

ANÁLISIS BAYESIANO DEL MODELO ARE(1) CON COEFICIENTE INDEPENDIENTE

(análisis bayesiano / distribuciones conjugadas / modelos autorregresivos de primer orden)

FCO. DEMETRIO LÓPEZ ROMERO

Departamento de Estadística e Investigación Operativa. Universidad de Málaga. Málaga
e-mail: demetrio@uma.es

Presentado por Fco. Javier Girón, 9 de abril de 1997.

RESUMEN

En este trabajo se introduce el modelo ARE(1) con indicador de nivel mínimo μ , parámetro que generaliza el modelo AR(1) con errores exponenciales y se analiza desde un punto de vista bayesiano, obteniéndose una familia de distribuciones conjugadas para el hiperparámetro que describe el modelo.

ABSTRACT

In this paper the model ARE(1) with *minimum level* μ is introduced, which generalizes the AR(1) model with exponential errors. The model is analysed from a Bayesian perspective and finally a conjugate family of distributions for the parameters of the model is obtained.

INTRODUCCIÓN

En numerosas situaciones experimentales aparecen series temporales observadas que toman sólo valores no negativos. En estos casos la hipótesis de normalidad puede llevar a modelos no adecuados ya que permitirían la aparición de datos negativos (aunque fuesen con probabilidad pequeña); por lo tanto, parece necesario plantear modelos de series temporales con distribuciones asociadas no gaussianas, en particular distribuciones con soporte no negativo.

En esta línea cabe destacar el artículo de Gaver & Lewis (1), donde se introducen los primeros modelos para series temporales no negativas, en este caso con estructura autorregresiva de primer orden y con distribución marginal exponencial; es decir, el modelo vendría definido por un proceso AR(1) de la forma

$$X_t = \theta X_{t-1} + \varepsilon_t, \quad (t = 0, \pm 1, \dots),$$

donde $0 \leq \theta < 1$ y la sucesión de errores $\{\varepsilon_t\}$ es una sucesión de variables aleatorias independientes y equidistribuidas con distribución tal que el proceso $\{X_t\}$ posee una distribución marginal exponencial. Estos procesos son llamados por los autores modelos EAR(1).

Sin embargo, para conseguir que la distribución de X_t sea exponencial los autores prueban, a través de la función característica, que la distribución de los errores debe ser una mezcla de una exponencial de parámetro λ y una distribución degenerada en cero, con pesos $1 - \theta$ y θ respectivamente, de tal forma que estas variables no son absolutamente continuas en el cero, lo que provoca rachas de ceros que pueden ser de difícil justificación en series de datos generados por una variable continua. Esta anomalía es lo que los autores denominan *cero-deficiencia*.

Lawrence & Lewis (2) propone una alternativa para evitar los problemas en el cero antes aludidos. En este caso el modelo, denominado NEAR(1), se complica con un nuevo parámetro y ya no sigue la estructura estándar de los procesos autorregresivos clásicos. El modelo propuesto es de la forma

$$X_t = \varepsilon_t + \begin{cases} 0 & \text{con probabilidad } p, \\ \theta X_{t-1} & \text{con probabilidad } 1-p \end{cases} \quad (t = 1, 2, \dots),$$

donde $0 \leq p \leq 1$, $0 < \theta < 1$ y $\{\varepsilon_t\}$ una sucesión de variables aleatorias e independientes definida por

$$\varepsilon_t = \begin{cases} E_t & \text{con probabilidad } \frac{1-\theta}{1-\theta p}, \\ \theta p E_t & \text{con probabilidad } \frac{\theta(1-p)}{1-\theta p} \end{cases} \quad (t = 1, 2, \dots),$$

donde $\{E_t\}$ es una sucesión de exponenciales independientes de parámetro λ , con lo que se consigue, como en el caso

anterior, que la distribución marginal del proceso sea exponencial de parámetro λ .

En resumen, los modelos descritos brevemente han sido desarrollados con el fin de conseguir que las variables del proceso tengan una distribución marginal dada y una estructura correlacional autorregresiva o de media móvil.

Otra vía para ajustar modelos autorregresivos a series no negativas o, en general, no normales es considerar procesos en los que los errores o ruidos constituyan una sucesión de variables aleatorias independientes y equidistribuidas con una distribución dada. Siguiendo esta idea Damsleth & El-Shaaravi (3) introducen los modelos ARMA con errores de tipo Laplace y estudian la distribución marginal y estacionaria del proceso, mientras que Domínguez (4) estudia el caso exponencial en el modelo AR(1) desde un punto de vista bayesiano.

Otros autores también han realizado aportaciones en torno a este tema. Por ejemplo, Bell & Smith (5) realizan inferencias sobre los parámetros de un modelo AR(1) con errores positivos y un nivel inicial X_0 dado, estudiando de manera específica el caso exponencial y el uniforme sobre $[0, a]$, para $a > 0$. Andêl (6) modifica el modelo exponencial de Bell & Smith (5) y considera X_0 como una variable inicial exponencial con la misma media que la distribución estacionaria del proceso. Davis & McCormick (7) estudian las distribuciones límite de los estimadores asociados a los modelos de este tipo, para ello imponen a las distribuciones de los ruidos ciertas condiciones de regularidad, en concreto exigen que sean *regularmente variantes* en cero con exponente α ($\alpha > 0$), para las distribuciones de ruidos positivos; es decir, debe existir $\alpha > 0$ tal que $\lim_{t \rightarrow 0} F(tx)/F(t) = x^\alpha$ ($x > 0$), donde F es la distribución de los errores; en particular, la distribución exponencial es regularmente variante.

En consecuencia, parece consolidada la necesidad de construir modelos para series temporales con distribuciones, bien la marginal del proceso o bien la de los ruidos, no normales y en particular de tipo exponencial, y así lo han considerado numerosos investigadores; sin embargo, todos los trabajos referidos, a excepción del artículo de Domínguez (4), plantean el análisis de estos modelos desde un punto de vista clásico, por lo que creemos de interés realizar un análisis preciso y completo del modelo con errores exponenciales bajo una perspectiva bayesiana. En esta línea es también interesante reseñar el trabajo de Amaral (8) sobre los modelos propuestos por Andêl (6).

2. DEFINICIÓN DEL MODELO ARE(1).

Consideremos el proceso autorregresivo $\{X_t\}$ satisfaciendo las ecuaciones

$$(2.1)$$

$$X_t = c + \theta X_{t-1} + \varepsilon_t \quad (t = 1, 2, \dots),$$

donde $c \in \mathbb{R}$, $0 \leq \theta < 1$ y $\{\varepsilon_t\}$ es una sucesión de variables aleatorias independientes y equidistribuidas con distribución exponencial de parámetro $\lambda > 0$, y donde X_0 es una variable aleatoria que determina el nivel inicial de la población e independiente de la sucesión de errores. A estos modelos los denotaremos por ARE(1).

Como bien se sabe esta clase de modelos se pueden expresar como suma de errores, de la forma

$$X_t = \mu + \theta^t (X_0 - \mu) + \sum_{j=0}^{t-1} \theta^j \varepsilon_{t-j} \quad (t = 1, 2, \dots), \quad (2.2)$$

donde $\mu \in \mathbb{R}$ y, como comprobaremos más adelante, $\mu = c/(1 - \theta)$.

Para modelos estacionarios en los que la función de valor medio de la sucesión de errores es cero, el parámetro μ es la media del proceso, como ocurre en el modelo clásico definido por Box & Jenkins (9), en el cual los errores son *ruidos blancos*, esto es, variables aleatorias independientes con distribución normal de media cero.

Si el modelo no es estacionario μ no tiene ningún significado especial, excepto el de un punto de referencia para el nivel del proceso.

