

## UNA SOLUCION BAYESIANA A LA PARADOJA DE STEIN

*J. R. Ferrándiz*  
*Departamento de Bioestadística*  
*Universidad de Valencia*

### SUMMARY

If we are interested in making inferences about the square norm of the mean in a multivariate normal model, the usual uniform prior for the mean is not sound, as revealed by Stein in his 1959 work.

This paper studies in what sense this prior must be modified by using the maximization of missing information procedure (Bernardo, 1979).

*Key Words:* Bayesian Inference; Stein's Paradox; Non-informative priors; Amount of Information.

### RESUMEN

Stein (1959), en su artículo sobre la gran discrepancia entre intervalos de confianza e intervalos fiduciales, puso de relieve el comportamiento poco convincente de la distribución inicial uniforme para el vector de medias de una normal multivariante si nuestro interés se centra en hacer inferencias sobre el cuadrado de su norma.

En este artículo, utilizando el método de maximización de la información desconocida (Bernardo, 1979), se estudia en qué sentido la distribución inicial del vector de medias debe ser "corregida" mediante la obtención de la distribución inicial mínimo-informativa para el cuadrado de su norma y se exponen los resultados a que dicha corrección conduce.

*Palabras Clave:* Inferencia Bayesiana; Paradoja de Stein; Distribuciones no Informativas; Cantidad de Información.

*Clasificación AMS (1980)*

Primaria: 62F15

Secundaris: 62A15, 62B10, 62F25

## 1. INTRODUCCION

Comenzaré haciendo una breve descripción de la paradoja de Stein (1959). Supongamos el vector aleatorio  $X \in R^k$  distribuido normalmente con vector de medias  $\mu \in R^k$  y matriz de dispersión identidad  $I$  i.e.

$$X \sim N_k (X | \mu, I) \quad (1.1)$$

Obviamente  $X$  es un estadístico suficiente para el parámetro  $k$ -dimensional  $\mu$ .

En el caso de parámetros multidimensionales, si la verosimilitud puede factorizarse en funciones que involucren cada una a una sola componente del parámetro y del estadístico suficiente, el método fiducial especifica que las inferencias a deducir del valor observado del estadístico son independientes para cada componente del parámetro (Kendall & Stuart, 1968, pp. 154-155, con sus referencias a los textos de Fisher y Quenouille).

Siendo la distribución fiducial para la media de una normal univariante de varianza unidad a su vez una normal univariante con la misma varianza centrada en el valor muestral observado (ibidem pp. 141-142) se deduce de lo anterior la siguiente distribución fiducial para  $\mu$  a partir de (1.1) y la observación  $X = x$ :

$$\mu \sim N_k (\mu | x, I) \quad (1.2)$$

Stein concluye pues que la distribución fiducial para  $\psi = \mu' \mu$  será entonces una Chi-cuadrado descentrada con  $k$  grados de libertad y parámetro de descentralización  $t = x' x$ , esto es

$$\psi \sim \chi^2 (\psi | k, t) \quad (1.3)$$

Sea ahora  $\mu_0$  el verdadero valor del parámetro  $\mu$ . Una interpretación libre de las ideas de Stein es la siguiente: podemos considerar la distribución de la variable aleatoria  $\psi$ , gobernada por  $T = X' X$ , cuya distribución a su vez depende de  $\psi_0 = \mu_0' \mu_0$ .

Como consecuencia del procedimiento seguido en la inferencia, la distribución fiducial de  $\psi$  sólo depende de  $\psi_0$  a través de  $T$  y se puede estudiar el comportamiento de  $p(\psi|T)$  conociendo el verdadero valor  $\psi_0$ .

Así tendremos:

$$X \sim N_k(X|\mu_0, I) \Rightarrow T = X'X \sim \chi^2(T|k, \psi_0 = \mu_0' \mu_0) \quad (1.4)$$

