



Casa abierta al tiempo

**UNIVERSIDAD AUTÓNOMA METROPOLITANA**  
UNIDAD IZTAPALAPA División de Ciencias Básicas e Ingeniería

---

**MODELOS DE MARKOV OCULTOS PARA  
DESCRIBIR LOS NIVELES DE OZONO EN LA  
ATMOSFERA DE LA CIUDAD DE MÉXICO**

**Tesis para obtener el grado de:**

*Maestra en ciencias (Matemáticas Aplicadas e Industriales)*

**Por:**

Araceli Ramírez López

**Asesor:**

Dra. Blanca Rosa Pérez Salvador.

**Jurado calificador:**

Presidente: Dr. Raúl Montes de Oca Machorro.

Secretario: Dra. Blanca Rosa Pérez Salvador.

Vocal: Dr. Hugo Adán Cruz Suárez.

Vocal: Dr. Juan González Hernández†.

Ciudad de México, 29 de marzo de 2017



Casa abierta al tiempo

**UNIVERSIDAD AUTÓNOMA METROPOLITANA**  
UNIDAD IZTAPALAPA División de Ciencias Básicas e Ingeniería

---

**MODELOS DE MARKOV OCULTOS PARA  
DESCRIBIR LOS NIVELES DE OZONO EN LA  
ATMOSFERA DE LA CIUDAD DE MÉXICO**

**Tesis para obtener el grado de:**

*Maestra en ciencias (Matemáticas Aplicadas e Industriales)*

**Por:**

Araceli Ramírez López

**Asesor:**

Dra. Blanca Rosa Pérez Salvador.

**Jurado calificador:**

Presidente: Dr. Raúl Montes de Oca Machorro.

Secretario: Dra. Blanca Rosa Pérez Salvador.

Vocal: Dr. Hugo Adán Cruz Suárez.

Vocal: Dr. Juan González Hernández<sup>†</sup>.

Ciudad de México, 29 de marzo de 2017



## *Agradecimientos*

*Mi agradecimiento a la Dra. Blanca Rosa por guiarme en la elaboración de este trabajo y por la oportunidad que me brindó para terminar mi tesis, sin su ayuda no hubiera sido posible.*

*A mis sinodales, Dr. Raúl Montes de Oca, Dr. Hugo Adán Cruz y Dr. Juan González<sup>†</sup>, gracias por el tiempo que me brindaron para la revisión de esta tesis.*

*A los profesores de la MCMAI por compartir sus conocimientos, fué una maravillosa experiencia.*

*A tí, Romerito, gracias por apoyarme en los momentos difíciles y sobretodo por ser mi incondicional compañero de vida.*

*A mi familia por apoyarme y cuidarme siempre.*



# Índice general

	Página
Lista de figuras	3
Lista de cuadros	5
Notación	6
Introducción	7
<b>1. Modelos de Markov ocultos</b>	<b>12</b>
1.1. Cadenas de Markov . . . . .	12
1.2. Modelo de Markov oculto . . . . .	16
1.3. Modelos de Markov oculto con espacio de observaciones discreto . . . . .	16
1.3.1. Elementos de un modelo de Markov oculto . . . . .	17
1.3.2. Los tres problemas que establece un modelo de Markov oculto . . . . .	19
1.3.3. Solución del problema 1 de un modelo de Markov oculto . . . . .	21
1.3.4. Solución del problema 2 de un modelo de Markov oculto . . . . .	25
1.3.5. Solución del problema 3 de un modelo de Markov oculto . . . . .	27
1.4. Modelos de Markov ocultos con espacio de observaciones continuo . . . . .	30
1.4.1. Elementos de un HMM con espacio de observaciones continuo . . . . .	31
1.4.2. Algunas características de un HMM con espacio de observaciones continuo . .	34
1.4.3. Algoritmo EM para estimar los parámetros de un HMM . . . . .	36
1.4.4. Estimación de los parámetros de un HMM bajo la distribución Gaussiana . . .	38
1.4.5. Estimación de los parámetros de un HMM bajo una mezcla de Gaussianas . .	41
1.4.6. Estimación de los parámetros de un HMM bajo la distribución Gamma . . . .	46
1.4.7. Estimación de los parámetros de un HMM bajo la distribución Lognormal mul- tivariable . . . . .	49
<b>2. Niveles de ozono en la Ciudad de México</b>	<b>52</b>
2.1. Conceptos . . . . .	52
2.2. Regímenes del ozono . . . . .	54
2.3. Datos . . . . .	55
2.4. Procedimiento . . . . .	55
2.5. Resultados . . . . .	57
<b>3. Conclusiones</b>	<b>80</b>

Bibliografía	82
Apéndice	85
Código del programa	85

# Índice de figuras

1.1.	Diagrama del ejemplo: Urnas de pelotas, un modelo de Markov oculto . . . . .	17
1.2.	Simulación de los estados de una cadena de Markov oculta con observaciones que obedecen a la distribución Poisson. . . . .	20
1.3.	Simulación de las observaciones de una cadena de Markov oculta con observaciones discretas. . . . .	20
1.4.	Simulación de los estados de una cadena de Markov oculta con observaciones continuas.	33
1.5.	Simulación de las observaciones de una cadena de Markov con observaciones continuas.	33
2.1.	Clasificación del ozono según su ubicación. . . . .	53
2.2.	Esquema de la producción de ozono troposférico. . . . .	53
2.3.	Esquema del número de modelos HMM ajustados a los datos. . . . .	56
2.4.	Secuencia de estados más probable asociada a la secuencia de observaciones de ozono del año 2002 en la zona noreste, se obtuvieron a partir de un HMM con 2 estados- distribución gaussiana, donde $m_1 = 28,39$ , $m_2 = 42,67$ , $M = 34,14$ . . . . .	59
2.5.	Secuencia de estados más probable asociada a la secuencia de observaciones de ozono en el año 2012 de la zona noreste, se obtuvieron con un HMM de 4 estados- distribución gaussiana, donde $m_1 = 21,05$ , $m_2 = 29,37$ , $m_3 = 38,18$ , $m_4 = 48,13$ , $M = 33,39$ . . . . .	61
2.6.	Secuencia de estados más probable asociada a la secuencia de observaciones de ozono del 2015 en el noreste, se obtuvieron con un HMM con 3 estados- distribución gaussiana, donde $m_1 = 18,89$ , $m_2 = 27,96$ , $m_3 = 40,22$ , $M = 27,23$ . . . . .	62
2.7.	En la parte superior, secuencias de estados de la serie de 1998 en la zona noreste, se obtuvo de un HMM con 4 estados- distribución gamma. En la parte inferior, la secuencia de estados de la serie del 2006, misma zona, se obtuvo con un HMM con 2 estados-distribución gamma. . . . .	65
2.8.	En la parte superior, secuencias de estados de la serie de 2014 en la zona suroeste, se obtuvo de un HMM con 3 estados- distribución gaussiana. En la parte inferior, la secuencia de estados de la serie del 2014, misma zona, se obtuvo con un HMM con 3 estados-distribución gamma. . . . .	71
2.9.	Contribución de la contaminación de fondo a la contaminación ambiental bajo los supuestos de un HMM con observaciones gaussiana y gamma. Evolución de 1992 a 2015 de las zonas noroeste y noreste. . . . .	75
2.10.	Contribución de la contaminación de fondo a la contaminación ambiental bajo los supuestos de un HMM con observaciones gaussiana y gamma. Evolución de 1992 a 2015 de las zonas centro y suroeste. . . . .	76
2.11.	Contribución de la contaminación de fondo a la contaminación ambiental bajo los supuestos de un HMM con observaciones gaussiana y gamma. Evolución de 1992 a 2015 de la zona sureste. Evolución de la contaminación ambiental media bajo un HMM con observaciones gamma y gaussiana, zona sureste. . . . .	77

2.12. Evolución de la contaminación ambiental media bajo un HMM con observaciones gamma y gaussiana, zona noroeste y noreste. . . . .	78
2.13. Evolución de la contaminación ambiental media bajo un HMM con observaciones gamma y gaussiana, zona centro y suroeste. . . . .	79

# Índice de cuadros

1.1.	Clasificación de los procesos de Markov. . . . .	14
2.1.	HMM-2 estados- distribución gaussiana, zona noreste, año 2002. De un lado los parámetros del modelo; del otro, en color azul la gráfica de la función de densidad asociada al estado 1, de color verde la función de densidad asociada al estado 2 y de color negro la mezcla de las funciones de densidad del estado 1 y 2. . . . .	58
2.2.	HMM-4 estados- distribución gaussiana, zona noreste, año 2012. De un lado los parámetros del modelo, del otro en color azul la densidad asociada al estado 1, de verde la densidad asociada al estado 2, de rojo la densidad asociada al estado 3, de morado la densidad asociada al estado 4 y de negro la mezcla de densidades. . . . .	60
2.3.	HMM-3 estados- distribución gaussiana, zona noreste, año 2015. De un lado, los parámetros del modelo, del otro de color azul la función de densidad asociada al estado 1, de color verde la función de densidad del estado 2, de rojo la densidad del estado 3 y de negro la mezcla de densidades. . . . .	63
2.4.	HMM-4 estados- distribución gamma, zona noreste, año 1998. De un lado, los parámetros; del otro, en azul la función de densidad asociada al estado 1, de verde la función de densidad asociada al estado 2, de rojo la función de densidad asociada al estado 3, de morado la función de densidad asociada al estado 4 y de negro la mezcla de funciones. . . . .	64
2.5.	HMM-2 estados- distribución gamma, zona noreste, año 2006. De un lado, los parámetros del modelo; del otro, en color azul la función de densidad asociada al estado 1, de color verde la función de densidad asociada al estado 2 y de negro la función de densidad asociada a la mezcla de densidades. . . . .	64
2.6.	HMM-3 estados- distribución gamma, zona noreste, año 2014. De un lado, los parámetros del modelo; del otro, en azul la función de densidad del estado 1, en verde la función de densidad asociada al estado 2, de rojo la función de densidad asociada al estado 3 y de negro la mezcla de densidades. . . . .	66
2.7.	Parámetros de los modelos HMM ajustados en la zona noroeste, en los años 2011, 2012 con distribución gaussiana y gamma, respectivamente. . . . .	67
2.8.	Parámetros de los modelos HMM ajustados en la zona centro en los años 2006, 2011, 2012 y 2015. . . . .	68
2.9.	Parámetros obtenidos de algunos modelos ajustados en la zona suroeste en los años 1992, 2012 y 2014. . . . .	70
2.10.	Parámetros obtenidos de algunos modelos ajustados en la zona sureste, en los años 2006,2007, 2011 y 2014. . . . .	73
2.11.	Relación de la contaminación de fondo y la contaminación ambiental, HMM- distribución gamma. . . . .	73
2.12.	Relación de la contaminación de fondo y la contaminación ambiental, HMM- distribución gaussiana. . . . .	74
2.13.	A)Prueba de medias pareadas con respecto a la contribución de la contaminación de fondo a la contaminación ambiental. Evalua la hipótesis $H_0$ : La contribución de la contaminación de fondo es la misma bajo el supuesto de un HMM con observaciones gamma y gaussiana. B) Prueba de medias pareadas sobre la estimación de la contaminación media ambiental. Evalua la hipótesis $H_0$ : El nivel medio de contaminación ambiental es el mismo cuando se asume que los niveles promedio diario se distribuyen de acuerdo a un HMM con observaciones gaussianas y observaciones gamma. . . . .	74



# Notación

Como referencia se presenta a continuación una lista de la notación utilizada.

## Capítulo 1. Modelos de Markov ocultos

$\mathbb{R}$	Conjunto de números reales.
$\Omega$	Espacio muestral.
$\mathcal{F}$	$\sigma$ -álgebra.
$\mathcal{B}(\mathbb{R})$	$\sigma$ -álgebra del Borel.
$T$	Conjunto de tiempos.
$X$	Variable aleatoria.
$S$	Espacio de estados.
$p_{ij}$	Probabilidad de transición del estado $i$ al estado $j$ en un paso.
$\mathbf{P}$	Matriz de probabilidades de transición.
$\pi^0$	Vector de distribución de probabilidad inicial.
HMM	Modelo de Markov oculto.
$\lambda$	Modelo de Markov oculto.
$\mathcal{O}$	Secuencia de observaciones.
$B$	Conjunto de distribuciones asociadas a un HMM.
$\mathcal{I}$	Secuencia de estados.
$\alpha_t(i)$	Variable forward.
$\beta_t(i)$	Variable backward.
$Q(\cdot)$	Función de Baum Welch.
$\sigma_i$	Parámetro de desviación estándar de la distribución gaussiana asociada al estado $i$ de un HMM.
$\mu_i$	Media de la distribución gaussiana asociada al estado $i$ de un HMM.
$\kappa_i$	Parámetro de forma de la distribución gamma asociada al estado $i$ de un HMM.
$\theta_i$	Parámetro de escala de la distribución gamma asociada al estado $i$ de un HMM.
$\lambda_c$	Modelo de Markov oculto con observaciones continuas.
$\Sigma_i$	Matriz de varianzas y covarianzas de la distribución gaussiana multivariable asociada al estado $i$ de un HMM.
$\mu_i$	Vector de medias de la distribución gaussiana multivariable asociada al estado $i$ de un HMM.

## Capítulo 2. Niveles de ozono en la Ciudad de México

$M$	Media anual de la contaminación ambiental.
$m_i$	Media de la contaminación ambiental asociada al estado $i$ .
IMECA	Índice Metropolitano de Calidad del Aire.
$ppm$	Partes por millón.



# Introducción

La teoría concerniente a los modelos de Markov ocultos (MOM), mejor conocidos por sus siglas en inglés como HMM (Hidden Markov Models), fue introducida al final de la década de 1960 y comienzos de la década de 1970 por Leonard E. Baum y sus colegas, Rabiner y Juang, (1986). Asimismo, las primeras aplicaciones de los HMM ocurrieron a mediados de la década de 1970 mostrando su utilidad en el procesamiento del habla; y es hasta finales de 1980 que tiene lugar la verdadera exposición de este modelo. Cabe mencionar que entre 1986 y 1989, Lawrence R. Rabiner presenta la teoría de los HMM, de Leonard E. Baum, con un lenguaje accesible lo que propició el desarrollo de aplicaciones en diversas áreas, Rabiner (1989). Por mencionar, los HMM han sido utilizados en el área de las biociencias para la clasificación automática de electrocardiogramas, la representación de la actividad neuronal de la corteza visual de simios sometidos a diferentes estímulos visuales; asimismo, en el modelado de series de tiempo de la cantidad de ataques epilépticos en personas y el análisis a gran escala del genoma humano, etc, Carol y Fiona (2009), Rodríguez y Bautista (2000). En epidemiología y biometría, se han desarrollado modelos de la secuencia del comportamiento de animales bajo observación Newman (2006), Cohen (1998), Durbin et al. (1998). También se ha mostrado su utilidad en el tratamiento de imágenes y diversos temas de climatología, etc, Aas et al. (1999), Sansom (1998).

Lo anterior señala que los HMM tienen una estructura matemática adecuada para el desarrollo de diversas aplicaciones. De acuerdo al trabajo seminal de Rabiner, Rabiner (1989), los HMM son:

... un doble proceso estocástico con un proceso subyacente que no es observable (oculto) pero que puede ser inspeccionado a través de otro proceso estocástico que genera la secuencia de observaciones...

Según Mohamed y Garder (2000):

...una cadena de Markov oculta es un proceso estocástico generado por dos mecanismos interrelacionados; por un lado, una cadena de Markov subyacente que tiene un número finito de estados, y por el otro un conjunto de funciones aleatorias, cada una asociada a un estado de la cadena. La cadena de Markov subyacente cambia de estado de acuerdo a su matriz de probabilidades de transición. El observador ve localmente la salida de las funciones aleatorias asociadas a cada estado actual, pero no puede observar directamente los estados de la cadena de Markov...

Considerando las bondades de los HMM en el desarrollo de aplicaciones, a lo largo de este documento se muestra la aplicación de los modelos de Markov oculto, como un complemento a las técnicas de análisis de contribución de fuentes de emisión (ACF) de contaminantes atmosféricos. Un análisis de contribución de fuentes consiste en estudiar el origen de las diversas fuentes de contaminación y la estimación de sus contribuciones a los niveles de contaminación atmosférica ambiente.

La información que provee un ACF es importante en el diseño de políticas públicas para la protección de la salud humana y del medio ambiente, así como para evaluar la efectividad de las estrategias para la reducción de contaminantes, Gomez et al. (2015). En la literatura se identifican tres tipos de técnicas de análisis de contribución de fuentes de emisión, Viana et al. (2008): a) métodos que

requieren la valoración de los datos capturados por los centros de monitoreo, b) métodos que dependen de inventarios de emisiones y de modelos de dispersión atmosférica y c) métodos que se basan en la evaluación estadística de datos sobre el material particulado. El primer grupo se apoya en el tratamiento estadístico básico de los datos, Belis et al. (2013), e incluye el modelado mediante series temporales de datos, Gomez et al. (2015). El segundo grupo incluye modelos para simular la formación de emisiones de aerosoles, transporte y deposición. El tercer grupo se emplea para el estudio del transporte del material particulado en el aire. Ante esta clasificación, Belis et al (2014) recomienda que previo al desarrollo de un ACF debe llevarse a cabo un análisis estadístico básico de los datos de estudio que puede incluir un análisis de tendencias de las observaciones o la estimación de la distribución estadística que mejor describe tal conjunto de datos, asimismo procedimientos sencillos como el análisis de correlaciones pueden utilizarse como una aproximación inicial.

Diversos autores señalan a los HMM como modelos flexibles de uso generalista para el análisis de series temporales univariantes y multivariantes que a su vez ofrecen una fácil interpretación de resultados, Cappé (2005), Zucchini y MacDonald (2009). Por lo anterior, los HMM pueden constituir un punto de partida para el desarrollo de los ACF.

La idea que subyace a esta propuesta puede explicarse de la siguiente manera: Los niveles de contaminación atmosférica en las áreas urbanas es el resultado de la presencia de contaminantes emitidos localmente (de diferentes fuentes), de otras contribuciones aportadas a nivel regional, y del transporte horizontal, Gomez et al. (2016). Por otra parte, el rango de los niveles de contaminación puede dividirse en intervalos consecutivos de tal manera que cada uno represente características específicas de los niveles de contaminación. El principal interés se encuentra en el primer intervalo (concentraciones más bajas) denominado la contaminación de fondo ya que es clave en estudios epidemiológicos y para valorar la exposición general de la población a la contaminación atmosférica. Asimismo, si se piensa que los estados ocultos de un HMM representan diferentes intervalos de concentración o regímenes de un contaminante, entonces dada una serie temporal de datos de un contaminante, el modelo de Markov oculto permite agrupar las observaciones en regímenes de concentración. Por lo que en cada régimen se agrupan observaciones de la serie temporal con concentración similar y que al mismo tiempo difieren en su concentración de otras observaciones agrupadas en otros regímenes. De tal manera que las observaciones agrupadas en el régimen con menor concentración corresponde a la contaminación de fondo. Por su parte, dado que los HMM proporcionan la media y desviación estándar de cada distribución asignada a cada uno de los regímenes, esto proporciona valiosa información al caracterizar una serie temporal de un contaminante.

En esta índole, los objetivos de este trabajo son:

- Presentar de manera detallada y con un lenguaje sencillo la estructura matemática de los modelos de Markov ocultos con la finalidad de que esta teoría sea comprendida por profesionistas con diversa formación e incentive su aplicación en áreas donde aún son desconocidos.

- Proponer el uso de los modelos de Markov ocultos homogéneos como una técnica estadística exploratoria y rutinaria para complementar otras técnicas de ACF con las que estimar las contribuciones de las fuentes a la formación de ozono en la Ciudad de México.

- Implementar la solución de los tres problemas básicos de un HMM en la librería de R (R Core Team, 2015).

- Determinar la evolución de la aportación de la contaminación de fondo a la contaminación ambiente del aire debida al ozono en la Ciudad de México.

Por otra parte, la hipótesis que se plantea es:

-Bajo un adecuado ajuste de un HMM homogéneo a la serie temporal del ozono en la Ciudad de México, la aportación de la contaminación de fondo a la contaminación ambiental del aire es invariante al asumir diferentes distribuciones asociadas a los estados ocultos del HMM.

En otras palabras se quiere verificar que el porcentaje de contaminación del aire que es explicado por la contaminación de fondo, cuando se asume que la distribución de las observaciones es gaussiana, no muestra diferencias estadísticas significativas al porcentaje explicado cuando las observaciones asumen la distribución gamma.

El documento está estructurado de la siguiente manera:

1. En el Capítulo 1, primero se presenta algunos conceptos sobre las cadenas de Markov; posteriormente, se desarrolla la teoría de los modelos de Markov ocultos homogéneos en el caso discreto y continuo; en particular se presentan las formulas para estimar los parámetros del modelo en los casos en que las observaciones siguen las distribuciones gaussianas, mezcla de gaussianas, gamma y lognormal, lo anterior empleando el método de máxima esperanza.
2. En el Capítulo 2, se expone la implementación de los HMM a la serie temporal del promedio diario de ozono en la Ciudad de México; se asume que las observaciones se comportan de acuerdo a la distribución gaussiana y a la distribución gamma. Se presenta un análisis de los porcentajes de contaminación ambiental explicados por la contaminación de fondo. El análisis se realiza para las cinco regiones de la Ciudad de México (centro, sureste, suroeste, noroeste y noreste) de forma anual para los años comprendidos entre 1992 y 2015.
3. En el Capítulo 3, se presentan las conclusiones del trabajo.
4. Finalmente, en el apéndice se muestra información adicional sobre el software empleado en el cálculo de resultados.

# Capítulo 1

## Modelos de Markov ocultos

Desde 1960 y 1970 comenzó el desarrollo de los modelos de Markov, poco después se desarrollaron los modelos de Markov ocultos conocidos como HMM por sus siglas en inglés de Hidden Markov Models. Aunque el desarrollo matemático de estos últimos modelos comenzó desde entonces, se han vuelto más populares en años recientes; la principal razón de esto es que los modelos de Markov ocultos poseen una estructura matemática que permite su aplicación en un amplio campo de conocimientos. Se pueden encontrar diversas aplicaciones de las cadenas de Markov ocultas por mencionar algunas; modelos sobre la dinámica de los consumidores, descripción del rendimiento de los activos financieros, modelos de expectativas de vida, modelos sobre la dinámica de poblaciones, etc. En este capítulo se proporciona algunos conceptos relevantes sobre los modelos de Markov, Hernandez-Lerma (1981); así como una revisión de los aspectos más relevantes de los modelos de Markov ocultos, Petrushin (2000), Psichogios y Ungar (2009).

### 1.1. Cadenas de Markov

El concepto básico sobre el que descansa el desarrollo de este capítulo es el de proceso estocástico.

**Definición 1.1.** Un *proceso estocástico* es un conjunto de variables aleatorias  $\{X(t), t \in T\}$ , donde  $T$  es un conjunto, finito o infinito, de tiempos. Cada una de estas variables aleatorias  $X(t) : \Omega \rightarrow \mathbb{R}$  está definida sobre un espacio medible  $(\Omega, \mathcal{F})$  y toma valores de otro espacio medible  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , en donde  $\mathcal{F}$  es la  $\sigma$ -álgebra sobre  $\Omega$  y  $\mathcal{B}(\mathbb{R})$  es la  $\sigma$ -álgebra de Borel sobre  $\mathbb{R}$ .

Asociado al proceso estocástico, se encuentra el *espacio de estados*, que permite clasificar a los procesos estocásticos en dos grandes grupos.

**Definición 1.2.** Sea  $\{X(t), t \in T\}$  un proceso estocástico, con  $X(t) : \Omega \rightarrow \mathbb{R}$  para cada  $t \in T$  fijo. El *espacio de estados* denotado por  $S$ , con  $S \subseteq \mathbb{R}$ , es el conjunto de todos los valores posibles que puede tomar la variable aleatoria  $X(t)$  para  $t \in T$  fijo. Es decir, se cumple que si  $i \in S$  entonces existe un  $w \in \Omega$  tal que  $X(w, t) = i$  para  $t \in T$  fijo. Cada elemento  $i \in S$  se llama un *estado del proceso*.

De esta manera, los procesos estocásticos se pueden clasificar en los siguientes grupos:

- Si el espacio de estados,  $S$ , es finito o numerable decimos que el proceso estocástico es discreto.
- Si el espacio de estados,  $S$ , es un intervalo, decimos que el proceso estocástico es continuo.

El proceso también se clasifica de acuerdo a la naturaleza del conjunto  $T$ .

- Si  $T$  es numerable, por ejemplo,  $T = \{0, 1, 2, 3, \dots\}$ ,  $T = \{0, \pm 1, \pm 2, \dots\}$ , etc., decimos que  $\{X_t\}$  es un proceso con parámetro (o en tiempo) discreto.
- Si  $T$  es un intervalo como  $T = [a, b]$ ,  $T = [0, \infty)$ ,  $T = (-\infty, \infty)$ , etc.,  $\{X_t\}$  es un proceso con parámetro (o en tiempo) continuo.

Cabe mencionar que algunos procesos estocásticos cumplen con una propiedad específica llamada propiedad de Markov y que será de utilidad más adelante.

**Definición 1.3.** Sea  $\{X(t), t \in T\}$ , un proceso estocástico con  $T \subseteq \mathbb{R}$  el conjunto de parámetros del tiempo y  $S \subseteq \mathbb{R}$  el conjunto de los estados; éste es un proceso de Markov si para cualquier  $x \in \mathbb{R}$  y  $s, t \in T$ , con  $s \leq t$ , se satisface que

$$P\{X(t) \leq x | X(r), r \leq s\} = P\{X(t) \leq x | X(s)\}. \quad (1.1)$$

La ecuación (1.1) se conoce como la propiedad de Markov, se puede interpretar  $s$  como el tiempo "presente",  $t > s$  como el "futuro" y  $r < s$  como el "pasado", entonces la ecuación (1.1) nos dice que dado el presente  $X(s)$  del proceso, el futuro  $X(t)$ ,  $t \geq s$ , es independiente del pasado  $X(r)$ ,  $r < s$ .

Se puede demostrar que la definición (1.3) es equivalente a lo siguiente, ver por ejemplo, Ash y Gardner (1975).

**Definición 1.4.** Sea  $\{X(t), t \in T\}$  un proceso estocástico con  $T \subseteq \mathbb{R}$  el conjunto de parámetros del tiempo y  $S \subseteq \mathbb{R}$  el conjunto de los estados.  $\{X(t), t \in \mathbb{R}\}$  es un proceso de Markov de primer orden<sup>1</sup> si y sólo si para cualquier entero  $n$  y tiempos  $t_1 < t_2 \dots < t_{n+1}$  en  $T$ , se satisface que la distribución condicional de  $X(t_{n+1})$  dados los valores de  $X(t_1), \dots, X(t_n)$ , sólo depende de  $X(t_n)$ ; es decir, para cualesquiera  $x, x_1, \dots, x_n$  elementos del espacio de estados,

$$P\{X(t_{n+1}) \leq x | X(t_1) = x_1, \dots, X(t_n) = x_n\} = P\{X(t_{n+1}) \leq x | X(t_n) = x_n\}. \quad (1.2)$$

Significa que dado el estado presente del proceso, el estado futuro del proceso es independiente de los estados pasados.

Los procesos de Markov, al igual que todo proceso estocástico, se clasifican de acuerdo a la naturaleza del parámetro de tiempo y a la naturaleza del espacio de estados. Con respecto al espacio de estados, por una parte se tiene el proceso de Markov con espacio de estados discreto y recibe el nombre de *cadena de Markov*; y, por otra parte se tiene el proceso de Markov con espacio de estados continuo. De manera similar, con respecto al tiempo, el proceso de Markov se clasifica en proceso de Markov en tiempo discreto y proceso de Markov en tiempo continuo.

Así, se tienen cuatro tipos básicos de procesos de Markov como se ilustra en el Cuadro 1.1. En este trabajo resulta de interés las cadenas de Markov en tiempo discreto.

Para tener una notación más práctica, de aquí en adelante se supondrá que  $T = \{0, 1, 2, 3, \dots\}$ . Esto permite describir la cadena de Markov  $\{X(t), t \in T\}$  como  $\{X_n, n = 0, 1, 2, \dots\}$  donde  $n$  indica el tiempo en el que se encuentra el proceso. Asimismo permite establecer la siguiente definición,

---

<sup>1</sup>El término "primer orden" se debe a que el estado futuro del proceso sólo depende del estado actual; en los procesos de Markov de "segundo orden" el estado futuro depende del estado actual y del estado inmediato anterior y así sucesivamente para otros procesos de Markov de mayor orden.

Parámetro (T)	Espacio de estados S	
	Discreto	Continuo
Discreto	Cadena de Markov en tiempo discreto	Proceso de Markov en tiempo discreto
Continuo	Cadena de Markov en tiempo continuo	Proceso de Markov en tiempo continuo

Cuadro 1.1: Clasificación de los procesos de Markov.

**Definición 1.5.** Una *cadena de Markov* es un proceso estocástico en tiempo discreto  $\{X_n, n = 0, 1, 2, \dots\}$ , con espacio de estados  $S$  discreto, y que satisface la propiedad de Markov; es decir, para cualquier entero  $n \geq 0$ , y para cualesquiera estados  $i, j, i_1, i_2, \dots$  se cumple que

$$P[X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0] = P[X_{n+1} = j | X_n = i]. \quad (1.3)$$

Existen otras formas equivalentes de expresar esta propiedad, por ejemplo, es posible demostrar que la condición (1.3) es equivalente a poder calcular la distribución de las variables  $X_0, X_1, \dots, X_n$  de la siguiente forma, ver Norris (2004).

$$P[X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0] = P[X_0 = i_0]P[X_1 = i_1 | X_0 = i_0] \cdots P[X_n = i | X_{n-1} = i_{n-1}]. \quad (1.4)$$

Sin pérdida de generalidad se tomará como espacio de estados de una cadena de Markov al conjunto discreto  $\{0, 1, 2, 3, \dots\}$  o cualquier subconjunto finito que conste de los primeros elementos de este conjunto.

A la probabilidad  $P[X_{n+1} = j | X_n = i]$ , de la ecuación (1.3), se le denota como  $p_{ij}(n, n+1)$  y se le conoce como *probabilidad de transición*. Esta probabilidad se interpreta como la probabilidad condicional de que el proceso estará en el estado  $j$  en el tiempo  $n + 1$  si en el tiempo  $n$  se encuentra en el estado  $i$ . Estas probabilidades también se conocen como *probabilidades de transición en un paso*.

