



UNIVERSIDAD NACIONAL AUTÓNOMA DE
MÉXICO

FACULTAD DE CIENCIAS

MODELO DE UNA CADENA DE MARKOV DE
ORDEN K NO HOMOGÉNEA Y BIVARIADA
APLICADA A EXCEDENTES DE OZONO Y PM_{10} EN
LA CIUDAD DE MÉXICO

T E S I S

QUE PARA OBTENER EL GRADO DE:
MATEMÁTICO

PRESENTA:
MARCO ANTONIO GALLEGOS HERRADA

DIRECTORA DE TESIS:
DRA. ELIANE REGINA RODRIGUES



CIUDAD UNIVERSITARIA, CDMX

MAYO, 2021

Modelo de una cadena de Markov de orden k no homogénea y bivariada aplicada a excedentes de ozono y PM_{10} en la Ciudad de México

por

Marco Antonio Gallegos Herrada

Tesis presentada para obtener el grado de

Matemático

en la

FACULTAD DE CIENCIAS

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Ciudad Universitaria, CDMX. Mayo, 2021

—Te quiero como todas las estrellas, porque son muchas y el cielo está lleno de ellas

—Yo te quiero como a todos los números.

—¿Como todos los números?

—Sí, porque no tienen fin...

A QUIÉN CORRESPONDA

Agradecimientos

Agradezco a mi familia, por estar siempre presentes incondicionalmente.

A la Dra. Eliane Regina Rodrigues, por su invaluable apoyo durante este proceso.

A las y los sinodales, , por sus correcciones y comentarios.

Al Instituto de Matemáticas de la UNAM, por haberme brindado una beca de lugar.

Índice general

1. Conceptos Básicos	3
1.1. Procesos Estocásticos	3
1.1.1. Procesos de Markov	5
1.2. Estadística bayesiana	8
1.2.1. Fórmula de Bayes	8
1.2.2. Familias Conjugadas	11
1.2.3. Selección de modelos	13
2. Cadenas de Markov de orden K	14
2.0.1. Distribución de \mathbf{W}_t	18
3. Inferencia bayesiana para un modelo Markoviano de orden K	22
3.1. Modelo Markoviano	22
3.2. Función de verosimilitud	23
3.3. Distribuciones a priori	27
3.4. Distribuciones a posteriori	29
4. Aplicación a datos de la red de monitoreo de la calidad del aire de la Ciudad de México	37
4.1. Descripción de los datos	37
4.2. Modelo matemático	42
4.3. Estimación de los parámetros	45

4.4. Selección del modelo	47
4.5. Resultados	48
5. Conclusiones	56
A.	58
A.1. Modelo en el caso $K = 2$	58
A.2. Probabilidades $P(W_t = i)$, $i = 0, 1, 2, 3$	61
A.3. Código Fuente	69

Índice de figuras

4.1. Datos de de la red de monitoreo de la CDMX de PM_{10} y ozono del periodo que comprende desde el 1 de enero de 1995 hasta el 31 de Diciembre del 2019. Las líneas horizontales corresponden a los límites permisibles para cada contaminante, i.e., $L = 75 \mu g/m^3$ para las observaciones de PM_{10} y $L = 0.095 ppm$ para el ozono.	39
4.2. Gráficas superiores: media (líneas gruesas continuas), máximo (línea punteada gruesa) y mínimo (línea punteada delgada) de las mediciones de ozono y PM_{10} por cada año del periodo observado. Gráficas inferiores: número de excedencias por año de los límites ambientales permisibles correspondientes para el ozono y PM_{10} para cada año del periodo observado.	41
4.3. Gráfica de los valores en logaritmo de $L(y K) \frac{\lambda^K}{K!}$ para $K = 0, 1, \dots, 8$. . .	46
4.4. Probabilidades de transición $P_{\bar{m}_j}^{(K)}(t)$ estimadas (líneas azules) y observadas (líneas punteadas negras) así como los intervalos de credibilidad del 95 % (líneas verdes y marrón) en el caso donde $K = 1$	49
4.5. Posterioris estimadas (líneas azules) y distribución de las observaciones (línea punteada) de $P(W_t = i)$, $t = 1, \dots, T$, así como los intervalos de credibilidad del 95 % (líneas verde y marrón) con $K = 1$	51
4.6. Valores absolutos de las diferencias entre las probabilidades $P(W_t = i)$ observadas y estimadas, $t = 1, \dots, T$, $i \in S$, de los casos $K = 0$ (líneas azules), $K = 1$ (líneas negras) y $K = 2$ (líneas rojas).	52

A.1. Distribución estimada (líneas azules) y distribución de las observaciones (línea punteada) de $P(W_t = i)$, $t = 1, \dots, T$, así como los intervalos de credibilidad del 95 % (líneas verde y marrón) con $K = 2$ 59

Introducción

Un riesgo para la salud presente en las grandes ciudades es la contaminación del aire. La Ciudad de México no es una excepción a esta problemática. Desde los años 50 se volvió notorio una disminución de la visibilidad en la ciudad ocasionada por los altos niveles de contaminación atmosférica. A finales de los 80 las autoridades ambientales de la Ciudad de México iniciaron con implementación de medidas para mitigar esta problemáticas, tales como la restricción de la circulación vehicular en el área metropolitana, regulación de una cantidad considerable de industrias, y la disminución del límite permitido de ozono en el ambiente para declarar contingencia ambiental (NOM-020-SSA1-2014). Esta última es una de las medidas que motivan la implementación de modelos como el que se presenta en este trabajo para el estudio de estos contaminantes y posterior predicción de excedencias a lo largo de los días de un año.

Este trabajo está organizado de la siguiente forma. En el capítulo 1 se abordan los conceptos básicos de cadenas de Markov, haciendo énfasis en las cadenas no homogéneas, así como los elementos primordiales que caracterizan a la estadística bayesiana. En el capítulo 2 se abordan las cadenas de Markov de orden K y definiciones pertinentes a estas. En el capítulo 3 se formula el modelo matemático a utilizar para el estudio de los contaminantes. El capítulo 4 está enfocado en la aplicación del modelo a datos de la red de monitoreo de la calidad del aire de la Ciudad de México y los resultados obtenidos a partir de la información utilizada. Los resultados presentados en este capítulo 4 dieron origen al artículo Gallegos-Herrada et al. (2021), que ya está siendo evaluado para su posible publicación. En el capítulo 5 se presentan las conclusiones del presente trabajo.

Finalmente, en el apéndice se presentan resultados alternativos a aquellos del capítulo 4, así como probabilidades estimadas de excedencia de los contaminantes estudiados a lo largo del año y también se presenta el código utilizado para la obtención de resultados.

Capítulo 1

Conceptos Básicos

En este capítulo se abordarán los conceptos y propiedades primordiales para definir las cadenas de Markov de orden K , $K \in \mathbb{N} \cup \{0\}$, las cuales se presentarán en el capítulo 2, y así poder trabajar con ellas. Asimismo, se darán las definiciones y conceptos básicos de la estadística bayesiana para poder hacer uso de modelos desde este paradigma estadístico. La información de este capítulo fue obtenida de (Ross, 1996), (Gamerman y Lopes, 2006) y (Karlin y Taylor, 1975).

1.1. Procesos Estocásticos

En esta sección presentamos los conceptos básicos de los procesos estocásticos y, en particular, de las cadenas de Markov. Se tomará en cuenta principalmente el caso de las cadenas de Markov no homogéneas. Empezaremos por la definición general de un proceso estocástico ((Ross, 1996), (Gamerman y Lopes, 2006)).

Definición 1.1.1 *Sea (Ω, \mathcal{F}, P) un espacio de probabilidad y (S, Σ) un espacio medible. Un **proceso estocástico** es una colección $\mathbf{X} = \{X_t : t \in T\}$ de variables aleatorias, tales que para toda $t \in T$,*

$$X_t : \Omega \rightarrow S.$$

El conjunto T es llamado el **conjunto de índices de \mathbf{X}** , y al conjunto S se llama el **espacio de estados de \mathbf{X}** . En lo siguiente, consideraremos $S \subseteq \mathbb{R}$ y $T \subseteq \mathbb{R}_+$.

Dependiendo de la forma de T y S , a un proceso estocástico \mathbf{X} se le puede clasificar de las siguientes formas:

1. Si T es finito o infinito numerable, se dice que el proceso \mathbf{X} es a tiempo discreto.
2. Si T es un intervalo de \mathbb{R} , se dice que el proceso \mathbf{X} es a tiempo continuo.
3. Si $S \subseteq \mathbb{Z}$, se dice que el proceso \mathbf{X} es de estado discreto.
4. Si S es un intervalo de \mathbb{R} , se dice que el proceso \mathbf{X} es de estado continuo.

Un ejemplo sencillo de un proceso estocástico son las conocidas caminatas aleatorias, (Ross, 1996) y (Gamerman y Lopes, 2006), que presentamos a continuación.

Ejemplo 1.1.1 Sea (Ω, \mathcal{F}, P) un espacio de probabilidad. Para cada $t \in \mathbb{N}$, vamos a considerar $X_t : \Omega \rightarrow \{-1, 1\}$ de tal forma que $\mathbf{X} = \{X_t : t \in \mathbb{N}\}$ es un conjunto de variables aleatorias idénticamente distribuidas que satisfacen:

$$P(X_t = i) = \begin{cases} p & \text{si } i = 1, \\ 1 - p & \text{si } i = -1. \end{cases}$$

Así, para cada $t \in \mathbb{N} \cup \{0\}$, definimos $Y_{t+1} = Y_t + X_{t+1}$, con $Y_0 \equiv 0$. Entonces tenemos que $\mathbf{Y} = \{Y_t : t \geq 0\}$ es un proceso estocástico a tiempo discreto y de estado discreto.

Notación 1.1.1 De ahora en adelante, consideraremos $\mathbb{N}^* := \mathbb{N} \cup \{0\}$.

Además del ejemplo 1.1.1, hay otros procesos estocásticos ya muy estudiados hoy en día y conocidos por excelencia, como lo son los procesos Poisson, las martingalas, los procesos de Lévy y los procesos de Markov. En este trabajo, nosotros nos enfocaremos en los procesos de Markov; más específicamente, en las *cadena de Markov*.

1.1.1. Procesos de Markov

La propiedad o dependencia de Markov es un concepto atribuido al matemático Ruso Andrei Andreivich Markov quién a principios del siglo XX investigaba la alternancia de las vocales y consonantes en el poema *Onegin* del autor Aleksandr Pushkin. Como fruto de sus trabajos, desarrolló un modelo probabilístico en el cuál resultados sucesivos dependían de sus predecesores sólo a través del más reciente de estos. Casi al mismo tiempo, el matemático francés Henri Poincaré estudiaba sucesiones de variables aleatorias que en realidad eran cadenas de Markov (Gamerman y Lopes, 2006).

A grandes rasgos, un proceso de Markov es un proceso estocástico $\mathbf{X} = \{X_t : t \in T\}$ tal que, dado el valor de X_t , la distribución de X_s , $t < s$, no dependen de los valores de X_u , $u < t$. Esto se puede traducir a que, si se sabe con certeza el estado actual del proceso, la probabilidad de cualquier comportamiento en el futuro no se verá alterada por información concerniente al pasado de dicho proceso (Karlin y Taylor, 1975). Formalmente, tenemos lo siguiente (Karlin y Taylor, 1975)

Definición 1.1.2 *Sea $\mathbf{X} = \{X_t : t \in T\}$ un proceso estocástico con espacio de estados S . Diremos que X es un **proceso de Markov** si y sólo si satisface la **propiedad de Markov**, es decir, si y sólo si para toda $t \in T$ y para cualesquiera $s_1, \dots, s_k \in T$ tales que $s_1 < \dots < s_k < t$, se cumple*

$$P(a < X_t \leq b \mid X_{s_1} = x_1, \dots, X_{s_k} = x_k) = P(a < X_t \leq b \mid X_{s_k} = x_k),$$

con $a < b$ y $x_1, \dots, x_k \in S$.

Así, habiendo formalizado lo que es un Proceso de Markov, estamos en condiciones de definir a las cadenas de Markov.

Definición 1.1.3 (Cadena de Markov) *Sea \mathbf{X} un proceso de Markov. Entonces, \mathbf{X} es una cadena de Markov si y sólo si su espacio de estados es finito o infinito numerable.*

Observación 1.1.1 *En la definición 1.1.3 no se menciona nada sobre el conjunto de índices de una cadena de Markov. Por lo tanto, podemos clasificar a las cadenas de Markov en dos grupos: las cadenas de Markov **continuas** (a tiempo continuo) y **discretas** (a tiempo discreto). En este trabajo, consideraremos que el conjunto de índices de cualquier cadena de Markov discreta estará dado por \mathbb{N} , y al momento de referirnos a una cadena de Markov, se supondrá que es discreta, a menos que se especifique lo contrario.*

A continuación, introduciremos algunos conceptos que nos serán de gran utilidad a lo largo de este trabajo:

Definición 1.1.4 *Sea $\mathbf{X} = \{X_t : t \in \mathbb{N}\}$ una cadena de Markov discreta con espacio de estados S . Entonces, dados $i, j \in S$, $n \in \mathbb{N}$, $m \in \mathbb{N}^*$, la **probabilidad de transición** del estado i al estado j en m pasos a partir del momento n , está dado por:*

$$P_{i,j}^{(n,n+m)} := P(X_{n+m} = j \mid X_n = i).$$

Dados $i, j \in S$, $n \in \mathbb{N}$, $m \in \mathbb{N}^*$, la probabilidad de transición $P_{i,j}^{(n,n+m)}$ no necesariamente depende del tiempo n . En tal situación, diremos que \mathbf{X} es una cadena **homogénea**, y en caso contrario, diremos que la cadena \mathbf{X} es **no homogénea**.

Notación 1.1.2 *Si $m = 1$, denotamos a la probabilidad de transición de ir del estado i al estado j en un paso al tiempo n como $P_{ij}^{(n,n+1)}$. Cuando tenemos una cadena homogénea entonces $P_{ij}^{(n,n+m)} = P_{ij}^{(m)}$ y $P_{ij}^{(n,n+1)} = P_{ij}$, donde $P_{ij}^{(m)}$ es la probabilidad de ir del estado i al estado j en m pasos a partir del momento $n = 1$ y P_{ij} es la probabilidad de ir del estado i al estado j en un paso a partir del momento $n = 1$.*

Observación 1.1.2 *De ahora en adelante, a menos que se indique lo contrario, estaremos considerando solamente cadenas no homogéneas.*

Proseguimos con la siguiente definición dada en (Karlin y Taylor, 1975) y (Ross, 1996):

Definición 1.1.5 Dada una cadena de Markov $\mathbf{X} = \{X_t : t \in \mathbb{N}\}$ no homogénea y discreta con espacio de estados S , definimos su **matriz de probabilidades de transición** del tiempo n al tiempo $n + 1$ como:

$$\mathcal{P}^{(n,n+1)} := (P_{i,j}^{(n,n+1)})_{i,j \in S}.$$

Definición 1.1.6 Sea $\mathbf{X} = \{X_t : t \in \mathbb{N}\}$ una cadena de Markov discreta con espacio de estados S y $n \in \mathbb{N}$. Para cada $x \in S$, definimos a la distribución de \mathbf{X} al tiempo n como $\pi^{(n)}(x) := P(X_n = x)$. Tomaremos a $\pi^{(n)}$ como el vector $\pi^{(n)} := (\pi^{(n)}(x))_{x \in S}$. En el caso $n = 1$, al vector $\pi^{(1)}$ se le conoce como la **distribución inicial** de la cadena \mathbf{X} .

Ahora, tenemos el siguiente teorema que es un clásico de la teoría de las cadenas de Markov (Gamerman y Lopes, 2006).

Teorema 1.1.1 (Ecuaciones de Chapman-Kolmogorov) Sean $n, m, r \in \mathbb{N}$ tales que $1 < r < m$. Entonces, dada una cadena de Markov $\mathbf{X} = \{X_t : t \in \mathbb{N}\}$ con espacio de estados S , tenemos que para cualesquiera $i, j \in S$ se cumple,

$$P_{i,j}^{(n,n+m)} = \sum_{k \in S} P_{i,k}^{(n,n+r)} P_{k,j}^{(n+r,m)}.$$

Demostración. Sean $n, m, r \in \mathbb{N}$ y \mathbf{X} como en las hipótesis y sean $i, j \in S$. Entonces,

$$\begin{aligned} P_{i,j}^{(n,n+m)} &= P(X_{n+m} = j \mid X_n = i) \\ &= \sum_{k \in S} P(X_{n+m} = j \mid X_n = i, X_{n+r} = k) P(X_{n+r} = k \mid X_n = i) \\ (\text{Prop. de Markov}) &= \sum_{k \in S} P(X_{n+m} = j \mid X_{n+r} = k) P(X_{n+r} = k \mid X_n = i) \\ &= \sum_{k \in S} P(X_{n+r} = k \mid X_n = i) P(X_{n+m} = j \mid X_{n+r} = k) \\ &= \sum_{k \in S} P_{i,k}^{(n,n+r)} P_{k,j}^{(n+r,n+m)}. \end{aligned}$$

Así, se cumple que $P_{i,j}^{(n,n+m)} = \sum_{k \in S} P_{i,k}^{(n,n+r)} P_{k,j}^{(n+r,n+m)}$. ■

1.2. Estadística bayesiana

En esta sección abordaremos los conceptos básicos de la inferencia bayesiana. La parte histórica proviene de la obra (Mcgrayne, 2011), mientras que la teoría y conceptos formales provienen de los textos (Gamerman y Lopes, 2006), (Carlin y Louis, 2009) y (Gelman *et al.*, 2013), entre otros.

1.2.1. Fórmula de Bayes

Desde su descubrimiento por el Reverendo Thomas Bayes, pasando por la popularidad que le dió Laplace al redescubrirla y desarrollarla hasta como la conocemos hoy en día, y hasta el auge de la estadística frecuentista con Fisher y Pearson, el solo hacer mención de la fórmula de Bayes o plantear su uso en el campo de las ciencias se consideraba un mero acto pagano; “la probabilidad de causas y eventos futuros, derivados de eventos pasados” (nombre acuñado por Laplace) era detestada por muchos (Mcgrayne, 2011).

Científicos y académicos, en su búsqueda por respuestas objetivas y precisas, la consideraban un método subjetivo y sin sustento científico alguno. Además, quién llegaba a hacer uso de ella, rara vez lo admitía o le cambiaba el nombre y alegaba usar otro método (cuando a John Tukey se le preguntó cómo funcionaban sus métodos de predicción en las elecciones estadounidenses, él decía que “tomaba fuerza prestada”, y Alan Turing, en la Segunda Guerra Mundial, trabajó con “bans”, que eran unidades de medida para un sistema bayesiano al que llamó *Banburismus*) (Mcgrayne, 2011).

Sin embargo, debido a los resultados de su aplicación en distintas áreas de la ciencia, se le ha comenzado a reconocer como una herramienta fundamental dentro de la probabilidad y la estadística. En términos actuales, esta regla se puede expresar de la siguiente forma (Gelman *et al.*, 2013):

Fórmula de Bayes. Sea (Ω, \mathcal{F}, P) un espacio de probabilidad, y sean $A, B \in \mathcal{F}$, con

$P(B) > 0$. Entonces, se satisface la siguiente igualdad:

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}.$$

Habiendo entrado un poco en contexto, ¿De qué va la Estadística bayesiana y cómo se relaciona con la fórmula de Bayes? Para poder hablar de ella, vamos a introducir los tres ingredientes que conforman un modelo bayesiano: la *función de verosimilitud*, la *distribución a priori* y la *distribución a posteriori*.

Gracias a la información que nuestros datos nos arrojen y la naturaleza de éstos, es posible hacer inferencia sobre ellos a través de un modelo asignado. Esto se hace a través de la función de densidad que describe los datos estudiados la cual deriva del modelo que se asume para describirlos y que posiblemente depende de un vector de parámetros θ . Si a dicha densidad en lugar de estar en función de los datos se toma en función de θ , entonces obtenemos a la función de verosimilitud correspondiente al conjunto de datos.

La función de verosimilitud de un modelo dada la muestra $\bar{x} = (x_1, \dots, x_n)$ es definida de manera formal a continuación (Gamerman y Lopes, 2006).

Definición 1.2.1 Dada una muestra $\bar{x} = (x_1, x_2, \dots, x_n)$ de $\bar{X} = (X_1, X_2, \dots, X_n)$, supongamos que esta puede ser descrita por un modelo representado por la función de densidad $p(\cdot|\theta)$, donde θ es el vector de parámetros. Entonces, la función de verosimilitud es la función $L(\bar{x}|\theta)$ dada por:

$$L(\bar{x}|\theta) \propto p(\bar{x}|\theta).$$

Es decir, $L(\bar{x}|\theta)$ es proporcional a la probabilidad de que tengamos la observación \bar{x} dada por la densidad $p(\cdot|\theta)$.

Ahora, lo que diferencia al método bayesiano del método frecuentista es la posibilidad de incluir información adicional acerca de nuestro parámetro desconocido θ . Esta información adicional sobre θ puede ser la opinión de gente experta en el tema a tratar, experiencias previas de acontecimientos relacionados con lo que se está modelando o

simple intuición.

Definición 1.2.2 La *distribución a priori* de θ o *distribución inicial*, denotada por $p(\theta)$, es la distribución de θ sin tomar en cuenta la información proporcionada por la muestra \bar{x} (Gamerman y Lopes, 2006).

Finalmente, la última pieza de un modelo bayesiano —la cuál es la meta de este método de inferencia— es la llamada distribución a posteriori o distribución final. En esta recopilamos la actualización de nuestras creencias (acumuladas en $p(\theta)$) que fueron reajustadas con la información de nuestras observaciones (encapsulada en $p(\cdot|\theta)$).

Definición 1.2.3 La *distribución a posteriori* o *distribución final* de θ , denotada por $p(\theta|\bar{x})$, es la distribución de θ cuando tomamos en cuenta nuestra creencia inicial sobre θ y le añadimos la información proporcionada por la muestra \bar{x} (Gamerman y Lopes, 2006).

Gracias al Teorema de Bayes podemos encontrar una conexión entre la distribución a posteriori con la distribución a priori y la función de verosimilitud. Dicha conexión es presentada a continuación (Gelman *et al.*, 2013).

Proposición 1.2.1 Sea $p(\cdot|\theta)$ una distribución de probabilidad asignada a un vector aleatorio $\bar{X} = (X_1, \dots, X_n)$ que modela un conjunto de observaciones, y sea $p(\theta)$ una distribución a priori para el vector de parámetros θ . Dada una muestra $\bar{x} = (x_1, \dots, x_n)$ de \bar{X} , tenemos que

$$p(\theta|\bar{x}) \propto L(\bar{x}|\theta)p(\theta).$$

Demostración. Sean $\bar{x}, p(\theta)$ y $p(\bar{x}|\theta)$ como en las hipótesis. Entonces, por propiedades de probabilidad condicional, tenemos

$$p(\theta|\bar{x}) = \frac{p(\theta, \bar{x})}{p(\bar{x})} \propto p(\theta, \bar{x}) = p(\bar{x}|\theta)p(\theta) \propto L(\bar{x}|\theta)p(\theta).$$

Esto se debe al hecho de que, como \bar{x} es una muestra fija de nuestro modelo \bar{X} , tenemos que $p(\bar{x})$ no depende de θ y, por tanto, es una constante con respecto a esta.

Podemos concluir que la posteriori es proporcional al producto de la priori con la función de verosimilitud. ■

Así, la estadística bayesiana puede resumirse como la creación de modelos que permitan incorporar información subjetiva para realizar inferencia sobre un vector de parámetros determinado, a la vez que tenemos la posibilidad de mejorar nuestras deducciones por medio del conjunto de datos que se modela.

1.2.2. Familias Conjugadas

Como vimos en la sección anterior, el objetivo primordial del método bayesiano es la obtención de la distribución a posteriori. En algunos casos es posible simplificar la obtención $p(\theta|\bar{x})$ al considerar lo que son conocidas como las **familias conjugadas** de distribuciones que nos permiten conocer las distribuciones a posteriori derivadas de un conjunto de funciones de verosimilitud y una clase de distribuciones iniciales. Su definición es dada a continuación (Gamerman y Lopes, 2006).

Definición 1.2.4 Sean \mathcal{P} y \mathcal{F} familias de distribuciones de probabilidad. Diremos que \mathcal{P} es conjugada de la familia \mathcal{F} si y sólo si para cualquier $\bar{X} = (X_1, \dots, X_n)$ de nuestro conjunto de observaciones tal que su función de densidad pertenece la familia \mathcal{F} , y cualquier distribución inicial $p(\theta)$ tal que $p(\theta) \in \mathcal{F}$, tenemos que, $p(\theta|\bar{x}) \in \mathcal{P}$, con \bar{x} una muestra de \bar{X} . Es decir, la distribución a posteriori del parámetro θ de un modelo bayesiano describiendo \bar{X} es de la misma familia que su distribución a priori.

Introducimos las siguientes distribuciones establecidas en (Gamerman y Lopes, 2006), las cuáles son miembros de unas familias conjugadas que nos serán de gran utilidad en el futuro de este trabajo.

Definición 1.2.5 Sean $n, k \in \mathbb{N}$, y sea $\bar{p} \in [0, 1]^k$, con $\bar{p} = (p_1, \dots, p_k)$, tal que $\sum_{i=1}^k p_i = 1$. Dado un vector aleatorio $X = (X_1, \dots, X_k)$, diremos que éste tiene una distribución **multinomial** con parámetros n y \bar{p} , y lo denotaremos por $X \sim \text{Multi}(n, \bar{p})$, si y sólo si

para cualquier vector $x \in \{0, \dots, n\}^k$, con $x = (x_1, \dots, x_k)$, su función de densidad es la siguiente:

$$f(x_1, \dots, x_n | n, p_1, \dots, p_k) = \frac{\Gamma(\sum_{i=1}^n x_i + 1)}{\prod_{j=1}^n \Gamma(x_j + 1)} \prod_{i=1}^k p_i^{x_i} \mathbb{I}_{\{\sum_{i=1}^k x_i = n\}}(x),$$

donde $\Gamma(\cdot)$ es la función gamma dada por

$$\Gamma(t) = \int_0^\infty x^{t-1} e^{-t} dx.$$

Observación 1.2.1 En el caso en que $t \in \mathbb{Z}_+$, tenemos $\Gamma(t) = (t - 1)!$

Definición 1.2.6 Sea $k \in \mathbb{N}$, con $k \geq 2$, y sean $\alpha_1, \dots, \alpha_k \in \mathbb{R}^+$. Dado $X = (X_1, \dots, X_k)$ vector aleatorio, diremos que éste tiene una distribución **Dirichlet** con vector de parámetros $\bar{\alpha} = (\alpha_1, \dots, \alpha_k)$, y lo denotaremos por $X \sim \text{Dir}(\bar{\alpha})$, si y sólo si para cualquier vector $(x_1, \dots, x_k) \in (0, 1)^k$ se tiene que su función de densidad está dada por:

$$f(x_1, \dots, x_k | \alpha_1, \dots, \alpha_k) = \begin{cases} \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{j=1}^k \Gamma(\alpha_j)} \prod_{i=1}^k x_i^{\alpha_i - 1}, & \text{si } \sum_{i=1}^k x_i = 1, \\ 0, & \text{cualquier otro caso.} \end{cases}$$

Al soporte de esta distribución se le denota por Δ_k y está dado por

$$\Delta_k = \left\{ (x_1, \dots, x_k) \in (0, 1)^k \mid \sum_{i=1}^k x_i = 1 \right\}. \quad (1.1)$$

Para dar finalidad a esta sección procederemos a demostrar que, en efecto, la clase de distribuciones Dirichlet es una familia conjugada del conjunto de las distribuciones muestrales multinomiales.

Proposición 1.2.2 El conjunto de distribuciones Dirichlet es una familia conjugada de la clase de las distribuciones muestrales multinomiales.

Demostración. Sea $n, k \in \mathbb{N}$, con $k \geq 2$, y sea $\bar{X} = (X_1, \dots, X_k)$ un modelo de un conjunto de observaciones a estudiar, con $\bar{X} \sim \text{Multi}(n, p)$, $\bar{p} = (p_1, \dots, p_k)$. Supongamos

que \bar{p} tiene distribución a priori una Dirichlet con parámetros $\bar{\alpha} = (\alpha_1, \dots, \alpha_k)$. Entonces, en este caso $\theta = \bar{p}$. Sea $\bar{x} = (x_1, \dots, x_k)$ una observación de \bar{X} . Tenemos que, por la proposición 1.2.1, para cualquier $\theta = (p_1, \dots, p_k)$, vale que

$$p(\theta|x) \propto f(\bar{x}|\theta)p(\theta) \propto \prod_{i=1}^k p_i^{x_i} \prod_{i=1}^k p_i^{\alpha_i-1} \propto \prod_{i=1}^k p_i^{x_i+\alpha_i-1}.$$

Así, tenemos que $\bar{p} | \bar{x}$ se distribuye $Dir(\alpha')$, donde $\alpha' = (\alpha_1 + x_1, \dots, \alpha_k + x_k)$. ■

1.2.3. Selección de modelos

Cuando se tiene un conjunto de datos x_1, \dots, x_n que se quiere estudiar usando un modelo estocástico, puede ser que se tengan varios modelos que podrían describir este conjunto de datos. De esta forma, uno se puede preguntar cuál de estos modelos describe mejor su comportamiento. En la inferencia bayesiana existen distintos métodos para la selección de modelos, los cuales dan preferencia de un modelo sobre otro. Uno de ellos es el siguiente (ver, por ejemplo, Barrios y Rodrigues 2015).

Definición 1.2.7 (Método SAD) *Dado un conjunto de observaciones x_1, \dots, x_n y $f(\cdot)$ la probabilidad asignada a dichas observaciones que corresponden a un modelo M , definimos la **suma de de las diferencias absolutas**, la cual vamos a denotar por *SAD* (por su nombre en inglés *Sum of Absolute values of the Differences*), como*

$$SAD = \sum_{i=1}^n |f(x_i) - \hat{f}(x_i)|,$$

donde $\hat{f}(x_i)$ son los valores estimados de $f(x_i)$.

Este método será utilizado en el capítulo 4 como una forma de seleccionar el modelo que mejor se ajusta al conjunto de datos que se estudiará en dicho capítulo.

Capítulo 2

Cadenas de Markov de orden K

Recordemos que lo que caracteriza a las cadenas de Markov (y más generalmente a los procesos de Markov) es la independencia del futuro con respecto al pasado, siempre y cuando se conozca el estado presente del proceso. Sin embargo, se puede suavizar este tipo de independencia para que el proceso dependa de más información anterior al presente, digamos, de los últimos K estados visitados (incluyendo al presente), para poder decir cosas sobre el futuro. Con esto la probabilidad de cualquier evento en el tiempo n sólo se ve afectada por los estados ocurridos a partir del tiempo $n - K$ y hasta el tiempo $n - 1$. Pasamos a dar la siguiente definición que formaliza esta idea (ver, por ejemplo, Boys y Henderson 2002, 2004, Rodrigues et al. 2015, Morales-Morillón 2018):

Definición 2.0.1 (Cadenas de Markov orden K) Sea $K \in \mathbb{N}$ y $\mathbf{W} = \{W_t : t \in \mathbb{N}\}$ un proceso estocástico a tiempo discreto y de espacio discreto S . Entonces, diremos que \mathbf{W} es una **cadena de Markov de orden K** si, y sólo si, el proceso $\mathbf{X}^{(K)} = \{X_t^{(K)} : t \in \mathbb{N}\}$, con $X_t^{(K)} = (W_t, \dots, W_{t+K-1})$, $t \in \mathbb{N}$, satisface

$$P(X_t^{(K)} = \bar{m}_{t+1} \mid X_t^{(K)} = \bar{m}_{t-1}, \dots, X_1^{(K)} = \bar{m}_1) = P(X_{t+1}^{(K)} = x_{t+1} \mid X_t^{(K)} = \bar{m}_t),$$

con $\bar{m}_1, \dots, \bar{m}_{t+1} \in S^K$.

Observación 2.0.1 Por la definición 2.0.1, tenemos que una cadena de Markov es en realidad una cadena de Markov de orden $K = 1$ y $X^{(1)} = \mathbf{W}$. Por otro lado, diremos que

un proceso estocástico $\mathbf{W} = \{W_t : t \in \mathbb{N}\}$ es una **cadena de Markov de orden 0** si y sólo si es una colección de variables aleatorias independientes.