$$\mu \sim N_k(\mu|X, I) \Rightarrow \psi \sim \chi^2(\psi|k, T) \quad (1.5)$$

y en consecuencia

$$E[\psi|\mu_0] = E[k + T|\mu_0] = 2k + \psi_0 \quad (1.6)$$

$$E[V(\psi|T)|\mu_0] = E[2k + 4T|\mu_0] = 6k + 4\psi_0 \quad (1.7)$$

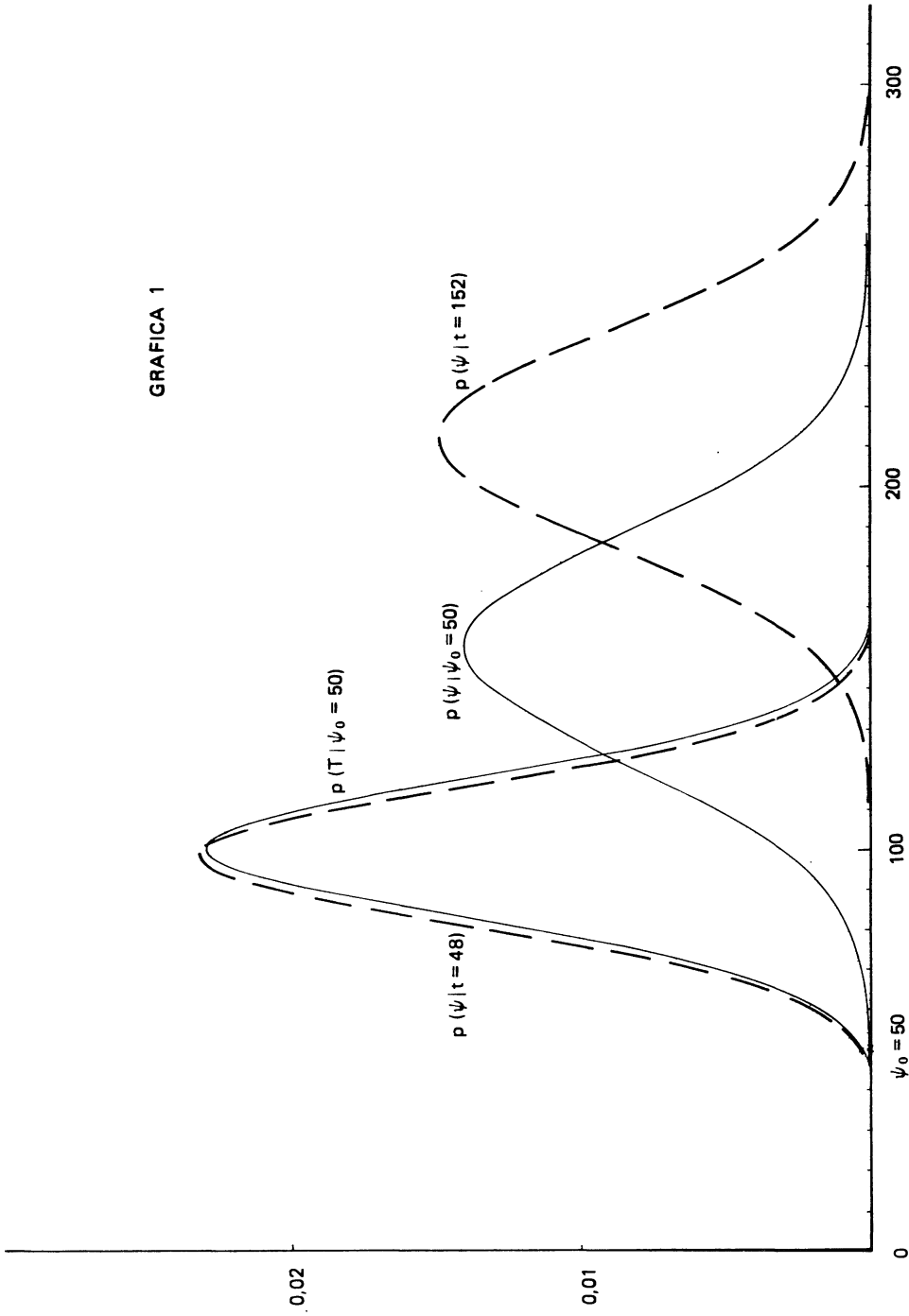
$$V[E(\psi|T)|\mu_0] = V[k + T|\mu_0] = V[T|\mu_0] = 2k + 4\psi_0 \quad (1.8)$$

$$V[\psi|\mu_0] = 8k + 8\psi_0 \quad (1.9)$$

Para  $k \uparrow \infty$  y si  $\psi_0$  es  $o(k^2)$  tenemos que la distribución fiducial para  $\psi$  cabe esperarla centrada en un valor  $\psi_0 + O(k)$  con desviación típica  $o(k)$  y por lo tanto “excesivamente” desplazada hacia los valores muy grandes de  $\psi$  con relación al verdadero valor  $\psi_0$ .

