Anuario Jurídico y Económico Escurialense, XLI (2008) 327-360 / ISSN: 1133-3677

El resurgir de Thomas Bayes

Agustín ALONSO RODRÍGUEZ
Real Centro Universitario
«Escorial-María Cristina»
San Lorenzo del Escorial

Resumen: En este trabajo presentamos el punto de vista de Thomas Bayes en la Teoría de la Probabilidad, y sus consecuencias en la probabilidad y estadística contemporáneas. Los nuevos métodos de cálculo numérico y la potencia de cálculo de los ordenadores han hecho posible la aplicación del famoso Teorema que lleva su nombre.

Abstract: In this paper we present the point of view of Bayes in the Theory of Probability, and its consequences in modern probability and statistics. The new numerical methods and the computing power of computers have made a reality the application of his famous Theorem.

Palabras clave: Teorema de Bayes, probabilidad a priori, probabilidad a posteriori, función de verosimilitud, métodos MCMC, programa R, proyecto BUGS.

Key words: Bayes theorem, probability a priori, probability a posteriori, likelihood function, MCMC numerical methods, R program, BUGS project.

Sumario:

I. Introducción

Nos acercamos a uno de los grandes pensadores en el campo de la probabilidad, y buena prueba de su estatura intelectual está en el hecho de que sigue siendo objeto de estudio, tanto su persona como su obra, ya que fue el primero en utilizar la probabilidad de forma inductiva.

Y, sin embargo, es sorprendente los pocos detalles sobre su vida que se conocen. Siguiendo a Barnard (1958), he aquí algunos datos. En el *The Dictionary of National Biography*, de finales del XIX, se hace mención de su padre, Joshua Bayes, uno de los seis *No-Conformistas* [1] ordenado ministro en Inglaterra, pero nada se dice de su hijo Thomas Bayes. Tampoco se hace referencia a Thomas Bayes en el *Imperial Dictionary of Universal Biography*, publicado en Glasgow, en 1865, siendo la *Encyclopedia Britannica* la primera referencia a su persona en una obra de consulta general.

Y en lo que se refiere a España, la *EnciclopediaUniversal Ilustrada Espasa* no menciona a Bayes hasta la edición de 2005. Con anterioridad a esa fecha, sí lo hacen la *Gran Enciclopedia Rialp* y la *Nueva Enciclopedia del Mundo*, de Durvan Ediciones (Club Internacional del Libro).

II. Apuntes biográficos

Siguiendo a Barnard (1958), podemos destacar los siguientes detalles biográficos.

Nació en Londres en 1702. Fue el mayor de los hijos de Ana y Joshua Bayes. Fue educado privadamente, como era costumbre entre los *No-Conformistas* de entonces.

1. No seguidores de la Church of England.

Cuando tenía unos doce años, Bernoulli escribía a Leibniz diciendo que el «pobre» De Moivre vivía en Londres teniendo que dar clases de matemáticas, lo que permite especular que quizás Thomas Bayes aprendió matemáticas de uno de los fundadores de la teoría de la probabilidad.

Tras ser ordenado ministro, Thomas comenzó su actividad pastoral ayudando a su padre, que por entonces recibió el cargo de ministro presbiteriano en la *casa de encuentros*, en Leather Lane, cerca de Holborn. Más tarde, pasó a ministrar en la capilla presbiteriana de Little Mount Sion, en Tunbridge Wells, abierta el primero de agosto del año 1720, si bien no parece que él fuera la persona que inaugurara la capilla. Allí se encontraba, en 1731, cuando publicó el tratado con título *Divine Benevolence or an attempt to prove that the Principle End of the Divine Providence and Government is the happiness of His Creatures*, o un intento de probar que la principal finalidad de la Divina Providencia y Gobierno era la felicidad de sus criaturas.

La obra fue publicada por el editor John Noon, quien, en 1736, publicó *An Introduction to the Doctrine of Fluxions and a Defence of the Mathematicians against the objections of the Author of the Analyst*. La obra apareció sin autor, si bien sus contemporáneos la atribuyeron a Thomas Bayes, y a él aparece atribuida en el Cátalogo del British Museum. Esta obra es una refutación del analista, el obispo anglicano George Berkeley, en sus ataques a los matemáticos, que en sus controversias y análisis no buscaban más que deducciones lógicas y no el hacer mejores a los hombres. Bayes argumenta que la labor de los matemáticos es hacer inteligibles y razonadas las nociones y conceptos que tratan. Es posible que este tratado tuviera algo que ver con su elección como *Fellow* de la Royal Society, en 1742.

William Whiston, sucesor de Newton en la cátedra «Lucasian», en la Universidad de Cambridge, refiere su admiración y respeto por Bayes al calificarlo como «a very good mathematician».

En opinión de Strange (1949), parece ser que Bayes quiso dejar su ministerio hacia 1749. No lo logró hasta 1752, cuando, tras diversos avatares, fue sucedido por el reverendo William Johnston, la persona que heredó la valiosa biblioteca de Bayes.

Bayes continuó viviendo en Tunbridge Wells hasta su muerte, el 17 de abril de 1761. Fue enterrado junto a sus padres y hermanos en el cementerio de los *No-Conformistas* de Bunhill Fields. En este cementerio reposan también los restos de su amigo el reverendo

Richard Price, quien un año después de la muerte de Bayes, 1763, publicó su famoso tratado: *An essay towards solving a problem in the Doctrine of Chances*.

En la reimpresión de este estudio publicada por *Biometrika*, volumen 45, diciembre de 1958, encontrará el lector la «Introducción» de G. A. Barnard, base de los rasgos biográficos aquí señalados, así como la presentación y comentarios de Richard Price en su escrito de solicitud al presidente de la Royal Society, John Canton para la publicación del ensayo de Bayes en *The Philosophical Transactions of the Royal Society*.

Hay que destacar que el escrito de solicitud de Price no es una mera carta de solicitud: es un comentario laborioso y documentado del pensamiento de Bayes para facilitar la decisión de John Canton, que pudiera no tener tiempo para una revisión a fondo del ensayo, antes de la publicación. Tanto es así que Price se considera el único responsable de los posibles errores que pudieran aparecer. Esto parece delatar las muchas horas de discusión y análisis entre ambos amigos. La petición de Price lleva la fecha del 10 de noviembre de 1763.

Price pone de relieve que el tratado de De Moivre sobre las leyes de la probabilidad (Laws of Chance) en su obra *Doctrine of Chances*, no eran suficientes para resolver la denominada probabilidad inversa, el *converse problem*, enfatizando Price que el problema resuelto por Bayes no había sido antes resuelto por nadie.

Price se permite afirmar que el propósito de Bayes era el poder pasar de las causas causadas a la existencia de la Causa Incausada, es decir, a la existencia de la Deidad [2].

---

2. En palabras de Price:

«The purpose I mean is to shew what reason we have for believing that there are in the constitution of things fixt laws according to which events happen, and that, therefore, the frame of the world must be the effect of the wisdom and power of an intelligent cause; and thus to confirm the argument taken from final causes for the existence of the Deity. It will be easy to see that the converse problem solved in this essay is more directly applicable to this purpose; for it shews us, with distinctness and precision, in every case of any particular order or recurrency of events, what reason there is to think that such recurrency or order is derived from stable causes or regulations in nature, and not from any of the irregularities of chance.»

