



UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

FACULTAD DE CIENCIAS

Estimación de un Proceso de Markov
discretamente observado

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIO

P R E S E N T A :

MIGUEL HERNÁNDEZ HERNÁNDEZ

TUTOR

DR. MOGENS BLADT PETERSEN





Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Hoja de Datos del Jurado

1. Datos del alumno
Hernández
Hernández
Miguel
56 12 70 50
Universidad Nacional Autónoma de México
Facultad de Ciencias
Actuaría
40104592-3
2. Datos del tutor
Dr
Mogens
Bladt
Petersen
3. Datos del sinodal 1
Dr
Ramsés Humberto
Mena
Chávez
4. Datos del sinodal 2
Dr
Eduardo Arturo
Gutiérrez
Peña
5. Datos del sinodal 3
Dr
Pablo
Padilla
Longoria
6. Datos del sinodal 4
Dr
Alberto
Contreras
Cristán
7. Datos del trabajo escrito
Estimación de un Proceso de Markov discretamente observado
98 p
2007

“Nunca se puede saber de antemano de qué son capaces las personas, hay que esperar, dar tiempo al tiempo, el tiempo es el que manda, el tiempo es quien está jugando al otro lado de la mesa y tiene en su mano todas las cartas de la baraja, a nosotros nos corresponde inventar los encartes con la vida, la nuestra.”

José Saramago

Por la oportunidad de vivir, mi madre, mi familia y mis amigos.

Gracias Dios.

A mi madre

Por tu amor, esfuerzo, dedicación y entrega. Nunca las palabras o actos de agradecimiento serán suficientes. Eres una mujer increíble. Te amo.

A mi abuelo

Por ser el hombre fuerte y tierno a quien seguir. Te extraño.

A mi tía

Por querer siempre hacer de mí alguien mejor. Gracias por tu cariño.

A mi tío

Por tus enseñanzas y consejos. Siempre influenciándome para bien.

A mis primos

Erika, Javier y Fernando

Por ser mis ejemplos y enseñarme a tener metas y sueños en la vida. Gracias por aceptarme como uno de ustedes, sus consejos y su ayuda.

A mis amigos

Por todo lo que he aprendido de ustedes, su apoyo y su amistad incondicional.

A la U.N.A.M. y profesores

Gracias por los conocimientos y la oportunidad de convertirme en alguien profesional.

ÍNDICE

PREFACIO	1
CAPÍTULO 1 . PROCESOS MARKOVIANOS	3
1.1 Cadenas de Markov en Tiempo Discreto	4
1.2 Cadenas de Markov en Tiempo Continuo	6
1.2.1 Tasas de Transición	7
1.3 La Función de Verosimilitud	12
1.4 Estimadores de Máxima Verosimilitud	13
1.5 Ejemplo	16
CAPÍTULO 2. METODOLOGÍA MONTE CARLO	
VÍA CADENAS DE MARKOV	19
2.1 Inferencia Bayesiana	19
2.1.1 Cálculo de Esperanzas	21
2.2 Integración de Monte Carlo	21
2.3 Cadenas de Markov	22
2.4 El algoritmo <i>Metropolis-Hastings</i>	23
2.5 Formas canónicas de distribuciones iniciales	25
2.5.1 El <i>Independence sampling</i>	25
2.5.2 El <i>Gibbs Sampling</i>	26
2.6 Cadenas de Markov en Espacio de Estados Generales	26
CAPÍTULO 3. ESTIMACIÓN DE MONTE CARLO	
VÍA CADENAS DE MARKOV	29
3.1 Simulación	32
3.2 Estimación	33

CAPÍTULO 4. LA INCLUSIÓN	46
4.1 Teorema	50
4.2 Ejemplo 1	54
4.3 Ejemplo 2	61
CAPÍTULO 5. RIESGO DE CRÉDITO	68
5.1 Calificaciones Moody's	68
5.2 Calificaciones de Largo Plazo	68
5.3 Calificaciones de Deuda a Largo Plazo	69
5.4 Estimación de las intensidades para las Calificaciones Moody's	71
CONCLUSIONES	94
BIBLIOGRAFÍA	97

PREFACIO

Este trabajo presenta un procedimiento de estimación para la matriz de intensidades de un Proceso de Markov basándose en la Teoría Bayesiana. Específicamente en el método de estimación de Monte Carlo vía Cadenas de Markov, el cual se basa en la teoría de los procesos estocásticos conocidos como Cadenas de Markov, evaluando las posibles soluciones para la estimación dadas observaciones puntuales de un Proceso de Markov y las situaciones que se pueden presentar en el proceso. Dicho trabajo está conformado por 5 capítulos:

Capítulo I, se proporciona una visión general de los Procesos Estocásticos conocidos como Markovianos, específicamente de las Cadenas de Markov en espacio de estado discreto, así como de los Procesos de Markov en espacio de estado discreto.

Capítulo II, se estudia de manera general el método de estimación Monte Carlo vía Cadenas de Markov. Con algunas de sus variantes más usadas: Algoritmo *Metropolis-Hastings* y el *Gibbs Sampling*. Así mismo se hace una breve introducción a las Cadenas de Markov en espacio de estados generales.

Capítulo III, se presenta un ejemplo de la estimación de la matriz de intensidades de un Proceso de Markov específico, dada la información discreta a diferentes tiempos. Se estudia la velocidad de convergencia del proceso, así como la autocorrelación de los datos. Las simulaciones y cálculos fueron generados en Matlab, así como los análisis estadísticos en SPSS.

Capítulo IV, se trata el problema de la inclusión, las posibles razones por las que la estimación de la matriz de intensidades puede no ser única debido a que una Cadena de Markov no define necesariamente uno y sólo un Proceso de Markov que incluya a la cadena.

Capítulo V, finalmente se aplica la estimación de la matriz de intensidades por medio del método de estimación Monte Carlo vía Cadenas de Markov para un caso práctico, basado en las calificaciones de crédito para una serie de empresas (o instrumentos de inversión) en la escala de Moody's.

CAPÍTULO 1

PROCESOS MARKOVIANOS

Los Procesos Markovianos se pueden clasificar de la siguiente manera:

	Tiempo Discreto	Tiempo Continuo
Espacio de Estados Discreto	Cadenas de Markov	Cadenas de Markov en Tiempo Continuo
Espacio de Estados Continuo	Cadenas de Markov en Espacio de Estados Generales	

En este capítulo se estudiarán los casos en los que el espacio de estados es discreto. En el capítulo 2 se revisarán las Cadenas de Markov en espacios de estados generales.

Un proceso de Markov es un proceso estocástico $\{X(t)|t \in T\}$ con un espacio de estados discreto E , tal que el pasado no influye en el futuro si el presente está especificado o es conocido. Esto significa lo siguiente:

Si $t_{n-1} \leq t_n$, entonces

$$P[X(t_n) = i | X(t), t \leq t_{n-1}] = P[X(t_n) = i | X(t_{n-1})],$$

con $i \in E$.

De esto sigue que si $t_1 < t_2 < \dots < t_n$, entonces

$$P[X(t_n) = i | X(t_{n-1}), \dots, X(t_1)] = P[X(t_n) = i | X(t_{n-1})].$$

Esta definición también es válida para procesos de tiempo discreto si $X(t_n)$ es reemplazado por X_n .

De lo anterior tenemos que $P[X(t_n)|X(t_{n-1}), \dots, X(t_1)] = P[X(t_n)|X(t_{n-1})]$, y de este modo obtenemos que $E[X(t_n)|X(t_{n-1}), \dots, X(t_1)] = E[X(t_n)|X(t_{n-1})]$, bajo el supuesto de que las probabilidades y los primeros momentos existen.

Además si el presente es conocido, entonces el pasado es independiente del futuro en el siguiente sentido: Si $k < m < n$, entonces

$$P[X(t_n), X(t_k)|X(t_m)] = P[X(t_n)|X(t_m)]P[X(t_k)|X(t_m)].$$

Un proceso de Markov $X(t_n)$ es llamado homogéneo si la probabilidad condicional $P[X(t_n)|X(t_{n-1})]$ es invariante con respecto a un cambio en el origen, pero su probabilidad $P[X(t_n)]$ puede depender de n .

La ecuación Chapman-Kolmogorov. La probabilidad condicional $P[X(t_n)|X(t_k)]$ puede ser expresada en términos de $P[X(t_n)|X(t_m)]$ y $P[X(t_m)|X(t_k)]$ para cualquiera $n > m > k$:

$$P[X(t_n)|X(t_k)] = \sum_{m \in E} P[X(t_n)|X(t_m)]P[X(t_m)|X(t_k)].$$

1.1 Cadenas de Markov en Tiempo Discreto

Una cadena de Markov en tiempo discreto es un proceso de Markov X_n que tiene un conjunto numerable de estados E . Una cadena de Markov está especificada en términos de sus probabilidades de estado

$$p_i[n] = P[x_n = i]$$

y sus probabilidades de transición

$$\pi_{ij}[n_1, n_2] = P[x_{n_2} = j | x_{n_1} = i].$$

Además,

$$\sum_j \pi_{ij}[n_1, n_2] = 1 \qquad \sum_i p_i[k] \pi_{ij}[k, n] = p_j[n].$$

Si $n_1 < n_2 < n_3$, entonces $\pi_{ij}[n_1, n_3] = \sum_r \pi_{ir}[n_1, n_2] \pi_{rj}[n_2, n_3]$. Esto es una forma discreta de la ecuación Chapman-Kolmogorov.

Si el proceso X_n es homogéneo, entonces las probabilidades de transición dependen sólo de la diferencia $m = n_2 - n_1$,

$$\pi_{ij}[m] = P[x_{n+m} = j | x_n = i].$$

Si $k = n_2 - n_1$, $n = n_3 - n_2$, obtenemos $\pi_{ij}[n+k] = \sum_{l \in E} \pi_{il}[k] \pi_{lj}[n]$.

Para una cadena de Markov con número de estados finito, lo anterior puede escribirse de la siguiente manera:

$$\Pi[n+k] = \Pi[n] \Pi[k]$$

donde $\Pi[n]$ es una matriz de Markov con elementos $\pi_{ij}[n]$. Esto lleva a que $\Pi[n] = \Pi^n$, donde $\Pi = \Pi[1]$ es la matriz de transición de un paso con elementos $\pi_{ij} = \pi_{ij}[1]$. Lo anterior es la solución para la ecuación recursiva $\Pi[n+1] = \Pi[n] \Pi$.

Sea $P[n]$ un vector cuyos elementos son las probabilidades de estado $p_i[n]$, entonces se concluye que $P[n] = \dots = P[n-k]\Pi^k = \dots = P[0]\Pi^n$. En general, $P[n]$ depende de n . Sin embargo, si el vector inicial $P[1] = P = [p_1, \dots, p_n]$ es tal que $P[2] = P$, entonces $P[n] = P$ para toda n . En este caso, el proceso homogéneo X_n , también es estacionario y el vector P es la solución para el sistema

$$P\Pi = P \quad \sum_i p_i = 1.$$

1.2 Cadenas de Markov en Tiempo Continuo

Una cadena de Markov en tiempo continuo es un proceso de Markov $x(t)$ que consiste de un conjunto de estados (discreto) E con discontinuidades en los puntos aleatorios t_n . Los valores $q_n = x(t_n^+)$ de $x(t)$ forman una cadena de Markov, que se dice incluida en el proceso $x(t)$.

Un proceso estocástico con espacio de estados discreto es llamado semi-Markov si no es markoviano pero la secuencia q_n incluida es una cadena de Markov.

Denotamos por

$$p_i(t) = P[X(t) = i]$$

las probabilidades de estado del proceso y por

$$\pi_{ij}[t_1, t_2] = P[X(t_2) = j | X(t_1) = i]$$

las probabilidades de transición. Estas cumplen que

$$\sum_j \pi_{ij}[t_1, t_2] = 1 \quad \sum_i p_i[t_1] \pi_{ij}[t_1, t_2] = p_j[t_2]$$

y satisfacen la ecuación de Chapman-Kolmogorov

$$\pi_{ij}[t_1, t_3] = \sum_r \pi_{ir}[t_1, t_2] \pi_{rj}[t_2, t_3] \quad t_1 < t_2 < t_3.$$

En problemas específicos las probabilidades de transición no son especificadas. Sin embargo, pueden ser determinadas por medio de las tasas de transición.

El proceso $X(t)$ es homogéneo si sus probabilidades de transición dependen de la diferencia $\tau = t_2 - t_1$

$$\pi_{ij}(\tau) = P[X(t + \tau) = j | x(t) = i].$$

De lo anterior y la ecuación de Chapman-Kolmogorov se tiene que si $\alpha = t_3 - t_2$ entonces

$$\pi_{ij}(\tau + \alpha) = \sum_{l \in E} \pi_{il}(\tau) \pi_{lj}(\alpha).$$

Esta es la ecuación Chapman-Kolmogorov para cadenas de Markov en tiempo continuo y puede ser escrita en forma matricial:

$$\Pi[\tau + \alpha] = \Pi[\tau] \Pi[\alpha] \quad \tau, \alpha \geq 0$$

donde $\Pi[\tau]$ es la matriz con elementos $\pi_{ij}(\tau)$.

1.2.1 Tasas de Transición

En el caso de tiempo discreto, se mostró que la matriz $\Pi[n]$ satisface la ecuación $\Pi[n + 1] = \Pi[n] \Pi$ y puede ser calculada en términos de su matriz de transición de un paso Π . A continuación se muestra que la matriz de transición $\Pi[\tau]$ de una cadena en tiempo continuo $X(t)$ satisface una ecuación diferencial y puede ser calculada por la matriz

$$\Pi'[0^+] = \Lambda \equiv \begin{bmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1N} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2N} \\ \dots & \dots & \dots & \dots \\ \lambda_{N1} & \lambda_{N2} & \dots & \lambda_{NN} \end{bmatrix}$$

en donde los elementos $\lambda_{ij} = \pi'_{ij}(0^+)$ son las derivadas por la derecha de los elementos $\pi_{ij}(\tau)$ de $\Pi[\tau]$. Estas derivadas serán las tasas de transición de $X(t)$. Claramente,

$$\sum_j \lambda_{ij} = 0 \text{ por que } \sum_j \pi_{ij}(\tau) = 1.$$

Como $\pi_{ij}(0) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$, se concluye con $\mu_i = -\lambda_{ii}$ que

$$\mu_i = \sum_{j \neq i} \lambda_{ij} \geq 0 \quad \lambda_{ij} \geq 0 \quad i \neq j.$$

En lo anterior, se supone que $\pi_{ij}(\tau)$ es diferenciable en $\tau = 0^+$. Esto sólo pasa si la probabilidad de que haya un punto de discontinuidad en el intervalo $(t, t + \Delta t)$ es proporcional a Δt :

$$P[X(t + \Delta t) = j | X(t) = i] = \begin{cases} 1 - \mu_i \Delta t & i = j \\ \lambda_{ij} \Delta t & i \neq j \end{cases}.$$

Derivando $\Pi[\tau + \alpha] = \Pi[\tau]\Pi[\alpha]$ con respecto a α , y fijando $\alpha = 0$, obtenemos

$$\Pi'[\tau] = \Pi[\tau]\Lambda \quad \Pi[0] = I.$$

Este es un sistema de ecuaciones diferenciales lineales con coeficientes constantes y su condición inicial es la matriz identidad. Resolviendo se tiene

$$\Pi[\tau] = e^{\Lambda \tau}.$$

De este modo se expresa $\Pi[\tau]$ en términos de la matriz de tasas de transición Λ .

Definimos W_t como la duración de tiempo que el proceso X permanece en el estado ocupado en el instante t . Entonces se tiene el siguiente resultado:

Para todo $i \in E$ y $t \geq 0$,

$$P[W_t > u | X_t = i] = e^{-\lambda(i)u}, \quad u \geq 0,$$

para algún número $\lambda(i) \in [0, \infty)$.

Por la homogeneidad de X , la probabilidad condicional en cuestión es independiente de t . Para una i fija, se denota como $f(u)$. Ya que el evento $\{W_t > u + v\}$ es igual al evento $\{W_t > u, W_{t+u} > v\}$, tenemos:

$$\begin{aligned} f(u + v) &= P[W_t > u + v | X_t = i] \\ &= P[W_t > u, W_{t+u} > v | X_t = i] \\ &= P[W_t > u | X_t = i] P[W_{t+u} > v | X_t = i, W_t > u]. \end{aligned}$$

Si $X_t = i$ y $W_t > u$, entonces $X_{t+u} = i$, y el saber esto hace todo el pasado inútil mientras el fin sea predecir el futuro. Entonces el segundo factor del lado derecho es igual a

$$P[W_{t+u} > v | X_{t+u} = i] = f(v).$$

La función f está acotada entre 0 y 1, y por lo anterior también se muestra que satisface $f(u + v) = f(u)f(v)$ para todo $u, v \geq 0$. Entonces, debe ser de la forma $f(u) = e^{-cu}$ para alguna constante $c \geq 0$; esta constante c podría depender del estado i fijo.

En términos de los tiempos de espera W_t desde t hasta el instante del siguiente cambio de estado, definimos

$$T_0 = 0; \quad T_{n+1} = T_n + W_{T_n}, \quad n \in N$$

$$Y_n = X_{T_n} \quad n \in N.$$

Los tiempos T_0, T_1, T_2, \dots son los instantes de transición para el proceso X , y Y_0, Y_1, Y_2, \dots son los estados sucesivos visitados por X . Si $Y_n = i$, el intervalo $[T_n, T_{n+1})$ se dice que es un intervalo de estancia en el estado i .

A continuación se muestra que $\{Y_n; n \in N\}$ es una cadena de Markov, y $T_{n+1} - T_n$ tiene una distribución exponencial con parámetro dependiente de Y_n .

Para cualquier $n \in N$, $j \in E$, y $u \in R^+$ tenemos

$$P[Y_{n+1} = j, T_{n+1} - T_n > u | Y_0, \dots, Y_n; T_0, \dots, T_n] = q_{ij} e^{-\lambda(i)u},$$

si $\{Y_n = i\}$ ocurre. Aquí, $q_{ij} \geq 0$, $q_{ii} = 0$, y $\sum_j q_{ij} = 1$.

Supongamos que n , j , y u están fijos. Primero se observa que T_n es un tiempo de paro: esto se sigue de que $T_n \leq t$ si y sólo si hay $0 < s_1 < \dots < s_n \leq t$ tal que $X_0 \neq X_{s_1} \neq \dots \neq X_{s_n}$, y la existencia de tal s_1, \dots, s_n puede ser determinado una vez que X_s es conocido para todo $s \leq t$.

Hacemos $T = T_n$ por simplicidad. Entonces $T_{n+1} - T_n = W_T$ y $Y_{n+1} = X_{T+W_T}$. Claramente, conocer $Y_0, \dots, Y_n, T_0, \dots, T_n$ es equivalente a conocer X_t para toda $t \leq T_n = T$. Entonces, el lado izquierdo de la igualdad es

$$P[X_{T+W_T} = j, W_T > u | X_t; t \leq T].$$

Por la propiedad fuerte de Markov, resulta que $g(X_T) = g(Y_n)$, donde

$$g(i) = P[X_{W_0} = j, W_0 > u | X_0 = i].$$

Ahora se puede escribir de la siguiente manera:

$$g(i) = P[W_0 > u | X_0 = i] P[X_{W_0} = j | X_0 = i, W_0 > u].$$

Como se mostró anteriormente $P[W_0 > u | X_0 = i] = e^{-\lambda(i)u}$.

Por otro lado, el evento $\{X_0 = i, W_0 > u\}$ es el mismo que $\{Y_s = i; s \leq u\}$ y dado que $W_0 > u$ tenemos $W_0 = u + W_u$. Entonces,

$$\begin{aligned} P[X_{W_0} = j | X_0 = i, W_0 > u] &= P[X_{u+W_u} = j | X_s = i; s \leq u] \\ &= P[X_{u+W_u} = j | X_u = i] \\ &= P[X_{W_0} = j | X_0 = i] \\ &= q_{ij} \end{aligned}$$

que es independiente de u , Lo anterior nos lleva a que $g(i) = q_{ij} e^{-\lambda(i)u}$.

La secuencia Y_0, Y_1, \dots de estados visitados sucesivamente forma una cadena de Markov con matriz de transición $\{q_{ij}\}_{i,j \in E}$.

De lo anterior se obtiene que $P[T_{n+1} - T_n > u | Y_n = i, Y_{n+1} = j] = e^{-\lambda(i)u}$ es independiente de j ; en otras palabras, la distribución de un tiempo de estancia depende solamente del estado en el que se encuentra y no del siguiente estado.

Para todo $n \in \mathbb{N}$, $i_0, \dots, i_n \in E$, y $u_0, \dots, u_n \in \mathbb{R}^+$ tenemos que

$$P[T_1 - T_0 > u_1, \dots, T_n - T_{n-1} > u_n | Y_0 = i_0, \dots, X_n = i_n] = e^{-\lambda(i_0)u_1} \dots e^{-\lambda(i_{n-1})u_n}.$$

Los tiempos entre saltos son condicionalmente independientes dado el sucesivo estado visitado, y cada estancia tiene una distribución exponencial con el parámetro dependiente del estado que se está visitando.

1.3 La Función de Verosimilitud

Sea X una proceso de Markov en un espacio de estados finito $E = \{1, \dots, m\}$ y una matriz de intensidades $\Lambda = \{\lambda_{ij}\}_{i,j \in E}$. Si X ha sido observada continuamente en el intervalo de tiempo $[0, \tau]$, se puede obtener una estimación de Λ por máxima verosimilitud.

$$\text{Sea } \Lambda = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1m} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2m} \\ \dots & \dots & \dots & \dots \\ \lambda_{m1} & \lambda_{m2} & \dots & \lambda_{mm} \end{bmatrix}.$$

Sabemos que $-\lambda_{ii}$ es la tasa total de salida del estado i y que $\lambda_{ii} = -\sum_{j \neq i} \lambda_{ij}$.

Además la probabilidad de saltar del estado i al estado j está dada por $-\frac{\lambda_{ij}}{\lambda_{ii}}$.

Entonces la función de verosimilitud está conformada por la probabilidad de mantenerse un tiempo específico en un estado i y la probabilidad de saltar de ese estado i al siguiente estado j :

$$\left(-\lambda_{ii} e^{\lambda_{ii} x_i^{(h)}}\right) \left(-\frac{\lambda_{ij}}{\lambda_{ii}}\right) = e^{\lambda_{ii} x_i^{(h)}} \lambda_{ij} = e^{\left(-\sum_{j \neq i} \lambda_{ij}\right) x_i^{(h)}} \lambda_{ij},$$

donde $x_i^{(h)}$ es el tiempo de estancia en el estado i en la h -ésima visita a ese estado. Sea $R_i(\tau) = \sum_h x_i^{(h)}$ el tiempo total de estancia en el estado i y $N_{ij}(\tau)$ el número total de saltos de i a j en el intervalo de tiempo $[0, \tau]$. Entonces al multiplicar las probabilidades de toda la cadena observada hasta el tiempo τ obtenemos:

$$\begin{aligned} L &= \left[\prod_{i=1}^m e^{\left(-\sum_{j \neq i} \lambda_{ij} \right) \left(\sum_h x_i^{(h)} \right)} \right] \left[\prod_{i=1}^m \prod_{j \neq i} \lambda_{ij}^{N_{ij}(\tau)} \right] = \\ &= \left[\prod_{i=1}^m \prod_{j \neq i} e^{-\lambda_{ij} R_i(\tau)} \right] \left[\prod_{i=1}^m \prod_{j \neq i} \lambda_{ij}^{N_{ij}(\tau)} \right] = \\ &= \prod_{i=1}^m \prod_{j \neq i} e^{-\lambda_{ij} R_i(\tau)} \lambda_{ij}^{N_{ij}(\tau)}. \end{aligned}$$

Por lo tanto la función de verosimilitud está dada por:

$$L(\Lambda; X) = \prod_{i=1}^m \prod_{j \neq i} e^{-\lambda_{ij} R_i(\tau)} \lambda_{ij}^{N_{ij}(\tau)}.$$

1.4 Estimadores de Máxima Verosimilitud

Se estimarán los elementos de la matriz de intensidades $\Lambda = \{\lambda_{ij}\}_{i,j \in E}$ por medio del método de máxima verosimilitud. Por propiedades de logaritmo, la función $\ln(L)$ se maximiza en los mismos puntos que L .

Entonces $\ln(L) = \sum_{i=1}^m \sum_{j \neq i} -\lambda_{ij} R_i(\tau) + N_{ij}(\tau) \ln(\lambda_{ij})$. Derivando:

$$\begin{aligned}
 \frac{\partial \ln(L)}{\partial \lambda_{kl}} &= \sum_{i=1}^m \sum_{j \neq i} \frac{\partial (-\lambda_{ij} R_i(\tau) + N_{ij}(\tau) \ln(\lambda_{ij}))}{\partial \lambda_{kl}} = \\
 &= \frac{\partial (-\lambda_{kl} R_k(\tau) + N_{kl}(\tau) \ln(\lambda_{kl}))}{\partial \lambda_{kl}} = \\
 &= -R_k(\tau) + \frac{N_{kl}(\tau)}{\lambda_{kl}}.
 \end{aligned}$$

Igualando las derivadas a cero:

$$\begin{aligned}
 -R_k(\tau) + \frac{N_{kl}(\tau)}{\lambda_{kl}} &= 0 \Rightarrow \frac{N_{kl}(\tau)}{\lambda_{kl}} = R_k(\tau) \\
 \Rightarrow \lambda_{kl} &= \frac{N_{kl}(\tau)}{R_k(\tau)}.
 \end{aligned}$$

Además si se calculara la matriz Hessiana se obtendría que los elementos de la diagonal serían de la forma $\frac{\partial^2 \ln(L)}{\partial \lambda_{kl}^2} = \frac{\partial}{\partial \lambda_{kl}} \left(-R_k(\tau) + \frac{N_{kl}(\tau)}{\lambda_{kl}} \right) = -\frac{N_{kl}(\tau)}{\lambda_{kl}^2}$, mientras que todos los elementos fuera de la diagonal son iguales a cero. Dadas las condiciones anteriores, los valores propios de la matriz Hessiana son iguales a los elementos de la diagonal. Estos elementos son siempre negativos, por lo que la matriz Hessiana es definida negativa, y por lo tanto los puntos encontrados determinan un máximo.

