Revista Colombiana de Estadística
Diciembre 2008, volumen 31, no. 2, pp. 145 a 155

Distribución predictiva bayesiana para modelos de pruebas de vida vía MCMC

The Bayesian Predictive Distribution in Life Testing Models via MCMC

Carlos Javier Barrera[1,a], Juan Carlos Correa[2,b]

[1]Facultad de Ciencias Básicas, Instituto Tecnológico Metropolitano Institución Universitaria (ITM), Medellín, Colombia

[2]Escuela de Estadística, Facultad de Ciencias, Universidad Nacional de Colombia, Medellín, Colombia

Resumen

En el estudio de la confiabilidad es muy frecuente el desconocimiento de parámetros poblacionales; por tanto, es necesario recoger información muestral relevante para la estimación de estos a través de distribuciones de probabilidad, conocidas como distribución a priori. Los métodos bayesianos permiten incorporar opiniones subjetivas acerca de incertidumbres con respecto al parámetro o vector de parámetros de interés. La incertidumbre acerca del verdadero valor de un parámetro de interés θ en la población es modelada por la función de densidad a priori $\pi(\theta)$, $(\theta \in \Theta)$. Para obtener las distribuciones predictivas bayesianas, se implementará la metodología MCMC, la cual exige calibración, diseño, implementación y validación de algoritmos apropiados.

***Palabras clave*:** a priori, distribución predictiva, fiabilidad, MCMC.

Abstract

In reliability studies it is common to not know the population parameters, therefore, it becomes necessary to collect a sample in order to estimate the parameters of the assumed probability distribution. Bayesian methods allow to incorporate subjective information about uncertainties regarding the parameter or parameters of interest. From the bayesian point of view, the uncertainty about the true value of a parameter of interest θ in the population, is modeled by the prior density function $\pi(\theta)$, $(\theta \in \Theta)$. We will implement the methodology MCMC to obtain the predictive bayesian distributions, which requires the calibration, design, implementation, in addition to the validation of appropriate algorithms.

***Key words*:** Prior, Predictive Distribution, Reliability, MCMC.

[a]Docente tiempo completo especial. E-mail: cjbarrer@unal.edu.co
[b]Profesor asociado. E-mail: jccorrea@unalmed.edu.co

1. Introducción

Los rápidos avances tecnológicos, los grandes desarrollos de productos sofisticados, la intensa competición global y las cada vez mayores expectativas de los clientes han puesto más presión en el procedimiento de manufactura de productos de alta calidad. Los intervalos predictivos permiten predecir la duración y el costo de garantía de un producto; además, permite hacer un seguimiento del producto en campo para proporcionar información de las posibles causas de falla y los métodos para mejorar la confiabilidad del producto. En los estudios de confiabilidad es muy frecuente el desconocimiento de parámetros poblacionales; por tanto es necesario recopilar información muestral relevante para los parámetros (Meeker & Escobar 1998).

La estadística bayesiana permite la incorporación de información no muestral sobre características desconocidas del fenómeno en estudio. Hewett (1968), Kalbfleisch (1971), Dunsmore (1974) y Christensen & Huffman (1985), entre otros, han realizado trabajos en el área de distribuciones predictivas.

La utilización de distribuciones conjugadas es una limitante para expresar el conocimiento a priori acerca de un parámetro desconocido; sin embargo, cuando se utilizan distribuciones a priori no conjugadas, el problema es computacional, ya que usualmente se llegan a expresiones muy complejas y sin solución cerrada de la distribución a posteriori. Por esta razón, las soluciones vía simulación son las más aceptadas para la solución de este tipo de problemas (Hill 2002).

En el presente artículo se construye la distribución predictiva bayesiana para datos de sobrevida a partir de distribuciones a priori no conjugadas, no informativas. Se emplea la *Markov chain Monte Carlo Methods* (metodología Monte Carlo por cadenas de Markov –MCMC–) para obtener los parámetros de las distribuciones a posteriori, la cual exige calibración, diseño, implementación y validación de algoritmos apropiados.