Por último, destacar que el ensayo de Bayes arranca directamente con el planteamiento del problema que desea tratar. En sus propios términos escribe:

> «PROBLEM
>
> *Given* the number of times in which an unknown event has happened and failed: *Required* the chance that the probability of its happening in a single trial lies somewhere between any two degrees of probability that can be named.»

Siguen luego una serie de definiciones y el resto de la argumentación.

Aquí comenzaremos deduciendo el teorema de Bayes, en términos de la teoría elemental de la probabilidad, para luego exponer la dirección que ha dado al teorema su vigente actualidad.

III. El teorema de Bayes

A fin de mostrar la simplicidad e inmediatez de este famoso teorema, es necesario recordar los fundamentos, comenzando por un par de conceptos.

En primer lugar, por *experimento aleatorio* se entiende aquel experimento cuyo resultado no es predecible con exactitud, aun repitiéndolo en igualdad de condiciones. El conjunto de todos los posibles resultados de un experimento aleatorio se conoce como el *espacio muestral,* simbolizado aquí por U*,* y por *evento*, se entiende uno cualquiera de los posibles resultados de un experimento aleatorio, simbolizados por las letras mayúsculas: A, B, etc.

El razonamiento lógico plausible requiere, como punto de partida, la compatibilidad con los axiomas establecidos por Kolmogorov (1933), que son los siguientes:

Axioma 1. Para todo evento $A \in U$, $P(A) \geq 0$, es decir, la probabilidad es una magnitud no negativa.

Axioma 2. $P(U) = 1$, la probabilidad del *espacio muestral*, el conjunto de todos los posibles resultados de un experimento aleatorio, es la unidad.

Axioma 3. Si $A \in U$ y $B \in U$ son dos eventos mutuamente excluyentes o *disjuntos*, entonces $P(A \cup B) = P(A) + P(B)$.

A partir de estos axiomas, tenemos las siguientes *reglas* de operación con probabilidades.

1. $P(\varnothing)=0$, la probabilidad del evento *vacío* es cero. El evento *vacío* es el evento resultante de la intersección de dos eventos excluyentes o disjuntos.

2. $P(\overline{A})=1-P(A)$, siendo $\overline{A}$ el complemento del evento A.

3. $P(A\cup B)=P(A)+P(B)-P(A\cap B)$, es decir la probabilidad de la unión de dos eventos cualesquiera es la suma de las probabilidades singulares menos la probabilidad de la intersección.

4. Si ambos eventos fueran independientes, entonces

$$P(A\cup B)=P(A)+P(B)$$

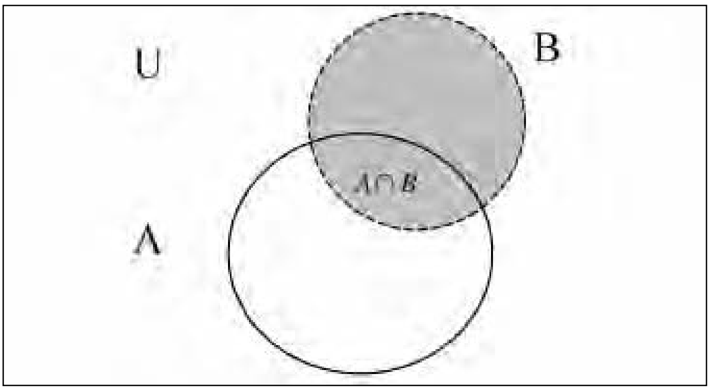
5. Para dos eventos A y B, la probabilidad marginal de A, por ejemplo, viene dada por $P(A)=P(A\cap B)+P(A\cap\overline{B})$, decir, su probabilidad es la suma de las probabilidades de los dos eventos disjuntos posibles.

6. Para dos eventos A y B, la probabilidad de B condicionada a A, se expresa como.

$$P(B\mid A)=\frac{P(A\cap B)}{P(A)},$$

En palabras, la probabilidad condicionada de B, una vez sabido que se ha dado el evento A, es proporcional a la probabilidad conjunta de ambos eventos, ponderada por la probabilidad del evento condicionante. La situación viene representada en la figura 1.

FIGURA 1



Conviene señalar que el evento B no es observable, mientras que A es el evento observable, que pasa a ser el espacio muestral *reducido* de este experimento. Cabría escribir la probabilidad condicionada como P(A | B), pero los eventos A y B no son simétricos, y por eso, la probabilidad condicionada se indicará como P(B | A).

Como corolario de esta regla, está el hecho de que de ser A y B eventos independientes, entonces la $P(B \mid A) = P(B)$, que pone de relieve como el conocimiento del evento A en nada afecta a la probabilidad del evento B.

7. Si reordenamos, al menos formalmente, los roles del los eventos A y B, en la regla anterior, tenemos:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \qquad (1)$$

Si bien es necesario repetir que A y B no son eventos equivalentes. Es decir, la ocurrencia o no ocurrencia de B no es observable, mientras que A es un evento observable, que puede ocurrir tanto si se da B como su complemento $\overline{B}$. No obstante, las posibilidades de darse A dependen de que sea B o $\overline{B}$ el que ocurra. En otras palabras, la ocurrencia de A está condicionada por la ocurrencia o no de B.

Si en (1) despejamos el numerador, llegamos a:

$$P(A \cap B) = P(B) \times P(A \mid B) \qquad (2)$$

relación que se conoce como la regla de la multiplicación, que reescribe la relación de la probabilidad condicional de un evento observable, dado el evento no observable, en una manera útil para obtener la probabilidad conjunta $P(A \cap B)$

De igual forma, podemos escribir:

$$P(A \cap \overline{B}) = P(\overline{B}) \times P(A \mid \overline{B}).$$

8.- La última regla para operar con probabilidades es la regla de la probabilidad total.

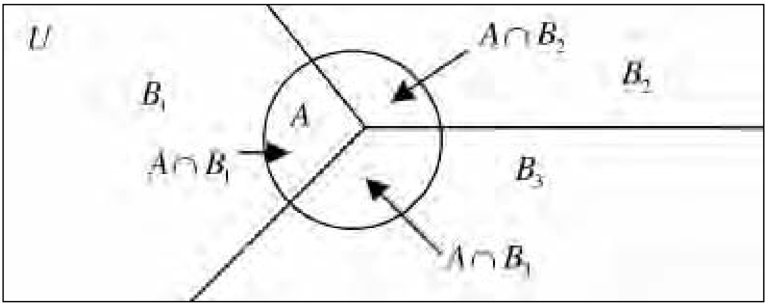
Dados los eventos $B_1, B_2, \ldots, B_n$ que particionan el espacio muestral U, y dado el evento A, entonces

$$P(A) = \sum_{i=1}^{n} P(B_i \cap A).$$

En palabras: la probabilidad del evento A es la suma de las probabilidades de sus elementos disjuntos.

Si fueran tres los eventos no observables B, que particionan el espacio muestral U, podríamos representar gráficamente lo dicho, como en la figura 2.