Luego entonces los estimadores de máxima de verosimilitud para la matriz de

intensidades están dadas por $\hat{\lambda}_{kl} = \frac{N_{kl}(\tau)}{R_k(\tau)}$ para $k, l \in E, k \neq l$.

Ahora, suponiendo que se tienen un número finito S de observaciones de series del proceso independientes e idénticamente distribuidas, la función de verosimilitud será de una forma muy similar. Sea $R_i^s(\tau_{s-1}, \tau_s)$ el tiempo total de estancia en el estado i y $N_{ij}^s(\tau_{s-1}, \tau_s)$ el número de saltos del estado i al estado j en la s -ésima observación durante el periodo de tiempo (τ_{s-1}, τ_s) , donde $\tau_0 = 0$ y $\tau_s = \tau$.

Así la función de verosimilitud para cada observación sería

$$L(\Lambda; X_{(s)}) = \prod_{i=1}^m \prod_{j \neq i} \lambda_{ij}^{N_{ij}^s(\tau_{s-1}, \tau_s)} e^{-\lambda_{ij} R_i^s(\tau_{s-1}, \tau_s)}.$$

Debido a que las observaciones son

independientes e idénticamente distribuidas la probabilidad de las observaciones en conjunto sería el producto de las probabilidades, es decir, el producto de las funciones de verosimilitud:

$$\begin{aligned} L(\Lambda; X) &= \prod_{s=1}^S L(\Lambda; X_{(s)}) \\ &= \prod_{s=1}^S \prod_{i=1}^m \prod_{j \neq i} \lambda_{ij}^{N_{ij}^s(\tau_{s-1}, \tau_s)} e^{-\lambda_{ij} R_i^s(\tau_{s-1}, \tau_s)} \\ &= \prod_{i=1}^m \prod_{j \neq i} e^{-\lambda_{ij} \sum_{s=1}^S R_i^s(\tau_{s-1}, \tau_s)} \lambda_{ij}^{\sum_{s=1}^S N_{ij}^s(\tau_{s-1}, \tau_s)} \\ &= \prod_{i=1}^m \prod_{j \neq i} e^{-\lambda_{ij} R_i^T(\tau)} \lambda_{ij}^{N_{ij}^T(\tau)}. \end{aligned}$$

Ya que $(0, \tau) = \bigcup_{s=1}^S (\tau_{s-1}, \tau_s)$, entonces $R_i^T(\tau)$ y $N_{ij}^T(\tau)$ son el tiempo total de estancia en el estado i y el número total de saltos del estado i a j en el intervalo de tiempo total $(0, \tau)$. De esta manera obtenemos que el estimador de máxima

verosimilitud para la matriz de transición es: $\hat{\lambda}_{kl} = \frac{N_{kl}^T(\tau)}{R_k^T(\tau)}$. Aún si las

observaciones no se hicieran de manera continua, este estimador seguiría siendo adecuado.

1.5 Ejemplo

Se simulará un Proceso de Markov a partir de una matriz de intensidades conocida y se efectuará la estimación de la misma por medio de los estimadores de máxima verosimilitud.

Realizar una simulación de un Proceso de Markov dada una matriz de intensidades no es algo difícil. Se sabe que los tiempos de estancia en un estado i

sigue una distribución exponencial con parámetro $-\frac{1}{\lambda_{ii}}$, además sabemos que la

intensidad de saltar de un estado i a uno j está dada por $-\frac{\lambda_{ij}}{\lambda_{ii}}$. Entonces la

simulación de un proceso de Markov consiste en generar variables aleatorias exponenciales con los parámetros dados y generar los saltos con la distribución especificada por la matriz.

Se simulará un proceso con un espacio de estados $E = \{1, \dots, 5\}$. La probabilidad de saltar será igual para todos los estados y los tiempos de estancia tendrán parámetro igual a 1. Esto es que la matriz de intensidades es la siguiente:

$$\Lambda = \begin{bmatrix} -1 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & -1 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & -1 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & -1 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 & -1 \end{bmatrix}.$$

Después de que el proceso fue simulado hasta que hubiese 5000 cambios de estado o lo que es lo mismo 5000 saltos, se obtuvieron los siguientes datos:

$$N = \begin{bmatrix} 0 & 258 & 236 & 257 & 265 \\ 248 & 0 & 218 & 283 & 243 \\ 244 & 223 & 0 & 227 & 253 \\ 253 & 266 & 244 & 0 & 260 \\ 271 & 245 & 249 & 256 & 0 \end{bmatrix}$$

$$R = \begin{bmatrix} 1030.9 \\ 972.0 \\ 911.7 \\ 1070.4 \\ 1023.8 \end{bmatrix}.$$

Donde N es la matriz con elementos $(N_{ij})_{i,j \in E}$ y R es el vector columna con elementos $(R_i)_{i \in E}$. Entonces la matriz de intensidades estimada por máxima verosimilitud es:

$$\hat{\Lambda} = \begin{bmatrix} -0.9855 & 0.2503 & 0.2289 & 0.2493 & 0.2571 \\ 0.2551 & -1.0206 & 0.2243 & 0.2912 & 0.2500 \\ 0.2676 & 0.2446 & -1.0386 & 0.2490 & 0.2775 \\ 0.2364 & 0.2485 & 0.2280 & -0.9557 & 0.2429 \\ 0.2647 & 0.2393 & 0.2432 & 0.2500 & -0.9973 \end{bmatrix}.$$

De esto obtenemos las probabilidades de saltar de un estado a otro:

i a j	1	2	3	4	5
1	0.0000	0.2539	0.2323	0.2530	0.2608
2	0.2500	0.0000	0.2198	0.2853	0.2450
3	0.2577	0.2355	0.0000	0.2397	0.2672
4	0.2473	0.2600	0.2385	0.0000	0.2542
5	0.2654	0.2400	0.2439	0.2507	0.0000

Aproximadamente una distribución uniforme, es decir muy parecida a la distribución equiprobable que se tenía en un principio.

CAPÍTULO 2

METODOLOGÍA MONTE CARLO VÍA CADENAS DE MARKOV

La metodología Monte Carlo vía Cadenas de Markov (MCMC por sus siglas en inglés) permite un modelaje estadístico realista. Hasta recientemente, conocer la gran complejidad y estructura en muchas aplicaciones era difícil y requería el desarrollo de metodología y software específicos. La alternativa era unificar el problema en una estructura simple de un método posible. Los métodos MCMC proveen una estructura en la que problemas complejos pueden ser analizados usando software genérico.

MCMC es esencialmente integración de Monte Carlo usando cadenas de Markov. Bayesianos, y algunas ocasiones también frecuentistas, necesitan integrar sobre distribuciones de probabilidad complejas para hacer inferencia sobre parámetros de un modelo o predicciones. La integración de Monte Carlo obtiene muestras de la distribución requerida y después calcula medias muestrales para aproximar esperanzas. MCMC obtiene estas muestras simulando una cadena de Markov especialmente construida por un largo tiempo. La estructura general de Metropolis y Hastings es una manera de generar estas cadenas.

2.1 Inferencia Bayesiana

La mayoría de las aplicaciones de MCMC están orientadas hacia la inferencia Bayesiana. Desde una perspectiva Bayesiana, no hay una distinción entre observaciones y parámetros ya que ambos son considerados cantidades aleatorias. Denotemos por D los datos observados, y θ los parámetros del modelo y datos faltantes. La inferencia formal requiere establecer una distribución conjunta $P(D, \theta)$ para los valores aleatorios. Esta distribución conjunta incluye dos partes:

una distribución *a priori* $P(\theta)$ y una verosimilitud $P(D|\theta)$. Especificando $P(\theta)$ y $P(D|\theta)$ se tiene un modelo probabilístico, en donde

$$P(D, \theta) = P(D|\theta)P(\theta).$$

Teniendo observaciones D , el teorema de Bayes es aplicado para determinar la distribución de θ dado D :

$$P(\theta|D) = \frac{P(\theta)P(D|\theta)}{\int P(\theta)P(D|\theta)d\theta}.$$

Esta es llamada la distribución *a posteriori* de θ , y es el objetivo de toda inferencia Bayesiana.

Cualquier uso de la distribución *a posteriori* es legítimo para la inferencia Bayesiana: momentos, cuantiles, etc. Estas cantidades pueden ser expresadas en términos de esperanzas *a posteriori* de funciones de θ . La esperanza *a posteriori* de una función f es

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

Las integrales en la expresión anterior han sido la fuente de la mayoría de las dificultades en la práctica de la inferencia Bayesiana, especialmente en muchas dimensiones. En muchas aplicaciones, la evaluación de $E[f(\theta)|D]$ es imposible. Propuestas alternativas incluyen evaluación numérica, lo cual es difícil e impreciso en más de 20 dimensiones; aproximaciones analíticas; e integración de Monte Carlo, incluyendo MCMC.

2.1.1 Cálculo de Esperanzas

El problema de calcular esperanzas en distribuciones de varias dimensiones también ocurre en áreas de inferencia frecuentista. Sea X un vector de k variables aleatorias, con distribución $\pi(\cdot)$. En aplicaciones Bayesianas, X incluye parámetros del modelo y datos faltantes; en aplicaciones frecuentistas, puede incluir datos o efectos aleatorios. Para la inferencia Bayesiana, $\pi(\cdot)$ será una distribución *a posteriori*. De cualquier modo, el objetivo es evaluar la esperanza

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

para las funciones $f(\cdot)$ de interés. Regularmente $\int \pi(x)dx$ es desconocida. Esta es una situación común en la práctica, por ejemplo, en la inferencia Bayesiana se sabe que $P(\theta|D) \propto P(\theta)P(D|\theta)$, pero no es sencillo evaluar la constante $\int P(\theta)P(D|\theta)d\theta$ para normalizar.

Se introduce MCMC como un método para evaluar expresiones de la forma

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

2.2 Integración de Monte Carlo

La integración de Monte Carlo evalúa $E[f(X)]$ obteniendo muestras $\{X_t, t = 1, \dots, n\}$ de una distribución $\pi(\cdot)$ y después aproximando $E[f(X)] \approx \frac{1}{n} \sum_{t=1}^n f(X_t)$.

Así, la media poblacional de $f(X)$ es estimada por una media muestral. Cuando las muestras de $\{X_t\}$ son independientes, las leyes de los grandes números

aseguran que la aproximación puede hacerse tan precisa como se desee, incrementando el tamaño n .

En general, generar $\{X_t\}$ independientes de $\pi(\cdot)$ no es viable. Sin embargo las muestras $\{X_t\}$ no deben ser necesariamente independientes. $\{X_t\}$ puede ser generada por cualquier proceso que saque muestras siempre sobre $\pi(\cdot)$ en las proporciones correctas. Una manera de hacer esto, es mediante una cadena de Markov apropiada en la que $\pi(\cdot)$ sea su distribución estacionaria. Esto es MCMC.

2.3 Cadenas de Markov

Supongamos que se genera una sucesión de variables aleatorias, $\{X_0, X_1, X_2, \dots\}$, tal que a cada tiempo $t \geq 0$, el siguiente estado X_{t+1} se obtiene de una distribución $P[X_{t+1}|X_t]$ que depende solamente del estado actual de la cadena, X_t . Esto es que dado X_t , el siguiente estado no depende de la historia de la cadena, $\{X_0, X_1, X_2, \dots, X_{t-1}\}$. Esto es una cadena de Markov, y asumiendo que la cadena es homogénea, entonces $P[\cdot|\cdot]$ no depende de t .

Cómo afecta el punto inicial X_0 a X_t concierne a la distribución de X_t dado X_0 , la cual denotamos por $P^{(t)}[X_t|X_0]$. Ya que no se tiene la información intermedia $\{X_1, X_2, \dots, X_{t-1}\}$, entonces X_t depende de X_0 . Sujeto a condiciones de regularidad, la cadena gradualmente ‘olvidará’ su estado inicial y $P^{(t)}[\cdot|X_0]$ eventualmente converge a una única distribución estacionaria que no depende ni de t o X_0 . Sea $\phi(\cdot)$ la distribución estacionaria. Entonces mientras aumente t los puntos $\{X_t\}$ parecerán cada vez más muestras de $\phi(\cdot)$.

Después de un suficiente número de iteraciones m , $\{X_t; t = m + 1, \dots, n\}$ serán aproximadamente muestras de $\phi(\cdot)$. Ahora ya se puede utilizar la cadena de Markov para estimar $E[f(X)]$, ya que X tiene una distribución $\phi(\cdot)$. Esto nos da un estimador de la siguiente forma

$$\bar{f} = \frac{1}{n - m} \sum_{t=m+1}^n f(X_t).$$

El cual es llamado un promedio ergódico. Esto debido a que la convergencia de dicha esperanza se fundamenta en el teorema ergódico.

2.4 El algoritmo Metropolis-Hastings

El promedio ergódico muestra cómo una cadena de Markov puede usarse para estimar $E[f(X)]$, donde la esperanza es tomada de su distribución estacionaria $\phi(\cdot)$. Esto parecería ser la solución para el problema, pero primero se necesita saber cómo construir una cadena de Markov tal que su distribución estacionaria $\phi(\cdot)$ sea precisamente la distribución de interés $\pi(\cdot)$.

Para el algoritmo Metropolis-Hastings, a cada tiempo t , el siguiente estado X_{t+1} se escoge primero generando un punto candidato Y de una distribución propuesta $q(\cdot|X_t)$. El punto candidato Y es entonces aceptado con probabilidad $\alpha(X_t, Y)$ donde

$$\alpha(X, Y) = \min\left(1, \frac{\pi(Y)q(X|Y)}{\pi(X)q(Y|X)}\right).$$

Si el punto candidato es aceptado, el siguiente estado es $X_{t+1} = Y$. Si se rechaza, la cadena no se mueve, $X_{t+1} = X_t$. Entonces el algoritmo Metropolis-Hastings es bastante simple:

Inicializa X_0 ; sea $t=0$

Repetir {

 Generar un punto Y de $q(\cdot|X_t)$

 Generar una variable Uniforme(0,1) U

 Si $U \leq \alpha(X_t, Y)$ entonces $X_{t+1} = Y$

 En otro caso $X_{t+1} = X_t$

 Incrementar t

}.

La distribución propuesta $q(\cdot|.)$ puede tener esencialmente cualquier forma y la distribución estacionaria de la cadena será $\pi(\cdot)$. Esto se ve en la probabilidad de transición de la cadena:

$$P(X_{t+1}|X_t) = q(X_{t+1}|X_t)\alpha(X_t, X_{t+1}) + I(X_{t+1} = X_t) \left[1 - \int q(Y|X_t)\alpha(X_t, Y)dY \right],$$

donde $I(\cdot)$ es una función indicadora. El primer término viene de la aceptación de que $Y = X_{t+1}$, y el segundo término se refiere al rechazo de todos los posibles valores Y . Usando el hecho de que

$$\pi(X_t)q(X_{t+1}|X_t)\alpha(X_t, X_{t+1}) = \pi(X_{t+1})q(X_t|X_{t+1})\alpha(X_{t+1}, X_t)$$

se obtiene que

$$\pi(X_t)P(X_{t+1}|X_t) = \pi(X_{t+1})P(X_t|X_{t+1}).$$

Integrando ambas partes con respecto a X_t , se tiene:

$$\int \pi(X_t)P(X_{t+1}|X_t)dX_t = \pi(X_{t+1}).$$

El lado izquierdo de la ecuación da la distribución marginal X_{t+1} suponiendo que X_t se distribuye $\pi(\cdot)$. La igualdad nos dice que si X_t se distribuye $\pi(\cdot)$, entonces X_{t+1} también. Entonces, una vez que se ha obtenido un valor de la distribución estacionaria, todos los valores subsecuentes que se generen tendrán esa distribución.

2.5 Formas canónicas de distribuciones iniciales

Cualquier distribución inicial $q(\cdot)$ finalmente obtendrá observaciones de la distribución objetivo $\pi(\cdot)$. Sin embargo, la tasa de convergencia a la distribución estacionaria dependerá crucialmente en la relación que exista entre $q(\cdot)$ y $\pi(\cdot)$. En problemas de muchas dimensiones, frecuentemente es necesario realizar análisis exploratorios para determinar a grosso modo la forma y orientación de $\pi(\cdot)$. Esto ayudará en la construcción de una inicial $q(\cdot)$, lo cual llevará a una rápida convergencia. El progreso en la práctica frecuentemente depende en la experimentación y construcción, aunque iniciales desatinadas pueden llegar a trabajar sorprendentemente bien. Para la eficiencia computacional, $q(\cdot)$ debería ser seleccionada de manera que se puedan obtener observaciones y evaluaciones de manera sencilla.

2.5.1 El *Independence sampler*

El *Independence Sampler* es una forma del algoritmo *Metropolis-Hastings*, en el cual $q(Y|X) = q(Y)$ no depende de X . Para este caso, la probabilidad de aceptación toma la forma

$$\alpha(X, Y) = \min\left(1, \frac{w(Y)}{w(X)}\right)$$

donde $w(X) = \frac{\pi(X)}{q(X)}$.

En general, este algoritmo puede trabajar muy bien o muy mal. Para que trabaje bien, $q(\cdot)$ debe ser una buena aproximación a $\pi(\cdot)$, pero es más seguro si $q(\cdot)$ tiene una cola más pesada que $\pi(\cdot)$. Para mostrar esto, supóngase que $q(\cdot)$ es de cola más ligera que $\pi(\cdot)$, y que X_t se encuentra en la cola de $\pi(\cdot)$. La mayoría de los candidatos no estarán en la cola, entonces $w(X_t)$ será mucho más grande que $w(Y)$ dando una baja probabilidad de aceptación. De este modo las distribuciones iniciales independientes con colas pesadas ayudan a evadir largos periodos en la cola.

2.5.2 El *Gibbs Sampling*

Para este algoritmo, la distribución para probabilidad de aceptación está dada por $q(Y|X) = \pi(Y)$. Substituyendo esto, se obtiene que la probabilidad de aceptación es 1; esto es, que los candidatos del *Gibbs Sampling* siempre se aceptan.

2.6 Cadenas de Markov en Espacio de Estados Generales

Se asume que la distribución π está definida en un conjunto E llamado espacio de estados. Frecuentemente E puede encontrarse en R^k , pero pueden ser espacios generales, con la restricción de que la colección formada por los subconjuntos de E sea una sigma álgebra.

Para una distribución ν y una función real h evaluada en E , la esperanza de h bajo ν se denota por

$$\nu h = \int h(x)\nu(dx),$$

donde $\nu(dx)$ es la probabilidad de un pequeño subconjunto medible $dx \in E$; $h(x)$ es el valor de h en dx ; y la integral está definida sobre dicho subconjunto dx . Entonces si ν es discreto con función de probabilidad $f(x)$, se tiene que

$$\nu h = \int h(x)\nu(dx) = \sum h(x)f(x);$$

si ν es continua con densidad $f(x)$, entonces

$$\nu h = \int h(x)\nu(dx) = \int h(x)f(x)dx.$$

La distribución de una cadena de Markov homogénea en tiempo $\{X_n\} = \{X_n, n = 1, 2, \dots\}$ en el espacio de estados E está especificada por su distribución inicial y su kernel de transición. El kernel de transición es una función $P(x, A)$ tal que para cualquier $n \geq 0$

$$P\{X_{n+1} \in A | X_n = x\} = P(x, A)$$

para todo $x \in E$ y $A \in E$. Esto es que $P(x, \cdot)$ es la distribución de la cadena de Markov después de un salto dado que comienza en x . Estrategias diferentes, tales como los algoritmos *Gibbs Sampling*, *Metropolis-Hastings*, etc., dan como resultado diferentes kernels de transición. Los kernels de transición son la versión en espacios generales de las matrices de transición de las cadenas discretas.

Para una distribución de probabilidad ν sobre E , se define la distribución νP como

$$\nu P(A) = \int P(x, A)\nu(dx).$$

La distribución νP es la distribución de la posición de una cadena de Markov con kernel de transición P y distribución inicial ν después de un paso. Una cadena de Markov tiene distribución invariante π si $\pi = \pi P$, también llamada distribución estacionaria.

Para una función real h evaluada en E , la esperanza está definida como

$$E[h(X_1)|X_0 = x] = \int P(x, dy)h(y).$$

El producto PQ de dos kernels de transición se define como

$$PQ(x, A) = \int P(x, dy)Q(y, A)$$

para todo $x, A \in E$, Entonces la n -ésima iteración P^n de P se define recursivamente por $P^n = PP^{n-1}$. Por convención, P^0 es el kernel identidad que asigna una probabilidad igual a 1 para el valor inicial. Usando esta notación, se dice que $P\{X_n \in A|X_0 = x\} = P^n(x, A)$ para cualquier $n \geq 0$.

CAPÍTULO 3

ESTIMACIÓN DE MONTE CARLO VÍA CADENAS DE MARKOV

Se presenta una propuesta para estimar los parámetros de un Proceso de Markov observado discretamente, la cual hace uso de la metodología MCMC. Considérese un Proceso de Markov $\{J(t)\}$ con $p = p_1 + p_2 + \dots + p_m$ estados y matriz de intensidades Λ . Un nuevo proceso $\{X(t)\}$ es definido de la siguiente manera:

$$X(t)=i \Leftrightarrow J(t) \in \{p_{i-1} + 1, \dots, p_i\}, \quad i = 1, 2, \dots, m;$$

donde $p_0 = 0$. De este modo se ha agrupado a los estados de J, y X indica en qué grupo de los estados se encuentra J en un tiempo cualquiera. El proceso $\{X(t)\}$, en general, no es un Proceso de Markov homogéneo, ya que los tiempos de estancia en los estados $1, 2, \dots, m$ no necesariamente son distribuidos exponencialmente.

Considerando observaciones en tiempo discreto de X, el propósito es estimar la matriz de intensidades Λ del proceso de Markov J. Si $p_i = 1$ para todo $i \in E$, entonces se podrán estimar los parámetros de Λ siempre que esta sea única, determinada por la distribución de la cadena de Markov incluida en el proceso. Si algún $p_i > 1$, entonces Λ deja de ser única, y no sería posible estimar todos los parámetros por medio de MCMC, lo cual será evidente de lo siguiente. Sin embargo, será posible hacer estimaciones prácticas que sean invariantes bajo diferentes representaciones. Un ejemplo son las tasas de transición dependientes del tiempo en los diferentes estados $1, 2, \dots, m$ del proceso X.

Descomponiendo la matriz de intensidades del siguiente modo:

$$\Lambda = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} & \dots & \Lambda_{1m} \\ \Lambda_{21} & \Lambda_{22} & \dots & \Lambda_{2m} \\ \dots & \dots & \dots & \dots \\ \Lambda_{m1} & \Lambda_{m2} & \dots & \Lambda_{mm} \end{bmatrix}.$$

Considérense las observaciones discretas $x = (x_1, \dots, x_n)$ del proceso continuo $\{X(t)\}$ observado en tiempos t_1, \dots, t_n hasta el tiempo τ ($t_1 = 0$ y $t_n = \tau$). La estructura es esencialmente Bayesiana. Se escoge una distribución *a priori* $\phi(\Lambda)$ y se está interesado en la distribución condicional de Λ dada la información x . Sin embargo, debemos estudiar el problema más general de encontrar la distribución condicional de (Λ, J) dado x , donde $J = \{J(t)\}_{0 \leq t \leq \tau}$ denota la observación continua de J .

Para este fin se emplea el método *Gibbs Sampling* con dos espacios Λ y J ; muestreando alternativamente la simulación de J dado (Λ, x) y de Λ dado (J, x) (x no es de importancia cuando se condiciona sobre J). Iterando el método resulta en una sucesión de variables (Λ_n, J_n) . Bajo condiciones deseables el método *Gibbs Sampling* finalmente producirá una sucesión estacionaria y ergódica, lo que es, después de descartar un cierto periodo de “calentamiento”, por ejemplo, las primeras $K-1$ iteraciones, la sucesión $(\Lambda_n, J_n)_{n \geq K}$ puede considerarse estacionaria y la distribución estacionaria es exactamente aquella de (Λ, J) dado x .

Si $p_i = 1$ para todo i , entonces por ergodicidad el promedio empírico

$$\frac{1}{N} \sum_{i=K}^{N+K} \Lambda_i$$

converge a la verdadera esperanza de Λ condicionada sobre x . Además, se pueden construir intervalos de credibilidad basados en la distribución empírica de $(\Lambda_n, J_n)_{n \geq K}$, y los cuantiles de dicha distribución pueden ser de interés también.

En situaciones donde Λ no es única con respecto a la distribución de la muestra en tiempo discreto, la esperanza de la distribución *a posteriori* puede no ser una cantidad significativa, aún así, pueden realizarse estimaciones prácticas.

Una elección apropiada de la distribución *a priori* es usualmente esencial para asegurar buenas propiedades de combinación de los datos y una *a posteriori* que no sea dominada por la *a priori*. Algunas veces los hiper-parámetros tendrán que ser especificados para asegurar una combinación satisfactoria; la experiencia muestra, sin embargo, que esto no es necesario para el presente caso. Se escoge la *a priori*,

$$\phi(\Lambda) \propto \prod_{i=1}^n \prod_{j \neq i} q_{ij}^{\alpha_{ij}-1} e^{-q_{ij}\beta_i},$$

donde $\alpha_{ij} > 0, i, j \in E$ y $\beta_i > 0, i \in E$ son constantes conocidas que se escogen convenientemente. Entonces $q_{ij} \sim \Gamma\left(\frac{1}{\beta_i}, \alpha_{ij}\right)$.