Podemos extender la noción de distribución inicial de una cadena de Markov (de orden $K = 1$) de la siguiente forma:

Definición 2.0.2 Sea $\mathbf{W} = \{W_t : t \in \mathbb{N}\}$ una cadena de Markov de orden K y espacio de estados S y $t \in \mathbb{N}$. Para cada $x \in S^K$ sea $\pi_t^{(K)}(x) = P(X_t^{(K)} = x)$. Definimos a la distribución inicial de la cadena \mathbf{W} como el vector

$$\pi_1^{(K)} := (\pi_1^{(K)}(x))_{x \in S^K}.$$

Definición 2.0.3 Sea $n \in \mathbb{N}$ y $S \subseteq \mathbb{N}^*$ tal que $S = \mathbb{N}^*_{\leq m} = \{s \in \mathbb{N}^* : s \leq m\}$, con $m \in \mathbb{N}^*$, y sea $|S|$ es la cardinalidad de S . Definimos la función $\Phi_{|S|,n} : S^n \rightarrow \mathbb{N}^*$ como

$$\Phi_{|S|,n}(x_1, \dots, x_n) = \sum_{i=0}^{n-1} x_{i+1} |S|^i = \sum_{i=0}^{n-1} x_{i+1} (m+1)^i.$$

Observación 2.0.2 Dados $n \in \mathbb{N}$ y $S = \mathbb{N}^*_{\leq m}$, $m \in \mathbb{N}^*$, sea $\Phi_{|S|,n}$ como en la definición 2.0.3. Entonces tenemos que $\text{Im}(\Phi_{|S|,n}) = \{0, \dots, |S|^n - 1\}$ y, por ende, $|\text{Im}(\Phi_{|S|,n})| = |S|^n$. Además, es fácil ver que $\Phi_{|S|,n}$ es inyectiva.

Sea $W = \{W_t : t \in \mathbb{N}\}$ un proceso con espacio de estados S y gobernado por una cadena de Markov $X^{(K)} = \{X_t^{(K)} : t \in \mathbb{N}\}$ no homogénea de orden K cuyo espacio de estados es $S_3^{(K)} = \{(y_1, y_2, \dots, y_K) \in (S)^K\}$. Note que podemos ver a los elementos de $S_3^{(K)}$ como elementos del conjunto $S_4^{(K)} = \{0, 1, \dots, |S|^K - 1\}$ por medio de la transformación $\Phi_{|S|,K}(\cdot)$. Cuando $\mathbf{x} = (y_1, y_2, \dots, y_K) \in S_3^{(K)}$ y $\bar{m} \in S_4^{(K)}$ son tales que $\Phi_{|S|,K}(\mathbf{x}) = \bar{m}$, también utilizaremos la notación $\mathbf{x} \leftrightarrow \bar{m}$ para indicar su respectiva correspondencia. Notemos que si $K = 1$, tenemos $S_4^{(1)} = S_3^{(1)}$. Cuando $K = 0$, también tenemos $S_4^{(0)} = S_3^{(0)}$.

Notación 2.0.1 A partir de este momento, vamos a tomar $\mathcal{S}^* := \mathcal{S} \cup \{0\}$, con $\mathcal{S} = \{1, \dots, M\}$, $M \in \mathbb{N}$. Asimismo, al momento de hablar de cadenas de Markov W de cualquier orden K , su espacio de estados estará denotado por S , a menos que se especifique lo contrario.

Ahora, vayamos con las probabilidades de transición de $X^{(K)} = \{X_t^{(K)} : t \in \mathbb{N}\}$ en el caso en que $K \in \mathcal{S}$.

Definición 2.0.4 Dados $x_t, \dots, x_{t+K-1}, x_{t+K} \in S$, tenemos que la probabilidad de ir del vector de estados $\mathbf{x} = (x_t, \dots, x_{t+K-1})$ al vector de estados $(y_{t+1}, \dots, y_{t+K}) \leftrightarrow \mathbf{y} \in S_4^{(K)}$ en el tiempo $t+1$ se denota de la siguiente forma:

$$P_{\mathbf{xy}}^{(K)}(t) = P\left(X_{t+1}^{(K)} = \mathbf{y} \mid X_t^{(K)} = \mathbf{x} = (x_t, \dots, x_{t+K-1})\right). \quad (2.1)$$

La probabilidad (2.1) resulta distinta de cero si y sólo si $y_i = x_i$, $i = t+1, \dots, t+K-1$. Por lo tanto, \mathbf{y} ocurre si y sólo si la observación que sigue a x_t, \dots, x_{t+K-1} es y_{t+K} . Es decir, de $X_t^{(K)} = (x_t, \dots, x_{t+K-1})$ se sigue que $W_{t+K} = y_{t+K}$. Así, teniendo $K \in \mathcal{S}$, la probabilidad (2.1) se puede reescribir de la siguiente forma:

$$P\left(W_{t+K} = y_{t+K} \mid X_t^{(K)} = (x_t, \dots, x_{t+K-1}) \leftrightarrow \bar{m}\right), \quad (2.2)$$

donde \bar{m} es el escalar en $S_4^{(K)}$ correspondiente al vector (x_t, \dots, x_{t+K-1}) en $S_3^{(K)}$. Así, procedemos a dar la siguiente definición.

Definición 2.0.5 Sea $K \in \mathcal{S}$ y $t \in \mathbb{N}$. Dados $j \in S$ y $\bar{m} \in S_4^{(K)}$, vamos a denotar por $P_{\bar{m}j}^{(K)}(t)$ a la probabilidad de transición correspondiente a j, \bar{m}, K y t . Es decir,

$$P_{\bar{m}j}^{(K)}(t) := P\left(W_{t+K} = j \mid X_t^{(K)} = \bar{m}\right).$$

De igual forma nos será de gran ayuda establecer lo siguiente.

Definición 2.0.6 Sea $K \in \mathcal{S}^*$. Para $K > 0$, $t \in \mathbb{N}$ y $\bar{m} \in S_4^{(K)}$, vamos a denotar por $Q_{\bar{m}}^{(K)}(t)$ a la probabilidad $P(X_t^{(K)} = \bar{m}) = \pi_t^{(K)}(\bar{m})$. En el caso $K = 0$, $t \in \mathbb{N}$ y $j \in S$ vamos a denotar por $Q_j^{(0)}(t)$ a la probabilidad $P(W_t = j) = \pi_t^{(0)}(j)$.

Observación 2.0.3 Observemos que $\pi_t^{(0)}(j) = \pi_t^{(1)}(j)$, $j \in S, t \in \mathbb{N}$.

Ya con estas notaciones establecidas, las matrices de probabilidades de transición, así como la distribución inicial y los vectores $\pi^{(n)}$, de nuestra cadena $X^{(K)}$ se pueden expresar de una manera más económica y cómoda.

Notación 2.0.2 Sean $K \in \mathcal{S}$ y $t \in \mathbb{N}$. Entonces, vamos a denotar por $P^{(K)}(t)$ a la matriz de probabilidades de transición $\left(P_{\bar{m}j}^{(K)}(t)\right)_{\bar{m} \in S_4^{(K)}, j \in S}$.

Notación 2.0.3 Sean $K \in \mathcal{S}^*$ y $t \in \mathbb{N}$. Entonces, vamos a denotar por $Q^{(K)}(t)$ al vector de probabilidades $\left(Q_{\bar{m}}^{(K)}(t)\right)_{\bar{m} \in S_4^{(K)}}$ donde $Q_{\bar{m}}^{(K)}(t) = P(X_t^{(K)} = \bar{m})$.

Observación 2.0.4 Usando (2.2) y dados $x_1, \dots, x_n, x_{n+1} \in S, \bar{m}_1, \dots, \bar{m}_{n-K+2} \in S_3^{(K)}$, con $\bar{m}_t = (x_t, \dots, x_{t+K-1}) \in S^K$, tenemos que

$$\begin{aligned}
P(W_{n+1} = x_{n+1} \mid W_n = x_n, \dots, W_1 = x_1) &= \frac{P(W_{n+1} = x_{n+1}, W_n = x_n, \dots, W_1 = x_1)}{P(W_n = x_n, \dots, W_1 = x_1)} \\
&= \frac{P(X_{n-K+2}^{(K)} = \bar{m}_{n-K+2}, \dots, X_1^{(K)} = \bar{m}_1)}{P(X_{n-K+1}^{(K)} = \bar{m}_{n-K+1}, \dots, X_1^{(K)} = \bar{m}_1)} \\
&= P(X_{n-K+2}^{(K)} = \bar{m}_{n-K+2} \mid X_{n-K+1}^{(K)} = \bar{m}_{n-K+1}, \dots, X_1^{(K)} = \bar{m}_1) \\
(\text{Propiedad de Markov de orden } K) &= P(X_{n-K+2}^{(K)} = \bar{m}_{n-K+2} \mid X_{n-K+1}^{(K)} = \bar{m}_{n-K+1}) \\
&= P(W_{n+1} = x_{n+1} \mid X_{n-K+1}^{(K)} = \bar{m}_{n-K+1}) \\
&= P(W_{n+1} = x_{n+1} \mid W_n = x_n, \dots, W_{n-K+1} = x_{n-K+1}).
\end{aligned}$$

Es decir, de la propiedad de la definición (2.0.1) se tiene que

$$P(W_{n+1} = x_{n+1} \mid W_n = x_n, \dots, W_1 = x_1) = P(W_{n+1} = x_{n+1} \mid W_n = x_n, \dots, W_{n-K+1} = x_{n-K+1}).$$

2.0.1. Distribución de W_t

En esta sección mostramos cómo obtener las distribuciones de $\mathbf{W} = \{W_t : t \in \mathbb{N}\}$ al tiempo t , $P(W_t = \cdot)$, $t \geq 1$. Todo esto lo haremos a partir de los renglones de las matrices de probabilidades de transición $P_{\overline{m}j}^{(K)}(t)$ y la distribución inicial $Q^{(K)}(1)$. El método dependerá del orden K correspondiente a la cadena $X^{(K)}$.

Distribución de W_t al tiempo t con $K = 0$

Si $K = 0$, tenemos que $P_{\overline{m}}^{(0)}(t) = Q_{\overline{m}}^{(0)}(t) = P(W_t = \overline{m})$, $t \geq 1$, $\overline{m} \in S_4^{(K)} = S$, por lo que la distribución final de la cadena W al tiempo t viene siendo la misma que aquella de $Q^{(0)}(t)$.

Distribución de W_t al tiempo t con $K = 1$

Para el caso $K = 1$, se procede a realizar lo siguiente. Dados $t \geq 2$, $i \in S$, ocurre que

$$P(W_t = i) = \sum_{j \in S} P(W_t = i, W_{t-1} = j) = \sum_{j \in S} P(W_t = i | W_{t-1} = j) P(W_{t-1} = j), \quad (2.3)$$

donde al tiempo $t = 1$ tenemos $P(W_1 = j) = Q_j^{(K)}(1)$. De esta forma observamos que podemos conseguir las distribuciones buscadas de manera recursiva.

Distribución de W_t al tiempo t con $K > 1$

Por el momento, coloquémonos en el escenario en que $K = 2$. Dado $i \in S$ y $t \geq 1$, para obtener $P(W_t = i)$ tenemos los siguientes casos:

- Caso 1: Para $t = 1, 2$, consideramos

$$P(W_t = i) = \begin{cases} \sum_{j \in S} P(W_t = i, W_{t+1} = j) = \sum_{\{\overline{m}=(i,j) : j \in S\}} Q_{\overline{m}}^{(K)}(1) & \text{si } t = 1, \\ \sum_{j \in S} P(W_{t-1} = j, W_t = i) = \sum_{\{\overline{m}=(j,i) : j \in S\}} Q_{\overline{m}}^{(K)}(1) & \text{si } t = 2. \end{cases}$$

- Caso 2: Para $t > 2$, se hará lo siguiente:

$$\begin{aligned}
P(W_t = i) &= \sum_{j,z \in S} P(W_t = i, W_{t-1} = j, W_{t-2} = z) \\
&= \sum_{j,z \in S} P(W_t = i \mid W_{t-1} = j, W_{t-2} = z) P(W_{t-1} = j, W_{t-2} = z) \\
&= \sum_{j,z \in S} P_{(j,z)i}^{(K)}(t-2) Q_{(j,z)}^{(K)}(t-2) \\
&= \sum_{\bar{m} \in S_4^{(K)}} P_{\bar{m}i}^{(K)}(t-2) Q_{\bar{m}}^{(K)}(t-2),
\end{aligned}$$

donde $\bar{m} = (j, z)$.

Dado que sólo conocemos las distribuciones de $P_{\bar{m}j}^{(K)}(t)$ y $Q_{\bar{m}}^{(K)}(1)$, hará falta obtener los valores para $Q_{\bar{m}}^{(K)}(t)$, $t \geq 2$. Esto lo podemos calcular de manera recursiva. Tomando $\bar{m} = (i, j)$ y $t \geq 2$ tenemos que

$$\begin{aligned}
Q_{\bar{m}}^{(K)}(t) &= P(W_{t+1} = j, W_t = i) \\
&= \sum_{z \in S} P(W_{t+1} = j, W_t = i, W_{t-1} = z) \\
&= \sum_{z \in S} P(W_{t+1} = j \mid W_t = i, W_{t-1} = z) P(W_t = i, W_{t-1} = z) \\
&= \sum_{z \in S} P_{(z,i)j}^{(K)}(t-1) Q_{(z,i)j}^{(K)}(t-1).
\end{aligned}$$

De manera general, tenemos que, dado $K \in \mathcal{S}$ con $K > 1$, para $t \in \{1, \dots, K\}$,

$$\begin{aligned}
P(W_t = i) &= \sum_{\substack{w_j \in S, 1 \leq j \leq K \\ j \neq t}} P(W_1 = w_1, \dots, W_{t-1} = w_{t-1}, W_t = i, W_{t+1} = w_{t+1}, \dots, W_K = w_K) \\
&= \sum_{\substack{\bar{m} = (w_1, \dots, w_{t-1}, i, w_{t+1}, \dots, w_K) \\ w_j \in S, 1 \leq j \leq K \\ j \neq t}} Q_{\bar{m}}^{(K)}(1),
\end{aligned}$$

mientras que para $t \geq K + 1$,

$$P(W_t = i) = \sum_{\bar{m} \in S_4^{(K)}} P(W_t = i, X_{t-K}^{(K)} = \bar{m})$$

$$\begin{aligned}
&= \sum_{\bar{m} \in S_4^{(K)}} P(W_t = i \mid X_{t-K}^{(K)} = \bar{m}) P(X_{t-K}^{(K)} = \bar{m}) \\
&= \sum_{\bar{m} \in S_4^{(K)}} P_{\bar{m}i}^{(K)}(t-K) Q_{\bar{m}}^{(K)}(t-K).
\end{aligned}$$

Para obtener $Q_{\bar{m}}^{(K)}(t)$, $t \geq 2$, con $\bar{m} = (w_t \dots, w_{t+K-2}, j)$, $\bar{m} \in S_4^{(K)}$, tenemos que

$$\begin{aligned}
Q_{\bar{m}}^{(K)}(t) &= P(X_t^{(K)} = \bar{m}) \\
&= P(W_t = w_t, \dots, W_{t+K-2} = w_{t+K-2}, W_{t+K-1} = j) \\
&= \sum_{z \in S} P(W_{t-1} = z, W_t = w_t, \dots, W_{t+K-2} = w_{t+K-2}, W_{t+K-1} = j) \\
&= \sum_{\substack{\bar{w}=(z, w_t, \dots, w_{t+K-2}) \\ z \in S}} P(W_{t+K-1} = j, X_{t-1}^{(K)} = \bar{w}) \\
&= \sum_{\substack{\bar{w}=(z, w_t, \dots, w_{t+K-2}) \\ z \in S}} P(W_{t+K-1} = j \mid X_{t-1}^{(K)} = \bar{w}) P(X_{t-1}^{(K)} = \bar{w}) \\
&= \sum_{\substack{\bar{w}=(z, w_t, \dots, w_{t+K-2}) \\ z \in S}} P_{\bar{w}j}^{(K)}(t-1) Q_{\bar{w}}^{(K)}(t-1) \\
&= \left([P^{(K)}(t-1)]_j^* \right)^\top \cdot [Q^{(K)}(t-1)]^*,
\end{aligned}$$

donde $\left([P^{(K)}(t-1)]_j^* \right)^\top$ es la transpuesta de la columna j correspondiente a la matriz $P^{(K)}(t-1)$ cuyas entradas estén asociadas a vectores \bar{w} de la forma $\bar{w} = (z, \dots, w_{t+K-2})$, $z \in S$, y $[Q^{(K)}(t-1)]^*$ son las entradas del vector $Q^{(K)}(t-1)$ filtradas por el mismo criterio mencionado anteriormente.

Para terminar con esta sección, demostraremos lo siguiente (Fan y Tsai, 1999).

Proposición 2.0.1 *Sea $K \in \mathbb{N}$, sea $\mathbf{W} = \{W_t : t \in \mathbb{N}\}$ una cadena de Markov de orden K , sea $n > K$ y sean $x_1, \dots, x_n \in S$. Entonces, se satisface la siguiente igualdad:*

$$P(W_1 = x_1, \dots, W_n = x_n) = P(X_1^{(K)} = \bar{m}_1) \prod_{t=1}^{n-K} P(W_{t+K} = x_{t+K} \mid X_t^{(K)} = \bar{m}_t). \quad (2.4)$$

donde $\bar{m}_t = (x_t, \dots, x_{t+K-1})$, $t = 1, \dots, n-K$.

Demostración. Procederemos a demostrar la proposición usando inducción.

PASO BASE. Para $n = K + 1$ tenemos que,

$$\begin{aligned} P(W_1 = x_1, \dots, W_{K+1} = x_{K+1} = \bar{x}_{K+1}) &= P(W_{K+1} = x_{K+1} \mid X_1^{(K)} = \bar{m}_1) P(X_1^{(K)} = \bar{m}_1) \\ &= P(X_1^{(K)} = \bar{m}_1) \prod_{t=1}^1 P(W_{K+1} = x_{K+1} \mid X_t^{(K)} = \bar{m}_t) \\ &= P(X_1^{(K)} = \bar{m}_1) \prod_{t=1}^{n-K} P(W_{K+1} = x_{K+1} \mid X_t^{(K)} = \bar{m}_t) \end{aligned}$$

PASO INDUCTIVO. Supongamos que la propiedad se cumple para n , con $n > K$,

$$P(W_1 = w_1, \dots, W_n = x_n) = P(X_1^{(K)} = \bar{m}_1) \prod_{t=1}^{n-K} P(W_{K+1} = x_{K+1} \mid X_t^{(K)} = \bar{m}_t).$$

Entonces, para $n + 1$ se tiene

$$\begin{aligned} P(W_1 = x_1, \dots, W_{n+1} = x_{n+1}) &= P(W_{n+1} = x_{n+1} \mid W_n = x_n, \dots, W_1 = x_1) \\ &\quad \times P(W_1 = x_1, \dots, W_n = x_n) \\ (\text{Propiedad de Markov de orden } K) &= P(W_{n+1} = x_{n+1} \mid X_{n+1-K}^{(K)} = \bar{m}_{n+1-K}) \\ &\quad \times P(W_1 = x_1, \dots, W_n = x_n) \\ (\text{Hipótesis de inducción}) &= P(W_{n+1} = x_{n+1} \mid X_{n+1-K}^{(K)} = \bar{m}_{n+1-K}) \\ &\quad \times P(X_1^{(K)} = \bar{m}_1) \prod_{t=1}^{n-K} P(W_{t+K} = x_{t+K} \mid X_t^{(K)} = \bar{m}_t) \\ &= P(X_1^{(K)} = \bar{m}_1) \prod_{t=1}^{(n+1)-K} P(W_{t+K} = x_{t+K} \mid X_t = \bar{m}_t), \end{aligned}$$

y por lo tanto

$$P(W_1 = x_1, \dots, W_{n+1} = w_{n+1}) = P(X_1^{(K)} = \bar{m}_1) \prod_{t=1}^{(n+1)-K} P(W_{t+K} = x_{t+K} \mid X_t = \bar{m}_t).$$

De esta forma, podemos concluir que esta propiedad se cumple para cualquier $n > K$. ■

Capítulo 3

Inferencia bayesiana para un modelo Markoviano de orden K

En este capítulo presentaremos el modelo que vamos a utilizar para el estudio de contaminantes atmosféricos, notación concerniente a éste y, dado que realizaremos inferencia bayesiana, las distribuciones iniciales de los parámetros a estimar y la función de verosimilitud de los datos a estudiar, así como las distribuciones a posteriori correspondientes. Aquí, también procedemos como en Rodrigues *et al.* (2015).

3.1. Modelo Markoviano

En un modelo de Markov de orden K el objetivo es el de realizar inferencia tanto con respecto al orden K de la cadena $X^{(K)}$ como de las probabilidades de transición y distribuciones iniciales $Q^{(K)}(1)$, $K \in \mathcal{S}$, además de las probabilidades $Q^{(0)}(t)$'s si $K = 0$.

A partir de este momento, salvo se indique lo contrario, estaremos considerando una cadena de Markov que evoluciona hasta un tiempo definido y fijo. Indicaremos este tiempo por T . De esta forma, la cadena de Markov estará definida para los tiempos $1, 2, \dots, T$. Por lo tanto cuando hagamos referencia a las matrices de probabilidades de transición, sólo se estarán contemplando los primeros $T - K$ tiempos, donde $K \in \mathcal{S}^*$, dado que la

transición al tiempo T depende del estado de la cadena en estos tiempos.

Las siguientes definiciones nos permitirán plantear de forma concreta el vector de parámetros a estimar en nuestro modelo.

Definición 3.1.1 Sea $A \in \mathbb{M}_{m \times n}(\mathbb{R})$, con $A = (a_{ij})$. Vamos a denotar por $\text{vec } A$ al vector de los renglones de la matriz A , es decir, $\text{vec } A = (a_{11}, \dots, a_{1n}, a_{21}, \dots, a_{2n}, \dots, a_{m1}, \dots, a_{mn})$.

Definición 3.1.2 Sea $K \in \mathcal{S}^*$. Para $K > 0$, vamos a definir a $P^{(K)}$ como el vector de los renglones de las matrices de probabilidades de transición:

$$P^{(K)} := (\text{vec } P^{(K)}(1), \dots, \text{vec } P^{(K)}(T - K)).$$

En el caso $K = 0$, vamos a denotar por $Q^{(0)}$ al vector de los $Q^{(0)}(t)$'s,

$$Q^{(0)} := (Q^{(0)}(1), \dots, Q^{(0)}(T)).$$

Definición 3.1.3 Sea $K \in \mathcal{S}$. Dado $t \in \mathbb{N}$ y $\bar{m} \in S_4^{(K)}$, vamos a denotar por $P_{\bar{m}}^{(K)}(t)$ al renglón correspondiente a \bar{m} en la matriz $P^{(K)}(t)$, es decir, $P_{\bar{m}}^{(K)}(t) := (P_{\bar{m}j}^{(K)}(t))_{j \in \mathcal{S}}$.

Definición 3.1.4 Sea $K \in \mathcal{S}^*$. Para $K > 0$ vamos a definir $\theta^{(K)} := (Q^{(K)}(1), P^{(K)})$, mientras que para $K = 0$, $\theta^{(0)} := Q^{(0)}$.

De esta forma, tenemos que el vector de parámetros a estimar es $\theta = (K, \theta^{(K)})$. Como la estimación de los parámetros será realizada adoptando el punto de vista bayesiano, necesitamos especificar la distribución a priori de θ y también la función de verosimilitud del modelo.

3.2. Función de verosimilitud

Supongamos que $K = k$. Regresemos con nuestra cadena $X^{(k)}$ y ubiquémonos en el escenario donde $k \in \mathcal{S}$. Por la proposición 2.0.1, dados $x_1, x_2, \dots, x_T \in S$, tenemos

$$P(W_1 = x_1, \dots, W_T = x_T) = P(X_1^{(k)} = \bar{m}_1) \prod_{t=1}^{T-k} P(W_{t+k} = x_{t+k} \mid X_t^{(k)} = \bar{m}_t), \quad (3.1)$$

donde para cada $t \in \{1, \dots, T - k\}$, $\bar{m}_t = (x_t, \dots, x_{t+k-1})$. La probabilidad (3.1) puede expresarse en términos de la distribución inicial $Q^{(k)}(1)$ y las matrices de probabilidades de transición de la cadena de Markov. Para hacer esto, nos serviremos de la siguiente definición.

Definición 3.2.1 *Sea \underline{x} un conjunto de observaciones de nuestra cadena W , con $k \in \mathcal{S}^*$. Dados $j \in S$, $\bar{m}, \bar{w} \in S_4^{(K)}$ y $t \in \mathbb{N}$,*

- *Para $k > 0$, definimos a $n_{\bar{m}j}^{(k)}(t)$ como el número de observaciones tales que los valores correspondientes al vector $(X_t^{(k)}, W_{t+k})$ son (\bar{m}, j) , y $n_{\bar{w}}^{(k)}$ hará alusión al número de observaciones cuyo valor correspondiente a $X_1^{(k)}$ sea \bar{w} .*
- *Para el caso $k = 0$, definimos a $n_j^{(0)}(t)$ como el número de observaciones en las que el valor obtenido en W_t es igual a j .*

Esto nos permitirá poder expresar a la probabilidad $P(W_1 = x_1, \dots, W_T = x_T)$ en términos de los renglones de las matrices de probabilidades de transición y de los valores que puede tomar la distribución inicial. Para dejar más claro esto, nos apoyaremos en el siguiente ejemplo.

Ejemplo 3.2.1 *Sea $\underline{y} = (\bar{y}_1, \dots, \bar{y}_6)$ un conjunto de seis muestras independientes de los primeros cuatro tiempos de un proceso $W = \{W_t : t \in \mathbb{N}\}$ con espacio de estados $S = \{0, 1\}$ y la cual está regida por una cadena de Markov $X^{(k)}$ de orden k . Supongamos que*

$$\begin{aligned} \bar{y}_1 &= (1, 0, 1, 0), & \bar{y}_2 &= (0, 1, 1, 0), & \bar{y}_3 &= (0, 0, 0, 1), \\ \bar{y}_4 &= (0, 1, 0, 0), & \bar{y}_5 &= (1, 0, 0, 1), & \bar{y}_6 &= (1, 1, 0, 0). \end{aligned}$$

Si el orden de X fuera $k = 2$, dado que $|S| = 2$, tenemos que $|S_4^{(K)}| = |S|^2 = 2^2 = 4$. Consideremos al vector inicial

$$Q^{(2)}(1) = \left(Q_{(0,0)}^{(2)}(1), Q_{(0,1)}^{(2)}(1), Q_{(1,0)}^{(2)}(1), Q_{(1,1)}^{(2)}(1) \right) \in \Delta_{|S_4^{(K)}|} = \Delta_4$$

y las matrices de probabilidades de transición $P^{(2)}(t)$, $t = 1, 2$, tales que

$$P_{\bar{m}}^{(2)}(t) = \left(P_{\bar{m}0}^{(2)}(t), P_{\bar{m}1}^{(2)}(t) \right) \in \Delta_{|S|} = \Delta_2,$$

con $\bar{m} \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}$. Calculando los valores $n_{\bar{m}}^{(2)}(t)$ tenemos que

(ii) para $t = 1$,

$$\begin{aligned} n_{(0,0)0}^{(2)}(1) &= 1, & n_{(0,1)0}^{(2)}(1) &= 1, & n_{(1,0)0}^{(2)}(1) &= 1, & n_{(1,1)0}^{(2)}(1) &= 1, \\ n_{(0,0)1}^{(2)}(1) &= 0, & n_{(0,1)1}^{(2)}(1) &= 1, & n_{(1,0)1}^{(2)}(1) &= 1, & n_{(1,1)1}^{(2)}(1) &= 0, \end{aligned}$$

(ii) mientras que para $t = 2$,

$$\begin{aligned} n_{(0,0)0}^{(2)}(2) &= 0, & n_{(0,1)0}^{(2)}(2) &= 1, & n_{(1,0)0}^{(2)}(2) &= 2, & n_{(1,1)0}^{(2)}(2) &= 1, \\ n_{(0,0)1}^{(2)}(2) &= 2, & n_{(0,1)1}^{(2)}(2) &= 0, & n_{(1,0)1}^{(2)}(2) &= 0, & n_{(1,1)1}^{(2)}(2) &= 0. \end{aligned}$$

En cuanto a las $n_{\bar{m}}^{(k)}$'s,

$$n_{(0,0)}^{(2)} = 1, \quad n_{(0,1)}^{(2)} = 2, \quad n_{(1,0)}^{(2)} = 2, \quad n_{(1,1)}^{(2)} = 1.$$

Dado que las muestras son independientes,

$$P(\underline{y} | k, Q^{(2)}(1), P^{(2)}) = \prod_{i=1}^6 P(y_i | k, Q^{(2)}(1), P^{(2)})$$

$$\begin{aligned} \text{Prop. 1.1.1} &= \left[Q_{(1,0)}^{(2)}(1) P_{(1,0)1}^{(2)}(1) P_{(0,1)0}^{(2)}(2) \right] \times \left[Q_{(0,1)}^{(2)}(1) P_{(0,1)1}^{(2)}(1) P_{(1,1)0}^{(2)}(2) \right] \\ &\quad \times \left[Q_{(0,0)}^{(2)}(1) P_{(0,0)0}^{(2)}(1) P_{(0,0)1}^{(2)}(2) \right] \times \left[Q_{(0,1)}^{(2)}(1) P_{(0,1)0}^{(2)}(1) P_{(1,0)0}^{(2)}(2) \right] \\ &\quad \times \left[Q_{(1,0)}^{(2)}(1) P_{(1,0)0}^{(2)}(1) P_{(0,0)1}^{(2)}(2) \right] \times \left[Q_{(1,1)}^{(2)}(1) P_{(1,1)0}^{(2)}(1) P_{(1,0)0}^{(2)}(2) \right] \\ &= \left[Q_{(0,0)}^{(2)}(1) \right]^{n_{(0,0)}^{(2)}} \left[Q_{(0,1)}^{(2)}(1) \right]^{n_{(0,1)}^{(2)}} \left[Q_{(1,0)}^{(2)}(1) \right]^{n_{(1,0)}^{(2)}} \left[Q_{(1,1)}^{(2)}(1) \right]^{n_{(1,1)}^{(2)}} \\ &\quad \times \left[P_{(0,0)0}^{(2)}(1) \right]^{n_{(0,0)0}^{(2)}(1)} \left[P_{(0,1)0}^{(1)}(1) \right]^{n_{(0,1)0}^{(2)}(1)} \\ &\quad \times \left[P_{(1,0)0}^{(1)}(1) \right]^{n_{(1,0)0}^{(2)}(1)} \left[P_{(1,1)0}^{(1)}(1) \right]^{n_{(1,1)0}^{(2)}(1)} \end{aligned}$$

$$\begin{aligned}
& \times [P_{(0,0)1}^{(1)}(1)]^{n_{(0,0)1}^{(2)}(1)} [P_{(0,1)1}^{(1)}(1)]^{n_{(0,1)1}^{(2)}(1)} \\
& \times [P_{(1,0)1}^{(1)}(1)]^{n_{(1,0)1}^{(2)}(1)} [P_{(1,1)1}^{(1)}(1)]^{n_{(1,1)1}^{(2)}(1)} \\
& \times [P_{(0,0)0}^{(2)}(2)]^{n_{(0,0)0}^{(2)}(2)} [P_{(0,1)0}^{(1)}(2)]^{n_{(0,1)0}^{(2)}(2)} \\
& \times [P_{(1,0)0}^{(1)}(2)]^{n_{(1,0)0}^{(2)}(2)} [P_{(1,1)0}^{(1)}(2)]^{n_{(1,1)0}^{(2)}(2)} \\
& \times [P_{(0,0)1}^{(1)}(2)]^{n_{(0,0)1}^{(2)}(2)} [P_{(0,1)1}^{(1)}(2)]^{n_{(0,1)1}^{(2)}(2)} \\
& \times [P_{(1,0)1}^{(1)}(2)]^{n_{(1,0)1}^{(2)}(2)} [P_{(1,1)1}^{(1)}(2)]^{n_{(1,1)1}^{(2)}(2)} \\
& = \prod_{(x,y) \in S_3^{(K)}} \left[[Q_{(x,y)}^{(2)}(1)]^{n_{(x,y)}^{(2)}} \prod_{t=1}^2 \prod_{j \in S} [P_{(x,y)j}^{(2)}(t)]^{n_{(x,y)j}^{(2)}(t)} \right] = L(\underline{y}|k, \theta^{(k)}).
\end{aligned}$$

