

**BANCO CENTRAL DE COSTA RICA
DIVISIÓN ECONÓMICA
DEPARTAMENTO DE INVESTIGACIONES ECONÓMICAS
DIE-NT-03-96**

INFERENCIA BAYESIANA Y BANDAS CAMBIARIAS

Juan E. Muñoz Giró

ABRIL, 1996

INFERENCIA BAYESIANA Y BANDAS CAMBIARIAS¹

La inferencia estadística clásica se distingue, fundamentalmente, por dos características que la han definido desde sus inicios en el siglo XVIII: por las distribuciones probabilísticas perfectas que se suponen en el análisis y por el supuesto de que el investigador no tiene conocimiento previo sobre el fenómeno que está analizando. La inferencia bayesiana, en contraste a la clásica, le da la posibilidad a las distribuciones de probabilidad de ser asimétricas -tener sesgos- y le permite al investigador incorporar la información previa que puede poseer sobre el fenómeno. En esta nota técnica se discuten los principales postulados de la teoría bayesiana y se procede a aplicarla en un contexto de política cambiaria en el que la percepción subjetiva de cada experto puede conducir a diferentes resultados.

INTRODUCCIÓN

Hacia finales del siglo XVIII, precisamente en 1763, fue publicado el Teorema de Bayes, así llamado por el nombre del monje cisterciense que lo desarrolló en respuesta a los postulados de la inferencia gaussiana. El estudio clásico de las distribuciones de probabilidad -o estadística gaussiana- supone funciones de densidad simétricas y bien definidas, así como la ausencia de cualquier tipo de conocimiento previo por parte del investigador. Bayes, en la justificación de su teoría argumentaba que los datos no necesariamente provenían de tales funciones de densidad, sino, todo lo contrario, que probablemente eran generados por leyes probabilísticas sujetas a formas asimétricas y sesgadas. En el tanto que el investigador conociera estas características, el procedimiento correcto de inferencia estadística debería incorporar, decía Bayes, esta información y, de esa forma, contar con un marco probabilístico más apropiado para la inferencia estadística.

Desde su aparición hace más de dos siglos, la inferencia bayesiana no pasó de constituir simplemente un interés teórico, sin ningún tipo de aplicación directa. Sin embargo, la mayor disponibilidad de recursos informáticos dio inicio a una *revolución* en la ciencia estadística que también involucró el resurgimiento de la teoría bayesiana. Es así como en 1973 Box y Tiao publican un texto bastante comprensivo sobre esta teoría: *Bayesian Inference in Statistical Analysis*. Desde entonces es posible observar cada vez más artículos sobre el análisis bayesiano en revistas estadísticas y matemáticas.

En el campo económico, la inferencia bayesiana comienza a tomar un lugar importante especialmente cuando se comienzan a modelar expectativas racionales: hay

¹ Autorizado por Claudio Ureña C.

información previa que el sujeto económico incorpora en sus decisiones y una vez que suceden los hechos sus decisiones pueden ser modificadas. Es un típico proceso bayesiano. En este aspecto, el presente documento está ilustrado con un ejemplo de aplicación práctica con elementos bayesianos, los cuales se discutirán más adelante. El documento también responde al interés del Departamento de Investigaciones Económicas en desarrollar y aplicar métodos de investigación que amplíen el nivel de conocimientos de los investigadores y economistas de la División Económica como un todo.

II. INFERENCIA BAYESIANA

Como se menciona en la introducción, la inferencia bayesiana constituye un proceso de análisis en el que el investigador puede hacer uso de la información que pueda tener sobre el comportamiento de la función de distribución de la que provienen los datos. Las principales características que se le pueden apuntar a la teoría bayesiana son las siguientes:

1. Para cualquier modelo probabilístico, el análisis bayesiano hace uso de toda la información previa que se tenga sobre los datos.
2. Los resultados de inferencia estadística que se consideren inaceptables probablemente lo sean porque provienen de supuestos inapropiados y no por errores del proceso de inferencia. En otras palabras, la distribución de probabilidades que se ha supuesto *a priori* puede ser incorrecta.
3. En vista de que el análisis bayesiano puede ser aplicado a cualquier modelo de probabilidades, menos atención se le debe dar a la conveniencia matemática del modelo y más atención a sus méritos científicos.
4. La inferencia bayesiana no presenta problemas en la escogencia de estimadores y de intervalos de confianza.
5. Proporciona una manera satisfactoria de introducir explícitamente y de darle seguimiento a los supuestos sobre el conocimiento previo. Por ejemplo, en la teoría clásica, se supone *a priori* que la distribución de probabilidad de los datos es exactamente *normal* y que cada observación tiene exactamente la *misma* variancia y que está distribuida *independientemente* de las otras observaciones. Si el análisis de los residuos proporciona evidencia de que estos supuestos inmutables de la teoría clásica no se verifican, entonces la teoría clásica no puede dar luz sobre la *verdadera* o *a posteriori* distribución de probabilidad. Para ello sería necesario utilizar la inferencia bayesiana.
6. El teorema de Bayes permite la actualización continua de la información sobre los parámetros de la distribución conforme se generan más observaciones.

7. El teorema describe un proceso de aprendizaje en el tiempo.
8. A diferencia de la inferencia clásica, la bayesiana no requiere de la evaluación de las propiedades de los estimadores obtenidos de muestras sucesivas.
9. La probabilidad de un evento está dada por el grado de confianza o creencia que tiene un individuo sobre la ocurrencia del evento.
10. Esta confianza depende de la información cuantitativa y cualitativa que se posea.

Teorema de Bayes: suponga un vector de N observaciones $Y' = (y_1, \dots, y_N)$ cuya distribución de probabilidad $f(Y|\theta)$ depende del valor de K parámetros $\theta' = (\theta_1, \dots, \theta_K)$. Suponga que θ también tiene una distribución de probabilidad $g(\theta)$. Entonces, la distribución conjunta de Y y θ está definida por:

$$(1) \quad h(\mathbf{q}, Y) = f(Y|\mathbf{q}) \cdot g(\mathbf{q}|y) \cdot f(y)$$

donde $f(Y)$ es la función de densidad marginal del vector de observaciones Y . Por tanto, la distribución condicional de θ dada la información de Y es:

$$(2) \quad g(\mathbf{q}|y) = \frac{f(y|\mathbf{q}) \cdot g(\mathbf{q})}{f(y)}$$

La función $g(\theta|Y)$ es la distribución de probabilidad de θ posterior a la información de Y . En términos de la función de verosimilitud, el teorema de Bayes se puede postular como:

$$(3) \quad g(\mathbf{q}|y) \propto \ell(\mathbf{q}|y) \cdot g(\mathbf{q})$$

donde la función $\ell(\theta|Y)$ es la función de verosimilitud. Esta forma funcional establece que $\ell(\theta|Y)$ es la función por medio de la cual Y modifica el conocimiento previo de θ ; en otras palabras, se puede considerar como la información sobre θ que proviene de los datos. El símbolo (\propto) indica que la relación expresada en la ecuación (3) es una proporción y no una igualdad. Es decir, la distribución de probabilidad de θ posterior a la información de Y es proporcional al producto de la distribución *a priori* de θ y la verosimilitud de θ dado Y .

III. EL MODELO LINEAL GENERAL Y LA INFERENCIA BAYESIANA

Con el propósito de facilitar la aplicación de la inferencia bayesiana considérese el modelo lineal general de la forma:

$$(4) \quad Y = X\mathbf{b} + U$$

donde Y es un vector de observaciones (y_1, \dots, y_N); X es una matriz de orden ($N \times K$), es decir K variables y N observaciones; β es un vector de K parámetros (β_1, \dots, β_K) y U es un vector de N errores estocásticos (u_1, \dots, u_N). También por simplicidad se supone que el vector de errores sigue una distribución normal multivariante con vector de medias iguales a cero y una matriz de variancias y covariancias equivalente a $\sigma^2 I_N$, donde I es la matriz identidad de orden ($N \times N$). Con la definición de este modelo lineal el vector de parámetros θ se convierte en $\theta' = (\beta', \sigma)$. La desviación estándar es un escalar. Con estas definiciones, el postulado matemático del teorema de Bayes al modelo lineal general toma la siguiente forma:

$$(5) \quad g(\mathbf{b}, \mathbf{s} | Y) \propto \ell(\mathbf{b}, \mathbf{s} | Y) \cdot g(\mathbf{b}, \mathbf{s})$$

