ESTADÍSTICA ESPAÑOLA
Vol. 39, Núm. 142, 1997, págs. 99 a 128

Funciones a priori imparciales undimensionales

por
JESÚS BASULTO SANTOS
Departamento de Economía Aplicada I
Universidad de Sevilla

RESUMEN

En el presente trabajo introducimos el concepto de función a priori imparcial para un parámetro unidimensional. Las funciones a priori imparciales permiten dar una nueva interpretación de la función a priori de Jeffreys y, además, pueden aplicarse a un mayor conjunto de modelos. Hemos aplicado el concepto a varias familias de modelos, obteniendo la correspondiente función a priori imparcial. Por último, utilizamos el concepto de función a priori imparcial como un método alternativo para obtener intervalos de confianza clásicos.

Palabras Claves: estimación bayesiana, funciones a priori imparciales, intervalos de confianza, regla de Jeffreys.

Clasificación AMS: 62F15, 62G15.

1. INTRODUCCIÓN

La aproximación bayesiana a la estimación de un parámetro $\theta \in \Theta \subset R$, correspondiente a un modelo con distribución de probabilidad $f(x|\theta)$, debe seleccionar una función positiva, $\pi(\theta)$, sobre el espacio paramétrico Θ, para obtener vía el teorema de Bayes la distribución a posteriori $\pi(\theta|x) \propto \pi(\theta)L(\theta|x)$, donde $L(\theta|x)$ es la función de verosimilitud para una muestra **x** de tamaño n.

La regla general de selección más conocida para elegir la función $\pi(\theta)$ es la de Jeffreys. Dicha regla general toma como función $\pi(\theta)$ la expresión siguiente $\pi(\theta) \propto \sqrt{I(\theta)}$,donde $I(\theta) = E\left[-\partial^2 L(\theta|x) / \partial\theta^2\right]$ es la información de Fisher y $l(\theta|x)$ es el logaritmo natural de la función de verosimilitud para una muestra x de tamaño unidad.

La regla general de Jeffreys presenta las siguientes ventajas: (1) es invariante a transformaciones biyectivas, con derivada continua, del parámetro θ; y (2) si $Pr[\theta \leq h(\alpha,x) | x] = \alpha$ es un intervalo unilateral calculado en la distribución a posteriori $\pi(\theta|x)$ con $\pi(\theta) \propto \sqrt{I(\theta)}$, entonces, bajo ciertas condiciones, puede verse (B. L. Welch and H. W. Peers,1963) que $Pr[\theta \leq h(\alpha,x) \mid \theta] = \alpha + O(1/n)$.

Los dos puntos anteriores son importantes debido a que el (1) evita el tener que seleccionar una parametrización específica para un modelo, mientras que el punto (2) se refiere al interés perseguido por Fisher en la construcción de la denominada inferencia fiducial (Zabell,1994). Este último punto contribuye a explicar el motivo de considerar a $\pi(\theta) \propto \sqrt{I(\theta)}$ como una distribución a priori objetiva.

Frente a las ventajas anteriores, la regla general de Jeffreys presenta los siguientes problemas: (i) no es aplicable a modelos no regulares, por ejemplo, el modelo uniforme $(0,\theta)$, con $\theta>0$; (ii) no es aplicable a modelos con parámetro discreto; (iii) no asegura que la distribución a posteriori sea propia cuando $\pi(\theta)$ es impropia; y (iv) la regla general puede conducir, en el caso de tener datos no independientes, a distribuciones a priori que dependan del tamaño muestral.

Para una revisión de las reglas que se utilizan para seleccionar distribuciones a priori, puede verse el trabajo de Kass, R. E. y Wasserman, L. (1995), donde se recogen, además, otras reglas propuestas por Jeffreys.

En el presente trabajo se retoma la regla propuesta en (Basulto,1995) para seleccionar funciones positivas a priori, se relaciona dicha propuesta con la regla de Jeffreys y se estudia la propuesta en relación a los problemas planteados más arriba.

A partir de aquí, el trabajo consiste en la sección 2, donde definimos la regla para obtener distribuciones a priori imparciales, en la sección 3, donde se relaciona nuestra regla con la regla general de Jeffreys y se obtienen para varios modelos generales, en la sección 4, donde se relacionan las funciones a priori imparciales con los intervalos de confianza clásicos y, por último, en la sección 5 se discute el contenido del presente trabajo.

2. CONCEPTOS Y DEFINICIONES

En la presente sección vamos a introducir el concepto de función a priori imparcial. Este concepto está basado en la teoría, de las distribuciones afectadas (weighted distributions), que fue introducida por Fisher (1934) y, posteriormente, desarrollada por Rao (1965,1994). A continuación recogemos la definición y ciertos resultados de la teoría de las distribuciones afectadas que utilizaremos en el desarrollo de las funciones a priori imparciales.

Sea X una variable aleatoria con función de densidad f(x), donde $X \in \Omega \subset R$. Supongamos que para todo valor x de la variable X, la probabilidad de que x sea **registrado** es $w(x,\lambda)$, donde $w(x,\lambda)$ es denominada función peso y verifica $0 \leq w(x,\lambda) \leq 1$, siendo λ un miembro de un cierto conjunto Λ. Diremos que un valor x de la variable X es **registrado** si $u \leq w(x,\lambda)$, donde u es una realización de una variable aleatoria uniforme (0,1) que es independiente de la variable X. Se prueba, que la función de densidad de la variable aleatoria, X^{λ}, de los valores registrados, es

$$f(x / \lambda) = \frac{f(x)w(x,\lambda)}{\int_{\Omega} f(x)w(x,\lambda)dx} \qquad [1]$$

donde la función $f(x/\lambda)$ es denominada **distribución afectada** de f(x) por la función peso $w(x,\lambda)$. El **grado de concordancia** de la variable aleatoria X con la función peso $w(x,\lambda)$, $GC(X, \lambda)$, es medido por la integral que aparece en el denominador de [1], es decir

$$GC(X,\lambda) = \int_{\Omega} f(x)w(x,\lambda)dx \qquad [2]$$

El grado de concordancia mide la probabilidad de registrar **al menos** un valor de la variable X. Más concretamente, si **x** es una muestra aleatoria de tamaño n de la variable aleatoria X, entonces el producto, $\{n.GC(X, \lambda)\}$, es el número esperado de valores registrados en la muestra.

Con las fórmulas [1] y [2] ya podemos pasar a exponer los conceptos de funciones a priori imparciales.

Sea $\{f(x/\theta), \theta \in \Theta \subset R\}$ una familia de funciones de densidad que depende de un parámetro. Si **x** es una muestra aleatoria de tamaño n, seleccionada de uno de los modelos de la familia, entonces la función de verosimilitud es

$$L(\theta / x) \propto \prod_{i=1}^{n} f(x_i / \theta)$$

Vamos a suponer que a partir de un cierto tamaño muestral n, la función de verosimilitud está acotada, es decir, $L(\theta/x) \leq K(x)$; simplificando supondremos que existe el estimador máximo verosímil $\hat{\theta}_n$, es decir, $L(\theta/x) \leq L(\hat{\theta}_n/x)$ para todo $\theta \in \Theta$.

Sea ahora la siguiente familia de funciones pesos

$$w(\theta, x) = \frac{l(\theta / x)}{L(\hat{\theta}_n / x)} \tag{3}$$

donde **x** será normalmente un miembro del conjunto de todas las muestras aleatorias de tamaño n, M_n, o de un cierto subconjunto de M_n.

Si un investigador propone una función a priori, $\pi(\theta)$, para el parámetro $\theta \in \Theta$, y si para la muestra **x** la función peso es [3], entonces la distribución afectada de $\pi(\theta)$, es, por [1], igual a la distribución a posteriori, es decir, $\pi(\theta/x) \propto \pi(\theta).L(\theta/x)$.

Para la distribución a priori $\pi(\theta)$ el grado de concordancia, es

$$GC(\pi(\theta), x) = \int_{\Theta} \pi(\theta) \frac{L(\theta / x)}{L(\hat{\theta}_n / x)} d\theta \tag{4}$$

que a partir de [1] y [2] podemos deducir las dos siguientes conclusiones: a) todo valor θ_1 seleccionado de forma aleatoria de $\pi(\theta)$ que sea **registrado** es también una realización aleatoria de $\pi(\theta/x)$ y b) si $\{\theta_1, \theta_2, \ldots, \theta_n\}$ es una muestra aleatoria de $\pi(\theta)$, entonces el producto de [4] por n, mide el número esperado de estos valores que son realizaciones aleatorias de $\pi(\theta/x)$.

