecto:

$$I[\lambda\, X^{(k)} + (1-\lambda)\, X^{(m)}; \theta] = \lambda\, I[X^{(k)}; \theta] + (1-\lambda)\, I[X^{(m)}; \theta] =$$

$$= \lambda\, k\, I[X; \theta] + (1-\lambda)\, m\, I[X; \theta] = n\, I[X; \theta] = I[X^{(n)}; \theta]$$

Ejemplo en el que no se verifican las C.R.

Sea $X{:}u(0, \theta)$; entonces $I[X; \theta] = \frac{1}{\theta^2}$ y $I[X^{(n)}; \theta] = \frac{n^2}{\theta^2}$ $\forall\, n \geqslant 1$;

luego $\qquad I[X^{(n)}; \theta] \neq n\, I[X; \theta] \quad \forall\, n \geqslant 2$

además, tomando por ejemplo $\lambda = \frac{1}{2}$ $k = 2$, $m = 4$, $n = \frac{1}{2}\, 2 + \frac{1}{2}\, 4 = 3$

$$I\left[\frac{1}{2}\, X^{(2)} + \frac{1}{2}\, X^{(4)}; \theta\right] = \frac{1}{2}\, \frac{2^2}{\theta^2} + \frac{1}{2}\, \frac{4^2}{\theta^2} = \frac{10}{\theta^2} > \frac{q}{\theta^2} = I[X^{(3)}; \theta]$$

luego en este ejemplo $I[\lambda\, X^{(k)} + (1-\lambda)\, X^{(m)}; \theta] \neq I[X^{(n)}; \theta]$.

Propiedad 5

Siempre se verifica $I[\lambda\ X^{(k)} + (1-\lambda)\ X^{(m)}; \theta] \geqslant I[X^{(n)}; \theta]$ siendo $n = \lambda\ k + (1-\lambda)\ m$, (se obtiene fácilmente de la propiedad 1).

Observación: Hemos obtenido por tanto que la función $I[X^{(n)}; \theta]$ es una función convexa de n. Es interesante observar que por lo contrario, con la información de Shannon, ésta función es cóncava (ver Lindley (9)). Así, en el caso en que no se verifican C.R. podemos obtener a partir de aquí un ejemplo en el cual las informaciones de Shannon y Fisher dan desigualdades contrarias.

Ejemplo en el que no se verifican las C.R.

Sea $X{:}u(0, \theta)$; $\Theta = (0, +\infty)$, $Y = \frac{1}{2} X^{(2)} + \frac{1}{2} X^{(4)}$, $Z = X^{(3)}$.

Puede comprobarse fácilmente que:

1°) $I[Y; \theta] > I[Z; \theta] \ \ \forall\, \theta \in \Theta$

2°) $J[Z; p(\theta)] \geqslant J[Y; p(\theta)]$ $\forall$ distribución a priori $p(\theta)$ sobre Θ

3°) $J[Z; p_0(\theta)] > J[Y; p_0(\theta)]$ tomando por ejemplo $p_0(\theta){:}\gamma(p)$ con $p = 7$.

2.5. Invariancia

Teorema 1

Sea $X = \{\mathbb{R}; \beta; P_\theta, \theta \in \Theta\}$ de densidad $p(x/\theta)$; sea T una transformación de $\mathbb{R} \to \mathbb{R}$ monótona estricta y derivable. Entonces $I(X; \theta) = I(T; \theta)$. (Trivial).

Corolario

En particular, una translación $T = a\,X + b$ conserva la información de Fisher: $I[a\,X + b; \theta] = I[X; \theta]$.

Teorema 2

Sea $X = \{\mathbb{R}; \beta; P_\theta, \theta \in \Theta\}$ y supongamos que $T = T(X)$ es una transformación de $\mathbb{R} \to \mathbb{R}$ con derivada continua y distinta de cero salvo en un conjunto de media nula; supongamos también que la familia $\{P_\theta, \theta \in \Theta\}$ es una familia completa de distribuciones. Entonces $p(x/\theta) = p(t/\theta) \cdot H(x)$ implica:

$$I(T; \theta) = I(X; \theta)$$

En efecto:

$$I[X; \theta] = \int_{\mathfrak{X}} p(x/\theta) \left[\frac{\partial}{\partial\theta} \ln p(x/\theta)\right]^2 dx =$$

$$= \int_T p(t/\theta)\, H(x(t)) \left[\frac{\partial}{\partial\theta} \ln p(t/\theta)\right]^2 x'(t)\, dt = (*)$$

por otro lado:

$$\int_T p(t/\theta)\, dt = 1$$

$$\int_{\mathfrak{X}} p(x/\theta)\, dx = \int_T p(t/\theta)\, H(x(t))\, x'(t)\, dt = 1$$

luego:

$$\int_T p(t/\theta)\, [1 - H(x(t))\, x'(t)]\, dt = 0 = \int_{\mathfrak{X}} \frac{p(x/\theta)}{H(x)} [1 - H(x)\, x']\, t'(x)\, dx$$

y por ser $\{P_\theta, \theta \in \Theta\}$ completa

$$\frac{1 - H(x)\, x'}{H(x)}\, t'(x) = 0$$

c.s. y al ser $t'(x) \neq 0$ c.s. se tiene $H(x)\, x' = 1$ c.s., luego:

$$(*) = \int_T p(t/\theta) \left[\frac{\partial}{\partial \theta} \ln p(t/\theta) \right]^2 dt = I[T; \theta]$$