Como el teorema se aplica sobre los parámetros que serán sujetos de inferencia, la ecuación (5) postula que la función de probabilidad conjunta de los parámetros β y σ , condicionada (*a posteriori*) por la muestra Y , es proporcional a la verosimilitud de haber obtenido los valores muestrales con los parámetros β y σ , ponderada por la distribución de probabilidad *a priori*. Bajo el supuesto de que los errores tienen una distribución normal, la función de verosimilitud se define:

$$(6) \quad \ell(\mathbf{b}, \mathbf{s} | Y) = (2\pi)^{-\frac{N}{2}} \mathbf{s}^{-N} e^{-\frac{1}{2\mathbf{s}^2} (Y - X\mathbf{b})'(Y - X\mathbf{b})}$$

Los estimadores que hacen máxima la función de verosimilitud son los siguientes:

$$(7) \quad \hat{\mathbf{b}} = (X'X)^{-1} X'Y$$

$$(8) \quad \hat{\mathbf{s}}^2 = \frac{(Y - X\hat{\mathbf{b}})'(Y - X\hat{\mathbf{b}})}{N - K}$$

En el caso del estimador de la variancia se hace la corrección por los grados de libertad ($N-K$) para evitar el sesgo del estimador máximo verosímil. Con estos estimadores, la función de verosimilitud se puede expresar como una proporción de la forma:

$$(9) \quad \ell(\mathbf{b}, \mathbf{s} | Y) \propto \mathbf{s}^{-N} e^{-\frac{1}{2\mathbf{s}^2} [(N-K)\hat{\mathbf{s}}^2 + (\mathbf{b} - \hat{\mathbf{b}})'X'X(\mathbf{b} - \hat{\mathbf{b}})]}$$

en el que la diferencia con la ecuación (6) es la constante que ha desaparecido. Para los efectos siguientes, la ecuación (9) será la función de verosimilitud relevante.

Hasta este punto se han planteado los elementos básicos del Teorema de Bayes, quedando por discutir el tipo de información previa que posee el investigador. En este caso la teoría propone dos posibilidades: el análisis con información *a priori* inexistente y con información conocida, dentro de la cual se pueden identificar la información conjugada natural, la información \mathbf{g} *a priori* y la distribución gama-t. A continuación se discuten estas posibilidades.

IV. ANÁLISIS SIN INFORMACIÓN A PRIORI

Cuando el investigador no tiene información *a priori* sobre la distribución de probabilidad conjunta de β y σ , el procedimiento comúnmente utilizado es suponer que:

$$(10) \quad g(\mathbf{b}, \mathbf{s}) \propto \mathbf{s}^{-1}$$

Este supuesto tiene las siguientes implicaciones:

1. *A priori*, los parámetros son independientes entre ellos.
2. Las funciones de densidades para los parámetros son proporcionales a una constante. La función de densidad marginal de la desviación estándar es proporcional al inverso de la desviación estándar.
3. Las densidades marginales son uniformes; es decir, se supone que los valores de los parámetros pueden ocurrir con igual probabilidad.

Una característica importante de la distribución de probabilidad de la ecuación (10) es que su valor integrado en el rango de los parámetros no es igual a 1. Es una función *impropia*. Esta característica no representaría ningún problema en el tanto que la distribución *a posteriori* integre a 1 (que sea *propia*). La justificación de la ecuación (10) yace en la ausencia de información propiamente, aunque también evita sesgos y es fácil de trabajar. Sin embargo, su uso puede conducir a funciones posteriores ambiguas y a estimadores inadmisibles. Con esta función *a priori*, la distribución de densidad conjunta y condicionada tiene la siguiente forma:

$$(11) \quad g(\mathbf{b}, \mathbf{s} | Y) \propto \mathbf{s}^{-(N+1)} e^{-\frac{1}{2\mathbf{s}^2}[(N-K)\hat{\mathbf{s}}^2 + (\mathbf{b} - \hat{\mathbf{b}})'X'X(\mathbf{b} - \hat{\mathbf{b}})]}$$