El grado de concordancia nos permite comparar diferentes funciones a priori. Por ejemplo, si un investigador selecciona la función $\pi_1(\theta)$ y resulta que para una muestra **x** de tamaño n, $GC(\pi_1(\theta),x)=0.5$, mientras que otro investigador selecciona $\pi_2(\theta)$, y resulta $GC(\pi_2(\theta),x)=0.3$, entonces podemos afirmar que para el primer investigador, su distribución está más de acuerdo con **x** que la distribución del otro investigador. Ahora bien, puede ocurrir que para otra muestra aleatoria **x·**, $GC(\pi_1(\theta), x\cdot)=0.56$ y

$GC(\pi_2(\theta), x\cdot)=0.6$, con lo que el segundo investigador está más de acuerdo con **x·** que el primero. Estos cambios en los grados de concordancia, que ocurren como consecuencia de observar diferentes muestras de tamaño n, impiden que podamos encontrar una función a priori con máximo grado de concordancia. En cambio, es razonable buscar funciones a priori que mantenga constante los grados de concordancia a cambios en las muestras **x** de tamaño n, y considerar que tales funciones se comportan de forma imparcial antes dichas muestras.

Una función positiva $\pi^*(\theta)$ que verifique

$$GC(\pi^*(\theta),x) = k(n) \qquad [5]$$

para todas las muestras $x \in M_n$ o para un cierto subconjunto de M_n, diremos que es una **función a priori imparcial**. En [5], k(n) es una constante finita que puede depender de n.

La función a priori imparcial, $\pi^*(\theta)$, suele ser impropia, ahora bien, al exigir que k(n) sea finita, conduce a una distribución a **posteriori** $\pi(\theta/x) \propto \pi(\theta).L(\theta/x)$ propia. Cuando $\pi^*(\theta)$ es impropia, la función a priori imparcial está determinada salvo una constante de proporcionalidad.

Vamos a ilustrar la definición [5] por medio del ejemplo siguiente.

Ejemplo 1

Supongamos que $f(x/\theta) = 1/\theta$ para $0<x\leq\theta$, $\theta>0$, y es nulo en otro caso. Para una muestra $x= \{x_1,x_2,\ldots,x_n\}$ de tamaño n, la función de verosimilitud es $L(\theta/\mathbf{x}) \propto 1/\theta^n$ para $\theta \geq y_n$, y nula en otro caso, donde $y_n=\text{máximo}\{x_1,x_2,\ldots,x_n\}$. Vamos a comprobar que la función $\pi^*(\theta)=1/\theta$ es una función a priori imparcial. En efecto:

$$GC\left(\frac{1}{\theta},x\right) = \int_{y_n}^{\infty} \theta^{-1} \frac{\theta^{-n}}{y_n^{-n}} d\theta = 1/n$$

para todo $\mathbf{x} \in M_n$.

Cuando la función de verosimilitud es invariante a translaciones que ocasionan las diferentes muestras de tamaño n, es decir, $L(\theta / x) = h(g(\theta) - t(x))$, Box-Tiao (1973) proponen la función a priori uniforme $\pi(\gamma)=1$ para el parámetro $\gamma=g(\theta)$. Box-Tiao suponen, además, que el espacio muestral es independiente del parámetro θ y que la familia de densidades admiten un estadístico suficiente, t(x). Un ejemplo de familia que es transladable por los datos es la función de densidad exponencial, es decir, $L(\theta / x) = \frac{r}{\theta} e^{-\frac{r}{\theta}}$, donde $r = \sum_{i=1}^{n} x_i$, y que puede ser escrita cómo

$$L(\theta / x) = \exp[(\log(r) - \log(\theta)) - \exp(\log(r) - \log(\theta))]$$

resultando una familia transladable.

Cuando la función de verosimilitud es transladable, entonces la cota de la función de verosimilitud, k(x), es decir $k(x) \geq L(\gamma/x)$, es constante es muestras de tamaño n, y además las funciones $L(\gamma/x)$ y $L(\gamma/x')$,donde **x** y **x'** son dos muestras de tamaño n diferentes, verifican que $L(\gamma/x)=L(\gamma'/x')$, siendo $\gamma'=\gamma+(t(x')-t(x))$. Como consecuencia de estos resultados, y si suponemos que existe la integral de la función de verosimilitud, entonces la función $\pi(\gamma)=1$ es imparcial al ser constante el grado de concordancia, es decir, GC(1,x) es constante para todas las muestras de M_n.

La definición [5] no es útil en muchas aplicaciones, debido a que no siempre existe una función a priori imparcial para cualquier tamaño muestral; una definición más débil que [5] es la siguiente: una función positiva $\pi^*(\theta)$ es una función a priori imparcial si

$$g(n).GC(\pi^*(\theta),x) \qquad [6]$$

converge en probabilidad a una cantidad c>0 , para todo $\theta \in \Theta$. La función g(n) suele ser g(n)=n o $g(n)=\sqrt{n}$. Si [6] es cierto, entonces para n grande se cumple que $GC(\pi^*(\theta),x) \approx c/g(n)$.

Vamos a ilustrar la definición [6] por medio del ejemplo siguiente.

Ejemplo 2

Sea la familia de Poisson $f(x /\theta) = \frac{\theta^x e^{-\theta}}{x!}$, $\theta>0$, x= 0, 1, 2,... . Para una muestra $x= \{x_1,x_2,\ldots,x_n\}$ de tamaño n, la función de verosimilitud es $L(\theta/x) \propto \theta^r e^{-n\theta}$, donde $r=\sum_{i=1}^{n} x_i$ y $\hat{\theta}_n=r/n$ es el estimador máximo verosímil. Vamos a comprobar que la función $\pi^*(\theta)=1/\sqrt{\theta}$ es una función a priori imparcial. En efecto:

$$GC\left(\frac{1}{\sqrt{\theta}}, x\right) = \int_0^{\infty} \theta^{-1/2} \frac{\theta^r e^{-n\theta}}{e^{-r}(r / n)^r} d\theta = e^r (r / n)^{-r} \frac{\Gamma(r + 0.5)}{n^{r+0.5}}$$

y por la aproximación de Stirling, $\Gamma(r + 0.5) \approx \sqrt{2\pi}\, e^{-(r+0.5)}(r + 0.5)^r$, $\sqrt{(n)}.GC(\pi^{*}(\theta),x)$ converge en probabilidad a $\sqrt{(2\pi/e)}$. Esto es también consecuencia de que $Pr[r+0.5 > k]$ converge en probabilidad a la unidad para todo $k > 0$.

Para la familia exponencial

$$f(x / \theta) = h(x)w(\theta)\exp(c(\theta)t(x))$$

donde el espacio muestral de X es un intervalo de R, independiente de θ, Box-Tiao (1973) aproximan el logaritmo de la función de verosimilitud, $l(\theta/x)$, hasta el término cuadrático de un desarrollo de Taylor, es decir,

$$l(\theta / x) \approx l\left(\hat{\theta}_n / x\right) - \frac{n}{2}\left(\theta - \hat{\theta}_n\right)^2\left(-\frac{1}{n}\frac{\partial^2 l}{\partial\theta^2}\right)_{\hat{\theta}_n}$$

donde $j\left(\hat{\theta}_n\right) = \left(-\frac{1}{n}\frac{\partial^2 l}{\partial\theta^2}\right)_{\hat{\theta}_n}$ depende únicamente del estimador máximo verosímil $\hat{\theta}_n$

Si ahora tomamos transformación $\phi=\phi(\theta)$ biyectiva con derivada continua; entonces

$$j(\hat{\phi}_n) = j(\hat{\theta}_n)\left(\frac{d\theta}{d\phi}\right)^2_{\hat{\theta}_n}$$

será constante si la transformación es de la forma

$$\left|\frac{d\theta}{d\phi}\right|_{\hat{\theta}_n} \propto j^{-1/2}(\hat{\theta}_n)$$

con lo que la función de verosimilitud en el nuevo parámetro ϕ es aproximadamente transladable. En consecuencia, $\pi(\phi)=1$ es, para n suficientemente grande, una función a priori imparcial. La aproximación de Taylor hasta el término cuadrático tiene un error de $1/\sqrt{(n)}$. Una mejor aproximación hasta el término cúbico, con un error de $1/n$, es dada por Kass (1990). Un ejemplo, es comprobar que para el modelo de Bernoulli, la función a priori imparcial es $\pi(\theta)\propto 1/\sqrt{(\theta(1-\theta))}$.

La definición [6] es muy útil, ahora bien, existen muchas aplicaciones donde, para una muestra **x** de tamaño n, el estadístico mínimo suficiente $(t_1, t_2, \ldots, t_n)$, $1\leq r \leq n$, es equivalente al estadístico $(\hat{\theta}_n, A)$, donde A es un vector de estadísticos conjuntamente auxiliares de dimensión n-1. En este caso no es necesario recurrir a grandes muestras para estimar el parámetro θ, debido a que dicha estimación puede realizarse por medio de la distribución condicionada $f(\hat{\theta}_n/A, \theta)$. En el siguiente ejemplo se ilustra la utilidad de esta distribución condicionada.