Teorema 3

Sea X un experimento sobre θ e $Y = b_1 X + b_2$ un experimento sobre $\omega = a_1 \theta + a_2$; entonces $I[Y; \omega] = \dfrac{1}{a_1^2} I[X; \theta]$, (trivial).

Corolario

Las translaciones en θ y X dejan invariante la información de Fisher.

Teorema 4

Sean los experimentos $X_1, \ldots, X_n$ sobre $(\mathbb{R}, \beta)$ y consideremos la transformación de $\mathbb{R}^n \to \mathbb{R}^n$

$$\begin{aligned} Y_1 &= Y_1(X_1 \ldots X_n) \\ &\vdots \\ Y_m &= Y_m(X_1 \ldots X_n) \end{aligned}$$

tal que la aplicación:

$$\begin{pmatrix} X_{n-m+1} \\ \vdots \\ X_n \end{pmatrix} \to \begin{pmatrix} Y_1 \\ \vdots \\ Y_m \end{pmatrix}$$

es biyectiva y con derivadas parciales primeras continuas $\forall\, X_1, \ldots, X_{n-m}$. Entonces $I[(X_1 \ldots X_n); \theta] = I[(X_1 \ldots X_{n-m}\ Y_1 \ldots Y_m); \theta]$. (Trivial).

2.6. Suficiencia

Teorema 1

Sea $X = \{\mathfrak{X}; \mathbf{A}; P_\theta, \theta \in \Theta\}$ un experimento y supongamos que se verifican las condiciones de regularidad; concretamente, E_θ independiente de θ y derivabilidad dos veces bajo el signo integral y sumatorio. Sea $\{E_i\}_{i=1,2,\ldots}$ una partición de $\mathfrak{X}$ por elementos del σ-álgebra $\mathbf{A}$. Entonces,

$$I[X;\theta] \geqslant \sum_i P_\theta(E_i) \left[\frac{\partial}{\partial\theta} \ln P_\theta(E_i)\right]^2$$

con igualdad si y solo si $\forall\, i = 1, 2, \ldots$ y $\forall\, x \in E_i$ $\dfrac{p(x/\theta)}{P_\theta(E_i)}$ es independiente de θ c.s. En efecto:

$$I[X;\theta] = \int_{\mathfrak{X}} p(x/\theta) \left[\frac{\partial}{\partial\theta} \ln p(x/\theta)\right]^2 dx =$$

$$= -\int_{\mathfrak{X}} p(x/\theta) \left[\frac{\partial^2}{\partial\theta^2} \ln p(x/\theta)\right] dx =$$

$$= -\sum_i \int_{E_i} \left[\frac{\partial^2}{\partial\theta^2} \ln p(x/\theta)\right] p(x/\theta)\, dx$$

por otro lado:

$$\sum_i P_\theta(E_i) \left[\frac{\partial}{\partial\theta} \ln P_\theta(E_i)\right]^2 = -\sum_i P_\theta(E_i) \left[\frac{\partial^2}{\partial\theta^2} \ln P_\theta(E_i)\right]$$

en efecto, derivando $\sum_i P_\theta(E_i) = 1$ respecto a θ.

$$\sum_i \frac{\partial}{\partial\theta} P_\theta(E_i) = 0 = \sum_i \left[\frac{\partial}{\partial\theta} \ln P_\theta(E_i)\right] P_\theta(E_i)$$

volviendo a derivar respecto a θ.

$$\sum_i \left[\frac{\partial^2}{\partial\theta^2} \ln P_\theta(E_i)\right] P_\theta(E_i) + \sum_i \left[\frac{\partial}{\partial\theta} \ln P_\theta(E_i)\right]^2 P_\theta(E_i) = 0$$

tomamos $g_i(x/\theta) = \dfrac{p(x/\theta)}{P_\theta(E_i)}$ $x \in E_i$ es una densidad en E_i teniendo en cuenta el t° 1 y las condiciones C.R.