FIGURA 2



En este caso, el espacio muestral es $U = B_1 \cup B_2 \cup B_3$, y la probabilidad de A es: $P(A) = P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3)$

Si para cada uno de los eventos disjuntos utilizamos la regla de la multiplicación, llegamos a

$$P(A) = \sum_{j=1}^{n} P(B_j) \times P(A \mid B_j) \quad (3)$$

Y recordando que la probabilidad condicionada $P(B_i \mid A)$;*i=1,2,…,n* se obtiene dividiendo cada probabilidad conjunta por la probabilidad del evento A, es decir:

$$P(B_i \mid A) = \frac{P(A \cap B_i)}{P(A)} \quad (4)$$

Al utilizar la regla de la multiplicación para evaluar el numerador de (4) y la regla de la probabilidad total para evaluar el denominador de (4), llegamos a

$$P(B_i \mid A) = \frac{P(B_i) \times P(A \mid B_i)}{\sum_{j=1}^{n} P(B_j) \times P(A \mid B_j)} \quad (5)$$

que representa la formulación del Teorema de Bayes para eventos publicada en 1763, (Bolstad 2004, p. 65).

Hay que repetir que en esta formulación los $B_i : i = 1,\ldots,n$ son eventos no observables, A es el evento observado, y $P(B_i)$ son las probabilidades establecidas *a priori*, o si se prefiere, las opiniones iniciales de los expertos. La presencia de estas probabilidades *a priori* ha sido la fuente de las discordias en la aplicación del teorema.

IV. El paso de eventos a teorías científicas

Las hipótesis científicas se expresan mediante distribuciones de probabilidad formuladas a partir de la observación de los datos. Estas distribuciones de probabilidad dependen de magnitudes desconocidas llamadas *parámetros*, en símbolos: θ. Desde el punto de vista bayesiano, el conocimiento sobre los *parámetros* es manifestado asignando una distribución de probabilidad a los mismos, es la distribución *a priori*, que en símbolos se escribe: $p(\theta)$.

Cuando nuevos datos, *y*, están disponibles, la información que contienen en relación a los *parámetros* es expresada en la *función de verosimilitud*, que es proporcional a la distribución de los datos observados, dados los parámetros, y que se representa como $p(y \mid \theta)$. Esta información es combinada con la opinión o probabilidad *a priori* para actualizar la distribución *a priori* $p(\theta)$.

El teorema de Bayes establece como esta actualización es matemáticamente realizada: la distribución *a posteriori* es proporcional a la distribución *a priori* multiplicada por la *función de verosimilitud,* o más precisamente:

$$p(\theta \mid y) = \frac{p(\theta) \times p(y \mid \theta)}{\int p(\theta) \times p(y \mid \theta) d\theta} \qquad (6)$$

expresión que se corresponde con (5). En teoría esta distribución *a posteriori* es evaluable, si bien puede ser muy complicada de obtener, y sólo modernamente se ha logrado mediante la obtención de muestras tomadas de $p(\theta \mid y)$ con ayuda de los ordenadores.

V. Otra formulación del Teorema de Bayes

En el numerador del teorema aparece la probabilidad *a priori* multiplicada por la función de verosimilitud, y en el denominador, la suma de probabilidades *a priori* multiplicada por la *función de verosimilitud*. Esta división hace que las probabilidades *a posteriori* sumen la unidad.

Ahora bien, si en el numerador multiplicamos cada *función de verosimilitud* por una constante, el denominador quedará multiplicado por la misma constante, y al efectuar la división se eliminará la constante, quedando la distribución de probabilidad *a posteriori*. Debido a esto, sólo necesitamos conocer la *función de verosimilitud* y la constante de proporcionalidad. Las ponderaciones relativas dadas por la *función de verosimilitud* a las distintas probabilidades es lo verdaderamente necesario.

Lo mismo cabe decir respecto a las probabilidades *a priori*, que pueden ser multiplicadas por una constante con lo que el denominador también quedaría multiplicado por la constante, y al dividir quedarían sólo las probabilidades *a posteriori*.

En definitiva, lo verdaderamente necesario es conocer las ponderaciones asignadas por la *función de verosimilitud* a las probabilidades.

Por esta razón, el Teorema de Bayes se suele expresar también como

$$posterior \propto prior \times verosimilitud \quad (7)$$

que se lee diciendo que la función de distribución de probabilidad *a posteriori* es proporcional a la función de probabilidad *a priori* multiplicada por la *función de verosimilitud*.

VI. Tras la herencia de Bayes

Aunque la metodología de Bayes fue seguida con entusiasmo por Laplace y otros notables matemáticos, sin embargo, cayó en desuso por los problemas inherentes al establecimiento de las probabilidades *a priori*. Así, en la primera mitad del siglo XX surgió el enfoque *frecuentista*, aunque la filosofía bayesiana no murió del todo. Pensadores como Bruno de Finetti, en Italia, y Harold Jeffreys en Inglaterra mantuvieron viva la llama.

El moderno movimiento bayesiano arranca en la segunda mitad del siglo XX, con nombres como Jimmy Savage en USA y Dennis Lindley en Inglaterra, si bien la realización práctica de la inferencia bayesiana tendrá que esperar hasta las últimas décadas del siglo, con el advenimiento de los ordenadores y los nuevos métodos de cálculo que los ordenadores han hecho posibles.

Estos nuevos métodos numéricos se engloban bajo las siglas MCMC (Marcov Chain Monte Carlo), y son métodos de integración que utilizan cadenas de Markov.

La integración Monte Carlo genera muestras de la distribución de probabilidad de interés y a partir de ellas evalúa los momentos de la distribución de probabilidad *a posteriori*. MCMC genera estas muestras por largo tiempo, y por ende las muestras son de gran tamaño.

Existen varias formas de construir estas cadenas, pero todas, incluido el método de Gibbs (Geman y Geman, 1984) son casos especiales del procedimiento desarrollado por Metropolis et al. (1953) y Hastings (1970). Llegados aquí, es necesario decir que han sido necesarios casi cuarenta años para que los métodos MCMC entraran en el ámbito de la estadística, desde el campo de la física, donde se originaron, y sólo en los últimos años es cuando se ha dejado notar su influjo en la estadística bayesiana e incluso en la estadística «clásica» (Gilks et al., 1996, p. 2).

VII. Inferencia bayesiana

La mayoría de las aplicaciones de los métodos MCMC están orientados a la estadística bayesiana. Desde esta perspectiva, no existe diferencia entre observables y no observables o parámetros. En un modelo estadístico bayesiano todas las magnitudes son aleatorias.

Si indicamos los datos observados mediante D, y los parámetros de un modelo, y las observaciones perdidas, por θ, es necesario formar la distribución de probabilidad conjunta $P(D,\theta)$. Esta distribución de probabilidad tiene dos componentes: $P(\theta)$, la distribución de probabilidad *a priori,* fruto de los conocimientos del investigador, y la *función de verosimilitud* $P(D \mid \theta)$. La especificación de ambas constituye el modelo probabilístico completo:

$$P(D,\theta) = P(D \mid \theta) \times P(\theta)$$

Una vez observados los datos D, el Teorema de Bayes permite establecer la distribución de probabilidad de θ condicionada por D:

$$P(\theta \mid D) = \frac{P(\theta) \times P(D \mid \theta)}{\int P(\theta) \times P(D \mid \theta) d\theta}$$

que, como sabemos, es la distribución de probabilidad *a posteriori* de θ.