Ejemplo 3 (Cox y Hinkley, 1974)

Sea la familia rectangular $f(x/\theta)=1$ para $\theta-0.5<x\leq\theta+0.5$, $\theta\in R$, y es nulo en otro caso. Para una muestra $x=\{x_1,x_2,\ldots,x_n\}$ de tamaño n, la función de verosimilitud es $L(\theta/x)\propto 1$ si $y_n-0.5\leq\theta\leq y_1+0.5$, y nula en otro caso. Los estadísticos y_1 e y_n son, respectivamente, el mínimo y el máximo de la muestra. Un estadístico mínimo suficiente es (y_1, y_n), que además es equivalente a $(\hat{\theta}_n, r)$, donde $\hat{\theta}_n=(y_1+y_n)/2$ es un estimador máximo verosímil y $r=y_n-y_1$ es el estadístico recorrido que es auxiliar(su distribución no depende del parámetro θ). Es fácil

Igualmente se prueba (Welsh,1996) que la distribución marginal $f(\hat{\theta}_n/\theta)$, es

$$f\left(\hat{\theta}_n/\theta\right)=\begin{cases} n2^{n-1}\left(\hat{\theta}_n-\theta+0.5\right)^{n-1} & \text{si } \theta-0.5\leq\hat{\theta}_n\leq\theta \\ n2^{n-1}\left(\theta+0.5-\hat{\theta}_n\right)^{n-1} & \text{si } \theta\leq\hat{\theta}_n\leq\theta+0.5 \end{cases}$$

Un intervalo central de confianza con probabilidad $1-\alpha$, calculado con la distribución primera (condicionada a r), es

$$\left[\hat{\theta}_n+(1-r)\frac{\alpha}{2}+\frac{r}{2}-\frac{1}{2}, \hat{\theta}_n-(1-r)\frac{\alpha}{2}-\frac{r}{2}+\frac{1}{2}\right] \qquad [7]$$

que se reduce a un intervalo central para la distribución marginal cuando r=0 (que es un supuesto absurdo). Si $\alpha=0$, entonces los intervalos contienen el valor desconocido con seguridad. Obsérvese que si r=1, entonces el intervalo anterior se reduce al verdadero valor.

Este último ejemplo ilustra cómo el **Principio de Condicionar** a estadísticos auxiliares, conduce a sacar conclusiones válidas sobre un cierto parámetro de interés.

A partir del **Principio de Condicionar** podemos generalizar la definición [5] a situaciones donde el estadístico mínimo suficiente es de la forma $(\hat{\theta}_n, A)$, siendo A un vector de estadísticos conjuntamente auxiliares. Bajo las condiciones indicadas, diremos que la función positiva $\pi^*(\theta)$ es una función a priori imparcial si

$$GC(\pi^*(\theta),x)=k(n, A) \qquad [8]$$

que se reduce a [5] si k(n, A)=k(n).

Vamos a ilustrar la definición [8] por medio de los siguientes ejemplos.

Ejemplo 3 (continuación).

En este modelo rectangular, $\pi^*(\theta)=1$ es una función a priori imparcial. En efecto:

$$GC(1,x) = \int_{y_n-0.5}^{y_1+0.5} \frac{L(\theta / x)}{L(\hat{\theta}_n / x)} d\theta = 1 - r$$

en consecuencia, el estadístico GC(1,x) es función del estadístico auxiliar r y, así, $\pi^*(\theta)=1$ es imparcial. Es fácil probar que la correspondiente distribución a posteriori, es

$$\pi(\theta / x) = \begin{cases} (1-r)^{-1} & \text{si } \hat{\theta}_n - 0.5 + r/2 \le \theta \le \hat{\theta}_n + 0.5 - r/2 \\ 0 & \text{en otro caso} \end{cases}$$

y, también, que un intervalo bayesiano central con probabilidad 1-α coincide con [7].

Obsérvese que la función $\pi^*(\theta)=1$ no nos protege frente a modelos falsos, ya que el grado de concordancia no es constante para todas las posibles muestras de tamaño n. Recordemos que los estadísticos auxiliares son útiles para criticar modelos; en nuestro ejemplo si r=1.4, necesariamente conduce a rechazar el modelo propuesto.

La definición [8] está relacionada con el trabajo de Kass(1990). Este autor generaliza el concepto de verosimilitud transladable de Box-Tiao(1973) al caso en el que el vector de estadísticos mínimo suficientes es de la forma $(\hat{\theta}_n, A)$. En este caso $L(\theta / x) = f(\hat{\theta}_n / \theta, A)g(A)$, y, siguiendo a Box-Tiao, Kass supone que $L(\theta / x) = h_A(\theta - \hat{\theta}_n)g(A)$, con lo que la función de verosimilitud es transladable para todas las muestras, **x**, con un valor de A(x) constante. Bajo este supuesto, la función $\pi(\theta)=1$ propuesta por Kass es también una función a priori imparcial.

Ejemplo 4

Sea el modelo de Laplace $f(x/\mu)=\frac{1}{2}e^{-|x-\mu|}$, donde $x\in R$ y $\mu\in R$. Para una muestra aleatoria **x**, de tamaño n, el estadístico mínimo suficiente es la muestra ordenada $(y_1,y_2,\ldots,y_n)$. Este estadístico es equivalente al estadístico $(\hat{\mu}, C_2, C_3, \ldots, C_n)$, donde los estadísticos $C_k=y_k-y_1$, k=2,3,...,n, son un conjunto de estadísticos auxiliares, en realidad son máximos auxiliares, y $\hat{\mu}$ es el estimador máximo verosímil de μ definido por $\hat{\mu} = y_1+C_{(n+1)/2}$ cuando n es impar, y $\hat{\mu} = y_1+0.5^*(C_{n/2+}\ C_{(n+2)/2})$ cuando n es par. Es fácil probar que la densidad de $\hat{\mu}$ condicionada a los estadísticos auxiliares $(C_2,C_3,\ldots,C_n)$ y el parámetro μ, es

$$f(\hat{\mu} / C_2, C_2, \ldots, C_n, \mu) = \frac{n!e^{-|\hat{\mu}-\mu-h(C_2,C_2,\ldots,C_n)|}e^{-\sum_{k=2}^{n}|C_k+\hat{\mu}-\mu-h(C_2,C_2,\ldots,C_n)|}}{2^n f(C_2,C_2,\ldots,C_n)}$$

donde $h(C_2,C_3,\ldots,C_n)= C_{(n+1)/2}$ cuando n es impar, y $h(C_2,C_3,\ldots,C_n)= 0.5^*(C_{n/2}+ C_{(n+2)/2})$ cuando n es par. Ahora, la función $\pi^*(\theta)=1$ es una función a priori imparcial. En efecto:

$$GC(1,x) = \int_{-\infty}^{\infty} \frac{f(\hat{\mu} / C_2, C_3, \ldots, c_n, \mu)}{f(\hat{\mu} / C_2, C_3, \ldots, c_n, \hat{\mu})} d\mu = \frac{f(C_2, C_3, \ldots, c_n)2^n}{n!e^{-|-h(C_2,C_3,\ldots,C_n)|}e^{-\sum|C_k-h(C_2,C_3,\ldots,C_n)|}}$$

que depende de n y de los estadísticos auxiliares $(C_2,C_3,\ldots,C_n)$, luego $\pi^*(\theta)=1$ es una función a priori imparcial.

Ejemplo 5 (Fisher, 1990)

Sea el vector de variables aleatorias independientes (X, Y), donde X e Y siguen modelos exponenciales con parámetros θ y $1/\theta$, respectivamente. Para una muestra (x, y), el estadístico $A=\sqrt{(x.y)}$ es un estadístico auxiliar como consecuencia de que $X\theta$ e Y/θ son variables aleatorias exponenciales con parámetro unidad. El estimador máximo verosímil es $\theta_n ==\sqrt{(y/x)}$ y la muestra (x, y) es equivalente al estadístico mínimo suficiente (θ_n, A).

Ahora, la función $\pi^*(\theta)=1/\theta$ es una función a priori imparcial. En efecto:

$$GC\left(\frac{1}{\theta}, (x,y)\right) = \int_0^{\infty} \frac{1}{\theta} \frac{e^{-A\left(\frac{\hat{\theta}}{\theta}+\frac{\theta}{\hat{\theta}}\right)}}{e^{-A2}} d\theta \propto K_0(2A)e^{2A}$$

donde K_0 es la función de Bessel.

Cuando el espacio paramétrico Θ es finito, es decir, $\Theta = \{\theta_1, \theta_2, \ldots, \theta_p\}$, entonces la definición [5] se expresa como,

$$\sum_{j=1}^{p} \pi^*(\theta_j)L(\theta_j / x) = k(n)L(\hat{\theta}_n / x) \qquad [9]$$

para todas las muestras $\mathbf{x} \in M_n$. Una aplicación relevante es determinar funciones a priori imparciales para el modelo hipergeométrico.