$$-\int_{E_i} g_i(x/\theta) \left[\frac{\partial^2}{\partial\theta^2} \ln g_i(x/\theta)\right] dx \geqslant 0$$

con = si y solo si $g_i(x/\theta)$ es independiente de θ c.s., luego:

$$-\int_{E_i} \frac{p(x/\theta)}{P_\theta(E_i)} \left[\frac{\partial^2}{\partial\theta^2} \ln p(x/\theta) - \frac{\partial^2}{\partial\theta^2} \ln P_\theta(E_i)\right] dx \geqslant 0$$

con = si y solo si $\dfrac{p(x/\theta)}{P_\theta(E_i)}$ es independiente de θ c.s., por lo tanto:

$$-\frac{1}{P_\theta(E_i)} \int_{E_i} p(x/\theta) \frac{\partial^2}{\partial\theta^2} \ln p(x/\theta)\, dx \geqslant -\frac{\partial^2}{\partial\theta^2} \ln P_\theta(E_i)$$

con = si y solo si $\dfrac{p(x/\theta)}{P_\theta(E_i)}$ es independiente de θ c.s. y como queríamos demostrar:

$$-\sum_i \int_{E_i} p(x/\theta) \frac{\partial^2}{\partial\theta^2} \ln p(x/\theta)\, dx \geqslant -\sum_i P_\theta(E_i) \frac{\partial^2}{\partial\theta^2} \ln P_\theta(E_i)$$

con igualdad si y solo si $\dfrac{p(x/\theta)}{P_\theta(E_i)}$ es independiente de θ c.s.

Teorema 2

Para todo estadístico T de la muestra $X^{(n)} = (X_1 \dots X_n)$ de X con $n > 1$ se verifica $I[T; \theta] \leqslant I[X^{(n)}; \theta]$ con igualdad si y solo si T es suficiente. Suponemos las C.R. suficientes para que las distintas cantidades de información se puedan expresar de la forma (1) y que la aplicación $x_1 \to T(x_1 \dots x_n)$ es biyectiva y con derivada primera continua $\forall\, x_2 \dots x_n$. En efecto:

Por el Teorema 4 de 2.5 $I[(X_1 \dots X_n); \theta] = I[(T\, X_2 \dots X_n); \theta] = I[T; \theta] + I[(X_2 \dots X_n)/T; \theta]$ por el teorema 2 de 2.2. Además, por el teorema 3 de 2.2.

$$I[(X_2 \dots X_n)/T; \theta] \geqslant 0$$

con = si y solo si $p(x_2 - x_n/t, \theta)$ es independiente de θ c.s. Luego:

$$I[X^{(n)}; \theta] \geqslant I[T; \theta]$$

con = si y solo si T es suficiente c.q.d.

Teorema 3

Si T es suficiente siempre se verifica $I(T; \theta) = I(X^{(n)}; \theta)$. En efecto, por el teorema 4 de 2.5.:

$$I[X^{(n)}; \theta] = I[(T\, X_2 \dots X_n); \theta] =$$

$$= \int_{T \cdot \mathfrak{X}^{(n-1)}} p(t/\theta)\, p(x_2 \dots x_n/t, \theta) \left[\frac{\partial}{\partial \theta} \ln p(t/\theta) + \right.$$

$$\left. + \frac{\partial}{\partial \theta} \ln p(x_2 \dots x_n/t, \theta) \right]^2 dt\, dx_2 \dots dx_n$$

y como por ser t suficiente $p(x_2 \dots x_n/t, \theta)$ es independiente de θ c.s.

$$= \int_{T} p(t/\theta) \left[\frac{\partial}{\partial \theta} \ln p(t/\theta) \right]^2 dt = I[T; \theta]$$

Teorema 4

Supongamos que la muestra $X^{(n)} = (X_1 \ldots X_n)$ de X admite un estadístico suficiente $R = R(X_1 \ldots X_n)$ y sea $T = T(X_1 \ldots X_n)$ otro estadístico tal que $\forall\, x_1 \ldots x_{n-2}$ la aplicación $\begin{pmatrix} x_{n-1} \\ x_n \end{pmatrix} \to \begin{pmatrix} r = R(x_1 \ldots x_n) \\ t = T(x_1 \ldots x_n) \end{pmatrix}$ es biyectiva y con derivadas parciales primeras continuas. Supongamos además que se verifican las C.R. suficientes para que las distintas cantidades de información que figuran a continuación puedan expresarse de la forma (1). Entonces $I[T/R; \theta] = 0$. En efecto, por el teorema 4 de 2.5

$$I[X^{(n)}; \theta] = I[(R, T, X_1 \ldots X_{n-2}); \theta] =$$

y por el teorema 2 de 2.2.:

$$= I[R; \theta] + I[T/R; \theta] + I[X^{(n-2)}/(R, T); \theta]$$

Además, los teoremas 1 y 3 de 2.2., nos aseguran que estas cuatro cantidades son no negativas. Por otro lado, por ser R suficiente, $I(R; \theta) = I[X^{(n)}; \theta]$; luego $I[T/R; \theta] = 0$, c.q.d.

Este teorema nos precisa la idea de que si se ha observado un estadístico suficiente R, la observación de cualquier otro estadístico T no aporta ninguna nueva información sobre θ.