Esta familia gamma de distribuciones *a priori* es conjugada para el modelo de observaciones continuas en el intervalo de tiempo $[0, \tau]$, la cual es una familia de procesos exponenciales. De hecho la distribución *a posteriori* es

$$p^*(\Lambda) = L_\tau(\Lambda)\phi(\Lambda) \\ \propto \prod_{i=1}^n \prod_{j \neq i} q_{ij}^{N_{ij} + \alpha_{ij} - 1} e^{-q_{ij}(R_i + \beta_i)},$$

donde L_τ es la función de verosimilitud. La simulación de J dado (Λ, x) se realiza simulando un proceso de Markov salto a salto en los intervalos $[t_k, t_{k+1}]$ tomando una condición inicial $X(t_k) = i$ tal que el proceso se encuentre en un estado $X(t_{k+1}) = j$, por ejemplo, en el tiempo t_{k+1} . Esto puede hacerse por rechazo, lo cual resulta bastante eficiente hasta en grandes dimensiones. Muestrear de una distribución de Λ dado (J, x) es simplemente simular de la distribución *a posteriori*.

3.1 Simulación

Retomando el ejemplo del Capítulo 1, se simulará un proceso de Markov con un espacio de estados $E = \{1, \dots, 5\}$ y la siguiente matriz de intensidades:

$$\Lambda = \begin{bmatrix} -1 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & -1 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & -1 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & -1 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 & -1 \end{bmatrix}$$

Así mismo, los tiempos de estancia en cada uno de los estados tienen una distribución exponencial con parámetro 1 y además probabilidades de saltar de un estado a otro son iguales en todos los casos:

i a j	1	2	3	4	5
1	0.00	0.25	0.25	0.25	0.25
2	0.25	0.00	0.25	0.25	0.25
3	0.25	0.25	0.00	0.25	0.25
4	0.25	0.25	0.25	0.00	0.25
5	0.25	0.25	0.25	0.25	0.00

En esta simulación se obtuvieron la matriz de saltos entre los estados y el vector de tiempos totales de estancia en un estado específico:

$$N = \begin{bmatrix} 0 & 43 & 58 & 50 & 58 \\ 56 & 0 & 46 & 48 & 42 \\ 49 & 57 & 0 & 58 & 43 \\ 49 & 42 & 64 & 0 & 46 \\ 56 & 50 & 39 & 45 & 0 \end{bmatrix} \quad R = \begin{bmatrix} 196.30 \\ 202.20 \\ 214.06 \\ 196.12 \\ 188.60 \end{bmatrix}.$$

Dados los datos anteriores se pueden estimar los elementos de la matriz de intensidades por máxima verosimilitud, dividiendo la cantidad de saltos entre el tiempo total de estancia en el estado de donde se salta.

$$\hat{\Lambda} = \begin{bmatrix} -1.0647 & 0.2190 & 0.2955 & 0.2547 & 0.2955 \\ 0.2770 & -0.9495 & 0.2275 & 0.2374 & 0.2077 \\ 0.2289 & 0.2663 & -0.09670 & 0.2710 & 0.2009 \\ 0.2498 & 0.2141 & 0.3263 & -1.0248 & 0.2345 \\ 0.2969 & 0.2651 & 0.2068 & 0.2386 & -1.0074 \end{bmatrix}.$$

Esta estimación es buena, pero tiene el inconveniente de que necesita conocer toda la información continua del proceso hasta un tiempo preciso. Debido a este problema se recurre a métodos de estimación como el de Monte Carlo vía Cadenas de Markov, el cual se implementará para obtener una estimación de la matriz de intensidades.

3.2 Estimación

Ahora se discretiza el proceso de Markov. Dicha información se separará por espacio de una unidad de tiempo y esto nos dará los datos para implementar el método MCMC. Hecho esto, se obtuvo una discretización de 998 parejas de datos

formadas por el tiempo en el que se discretizó y el estado en el que se encontraba el proceso en ese momento, de la cual se muestra sólo una parte:

Tiempo	Estado	Tiempo	Estado	Tiempo	Estado	Tiempo	Estado
0	5	25	4	50	4	75	4
1	3	26	2	51	4	76	1
2	4	27	3	52	2	77	4
3	3	28	3	53	5	78	4
4	3	29	3	54	3	79	3
5	4	30	2	55	5	80	3
6	5	31	3	56	1	81	5
7	2	32	3	57	3	82	1
8	1	33	3	58	4	83	1
9	1	34	5	59	2	84	3
10	1	35	5	60	2	85	3
11	3	36	5	61	2	86	2
12	1	37	1	62	4	87	3
13	5	38	5	63	5	88	4
14	5	39	2	64	2	89	2
15	5	40	5	65	3	90	4
16	3	41	5	66	3	91	4
17	5	42	3	67	2	92	4
18	5	43	3	68	1	93	1
19	2	44	5	69	5	94	3
20	1	45	2	70	1	95	2
21	1	46	2	71	4	96	2
22	5	47	2	72	4	97	1
23	1	48	4	73	2	98	2
24	3	49	4	74	5	99	5

Lo que se pretende con el método MCMC es simular un proceso de Markov que pase por todos los puntos determinados por la información discreta con la que se cuenta, basándose en una simulación en los espacios donde falta la información, comenzando en el punto que se tiene y rechazando simulaciones hasta que se cumpla que termine en el punto final especificado. Por ejemplo, en el tiempo $t=0$ el proceso se encuentra en el estado 5, se calculan las probabilidades de transición basándose en la matriz de intensidades *a priori* y se simulan los saltos con respecto a esta distribución, los tiempos de estancia se simulan como una distribución exponencial con el respectivo parámetro dependiendo del estado del que haya saltado. Si al haber hecho esto la simulación del proceso al tiempo $t=1$

no se encuentra en el estado 3, esta simulación se elimina y se repite hasta que se cumpla dicha condición. Así se sigue para cada bloque hasta llegar al tiempo final del que se cuenta información.

Lo anterior se realiza durante un número específico de iteraciones. Tomando en la primera iteración una matriz de intensidades aleatoria en la que las entradas

$\lambda_{ij} \sim \Gamma\left(\frac{1}{\beta_i}, \alpha_{ij}\right)$ $i, j \in E$ $i \neq j$ tienen una distribución gamma con todos los

parámetros iguales a 1 o lo que es lo mismo toman una distribución exponencial con parámetro 1, los elementos de la diagonal son también aleatorios pues están determinados por los demás elementos de su respectiva fila. Esta matriz es la que

se toma por *a priori*. Habiendo simulado el proceso se obtienen N_{ij} $i, j \in E$ y

R_i $i \in E$ que son el número de saltos del estado i al j y el tiempo total de estancia en el estado i respectivamente, y con esto se generan las nuevas variables aleatorias que serán los elementos de la matriz *a posteriori*. Entonces las entradas

$\lambda_{ij} \sim \Gamma\left(\frac{1}{R_i + \beta_i}, N_{ij} + \alpha_{ij}\right)$ $i, j \in E$ $i \neq j$ toman nuevamente una distribución gamma,

pero dado que $\alpha_{ij} = \beta_i = 1$ se tiene que $\lambda_{ij} \sim \Gamma\left(\frac{1}{R_i + 1}, N_{ij} + 1\right)$ $i, j \in E$ $i \neq j$. La matriz

a posteriori de la iteración anterior se convierte en la matriz *a priori* para la siguiente iteración.

Para la simulación de este ejemplo se tomó la siguiente matriz inicial generada aleatoriamente con las distribuciones consideradas:

$$M_0 = \begin{bmatrix} -1.7936 & 0.3597 & 0.7001 & 0.4152 & 0.3187 \\ 0.3824 & -2.7345 & 0.1056 & 1.0730 & 1.1735 \\ 1.1948 & 0.1508 & -2.7606 & 1.2388 & 0.1761 \\ 0.6131 & 0.1582 & 0.4386 & -1.7755 & 0.5655 \\ 1.8913 & 0.5216 & 0.2009 & 0.6272 & -3.2411 \end{bmatrix}.$$

Después de haber hecho la simulación tomando a esta como la matriz de intensidades y obtenido los saltos totales de i a j (N_{ij} $i, j \in E$), así como el tiempo total de estancia en el estado i (R_i $i \in E$), se generan valores aleatorios con una distribución gamma $\lambda_{ij} \sim \Gamma\left(\frac{1}{R_i+1}, N_{ij}+1\right)$ $i, j \in E$ $i \neq j$ para todos los elementos de la matriz, a excepción de la diagonal. Esta matriz se tomará como la nueva matriz de intensidades, es decir, la matriz *a posteriori* para la iteración actual o *a priori* para la siguiente:

$$M_1 = \begin{bmatrix} -2.2357 & 0.5942 & 0.9003 & 0.4439 & 0.2974 \\ 0.3472 & -1.9779 & 0.0986 & 0.6264 & 0.9057 \\ 0.8430 & 0.0937 & -2.1215 & 1.0463 & 0.1385 \\ 0.7077 & 0.1425 & 0.6076 & -2.1235 & 0.6657 \\ 1.6834 & 0.6167 & 0.2775 & 0.5007 & -3.0784 \end{bmatrix}.$$

Así sucesivamente la matriz *a posteriori* de una iteración se convertirá en la *a priori* para la siguiente. Para la estimación se tomarán los valores generados aleatoriamente en cada iteración, sin tomar en cuenta un “calentamiento”, para aminorar la variación de los datos. El “calentamiento” se toma como el tiempo en que la distribución de la cadena de Markov del método se estaciona y este es el momento en que las observaciones que se toman siguen esta distribución estacionaria de la cadena. De este modo se puede hacer una estimación de la distribución, o en este caso, de la matriz de intensidades.

Sea $\lambda_{ij}^{(k)}$ $i, j \in E$ $i \neq j$ la observación del elemento (i, j) en la k -ésima iteración. Entonces la estimación de cada elemento de la matriz se hará como un promedio ergódico de las observaciones con las que se cuenta:

$$\bar{\lambda}_{ij} = \frac{1}{n - m} \sum_{k=m+1}^n \lambda_{ij}^{(k)} .$$

Regresando al ejemplo en cuestión, después de iterar el método en 6000 ocasiones, se cuenta con sucesión de datos de 6000 observaciones para cada una de las entradas de la matriz, aunque obviamente la diagonal de la matriz de intensidades puede ser fácilmente determinada por los demás elementos de la fila correspondiente. Por lo anterior se tomó un “calentamiento” de 500 iteraciones, determinado por medio de la gráfica de las iteraciones, que se muestran más adelante. El estimador respectivamente para cada elemento de la matriz se calculó sumando desde la observación 501 hasta la 6000 para cada entrada (i , j) y dividiendo cada suma entre 5500:

$$\bar{\lambda}_{ij} = \frac{1}{5500} \sum_{k=501}^{6000} \lambda_{ij}^{(k)} .$$

Hecho lo anterior se obtuvo la siguiente estimación de la matriz de intensidades:

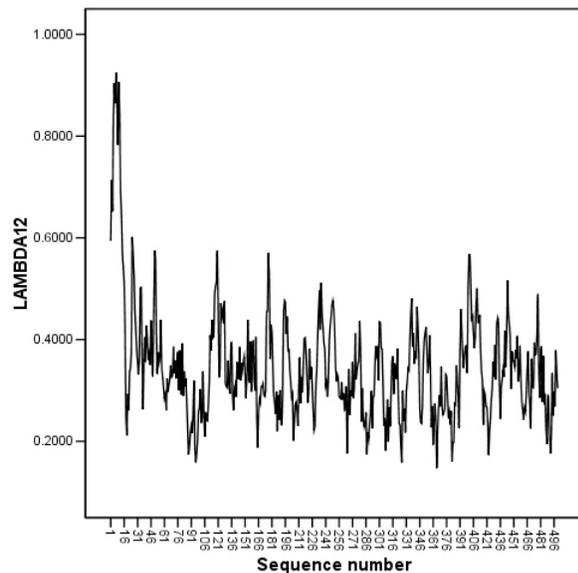
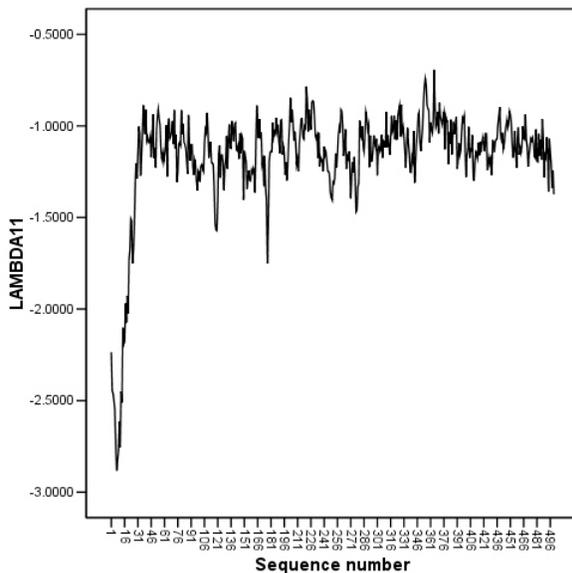
$$\hat{\Lambda} = \begin{bmatrix} -1.1028 & 0.3472 & 0.2123 & 0.3118 & 0.2315 \\ 0.2576 & -1.0323 & 0.2201 & 0.2742 & 0.2804 \\ 0.0884 & 0.2807 & -1.0417 & 0.3895 & 0.2831 \\ 0.3476 & 0.2398 & 0.4810 & -1.3940 & 0.3256 \\ 0.4252 & 0.2644 & 0.2468 & 0.3702 & -1.3066 \end{bmatrix} .$$

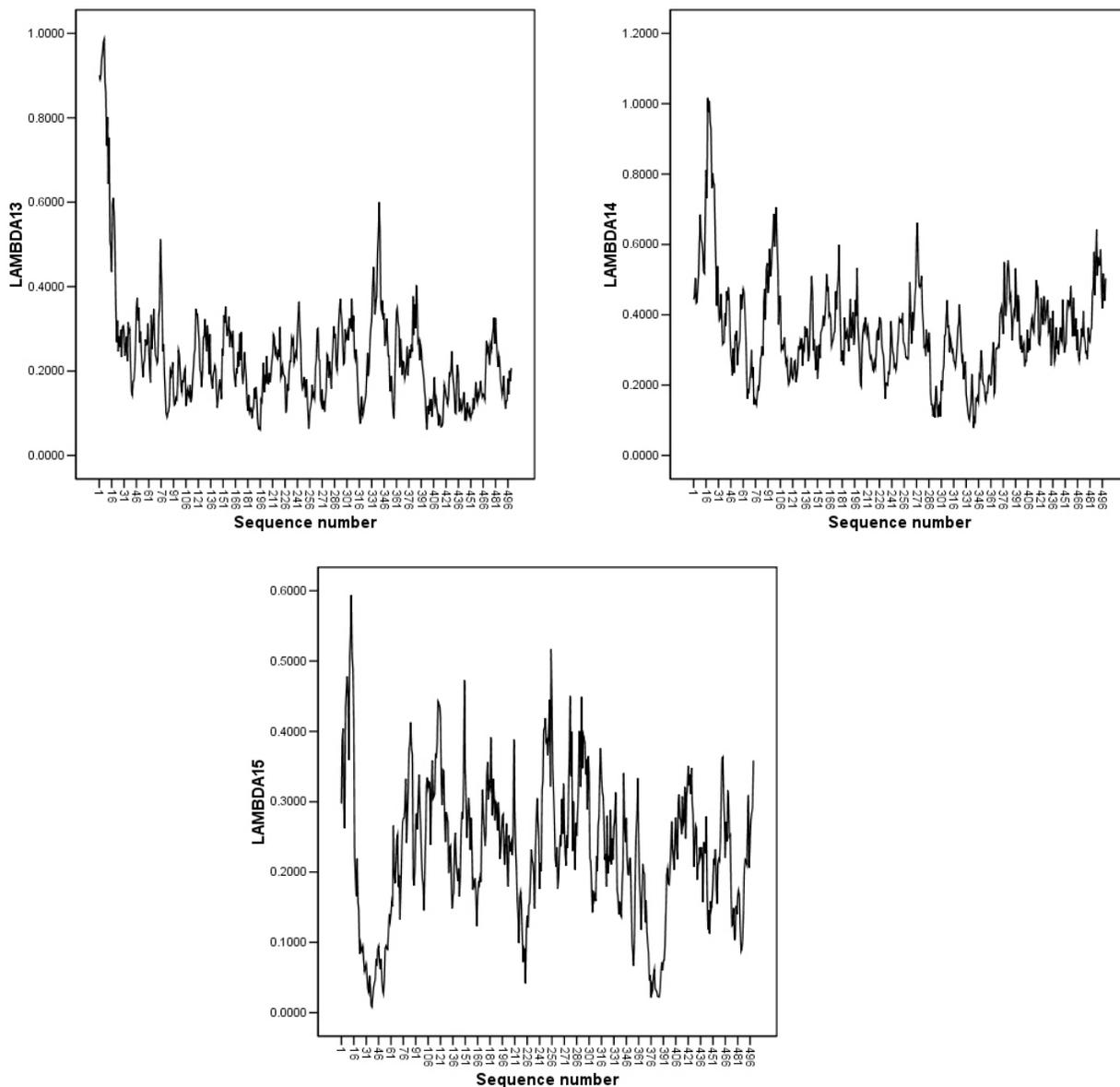
De la cual se obtienen las siguientes probabilidades de saltar del estado i al j:

i a j	1	2	3	4	5
1	0.0000	0.3149	0.1925	0.2828	0.2099
2	0.2496	0.0000	0.2132	0.2656	0.2716
3	0.0849	0.2695	0.0000	0.3739	0.2718
4	0.2494	0.1720	0.3451	0.0000	0.2336
5	0.3254	0.2024	0.1889	0.2834	0.0000

Esta no es la distribución uniforme que se tenía y que se podía esperar, pero aún así los elementos de la diagonal de la matriz de intensidades estimada resultaron ser muy parecidos a las intensidades de los tiempos de estancia que se tenía en un principio, es decir, los elementos de la diagonal son muy cercanos a 1, independientemente de la matriz inicial. En este sentido tenemos una buena estimación de las intensidades para el proceso. En general, podemos decir que la estimación conserva la misma distribución de los tiempos de estancia que la matriz original del proceso.

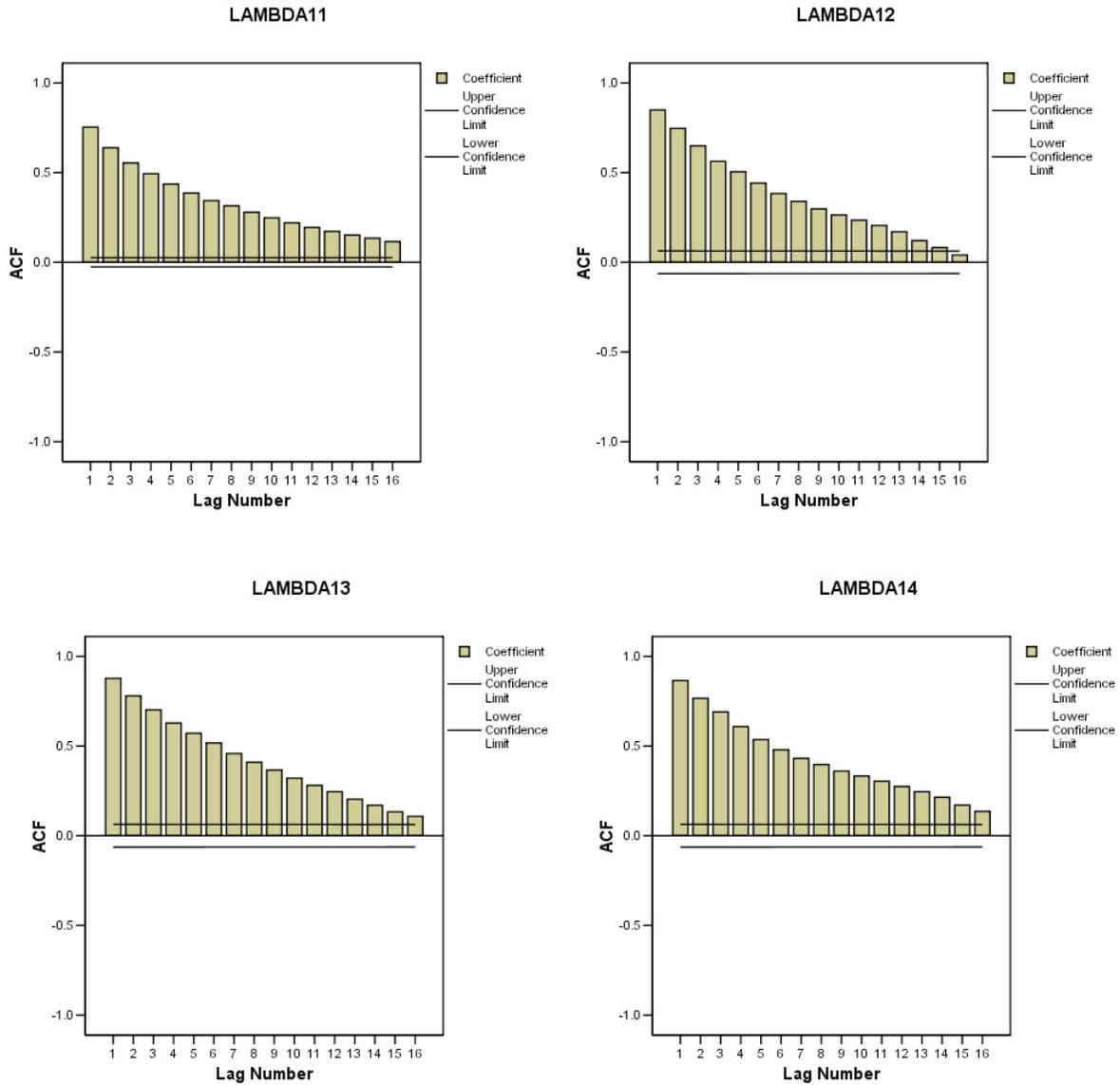
Ahora prestemos atención al “calentamiento” para tener una referencia de la velocidad de convergencia o velocidad para llegar a la distribución estacionaria que se busca con el método Monte Carlo vía Cadenas de Markov. Esto lo estudiaremos con las siguientes gráficas, tomando las primeras 500 observaciones que se definieron como el “calentamiento”.

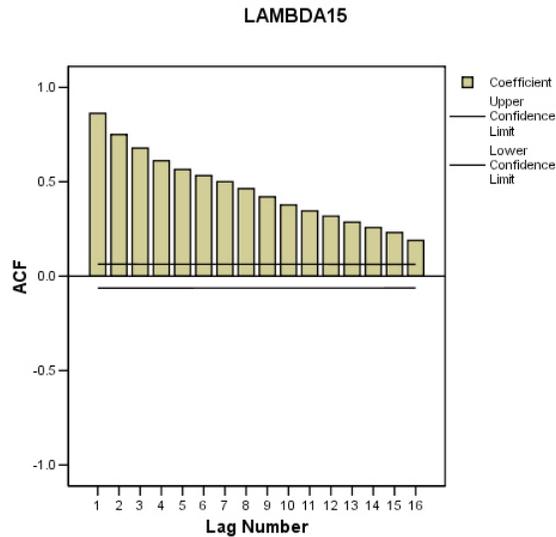




Estas gráficas muestran los valores de las primeras 500 iteraciones para los elementos de la fila 1 de las matrices obtenidas del proceso. Se puede observar que dichos elementos tienen una buena convergencia después de aproximadamente 100 iteraciones, lo cual es una buena velocidad de convergencia en este ejemplo.

Además, también podemos estudiar la autocorrelación del proceso para cada una de las entradas, pero sólo se mostrará lo respectivo a la primera fila de la matriz. A continuación se muestran las gráficas de autocorrelación de dicha fila:





En términos generales se espera que la autocorrelación de una serie se encuentre dentro de las bandas de confianza para poder concluir que la autocorrelación es cero o casi nula, en caso de que esto suceda se puede decir que los datos de la serie según va avanzando el tiempo dejan de depender de observaciones pasadas. En este ejemplo se observa que la autocorrelación va disminuyendo pero ni siquiera entra en las bandas de confianza. Esto es un indicio de que hay problemas en la simulación. El problema podría ser que la matriz de intensidades que determina el proceso que se ajusta a la información discreta no es única y por lo tanto depende de las simulaciones anteriores. Así mismo habría que estudiar más a fondo la razón por la que esto sucede.

