INFORMACION ESPERADA PROPORCIONADA POR UN EXPERIMENTO ASOCIADO A UN DISEÑO ESTRATIFICADO

Ernesto Veres Ferrer
Instituto Nacional de Estadística

Recibido: 5 mayo 1982

PRESENTADO POR EL ACADÉMICO NUMERARIO D. FRANCISCO AZORÍN

El concepto de información esperada definido por Lindley (1956) puede generalizarse a experimentos más complejos que el considerado por dicho autor. Dentro de esta línea está la generalización (Bernardo, 1978) que define la información esperada útil como aquella que un experimento simple proporciona sobre una nueva cantidad de interés transformada de la primitiva. En el presente trabajo generalizamos la definición de Shannon-Lindley a experimentos asociados a diseños estratificados.

La posible importancia de las generalizaciones de este tipo reside en que las medidas de información pueden utilizarse en Inferencia Estadística, en cuanto que ésta se ocupa, precisamente, de cuantitativizar la información que se posee. En efecto, dicha rama de las matemáticas puede concebirse como cierto problema de decisión en el que el espacio de decisiones es el conjunto de posibles distribuciones finales que describen el comportamiento de la magnitud sobre la que se desea efectuar inferencia, y en el que la función de utilidad viene definida en términos de la información esperada conseguida. En consecuencia, al maximizar estas funciones de la información esperada a fin de diseñar un experimento óptimo se está aplicando el principio Bayesiano general de maximización de la utilidad esperada. Además, éste parece ser el único procedimiento razonable si el objetivo pretendido con la experimentación es exclusivamente el de aumentar nuestros conocimientos sobre la cantidad de interés (Bernardo, 1976).

1. Introducción y notación. Información esperada global

Supondremos, primeramente, que el objetivo pretendido por cierta investigación es el de mejorar nuestro conocimiento sobre el valor de cierta cantidad o parámetro de interés. Para ello el científico diseñará un experimento de forma que la distribución de sus resultados dependerá de la cantidad de interés de una manera conocida, de manera que la observación de tales resultados proporcionará cierta cantidad de información.

El modelo matemático apropiado para describir un experimento que supone la existencia de un diseño de estratificación contiene un espacio muestral X en el que se ha establecido una partición en L subconjuntos o estratos:

$$X = \bigcup_{i=1}^{L} X_i; \quad X_i \cap X_j = \emptyset \quad \text{para todo} \quad i \neq j$$

Cada uno de estos subespacios está dotado de una apropiada σ-álgebra sobre la que se define sendas familias de medidas de probabilidad. Las densidades de cada una de ellas se individualizan mediante un parámetro θ_i que toma valores en su respectivo espacio paramétrico Θ_i.

Para cada estrato se supone definido el experimento

$$E_i = \{X_i, \Theta_i, p(x_i/\theta_i)\}$$

que consiste en una observación de la variable aleatoria $x_i \in X_i$ que se distribuye, para algún $\theta_i \in \Theta_i$ dado, de acuerdo con la densidad $p(x_i/\theta_i)$. En estas condiciones definiremos el experimento compuesto $E^s(\mathbf{n})$ como aquel que está formado por los L experimentos componentes $E(n_i)$:

$$E^s(\mathbf{n}) = \{E(n_i)\}_{i=1}^{L}$$

y donde $E(n_i)$ es el experimento que consiste en n_i repeticiones independientes (dado θ_i) del experimento E_i. Con $\mathbf{n} = (n_1, n_2, \ldots, n_L)$

denotaremos al vector cuyas componentes son los respectivos tamaños muestrales de las muestras extraídas en sus estratos respectivos y que llamaremos el vector tamaño muestral.

La densidad de probabilidad que describe el comportamiento de la muestra

$$\mathbf{z}_i = (x_{i_1}, x_{i_2}, \ldots, x_{in_i})$$

resultado de la realización de E (n_i) viene dada a través del productorio

$$p\,(z_i/\theta_i) = \prod_{j=1}^{n_i} p\,(x_{ij}/\theta_i)$$

La muestra total obtenida tras la realización del experimento global $E^s(\mathbf{n})$ está formada por el conjunto de muestras parciales obtenidas de sendos estratos. La denotaremos por $z = (z_1, z_2, \ldots, z_L)$. Dado que las realizaciones de los experimentos E (n_i) en sus respectivos estratos son independientes entre sí, la densidad de probabilidad que describe el comportamiento de la muestra global z en función de todas las cantidades θ_i viene dada a través del doble productorio

$$p\,(z/\boldsymbol{\theta}) = \prod_{i=1}^{L} p\,(z_i/\theta_i) = \prod_{i=1}^{L} \prod_{j=1}^{n_i} p\,(x_{ij}/\theta_i)$$

y en donde con $\boldsymbol{\theta}$ denotamos el vector cuyas componentes son las respectivas cantidades de interés θ_i, esto es, $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_L)$. Dicho vector recibirá el nombre de vector cantidad de interés o vector paramétrico y tomará valores en el espacio paramétrico L-dimensional

$$\Theta = \prod_{i=1}^{L} \Theta_i .$$

Al fin de terminar de describir los elementos del problema, el argumento Bayesiano extiende este modelo básico y supone que $\boldsymbol{\theta}$ es soporte de una σ-álgebra, de forma que las opiniones iniciales del investigador antes de realizar $E^s(\mathbf{n})$ se describen a través de la densidad de probabilidad inicial $p\,(\boldsymbol{\theta})$ (DeFinetti, 1937). Dicha densidad pone de manifiesto la posible relación de dependencia y correlación entre las distintas componentes del vector paramétrico.

Una forma alternativa para efectuar la descripción sobre θ puede venir expresada a través de cierta familia de densidades iniciales $p(\theta_i/\varphi)$ $(i = 1, 2, \ldots, L)$, cuyas componentes individuales describen el comportamiento de cada parámetro θ_i por separado y en función de un único y nuevo parámetro φ. La densidad inicial sobre φ terminaría de describir el problema, al verificarse que

$$p(\theta) = \int p(\varphi) \cdot \prod_{i=1}^{L} p(\theta_i/\varphi) \cdot d\varphi$$

Siendo $p(\theta)$ la densidad inicial de θ, la densidad predictiva de los resultados experimentales proporcionados por $E^s(\mathbf{n})$ se define como

$$p(z) = \int p(z/\theta) \cdot p(\theta) \cdot d\theta$$

de forma que realizado $E^s(\mathbf{n})$ y obtenido z, las opiniones sobre θ del científico deberán describirse por la densidad posterior de θ dada por el Teorema de Bayes, esto es,

$$p(\theta/z) = p(z/\theta) \cdot p(\theta)/p(z)$$

Supondremos que, salvo indicación contraria, las integrales que aparecen en este trabajo están extendidas a todos los respectivos recintos de definición, así como su convergencia en todos los casos. Con todos estos elementos estamos ya en condiciones de definir una medida de la información esperada que sobre θ proporciona el experimento global $E^s(\mathbf{n})$.

DEFINICIÓN.—La información esperada sobre θ proporcionada por el experimento $E^s(\mathbf{n})$, cuando la densidad inicial es $p(\theta)$, se define a través de la doble integral

$$I^{\theta}\{E^s(\mathbf{n}),\ p(\theta)\} = \iint p(z)\, p(\theta/z) \cdot \log \frac{p(\theta/z)}{p(\theta)} \cdot d\theta \cdot dz$$

La anterior información recibirá el nombre de información esperada global y es la correspondiente generalización de la información esperada definida por Lindley (1956), siguiendo el trabajo de Shannon (1948). Respecto a la base de los logaritmos utilizados en la definición y por comodidad matemática, será la base e de los logaritmos neperianos. Notemos que la información definida es una doble esperanza del logaritmo.

$$\log \left(p\,(\boldsymbol{\theta}/z)/p\,(\boldsymbol{\theta}) \right)$$

que expresa, a su vez, una medida de la información que sobre $\boldsymbol{\theta}$ proporciona un resultado concreto z.

