

UNIVERSIDAD CENTROCCIDENTAL
“LISANDRO ALVARADO”

Decanato de Ciencias y Tecnología
Licenciatura en Ciencias Matemáticas



“ESTIMACIÓN POR MÁXIMA CUASI-VEROSIMILITUD EN
MODELOS CON VOLATILIDAD ESTOCÁSTICA”

TRABAJO ESPECIAL DE GRADO PRESENTADO POR

JAVIER ALBERTO RAMÍREZ BRIZUELA

COMO REQUISITO FINAL
PARA OBTENER EL TÍTULO DE LICENCIADO
EN CIENCIAS MATEMÁTICAS
ÁREA DE CONOCIMIENTO: ESTADÍSTICA Y PROBABILIDAD.
TUTOR: MSc. JHONNY ESCALONA

*Dedicado a mis padres, a mi hermana y a
mi esposa.*

AGRADECIMIENTOS

Primeramente agradezco a Dios, quien ha sido mi luz y mi guía en este largo camino, y quien me ha dado la fuerza y la sabiduría para seguir adelante en cada paso de mi vida.

A mis padres, por estar siempre conmigo en cada instante de mi vida, por sus consejos y por su apoyo incondicional, y a mi hermana, porque juntos formamos esta maravillosa familia en la que vivo el día a día. Gracias por todo lo que me han enseñado y gracias por su compañía.

A mis amigos, amigas y compañeros de estudio, El Gocho (Juan Carlos) y Elvis con quienes compartí esta vida universitaria desde el inicio de la carrera, siempre estudiando, hasta momentos antes de cada examen. A Beatriz, Aura Rosa, y Massiel, quienes nos ayudaban en momentos de dificultades. Gracias por ser parte de esta experiencia, por vivir grandes y difíciles momentos, la incertidumbre de una evaluación, pero que al final, juntos culminamos nuestra carrera.

A Johela por brindarme su amistad y colaboración en mi trabajo, y a Joan, Joelviz, Datsy, Dannymar, Glendy, Kissy, Nataly, Katherina, Rosmary, gracias a todos por brindarme su mano amiga y amistad incondicional.

A los profesores Jhonny Escalona, quien fue mi tutor y colaborador en mi trabajo de grado; también a los profesores Rómulo Castillo, Mario Rodríguez, Javier Hernández, Ismael Huerta, Eibar Henández, gracias por su excelente educación y por ser parte de mi formación profesional.

Y finalmente, a mi esposa, mi novia, mi amiga y mi compañera fiel en todo momento, el amor de mi vida, Esthefany, por sus hermosas palabras en momentos difíciles y su apoyo incondicional.

A TODOS MIL GRACIAS, Y QUE DIOS LOS BENDIGA.

“ESTIMACIÓN POR MÁXIMA CUASI-VEROSIMILITUD EN MODELOS CON VOLATILIDAD ESTOCÁSTICA”

RESUMEN

En este trabajo se estudió el Algoritmo EM, como herramienta para obtener estimadores de Máxima Verosimilitud en un modelo con volatilidad estocástica. Para ello se implementaron dos versiones de algoritmo EM: primero se estudio el algoritmo EM para modelos de espacio-estado lineales, utilizando Filtros y Suavizado de Kalman. Segundo, se estudió una versión Monte Carlo del algoritmo EM, que permite obtener los estimadores de maxima verosimilitud del modelo, sin la necesidad de linealizar el mismo.

Se realizó un análisis de los métodos aplicados, con el fin de estudiar los resultados obtenidos, así como también una comparación entre los tiempos de ejecución y consumo computacional de ambos métodos.

ÍNDICE

| | |
|---|-----------|
| Agradecimientos | i |
| Resumen | ii |
| Introducción | 1 |
| 1. Preliminares. | 2 |
| 1.1. Esperanza y Covarianza de Variables Aleatorias | 2 |
| 1.2. Estimación por Máxima Verosimilitud | 5 |
| 1.3. La Distribución Normal Vectorial | 6 |
| 1.4. Distribuciones de Probabilidad (Familia Exponencial) | 6 |
| 1.5. Diferenciación de la traza de una matriz | 8 |
| 1.6. Métodos Monte Carlo | 9 |
| 1.6.1. Integración Monte Carlo | 10 |
| 1.7. Algoritmo Metropolis-Hasting | 10 |
| 1.7.1. Cadena de Markov | 10 |
| 1.7.2. Ecuación de Chapman-Kolmogorov | 11 |
| 1.7.3. Distribuciones Límites | 12 |
| 1.7.4. Algoritmo Metropolis Hasting | 13 |
| 2. Filtros y Suavizado de Kalman. | 14 |
| 2.1. Modelos de Estado-Espacio | 14 |
| 2.2. Filtros de Kalman | 15 |
| 2.3. Estimadores de Máxima Verosimilitud. | 23 |
| 3. Algoritmo EM | 25 |
| 3.1. Definición de algoritmo EM | 25 |
| 3.1.1. Propiedades del algoritmo EM | 27 |
| 3.2. Representación del algoritmo EM para modelos de espacio-estado . . | 28 |

| | |
|--|-----------|
| 3.3. Aplicación del Algoritmo EM | 31 |
| 3.3.1. Resultados | 34 |
| 4. Algoritmo MCEM | 36 |
| 4.0.2. Muestreador de Gibbs | 37 |
| 4.0.3. Algoritmo Muestreador de Corte | 37 |
| 4.1. Aplicación del algoritmo MCEM | 38 |
| 4.1.1. Resultados | 40 |
| 5. Conclusiones | 43 |
| A. Apéndice. | 44 |
| A.1. Codigos Matlab para la estimación por Máxima Verosimilitud, uti- lizando filtros y suavizado de Kalman | 44 |
| A.1.1. Función que ejecuta el modelo | 44 |
| A.1.2. Función para la simulación del modelo | 46 |
| A.1.3. Filtro de Kalman | 46 |
| A.1.4. Suavizado de Kalman | 47 |
| A.2. Función que calcula los estimadores de Máxima Verosimilitud | 48 |
| A.3. Codigos Matlab para la estimación por Máxima Verosimilitud, uti- lizando el algoritmo MCEM | 49 |
| A.3.1. Función que ejecuta el modelo | 49 |
| A.3.2. Función para el muestreador de corte. | 51 |
| A.3.3. Función Log-Verosimilitud | 52 |
| Referencias bibliográficas. | 53 |

Introducción.

El estudio de los modelos con volatilidad en los mercados financieros y cambiarios ha ido creciendo en las últimas décadas. Para el estudio de la evolución dinámica de la volatilidad en modelos econométricos se han propuesto básicamente dos tipos de modelos: los modelos de heterocedasticidad condicional autorregresiva (modelos ARCH, propuestos por Engle (1982) y modelos GARCH, propuestos por Bollerslev(1986)), y todas las variantes que a partir de ellos han ido surgiendo y los modelos de volatilidad estocástica (modelos SV, propuestos por Taylor (1986)) y sus variantes.

De estos modelos, los más utilizados han sido los modelos de heterocedasticidad condicional autorregresiva, ya que son más sencillos de estimar y están implementados en la mayoría de los software econométricos.

La diferencia entre los modelos ARCH y GARCH y modelos SV es que los modelos ARCH y GARCH no tienen ruidos en la estructura de la volatilidad, en otras palabras, su volatilidad es determinística. Los modelos de volatilidad estocástica son más complicados de estimar puesto que es difícil construir su función de verosimilitud. Sin embargo, siguiendo el trabajo de Kim (1997), es posible obtener estimadores de máxima casi-verosimilitud (es decir, estimadores a una aproximación de la función de verosimilitud).

CAPÍTULO 1

PRELIMINARES.

En este capítulo, presentaremos una serie de definiciones y resultados, que serán la herramienta principal para el desarrollo de este trabajo. Se estudiarán algunas propiedades y definiciones básicas de estadística y probabilidad ([5]), también se presentan algunas propiedades y resultados del Algebra de Matrices ([3]), definiciones del Metodo Monte Carlo y Cadenas de Markov([1]) las cuales son la base principal del método MCEM, el cual describiremos más adelante.

§1.1. Esperanza y Covarianza de Variables Aleatorias

DEFINICIÓN 1.1. Sea X una variable aleatoria discreta, con posibles valores x_1, x_2, \dots, x_n y con función de probabilidad $p(x_i)$, $i = 1, \dots, n$, se define la esperanza como

$$E[X] = \sum_{i=1}^n x_i p(x_i) \quad (1.1)$$

Y para una variable aleatoria continua, con función de densidad $f(x)$, la esperanza se define como

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx \quad (1.2)$$

PROPIEDAD 1.1. Sean X y Y dos variables aleatorias, a y b constantes, entonces:

- $E[aX + b] = aE[X] + b$
- $E[aX + bY] = aE[X] + bE[Y]$
- Si X, Y son independientes entonces $E[XY] = E[X]E[Y]$

DEFINICIÓN 1.2. Sea X una variable aleatoria con valor esperado $E[X] = \mu$, se define la varianza como:

$$Var(X) = E[(X - \mu)^2]$$

Y si desarrollamos el lado derecho de la igualdad, podemos escribir la varianza de la siguiente manera:

$$\text{Var}(X) = E[X^2] - \mu^2$$

PROPIEDAD 1.2. Sea X y Y dos variables aleatorias, a y b constantes, entonces:

- $\text{Var}(a) = 0$
- $\text{Var}(aX + b) = a^2\text{Var}(X)$
- $\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y) \pm 2\text{Cov}(X, Y)$ donde $\text{Cov}(X, Y)$ es la covarianza de X e Y

DEFINICIÓN 1.3. Sean X e Y dos variables aleatorias, la covarianza de X e Y ($\text{Cov}(X, Y)$ o S_{XY}) se define como:

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$$

PROPIEDAD 1.3. Sean X e Y dos variables aleatorias, a y b constantes, entonces:

- $\text{Cov}(X, a) = 0$
- $\text{Cov}(X, X) = \text{Var}(X)$
- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
- $\text{Cov}(aX, bY) = ab\text{Cov}(X, Y)$
- $\text{Cov}(X + a, Y + b) = ab\text{Cov}(X, Y)$
- $\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$

Ahora bien, para los casos n -dimensionales tenemos las siguientes definiciones.

Sea $X = (X_1, \dots, X_n)$ un n -vector estocástico, donde el valor esperado $E[X_i]$ existe para todo i , entonces definimos:

DEFINICIÓN 1.4. El vector esperanza $E(X)$ de un n -vector estocástico $X = (x_1, \dots, x_n)$ se define como el vector $E(X) = (E(X_1), \dots, E(X_n))$.

DEFINICIÓN 1.5. Si X, Y son dos vectores estocásticos, n -dimensionales, entonces podemos definir la matriz de Covarianza de X de la siguiente manera:

$$\Sigma(X) = \begin{pmatrix} Cov(X_1, X_1) & \cdots & Cov(x_1, X_n) \\ Cov(X_2, X_1) & \cdots & Cov(x_2, X_n) \\ \vdots & \cdots & \vdots \\ Cov(X_n, X_1) & \cdots & Cov(X_n, X_n) \end{pmatrix}$$

A causa de que $Cov(x_i, x_j) = Cov(x_j, x_i)$, así la matriz es

Y la matriz de Covarianza de XY , como sigue:

$$\Sigma_{XY} = \begin{pmatrix} Cov(X_1, Y_1) & \cdots & Cov(X_1, Y_n) \\ Cov(X_2, Y_1) & \cdots & Cov(X_2, Y_n) \\ \vdots & \cdots & \vdots \\ Cov(X_n, Y_1) & \cdots & Cov(X_n, Y_n) \end{pmatrix}$$

simétrica.

Las definiciones del caso n -dimensional, conservan las propiedades descritas en (1.1), (1.2) y (1.3)

TEOREMA 1.1. Sean $x = (x_1, \dots, x_n)'$ y $y = (y_1, \dots, y_m)'$ dos vectores aleatorios con distribución normal Gaussiana, y supongamos que el vector $(y' + x)'$ multivariado de tamaño $(n + m) \times 1$ tiene una distribución normal. Entonces

$$E(y|x) = \mu_y + \sigma_{yx} \Sigma_{xx}^{-1} (x - \mu_x) \tag{1.3}$$

$$Var(y|x) = \Sigma_{yy} - \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy} \tag{1.4}$$

donde $\mu_y = E(y)$, es de orden $m \times 1$, $\Sigma_{yy} = Var(y)$ de orden $m \times m$, $\mu_x = E(x)$, de orden $n \times 1$, $\Sigma_{yx} = \Sigma'_{xy} = Cov(y, x)$ es de orden $m \times n$, y $\Sigma_{xx} = Var(x)$ es una matriz no singular de orden $n \times n$

Demostración. Ver apendice (B) en ([7]).

DEFINICIÓN 1.6. (Probabilidad Condicional)

Dados X y Y eventos probabilísticos, la probabilidad condicional de X dado Y es definida siempre y cuando $P(Y) > 0$ por:

$$P(X|Y) = \frac{P(XY)}{P(Y)}. \quad (1.5)$$

Así, si X y Y son variables aleatorias, entonces se define la función de masa de probabilidad condicional de X dado $Y = y$ por:

$$\begin{aligned} P_{(X|Y)}(x|y) &= P(X = x|Y = y) \\ &= \frac{P(X = x, Y = y)}{P(Y = y)} \\ &= \frac{P(x, y)}{P_Y(y)}. \end{aligned}$$

TEOREMA 1.2. (Teorema de Bayes.)

Suponga que se observa un variable aleatoria y , y se desea hacer inferencias acerca de otra variable aleatoria θ , donde θ se extrae de alguna distribución $p(\theta)$. De la definición de probabilidad condicional, se tiene:

$$P(\theta|y) = \frac{P(y, \theta)}{P(y)}. \quad (1.6)$$

§1.2. Estimación por Máxima Verosimilitud

La función de verosimilitud (o, simplemente, verosimilitud) es una función de los parámetros de un modelo estadístico que permite realizar inferencias acerca de su valor a partir de un conjunto de observaciones. La estimación vía Máxima Verosimilitud es una herramienta muy usada para la estimación de parámetros en modelos estadísticos y a continuación, presentaremos una definición de la función Verosimilitud y el Método de Máxima Verosimilitud.

DEFINICIÓN 1.7. Sea X_1, X_2, \dots, X_n una muestra de variables aleatorias, independientes e idénticamente distribuidas, con función de densidad $f_\theta(x_i)$ para cada i , donde θ es el vector de parámetros de la distribución que además es desconocido, Se define la función de Verosimilitud como

$$L(\theta) = L(\theta|X_1, \dots, X_n) = \prod_i f_\theta(x_i)$$

DEFINICIÓN 1.8 (Estimador de Máxima Verosimilitud). *Sea X_1, X_2, \dots, X_n una muestra de variables aleatorias, independientes e idénticamente distribuidas, con función de densidad $f_\theta(x_i)$ para cada i con parámetro θ desconocido, y supongamos que deseamos estimar $\hat{\theta}$ como el valor que está mas próximo al valor verdadero de θ . En este sentido, definimos $\hat{\theta}$ como el valor (o los valores) que maximizan la función de Verosimilitud $L(\theta|X_1, \dots, X_n)$. Este se conoce como Estimador de Máxima Verosimilitud*

§1.3. La Distribución Normal Vectorial

DEFINICIÓN 1.9. *Diremos que un vector estocástico X sigue una distribución normal, con vector esperanza $\mu = E[X]$ y matriz de Covarianza Σ si su función de densidad esta dada por*

$$f_X(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} (X - \mu)' \Sigma^{-1} (X - \mu) \right\}$$

Y para el caso uni-dimensional, diremos que una variable aleatoria x tiene una distribución normal, o distribución gaussiana, con parámetros μ y σ^2 si su función de densidad viene dada por:

$$f_x(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2} \right\}$$

y su grafica se muestra en la figura (1.1).