En la primera parte de este artículo, se presenta el desarrollo teórico de la metodología MCMC. Posteriormente, se realiza una aplicación para datos de sobrevida de computadores donde se implementará la anterior metodología. Se utilizarán distribuciones a priori no conjugadas y se obtendrán las respectivas distribuciones predictivas bayesianas para el modelo Weibull en presencia de observaciones censuradas. En la tercera sección, se verifica la convergencia de los algoritmos utilizados por el proceso MCMC y, en la última sección, se presentan las conclusiones respectivas.

2. Metodología

El problema que se estudiará puede tipificarse así: sea X una variable aleatoria que representa el tiempo de vida de un producto con fdp (función de densidad de probabilidad) $p(x \mid \theta)$ $(x \in \mathbb{X}; \theta \in \Theta)$, donde θ es un parámetro o vector de parámetros que caracteriza la distribución. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de esta distribución. Sea además $Y_1, Y_2, \ldots, Y_N$ una segunda muestra

aleatoria de observaciones futuras independientes de una distribución con función de densidad de probabilidad $p(y \mid \theta)$ $(y \in \mathbb{Y}; \theta \in \Theta)$. Se desea hacer predicciones acerca de alguna función de $Y_1, Y_2, \ldots, Y_N$. Se asume que las dos distribuciones contienen el mismo parámetro θ (Dunsmore 1974).

El conocimiento previo acerca de los parámetros de una distribución de interés se expresa en una distribución de probabilidad, la cual se conoce como distribución a priori. La determinación de la distribución a priori es un problema fundamental en la estadística bayesiana. En muchos casos, es necesario utilizar procesos de elicitación para obtener estas distribuciones.

La distribución a posteriori se calcula mediante el teorema de Bayes como:

$$\pi(\theta \mid Datos) \propto \pi(\theta)\, L(\theta \mid Datos)$$

Si Y es una muestra aleatoria de observaciones futuras de determinado proceso, con fdp $p(y \mid \theta)$, entonces la distribución predictiva de Y es:

$$p(y \mid Datos) = \int_{\boldsymbol{\Theta}} p(y \mid \theta)\pi(\theta \mid Datos)\, d\theta = E_{\theta|Datos}\left[p(y \mid \theta)\right]$$

con $\theta \in \Theta$ (Hill 2002).

Cuando no es factible el cálculo directo de la distribución a posteriori, se utiliza la metodología MCMC.

Cuando las distribuciones a posteriori son de alta dimensión o cuando no tienen una forma distribucional conocida, las soluciones analíticas o numéricas comunes no pueden obtenerse. Una solución es generar muestras para los parámetros de interés considerando un procedimiento MCMC, donde se simula una cadena de Markov con distribución estacionaria dada por la distribución a posteriori $\pi(\theta \mid Datos)$ (Hill 2002).

Los métodos MCMC son algoritmos iterativos que se utilizan cuando no es factible el muestreo directo de una distribución de interés $\pi(\theta \mid Datos)$.

Una cadena de Markov es generada muestreando

$$\theta^{(t+1)} \sim \pi(\theta \mid \theta^{(t)})$$

donde π es llamado el *kernel de transición* de la cadena de Markov. En una cadena de Markov, la muestra $\theta^{(t+1)}$ depende solo de $\theta^{(t)}$ y no de $\theta^{(0)}, \theta^{(1)}, \ldots, \theta^{(t-1)}$ (Kao 1997).

El muestreador de Gibbs es un kernel de transición creado por una serie de distribuciones condicionales completas, es decir, un esquema de actualización markoviana basada en todas las probabilidades condicionales declaradas.

Si la distribución límite de interés es $\pi(\theta)$, donde θ es un vector de dimensión k de parámetros a estimar, entonces el objetivo es producir una cadena de Markov que, a través de ciclos, garantice que las declaraciones condicionales se muevan alrededor de esta distribución. Sea Θ el conjunto de todas las distribuciones condicionales para θ, las cuales se definen como: $\pi(\Theta) = \pi(\theta_i \mid \theta_{-i})$, para $i = 1, 2, \ldots, k$, donde la notación θ_{-i} indica una forma paramétrica específica de Θ sin el coeficiente θ_i (Casella & George 1992).