En la gráfica 1 se ejemplifica este fenómeno para  $k = 50$  y  $\psi_0 = 50$  aproximando las Chi-cuadrado descentradas por normales con los dos primeros momentos iguales. Incluso para este valor no muy grande de  $k$  se observa claramente como las distribuciones fiduciales de  $\psi$  quedarán, con gran probabilidad, desplazadas hacia la derecha del verdadero valor  $\psi_0$ .

Lo que C. Stein (1959) demuestra en su artículo es que, supuestas las condiciones ya mencionadas, la probabilidad “frecuencial” de que un intervalo  $[\psi_\alpha(T), \infty[$  de contenido probabilístico fiducial  $\alpha$  cualquiera, contenga al verdadero valor del parámetro  $\psi_0$  tiende a cero conforme  $k$  tiende a infinito.



GRAFICA 1

Dada la monotonicidad decreciente de la función de distribución  $F(\psi|T)$  con respecto a  $T$  (V. Johnson & Kotz, 1970, vol. 2, pp. 135), tenemos la relación:

$$T < T' \Rightarrow \psi_\alpha(T) < \psi_\alpha(T') \quad 0 < \alpha < 1 \quad (1.10)$$

y por tanto, si denominamos  $T_\alpha$  a aquel valor para el cual  $F(\psi = \psi_0 | T_\alpha) = 1 - \alpha$ , la probabilidad de que el intervalo  $[\psi_\alpha, \infty[$  contenga a  $\psi_0$  será  $F(T_\alpha | \psi_0)$ .

En nuestro ejemplo, y aproximando las Chi-cuadrado por normales, tenemos para  $\alpha = 0,99$ ,  $T_{0,99} \simeq 36,5$ , y la probabilidad de que  $[\psi_{0,99}, \infty[$  contenga a  $\psi_0 = 50$  será aproximadamente 0,00012.

La argumentación de Stein revela el “mal comportamiento” de la metodología fiducial para este problema en el sentido del principio débil de repetición muestral (Cox & Hinkley, 1974, pp. 45-46).

Pero la distribución (1.2) es también la que se obtiene utilizando la metodología Bayesiana con la distribución inicial impropia uniforme para  $\mu$ . Y es más, el comportamiento anómalo de dicha distribución final no se debe en este caso a la utilización de una inicial impropia. Con una inicial uniforme en una esfera  $k$ -dimensional centrada en el origen y radio grande, o incluso con una inicial normal de dispersión grande, se obtienen resultados similares (Cox & Hinkley, 1974, pp. 383-384).

Este artículo se plantea la necesidad de corregir la distribución inicial usual para  $\mu$  de modo que desaparezca la anomalía evidenciada por Stein. La técnica utilizada para cubrir dicho objetivo es el método de maximización de la información desconocida (Bernardo, 1979) cuya breve descripción y aplicación se encomienda al apartado 2.

En el apartado 3 se exponen los resultados a que nos conduce dicho método que son comentados en el apartado 4.

## 2. DISTRIBUCION INICIAL MINIMO INFORMATIVA

### 2.1. El método de maximización de la información desconocida MID

Bernardo (1979) propone un nuevo método de obtención de distribuciones iniciales a utilizar en aquellos casos en que no se desee introducir información inicial sobre el parámetro de interés más allá de la que implica la selección de una determinada clase  $C$  de distribuciones iniciales.

La idea básica del método consiste en la búsqueda de la distribución inicial en  $C$  que maximice la información esperada del experimento cuando el tamaño muestral crece indefinidamente, entendiendo por información esperada del experimento una generalización del concepto que propuso Lindley (1956).

Bernardo, utilizando un argumento heurístico, comprueba que su método reproduce la regla de Jeffreys para aquellos modelos que verifican ciertas condiciones de regularidad (esencialmente las mismas que se exigen para la consistencia y normalidad asintótica del estimador máximo verosímil, ver eg. Walker, 1969) siendo  $C$  una clase suficientemente amplia.

Para aquellos casos en que deseemos hacer inferencias solamente respecto de unos determinados componentes del vector de parámetros  $\theta \in \Theta \subset R^k$ , es decir, cuando consideramos la descomposición  $\theta' = (\psi', \omega')$ ,  $\psi \in \Phi \subset R^r$ ,  $\omega \in \Omega \subset R^{k-r}$ , de modo que  $\omega$  representa aquella parte del parámetro sobre la que no estamos interesados pero que no podemos eludir en el modelo muestral  $p(X|\theta)$ , Bernardo propone la aplicación del método M.I.D. al modelo reducido:

$$p(X|\psi) = \int_{\Omega} p(X|\theta) p(\omega|\psi) d\omega \quad (2.1.1)$$

donde  $p(\omega|\psi)$  es una densidad inicial que se puede escoger de modo que refleje la información inicial del investigador o bien puede ser determinada aplicando a su vez el método M.I.D. al modelo completo  $p(X|\omega, \psi)$  pero considerando conocido el valor de  $\psi$ .

