TRABAJOS DE ESTADISTICA Y
DE INVESTIGACION OPERATIVA
Vol. 32, Núm. 3, 1981, pp. 18 a 31

INFERENCIA BAYESIANA SOBRE EL COEFICIENTE DE CORRELACION DE UNA POBLACION NORMAL BIVARIANTE

María J. Bayarri
Departamentos de Bioestadística-Estadística
Universidad de Valencia

RESUMEN

En el presente artículo, se utiliza el método de maximización de la información desconocida para deducir las distribuciones inicial y final de referencia cuando se trata de hacer inferencias sobre el coeficiente de correlación de una población normal bivariante.

Palabras clave: Coeficiente de correlación; Distribución Normal Bivariante; Distribuciones de Referencia; Distribuciones Mínimo Informativas; Inferencia Bayesiana.

Clasificación según la AMS (1980): Primaria: 62F15; Secundaria: 62B10, 62A15.

1. INTRODUCCION

En la Metodología Estadística, el estudio de la relación existente entre dos cantidades aleatorias que se distribuyen conjuntamente normales, es un caso muy frecuente y de indudable interés práctico.

Una descripción posible de tal relación podría ser la covarianza, pero esta depende de las unidades de medida. Se presenta entonces la necesidad de crear una medida adimensional de la relación existente entre dos o más cantidades aleatorias. De esta forma se introdujo el coeficiente de correlación ρ. Posteriormente, Katz L. (1975) justificaría su empleo mostrando que ρ constituye de hecho una medida razonable de la colinealidad de dos cantidades aleatorias al minimizar la media de los cuadrados de las distancias ortogonales a la recta de mínimos cuadrados.

Consideremos, pues, el tipo de inferencias que pueden realizarse sobre el coeficiente de correlación. Desde un punto de vista ortodoxo el problema fue tratado de forma exhaustiva pero los resultados no fueron del todo satisfactorios: la intratabilidad matemática de la distribución en el muestreo $p\,(r\,|\,\theta)$ del coeficiente de correlación muestral r, da lugar a una numerosa serie de aproximaciones sin indicaciones claras (a no ser que se utilice la información subjetiva que el estadístico posee sobre ρ) de cuando utilizar una u otra; se han deducido incontables estimadores insesgados y, por último, el estimador máximo verosimil puede ser la solución de una ecuación de tercer grado y, por lo tanto, no ser único. Para un estudio más detallado ver Bayarri (1979).

Desde una perspectiva Bayesiana, la dependencia de los resultados con respecto de la distribución inicial utilizada, ha sido uno de los puntos más criticados. Aunque tal aspecto de la Inferencia Bayesiana no debería molestar por ser altamente realista, se consideró interesante estudiar distribuciones iniciales que representen en algún sentido, un estado de ignorancia inicial, llamándolas distribuciones iniciales no informativas, que, combinadas con la función de verosimilitud mediante el Teorema de Bayes, dan lugar a las distribuciones finales mínimo informativas. Tales distribuciones sirvieron como un argumento más a favor de la Inferencia Bayesiana al reproducir numéricamente en muchos casos los resultados clásicos. Sin embargo dieron lugar a una serie de dificultades como las paradojas de marginalización de Dawid, Stone & Zidek (1973).

Posteriormente Bernardo (1976, 1979), partiendo de la idea de que las iniciales no informativas dependen específicamente del parámetro de interés (y no sólo del modelo) desarrolló, utilizando Teoría de la Información, un método general que salva las dificultades mencionadas, y es, por tanto, el método más prometedor aparecido hasta el momento.

En este trabajo, estudiamos las consecuencias de aplicar el *método* *de maximización de la información desconocida* propuesto por Bernardo (1979) cuando queremos hacer inferencias sobre el coeficiente de correlación ρ de una población normal bivariante.

2. NOTACION

Supongamos que X, *espacio muestral* de elementos x (resultados experimentales), está dotado de una σ-álgebra apropiada, sobre la que se ha definido una familia de distribuciones de probabilidad identificadas mediante un parámetro θ perteneciente a un *espacio paramétrico* Θ. Tales medidas de probabilidad las supondremos dominadas por una medida σ-finita de forma que pueden ser descritas mediante sus funciones de densidad $p(x \mid \theta)$ respecto de dicha medida dominante.

