Comportamiento asintótico de la información proporcionada por un experimento asociado a un diseño de estratificación

Por Ernesto Veres Ferrer

Recibido: 5 de mayo de 1982.

Presentado por el académico numerario Francisco Azorín Poch.

Resumen

La generalización del concepto de información esperada a un experimento $E^s(\mathbf{n})$ que implica la existencia de estratificación en el espacio muestral (Veres, 1983*a*) introduce una mayor dificultad en el cálculo exacto de la misma. Si esta dificultad es ya notoria cuando la cantidad de interés es una variable aleatoria unidimensional y la muestra resultado de cierto experimento se extrae de un único espacio muestral, es evidente entonces la dificultad adicional que supone manejar un vector paramétrico como cantidad de interés y una muestra que es un conjunto de submuestras extraídas de sendas poblaciones distintas.

Precisamente, en este trabajo va a estudiarse el comportamiento en muestras grandes de la información que proporciona el resultado de cierto experimento asociado a un diseño de estratificación, así como el de su valor esperado o información esperada global. El objeto perseguido es el de obtener expresiones relativamente sencillas que describan a esta información esperada global con más o menos grado de aproximitividad, a fin de conocer al menos su comportamiento asintótico, lo que nos permitirá diseñar adecuadamente al experimento $E^s(\mathbf{n})$. Finalmente, la fórmula de aproximación obtenida se aplica al estudio del modelo binomial generalizado.

1. INTRODUCCION Y DEFINICIONES

Supongamos primeramente que el objetivo pretendido por un investigador al diseñar cierta investigación es el de mejorar su conocimiento sobre el valor de cierta cantidad o parámetro de interés sobre el que se desea, en último término, efectuar inferencia. Para ello diseñará un experimento de forma que la distribución de sus resultados dependerá de la cantidad de interés de una manera conocida, proporcionando así cierta cantidad de información.

A fin de describir un experimento asociado a un diseño de estratificación, el modelo matemático básico a considerar contiene un espacio muestral X en el que existe una partición en L subconjuntos o estratos X_i ($i = 1, 2, \dots, L$). Cada uno de estos subespacios está dotado de una apropiada σ-álgebra sobre las que se definen 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 $E_i = \{X_i, \Theta_i, p(x_i/\vartheta_i)\}$ que consiste en una observación de la variable aleatoria $x_i \in X_i$ que se distribuye, para algún $\vartheta_i \in \Theta_i$ dado, de acuerdo con la densidad $p(x_i/\vartheta_i)$. En estas condiciones definiremos el experimento compuesto $E^s(\mathbf{n})$ como aqeúl que está formado por los L experimentos componentes independientes $E(n_i)$:

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

y donde $E(n_i)$ representa al experimento, que consiste en n_i repeticiones independientes (dado ϑ_i) del experimento E_i. Con $\mathbf{n}' = (n_1, n_2, \ldots, n_L)$ denotamos al vector cuyas componentes son los respectivos tamaños de las muestras extraídas en sus estratos respectivos y que llamaremos vector tamaño muestral.

La densidad de probabilidad que describe el comportamiento de la muestra global z resultado de $E^s(\mathbf{n})$ —muestra formada por el conjunto de submuestras parciales $z = (z_1, z_2, \ldots, z_L)$ extraídas en los estratos y que resultan, a su vez, de sendos experimentos componentes $E(n_i)$— viene dada a través del doble productorio

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

en el que con $\boldsymbol{\vartheta}$ denotamos al vector cuyas componentes son las respectivas cantidades de interés ϑ_i, esto es, $\boldsymbol{\vartheta}' = (\vartheta_1, \vartheta_2, \ldots, \vartheta_L)$. Dicho vector recibirá el nombre de vector cantidad de interés o paramétrico y tomará valores en el espacio paramétrico L-dimensional

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

A fin de terminar de describir los elementos del problema, el argumento Bayesiano extiende este modelo básico al suponer que $\boldsymbol{\Theta}$ es soporte de una apropiada 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 $p(\boldsymbol{\vartheta})$ (DeFinetti, 1937). Dicha densidad pone de manifiesto la posible relación de dependencia y correlación entre las distintas componentes del vector paramétrico.

Una vez obtenido el resultado z tras la realización de $E^s(\mathbf{n})$, los conocimientos del investigador sobre el valor de $\boldsymbol{\vartheta}$ se describirán por la densidad posterior $p(\boldsymbol{\vartheta}/z)$ obtenida mediante el Teorema de Bayes:

$$p(\boldsymbol{\vartheta}/z) = p(z/\boldsymbol{\vartheta}) \cdot p(\boldsymbol{\vartheta})/p(z)$$

donde

$$p(z) = \int p(z/\boldsymbol{\vartheta}) \cdot p(\boldsymbol{\vartheta}) \cdot d\boldsymbol{\vartheta}$$

es la densidad predictiva del resultado z. Supondremos en lo que sigue, y salvo indicación en contra, que cuantas integrales aparecen en este trabajo existen y están extendidas a todos sus respectivos dominios de definición.

Definición. La información proporcionada por el resultado z del experimento $E^s(\mathbf{n})$ sobre el valor del vector de interés ϑ y cuando la densidad inicial es $p(\vartheta)$, viene dada por

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

Notemos que $I^{\vartheta}\{z, p(\vartheta)\}$ depende de los resultados experimentales únicamente a través de la densidad $p(z/\vartheta)$. Además, considerando el logaritmo del cociente, la anterior definición proporciona una medida entre las densidades de probabilidad atribuidas al verdadero valor del parámetro antes y después de observar los resultados del experimento $E^s(\mathbf{n})$. Por otra parte, se trata de una expresión invariante ante transformaciones $\psi = \psi(\vartheta)$ biyectivas del parámetro ϑ. En efecto,

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

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

y en donde $p(\psi)$ y $p(\psi/z)$ son las correspondientes densidades inicial y posterior deducidas de las correspondientes $p(\vartheta)$ y $p(\vartheta/z)$ por la transformación biyectiva.

El valor esperado sobre todas los posibles resultados del experimento $E^s(\mathbf{n})$ proporciona la siguiente

Definición. La información esperada global sobre ϑ proporcionada por el experimento $E^s(\mathbf{n})$, cuando la densidad inicial es $p(\vartheta)$, viene dada por la doble integral

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

$$= \iint p(z) \cdot p(\vartheta/z) \cdot \log \frac{p(\vartheta/z)}{p(\vartheta)} \cdot d\vartheta \cdot dz$$

Para un experimento $E(n)$ que proporciona una muestra aleatoria simple, la expresión anterior aparece ya en Shannon (1948) y sus propiedades fueron estudiadas por Lindley (1956) y generalizadas por Bernardo (1978) al caso de que existan parámetros marginales. El estudio de las propiedades satisfechas por $I^{\vartheta}\{E^s(\mathbf{n}), p(\vartheta)\}$ puede encontrarse en Veres (1983*a*).