Cuando en la cadena de Markov la probabilidad  $p_{ij}(n, n + 1)$  depende de  $n$  se dice que la cadena es *no homogénea en el tiempo*. Por otra parte, cuando los números  $p_{ij}(n, n + 1)$  no dependen de  $n$  se dice que la cadena es *homogénea en el tiempo*; en este caso a  $p_{ij}(n, n + 1)$  se le denota simplemente como  $p_{ij}$ . Por simplicidad se asume que estudiaremos cadenas de Markov homogéneas en el tiempo, lo que implica que

$$\begin{aligned} P[X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0] &= P[X_{n+1} = j | X_n = i] \\ &= P[X_1 = j | X_0 = i] \\ &= p_{ij}. \end{aligned} \quad (1.5)$$

Al variar los índices  $i, j$  en la fórmula  $P[X_{n+1} = j | X_n = i] = p_{ij}$  se obtiene la *matriz de probabilidades de transición en un sólo paso*,  $\mathbf{P}$ .

$$\mathbf{P} = \begin{pmatrix} p_{00} & p_{01} & p_{02} & \cdots \\ p_{10} & p_{11} & p_{12} & \cdots \\ p_{20} & p_{21} & p_{22} & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix} \quad (1.6)$$

La entrada  $(i, j)$  de esta matriz es la probabilidad de transición  $p_{ij}$ , es decir, la probabilidad de pasar del estado  $i$  al estado  $j$  en una unidad de tiempo.

Otra definición importante asociada a una cadena de Markov es la distribución de probabilidad inicial.

**Definición 1.6.** Para una cadena de Markov con espacio de estados finito  $\{0, 1, 2, \dots, N\}$  la *distribución de probabilidad inicial* que se denota por  $\pi^0 = \{\pi_0^0, \pi_1^0, \dots, \pi_N^0\}$  se define como

$$\begin{aligned}\pi_0^0 &= P[X_0 = 0] \\ \pi_1^0 &= P[X_0 = 1] \\ &\vdots \\ \pi_N^0 &= P[X_0 = N]\end{aligned}$$

y además  $\sum_{i=0}^N \pi_i^0 = 1$ .

De manera similar para una cadena de Markov con espacio de estados no finito.

La distribución de probabilidad inicial nos dice la probabilidad con la que la cadena inicia en cada uno de los estados. Una vez transcurrida la primera unidad de tiempo, la cadena se encuentra en uno de los posibles estados  $\{0, 1, 2, \dots, N\}$  de acuerdo con la distribución de probabilidad,

$$\pi^1 = \{\pi_0^1, \pi_1^1, \dots, \pi_N^1\} \quad (1.7)$$

en donde la  $j$ -ésima entrada es

$$\begin{aligned}\pi_j^1 &= P[X_1 = j] \\ &= \sum_{i=0}^N P[X_1 = j | X_0 = i] P[X_0 = i] \\ &= \pi_0^0 p_{0j} + \pi_1^0 p_{1j} + \dots + \pi_N^0 p_{Nj} \\ &= (\pi^0 \mathbf{P})_j\end{aligned} \quad (1.8)$$

es decir,  $\pi_j^1$  es la  $j$ -ésima entrada de la matriz que resulta de multiplicar  $\pi^0 P$ . Así  $\pi^1$  se obtiene a partir de  $\pi^0$  y de la matriz de transición  $P$  a través de la fórmula  $\pi^1 = \pi^0 P$ ,

$$(\pi_0^1, \pi_1^1, \dots, \pi_N^1) = (\pi_0^0, \pi_1^0, \dots, \pi_N^0) \begin{pmatrix} p_{00} & p_{01} & \cdots & p_{0N} \\ p_{10} & p_{11} & \cdots & p_{1N} \\ \vdots & \vdots & \vdots & \vdots \\ p_{N0} & p_{N1} & \cdots & p_{NN} \end{pmatrix} \quad (1.9)$$

Por consiguiente,  $\pi^2 = \pi^1 \mathbf{P} = \pi^0 \mathbf{P}^2$ , y así sucesivamente. En general,  $\pi^{n+1} = \pi^n \mathbf{P} = \pi^0 \mathbf{P}^{n+1}$ , lo que define una sucesión infinita de distribuciones de probabilidad  $\pi^0, \pi^1, \pi^2, \dots$ , en donde cada una de ellas se obtiene de multiplicar la distribución anterior por la matriz de probabilidades de transición en un paso.

Las definiciones anteriores serán de utilidad para el desarrollo de la teoría de los modelos de Markov ocultos.

## 1.2. Modelo de Markov oculto

**Definición 1.7.** Un *modelo de Markov oculto* es definido como un proceso bivariado  $\{(X_t, Z_t), t \geq 0\}$  tal que

- $\{X_t, t \geq 0\}$  es una cadena de Markov con matriz de probabilidad de transición  $A$  y vector de distribución de probabilidad inicial  $\pi$ .
- En el proceso de observaciones  $\{Z_t, t \geq 0\}$ , la distribución condicional de  $Z_t$  depende de  $X_t$ .

La variable aleatoria  $\{Z_t\}$  puede ser una variable aleatoria discreta, en cuyo caso decimos se trata de un *modelo de Markov oculto con espacio de observaciones discreto*; o bien, puede ser una variable aleatoria continua y decimos que se trata de un *modelo de markov oculto con espacio de observaciones continuo*.

En la primer parte de este capítulo, se aborda los modelos de Markov con espacio de observaciones discreto y más adelante se establecen las características de los modelos de Markov con espacio de observaciones continuo.

## 1.3. Modelos de Markov oculto con espacio de observaciones discreto

En este apartado se presenta una breve descripción de los modelos de Markov ocultos con espacio de observaciones discreto, para ello se considera el ejemplo siguiente.

■ **Ejemplo 1.1. Urna de pelotas.** Suponga un sistema compuesto por  $N$  urnas que contiene un gran número de pequeñas pelotas de colores, como se muestra en la figura 1.1. Existen  $M$  posibles colores para cada pelota en cada urna. Se realiza el siguiente experimento:

1. De acuerdo a un procedimiento aleatorio se elige una urna inicial. De esta urna se elige aleatoriamente una pelota, cuyo color será la primera observación y se regresa la pelota en la urna a la que pertenece.
2. Enseguida, se selecciona una nueva urna según un procedimiento aleatorio asociado con la actual urna y se repite el proceso de extracción de la pelota.
3. Se repite el paso 2 tantas veces como observaciones se necesiten.

El proceso total genera una secuencia finita de observaciones de colores. Este experimento constituye un modelo de Markov oculto, donde se pueden identificar los siguientes elementos:

- El proceso de elección de las urnas constituye una cadena de Markov donde cada urna representa un estado de la cadena.
- El procedimiento aleatorio de elegir la primera urna está determinado por la distribución inicial de la cadena de Markov sobre cada uno de los estados, es decir sobre las urnas.
- Como cada urna tiene una gran cantidad de pelotas de colores, existe una distribución de probabilidad, para la extracción de una pelota, asociada a cada una de las urnas.

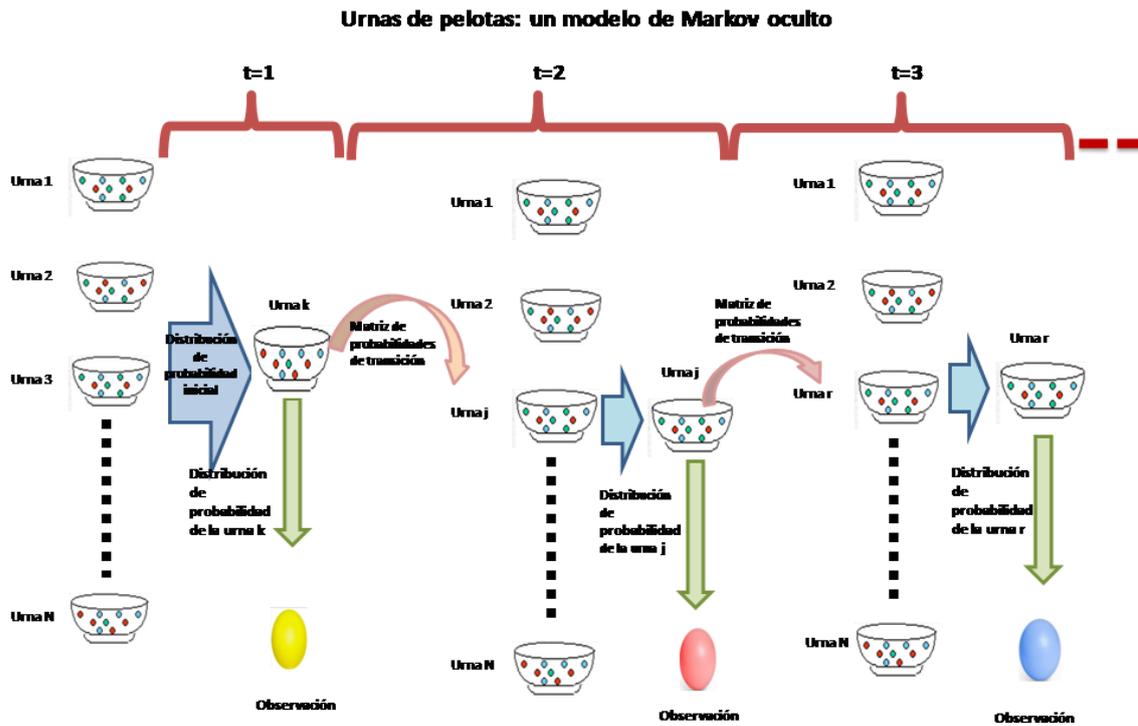


Figura 1.1: Diagrama del ejemplo: Urnas de pelotas, un modelo de Markov oculto

- Una vez elegida la primera urna y haber observado la extracción de pelota, se elige la siguiente urna o estado de la cadena a partir de la matriz de transición de la cadena de Markov. Una vez alcanzado el siguiente estado o urna se realiza la extracción de la pelota de acuerdo a la distribución de probabilidad de las pelotas asociada a esa nueva urna. El proceso continúa en este mismo orden.
- Este ejemplo, como todo modelo de Markov oculto está compuesto por dos procesos, el primero que es la selección de las urnas representa la cadena de Markov; el segundo proceso está descrito por la extracción de las pelotas.

### 1.3.1. Elementos de un modelo de Markov oculto

Con la idea en mente que proporciona el ejemplo anterior; enseguida se describen los elementos esenciales de cualquier modelo de Markov oculto con espacio de observaciones discreto.

1. Sea  $T$  el número total de tiempos en los que se observa el modelo de Markov oculto.
2. En el modelo de Markov oculto ocurren dos procesos de forma simultánea, el primero de ellos es una cadena de Markov, la cual será denotada por  $\{X_t, t \geq 0\}$ . Ésta tendrá un número finito  $N$  de estados y el conjunto de posibles estados será denotado como

$$Q = \{q_1, q_2, q_3, \dots, q_N\}. \quad (1.10)$$

La cadena de Markov tendrá asociada una matriz de probabilidades de transición denotada por  $A = \{a_{q_i, q_j}\}$ , donde

$$a_{q_i, q_j} = P[X_{t+1} = q_j | X_t = q_i]. \quad (1.11)$$

Observe que la ecuación (1.11) corresponde a una cadena de Markov homogénea en el tiempo. Además la cadena de Markov tendrá una distribución de probabilidad inicial denotada por

$\pi = \{\pi_{q_i}\}$ , donde

$$\pi_{q_i} = P[X_1 = q_i]. \quad (1.12)$$

La distribución de probabilidad inicial determinará el estado en el cual comienza la cadena de Markov; después para cada tiempo  $t$ , se hará una transición a un nuevo estado, esto de acuerdo a la matriz de probabilidades de transición. Como transición puede entenderse también permanecer en el mismo estado.

3. El segundo proceso que se presenta en un modelo de Markov oculto, es el proceso de observaciones que será denotado como  $\{Z_t\}$ . En este proceso la variable  $Z_t$  es una variable aleatoria discreta cuyo soporte está compuesto de  $M$  valores y será denotado como

$$V = \{v_1, v_2, v_3, \dots, v_M\}. \quad (1.13)$$

4. Los procesos  $\{X_t\}$  y  $\{Z_t\}$  están relacionados de la siguiente manera: Cada estado de la cadena de Markov  $\{X_t\}$  tiene asociada una distribución de probabilidad para la variable aleatoria  $Z_t$ . Por lo tanto, se tienen  $N$  distribuciones de probabilidad para la variable  $Z_t$ , una por cada estado de la cadena de Markov; este conjunto de distribuciones será denotado como

$$B = \{b(Z_t|q_j)\} \text{ para } j = 1, 2, 3, \dots, N. \quad (1.14)$$

donde para cada  $j = 1, 2, 3, \dots, N$  se tiene,

$$\begin{aligned} b(v_1|q_j) &= P[Z_t = v_1|X_t = q_j] \\ b(v_2|q_j) &= P[Z_t = v_2|X_t = q_j] \\ &\vdots \\ b(v_M|q_j) &= P[Z_t = v_M|X_t = q_j] \end{aligned}$$

Se puede observar que la distribución de probabilidad de la variable  $Z_t$  está determinada únicamente por el estado  $X_t$  por lo que  $Z_t$  es independiente del estado  $X_{t'}$  y de la variable  $Z_{t'}$  siempre que  $t \neq t'$ . Cabe mencionar que cuando  $X_t$  sea desconocido,  $Z_t$  depende de  $X_{t-1}$  debido a la propiedad de Markov.

Con los elementos de un modelo de Markov oculto ya descritos, el proceso para generar un conjunto de observaciones es el siguiente:

1. Definir  $T$ , el número de observaciones deseadas del modelo de Markov oculto.
2. Fijar el estado inicial de la cadena de Markov  $x_1$ , una realización de la variable aleatoria  $X_1$ , de acuerdo a la distribución de probabilidad inicial  $\pi$  de la cadena.
3. Fijar  $t = 1$
4. Elegir una observación  $z_t$ , es decir una realización de la variable aleatoria  $Z_t$ , de acuerdo a la distribución de probabilidad asociada al estado  $x_t$ , es decir  $b(Z_t|x_t)$ .
5. Elegir  $x_{t+1}$ , es decir la realización de la variable  $X_{t+1}$ , a partir de la matriz de probabilidades de transición  $A$ .

6. Fijar  $t = t + 1$ .
7. Si  $t < T$  regresar al paso 4.
8. Al final el proceso genera un conjunto de estados denotado como  $\mathcal{I} = \{x_1, x_2, x_3, \dots, x_T\}$ , así como el conjunto de observaciones  $\mathcal{O} = \{z_1, z_2, z_3, \dots, z_T\}$ .

Para ilustrar el proceso enseguida se presenta un ejemplo de modelo de Markov oculto con espacio de observaciones discreto.

■ **Ejemplo 1.2.** Considere un modelo de Markov oculto que consta de una cadena de Markov  $\{X_t\}$  con dos posibles estados,  $Q = \{1, 2\}$ , cada estado tiene asociada una distribución de Poisson para el proceso de observaciones del modelo  $\{Z_t\}$ . La distribución de probabilidad inicial y la matriz de transición de la cadena de Markov están dadas por

$$\pi = (P[X_1 = 1], P[X_1 = 2]) = (0,7, 0,3) \quad (1.15)$$

$$A = \begin{pmatrix} P[X_{t+1} = 1|X_t = 1] & P[X_{t+1} = 2|X_t = 1] \\ P[X_{t+1} = 1|X_t = 2] & P[X_{t+1} = 2|X_t = 2] \end{pmatrix} = \begin{pmatrix} 0,95 & 0,05 \\ 0,02 & 0,98 \end{pmatrix}. \quad (1.16)$$

Además, las distribuciones de probabilidad asociadas a los estados son:

Distribución de probabilidad de Poisson con media  $\lambda = 10$ .

$$b(z_t|q_1 = 2) = \frac{e^{-\lambda}\lambda^{z_t}}{z_t!} = \frac{e^{-10}10^{z_t}}{z_t!}, \quad (1.17)$$

Distribución de probabilidad de Poisson con media  $\lambda = 1$ .

$$b(z_t|q_1 = 1) = \frac{e^{-\lambda}\lambda^{z_t}}{z_t!} = \frac{e^{-1}1^{z_t}}{z_t!}. \quad (1.18)$$

Seguendo a Laverty et al. (2002), se generaron un conjunto de 200 observaciones del modelo de Markov oculto. En la figura (1.2) se muestra la gráfica de los estados de la cadena de Markov. Asimismo, en la figura (1.3) se muestra las 200 observaciones que se obtuvieron en la simulación de modelo.

### 1.3.2. Los tres problemas que establece un modelo de Markov oculto

Con base en el apartado anterior, para especificar un modelo de Markov oculto se requiere conocer la distribución de probabilidad inicial de la cadena  $\pi$ , el conjunto de distribuciones de probabilidad asociadas a cada estado  $B$  y la matriz de transición  $A$ . Por consiguiente, en adelante un modelo de Markov oculto será denotado como  $\lambda = (A, B, \pi)$ .

En todo modelo de Markov oculto se presentan tres problemas de interés que se explican en este apartado.

**Problema 1.** Dado un modelo de Markov oculto  $\lambda = (A, B, \pi)$  y una secuencia de observaciones  $\mathcal{O} = \{z_1, z_2, z_3, \dots, z_T\}$ , es de interés conocer la probabilidad de que dicha secuencia de observaciones provenga de tal modelo. En otras palabras, ¿cómo determinar  $P[\mathcal{O}|\lambda]$ ?

**Problema 2.** Dado un modelo de Markov oculto  $\lambda = (A, B, \pi)$  y una secuencia de observaciones  $\mathcal{O} = \{z_1, z_2, z_3, \dots, z_T\}$  se busca conocer la secuencia de estados de la cadena de Markov,  $\mathcal{I} = \{x_1, x_2, x_3, \dots, x_T\}$ , que más probablemente generaron la secuencia de observaciones  $\mathcal{O}$ .

**Problema 3.** Dada una secuencia de observaciones  $\mathcal{O} = \{z_1, z_2, z_3, \dots, z_T\}$  ¿cómo estimar los parámetros del modelo  $\lambda = (A, B, \pi)$  de donde proviene dicha secuencia de observaciones?

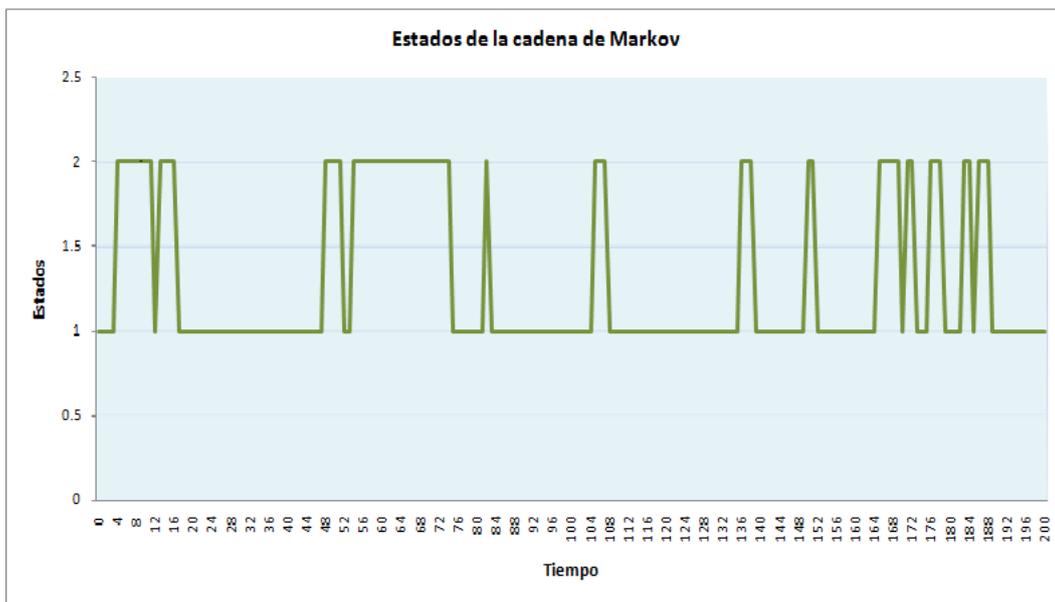


Figura 1.2: Simulación de los estados de una cadena de Markov oculta con observaciones que obedecen a la distribución Poisson.

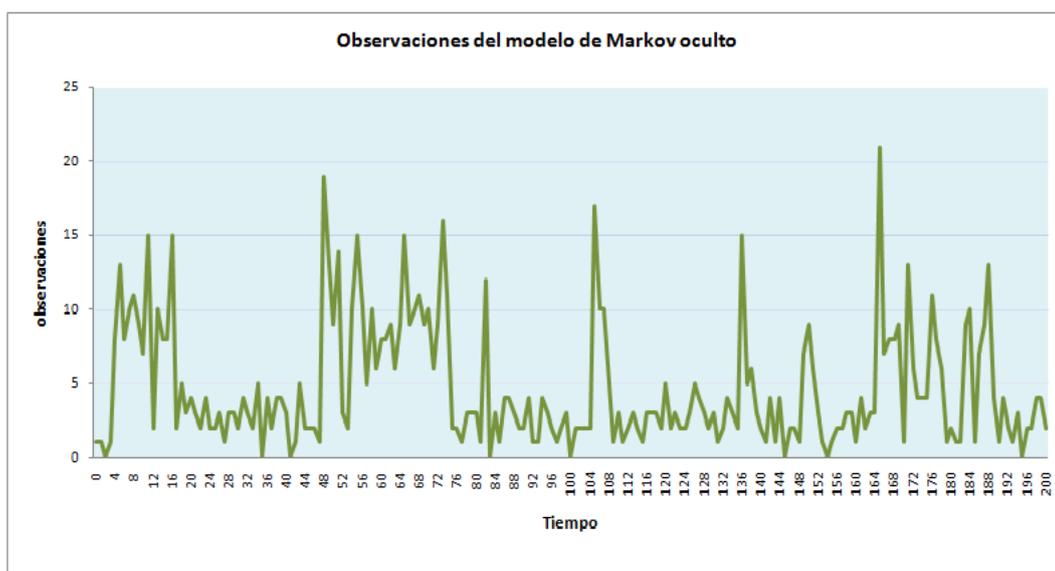


Figura 1.3: Simulación de las observaciones de una cadena de Markov oculta con observaciones discretas.

### 1.3.3. Solución del problema 1 de un modelo de Markov oculto

El problema 1 consiste en determinar la probabilidad de que el modelo  $\lambda = (A, B, \pi)$  sea capaz de generar la secuencia de observaciones  $\mathcal{O} = \{z_1, z_2, z_3, \dots, z_T\}$ , para ello se prosigue de la siguiente manera:

Sea  $\mathcal{O} = \{z_1, z_2, z_3, \dots, z_T\}$  la secuencia de observaciones generadas por el modelo de Markov oculto  $\lambda = (A, B, \pi)$ . Supongamos que  $\mathcal{I} = \{x_1, x_2, x_3, \dots, x_T\}$  es una secuencia fija de estados que ocurren durante el proceso mediante el cual se genera la secuencia de observaciones  $\mathcal{O}$ .

Dado que  $Z_t$  sólo depende de  $X_t$ , la probabilidad de obtener la secuencia de observaciones  $\mathcal{O}$  dada la secuencia fija de estados  $\mathcal{I}$  y el modelo  $\lambda$  está dada por

$$\begin{aligned} P[\mathcal{O}|\mathcal{I}, \lambda] &= P[Z_1 = z_1, Z_2 = z_2, Z_3 = z_3, \dots, Z_T = z_T | X_1 = x_1, X_2 = x_2, X_3 = x_3, \dots, X_T = x_T, \lambda] \\ &= P[Z_1 = z_1 | X_1 = x_1] P[Z_2 = z_2 | X_2 = x_2] \cdots P[Z_T = z_T | X_T = x_T] \\ &= b(z_1|x_1) b(z_2|x_2) \cdots b(z_T|x_T). \end{aligned} \quad (1.19)$$

Por otra parte, la probabilidad de la secuencia fija de estados  $\mathcal{I}$  dado el modelo  $\lambda$  es

$$\begin{aligned} P[\mathcal{I}|\lambda] &= P[X_1 = x_1, X_2 = x_2, X_3 = x_3, \dots, X_T = x_T | \lambda] \\ &= P[X_T = x_T | X_{T-1} = x_{T-1}, \dots, X_1 = x_1] P[X_{T-1} = x_{T-1} | X_{T-2} = x_{T-2}, \dots, X_1 = x_1] \cdots \\ &\quad \cdots P[X_2 = x_2 | X_1 = x_1] P[X_1 = x_1 | \lambda] \\ &= \pi_{x_1} a_{x_1, x_2} a_{x_2, x_3} \cdots a_{x_{T-2}, x_{T-1}} a_{x_{T-1}, x_T}. \end{aligned} \quad (1.20)$$

En consecuencia,

$$\begin{aligned} P[\mathcal{O}, \mathcal{I}|\lambda] &= P[\mathcal{O}|\mathcal{I}, \lambda] P[\mathcal{I}|\lambda] \\ &= \pi_{x_1} b(z_1|x_1) a_{x_1, x_2} b(z_2|x_2) a_{x_2, x_3} \cdots b(z_{T-1}|x_{T-1}) a_{x_{T-1}, x_T} b(z_T|x_T). \end{aligned} \quad (1.21)$$

Aplicando el teorema de la probabilidad total, la probabilidad de la secuencia de observaciones  $\mathcal{O}$  dado el modelo  $\lambda$  se obtiene sumando la probabilidad (1.21) sobre todas las posibles secuencias de estados,

$$\begin{aligned} P[\mathcal{O}|\lambda] &= \sum_{\mathcal{I}} P[\mathcal{O}, \mathcal{I}|\lambda] \\ &= \sum_{x_1, x_2, \dots, x_T} \pi_{x_1} b(z_1|x_1) a_{x_1, x_2} b(z_2|x_2) a_{x_2, x_3} \cdots b(z_{T-1}|x_{T-1}) a_{x_{T-1}, x_T} b(z_T|x_T). \end{aligned} \quad (1.22)$$

Al analizar la ecuación (1.22), se puede observar que para calcular la probabilidad de  $P[\mathcal{O}|\lambda]$  a partir de la definición son necesarias  $(2T - 1)N^T$  multiplicaciones y  $N^T - 1$  sumas. Esta forma de calcular la probabilidad  $P[\mathcal{O}|\lambda]$  tiene un costo computacional enorme. Para resolver el inconveniente existe un procedimiento más eficiente llamado *forward-backward procedure*.

El procedimiento *forward-backward* es una técnica recursiva basada en la definición de dos variables llamadas forward y backward, las cuales se pueden escribir recursivamente lo que facilita calcular la probabilidad  $P[\mathcal{O}|\lambda]$ .

### Cálculo de $P[\mathcal{O}|\lambda]$ mediante la variable forward

**Definición 1.8.** Dado un modelo de Markov oculto  $\lambda = (A, B, \pi)$  y dada la secuencia de observaciones  $\mathcal{O} = \{z_1, z_2, z_3, \dots, z_T\}$  así como el estado en el tiempo  $t$  de la cadena de Markov asociada,  $X_t = q_i$ . La *variable forward* o hacia adelante, denotada por  $\alpha_t(i)$  se define como

$$\alpha_t(i) = P[Z_1 = z_1, Z_2 = z_2, Z_3 = z_3, \dots, Z_t = z_t, X_t = q_i | \lambda]. \quad (1.23)$$

**Proposición 1.1.** Para  $j = 1, 2, 3, \dots, N$  y  $t = 2, 3, \dots, T - 1$ , la variable forward  $\alpha_{t+1}(j)$  satisface la relación de recursividad

$$\alpha_{t+1}(j) = \sum_{i=1}^N \alpha_t(i) a_{q_i, q_j} b(z_{t+1} | q_j) \quad (1.24)$$

**Demostración.** La demostración se realiza por el método de inducción matemática.

Para  $t = 1$ ,  $1 \leq j \leq N$ , de la definición se sigue que

$$\begin{aligned} \alpha_1(j) &= P[Z_1 = z_1, X_1 = q_j | \lambda] \\ &= P[Z_1 = z_1 | X_1 = q_i, \lambda] P[X_1 = q_j | \lambda] \\ &= \pi_{q_j} b(z_1 | q_j). \end{aligned} \quad (1.25)$$

Luego para  $t = 2$ ,  $1 \leq j \leq N$  se obtiene

$$\begin{aligned} \alpha_2(j) &= P[Z_1 = z_1, Z_2 = z_2, X_2 = q_j | \lambda] \\ &= \sum_{i=1}^N P[Z_1 = z_1, Z_2 = z_2, X_1 = q_i, X_2 = q_j | \lambda] \\ &= \sum_{i=1}^N P[Z_1 = z_1 | Z_2 = z_2, X_1 = q_i, X_2 = q_j, \lambda] P[Z_2 = z_2 | X_1 = q_i, X_2 = q_j, \lambda] \\ &\quad P[X_2 = q_j | X_1 = q_i, \lambda] P[X_1 = q_i | \lambda] \\ &= \sum_{i=1}^N b(z_1 | q_i) \pi_{q_i} a_{q_i, q_j} b(z_2 | q_j) \\ &= \sum_{i=1}^N \alpha_1(i) a_{q_i, q_j} b(z_2 | q_j). \end{aligned} \quad (1.26)$$

Ahora suponemos que la afirmación es verdadera para el tiempo  $t$ ; es decir, que para  $1 \leq j \leq N$  se satisface

$$\alpha_t(j) = \sum_{i=1}^N \alpha_{t-1}(i) a_{q_i, q_j} b(z_t | q_j). \quad (1.27)$$

Por lo que es necesario demostrar que la recurrencia es verdadera para el tiempo  $t + 1$ ;

$$\begin{aligned} \alpha_{t+1}(j) &= P[Z_1 = z_1, Z_2 = z_2, \dots, Z_{t+1} = z_{t+1}, X_{t+1} = q_j | \lambda] \\ &= \sum_{i=1}^N P[Z_1 = z_1, Z_2 = z_2, \dots, Z_{t+1} = z_{t+1}, X_t = q_i, X_{t+1} = q_j | \lambda] \\ &= \sum_{i=1}^N P[Z_{t+1} = z_{t+1} | Z_1 = z_1, Z_2 = z_2, \dots, Z_t = z_t, X_t = q_i, X_{t+1} = q_j, \lambda] \\ &\quad P[X_{t+1} = q_j | X_t = q_i, Z_1 = z_1, Z_2 = z_2, \dots, Z_t = z_t, \lambda] P[Z_1 = z_1, Z_2 = z_2, \dots, Z_t = z_t, X_t = q_i | \lambda] \\ &= \sum_{i=1}^N \alpha_t(i) a_{q_i, q_j} b(z_{t+1} | q_j) \end{aligned} \quad (1.28)$$

Esto demuestra la proposición.

