

UNA FAMILIA DE DISTRIBUCIONES CONJUGADAS PARA UN PROCESO ARE (1)*

E. CARO, J. I. DOMÍNGUEZ y F. J. GIRÓN
Dpto. de Matemáticas
(U.D. Estadística E. I.O.)
Facultad de Ciencias
Universidad de Málaga

RESUMEN

En este artículo se estudia, desde una perspectiva bayesiana, un proceso AR(1) con errores exponenciales, ARE(1); para ello se construye una nueva familia de distribuciones conjugadas, denotada por CDG, que permite construir una especie de filtro de Kalman para la estimación recursiva de los parámetros del modelo.

Palabras clave: Distribuciones conjugadas; Errores exponenciales; Proceso autorregresivo de primer orden.

Clasificación AMS (1980): 62M10; 62M20; 62F15.

Recibido: Octubre 1990.
Revisado: Enero 1991.

ABSTRACT

In this article estimation of autoregressive processes AR(1) with exponential errors, denoted ARE(1), is considered from a Bayesian perspective. For these processes a new family of conjugate distributions, CDG, is shown to exist which allows for recursive estimation in a Kalman filter-like fashion.

Key words: Conjugate distributions; Exponential errors; First order autoregressive processes.

AMS subject classification: 62M10; 62M20; 62F15.

Title: A conjugate family for an ARE(1) process*.

* Este artículo ha sido parcialmente subvencionado con la ayuda de la *Dirección General de Investigación Científica y Técnica (DGICYT)* como parte del proyecto de referencia PB-87-0607-C02-02, por la *Consejería de Educación de la Junta de Andalucía* como parte del proyecto acogido al *Plan Andaluz para la Consolidación de Grupos de Investigación y Desarrollo Tecnológico* y por la *Universidad de Málaga* como parte de un proyecto acogido al *Plan de Acciones Concertadas (Grupos Precompetitivos)*.

1. INTRODUCCION

Sea $\{Y_t: t = 1, 2, \dots\}$ un proceso de tipo AR(1), es decir $Y_t = \theta Y_{t-1} + \varepsilon_t$ donde el parámetro autorregresivo es desconocido y los errores son v.a. independientes e igualmente distribuidas. Nuestro objetivo será estimar los parámetros (autorregresivo y los asociados a la distribución de los errores) así como encontrar la distribución predictiva del proceso.

En la literatura existente sobre este tema, se suele plantear el problema bajo la condición de que los errores sean normales, casi siempre de media conocida y desviación desconocida, lo que permite utilizar una distribución a priori del tipo Normal-Gamma y hacer un análisis bayesiano del problema. Véase, por ejemplo, Zellner (1971) y Broemeling (1985).

Recientemente, Domínguez (1990), ante las posibles limitaciones que se presentan al estudiar este tipo de problemas bajo dicho punto de vista, como sucede en Broemeling (loc. cit.) al proponer la «estacionariedad aproximada», y con idea de introducir modelos con errores no normales (Damsleth & El-Shaarawi, 1989), los introduce y estudia directamente para el caso en que los errores sean del tipo exponencial de parámetro desconocido. Precisamente, al observar los resultados obtenidos en la citada referencia, en donde el modelo se desarrolla para el caso de estacionariedad con dependencia autorregresiva positiva, se ve como este modelo puede generalizarse considerando para la verosimilitud una familia de distribuciones conjugadas, que llamaremos distribuciones del tipo CDG, y que surgen allí de una forma natural. De igual modo a como se propone en Domínguez (loc. cit.), el modelo puede generalizarse al caso de variables negativas e incluso al caso de que el proceso sea «explosivo».

En el apartado dos se justifica, introduce y estudia la familia de distribuciones del tipo CDG. En el tres planteamos, desde una perspectiva bayesiana, el modelo objeto de estudio y obtenemos el filtro adecuado. En el último apartado se realiza una simulación para contrastar los resultados obtenidos, así como algunos comentarios sobre generalizaciones, derivaciones y aplicaciones del modelo propuesto.

2. DISTRIBUCION CDG

Consideremos el modelo autorregresivo original de forma que los errores se distribuyen según una distribución exponencial de parámetro λ desconocido.

