DISTRIBUCIONES NO INFORMATIVAS EN MODELOS TRUNCADOS NO REGULARES: MODELOS DE BÚSQUEDA DE EMPLEO.

BASULTO SANTOS, Jesús
Departamento de Economía Aplicada I
Universidad de Sevilla
Correo-e: basulto@us.es

ORTEGA IRIZO, Fco. Javier
Departamento de Economía Aplicada I
Universidad de Sevilla
Correo-e: fjortega@us.es

RESUMEN

Ofrecemos una vía de generalización de la Regla de Jeffreys para la obtención de distribuciones a priori, que puede aplicarse también a modelos no regulares. El análisis del caso unidimensional nos permitirá también obtener las distribuciones a priori en modelos multidimensionales, al menos en algunas situaciones específicas. A partir de estas distribuciones, podemos aplicar las técnicas de la Inferencia Bayesiana a aquellos modelos económicos en los que aparecen “parámetros no regulares”, como son los modelos de búsqueda de empleo, de subastas del sector público, etc.

Palabras Clave: Inferencia Bayesiana, Distribución a priori no informativa, modelo no regular, modelos de búsqueda de empleo.

1. Introducción.

Dentro del marco de la Inferencia Bayesiana, la obtención de una distribución a priori no informativa es un problema interesante, que ha generado gran cantidad de literatura. En la única situación en la que parece haber total acuerdo es en el caso de modelos uniparamétricos regulares, en el que suele optarse por la distribución obtenida a partir de la Regla de Jeffreys. Cuando los modelos tienen más de un parámetro, y más aún si no se cumplen las condiciones de regularidad usuales, existen multitud de propuestas para obtener las distribuciones a priori; cabe destacar la referencia a priori de Bernardo (Bernardo y Smith, 1994) por ser quizás la opción más habitualmente utilizada en estos últimos casos.

La distribución a priori de Jeffreys se basa, como es conocido, en el concepto de Información de Fisher, que adquiere sentido sólo bajo condiciones de regularidad. Por ello, nuestra propuesta de obtención de distribuciones a priori en modelos no regulares se basará en una medida de Información aplicable a modelos no regulares (denominada Información de Akahira) y a su relación con la Información de Fisher. De esta forma, ofrecemos una generalización de la Regla de Jeffreys para modelos no regulares uniparamétricos. La construcción de la distribución a priori para modelos de más de un parámetro, se hará siguiendo los caminos más usuales utilizados hasta ahora en la literatura sobre el tema, que en definitiva consisten básicamente en obtener la distribución conjunta a partir de distribuciones unidimensionales (bien marginales o bien condicionadas), si bien este es un tema que permanece aún bastante abierto y al que no se ha ofrecido una respuesta general satisfactoria.

Existen multitud de ejemplos de modelos no regulares en economía (Sareen, 2003). Nosotros vamos a ocuparnos de un modelo estacionario de búsqueda de empleo propuesto en Lancaster, 1997. En este modelo, uno de los parámetros desconocidos es el denominado “salario de reserva” (ξ), que es el valor mínimo a partir del cual el trabajador acepta la oferta de trabajo. Así, un trabajador sólo acepta una oferta con salario W si se verifica $W > \xi$; esta dependencia del recorrido de la variable W del parámetro desconocido ξ hace que el modelo no verifique las condiciones usuales de regularidad. Ofrecemos una solución parcial para este modelo, en el caso de que supongamos que la variable W sigue una distribución uniforme de extremos desconocidos; la obtención de la distribución a priori para otros casos es una cuestión que queda abierta para futuros trabajos.

A partir de aquí, el trabajo se estructura como sigue. En la sección 2, definimos la Información de Fisher y la de Akahira, así como la relación fundamental que puede establecerse entre ambas. En la sección 3, ofrecemos nuestra propuesta de obtención de la distribución a priori no informativa en el caso uniparamétrico, una fórmula cómoda de cálculo aplicable bajo ciertas condiciones y analizamos las propiedades de la distribución obtenida. En la sección 4, se describe la obtención de la distribución conjunta a partir de las condicionadas unidimensionales, ofreciéndose una solución parcial aplicable en ciertos casos interesantes. En la sección 5, se expone brevemente el modelo de búsqueda de empleo propuesto por Lancaster, 1997 y se obtiene la distribución a priori no informativa conjunta, a partir de la cual podemos obtener las distribuciones marginales a posteriori de forma exacta. Por último, en la sección 6 se exponen las conclusiones fundamentales y las cuestiones no resueltas.

2. Información de Fisher e Información de Akahira.

2.1. Información de Fisher

Consideremos una familia de distribuciones P_θ, $\theta \in \Theta \subseteq \mathbb{R}^k$ con funciones de densidad $f(x|\theta)$, verificando las condiciones estándar de regularidad (Azzalini,1996, p.71). Como es conocido, a partir de las hipótesis de regularidad pueden obtenerse las dos propiedades siguientes:

$$1.\ E_\theta\left[\frac{\partial \log f(x|\theta)}{\partial\theta}\right]=0 \quad \forall\theta \qquad y \qquad 2.\ E_\theta\left[\left(\frac{\partial \log f(x|\theta)}{\partial\theta}\right)^2\right]=-E_\theta\left[\frac{\partial^2 \log f(x|\theta)}{\partial\theta^2}\right] \quad \forall\theta .$$

Definición: Se llama Información de Fisher que la variable X proporciona sobre el parámetro $\theta \in \Theta \subseteq \mathbb{R}^k$ a

$$I_X(\theta)=E_\theta\left[\left(\frac{\partial \log f(x|\theta)}{\partial\theta}\right)^2\right].$$

Bajo las hipótesis de regularidad, aplicando la propiedad 2, se obtiene que

$$I_X(\theta) = -E_\theta\left[\frac{\partial^2 \log f\,(x\,|\,\theta)}{\partial\theta^2}\right].$$

Nota: En el caso de más de un parámetro, bajo las hipótesis de regularidad, podemos definir la matriz de Información de Fisher como:

$$I(\theta) = E_\theta\left[\left(\frac{\partial \log f(x\,|\,\theta)}{\partial\theta}\right)\left(\frac{\partial \log f(x\,|\,\theta)}{\partial\theta}\right)^t\right] = -E_\theta\left[\frac{\partial^2 \log f(x\,|\,\theta)}{\partial\theta\partial\theta^t}\right],$$

es decir, $\left(I(\theta)\right)_{i,j} = -\,E_\theta\left[\frac{\partial^2 \log f(x\,|\,\theta)}{\partial\theta_i\partial\theta_j}\right].$

A partir de los trabajos de Jeffreys, 1946, 1961, el concepto de Información de Fisher tomó relevancia dentro del enfoque Bayesiano, ya que dada una reparametrización biyectiva y regular $\varphi=\varphi(\theta)$ (donde suponemos que ambos parámetros son unidimensionales), se verifica que $I(\varphi) = I(\theta)\left|\partial\theta/\partial\varphi\right|^2$ y por tanto, la regla de obtención de distribuciones a priori consistente en tomar $\pi(\theta) \propto \sqrt{I(\theta)}$ es invariante ante reparametrizaciones, ya que verifica

$$\pi(\varphi) \propto \sqrt{I(\varphi)} = \sqrt{I(\theta)}\left|\frac{\partial\theta}{\partial\varphi}\right| = \pi(\theta)\left|\frac{\partial\theta}{\partial\varphi}\right|.$$

Así, este autor propuso la llamada regla de Jeffreys para construir distribuciones a priori no informativas basándose en el concepto de Información de Fisher. En el caso univariante, esta es la opción actualmente más aceptada.