En nuestro caso la media del proceso se puede obtener directamente de (2.2), en concreto se tiene

$$E[X_t] = \mu + \theta^t (E[X_0] - \mu) + \frac{1 - \theta^t}{\lambda(1 - \theta)}.$$

Nótese que, en realidad, la estacionaridad se alcanza para t grande o bien cuando la distribución de X_0 es la estacionaria, en cuyo caso la media de dicha distribución viene dada por

$$E[X_t] = \mu + \frac{1}{\lambda(1 - \theta)};$$

obsérvese que μ no es la media del proceso pero si es indicador de un *nivel mínimo* del mismo

Si tomamos esperanzas en (2.1), bajo la hipótesis de estacionaridad del modelo, se llega a

$$\mu + \frac{1}{\lambda(1 - \theta)} = c + \theta \left[\mu + \frac{1}{\lambda(1 - \theta)} \right] + \frac{1}{\lambda}$$

de donde se sigue $\mu = c/(1 - \theta)$.

Si ahora consideramos las desviaciones de X_t respecto de μ , que denotaremos por \tilde{X}_t ; es decir, $\tilde{X}_t = X_t - \mu$ ($t = 0, 1, \dots$), podemos expresar el modelo ARE(1) en la forma

$$\tilde{X}_t = \theta \tilde{X}_{t-1} + \varepsilon_t \quad (t = 1, 2, \dots), \quad (2.3)$$

que es la estándar para los modelos autorregresivos de primer orden.

En cuanto a la restricción que se hace sobre el parámetro autorregresivo ($0 \leq \theta < 1$), nos parece adecuado centrar nuestra aportación al caso no negativo, ya que lo que intentamos explicar con este modelo son series que siempre permanezcan por encima de un nivel mínimo, más concretamente series de datos no negativos ($\mu = 0$).

Además, hay que tener en cuenta que ajustar una serie con observaciones por debajo del nivel mínimo sería imposible si en la serie aparecen al menos dos datos consecutivos menores que μ , aún considerando la posibilidad de que el parámetro autorregresivo tome valores negativos; incluso en el caso de que no aparecieran rachas de datos por debajo del nivel mínimo se puede tener dificultades al realizar el ajuste.

Por otra parte, si la serie sólo posee datos por encima de un nivel mínimo, el estimador de máxima verosimilitud de θ es positivo, y la función de densidad a posteriori en un análisis bayesiano tendrá la mayor parte de su masa en \mathbb{R}^+ , incluido el estimador de máxima verosimilitud generalizado (moda de la distribución a posteriori), y esto ocurre aunque considerásemos a priori una distribución con cierta masa en \mathbb{R}^- .

Por todo esto, queremos destacar el mayor interés que presentan los modelos tal como lo hemos definido en (2.1) o (2.3) y de ahí que los resultados obtenidos se refieren al modelo bajo las condiciones dadas, sin pretender por ello restringirlo de manera taxativa a este caso en todas las situaciones que se pudieran presentar.

Hemos obviado también el caso explosivo ($\theta \geq 1$) si bien el análisis en esta situación apenas se diferenciaría del que realizamos pues bastaría, llegado el momento, considerar una distribución a priori no informativa que contemplara esta posibilidad.

3. ANÁLISIS DEL MODELO CUANDO EL NIVEL MÍNIMO ES CONOCIDO

En una primera fase supondremos que el nivel mínimo del proceso μ es conocido, generalmente $\mu = 0$, por lo que consideraremos el proceso ARE(1) tal como se describe en la expresión (2.3) y estimaremos el coeficiente autorregresivo θ y el parámetro λ asociado a las distribuciones de los errores.

Como es bien conocido, en el análisis bayesiano es deseable encontrar una familia de distribuciones conjugadas adecuada que permita llegar a determinar la distribución a posteriori de los parámetros con relativa facilidad. Caro, Domínguez & Girón (10) proponen la familia de distribuciones gamma-Pareto bilateral truncada (*GBTP*) para el análisis del modelo ARE(1) cuando $\theta \in \mathbb{R}$. Sin embargo, atendiendo a las consideraciones mencionadas en la sección precedente, podemos restringir θ al intervalo $[0, 1)$; con lo

cual se reduce la dimensión del estadístico suficiente y consecuentemente el número de parámetros que caracterizan este tipo de distribuciones.

Comencemos por calcular la función de densidad conjunta para las n primeras observaciones x_1, \dots, x_n de un modelo ARE(1), dado un valor inicial conocido x_0 y el nivel mínimo $\mu = \mu_0$; para ello, nos basta con obtener la distribución condicionada de X_t dado x_0, \dots, x_{t-1} y μ_0 para cada instante de tiempo t , que es fácil de determinar si tenemos en cuenta que el proceso autorregresivo de primer orden es una cadena de Markov, que la distribución de los errores es exponencial y que estos errores se pueden expresar en la forma

$$\varepsilon_t = \tilde{X}_t - \theta \tilde{X}_{t-1} \quad (t = 1, 2, \dots),$$

con lo que tendremos que

$$\begin{aligned} f(x_t | x_0, \dots, x_{t-1}, \mu_0, \theta, \lambda) &= f(x_t | x_{t-1}, \mu_0, \theta, \lambda) \\ &= \lambda \exp\{-\lambda(\tilde{x}_t - \theta \tilde{x}_{t-1})\} I(\mu_0 + \theta \tilde{x}_{t-1}, \infty)(x_t), \end{aligned}$$

con $\tilde{x}_t = x_t - \mu_0$ ($t = 0, 1, \dots$), de donde se sigue que la densidad conjunta para $x = (x_1, \dots, x_n)$ viene dada por

$$f(x | x_0, \mu_0, \theta, \lambda) = \prod_{t=1}^n f(x_t | x_{t-1}, \mu_0, \theta, \lambda) = \lambda^n \exp\{-\lambda(S_0 - \theta S_1)\} I_A(x),$$

donde

$$S_i = \sum_{t=1}^n \tilde{x}_{t-i} \quad (i = 0, 1) \tag{3.1}$$

y el soporte de x viene dado por el conjunto

$$A = \{x \in \mathbb{R}^n: x_t - \mu_0 > \theta(x_{t-1} - \mu_0), t = 1, \dots, n\}.$$

Entonces la función de verosimilitud para el vector de parámetros $\phi = (\theta, \lambda)$ vendrá dada por

$$l(\phi) = \lambda^n \exp\{-\lambda(S_0 - \theta S_1)\} I_{[0, \theta_0] \times (0, \infty)}(\phi), \tag{3.2}$$

donde

$$\theta_0^* = \min_{1 \leq t \leq n} \left\{ 1, \frac{\tilde{x}_t}{\tilde{x}_{t-1}} \right\}.$$

Teniendo en cuenta la forma funcional de esta función de verosimilitud, la familia de distribuciones conjugadas naturales que proponemos es del tipo

$$f(\phi | \alpha, \beta_0, \beta_1, \theta_0) \propto \lambda^{\alpha-1} \exp\{-\lambda(\beta_0 - \theta\beta_1)\} I_{(0, \infty) \times [0, \theta_0]}(\phi),$$

con $\alpha > 1$, $\beta_0 - \theta_0\beta_1 > 0$, $\beta_0 > 0$, $\beta_1 > 0$, $0 \leq \theta_0 < 1$ y constante de normalización dada por

$$\begin{aligned} k(\alpha, \beta_0, \beta_1, \theta_0) &= \left[\int_0^{\theta_0} \int_0^{\infty} \lambda^{\alpha-1} \exp\{-\lambda(\beta_0 - \theta\beta_1)\} d\lambda d\theta \right]^{-1} \\ &= \frac{\beta_1}{\Gamma(\alpha-1)} \left[(\beta_0 - \theta_0\beta_1)^{-\alpha+1} - \beta_0^{-\alpha+1} \right]^{-1}. \end{aligned} \tag{3.4}$$