3. Estimación de la distribución predictiva vía MCMC

Para obtener la distribución predictiva bayesiana vía MCMC, se utilizó el muestreador de Gibbs, con el cual generamos muestras de la distribución a posteriori de los parámetros. En el caso estudiado, donde se asume que las observaciones muestrales se distribuyen Weibull, la distribución a posteriori presenta dos parámetros de interés (α y β). Así, se hacen m iteraciones del muestreador de Gibbs para obtener m pares de valores de los parámetros (α_i, β_i). Para una secuencia de valores que se hallen en el rango de posibles resultados para las observaciones futuras y, se remplazaron los pares (α_i, β_i) en $p(y \mid \alpha_i, \beta_i)$, $i = 1, 2, \ldots, m$, formando m funciones. La fdp predictiva bayesiana se obtuvo calculando el promedio ergódico de las anteriores m funciones de distribución, es decir, calculando

$$\frac{1}{m}\sum_{i=1}^{m} p(y \mid \alpha_i, \beta_i)$$

que es una mezcla de m densidades.

Este promedio es una densidad, ya que es una combinación lineal convexa de densidades:

$$\sum a_i p_i, \quad \text{con } a_i > 0, \quad \text{y}$$

$$\sum a_i = 1 \implies \int_{\mathbf{Y}} \sum a_i p_i \, dy = \sum a_i \int_{\mathbf{Y}} p_i \, dy = \sum a_i = 1$$

donde a_i es una constante. En general, la fdp predictiva se obtiene así:

$$p(y \mid x) = E_{\theta|X}\left[p(y \mid \theta_{1_i}, \theta_{2_i})\right] \approx \frac{1}{m}\sum_{i=1}^{m} p(y \mid \theta_{1_i}, \theta_{2_i})$$

donde X es el vector de datos y θ el vector de parámetros.

Usualmente las distribuciones predictivas resultan de forma cerrada o se resuelve la integral por medio de métodos numéricos. En este problema, la distribución predictiva resulta de alta densidad; por tanto, se obtiene a través de un promedio de densidades, el cual no aparece en la literatura y es muy simple de hallar computacionalmente (Hill 2002, Hewett 1968, Kalbfleisch 1971, Dunsmore 1974, Christensen & Huffman 1985, Komaki 2001, entre otros).

4. Aplicación

Para llevar a cabo esta aplicación, se utilizaron los datos proporcionados por la sección de bienes y suministros, y la oficina de soporte técnico de la Universidad Nacional de Colombia, sede Medellín, donde se tomaron 72 observaciones, de las cuales 55 fueron censura. Estas observaciones registran los tiempos para la primera falla física de los computadores de escritorio del bloque 21 en la sede. No se

consideraron fallas debido al usuario; por ejemplo, si el computador presenta una falla debido a que el usuario ha borrado un archivo del sistema, entonces esa falla no se registra en nuestra base de datos. Todos los computadores en estudio son Pentium IV de 1.5 GHZ, Dell Optiplex, modelo GX240, comprados en la misma fecha (febrero 4 de 2002). Estos equipos, desde el mes de compra, han estado diariamente en funcionamiento, realizando actividades comunes en salas de cómputo del bloque 21. Los datos son los presentados en la tabla 1.

Tabla 1: Tiempos (en meses) para la primera falla de 72 computadores.

14.07	17.80	19.43	21.33	24.60	28.97	29.63	33.73	37.60	37.67	40.87	52.40
53.97	60.57	64.27	65.43	65.43	66^+	66^+	66^+	66^+	66^+	66^+	66^+
66^+	66^+	66^+	66^+	66^+	66^+	66^+	66^+	66^+	66^+	66^+	66^+
66^+	66^+	66^+	66^+	66^+	66^+	66^+	66^+	66^+	66^+	66^+	66^+
66^+	66^+	66^+	66^+	66^+	66^+	66^+	66^+	66^+	66^+	66^+	66^+
66^+	66^+	66^+	66^+	66^+	66^+	66^+	66^+	66^+	66^+	66^+	66^+

Las observaciones censuradas (66^+) indican que en el momento de obtener los datos para realizar el presente estudio, es decir, a los 66 meses de haber adquirido los computadores, estos equipos no habían presentado la primera falla física.

El objetivo de esta aplicación es modelar el tiempo desde que se compra un computador específico hasta que presente la primera falla, empleando la metodología MCMC para obtener los parámetros de la distribución a posteriori. Los procedimientos utilizados para generar las gráficas de las distribuciones a posteriori y predictiva, y el chequeo de convergencia de las a posteriori, fueron realizados con el programa R (R Development Core Team 2007).