Ejemplo 6

Para una población finita de tamaño conocido N, sea $R \in \{0,1,\ldots,N\}$ el número de elementos de la población que verifican una cierta condición. Si para estimar R seleccionamos una muestra de tamaño n con un diseño aleatorio simple sin reposición, entonces la distribución del número, y, de elementos de la muestra que verifican la condición, se distribuye como un modelo hipergeométrico, siendo su función de verosimilitud

$$L(R / N, n, y) \propto \binom{R}{y}\binom{N-R}{n-y}$$

cuando $R \in [y, N\text{-}n\text{+}y]$ y cero en otro caso. La definición [9] conduce al siguiente sistema de ecuaciones:

$$H^{(n+1)x(N+1)}\vec{\pi} = \vec{b} \qquad [10]$$

donde el elemento (i, j) de la matriz H es $L(j / N, n, i) / \binom{N}{n}$, el vector $\vec{\pi}$ contiene los valores de la función imparcial que se busca, y el elemento i-ésimo del vector $\vec{b}$ es $\dfrac{b(i)}{\sum_{k=0}^{n} b(k)}$, donde $b(i) = \binom{\hat{R}}{i}\binom{N-\hat{R}}{n-i}$ y $\hat{R}$ es el estimador máximo verosímil de R.

Si definimos $H\vec{\pi} = \vec{w}$, entonces [10] es equivalente a resolver el siguiente problema de optimización,

$$\mathrm{Min}_{\vec{w}} \sum_{i=0}^{n} b(i) \log\left(\frac{b(i)}{w(i)}\right)$$

donde $w(i)=\sum_{j=0}^{N}\pi(i)\frac{L(j/N,n,i)}{\binom{N}{n}}$. El mínimo se hace nulo cuando $\dot{w}=\dot{b}$. Este último problema es equivalente a maximizar la expresión

$$\mathrm{Max}\sum b(i)\log(w(i)) \qquad [11]$$

Ahora, [11] es equivalente a maximizar la función de verosimilitud de un modelo multinomial con probabilidades,

$$p_{ij}=\pi(j)\frac{L(j/N,n,i)}{\binom{N}{n}}$$

y donde sólo se observan los totales de las filas n_i tales que $\frac{n_{i.}}{n}=b(i)$ para i=0,1,...,n; las frecuencias absolutas, n_{ij} ,del modelo multinomial no son observadas.

En conclusión, [11] es un problema de **datos incompletos** y puede ser resuelto por el siguiente algoritmo EM (G.J.Mclachlan and T.Krishnan,1997):

(a) Partir de unos valores iniciales $\pi^0(j)$, j=0,1,...,N.

(b) Calcular nuevos valores a partir de la fórmula siguiente:

$$\pi^1(j)=\pi^0(j)\sum_{i=0}^{n}\frac{b(i)L(j/N,n,i)}{\sum_{k=0}^{N}\pi^0(k)L(k/N,n,i)}$$

y repetir el proceso hasta lograr la aproximación deseada.

Si el sistema de ecuaciones [10] no tuviera una solución, el algoritmo EM nos proporcionaría una solución factible, $\pi(j)>0$ con j=0,1,...,N, que sería, además, óptima en el sentido de [11].

Puede probarse que el sistema [10] posee infinitas soluciones factibles, que se aproximan a una única solución a medida que **n** aumenta.

El algoritmo EM utilizado en el modelo hipergeométrico puede ser extendido al caso continuo y, así, aplicarse a la resolución de Ecuaciones Integrales (Vardi y Lee, 1993). El siguiente algoritmo calcula la función f(x) no negativa de la Ecuación Integral

$$g(y) = \int_{D_f} h(x, y)\, f(x)dx$$

para $y \in D_g$, donde D_f y D_g son los dominios de las funciones no negativas f y g, respectivamente. Los pasos del algoritmo son:

(a) Partir de unos valores iniciales $f^0(x)>0$, $x \in D_f$.

(b) Actualizar la solución inicial con la siguiente expresión,

$$f^1(x) = f^0(x) \int_{D_g} \frac{h(x, y)}{\int_{D_f} (s)h(s, y)ds} g(y)dy$$

y repetir el proceso hasta lograr la aproximación deseada.

Vamos a ilustrar este algoritmo de Vardi y Lee (1993) con algunos ejemplos.

Ejemplo 7

En el Ejemplo 1, la ecuación integral [5] puede escribirse,

$$\int_{y_n}^{\infty} \pi(\theta) f(y_n / \theta) d\theta \propto f(y_n / y_n)$$

donde la función de densidad del estimador máximo verosímil, y_n, es

$$f(y_n / \theta) = \frac{ny^{n-1}}{\theta^n}$$

siendo $0< y_n \leq \theta$.Si ahora, partimos de una función a priori inicial $\pi^0(\theta)=1$, el paso siguiente nos conduce a

$$\pi^1(\theta) \propto \pi^0(\theta) \int_0^{\theta} f(y_n / y_n) \frac{f(y_n / \theta)}{\int_{y_n}^{\infty} f(y_n / \theta) d\theta} dy_n \propto$$

$$\propto \int_0^{\theta} y_n^{-1} \frac{\dfrac{y_n^{n-1}}{\theta^n}}{\dfrac{n}{n-1}} dy_n \propto \theta^{-1}$$

luego $\pi^1(\theta)=1/\theta$, que es precisamente la función a priori imparcial.

Ejemplo 8

En el Ejemplo 3, la ecuación integral [5] puede escribirse,

$$\int_{y_n}^{\infty} \pi(\theta) f(\hat{\theta}_n / r, \theta) d\theta \propto f(\hat{\theta}_n / \hat{\theta}_n)$$

Si ahora,partimos de una función a priori inicial $\pi^0(\theta)=1$, el paso siguiente nos conduce a

$$\pi^1(\theta) \propto \int_{\theta-0.5+r/2}^{\theta+0.5-r/2} f(\hat{\theta}_n / r, \theta) \frac{f(\hat{\theta}_n / r, \theta)}{\int_{\hat{\theta}_n-0.5+r/2}^{\hat{\theta}_n+0.5-r/2} f(\hat{\theta}_n / r, \theta) d\theta} d\hat{\theta}_n \propto$$

$$\propto \int_{\theta-0.5+r/2}^{\theta+0.5-r/2} f(\hat{\theta}_n / r, \theta) \frac{f(\hat{\theta}_n / r, \theta)}{1} d\hat{\theta}_n \propto (1-r)^{-1}$$

luego $\pi^1(\theta)=1$, debido a que r es constante; en consecuencia se trata de la función a priori imparcial.

Ejemplo 9

Sea la siguiente familia de densidades, $f(x / \theta) = \frac{g(x)}{h(\theta)}$, $a(\theta) \leq x \leq b(\theta)$, donde $a(\theta)$ es una función creciente y $b(\theta)$ es decreciente, $-\infty < \theta < \lambda$, siendo $a(\lambda)=b(\lambda)$, $g(x)>0$y $h(\theta)$ es decreciente estricta con $h(-\infty)=\infty$ y $h(\lambda)=0$.

Para una muestra aleatoria **x** de tamaño n, el estimador máximo verosímil para θ, es $\hat{\theta}_n = \min\{a^{-1}(y_1)), b^{-1}(y_n))\}$, y tiene la siguiente distribución en el muestreo,

$$f(\hat{\theta}_n / \theta) = \frac{-nh'(\hat{\theta}_n) h^{n-1}(\hat{\theta}_n)}{h^n(\theta)}$$

donde $\hat{\theta}_n \geq \theta$.

Ahora, la ecuación integral que define la función a priori imparcial, es

$$\int_{-\infty}^{\theta_n} \pi(\theta) f(\hat{\theta}_n / \theta) d\theta \propto f(\hat{\theta}_n / \hat{\theta}_n)$$

que si partimos de la función $\pi^0(\theta) \propto -h'(\theta)$ el paso siguiente del algoritmo nos conduce a

$$\pi^1(\theta) \propto \pi^0(\theta) \int_{\hat{\theta}} f(\hat{\theta}_n / \hat{\theta}_n) \frac{f(\hat{\theta}_n / \theta)}{\int_n \pi^0(\theta) f(\hat{\theta}_n / \theta) d\theta} d\hat{\theta}_n \propto$$

$$\propto -h'(\theta) \int_{\hat{\theta}} (\hat{\theta}_n / \hat{\theta}_n) \frac{f(\hat{\theta}_n / \theta)}{-h'(\hat{\theta}_n)} d\hat{\theta}_n \propto -\frac{h'(\theta)}{h(\theta)}$$

que produce la función a priori imparcial correspondiente.

3. RESULTADOS GENERALES

En este apartado recogemos varias familias de distribuciones de las que hemos podido calcular las correspondientes funciones a priori imparciales.

Si el espacio paramétrico Θ es finito, entonces la distribución a priori uniforme es una función a priori imparcial. La prueba de este resultado se deduce del Teorema siguiente:

Teorema 1. Si $\Theta = \{\theta_1, \theta_2, \ldots, \theta_m\}$ es el espacio paramétrico de los modelos $\{f(x|\theta) : \theta \in \Theta\}$, y si estos modelos verifican la condición siguiente:

$$\int_S \left| f(x / \theta_i) - f(x / \theta_j) \right| f(x / \theta_j) dx > 0$$

para todo $i \neq j$. Entonces para una distribución a priori positiva $\pi(\theta_i)>0$ para $i = 1, 2, \ldots, m$; se verifica que $GC(\pi, x)$ converge con "probabilidad uno" al valor $\pi(\theta^*)$, donde θ^* es el verdadero valor de θ. La demostración del Teorema puede verse en (Basulto, 1996).