Cualquier característica de esta distribución *a posteriori* puede ser analizada, a saber: los momentos, los cuantiles, las regiones de mayor probabilidad, etc. Así, por ejemplo:

$$E[f(\theta \mid D)] = \frac{\int f(\theta) P(\theta) P(D \mid \theta) d\theta}{\int P(\theta) P(D \mid \theta) d\theta}$$

La evaluación analítica de esta expresión puede ser prácticamente imposible, y sólo gracias a los procedimientos MCMC se logra hacerlo.

En un marco de referencia más general, sea X un vector de *k* variables aleatorias, con distribución de probabilidad $\pi(\cdot)$, y se trata de evaluar el valor esperado:

$$E[f(X)] = \frac{\int f(x) \pi(x) dx}{\int \pi(x) dx}$$

para una función $f(\cdot)$ de interés. Con la integración Monte Carlo, la evaluación de $E[f(X)]$ se obtiene a partir de las muestras $\{X_t; t = 1, 2, \ldots, n\}$ obtenidas de $\pi(\cdot)$ y promediándolas

$$E[f(X)] = \frac{1}{n} \sum_{t=1}^{n} f(X_t)$$

Es decir aproximamos la media poblacional mediante la media muestral.

Si las muestras $\{X_t\}$ son independientes, entonces la ley de los grandes números asegura que la aproximación puede ser tan precisa como se quiera, aumentando el tamaño de la muestra: *n*.

En general, la obtención de muestras de $\pi(\cdot)$ independientes no es factible dado que puede que no sea una distribución estándar. Sin embargo, las muestras $\{X_t\}$ no tienen que ser necesariamente independientes. Lo que sí se requiere es que provengan de $\pi(\cdot)$ de una manera apropiada, y una manera de lograrlo es a partir de cadenas de Markov que tengan a $\pi(\cdot)$ como distribución estacionaria de base, lo que se logra mediante los procedimientos de MCMC.

VIII. Cadenas de Markov

Sea la secuencia de variables aleatorias $\{X_0,X_1,X_2,...\}$ tal que en todo momento $t\geq 0$, el siguiente estado X_{t+1} es muestreado (tomado) de la distribución $P(X_{t+1} \mid X_t)$, que depende sólo del estado actual de la cadena: X_t y no de la historia de la cadena $\{X_0,X_1,...X_{t-1}\}$. Este tipo de secuencia recibe el nombre de cadena de Markov, y $P(\cdot \mid \cdot)$ se conoce como el *kernel* de transición de la cadena. Se supone que $P(\cdot \mid \cdot)$ es independiente de *t*.

Ahora bien, ¿cómo afecta X_0 a X_t? Todo depende de la distribución de X_t dado X_0, que podemos indicar mediante $P^t(X_t \mid X_0)$ sin tener en cuenta las demás variables de la cadena. En otras palabras, X_t depende directamente de X_0. En condiciones generales de regularidad, la cadena irá olvidando su estado inicial, y $P^t(\cdot \mid X_0)$ terminará convergiendo a una distribución estacionaria o invariante, independiente de *t* o de X_0. Sea esta distribución $\phi(\cdot)$.

Por tanto, al aumentar *t*, los valores muestrales $\{X_t\}$ irán asemejándose a los valores muestrales de $\phi(\cdot)$. Así, a partir de un largo periodo de adaptación (*burn-in*), por ejemplo de *m* repeticiones, los valores de $\{X_t;t=m+1,m+2,...,n\}$ serán valores que se aproximarán a los de la distribución $\phi(\cdot)$.

Volviendo a la evaluación del valor esperado $E[f(X)]$, teniendo X la distribución estacionaria $\phi(\cdot)$, y tras eliminar los valores iniciales de *burn-in*, tenemos el estimador

$$\bar{f} = \frac{1}{n-m} \sum_{i=m+1}^{n} f(X_i) \qquad (8)$$

que se conoce como el promedio ergódico, media ergódica, siendo el teorema ergódico el requisito que asegura la convergencia al valor esperado. Gilks et al. (1996), Roberts (1995), Tierney (1995).

La expresión en (8) muestra como una cadena de Markov puede ser utilizada para estimar $E[f(X)]$, tomando como base la distribución estacionaria $\phi(\cdot)$.

Ahora bien, la cuestión es ¿cómo construir una cadena de Markov tal que su distribución estacionaria $\phi(\cdot)$ se corresponda con nuestra distribución de interés $\pi(\cdot)$?

Para lograrlo existe el procedimiento debido a Hastings (1970), que es una generalización del procedimiento propuesto por Metropolis et al. (1953). Un caso especial del procedimiento Metropolis-Hastings lleva el nombre de *Gibbs sampler*, nombre dado por Geman y Geman (1984).

No es este el lugar para entrar en detalles sobre estos procedimientos. El lector interesado encontrará un resumen en Gilks et al. (1996), pp. 5-15. Véase también Albert (2007), Congdon (2003, 2005, 2006), Gamerman et al. (2006), Gelman et al. (2004), Gelman et al. (2007), Gill (2002), Geweke (2005), Lancaster (2004), Lynch (2007), Rizzo (2008), Rossi et al. (2005).

En estadística bayesiana, los resultados se presentan como resúmenes de la distribución *a posteriori* $\pi(\cdot)$, en términos de medias, desviaciones estándar, correlaciones, regiones de mayor probabilidad, etc., utilizando los estadísticos muestrales resultantes de los procedimientos MCMC.

IX. El enfoque bayesiano en el entorno R

Pasemos al lado aplicado de lo aquí expuesto.

R es un lenguaje y programa de software libre. O más en concreto, R es un entorno numérico, estadístico y gráfico para el análisis de datos. R se está convirtiendo en un fenómeno de alcance mundial. Por esta razón, son numerosas las aportaciones que extienden la operatividad de R a campos inicialmente no considerados, como pueden ser, el análisis del ADN, la minería de datos, etc., y el análisis bayesiano no podía ser menos. Tanto es así, que el número de marzo de 2006 de *R-News* está dedicado a la inferencia bayesiana y a la simulación Monte Carlo en el entorno R.

Para aquél entonces ya se contaba con 33 paquetes o bibliotecas de programas dedicados al tema, que a finales de septiembre de 2007 ya se aumentaba hasta 45. Estos paquetes, una vez instalados en R son totalmente operativos, tanto en lo que es específico del paquete como en lo que es propio de R. Además, todos los paquetes son revisados y actualizados continuamente.

9.1. *El proyecto BUGS*

Para comenzar, es obligatorio hacer referencia al Proyecto BUGS: *Bayesian inference Using Gibbs Sampling*. El proyecto comenzó en 1989, en la Unidad de Bioestadística del Medical Research Council, Cambridge. Posteriormente, en colaboración con la Escuela de Medicina de St. Mary's, Imperial College, Londres, pasa a denominarse WinBUGS. Más tarde, la Universidad de Helsinki, se une al Proyecto para el desarrollo de la nueva adaptación llamada OpenBUGS.

BUGS describe los modelos estadísticos mediante distribuciones conjuntas de probabilidad. Esta descripción es a la vez muy general y muy explícita. Por ser lo primero, es posible combinar modelos para lograr un modelo de mayor dimensión. Por ser muy explícita, impide la confusión entre modelos. En cuanto lenguaje de programación BUGS facilita a especialistas y profanos la práctica del análisis bayesiano.