§1.4. Distribuciones de Probabilidad (Familia Exponencial)

DEFINICIÓN 1.10. *Diremos que una función de distribución F pertenece a una familia exponencial si la función de densidad puede expresarse de la forma*

$$f_{\mathbf{X}}(x; \theta) = b(x) \exp(\eta(\theta)T(x) - A(\theta)) \quad (1.7)$$

donde $b(x), T(x), \eta(\theta)$ y $A(\theta)$ son funciones conocidas. θ se conoce como parámetro de la familia y \mathbf{X} es un vector de observaciones.

PROPOSICIÓN 1.1. *La función de distribución normal con media μ y varianza σ^2 desconocida, pertenece a la familia exponencial*

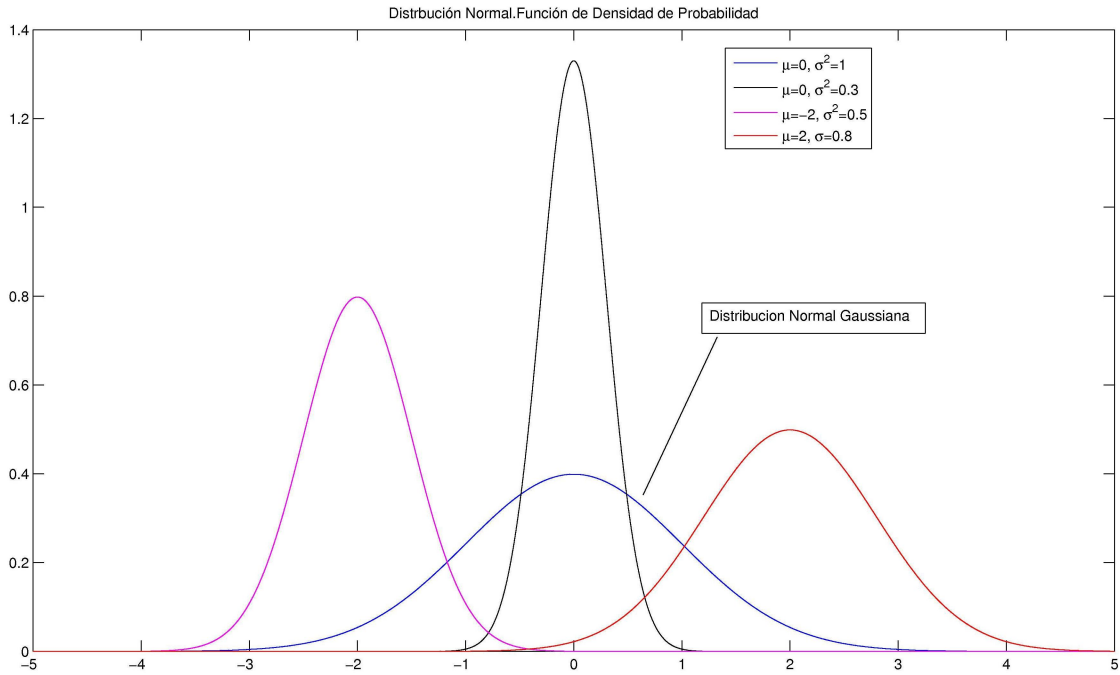


FIGURA 1.1: FUNCIÓN DE DENSIDAD

Demostración:

En efecto, la función de densidad de una distribución normal con media μ y varianza σ^2 viene dada por

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{\mu - x}{\sigma}\right)^2\right) \quad (1.8)$$

Definimos

$$\eta(\theta) = \theta = \left(\frac{\mu}{\sigma^2}, \frac{1}{\sigma^2}\right)'$$

$$b(x) = \frac{1}{\sqrt{2\pi}}, \quad T(x) = \left(x, -\frac{x^2}{2}\right)'$$

y

$$A(\theta) = \frac{\mu^2}{2\sigma^2} - \ln\left(\frac{1}{\sigma}\right)$$

Veamos que la expresión $f(x) = b(x) \exp(\eta(\theta)T(x) - A(\theta))$ se reduce a (1.8).

$$\begin{aligned}
 \theta T(x) &= \frac{\mu x}{\sigma^2} - \frac{x^2}{2\sigma^2} \\
 \Rightarrow \theta T(x) - A(\theta) &= \frac{\mu x}{\sigma^2} - \frac{x^2}{2\sigma^2} - \frac{\mu^2}{2\sigma^2} + \ln\left(\frac{1}{\sigma}\right) \\
 &= \frac{1}{2\sigma^2} (2\mu x - x^2 - \mu^2) + \ln\left(\frac{1}{\sigma}\right) \\
 &= -\frac{1}{2\sigma^2} (\mu^2 - 2\mu x + x^2) + \ln\left(\frac{1}{\sigma}\right) \\
 &= -\frac{1}{2\sigma^2} (\mu - x)^2 + \ln\left(\frac{1}{\sigma}\right)
 \end{aligned}$$

Así,

$$\begin{aligned}
 \exp(\eta(\theta)T(x) - A(\theta)) &= \exp\left(-\frac{1}{2\sigma^2}(\mu - x)^2 + \ln\left(\frac{1}{\sigma}\right)\right) \\
 &= \exp\left(-\frac{1}{2}\left(\frac{\mu - x}{\sigma}\right)^2\right) \exp\left(\ln\left(\frac{1}{\sigma}\right)\right) \\
 &= \exp\left(-\frac{1}{2}\left(\frac{\mu - x}{\sigma}\right)^2\right) \frac{1}{\sigma}
 \end{aligned}$$

Luego,

$$\begin{aligned}
 f(x) &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{\mu - x}{\sigma}\right)^2\right) \frac{1}{\sigma} \\
 &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{\mu - x}{\sigma}\right)^2\right)
 \end{aligned}$$

§1.5. Diferenciación de la traza de una matriz

TEOREMA 1.3. *Sea $F = \{f_i\}$ una matriz de funciones de orden $p \times q$, definida sobre un conjunto S , y $x = (x_1, x_2, \dots, x_m)$ un vector de m variables. Entonces, para algún punto interior c de S en el cual F es continuamente diferenciable, se tiene que $\text{tr}(F)$ es continuamente diferenciable y*

$$\frac{\partial \text{tr}(F)}{\partial x_j} = \text{tr}\left(\frac{\partial F}{\partial x_j}\right)$$

COROLARIO 1.1. *Sea $F = \{f_i\}$ una matriz de funciones de orden $p \times q$, definida sobre un conjunto S , $x = (x_1, x_2, \dots, x_m)$ un vector de m variables y A es una matriz*

constante, o que no varia respecto a x_j . Entonces, para algun punto interior c de S en el cual F es continuamente diferenciable se tiene que $\text{tr}(AF)$ ó $\text{tr}(FA)$ es continuamente diferenciable y

$$\frac{\partial \text{tr}(AF)}{\partial x_j} = \frac{\partial \text{tr}(FA)}{\partial x_j} = \text{tr} \left(\frac{\partial F}{\partial x_j} A \right) = \text{tr} \left(A \frac{\partial F}{\partial x_j} \right)$$

En particular al resultado anterior, si consideramos $X = \{x_{ij}\}$ una matriz de variables independientes, de orden $(m \times n)$ y $A = \{a_{ij}\}$ una matriz constante de orden $n \times m$, entonces tenemos que

$$\frac{\partial \text{tr}(AX)}{\partial x_{ij}} = \frac{\partial \text{tr}(XA)}{\partial x_{ij}} = a_{ij}$$

TEOREMA 1.4. Sea $X = \{x_{ij}\}$ una matriz de orden $m \times m$ definida sobre un conjunto S , continuamente diferenciable en un punto $c \in \text{int}(S)$, entonces, $\log(\det(X))$ es continuamente diferenciable en c , y además,

$$\frac{\partial \log(\det(X))}{\partial x_{ij}} = y_{ji}$$

donde y_{ji} es el ji -elemento de X^{-1} , o equivalentemente el ij -elemento de $(X^{-1})'$.

Ademas, este resultado puede ser reescrito como:

$$\frac{\partial \log(\det(X))}{\partial x_{ij}} = (X^{-1})'$$

Sea $X = \{x_{ij}\}$ una matriz de orden $m \times n$ de variables independientes, entonces, para cualquier matriz A constante de orden $m \times m$ se tiene

$$\frac{\partial X'AX}{\partial X} = (A + A')X$$

■

§1.6. Métodos Monte Carlo

El método Monte Carlo es un método no determinístico o estadístico numérico, usado para aproximar expresiones matemáticas complejas y costosas de evaluar con exactitud. El método se llamó así en referencia al Casino de Monte Carlo (Principe de

Mónaco) por ser “la capital del juego de azar”, al ser la ruleta un generador simple de números aleatorios. La invención, el nombre y el desarrollo sistemático de los métodos de Monte Carlo datan aproximadamente de 1944 y se les atribuye a Stanislaw Ulam y a John Von Neumann, y posteriormente fueron mejorados enormemente con el desarrollo de la computadora.

§1.6.1. Integración Monte Carlo

La integración, en general, es muy útil en el cálculo de probabilidades y esperanzas, claro está que cuando se quiere calcular una esperanza, lo primero que intentamos es resolver la integral de forma analítica. Cuando la evaluación analítica o el cálculo de la integral es casi imposible o muy complicada, la integración numérica es una opción, y los métodos Monte Carlo proponen una forma de integración numérica.

Consideremos ξ_1, ξ_2, \dots una sucesión de variables aleatorias, idénticamente distribuidas, con distribución de probabilidad π , y f una función. entonces el estimador

$$\hat{\pi}(f) = N^{-1} \sum_{i=1}^N f(\xi_i) \quad (1.9)$$

converge a $\pi(f)$ para toda función f π -integrable, con N suficientemente grande.

§1.7. Algoritmo Metropolis-Hasting

§1.7.1. Cadena de Markov

DEFINICIÓN 1.11. *Se define como un proceso estocástico que cumple con la propiedad de Márkov, es decir, si se conoce la historia del sistema hasta su instante actual, su estado presente resume toda la información relevante para describir en probabilidad su estado futuro.*

Una cadena de Markov es una secuencia X_1, X_2, X_3, \dots de variables aleatorias. El rango de estas variables, es llamado espacio-estado, el valor de X_n es el estado del proceso en el tiempo n . Si la distribución de probabilidad condicional de X_{n+1} en estados pasados es una función de X_n por sí sola, entonces:

$$P(X_{n+1} = j | X_0 = i_0, X_1 = i_1, \dots, X_n = i) = P(X_{n+1} = j | X_n = i)$$

a esta identidad se le conoce como propiedad de Markov

- Si $X_n = i$, diremos que la cadena de Markov está en el estado i a tiempo n o en el paso n .
- $P(X_{n+1} = j|X_n = i)$ se denomina probabilidad de transición.
- Si $P(X_{n+1} = j|X_n = i)$ es independiente de n , entonces diremos que la cadena de Markov posee probabilidades de transición estacionaria.

DEFINICIÓN 1.12. Una cadena de Markov se dice homogénea si la probabilidad de ir del estado i al estado j no depende del tiempo en el que se encuentre la cadena, es decir:

$$P(X_n = j|X_{n-1} = i) = P(X_1 = j|X_0 = i)$$

para todo n y para todo i, j . Ahora, si para alguna pareja de espacios y para algún tiempo n esta propiedad no se cumple, entonces diremos que la cadena de Markov es no homogénea

DEFINICIÓN 1.13. Sea $\{X_n; n \geq 0\}$ una cadena de Markov homogénea con espacio de estados $E = \{0, 1, 2, \dots\}$, entonces $p_{ij} = P(X_{n+1} = j|X_n = i)$, $i \geq 0$, $j \geq 0$ es independiente de n . La Matriz de transición de $\{X_n; n \geq 0\}$ se define como

$$P = [p_{ij}] = \begin{pmatrix} p_{00} & p_{01} & p_{02} & \cdots \\ p_{10} & p_{11} & p_{12} & \cdots \\ p_{20} & p_{21} & p_{22} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

donde $p_{ij} \geq 0$, $\sum_{j=1}^{\infty} p_{ij} = 1$, $i = 0, 1, 2, \dots$

§1.7.2. Ecuación de Chapman-Kolmogorov

Una caracterización importante de las Cadenas de Markov es que la distribución de probabilidad de $\{X_n; n \geq 0\}$ se puede obtener a través de calculos matriciales.

DEFINICIÓN 1.14. Sea $P = [p_{ij}]$ la matriz de transición de una Cadena de Markov $\{X_n; n \geq 0\}$. Se define la matriz potencia como

$$P^2 = PP$$

con elementos (i, j) dados por

$$p_{ij}^{(2)} = \sum_k p_{ik} p_{kj}$$

Observe que cuando el espacio de estado E es finito la serie $\sum_k^{\infty} p_{ik} p_{kj}$ converge, ya que

$$\sum_k p_{ik} p_{kj} \leq \sum_k p_{ik} = 1$$

De igual forma $P^3 = PP^2$ con elementos (i, j)

$$p_{ij}^{(3)} = \sum_k p_{ik} p_{kj}^{(2)}$$

Y en general $P^{n+1} = PP^n$ tiene elementos (i, j)

$$p_{ij}^{(n+1)} = \sum_k p_{ik} p_{kj}^{(n)}$$

Notación: $P^0 = I$ la matriz identidad

DEFINICIÓN 1.15. La probabilidad de transición del paso n para una Cadena de Markov homogénea $\{X_n; n \geq 0\}$ está definida como

$$P(X_{n+1} = j | X_0 = i)$$

DEFINICIÓN 1.16. Sea $\{X_n; n \geq 0\}$ una Cadena de Markov homogénea, con matriz de transición P , se define la ecuación Chapman-Kolmogorov como

$$P^{m+n} = P^n P^m$$

§1.7.3. Distribuciones Límites

DEFINICIÓN 1.17. Una Cadena de Markov es llamada regular, si existe un entero positivo m tal que después de m pasos, cada estado tiene probabilidad no nula, sin importar el estado inicial.

Notación: $A > 0$ denotara que cada elemento a_{ij} de A satisface que $a_{ij} > 0$

Entonces para una Cadena de Markov regular con matriz de transición P , existe $m > 0$ tal que $P^m > 0$

TEOREMA 1.5. *Sea $\{X_n; n \geq 0\}$ una Cadena de Markov homogénea regular con estados finitos y matriz de transición P . Entonces*

$$\lim_{n \rightarrow \infty} P^n = \widehat{P}$$

donde \widehat{P} es la matriz donde las filas son igual a la distribución estacionaria \widehat{p} para la Cadena de Markov

§1.7.4. Algoritmo Metropolis Hasting

Como consecuencia de la aplicación de la integración Monte Carlo, supongamos que queremos simular una sucesión de variables aleatorias $\{\xi_k\}_{k \geq 0}$ con función de probabilidad π compleja, y que además que cumpla (1.9). El algoritmo Metropolis Hasting genera cadenas de Markov con función de probabilidad π , y que satisface (1.9) para alguna función f .

Este algoritmo se describe de la siguiente manera: Dado ξ^i ,

- Generamos $\xi \sim r(\xi^i, \cdot)$;
- Seleccionamos

$$\xi^{i+1} = \begin{cases} \xi, & \text{con probabilidad } \alpha(\xi^i, \xi) = \frac{\pi(\xi)r(\xi, \xi^i)}{\pi(\xi^i)r(\xi^i, \xi)} \\ \xi^i, & \text{en otros casos} \end{cases} \quad (1.10)$$

Donde el valor ξ^1 es tomado arbitrariamente, y $r(\cdot, \cdot)$ es una función propuesta. Esta función r proviene del hecho de proponer una matriz de probabilidad de transición de Markov, de modo que $r(\xi^n, \xi^m)$ es la probabilidad de ir del estado ξ^n al estado ξ^m .