El argumento Bayesiano extiende este modelo básico y supone que θ soporta una σ-álgebra y una medida de probabilidad que describiremos mediante su función de densidad $p(\theta)$ respecto de cierta medida σ-finita dominante definida en Θ.

El caso más general es aquel en que se trata de hacer inferencias sobre $\psi = \psi(\theta)$, una función conocida de θ que, en particular, puede ser el propio parámetro θ. Sin pérdida de generalidad (efectuando el cambio de variable adecuado) podemos suponer que $\theta = (\psi, \omega)$ donde ψ es la *magnitud de interés*, y ω el *parámetro marginal*.

El modelo con el que vamos a trabajar es la normal bivariante cuya función de verosimilitud en función de $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$ es:

$$p(x, y \mid \mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = N_2 \left[\begin{bmatrix} x \\ y \end{bmatrix} \middle| \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\,\sigma_1\sigma_2 \\ \rho\,\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}\right] =$$

$$= \frac{1}{2\pi\,\sigma_1\sigma_2(1-\rho^2)^{1/2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(\mu_1 - x)^2}{\sigma_1^2} + \right.\right.$$

$$\left.\left. + \frac{(\mu_2 - y)^2}{\sigma_2^2} - \frac{2\rho(\mu_1 - x)(\mu_2 - y)}{\sigma_1\sigma_2}\right]\right\}$$

3. DISTRIBUCIONES INICIALES DE REFERENCIA

Un estudio del método de maximización de la información desconocida sugiere, que, en principio podría aplicarse de dos formas alternativas:

Con $\theta = \{\psi, \omega\}$, podemos aplicar el método para obtener $\pi(\omega/\psi)$ y utilizar este resultado para obtener $\pi(\psi)$, con lo que la distribución inicial de referencia $\pi(\theta)$ para θ es, precisamente:

$$\pi_\psi(\theta) \propto \pi(\psi) \cdot \pi(\omega | \psi)$$

Este es el tratamiento que parece natural y le llamaremos tratamiento *'global'* porque maneja el parámetro marginal ω como un todo.

El método alternativo (Bernardo 1976) consiste en obtener independientemente la inicial de referencia, no de todo el vector ω condicionado a ψ, sino de cada componente ω_i de ω condicionado al resto de las componentes de θ con lo que:

$$\pi_\psi(\theta) \propto \pi(\psi) \prod_{i=1}^{n} \pi_i(\omega_i | \psi, \omega_1, \omega_2, \ldots, \omega_{i-1})$$

A esta forma de aplicar el método lo llamaremos tratamiento *'secuencial'*. Las implicaciones que pudiera tener una reordenación de las componentes de θ, una reparametrización o un conflicto entre las iniciales deducidas según se utilice uno u otro tratamiento, no van a discutirse aquí excepto en lo que al coeficiente de correlación se refiere.

Situémonos en el caso más general en que todos los parámetros son desconocidos. En este caso:

$$\theta = \{\mu_1, \mu_2, \sigma_1, \sigma_2, \rho\} \qquad \psi = \rho \qquad \omega = \{\mu_1, \mu_2, \sigma_1, \sigma_2\}$$

Mediante una serie de cálculos sencillos, la matriz de Información de Fisher, $F(\theta)$, cuyo elemento general es de la forma:

$$F_{ij}(\theta) = -\int p(z|\theta) \frac{\partial^2}{\partial\theta_i\theta_j} \{\log p(z|\theta)\}\, dz$$

donde z representa una muestra de tamaño n, de una normal bivariante en nuestro caso, resulta ser:

$$F(\theta) = \begin{bmatrix} A(\theta) & 0 \\ 0 & B(\theta) \end{bmatrix}$$

con:

$$A(\theta) = \begin{bmatrix} \dfrac{1}{\sigma_1^2 (1-\rho^2)} & -\dfrac{\rho}{\sigma_1 \sigma_2 (1-\rho^2)} \\[2ex] -\dfrac{\rho}{\sigma_1 \sigma_2 (1-\rho^2)} & \dfrac{1}{\sigma_2^2 (1-\rho^2)} \end{bmatrix}$$