El algoritmo para calcular  $P[\mathcal{O}|\lambda]$  empleando la variable forward consiste en:

1. **Inicialización:** Para  $1 \leq i \leq N$  hacer

$$\alpha_1(i) = \pi_{q_i} b(z_1|q_i). \quad (1.29)$$

2. **Inducción:** Para  $1 \leq t \leq T - 1$  y  $1 \leq j \leq N$  hacer

$$\alpha_{t+1}(j) = b(z_{t+1}|q_j) \sum_{i=1}^N \alpha_t(i) a_{q_i, q_j}. \quad (1.30)$$

3. **Finalización:** La probabilidad deseada es igual a

$$\begin{aligned} P[\mathcal{O}|\lambda] &= P[Z_1 = z_1, Z_2 = z_2, \dots, Z_T = z_T|\lambda] \\ &= \sum_{i=1}^N P[Z_1 = z_1, Z_2 = z_2, \dots, Z_T = z_T, X_T = q_i|\lambda] \\ &= \sum_{i=1}^N \alpha_T(i). \end{aligned} \quad (1.31)$$

### Cálculo de $P[\mathcal{O}|\lambda]$ mediante la variable backward

La otra forma de calcular la probabilidad  $P[\mathcal{O}|\lambda]$  es mediante el uso de la variable backward como se muestra enseguida.

**Definición 1.9.** Dado un modelo de Markov oculto  $\lambda = (A, B, \pi)$  y una secuencia de observaciones  $\mathcal{O} = \{z_1, z_2, z_3, \dots, z_T\}$  así como el estado en el tiempo  $t$  de la cadena de Markov asociada,  $X_t = q_i$ . La *variable backward* o hacia atrás, denotada por  $\beta_t(i)$  se define como

$$\beta_t(i) = P[Z_{t+1} = z_{t+1}, Z_{t+2} = z_{t+2}, \dots, Z_T = z_T | X_t = q_i, \lambda] \quad (1.32)$$

para  $t = T - 1, T - 2, \dots, 1$  y  $1 \leq i \leq N$  además  $\beta_T(i) = 1$  para  $1 \leq i \leq N$ .

**Proposición 1.2.** Para  $i = 1, 2, 3, \dots, N$  y  $t = 1, 2, \dots, T - 1$ , la variable backward  $\beta_{t+1}(i)$  satisface la relación de recursividad

$$\beta_t(i) = \sum_{j=1}^N \beta_{t+1}(j) a_{q_i, q_j} b(z_{t+1}|q_j). \quad (1.33)$$

**Demostración.** La demostración nuevamente se realiza empleando el método de inducción matemática.

El caso  $t = T - 1$  y  $1 \leq i \leq N$  se sigue de la definición

$$\begin{aligned} \beta_{T-1}(i) &= P[Z_T = z_T | X_{T-1} = q_i, \lambda] \\ &= \sum_{j=1}^N P[Z_T = z_T, X_T = q_j | X_{T-1} = q_i, \lambda] \\ &= \sum_{j=1}^N P[Z_T = z_T | X_T = q_j, X_{T-1} = q_i, \lambda] P[X_T = q_j | X_{T-1} = q_i, \lambda] \\ &= \sum_{j=1}^N \beta_T(j) a_{q_i, q_j} b(z_T|q_j). \end{aligned} \quad (1.34)$$

Luego, para  $t = T - 2$  y  $1 \leq i \leq N$  se tiene

$$\begin{aligned}
\beta_{T-2}(i) &= P[Z_{T-1} = z_{T-1}, Z_T = z_T | X_{T-2} = q_i, \lambda] \\
&= \sum_{j=1}^N P[Z_{T-1} = z_{T-1}, Z_T = z_T, X_{T-1} = q_j | X_{T-2} = q_i, \lambda] \\
&= \sum_{j=1}^N P[Z_T = z_T | Z_{T-1} = z_{T-1}, X_{T-1} = q_j, X_{T-2} = q_i, \lambda] \\
&\quad P[Z_{T-1} = z_{T-1} | X_{T-1} = q_j, X_{T-2} = q_i, \lambda] P[X_{T-1} = q_j | X_{T-2} = q_i, \lambda] \\
&= \sum_{j=1}^N \beta_{T-1}(j) a_{q_i, q_j} b(z_{T-1} | q_j). \tag{1.35}
\end{aligned}$$

Suponemos que la proposición es válida para  $t = T - k + 1$  con  $4 \leq k \leq T$  y  $1 \leq i \leq N$ ; es decir se cumple que

$$\beta_{T-k+1}(i) = P[Z_{T-k+2} = z_{T-k+2}, Z_{T-k+3} = z_{T-k+3}, \dots, Z_T = z_T | X_{T-k+1} = q_i, \lambda]. \tag{1.36}$$

Por lo que se debe demostrar que la proposición es válida para  $t = T - k$  con  $4 \leq k \leq T - 1$  y  $1 \leq i \leq N$ ;

$$\begin{aligned}
\beta_{T-k}(i) &= P[Z_{T-k+1} = z_{T-k+1}, Z_{T-k+2} = z_{T-k+2}, \dots, Z_T = z_T | X_{T-k} = q_i, \lambda] \\
&= \sum_{j=1}^N P[Z_{T-k+1} = z_{T-k+1}, Z_{T-k+2} = z_{T-k+2}, \dots, Z_T = z_T, X_{T-k+1} = q_j | X_{T-k} = q_i, \lambda] \\
&= \sum_{j=1}^N P[Z_{T-k+2} = z_{T-k+2}, \dots, Z_T = z_T | Z_{T-k+1} = z_{T-k+1}, X_{T-k+1} = q_j, X_{T-k} = q_i, \lambda] \\
&\quad P[Z_{T-k+1} = z_{T-k+1} | X_{T-k+1} = q_j, X_{T-k} = q_i, \lambda] P[X_{T-k+1} = q_j | X_{T-k} = q_i, \lambda] \\
&= \sum_{j=1}^N \beta_{T-k+1}(j) a_{q_i, q_j} b(z_{T-k+1} | q_j). \tag{1.37}
\end{aligned}$$

Lo que demuestra la proposición.

El algoritmo para calcular la  $P[\mathcal{O} | \lambda]$  empleando la variable backward es:

1. **Inicialización.** Para  $i = 1, 2, 3, \dots, N$  hacer

$$\beta_T(i) = 1. \tag{1.38}$$

2. **Inducción.** Para  $t = T - 1, T - 2, \dots, 1$  y  $i = 1, 2, 3, \dots, N$  calcular

$$\beta_t(i) = \sum_{j=1}^N \beta_{t+1}(j) a_{q_i, q_j} b(z_{t+1} | q_j). \tag{1.39}$$

3. Terminación. La probabilidad deseada es

$$\begin{aligned}
P[\mathcal{O}|\lambda] &= P[Z_1 = z_1, Z_2 = z_2, Z_3 = z_3, \dots, Z_T = z_T|\lambda] \\
&= \sum_{i=1}^N P[Z_1 = z_1, Z_2 = z_2, Z_3 = z_3, \dots, Z_T = z_T, X_1 = q_j|\lambda] \\
&= \sum_{i=1}^N P[Z_2 = z_2, Z_3 = z_3, \dots, Z_T = z_T|Z_1 = z_1, X_1 = q_j, \lambda] \\
&\quad P[Z_1 = z_1|X_1 = q_j, \lambda]P[X_1 = q_j|\lambda] \\
&= \sum_{i=1}^N \beta_1(i)b(z_1|q_i)\pi_{q_i}.
\end{aligned} \tag{1.40}$$

### 1.3.4. Solución del problema 2 de un modelo de Markov oculto

El problema 2 consiste en encontrar una secuencia de estados  $\mathcal{I} = \{x_1, x_2, x_3, \dots, x_T\}$  que explique de la mejor manera la ocurrencia de una secuencia de observaciones  $\mathcal{O} = \{z_1, z_2, z_3, \dots, z_T\}$  bajo un modelo de Markov oculto  $\lambda = (A, B, \pi)$ . La dificultad recae en encontrar el criterio o método que permita determinar la secuencia de estados óptimos. En este apartado se revisará dos métodos que permiten abordar el problema: decodificación posterior y decodificación Viterbi.

#### Decodificación posterior

El método decodificación posterior consiste en elegir en cada tiempo  $t$ , el estado óptimo que de forma individual maximiza la posibilidad de la emisión de la observación  $z_t$ . Enseguida se detalla el proceso.

**Definición 1.10.** Dado un modelo de Markov oculto  $\lambda = (A, B, \pi)$  y una secuencia de observaciones  $\mathcal{O} = \{z_1, z_2, z_3, \dots, z_T\}$  del mismo, la variable  $\gamma_t(q_i)$  se define como la probabilidad de que la cadena de Markov  $\{X_t\}$  asociada al modelo  $\lambda$  visite el estado  $q_i$  en el tiempo  $t$  dada la secuencia de observaciones  $\mathcal{O}$  y el modelo  $\lambda$ ; es decir,

$$\begin{aligned}
\gamma_t(q_i) &= \frac{P[X_t = q_i|\mathcal{O}, \lambda]}{P[\mathcal{O}|\lambda]} \\
&= \frac{P[X_t = q_i, \mathcal{O}|\lambda]}{P[\mathcal{O}|\lambda]}
\end{aligned} \tag{1.41}$$

donde  $t = 1, 2, \dots, T$ ,  $i = 1, 2, 3, \dots, N$ .

**Proposición 1.3.** La variable  $\gamma_t(q_i)$  puede reescribirse en término de las variables forward  $\alpha_t(i)$  y backward  $\beta_t(i)$  de la siguiente manera

$$\gamma_t(q_i) = \frac{\alpha_t(i)\beta_t(i)}{P[\mathcal{O}|\lambda]} \tag{1.42}$$

para  $t = 1, 2, \dots, T$ ,  $i = 1, 2, \dots, N$ .

**Demostración.** Sea  $\mathcal{O} = \{z_1, z_2, z_3, \dots, z_T\}$  la secuencia de observaciones de modelo de Markov oculto  $\lambda = (A, B, \pi)$ , observe que,

$$\begin{aligned}
P[X_t = q_i, \mathcal{O}|\lambda] &= P[Z_1 = z_1, Z_2 = z_2, \dots, Z_T = z_T, X_t = q_i|\lambda] \\
&= P[Z_{t+1} = z_{t+1}, Z_{t+2} = z_{t+2}, \dots, Z_T = z_T|Z_1 = z_1, Z_2 = z_2, \dots, Z_t = z_t, X_t = q_i, \lambda] \\
&\quad P[Z_1 = z_1, Z_2 = z_2, \dots, Z_t = z_t, X_t = q_i|\lambda] \\
&= \alpha_t(i)\beta_t(i).
\end{aligned} \tag{1.43}$$

Sustituyendo la ecuación (1.43) en (1.41) se obtiene el resultado deseado,

$$\begin{aligned}
\gamma_t(q_i) &= \frac{\alpha_t(i)\beta_t(i)}{P[\mathcal{O}|\lambda]} \\
&= \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^N P[X_t = q_j, \mathcal{O}|\lambda]} \\
&= \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^N \alpha_t(j)\beta_t(j)}.
\end{aligned} \tag{1.44}$$

El conjunto  $\gamma_t(q_i)$  con  $i = 1, 2, \dots, N$  es una distribución de probabilidad ya que  $\gamma_t(q_i) \geq 0$  y

$$\sum_{i=1}^N \gamma_t(q_i) = \sum_{i=1}^N \left( \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^N \alpha_t(j)\beta_t(j)} \right) = 1. \tag{1.45}$$

Con algunas características de la variable  $\gamma_t(q_i)$  ya mencionadas, se define el método decodificación posterior.

**Definición 1.11.** Dado un modelo de Markov oculto  $\lambda = (A, B, \pi)$ , una secuencia de observaciones  $\mathcal{O} = \{z_1, z_2, z_3, \dots, z_T\}$  y la variable  $\gamma_t(q_i)$  como se definió en (1.41); el método *decodificación posterior* establece que la secuencia de estados  $\mathcal{I} = \{x_1, x_2, x_3, \dots, x_T\}$  que explica de mejor manera la secuencia de observaciones  $\mathcal{O}$  es

$$X_t = \arg \max_{1 \leq i \leq N} [\gamma_t(q_i)] \tag{1.46}$$

para  $t = 1, 2, \dots, T$ .

### Decodificación Viterbi

El ingeniero y hombre de negocios Andrew Viterbi desarrolló en 1976 el algoritmo que ahora lleva su nombre encontrando diversas aplicaciones como el reconocimiento de voz, lingüística computacional y bioinformática, Forney (1973). La idea clave del algoritmo es que cuando se tienen diversos caminos entre dos puntos, trata de encontrar cuál de todos los caminos es el mejor, para reconstruir el camino más probable de una secuencia de entrada. El método decodificación Viterbi está basado en los conceptos de programación dinámica como se muestra enseguida.

**Definición 1.12.** Dado un modelo de Markov oculto  $\lambda = (A, B, \pi)$  y una secuencia de observaciones  $\mathcal{O} = \{z_1, z_2, z_3, \dots, z_T\}$ , se define la variable  $\delta_t(q_i)$ , como

$$\delta_t(q_i) = \max_{x_1, x_2, \dots, x_{t-1}} P[X_1 = x_1, X_2 = x_2, \dots, X_{t-1} = x_{t-1}, X_t = q_i, Z_1 = z_1, Z_2 = z_2, \dots, Z_t = z_t | \lambda] \tag{1.47}$$

para  $i = 1, 2, \dots, N$  y  $t = 2, 3, \dots, T$ . Además para  $i = 1, 2, \dots, N$  se tiene que

$$\begin{aligned}
\delta_1(q_i) &= P[X_1 = q_i, Z_1 = z_1 | \lambda] \\
&= P[Z_1 = z_1 | X_1 = q_i, \lambda] P[X_1 = q_i | \lambda] \\
&= \pi_{q_i} b(z_1 | q_i).
\end{aligned} \tag{1.48}$$

**Proposición 1.4.** Dado un modelo de Markov oculto  $\lambda = (A, B, \pi)$  y una secuencia de observaciones  $\mathcal{O} = \{z_1, z_2, z_3, \dots, z_T\}$ , la variable  $\delta_t(q_i)$  satisface la relación de recursividad

$$\delta_t(q_j) = \max_{q_i} \delta_{t-1}(q_i) a_{q_i, q_j} b(z_t | q_j). \tag{1.49}$$

para  $t = 2, 3, \dots, T$ .

**Demostración.** Sea  $\mathcal{O} = \{z_1, z_2, z_3, \dots, z_T\}$  la secuencia de observaciones del modelo de Markov oculto  $\lambda = (A, B, \pi)$ ; de la definición de  $\delta_t(q_j)$  se sigue

$$\begin{aligned}
\delta_t(q_j) &= \max_{x_1, x_2, \dots, x_{t-1}} P[X_1 = x_1, X_2 = x_2, \dots, X_{t-1} = x_{t-1}, X_t = q_j, Z_1 = z_1, Z_2 = z_2, \dots, Z_t = z_t | \lambda] \\
&= \max_{q_i} \max_{x_1, x_2, \dots, x_{t-2}} P[X_1 = x_1, \dots, X_{t-2} = x_{t-2}, X_{t-1} = q_i, X_t = q_j, Z_1 = z_1, \dots, Z_t = z_t | \lambda] \\
&= \max_{q_i} \max_{x_1, x_2, \dots, x_{t-2}} P[Z_t = z_t | X_1 = x_1, \dots, X_{t-2} = x_{t-2}, X_{t-1} = q_i, X_t = q_j, Z_1 = z_1, \dots, Z_{t-1} = z_{t-1}, \lambda] \\
&\quad P[X_t = q_j | X_1 = x_1, \dots, X_{t-2} = x_{t-2}, X_{t-1} = q_i, Z_1 = z_1, \dots, Z_{t-1} = z_{t-1}, \lambda] \\
&= \max_{q_i} \delta_{t-1}(q_i) a_{q_i, q_j} b(z_t | q_j). \tag{1.50}
\end{aligned}$$

Una vez definida la variable  $\delta_t(q_i)$ ; el algoritmo del método Viterbi es:

1. **Inicialización.** Para  $i = 1, 2, \dots, N$  hacer

$$\begin{aligned}
\delta_1(q_i) &= \pi_{q_i} b(z_1 | q_i), \\
\psi_1(i) &= 0. \tag{1.51}
\end{aligned}$$

2. **Recursión.** Para  $t = 2, 3, \dots, N$  y  $j = 1, 2, \dots, N$  calcular

$$\delta_t(q_j) = \max_{q_i} \delta_{t-1}(q_i) a_{q_i, q_j} b(z_t | q_j). \tag{1.52}$$

$$\psi_t(q_j) = \arg \max_{q_i} \delta_{t-1}(q_i) a_{q_i, q_j}. \tag{1.53}$$

3. **Terminación.**

$$x_T^* = \arg \max_{q_i} \delta_T(q_i). \tag{1.54}$$

4. **Trayectoria de estados.** Para  $t = T - 1, T - 2, \dots, 2, 1$ , hacer

$$x_t^* = \psi_{t+1}(q_{t+1}^*). \tag{1.55}$$

La trayectoria de estados deseada es:  $\{x_1^*, x_2^*, \dots, x_T^*\}$ .

### 1.3.5. Solución del problema 3 de un modelo de Markov oculto

El último y más difícil problema que se presenta en un modelo de Markov oculto es la estimación de los parámetros del modelo. Dada la secuencia de observaciones  $\mathcal{O} = \{z_1, z_2, z_3, \dots, z_T\}$  se quiere encontrar los parámetros del modelo  $\lambda = (A, B, \pi)$  que mejor explique la secuencia de observaciones. Es decir, se quiere encontrar  $\hat{\lambda}$  que maximiza la siguiente probabilidad  $P[\mathcal{O} | \lambda]$ ,

$$\hat{\lambda} = \arg \max_{\lambda} P[\mathcal{O} | \lambda]. \tag{1.56}$$

El método para tratar el problema 3 de un modelo de Markov oculto es el *algoritmo Baum-Welch*.

## Algoritmo Baum-Welch

El *algoritmo Baum-Welch* es un método de maximización local y es un caso especial del método EM por las siglas en inglés del método de máxima esperanza. Se trata de un proceso iterativo en el que a cada paso se obtiene un estimador  $\hat{\lambda}$  que incrementa la probabilidad  $P[\mathcal{O}|\lambda]$ .

La idea principal del algoritmo es elegir arbitrariamente los valores iniciales de los parámetros del modelo  $\lambda = (A, B, \pi)$ . Usando estos valores asignados y la secuencia de observaciones se estima el número esperado de transiciones que ocurren del estado  $q_i$  al estado  $q_j$  para todos los posibles  $q_i$  y  $q_j$  estados de la cadena de Markov. Asimismo, se estima el número de emisiones del símbolo  $v_k$  en el estado  $q_i$  para  $k = 1, 2, \dots, M$  y todo  $q_i$  estado de la cadena. Estos cálculos son utilizados para recalcular los parámetros del modelo. Este proceso continúa hasta que se alcanza cierto criterio de paro.

Para trabajar iterativamente el modelo  $\lambda = (A, B, \pi)$  es necesario definir algunas variables intermedias y los estimadores a emplear.

**Definición 1.13.** Dado un modelo de Markov oculto  $\lambda = (A, B, \pi)$ , una secuencia de observaciones  $\mathcal{O} = \{z_1, z_2, z_3, \dots, z_T\}$  y los estados  $q_i, q_j$  de la cadena de Markov asociada al modelo; la probabilidad de transición del estado  $q_i$  al estado  $q_j$  en el tiempo  $t$  dada la secuencia de observaciones  $\mathcal{O}$  y el modelo  $\lambda$ , es igual a

$$\begin{aligned} \xi_t(q_i, q_j) &= P[X_t = q_i, X_{t+1} = q_j | \mathcal{O}, \lambda] \\ &= \frac{P[X_t = q_i, X_{t+1} = q_j, \mathcal{O}, \lambda]}{P[\mathcal{O}, \lambda]} \\ &= \frac{P[X_t = q_i, X_{t+1} = q_j, \mathcal{O}, \lambda]}{P[\mathcal{O}|\lambda]P[\lambda]} \end{aligned} \quad (1.57)$$

**Proposición 1.5.** La variable  $\xi_t(q_i, q_j)$  puede reescribirse como

$$\xi_t(q_i, q_j) = \frac{\alpha_t(i)a_{q_i, q_j}b(z_{t+1}|q_j)\beta_{t+1}(j)}{P[\mathcal{O}|\lambda]}. \quad (1.58)$$

**Demostración.** Sea  $\mathcal{O} = \{z_1, z_2, z_3, \dots, z_T\}$  la secuencia de observaciones del modelo  $\lambda = (A, B, \pi)$  y sean  $q_i, q_j$  cualesquiera estados de la cadena de Markov asociada al modelo. Observe que el numerador de la ecuación (1.57) puede escribirse como

$$\begin{aligned} P[X_t = q_i, X_{t+1} = q_j, \mathcal{O}, \lambda] &= P[Z_{t+1} = z_{t+1} | X_{t+1} = q_j, Z_1 = z_1, Z_2 =, \dots, Z_t = z_t, X_t = q_i, \\ &\quad Z_{t+2} = z_{t+2}, \dots, Z_T = z_T, \lambda] P[Z_{t+2} = z_{t+2}, \dots, Z_T = z_T | X_{t+1} = q_j, \\ &\quad Z_1 = z_1, Z_2 =, \dots, Z_t = z_t, X_t = q_i, \lambda] \\ &= P[X_{t+1} = q_j | X_t = q_i, Z_1 = z_1, Z_2 =, \dots, Z_t = z_t, \lambda] \\ &\quad P[Z_1 = z_1, Z_2 =, \dots, Z_t = z_t, X_t = q_i | \lambda] P[\lambda] \\ &= \alpha_t(i)a_{q_i, q_j}b(z_{t+1}|q_j)\beta_{t+1}(j)P[\lambda] \end{aligned} \quad (1.59)$$

Sustituyendo la ecuación (1.59) en (1.57) se obtiene el resultado deseado. ■

Es importante mencionar lo siguiente:

- De la definición de la variable  $\gamma_t(q_i)$  en la ecuación (1.41), se observa que

$$\begin{aligned}
 \gamma_t(q_i) &= P[X_i = q_i | \mathcal{O}, \lambda] \\
 &= \sum_{j=1}^N P[X_i = q_i, X_{t+1} = q_j | \mathcal{O}, \lambda] \\
 &= \sum_{j=1}^N \xi_t(q_i, q_j)
 \end{aligned} \tag{1.60}$$

- La expresión  $\sum_{t=1}^{T-1} \xi_t(q_i, q_j)$  se puede interpretar como el número esperado de transiciones del estado  $q_i$  al estado  $q_j$  a lo largo del tiempo.
- Asimismo,  $\sum_{t=1}^{T-1} \gamma_t(q_i)$  puede interpretarse como el número esperado de veces que el estado  $q_i$  es visitado a lo largo del tiempo; es decir, el número esperado de transiciones hechas desde el estado  $q_i$ .

Nos interesa determinar estimadores de la matriz de transición  $A = \{a_{q_i, q_j}\}$ , del conjunto de distribuciones  $B = \{b(v_k | q_j)\}$  y de la distribución de probabilidad inicial  $\pi = \{\pi_{q_i}\}$  para  $q_i, q_j$  estados de la cadena de Markov asociada al modelo. En seguida se definen estos estimadores.

**Definición 1.14.** En un modelo de Markov oculto, el estimador de la matriz de probabilidad de transición de la cadena de Markov, será denotada como  $\hat{A} = \{\hat{a}_{q_i, q_j}\}$ . Donde  $\hat{a}_{q_i, q_j}$  es el estimador de la probabilidad de transición del estado  $q_i$  al estado  $q_j$ , para cualesquiera  $q_i, q_j$  estados de la cadena de Markov y se define como:

$$\hat{a}_{q_i, q_j} = \frac{\text{Número esperado de transiciones del estado } q_i \text{ al estado } q_j.}{\text{Número esperado de transiciones desde el estado } q_i.} \tag{1.61}$$

**Definición 1.15.** En un modelo de Markov oculto, el estimador de las distribuciones de probabilidad asociadas a los estados de la cadena de Markov será denotado como  $\hat{B} = \{\hat{b}(v_k | q_i)\}$  para  $q_i, q_j$  estados de la cadena de Markov y para todo  $v_k$  elemento del soporte de las distribuciones. Donde  $\hat{b}(v_k | q_i)$  es el estimador de la probabilidad de observar  $v_k$  dado que la cadena se encuentra en el estado  $q_i$  y se define como:

$$\hat{b}(v_k | q_i) = \frac{\text{Número esperado de transiciones desde el estado } q_i \text{ y con observación } v_k.}{\text{Número esperado de transiciones desde el estado } q_i.} \tag{1.62}$$

**Definición 1.16.** En un modelo de Markov oculto, el estimador de la distribución inicial de probabilidad de la cadena de Markov  $\{X_t\}$  será denotado como  $\hat{\pi} = \{\hat{\pi}_{q_i}\}$ , para todo  $q_i$  estado de la cadena, donde

$$\hat{\pi}_{q_i} = P[X_1 = q_i | \mathcal{O}, \lambda] = \gamma_1(q_i). \tag{1.63}$$

■

Una vez definida la notación para los estimadores del modelo de Markov oculto, en seguida se presenta el algoritmo de Baum-Welch.

**Algoritmo Baum-Welch.**

1. **Inicialización:** Elegir aleatoriamente los estimadores iniciales del modelo  $\hat{A}, \hat{B}, \hat{\pi}$ . Fijar  $P_1 = 0$ .

2. Recursión:

- Para  $i = 1, 2, \dots, N$  y  $t = 1, 2, \dots, T$  calcular  $\alpha_t(i)$  y  $\beta_t(i)$  usando los valores de  $\hat{A}$ ,  $\hat{B}$ ,  $\hat{\pi}$ .
- Para  $i, j = 1, 2, \dots, N$  calcular

$$\sum_{t=1}^{T-1} \xi_t(q_i, q_j) = \sum_{t=1}^{T-1} \frac{\alpha_t(i) \hat{a}_{q_i, q_j} \hat{b}(z_{t+1}|q_j) \beta_{t+1}(j)}{P[\mathcal{O}|\lambda]} \quad (1.64)$$

$$\sum_{t=1}^{T-1} \gamma_t(q_i) = \sum_{t=1}^{T-1} \frac{\alpha_t(i) \beta_t(i)}{P[\mathcal{O}|\lambda]} \quad (1.65)$$

- Para  $i = 1, 2, \dots, N$  y  $k = 1, 2, \dots, M$  calcular

$$\sum_{t|Z_t=v_k, t=1,2,\dots,T-1} \gamma_t(q_i) = \sum_{t|Z_t=v_k, t=1,2,\dots,T-1} \frac{\alpha_t(i) \beta_t(i)}{P[\mathcal{O}|\lambda]} \quad (1.66)$$

$$\gamma_1(q_i) = \frac{\alpha_1(i) \beta_1(i)}{P[\mathcal{O}|\lambda]} \quad (1.67)$$

- Nueva estimación de los parámetros del modelo: Para  $i, j = 1, 2, \dots, N$  y  $k = 1, 2, \dots, M$  calcular

$$\hat{a}_{q_i, q_j} = \frac{\sum_{t=1}^{T-1} \xi_t(q_i, q_j)}{\sum_{t=1}^{T-1} \gamma_t(q_i)} \quad (1.68)$$

$$\hat{b}(v_k|q_i) = \frac{\sum_{t|Z_t=v_k, t=1,2,\dots,T-1} \gamma_t(q_i)}{\sum_{t=1}^{T-1} \gamma_t(q_i)} \quad (1.69)$$

$$\hat{\pi}_{q_i} = \gamma_1(q_i). \quad (1.70)$$

3. Finalización: Calcular  $P_2 = P[\mathcal{O}|\lambda]$  usando los estimadores anteriores.

Comparar los valores de  $P_1$  y  $P_2$ :

- Si  $P_2 - P_1 > cota$  hacer  $P_1 = P_2$  y regresar al paso 2.
- De lo contrario se ha terminado el proceso y los estimadores del modelo de Markov oculto son  $\hat{A}$ ,  $\hat{B}$ ,  $\hat{\pi}$ .

## 1.4. Modelos de Markov ocultos con espacio de observaciones continuo

En esta sección se describe un modelo de Markov oculto con espacio de observaciones continuo. En la primer parte se menciona los elementos del modelo sin considerar una distribución de probabilidad en particular para las observaciones del modelo. Más adelante se obtienen estimadores de los parámetros del modelo suponiendo algunas distribuciones de probabilidad, en particular se obtienen resultados para la distribución gaussiana multivariable, mezcla de gaussianas multivariadas, la distribución gamma y distribución lognormal.

### 1.4.1. Elementos de un HMM con espacio de observaciones continuo

Como se verá, un HMM con espacio de observaciones continuo tiene una estructura casi idéntica al caso discreto, sin embargo existen diferencias fundamentales en el proceso de estimar los parámetros del modelo. Recordando la definición al inicio de este capítulo; un *modelo de Markov oculto con espacio de observaciones continuo* es un proceso bivariado  $\{(X_t, Z_t), t \geq 0\}$  tal que  $\{X_t, t \geq 0\}$  es una cadena de Markov y  $\{Z_t, t \geq 0\}$  es el proceso de observaciones del modelo, donde la variable aleatoria  $\{Z_t\}$  es una variable aleatoria continua.