4.1. Estimación de las distribuciones a priori y a posteriori

Las distribuciones a priori y a posteriori se estimaron asumiendo que las observaciones muestrales siguen una distribución Weibull en presencia de observaciones censuradas.

4.1.1. Distribución Weibull cuando hay censura

Suponga que la distribución de las observaciones experimentales es una Weibull; por tanto,

$$p(x \mid \alpha, \beta) = \frac{\alpha}{\beta} x^{\alpha-1}\, e^{-x^{\alpha}/\beta}, \qquad 0 \le x < \infty, \qquad \alpha > 0, \qquad \beta > 0$$

La estimación de la distribución a priori de Jeffreys lleva a cálculos complejos de los parámetros de la distribución a posteriori, la cual viene dada por la expresión

$$\pi(\alpha, \beta \mid Datos) \propto \left| E\left[\frac{\alpha^2 x^{\alpha}\, \ln(x)^2 (x^{\alpha} - \beta) + \beta(2x^{\alpha} - \beta)}{\alpha^2 \beta^4} \right] \right|^{1/2} \frac{\alpha^n}{\beta^n}\, e^{-(\sum_{i=1}^{n} x_i^{\alpha})/\beta} \prod_{i=1}^{n} x_i^{\alpha-1}$$

La complejidad de esta distribución implica problemas computacionales en el cálculo de los parámetros a partir de las distribuciones condicionales. Se utilizará la distribución a priori no informativa de Laplace, donde

$$\pi(\alpha, \beta) = 1$$

La verosimilitud, cuando se consideran los datos censurados, viene dada por

$$L(\alpha, \beta \mid Datos) = \frac{\alpha^{17}}{\beta^{17}} e^{-\left(\sum_{i=1}^{72} x_i^{\alpha}\right)/\beta} \prod_{i=1}^{17} x_i^{\alpha-1}$$

Entonces, la distribución a posteriori para α y β está dada por

$$\pi(\alpha, \beta \mid Datos) \propto \frac{\alpha^{17}}{\beta^{17}} e^{-\left(\sum_{i=1}^{72} x_i^{\alpha}\right)/\beta} \prod_{i=1}^{17} x_i^{\alpha-1}$$

Posteriormente, se construye la respectiva fdp predictiva bayesiana.

4.2. Cálculo de los parámetros de las distribuciones a posteriori por medio de la metodología MCMC

La distribución a posteriori, cuando hay censura, viene dada por

$$\pi(\alpha, \beta \mid Datos) \propto \frac{\alpha^{17}}{\beta^{17}} e^{-\left(\sum_{i=1}^{72} x_i^{\alpha}\right)/\beta} \prod_{i=1}^{17} x_i^{\alpha-1}$$

Se usó el muestreador de Gibbs para generar muestras a partir de las siguientes distribuciones condicionales de α y β.

La distribución condicional de α, dados β y los datos, es

$$\pi(\alpha \mid \beta, Datos) \propto \alpha^{17} e^{-\left(\sum_{i=1}^{72} x_i^{\alpha}\right)/\beta} \prod_{i=1}^{17} x_i^{\alpha-1}$$

La condicional de β, dados α y los datos, es

$$\pi(\beta \mid \alpha, Datos) \propto \frac{1}{\beta^{17}}\, e^{-\left(\sum_{i=1}^{72} x_i^{\alpha}\right)/\beta}$$

La figura 1 muestra la distribución a posteriori estimada para α y β.

Ahora, como

$$p(y \mid x) = E_{\theta|Datos}\left[\frac{\alpha}{\beta} y^{\alpha-1}\, e^{-(y^{\alpha})/\beta}\right] \approx \frac{1}{m} \sum_{i=1}^{m} \frac{\alpha}{\beta} y^{\alpha_i - 1}\, e^{-y^{\alpha_i}/\beta_i}$$

donde $\theta = (\alpha, \beta)$.