En este último caso, y bajo las condiciones de regularidad mencionadas, su método le conduce a la siguiente regla:

$$\pi(\omega|\psi) \propto |J_{\omega, \omega}(\psi, \omega)|^{1/2} \quad (2.1.2)$$

$$\pi(\psi) \propto \exp\left(\int_{\Omega} \log(|H_{\psi, \psi}(\psi, \omega)|^{-1/2}) \pi(\omega|\psi) d\omega\right) \quad (2.1.3)$$

donde  $J_{\omega, \omega}(\psi, \omega)$  es la submatriz correspondiente a  $\omega$  en la matriz de información de Fisher para el modelo  $p(X|\theta)$  de elemento típico  $E[(\partial \log p(X|\theta)/\partial \theta_i)(\partial \log p(X|\theta)/\partial \theta_j)|\theta]$  y  $H_{\psi, \psi}(\psi, \omega)$  es la submatriz correspondiente a  $\psi$  en la inversa de la misma.

Si  $|J_{\omega, \omega}(\psi, \omega)|^{1/2}$  y  $|H_{\psi, \psi}(\psi, \omega)|^{-1/2}$  factorizan en la forma:

$$|J_{\omega, \omega}(\psi, \omega)|^{1/2} = h_1(\psi) h_2(\omega) \quad (2.1.4)$$

$$|H_{\psi, \psi}(\psi, \omega)|^{-1/2} = g_1(\psi) g_2(\omega) \quad (2.1.5)$$

las expresiones (2.1.2) y (2.1.3) se reducen a:

$$\pi(\omega|\psi) = \pi(\omega) \propto h_2(\omega) \quad (2.1.6)$$

$$\pi(\psi) \propto g_1(\psi) \quad (2.1.7)$$

## 2.2. Distribución inicial mínimo informativa para $\psi = \mu^T \mu$

La matriz de información de Fisher para  $X \sim N_k(X|\mu, I)$  es

$$J(\mu) = I \quad (2.2.1)$$

Si ahora hacemos la transformación a coordenadas esféricas de  $R^k \rightarrow (0, \infty) \times (0, \pi)^{k-2} \times (0, 2\pi)$ , que transforma  $\mu$  en  $\psi, \omega$ :

$$\mu = \sqrt{\psi} S(\omega) \quad (2.2.2)$$

donde  $\psi = \mu^T \mu$  y  $S(\omega)' = (\cos \omega_1, \sin \omega_1 \cos \omega_2, \dots, \sin \omega_1 \dots \sin \omega_{k-1})$ , tendremos entonces:

$$J(\psi, \omega) = (\partial(\mu)/\partial(\psi, \omega))' J(\mu) (\partial(\mu)/\partial(\psi, \omega)) \quad (2.2.3)$$

$$H(\psi, \omega) = (\partial(\psi, \omega)/\partial(\mu)) H(\mu) (\partial(\psi, \omega)/\partial(\mu))' \quad (2.2.4)$$

donde  $(\partial(\mu)/\partial(\psi, \omega))$  es la matriz del jacobiano de la transformación, de modo que:

$$|\partial(\mu)/\partial(\psi, \omega)| = \frac{1}{2} \psi^{k/2-1} \prod_{i=1}^{k-2} \text{sen}^{k-i-1} \omega_i \quad (2.2.5)$$

$$H_{\psi, \psi}(\psi, \omega) = 4 \psi \quad (2.2.6)$$

$$|J(\psi, \omega)| = \left[ \frac{1}{2} \psi^{k/2-1} \prod_{i=1}^{k-2} \text{sen}^{k-i-1} \omega_i \right]^2 \quad (2.2.7)$$

$$\begin{aligned} |J_{\omega, \omega}(\psi, \omega)| &= |J(\psi, \omega)| \cdot |H_{\psi, \psi}(\psi, \omega)| = \\ &= \psi^{k-1} \left[ \prod_{i=1}^{k-2} \text{sen}^{k-i-1} \omega_i \right]^2 \end{aligned} \quad (2.2.8)$$

utilizando en la obtención de (2.2.8) el resultado general para matrices definidas positivas particionadas,