Teorema 5

Sea $X = \{\mathfrak{X}; \mathsf{A}; P_\theta, \theta \in \Theta\}$ un experimento, donde $p(x/\theta) = dP_\theta/dx$ y sea $T = \{y; \beta; Q_\theta, \theta \in \Theta\}$ un estadístico de X de densidad $q(t/\theta) = dQ_\theta/dt$. Supongamos que tanto X como T verifican las C.R. Entonces, $I[X; \theta] \geqslant I[T; \theta]$ con igualdad si y solo si T es suficiente.

En efecto, comenzamos demostrando que si X verifica las C.R., entonces:

$$\forall A \in \mathbf{A} \quad \frac{\partial}{\partial\theta} \int_A p(x/\theta)\, dx = \int_A \frac{\partial}{\partial\theta} p(x/\theta)\, dx$$

Por hipótesis:

$$\exists \int_{\mathfrak{X}} \frac{\partial}{\partial\theta} p(x/\theta)\, dx, \quad \text{luego} \; \exists \int_A \frac{\partial}{\partial\theta} p(x/\theta)\, dx$$

$$\int_A \frac{\partial}{\partial\theta} p(x/\theta)\, dx = \int_{\mathfrak{X}} \frac{\partial}{\partial\theta} p(x/\theta) \;\; I_A \; dx = (*)$$

y como:

$$\left[\frac{\partial}{\partial\theta} p(x/\theta)\right] I_A = \frac{\partial}{\partial\theta} [p(x/\theta)\, I_A] = \begin{cases} \frac{\partial}{\partial\theta} p(x/\theta) & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases}$$

$$(*) = \int_{\mathfrak{X}} \frac{\partial}{\partial\theta} [p(x/\theta)\, I_A]\, dx$$

Como demuestra Crámer (3), si $\dfrac{\partial g(x, \theta)}{\partial\theta}$ c.s. en x y $\left|\dfrac{\partial g(x, \theta)}{\partial\theta}\right| < G(x)$ donde $G(x)$ es integrable $\mathfrak{X}$, $\forall\, \theta$ en un intervalo abierto (a, b), entonces:

$$\frac{d}{d\theta} \int_{\mathfrak{X}} g(x, \theta)\, dF(x) = \int_{\mathfrak{X}} \frac{\partial g(x, \theta)}{\partial\theta}\, dF(x) \quad \forall\, t \in (a, b)$$

En nuestro caso, tomando $g(x, \theta) = p(x/\theta)\, I_A$ para cada $A \in \mathbf{A}$: $\exists \dfrac{\partial g(x, \theta)}{\partial\theta}$ c.s. pues $\exists \displaystyle\int_{\mathfrak{X}} \frac{\partial g(x, \theta)}{\partial\theta}\, dx$; además $\left|\dfrac{\partial g(x, \theta)}{\partial\theta}\right| < \left|\dfrac{\partial g(x, \theta)}{\partial\theta}\right| + \dfrac{1}{e^{\|x\|^2}} = G(x)$ integrable $\mathfrak{X}$, $\forall\, \theta \in \Theta$ abierto de $\mathbb{R}$. Luego:

$$\int_{\mathfrak{X}} \left[\frac{\partial}{\partial\theta} p(x/\theta)\, I_A\right] dx = \frac{d}{d\theta}\int_{\mathfrak{X}} p(x/\theta)\, I_A\, dx \quad \forall\, \theta \in \Theta$$

es decir:

$$\int_A \frac{\partial}{\partial\theta} p(x/\theta)\, dx = \frac{\partial}{\partial\theta}\int_A p(x/\theta)\, dx$$

De aquí, como demuestra Rao (12) en 5a.4, resulta que $I[X;\theta] - I[T;\theta] \geqslant 0$ con igualdad si y solo si $\dfrac{p'(x/\theta)}{p(x/\theta)} = \dfrac{q'(t/\theta)}{q(t/\theta)}$ c.s., es decir $I[X;\theta] \geqslant I[T;\theta]$ con igualdad si y solo si T es suficiente, c.q.d.

Observación: El teorema 2 de 2.6. puede obtenerse ahora también como caso particular de éste.

Corolario

En las condiciones de regularidad C.R., si T es un estadístico suficiente de X y g una transformación monótona estricta y derivable de T, entonces $R = g(T)$ es también un estadístico suficiente de X.

En efecto, por el teorema 5 de 2.6. al ser T suficiente de x, $I[T;\theta] = I[X;\theta]$, por el teorema 1 de 2.5., $I[R;\theta] = I[T;\theta]$, luego por el teorema de 2.6., al ser $I[R;\theta] = I[X;\theta]$, resulta que $R = g(T(X))$ es también un estadístico suficiente de x.

2.7. Relación entre la cantidad de información de Fisher y las teorías de estimación y tests estadísticos

La cantidad de información de Fisher aparece subyacente en muchos resultados clásicos de la teoría de la estimación y contraste de hipótesis estadísticas.