Un caso particular de todo el desarrollo anterior y de gran importancia es aquel que establece como hipótesis adicional la de la independencia entre las componentes del vector de interés. Bajo esta hipótesis, la densidad de probabilidad que describe a las opiniones iniciales del investigador vendrá dada por un producto de densidades de probabilidad que describen dichas opiniones sobre cada componente θ_i por separado, esto es,

$$p\,(\boldsymbol{\theta}) = \prod_{i=1}^{L} p\,(\theta_i)$$

Supuesta esta independencia y dado que las realizaciones de los experimentos componentes del experimento global son independientes, la muestra z_i resultado del experimento E (n_i) componente i-ésimo de $E^s\,(\mathbf{n})$ no contendrá información sobre aquellas cantidades θ_j distintas de θ_i. Esto implica, a su vez, la existencia de una aditividad perfecta entre la información esperada global y el conjunto de informaciones esperadas proporcionadas por los experimentos componentes. Este resultado es el contenido en el siguiente:

TEOREMA 1.—Para un experimento $E^s\,(\mathbf{n})$ y una densidad inicial

$$p\,(\boldsymbol{\theta}) = p\,(\theta_1) \cdot p\,(\theta_2) \ldots p\,(\theta_L),$$

la información esperada global es suma de las informaciones esperadas que sobre cada θ_i por separado proporcionan los respectivos experimentos componentes del experimento global.

DEMOSTRACIÓN.—Bajo la hipótesis de independencia

$$p\ \theta/z = \frac{1}{p\,(z)} \prod_{i=1}^{L} p\,(z_i/\theta_i) \cdot p\,(\theta_i)$$

con

$$p\,(z) = \prod_{i=1}^{L} \int p\,(z_i/\theta_i)\, p\,(\theta_i) \cdot d\,\theta_i = \prod_{i=1}^{L} p\,(z_i)$$

De donde

$$p\,(\theta/z) = \prod_{i=1}^{L} p\,(\theta_i/z_i)$$

Así pues,

$$\log \frac{p\,(\theta/z)}{p\,(\theta)} = \sum_{i=1}^{L} \log \frac{p\,(\theta_i/z_i)}{p\,(\theta_i)}$$

Finalmente, por la misma definición de información esperada global,

$$I^{\theta} \left\{ E^{s}(\mathbf{n}),\, p\,(\theta) \right\} = \iint \prod_{i=1}^{L} p\,(\theta_i) \cdot p\,(z_i/\theta_i) \cdot$$

$$\cdot \left\{ \sum_{i=1}^{L} \log \frac{p\,(\theta_i/z_i)}{p\,(\theta_i)} \right\} d\,(\theta) \cdot d\,z = \sum_{i=1}^{L} \iint p\,(\theta_i) \cdot p\,(z_i/\theta_i) \cdot$$

$$\cdot \log \frac{p\,(\theta_i/z_i)}{p\,(\theta_i)} \cdot d\,\theta_i \cdot d\,z_i = \sum_{i=1}^{L} I^{\theta_i} \left\{ E\,(n_i),\, p\,(\theta_i) \right\}$$

(c. q. d.).

Este teorema relaciona, bajo la hipótesis adicional de independencia de las componentes del vector cantidad de interés, una información esperada global con un conjunto de informaciones esperadas definidas por Lindley (1956). Bajo condiciones generales de no independencia no existe esta aditividad, tal como se pondrá de manifiesto en la sección tercera de este trabajo.

2. Propiedades matemáticas de la información esperada global

Vamos a comprobar que las cuatro propiedades fundamentales satisfechas por la información esperada siguen verificándose por la información esperada global, mediante una generalización de los métodos utilizados por Lindley (1956).

TEOREMA 2.—Para un experimento $E^s(\mathbf{n})$ y densidad inicial $p(\theta)$,

$$I^{\theta}\{E^s(\mathbf{n}),\ p(\theta)\} \geqq 0$$

con igualdad si y sólo si $p(\theta/z)$ es independiente del resultado muestral z excepto, a lo sumo, en un conjunto de valores de θ de medida nula.

DEMOSTRACIÓN.—La información esperada global puede escribirse como

$$I^{\theta}\{E^s(\mathbf{n}), p(\theta)\} = \iint g(z,\theta)\cdot \log g(z,\theta)\cdot p(z)\cdot p(\theta)\cdot dz\cdot d(\theta)$$

donde

$$g(z,\theta) = p(\theta/z)\ /\ p(\theta)$$

Pero al ser $x\cdot \log x$ una función convexa y al tomar esperanzas respecto z y θ, resulta:

$$\iint g(z,\theta)\cdot \log g(z,\theta)\cdot p(z)\cdot p(\theta)\cdot dz\cdot d\theta \geqq$$

$$\geqq \iint g(z,\theta)\cdot p(z)\cdot p(\theta)\cdot dz\cdot d\theta\cdot \log \iint g(z,\theta)\cdot p(z)\cdot p(\theta)\cdot dz\cdot d\theta$$

con igualdad si y sólo si $g(z,\theta)$ es una constante salvo un conjunto de medida nula, esto es, cuando $p(\theta/z) = p(\theta)$ salvo en dicho conjunto. Además:

$$\iint g(z,\theta)\cdot p(z)\cdot p(\theta)\cdot dz\cdot d\theta = \iint p(\theta/z)\cdot p(z)\cdot dz\cdot d\theta = 1$$

de forma que el segundo miembro de la desigualdad anterior se anula al tomar logaritmos en la doble integral. Esto completa la demostración (c. q. d.).

El teorema indica que, bajo condición de que la densidad posterior de θ varíe con z, el experimento $E^s(\mathbf{n})$ resulta, por término medio, informativo. Para ello, basta que la muestra de un solo estrato modifique las opiniones iniciales. Inversamente, si realizado $E^s(\mathbf{n})$ la información contenida en su resultado z no modifica las opiniones iniciales representadas a través de $p(\theta)$, entonces no existe ganancia de información por el hecho de realizar la experimentación.

Supongamos ahora los experimentos

$$E^s(\mathbf{n}) = \} E(n_i) \{_{i=1,2\ldots,L} \quad \text{y} \quad E^s(\mathbf{n}+1) = \} E(n_i+1) \{_{i1,2,\ldots,L}$$

Definimos el experimento

$$E^s \} (\mathbf{n}+1)^a / z(\mathbf{n}) \{ = \} E(n_i+1)^a / z_i \{_{i=1,\ldots,L}$$

cuyos experimentos componentes

$$E(n_i+1)^a / z_i = \} X_i\, \theta_i,\ p(x_{iu_{i+1}} / \theta_i,\ z_i) \{$$

consisten en la obtención de la $(n_i + 1)$-ésima observación supuestamente conocido el resultado de $E(n_i)$. Consideremos la información que el experimento

$$E^s \} (\mathbf{n}+1)^a / z(\mathbf{n}) \{$$

proporciona sobre θ una vez realizado $E^s(\mathbf{n})$ y obtenido el resultado concreto $z(\mathbf{n})$. A su valor esperado

$$I^{\theta} \} E^s(\mathbf{n}+1)^a / E^s(\mathbf{n}), p(\theta) \{$$

respecto las posibles muestras estratificadas $z(\mathbf{n})$ se le llamará la información esperada sobre θ proporcionada por $E^s(\mathbf{n}+1)^a$ una vez que se ha realizado $E^s(\mathbf{n})$. Una demostración en línea a las del Teorema 2 establece que

$$I^{\theta} \} E^s(\mathbf{n}+1)^a / E^s(\mathbf{n}), p(\theta) \{$$

es no negativa. En lo que sigue emplearemos la notación siguiente: $z(\mathbf{n})$ expresará la muestra estratificada

$$(z_1(n_1),\ z_2(n_2),\ \ldots,\ z_L(n_L)),$$

donde

$$z_i(n_i) = (x_{i1},\ x_{i2},\ \ldots,\ x_{in_i})$$

mientras que con z_n representaremos al conjunto de L unidades muestrales extraídas en sendos estratos: $x_{1\,n_1}, x_{2\,n_2}, \ldots, x_{L\,n_L}$.