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = B^{-1} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}^{-1} \Rightarrow |A| = |A_{22}| |B_{11}|^{-1} \quad (2.2.9)$$

(Anderson, 1974, pp. 344)

Aplicando ahora (2.1.6) y (2.1.7) resulta)

$$\pi(\omega | \psi) \propto \prod_{i=1}^{k-2} \text{sen}^{k-i-1} \omega_i \quad (2.2.10)$$

$$\pi(\psi) \propto \psi^{-1/2} \quad (2.2.11)$$



### 3. DISTRIBUCION FINAL DE REFERENCIA PARA $\psi = \mu' \mu$

Antes de proseguir necesitamos el resultado previo que paso a enunciar:

*LEMA.* Sean  $L \in R^k$ ,  $S(\omega)' = (\cos \omega_1, \text{sen } \omega_1 \cos \omega_2, \dots, \text{sen } \omega_1 \dots \text{sen } \omega_{k-1})$  donde  $\omega \in \Omega = (0, \pi)^{k-2} \times (0, 2\pi)$ , entonces

$$\int_{\Omega} \exp \{L' S(\omega)\} \prod_{i=1}^{k-2} \text{sen}^{k-i-1} \omega_i d\omega =$$

$$= 2 \pi^{k/2} \sum_{i=0}^{\infty} (L' L/4)^i / (i! \Gamma(k/2 + i)) \quad (3.1)$$

La comprobación de la expresión (3.1) es sencilla considerando el vector aleatorio  $U \sim N_k(U|m, I)$  y realizando una transformación a coordenadas esféricas similar a (2.2.2)  $U = \sqrt{r} S(\omega)$ .

La densidad marginal para  $r = U' U$  será:

$$(2\pi)^{-k/2} (1/2) r^{k/2-1} \exp \{-(m' m + r)/2\}$$

$$\int_{\Omega} \exp \{\sqrt{r} m' S(\omega)\} \prod_{i=1}^{k-2} \text{sen}^{k-i-1} \omega_i d\omega$$

que debe coincidir con la  $\chi^2(r|k, m' m)$ . Llamando  $L = \sqrt{r} m$  se obtiene (3.1).

La distribución final de referencia para  $\psi$  a partir de la inicial impropia definida por (2.2.10) - (2.2.11); y de la verosimilitud para una muestra de tamaño  $n$  del modelo  $X \sim N_k(X|\mu, I)$  será la marginal de:

$$\pi(\psi, \omega | x_1 \dots x_n) \propto \psi^{-1/2} \exp \{-(n\psi + n t)/2 + n \sqrt{\psi} \bar{x}' S(\omega)\}$$

$$\prod_{i=1}^{k-2} \text{sen}^{k-i-1} \omega_i \quad (3.2)$$

donde  $\bar{x} = n^{-1} \sum_{i=1}^n x_i$  y  $t = \bar{x}' \bar{x}$

Utilizando (3.1):

$$\pi(\psi | x_1 \dots x_n) \propto \psi^{-1/2} \exp\{-n\psi/2\} \sum_{i=0}^{\infty} (n^2 \psi t/4)^i / (i! \Gamma(k/2 + i)) \quad (3.3)$$

que es precisamente la que obtiene Bernardo (1979) a partir del modelo  $\mathcal{X}^2(n | k, n \psi)$ . Merece la pena destacar este hecho por cuanto implica que, en este caso, el método M.I.D. evita la paradoja de marginalización (Dawid, Stone & Zidek, 1973) que podría resultar al ser  $p(t | \psi, \omega) = p(t | \psi)$ . Por tanto, la distribución inicial (2.2.10) es no informativa respecto de  $\psi$  en el sentido que le da Jaynes (1980).

