Stochastica XIII-1 (1992), 75-87

LAS f^*–DIVERGENCIAS COMO CRITERIO BAYESIANO DE COMPARACION DE EXPERIMENTOS

J.A. PARDO (*), M.L. MENÉNDEZ (**) & L. PARDO (***)

ABSTRACT

In this paper a bayesian criterion for comparing different experiments based on the maximization of the f^*-Divergence, is proposed and studied. After a general setting of the criterion, we prove that this criterion verifies the main properties that a criterion for comparing experiment must satisfy.

1. Introduction.

Sea $(\chi, \beta_\chi, P_\theta)_{\theta\in\Theta}$, el espacio estadístico asociado a algún experimento aleatorio X y denotemos por $f(x/\theta)$ la función de densidad de P_θ respecto de una medida conveniente μ_1. Supondremos que el valor del parámetro θ es el resultado de una variable aleatoria y $(\Theta, \beta_\Theta, Q)_{Q\in Q^*}$ es el espacio de probabilidad asociado a tal variable aleatoria. Denotaremos por $p(\theta)$ la densidad de Q respecto de una medida λ sobre (Θ, β_Θ) y por Θ^*, el conjunto de las distribuciones a priori sobre Θ. Supondremos que Θ y χ son subconjuntos de un espacio euclideo k–dimensional y β_θ, β_χ las σ–álgebras de Borel. Finalmente, denotaremos por $g(x,\theta)$ la densidad conjunta de la variable aleatoria (X,Θ) sobre el espacio medible $(\chi\times\Theta, \sigma(\beta_\chi\times\beta_\Theta))$, por $p(\theta/x)$ la distribución a posteriori de θ observado x y $f(x)$ la distribución predictiva o marginal de X.

Una vez establecido el esquema bayesiano en el que nos vamos a mover a lo largo del presente trabajo, pasamos a dar la definición de f^*–divergencia que nos permitirá cuantificar la información que un experimento proporciona o proporcionará acerca de un parámetro desconocido. La definición que aquí se presenta es una adaptación dada por Morales, Pardo y Quesada (1985) al contexto bayesiano de la dada por Csiszar (1976), (en un contexto no bayesiano) para dos distribuciones cualesquiera.

Definición 1. La cantidad de información acerca del valor de θ proporcionada por el valor observado x y cuantificada en términos de la f^*–Divergencia viene dada por

$$D_{f^*}\left(p(\theta/x), p(\theta)\right) = \int_\theta p(\theta) f^*(p(\theta/x)/p(\theta))\; d\lambda(\theta) \tag{1}$$

donde f^* es una función convexa definida en $(0, +\infty)$, verificando:

i) $f^*(0) = \lim_{u \to 0} f^*(u)$. ii) $Of^*(0/0) = 0$.

ii) $Of^*(a/0) = \lim_{\epsilon \to 0^+} \epsilon f' x(a/\epsilon) = a \lim_{u \to \infty} f^*(u)/u$.

Conviene hacer notar que mientras que i), ii) y iii) se imponen para evitar expresiones sin sentido, la convenxidad será esencial a lo largo del trabajo debido a que la mayoría de las desigualdades que aparecen en las demostraciones posteriores se basan en la utilización de la desigualdad de Jensen. La convexidad estricta de f^* será una condición necesaria para alcanzar determinadas igualdades.

Obsérvese que si se toma $f^*(x)$ como -log x, -xlog x, $(1 - x^{1/2})^2$, $x(1 - 1/x)^2$ y $x|1 - 1/x|^\alpha$ se obtienen las adaptaciones al contexto bayesiano de las medidas de Kullback–Leiber, Kullback–Leiber modificada, distancia de Matusita, medida de Kagan y medida de Vajda de grado α respectivamente. Estas medidas han tenido y siguen teniendo un gran interés, tanto teórico como práctico, dentro de la Teoría de la Información; en este sentido los resultados que se obtienen en este trabajo ponen de manifiesto, de forma conjunta, que estas medidas tienen un buen comportamiento a la hora de comparar experimentos en un contexto bayesiano.