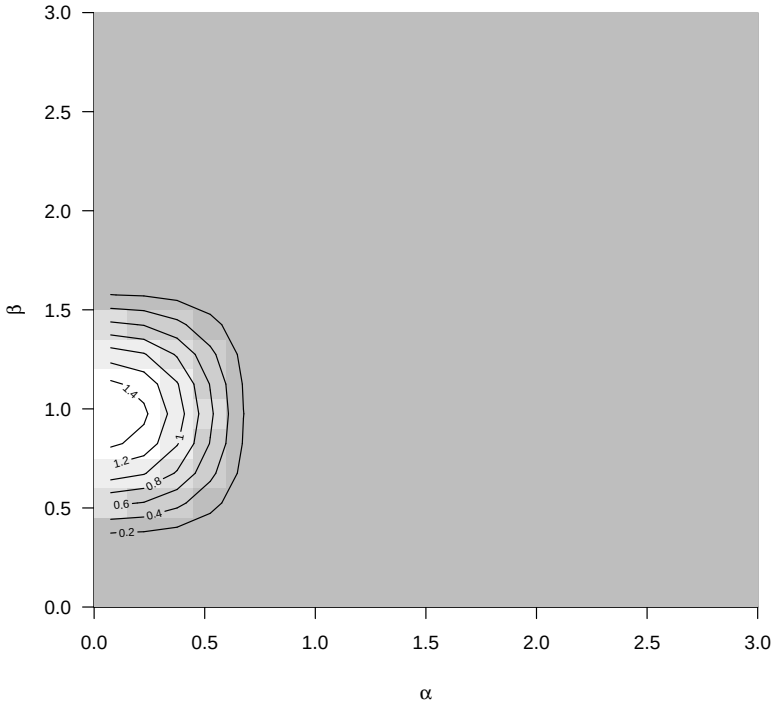

FIGURA 1: Estimación de la distribución a posteriori para α y β.

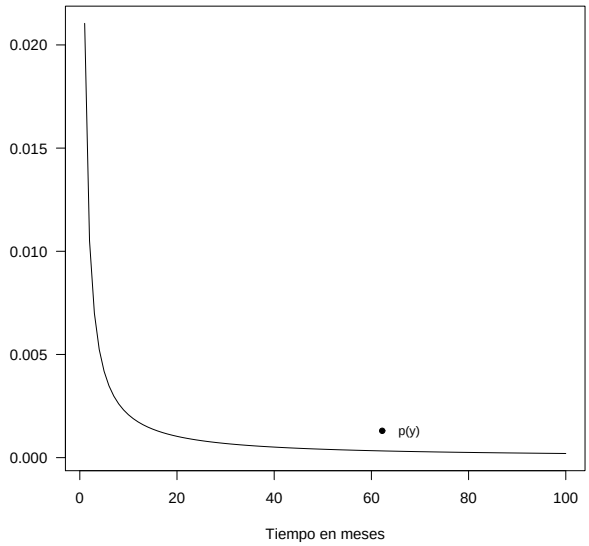

FIGURA 2: fdp predictiva bayesiana, cuando hay censura.

Entonces puede aproximarse la predictiva por este promedio ergódico.

La figura 2 muestra la fdp predictiva bayesiana.

Si se quiere construir el intervalo de probabilidad al 95 %, que es equivalente al intervalo de confianza en la estadística clásica, para el tiempo promedio de ocurrencia de la primera falla física de un computador de escritorio, entre los 72 equipos en estudio, este sería 1.85 meses.

4.3. Chequeo de convergencia

Para verificar la convergencia de los algoritmos, se utilizó el test KPSS (Kwiatkowski-Phillips-Schmidt-Shin) que viene incorporado en la función `kpss.test` del paquete `tseries` del lenguaje R (R Development Core Team 2007). Este test tiene el siguiente juego de hipótesis:

$H_0 =$ La cadena de Markov ha alcanzado la distribución estacionaria

vs.

$H_1 =$ La cadena de Markov no ha alcanzado la distribución estacionaria

En la prueba se utiliza el estadístico LM desarrollado por Kwiatkowski et al. (1992).