Ahora, si el orden de la cadena fuera $k = 0$, entonces vamos a considerar a los vectores $Q^{(0)}(t) = (Q_0^{(0)}(t), Q_1^{(0)}(t)) \in \Delta_{|S|} = \Delta_2$, $t = 1, \dots, 4$. Calculando los conteos $n_j^{(0)}(t)$, $j = 0, 1$, tenemos que

(i) para $t = 1$: $n_0^{(0)}(1) = 3, n_1^{(0)}(1) = 3$,

(ii) para $t = 2$: $n_0^{(0)}(2) = 3, n_1^{(0)}(2) = 3$,

(iii) para $t = 3$: $n_0^{(0)}(3) = 4, n_1^{(0)}(3) = 2$,

(iv) para $t = 4$: $n_0^{(0)}(4) = 4, n_1^{(0)}(4) = 2$,

y entonces

$$\begin{aligned}
P(\underline{y}|k=0, Q^{(0)}) &= \prod_{i=1}^6 P(y_i|k=0, Q^{(0)}) \\
&= [Q_1^{(0)}(1) Q_0^{(0)}(2) Q_1^{(0)}(3) Q_0^{(0)}(4)] \\
&\quad \times [Q_0^{(0)}(1) Q_1^{(0)}(2) Q_1^{(0)}(3) Q_0^{(0)}(4)] \\
&\quad \times [Q_0^{(0)}(1) Q_0^{(0)}(2) Q_0^{(0)}(3) Q_1^{(0)}(4)] \\
&\quad \times [Q_0^{(0)}(1) Q_1^{(0)}(2) Q_0^{(0)}(3) Q_0^{(0)}(4)] \\
&\quad \times [Q_1^{(0)}(1) Q_0^{(0)}(2) Q_0^{(0)}(3) Q_1^{(0)}(4)]
\end{aligned}$$

$$\begin{aligned}
& \times [Q_1^{(0)}(1) Q_1^{(0)}(2) Q_0^{(0)}(3) Q_0^{(0)}(4)] \\
& = [Q_0^{(0)}(1)]^{n_0^{(0)}(1)} [Q_1^{(0)}(1)]^{n_1^{(0)}(1)} [Q_0^{(0)}(2)]^{n_0^{(0)}(2)} [Q_1^{(0)}(2)]^{n_1^{(0)}(2)} \\
& \quad \times [Q_0^{(0)}(3)]^{n_0^{(0)}(3)} [Q_1^{(0)}(3)]^{n_1^{(0)}(3)} [Q_0^{(0)}(4)]^{n_0^{(0)}(4)} [Q_1^{(0)}(4)]^{n_1^{(0)}(4)} \\
& = \prod_{t=1}^4 \prod_{j \in S} [Q_j^{(0)}(t)]^{n_j^{(0)}(t)} \\
& = L(\underline{y} | k = 0, \theta^{(0)}).
\end{aligned}$$

Así, basándonos en el ejemplo 3.2.1, podemos reescribir (3.1) de la siguiente forma:

$$P(W_1 = x_1, \dots, W_T = x_T) = \prod_{\bar{m} \in S_4^{(K)}} \left[[Q_{\bar{m}}^{(k)}(1)]^{n_{\bar{m}}^{(k)}} \prod_{t=1}^{T-k} \prod_{j \in S} [P_{\bar{m}j}^{(k)}(t)]^{n_{\bar{m}j}^{(k)}(t)} \right],$$

donde $t \in \{1, \dots, T - k\}$, $\bar{m} \in S_4^{(K)}$, $j \in S$. De esta forma, la función de verosimilitud del modelo está dada por los siguientes casos:

(i) Para $k > 0$, la función de verosimilitud de \underline{y} estará dada por

$$L(\underline{y} | k, \theta^{(k)}) = \prod_{\bar{m} \in S_4^{(K)}} \left[(Q_{\bar{m}}^{(k)}(1))^{n_{\bar{m}}^{(k)}} \prod_{t=1}^{T-k} \prod_{j \in S} [P_{\bar{m}j}^{(k)}(t)]^{n_{\bar{m}j}^{(k)}(t)} \right].$$

(ii) En el caso $k = 0$, tomaremos

$$L(\underline{y} | k = 0, \theta^{(0)}) = \prod_{t=1}^T \prod_{j \in S} [Q_j^{(0)}(t)]^{n_j^{(0)}(t)}.$$

3.3. Distribuciones a priori

Para el orden de la cadena, K , consideraremos la distribución de una Poisson truncada con parámetro $\lambda > 0$:

$$P(K = k) \propto \frac{\lambda^k}{k!} \mathbb{I}_{\mathcal{S}^*}(k),$$

donde $\mathbb{I}_{\mathcal{S}^*}(\cdot)$ es la función indicadora del conjunto \mathcal{S}^* . Es decir, $\mathbb{I}_{\mathcal{S}^*}(x) = 1$ si $x \in \mathcal{S}^*$ y es cero en otro caso. Sea $K \in \mathcal{S}^*$. Para $K > 0$, asumiremos una independencia a priori de las matrices $P^{(K)}(t)$ y la distribución inicial $Q^{(K)}(1)$ condicionada a K , es decir,

$$P(\theta^{(K)}) = P(Q^{(K)}(1)|K) \prod_{t=1}^{T-K} P(\text{vec } P^{(K)}(t)|K) P(K).$$

A su vez, para cada $t \in \{1, \dots, T - K\}$ vamos a asumir que los renglones de la matriz $P^{(K)}(t)$ son independientes entre sí, y a cada uno de estos le vamos a asignar una distribución Dirichlet con vector de hiperparámetros $(\alpha_{\bar{m}j}^{(K)}(t))_{j \in \mathcal{S}}$, es decir, para cada $t \in \{1, \dots, T - K\}$ y para cada $\bar{m} \in S_4^{(K)}$,

$$P_{\bar{m}}^{(k)}(t) \sim \text{Dirichlet} \left((\alpha_{\bar{m}j}^{(k)}(t))_{j \in \mathcal{S}} \right),$$

por lo que para cada $t \in \{1, \dots, T - K\}$

$$P(\text{vec } P^{(K)}(t)|K) = \prod_{\bar{m} \in S_4^{(K)}} \left[\frac{\Gamma(\sum_{j \in \mathcal{S}} \alpha_{\bar{m}j}^{(K)}(t))}{\prod_{j \in \mathcal{S}} \Gamma(\alpha_{\bar{m}j}^{(K)}(t))} \prod_{j \in \mathcal{S}} [P_{\bar{m}j}^{(K)}(t)]^{\alpha_{\bar{m}j}^{(K)}(t)-1} \right].$$

A $Q^{(K)}(1)$ le asignaremos una distribución Dirichlet con parámetros $(\alpha_{\bar{m}}^{(K)})_{\bar{m} \in S_4^{(K)}}$,

$$P(Q^{(K)}(1)|K) = \frac{\Gamma(\sum_{\bar{m} \in S_4^{(K)}} \alpha_{\bar{m}}^{(K)})}{\prod_{\bar{m} \in S_4^{(K)}} \Gamma(\alpha_{\bar{m}}^{(K)})} \prod_{\bar{m} \in S_4^{(K)}} [Q_{\bar{m}}^{(K)}(1)]^{\alpha_{\bar{m}}^{(K)}-1}.$$

En el caso $K = 0$ se asumirá lo siguiente:

$$P(\theta^{(0)}) = \prod_{t=1}^T P(Q^{(0)}(t)|K) P(K),$$

y para cada $t \in \{1, \dots, T\}$, consideraremos

$$Q^{(0)}(t) \sim \text{Dirichlet} \left((\alpha_j^{(0)}(t))_{j \in \mathcal{S}} \right).$$

Es decir,

$$P(Q^{(0)}(t)|K=0) = \frac{\Gamma(\sum_{j \in \mathcal{S}} \alpha_j^{(0)}(t))}{\prod_{j \in \mathcal{S}} \Gamma(\alpha_j^{(0)}(t))} \prod_{j \in \mathcal{S}} [Q_j^{(0)}(t)]^{\alpha_j^{(0)}(t)-1}.$$

Recordemos que si $X = (X_1, \dots, X_n)$ es un vector aleatorio tal que $X \sim \text{Dirichlet}(\bar{\alpha})$, de acuerdo a (1.1), su soporte está dado por

$$\Delta_n := \left\{ (x_1, \dots, x_n) \in [0, 1]^n : \sum_{j=1}^n x_j = 1 \right\}.$$

Entonces, dadas las distribuciones que asumimos para los renglones de las matrices de probabilidades de transición y distribuciones iniciales, tenemos que:

- Si $K \in \mathcal{S}$, dados $t \in \{1, \dots, T-K\}$ y $\bar{m} \in S_4^{(K)}$, el soporte de $P_{\bar{m}}^{(K)}(t)$ es el conjunto $\Delta_{|S|}$ y aquel de $Q_{\bar{m}}^{(K)}(1)$ está dado por $\Delta_{|S_4^{(K)}|}$, los cuales están definidos de manera análoga a Δ_n pero considerando $n = |S|$ y $n = |S_4^{(K)}|$ respectivamente.
- Si $K = 0$, ocurre que para cada $t \in \{1, \dots, T\}$ el soporte de $Q^{(0)}(t)$ está dado por $\Delta_{|S|}$, donde $\Delta_{|S|}$ está construido de manera análoga a Δ_n considerando $n = |S|$.

Así, para cada $K \in \mathcal{S}$ vamos a definir $\Theta^{(K)} := \Delta_{|S_4^{(K)}|} \times \Delta_{|S|}^{(T-K)|S_4^{(K)}|}$, mientras que para $K = 0$, $\Theta^{(0)} := \Delta_{|S|}^T$. De este modo, tenemos que el espacio muestral de θ es

$$\Theta = \bigcup_{K=0}^M (\{K\} \times \Theta^{(K)}).$$

3.4. Distribuciones a posteriori

Antes de proseguir, vamos a introducir la siguiente notación que nos permitirá simplificar los cálculos a realizarse. Sea $K \in \mathcal{S}^*$, $t \in \{1, \dots, T-K\}$, $j \in S$, $\bar{m} \in S_4^{(K)}$. Entonces,

(i) Para $K > 0$, sean

- $\xi_{\bar{m}}^{(K)} := \alpha_{\bar{m}}^{(K)} + n_{\bar{m}}^{(K)}$
- $\xi_{\bar{m}j}^{(K)}(t) := \alpha_{\bar{m}j}^{(K)}(t) + n_{\bar{m}j}^{(K)}(t)$
- $\Theta_P^{(K)} := \Delta_{|S|}^{(T-K)|S_4^{(K)}|}$

- $P^{(K)}(-i) := (\text{vec } P^{(K)}(1), \dots, \text{vec } P^{(K)}(i-1), \text{vec } P^{(K)}(i+1), \dots, \text{vec } P^{(K)}(T-K))$. Es decir, el vector $P^{(K)}(-i)$ serán los renglones concatenados de cada una de las matrices de probabilidades de transición $P^{(K)}(t)$ con la excepción de los renglones de la matriz de probabilidades de transición $P^{(K)}(i)$.
- Denotaremos al dominio de $(Q^{(K)}(1), P^{(K)}(-i))$ por $\Theta^{(K)}(-i)$
- Denotaremos al dominio de $P^{(K)}(-i)$ por $\Theta_{P^{(K)}(-i)}^{(K)}$
- $P_{-\bar{m}}^{(K)}(t)$ será el vector $\text{vec } P^{(K)}(t)$ sin las entradas correspondientes al renglón \bar{m} de $P^{(K)}(t)$.

(ii) Para $K = 0$, sean

- $\xi_j^{(0)}(t) := \alpha_j^{(0)}(t) + n_j^{(0)}(t)$
- $Q^{(0)}(-i) = (Q^{(0)}(1), \dots, Q^{(0)}(i-1), Q^{(0)}(i+1), \dots, Q^{(0)}(T))$. Es decir, el vector $Q^{(0)}(-i)$ serán las concatenaciones de los vectores $Q^{(0)}(t)$ exceptuando al vector $Q^{(0)}(i)$.
- $\Theta_{Q^{(0)}(-i)}^{(0)}$ es el dominio de $Q^{(0)}(-i)$.

Con esta nueva notación, procedemos a demostrar la siguiente proposición.

Proposición 3.4.1 *Sea $K \in \mathcal{S}^*$. Entonces:*

(i) *Para $K > 0$, tenemos que*

$$P(Q^{(K)}(1), P^{(K)} | K, \underline{y}) \propto \prod_{\bar{m} \in \mathcal{S}_4^{(K)}} \left[[Q_{\bar{m}}^{(K)}(1)]^{\xi_{\bar{m}}^{(K)} - 1} \prod_{t=1}^{T-K} \prod_{j \in \mathcal{S}} [P_{\bar{m}j}^{(K)}(t)]^{\xi_{\bar{m}j}^{(K)}(t) - 1} \right].$$

(ii) *Para $K = 0$, se tiene*

$$P(Q^{(0)} | K = 0, \underline{y}) \propto \prod_{t=1}^T \prod_{j \in \mathcal{S}} [Q_j^{(0)}(t)]^{\xi_j^{(0)}(t) - 1}.$$

Demostración. Sea $K \in \mathcal{S}^*$. Si $K \in \mathcal{S}$, entonces,

$$\begin{aligned}
P(Q^{(K)}(1), P^{(K)} | K, \underline{y}) &\propto L(\underline{y} | K, Q^{(K)}(1), P^{(K)}) P(Q^{(K)}(1), P^{(K)} | K) \\
&= L(\underline{y} | K, Q^{(K)}(1), P^{(K)}) P(Q^{(K)}(1) | K) P(P^{(K)} | K) \\
&\propto \prod_{\bar{m} \in S_4^{(K)}} \left[[Q_{\bar{m}}^{(K)}(1)]^{n_{\bar{m}}^{(K)}} \prod_{t=1}^{T-K} \prod_{j \in S} [P_{\bar{m}j}^{(K)}(t)]^{n_{\bar{m}j}^{(K)}(t)} \right] \prod_{\bar{m} \in S_4^{(K)}} [Q_{\bar{m}}^{(K)}(1)]^{\alpha_{\bar{m}}^{(K)}-1} \\
&\quad \times \prod_{t=1}^{T-K} \prod_{\bar{m} \in S_4^{(K)}} \left[\frac{\Gamma(\sum_{j \in S} \alpha_{\bar{m}j}^{(K)}(t))}{\prod_{j \in S} \Gamma(\alpha_{\bar{m}j}^{(K)}(t))} \prod_{j \in S} [P_{\bar{m}j}^{(K)}(t)]^{\alpha_{\bar{m}j}^{(K)}(t)-1} \right] \\
&\propto \prod_{\bar{m} \in S_4^{(K)}} \left[[Q_{\bar{m}}^{(K)}(1)]^{\xi_{\bar{m}}^{(K)}-1} \prod_{t=1}^{T-K} \prod_{j \in S} [P_{\bar{m}j}^{(K)}(t)]^{\xi_{\bar{m}j}^{(K)}(t)-1} \right].
\end{aligned}$$

Por otro lado, si $K = 0$,

$$\begin{aligned}
P(Q^{(0)} | K = 0, \underline{y}) &\propto L(\underline{y} | Q^{(0)}) P(Q^{(0)} | K = 0) \\
&\propto \left[\prod_{t=1}^T \prod_{j \in S} [Q_j^{(0)}(t)]^{n_j^{(0)}(t)} \right] \left[\prod_{t=1}^T \prod_{j \in S} [Q_j^{(0)}(t)]^{\alpha_j^{(0)}(t)-1} \right] \\
&= \prod_{t=1}^T \prod_{j \in S} [Q_j^{(0)}(t)]^{\xi_j^{(0)}(t)-1}.
\end{aligned}$$

■

Con esta última demostración finalmente tenemos todo lo necesario para exhibir las distribuciones a posteriori de los renglones de las matrices $P^{(K)}(t)$ y las distribuciones iniciales $Q^{(K)}(1)$.

Proposición 3.4.2 *Sea $K \in \mathcal{S}^*$. Para $K > 0$, la distribución final de $Q^{(K)}(1)$ condicionada a K es una Dirichlet con vector de parámetros $(\xi_{\bar{m}}^{(K)})_{\bar{m} \in S_4^{(K)}}$, mientras que para cada $t \in \{1, \dots, T - K\}$, $\bar{m} \in S_4^{(K)}$, $P_{\bar{m}}^{(K)}(t)$ posee una distribución Dirichlet con vector de parámetros $(\xi_{\bar{m}j}^{(K)}(t))_{j \in S}$. En el caso $K = 0$, tenemos que para toda $t \in \{1, \dots, T\}$, la distribución final de $Q^{(0)}(t)$ es una Dirichlet con vector de parámetros $(\xi_j^{(0)}(t))_{j \in S}$.*

Demostración. Sea $K \in \mathcal{S}^*$. Si $K > 0$, con $Q^{(K)}(1)$ ocurre lo siguiente:

$$\begin{aligned}
P\left(Q^{(K)}(1) \mid K, \underline{y}\right) &= \int_{\Theta_P^{(K)}} P\left(Q^{(K)}(1), P^{(K)} \mid K, \underline{y}\right) dP^{(K)} \\
\text{Prop. 3.4.1} \quad &\propto \int_{\Theta_P^{(K)}} \prod_{\bar{m} \in S_4^{(K)}} \left[\left[Q_{\bar{m}}^{(K)}(1) \right]^{\xi_{\bar{m}}^{(K)} - 1} \prod_{t=1}^{T-K} \prod_{j \in S} \left[P_{\bar{m}j}^{(K)}(t) \right]^{\xi_{\bar{m}j}^{(K)}(t) - 1} \right] dP^{(K)} \\
&= \prod_{\bar{m} \in S_4^{(K)}} \left[Q_{\bar{m}}^{(K)}(1) \right]^{\xi_{\bar{m}}^{(K)} - 1} \left[\int_{\Theta_P^{(K)}} \prod_{t=1}^{T-K} \prod_{j \in S} \left[P_{\bar{m}j}^{(K)}(t) \right]^{\xi_{\bar{m}j}^{(K)}(t) - 1} dP^{(K)} \right] \\
&= \prod_{\bar{m} \in S_4^{(K)}} \left[\left[Q_{\bar{m}}^{(K)}(1) \right]^{\xi_{\bar{m}}^{(K)} - 1} \prod_{t=1}^{T-K} \int_{\Delta_{|S|}} \prod_{j \in S} \left[P_{\bar{m}j}^{(K)}(t) \right]^{\xi_{\bar{m}j}^{(K)}(t) - 1} dP_{\bar{m}}^{(K)}(t) \right] \\
&\propto \prod_{\bar{m} \in S_4^{(K)}} \left[Q_{\bar{m}}^{(K)}(1) \right]^{\xi_{\bar{m}}^{(K)} - 1},
\end{aligned}$$

por lo que, dado K y \underline{y} , $Q^{(K)}(1)$ tiene distribución *Dirichlet* $\left(\left(\xi_{\bar{m}}^{(K)} \right)_{\bar{m} \in S_4^{(K)}} \right)$. Ahora, dada $i \in \{1, \dots, T - K\}$, tenemos que la distribución a posteriori de la matriz $P^{(K)}(i)$ dado K es la siguiente:

$$\begin{aligned}
P\left(\text{vec } P^{(K)}(i) \mid K, \underline{y}\right) &= \int_{\Theta^{(K)}(-i)} P\left(Q^{(K)}(1), P^{(K)} \mid K, \underline{y}\right) dQ^{(K)}(1) dP^{(K)}(-i) \\
\text{Prop. 3.4.1} \quad &\propto \int_{\Theta^{(K)}(-i)} \prod_{\bar{m} \in S_4^{(K)}} \left[\left[Q_{\bar{m}}^{(K)}(1) \right]^{\xi_{\bar{m}}^{(K)} - 1} \prod_{t=1}^{T-K} \prod_{j \in S} \left[P_{\bar{m}j}^{(K)}(t) \right]^{\xi_{\bar{m}j}^{(K)}(t) - 1} \right] dQ^{(K)}(1) dP^{(K)}(-i) \\
(1) \quad &\propto \int_{\Theta_{P(-i)}^{(K)}} \prod_{\bar{m} \in S_4^{(K)}} \left[\prod_{t=1}^{T-K} \prod_{j \in S} \left[P_{\bar{m}j}^{(K)}(t) \right]^{\xi_{\bar{m}j}^{(K)}(t) - 1} \right] dP^{(K)}(-i) \\
&= \prod_{\bar{m} \in S_4^{(K)}} \left[\prod_{j \in S} \left[P_{\bar{m}j}^{(K)}(i) \right]^{\xi_{\bar{m}j}^{(K)}(i) - 1} \prod_{\substack{t=1 \\ t \neq i}}^{T-K} \left[\int_{\Delta_{|S|}} \prod_{j \in S} \left[P_{\bar{m}j}^{(K)}(t) \right]^{\xi_{\bar{m}j}^{(K)}(t) - 1} dP_{\bar{m}}^{(K)}(t) \right] \right] \\
&\propto \prod_{\bar{m} \in S_4^{(K)}} \prod_{j \in S} \left[P_{\bar{m}j}^{(K)}(i) \right]^{\xi_{\bar{m}j}^{(K)}(i) - 1},
\end{aligned}$$

donde (1) se da por la independencia condicionada a K de $Q^{(K)}(1)$ con las matrices de probabilidades de transición y por el hecho de que se está integrando sobre el soporte de $Q^{(K)}(1)$. De la proposición 3.4.2, se puede deducir de manera trivial las distribuciones

finales de los renglones de $P^{(K)}(i)$. Sea $\bar{w} \in S_4^{(K)}$. Entonces,

$$\begin{aligned}
P\left(P_{\bar{w}}^{(K)}(i) \mid K, \underline{y}\right) &= \int_{\Delta_{|S|}^{|S_4^{(K)}|_{-1}}} P\left(\text{vec } P^{(K)}(i) \mid K, \underline{y}\right) dP_{-\bar{w}}^{(K)}(i) \\
&\propto \left[\prod_{j \in S} \left[P_{\bar{w}j}^{(K)}(i) \right]^{\xi_{\bar{w}j}^{(K)}(i)-1} \right] \prod_{\substack{\bar{m} \in S_4^{(K)} \\ \bar{m} \neq \bar{w}}} \left[\int_{\Delta_{|S|}} \prod_{j \in S} \left[P_{\bar{m}j}^{(K)}(i) \right]^{\xi_{\bar{m}j}^{(K)}(i)-1} dP_{\bar{m}}^{(K)}(i) \right] \\
&= \prod_{j \in S} \left[P_{\bar{w}j}^{(K)}(i) \right]^{\xi_{\bar{w}j}^{(K)}(i)-1}.
\end{aligned}$$

Por lo tanto, para cada $t \in \{1, \dots, T - K\}$ y $\bar{m} \in S_4^{(K)}$, la distribución final de $P_{\bar{m}}^{(K)}(t)$ dado K corresponde a una Dirichlet con vector de parámetros $\left(\xi_{\bar{m}j}^{(K)}(t)\right)_{j \in S}$. En cuanto a $K = 0$, dado $i \in \{1, \dots, T\}$,

$$\begin{aligned}
P\left(Q^{(0)}(i) \mid K = 0, \underline{y}\right) &= \int_{\Theta_{Q^{(0)}}^{(0)}} P\left(Q^{(0)} \mid K = 0, \underline{y}\right) dQ^{(0)}(-i) \\
\text{Prop. 3.4.1} \quad &\propto \int_{\Theta_{Q^{(0)}}^{(0)}} \prod_{t=1}^T \prod_{j \in S} \left[Q_j^{(0)}(t) \right]^{\xi_j^{(0)}(t)-1} dQ^{(0)}(-i) \\
&= \prod_{j \in S} \left[Q_j^{(0)}(i) \right]^{\xi_j^{(0)}(i)-1} \left[\prod_{\substack{t=1 \\ t \neq i}}^T \int_{\Delta_{|S|}} \prod_{j \in S} \left[Q_j^{(0)}(t) \right]^{\xi_j^{(0)}(t)-1} dQ^{(0)}(-i) \right] \\
&= \prod_{j \in S} \left[Q_j^{(0)}(i) \right]^{\xi_j^{(0)}(i)-1},
\end{aligned}$$

lo cual nos indica que la distribución a posteriori de las $Q^{(0)}(t)$'s es una Dirichlet con vector de parámetros $\left(\xi_j^{(0)}(t)\right)_{j \in S}$, $1 \leq t \leq T$. ■

A continuación exhibimos cuál será la distribución final del orden K . Utilizando la fórmula de Bayes, observamos que la distribución a posteriori de K se puede expresar en términos de la función de verosimilitud de \underline{y} condicionada a $K = k$ y la distribución a priori de K ,

$$p(K = k \mid \underline{y}) \propto L(\underline{y} \mid K = k) p(K = k) = L(\underline{y} \mid K = k) \frac{\lambda^k}{k!} \mathbb{I}_{\mathcal{S}^*}(k).$$

Así, bastará con calcular $p(\underline{y} \mid K = k)$ para cada $k \in \mathcal{S}^*$ para conocer $p(K = k \mid \underline{y})$.

Proposición 3.4.3 Sea $K \in \mathcal{S}^*$. Entonces, para $K > 0$ tenemos que

$$p(\underline{y}|K = k) = \frac{\Gamma\left(\sum_{\bar{m} \in S_4^{(K)}} \alpha_{\bar{m}}^{(K)}\right)}{\Gamma\left(\sum_{\bar{m} \in S_4^{(K)}} \alpha_{\bar{m}}^{(K)} + n_{\bar{m}}^{(K)}\right)} \left[\prod_{\bar{m} \in S_4^{(K)}} \frac{\Gamma\left(\alpha_{\bar{m}}^{(K)} + n_{\bar{m}}^{(K)}\right)}{\Gamma\left(\alpha_{\bar{m}}^{(K)}\right)} \right] \\ \times \prod_{t=1}^{T-K} \prod_{\bar{m} \in S_4^{(K)}} \left[\frac{\Gamma\left(\sum_{j \in S} \alpha_{\bar{m}j}^{(K)}(t)\right)}{\Gamma\left(\sum_{j \in S} \alpha_{\bar{m}j}^{(K)}(t) + n_{\bar{m}j}^{(K)}(t)\right)} \left[\prod_{j \in S} \frac{\Gamma\left(\alpha_{\bar{m}j}^{(K)}(t) + n_{\bar{m}j}^{(K)}(t)\right)}{\Gamma\left(\alpha_{\bar{m}j}^{(K)}(t)\right)} \right] \right],$$

mientras que para $K = 0$,

$$p(\underline{y}|K = 0) = \prod_{t=1}^T \left[\frac{\Gamma\left(\sum_{j \in S} \alpha_j^{(0)}(t)\right)}{\Gamma\left(\sum_{j \in S} \alpha_j^{(0)}(t) + n_j^{(0)}(t)\right)} \prod_{j \in S} \frac{\Gamma\left(\alpha_j^{(0)}(t) + n_j^{(0)}(t)\right)}{\Gamma\left(\alpha_j^{(0)}(t)\right)} \right].$$

Demostración. Sea $K \in \mathcal{S}^*$. Para $K > 0$, empecemos por calcular $p(\underline{y}|K = k)$:

$$p(\underline{y}|K = k) = \int_{\Theta^{(K)}} L(\underline{y}|K, \theta^{(K)}) P\left(Q^{(K)}(1), P^{(K)}|K\right) dQ^{(K)}(1) dP^{(K)} \\ \propto \int_{\Theta^{(K)}} P\left(Q^{(K)}(1), P^{(K)}|K, \underline{y}\right) dQ^{(K)}(1) dP^{(K)} \\ \text{Prop. 3.4.1} \quad \propto \int_{\Theta^{(K)}} \prod_{\bar{m} \in S_4^{(K)}} \left[\left[Q_{\bar{m}}^{(K)}(1) \right]^{\xi_{\bar{m}}^{(K)} - 1} \prod_{t=1}^{T-K} \prod_{j \in S} \left[P_{\bar{m}j}^{(K)}(t) \right]^{\xi_{\bar{m}j}^{(K)}(t) - 1} \right] dQ^{(K)}(1) dP^{(K)} \\ (2) \quad = \frac{\prod_{\bar{m} \in S_4^{(K)}} \Gamma\left(\alpha_{\bar{m}}^{(K)} + n_{\bar{m}}^{(K)}\right)}{\Gamma\left(\sum_{\bar{m} \in S_4^{(K)}} \alpha_{\bar{m}}^{(K)} + n_{\bar{m}}^{(K)}\right)} \prod_{t=1}^{T-K} \prod_{\bar{m} \in S_4^{(K)}} \left[\frac{\prod_{j \in S} \Gamma\left(\alpha_{\bar{m}j}^{(K)}(t) + n_{\bar{m}j}^{(K)}(t)\right)}{\Gamma\left(\sum_{j \in S} \alpha_{\bar{m}j}^{(K)}(t) + n_{\bar{m}j}^{(K)}(t)\right)} \right].$$

La constante normalizadora c tal que

$$p(\underline{y}|K = k) = c \cdot \int_{\Theta^{(K)}} \prod_{\bar{m} \in S_4^{(K)}} \left[\left[Q_{\bar{m}}^{(K)}(1) \right]^{\xi_{\bar{m}}^{(K)} - 1} \prod_{t=1}^{T-K} \prod_{j \in S} \left[P_{\bar{m}j}^{(K)}(t) \right]^{\xi_{\bar{m}j}^{(K)}(t) - 1} \right] dQ^{(K)}(1) dP^{(K)}$$

está dada por

$$c = \frac{\Gamma\left(\sum_{\bar{m} \in S_4^{(K)}} \alpha_{\bar{m}}^{(K)}\right)}{\prod_{\bar{m} \in S_4^{(K)}} \Gamma\left(\alpha_{\bar{m}}^{(K)}\right)} \times \prod_{t=1}^{T-K} \prod_{\bar{m} \in S_4^{(K)}} \left[\frac{\Gamma\left(\sum_{j \in S} \alpha_{\bar{m}j}^{(K)}(t)\right)}{\prod_{j \in S} \Gamma\left(\alpha_{\bar{m}j}^{(K)}(t)\right)} \right].$$

Reacomodando los términos se tiene,

$$p(\underline{y}|K = k) = \frac{\Gamma\left(\sum_{\bar{m} \in S_4^{(K)}} \alpha_{\bar{m}}^{(K)}\right)}{\Gamma\left(\sum_{\bar{m} \in S_4^{(K)}} \alpha_{\bar{m}}^{(K)} + n_{\bar{m}}^{(K)}\right)} \left[\prod_{\bar{m} \in S_4^{(K)}} \frac{\Gamma\left(\alpha_{\bar{m}}^{(K)} + n_{\bar{m}}^{(K)}\right)}{\Gamma\left(\alpha_{\bar{m}}^{(K)}\right)} \right] \\ \times \prod_{t=1}^{T-K} \prod_{\bar{m} \in S_4^{(K)}} \left[\frac{\Gamma\left(\sum_{j \in S} \alpha_{\bar{m}j}^{(K)}(t)\right)}{\Gamma\left(\sum_{j \in S} \alpha_{\bar{m}j}^{(K)}(t) + n_{\bar{m}j}^{(K)}(t)\right)} \left[\prod_{j \in S} \frac{\Gamma\left(\alpha_{\bar{m}j}^{(K)}(t) + n_{\bar{m}j}^{(K)}(t)\right)}{\Gamma\left(\alpha_{\bar{m}j}^{(K)}(t)\right)} \right] \right].$$

En cuanto a $p(\underline{y}|K = 0)$,

$$p(\underline{y}|K = 0) = \int_{\Theta^{(0)}} L(\underline{y}|K = 0, Q^{(0)}) P(Q^{(0)}|K = 0) dQ^{(0)} \\ = \int_{\Theta^{(0)}} \prod_{t=1}^T \frac{\Gamma\left(\sum_{j \in S} \alpha_j^{(0)}(t)\right)}{\prod_{j \in S} \Gamma\left(\alpha_j^{(0)}(t)\right)} \prod_{j \in S} [Q_j^{(0)}(t)]^{\xi_j^{(0)}(t)-1} dQ^{(0)} \\ = \prod_{t=1}^T \left[\frac{\Gamma\left(\sum_{j \in S} \alpha_j^{(0)}(t)\right)}{\prod_{j \in S} \Gamma\left(\alpha_j^{(0)}(t)\right)} \int_{\Delta_{|S|}} \prod_{j \in S} [Q_j^{(0)}(t)]^{\xi_j^{(0)}(t)-1} dQ^{(0)} \right] \\ = \prod_{t=1}^T \frac{\Gamma\left(\sum_{j \in S} \alpha_j^{(0)}(t)\right) \prod_{j \in S} \Gamma\left(\alpha_j^{(0)}(t) + n_j^{(0)}(t)\right)}{\prod_{j \in S} \Gamma\left(\alpha_j^{(0)}(t)\right) \Gamma\left(\sum_{j \in S} \alpha_j^{(0)}(t) + n_j^{(0)}(t)\right)}.$$

Así,

$$p(\underline{y}|K = 0) = \prod_{t=1}^T \left[\frac{\Gamma\left(\sum_{j \in S} \alpha_j^{(0)}(t)\right)}{\Gamma\left(\sum_{j \in S} \alpha_j^{(0)}(t) + n_j^{(0)}(t)\right)} \prod_{j \in S} \frac{\Gamma\left(\alpha_j^{(0)}(t) + n_j^{(0)}(t)\right)}{\Gamma\left(\alpha_j^{(0)}(t)\right)} \right]$$

■

De esta forma, dada $K \in \mathcal{S}^*$, tenemos que

$$p(K|\underline{y}) = \begin{cases} \frac{\lambda^K}{K!} \cdot \frac{\Gamma\left(\sum_{\bar{m} \in S_4^{(K)}} \alpha_{\bar{m}}^{(K)}\right)}{\Gamma\left(\sum_{\bar{m} \in S_4^{(K)}} \alpha_{\bar{m}}^{(K)} + n_{\bar{m}}^{(K)}\right)} \left[\prod_{\bar{m} \in S_4^{(K)}} \frac{\Gamma\left(\alpha_{\bar{m}}^{(K)} + n_{\bar{m}}^{(K)}\right)}{\Gamma\left(\alpha_{\bar{m}}^{(K)}\right)} \right] \times \\ \prod_{t=1}^{T-K} \prod_{\bar{m} \in S_4^{(K)}} \left[\frac{\Gamma\left(\sum_{j \in S} \alpha_{\bar{m}j}^{(K)}(t)\right)}{\Gamma\left(\sum_{j \in S} \alpha_{\bar{m}j}^{(K)}(t) + n_{\bar{m}j}^{(K)}(t)\right)} \left[\prod_{j \in S} \frac{\Gamma\left(\alpha_{\bar{m}j}^{(K)}(t) + n_{\bar{m}j}^{(K)}(t)\right)}{\Gamma\left(\alpha_{\bar{m}j}^{(K)}(t)\right)} \right] \right] & \text{si } K > 0, \\ \\ \prod_{t=1}^T \left[\frac{\Gamma\left(\sum_{j \in S} \alpha_j^{(0)}(t)\right)}{\Gamma\left(\sum_{j \in S} \alpha_j^{(0)}(t) + n_j^{(0)}(t)\right)} \prod_{j \in S} \frac{\Gamma\left(\alpha_j^{(0)}(t) + n_j^{(0)}(t)\right)}{\Gamma\left(\alpha_j^{(0)}(t)\right)} \right] & \text{si } K = 0, \end{cases}$$

donde $c = \sum_{k=0}^M p(\underline{y}|K = k) \frac{\lambda^K}{K!}$.

