

*Estimación del orden en un modelo de cadena de markov
oculta no homogéneo con presencia de co-variables*

ANDRYU ENRIQUE MENDOZA BELTRÁN
ESTADÍSTICO, M. Sc. (C)
CÓDIGO: 2949727195



UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE ESTADÍSTICA
BOGOTÁ, D.C.
MAYO DE 2016

*Estimación del orden en un modelo de cadena de markov
oculta no homogéneo con presencia de co-variables*

ANDRYU ENRIQUE MENDOZA BELTRÁN
ESTADÍSTICO, M. Sc. (C)
CÓDIGO: 2949727195

DISERTACIÓN PRESENTADA PARA OPTAR AL TÍTULO DE
MAGISTER EN CIENCIAS - ESTADÍSTICA

DIRECTOR
JUAN MANUEL JULIO ROMÁN, M.Sc.
MAGISTER EN ESTADÍSTICA

LÍNEA DE INVESTIGACIÓN
PROCESOS ESTOCÁSTICOS Y MODELOS EN SERIES TEMPORALES



UNIVERSIDAD NACIONAL DE COLOMBIA
FACULTAD DE CIENCIAS
DEPARTAMENTO DE ESTADÍSTICA
BOGOTÁ, D.C.
MAYO DE 2016

Título en español

Estimación del orden en un modelo de cadena de markov oculta no homogéneo con presencia de co-variables

Title in English

Estimate the order in a Non Homogeneous Hidden Markov Model with precense of co-variables

Resumen: El presente documento muestra la estimación del orden o número de estados de la cadena, en un modelo en cadenas de markov ocultas no homogéneas usando la inferencia bayesiana. Para la estimación, se usa el método de Markov Chain Monte Carlo (MCMC), tal que la simulación se realiza de manera conjunta con los demás parámetros del modelo. Adicionalmente cada variable del proceso observado pertenece a la familia exponencial. El uso de esta metodología establece el modelo que mejor ajusta los datos. Estos valores son generados de distribuciones no pseudo a priori, obteniendo convergencia e independencia a un gran número de iteraciones.

Abstract: This document shows the estimation of the order or number of states of the Chain, in a non homogeneous hidden markov model using the bayesian inference. For the estimation, we used the Markov Chain Monte Carlo method's (MCMC) such that the simulation was performed in conjunction with the other parameters of the model, additionally each variable of the observed process belongs to the exponential family. The use of this method select the best model, This values was generated by the non pseudo prior distribution, obtaining convergence and not autocorrelation to a large number of iterations

Palabras clave: Modelos en Cadenas de Markov Ocultos No Homógeneos, Inferencia Bayesiana, Métodos de Markov Chain Monte Carlo (MCMC), Distribucion de la familia Exponencial

Keywords: Non Homogeneous Hidden Markov Model, Bayesian Inference, Markov Chain Monte Carlo Methods (MCMC), Exponential Family Distribution

Nota de aceptación

Trabajo de tesis

..

“Mención ..”

Jurado

..

Jurado

..

Jurado

..

Director

..

Codirector

..

Bogotá, D.C., 25 de Noviembre de 2016

Dedicado a

A quien le interese descubrir el mundo.

Agradecimientos

A la comprensión de mi familia, mi novia y mis amigos. Adicionalmente al apoyo recibido por parte de mi director, así como al departamento de Estadística por brindarme esta oportunidad.

Índice general

Índice general	I
Índice de tablas	III
Índice de figuras	IV
Introducción	V
1. Objetivos	1
1.1. Objetivo general	1
1.2. Objetivos específicos	1
2. Marco teórico	2
2.1. Modelo en Cadena de Markov oculto homogéneo	2
2.1.1. Estimación	3
2.1.2. Función de filtro y suavizamiento	4
2.2. Conjunto de distribuciones de la familia exponencial	5
2.3. Inferencia bayesiana	5
2.4. Métodos de simulación	7
2.4.1. Aproximación no iterativa	7
2.4.2. Aproximación MCMC	7
3. Modelo	10
3.1. Función de verosimilitud	11
3.2. Características	12
4. Inferencia	15
4.1. Inferencia Bayesiana de los NHHMM	15

4.2. Distribuciones a prior	16
4.3. Estimación y simulación de los parámetros	18
4.3.1. Distribuciones <i>a posteriori</i> de cada parámetro condicional a los demás parámetros	19
4.3.2. Proceso de simulación	23
5. Exploración de la inferencia Bayesiana	26
5.1. Caso Poisson	26
5.2. Caso Binomial	27
5.3. Caso exponencial	28
5.4. Exploración de la metodología	29
5.4.1. Resultados en la inferencia	30
Conclusiones	39
Trabajo futuro	40
Bibliografía	41

Índice de tablas

5.1. Frecuencia absoluta y relativa de las iteraciones simuladas del número de estados con la serie de tres y cuatro estados.	31
5.2. Resumen estadístico de las simulaciones realizadas para los parámetros Λ en cada uno de los modelos después de estimar con tres y cuatro estados para las series trabajadas.	33

Índice de figuras

5.1. Serie de tiempo simulada a partir de un Poisson-NHHMM con 3 estados, así como la función de densidad y el autorrelograma (Fuente propia).	30
5.2. Serie de tiempo simulada a partir de un Poisson-NHHMM con 4 estados, así como la función de densidad y el autorrelograma (Fuente propia).	31
5.3. Series de tiempo simuladas a partir de NHHMM variando la estructura de las co-variables que ingresan al modelo (Fuente propia).	32
5.4. Simulación del número de estados en un total de 50000 iteraciones al inferir sobre dos series de tiempo de 3 y 4 estados (Fuente propia).	33
5.5. Gráfico de densidad de la simulación <i>a posteriori</i> de los parámetros Lambda tomando en cuenta que los Lambda no fueron de una Pseudo prior con la simulación con la serie de 3 estados y la distribución prior para cada uno de ellos (Fuente propia).	34
5.6. Gráfico de densidad de la simulación <i>a posteriori</i> de los parámetros Lambda tomando en cuenta las que no fueron de una pseudo <i>a priori</i> con la simulación de una serie de 4 estados y la distribución prior para cada uno de ellos (Fuente propia).	35
5.7. Gráfico de densidad de la simulación <i>a posteriori</i> de los parámetros Lambda tomando en cuenta los valores simulados que no fueron de una pseudo <i>a priori</i> con la simulación de la serie de 3 estados y la distribución prior para cada uno de ellos (Fuente propia).	36
5.8. Gráfico de densidad de la simulación <i>a posteriori</i> de los parámetros Lambda tomando en cuenta las que no fueron de una pseudo <i>a priori</i> con la simulación de la serie de 4 estados y la distribución a prior i para cada uno de ellos (Fuente propia).	37
5.9. Gráfico de contraste entre los parámetros simulados de los β para la serie con tres estados (Fuente propia).	38

Introducción

Los modelos en cadena de markov oculta (*HMM*), así como su generalización a los modelos no homogéneos (*NHHMM*) usan las propiedades y características de las Cadenas de Markov para capturar la multimodalidad, la alta curtosis, la dependencia y otros comportamientos que se nota en la exploración de algunos procesos observados. Dicho uso se hace mediante la interacción del proceso estocástico $\{Y_t\}_{t \in N}$ y una Cadena de Markov $\{Z_t\}_{t \in N}$ discreta, tal que la distribución de Y_t depende del valor de $Z_t = z_t$; es decir, $Y_t|Z_t = i \sim f_i(Y|\theta_i)$.

En vista de lo anterior, y trajando con una Cadena de Markov discreta, esta se caracteriza por el número de estados K , la matriz de transición y las probabilidades de estado inicial. Al desconocer la Cadena, el número de estados tambien es desconocido y su estimación es de gran inpacto a la inferencia de los demás parámetros del modelo.

Usualmente el parámetro K asocia un modelo con cada posible número de estados, así que la inferencia del mismo se puede ver como el contraste de modelos potenciales definidos mediante algún criterio de selección. El cálculo, la alta cantidad de parámetros y la no identificabilidad del modelo hace que no sea consistente la estimación, principalmente cuando se modelan series categóricas. Este documento muestra la estimación del número de estados de la Cadena en un (*NHHMM*) mediante el uso de la inferencia Bayesiana, donde la estimación puntual del parámetro se realiza a través de la metodología expuesta en Carlin y Chib (1995 [3]).

Alrededor de este tema se han publicado varias investigaciones, entre las que se encuentra el libro de Zucchini y MacDonald(2009 [14]) donde se plantea el uso de las estadísticas BIC y AIC para la inferencia del número de estado en un *HMM*; una vez el número de estado se estima, la inferencia de los demás parámetros se puede realizar de manera clásica, maximizando la función de verosimilitud o mediante el uso del algoritmo Baum y Welch(2003 [13]). Por otra parte MacKay(2000 [12]) establece un estimador consistente del orden de la Cadena de Markov mediante el uso de las distancias entre las distribuciones de probabilidad teórica con la probabilidad empírica. Más adelante Chopin y Pelgrin (2004 [5]) y Chopin (2007 [4]) muestran la inferencia del número de estados a través de la construcción de un proceso alterno desconocido, donde se actualiza la estimación a medida que aumenta una nueva observacion.

En los *NHHMM*, Hughes, Guttorp y Charles(1999 [10]) usan el contraste de modelos bajo un criterio de selección para la estimación del número de estado de la Cadena. En esa investigación se modela un proceso dicotómico que relaciona la precipitación en diversos territorios. También Spezia (2006 [11]) muestra la inferencia Bayesiana de los *NHHMM Gaussiano* con un enlace logístico, usando los factores de Bayes para la estimación del número de estado.

Partiendo de los planteamientos anteriores y de los intereses de esta investigación, el presente documento se dividió en cinco capítulos: el primero aborda los objetivos de este trabajo; el segundo presenta el marco teórico de la investigación, en él se describen los principales desarrollos obtenidos en estos modelos y de manera general las herramientas teóricas para el cumplimiento de los objetivos; el tercer capítulo especifica y profundiza en el modelo a trabajar en la investigación, sus supuestos y características principales; la cuarta sección expone el desarrollo inferencial propuesto para el número de estados; más adelante en el quinto capítulo se realiza la exploración de la metodología con un par de simulaciones realizadas, en este caso se usa un *NHHMM Poisson*. Finalmente se presentan los hallazgos mediante un apartado de conclusiones y uno de trabajos futuros.

CAPÍTULO 1

Objetivos

1.1. Objetivo general

Estimar el orden o número de estados en un modelo de Cadenas de Markov oculto no homogéneo dada la presencia de co-variables cuando la distribución del proceso observado pertenece a la familia exponencial.

1.2. Objetivos específicos

- Plantear y desarrollar la estructura teórica y práctica de la distribución a posteriori conjunta de los parámetros en un Modelo en Cadenas de Markov oculta no Homogénea así como las distribuciones a posteriori condicionales para la implementación del Muestrador de Gibbs.
- Plantear las estructuras teóricas y prácticas de las distribuciones a priori y pseudo a priori en la inferencia Bayesiana de los parámetros de un Modelo en Cadenas de Markov ocultas no homogéneas siguiendo el planteamiento mostrado en Carlin B.P. and Chib S. (1995 [3]).
- Explorar y determinar las estadísticas resumen para la estimación puntual de los parámetros en un *NHMM* así como las características de las simulaciones para la inferencia de cada uno de ellos.

Marco teórico

Durante la investigación interactuaron un conjunto de herramientas teóricas y computacionales que son base y línea en el desarrollo de los objetivos planteados. En el presente capítulo se muestran estas herramientas de manera general, ya que son referenciados principalmente en el tercero y cuarto capítulo, donde se aborda el entorno inferencial de los *NHHMM* cuando la distribución condicional de una variable aleatoria del proceso pertenece a la familia exponencial.

Ya que esta investigación se fundamenta en la generalización de los *HMM* a los *NHHMM*, el primer apartado de este capítulo resume algunos componentes teóricos de los *HMM* tomando como base a Frühwirth-Schnatter (2006, cap. 10 y 11 [7]) y Zucchini y MacDonald(2009 [14]), y algunos artículos adicionales que muestran las metodologías inferenciales de estos modelos. El segundo momento, expone de manera general el entorno en los modelos lineales generalizados (GLM), principalmente la forma funcional de distribuciones de probabilidad que pertenecen a la familia exponencial, los conceptos mostrados en este tema toman como base a Dobson y Barnett (2008 [6]). Más adelante el tercer apartado muestra el entorno general de la inferencia Bayesiana y el quinto y último aspecto aborda los algoritmos de simulación usados en la investigación. Las fuentes principales que se han consultado y referenciado sobre este tema son tomadas de Gamerman y Lopes (2006 [8]), Gelman, Carlin, Stern, Dunson, Vehtari y Rubin(2009 [9]), y Albert (2011 [1]).

2.1. Modelo en Cadena de Markov oculto homogéneo

Los modelos en Cadenas de Markov ocultos, no son más que la generalización de los modelos de mixturas independientes a modelos de mixturas dependientes, de modo que la dependencia es guiada con una Cadena de Markov no observada. Es decir, si $Y_{1:T} = \{Y_t\}_{t=1:T}$ es una muestra de T periodos y $Z_{1:T} = \{Z_t\}_{t=1:T}$ una Cadena de Markov con espacios medibles $(\mathcal{Y}, \mathfrak{S}_Y)$ y $(\mathcal{Z} = \{1, \dots, K\}, \mathfrak{S}_Z)$ para cada variable aleatoria Y_t y Z_t respectivamente, entonces $P(Y_t = y_t | Z_t = k) \sim f(y_t | \theta_k)$ con $y_t \in \mathcal{Y}$ y $k \in \mathcal{Z}$.

Una forma lineal básica de estos modelos se presenta en la ecuación 2.1.

$$Y_t|z_t = \sum_{i=1}^K \mu_i I_{z_t=i}(Z_t) + \sum_{i=1}^K \epsilon_i I_{z_t=i}(Z_t) \quad (2.1)$$

Donde $y_{1:T} = \{y_t\}_{t=1:T} \in \mathcal{Y}^T$ es una realización de proceso de interés y $z_{1:T} = \{z_t\}_{t=1:T} \in \mathcal{Z}^T$ es la ejecución de la Cadena de Markov. Al suponer que $\epsilon_i \sim N(0, \sigma_i)$ se tiene que $Y_t|Z_t = k \sim N(\mu_k, \sigma_k), \forall i \in \mathcal{Z}$. La ecuación 2.1 relaciona un modelo conocido como *HMM-Gaussiano*.

El conjunto de parámetros de los *HMM* se dividen en tres partes: por un lado el orden o número de estados de la Cadena de Markov oculta (K), condicionado K , el segundo conjunto de parámetros relaciona la distribución del proceso observado $\underline{\theta} = (\underline{\theta}_1, \underline{\theta}_2, \dots, \underline{\theta}_K)$, para el modelo 2.1 $\underline{\theta}_i = (\mu_i, \sigma_i)$, y finalmente el tercer conjunto de parámetros está conformado por $\Gamma = \{\gamma_{ij}\}_{i,j=1,\dots,K}$ y $\underline{\delta} = (\delta_1, \dots, \delta_K)$ que son los parámetros de la Cadena de Markov oculta. Al suponer la independencia de cada variable aleatoria condicional $Y_t|Z_t = z_t$ con otra variable aleatoria condicional en otro momento, la función de verosimilitud es dada por la ecuación 2.2 dado y_1^T .

$$\begin{aligned} P(Y_{1:T}|\xi_1^K) &= \sum_{z_1^T \in \mathcal{Z}} P(Y_{1:T}|Z_{1:T} = z_{1:T}, \xi_1^K) P(Z_{1:T} = z_{1:T}|\xi_1^K) \quad (2.2) \\ &= \sum_{z_{1:T} \in \mathcal{Z}} \prod_{t=1}^T P(Y_t|Z_t = z_t, \xi_1^K) P(Z_0 = z_0) \prod_{t=1}^T P(Z_t|Z_{t-1} = z_t, \xi_1^K) \\ &= \sum_{z_{1:T} \in \mathcal{Z}} \prod_{i=1}^K \prod_{j=1}^{n_i} P(Y_j|Z_t = i, \theta_i) P(Z_0 = z_0) \prod_{i=1}^K \prod_{j=j}^{K-1} \gamma_{ij}^{n(i,j)} \end{aligned}$$

En este caso $n(i, j)$ y n_i representan el número de veces que pasa la cadena de Markov del estado i al estado j y el número de veces que se encuentra la cadena en el estado i .

2.1.1. Estimación

La inferencia de los parámetros en un *HMM* se puede realizar de manera clásica al encontrar los valores que maximizan la función de verosimilitud. Dada la complejidad de la función y las múltiples sumas que se presentan a medida que el proceso tiene más observaciones, en Zucchini y MacDonald (2009, cap. 2, eq. 2.12 [14]) se escribe la función de verosimilitud de forma matricial como se puede ver a continuación:

$$L_T = P(Y_{1:T} = y_{1:T}) = \delta \Gamma P(y_1) \Gamma P(y_2) \dots \Gamma P(y_T) \underline{1}' \quad (2.3)$$

Esta forma eficaz permite calcular la función de manera iterativa tal que $L_t = \alpha_t \underline{1}'$ donde $\alpha_t = \alpha_{t-1} \Gamma P(y_t)$ y $P(y_t)$ es una matriz diagonal con valor i -ésimo de la diagonal igual a $P(Y_t = y_t|Z_t = i)$. Con esta forma de escribir la función de verosimilitud se aplica

algún algoritmo de maximización tal como lo muestra Zucchini y MacDonald (2009, cap. 3, eq. 3.1 [14]). Otra manera de calcular la estimación de los parámetro es mediante el algoritmo de Baum Welch, el cual se fundamenta en el método EM para estos modelos, una profundización se puede observar en Zucchini y MacDonald (2009, cap. 4 [14]).