En la teoría estadística, esta distribución *a posteriori* es de la forma normal-gama, la cual permite expresar la densidad conjunta como el producto de las densidades marginales condicionadas. Es decir:

$$(12) \quad g(\mathbf{b}, \mathbf{s} | Y) = g(\mathbf{b} | \mathbf{s}, Y) \cdot g(\mathbf{s} | Y)$$

donde el primer factor multiplicativo a la derecha de la igualdad es la función de densidad marginal de los parámetros condicionada a la desviación estándar y a la información, en tanto que el segundo factor es la densidad marginal de la desviación estándar condicionada a la información. La función de distribución marginal del vector de los parámetros β es la siguiente:

$$(13) \quad g(\mathbf{b} | \mathbf{s}, Y) = (2\mathbf{p})^{\frac{K}{2}} \mathbf{s}^{-K} |X'X|^{-\frac{1}{2}} e^{-\frac{1}{2\mathbf{s}^2}(\mathbf{b} - \hat{\mathbf{b}})'X'X(\mathbf{b} - \hat{\mathbf{b}})}$$

la cual es una distribución normal con media β y matriz de variancias y covariancias equivalente a $\sigma^2(X'X)^{-1}$. La normalidad de la función está definida por el *kernel*, el cual corresponde a la potencia a la que está elevado el número \mathbf{e} de la ecuación (13), en tanto que

el factor que permite la integración a 1 son los tres factores multiplicativos que anteceden al número **e**.

La distribución de probabilidad de la desviación estándar condicionada a la información Y está dada por:

$$(14) \quad g(\mathbf{s}|Y) = \frac{2}{\Gamma\left(\frac{N-K}{2}\right)} \left[\frac{(N-K)}{2} \hat{\mathbf{s}}^2 \right]^{\frac{N-K}{2}} \mathbf{s}^{-(N-K+1)} e^{-\frac{N-K \hat{\mathbf{s}}^2}{2 \mathbf{s}^2}}$$

la cual corresponde a una distribución gama invertida definida por el *kernel* (los dos últimos factores multiplicativos a la derecha de la igualdad) y por los dos primeros factores multiplicativos que aseguran su integración a 1.

En vista de que la desviación estándar poblacional es desconocida, su factorización de la ecuación (13) permite obtener la distribución marginal *a posteriori* de los coeficientes β :

$$(15) \quad g(\mathbf{b}|Y) = \int_0^{\infty} g(\mathbf{b}, \mathbf{s}|Y) \mathcal{I}_{\mathbf{s}} = \int_0^{\infty} g(\mathbf{b}, \mathbf{s}, Y) \mathcal{I}_{\mathbf{s}}$$

la cual es proporcional a:

$$(16) \quad g(\mathbf{b}|Y) \propto \left[1 + \frac{1}{(N-K)\mathbf{s}^2} (\mathbf{b} - \hat{\mathbf{b}})' X' X (\mathbf{b} - \hat{\mathbf{b}}) \right]^{-\frac{N}{2}}$$

Ello indica que el vector de estos coeficientes tiene una distribución t multivariante con media β y matriz de variancias y covariancias equivalente a $(N-K)/(N-K-2) \sigma^2 (X'X)^{-1}$. Para un parámetro en particular, dígase β_1 , la densidad *a posteriori* se obtiene al integrar la ecuación (16) con respecto a β_2, \dots, β_K :

$$(17) \quad g(\beta_1|Y) \propto \left[1 + \frac{1}{N-K} \left(\frac{\beta_1 - \hat{\beta}}{\hat{\mathbf{s}} \sqrt{a_{11}}} \right)^2 \right]^{-\frac{(N-K+1)}{2}}$$

la cual define una distribución t univariante con media β_1 y variancia $(N-K)/(N-K-2) \sigma^2 a_{11}$, donde el elemento a_{11} corresponde al primer escalar de la diagonal de la matriz $(X'X)^{-1}$.

V. ANÁLISIS CON INFORMACIÓN CONJUGADA NATURAL A PRIORI

Una densidad *a priori* con información conjugada natural es altamente conveniente ya que su aplicación resulta en una densidad *a posteriori* de la misma forma. Es decir, es una información muy flexible. Por ejemplo, suponga una función de verosimilitud *a priori* equivalente al producto de dos funciones marginales:

$$(18) \quad \ell(\mathbf{b}, \mathbf{s} | Y) \propto h_1(\mathbf{b} | \mathbf{s}, Y) \cdot h_2(\mathbf{s} | Y)$$