Prosiguiendo, en este modelo se identifican los siguiente elementos:

1. Sea  $T$  el número total de tiempos en los que se observa el modelo de Markov oculto.
2. En el modelo ocurren dos procesos de forma simultánea, el primero de ellos es una cadena de Markov,  $\{X_t\}$ , que tiene un número finito  $N$  de estados que serán denotados como

$$Q = \{q_1, q_2, q_3, \dots, q_N\}. \quad (1.71)$$

La cadena de Markov tiene asociada una matriz de probabilidades de transición denotada por  $A = \{a_{q_i, q_j}\}$ , donde

$$a_{q_i, q_j} = P[X_{t+1} = q_j | X_t = q_i]. \quad (1.72)$$

Además, la cadena de Markov tiene una distribución de probabilidad inicial denotada por  $\pi = \{\pi_{q_i}\}$ , donde

$$\pi_{q_i} = P[X_1 = q_i]. \quad (1.73)$$

3. El segundo proceso que se presenta en el modelo, es el proceso de observaciones denotado como  $\{Z_t\}$ . En este proceso la variable  $Z_t$  es una variable aleatoria continua.
4. Cada estado de la cadena de Markov  $\{X_t\}$  tiene asociada una función de densidad de probabilidad para la variable aleatoria  $Z_t$ . Por lo tanto, se tienen  $N$  funciones de densidad de probabilidad para la variable  $Z_t$ , una por cada estado de la cadena de Markov; este conjunto de funciones es denotado como

$$\mathcal{B} = \{b(z_t | q_j)\} \text{ para } j = 1, 2, 3, \dots, N. \quad (1.74)$$

donde para cada  $j = 1, 2, 3, \dots, N$  se tiene,

$$b(z_t | q_j) = f_{Z_t | X_t = q_j}(z_t) \quad (1.75)$$

y  $f_{Z_t | X_t = q_j}(z_t)$  es la función de densidad condicional de la variable aleatoria  $Z_t$  dado que  $X_t = q_j$ . Cabe mencionar que la densidad de probabilidad de la variable  $Z_t$  está determinada únicamente por el estado  $X_t$  por lo que  $Z_t$  es independiente del estado  $X_{t'}$  y de la variable  $Z_{t'}$  siempre que  $t \neq t'$ . Cabe mencionar que cuando  $X_t$  es desconocido,  $Z_t$  depende de  $X_{t-1}$  debido a la propiedad de Markov.

Considerando los elementos del modelo de Markov oculto con espacio de observaciones continuo, de aquí en adelante éste será denotado como  $\lambda_c = (A, \mathcal{B}, \pi)$  para hacer diferencia con el caso discreto.

Al igual que en el caso discreto, el proceso para generar un conjunto de observaciones es el siguiente:

1. Definir  $T$ , el número de observaciones deseadas del modelo de Markov oculto.
2. Fijar el estado inicial de la cadena de Markov  $x_1$ , una realización de la variable aleatoria  $X_1$ , de acuerdo a la distribución de probabilidad inicial  $\pi$  de la cadena.
3. Fijar  $t = 1$
4. Elegir una observación  $z_t$ , una realización de la variable aleatoria  $Z_t$ , de acuerdo a la función de densidad de probabilidad asociada al estado  $x_t$ , es decir  $b(z_t|x_t)$ .
5. Elegir  $x_{t+1}$ , es decir la realización de la variable  $X_{t+1}$ , a partir de la matriz de probabilidades de transición  $A$ .
6. Fijar  $t = t + 1$ .
7. Si  $t < T$  regresar al paso 4.
8. Al final el proceso genera un conjunto de estados denotado como  $\mathcal{I} = \{x_1, x_2, x_3, \dots, x_T\}$ , así como el conjunto de observaciones  $\mathcal{O} = \{z_1, z_2, z_3, \dots, z_T\}$ .

Enseguida se presenta un ejemplo de modelo de Markov oculto con espacio de observaciones continuo.

■ **Ejemplo 1.3.** Considere un modelo de Markov oculto que consta de una cadena de Markov  $\{X_t\}$  con dos posibles estados,  $Q = \{1, 2\}$ , cada estado tiene asociada una función de densidad de probabilidad para el proceso de observaciones del modelo  $\{Z_t\}$ . La distribución de probabilidad inicial y la matriz de transición de la cadena de Markov están dadas por

$$\pi = (P[X_1 = 1], P[X_1 = 2]) = (0,7, 0,3) \quad (1.76)$$

$$A = \begin{pmatrix} P[X_{t+1} = 1|X_t = 1] & P[X_{t+1} = 2|X_t = 1] \\ P[X_{t+1} = 1|X_t = 2] & P[X_{t+1} = 2|X_t = 2] \end{pmatrix} = \begin{pmatrix} 0,95 & 0,05 \\ 0,02 & 0,98 \end{pmatrix}. \quad (1.77)$$

Además, las funciones de densidad asociadas a los estados son:

Distribución de probabilidad Gaussiana con media  $\mu = 4$  y desviación estándar  $\sigma = 10$ .

$$b(z_t|q_1 = 1) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z_t-\mu}{\sigma}\right)^2} = \frac{1}{10\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z_t-4}{10}\right)^2}, \quad (1.78)$$

Distribución de probabilidad Gaussiana con media  $\mu = 4$  y desviación estándar  $\sigma = 1$ .

$$b(z_t|q_1 = 2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z_t-\mu}{\sigma}\right)^2} = \frac{1}{1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z_t-4}{1}\right)^2} \quad (1.79)$$

Siguiendo a Laverty et al. (2002), se generaron un conjunto de 200 observaciones del modelo de Markov oculto con espacio de observaciones continuo. En la figura (1.4) se muestra la gráfica de los estados de la cadena de Markov. Asimismo, en la figura (1.5) se muestra las 200 observaciones que se obtuvieron en la simulación de modelo.

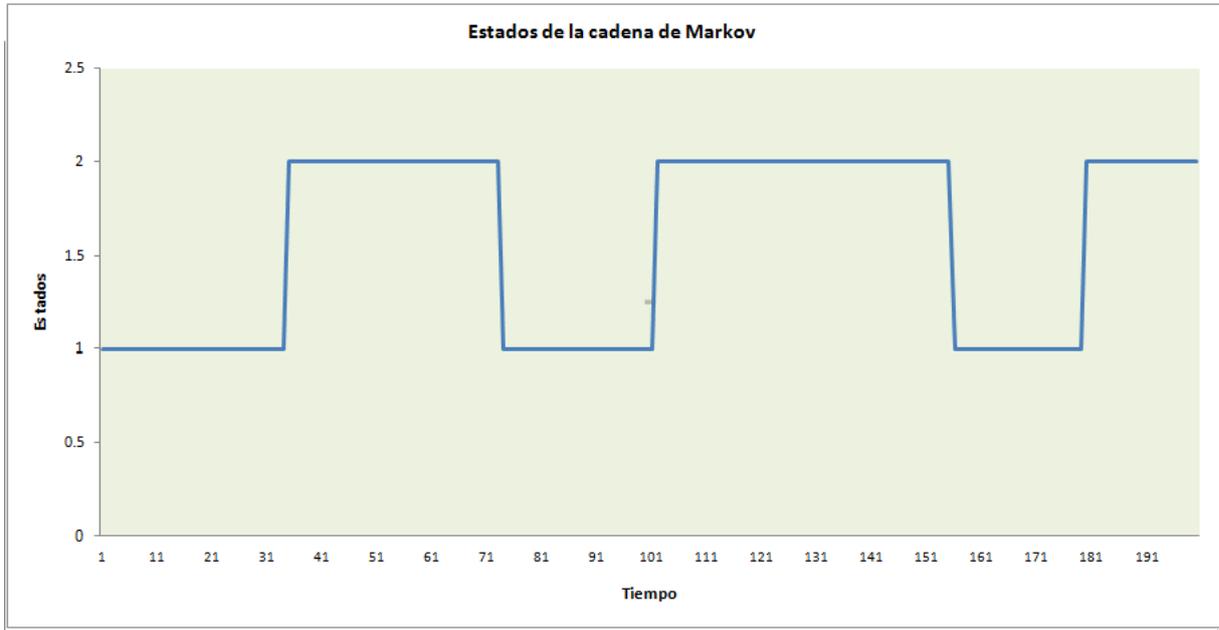


Figura 1.4: Simulación de los estados de una cadena de Markov oculta con observaciones continuas.

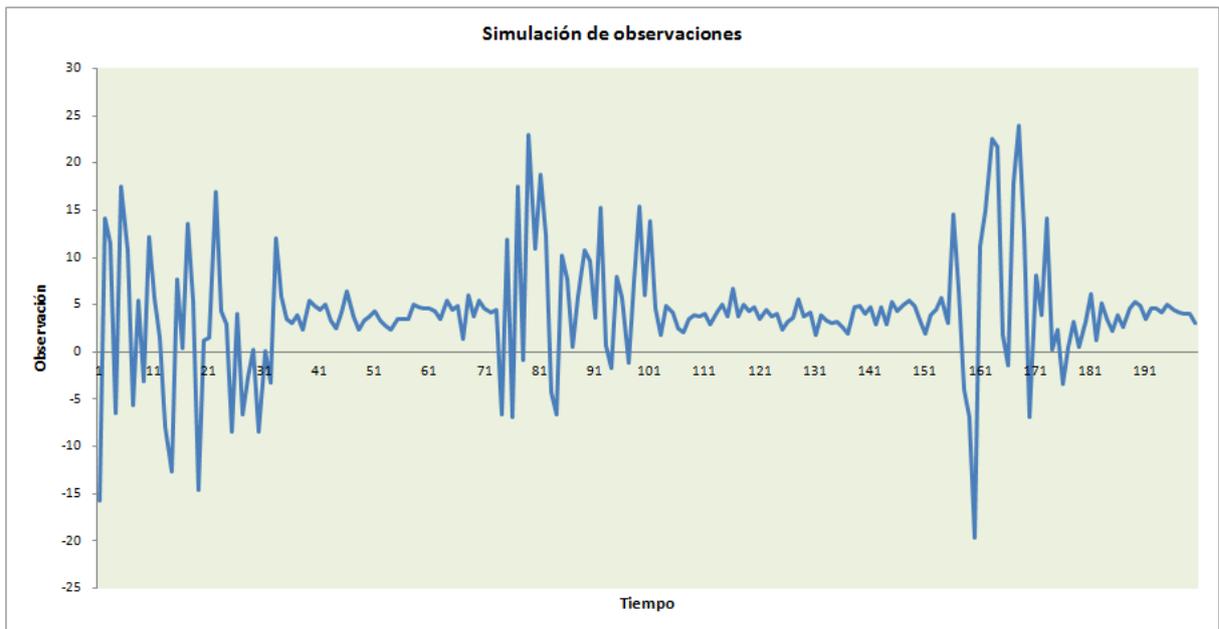


Figura 1.5: Simulación de las observaciones de una cadena de Markov con observaciones continuas.

### 1.4.2. Algunas características de un HMM con espacio de observaciones continuo

Enseguida se definen algunos conceptos que serán útiles más adelante cuando se determinen los estimadores de los parámetros de modelo.

**Definición 1.17.** Dado un modelo de Markov oculto con espacio de observaciones continuo,  $\lambda_c = (A, \mathcal{B}, \pi)$ , la secuencia de estados de la cadena de Markov asociada al modelo  $\mathcal{I} = \{x_1, x_2, x_3, \dots, x_T\}$ , así como un conjunto de observaciones del modelo  $\mathcal{O} = \{z_1, z_2, z_3, \dots, z_T\}$ . La *densidad conjunta* de  $(\mathcal{O}, \mathcal{I})$  es, Aspirot y Bermole (2004):

$$p(z_1, z_2, z_3, \dots, z_T, x_1, x_2, x_3, \dots, x_T | \lambda_c) = p(z_1, x_1) \prod_{t=2}^T p(z_t, x_t | x_{t-1}) \quad (1.80)$$

donde,

$$p(z_1, x_1) = b(z_1 | x_1) P(X_1 = x_1) = f_{Z_1 | X_1 = x_1}(z_1) \pi_{x_1}, \quad (1.81)$$

$$p(z_t, x_t | x_{t-1}) = b(z_t | x_t) P(X_t = x_t | X_{t-1} = x_{t-1}) = f_{Z_t | X_t = x_t}(z_t) a_{x_{t-1}, x_t}. \quad (1.82)$$

■

De la definición anterior se sigue que la función de verosimilitud para la secuencia de observaciones  $\mathcal{O} = \{z_1, z_2, z_3, \dots, z_T\}$  es:

$$p(z_1, z_2, z_3, \dots, z_T) = \sum_{\mathcal{I}} \pi_{x_1} b(z_1 | x_1) \prod_{t=1}^{T-1} a_{x_t, x_{t+1}} b(z_{t+1} | x_{t+1}). \quad (1.83)$$

Es importante establecer que la función de verosimilitud descrita puede ser expresada en términos de otras funciones de densidad conocidas como densidad hacia adelante y densidad hacia atrás.

**Definición 1.18.** Dado un modelo de Markov oculto con espacio de observacion continuo  $\lambda_c = (A, \mathcal{B}, \pi)$ , una secuencia de observaciones  $\mathcal{O} = \{z_1, z_2, z_3, \dots, z_T\}$  y la variable  $X_t = q_i$ , entonces la *función de densidad hacia adelante* se denota como  $\alpha_t(i)_c$  y se define así

$$\alpha_t(i)_c = p(z_1, z_2, z_3, \dots, z_t, X_t = q_i | \lambda_c), \quad (1.84)$$

para  $t = 1, 2, \dots, T$  y  $i = 1, 2, \dots, N$ . ■

De la Definición 1.18 y siguiendo el proceso por el cual se demostró la Proposición 1.1 se demuestra que la densidad hacia adelante satisface lo siguiente, Bilmes (1998):

$$\alpha_{t+1}(j)_c = \begin{cases} \pi_{q_j} b(z_1 | q_j) & \text{si } t = 0, \\ b(z_{t+1} | q_j) \sum_{i=1}^N \alpha_t(i)_c a_{q_i, q_j} & \text{si } t = 1, 2, \dots, T - 1. \end{cases} \quad (1.85)$$

**Definición 1.19.** Dado un modelo de Markov oculto con espacio de observacion continuo  $\lambda_c = (A, \mathcal{B}, \pi)$ , una secuencia de observaciones  $\mathcal{O} = \{z_1, z_2, z_3, \dots, z_T\}$  y la variable  $X_t = q_i$ , entonces la *función de densidad hacia atrás* se denota como  $\beta_t(i)_c$  y se define de la siguiente manera

$$\beta_t(i)_c = p(z_{t+1}, z_{t+2}, z_{t+3}, \dots, z_T | X_t = q_i, \lambda_c), \quad (1.86)$$

para  $t = 1, 2, \dots, T$  y  $i = 1, 2, \dots, N$ . ■

De la Definición 1.19 y siguiendo el proceso por el cual se demostró la Proposición 1.2 se demuestra que la función de densidad hacia atrás satisface,

$$\beta_t(j)_c = \begin{cases} 1 & \text{si } t = T, \\ \sum_{i=1}^N \beta_{t+1}(i)_c a_{q_i, q_j} b(z_{t+1}|q_i) & \text{si } t = T-1, T-2, \dots, 1. \end{cases} \quad (1.87)$$

De esta manera la función de verosimilitud de la secuencia de observaciones  $\mathcal{O} = \{z_1, z_2, z_3, \dots, z_T\}$  puede escribirse como

$$p(z_1, z_2, z_3, \dots, z_T | \lambda_c) = \sum_{i=1}^N \alpha_T(i)_c, \quad (1.88)$$

ó bien como

$$p(z_1, z_2, z_3, \dots, z_T | \lambda_c) = \sum_{i=1}^N \beta_1(i) b(z_1|q_i) \pi_{q_i}. \quad (1.89)$$

Ahora se define el número esperado de veces que una cadena de Markov asociada a un HMM visita cada uno de sus posibles estados, así como el número de transiciones que se realizan entre los estados; esta notación será útil para simplificar los estimadores del modelo más adelante.

**Definición 1.20.** Dado un modelo de Markov oculto con espacio de observación continuo  $\lambda_c = (A, \mathcal{B}, \pi)$ , una secuencia de observaciones  $\mathcal{O} = \{z_1, z_2, z_3, \dots, z_T\}$ ; la función de densidad condicional de la variable  $X_t = q_i$  dada la secuencia  $\mathcal{O}$  y el modelo  $\lambda_c$  se denota como  $\gamma_t(q_i)_c$  y es igual a

$$\gamma_t(q_i)_c = p(X_t = q_i | \mathcal{O}, \lambda_c), \quad (1.90)$$

para  $i = 1, 2, 3, \dots, N$  y  $t = 1, 2, 3, \dots, T$ . ■

Observe que además,

$$\begin{aligned} \gamma_t(q_i)_c &= \frac{p(\mathcal{O}, X_t = q_i)}{p(\mathcal{O} | \lambda_c)} \\ &= \frac{p(z_1, z_2, \dots, z_t, X_t = q_i | \lambda_c) p(z_{t+1}, z_{t+2}, \dots, z_T | X_t = q_i, \lambda_c)}{\sum_{j=1}^N p(\mathcal{O}, X_t = q_j | \lambda_c)} \\ &= \frac{\alpha_t(i)_c \beta_t(i)_c}{\sum_{j=1}^N \alpha_t(j)_c \beta_t(j)_c}. \end{aligned} \quad (1.91)$$

**Definición 1.21.** Dado un modelo de Markov oculto con espacio de observación continuo  $\lambda_c = (A, \mathcal{B}, \pi)$ , una secuencia de observaciones  $\mathcal{O} = \{z_1, z_2, z_3, \dots, z_T\}$ , la función de densidad conjunta de las variables  $X_t = q_i, X_{t-1} = q_j$  condicionada a la secuencia  $\mathcal{O}$  y el modelo  $\lambda_c$  se denota como  $\xi_t(q_i, q_j)_c$  y es igual a

$$\xi_t(q_i, q_j)_c = p(X_t = q_i, X_{t+1} = q_j | \mathcal{O}, \lambda_c), \quad (1.92)$$

para  $i, j = 1, 2, 3, \dots, N$  y  $t = 1, 2, 3, \dots, T-1$ . ■

Además,

$$\begin{aligned} \xi_t(q_i, q_j)_c &= \frac{p(X_t = q_i, X_{t+1} = q_j, \mathcal{O} | \lambda_c)}{p(\mathcal{O} | \lambda_c)} \\ &= \frac{\alpha_t(i)_c a_{q_i, q_j} b(z_{t+1}|q_j) \beta_{t+1}(j)_c}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i)_c a_{q_i, q_j} b(z_{t+1}|q_j) \beta_{t+1}(j)_c}. \end{aligned} \quad (1.93)$$

Ó bien,

$$\begin{aligned}\xi_t(q_i, q_j)_c &= \frac{p(X_t = q_i)p(z_{t+1}, \dots, z_T, X_{t+1} = q_j | X_t = q_i, \lambda_c)}{p(z_{t+1}, \dots, z_T | X_t = q_i, \lambda_c)} \\ &= \frac{\gamma_t(q_i)_c a_{q_i, q_j} b(z_{t+1} | q_j) \beta_{t+1}(j)_c}{\beta_t(i)_c}.\end{aligned}\quad (1.94)$$

### 1.4.3. Algoritmo EM para estimar los parámetros de un HMM

Al igual que en el caso discreto, es importante determinar los parámetros de un modelo de Markov oculto con espacio de observaciones continuo. Se verá que el camino apropiado para realizar la estimación es empleando el método EM o máxima esperanza por ese motivo se desarrolla a grandes rasgos el principio por el que funciona dicho método.

#### Estimación de máxima verosimilitud

Como antecedente al método EM, explicaremos el problema de estimación de máxima verosimilitud. Consideré un conjunto de vectores de datos denotado como  $\mathcal{W} = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_M\}$  donde  $\mathbf{w}_i \in \mathbb{R}^n$  es un vector de datos. También suponga que el conjunto de vectores son independientes e idénticamente distribuidos con distribución  $p(\mathbf{w}|\Theta)$ , donde  $\Theta$  es un conjunto de parámetros de la distribución. En consecuencia, la densidad conjunta de  $\mathcal{W}$  es

$$p(\mathcal{W}|\Theta) = \prod_{i=1}^M p(\mathbf{w}_i|\Theta) = \mathcal{L}(\Theta|\mathcal{W}).\quad (1.95)$$

La función  $\mathcal{L}(\Theta|\mathcal{W})$  se llama *función de verosimilitud* de los parámetros  $\Theta$  dados los datos  $\mathcal{W}$ . Se visualiza como una función de los parámetros  $\Theta$  donde los datos  $\mathcal{W}$  son fijos. Prosiguiendo, el problema de estimación de máxima verosimilitud tiene el objetivo de encontrar  $\Theta$  tal que maximice  $\mathcal{L}$ . Es decir, se desea encontrar  $\Theta^*$  donde

$$\Theta^* = \arg \max \mathcal{L}(\Theta|\mathcal{W}).\quad (1.96)$$

En ocasiones se maximiza  $\log(\mathcal{L}(\Theta|\mathcal{W}))$  ya que resulta más fácil de tratar. Cabe mencionar que dependiendo de la forma de  $p(\mathbf{w}_i|\Theta)$  es la dificultad del problema. Por ejemplo, si  $p(\mathbf{w}_i|\Theta)$  es la densidad Gaussiana donde  $\Theta = (\mu, \sigma^2)$ , entonces únicamente es necesario determinar las derivadas de  $\log(\mathcal{L}(\Theta|\mathcal{W}))$  con respecto a los parámetros e igualar a cero y finalmente resolver para  $\mu$  y  $\sigma^2$ . Desafortunadamente, para muchos problemas no es posible encontrar expresiones analíticas por lo que se debe recurrir a técnicas más elaboradas.

#### Algoritmo EM

El algoritmo EM es una técnica más elaborada que el método de estimación de máxima verosimilitud. El algoritmo EM, Dempster et al. (1977), Ghahramani y Jordan (1995), Jordan y Jacobs (1994), es un método general para encontrar los estimadores de máxima verosimilitud cuando la información sobre los datos es incompleta.

Existen dos principales aplicaciones del algoritmo EM:

- Cuando la información sobre los datos es incompleta debido a limitaciones en el proceso de observación de la muestra.

- Cuando la optimización de la función de verosimilitud es analíticamente intratable; sin embargo, es posible simplificar la función de verosimilitud suponiendo la existencia de más variables no observadas.

Antes de mencionar los pasos del algoritmo EM considere la siguiente notación.

Sea  $\mathcal{H} = (\mathcal{W}, \mathcal{Y})$  un conjunto de vectores de datos, con  $\mathbf{h} = (\mathbf{w}, \mathbf{y})$  donde  $\mathbf{h} \in \mathcal{H}$ ,  $\mathbf{w} \in \mathcal{W}$  y  $\mathbf{y} \in \mathcal{Y}$ . Además,  $\mathcal{W}$  es el conjunto de vectores de datos observados y  $\mathcal{Y}$  es un conjunto de vectores de datos no observados por limitaciones en el proceso de observación. A  $\mathcal{W}$  se le conoce como el conjunto incompleto de vectores de datos y a  $\mathcal{H}$  se le conoce como el conjunto completo de vectores de datos. También, sea  $p(\mathbf{h}|\Theta)$  la función de densidad de  $\mathbf{h} \in \mathcal{H}$  con

$$p(\mathbf{h}|\Theta) = p(\mathbf{w}, \mathbf{y}|\Theta) = p(\mathbf{y}|\mathbf{w}, \Theta)p(\mathbf{w}|\Theta). \quad (1.97)$$

donde  $p(\mathbf{w}|\Theta)$  es la función de densidad marginal de los datos incompletos,  $p(\mathbf{y}|\mathbf{w}, \Theta)$  es la relación probabilística que se asume entre las variables observadas y no observadas. Ambas funciones dependen de un conjunto  $\Theta$  de parámetros.

A partir de la función de densidad (1.97), es posible definir una nueva función de verosimilitud, llamada la función de *verosimilitud de los datos completos*

$$\mathcal{L}(\Theta|\mathcal{H}) = \mathcal{L}(\Theta|\mathcal{W}, \mathcal{Y}) = p(\mathcal{W}, \mathcal{Y}|\Theta). \quad (1.98)$$

La función (1.98) es una variable aleatoria ya que la información contenida en  $\mathcal{Y}$  es desconocida, aleatoria, y determinada por alguna distribución que se asume. Por consiguiente, se puede escribir

$$\mathcal{L}(\Theta|\mathcal{W}, \mathcal{Y}) = G_{\mathcal{W}, \Theta}(\mathcal{Y}), \quad (1.99)$$

para alguna función  $G_{\mathcal{W}, \Theta}(\cdot)$  donde  $\mathcal{W}$  y  $\Theta$  son constantes y  $\mathcal{Y}$  es aleatorio.

Una vez definida la función de verosimilitud de datos completos, los *pasos del algoritmo EM* son:

- Paso E del algoritmo EM

El paso E consiste en determinar la función  $Q(\Theta, \Theta^{(i-1)})$ , definida como

$$\begin{aligned} Q(\Theta, \Theta^{(i-1)}) &= E[\log p(\mathcal{W}, \mathcal{Y}|\Theta)|\mathcal{W}, \Theta^{(i-1)}] \\ &= \int_{\mathbf{y} \in \vartheta} \log p(\mathcal{W}, \mathbf{y}|\Theta) f(\mathbf{y}, \mathcal{W}|\Theta^{(i-1)}) d\mathbf{y} \\ &= \int_{\mathbf{y} \in \vartheta} \log p(\mathcal{W}, \mathbf{y}|\Theta) f(\mathbf{y}|\mathcal{W}, \Theta^{(i-1)}) f(\mathcal{W}|\Theta^{(i-1)}) d\mathbf{y} \end{aligned} \quad (1.100)$$

El primer argumento  $\Theta$  corresponde a los parámetros que se ajustan para maximizar la esperanza matemática del logaritmo de la función de verosimilitud en el segundo paso del algoritmo EM. El segundo argumento  $\Theta^{(i-1)}$  corresponde a los parámetros que se emplean para evaluar la esperanza matemática en (1.100).

Cabe mencionar que en la ecuación (1.100),  $\mathcal{W}$  y  $\Theta^{(i-1)}$  son constantes y  $\mathcal{Y}$  es aleatorio. Asimismo, la función  $f(\mathbf{y}|\mathcal{W}, \Theta^{(i-1)})$  es la densidad marginal de los datos no observados y como puede notarse depende de los datos observados  $\mathcal{W}$  y de los parámetros  $\Theta^{(i-1)}$ . La función  $f(\mathcal{W}|\Theta^{(i-1)})$  es la densidad de los datos observados y depende de los parámetros  $\Theta^{(i-1)}$ . Finalmente,  $\vartheta$  es el espacio de los posibles valores que puede tomar  $\mathbf{y} \in \mathcal{Y}$ .

- Paso M del algoritmo EM El segundo paso del algoritmo EM es maximizar la esperanza matemática calculada en el paso E con respecto a los parámetros  $\Theta$ . Para ello se define,

$$\Theta^{(i)} = \arg \max_{\Theta} Q(\Theta, \Theta^{(i-1)}). \quad (1.101)$$

Una vez determinado  $\Theta^{(i)}$ , se repite el paso E del algoritmo EM pero ahora se emplea  $\Theta^{(i)}$  en vez de  $\Theta^{(i-1)}$ . Cada iteración incrementa el valor de la función de log-verosimilitud, además el algoritmo garantiza la convergencia a un máximo local de la función de verosimilitud. Lo último se establece en diversos artículos donde se aborda sobre la convergencia del algoritmo EM, Wu (1984), Xu y Jordan (1996).

De esta manera se ha presentado de forma general el algoritmo EM, los detalles que se requieren en cada paso para realizar los cálculos del algoritmo EM dependen de la aplicación en particular. Enseguida se verá su aplicación para determinar los parámetros de un modelo de Markov oculto.

#### 1.4.4. Estimación de los parámetros de un HMM bajo la distribución Gaussiana

En este apartado se obtienen los estimadores de los parámetros de un modelo de Markov oculto con espacio de observaciones continuo y se asume la distribución gaussiana multivariable como la distribución de las observaciones del modelo. Para ello se emplea el algoritmo EM con el fin de encontrar los estimadores de máxima verosimilitud de los parámetros del modelo. El algoritmo EM aplicado a un modelo de Markov oculto es conocido como algoritmo de Baum-Welch.

El problema que se plantea en términos del algoritmo EM aplicado a un modelo de Markov oculto es: Se tiene un conjunto de datos conocido que es la secuencia de observaciones del modelo  $\mathcal{O} = (z_1, z_2, z_3, \dots, z_T)$ ; y un conjunto de datos desconocido, aleatorio, que es la secuencia de estados de la cadena de Markov  $\mathcal{I} = (x_1, x_2, \dots, x_T)$ ; y, se busca estimar un conjunto de parámetros  $\lambda_c = (A, \mathcal{B}, \pi)$ . Para la estimación de los parámetros se supone una relación probabilística entre los datos conocidos y desconocidos.

Antes de establecer el algoritmo EM aplicado al modelo de Markov oculto se dan algunas definiciones necesarias.

**Definición 1.22.** Sea  $Y_t = \begin{pmatrix} Y_{t,1} \\ Y_{t,2} \\ \vdots \\ Y_{t,d} \end{pmatrix}$  un vector aleatorio de dimensión  $d$ , se dice que  $Y_t$  tiene una *distribución gaussiana multivariada* si su función de densidad, denotada como  $f_{y_t}(\mu, \Sigma)$  está dada por

$$f_{y_t}(\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(y_t - \mu)^\top \Sigma^{-1} (y_t - \mu)} \quad (1.102)$$

donde  $\mu$  es la media del vector aleatorio  $Y_t$  por lo que es un vector de dimensión  $d$ ,  $\Sigma$  es la matriz de varianza-covarianza de la variable aleatoria  $Y_t$  por lo que es una matriz de dimensión  $d$  por  $d$ . Asimismo,  $|\Sigma|$  denota el determinante de la matriz de varianza y  $\top$  denota la traspuesta de la matriz.

**Definición 1.23.** Sea  $\lambda_c = (A, \mathcal{B}, \pi)$  un modelo de Markov oculto con espacio de observaciones continuo y sea  $\mathcal{O} = (z_1, z_2, z_3, \dots, z_T)$  una secuencia de observaciones conocida, asociado al modelo de

Markov existe una secuencia de estados desconocida  $\mathcal{I} = (x_1, x_2, \dots, x_T)$ ; de la ecuación (1.80) la función de densidad conjunta de  $(z_1, z_2, z_3, \dots, z_T, x_1, x_2, \dots, x_T)$  está dada por

$$p(z_1, z_2, z_3, \dots, z_T, x_1, x_2, \dots, x_T | \lambda_c^i) = \pi_{x_1} b(z_1 | x_1) \prod_{t=1}^{T-1} a_{x_t, x_{t+1}} b(z_{t+1} | x_{t+1}). \quad (1.103)$$

donde  $Z_t$  es un vector aleatorio de dimensión  $d$  y  $b(z_{t+1} | x_{t+1})$  es la función de densidad gaussiana multivariable de dimensión  $d$ .