La importancia del concepto de Información de Fisher es indudable, aunque eso sí, recordando siempre que sólo es aplicable a modelos regulares. Por ello, es natural plantearse si habrá alguna forma de generalizar este concepto o al menos definir una medida de información aplicable a modelos no regulares y que tenga en esencia todas las propiedades de la Información de Fisher.

2.1. Información de Akahira.

Existen varias definiciones alternativas de medidas de información aplicables a modelos no regulares, aunque nosotros vamos a trabajar sólo con la que utilizan Akahira y Takeuchi, 1991.

Consideremos una familia de distribuciones de probabilidad cuyas funciones de densidad, con respecto a la medida de Lebesgue, sea $\{f(x \mid \theta), \theta \in \Theta \subseteq \mathbb{R}\}$ y $X_1, ..., X_n$ una muestra aleatoria simple del modelo $f(x \mid \theta)$. Definimos la cantidad de información entre $f(\bullet \mid \theta_1)$ y $f(\bullet \mid \theta_2)$ como:

$$J_{X_1}(\theta_1, \theta_2) = -8 \log \int f(x \mid \theta_1)^{1/2} f(x \mid \theta_2)^{1/2} dx .$$

La integral que aparece en la definición es conocida como la afinidad entre $f(\bullet \mid \theta_1)$ y $f(\bullet \mid \theta_2)$ (que llamaremos $A_{X_1}(\theta_1, \theta_2)$). Podemos observar de forma trivial que: i) Si $\theta_1 = \theta_2$, entonces la afinidad es uno (es decir, la afinidad de una variable consigo misma es uno); ii) Si $\mathrm{sop}(X \mid \theta_1) \cap \mathrm{sop}(X \mid \theta_2) = \varnothing$, entonces la afinidad entre las distribuciones es cero (donde $\mathrm{sop}(X \mid \theta_i)$ representa el soporte de la densidad $f(x \mid \theta_i)$).

La afinidad es una medida de “similitud” entre distribuciones, que toma valores entre 0 y 1 (Matusita, 1955, Le Cam, 1990). Al ser la información una función decreciente de la afinidad, obtendremos que la información es una medida de “discrepancia” entre distribuciones. Remarquemos también que la información entre dos variables estará comprendida entre 0 e infinito, alcanzándose estos valores en los dos casos extremos reseñados anteriormente.

Indiquemos que la Información de Akahira, reproduce las tres propiedades más importantes de la información de Fisher (Ortega y Basulto, 2003).

Ejemplo 2.1: Consideremos el modelo Exponencial de parámetro $\theta>0$, cuya función de densidad es $f(x \mid \theta) = \theta e^{-\theta x}$, $x > 0$, $\theta > 0$.

En este caso, $\forall \theta_1, \theta_2 > 0$, tenemos $A(\theta_1, \theta_2) = 2\sqrt{\theta_1 \theta_2} / (\theta_1 + \theta_2)$ y por tanto, $J(\theta_1, \theta_2) = -8 \log\left(2\sqrt{\theta_1 \theta_2} / (\theta_1 + \theta_2)\right)$.

Ejemplo 2.2: Consideremos el modelo $U(\theta-½,\theta+½)$, $\theta\in\mathbb{R}$. Dados dos valores θ_1 y θ_2, en este caso obtenemos:

$$J(\theta_1,\theta_2)=\begin{cases}-8\log(1+\theta_2-\theta_1) & \text{si } \theta_2\geq\theta_1 \text{ y } \mathrm{sop}(X_1)\cap \mathrm{sop}(X_2)\neq\varnothing\\ -8\log(1+\theta_1-\theta_2) & \text{si } \theta_1\geq\theta_2 \text{ y } \mathrm{sop}(X_1)\cap \mathrm{sop}(X_2)\neq\varnothing\\ +\infty & \text{si } \mathrm{sop}(X_1)\cap \mathrm{sop}(X_2)=\varnothing\end{cases}$$

donde $X_i\sim U(\theta_i-½,\theta_i+½)$, $i=1,2.$ Esta fórmula, puede ser resumida en

$$J(\theta_1,\theta_2)=\begin{cases}-8\log(1-|\theta_1-\theta_2|) & \text{si } |\theta_1-\theta_2|\leq 1\\ +\infty & \text{si } |\theta_1-\theta_2|>1.\end{cases}$$

2.2.3 Relación entre las Informaciones de Fisher y de Akahira.

En el caso de modelos regulares, existe una conexión entre ambas medidas de información reflejada en la siguiente proposición, cuya demostración puede verse en Akahira y Takeuchi, 1991.

Proposición: En los modelos regulares, para h suficientemente pequeño se verifica:

$$J(\theta,\theta+h)=I(\theta)h^2+o(h^2)\ .$$

A partir de la proposición, podemos establecer inmediatamente el siguiente corolario:

Corolario: En los modelos regulares, se verifica

$$I(\theta)=\lim_{h\to 0}\frac{J(\theta,\theta+h)}{h^2}\ .$$

De este corolario, podemos extraer las dos conclusiones siguientes:

a) Tenemos que $J(\theta,\theta+h)$ tiende a cero cuando h tiende a cero (para cualquier valor de θ) y además la velocidad de esta convergencia es del orden de h^2.

b) Dado un h pequeño y fijo, mientras mayor sea $I(\theta)$, "más distintas" serán $f(x\,|\,\theta)$ y $f(x\,|\,\theta+h)$ y, por tanto, mayor será la "capacidad de discriminación" entre θ y θ+h.

Ejemplo 2.1 (continuación): Consideremos el modelo Exponencial de parámetro θ>0, cuya función de densidad es $f(x\,|\,\theta)=\theta e^{-\theta x},\ x>0$.

Hemos obtenido $J(\theta_1,\theta_2)$, de donde deducimos que $J(\theta,\theta+h)=-8\log\left(2\sqrt{\theta(\theta+h)}\big/(2\theta+h)\right)$.

Además, puede comprobarse que la primera y segunda derivadas respecto de h son:

$$\frac{\partial J(\theta,\theta+h)}{\partial h}=J'(\theta,\theta+h)=\frac{4h}{(h+\theta)(h+2\theta)} \quad \text{y} \quad \frac{\partial^2 J(\theta,\theta+h)}{\partial h^2}=J''(\theta,\theta+h)=\frac{4(2\theta^2-h^2)}{(h+\theta)^2(h+2\theta)^2}.$$

Así, $\lim_{h\to 0}\frac{J(\theta,\theta+h)}{h}=\lim_{h\to 0}J'(\theta,\theta+h)=0$ y $\lim_{h\to 0}\frac{J(\theta,\theta+h)}{h^2}=\lim_{h\to 0}J''(\theta,\theta+h)=1/\theta^2=I_{X_1}(\theta)$.