TEOREMA 3.—Siendo $p(\theta)$ la densidad inicial y para los experimentos $E^s(\mathbf{n})$ y $E^s(\mathbf{n}+\mathbf{1})$, se verifica que

$$I^{\theta}\{E^s(\mathbf{n}+\mathbf{1}), p(\theta)\} = I^{\theta}\{E^s(\mathbf{n}), p(\theta)\} + I^{\theta}\{E^s(\mathbf{n}+\mathbf{1})^a/E^s(\mathbf{n}), p(\theta)\}$$

DEMOSTRACIÓN.—De sus definiciones respectivas:

$$I^{\theta}\{E^s(\mathbf{n}), p(\theta)\} = \iint p(\theta) \cdot p(z(\mathbf{n})/\theta) \cdot \log \frac{p(\theta/z(\mathbf{n}))}{p(\theta)} \cdot d\theta \cdot dz(\mathbf{n}) =$$

$$= \iint p(\theta) \cdot p(z(\mathbf{n}+\mathbf{1})/\theta) \cdot \log \frac{p(\theta/z(\mathbf{n}))}{p(\theta)} \cdot d\theta \cdot dz(\mathbf{n}+\mathbf{1})$$

$$I^{\theta}\{E^s(\mathbf{n}+\mathbf{1})^a/E^s(\mathbf{n}), p(\theta)\} =$$

$$= \int p(z(\mathbf{n})) \iint p(\theta)/z(\mathbf{n})) \cdot p(z_{n+1}/z(\mathbf{n}), \theta) \cdot \log \frac{p(\theta/z_{n+1}, z(\mathbf{n}))}{p(\theta/z, (\mathbf{n}))} \cdot d\theta \cdot$$

$$\cdot\, dz_{n+1} \cdot dz(\mathbf{n}) = \iint p(\theta) \cdot p(z(\mathbf{n}+\mathbf{1})/\theta) \cdot \log \frac{p(\theta/z(\mathbf{n}+\mathbf{1}))}{p(\theta/z(\mathbf{n}))} \cdot d\theta \cdot dz(\mathbf{n}+\mathbf{1})$$

Sumando ambas informaciones esperadas resulta finalmente

$$\iint p(\theta) \cdot p(z(\mathbf{n}+\mathbf{1})/\theta) \cdot \log \frac{p(\theta/z(\mathbf{n}+\mathbf{1}))}{p(\theta)} \cdot d\theta \cdot dz(\mathbf{n}+\mathbf{1}) =$$

$$= I^{\theta}\{E^s(\mathbf{n}+\mathbf{1}), p(\theta)\}$$

(c. q. d.).

Obviamente este resultado puede generalizarse a un número finito de experimentos con espacio paramétrico común Θ. Resultaría como expresión generalizada:

$$\mathrm{I}^{\theta} \{E^{s}(\mathbf{n}+\mathbf{k}),\, p(\theta) = \mathrm{I}^{\theta} \{E^{s}(\mathbf{n}),\, p(\theta)\} +$$

$$+ \sum_{i=1}^{k} \mathrm{I}^{\theta} \{E^{s}(\mathbf{n}+\mathbf{i})^{a}/E^{s}(\mathbf{n}+\mathbf{i}-\mathbf{1}),\, p(\theta)\}$$

proporcionando así una regla necesaria para la adición de grupos de información provenientes de diferentes caminos. Como importante e inmediata consecuencia es el siguiente:

Corolario.—Siendo $z(\mathbf{n})$ suficiente para

$$z(\mathbf{n}+\mathbf{1}) = (z_{n+1},\, z(\mathbf{n}))$$

con respecto a θ, en el sentido de que

$$p(\theta/z(\mathbf{n}+\mathbf{1})) = p(\theta/z(\mathbf{n})),$$

entonces

$$\mathrm{I}^{\theta} \{E^{s}(\mathbf{n}+\mathbf{1}),\, p(\theta)\} = \mathrm{I}^{\theta} \{E^{s}(\mathbf{n}),\, p(\theta)\}$$

Demostración.—Por misma definición de

$$\mathrm{I}^{\theta} \{E^{s}(\mathbf{n}+\mathbf{1})^{a}/E^{s}(\mathbf{n}),\, p(\theta)\}, \quad \text{si} \quad p(\theta/z(\mathbf{n}+\mathbf{1})) = p(\theta/z(\mathbf{n})),$$

dicha información es nula por el Teorema 2. El resultado, pues, se sigue de forma inmediata del Teorema 3 (c. q. d).

Este corolario establece que no existe pérdida de información si ésta se ciñe a la observación de un estadístico suficiente con respecto a la cantidad de interés. E inversamente, ante un estadístico insuficiente en el sentido descrito existe pérdida de información al ser

$$\mathrm{I}^{\theta} \{E^{s}(\mathbf{n}+\mathbf{1})^{a}/E^{s}(\mathbf{n}),\, p(\theta)\}$$

estrictamente positiva.

El comportamiento de la información esperada global como función del vector tamaño muestral es análogo al seguido por la información esperada de Shannon-Lindley. Y así, el crecimiento y concavidad como función de $\mathbf{n}$ satisfechos por

$$I^{\theta} \{E^s(\mathbf{n}), p(\theta)\}$$

no hace sino confirmar el conocido principio de la utilidad marginal decreciente de sucesivas observaciones independientes y equidistribuidas.

Teorema 4.—Para un experimento $E^s(\mathbf{n})$ y densidad inicial $p(\theta)$, la información esperada global es una función creciente y cóncava del vector tamaño muestral.

Demostración.—Por el Teorema 3 es conocido ya que

$$I^{\theta} \{E^s(\mathbf{n}+1), p(\theta)\} - I^{\theta} \{E^s(\mathbf{n}), p(\theta)\} = I^{\theta} \{E^s(\mathbf{n}+1)^a/E^s(\mathbf{n}), p(\theta)\} \geq 0$$

lo que asegura el crecimiento de la información esperada global como función del vector $\mathbf{n}$. Para probar su concavidad es suficiente demostrar que la expresión

$$\Delta = 2.I^{\theta} \{E^s(\mathbf{n}), p(\theta)\} - I^{\theta} \{E^s(\mathbf{n}+1), p(\theta)\} - I^{\theta} \{E^s(\mathbf{n}-1), p(\theta)\}$$

es positiva. Utilizando nuevamente las expresiones desarrolladas en el Teorema 3 resulta:

$$\Delta = I^{\theta} \{E^s(\mathbf{n})^a/E^s(\mathbf{n}-1), p(\theta)\} - I^{\theta} \{E^s(\mathbf{n}+1)^a/E^s(\mathbf{n}), p(\theta)\} =$$

$$= \iint p(\theta) \cdot p(z(\mathbf{n}+1)/\theta) \cdot \log f(\theta, z(\mathbf{n}+1)) \cdot d\theta \cdot dz(\mathbf{n}+1)$$

donde

$$f(\theta, z(\mathbf{n}+1)) = \frac{p(\theta/z(\mathbf{n}))^2}{p(\theta/z(\mathbf{n}-1)) \cdot p(\theta/z(\mathbf{n}+1))} =$$