Excepto en algunos casos especiales el cálculo exacto de $I^{\vartheta}\{z, p(\vartheta)\}$ e $I^{\vartheta}\{E^s(\mathbf{n}), p(\vartheta)\}$ es difícil. Si este hecho es ya conocido para situaciones planteadas por experimentos simples, resulta evidente el incremento de complejidad introducido por la existencia de un vector como cantidad de interés y de una muestra que es un conjunto de ellas extraídas de sendas poblaciones distintas. En línea a simplificar esa complejidad puede intentarse reducir la situación expresada por el experimento $E^s(\mathbf{n})$ a otra planteada por un experimento simple equivalente (Veres, 1983*b*). Sin embargo, es esencial conocer al menos su comportamiento asintótico para, por ejemplo, diseñar óptimamente el experimento. Precisamente es objetivo de este trabajo estudiar tal comportamiento asintótico para aplicarlo, finalmente, al modelo binomial generalizado para su correspondiente ejemplarización. Se trata, pues, de la generalización al experimento $E^s(\mathbf{n})$ del resultado encontrado por Bernardo (1979) para experimentos simples.

2. COMPORTAMIENTO ASINTOTICO DE $I^{\vartheta}\{E^s(\mathbf{n}), p(\vartheta)\}$

Ya hemos indicado que la información proporcionada sobre $\boldsymbol{\vartheta}$ es la misma que la proporcionada sobre cualquier cantidad transformada por una biyección $\psi = \psi(\boldsymbol{\vartheta})$. En consecuencia, puede simplificarse el cálculo de I^{ϑ} considerando una transformación $\boldsymbol{\psi} = \psi(\boldsymbol{\vartheta})$ cuya densidad inicial $p(\boldsymbol{\psi}) = \{p(\boldsymbol{\vartheta})\}_{\vartheta=\psi} \times \times\ |\partial\boldsymbol{\vartheta}/\partial\boldsymbol{\psi}|$ sea aproximadamente multinormal y calculando I^{ψ} en su lugar. Evidentemente que existe siempre una transformación de $\boldsymbol{\vartheta}$ cuya distribución es exactamente normal pero, en general, tal transformación exacta es complicada. Puesto que lo que se pretende es proporcionar un resultado manejable generalmente resultará más práctico utilizar una transformación sencilla aunque su densidad $p(\boldsymbol{\psi})$ sea tan sólo aproximadamente multinormal.

Siendo $\boldsymbol{\psi} = \psi(\boldsymbol{\vartheta})$ biyectiva, puede escribirse que $\boldsymbol{\vartheta} = F(\boldsymbol{\psi})$ donde $F^{-1} = \psi$ es la correspondiente transformación inversa. Particularizando para cada componente escribiremos que $\vartheta_i = F_i(\boldsymbol{\psi})$, de forma que resultará:

$$\boldsymbol{\vartheta} = (F_1(\boldsymbol{\psi}), F_2(\boldsymbol{\psi}), \ldots, F_L(\boldsymbol{\psi}))$$

pudiendo escribirse que

$$p(z/\boldsymbol{\vartheta}) = p(z/F_1(\boldsymbol{\psi}), F_2(\boldsymbol{\psi}), \ldots, F_L(\boldsymbol{\psi})) = \prod_{i=1}^{L} p(z_i/F_i(\boldsymbol{\psi}))$$

Teorema 1

Siendo $p(\boldsymbol{\psi})$ aproximadamente multinormal de vector media $\boldsymbol{\psi}_o$ y matriz de precisión T_o y para valores grandes en todas sus componentes del vector tamaño muestral $\mathbf{n}$, se verifica que $p(\boldsymbol{\psi}/z)$ es también aproximadamente normal multivariante con vector media $\boldsymbol{\psi}^*$ y matriz de precisión $T_o + H(z)$, donde

$$\boldsymbol{\psi}^* = (T_o + H(z))^{-1}\cdot(T_o\boldsymbol{\psi}_o + H(z)\cdot\hat{\boldsymbol{\psi}})$$

$$H(z) = \sum_{i=1}^{L} H^i(z_i)$$

$H^i(z_i)$ es la matriz de información de elementos iguales a

$$-\frac{\partial^2}{\partial\psi_p\partial\psi_q}\log p(z_i/F_i(\psi))\Big|_{\hat{\psi}^i} \quad ; \quad p, q = 1, 2, \ldots, L$$

$$\hat{\psi} = H(z)^{-1}\sum_{i=1}^{L} H^i(z_i)\cdot\hat{\psi}_i$$

y $\hat{\psi}^i$ es el estimador de máxima verosimilitud de ψ calculado a partir de la densidad de $p(z_i/F_i(\psi))$.

Demostración. Por Bayes, y al ser $p(\psi)$ aproximadamente normal:

$$p(\psi/z) \propto p(\psi)\cdot p(z/\psi) \simeq \frac{|T_o|^{\frac{1}{2}}}{(\sqrt{2\pi})^L}\cdot e^{-\frac{1}{2}(\psi - \psi_o)'T_o(\psi - \psi_o)}\cdot p(z/\psi) \propto$$

$$\propto p(z/\psi)\cdot\exp\left(-\tfrac{1}{2}(\psi - \psi_o)'T_o(\psi - \psi_o)\right) =$$

$$= \exp\left(\log p(z/\psi) - \tfrac{1}{2}(\psi - \psi_o)'T_o(\psi - \psi_o)\right)$$

Consideremos el factor $\exp(\log p(z/\psi))$. Resulta ser

$$\log p(z/\psi) = \sum_{i=1}^{L}\log p(z_i/F_i(\psi))$$

esto es, una suma de logaritmos de densidades de probabilidad con ámbito de aplicación en un único estrato. Para cada uno de esos sumandos es conocido (Lindley, 1965) que, siendo $\hat{\psi}^i$ el estimador de máxima verosimilitud de ψ calculado a partir de su respectiva densidad de probabilidad $p(z_i/F_i(\psi))$, se cumple en un entorno de dicha estimación que

$$\log p(z_i/F_i(\psi)) = \log p(z_i/\hat{\psi}^i) - \tfrac{1}{2}(\psi - \hat{\psi}^i)'H^i(z_i)(\psi - \hat{\psi}^i) + R_i$$

donde R_i es, en muestras grandes, despreciable frente al segundo término de la igualdad anterior (es del orden de $0(n_i^{-1/2})$) y

$$\hat{\psi}^{i\prime} = (\hat{\psi}^i_1, \hat{\psi}^i_2, \ldots, \hat{\psi}^i_L) \quad ; \quad \lim_{n_i\to\infty} R_i = 0$$

y $H^i(z_i)$ es la matriz simétrica de información, de elementos $h^i_{jk}(z_i)$ definidos como:

$$h^i_{jk}(z_i) = -\frac{\partial^2}{\partial\psi_j\partial\psi_k}\log p(z_i/F_i(\psi))\Big|_{\hat{\psi}^i}$$

verificándose además por la ley fuerte de los grandes números que

$$\lim_{n_i \to \infty} \frac{1}{n_i} \log p(z_i/F_i(\psi)) = \int p(x_{ij}/F_i(\psi)) \cdot \log p(x_{ij}/F_i(\psi)) \cdot dx_{ij}$$

$$\lim_{n_i \to \infty} \frac{1}{n_i} \frac{\partial^r}{\partial \psi_1^{m1} \ldots \partial \psi_L^{m_L}} \log p(z_i/F_i(\psi)) =$$

$$= \int p(x_{ij}/F_i(\psi)) \cdot \frac{\partial^r}{\partial \psi_1^{m_1} \ldots \partial \psi_L^{m_L}} \log p(x_{ij}/F_i(\psi)) \cdot dx_{ij}$$

con $r = \sum_{i=1}^{L} m_i$.