Se obtiene que la autocorrelación va disminuyendo en medida que el número de iteraciones crece pero al final no entra en las bandas de confianza. Esto podría ser debido a que el problema de la inclusión está presente y la solución determinada por los datos discretos no es única. Para observar una manera de solucionar esto para el ejemplo anterior, se discretizó el proceso de Markov simulado a espacios de 0.5 unidades de tiempo. Se muestran las primeras 50 unidades de tiempo del proceso de Markov simulado anteriormente, esto con la finalidad de obtener más información del proceso para lograr una mejor estimación de la matriz de transición:

Tiempo	Estado								
0	5	10.5	3	21	1	31.5	3	42	3
0.5	3	11	3	21.5	5	32	3	42.5	3
1	3	11.5	1	22	5	32.5	3	43	3
1.5	4	12	1	22.5	5	33	3	43.5	5
2	4	12.5	5	23	1	33.5	5	44	5
2.5	3	13	5	23.5	2	34	5	44.5	2
3	3	13.5	5	24	3	34.5	5	45	2
3.5	3	14	5	24.5	4	35	5	45.5	2
4	3	14.5	5	25	4	35.5	5	46	2
4.5	4	15	5	25.5	4	36	5	46.5	2
5	4	15.5	3	26	2	36.5	1	47	2
5.5	2	16	3	26.5	3	37	1	47.5	4
6	5	16.5	5	27	3	37.5	5	48	4
6.5	2	17	5	27.5	3	38	5	48.5	4
7	2	17.5	5	28	3	38.5	5	49	4
7.5	1	18	5	28.5	3	39	2	49.5	4
8	1	18.5	2	29	3	39.5	5	50	4
8.5	1	19	2	29.5	3	40	5	50.5	4
9	1	19.5	4	30	2	40.5	5	51	4
9.5	3	20	1	30.5	2	41	5	51.5	4
10	1	20.5	1	31	3	41.5	3	52	2

Ahora se iteró sobre estos datos y se utilizó la misma matriz inicial:

$$M_0 = \begin{bmatrix} -1.7936 & 0.3597 & 0.7001 & 0.4152 & 0.3187 \\ 0.3824 & -2.7345 & 0.1056 & 1.0730 & 1.1735 \\ 1.1948 & 0.1508 & -2.7606 & 1.2388 & 0.1761 \\ 0.6131 & 0.1582 & 0.4386 & -1.7755 & 0.5655 \\ 1.8913 & 0.5216 & 0.2009 & 0.6272 & -3.2411 \end{bmatrix} .$$

Se realizaron, igual que en el ejemplo anterior, 6000 iteraciones y se tomó un “calentamiento” de 500 iteraciones. Se obtuvo la estimación para la matriz de intensidades del mismo proceso de Markov pero ahora con la diferencia de que la distancia entre los datos no era de una unidad de tiempo, sino de 0.5:

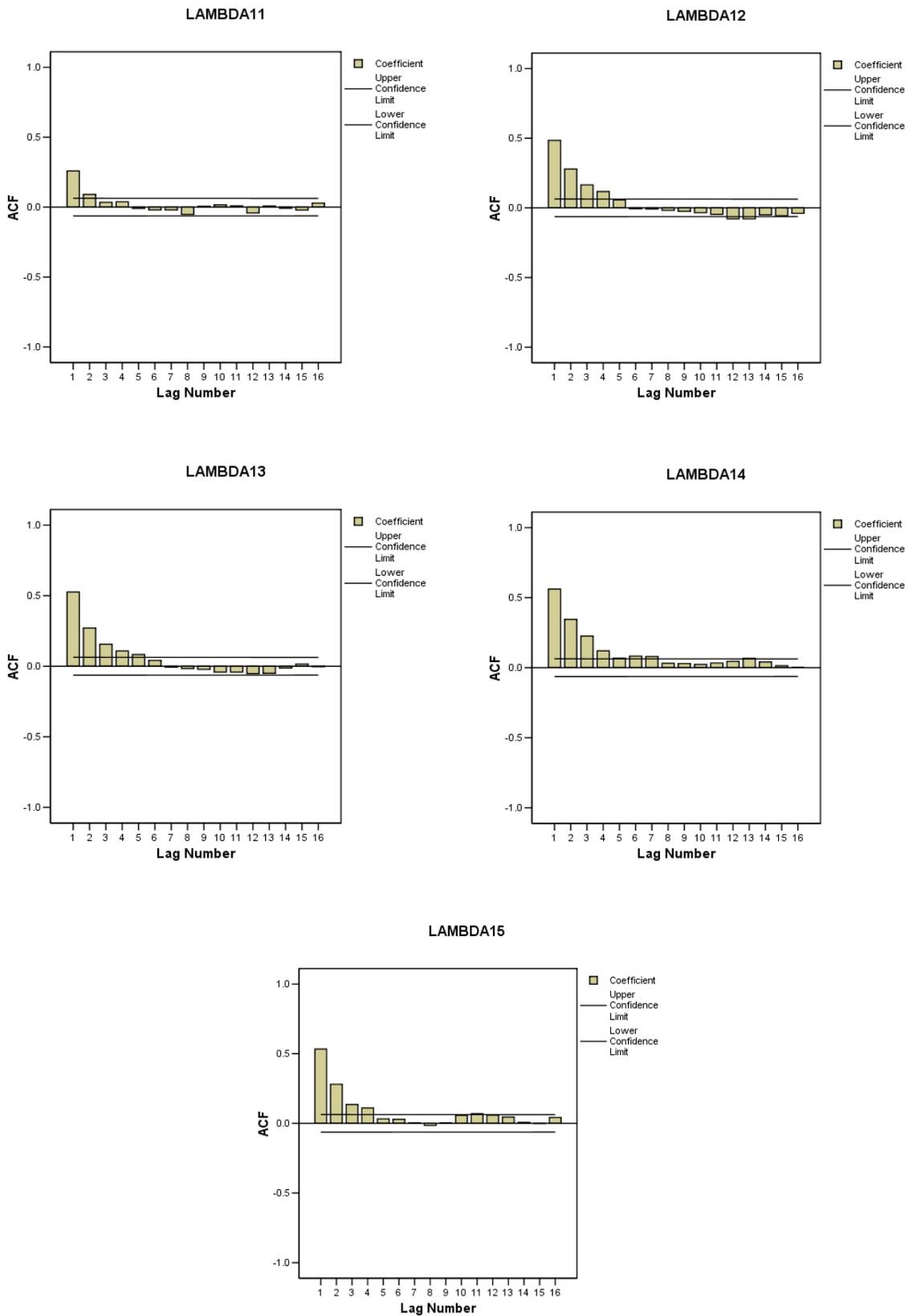
$$\hat{\Lambda} = \begin{bmatrix} -1.0462 & 0.2923 & 0.2471 & 0.2845 & 0.2224 \\ 0.2397 & -0.9964 & 0.2368 & 0.2755 & 0.2445 \\ 0.1814 & 0.2884 & -0.9899 & 0.3114 & 0.2087 \\ 0.2548 & 0.2053 & 0.4227 & -1.1484 & 0.2656 \\ 0.3210 & 0.2490 & 0.1911 & 0.2729 & -1.0341 \end{bmatrix}.$$

De esta se determinan las probabilidades de transición entre estados:

i j	1	2	3	4	5
1	0.0000	0.2794	0.2362	0.2720	0.2126
2	0.2405	0.0000	0.2377	0.2765	0.2453
3	0.1833	0.2913	0.0000	0.3146	0.2109
4	0.2219	0.1788	0.3681	0.0000	0.2313
5	0.3104	0.2408	0.1848	0.2639	0.0000

En este sentido la estimación mejoró en gran medida pues se aproxima más a la distribución original, la cual era uniforme. Como caso específico, la probabilidad de saltar del estado 3 al estado 1 en el ejemplo anterior había sido muy pequeña (.0884) comparada con la real (0.25), en este ejemplo dicha probabilidad mejora pues se aproxima más a la original (0.1833).

La autocorrelación de los datos debió haber mejorado también pues la menor distancia entre las observaciones del proceso original proporciona más información. Se usaron las últimas 1000 iteraciones para obtener las gráficas de autocorrelación. Estos fueron los resultados logrados:



La autocorrelación no sólo disminuyó notablemente desde las primeras observaciones sobre las que se hizo la gráfica, sino que también cayeron dentro de las bandas de confianza. Esto quiere decir que la autocorrelación se vuelve nula con respecto al tiempo, o en este caso, el número de iteraciones del proceso. Al comparar la autocorrelación del ejemplo anterior contra ésta, se observa una gran diferencia pues en el primer el ejemplo la autocorrelación disminuía muy lentamente y nunca quedó dentro de las bandas de confianza, por eso que no pudiera considerarse nula; en este nuevo ejemplo, la autocorrelación entra en las bandas de confianza casi desde el principio dando una mayor seguridad de que la autocorrelación era nula o casi cero. Lo anterior es debido a que la separación entre observaciones al volverse menor permite en todos sentidos una mejor estimación y mayor confiabilidad en los resultados obtenidos.

Más adelante se presenta el problema de la inclusión, en el cual se retomará este mismo proceso de Markov, pero con una distancia mayor entre las observaciones. Esto servirá para ilustrar este problema y darse cuenta de que la solución de un proceso de Markov discretizado no siempre es única.

CAPÍTULO 4

LA INCLUSIÓN

Se tiene interés en hacer inferencia sobre la matriz de intensidades Λ basándose en una muestra de observaciones de X , un proceso de Markov, en tiempos discretos, es decir, $\{X(t_1), \dots, X(t_n)\}$. También para observaciones discretas la función de verosimilitud es en teoría simple. El proceso $Y_i = X(t_i)$ es una cadena de Markov en tiempo discreto, en general no homogénea con respecto al tiempo, para la cual la matriz de transición al tiempo i es $P^{\Delta_i}(\Lambda)$, donde $\Delta_i = t_{i+1} - t_i$ y

$$P^t(\Lambda) = \exp(t\Lambda), \quad t > 0,$$

con $\exp(\cdot)$ denotando la función exponencial de matrices. Entonces la función de verosimilitud para los datos discretos está dada por

$$L_n(\Lambda) = \prod_{i=1}^{n-1} P^{\Delta_i}(\Lambda)_{x_i x_{i+1}}, \quad \Lambda \in \ell$$

donde x_1, \dots, x_n denota los valores observados de X . Para una matriz A se denota la entrada ij por A_{ij} . El conjunto de todas las matrices de intensidades se denota por ℓ . Este es el conjunto de las matrices cuyos elementos fuera de la diagonal son no negativos y la suma de las entradas en cada fila es igual a cero. En el caso de observaciones a tiempos equidistantes, es decir, $\Delta_i = \Delta$ para alguna $\Delta > 0$, la cadena de Markov Y es homogénea con respecto al tiempo con matriz de transición $P^\Delta(\Lambda)$, así que la función de verosimilitud se simplifica a

$$L_n(\Lambda) = \prod_{i=1}^m \prod_{j=1}^m P^\Delta(\Lambda)_{ij}^{K_{ij}(n)}, \quad \Lambda \in \ell$$

donde $K_{ij}(n)$ es el número de transiciones del estado i al j en la cadena de Markov en tiempo discreto $\{X(t_1), \dots, X(t_n)\}$. Se considerará el caso de observaciones a tiempos equidistantes.

Para la clase completa de cadenas de Markov homogéneas con respecto al tiempo con espacio de estados $\{1, \dots, m\}$, la función de verosimilitud basada en observaciones del estado de la cadena en los primeros n puntos de tiempo es

$$L_n(P) = \prod_{i=1}^m \prod_{j=1}^m P_{ij}^{K_{ij}(n)}, \quad P \in \mathcal{P}$$

donde $K_{ij}(n)$ es de nuevo el número de transiciones de i a j hasta el tiempo n , y \mathcal{P} denota el conjunto de las matrices de transición de $m \times m$ (matrices estocásticas), es decir, matrices de $m \times m$ con entradas no negativas y además la suma de cada una de las filas es igual a 1. Esta función de verosimilitud es idéntica a la de m distribuciones multinomiales independientes, así que el estimador de máxima verosimilitud del parámetro P es

$$\hat{P}_{ij} = K_{ij}(n) / K_i(n), \quad \text{donde} \quad K_i(n) = \sum_{j=1}^m K_{ij}(n).$$

Se define $\mathcal{P}_0 = \{P^\Delta(\Lambda) \mid \Lambda \in \ell\}$ el conjunto de las matrices de transición que corresponde a la observación discreta de un proceso de Markov en tiempo continuo. Ahora supongamos que se calcula \hat{P} como se muestra anteriormente basado en las observaciones discretas de un proceso de Markov en tiempo continuo. Si $\hat{P} \in \mathcal{P}_0$, ahí existe un $\Lambda \in \ell$ tal que $P^\Delta(\hat{\Lambda}) = \hat{P}$, y la función de

verosimilitud $L_n(\Lambda)$ alcanza su valor máximo en $\hat{\Lambda}$, el cual es por lo tanto el estimador de máxima verosimilitud. Sin embargo, existen dos problemas aquí. Uno es que el conjunto P_0 es muy complicado (excepto cuando $m = 2$); el otro es que la función exponencial de matrices no es inyectiva en todas las partes de su dominio, así que $\hat{\Lambda}$ no es necesariamente única. Cuando $\hat{P} \notin P_0$ la situación no es clara debido a la complicada estructura de P_0 , pero parece ser que no es inusual que el estimador de máxima de verosimilitud no exista, en particular cuando el tiempo entre observaciones Δ es grande. Resultados generales acerca de la existencia y unicidad del estimador de máxima verosimilitud se resumen en un teorema que se enuncia más adelante. En particular la probabilidad de que $\hat{P} \in P_0$ tiende a uno cuando n tiende a infinito. Se dará un acercamiento al caso cuando $m = 2$, donde el estimador de máxima verosimilitud no existe cuando $\hat{P} \notin P_0$.

El problema de identificar el conjunto P_0 tiene una larga historia y fue primeramente estudiado por Elfving (1937). Regularmente se refiere a él como el *problema de inclusión* para cadenas de Markov espacio de estados finito. Kingman (1962) mostró que $P_0 = P_+$ cuando $m = 2$, donde

$$P_+ = \{P \in P \mid \det(P) > 0\},$$

y deriva los siguientes resultados acerca de P_0 . Para $m \geq 3$, P_0 es relativamente un conjunto cerrado de P_+ con una compleja forma geométrica. En particular, no es convexo. Su interior relativo como un subconjunto de P es no vacío, así que su dimensión es $m(m-1)$. ∂P_0 denota la frontera de P_0 relativa a P_+ . Entonces

$$\partial P_0 = \left(\bigcup_{i \neq j} E_{ij} \right) \cup \mathcal{E},$$

donde E_{ij} es un subconjunto no vacío del conjunto de exponenciales de matrices con $\lambda_{ij} = 0$, y ε es un subconjunto no vacío de las matrices de $m \times m$ con menos de m eigen-valores distintos.

El segundo problema es si existen dos o más matrices de intensidades, Λ , para las cuales la correspondiente matriz de transición, $\exp(\Delta\Lambda)$ es la misma, es decir, existen dos o más Procesos de Markov en tiempo continuo para los cuales la muestra de estados $\{X(\Delta), \dots, X(n\Delta)\}$ tiene la misma distribución. En términos estadísticos esta es la cuestión si la parametrización de la distribución de los datos $X(\Delta), \dots, X(n\Delta)$ por Λ es identificable o no. Sea P_{00} , el subconjunto de P_0 , de las matrices $P \in P_0$ para la cuales Λ está determinada únicamente por $P = \exp(\Lambda)$. Para $m = 2$, $P_{00} = P_0 = P_+$. La caracterización del conjunto P_{00} es el clásico problema acerca de cuándo el logaritmo real de una matriz es único, el cual fue resuelto para matrices generales por Culver (1966). Su resultado general es que P_{00} consiste en las matrices de transición $P \in P_0$ para las cuales todos los eigen-valores son positivos y no divisores elementales (bloque de Jordan) que pertenece a cualquier eigen-valor de P más de una vez. Entonces, una vez que P ha sido calculado por máxima verosimilitud, en principio es fácil comprobar si determina un estimador único de la matriz de intensidades (proporcionado por $\hat{P} \in P_0$). Si $P \notin P_{00}$, hay una infinidad de soluciones X para la ecuación $P = \exp(X)$, de las cuales no todas pertenecen a P_0 . El conjunto de soluciones es contable si todos los eigen-valores reales de P son positivos con sus bloques de Jordan apareciendo una sola vez y cualquier eigen-valor complejo pertenece a sólo un bloque de Jordan. De otra manera, hay una cantidad incontable de soluciones. Cuthbert (1973) mostró que en el caso contable sólo un subconjunto finito de las soluciones está en P_0 .

Condiciones necesarias simples para que una matriz de transición P pertenezca a P_{00} las dio Cuthbert. Una condición simple, pero cruda, para que $P \in P_0$ pertenezca a P_{00} es que

$$\inf_i P_{ii} \geq \frac{1}{2}.$$

Un criterio menos crudo consiste en lo siguiente:

$$\left(\inf_i P_{ii}\right) \cdot \det(P) \geq e^{-\pi} \prod_i P_{ii}.$$

Los resultados de la existencia y unicidad para el estimador de máxima verosimilitud se resumen en el siguiente teorema.

4.1 TEOREMA

Si \hat{P} estimado por máxima verosimilitud pertenece a P_0 , entonces el estimador de máxima verosimilitud de la matriz de intensidad $\hat{\Lambda}$ existe y es la solución para $\hat{P} = \exp(\Delta \hat{\Lambda})$. Si $\hat{P} \notin P_0$, entonces el estimador de máxima verosimilitud $\hat{\Lambda}$ existe y satisface que $\exp(\Delta \hat{\Lambda}) \in \partial P_0$, como se definió anteriormente, o la función de verosimilitud no tiene máximo en ℓ . Si la verdadera matriz de intensidades Λ_0 satisface que $\exp(\Delta \Lambda_0) \in \text{int } P_0$, y si el proceso de Markov es ergódico, entonces la probabilidad de que el estimador de máxima verosimilitud exista tiende a uno cuando $n \rightarrow \infty$, y $\exp(\Delta \hat{\Lambda}) \rightarrow \exp(\Delta \Lambda_0)$ casi seguramente. Aún más, si Λ_0 satisface que $\exp(\Delta \Lambda_0) \in \text{int } P_{00}$, entonces la probabilidad de que el estimador de máxima verosimilitud es único tiende a uno y $\hat{\Lambda} \rightarrow \Lambda_0$ casi seguramente cuando $n \rightarrow \infty$. La condición $\exp(\Delta \Lambda_0) \in \text{int } P_{00}$ se satisface cuando Δ es suficientemente pequeño.

Prueba. La situación donde $\hat{P} \in P_0$ es trivial y fue discutida anteriormente. Ahora se asume que $\hat{P} \notin P_0$ y se define el conjunto

$$P_c = \{P \in P \mid \log L(P) \geq -c\},$$

donde $L(P)$ es la función de verosimilitud para toda la clase de cadenas de Markov dadas por $L_n(P) = \prod_{i=1}^m \prod_{j=1}^m P_{ij}^{K_{ij}(n)}$ y $c > 0$. Considerando el conjunto compacto $P_c \cap \bar{P}_0$ para alguna $c > 0$ suficientemente grande que $P_c \cap \bar{P}_0$ no sea vacío. Aquí \bar{P}_0 el conjunto $\bar{P}_0 = P_0 \cup \{P \in P \mid \det(P) = 0\}$. La función continua $L(P)$ tiene un máximo \tilde{P} en $P_c \cap \bar{P}_0$, y ya que $L(P)$ se incrementa siempre que P se mueve en la dirección de \hat{P} , \tilde{P} está en la frontera de $P_c \cap \bar{P}_0$. Entonces, $\tilde{P} \in \partial P_0$, en cuyo caso ahí existe una Λ tal que $\exp(\Delta \hat{\Lambda}) = \tilde{P}$ (P_0 es cerrado con relación a P_+), o $\det(\tilde{P}) = 0$, en cuyo caso la función de verosimilitud no tiene máximo en ℓ .

Ahora se asume que $\exp(\Delta \Lambda_0) \in \text{int } P_0$. Para procesos de Markov, se sabe que $\hat{P} \rightarrow \exp(\Delta \Lambda_0) \in \text{int } P_0$ casi seguramente cuando $n \rightarrow \infty$. Luego entonces la probabilidad de que $\hat{P} \in \text{int } P_0$ tiende a uno cuando $n \rightarrow \infty$. La unicidad y consistencia del estimador de máxima verosimilitud se muestra de la misma manera. El hecho de que $\exp(\Delta \hat{\Lambda}) \in \text{int } P_0$ cuando Δ es suficientemente pequeño se sigue de que $\inf_i P_{ii} \geq \frac{1}{2}$.

†

La situación de que el $\det(\tilde{P}) = 0$, donde el estimador de máxima verosimilitud no existe, es más probable cuando el determinante de $\exp(\Delta \Lambda_0)$ está cerca de cero. Cuando el proceso de Markov es ergódico, $\exp(\Delta \Lambda_0)$ converge cuando $\Delta \rightarrow \infty$ a la

matriz singular, donde todas las filas son iguales al vector π dado por $\pi\Lambda_0 = 0$ (la distribución estacionaria). Entonces la posibilidad de que el estimador de máxima verosimilitud no exista crece con Δ (por lo menos cuando Δ es suficientemente grande).

Para una muestra finita solamente se puede decir de cierto acerca de la unicidad que el estimador de máxima verosimilitud es único cuando $\hat{P} \in P_{00}$ y que éste no es único cuando $\hat{P} \in P_0 - P_{00}$. Si $\hat{P} \notin P_0$, no se puede estar seguro de que el estimador de máxima verosimilitud sea único, ni siquiera cuando $\hat{P} \in P_0$, por la complicada estructura geométrica del conjunto P_0 .

Consideremos más a detalle el caso de un proceso de Markov con dos estados. Este caso es más simple que cuando $m > 2$ porque $P_0 = P_+$, pero los problemas estadísticos ocurren en la frontera cuando $\det(P) = 0$, así que este problema es instructivo.

Para una matriz de intensidades

$$\begin{pmatrix} -\alpha & \alpha \\ \beta & -\beta \end{pmatrix}$$

donde $\alpha, \beta \geq 0$, los eigen-valores son 0 y $(\alpha + \beta)$. La matriz de transición correspondiente es

$$P^\Delta(\Lambda) = \frac{1}{\alpha + \beta} \begin{pmatrix} \beta + \alpha e^{-\Delta(\alpha + \beta)} & \alpha(1 - e^{-\Delta(\alpha + \beta)}) \\ \beta(1 - e^{-\Delta(\alpha + \beta)}) & \alpha + \beta e^{-\Delta(\alpha + \beta)} \end{pmatrix}$$

que tiene eigen-valores 1 y $\rho = \exp(-\Delta(\alpha + \beta))$. Es conveniente reparametrizar el modelo:

$$\pi_{11} = 1 - \frac{(1-\rho)\alpha}{\alpha + \beta} \text{ y } \pi_{21} = \frac{(1-\rho)\beta}{\alpha + \beta}.$$

Ignoramos el caso trivial donde $\alpha = \beta = 0$. El conjunto de valores de los parámetros es

$$\Pi_0 = \{(\pi_{11}, \pi_{21}) \mid 0 \leq \pi_{11} < \pi_{21} \leq 1\}.$$

Nótese que Π_0 es una parametrización de P_0 , mientras $P = [0,1]^2$. El determinante de $P^\Delta(\Lambda)$ es igual a $\pi_{11} - \pi_{21}$ corresponde al problema de la frontera de P_0 , donde $\det(P) = 0$. La función de verosimilitud es

$$L(\pi_{11}, \pi_{21}) = \pi_{11}^{K_{11}(n)} (1 - \pi_{11})^{K_{12}(n)} \pi_{21}^{K_{21}(n)} (1 - \pi_{21})^{K_{22}(n)},$$

así el estimador de máxima de verosimilitud de π_{11} es $\hat{\pi}_{11} = \frac{K_{11}(n)}{K_{1.}(n)}$.

$\frac{K_{21}(n)}{K_{2.}(n)} < \frac{K_{11}(n)}{K_{1.}(n)}$, es decir si $\hat{P} \in P_0$, entonces $\hat{\pi}_{21} = \frac{K_{21}(n)}{K_{2.}(n)}$. De otro modo, la

verosimilitud parcial $\tilde{L}(\pi_{21}) = L(\hat{\pi}_{11}, \pi_{21})$, donde $0 \leq \pi_{21} < \hat{\pi}_{11}$, sigue creciendo mientras π_{21} se acerca al punto frontera $\hat{\pi}_{11}$. Entonces en este caso la función de

verosimilitud no alcanza su máximo en Π_0 , y el estimador de máxima verosimilitud no existe. La situación es más posible cuando los valores reales de π_{11} y π_{21} están

cerca, lo cual pasa cuando $\Delta(\alpha + \beta)$ es grande porque entonces ambas probabilidades son cercanas a la probabilidad del estado 1 en la distribución

estacionaria, $\frac{\beta}{\alpha + \beta}$.

Debido a que $\alpha + \beta = -\log(\pi_{11} - \pi_{21}) / \Delta$, la función de verosimilitud crece (ligeramente) cuando $\alpha + \beta \rightarrow \infty$. Si por alguna razón se piensa que $\alpha + \beta$ no es grande, se puede acercarse al problema penalizando la verosimilitud con una distribución *a priori*, por ejemplo:

$$\phi(\alpha, \beta) \propto \alpha^a e^{-b\alpha} \beta^c e^{-d\beta}$$

la cual es la distribución *a priori* conjugada para el modelo continuo con función de verosimilitud $\prod_{i=1}^m \prod_{j \neq i} e^{-\lambda_{ij} R_i^T(\tau)} \lambda_{ij}^{N_{ij}^T(\tau)}$. Las exponenciales aseguran que la distribución *a posteriori* tienda a cero en la frontera cuando $\pi_{11} = \pi_{21}$ de tal modo que un estimador que maximice la distribución *a posteriori* exista también cuando $\frac{K_{21}(n)}{K_2(n)} \geq \frac{K_{11}(n)}{K_1(n)}$, es decir, cuando $\hat{P} \notin P_0$. Este estimador no es explícito, pero debe ser encontrado numéricamente.

Los valores propios de $\exp(\Delta\Lambda)$ son $e^{\Delta\gamma_i}$, $i=1, \dots, m$, donde $\{\gamma_i\}$ son los eigenvalores de Λ . Entonces, cuando $\exp(\Delta\Lambda)$ se acerca a la frontera crítica, donde $\det(\exp(\Delta\Lambda)) \rightarrow 0$, uno o más de los eigenvalores de Λ debe tender a menos infinito (Δ es fijo). Entonces la idea presentada en el ejemplo anterior de penalizar la función de verosimilitud discreta, que está acotada, por la distribución *a priori* conjugada por la función de verosimilitud continua, en general asegurará que no hay problemas con la existencia de un estimador que maximice la distribución *a posteriori*.