La adaptación a R de BUGS tiene varias denominaciones, dependiendo de las distintas colaboraciones. Así, OpenBUGS en R se denomina BRugs, y para WinBUGS, tenemos R2WinBUGS

Independientemente del proyecto BUGS, está JAGS *(Just Another Gibbs Sampler)* desarrollado por Martyn Plummer y colaboradores, en el Science and Technology Facilities Council, Reino Unido. JAGS es otra implementación de los métodos MCMC, utilizando la descripción de modelos dada por BUGS.

Podemos clasificar las distintas adaptaciones de BUGS en dos grandes apartados, a saber, adaptaciones ordenadas a instrucciones y adaptaciones orientadas a modelos. Las primeras siguen de cerca el lenguaje BUGS, las segundas ponen la atención en los modelos, evitando al usuario el tener que conocer las instrucciones de BUGS.

9.2. *Los paquetes BRugs y MCMCpack*

Como ejemplo de las aplicaciones orientadas a instrucciones tenemos OpenBUGS, con sus dos adaptaciones en R: BRugs y R2WinBugs, y como ejemplo de aplicaciones orientadas a modelos tenemos MCMCpack (Andrew D. Martin, Kevin M. Quinn, Jong Hee Park) y Zelig (Kosuke Imai, Gary King, Olivia Lau), si bien este último es un completo entorno de cálculos estadísticos, bayesianos y no bayesianos, que permite la estimación de los modelos bayesianos utilizando MCMCpack.

Dadas las facilidades que los modernos ordenadores permiten, ya existen dos aplicaciones en un entorno totalmente bayesiano: el paquete rv (simulación basada en variables aleatorias) y Umacs: the *Universal Markov Chain Sampler,* frutos del esfuerzo de Jouni Kerman y Andrew Gelman.

Por último, hay que mencionar la existencia de CODA: *Convergence Diagnosis and Output Analysis* (Martyn Plummer, Nick Best, Kate Cowles, Karen Vines) y BOA: *Bayesian Output Analysis program* (Brian J. Smith). Estos dos paquetes están diseñados para completar y validar los resultados generados por los procedimientos MCMC. Ambos permiten análisis estadísticos, representaciones gráficas, aplicación de los criterios de convergencia existentes (gelman, geweke, heidel, raftery, etc.), establecimiento de las regiones de mayor probabilidad para los parámetros, etc.

Los programas citados son ejemplo de entornos completos de análisis bayesiano. Al margen de ellos, existen otros programas para aspectos específicos del enfoque bayesiano, como por ejemplo: BMA: *Bayesian Model Averaging* (Adrian Raftery, Jennifer Hoeting, Chris Volinsky, Ian Painter, Ka Yee Yeung) Cf. Alonso Rodríguez, 2006.

Todos los ejemplos citados son operativos en R, y descargables desde CRAN: *The Comprehensive R Archive Network*, en Viena. La excepción es JAGS, versión 1.0, todavía en fase de pruebas, al momento de escribir este artículo.

X. ILUSTRACIÓN NUMÉRICA

En lo que sigue, vamos a ilustrar el uso de MCMCpack y de BRugs, utilizando los datos del ejemplo propuesto por Wonnacott-Wonnacott en su texto *Econometrics*, segunda edición, p. 72, 1979.

10.1. *Los datos y el modelo*

El ejemplo propuesto por Wonnacott-Wonnacott hace referencia a la cosecha **y**, en kilos por ha., observada en una explotación agrícola en función del abono **x2**, en kilos por ha. y de la lluvia, **x3**, en litros por ha. (Las unidades de referencia están adaptadas al campo español: kilos, hectáreas y litros.)

Las observaciones para estas variables son:

	y	x2	x3
1	40	100	10
2	50	200	20
3	50	300	10
4	70	400	30
5	65	500	20
6	65	600	20
7	80	700	30

El modelo a estimar con estas variables es:

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t ; t = 1, \ldots, 7$$

Y se supone que se satisfacen los supuestos del modelo lineal general.

Utilizando R con sus posibilidades estadísticas básicas, tenemos el siguiente resultado:

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 28.095238   2.491482  11.277 0.000352 ***
x2           0.038095   0.005832   6.532 0.002838 **
x3           0.833333   0.154303   5.401 0.005690 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '

Residual standard error: 2.315 on 4 degrees of freedom
Multiple R-Squared: 0.9814,     Adjusted R-squared: 0.972
F-statistic: 105.3 on 2 and 4 DF,  p-value: 0.0003472
```

Es decir, las estimaciones puntuales de los parámetros son:

$$\hat{\beta}_1 = 28.095238; \hat{\beta}_2 = 0.038095; \hat{\beta}_3 = 0.833333; \hat{\sigma} = 2.315$$

10.2. *Aplicación del paquete MCMCpack*

Para utilizar el paquete MCMCpack, es necesario, una vez instalado, hacerlo activo en R, mediante la instrucción: `library(MCMCpack)`, que nos indica los paquetes incorporados para el correcto funcionamiento de MCMCpack, así como los créditos. Por tanto en pantalla aparece.

```
library(MCMCpack)
Loading required package: coda
Loading required package: lattice
Loading required package: MASS
##
## Markov Chain Monte Carlo Package (MCMCpack)
## Copyright (C) 2003-2008 Andrew D. Martin, Kevin M. Quinn,
   and Jong Hee Park
##
## Support provided by the U.S. National Science Foundation
## (Grants SES-0350646 and SES-0350613)
##
```

Una vez activado el paquete, pedimos que nos estime la distribución *a posteriori* de los parámetros:

```
posterior = MCMCregress(y ~ x2 + x3, data = datos)
```

Y al teclear `summary(posterior)`, obtenemos:

```
Iterations = 1001:11000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

                Mean        SD  Naive SE Time-series SE
(Intercept) 28.10882  3.443673 3.444e-02      3.272e-02
x2           0.03791  0.008035 8.035e-05      9.505e-05
x3           0.83728  0.213883 2.139e-03      2.349e-03
sigma2      10.30811 16.369795 1.637e-01      3.126e-01