Capítulo 4

Aplicación a datos de la red de monitoreo de la calidad del aire de la Ciudad de México

En este capítulo presentamos la aplicación del modelo descrito en el capítulo 2 para las mediciones de ozono y partículas con diámetro menor a 10 micras (PM_{10}) obtenidas de la red de monitoreo de la calidad del aire establecida en el área metropolitana de la Ciudad de México.

4.1. Descripción de los datos

Los contaminantes considerados en este estudio son el ozono y las partículas con diámetro menor a 10 micras (PM_{10}), los cuales están dados en partes por millón (ppm) y en microgramos por metro cúbico ($\mu g/m^3$), respectivamente. Las mediciones de estos contaminantes fueron recolectadas desde el 1 de enero de 1995 al 31 de diciembre del 2019 por la red de monitoreo de la Ciudad de México. Estas mediciones fueron recolectadas en cada estación de monitoreo minuto a minuto y el promedio por hora es reportado.

Observación 4.1.1 El monitoreo del ozono en el área metropolitana de la Ciudad de México data desde finales de la década de los 80. Sin embargo, dado que el seguimiento sistemático del PM_{10} inició en 1995, es a partir de este año del que datan nuestras observaciones.

Nuestro objetivo es estudiar la probabilidad de que las mediciones de ozono y PM_{10} rebasen sus límites ambientales correspondientes. Esto lo haremos tomando en cuenta el comportamiento de estos contaminantes a lo largo del año. Con esto también será posible identificar en qué periodo del año se tiene una alta o una baja en el número de excedencias. En lugar de considerar los casos por regiones, tomaremos en cuenta los valores globales de los niveles de contaminantes. Es decir, se tomarán los máximos diarios en toda la zona metropolitana. Esto es realizado de la siguiente forma.

En el caso del ozono el máximo diario es el máximo de todas las mediciones por hora en un periodo de 24 horas tomadas en todas las estaciones de monitoreo. Si tomamos en cuenta las mediciones de PM_{10} , entonces primero se obtiene la media de las mediciones horarias reportadas en el día. El máximo diario es el máximo de todos estos promedios tomando en cuenta todas las estaciones de monitoreo.

Observación 4.1.2 En el caso del PM_{10} se toma el promedio de 24 horas primero porque los alertas ambientales son con base en estos promedios.

Los límites ambientales considerados en este trabajo son aquellos especificados en (NOM-020-SSA1-2014) y (NOM-025-SSA1-2014), es decir, son $L = 75 \mu g/m^3$ para los datos de PM_{10} y $L = 0.095 ppm$ para los datos de ozono.

En la Figura 4.1 se tiene las gráficas con las mediciones diarias de ozono y PM_{10} . Las líneas horizontales representan los límites ambientales correspondientes a cada contaminante. Viendo la gráfica superior de la figura 4.1 que corresponde a las mediciones de PM_{10} , podemos ver que hay varios episodios con repuntes altos al inicio del periodo observado y, posicionándonos al final de dicho periodo, a pesar de picos recurrentes, estos no son tan elevados como al principio.

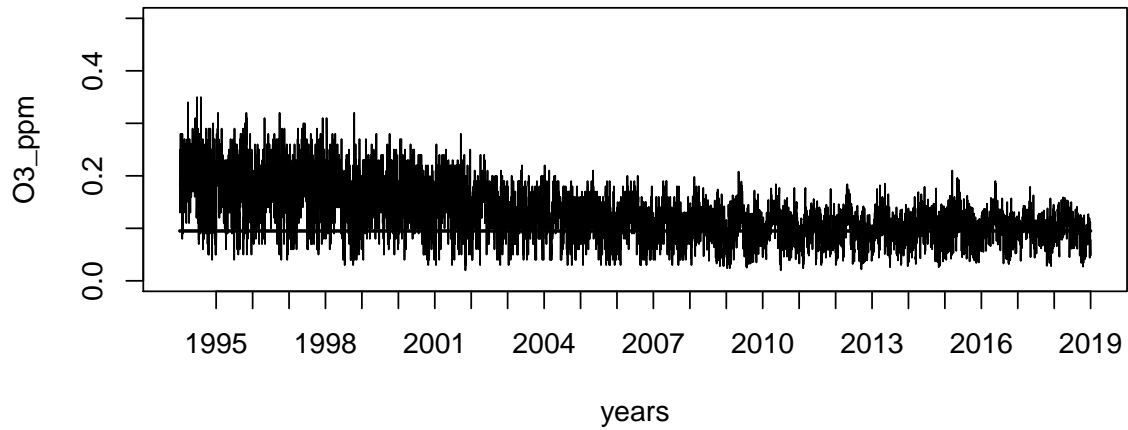
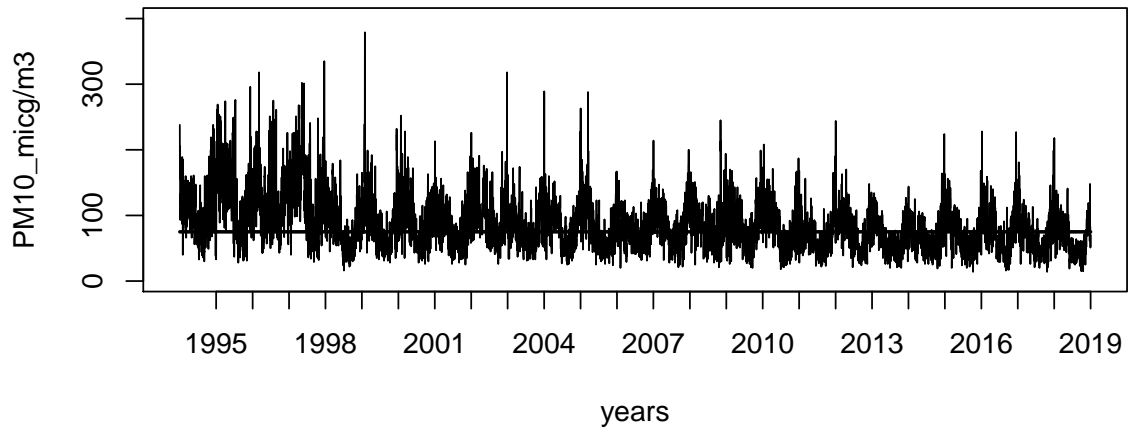


Figura 4.1: Datos de de la red de monitoreo de la CDMX de PM_{10} y ozono del periodo que comprende desde el 1 de enero de 1995 hasta el 31 de Diciembre del 2019. Las líneas horizontales corresponden a los límites permisibles para cada contaminante, i.e., $L = 75 \mu g/m^3$ para las observaciones de PM_{10} y $L = 0.095 ppm$ para el ozono.

En el caso del ozono, cuyas mediciones corresponden a la gráfica inferior en la figura 4.1, también vemos concentraciones más altas en el inicio del periodo observado. Sin embargo, hacia el final de dicho periodo, las mediciones muestran valores más pequeños

que se mueven alrededor del límite permisible $L = 0.095 ppm$. De esto podemos concluir que las medidas implementadas por las autoridades ambientales pueden haber contribuido a la disminución de los niveles de ozono y PM_{10} , incluyendo la altura de los picos, en el área metropolitana. Los valores máximo, mínimo y promedio global de los contaminantes considerados en este estudio están dados en la tabla 4.1 dada a continuación.

Contaminante	Mínimo global	Máximo global	Promedio global
ozono	0.02 ppm	0.35 ppm	0.128 ppm
PM_{10}	14 $\mu g/m^3$	379 $\mu g/m^3$	86.92 $\mu g/m^3$

Tabla 4.1: Máximo, mínimo y media de las concentraciones de ozono y PM_{10} recolectadas de la red de monitoreo de la Ciudad de México durante el periodo del 01 de enero de 1995 hasta el 31 de diciembre de 2019.

Observando las mediciones de ozono y PM_{10} presentadas en las gráficas de la figura 4.1, podemos ver los días en los cuales se rebasó los límites permisibles establecidos. El número de días por año en que se sobrepasó los límites ambientales permisibles varía desde 124 días hasta 312 en el caso del PM_{10} y desde 178 hasta 341 en el caso del ozono. El número más alto de excedencias en un año ocurrió en los años 1997 y 1996 de los contaminantes PM_{10} y ozono, respectivamente, y los valores más pequeños fueron en los años 2014 y 2012. En total hubo 6,659 y 5,014 días con excedentes de ozono y PM_{10} , respectivamente, y que vienen siendo aproximadamente el 72.93 % y 55.9 % de los días observados de cada contaminante.

En la figura 4.2, gráficas superiores, podemos ver los comportamientos de los máximos (líneas punteadas gruesas), mínimos (líneas punteadas delgadas) y promedios anuales (líneas gruesas continuas) tanto del ozono como del PM_{10} a lo largo del periodo observado. En dichas gráficas es posible notar un decremento año por año de la media y máximo anual entre 1995 y 2008 en las mediciones de ozono, probablemente resultado de las medidas implementadas por autoridades ambientales con el fin de disminuir los niveles de este.

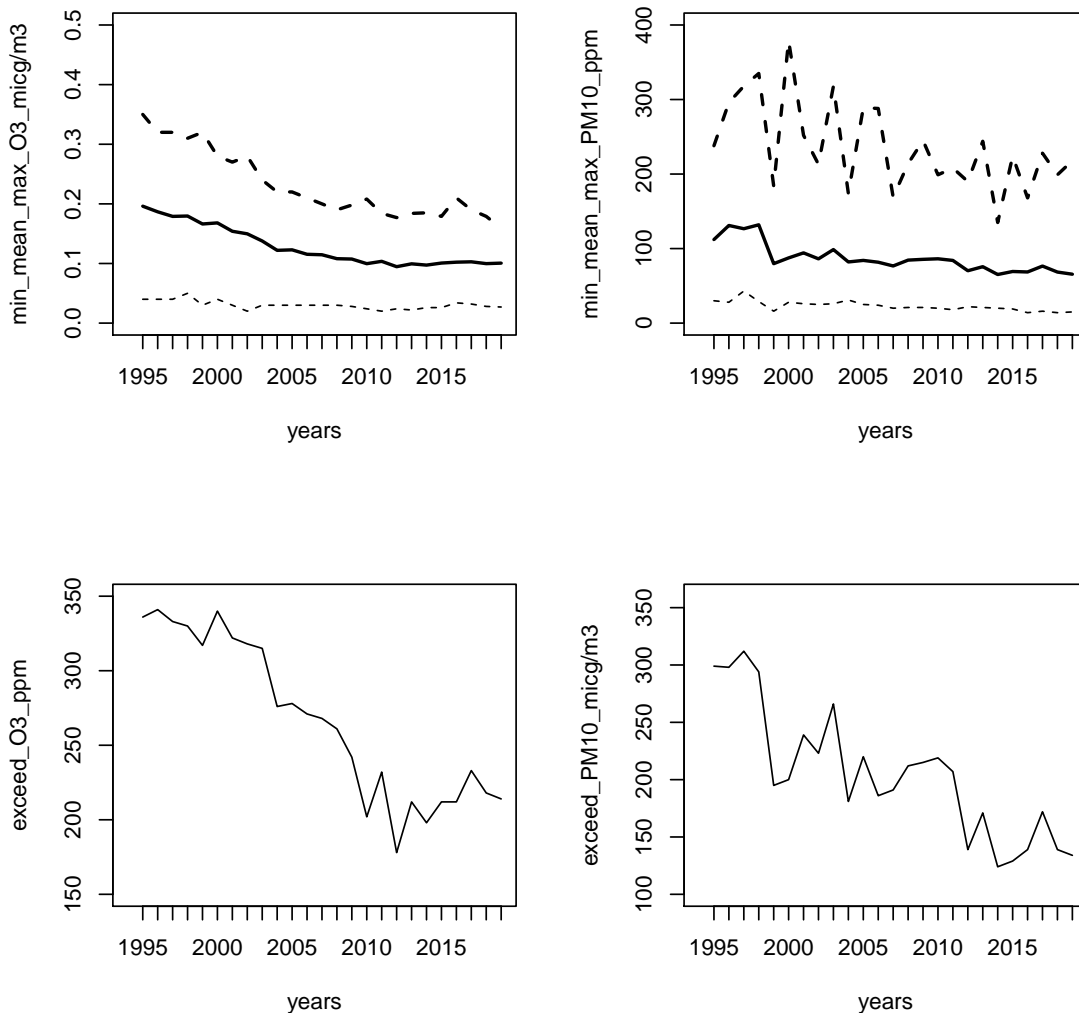


Figura 4.2: Gráficas superiores: media (líneas gruesas continuas), máximo (línea punteada gruesa) y mínimo (línea punteada delgada) de las mediciones de ozono y PM_{10} por cada año del periodo observado. Gráficas inferiores: número de excedencias por año de los límites ambientales permisibles correspondientes para el ozono y PM_{10} para cada año del periodo observado.

A su vez, estas acciones también contribuyeron a una merma de los niveles de PM_{10} ya que impactan en las emisiones de industrias, automóviles, autobuses, taxis y camiones. Esto se ve reflejado en las medias anuales. Sin embargo, se observa una menor disminución de los valores máximos anuales con respecto de las medias, probablemente consecuencia de

los picos ocurridos en el periodo. En la figura 4.2, gráficas inferiores, se muestra el número de excedencias año por año de ambos contaminantes. Observándolas podemos ver que en el año 2000 también teníamos un alto número de excedentes de ozono. De cualquier forma, esto es más pequeño que la cantidad ocurrida en 1996. También, en las mismas gráficas, podemos ver que tanto los excedentes de ozono como de PM_{10} disminuyen rápidamente hasta alrededor de 2008 y después se mantienen en valores cercanos al de este año, lo cual es reflejo de la disminución de concentración de los contaminantes a lo largo de los años como se muestra en la figura 4.1 y en el decremento de los máximos y promedios anuales mostrados en las gráficas superiores de la figura 4.2.

Así, podemos ver una mejora de la calidad del aire relacionada con el ozono y PM_{10} en años tardíos del periodo de observación. Sin embargo, esto no significa que la calidad del aire sea ideal dado que las alertas ambientales aún siguen pasando en la ciudad.

4.2. Modelo matemático

En esta sección hacemos una adaptación del modelo presentado en el capítulo 2 así como el análisis bayesiano presentado en el capítulo 3 para estudiar conjuntamente la probabilidad de la ocurrencias de excedencias de los límites ambientales de las concentraciones de ozono y PM_{10} .

Sea $s \in \mathbb{N}$ el total de contaminantes a estudiar y sea $L = (L_1, \dots, L_s) \in \mathbb{R}^s$ donde, para cada $i \in \{1, \dots, s\}$, L_i representa el límite aceptado de la concentración de un contaminante i . Por otro lado, sea $N \in \mathbb{N}$ el número de años en el que las mediciones de los contaminantes a estudiar fueron tomadas y, para cada $j \in \{1, \dots, N\}$, sea T_j el número de días correspondiente al j –ésimo año.

Sea $i \in \{1, \dots, N\}$. Dado $t \in \{1, \dots, T_i\}$, vamos a considerar al vector $Z^{(i)}(t)$ como $Z^{(i)}(t) = (Z_1^{(i)}(t), \dots, Z_s^{(i)}(t)) \in \mathbb{R}^s$, donde para cada $j \in \{1, \dots, s\}$, $Z_j^{(i)}(t)$ representa la medición del contaminante j recolectado en el día t del i –ésimo año. Así, para cada $i \in \{1, \dots, N\}$, vamos a considerar a $Z^{(i)} = (Z^{(i)}(t))_{1 \leq t \leq T_i}$ como el conjunto de

mediciones diarias de los s contaminantes a estudiar durante el año i .

Para cuestiones prácticas en este trabajo, basándonos en (Drton *et al.*, 2003), asumiremos que, para todo $j \in \{1, \dots, N\}$, $T_j = 366$, en donde, en caso de que el año j no sea bisiesto, $Z^{(j)}(366) = \bar{0}$.

Observación 4.2.1 Dado lo anterior, dejaremos de usar índices para referirnos a los días de cualquier año en cuestión. Es decir, para toda $j \in \{1, \dots, N\}$, $T_j = T$, donde $T = 366$.

Ahora, para cada $t \in \{1, \dots, T\}$ consideremos a $Y(t) = (Y_1(t), \dots, Y_s(t))$ donde $Y_i(t)$, $1 \leq i \leq s$, están definidos de la siguiente forma:

$$Y_i(t) = \begin{cases} 0 & \text{si } Z_i(t) < L_i, \\ 1 & \text{si } Z_i(t) \geq L_i. \end{cases}$$

Sea $y = \{Y(t) : 1 \leq t \leq T\}$. Su espacio de estados es $\mathcal{X} = \{(x_1, x_2, \dots, x_s) \in \{0, 1\}^s\}$. Utilizando la función $\Phi_{\{0,1\},s}$ definida en 2.0.3, es posible convertir al proceso dado por $y = \{Y(t) : 1 \leq t \leq T\}$ que toma valores en un espacio s -dimensional a un proceso unidimensional que toma valores en $S = \text{Im}(\Phi_{\{0,1\},s}) = \{0, 1, \dots, 2^s - 1\}$. Sea $W = \{W_t : t \in \mathbb{N}\}$ dicho proceso. Vamos a asumir que dicho proceso está gobernado por una cadena de Markov $X^{(K)} = \{X_t^{(K)} : t \in \mathbb{N}\}$ no homogénea de orden K y espacio de estados $S_3^{(K)} = \{(y_1, y_2, \dots, y_K) \in S^K\}$, donde $K \in \mathcal{S} \cup \{0\}$ con $\mathcal{S} = \{1, \dots, M\}$ y $M \leq T$. Recordemos que por la definición 2.0.1 de la cadena $X^{(K)}$, tenemos que para cada $n \in \mathbb{N}$, $X_n^{(K)} = (W_n, \dots, W_{n+K-1})$. En el caso en que $K \in \mathcal{S}$, al igual que con el espacio de estados $S_3^{(K)}$ de los contaminantes a estudiar y como se mostró en el capítulo 2, podemos ver a los elementos de $S_3^{(k)}$ como elementos del conjunto $S_4^{(k)} = \{0, 1, \dots, |S|^k - 1\}$ por medio de la transformación $\Phi_{|S|,k}(\cdot)$ definida en el capítulo 1. También, recordemos la notación presentada: dados $\mathbf{x} = (y_1, y_2, \dots, y_k) \in S_3^{(K)}$, $\bar{m} \in S_4^{(K)}$ tales que $\Phi_{|S|,k}(\mathbf{x}) = \bar{m}$, $\mathbf{x} \leftrightarrow \bar{m}$ indicará su respectiva correspondencia. Si $K = 0$, para la cadena $X^{(0)}$ tenemos $S_4^{(K)} = S_3^{(K)}$. Para la implementación del modelo, vamos a considerar lo siguiente:

- El número de contaminantes s es dos: ozono y PM_{10} .
- El número de años que vamos a considerar es $N = 25$, donde la recolección de nuestros datos empieza desde el 1 de enero de 1995 y va hasta el 31 de diciembre de 2019.
- El vector de mediciones al día t en el año i será $Z^{(i)}(t) = (Z_1^{(i)}(t), Z_2^{(i)}(t))$, $1 \leq i \leq 25$, $t = 1, \dots, 366$, donde $Z_1^{(i)}(t)$ y $Z_2^{(i)}(t)$ son los máximos diarios de acuerdo a la sección 4.1.
- Consideraremos $L_1 = 0.095\text{ppm}$ como límite permisible para las emisiones de ozono, y $L_2 = 75\mu\text{g}/\text{m}^3$ para aquellas de PM_{10} , respectivamente, esto de acuerdo a lo establecido en (NOM-020-SSA1-2014) y (NOM-025-SSA1-2014).
- El máximo del conjunto \mathcal{S} de valores para el orden K de la cadena pasará a ser $M = 8$. Es decir, $\mathcal{S} \cup \{0\} = \{0, 1, \dots, 8\}$.

Así, el espacio de estados de la cadena $Y = \{\Phi_{\{0,1\},s}(Y(t)) : 1 \leq t \leq T\}$ resulta ser el conjunto $\{0, 1, 2, 3\}$, los cuales se generan de la siguiente forma:

$$\begin{aligned}\Phi_{1,2}(0, 0) &= 0 \cdot 2^0 + 0 \cdot 2^1 = 0, & \Phi_{1,2}(1, 0) &= 1 \cdot 2^0 + 0 \cdot 2^1 = 1, \\ \Phi_{1,2}(0, 1) &= 0 \cdot 2^0 + 1 \cdot 2^1 = 2, & \Phi_{1,2}(1, 1) &= 1 \cdot 2^0 + 1 \cdot 2^1 = 3.\end{aligned}$$

Los hiperparámetros de las distribuciones a priori Dirichlet van a ser elegidos para reflejar las ocurrencias de los eventos a los cuales están asociados. Por ejemplo, en el caso donde $K = 1$, los eventos que no han sido observados pero que son parte del espacio de posibles eventos tendrán una pequeña probabilidad asignada, mientras que aquellos que han sido observados tendrán una probabilidad que absorba tanto la contribución de los datos observados como las distribuciones iniciales. Dicho esto, presentamos los hiperparámetros para las distribuciones a priori:

- Para $K > 0$, dado $1 \leq t \leq T - K$ y $\bar{m} \in S_4^{(K)}$ vamos a considerar al vector

$(\alpha_{\bar{m}0}^{(K)}(t), \alpha_{\bar{m}1}^{(K)}(t), \alpha_{\bar{m}2}^{(K)}(t), \alpha_{\bar{m}3}^{(K)}(t)) = \left(\frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}\right)$, mientras que a $Q^{(K)}(1)$, para cada $\bar{m} \in S_4^{(K)}$ vamos a tomar $\alpha_{\bar{m}}^{(K)} = \frac{1}{16}$.

- Para $K = 0$, tomaremos $(\alpha_0^{(0)}(t), \alpha_1^{(0)}(t), \alpha_2^{(0)}(t), \alpha_3^{(0)}(t)) = \left(\frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}\right)$, $1 \leq t \leq T$.

- Para la distribución del orden K tomaremos $\lambda = 1$.

Entonces, dado $K > 0$ tenemos que la distribución a posteriori de $Q^{(K)}(1)$ es una *Dirichlet* $\left(\left(\frac{1}{16} + n_{\bar{m}}^{(2)}\right)_{\bar{m} \in S_4^{(K)}}\right)$, y la de $P_{\bar{m}j}^{(K)}(t)$ es una *Dirichlet* $\left(\left(\frac{1}{16} + n_{\bar{m}}^{(K)}(t)\right)_{j \in S}\right)$, mientras que dado $K = 0$, la distribución a posteriori de $Q^{(K)}(t)$ está dada por una *Dirichlet* $\left(\frac{1}{16} + n_0^{(0)}(t), \frac{1}{16} + n_1^{(0)}(t), \frac{1}{16} + n_2^{(0)}(t), \frac{1}{16} + n_3^{(0)}(t)\right)$, $1 \leq t \leq T$.

4.3. Estimación de los parámetros

Recordemos que la distribución a posteriori del orden de la cadena está dada de la siguiente forma:

$$p(K|\underline{y}) \propto p(\underline{y}|K) p(K) \propto L(\underline{y}|K) \frac{\lambda^K}{K!}$$

la cual , con los parámetros asignados previamente, cumple que

$$p(K|\underline{y}) \propto \begin{cases} \frac{1}{K!} \cdot \frac{\Gamma\left(\sum_{i=1}^{4^K} \frac{1}{16}\right)}{\Gamma\left(\sum_{\bar{m} \in S_4^{(K)}} \frac{1}{16} + n_{\bar{m}}^{(K)}\right)} \left[\prod_{\bar{m} \in S_4^{(K)}} \frac{\Gamma\left(\frac{1}{16} + n_{\bar{m}}^{(K)}\right)}{\Gamma\left(\frac{1}{16}\right)} \right] \times \\ \prod_{t=1}^{T-K} \prod_{\bar{m} \in S_4^{(K)}} \left[\frac{\Gamma\left(\sum_{j=0}^3 \frac{1}{16}\right)}{\Gamma\left(\sum_{j=0}^3 \frac{1}{16} + n_{\bar{m}j}^{(K)}(t)\right)} \left[\prod_{j \in S} \frac{\Gamma\left(\frac{1}{16} + n_{\bar{m}j}^{(K)}(t)\right)}{\Gamma\left(\frac{1}{16}\right)} \right] \right] & \text{si } K > 0, \\ \prod_{t=1}^T \left[\frac{\Gamma\left(\sum_{j=0}^3 \frac{1}{16}\right)}{\Gamma\left(\sum_{j=0}^3 \frac{1}{16} + n_j^{(0)}(t)\right)} \prod_{j \in S} \frac{\Gamma\left(\frac{1}{16} + n_j^{(0)}(t)\right)}{\Gamma\left(\frac{1}{16}\right)} \right] & \text{si } K = 0. \end{cases}$$

Para auxiliarnos en el cálculo de estos valores, consideramos el logaritmo de estos, ya que esta función respeta el orden. Haciendo la computación correspondiente, obtenemos lo siguiente.

$K =$	$\log(L(\underline{y} K) \frac{\lambda^K}{K!})$	$K =$	$\log(L(\underline{y} K) \frac{\lambda^K}{K!})$	$K =$	$\log(L(\underline{y} K) \frac{\lambda^K}{K!})$
0	-11952.01	3	-11808.59	6	-11837.81
1	-11503.57	4	-11806.84	7	-11906.25
2	-11786.97	5	-11818.90	8	-12026.29

Tabla 4.2: Valores en logaritmo de $L(\underline{y}|K) \frac{\lambda^K}{K!}$ para $K = 0, 1, \dots, 8$.

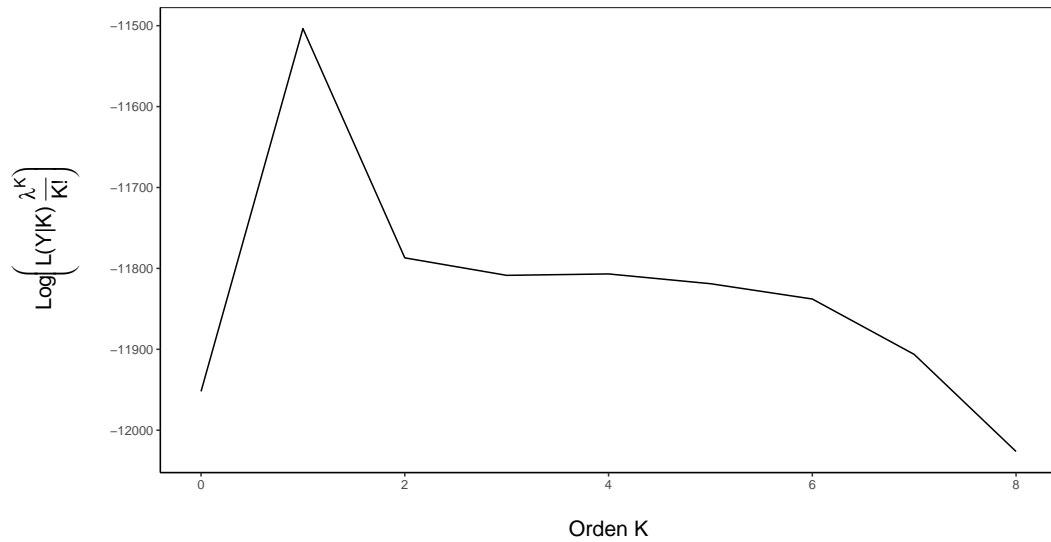


Figura 4.3: Gráfica de los valores en logaritmo de $L(\underline{y}|K) \frac{\lambda^K}{K!}$ para $K = 0, 1, \dots, 8$

De la tabla 4.2 así como de la gráfica 4.3 se puede observar que, a partir de $K = 3$, hay un decremento constante de los los valores de $L(\underline{y}|K) \frac{\lambda^K}{K!}$.

Finalmente, una vez que seleccionamos el orden K de nuestra cadena, se procedió a generar 1,000 muestras de las distribuciones finales de $Q^{(K)}(1)$ y las probabilidades de transición $P_{\bar{m}_j}^{(K)}(t)$, $\bar{m} \in S_4^{(K)}$, $j \in S$, $1 \leq t \leq T - K$. Dadas estas 1,000 muestras, se

consideró el promedio de los valores simulados como el valor estimado de cada una de las probabilidades de transición $P_{\bar{m}j}^{(K)}(t)$ así como la distribución inicial $Q^{(K)}(1)$.

4.4. Selección del modelo

Como vimos en la sección 4.3, el orden que maximiza la distribución a posteriori de K es $K = 1$. Con esto el modelo seleccionado es aquel en que se asume orden $K = 1$. Otra forma de seleccionar el modelo que proporciona las distribuciones iniciales y de transición que mejor describen nuestros datos es usar el método descrito en el capítulo 1 en la sección 1.2.3. De esta forma, vamos a considerar lo siguiente.

Retomando la definición del capítulo 1, tenemos que el método SAD está calculado como

$$\text{SAD} = \sum_{i=1} |p(i) - \hat{p}(i)|.$$

En nuestro caso las probabilidades $p(\cdot)$ serán las probabilidades empíricas y las $\hat{p}(\cdot)$ serán las probabilidades estimadas usando el método descrito en la sección 2.0.1. Es decir, dado $K \in S^*$, tenemos que:

- Si $K > 0$, entonces

$$\text{SAD} = \sum_{\bar{m} \in S_4^{(K)}} \left| Q_{\bar{m}}^{(0)}(1) - \hat{Q}_{\bar{m}}^{(0)}(1) \right| + \sum_{t=1}^{T-K} \sum_{\bar{m} \in S_4^{(K)}} \sum_{j=0}^3 \left| P_{\bar{m}j}^{(K)}(t) - \hat{P}_{\bar{m}j}^{(K)}(t) \right|,$$

donde tenemos probabilidades empíricas y estimadas como se mencionó anteriormente.

