CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS

INFERENCIA BAYESIANA SOBRE EL COEFICIENTE DE VARIACION: UNA SOLUCION A LA PARADOJA DE MARGINALIZACION *

José-Miguel Bernardo

PUBLICADO EN LA REVISTA
TRABAJOS DE ESTADISTICA
Y DE
INVESTIGACION OPERATIVA

Vol. XXVIII
Madrid 1977

TRABAJOS DE ESTADISTICA Y
DE INVESTIGACION OPERATIVA
Vol XXVIII Cuad 2 y 3 Madrid, 1977

INFERENCIA BAYESIANA SOBRE EL COEFICIENTE DE VARIACION: UNA SOLUCION A LA PARADOJA DE MARGINALIZACION *

José-Miguel Bernardo
Departamento de Estadística
Universidad de Valencia
Paseo al Mar 13. Valencia-10

Sumario

En Inferencia Bayesiana, el uso indiscriminado de distribuciones iniciales impropias "no informativas" da lugar a ciertos resultados insatisfactorios conocidos como paradojas de marginalización. Utilizando un nuevo método para la construcción de distribuciones iniciales de referencia desarrollado por el autor, se ejemplifica la solución de tales paradojas con el análisis del problema de inferencia planteado por el coeficiente de variación

Frases clave.

Distribuciones Iniciales Difusas, Distribuciones de Referencia, Inferencia Bayesiana, Información Esperada, Paradojas de Marginalización.

Clasificación según la AMS (1970).

Primaria. 62F15, 62A15; *Secundaria:* 62B10

---

* Presentado en la IX Reunión Nacional de Investigación Operativa, celebrada en Barcelona, 27-29 septiembre 1976.

1. El problema

El uso indiscriminado de distribuciones iniciales impropias "no-informativas" en el análisis estadístico Bayesiano ha sido recientemente criticado por numerosos autores. Mediante una colección de ejemplos, Stone & Dawid (1972) y Dawid, Stone & Zidek (1973) han puesto de manifiesto que el uso de una *única* distribución inicial "no-informativa" para cualquier problema de inferencia en un determinado modelo da lugar a ciertos resultados insatisfactorios conocidos como paradojas de marginalización.

Un ejemplo particularmente sencillo de este comportamiento, que ocurre a menudo en la práctica estadística y cuya solución tiene por tanto interés en sí misma, aparece al intentar realizar inferencias sobre el coeficiente de variación de una distribución normal.

En efecto, supóngase que se dispone de una muestra aleatoria $z = \{x_1, x_2, \ldots, x_n\}$ de una población normal de media μ y varianza σ^2 desconocidas y que, basándose únicamente en la información proporcionada por tal muestra, se pretende hacer inferencias sobre el coeficiente de variación $\psi = \mu/\sigma$.

La distribución inicial "no-informativa" comúnmente aceptada para el modelo normal con ambos parámetros desconocidos, ya propuesta por Jeffreys (1939/67, p. 138) y con otros argumentos por Barnard (1952) es $p(\mu, \sigma) = \sigma^{-1}$.

Con tal distribución inicial, la distribución posterior de μ y σ tras observar la muestra, $p(\mu, \sigma | z)$, resulta, en virtud del teorema de Bayes, proporcional a

$$\left(\frac{1}{\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\,\Sigma\,(x_i-\mu)^2\right\}\sigma^{-1} = \sigma^{-(n+1)} \exp\left\{-\frac{n}{2\sigma^2}(\mu-m)^2 - \frac{ns^2}{2\sigma^2}\right\}$$

donde m y s^2 son, respectivamente, la media y la varianza muestrales. En consecuencia, la distribución posterior de ψ y σ, $p(\psi, \sigma | z)$, es proporcional a

$$\sigma^{-n} \exp\left\{-\frac{n\psi^2}{2} + \frac{nm\psi}{\sigma} - \frac{n}{2\sigma^2}(m^2+s^2)\right\}$$