Caracterización. Esta familia de distribuciones se puede caracterizar como sigue: la distribución condicionada de λ dado θ es una gamma con parámetros α y $\beta_0 - \theta\beta_1$; es decir,

$$f(\lambda|\theta) = \frac{(\beta_0 - \theta\beta_1)^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp\{-\lambda(\beta_0 - \theta\beta_1)\} I_{(0,\infty)}(\lambda) \quad (0 \leq \theta < \theta_0),$$

y la distribución marginal de θ es de la forma

$$f(\theta) = k(\alpha, \beta_0, \beta_1, \theta_0) \Gamma(\alpha) (\beta_0 - \theta\beta_1)^{-\alpha} I_{[\theta_0, \theta_0]}(\theta),$$

donde se puede observar que la variable $\beta_0 - \theta\beta_1$ sigue una distribución de Pareto truncada de parámetros $\alpha - 1$ y $\beta_0 - \theta_0\beta_1$ y con soporte dado por el intervalo $(\beta_0 - \theta_0\beta_1, \beta_0]$, por lo que llamaremos a la distribución de θ *Pareto truncada opuesta*, y lo indicaremos por $\theta \sim \mathcal{PTO}(\alpha - 1, \beta_0, \beta_1, \theta_0)$. A la distribución conjunta la llamaremos *gamma-Pareto truncada opuesta* y la denotaremos por $\phi \sim \mathcal{GPTO}(\alpha, \beta_0, \beta_1, \theta_0)$.

Esta caracterización nos lleva a definir dos nuevas distribuciones, unidimensional y bidimensional respectivamente.

Definición 1. Diremos que una variable aleatoria Y sigue una distribución de Pareto truncada opuesta con parámetros α, β_0, β_1 e y_0 si su función de densidad es de la forma

$$f(y|\alpha, \beta_0, \beta_1, y_0) = \frac{\alpha\beta_1}{(\beta_0 - \beta_1 y_0)^{-\alpha}} \left(\frac{1}{\beta_0 - \beta_1 y}\right)^{\alpha+1} I_{[0, y_0]}(y), \tag{3.5}$$

con $\alpha > 0, \beta_0 - \beta_1 y_0 > 0, \beta_0 > 0, \beta_1 > 0$ e $y_0 > 0$.

Lo indicaremos por $Y \sim \mathcal{PTO}(\alpha, \beta_0, \beta_1, y_0)$.

Como ya se ha comentado, Y es tal que $Z = \beta_0 - \beta_1 Y$ sigue una distribución Pareto truncada de parámetros $\alpha, \beta_0 - \beta_1 y_0$ y β_0 , de acuerdo con la definición dada en Kotz & Johnson (11); es decir, Z posee una función de densidad dada por

$$f(z) = \frac{\alpha}{\beta_0 - \beta_1 y_0} \left[1 - \left(\frac{\beta_0 - \beta_1 y_0}{\beta_0}\right)^\alpha\right]^{-1} \left(\frac{\beta_0 - \beta_1 y_0}{z}\right)^{\alpha+1} I_{[\beta_0 - \beta_1 y_0, \beta_0]}(z).$$

Definición 2. Diremos que un vector aleatorio (X, Y) sigue una distribución gamma-Pareto truncada opuesta con parámetros α, β_0, β_1 e y_0 si su función de densidad conjunta es de la forma

$$f(x, y|\alpha, \beta_0, \beta_1, y_0) = k(\alpha, \beta_0, \beta_1, y_0) \alpha^{\alpha-1} e^{-(\beta_0 - \beta_1 y)x} I_{(0, \infty) \times [0, y_0]}(x, y), \tag{3.6}$$

con $\alpha > 1, \beta_0 - \beta_1 y_0 > 0, \beta_0 > 0, \beta_1 > 0, y_0$ y $k(\alpha, \beta_0, \beta_1, y_0)$ es una constante positiva dada por (3.4).

Lo indicaremos por $(X, Y) \sim \mathcal{GPTO}(\alpha, \beta_0, \beta_1, y_0)$.

En la definición de esta familia hay que hacer una primera observación relativa al soporte del parámetro α ; este so-

porte puede ser extendido a $0 < \alpha \leq 1$; sin embargo, se pierde la caracterización relacionada con la distribución marginal de Y , pues $\alpha - 1$ serían menor que 0 y no puede considerarse entonces como una transformación de una distribución de Pareto; tampoco la marginal de X puede ser expresada en términos de distribuciones usuales.

Por supuesto, esto no es óbice para usar esta extensión en caso necesario; de hecho, la distribución impropia a priori no informativa se puede obtener como un caso límite de esta familia cuando α tiende a 0, siendo la distribución marginal de Y , en el límite, una uniforme.

Por otro lado, también podemos extender la constante de proporcionalidad (3.4) a valores de α comprendidos entre 0 y 1; es decir, la familia de \mathcal{GPTO} lleva asociada la siguiente función

$$k(\alpha, \beta_0, \beta_1, y_0) = \left[\int_0^{y_0} \int_0^\infty x^{\alpha-1} \exp\{-(\beta_0 - \beta_1 y_0)x\} dx dy \right]^{-1};$$

definida para $\alpha > 0, \beta_0 - \beta_1 y_0 > 0, \beta_0 > 0, \beta_1 > 0$ e $y_0 > 0$; obteniéndose que

$$k(\alpha, \beta_0, \beta_1, y_0) = \begin{cases} \frac{\beta_1}{\Gamma(\alpha-1)} [(\beta_0 - \beta_1 y_0)^{-\alpha+1} - \beta_0^{-\alpha+1}]^{-1} & \text{si } \alpha > 1, \\ \beta_1 \left\{ \log \left[\frac{\beta_0}{\beta_0 - \beta_1 y_0} \right] \right\}^{-1} & \text{si } \alpha = 1, \\ \frac{(\alpha-1)\beta_1}{\Gamma(\alpha)} [(\beta_0 - \beta_1 y_0)^{-\alpha+1} - \beta_0^{-\alpha+1}]^{-1} & \text{si } 0 < \alpha < 1. \end{cases} \tag{3.7}$$

No es difícil comprobar que si $(X, Y) \sim \mathcal{GPTO}(\alpha, \beta_0, \beta_1, y_0)$, entonces la función de densidad marginal de X es

$$f(x) = \frac{k(\alpha, \beta_0, \beta_1, y_0)}{\beta_1} x^{\alpha-2} (e^{-(\beta_0 - \beta_1 y_0)x} - e^{-\beta_0 x}) I_{(0, \infty)}(x)$$

que forma una mezcla de dos distribuciones gammas —donde es negativo y el otro mayor que uno— dada por

$$x - \frac{k(\alpha, \beta_0, \beta_1, y_0) \Gamma(\alpha-1)}{\beta_1 (\beta_0 - \beta_1 y_0)^{\alpha-1}} \mathcal{G}(\alpha-1, \beta_0 - \beta_1 y_0) - \frac{k(\alpha, \beta_0, \beta_1, y_0)}{\beta_1 \beta_0^{\alpha-1}} \mathcal{G}(\alpha-1, \beta_0).$$

La distribución marginal para Y es una $\mathcal{PTO}(\alpha - 1, \beta_0, \beta_1, y_0)$.

En la figura 1 se representa la función de densidad conjunta de un vector aleatorio (X, Y) con distribución gamma-Pareto truncada opuesta con parámetros $\alpha = 2, \beta_0 = 5, \beta_1 = 5$ e $y_0 = 0.5$. En la figura 2 se representa las funciones de densidad marginal de la distribución bidimensional representada en la figura 1. La distribución de X resulta una mezcla de distribuciones gammas, en este caso dos exponenciales, $X \sim 2 \text{Exp}(2.5) - \text{Exp}(5)$, mientras que $Y \sim \mathcal{PTO}(1, 5, 5, 0.5)$.