**Definición 1.24.** Dado un modelo de Markov oculto con espacio de observaciones continuo  $\{X_t, Z_t\}$ ,  $t > 0$  y parámetros  $\lambda_c = (A, \mathcal{B}, \pi)$  se dice que se asume la *distribución gaussiana multivariable para la variable de observaciones*  $Z_t$  si para cada posible estado de la cadena de Markov asociada al modelo, la función de densidad condicional de la variable  $Z_t$  dado que  $X_t = q_i$  para  $i = 1, 2, \dots, N$  está dada por

$$b(z_t | q_i) = f_{z_t}(\mu_i, \Sigma_i) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma_i|^{\frac{1}{2}}} e^{-\frac{1}{2}(z_t - \mu_i)^\top \Sigma_i^{-1} (z_t - \mu_i)}. \quad (1.104)$$

**Definición 1.25.** Considere un modelo de Markov oculto con espacio de observaciones continuo  $\lambda_c = (A, \mathcal{B}, \pi)$ , una secuencia de observaciones  $\mathcal{O} = (z_1, z_2, z_3, \dots, z_T)$  conocida y  $\mathcal{I} = (x_1, x_2, \dots, x_T)$  denota una posible secuencia de estados de la cadena de Markov asociada al modelo. La *función de Baum Welch* denotada como  $Q(\lambda_c, \lambda_c^i)$  se define así

$$Q(\lambda_c, \lambda_c^i) = E[\ln p(\mathcal{O}, \mathcal{I} | \lambda_c) | \mathcal{O}, \lambda_c^i] \quad (1.105)$$

donde  $\lambda_c^i$  son un estimador de los parámetros del modelo de Markov oculto y son conocidos. Cabe mencionar que la esperanza matemática condicional se toma con respecto a la variable  $I$  y que se encuentra condicionada a  $\mathcal{O}$  y a  $\lambda_c^i$  ya que estos son conocidos y deterministas.

Con la notación y definición establecidas, se procede a calcular los dos pasos del algoritmo EM.

### Paso E del algoritmo EM

El primer paso (paso E) del algoritmo EM es construir la función de Baum Welch como sigue

$$\begin{aligned} Q(\lambda_c, \lambda_c^i) &= E[\ln p(\mathcal{O}, \mathcal{I} | \lambda_c) | \mathcal{O}, \lambda_c^i] \\ &= \sum_{\mathcal{I}} \ln (p(\mathcal{O}, \mathcal{I} | \lambda_c)) p(\mathcal{O}, \mathcal{I} | \lambda_c^i) \\ &= \sum_{\mathcal{I}} \ln (\pi_{x_1} b(z_1 | x_1) \prod_{t=1}^{T-1} a_{x_t, x_{t+1}} b(z_{t+1} | x_{t+1})) p(\mathcal{O}, \mathcal{I} | \lambda_c^i) \\ &= \sum_{\mathcal{I}} \ln(\pi_{x_1}) p(\mathcal{O}, \mathcal{I} | \lambda_c^i) + \sum_{\mathcal{I} \in \mathcal{J}} \left( \sum_{t=1}^T \ln b(z_t | x_t) \right) p(\mathcal{O}, \mathcal{I} | \lambda_c^i) + \sum_{\mathcal{I} \in \mathcal{J}} \left( \sum_{t=1}^{T-1} \ln a_{x_t, x_{t+1}} \right) p(\mathcal{O}, \mathcal{I} | \lambda_c^i) \\ &= \sum_{i=1}^N \ln \pi_{q_i} p(\mathcal{O}, X_1 = q_i | \lambda_c^i) + \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^{T-1} \ln a_{q_i, q_j} p(\mathcal{O}, X_t = q_i, X_{t+1} = q_j | \lambda_c^i) \\ &\quad + \sum_{i=1}^N \sum_{t=1}^T \left( -\frac{d}{2} \ln 2\pi - \frac{1}{2} \ln |\Sigma_i| - \frac{1}{2} (z_t - \mu_i)^\top \Sigma_i^{-1} (z_t - \mu_i) \right) p(\mathcal{O}, X_t = q_i | \lambda_c^i) \quad (1.106) \end{aligned}$$

Evidentemente se ha asumido que las observaciones del modelo de Markov oculto tienen una distribución gaussiana multivariable de dimensión  $d$ . Además se ha asumido que  $\mathcal{O}$  y  $\lambda_c^i$  son conocidos.

### Paso M del algoritmo EM

Una vez determinada la función  $Q(\lambda_c, \lambda_c^i)$ , se procede a realizar el paso M del algoritmo EM, por lo que se debe resolver el problema de optimización,

$$\lambda_c^{i+1} = \arg \max_{\lambda_c} Q(\lambda_c, \lambda_c^i) \quad (1.107)$$

Específicamente, se debe resolver

$$\lambda_c^{i+1} = \arg \max_{\pi_{q_i}, a_{q_i, q_j}, \Sigma_i, \mu_i, \text{ para } i, j=1, 2, \dots, N} Q(\lambda_c, \lambda_c^i) \quad (1.108)$$

Sujeto a

$$\sum_{i=1}^N \pi_{q_i} = 1$$

$$\sum_{j=1}^N a_{q_i, q_j} = 1 \text{ para } i = 1, 2, \dots, N$$

Al tratarse de un problema de maximización con restricciones de igualdad es necesario emplear el método de multiplicadores de Lagrange para resolverlo. El lagrangiano correspondiente al problema (1.108) será denotado como  $L(\pi_{q_i}, a_{q_i, q_j}, \Sigma_i, \mu_i, \eta_1, \eta_{2,1}, \eta_{2,2}, \dots, \eta_{2,N})$  y está dado por

$$L(\pi_{q_i}, a_{q_i, q_j}, \Sigma_i, \mu_i, \eta_1, \eta_{2,1}, \eta_{2,2}, \dots, \eta_{2,N}) = Q(\lambda_c, \lambda_c^i) + \eta_1 \left( \sum_{i=1}^N \pi_{q_i} - 1 \right) + \sum_{i=1}^N \eta_{2,i} \left( \sum_{j=1}^N a_{q_i, q_j} - 1 \right) \quad (1.109)$$

donde  $\eta_1, \eta_{2,1}, \eta_{2,2}, \dots, \eta_{2,N}$  son los multiplicadores de Lagrange.

Las condiciones de primer orden son

$$\frac{\partial L}{\partial \pi_{q_i}} = \frac{p(\mathcal{O}, X_1 = q_i | \lambda_c^i)}{\pi_{q_i}} + \eta_1 = 0 \quad \text{para } i = 1, 2, \dots, N \quad (1.110)$$

$$\frac{\partial L}{\partial \eta_1} = \sum_{i=1}^N \pi_{q_i} - 1 = 0 \quad (1.111)$$

$$\frac{\partial L}{\partial a_{q_i, q_j}} = \frac{\sum_{t=1}^{T-1} p(\mathcal{O}, X_t = q_i, X_{t+1} = q_j | \lambda_c^i)}{a_{q_i, q_j}} + \eta_{2,i} = 0 \quad \text{para } i, j = 1, 2, \dots, N \quad (1.112)$$

$$\frac{\partial L}{\partial \eta_{2,i}} = \sum_{j=1}^N a_{q_i, q_j} - 1 = 0 \quad \text{para } i = 1, 2, \dots, N \quad (1.113)$$

$$\frac{\partial L}{\partial \mu_i} = \sum_{t=1}^T (z_t - \mu_i) p(\mathcal{O}, X_t = q_i | \lambda_c^i) = 0 \quad \text{para } i = 1, 2, \dots, N \quad (1.114)$$

$$\frac{\partial L}{\partial \Sigma_i^{-1}} = \sum_{t=1}^T (\Sigma_i - (z_t - \mu_i)(z_t - \mu_i)^\top) p(\mathcal{O}, X_t = q_i | \lambda_c^i) = 0 \quad \text{para } i = 1, 2, \dots, N \quad (1.115)$$

De las ecuaciones (1.110) y (1.111) se obtiene

$$\pi_{q_i} = \frac{p(\mathcal{O}, X_1 = q_i | \lambda_c^i)}{p(\mathcal{O} | \lambda_c^i)}, \quad \text{para } i = 1, 2, \dots, N. \quad (1.116)$$

De las ecuaciones (1.112) y (1.113) se tiene

$$a_{q_i, q_j} = \frac{\sum_{t=1}^{T-1} p(\mathcal{O}, X_t = q_i, X_{t+1} = q_j | \lambda_c^i)}{\sum_{t=1}^{T-1} p(\mathcal{O}, X_t = q_i | \lambda_c^i)} \quad \text{para } i, j = 1, 2, \dots, N. \quad (1.117)$$

De la ecuación (1.114) se sigue que,

$$\mu_i = \frac{\sum_{t=1}^T p(\mathcal{O}, X_t = q_i | \lambda_c^i) z_t}{\sum_{t=1}^T p(\mathcal{O}, X_t = q_i | \lambda_c^i)}, \quad \text{para } i = 1, 2, \dots, N. \quad (1.118)$$

Finalmente, de la ecuación (1.115) se obtiene,

$$\Sigma_i = \frac{\sum_{t=1}^T p(\mathcal{O}, X_t = q_i | \lambda_c^i) (z_t - \mu_i)(z_t - \mu_i)^\top}{\sum_{t=1}^T p(\mathcal{O}, X_t = q_i | \lambda_c^i)} \quad \text{para } i = 1, 2, \dots, N. \quad (1.119)$$

Considerando las definiciones 1.20 y 1.21, se sigue que los estimadores de los parámetros de un modelo de Markov oculto con espacio de observaciones continuo bajo el supuesto que la distribución de la variable de observaciones es la distribución gaussiana multivariable, son:

$$\pi_{q_i} = \gamma_1(q_i)_c, \quad \text{para } i = 1, 2, \dots, N. \quad (1.120)$$

$$a_{q_i, q_j} = \frac{\sum_{t=1}^{T-1} \xi_t(q_i, q_j)_c}{\sum_{t=1}^{T-1} \gamma_t(q_i)_c} \quad \text{para } i, j = 1, 2, \dots, N. \quad (1.121)$$

$$\mu_i = \frac{\sum_{t=1}^T \gamma_t(q_i)_c z_t}{\sum_{t=1}^T \gamma_t(q_i)_c}, \quad \text{para } i = 1, 2, \dots, N. \quad (1.122)$$

$$\Sigma_i = \frac{\sum_{t=1}^T \gamma_t(q_i)_c (z_t - \mu_i)(z_t - \mu_i)^\top}{\sum_{t=1}^T \gamma_t(q_i)_c} \quad \text{para } i = 1, 2, \dots, N. \quad (1.123)$$

Si bien, las ecuaciones obtenidas corresponden a un modelo de Markov oculto con espacio de observaciones continuo donde cada posible observación es un vector de dimensión  $d$ , en la aplicación que se realiza del modelo en el siguiente capítulo, la observación es un escalar por lo que las formulas empleadas son un caso simplificado del modelo presentado que se obtienen de hacer  $d = 1$ .

### 1.4.5. Estimación de los parámetros de un HMM bajo una mezcla de Gaussianas

En este apartado se obtienen los estimadores de un modelo de Markov oculto con espacio de observaciones continuo, donde se asume una mezcla de gaussianas multivariable como la distribución de las observaciones del modelo. Para ello se emplea el algoritmo EM con el fin de encontrar los estimadores de máxima verosimilitud de los parámetros del modelo.

El problema que se plantea en términos del algoritmo EM aplicado a un modelo de Markov oculto es: Se tiene un conjunto de datos conocido que es la secuencia de observaciones del modelo  $\mathcal{O} = (z_1, z_2, z_3, \dots, z_T)$ ; y dos conjuntos de datos desconocidos; por un lado, la secuencia de estados de la cadena de Markov  $\mathcal{I} = (x_1, x_2, \dots, x_T)$  y por otro lado, una secuencia de componentes de las mezclas de gaussianas  $\mathbf{M} = \{m_{x_1}, m_{x_2}, \dots, m_{x_t}, \dots, m_{x_T}\}$  que generaron el conjunto de observaciones y, se busca estimar un conjunto de parámetros  $\lambda_c = (A, \mathcal{B}, \pi)$  suponiendo una relación probabilística entre los datos conocidos y desconocidos.

Antes de proceder con la aplicación del algoritmo EM se consideran algunas definiciones necesarias.

**Definición 1.26.** Sea  $Y_t$  un vector aleatorio de dimensión  $d$ , se dice que  $Y_t$  tiene como distribución la *mezcla de  $R$  gaussianas multivariadas* si su función de densidad, denotada como  $f_{y_t}(\mu_1, \Sigma_1, \dots, \mu_R, \Sigma_R)$  está dada por

$$f_{y_t}(\mu_1, \Sigma_1, \dots, \mu_R, \Sigma_R) = \sum_{k=1}^R C_k \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma_k|^{\frac{1}{2}}} e^{-\frac{1}{2}(y_t - \mu_k)^\top \Sigma_k^{-1} (y_t - \mu_k)} \quad (1.124)$$

donde  $\mu_k$  y  $\Sigma_k$  son la media y matriz de varianza-covarianza de algún vector aleatorio; además se cumple que  $\sum_{k=1}^R C_k = 1$  y  $C_k$  es el ponderador o peso de la  $k$ -ésima función de densidad en la mezcla que se realiza.

**Definición 1.27.** Dado un modelo de Markov oculto con espacio de observaciones continuo  $\{X_t, Z_t\}$ ,  $t > 0$  y parámetros  $\lambda_c = (A, \mathcal{B}, \pi)$  se dice que se asume una *mezcla de  $R$  gaussianas multivariadas de dimensión  $d$  como la distribución para la variable de observaciones  $Z_t$*  si para cada posible estado de la cadena de Markov asociada al modelo, la función de densidad condicional de la variable  $Z_t$  dado que  $X_t = q_i$  para  $i = 1, 2, \dots, N$  está dada por

$$b(z_t | q_i) = f_{z_t}(\mu_{i,1}, \Sigma_{i,1}, \dots, \mu_{i,R}, \Sigma_{i,R}) = \sum_{k=1}^R C_{i,k} \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma_{i,k}|^{\frac{1}{2}}} e^{-\frac{1}{2}(z_t - \mu_{i,k})^\top \Sigma_{i,k}^{-1} (z_t - \mu_{i,k})}. \quad (1.125)$$

**Definición 1.28.** Dado un modelo de Markov oculto con espacio de observaciones continuo  $\{X_t, Z_t\}$ ,  $t > 0$  y parámetros  $\lambda_c = (A, \mathcal{B}, \pi)$ , que asume una mezcla de  $R$  gaussianas multivariadas de dimensión  $d$  como la distribución de las observaciones del modelo. Se define  $M_{X_t}$  como una variable aleatoria discreta cuyo soporte es el conjunto de valores  $\{1, 2, 3, \dots, R\}$  y cuya distribución de probabilidad depende de la variable  $X_t$ . Tal variable sirve para referirse a los términos en una mezcla de gaussianas.

En otras palabras, la variable  $M_{q_i} = k$ , con  $k$  en el conjunto soporte de la variable, sirve para referirse al  $k$ -ésimo sumando que forma la mezcla de  $R$  gaussianas multivariadas en la ecuación (1.125).

**Definición 1.29.** Sea  $\lambda_c = (A, \mathcal{B}, \pi)$  un modelo de Markov oculto con espacio de observaciones continuo y sean  $\mathcal{O} = (z_1, z_2, z_3, \dots, z_T)$  una secuencia de observaciones conocida,  $\mathcal{I} = (x_1, x_2, \dots, x_T)$  la secuencia de estados desconocida asociado al modelo de Markov y  $\mathbf{M} = \{m_{x_1}, m_{x_2}, \dots, m_{x_T}\}$  una posible secuencia de los componentes de las mezclas de gaussianas. La función de densidad conjunta de  $(z_1, z_2, z_3, \dots, z_T, x_1, x_2, \dots, x_T, m_{x_1}, m_{x_2}, \dots, m_{x_T})$  está dada por

$$p(z_1, z_2, \dots, z_T, x_1, x_2, \dots, x_T, m_{x_1}, m_{x_2}, \dots, m_{x_T} | \lambda_c^i) = \pi_{x_1} b(z_1 | x_1, m_{x_1}) \prod_{t=1}^{T-1} a_{x_t, x_{t+1}} b(z_{t+1} | x_{t+1}, m_{x_{t+1}}) p(m_{x_{t+1}} | x_{t+1}) \quad (1.126)$$

donde  $Z_t$  es un vector aleatorio de dimensión  $d$  y  $b(z_t | x_t, m_{x_t}) = p(z_t | x_t, m_{x_t})$  es la función de densidad de  $z_t$  condicional a  $x_t$  y a  $m_{x_t}$ .

Con la notación y definición establecidas, se procede a calcular los dos pasos del algoritmo EM.

### Paso E del algoritmo EM

El primer paso (paso E) del algoritmo EM es construir la función de Baum Welch

$$\begin{aligned}
Q(\lambda_c, \lambda_c^i) &= E[\ln p(\mathcal{O}, \mathcal{I}, \mathbf{M}|\lambda_c)|\mathcal{O}, \lambda_c^i] \\
&= \sum_{\mathcal{I}} \sum_{\mathbf{M}} \ln (p(\mathcal{O}, \mathcal{I}, \mathbf{M}|\lambda_c)) p(\mathcal{O}, \mathcal{I}, \mathbf{M}|\lambda_c^i) \\
&= \sum_{\mathcal{I}} \sum_{\mathbf{M}} \ln (\pi_{x_1} b(z_1|x_1, m_{x_1}) p(m_{x_1}|x_1) \prod_{t=1}^{T-1} a_{x_t, x_{t+1}} b(z_{t+1}|x_{t+1}, m_{x_{t+1}}) p(m_{x_{t+1}}|x_{t+1})) p(\mathcal{O}, \mathcal{I}, \mathbf{M}|\lambda_c^i) \\
&= \sum_{\mathcal{I}} \sum_{\mathbf{M}} \ln (\pi_{x_1}) p(\mathcal{O}, \mathcal{I}, \mathbf{M}|\lambda_c^i) + \sum_{\mathcal{I}} \sum_{\mathbf{M}} \sum_{t=1}^T \ln (b(z_t|x_t, m_{x_t}) p(m_{x_t}|x_t)) p(\mathcal{O}, \mathcal{I}, \mathbf{M}|\lambda_c^i) \\
&\quad + \sum_{\mathcal{I}} \sum_{\mathbf{M}} \sum_{t=1}^{T-1} \ln a_{x_t, x_{t+1}} p(\mathcal{O}, \mathcal{I}, \mathbf{M}|\lambda_c^i) \\
&= \sum_{i=1}^N \ln \pi_{q_i} p(\mathcal{O}, X_1 = q_i|\lambda_c^i) + \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^{T-1} \ln a_{q_i, q_j} p(\mathcal{O}, X_t = q_i, X_{t+1} = q_j|\lambda_c^i) \\
&\quad + \sum_{i=1}^N \sum_{k=1}^R \sum_{t=1}^T \ln (p(m_{x_t} = k|x_t = q_i) b(z_t|x_t = q_i, m_{x_t} = k)) p(\mathcal{O}, X_t = q_i, M_{X_t} = k|\lambda_c^i) \\
&= \sum_{i=1}^N \ln \pi_{q_i} p(\mathcal{O}, X_1 = q_i|\lambda_c^i) + \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^{T-1} \ln a_{q_i, q_j} p(\mathcal{O}, X_t = q_i, X_{t+1} = q_j|\lambda_c^i) \\
&\quad + \sum_{i=1}^N \sum_{k=1}^R \sum_{t=1}^T \ln C_{i,k} f_{z_t}(\mu_{i,k}, \Sigma_{i,k}) p(\mathcal{O}, X_t = q_i, M_{X_t} = k|\lambda_c^i) \\
&= \sum_{i=1}^N \ln \pi_{q_i} p(\mathcal{O}, X_1 = q_i|\lambda_c^i) + \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^{T-1} \ln a_{q_i, q_j} p(\mathcal{O}, X_t = q_i, X_{t+1} = q_j|\lambda_c^i) \\
&\quad + \sum_{i=1}^N \sum_{k=1}^R \sum_{t=1}^T \left( -\frac{d}{2} \ln 2\pi - \frac{1}{2} \ln |\Sigma_{i,k}| - \frac{1}{2} (z_t - \mu_{i,k})^T \Sigma_{i,k}^{-1} (z_t - \mu_{i,k}) \right) p(\mathcal{O}, X_t = q_i, M_{X_t} = k|\lambda_c^i) \\
&\quad + \sum_{i=1}^N \sum_{k=1}^R \sum_{t=1}^T \ln C_{i,k} p(\mathcal{O}, X_t = q_i, M_{X_t} = k|\lambda_c^i) \tag{1.127}
\end{aligned}$$

En esta ecuación  $\lambda_c^i$  es un conjunto de parámetros del modelo de Markov oculto y son conocidos y por otra parte  $\lambda_c$  son los nuevos parámetros del modelo que se pretenden estimar en el segundo paso del algoritmo EM.

### Paso M del algoritmo EM

Una vez determinada la función de Baum-Welch, el paso M del algoritmo EM consiste en resolver el problema de optimización

$$\lambda_c^{i+1} = \arg \max_{\pi_{q_i}, a_{q_i, q_j}, \Sigma_{i,k}, \mu_{i,k}, C_{i,k}} Q(\lambda_c, \lambda_c^i) \quad (1.128)$$

para  $i, j=1, 2, \dots, N$  y  $k=1, 2, \dots, R$

Sujeto a

$$\sum_{i=1}^N \pi_{q_i} = 1$$

$$\sum_{j=1}^N a_{q_i, q_j} = 1 \text{ para } i = 1, 2, \dots, N$$

$$\sum_{k=1}^R C_{i,k} = 1 \text{ para } i = 1, 2, \dots, N$$

Se trata de un problema de maximización con restricciones de igualdad, para su solución se emplea el método de multiplicadores de Lagrange. El lagrangiano correspondiente al problema (1.128) será denotado como  $L(\pi_{q_i}, a_{q_i, q_j}, \Sigma_{i,k}, \mu_{i,k}, C_{i,k}, \eta_1, \eta_{2,1}, \dots, \eta_{2,N}, \eta_{3,1}, \dots, \eta_{3,N})$  y está dado por

$$\begin{aligned} L(\pi_{q_i}, a_{q_i, q_j}, \Sigma_{i,k}, \mu_{i,k}, \eta_1, \eta_{2,1}, \dots, \eta_{2,N}, \eta_{3,1}, \dots, \eta_{3,N}) = & Q(\lambda_c, \lambda_c^i) + \eta_1 \left( \sum_{i=1}^N \pi_{q_i} - 1 \right) \\ & + \sum_{i=1}^N \eta_{2,i} \left( \sum_{j=1}^N a_{q_i, q_j} - 1 \right) \\ & + \sum_{i=1}^N \eta_{3,i} \left( \sum_{k=1}^R C_{i,k} - 1 \right) \end{aligned} \quad (1.129)$$

donde  $\eta_1, \eta_{2,1}, \eta_{2,2}, \dots, \eta_{2,N}, \eta_{3,1}, \eta_{3,2}, \dots, \eta_{3,N}$  son los multiplicadores de Lagrange.

Las condiciones de primer orden del problema (1.129) son

$$\frac{\partial L}{\partial \pi_{q_i}} = \frac{p(\mathcal{O}, X_1 = q_i | \lambda_c^i)}{\pi_{q_i}} + \eta_1 = 0 \quad \text{para } i = 1, 2, \dots, N \quad (1.130)$$

$$\frac{\partial L}{\partial \eta_1} = \sum_{i=1}^N \pi_{q_i} - 1 = 0 \quad (1.131)$$

$$\frac{\partial L}{\partial a_{q_i, q_j}} = \frac{\sum_{t=1}^{T-1} p(\mathcal{O}, X_t = q_i, X_{t+1} = q_j | \lambda_c^i)}{a_{q_i, q_j}} + \eta_{2,i} = 0 \quad \text{para } i, j = 1, 2, \dots, N \quad (1.132)$$

$$\frac{\partial L}{\partial \eta_{2,i}} = \sum_{j=1}^N a_{q_i, q_j} - 1 = 0 \quad \text{para } i = 1, 2, \dots, N \quad (1.133)$$

$$\frac{\partial L}{\partial C_{i,k}} = \frac{\sum_{t=1}^T p(\mathcal{O}, X_t = q_i, M_{X_t} = k | \lambda_c^i)}{C_{i,k}} - \eta_{3,i} = 0 \quad (1.134)$$

para  $i = 1, \dots, N$  y  $k = 1, \dots, R$

$$\frac{\partial L}{\partial \eta_{3,i}} = \sum_{k=1}^R C_{i,k} - 1 = 0 \quad \text{para } i = 1, 2, \dots, N \quad (1.135)$$

$$\frac{\partial L}{\partial \mu_{i,k}} = \sum_{t=1}^T (z_t - \mu_{i,k}) p(\mathcal{O}, X_t = q_i, M_{X_t} = k | \lambda_c^i) = 0 \quad (1.136)$$

para  $i = 1, 2, \dots, N$  y  $k = 1, 2, \dots, R$

$$\frac{\partial L}{\partial \Sigma_{i,k}^{-1}} = \sum_{t=1}^T (\Sigma_{i,k} - (z_t - \mu_{i,k})(z_t - \mu_{i,k})^T) p(\mathcal{O}, X_t = q_i, M_{X_t} = k | \lambda_c^i) = 0 \quad (1.137)$$

para  $i = 1, 2, \dots, N$  y  $k = 1, 2, \dots, R$ .

De las ecuaciones (1.130) y (1.131) se obtiene

$$\pi_{q_i} = \frac{p(\mathcal{O}, X_t = q_i | \lambda_c^i)}{p(\mathcal{O} | \lambda_c^i)}, \quad \text{para } i = 1, 2, \dots, N. \quad (1.138)$$

De las ecuaciones (1.132) y (1.133) se tiene

$$a_{q_i, q_j} = \frac{\sum_{t=1}^{T-1} p(\mathcal{O}, X_t = q_i, X_{t+1} = q_j | \lambda_c^i)}{\sum_{t=1}^{T-1} p(\mathcal{O}, X_t = q_i | \lambda_c^i)} \quad \text{para } i, j = 1, 2, \dots, N. \quad (1.139)$$

De las ecuaciones (1.134) y (1.135) se tiene

$$C_{i,k} = \frac{\sum_{t=1}^T p(\mathcal{O}, X_t = q_i, M_{X_t} = k | \lambda_c^i)}{\sum_{t=1}^T \sum_{k=1}^R p(\mathcal{O}, X_t = q_i, M_{X_t} = k | \lambda_c^i)} \quad \text{para } i = 1, 2, \dots, N \text{ y } k = 1, 2, \dots, R. \quad (1.140)$$

De la ecuación (1.136) se sigue que,

$$\mu_{i,k} = \frac{\sum_{t=1}^T z_t p(\mathcal{O}, X_t = q_i, M_{X_t} = k | \lambda_c^i)}{\sum_{t=1}^T p(\mathcal{O}, X_t = q_i, M_{X_t} = k | \lambda_c^i)}, \quad \text{para } i = 1, 2, \dots, N \text{ y } k = 1, 2, \dots, R. \quad (1.141)$$

Finalmente, de la ecuación (1.137) se obtiene,

$$\Sigma_{i,k} = \frac{\sum_{t=1}^T p(\mathcal{O}, X_t = q_i, M_{X_t} = k | \lambda_c^i) (z_t - \mu_{i,k})(z_t - \mu_{i,k})^T}{\sum_{t=1}^T p(\mathcal{O}, X_t = q_i, M_{X_t} = k | \lambda_c^i)} \quad \text{para } i = 1, 2, \dots, N \text{ y } k = 1, 2, \dots, R. \quad (1.142)$$

Los estimadores obtenidos se pueden expresar de una forma más sencilla a partir de la siguiente definición.

**Definición 1.30.** Dado un modelo de Markov oculto con espacio de observacion continuo  $\lambda_c = (A, \mathcal{B}, \pi)$  y una secuencia de observaciones  $\mathcal{O} = \{z_1, z_2, z_3, \dots, z_T\}$ ; la función de densidad conjunta de la variable  $X_t = q_i$  y  $M_{X_t} = k$  dada la secuencia  $\mathcal{O}$  y el modelo  $\lambda_c$  se denota como  $\gamma_t(q_i, k)_c$  y es igual a

$$\begin{aligned} \gamma_t(q_i, k)_c &= p(X_t = q_i, M_{X_t} = k | \mathcal{O}, \lambda_c) \\ &= \gamma_t(q_i)_c \frac{C_{i,k} f_{z_t}(\mu_{i,k}, \Sigma_{i,k})}{\sum_{k=1}^R C_{i,k} f_{z_t}(\mu_{i,k}, \Sigma_{i,k})} \end{aligned} \quad (1.143)$$

para  $i = 1, 2, 3, \dots, N$ ,  $k = 1, 2, \dots, R$  y  $t = 1, 2, 3, \dots, T$ . ■

Retomando las definiciones 1.20, 1.21 y 1.30, los estimadores pueden escribirse como sigue,

$$\pi_{q_i} = \gamma_1(q_i)_c, \quad \text{para } i = 1, 2, \dots, N. \quad (1.144)$$

$$a_{q_i, q_j} = \frac{\sum_{t=1}^{T-1} \xi_t(q_i, q_j)_c}{\sum_{t=1}^{T-1} \gamma_t(q_i)_c} \quad \text{para } i, j = 1, 2, \dots, N. \quad (1.145)$$

$$C_{i,k} = \frac{\sum_{t=1}^T \gamma_t(q_i, k)_c}{\sum_{t=1}^T \gamma_t(q_i)_c} \quad \text{para } i = 1, 2, \dots, N \text{ y } k = 1, 2, \dots, R. \quad (1.146)$$

$$\mu_{i,k} = \frac{\sum_{t=1}^T \gamma_t(q_i, k)_c z_t}{\sum_{t=1}^T \gamma_t(q_i, k)_c}, \quad \text{para } i = 1, 2, \dots, N \text{ y } k = 1, 2, \dots, R. \quad (1.147)$$

$$\Sigma_{i,k} = \frac{\sum_{t=1}^T \gamma_t(q_i, k)_c (z_t - \mu_{i,k})(z_t - \mu_{i,k})^\top}{\sum_{t=1}^T \gamma_t(q_i, k)_c} \quad \text{para } i = 1, 2, \dots, N \text{ y } k = 1, 2, \dots, R. \quad (1.148)$$

Las ecuaciones obtenidas corresponden a un modelo de Markov oculto con espacio de observaciones continuo donde cada posible observación es un vector de dimensión  $d$ .