puesto que el Jacobiano de la transformación es σ^{-1}. Integrando σ, la distribución posterior de ψ resulta de la forma

$$p(\psi | z) \propto \exp\left\{-\frac{n\psi^2}{2}\right\} \int_0^\infty t^{n-2} \exp\left\{-\frac{1}{2}t^2 + r\psi t\right\} dt \qquad (1)$$

donde $r = (\Sigma x_i)/(\Sigma x_i^2)^{1/2}$. Consecuentemente, la distribución final de ψ depende de la muestra únicamente a través de r, i.e. r es suficiente para realizar inferencias sobre ψ.

Por otra parte (Stone & Dawid, 1972), la distribución de r depende de μ y σ únicamente a través de ψ. Específicamente,

$$p(r | \mu, \sigma) = p(r | \psi) \propto$$

$$\propto \exp\left\{-\frac{n\psi^2}{2}\right\}\left(1-\frac{r^2}{n}\right)^{(n-3)/2} \int_0^\infty t^{n-1} \exp\left\{-\frac{t^2}{2} + r\psi t\right\} dt \qquad (2)$$

Por tanto, explotando este hecho, deberían poderse hacer inferencias sobre el valor de ψ asignando una distribución inicial para ψ, $p(\psi)$, y combinándola con (2) por medio del teorema de Bayes para obtener $p(\psi | z) = p(\psi | r)$.

Sin embargo, como función de ψ, (2) *no* es un múltiplo de (1) y en consecuencia no existe *ninguna* distribución inicial para ψ que permita reproducir las inferencias sobre ψ que se deducirían de (1). Este resultado arroja serias dudas sobre la validez de (1) como descripción de las inferencias que se pueden realizar sobre ψ basándose en la muestra z. En la Sección 3, presentamos una solución alternativa que no presenta tal dificultad.

2. Definición de una distribución inicial de referencia

Nuestra postura, desarrollada en Bernardo (1976), es que para *cada* problema de inferencia en un determinado modelo existe una distribución inicial de referencia que puede utilizarse como "no-informativa" en el sentido de que la correspondiente distribución posterior $p(\psi | z)$ del parámetro de interés ψ describe la clase de posibles valores de ψ que son sostenidos por la muestra z.

La idea básica subyacente a nuestra construcción de una distribución inicial de referencia es como sigue. Considérese un vector θ de parámetros y un experimento E cuyo resultado z proporciona información sobre θ. Supóngase que estamos interesados en el valor de cierta función *real* $\psi = \psi(\theta)$ de los parámetros. Sin pérdida de generalidad, podemos suponer que ψ es la primera componente de θ, puesto que en otro caso siempre podemos hacer la transformación adecuada para obtener tal situación. Por tanto, escribiremos $\theta = \{\psi, \omega\}$, $\omega = \{\omega_1, \omega_2, \ldots, \omega_k\}$.

Ante todo, generalizando la definición original de Lindley (1956), la información que puede esperarse sobre el valor de ψ de la realización de E cuando la distribución inicial de θ es $p(\theta)$, se define como

$$I^{\psi}\{E, p(\theta)\} = \int\int p(z, \theta) \log \frac{p(\psi|z)}{p(\psi)} d\theta\, dz$$

cuando tal integral existe. Análogamente, la información residual esperada sobre cada uno de los ω_i cuando ψ es conocido puede ser definida como

$$I^{\omega_i|\psi}\{E, p(\theta)\} = \int\int p(z, \theta) \log \frac{p(\omega_i|\psi, z)}{p(\omega_i|\psi)} d\theta\, dz$$