Supongamos que hemos realizado n observaciones $\mathbf{y}_n = (y_1, \dots, y_n)$, y sea y_0 un valor conocido inicial. Entonces la verosimilitud asociada (Domínguez, loc. cit.) será:

$$l(\theta, \lambda | y_0, \mathbf{y}_n) = \prod_{t=1}^n \lambda \exp \{-\lambda(y_t - \theta y_{t-1})\} = \lambda^n \exp \{-\lambda(S_n - \theta R_n)\} \quad (2.1)$$

donde $y_t \in R$, $S_n = \sum_{1 \leq t \leq n} y_t$, $R_n = S_{n-1} + y_0$ y además $y_t - \theta y_{t-1} \geq 0 \forall t$. Esta forma funcional de la verosimilitud nos lleva a considerar una familia conjugada de distribuciones para (θ, λ) del tipo:

$$f(\theta, \lambda | p, a, b, \theta_1, \theta_2) \propto \lambda^{p-1} \exp \{-\lambda(a - b\theta)\} I[\theta_1, \theta_2](\theta) \quad (2.2)$$

y dar la siguiente

Definición 2.1

Dado un vector bidimensional (X, Y) absolutamente continuo, diremos que sigue una distribución CDG de parámetros p, a, b, θ_1 y θ_2 , y la denotaremos por CDG $(x, y | p, a, b, \theta_1, \theta_2)$ si su función de densidad es de la forma:

$$f(x, y) = \begin{cases} k(p, a, b, \theta_1, \theta_2) x^{p-1} e^{-(a-by)x} & (x > 0, \theta_1 \leq y \leq \theta_2) \\ 0 & \text{resto} \end{cases} \quad (2.3)$$

donde $\theta_1 < \theta_2 \in R$, $p > 1$, $a > b\theta_2$ y $k(p, a, b, \theta_1, \theta_2)$ es una constante positiva cuyo valor se determina bajo la condición de que $\int \int_{S^*} f(x, y) dx dy = 1$, siendo S^* el soporte del vector con densidad dada en (2.3), es decir el rectángulo $R^+ \times [\theta_1, \theta_2]$. En consecuencia:

$$k(p, a, b, \theta_1, \theta_2) = \frac{b}{\Gamma(p-1)} \frac{1}{(a - b\theta_2)^{-p+1} - (a - b\theta_1)^{-p+1}} \quad (2.4)$$

En lo que sigue, se dan los principales resultados de interés sobre la familia CDG; en concreto, las distribuciones marginales, las condicionadas, los valores esperados, moda y las varianzas y covarianza. Estos

resultados se utilizarán en los próximos apartados. Para abreviar la notación, y mientras no exista confusión en los valores de los parámetros, denotaremos $k(p) = k(p, a, b, \theta_1, \theta_2)$. Entonces:

— Distribuciones marginales:

A partir de la distribución bidimensional CDG con densidad (2.3), es fácil obtener

$f_Y(y) = k(p)\Gamma(p)(a - by)^{-p}I_{[\theta_1, \theta_2]}(y)$ cuyo valor esperado y varianza son, respectivamente:

$$E(Y) = \frac{a}{b} - \frac{p-1}{b} \frac{k(p)}{k(p-1)}$$

$$\text{Var}(Y) = \frac{(p-1)(p-2)}{b^2} \frac{k(p)}{k(p-2)} - \frac{(p-1)^2}{b^2} \frac{k^2(p)}{k^2(p-1)}$$

De igual modo

$$\begin{aligned} f_X(x) &= k(p)b^{-1}x^{p-2} [\exp\{-x(a - b\theta_2)\} - \exp\{-x(a - b\theta_1)\}] = \\ &= \mu(p)G(p-1, a - b\theta_2) + (1 - \mu(p))G(p-1, a - b\theta_1) \end{aligned}$$

que es una «falsa mixtura» de distribuciones gammas (los coeficientes suman uno, pero puede ser negativo uno de ellos) con pesos $\mu(p)$ y $(1 - \mu(p))$ tales que $\mu(p) = k(p)b^{-1}\Gamma(p-1)(a - b\theta_2)^{-p+1}$. En este caso, el valor esperado y varianza son, respectivamente:

$$E(X) = \frac{k(p)}{k(p+1)} \quad \text{Var}(X) = \frac{k(p)}{k(p+2)} - \frac{k^2(p)}{k^2(p+1)}$$

Por último, a partir de la densidad (2.3), la moda de la distribución CDG es

$$\hat{x} = \frac{p-1}{a - b\hat{y}} \quad \hat{y} = \begin{cases} \theta_1 & \text{si } b < 0 \\ \theta_2 & \text{si } b > 0 \end{cases}$$