2.1.2. Función de filtro y suavizamiento

La estimación de la Cadena de Markov no observada, así como la inferencia sobre los parámetros que interactúan con ella son de gran importancia para el pronóstico, la evaluación de la estructura del proceso observado y el ajuste del modelo, entre otras utilidades. Su inferencia se fundamenta en el análisis de la probabilidad de la cadena, dado el conocimiento de los demás parámetros y el proceso observado.

$$P(Z_t | \underline{\Theta}, y_1^T) \quad (2.4)$$

La probabilidad 2.4 depende de la relación entre t y τ , ya que si $t = \tau$ se conoce como la distribución de **Filtro**, si $t < \tau$ se conoce como la función de **Suavizamiento** y si $t > \tau$ la probabilidad 2.4 se conoce como la función de **pronóstico**. Para encontrar las probabilidades de filtro, se usa un algoritmo en dos etapas de manera iterativa, tal como se puede ver en el contenido subsiguiente:

a) Para un momento t se obtiene $P(Z_1^t = z_1^t | y_1^{t-1}, \underline{\Theta})$

$$P(Z_1^t = z_1^t | y_1^{t-1}, \underline{\Theta}) = P(Z_t = z_t | Z_{t-1} = z_{t-1}, y_{t-1}, \underline{\Theta}) P(Z_1^{t-1} = z_1^{t-1} | y_1^{t-1}, \underline{\Theta}) \quad (2.5)$$

b) Con el conocimiento de 2.5, calculamos la probabilidad de filtro tal como se muestra en 2.6.

$$\begin{aligned} P(Z_1^t = z_1^t | y_1^t, \underline{\Theta}) &= \frac{P(Z_1^t = z_1^t, y_t | y_1^{t-1}, \underline{\Theta})}{P(y_t | y_1^{t-1}, \underline{\Theta})} \\ &= \frac{P(y_t | Z_1^t = z_1^t, y_1^{t-1}, \underline{\Theta}) P(Z_1^t = z_1^t | y_1^{t-1}, \underline{\Theta})}{\sum_{k=1}^K P(y_t, Z_t = k | y_1^{t-1}, \underline{\Theta})} \\ &= \frac{P(y_t | Z_1^t = z_1^t, y_1^{t-1}, \underline{\Theta}) P(Z_1^t = z_1^t | y_1^{t-1}, \underline{\Theta})}{\sum_{k=1}^K P(y_t | Z_t = k, y_1^{t-1}, \underline{\Theta}) P(Z_t = k | y_1^{t-1}, \underline{\Theta})} \end{aligned} \quad (2.6)$$

El cálculo inicia con la distribución $P(Z_1 = z_1 | y_0, \underline{\Theta}) = \sum_{k=1}^K P(Z_1 = z_1 | Z_0 = k) P(Z_0 = k)$, y se realiza de manera iterativa hasta encontrar el valor de $P(Z_1^t | Y_1^t, \underline{\Theta})$.

Una vez se obtiene la distribución de Filtro en cada momento del proceso, se logra el cálculo de las probabilidades de la Cadena tomando en cuenta toda la información y_1^T . Esto requiere el cálculo de las probabilidades de **suavizamiento**, el cual contempla el siguiente proceso iterativo:

a) Una vez calculadas las probabilidades de filtro se calculan las probabilidades de suavizamiento, iterativamente, desde el último periodo T al periodo 1, se debe iniciar con $P(Z_{1:T} = z_{1:T} | Y_{1:T}, \underline{\Theta})$.

b) Se obtienen las probabilidades de suavizamiento mediante la ecuación 2.7.

$$P(Z_t = z_t | y_1^T, \Theta) = \sum_{k=1}^K \frac{P(Z_{t+1} = k | Z_t = z_t, y_{1:t}, \Theta) P(Z_t | y_{1:t}, \Theta) P(Z_{t+1} | y_{1:T}, \Theta)}{\sum_{j=1}^K P(Z_{t+1} = k | Z_t = j, y_{1:t}, \Theta) P(Z_t = j | y_{1:t}, \Theta)} \quad (2.7)$$

La demostración de la ecuación 2.7 se realiza en Frühwirth-Schnatter (2006, cap. 11.2.4 [7]), proceso durante el cual se obtiene que:

$$P(Z_{1:t} | Z_{t+1} = k, y_{1:T}, \Theta) = \frac{P(Z_{t+1} = k | Z_{1:t} = z_{1:t}, y_{1:t}, \Theta) P(Z_{1:t} = z_{1:t} | y_{1:t}, \Theta)}{\sum_{j=1}^K K P(Z_{t+1} = k | Z_t = j, y_{1:t}, \Theta) P(Z_t = j | y_{1:t}, \Theta)} \quad (2.8)$$

El desarrollo de esta función es de gran importancia en la inferencia Bayesiana de los parámetros y es abordado con profundidad en Frühwirth-Schnatter (2006, cap. 11.2.4, eq. 11.10 [7]).

2.2. Conjunto de distribuciones de la familia exponencial

Una de las ventajas y facilidades que tienen los *HMM* es la exploración, descripción y pronóstico de procesos estocásticos discretos, tal como lo son las series de conteo o las series de presencia y ausencia de una característica de un individuo en un entorno temporal, espacial o espacio-temporal. De forma análoga los *NHMM* facilitan el modelamiento sobre procesos que pertenecen a la familia exponencial vinculando la influencia de co-variables al proceso.

Basados en la función exponencial, dicha familia está conformada por un conjunto de distribución de probabilidad que cumple con algunas propiedades en común para el desarrollo inferencial de los mismos. Una distribución de probabilidad que pertenece a esta familia se puede escribir como:

$$f(y|\theta) = s(y)t(\theta)e^{a(y)b(\theta)} = e^{a(y)b(\theta)+c(\theta)+d(y)} \quad (2.9)$$

Donde a, b, s y t son funciones conocidas y a partir de esta forma estándar se puede explicar que $E(a(Y)) = -c(\theta)' / b(\theta)'$ y $Var(a(Y)) = (b(\theta)''c(\theta)' - c(\theta)''b(\theta)') / (b(\theta)')^3$. La descripción fue tomada de Dobson y Barnett (2008 [6]).

2.3. Inferencia bayesiana

El presente capítulo menciona algunos conceptos básicos de la inferencia Bayesiana. Para mayor profundización, se recomienda consultar a Gamerman y Lopes (2006, cap. 2 [8]) o Gelman, Carlin, Stern, Dunson, Vehtari y Rubin (2009, cap. 3 [9]), los cuales muestran a mayor profundidad el desarrollo de estos temas y el uso de las herramientas tecnológicas para la inferencia.

Uno de los objetivos inferenciales es lograr conocer, bajo un grado de incertidumbre, el comportamiento de la población identificando los valores de una muestra de ella. Dada esta necesidad, la inferencia clásica se fundamenta en el análisis de la función de verosimilitud, $f(\underline{y}|\underline{\theta})$, que obtiene como estimación los valores de los parámetros $\underline{\theta}$ y las características que maximicen la función.

Así como la inferencia clásica vincula los valores de la muestra por medio de la distribución de verosimilitud, la **inferencia Bayesia** también lo toma en cuenta, sin embargo la metodología cambia un poco al asumir que los parámetros son variables aleatorias. Al ser variables aleatorias presentan una distribución de probabilidad, ya que las características de estas dan información de la estimación de los parámetros una vez se conozcan los valores de la muestra. En este caso, se toma como evento de interés la ocurrencia de los parámetros y como evento condicional al conjunto de valores de la muestra, que es observado, de esta manera la distribución de los parámetros es dada por:

$$P(\underline{\theta}|\underline{y}) = \frac{P(\underline{\theta}, \underline{y})}{P(\underline{y})} = \frac{P(\underline{y}|\underline{\theta})P(\underline{\theta})}{P(\underline{y})} \quad (2.10)$$

El denominador de la ecuación que se puede ver en la figura 2.10 representa la distribución marginal de los datos, que corresponde a una constante para la distribución de los parámetros. Por ende se tiene que $P(\underline{\theta}|\underline{y}) \propto P(\underline{y}|\underline{\theta})P(\underline{\theta})$. La componente $P(\underline{\theta}|\underline{y})$ se conoce como la distribución **a posteriori**, la función $P(\underline{y}|\underline{\theta})$ es la distribución de **verosimilitud** y a la distribución $P(\underline{\theta})$ se le conoce como **a prior**.

Jeffrey (1961), propone y desarrolla el entorno inferencial Bayesiano basado en una distribución *a priori* constante; la cual es conocida como la distribución *a priori* de Jeffrey $P(\underline{\theta}) \propto |I(\underline{\theta})|^{1/2}$, donde se plantea:

$$I(\underline{\theta}) = E\left[-\frac{d \log(f(\underline{y}|\underline{\theta}))}{d\theta d\theta^2} \mid \underline{\theta}\right] \quad (2.11)$$

Es usual tomar como distribución *a priori* una función cuya distribución *a posteriori* pertenezca a la misma familia. A dichas distribuciones *a priori* se les conocen como **Conjugadas**, en Gelman, Carlin, Stern, Dunson, Vehtari y Rubin(2009, cap 2, pg 37 [9]) se muestra de manera general una a distribución *a priori* Conjugada cuando la distribución del proceso observado pertenece a la familia exponencial. De esta manera, si y pertenece a la familia exponencial, entonces se propone $f(\theta) \propto t(\theta)^\eta e^{b(\theta)\nu}$, de tal manera que la distribución *a posteriori* sea como se evidencia en la figura 2.12.

$$P(\theta|y) \propto t(\theta)^{\eta+n} e^{b(\theta)(\nu+a(y))} \quad (2.12)$$

2.4. Métodos de simulación

Tal como mencionamos anteriormente, el objetivo de la inferencia Bayesiana se fundamenta en el análisis de la distribución de los parámetros, específicamente la *a posteriori*. Esta distribución se obtiene con la multiplicación de la distribución *a priori* y la función de verosimilitud, es decir $h(\theta) = f(y|\theta)g(\theta)$. Entonces, el resumen estadístico consiste en el cálculo de la integral de una transformación de los parámetros $T(\theta)$, bajo esta distribución, es decir:

$$E[t(\Theta|y)] = \frac{\int t(\theta)g(\theta)f(y|\theta)}{\int g(\theta)f(y|\theta)} \quad (2.13)$$

Por ejemplo, al tener $T(\theta) = \theta$ o $T(\theta) = \theta^2$ se calcula el promedio o el segundo momento de la distribución *a posteriori*. Dentro de las transformaciones más usadas para la inferencia se encuentra el cálculo de los percentiles y las distribuciones marginales, cuando se cuenta con un vector de parámetros, entre otras.

Partiendo de que la estructura de la distribución *a posteriori* consta de la multiplicación de distribuciones de probabilidad, que no necesariamente pertenecen a la misma familia, su forma puede ser no conocida y el cálculo analítico en 2.13 imposible de realizar. Por ende, se requiere de la utilización de métodos computacionales que permitan el cálculo de 2.13 y desarrollos adicionales para la estimación e inferencia de los parámetros.

2.4.1. Aproximación no iterativa

Al obtener la distribución *a posteriori* $\pi(\underline{\theta})$ de un conjunto de parámetros $\underline{\theta}$, una vez se tenga el conocimiento de una muestra de datos $\underline{y} = (y_1, y_2, \dots, y_n)$ con función verosímil $f(\underline{y}|\underline{\theta})$ y distribución *a priori* $g(\underline{\theta})$, se obtiene la aproximación de Taylor de segundo grado del logaritmo de la distribución alrededor de la moda *mod*, mediante la siguiente ecuación:

$$\log(\pi(\underline{\theta})) \propto \log(\pi(mod)) - \frac{1}{2}(\underline{\theta} - mod)\left[-\frac{\partial^2 \log \pi(mod)}{\partial \underline{\theta} \partial \underline{\theta}'}\right](\underline{\theta} - mod)' \quad (2.14)$$

Esta aproximación permite identificar que $\underline{\theta} \sim N(\mu = mod, Var = V)$, donde $V = \left(-\frac{\partial^2 \log \pi(mod)}{\partial \underline{\theta} \partial \underline{\theta}'}\right)^{-1}$. Una vez obtenida la aproximación, la estimación puntual del parámetro se realiza con el cálculo de la moda. Una forma de obtener el estadístico es por medio del algoritmo de Newton Raphson o mediante el algoritmo de Nelder Mead para el caso multivariado.

2.4.2. Aproximación MCMC

Una segunda manera para calcular la integral en 2.13 es por medio del método de Montecarlo, donde se simula una gran cantidad de valores de la distribución *a posteriori*, de este modo el cálculo de la media de una transformación es dada por:

$$E(h(\underline{\theta}|y)) = \int h(\underline{\theta})\pi(\underline{\theta})d\underline{\theta} \propto \frac{\sum_{j=1}^m h(\underline{\theta}_j)}{m} = \bar{h} \quad (2.15)$$

Donde $h(\underline{\theta}_j)$ es la transformación del valor $\underline{\theta}$ simulado de la distribución *a posteriori*. Adicionalmente se puede obtener una estimación del error estándar por medio de la simulación de valores a partir del cálculo que podemos ver a continuación:

$$se_{\bar{h}} = \sqrt{\frac{\sum_{j=1}^m (h(\underline{\theta}_j) - \bar{h})^2}{(m-1)m}}$$

Con el cálculo para la determinación de las características de la distribución *a posteriori*, el cómputo se fundamenta en la simulación de un conjunto de valores que provengan de una distribución de probabilidad. En algunos casos, la simulación es bastante simple, siempre y cuando esta tenga una forma usual conocida; no obstante, es común encontrar el caso en el que la distribución no sea una forma conocida y es necesario el uso de algoritmos para descubrirla. Uno de ellos es el algoritmo de **Metropolis Hastings** el cual simula una Cadena de Markov continua *irreducible* y *aperiódica* de tal modo que la distribución estacionaria es la distribución *a posteriori*. Para ello es importante tener en cuenta que si una Cadena de Markov es irreducible y aperiódica el vector estacionario de valores es único.

El algoritmo de **Metropolis Hasting** simula iterativamente un conjunto de valores, iniciando en un valor propuesto de $\underline{\theta}_0$ de tal manera que para la iteración t , se contrastan los valores $\underline{\theta}_t$ con $\underline{\theta}_{t-1}$ bajo una regla de decisión. Dicho algoritmo propone una distribución Q de la cual se simulan los valores para la determinación de la pertenencia a la distribución; dado el conocimiento del valor de los parámetros $\underline{\theta}_{t-1}$, la simulación del siguiente valor se da a través de estos pasos:

- Se simula el valor de $\underline{\theta}^*$, a partir de la distribución $Q(\underline{\theta}^*|\underline{\theta}_{t-1})$
- Se calcula la razón a través de esta ecuación:

$$R = \frac{\pi(\underline{\theta}^*)Q(\underline{\theta}_{t-1}|\underline{\theta}^*)}{\pi(\underline{\theta}_{t-1})Q(\underline{\theta}^*|\underline{\theta}_{t-1})} \quad (2.16)$$

- Calculamos la probabilidad de aceptación $P = \min(1, R)$.
- Se muestrea el valor de $\underline{\theta}^t$, ya que $\underline{\theta}^t = \underline{\theta}^{ast}$ con probabilidad P . De otra manera se puede presentar así: $\underline{\theta}^t = \underline{\theta}^{ast}$.

Dentro de las alternativas para la función Q se encuentra una distribución independiente $Q(\underline{\theta}^*|\underline{\theta}_{t-1}) = Q(\underline{\theta}^*)$ o una distribución simétrica $Q(\underline{\theta}^*|\underline{\theta}_{t-1}) = Q(\underline{\theta}^* - \underline{\theta}_{t-1})$. El segundo caso se conoce como un algoritmo de Metropolis Hastings en ruido blanco. En este caso se toma:

$$\underline{\theta}^* = \underline{\theta}_{t-1} + HZ \quad (2.17)$$

Donde H se conoce como un valor de escala y Z es un cuantil de la distribución normal con media cero y varianza V . En el caso de haber propuesto una distribución independiente se simula $\underline{\theta}^*$ de una normal con media μ y varianza V .

Una vez simulados muchos valores de la distribución *a posteriori* se toman algunos de estos a partir de cierto punto, de tal manera que se garantice la no correlación entre ellos, para usarlos en el cálculo de la integral mediante el método de Montecarlo. Recordemos que los valores que se simularon corresponden a una Cadena de Markov con distribución estacionaria $\pi(\underline{\theta})$; por ende el método se conoce como el método de *Montecarlo Markov Chain (MCMC)*.

Uno de los grandes inconvenientes se encuentra cuando el número de parámetros es bastante grande y los métodos de simulación conjunta presentan inconsistencias o errores en la convergencia. Para esos casos se recomienda el uso del algoritmo de **Gibbs**. El algoritmo permite la simulación de un vector de parámetros de la distribución conjunta a partir de la simulación de valores de las distribuciones *a posteriori* de cada parámetro condicionado al resto. Es decir, si queremos simular valores de $\underline{\theta} = (\theta_1, \theta_2, \dots, \theta_m)$ a partir de la distribución conjunta $P(\underline{\theta} = (\theta_1, \theta_2, \dots, \theta_m) | \underline{y})$, donde \underline{y} es una muestra de valores observada, se empieza con el planteamiento de valores iniciales $\underline{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_m^{(0)})$ para cada uno de ellos. La simulación en la iteración t para el parámetro $\theta_i^{(t)}$ se realiza de acuerdo con la distribución condicional:

$$P(\theta_i | \theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_{i-1}^{(t)}, \theta_{i+1}^{(t-1)}, \dots, \theta_m^{(t-1)}) \quad (2.18)$$

De esta manera el algoritmo finaliza ciclos de simulación en cada iteración. El principio del algoritmo se fundamenta en que una función conjunta que no presenta una forma conocida o fácil de simular, pero es proporcional a las distribuciones condicionales de cada parámetro, o sea que cada una de ellas puede pertenecer a una forma conocida o fácil de simular, de esta manera puede simular valores de los componentes alternamente. No obstante, puede que la simulación de la ecuación 2.18 no sea conocida y usemos el algoritmo de Metropolis Hastings; en este caso usamos un Metropolis dentro de un Gibbs.