Dado el radio de convergencia infinito de la serie en (3.3), resulta fácil la expresión de las características de la distribución final de referencia al ser lícito el intercambio de los signos de sumación e integración. Utilizando la nomenclatura de Abramowitz & Stegun (1972, p. 504) para las funciones hipergeométricas, introducimos:

$$M(a, b, z) = (\Gamma(b)/\Gamma(a)) \sum_{i=0}^{\infty} \Gamma(a+i) z^i / (i! \Gamma(b+i)) \quad (3.4)$$

con lo que la constante de proporcionalidad para (3.3) resulta ser:

$$C(t) = \sqrt{2\pi/n} M(1/2, k/2, n t/2) / \Gamma(k/2) \quad (3.5)$$

La función generatriz y los momentos:

$$E[e^{r\psi} | t] = \sqrt{n/(n-2r)} M(1/2, k/2, n t/(2(n-2r))) / M(1/2, k/2, n t/2) \quad (3.6)$$

$$E[\psi^r | t] = (2/n)^{r+1/2} \Gamma(r+1/2) M(r+1/2, k/2, n t/2) / (\Gamma(k/2) C(t)) \quad (3.7)$$

y la función de distribución:

$$F(\psi|t) = \sqrt{2/n} \sum_{i=0}^{\infty} A_i(t, k) \gamma(i + 1/2, n \psi/2)/C(t) \quad (3.8)$$

donde:

$$A_i(t, k) = (n t/2)^i / (i! \Gamma(k/2 + i)) \quad (3.9)$$

$$\gamma(a, x) = \int_0^x y^{a-1} \exp(-y) dy \quad (3.10)$$

Resulta fácil comprobar que  $F(\psi|t)$  es monótona decreciente con  $t$  como consecuencia de que el signo de  $\partial F(\psi|t)/\partial t$  es el de la expresión:

$$\begin{aligned} & \sum_{j \geq 1} A_0 A_j \Gamma(1/2) \gamma(j + 1/2, n \psi/2) - \\ & - \sum_{i \geq 1} A_i^2 A_0 \Gamma(i + 1/2) \gamma(1/2, n \psi/2) \end{aligned} \quad (3.11)$$

donde el acento representa derivación respecto a  $t$ .

La negatividad de (3.11) es consecuencia de ser el minuendo menor que el sustraendo término a término pues:

$$\gamma(j + 1/2, n \psi/2)/\Gamma(j + 1/2) < \gamma(1/2, n \psi/2)/\Gamma(1/2) \quad \psi > 0 \quad (3.12)$$

que se puede deducir fácilmente a partir de la relación de recurrencia:

$$\gamma(a + 1, x) = a \gamma(a, x) - x^a e^{-x} \quad (3.13)$$

(Abramowitz & Stegun, 1972, p. 262).

Para el ejemplo planteado en el apartado 2 obtenemos los resultados de la gráfica 2, contrarréplica de la gráfica 1. En ella podemos observar que  $E[\pi(\psi|T)|\psi_0] = \pi(\psi|\psi_0)$  está centrada en  $\psi_0$  con  $E(\psi|\psi_0) = 50,7$ ,  $V[E(\psi|T)|\psi_0] = 317$ ,  $E[V(\psi|T)|\psi_0] = 307$  y  $V(\psi|\psi_0) = 624$ , valores obtenidos por procedimientos numéricos.

Por último, la comparación del contenido probabilístico de los intervalos  $[\psi_\alpha(T), \infty[$  con la probabilidad “frecuencial” de que dichos intervalos contengan al verdadero valor del parámetro se reduce, de hecho, dada la monotonicidad de  $F(\psi|t)$  respecto de  $t$  y por tanto la validez de (1.10), a la comparación de  $1 - F(\psi|t)$  con  $F(t|\psi)$ . La tabla 1 recoge los resultados numéricos para unos cuantos valores de  $\psi$  con  $t = 100 = E(T|\psi_0 = 50)$ .

Los valores en la primera columna representan el contenido probabilístico  $\alpha$  a posteriori del intervalo  $[\psi, \infty[$  dada la observación  $t = 100$ .

Los valores en la segunda columna representan la probabilidad “frecuencial” de que el intervalo  $[\psi_\alpha, \infty[$  contenga al verdadero valor supuesto igual a  $\psi$ .