Hay que hacer notar que para la construcción de la familia conjugada natural que se propone, como la hacen en su

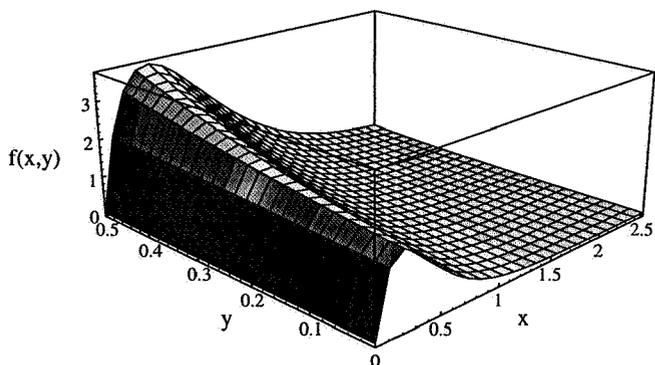


Figura 1. Distribución $GPTO(1, 5, 5, 0.5)$.

artículo Caro et al (10) y al igual que ocurre con observaciones independientes, se ha tenido en cuenta la existencia de un estadístico suficiente de dimensión fija; véase, por ejemplo, DeGroot (12) o Box & Tiao (13).

Con lo anteriormente expuesto podemos dar el siguiente teorema, que especifica la familia de distribuciones conjun-

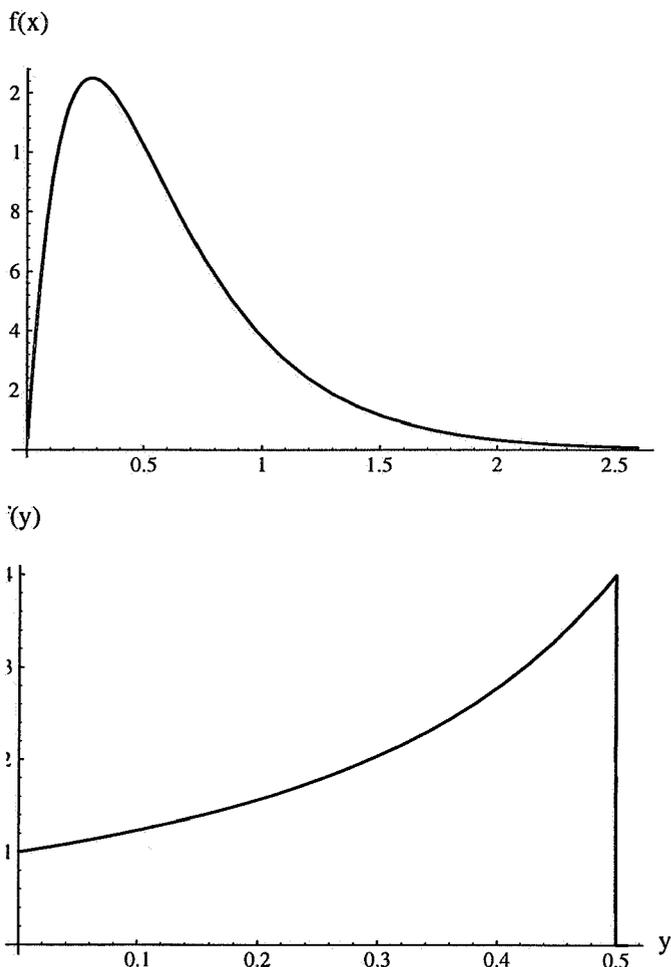


Figura 2. Distribuciones marginales de la distribución conjunta de la figura 1.

gadas para un modelo ARE(1) con nivel mínimo conocido.

Teorema 1. Si en el modelo AR(1) con errores exponenciales y nivel mínimo conocido ($\mu = \mu_0$), definido por la ecuación (2.3), la distribución conjunta a priori de ϕ es una distribución $GPTO(\alpha, \beta_0, \beta_1, \theta_0)$, entonces la distribución a posteriori dado $\tilde{x}_0, \tilde{x}_1, \dots, \tilde{x}_n$ donde $\tilde{x}_0 \geq 0$ es un valor inicial conocido y $\tilde{x}_1, \dots, \tilde{x}_n$, una realización del proceso, es $GPTO(\alpha^{(n)}, \beta_0^{(n)}, \beta_1^{(n)}, \theta_0^{(n)})$ con

$$\begin{aligned} \alpha^{(n)} &= \alpha + n, \\ \beta_i^{(n)} &= \beta_i + S_i \quad (i=0, 1), \\ \theta_0^{(n)} &= \min\{\theta_0, \theta_0^*\}. \end{aligned}$$

donde S_0, S_1 , y θ_0^* se define en (3.1) y (3.3), respectivamente.

Demostración. Haciendo uso del teorema de Bayes, la función de densidad a posteriori es proporcional al producto de la función de verosimilitud por la función de densidad a priori; es decir, $f(\phi | \tilde{x}_0, \dots, \tilde{x}_n) \propto f(\tilde{x}_1, \dots, \tilde{x}_n | \tilde{x}_0, \phi) f(\phi) = l(\phi) f(\phi)$, donde $l(\phi)$ está dada en (2.7) y la distribución a priori $f(\phi)$ es de la familia $GPTO$, se obtiene directamente el resultado deseado,

$$f(\phi | \tilde{x}_0, \dots, \tilde{x}_n) \propto \lambda^{\alpha+n-1} \exp\{-\lambda[\beta_0 + S_0 - \theta(\beta_1 + S_1)]\} I_{(0, \infty) \times [0, \min\{\theta_0, \theta_0^*\}]}(\phi).$$

Es decir, la familia de distribuciones $GPTO$, dada en la definición 2, es una familia de distribuciones conjugadas para ϕ .

En lo que sigue, nuestro objetivo es calcular las distribuciones marginales a posteriori de los parámetros del modelo partiendo de una distribución conjunta a priori no informativa, que proponemos a continuación.

Se tiene que la función de densidad a priori no informativa para λ es la usual para parámetros de escala; esto es, proporcional a λ^{-1} ; mientras que para θ , y puesto que $\theta \tilde{x}_0$ es un parámetro de localización con \tilde{x}_0 constante en nuestro modelo, se puede elegir una distribución a priori uniforme en $[0, 1]$ —véase, por ejemplo, Berger (14)— esto es posible bajo la hipótesis de estacionaridad del modelo (si no contemplamos esta hipótesis debemos considerar a priori una distribución impropia uniforme en \mathbb{R}^+ , que no modifica sustancialmente el análisis siguiente). Con todo esto, si asumimos la hipótesis de independencia a priori entre los parámetros del modelo, podemos entonces considerar una distribución a priori conjunta impropia no informativa dada por $f(\phi) \propto \lambda^{-1} I_{(0, \infty) \times [0, 1]}(\phi)$.

La distribución a priori considerada, es un caso límite (distribución impropia) de la familia de distribuciones conjugadas propuesta, cuando ésta se extiende a valores de α en el intervalo $(0, 1]$, con α, β y β_1 tendiendo a 0, mientras $\theta_0 = 1$; con lo cual, aplicando el teorema 1 se tendrá que la distribución conjunta a posteriori es una distribución $GPTO(n, S_0, S_1, \theta_0^*)$ cuya función de densidad viene dada por

$$f(\phi|D_n) = k(n, S_0, S_1, \theta_0^*) \lambda^{n-1} \exp\{-\lambda(S_0 - \theta S_1)\} I_{(0, \infty) \times [0, \theta_0^*]}(\phi).$$

En lo sucesivo, siempre que no de lugar a equívocos, a la constante de normalización $k(n, S_0, S_1, \theta_0^*)$ la denotaremos por $k(n)$ y en general podemos definir

$$k(t) = k(n, S_0, S_1, \theta_0^*) \quad (t = 1, 2, \dots). \tag{3.8}$$

Hay que hacer notar, a la vista de esta densidad, que es deseable contar al menos con dos observaciones en la serie de datos para que la distribución a posteriori pertenezca a la familia de distribuciones *GPFO*, tal como se expresa en la definición 2, y podamos hacer uso de los resultados anteriores.