$$B(\theta) = \begin{bmatrix} \dfrac{2-\rho^2}{\sigma_1^2 (1-\rho^2)} & -\dfrac{\rho^2}{\sigma_1 \sigma_2 (1-\rho^2)} & -\dfrac{\rho}{\sigma_1 (1-\rho^2)} \\[2ex] -\dfrac{\rho^2}{\sigma_1 \sigma_2 (1-\rho^2)} & \dfrac{2-\rho^2}{\sigma_2^2 (1-\rho^2)} & -\dfrac{\rho}{\sigma_2 (1-\rho^2)} \\[2ex] -\dfrac{\rho^2}{\sigma_1 (1-\rho^2)} & -\dfrac{\rho}{\sigma_2 (1-\rho^2)} & \dfrac{1+\rho^2}{(1-\rho^2)^2} \end{bmatrix}$$

y su inversa es:

$$F^{-1}(\theta)=\begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & 0 & 0 & 0 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & 0 & 0 & 0 \\ 0 & 0 & \frac{\sigma_1^2}{2} & \frac{\rho^2\sigma_1\sigma_2}{2} & \frac{\rho\sigma_1(1-\rho^2)}{2} \\ 0 & 0 & \frac{\rho^2\sigma_1\sigma_2}{2} & \frac{\sigma_2^2}{2} & \frac{\rho\sigma_2(1-\rho^2)}{2} \\ 0 & 0 & \frac{\rho\sigma_1(1-\rho^2)}{2} & \frac{\rho\sigma_2(1-\rho^2)}{2} & (1-\rho^2)^2 \end{bmatrix}$$

Como θ es asintóticamente normal con precisión $n\,F(\theta)$, ρ es asintóticamente normal con precisión $h_0(\theta)$ donde:

$$[h_0(\theta)]^{1/2}=\left(\frac{1}{(1-\rho^2)}\right)^{1/2}\propto \pi(\rho)\,f(\omega) \qquad (1)$$

Asimismo, $\omega|\rho$ es asintóticamente normal con matriz de precisión $h_1(\theta)$ cuyo determinante elevado a 1/2 es:

$$|h_1(\theta)|^{1/2}=\left(\frac{1}{\sigma_1^4\,\sigma_2^4\,(1-\rho^2)^2}\right)^{1/2}\propto \pi(\omega|\rho)\,f(\rho) \qquad (2)$$

Puesto que existe normalidad asintótica a posteriori de θ y la precisión (matriz de precisión) asintótica final de ρ ($\omega|\rho$) factoriza en la forma indicada por (1) ((2)), el tratamiento global del método de maximización de la información desconocida, da lugar a la siguiente distribución inicial de referencia:

$$\pi_\rho(\theta)\propto\pi(\rho)\,\pi(\omega|\rho)\propto\frac{1}{(1-\rho^2)}\,\frac{1}{\sigma_1^2\,\sigma_2^2} \qquad \text{con } \pi(\rho)\propto\frac{1}{1-\rho^2}$$

Análogamente, si utilizamos el tratamiento secuencial del método, y si

$$\{h_0(\theta)\}^{1/2} = \pi(\rho)\, f(\omega)$$

$$\{h_i'(\theta)\}^{1/2} = \pi_i(\omega_i | \rho, \omega_1, ..., \omega_{i-1})\, f(\theta_i^*)$$

donde:

$h_0(\theta)$ es la precisión a posteriori de ρ

$h_i'(\theta)$ es la precisión a posteriori de $\omega_i | \rho, \omega_i, ..., \omega_{i-1}$

$\theta_i^* = \theta \sim \omega_i$

$f(\omega), f(\theta_i^*)$ son funciones únicamente de ω y θ_i^* respectivamente.

Entonces:

$$\pi_\rho(\theta) \propto \pi(\rho) \prod_{i=1}^{4} \pi_i(\omega_i | \rho, \omega_1, \omega_2, \omega_3)$$

Todas las precisiones buscadas se obtienen fácilmente mediante los resultados sobre marginalización y condicionamiento en una normal multivariante, obteniéndose finalmente:

$$\pi_\rho(\theta) \propto \frac{1}{\sigma_1\, \sigma_2\, (1 - \rho^2)} \qquad \text{con } \pi(\rho) \propto (1 - \rho^2)^{-1}$$

que difiere de la anterior únicamente en el exponente de σ_1 y σ_2, y coincide con la propuesta por Lindley.