— Distribuciones condicionadas:

$f(x|y) = G(p, a - by)$, y valor esperado $E(X|y = y) = p/(a - by)$

Así mismo

$$f(y|x) = bx [\exp(bxy)] / [\exp(bx\theta_2) - \exp(bx\theta_1)]$$

y valor esperado:

$$E(Y|X=x) = \frac{\theta_2 \exp(bx\theta_2) - \theta_1 \exp(bx\theta_1)}{\exp(bx\theta_2) - \exp(bx\theta_1)} - \frac{1}{bx}$$

Por último, es fácil ver que la covarianza viene dada por:

$$\text{Cov}(X, Y) = \left[\frac{a}{b} - E(Y) \right] E(X) - \frac{p}{b}$$

A la vista de la forma funcional que tienen las distribuciones marginales y condicionadas de la distribución CDG, ésta puede ser considerada como un tipo de distribución «Gamma-Pareto truncada» ya que la distribución de $X|Y$ es Gamma mientras que la ley de Y recuerda un tipo de distribución de Pareto truncada con la salvedad de que puede estar definida para valores negativos.

Merece la pena señalar que para la construcción de la familia conjugada CDG, hemos utilizado la definición «clásica» (véase, por ejemplo, De Groot (1970)) de las mismas basada en la existencia de estadísticos suficientes de dimensión fija. La caracterización de familias conjugadas dada por Diaconis and Ylvisaker (1979) no es directamente aplicable a procesos dependientes ya que la esperanza a posteriori de los parámetros no es lineal en las observaciones, por lo que sería interesante estudiar el problema de caracterizar la clase de procesos AR(1) para los cuales existe un estadístico suficiente de dimensión fija. El proceso AR(1) con errores normales i.i.d. y el ARE(1) pertenecen a esta clase, sin embargo si los errores del AR(1) son de tipo gamma (no exponencial) entonces ya no existe estadístico suficiente.

3. ANALISIS BAYESIANO DEL MODELO

El resultado que a continuación se demuestra, justifica la elección de la familia CDG como una familia de distribuciones conjugadas para el proceso ARE(1) que nos permite, además, construir una especie de filtro de Kalman para la estimación recursiva de los parámetros del modelo a medida que se disponga de nuevas observaciones del proceso.

Teorema 3.1

La familia de distribuciones CDG definida por (2.3) es una familia conjugada para el proceso ARE (1).

Demostración:

Sea $\mathbf{y}_n = (y_1, \dots, y_n)$ una realización del proceso ARE(1), con condición inicial y_0 . La función de verosimilitud viene dada por (2.1), es decir

$$l(\lambda, \theta | y_0, \mathbf{y}_n) \propto \lambda^n \exp \{ -\lambda(S_n - \theta R_n) \} I_{[\theta_1^{(n)}, \theta_2^{(n)}]}(\theta) \quad (3.1)$$

donde

$$\theta_1^{(n)} = \max_{y_{n-1} < 0} \left\{ -\infty, \frac{y_n}{y_{n-1}} \right\} ; \quad \theta_2^{(n)} = \min_{y_{n-1} > 0} \left\{ +\infty, \frac{y_n}{y_{n-1}} \right\}$$

Si, a priori $(\lambda, \theta) \approx \text{CDG}(p_0, a_0, b_0, \theta_1^{(0)}, \theta_2^{(0)})$, por el teorema de Bayes se tiene que la densidad conjunta

$$f(\lambda, \theta | y_0, \mathbf{y}_n) \propto \lambda^{n+p_0-1} \exp \{ -\lambda[S_n + a_0 - (R_n + b_0)\theta] \} I_{[\max\{\theta_1^{(0)}, \theta_1^{(n)}\}, \min\{\theta_2^{(0)}, \theta_2^{(n)}\}]}(\theta)$$

de donde resulta que para todo n

$$\lambda, \theta | y_0, \mathbf{y}_n \approx$$

$$\approx \text{CDG} \left(p_0 + n, a_0 + S_n, b_0 + R_n, \max_{y_{n-1} < 0} \left\{ \theta_1^{(0)}, \frac{y_n}{y_{n-1}} \right\}, \min_{y_{n-1} > 0} \left\{ \theta_2^{(0)}, \frac{y_n}{y_{n-1}} \right\} \right)$$

lo que demuestra el teorema.