En la práctica, basta considerar un radio aceptación, esto es, al generar ξ en el primer paso en una iteración i , generamos $U \sim ([0, 1])$, y aceptamos ξ solo si $U \leq \alpha(\xi^i, \xi)$.

CAPÍTULO 2

FILTROS Y SUAVIZADO DE KALMAN.

§2.1. Modelos de Estado-Espacio

Los modelos estado-espacio o modelo lineal dinámico, fueron introducidos por Kalman (1960) y Kalman y Bucy (1961), representando un gran interés al igual que la regresión lineal, aunque el modelo originalmente se introdujo como un método principalmente para ser usado en la investigación relacionada con el sector aeroespacial, se ha aplicado para el modelado de datos de la economía. Un excelente tratamiento moderno de análisis de series temporales basado en el modelo de espacio de estado es el texto por Shumway y Stoffer [6].

El modelo de estado-espacio o modelo lineal dinámico (DLM), en su forma básica, contiene un vector autorregresión de primer orden como la ecuación de estado.

$$x_t = \Phi x_{t-1} + w_t \quad (2.1)$$

Donde la ecuación de estado determina la regla para la generación del vector de estado x_t del pasado x_{t-1} , para los puntos a tiempo $t = 1, \dots, n$.

Supondremos que w_t , con $t = 1, \dots, n$ son vectores aleatorios normales independientes idénticamente distribuidos con media cero y matriz de covarianza Q . En el Modelo Lineal Dinámico supondremos que x_0 es un vector aleatorio normal con media μ_0 y matriz de covarianza Σ_0 de orden $p \times p$. El Modelo Lineal Dinámico posee una ecuación adicional denominada ecuación de observaciones, en el caso en que el vector de estado no sea observable directamente, sino a través de una transformación digamos

$$y_t = M_t x_t + v_t \quad (2.2)$$

Donde M_t es una matriz de orden $q \times p$, y_t es un vector de observación de orden

$q \times 1$ y v_t es el vector normal con media cero y matriz de covarianza R de orden $q \times q$.

Supondremos que $\{w_t\}$ y $\{v_t\}$ son no correlacionados.

§2.2. Filtros de Kalman

En la práctica el principal objetivo en el análisis de un modelo lineal dinámico es producir estimadores para los vectores no observados x_t a partir de las observaciones $Y_r = (y_1, \dots, y_r)$ al tiempo r .

Cuando $r < t$: El problema es denominado Predicción.

Cuando $r = t$: El problema es denominado Filtrado.

Cuando $r > t$: El problema es denominado Suavizado.

A lo largo de este capítulo utilizaremos las siguientes notaciones:

$$x_t^r = E(x_t | Y_r) \quad (2.3)$$

y

$$P_{t_1, t_2}^r = Cov(x_{t_1}, x_{t_2} | Y_r) = E\{(x_{t_1} - x_{t_1}^r)(x_{t_2} - x_{t_2}^r)^T\} \quad (2.4)$$

Cuando $t_1 = t_2 = t$, la ecuación definida en (2.4) nos queda.

$$P_t^r = Cov(x_t | Y_r) = E\{(x_t - x_t^r)(x_t - x_t^r)^T\} \quad (2.5)$$

El Filtro de Kalman nos da la ecuación de filtros y predicción, el nombre filtro viene por el hecho que x_t^t es un filtro lineal de las observaciones y_1, y_2, \dots, y_t es decir:

$$x_t^t = \sum_{r=1}^t B_r y_r \quad (2.6)$$

Para matrices adecuadamente elegidas B_r de orden $p \times q$.

La ventaja del filtro de Kalman es que éste especifica como actualizar el filtro desde x_{t-1}^{t-1} a x_t^t , a partir de y_t sin tener que procesar toda los datos y_1, y_2, \dots, y_t .

PROPIEDAD 2.1. (El Filtro de Kalman.)

Para el modelo lineal dinámico dado por las ecuaciones (2.1) y (2.2) con condiciones iniciales $x_0^0 = \mu_0$ y $P_0^0 = \sum_0$ para $t = 1, \dots, n$ se tiene

$$x_{t-1}^t = \Phi x_{t-1}^{t-1} \quad (2.7)$$

$$P_{t-1}^t = \Phi P_{t-1}^{t-1} \Phi' + Q \quad (2.8)$$

con

$$x_t^t = x_t^{t-1} + (y_t - M_t x_t^{t-1}) \quad (2.9)$$

$$P_t^t = [I - K_t M_t] P_t^{t-1} \quad (2.10)$$

Donde

$$K_t = P_t^{t-1} M_t' [M_t P_t^{t-1} M_t' + R]^{-1} \quad (2.11)$$

Es denominada la ganancia de kalman.

Demostración. Haciendo uso de algunas definiciones y cálculos ya vistos tenemos que

$$\begin{aligned} x_t^{t-1} &= E(x_t | Y_{t-1}) \\ &= E(\Phi s_{t-1} + w_t | Y_{t-1}) \\ &= E(\Phi s_{t-1} | Y_{t-1}) + E(w_t | Y_{t-1}) \\ &= \Phi E(s_{t-1} | Y_{t-1}) \\ x_t^{t-1} &= \Phi x_{t-1}^{t-1} \end{aligned}$$

luego tenemos que

$$P_t^{t-1} = E\{(x_t - s_t^{t-1})(x_t - x_t^{t-1})^T\} \quad (2.12)$$

$$= E\{\Phi x_{t-1} + w_t - \Phi x_{t-1}^{t-1}\}(\Phi x_{t-1} + w_t - \Phi x_{t-1}^{t-1})'$$

$$P_t^{t-1} = E\{\Phi(x_{t-1} - x_{t-1}^{t-1}) + w_t\}(\Phi(x_{t-1} - x_{t-1}^{t-1}) + w_t)'$$

$$P_t^{t-1} = E\{\Phi(x_{t-1} - x_{t-1}^{t-1})(x_{t-1} - x_{t-1}^{t-1})'\Phi'\} + E\{\Phi(x_{t-1} - x_{t-1}^{t-1})w_t'\}$$

$$+ E\{w_t(x_{t-1} - x_{t-1}^{t-1})'\Phi'\} + E\{w_t w_t'\}$$

$$= \Phi P_{t-1}^{t-1} \Phi' + \Phi E\{(x_{t-1} - x_{t-1}^{t-1})w_t'\} + E\{w_t(x_{t-1} - x_{t-1}^{t-1})'\}\Phi'$$

$$+ E\{w_t w_t'\}$$

como $w_t \sim N(0, Q)$ con $t = 1, \dots, n$, siendo ellas variables aleatorias independientes e idénticamente distribuidas tenemos que

$$P_t^{t-1} = \Phi P_{t-1}^{t-1} \Phi^T + Q$$

Luego para obtener la ecuación (2.9) primero definimos las innovaciones como

$$\begin{aligned} \varepsilon_t &= y_t - E(y_t | Y_{t-1}) \\ &= y_t - E(M_t s_t + v_t | Y_{t-1}) \\ &= y_t - E(M_t s_t | Y_{t-1}) - E(v_t | Y_{t-1}) \\ &= y_t - M_t E(s_t | Y_{t-1}) \end{aligned}$$

Así,

$$\varepsilon_t = y_t - M_t x_t^{t-1} \tag{2.13}$$

Para $t = 1, \dots, n$ se tiene que

$$\begin{aligned} E(\varepsilon_t) &= E(y_t - E(y_t | Y_{t-1})) \\ &= E(y_t) - E(E(y_t | Y_{t-1})) \\ &= E(y_t) - E(y_t) \\ \Rightarrow E(\varepsilon_t) &= 0 \end{aligned}$$

$$\begin{aligned} \Sigma_t &= Var(\varepsilon_t) \\ &= Var(y_t - M_t s_t^{t-1}) \\ &= Var(M_t x_t + v_t - M_t s_t^{t-1}) \\ &= Var(M_t(x_t - x_t^{t-1}) + v_t) \\ &= Cov[(M_t(x_t - x_t^{t-1}) + v_t), (M_t(x_t - x_t^{t-1}) + v_t)'] \\ &= Cov[M_t(x_t - s_t^{t-1}), (x_t - x_t^{t-1})^T M_t^T] + Cov[M_t(x_t - x_t^{t-1}), v_t'] \\ &\quad + Cov[v_t, (x_t - x_t^{t-1})M_t'] + Cov[v_t, v_t'] \end{aligned}$$

Además, puesto que v_t son variables independiente e idénticamente distribuidos con $v_t \sim N(0, R)$ y usando las definiciones y propiedades de covarianza y esperanza condicional nos queda que

$$\Sigma_t = M_t P_t^{t-1} M_t' + R$$

Donde, $E(\varepsilon_t, y'_r) = 0$ para $r < t$.

En vista de que la sucesión de innovación es un proceso gaussiano, esto implica que las innovaciones son independientes de las observaciones (observaciones pasadas). Además la condición de covarianza entre x_t y ε_t , dado Y_{t-1} es:

$$\begin{aligned}
 Cov(x_t, \varepsilon_t | Y_{t-1}) &= Cov(x_t, y_t - M_t x_t^{t-1} | Y_{t-1}) \\
 &= Cov(x_t, y_t - M_t x_t^{t-1} | Y_{t-1}) - Cov(x_t^{t-1}, y_t - M_t x_t^{t-1} | Y_{t-1}) \\
 &= Cov((x_t - x_t^{t-1}), y_t - M_t x_t^{t-1} | Y_{t-1}) \\
 &= Cov((x_t - x_t^{t-1}), M_t(x_t - x_t^{t-1}) + v_t) \\
 &= Cov((x_t - x_t^{t-1}), M_t(x_t - x_t^{t-1})) + Cov((x_t - x_t^{t-1}), v_t) \\
 &= E\{(x_t - x_t^{t-1})(M_t(x_t - x_t^{t-1}))'\} - E(x_t - x_t^{t-1})M_t E(x_t - x_t^{t-1}) \\
 &+ E\{(x_t - x_t^{t-1})(v_t)'\} - E(x_t - x_t^{t-1})E(v_t) \\
 &= E\{(x_t - x_t^{t-1})(x_t - x_t^{t-1})'\}M_t' \\
 \Rightarrow Cov(x_t, \varepsilon_t | Y_{t-1}) &= P_t^{t-1}M_t'
 \end{aligned}$$

Del uso de estos resultados tenemos que la distribución condicional conjunta de s_t y ε_t dado y_{t-1} es normal.

$$X \sim N(\mu, \sigma^2)$$

con

$$X = \begin{pmatrix} x_t \\ \varepsilon_t \end{pmatrix}$$

Además como

$$\begin{pmatrix} x_t \\ \varepsilon_t \end{pmatrix} | Y_{t-1} \sim N \left[\begin{pmatrix} E(x_t) \\ E(\varepsilon_t) \end{pmatrix}, \begin{pmatrix} Var(x_t) & Cov(x_t, \varepsilon_t) \\ Cov(\varepsilon_t, x_t) & Var(\varepsilon_t) \end{pmatrix} \right]$$

En conclusión nos queda que:

$$\begin{pmatrix} x_t \\ \varepsilon_t \end{pmatrix} | Y_{t-1} \sim N \left[\begin{pmatrix} x_t^{t-1} \\ 0 \end{pmatrix}, \begin{pmatrix} P_t^{t-1} & P_t^{t-1}M_t^T \\ M_t^T P_t^{t-1} & \Sigma_t \end{pmatrix} \right] \quad (2.14)$$

Por otra parte calculemos x_t^t

$$\begin{aligned} x_t^t &= E(x_t|y_1, \dots, y_{t-1}, y_t) \\ &= E(x_t|y_{t-1}, y_t) \end{aligned}$$

como $\varepsilon_t = y_t - M_t x_t^{t-1}$ entonces despejando tenemos que $y_t = \varepsilon_t + M_t x_t^{t-1}$

Así nos queda que

$$x_t^t = E(x_t|y_{t-1}, \varepsilon_t + M_t x_t^{t-1})$$

Así usando la siguiente definición de $E(x|y)$ en x_t^t

$$E(x|y) := \mu_x + \Sigma_{xy} \Sigma_{yy}^{-1} (y - \mu_y)$$

$$x_t^t = E(x_t|y_{t-1}) + \Sigma_{x_t, \varepsilon_t + M_t x_t^{t-1}} \Sigma_{\varepsilon_t + M_t x_t^{t-1}}^{-1} [\varepsilon_t + M_t x_t^{t-1} - E(\varepsilon_t + M_t x_t^{t-1})]$$

Pero

$$\begin{aligned} \Sigma_{x_t, \varepsilon_t + M_t x_t^{t-1}} &= \text{cov}(s_t, \varepsilon_t + M_t x_t^{t-1} | Y_{t-1}) \\ &= \text{cov}(x_t, \varepsilon_t | Y_{t-1}) + \text{cov}(x_t, M_t x_t^{t-1} | Y_{t-1}) \\ &= P_t^{t-1} M_t' + E[(x_t - E(x_t | Y_{t-1})) (M_t x_t^{t-1} - E(M_t x_t^{t-1} | Y_{t-1}))] \\ &= P_t^{t-1} M_t' + E[(x_t - x_t^{t-1}) (M_t x_t^{t-1} - M_t E(x_t^{t-1} | Y_{t-1}))] \\ &= P_t^{t-1} M_t' + E[(x_t - x_t^{t-1}) (M_t x_t^{t-1} - M_t x_t^{t-1})] \\ &= P_t^{t-1} M_t' \end{aligned}$$

$$\begin{aligned} \Sigma_{\varepsilon_t + M_t x_t^{t-1}} &= \text{cov}(\varepsilon_t + M_t x_t^{t-1}, \varepsilon_t + M_t x_t^{t-1} | Y_{t-1}) \\ &= \text{var}(\varepsilon_t + M_t x_t^{t-1} | Y_{t-1}) \\ &= \text{var}(\varepsilon_t | Y_t) \\ &= M_t P_t^{t-1} M_t' + R \end{aligned}$$

Así de esto y de la definición y propiedades de esperanza, esperanza condicional y puesto que $E(\varepsilon_t) = 0$ tenemos que

$$x_t^t = x_t^{t-1} + (P_t^{t-1} M_t') (M_t P_t^{t-1} M_t' + R)^{-1} [\varepsilon_t + M_t x_t^{t-1} - E(\varepsilon_t) - M_t' E(x_t^{t-1})]$$

$$s_t^t = s_t^{t-1} + (P_t^{t-1} M_t') (M_t P_t^{t-1} M_t' + R)^{-1} \varepsilon_t$$

A partir de la ecuación (2.14) y utilizando la definición $var(X|Y) = \sum_{xx} - \sum_{xy} \sum_{yy}^{-1} \sum_{yx}$ podemos calcular a P_t^t , esto es:

$$\begin{aligned} P_t^t &= Cov(s_t|y_{t-1}, y_t) \\ P_t^t &= Cov(s_t|y_{t-1}, \varepsilon_t + M_t x_t^{t-1}) \\ P_t^t &= \Sigma_{x_t, x_t|y_{t-1}} - \Sigma_{x_t, y_t|y_{t-1}} \Sigma_{y_t, y_t|y_{t-1}}^{-1} \Sigma_{y_t, x_t|y_{t-1}} \end{aligned}$$

Recordando que $y_t = \varepsilon_t + \tilde{A}_t x_t^{t-1}$ y por (2.14)

$$\begin{aligned} P_t^t &= Var(x_t|y_{t-1}) - (P_t^{t-1} M_t')(M_t P_t^{t-1} M_t' + R)^{-1} (M_t P_t^{t-1}) \\ P_t^t &= P_t^{t-1} - (P_t^{t-1} M_t')(M_t P_t^{t-1} M_t' + R)^{-1} (M_t P_t^{t-1}) \end{aligned}$$

PROPIEDAD 2.2. (Kalman Suavizado.)