### 1.4.6. Estimación de los parámetros de un HMM bajo la distribución Gamma

En este apartado se obtienen los estimadores de los parámetros de un modelo de Markov oculto con espacio de observaciones continuo cuando se asume la distribución gamma como la distribución de las observaciones del modelo. Se emplea el algoritmo EM para encontrar los parámetros del modelo. Antes de establecer el algoritmo EM aplicado al modelo de Markov oculto se dan algunas definiciones.

**Definición 1.31.** Sea  $Y_t$  una variable aleatoria, se dice que  $Y_t$  tiene una *distribución gamma* si su función de densidad, denotada como  $f_{y_t}(\theta, \kappa)$  está dada por

$$f_{y_t}(\theta, \kappa) = \frac{y_t^{\kappa-1}}{\theta^\kappa \Gamma(\kappa)} e^{-\frac{y_t}{\theta}} \quad (1.149)$$

siempre que  $y_t > 0$ , y  $f_{y_t}(\theta, \kappa) = 0$  en otro caso, donde  $\theta$  es el parámetro de escala porque donde más influye es en la dispersión de la distribución y  $\kappa$  es el parámetro de forma porque es el que determina el grado de asimetría y apuntamiento de la densidad.

**Definición 1.32.** Dado un modelo de Markov oculto con espacio de observaciones continuo  $\{X_t, Z_t\}$ ,  $t > 0$  y parámetros  $\lambda_c = (A, \mathcal{B}, \pi)$  se dice que se asume la *distribución gamma para la variable de observaciones*  $Z_t$  si para cada posible estado de la cadena de Markov asociada al modelo, la función de densidad condicional de la variable  $Z_t$  dado que  $X_t = q_i$ , para  $i = 1, 2, \dots, N$ , está dada por

$$b(z_t|q_i) = f_{z_t}(\theta_i, \kappa_i) = \frac{z_t^{\kappa_i-1}}{\theta_i^{\kappa_i} \Gamma(\kappa_i)} e^{-\frac{z_t}{\theta_i}}. \quad (1.150)$$

Con la notación y definición establecidas, se procede a calcular los dos pasos del algoritmo EM.

### Paso E del algoritmo EM

El primer paso (paso E) del algoritmo EM es construir la función de Baum Welch como sigue

$$\begin{aligned} Q(\lambda_c, \lambda_c^i) &= E[\ln p(\mathcal{O}, \mathcal{I}|\lambda_c)|\mathcal{O}, \lambda_c^i] \\ &= \sum_{\mathcal{I}} \ln (p(\mathcal{O}, \mathcal{I}|\lambda_c)) p(\mathcal{O}, \mathcal{I}|\lambda_c^i) \\ &= \sum_{\mathcal{I}} \ln (\pi_{x_1} b(z_1|x_1) \prod_{t=1}^{T-1} a_{x_t, x_{t+1}} b(z_{t+1}|x_{t+1})) p(\mathcal{O}, \mathcal{I}|\lambda_c^i) \\ &= \sum_{\mathcal{I}} \ln(\pi_{x_1}) p(\mathcal{O}, \mathcal{I}|\lambda_c^i) + \sum_{\mathcal{I} \in \mathcal{J}} \left( \sum_{t=1}^T \ln b(z_t|x_t) \right) p(\mathcal{O}, \mathcal{I}|\lambda_c^i) + \sum_{\mathcal{I} \in \mathcal{J}} \left( \sum_{t=1}^{T-1} \ln a_{x_t, x_{t+1}} \right) p(\mathcal{O}, \mathcal{I}|\lambda_c^i) \\ &= \sum_{i=1}^N \ln \pi_{q_i} p(\mathcal{O}, X_1 = q_i|\lambda_c^i) + \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^{T-1} \ln a_{q_i, q_j} p(\mathcal{O}, X_t = q_i, X_{t+1} = q_j|\lambda_c^i) \\ &\quad + \sum_{i=1}^N \sum_{t=1}^T \left( (\kappa_i - 1) \ln z_t - \kappa_i \ln \theta_i - \ln \Gamma(\kappa_i) - \frac{z_t}{\theta_i} \right) p(\mathcal{O}, X_t = q_i|\lambda_c^i) \end{aligned} \quad (1.151)$$

### Paso M del algoritmo EM

Una vez determinada la función  $Q(\lambda_c, \lambda_c^i)$ , se procede a realizar el paso M del algoritmo EM, por lo que se debe resolver el problema

$$\lambda_c^{i+1} = \arg \max_{\pi_{q_i}, a_{q_i, q_j}, \kappa_i, \theta_i \text{ para } i, j=1, 2, \dots, N} Q(\lambda_c, \lambda_c^i) \quad (1.152)$$

Sujeto a

$$\begin{aligned} \sum_{i=1}^N \pi_{q_i} &= 1 \\ \sum_{j=1}^N a_{q_i, q_j} &= 1 \text{ para } i = 1, 2, \dots, N \end{aligned}$$

El lagrangiano correspondiente al problema (1.152) será denotado como  $L(\pi_{q_i}, a_{q_i, q_j}, \kappa_i, \theta_i, \eta_1, \eta_{2,1}, \eta_{2,2}, \dots, \eta_{2,N})$  y está dado por

$$L(\pi_{q_i}, a_{q_i, q_j}, \kappa_i, \theta_i, \eta_1, \eta_{2,1}, \eta_{2,2}, \dots, \eta_{2,N}) = Q(\lambda_c, \lambda_c^i) + \eta_1 \left( \sum_{i=1}^N \pi_{q_i} - 1 \right) + \sum_{i=1}^N \eta_{2,i} \left( \sum_{j=1}^N a_{q_i, q_j} - 1 \right) \quad (1.153)$$

donde  $\eta_1, \eta_{2,1}, \eta_{2,2}, \dots, \eta_{2,N}$  son los multiplicadores de Lagrange.

Las condiciones de primer orden son

$$\frac{\partial L}{\partial \pi_{q_i}} = \frac{p(\mathcal{O}, X_1 = q_i | \lambda_c^i)}{\pi_{q_i}} + \eta_1 = 0 \quad \text{para } i = 1, 2, \dots, N \quad (1.154)$$

$$\frac{\partial L}{\partial \eta_1} = \sum_{i=1}^N \pi_{q_i} - 1 = 0 \quad (1.155)$$

$$\frac{\partial L}{\partial a_{q_i, q_j}} = \frac{\sum_{t=1}^{T-1} p(\mathcal{O}, X_t = q_i, X_{t+1} = q_j | \lambda_c^i)}{a_{q_i, q_j}} + \eta_{2,i} = 0 \quad \text{para } i, j = 1, 2, \dots, N \quad (1.156)$$

$$\frac{\partial L}{\partial \eta_{2,i}} = \sum_{j=1}^N a_{q_i, q_j} - 1 = 0 \quad \text{para } i = 1, 2, \dots, N \quad (1.157)$$

$$\frac{\partial L}{\partial \kappa_i} = \sum_{t=1}^T \left( \ln z_t - \ln \theta_i - \frac{\Gamma'(\kappa_i)}{\Gamma(\kappa_i)} \right) p(\mathcal{O}, X_t = q_i | \lambda_c^i) = 0 \quad \text{para } i = 1, 2, \dots, N \quad (1.158)$$

$$\frac{\partial L}{\partial \theta_i} = \sum_{t=1}^T \left( -\frac{\kappa_i}{\theta_i} + \frac{z_t}{\theta_i^2} \right) p(\mathcal{O}, X_t = q_i | \lambda_c^i) = 0 \quad \text{para } i = 1, 2, \dots, N \quad (1.159)$$

De las ecuaciones (1.154) y (1.155) se obtiene

$$\pi_{q_i} = \frac{p(\mathcal{O}, X_1 = q_i | \lambda_c^i)}{p(\mathcal{O} | \lambda_c^i)}, \quad \text{para } i = 1, 2, \dots, N. \quad (1.160)$$

De las ecuaciones (1.156) y (1.157) se tiene

$$a_{q_i, q_j} = \frac{\sum_{t=1}^{T-1} p(\mathcal{O}, X_t = q_i, X_{t+1} = q_j | \lambda_c^i)}{\sum_{t=1}^{T-1} p(\mathcal{O}, X_t = q_i | \lambda_c^i)} \quad \text{para } i, j = 1, 2, \dots, N. \quad (1.161)$$

De la ecuación (1.159) se sigue que,

$$\theta_i = \frac{\sum_{t=1}^T p(\mathcal{O}, X_t = q_i | \lambda_c^i) z_t}{\kappa_i \sum_{t=1}^T p(\mathcal{O}, X_t = q_i | \lambda_c^i)}, \quad \text{para } i = 1, 2, \dots, N. \quad (1.162)$$

Finalmente, de la ecuación (1.158) se obtiene,

$$-\ln \theta_i - \Psi(\kappa_i) = -\frac{\sum_{t=1}^T \ln z_t p(\mathcal{O}, X_t = q_i | \lambda_c^i)}{\sum_{t=1}^T p(\mathcal{O}, X_t = q_i | \lambda_c^i)} \quad \text{para } i = 1, 2, \dots, N. \quad (1.163)$$

donde,  $\Psi(\kappa_i)$  es la función digamma. Considerando la definición 1.20 se sigue que los estimadores de los parámetros del modelo son:

$$\pi_{q_i} = \gamma_1(q_i)_c, \quad \text{para } i = 1, 2, \dots, N. \quad (1.164)$$

$$a_{q_i, q_j} = \frac{\sum_{t=1}^{T-1} \xi_t(q_i, q_j)_c}{\sum_{t=1}^{T-1} \gamma_t(q_i)_c} \quad \text{para } i, j = 1, 2, \dots, N. \quad (1.165)$$

$$\theta_i = \frac{\bar{z}_t}{\kappa_i} \quad \text{para } i = 1, 2, \dots, N. \quad (1.166)$$

$$\ln \kappa_i - \Psi(\kappa_i) = \ln \bar{z}_t - \overline{\ln z_t} \quad \text{para } i = 1, 2, \dots, N. \quad (1.167)$$

Donde,

$$\bar{z}_t = \frac{\sum_{t=1}^T \gamma_t(q_i)_c z_t}{\sum_{t=1}^T \gamma_t(q_i)_c}, \quad \text{para } i = 1, 2, \dots, N. \quad (1.168)$$

$$\overline{\ln z_t} = \frac{\sum_{t=1}^T \ln z_t \gamma_t(q_i)_c}{\sum_{t=1}^T \gamma_t(q_i)_c} \quad \text{para } i = 1, 2, \dots, N. \quad (1.169)$$

### 1.4.7. Estimación de los parámetros de un HMM bajo la distribución Lognormal multivariable

En este apartado se obtienen los estimadores de los parámetros de un modelo de Markov oculto con espacio de observaciones continuo cuando se asume la distribución Lognormal como la distribución de las observaciones del modelo. Se emplea el algoritmo EM para encontrar los parámetros del modelo. Antes de establecer el algoritmo EM aplicado al modelo de Markov oculto se dan algunas definiciones.

**Definición 1.33.** Sea  $Y_t = (Y_{t,1}, Y_{t,2}, \dots, Y_{t,d})^\top$  un vector aleatorio de dimensión  $d$ , se dice que  $Y_t$  tiene una *distribución Lognormal multivariable* si su función de densidad, denotada como  $f_{y_t}(\mu, \Sigma)$  está dada por

$$f_{y_t}(\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}} \prod_{k=1}^d y_{t,k}} e^{-\frac{1}{2}(\ln y_t - \mu)^\top \Sigma^{-1} (\ln y_t - \mu)} \quad (1.170)$$

donde el vector  $\mu$  de dimensión  $d$  y la matriz  $\Sigma$  son los parámetros de la distribución.

**Definición 1.34.** Dado un modelo de Markov oculto con espacio de observaciones continuo  $\{X_t, Z_t\}$ ,  $t > 0$  y parámetros  $\lambda_c = (A, \mathcal{B}, \pi)$  se dice que se asume la *distribución Lognormal multivariable para la variable de observaciones*  $Z_t$  si para cada posible estado de la cadena de Markov asociada al modelo, la función de densidad condicional de la variable  $Z_t$  dado que  $X_t = q_i$  para  $i = 1, 2, \dots, N$  está dada por

$$b(z_t|q_i) = f_{z_t}(\mu_i, \Sigma_i) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma_i|^{\frac{1}{2}} \prod_{k=1}^d z_{t,k}} e^{-\frac{1}{2}(\ln z_t - \mu_i)^\top \Sigma_i^{-1} (\ln z_t - \mu_i)}. \quad (1.171)$$

Con la notación y definición establecidas, se procede a calcular los pasos del algoritmo EM.

#### Paso E del algoritmo EM

Enseguida se presenta la función de Baum Welch

$$\begin{aligned} Q(\lambda_c, \lambda_c^i) &= E[\ln p(\mathcal{O}, \mathcal{I}|\lambda_c)|\mathcal{O}, \lambda_c^i] \\ &= \sum_{\mathcal{I}} \ln (p(\mathcal{O}, \mathcal{I}|\lambda_c)) p(\mathcal{O}, \mathcal{I}|\lambda_c^i) \\ &= \sum_{\mathcal{I}} \ln (\pi_{x_1} b(z_1|x_1) \prod_{t=1}^{T-1} a_{x_t, x_{t+1}} b(z_{t+1}|x_{t+1})) p(\mathcal{O}, \mathcal{I}|\lambda_c^i) \\ &= \sum_{\mathcal{I}} \ln(\pi_{x_1}) p(\mathcal{O}, \mathcal{I}|\lambda_c^i) + \sum_{\mathcal{I} \in \mathcal{J}} \left( \sum_{t=1}^T \ln b(z_t|x_t) \right) p(\mathcal{O}, \mathcal{I}|\lambda_c^i) + \sum_{\mathcal{I} \in \mathcal{J}} \left( \sum_{t=1}^{T-1} \ln a_{x_t, x_{t+1}} \right) p(\mathcal{O}, \mathcal{I}|\lambda_c^i) \\ &= \sum_{i=1}^N \sum_{t=1}^T \left( -\frac{d}{2} \ln 2\pi - \frac{1}{2} \ln |\Sigma_i| - \sum_{r=1}^d \ln z_{t,r} - \frac{1}{2} (\ln z_t - \mu_i)^\top \Sigma_i^{-1} (\ln z_t - \mu_i) \right) p(\mathcal{O}, X_t = q_i|\lambda_c^i) \\ &\quad + \sum_{i=1}^N \ln \pi_{q_i} p(\mathcal{O}, X_1 = q_i|\lambda_c^i) + \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^{T-1} \ln a_{q_i, q_j} p(\mathcal{O}, X_t = q_i, X_{t+1} = q_j|\lambda_c^i). \end{aligned} \quad (1.172)$$

### Paso M del algoritmo EM

Una vez determinada la función  $Q(\lambda_c, \lambda_c^i)$ , se procede a realizar el paso M del algoritmo EM, el cual consiste en resolver el problema,

$$\lambda_c^{i+1} = \arg \max_{\pi_{q_i}, a_{q_i, q_j}, \Sigma_i, \mu_i, \text{ para } i, j=1, 2, \dots, N} Q(\lambda_c, \lambda_c^i) \quad (1.173)$$

Sujeto a

$$\sum_{i=1}^N \pi_{q_i} = 1$$

$$\sum_{j=1}^N a_{q_i, q_j} = 1 \text{ para } i = 1, 2, \dots, N$$

El lagrangiano correspondiente al problema (1.173) es  $L(\pi_{q_i}, a_{q_i, q_j}, \Sigma_i, \mu_i, \eta_1, \eta_{2,1}, \eta_{2,2}, \dots, \eta_{2,N})$  y está dado por

$$L(\pi_{q_i}, a_{q_i, q_j}, \Sigma_i, \mu_i, \eta_1, \eta_{2,1}, \eta_{2,2}, \dots, \eta_{2,N}) = Q(\lambda_c, \lambda_c^i) + \eta_1 \left( \sum_{i=1}^N \pi_{q_i} - 1 \right) + \sum_{i=1}^N \eta_{2,i} \left( \sum_{j=1}^N a_{q_i, q_j} - 1 \right) \quad (1.174)$$

donde  $\eta_1, \eta_{2,1}, \eta_{2,2}, \dots, \eta_{2,N}$  son los multiplicadores de Lagrange.

Las condiciones de primer orden son

$$\frac{\partial L}{\partial \pi_{q_i}} = \frac{p(\mathcal{O}, X_1 = q_i | \lambda_c^i)}{\pi_{q_i}} + \eta_1 = 0 \quad \text{para } i = 1, 2, \dots, N \quad (1.175)$$

$$\frac{\partial L}{\partial \eta_1} = \sum_{i=1}^N \pi_{q_i} - 1 = 0 \quad (1.176)$$

$$\frac{\partial L}{\partial a_{q_i, q_j}} = \frac{\sum_{t=1}^{T-1} p(\mathcal{O}, X_t = q_i, X_{t+1} = q_j | \lambda_c^i)}{a_{q_i, q_j}} + \eta_{2,i} = 0 \quad \text{para } i, j = 1, 2, \dots, N \quad (1.177)$$

$$\frac{\partial L}{\partial \eta_{2,i}} = \sum_{j=1}^N a_{q_i, q_j} - 1 = 0 \quad \text{para } i = 1, 2, \dots, N \quad (1.178)$$

$$\frac{\partial L}{\partial \mu_i} = \sum_{t=1}^T (\ln z_t - \mu_i) p(\mathcal{O}, X_t = q_i | \lambda_c^i) = 0 \quad \text{para } i = 1, 2, \dots, N \quad (1.179)$$

$$\frac{\partial L}{\partial \Sigma_i^{-1}} = \sum_{t=1}^T (\Sigma_i - (\ln z_t - \mu_i)(\ln z_t - \mu_i)^\top) p(\mathcal{O}, X_t = q_i | \lambda_c^i) = 0 \quad \text{para } i = 1, 2, \dots, N \quad (1.180)$$

De las ecuaciones (1.175) y (1.176) se obtiene

$$\pi_{q_i} = \frac{p(\mathcal{O}, X_1 = q_i | \lambda_c^i)}{p(\mathcal{O} | \lambda_c^i)}, \quad \text{para } i = 1, 2, \dots, N. \quad (1.181)$$

De las ecuaciones (1.177) y (1.178) se tiene

$$a_{q_i, q_j} = \frac{\sum_{t=1}^{T-1} p(\mathcal{O}, X_t = q_i, X_{t+1} = q_j | \lambda_c^i)}{\sum_{t=1}^{T-1} p(\mathcal{O}, X_t = q_i | \lambda_c^i)} \quad \text{para } i, j = 1, 2, \dots, N. \quad (1.182)$$

De la ecuación (1.179) se sigue que,

$$\mu_i = \frac{\sum_{t=1}^T p(\mathcal{O}, X_t = q_i | \lambda_c^i) \ln z_t}{\sum_{t=1}^T p(\mathcal{O}, X_t = q_i | \lambda_c^i)}, \quad \text{para } i = 1, 2, \dots, N. \quad (1.183)$$

Finalmente, de la ecuación (1.180) se obtiene,

$$\Sigma_i = \frac{\sum_{t=1}^T p(\mathcal{O}, X_t = q_i | \lambda_c^i) (\ln z_t - \mu_i) (\ln z_t - \mu_i)^\top}{\sum_{t=1}^T p(\mathcal{O}, X_t = q_i | \lambda_c^i)} \quad \text{para } i = 1, 2, \dots, N. \quad (1.184)$$

Considerando las definiciones 1.20 y 1.21, se sigue que los estimadores de los parámetros del modelo son:

$$\pi_{q_i} = \gamma_1(q_i)_c, \quad \text{para } i = 1, 2, \dots, N. \quad (1.185)$$

$$a_{q_i, q_j} = \frac{\sum_{t=1}^{T-1} \xi_t(q_i, q_j)_c}{\sum_{t=1}^{T-1} \gamma_t(q_i)_c} \quad \text{para } i, j = 1, 2, \dots, N. \quad (1.186)$$

$$\mu_i = \frac{\sum_{t=1}^T \gamma_t(q_i)_c \ln z_t}{\sum_{t=1}^T \gamma_t(q_i)_c}, \quad \text{para } i = 1, 2, \dots, N. \quad (1.187)$$

$$\Sigma_i = \frac{\sum_{t=1}^T \gamma_t(q_i)_c (\ln z_t - \mu_i) (\ln z_t - \mu_i)^\top}{\sum_{t=1}^T \gamma_t(q_i)_c} \quad \text{para } i = 1, 2, \dots, N. \quad (1.188)$$

Una vez definidos los parámetros del modelo de Markov oculto bajo el supuesto de que las observaciones del modelo asumen diversas distribuciones de probabilidad, en el siguiente capítulo se presenta la aplicación del modelo para analizar el comportamiento de los niveles de ozono en la Ciudad de México.

# Capítulo 2

## Niveles de ozono en la Ciudad de México

En este capítulo, en primera instancia se describe el proceso de formación del ozono en la atmósfera y se justifica el por qué de su estudio. Asimismo, se describe el proceso de modelado de series temporales mediante modelos ocultos de Markov; en particular, se ilustra el modelado de la serie de promedios diarios de los niveles de ozono en la Ciudad de México a través de un modelo oculto de Markov homogéneo. Por último, se presentan los resultados obtenidos y su interpretación.

### 2.1. Conceptos

El ozono ( $O_3$ ) es un gas que se encuentra en dos partes de la atmósfera; la estratosfera y troposfera. El de la atmósfera superior, o estratosfera, es un gas esencial que ayuda a proteger a la Tierra de los dañinos rayos ultravioletas del sol. En contraste, el ozono hallado cerca de la superficie, en la troposfera, tiene efectos nocivos a la salud humana, la vegetación y los materiales, CEC. (2008). Por esta razón, el ozono se describe a menudo como “bueno arriba y malo de cerca”. Figura 2.1

El ozono troposférico, también llamado ozono ambiental u ozono de bajo nivel, es un gas incoloro y muy irritante creado por reacciones fotoquímicas a la luz del sol entre los óxidos de nitrógeno ( $NO_x$ ) y los compuestos orgánicos volátiles (COV), especialmente hidrocarburos. Entre las principales fuentes para la emisión de estos componentes se encuentran la quema de combustibles fósiles en el transporte, la industria y las centrales eléctricas; asimismo, en la evaporación de combustibles líquidos y solventes, Velázquez y Jiménez (2007). El ozono también se forma en niveles bajos proveniente de emisiones naturales que incluye las emisiones de plantas y suelos, los incendios forestales y los rayos durante las tormentas eléctricas. Figura 2.2.

El presente trabajo se enfoca al estudio del ozono troposférico debido a su impacto en la calidad de vida de la población en la Ciudad de México.

El ozono troposférico, componente básico del smog, se considera un problema “sin umbral” porque incluso en muy pequeñas cantidades en el aire tiene efectos nocivos en la salud humana, principalmente en los sistemas cardiovascular y respiratorio; por mencionar, se desprenden daños para el ser humano en ojos, fosas nasales, garganta y bronquios, provocando inflamaciones en las mucosas y conjuntiva, Velázquez y Jiménez (2007). Asimismo, la exposición al ozono se ha vinculado con la mortalidad prematura y una gama de cuestiones de morbilidad, como admisiones en hospitales, al ser causante de asma y agravar las dolencias de grupos de población más vulnerable como niños, ancianos o enfermos crónicos, Sousa et al. (2007) y Filleul et al. (2006). En general, el ozono troposférico también daña a los animales y a la vegetación; con respecto a esta última, afecta la productividad de los cultivos, las

**Relación tipo de ozono - capas de la Tierra**

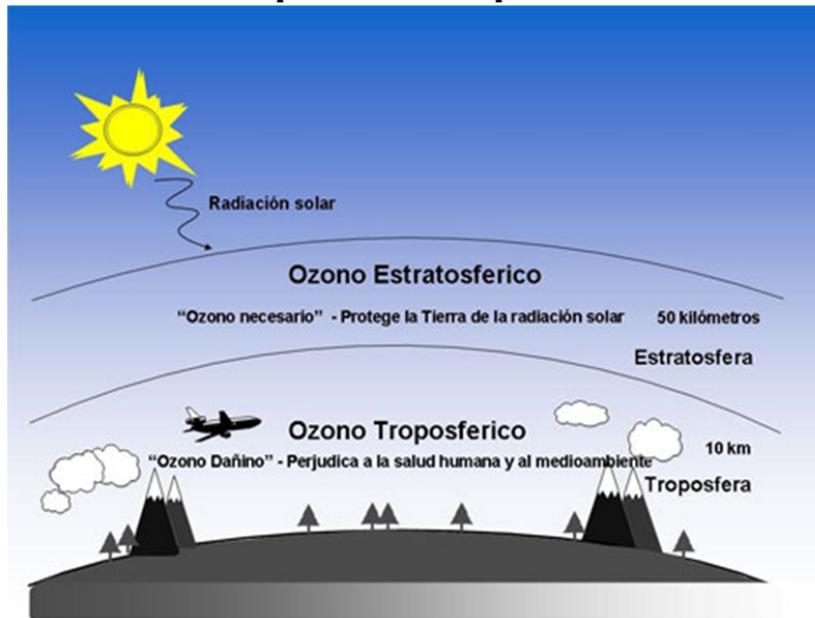


Figura 2.1: Clasificación del ozono según su ubicación.

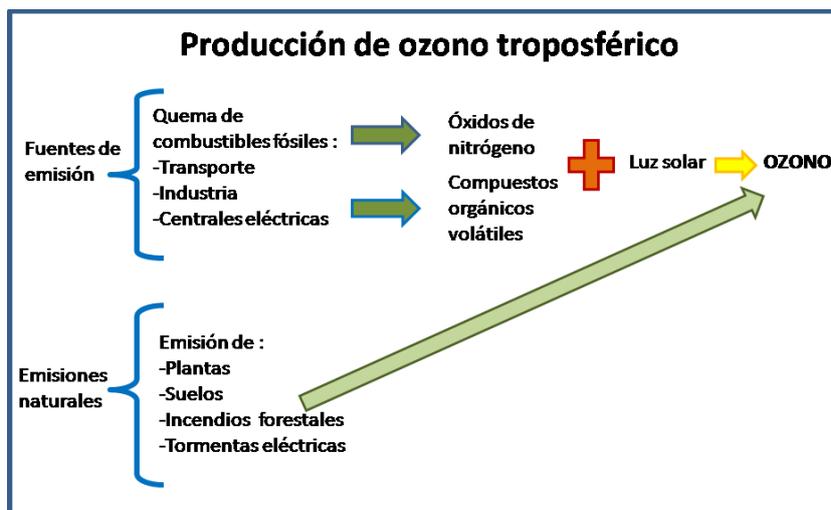


Figura 2.2: Esquema de la producción de ozono troposférico.

flores, los matorrales y los bosques; esto se debe a que reduce la fotosíntesis, lo que a su vez origina una reducción en el rendimiento de las plantas cultivadas, Davis et al. (2006) y Scebba et al. (2006). Es más, puede deteriorar el algodón y los materiales sintéticos, producir roturas en el caucho, acelerar el desteñido de ciertas pinturas y recubrimientos, Escobedo y Chacalo (2008).

Debido a los efectos mencionados, en México la Norma Oficial Mexicana (NOM-020-SSA1-2014) recomienda mantener concentraciones menores a 0,095 *ppm* para el promedio de 1 hora, y menores a 0,070 *ppm* para el promedio de 8 horas (máximo anual). Además, de acuerdo con la Norma Oficial Mexicana (NOM-036-SEMARNAT-1993) las unidades habituales para expresar la concentración del ozono en el aire ambiente son microgramos/metro cúbico,  $\frac{\mu g}{m^3}$ , miligramo/metro cúbico,  $\frac{mg}{m^3}$ , partes por millón, *ppm*, IMECA (Índice Metropolitano de Calidad del Aire), entre otras. La concentración expresada en *ppm* se refiere a la cantidad de unidades de ozono por cada millón de unidades de aire; por ejemplo, 5 *ppm* de ozono significa 5 litros de ozono en cada millón de litros de aire. Por otro lado, el IMECA es una escala arbitraria, normalizada entre 0 y 500 que permite a la población interpretar el grado de peligrosidad que exhibe la atmósfera en un tiempo y lugar dados. La relación entre la concentración expresada en *ppm* e IMECA siempre es lineal;

$$\text{IMECA } O_3 = \frac{100 C}{0,11}, \quad (2.1)$$

donde, IMECA  $O_3$  es la concentración expresada en unidades IMECA y  $C$  es la concentración en unidades de *ppm*.

Cabe mencionar, que de acuerdo a la Norma ya mencionada, el método para determinar la concentración de ozono en el aire ambiente es el de fotometría en la región de radiación ultravioleta; éste se basa en el principio fotométrico de la absorción de luz en el rango de la radiación ultravioleta por el ozono. Para reportar los valores de calidad del aire al público, se calculan las concentraciones promedio por minuto expresadas en unidades de *ppm*, a partir de estos se determinan los promedios horarios, siendo éstos últimos, expresados en IMECA, los reportados en cada estación de monitoreo.

## 2.2. Regímenes del ozono

Numerosos estudios relacionados con los aerosoles atmosféricos basan sus conclusiones en el desarrollo de análisis de contribución de fuentes (ACF); el cual consiste en estudiar el origen de las diversas fuentes de contaminación y la estimación de sus contribuciones a los niveles de la contaminación atmosférica ambiente. La información proporcionada por estos análisis es crucial para el desarrollo e implementación de políticas para la protección de la salud humana y el medio ambiente, Gomez et al. (2015). En este capítulo se propone el uso de los HMM homogéneos como una técnica estadística exploratoria para complementar la elaboración de un ACF.