Modelo

Sea $\{Y_t\}_{t \in \mathcal{N}}$ un proceso estocástico observado y $\{Z_t\}_{t \in \mathcal{N}}$ una Cadena de Markov no homogénea no observada donde las probabilidades de transición son guiadas por un proceso multivariado $\{\mathbf{x}_t\}$ exógeno observado. Adicionalmente, la distribución de probabilidad de Y_t es una de las distribuciones K que dependen del comportamiento de Z_t donde $Z_t = z_t \in \mathfrak{Z} = \{1, 2, \dots, K\}$, es decir:

$$Y_t | Z_t = z_t \sim F(Y_t | \varrho_{z_t}) \quad (3.1)$$

$$P(Z_0 = i) = \delta_i, \quad \sum_{i=1}^K \delta_i = 1 \quad \forall k \in \mathfrak{Z} \quad (3.2)$$

$$P(Z_t = i | Z_{t-1} = j) = \gamma_{ij}^t = \begin{cases} \frac{1}{1 + \sum_{l \neq j}^K e^{x_t \beta_{lj}^t}} & \text{Si } i = j \\ \frac{e^{x_t \beta_{ij}^t}}{1 + \sum_{l \neq j}^K e^{x_t \beta_{lj}^t}} & \text{Si } i \neq j \end{cases} \quad (3.3)$$

Para la inferencia se plantean los siguientes supuestos:

Supuesto S1 Cada variable aleatoria del proceso estocástico $Y_1^T = \{Y_t\}_{t=1, \dots, T}$ es independiente de las demás dada la Cadena de Markov $Z_1^{(K)T} = z_1^{T(K)}$ correspondiente. Veamos:

$$P(Y_1^T = y_1^T | Z_1^T = z_1^T) = \prod_{t=1}^T P(Y_t = y_t | Z_t = z_t) \quad (3.4)$$

Supuesto S2 La distribución de probabilidad $F(Y_t | \cdot)$ pertenece a la familia exponencial. Observe la siguiente ecuación:

$$P(Y_t = y_t | Z_t = i) = s(y_t) t(\varrho_i) e^{a(y_t) b(\varrho_i)} = e^{a(y_t) b(\varrho_i) + c(\varrho_i) + d(y_t)} \quad (3.5)$$

El conjunto de parámetros en un *NHHMM* se puede clasificar en tres grupos: el primero está conformado por el número de estados de la Cadena de Markov ($K = k$), de manera que $k \in [1, M]$ y M es conocido. Dado $K = k$, el segundo conjunto ($\xi_Z^{(k)}$) relaciona los parámetros de la Cadena de Markov y corresponden a las probabilidades de estado inicial $\underline{\delta} = (\delta_1^{(k)}, \dots, \delta_k^{(k)})$ con $i = 1, \dots, k$, tal que $\delta \in (0, 1)^k$ y los parámetros que impactan a la matriz de transición $\underline{\beta}^{(k)}$, con:

$$\underline{\beta}^{(k)} = \begin{pmatrix} & \underline{\beta}_{12}^{(k)} & \dots & \underline{\beta}_{1k}^{(k)} \\ \underline{\beta}_{21}^{(k)} & & \dots & \underline{\beta}_{2k}^{(k)} \\ \dots & \dots & \dots & \dots \\ \underline{\beta}_{k1}^{(k)} & \underline{\beta}_{k2}^{(k)} & \dots & \dots \end{pmatrix} \quad (3.6)$$

Donde $\underline{\beta}_{ij}^{(k)} = (\beta_{ij1}^{(k)}, \dots, \beta_{ijq}^{(k)})$ es el vector de parámetros que interactúa en la influencia de las q variables exógenas a la probabilidad de pasar del estado i al j en un momento dado, para $\xi_Z^{(k)}$, el vector de parámetros $\underline{\beta}_{ij}^{(k)} \in \mathcal{R}^q$. El tercer grupo ($\xi_Y^{(k)}$) son los parámetros asociados al proceso $\{Y_t\}$ con $\xi_Y^{(k)} = (\underline{\varrho}_1^{(k)}, \dots, \underline{\varrho}_k^{(k)})$; este arreglo de vectores de parámetros depende de la distribución de Y_t y por ende su espacio paramétrico depende de esa distribución. Con relación a la notación, el superíndice $^{(k)}$ indica que en ese caso el número de estados es conocido y se define como k .

3.1. Función de verosimilitud

Al observar una muestra $y_1^T = \{y_t\}_{1, \dots, T}$ del proceso estocástico $Y_1^T = \{Y_t\}_{1, \dots, T}$, la función de verosimilitud condicional a la Cadena de Markov con k estados para un *NHHMM* es dada por 3.7, mientras que la función de verosimilitud incondicional es dada por 3.8.

$$\begin{aligned} P(Y_{1:T} = y_{1:T} | K = k, Z_{1:T} = z_{1:T}, \xi_Y^{(k)}, \xi_Z^{(k)}) &= \prod_{t=1}^T P(Y_t = y_t | Z_t = z_t, \xi_Y^{(k)}, \xi_Z^{(k)}) \\ &= \prod_{i=1}^k \prod_{j=1}^{n_i} e^{a(y_j)b(\underline{\varrho}_i^{(k)}) + c(\underline{\varrho}_i^{(k)}) + d(y_j)} \\ &= \prod_{i=1}^k e^{b(\underline{\varrho}_i^{(k)}) \sum_{j=1}^{n_i} a(y_j) + n_i c(\underline{\varrho}_i^{(k)}) + \sum_{j=1}^{n_i} d(y_j)} \end{aligned} \quad (3.7)$$

$$\begin{aligned}
P(Y_{1:T} = y_{1:T} | K = k, \xi_Y^{(k)}, \xi_Z^{(k)}) &= \sum_{z_{1:T} \in \mathfrak{Z}} P(Y_{1:T} = y_{1:T}, z_{1:T} | K = k, \xi_Y^{(k)}, \xi_Z^{(k)}) \\
&= \sum_{z_{1:T} \in \mathfrak{Z}} P(Y_{1:T} = y_{1:T} | z_{1:T}, K = k, \xi_Y^{(k)}, \xi_Z^{(k)}) P(z_{1:T} | K = k, \xi_Y^{(k)}, \xi_Z^{(k)}) \\
&= \sum_{z_{1:T} \in \mathfrak{Z}} \prod_{i=1}^k \prod_{j=1}^{n_i} e^{a(y_j) b(\underline{\varrho}_i^{(k)}) + c(\underline{\varrho}_i^{(k)}) + d(y_j)} P(z_{1:T} | K = k, \xi_Y^{(k)}, \xi_Z^{(k)}) \\
&= \sum_{z_{1:T} \in \mathfrak{Z}} \prod_{i=1}^k \prod_{j=1}^{n_i} e^{a(y_j) b(\underline{\varrho}_i^{(k)}) + c(\underline{\varrho}_i^{(k)}) + d(y_j)} P(Z_0 = z_0) \prod_{t=1}^T P(Z_t = s | Z_{t-1} = q, \xi_Z^{(k)}) \\
&= \sum_{z_{1:T} \in \mathfrak{Z}} \prod_{i=1}^k e^{b(\underline{\varrho}_i) \sum_{j=1}^{n_i} a(y_j) + nc(\underline{\varrho}_i) + \sum_{j=1}^{n_i} d(y_j)} P(Z_0 = z_0) \prod_{q=1}^k \prod_{s=1}^{k-1} \gamma_{qs}^{n_{q,s}}
\end{aligned} \tag{3.8}$$

En el que n_i es el número de veces que está la cadena en el estado i y $n_{q,s}$ el número de veces que pasa la cadena del estado q al s . La productoria en la primera línea de 3.7, así como la productoria de la tercera en 3.8 se cumple gracias a 3.4. Entonces, la inferencia de la Cadena de Markov puede realizarse con el cálculo de la función de Filtro y suavizamiento del proceso tal como se muestra en Frühwirth-Schnatter (2006, pg. 331, eq 11.16 [7]).

3.2. Características

Cuando se conoce la Cadena de Markov oculta el valor esperado de una transformación de la variable de interés es igual a las características mostradas por Dobson y Barnett (2008, eq. 3.9 y eq. 3.12, cap. 3, cap. 4 y cap. 5 [6]), donde se plantea que $E[Y_t | Z_t = z_t] = \int_{\nu} Y_t P(Y_t | Z_t) dY_t = -c'(\underline{\varrho}_{z_t}) / b'(\underline{\varrho}_{z_t})$ y la varianza es dada por $Var[Y_t | Z_t = z_t] = (b''(\underline{\varrho}_{z_t}) c'(\underline{\varrho}_{z_t}) - c''(\underline{\varrho}_{z_t}) b'(\underline{\varrho}_{z_t})) / (b'(\underline{\varrho}_{z_t}))^3$.

Cuando no se conoce la Cadena de Markov, el valor esperado de una transformación del proceso de interés se puede ver como:

$$\begin{aligned}
E[h(Y_t)] &= \int_{\nu} h(Y_t) P(Y_t) dY_t = \int_{\nu} h(Y_t) \sum_{k_t=1}^K P(Y_t | Z_t = k_t) P(Z_t = k_t) dY_t \\
&= \int_{\nu} \sum_{k_t=1}^K h(Y_t) P(Y_t | Z_t = k_t) P(Z_t = k_t) dY_t = \sum_{k_t=1}^K P(Z_t = k_t) \int_{\nu} h(Y_t) P(Y_t | Z_t = k_t) dY_t \\
&= \sum_{k_t=1}^K P(Z_t = k_t) E(h(Y_t) | Z_t = k_t)
\end{aligned} \tag{3.9}$$

Adicionalmente, se puede observar que:

$$\begin{aligned}
E(h(Y_t, Y_{t+s})) &= \int_{\nu_1} \int_{\nu_2} h(Y_t, Y_{t+s}) P(Y_t, Y_{t+s}) dY_t d(Y_{t+s}) \\
&= \int_{\nu_1} \int_{\nu_2} \sum_{i=1}^k \sum_{j=1}^k h(Y_t, Y_{t+s}) P(Y_t, Y_{t+s}, Z_t = i, Z_{t+s} = j) dY_t dY_{t+s} \\
&= \int_{\nu_1} \int_{\nu_2} \sum_{i=1}^k \sum_{j=1}^k h(Y_t, Y_{t+s}) P(Y_t, Y_{t+s} | Z_t = i, Z_{t+s} = j) P(Z_t = i, Z_{t+s} = j) dY_t dY_{t+s} \\
&= \sum_{i=1}^k \sum_{j=1}^k P(Z_t = i, Z_{t+s} = j) \int_{\nu_1} \int_{\nu_2} h(Y_t, Y_{t+s}) P(Y_t, Y_{t+s} | Z_t = i, Z_{t+s} = j) dY_t dY_{t+s} \\
&= \sum_{i=1}^k \sum_{j=1}^k P(Z_t = i, Z_{t+s} = j) E(h(Y_t, Y_{t+s}) | Z_t = i, Z_{t+s} = j).
\end{aligned} \tag{3.10}$$

Con la ecuación representada en la figura 3.11, la interacción entre dos momentos depende de la lejanía en la que está el uno del otro, la cual va a depender de la matriz de transición. A partir de la ecuación 3.3 podemos obtener:

- El valor esperado:

$$\begin{aligned}
E[Y_t] &= \sum_{k_t=1}^K P(Z_t = k_t) E(h(Y_t) | Z_t = k_t) = \sum_{k_t=1}^K P(Z_t = k_t) - c'(\underline{\varrho}_{k_t}) / b'(\underline{\varrho}_{k_t}) \\
&\quad \sum_{k_t=1}^K \sum_{k_{t-1}=1}^K -P(Z_t = k_t | Z_{t-1} = k_{t-1}) P(Z_{t-1} = k_{t-1}) c'(\underline{\varrho}_{k_t}) / b'(\underline{\varrho}_{k_t}) \tag{3.11} \\
&\quad \sum_{k_t=1}^K \sum_{k_{t-1}=1}^K -P(Z_{t-1} = k_{t-1}) \frac{e^{\underline{x}_t \beta_{t_{k-1} t_k} c'(\underline{\varrho}_{k_t})}}{(1 + \sum_{j=1}^K e^{\underline{x}_t \beta_{j t_k} b'(\underline{\varrho}_{k_t})})}
\end{aligned}$$

- La varianza:

$$\begin{aligned}
Var[Y_t] &= E[Y_t^2] - (E[Y_t])^2 = \sum_{k=1}^K P(Z_t = k)E[Y_t^2|Z_t = k] - \left(\sum_{k=1}^K P(Z_t = k)E[Y_t|Z_t = k]\right)^2 \\
&= \sum_{k=1}^K P(Z_t = k)(Var(Y_t|Z_t = k) + (E(Y_t|Z_t = k))^2) - \sum_{k=1}^K P(Z_t = k)^2(E[Y_t|Z_t = k])^2 \\
&\quad - 2 \sum_{k_1=1}^K \sum_{k_2=1}^K P(Z_t = k_1)P(Z_t = k_2)E[Y_t|Z_t = k_1]E[Y_t|Z_t = k_2] \\
&= \sum_{k=1}^K P(Z_t = k)Var(Y_t|Z_t = k) + P(Z_t = k)(1 - P(Z_t = k))(E[Y_t|Z_t = k])^2 \\
&\quad - 2 \sum_{k_1=1}^K \sum_{k_2=1}^K P(Z_t = k_1)P(Z_t = k_2)E[Y_t|Z_t = k_1]E[Y_t|Z_t = k_2]
\end{aligned} \tag{3.12}$$

La exploración, el análisis y el tratamiento de la función de verosimilitud son de gran importancia, principalmente en la inferencia clásica. Un estimador de los parámetros se encuentra con los valores que maximizan los resultados de la función de verosimilitud, para esto es necesario alguna transformación funcional que permita el uso de métodos computacionales, tal como se muestra en Zucchini y MacDonald (2009 , cap. 3, pg. 46 y 47, eq 3.1 [14]) , solucionando problemas como la tendencia a cero en momentos iniciales del proceso, los valores en la estimación que no pertenezcan al dominio de los valores del parámetro y encontrar los valores de maximización que no sean, por ser valores en la función, muy cercanos a cero. Adicionalmente, se usan algoritmos como el de *Viterbi* para la inferencia de la Cadena.

Cabe resaltar que el número de estados de la Cadena es desconocido, pero se conoce el dominio de valores que puede tomar, en el que cada posible valor establece un potencial modelo y por ende un conjunto de parámetros $\theta^{(i)}$ con $i = 1, \dots, M$. Es importante resaltar, ya para finalizar que en el presente capítulo llamaremos $\underline{\Theta} = (\underline{\theta}^{(1)}, \underline{\theta}^{(2)}, \dots, \underline{\theta}^{(M)})$ al arreglo de todos los parámetros para el conjunto de los posibles modelos.

CAPÍTULO 4

Inferencia

El presente capítulo da a conocer un desarrollo investigativo que se elabora en el marco de la maestría en estadística y se fundamenta en la inferencia del número de estados de la Cadena de Markov. La base de la metodología propuesta en este documento se obtiene de los planteamientos mostrados en Carlin y Chib (1995 [3]), Carlin y Chib(1996 [2]) y Spezia(2006 [11]).

Este capítulo se divide en tres partes: la primera presenta la distribución *a posteriori* de los parámetros del modelo, la segunda muestra la estructura de las distribuciones *a priori* que se usan en la investigación y la última parte expone el desarrollo de las diversas distribuciones *a posteriori* condicionales para la implementación del algoritmo de Gibbs en la simulación, siguiendo la metodología mostrada en Carlin y Chib (1995 [3]).

4.1. Inferencia Bayesiana de los NHHMM

La inferencia Bayesiana se fundamenta en el análisis de la distribución *a posteriori* de los parámetros del modelo; tal como se mencionó en el capítulo anterior, el conjunto de parámetros de un *NHHMM* está dado por el número de estados de la cadena $K = l$, el conjunto de parámetros $\underline{\theta}^{(l)} = (Z_1^{T(l)}, \underline{\xi}_Z^{(l)})$ y el conjunto $\underline{\xi}_Y^{(l)}$. Dada una muestra y_1^T del proceso estocástico Y_1^T , la distribución *a posteriori* conjunta, con $l = 1, \dots, K$, es dada por:

$$\begin{aligned} \pi(K = l, \underline{\theta}^{(l)}) &= P(K = l, \underline{\theta}^{(l)} | Y_1^T = y_1^T) = P(K = l, Z_1^{T(l)}, \underline{\xi}_Z^{(l)}, \underline{\xi}_Y^{(l)} | Y_1^T = y_1^T) \\ &\propto f(Y_1^T = y_1^T | Z_1^{T(l)}, \underline{\xi}_Z^{(l)}, \underline{\xi}_Y^{(l)}, K = l) P(Z_1^{T(l)} | \underline{\xi}_Z^{(l)}, \underline{\xi}_Y^{(l)}, K = l) P(\underline{\xi}_Z^{(l)} \underline{\xi}_Y^{(l)} | K = l) P(K = l) \end{aligned} \quad (4.1)$$

De tal manera que la probabilidad de la Cadena de Markov está dada por:

$$P(Z_1^{T(l)} | \xi_Z^{(l)}, \xi_Y^{(l)}, K = l) = P(Z_0^{(l)} = z_0^{(l)}) \prod_{i=1}^T P(Z_t^{(l)} = z_t^{(l)} | Z_{t-1}^{(l)} = z_{t-1}^{(l)}, \xi_Z^{(l)}, K = l) \quad (4.2)$$

La expresión $P(Y_{1:T} = y_{1:T} | Z_{1:T}^{(l)}, \xi_Z^{(l)}, \xi_Y^{(l)}, K = l)$ corresponde a la función de verosimilitud condicional a la Cadena de Markov, que se puede observar en la ecuación 3.7 3.7. Las últimas componentes de la distribución *a posteriori*, $P(\xi_Z^{(l)}, \xi_Y^{(l)} | K = l)$ y $P(K = l)$, corresponden a las distribuciones *a priori* de los parámetros condicionado al número de estados de la cadena y la distribución *a priori* del número de estado.

4.2. Distribuciones a prior

Tal como se muestra en la ecuación 4.1, la distribución conjunta *a priori* de los parámetros se puede ver como la multiplicación de la distribución conjunta de la Cadena de Markov condicionada a los demás parámetros, por la distribución conjunta de los parámetros de la cadena y los parámetros del proceso $\{Y_t\}$ condicionado al número de estados, por la distribución *A prior* de K . Para la investigación realizada se suponen las siguientes condiciones sobre los parámetros $(Z_{1:T}^{(k)}, \xi_Z^{(k)}, \xi_Y^{(k)}, K = k)$.

Supuesto P1 El vector conjunto de parámetros $\xi_Z^{(k)}$ es independiente del vector conjunto de parámetros $\xi_Y^{(k)}$, así como la independencia de cada elemento de $\xi_Z^{(k)}$ con cada elemento de $\xi_Y^{(k)}$ de manera pareada dado el número de estados de la Cadena Markoviana $K = k$; de esta forma:

$$P(\xi_Z^{(k)}, \xi_Y^{(k)} | K = l) = P(\xi_Z^{(k)} | K = l) P(\xi_Y^{(k)} | K = l) \quad (4.3)$$

Supuesto P2 Cada parámetro $\underline{\varrho}_i^{(k)}$ es independiente de $\underline{\varrho}_j^{(k)} \forall i \neq j$ y $i = 1, \dots, k$ y $j = 1, \dots, k$ dado el número de estados $K = k$.