Si ahora tomamos una distribución a priori uniforme $\pi^*(\theta_i) = 1/m$, para $i = 1, 2\ldots, m$; entonces $GC(\pi^*, x)$ converge con "probabilidad uno" al valor $1/m$, para todo $\theta \in \Theta$.

En el caso de que el espacio paramétrico Θ es un intervalo abierto de R y que la familia de modelos $\{f(x|\theta) : \theta \in \Theta\}$ verifican las condiciones de regularidad, utilizadas en la demostración de la normalidad asintótica del estimador máximo verosímil (Serflin, 1980), la función a priori de Jeffreys $\pi^{*}(\theta) \propto \sqrt{I_1(\theta)}$ es una función a priori imparcial, donde $I_1(\theta)$ es la información de Fisher para una muestra de tamaño n=1. Para un enunciado y una demostración puede verse (Basulto, 1996).

Cuando la familia de densidades es no regular, el cálculo de la función a priori imparcial puede ser difícil. Vamos a considerar la familia estudiada por Ghosal y Samanta (1997), donde la distribución a posteriori, centrada y normalizada, converge a una exponencial.

Supongamos que $f(x|\theta)$ es estrictamente positiva en el intervalo cerrado $S(\theta)=[a(\theta),b(\theta)]$,donde $a(\theta)$ es una función creciente y $b(\theta)$ es decreciente y que el conjunto $S(\theta)$ es decreciente. Puede verse en Ghosal y Samanta (1997), que bajo ciertas condiciones, por ejemplo, que $c(\theta)=E\left[\frac{\partial \ln(f(x/\theta))}{\partial\theta} / \theta\right]$ sea finita, se demuestra (ver el punto 2.12 de Ghosal y Samanta) que la función $\pi(\theta) \propto |c(\theta)|$ es una función a priori imparcial cuando aplicamos la definición (6).

Familias de densidades más pequeñas pueden ser abordadas directamente. Vamos a recoger algunas familias donde es directo obtener la función a priori imparcial.

El siguiente Lema afirma que para modelos donde θ es un parámetro de **localización**, la función a priori imparcial es $\pi(\theta)=1$.

Lema 1. Para el modelo $f(x/\theta)=h(x-\theta)$, $\theta \in R$. Si para todo x la función de verosimilitud $L(\theta/x)$ tiene un máximo, entonces la función a priori imparcial es $\pi(\theta)=1$.

Demostración:

Debido a que la muestra ordenada, $y=(y_1,y_2,\dots,y_n)$, es suficiente para θ, la transformación $C_k=y_k-y_1$, $k=2,3,\dots,n$, permite expresar la función de densidad conjunta cómo:

$$f(y_1,C_2,C_3,\dots,C_n/\theta) = f(y_1/C_2,C_3,\dots,C_n,\theta).f(C_2,C_3,\dots,C_n)$$

donde la función de densidad marginal no depende del parámetro θ porque los estadísticos son invariantes a las traslaciones. Si $\hat{\theta}$ es el estimador máximo verosímil, puede probarse que es homoscedástico, es decir,

$$\hat{\theta}(a + x_1, a + x_2, \ldots, a + x_n) = a + \hat{\theta}(x_1, x_2, \ldots, x_n)$$

para todo $a \in R$. Además el estimador máximo verosímil puede expresarse como $\hat{\theta} = y_1 + g(C_2, C_3, \ldots, C_n)$, en consecuencia el grado de concordancia para función $\pi(\theta)=1$ y la muestra **y** puede ser escrito cómo,

$$\int \frac{f(\hat{\theta} / C_2, C_3, \ldots C_n, \theta)}{f(\hat{\theta} / C_2, C_3, \ldots C_n, \hat{\theta})} d\theta =$$

$$= \frac{f(C_2, C_3, \ldots C_n)}{n!\, h(-g(C_2, C_3, \ldots C_n)) \prod_{i=2}^{n} h(C_i - g(C_2, C_3, \ldots C_n))} \int f(\hat{\theta} / C_2, C_3, \ldots C_n, \theta) d\theta$$

donde la última integral vale la unidad. Luego, el grado de concordancia es función de un conjunto de estadísticos auxiliares, lo que prueba nuestro enunciado.

Un lema análogo afirmar que si θ es un parámetro de **escala**, entonces la función a priori imparcial es $\pi(\theta)=1/\theta$.

Lema 2. Para el modelo $f(x/\theta)=h(x/\theta)/\theta$, $\theta > 0$. Si para todo x la función de verosimilitud $L(\theta/x)$ tiene un máximo, entonces la función a priori imparcial es $\pi(\theta)=1/\theta$.

Demostración:

La prueba es análoga que en el caso del modelo con un parámetro de localización.

La siguiente familia de modelos es la más general que admite un estadístico suficiente y el rango de la variable depende del parámetro θ (Kosmas K. Ferentinos, 1990) y (Kendall's Advanced Theory of Statistics, volume 2, 1991):

$$f(x/\theta)=g(x)/h(\theta), \quad \text{donde} \quad a(\theta) \le x \le b(\theta) \qquad [12]$$

siendo $a(\theta)$ una función monótona creciente y $b(\theta)$ una función monótona decreciente. Las funciones g y h deben ser positivas y además la función h es una función monótona decreciente. Una segunda familia es tomar $a(\theta)$ como una función monótona decreciente y $b(\theta)$ como una función monótona creciente, siendo ahora $h(\theta)$ una función monótona creciente. Estas familias son casos particulares de la familia investigada por Ghosal y Samanta (1997), ahora bien, en estas familias más concretas podemos obtener las funciones a priori imparciales con métodos directos y, además, podemos valorar mejor cuándo el resultado es válido para n finito o para n infinito.

El siguiente lema proporciona la función a priori imparcial para el parámetro de la primera familia de modelos.

Lema 3. La función $\pi(\theta)=\left|\frac{\partial\ \ln(h(\theta))}{\partial\theta}\right|$ es la función a priori imparcial para el parámetro de la primera familia.

Demostración:

Cuando $h(-\infty)=\infty$,la aplicación del primer criterio (2) conduce directamente al resultado. Cuando $h(-\infty)=c$, donde $c>0$, se prueba (Kosmas K. Ferentinos, 1990) que $(1 - n\ .GC(\pi,x))=(h(\hat{\theta})/c)^n$ sigue una distribución uniforme $(0, (h(\theta)/c)^n)$, donde $\hat{\theta}$ es el estimador máximo verosímil. En consecuencia, $(1\text{-}n\ .GC(\pi,x))$ converge en probabilidad a cero. Esto último prueba el resultado al aplicar el tercer criterio que dimos para definir funciones a priori imparciales.

Se demuestra igualmente que la función $\pi(\theta)=\left|\frac{\partial\ \ln(h(\theta))}{\partial\theta}\right|$ es también una función a priori imparcial para la segunda familia.

Ejemplos incluidos en la familia [12] son los denominados modelos truncados: Sea X una variable aleatoria con función de densidad conocida g(x) sobre R, tomemos una nueva variable aleatoria Y = X cuando $X>\theta$, donde θ es un parámetro desconocido con valores en R. La función de densidad de Y es ahora: $f(y/\theta)=g(y)/(1\text{-}G(\theta))$, con $y > \theta$. La función $G(\theta)$ es la función de distribución de X. En este ejemplo $h(\theta)=1\text{-}G(\theta)$ y ahora $h(-\infty)=1$. El último ejemplo que vamos a dar de la familia de modelos (5) es: $f(x / \theta)= k.\theta/((\theta\text{-} k).x^2)$, donde $0 <k<x<\theta$ y k es una constante conocida. Este modelo es equivalente a considerar una variable aleatoria Z con función de densidad $g(z)=k/z^2$, con $z > k$. Ahora la variable X=Z con $X<\theta$, siendo $G(\theta)=\theta/(\theta\text{-} k)$, tiene una función de densidad igual que la del modelo bajo estudio.

Un ejemplo de un modelo no incluido en la familia [12] es: Sea X una variable aleatoria normal de media μ desconocida y varianza unidad. Definamos la siguiente variable aleatoria truncada Y=X cuando $X>\mu$. La función de densidad de esta nueva variable aleatoria es $f(y/\mu)=\frac{2}{\sqrt{2\pi}}e^{-0.5(y-\mu)^2}$ con $y>\mu$. Este modelo tiene como estadístico mínimo suficiente $(m, \bar{y})$, donde m es el mínimo de los datos. Además, se prueba, que el estadístico $\bar{y}$ -m es auxiliar, luego un estadístico mínimo suficiente es también $(m, \bar{y}-m)$. Ahora, es fácil probar que $\pi(\mu)=1$ es una función a priori imparcial porque GC[1,y] depende del estadístico auxiliar $\bar{y}$ -m.