De ahí que

$$\sum_{i=1}^{L} \log p(z_i/F_i(\psi)) = \sum_{i=1}^{L} \log p(z_i/\hat{\psi}^i) - \tfrac{1}{2} \sum_{i=1}^{L} (\psi - \hat{\psi}^i)' H^i(z_i)(\psi - \hat{\psi}^i) +$$

$$+ \sum_{i=1}^{L} R_i = \sum_{i=1}^{L} \log p(z_i/\hat{\psi}^i) - \tfrac{1}{2}(\psi - \hat{\psi})' H(z)(\psi - \hat{\psi}) + Q + \sum_{i=1}^{L} R_i$$

donde Q es una suma de términos en los que no aparece ψ y

$$H(z) = \sum_{i=1}^{L} H^i(z_i) \quad ; \quad \hat{\psi} = H(z)^{-1} \cdot \sum_{i=1}^{L} H^i(z_i) \cdot \hat{\psi}^i$$

Así pues:

$$p(z/\psi) = e^{\log p(z/\psi)} \propto \prod_{i=1}^{L} p(z_i/\hat{\psi}^i) \cdot e^{-\frac{1}{2}(\psi - \hat{\psi})' H(z)(\psi - \hat{\psi}) + \sum_{i=1}^{L} R_i} \cdot e^{Q} \propto$$

$$\propto e^{-\frac{1}{2}(\psi - \hat{\psi})' H(z)(\psi - \hat{\psi})}$$

puesto que ni Q ni el productorio $\prod_{i=1}^{L} p(z_i/\hat{\psi}^i)$ dependen de ψ.

Finalmente, sustituyendo el resultado anterior en el producto de densidades $p(\psi) \cdot p(z/\psi)$ resulta:

$$p(\psi/z) \propto e^{-\frac{1}{2}(\psi - \hat{\psi})' H(z)(\psi - \hat{\psi})} \cdot e^{-\frac{1}{2}(\psi - \psi_o)' T_o(\psi - \psi_o)} \cdot e^{\sum_{i=1}^{L} R_i} \propto$$

$$\propto \exp\left\{ -\tfrac{1}{2}(\psi - (H(z) + T_o)^{-1} \cdot (H(z)\hat{\psi} + T_o\psi_o))'(T_o + H(z)) \cdot \right.$$

$$\left. \cdot (\psi - (H(z) + T_o)^{-1}(H(z)\hat{\psi} + T_o\psi_0)) + \sum_{i=1}^{L} R_i \right\} =$$

$$= \exp\left\{ -\tfrac{1}{2}(\psi - \psi^*)'(T_o + H(z))(\psi - \psi^*) + \sum_{i=1}^{L} R_i \right\}$$

y donde

$$\sum_{i=1}^{L} R_i = \sum_{i=1}^{L} O(n_i^{-1/2}) = O(\mathbf{n}^{-\frac{1}{2}})$$

Comparando la última expresión con la definición de una distribución multinormal se sigue el resultado apetecido.

(c.q.d.)

Utilizando este resultado vamos a estimar la integral que define a $I^{\vartheta}\{z, p(\vartheta)\}$ y obtener, de esta forma, una aproximación válida para muestras grandes en todos los estratos.

Teorema 2

Si $\psi = \psi(\vartheta)$ es una transformación biyectiva del vector ϑ con densidad aproximadamente normal y con matriz de precisión T_o, entonces

$$I^{\vartheta}\{z, p(\vartheta)\} = \tfrac{1}{2}\log\frac{|T_o + H(z)|}{|T_o|} - \tfrac{1}{2}\sum_{p=1}^{L}\sum_{q=1}^{L} h_{pq}(z)\cdot\frac{|Adj(t_{qp} + h_{qp}(z))|}{|T_o + H(z)|} +$$

$$+ \tfrac{1}{2}(\hat{\psi} - \psi_o)'H(z)(T_o + H(z))^{-1}T_o(T_o + H(z))^{-1}H(z)(\hat{\psi} - \psi_o) + \varepsilon$$

donde ε es una cantidad despreciable, en muestras grandes, frente a $I^{\vartheta}\{z, p(\vartheta)\}$.

Demostración. Debido a la invarianza ante transformaciones biyectivas del vector de interés puede escribirse:

$$I^{\vartheta}\{z, p(\vartheta)\} = I^{\psi}\{z, p(\psi)\} = \int N^L(\psi^*, T_o + H(z))\cdot$$

$$\log\frac{N^L(\psi^*, T_o + H(z))}{N^L(\psi_o, T_o)}\cdot d\psi + \varepsilon$$

en donde las expresiones de las multinormales L-dimensionales vienen dadas a través de las respectivas matrices de precisión. La integral anterior es igual a la suma

$$-\tfrac{1}{2}\int N^L(\psi^*, T_o + H(z))\cdot(\psi - \psi^*)'(T_o + H(z))(\psi - \psi^*)\cdot d\psi +$$

$$+ \tfrac{1}{2}\log\frac{|T_o + H(z)|}{|T_o|} + \tfrac{1}{2}\int N^L(\psi^*, T_o + H(z))\cdot(\psi - \psi_o)'T_o(\psi - \psi_o)\cdot d\psi =$$

$$= \tfrac{1}{2}\log\frac{|T_o + H(z)|}{|T_o|} - \tfrac{1}{2}\sum_{p=1}^{L}\sum_{q=1}^{L}(t_{pq} + h_{pq}(z))\cdot\frac{|Adj(t_{qp} + h_{qp}(z))|}{|T_o + H(z)|} +$$

$$+ \tfrac{1}{2}\int N^L(\psi^*, T_o + H(z))\cdot(\psi - \psi^*)'T_o(\psi - \psi^*)\cdot d\psi +$$

$$+ \tfrac{1}{2}\int N^L(\psi^*, T_o + H(z))\cdot(\psi - \psi^*)'T_o(\psi^* - \psi_o)\cdot d\psi +$$

$$+ \tfrac{1}{2}\int N^L(\psi^*, T_o + H(z))\cdot(\psi^* - \psi_o)'T_o(\psi - \psi^*)\cdot d\psi +$$

$$+ \tfrac{1}{2}\int N^L(\psi^*, T_o + H(z))\cdot(\psi^* - \psi_o)'T_o(\psi^* - \psi_o)\cdot d\psi =$$

$$= \tfrac{1}{2}\log\frac{|T_o + H(z)|}{|T_o|} - \tfrac{1}{2}\sum_{p=1}^{L}\sum_{q=1}^{L} h_{pq}(z)\cdot\frac{|Adj\,(t_{qp} + h_{qp}(z))|}{|T_o + H(z)|} +$$

$$+ \tfrac{1}{2}(\psi^* - \psi_o)'T_o(\psi^* - \psi_o)$$

El resultado se sigue sin más que considerar que

$$\psi^* - \psi_o = (T_o + H(z))^{-1}H(z)(\hat{\psi} - \psi_o)$$

de forma que el último término en la cadena de igualdades anterior resulta:

$$\tfrac{1}{2}(\hat{\psi} - \psi_o)'H(z)(T_o + H(z))^{-1}T_o(T_o + H(z))^{-1}H(z)(\hat{\psi} - \psi_o)$$

(c.q.d.)