Antes de observar la realización muestral x, la información que se espera obtener por la observación del experimento X, viene dada por:

Definición 2. La cantidad de información, acerca de θ, que se espera obtener por la realización del experimento X viene dada, supuesto que exista, por

$$\begin{aligned} D_{f^*}[X, p(\theta)] &= E_X\left[D_{f^*}[p(\theta/x), p(\theta)]\right] = \\ &= \int_{\mathcal{X}} \left[\int_{\Theta} p(\theta) f^*(p(\theta/x)/p(\theta))\, d\lambda(\theta)\right] f(x)\, d\mu_1(x) = \\ &= \int_{\Theta} \left[\int_{\mathcal{X}} f^*(f(x/\theta)/f(x)) f(x)\, d\mu_1(x)\right] p(\theta)\, d\lambda(\theta) = \end{aligned}$$

que en lo sucesivo se denominará f^*–Divergencia esperada acerca de θ.

2. Las f^*–divergencias y la comparación de experimento.

En caso de que el estadístico tenga a su alcance distintas variables observables, es decir, diversos experimentos, y tenga la necesidad de elegir entre ellos se planteará lógicamente "Cual de ellos elegir". El primero que se planteó el problema de la comparación de experimentos fue Blackwell (1951) quien introdujo la noción de experimento suficiente, así como la de experimento más informativo estudiando la relación entre ambos conceptos. Lehmann (1959) estableció una comparación de experimentos aplicable a los problemas de test de hipótesis. Finalmente cabe destacar en esta línea de dar métodos para comparar experimentos los trabajos de De Groot (1960, 1962, 1966), Raiffa and Schlaifer (1961), Le Cam (1964), Lindley (1956, 1968), García–Carrasco (1978, 1982).

En este apartado se analizará como la maximización de la f^*–divergencia, cumple las propiedades más importantes, a nuestro juicio, que debe verificar un criterio de comparación de experimentos.

Definición 3. Sean $X_1 = (\chi_1, \beta_{\chi_1}, f(x_1/\theta))_{\theta\in\Theta'}$ $X_2 = (\chi_2, \beta_{\chi_2}, f(x_2/\theta))_{\theta\in\Theta}$ dos experimentos aleatorios. Se dice que el experimento X_1 es preferido al experimento X_2, según el criterio de la f^*–divergencia, y respecto a la distribución a priori $p(\theta)$ si y sólo si

$$D_{f^*}[X_1, p(\theta)] \geq D_{f^*}[X_2, p(\theta)] \tag{1}$$

(lo denotaremos por $X_1 \geq X_2$). Diremos que X_1 es indiferente a X_2 respecto a la distribución a priori $p(\theta)$ si y sólo si $X_1 \geq X_2$ y $X_2 \geq X_1$.

Es sencillo establecer que la relación $\geq$ es un preorden completo. En caso de que en la definición anterior no estuviese fija $p(\theta)$ se tendría un criterio no bayesiano ya que no tendría en cuenta el particular conocimiento que se tiene acerca de θ. Es sencillo comprobar que en este caso la relación $\geq$ es un preorden parcial.

Los resultados siguientes muestran las propiedades del criterio de comparación entre experimentos dado en la definición 3. Comencemos dando la siguiente definición.

Definición 4. Se dice que el experimento $N = (\mathcal{N}, \beta_{\mathcal{N}}, P)$ es un experimento nulo si P no depende de θ.

Teorema 1. Dado cualquier experimento $X = (\chi, \beta_\chi, f(x/\theta)_{\theta\in\Theta}$, se verifica que $\chi \geq \mathcal{N}$, $\forall\ p(\theta) \in \Theta^*$. En caso de que f^* sea estrictamente convexa, se de la igualdad si y sólo si $f(x/\theta)$ es independiente de θ c.s. μ_1.

Demostración.

Al ser,

$$\begin{aligned} D_{f^*}\left[N, p(\theta)\right] &= \int_{\mathcal{N}} f(n) \left[\int_\Theta f^*(p(\theta/n)/p(\theta))\ d\lambda(\theta)\right]\ d\mu_1(n) = \\ &= \int_{\mathcal{N}} f(n) \left[\int_\Theta f^*(f(n/\theta)/f(n))\ d\lambda(\theta)\right]\ d\mu_1(n) = \\ &= \int_{\mathcal{N}} \left[\int_\Theta f^*(1)p(\theta)f(n)\ d\lambda(\theta)\right]\ d\mu_1(n) = f^*(1) \end{aligned}$$