Indiquemos que Pitman, 1979, adopta como definición de modelo regular la existencia de $\lim_{h\to 0}\left(J(\theta,\theta+h)/h^2\right)$, resaltando que para que se cumpla esta propiedad no es necesario que el recorrido de la variable no dependa del parámetro. En efecto, si consideramos el modelo cuya densidad es:

$$f(x\,|\,\theta)=\frac{1}{2}e^{(\theta-x)}(x-\theta)^2 \qquad x\geq\theta,$$

es fácil comprobar que se verifican:

1. $\lim_{h\to 0}\frac{J(\theta,\theta+h)}{|h|}=E\left[\frac{\partial\log f(x\,|\,\theta)}{\partial\theta}\right]=0.$

2. $\lim_{h\to 0}\frac{J(\theta,\theta+h)}{h^2}=1=E\left[\left(\frac{\partial\log f(x\,|\,\theta)}{\partial\theta}\right)^2\right]=-E\left[\frac{\partial^2\log f(x\,|\,\theta)}{\partial\theta^2}\right],$

y por tanto este modelo sería regular, aunque el recorrido dependa del parámetro.

Nota 1: La definición que hemos utilizado puede ampliarse de la forma:

$$J_{X_1}(\theta_1,\theta_2) = -\frac{8}{\left(1-\alpha^2\right)}\log\int f(x\,|\,\theta_1)^{\frac{1-\alpha}{2}}\, f(x\,|\,\theta_2)^{\frac{1+\alpha}{2}}\, dx\ , \quad -1<\alpha<1\,,$$

aunque nosotros sólo trabajaremos con el caso $\alpha = 0$.

Nota 2: Obsérvese cómo la información puede calcularse sin problemas en el caso de más de un parámetro. De hecho, la definición de información puede establecerse de forma más general usando directamente medidas de probabilidad. En concreto, en Akahira y Takeuchi, 1991, la definición que se ofrece es:

Dada una variable aleatoria X definida sobre un espacio muestral χ y P y Q medidas absolutamente continuas respecto a una medida σ-finita μ, definimos la cantidad de información entre P y Q como:

$$J(P,Q) = -8\log\int\left(\frac{dP}{d\mu}\right)^{1/2}\left(\frac{dQ}{d\mu}\right)^{1/2} d\mu\ .$$

La definición ofrecida inicialmente en este trabajo no es más que un caso particular, donde P y Q son las medidas de probabilidad inducidas por las variables correspondientes a los parámetros θ_1, θ_2, la medida σ-finita considerada es la de Lebesgue y las derivadas respecto a esta medida son las funciones de densidad de las variables.

3. Obtención de distribuciones a priori no informativas. Caso uniparamétrico.

3.1. Elección de la distribución a priori.

Tras observar la relación entre ambas medidas de información en los modelos regulares, nos podemos preguntar qué ocurriría en el caso de considerar un modelo no regular. Antes de pasar a resultados generales, vamos a ver un ejemplo.

Ejemplo 3.1: Consideremos el modelo $U(0,\theta),\ \theta \in (0,+\infty)$, con función de densidad $f(x\,|\,\theta) = \theta^{-1},\quad 0 \le x \le \theta$.

Como es conocido, este modelo es no regular, puesto que

$$\int \frac{\partial}{\partial\theta} f(x,\theta)dx = -\frac{1}{\theta} \neq 0 = \frac{\partial}{\partial\theta}\int f(x,\theta)dx \ .$$

Para h>0 obtendremos $J(\theta,\theta+h) = -4\log\left(\theta/\theta+h\right)$ mientras que para h<0 se tendrá $J(\theta,\theta+h) = -4\log\left(\theta/\theta+h\right)$, por lo que en este modelo se verifica

$$\lim_{h\to 0}\frac{J(\theta,\theta+h)}{h^2} = +\infty \ .$$

Es decir, J(θ,θ+h) converge a cero cuando h tiende a cero, pero la velocidad de esta convergencia es inferior a la de h^2. Sin embargo, podemos comprobar que dicha convergencia es tan rápida como la de $|h|$. En efecto, cálculos elementales de límites nos llevan a:

$$\lim_{h\to 0^+}\frac{J(\theta,\theta+h)}{|h|} = \lim_{h\to 0^-}\frac{J(\theta,\theta+h)}{|h|} = 4/\theta \ .$$

Hemos visto así que en ambos casos (modelo regular y no regular) se tiene $\lim_{h\to 0} J(\theta,\theta+h) = 0$ (recordar que $J(\theta,\theta) = 0$), si bien la velocidad de esta convergencia es más rápida en los modelos regulares que en el modelo uniforme.

Sabemos que en los modelos regulares (con un único parámetro) la distribución a priori no informativa comúnmente aceptada es la de Jeffreys, a saber, $\pi(\theta) \propto \left(I(\theta)\right)^{1/2}$ que podemos escribir a partir del corolario como

$$\pi(\theta) \propto \left(\lim_{h\to 0}\frac{J(\theta,\theta+h)}{|h|^2}\right)^{1/2} \ .$$

Según el ejemplo visto anteriormente de la distribución uniforme, y puesto que la convergencia es del orden de $|h|$, proponemos como distribución a priori para el parámetro θ

$$\pi(\theta) \propto \lim_{h\to 0}\frac{J(\theta,\theta+h)}{|h|} \ .$$

De forma global, nuestra propuesta es la siguiente:

1. Obtener k tal que

$$\lim_{h\to 0}\left(\frac{J(\theta,\theta+h)}{h^k}\right)=C(\theta),$$

donde C(θ) es una función que puede ser constante (pero no idénticamente nula ni infinito).

2. Elegir como distribución a priori

$$\pi(\theta)\propto \lim_{h\to 0}\left(\frac{J(\theta,\theta+h)}{|h|^k}\right)^{1/k}.$$

Ejemplo 3.1. (continuación): En el ejemplo anterior de la distribución uniforme en (0,θ), como obtuvimos que el límite era 4/θ, tendremos $\pi(\theta)\propto\theta^{-1}$ que es la distribución que se acepta comúnmente como no informativa para este modelo, y que coincide con la distribución a priori de referencia de Bernardo y Smith, 1994 y con la distribución imparcial de Basulto, 1997.

Ejemplo 3.2: Consideremos el modelo $U(\theta-1/2,\theta+1/2),\ \theta\in\mathbb{R}$, con función de densidad

$$f(x,\theta)=1,\quad \theta-\frac{1}{2}\le x\le\theta+\frac{1}{2}.$$

Para h>0, obtenemos $J(\theta,\theta+h)=-8\log(1-h)$, mientras que para h<0 la información es $J(\theta,\theta+h)=-8\log(1+h)$. Puede comprobarse sin dificultad que también en este caso la velocidad de convergencia a 0 es del orden de $|h|$ y que se verifica

$$\lim_{h\to 0}\frac{J(\theta,\theta+h)}{|h|}=8,$$

y por tanto, la distribución a priori sería para este caso $\pi(\theta)\propto 1$, que coincide con la distribución a priori de referencia de Bernardo y Smith, 1994 y con la distribución a priori imparcial de Basulto, 1997.

3.2. Expresión alternativa para la obtención de la distribución a priori.

En esta sección vamos a considerar una familia de modelos para los cuales vamos a deducir una expresión alternativa de la distribución a priori, que permite su cálculo con mayor facilidad. Sean $X_1,...,X_n$ independientes e idénticamente distribuidas con distribución P_θ y con densidad f(x|θ) respecto a la medida de Lebesgue en $\mathbb{R}$, donde $\theta \in \Theta \subseteq \mathbb{R}$ siendo Θ abierto. Suponemos que:

i) ∀θ∈Θ, f(•|θ) es estrictamente positiva en un intervalo cerrado (acotado o no) $S(\theta) = \left[a_1(\theta), a_2(\theta)\right]$ y vale cero fuera de él. Está permitido que uno de los extremos sea constante y puede ser más o menos infinito.

ii) Los conjuntos S(θ) son crecientes o decrecientes en θ (es decir, que $\forall\ \theta_1 < \theta_2$ se verifica $S(\theta_1) \subseteq S(\theta_2)$ o bien $S(\theta_1) \supseteq S(\theta_2)$).

iii) Las funciones $a_1(\theta)$ y $a_2(\theta)$ son estrictamente monótonas y continuamente diferenciables a menos que sean constantes o valgan más o menos infinito.

iv) En el conjunto $R(x,\theta) = \left\{(x,\theta) \in \mathbb{R}^2 : x \in S(\theta)\right\}$, tanto $f(x\,|\,\theta)$ como $\partial f(x\,|\,\theta)/\partial\theta$ son conjuntamente continuas en (x,θ).