A partir de la anterior aproximación puede estudiarse el comportamiento asintótico de la información esperada global. Para ello veamos previamente el siguiente

Lema

Si el vector tamaño muestral **n** es suficientemente grande en todas sus componentes, dado ψ, la distribución de $\hat{\psi}$ es aproximadamente normal multivariante con vector media igual a ψ y matriz de precisión

$$I(\psi) = \sum_{i=1}^{L} n_i\cdot I^i(\psi)$$

donde $I^i(\psi)$ es la matriz de elementos $c^i_{pq}(\psi)$ definidos como

$$c^i_{pq}(\psi) = -\int p(x_{ij}/F_i(\psi))\frac{\partial^2}{\partial\psi_p\partial\psi_q}\log p(x_{ij}/F_i(\psi))\cdot dx_{ij}$$

Demostración. Dado ψ, las distintas estimaciones máximo-verosímiles $\hat{\psi}^i$ son independientes entre sí, toda vez que así lo son las submuestras parciales sobre las que están definidas. Por otra parte, es conocido (Bernardo, 1979), que, según la ley fuerte de los grandes números y para valores grandes de n_i, la constante $h^i_{pq}(z_i)$ se comporta, para todo p, $q = 1, 2, \ldots, L$ y dado ψ, como n_i veces la constante

$$c^i_{pq}(\psi) = -\int p(x_{ij}/F_i(\psi))\cdot\frac{\partial^2}{\partial\psi_p\partial\psi_q}\log p(x_{ij}/F_i(\psi))\cdot dx_{ij}$$

Sea la matriz $I^i(\psi)$ definida con los elementos $c^i_{pq}(\psi)$ $(p, q = 1, 2, \ldots, L)$. Del comentario anterior deducimos que si n_i es grande, la matriz $H^i(z_i)$ se comportará, dado ψ, como n_i veces la matriz $I^i(\psi)$. Así pues, siendo el vector $\mathbf{n}$ lo suficientemente grande en todas sus componentes resultará que

$$H(z) = \sum_{i=1}^{L} H^i(z_i) \simeq \sum_{i=1}^{L} n_i \cdot I^i(\psi) = I(\psi)$$

Podrá, pues, escribirse que

$$\hat{\psi} = H(z)^{-1} \sum_{i=1}^{L} h^i(z_i) \cdot \hat{\psi}^i = \sum_{i=1}^{L} H(z)^{-1} \cdot H^i(z_i) \cdot \hat{\psi}^i$$

Además y bajo ciertas débiles condiciones de regularidad es conocido (Cox y Hinkley, 1974) que dado ψ la distribución del estimador máximo-verosímil $\hat{\psi}^i$ es, en muestras grandes, aproximadamente multinormal, con vector media ψ y matriz de precisión

$$n_i \cdot I^i(\psi)$$

Así pues, cuando n_i es grande:

$$p(\hat{\psi}^i/\psi) = N^L(\psi,\ n_i \cdot I^i(\psi) \text{ como matriz de precisión})$$

De ahí que siendo $\mathbf{n}$ lo suficientemente grande en todas sus componentes:

$$H(z)^{-1} \cdot H^i(z_1) \cdot \hat{\psi}^i \simeq I(\psi)^{-1} \cdot n_i \cdot I^i(\psi) \cdot \hat{\psi}^i$$

Y así resulta que, dado ψ, la densidad $p(I(\psi)^{-1} n_i I^i(\psi) \psi^i/\psi)$ es aproximadamente nomal multivariante:

$$N^L(I(\psi)^{-1} n_i I^i(\psi)\psi,\ \text{matriz de momentos} = I(\psi)^{-1} n_i I^i(\psi) I(\psi)^{-1})$$

Como $\hat{\psi}$ es, dado ψ, una suma de multinormales independientes resulta (Graybill, 1961) finalmente que siendo $\mathbf{n}$ lo suficientemente grande como para asegurar que lo sean todas sus componentes

$$p(\hat{\psi}/\psi) \simeq N^L\left(\sum_{i=1}^{L} I(\psi)^{-1} n_i I^i(\psi)\psi,\ \sum_{i=1}^{L} I(\psi)^{-1} n_i I^i(\psi) I(\psi)^{-1} \right) =$$

$$= N^L(\psi,\ I(\psi)^{-1} = \text{matriz de momentos})$$

(c.q.d.)

Estamos ya en condiciones de poder encontrar una sencilla expresión que se aproxime asintóticamente a $I^{\vartheta}\{E^s(\mathbf{n}), p(\vartheta)\}$.

Teorema 3

Siendo $\psi = \psi(\vartheta)$ una transformación biyectiva de la cantidad de interés ϑ con densidad aproximadamente multinormal de matriz de precisión T_o, entonces

$$I^{\vartheta}\{E^s(\mathbf{n}), p(\vartheta)\} = \tfrac{1}{2} \log \frac{|T_o + R|}{|T_o|} + \varepsilon$$

donde R es la matriz esperanza de la $I(\psi)$ definida en el lema anterior:

$$R = \int p(\psi) \cdot I(\psi) \cdot d\psi$$

y ε es una cantidad que, en muestras grandes, es despreciable frente al logaritmo del cociente $|T + R|/|T_o|$.

Demostración. Por el lema anterior, siendo $\mathbf{n}$ suficientemente grande y dado ψ, $H(z)$ se comporta como la matriz

$$I(\psi) = \sum_{i=1}^{L} n_i I^i(\psi)$$

De ahí que siendo $\mathbf{n}$ grande, el valor esperado de cualquier función de $H(z)$ sea, aproximadamente, igual al valor que toma la función en el valor esperado de $H(z)$, esto es, en

$$\int p(z) \cdot H(z) \cdot dz = \int p(\psi) \cdot I(\psi) \cdot d\psi = R$$

verificándose las igualdades

$$\begin{aligned} \int p(z) \cdot H(z) \cdot dz &= \int \sum_{i=1}^{L} H^i(z_i) \cdot p(z) \cdot dz = \sum_{i=1}^{L} \int H^i(z_i) \cdot p(z_i) \cdot dz_i = \\ &= \sum_{i=1}^{L} n_i \int p(\psi) \cdot I^i(\psi) \cdot d\psi = \int \left\{ \sum_{i=1}^{L} n_i I^i(\psi) \right\} p(\psi) \cdot d\psi = \\ &= \int I(\psi) \cdot p(\psi) \cdot d\psi \end{aligned}$$

De ahí que

$$\begin{aligned} I^{\vartheta}\{E^s(\mathbf{n}), p(\vartheta)\} &= \int p(z) \cdot I^{\vartheta}\{z, p(\vartheta)\} \, dz = \\ &= \int p(z) \cdot I^{\psi}\{z, p(\psi)\} \cdot dz = I^{\psi}\{E^s(\mathbf{n}), p(\psi)\} \end{aligned}$$