Supuesto P3 Cada elemento de la matriz de parámetros en $\underline{\beta}^{(k)}$ es independiente del vector probabilidades de estado inicial de la Cadena dado el número de estados $K = k$.

Supuesto P4 Dado $K = k$, cada vector conjunto de parámetros $\underline{\beta}_i^{(k)}$ es independiente al vector conjunto de parámetros $\underline{\beta}_j^{(k)}$ que son vectores correspondientes a fila de la matriz de parámetros $\underline{\beta}^{(k)}$, donde $\underline{\beta}_h^{(k)} = (\underline{\beta}_{h1}^{(k)}, \dots, \underline{\beta}_{hk}^{(k)})$

A partir del supuesto P2, la distribución *a priori* conjunta de los parámetros del proceso observado $\xi_Y^{(k)} = (\underline{\varrho}_1^{(k)}, \dots, \underline{\varrho}_k^{(k)})$ se muestra en 4.4. Adicionalmente se hace uso de distribuciones conjugadas para la familia exponencial tal como se presenta en Gelman, Carlin, Stern, Dunson, Vehteri y Rubin (2009, cap. 2, pg 37 [9]).

$$P(\xi_Y^{(k)} | K = k) = \prod_{i=1}^K P(\underline{\varrho}_i^{(k)} | K = k) = \prod_{i=1}^K t(\underline{\varrho}_i^{(k)})^\eta e^{b(\underline{\varrho}_i^{(k)})^\nu} \quad (4.4)$$

Donde las funciones $t(\cdot)$ y $b(\cdot)$ corresponden a las mencionadas en el capítulo (2.2), con η y ν hiperparámetros conocidos. En el caso de tener un *Poisson-NHHMM* cada componente $\underline{\varrho}_i^{(k)} = \lambda_i^{(k)}$, con funciones conjugadas dadas por $\lambda_i^{(l)}|K = l \sim \text{Gamma}(\alpha_i, \nu_i)$. En Spezia (2006, pg. 715 [11]) se proponen distribuciones normales truncadas para cada parámetro μ_i en un *Gaussian-NHHMM* con el fin de controlar la no identificabilidad en el modelo a partir de suponer $\mu_i < \mu_j$ cuando $i < j$, para controlar la no identificabilidad en un *Poisson - NHHMM* se puede suponer que $\alpha_i^{(k)} < \alpha_j^{(k)}$ con $\nu_i = \nu_j$ cuando $i < j$.

Para la Cadena de Markov, el conjunto de parámetros $\xi_Z^{(k)}$ contempla la matriz $\beta^{(k)}$ y el vector de probabilidades de estado inicial $\underline{\delta}^{(k)}$. Con el supuesto P4 tenemos que $P(\xi_Z^{(k)}|K = k) = P(\beta^{(k)}, \underline{\delta}^{(k)}|K = k) = P(\beta^{(k)}|K = k)P(\underline{\delta}^{(k)}|K = k) = \prod_{i=1}^k P(\underline{\beta}_i^{(k)}|K = k)P(\underline{\delta}^{(k)}|K = k)$. En este caso se propone $\underline{\delta}^{(k)}|K = k \sim \text{Dirichlet}(a_1, \dots, a_{k-1})$ donde a_1, \dots, a_{k-1} son hiperparámetros conocidos y para $\underline{\beta}^{(l)} = \{\underline{\beta}_{ij}^{(l)}\}_{i \neq j}$ se plantea que $\underline{\beta}_{ij}^{(l)}|K = l \sim N(\mu_{\beta_{ij}}, \sigma_{\beta_{ij}})$. De esta manera tenemos que:

$$\begin{aligned} P(\xi_Z^{(k)}|K = k) &= \prod_{i=1}^k \prod_{i \neq j}^{k-1} P(\underline{\beta}_{ij}^{(k)}|K = k)P(\underline{\delta}^{(k)}|K = k) \\ &= \prod_{i=1}^k \prod_{i \neq j}^{k-1} \Phi(\underline{\beta}_{ij}^{(k)}|\mu_{\beta_{ij}}, \sigma_{\beta_{ij}}) \left(\frac{1}{B(\underline{a})} \prod_{i=1}^k \delta_i^{a_i-1} \right) \\ &\propto \prod_{i=1}^k \prod_{i \neq j}^{k-1} e^{-1/2(\underline{\beta}_{ij}^{(k)} - \mu_{\beta_{ij}})\sigma_{\beta_{ij}}^{-1}(\underline{\beta}_{ij}^{(k)'} - \mu_{\beta_{ij}})} \prod_{i=1}^{k-1} \delta_i^{a_i-1} \end{aligned} \quad (4.5)$$

Por último, se formula una distribución uniforme discreta para el número de estados $K = k$, aunque se puede asumir que sigue una distribución *a priori* binomial con hiperparámetros conocidos, es decir $P(K) \propto \pi^K(1 - \pi)^{n-K}$. En las ecuaciones 4.4 y 4.5 la distribución conjunta *A priori* es dada por:

$$\begin{aligned} P(\xi_Z^{(k)}, \xi_Y^{(k)}, K = k) &= \prod_{i=1}^K t(\underline{\varrho}_i^{(k)})\eta e^{b(\underline{\varrho}_i^{(k)})\nu} \\ &\prod_{i=1}^k \prod_{i \neq j}^{k-1} e^{-1/2(\underline{\beta}_{ij}^{(k)} - \mu_{\beta_{ij}})\sigma_{\beta_{ij}}^{-1}(\underline{\beta}_{ij}^{(k)'} - \mu_{\beta_{ij}})} \prod_{i=1}^{k-1} \delta_i^{a_i-1} (1/K) \end{aligned} \quad (4.6)$$

Así la distribución *a posteriori* de los *NHHMM* mostrada en 4.1 es dada por 4.7 a partir de 4.6 y 4.2.

$$\begin{aligned}
\pi(K = k, \underline{\theta}^{(k)}) &\propto P(Y_{1:T} = y_{1:T} | Z_{1:T}^{(k)}, \xi_Z^{(k)}, \xi_Y^{(k)}, K = k) P(Z_{1:T}^{(k)} | \xi_Z^{(k)}, \xi_Y^{(k)}, K = k) \\
&P(\xi_Z^{(k)}, \xi_Y^{(k)} | K = k) P(K = k) \\
&\propto \left(\prod_{i=1}^k t(\underline{\rho}_i)^{n_i} e^{b(\underline{\rho}_i) \sum_{j=1}^{n_i} a(y_j)} \right) (P(Z_0 = z_0^{(k)}) \prod_{i=1}^T P(Z_t^{(k)} = z_t^{(k)} | Z_{t-1}^{(k)} = z_{t-1}^{(k)}, \xi_Z^{(k)}, K = k)) \\
&\left(\prod_{i=1}^k t(\underline{\rho}_i^{(k)}) \eta e^{b(\underline{\rho}_i^{(k)}) \nu} \right) \left(\prod_{i=1}^k \prod_{i \neq j}^{k-1} e^{-1/2(\beta_{ij}^{(k)} - \mu_{\beta_{ij}}) \sigma_{\beta_{ij}}^{-1} (\beta_{ij}^{(k)} - \mu_{\beta_{ij}})'} \prod_{i=1}^{k-1} \delta_i^{a_i - 1} \right) (1/K) \\
&\propto \left(\prod_{i=1}^k t(\underline{\rho}_i)^{n_i + \eta} e^{b(\underline{\rho}_i) (\nu + \sum_{j=1}^{n_i} a(y_j))} \right) \left(\prod_{i=1}^k \prod_{i \neq j}^{k-1} \frac{e^{-1/2(\beta_{ij}^{(k)} - \mu_{\beta_{ij}}) \sigma_{\beta_{ij}}^{-1} (\beta_{ij}^{(k)} - \mu_{\beta_{ij}})'} + \sum_{t=1}^{n_{ij}} x_t' \beta_{ij}}{\prod_{t=1}^{n_{ij}} (1 + e^{x_t' \beta_{ij}})} \left(\frac{\sum_{t=1}^{n_{ii}} x_t' \beta_{ii}}{\prod_{t=1}^{n_{ij}} (1 + e^{x_t' \beta_{ij}})} \right) \right) \\
&(P(Z_0 = z_0^{(l)}) \prod_{i=1}^{k-1} \delta_i^{a_i - 1}) P(K = k)
\end{aligned} \tag{4.7}$$

La ecuación 4.7 muestra la forma funcional de la distribución conjunta *a posteriori* en estos modelos, el cual es una productoria de las correspondientes funciones de parámetros. En esta ecuación n_i , n_{ii} y n_{ij} se representan el número de veces que se encuentra la cadena en el estado i , el número de veces que la cadena está en el estado i y se queda en ese estado, y el número de veces que la cadena pasa del estado i al j , respectivamente.

4.3. Estimación y simulación de los parámetros

Una manera general para resumir las características y estadísticos de la distribución posterior es mediante la media de una transformación de los parámetros bajo esta distribución, la cual es dada por 4.8, para más especificación al respecto es relevante revisar Albert (2011, cap 5.7, pg 97 [1]).

$$E(h(K, \underline{\theta}^{(K)}) | y_1^T) = \int h(K, \underline{\theta}^{(K)}) \pi(K, \underline{\theta}^{(K)}) \tag{4.8}$$

Ya que el desarrollo analítico de la integral es bastante complejo, usamos el método de Montecarlo Markov Chain (MCMC) para obtener los cálculos correspondientes. Esto requiere la simulación de valores de cada parámetro de la distribución *a posteriori*, el cual se realiza con el algoritmo de *Gibbs* siguiendo el desarrollo mostrado en Carlin y Chib (1995[3]), ya que se busca la inferencia de un modelo y el número de parámetros es bastante amplio.

El algoritmo simula valores de $(K, \underline{\Theta})$ a partir de la simulación de valores de cada distribución *a posteriori* condicionada a los demás. Para la investigación se toma $\underline{\Theta} = (\underline{\theta}^{(1)}, \underline{\theta}^{(2)}, \dots, \underline{\theta}^{(M)})$, donde cada componente $\underline{\theta}^{(i)} = (Z_{1:T}^{(i)}, \xi_Z^{(i)}, \xi_Y^{(i)})$, tal que $Z_{1:T}^{(i)}$ es la cadena de Markov con (i) estados, $\xi_Z^{(i)} = (\delta^{(i)}, \beta^{(i)})$ son los parámetros de estado inicial y la matriz de parámetros de las co-variables que influyen la matriz de transición, y

$\xi_Y^{(i)} = (\underline{\varrho}_1, \dots, \underline{\varrho}_i)$ son los parámetros del proceso observado dado que tenemos i estados. Algunas distribuciones *a posteriori* son proporcionales a formas funcionales conocidas y la simulación se realiza de manera directa en el software R, mientras que en otros casos la función es desconocida y difícil de simular y por esto se usa el algoritmo de *Metropolis-Hastings* dentro del *Gibbs* construyendo Cadenas de Markov irreducibles y aperiódicas de valores en estos parámetros.

4.3.1. Distribuciones *a posteriori* de cada parámetro condicional a los demás parámetros

El presente capítulo detalla las distribuciones *a posteriori* de cada parámetro condicionando a los demás para el desarrollo del algoritmo de Gibbs. La implementación en el modelo se fundamenta en los planteamientos mostrados en Carlin y Chib (1995 [3]) ; inicialmente se plantea un número conocido máximo de estados M , asociados a un conjunto de M posibles modelos. Las distribuciones *a posteriori* que se despliegan a continuación toman como insumo los valores conocidos en la $(r - 1)$ -ésima iteración de los parámetros:

$$(K_{(r-1)}, (Z_{1:T(r-1)}^{(1)}, \xi_{Z(r-1)}^{(1)}, \xi_{Y(r-1)}^{(1)}), \dots, (Z_{1:T(r-1)}^{T(M)}, \xi_{Z(r-1)}^{(M)}, \xi_{Y(r-1)}^{(M)}))$$

Conociendo estos valores simulamos los valores de cada parámetro en la (r) -ésima iteración a partir de las siguientes distribuciones:

a) Distribución condicional completa de K para la simulación del valor de $K_{(r)}$, así:

$$\begin{aligned} P(K = l | \underline{\Theta}_{(r-1)}, y_1^T) &= \frac{P(K = l, \underline{\Theta}_{(r-1)}, y_1^T)}{P(\underline{\Theta}_{(r-1)}, y_1^T)} \\ &= \frac{P(y_1^T | K = l, \underline{\Theta}_{(r-1)}) P(\underline{\Theta}_{(r-1)} | K = l) P(K = l)}{\sum_{h=1}^K P(\underline{\Theta}_{(r-1)}, y_1^T, K = h)} \quad (4.9) \\ &= \frac{P(y_1^T | \underline{\theta}_{(r-1)}^l, K = l) (\prod_{j=1}^M P(\underline{\theta}_{(r-1)}^{(j)} | K = l)) P(K = l)}{\sum_{h=2}^K P(y_1^T | \underline{\theta}_{(r-1)}^{(h)}) (\prod_{j=1}^M P(\underline{\theta}_{(r-1)}^{(j)} | K = h)) P(K = h)} \end{aligned}$$

Tal como se puede ver en la ecuación anterior, la distribución condicional completa de K es de tipo discreto lo que permite el cálculo de cada uno de las probabilidades con $l = 1, \dots, M$, donde el resultado del numerador es un componente de la sumatoria en el denominador; esta distribución se expone en Carlin y Chib (1996, cap 2.1, eq. 2 [3]). Cada componente está conformado por tres expresiones: la primera corresponde a la función de verosimilitud mostrada en 3.7, la segunda es una productoria que se obtiene con el supuesto de independencia conjunta entre $\underline{\theta}^{(i)}$ y $\underline{\theta}^{(j)}$ con $i, j = 1, \dots, k$ y $i \neq j$ condicionado a que $K = k$, adicionalmente cada variable aleatoria del proceso Y_1^T es independiente de $\underline{\theta}^{(j)}$ cuando θ_j es representativo para y_1^T , la tercera y última expresión en la ecuación corresponde a la distribución *a priori* del orden o número de estados.

Cabe anotar que la productoria en la segunda expresión contempla las distribuciones de $\underline{\theta}_{(i)}$ con $i \neq k = K$; en este caso se usan distribuciones **pseudo a priori** que ayudan a conectar el proceso iterativo de simulación de todos los parámetros condicionado al valor conocido del número de estados, para esta distribución se recomiendan formas conocidas que faciliten el cálculo.

Una vez generamos el valor $K_{(r)} = k$ en la r -ésima iteración, la simulación de los demás parámetros $\underline{\Theta}_r = (\underline{\theta}_{(r)}^{(1)}, \underline{\theta}_{(r)}^{(2)}, \dots, \underline{\theta}_{(r)}^{(M)})$ se realiza por medio de la distribución 4.10 como se aborda en Carlin y Chib (1996, cap. 2, eq 1 [3]). Esta distribución es por partes, la primera corresponde a la distribución de $\underline{\theta}_{(r)}^{(k)}$ cuando $K_{(r)} = k$, la segunda cuando se presenta $K_{(r)} \neq k$. Inicialmente obtenemos la distribución condicional completa cuando el espacio paramétrico de los valores sea igual al número de estados de la cadena en esa iteración.

$$P(\underline{\Theta}|K_{(r)} = l, y_1^T) \propto \begin{cases} p(Y_1^T = y_1^T | \underline{\theta}^{(l)}, K_{(r)} = l) P(\underline{\theta}^{(l)} | K_{(r)} = l) & \text{Si } l = i \\ P(\underline{\theta}^{(i)} | K_{(r)} = l, y_1^T) & \text{Si } l \neq i \end{cases} \quad (4.10)$$

- Distribución condicional de $\underline{\theta}^{(k)} = (Z_1^{T(k)}, \xi_Z^{(k)}, \xi_Y^{(k)})$ para la simulación de $\underline{\theta}_{(r)}^{(k)} = (Z_{1(r)}^{T(k)}, \xi_{Z(r)}^{(k)}, \xi_{Y(r)}^{(k)})$ dado el valor de $K_{(r)} = k$. Adicionalmente $\underline{\theta}_{(r-1)}^{(s)}$ si $s > k$ y $\underline{\theta}_{(r)}^{(s)}$ si $s < k$.

- b) Dado $K_{(r)} = k$, la distribución condicional completa de $\xi_Y^{(k)} = (\underline{\varrho}_1^{(k)}, \underline{\varrho}_2^{(k)}, \dots, \underline{\varrho}_k^{(k)})$, para simular los valores de $\xi_{Y(r)}^{(k)} = (\underline{\varrho}_{1(r)}^{(k)}, \underline{\varrho}_{2(r)}^{(k)}, \dots, \underline{\varrho}_{k(r)}^{(k)})$, se puede ver de manera independiente para cada componente $\underline{\varrho}_s^{(k)}$ dados los supuestos anteriores. En este caso la distribución condicional completa de cada $\underline{\varrho}_h^{(k)}$ que depende de la distribución de probabilidad que se muestrea en Y_t con $h = 1, \dots, k$ también se presenta en 4.11, para que esto sea $l \neq k$.

$$\begin{aligned} P(\underline{\varrho}_h^{(k)} | K_{(r)} = k, \underline{\theta}_{(r)}^{(1)}, \dots, \underline{\theta}_{(r)}^{(k-1)}, \underline{\varrho}_{1(r)}, \dots, \underline{\varrho}_{h-1(r)}, \underline{\varrho}_{h+1(r-1)}, \dots, \underline{\varrho}_{k(r)}, \xi_{Z(r-1)}^{(k)}, Z_{1(r-1)}^{T(k)}, y_1^T) \\ \propto P(\underline{\varrho}_h^{(k)}, K_{(r)} = k, \underline{\theta}_{(r)}^{(1)}, \dots, \xi_{Z(r-1)}^{(k)}, Z_{1(r-1)}^{T(k)} | y_1^T) \\ \propto g(\underline{\varrho}_h^{(k)})^{n_h} e^{\phi(\underline{\varrho}_h^{(k)}) \sum_{j=1}^{n_h} t(y_j)} g(\underline{\varrho}_h^{(k)})^\eta e^{b(\underline{\varrho}_h^{(k)}) \nu} \\ \propto g(\underline{\varrho}_h^{(k)})^{n_h + \eta} e^{b(\underline{\varrho}_h^{(k)}) (\nu + \sum_{j=1}^{n_h} t(y_j))} \end{aligned} \quad (4.11)$$

En este caso tenemos que n_h es el número de veces que se encuentra la Cadena de Markov en el estado h , conociendo la Cadena de Markov con k número de estados en la iteración $(r - 1)$, el resultado de la tercera fila en 4.11 es dado por el planteamiento obtenido en la distribución *a posteriori* mostrada en 4.7. Al tomar en cuenta una distribución *a priori* conjugada para estos parámetros se obtiene una forma conocida ya que $y_i \sim F$ tal como se muestra en la cuarta línea de 4.11.