- Si $K = 0$, entonces

$$\text{SAD} = \sum_{t=1}^T \sum_{i=0}^3 \left| Q_i^{(0)}(t) - \hat{Q}_i^{(0)}(t) \right|,$$

con $Q_i^{(0)}(t)$ la probabilidad empírica y $\hat{Q}_i^{(0)}(t)$ la probabilidad estimada.

4.5. Resultados

Como vimos en la figura 4.3, a partir del orden $K = 3$ los valores de la distribución a posteriori de K son bastante menores y, por lo tanto, tienen una probabilidad bastante baja de ocurrir. De esta forma, vamos a usar el criterio SAD para la selección del modelo contemplando solamente a los órdenes $K = 0, 1, 2$. Los valores obtenidos están dados en la tabla 4.3:

$K =$	0	1	2
SAD	3.75	263.58	3393.4

Tabla 4.3: Valores obtenidos a través del criterio SAD para los modelos con orden $K = 0, 1, 2$.

Así, tenemos que el modelo con el menor valor SAD es aquel correspondiente al orden $K = 0$. Sin embargo, algo importante a tomar en cuenta en este caso es que las probabilidades estimadas a partir de la distribución a posteriori se comportan como las probabilidades observadas: estas tienen una fuerte influencia de los datos dado que los parámetros de la distribución Dirichlet hacen que mayor peso sea asociado a los estados con mayor conteo. Por lo tanto, las probabilidades finales estimadas reciben mayor peso de los datos, que es en lo que se basan las probabilidades empíricas. El segundo menor valor corresponde al caso $K = 1$. Por ende, el modelo que vamos a elegir será aquel con orden $K = 1$. De cualquier forma, para ilustrar sobre cómo utilizar el modelo para órdenes K mayores a 1, en la sección A.1 del apéndice se incluyeron los resultados obtenidos del modelo para el orden $K = 2$. Así, vamos a considerar de ahora en adelante el caso $K = 1$. En la figura 4.4 tenemos las gráficas en función del tiempo t de las probabilidades de transición $P_{ij}^{(1)}(t)$ estimadas (líneas azules) y las observadas (líneas negras punteadas) así como los intervalos de credibilidad del 95 % (líneas verdes y marrón).

Viendo la figura 4.4, es posible observar que cuando las transiciones observadas tienen valor cero o casi cero, debido a la falta de ocurrencias o casi nula aparición de estas en el

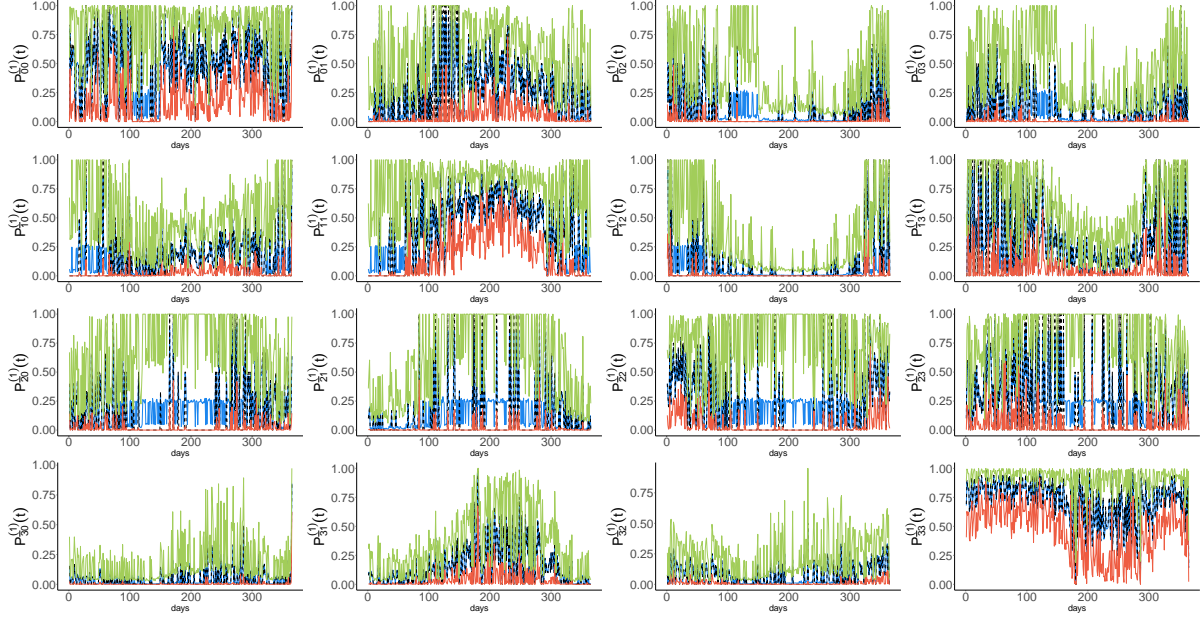


Figura 4.4: Probabilidades de transición $P_{m_j}^{(K)}(t)$ estimadas (líneas azules) y observadas (líneas punteadas negras) así como los intervalos de credibilidad del 95 % (líneas verdes y marrón) en el caso donde $K = 1$.

periodo observado, la estimación generada por las muestras a posteriori de las probabilidades de transición asigna valores pequeños a estos eventos. De esta forma, eventos con muy baja o nula probabilidad de ocurrencia en el pasado pueden ocurrir en un futuro. Adicionalmente, podemos ver que el valor más grande dado a eventos no observados es aproximadamente 0.25. En los casos donde tenemos transiciones observadas, las probabilidades de transición estimadas se ajustan bien a lo observado.

También de la figura 4.4 podemos ver que las transiciones del estado 0 (i.e., no hay excedencias de ninguno de los contaminantes) al estado 0 son más altas alrededor del fin de junio y principios de diciembre. Este periodo corresponde en parte a las estaciones de verano y otoño, y dado que en verano hay días lluviosos la mayor parte del tiempo, tenemos una menor probabilidad de que ambos contaminantes sobrepasen los límites ambientales respectivos. Por otro lado, vemos una probabilidad alta de ir del estado 0 al estado 1 (i.e., cuando sólo hay excedencia de ozono) durante el mismo periodo de valores elevados en $P_{00}^{(1)}(t)$. Así, si ocurren excedencias en aquel periodo es más probable que

sean causadas por altos niveles de ozono que de PM_{10} . Esto se reafirma al observar las probabilidades de transición de ir del estado 0 al estado 2 (el cual corresponde a sólo excedencia de PM_{10}).

Otras transiciones que llaman la atención por sus valores relativamente más altos son aquellos relacionados con las probabilidades $P_{11}^{(1)}(t)$ y $P_{33}^{(1)}(t)$, es decir, la probabilidad de que haya excedentes de ozono al tiempo $t + 1$ dado que en el tiempo presente t ya hay excedencias y la probabilidad de que haya excedencias de ozono y PM_{10} al tiempo $t + 1$ dado que en el tiempo t ya hay excedencias de ambos, respectivamente. Los altos valores del primero ocurren en un periodo similar a aquel de la transición 0 a 1. El último ocurre prácticamente durante todo el año. Así, una vez que tenemos excedentes de PM_{10} durante primavera/verano/otoño hay una alta probabilidad de que siga ocurriendo. Similarmente, una vez ambos contaminantes exceden sus respectivos límites ambientales, difícilmente podrán disminuir.

Otro interés aquí es la estimación de la probabilidad de ocurrencia de uno de los estados de S a lo largo del año. Para la obtención de los valores, dada una muestra de los valores de la distribución inicial $Q^{(k,l)}(1)$ y de las probabilidades de transición $P_{\bar{m}j}^{(k,l)}(t)$, $1 \leq l \leq 1,000$, se utilizó la fórmula recursiva (2.3) de la sección 2.0.1, y se tomó como estimación a

$$\hat{P}(W_t = i) = \frac{1}{1,000} \sum_{l=1}^{1,000} P^{(l)}(W_t = i).$$

Los valores estimados de $P(W_t = i)$, $i \in S$, $1 \leq t \leq 365$ se anexarán en la sección A.2 del apéndice. En la figura 4.5 tenemos las gráficas de las distribuciones $P(W_t = i)$ estimadas (líneas azules) y las observadas (líneas negras punteadas) así como los intervalos de credibilidad del 95 % (líneas verdes y marrón) en el caso $K = 1$.

Es posible notar al observar la figura 4.5 que el ajuste de las probabilidades estimadas y empíricas es muy buena. Además, las probabilidades con valor diferente de cero pero con pequeños valores fueron asignados a eventos que habían observado probabilidad empírica cero o aproximadamente cero. Esto se puede ver, por ejemplo, en el caso de $P(W_t = 2)$ (gráfica inferior izquierda) para el periodo que abarca desde mediados de abril hasta

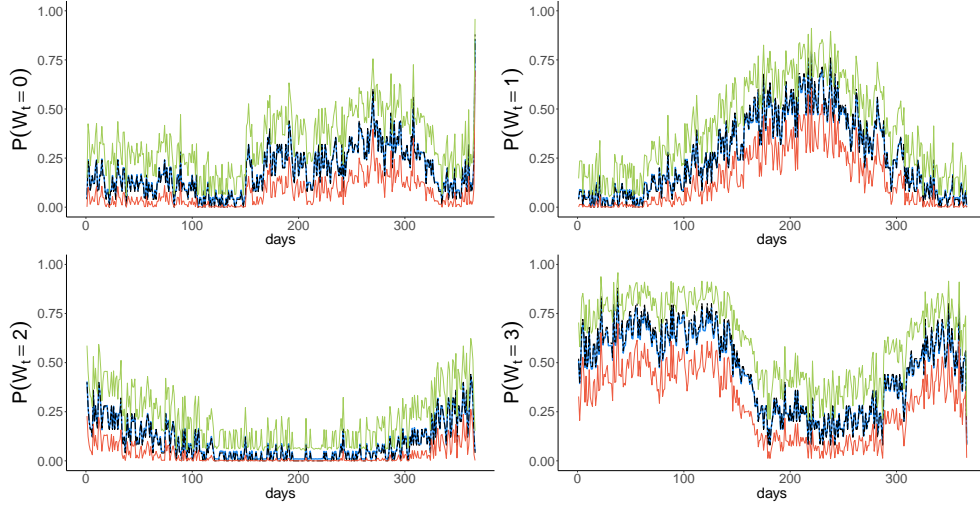


Figura 4.5: Posteriors estimadas (líneas azules) y distribución de las observaciones (línea punteada) de $P(W_t = i)$, $t = 1, \dots, T$, así como los intervalos de credibilidad del 95 % (líneas verde y marrón) con $K = 1$.

mediados de octubre, y un pequeño número de días en el caso de $P(W_t = 0)$ (gráfica superior izquierda). En ese sentido, incluso si los eventos no se han observado hasta ahora, las probabilidades estimadas correspondientes han sido valores positivos y pequeños y estos permiten que ocurra, incluso con pequeña probabilidad.

Algo que viene a ratificar la bondad de ajuste del modelo en el caso $K = 1$ es la figura 4.6 con las gráficas de $|P(W_t = i) - \hat{P}(W_t = i)|$, $t = 1, 2, \dots, T$, $i \in S$, donde $P(W_t = i)$ y $\hat{P}(W_t = i)$ indican respectivamente, las probabilidades empíricas y estimadas utilizando los valores generados de las respectivas distribuciones a posteriori de la distribución inicial y probabilidades de transición. Presentamos los casos $K = 0$ (líneas azules), $K = 1$ (líneas negras) y $K = 2$ (líneas rojas).

Observando la figura 4.6, podemos ver que, comparado con el caso $K = 2$, un mejor ajuste se obtiene tomando $K = 1$. Cuando $K = 0$, las diferencias son incluso más pequeñas que aquellas del caso $K = 1$. Sin embargo, notemos que cuando $K = 0$ las probabilidades finales estimadas reciben mayor peso de los datos, que predomina en las probabilidades empíricas, y esto puede inducir a un sobreajuste. Así, esto reafirma la selección del modelo con $K = 1$. También podemos ver que cuando $K = 2$, las

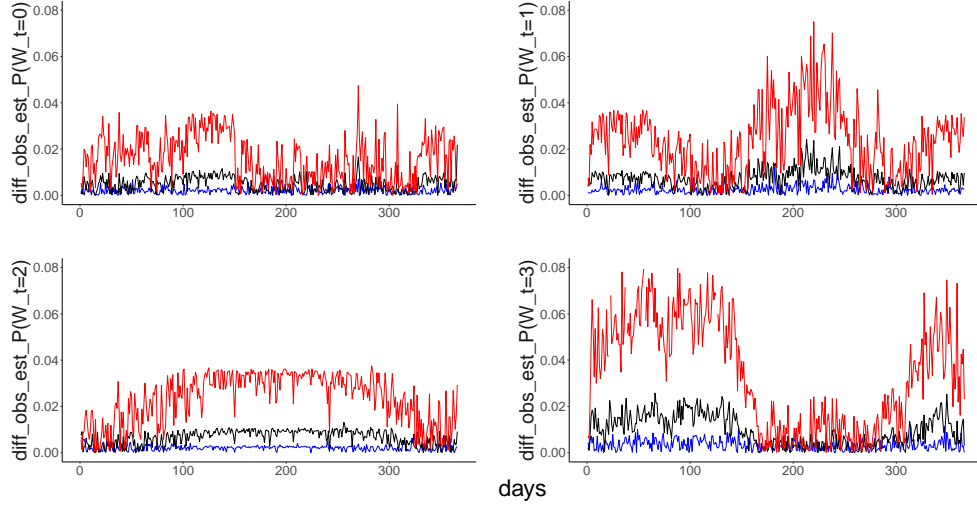


Figura 4.6: Valores absolutos de las diferencias entre las probabilidades $P(W_t = i)$ observadas y estimadas, $t = 1, \dots, T$, $i \in S$, de los casos $K = 0$ (líneas azules), $K = 1$ (líneas negras) y $K = 2$ (líneas rojas).

diferencias a lo largo de los años tampoco son muy elevadas. El valor más alto ocurrió en el periodo comprendido de junio a octubre para $P(W_t = 2)$, y de enero a junio así como de noviembre a fines de año para $P(W_t = 3)$. Sin embargo, incluso en estos casos las diferencias son pequeñas.

Volviendo con los resultados de la gráfica 4.5, tenemos otro hecho a observar: la probabilidad de ocurrencia de los estados 0, 1 y 2 están mayormente abajo de 0.75 con valores menores a 0.5 en el caso de los estados 0 y 2. Además, en la primera mitad del año la probabilidad de ocurrencia del estado 3 es mayor de 0.25 y casi todo el año se encuentra en el intervalo $(0.5, 0.75)$, el cual, comparado con la probabilidad de ocurrencia de los estados restantes, es muy alto. Así, el modelo y los datos corroboran el hecho de que durante el invierno y primavera hay una mayor probabilidad de que tanto las cantidades de ozono como de PM_{10} superen los límites ambientales debido a la temporada seca y cielos despejados. Asimismo, en marzo/abril hay una cantidad considerable de incendios forestales en el sur de México, con el humo viajando a la capital, lo cual puede generar excedencias de PM_{10} durante ese periodo.

Fijando la atención en las excedencias de ambos contaminantes, i.e., la ocurrencia del

estado 3, estas son más bajas durante la temporada de lluvias y especialmente durante el periodo que comprende principios de julio hasta finales de octubre. Una probabilidad estimada muy baja —prácticamente cero— se observa en el estado 2 (excedencias sólo de PM_{10}) durante el periodo que abarca inicios de abril a mediados de octubre. Además, en cuanto a las ocurrencias del estado 1 (sólo excedencias de ozono), se encuentran arriba de 0.5 en el periodo comprendido de mediados de junio hasta principios de septiembre. En contraste, la probabilidad de ocurrencia del estado 3 es baja (alrededor de 0.25) en el mismo periodo. Además, podemos ver que desde el 1 de enero hasta principios de junio la probabilidad de excedencia de ambos contaminantes, es decir, la ocurrencia del estado 3, está por encima de 0.5, a diferencia de los otros estados que están por debajo de 0.3 la mayor parte del periodo. Así, es más probable tener excedencias de ambos contaminantes al mismo tiempo que tener excedencia de sólo uno o no tener excedencias como tal. Algunos ejemplos concretos de los resultados presentados aquí pueden ser usados para obtener información sobre la ocurrencia de excedencias de ningún, uno o más contaminantes en un día dado del año. Esto se realiza utilizando la probabilidad estimada $P(W_t = i)$, $i \in S_2$, $t = 1, \dots, T$.

Procedemos de la siguiente forma. Si queremos conocer la probabilidad de tener excedentes tanto de ozono como de PM_{10} al tiempo dado t , $t = 1, \dots, T$, sólo necesitamos observar los valores de $P(W_t = 3)$. Por ejemplo, tomando $t = 90$, que corresponde al 31 de marzo en un año no bisiesto y el 30 de marzo en uno bisiesto, tenemos que esta probabilidad es aproximadamente 0.777 mientras que la probabilidad de tener sólo excedencias de ozono es $P(W_{90} = 1) \approx 0.127$ y solo PM_{10} $P(W_{90} = 2) \approx 0.009$. También tenemos que $P(W_{90} = 0) \approx 0.086$.

Otra cantidad que puede ser obtenida usando el modelo es la probabilidad de tener excedencias de ningún, algún o ambos contaminantes en un tiempo dado cuando tenemos información de unos días anteriores. Por ejemplo, asumiendo que queremos conocer la probabilidad de que al día $t = 100$ (un día a principios de abril) haya excedencias de ambos contaminantes, es decir, una ocurrencia del estado 3, y dado que en los tres días

anteriores tenemos la secuencia de estados: $W_{97} = 2, W_{98} = 1, W_{99} = 2, W_{100} = 3$. Es decir, queremos obtener

$$\begin{aligned}
 P(W_{97} = 2, W_{98} = 1, W_{99} = 2, W_{100} = 3) &= P(W_{100} = 3 \mid W_{99} = 2) \\
 &\quad \times P(W_{99} = 2 \mid W_{98} = 1) \\
 &\quad \times P(W_{98} = 1 \mid W_{97} = 2) \times P(W_{97} = 2) \\
 &\approx 0.917 \times 0.026 \times 0.045 \times 0.048 \\
 &\approx 5.1499\text{E} - 05.
 \end{aligned}$$

Si consideramos los estados 0, 1 o 2 en lugar de 3 y tomando en cuenta la misma sucesión de estados en los tiempos $t = 97, 98, 99$, tenemos

$$P(W_{97} = 2, W_{98} = 1, W_{99} = 2, W_{100} = i) \approx \begin{cases} 0.027 \times 0.026 \times 0.045 \times 0.048 & \text{si } i = 0 \\ 0.03 \times 0.026 \times 0.045 \times 0.048 & \text{si } i = 1 \\ 0.025 \times 0.026 \times 0.045 \times 0.048 & \text{si } i = 2 \end{cases}$$

es decir,

$$P(W_{97} = 2, W_{98} = 1, W_{99} = 2, W_{100} = i) \approx \begin{cases} 1.516\text{E} - 06 & \text{si } i = 0 \\ 1.685\text{E} - 06 & \text{si } i = 1 \\ 1.404\text{E} - 06 & \text{si } i = 2 \end{cases}$$

Así, la probabilidad de tener ninguna excedencia o solo de un tipo de contaminante es de orden diez veces más bajo que tener excedencias de ambos contaminantes. Recordemos que en marzo/abril es primavera y hay muchos incendios forestales.

Ahora, si queremos conocer la probabilidad de tener excedencias de ozono en el día

$t = 250$, entonces sólo tenemos que utilizar los valores de $P(W_{250} = 1)$ y $P(W_{250} = 3)$. Esto dado que el evento {excedencia de ozono presente} puede ser escrito como la unión de los eventos {solamente excedencia de ozono} y {excedencia de ambos contaminantes}, i.e., {excedencia de ozono presente} = {ocurrencia de (1, 0)} \cup {ocurrencia de (1, 1)},

$$\begin{aligned} P(\{\text{excedencia de ozono al día } t = 250\}) &= P(W_{250} = 1) + P(W_{250} = 3) \\ &\approx 0.059 + 0.126 \\ &= 0.635. \end{aligned}$$

Usando el modelo descrito aquí también podemos estudiar de manera conjunta más contaminantes. Por ejemplo, se podría incluir $\text{PM}_{2.5}$, dióxido de nitrógeno, monóxido de carbono, entre otros.

Capítulo 5

Conclusiones

En el trabajo presente se describieron los conceptos básicos de cadenas de Markov no homogéneas, estadística bayesiana y cadenas de orden K con el objetivo de utilizar una cadena no homogénea multivariada de orden K para el estudio de excedencias de contaminantes de interés. El modelo fue aplicado en su versión bivariada para datos recopilados de ozono y PM_{10} en la Ciudad de México, el cual permite la estimación del orden K y las probabilidades iniciales así como las de transición asociadas utilizando el punto de vista bayesiano.

El método utilizado para tratar las observaciones fue transformar el vector de registros de ocurrencias en un indicador unidimensional. Estos constituyen los valores de cada elemento de la información observada. Para estos datos el espacio de estados $S_4^{(K)}$ de la cadena de Markov $X^{(K)}$ así como de las distribuciones iniciales y probabilidades de transición correspondientes estuvieron definidos. La elección del orden de la cadena se llevó a cabo utilizando el máximo a posteriori y el valor seleccionado fue $K = 1$. Dicha elección se reforzó con la aplicación del método SAD a los modelos con mayor log-probabilidad a posteriori. La distribución inicial así como las probabilidades de transición fueron obtenidas usando la media de los valores generados a partir de las correspondientes distribuciones a posteriori.

Usando estas estimaciones, el valor de las probabilidades de tener excedentes de uno,

ambos o ninguno de los contaminantes fue obtenido. Esto fue hecho tomando ventaja de la propiedad de Markov del modelo así como la fórmula recursiva (2.3) dada en la sección 2.0.1 del capítulo 2 que utiliza los valores generados de la distribución inicial y las probabilidades de transición. Los resultados muestran un buen ajuste de los valores estimados a los observados (ver figura 4.5).

Finalmente, se mostró cómo obtener la probabilidad de tener excedencias de ningún, algún o ambos contaminantes en un tiempo dado cuando tenemos información de unos días anteriores, así como otras probabilidades de interés.

De manera general, realizar este proyecto de tesis me permitió un primer acercamiento al trabajo de investigación, siendo esta mi primera experiencia trabajando con modelos bayesianos. A partir de esto, adquirí conocimiento sobre la estadística bayesiana, las cadenas de Markov y todo esto me ha motivado para continuar en el área de la investigación enfocada en estadística.

Apéndice A

En este apéndice presentaremos como proceder en el caso en que el orden de la cadena es mayor que uno. También daremos las tablas con las probabilidades $P(W_t = i), t = 0, 1, 2, 3$, para los casos $K = 1, 2$. Finalmente, presentaremos el código del programa utilizado para la estimación de los parámetros involucrados en el modelo.

A.1. Modelo en el caso $K = 2$

En esta sección presentaremos, para fines ilustrativos, el caso donde tenemos $K = 2$. A pesar de que el caso $K = 2$ no fue elegido, daremos algunos de sus resultados con el fin de ilustrar cómo proceder para obtener las probabilidades $P(W_t = i), i = 0, 1, 2, 3, t \geq 1$, si el modelo seleccionado asigna a la cadena un orden mayor o igual a 2. Cuando $K = 2$, tenemos que $S_4^{(2)} = \{0, 1, \dots, 15\}$.

Así como lo hemos hecho hasta ahora, usando la correspondiente distribución a posteriori Dirichlet de la distribución inicial $Q_{\bar{m}}^{(2)}(1)$ y las probabilidades de transición $P_{\bar{m}_j}^{(2)}(t)$ obtenemos muestras de estas probabilidades y las medias de estas son usadas para estimar las probabilidades respectivas, $\bar{m} \in S_4^{(2)}, j \in S, t = 3, 4, \dots, T - 2$.

Consideremos ahora la probabilidad estimada $P(W_t = i)$ de la ocurrencia de uno de los estados en $S = \{0, 1, 2, 3\}$ a lo largo del año, $i \in S, t = 1, 2, \dots, T$. Así como en el caso cuando $K = 1$, vamos a tomar en cuenta la propiedad de Markov así como la distribución inicial estimada. Como se describió en la sección 4.5, sea $Q_{(i,j)}^{(2,l)}(t) =$

$P^{(l)}(X_t^{(2)} = (i, j))$, $\bar{m} \leftrightarrow (i, j)$, $i, j \in S$, el l -ésimo elemento de las muestras de la distribución a tiempo t de la cadena con orden $K = 2$. Así, para $t = 1, 2$ tenemos que $P^{(l)}(W_1 = j) = \sum_{j \in S} P^{(l)}(X_1^{(2)} = (i, j))$ y $P^{(l)}(W_2 = j) = \sum_{i \in S_2} P^{(l)}(X_1^{(2)} = (i, j))$. Observemos que para $t \geq 3$, podemos escribir para $j, k \in S_2$,

$$\begin{aligned}
P^{(l)}(W_t = k) &= \sum_{i \in S_2} \sum_{j \in S_2} P^{(l)}(W_t = k \mid X_{t-2}^{(2)} = (i, j)) P^{(l)}(X_{t-2}^{(2)} = (i, j)) \\
&= \sum_{i \in S_2} \sum_{j \in S_2} P_{(i,j)k}^{(2,l)}(t-2) P^{(l)}(X_{t-2}^{(2)} = (i, j)), \\
\text{donde } \begin{cases} P^{(l)}(X_1^{(2)} = (i, j)) &= Q_{(i,j)}^{(2,l)}(1), \\ P^{(l)}(X_1^{(2)} = (i, j)) &= \sum_{l \in S_2} P_{(l,i)}^{(2,l)}(t-1) P^{(l)}(X_{t-1}^{(2)} = (l, i)) \quad \text{si } t \geq 2. \end{cases}
\end{aligned}$$

Después de esto, las probabilidades estimadas $\hat{P}(W_t = i), i \in S_2, t = 1, \dots, T$, son obtenidas usando las medias de las muestras simuladas. En la figura A.1 tenemos estas probabilidades estimadas (líneas azules) así como las observadas (líneas negras punteadas) y los intervalos de credibilidad del 95 % (líneas verde y marrón).

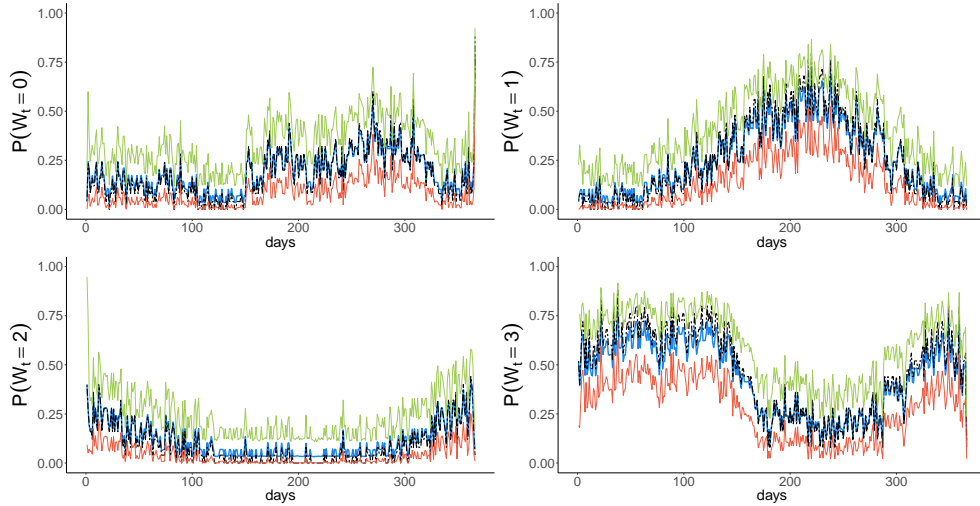


Figura A.1: Distribución estimada (líneas azules) y distribución de las observaciones (línea punteada) de $P(W_t = i)$, $t = 1, \dots, T$, así como los intervalos de credibilidad del 95 % (líneas verde y marrón) con $K = 2$.

De esta forma, podemos hacer estimaciones como aquellas realizadas para el caso $K = 1$. Por ejemplo, si queremos saber el valor de la probabilidad de obtener la secuencia $W_{97} = 2, W_{98} = 1, W_{99} = 2, W_{100} = 3$, entonces necesitamos obtener

$$\begin{aligned}
 P(W_{97} = 2, W_{98} = 1, W_{99} = 2, W_{100} = 3) &= P(W_{100} = 3 \mid W_{99} = 2, W_{98} = 1) \\
 &\quad \times P(W_{99} = 2 \mid W_{98} = 1, W_{97} = 2) \\
 &\quad \times P(W_{98} = 1, W_{97} = 2) \\
 &= P_{(1,2)3}^{(2)}(98) \times P_{(2,1)2}^{(2)}(97) \times Q_{(2,1)}^{(2)}(97),
 \end{aligned}$$

donde las probabilidades de transición son aquellas estimadas y $Q_{(2,1)}^{(2)}(97)$ se obtiene usando la forma recursiva presentada en el capítulo 4. De esta forma, tenemos

$$\begin{aligned}
 P(W_{97} = 2, W_{98} = 1, W_{99} = 2, W_{100} = 3) &= P(W_{100} = 3 \mid W_{99} = 2, W_{98} = 1) \\
 &\approx 0.262 \times 0.243 \times 0.089 \\
 &\approx 5.666\text{E} - 04.
 \end{aligned}$$

A.2. Probabilidades $P(W_t = i)$, $i = 0, 1, 2, 3$

Aquí presentamos las tablas con las probabilidades $P(W_t = i)$ para los casos $K = 1, 2$.