Considérese ahora el experimento $E(n)$ que consiste en n repeticiones independientes de E, de forma que $I^{\psi}\{E(n), p(\theta)\}$ es la información sobre ψ que puede esperarse de $E(n)$ cuando la distribución inicial de $\theta = \{\psi, \omega\}$ es $p(\theta) = p(\psi)\, p(\omega|\psi)$. Si se realizasen infinitas repeticiones de E se llegaría a conocer θ, y por tanto ψ, perfectamente. Por tanto, $I^{\psi}\{E(\infty), p(\theta)\}$, si existe, mide la cantidad de *información desconocida* sobre el valor de ψ, cuando la información inicial sobre θ está descrita por $p(\theta)$, que podría obtenerse repitiendo E. Parece natural definir una distribución difusa de ψ con respecto a E, $\pi(\psi)$ como aquélla que maximiza la información desconocida sobre ψ para cualquier $p(\omega|\psi)$.

Todas las densidades de la forma $\pi(\psi)\, p(\omega|\psi)$ serán llamadas ψ-difusas relativamente a E. Difieren en el tipo de opiniones que describen sobre el valor de ω dado ψ. Parece natural seleccionar como distribución inicial de referencia la más difusa entre ellas, i.e. la que maximiza cada una de las informaciones residuales desconocidas, $I^{\omega_i|\psi}\{E(\infty), p(\theta)\}$, $i = 1, \ldots, k$. La distribución posterior de ψ que se obtenga utilizando tal distribución inicial, será una descripción completa, argüimos, de las inferencias sobre ψ que pueden realizarse utilizando exclusivamente los resultados experimentales.

Generalizando ligeramente el trabajo de Stone (1958), se encuentra que si la distribución posterior de ψ es asintóticamente normal con precisión $nh(\hat{\theta})$, donde $\hat{\theta}$ es el estimador máximo-verosímil de θ, entonces para cualquier distribución inicial positiva (i.e. $p(\theta) > 0, \forall\theta$), y $n \to \infty$,

$$I^{\psi}\{E(n), p(\theta)\} = \frac{1}{2} \log \frac{n}{2\pi e} + \int p(\theta) \log \frac{h(\theta)^{1/2}}{p(\psi)} d\theta + o(1)$$

Se deduce de ello (Bernardo, 1976) que si $h(\theta)$ puede descomponerse en la forma

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

entonces la información desconocida sobre ψ, $I^{\psi}\{E(\infty), p(\theta)\}$ es maximizada para todo $p(\omega|\psi)$ cuando $p(\psi) = \pi(\psi)$.

Análogamente, si la distribución posterior de ω dado ψ es asintóticamente normal con precisión $nh_i(\hat{\theta})$ y

$$h_i(\theta)^{1/2} = \pi(\omega_i)\, g_i(\theta_i^*), \quad \theta_i^* = \theta - \{\omega_i\}$$

entonces la información residual desconocida es maximizada cuando $p(\omega_i|\psi) = \pi(\omega_i)$. En tal caso, la distribución inicial de referencia para realizar inferencias sobre ψ viene dada por $\pi(\psi) \prod_i \pi(\omega_i)$.

3. Distribución posterior del coeficiente de variación.

Considérese de nuevo el problema del coeficiente de variación. La función de verosimilitud del modelo normal puede escribirse, en función de $\theta = \{\psi, \sigma\}$, $\psi = \mu/\sigma$, como

$$p(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{(2\pi)}} \exp\left\{-\frac{1}{2\sigma^2}(x - \psi\sigma)^2\right\}$$

Se sabe (Walker, 1969) que la distribución posterior de es asintóticamente normal con matriz de precisión $nF(\hat{\theta})$, donde $\hat{\theta}$ es el estimador máximo-verosímil de θ y $F(\theta)$ la llamada matriz de información de Fisher, de elemento típico

$$-\int p(x|\theta)\,\frac{\partial^2}{\partial\theta_i\,\partial\theta_j}\log p(x|\theta)\,dx$$