Además de caracterizar en ciertas condiciones de regularidad los estadísticos suficientes (teorema 5 d e 2.6.), surge de modo natural en la acotación de Crámer-Rao, estando de este modo íntimamente ligada al concepto de eficiencia. Podríamos por tanto enunciar muchas de las propiedades y teoremas clásicos relacionados con los estadísticos suficientes y eficientes en términos de la cantidad de información de Fisher.

Además, debido al conocido resultado, que asegura en ciertas condiciones de regularidad la existencia de un estimador de máxima verosimilitud que se distribuye asintóticamente normal con media la media poblacional y varianza el inverso de la información de Fisher resulta que esta cantidad juega también un papel importante en la teoría y métodos, tanto de estimación puntual como de intervalos de confianza, para muestras grandes.

En cuanto a la teoría de Tests de hipótesis son conocidas en la literatura estadística, varios conceptos de la eficiencia asintótica relativa de dos tests, así como de sus relaciones con la información de Fisher (ver Rao (11) y (12)). También la relación entre eficiencia de un test e información de Fisher ha sido estudiada con frecuencia en la teoría de los tests de orden (Capon (1) y Duran (4)). De Groot (7) demuestra también algunas relaciones fundamentales entre la eficiencia de un test estadístico y la información de Fisher, así como las relaciones entre la eficiencia relativa asintótica de Pitman y la eficiencia relativa asintótica de Fisher, definida esta última como límite de cociente de informaciones.

Esta íntima relación de la cantidad de información de Fisher con la estadística clásica, es una de las razones que nos inclinan a elegir esta cantidad como punto de partida de un criterio para comparar experimentos, ya que además, como hemos visto anteriormente, en ciertas condiciones de regularidad, se comporta como una buena medida de información.

2.8. Generalizaciones de la información de Fisher

2.8.1. Información de Fisher en el caso discreto

En todo el desarrollo anterior hemos empleado experimentos de tipo continuo, y aunque los resultados allí obtenidos son igualmente válidos para el caso discreto, creemos conveniente hacer las siguientes observaciones:

(I) En este caso nuestros experimentos serán del tipo $X = \{\mathfrak{X}; \mathbf{A}; P_\theta, \theta \in \Theta\}$ donde Θ es un abierto de $\mathbb{R}$, $\mathfrak{X} = \{x_n\}_{n \in N}$ y P_θ viene descrita para cada $\theta \in \Theta$ por las probabilidades $p(x_n/\theta)$, tales que:

$$p(x_n/\theta) \geqslant 0 \quad \forall\, n \in N \quad \text{y} \quad \sum_{n \in N} p(x_n/\theta) = 1$$

Definimos la cantidad de información dada por X sobre θ a la cantidad, si existe:

$$I[X; \theta] = E \left\{ \frac{\partial}{\partial \theta} \ln p(x/\theta) \right\}^2 = \sum_{n=1}^{\infty} p(x_n/\theta) \left[\frac{\partial}{\partial \theta} \ln p(x_n/\theta) \right]^2$$

(II) Las C.R. aquí consideradas serán:

a) $E_\theta = \{x \in \mathfrak{X} / p(x/\theta) > 0\}$ independientes de θ.

b) $\sum_{n=1}^{\infty} p(x/\theta) = 1$, derivable dos veces bajo el sumatorio.

Observamos que si el número de elementos de $\mathfrak{X}$ es finito las C.R. se reducen a a).

Bajo estas C.R. se sigue verificando que $I[X; \theta] = \operatorname{Var}\left(\frac{\partial}{\partial \theta} \ln p(x/\theta) \right) = -E\left(\frac{\partial^2}{\partial \theta^2} \ln p(x/\theta) \right)$.

Las propiedades principales de no negatividad y aditividad, siguen siendo ciertas en condiciones análogas a las consideradas en el caso continuo.

(III) El estudio de la información en muestras de tamaño n para el caso continuo, puede hacerse de modo totalmente análogo en el discreto, si bien en este caso puede complicarse la notación.

(IV) En cuanto al estudio de la invarianza y suficiencia, si bien la mayoría de los resultados obtenidos en el caso continuo son trasladables al discreto, conviene observar como se simplifican las cosas en este caso.

La diferencia fundamental estriba en que al considerar un estadístico T de un experimento X, podemos considerar la cantidad de información de Fisher de la variable aleatoria bidimensional (X, T) sobre θ, ya que en este caso sigue siendo discreta mientras que en el caso continuo se concentraba a lo largo de la curva $y = T(X)$ de modo que al no ser su distribución continua ni discreta no estaba definida la información de Fisher.

2.8.2. Generalización de la información de Fisher en el caso de no verificarse las condiciones de regularidad

Una vez estudiada la medida de información de Fisher y viendo que solo bajo ciertas condiciones de regularidad se comporta como tal, sería conveniente dar una definición más amplia de esta cantidad de información de modo que conservando las propiedades principales en el caso general coincidiera con la información de Fisher primeramente definida en caso de verificarse las C.R. consideradas.