se probarà que $D_{f^*}\left[X, p(\theta)\right] \geq f^*(1)$, $\forall p(\theta) \in \Theta^*$. En efecto, consideremos la variable aleatoria $\eta_\theta(x) = p(\theta/x)/p(\theta)$, θ fijo, con función de densidad $f(x)$ y la función $f^* : \mathbb{R}^+ \to R$ convexa, se tiene entonces

$$E_X(\eta_\theta(x)) = \int_\chi (p(\theta/x)/p(\theta))f(x)\ d\mu_1(x) = 1$$

y

$$f^*\left[E_X(\eta_\theta(x))\right] = f^*(1).$$

Además,

$$f^*(\eta_\theta(x)) = f^*(p(\theta/x)/p(\theta))$$

y

$$E_X\left[f^*(\eta_\theta(x))\right] = \int_X f^*(p(\theta/x)/p(\theta))\ f(x)\ d\mu_1(x)$$

Por la desigualdad de Jensen se tiene que

$$E_X\left[f^*(\eta_\theta(x))\right] \geq f^*\left[E_X(\eta_\theta(x))\right] \qquad \forall\theta \in \Theta,$$

luego

$$\int_X f^*(p(\theta/x)/p(\theta))f(x)\ d\mu_1(x) \geq f^*(1) \qquad \forall\theta \in \Theta, \tag{2}$$

es decir

$$\int_X p(\theta)f^*(p(\theta/x)/p(\theta))f(x)\ d\mu_1(x) \geq f^*(1)p(\theta) \qquad \forall\theta \in \Theta$$

e integrando a lo largo de Θ

$$\int_X \left[\int_\Theta p(\theta)f^*(p(\theta/x)/p(\theta))\ d\lambda(\theta)\right] f(x)\ d\mu_1(x) \geq f^*(1)$$

con lo cual

$$D_{f^*}\left[X, p(\theta)\right] \geq f^*(1) \quad \forall p(\theta) \in \Theta^*$$

Si f^* es estrictamente convexa se dará la igualdad si y sólo si $P(\eta_\theta(x) = E_X(\eta_\theta(x))) = 1$ c.s. μ_1; es decir, si y sólo si, $p(\theta/x) = p(\theta)$ c.s. μ_1 o lo que es lo mismo si y sólo si $f(x/\theta)$ es independiente de θ c.s. μ_1.

Así pues, como era de esperar, cualquier experimento cuyos resultados dependen del parámetro desconocido, proporciona más información acerca de éste que otro cuyos resultados no dependen del parámetro de interés.

Definición 5. Sean $X_1 = (\chi_1, \beta_{\chi_1}, f(x_1/\theta))_{\theta\in\Theta}$ y $X_2 = (\chi_2, \beta_{\chi_2}, f(x_2/\theta))_{\theta\in\Theta}$ dos experimentos asociados al mismo problema estadístico. Se denomina experimento compuesto o experimento suma al experimento $X_1 \times X_2 = (\chi_1 \times \chi_2, \beta_{\chi_1\times\chi_2},\ f(x_1, x_2/\theta))_{\theta\in\Theta}$ en el que β_{χ_1} es la σ–álgebra inducida a partir de $\beta_{\chi_1\times\chi_2}$ sobre χ_1 por la proyección de índice i $(i = 1, 2)$ y $f(x_1/\theta)$ es la densidad sobre β_{χ_1} inducida a partir de $f(x_1, x_2/\theta)$.

Una vez vista esta definición vamos a establecer un resultado que indica que, el experimento compuesto proporciona más información acerca de θ que cualquiera de los experimentos marginales.

Teorema 2. $X_1 \times X_2 \geq X_1$, $\forall p(\theta) \in \Theta^*$. Además, en el caso de que f^* sea estrictamente convexa, se da la igualdad si y sólo si $f(x_2/x_1, \theta)$ es independiente de θ μ_2 c.s.

Demostración.