Por el teorema 2:

$$\begin{aligned} I^{\vartheta}\{E^s(\mathbf{n}), p(\vartheta)\} \simeq {} & \tfrac{1}{2} \log \frac{|T_o + R|}{|T_o|} - \tfrac{1}{2} \sum_{p=1}^{L} \sum_{q=1}^{L} r_{pq} \cdot \frac{|Adj\,(t_{qp} + r_{qp})|}{|T_o + R|} + \\ & + \tfrac{1}{2} \int p(z) \cdot (\hat{\psi} - \psi_o)' R (T_o + R)^{-1} T_o (T_o + R)^{-1} R (\hat{\psi} - \psi_o) \cdot dz \end{aligned}$$

donde r_{pq} es el elemento (p, q) de la matriz R, es es,

$$r_{pq} = \sum_{i=1}^{L} n_i \int c_{pq}^i(\psi)\cdot p(\psi)\cdot d\psi$$

Consideremos la penúltima integral:

$$\int p(z)\cdot(\hat{\psi} - \psi_o)'R(T_o + R)^{-1}T_o(T_o + R)^{-1}R(\hat{\psi} - \psi_o)\cdot dz =$$

$$= \int p(\hat{\psi})\cdot(\hat{\psi} - \psi_o)'R(T_o + R)^{-1}T_o(T_o + R)^{-1}R(\hat{\psi} - \psi_o)\cdot d\hat{\psi} =$$

$$= \int p(\psi)\int p(\hat{\psi}/\psi)\cdot(\hat{\psi} - \psi_o)'R(T_o + R)^{-1}T_o(T_o + R)^{-1}R(\hat{\psi} - \psi_o)\cdot d\hat{\psi}\cdot d\psi =$$

$$= \int p(\psi)\int p(\hat{\psi}/\psi)\cdot(\hat{\psi} - \psi)'R(T_o + R)^{-1}T_o(T_o + R)^{-1}R(\hat{\psi} - \psi)\cdot d\hat{\psi}\cdot d\psi \quad (1) +$$

$$+ \int p(\psi)\int p(\hat{\psi}/\psi)\cdot(\hat{\psi} - \psi)'R(T_o + R)^{-1}T_o(T_o + R)^{-1}R(\psi - \psi_o)\cdot d\hat{\psi}\cdot d\psi +$$

$$+ \int p(\psi)\int p(\hat{\psi}/\psi)\cdot(\psi - \psi_o)'R(T_o + R)^{-1}T_o(T_o + R)^{-1}R(\hat{\psi} - \psi)\cdot d\hat{\psi}\cdot d\psi +$$

$$+ \int p(\psi)\int p(\hat{\psi}/\psi)\cdot(\psi - \psi_o)'R(T_o + R)^{-1}T_o(T_o + R)^{-1}R(\psi - \psi_o)\cdot d\hat{\psi}\cdot d\psi \; (2) =$$

$$= (1) + (2)$$

Por el lema es conocido que $p(\hat{\psi}/\psi) \simeq N^L(\psi$, matriz de precisión $I(\psi))$. Así pues, dado ψ, la nueva variable $\mathbf{Y} = (T_o + R)^{-1}R(\hat{\psi} - \psi)$ se distribuirá según la multinormal

$p(\mathbf{Y}/\psi) \simeq N^L(\mathbf{O}$, matriz de momentos $(T_o + R)^{-1}R\cdot I(\psi)^{-1}\cdot R(T_o + R)^{-1})$

De ahí que

$$(1) = \int p(\psi)\int p(\mathbf{Y}/\psi)\cdot \mathbf{Y}'T_o\mathbf{Y}\cdot d\mathbf{Y}\cdot d\psi =$$

$$= \int p(\psi) \sum_{p=1}^{L}\sum_{q=1}^{L} t_{pq}a_{pq}(\psi)\cdot d\psi = \sum_{p=1}^{L}\sum_{q=1}^{L} t_{pq}b_{pq}$$

donde a_{pq} es el elemento (p, q) de la matriz $(T_o + R)^{-1}R\cdot I(\psi)^{-1}R(T_o + R)^{-1}$; t_{pq} el correspondiente de la matriz T_o y b_{pq}, el elemento (p, q) de la matriz $(T_o + R)^{-1}R(T_o + R)^{-1}$, toda vez que

$$\int p(\psi)\cdot I(\psi)^{-1}\cdot d\psi = R^{-1}$$

Sea la nueva variable $\mathbf{X} = (T_o + R)^{-1}R(\psi - \psi_o)$. Dado que la distribución que describe las opiniones iniciales que el investigador posee sobre ψ es, por hipótesis, aproximadamente multinormal, resulta:

$p(\mathbf{X}) \simeq N^L(\mathbf{O}$, matriz de momentos $(T_o + R)^{-1}RT_o^{-1}R(T_o + R)^{-1})$

De ahí que

$$(2) = \int p(\psi)\cdot(\psi - \psi_o)'R(T_o + R)^{-1}T_o(T_o + R)^{-1}R(\psi - \psi_o)\cdot d\psi =$$

$$= \int p(\mathbf{X})\cdot\mathbf{X}'T_o\mathbf{X}\cdot d\mathbf{X} = \sum_{p=1}^{L}\sum_{q=1}^{L} t_{pq}d_{pq}$$

donde d_{pq} es el elemento (p, q) de la matriz $(T_o + R)^{-1}RT_oR(T_o + R)^{-1}$. Siendo las matrices T_o y R simétricas, reuniendo todos los resultados parciales obtenidos:

$$I^{\vartheta}\{E^s(\mathbf{n}), p(\vartheta)\} \simeq \tfrac{1}{2}\log\frac{|T_o + R|}{|T_o|} - \tfrac{1}{2}\sum_{p=1}^{L}\sum_{q=1}^{L} r_{pq}\cdot\frac{|Adj\,(t_{pq} + r_{pq})|}{|T_o + R|} +$$

$$+ \sum_{p=1}^{L}\sum_{q=1}^{L} t_{pq}(b_{pq} + d_{pq})$$

Y al ser

$$\sum_{p=1}^{L}\sum_{q=1}^{L} r_{pq}\cdot\frac{|Adj\,(t_{pq} + r_{pq})|}{|T_o + R|} = \frac{L}{2} - \sum_{p=1}^{L}\sum_{q=1}^{L} t_{pq}\cdot\frac{|Adj\,(t_{pq} + r_{pq})|}{|T_o + R|}$$

y denotando por g_{pq} al elemento (p, q) de la matriz $(T_o + R)^{-1}$:

$$I^{\vartheta}\{E^s(\mathbf{n}), p(\vartheta)\} \simeq \tfrac{1}{2}\log\frac{|T_o + R|}{|T_o|} - \frac{L}{2} + \tfrac{1}{2}\sum_{p=1}^{L}\sum_{q=1}^{L} t_{pq}(b_{pq} + d_{pq} + g_{pq})$$

donde $b_{pq} + d_{pq} + g_{pq}$ es el elemento (p, q) de la matriz

$$(T_o + R)^{-1}R(T_o + R)^{-1} + (T_o + R)^{-1}RT_o^{-1}R(T_o + R)^{-1} + (T_o + R)^{-1} =$$

$$= (T_o + R)^{-1}\{R + RT_o^{-1}R + T_o + R\}(T_o + R)^{-1} =$$

$$= (T_o + R)^{-1}\{(T_o + R)T_o^{-1}(T_o + R)\}(T_o + R)^{-1} = T_o^{-1}$$

Por tanto

$$b_{pq} + d_{pq} + g_{pq} = \frac{|Adj\,(t_{qp})|}{|T_o|} = \frac{|Adj\,(t_{pq})|}{|T_o|}$$