Es fácil observar que las densidades marginales a posteriori para λ y θ son, respectivamente,

$$f(\lambda|D_n) = \frac{k(n)}{S_1} \lambda^{n-2} \left(e^{-\lambda(S_0 - \theta_0^* S_1)} - e^{-\lambda S_0} \right) I_{(0, \infty)}(\lambda), \tag{3.9}$$

que se puede expresar como una mixtura de dos distribuciones gamma, en la forma

$$\lambda|D_n \sim \frac{k(n)\Gamma(n-1)}{S_1(S_0 - \theta_0^* S_1)^{n-1}} \mathcal{G}(n-1, S_0 - \theta_0^* S_1) - \frac{k(n)\Gamma(n-1)}{S_1(S_0)^{n-1}} \mathcal{G}(n-1, S_0);$$

y

$$f(\theta|D_n) = k(n)\Gamma(n) (S_0 - \theta S_1)^{-n} I_{[0, \theta_0^*]}(\theta), \tag{3.10}$$

que es una distribución *PFO*($n-1, S_0, S_1, \theta_0^*$).

4. ANÁLISIS DEL MODELO CUANDO EL NIVEL MÍNIMO SE DESCONOCE

Una de las innovaciones que planteamos en el modelo ARE(1) en relación a otros modelos que también analizan series temporales no negativas —por ejemplo, los modelos EAR(1) y NEAR(1) descritos en la introducción o el propio modelo autorregresivo con errores exponenciales introducido por Domínguez (4)— es la incorporación al modelo de un coeficiente independiente c de acuerdo con la ecuación (2.1), lo que nos lleva a definir el *nivel mínimo del proceso* como $\mu = c/(1 - \theta)$; es decir, todos los datos de la serie son valores superiores a μ , de esta manera generalizamos los modelos mencionados.

Hasta ahora el análisis bayesiano planteado en la sección precedente se ha realizado bajo la hipótesis de un nivel mínimo conocido. De manera general, podemos suponer que μ es igual a cero, lo que nos lleva a series de datos no negativos; trabajar bajo esta hipótesis no plantea inconvenientes graves si el verdadero valor del parámetro no difiere de manera significativa de cero. Sin embargo, si $|\mu|$ es grande, el hecho de considerar el nivel mínimo nulo ocasiona graves

problemas en las estimaciones de los otros dos parámetros del modelo, en particular de la autocorrelación del proceso, con lo que el modelo queda desvirtuado.

En el análisis bayesiano conjunto de los tres parámetros $\phi^* = (\mu, \lambda, \theta)$ nos encontramos con el grave inconveniente de su complejidad, debido a la dependencia que la muestra ocasiona sobre el soporte de la distribución a posteriori, aunque se pueda encontrar una familia de distribuciones conjugadas naturales para ϕ^* .

De manera análoga al caso anterior no es difícil obtener la función de verosimilitud para ϕ^* , que viene dada por

$$l(\phi^*) = \lambda^n \exp\{-n\lambda[\bar{x}_0 - \mu - \theta(\bar{x}_1 - \mu)]\} I_R(\phi^*), \tag{4.1}$$

donde

$$\bar{x}_i = \frac{1}{n} \sum_{t=1}^n x_{t-i} \quad (i = 0, 1)$$

y el soporte R^* es el subconjunto de \mathbb{R}^3 dado por

$$R^* = \left\{ (\mu, \theta, \lambda) \in \mathbb{R}^3: \mu \leq x_0, 0 \leq \theta < \min_{1 \leq i \leq n} \left\{ 1, \frac{x_i - \mu}{x_{i-1} - \mu} \right\}, \lambda > 0 \right\},$$

y de aquí surge de manera natural la familia de distribuciones conjugadas dada por la función de densidad

$$f(\phi^*|\alpha_0, \alpha_1, \beta_0, \beta_1, \mu_0, \Theta(\cdot)) \propto \lambda^{\alpha_0-1} \exp\{-\alpha_1 \lambda[\beta_0 - \mu - \theta(\beta_1 - \mu)]\} I_R(\phi^*), \tag{4.2}$$

donde

$$\begin{aligned} \alpha_0 &> 1, \\ \alpha_1 &> 0, \\ \beta_1 - \mu &> 0 \quad (i=0, 1, \mu \leq \mu_0), \\ \beta_0 - \mu - \Theta(\mu) (\beta_1 - \mu) &> 0 \quad (\mu \leq \mu_0), \end{aligned}$$

y el soporte de la distribución está dado por

$$R = \{\phi^* \in \mathbb{R}^3: \mu \leq \mu_0, 0 \leq \theta < \Theta(\mu), \lambda > 0\},$$

donde $\mu_0 \in \mathbb{R}$ y $\Theta(\cdot)$ es una función continua de $(-\infty, \mu_0]$ en $[0, 1]$.

Como se puede observar la distribución posee un soporte de gran complejidad delimitado por una función $\Theta(\cdot)$ que, como se verá en el análisis posterior, no es derivable y cuya dificultad aumenta a medida que aumenta el tamaño muestral. Todo esto hace prácticamente prohibitivo obtener resultados analíticos, con lo que, inicialmente, sólo se pueden obtener estimaciones de los parámetros utilizando métodos numéricos.

En cualquier caso es posible plantear el siguiente teorema.

Teorema 2. Si en el modelo ARE(1), con un valor inicial x_0 fijado, la distribución conjunta a priori de $\phi^* = (\mu, \theta, \lambda)$

tiene una función de densidad como la definida en la expresión (4.2) con parámetros $\alpha_0, \alpha_1, \beta_0, \beta_1$ y soporte R , con $\mu_0 \leq x_0$, entonces la distribución a posteriori dado D_n es del mismo tipo, con parámetros

$$\alpha_i^{(n)} = \alpha_i + n \quad (i = 0, 1),$$

$$\beta_i^{(n)} = \beta_i + \bar{x}_i \quad (i = 0, 1),$$

y soporte

$$R^{(n)} = \{\phi^* \in \mathbb{R}^3: \mu \leq m_0^{(n)}, 0 \leq \theta < \Theta^{(n)}(\mu), \lambda > 0\},$$

donde

$$\mu_0^{(n)} = \min_{1 \leq i \leq n} \{\mu_0, x_i\}$$

y

$$\Theta^{(n)}: (-\infty, \mu_0^{(n)}) \rightarrow [0, 1]$$

definida por

$$\Theta^{(n)}(\mu) = \min_{1 \leq i \leq n} \left\{ \Theta(\mu), \frac{x_i - \mu}{x_{i-1} - \mu} \right\}.$$

Demostración. Al igual que el teorema 1 el resultado es consecuencia del teorema de Bayes, teniendo en cuenta que la función de verosimilitud viene dada por la expresión (4.1). ■

La restricción sobre μ_0 ($\mu_0 \leq x_0$) dada en la hipótesis del teorema es consecuencia de la propia naturaleza del modelo, de acuerdo con lo expuesto en la sección 2. Por otro lado, obsérvese que no hemos impuesto ninguna restricción por la izquierda al nivel mínimo, ya que no hay ningún impedimento teórico para ello; sin embargo, es posible que en determinados problemas sea necesario restringir el soporte a priori de μ a la recta real positivo. En cualquier caso el análisis del modelo no difiere de manera sustancial del que hacemos a continuación.