Consideremos la variable aleatoria $\eta_{\theta,x_1}(x_2) = p(\theta/x_1, x_2)/p(\theta)$ con $\theta \in \Theta$ y $x_1 \in \chi_1$ fijos y función de densidad $f(x_2/x_1)$. Sea $f^* : \mathbb{R}^+ \to R$ una función convexa, se tiene entonces:

$$E\left(\eta_{\theta,x_1}(x_2)\right) = \int_{\chi_2} (p(\theta/x_1, x_2)/p(\theta)) f(x_2/x_1)\ d\mu_2(x_2) = p(\theta/x_1)/p(\theta)$$

y

$$f^*\left[E\left(\eta_{\theta,x_1}(x_2)\right)\right] = f^*(p(\theta/x_1)/p(\theta)).$$

Además,,

$$f^*\left[\eta_{\theta,x_1}(x_2)\right] = f^*(p(\theta/x_1, x_2)/p(\theta))$$

y

$$E\left[f^*\left(\eta_{\theta,x_1}(x_2)\right)\right] = \int_{\chi_2} f^*(p(\theta/x_1, x_2)/p(\theta)) f(x_2/x_1)\ d\mu_2(x_2)$$

Por la desigualdad de Jensen se tiene:

$$E\left[f^*\left(\eta_{\theta,x_1}(x_2)\right)\right] \geq f^*\left[E\left(\eta_{\theta,x_1}(x_2)\right)\right] \qquad \forall \theta \in \Theta \quad \text{y} \quad \forall x_1 \in \chi_1,$$

luego

$$\int_{\chi_2} f^*(p(\theta/x_1, x_2)/p(\theta))f(x_2/x_1)\, d\mu_2(x_2) \geq f^*(p(\theta/x_1)/p(\theta))$$

multiplicando por $f(x_1)$, integrando en χ_1 y multiplicando por $p(\theta)$ e integrando en Θ se tiene:

$$\int_{\chi_1}\int_{\chi_2}\left(\int_{\Theta} p(\theta)f^*(p(\theta/x_1, x_2)/p(\theta))d\lambda(\theta)\right] f(x_1, x_2)d\mu_2(x_2)\, d\mu_1(x_1) \geq$$
$$\geq \int_{\chi_1}\left(\int_{\Theta} p(\theta)f^*(p(\theta/x_1)/p(\theta))\, d\lambda(\theta)\right) f(x_1)\, d\mu_1(x_1)$$

es decir,

$$D_{f^*}(X_1 \times X_2; p(\theta)) \geq D_{f^*}(X_1; p(\theta)) \quad \forall p(\theta) \in \Theta^*$$

Si f^* es estrictamente convexa se dará la igualdad si y sólo si $\forall\theta\ \forall x_1$, se tiene $E_{X_2}(\eta_{\theta,x_1}(x_2)) = p(\theta/x_2)$ c.s. μ_2. Es decir si y sólo si $p(\theta/x_1, x_2) = p(\theta/x_2)$ c.s. μ_2. Ahora bien como

$$p(\theta/x_1, x_2) = \frac{f(x_2/x_1, \theta)}{f(x_1, x_2)} f(x_1, \theta) \quad \text{ y } \quad p(\theta/x_1) = \frac{f(x_1, \theta)}{f(x_1)}$$

la igualdad se dará si y solo si $f(x_2/x_1, \theta) = f(x_2/x_1)$ c.s. μ_2.

Así pues dada una muestra aleatoria simple de tamaño n (experimento compuesto asociado a n experimentos independientes e idénticamente distribuidos), $X^{(n)} = X \times \overset{n)}{\ldots} \times X$, se establece, de forma análoga a la demostración desarrollada en el teorema anterior, que

$$X^{(n)} \geq X^{(n-1)} \quad \forall p(\theta) \in \theta^* \quad \forall n \in \mathbb{N}.$$

Es decir, cuanto más experimentos se observen mayor información se tendrá acerca de θ.

El siguiente resultado relaciona la información proporcionada por una muestra aleatoria simple de tamaño n, $X^{(n)} = X \times \overset{n)}{\ldots} \times X$, con la proporcionada por un estadístico, $T = (\chi^n, \beta_{\chi^n}, P_\theta \to (R^k, B(R^k), P_\theta^T)$, de la muestra.

Teorema 3.

$$X^{(n)} \geq T \qquad \forall p(\theta) \in \Theta^*.$$

Además en el caso en que f^* sea estrictamente convexa, se da la igualdad si y solo si T es suficiente para θ.

Demostración.