Proposición. Bajo las condiciones señaladas anteriormente, se verifica:

$$\lim_{h\to 0}\frac{J(\theta,\theta+h)}{|h|} \quad = \quad \left|4E\left[\frac{\partial \log f(x\,|\,\theta)}{\partial\theta}\right]\right|.$$

Como consecuencia de la proposición, la distribución a priori que elegiríamos en este caso sería:

$$\pi(\theta) \propto \left|E\left[\frac{\partial \log f(x\,|\,\theta)}{\partial\theta}\right]\right|.$$

Dmt:

Para la demostración, vamos a suponer, sin pérdida de generalidad, que los conjuntos S(θ) son crecientes en θ; así, $a_1'(\theta) < 0$ y $a_2'(\theta) > 0$ (a menos que alguno de los extremos de los intervalos sea constante). En esta situación, la función $\log f(x|\theta)$ es decreciente en θ, y por tanto, debemos probar que $\pi(\theta) \propto -E\left[\partial \log f(x|\theta)/\partial\theta\right]$. En el caso contrario, la demostración sería similar, aunque obtendríamos $\pi(\theta) \propto E\left[\partial \log f(x|\theta)/\partial\theta\right]$, puesto que la función $\log f(x|\theta)$ sería creciente en θ.

A) Vamos a demostrar que $\lim_{h\to 0^+} \frac{J(\theta,\theta+h)}{h} = -4E\left[\frac{\partial \log f(x|\theta)}{\partial\theta}\right]$. Sea h>0.

Definimos $g_\theta(x,h) = \sqrt{f(x|\theta) f(x|\theta+h)}$ y $G_\theta(h) = \int_{a_1(\theta)}^{a_2(\theta)} g_\theta(x,h)dx$. Puesto que $G_\theta(h)$ será continua en un intervalo $[0,\varepsilon]$, se verifica $\lim_{h\to 0} G_\theta(h) = G_\theta(0) = 1$, y por tanto,

$$\lim_{h\to 0}\frac{J(\theta,\theta+h)}{h} = \lim_{h\to 0}\frac{-8\log G_\theta(h)}{h} = -8\lim_{h\to 0}\frac{\partial \log G_\theta(h)}{\partial h} = -8\lim_{h\to 0}\frac{\frac{\partial G_\theta(h)}{\partial h}}{G_\theta(h)} = -8\lim_{h\to 0}\frac{\int_{a_1(\theta)}^{a_2(\theta)}\frac{\partial g_\theta(x,h)}{\partial h}}{G_\theta(h)} =$$

$$= -8\lim_{h\to 0}\frac{\int_{a_1(\theta)}^{a_2(\theta)}\sqrt{f(x|\theta)}\frac{\partial f(x|\theta+h)/\partial h}{2\sqrt{f(x|\theta+h)}}dx}{G_\theta(h)} = -4\int_{a_1(\theta)}^{a_2(\theta)}\frac{\partial f(x|\theta)}{\partial\theta}dx = -4E\left[\frac{\partial \log f(x|\theta)}{\partial\theta}\right].$$

B) Vamos a demostrar que $\lim_{h\to 0^-} \frac{J(\theta,\theta+h)}{-h} = -4E\left[\frac{\partial \log f(x|\theta)}{\partial\theta}\right]$. Sea h>0.

Definamos ahora $q_\theta(x,h) = \sqrt{f(x|\theta) f(x|\theta-h)}$ y $Q_\theta(h) = \int_{a_1(\theta-h)}^{a_2(\theta-h)} q_\theta(x,h)dx$. Puesto que $G_\theta(h)$ será continua en un intervalo $[0,\varepsilon]$, se verifica $\lim_{h\to 0} G_\theta(h) = G_\theta(0) = 1$, y por tanto,

$$\lim_{h\to 0}\frac{J(\theta,\theta+h)}{-h} = \lim_{h\to 0}\frac{8\log Q_\theta(h)}{h} = 8\lim_{h\to 0}\frac{\partial \log Q_\theta(h)}{\partial h} = 8\lim_{h\to 0}\frac{\frac{\partial Q_\theta(h)}{\partial h}}{Q_\theta(h)} = 8\left.\frac{\partial Q_\theta(h)}{\partial h}\right|_{h=0}.$$

En este caso, puesto que los límites de integración dependen de h, para calcular $\partial Q_\theta(h)/\partial h$, aplicamos la fórmula de Leibniz (Apóstol, 1960), obteniendo:

$$\frac{\partial Q_\theta(h)}{\partial h} = \int_{a_1(\theta-h)}^{a_2(\theta-h)}\sqrt{f(x|\theta)}\frac{\partial\sqrt{f(x|\theta-h)}}{\partial h}dx + a_2^{'}(\theta-h)\sqrt{f(a_2(\theta-h)|\theta)}\sqrt{f(a_2(\theta-h)|\theta-h)} -$$
$$- a_1^{'}(\theta-h)\sqrt{f(a_1(\theta-h)|\theta)}\sqrt{f(a_1(\theta-h)|\theta-h)}$$

y así obtenemos:

$$\lim_{h\to 0}\frac{J(\theta,\theta+h)}{h}=8\left\{\int_{a_1(\theta)}^{a_2(\theta)}\sqrt{f(x\mid\theta)}\,\frac{\partial\sqrt{f(x\mid\theta)}}{\partial\theta}\,dx+a_2^{'}(\theta)f(a_2(\theta)\mid\theta)-a_1^{'}(\theta)f(a_1(\theta)\mid\theta)\right\}.$$

Aplicando de nuevo la fórmula de Leibniz, obtenemos que:

$$a_2^{'}(\theta)f(a_2(\theta)\mid\theta)-a_1^{'}(\theta)f(a_1(\theta)\mid\theta)=-\int_{a_1(\theta)}^{a_2(\theta)}\frac{\partial f(x\mid\theta)}{\partial\theta}dx$$

y por tanto:

$$\lim_{h\to 0}\frac{J(\theta,\theta+h)}{h}=4\int_{a_1(\theta)}^{a_2(\theta)}\frac{\partial f(x\mid\theta)}{\partial\theta}dx-8\int_{a_1(\theta)}^{a_2(\theta)}\frac{\partial f(x\mid\theta)}{\partial\theta}dx=-4E\left[\frac{\partial\log f(x\mid\theta)}{\partial\theta}\right].$$

Nota: Los modelos más importantes que pertenecen a esta familia son:

1. La familia de localización: $f(x,\theta)=f_0(x-\theta),\ \theta\in\mathbb{R},$ donde $f_0(z)$ es una densidad en el intervalo $[0,+\infty)$. En este caso, $a_1(\theta)=\theta$ y $a_2(\theta)\equiv+\infty$.