Para el modelo espacio-estado definido en (2.1) y (2.2) con condiciones iniciales x_n^n y P_n^n obtenidas a travez de la propiedad (2.1) del Filtro de kalman, para $t = n, n-1, \dots, 1$ se tiene

$$x_{t-1}^n = x_{t-1}^{t-1} + J_{t-1}(x_t^n - x_t^{t-1}) \tag{2.15}$$

$$P_{t-1}^n = P_{t-1}^{t-1} + J_{t-1}(P_t^n - P_t^{t-1})J_{t-1}' \tag{2.16}$$

Donde

$$J_{t-1} = P_{t-1}^{t-1} \Phi' [P_t^{t-1}]^{-1} \tag{2.17}$$

Demostración. Para $1 \leq t \leq n$ definimos $Y_{t-1} = \{y_1, \dots, y_{t-1}\}$ y $\eta_t = \{v_1, \dots, v_n, w_1, \dots, w_n\}$, con Y_0 vacio.

Sea

$$q_{t-1} = E(x_{t-1}|Y_{t-1}, x_t - x_t^{t-1}, \eta_t) \tag{2.18}$$

Ahora bien, puesto que $Y_{t-1}, (x_t - x_t^{t-1}), \eta_t$ son mutuamente independientes y x_{t-1} y η_t son independientes, usando la definición $E(x|y) = \mu_x + \Sigma_{xy} \Sigma_{yy}^{-1} (y - \mu_y)$ en (2.18) se tiene que:

$$\begin{aligned} q_{t-1} &= E(x_{t-1}|Y_{t-1}, \eta_t) + \Sigma_{x_{t-1}, (x_t - x_t^{t-1})|Y_{t-1}, \eta_t} \Sigma_{(x_t - x_t^{t-1}), (x_t - x_t^{t-1})|Y_{t-1}, \eta_t}^{-1} [x_t - x_t^{t-1} \\ &\quad - E(x_t - x_t^{t-1}|Y_{t-1}, \eta_t)] \\ &= E(x_{t-1}|Y_{t-1}) + \Sigma_{x_{t-1}, \eta_t|Y_{t-1}} \Sigma_{\eta_t \eta_t|Y_{t-1}}^{-1} (\eta_t - E(\eta_t)) \\ &\quad + \Sigma_{x_{t-1}, (x_t - x_t^{t-1})|Y_{t-1}, \eta_t} \Sigma_{(x_t - x_t^{t-1}), (x_t - x_t^{t-1})|Y_{t-1}, \eta_t}^{-1} [(x_t - x_t^{t-1}) \\ &\quad - E(x_t - x_t^{t-1}|Y_{t-1}, \eta_t)] \end{aligned} \tag{2.19}$$

Por otro lado, tenemos que

$$E(x_{t-1}|Y_{t-1}) = x_{t-1}^{t-1} \quad \text{Por (2,3)}$$

Además, como x_{t-1} y η_t son independientes, se tiene que $Cov(x_{t-1}, \eta_t) = 0$ así, $Cov(x_{t-1}, \eta_t|Y_{t-1}) = 0$, por lo tanto

$$\Sigma_{x_{t-1}, \eta_t|Y_{t-1}} \Sigma_{\eta_t \eta_t|Y_{t-1}} (\eta_t - E(\eta_t)) = 0$$

Ahora bien,

$$\begin{aligned} \Sigma_{x_{t-1}, (x_t - x_t^{t-1})|Y_{t-1}, \eta_t} &= Cov(x_{t-1}, (x_t - x_t^{t-1})|Y_{t-1}, \eta_t) & (2.20) \\ &= E[(x_{t-1} - E(x_{t-1})|Y_{t-1}, \eta_t))(x_t - x_t^{t-1} - E(x_t - x_t^{t-1})|Y_{t-1}, \eta_t)'] \\ &= E[(x_{t-1} - x_{t-1}^{t-1})(x_t - x_t^{t-1} - (E(x_t|Y_{t-1}, \eta_t) - E(x_t^{t-1}|Y_{t-1}, \eta_t)))'] \\ &= E[(x_{t-1} - x_{t-1}^{t-1})(x_t - x_t^{t-1} - x_t^{t-1} + x_t^{t-1})'] \\ &= E[(x_{t-1} - x_{t-1}^{t-1})(x_t - x_t^{t-1})'] \\ &= E[(x_{t-1} - x_{t-1}^{t-1})(\Phi x_{t-1} + w_t - \Phi x_{t-1}^{t-1})'] \quad \text{Por(2,1 y 2,7)} \\ &= E[(x_{t-1} - x_{t-1}^{t-1})(\phi(x_{t-1} - x_{t-1}^{t-1}))' + E[(x_{t-1} - x_{t-1}^{t-1})w_t']] \\ &= E[(x_{t-1} - x_{t-1}^{t-1})(x_{t-1} - x_{t-1}^{t-1})'\Phi' \quad (E(w_t) = 0) \\ &= P_{t-1}^{t-1}\Phi' \quad \text{Por(2,5)} \end{aligned}$$

Por otra parte,

$$\begin{aligned} E(x_t - x_t^{t-1}|Y_{t-1}, \eta_t) &= E(x_t|Y_{t-1}, \eta_t) - E(x_t^{t-1}|Y_{t-1}, \eta_t) \\ &= x_t^{t-1} - x_t^{t-1} = 0 \end{aligned}$$

Y finalmente,

$$\begin{aligned} \Sigma_{(x_t - x_t^{t-1}), (x_t - x_t^{t-1})|Y_{t-1}, \eta_t} &= Cov(x_t - x_t^{t-1}|Y_{t-1}, \eta_t) \\ &= E((x_t - x_t^{t-1})(x_t - x_t^{t-1})') \\ &= E((\Phi x_{t-1} + w_t - \Phi x_{t-1}^{t-1})(\Phi x_{t-1} + w_t - \Phi x_{t-1}^{t-1})') \\ &= E[(\Phi(x_{t-1} - x_{t-1}^{t-1}) + w_t)(\Phi(x_{t-1} - x_{t-1}^{t-1}) + w_t)'] \\ &= \Phi P_{t-1}^{t-1}\Phi' + Q \quad \text{Por(2,5)} \\ &= P_t^{t-1} \quad \text{Por(2,8)} \end{aligned}$$

Luego, sustituyendo estos resultados en (2.19) tenemos que:

$$q_{t-1} = x_{t-1}^{t-1} + J_{t-1}(x_t - x_t^{t-1})$$

Donde $J_{t-1} = P_{t-1}^{t-1} \Phi' [P_t^{t-1}]^{-1}$

Sabiendo que $Y_{t-1}, (x_t - x_t^{t-1})$ y η_t generan $Y_n = \{y_1, \dots, y_n\}$, entonces

$$\begin{aligned} x_{t-1}^n &= E[x_{t-1}|Y_n] \quad \text{Por (2,3)} \\ &= E[q_{t-1}|Y_n] \\ &= x_{t-1}^{t-1} + J_{t-1}(x_t^n - x_t^{t-1}) \end{aligned} \quad (2.21)$$

Ahora bien, para la recursión de la covarianza, multiplicamos por menos uno y sumamos x_{t-1} en ambos lados de la ecuación (2.21), para obtener

$$\begin{aligned} x_{t-1}^n &= x_{t-1}^{t-1} + J_{t-1}(x_t^n - x_t^{t-1}) \\ \Rightarrow x_{t-1} - x_{t-1}^n &= x_{t-1} - x_{t-1}^{t-1} - J_{t-1}(x_t^n - \Phi x_{t-1}^{t-1}) \\ \Rightarrow x_{t-1} - x_{t-1}^n &= x_{t-1} - x_{t-1}^{t-1} - J_{t-1}x_t^n + J_{t-1}\Phi x_{t-1}^{t-1} \\ \Rightarrow (x_{t-1} - x_{t-1}^n) + J_{t-1}x_t^n &= (x_{t-1} - x_{t-1}^{t-1}) + J_{t-1}\Phi x_{t-1}^{t-1} \end{aligned}$$

Luego, multiplicando a cada lado de la ecuación anterior, por la transposición de si misma se tiene que:

$$\begin{aligned} &[(x_{t-1} - x_{t-1}^n) + J_{t-1}x_t^n][(x_{t-1} - x_{t-1}^n)' + (x_t^n)'J_{t-1}'] \\ &= [(x_{t-1} - x_{t-1}^{t-1}) + J_{t-1}\Phi x_{t-1}^{t-1}][(x_{t-1} - x_{t-1}^{t-1})' + (x_{t-1}^{t-1})'\Phi'J_{t-1}'] \end{aligned}$$

Multiplicando y aplicando esperanza de un lado de la igualdad, obtenemos

$$\begin{aligned} &E[(x_{t-1} - x_{t-1}^n)(x_{t-1} - x_{t-1}^n)'] + E[(x_{t-1} - x_{t-1}^n)(x_t^n)'(J_{t-1})'] \\ &+ E[J_{t-1}x_t^n(x_{t-1} - x_{t-1}^n)'] + E[J_{t-1}x_t^n(x_t^n)'(J_{t-1})'] \\ &= p_{t-1}^n + J_{t-1}E[x_t^n(x_t^n)'](J_{t-1})' \quad (I) \end{aligned}$$

Aplicando el mismo procedimiento del otro lado de la igualdad,

$$\begin{aligned} &E[(x_{t-1} - x_{t-1}^{t-1})(x_{t-1} - x_{t-1}^{t-1})'] + E[(x_{t-1} - x_{t-1}^{t-1})(x_{t-1}^{t-1})'\Phi'J_{t-1}'] \\ &+ E[J_{t-1}\Phi x_{t-1}^{t-1}(x_{t-1} - x_{t-1}^{t-1})'] + E[J_{t-1}\Phi x_{t-1}^{t-1}(x_{t-1}^{t-1})'\Phi'J_{t-1}'] \\ &= P_{t-1}^{t-1} + J_{t-1}\Phi E[x_{t-1}^{t-1}(x_{t-1}^{t-1})']\Phi'J_{t-1} \quad (II) \end{aligned}$$

Sustituyendo (I) y (II) en la ecuación (2.22), tenemos que:

$$\begin{aligned} p_{t-1}^n + J_{t-1}E[x_t^n(x_t^n)'](J_{t-1})' &= P_{t-1}^{t-1} + J_{t-1}\Phi E[x_{t-1}^{t-1}(x_{t-1}^{t-1})']\Phi' J_{t-1}' \\ \Rightarrow p_{t-1}^n &= P_{t-1}^{t-1} + J_{t-1}\Phi E[x_{t-1}^{t-1}(x_{t-1}^{t-1})']\Phi' J_{t-1}' - J_{t-1}E[x_t^n(x_t^n)'](J_{t-1})' \\ \Rightarrow p_{t-1}^n &= P_{t-1}^{t-1} + J_{t-1}[\Phi E[x_{t-1}^{t-1}(x_{t-1}^{t-1})']\Phi' - E[x_t^n(x_t^n)']](J_{t-1})' \end{aligned}$$

usando el hecho de que:

$$E[x_{t-1}^{t-1}(x_{t-1}^{t-1})'] = E[x_{t-1}(x_{t-1})'] - P_{t-1}^{t-1}$$

y

$$E[x_t^n(x_t^n)'] = E[x_t(x_t)'] - P_t^n = \Phi E[x_{t-1}(x_{t-1})']\Phi' + QP_t^n$$

obtenemos,

$$P_{t-1}^n = P_{t-1}^{t-1} + J_{t-1}(P_t^n - P_t^{t-1})J_{t-1}'$$

§2.3. Estimadores de Máxima Verosimilitud.

La estimación de los parámetros que especifica el modelo de Espacio-Estado definidos en (2.1) y (2.2) es bastante complicado. Sea $\theta = \{\mu_0, \Sigma_0, \Phi, Q, R\}$ el vector de parámetro que contiene los elementos de la media inicial μ_0 y covarianza Σ_0 , la matriz de transición Φ y matrices de covarianzas Q y R de la ecuación de estado y observación respectivamente.

Utilizaremos máxima verosimilitud bajo el supuesto que el estado inicial es normal, $s_0 \sim N(\mu_0, \sigma_0)$ y los errores w_1, w_2, \dots, w_n i.i.d y v_1, v_2, \dots, v_n i.i.d son normales conjuntamente independientes.

Continuamos por suponer, por simplicidad que $\{W_t\}$ y $\{V_t\}$ son no correlacionados.

La verosimilitud es calculada utilizando las innovaciones $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ definidas en la ecuacion (2.13)

$$\varepsilon_t = y_t - M_t x_t^{t-1}$$

En la sección anterior se probó que $E(\varepsilon_t) = 0$ y $\Sigma_t = \text{var}(\varepsilon_t) = M_t P_t^{t-1} M_t' + R$.

Por lo tanto podemos escribir el log de la verosimilitud, $L_Y(\theta)$ como:

$$-\ln L_Y(\theta) = \frac{1}{2} \sum_{t=1}^n \log |\Sigma_t(\theta)| + \frac{1}{2} \sum_{t=1}^n \varepsilon_t(\theta)' \Sigma_t(\theta)^{-1} \varepsilon_t(\theta) + \text{constante} \quad (2.22)$$

Donde se resalta la dependencia de las innovaciones sobre los parámetros θ . Por supuesto la ecuación (2.22) es una función no lineal de los parámetros desconocidos.

■

CAPÍTULO 3

ALGORITMO EM

§3.1. Definición de algoritmo EM

Sea $p(\mathbf{x}, \Theta)$ una función de densidad determinada por el vector de parámetros Θ . Supongamos que $\mathcal{X} = \{X_1, \dots, X_n\}$ es una muestra aleatoria independiente idénticamente distribuida con distribución p . La función de densidad conjunta de la muestra \mathcal{X} viene dada por

$$p(\mathbf{x}|\Theta) = \prod_{i=1}^n p(x_i|\Theta)$$

donde $\mathbf{x} = (x_1, \dots, x_n)$ son datos observados de la muestra \mathcal{X} , la función \mathcal{L} es llamada función de verosimilitud. El problema de máxima verosimilitud consiste en hallar Θ que maximize a \mathcal{L} . Es decir, se desea hallar Θ^* donde,

$$\Theta^* = \arg \max_{\Theta} f(\Theta|\mathbf{x}).$$

Dependiendo de la forma de $p(\mathbf{x}, \Theta)$, hallar Θ^* puede ser fácil o complicado. Por ejemplo, si $p(\mathbf{x}, \Theta)$ es normal con $\Theta = (\mu, \sigma^2)$, entonces es fácil ver que

$$\Theta^* = (\bar{x}, s^2) = \left(\frac{1}{n} \sum_{i=1}^n x_i, \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right).$$

Sin embargo, en muchos problemas, no es posible encontrar tales expresiones analíticas y debemos recurrir a técnicas más elaboradas. El algoritmo EM es una de esas técnicas.