Un ejemplo de gran interés en el campo de la Econometría es el siguiente modelo AR(1):

$$Y_t = \rho Y_{t-1} + \varepsilon_t$$

donde las v.a. $\{\varepsilon_t$: t=1,2,...,n$\}$ son independientes y normales con media cero y varianza conocida σ^2. Vamos a suponer que Y_0 es conocido y que $\rho \in (-1,1)$. Para una función a priori $\pi(\rho)$, el grado de concordancia es

$$GC[\pi, x] = \int_{-1}^{1} \pi(\rho) \operatorname{Exp}\left[-\frac{0.5n}{\sigma^2}(\rho - \hat{\rho})^2 \sum_1^n y_{t-1}^2 / n\right] d\rho \quad (*)$$

la función a priori de Jeffreys para una muestra de tamaño n es igual a la expresión siguiente (P.C.B. Phillips, 1991):

$$\pi_j(\rho) \propto \left\{\frac{n}{(1-\rho^2)} - \frac{(1-\rho^{2n})}{(1-\rho^2)^2} + y_0^2 \frac{(1-\rho^{2n})}{\sigma^2(1-\rho^2)}\right\}$$

que depende del tamaño de la muestra n. La función a priori de Jeffreys se obtiene a partir de tomar el valor esperado de la expresión: $\sum_1^n y_{t-1}^2 / n$. Pero esta última expresión converge en probabilidad a $\frac{\sigma^2}{(1-\rho^2)}$ y el estimador máximo verosímil $\hat{\rho}$ converge en probabilidad al valor verdadero $\rho \in (-1,1)$. Si ahora aplicamos a (*) la aproximación de Laplace (Olver, F. W. J., 1974), la integral se aproxima, para n grande, a

$$\sqrt{\frac{2\pi}{n}} \left\{ \frac{\pi(\hat{\rho})\sigma}{\sqrt{\sum_1^n y_{t-1}^2 / n}} \right\}$$

y la expresión, entre llaves, converge en probabilidad a $\frac{\pi(\rho)}{\left(1/\sqrt{(1-\rho^2)}\right)}$. En consecuencia la función a priori imparcial es $\pi(\rho) \propto \sqrt{\frac{1}{(1-\rho^2)}}$.

4. FUNCIONES A PRIORI IMPARCIALES E INTERVALOS DE CONFIANZA CLÁSICOS

La obtención de un intervalo de confianza clásico para un cierto parámetro puede presentar problemas cuando el pivote se construye con un estadístico **no suficiente**. El siguiente ejemplo ilustra el problema.

Sea X una variable aleatoria con función de densidad $f(x/\theta)= \exp[\theta -x]$ cuando $x>\theta$ y es cero en otro caso. Para una muestra aleatoria **x** de tamaño n, la variable aleatoria $q=\sum_{i=1}^{n}(x_i-1)/n-\theta+1$ sigue la función de densidad $q^{n-1}\exp[-q]$ donde $q > 0$ (E. T. Jaynes, 1976). En consecuencia q es un pivote. Jaynes prueba que para la muestra {12, 14, 16}, el intervalo de longitud mínima, con un coeficiente de confianza del 90 por ciento, es [12.14, 13.82]. El intervalo resulta un **absurdo** debido a que por definición $\theta < 12$. La causa de este desagradable resultado se debe a que el estadístico $\sum_{i=1}^{n}(x_i-1)/n$ no es suficiente para el parámetro θ. Un estadístico suficiente es el estimador máximo verosímil θ=m, donde **m** es el menor de los datos observado en **x.** En Kosmas K. Ferentinos (1990) se prueba que el pivote $p=h(m)/h(\theta)$, donde $h(\theta) = \exp[-\theta]$, sigue la función de densidad $n\,.p^{n-1}$, con $0<p<1$. El intervalo de mínima longitud y coeficiente 90 por ciento es [11.23, 12] (Kosmas K. Ferentinos, 1990). Jaynes demuestra que con una función a priori imparcial $\pi(\theta)=1$, el intervalo clásico [11.23, 12] es un intervalo en la distribución a posteriori con probabilidad 0.90.

Uno de los trabajos más importantes que relaciona intervalos unilaterales clásicos con bayesianos, es el realizado por Welch, B. L. y Peers, H. W. (1963). Estos autores probaron el resultado siguiente: Para la función a priori de Jeffreys, $\pi(\theta)\propto\sqrt{I_1(\theta)}$, un intervalo a posteriori unilateral de la forma $\{\theta < g(x,\pi,\alpha)\}$, donde $P_r[\{\theta < g(x,\pi,\alpha)\} | x] = \alpha$, se comporta como un intervalo de confianza con un coeficiente de confianza igual a $\alpha + O(1/n)$. El símbolo O debe interpretarse en probabilidad.

El resultado de Welch, B. L. y Peers, H. W. (1963) tiene un doble interés: en primer lugar permite interpretar la función a priori de Jeffreys como un método para construir la Inferencia Bayesiana sobre una base **Objetiva**; y, en segundo lugar, abre un camino alternativo para construir intervalos de confianza clásicos. También, que el intervalo del ejemplo, [11.23, 12], sea válido tanto como un intervalo de confianza clásico como un intervalo bayesiana, logra lo que buscaba la inferencia fiducial de Fisher (Zabell,1994).

El trabajo de Welch, B. L. y Peers, H. W. (1963) no cubre, por ejemplo, modelos $f(x/\theta)$, tales que la derivada del logaritmo de la correspondiente función de verosimilitud es diferente de cero en su máximo. A continuación aplicamos las ideas de Welch. y Peers a una familia de modelos no considerada por dichos autores.

La siguiente familia de modelos es la de Kosmas K. Ferentinos (1990), donde $a(\theta)$ es monótona creciente y $b(\theta)$ monótona decreciente. Del lema 3 sabemos que la función $\pi(\theta)=\left|\frac{\partial \ln(h(\theta))}{\partial\theta}\right|$ es la función a priori imparcial para el parámetro. El siguiente lema obtiene el pivote propuesto por Kosmas K. Ferentinos.

Lema 4. Si $h(-\infty)=\infty$, un intervalo unilateral bayesiano es un intervalo unilateral clásico para la función a priori imparcial $\pi(\theta)=\left|\frac{\partial \ln(h(\theta))}{\partial\theta}\right|$.

Demostración:

En primera lugar se prueba que la distribución a posteriori para θ es la función

$$\pi(\theta / x) = n\frac{-h'(\theta)}{h(\theta)}\left(\frac{h(\hat{\theta})}{h(\theta)}\right)^n$$

y ahora $r(\pi,\theta) = \int_{-\infty}^{\theta} \pi(t / x)dt = \left(\frac{h(\hat{\theta})}{h(\theta)}\right)^n$ es una variable aleatoria con función de densidad uniforme en (0.1). Esto último es consecuencia de que $q= h(\hat{\theta})/h(\theta)$ sigue la función de densidad $n\,.q^{n-1}$, con $0<q<1$ (Kosmas K. Ferentinos, 1990).

Un resultado equivalente, se obtiene con la segunda familia de modelos si tomamos $h(\infty)=\infty$.

El siguiente resultado relaciona intervalos bayesianos con intervalos de confianza en la primera familia de modelos considerados por (Kosmas K. Ferentinos, 1990), en el caso de que $h(-\infty)=c$, donde $c>0$.

Lema 5. Si $h(-\infty)=c$, donde $c>0$, entonces la v.a. $r(\pi,\theta)$ se aproxima, para n grande, a una distribución uniforme (0,1).

Demostración:

$$r(\pi,\theta) = \int_{-\infty}^{\theta} \pi(t / x)dt = \frac{\left(\frac{h(\hat{\theta})}{h(\theta)}\right)^n - \left(\frac{h(\hat{\theta})}{c}\right)^n}{1-\left(\frac{h(\hat{\theta})}{c}\right)^n}$$

y al ser $\left(\frac{h(\hat{\theta})}{h(\theta)}\right)^n$ una v.a. uniforme (0.1), entonces:

$$P_r[r \leq r_0 / \theta] = \frac{r_0}{1-\left(\frac{h(\theta)}{c}\right)^n + r_0\left(\frac{h(\theta)}{c}\right)^n}$$

prueba el resultado que buscamos.

Por último, vamos a recoger dos resultados generales, el primero sobre el modelo $f(x/\theta)=h(x-\theta)$, $\theta \in R$, y el segundo sobre el modelo $f(x/\theta)=h(x/\theta)/\theta$, $\theta > 0$.

Lema 6. Para el modelo $f(x/\theta)=h(x-\theta)$, $\theta \in R$. Si para todo x la función de verosimilitud $L(\theta/x)$ tiene un máximo y $\pi(\theta)=1$, entonces $r(\pi, \theta) = \int_{-\infty}^{\theta} \pi(t/x)dt$ define una variable aleatoria cuya distribución, condicionada a los estadísticos subsidiarios $\{C_2,\ldots,C_n\}$, es uniforme en (0,1), para todo θ.

Demostración:

Puede verse en Welch, B. L. y Peers, H. W. (1963).

Lema 7. Para el modelo $f(x/\theta)=h(x/\theta)/\theta$, $\theta > 0$. Si para todo x la función de verosimilitud $L(\theta/x)$ tiene un máximo y $\pi(\theta)=1/\theta$, entonces $r(\pi, \theta) = \int_{-\infty}^{\theta} \pi(t/x)dt$ define una variable aleatoria uniforme en (0,1), para todo θ.