Así, por ejemplo, si consideramos una distribución a priori impropia no informativa, que es un caso límite de la familia propuesta con $\alpha_0 = \alpha_1 = 0, \beta_0 = \beta_1 = 0, \mu_0 = x_0$ y $\Theta(\mu) = I_{(-\infty, x_0)}(\mu)$; es decir, con función de densidad dada por

$$f(\phi^*) \propto \lambda^{-1} I_{(-\infty, x_0] \times (0, \infty) \times [0, 1]}(\phi^*),$$

se sigue del teorema 2 que la distribución a posteriori está determinada por

$$f(\phi^* | D_n) c_n \frac{n^n}{\Gamma(n-1)} \lambda^{n-1} \exp\{-n\lambda[\bar{x}_0 - \mu - \theta(\bar{x}_1 - \mu)]\} I_R(\phi^*), \tag{4.3}$$

con

$$c_n = \Gamma(n-1) \left[\int_{-\infty}^{\mu_0^{(n)}} \frac{1}{k(n, \bar{x}_0 - \mu, \bar{x}_1 - \mu, \Theta^*(\mu))} d\mu \right]^{-1}, \tag{4.4}$$

y soporte dado por

$$R^* = \{\phi^* \in \mathbb{R}^3: \mu \leq \mu_0^*, 0 \leq \theta < \Theta^*(\mu), \lambda > 0\},$$

donde

$$\mu_0^* = \min_{0 \leq i \leq n} \{x_i\} \quad \text{y} \quad \Theta^*(\mu) = \min_{1 \leq i \leq n} \left\{ 1, \frac{x_i - \mu}{x_{i-1} - \mu} \right\} \quad (\mu \leq \mu_0^*).$$

Con esto, podemos expresar las funciones de densidad marginales a posteriori en términos de c_n , en la forma

$$f(\mu | D_n) = \frac{c_n}{\Gamma(n-1)} \frac{1}{k(n, \bar{x}_0 - \mu, \bar{x}_1 - \mu, \Theta^*(\mu))} I_{(-\infty, \mu_0^*]}(\mu),$$

$$f(\lambda | D_n) = \frac{n^{n-1}}{\Gamma(n-1)} c_n \lambda^{n-2} \cdot \left[\int_{-\infty}^{\mu_0^*} \frac{(e^{-\lambda(\bar{x}_0 - \mu) - \theta^*(\mu)(\bar{x}_1 - \mu)} - e^{-\lambda(\bar{x}_0 - \mu)})}{\bar{x}_1 - \mu} d\mu \right] I_{(0, \infty)}(\lambda), \tag{4.5}$$

$$f(\theta | D_n) = c_n (n-1)^{n-1} \left(\frac{1 - \theta}{\bar{x}_0 - \theta \bar{x}_1 - \min_{1 \leq i \leq n} \{x_i - \theta x_{i-1}\}} \right)^{n-1} I_{[0, 1]}(\theta).$$

Como se observa, al trabajar con mínimos de funciones, nos vemos obligados a utilizar métodos numéricos para calcular los estimadores (medias de las distribuciones marginales a posteriori). En cualquier caso las estimaciones de θ , y sobre todo de λ , son extremadamente complicadas a partir de estas densidades por lo que es aconsejable, una vez realizada la estimación de μ , que notaremos por $\hat{\mu}_B$, completar el análisis del modelo de acuerdo a lo analizado en la sección 3, considerando $\bar{x}_i = x_i - \hat{\mu}_B$.

Hay que hacer notar que el soporte marginal de θ es todo el intervalo $[0, 1)$ debido a que no existe restricción por la izquierda para el nivel mínimo. En cambio, si exigimos que μ sea no negativo, entonces el soporte de θ se modifica con la muestra, como ocurre con el caso en que μ es conocido; en concreto, el intervalo donde se mueve θ para este caso es $[0, \Theta^*(0))$.

Como ya hemos comentado, en algunas situaciones podemos suponer que $\mu = 0$. Por lo que sería interesante, siempre que fuera posible, contrastar esta hipótesis; es decir, plantear el contraste de hipótesis

$$H_0: \mu = 0,$$

$$H_1: \mu \neq 0.$$

Si la hipótesis nula es aceptada, nuevamente aconsejamos retomar lo expuesto en la sección precedente, bajo el supuesto de que μ es conocido; en caso contrario, nos vemos obligados a estimar el nivel mínimo a partir de la distribución marginal a posteriori.

Para realizar el contraste de hipótesis podemos hacer uso del ratio α_0/α_1 —véase Berger (14)—, donde α_0 y α_1 son las probabilidades a posteriori de las hipótesis nula y alternativa, respectivamente. Si este ratio está cercano a 0, convendrá concluir que la hipótesis alternativa es la más creíble. Si, por el contrario, α_0/α_1 toma un valor alto, por ejemplo 10, parece lógico tomar como conclusión que la hipótesis H_0 es «10 veces más verosímil» que la hipótesis H_1 . Si el resultado es un valor cercano a 1 la decisión a tomar, ya no es tan clara.

En ocasiones también es interesante el uso del factor de Bayes, definido por

$$B = \frac{\alpha_0 \pi_1}{\alpha_1 \pi_0},$$

donde π_0 y π_1 son las probabilidades a priori de las hipótesis nula y alternativa respectivamente; generalmente, supondremos que $\pi_0 = \pi_1 = \frac{1}{2}$, con lo que el factor de Bayes es el ratio α_0/α_1 .

Para un estudio en profundidad de ambos cocientes véase la obra de Berger (14) o de DeGroot (12).

Con todo esto y para nuestro caso, si partimos de una realización del proceso $x = (x_1, \dots, x_n)$, se tiene que

$$\alpha_0 = \Pr(\mu = 0 | D_n) = \frac{\pi_0 m_0(x)}{\pi_0 m_0(x) + \pi_1 m_1(x)} = \left(1 + \frac{\pi_1 m_1(x)}{\pi_0 m_0(x)} \right)^{-1},$$

con

$$m_0(x) = \int_0^{\Theta(0)} \int_0^\infty f(x|x_0, \mu = 0, \theta, \lambda) f_0(\phi^*) d\lambda d\theta,$$

$$m_1(x) = \int_{-\infty}^{\mu(0)} \int_0^{\Theta(\mu)} \int_0^\infty f(x|x_0, \mu, \theta, \lambda) f_1(\phi^*) d\lambda d\theta d\mu,$$

donde $f_0(\phi^*)$ y $f_1(\phi^*)$ son las funciones de densidad a priori bajo la hipótesis H_0 y H_1 , respectivamente.

Si asumimos la independencia a priori entre μ y $\phi = (\theta, \lambda)$, denotamos por $f(\phi)$ la función de densidad de ϕ y a priori suponemos que la probabilidad de que $\mu = 0$ es π_0 , mientras que la probabilidad de que $\mu \neq 0$ es π_1 de acuerdo a la función de densidad $f(\mu)$; entonces m_0 y m_1 puede expresarse, respectivamente, en la forma

$$m_0(x) = \int_0^{\Theta(0)} \int_0^\infty f(x|x_0, \mu = 0, \theta, \lambda) f(\phi) d\lambda d\theta,$$

$$m_1(x) = \int_{-\infty}^{\mu(0)} \int_0^{\Theta(\mu)} \int_0^\infty f(x|x_0, \mu, \theta, \lambda) f(\phi) f(\mu) d\lambda d\theta d\mu,$$

Entonces, teniendo en cuenta que $\alpha_1 = 1 - \alpha_0$, es fácil llegar a

$$\frac{\alpha_0}{\alpha_1} = \frac{\pi_0 m_0(x)}{\pi_1 m_1(x)},$$

concluyendo que, para $\pi_0 = \pi_1$, el factor de Bayes es

$$B = \frac{m_0(x)}{m_1(x)}.$$

5. ESTUDIO DE UN EJEMPLO

Como aplicación se ha generado una serie de 50 observaciones a partir de una simulación de un modelo ARE(1). En concreto, el modelo simula es $X_t = 0.8X_{t-1} + \varepsilon_t$, donde la media de ε_t es 2 ($\lambda = 0.5$) y como dato inicial se ha elegido la media de la distribución estacionaria,

$$X_0 = E[X_t] = \mu + \frac{1}{\lambda(1-\theta)} = 10.$$

Obtendremos, primero, las distribuciones a posteriori para λ y θ bajo el supuesto de que μ es conocido ($\mu = 0$). Utilizaremos para ello una distribución a priori impropia no informativa.