Si ahora aplicamos la versión secuencial del teorema de Bayes, teniendo en cuenta el resultado anterior resulta la siguiente recurrencia:

Si la distribución a posteriori de (λ, θ) dado (y_0, \mathbf{y}_{n-1}) es del tipo CDG $(p_{n-1}, a_{n-1}, b_{n-1}, \theta_1^{(n-1)}, \theta_2^{(n-1)})$, entonces la distribución a posteriori de (λ, θ) dado $(y_0, \mathbf{y}_{n-1}, y_n) \equiv (y_0, \mathbf{y}_n)$ es también una distribución del tipo CDG $(p_n, a_n, b_n, \theta_1^{(n)}, \theta_2^{(n)})$ donde los parámetros vienen dados por

$$p_n = p_{n-1} + 1 \quad a_n = a_{n-1} + y_n \quad b_n = b_{n-1} + y_{n-1}$$

$$\theta_1^{(n)} = \begin{cases} \text{máx} \left\{ \theta_1^{(n-1)}, \frac{y_n}{y_{n-1}} \right\} & \text{si } y_{n-1} < 0 \\ \theta_1^{(n-1)} & \text{si } y_{n-1} > 0 \end{cases} \quad (3.2)$$

$$\theta_2^{(n)} = \begin{cases} \text{mín} \left\{ \theta_2^{(n-1)}, \frac{y_n}{y_{n-1}} \right\} & \text{si } y_{n-1} > 0 \\ \theta_2^{(n-1)} & \text{si } y_{n-1} < 0 \end{cases}$$

Es importante reseñar que la consideración de la familia CDG permite incorporar automáticamente la hipótesis de estacionariedad del proceso.

El problema de hacer predicciones sobre observaciones futuras se resuelve a través de la función predictiva. Para el caso de la distribución conjunta de varias observaciones futuras la distribución multivariante predictiva no se puede calcular de manera analítica sencilla. Sin embargo la distribución predictiva de una observación futura $y_{n+1} | y_0, \mathbf{y}_n$ se calcula fácilmente, aunque laboriosamente, a partir de:

$$\begin{aligned} f(y_{n+1} | y_0, \mathbf{y}_n) &= \int \int f(y_{n+1} | y_0, \mathbf{y}_n, \lambda, \theta) f(\lambda, \theta | y_0, \mathbf{y}_n) d\lambda d\theta = \\ &= \int \int k(p_n, a_n, b_n, \theta_1^{(n)}, \theta_2^{(n)}) \lambda p_n^{-1} \\ &\quad e^{-(a_n - b_n \theta)} \lambda e^{-\lambda(y_{n+1} - \theta y_n)} d\lambda d\theta = \\ &= \begin{cases} \frac{k(p_n, a_n, b_n, \theta_1^{(n)}, \theta_2^{(n)})}{k\left(p_{n+1}, a_n + y_{n+1}, b_n + y_n, \theta_1^{(n)}, \text{mín} \left\{ \theta_2^{(n)}, \frac{y_{n+1}}{y_n} \right\}\right)} & \text{si } y_n > 0 \\ \frac{k(p_n, a_n, b_n, \theta_1^{(n)}, \theta_2^{(n)})}{k\left(p_{n+1}, a_n + y_{n+1}, b_n + y_n, \text{máx} \left\{ \theta_1^{(n)}, \frac{y_{n+1}}{y_n} \right\}, \theta_2^{(n)}\right)} & \text{si } y_n < 0 \end{cases} \quad (3.3) \end{aligned}$$

y de la cual pueden obtenerse, por ejemplo, intervalos de predicción de máxima densidad a posteriori (M.D.P.).

La media de la distribución predictiva puede obtenerse directamente de (3.3), por integración numérica. Sin embargo, a partir del modelo