La razón por la cual emplear los HMM en este tipo de análisis radica en que los HMM permiten agrupar las observaciones de un contaminante en subconjuntos; estos subconjuntos, que se llaman de aquí en adelante como regímenes de concentración o simplemente regímenes, se caracterizan porque en cada uno se ubican las observaciones con concentraciones similares y que al mismo tiempo difieren en su concentración de otras observaciones agrupadas en otros regímenes. En realidad, los HMM permiten asociar cada estado de la cadena de Markov con un régimen de concentración y con ello caracterizar la distribución de las observaciones en cada régimen. En particular, será de interés la media de la

distribución de cada estado o régimen, ya que este valor representará la concentración de diferentes fuentes de contaminantes; al respecto, Lenschow et al. (2001) establece que:

- 1) La media de la distribución asociada al primer estado o régimen representa la concentración característica del área de estudio, del cual no se esperan grandes cambios a lo largo del tiempo, bajo condiciones atmosféricas estables; puede ser considerada un indicador de la exposición real al contaminante.
- 2) La media de la distribución asociada al segundo estado puede interpretarse como la concentración en días afectados por un incremento en las contribuciones de origen antropogénico, debido a actividades que se realizan en la región de estudio. Este valor podría estar sujeto a más variaciones en comparación con la media en el primer estado.
- 3) La media de la distribución asociada al tercer régimen es la concentración de aquellos días afectados por contribuciones muy variables en concentración y son las principales responsables de la superación del valor límite diario establecido por las normas oficiales para algún contaminante.
- 4) Por último, la media de la distribución de las observaciones en el cuarto estado de la cadena de Markov, representa la concentración en aquellos días afectados por episodios severos, y por lo general poco frecuentes.

## 2.3. Datos

Para ejemplificar el uso de los HMM como auxiliar en el análisis de fuentes de emisión de contaminantes se emplean datos de los niveles de ozono en la Ciudad de México. La información fue obtenida en la página electrónica del Sistema de Monitoreo Atmosférico de la Ciudad de México <http://www.aire.df.gob.mx/>; donde se reportan, desde 1992, la concentración promedio horaria de ozono, en unidades IMECA, para las cinco regiones de la ciudad; es decir, noroeste, noreste, centro, sureste y suroeste. Con esta información, se obtuvo el promedio diario de concentración de ozono en unidades IMECA; cabe mencionar que se descartaron algunos días debido a que presentaban menos del 80 por ciento de registros. Por consiguiente, para cada región de la ciudad, se construyeron 24 series de tiempo, una por cada año desde 1992 hasta 2015, dando un total de 120 series.

## 2.4. Procedimiento

Primero se ordenó la información y se realizaron las pruebas de bondad de ajuste tipo Pearson, Altman (2004), para cada una de las 120 series, con la finalidad de determinar si los modelos ocultos de Markov homogéneos son idóneos para modelar los datos. Si bien en la mayoría de las aplicaciones de los HMM se omite o no se menciona esta parte, es importante verificar los supuestos antes de ajustar el modelo. Una vez que se realizaron las pruebas de bondad de ajuste, se procede a realizar el ajuste de los modelos, donde los principales aspectos a considerar son: cómo determinar el número de estados de la cadena de Markov asociada al HMM, cuál es la distribución que se asume para las observaciones y cómo determinar los parámetros iniciales y finales del HMM.

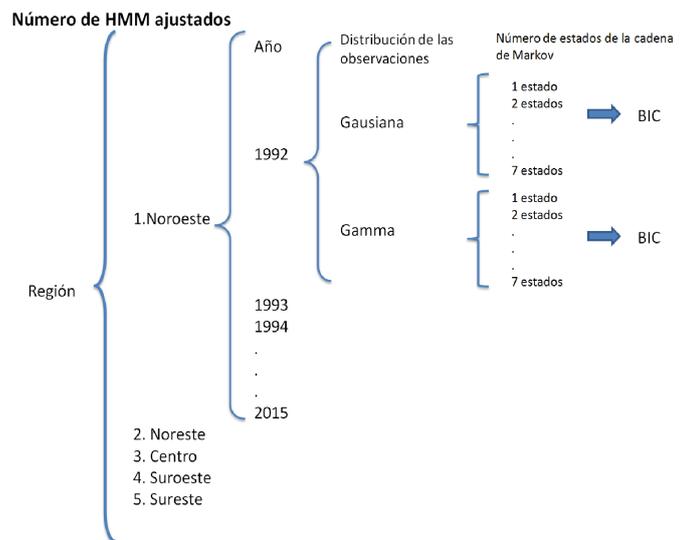


Figura 2.3: Esquema del número de modelos HMM ajustados a los datos.

### Distribución de las observaciones

Debido a la enorme cantidad de información que se tiene, en este trabajo se consideró sólo trabajar con HMM donde las observaciones asumen las distribuciones gaussiana y gamma. Figura 2.3.

### Número de estados de la cadena de Markov

Se consideraron 7 posibles modelos; con uno, dos, tres, cuatro, cinco, seis y siete estados en la cadena de Markov asociada al HMM. Para seleccionar al mejor modelo se empleó el criterio de información bayesiana, BIC. Figura 2.3.

### Valores iniciales de los parámetros del modelo HMM

En general, se considera el ajuste de un HMM con una cadena de Markov de  $k$  estados y distribución de observaciones gaussiana o gamma para la serie de datos<sup>1</sup>  $R^\star$ . Se debe determinar los valores iniciales del vector de probabilidad inicial de los estados, de la matriz de transición y de los parámetros de las distribuciones gamma y gaussiana para iniciar el algoritmo recursivo EM.

Dado que se tienen  $k$  estados en la cadena de Markov, entonces las observaciones son el resultado de la mezcla de  $k$  distribuciones, gaussiana o gamma, según sea el caso; por lo que en primera instancia, se dividió el rango de los datos de la serie  $R^\star$  en  $k$  intervalos; los datos en estos intervalos se asociaron a las  $k$  distribuciones correspondientes a cada uno de los  $k$  estados. Cabe mencionar que esta asociación permite generar una secuencia de estados a partir de la secuencia de observaciones. Considerando esta agrupación se obtienen los siguientes valores.

Se asume que en la cadena de Markov, todos los estados tienen la misma probabilidad de iniciar la secuencia de observaciones, por lo que la probabilidad inicial de cada estado es igual a  $1/k$ . Con respecto a los valores iniciales de la matriz de transición, el valor inicial de la probabilidad de transición del estado  $i$  al estado  $j$  se estima como la frecuencia relativa de pasar del estado  $i$  al estado  $j$ ; es decir,

<sup>1</sup>Hace referencia a alguna serie de datos entre los años de 1992-2015 para alguna región de la ciudad (noroeste, noreste, centro, suroeste o sureste).

a partir de la secuencia de estados, que se menciona en el párrafo anterior, se contó el número de veces que se pasó del estado  $i$  al estado  $j$  y se dividió entre el número de veces que se visitó el estado  $i$ . Por último, dado que cada uno de los  $k$  estados tiene asociado un grupo de observaciones, los valores iniciales de los parámetros de las  $k$  distribuciones se calculan empleando los estimadores de máxima verosimilitud sobre estos conjuntos. Cuando se asume que las observaciones son de tipo gaussianas se emplean los siguientes estimadores

$$\hat{\mu}_i = \bar{z} \quad (2.2)$$

$$\hat{\sigma}_i^2 = \sum_{i=1}^n \frac{z_i^2}{n-1} - \frac{n\bar{z}^2}{n-1}, \quad (2.3)$$

donde  $\hat{\mu}_i$  y  $\hat{\sigma}_i^2$  son estimadores de la media y varianza, respectivamente, de la distribución asociada al estado  $i$  para  $i = 1, 2, \dots, k$ . Y cuando se asume que las observaciones son de tipo gamma se emplean los estimadores,

$$\theta_i = \frac{\bar{z}}{\kappa_i} \quad (2.4)$$

$$\kappa_i = \frac{\bar{z}^2}{s_i^2}, \quad (2.5)$$

donde  $s_i^2$  es la varianza muestral de los datos que pertenecen al intervalo asociado al estado  $i$ .  $\theta_i$  y  $\kappa_i$  son los parámetros de escala y de forma, respectivamente.

### Valores finales de los parámetros del modelo HMM

Una vez que se tienen los valores iniciales del modelo HMM se realiza el proceso iterativo del método EM, descrito en el capítulo anterior, hasta obtener los parámetros finales del modelo. Los estimadores obtenidos al final de este proceso y que es importante tener presente para comprender la sección siguiente son; el estimador del vector de probabilidades inicial,  $\hat{\pi}$ , estimador de la matriz de probabilidades de transición,  $\hat{A}$ , media de la distribución asociada al régimen de contaminación o estado  $i$  que será denotada como  $m_i$ , así como la respectiva desviación estándar; nivel medio de exposición anual de la población que será denotado como  $M$ , así como su respectiva desviación estándar,  $SD$ . También se obtuvieron el rango más probable de las observaciones asociadas a cada estado y el número de observaciones de la serie que se agruparon en cada estado. El algoritmo para el cálculo de los estimadores mencionados se implementó en el software R, <https://www.r-project.org/>.

## 2.5. Resultados

Con respecto a la prueba de bondad de ajuste, únicamente en 9 series de las 120 se rechazó que los HMM fueran adecuados para modelar su comportamiento, es importante señalar que estas 9 series corresponden a los primeros años de medición y registro de los niveles de ozono en la ciudad y no se reportaron las mediciones durante todas las horas, en consecuencia faltan observaciones; puede deberse a esto que los HMM no hayan sido adecuados para modelar estas series.

Debido a que a las 111 series restantes se ajustó un HMM con distribución gaussiana y gamma en las observaciones, se obtuvo un total de 222 modelos, por lo que el total de resultados obtenidos es extenso. Por lo anterior, en este apartado únicamente se muestran algunos modelos que permiten

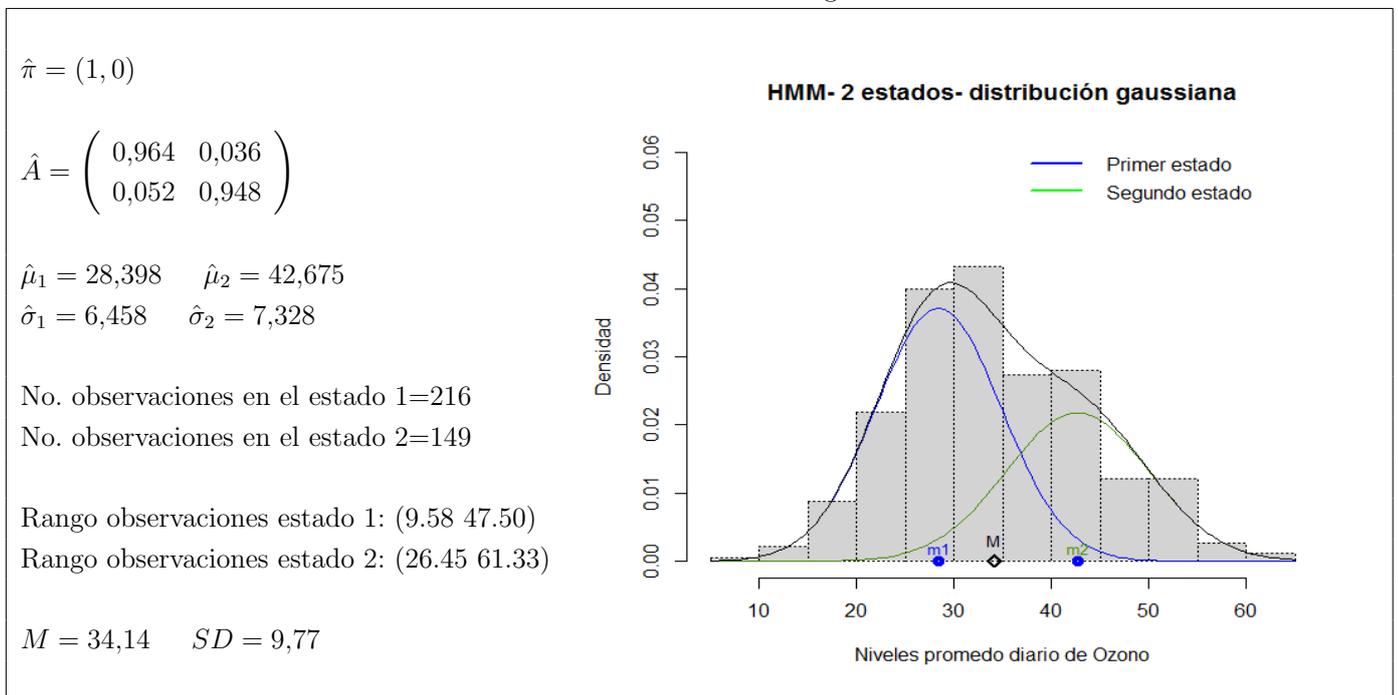
ilustrar los principales hallazgos; los modelos que se mencionan enseguida corresponden a los años cuando el HMM es un mejor modelo, de acuerdo con el criterio de información bayesiana.

**Zona noreste**

Para la zona noreste se muestran los resultados con mayor detalle, porque es aquella donde se presentó una mayor variedad en el número de estados del HMM ajustado entre un año y otro. Primero, se presenta el HMM, con distribución gaussiana, ajustado en los años 2002, 2012 y 2015; posteriormente, el modelo HMM, con distribución gamma, ajustado en los años 1998, 2006, y 2014.

HMM- 2 estados- distribución gaussiana, año 2002. En el cuadro 2.1, se muestra los valores finales de los parámetros de un HMM con dos estados, distribución gaussiana en las observaciones, para el año 2002; también se muestra la gráfica de la función de densidad de las distribuciones asociadas a cada uno de los estados. El primer estado, contaminación de fondo, tiene una media de  $m_1 = 28,39$  y el segundo estado tiene media igual a  $m_2 = 42,675$ ; siendo la desviación estándar mayor para la distribución en el estado 2. Cabe mencionar que la media de la serie igual a  $M = 34,14$  es el valor promedio de la contaminación ambiental que toma encuentra la dependencia temporal de las observaciones. Asimismo, de la matriz de probabilidades de transición, se observa que el estado 1 fue el que generó un mayor número de observaciones, 216 en total. En la Figura 2.4 se ilustra la trayectoria de estados más probable que haya generado la secuencia de observaciones de ozono en el año 2002.

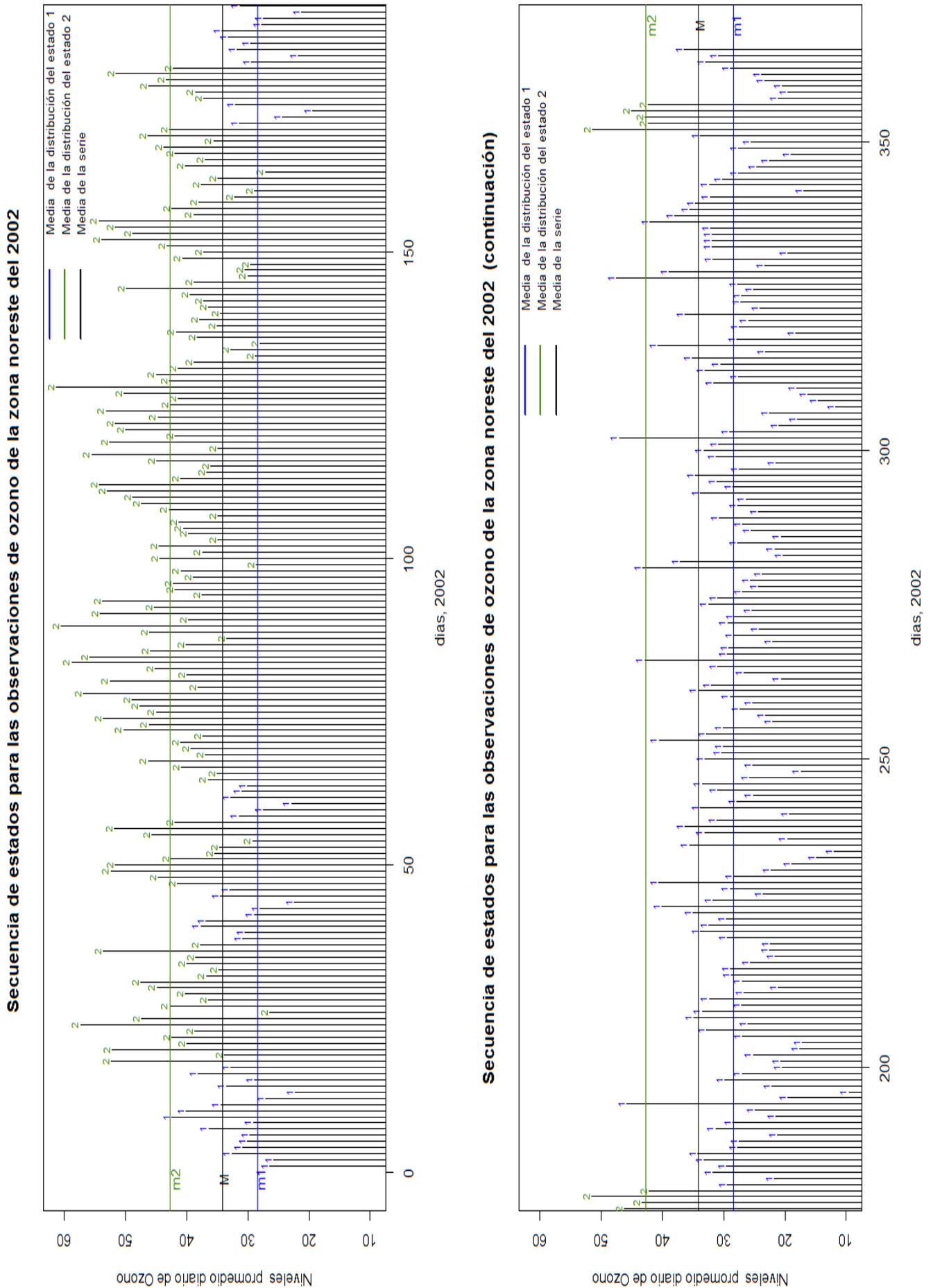
HMM 2 estados- distribución gaussiana.



Cuadro 2.1: HMM-2 estados- distribución gaussiana, zona noreste, año 2002. De un lado los parámetros del modelo; del otro, en color azul la gráfica de la función de densidad asociada al estado 1, de color verde la función de densidad asociada al estado 2 y de color negro la mezcla de las funciones de densidad del estado 1 y 2.

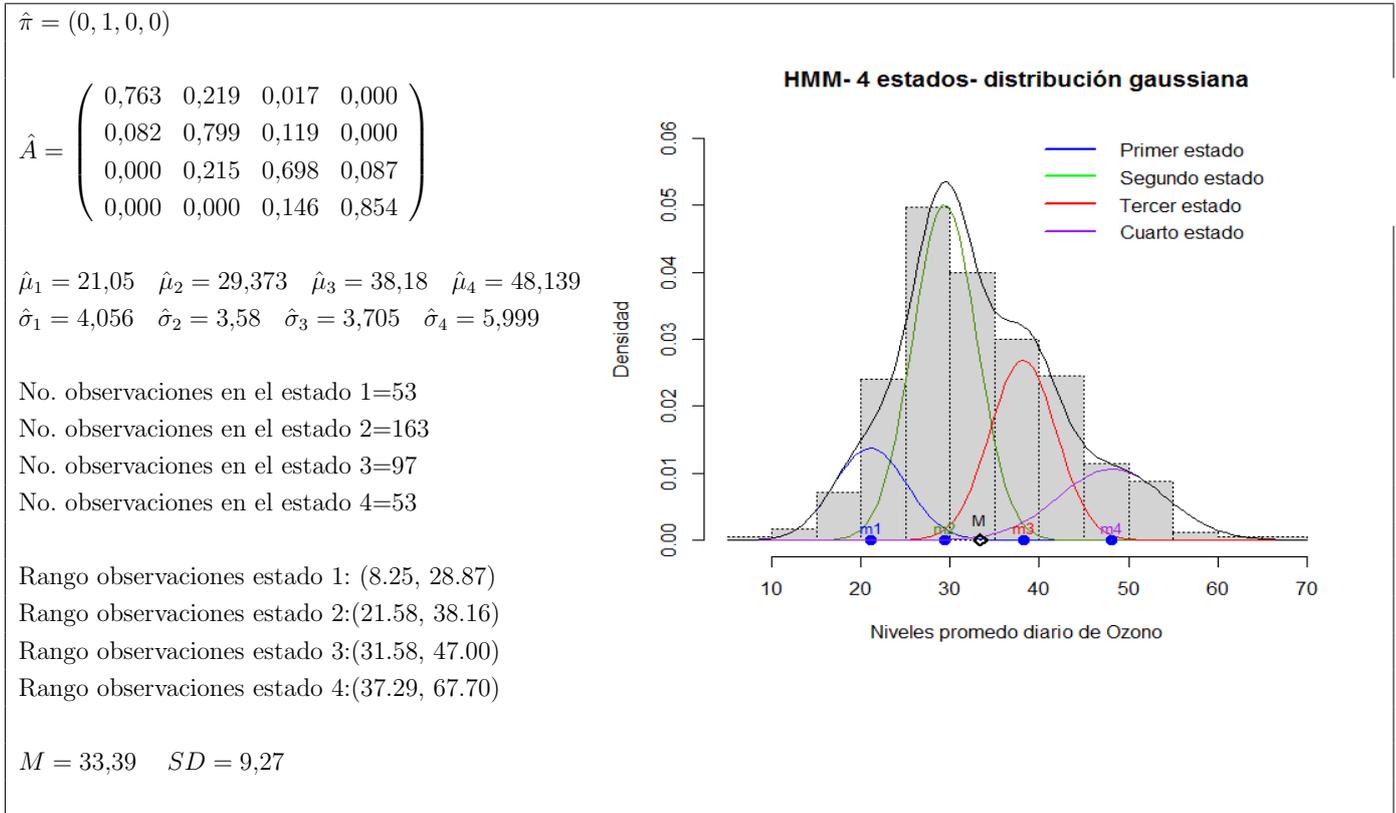
HMM- 4 estados- distribución gaussiana, año 2012. En el año 2012, el modelo que mejor ajustó fue un HMM con 4 estados y distribución gaussiana en las observaciones; los parámetros del modelo se muestran en el Cuadro 2.2 así como la gráfica de las funciones de densidad asociadas a cada uno de los estados. La media de la distribución asociada al estado 1 igual a  $m_1 = 21,05$  es la contaminación de fondo promedio a lo largo del año 2012. Cabe mencionar, que al estado 4 está asociada la distribución

Figura 2.4: Secuencia de estados más probable asociada a la secuencia de observaciones de ozono del año 2002 en la zona noreste, se obtuvieron a partir de un HMM con 2 estados- distribución gaussiana, donde  $m_1 = 28,39$ ,  $m_2 = 42,67$ ,  $M = 34,14$ .



con mayor media y mayor desviación estándar lo que concuerda con la idea de que en este estado ocurren las observaciones con valores de ozono más altas. En la Figura 2.5, se ilustra la secuencia de estados más probable que explica la secuencia de observaciones de ozono para este año; se puede apreciar que el estado 4, que corresponde a días con contaminación extrema ocurrió entre los días 102 y 155; es decir de mediados de abril y hasta finales de mayo.

HMM 4 estados- distribución gaussiana



Cuadro 2.2: HMM-4 estados- distribución gaussiana, zona noreste, año 2012. De un lado los parámetros del modelo, del otro en color azul la densidad asociada al estado 1, de verde la densidad asociada al estado 2, de rojo la densidad asociada al estado 3, de morado la densidad asociada al estado 4 y de negro la mezcla de densidades.

HMM- 3 estados- distribución gaussiana, año 2015. En el Cuadro 2.3 están los parámetros del modelo que mejor ajustó en la zona noreste en el año 2015, se trata de un HMM con 3 estados y distribución gaussiana en las observaciones. Se puede ver que los estados 1 y 2 explican el mayor número de observaciones de este año con 135 y 156, respectivamente. En la matriz de transiciones se observan valores altos en la diagonal de la matriz lo que señala que la cadena de Markov permanece durante varios días en el mismo estado, Figura 2.6, además valores cercanos a cero en los extremos señala, por ejemplo, que transiciones del estado 1 al 3 y viceversa son poco probables de observar; lo cual tiene que ver con el hecho de que no es posible pasar de un nivel muy alto de contaminación promedio a uno muy bajo en tan sólo un día.

HMM- 4 estados- distribución gamma, año 1998. Para el año 1998, el modelo que mejor ajustó fue un HMM con 4 estados y distribución gamma en las observaciones. De la matriz de probabilidades de transición se observa que transiciones del estado 3 al 1, del estado 4 al 1, del estado 4 al 2 y viceversa, son poco probables; asimismo, en el estado 2, es donde se produce el mayor número de observaciones, es decir, un mayor número de días los niveles promedio diario de ozono se encontraron en un rango de 22.16 a 59.16 IMECAS. Es de observar que el valor medio del promedio diario de ozono ambiental en este año fue de  $M = 46,41$  bastante superior con respecto a años más recientes, Cuadro 2.4. En la

Figura 2.5: Secuencia de estados más probable asociada a la secuencia de observaciones de ozono en el año 2012 de la zona noreste, se obtuvieron con un HMM de 4 estados- distribución gaussiana, donde  $m_1 = 21,05$ ,  $m_2 = 29,37$ ,  $m_3 = 38,18$ ,  $m_4 = 48,13$ ,  $M = 33,39$ .

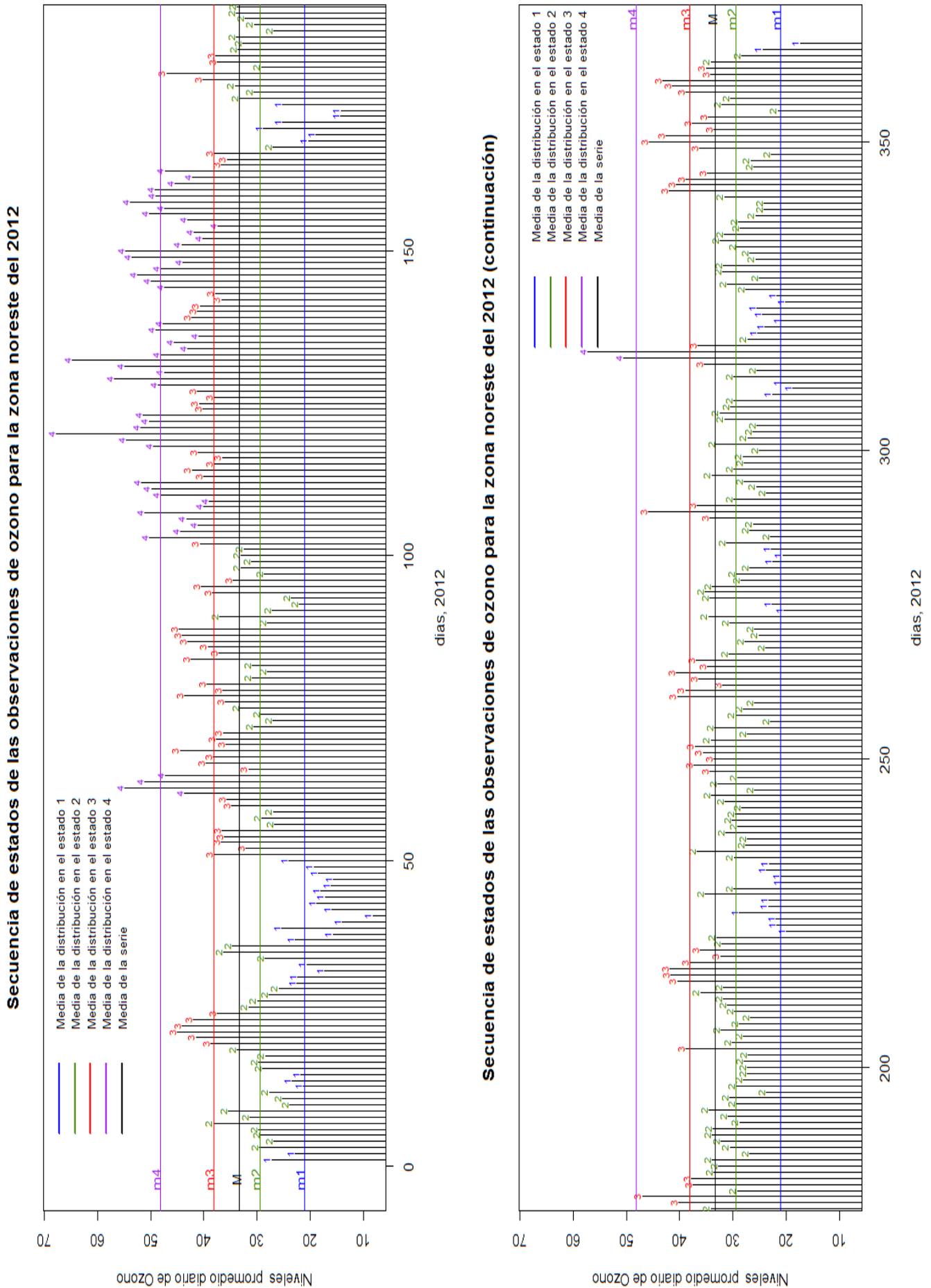
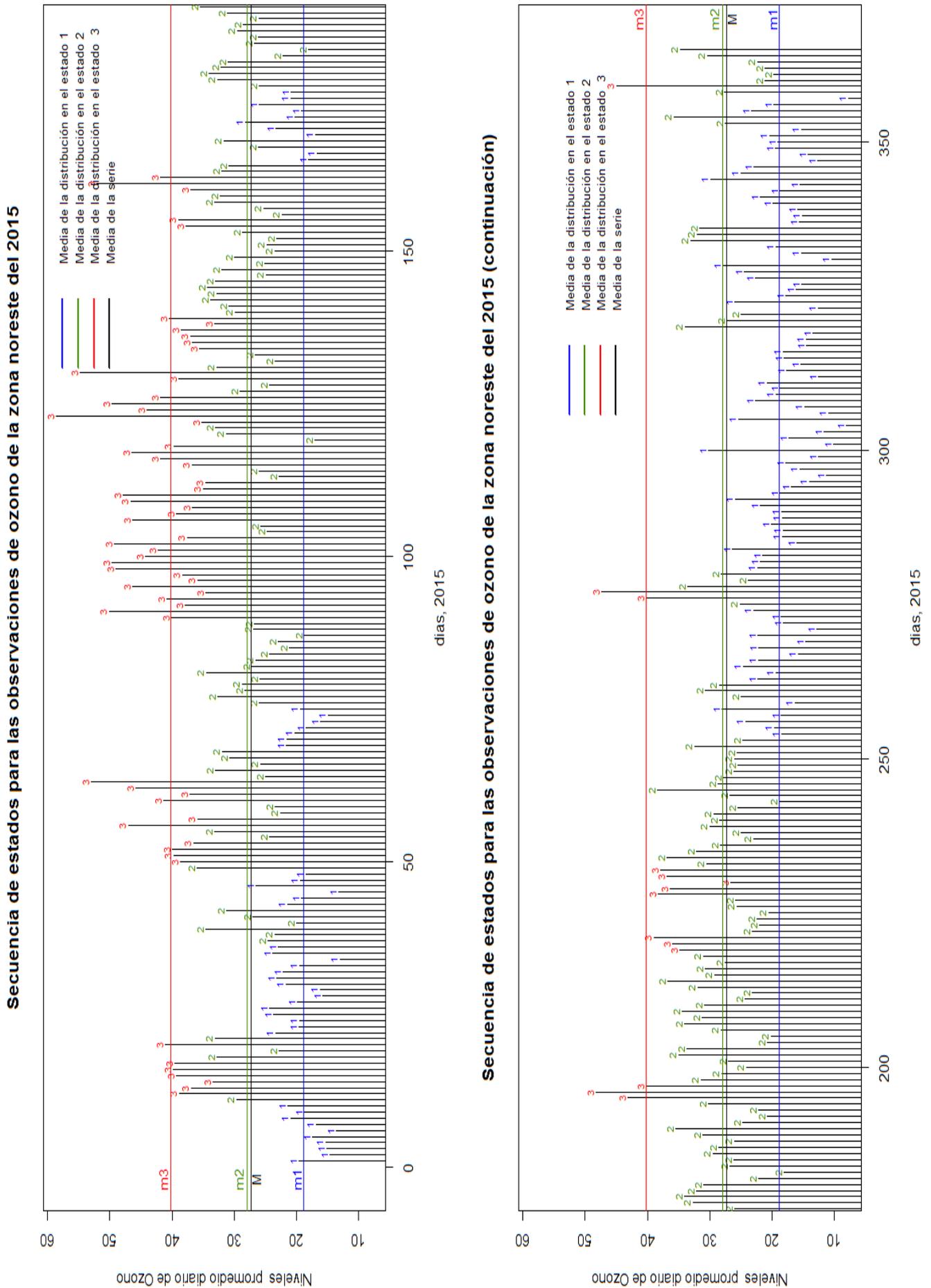
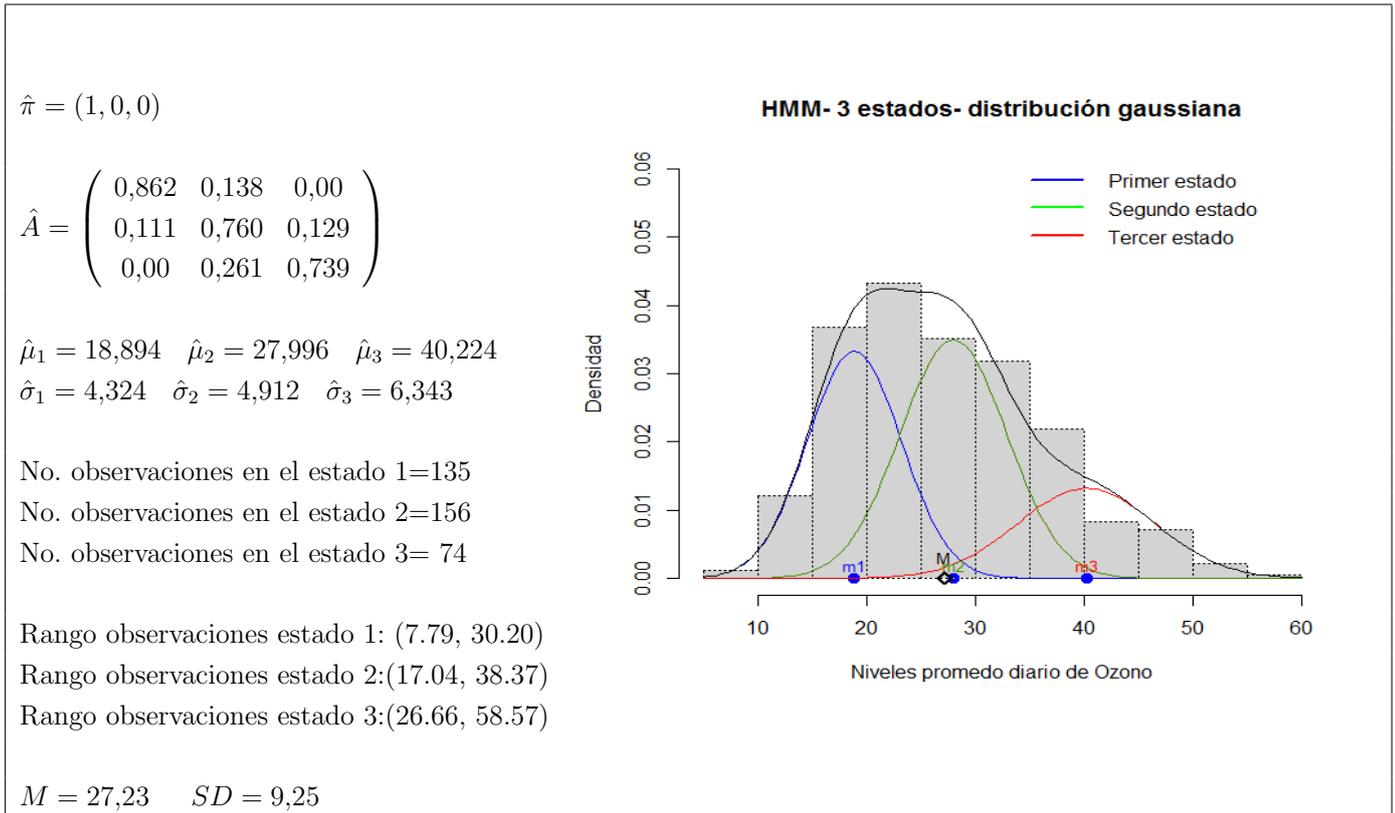