Se tiene que, en nuestro caso, el estadístico suficiente de dimensión fija (n, S_0, S_1, θ_0^*), que describe la función de verosimilitud, es (50, 548.584, 550.693, 0.805295) y, de acuerdo con el teorema 1, llegamos a que la distribución conjunta a posteriori para los parámetros del modelo es una distribución $\mathcal{GPTO}(50, 548.584, 550.693, 0.805295)$, mientras que las marginales son $\mathcal{G}(49, 105.114)$ y $\mathcal{PTO}(49, 548.584, 550.693, 0.805295)$, respectivamente. El valor de la constante de normalización es $k(50) = 5.10934 \cdot 10^{40}$.

Obsérvese que la distribución marginal de λ es una gamma y no una mixtura, puesto que el segundo término es despreciable para tamaños de muestra «suficientemente» grandes. De manera general se puede comprobar que

$$\frac{k(n)\Gamma(n-1)}{S_1(S_0 - \theta_0^* S_1)^{n-1}} \rightarrow 1 \quad \text{cuando} \quad n \rightarrow \infty.$$

Si bien, tanto la moda (estimador de máxima verosimilitud generalizado) como la mediana (estimador bayesiano bajo función de pérdida definida por el valor absoluto de la diferencia entre el parámetro y la decisión a toma) de la distribución a posteriori pueden ser utilizados como estimadores bayesianos de los parámetros del modelo —véase; por ejemplo, DeGroot (12) o Berger (14)—, sin duda el estimador más utilizado en análisis bayesiano es la media de dicha distribución, que en Teoría de la Decisión minimiza el riesgo bajo función de pérdida cuadrática, y cuya precisión o error de estimación viene dada por la inversa de la matriz de covarianza. Por tanto, y siguiendo esta

pauta, realizaremos nuestra estimación a través de la esperanza a posteriori.

A partir de las esperanzas de las funciones de densidad marginal, se tiene que las estimaciones de λ y θ son $\hat{\lambda}_B = 0.466161$ y $\hat{\theta}_B = 0.801319$.

Podemos también determinar los intervalos MDP. Al 95% se tiene que el intervalo de θ es (0.793262, 0.805295), mientras que el de λ es (0.339164, 0.598557).

En la figura 3 se representa la distribución conjunta a priori y en la figura 4 las marginales junto con los intervalos MDP, que aparecen sombreados.

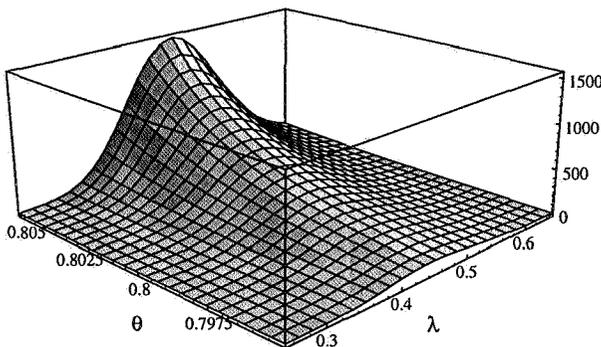


Figura 3. Distribución conjunta a posteriori para el ejemplo.

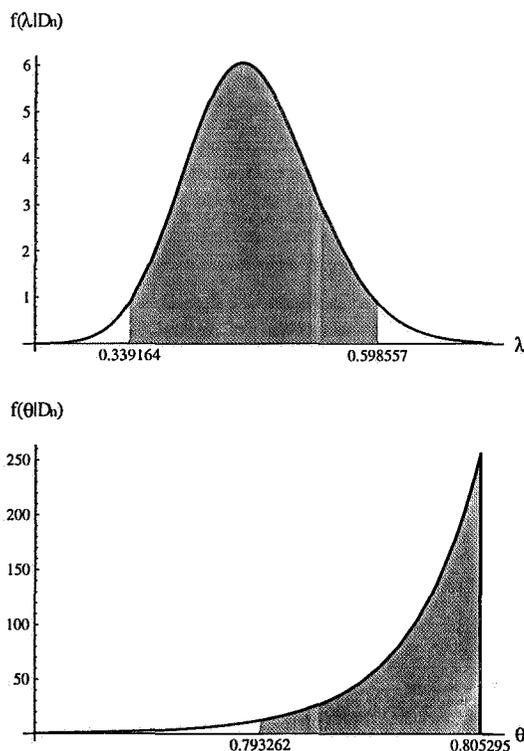


Figura 4. Distribuciones marginales a posteriori e intervalo MDP al 95%.

Estimaremos, ahora, el nivel mínimo del proceso a partir de la función de densidad marginal a posteriori de μ dada en (4.5).

Supondremos que los datos son no negativos, lo que nos lleva a considerar al intervalo $[0, \alpha_0]$ como soporte para la distribución a priori de μ ; es decir, la distribución a priori no informativa, viene dada por la función de densidad impropia

$$f(\phi^*) \propto \lambda^{-1} I_{[0, 10] \times (0, \infty) \times [0, 1]}(\phi^*),$$

de donde se sigue que la función de densidad marginal a posteriori, cuyo soporte es $[0, \mu_0^*]$ con $\mu_0^* = 5.44774$, está dada por

$$f(\mu | D_n) = \frac{c_n}{\Gamma(49)} \frac{1}{k(50, 10.9717 - \mu, 11.0139 - \mu, \Theta^*(\mu))} I_{[0, \mu_0^*]}(\mu),$$

donde $c_n = 9.32078 \cdot 10^{16}$, de acuerdo con la expresión (4.4), y

$$\Theta^*(\mu) = \min_{1 \leq i \leq 50} \left\{ 1, \frac{x_i - \mu}{x_{i-1} - \mu} \right\} \quad (0 \leq \mu \leq \mu_0^*).$$

En la figura 5 está dibujada la función de densidad a posteriori de μ y en la figura 6 se representa la función $\Theta^*(\cdot)$, que delimita el soporte conjunto a posteriori de (μ, θ) que aparece sombreado.

Entonces una estimación bayesiana de μ es el valor esperado de esta distribución que se ha de calcular, también, mediante métodos numéricos, resultando $\hat{\mu}_B = 0.498552$. En este caso la moda de la distribución, estimador de máxima verosimilitud generalizado, nos ofrece una mejor estimación, $\hat{\mu} = 0.135529$.

A partir de estas estimaciones, media y moda de la distribución a posteriori, podemos obtener las estimaciones de θ y λ , haciendo uso de (3.9) y (3.10) con $\bar{x}_i = x_i - \mu_B$ y $\bar{x}_i = x_i - \hat{\mu}$, respectivamente. Se obtiene, utilizando $\hat{\mu}_B$, que $\hat{\theta}_B = 0.789726$ y $\hat{\lambda}_B = 0.461251$; mientras que si utilizamos el estimador de máxima verosimilitud generalizado, resulta $\hat{\theta}_B = 0.799521$ y $\hat{\lambda}_B = 0.467768$.

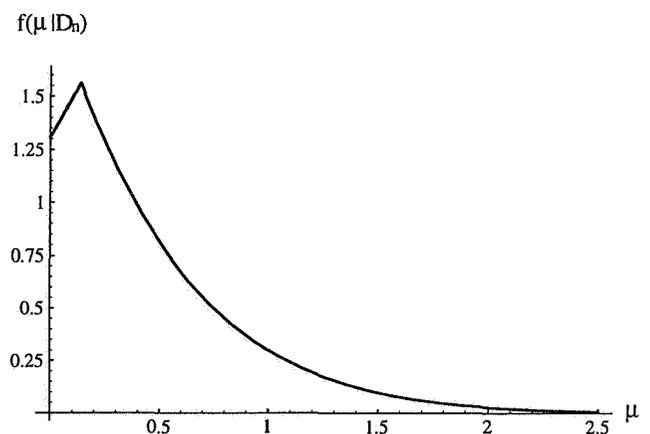


Figura 5. Función de densidad marginal a posteriori de μ .