- c) Dado $K_{(r)} = k$ y los valores de $\xi_{Y_{(r)}}^{(k)}$, la distribución *a posteriori* condicional de $\xi_{Z_{(r)}}^{(k)} = (\underline{\beta}_{(r)}^{(k)}, \underline{\delta}_{(r)}^{(k)})$ para simular los valores de $\xi_{Z_{(r)}}^{(k)} = (\underline{\beta}_{(r)}^{(k)}, \underline{\delta}_{(r)}^{(k)})$ se brinda de manera independiente. La distribución condicional completa de $\underline{\beta}^{(k)}$ está conformada por la productoria de cada uno de sus componentes $\underline{\beta}_{ij}^{(k)} \in \underline{\beta}_{(r)}^{(k)}$ teniendo en cuenta la independencia de los parámetros de la fila en la matriz. Para esto, sea $\underline{\beta}_{.i}^{(k)} = (\underline{\beta}_{i1}^{(k)}, \dots, \underline{\beta}_{ik}^{(k)})$, con $k \neq i$, el vector de parámetros que reúne los componentes de la fila i de la matriz $\underline{\beta}^{(k)}$ de tal forma que la distribución condicional completa de $\underline{\beta}_{ij}^{(k)}$ es dada por:

$$\begin{aligned}
P(\underline{\beta}_{ij}^{(k)} | K_{(r)} = k, \underline{\theta}_{(r)}^{(1)}, \dots, \underline{\theta}_{(r)}^{(k-1)}, \xi_{Y_{(r)}}^{(k)}, \underline{\beta}_{1.(r)}^{(k)}, \dots, \underline{\beta}_{i-1.(r)}^{(k)}, \underline{\beta}_{i1(r)}^{(k)}, \dots, \underline{\beta}_{ij-1(r)}^{(k)}, \underline{\beta}_{ij+1(r-1)}^{(k)}, \dots, \\
\underline{\beta}_{ik(r-1)}^{(k)}, \underline{\beta}_{i+1.(r-1)}^{(k)}, \dots, \underline{\beta}_{k.(r-1)}^{(k)}, \underline{\delta}_{(r-1)}^{(k)}, Z_{2(r-1)}^{T(k)}, y_1^T) \\
\propto P(Z_{2(r-1)}^{T(k)} | \xi_Z^{(k)}, K = k, y_1^T, \dots) P(\underline{\beta}_{ij}^{(k)} | K = k) \\
\propto \prod_{t=2}^{n_i} P(Z_t^{(k)} | Z_{t-1}^{(k)} = i, K = k, \underline{\beta}_{.i}^{(k)}) P(\underline{\beta}_{ij}^{(k)} | K = k) \\
= \left(\prod_{g=1, g \neq j}^k \prod_{t=2}^{n_{i,g}} P(Z_t^k = g | Z_{t-1}^k = i, \underline{\beta}_{ij}^{(k)}, \dots) \right) \left(\prod_{t=2}^{n_{i,j}} P(Z_t = j | Z_i = i, \underline{\beta}_{ij}^{(k)}, \dots) \right) (P(\underline{\beta}_{ij}^{(k)}, K = k)) \\
= \left(\prod_{g=1, g \neq j}^k \prod_{t=2}^{n_{i,g}} \left(\frac{e^{x_t' \underline{\beta}_{ig}^{(k)}}}{1 + e^{x_t' \underline{\beta}_{ij}^{(k)}} + \sum_{g \neq j} e^{x_t' \underline{\beta}_{ig}^{(k)}}} \right) \right) \left(\prod_{t=2}^{n_{i,j}} \left(\frac{e^{x_t' \underline{\beta}_{ij}^{(k)}}}{1 + e^{x_t' \underline{\beta}_{ij}^{(k)}} + \sum_{g \neq j} e^{x_t' \underline{\beta}_{ig}^{(k)}}} \right) \right) \\
e^{(\underline{\beta}_{ij}^{(k)} - \mu_{\beta_{ij}}) \sigma_{\beta_{ij}} (\underline{\beta}_{ij}^{(k)} - \mu_{\beta_{ij}})' }
\end{aligned} \tag{4.12}$$

La distribución anterior toma en cuenta los parámetros de $\underline{\beta}_{ij}^{(k)}$ cuando $i \neq j$ y se puede ver que la distribución *a posteriori* condicional para estos parámetros no tiene forma conocida; por lo tanto usamos el algoritmo de *Metropolis* dentro del *Gibbs* que se está aplicando. Para esto se genera un valor $\beta_{ij}^{(k)\star}$ de una distribución propuesta $u(\beta_{ij}^{(k)\star} | \beta_{ij(r-1)}^{(k)})$ a partir de la distribución condicional completa que se está trabajando, con este valor la probabilidad de que $\beta_{ij}^{(k)} = \beta_{ij}^{(k)\star}$ es definida por:

$$\min \left\{ 1, \frac{P(\beta_{ij}^{(k)\star} | \dots) u(\beta_{ij(r-1)}^{(k)} | \beta_{ij}^{(k)\star})}{P(\beta_{ij(r-1)}^{(k)} | \dots) u(\beta_{ij}^{(k)\star} | \beta_{ij(r-1)}^{(k)})} \right\} \tag{4.13}$$

Con esta probabilidad se determina si $\beta_{ij}^{(k)} = \beta_{ij}^{(k)\star}$ o $\beta_{ij}^{(k)} = \beta_{ij(r-1)}^{(k)}$. Es importante agregar que la simulación de $\delta_{(r)}^{(k)}$ se realiza mediante la distribución a la que son proporcional los siguientes componentes $P(Z_0 = z_0^{(k)}) \prod_{i=1}^{k-1} \delta_i^{a_i-1}$.

- d) Obteniendo los valores de $K_{(r)} = k$, $\xi_{Y_{(r)}}^{(k)}$ y $\xi_{Z_{(r)}}^{(k)}$ se genera la Cadena de Markov $z_{1(r)}^{T(k)}$; en este caso usamos el algoritmo *Forward-Filtering-Backward-Smoothing*, el cual se puede observar en Frühwirth-Schnatter (2006, cap. 11.2 pg. 325 [7]) para los *HMM* y en Spezia (2006, cap. 3, pg. 716-717 [11]) para los *NHHMM*. Para este fin, la distribución *a posteriori* de la cadena se define por medio de la siguiente factorización:

$$\begin{aligned}
P(Z_1^{T(k)} = z_1^{T(k)} | K_{(r)} = k, \underline{\theta}_{(r)}^{(1)}, \dots, \underline{\theta}_{(r)}^{(k)}, \underline{\theta}_{(r-1)}^{(k+1)}, \dots, \underline{\theta}_{(r-1)}^{(M)}) \\
= \prod_{t=1}^{T-1} P(Z_t^{(k)} = z_t^{(k)} | Z_{t+1}^T = z_{t+1}^{T(k)}, y_1^T, K_{(r)} = k, K = k, \underline{\theta}_{(r)}^{(k)}) \\
P(Z_T^{(k)} = z_T^{(k)}, y_1^T, K = k, \underline{\theta}_{(r)}^{(k)})
\end{aligned} \tag{4.14}$$

La ecuación 4.14 presenta la distribución *a posteriori* como una productoria de la distribución para cada uno de los momentos de la Cadena de Markov, que depende de los siguientes momentos hasta el último $P(Z_T^{(k)} = z_T^{(k)}, y_1^T, K = k, \underline{\theta}_{(r)}^{(k)})$, siendo el último componente de la productoria la función de probabilidad de *Filtro* para ese momento. Es así como la generación de la cadena plantea el cálculo iterativo de todas las probabilidad de filtro, y una vez se obtiene la última probabilidad empieza la simulación en *Retrosceso*, para lo cual la distribución de probabilidad condicional de la que se generan los demás momentos, que corresponden a los demás componentes de la distribución condicional completa de la Cadena de Markov, está dada por:

$$\begin{aligned}
P(Z_t^{(k)} = z_t^{(k)} | Z_{t+1}^{T(k)} = z_{t+1}^{T(k)}, y_1^T, K^{(r)} = k, \underline{\theta}_{(r)}^{(k)}) \\
\propto P(Z_{t+1}^{T(k)} = z_{t+1}^{T(k)}, y_{t+1}^T | Z_t^{(k)} = z_t^{(k)}, K^{(r)} = k, \underline{\theta}_{(r)}^{(k)}, y_1^t) P(Z_t^{(k)} = z_t^{(k)} | K^{(r)} = k, \underline{\theta}_{(r)}^{(k)}, y_1^t) \\
\propto P(Z_{t+2}^{T(k)} = z_{t+2}^{T(k)}, y_{t+1}^T | Z_{t+1}^{(k)} = z_{t+1}^{(k)}, Z_t^{(k)} = z_t^{(k)}, K^{(r)} = k, \underline{\theta}_{(r)}^{(k)}, y_1^t) \\
P(Z_{t+1}^{(k)} = z_{t+1}^{(k)} | Z_t^{(k)} = z_t^{(k)}, K^{(r)} = k, \underline{\theta}_{(r)}^{(k)}, y_1^t) P(Z_t^{(k)} = z_t^{(k)} | K^{(r)} = k, \underline{\theta}_{(r)}^{(k)}, y_1^t) \\
\propto P(Z_{t+1}^{(k)} = z_{t+1}^{(k)} | Z_t^{(k)} = z_t^{(k)}, K^{(r)} = k, \underline{\theta}_{(r)}^{(k)}, y_1^t) P(Z_t^{(k)} = z_t^{(k)} | K^{(r)} = k, \underline{\theta}_{(r)}^{(k)}, y_1^t)
\end{aligned} \tag{4.15}$$

El planteamiento anterior es realizado en Carlin y Chib(1996, cap.2, pg. 83, eq. 7 [2]), donde el resultado de la tercera línea se debe a la independencia entre $Z_t^{(k)}$ y $(Z_{t+2}^{T(k)} = z_{t+2}^{T(k)}, y_{t+1}^T | Z_{t+1}^{(k)} = z_{t+1}^{(k)}, Z_t^{(k)} = z_t^{(k)}, K^{(r)} = k, \underline{\theta}_{(r)}^{(k)}, y_1^t)$. Este resultado no es más que la productoria entre la función de *filtro* en Z_t y la probabilidad de transición conociendo $Z_{t+1}^{(k)} = z_{t+1}^{(k)}$. Para encontrar las funciones de probabilidad de filtro en cada momento se usa el proceso iterativo que aparece a continuación y que se puede ver en Frühwirth-Schnatter (2006, cap. 11 [7]).

i Conociendo $P(Z_{t-1}^{(k)} = i | \underline{\theta}_{(r)}^{(k)}, K_{(r)} = k, y_1^{t-1})$ se calcula $P(Z_t^{(k)} = l | \underline{\theta}_{(r)}^{(k)}, K_{(r)} = k, y_1^{t-1})$ mediante:

$$\begin{aligned}
P(Z_t^{(k)} = l | \underline{\theta}_{(r)}^{(k)}, K_{(r)} = k, y_1^{t-1}) \\
= \sum_{i=1}^K P(Z_t^{(k)} = l | Z_{t-1}^{(k)} = i, \underline{\theta}_{(r)}^{(k)}, K_{(r)} = k, y_1^{t-1}) P(Z_{t-1}^{(k)} = i | \underline{\theta}_{(r)}^{(k)}, K_{(r)} = k, y_1^{t-1})
\end{aligned} \tag{4.16}$$

ii A partir del paso anterior se deduce $P(Z_t^{(k)} = l | \underline{\theta}_{(r)}^{(k)}, K_{(r)} = k, y_1^{t-1})$, mediante la siguiente ecuación:

$$\begin{aligned} P(Z_t^{(k)} = l | \underline{\theta}_{(r)}^{(k)}, K_{(r)} = k, y_1^t) \\ = \frac{P(y_t | Z_t^{(k)} = l, \underline{\theta}_{(r)}^{(k)}, K_{(r)} = k, y_1^{t-1}) P(Z_t^{(k)} = l | \underline{\theta}_{(r)}^{(k)}, K_{(r)} = k, y_1^{t-1})}{\sum_{i=1}^k P(y_t | Z_t^{(k)} = i, \underline{\theta}_{(r)}^{(k)}, K_{(r)} = k, y_1^{t-1}) P(Z_t^{(k)} = i | \underline{\theta}_{(r)}^{(k)}, K_{(r)} = k, y_1^{t-1})} \end{aligned} \quad (4.17)$$

Partiendo de $\delta_{(r)}^k$, se calculan las diversas funciones de filtración mediante (i) y (ii) hasta llegar a $P(Z_T^{(k)} = l | \underline{\theta}_{(r)}^{(k)}, K_{(r)} = k, y_1^T)$, de esta probabilidad se puede generar $Z_{T(r)}^{(k)}$. Conociendo $Z_{t+1(r)}^{(k)}$ se crea el valor de $Z_{t(r)}^{(k)}$, a partir de la ecuación 4.15 esto para cada t desde $T-1$ a 1.

Con la generación de la cadena de Markov en la iteración (r) ya se tienen los diversos valores de $\underline{\theta}_{(r)}^{(k)}$ donde $K_{(r)} = k$, para la simulación de los demás parámetros $\underline{\theta}_{(r)}^{(g)}$, cuando $g \neq k$, se usa la forma que presenta a continuación.

- Distribución *a posteriori* condicional de los parámetros $\underline{\theta}^{(g)} = (Z_{1(r)}^{T(g)}, \xi_Z^{(g)}, \xi_Y^{(g)})$ conociendo el valor de $K_{(r)}$ tal que $K_{(r)} = k$ y $g \neq k$; en este caso se toman distribuciones **Pseudo-Aprior** independientes para los parámetros $\xi_Z^{(g)}$ y $\xi_Y^{(g)}$. Con las simulaciones de los anteriores parámetros las cadenas se suponen tal como se realizó en el paso (d) cambiando los parámetros por los simulados de las correspondientes Pseudo-Aprior .
- e) La distribución Pseudo-Aprior para cada componente de $\xi_Z^{(g)} = (\delta_j^{(g)}, \underline{\beta}^{(g)})$ se toma de manera independiente. En el desarrollo de esta investigación se usa como $\underline{\beta}_{ij}^{(g)} \sim N(\mu_{ij}^{(g)}, \Sigma_{ij}^{(g)})$, el cual toma en cuenta un conocimiento preliminar y una distribución a plantear.
- f) La distribución Pseudo-Aprior para $\xi_Y^{(g)} = (\underline{\varrho}_{1(r)}^{(g)}, \underline{\varrho}_{2(r)}^{(g)}, \dots, \underline{\varrho}_{g(r)}^{(g)})$ se realiza de manera independiente de cada componente tomando distribuciones conocidas dependiendo de la distribución del proceso observado que se está asumiendo. Por ejemplo, si se está modelando con un proceso observado y_1^T con un *NHHMM-Poisson* $\underline{\varrho}_{i(r)}^{(g)} = \lambda_{i(r)}^{(g)}$ se puede tomar una distribución Pseudo-Aprior Gamma o Normal truncada.

4.3.2. Proceso de simulación

Para la simulación de los correspondientes parámetros se plantea un número máximo de posibles estados M , para ello es importante recordar que por cada posible número de estados hay un modelo que asocia sus correspondientes conjuntos de parámetros $(\xi_Z^{(k)}, \xi_Y^{(k)})$. El proceso de simulación comienza con el planteamiento de valores iniciales para cada conjunto de parámetros asociado a cada uno de los posibles números de estados.

- Con $K = 1$, $(\xi_{Z(0)}^{(1)}, \xi_{Y(0)}^{(1)}) = ((\underline{\beta}_{(0)}^{(1)}), (\underline{\varrho}_{(0)}^{(1)}))$

- Con $K = 2$, $(\xi_{Z(0)}^{(2)}, \xi_{Y(0)}^{(2)}) = ((\delta_{(0)}^{(2)} \underline{\beta}_{(0)}^{(2)}), (\underline{\varrho}_{1(0)}^{(2)}, \underline{\varrho}_{2(0)}^{(2)}))$
- Con $K = M$, $(\xi_{Z(0)}^{(M)}, \xi_{Y(0)}^{(M)}) = ((\delta_{(0)}^{(M)} \underline{\beta}_{(0)}^{(M)}), (\underline{\varrho}_{1(0)}^{(2)}, \dots, \underline{\varrho}_{M(0)}^{(2)}))$

Partiendo de los valores iniciales y conociendo un proceso observado y_1^T de un proceso estocástico Y_1^T se simulan cada una de las cadenas de Markov ocultas siguiendo el numeral **(d)** del capítulo anterior, para de este modo obtener $Z_{1(0)}^{T(1)}, Z_{1(0)}^{T(2)}, \dots, Z_{1(0)}^{T(M)}$.