TABLA 1

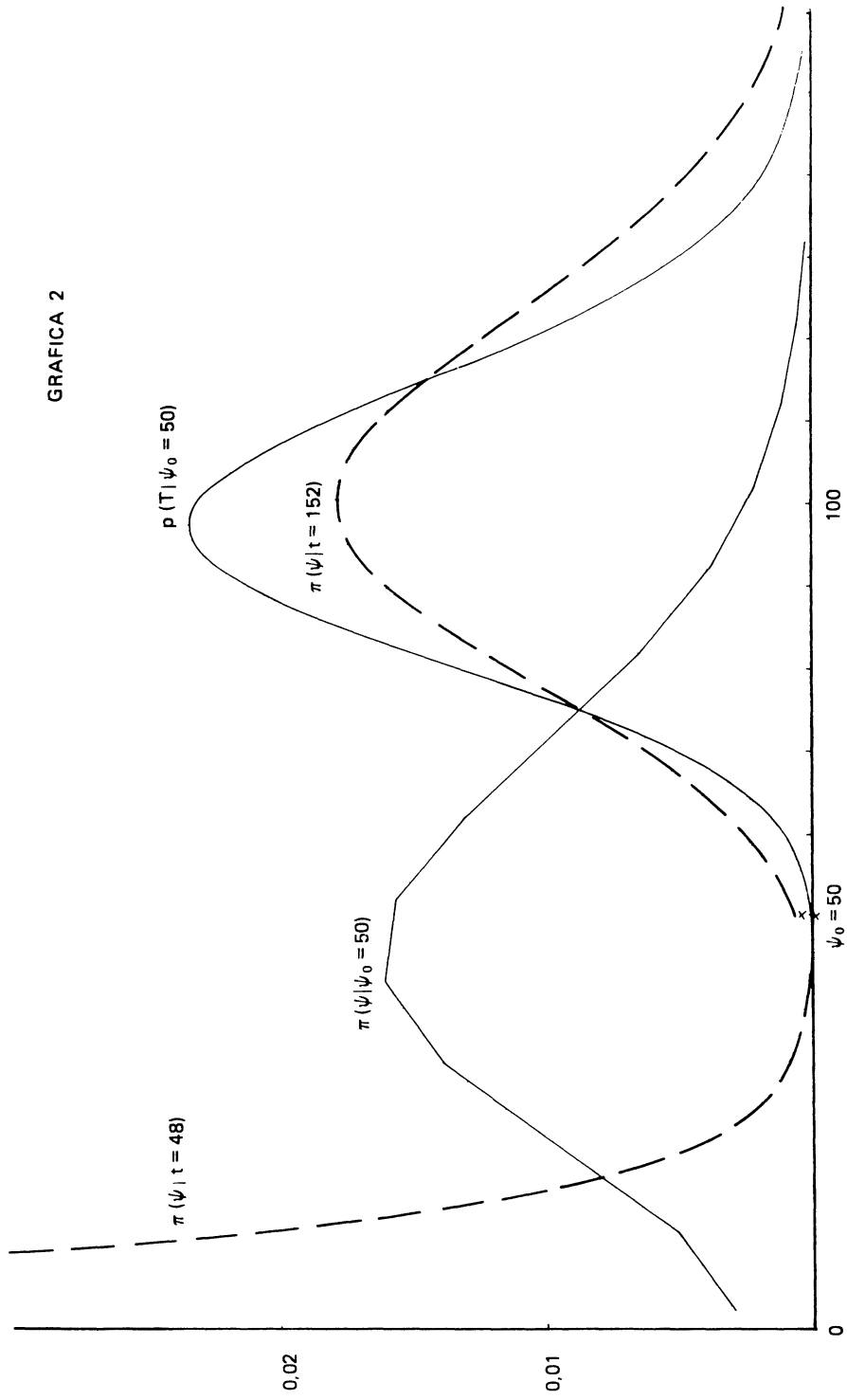
$\psi$	$1 - F(\psi t = 100)$	$F(T = 100 \psi)$
10	0,997	0,998
20	0,974	0,980
30	0,890	0,905
40	0,720	0,743
50	0,495	0,520

Así, suponiendo  $\psi_0 = 50$ , observamos en la tabla 1 que el intervalo  $[\psi_{0,495}, \infty[$  tiene una probabilidad “frecuencial” 0,52 de contener al verdadero valor. Si queremos comparar directamente el resultado del ejemplo en el apartado 2 tenemos que alcanzar  $t = 144,9$  para que  $1 - F(\psi = 50|t) = 0,9900$ , con lo que  $F(T = 144,9|\psi_0 = 50) = 0,9909$ . Es decir, si  $\psi_0 = 50$ , la probabilidad “frecuencial” de que el intervalo a posteriori de la forma  $[\psi_{0,99}(T), \infty[$  cubra al verdadero valor es aproximadamente 0,99.

#### 4. CONCLUSIONES

A la vista de los resultados del apartado anterior cabría decir que la distribución inicial propuesta en (2.2.10) - (2.2.11) “corrige” el com-

GRAFICA 2



portamiento poco convincente de la distribución (2.2). En realidad, la distribución inicial propuesta resulta ser la uniforme para  $\sqrt{\psi}$ , i.e., para la distancia euclídea al origen del parámetro  $\mu$ . Así, la distribución final que se obtiene no es más que la verosimilitud de  $\sqrt{\psi}$  normalizada, por lo que el comportamiento de la misma será tanto mejor cuanto mayor sea el tamaño muestral  $n$ .