2. $f(x\mid\theta)=c(x)/g(\theta),\quad a(\theta)\le x\le b(\theta)$ (donde las funciones verifican las hipótesis especificadas anteriormente).

Observemos que esta familia incluye a gran cantidad de modelos, entre los que podemos destacar los modelos uniformes con soporte en (i) [θ,θ], θ>0, (ii) [-θ,θ],θ>0, (iii)[θ,1/θ], 0<θ<1, así como la familia truncada $f(x,\theta)=g(x)/G(\theta),\quad x>\theta,$ donde g(•) es una densidad en $(0,+\infty)$ y $G(x)=\int_x^{+\infty}g(t)dt$. Indiquemos que el modelo de Pareto pertenece a la familia truncada, ya que en este caso la función de densidad es $f(x,\theta)=\alpha\theta^{\alpha}x^{-(1+\alpha)},\ x>\theta$ y así estamos en la situación descrita tomando $g(x)=\alpha x^{1-\alpha}$ y $G(\theta)=\int_x^{+\infty}g(t)dt=\theta^{-\alpha}$.

Señalemos también que modelos tales como los uniformes en $[\theta-½,\theta+½]$ o en $[\theta,2\theta]$ no están en esta familia por no ser los soportes ni crecientes ni decrecientes en θ (es decir, dados $\theta_1<\theta_2$ en general no se verifica ni $sop(\theta_1)\subseteq sop(\theta_2)$ ni $sop(\theta_1)\supseteq sop(\theta_2)$).

Ejemplo 3.3: Consideremos la familia de localización: $f(x,\theta) = f_0(x-\theta),\ \theta \in \mathbb{R}$, donde $f_0(z)$ es una densidad en el intervalo $[0,+\infty)$.

En este caso, al aplicar el resultado obtenido en la proposición tenemos lo siguiente:

$$\frac{\partial \log f(x,\theta)}{\partial\theta} = \frac{-f_0^{'}(x-\theta)}{f_0(x-\theta)} \quad \Rightarrow E\left[\frac{\partial \log f(x,\theta)}{\partial\theta}\right] = -\int_0^{+\infty} f_0^{'}(z)dz \ = f_0(0+).$$

De esta forma, la distribución a priori sería $\pi(\theta) \propto 1$ siempre que $f_0(0+) \neq 0$, en cuyo caso lo que ocurre es que la convergencia de la información es de orden dos, con lo que tendríamos que calcular el límite dividiendo por h^2.

3.3 Propiedades de la distribución a priori propuesta.

3.3.1 Invarianza ante reparametrizaciones.

Proposición: La distribución a priori obtenida según la propuesta anterior es invariante ante reparametrizaciones biyectivas.

Dmt:

La demostración de la invarianza es consecuencia de que la definición de la distribución a priori se basa en una medida de información que es una característica intrínseca de las distribuciones de probabilidad, independiente de la parametrización elegida. Así, dados dos valores θ_1, θ_2 y siendo $\lambda = \lambda(\theta)$ biyectiva (continua y diferenciable), se verificará $J(\theta_1,\theta_2) = J(\lambda_1,\lambda_2)$, donde $\lambda_i = \lambda(\theta_i), i = 1,2$.

Si consideramos ahora los valores θ e $\Delta\theta$, y siendo $\lambda = \lambda(\theta)$ e $\Delta\lambda = \lambda(\theta+\Delta\theta) - \lambda(\theta)$, tendremos que $J(\theta,\theta+\Delta\theta) = J(\lambda(\theta),\lambda(\theta+\Delta\theta)) = J(\lambda,\lambda+\Delta\lambda)$. Por tanto,

$$\frac{J(\theta,\theta+\Delta\theta)}{\Delta\theta} = \frac{J(\lambda,\lambda+\Delta\lambda)}{\Delta\lambda}\frac{\Delta\lambda}{\Delta\theta} = \frac{J(\lambda,\lambda+\Delta\lambda)}{\Delta\lambda}\frac{\lambda(\theta+\Delta\theta)-\lambda(\theta)}{\Delta\theta},$$

y tomando límite cuando $\Delta\theta \to 0$, obtenemos

$$\pi(\theta) = \pi(\lambda)\left|\frac{\partial\lambda}{\partial\theta}\right|,$$

lo cual concluye la demostración.

3.3.2 Propiedades frecuencialistas de los intervalos Bayesianos.

Uno de los argumentos más usados en la literatura para construir distribuciones a priori no informativas (o para decidir si una determinada distribución a priori no informativa es una elección buena) es poder calcular con dichas distribuciones intervalos bayesianos de probabilidad 1-α cuyo nivel de confianza, en el sentido de la estadística clásica, sea también 1-α (o al menos, de forma aproximada).

El primer trabajo que puede considerarse en este sentido es el de Welch y Peers ,1963, en el que se demuestra que en modelos regulares y con un sólo parámetro la distribución de Jeffreys es la única que verifica $P\left[\theta < g(S,\alpha) \mid \theta\right] = 1-\alpha + O(n^{-1})$ donde $g(S,\alpha)$ es el extremo superior del intervalo bayesiano unilateral de probabilidad 1-α obtenido a partir de una muestra S de la variable X , es decir, $P\left[\theta < g(S,\alpha) \mid S\right] = 1-\alpha$, o lo que es lo mismo, $g(S,\alpha)$ es el percentil de orden $1-\alpha$ de la distribución a posteriori de θ dada la muestra S.

En Ghosal, 1999, se demuestra que, bajo las condiciones descritas en el epígrafe 3.2, cualquier distribución a priori diferenciable lleva a intervalos unilaterales con probabilidad de cubrimiento en sentido frecuencialista $1-\alpha+O(n^{-1})$; sin embargo, también se establece en dicho artículo que la única distribución a priori que verifica que los intervalos bayesianos unilaterales de probabilidad $1-\alpha$ tienen probabilidad de cubrimiento $1-\alpha+O(n^{-2})$, es la obtenida según nuestra propuesta.

Hagamos notar que en esta situación no es aconsejable trabajar con intervalos bilaterales, ya que tanto la distribución a posteriori como la distribución muestral de $\hat{\theta}$, en el límite, son muy asimétricas y están muy concentradas en uno de los extremos de su recorrido.

Aunque este resultado general es bastante importante, existen varios ejemplos en los que la coincidencia de resultados entre la inferencia bayesiana y clásica es exacta (Ortega y Basulto, 2003)

4. Distribuciones multiparamétricas.

Cuando hay más de un parámetro, es decir, cuando $\theta \in \Theta \subseteq \mathbb{R}^m$, la regla general de Jeffreys, aplicable al caso regular, consiste en tomar $\pi(\theta) \propto \sqrt{|I(\theta)|}$, donde $|I(\theta)|$ representa el determinante de la matriz de información, lo que sigue asegurando la invarianza ante reparametrizaciones arbitrarias. No obstante, esta opción presenta deficiencias importantes, que hacen que en muchos casos no sea la opción habitualmente elegida (Jeffreys, 1961, Ortega y Basulto, 2003). Para evitar estos inconvenientes, Jeffreys sugirió una modificación para su regla general en el caso multiparamétrico que debía aplicarse en los modelos con parámetros de localización y escala y que resulta equivalente a obtener la distribución de cada parámetro suponiendo que los otros son fijos y posteriormente, la distribución multiparamétrica será el producto de las correspondientes unidimensionales (Jeffreys, 1961, p.182-183).