Designemos por $f(t/\theta)$ la densidad de T y por $K_T = (K_T(t))_t$ la partición determinada por T en $X^{(n)}$, siendo $K_T(t) = \{x/T(x) = t\}$ $t \in \mathrm{Im}(T) = Y$. Consideremos $\theta \in \Theta$ y $t \in \mathrm{Im}(T)$, fijos, y se define la variable aleatoria

$$\eta_{t,\theta}(x_1, \ldots, x_n) = \frac{p(\theta/x_1, \ldots, x_n)}{p(\theta)}$$

con función de densidad $f(x_1, \ldots, x_n/t)$. Sea $f^* : R^+ \to R$ una función convexa, se tiene que

$$E\left(\eta_{t,\theta}(x_1, \ldots, x_n)\right) = \int_{X^{(n)}} \frac{p(\theta/x_1, \ldots, x_n)}{p(\theta)} f(x_1, \ldots, x_n/t)\ d\mu(x_1, \ldots, x_n) = \\ = \frac{p^T(\theta/t)}{p(\theta)}$$

$$f^*\left(E\left(\eta_{t,\theta}(x_1, \ldots, x_n)\right)\right) = f^*\left(\frac{p^T(\theta/t)}{p(\theta)}\right)$$

Además,

$$f^*\left(\eta_{t,\theta}(x_1, \ldots, x_n)\right) = f^*\left(\frac{p(\theta/x_1, \ldots, x_n)}{p(\theta)}\right)$$

$$E\left[f^*\left(\eta_{t,\theta}(x_1, \ldots, x_n)\right)\right] = \int_{X^{(n)}} f^*\left(\frac{p(\theta/x_1, \ldots, x_n)}{p(\theta)}\right) f(x_1, \ldots, x_n/t)\ d\mu(x_1, \ldots, x_n)$$

A través de la desigualdad de Jensen se llega a que

$$E\left[f^*\left(\eta_{t,\theta}(x_1, \ldots, x_n)\right)\right] \geq f^*\left[E\left(\eta_{t,\theta}(x_1, \ldots, x_n)\right)\right] \quad \forall\theta, \forall t,$$

luego

$$f(t)p(\theta)\int_{\mathcal{X}^{(n)}} f^*\left(\frac{p(\theta/x_1,\ldots,x_n)}{p(\theta)}\right) f(x_1,\ldots,x_n/t)\; d\mu(x_1,\ldots,x_n) \geq$$
$$\geq f(t)p(\theta)f^*\left(\frac{p^T(\theta/t)}{p(\theta)}\right) \tag{4}$$

e integrando en Y se tiene

$$\int_{\mathcal{X}^{(n)}} p(\theta)f^*\left(\frac{p(\theta/x_1,\ldots,x_n)}{p(\theta)}\right) f(x_1,\ldots,x_n)\; d\mu(x_1,\ldots,x_n) \geq$$
$$\int_Y p(\theta)f^*\left(\frac{p^T(\theta/t)}{p(\theta)}\right) f(t)\; d\mu(t) \forall\theta \in \Theta \tag{5}$$

e integrando en Θ se tiene

$$\int_{\mathcal{X}^{(n)}} \left[\int_\Theta p(\theta)f^*\left(\frac{p(\theta/x_1,\ldots,x_n)}{p(\theta)}\right)\; d\lambda(\theta)\right] f(x_1,\ldots,x_n) d\mu(x_1,\ldots,x_n) \geq$$
$$\geq \int_Y \left[\int_\Theta p(\theta)f^*\left(\frac{p^T(\theta/t)}{p(\theta)}\right) d\lambda(\theta)\right) f(t)\; d\mu(t)$$

es decir,

$$D_{f^*}\left(X^{(n)}; p(\theta)\right) \geq D_{f^*}\left(T; p(\theta)\right)\; \forall p(\theta) \in \Theta^*$$

Si f^* es estrictamente convexa se dará la igualdad si y sólo si $p(\eta_{\eta,\theta}(x_1,\ldots,x_n) = p^T(\theta/t)) = 1\quad \forall(x_1,\ldots,x_n) \in K_T(t)$. Es decir, si $p^T(\theta/t) = p(\theta/x_1,\ldots,x_n)$ $\forall(x_1,\ldots,x_n) \in K_T(t)\,\mu$ c.s. o lo que es lo mismo si T es suficiente para θ.

Así pues, este teorema recoge el resultado intuitivo de que cualquier muestra es más informativa que cualquier función medible de la misma, salvo que el estadístico sea suficiente.

Teorema 4. Sea X un experimento cuyo espacio estadístico asociado es $(R, \beta(R), f(x/\theta))_{\theta\in\Theta}$ y sea g una función medible estrictamente monótona de $\mathbb{R}$ en $\mathbb{R}$. Entonces X es indiferente a $g(X)$.