En nuestro caso, es fácil comprobar que tal matriz resulta ser

$$F\begin{pmatrix}\psi\\ \sigma\end{pmatrix}=\begin{pmatrix}1 & \psi/\sigma\\ \psi/\sigma & (2+\psi^2)/\sigma^2\end{pmatrix}$$

y, en consecuencia, las distribuciones asintóticas posteriores de ψ y de σ condicionada a ψ son normales con precisiones respectivas $nh(\hat{\theta})$ y $nh_1(\hat{\theta})$ donde

$$h(\theta)=\left(1+\frac{\psi^2}{2}\right)^{-1}$$

$$h_1(\theta)=(2+\psi^2)\,\sigma^{-2}$$

Por tanto, haciendo uso de los resultados descritos en la sección anterior, las densidades ψ-difusas son aquéllas de la forma $p(\psi,\sigma)\propto \propto (1+\psi^2/2)^{-1/2}p(\sigma|\psi)$ y la distribución inicial de referencia para hacer inferencias sobre ψ es $\pi(\psi,\sigma)=(1+\psi^2/2)^{-1/2}\sigma^{-1}$ o, en términos de la métrica inicial,

$$\pi(\mu,\sigma)=\left(1+\frac{\mu^2}{2\sigma^2}\right)^{-1/2}\sigma^{-2} \qquad (3)$$

Utilizando (3) como distribución inicial, la distribución posterior de ψ resulta ser

$$\pi(\psi|z)\propto\left(1+\frac{\psi^2}{2}\right)^{-1/2}\exp\left\{-\frac{n\psi^2}{2}\right\}$$

$$\times\int_0^\infty t^{n-1}\exp\left\{-\frac{1}{2}t^2+r\psi t\right\}dt \qquad (4)$$

Como habíamos anticipado, (4) y (2) son proporcionales como funciones de ψ, siendo $\pi(\psi)=(1+\psi^2/2)^{-1/2}$ el correspondiente factor de proporcionalidad Consecuentemente, la distribución final (4), puede reproducirse a partir de (2) utilizando $\pi(\psi)$ como distribución inicial "no-informativa" y, por tanto, la paradoja de marginalización ha sido resuelta

Es interesante observar que, para valores grandes de ψ, $(1+\psi^2/2)^{-1/2}$ se comporta como $|\psi|^{-1}$, mostrando una conexión con el tratamiento de Stone & Dawid (1972)

4. Discusión

La distribución posterior (4) debe entenderse como una distribución de referencia, que describe el tipo de inferencias que pueden hacerse sobre ψ basándose *exclusivamente* en los resultados experimentales

En cualquier problema real, se dispone de cierta información inicial sobre los valores de μ y σ Tal información debe traducirse en una distribución inicial que, combinada con la información proporcionada por los datos experimentales mediante el teorema de Bayes, dé lugar a la correspondiente distribución posterior Esta distribución posterior, $p(\psi|z)$, que describe las conclusiones finales sobre el valor de ψ, puede entonces compararse con (4) obteniéndose de ello la necesaria información sobre la importancia relativa de los datos experimentales y la información inicial en las conclusiones finales

5. Referencias

BARNARD, G A (1952) The frequency justification of certain sequential tests *Biometrika 39*, 144-50

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

DAWID, A P, STONE, M & ZIDEK, J V (1973) Marginalization Paradoxes in Bayesian and Structural Inference *J Roy Statist Soc Ser B 35*, 189-233 (Con discusión)

JEFFREYS, H. (1939/61). *Theory of Probability*. Oxford: Clarendon Press.

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

STONE, M. (1958). *Studies with a Measure of Information*. Tesis Doctoral. Universidad de Cambridge.

STONE, M. & DAWID, A. P. (1972). Un-Bayesian implications of improper Bayesian inference in routine statistical problems. *Biotrika 59*, 369-75.

WALKER, A. M. (1969). On the asymptotic behaviour of posterior distributions. *J. Roy. Statist. Soc. Ser. B 31*, 80-8.