ARE (1), se puede dar una fórmula explícita de la esperanza de la distribución predictiva, como sigue:

$$\begin{aligned}
 E(Y_{n+1} | y_0, \mathbf{y}_n) &= E(\theta Y_n + \varepsilon_{n+1} | y_0, \mathbf{y}_n) = y_n E(\theta | y_0, \mathbf{y}_n) + E(\varepsilon_{n+1} | y_0, \mathbf{y}_n) = \\
 &= y_n E(\theta | y_0, \mathbf{y}_n) + E_\lambda(E(\varepsilon_{n+1} | \lambda, y_0, \mathbf{y}_n)) = \\
 &= y_n E(\theta | y_0^0, \mathbf{y}_n) + E_\lambda(\lambda^{-1} | y_0, \mathbf{y}_n) = \\
 &= \left(y_n - \frac{b_n}{p_n - 1} \right) E(\theta | y_0, \mathbf{y}_n) + \frac{a_n}{p_n - 1} = \\
 &= \frac{a_n}{b_n} y_n + (b_n - p_{n-1} y_n) \frac{k(p_n)}{p_n k(p_n - 1)} \tag{3.4}
 \end{aligned}$$

Por último, para obtener la distribución no informativa de ambos parámetros, y debido a la dependencia del soporte de la verosimilitud de la muestra, la regla de Jeffreys no es directamente aplicable, por lo que si hacemos la hipótesis de independencia a priori de ambos parámetros, la distribución no informativa de λ es la usual, es decir $p_0(\lambda) \propto \lambda^{-1}$, mientras que para el parámetro autorregresivo θ (sin hipótesis de estacionariedad) tomamos como distribución no informativa una localmente uniforme, es decir, $p_0(\theta) \propto \text{cte}$. De aquí resulta, con las hipótesis anteriores, que la distribución no informativa impropia sería $p_0(\lambda, \theta) \propto \lambda^{-1}$, que es un caso límite de la familia CDG cuando $a \rightarrow 0$, $b \rightarrow 0$, $p \rightarrow 0$, $\theta_1^{(0)} \rightarrow -\infty$, $\theta_2^{(0)} \rightarrow +\infty$.

Para esta distribución de referencia a priori, la distribución a posteriori de referencia (que sería propia para tamaños muestrales $n \geq 2$) vendría dada por una del tipo

$$\text{CDG} \left(n, S_n, R_n, \max_{y_{n-1} < 0} \left\{ -\infty, \frac{y_n}{y_{n-1}} \right\}, \min_{y_{n-1} > 0} \left\{ +\infty, \frac{y_n}{y_{n-1}} \right\} \right)$$

4. COMENTARIOS, DISCUSION Y RESULTADOS DE LA SIMULACION

El modelo considerado ARE(1) es susceptible de ser generalizado a procesos autorregresivos de orden p , ARE(p). Su tratamiento estadístico presenta, sin embargo, problemas debido a la no existencia de estadísticos suficientes de dimensión fija, a diferencia del ARE(1).

Por otra parte el modelo ARE(1) es muy sensible a la falta de especificación del modelo, como hemos comprobado en los experimentos de simulación. Así, por ejemplo, el modelo es poco robusto ante la presencia de observaciones anómalas y/o desviaciones de los errores del modelo exponencial, lo cual nos conduce a extender nuestro estudio al caso de errores generados por un mecanismo más complejo como puede ser, por ejemplo, una mezcla de distribuciones exponenciales que, en parte, obvia los inconvenientes que acabamos de señalar.

Los resultados de la simulación demuestran también que cuando el modelo es correcto, la estimación bayesiana de los parámetros del mismo es muy eficiente, en particular la estimación del parámetro autorregresivo. De igual forma el análisis de los residuos nos proporciona un test de carácter informal, sobre la validez modelo, como se ilustra en el ejemplo de simulación que analizamos a continuación.

— La figura 1 representa una serie de 250 observaciones simuladas de un proceso ARE(1) con $\lambda = .5$, $\theta = .5$ y con valor inicial $y_0 = 0$.

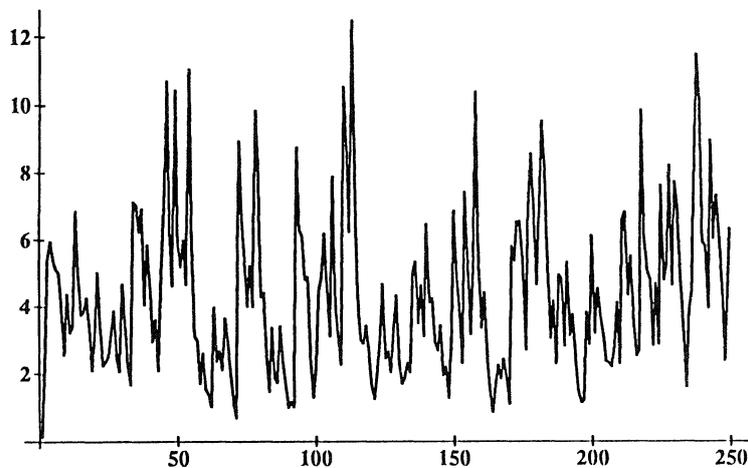


Figura 1

— La figura 2 representa la estimación secuencial de la media del parámetro λ del error exponencial que, como se observa, sigue el patrón habitual de la estimación bayesiana de una muestra de una exponencial.