Nótese que a la vista de la estructura de $F(\theta)$, es claro que $\pi_\rho(\theta)$ no depende del orden en que consideremos las marginales, es decir, es independiente del orden que consideremos en las componentes de ω.

4. DISTRIBUCIONES FINALES DE REFERENCIA

Empezaremos calculando la distribución final de referencia con una inicial general de la forma:

$$\pi(\theta) \propto \frac{\pi(\rho)}{\sigma_1^a \ \sigma_2^a} \quad \text{con } \pi(\rho) \propto (1 - \rho^2)^{-1} \tag{3}$$

Si $z = \{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$ es una muestra de tamaño n de una distribución normal bivariante, la distribución final utilizando (3) como inicial es:

$$\pi(\rho | z) \propto \int \frac{\pi(\rho)}{\sigma_1^a \ \sigma_2^a} \, p(z | \theta) \, d\omega$$

La integración respecto de (μ_1, μ_2) es sencilla utilizando los resultados de una normal bivariante. Para integrar respecto de σ_1, σ_2 utilizaremos la transformación propuesta por Fisher (1915):

$$\sigma_2^2 = \xi^2 \ \sigma_1^2 \ \frac{S_2^2}{S_1^2} \quad \text{con } S_2^2 = \Sigma \ (x_i - \bar{x})^2 \quad S_2^2 = \Sigma \ (y_i - \bar{y})^2$$

con lo que finalmente:

$$\pi(\rho | z) \propto \frac{\pi(\rho) \, (1 - \rho^2)^{\frac{n + 2a - 3}{2}}}{(1 - \rho \, r)^{n + a - (5/2)}} \int_0^1 (1 - u)^{n + a - 3} \ u^{-1/2} \ .$$

$$\cdot \left[1 - \frac{u \, (1 + \rho \, r)}{2} \right]^{-1/2} du \tag{4}$$

donde u es tal que $\xi - 2 \, r + \xi^{-1} = 2 \, \dfrac{1 - r}{1 - u}$.

La integral que aparece en (4) fue estudiada por Hotelling (1953) de forma que si denotamos la conocida función hipergeométrica (ver Abramowitz & Stegun pag. 556) por:

$$F(a, b, c, d) = \frac{\Gamma(c)}{\Gamma(a)\,\Gamma(b)} \sum_{n=0}^{\infty} \frac{\Gamma(a+n)\,\Gamma(b+n)}{\Gamma(c+n)} \frac{d^n}{n!}$$

puede demostrarse que tal integral se reduce a:

$$F\left(\frac{1}{2}, \frac{1}{2}, n + a - \frac{3}{2}, \frac{1}{2}(1 + \rho r)\right)$$

que converge rápidamente a 1 y, si n es grande, puede por tanto ser aproximada por la unidad.

Tenemos, en consecuencia, dos distribuciones finales de referencia:

a) Si deducimos la inicial aplicando el tratamiento *global,* la inicial resultante es de la forma (3) con $a = 2$, y, por tanto:

$$\pi(\rho|z) \propto \frac{(1-\rho^2)^{(n-1)/2}}{(1-\rho r)^{n-1/2}} F\left(\frac{1}{2}, \frac{1}{2}, n + \frac{1}{2}, \frac{1}{2}(1 + \rho r)\right) \quad (5)$$

b) Si utilizamos el tratamiento secuencial, resulta una inicial de referencia de la forma (3) con $a = 1$ y, por tanto:

$$\pi(\rho|z) \propto \frac{(1-\rho^2)^{(n-3)/2}}{(1-\rho r)^{n-3/2}} F\left(\frac{1}{2}, \frac{1}{2}, n - \frac{1}{2}, \frac{1}{2}(1 + \rho r)\right) \quad (6)$$

Obviamente, un tamaño muestral moderadamente grande va a hacer que ambas distribuciones finales sean prácticamente idénticas. Sin embargo, es inevitable plantearse preguntas como: i) ¿Qué distribución final debemos emplear? ii) ¿Hay argumentos suficientes para rechazar la distribución inicial para ρ que se deduce de aplicar la regla de Jeffreys multivariante: $\pi_\rho(\theta) \propto (1-\rho^2)^{-3/2}\,\sigma_1^{-1}\,\sigma_2^{-1}$?