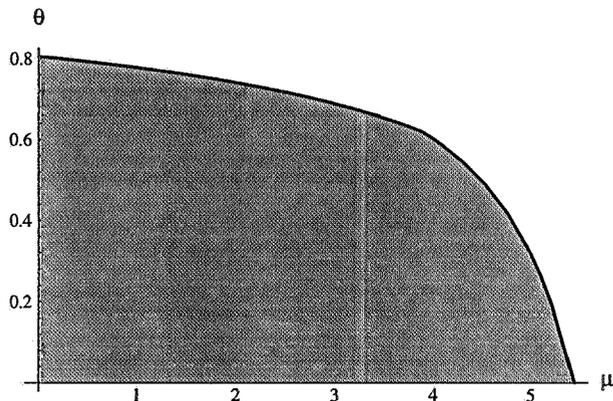


Figura 6. Soporte de (μ, θ) en la distribución conjunta a posteriori para el ejemplo.

En la tabla I, se puede observar los valores estimados de los parámetros θ y λ , de acuerdo a la decisión tomada acerca de μ . Planteamos tres posibles alternativas: I) aceptamos la hipótesis nula, lo que nos lleva a considerar $\mu = 0$; II) elegimos la media de la distribución a posteriori como estimador bayesiano; y III) elegimos la moda, estimador de máxima verosimilitud generalizado, para la estimación del nivel mínimo. Como se puede observar, en esta ocasión, la moda es mejor estimador que la media y a su vez provoca mejores estimaciones en los otros parámetros.

		$\hat{\theta}_b$	$\hat{\lambda}_b$
Si aceptamos H_0	$\mu = 0$	0.801319	0.466161
Si estimamos μ :	$\hat{\mu}_b = 0.498552$	0.789726	0.461251
	$\hat{\mu} = 0.135529$	0.799521	0.467768

Tabla I. Estimaciones bayesianas de los parámetros del modelo ARE(1) del ejemplo.

Por lo que respecta al contraste de hipótesis, podemos considerar la independencia a priori entre μ respecto de ϕ , y elegir para la hipótesis nula $\pi_0 = \Pr(\mu = 0) = \frac{1}{2}$ como probabilidad a priori y para la hipótesis alternativa se tendría $\pi_1 = \Pr(\mu \in (0, x_0]) = \frac{1}{2}$ de acuerdo a una distribución uniforme en este intervalo; es decir, con función de densidad $f(\mu) = 0.1 I_{(0, 10]}$, mientras que para ϕ optaremos por la distribución habitual; esto es, la distribución impropia dada por $f(\phi) \propto \lambda^{-1} I_{(0, 1) \times [0, \infty)}$.

Con lo que se obtiene que

$$\alpha_0 = \Pr(\mu = 0 | D_n) = \left(1 + \frac{\pi_1}{\pi_0} \frac{m_1(x)}{m_0(x)} \right)^{-1}$$

donde

$$\begin{aligned} \frac{m_0(x)}{m_1(x)} &= \frac{x_0 \int_0^{\Theta^*(0)} \int_0^\infty \lambda^{n-1} \exp\{-n\lambda[\bar{x}_0 - \theta(\bar{x}_1)]\} d\lambda d\theta}{\int_0^{\mu_0} \int_0^{\Theta^*(\mu)} \int_0^\infty \lambda^{n-1} \exp\{-n\lambda[\bar{x}_0 - \mu - \theta(\bar{x}_1 - \mu)]\} d\lambda d\theta d\mu} \\ &= \frac{x_0 c_n n_n}{\Gamma(n-1) k(n, S_0, S_1, \Theta^*(0))} \\ &= 13.0521, \end{aligned}$$

lo que nos da una probabilidad a posteriori para la hipótesis nula de $\alpha_0 = 0.928836$; por lo que es razonable aceptar esta hipótesis ($H_0: \mu = 0$), bajo las condiciones a priori que hemos impuesto.

Como se observa, no hemos entrado en consideraciones acerca de las condiciones a priori del test, ya que se aparta de nuestro verdadero objetivo. En cualquier caso, estas condiciones sobre las distribuciones a priori han sido siempre el gran caballo de batalla de los estadísticos bayesianos y siempre han suscitado gran polémica las decisiones a tomar en este punto, esto se hace más patente en los contrastes de hipótesis. Nos remitimos, por tanto, a la numerosa bibliografía que existe sobre este tema y en particular al texto de Berger (14) mencionado en varias ocasiones.

En la figura 7 se puede observar la probabilidad a posteriori de la hipótesis nula dependiendo de la probabilidad π_0 que elijamos a priori.

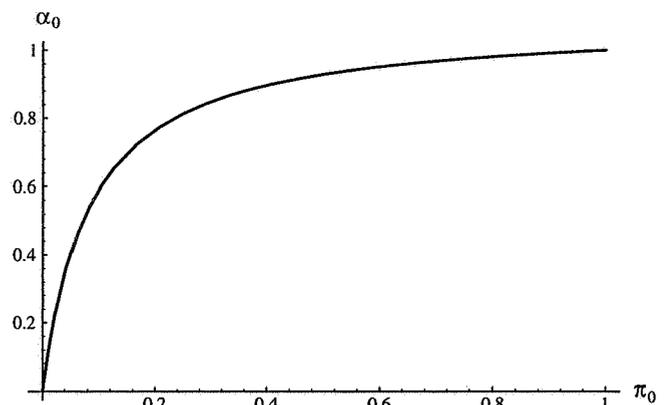


Figura 7. Probabilidad a posteriori de H_0 en función de la probabilidad a priori.

BIBLIOGRAFÍA

- Gaver, D.P. & Lewis, P.A.W. (1980) First-order autoregressive gamma sequences and point processes, *Adv. Appl. Prob.* 12, 727-745.
- Lawrence, A.J. & Lewis, P.A.W. (1981) A new autoregressive time series models in exponential variables (NEAR(1)). *Adv. Appl. Prob.* 13, 826-845.

3. Damsleth E. & El-Shaaravi, A.H. (1989) ARMA models with double exponentially distributed noise *J.R. Statist. Soc. B* **51**, 61-69.
4. Domínguez, J.I. (1990) Análisis bayesiano de un modelo AR(1) con errores exponenciales, *Estadística Española*, **32**, **124**, 401-416.
5. Bell, C.B. & Smith, E.P. (1986) Inference for non-negative autoregressive schemes. *Commun. Statist.-Theory Meth.* **15**, 2267-2293.
6. Andel, J. (1988) On AR(1) processes with exponential white noise. *Commun. Statist.-Theory Meth.* **17**, 1481-1495.
7. Davis, R.A. & McCormick, P. (1989) Estimation for first-order autoregressive processes with positive or bounded innovations. *Stoch. Proceses Appl.* **31**, 237-250.
8. Amaral, M.A. (1990) Bayesian analysis of an autoregressive process with exponential white noise. *Statistics* **21**, 601-608.
9. Box, G. E. P. & Jenkins, G.M. (1976) *Time Series Analysis: Forecasting and Control*. Revised Edition. Holden-Day, Oakland.
10. Caro, E., Domínguez, J.I. & Girón, F.J. (1994) A. Conjugate family for AR(1) processes with exponential errors. *Commun. Statist. Theory Meth.* **23**, 1777-1784.
11. Kotz, S. & Johnson, N.L. (1970) *Continuous Univariate distributions-I*. Willey, New York.
12. DeGroot, M.H. (1970) *Optimal Statistical Decisions*. McGraw-Hill, New York.
13. Box, G.E.P. & Tiao, G.C. (1973) *Bayesian Inference in Statistical Analysis*. Addison-Wesley, Reading, Massachusetts.
14. Berger, J.O. (1985) *Statistical Decisions and Bayesian Analysis*. Second Edition. Springer-Verlag, New York.