4.2 EJEMPLO 1

Para hacer lo anterior más claro tomemos un caso específico. Consideremos la cadena de Markov $Y_n = (-1)^n$, $n=1, 2, \dots, N$. Después de la estimación por el

método MCMC aplicada a esta cadena discreta se espera un estimador dado aproximadamente por

$$\hat{\Lambda} = \begin{pmatrix} -\hat{\alpha} & \hat{\alpha} \\ \hat{\alpha} & -\hat{\alpha} \end{pmatrix}.$$

Ambos estados de la cadena tienen el mismo número de observaciones en la serie de datos discretos y están separados a la misma distancia por lo tanto no hay ninguna distinción entre ellos. Esta es la razón intuitiva para esperar este tipo de resultado. Ahora tomaremos distintas distribuciones *a priori* para comprobar si esta cadena es incluíble o no. Pues si distintas distribuciones *a priori* llevan aproximadamente al mismo resultado entonces podríamos determinar que esta cadena es incluíble. Se tomaron como *a priori* matrices de la siguiente forma:

$$M_0 = \begin{pmatrix} -\alpha_i & \alpha_i \\ \beta_i & -\beta_i \end{pmatrix}.$$

Se usaron tres distribuciones *a priori*, en donde α_i y β_i se simularon de manera aleatoria con una distribución exponencial ($\alpha_i, \beta_i \sim \exp(\theta_i)$) con tres diferentes valores para los parámetros, 0.1, 1 y 10. Además se tomó una cadena con 1000 elementos, es decir, la cadena de Markov $Y_n = (-1)^n$, $n = 1, 2, \dots, 1000$. Entonces se obtuvieron las siguientes matrices *a priori* para cada uno de los casos:

$$\theta_1 = 10, \quad M_0^{\theta_1} = \begin{pmatrix} -0.0784 & 0.0784 \\ 0.0272 & -0.0272 \end{pmatrix}$$

$$\theta_2 = 1, \quad M_0^{\theta_2} = \begin{pmatrix} -1.5963 & 1.5963 \\ 0.6440 & -0.6440 \end{pmatrix}$$

$$\theta_3 = 0.1, \quad M_0^{\theta_3} = \begin{pmatrix} -4.9949 & 4.9949 \\ 14.6470 & -14.6470 \end{pmatrix}.$$

Se aplicó el método MCMC con cada una de estas matrices *a priori* con sus respectivos parámetros iniciales. Se hicieron 2000 iteraciones para cada caso y las primeras 500 iteraciones fueron el “calentamiento”. Se obtuvieron las estimaciones tomando el promedio de las últimas 1500 iteraciones para cada uno de los elementos de la matriz. Estos fueron los resultados para cada uno de los casos:

$$\hat{\Lambda}_1 = \begin{pmatrix} -2.3589 & 2.3589 \\ 2.3486 & -2.3486 \end{pmatrix}$$

$$\hat{\Lambda}_2 = \begin{pmatrix} -4.2982 & 4.2982 \\ 4.2919 & -4.2919 \end{pmatrix}$$

$$\hat{\Lambda}_3 = \begin{pmatrix} -7.3486 & 7.3486 \\ 7.3218 & -7.3218 \end{pmatrix}.$$

Efectivamente en cada caso la estimación es de una matriz donde el valor absoluto de sus elementos son prácticamente idénticos, en términos de estimación. El problema resultante de esto es que ninguna de las tres estimaciones es parecida a la otra. Esto contradice la teoría de la metodología MCMC, es decir, el hecho de que la estimación no depende de la distribución *a priori* que se escoja, ya que después del “calentamiento” se “olvida” de la *a priori* y alcanza la distribución estacionaria de la cadena del Markov del método. Una matriz inicial de elementos pequeños da como resultado una matriz con intensidades pequeñas y a medida que los elementos de la matriz *a priori* crezcan la estimación de las intensidades resultante también crecerá. Si tomamos en cuenta la matriz y su exponencial

$$\Lambda = \begin{pmatrix} -\alpha & \alpha \\ \alpha & -\alpha \end{pmatrix},$$

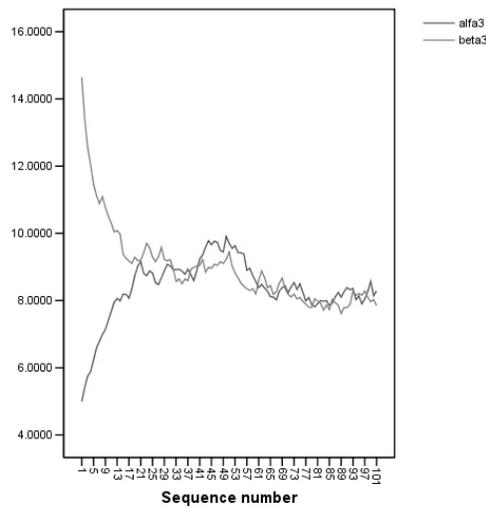
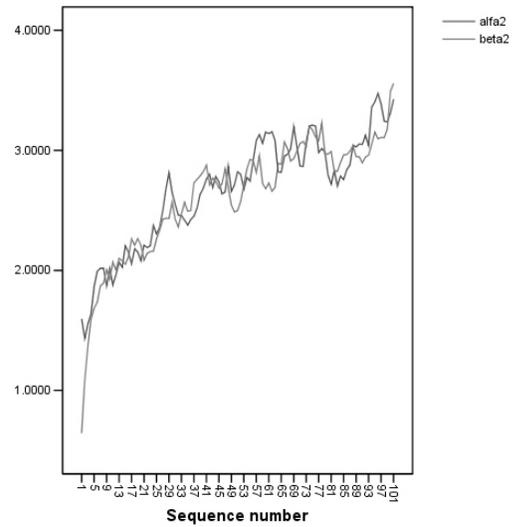
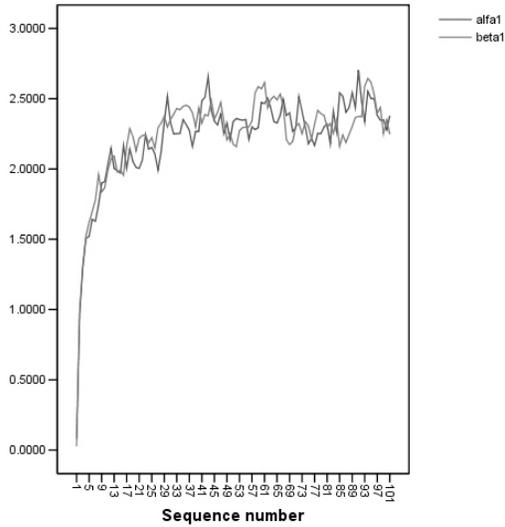
$$\exp(\Delta\Lambda) = \frac{1}{2\alpha} \begin{pmatrix} \alpha(1 + e^{-2\alpha\Delta}) & -\alpha(e^{-2\alpha\Delta} - 1) \\ -\alpha(e^{-2\alpha\Delta} - 1) & \alpha(1 + e^{-2\alpha\Delta}) \end{pmatrix}.$$

Entonces al hacer tender Δ a infinito, se obtiene la distribución estacionaria de este proceso:

$$\lim_{\Delta \rightarrow \infty} [\exp(\Delta\Lambda)] = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}.$$

Entonces se tiene que la distribución estacionaria es $\pi = (1/2, 1/2)$. Esto ejemplifica lo que se mostró anteriormente, cuando se da el caso que $\pi_{11} = \pi_{21}$. Se penaliza la función de verosimilitud discreta como se indica, que es precisamente la distribución utilizada por el método de estimación usado. Efectivamente se encuentra una solución numérica para el problema, pero no es una solución consistente. Cualquier matriz de intensidades con esta forma tiene la misma distribución estacionaria, lo cual es la principal razón para que el estimador no sea único, y para el caso de este método de estimación depende totalmente de la matriz de intensidades *a priori* que se tome.

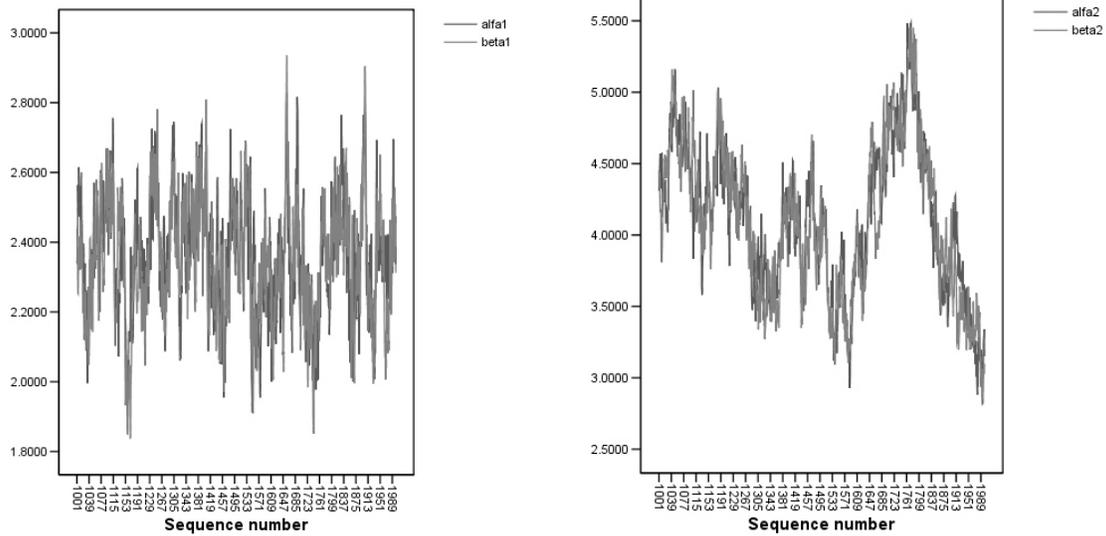
Se tiene interés en estudiar el “calentamiento” para saber que tan rápido se aproximan los estimadores para cada una de las matrices. Aquí se presentan las gráficas de las primeras 100 iteraciones:

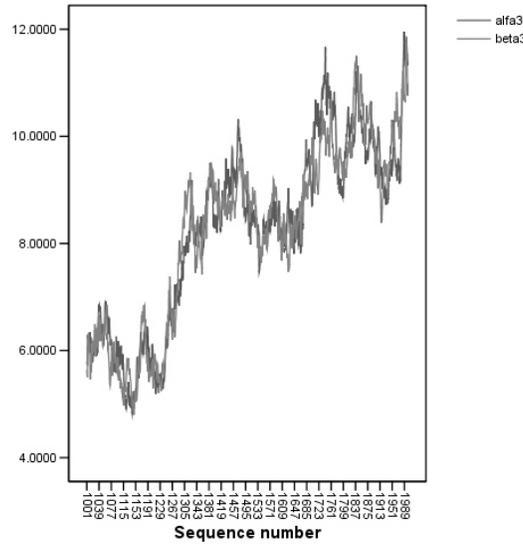


La tendencia en los tres casos es que los estimadores se aproximen el uno al otro después de un número pequeño de iteraciones, lo cual nos indica que las observaciones discretas de la cadena marcan el equilibrio entre los dos estados, pues los resultados no hacen distinciones entre uno y otro. Los estimadores tienen una fuerte correlación pues cuando uno crece el otro también y viceversa. Aparentemente de las gráficas se podría creer que los estimadores están

convergiendo y tienden a estacionarse, pero 100 observaciones de 2000 no son suficientes para determinar que el método converja a una distribución estacionaria que es lo que se busca. En el caso de que se convergiera a esta distribución el promedio ergódico de las últimas 1500 iteraciones proporcionaría un buen estimador en cada caso aislado, pero es obvio que aunque el estimador en cada caso fuese confiable, ninguno de las tres estimaciones se aproximan entre sí y por lo tanto no se llegaría a un resultado enteramente confiable.

Para comprobar lo anterior se obtuvieron las gráficas de las últimas 1000 iteraciones del método, y de este modo determinar si el promedio ergódico daría una buena estimación en cada caso:





Para el caso $\theta_1 = 10$ las iteraciones tienen una cierta estabilidad entre 2 y 2.8, se podría creer que efectivamente se vuelve estacionaria pero el rango en el que se mueven los datos es relativamente grande y no se puede saber con certeza la exactitud con la que se estimaría si se tomara el promedio ergódico. Cuando $\theta_2 = 1$, el rango va desde 3 hasta 5.5, la fluctuación es más diversa y no sigue una tendencia específica, es decir no es creciente ni decreciente; los datos varían de una manera irregular, por lo tanto el promedio ergódico no sería un buen estimador para el parámetro en el caso de que la cadena en estudio fuera incluíble. Ahora cuando se toma $\theta_3 = 0.1$ las iteraciones siguen una tendencia creciente y se mueven en un intervalo entre 5 y 12, lo cual no es bueno tampoco para la estimación, pues en este caso tampoco sería factible usar el promedio como un estimador confiable.

Finalmente podemos concluir que la cadena de Markov $Y_n = (-1)^n$ no es incluíble, pues la información proporcionada no hace ninguna distinción entre los dos valores que puede tomar. Con probabilidad 1 se sale de un estado para entrar al otro, y ya que la discretización es prácticamente la misma para ambos casos, las intensidades para los tiempos de estancia que pudiese estimar el método MCMC serán iguales, si no iguales por lo menos muy cercanas.

4.3 EJEMPLO 2

Retomando el ejemplo del capítulo anterior (Estimación), un proceso de Markov con la matriz de intensidades:

$$\Lambda = \begin{bmatrix} -1 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & -1 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & -1 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & -1 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 & -1 \end{bmatrix},$$

con probabilidades de transición,

i a j	1	2	3	4	5
1	0.00	0.25	0.25	0.25	0.25
2	0.25	0.00	0.25	0.25	0.25
3	0.25	0.25	0.00	0.25	0.25
4	0.25	0.25	0.25	0.00	0.25
5	0.25	0.25	0.25	0.25	0.00

En los ejemplos anteriores se discretizó el proceso desde el tiempo 0 hasta el 1000, a distancias 0.5 y 1. En este se hizo a 2 unidades de tiempo, es decir 500 parejas de información tiempo-estado. Por lo tanto se tiene una discretización con menos información; adelante se muestra sólo una pequeña parte de la discretización:

Tiempo	Estado								
0	5	10	1	20	1	30	2	40	5
2	4	12	1	22	5	32	3	42	3
4	3	14	5	24	3	34	5	44	5
6	5	16	3	26	2	36	5	46	2
8	1	18	5	28	3	38	5	48	4

Con esta base de información se utilizó el método MCMC para estimar la matriz de intensidades como en los casos anteriores. Se utilizó la misma matriz de intensidades *a priori* para la simulación:

$$M_0 = \begin{bmatrix} -1.7936 & 0.3597 & 0.7001 & 0.4152 & 0.3187 \\ 0.3824 & -2.7345 & 0.1056 & 1.0730 & 1.1735 \\ 1.1948 & 0.1508 & -2.7606 & 1.2388 & 0.1761 \\ 0.6131 & 0.1582 & 0.4386 & -1.7755 & 0.5655 \\ 1.8913 & 0.5216 & 0.2009 & 0.6272 & -3.2411 \end{bmatrix}.$$

Se realizaron 6000 iteraciones utilizando el método como se explicó en el capítulo anterior. El “calentamiento” fue de 500 iteraciones. Es el mismo procedimiento utilizado en los ejemplos anteriores, y se obtuvo la estimación para la matriz de intensidades:

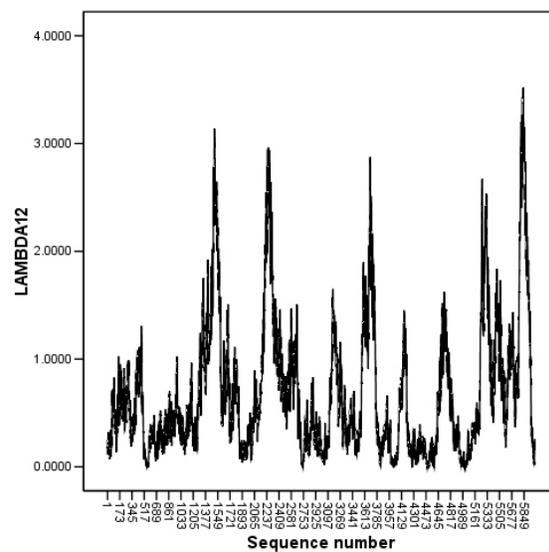
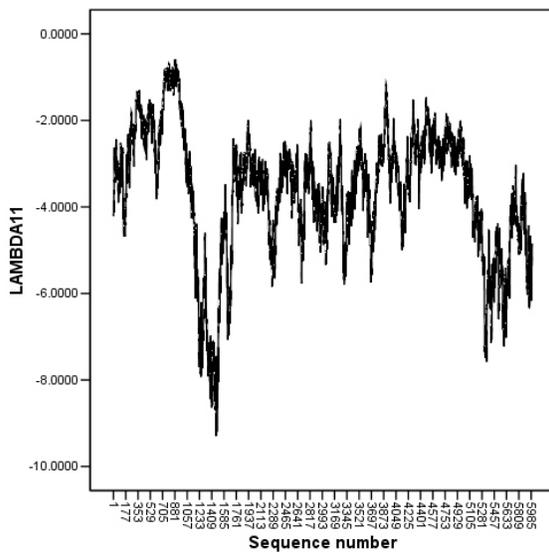
$$\hat{\Lambda} = \begin{bmatrix} -3.7288 & 0.6978 & 0.9590 & 1.1789 & 0.8931 \\ 0.9834 & -3.7697 & 1.1733 & 0.8653 & 0.7476 \\ 0.6885 & 1.3200 & -4.0409 & 0.8572 & 1.1752 \\ 0.8869 & 0.9224 & 1.0960 & -4.0940 & 1.1887 \\ 1.2018 & 1.1180 & 1.5032 & 0.9807 & -4.8037 \end{bmatrix}.$$

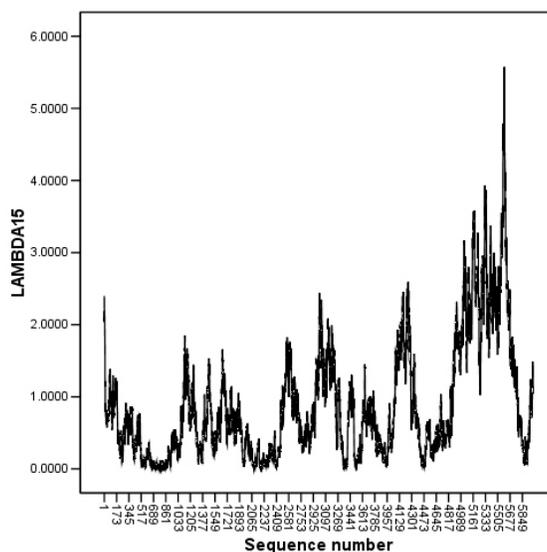
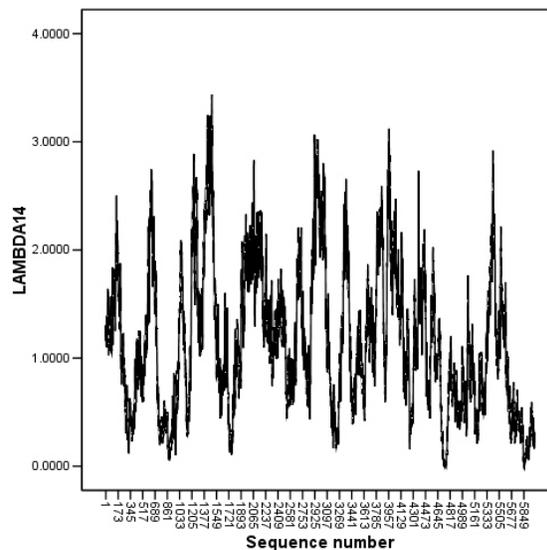
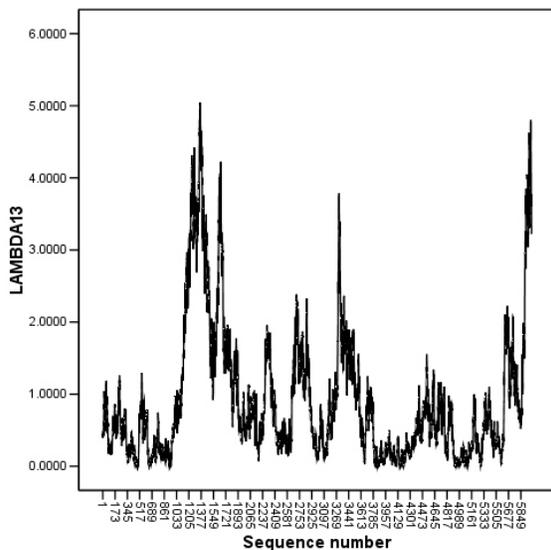
De esta estimación se obtienen las probabilidades de transición estimadas:

i a j	1	2	3	4	5
1	0.0000	0.1871	0.2572	0.3162	0.2395
2	0.2609	0.0000	0.3113	0.2295	0.1983
3	0.1704	0.3267	0.0000	0.2121	0.2908
4	0.2166	0.2253	0.2677	0.0000	0.2903
5	0.2502	0.2327	0.3129	0.2042	0.0000

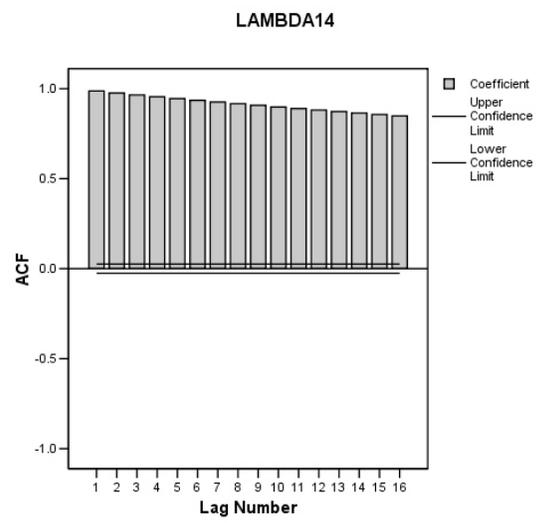
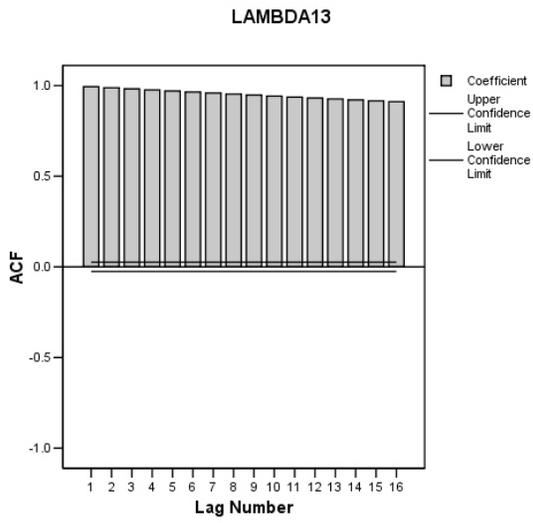
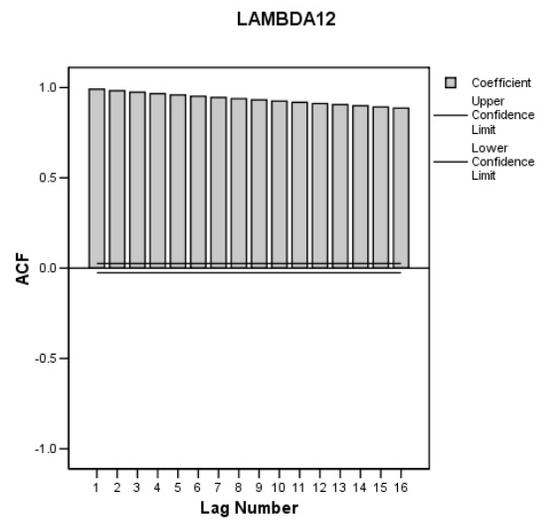
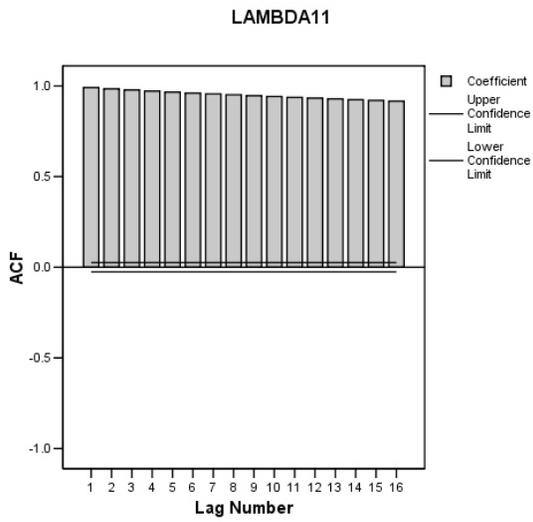
La estimación es mala pues no se acerca a la matriz de intensidades, ni siquiera es parecida, pero las probabilidades de transición son parecidas. Esto hace creer que las observaciones discretas del proceso no determinan una solución única. La distancia entre los datos no aporta suficiente información y por esto la solución no es única. Es decir, puede existir más de una matriz de intensidades que tenga la misma distribución estacionaria.

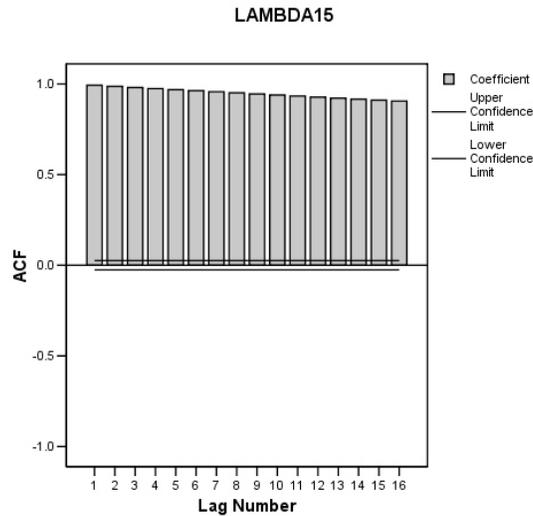
Como en los casos anteriores se estudiarán los resultados con respecto a la primera fila de la matriz de intensidades estimada. En el “calentamiento” no se observa que la cadena tenga tendencia a estacionarse, entonces no habría estimación o no tendría caso.