De ningún modo se pretende reivindicar que dicha distribución sea “la” distribución inicial que resuelve la paradoja de Stein. La conclusión a extraer de los resultados anteriores es que el método M.I.D. proporciona una distribución inicial que es rápidamente dominada por la información experimental, de modo que la distribución final a que conduce puede servir de patrón de comparación para estudiar la importancia de la información que contienen otras distribuciones iniciales, y que, por ejemplo, en el caso de la uniforme para  $\mu$  resulta ser determinante cuando el tamaño muestral  $n$  es pequeño en comparación con la dimensión  $k$  del modelo.

No se trata aquí de ningún “principio de no coherencia”, como el propuesto por Wilkinson (1977) desde una óptica fiducial para soslayar la paradoja de Stein, que nos obligue a cambiar la distribución inicial según la transformación del parámetro en que estemos interesados. Se trata simplemente de una ayuda a la hora de comprender las implicaciones contenidas en la selección de determinadas distribuciones iniciales, que oriente al decisor en la especificación de su propia distribución inicial.

Por otro lado es importante destacar que, en este caso, el método propuesto por Bernardo evita la paradoja de marginalización que podría haberse producido dadas las características del modelo. Resulta verdaderamente interesante comprobar que la distribución (2.2.10) para los parámetros marginales es no informativa para el parámetro de interés en el sentido indicado por Jaynes (1980). La investigación de la relación entre el método de maximización de la información desconocida y las paradojas de marginalización promete ser fructífera a la luz de las ideas expuestas por dicho autor.

Finalmente, no hay que olvidar que la versión “clásica” de la paradoja de Stein estudiada en este artículo se concreta en la inadmisibilidad del estimador media muestral en el modelo normal multivariante con

función de pérdida cuadrática (Stein, 1956). En la interpretación bayesiana que el propio Stein (1962) hace de sus estimadores de simetría esférica, considera una distribución inicial para  $\mu' \mu$  muy parecida a (2.2.11), y él mismo reconoce que ésta última, que le fue sugerida por I.J. Good, debe conducir a resultados cualitativamente similares. El desarrollo de este tema será objeto de un próximo artículo.

## REFERENCIAS

- ABRAMOWITZ, M. & STEGUN, I.A. (1965/72). *Handbook of Mathematical Functions*. New York: Dover.
- ANDERSON, T.W. (1958/74). *An Introduction to Multivariate Statistical Analysis*. New York: John Wiley & Sons.
- BERNARDO, J.M. (1979). "Reference Posterior Distributions for Bayesian Inference" (with Discussion). *J. Roy. Statist. Soc., Ser B*, 41, 113-147.
- COX, D.R. & HINKLEY, D.V. (1974). *Theoretical Statistics*. London: Chapman & Hall.
- 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.
- JAYNES, E.T. (1980). "Marginalization and Prior Probabilities", en *Bayesian Analysis in Econometrics and Statistics* (Zellner A. ed.). Amsterdam: North Holland.
- JOHNSON, N.L. & KOTZ, S. (1970). *Continuous Univariate Distributions. Vol. 2*. New York: Houghton Mifflin.
- KENDALL, M.G. & STUART, A. (1963/68). *The Advanced Theory of Statistics, vol. 2*. London: Charles Griffin & Co.
- LINDLEY, D.V. (1956). "On a Measure of Information Provided by an Experiment". *Ann. Math. Statist.*, 27, 986-1005.
- STEIN, C. (1956). "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution", en *Proc. III Berkeley Symp. I*, 197-206.
- STEIN, C. (1959). "An Example of Wide Discrepancy Between Fiducial and Confidence Interval". *Ann. Math. Statist.*, 30, 877-880.
- STEIN, C. (1962). "Confidence Stes for the Mean of a Multivariate Normal Distribution". *J. Roy. Statist. Soc., Ser. B*, 24, 265-296.

- WALKER, D.M. (1969). "On the Asymptotic Behaviour of a Posterior Distribution".  
*J. Roy. Statist. Soc., Ser. B*, 31, 80-88.
- WILKINSON, G.N. (1977). "On resolving the Controversy in Statistical Inference".  
*J. Roy. Statist. Soc., Ser. B*, 39, 119-171.