Demostración:

Para la muestra ordenada $y=(y_1,y_2,\ldots,y_n)$, su función de densidad conjunta es, $f(y_1,y_2,\ldots,y_n/\theta) = n!\,\theta^{-n}\prod_{i=1}^{n} h\left(\frac{y_i}{\theta}\right)$. La transformación $y_1=y_1, t_k=y_k/y_1$ para $k=2,3,\ldots,n$; conduce a siguiente función de densidad condicionada,

$$f(y_1/t_2,t_3,\ldots,t_n,\theta) = \frac{n!\,\theta^{-n} h\left(\frac{y_1}{\theta}\right)\prod_{k=2}^{n} h\left(\frac{t_k y_1}{\theta}\right)|y_1|^{n-1}}{f(t_2,t_3,\ldots,t_n)}$$

donde la función de densidad del denominador no depende del parámetro θ, porque los estadísticos $\{t_k,\ k=2,3,\ldots,n\}$ son conjuntamente auxiliares. Si $\hat{\theta}$ es el estimador máximo verosímil de θ, entonces verifica $\frac{y_1}{\hat{\theta}} = g(t_1, t_2, \ldots, t_n)$, ya que los estadísticos $\{t_k,\ k=2,3,\ldots,n\}$ son máximos auxiliares. Ahora, la función de densidad de $\hat{\theta}$ condicionada a los estadísticos auxiliares, es

$$f(\hat{\theta} / t_2, t_3, \ldots, t_n, \theta) = \frac{n!\,\theta^{-n} h\left(\frac{\hat{\theta}\, g(t_1, t_2, \ldots, t_n)}{\theta}\right) \prod_{k=2}^{n} h\left(\frac{t_k \hat{\theta}\, g(t_1, t_2, \ldots, t_n)}{\theta}\right) \hat{\theta}^{\,n-1} \left|g(t_1, t_2, \ldots, t_n)\right|^n}{f(t_2, t_3, \ldots, t_n)}$$

y la distribución a posteriori es

$$\pi(\theta / x) \propto \theta^{-1} f(\hat{\theta} / t_2, t_3, \ldots, t_n, \theta)$$

A partir de aquí, es fácil probar que la función de distribución a posteriori en θ_1, $F(\theta_1/x)$, es un estadístico que sigue una distribución uniforme (0, 1).

A continuación vamos a ver que el resultado de Welch, B. L. y Peers, H. W. (1963) es exacto en el modelo de Pareto. Sea X una variable aleatoria de Pareto con función de densidad,

$$f(x / \alpha, \beta) = \alpha . \beta^{-\alpha} x^{-(1+\alpha)}$$

donde $\alpha>0$, $\beta>0$ y $x \geq 1/\beta$. Vamos a suponer que β es conocido.

Para una muestra aleatoria **x** de tamaño n, la función de verosimilitud se puede expresar: $L(\alpha/x) \propto \alpha^n e^{n-\alpha(v-n\,t)}$, donde $t =- \log(\beta)$ y $v = \sum_{i=1}^{n} \log(x_i)$. Se prueba directamente que la función a priori imparcial es $\pi(\alpha) \propto 1/\alpha$ y la distribución a posteriori para α es una Gamma(n, v -n t). Vamos a probar que la variable aleatoria $r(\pi,\alpha)$ sigue una distribución uniforme (0, 1) para todo α:

$$r(\pi, \alpha) = \int_0^{\alpha} \frac{(v-n)^n}{\Gamma(n)} z^{n-1} e^{-z(v-nt)} dz = \int_0^{2\alpha(v-nt)} \frac{1}{\Gamma(n) 2^n} w^{n-1} e^{-w/2} dw$$

La última integral es igual al valor de la función de distribución de una Ji - cuadrado con 2n grados de libertad calculada en $2\alpha(v\ -n\ t)$. Pero la variable aleatoria $2\alpha(v\ -n\ t)$ sigue una Ji - cuadrado con 2n grados de libertad; con lo que la variable aleatoria $r(\pi,\alpha)$ sigue una distribución uniforme (0, 1), para todo α.

A continuación, vamos a aplicar el resultado de Welch, B. L. y Peers, H. W. (1963) al modelo de Bernoulli, que no es exacto y sólo es válido para grandes muestras.

Recogimos en el apartado 2, que $\pi(\theta) \propto 1/\sqrt{(\theta(1-\theta))}$ es, para grandes muestras, la función a priori imparcial para θ. Si ahora tomamos una muestra aleatoria **x** de tamaño n, es fácil ver que la variable aleatoria $r(\pi,\theta)$ es:

$$r(\pi,\theta)=\frac{\int_0^\theta z^{r-1/2}(1-z)^{n-r-1/2}\,dz}{\int_0^1 z^{r-1/2}(1-z)^{n-r-1/2}\,dz}$$

que se aproxima, según Welch, B. L. y Peers, H. W. (1963), para n grande a un modelo uniforme (0,1). En la siguiente tabla recogemos para diferentes tamaños de la muestra y para varios valores del parámetro θ, las probabilidades asociadas a intervalos unilaterales bayesianos (.90, .95 y .99) y sus correspondientes coeficientes de confianza que logran cuando se utilizan como intervalos de confianza clásicos.

Tabla 1

	.90	*.95*	*.99*		*.90*	*.95*	*.99*
n = 10				n = 20			
θ = .2	.892	.892	1	θ = .2	.930	.930	.988
.6	.833	.945	.987	.6	.872	.943	.993
.8	.879	.967	.993	.8	.913	.967	.990
.1	1	1	1	.1	.878	.878	1
.9	.929	.929	.987	.9	.867	.956	.988
.05	1	1	1	.05	1	1	1
.95	.913	.913	.98	.95	.924	.924	.984
n = 40				n = 80			
θ = .2	.924	.924	.992	θ = .2	.924	.924	.992
.6	.870	.960	.991	.6	.870	.960	.991
.8	.912	.956	.992	.8	.912	.956	.992
.1	.919	.919	.985	.1	.919	.919	.985
.9	.900	.958	.984	.9	.899	.946	.987
.05	.871	.871	1	.05	.871	.871	1
.95	.861	.951	.986	.95	.894	.953	.993

En la tabla adjunta, las probabilidades dentro de las celdillas son los coeficientes que alcanzan los intervalos bayesianos para las probabilidades de .90, .95 y .99. Los resultados que recogemos en la tabla confirman el trabajo de Welch, B. L. y Peers, H. W. (1963). Los cálculos se han realizado con el programa Mathematica.

Sólo resaltamos que para valores pequeños del parámetro, la aproximación necesita un valor mayor del tamaño de la muestra.

Por último, debemos resaltar la robustez que poseen los métodos bayesianos en relación a la elección de la función a priori, como consecuencia de los teoremas que demuestran que, bajo condiciones bastantes generales, la distribución a posteriori, centrada y normalizada, converge, para n tendiendo a infinito, a una cierta distribución, para **toda función a priori positiva y continua**.

Estos últimos resultados nos conducen a preguntarnos sobre, por ejemplo, por qué investigar las funciones a priori imparciales. Nuestra repuesta es doble: (a) el concepto de imparcialidad es más evidente que otros conceptos tales como, por ejemplo, el de funciones a priori no informativas, y (b) los intervalos bayesianos pueden comportarse como intervalos clásicos para todo tamaño muestral o para un tamaño muestral moderadamente grande.

El siguiente ejemplo tomado de Ghosal y Samanta (1997) aclarará el contenido de (b).

Ejemplo 10 (Ghosal and Samanta (1997))

Sea la familia del Ejemplo 1. Para la función a priori imparcial, $\pi_1(\theta)=\theta^{-1}$,y una muestra aleatoria **x** de tamaño n, la distribución a posteriori, es

$$\pi_1(\theta / x) = \frac{n y_n^n}{\theta^{n+1}} ; \theta \geq y_n$$

Para la función a priori, $\pi_2(\theta)=e^{-\theta}$, y una muestra aleatoria **x** de tamaño n, la distribución a posteriori, es

$$\pi_2(\theta / x) = \frac{e^{-\theta}\theta^{-n} y_n^{n-1}}{\int_1^{\infty} \frac{e^{-t y_n}}{t^n} dt} \quad ; \theta \geq y_n$$

Un intervalo bayesiano con probabilidad 1-α en $\pi_1(\theta/x)$, es

$$\Pr[\theta \leq \theta_\alpha / x] = 1 - \alpha$$

es decir, $\theta_\alpha = \frac{y_n}{\alpha^{1/n}}$.

En nuestro trabajo probamos que $\Pr[\theta \leq \theta_\alpha / \theta] = 1 - \alpha$, $\forall\theta>0$.

Un intervalo bayesiano con probabilidad 1-α en $\pi_2(\theta/\mathbf{x})$, verifica

$$\int_{Y_n}^{\theta_\alpha} \pi_2(\theta / x) d\theta = 1 - \alpha$$

dando lugar al intervalo $\{\theta \leq \hat{\theta}_\alpha\}$.