Finalmente, pues,

$$I^{\vartheta}\{E^s(\mathbf{n}), p(\vartheta)\} \simeq \tfrac{1}{2}\log\frac{|T_o + R|}{|T_o|} - \frac{L}{2} + \tfrac{1}{2}\sum_{p=1}^{L}\sum_{q=1}^{L} t_{pq}\cdot\frac{|Adj\,(t_{pq})|}{|T_o|} =$$

$$= \tfrac{1}{2}\log\frac{|T_o + R|}{|T_o|}$$

(c.q.d.)

Tal como era de esperar la coherencia del anterior resultado queda comprobada al aplicar la aproximación anterior al modelo normal y ver la coincidencia de los resultados obtenidos. Supondremos, pues, el modelo definido por las dos siguientes hipótesis básicas:

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

2.ª $p(\boldsymbol{\vartheta})$ es multinormal, $N^L(\boldsymbol{\vartheta}_o, H_o)$.

Así pues, siendo $p(\boldsymbol{\vartheta})$ ya multinormal, consideraremos la transformación identidad $\psi_i = \vartheta_i$, $i = 1, 2, \ldots, L$. Para ella,

$$c^i_{pq}(\psi) = c^i_{pq}(\vartheta_i) = 0 \text{ para todo } p, q \text{ tales que } p \neq i \text{ ó } q \neq i \quad ;$$

$$c^i_{ii}(\vartheta_i) = 1/\sigma^2_i$$

De ahí que las matrices $I(\psi)$ e $I(\boldsymbol{\vartheta})$ coincidan y sean iguales a una matriz diagonal de elementos n_i/σ^2_i. Esto es, ambas matrices son iguales al producto NH^{-1}, donde H es otra matriz diagonal de elementos σ^2_i. Así pues,

$$R = \int p(\psi)\cdot I(\psi)\cdot d\psi = NH^{-1}$$

Finalmente,

$$I^{\boldsymbol{\vartheta}}_{\text{aproximada}}\{E^s(\mathbf{n}), N^L(\boldsymbol{\vartheta}_o, T_o^{-1} = H_o)\} = \tfrac{1}{2}\log\frac{|H_o^{-1} + R|}{|H_o^{-1}|} =$$

$$= \tfrac{1}{2}\log\frac{|H_o^{-1} + NH^{-1}|}{|H_o^{-1}|} = \tfrac{1}{2}\log\frac{|H + H_oN|}{|H|} = I^{\boldsymbol{\vartheta}}\{E^s(\mathbf{n}), N^L(\boldsymbol{\vartheta}_o, H_o)\}$$

esto es, la información esperada global exacta (Veres, 1983*a*).

3. EJEMPLO DE CALCULO APROXIMADO

Supongamos cierta investigación que pretende desterminar la distribución en porcentaje y entre los diferentes estratos de cierta característica cualitativa. Así pues, la descripción del comportamiento ante dicha cualidad del individuo j-ésimo de estrato i-ésimo se efectuará a través de la distribución de Bernouilli:

$$p(x_{ij}/\vartheta_i) = \vartheta_i^{x_{ij}}(1 - \vartheta_i)^{1-x_{ij}} \qquad x_{ij} \in \{0, 1\}$$

Supongamos que la información inicial que se posee sobre el vector paramétrico $\boldsymbol{\vartheta}$ puede describirse mediante cierta densidad de Dirichlet. Concretamente, supongamos que $p(\boldsymbol{\vartheta}) = D^L(\boldsymbol{\alpha})$, con $\boldsymbol{\alpha}' = (\alpha_1, \alpha_2, \ldots, \alpha_{L+1})$ y α_i no muy pequeño para todo $i = 1, 2, \ldots, L + 1$. Así pues, el vector $\boldsymbol{\vartheta}$ tomará valores en el cuadrado L-dimensional abierto $(0, 1)^L$, exigiéndose además que

$$\sum_{i=1}^{L} \vartheta_i < 1$$

Una interpretación práctica del modelo aquí mencionado puede encontrarse en veres (1981).

La distribución de Dirichlet es, en cierto sentido, una distribución beta multivariante, lo que nos permite considerar a este modelo como una generalización del binomial. De hecho, si cierta variable aleatoria x tiene una distribución beta, el vector $y' = (x, 1 - x)$ tiene una distribución de Dirichlet. Inversamente, si el vector aleatorio $\boldsymbol{\vartheta}$ se distribuye según una Dirichlet de vector paramétrico $\boldsymbol{\alpha}$ entonces la distribución marginal de cualquier componente de $\boldsymbol{\vartheta}$, por ejemplo, ϑ_j, es una distribución beta con parámetros α_j y $\alpha_o - \alpha_j$. Esto implica que los modelos para cada uno de los estratos x_i por separado responden al modelo binomial estudiado con exhaustividad por Bernardo (1976) y por Basulto y Bernardo (1978). Así pues, denotando por I^{ϑ_i} a la información esperada que el experimento componente $E(n_i)$ proporciona sobre ϑ_i, componente i-ésima de $\boldsymbol{\vartheta}$, y que está definida a partir de las respectivas marginales de las densidades incial y posterior generales (Veres, 1983*a*), puede escribirse que:

$$I^{\vartheta_i}\{E(n_i), m_i(D^L(\boldsymbol{\alpha}))\} = H\{Be(\alpha_i, \alpha_o - \alpha_i)\} -$$

$$- \sum_{r=0}^{L} H\{Be(\alpha_i + r, \alpha_o - \alpha_i + n - r) \cdot Bb_r(r/\alpha_i, \alpha_o - \alpha_i, n)\}$$

donde m_i representa la marginal i-ésima de $D^L(\boldsymbol{\alpha})$ y Bb_r una densidad predictiva beta-binomial, y H la entropía de las densidades beta (Lindley, 1957).