Se observa en las gráficas de la secuencia que siguieron los elementos de la matriz de intensidades que en el “calentamiento” no se converge, es más, en toda la cadena la secuencia no sigue una tendencia y en ningún momento se llega a la distribución estacionaria que es lo que busca el método MCMC.





También la autocorrelación de los datos es extremadamente alta; esto quiere decir que las iteraciones dependen completamente de las anteriores y por lo tanto la estimación no es correcta. Por lo tanto no tendría caso obtener intervalos de credibilidad para la matriz de intensidades.

La distribución estacionaria del proceso de Markov en estudio es

$\pi = (0.2, 0.2, 0.2, 0.2, 0.2)$. Los resultados anteriores vienen de la proximidad a

la distribución estacionaria:

$$\exp(2\Lambda) = \begin{bmatrix} 0.2657 & 0.1836 & 0.1836 & 0.1836 & 0.1836 \\ 0.1836 & 0.2657 & 0.1836 & 0.1836 & 0.1836 \\ 0.1836 & 0.1836 & 0.2657 & 0.1836 & 0.1836 \\ 0.1836 & 0.1836 & 0.1836 & 0.2657 & 0.1836 \\ 0.1836 & 0.1836 & 0.1836 & 0.1836 & 0.2657 \end{bmatrix}.$$

Esta es la matriz estocástica del proceso al tiempo 2. No se alcanza todavía la distribución estacionaria pero se aproxima mucho a ella. Una discretización a esta distancia es representativa de la distribución mencionada, entonces no es la representación de un proceso único sino de un número de procesos de Markov

cuya matriz de intensidades tiene esta distribución. Esto demuestra que no cualquier discretización del proceso es buena para la estimación, pues si la distancia no es suficientemente pequeña se pierde información necesaria. En este ejemplo se entiende cómo los datos discretos a una distancia pequeña puede lograr una muy buena estimación de la matriz de intensidades para el proceso de Markov. A una distancia de $\Delta = 0.5$, se obtiene una excelente estimación mientras que con $\Delta = 2$ la estimación ni siquiera existe, pues no cuenta con los requerimientos necesarios para tomarla como tal, no es estacionaria y es completamente dependiente de las iteraciones anteriores.

CAPÍTULO 5

RIESGO DE CRÉDITO

5.1 Calificaciones de Moody's

Una calificación es una evaluación de la capacidad y compromiso jurídico de un emisor para efectuar los pagos de intereses y amortización del principal en el plazo previsto en las condiciones de emisión de un título de renta fijo específico.

La calificación mide la probabilidad de que un emisor incurra en incumplimiento (“default”) con respecto al valor o título durante la duración del mismo, la cual, dependiendo del instrumento, podrá ser de unos cuantos días o hasta 30 años o más. Asimismo, las calificaciones de largo plazo incorporan una evaluación de cuál sería la pérdida monetaria esperada en caso de producirse un incumplimiento.

5.2 CALIFICACIONES DE LARGO PLAZO

Los símbolos de calificación Aaa a C de Moody's se utilizan para calificar la solvencia crediticia de títulos con plazo de un año o más, es decir, de “largo plazo”. En la parte más alta de la escala, la “triple A” (Aaa) de Moody's representa una calidad crediticia de “primera categoría”, lo que significa que el valor es muy superior a las demás categorías en cuanto a los márgenes de protección contra el incumplimiento bajo las condiciones económicas más severas. La calificación más baja, C, indica el más alto nivel de riesgo de crédito, y significa que un título tiene muy pocas posibilidades de representar cualquier valor de inversión real. Ubicados en el sector medio de la escala, los títulos con calificación de Baa y superior generalmente se consideran como aptos para la inversión; los de calificación Ba e inferior se consideran especulativos. Los modificadores numéricos (1, 2 y 3) que

se agregan a las categorías de calificación de Aa a Caa indican graduaciones más precisas de riesgo, como se muestra más adelante.

5.3 CALIFICACIONES DE DEUDA A LARGO PLAZO

Aaa. Los títulos calificados *Aaa* se consideran como los de mejor calidad. Conllevan el menor grado de riesgo de inversión y por lo general se conocen como de “primera categoría”. Los pagos de intereses están protegidos por un margen muy grande o excepcionalmente estable y el cobro del principal es seguro. Aunque es probable que cambien los diversos elementos de protección, se considera que los cambios que pueden preverse tienen poca probabilidad de afectar la solidez fundamental de estas emisiones.

Aa. Los títulos calificados *Aa* se consideran de alta calidad según todas las normas. Junto con el grupo de bonos calificados *Aaa*, comprenden lo que se conoce como títulos de alta calidad. Tienen menor calificación que los títulos *Aaa* por ser sus márgenes de protección menos amplios, por tener mayor fluctuación de los elementos de protección, o porque existen otros elementos que pueden hacer que el riesgo a largo plazo parezca algo mayor que el de los títulos calificados *Aaa*.

A. Los títulos calificados *A* poseen muchos atributos de inversión favorables y deben ser considerados como obligaciones de grado medio superior. Los factores que dan protección al principal y los intereses se consideran adecuados, pero pueden existir elementos que sugieren una susceptibilidad al deterioro en algún momento futuro.

Baa. Los títulos calificados *Baa* se consideran obligaciones de grado medio (es decir, no tienen ni mucha ni poca protección). Los pagos de intereses y la protección del principal parecen ser apropiados por el momento, pero ciertos elementos protectores pueden faltar o ser menos fiables a largo plazo. Estos

bonos no tienen características de inversión sobresalientes y, de hecho, también tienen características especulativas.

Ba. Los títulos calificados Ba se consideran como títulos con elementos especulativos; su futuro no es muy seguro. Muchas veces la protección de los pagos de intereses y principal puede ser muy moderada y por lo tanto pueden no estar bien protegidos. Su situación de incertidumbre es una característica de los bonos de esta clase.

B. Los títulos calificados B por lo general no poseen cualidades deseables como instrumentos de inversión. La seguridad en el cumplimiento de los pagos de intereses o principal, u otros compromisos contractuales, puede ser limitada a largo plazo.

Caa. Los títulos calificados Caa son de baja calidad. Estas emisiones pueden haber incumplido los pagos o pueden contener elementos de riesgo con respecto al cobro del principal e intereses.

Ca. Los títulos calificados Ca representan obligaciones con alto grado de especulación. Estas emisiones a menudo han incumplido pagos o tienen marcadas deficiencias.

C. Los títulos calificados C pertenecen a la categoría más baja y la posibilidad de que estas emisiones alcancen alguna vez valor de inversión es remota. Las calificaciones de títulos de Moody's, cuando se especifica, se aplican a obligaciones bancarias con derecho de prioridad (u obligaciones bancarias "senior") y obligaciones hacia tenedores de pólizas con derecho de prioridad y reclamaciones de compañías de seguros con un vencimiento original de más de un año. Las obligaciones que dependen de mecanismos de apoyo, tales como cartas de crédito y bonos o fianzas de indemnización, se excluyen a menos que se califiquen explícitamente. Las obligaciones de una sucursal de un banco se

consideran como domiciliadas en el país en donde se encuentra ubicada la sucursal. Salvo excepción en contrario, la calificación de Moody's sobre la capacidad de un banco de reintegrar las obligaciones con derecho de prioridad se extiende sólo a sucursales ubicadas en países que tienen una Calificación Soberana para Depósitos Bancarios de Moody's. A las obligaciones de sucursales se les da la calificación que sea más baja entre la calificación del banco, o la Calificación Soberana para Depósitos Bancarios de Moody's para el país en donde se encuentra ubicada la sucursal. Cuando la obligación está denominada en una moneda distinta a la moneda del país en donde se encuentra domiciliada la obligación, las calificaciones de Moody's no incorporan una opinión en cuanto a si el pago de la obligación se verá afectado por las acciones del gobierno que controla la moneda de denominación. Asimismo, las calificaciones de Moody's tampoco incorporan el riesgo relacionado con conflictos bilaterales entre el país de domicilio de un inversor por una parte y el país de domicilio del emisor o el país en donde se encuentra ubicada una sucursal del emisor por otra parte. Moody's no realiza ninguna afirmación acerca del hecho de que las obligaciones bancarias u obligaciones de compañías de seguros calificadas estén exentas de la inscripción correspondiente conforme a la Ley de Títulos de los Estados Unidos (U.S. Securities Act) de 1933 o que han sido emitidas de conformidad con cualquier otra ley o reglamento aplicable. Las calificaciones de Moody's tampoco implican representaciones sobre la posibilidad de ejecutar legalmente la obligación específica de un banco o una compañía de seguros, o sobre el hecho de que la emisión constituye una obligación con derecho de prioridad válido del emisor calificado.

5.4 ESTIMACIÓN DE LAS INTENSIDADES PARA LAS CALIFICACIONES MOODY'S

Después de haber estudiado el método MCMC para estimar la matriz de intensidades para un proceso de Markov homogéneo, se obtiene un modelo para el riesgo de crédito de las empresas. Moodys maneja una escala de calificaciones

de riesgo de crédito sobre la capacidad que tienen las empresas para afrontar sus responsabilidades en el ámbito financiero. Se tomarán estas calificaciones como los estados del proceso y la unidad de tiempo será un año. Las calificaciones se simplificarán, es decir, sólo se tomarán en cuenta las calificaciones estrictas sin sus subcategorías, así mismo, las calificaciones Caa, Ca, C se agruparán en una sola (C). Entonces tenemos un proceso de Markov X en un espacio de calificaciones E , donde

$$E = \{Aaa, Aa, A, Baa, Ba, B, C, D\},$$

sin pérdida de generalidad se define el espacio de estados

$$E' = \{1, 2, 3, 4, 5, 6, 7, 8\},$$

donde existe la correspondencia única entre estados y calificaciones definida de la siguiente manera:

Calificación	Aaa	Aa	A	Baa	Ba	B	C	D
Estado	1	2	3	4	5	6	7	8

El proceso se supondrá homogéneo con respecto al tiempo, y el estado D (default) como un estado absorbente, después de saltar a este estado, con probabilidad 1 se mantendrá ahí el proceso por todo el tiempo restante. Se busca estimar la matriz de intensidades para el proceso X , de tal modo que se pueda obtener una probabilidad de *default* para las empresas.

Se utilizará una población de 1250 empresas. Por confidencialidad, todos los datos que se emplearán para este ejemplo serán simulados, aunque basados en datos reales. Lo primero que se obtendrá son los estados iniciales para las empresas y después de esto se simulará la historia de las calificaciones de las

empresas durante cinco años. Para lo anterior tomaremos una distribución de las calificaciones iniciales de la siguiente manera:

Calificación	Aaa	Aa	A	Baa	Ba	B	C
Proporción	0.0360	0.0999	0.3389	0.2566	0.0999	0.1535	0.0152

Para la historia de calificaciones de las empresas, se usará la siguiente matriz de intensidades obtenida de datos reales:

$$\Lambda = \begin{pmatrix} -0.0592 & 0.0592 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0036 & -0.0849 & 0.0812 & 0 & 0 & 0 & 0 & 0 \\ 0.0233 & -0.0788 & 0.0514 & 0.0041 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.0463 & -0.0928 & 0.0445 & 0.0010 & 0 & 0.0009 \\ 0 & 0 & 0.0018 & 0.1182 & -0.1809 & 0.0563 & 0 & 0.0046 \\ 0 & 0 & 0.0032 & 0.0053 & 0.0840 & -0.2193 & 0.0559 & 0.0710 \\ 0 & 0 & 0 & 0 & 0.0097 & 0.0756 & -0.4651 & 0.3798 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Se observa que en la última fila de la matriz, correspondiente a la calificación *default*, las entradas son cero; esto es debido a que dicho estado es absorbente, una vez que se entra en dicho estado no se puede salir de él. Para cada una de las empresas se tomó la calificación o el estado inicial al tiempo 0; partiendo de esta calificación se simuló un proceso de Markov con la matriz de intensidades Λ durante cinco años, es decir, hasta el tiempo 5. Se obtuvieron los siguientes estados iniciales para las 1250 empresas:

Calificación	Aaa	Aa	A	Baa	Ba	B	C
Cantidad	43	122	451	313	121	174	26

No se tienen calificaciones D, pues esto implicaría que una empresa en *default* se tomara en cuenta para el estudio, pero esto no tiene sentido pues ya no tiene participación en el medio.

Los tiempos de estancia en cada calificación se simularon con una distribución exponencial con parámetro $-\frac{1}{\lambda_{ij}}$, donde λ_{ij} es la entrada (i,j) de la matriz de intensidades Λ . Entonces se tienen los parámetros:

Estado	1	2	3	4	5	6	7
Calificación	Aaa	Aa	A	Baa	Ba	B	C
Parámetro	16.886	11.783	12.692	10.778	5.5282	4.5604	2.15

Así mismo se tienen las probabilidades de saltar del estado i al j, en el momento en que dicho salto exista:

Probabilidades de i a j	1	2	3	4	5	6	7	8
1	0	1.0000	0	0	0	0	0	0
2	0.0429	0	0.9571	0	0	0	0	0
3	0	0.2964	0	0.6520	0.0517	0	0	0
4	0	0	0.4993	0	0.4802	0.0111	0	0.0095
5	0	0	0.0099	0.6534	0	0.3112	0	0.0255
6	0	0	0.0144	0.0243	0.3829	0	0.2548	0.3237
7	0	0	0	0	0.0209	0.1625	0	0.8166
8	0	0	0	0	0	0	0	1.0000

Con lo anterior se tiene toda la información necesaria para simular cinco años de calificaciones para cada una de las 1250 empresas. En cada caso se simula el tiempo de estancia en el estado inicial con una distribución exponencial con el parámetro correspondiente. El siguiente estado es determinado por la distribución dada por la tabla de probabilidades que se muestra. Con respecto a dichas distribuciones, se observa que hay saltos que no sucederán en la simulación pues tienen probabilidad 0.

Después de haber simulado el proceso continuo para cada una de las empresas, se discretizó el proceso a tres distancias: 3 meses, 6 meses y 1 año. Para

ejemplificar lo anterior se muestran las discretizaciones de las primeras 20 empresas para las tres distancias:

3 MESES

Empresa	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Mes 0	3	6	5	3	4	6	5	7	4	3	3	7	6	3	3	3	3	3	3	4
Mes 3	3	6	5	3	4	6	5	7	4	3	3	7	6	3	3	3	3	3	3	4
Mes 6	3	6	4	4	4	6	5	7	4	3	3	7	3	3	3	3	3	3	3	4
Mes 9	3	6	4	4	4	6	5	7	4	3	3	7	3	3	3	3	3	3	3	4
Mes 12	3	6	4	4	4	6	5	7	4	3	3	7	3	3	3	3	3	3	3	4
Mes 15	3	6	4	4	4	6	5	7	4	3	3	8	3	3	3	3	3	3	3	3
Mes 18	3	6	4	4	4	6	5	8	4	3	3	8	3	3	3	3	3	3	3	3
Mes 21	3	6	4	4	4	6	5	8	4	3	3	8	3	3	3	3	3	3	3	3
Mes 24	3	6	4	4	4	6	5	8	4	3	3	8	3	3	3	3	3	3	3	3
Mes 27	3	6	4	4	4	6	5	8	4	3	3	8	3	3	3	3	3	3	3	3
Mes 30	3	6	4	4	4	6	5	8	4	3	3	8	3	3	3	3	3	3	3	3
Mes 33	3	6	4	4	4	6	5	8	4	3	3	8	3	3	3	3	3	3	3	3
Mes 36	3	6	4	4	4	6	5	8	4	3	3	8	3	3	3	3	3	4	3	4
Mes 39	3	6	4	4	4	6	5	8	4	3	3	8	3	3	4	3	3	4	3	4
Mes 42	3	6	4	4	4	6	5	8	5	3	3	8	3	3	4	3	3	4	3	4
Mes 45	3	6	4	4	4	6	4	8	5	3	3	8	3	3	4	3	3	4	3	4
Mes 48	3	6	4	4	4	6	4	8	5	3	3	8	3	3	4	3	3	4	3	4
Mes 51	3	6	4	4	4	6	4	8	5	3	3	8	3	3	4	3	3	4	3	4
Mes 54	3	6	4	4	4	6	4	8	5	3	3	8	3	3	4	3	3	4	3	4
Mes 57	3	6	4	4	4	6	4	8	5	3	3	8	3	3	4	3	3	4	3	4
Mes 60	5	6	4	4	4	6	4	8	5	3	3	8	3	3	4	3	3	4	3	4

6 MESES

Empresa	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Mes 0	3	6	5	3	4	6	5	7	4	3	3	7	6	3	3	3	3	3	3	4
Mes 6	3	6	4	4	4	6	5	7	4	3	3	7	3	3	3	3	3	3	3	4
Mes 12	3	6	4	4	4	6	5	7	4	3	3	7	3	3	3	3	3	3	3	4
Mes 18	3	6	4	4	4	6	5	8	4	3	3	8	3	3	3	3	3	3	3	3
Mes 24	3	6	4	4	4	6	5	8	4	3	3	8	3	3	3	3	3	3	3	3
Mes 30	3	6	4	4	4	6	5	8	4	3	3	8	3	3	3	3	3	3	3	3
Mes 36	3	6	4	4	4	6	5	8	4	3	3	8	3	3	3	3	3	4	3	4
Mes 42	3	6	4	4	4	6	5	8	5	3	3	8	3	3	4	3	3	4	3	4
Mes 48	3	6	4	4	4	6	4	8	5	3	3	8	3	3	4	3	3	4	3	4
Mes 54	3	6	4	4	4	6	4	8	5	3	3	8	3	3	4	3	3	4	3	4
Mes 60	5	6	4	4	4	6	4	8	5	3	3	8	3	3	4	3	3	4	3	4

1 AÑO

Empresa	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Año 0	3	6	5	3	4	6	5	7	4	3	3	7	6	3	3	3	3	3	3	4
Año 1	3	6	4	4	4	6	5	7	4	3	3	7	3	3	3	3	3	3	3	4
Año 2	3	6	4	4	4	6	5	8	4	3	3	8	3	3	3	3	3	3	3	3
Año 3	3	6	4	4	4	6	5	8	4	3	3	8	3	3	3	3	3	4	3	4
Año 4	3	6	4	4	4	6	4	8	5	3	3	8	3	3	4	3	3	4	3	4
Año 5	5	6	4	4	4	6	4	8	5	3	3	8	3	3	4	3	3	4	3	4

Una vez que se cuenta con la información discreta se aplicó el método MCMC para estimar la matriz de intensidades en cada una de las distintas distancias. Las empresas son independientes entre sí, entonces sin pérdida de generalidad se simulará por MCMC para cada una por separado en cada iteración. Al final de cada iteración se contarán todos los saltos del total de las empresas y de igual manera se hará con el tiempo total de estancia en cada estado (calificación).

Se toma una matriz de intensidades *a priori* de manera aleatoria donde las entradas tienen una distribución exponencial con parámetro 1. Todas las entradas se toman de este modo excepto la diagonal pues estos elementos están determinados por la suma de las otras entradas en la fila. A partir de esta matriz *a priori* se simula el proceso para cada una de las empresas y se obtienen los totales de saltos y tiempos de estancia en la primera iteración. Las entradas de la siguiente matriz de intensidades se simulan como una distribución gamma. Sea

$R_i^{(h)}$ $i \in E$, $h \in \{1,2,\dots,1250\}$, el tiempo de estancia en el estado i en el proceso de la empresa h y $N_{ij}^{(h)}$ $i, j \in E$, $i \neq j$, $h \in \{1,2,\dots,1250\}$, el número de saltos del estado i al estado j en la empresa h . El tiempo de estancia total y el número total de saltos está dado por

$$R_i = \sum_{h=1}^{1250} R_i^{(h)} \quad N_{ij} = \sum_{h=1}^{1250} N_{ij}^{(h)} .$$

Los parámetros de la distribución gamma para simular la siguiente matriz de intensidades están dados de la siguiente manera:

$$\lambda_{ij} \sim \Gamma\left(\frac{1}{R_i + 1}, N_{ij} + 1\right) \quad i, j \in E \quad i \neq j.$$

Una vez que se simulan las entradas de la nueva matriz de intensidades se vuelve a simular el proceso de saltos que pase específicamente por todos los puntos determinados por la información discreta. Nuevamente se obtienen el total de saltos y tiempos de estancia por empresa, se suman respectivamente y se vuelven a simular las entradas de una nueva matriz de intensidades con la que se repetirá el proceso en la siguiente iteración.

Para la estimación de la matriz de intensidades de las calificaciones de riesgo de crédito se realizaron 5000 iteraciones para cada una de las tres series de datos discretos. Se utilizó la misma distribución para los tres casos:

$$M_o = \begin{pmatrix} -7.4254 & 0.3514 & 3.2653 & 0.0188 & 1.2628 & 0.9600 & 1.0890 & 0.4781 \\ 0.6898 & -3.0320 & 0.0728 & 0.5392 & 0.5108 & 0.6947 & 0.2240 & 0.3007 \\ 0.7149 & 1.0140 & -8.9623 & 0.2799 & 2.9049 & 0.4325 & 3.5525 & 0.0636 \\ 0.5415 & 0.1384 & 1.7357 & -8.8401 & 0.7244 & 2.4234 & 2.4038 & 0.8728 \\ 2.2186 & 0.6919 & 0.4381 & 2.4511 & -7.0383 & 0.0635 & 0.4257 & 0.7493 \\ 1.8123 & 0.0405 & 0.7379 & 3.1404 & 0.8433 & -8.5867 & 0.8357 & 1.1765 \\ 0.7814 & 0.0414 & 0.6885 & 1.0704 & 1.4416 & 0.3603 & -4.4448 & 0.0613 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

De esta matriz de intensidades obtenemos las probabilidades de transición con el que se simularán los saltos de estados en el momento en que estos ocurran para la primera iteración. Se observa que el renglón correspondiente al estado 8, es decir, calificación *default* no se simula pues es un estado que, con anterioridad se sabe, es absorbente y por lo tanto no tiene caso hacer estimación sobre éste.

Probabilidades de i a j	1	2	3	4	5	6	7	8
1	0	0.0473	0.4397	0.0025	0.1701	0.1293	0.1467	0.0644
2	0.2275	0	0.0240	0.1779	0.1685	0.2291	0.0739	0.0992
3	0.0798	0.1131	0	0.0312	0.3241	0.0483	0.3964	0.0071
4	0.0613	0.0157	0.1964	0	0.0819	0.2741	0.2719	0.0987
5	0.3152	0.0983	0.0622	0.3483	0	0.0090	0.0605	0.1065
6	0.2111	0.0047	0.0859	0.3657	0.0982	0	0.0973	0.1370
7	0.1758	0.0093	0.1549	0.2408	0.3243	0.0811	0	0.0138
8	0	0	0	0	0	0	0	1.0000

Por el método MCMC se simularon 5000 iteraciones y se tomaron las primeras 500 como “calentamiento”. La estimación de la matriz de intensidades está dada por

$$\hat{\Lambda} = \frac{1}{4500} \sum_{k=501}^{5000} M_k ,$$

donde M_k es la matriz de intensidades obtenida en la k -ésima iteración del proceso. De este modo se obtienen las estimaciones para las matrices de intensidades de las tres series de datos discretos:

3 MESES	Aaa	Aa	A	Baa	Ba	B	C	D
Aaa	-0.0985	0.0610	0.0102	0.0055	0.0055	0.0054	0.0056	0.0054
Aa	0.0031	-0.0759	0.0637	0.0016	0.0030	0.0015	0.0015	0.0015
A	0.0005	0.0193	-0.0801	0.0535	0.0053	0.0005	0.0005	0.0005
Baa	0.0006	0.0011	0.0463	-0.0953	0.0443	0.0018	0.0006	0.0006
Ba	0.0015	0.0016	0.0038	0.1195	-0.2084	0.0714	0.0030	0.0076
B	0.0016	0.0016	0.0092	0.0047	0.0611	-0.2270	0.0709	0.0779
C	0.0099	0.0099	0.0112	0.0117	0.0209	0.1169	-0.6205	0.4401
D	0	0	0	0	0	0	0	0

6 MESES	Aaa	Aa	A	Baa	Ba	B	C	D
Aaa	-0.0974	0.0650	0.0054	0.0053	0.0054	0.0054	0.0055	0.0054
Aa	0.0031	-0.0797	0.0690	0.0015	0.0015	0.0015	0.0015	0.0015
A	0.0005	0.0205	-0.0815	0.0546	0.0045	0.0005	0.0005	0.0005
Baa	0.0006	0.0006	0.0464	-0.0970	0.0461	0.0021	0.0006	0.0006
Ba	0.0016	0.0016	0.0040	0.1206	-0.2075	0.0695	0.0017	0.0085
B	0.0016	0.0016	0.0093	0.0039	0.0609	-0.2251	0.0667	0.0812
C	0.0096	0.0096	0.0103	0.0098	0.0193	0.1049	-0.5707	0.4071
D	0	0	0	0	0	0	0	0

1 AÑO	Aaa	Aa	A	Baa	Ba	B	C	D
Aaa	-0.0986	0.0612	0.0095	0.0056	0.0056	0.0056	0.0055	0.0055
Aa	0.0032	-0.0782	0.0639	0.0034	0.0030	0.0016	0.0015	0.0015
A	0.0005	0.0199	-0.0799	0.0530	0.0049	0.0005	0.0005	0.0005
Baa	0.0006	0.0010	0.0458	-0.0954	0.0443	0.0024	0.0007	0.0006
Ba	0.0016	0.0017	0.0033	0.1212	-0.2089	0.0722	0.0029	0.0061
B	0.0016	0.0016	0.0091	0.0036	0.0628	-0.2276	0.0658	0.0831
C	0.0093	0.0095	0.0118	0.0109	0.0212	0.1011	-0.5590	0.3951
D	0	0	0	0	0	0	0	0

Se obtienen las probabilidades de transición al momento de existir un salto en el proceso:

3 MESES	Aaa	Aa	A	Baa	Ba	B	C	D
Aaa	0	0.6673	0.0556	0.0544	0.0556	0.0559	0.0561	0.0552
Aa	0.0395	0	0.8661	0.0188	0.0187	0.0192	0.0191	0.0186
A	0.0059	0.2518	0	0.6697	0.0552	0.0058	0.0058	0.0058
Baa	0.0063	0.0062	0.4783	0	0.4751	0.0213	0.0065	0.0063
Ba	0.0075	0.0076	0.0194	0.5811	0	0.3350	0.0084	0.0409
B	0.0070	0.0070	0.0415	0.0171	0.2705	0	0.2963	0.3606
C	0.0168	0.0169	0.0181	0.0171	0.0338	0.1839	0	0.7134
D	0	0	0	0	0	0	0	1

6 MESES	Aaa	Aa	A	Baa	Ba	B	C	D
Aaa	0	0.6187	0.1037	0.0556	0.0558	0.0551	0.0563	0.0548
Aa	0.0406	0	0.8392	0.0213	0.0393	0.0199	0.0200	0.0198
A	0.0061	0.2415	0	0.6682	0.0664	0.0061	0.0058	0.0059
Baa	0.0064	0.0113	0.4854	0	0.4650	0.0189	0.0065	0.0064
Ba	0.0074	0.0077	0.0181	0.5733	0	0.3427	0.0142	0.0366
B	0.0071	0.0071	0.0406	0.0205	0.2691	0	0.3124	0.3432
C	0.0159	0.0159	0.0180	0.0189	0.0336	0.1884	0	0.7093
D	0	0	0	0	0	0	0	1

1 AÑO	Aaa	Aa	A	Baa	Ba	B	C	D
Aaa	0	0.6211	0.0968	0.0568	0.0570	0.0572	0.0554	0.0557
Aa	0.0405	0	0.8175	0.0441	0.0390	0.0199	0.0197	0.0192
A	0.0063	0.2496	0	0.6639	0.0618	0.0062	0.0061	0.0061
Baa	0.0066	0.0107	0.4806	0	0.4639	0.0248	0.0068	0.0065
Ba	0.0076	0.0080	0.0157	0.5801	0	0.3454	0.0141	0.0291
B	0.0071	0.0071	0.0400	0.0156	0.2758	0	0.2893	0.3651
C	0.0167	0.0169	0.0211	0.0195	0.0380	0.1809	0	0.7069
D	0	0	0	0	0	0	0	1

El resultado es que las matrices obtenidas por el método MCMC para las tres distancias (3 meses, 6 meses y 1 año) son bastante similares entre sí. Se observa que los parámetros que en la matriz de intensidades original son cero, en el caso de la estimación no lo son, aunque aparentemente son suficientemente pequeños. Esto es un indicio de que dichos parámetros deberían de ser cero, aunque debido a la técnica de estimación en el proceso esto no es posible, ya que los datos están dados por una simulación de una variable aleatoria y esta acción ya lleva consigo una penalización, pues la simulación por construcción no arroja valores iguales a cero. Pese a que los parámetros dados para la simulación de las variables fueran muy cercanos a cero, la simulación no arrojaría dicho valor puesto que la simulación de una variable aleatoria (Gamma, en este caso) nunca es cero.