Valores de las probabilidades $P(W_t = i)$ con $k = 1$

	$P(W_t = 0)$	$P(W_t = 1)$	$P(W_t = 2)$	$P(W_t = 3)$
$t = 1$	0.0417	0.0438	0.3924	0.5221
$t = 2$	0.2406	0.0875	0.2709	0.4010
$t = 3$	0.2022	0.0848	0.2003	0.5127
$t = 4$	0.0881	0.0851	0.1625	0.6644
$t = 5$	0.1254	0.0098	0.1631	0.7016
$t = 6$	0.1623	0.0886	0.1262	0.6229
$t = 7$	0.1261	0.0480	0.3562	0.4698
$t = 8$	0.1251	0.0505	0.2037	0.6207
$t = 9$	0.2020	0.0494	0.2813	0.4673
$t = 10$	0.2386	0.0477	0.2030	0.5108
$t = 11$	0.0852	0.0868	0.1627	0.6653
$t = 12$	0.0866	0.0096	0.3920	0.5117
$t = 13$	0.1626	0.0088	0.3575	0.4710
$t = 14$	0.1597	0.0864	0.2815	0.4724
$t = 15$	0.2400	0.0102	0.1599	0.5898
$t = 16$	0.1615	0.0097	0.2382	0.5906
$t = 17$	0.1629	0.0864	0.2713	0.4795
$t = 18$	0.0851	0.1230	0.2798	0.5120
$t = 19$	0.0461	0.0088	0.2794	0.6657
$t = 20$	0.0885	0.0833	0.2797	0.5485
$t = 21$	0.0859	0.1596	0.2025	0.5521
$t = 22$	0.0095	0.0088	0.1641	0.8175
$t = 23$	0.0854	0.0089	0.2045	0.7011
$t = 24$	0.0857	0.0475	0.2790	0.5878
$t = 25$	0.1634	0.0461	0.1597	0.6308
$t = 26$	0.0868	0.0091	0.2767	0.6274
$t = 27$	0.1233	0.0489	0.1631	0.6647
$t = 28$	0.0496	0.0473	0.2350	0.6682
$t = 29$	0.1623	0.0095	0.2393	0.5889
$t = 30$	0.2408	0.0096	0.1650	0.5846
$t = 31$	0.1635	0.0091	0.2422	0.5851
$t = 32$	0.1614	0.0094	0.2445	0.5847
$t = 33$	0.1237	0.0483	0.3198	0.5082
$t = 34$	0.0848	0.0107	0.1252	0.7793
$t = 35$	0.2008	0.1245	0.0857	0.5890
$t = 36$	0.1981	0.0880	0.1219	0.5920
$t = 37$	0.1211	0.0884	0.0478	0.7427
$t = 38$	0.0094	0.0090	0.1250	0.8565

	$P(W_t = 0)$	$P(W_t = 1)$	$P(W_t = 2)$	$P(W_t = 3)$
$t = 39$	0.1242	0.0095	0.2417	0.6246
$t = 40$	0.1594	0.0865	0.1238	0.6303
$t = 41$	0.1252	0.0882	0.1605	0.6260
$t = 42$	0.0879	0.1234	0.0869	0.7018
$t = 43$	0.1633	0.0500	0.2406	0.5460
$t = 44$	0.1250	0.0860	0.1979	0.5912
$t = 45$	0.1221	0.0880	0.1991	0.5908
$t = 46$	0.0863	0.0089	0.2340	0.6708
$t = 47$	0.1198	0.0858	0.0870	0.7074
$t = 48$	0.1587	0.0476	0.1261	0.6675
$t = 49$	0.0853	0.0865	0.0888	0.7394
$t = 50$	0.1210	0.0481	0.0887	0.7422
$t = 51$	0.1237	0.0830	0.1671	0.6261
$t = 52$	0.1619	0.0093	0.2030	0.6257
$t = 53$	0.1228	0.0851	0.0856	0.7065
$t = 54$	0.1267	0.0098	0.1200	0.7435
$t = 55$	0.0483	0.0103	0.1595	0.7819
$t = 56$	0.0476	0.0488	0.1240	0.7796
$t = 57$	0.1636	0.0103	0.1993	0.6268
$t = 58$	0.0466	0.0093	0.2409	0.7031
$t = 59$	0.1235	0.0497	0.2031	0.6237
$t = 60$	0.0861	0.0476	0.1239	0.7423
$t = 61$	0.0865	0.0094	0.1994	0.7048
$t = 62$	0.1253	0.0480	0.1229	0.7037
$t = 63$	0.0486	0.1239	0.0493	0.7782
$t = 64$	0.0496	0.1244	0.1259	0.7000
$t = 65$	0.0881	0.1245	0.1230	0.6643
$t = 66$	0.0476	0.0495	0.1686	0.7343
$t = 67$	0.1273	0.0883	0.1258	0.6586
$t = 68$	0.1269	0.0881	0.0863	0.6986
$t = 69$	0.2021	0.0861	0.0851	0.6266
$t = 70$	0.1260	0.1642	0.1223	0.5876
$t = 71$	0.1661	0.1230	0.0466	0.6643
$t = 72$	0.1647	0.1654	0.0469	0.6230
$t = 73$	0.1640	0.1246	0.0476	0.6639
$t = 74$	0.2403	0.0880	0.0463	0.6253
$t = 75$	0.1263	0.0837	0.2386	0.5514
$t = 76$	0.2048	0.0888	0.1618	0.5445

	$P(W_t = 0)$	$P(W_t = 1)$	$P(W_t = 2)$	$P(W_t = 3)$
$t = 77$	0.1254	0.2069	0.2027	0.4649
$t = 78$	0.0854	0.1259	0.1967	0.5919
$t = 79$	0.1972	0.0488	0.0478	0.7061
$t = 80$	0.1988	0.1677	0.1248	0.5086
$t = 81$	0.1225	0.1671	0.1598	0.5507
$t = 82$	0.0854	0.1637	0.1971	0.5539
$t = 83$	0.0085	0.1220	0.1626	0.7069
$t = 84$	0.0862	0.1242	0.0880	0.7016
$t = 85$	0.0851	0.2792	0.0087	0.6270
$t = 86$	0.1636	0.1234	0.0862	0.6268
$t = 87$	0.1655	0.1249	0.1242	0.5854
$t = 88$	0.0882	0.0864	0.0478	0.7776
$t = 89$	0.2789	0.0492	0.0469	0.6250
$t = 90$	0.0863	0.1272	0.0092	0.7773
$t = 91$	0.1232	0.0857	0.1631	0.6280
$t = 92$	0.1267	0.2059	0.0476	0.6199
$t = 93$	0.0892	0.1669	0.1583	0.5856
$t = 94$	0.0889	0.1280	0.0867	0.6963
$t = 95$	0.0869	0.1659	0.0097	0.7375
$t = 96$	0.0482	0.1644	0.0876	0.6999
$t = 97$	0.1282	0.1247	0.0480	0.6992
$t = 98$	0.1277	0.0878	0.0838	0.7007
$t = 99$	0.0869	0.0878	0.0857	0.7395
$t = 100$	0.1647	0.0857	0.0093	0.7402
$t = 101$	0.0886	0.2402	0.0101	0.6611
$t = 102$	0.1277	0.1241	0.1279	0.6203
$t = 103$	0.0498	0.2385	0.1631	0.5486
$t = 104$	0.1618	0.2027	0.0483	0.5872
$t = 105$	0.0094	0.1648	0.1252	0.7006
$t = 106$	0.0097	0.1252	0.1663	0.6988
$t = 107$	0.0453	0.2090	0.0491	0.6966
$t = 108$	0.0475	0.2806	0.0476	0.6243
$t = 109$	0.0101	0.3185	0.0473	0.6241
$t = 110$	0.0104	0.2431	0.0094	0.7371
$t = 111$	0.0882	0.2018	0.0472	0.6628
$t = 112$	0.0087	0.4000	0.0102	0.5812
$t = 113$	0.0873	0.2778	0.0100	0.6250
$t = 114$	0.0880	0.1629	0.0871	0.6621
$t = 115$	0.1234	0.1634	0.0472	0.6660
$t = 116$	0.0107	0.2007	0.0848	0.7038
$t = 117$	0.0477	0.0850	0.0884	0.7789
$t = 118$	0.0088	0.2029	0.1264	0.6619
$t = 119$	0.0476	0.1999	0.0863	0.6662
$t = 120$	0.0473	0.2376	0.0483	0.6668
$t = 121$	0.0098	0.2763	0.0097	0.7042
$t = 122$	0.0840	0.1244	0.0096	0.7820
$t = 123$	0.0474	0.2376	0.0092	0.7058
$t = 124$	0.0478	0.1977	0.0084	0.7462

	$P(W_t = 0)$	$P(W_t = 1)$	$P(W_t = 2)$	$P(W_t = 3)$
$t = 125$	0.0095	0.2757	0.0095	0.7052
$t = 126$	0.0473	0.1257	0.0465	0.7805
$t = 127$	0.0095	0.2781	0.0475	0.6649
$t = 128$	0.0870	0.2421	0.0472	0.6238
$t = 129$	0.0095	0.2817	0.0480	0.6609
$t = 130$	0.0106	0.2402	0.0092	0.7399
$t = 131$	0.0098	0.2060	0.0484	0.7358
$t = 132$	0.0479	0.2848	0.0101	0.6572
$t = 133$	0.0095	0.3942	0.0106	0.5857
$t = 134$	0.0870	0.3917	0.0091	0.5122
$t = 135$	0.0474	0.2796	0.0475	0.6255
$t = 136$	0.0495	0.3559	0.0487	0.5458
$t = 137$	0.0906	0.3578	0.0101	0.5415
$t = 138$	0.0516	0.2399	0.0104	0.6981
$t = 139$	0.0469	0.3514	0.0487	0.5531
$t = 140$	0.0090	0.4294	0.0093	0.5523
$t = 141$	0.0092	0.2364	0.0480	0.7064
$t = 142$	0.0476	0.2036	0.0479	0.7010
$t = 143$	0.0484	0.2783	0.0087	0.6645
$t = 144$	0.0488	0.3544	0.0106	0.5863
$t = 145$	0.0504	0.3159	0.0108	0.6230
$t = 146$	0.0875	0.4366	0.0484	0.4276
$t = 147$	0.0470	0.3539	0.0103	0.5888
$t = 148$	0.0092	0.4700	0.0094	0.5114
$t = 149$	0.0494	0.4312	0.0472	0.4722
$t = 150$	0.0485	0.2767	0.1239	0.5509
$t = 151$	0.2387	0.2386	0.0490	0.4737
$t = 152$	0.2411	0.2367	0.0103	0.5118
$t = 153$	0.3190	0.2386	0.0093	0.4331
$t = 154$	0.2398	0.3224	0.0486	0.3893
$t = 155$	0.1981	0.3205	0.0092	0.4722
$t = 156$	0.2350	0.2424	0.0488	0.4738
$t = 157$	0.0866	0.4644	0.0098	0.4392
$t = 158$	0.1248	0.4323	0.0092	0.4336
$t = 159$	0.0864	0.4732	0.0104	0.4300
$t = 160$	0.0878	0.4324	0.0485	0.4313
$t = 161$	0.1675	0.3923	0.0094	0.4307
$t = 162$	0.0870	0.5083	0.0101	0.3946
$t = 163$	0.1238	0.4322	0.0091	0.4349
$t = 164$	0.1638	0.4313	0.0105	0.3944
$t = 165$	0.1610	0.4355	0.0469	0.3566
$t = 166$	0.0865	0.5441	0.0097	0.3596
$t = 167$	0.1628	0.5468	0.0105	0.2800
$t = 168$	0.2792	0.4320	0.0091	0.2797
$t = 169$	0.2369	0.4322	0.0104	0.3205
$t = 170$	0.2011	0.5854	0.0105	0.2030
$t = 171$	0.3573	0.3137	0.0871	0.2419
$t = 172$	0.3198	0.3902	0.0473	0.2427

	$P(W_t = 0)$	$P(W_t = 1)$	$P(W_t = 2)$	$P(W_t = 3)$
$t = 173$	0.3985	0.3109	0.0097	0.2809
$t = 174$	0.1638	0.5130	0.0483	0.2750
$t = 175$	0.1650	0.6658	0.0098	0.1594
$t = 176$	0.2858	0.4666	0.0091	0.2385
$t = 177$	0.2806	0.3969	0.0471	0.2753
$t = 178$	0.1684	0.5090	0.0472	0.2754
$t = 179$	0.2425	0.5876	0.0844	0.0855
$t = 180$	0.3158	0.5146	0.0095	0.1601
$t = 181$	0.2780	0.6259	0.0098	0.0863
$t = 182$	0.3176	0.5074	0.0093	0.1657
$t = 183$	0.2786	0.5061	0.0094	0.2059
$t = 184$	0.3214	0.3525	0.0828	0.2432
$t = 185$	0.2011	0.4370	0.0468	0.3150
$t = 186$	0.1989	0.5513	0.0103	0.2396
$t = 187$	0.2014	0.4725	0.0106	0.3154
$t = 188$	0.3164	0.3598	0.0492	0.2746
$t = 189$	0.2372	0.5169	0.0096	0.2363
$t = 190$	0.2409	0.3905	0.0875	0.2811
$t = 191$	0.4340	0.3117	0.0870	0.1673
$t = 192$	0.3931	0.3559	0.0102	0.2408
$t = 193$	0.3153	0.5104	0.0094	0.1650
$t = 194$	0.2410	0.5092	0.0488	0.2010
$t = 195$	0.1658	0.4324	0.0090	0.3929
$t = 196$	0.1266	0.6274	0.0096	0.2363
$t = 197$	0.2384	0.5116	0.0099	0.2400
$t = 198$	0.2830	0.4706	0.0093	0.2372
$t = 199$	0.2772	0.5073	0.0092	0.2063
$t = 200$	0.1615	0.5516	0.0092	0.2778
$t = 201$	0.2024	0.5840	0.0105	0.2031
$t = 202$	0.2775	0.4695	0.0099	0.2431
$t = 203$	0.2012	0.5867	0.0089	0.2032
$t = 204$	0.2046	0.5087	0.0085	0.2781
$t = 205$	0.2784	0.3560	0.0091	0.3565
$t = 206$	0.2402	0.5479	0.0098	0.2021
$t = 207$	0.2004	0.3620	0.0847	0.3530
$t = 208$	0.0891	0.6597	0.0094	0.2419
$t = 209$	0.1264	0.6627	0.0095	0.2014
$t = 210$	0.1256	0.5921	0.0089	0.2734
$t = 211$	0.1257	0.5860	0.0485	0.2399
$t = 212$	0.0850	0.6233	0.0108	0.2810
$t = 213$	0.0902	0.6558	0.0091	0.2449
$t = 214$	0.1276	0.6210	0.0098	0.2415
$t = 215$	0.2425	0.4660	0.0100	0.2815
$t = 216$	0.2786	0.5088	0.0101	0.2024
$t = 217$	0.1605	0.7418	0.0097	0.0880
$t = 218$	0.2416	0.5104	0.0106	0.2374
$t = 219$	0.3227	0.4647	0.0096	0.2030
$t = 220$	0.1303	0.7763	0.0092	0.0841

	$P(W_t = 0)$	$P(W_t = 1)$	$P(W_t = 2)$	$P(W_t = 3)$
$t = 221$	0.2031	0.4702	0.0098	0.3169
$t = 222$	0.3159	0.4713	0.0096	0.2032
$t = 223$	0.2016	0.6241	0.0086	0.1657
$t = 224$	0.1997	0.5498	0.0505	0.2000
$t = 225$	0.3152	0.5105	0.0096	0.1647
$t = 226$	0.2802	0.4703	0.0480	0.2015
$t = 227$	0.1620	0.6278	0.0094	0.2008
$t = 228$	0.2371	0.6686	0.0094	0.0849
$t = 229$	0.1980	0.6681	0.0093	0.1246
$t = 230$	0.1260	0.7036	0.0096	0.1608
$t = 231$	0.2421	0.6639	0.0097	0.0843
$t = 232$	0.2412	0.4715	0.0464	0.2409
$t = 233$	0.2814	0.5481	0.0091	0.1613
$t = 234$	0.3211	0.5087	0.0476	0.1225
$t = 235$	0.3532	0.4350	0.0490	0.1628
$t = 236$	0.2429	0.5094	0.0094	0.2382
$t = 237$	0.2016	0.5840	0.0096	0.2048
$t = 238$	0.1612	0.7394	0.0097	0.0897
$t = 239$	0.1993	0.5889	0.0451	0.1667
$t = 240$	0.1620	0.4706	0.0857	0.2818
$t = 241$	0.3186	0.4304	0.0494	0.2016
$t = 242$	0.0852	0.6674	0.1594	0.0880
$t = 243$	0.2779	0.5130	0.0474	0.1616
$t = 244$	0.1632	0.5842	0.0486	0.2040
$t = 245$	0.2431	0.6189	0.0104	0.1276
$t = 246$	0.2782	0.4331	0.0472	0.2415
$t = 247$	0.2789	0.5088	0.0503	0.1620
$t = 248$	0.3547	0.4713	0.0098	0.1642
$t = 249$	0.3132	0.5090	0.0105	0.1673
$t = 250$	0.3541	0.5099	0.0097	0.1263
$t = 251$	0.2769	0.4714	0.0466	0.2051
$t = 252$	0.2766	0.5125	0.0100	0.2009
$t = 253$	0.3226	0.3917	0.0101	0.2756
$t = 254$	0.3231	0.4673	0.0468	0.1627
$t = 255$	0.3210	0.3902	0.0101	0.2787
$t = 256$	0.2798	0.3866	0.0531	0.2805
$t = 257$	0.4303	0.2783	0.0514	0.2400
$t = 258$	0.3956	0.3154	0.0885	0.2005
$t = 259$	0.3972	0.4681	0.0107	0.1241
$t = 260$	0.3572	0.3965	0.0090	0.2373
$t = 261$	0.3580	0.3936	0.0094	0.2390
$t = 262$	0.2379	0.5923	0.0091	0.1607
$t = 263$	0.2796	0.5510	0.0097	0.1597
$t = 264$	0.2405	0.4762	0.0825	0.2008
$t = 265$	0.3949	0.2814	0.0493	0.2743
$t = 266$	0.2806	0.4316	0.0479	0.2399
$t = 267$	0.3560	0.3544	0.0481	0.2415
$t = 268$	0.3936	0.3133	0.0104	0.2828

	$P(W_t = 0)$	$P(W_t = 1)$	$P(W_t = 2)$	$P(W_t = 3)$
$t = 269$	0.4681	0.3937	0.0103	0.1280
$t = 270$	0.5834	0.2056	0.0485	0.1625
$t = 271$	0.4299	0.2819	0.0487	0.2395
$t = 272$	0.3598	0.4340	0.0479	0.1583
$t = 273$	0.4363	0.3582	0.0094	0.1962
$t = 274$	0.4337	0.2798	0.0888	0.1977
$t = 275$	0.2807	0.3582	0.0852	0.2759
$t = 276$	0.3929	0.3207	0.0470	0.2394
$t = 277$	0.3957	0.2391	0.0467	0.3184
$t = 278$	0.2792	0.3161	0.0465	0.3582
$t = 279$	0.3548	0.3906	0.0894	0.1652
$t = 280$	0.2810	0.4305	0.0497	0.2387
$t = 281$	0.1628	0.3945	0.0872	0.3555
$t = 282$	0.3151	0.5488	0.0109	0.1252
$t = 283$	0.3515	0.4727	0.0098	0.1660
$t = 284$	0.3177	0.4349	0.0101	0.2374
$t = 285$	0.3174	0.3943	0.0466	0.2417
$t = 286$	0.3211	0.3899	0.1264	0.1627
$t = 287$	0.4722	0.3554	0.0854	0.0870
$t = 288$	0.2045	0.4323	0.0090	0.3542
$t = 289$	0.2776	0.2419	0.0484	0.4321
$t = 290$	0.4320	0.2043	0.0102	0.3535
$t = 291$	0.2769	0.2827	0.0472	0.3931
$t = 292$	0.3160	0.2017	0.0465	0.4358
$t = 293$	0.3583	0.2035	0.0840	0.3543
$t = 294$	0.3165	0.2036	0.1261	0.3539
$t = 295$	0.4330	0.0856	0.0892	0.3922
$t = 296$	0.4337	0.1270	0.0099	0.4293
$t = 297$	0.2819	0.2391	0.0498	0.4292
$t = 298$	0.2420	0.2394	0.1638	0.3548
$t = 299$	0.2780	0.2420	0.0878	0.3921
$t = 300$	0.2435	0.2008	0.1224	0.4332
$t = 301$	0.2824	0.2763	0.1240	0.3173
$t = 302$	0.3175	0.3138	0.0506	0.3181
$t = 303$	0.2361	0.3235	0.0096	0.4308
$t = 304$	0.2365	0.3617	0.0472	0.3546
$t = 305$	0.1992	0.2804	0.1246	0.3959
$t = 306$	0.3589	0.2374	0.0882	0.3155
$t = 307$	0.3166	0.3154	0.1666	0.2013
$t = 308$	0.5471	0.0870	0.0848	0.2811
$t = 309$	0.2785	0.2026	0.1589	0.3599
$t = 310$	0.2799	0.1245	0.1243	0.4714
$t = 311$	0.2402	0.1980	0.0861	0.4757
$t = 312$	0.3176	0.0868	0.1634	0.4322
$t = 313$	0.2768	0.1314	0.0853	0.5065
$t = 314$	0.2406	0.1272	0.0836	0.5487
$t = 315$	0.2778	0.2008	0.0865	0.4349
$t = 316$	0.2776	0.1281	0.1593	0.4350
$t = 317$	0.2788	0.0835	0.1254	0.5123

	$P(W_t = 0)$	$P(W_t = 1)$	$P(W_t = 2)$	$P(W_t = 3)$
$t = 318$	0.2800	0.0853	0.1241	0.5106
$t = 319$	0.2407	0.0871	0.1633	0.5089
$t = 320$	0.1248	0.2429	0.1646	0.4677
$t = 321$	0.2417	0.1981	0.0851	0.4751
$t = 322$	0.2423	0.1207	0.1250	0.5121
$t = 323$	0.1995	0.1622	0.1242	0.5141
$t = 324$	0.2395	0.1652	0.0481	0.5472
$t = 325$	0.1595	0.0092	0.2449	0.5864
$t = 326$	0.2372	0.1267	0.2017	0.4343
$t = 327$	0.1614	0.0460	0.0889	0.7038
$t = 328$	0.1277	0.0828	0.2335	0.5560
$t = 329$	0.1227	0.1235	0.1264	0.6274
$t = 330$	0.1239	0.0847	0.2828	0.5086
$t = 331$	0.1238	0.0483	0.3175	0.5105
$t = 332$	0.0860	0.1644	0.2046	0.5451
$t = 333$	0.0893	0.0855	0.1599	0.6652
$t = 334$	0.0893	0.0472	0.2771	0.5865
$t = 335$	0.0099	0.2395	0.2794	0.4712
$t = 336$	0.1286	0.0468	0.2413	0.5834
$t = 337$	0.1242	0.0847	0.2068	0.5843
$t = 338$	0.0880	0.0858	0.1641	0.6621
$t = 339$	0.1663	0.0469	0.0862	0.7006
$t = 340$	0.2053	0.0105	0.2389	0.5452
$t = 341$	0.0477	0.0452	0.3213	0.5858
$t = 342$	0.0870	0.0100	0.2788	0.6241
$t = 343$	0.1253	0.0487	0.1275	0.6986
$t = 344$	0.1271	0.0478	0.1667	0.6584
$t = 345$	0.0486	0.0492	0.3169	0.5852
$t = 346$	0.1265	0.0861	0.3163	0.4711
$t = 347$	0.1999	0.0483	0.2401	0.5117
$t = 348$	0.0476	0.0098	0.3574	0.5852
$t = 349$	0.0882	0.0478	0.0892	0.7748
$t = 350$	0.0870	0.0480	0.2821	0.5829
$t = 351$	0.0865	0.0861	0.2799	0.5475
$t = 352$	0.0477	0.0862	0.1988	0.6673
$t = 353$	0.0497	0.1556	0.2074	0.5873
$t = 354$	0.1967	0.0100	0.2824	0.5109
$t = 355$	0.1269	0.0471	0.3927	0.4333
$t = 356$	0.1251	0.0491	0.2750	0.5507
$t = 357$	0.2398	0.0462	0.2376	0.4764
$t = 358$	0.0856	0.0456	0.3526	0.5161
$t = 359$	0.0836	0.0096	0.1204	0.7864
$t = 360$	0.3123	0.0881	0.2032	0.3963
$t = 361$	0.0857	0.1211	0.3185	0.4747
$t = 362$	0.1213	0.0834	0.4391	0.3561
$t = 363$	0.0850	0.0095	0.3943	0.5112
$t = 364$	0.1268	0.0493	0.3164	0.5074
$t = 365$	0.2464	0.0844	0.1236	0.5456
$t = 366$	0.8581	0.0098	0.0491	0.0830

Valores de las probabilidades $P(W_t = i)$ con $k = 2$

	$P(W_t = 0)$	$P(W_t = 1)$	$P(W_t = 2)$	$P(W_t = 3)$
$t = 1$	0.0454	0.0475	0.3943	0.5128
$t = 2$	0.2433	0.0841	0.2787	0.3938
$t = 3$	0.2091	0.1020	0.1914	0.4975
$t = 4$	0.0999	0.0977	0.1691	0.6332
$t = 5$	0.1389	0.0316	0.1756	0.6539
$t = 6$	0.1724	0.1020	0.1383	0.5873
$t = 7$	0.1379	0.0684	0.3465	0.4471
$t = 8$	0.1365	0.0687	0.2084	0.5864
$t = 9$	0.2072	0.0687	0.2742	0.4500
$t = 10$	0.2415	0.0675	0.2091	0.4820
$t = 11$	0.1005	0.1040	0.1781	0.6174
$t = 12$	0.1018	0.0345	0.3818	0.4818
$t = 13$	0.1695	0.0340	0.3486	0.4479
$t = 14$	0.1737	0.1027	0.2801	0.4435
$t = 15$	0.2434	0.0355	0.1744	0.5467
$t = 16$	0.1750	0.0345	0.2404	0.5502
$t = 17$	0.1745	0.1060	0.2748	0.4447
$t = 18$	0.1051	0.1376	0.2778	0.4796
$t = 19$	0.0665	0.0342	0.2810	0.6182
$t = 20$	0.1061	0.1001	0.2755	0.5182
$t = 21$	0.1047	0.1691	0.2079	0.5183
$t = 22$	0.0342	0.0360	0.1750	0.7548
$t = 23$	0.1061	0.0355	0.2064	0.6520
$t = 24$	0.1047	0.0697	0.2728	0.5529
$t = 25$	0.1700	0.0677	0.1718	0.5905
$t = 26$	0.1013	0.0367	0.2747	0.5872
$t = 27$	0.1372	0.0689	0.1727	0.6212
$t = 28$	0.0710	0.0672	0.2415	0.6203
$t = 29$	0.1736	0.0344	0.2407	0.5514
$t = 30$	0.2430	0.0366	0.1719	0.5485
$t = 31$	0.1722	0.0325	0.2421	0.5532
$t = 32$	0.1757	0.0350	0.2361	0.5532
$t = 33$	0.1401	0.0666	0.3082	0.4852
$t = 34$	0.1030	0.0348	0.1403	0.7220
$t = 35$	0.2080	0.1373	0.1015	0.5532
$t = 36$	0.2113	0.1018	0.1391	0.5477
$t = 37$	0.1402	0.1006	0.0707	0.6885
$t = 38$	0.0358	0.0356	0.1377	0.7909
$t = 39$	0.1375	0.0345	0.2393	0.5886
$t = 40$	0.1747	0.1037	0.1376	0.5841
$t = 41$	0.1380	0.1015	0.1709	0.5896
$t = 42$	0.1063	0.1364	0.0975	0.6598
$t = 43$	0.1728	0.0716	0.2384	0.5172

	$P(W_t = 0)$	$P(W_t = 1)$	$P(W_t = 2)$	$P(W_t = 3)$
$t = 44$	0.1402	0.1047	0.2052	0.5499
$t = 45$	0.1377	0.1031	0.2097	0.5495
$t = 46$	0.1003	0.0343	0.2428	0.6226
$t = 47$	0.1379	0.1006	0.1033	0.6582
$t = 48$	0.1725	0.0695	0.1367	0.6213
$t = 49$	0.1019	0.1018	0.1063	0.6900
$t = 50$	0.1371	0.0698	0.1049	0.6881
$t = 51$	0.1397	0.1051	0.1703	0.5849
$t = 52$	0.1733	0.0349	0.2067	0.5852
$t = 53$	0.1379	0.1065	0.1056	0.6500
$t = 54$	0.1375	0.0369	0.1399	0.6857
$t = 55$	0.0702	0.0351	0.1741	0.7206
$t = 56$	0.0703	0.0707	0.1391	0.7199
$t = 57$	0.1732	0.0349	0.2090	0.5829
$t = 58$	0.0676	0.0365	0.2474	0.6485
$t = 59$	0.1407	0.0679	0.2059	0.5855
$t = 60$	0.1073	0.0692	0.1388	0.6848
$t = 61$	0.1044	0.0329	0.2123	0.6504
$t = 62$	0.1390	0.0701	0.1407	0.6502
$t = 63$	0.0694	0.1381	0.0700	0.7225
$t = 64$	0.0692	0.1366	0.1432	0.6510
$t = 65$	0.1020	0.1378	0.1367	0.6235
$t = 66$	0.0676	0.0698	0.1736	0.6889
$t = 67$	0.1343	0.1015	0.1384	0.6258
$t = 68$	0.1363	0.1010	0.1031	0.6595
$t = 69$	0.2035	0.1050	0.1014	0.5901
$t = 70$	0.1331	0.1752	0.1358	0.5559
$t = 71$	0.1711	0.1384	0.0659	0.6247
$t = 72$	0.1742	0.1707	0.0650	0.5901
$t = 73$	0.1725	0.1380	0.0679	0.6216
$t = 74$	0.2385	0.1029	0.0693	0.5893
$t = 75$	0.1344	0.1053	0.2448	0.5155
$t = 76$	0.2082	0.1020	0.1725	0.5174
$t = 77$	0.1360	0.2064	0.2085	0.4491
$t = 78$	0.1049	0.1363	0.2105	0.5482
$t = 79$	0.2104	0.0680	0.0690	0.6527
$t = 80$	0.2101	0.1735	0.1378	0.4786
$t = 81$	0.1381	0.1782	0.1658	0.5179
$t = 82$	0.0994	0.1741	0.2092	0.5173
$t = 83$	0.0346	0.1381	0.1708	0.6565
$t = 84$	0.1068	0.1341	0.1063	0.6528
$t = 85$	0.1031	0.2824	0.0333	0.5813
$t = 86$	0.1751	0.1388	0.1070	0.5791

	$P(W_t = 0)$	$P(W_t = 1)$	$P(W_t = 2)$	$P(W_t = 3)$
$t = 87$	0.1727	0.1409	0.1401	0.5464
$t = 88$	0.1048	0.1029	0.0719	0.7203
$t = 89$	0.2834	0.0677	0.0670	0.5818
$t = 90$	0.1055	0.1382	0.0337	0.7226
$t = 91$	0.1444	0.1048	0.1718	0.5790
$t = 92$	0.1387	0.2086	0.0662	0.5865
$t = 93$	0.1037	0.1745	0.1750	0.5469
$t = 94$	0.1035	0.1407	0.1022	0.6536
$t = 95$	0.1010	0.1682	0.0345	0.6963
$t = 96$	0.0694	0.1706	0.1039	0.6562
$t = 97$	0.1352	0.1400	0.0689	0.6560
$t = 98$	0.1365	0.1036	0.1045	0.6554
$t = 99$	0.1037	0.1034	0.1035	0.6894
$t = 100$	0.1692	0.1065	0.0354	0.6890
$t = 101$	0.1011	0.2426	0.0361	0.6203
$t = 102$	0.1385	0.1375	0.1390	0.5850
$t = 103$	0.0706	0.2388	0.1731	0.5175
$t = 104$	0.1767	0.2026	0.0679	0.5527
$t = 105$	0.0335	0.1748	0.1380	0.6537
$t = 106$	0.0339	0.1345	0.1757	0.6559
$t = 107$	0.0687	0.2063	0.0696	0.6554
$t = 108$	0.0677	0.2695	0.0696	0.5932
$t = 109$	0.0353	0.3134	0.0662	0.5852
$t = 110$	0.0346	0.2451	0.0330	0.6873
$t = 111$	0.1045	0.2076	0.0694	0.6185
$t = 112$	0.0335	0.3776	0.0348	0.5542
$t = 113$	0.1023	0.2777	0.0330	0.5869
$t = 114$	0.1026	0.1723	0.1033	0.6218
$t = 115$	0.1387	0.1731	0.0677	0.6205
$t = 116$	0.0324	0.2088	0.1037	0.6552
$t = 117$	0.0692	0.1040	0.1045	0.7223
$t = 118$	0.0351	0.2074	0.1351	0.6224
$t = 119$	0.0697	0.2063	0.0999	0.6241
$t = 120$	0.0678	0.2389	0.0706	0.6227
$t = 121$	0.0336	0.2772	0.0337	0.6555
$t = 122$	0.1036	0.1386	0.0346	0.7233
$t = 123$	0.0692	0.2442	0.0326	0.6540
$t = 124$	0.0678	0.2059	0.0354	0.6908
$t = 125$	0.0341	0.2772	0.0367	0.6520
$t = 126$	0.0710	0.1405	0.0693	0.7192
$t = 127$	0.0364	0.2741	0.0695	0.6201
$t = 128$	0.1026	0.2409	0.0690	0.5875
$t = 129$	0.0355	0.2757	0.0692	0.6197
$t = 130$	0.0346	0.2407	0.0349	0.6898
$t = 131$	0.0343	0.2095	0.0680	0.6882
$t = 132$	0.0681	0.2789	0.0349	0.6181
$t = 133$	0.0350	0.3770	0.0354	0.5526
$t = 134$	0.1023	0.3804	0.0358	0.4814