El algoritmo EM es una técnica utilizada para encontrar el estimador de máxima verosimilitud de una distribución de un conjunto de datos incompletos. El término “datos incompletos” en su forma general implica la existencia de dos muestras aleatorias \mathcal{X} , \mathcal{Y} y una función de \mathcal{X} a \mathcal{Y} . Sea \mathbf{y} los datos observados de la muestras \mathcal{Y} , el correspondiente \mathbf{x} de \mathcal{X} no es directamente observable, pero si lo es indirectamente a través de \mathbf{y} . Específicamente, supondremos que existe una función $\mathbf{x} \rightarrow \mathbf{y}(\mathbf{x})$ de

\mathcal{X} a \mathcal{Y} y \mathbf{x} es conocido en $\mathcal{X}(\mathbf{y})$, el subconjunto de \mathcal{X} determinado por la ecuación $\mathbf{y} = \mathbf{y}(\mathbf{x})$ donde \mathbf{y} son los datos observados. A \mathbf{x} lo llamaremos datos completos. Observemos que

$$\mathcal{L}(\Theta|\mathbf{y}) = \int_{\mathcal{X}(\mathbf{y})} f(\Theta|\mathbf{x})d\mathbf{x}$$

El Algoritmo EM encuentra el valor de Θ que maximiza a $\mathcal{L}(\Theta|\mathbf{y})$ comenzando con un valor inicial para Θ . En cada iteración i el algoritmo se realiza en dos pasos, en el primer paso se calcula la esperanza $Q(\Theta|\Theta^{i-1}) = E[\log(\mathcal{L}(\Theta|\mathbf{x}))|\mathbf{y}, \Theta^{i-1}]$, donde Θ^{i-1} es el estimado de Θ en la iteración $i-1$ y en el segundo paso maximizamos $Q(\Theta|\Theta^{i-1})$. Para fijar ideas, mostraremos un ejemplo indicando los pasos del algoritmo.

Ejemplo: 197 individuos estan distribuidos multinomialmente en cuatro categorias, los datos observados son:

$$\mathbf{y} = (y_1, y_2, y_3, y_4) = (125, 18, 20, 34)$$

supongamos que la probabilidades de que una observación este en cada una de las categorias es respectivamente

$$(p_1, p_2, p_3, p_4) = \left(\frac{1}{2} + \frac{\pi}{4}, \frac{1}{4}(1 - \pi), \frac{1}{4}(1 - \pi), \frac{\pi}{4} \right)$$

para algún $\pi \in [0, 1]$ desconocido y el cual se desea estimar. Así,

$$\mathcal{L}(\pi|\mathbf{y}) = \frac{(y_1 + y_2 + y_3 + y_4)!}{y_1!y_2!y_3!y_4!} \left(\frac{1}{2} + \frac{\pi}{4} \right)^{y_1} \left(\frac{1}{4}(1 - \pi) \right)^{y_2} \left(\frac{1}{4}(1 - \pi) \right)^{y_3} \left(\frac{\pi}{4} \right)^{y_4}$$

Para ilustrar el algoritmo EM, representamos a \mathbf{y} como los datos incompletos de una población multinomial de 5 categorias y las probabilidades de que un individuo pertenezca a cada categoria son $(\frac{1}{2}, \frac{\pi}{4}, \frac{1}{4}(1 - \pi), \frac{1}{4}(1 - \pi), \frac{\pi}{4})$, la idea es dividir la primera de las cuatro categorias en dos categorias. Así, los datos completos son $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5)$, donde $y_1 = x_1 + x_2$, $y_2 = x_3$, $y_3 = x_4$, $y_4 = x_5$ y la verosimilitud de los datos completos es

$$\mathcal{L}(\pi|\mathbf{x}) = \frac{(x_1 + x_2 + x_3 + x_4 + x_5)!}{x_1!x_2!x_3!x_4!x_5!} \left(\frac{1}{2} \right)^{x_1} \left(\frac{\pi}{4} \right)^{x_2} \left(\frac{1}{4}(1 - \pi) \right)^{x_3} \left(\frac{1}{4}(1 - \pi) \right)^{x_4} \left(\frac{\pi}{4} \right)^{x_5}$$

Aplicando log obtenemos

$$\log \mathcal{L}(\pi|\mathbf{x}) = K + (x_2 + x_5) \log \left(\frac{\pi}{4} \right) + (x_3 + x_4) \log \left(\frac{1}{4}(1 - \pi) \right)$$

así,

$$E[\log \mathcal{L}(\pi|\mathbf{x})|\mathbf{y}, \pi^{i-1}] = K + E(x_2|y_1, \pi^{i-1}) \log\left(\frac{\pi}{4}\right) + y_4 \log\left(\frac{\pi}{4}\right) + (y_2 + y_3) \log\left(\frac{1}{4}(1 - \pi)\right)$$

donde K no depende de π . Ahora, utilizando el hecho que x_2 y y_1 tienen distribución binomial obtenemos

$$E(x_2|y_1, \pi) = y_1 \frac{\frac{\pi^{i-1}}{4}}{\frac{1}{2} + \frac{\pi^{i-1}}{4}}$$

por lo tanto el primer paso del algoritmo es

$$Q(\pi|\pi^{i-1}) = E[\log \mathcal{L}(\pi|\mathbf{x})|\mathbf{y}, \pi] = K + y_1 \frac{\frac{\pi^{i-1}}{4}}{\frac{1}{2} + \frac{\pi^{i-1}}{4}} \log\left(\frac{\pi}{4}\right) + y_4 \log\left(\frac{\pi}{4}\right) + (y_2 + y_3) \log\left(\frac{1}{4}(1 - \pi)\right)$$

En el segundo paso maximizamos $Q(\pi|\pi^{i-1})$, entonces derivando con respecto a π obtenemos

$$Q'(\pi|\pi^{i-1}) = 0 + y_1 \frac{\frac{\pi^{i-1}}{4}}{\frac{1}{2} + \frac{\pi^{i-1}}{4}} \frac{1}{\pi} + y_4 \frac{1}{\pi} - (y_2 + y_3) \frac{1}{1 - \pi} = 0$$

Despejando π obtenemos

$$\pi = \pi^i = \frac{(y_1 + y_4)\pi^{i-1} + 2y_4}{(y_1 + y_2 + y_3 + y_4)\pi^{i-1} + 2(y_2 + y_3 + y_4)}$$

Comenzando con $\pi^0 = 0,5$ obtenemos $\pi^* = 0,626821484$ en 8 iteraciones y con una tolerancia de 10^{-6} .

§3.1.1. Propiedades del algoritmo EM

El logaritmo de la distribución condicional de x dada y y Θ viene dada por

$$\log k(x|y; \Theta) = \log f_x(\Theta) - \log g_y(\Theta),$$

donde $f_x(\Theta)$ es la función de densidad conjunta de los datos completos \mathbf{x}, \mathbf{y} y $g_y(\Theta)$ es la función de densidad de las observaciones \mathbf{y} . Además, sabiendo que $\mathcal{L}(\Theta) = \log g(\Theta)$, aplicando esperanza sobre $\mathcal{L}(\Theta)$, condicionada por \mathbf{y} obtenemos

$$E\{\mathcal{L}(\Theta|y)|y\} = E_{\Theta^k}\{\log f_x(\Theta)|y\} - E_{\Theta^k}\{\log k(x|y; \Theta)|y\}$$

lo que implica:

$$\begin{aligned} \mathcal{L}(\Theta|y) = E_{\Theta^k}\{\mathcal{L}(\Theta|y)|y\} &= E_{\Theta^k}\{\log f(\Theta)|y\} - E_{\Theta^k}\{\log k(x|y; \Theta)|y\} \\ &= Q(\Theta; \Theta^{(k)}) - H(\Theta, \Theta^{(k)}) \end{aligned}$$

Luego,

$$\begin{aligned} \mathcal{L}(\Theta^{(k+1)}|y) - \mathcal{L}(\Theta^{(k)}|y) &= \{Q(\Theta^{(k+1)}; \Theta^{(k)}) - Q(\Theta^{(k)}; \Theta^{(k)})\} \\ &\quad - \{H(\Theta^{(k+1)}, \Theta^{(k)}) - H(\Theta^{(k)}, \Theta^{(k)})\} \end{aligned}$$

El primer término de la diferencia en la anterior ecuación es no negativo ya que en el paso M del algoritmo EM para la estimación elegimos $\Theta^{(k+1)}$ de tal forma que

$$Q(\Theta^{(k+1)}; \Theta^{(k)}) \geq Q(\Theta^{(k)}; \Theta^{(k)})$$

Luego, $\mathcal{L}(\Theta^{(k+1)}|y) \geq \mathcal{L}(\Theta^{(k)}|y)$ si

$$H(\Theta^{(k+1)}, \Theta^{(k)}) \leq H(\Theta^{(k)}, \Theta^{(k)})$$

con lo que obtenemos que $\mathcal{L}(\Theta^{(k+1)}|y) \geq \mathcal{L}(\Theta^{(k)}|y)$ en cada paso del algoritmo, así, la sucesión $\{\mathcal{L}(\Theta^{(k)}|y)\}$ es creciente, entonces si dicha sucesión es acotada, es convergente. Ahora, demostremos que

$$H(\Theta^{(k+1)}, \Theta^{(k)}) \leq H(\Theta^{(k)}, \Theta^{(k)}),$$

para todo Θ tenemos que:

$$\begin{aligned} H(\Theta, \Theta^{(k)}) - H(\Theta^{(k)}, \Theta^{(k)}) &= E_{\Theta^{(k)}}[\log\{k(X|y; \Theta)/k(X|y; \Theta^{(k)})|y\}] \\ &\leq \log[E_{\Theta^{(k)}}\{k(X|y; \Theta)/k(X|y; \Theta^{(k)})|y\}] \\ &= \log \int_{\mathcal{X}(y)} k(X|y; \Theta) dx = 0 \end{aligned}$$

§3.2. Representación del algoritmo EM para modelos de espacio-estado

R.H Shumway y D. S Stoffer, 1982 ver([6]), proponen la implementación del algoritmo *EM* en conjunto con los estimadores de Kalman para generar un procedimiento recursivo para obtener los estimadores de máxima verosimilitud en un Modelo de Espacio-Estado cuando en este se presentan datos faltantes (datos incompletos). El modelo está descrito por las siguientes ecuaciones:

$$y_t = M_t x_t + v_1, \quad t = 1, 2, \dots, n \tag{3.1}$$

$$x_t = \Phi x_{t-1} + w_1, \quad t = 1, 2, \dots, n \tag{3.2}$$

Donde x_t es un vector de orden $p \times 1$, w_t , con $t = 1, \dots, n$ son vectores aleatorios normales independientes e idénticamente distribuidos con medida cero y matriz de covarianza Q , x_0 es un vector aleatorio normal con media μ_0 y matriz de covarianza Σ_0 de orden $p \times p$, M_t es un matriz de orden $q \times p$, y_t es un vector de observación de orden $q \times 1$ y v_t es un vector normal con media cero y matriz de covarianza R de orden $q \times q$, y además $\{w_t\}$ y $\{v_t\}$ son no correlacionados.

Para desarrollar un procedimiento para estimar los parámetros en el modelo espacio-estado definido por (3.1) y (3.2), notamos primero que la log verosimilitud conjunta de los datos completos $x_1, \dots, x_n, y_1, \dots, y_n$ puede ser escrita de la forma

$$\begin{aligned} \log L \doteq & -\frac{1}{2} \log |\Sigma| - \frac{1}{2} (\mathbf{x}_0 - \mu)' \Sigma^{-1} (\mathbf{x}_0 - \mu) \\ & - \frac{n}{2} \log |\mathbf{Q}| - \frac{1}{2} \sum_{t=1}^n (\mathbf{x}_t - \Phi \mathbf{x}_{t-1})' \mathbf{Q}^{-1} (\mathbf{x}_t - \Phi \mathbf{x}_{t-1}) \\ & - \frac{n}{2} \log |\mathbf{R}| - \frac{1}{2} \sum_{t=1}^n (\mathbf{y}_t - \mathbf{M}_t \mathbf{x}_t)' \mathbf{R}^{-1} (\mathbf{y}_t - \mathbf{M}_t \mathbf{x}_t) \end{aligned} \quad (3.3)$$

Donde $\log L$ debe ser maximizado respecto a los parámetros μ, Σ, Φ, Q y R . Ya que la log *verosimilitud* en (3.3) depende de la serie de datos no-observados \mathbf{x}_t , $t = 0, 1, \dots, n$, consideraremos la aplicación de algoritmo EM condicionalmente respecto a las series observadas y_1, y_2, \dots, y_n . Esto es, se definen los parámetros estimados en el $(r + 1)$ -iterado como los valores μ, Σ, Φ, Q, R , que maximizan

$$G(\mu, \Sigma, \Phi, Q, R) = E_r(\log L | y_1, \dots, y_n) \quad (3.4)$$

donde E_r denota la esperanza condicional relativa a una densidad que contiene a los r -valores iterados $\mu(r), \Sigma(r), \Phi(r), Q(r)$ y $R(r)$. Un procedimiento iterativo se define como una secuencia de pasos, como se ha demostrado en Dempster (en 1977) para obtener una verosimilitud no decreciente, con un punto fijo definido como un punto estacionario de la función de verosimilitud.

Con el fin de calcular esperanza condicional definida en (3.4), es conveniente definir la media condicional

$$\mathbf{x}_t^s = E(\mathbf{x}_t | y_1, \dots, y_s) \quad (3.5)$$

la función covarianza

$$\mathbf{P}_t^s = cov(\mathbf{x}_t | y_1, \dots, y_s) \quad (3.6)$$

y

$$\mathbf{P}_{t,t-1}^s = cov(\mathbf{x}_t, \mathbf{x}_{t-1} | y_1, \dots, y_s) \quad (3.7)$$

Ahora bien, tomando esperanza condicional en (3.3) tenemos :

$$\begin{aligned} G(\mu, \Sigma, \Phi, Q, R) &= -\frac{1}{2} \log |\Sigma| - \frac{1}{2} tr \{ \Sigma^{-1} (P_0^n + (x_0^n - \mu)(x_0^n - \mu)') \} \\ &= -\frac{n}{2} \log |Q| - \frac{1}{2} tr \{ Q^{-1} (C - B\Phi' - \Phi B' + \Phi A\Phi') \} \\ &= -\frac{n}{2} \log |R| \\ &\quad - \frac{1}{2} tr \left\{ R^{-1} \sum_{t=1}^n [(y_t - M_t x_t^n)(y_t - M_t x_t^n)' + M_t P_t^n M_t'] \right\} \end{aligned} \quad (3.8)$$

donde tr denota la traza y

$$A = \sum_{t=1}^n (P_{t-1}^n + x_{t-1}^n x_{t-1}^{n'}). \quad (3.9)$$

$$B = \sum_{t=1}^n (P_{t,t-1}^n + x_{t-1}^n x_{t-1}^{n'}). \quad (3.10)$$

$$y \quad C = \sum_{t=1}^n (P_t^n + x_t^n x_t^{n'}). \quad (3.11)$$

Los terminos del Filtro de Kalman x_t^n , P_t^n y $P_{t,t-1}^n$ son calculados bajo los valores de los parámetros $\mu(r)$, $\Phi(r)$, $Q(r)$, $R(r)$ usando las recursiones descritas en el capítulo anterior. Además si derivamos (3.8) en función de cada uno de los parámetros, e igualando a cero, es facil ver que

$$\Phi(r+1) = BA' \quad (3.12)$$

$$Q(r+1) = n^{-1}(C - BA^{-1}B') \quad (3.13)$$

$$y \quad R(r+1) = n^{-1} \sum_{t=1}^n [(y_t - M_t x_t^n)(y_t - M_t x_t^n)' + M_t P_t^n M_t'] \quad (3.14)$$

Y además, el valor de la función log-verosimilitud puede ser calculado en cada paso, usando las innovaciones, a través de la siguiente expresión:

$$\begin{aligned} \log L &= -\frac{1}{2} \sum_{t=1}^n | M_t P_t^{t-1} M_t' + R_t | \\ &\quad - \frac{1}{2} \sum_{t=1}^n (y_t - M_t x_t^{t-1})(M_t P_t^{t-1} M_t' + R_t)^{-1} (y_t - M_t x_t^{t-1})' \end{aligned} \quad (3.15)$$

§3.3. Aplicación del Algoritmo EM

Consideremos el siguiente modelo:

Sea $x_0 \sim N(0, \frac{\sigma^2}{1-\phi^2})$ y para $t = 1, \dots, n$ se define

$$x_t = \phi x_{t-1} + \sigma u_t \tag{3.16}$$

$$y_t = \beta \exp(x_t/2) v_t \tag{3.17}$$

donde u_t y v_t son independientes con distribución $N(0, 1)$

Ahora, tomando logaritmos en las observaciones al cuadrado se obtiene

$$r_t = \log y_t^2 = \log(\beta^2) + x_t + \log(v_t^2) \tag{3.18}$$

y si aproximamos $\log(v^2)$ a una distribución normal, podemos utilizar filtros de Kalman para calcular una aproximación a la función de verosimilitud para el modelo lineal dinamico representado por la ecuaciones (3.16) y (3.18).