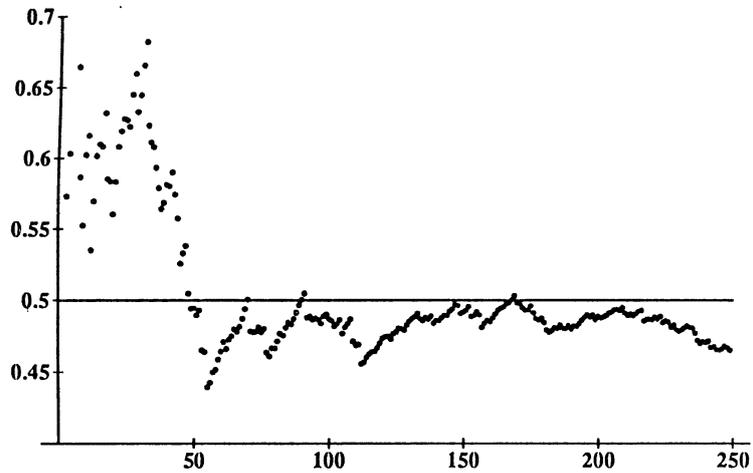


Figura 2

— Las figuras 3a y 3b representan la estimación secuencial de la media del parámetro autorregresivo y la estimación por el método de máxima verosimilitud para dicho parámetro, respectivamente. Es de reseñar el paralelismo existente entre ambas estimaciones, aunque el caso de máxima verosimilitud (3b) sobreestima el parámetro autorregresivo por la propia definición del proceso mientras que la estimación bayesiana (3a) es más eficiente.

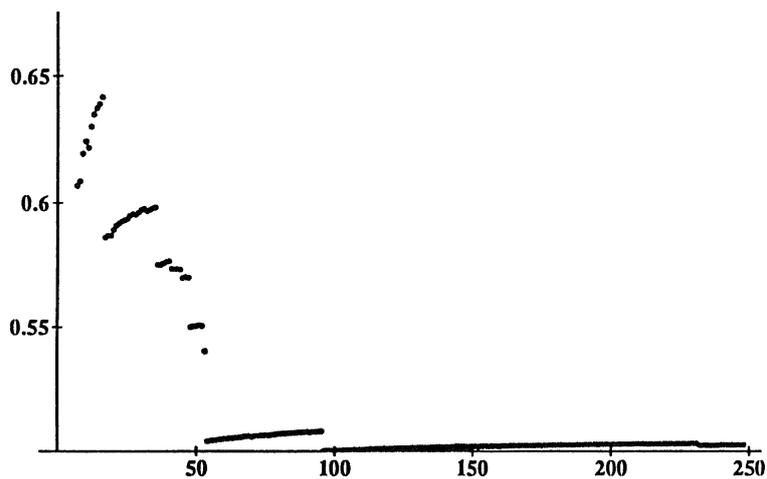


Figura 3a

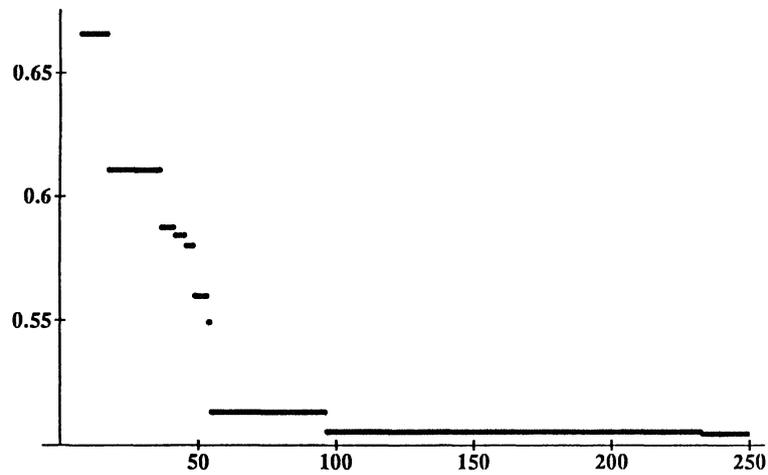


Figura 3b

— Las figuras 4a y 4b representan los residuos «ordinarios» y secuenciales definidos, respectivamente, por