	$P(W_t = 0)$	$P(W_t = 1)$	$P(W_t = 2)$	$P(W_t = 3)$
$t = 135$	0.0678	0.2780	0.0690	0.5852
$t = 136$	0.0710	0.3467	0.0675	0.5147
$t = 137$	0.1069	0.3477	0.0359	0.5096
$t = 138$	0.0700	0.2398	0.0351	0.6550
$t = 139$	0.0675	0.3455	0.0674	0.5197
$t = 140$	0.0346	0.4119	0.0352	0.5183
$t = 141$	0.0341	0.2449	0.0675	0.6536
$t = 142$	0.0688	0.2088	0.0690	0.6534
$t = 143$	0.0690	0.2750	0.0349	0.6212
$t = 144$	0.0696	0.3440	0.0351	0.5512
$t = 145$	0.0698	0.3043	0.0354	0.5905
$t = 146$	0.1032	0.4147	0.0688	0.4133
$t = 147$	0.0696	0.3433	0.0334	0.5538
$t = 148$	0.0349	0.4525	0.0357	0.4770
$t = 149$	0.0704	0.4133	0.0672	0.4491
$t = 150$	0.0668	0.2774	0.1354	0.5204
$t = 151$	0.2424	0.2413	0.0684	0.4478
$t = 152$	0.2430	0.2409	0.0354	0.4807
$t = 153$	0.3081	0.2432	0.0349	0.4139
$t = 154$	0.2431	0.3082	0.0681	0.3806
$t = 155$	0.2085	0.3067	0.0341	0.4507
$t = 156$	0.2407	0.2368	0.0682	0.4543
$t = 157$	0.1038	0.4491	0.0334	0.4136
$t = 158$	0.1378	0.4138	0.0333	0.4151
$t = 159$	0.1036	0.4504	0.0334	0.4126
$t = 160$	0.1029	0.4153	0.0682	0.4136
$t = 161$	0.1696	0.3797	0.0349	0.4157
$t = 162$	0.1042	0.4808	0.0337	0.3813
$t = 163$	0.1377	0.4110	0.0349	0.4164
$t = 164$	0.1708	0.4136	0.0341	0.3814
$t = 165$	0.1705	0.4104	0.0704	0.3488
$t = 166$	0.1013	0.5143	0.0361	0.3483
$t = 167$	0.1735	0.5143	0.0341	0.2781
$t = 168$	0.2769	0.4159	0.0356	0.2717
$t = 169$	0.2422	0.4121	0.0342	0.3116
$t = 170$	0.2045	0.5541	0.0345	0.2070
$t = 171$	0.3466	0.3095	0.1046	0.2393
$t = 172$	0.3108	0.3812	0.0687	0.2393
$t = 173$	0.3781	0.3123	0.0347	0.2748
$t = 174$	0.1690	0.4834	0.0693	0.2783
$t = 175$	0.1718	0.6199	0.0350	0.1733
$t = 176$	0.2790	0.4439	0.0349	0.2422
$t = 177$	0.2784	0.3760	0.0700	0.2755
$t = 178$	0.1726	0.4798	0.0704	0.2772
$t = 179$	0.2456	0.5460	0.1061	0.1023
$t = 180$	0.3108	0.4822	0.0353	0.1716
$t = 181$	0.2741	0.5901	0.0338	0.1020
$t = 182$	0.3100	0.4825	0.0348	0.1727

	$P(W_t = 0)$	$P(W_t = 1)$	$P(W_t = 2)$	$P(W_t = 3)$
$t = 183$	0.2770	0.4782	0.0362	0.2086
$t = 184$	0.3105	0.3481	0.1012	0.2402
$t = 185$	0.2078	0.4153	0.0683	0.3086
$t = 186$	0.2062	0.5185	0.0359	0.2394
$t = 187$	0.2046	0.4540	0.0353	0.3061
$t = 188$	0.3087	0.3504	0.0703	0.2706
$t = 189$	0.2443	0.4825	0.0354	0.2379
$t = 190$	0.2403	0.3858	0.1015	0.2725
$t = 191$	0.4166	0.3090	0.1026	0.1718
$t = 192$	0.3836	0.3439	0.0343	0.2382
$t = 193$	0.3089	0.4878	0.0325	0.1708
$t = 194$	0.2429	0.4816	0.0672	0.2083
$t = 195$	0.1757	0.4105	0.0343	0.3795
$t = 196$	0.1420	0.5868	0.0335	0.2378
$t = 197$	0.2433	0.4800	0.0352	0.2415
$t = 198$	0.2770	0.4455	0.0360	0.2414
$t = 199$	0.2757	0.4832	0.0362	0.2049
$t = 200$	0.1716	0.5161	0.0356	0.2768
$t = 201$	0.2043	0.5512	0.0338	0.2107
$t = 202$	0.2750	0.4451	0.0345	0.2455
$t = 203$	0.2031	0.5529	0.0333	0.2106
$t = 204$	0.2075	0.4778	0.0335	0.2812
$t = 205$	0.2716	0.3467	0.0350	0.3468
$t = 206$	0.2426	0.5144	0.0348	0.2082
$t = 207$	0.2088	0.3471	0.1010	0.3430
$t = 208$	0.1044	0.6198	0.0352	0.2406
$t = 209$	0.1346	0.6251	0.0346	0.2057
$t = 210$	0.1423	0.5501	0.0343	0.2733
$t = 211$	0.1396	0.5543	0.0688	0.2373
$t = 212$	0.1008	0.5891	0.0331	0.2770
$t = 213$	0.1012	0.6233	0.0338	0.2418
$t = 214$	0.1375	0.5868	0.0332	0.2425
$t = 215$	0.2413	0.4460	0.0319	0.2808
$t = 216$	0.2797	0.4819	0.0343	0.2041
$t = 217$	0.1730	0.6911	0.0330	0.1029
$t = 218$	0.2426	0.4815	0.0345	0.2414
$t = 219$	0.3095	0.4477	0.0352	0.2077
$t = 220$	0.1370	0.7250	0.0339	0.1041
$t = 221$	0.2056	0.4528	0.0340	0.3075
$t = 222$	0.3098	0.4525	0.0359	0.2019
$t = 223$	0.2097	0.5827	0.0330	0.1747
$t = 224$	0.2079	0.5126	0.0693	0.2102
$t = 225$	0.3096	0.4783	0.0362	0.1759
$t = 226$	0.2752	0.4432	0.0727	0.2089
$t = 227$	0.1703	0.5859	0.0338	0.2101
$t = 228$	0.2441	0.6187	0.0353	0.1018
$t = 229$	0.2100	0.6177	0.0353	0.1370
$t = 230$	0.1381	0.6546	0.0351	0.1722

	$P(W_t = 0)$	$P(W_t = 1)$	$P(W_t = 2)$	$P(W_t = 3)$
$t = 231$	0.2400	0.6219	0.0339	0.1042
$t = 232$	0.2394	0.4488	0.0719	0.2399
$t = 233$	0.2772	0.5154	0.0338	0.1736
$t = 234$	0.3085	0.4848	0.0700	0.1367
$t = 235$	0.3451	0.4135	0.0689	0.1726
$t = 236$	0.2447	0.4829	0.0332	0.2392
$t = 237$	0.2098	0.5507	0.0353	0.2043
$t = 238$	0.1743	0.6898	0.0326	0.1033
$t = 239$	0.2077	0.5528	0.0706	0.1689
$t = 240$	0.1747	0.4542	0.1013	0.2698
$t = 241$	0.3097	0.4158	0.0685	0.2059
$t = 242$	0.1018	0.6262	0.1688	0.1032
$t = 243$	0.2746	0.4840	0.0705	0.1709
$t = 244$	0.1734	0.5509	0.0699	0.2057
$t = 245$	0.2414	0.5893	0.0348	0.1344
$t = 246$	0.2788	0.4111	0.0693	0.2408
$t = 247$	0.2741	0.4851	0.0675	0.1734
$t = 248$	0.3426	0.4483	0.0324	0.1767
$t = 249$	0.3125	0.4837	0.0337	0.1701
$t = 250$	0.3401	0.4885	0.0352	0.1362
$t = 251$	0.2692	0.4545	0.0692	0.2071
$t = 252$	0.2708	0.4866	0.0358	0.2069
$t = 253$	0.3127	0.3786	0.0339	0.2748
$t = 254$	0.3098	0.4534	0.0680	0.1687
$t = 255$	0.3128	0.3795	0.0327	0.2750
$t = 256$	0.2768	0.3800	0.0693	0.2739
$t = 257$	0.4120	0.2775	0.0694	0.2411
$t = 258$	0.3832	0.3084	0.1047	0.2037
$t = 259$	0.3825	0.4470	0.0347	0.1357
$t = 260$	0.3469	0.3792	0.0346	0.2393
$t = 261$	0.3463	0.3788	0.0338	0.2411
$t = 262$	0.2389	0.5539	0.0336	0.1736
$t = 263$	0.2750	0.5214	0.0331	0.1705
$t = 264$	0.2420	0.4503	0.1032	0.2044
$t = 265$	0.3828	0.2789	0.0663	0.2720
$t = 266$	0.2792	0.4142	0.0684	0.2382
$t = 267$	0.3482	0.3445	0.0677	0.2396
$t = 268$	0.3805	0.3122	0.0343	0.2729
$t = 269$	0.4527	0.3792	0.0341	0.1340
$t = 270$	0.5526	0.2085	0.0674	0.1716
$t = 271$	0.4146	0.2776	0.0673	0.2405
$t = 272$	0.3450	0.4139	0.0692	0.1718
$t = 273$	0.4170	0.3435	0.0331	0.2064
$t = 274$	0.4125	0.2768	0.1080	0.2027
$t = 275$	0.2741	0.3466	0.1029	0.2764
$t = 276$	0.3803	0.3102	0.0692	0.2404
$t = 277$	0.3751	0.2415	0.0694	0.3140
$t = 278$	0.2748	0.3071	0.0676	0.3505

	$P(W_t = 0)$	$P(W_t = 1)$	$P(W_t = 2)$	$P(W_t = 3)$
$t = 279$	0.3430	0.3807	0.1038	0.1725
$t = 280$	0.2757	0.4127	0.0675	0.2442
$t = 281$	0.1720	0.3797	0.1025	0.3458
$t = 282$	0.3153	0.5144	0.0338	0.1365
$t = 283$	0.3467	0.4460	0.0374	0.1699
$t = 284$	0.3075	0.4165	0.0350	0.2410
$t = 285$	0.3102	0.3815	0.0688	0.2395
$t = 286$	0.3073	0.3815	0.1383	0.1728
$t = 287$	0.4472	0.3473	0.1035	0.1021
$t = 288$	0.2052	0.4150	0.0345	0.3452
$t = 289$	0.2761	0.2389	0.0703	0.4147
$t = 290$	0.4157	0.2045	0.0345	0.3453
$t = 291$	0.2768	0.2751	0.0684	0.3797
$t = 292$	0.3108	0.2060	0.0680	0.4153
$t = 293$	0.3440	0.2046	0.1037	0.3478
$t = 294$	0.3096	0.2043	0.1391	0.3469
$t = 295$	0.4140	0.0996	0.1051	0.3813
$t = 296$	0.4128	0.1404	0.0345	0.4123
$t = 297$	0.2744	0.2453	0.0678	0.4124
$t = 298$	0.2426	0.2444	0.1712	0.3418
$t = 299$	0.2742	0.2441	0.1023	0.3795
$t = 300$	0.2385	0.2088	0.1351	0.4176
$t = 301$	0.2753	0.2718	0.1369	0.3161
$t = 302$	0.3081	0.3085	0.0681	0.3153
$t = 303$	0.2400	0.3102	0.0342	0.4156
$t = 304$	0.2388	0.3450	0.0668	0.3494
$t = 305$	0.2057	0.2762	0.1397	0.3784
$t = 306$	0.3458	0.2439	0.1041	0.3062
$t = 307$	0.3149	0.3078	0.1723	0.2049
$t = 308$	0.5205	0.1020	0.1018	0.2756
$t = 309$	0.2769	0.2073	0.1743	0.3415
$t = 310$	0.2760	0.1344	0.1397	0.4498
$t = 311$	0.2443	0.2039	0.1028	0.4490
$t = 312$	0.3102	0.1028	0.1712	0.4158
$t = 313$	0.2779	0.1383	0.1024	0.4815
$t = 314$	0.2435	0.1381	0.1052	0.5132
$t = 315$	0.2764	0.2052	0.1063	0.4121
$t = 316$	0.2755	0.1409	0.1763	0.4073
$t = 317$	0.2765	0.1055	0.1373	0.4807
$t = 318$	0.2747	0.1037	0.1415	0.4801
$t = 319$	0.2408	0.1035	0.1707	0.4851
$t = 320$	0.1390	0.2420	0.1742	0.4448
$t = 321$	0.2431	0.2048	0.1031	0.4490
$t = 322$	0.2419	0.1385	0.1388	0.4809

	$P(W_t = 0)$	$P(W_t = 1)$	$P(W_t = 2)$	$P(W_t = 3)$
$t = 323$	0.2095	0.1704	0.1382	0.4819
$t = 324$	0.2423	0.1712	0.0688	0.5176
$t = 325$	0.1727	0.0342	0.2423	0.5508
$t = 326$	0.2407	0.1387	0.2036	0.4169
$t = 327$	0.1751	0.0691	0.1049	0.6510
$t = 328$	0.1372	0.1059	0.2436	0.5133
$t = 329$	0.1367	0.1371	0.1381	0.5881
$t = 330$	0.1382	0.1049	0.2733	0.4836
$t = 331$	0.1368	0.0694	0.3056	0.4881
$t = 332$	0.1031	0.1684	0.2037	0.5248
$t = 333$	0.1018	0.1001	0.1726	0.6256
$t = 334$	0.1026	0.0664	0.2760	0.5551
$t = 335$	0.0331	0.2388	0.2814	0.4467
$t = 336$	0.1372	0.0662	0.2469	0.5497
$t = 337$	0.1391	0.1024	0.2109	0.5476
$t = 338$	0.1053	0.1027	0.1724	0.6195
$t = 339$	0.1739	0.0679	0.1054	0.6528
$t = 340$	0.2082	0.0340	0.2415	0.5162
$t = 341$	0.0683	0.0694	0.3062	0.5561
$t = 342$	0.1031	0.0323	0.2748	0.5898
$t = 343$	0.1375	0.0685	0.1346	0.6594
$t = 344$	0.1382	0.0676	0.1716	0.6226
$t = 345$	0.0700	0.0703	0.3067	0.5529
$t = 346$	0.1384	0.1039	0.3096	0.4482
$t = 347$	0.2060	0.0703	0.2377	0.4860
$t = 348$	0.0680	0.0346	0.3375	0.5599
$t = 349$	0.1061	0.0676	0.1009	0.7254
$t = 350$	0.1036	0.0723	0.2764	0.5477
$t = 351$	0.1069	0.1030	0.2772	0.5130
$t = 352$	0.0691	0.1075	0.2063	0.6172
$t = 353$	0.0695	0.1764	0.2032	0.5509
$t = 354$	0.2071	0.0346	0.2747	0.4836
$t = 355$	0.1375	0.0691	0.3733	0.4201
$t = 356$	0.1358	0.0689	0.2737	0.5216
$t = 357$	0.2425	0.0662	0.2389	0.4525
$t = 358$	0.1031	0.0669	0.3445	0.4855
$t = 359$	0.1041	0.0337	0.1354	0.7268
$t = 360$	0.3136	0.1037	0.2045	0.3781
$t = 361$	0.1039	0.1386	0.3103	0.4472
$t = 362$	0.1395	0.1034	0.4132	0.3439
$t = 363$	0.1056	0.0353	0.3817	0.4774
$t = 364$	0.1392	0.0693	0.3099	0.4816
$t = 365$	0.2428	0.1055	0.1364	0.5153
$t = 366$	0.7938	0.0338	0.0693	0.1031

A.3. Código Fuente

Simulación ("fcts&DataToUse.R")

```
1 library(readr) #It helps to read and write csv in such a faster way
2
3 library(gtools) #Used for all the functions
4
5 library(bayestestR) #We will use the function ci to obtain the
6 #inferior and superior limit of the correspondent credible intervals
7
8 library(ggplot2) #we will need this library to plot the data from
9 #the dataframes
10
11 library(gridExtra) #This will help us to arrange the possible
12 #values in one big arrange
13
14 vec2scal <- function(vec,maxStateSpace=1,
15                       totalValSS=maxStateSpace+1){
16   #if(any(vec > maxStateSpace)){
17     # warning("One of the values is out of the space state")
18   #}
19   powerLevels <- rep(0,length(vec))
20   for(i in 1:length(vec)){
21     powerLevels[i] <- totalValSS^(i-1)
22   }
23   goodOrder <- sort(powerLevels,T)
24   #Sin rev(vec) da los escalares provenientes del vector (O3,PM10)
25   #return(sum(goodOrder*vec))
26
27   #Con rev(vec) da los escalares provenientes de los vectores
28   #\overline{m}
29   return(sum(goodOrder*rev(vec)))
30 }
31
32
33 #####First function####
34
35 #####Function wich gives me the counts of different ocurrences respect
36 #to a given matrix, the Order of the Markov Chain, the maximum value
37 #and the total of elements in our State Space###
38 listOfCountsGen <- function(givenmat,EstimOrder=3,
39                             maxStateSpace=3,
40                             totalValSS=maxStateSpace+1){
```

```

41 #Here we will store the final result
42 result <- array(0,c(totalValSS^EstimOrder,totalValSS,ncol(givenmat)-EstimOrder))
43 #With this we will count how long does it take the function to obtain
44 #the matrices (This is merely optional and can be easily removed)
45 progbar <- txtProgressBar(min=0,max=(ncol(givenmat)-EstimOrder),
46                           style=3,char="=")
47 #For each t in time, we will proceed to obtain the respective count
48 #matrix
49 if(EstimOrder==0){
50   #result <- array(0,c(totalValSS,ncol(givenmat)))
51   for(t in 1:ncol(givenmat)){
52     #For each t in time, we fix the occurrences of the possible values
53     #of the State Space, and we save it in the array
54
55     #result[,t] <- table(givenmat[,t])
56
57     counter <- rep(0,totalValSS)
58     #Here we count each of the possible values at the t+EstimOrder column
59     respectiveValues <- table(givenmat[,t])
60     for(k in 1:totalValSS){
61       #Here we save the total of each possible value at the given selected column
62       counter[as.numeric(names(respectiveValues[k]))+1] <-
63         as.numeric(respectiveValues[k])
64     }
65
66     #Finally, we store the matrix we created to his correspondent place
67     #in our array
68     result[,t] <- counter
69
70     #This is for observing how the algorithm works with the time
71     setTxtProgressBar(progbar,value=t)
72   }
73   #We add some names to indicate the correspondent sum of each
74   #State Space value
75   #dimnames(result) <- list(c(),0:maxStateSpace)
76   #dimnames(result) <- list(c(),c(0:maxStateSpace))
77   return(result)
78 }
79 else if(EstimOrder!=0){
80   ###With this I obtain the (totalValSS)^EstimOrder EstimOrder-Vectors
81   #with possible values {0,1,...,maxStateSpace}
82
83   mykStateLevels <- permutations(totalValSS,EstimOrder,
84                                 v=0:maxStateSpace,
85                                 repeats.allowed = TRUE)

```

```

86
87   if(EstimOrder!=1){
88     mykStateLevels <- t(apply(mykStateLevels,1,rev))
89
90   }
91
92
93   for(t in 1:(ncol(givenmat)-EstimOrder)){
94
95     #Here we'll save our EstimOrder-column selected-correspondent count
96     #matrix
97     CounterMat <- matrix(0,nrow(mykStateLevels),totalValSS)
98
99
100    #Here we obtain the different values of the three columns selected
101    #with respect to time and we classify them as a factor
102    simplifyCols <- apply(as.matrix(givenmat[,t:(t+EstimOrder-1)]),
103      1,paste,collapse=".")
104
105    simplifyCols_fac <- factor(simplifyCols,
106      levels=apply(mykStateLevels,1,paste,
107        collapse="."),
108      labels=1:nrow(mykStateLevels))
109
110    #Here we obtain the values from simplifyCols_fac with no repetition
111    uniqueObs <- unique(c(simplifyCols_fac))
112    #And here we start to modify our CounterMat matrix only in the rows
113    #we know they appear as one of the EstimOrder-columns
114    for(j in uniqueObs){
115      #Here we will save the occurrences of each value of the correspondent
116      #EstimOrder-column
117      counter <- rep(0,totalValSS)
118      #Here we count each of the possible values at the t+EstimOrder column
119      respectiveValues <- table(givenmat[simplifyCols_fac == j,(t+EstimOrder)])
120      for(k in 1:totalValSS){
121        #Here we save the total of each possible value at the given selected column
122        counter[as.numeric(names(respectiveValues[k]))+1] <-
123          as.numeric(respectiveValues[k])
124      }
125      #Finally, we start to bind the obtained row with the corresponding
126      #value of (totalValSS)^EstimOrder possible choices
127      CounterMat[j,] <- counter
128
129    }
130    #Finally, we store the matrix we created to his correspondent place

```

```

131     #in our array
132     result[, ,t] <- CounterMat
133     setTxtProgressBar(progbar,value=t)
134
135 }
136 dimnames(result) <- list(apply(mykStateLevels,1,
137                               paste,collapse=""),c(0:maxStateSpace))
138 #We add the resulted count matrix to our list of count matrices
139 return(result)
140 }
141 }
142
143
144 #####Second function####
145
146 #Function that generates for each time, four matrices, each one corresponding
147 #to one specific value for a vector m at time t.
148 my4MatsGen <- function(givenmat, countsList,EstimOrder=3,
149                        numSamples=1000,
150                        maxStateSpace=3,
151                        totalValSS=maxStateSpace+1){
152
153     #This is just for looking the time the function takes to make all
154     #the proceed (not really necessary)
155     progbar <- txtProgressBar(min=0,max=(ncol(givenmat)-EstimOrder),
156                              style=3,char="=")
157
158     if(EstimOrder!=0){
159         #Like in the first function, we set the (totalValSS)^EstimOrder
160         #possible EstimOrder-vectors with values {0,1,...,maxStateSpace}
161         mykStateLevels <- permutations(totalValSS,EstimOrder,
162                                       v=0:maxStateSpace,
163                                       repeats.allowed = TRUE)
164
165         #We define an array where we will save our desired matrices
166         my4Mats <- array(NA,c(nrow(mykStateLevels),
167                               totalValSS,4,ncol(givenmat)-EstimOrder))
168
169         #For each time t, we will create four matrices that will count
170         #different things with respect each EstimOrder-vector and each
171         #sample of 1000 observations:
172         #the Median, the sd, the inf. limit. cred. int. (95%) and the
173         #sup. lim. cred. int. (5%)
174         for(t in 1:(ncol(givenmat)-EstimOrder)){
175             for(i in 1:nrow(mykStateLevels)){

```



```

176     #We sample 1000 observation of what it will be the posterior
177     #distribution of our model
178     if(EstimOrder==1){
179         myDirSamp <- rdirichlet(1000,1/4 + countsList[i,,t])
180
181     } else{
182         myDirSamp <- rdirichlet(1000,1/16 + countsList[i,,t])
183     }
184     #Median Matrix
185     my4Mats[i,,1,t] <- apply(myDirSamp,MARGIN=2,FUN=mean)
186     #Sd Matrix
187     my4Mats[i,,2,t] <- apply(myDirSamp,MARGIN=2,FUN=sd)
188     #inferior limit credible interval 95%
189     my4Mats[i,,3,t] <- apply(myDirSamp,MARGIN=2,
190                             FUN=function(x) return(ci(x,
191                                                         method="ETI",
192                                                         ci=.95)[[2]]))
193     #superior limit credibe interval 95%
194     my4Mats[i,,4,t] <- apply(myDirSamp,MARGIN=2,
195                             FUN=function(x) return(ci(x,
196                                                         method="ETI",
197                                                         ci=.95)[[3]]))
198     }
199     #This is part of the progbar value, and must be removed if progbar
200     #is as well
201     setTxtProgressBar(progbar,value=t)
202 }
203 dimnames(my4Mats) <- list(apply(mykStateLevels,1,paste,collapse=""),
204                          0:maxStateSpace,
205                          c("Median_Matrix","sd_Matrix",
206                            "Inf_lim_CI_95%","Sup_lim_CI_95%"),c())
207 #Finally, we tell the function to return the modified array we
208 #just made
209 return(my4Mats)
210 }
211
212 else if(EstimOrder==0){
213     my4Mats <- array(NA,c(1,totalValSS,4,ncol(givenmat)))
214
215     for(t in 1:(ncol(givenmat))){
216
217         #We sample 1000 observation of what it will be the posterior
218         #distribution of our model
219         myDirSamp <- rdirichlet(1000,1/4 + countsList[1,,t])
220         #Median Matrix

```

```

221 my4Mats[1,,1,t] <- apply(myDirSamp,MARGIN=2,FUN=mean)
222 #Sd Matrix
223 my4Mats[1,,2,t] <- apply(myDirSamp,MARGIN=2,FUN=sd)
224 #inferior limit credible interval 95%
225 my4Mats[1,,3,t] <- apply(myDirSamp,MARGIN=2,
226 FUN=function(x) return(ci(x,
227 method="ETI",
228 ci=.95)[[2]]))
229 #superior limit credibe interval 95%
230 my4Mats[1,,4,t] <- apply(myDirSamp,MARGIN=2,
231 FUN=function(x) return(ci(x,
232 method="ETI",
233 ci=.95)[[3]]))
234
235 #This is part of the progbar value, and must be removed if progbar
236 #is as well
237 setTxtProgressBar(progbar,value=t)
238 }
239 dimnames(my4Mats) <- list(c(),0:maxStateSpace,
240 c("Median□Matrix","sd□Matrix",
241 "Inf□lim□CI□95%","Sup□lim□CI□95%"),c())
242 return(my4Mats)
243 }
244
245 }
246
247 #With this functions, we also generate 1000 draws of P(X_t=i) for
248 #k>=1 and the we consider only the mean of this value.
249 my4MatsGen_with_X_t <- function(givenmat, countsList,OrderChain=3,
250 numSamples=1000,
251 maxStateSpace=3){
252
253 totalValSS <- maxStateSpace+1
254 X_t <- matrix(NA,ncol(givenmat),totalValSS)
255 #Q_ts <- array(NA,c(numSamples,ncol(givenmat)-OrderChain,totalValSS^OrderChain))
256 Q_ts <- matrix(NA,ncol(givenmat)-OrderChain,totalValSS^OrderChain)
257
258 CI.95_X_t <- array(NA,c(366,2,totalValSS))
259
260 #This is just for looking the time the function takes to make all
261 #the proceed (not really necessary)
262 progbar <- txtProgressBar(min=0,max=(ncol(givenmat)-OrderChain),
263 style=3,char="=")
264
265 if(OrderChain==0){

```

```

266 my4Mats <- array(NA,c(1,totalValSS,4,ncol(givenmat)))
267
268 for(t in 1:(ncol(givenmat))){
269
270     #We sample 1000 observation of what it will be the posterior
271     #distribution of our model
272     #myDirSamp <- rdirichlet(1000,1/4 + countsList[1,,t])
273     myDirSamp <- rdirichlet(1000,1/16 + countsList[1,,t])
274
275     #Median Matrix
276     my4Mats[1,,1,t] <- apply(myDirSamp,MARGIN=2,FUN=mean)
277     #Sd Matrix
278     my4Mats[1,,2,t] <- apply(myDirSamp,MARGIN=2,FUN=sd)
279     #inferior limit credible interval 95%
280     my4Mats[1,,3,t] <- apply(myDirSamp,MARGIN=2,
281                             FUN=function(x) return(ci(x,
282                                                         method="ETI",
283                                                         ci=.95)[[2]]))
284     #superior limit credibe interval 95%
285     my4Mats[1,,4,t] <- apply(myDirSamp,MARGIN=2,
286                             FUN=function(x) return(ci(x,
287                                                         method="ETI",
288                                                         ci=.95)[[3]]))
289
290     #This is part of the progbar value, and must be removed if progbar
291     #is as well
292     setTxtProgressBar(progbar,value=t)
293 }
294 dimnames(my4Mats) <- list(c(),0:maxStateSpace,
295                          c("Median_Matrix","sd_Matrix",
296                            "Inf_lim_CI_95%","Sup_lim_CI_95%"),c())
297 return(my4Mats)
298
299 } else if(OrderChain!=0){
300
301     #Like in the first function, we set the (totalValSS)^OrderChain
302     #possible OrderChain-vectors with values {0,1,...,maxStateSpace}
303     mykStateLevels <- permutations(totalValSS,OrderChain,
304                                   v=0:maxStateSpace,
305                                   repeats.allowed = TRUE)
306
307     if(OrderChain !=1){
308         mykStateLevels <- t(apply(mykStateLevels,1,rev))
309     }
310

```

```

311 #print(mykStateLevels)
312
313 #correct_tags <- apply(mykStateLevels,1,vec2scal,c(maxStateSpace=maxStateSpace)) + 1
314
315 #print(correct_tags)
316
317 #We define an array where we will save our desired matrices
318 my4Mats <- array(NA,c(nrow(mykStateLevels),
319                       totalValSS,4,ncol(givenmat)-OrderChain))
320
321 #For each time t, we will create four matrices that will count
322 #different things with respect each OrderChain-vector and each
323 #sample of 1000 observations:
324 #the Median, the sd, the inf. limit. cred. int. (95%) and the
325 #sup. lim. cred. int. (5%)
326
327
328 first_n_m <- table(apply(as.matrix(givenmat[,1:OrderChain]),
329                          1,vec2scal,c(maxStateSpace=maxStateSpace))) #<----
330 print(first_n_m)
331 print("tamaño de first_n_m")
332 print(length(first_n_m))
333 all_ms <- rep(0,totalValSS^OrderChain) #<----
334
335 for(name in names(first_n_m)){
336   all_ms[as.integer(name)+1] <- first_n_m[name] # <-----
337 }
338 print(all_ms)
339 #Sys.sleep(30)
340 #Estos van de la mano, sirven para X_1,X_2,...,X_k
341 if(OrderChain==1){
342   #toStart <- rdirichlet(1000,1/4 + all_ms)
343   toStart <- rdirichlet(1000,1/16 + all_ms)
344
345   #Aquí damos el valor para ordenes k>=2
346 } else {
347   #print("1/16")
348   toStart <- rdirichlet(1000,1/16 + all_ms)
349 }
350
351 allSamples_X_t <- array(NA,c(numSamples,totalValSS,ncol(givenmat)))
352 forFirstX_t <- apply(toStart,2,mean)
353 Q_ts[1,] <- forFirstX_t
354
355 #Esto es para los CI's del 95% para X_1,X_2,...,X_k

```

```

356 first_inf_CI <- apply(toStart,MARGIN=2,
357 FUN=function(x) return(ci(x,
358 method="ETI",
359 ci=.95)[[2]]))
360 #superior limit credibe interval 95%
361 first_sup_CI <- apply(toStart,MARGIN=2,
362 FUN=function(x) return(ci(x,
363 method="ETI",
364 ci=.95)[[3]]))
365
366
367 #Aquí guardamos P(X_l=i),l=1,..k,i=0,..,3
368 #También vamos a guardar las 1000 muestras correspondientes a los primeros
369 #P(X_l=i),l=1,...,k, i=0,...,3
370 for(l in 1:OrderChain){
371   for(m in 0:maxStateSpace){
372     #eggs <- which(mykStateLevels[,l] == m)
373     #print(eggs)
374     #print(dim(toStart))
375     allSamples_X_t[,m+1,l] <- apply(as.matrix(toStart[,which(mykStateLevels[,l] == m)]),
376 1,sum)
377     X_t[l,m+1] <- sum(forFirstX_t[which(mykStateLevels[,l] == m)])
378     CI.95_X_t[l,1,m+1] <- sum(first_inf_CI[which(mykStateLevels[,l] == m)])
379     CI.95_X_t[l,2,m+1] <- sum(first_sup_CI[which(mykStateLevels[,l] == m)])
380
381   }
382 }
383
384
385 #Aquí vamos a estar guardando las 1,0000 simulaciones de las probs. de transición
386 #dado el tiempo t
387 toUseforX_t <- array(NA,c(numSamples,totalValSS,totalValSS^OrderChain))
388 #print(dim(toUseforX_t))
389 #Aquí vamos a estar guardando las 1000 simulaciones de las Q_(t),t=1,...,T-k
390
391 #####
392 ##Se necesita agregar el tiempo ya que las Q_(t)'s dependen de las Q_(t-1)'s.##
393 #####
394 #toUseforQ_t <- array(NA,c(numSamples,totalValSS^OrderChain))
395
396 toUseforQ_t <- array(NA,c(numSamples,ncol(givenmat)-OrderChain,totalValSS^OrderChain))
397 toUseforQ_t[,1,] <- toStart
398
399 #Aquí guardaremos todas las muestras que vamos a generar
400 allSamples_P_mj <- array(NA,c(numSamples,