A fin de poder aplicar la aproximación deducida en la sección anterior a la información esperada global, consideremos la siguiente transformación biyectiva:

$$\psi_i = \log \frac{\vartheta_i}{1 - \sum_{j=1}^{L} \vartheta_j} \quad ; \quad i = 1, 2, \ldots, L$$

de jacobiano igual a:

$$J = \left|\frac{\partial \boldsymbol{\psi}}{\partial \boldsymbol{\vartheta}}\right| = 1/\left\{\prod_{i=1}^{L} \vartheta_i\right\}\left(1 - \sum_{j=1}^{L} \vartheta_j\right) \neq 0$$

La correspondiente transformación inversa es

$$\vartheta_i = e^{\psi_i}/\left(1 + \sum_{j=1}^{L} e^{\psi_j}\right) \quad ; \quad i = 1, 2, \ldots, L$$

de jacobiano igual a:

$$J = \left|\frac{\partial \boldsymbol{\vartheta}}{\partial \boldsymbol{\psi}}\right| = \prod_{i=1}^{L} e^{\psi_i} / \left(1 + \sum_{j=1}^{L} e^{\psi_j}\right)^{L+1}$$

De ahí que la densidad de $\boldsymbol{\psi}$ venga dada por:

$$p(\boldsymbol{\psi}) = \{p(\boldsymbol{\vartheta})\}_{\boldsymbol{\vartheta}=F(\boldsymbol{\psi})} \cdot \left|\frac{\partial \boldsymbol{\vartheta}}{\partial \boldsymbol{\psi}}\right| =$$

$$= \frac{\Gamma(\alpha_o)}{\prod_{i=1}^{L+1} \Gamma(\alpha_i)} \cdot \frac{1}{\left(1 + \sum_{j=1}^{L} e^{\psi_j}\right)^{\alpha_0}} \cdot \prod_{i=1}^{L} e^{\psi_i \alpha_i}$$

siendo $\alpha_o = \sum_{i=1}^{L+1} \alpha_i$.

Esta densidad puede aproximarse por una multinormal L-dimensional $N^L(\boldsymbol{\psi}_o, T_o)$, siendo

$$\psi_{oi} = \log \frac{\alpha_i}{\alpha_{L+1}} \quad ; \quad i = 1, 2, \ldots, L$$

y

$$t_{ii} = \alpha_i - \frac{\alpha_i^2}{\alpha_o} \quad i = 1, 2, \ldots, L \quad ; \quad t_{ij} = -\frac{\alpha_i \alpha_j}{\alpha_o} \quad i \neq j = 1, 2, \ldots, L$$

En efecto,

$$p(\boldsymbol{\psi}) = \frac{\Gamma(\alpha_o)}{\prod_{i=1}^{L+1} \Gamma(\alpha_i)} \cdot e^{\left\{\sum_{i=1}^{L} \psi_i \alpha_i - \alpha_o \cdot \log\left(1 + \sum_{j=1}^{L} e^{\psi_j}\right)\right\}} \propto e^{Q}$$

Desarrollando el exponente Q en serie de Taylor alrededor de su máximo, que resulta ser $\boldsymbol{\psi}_o$ deducido de las condiciones

$$\frac{\partial Q}{\partial \psi_i} = \alpha_i - \alpha_o \cdot \frac{e^{\psi_i}}{1 + \sum_{j=1}^{L} e^{\psi_j}} = 0 \quad ; \quad i = 1, 2, \ldots, L$$

se obtiene que

$$p(\boldsymbol{\psi}) \simeq \frac{\Gamma(\alpha_o)}{\prod_{i=1}^{L+1} \Gamma(\alpha_i)} \cdot \exp\left\{\sum_{i=1}^{L} \alpha_i \cdot \log \frac{\alpha_i}{\alpha_{L+1}} - \alpha_o \cdot \log \frac{\alpha_o}{\alpha_{L+1}} - \right.$$

$$\left. - \tfrac{1}{2}(\boldsymbol{\psi} - \boldsymbol{\psi}_o)' T_o (\boldsymbol{\psi} - \boldsymbol{\psi}_o)\right\} \propto e^{-\frac{1}{2}(\boldsymbol{\psi} - \boldsymbol{\psi}_o)' T_o (\boldsymbol{\psi} - \boldsymbol{\psi}_o)}$$

donde T_o es una matriz de precisión, y donde sus elementos se obtienen de las condiciones

$$\left.\frac{\partial^2 Q}{\partial\psi_i\partial\psi_j}\right|_{\psi_0} \quad ; \quad i, j = 1, 2, \ldots, L$$

Notemos que

$$|T_o| = \frac{1}{\alpha_o}\cdot\prod_{i=1}^{L+1}\alpha_i$$

Bajo las hipótesis anteriores pasemos a aproximar la información exacta $I^9\{E^s(\mathbf{n}),\ D^L(\boldsymbol{\alpha})\}$ mediante el resultado del teorema 3. Para ello, y para un cierto i tal que $1 \leqslant i \leqslant L$, resulta que:

$$p(x_{ij}/F_i(\boldsymbol{\psi})) = \frac{e^{\psi_i x_{ij}}\left(1 + \sum\limits_{j\neq 1}^{L} e^{\psi_j}\right)^{1-x_{ij}}}{1 + \sum\limits_{j=1}^{L} e^{\psi_j}}$$

Así pues:

$$\frac{\partial^2 \log p(x_{ij}/F_i(\boldsymbol{\psi}))}{\partial\psi_p\partial\psi_q} = -(1 - x_{ij})\cdot\frac{e^{\psi_p}\cdot e^{\psi_q}}{\left(1 + \sum\limits_{j\neq i}^{L} e^{\psi_j}\right)^2} +$$

$$+ \frac{e^{\psi_p}\cdot e^{\psi_q}}{\left(1 + \sum\limits_{j=1}^{L} e^{\psi_j}\right)^2} \qquad \text{si } p, q \neq i \text{ y } p \neq q$$

$$\frac{\partial^2 \log p(x_{ij}/F_i(\boldsymbol{\psi}))}{\partial\psi_p^2} = (1 - x_{ij})\cdot\frac{e^{\psi_p}\left(1 + \sum\limits_{j\neq p,i}^{L} e^{\psi_j}\right)}{\left(1 + \sum\limits_{j\neq i}^{L} e^{\psi_j}\right)^2} -$$

$$- \frac{e^{\psi_p}\left(1 + \sum\limits_{j\neq p}^{L} e^{\psi_j}\right)}{\left(1 + \sum\limits_{j=1}^{L} e^{\psi_j}\right)^2} \qquad \text{si } p \neq i$$

$$\frac{\partial^2 \log p(x_{ij}/F_i(\boldsymbol{\psi}))}{\partial\psi_p\partial\psi_i} = \frac{e^{\psi_i}\cdot e^{\psi_p}}{\left(1 + \sum\limits_{j=1}^{L} e^{\psi_j}\right)^2} \qquad \text{si } p \neq i$$

$$\frac{\partial^2 \log p(x_{ij}/F_i(\boldsymbol{\psi}))}{\partial\psi_i^2} = -\frac{\exp(\psi_i)\cdot\left(1 + \sum\limits_{j\neq i}^{L} e^{\psi_j}\right)}{\left(1 + \sum\limits_{j=1}^{L} e^{\psi_j}\right)^2}$$