$$o_{t+1} = y_{t+1} - \theta_n^* y_t \quad \text{y} \quad r_{t+1} = y_{t+1} - \theta_t^* y_t$$

donde θ_t^* es la media a posteriori de θ dado $D_t = \{y_0, \mathbf{y}_t\}$, es decir $\theta_t^* = E(\theta | D_t)$ ($2 \leq t \leq n$).

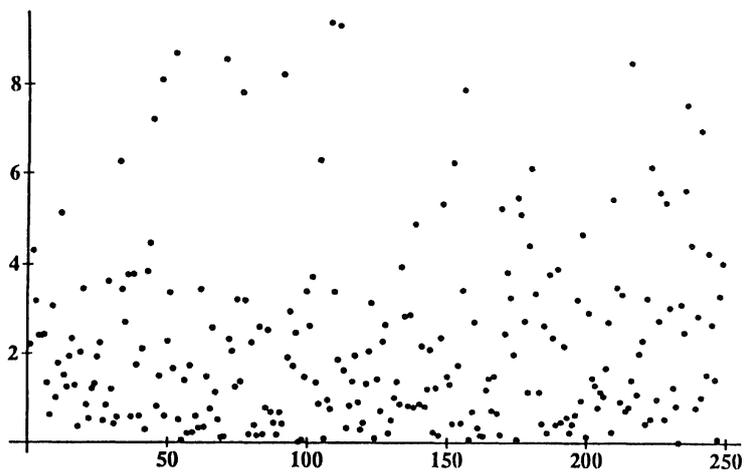


Figura 4a

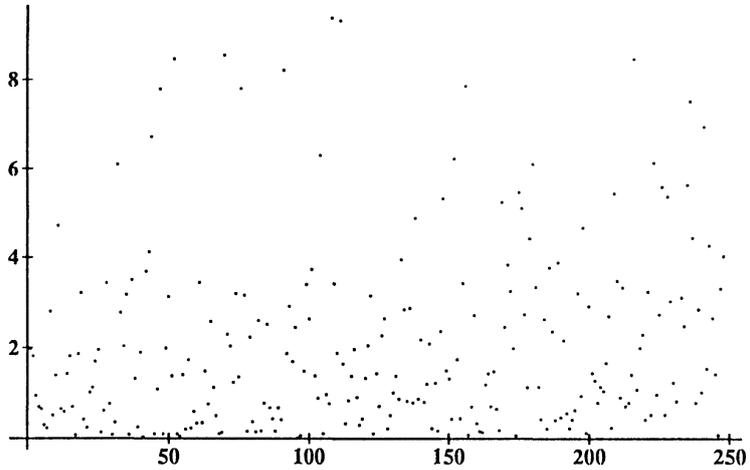


Figura 4b

Debido a la rápida convergencia de la distribución a posteriori de θ , puede observarse que a partir de un tamaño de serie moderadamente pequeño no existe diferencia entre ambos tipos de residuos.

— Las figuras 5 y 6 muestran la densidad a posteriori de los parámetros λ y θ , respectivamente, condicionada a los instantes de tiempo $t = 10, 50$ y 250 representadas, respectivamente, por una línea a trazos, una de puntos y una continua. Ellas reflejan el aumento de