Para el chequeo de convergencia de la cadena de Markov generada de la distribución condicional de α, cuando hay censura y cuando se considera que los datos se distribuyen Weibull, la condicional para el parámetro α es:

$$\pi(\alpha \mid \beta, Datos) \propto \alpha^n \prod_{i=1}^{n} x_i^{\alpha-1}\, e^{-\left(\sum_{i=1}^{n} x_i^{\alpha}\right)/\beta}$$

Ahora, generando 5000 muestras a partir de esta distribución y quemando las primeras 1000, se tiene, que el nivel KPSS es 0.1035, el parámetro de truncamiento es 6 y el valor p es 0.1. Por tanto, no se rechaza la hipótesis nula de que la cadena de Markov haya alcanzado la distribución estacionaria; sin embargo, se realizan diagnósticos gráficos y se verifica la correlación existente entre los valores generados para α con distintos rezagos.

Las autocorrelaciones para α con diferentes rezagos se muestran en la tabla 2.

Tabla 2: Autocorrelaciones de los α con diferentes rezagos.

	Autocorrelación de α con			
	1 rezago	5 rezagos	10 rezagos	50 rezagos
α	0.02955384	−0.03599145	0.06363447	0.03170772

En la tabla 2 se observa que no existe una fuerte asociación entre los valores de los parámetros generados con diferentes rezagos.

La figura 3 muestra, respectivamente, los promedios móviles y la densidad para la cadena de valores generados de la distribución condicional de α.

Con base en la figura 3, los resultados de la prueba KPSS y las autocorrelaciones con los diferentes rezagos, puede afirmarse que la distribución para el parámetro α es estacionaria.

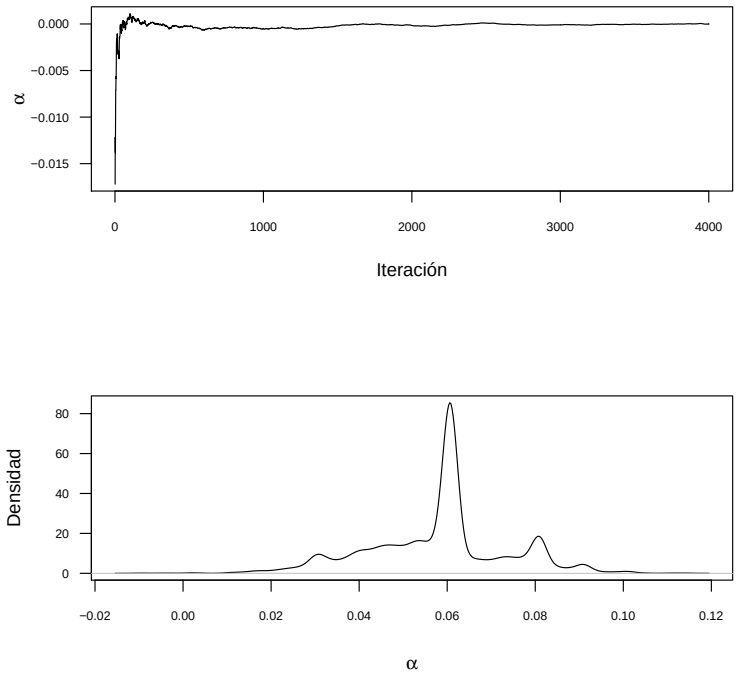

FIGURA 3: Promedios móviles y densidad para α, respectivamente.

Para el chequeo de convergencia de la cadena de Markov generada de la distribución condicional de β, cuando hay censura y cuando se considera que los datos se distribuyen Weibull, la distribución condicional para el parámetro β es

$$\pi(\beta \mid \alpha, Datos) \propto \frac{1}{\beta^n}\, e^{-\left(\sum_{i=1}^{n} x_i^{\alpha}\right)/\beta}$$

A partir de esta distribución, se genera igual número de muestras que en los casos anteriores y se queman las primeras 1000 observaciones.

Las autocorrelaciones para los valores generados de β con diferentes rezagos se muestran en la tabla 3.

TABLA 3: Autocorrelaciones de los β con diferentes rezagos.

	Autocorrelación de β con			
	1 rezago	5 rezagos	10 rezagos	50 rezagos
β	0.03476958	−0.01109447	−0.03358050	0.001319354

En la tabla 3, se observa que no hay una fuerte asociación de los parámetros generados por la cadena de Markov con distintos rezagos.

El estadístico KPSS es 0.0805, el parámetro de truncamiento es 6 y el valor p es 0.1. Por tanto, hay evidencia de que ya se ha alcanzado la distribución límite para β.