Demostración.

La función medible g induce en R una medida de probabilidad $P_\theta^g(A) = P_\theta(g^{-1}(A))$ (siendo P_θ la medida de probabilidad asociada a $f(x/\theta)$ dando lugar al espacio estadístico $(R, \beta(R), P_\theta^g)_{\theta\in\Theta}$. Designaremos por $h^g(y/\theta)$ la función de densidad de P_θ^g respecto de una medida conveniente μ_g. Evidentemente,

$$h^g(y/\theta) = f(g^{-1}(y)/\theta)\ |(g^{-1}(y))'|, \quad p^g(y) = p(\theta/g^{-1}(y))$$

y

$$h^g(y) = \int_\Theta h^g(y/\theta)p(\theta(\ d\lambda(\theta) = f(g^{-1}(y))|(g^{-1}(y))'|$$

Por tanto,

$$\int_R h^g(y)\left[\int_\Theta p(\theta)f^*\left(\frac{p^g(\theta/y)}{p(\theta)}\right)\ d\lambda(\theta)\right]\ d\mu_g(y) =$$
$$= \int_R \left[\int_\Theta f(g^{-1}(y))|(g^{-1}(y))'|p(\theta)f^*\left(\frac{p(\theta)(g^{-1}(y))}{p(\theta)}\right)d\lambda(\theta)\right]d\mu_g(y)$$

es decir,

$$X \sim g(X)$$

El siguiente resultado establece que al efectuar una partición en el espacio muestral y agrupar en uno todos los elementos de cada clase, se pierde, por lo general, información, salvo en el caso en el que,

$$p(\theta/x) = p(\theta/E_1)\ c.s.\ \forall i \in \mathbb{N} \quad \text{y} \quad \forall x \in E_1$$

Teorema 5. Sea X_1 un experimento con espacio estadístico asociado $(X_1, \beta_{\chi_1}, f(x_1/\theta))_{\theta\in\Theta}$ y sea $(E_1)_{i\in\mathbb{N}}$ una partición medible de χ_1. Consideremos el nuevo experimento $X_2 = (\chi_2, \beta_{\chi_2}, Q_\theta)$, donde $\chi_2 = (E_1\ /\ i \in \mathbb{N})$, β_{χ_2} es la σ-álgebra engendrada por las $\{E_1\}_{i\in\mathbb{N}}$ y Q_θ es tal que $Q_\theta(E_1) = P_\theta(E_1)$. Entonces $X_1 \geq X_2$. Además si f^* es estrictamente convexa se da la igualdad si y sólo si $p(\theta/x) = p(\theta/E_1)\ \mu_1$ c.s. $\forall i \in \mathbb{N}$ y $x \in E_1$.

Demostración.

Por una parte

$$D_{f^*}[X_1, p(\theta)] = \int_\Theta p(\theta)\left[\int_{X_1} f(x) f^*\left(\frac{f(x/\theta)}{f(x)}\right) d\mu_1(x)\right] d\lambda(\theta)$$

por otra,

$$D_{f^*}[X_2, p(\theta)] = \sum_{i\in\mathbb{N}} P(E_1)\int_\Theta p(\theta) f^*\left(\frac{p(\theta/E_1)}{p(\theta)}\right] d\lambda(\theta) =$$

$$\int_\Theta p(\theta)\left[\sum_{i\in\mathbb{N}} P(E_1) f^*\left(\frac{p_\theta(E_1)}{p(E_1)}\right)\right] d\lambda(\theta)$$

y aplicando un resultado de Csiszar (1976), se tiene:

$$\int_{E_1} f(x) f^*\left(\frac{f(x/\theta)}{f(x)}\right) d\mu_1(x) \geq \int_{E_1} f(x) d\mu_1(x) f^*\left[\frac{\int_{E_1} f(x/\theta)\mu_1(x)}{\int_{E_1} f(x) d\mu_1(x)}\right] =$$

$$= P_{f(x)}(E_1) f^*\left[\frac{P_{f(x/\theta)}(E_1)}{P_{f(x)}(E_1)}\right]$$

Sumando los resultados correspondientes a todos los índes i, se llega a

$$\int_{X_1} f(x) f^*\left(\frac{f(x/\theta)}{f(x)}\right) d\mu_1(x) \geq \sum_{i\in\mathbb{N}} P_{f(x)}(E_1) f^*\left[\frac{P_{f(x/\theta)}(E_1)}{P_{f(x)}(E_1)}\right]$$