El camino más usual para la construcción de distribuciones a priori multidimensionales consiste en obtener las mismas a partir de ciertas distribuciones unidimensionales (marginales o condicionadas), dependiendo el proceso seguido y la distribución a priori obtenida de si se considera que alguno de los parámetros es el de interés, siendo el resto parámetros "perturbadores", o si, por el contrario, se considera que todos los parámetros son de interés (Bernardo y Smith, 1994, Nicolau, 1993).

Supongamos por simplicidad que $\theta = (\theta_1, \theta_2) \in \Theta \subseteq \mathbb{R}^2$; además, vamos a suponer también que en nuestro problema ambos parámetros son de interés. En este caso, un camino posible es obtener las distribuciones condicionadas $\pi_{1|2}(\theta_1|\theta_2)$ y $\pi_{2|1}(\theta_2|\theta_1)$ aplicando la regla de obtención de distribuciones unidimensionales y buscar posteriormente una distribución conjunta $\pi(\theta_1, \theta_2)$ compatible con ambas condicionadas, que no siempre tiene por qué existir (Arnold y otros, 1999). Un problema importante de este tipo de procedimientos es que, por ejemplo, la distribución $\pi_{1|2}(\theta_1|\theta_2)$ quedará determinada salvo una constante arbitraria que puede depender de θ_2, y que posteriormente influye en la obtención de la distribución conjunta.

De esta forma, al aplicar una regla uniparamétrica al parámetro θ_1 suponiendo que θ_2 es fijo, obtendremos una expresión del tipo $\pi_{1|2}(\theta_1 | \theta_2) \propto g_{1|2}(\theta_1, \theta_2) C_{1|2}(\theta_2)$, donde $C_{1|2}(\theta_2)$ es una función arbitraria. De forma análoga, tendremos $\pi_{2|1}(\theta_2 | \theta_1) \propto g_{2|1}(\theta_1, \theta_2) C_{2|1}(\theta_1)$. En el caso particular de que se verifiquen:

$$g_{1|2}(\theta_1, \theta_2) C_{1|2}(\theta_2) = h_{1|2}(\theta_1) h^*_{1|2}(\theta_2) \ \text{ y } \ g_{2|1}(\theta_1, \theta_2) C_{2|1}(\theta_1) = h_{2|1}(\theta_2) h^*_{2|1}(\theta_1)$$

y siguiendo a Nicolau, 1993 y Ghosal, 1999, proponemos elegir $\pi(\theta_1, \theta_2) \propto h_{1|2}(\theta_1) h_{2|1}(\theta_2)$. Observemos que en este caso, como indica Nicolau, 1993, podemos considerar que ambos parámetros son independientes a priori; la elección de tomar ambas constantes arbitrarias iguales a 1 es la que conlleva a que las distribuciones condicionadas coincidan con las correspondientes marginales.

Aunque la solución ofrecida es parcial, es necesario observar que esta situación es interesante y muy frecuente en la práctica (Ghosal, 1999) , sobre todo cuando en el modelo uno de los parámetros es regular y el otro no regular.

Ejemplo 4.1: Consideremos el modelo de Pareto, cuya densidad viene dada por $f(x | \eta, \varphi) = \varphi \eta^{\varphi} x^{-(1+\varphi)}$, $x > \eta$, $\eta, \varphi > 0$. Si consideramos que η es conocido (y por tanto fijo), el modelo cumple las condiciones de regularidad; diremos que el modelo es regular con respecto al parámetro φ o que “el parámetro φ es regular”. Por el contrario, si consideramos conocido el valor de φ, el modelo no verifica las hipótesis de regularidad, por lo que diremos que “el parámetro η es no regular”.

Ahora, podemos obtener los núcleos de las distribuciones a priori condicionadas, aplicando la regla univariante. Sea $\ell(\eta, \varphi | x) = \log f\left(x | \eta, \varphi\right)$.

A) Si consideramos η conocido, por ser φ regular, la distribución a priori de $\varphi | \eta$ será proporcional a la Información de Fisher del modelo (cuando η es conocido). Es fácil comprobar que $E\left[\partial^2 \ell / \partial \varphi^2\right] = -\varphi^{-2}$, y por tanto, obtendremos $\pi\left(\varphi | \eta\right) \propto \varphi^{-1} C_1(\eta)$.

B) Si consideramos φ conocido, el modelo es no regular y verifica las condiciones señaladas en la sección 3.2. Teniendo en cuenta que $E\left[\partial\ell/\partial\eta\right]=\varphi/\eta$, obtendremos $\pi(\eta\,|\,\varphi)\propto\eta^{-1}C_2(\varphi)$.

Como puede apreciarse, ambas condicionadas pueden descomponerse en producto de funciones que dependen cada una de un solo parámetro, por lo que la distribución a priori conjunta que proponemos es $\pi\left(\eta,\varphi\right)\propto\eta^{-1}\varphi^{-1}$.

Ejemplo 4.2: Consideremos la familia de localización-escala:

$f(x,\theta)=\varphi^{-1}f_0((x-\theta)/\varphi),\ \theta\in\mathbb{R},\varphi>0$ donde $f_0(z)$ es una densidad en el intervalo $[0,+\infty)$. Supongamos que el parámetro φ es regular y que $f_0(0+)\neq 0$, es decir, que el parámetro θ es no regular. En este caso, se verifican:

$$E\left[\frac{\partial\ell}{\partial\theta}\right]=\varphi^{-1}f_0(0+);\qquad -E\left[\frac{\partial^2\ell}{\partial\varphi^2}\right]=k\varphi^{-2},\ \text{donde } k=\int\left[1+xf_0^{'}(x)/f_0(x)\right]^2 f_0(x)dx$$

por lo que $\pi\left(\theta\,|\,\varphi\right)\propto 1{\cdot}C_1(\varphi)$ y $\pi\left(\varphi\,|\,\theta\right)\propto\varphi^{-1}{\cdot}C_2(\theta)$, es decir, las condicionadas pueden descomponerse en producto de funciones dependientes de cada uno de los parámetros y por lo tanto la distribución a priori conjunta será $\pi\left(\theta,\varphi\right)\propto\varphi^{-1}$.

En el caso de que no se cumpla esta propiedad, siempre quedaría la opción de buscar una reparametrización para la cual sí se verifique.

Ejemplo 4.3: Consideremos el modelo Uniforme en (α,β), cuya densidad viene dada por $f(x\,|\,\alpha,\beta)=(\beta-\alpha)^{-1},\ \alpha\le x\le\beta,\ \alpha<\beta\in\mathbb{R}$. En este caso, ambos parámetros son no regulares, verificándose las condiciones de la sección 3.2 si consideramos alguno de los parámetros fijo. Siendo $\ell(\alpha,\beta\,|\,x)=\log f(x\,|\,\alpha,\beta)$, podemos comprobar fácilmente que $\left|E\left[\partial\ell/\partial\alpha\right]\right|=\left|E\left[\partial\ell/\partial\beta\right]\right|=(\beta-\alpha)^{-1}$, con lo que obtendríamos $\pi(\alpha\,|\,\beta)\propto(\beta-\alpha)^{-1}C_1(\beta)$ y $\pi(\beta\,|\,\alpha)\propto(\beta-\alpha)^{-1}C_2(\alpha)$, es decir, las densidades a priori condicionadas no pueden descomponerse en producto de funciones dependientes de un solo parámetro.