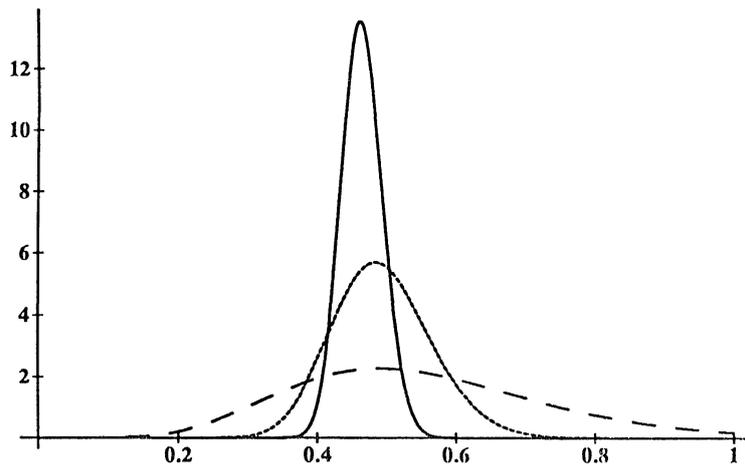


Figura 5

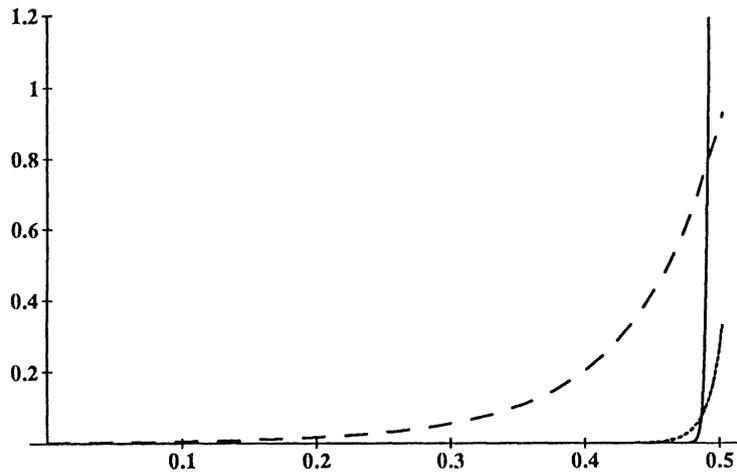


Figura 6

precisión en la estimación de ambos parámetros a medida que se condicionan por un número mayor de datos de la serie. Es de reseñar la gran rapidez de convergencia de la distribución del parámetro autorregresivo.

— La figura 7 representa la densidad predictiva de una observación futura de la serie, la cual también converge hacia una especie de exponencial trasladada. La media de la distribución predictiva es 5.33 y el intervalo de predicción M.D.P. al 95 % es (3.18; 9.65).

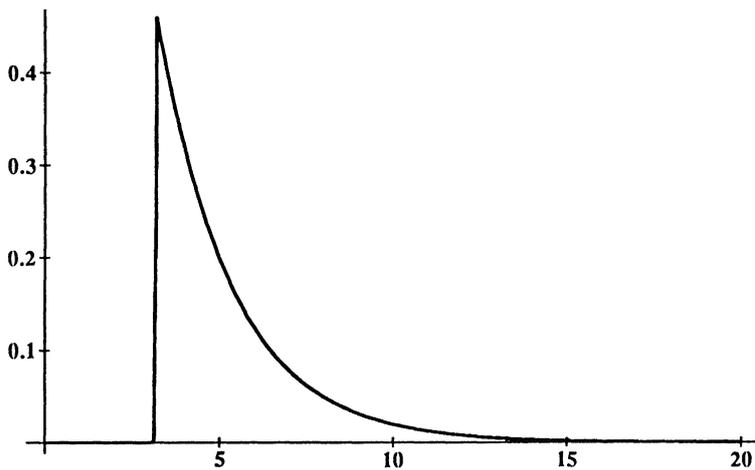


Figura 7

REFERENCIAS

- BOX, G. E. P., y JENKINS, G. M. (1976): *Time series Analysis, Forecasting and Control*, Holden-Day, San Francisco.
- BROEMELING, L. D. (1985): *Bayesian Analysis of Linear Models*, Marcel Dekker, Nueva York.
- DE GROOT, M. H. (1970): *Optimal Statistical Decisions*, McGraw-Hill, Nueva York.
- DIACONIS, P., y YLVISAKER, D. (1979): «Conjugate priors for exponential families», *Ann. Statist.*, **7**, 267-281.
- DOMINGUEZ, J. I. (1990): «Análisis Bayesiano de un Modelo AR(1) con Errores Exponenciales», *Estadística Española*, vol. 32, 124, 401-416.
- DAMSLETH, E., y EL-SHAARAWI, A. H. (1989): ARMA Models with Double-Exponentially Distributed Noise», *J. R. Statist. Soc. B*, **51**, 61-69.
- PEÑA, E. A., y GRUPTA, A. K. (1990): Bayes Estimation for the Marshall-Olkin Exponential Distribution», *J. R. Statist Soc. B*, **52**, 379-389.
- ZELLNER, A. (1971): *An introduction to Bayesian Inference in Econometrics*, J. Wiley.