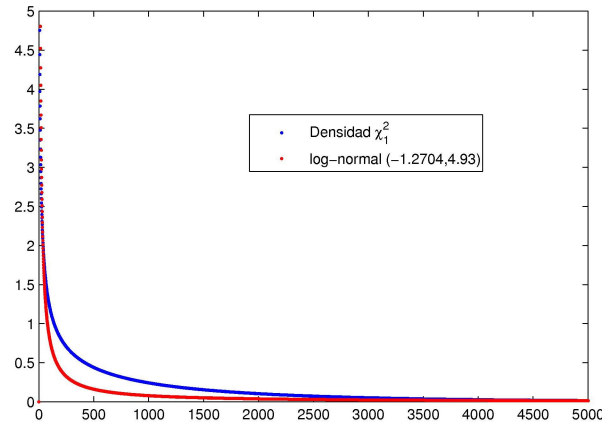


FIGURA 3.1: APROXIMACIÓN LOG-NORMAL A UNA χ_1^2

Si definimos $\eta_t = \sigma u_t$, $\alpha_t = \log(v_t^2)$ y $\varphi = \log(\beta^2)$ entonces las ecuaciones (1) y (3) se pueden reescribir de la siguiente manera:

$$\eta_t = x_t - \phi x_{t-1} \tag{3.19}$$

$$\alpha_t = r_t - x_t - \varphi \quad (3.20)$$

donde $\eta_t \sim N(0, \sigma^2)$ y $\alpha_t \sim N(-1,2704, 4,93)$.

Así, $\theta = (\sigma, \varphi, \phi)$ es el vector de parámetros a estimar en la ecuación de quasi-verosimilitud dada por la siguiente expresión:

$$\begin{aligned} \log L_z(\theta) &= -\frac{1}{2} \log \left(\frac{\sigma^2}{1 - \phi^2} \right) - \frac{1}{2} \left(\frac{(1 - \phi^2)x_0^2}{\sigma^2} \right) \\ &\quad - \frac{n}{2} \log \sigma^2 - \frac{1}{2} \sum_{t=1}^n \left(\frac{x_t - \phi x_{t-1}}{\sigma} \right)^2 \\ &\quad - \frac{1}{2} \sum_{t=1}^n \left(\frac{r_t - \varphi - x_t + 1,2704}{\sqrt{4,93}} \right)^2 \end{aligned}$$

Luego, aplicando la definición del paso E , descrito en el algoritmo EM, definimos entonces

$$G(\sigma, \varphi, \phi) = E_r(\log L_z(\theta) | y_1, \dots, y_n)$$

y usando las definiciones (2.3), (2.4) y (2.5) del capítulo 2, se obtiene que

$$\begin{aligned} G(\theta) &= -\frac{1}{2} \log \sigma^2 + \frac{1}{2} \log(1 - \phi^2) - \frac{1}{2\sigma^2} ((1 - \phi^2)(P_0^n + (x_0^n)^2)) \\ &\quad - \frac{n}{2} \log \sigma^2 - \frac{1}{\sigma^2} (C - 2\phi B + \phi^2 A) \\ &\quad - \frac{1}{9,86} \sum_{t=1}^n [(r_t - \varphi - x_t^n + 1,2704)^2 + P_t^n] \end{aligned}$$

Donde

$$\begin{aligned} A &= \sum_{t=1}^n (P_{t-1}^n + x_{t-1}^n (x_{t-1}^n)') \\ B &= \sum_{t=1}^n (P_{t,t-1}^n + x_t^n (x_{t-1}^n)') \end{aligned}$$

y

$$C = \sum_{t=1}^n (P_t^n + x_t^n (x_t^n)')$$

Ahora bien, derivando la función G respecto a los parámetros e igualando a cero, obtendremos los estimadores $\sigma^2(r+1)$, $\varphi(r+1)$, $\phi(r+1)$. En efecto:

$$\begin{aligned} \frac{\partial G(\theta)}{\partial \sigma^2} &= -\frac{1}{2} \frac{1}{\sigma^2} + \frac{(1 - \phi^2)(P_0^n + (x_0^n)^2)}{2\sigma^4} - \frac{n}{2} \frac{1}{\sigma^2} + \frac{(C - 2\phi B + \phi^2 A)}{\sigma^4} \\ &= -\left(\frac{n+1}{2} \right) \frac{1}{\sigma^2} + \frac{(1 - \phi^2)(P_0^n + (x_0^n)^2) + 2(C - 2\phi B + \phi^2 A)}{2\sigma^4} \end{aligned}$$

$$\begin{aligned}
 \frac{\partial G(\theta)}{\partial \phi} &= \frac{2\phi}{2(1-\phi^2)} + \frac{2\phi}{2\sigma^2}(P_0^n + (x_0^n)^2) \\
 &\quad - \frac{1}{2\sigma^2}(-2B + 2A\phi) \\
 &= \frac{\phi}{1-\phi^2} + \frac{\phi}{\sigma^2}(P_0^n + (x_0^n)^2) \\
 &\quad + \frac{1}{\sigma^2}(B - A\phi)
 \end{aligned}$$

$$\begin{aligned}
 \frac{\partial G(\theta)}{\partial \varphi} &= -\frac{1}{9,86} \sum_{t=1}^n 2(r_t - \varphi - x_t^n + 1,2704)(-1) \\
 &= \frac{1}{4,93} \sum_{t=1}^n (r_t - x_t^n + 1,2704) - n\varphi
 \end{aligned}$$

Luego, igualando cada derivada parcial a cero, y despejando $\sigma^2(r+1)$, $\varphi(r+1)$ respectivamente, obtenemos que:

$$\sigma^2(r+1) = \frac{(1-\phi^2)(P_0^n + (x_0^n)^2 + 2(C - 2\phi B + \phi^2 A))}{n+1} \quad (3.21)$$

$$\varphi(r+1) = \frac{1}{n} \frac{\sum_{t=1}^n (r_t - x_t^n + 1,2704)}{4,93} \quad (3.22)$$

además $\phi(r+1)$ se obtiene de la siguiente ecuación y se selecciona el que su valor absoluto sea menor que uno

$$\phi^3(A - P_0^n - (x_0^n)^2) - \phi^2(B) + \phi(\sigma^2 + P_0^n + (x_0^n)^2 - A) + B = 0 \quad (3.23)$$

y la función *log*-Verosimilitud, usando las innovaciones, viene dada por la siguiente expresión:

$$\begin{aligned}
 \log L_{\varepsilon_t} &= -\frac{1}{2} \sum_{t=1}^n \log |P_t^{t-1} + 4,93| \\
 &= -\frac{1}{2} \sum_{t=1}^n (r_t - \varphi - x_t^{t-1} + 1,2704)(P_t^{t-1} + 4,93)^{-1}(r_t - \varphi - x_t^{t-1} + 1,2704)
 \end{aligned}$$

■

§3.3.1. Resultados

El algoritmo EM se programó en un computador con las siguientes características: Procesador Intel(R) Core(TM) 2 Duo CPU T5750 2.00GHz 997MHz, 1GB de Ram. Se utilizó el Software de Matlab, Versión 7.6.0.324 (R2008a), Febrero 10,2008.

Los datos utilizados para la implementación del algoritmo EM con filtros de Kalman, fueron obtenidos de "*the British pound/US dollar exchange rate historical series*", desde 1971-01-01 al 2010-04-01, donde el $\log - return$ de la serie de estos datos, representa las observaciones $\{y_t\}_{t \geq 0}$, con 472 observaciones,

Para un total de 200 iteraciones, se obtuvieron los siguientes resultados

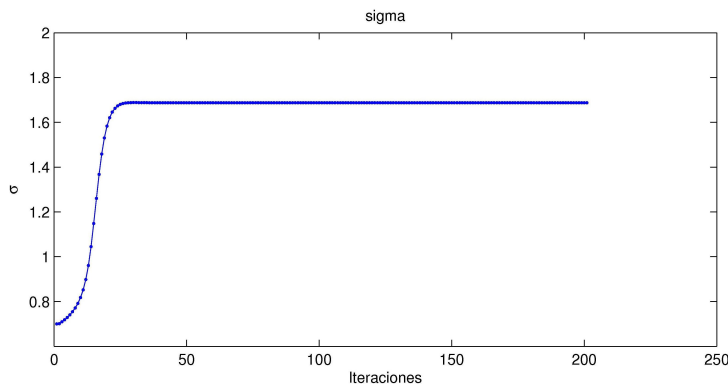


FIGURA 3.2: TRAYECTORIA DE ALGORITMO EM, PARA EL ESTIMADOR DE σ . (VALOR ESTIMADO = 1.64)

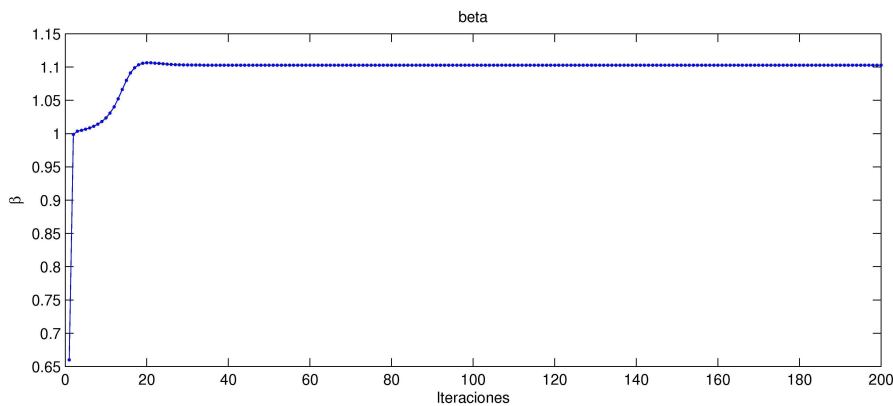


FIGURA 3.3: TRAYECTORIA DEL ALGORITMO EM, PARA EL ESTIMADOR DE β . (VALOR ESTIMADO ES = 1.104)

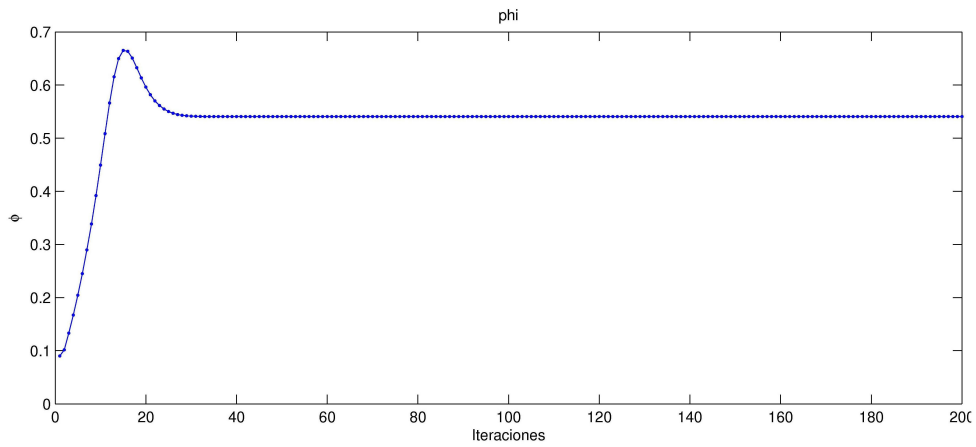


FIGURA 3.4: TRAYECTORIA DEL ALGORITMO EM, PARA EL ESTIMADOR DE ϕ .
(VALOR ESTIMADO = 0.64)

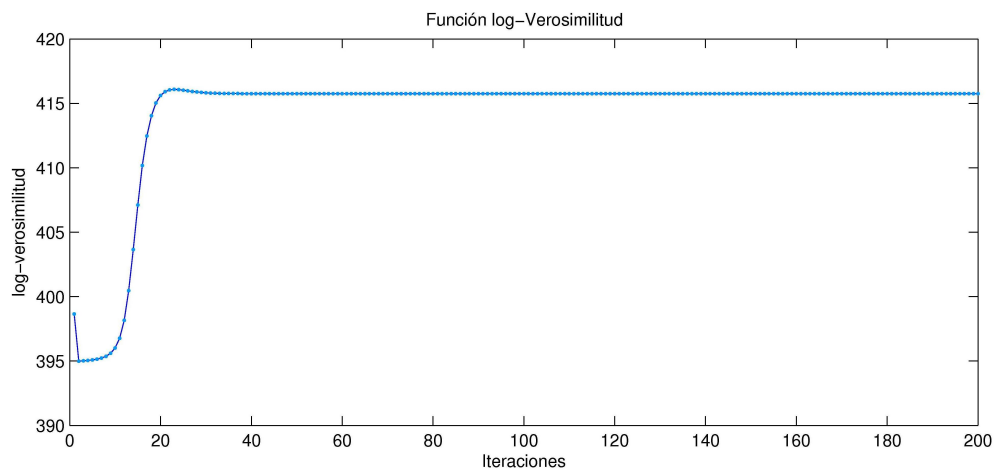


FIGURA 3.5: TRAYECTORIA DE ALGORITMO EM, PARA EL log-VEROSIMILITUD,
(VALOR MÁXIMO = 415.8)

CAPÍTULO 4

ALGORITMO MCEM

A lo largo de esta sección, mantendremos la noción de datos incompletos, introducidos en el capítulo anterior. Recordemos que el paso E del algoritmo EM consiste en evaluar la función definida en (3.4) y aquí, consideraremos el caso en que la evaluación numérica directa de esta esperanza es complicada.

Este principio fue propuesto por Wei y Tanner (1991) y Tanner (1993), y consiste en aplicar el Método de Monte Carlo para aproximar los insolubles E -pasos del algoritmo EM con una media empírica sobre la base de los datos simulados, esto es

$$\tilde{G}(\mu, \Sigma, \Phi, Q, R) = \frac{1}{m} \sum_1^m (\log L(\xi_i) | y_1, \dots, y_n) \quad (4.1)$$

donde ξ_1, \dots, ξ_m son idénticamente distribuidas con función de probabilidad $p(x, \theta)$. Este método es llamado Algoritmo MCEM donde se modifica el algoritmo EM introduciendo la aplicación del Método Monte Carlo, dicho método consiste en realizar los siguientes pasos:

1. Calculamos la función de Verosimilitud (L) de los datos completos.
2. Aplicamos logaritmo a la función de Verosimilitud ($\log(L)$).
3. (Paso E) Aplicamos Esperanza a la función $\log(L)$ y aproximamos mediante el método Monte Carlo.
4. (Paso M) Maximizamos el resultado anterior en función de los parámetros del modelo.