De la definición de $c^i_{pq}(\psi)$, con p, $q = 1, 2, \ldots, L$, resulta:

$$c^i_{pq}(\psi) = -\sum_{x_{ij}=0,1} p(x_{ij}/F_i(\psi))\cdot\frac{\partial^2}{\partial\psi_p\partial\psi_q}\log p(x_{ij}/F_i(\psi))$$

de forma que la matriz $I^i(\psi)$ será la de los elementos $c^i_{pq}(\psi)$ siguientes:

$$c^i_{ii}(\psi) = \frac{e^{\psi_i}\left(1+\sum_{j\neq i}^{L} e^{\psi_j}\right)}{\left(1+\sum_{j=1}^{L} e^{\psi_j}\right)^2}$$

$$c^i_{pi}(\psi) = c^i_{ip}(\psi) = -\frac{e^{\psi_i}\cdot e^{\psi_p}}{\left(1+\sum_{j=1}^{L} e^{\psi_j}\right)^2} \qquad \text{con } p \neq i$$

$$c^i_{pp}(\psi) = \frac{e^{\psi_i}\cdot e^{2\psi_p}}{\left(1+\sum_{j=1}^{L} e^{\psi_j}\right)^2\cdot\left(1+\sum_{j\neq i}^{L} e^{\psi_j}\right)} \qquad \text{con } p \neq i$$

$$c^i_{pq}(\psi) = c^i_{qp}(\psi) = \frac{e^{\psi_i}\cdot e^{\psi_p}\cdot e^{\psi_q}}{\left(1+\sum_{j\neq i}^{L} e^{\psi_j}\right)\cdot\left(1+\sum_{j=1}^{L} e^{\psi_j}\right)^2} \qquad \text{con } p \neq q \quad , \quad p, q \neq i$$

Denotando por R^i a la matriz esperanza de la $I^i(\psi)$ es evidente que

$$R = \sum_{i=1}^{L} n_i\cdot R^i$$

Así pues, R^i será la matriz de elementos r^i_{pq} donde

$$r^i_{pi} = r^i_{ip} = \int c^i_{pi}(\psi)\cdot p(\psi)\cdot d\psi = -\frac{\alpha_i\alpha_p}{\alpha_o(\alpha_o+1)} \qquad \text{si } p \neq i$$

$$r^i_{pp} = \int c^i_{pp}(\psi)\cdot p(\psi)\cdot d\psi = \frac{\alpha_i\alpha_p(\alpha_p+1)}{\alpha_o(\alpha_o+1)(\alpha_o-\alpha_i+1)} \qquad \text{si } p \neq i$$

$$r^i_{pq} = r^i_{qp} = \int c^i_{pq}(\psi)\cdot p(\psi)\cdot d\psi = \frac{\alpha_i\alpha_p\alpha_q}{\alpha_o(\alpha_o+1)(\alpha_0-\alpha_i+1)} \text{ si } p \neq q \text{ y } p, q \neq i$$

$$r^i_{ii} = \int c^i_{ii}(\psi)\cdot p(\psi)\cdot d\psi = \frac{\alpha_i(\alpha_o-\alpha_i)}{\alpha_o(\alpha_o+1)}$$

Finalmente, la matriz R será la matriz de elementos r_{pq}:

$$r_{pq} = \sum_{i=1}^{L} n_i \cdot r_{pq}^i$$

y la matriz $T_o + R$ tendrá como elementos $s_{pq} = r_{pq} + t_{pq}$ donde:

$$(s_{ij}) = \begin{cases} s_{ii} = \dfrac{n_i\alpha_i(\alpha_o - \alpha_i)}{\alpha_o(\alpha_o + 1} + \displaystyle\sum_{j \neq i}^{L} \dfrac{n_j\alpha_i\alpha_j(\alpha_i + 1)}{\alpha_o(\alpha_o + 1)(\alpha_o - \alpha_j + 1)} + \dfrac{\alpha_i(\alpha_0 - \alpha_i)}{\alpha_o} \\ s_{ij} = -\dfrac{(n_i + n_j)\alpha_i\alpha_j}{\alpha_o(\alpha_o + 1)} + \displaystyle\sum_{h \neq i,j}^{L} \dfrac{n_h\alpha_h\alpha_i\alpha_j}{\alpha_o(\alpha_o + 1)(\alpha_o - \alpha_h + 1)} - \dfrac{\alpha_i\alpha_j}{\alpha_o} \end{cases}$$

Aplicando, pues, el teorema 3 resulta en último término que

$$I^{9}\{E^s(\mathbf{n}), D^L(\boldsymbol{\alpha})\} \simeq \tfrac{1}{2}\log\frac{|T_o + R|}{|T_o|} = \tfrac{1}{2}\log\frac{\alpha_o \cdot |T_o + R|}{\prod_{i=1}^{L+1} \alpha_i}$$

y donde las matrices T_o y R han quedado definidas en todo el desarrollo anterior.

Tal como era de esperar la coherencia de este resultado queda comprobada al aplicarlo al modelo binomial y ver la coincidencia con la aproximación obtenida para dicho modelo por Bernardo (1979). En efecto, considerando un único estrato y siendo la distribución inicial una beta de parámetros α_1 y α_2, resulta:

$$T_o = \frac{\alpha_1\alpha_2}{\alpha_1 + \alpha_2} = \frac{\alpha_1\alpha_2}{\alpha_o} \quad ; \quad R = n_1 \cdot \frac{\alpha_1\alpha_2}{\alpha_o(\alpha_o + 1)}$$

de forma que

$$I^{9_1}_{\text{aproximada}}\{E(n_1), Be(\alpha_1, \alpha_2)\} \simeq \tfrac{1}{2}\log\frac{\alpha_1 + \alpha_2}{\alpha_1\alpha_2}\left(n_1 \cdot \frac{\alpha_1\alpha_2}{\alpha_o(\alpha_o + 1)} + \frac{\alpha_1\alpha_2}{\alpha_o}\right) =$$

$$= \tfrac{1}{2}\log\left(1 + \frac{n_1}{\alpha_1 + \alpha_2 + 1}\right)$$

resultado que coincide con la aproximación obtenida por Bernardo (1979).

REFERENCIAS

BASULTO, J., y BERNARDO, J. M.: «Análisis Bayesiano de un proceso binomial». *Trabajos de Estadística y de Investigación Operativa, 29*, 3-27, 1978.

BERNARDO, J. M.: *The use of information in the design and analysis of scientific experimenttion.* Tesis, Universidad de Londres, 1976.

——: «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, 1978.

——: «Comportamiento asintótico de la información proporcionada por un experimento». *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales de Madrid, 73*, 491-502, 1979.

COXD. R., y HINKLEY, D. V.: *Theoretical Statistics*, Champmand and Hall, Londres, 1974.

DEFINETTI, B.: «La prévision, ses lois logiques, ses sources subjectives». Trabajo aparecido en Kyburg y Smokler (eds., 1964): *Studies in Subjetive Probability*, MacGraw-Hill, Nueva York, 1937.

GRAYBILL, F. A.: *An Introduction to Linnear Statistical Models*, McGraw-Hill, Nueva York, 1961.

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

——: «Binomial Sampling and the Concept of Information». *Biometrika, 44*, 179-46, 1957.

——: *Introduction to Probability and Statistics*, Cambridge University Press, 1965.

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

VERES, E.: «Información esperada proporcionada por un experimento asociado a un diseño estratificado». *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales de Madrid*, 77, 569-595, 1983*a*.

——: «Métodos de obtención de la información esperada global». *Trabajos de Estadística y de Investigación Operativa*, 34, 79-94, 1983*b*.

——: *Diseño Bayesiano de muestras, posiblemente estratificadas, cuando la utilidad terminal es función de la información conseguida.* Tesis, Universidad de Valencia, 1981.

Instituto Nacional de Estadística. Valencia