Los parámetros que originalmente son iguales a cero, influyen de manera directa en el caso de la estimación de las tasas totales (diagonal) además de afectar las probabilidades de transición. Por ejemplo, en el caso de la matriz de intensidades original implicaba las siguientes probabilidades de saltos:

Probabilidades de i a j	Aaa	Aa	A	Baa	Ba	B	C	D
Aaa	0	1.0000	0	0	0	0	0	0
Aa	0.0429	0	0.9571	0	0	0	0	0
A	0	0.2964	0	0.6520	0.0517	0	0	0
Baa	0	0	0.4993	0	0.4802	0.0111	0	0.0095
Ba	0	0	0.0099	0.6534	0	0.3112	0	0.0255
B	0	0	0.0144	0.0243	0.3829	0	0.2548	0.3237
C	0	0	0	0	0.0209	0.1625	0	0.8166
D	0	0	0	0	0	0	0	1.0000

Mientras que la estimación en el caso de la distancia a 3 meses, implica las probabilidades de la siguiente manera:

3 MESES	Aaa	Aa	A	Baa	Ba	B	C	D
Aaa	0	0.6673	0.0556	0.0544	0.0556	0.0559	0.0561	0.0552
Aa	0.0395	0	0.8661	0.0188	0.0187	0.0192	0.0191	0.0186
A	0.0059	0.2518	0	0.6697	0.0552	0.0058	0.0058	0.0058
Baa	0.0063	0.0062	0.4783	0	0.4751	0.0213	0.0065	0.0063
Ba	0.0075	0.0076	0.0194	0.5811	0	0.3350	0.0084	0.0409
B	0.0070	0.0070	0.0415	0.0171	0.2705	0	0.2963	0.3606
C	0.0168	0.0169	0.0181	0.0171	0.0338	0.1839	0	0.7134
D	0	0	0	0	0	0	0	1

El caso más relevante es la calificación Aaa, pues originalmente sólo saltaba a la calificación Aa con probabilidad 1, y en el caso de la estimación el salto tiene probabilidad 0.67 y también los saltos a las demás calificaciones tienen una probabilidad positiva, lo cual no debería suceder pues estos saltos no pasan en los datos originales. Por lo tanto se tiene que determinar qué estimaciones de las intensidades son significativas y cuáles no. En este caso no se podrán evaluar por medio de pruebas de hipótesis puesto que no se cuenta con la distribución explícita de los parámetros.

Se empleará un método el cual consiste en evaluar una distribución al hacer cortes de la misma a diferentes niveles, y de este modo determinar los puntos donde se corta la gráfica de la distribución, determinando de este modo las abscisas, y posteriormente las ordenadas. Conociendo estos últimos valores se

calcula la masa contenida entre estos puntos para obtener un intervalo de credibilidad. Para ejemplificar esto se tomará una distribución normal estándar, la cual tiene la forma $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, después se obtendrá el valor de x_1 dado el valor $f(x_1)$, y se determinará la probabilidad acumulada entre $-x_1$ y x_1 , así hasta obtener la masa $1 - \alpha$ para obtener un intervalo de “credibilidad” del $(1 - \alpha)\%$. Al evaluar la distribución normal se obtiene lo siguiente:

f(x)	x	F(x)	Credibilidad	Intervalo
0.3500	0.5115	0.3045	39%	[0.5115,-0.5115]
0.2500	0.9670	0.1668	67%	[0.967,-0.967]
0.1500	1.3985	0.0810	84%	[1.3985,-1.3985]
0.1031	1.6450	0.0500	90%	[1.645,-1.645]
0.0584	1.9600	0.0250	95%	[1.96,-1.96]

dando como resultado los intervalos correspondientes. Esto ejemplifica el método y da como resultado los intervalos de confianza para una distribución normal estándar. Ahora este mismo método se aplicará a las estimaciones de las entradas de la matriz de intensidades.

Sabemos que el estimador de las intensidades es el promedio de un cierto número de variables aleatorias distribuidas gamma, así mismo, empíricamente se tomará el promedio de las distribuciones de cada una de las entradas como la distribución del parámetro a estimar. Se evaluará de la misma manera que en el ejemplo anterior y se obtendrán para los casos de interés, es decir, los parámetros pequeños, su intervalo de “credibilidad” y en los casos para los cuales el cero quede contenido en dicho intervalo se concluirán como no significativos.

Tomemos en cuenta que la distribución *a posteriori* de cada una de las entradas de la matriz de intensidades se distribuye gamma para cada una de las

iteraciones, es decir, $\lambda_{ij}^{(k)} \sim \Gamma\left(\frac{1}{R_{ij}^{(k)} + 1}, N_{ij}^{(k)} + 1\right)$ $i, j \in E$ $i \neq j$ en la k-ésima

iteración. Entonces se estimará el promedio de las distribuciones *a posteriori* como

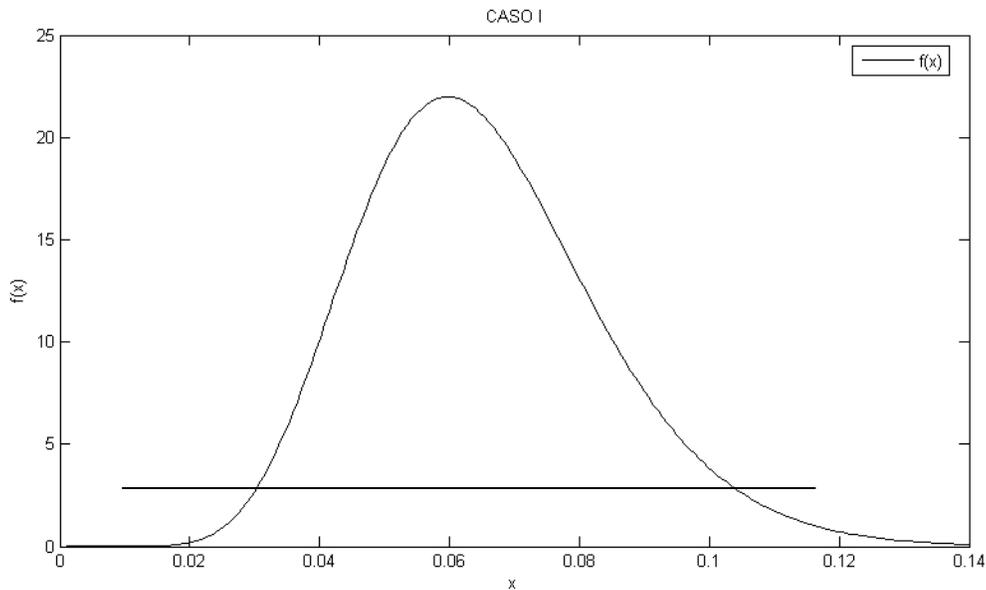
la estimación de la distribución de cada una de las entradas de la matriz de

intensidades: $\hat{f}_{ij}(x) = \frac{1}{N} \sum_{k=1}^N f_{ij}^{(k)}(x)$.

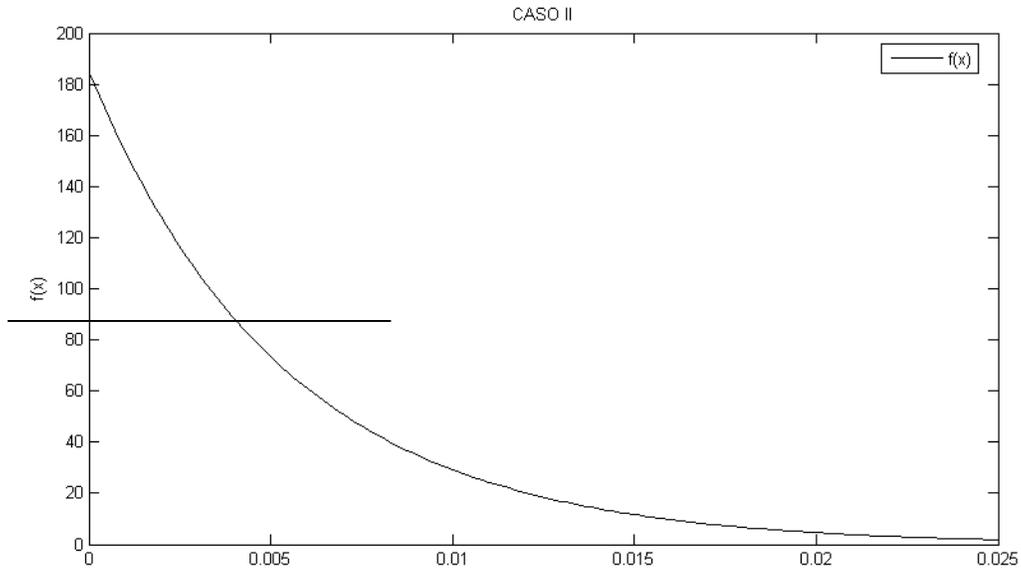
Por lo tanto, la distribución tiene la siguiente forma:

$$\hat{f}_{ij}(x) = \frac{1}{N} \sum_{k=1}^N \frac{(R_i^{(k)} + 1)^{N_{ij}^{(k)} + 1}}{\Gamma(N_{ij}^{(k)} + 1)} x^{N_{ij}^{(k)}} e^{-(R_i^{(k)} + 1)x}.$$

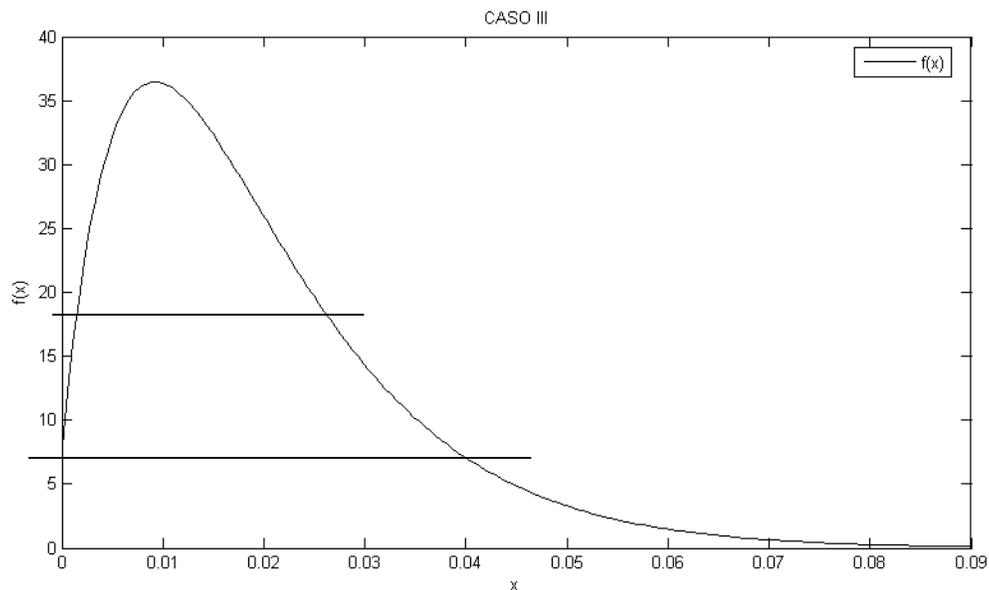
Esta función es una densidad y sobre ésta se determinará la significancia de las estimaciones de la matriz de intensidades, si el valor 0 queda contenido en el intervalo obtenido por el método, se dirá que la estimación no es significativa. Para hacer esta evaluación existen tres casos, los cuales se explican con su gráfica respectivamente:



- I. El caso en que la distribución tiene una gráfica de esta forma y $\hat{f}_{ij}(0) = 0$, por lo tanto el intervalo del 95% no contendrá al cero evidentemente. Esta es la razón por la que se considera significativa.



II. En este caso $\hat{f}_{ij}(0) > 0$, además $\hat{f}_{ij}(x)$ es estrictamente decreciente, el método que se aplica hace un corte a una cierta altura, pero corta en un solo punto de la gráfica y por esto el cero queda contenido en el intervalo. Estos casos se consideran no significativos.



III. Este caso es el que requiere más análisis pues $\hat{f}_{ij}(0) > 0$, pero el corte determina dos ordenadas. Para determinar la significancia en estos casos se

encontró x tal que $\int_0^x \hat{f}_{ij}(y)dy \approx .95$. Si $\hat{f}_{ij}(0) \geq \hat{f}_{ij}(x)$ se determina que no es significativa pues el intervalo es de la forma $[0,x]$, pero en el caso en que $\hat{f}_{ij}(0) < \hat{f}_{ij}(x)$ la estimación resulta ser significativa pues el intervalo es de la forma $[x_1,x]$ donde $x_1 \neq 0$.

En la siguiente tabla se observan las intensidades que son significativas (i), a las que se les aplicó el método y resultaron no significativas (ii); las que resultaron significativas por el método (iii) :

(i,j)	1	2	3	4	5	6	7	8
1	-	i	ii	ii	ii	ii	ii	ii
2	iii	-	i	ii	ii	ii	ii	ii
3	ii	i	-	i	iii	ii	ii	ii
4	ii	ii	i	-	i	iii	ii	ii
5	ii	ii	ii	i	-	i	ii	i
6	ii	ii	i	ii	i	-	ii	ii
7	ii	ii	ii	ii	ii	i	-	i
8	-	-	-	-	-	-	-	-

Finalmente se volvió a usar el método MCMC para estimar la matriz de intensidades del Proceso de Markov determinado por las calificaciones de riesgo de crédito que se tenían de manera discreta cada 3 meses. En este caso se consideraron las entradas no significativas como intensidades iguales a cero, es decir, para estos casos no hubo estimación ni simulación.

Se usó la siguiente matriz de intensidades inicial:

$$M_o = \begin{pmatrix} -0.8559 & 0.8559 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0449 & -0.3070 & 0.2621 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.7948 & -4.0908 & 0.8934 & 1.4026 & 0 & 0 & 0 \\ 0 & 0 & 0.4770 & -1.8969 & 0.5321 & 0.8878 & 0 & 0 \\ 0 & 0 & 0 & 0.4928 & -2.4927 & 1.2520 & 0 & 0.7479 \\ 0 & 0 & 0.4463 & 0 & 0.7662 & -1.7241 & 0.4094 & 0.1022 \\ 0 & 0 & 0 & 0 & 0 & 0.6871 & -1.4833 & 0.7962 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

De esta matriz se obtienen las probabilidades de salto del estado i a j :

Probabilidades de i a j	1	2	3	4	5	6	7	8
1	0	1.0000	0	0	0	0	0	0
2	0.1462	0	0.8538	0	0	0	0	0
3	0	0.4388	0	0.2184	0.3429	0	0	0
4	0	0	0.2515	0	0.2805	0.4680	0	0
5	0	0	0	0.1977	0	0.5023	0	0.3001
6	0	0	0.2588	0	0.4444	0	0.2375	0.0593
7	0	0	0	0	0	0.4633	0	0.5368
8	0	0	0	0	0	0	0	1.0000

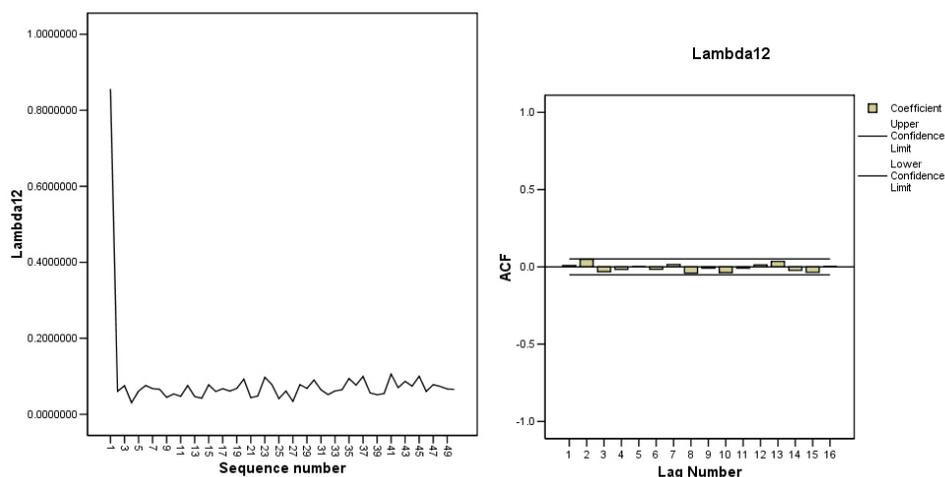
De esta matriz inicial se realizaron 5000 iteraciones del MCMC con los mismos datos discretos a 3 meses, tomando un “calentamiento” de 500 iteraciones, obteniendo como resultado la siguiente matriz de intensidades:

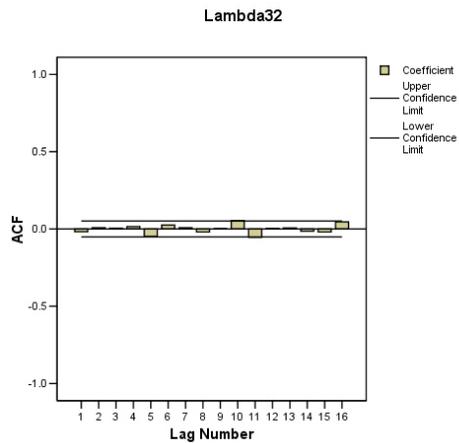
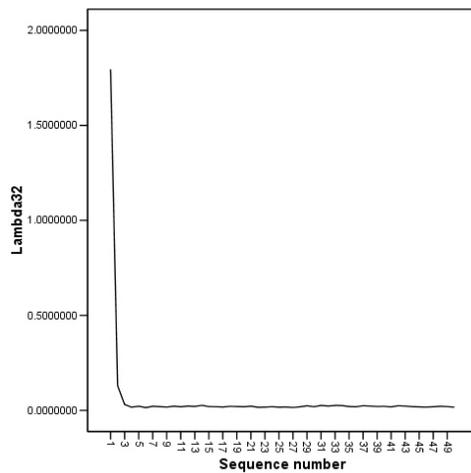
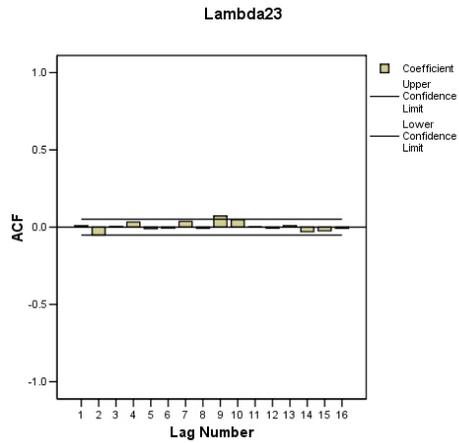
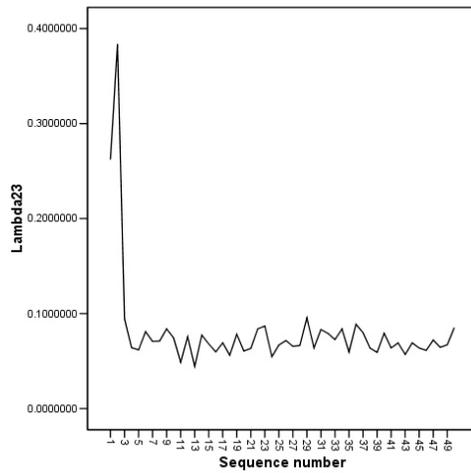
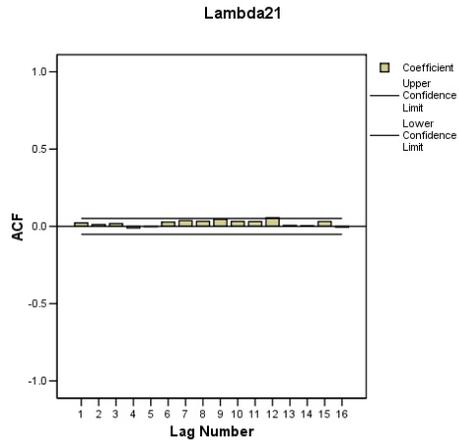
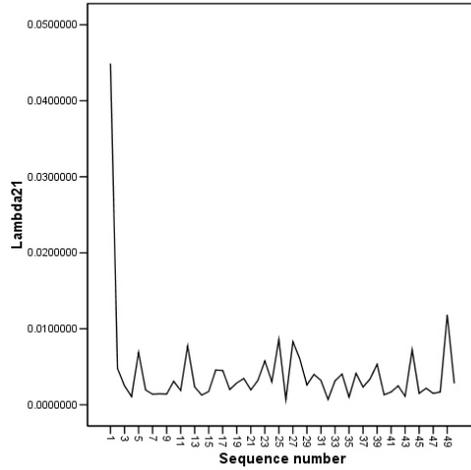
Intensidades	Aaa	Aa	A	Baa	Ba	B	C	D
Aaa	-0.0658	0.0658	0	0	0	0	0	0
Aa	0.0030	-0.0721	0.0691	0	0	0	0	0
A	0	0.0206	-0.0797	0.0546	0.0045	0	0	0
Baa	0	0	0.0474	-0.0957	0.0462	0.0021	0	0
Ba	0	0	0	0.1244	-0.2030	0.0699	0	0.0086
B	0	0	0.0099	0	0.0646	-0.2223	0.0665	0.0814
C	0	0	0	0	0	0.1136	-0.5181	0.4045
D	0	0	0	0	0	0	0	0