Un camino de interés en este sentido es el seguido por Vajda (13), a partir del concepto de sensibilidad introducido por Rao (12), pero cuyas condiciones de regularidad de partida no coinciden exactamente con las aquí consideradas.

Otra generalización interesante de la cantidad de información de Fisher es la debida a Losfeld (10), y que se aplica en particular al caso de familias de probabilidad en las que $E_\theta = \{x/p(x/\theta) > 0\}$ depende de θ; surge de las definiciones de información y divergencia de Kullback generalizadas.

También un camino lógico en principio y que parece ajustarse bien a las condiciones de regularidad aquí consideradas, es tomar por definición como medida de información el denominador que aparece en la desigualdad de Chapman (2), generalización clásica de la cota de Cramer Rao.

Por último, otra idea tambien factible y cuyo estudio dejamos abierto, sería aproximar la cantidad de información de Fisher de un experimento X que no se comporte regularmente por las cantidades de información de una sucesión de experimentos regulares X_n, tales que X_n convergieran a X de un modo conveniente.

2.8.3. Información de Fisher cuando la distribución a priori es conocida

Vamos a considerar ahora la información de Fisher en un problema bayesiano. Es decir, supongamos la existencia de una distribución $\Pi(\theta)$ sobre un σ-álgebra de conjuntos F sobre Θ. Trataremos concretamente el caso en que la distribución $H(\theta)$ viene descrita por la densidad $p(\theta)$.

En este caso sería natural definir la cantidad media de información de Fisher dada por un experimento X sobre θ, cuando la distribución a priori es $p(\theta)$, como:

$$I[X; p(\theta)] = E_\theta \, I[X; \theta] = \int_\Theta p(\theta) \, I[X; \theta] \, d\theta$$

Todas las propiedades de no negatividad, aditividad, información en muestras de tamaño n, invarianza y suficiencia, se obtienen inmediatamente para esta cantidad media de información a partir de las correspondientes estudiadas para la información de Fisher, por ser la esperanza un operador lineal.

3. Criterio de comparación de experimentos como maximización de la cantidad de información de Fisher

3.1. Planteamiento

Una vez estudiadas las propiedades de la cantidad de información de Fisher y visto su bien comportamiento en ciertas condiciones de regularidad, así como su fuerte relación con la estadística clásica, nos parece una buena medida para dar a partir de ella un criterio de comparación de experimentos, ya que esta cantidad de información tiene además la ventaja de venir definida exclusivamente a partir del experimento, comparando así éstos, de una manera absoluta y no en función de los términos concretos del problema de partida.

Nos vamos a limitar en este estudio al caso de experimentos que verifican las C.R. y cuyas probabilidades vienen descritas por densidades. Para el caso discreto, teniendo en cuenta el apartado 2.8.1. podría hacerse de modo análogo. También, en el caso de no verificarse las C.R., pueden definirse criterios de comparación análogos a partir de cualquiera de las definiciones de información de Fisher generalizada 2.8.2. Por último, si existen probabilidades a priori sería lógico definir un criterio, más débil que el aquí considerado, a partir de la información media definida en 2.8.3., obteniendo así un preorden completo.

Como hemos señalado sucesivas veces lo anterior solo tiene sentido para Θ un abierto de $\mathbb{R}$. En el caso de ser $\Theta = \{\theta_1 \dots \theta_N\}$ finito, podría definirse a partir de la desigualdad de Chapman (2) y por ser ésta válida para cualquier $\Theta \subset \mathbb{R}$, una medida de información análoga a la de Fishe del siguiente modo:

$$I[X; \theta_i] = \min_{\theta_j \neq \theta_i} \int_{\mathfrak{X}} \left[\frac{p(x/\theta_j) - p(x/\theta_i)}{(\theta_j - \theta_i)\, p(x/\theta_i)} \right]^2 p(x/\theta_i)\, dx \quad \forall\, i = 1 \dots N$$

donde el mínimo se extiende a todos los $\theta_j \in \Theta$, $\theta_j \neq \theta_i$ tales que $E_{\theta_j} \subset E_{\theta_i}$. A partir del estudio de sus propiedades podría definirse un criterio de comparación análogo al aquí considerado, cuyas relaciones con otros criterios, y en particular con el de Blackwell podrían ser de interés.

Del mismo modo para cualquier $\Theta \subset \mathbb{R}$ podríamos definir a partir de la cota de Chapman una cantidad de información y el criterio de comparación correspondiente.

El caso $\theta = (\theta_1 \ldots \theta_s)$ podría ser tratado a partir de las matrices de información de Fisher (ver por ejemplo Rao (12), Fourgeaud (5)) considerando C.R. análogas.

Así, bastaría para cada uno de estos casos hacer un estudio detallado de sus propiedades y a partir de ellas ver el comportamiento del criterio de comparación correspondiente.