$$= \frac{p(z(\mathbf{n})/\theta)^2}{p(z(\mathbf{n}-1)/\theta) \cdot p(z(n+1)/\theta)} \cdot \frac{p(z(\mathbf{n}-1)) \cdot p(z(\mathbf{n}+1))}{p(z(\mathbf{n}))^2} =$$

$$= \prod_{i=1}^{L} \frac{p(x_{in_i}/\theta_i)}{p(x_{in_i+1}/\theta_i)} \cdot \frac{p(z(\mathbf{n}-1))}{p(z_n, z(\mathbf{n}-1))} \cdot \frac{p(z_{n+1}, z(\mathbf{n}))}{p(z(\mathbf{n}))} =$$

$$= \prod_{i=1}^{L} \frac{p(x_{in_i}/\theta_i)}{p(x_{in_i+1}/\theta_i)} \cdot \frac{p(z_{n+1}/z(\mathbf{n}))}{p(z_n/z(\mathbf{n}-1))}$$

y en donde las cantidades aleatorias x_{in_i} y x_{in_i+1} están equidistribuidas. Así pues:

$$\Delta = \sum_{i=1}^{L} \int\int p(\theta) \cdot p(z(\mathbf{n}+1)/\theta) \cdot \log \frac{p(x_{in_i}/\theta_i)}{p(x_{in_i+1}/\theta_i)} \cdot d\theta \cdot dz(\mathbf{n}+1) +$$

$$+ \int\int p(\theta) \cdot p(z(\mathbf{n}+1)/\theta) \cdot \log \frac{p(z_{n+1}/z(\mathbf{n}))}{p(z_n/z(\mathbf{n}-1))} \cdot d\theta \cdot dz(\mathbf{n}+1) =$$

$$= \int\int p(z(\mathbf{n})) \cdot p(z_{n+1}/z(\mathbf{n})) \cdot \log \frac{p(z_{n+1}/z(\mathbf{n}))}{p(z_{n+1}/z(\mathbf{n}-1))} \cdot dz_{n+1} \cdot dz(\mathbf{n})$$

expresión que puede interpretarse como la información esperada que sobre z_{n+1} proporciona z_n y que, por el Teorema 2, es positiva (c. q. d.).

El comportamiento de la información esperada global como funcional de la densidad inicial queda recogido en el siguiente:

TEOREMA 5.—$I^{\theta}\{E^s(\mathbf{n}), p(\theta)\}$ es una funcional cóncava de la densidad inicial $p(\theta)$.

Demostración.—Sean $p_1(\theta)$ y $p_2(\theta)$ dos densidades de probabilidad iniciales para θ, y $\lambda \in [0, 1]$. Denotamos por $p(\theta)$ a la densidad combinación lineal

$$p(\theta) = \lambda\, p_1(\theta) + (1-\lambda)\, p_2(\theta)$$

Se verifica que

$$I^{\theta}\{E^s(\mathbf{n}), p(\theta)\} - \lambda I^{\theta}\{E^s(\mathbf{n}), p_1(\theta)\} - (1-\lambda) I^{\theta}\{E^s(\mathbf{n}), p_2(\theta)\} =$$

$$= \int\int p(z(\mathbf{n})/\theta) \cdot (\lambda p_1(\theta) + (1-\lambda) p_2(\theta)) \cdot \log \frac{p(z(\mathbf{n})/\theta)}{p(z(\mathbf{n}))} \cdot d\theta \cdot dz(\mathbf{n}) -$$

$$- \lambda \int\int p(z(\mathbf{n})/\theta) \cdot p_1(\theta) \cdot \log \frac{p(z(\mathbf{n})/\theta)}{p_1(z(\mathbf{n}))} \cdot d\theta \cdot dz(\mathbf{n}) -$$

$$-(1-\lambda)\iint p(z(\mathbf{n})/\theta)\cdot p_2(\theta)\cdot\log\frac{p(z(\mathbf{n})/\theta)}{p_2(z(\mathbf{n}))}\cdot d\theta\cdot dz(\mathbf{n})=$$

$$=\lambda\int p_1(z(\mathbf{n}))\cdot\log\frac{p_1(z(\mathbf{n}))}{p(z(\mathbf{n}))}\cdot dz(\mathbf{n})+(1-\lambda)\int p_2(z(\mathbf{n}))\cdot$$

$$\cdot\log\frac{p_2(z(\mathbf{n}))}{p(z(\mathbf{n}))}\cdot dz(\mathbf{n})$$

donde

$$p_i(z(\mathbf{n}))=\int p(z(\mathbf{n})/\theta)\cdot p_i(\theta)\cdot d\theta \qquad i=1,2$$

$$p(z(\mathbf{n}))=\lambda p_1(z(\mathbf{n}))+(1-\lambda)p_2(z(\mathbf{n}))$$

El resultado del teorema se sigue sin más que considerar que las dos últimas integrales son no negativas, empleando un razonamiento análogo al de la demostración del Teorema 2 (c. q. d.).

Del conjunto de teoremas anterior queda establecida la coincidencia entre las propiedades satisfechas por la información esperada proporcionada por un experimento simple E (n) y las satisfechas por la información esperada global.

3. Información esperada proporcionada por $E^s(n)$ sobre un parámetro global y sobre una componente concreta del vector de interés

En las invesigaciones prácticas se suele estar interesado en el conocimiento de cierto parámetro global φ que responde a cierta característica válida para todos los elementos del espacio muestral. Así pues, supongamos que al investigador le interesa adquirir conocimiento sobre φ y, para ello, idea un experimento de la forma $E^s(\mathbf{n})$, esto es, asociado a la obtención de una muestra estratificada. Planteado así el problema, la estratificación va a afectar doblemente, por una parte al espacio muestral en el que se establece una partición en estratos; por otra, va a afectar íntimamente a la naturaleza y descripción del parámetro φ. En efecto, en cada uno de los estratos el parámetro θ_i asociado al experimento componente E (n_i) tiene ahora la misma interpretación conceptual que el parámetro global φ, pero con el ámbito de aplicación restringido a su estrato respectivo. Por ejemplo, siendo φ el nivel de paro como cantidad de interés a inferir y ante una posible estratificación regional, θ_i representaría dicha proporción de parados pero referida a la región i-ésima. Deducimos con esto dos importantes consecuencias. En primer lugar y a pesar de la coincidencia conceptual en su definición, las cantidades φ y θ_i ($i = 1, 2, \ldots, L$) son todas distintas entre sí, pudiendo incluso (no sería, evidentemente, el caso más común) tomar valores excluyentes; en segundo lugar, el conocimiento de todas las componentes θ_i del vector $\boldsymbol{\theta}$ permite, a su vez, conocer al parámetro global φ. Esto es, supondremos la existencia de una cierta relación funcional $\varphi = f(\boldsymbol{\theta})$ que liga al parámetro global φ con el vector paramétrico $\boldsymbol{\theta}$ cuyas componentes θ_i admiten, como hemos dicho, precisamente la misma interpretación conceptual que φ pero con ámbito de aplicación en su estrato respectivo. Por regla general las características que son objeto de estudio en las investigaciones estadísticas vienen representadas por un total, un valor medio o una proporción. En estos casos, la relación funcional que liga a las cantidades θ_i ($i = 1, 2, \ldots, L$) con el parámetro global φ es lineal, si bien en el presente trabajo se supondrá, y salvo especificación contraria, una relación cualquiera a fin de conseguir una mayor generalidad. Finalmente, al estar interesado en inferir el valor de φ, el investigador deberá cuantificar la infromación esperada que sobre dicho parámetro proporciona el experimento global $E^s(\mathbf{n})$.