Ahora bien, podemos considerar la reparametrización $\mu=\alpha,\ \sigma=\beta-\alpha$, es decir, los parámetros pasan a ser el punto inicial de la distribución y la longitud del intervalo (obsérvese que μ es un parámetro de localización y σ de escala). Ahora, $\ell(\mu,\sigma \mid x)=\log(\sigma^{-1}),\ \mu \le x \le \mu+\sigma$. Si consideramos σ fijo, estamos básicamente en la situación del ejemplo 3.2, obteniéndose $\pi(\mu \mid \sigma) \propto 1 \bullet C_1(\sigma)$, mientras que al considerar μ fijo estaríamos básicamente en el caso del ejemplo 3.1, obteniendo $\pi(\sigma \mid \mu) \propto \sigma^{-1} C_2(\mu)$. Por tanto, la distribución a priori conjunta que proponemos en este modelo es $\pi(\mu,\sigma) \propto \sigma^{-1}$ (que, deshaciendo el cambio, resulta ser $\pi(\alpha,\beta) \propto (\beta-\alpha)^{-1}$).

5. Un modelo de búsqueda de empleo.

En Lancaster, 1997, se propone un modelo de búsqueda de empleo adecuado a la situación en la que los datos observados son la duración de la búsqueda (T) y el salario aceptado (W), cuya verosimilitud para una muestra $((t_1,w_1),\dots,(t_n,w_n))$ vendría dada por:

$$L(\theta,\lambda,\xi)=\lambda^n \exp\left\{-\lambda \overline{F}(\xi \mid \theta) T\right\} \prod_{i=1}^{n} f\left(w_i \mid \theta\right), \quad b<\xi<w$$

donde λ es la tasa de aparición de ofertas de trabajo, $T=\sum_{i=1}^{n} t_i$, $f(w \mid \theta)$ es la función de densidad de la oferta salarial W (θ puede ser de dimensión superior a uno), $\overline{F}(w \mid \theta)$ es la función de supervivencia de W, ξ es el salario de reserva, es decir el valor salarial por debajo del cual el trabajador no acepta la oferta de trabajo y b es la tasa de beneficio del desempleo, cuyo valor se supone conocido.

Si aceptamos la hipótesis de que los desempleados actúan buscando maximizar la esperanza de sus ingresos bajo un horizonte infinito, entonces, tendríamos la restricción:

$$\xi=b+\frac{\lambda}{\rho}\int_{\xi}^{+\infty} \overline{F}(w \mid \theta)dw ,$$

donde ρ es la tasa de descuento, que puede suponerse conocida (si fuese desconocida habríamos de añadir un parámetro más al modelo). Esta restricción, puede expresarse en la forma $\xi=g(\theta,\lambda)$ y así la restricción de la verosimilitud quedaría de la forma

$b < g(\theta,\lambda) < w$. En este caso, el parámetro ξ "desaparecería" en el modelo, pues lo sustituiríamos por su expresión $g(\theta,\lambda)$. Aun cuando los parámetros θ y λ tuviesen un comportamiento regular en el modelo inicial, bajo esta hipótesis pasarían a tener, en general, un comportamiento no regular (debido a que el recorrido de la variable W dependería de ambos parámetros). Si abandonamos la restricción debida a la optimalidad en el comportamiento, el parámetro ξ deja de estar relacionado funcionalmente con (θ,λ) y pasa a ser un parámetro más del modelo. Observemos que en el caso de que el modelo de las ofertas salariales $f(w\,|\,\theta)$ fuese regular, tendríamos que el vector paramétrico (θ,λ) tendría un comportamiento regular, mientras que ξ sería no regular.

Vamos a suponer que la distribución de las ofertas salariales es uniforme en (α,β), con ambos extremos desconocidos (en Lancaster, 1997 se elige una distribución log-Normal). En este caso, podemos simplificar el modelo, pues el salario aceptado, que es superior al salario de reserva ξ, se distribuirá uniformemente en el intervalo (ξ,β), que no depende del extremo inferior del intervalo α (observemos que, en general, si la distribución de las ofertas salariales depende de dos parámetros α y β, entonces la distribución truncada del salario aceptado dependerá de α, β y ξ). Haciendo la reparametrización $\sigma = \beta - \xi$, el salario aceptado sigue una distribución $U(\xi, \xi+\sigma)$.

CASO A: El parámetro ξ no está relacionado funcionalmente con σ.

En este caso, $f(w\,|\,\xi,\sigma) = \sigma^{-1}$, $\xi < w < \xi + \sigma$, $\overline{F}(\xi\,|\,\sigma) = 1$, y la verosimilitud sería:

$$L(\lambda,\xi,\sigma) = \lambda^n \exp\{-\lambda T\}\sigma^{-n}, \quad b < \xi < w_{(1)},\ \xi+\sigma > w_{(n)},$$

donde $w_{(1)}$ y $w_{(n)}$ son, respectivamente, los valores mínimo y máximo de $\{w_1,...,w_n\}$.

Para obtener la distribución a priori conjunta, teniendo en cuenta que el logaritmo de la verosimilitud para una muestra de tamaño 1 sería $\ell(\lambda,\xi,\sigma) = \mathrm{Log}\lambda - \lambda t - \mathrm{Log}\sigma$, $\xi < w < \xi + \sigma$, aplicamos las reglas de obtención de las distribuciones condicionadas unidimensionales. El parámetro λ es regular y, puesto que $-\partial^2\ell/\partial\lambda^2 = \lambda^{-2}$, obtenemos $\pi(\lambda\,|\,\xi,\sigma) \propto \lambda^{-1}C_1(\xi,\sigma)$. Los parámetros ξ y σ tienen comportamiento no regular; Como en el ejemplo 4.3, obtendremos $\pi\left(\xi \mid \lambda, \sigma\right) \propto 1 \cdot C_2(\lambda, \sigma)$ y $\pi\left(\sigma \mid \lambda, \xi\right) \propto \sigma^{-1} \cdot C_3(\lambda, \xi)$. Así, estamos en la situación en la que podemos considerar los parámetros independientes a priori, siendo la distribución conjunta:

$$\pi\left(\lambda, \xi, \sigma\right) \propto \lambda^{-1} \sigma^{-1}.$$

Con esta distribución a priori, podemos llevar a cabo nuestra inferencia algebraicamente y de forma exacta. En efecto, la distribución a posteriori es

$$\pi(\lambda, \xi, \sigma \mid \vec{t}, \vec{w}) = \lambda^{n-1} \exp\left\{-\lambda T\right\} \sigma^{-(n+1)}, \quad b < \xi < w_{(1)},\ \xi + \sigma > w_{(n)},$$

donde $\vec{t}$ y $\vec{w}$ representan los tiempos de búsqueda y los salarios observados respectivamente. Esta distribución es propia; concretamente, la constante de integración viene dada por:

$$K = nT^{n} \left((w_{(n)} - b)^{-(n-1)} - r^{-(n-1)} \right) \Big/ (n-2)!,$$

donde $r = w_{(n)} - w_{(1)}$ representa el recorrido de los salarios aceptados. Las distribuciones marginales a posteriori de los parámetros también pueden obtenerse explícitamente. Concretamente, la distribución a posteriori del parámetro λ resulta ser una Gamma de parámetros (n,T); La densidad a posteriori de ξ viene dada por $\pi(\xi \mid \vec{t}, \vec{w}) \propto (w_{(n)} - \xi)^{-n}, \quad b < \xi < w_{(1)}$ mientras que la distribución marginal de σ es:

$$\pi(\sigma \mid \vec{t}, \vec{w}) \propto \begin{cases} \sigma^{-(n+1)}(\sigma - r) & \text{si} \quad r \le \sigma \le w_{(n)} - b \\ \sigma^{-(n+1)} & \text{si} \quad w_{(n)} - b \le \sigma < +\infty \end{cases}.$$

Las constantes de integración de las densidades marginales de ξ y σ también pueden calcularse sin mayor dificultad.