A continuación nos limitaremos, como hemos señalado antes, al caso de experimentos que verifican las C.R. y cuyas probabilidades vienen descritas por densidades, por ser el caso para el cual ha sido estudiado con más detalle la medida de información definida.

Definición

Dados los experimentos $X = \{\mathfrak{X}; \mathbf{A}; P_\theta, \theta \in \Theta\}$ e $Y = \{y; \beta; Q_\theta, \theta \in \Theta\}$ sobre el mismo Θ, diremos que X es más informativo que $Y (X \geqslant^F Y)$ si $I[X; \theta] \geqslant I[Y; \theta]$ $\theta \in \Theta$.

Diremos que $X \overset{F}{\sim} Y$ si y solo si $X \overset{F}{\geqslant} Y$ e $Y \overset{F}{\geqslant} X$.

3.2. Propiedades

I – Para todo experimento X, $X \overset{F}{\geqslant} N$ donde $N = \{y; \beta; Q\}$ es el experimento nulo es decir, aquel en que Q no depende de θ (consecuencia del teorema 1 de 2.2.).

II – La relación $\overset{F}{\geqslant}$ es un preorden parcial.

En general dados dos experimentos X e Y sobre Θ puede que ni $X \overset{F}{\geqslant} Y$ ni $Y \overset{F}{\geqslant} X$, aunque a veces podríamos conformarnos con comparar las cantidades de información $\forall\, \theta \in \Theta' \subset \Theta$, de modo que Θ' fuera elegido convenientemente.

III – $(X_1 X_2) \overset{F}{\gtrless} X_1$ con $\overset{F}{\sim}$ si y solo si $p(x_2/x_1, \theta)$ es independiente de θ c.s. (consecuencia inmediata del corolario del teorema 3 de 2.2.).

IV – $X^{(n+1)} \geqslant X^{(n)}$ $\forall\, n \geqslant 1$ (consecuencia inmediata de la propiedad 3 de 2.4.).

V – Si X_1, X_2 y X_3 son tres experimentos sobre el mismo Θ y X_3 es independiente de ambos X_1 y X_2, entonces $X_1 \overset{F}{\gtrless} X_2 \Rightarrow (X_1 X_3) \overset{F}{\gtrless} (X_2 X_3)$.

En efecto, por el corolario del teorema 4 de 2.2. y las hipótesis:

$$I[(X_1 X_3); \theta] = I[X_1; \theta] + I[X_3; \theta] \geqslant I[X_2; \theta] + I[X_3; \theta] =$$
$$= I[(X_2 X_3); \theta]$$

VI – Si X_1, X_2, X_3 y X_4 son cuatro experimentos sobre el mismo Θ tales que $X_1 \overset{F}{\gtrless} X_2$, $X_3 \overset{F}{\gtrless} X_4$, X_1 es independiente de X_3 y X_2 de X_4, entonces $(X_1 X_3) \overset{F}{\gtrless} (X_2 X_4)$.

En efecto, por el corolario del teorema 4 de 2.2. y las hipótesis:

$$I[(X_1 X_3); \theta] = I[X_1; \theta] + I[X_3; \theta] \geqslant I[X_2; \theta] + I[X_4; \theta] =$$
$$= I[(X_2 X_4); \theta]$$

VII – Sea $X = \{\mathfrak{X}; \mathbf{A}; P_\theta, \theta \in \Theta\}$ un experimento donde P_θ, $\theta \in \Theta$, viene descrito por la densidad $p(x/\theta)$. Sea $\{E_i\}_{i \in N}$ una partición de $\mathfrak{X}$ por elementos del σ-álgebra $\mathbf{A}$. Consideremos el nuevo experimento $Y = \{y; \beta; Q_\theta, \theta \in \Theta\}$ donde $y = \{E_i\}_{i \in N}$, β es la σ-álgebra engendrada por los E_i y Q_θ es tal que $Q_\theta(E_i) = P_\theta(E_i)$. Entonces $X \overset{F}{\gtrless} Y$ (inmediato del teorema 1 de 2.6. y del apartado 2.8.1.).

VIII – Para todo estadístico $T = T(X^{(n)})$ de la muestra de tamaño n, $X^{(n)} \overset{F}{\gtrless} T$ con $\overset{F}{\sim}$ si y solo si T es un estadístico suficiente (inmediato del teorema 5 de 2.6.).

IX – Sean S y T dos estadísticas suficientes de $X^{(n)}$ e $Y^{(m)}$ respectivamente, entonces $X^{(n)} \overset{F}{\gtrless} Y^{(m)}$ si y solo si $S \overset{F}{\gtrless} T$ (consecuencia de VIII y ser $\overset{F}{\gtrless}$ transitiva).

X – Sea $X = \{\mathbb{R}; \beta; P_\theta, \theta \in \Theta\}$, si T es una transformación de $R \to R$ monótona estricta y derivable, entonces $X \overset{F}{\sim} T$ (consecuencia inmediata del teorema 1 de 2.5.).