Un tratamiento alternativo del modelo va a conducirnos en el próximo apartado a rechazar la solución (5) y a contestar afirmativamente a la segunda pregunta.

5. RESULTADOS FINALES

Como podemos ver en (4), la distribución final de ρ no depende de la muestra z más que a través del estadístico r, y esto para cualquier distribución inicial para ρ, $\pi(\rho)$. En consecuencia, r es en nuestras condiciones, un estadístico suficiente (para una discusión de este punto ver Jaynes (1980) y Dawid (1980)). Además, la distribución de r es conocida y depende de θ únicamente a través de ρ:

$$p(r|\theta) = p(r|\rho) \propto \frac{(1-\rho^2)^{(n-1)/2}\,(1-r^2)^{(n-4)/2}}{(1-\rho\, r)^{n-3/2}}$$

$$F\left(\frac{1}{2}, \frac{1}{2}, n-\frac{1}{2}, \frac{1+\rho\, r}{2}\right)$$

por tanto debería cumplirse que:

$$\pi(\rho|z) = \pi(\rho|r) \propto \pi(\rho)\, p(r|\rho)$$

Puesto que cualquiera que sea el tratamiento que utilicemos (global o secuencial) la distribución inicial para ρ resulta ser $\pi(\rho) \propto$ $\propto (1-\rho^2)^{-1}$, y puesto que una vez obtenida la muestra, r es una constante, el método de maximización de la información desconocida da lugar a la distribución final de referencia:

$$\pi(\rho|r) \propto \frac{(1-\rho^2)^{(n-3)/2}}{(1-\rho\, r)^{n-3/2}} F\left(\frac{1}{2}, \frac{1}{2}, n-\frac{1}{2}, \frac{1+\rho\, r}{2}\right) \qquad (7)$$

Una comparación de (5) y (7) pone de manifiesto una paradoja de marginalización puesto que $\pi(\rho|z)$ no es proporcional a $\pi(\rho)\, p(r|\rho)$.

Tal resultado conduce a rechazar (7) y a adoptar (6) como distribución final de referencia para ρ, independientemente de los problemas intrínsecos que el tratamiento secuencial del método de maximización de la información desconocida pudiera tener.

Por otra parte, podemos utilizar la distribución de r para tratar el problema de obtención de $\pi(\rho)$ como si se tratase del modelo unidimensional $p(r|\rho)$, en los que el método de maximización de la información desconocida no plantea la disyuntiva esbozada en la sección 3, y comparar el resultado con los obtenidos hasta ahora.

Consideremos, por tanto, que nuestro modelo no es $p(z|\theta)$, sino $p(r|\rho)$. Utilizando la expresión (Bernardo, 1979):

$$\pi(\rho) \propto \exp \int p(r|\rho) \log p^*(\rho|r)\, dr$$

donde $p^*(\rho|r)$ es la distribución asintótica final de ρ, empleando para ello la transformación propuesta por Fisher (1921):

$$\rho = \tanh \delta \qquad r = \tanh t$$

y el hecho (Lindley 1965/70, pág. 215) de que la distribución final de δ cuando $n \to \infty$ es una normal de media $\tanh^{-1} r$ y precisión n, es decir:

$$p^*(\delta|t) = N(\delta|t, n)$$

de forma que:

$$\log p^*(\delta|t) = \frac{1}{2} \log \frac{n}{2\pi} \quad - (t - \delta)^2$$

Por otra parte (Fisher 1921), si n es grande (recuérdese que trabajamos en distribuciones asintóticas):

$$p(t|\delta) \cong N(t|\delta, n-3)$$

de forma que la distribución inicial de referencia de δ, $\pi(\delta)$ es, por tanto:

$$\pi(\delta) \propto \exp\left\{\frac{1}{2}\log\frac{n}{2} - E_t\,(t-\delta)^2\right\}$$

y puesto que $E_t\ (t-\delta)^2 = 1/(n-3)$ concluimos:

$$\pi(\delta) \propto \text{constante}$$

por lo tanto:

$$\pi(\rho) \propto \pi(\delta)\left|\frac{d\delta}{d\rho}\right| \propto \frac{1}{1-\rho^2}$$

que es exactamente la misma inicial obtenida antes, y que nos lleva a descartar la que viene usándose hasta ahora:

$$\pi(\rho) \propto (1-\rho^2)^{-3/2}$$

Además, $\pi(\rho) \propto (1-\rho^2)^{-1}$ tiene una interpretación atractiva como límite de distribuciones propias. En efecto: si trabajamos con la familia conjugada de la multinormal (la normal Wishart), tal distribución inicial equivale a tomar para Σ^{-1} (inversa de la matriz de varianzas-covarianzas) una Wishart inicial con matriz de precisión B verificando $tr\,(B\,\Sigma) = 0$, y con $\alpha = 1$ grados de libertad (DeGroot 1970) lo cual tiene un claro sentido intuitivo como 'límite' de distribuciones iniciales propias ($\alpha > 1$). No ocurre así con la deducida de la regla de Jeffreys multivariante, equivalente al caso en que $\alpha = 0$ que no sólo es impropia, sino que además es inalcanzable como límite de distribuciones Wishart.

6. CONCLUSIONES

Podemos concluir que *la* distribución inicial de referencia de ρ cuando queremos hacer inferencias sobre ρ en un modelo Normal con todos los parámetros desconocidos, es precisamente, $\pi(\rho) \propto (1 - \rho^2)^{-1}$ y la correspondiente de θ: $\pi_\rho(\theta) \propto \sigma_1^{-1}\, \sigma_2^{-1}\, (1 - \rho^2)^{-1}$. Por tanto *la* distribución final de referencia es:

$$\pi(\rho|z) \propto \frac{(1 - \rho^2)^{(n-3)/2}}{(1 - \rho\, r)^{n-3/2}}\, F\left(\frac{1}{2}, \frac{1}{2}, n - \frac{1}{2}, \frac{1}{2}(1 + \rho\, r)\right)$$

y si además n es moderadamente grande de forma que no haya error apreciable al aproximar la serie hipergeométrica por la unidad, puede utilizarse la aproximación:

$$\pi(\rho|z) \propto \frac{(1 - \rho^2)^{(n-3)/2}}{(1 - \rho\, r)^{n-3/2}}$$

Asimismo concluimos que es necesario tratar con cuidado el método de maximización de la información desconocida para construir distribuciones iniciales de referencia cuando existan parámetros marginales, especialmente si se utiliza el tratamiento global pues, como hemos visto, da lugar a resultados insatisfactorios.

REFERENCIAS

Bayarri, M.J. (1979). *Inferencia sobre el coeficiente de correlación de una población normal bivariante.* Tesis de Licenciatura. Universidad de Valencia.

Bernardo, J.M. (1976). *The use of information in the design and analysis of scientific experimentation.* Ph. D. Thesis. University of London.

Bernardo, J.M. (1979). *Reference Posterior Distribution for Bayesian Inference.* J. Roy. Stistist. Soc. B, 41.

Dawid, A.P. (1980). *A Bayesian look at nuisance parameters.* En Bayesian Statistics (Bernardo, J.M. et al. eds.). Valencia: Imprenta Universitaria (En prensa).

Dawid, A.P., Stone, M. & Zidek, J.V. (1973). *Marginalization paradoxes in Bayesian and structural inference.* J. Roy. Statist. Soc. B, 35, 189-233.

DeGroot, M.H. (1970). *Optimal Statistical Decisions.* New York: McGraw-Hill.

Fisher, R.A. (1915). *Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population.* Biometrika, 10, 507-521.

Fisher, R.A. (1921). *On the "probable error" of a coefficient of correlation deduced from a small sample.* Metron 1(4), 3-32.

Hotelling, H. (1953). *New light on the correlation coefficient and its transforms.* J. Roy. Statist, Soc. B, 193-232.

Jaynes, E.T. (1980). *Marginalization and prior probabilities.* En Studies in Bayesian Statistics (Zellner A. ed.). Amsterdam: North Holland (En prensa).

Katz, L. (1975). *A direct development of the correlation coefficient.* The American Statician, Vol. 29, n° 4, 170.

Lindley, D.V. (1965). *Introduction to Probability and Statistics from a Bayesian Viewpoint.* Cambridge: Cambridge University Press.

Oberhettinger, F. (1965). *Hypergeometric Functions.* En Handbook of Mathematical Functions (Abramowitz & Stegun eds.) New York: Dover.