Notemos que tal como ha quedado establecida esta nueva situación la cuantificación de la información que sobre φ proporciona $E^s(\mathbf{n})$ y que denotaremos por $I^{\varphi}\{E^s(\mathbf{n}), p(\varphi)\}$, no se trata sino de la aplicación para el experimento global del concepto de información esperada útil definido por Bernardo (1978). Dicha información esperada sobre φ vendría dada por la doble integral

$$I^{\varphi}\{E^s(\mathbf{n}), p(\theta)\} = \iint p(z) \cdot p(\varphi/z) \cdot \log \frac{p(\varphi/z)}{p(\varphi)} \cdot d\varphi \cdot dz$$

donde:

z es el resultado del experimento global $E^s(\mathbf{n})$;

$p(\varphi)$ es la densidad inicial deducida de $p(\boldsymbol{\theta})$ por la transformación $\varphi = f(\boldsymbol{\theta})$, y

$p(\varphi/z)$ es la densidad posterior deducida de $p(\boldsymbol{\theta}/z)$ por la transformación $\varphi = f(\boldsymbol{\theta})$, y que llamaremos información esperada global útil sobre φ frente a

$$\mathrm{I}^{\boldsymbol{\theta}}\{\mathrm{E}^s(\mathbf{n}), p(\boldsymbol{\theta})\}$$

que sería la información esperada global total.

La situación planteada por la cuantificación de la información esperada que el experimento global proporciona sobre una determinada cantidad de interés θ_i, componente i-ésima del vector de interés $\boldsymbol{\theta}$, es un caso particular de información esperada global útil, donde la transformación entre parámetros es la correspondiente función marginal i-ésima. La expresión de esta información esperada —que denotaremos por

$$\mathrm{I}^{\boldsymbol{\theta}_i}\{\mathrm{E}^s(\mathbf{n}), p(\boldsymbol{\theta})\} \text{ —}$$

vendrá dada por la doble integral

$$\mathrm{I}^{\theta_i}\{\mathrm{E}^s(\mathbf{n}), p(\boldsymbol{\theta})\} = \iint p(z) \cdot p(\theta_i/z) \cdot \log \frac{p(\theta_i/z)}{p(\theta_i)} \cdot d\theta_i \cdot dz$$

donde

z es el resultado del experimento global $\mathrm{E}^s(\mathbf{n})$;

$p(\theta_i)$ es la correspondiente densidad marginal i-ésima de la inicial $p(\boldsymbol{\theta})$, y

$p(\theta_i/z)$ es la correspondiente densidad marginal i-ésima de la posterior $p(\boldsymbol{\theta}/z)$.

Siendo

$$\mathrm{I}^{\varphi}\{\mathrm{E}^s(\mathbf{n}), p(\boldsymbol{\theta})\} \text{ e } \mathrm{I}^{\boldsymbol{\theta}_i}\{\mathrm{E}^s(\mathbf{n}), p(\boldsymbol{\theta})\}$$

dos informaciones esperadas globales y útiles y empleando idénticas técnicas de demostración que las contenidas en Bernardo (1978), puede demostrarse que ambas verifican las tres primeras propiedades de la información esperada global. En particular, ambas son no negativas y solamente se anulan cuando la muestra z no modifica las opiniones iniciales que sobre φ ó θ_i, respectivamente, posee el investigador; verifican idénticos criterios de aditividad, al ser

$$\mathrm{I}^{\varphi}\{\mathrm{E}^s(\mathbf{n}+1), p(\theta)\} = \mathrm{I}^{\varphi}\{\mathrm{E}^s(\mathbf{n}), p(\theta)\} + \mathrm{I}^{\varphi}\{\mathrm{E}^s(\mathbf{n}+1)^a/\mathrm{E}^s(\mathbf{n}), p(\theta)\}$$

$$\mathrm{I}^{\theta_i}\{\mathrm{E}^s(\mathbf{n}+1), p(\theta)\} = \mathrm{I}^{\theta_i}\{\mathrm{E}^s(\mathbf{n}), p(\theta)\} + \mathrm{I}^{\theta_i}\{\mathrm{E}^s(\mathbf{n}+1)^a/\mathrm{E}^s(\mathbf{n}), p(\theta)\}$$

de forma que no existe pérdida de información cuando la muestra z se sustituye por la observación de estadísticos suficientes respecto φ o θ_i; finalmente, consideradas como funciones del vector tamaño muestral, ambas son crecientes y cóncavas. Pero además, y por su misma definición en términos de marginales i-ésimas,

$$\mathrm{I}^{\theta_i}\{\mathrm{E}^s(\mathbf{n}), p(\theta)\}$$

es una funcional cóncava de la densidad inicial $p(\theta)$, propiedad ésta no satisfecha por cualquier información esperada útil.

Teorema 6.—$\mathrm{I}^{\theta_i}\{\mathrm{E}^s(\mathbf{n}), p(\theta)\}$ es una funcional cóncava de la densidad inicial $p(\theta)$.

Demostración.—Basta aplicar el razonamiento seguido en el Teorema 5 sin más que considerar que la igualdad

$$p(\theta) = \lambda p_1(\theta) + (1-\lambda) p_2(\theta)$$

conduce a

$$p(\theta_i) = \int p(\theta)\cdot d\theta' = \lambda \int p_1(\theta)\cdot d\theta' + (1-\lambda)\int p_2(\theta)\cdot d\theta' = \\ = \lambda p_1(\theta_i) + (1-\lambda) p_2(\theta_i)$$

donde las integrales anteriores están extendidas al conjunto

$$\Theta' = \prod_{j \neq i}^{L} \Theta_j$$

y

$$d\theta' = \prod_{j \neq i}^{L} d\theta_j$$

(c. q. d.).

Siendo $\psi = \psi(\theta)$ una transformación cualquiera del vector cantidad de interés, una forma alternativa para definir la información esperada global útil viene dada por la doble integral:

$$\mathrm{I}^{\psi}\{\mathbf{E}^s(\mathbf{n}), p(\theta)\} = \iint p(z/\theta) \cdot p(\theta) \cdot \log \frac{p(\psi(\theta)/z)}{p(\psi(\theta))} \cdot d\theta \cdot dz$$

que sirve para poner de manifiesto el hecho de que la información esperada global útil es invariante ante transformaciones biyectivas del vector de interés. En efecto:

Teorema 7.—Siendo $\psi = \psi(\theta)$ una transformación biyectiva, las informaciones esperadas globales que el experimento $\mathrm{E}^s(\mathbf{n})$ proporciona sobre ψ y sobre θ son iguales.

Demostración.—Siendo $\psi = \psi(\theta)$ biyectiva

$$p(\psi) = \{p(\theta)\}_{\theta=\psi} \cdot |\mathrm{J}| \quad \text{y} \quad p(\psi/z) = \{p(\theta/z)\}_{\theta=\psi} \cdot |\mathrm{J}|$$

donde

$$|\mathrm{J}| = \left| \frac{\partial \theta}{\partial \psi} \right| = \begin{vmatrix} \dfrac{\partial \theta_1}{\partial \psi_1} & \cdots & \dfrac{\partial \theta_L}{\partial \psi_1} \\ \cdots & \cdots & \cdots \\ \dfrac{\partial \theta_1}{\partial \psi_L} & \cdots & \dfrac{\partial \theta_L}{\partial \psi_L} \end{vmatrix}$$

Definiendo

$$\mathrm{I}^{\theta}\{z, p(\theta)\} = \int p(\theta/z) \cdot \log \frac{p(\theta/z)}{p(\theta)} \cdot d\theta$$

resulta ser

$$\mathrm{I}^{\theta}\{z, p(\theta)\} = \int p(\theta/z) \cdot \log \frac{p(\theta/z)\ |\mathrm{J}|}{p(\theta)\ \cdot |\mathrm{J}|}\, d\theta =$$

$$= \int p(\psi/z) \cdot \log \frac{p(\psi/z)}{p(\psi)}\, d\psi = \mathrm{I}^{\psi}\{z, p(\psi)\}$$

De ahí que trivialmente

$$\mathrm{I}^{\theta}\,\}\mathrm{E}^{s}(\mathbf{n}), p(\theta)\{ = \int p(z)\cdot \mathrm{I}^{\theta}\,\}z, p(\theta)\{\cdot dz =$$

$$= \int p(z)\cdot \mathrm{I}^{\psi}\,\{z, p(\psi)\}\cdot dz = \mathrm{I}^{\psi}\,\}\mathrm{E}^{s}(\mathbf{n}), p(\theta)\{$$

(c. q. d.).