Definanse las funciones marginales con las siguientes formas:

$$(19) \quad h_1(\mathbf{b} | \mathbf{s}, Y) = e^{-\frac{1}{2s^2}(\mathbf{b} - \hat{\mathbf{b}})' X' X (\mathbf{b} - \hat{\mathbf{b}})}$$

$$(20) \quad h_2(\mathbf{s} | Y) = \mathbf{s}^{-N} e^{-\frac{N-K \bar{s}^2}{2 s^2}}$$

de forma tal que la distribución de información conjugada natural conjunta para (β, σ) es una donde la función condicionada de β dado σ es normal multivariante y la distribución marginal de σ es gama invertida. Esas densidades *a priori* se pueden definir, entonces, como:

$$(21) \quad g(\mathbf{b} | \mathbf{s}) = (2\mathbf{p})^{-\frac{K}{2}} \mathbf{s}^{-K} |A|^{1/2} e^{-\frac{1}{2s^2}(\mathbf{b} - \bar{\mathbf{b}})' A (\mathbf{b} - \bar{\mathbf{b}})}$$

$$(22) \quad g(\mathbf{s}) = \frac{2}{\Gamma\left(\frac{N-K}{2}\right)} \left[\frac{(N-K)\bar{s}^2}{2} \right]^{\frac{N-K}{2}} \mathbf{s}^{-(N-K+1)} e^{-\frac{N-K \bar{s}^2}{2 s^2}}$$

donde los parámetros β y \bar{s}^2 (con barra superior) y $(N-K)$ y A son valores de información *a priori* subjetiva. Con esta especificación, $E(\beta | \sigma) = E(\beta) = \beta$ y $\text{cov}(\beta | \sigma) = \sigma^2 A^{-1}$. En el caso de los parámetros de la distribución de la variancia, éstos son:

$$(23) \quad E(\mathbf{s}) = \frac{\Gamma\left(\frac{N-K-1}{2}\right)}{\Gamma\left(\frac{N-K}{2}\right)} \left(\frac{N-K}{2}\right)^{1/2} \bar{s}$$

$$(24) \quad E(\mathbf{s}^2) = \frac{N-K}{N-K-2} \bar{s}^2$$

$$(25) \quad Mo(\mathbf{s}^2) = \left(\frac{N-K}{N-K+1} \right)^{\frac{1}{2}} \bar{s}^2$$

donde la ecuación (25) es la moda de la distribución. Con esta información la densidad conjunta *a priori* es:

$$(26) \quad g(\mathbf{b}, \mathbf{s}) \propto \mathbf{s}^{-(N+1)} e^{-\frac{1}{2\mathbf{s}^2}[(N-K)\bar{s}^2 + (\mathbf{b}-\bar{\mathbf{b}})'A(\mathbf{b}-\bar{\mathbf{b}})]}$$

con la cual se puede definir la función marginal de β :

$$(27) \quad g(\mathbf{b}) \propto \left[1 + \frac{1}{(N-K)\bar{s}^2} (\mathbf{b}-\bar{\mathbf{b}})' A (\mathbf{b}-\bar{\mathbf{b}}) \right]^{-\frac{N}{2}}$$

Para obtener la función de distribución conjunta *a posteriori* habría que combinar la información conjugada natural (ecuaciones 21 y 22) con la función de verosimilitud (ecuación 9):

$$(28) \quad g(\mathbf{b}, \mathbf{s} | Y) \propto \mathbf{s}^{-K} e^{-\frac{1}{2\mathbf{s}^2}(\mathbf{b}-\bar{\mathbf{b}})'(A+X'X)(\mathbf{b}-\bar{\mathbf{b}})} \mathbf{s}^{-(N-K+1)} e^{-\frac{N-K}{2} \frac{s^2}{\mathbf{s}^2}}$$

Esta función es una normal gama en la que el vector β tiene una distribución t-student multivariante y β_i una distribución t-student univariante .

No obstante la utilidad práctica que tiene el suponer información conjugada natural, el investigador puede enfrentar un problema para definir el valor del parámetro $|A|$, por lo que podría ser conveniente definir una alternativa como la que se discute a continuación.