Para la simulación de los valores en la primera iteración se comienza con la simulación del número de estados $K_{(1)}$, a partir de la distribución condicional completa $P(K = h | (Z_{1(0)}^{T(2)}, \dots, Z_{1(0)}^{T(M)}, \xi_{Z(0)}^{(1)}, \xi_{Y(0)}^{(1)}), (\xi_{Z(0)}^{(2)}, \xi_{Y(0)}^{(2)}), \dots, (\xi_{Z(0)}^{(M)}, \xi_{Y(0)}^{(M)}), y_1^T)$ descrita en el numeral **(a)**. Esta distribución es discreta por lo que se calculan las probabilidades para cada valor de h con $h = 1, \dots, M$ y se simula un valor con estos resultados. Una vez se obtiene $K_{(1)}$ suponemos cada componente de $\xi_{Z(1)}^{(k)}$ y $\xi_{Y(1)}^{(k)}$, de tal forma que si $K_{(1)} = k$ se simula de las siguientes distribuciones completas mencionadas en los numerales **(b)** y **(c)** del capítulo anterior.

$$P(\underline{\varrho}_{i(1)}^{(k)} | \xi_{Z(0)}^{(k)}, Z_{1(0)}^{T(k)}, K_{(1)} = k)$$

$$P(\underline{\beta}_{ij}^{(k)} | \xi_{Y(1)}^{(k)}, Z_{1(0)}^{T(k)}, K_{(1)} = k)$$

Las distribuciones condicionales completas de $\underline{\varrho}_{i(1)}^{(k)}$ son formas conocidas de la familia exponencial tal como se muestra en el numeral **(b)** por lo que se simulan de manera directa con el software, mientras que la simulación de los parámetros $\underline{\beta}_{ij}$ se usa el algoritmo de *Metropolis - Hastings usando un ruido blanco gaussiano para los valores propuestos de contraste en la simulación*.

En el caso de que $K_{(1)} \neq k$, la simulación de $\xi_{Z(1)}^{(g)}$ y $\xi_{Y(1)}^{(g)}$ se realiza de las distribuciones **pseudo a prior** descritas en el numeral **(e)** y **(f)**. Con las simulaciones realizadas de los parámetros $\xi_{Z(1)}^{(h)}$ y $\xi_{Y(1)}^{(h)}$ para todo $h = 1, \dots, M$ se simulan las posibles Cadenas de Markov $Z_{1(1)}^{T(1)}, Z_{1(1)}^{T(2)}, \dots, Z_{1(1)}^{T(M)}$ del numeral **(d)**; de esta manera obtenemos los valores en la iteración **(1)**:

$$(K_{(1)}, (Z_{1(1)}^{T(1)}, \xi_{Z(1)}^{(1)}, \xi_{Y(1)}^{(1)}), \dots, (Z_{1(1)}^{T(M)}, \xi_{Z(1)}^{(M)}, \xi_{Y(1)}^{(M)}))$$

Se simulan los valores iterativamente hasta tener un número bastante amplio de valores simulados para la inferencia realizada, para garantizar la convergencia en los valores simulados. Una vez se obtiene una alta cantidad de valores simulados del parámetro K así como cada parámetro que contempla cada uno de los modelos, se comienza la estimación con la del número de estados $\hat{K} = K_c$ usando el valor modal de los valores obtenidos de la simulación. Esto debido a que su distribución condicional completa es de tipo discreta y con la estimación de este número de estados se determina el modelo que mejor se ajusta los datos.

Con la estimación del número de estados se realiza la estimación de los parámetros correspondientes al modelo que representa la estimación de \hat{K} , es decir el modelo que mejor ajusta los datos con \hat{K} . Dado que $\hat{K} = K_c$ la inferencia de los parámetros se realiza con la información simulada de cada uno de ellos correspondiente a iteraciones no provenientes de pseudo a prior s, los demás valores en la cadena no se toman en cuenta para la inferencia. De forma análoga la construcción de intervalos de credibilidad y otra estadística de la que requiera el cálculo.

Exploración de la inferencia Bayesiana

Este capítulo presenta algunos resultados teóricos de *NHHMM-Poisson*, *NHHMM-Binomial* y *NHHMM-Exponencial*. Adicionalmente, se muestra la estimación de los parámetros, según la metodología propuesta, en dos series de tiempo generadas de un *NHHMM-Poisson* con tres estados y de un *NHHMM-Poisson* con cuatro estados.

5.1. Caso Poisson

Sea $\{y_t\}_{t=1,\dots,T}$ una serie de tiempo observada de forma que $Y_t \sim \text{Poisson}(\lambda_i)$ con $i = 1, \dots, K$, adicionalmente sea $Z_1^T = z_1^T$ una Cadena de Markov no homogénea tal que las probabilidades de transición dependen del comportamiento de variables exógenas en cada momento mediante la función logística. Es importante recordar que esta cadena influencia la distribución de probabilidad del proceso observado en cada momento y de esta manera el conjunto de parámetros es $\xi_Y^{(K)} = (\lambda_1, \lambda_2, \dots, \lambda_K)$ y $\xi_Z^{(K)} = (\underline{\beta}^{(K)}, \underline{\delta}^{(K)})$. Además, la función de verosimilitud es dada por:

$$\begin{aligned}
 P(Y_1^T = y_1^T | K, Z_1^{T(K)} = z_1^T, \xi_Y^{(K)}, \xi_Z^{(K)}) &= \prod_{t=1}^T \frac{e^{-\lambda_{z_t}} \lambda_{z_t}^{y_t}}{y_t!} \\
 &= \prod_{i=1}^K \frac{e^{-n_i \lambda_i} \lambda_i^{\sum_{j=1}^{n_i} y_j}}{\prod_{j=1}^{n_i} y_j!} \\
 &= \frac{e^{-\sum_{i=1}^K n_i \lambda_i} \prod_{i=1}^K \lambda_i^{\sum_{j=1}^{n_i} y_j}}{\prod_{i=1}^K \prod_{j=1}^{n_i} y_j!}
 \end{aligned} \tag{5.1}$$

En este caso se puede tomar una distribución *a priori* Gamma truncada con $\alpha_i, \eta_i, i = 1, \dots, K$ tal que $\alpha_i > \alpha_j$ y $\eta_i > \eta_j$ cuando $i > j$ con el fin de evitar la no identificabilidad en el modelo. De esta manera se toma como distribución *a priori* $P(\lambda_i | \alpha_i, \eta_i) \propto \lambda_i^{\alpha_i - 1} e^{-\lambda_i \eta_i}$, la cual es conjugada para cada parámetro λ_i con $i = 1, \dots, k$, por lo tanto la distribución *a posteriori* es $P(\lambda_i) \propto \lambda_i^{\sum_{l=1}^{n_i} y_l + \alpha - 1} e^{-\lambda_i(\eta + n_i)}$ que corresponde al Kernel de una distribución

Gamma con $\alpha_{pos} = \sum_{l=1}^{n_i} y_l + \alpha$ y $\eta_{pos} = \eta + n_i$. Por esta razón, la distribución *a posteriori* conjunta de todos los parámetros del modelo es generada por:

$$\begin{aligned}
\pi(K = k, \theta^{(k)}) &\propto P(Y_1^T = y_1^T | Z_1^{T(k)}, \xi_Z^{(k)}, \xi_Y^{(k)}, K = k) \\
&P(Z_1^{T(k)} | \xi_Z^{(k)}, \xi_Y^{(k)}, K = k) P(\xi_Z^{(k)}, \xi_Y^{(k)} | K = k) P(K = k) \\
&\propto \left(\frac{e^{-\sum_{i=1}^K n_i \lambda_i} \prod_{i=1}^K \lambda_i^{\sum_{j=1}^{n_i} y_j}}{\prod_{i=1}^K \prod_{j=1}^{n_i} y_j!} \right) \left(\prod_{i=1}^k \prod_{i \neq j}^{k-1} \frac{e^{-1/2(\beta_{ij}^{(k)} - \mu_{\beta_{ij}}) \sigma_{\beta_{ij}}^{-1} (\beta_{ij}^{(k)} - \mu_{\beta_{ij}})' + \sum_{t=1}^{n_{ij}} x'_t \beta_{ij}}}{\prod_{t=1}^{n_{ij}} (1 + e^{x'_t \beta_{ij}})} \right) \left(\frac{\sum_{t=1}^{n_{ii}} x'_t \beta_{ii}}{\prod_{t=1}^{n_{ij}} (1 + e^{x'_t \beta_t})} \right) \\
&\propto (P(Z_0 = z_0^{(l)}) \prod_{i=1}^{k-1} \delta_i^{a_i - 1}) P(K = k) \\
&\propto e^{-\sum_{i=1}^K n_i \lambda_i} \prod_{i=1}^k \frac{\lambda_i^{\sum_{j=1}^{n_i} y_j}}{\prod_{j=1}^{n_i} y_j!} \prod_{i \neq j}^{k-1} \frac{e^{-1/2(\beta_{ij}^{(k)} - \mu_{\beta_{ij}}) \sigma_{\beta_{ij}}^{-1} (\beta_{ij}^{(k)} - \mu_{\beta_{ij}})' + \sum_{t=1}^{n_{ij}} x'_t \beta_{ij}}}{\prod_{t=1}^{n_{ij}} (1 + e^{x'_t \beta_{ij}})} \left(\frac{\sum_{t=1}^{n_{ii}} x'_t \beta_{ii}}{\prod_{t=1}^{n_{ij}} (1 + e^{x'_t \beta_t})} \right) \\
&(P(Z_0 = z_0^{(l)}) \prod_{i=1}^{k-1} \delta_i^{a_i - 1}) P(K = k)
\end{aligned} \tag{5.2}$$

La distribución *a posteriori* condicional para cada parámetro del proceso es dada por:

$$P(\lambda_i | Z_{1:T}^{(K)}, \xi_Z^{(K)}, K, y_{1:T}) \propto e^{-\sum_{i=1}^K n_i \lambda_i} \frac{\lambda_i^{\sum_{j=1}^{n_i} y_j}}{\prod_{j=1}^{n_i} y_j!} \tag{5.3}$$

5.2. Caso Binomial

La distribución Binomial hace parte de la familia exponencial donde p_i corresponde a la probabilidad de éxito y \mathfrak{n} el total ensayos evaluados y lo que se quiere calcular es la probabilidad de que cierta cantidad de ellos tengan éxito. De esta manera un *Binomial-NHHMM* contempla el vector de parámetros $\varrho^{(k)} = (p_1^{(k)}, p_2^{(k)}, \dots, p_k^{(k)})$; en este caso se proponen una distribución *a priori* *Beta*(a_i, b_i) tal que $b_i < b_j$ si $i > j$ con el fin de controlar el problema de la no identificabilidad del modelo.

Una vez observada una muestra $\{Y_t\}_1^T$ de un proceso estocástico $\{Y_t\}_1^T$ de manera tal que $Y_t = y_t | Z_t = z_t \sim \text{Binomial}(\mathfrak{n}, p_{z_t})$, entonces la función de verosimilitud condicional al conocimiento de la cadena en este caso estaría dada por:

$$\begin{aligned}
P(Y_1^T = y_1^T | K, Z_1^{T(K)} = z_1^T, \xi_Y^{(K)}, \xi_Z^{(K)}) &= \prod_{t=1}^T \binom{\mathfrak{n}}{y_t} p_i^{y_t} (1 - p_i)^{\mathfrak{n} - y_t} \\
&\propto \prod_{i=1}^K p_i^{\sum_{j=1}^{n_i} y_j} (1 - p_i)^{\mathfrak{n}^2 - \sum_{j=1}^{n_i} y_j} \prod_{j=1}^{n_i} \binom{\mathfrak{n}}{y_j}
\end{aligned} \tag{5.4}$$

Para este escenario la distribución conjunta *a posteriori* es dada por:

$$\begin{aligned}
\pi(K = k, \theta^{(k)}) &\propto P(Y_1^T = y_1^T | Z_1^{T(k)}, \xi_Z^{(k)}, \xi_Y^{(k)}, K = k) \\
&P(Z_1^{T(k)} | \xi_Z^{(k)}, \xi_Y^{(k)}, K = k) P(\xi_Z^{(k)}, \xi_Y^{(k)} | K = k) P(K = k) \\
&\propto \left(\prod_{i=1}^K p_i^{\sum_{j=1}^{n_i} y_j} (1 - p_i)^{\mathbb{n}^2 - \sum_{j=1}^{n_i} y_j} \prod_{j=1}^{n_i} \binom{\mathbb{n}}{y_j} \right) \left(\prod_{i=1}^k \prod_{i \neq j}^{k-1} \frac{e^{-1/2(\beta_{ij}^{(k)} - \mu_{\beta_{ij}}) \sigma_{\beta_{ij}}^{-1} (\beta_{ij}^{(k)} - \mu_{\beta_{ij}})' + \sum_{t=1}^{n_{ij}} x'_t \beta_{ij}}}{\prod_{t=1}^{n_{ij}} (1 + e^{x'_t \beta_{ij}})} \right) \\
&\left(\frac{\sum_{t=1}^{n_{ii}} x'_t \beta_{ii}}{\prod_{t=1}^{n_{ij}} (1 + e^{x'_t \beta_{ij}})} \right) \propto (P(Z_0 = z_0^{(l)}) \prod_{i=1}^{k-1} \delta_i^{a_i - 1}) P(K = k)
\end{aligned} \tag{5.5}$$

La distribución condicional completa de cada p_i es proporcionada por:

$$P(p_i | Z_1^{T(K)}, \xi_Y^{(K)}, y_1^T, K) \propto p_i^{\sum_{j=1}^{n_i} y_j} (1 - p_i)^{\mathbb{n}^2 - \sum_{j=1}^{n_i} y_j} \tag{5.6}$$

Antes de avanzar es relevante recordar que la distribución a prior $Beta(a_i, b_i)$ es conjugada con relación a p_i .

5.3. Caso exponencial

Esta distribución es una de las distribuciones continuas que hacen parte de la familia exponencial usada para describir la probabilidad del tiempo que tarda un elemento en realizar una actividad. Sea $\{y_t\}_1^T$ una muestra de un proceso estocástico $\{Y_t\}_1^T$ de forma tal que $Y_t | Z_t = z_t \sim Exp(v_{z_t})$, por ende $P(Y_t = y_t | Z_t = z_t) = v_{z_t} e^{-v_{z_t} y_t}$. En este caso el conjunto de parámetros es $\xi_Y^{(k)} = (v_1^{(k)}, v_2^{(k)}, \dots, v_k^{(k)})$. Una vez observada la muestra la función de verosimilitud es dada en:

$$\begin{aligned}
P(Y_1^T = y_1^T | K, Z_1^{T(K)} = z_1^T, \xi_Y^{(K)}, \xi_Z^{(K)}) &= \prod_{t=1}^T P(Y_t = y_t | K, Z_1^{T(K)} = z_1^{T(K)}, \xi_Y^{(K)}, \xi_Z^{(K)}) \\
&= \prod_{t=1}^T v_{z_t} e^{-v_{z_t} y_t} = \prod_{i=1}^K \prod_{j=1}^{n_i} v_i e^{-v_i y_t} = \prod_{i=1}^K n_i v_i e^{-v_i \sum_j^{n_i} y_j}
\end{aligned} \tag{5.7}$$

En este caso se recomienda el uso de una distribución a prior $Gamma(\alpha, \eta)$ tal que la distribución conjunta *a posteriori* es:

$$\begin{aligned}
\pi(K = k, \theta^{(k)}) &\propto P(Y_1^T = y_1^T | Z_1^{T(k)}, \xi_Z^{(k)}, \xi_Y^{(k)}, K = k) \\
P(Z_1^{T(k)} | \xi_Z^{(k)}, \xi_Y^{(k)}, K = k) &P(\xi_Z^{(k)}, \xi_Y^{(k)} | K = k) P(K = k) \\
&\propto \left(\prod_{i=1}^K n_i v_i^{\alpha_i} e^{-v_i(\sum_j^{n_i} y_j + \eta_i)} \right) \left(\prod_{i=1}^k \prod_{i \neq j}^{k-1} \frac{e^{-1/2(\beta_{ij}^{(k)} - \mu_{\beta_{ij}}) \sigma_{\beta_{ij}}^{-1} (\beta_{ij}^{(k)} - \mu_{\beta_{ij}})' + \sum_{t=1}^{n_{ij}} x_t' \beta_{ij}}}{\prod_{t=1}^{n_{ij}} (1 + e^{x_t' \beta_{ij}})} \right) \left(\frac{\sum_{t=1}^{n_{ii}} x_t' \beta_{ii}}{\prod_{t=1}^{n_{ij}} (1 + e^{x_t' \beta_{ij}})} \right)
\end{aligned} \tag{5.8}$$

Y la distribución condicional completa de cada λ_i es dada por:

$$P(v_i | Z_1^{T(K)}, \xi_Y^{(K)}, y_1^T, K) \propto v_i^{\alpha_i + 1 - 1} e^{-v_i(\eta + \sum_{j=1}^{n_i} y_j)} \tag{5.9}$$

La cual representa el Kernel de una distribución Gamma con parámetros $\alpha_i + 1$ y $\eta + \sum_{j=1}^{n_i} y_j$.

5.4. Exploración de la metodología

Con el fin de explorar la metodología propuesta en el capítulo anterior, se generaron dos procesos estocástico obtenidos de un *NHHMM-Poisson* con $K = 3$ y $K = 4$. Las figuras 5.1 y 5.2 muestran las series de tiempo, de tal forma que encima se encuentran las cadenas de Markov (Color Azul) de la que se generó. Es relevante agregar que las series se generan a partir de los siguientes valores $\lambda_3 = (10, 20, 30)$, $\lambda_4 = (10, 20, 30, 40)$, $\delta_3 = (0.33, 0.33, 0.33)$, $\delta_4 = (0.25, 0.25, 0.25, 0.25)$, β_3 y β_4 , lo cual es igual a:

$$\beta_3 = \begin{pmatrix} 2 & 1.5 & 2 & -0.1 \\ 1.2 & -1.5 & 3.1 & 2 \\ 0.5 & 2 & 1 & 3 \end{pmatrix}$$

$$\beta_4 = \begin{pmatrix} 2 & 1.3 & 3.1 & -0.1 & -0.4 & -1.5 \\ 2.1 & 1.3 & 4 & 3.2 & -1.3 & 2 \\ -2 & 1.2 & -1.2 & 3.1 & 1 & 1.1 \\ 3.1 & 1.2 & -1.4 & 1.5 & 4 & 1.5 \end{pmatrix}$$

Con el fin de observar el impacto de la forma en las co-variables al proceso observado, se generan cuatro series adicionales. El primer caso evalúa el comportamiento de la serie cuando la covariable proviene de una distribución normal, el segundo cuando la covariable sigue un modelo $AR(1, \phi = 0.5)$, el tercero cuando la covariable persigue un comportamiento $AR(1, \phi = 1)$ y por último cuando la variable regresora sigue un modelo $AR(1, \phi = 1.2)$. La figura 5.3 muestra la alta sensibilidad de la serie observada dependiendo de la estructura de las co-variables, cuando el valor de estas crece constante la Cadena de Markov se vuelve absorbente en un punto.