XI – Sean $X_1 \ldots X_n$ experimentos sobre $(\mathbb{R}, \beta)$, y consideremos una transformación de $\mathbb{R}^n \to \mathbb{R}^m$

$$\begin{matrix} Y_1 = Y_1(X_1 \ldots X_n) \\ \vdots \\ Y_m = Y_m(X_1 \ldots X_n) \end{matrix}$$

tal que la aplicación

$$\begin{pmatrix} X_{n-m+1} \\ \vdots \\ X_n \end{pmatrix} \to \begin{pmatrix} Y_1 \\ \vdots \\ Y_m \end{pmatrix}$$

es biyectiva y con derivadas parciales primeras continuas $\forall x_1 \ldots x_{n-m}$. Entonces $(X_1 \ldots X_n) \overset{F}{\sim} (X_1 \ldots X_{n-m} \;\; Y_1 \ldots Y_m)$ (inmediato del teorema 4 de 2.5.).

3.3. Relaciones entre los criterios de suficiencia y de Fisher

Teorema 1

Sean X e Y dos experimentos sobre el mismo Θ. Si X es preferido a Y con el criterio de suficiencia de Lhemann [ver Gª Carrasco (6)] entonces $X \overset{F}{\gtrless} Y$.

En efecto, $X \overset{\pounds h}{\geqslant} Y \Rightarrow$ una v.a. U independiente de X y una función h medible tal que $H = h(X, U)$ tiene la misma distribución que $Y \; \forall \, \theta \in \Theta$.

Además se demostró en el teorema 1 de 1.1. del artículo antes citado que $P_r(H/\theta, X = x)$ es independiente de θ.

Por otro lado, por el corolario del teorema 3 de 2.2., $I[(X\,H);\theta] \geqslant \geqslant I[X;\theta]$ con igualdad si y solo si $p(h/\theta, x)$ es independiente de θ c.s. Luego:

$$I[Y;\theta] = I[H;\theta] \leqslant I[(X\,H);\theta] = I[X;\theta] \quad \forall\, \theta \in \Theta$$

Teorema 2

Sean $X = \{\mathfrak{X};\, \mathbf{A};\, P_\theta,\, \theta \in \Theta\}$ e $Y = \{y;\, \beta;\, Q_\theta,\, \theta \in \Theta\}$ dos experimentos sobre Θ, tales que $\{P_\theta\}_{\theta \in \Theta}$ es una familia completa de distribuciones. Entonces, si X es suficiente para Y [ver G^a Carrasco (6)], X es preferido a Y con el criterio de Fisher.

En efecto, en estas hipótesis se demostró en el teorema 2 de 2.3. del artículo antes citado que $P_\theta^*\{(x\,y)/p(y/x,\theta)$ depende de $\theta\} = 0$ $\forall\, \theta \in \Theta$, luego, por el corolario del teorema 3 de 2.2.,

$$I[Y;\theta] \leqslant I[(X\,Y);\theta] = I[X;\theta]$$

BIBLIOGRAFIA

1.– Capon, J.: "*Asymptotic efficiency of certain locally most powerful rank tests*", Ann. Math. Statist. 32, 88-100, (1961).

2.– Chapman, D.G. y Robbins, H.: "*Minimum Variance Estimation without Regularity Assymptions*", Ann. Math. Statist. 22, 581-586, (1951).

3.– Crámer, H.: "*Métodos matemáticos de Estadística*", Aguilar (1963).

4.– Duran, B.S. y Mielke, P.W.: "*Robustness of sum of squared rank tests*", Jour. Amer. Statist. Assoc. 63, 338-344 (1968).

5.– Fourgeaud, G. y Fuchs, A.: "*Statistique*", Dunod (1972).

6.– García-Carrasco, M.P.: "*Criterios para la comparación de experimentos*". Trabajos de Estadística y de Investigación Operativa. Vol. XXX (1979).

7.– De Groot, M.H. y Raghavachari, M.: "*Relations between Pitman efficiency and Fisher information*", Sankhya. A., 319-324, (1970).

8.– Lehmann, E.L.: "*Testing Statistical Hypotheses*", John Wiley, (1959).

9. Lindley, D.V.: "*On a Measure of the Information Provided by an Experiment*" Annals of Mathematical-Statistics 27, 986-1005 (1956).

10. Losfeld, J.: "*Une generalisation de la quantité d'information de Fisher et de l'inégalité de Cramer-Rao*" Rev. Roumaine Math. Pures Appl. 16, 503-515, (1971).

11. Rao, C.R.: "*Asymptotic efficiency and limiting information*", Proc. Fourth Berkeley Symp. Math. Statis. Probability 1, 531-545, (1961).

12. Rao, C.R.: "*Linear Statistical Inference and its Applications*", John Wiley and Sons, Inc., New York (1965).

13. Vajda, I.: "*Divergence and generalized Fisher's information*", Sixth Prague Conference on Information Theory, Statistical Decision Function, Random Processes, 873-886, Academia Prague (1973).