VII. LA DISTRIBUCIÓN *g* A PRIORI

Como solución alternativa a la definición de una matriz A en el método de información conjugada, Zellner propone el uso de una *distribución g* que permita definir a A como:

$$(29) \quad A = g_0 X' X$$

donde g_0 es un parámetro *a priori*. Si se supone además la observancia de la ecuación (10), entonces la densidad conjunta estaría definida por:

$$(30) \quad g(\mathbf{b}, \mathbf{s} | Y) \propto \mathbf{s}^{-K} e^{-\frac{1+g_0}{2s^2} (\mathbf{b}-\bar{\mathbf{b}})' X' X (\mathbf{b}-\bar{\mathbf{b}})} \mathbf{s}^{-(N+1)} e^{-\frac{N}{2} \frac{s^2}{s^2}}$$

donde los parámetros con doble barra se definen como:

$$(31) \quad N \bar{s}^2 = (Y - X \bar{\mathbf{b}})' (Y - X \bar{\mathbf{b}}) + g_0 (\bar{\mathbf{b}} - \hat{\mathbf{b}})' X' X (\bar{\mathbf{b}} - \hat{\mathbf{b}})$$

$$(32) \quad \bar{\mathbf{b}} = \frac{g_0 \bar{\mathbf{b}} + \hat{\mathbf{b}}}{1 + g_0}$$

Es decir, los nuevos parámetros se definen como promedios ponderados entre los estimadores de información conjugada natural y de máxima verosimilitud.

VII. DISTRIBUCIÓN Gama-t INDEPENDIENTE

Este enfoque supone *a priori* que los parámetros β y σ son independientes con funciones de densidad proporcionales a:

$$(33) \quad g(\mathbf{b}) \propto \left[1 + \frac{1}{v^*} (\mathbf{b} - \bar{\mathbf{b}})' A (\mathbf{b} - \bar{\mathbf{b}}) \right]^{-\frac{K+v^*}{2}}$$

$$(34) \quad g(\mathbf{s}) \propto \frac{1}{\mathbf{s}^{\bar{v}+1}} e^{-\frac{\bar{v}}{2} \frac{s^2}{s^2}}$$

donde v^* , A , $\bar{\beta}$ y \bar{v} (estos dos últimos con barra) son parámetros *a priori* por especificar. La función de densidad *a priori* para el vector β , una vez definida la densidad conjunta apropiadamente, está dada por:

$$(35) \quad g(\mathbf{b} | Y) \propto \left[1 + \frac{1}{\bar{v}} \frac{1}{s} (\mathbf{b} - \hat{\mathbf{b}})' X' X (\mathbf{b} - \hat{\mathbf{b}}) \right]^{-\frac{K+v}{2}} \left[1 + \frac{1}{v^*} (\mathbf{b} - \bar{\mathbf{b}})' A (\mathbf{b} - \bar{\mathbf{b}}) \right]^{-\frac{K+v^*}{2}}$$

la cual define una distribución t-student multivariante, donde los parámetros revisados se definen como:

$$(36) \quad \bar{v} = \bar{v} + (N - K)$$

$$(37) \quad v s^2 = v \hat{\mathbf{S}}^2 + \overline{v s^2}$$

$$(38) \quad v = N - K$$

$$(39) \quad v \hat{\mathbf{S}}^2 = (Y - X \hat{\mathbf{b}})'(Y - X \hat{\mathbf{b}})$$

VIIIIA FUNCIÓN DE PERDIDA

La función de distribución *a posteriori* resume toda la información necesaria y suficiente sobre el parámetro estimado. Sin embargo, la escogencia del estimador de β *a priori* induce, como también sucede en la inferencia clásica, a una pérdida de eficiencia por no contar con el verdadero valor paramétrico. Para capturar la magnitud de esta pérdida es importante encontrar un estimador de β que minimice la función de pérdida. Entre las funciones de pérdida comúnmente usadas están las siguientes:

$$(40) \quad L_1 = c \left| \hat{\mathbf{b}} - \mathbf{b} \right|$$

$$(41) \quad L_2 = c (\hat{\mathbf{b}} - \mathbf{b})^2$$

La primera de ellas representa una pérdida absoluta que se minimiza con el valor muestral de la mediana, mientras que la segunda especificación alcanza una pérdida mínima cuadrática cuando se usa la media muestral.