CASO B: El parámetro ξ está relacionado funcionalmente con σ.

Según hemos indicado, la relación funcional sería $\xi = b + \dfrac{\lambda}{\rho} \displaystyle\int_{\xi}^{+\infty} \overline{F}(w \mid \theta) dw$, donde tanto b como ρ son conocidos. Siendo la distribución de las ofertas salariales uniforme, es fácil comprobar que $\int_{\xi}^{+\infty} \overline{F}(w \mid \theta)dw = \sigma / 2$, por lo que la relación entre los parámetros es $\xi = b + \lambda\sigma/(2\rho)$. Así, el logaritmo de la verosimilitud para una muestra de tamaño 1 sería $\ell(\lambda, \sigma) = \text{Log}\lambda - \lambda t - \text{Log}\sigma,\ \ b + \lambda\sigma/(2\rho) < w < b + \lambda\sigma/(2\rho) + \sigma$.

Para obtener la distribución de $\lambda \mid \sigma$, observemos que $\partial\ell/\partial\lambda = \lambda^{-1} - t$, y que bajo el supuesto en el que estamos trabajando, T sigue una distribución exponencial de parámetro λ, por lo que $E\left[\partial\ell/\partial\lambda\right] = \lambda^{-1} - E[T] = \lambda^{-1} - \lambda^{-1} = 0$. Es decir, aunque el recorrido de W depende del parámetro λ, el comportamiento es regular. Por tanto, tomaremos $\pi\left(\lambda \mid \sigma\right) \propto \sqrt{-E\left[\partial^2\ell/\partial\lambda^2\right]}$, es decir, $\pi\left(\lambda \mid \sigma\right) \propto \lambda^{-1} C_1(\sigma)$. Por otra parte, $E\left[\partial\ell/\partial\sigma\right] = \sigma^{-1}$, y por tanto $\pi\left(\sigma \mid \lambda\right) \propto \sigma^{-1} C_2(\lambda)$. En definitiva, la distribución a priori conjunta resultaría ser $\pi\left(\lambda, \sigma\right) \propto \lambda^{-1}\sigma^{-1}$, que como vemos es la misma que se obtuvo en el caso A. Indiquemos que las distribuciones marginales a posteriori también pueden obtenerse de manera similar.

6. Conclusiones y cuestiones abiertas.

La propuesta de obtención de distribuciones a priori no informativas para modelos no regulares uniparamétricos, se muestra bastante satisfactoria, ya que a su sencillez de cálculo se unen sus buenas propiedades de invarianza ante reparametrizaciones y de "comportamiento frecuencialista" de los intervalos de probabilidad bayesianos que se obtienen a partir de ella.

En el caso de más de un parámetro, hemos ofrecido una solución parcial satisfactoria para un conjunto de modelos importante. No obstante, permanece abierto el problema general de obtención de la distribución conjunta en cualquier situación; debemos resaltar que este problema permanece abierto aún en el caso de modelos regulares. Por ejemplo, la solución ofrecida en Bernardo y Smith (1994) depende de la ordenación que se considere sobre el conjunto de parámetros; si el vector (θ_1, θ_2) "está ordenado", es decir θ_1 es el parámetro de interés mientras que θ_2 es un parámetro perturbador, estos autores sugieren obtener la distribución conjunta como $\pi(\theta_2 \mid \theta_1)\pi(\theta_1)$; en Nicolau, 1993, se considera el caso de parámetros ortogonales y se propone una distribución de la que el autor dice (pág. 378): "esencialmente, esta distribución es sólo la distribución a priori del parámetro de interés dado que el parámetro perturbador está determinado", puesto que se considera una solución del tipo $\pi(\theta_1 | \theta_2)\pi(\theta_2)$, donde $\pi(\theta_2)$ puede elegirse arbitrariamente.

Para el modelo de búsqueda de empleo, se ofrece una solución sólo en el caso de que la distribución de las ofertas salariales se suponga uniforme (de extremos desconocidos). Para otro tipo de distribuciones (por ejemplo, cuando las ofertas salariales siguen un modelo log-Normal), al construir las distribuciones condicionadas no se verifica la condición de "separabilidad" necesaria para aplicar nuestra propuesta, siendo este un problema a resolver. Lancaster, 1997, asume independencia a priori entre los parámetros, siendo este una posible solución a considerar.

BIBLIOGRAFÍA

1. Akahira, M. and Takeuchi, K. (1991), "A Definition of Information Amount Applicable to Non-Regular Cases", *Journal of Computing and Information*, **2**, pp.71-92.

2. Apóstol, T.M.(1960), "Análisis Matemático", Reverté, Madrid.

3. Arnold, B.C., Castillo, E. and Sarabia, J.M. (1999), *Conditional Specification of Statistical Models*, Springer Series in Statistics, Springer-Verlag, New York.

4. Azzalini, A. (1996), "Statistical Inference Based on the Likelihood", Chapman and Hall, London.

5. Basulto, J. (1997), "Funciones a Priori Imparciales Unidimensionales", *Estadística Española,* **142,** pp.99-128.

6. Bernardo, J.M. and Smith, A.F.M. (1994), "Bayesian Theory", John Wiley and Sons, Chichester.

7. Ghosal, S. (1999), "Probability Matching Priors for Non-Regular Cases", *Biometrika,* **86,** pp.956-964.

8. Jeffreys, H. (1946), “An Invariant Form for the Prior Probability in Estimation Problems”, *Proc. Roy. Soc. (London), Ser. A*, **186**, pp.453-461.

9. Jeffreys, H. (1961), “Theory of Probability”, 3rd. edition, Oxford University Press, London.

10. Kosmas, K.F. (1990), “Shortest Confidence Intervals for Families of Distributions Involving Truncation Parameters”, *The American Statician*, **44**, pp.167-168.

11. Lancaster, T. (1997), “Exact Structural Inference in Optimal Job-Search Models”, *Journal of Business and Economic Statistics*, **15**, pp.165-179.

12. Le Cam, (1990), “On the standard asymptotic confidence ellipsoids of Wald”, *International Satatistical Review*, **58**, pp.129-152.

13. Matusita, K. (1955), “Decisión Rules Based on the Distance for Problems of Fit, two Samples and Estimation”, *Ann. Math. Statist.*, ***26***, pp.631-640.

14. Nicolau, (1993), “Bayesian Intervals with Good Frequentist Behaviour in the Presence of Nuisance Parameters”, *J.R. Statist. Soc., Ser. B*, **55**, pp.377-390.

15. Ortega, F.J. y Basulto, J. (2003), “Distribuciones a priori unidimensionales en Modelos No Regulares: Medidas de Información”, *Estadística Española*, **154**, pp.363-383.

16. Pitman, E.J. (1979), “Some Basic Theory for Statistical Inference”, Chapman and Hall, London.

17. Sareen, S. (2003), “Reference Bayesian inference in nonregular models”, *Journal of Econometrics*, **113**, pp.265-288.

18. Welch, B.L. and Peers, H.W. (1963), “On Formulae for Confidence Points Based on Integral of Weithed Likelihoods”, *J.R. Statist. Soc., Ser. B*, **25**, pp.318-329.