■

§4.0.2. Muestreador de Gibbs

El muestreador de Gibbs (Gibbs Sampling), es un caso particular del Algoritmo Metropolis-Hasting, que permite generar cadenas de variables aleatorias que cumplen la propiedad de Markov. Por nuestro interes, a continuación veremos el principio del "Muestreador de Gibb" a través de un algoritmo denominado "Muestreador de Corte":

§4.0.3. Algoritmo Muestreador de Corte

Iniciamos en un punto arbitrario (ξ^1, u^1) sobre \mathcal{I}_π , luego para $i \geq 1$ simulamos,

1. $u^{i+1} \sim U([0, \pi(\xi^i)]);$
2. $\xi^{i+1} \sim U(S(u^{i+1})),$ donde $S(u) = \{x : \pi(x) \geq u\}$

Para el modelo descrito por las siguientes ecuaciones,

$$x_t = \phi x_{t-1} + \sigma u_t$$

$$y_t = \beta \exp(x_t/2) v_t$$

con $x_0 \sim N(0, \frac{\sigma^2}{1-\phi^2})$, $\{u_t\}_{t \geq 0}$ y $\{v_t\}_{t \geq 0}$ independientes con distribución normal gaussiana, si consideramos la distribución condicional de x_k , dados x_{k-1}, x_{k+1}, Y_k , ignorando constantes, es proporcional a

$$\exp \left[- \left\{ \frac{(x_{k+1} - \phi x_k)^2}{\sigma^2} + \frac{(x_k - \phi x_{k-1})^2}{\sigma^2} \right\} \right] \frac{1}{\beta \exp(x/2)} \exp \left[- \frac{y_k^2}{2\beta^2 \exp x} \right]$$

Y además, completando cuadrados y simplificando la expresión anterior se obtiene

$$\pi(x|x_{k-1}, x_{k+1}) \propto \exp \left[- \frac{1 + \phi^2}{2\sigma^2} \left\{ (x - \mu_k)^2 + \frac{y_k^2 \sigma^2}{(1 + \phi^2)\beta^2} \exp(-x) \right\} \right]$$

donde

$$\mu_k = \frac{\phi(x_{k+1} + x_{k-1}) - \frac{\sigma^2}{2}}{1 + \phi^2}$$

Ahora, definiendo

$$\alpha_k = \frac{y_k^2 \sigma^2 \exp(\mu_k)}{(1 + \phi^2)\beta^2} \quad y \quad \rho = \frac{1 + \phi^2}{2\sigma^2}$$

$\pi(x|x_{k-1})$ es proporcional a

$$\exp[-\rho \{ (x - \mu_k)^2 + \alpha_k \exp[-(x - \mu)] \}]$$

Y finalmente, el conjunto $S(u)$, viene dado de la siguiente manera

$$S(u) = \{x \mid \exp[-\rho\{(x - \mu_k)^2 + \alpha_k \exp[-(x - \mu_k)]\}] \geq u\}$$

No es necesario, resolver la expresión anterior de manera explicita, en este sentido para cada iterado de ξ^{i+1} , basta generarlo uniformemente sobre un conjunto grande y aceptarlo solo si $\xi^{i+1} \in S(u)$

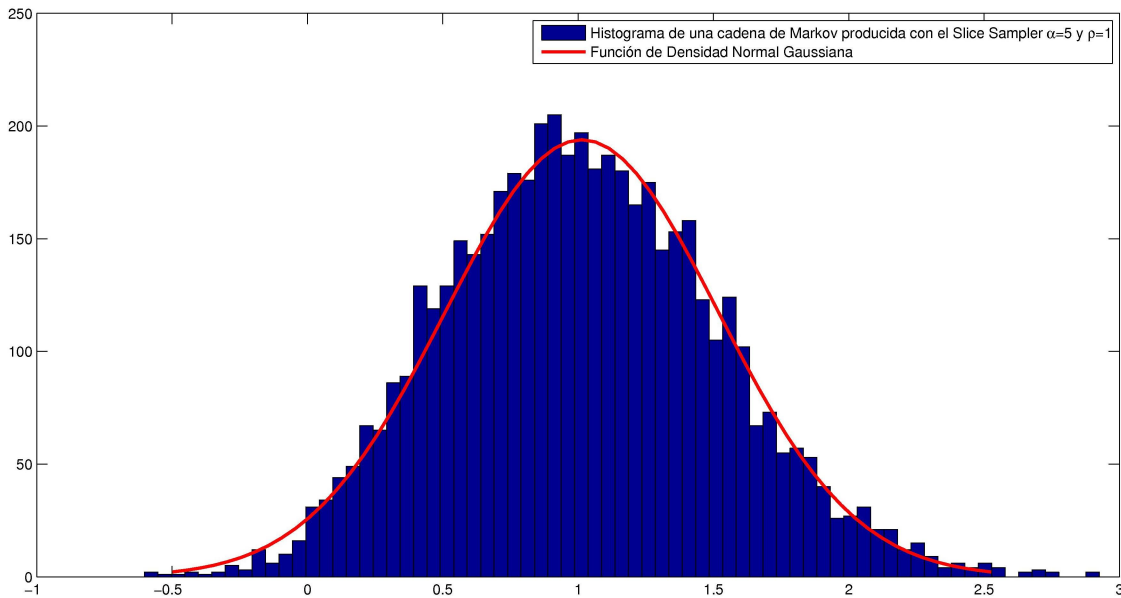


FIGURA 4.1: DENSIDAD DE UNA CADENA DE MARKOV GENERADA CON EL MUESTREADOR DE CORTE, CON $\alpha = 5$ Y $\rho = 1$

Notese que la recursión del muestreador de corte, genera una cadena de Markov, y que además posee un perfil Gaussiano, lo cual es necesario en este modelo.

§4.1. Aplicación del algoritmo MCEM

Consideremos el modelo descrito en el capítulo anterior por las ecuaciones (3.16) y (3.17),

$$x_t = \phi x_{t-1} + \sigma u_t \tag{4.2}$$

$$y_t = \beta \exp(x_t/2) v_t \tag{4.3}$$

para $t = 1, \dots, n$. Y además, $x_0 \sim N(0, \frac{\sigma^2}{1-\phi^2})$, $\{u_t\}_{t \geq 0}$ y $\{v_t\}_{t \geq 0}$ son independientes con distribución $N(0, 1)$

Luego, la función de log –Verosimilitud de los datos completos, dado el vector de parámetros $\theta = (\sigma, \beta, \phi)$, viene dada por la siguiente expresión:

$$\begin{aligned} \log \mathcal{L}(X, Y|\theta) &= -\frac{n+1}{2} \log \beta^2 - \frac{n+1}{2} \log \sigma^2 - \frac{1}{2\beta^2} s_4 \\ &\quad + \frac{1}{2} \log(1 - \phi^2) - \frac{(1 - \phi^2)s_0}{2\sigma^2} - \frac{1}{2\sigma^2} (s_2 - 2\phi s_3 + \phi^2 s_1) \end{aligned} \quad (4.4)$$

donde

$$\begin{aligned} s_0(x_{0:n}) &= x_0^2, & s_1(x_{0:n}) &= \sum_{k=0}^{n-1} x_k^2, & s_2(x_{0:n}) &= \sum_{k=1}^n x_k^2 \\ s_3(x_{0:n}) &= \sum_{k=1}^n x_k x_{k-1} & s_4(x_{0:n}) &= \sum_{k=0}^n y_k^2 \exp(-x_k) \end{aligned}$$

Ahora bien, aplicando el paso E , como se describe en este algoritmo, obtenemos la siguiente expresión:

$$\begin{aligned} \widehat{Q}_m(\theta|\theta_{k-1}) &= -\frac{n+1}{2} \log \beta^2 - \frac{n+1}{2} \log \sigma^2 - \frac{1}{2\beta^2} \widehat{S}_4 \\ &\quad + \frac{1}{2} \log(1 - \phi^2) - \frac{(1 - \phi^2)\widehat{S}_0}{2\sigma^2} - \frac{1}{2\sigma^2} (\widehat{S}_2 - 2\phi\widehat{S}_3 + \phi^2\widehat{S}_1) \end{aligned} \quad (4.5)$$

con

$$\begin{aligned} \widehat{S}_0 &= \frac{1}{m} \sum_{i=1}^m s_{0i}, & \widehat{S}_1 &= \frac{1}{m} \sum_{i=1}^m s_{1i}, & \widehat{S}_2 &= \frac{1}{m} \sum_{i=1}^m s_{2i} \\ \widehat{S}_3 &= \frac{1}{m} \sum_{i=1}^m s_{3i} & \widehat{S}_4 &= \frac{1}{m} \sum_{i=1}^m s_{4i} \end{aligned}$$

Y aplicando el paso M , maximizamos $\widehat{Q}_m(\theta|\theta_{k-1})$ en función de los parámetros para obtenemos las siguientes ecuaciones:

$$\beta = \sqrt{\frac{\widehat{S}_4}{n+1}} \quad (4.6)$$

$$\sigma^2 = \frac{1}{n+1} \{(\widehat{S}_0 + \widehat{S}_2) - 2\phi\widehat{S}_3 + \phi^2(\widehat{S}_1 - \widehat{S}_0)\} \quad (4.7)$$

$$\begin{aligned} & \phi^3[n(\widehat{S}_1 - \widehat{S}_0)] + \phi^2[(1-n)\widehat{S}_3] \\ & + \phi[-\widehat{S}_2 + n\widehat{S}_0 - (n+1)\widehat{S}_1] + (n+1)\widehat{S}_3 = 0 \end{aligned} \quad (4.8)$$

En consecuencia, el paso- M implica los siguientes calculos: encontrar ϕ , como solución de (4.8), seleccionando el que, en valor absoluto sea menor que 1, determinar ϕ , usando(4.7) y β dado por (4.6)

§4.1.1. Resultados

El algoritmo *MCEM* fue programado en un computador con las siguientes características: Procesador Intel(R) Core(TM) 2 Duo CPU T5750 2.00GHz 997MHz, 1GB de Ram. Se utilizó el Software de Matlab, Versión 7.6.0.324 (R2008a), Febrero 10,2008.

Los datos utilizados para la implementación del algoritmo *MCEM*, fueron obtenidos de "*the British pound/US dollar exchange rate historical series*", desde 1971-01-01 al 2010-04-01, donde el *log - return* de la serie de estos datos, representa las observaciones $\{y_t\}_{t \geq 0}$, con 472 observaciones, y $\{x_t\}_{t \geq 0}$, fue muestreado utilizando el muestreador de corte, descrito en este capítulo.

Para un total de 200 iteraciones, y $m = 10000$ en el paso Monte Carlo, los resultados obtenidos fueron los siguientes:

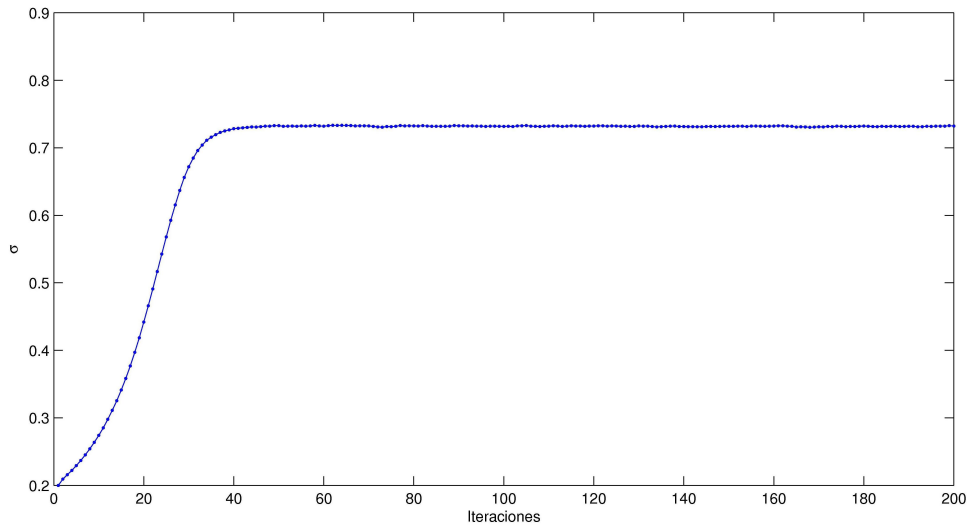


FIGURA 4.2: TRAYECTORIA DE ALGORITMO MCEM, PARA EL ESTIMADOR DE σ . (VALOR ESTIMADO = 0.73)

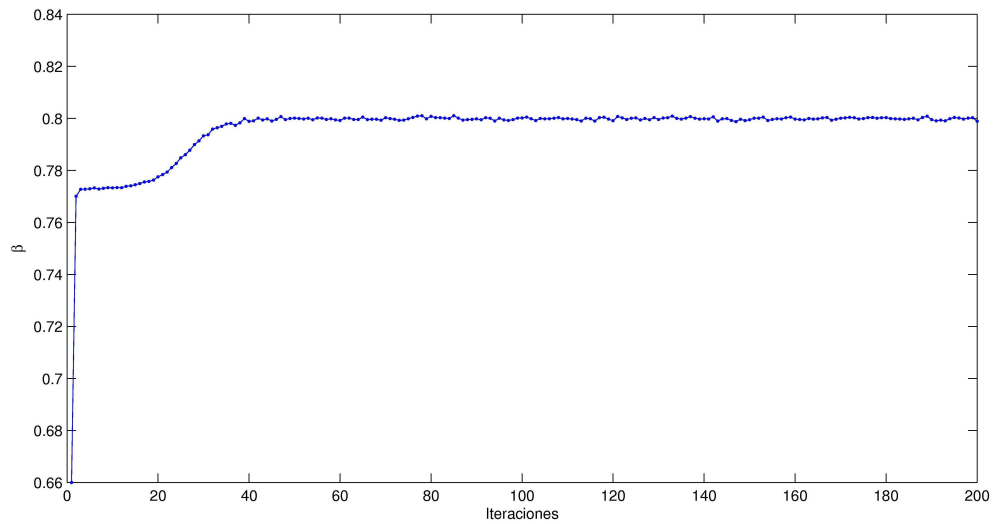


FIGURA 4.3: TRAYECTORIA DEL ALGORITMO MCEM, PARA EL ESTIMADOR DE β . (VALOR ESTIMADO = 0.8)

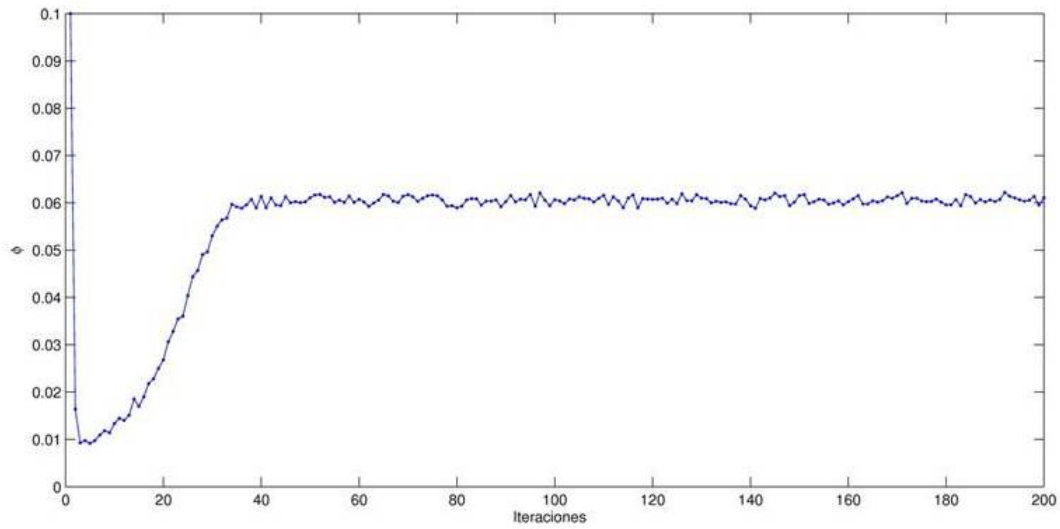


FIGURA 4.4: TRAYECTORIA DEL ALGORITMO MCEM, PARA EL ESTIMADOR DE ϕ . (VALOR ESTIMADO = 0.061)