Como era de esperar, la información esperada global útil no puede ser mayor que la información esperada global total. Para demostrar esta afirmación, sigamos la misma técnica empleada por Bernardo (1978).

Teorema 8.—Para un experimento $\mathrm{E}^{s}(\mathbf{n})$, densidad inicial $p(\theta)$ y transformación $\psi = \psi(\theta)$

$$\mathrm{I}^{\psi}\,\}\mathrm{E}^{s}(\mathbf{n}), p(\theta)\{ \leq \mathrm{I}^{\theta}\,\}\mathrm{E}^{s}(\mathbf{n}), p(\theta)\{$$

Demostración.—Sea $\eta = \{\psi, \omega\}$ una transformación biyectiva del vector paramétrico θ. En virtud del Teorema 7:

$$\mathrm{I}^{\theta}\,\{\mathrm{E}^{s}(\mathbf{n}), p(\theta)\} = \mathrm{I}^{\eta}\,\{\mathrm{E}^{s}(\mathbf{n}), p(\eta)\}$$

de donde

$$\mathrm{I}^{\theta}\,\{\mathrm{E}^{s}(\mathbf{n}), p(\theta)\} - \mathrm{I}^{\psi}\,\{\mathrm{E}^{s}(\mathbf{n}), p(\theta)\} = \mathrm{I}^{\eta}\,\{\mathrm{E}^{s}(\mathbf{n}), p(\eta)\} -$$

$$- \mathrm{I}^{\psi}\,\{\mathrm{E}^{s}(\mathbf{n}), p(\eta)\} = -\int p(\eta)\cdot\log p(\eta)\cdot d\eta +$$

$$+ \iint p(\eta)\cdot p(z/\eta)\cdot\log p(\eta/z)\cdot d\eta\cdot dz + \int p(\psi)\cdot\log p(\psi)\cdot d\psi -$$

$$- \iint p(\psi)\cdot p(z/\psi)\cdot\log p(\psi/z)\cdot d\psi\cdot dz$$

Por otra parte es evidente que

$$\int p(\eta)\cdot\log p(\eta)\cdot d\eta = \int p(\psi)\cdot\log p(\psi)\cdot d\psi +$$

$$+ \iint p(\psi)\cdot p(\omega/\psi)\cdot\log p(\omega/\psi)\cdot d\omega\cdot d\psi$$

de forma que sustituyendo

$$\mathrm{I}^{\theta} \{ \mathrm{E}^s(\mathbf{n}), p(\theta) \} - \mathrm{I}^{\psi} \{ \mathrm{E}^s(\mathbf{n}), p(\theta) \} =$$

$$= \int p(\psi) \iint p(\omega/\psi) \cdot p(z/\omega, \psi) \cdot \log \frac{p(\omega/\psi, z)}{p(\omega/\psi)} \cdot d\omega \cdot dz \cdot d\psi$$

que es la información esperada sobre ω proporcionada por $\mathrm{E}^s(\mathbf{n})$ una vez que ψ es conocido. Y dicha información es positiva por un argumento similar al usado para demostrar el Teorema 2 (c. q. d.).

La comparación entre las informaciones esperadas que sobre θ_i proporcionan los experimentos $\mathrm{E}^s(\mathbf{n})$ y $\mathrm{E}(n_i)$ pone de manifiesto el también resultado lógico de que la muestra estratificada z contiene más información, por término medio, sobre θ_i que la respectiva submuestra z_i.

TEOREMA 9.

$$\mathrm{I}^{\theta_i} \{ \mathrm{E}^s(\mathbf{n}), p(\theta) \} \geq \mathrm{I}^{\theta_i} \{ \mathrm{E}(n_i), p(\theta_i) \}$$

DEMOSTRACIÓN.—De sus mismas definiciones

$$\Delta = \mathrm{I}^{\theta_i} \{ \mathrm{E}^s(\mathbf{n}), p(\theta) \} - \mathrm{I}^{\theta_i} \{ \mathrm{E}(n_i), p(\theta_i) \} =$$

$$= \iiint p(\theta_i, z', z_i) \cdot \log \frac{p(\theta_i / z', z_i)}{p(\theta_i / z_i)} \cdot d\theta_i \cdot dz_i \cdot dz'$$

donde

$$z' = (z_1, z_2, \ldots, z_{i-1} z_{i+1}, \ldots, z_L)$$

Llamando

$$g(\theta_i, z_i, z') = \frac{p(\theta_i/z_i, z')}{p(\theta_i/z_i)}$$

resulta finalmente

$$\Delta = \iint g(\theta_i, z_i, z') \cdot \log g(\theta_i, z_i, z') \cdot p(z) \cdot p(\theta_i/z_i) \cdot d\theta_i \cdot dz$$

expresión que es no negativa por idéntica razón a la utilizada en la demostración del Teorema 2 (c. q. d.).

La hipótesis adicional de independencia de las componentes del vector de interés $\boldsymbol{\theta}$ se traduce en el resultado esperado de que la información sobre θ_i viene canalizada exclusivamente a través del resultado muestral z_i.

TEOREMA 10.—Siendo la densidad inicial $p(\boldsymbol{\theta})$ un producto de densidades definidas sobre cada una de sus componentes, entonces

$$\mathrm{I}^{\theta_i}\{\mathrm{E}(n_i), p(\theta_i)\} = \mathrm{I}^{\theta_i}\{\mathrm{E}^s(\mathbf{n}), p(\boldsymbol{\theta})\}$$

DEMOSTRACIÓN.—Bajo la hipótesis de independencia

$$p(\theta_i/z) = p(\theta_i/z_i) \quad \text{y} \quad p(z) = \prod_{i=1}^{L} p(z_i)$$

de donde

$$\mathrm{I}^{\theta_i}\{\mathrm{E}^s(\mathbf{n}), p(\boldsymbol{\theta})\} = \iint p(z) \cdot p(\theta_i/z) \cdot \log \frac{p(\theta_i/z)}{p(\theta_i)} \cdot d\theta_i \cdot dz =$$
$$= \iint p(z_i) \cdot p(\theta_i/z) \cdot \log \frac{p(\theta_i/z_i)}{p(\theta_i)} \cdot d\theta_i \cdot dz_i = \mathrm{I}^{\theta_i}\{\mathrm{E}(n_i), p(\theta_i)\}$$

(c. q. d.).