2. Quantiles for each variable:

                2.5%     25%      50%      75%    97.5%
(Intercept) 21.21276 26.2700 28.12451 29.94507 34.99823
x2           0.02188  0.0337  0.03797  0.04216  0.05386
x3           0.41796  0.7222  0.83681  0.95189  1.25898
sigma2       1.92829  4.0114  6.36786 11.08251 41.52165
```

Este resultado nos muestra ahora, no el valor único de los parámetros estimados, como anteriormente, sino toda la distribución *a posteriori* de los mismos, con sus momentos: media, desviación estándar, y los cuantiles, con diversos porcentajes, estimaciones obtenidas sobre una cadena, que por defecto tiene una longitud igual a 10.000 muestras o valores.

El paquete MCMC está orientado a modelos, por esta razón, la instrucción para obtener los resultados es simplemente `MCMCregress`, seguida de las variables integrantes del modelo y del lugar donde se encuentran.

Con ayuda del paquete CODA, podemos obtener otra serie de resultados, por ejemplo la representación gráfica de las distribuciones *a posteriori* de los parámetros. Las tareas que CODA realiza se ven facilitadas por el uso de un *menú*.

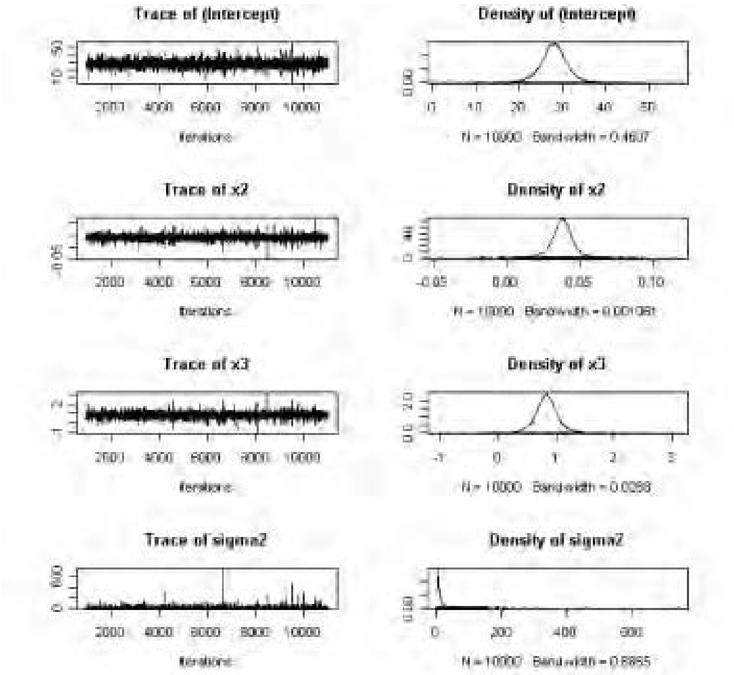
Dado que CODA ha sido incorporado en memoria al activar MCMCpack, para activar el menú, basta con teclear `codamenu()` y en pantalla aparece:

```
CODA startup menu

1: Read BUGS output files
2: Use an mcmc object
3: Quit

Selección: 2

Enter name of saved object (or type "exit" to quit)
1:posterior
Checking effective sample size ...OK
CODA Main Menu

1: Output Analysis
2: Diagnostics
3: List/Change Options
4: Quit
```

Por razones de espacio no podemos seguir mostrando el resto de opciones del menú. Nos limitaremos al siguiente ejemplo generado con su ayuda.

Al seleccionar `Plots,` obtenemos, figura 3:

FIGURA 3



que nos da el gráfico de las distribuciones de probabilidad *a posteriori* de los parámetros, así como las trazas mezcladas de los valores simulados por $\phi(\cdot)$ y los teóricos de la distribución de interé $\pi(\cdot)$. CODA permite examinar los criterios de convergencia desarrollados (Geweke, etc) para juzgar el gráfico de las trazas.

El programa nos brinda la posibilidad de guardar el gráfico como postscript.

10.2. *Aplicación del paquete BRugs*

La obtención de la distribución de probabilidad *a posteriori* de los parámetros es ahora un poco más laboriosa. Hay que comenzar escribiendo el modelo con la sintaxis de BUGS, almacenarlo en el *working directory* y dar las instrucciones según BUGS.

El modelo para nuestro caso comporta el siguiente conjunto de instrucciones, comenzando con la palabra *model*:

```
model
{
      for(i in 1:N){
            y[i] ~ dnorm(mu[i], tau)
            mu[i] <- beta1 + beta2*x2[i] + beta3*x3[i]
      }
tau ~ dgamma(0.001, 0.001)
sigma <- 1/sqrt(tau)
beta1 ~ dnorm(0.0, 1.0E-6)
beta2 ~ dnorm(0.0, 1.0E-6)
beta3 ~ dnorm(0.0, 1.0E-6)
}
```

Una vez escrito se guarda en el *working directory* como fichero ASCII y nombre *cosechamodelo.txt*.

El bucle iniciado por *for* establece que para cada *i*, la variable dependiente de nuestro caso, *y*, se distribuye normalmente, con media *mu* y precisión *tau*, y que *mu* varía a lo largo de todas las observaciones como una función lineal de x2 y x3. Es decir, *mu* es la función: $\beta_1 + \beta_2 x_{2_i} + \beta_3 x_{3_i}$. El bucle constituye la *función de verosimilitud*. Fuera del bucle, pero dentro del modelo, se indican las probabilidades a priori de los parámetros, que en nuestro caso se distribuyen (~) normalmente, salvo la *precisión* que lo hace como una variable con distribución *gamma*. Estas son las características estándar para el modelo lineal con errores homoscedásticos incorrelacionados. Cf. Lancaster, T. (2004), p. 375.

Los datos se guardan también como un fichero ASCII, y nombre *cosecha.txt*

Por último, hay que crear otro fichero con los valores iniciales de los parámetros, y guardarlo como fichero ASCII, y nombre *cosechainits.txt*. En este fichero se ha grabado:

```
list(beta1 = 0, beta2 = 0, beta3 = 0, tau = 1)
```

Una vez hecho lo anterior, entramos en R, y tecleamos las siguientes instrucciones. Para distinguir las instrucciones de los comentarios, escribiremos las primeras en `courier new`, **negrita**, y los comentarios en Times New Roman. Los mensajes de algunas de las instrucciones aparecen en pantalla en `courier new.`

`library(BRugs)`

```
Loading required package: coda
Loading required package: lattice
Welcome to BRugs running on OpenBUGS version 3.0.2
```

`setwd("c:/Archivos de programa/R/R-2.6.1/temp")`

esta instrucción fija el working directory, en donde se guardan los anteriores ficheros ASCII

`modelCheck("cosechamodelo.txt")`

pide que compruebe la corrección sintáctica del modelo; BRugs devuelve el mensaje:

```
model is sintactically correct
```

`datos = read.table("cosecha.txt", header=T)`

y al teclear **datos**, en pantalla aparece:

```
  y  x2  x3
1 40 100 10
2 50 200 20
3 50 300 10
4 70 400 30
5 65 500 20
6 65 600 20
7 80 700 30
```

Seguidamente, reasignamos los valores a las variables:

```
y = datos$y
x2 = datos$x2
x3 = datos$x3
N = nrow(datos)
```

y tecleamos:

```
data = list("y", "x2", "x3", "N")
bugsData(data, fileName=file.path(getwd(),
"cosedata.txt"),digits=5)
```

esta instrucción crea el fichero de datos **cosedata.txt** requerido por BRugs

```
modelData("cosedata.txt")
```

y en pantalla sale:

```
data loaded
```

```
modelCompile(numChains=2)
```

esta instrucción compila el modelo y genera dos cadenas de muestras

```
modelInits(rep("cosechainits.txt",2))
```

y en pantalla aparece:

```
Initializing chain 1: initial values loaded but this
or another chain contain uninitialized variables
Initializing chain 2: model is initialized
```

Volvemos a repetir la instrucción:

```
modelInits(rep("cosechainits.txt",2))
```

y aparece en pantalla

```
Initializing chain 1: model is initialized
Initializing chain 2: model is initialized
```

Para ver los objetos generados por el programa, teclear:

modelNames()

que devuelve los siguientes objetos

```
[1] "N"       "beta1"  "beta2"  "beta3"  "deviance" "mu"
 [7] "sigma"  "tau"    "x2"     "x3"     "y"
```

y para conservar a nuestra disposición estos resultados, hay que *monitorizarlos* tecleando:

samplesSet(c("beta1","beta2","beta3","deviance","tau","mu"))

que envía a pantalla:

```
monitor set for variable 'beta1'
monitor set for variable 'beta2'
monitor set for variable 'beta3'
monitor set for variable 'deviance'
monitor set for variable 'tau'
monitor set for variable 'mu'
```

modelUpdate(1000)

se simulan 1000 observaciones más

Para ver los estadísticos simulados:

samplesStats("*")

que envía a la pantalla del ordenador:

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
beta1	28.15000	4.90100	0.4704000	21.06000	27.86000	36.90000	1	2000
beta2	0.03789	0.01003	0.0004988	0.02193	0.03793	0.05398	1	2000
beta3	0.83640	0.30290	0.0273100	0.28460	0.84240	1.27500	1	2000
deviance	33.77000	5.29100	0.4754000	28.37000	32.42000	51.14000	1	2000
mu[1]	40.30000	2.96900	0.2401000	35.79000	40.10000	45.97000	1	2000
mu[2]	52.45000	2.41500	0.0931500	48.43000	52.38000	56.99000	1	2000
mu[3]	47.87000	3.10900	0.2849000	43.15000	47.74000	54.45000	1	2000
mu[4]	68.38000	3.22000	0.2323000	62.56000	68.40000	73.53000	1	2000
mu[5]	63.80000	1.98800	0.0835400	60.88000	63.78000	67.07000	1	2000
mu[6]	67.59000	2.71100	0.1279000	63.57000	67.54000	71.78000	1	2000
mu[7]	79.74000	2.88800	0.1510000	75.06000	79.84000	84.26000	1	2000
tau	0.18990	0.13680	0.0061840	0.01054	0.15880	0.53500	1	2000

Para otros estadisticos de interés, como el DIC (Deviance Information Criterion), primero se monitoriza mediante

```
dicSet()
```

y en pantalla aparece el mensaje:

```
deviance set
```

luego se actualiza el modelo con otras 1000 muestras

```
modelUpdate(1000)
```

y aparece el mensaje:

```
1000 updates took 0 s
```

y para ver el resultado,

```
dicStats()
```

que pone en pantalla

```
        Dbar  Dhat   DIC    pD
y      33.29 28.67 37.91 4.619
total  33.29 28.67 37.91 4.619
```

En esta ocasión hemos procedido instrucción por instrucción. Más correcto hubiera sido reunir todas las instrucciones en un fichero de instrucciones, técnicamente denominado *script*, y ejecutarlas de un golpe, solicitando los resultados según convenga.

Para utilizar el programa CODA necesitamos adaptar los resultados anteriores al formato apropiado. Para ello

```
Uno = buildMCMC("beta1")
Dos = buildMCMC("beta2")
Tres = buildMCMC("beta3")
Cuatro = buildMCMC("tau")
Cinco = buildMCMC("mu")
```

Y seguidamente, tecleamos

```
codamenu()
```

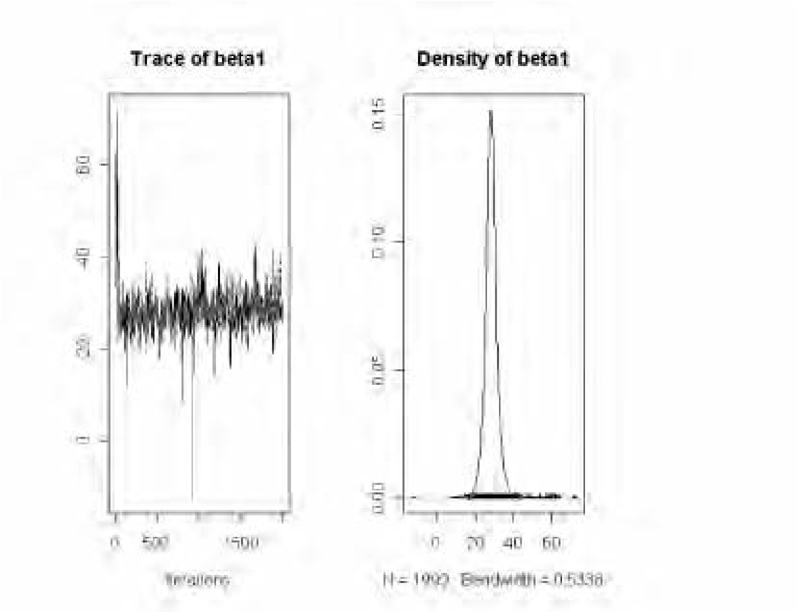
apareciendo en pantalla:

```
CODA startup menu

1: Read BUGS output files
2: Use an mcmc object
3: Quit

Selección: 2

Enter name of saved object (or type "exit" to quit)
1:Uno
Checking effective sample size ...
```

De nuevo, razones de espacio nos impide mostrar el menú y sus posibilidades. No obstante, veamos un par de resultados.

El gráfico de la distribución de probabilidad del parámetro **beta1**, y la mezcla de las trazas correspondientes, aparece en la figura 4

FIGURA 4



Si elegimos:

2: Statistics.

el resultado para cada cadena (se generaron dos) es:

```
chain1

Iterations = 2:2000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 1999

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

          Mean             SD       Naive SE Time-series SE
       28.5765         4.5353         0.1014         0.5618

2. Quantiles for each variable:

 2.5%   25%   50%   75% 97.5%
22.08 26.32 28.14 30.00 37.90

chain2

Iterations = 2:2000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 1999

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

          Mean             SD       Naive SE Time-series SE
      28.36203        3.64642        0.08156        0.31261

2. Quantiles for each variable:

 2.5%   25%   50%   75% 97.5%
21.31 26.66 28.36 30.11 35.25
```

El otro paquete para analizar los resultados generados por el método MCMC es BOA. Si queremos utilizar el paquete BOA para el análisis del output generado por BRugs, tecleemos:

```
library(boa)

boa.menu()
```

y tras los créditos y avisos oportunos, aparece:

```
BOA MAIN MENU
*************

1: File              >>
2: Data Management   >>
3: Analysis          >>
4: Plot              >>
5: Options           >>
6: Window            >>
```

Y se brinda la posibilidad de seleccionar una de las tareas:

```
Selección: 1
```

Un nuevo menú parece en pantalla, brindando la oportunidad de seleccionar otra tarea.

```
FILE MENU
=========

1: Back
2: ----------------------+
3: Import Data        >> |
4: Save Session          |
5: Load Session          |
6: Exit BOA              |
7: ----------------------+
```

```
Selección: 3
```

Otro menú nos permite importar el tipo de objeto a examinar:

```
IMPORT DATA MENU
----------------

1: Back
2: --------------------------+
3: CODA Output Files         |
4: Flat ASCII Files          |
5: Data Matrix Objects       |
6: View Format Specifications|
7: Options...                |
8: --------------------------+