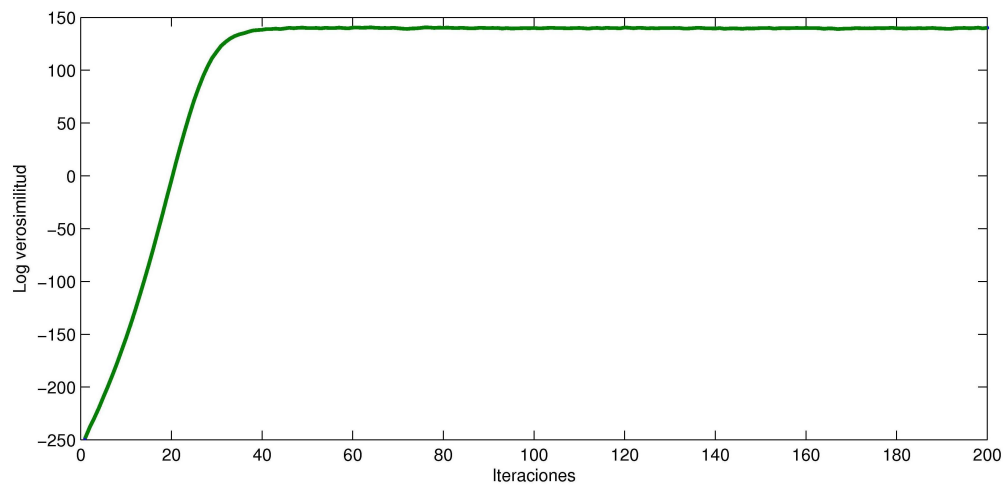


FIGURA 4.5: TRAYECTORIA DE ALGORITMO MCEM, PARA EL \log -VEROSIMILITUD. (VALOR MÁXIMO = 140)

CAPÍTULO 5

CONCLUSIONES

- Se obtuvieron los estimadores de Máxima Verosimilitud para los métodos EM con Filtros de Kalman y Monte Carlo EM.
- Para los estimadores con el método Monte Carlo EM, se observa la convergencia en un promedio de 50 iteraciones obteniendo los resultados requeridos y para los estimadores con el método EM con Filtros de Kalman, se obtuvo una la convergencia con un promedio de 30 iteraciones obteniendo los resultados deseados, observandose una diferencia entre los tiempos de ejecución de ambos metodos con un promedio de 10 a 15 minutos, siendo el método MCEM el de más larga duración.
- Dicha diferencia de tiempo esta justificada por la estructura de cada método, entre los cuales podemos mencionar para el algoritmo EM con Filtros de Kalman, la aproximación de una variable χ_1^2 a una log-normal, como se muestra en (3.1) para poder escribir el modelo de forma lineal y así usar filtros y suavizado de Kalman y Las recursiones lineales de los filtros de Kalman para la estimación de los parámetros. Por otra parte, el algoritmo MCEM utiliza el modelo original, siendo este no lineal y estocástico, donde al aplicar el principio de la integración Monte Carlo trae como consecuencia cálculos de promedios de la serie de las variables de estado y además se implemento el muestreador de corte para estas variables, el cual implica un consumo de computo mas largo (en promedio) que el método anterior.

APÉNDICE A

APÉNDICE.

§A.1. Codigos Matlab para la estimación por Máxima Verosimilitud, utilizando filtros y suavizado de Kalman

§A.1.1. Función que ejecuta el modelo

```
%INICIAMOS CON UN ESPACIO EN LIMPIO
clear all;
clc;
%Parámetros iniciales
sigma0=0.7;
beta0=0.66;
phi0=0.09;
% %función para la simulación
% [r,x]=simul(0.2,0.8,0.05,300);
% datos=r;

load Y; % Y.mat guarda los datos reales para la
% implemetación del algoritmo
datos=log(Y.^2);

n=200; %número de simulaciones

x0=normrnd(0,sigma0^2/(1-phi0^2));
P0=0.4;
%Inicialización del espacio vectorial
LL1=zeros(n,1);
```

```
sigma=zeros(n+1,1);
beta=zeros(n+1,1);
phi=zeros(n+1,1);
sigma(1)=sigma0;
beta(1)=beta0;
phi(1)=phi0;
k=2;
tic
while k<=n+1;
    %Filtros de Kalman
    [xtt ,xtt1 ,Ptt ,Ptt1 ,K,L]=kalmanf(sigma0 , beta0 , phi0 , x0 ,0 , datos );
    LL1(k-1)=L;
    %Suavizado de Kalman
    [J ,x1t ,P1t ,P12t]=kalmans(phi0 , x0 , P0 , xtt , xtt1 , Ptt , K , Ptt1 , datos );
    %Función de los estimadores de Máxima verosimilitud
    [s b p]=estimador(P1t , P12t , x1t , xtt , Ptt , datos , n , sigma0 );

    sigma(k)=s;
    beta(k)=b;
    phi(k)=p;

    sigma0=s;
    beta0=b;
    phi0=p;
    k=k+1

end
toc
%Gráficas para la seri de los resultados obtenidos
figure(1);
subplot(2,2,1); plot(LL1); title('Función log-Verosimilitud')
subplot(2,2,2); plot(sigma); title('sigma')
subplot(2,2,3); plot(beta); title('beta')
```

```
subplot(2,2,4); plot(phi); title('phi')
```

§A.1.2. Función para la simulación del modelo

```
%Función simulacion para el modelo linealizado, con parametros
%iniciales sigma, beta y phi, y n como tamaño delos datos
```

```
x=zeros(n,1);
r=x;
x0=normrnd(0,(sigma)^2/(1-(phi)^2));
vphi=log(beta);
for k=1:n
    x(k)=phi*x0+sigma*normrnd(0,sigma);
    r(k)=vphi+x(k)+normrnd(-1.2704,sqrt(4.93));
    x0=x(k);
end
```

§A.1.3. Filtro de Kalman

```
function [xtt,xtt1,Ptt,Ptt1,K,L]=kalmanf(sigma,beta,phi,x0,P0,datos)
% Filtro de Kalman
% parámetros iniciales sigma, beta, phi, x0=mu0 y P0=sigma0

[n,m]=size(datos);
%Inicialización de los espacios vectoriales
xtt=zeros(n,m);
xtt1=xtt;
Ptt=zeros(n,m);
Ptt1=zeros(n,m);
K=zeros(n,m);
LL=zeros(n,m);
vphi=log(beta);
for k=1:n
    xtt1(k)=phi*x0;
    Ptt1(k)=phi*P0*phi'+sigma^2;
```

```

%Función log-verosimilitud
LLL=-0.5*log(abs(Ptt1(k)+4.93))...
-0.5*(datos(k)-vphi-xtt1(k)+1.2704)'inv(Ptt1(k)+4.93)*...
(datos(k)-vphi-xtt1(k)+1.2704);
LL(k)=-LLL;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
K(k)=Ptt1(k)*inv(Ptt1(k)+4.93);
xtt(k)=xtt1(k)+K(k)*(datos(k)-vphi-xtt1(k));
Ptt(k)=Ptt1(k)-K(k)*Ptt1(k);
x0=xtt(k);
P0=Ptt(k);
end
L=sum(LL);

```

§A.1.4. Suavizado de Kalman

```

function [J,x1t,P1t,P12t]=kalmans(phi,x0,P0,xtt,xtt1,Ptt,K,Ptt1,datos)
% Filtro de Kalman suavizadovalores iniciales
%Con valores iniciales phi,x0,P0,xtt,xtt1,Ptt,K,Ptt1.

[n,m]=size(datos);
%Inicialización de los espacios vectoriales
x1t=zeros(n,m);
J=zeros(n,m);
P1t=zeros(n,m);
P12t=zeros(n,m);
for k=n:-1:2
    J(k)=Ptt(k-1)*phi'inv(Ptt1(k));
    x1t(k)=xtt(k-1)+J(k)*(xtt(k)-phi*xtt1(k-1));
    P1t(k)=Ptt(k-1)+J(k)*(Ptt(k)-Ptt1(k))*J(k)';
end
J(1)=P0*phi*inv(Ptt1(1));
x1t(1)=x0+J(1)*(x1t(2)-phi*xtt1(1));

```

```

P1t(1)=P0+J(1)*(P1t(2)-Ptt1(1))*J(1)';
P12t(n)=(1-K(n))*phi*Ptt(n-1);
for k=n:-1:2
    P12t(k-1)=Ptt(k-1)*J(k-1)'+J(k)*(P12t(k)-phi*Ptt(k-1))*J(k-1)';
end
end

```

§A.2. Función que calcula los estimadores de Máxima Verosimilitud

```

%Iniciamos calculando los términos coeficientes A,B, C
sum=0;
for k=1:n
    A=sum + P1t(k) +x1t(k)*x1t(k)';
    sum=A;
end
sum=0;
for k=1:n-1
    B=sum + P12t(k) +x1t(k+1)*x1t(k)';
    sum=B;
end
B=B+ P12t(n) +xtt(n)*x1t(n)';
sum=0;
for k=2:n
    C=sum + P1t(k) +x1t(k)*x1t(k)';
sum=C;
end
C=C + Ptt(n)+xtt(n)*xtt(n)';

% Estimador de Máxima verosimilitud para phi
H=[A-x1t(1)-P1t(1),-B,sigma0^2+x1t(1)+P1t(1)-A,B];
q=roots(H); % roots calcula las raices del polinomio, y H contiene los

```

```

% coheficientes del mismo
J=find(abs(q)<1);
phi=max(q(J));

% Estimador de Máxima verosimilitud para sigma
ss=((1-phi^2)*((P1t(1)+ (x1t(1)^2)))+ ...
2*(C - (2*phi*B)+(phi^2 *A)))/(n+1);
sigma=sqrt(ss);

% Estimador de Máxima verosimilitud para beta
sum=0;
for k=1:n-1
    PP=sum + datos(k) - x1t(k+1) + 1.2704;
    sum=PP;
end
P = PP+datos(n)- xtt(n) + 1.2704 ;
bb=P/(n*4.93);

beta=sqrt(exp(bb));
end

```

§A.3. Codigos Matlab para la estimación por Máxima Verosimilitud, utilizando el algoritmo MCEM

§A.3.1. Función que ejecuta el modelo

```

function [sigma ,beta ,phi ,LL]=MCEM(sigma0 ,beta0 ,phi0 ,iter ,m)
%MCEM es la función principal que simula el modelo con
%parámetros de entrada sigma0,beta0,phi0; iter es el número
% de iteraciones y m el tamaño de la serie para el paso MCMC
%[Y,X]=simmodel(300,0.7,0.65,0.09); % Función para los datos simulados
load Y % lectura de los datos reales
n=length(Y)-1;
%Inicialización del espacio vectorial

```

```
sigma=zeros(iter,1);
beta=zeros(iter,1);
phi=beta;
beta(1)=beta0;
sigma(1)=sigma0;
phi(1)=phi0;
i=1;
LL=zeros(iter,1);

while i<=iter
tic
    s0=0; s11=0; s22=0; s33=0; s44=0;
for j= 1:m;
    Xk=gibbs(Y,sigma0,beta0,phi0); %Muestreador de Gibbs para
    %la serie de las variables de estado
    s0=s0+Xk(1).^2;
    s1=sum(Xk(1:n).^2);
    s2=sum(Xk(2:n+1).^2);
    s3=sum(Xk(1:n).*Xk(2:n+1));
    s4=sum((Y.^2).*exp(-Xk));

    s11=s11+s1;
    s22=s22+s2;
    s33=s33+s3;
    s44=s44+s4;
end
s0=s0/m;
s11=s11/m;
s22=s22/m;
s33=s33/m;
s44=s44/m;

%Cálculo del estimador de Máxima verosimilitud para beta
```

```

b=sqrt(s44/(n+1));

%Cálculo del estimador de Máxima verosimilitud para phi
C=[n*(s11-s0),(1-n)*s33, -s22+n*s0-(n+1)*s11,(n+1)*s33];
q=roots(C);
J=find(abs(q)<1);
p=max(q(J));

%Cálculo del estimador de Máxima verosimilitud para sigma
s=sqrt((1/(n+1))*(s0+s22-2*p*s33+p^2*(s11-s0)));
%Valor de la función log verosimilitud
L0=-(funq(s0,s11,s22,s33,s44,s,b,p,n));

LL(i)=L0;
i=i+1
    beta(i)=b;
    phi(i)=p;
    sigma(i)=s;
    sigma0=s;
    beta0=b;
    phi0=p;
toc
end

% Graficas para la serie de los parámetros estimados
figure(1);
    subplot(3,2,1); plot(LL); title('Función log-Verosimilitud')
    subplot(3,2,2); plot(sigma); title('sigma')
    subplot(3,2,3); plot(beta); title('beta')
    subplot(3,2,4); plot(phi); title('phi')

```

§A.3.2. Función para el muestreador de corte.

```

function X=gibbs(Y,sigma,beta,phi)
%gibbs, (muestreador de corte) muestrea la serie de

```

```

%las variables de estado con parametros iniciales
%sigma,beta,phi, y Y la serie de datos observados
n=length(Y);
Xo=normrnd(0,sigma^2/(1-phi^2));
X=zeros(n,1);
j=2;
X(1)=Xo;
r=(1+phi^2)/(2*sigma^2);
a=(Y(1)^2*sigma^2)/(beta^2);
u=rand*exp(-r*((Xo)^2+a*exp(-Xo)));
while j<=n
    X1=-2+4*rand;
    if exp(-r*((X1)^2+a*exp(-X1)))>=u;

        X(j)=X1;
a=(Y(j)^2*sigma^2)/(beta^2);
j=j+1;
    Xo=X1;
    u=rand*exp(-r*((Xo)^2+a*exp(-Xo)));
end
end

```

§A.3.3. Función Log-Verosimilitud

```

function L=funq(s0,s11,s22,s33,s44,sigma,beta,phi,n)
%Función Logaritmo de la verosimilitud de los datos completos
L=(-(n+1)/2)*(log(beta^2))-(1/(2*(beta^2)))*s44-...
((n+1)/2)*(log(sigma^2))+0.5*log((1-phi^2))-...
((1-phi^2)*s0)/(2*(sigma^2))...
-1/(2*sigma^2)*(s22-2*phi*s33+(phi^2)*s11);

```

REFERENCIAS BIBLIOGRÁFICAS.

- [1] OLIVER CAPPÉ AND ERIC MOULINES AND TOBIAS RYDÉN. “*Inference in Hidden Markov Models*”. Springer Series in Statistics, Cap-6 y Cap-10, May 22, 2007.
- [2] A. P. DEMPSTER, N. M. LAIRD, AND D. B. RUBIN. “*Maximum likelihood from incomplete data via the EM algorithm*”. Journal of the Royal Statistical Society Series B, 39:1-38, 1977.
- [3] DAVID A HARVILLE. “*Matrix Algebra from a Statistician’s perspective*”. Springer, 1997.
- [4] S. KIM, N. SHEPPARD AND C. SIDDHARTHA. “*Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models*”. The Review of Economic Studies, Vol. 65, No. 3 (Jul., 1998), pp. 361-393.
- [5] WILLIAN MENDENHALL. “*Estadística Matemática con aplicaciones*”.
- [6] ROBERT. H. SHUMWAY, DAVID. S. STOFFER. “*An Approach to Time Series Smoothing And Forecasting Using EM Algorithm*”. Journal of Time Series Analysis. Vol. 3, No. 4, pp.253-263, 1982.
- [7] ROBERT. H. SHUMWAY, DAVID. S. STOFFER. “*Time series Analysis and Its Applications*”. Springer. Second Edition, 2006.