En las siguientes tablas recogemos los resultados obtenidos por simulación, de los intervalos $\{\theta \leq \theta_\alpha\}$ y $\{\theta \leq \bar{\theta}_\alpha\}$, para $\theta=9$, $1-\alpha=0.90$, diferentes tamaño muestrales n y varios replicaciones de las muestras. Todos los cálculos han sido realizados con Mathematical.

Tabla 2

n=5	$Pr[\theta \leq \theta_\alpha/\theta]$	$Pr[\theta \leq \bar{\theta}_\alpha/\theta]$
1000	0.897	0.564
2000	0.905	0.592
5000	0.897	0.594
n=10	$Pr[\theta \leq \theta_\alpha/\theta]$	$Pr[\theta \leq \bar{\theta}_\alpha/\theta]$
1000	0.892	0.714
2000	0.895	0.721
n=25	$Pr[\theta \leq \theta_\alpha/\theta]$	$Pr[\theta \leq \bar{\theta}_\alpha/\theta]$
1000	0.892	0.781
2000	0.900	0.824

Cuando $\theta=1$, hemos obtenido la siguiente tabla:

n=5	$Pr[\theta \leq \theta_\alpha/\theta]$	$Pr[\theta \leq \bar{\theta}_\alpha/\theta]$
1000	0.888	0.889

donde se observa que en este caso ambas funciones a priori conducen a los mismos resultados a partir de pequeñas muestras.

Cuando $\theta<1$, se observa que $Pr[\theta \leq \bar{\theta}_\alpha/\theta] \geq 1-\alpha$. Para valores altos de θ, $Pr[\theta \leq \hat{\theta}_\alpha/\theta]$ se aproxima a cero cuando n se mantiene. Estos comportamientos observados son debidos a que el **grado de concordancia** de la función a priori $\pi_2(\theta)=e^{-\theta}$ posee un máximo en el entorno de $\theta=1$ y disminuye rápidamente a cero a medida que se acerca a cero o se aleja hacia infinito.

Un último ejemplo, es

Ejemplo 11

Supongamos las funciones de densidades $f(x/\theta)=\frac{k}{x^2h(\theta)}$ cuando $0<k\leq x\leq\theta$ y es nula en otro caso, y donde $h(\theta)=\frac{\theta-k}{\theta}$. En este ejemplo, la función a priori imparcial es $\pi(\theta)=h'(\theta)/h(\theta)$, y el grado de concordancia es igual a

$$GC[\pi(\theta),x]=\frac{1}{n}\left(1-h^n(\hat{\theta})\right)$$

donde $\hat{\theta}$ es el estimador máximo verosímil que coincide con el máximo valor de la muestra. Si ahora seleccionamos un intervalo bayesiano, $\{\theta\leq\theta\leq\theta_1\}$ con probabilidad 1-α, es fácil probar que

$$Pr\left[\hat{\theta}\leq\theta\leq\theta_1/\theta\right]=1-\frac{\alpha}{1-(1-\alpha)h^n(\theta)}$$

donde se observa que cuando θ es pequeño, entonces el intervalo bayesiano se comporta mejor como un intervalo de confianza clásico, consecuencia de ser el grado de concordancia alto. Cuando θ es grande, debemos aumentar el tamaño muestral si queremos que el intervalo bayesiano se comporte como un intervalo clásico, debido a que el grado de concordancia es muy pequeño.

5. RESUMEN Y DISCUSIÓN

En el presente trabajo hemos introducido el concepto de función a priori imparcial para un parámetro unidimensional. Las funciones a priori imparciales permiten dar una nueva interpretación a la función a priori de Jeffreys y, además, pueden aplicarse a un mayor conjunto de modelos. Hemos aplicado el concepto de función a priori imparcial a varias familias de modelos, obteniendo la correspondiente función a priori. Por último, hemos utilizado el concepto de función a priori imparcial como un método alternativo para obtener intervalos de confianza clásicos.

Varios cuestiones quedan pendientes: (a) existencia, (b) unicidad y (c) solución general, de la ecuación integral que define el concepto de función a priori imparcial.

Los puntos (a) y (c) admiten ser resulto a partir del algoritmo de Vardi y Lee (1993), al darnos la mejor aproximación en el sentido de la expresión [11]. En cuanto al problema (b), tiende a perder importancia a medida que el tamaño de la muestra aumenta. Cuando el tamaño de la muestra es pequeño, deberemos estudiar el comportamiento de la distribución a posteriori a modificaciones en las funciones a priori imparciales.

La generalización del concepto de función a priori imparcial a dos o más parámetros admite varias posibilidades. En nuestro trabajo (Basulto, 1995) pueden verse algunas propuestas. Nos ha parecido que en el presente trabajo nos centremos en el caso univariante, para evitar complicaciones y porque muchos de los problemas multivariante pueden ser abordados a partir de los métodos univariantes.

REFERENCIAS

BASULTO, J. (1995), «Una Propuesta para Obtener Distribuciones a Priori Imparciales». *Publicación del Departamento de Economía Aplicada I de la Universidad de Sevilla.*

BASULTO, J. (1996), «Univariate Public Prior Distributions». *Publicación del Departamento de Economía Aplicada I de la Universidad de Sevilla.*

BOX G .E. P. and TIAO. G. (1973), «Bayesian Inference in Statistical Analysis». *Addison-Wesley.*

COX, D. R. and HINKLEY (1974), «Theoretical Statistics». *Chapman and Hall.* London.

FISHER , R. A. (1934), «The effect of methods of ascertainment upon the estimation of frequencies». *Annals of Eugenics* 6, 13-25.

GHOSAL, S. and SAMANTA, T. (1997), «Asymtotic Expansions of Posterior Distributions in Nonregular Cases». *Ann. Inst. Statist. Math. Vol 49, Nº 1, 181-197.*

JAYNES, E. T. (1976), «Confidence Intervals vs Bayesian Intervals». *Foundations of Probability Theory Statistics Inference, and Statistical Theories of Science.* W. L. Harper and C. A. Hooker, editors; D. Reidel Publishing Co,. Dordrecht. Holand.

KASS, R. E. (1990), «Data-translated likelihood and Jeffreys'rules». *Biometrika 77, 107-114.*

KASS, R. E. and WASSERMAN, L. (1995), «The Selection of Prior Distributions by Formal Rules. Technical Report # 583». *Department of Statistics. Carnegie Mellon University.* Pittsburgh, Pennsylvania.

FISHER, R. A. (1990), «Statistical Methods. Experimental Design and Scientific Inference». *Oxford Science Publications.*

KOSMAS, K. FERENTINOS. (1990), «Shortest Confidence Intervals for Families of Distributions Involving Truncation Parameters». *The American Statistician*, 44, 167-168.

MCLACHLAN, G. J. and KRISHNAN, T. (1997), «The EM Algorithm and Extensions». Wiley.

OLVER, F. W. J. (1974), «Asymptotic and Special Functions». *Academic Press.* New York.

PHILLIPS, P. C. B. (1991), «To criticize the critics: an objectives Bayesian analysis of stochastics trends». *Journal of Applied Econometrics, Vol. 6*, nº 4, 331-364

RAO, C. R. (1965), «On discreta distributions arising out of methods of ascertainment». *Clasical and Contagious Discrete Disributions* (G. P. Patil, de). Calcuta Statistical Publishing Society , 320-333.

RAO, C. R. (1994), «Estadística y Verdad. Aprovechando el Azar». *PPU.* Barcelona.

SERFLING, R. J. (1980), «Approximations Theorems of Mathematical Statistics». Wiley.

STUART, A. and ORD, J. K. (1991), «Kendall' Advanced Theory of Statistics». Volume 2. Fifth Edition.

VARDI, Y. and LEE, D. (1993), «From Image Deblurring to Optimal Investments:Maximum Likelihood Solutions for Positive Linear Inverse Problem». *J.R.Statist. Soc. B. 55, nº 3, 569-612.*

WELCH, B. L. and PEERS, H. W. (1963), «On Formulae for Confidence Points Based on Integrals of Weighted Likelihoods». *Journal of the Royal Statistical Society, Ser. B,25*, 318-329.

WELSH, A. H. (1996), «Aspects of Statistical Inference». Wiley.

ZABELL, S. L. (1992), «R.A. Fisher and the Fiducial argument». *Statistical Science, 7*, 369-387.

UNIDIMENSIONAL IMPARTIAL A PRIORI FUNCTIONS

SUMMARY

In this paper, we introduce the concept of impartial a priori funcions for a unidimensional parameter. These functions allow us a new interpretation of Jeffreys´s a priori function and, yet, can be applied to a larger set of models. We apply this concept to several families of models, thus obtaining the corresponding impartial a priori public functions. Finally, we use the concept of impartial a priori functions as an alternative method for obtaining classic confidence intervals.

Keywords: bayesian estimation, impartial a priori functions, confidence intervals, Jeffreys's rule.

AMS Classification: 62F15, 62G15.