En el procedimiento bayesiano, el estimador puntual es aquel valor de β que minimiza la pérdida *a posteriori* definida como:

$$(42) \quad E_{\mathbf{b}|Y} \left[L(\mathbf{b}, \hat{\mathbf{b}}) \right] = \int L(\mathbf{b}, \hat{\mathbf{b}}) \cdot g(\mathbf{b}|Y) \mathbb{1} \mathbf{b}$$

Como se mencionó anteriormente, la función de pérdida absoluta se minimiza con el valor muestral de la mediana, en tanto que la función cuadrática se minimiza con la media muestral. Ambos estimadores coinciden cuando la distribución es normal.

El estimador bayesiano con una función de pérdida cuadrática está definido cuando se busca un estimador β que minimice:

$$(43) \quad E_{\mathbf{b}|Y} [L_2] = \int c (\hat{\mathbf{b}} - \mathbf{b})^2 \cdot g(\mathbf{b}|Y) \mathbb{1} \mathbf{b}$$

el cual corresponde a la solución de primer orden:

$$(44) \quad \frac{\int \mathbb{1} E_{b|Y}[L_2] \mathbb{1} b}{\int \mathbb{1} b} = \int 2c(\hat{\mathbf{b}} - \mathbf{b}) \cdot g(\mathbf{b}|Y) \mathbb{1} b = 0$$

que arreglada de la siguiente forma:

$$(45) \quad \hat{\mathbf{b}} \int g(\mathbf{b}|Y) \mathbb{1} b = \int \mathbf{b} g(\mathbf{b}|Y) \mathbb{1} b$$

donde el valor de la integral matemática a la izquierda de la igualdad es 1, proporciona la definición de la media de la distribución:

$$(46) \quad \hat{\mathbf{b}} = E(\mathbf{b}|Y) = \int \mathbf{b} g(\mathbf{b}|Y) \mathbb{1} b$$

Por su parte, el estimador bayesiano de la función de pérdida absoluta se obtiene al minimizar:

$$(47) \quad E_{b|Y}[L_1] = \int c|\hat{\mathbf{b}} - \mathbf{b}| \cdot g(\mathbf{b}|Y) \mathbb{1} b$$

lo cual es equivalente a la minimización por intervalos de integración:

$$(48) \quad E_{b|Y}[L_1] = \int_{-\infty}^{\hat{\mathbf{b}}} c(\hat{\mathbf{b}} - \mathbf{b}) g(\mathbf{b}|Y) \mathbb{1} b + \int_{\hat{\mathbf{b}}}^{\infty} c(\mathbf{b} - \hat{\mathbf{b}}) g(\mathbf{b}|Y) \mathbb{1} b$$

La solución de primer orden es igual a:

$$(49) \quad \frac{\int \mathbb{1} E_{b|Y}[L_1] \mathbb{1} b}{\int \mathbb{1} b} = \int_{-\infty}^{\hat{\mathbf{b}}} c \cdot g(\mathbf{b}|Y) \mathbb{1} b + \int_{\hat{\mathbf{b}}}^{\infty} c \cdot g(\mathbf{b}|Y) \mathbb{1} b$$

que, una vez igualada a cero, proporciona la siguiente solución:

$$(50) \quad \int_{-\infty}^{\hat{\mathbf{b}}} g(\mathbf{b}|Y) \mathbb{1} b = \int_{\hat{\mathbf{b}}}^{\infty} g(\mathbf{b}|Y) \mathbb{1} b = \frac{1}{2}$$

Esta solución indica que el estimador que minimiza la función de pérdida absoluta acumula una probabilidad del 50 por ciento a la izquierda y a la derecha de su valor muestral. En otras palabras, ese estimador es la mediana de la muestra.

IX. UNA APLICACIÓN BÁSICA

Originalmente se tenía prevista la aplicación de un paquete de análisis econométrico llamado *Bayesian Regression Analysis Program* (BRAP). Sin embargo, no ha sido posible estimar las regresiones por la ausencia de uno de los archivos del *software*. No obstante este inconveniente, se considera de interés presentar un ejemplo de actualización informativa al estilo bayesiano. El ejemplo es con propósito ilustrativo y no pretende ser vinculante para los objetivos de política económica. Supóngase, en primer lugar, que se le pregunta a un economista cuáles podrían ser las eventuales bandas de flotación para un tipo de cambio que en el día había sido observado en ¢200 por US\$1. Su respuesta fue de 10 por ciento por arriba y por debajo de ese valor. Ello implica que los niveles de las bandas serían ¢180 y ¢220.