La serie en azul muestra las Cadenas de Markov que generaron las series observadas.

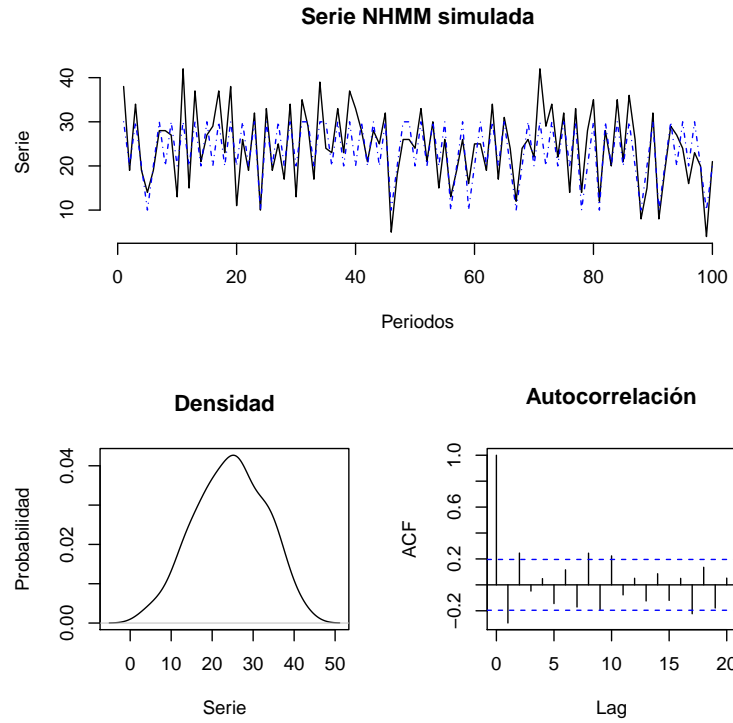


FIGURA 5.1. Serie de tiempo simulada a partir de un Poisson-NHHMM con 3 estados, así como la función de densidad y el autorrelograma (Fuente propia).

5.4.1. Resultados en la inferencia

Con las series de tiempo obtenidas en el subcapítulo anterior, se aplica la metodología propuesta en la investigación para la inferencia de los parámetros en un entorno controlado. Para la inferencia se toma como número máximo de estados $M = 5$ en cada serie. Después se plantean valores iniciales para cada conjunto de parámetros $\underline{\lambda}^{(i)}$. Con los valores iniciales se procede a la simulación de valores siguiendo el algoritmo de Gibbs hasta obtener 50.000 iteraciones de cada uno de ellos. En el caso del número de estados se estima el valor mediante el cálculo de la moda con los resultados obtenidos de las 50.000 iteraciones. Recordemos que en algunas iteraciones los valores provienen de distribuciones pseudo *a priori*.

- $\lambda_{2(0)}^{(3)} = (7, 20)$, $\lambda_{3(0)}^{(3)} = (15.63, 25.38, 34.09)$, $\lambda_{4(0)}^{(3)} = (14.48, 22.77, 27.63, 35.20)$ y $\lambda_{4(0)}^{(5)} = (7.57, 11.2, 14.84, 19.78, 25.58)$.
- $\lambda_{2(0)}^{(4)} = (7, 20)$, $\lambda_{3(0)}^{(4)} = (17.36, 28.45, 38.84)$, $\lambda_{4(0)}^{(4)} = (15.23, 25.23, 32.04, 40.33)$ y $\lambda_{4(0)}^{(5)} = (7.57, 11.2, 14.84, 19.78, 25.58)$.
- Para los valores iniciales de δ se usan vectores de igual número de componentes para la serie de 3 como de 4 estados, de tal manera que se suma en su agregado.
- Para los valores iniciales de β se toman matrices $\beta_{i(0)}^{(3)} = \beta_{i(0)}^{(4)} = \{\beta_{ij}(0)^{(k)}\} = \{1.2\}$.

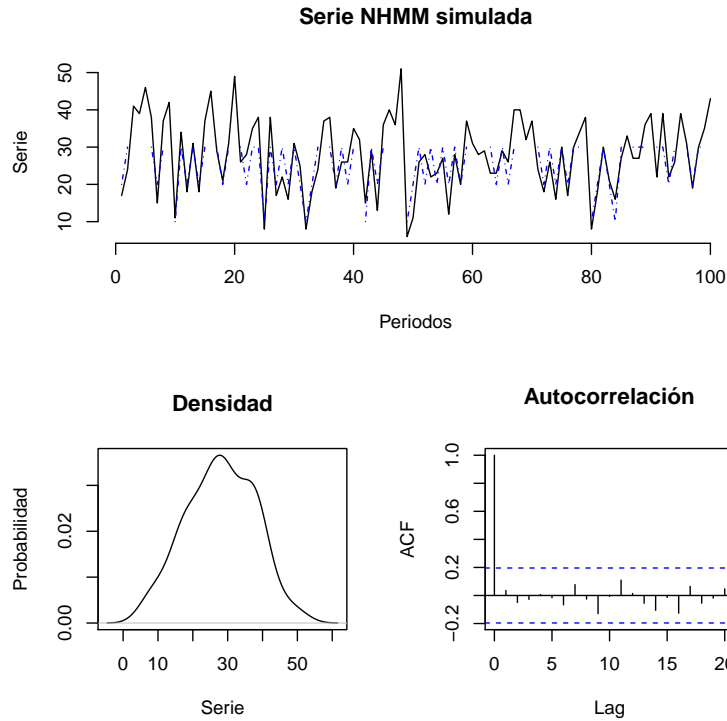


FIGURA 5.2. Serie de tiempo simulada a partir de un Poisson-NHHMM con 4 estados, así como la función de densidad y el autorrelograma (Fuente propia).

Dentro de la metodología propuesta se asumen distribuciones **pseudo a prior** normales bivariadas para cada $\beta_{ij}^{(k)}$ y distribuciones Gamma con parámetros $\alpha = \lambda_{i(0)}^{(k)}$ y $\gamma = 1$. Estas distribuciones son funciones que conectan valores en la simulación y generan formas conocidas para la estimación de los parámetros (*Es importante recalcar que la inferencia de los parámetros se realiza solamente con valores simulados de una distribución no de una pseudo prior*). El primer parámetro dentro del algoritmo de Gibbs es el número de estados. La figura (refgraficoestado1) muestra las frecuencias relativas del número de estados en las simulaciones en cada una de las series y el resumen estadístico se observa en la tabla ??.

La siguiente tabla muestra la frecuencia y frecuencia relativa del número de estados en la inferencia de las dos series de tiempo.

Resumen de la estimación del número de estados										
	Con la serie de 3 estados					Con la serie de 4 estados				
Número estados	1	2	3	4	5	1	2	3	4	5
Absolutas	0	0	21945	16803	11251	0	0	1176	27512	11311
Relativas	0	0	43.89 %	33.60 %	22.50 %	0	0	2.35 %	75.02 %	22.62 %

TABLA 5.1. Frecuencia absoluta y relativa de las iteraciones simuladas del número de estados con la serie de tres y cuatro estados.

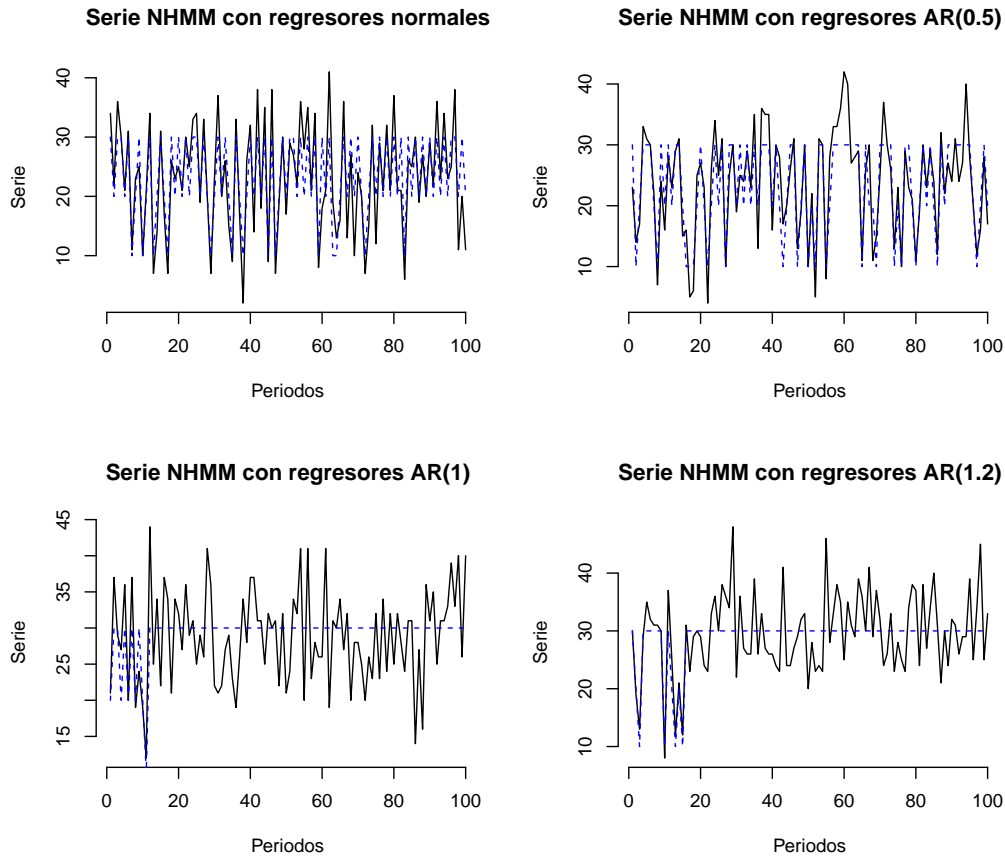


FIGURA 5.3. Series de tiempo simuladas a partir de NHHMM variando la estructura de las co-variables que ingresan al modelo (Fuente propia).

Con los resultados obtenidos en la estimación del número de estados en la de 3 estados y 4 estados, se observa estimaciones iguales a los generados en las simulaciones realizadas. *Tal como se menciona en los capítulos anteriores se tomaron en cuenta para la simulación de cada uno de ellos los valores de las iteraciones en las que el número de estados se encuentra en 3 o 4 estados.* Esto implica que los valores de la simulación para la inferencia de los parámetros se obtienen a partir de los valores que se simularon de distribuciones que no fueron pseudo a priori. Si se toman en cuenta todos los valores de la simulación se logran valores que con correlación, no hay convergencia generando estimaciones inconsistentes este resultado se analiza de lo encontrado en las simulaciones.

La tabla 5.2 muestra el resumen estadístico de las simulaciones de los parámetros Λ_i teniendo en cuenta las iteraciones que no provienen de una pseudo *a priori* para el ajuste en cada modelo.

Análogamente, se muestran las figuras con las funciones de densidad *a posteriori* tanto para el ajuste del modelo con tres estados como en el de cuatro estados, es necesario recalcar que esta distribución solo toma en cuenta los valores provenientes de distribuciones no pseudo *a priori*. Se observan distribuciones *a posteriori* algo simétricas muy diferentes a la distribución *a priori* tomada en cuenta. En el caso de tener en cuenta todas las

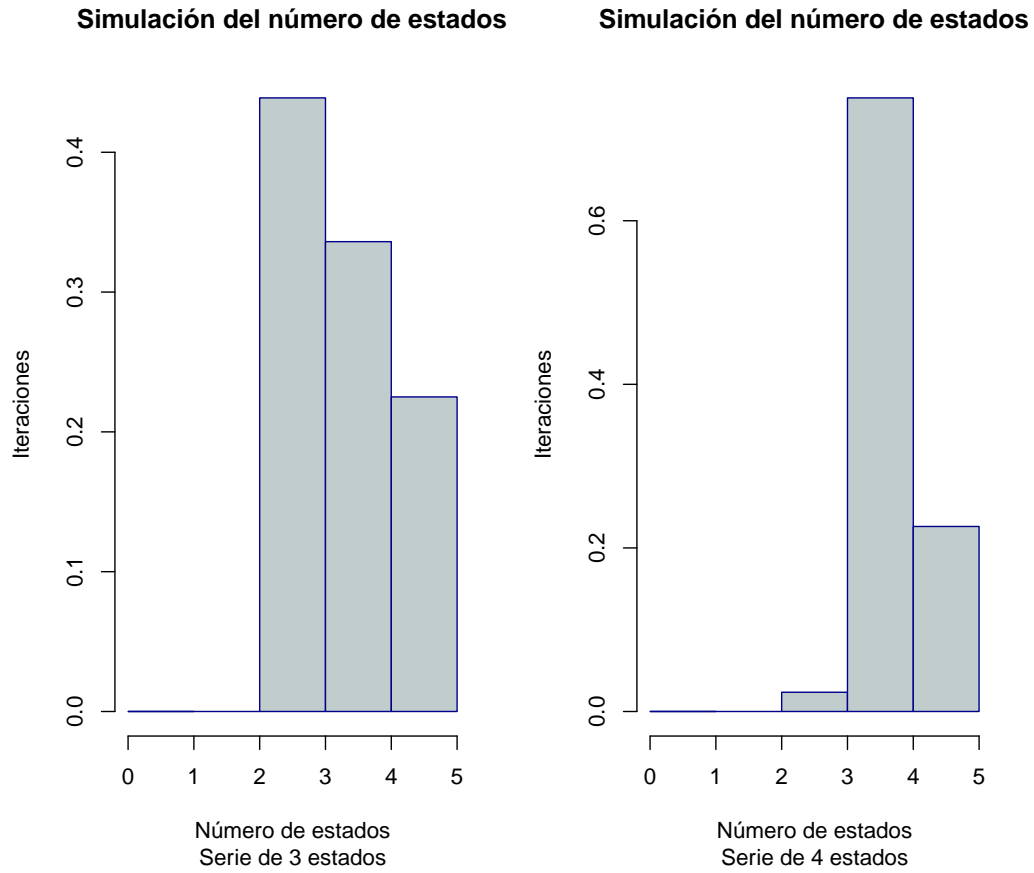


FIGURA 5.4. Simulación del número de estados en un total de 50000 iteraciones al inferir sobre dos series de tiempo de 3 y 4 estados (Fuente propia).

Resumen de los Lambda							
	De tres estados			De cuatro estados			
Estadístico	Lambda1	Lambda2	Lambda 3	Lambda1	Lambda2	Lambda3	Lambda4
Min	8.13	14.33	25.00	6.13	8.62	20.21	10.53
P2.5	11.32	17.57	28.05	11.12	12.72	27.62	31.11
P25	14.18	19.40	29.9	16.40	15.29	30.21	34.09
Mediana	16.31	20.27	30.59	18.41	16.77	31.05	35.36
Promedio	16.09	20.38	30.53	18.49	17.05	30.98	35.31
P75	17.95	21.20	31.26	20.72	18.32	31.88	36.61
P97.5	20.97	24.24	32.56	25.37	24.09	33.56	39.23
Max	26.72	27.97	34.56	36.84	42.78	36.99	44.45
Lim inf HPD	11.14	16.70	28.44	11.34	11.97	28.16	31.28
Lim sup HPD	20.84	24.25	32.79	25.62	22.40	33.91	39.37

TABLA 5.2. Resumen estadístico de las simulaciones realizadas para los parámetros Λ en cada uno de los modelos después de estimar con tres y cuatro estados para las series trabajadas.

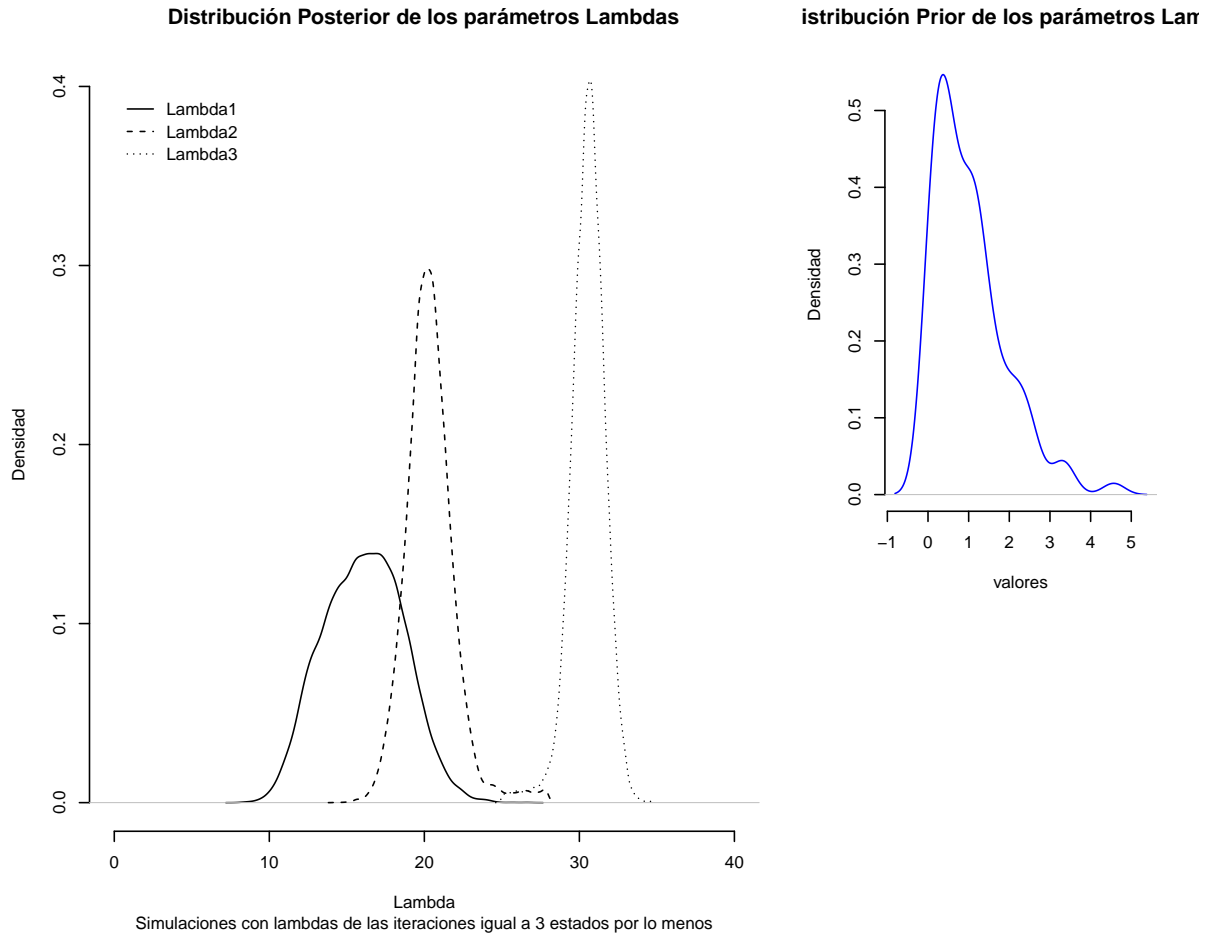


FIGURA 5.5. Gráfico de densidad de la simulación *a posteriori* de los parámetros Lambda tomando en cuenta que los Lambda no fueron de una Pseudo prior con la simulación con la serie de 3 estados y la distribución prior para cada uno de ellos (Fuente propia).

iteraciones no solamente se obtienen distribuciones *a posteriori* multimodales, asimétricas, correlacionadas, entre otras características que distorsionan el cálculo de las estimaciones.