De los resultados de los Teoremas 1 y 10 deducimos finalmente que ante la independencia de las componentes del vector $\boldsymbol{\theta}$ se verifica que

$$\sum_{i=1}^{L} \mathrm{I}^{\theta_i}\{\mathbf{E}(n_i), p(\theta_i)\} = \sum_{i=1}^{L} \mathrm{I}^{\theta_i}\{\mathrm{E}^s(\mathbf{n}), p(\boldsymbol{\theta})\} = \mathrm{I}^{\boldsymbol{\theta}}\{\mathbf{E}^s(\mathbf{n}), p(\boldsymbol{\theta})\}$$

4. El modelo normal

La extensión del modelo normal a un experimento de la forma $\mathrm{E}^s(\mathbf{n})$ supone considerar un conjunto de L estratos o subpoblaciones sobre las que se definen sendas densidades normales y dependientes de cada componente θ_i del vector de interés $\boldsymbol{\theta}$, N (θ_i, σ_i) $i = 1, 2, \ldots, L$, y en donde todas las desviaciones σ_i se suponen conocidas. Debido a la independencia de las muestras parciales, la densidad conjunta que describe el comportamiento de la muestra global z viene dada por

$$p\,(z/\boldsymbol{\theta}) = \prod_{i=1}^{L} \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi}\,\sigma_i} \exp\left(-\frac{1}{2}\left(\frac{x_{ij}-\theta_i}{\sigma_i}\right)^2\right)$$

La densidad inicial $p\,(\boldsymbol{\theta})$ termina de especificar completamente al modelo. En general supondremos que $p\,(\boldsymbol{\theta})$ es también una multinormal de vector media $\boldsymbol{\theta}_0$ y matriz de momentos H_0, de componentes σ^0_{ij}. Así pues,

$$p\,(\boldsymbol{\theta}) = \frac{1}{(\sqrt{2\pi})^L\,|H_0|^{1/2}} \exp\left(-\frac{1}{2}\,(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\,H_0^{-1}\,(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\right)$$

Esta hipótesis sobre la distribución de $\boldsymbol{\theta}$ abarca también una serie de casos en los que el investigador prefiere describir sus opiniones iniciales sobre cada componente θ_i por separado a través de un conjunto de L densidades $p\,(\theta_i/\varphi)$, donde φ es un nuevo parámetro que expresa la dependencia entre los θ_i. En efecto, siendo las densidades $p\,(\theta_i/\varphi)$ normales de media φ y varianza conocida σ_i^2 y distribuyéndose φ según una N (φ_0, σ_0), la distribución conjunta de las cantidades θ_i es (Veres, 1981) multinormal L-dimensional $N^L\,(\boldsymbol{\varphi}_0, S^{-1})$, donde $\boldsymbol{\varphi}_0$ es el vector de componentes iguales a φ_0 y S es la matriz de elementos s_{pq}

$$s_{pp} = \sum_{\substack{i \geq 0 \\ i \neq p}}^{L} \prod_{j \neq p,\, i}^{L} \sigma_j^2 \Big/ \sum_{i=0}^{L} \prod_{j \neq i}^{L} \sigma_j^2 \qquad p = 1, 2, \ldots, L$$

$$s_{pq} = -\prod_{\substack{i \geq 0 \\ i \neq p,\, q}}^{L} \sigma_i^2 \Big/ \sum_{i=0}^{L} \prod_{j \neq i}^{L} \sigma_j^2 \qquad p, q = 1, 2, \ldots, L \quad p \neq q$$

En definitiva, consideraremos como hipótesis básicas del modelo normal asociado al experimento $E^s\,(\mathbf{n})$ las dos siguientes:

I. $p(x_i/\theta_i)$ es $N(\theta_i, \sigma_i)$, $i = 1, 2, \ldots, L$.

II. $p(\boldsymbol{\theta})$ es multinormal $N(\boldsymbol{\theta}_0, H_0)$.

De esta forma el experimento $E^s(\mathbf{n})$ se definirá a partir de la familia de L experimentos $E(n_i)$, donde

$$E_i = \{X_i, \theta_i, N(\theta_i, \sigma_i)\}\ i = 1, 2, \ldots, L$$

Denotemos por H a la matriz diagonal L × L cuyos elementos son las varianzas σ_i^2. Consideremos las respectivas matrices de precisión R y T (de elementos r_p y t_{pq}, respectivamente) de las matrices de momentos H y H_0. Sea N la matriz diagonal L × L de elementos los tamaños muestrales de las submuestras parciales extraídas de sendos estratos. Finalmente, denotemos por $\bar{\mathbf{x}}$ el vector columna L × 1 cuyas componentes son las medias muestrales de las submuestras parciales. Bajo estas condiciones demostremos el siguiente:

LEMA.—Bajo las hipótesis del modelo normal,

$$p(\boldsymbol{\theta}/z) = N^L(\boldsymbol{\theta}^*, (T + NR)^{-1})$$

con

$$\boldsymbol{\theta}^* = (T + NR)^{-1}(T\boldsymbol{\theta}_0 + NR\bar{\mathbf{x}})$$

DEMOSTRACIÓN.—La función de verosimilitud satisface la siguien te relación:

$$p(z/\boldsymbol{\theta}) \propto \exp\left(-\frac{1}{2}\sum_{i=1}^{L}\sum_{j=1}^{n_i}\left(\frac{x_{ij}-\theta_i}{\sigma'_i}\right)^2\right) =$$

$$= \prod_{i=1}^{L} \exp\left(-\frac{1}{2}\sum_{j=1}^{n_i} r_i (x_{ij}-\theta_i)^2\right)$$

Pero al ser

$$\sum_{j=1}^{n_i}(x_{ij}-\theta_i)^2 = n_i(\theta_i - \bar{x}_i)^2 + \sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)^2 \quad \text{con} \quad \bar{x}_i = \sum_{j=1}^{n_i}\frac{x_{ij}}{n_i}$$

resulta finalmente

$$p(z/\theta) \propto \prod_{i=1}^{L} \exp\left(-\frac{1}{2} n_i r_i (\theta_i - \bar{x}_i)^2\right) = \exp\left(-\frac{1}{2} \sum_{i=1}^{L} n_i r_i (\theta_i - \bar{x}_i)^2\right) =$$

$$= \exp\left(-\frac{1}{2} (\theta - \overline{\mathbf{x}})' \, \mathrm{NR} \, (\theta - \overline{\mathbf{x}})\right)$$

Dado que

$$p(\theta) \propto \exp\left(-\frac{1}{2} (\theta - \theta_0)' \, \mathrm{T} \, (\theta - \theta_0)\right)$$

la tesis del lema se alcanza sin más que seguir idéntica técnica de demostración que la utilizada por DeGroot (1970), pág. 176 (c. q. d.).

TEOREMA 11.—En el modelo normal,

$$\mathrm{I}^{\theta} \{\mathcal{E}^s(\mathbf{n}), \mathrm{N}^{\mathrm{L}}(\theta_0, \mathrm{H}_0)\} = \frac{1}{2} \log \frac{|\mathrm{H}_0\mathrm{N} + \mathrm{H}|}{|\mathrm{H}|}$$

DEMOSTRACIÓN.—Aplicando la definición de información esperada global sobre le cociente

$$\log \frac{\mathrm{N}^{\mathrm{L}}(\theta^*, (\mathrm{T} + \mathrm{NR})^{-1})}{\mathrm{N}^{\mathrm{L}}(\theta_0, \mathrm{H}_0)} = \log \sqrt{\frac{|\mathrm{T} + \mathrm{NR}|}{|\mathrm{H}_0|}} - \frac{1}{2} (\theta - \theta^*)' (\mathrm{T} + \mathrm{NR}) (\theta - \theta^*) +$$

$$+ \frac{1}{2} (\theta - \theta_0)' \, \mathrm{H}_0^{-1} \, (\theta - \theta_0)$$

resulta:

$$\mathrm{I}^{\theta} \{\mathrm{E}^s(\mathbf{n}), \mathrm{N}^{\mathrm{L}}(\theta_0, \mathrm{H}_0)\} = \frac{1}{2} \log \frac{|\mathrm{T} + \mathrm{NR}|}{|\mathrm{H}_0|} -$$

$$- \frac{1}{2} \int p(z) \int p(\underset{\sim}{\theta}/z) (\theta - \theta^*)' (\mathrm{T} + \mathrm{NR}) (\theta - \theta^*) \cdot d\theta \cdot dz +$$

$$+\frac{1}{2}\int p(\theta)\,(\theta-\theta_0)'\,H_0^{-1}\,(\theta-\theta_0)\cdot d\,\theta=$$

$$=\frac{1}{2}\log\frac{|\,T+NR\,|}{|\,H_0\,|}-\frac{L}{2}+\frac{L}{2}$$

El resultado se sigue de forma inmediata sin más que considerar que

$$\frac{|\,T+NR\,|}{|\,H_0\,|}=\frac{|\,H_0^{-1}+NH^{-1}\,|}{|\,H_0\,|}=\frac{|\,H+H_0N\,|}{|\,H\,|}$$

(c. q. d.).