Selección: 5

Enter object name [none]
1: Uno
Read 1 item
+++ Object successfully imported +++
```

Apareciendo el mensaje de que el objeto ha sido importado correctamente.

Y seleccionando `Back` dos veces se vuelve al menú principal para realizar el análisis, los gráficos, etc. Así, el gráfico de la distribución de probabilidad *a posteriori* de **beta1**, es el siguiente, figura 5:

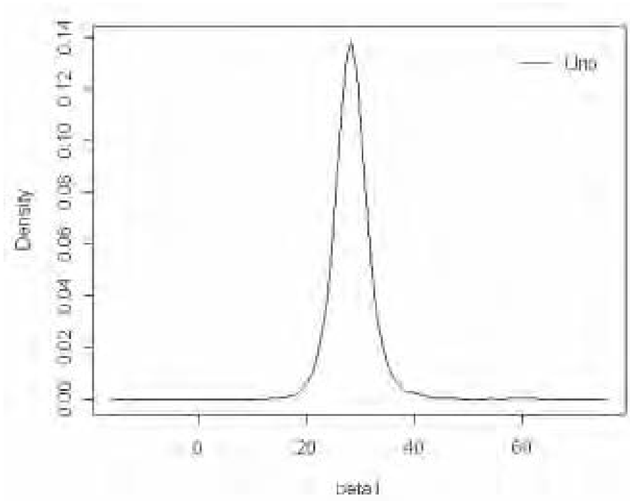


FIGURA 5

El menú sigue ofreciendo otras posibilidades, pero aquí lo dejamos.

XI. RESUMEN

Dentro de las limitaciones de espacio inherentes a todo artículo, esperamos haber mostrado al lector cómo los ordenadores han permitido hacer viable el enfoque bayesiano, presenciando, por tanto, el momento presente un verdadero resurgir de Thomas Bayes.

XII. BIBLIOGRAFÍA

(Nota. La indicación ® al final de una referencia indica la utilización del Programa R, o una adaptación operativa en R, dentro de la misma.)

ALBERT, J., *Bayesian computation with R*, Springer, New York 2007 ®.

ALONSO RODRÍGUEZ, A., «Un procedimento bayesiano para la selección de modelos», *Anuario Jurídico y Económico Escurialense*, XXXIX (2006) 427-442 ®.

BAYES, Th., *An Essay towards solving a problem in the Doctrine of Chances*, 1763; reimpreso en *Biometrika*, 45 (1958) 296-315.

BARNARD, G. A., «Thomas Bayes, a biographical note», *Biometrika* (1958) 293-295.
BERNARDO, J. M., y SMITH, A. F. M., *Bayesian Theory*, John Wiley, New York 1994.
BMA, contributed package, en CRAN: The Comprehensive R Archive Network, Viena.
BOA, contributed package, en CRAN: The Comprehensive R Archive Network, Viena.
BOLSTAD, W. M., *Introduction to Bayesian Statistics*, John Wiley, Haboken, New Jersey 2004 ®.
BOX, G. E. P., y TIAO, G. C., *Bayesian Inference in Statistical Analysis*, Addison-Wesley Pub. Co., Reading. Massachusetts 1973.
BRugs, contributed package, en CRAN: The Comprehensive R Archive Network, Viena.
BUGS, ver Lunn et al.
CHAMBERS, J. M., *Programming with Data, A Guide to the S Language*, Springer, New York 1998.
CODA, contributed package, en CRAN: The Comprehensive R Archive Network, Viena.
CONGDON, P., *Applied Bayesian Modelling*, John Wiley, Chichester, West Sussex 2003 ®.
–*Bayesian Models for Categorical Data*, John Wiley, Chichester, West Sussex 2005 ®.
–*Bayesian Statistical Modelling*, segunda ed., John Wiley, Chichester, West Sussex 2006 ®.
CRAN: The Comprehensive R Archive Network, http://cran.r-project.org/
GAMERMAN, D., y LOPES, H. F., *Markov Chain Monte Carlo, Stochastic Simulation for Bayesian Inference*, 2.ª ed., Chapman & Hall/CRC, Boca Raton, Florida 2006 ®.
GELMAN, A.; CARLIN, J. B.; STERN, H. S., y RUBIN, D. B., *Bayesian Data Analysis*, 2.ª ed., Chapman & Hall/CRC, Boca Raton, Florida 2004 ®.
GELMAN, A., y HILL, J., *Data Analysis using Regression and Multilevel/Hierarchical Models*, Cambridge University Press, New York 2007 ®.
GEMAN, S., y GEMAN, D., «Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images», en *IEEE Transactions in Pattern Analysis and Machine Intellegence*, 6 (1984) 721-741.
GEWEKE, J., *Contemporary Bayesian Econometrics and Statistics*, John Wiley, Haboken, New Jersey 2005.
GILKS, W. R.; RICHARDSON, S., y SPIEGELHALTER, D. J., *Markov Chain Monte Carlo in Practice*, Chapman & Hall/CRC, Boca Ratón, Florida 1996 ®.
GILL, J., *Bayesian Methods: A Social and Behavioral Sciences Approach*, Chapman & Hall/CRC, Boca Raton, Florida 2002 ®.

HASTINGS, W. K., «Monte Carlo sampling methods using Markov chains and their applications», *Biometrika*, 57 (1970) 97-109.

JEFFREYS, M., *Theory of Probability*, 3.ª ed., Oxford University Press, Oxford 1961.

KOLMOGOROV, A. N., *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Springer-Verlag, Berlín 1933.

LANCASTER, T., *An Introduction to Modern Bayesian Econometrics*, Blackwell Publishing, Oxford 2004 ®.

LUNN, D. J.; THOMAS, A.; BEST, N.,y SPIEGELHALTER, D., «WinBUGS — a Bayesian modelling framework: concepts, structure, and extensibility», en *Statistics and Computing*, 10 (2000) 325-337.

LYNCH, S. M., *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*, Springer, New York 2007 ®

MCMCpack, contributed package, en CRAN: The Comprehensive R Archive Network, Viena.

METROPOLIS, N.; ROSENBLUTH, T. H.; ROSENBLUTH, M. N.; TELLER, A. H., y TELLER, E., «Equations of state calculations by fast computing machines», *Journal of Chemical Physics*, 21 (1953) 1087-1091.

R Development Core Team (2007). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria 1953. ISBN 3-900051-07-0, URL http://www.R-project.org.

RIZZO, M. L., *Statistical Computing with R*, Chapman & Hall/CRC, Boca Raton, Florida 2008 ®.

R-News, en CRAN: The Comprehensive R Archive Network, Viena ®.

ROBERTS, G. O., «Markov chain concepts related to sampling algorithms», en GILKS et al, *Markov Chain Monte Carlo in Practice,* Chapman&Hall/CRC, Boca Ratón, Florida 1996 ®.

ROSSI, P. E.; ALLENBY, G. M., y MCCULLOCH, R., *Bayesian Statistics and Marketing*, John Wiley, Chichester, West Sussex 2005 ®.

R2WinBugs, contributed package, en CRAN: The Comprehensive R Archive Network, Viena.

rv, contributed package, en CRAN: The Comprehensive R Archive Network, Viena.

S, ver CHAMBERS, J. M.

SAVAGE, L. J., *The Foundations of Statistics*, Dover Publications, New York 1972.

STRANGE, C. H., *Nonconformity in Tunbridge Wells*, Tunbridge Wells 1949, citado por BARNARD, o.c.

TIERNEY, L., «Introduction to general state-space Markov chain theory», en GILKS et al, *Markov Chain Monte Carlo in Practice,* Chapman & Hall/CRC, Boca Ratón, Florida 1996 ®.

Umacs, contributed package, en CRAN: The Comprehensive R Archive Network, Viena.

WONNACOTT, R. J., y WONNACOTT, Th. H., *Econometrics*, 2.ª ed., John Wiley, New York 1979.

Zelig, contributed package, en CRAN: The Comprehensive R Archive Network, Viena.

ZELLNER, A., *An Introduction to Bayesian Inference in Econometrics*, John Wiley, New York 1971.