```

```

401     totalValSS, nrow(mykStateLevels),
402     (ncol(givenmat)-OrderChain))
403
404   if(OrderChain==1){
405     for(t in 1:(ncol(givenmat)-OrderChain)){
406
407       for(i in 1:nrow(mykStateLevels)){
408         #We sample 1000 observation of what it will be the posterior
409         #distribution of our model
410         #myDirSamp <- rdirichlet(numSamples,1/4 + countsList[i,,t])
411         myDirSamp <- rdirichlet(numSamples,1/16 + countsList[i,,t])
412
413         toUseforX_t[, ,i] <- myDirSamp
414
415         allSamples_P_mj[, ,i,t] <- myDirSamp
416         #Median Matrix
417         my4Mats[i,,1,t] <- apply(myDirSamp, MARGIN=2, FUN=mean)
418         #Sd Matrix
419         my4Mats[i,,2,t] <- apply(myDirSamp, MARGIN=2, FUN=sd)
420         #inferior limit credible interval 95%
421         my4Mats[i,,3,t] <- apply(myDirSamp, MARGIN=2,
422                                 FUN=function(x) return(ci(x,
423                                                         method="ETI",
424                                                         ci=.95)[[2]]))
425         #superior limit credibe interval 95%
426         my4Mats[i,,4,t] <- apply(myDirSamp, MARGIN=2,
427                                 FUN=function(x) return(ci(x,
428                                                         method="ETI",
429                                                         ci=.95)[[3]]))
430       }
431
432       for(b in 0:maxStateSpace){
433         actual_Row <- apply(as.matrix(toUseforX_t[,b+1,]*toUseforQ_t[,t,]),
434                             MARGIN = 1, sum)
435         allSamples_X_t[,b+1,t+OrderChain] <- actual_Row
436         X_t[t+OrderChain,b+1] <- mean(actual_Row)
437         CI.95_X_t[t+OrderChain,1,b+1] <- ci(actual_Row, method="ETI", ci=.95)[[2]]
438         CI.95_X_t[t+OrderChain,2,b+1] <- ci(actual_Row, method="ETI", ci=.95)[[3]]
439       }
440
441       if(t!=(ncol(givenmat)-OrderChain)){
442         for(j in 1:nrow(mykStateLevels)){
443
444
445           toObtain <- c()

```

```

446     selected_ms <- c()
447
448     for(e in 0:maxStateSpace){
449         #print(vec2scal(c(e,toObtain),OrderChain)+1)
450         #agregar <- vec2scal(c(e,toObtain),maxStateSpace)
451         selected_ms[e+1]<- vec2scal(c(e,toObtain),maxStateSpace)+1
452         #selected_ms[e+1]<- agregar+1
453
454     }
455
456     #print(selected_ms)
457     #Aquí guardamos el promedio de los valores que obtuvimos
458     #con las simulaciones
459     toUseforQ_t[,t+1,j] <- apply(as.matrix(toUseforX_t[,
460         mykStateLevels[j,OrderChain]+1,selected_ms]*toUseforQ_t[,t,selected_ms]),
461         1,sum)
462
463     }
464
465     Q_ts[t+1,] <- apply(toUseforQ_t[,t+1,],2,mean)
466
467 }
468
469
470 #This is part of the progbar value, and must be removed if progbar
471 #is as well
472 setTxtProgressBar(progbar,value=t)
473 }
474
475 } else {
476     for(t in 1:(ncol(givenmat)-OrderChain)){
477
478         for(i in 1:nrow(mykStateLevels)){
479             #We sample 1000 observation of what it will be the posterior
480             #distribution of our model
481             myDirSamp <- rdirichlet(numSamples,1/16 + countsList[i,,t])
482
483             toUseforX_t[:,i] <- myDirSamp
484             allSamples_P_mj[:,i,t] <- myDirSamp
485             #Median Matrix
486             my4Mats[i,,1,t] <- apply(myDirSamp,MARGIN=2,FUN=mean)
487             #Sd Matrix
488             my4Mats[i,,2,t] <- apply(myDirSamp,MARGIN=2,FUN=sd)
489             #inferior limit credible interval 95%
490             my4Mats[i,,3,t] <- apply(myDirSamp,MARGIN=2,

```

```

491             FUN=function(x) return(ci(x,
492                                     method="ETI",
493                                     ci=.95)[[2]]))
494     #superior limit credibe interval 95%
495     my4Mats[i,,4,t] <- apply(myDirSamp,MARGIN=2,
496                             FUN=function(x) return(ci(x,
497                                                         method="ETI",
498                                                         ci=.95)[[3]]))
499 }
500
501 for(b in 0:maxStateSpace){
502     actual_Row <- apply(as.matrix(toUseforX_t[,b+1,]*toUseforQ_t[,t,]),
503                         MARGIN = 1,sum)
504     allSamples_X_t[,b+1,t+OrderChain] <- actual_Row
505     X_t[t+OrderChain,b+1] <- mean(actual_Row)
506     CI.95_X_t[t+OrderChain,1,b+1] <- ci(actual_Row,method="ETI",ci=.95)[[2]]
507     CI.95_X_t[t+OrderChain,2,b+1] <- ci(actual_Row,method="ETI",ci=.95)[[3]]
508 }
509
510 if(t!=(ncol(givenmat)-OrderChain)){
511     #print(t)
512     for(j in 1:nrow(mykStateLevels)){
513         #print(j)
514         toObtain <- as.vector(mykStateLevels[j,1:(OrderChain-1)])
515         selected_ms <- c()
516
517         for(e in 0:maxStateSpace){
518             #print(vec2scal(c(e,toObtain),OrderChain)+1)
519             #agregar <- vec2scal(c(e,toObtain),maxStateSpace)
520             selected_ms[e+1]<- vec2scal(c(e,toObtain),maxStateSpace)+1
521             #selected_ms[e+1]<- vec2scal(c(toObtain,e),maxStateSpace)+1
522             #selected_ms[e+1]<- agregar+1
523         }
524
525         #print(selected_ms)
526
527         #print(selected_ms)
528         #Aquí guardamos el promedio de los valores que obtuvimos con las simulaciones
529         toUseforQ_t[,t+1,j] <- apply(as.matrix(toUseforX_t[,mykStateLevels[j,
530                                             OrderChain]+1,selected_ms]*toUseforQ_t[,t,selected_ms]),1,sum)
531
532     }
533
534     Q_ts[t+1,] <- apply(toUseforQ_t[,t+1,],2,mean)
535

```



```

536     }
537
538
539     #This is part of the progbar value, and must be removed if progbar
540     #is as well
541     setTxtProgressBar(progbar,value=t)
542 }
543
544 }
545
546 dimnames(my4Mats) <- list(apply(mykStateLevels,1,paste,collapse=""),
547                           0:maxStateSpace,
548                           c("Median_Matrix","sd_Matrix",
549                             "Inf_lim_CI_95%","Sup_lim_CI_95%"),c())
550 #Finally, we tell the function to return the modified array we
551 #just made
552
553 return(list(fourMats = my4Mats,
554            X_t = X_t,
555            Q_t = Q_ts,
556            CI.95=CI.95_X_t,
557            samples_P_mj=allSamples_P_mj,
558            samples_Q_t=toUseforQ_t,
559            Samples_X_t=allSamples_X_t))
560 }
561
562 }
563
564
565
566 #This help us to plot all the P_mj's
567 OrderedbyTime <- function(countMatGen,count4Mats,givenvec,
568                           EstimOrder=length(givenvec),
569                           maxStateSpace=3){
570
571     totalValSS <- maxStateSpace + 1
572
573     #Here we will start the store all the values we need in order to
574     #make our desired dataframe
575     listbyVal <- list()
576
577     #As before, we can see how the program evolves during time
578     progbar <- txtProgressBar(min=0,max=(length(count4Mats[1,1,1,])-EstimOrder),
579                              style=3,char="=")
580

```

```

581   if(EstimOrder!=0){
582
583     #Same as the other functions...
584     mykStateLevels <- permutations(totalValSS,EstimOrder ,
585                                   v=0:maxStateSpace ,
586                                   repeats.allowed = TRUE)
587
588     if(EstimOrder !=1){
589       mykStateLevels <- t(apply(mykStateLevels,1,rev))
590     }
591
592     #Here we obtain the correspondent index of the given EstimOrder-vector
593     #in our count matrix
594     myindex <- which(apply(mykStateLevels,1,paste ,
595                           collapse="")==paste(givenvec ,collapse=""))
596
597
598     #We start by making counts in the total Values of our State Space
599     for(k in 1:totalValSS){
600       #Here we will create a matrix of dimension 4x(totalTimes - EstimOrder)
601       matbyTime <- matrix(NA,4,length(count4Mats[1,1,1]))
602       #Then, for each time t, we will store values in the t row of our
603       #matrix matbyTime that correspond to the four values we mention before
604       for(t in 1:length(count4Mats[1,1,1])){
605         #Here we store the values obtained in the count4Mats function, but
606         #let's remember we want now the Empiric Mean instead of the sd value
607         #so we will fix this later
608         matbyTime[,t] <- count4Mats[myindex,k,,t]
609         #Here is where we start to replace the sd values by the emp. mean
610         #values, but we must be careful in the case where the sum of our
611         #observations is zero
612         if(sum(countMatGen[myindex,,t])!=0){
613           matbyTime[2,t] <- countMatGen[myindex,k,t]/sum(countMatGen[myindex,,t])
614         } else {
615           matbyTime[2,t] <- 0
616         }
617         #This is part of the progbar value
618         setTxtProgressBar(progbar ,value=t)
619       }
620       #We make a dataframe from the matbyTime matrix
621       matbyTime <- as.data.frame(t(matbyTime))
622       #And we set names to the columns that correspond of what they
623       #represent
624       names(matbyTime) <- c("Estim. Mean", "Empiric Mean", "CI Inf Limit",
625                             "CI Sup Limit")

```

```

626     #Here we store the dataframe for the k possible value, where
627     # k is in {0,1,...,maxStateSpace}
628     listbyVal[[k]] <- matbyTime
629   }
630   #Finally, we label each dataframe to its corresponding value in the
631   #State Space
632   names(listbyVal) <- 0:maxStateSpace
633   return(listbyVal)
634 }
635 else if(EstimOrder==0){
636   #We start by making counts in the total Values of our State Space
637   for(k in 1:totalValSS){
638     #Here we will create a matrix of dimension 4x(totalTimes - EstimOrder)
639     matbyTime <- matrix(NA,4,length(count4Mats[1,1,1]))
640     #Then, for each time t, we will store values in the t row of our
641     #matrix matbyTime that correspond to the four values we mention before
642     for(t in 1:length(count4Mats[1,1,1])){
643       #Here we store the values obtained in the count4Mats function, but
644       #let's remember we want now the Empiric Mean instead of the sd value
645       #so we will fix this later
646       matbyTime[,t] <- count4Mats[1,k,,t]
647       #Here is where we start to replace the sd values by the emp. mean
648       #values, but we must be careful in the case where the sum of our
649       #observations is zero
650       if(sum(countMatGen[1,,t])!=0){
651         matbyTime[2,t] <- countMatGen[1,,t][k]/sum(countMatGen[1,,t])
652       } else {
653         matbyTime[2,t] <- 0
654       }
655       #This is part of the progbar value
656       setTxtProgressBar(progbar,value=t)
657     }
658     #We make a dataframe from the matbyTime matrix
659     matbyTime <- as.data.frame(t(matbyTime))
660     #And we set names to the columns that correspond of what they
661     #represent
662     names(matbyTime) <- c("Estim_□Mean","Empiric_□Mean","□□CI_□Inf_□Limit",
663                          "CI_□Sup_□Limit")
664     #Here we store the dataframe for the k possible value, where
665     # k is in {0,1,...,maxStateSpace}
666     listbyVal[[k]] <- matbyTime
667   }
668   names(listbyVal) <- 0:maxStateSpace
669   return(listbyVal)
670 }

```

```

671 }
672
673 #####
674
675 #NOTE : The reason we transformed the matbyTime matrix to a dataframe
676 #is because it will make the corresponding graphs easier to plot
677
678 #####
679
680
681 #Function that gives us the line graphs of the Estim. Mean, Emp. Mean,
682 #CI inf lim and CI sup lim of a given EstimOrder-vector
683 plotVectorGraphs <- function(countMatGen,count4Mats,givenvec,
684                               EstimOrder=length(givenvec),maxStateSpace=3,
685                               totalValSS=maxStateSpace+1,numrow=2){
686
687   #Here we'll store our desired graphs
688   myGraphslist <- list()
689   #First, we obtain the totalValSS possible matrices of the given
690   #EstimOrder-vector
691   byTimeMat <- OrderedbyTime(countMatGen,count4Mats,givenvec,
692                               EstimOrder,maxStateSpace)
693
694   #Then, for each value in the State Space, we select the correspondent
695   #matrix and we fix the attributes of our desired plot.
696   for(i in 1:length(byTimeMat)){
697     actualInd <- i-1
698     convVec <- vec2scal(givenvec,maxStateSpace = 3)
699     valueDf <- byTimeMat[[i]]
700     myGraphslist[[i]] <- ggplot(valueDf,aes(x=1:nrow(valueDf),
701                                             colour=Values))+
702       geom_line(aes(y='Estim. Mean',color="Estim_□Mean"),size=1) +
703       geom_line(aes(y='Empiric Mean', color="Empiric_□Mean"),linetype="dashed",size=1) +
704       geom_line(aes(y=' CI Inf Limit',color="CI_□Inf_□Limit")) +
705       geom_line(aes(y='CI Sup Limit',color="CI_□Sup_□Limit")) +
706       ylab(bquote(P [bar(. (convVec))*. (actualInd)]^(.(EstimOrder))* (t))) + xlab("Days") +
707       xlim(0,nrow(valueDf))+
708       #ggtitle(paste("Graph for the value ",actualInd)) +
709
710       #theme(plot.title = element_text(size=20)) +
711       #If we want to personalize the colours of each line vlaue, we
712       #have to set
713       scale_colour_manual(values=c("tomato2","darkolivegreen3","black","dodgerblue2")) +
714       theme(legend.position = "none")
715

```

```

716 }
717 #Finally, in another list, we save to values. In "Graphs" we store
718 #the arrange of all the plots we made before, and in "toSave", we
719 #store an object that will help us save the graphs with the function
720 #ggsave()
721 return(list("Graphs"=do.call("grid.arrange",
722                             c(myGraphslist,nrow=numrow)),
723         "toSave"=do.call("arrangeGrob",c(myGraphslist,nrow=numrow))))
724 }
725
726 #This function will help us calculate the logarithm of gamma
727 #values, due to the lack of capacity of R for the computation of
728 #big values
729 logGamma <- function(x){
730   if(x>171){
731     counter <- x
732     inlog <- 0
733     while(counter>171){
734       inlog <- inlog + log(counter-1)
735       counter <- counter-1
736     }
737     return(inlog+log(gamma(counter)))
738   }
739   else{
740     return(log(gamma(x)))
741   }
742 }
743
744
745 #This is the original function we used before to obtain the
746 #log-Prob, which now it won't work for what we need at this moment
747 OrderProb_forAllparam_4 <- function(givenmat,countMatGen,maxStateSpace=3,
748                                   totalValSS=maxStateSpace+1){
749   Order <- ncol(givenmat) - dim(countMatGen)[3]
750   if(Order==0){
751
752     result <- -nrow(givenmat)*totalValSS*logGamma(1/totalValSS) -
753             nrow(givenmat)*logGamma(nrow(givenmat) + 1)
754     toMultiplyMat <- matrix(NA,nrow(givenmat),totalValSS)
755
756     #In the case of Order=0, we will have the product of the
757     #gamma(n^{0}_{m}(t)+alpha^{0}_{m}(t))/gamma(alpha^{0}_{m}(t)), t in {1,...,T},
758     #m in {1,...,totalValSS}
759     #toMultiplyMat <- matrix(NA,dim(countMatGen)[3],dim(countMatGen)[2])
760

```

```

761   for(i in 1:ncol(toMultiplyMat)){
762     for(t in 1:nrow(toMultiplyMat)){
763       toMultiplyMat[t,i] <- logGamma(countMatGen[1,i,t]+1/totalValSS)
764     }
765   }
766   result <- result + sum(toMultiplyMat)
767   #Now, we must add the next values
768   return(result)
769 }
770 else{
771
772   result <- -logGamma(ncol(givenmat)-dim(countMatGen)[3]+1) -
773     (totalValSS^(ncol(givenmat)-dim(countMatGen)[3]))*(4*dim(countMatGen)[3] +
774     1)*logGamma(1/totalValSS) +
775     logGamma(totalValSS^(ncol(givenmat)-dim(countMatGen)[3]-1)) -
776     logGamma(totalValSS^(ncol(givenmat)-dim(countMatGen)[3]-1) + nrow(givenmat))
777
778
779   #If Order is not zero, then we will need to calculate the product
780   #of the  $\gamma(n^{\{k\}_m} + \alpha^{\{k\}_m}) / \gamma(\alpha^{\{k\}_m})$ ,
781   #m in 1:totalValSS^EstimOrder
782   toMultiplyVec <- c()
783   for(i in 1:length(countMatGen[,1,1])){
784     toMultiplyVec[i] <- logGamma(sum(countMatGen[i,,1])+(1/totalValSS))
785   }
786   result <- result + sum(toMultiplyVec)
787
788   #Next, we do the calculations for the other product,
789   #more specifically,
790   #the value  $1/\gamma(n^{\{0\}_m}(t)+1)$ *
791   # $\prod_{j=0}^{\{3\}} \gamma(n^{\{k\}_{mj}}(t) + \alpha^{\{k\}_{mj}}(t)) / \gamma(\alpha^{\{k\}_{mj}}(t))$ 
792   toMultiplyMat <- matrix(NA,dim(countMatGen)[3],dim(countMatGen)[1])
793
794   for (j in 1:dim(countMatGen)[1]){
795     for(t in 1:nrow(toMultiplyMat)){
796       toMultiplyMat[t,j] <- -logGamma(sum(countMatGen[j,,t])+1)
797       for(s in 1:totalValSS){
798         toMultiplyMat[t,j] <- toMultiplyMat[t,j] +
799           logGamma(countMatGen[j,s,t]+(1/totalValSS))
800       }
801     }
802
803
804   }
805   #Finally, we make the sum of the vector and the matrix and we

```

```

806     #subtract the logarithm of the factorial of the Order
807     #ProbInLog <- sum(toMultiplyVec) + sum(toMultiplyMat)-logGamma(Order+1)
808     result <- result + sum(toMultiplyMat)
809     #return(ProbInLog)
810     return(result)
811
812 }
813 }
814
815 #Esta función es importante para scal2vec, se debe mantener.
816 bin_to_base4 <- function(x){
817   x <- strsplit(x, '')
818   vapply(x, function(bits){
819     bits <- as.integer(bits)
820     paste(2 * bits[c(TRUE, FALSE)] + bits[c(FALSE, TRUE)], collapse = '')
821   }, character(1))
822 }
823
824 #Nos convierte escalares en vectores de la forma en que queramos.
825 #Función lista para cualquier k !!!!#
826 scal2vec <- function(x,Order){
827   toConvert <- paste(rev(as.integer(base::intToBits(x))),collapse="")
828
829   #Esta forma es la indicada conseguir los vectores {m}.
830   return(as.integer(strsplit(substr(intToUtf8(rev(utf8ToInt(bin_to_base4(toConvert)))),
831     start=1,stop=Order),split="")[[1]])),
832     #Este te da vectores para regresar a valores de Ozono y PM10
833   }
834
835 #####
836 ### Función para calcular el valor SAD para cada orden K ###
837 #####
838
839 SAD_estimator <- function(countMat, count4Mats, Q_1, Order, maxStateSpace=3){
840
841   totalValSS <- maxStateSpace+1
842
843
844   if(Order==0){
845     toSum <- c()
846
847     P_mj_0 <- OrderedbyTime(countMat, count4Mats, c(1), EstimOrder = 0)
848
849     for(i in 1:totalValSS){
850

```

```

851     toSum <- c(toSum, sum(abs(P_mj_0[[i]][,1] - P_mj_0[[i]][,2])))
852   }
853   #print(length(toSum))
854   print(toSum)
855   return(sum(toSum))
856 }
857
858 else {
859
860   toSum <- c()
861
862   for(i in 1:(totalValSS^Order)){
863     toAdd <- OrderedbyTime(countMat,
864       count4Mats, givenvec=scal2vec(i-1, Order), EstimOrder = Order)
865
866     for(j in 0:maxStateSpace){
867       element <- toAdd[[j+1]]
868       toSum <- c(toSum, sum(abs(element[,1] - element[,2])))
869     }
870   }
871
872   #Esto es para calcular el SAD para Q(1)
873   first_n_m <- apply(countMat[, , 1], 1, sum)
874   print(length(toSum))
875   SAD_all_P_mj <- sum(toSum)
876   print(toSum)
877
878   SAD_Q_1 <- sum(abs(Q_1[1,] - first_n_m/sum(first_n_m)))
879   print(SAD_Q_1)
880   print(abs(Q_1[1,] - first_n_m/sum(first_n_m)))
881
882   return(SAD_Q_1 + SAD_all_P_mj)
883
884 }
885
886 }

```

Gráficas ("generador_graficas.R")

```
1 source("fcts&DataToUse.R")
2
3 ###Para graficar P(W_t=i)###
4
5 plot_W_ts <- function(X_ts,Empiric_Xts,CIs,EstimOrder,
6                       maxStateSpace=3){
7
8   totalValSS <- maxStateSpace + 1
9
10  #Here we'll store our desired graphs
11  myGraphslist <- list()
12
13  #Then, for each value in the State Space, we select the correspondent
14  #matrix and we fix the attributes of our desired plot.
15  for(i in 1:totalValSS){
16
17    actualInd <- i-1
18    toUse <- as.data.frame(Empiric_Xts[,i])
19    names(toUse) <- c("Empiric_Mean")
20    toUse$XtoUse <- X_ts[,i]
21    toUse$inf_CI <- CIs[,1,i]
22    toUse$sup_CI <- CIs[,2,i]
23    myGraphslist[[i]] <- ggplot(toUse,aes(x=1:nrow(X_ts)))+
24      geom_line(aes(y='XtoUse',color="X_t_estimado"),size=1)+
25      geom_line(aes(y='Empiric Mean',color="X_t_empirico"),linetype="twodash",size=.8)+
26      geom_line(aes(y='inf_CI',color="CI_Inf_Limit")) +
27      geom_line(aes(y='sup_CI',color="CI_Sup_Limit")) +
28
29      ylab(bquote(P(W[t]==.(actualInd))))+
30      xlab("days") + xlim(0,nrow(X_ts))+
31      ylim(c(0,1))+
32
33      scale_colour_manual(values=c("tomato2","darkolivegreen3","black","dodgerblue2")) +
34      theme(legend.position = "none",axis.title.y = element_text(size=30),
35            axis.title.x = element_text(size=22),axis.text = element_text(size=20))+
36      theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
37            panel.background = element_blank(), axis.line = element_line(colour = "black"))
38
39  }
40  #Finally, in another list, we save to values. In "Graphs" we store
41  #the arrange of all the plots we made before, and in "toSave", we
42  #store an object that will help us save the graphs with the function
43  #ggsave()
```

```

44     return(list("Graphs"=do.call("grid.arrange",
45                               myGraphslist),
46                               "toSave"=do.call("arrangeGrob",myGraphslist)))
47 }
48
49
50 #W_t_fork0 <- plot_W_ts(test$X_t,Empiric_X_t,test$CI.95,EstimOrder = 1)
51
52
53 #####
54 ###Aquí guardamos gráfica de P(W_t=i) para k=0###
55 #####
56
57 W_t_fork0 <- plot_W_ts(probs_X_t_fork0,Empiric_X_t,CI.95_fork0,EstimOrder = 0)
58
59
60 #####
61 ###Aquí guardamos gráfica de P(W_t=i) para k=1###
62 #####
63
64 W_t_fork1 <- plot_W_ts(probs_X_t_fork1,Empiric_X_t,CI.95_fork1,EstimOrder = 1)
65
66
67
68 #####
69 ### Esto nos permite graficar las probabilidades de transición P_mj a como ###
70 ### se necesita en el artículo ###
71 #####
72
73 plotVectorGraphsbyRows <- function(countMatGen,count4Mats,givenvec,
74                                   EstimOrder=length(givenvec),maxStateSpace=3,
75                                   totalValSS=maxStateSpace+1,numrow=1){
76
77     #Here we'll store our desired graphs
78     myGraphslist <- list()
79     #First, we obtain the totalValSS possible matrices of the given
80     #EstimOrder-vector
81     byTimeMat <- OrderedbyTime(countMatGen,count4Mats,givenvec,
82                               EstimOrder,maxStateSpace)
83
84     #Then, for each value in the State Space, we select the correspondent
85     #matrix and we fix the attributes of our desired plot.
86     for(i in 1:length(byTimeMat)){
87         actualInd <- i-1
88         convVec <- vec2scal(givenvec,maxStateSpace = 3)

```

```

89     valueDf <- byTimeMat[[i]]
90     myGraphslist[[i]] <- ggplot(valueDf, aes(x=1:nrow(valueDf),
91                                           colour=Values))+
92       geom_line(aes(y='Estim. Mean', color="Estim_□Mean"), size=1) +
93       geom_line(aes(y='Empiric Mean', color="Empiric_□Mean"), linetype="dashed", size=1) +
94       geom_line(aes(y=' CI Inf Limit', color="CI_□Inf_□Limit")) +
95       geom_line(aes(y=' CI Sup Limit', color="CI_□Sup_□Limit")) +
96       ylab(bquote(P [bar(. (convVec)) * . (actualInd)] ^ (. (EstimOrder)) * (t))) + xlab("days") +
97       xlim(0, nrow(valueDf))+
98       #ggtitle(paste("Graph for the value ", actualInd)) +
99
100      scale_colour_manual(values=c("tomato2", "darkolivegreen3", "black", "dodgerblue2")) +
101      theme(legend.position = "none", axis.title.y = element_text(size=35),
102            axis.title.x = element_text(size=22), axis.text = element_text(size=30))+
103      theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
104            panel.background = element_blank(),
105            axis.line = element_line(colour = "black"))
106
107  }
108  #Finally, in another list, we save to values. In "Graphs" we store
109  #the arrange of all the plots we made before, and in "toSave", we
110  #store an object that will help us save the graphs with the function
111  #ggsave()
112  return(list("Graphs"=do.call("grid.arrange",
113                               c(myGraphslist, nrow=numrow)),
114           "toSave"=do.call("arrangeGrob", c(myGraphslist, nrow=numrow))))
115 }
116
117
118 #####
119 ### Aquí vamos a generar y guardar las gráficas de las probabilidades de transición###
120 #####
121
122 #####
123 #####      k = 1      #####
124 #####
125
126 #####
127 # Con Dirichlet(1/4) #
128 #####
129
130 #four_mats1 <- readRDS("4_matrices_k=1")
131
132
133 #####

```

```

134 # Con Dirichlet(1/16) #
135 #####
136
137 four_mats1 <- readRDS("4_matrices_k=1_con_1_16.rds")
138
139
140 #####
141 #####      k = 2      #####
142 #####
143
144 four_mats2 <- readRDS("4_matrices_k=2")
145
146 mylist <- list()
147 for(i in 0:15){
148   print(scal2vec(i,2))
149   toPlot <- plotVectorGraphsbyRows(n_mj_fork2,
150     four_mats2,scal2vec(i,Order = 2),EstimOrder = 2)
151   mylist[[i+1]] <- toPlot$toSave
152 }
153
154 #####
155 ### Graficas para k=2 en cuatro fragmentos distintos de 4x4 ###
156 #####
157
158 allP_2s_Part1 <- do.call("arrangeGrob",c(mylist[1:4],nrow=4))
159
160 allP_2s_Part2 <- do.call("arrangeGrob",c(mylist[5:8],nrow=4))
161
162 allP_2s_Part3 <- do.call("arrangeGrob",c(mylist[9:12],nrow=4))
163
164 allP_2s_Part4 <- do.call("arrangeGrob",c(mylist[13:16],nrow=4))
165
166
167 #####
168 ###Aquí tenemos todas las gráficas para k=2 en una sola imagen de 16x4###
169 #####
170
171 allP_2s <- do.call("arrangeGrob",c(mylist,nrow=16))
172
173 #####
174 ### Gráfica de los log-probs ###
175 #####
176
177 probs= c(-11952.01,
178          -11503.57,

```

```

179         -11786.97,
180         -11808.59,
181         -11806.84,
182         -11818.90,
183         -11837.81,
184         -11906.25,
185         -12026.29)
186
187 plot(probs, type="l")
188
189 library(ggplot2)
190
191 ordersK <- ggplot(data.frame(x=0:8,y=probs),
192                   aes(x=x,y=y),
193 ) + geom_line() +
194     theme_bw() +
195     theme(axis.line = element_line(colour = "black"),
196           panel.grid.major = element_blank(),
197           panel.grid.minor = element_blank(),
198           #panel.border = element_blank(),
199           panel.background = element_blank()
200 ) +
201     labs(#title = "Ontario mobility trends",
202          #subtitle = paste0("Mobility reports between ", " and "),
203          x = "Order_K",
204          #y= paste0("L(Y|K)", expression(Lambda))
205          y=expression(paste("Log", bgroup("(", paste("L(Y|K)", frac(lambda~K, "K!")), ")")))
206 ) +
207     theme(axis.title.x = element_text(size=15,
208                                       hjust = .5,
209                                       margin = margin(t=20)
210 ),
211          axis.title.y = element_text(size=15,
212                                       hjust = .5,
213                                       margin = margin(r=20),
214                                       #angle=0,
215                                       vjust=.5
216 ),
217          #axis.text.x = element_text(angle=90),
218          panel.spacing = unit(2, "lines"))

```

Bibliografía

- BARRIOS, J.M. Y RODRIGUES, E.R. A queueing model to study the occurrence and duration of ozone exceedances in Mexico City. *Journal of Applied Statistics* **42**(1):214–230 (2015). <https://doi.org/10.1080/02664763.2014.939613>
- BOYS, R.J. Y HENDERSON, D.A. A Bayesian approach to DNA sequence segmentation. *Biometrics* **60**(3):573–581 (2004). <https://doi.org/10.1111/j.0006-341X.2004.00206.x>
- BOYS, R. Y HENDERSON, D. On Determining the Order of Markov Dependence of an Observed Process Governed by a Hidden Markov Model. *Scientific Programming* **10**:241–251 (2002). <https://doi.org/10.1155/2002/683164>
- CARLIN, B.P. Y LOUIS, T.A. *Bayesian Methods for Data Analysis*. 2^a edición. Chapman & Hall/CRC press (2009)
- DRTON, M., MARZBAN, C., GUTTORP, P., Y SCHAEFER, J.T. A Markov Chain Model of Tornadoic Activity. *Monthly Weather Review* **131**(12):2941 – 2953 (2003). [https://doi.org/10.1175/1520-0493\(2003\)131<2941:AMCMOT>2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131<2941:AMCMOT>2.0.CO;2)
- FAN, T.H. Y TSAI, C.A. A Bayesian method in determining the order of a finite state Markov chain. *Communications in Statistics - Theory and Methods* **28**:1711–1730 (1999). <https://doi.org/10.1080/03610929908832381>
- GALLEGOS-HERRADA, M.A., RODRIGUES, E.R., TARUMOTO, M.H., Y TZINTZUN,

- G. A multi-dimensional non-homogeneous Markov chain of order K to jointly study multi-pollutant exceedances. *Environmental Model & Assessment* (2021)
- GAMERMAN, D. Y LOPES, H.F. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. 2^a edición. Chapman & Hall/CRC (2006)
- GELMAN, A., CARLIN, J., STERN, H., DUNSON, D., VEHTARI, A., Y RUBIN, D. *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis (2013)
- KARLIN, S. Y TAYLOR, H.M. *A First Course in Stochastic Processes*. 2^a edición. Academic Press (1975)
- MCGRAYNE, S.B. *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy*. Yale University Press (2011)
- MORALES-MORILLÓN, J.A. Cadenas de Markov no-homogéneas para el mapeo genético de poblaciones mezcladas (Admixture Mapping). Tesis de licenciatura. Facultad de Ciencias, UNAM. (2018)
- NOM-020-SSA1-2014. “Salud ambiental. Valor ímite permisible para la concentración de ozono (O₃) en el aire ambiente y criterios para su evaluación. Norma Oficial Mexicana”. Informe técnico, Diario Oficial de la Federación, México (2014)
- NOM-025-SSA1-2014. “Salud ambiental. Valores límite permisibles para la concentración de partículas suspendidas PM₁₀ y PM_{2.5} en el aire ambiente y criterios para su evaluación.”. Informe técnico, Diario Oficial de la Federación, México (2014)
- RODRIGUES, E.R., TARUMOTO, M.H., Y TZINTZUN, G. A Non-Homogeneous Markov Chain Model to Study Ozone Exceedances in Mexico City. En F. Nejadkoorki (editor), *Current Air Quality Issues*, capítulo 17. IntechOpen, Rijeka (2015). <https://doi.org/10.5772/59728>

Ross, S.M. *Stochastic Processes*. 2^a edición. John Wiley & Sons, Inc (1996)