luego,

$$p(\theta)\int_{X_1} f(x) f^*\left(\frac{f(x/\theta)}{f(x)}\right) d\mu_1(x) \geq p(\theta)\sum_{i\in\mathbb{N}} P_{f(x)}(E_1) f^*\left(\frac{P_{f(x/\theta)}(E_1)}{P_{f(x)}(E_1)}\right)$$

y si integramos sobre Θ se sigue que

$$\int_\Theta p(\theta)\left(\int_{X_1} f(x) f^*\left(\frac{f(x/\theta)}{f(x)}\right) d\mu_1(x)\right) d\lambda(\theta) \geq$$

$$\geq \int_\Theta p(\theta)\left(\sum_{i\in N} P_{f(x)}(E_1) f^*\left(\frac{P_{f(x/\theta)}(E_1)}{P_{f(x)}(E_1)}\right)\right) d\lambda(\theta) =$$

$$= \int_\Theta p(\theta)\left(\sum_{i\in N} P_{(E_1)} f^*\left(\frac{P(E_1/\theta)}{P(E_1)}\right)\right) d\lambda(\theta)$$

es decir

$$X_1 \geq X_2$$

Además se da la igualdad si para cada i

$$\frac{f(x/\theta)}{f(x)} = \frac{P_{f(x/\theta)}(E_1)}{P_{f(x)}(E_1)}$$

es decir, se da la igualdad si para $i \in \mathbb{N}$,

$$p(\theta/x) = p(\theta/E_1) \quad \forall x \in E_1 \quad \mu_1 \text{c.s.}$$

Señalemos finalmente el papel relevante que juega en los resultados establecidos el hecho de que f^* sea estrictamente convexa ya que sin esta condición no podrían establecerse resultados relevantes como el relacionado con la información proporcionada por una muestra y un estadístico suficiente de ella, entre otros.

Un punto interesantes, que se deja para trabajos posteriores, es el de analizar la relación de este criterio con los de Lhemann, Blackwell y De Groot.

Referencias

[1] Blackwell, D., Comparison of Experiment. Proc. Sec. Berkeley Symp. on Math. Stat. and Probability, University California Press, 93-102 (1951).

[2] Csiszar, I., Information type measures of difference of probability distributions and indirect observations. Studia Scientarum Mathematicarum Hungarica 2, pp. 299-318 (1967)

[3] De Groot, M.H., Optimal Statistical Decisions. McGraw-Hill (1970)

[4] García-Carrasco, M.P., Criterios para la comparación de experimentos. Trabajos de Estad. e Inv. Operativa, 29, 28-51 (1978).

[5] Lindley, D.V., On a measure of the information provided by an experiment. Ann. Math. Statist. 27, 986-1005 (1956).

[6] Lindley, D.V., Binomial sampling and the concept of information. Biometrika 44, 179-186 (1957).

[7] Morales, D., Pardo, L. y Quesada, V., Las medidas de f^*–divergencia en el diseño secuencial de experimentos en un contexto Bayesiano. Trabajos de Estadística, Vol. 1, n.2, pp. 95-109 (1986).

[8] Pardo, J.A., Las f^*–divergencias como medidas de información en la inferencia bayesiana. Tesina presentada en la Univ. Complutense de Madrid, (1985).

[9] Pardo, L., Plan de muestreo secuencial basado en la Energía Informacional en el modelo de Bernouilli. Estadística Española, 104, 27-49, (1984).

[10] Pardo, L., The Order α Information Energy Gain in Sequential Design of Experiments. Actas del Third European Young Statisticians Meeting, 140-147, (1983).

[11] Pardo, L y Menéndez, M.L., Binominal Sampling Schemes to obtain a prescribed amount of Onnicescu's Information Energy. Actas del 16th. European meeting of Statisticians. Marburg (1984).

(*) Escuela Universitaria de Estadística.
Universidad Complutense de Madrid
(**) Departamento de Matemática Aplicada
Escuela Técnica Superior de Arquitectura
Universidad Politécnica de Madrid
(***) Departamento de Estadística e Investigación Operativa
Facultad de Matemáticas
Universidad Complutense de Madrid