Notemos que bajo la hipótesis de independencia de las componentes del vector de interés, la matriz de momentos H_0 es diagonal. Así pues, $H + H_0 N$ sería la matriz diagonal de elementos $\sigma_i^2 +$ $+ \sigma^0_{ii}\, n_i$. De ahí que la información esperada global en este caso de independencia tomaría el valor

$$I^{\theta}\ \{E^s(\mathbf{n}), N^L\ (\theta_0, H_0=\text{diagonal})\}=\frac{1}{2}\log\prod_{i=1}^{L}\frac{(\sigma_i^2+\sigma_{ii}^2\,n_i)}{\sigma_i^2}=$$

$$=\sum_{i=1}^{L}\frac{1}{2}\log\left(1+n_i\frac{\sigma^2_{ii}}{\sigma_i^2}\right)$$

esto es, una suma de informaciones esperadas que sobre cada θ_i proporciona el respectivo experimento E (n_i). Este resultado no es sino confirmación del Teorema 1.

Consideremos ahora un parámetro global φ, relacionado con el vector θ a través de una función lineal de sus componentes:

$$\varphi=J'\,\theta\qquad\text{con}\qquad J'=(h_1, h_2, \ldots, h_L)$$

Bajo las hipótesis del modelo normal resultarán las siguientes densidades inicial y posterior para φ:

$$p(\varphi)=N\ (J'\,\theta_0, J'\,H_0\,J)$$

$$p(\varphi/z)=N\,(J'\,(H_0^{-1}+NH^{-1})^{-1}\,(H_0^{-1}\,\theta_0+NH^{-1}\,\overline{\mathbf{x}}),\ \ J'\,(H_0^{-1}+NH^{-1})^{-1}\,J)$$

TEOREMA 12.—En el modelo normal, siendo $\varphi = J'\theta$, resulta ser la información esperada global útil igual a

$$I^{\varphi}\{E^{s}(\mathbf{n}), N^{L}(\theta_0, H_0)\} = \frac{1}{2}\log\frac{J' H_0 J}{J'(H_0^{-1} + NH^{-1})^{-1} J}$$

DEMOSTRACIÓN.—Aplicando la definición de información esperada global útil sobre el logaritmo

$$\log\frac{N(J'\theta^*, J'(H_0^{-1} + NH^{-1})^{-1} J)}{N(J'\theta_0, J' H_0 J)} = \log\sqrt{\frac{J' H_0 J}{J'(H_0^{-1} + NH^{-1})^{-1} J}} -$$

$$-\frac{1}{2}\int p(z)\int p(\varphi/z)(\varphi - J'\theta^*)'(J'(H_0^{-1} + NH^{-1})^{-1} J)^{-1}(\varphi - J'\theta^*)\cdot$$

$$\cdot\, d\varphi\cdot dz + \frac{1}{2}\int p(\varphi)(\varphi - J'\theta_0)'(J' H_0 J)^{-1}(\varphi - J'\theta_0)\cdot d\varphi =$$

$$= \frac{1}{2}\log\frac{J' H_0 J}{J'(H_0^{-1} + NH^{-1})^{-1} J} = \frac{1}{2} + \frac{1}{2}$$

(c. q. d.).

Notemos que bajo la hipótesis de independencia con H_0 diagonal

$$J' H_0 J = \sum_{i=1}^{L} h^2_i \sigma^0_{ii}$$

$$J'(H_0^{-1} + NH^{-1})^{-1} J = \sum_{i=1}^{L} h_i^2 \frac{\sigma_i^2 \sigma^0_{ii}}{\sigma_i^2 + n_i \sigma^0_{ii}}$$

de donde

$$I^{\varphi}\{E^{s}(\mathbf{n}), N^{L}(\theta_0, H_0 = \text{diagonal})\} = \frac{1}{2}\log\frac{\sum_{i=1}^{L} h_i^2 \sigma^0_{ii}}{\sum_{i=1}^{L} h_i^2 \frac{\sigma_i^2 \sigma^0_{ii}}{\sigma_i^2 + n_i \sigma^0_{ii}}}$$

Finalmente, como caso particular con $J' = (0, ..., 0, 1, 0, ..., 0)$ resulta:

COROLARIO.—La información esperada global útil sobre θ_i, componente i-ésima del vector de interés $\boldsymbol{\theta}$, es

$$I^{\theta_i}\{E^s(\mathbf{n}), N^L(\boldsymbol{\theta}_0, H_0)\} = \frac{1}{2} \log \frac{\sigma^0_{ii} \cdot | H_0^{-1} + N H^{-1} |}{| \operatorname{Adj} m_{ii} |}$$

DEMOSTRACIÓN.—Basta sustituir el valor particular de J en el Teorema anterior. Resulta ser:

$$J' H^0 J = \sigma^0_{ii} \quad \text{y} \quad J' (H_0^{-1} + N H^{-1})^{-1} J = \frac{| \operatorname{Adj} m_{ii} |}{| H_0^{-1} + N H^{-1} |}$$

con m_{ii} el elemento (i, i) de la matriz $H_0^{-1} + N H^{-1}$ (c. q. d.).

Nuevamente, bajo la hipótesis de independencia entre las θ_i, resulta:

$$J' (H_0^{-1} + N H^{-1})^{-1} J = \frac{\sigma_i^2 \sigma^0_{ii}}{\sigma_i^2 + n_i \sigma^0_{ii}}$$

de donde

$$I^{\theta_i}\{E^s(\mathbf{n}), N^L(\theta_0, H_0 = \text{diagonal})\} = \frac{1}{2} \log \left(1 + n_i \frac{\sigma^0_{ii}}{\sigma_i^2}\right) = I^{\theta_i}\{E(n), N(\theta_i, \sigma_i)\}$$

resultado que no es sino confirmación del contenido en el Teorema 10.

Referencias

BERNARDO, J. M. (1976). The use of information in the design and analysis of scientific experimentation. Ph. D. Tesis. Universidad de Londres.

BERNARDO, J. M. (1978). Una medida de la información útil proporcionada por un experimento. *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales de Madrid*, **72**, 419-440.

DEFINETTI, B. (1937). La prévision, ses lois logiques, ses sources subjectives. Trabajo aparecido en *Studies in Subjective Proba-*

bility (Kyburg & Smokler eds., 1964). New York: Mac Graw Hill.

DEGROOT, M. H. (1970). *Opimal Statistical Decisions*. New York: Mc Graw.

LINDLEY, D. V. (1956). On a Measure of the Information provided by an Experiment. *Ann. Math. Statist.*, **27**, 986-1005.

SHANNON, C. E. (1948). A Mathematical Theory of Communication. *Bell. System Tech. J.*, **27**, 379-423, 623-56.

VERES, E. (1981). Diseño Bayesiano de muestras, posiblemene estratificadas, cuando la utilidad terminal es función de la información conseguida. Tesis. Universidad de Valencia.