Tanto en la simulación de cuatro estados como de tres hay estimaciones que pueden distar del parámetro real, no obstante, en algunos casos se encuentra en un intervalo de credibilidad.

Por último, se presentan los gráficos de las cadenas generadas en la simulación y los autocorrelogramas para realizar la inferencia en estos parámetros. Uno de los puntos importantes a tener en cuenta es que la presencia de una leve autocorrelación que puede indicar la toma de unos valores periódicos en la simulación para el cálculo de la inferencia. El primero de ellos relaciona el gráfico con un modelo de tres estados y el siguiente los resultados para un modelo de cuatro estados.

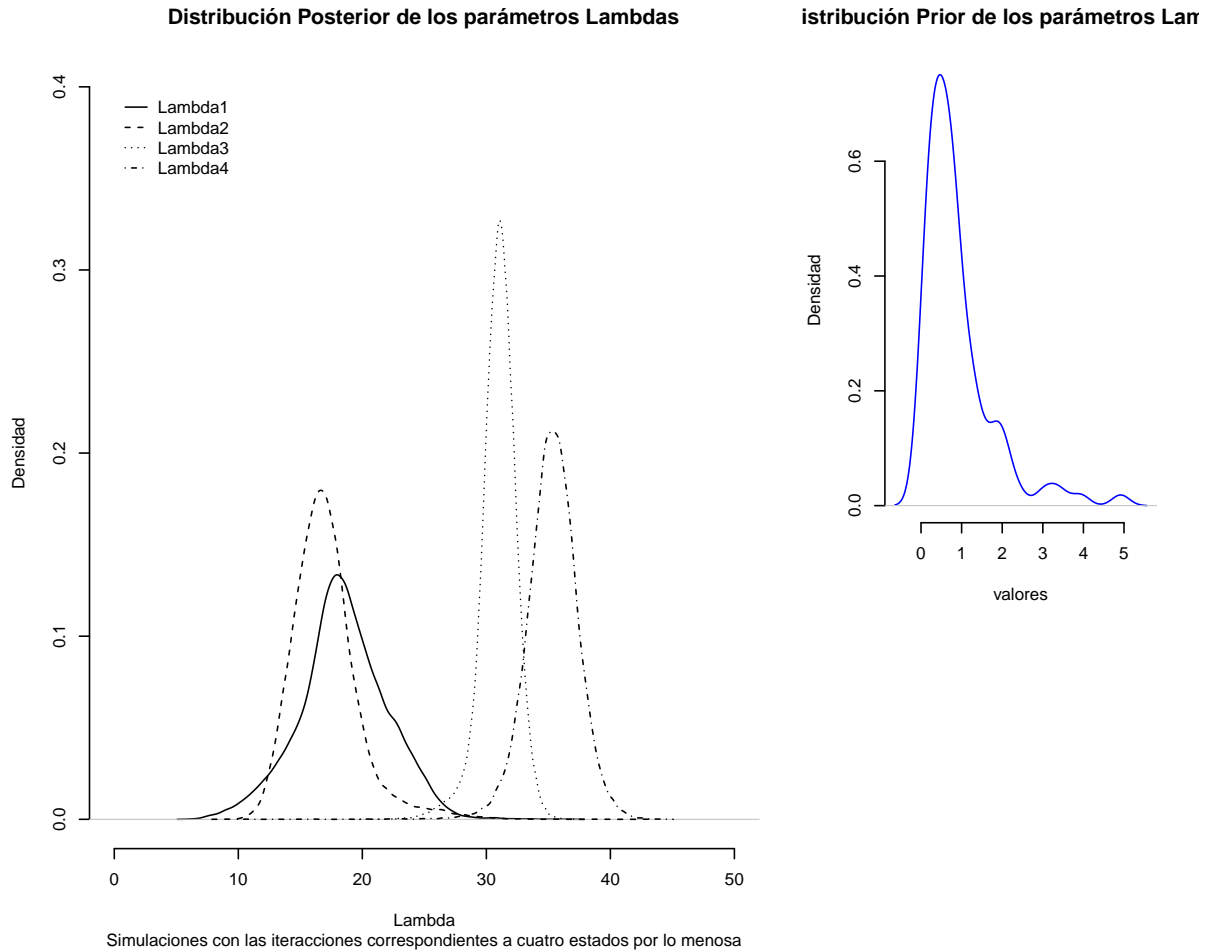


FIGURA 5.6. Gráfico de densidad de la simulación *a posteriori* de los parámetros Lambda tomando en cuenta las que no fueron de una pseudo *a priori* con la simulación de una serie de 4 estados y la distribución prior para cada uno de ellos (Fuente propia).

Exploración de los parámetros β en el proceso

Una parte que impacta bastante al proceso y convergencia de la simulación son los parámetros β . La estimación de estos presentan bastantes inconsistencias, no obstante, para el proceso se asumieron conocidos y se realizó una pequeña exploración con algunas pruebas obtenidas. La figura anterior muestra los contornos para estos parámetros con la simulación de los mismos. El punto azul representa el valor real mientras que el punto rojo representa la estimación. Además, la estimación de estos parámetros es bastante compleja por ser un vector que crece a grandes rangos a medida que aumenta el número de estados.

En la investigación se utilizaron diversos valores de parámetros iniciales y diversas distribuciones pseudo *a priori* con distintos valores de hiperparámetros. Dentro de los principales resultados obtenidos se encuentra la alta influencia de la forma de dichas distribuciones principalmente en la influencia de los valores de los hiperparámetros, ya que al tomar varianzas muy altas sobre estos se genera convergencia muy pronta del número

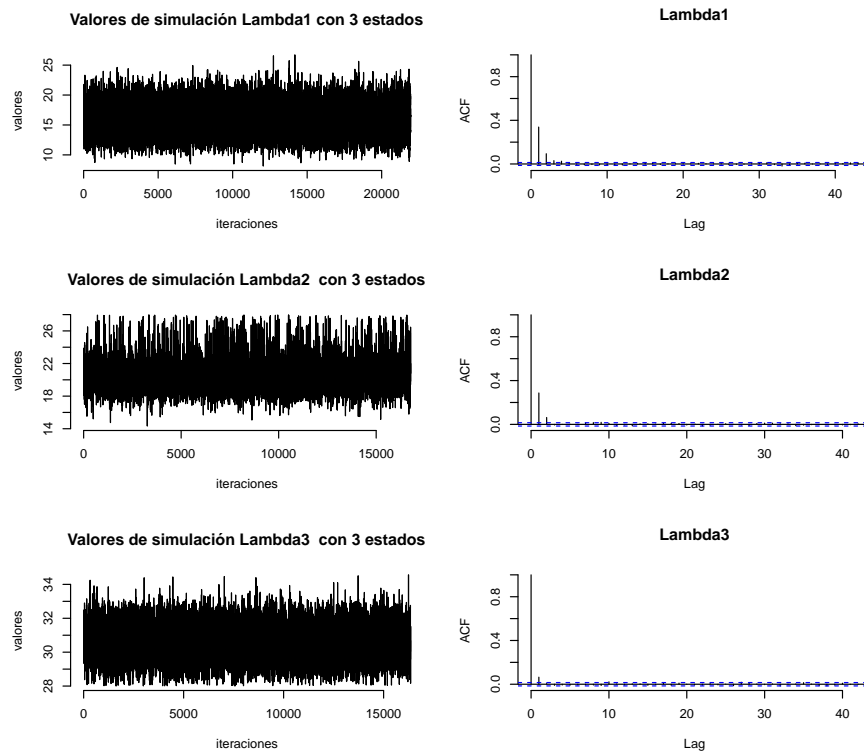


FIGURA 5.7. Gráfico de densidad de la simulación *a posteriori* de los parámetros Lambda tomando en cuenta los valores simulados que no fueron de una pseudo *a priori* con la simulación de la serie de 3 estados y la distribución prior para cada uno de ellos (Fuente propia).

de estados sobre un solo estado. Por lo tanto, es aconsejable no usar distribuciones con varianzas grandes en estos modelos.

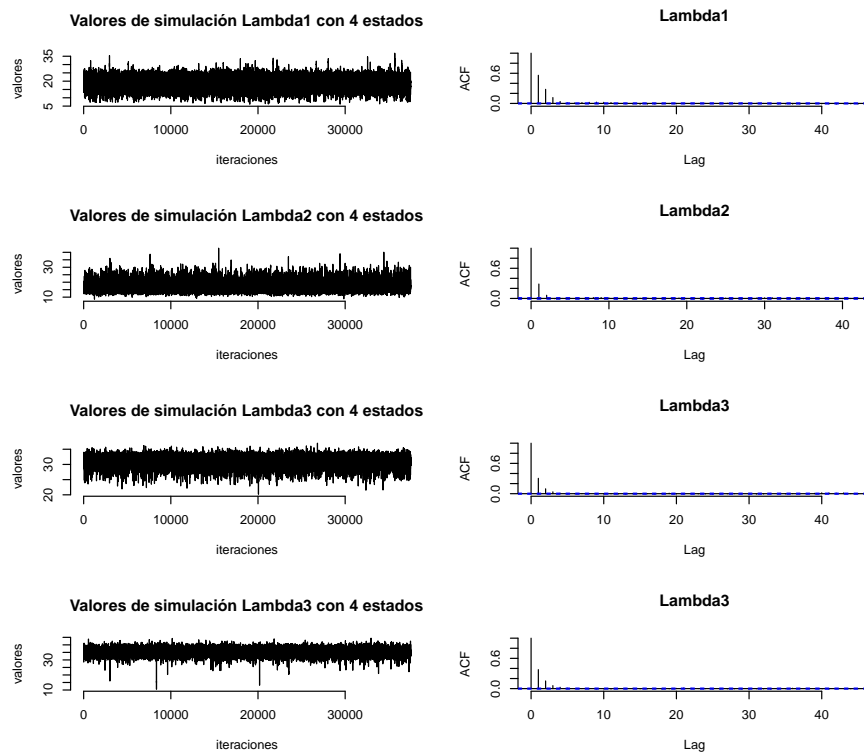


FIGURA 5.8. Gráfico de densidad de la simulación *a posteriori* de los parámetros Lambda tomando en cuenta las que no fueron de una pseudo *a priori* con la simulación de la serie de 4 estados y la distribución a prior i para cada uno de ellos (Fuente propia).

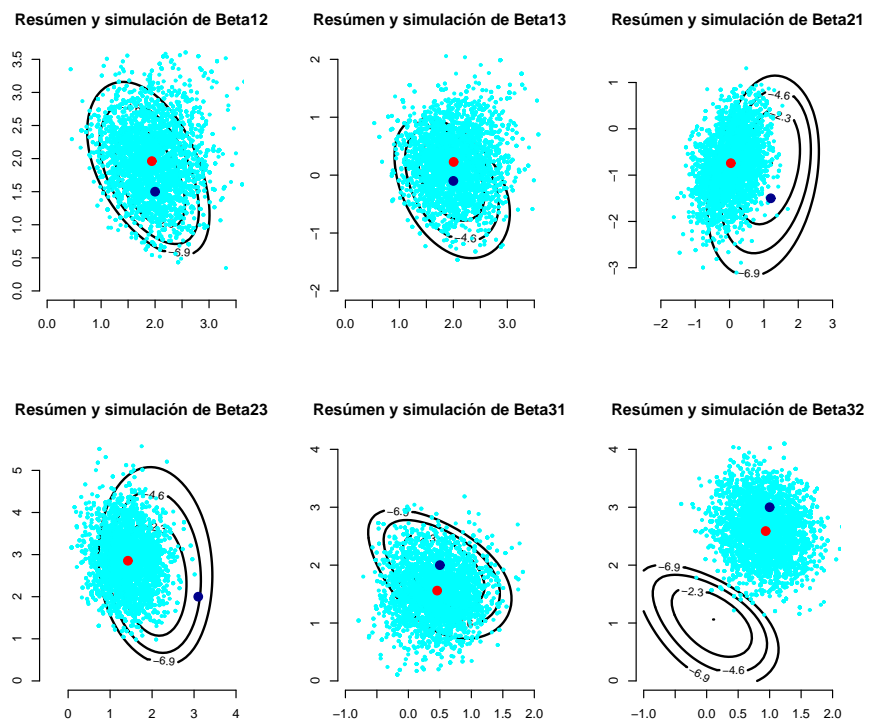


FIGURA 5.9. Gráfico de contraste entre los parámetros simulados de los β para la serie con tres estados (Fuente propia).

Conclusiones

- Para la inferencia del número de estados, la moda es un buen estadístico resumen en la estimación de este parámetro, ya que al inicio de las iteraciones hay una alta dispersión de valores mientras que a altas iteraciones se observan prevalencias en solo un número. Al ser una distribución discreta al menos uno de los valores del intervalo de credibilidad tiende a ser la misma moda. La implementación del método propuesto por Carlin y Chib (1995 [3]) genero estimaciones concordantes con los valores propuestos en las simulaciones realizadas.
- En los *NHHMM* la estimación de la Cadena de Markov mediante el Filtro utilizado, depende de la inferencia de los componentes de la matriz de parámetros β , y a su vez influyen de manera determinante el número de estados. Por ende, proponer distribuciones *a priori* y pseudo *a priori* con bajo conocimiento del problema puede generar inconvenientes en la simulación como convergencia en un único estado o incoherencias en el proceso iterativo.
- En el resumen estadístico para la estimación de los parámetros se toman en cuenta solo los valores de la simulación pertenecientes a distribuciones no pseudo *a priori*, ya que al tomar en cuenta todos los valores hay una alta variabilidad, datos extremos y valores con autocorrelaciones en las simulaciones.
- Para la convergencia en la simulación de los parámetros es importante realizar un gran número de iteraciones que garanticen la consistencia de los estimadores, ya que en la exploración realizada se encontraron algunas estimaciones con intervalos de credibilidad que incluso no alcanzan a contener el valor del parámetro del que provienen las series utilizadas.

Trabajo futuro

- Exploración, desarrollo e implementación de la inferencia del número de estados en un *NHHMM* por medio de la generalización de la inferencia propuesta por Chopin (2007 [4]) en los *HMM*. Esta metodología construye un proceso estocástico alterno para la estimación dinámica del número de estados.
- Exploración, desarrollo e implementación de la inferencia del número de estados en un *NHHMM* con los métodos propuestos en el presente documento cuando la Cadena de Markov oculta sea bivariada, obteniendo la generalización y aplicación al modelamiento de procesos punto en el espacio o modelos espaciales sobre variables que pertenezcan a la familia exponencial.
- Explorar y desarrollar métodos inferenciales Bayesianos del número de estados de la Cadena de Markov oculta en las diversas generalizaciones de los *NHHMM* Zucchini y MacDonald(2009, cap 8 [14]).
- Utilización de los métodos desarrollados para la estimación del número de estados de la Cadena de Markov en un *NHHMM* en la aplicación de la modelación de procesos estocásticos observados, de tal forma que la variable de interés pertenezca a la familia exponencial en los diversos escenarios investigativos.

Bibliografía

- [1] J. Albert, *Bayesina Computation with R.*, segunda ed., Springer, USA, 2011.
- [2] B. P. Carlin and S. Chib, *Calculating posterior distribution and modal estimation in Markov mixture models*, *Journal of Econometrics* **75** (1996), 79–97.
- [3] B.P. Carlin and S. Chib, *Bayesian Model Choise via Markov Chain Monte Carlo Methods*, *Journal of the Royal Statistical Society* **57** (1995), no. 3, 473–484.
- [4] N. Chopin, *Inference and model choice for sequentially ordered hidden Markov models*, *Journal of the Royal Statistical Society* **69** (2007), no. 2, 269–284.
- [5] N. Chopin and F. Pelgrin, *Bayesian inference and state number determination for hidden markov models: an application to the information content of the yield curve about inflation*, *Journal of Econometrics* **126** (2004), no. 2, 327–344.
- [6] A. J. Dobson and A. G. Barnett, *An introduction to generalized linear Models*, third ed., Chapman and Hall, 2008.
- [7] S. Frühwirth-Schnatter, *Finite Mixture and Markov Switching Models*, Springer, 2006.
- [8] D. Gamerman and H. F. Lopes, *Markov chain monte carlo stochastic simulation for bayesian inference*, segunda ed., Chapman and Hall/CRC, 2006.
- [9] Carlin John B. Stern Hal L. Dunson David B. Vehtari Aki Gelman, A. and Rubin Donald B., *Bayesian Data Analysis*, fourth ed., Chapman and Hall/CRC, 2009.
- [10] Guttorp P. Hughes James, P. and P. Charles Stephen, *A Nonhomogeneous Hidden Markov Model for Presipitation*, *Royal Statistical Society* **48** (1999), no. 1, 15–30.
- [11] Spezia Luigi, *Bayesian analysis of Non-Homogeneous hidden Markov Models*, *Journal of Statistical Computation and Simulation* **76** (2006), no. 8, 713–725.
- [12] J. MacKay Rachel, *Estimating the Order of a Hidden Markov Model*, *The Canadian Journal of Satatistics* **30** (2000), no. 4, 573–589.
- [13] R.L. Welch, *Hidden Markov models and the Baum Welch algorithm*, *IEEE Inform. Soc. Newsl.* **53** (2003), no. 1, 10–13.
- [14] W. Zucchini and L. MacDonald Iain, *Hidden Markov Models fo Time series, an introduction using R*, Chapman and Hall/CRC, 2009.