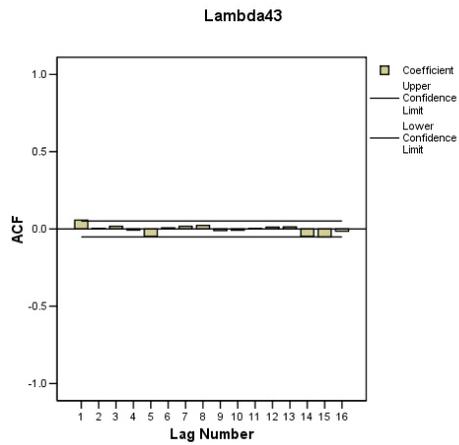
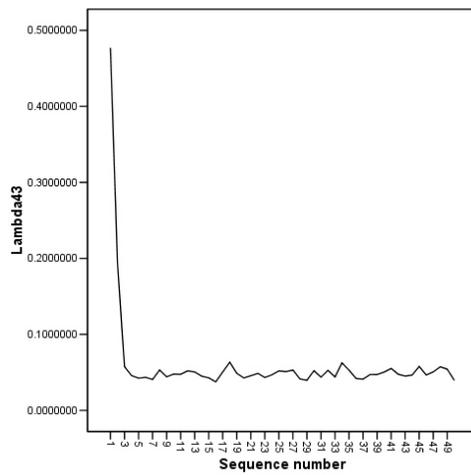
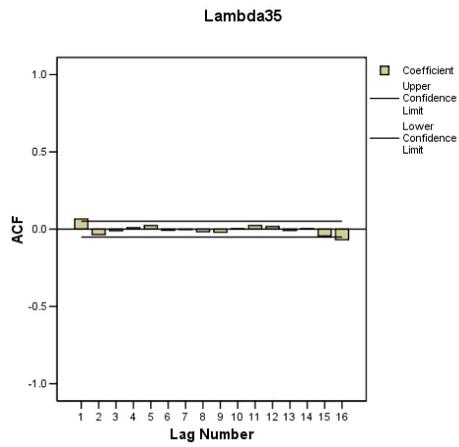
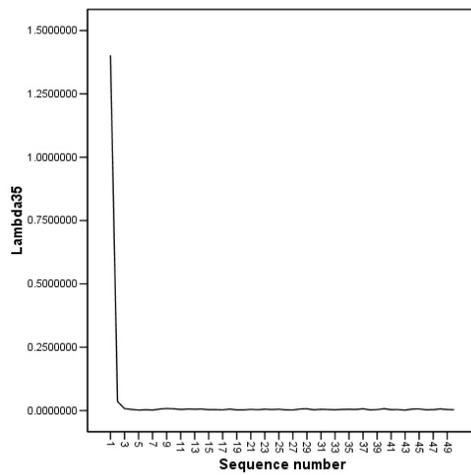
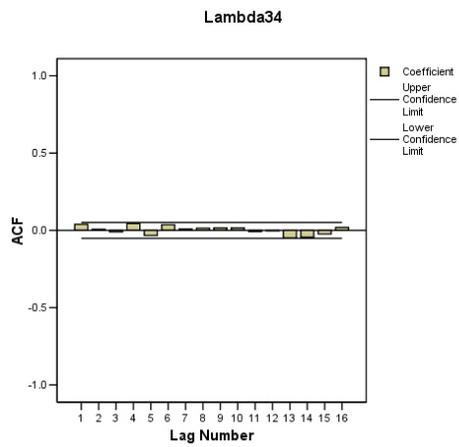
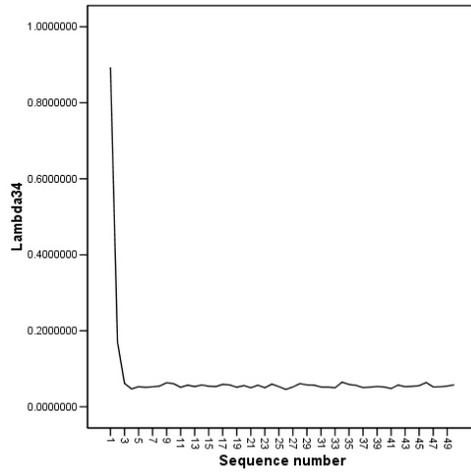
Esta matriz presenta probabilidades de transición muy similares a la matriz de intensidades original, con la que se simuló todo el proceso a estimar:

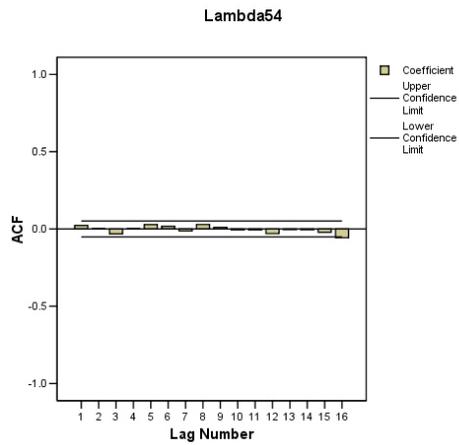
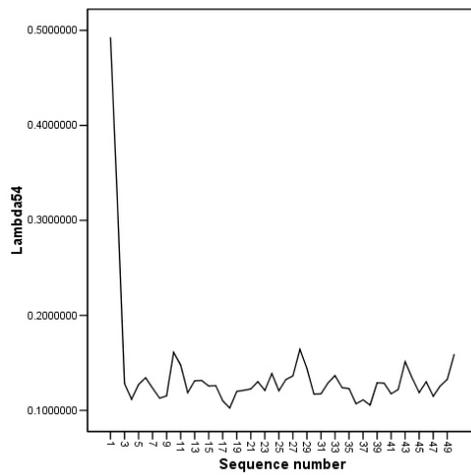
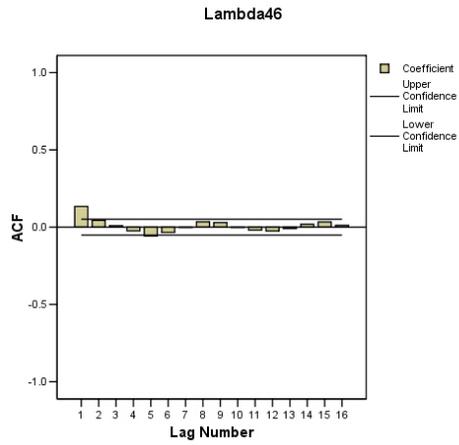
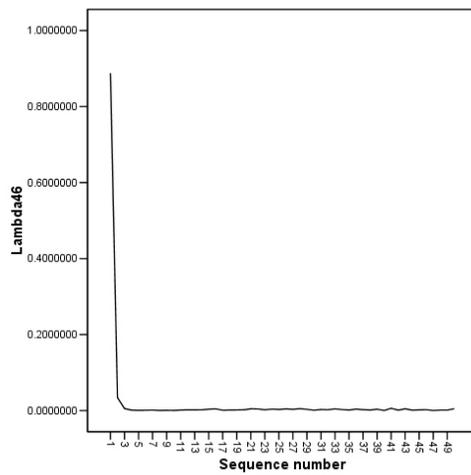
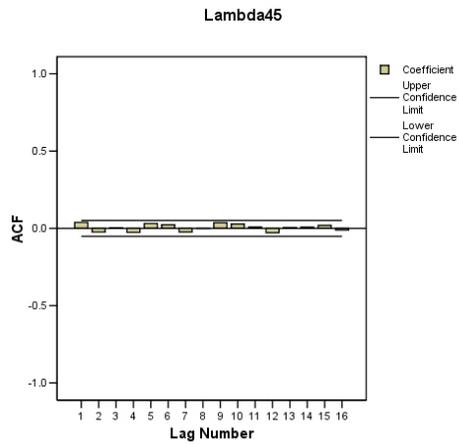
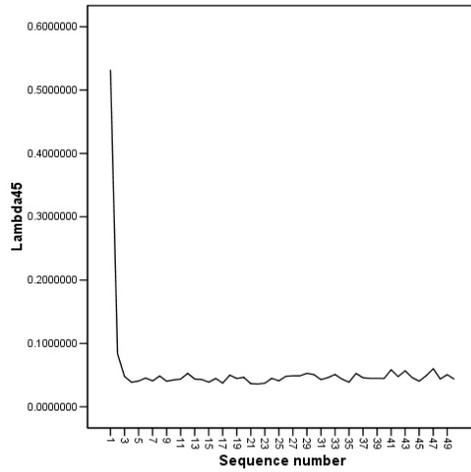
Probabilidades de i a j	1	2	3	4	5	6	7	8
1	0	1.0000	0	0	0	0	0	0
2	0.0417	0	0.9583	0	0	0	0	0
3	0	0.2581	0	0.6849	0.0570	0	0	0
4	0	0	0.4957	0	0.4828	0.0215	0	0
5	0	0	0	0.6130	0	0.3444	0	0.0426
6	0	0	0.0444	0	0.2904	0	0.2992	0.3659
7	0	0	0	0	0	0.2193	0	0.7807
8	0	0	0	0	0	0	0	1.0000

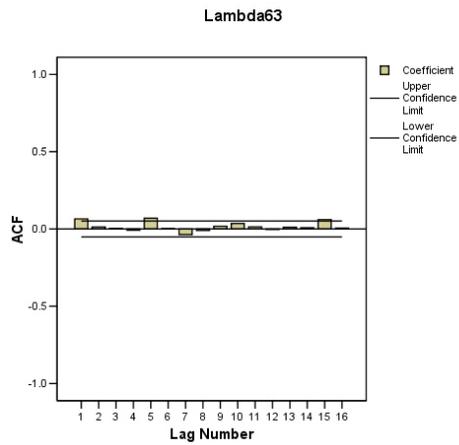
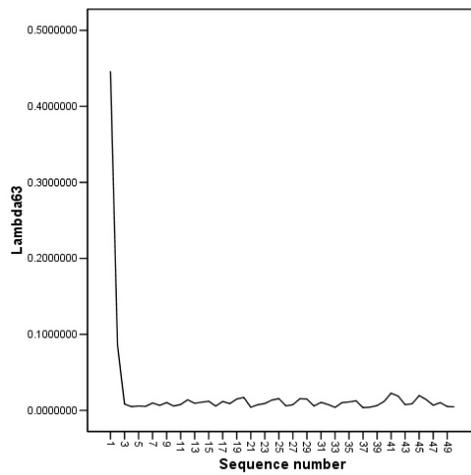
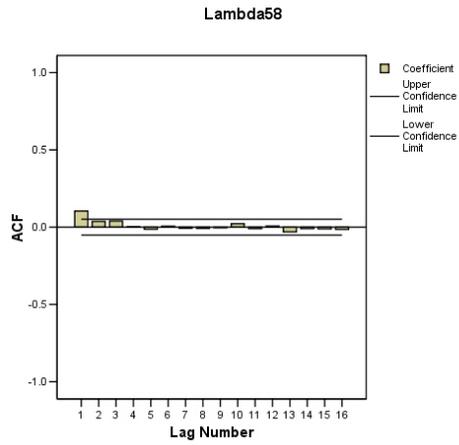
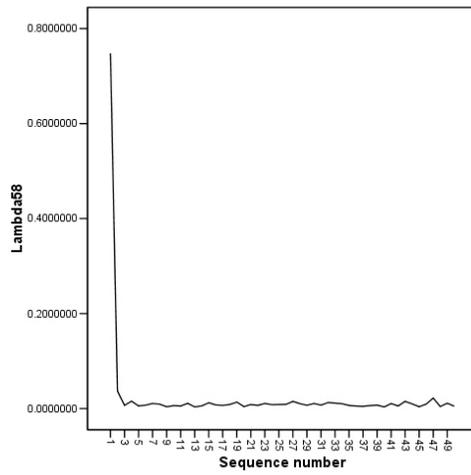
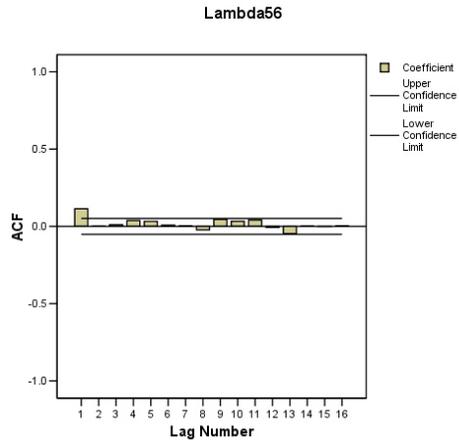
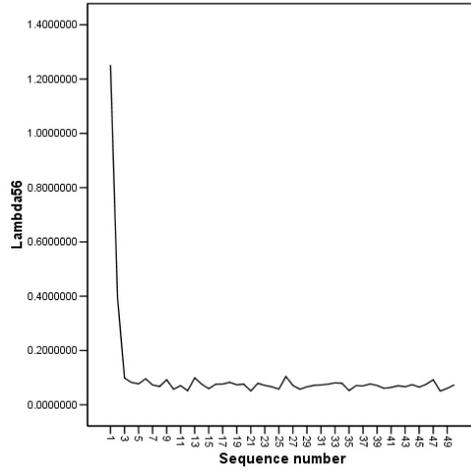
Observemos la velocidad de convergencia de las entradas significativas y la autocorrelación de los datos simulados. El “calentamiento” fue de 500 iteraciones, pero en la gráfica se muestran únicamente las primeras 50 pues éstas son suficientes para mostrar la velocidad de convergencia. Para la autocorrelación se tomaron las últimas 1500 iteraciones.

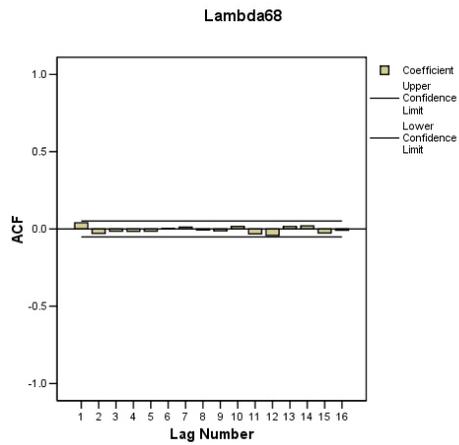
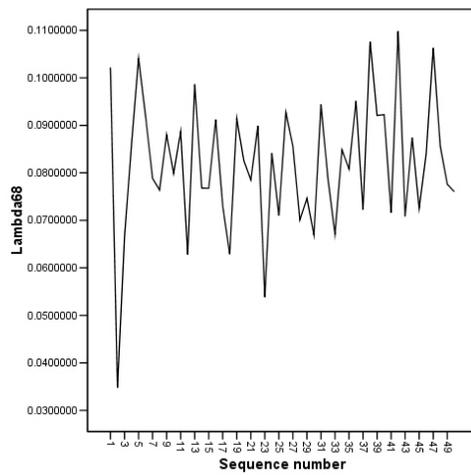
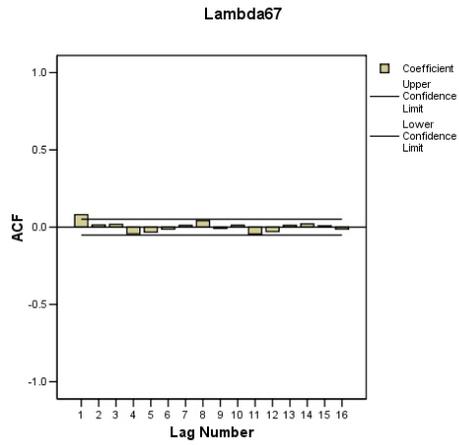
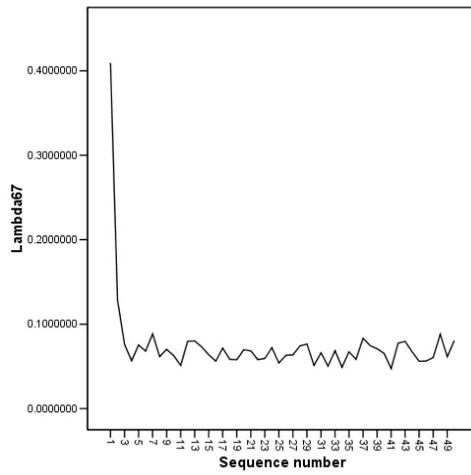
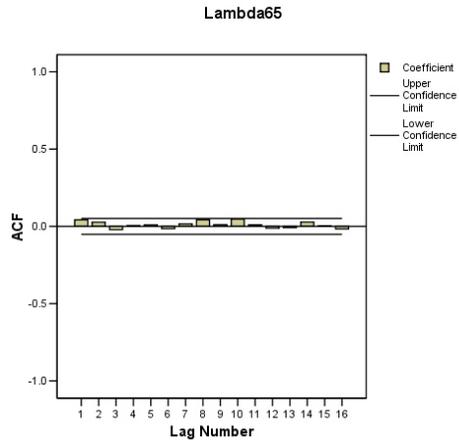
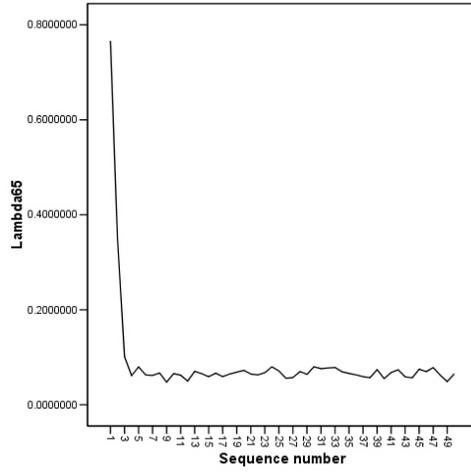


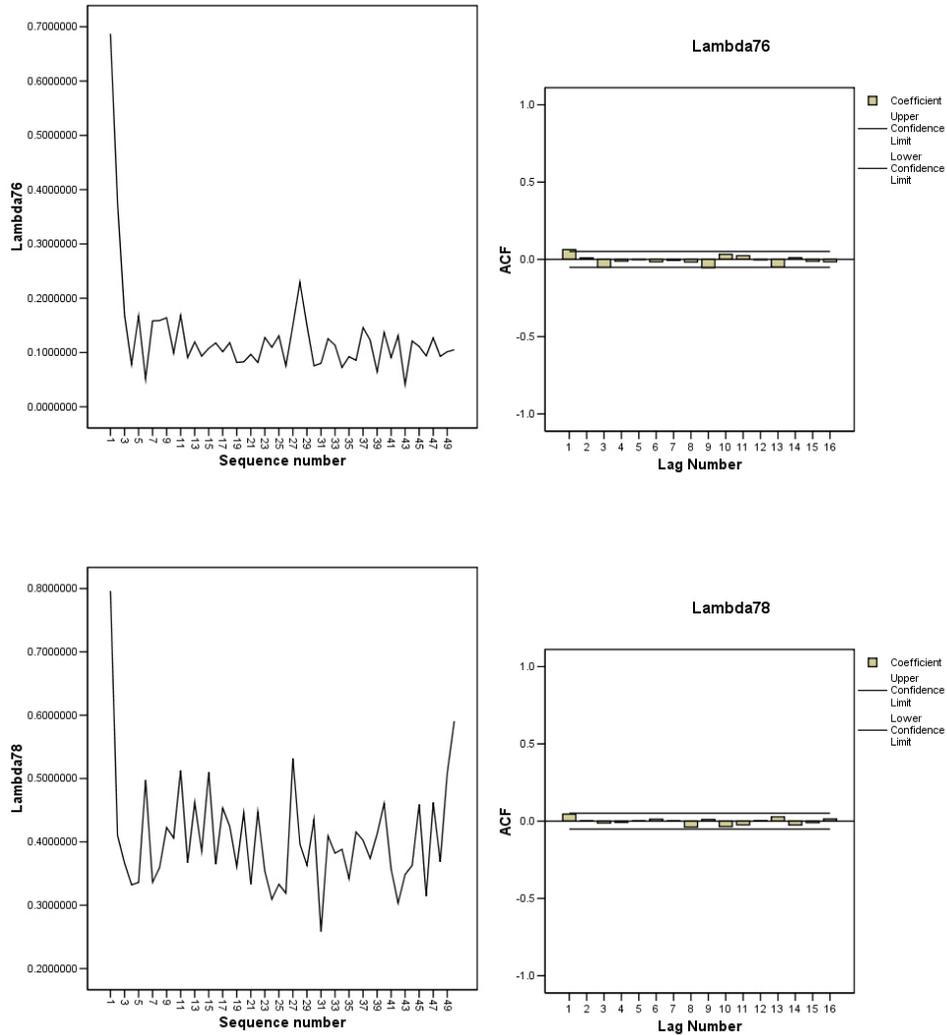












Estas gráficas demuestran que la velocidad de convergencia del método de estimación utilizado es suficientemente rápido, ya que prácticamente en las primeras iteraciones el estimador se vuelve estacional. Las gráficas de autocorrelación muestran que los datos son confiables y que no hay sesgos por alguna dependencia que pudiera existir con las iteraciones anteriores.

Finalmente se cuenta con una buena estimación para la matriz de intensidades del Proceso de Markov para las calificaciones de riesgo de crédito.

CONCLUSIONES

Se han estudiado las Cadenas de Markov y sus propiedades, así como los Procesos de Markov, dentro de los cuales encontramos una propiedad importante: los tiempos de estancia en un estado del proceso antes de saltar a otro siguen una distribución exponencial con un parámetro fijo, el cual sólo depende del estado en el que se encuentre. Los resultados permiten llegar a una función de verosimilitud para los Procesos de Markov. Cuando se cuenta con toda la información del proceso (tiempos de estancia y saltos entre estados de manera continua) es posible estimar la matriz de intensidades por el método de máxima verosimilitud. Esto se complica cuando no se tiene la información de manera continua y sólo se cuenta con observaciones discretas de la cadena; para esto se buscó una manera alternativa de estimación.

Se ha revisado de manera general el método de Monte Carlo vía Cadenas de Markov para la estimación de parámetros de un problema específico, tomando las propiedades de las Cadenas de Markov y fundamentado en la Estadística Bayesiana. De manera iterativa se construye una Cadena de Markov la cual tiene como distribución estacionaria la distribución *a posteriori* de los parámetros de interés. Se tiene de este modo una manera eficiente y factible para estimar parámetros de modelos, que posteriormente servirán para explicar algún fenómeno del que se tenga interés en estudiar. Este trabajo buscó la estimación de la matriz de intensidades de un Proceso de Markov dadas las observaciones discretas que se tengan del mismo. El método MCMC para este caso se basa en una Cadena de Markov en un espacio de Matrices de Intensidades que cumplan con las condiciones dadas por la información discreta con la que se cuenta. El método iterativo nos permite llegar a una estimación eficiente para la matriz de intensidades, siempre que se cumplan las condiciones suficientes para determinar una solución única.

De manera general se estudian las condiciones bajo las cuales es posible obtener una estimación de la matriz de intensidades de un Proceso de Markov observado de manera discreta. Lo anterior debido a que todo Proceso de Markov define una única Cadena de Markov incluida, pero esto no necesariamente se cumple en el sentido inverso. Por esto la importancia de tener conocimiento de las condiciones que determinan la unicidad, para poder concluir si la estimación obtenida tiene sentido. Esto depende de manera importante de la distancia entre las observaciones y el número de ellas con las que se cuente. Si la discretización del Proceso de Markov no proporciona información suficiente se corre el riesgo de que el método MCMC nunca converja a la distribución estacionaria, en cuyo caso el promedio ergódico que se utiliza no sería una estimación adecuada de la matriz de intensidades. Por otro lado, si la discretización no contiene suficiente información, las iteraciones de las distribuciones *a posteriori* pueden volverse dependientes de la distribución *a priori*, y por lo tanto la estimación no sería única, pues para distintas distribuciones *a priori* que se tomaran se tendrían diferentes estimaciones.

Existen una gran diversidad de entidades e instrumentos financieros, que conllevan un riesgo de crédito, esto es que existe la posibilidad de que dichas entidades o instrumentos incumplan sus obligaciones. Por esta razón se tiene interés en estudiar este fenómeno. Moody's nos proporciona una escala de calificaciones de riesgo de crédito y basándose en estas calificaciones se propone modelar este fenómeno como un Proceso de Markov. Debido a que las calificaciones de instrumentos reales es confidencial se tuvieron que simular los datos discretos. La primera estimación propone algunas entradas con valores muy pequeños de la matriz de intensidades para este proceso. Para estos casos se propuso una manera alternativa de determinar si dichos parámetros eran cero. En un segundo intento, después de haber determinado los parámetros que son iguales a cero se volvió a estimar los parámetros restantes, obteniendo una estimación más cercana a la matriz de intensidades original. Este trabajo proporciona un modelo de riesgo de crédito para los instrumentos financieros.

Existen otros fenómenos que se pueden modelar como un Proceso de Markov; para esto es necesario contar con la matriz de intensidades correspondiente al modelo. Esto se puede estimar de manera relativamente sencilla si se cuenta con la información continua. Desafortunadamente existen casos en los que sólo es posible contar con observaciones discretas del fenómeno; para éstos se puede usar el método de estimación propuesto en este trabajo.

BIBLIOGRAFÍA

- Manuel Ammann, **Credit Risk Valuation**, Alemania, 2001, Springer Verlag.
- Papoulis Athanasios, **Probability, Random Variables, and Stochastic Processes**, EUA, Ed. Mc.Graw Hill.
- Mogens Bladt, Michael Sørensen, **Statistical inference for discretely observed Markov jump processes**, 2004.
- Christian Bluhm, Ludger Overbeck, Christoph Wagner, **An introduction to Credit Risk Modeling**, EUA, 2003, Chapman & Hall/CRC.
- Jens Christensen, David Lando, **Confidence sets for continuous-time rating transition probabilities**, 2002.
- Erhan Çinlar, **Introduction to Stochastic Processes**, EUA, 1975, Prentice-Hall, Inc.
- W. J. Culver, **On the existence and uniqueness of the real logarithm of a matrix**, 1966, Proc. Am. Math. Soc.
- J. R. Cuthbert, **The logarithm function for finite-state Markov semi-groups**, 1973, J. London Math. Soc.
- G. Elfving, **Zur Theorie der Markoffschen Ketten**, 1937, Acta Soc. Sci. Fennicae A, 2.
- W. R. Gilks, S. Richardson, D. J. Spiegelhalter, **Markov Chain Monte Carlo in Practice**, Gran Bretaña, 1996, Chapman & Hall.
- J. F. C. Kingman, **The imbedding problem for finite Markov chains**, 1962, Z. Wahrscheinlichkeitstheorie.
- David Lando, Torben M. Skødeberg, **Analyzing rating transitions and rating drift with continuous observations**, Journal of Banking & Finance 26, 2002, pp. 423 – 444.
- B. Walsh, **Introduction to Bayesian Analysis**, 2002, Lecture Notes for EEB 596z.

- <http://www.mathworks.com/access/helpdesk/help/toolbox/mbc/model/techdoc9.html>
- http://www.mathworks.com/cgi-bin/tehis/webinator/search/?db=MSS&prox=page&rorder=750&rprox=750&rdfreq=500&rwfreq=500&rlead=250&sufs=0&order=r&is_summary_on=1&ResultCount=10&query=gamma+distribution
- <http://www.moody.com.mx/mexico/pdf/Introduccion%20a%20las%20calificaciones%20de%20Moody.pdf>
- <http://www.mathworks.com/access/helpdesk/help/toolbox/stats/index.html?/access/helpdesk/help/toolbox/stats/f14937.html>