De igual manera que en el caso anterior, de los promedios móviles y la densidad de β, se obtuvieron las gráficas, las cuales se muestran en la figura 4.

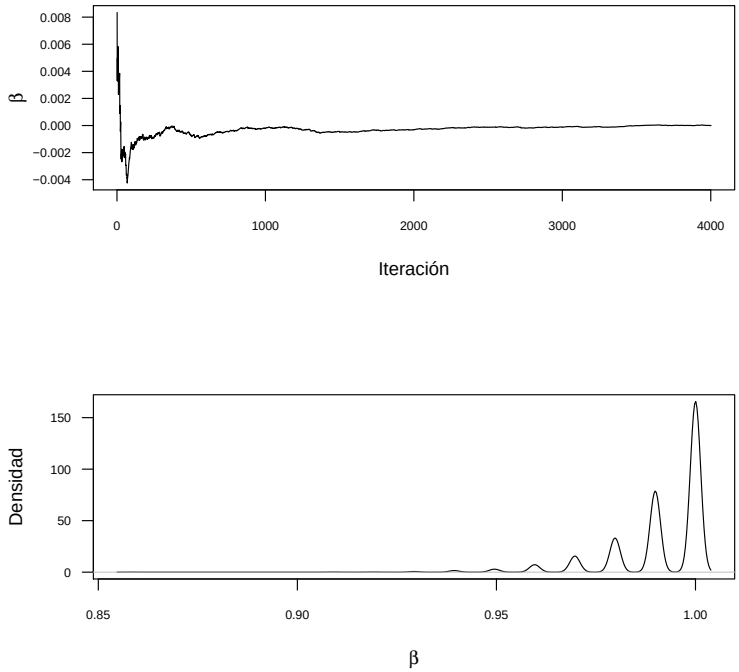

FIGURA 4: Promedios móviles y densidad para β, respectivamente.

Los gráficos y las pruebas teóricas permiten afirmar que la distribución para β es estacionaria, es decir, está muestreándose de la distribución límite para β.

5. Conclusiones

Es posible hallar la distribución predictiva bayesiana a partir de a prioris no conjugadas.

Se ha propuesto una metodología basada en el algoritmo MCMC, que permite calcular la distribución predictiva en situaciones en que esta distribución no puede hallarse en forma analítica y no sea fácilmente tratable con los métodos de simulación tradicionales. Estos casos pueden presentarse cuando se trabaja con distribuciones no conjugadas o con problemas de alta densidad. Esta metodología tiene la ventaja de ser implementable con facilidad, pero hay que tener cuidado con los problemas de convergencia, en los cuales hay que recurrir a diversos diagnósticos (Kwiatkowski et al. 1992).

En este caso, se utiliza la prueba KPSS para estacionariedad, la cual no aparece en la literatura utilizada con este propósito.

[Recibido: febrero de 2008 — Aceptado: septiembre de 2008]

Referencias

Casella, G. & George (1992), 'Explaining the Gibbs Sampler', *The American Statistician* **46**(3), 167–174.

Christensen, R. & Huffman, M. (1985), 'Bayesian Point Estimation Using the Predictive Distribution', *The American Statistician* **39**(4), 319–321.

Dunsmore, I. (1974), 'The Bayesian Predictive Distribution in Life Testing Models', *Technometrics* **16**(3), 455–460.

Hewett, J. (1968), 'A Note on Prediction Intervals Based on Tartial Observations in Certain Life Test Experiments', *Technometrics* **10**, 850–853.

Hill, G. (2002), *Bayesian Methods*, Chapman and Hall.

Kalbfleisch, J. D. (1971), Likelihood Methods of Prediction, *in* V. P. Godambe & D. A. Sprott, eds, 'Foundations of Statistical Inference', pp. 378–392.

Kao, E. (1997), *An Introduction to Stochastic Processes*, Duxbury Press.

Komaki, F. (2001), 'A Shrinkage Predictive Distribution for Multivariate Normal Observables', *Biometrika* **88**(3), 859–864.

Kwiatkowski, D., Phillips, P. & Schmidt, P. (1992), 'Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root', *Journal of Econometrics* **54**, 159–178.

Meeker, W. & Escobar, L. (1998), *Statistical Methods For Reliability Data*, Wiley Interscience.

R Development Core Team (2007), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
*http://www.R-project.org