Aun cuando ese economista no dio ninguna otra explicación, en el fondo su opinión representa todo un sistema de pronóstico sustentado en su experiencia laboral y académica. Tiene conocimiento previo que, de ser consultado, puede ser vinculante en una decisión futura. Ahora bien, como lo que implica dinámica económica, puede que la opinión de él esté sujeta a un nivel de error dado. Para simplificar el ejemplo, se supone también que él tiene una confianza privada del 95 por ciento de que el tipo de cambio flotaría entre esas bandas. En términos estadísticos simples, las bandas cambiarias subjetivas se podrían expresar como:

$$P(180 < e < 220) = 0,95$$

Adicionalmente, si las relaciones entre los *fundamentals* del tipo de cambio muestran, según este economista, un comportamiento *normal* (que es lo que siempre se supone en la regresión clásica), entonces el valor tabular de la distribución normal asociada con el intervalo anterior sería equivalente a 1,96 bajo un 1 por ciento de significancia. Con esta información el valor medio del tipo de cambio es ¢200 (información que fue dada desde un principio), la desviación estándar es de 10,20 y la variancia de 104,04.

Considérese ahora la opinión de otro economista con más experiencia laboral y académica que el primero. ¿Cómo afectaría la opinión de este último a las bandas del primero? De acuerdo con el segundo, las bandas cambiarias deberían de estar 5 por ciento por arriba y por debajo del nivel de ¢200 por US\$1. Es decir, en términos estadísticos:

$$P(190 < e < 210) = 0,95$$

Bajo los mismos supuestos establecidos para el caso de la estructura probabilística del primer economista, la variancia implícita del tipo de cambio en el caso del segundo es 26,03 y su desviación estándar es de 5,10. La información conjugada para ambos economistas permite estimar un tipo de cambio y su variancia con las siguientes fórmulas:

$$(51) \quad \bar{e} = \frac{\frac{1}{s_1^2} e_1 + \frac{1}{s_2^2} e_2}{\frac{1}{s_1^2} + \frac{1}{s_2^2}}$$

$$(52) \quad \bar{s}^2 = \frac{1}{\frac{1}{s_1^2} + \frac{1}{s_2^2}}$$

Estas ecuaciones representan ajustes armónicos para la media y la variancia. En el caso del promedio, la media conjugada permanece en €200 por US\$1, mientras que la variancia conjugada se sitúa en 20,83 con una desviación estándar de 4,56. Nótese que la variancia y la desviación estándar no se sitúan entre los valores implícitos en las opiniones de ambos economistas, sino que, todo lo contrario, están por debajo de ambas opiniones. Esta es una de las principales características de la inferencia bayesiana. La consideración de valores muestrales adicionales **siempre** mejoran la precisión (menor variancia) de los estimadores. Con la información conjugada, las bandas cambiarias se situarían en:

$$P(191 < e < 209) = 0,95$$

En otras palabras, las bandas cambiarias resultantes de la opinión de dos expertos serían equivalentes a 4,5 por ciento por arriba y por debajo de €200 por US\$1. Como se mencionó desde un principio, este es un simple ejemplo que puede ser aplicado formalmente una vez que se cuente con el manual del paquete de econometría bayesiana.

X.BIBLIOGRAFÍA

Billingsley, Patrick (1986). **Probability and Measure**. Second Edition. Nueva York: John Willey & Sons, The University of Chicago, 622 páginas.

Box, George E. P. y George C Tiao (1973). **Bayesian Inference in Statistical Analysis**. Massachussetts: Addison-Wesley Publishing Company, 588 páginas.

Judge, George G., R. Carter Hill, William E. Griffiths, Helmut Lutkepohl y Tsoung-Chao Lee (1988). **Introduction to the Theory and Practice of Econometrics**. Second Edition. Nueva York: John Wiley & Sons, 1024 páginas.

Judge, George G., R. Carter Hill, William E. Griffiths, Helmut Lutkepohl y Tsoung-Chao Lee (1985). The Theory and Practice of Econometrics. Second Edition. Nueva York: John Wiley & Sons, 1019 páginas.

F:\INVESTIG\DIE\NT\NT96\NT0396.DOC