Figura 2.6: Secuencia de estados más probable asociada a la secuencia de observaciones de ozono del 2015 en el noreste, se obtuvieron con un HHM con 3 estados- distribución gaussiana, donde  $m_1 = 18,89$ ,  $m_2 = 27,96$ ,  $m_3 = 40,22$ ,  $M = 27,23$ .



HMM 3 estados- distribución gaussiana.



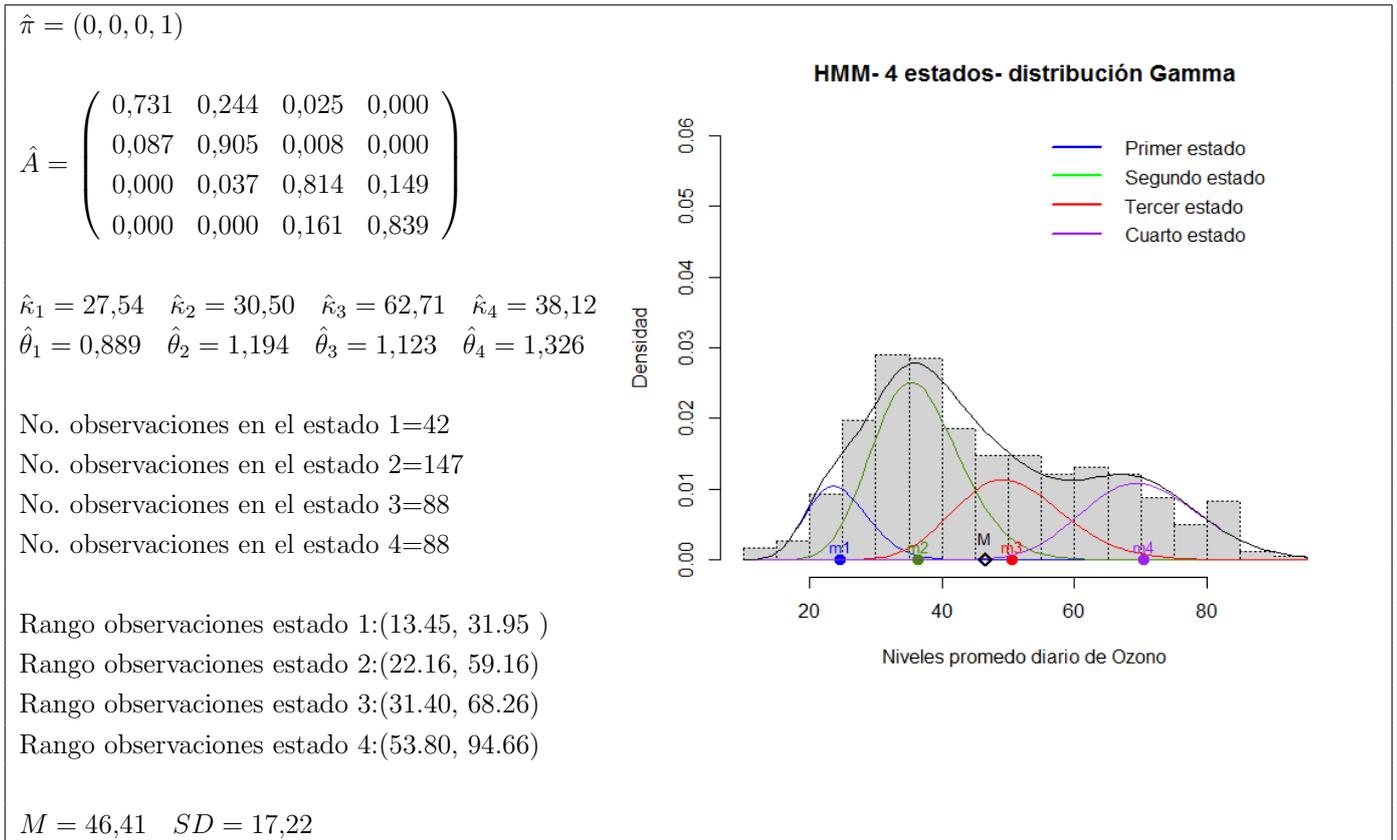
Cuadro 2.3: HMM-3 estados- distribución gaussiana, zona noreste, año 2015. De un lado, los parámetros del modelo, del otro de color azul la función de densidad asociada al estado 1, de color verde la función de densidad del estado 2, de rojo la densidad del estado 3 y de negro la mezcla de densidades.

parte superior de la Figura 2.7 se presenta la secuencia de estados que mejor explica la secuencia de observaciones de ozono; los estados 3 y 4 ocurrieron con más frecuencia en la primer parte del año, mientras que los estados 1 y 2 explican las observaciones en la parte restante del año, esto a diferencia de los modelos explicados anteriormente donde se observa una mayor alternancia entre estados.

HMM- 2 estados- distribución gamma, año 2006. En el Cuadro 2.5 se encuentran los parámetros del modelo ajustado en el 2006; se trata de un HMM con dos estados; en el estado 1 que corresponde a la contaminación de fondo se encuentra la mayoría de las observaciones de ese año; lo que significa que en 60 por ciento de los días los niveles promedio diario de ozono estuvieron en un rango de 12.45 a 47.29 IMECAS. Asimismo, los niveles altos de ozono asociados al estado 2 de la cadena de Markov ocurrieron en su mayoría en la primera mitad del año, mientras que los niveles bajos ocurrieron en la segunda mitad del año, como lo muestra la secuencia de estados en la parte inferior de la Figura 2.7.

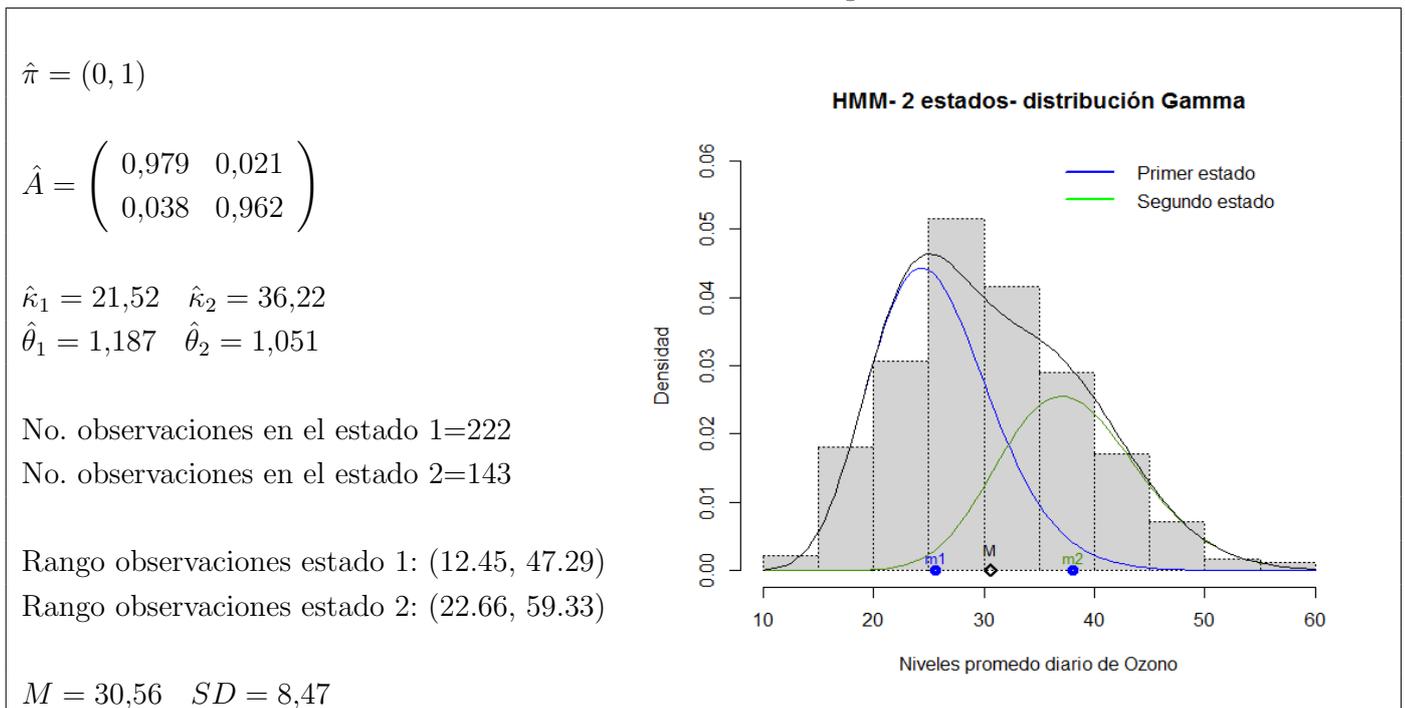
HMM- 3 estados- distribución gamma, año 2014. Por último, en el Cuadro 2.6 están los parámetros finales del modelo HMM de 3 estados ajustado en el año 2014. Del vector de probabilidades iniciales, es más probable que la cadena de Markov comience con el estado 2, que es también el estado en el que se producen el mayor número de observaciones en ese año. En el estado 1, al que pertenecen observaciones del régimen de contaminación de fondo se encuentra una cantidad pequeña de observaciones, únicamente 62. Al igual que en otros años, transiciones entre los estados 3 y 1, y viceversa, son poco probables.

HMM 4 estados- distribución gamma.



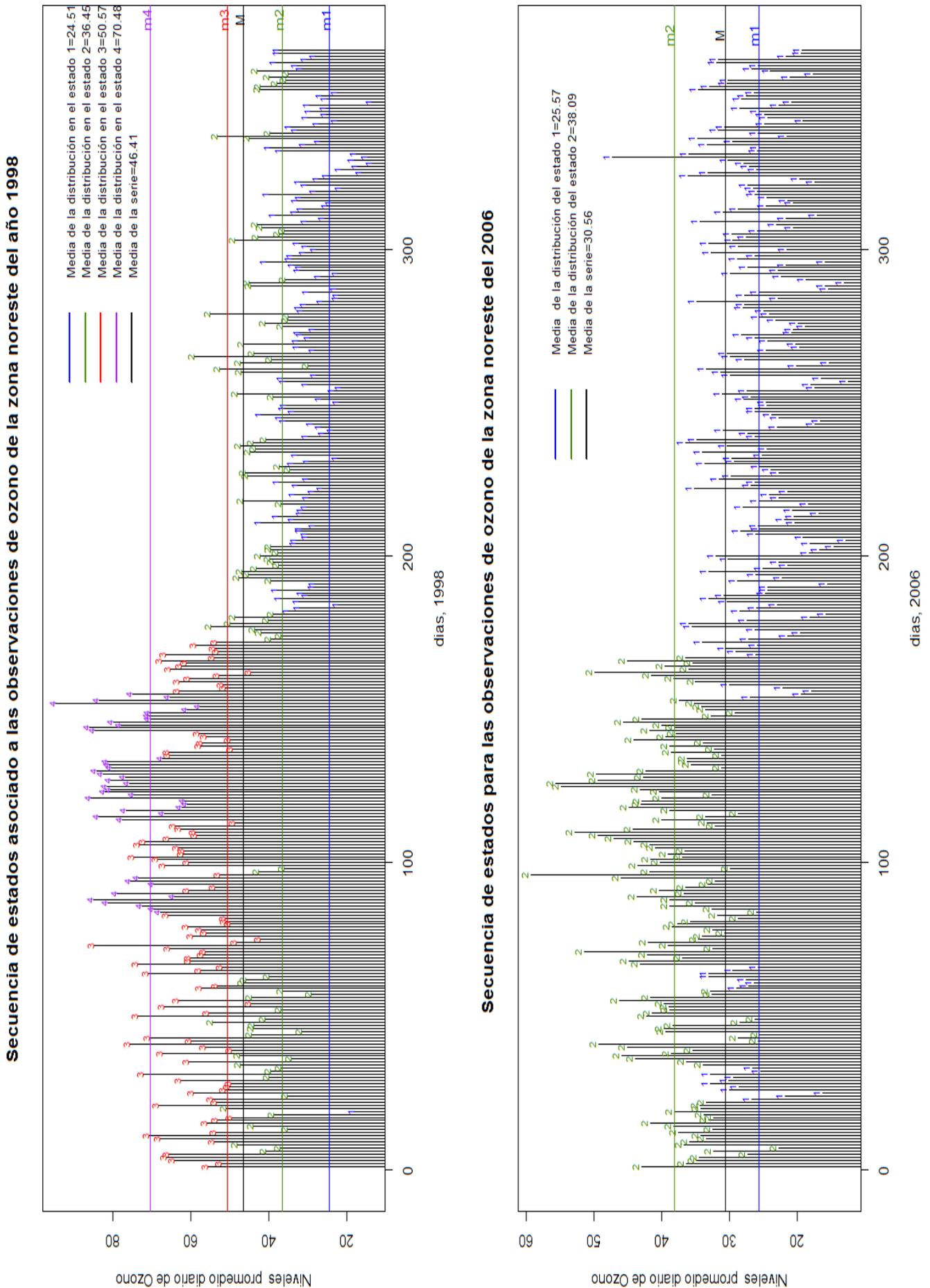
Cuadro 2.4: HMM-4 estados- distribución gamma, zona noreste, año 1998. De un lado, los parámetros; del otro, en azul la función de densidad asociada al estado 1, de verde la función de densidad asociada al estado 2, de rojo la función de densidad asociada al estado 3, de morado la función de densidad asociada al estado 4 y de negro la mezcla de funciones.

HMM 2 estados-distribución gamma.

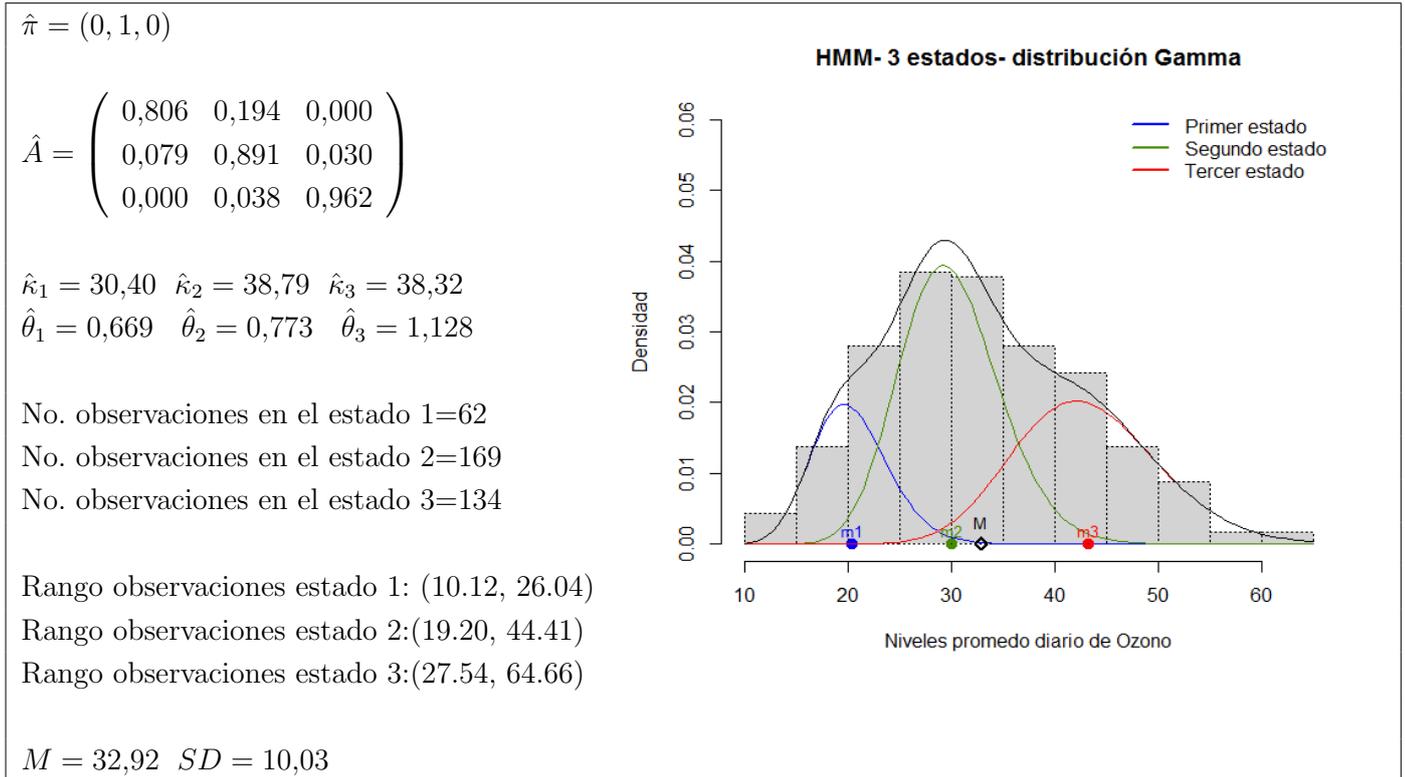


Cuadro 2.5: HMM-2 estados- distribución gamma, zona noreste, año 2006. De un lado, los parámetros del modelo; del otro, en color azul la función de densidad asociada al estado 1, de color verde la función de densidad asociada al estado 2 y de negro la función de densidad asociada a la mezcla de densidades.

Figura 2.7: En la parte superior, secuencias de estados de la serie de 1998 en la zona noreste, se obtuvo de un HMM con 4 estados-distribución gamma. En la parte inferior, la secuencia de estados de la serie del 2006, misma zona, se obtuvo con un HMM con 2 estados-distribución gamma.



HMM 3 estados- distribución gamma.



Cuadro 2.6: HMM-3 estados- distribución gamma, zona noreste, año 2014. De un lado, los parámetros del modelo; del otro, en azul la función de densidad del estado 1, en verde la función de densidad asociada al estado 2, de rojo la función de densidad asociada al estado 3 y de negro la mezcla de densidades.

**Zona noroeste**

Para la zona noroeste, de forma más breve, se presentan los modelos de los años donde mejor ajustó un HMM. Se trata de un HMM con 3 estados y distribución gaussiana para el año 2011 y un HMM con 3 estados y distribución gamma en el año 2012.

HMM- 3 estados- distribución gaussiana, año 2011. En el cuadro 2.7 se muestra a detalle los parámetros finales del HMM ajustado para el año 2011; donde el vector de probabilidades inicial señala que el estado más probable para iniciar la cadena de Markov, es el estado 1. Además, en el régimen de contaminación determinado por el estado 2, se encuentra un mayor número de observaciones y en el régimen determinado por el estado 3, se encuentran un menor número de observaciones, esto se atribuye a que en ese año fueron menos los días en que se observaron promedios diarios muy altos de ozono. Del estimador de la matriz de probabilidades de transición, se observa que transiciones entre los estados 3 a 1 y del estado 3 a 2 y viceversa, tienen poca probabilidad de ocurrir.

HMM- 3 estados- distribución gamma, año 2012. También, en el Cuadro 2.7 se presentan los parámetros finales que se obtuvieron del modelo ajustado en el año 2012, un HMM con 3 estados y distribución gamma. Al igual que en el modelo anterior, es más probable que la cadena de Markov inicie en el estado 1, además de la matriz de probabilidades de transición se espera que la cadena de Markov permanezca varios días en el mismo estado, y que transiciones del estado 3 al 1 se observen muy poco.

Zona noroeste

HMM-3 estados-distribución gaussiana, año 2011	HMM- 3 estados- distribución gamma, año 2012
$\hat{\pi} = (1, 0, 0)$	$\hat{\pi} = (1, 0, 0)$
$\hat{A} = \begin{pmatrix} 0,927 & 0,073 & 0,000 \\ 0,047 & 0,946 & 0,006 \\ 0,012 & 0,000 & 0,988 \end{pmatrix}$	$\hat{A} = \begin{pmatrix} 0,849 & 0,151 & 0,000 \\ 0,091 & 0,895 & 0,014 \\ 0,000 & 0,035 & 0,965 \end{pmatrix}$
$\hat{\mu}_1 = 21,81 \quad \hat{\mu}_2 = 29,41 \quad \hat{\mu}_3 = 39,53$ $\hat{\sigma}_1 = 5,53 \quad \hat{\sigma}_2 = 4,99 \quad \hat{\sigma}_3 = 6,07$	$\hat{\kappa}_1 = 17,88 \quad \hat{\kappa}_2 = 32,71 \quad \hat{\kappa}_3 = 53,09$ $\hat{\theta}_1 = 1,019 \quad \hat{\theta}_2 = 0,867 \quad \hat{\theta}_3 = 0,735$
No. obs. en el estado 1=121 No. obs. en el estado 2=160 No. obs. en el estado 3=84	No. obs. en el estado 1=108 No. obs. en el estado 2=186 No. obs. en el estado 3=72
Rango obs. estado 1: (3.66,35.37) Rango obs. estado 2: (17.83,47.29) Rango obs. estado 3: (24.91,53.58)	Rango obs. estado 1: (5.62, 27.70) Rango obs. estado 2: (17.87, 42.62) Rango obs. estado 3: (26.95, 51.37)
$M = 29,2 \quad SD = 8,45$	$M = 27,6 \quad SD = 8,84$

Cuadro 2.7: Parámetros de los modelos HMM ajustados en la zona noroeste, en los años 2011, 2012 con distribución gaussiana y gamma, respectivamente.

**Zona centro**

Con respecto a la zona centro, se muestra los resultados obtenidos en 4 años; a decir, los modelos ajustados en 2006, 2011, 2012 y 2014. Detalles al respecto se muestran en el Cuadro 2.8.

HMM- 2 estados- distribución gaussiana, año 2006. En la parte superior derecha del Cuadro 2.8 se muestra los parámetros de un HMM con 2 estados y distribución gaussiana, ajustado con datos del año 2006. En esta ocasión la cadena de Markov inició con el estado 2, es poco frecuente que esto se observe; ya que es en el régimen de contaminación determinado por el estado 1 donde se encuentra el 57 por ciento de las observaciones registradas. El rango de los niveles promedio diario de ozono que se agruparon en el estado 1 va de 3.95 a 35.20; y el rango en el estado 2 va de 14.58 a 50.37, por lo que hay un importante traslape de las funciones de densidad. Cabe mencionar, que en este año el nivel medio de contaminación ambiental fue de  $M = 25,03$ , bajo comparado con el que se reportó en ese mismo año en la zona noreste.

HMM- 3 estados- distribución gaussiana, año 2011. En la parte superior izquierda del Cuadro 2.8 se muestra los principales resultados del modelo HMM con 3 estados y distribución gaussiana para el año 2011. Los aspectos observados son muy similares a los HMM de 3 estados ajustados para la zona noreste. Entre ellos, persistencia de la cadena de Markov en un estado por varios días, poca o nula transición entre los estados 3 y 1. Asimismo, se observa un nivel medio de contaminación ambiental menor con respecto a la zona noroeste.

HMM- 3 estados- distribución gamma, año 2012. En la parte inferior izquierda del Cuadro 2.8 están los parámetros del modelo ajustado en el año 2012; se trata de un HMM con 3 estados y distribución gamma. Al respecto, en el estado 1, se agruparon 74 observaciones, donde la distribución tiene una media de 12.83; en el estado 2 se agruparon 191 observaciones, donde la media de la distribución

asociada a este estado es de 22.60; y por último, en el estado 3 se agruparon 101 observaciones, donde la media de la distribución tiene un valor de 35.18.

HMM- 3 estados- distribución gamma, año 2015. Por último, en la parte inferior derecha del Cuadro 2.8 se presentan los resultados del modelo ajustado en el año 2015. Bajo un HMM con 3 estados y distribución gamma, el estado 1 de la cadena de Markov agrupó 123 observaciones y la distribución asociada a este estado tiene media igual a 14.61; en el estado 2 se agruparon 148 observaciones y la distribución tiene media igual a 23.393, por último el estado 3 con 94 observaciones tiene media en su distribución igual a 33.53. Cabe mencionar que el valor medio de la contaminación ambiental para este año fue de  $M = 14,65$ , un valor muy bajo con respecto a otros años y otras zonas.

Zona centro

<p>HMM- 3 estados- distribución gaussiana, año 2011  <math>\hat{\pi} = (1, 0, 0)</math></p> $\hat{A} = \begin{pmatrix} 0,947 & 0,053 & 0,000 \\ 0,078 & 0,905 & 0,017 \\ 0,000 & 0,023 & 0,977 \end{pmatrix}$ <p><math>\hat{\mu}_1 = 18,35 \quad \hat{\mu}_2 = 26,82 \quad \hat{\mu}_3 = 39,90</math>  <math>\hat{\sigma}_1 = 4,96 \quad \hat{\sigma}_2 = 4,39 \quad \hat{\sigma}_3 = 7,05</math></p> <p>No. obs. en el estado 1=164            No. obs. en el estado 2=113            No. obs. en el estado 3=88</p> <p>Rango obs. estado 1: (1.66,29.54)            Rango obs. estado 2: (11.50,37.20)            Rango obs. estado 3: (21.47,58.58)</p> <p><math>M = 25,85 \quad SD = 9,95</math></p>	<p>HMM- 2 estados-distribución gaussiana, año 2006  <math>\hat{\pi} = (0, 1)</math></p> $\hat{A} = \begin{pmatrix} 0,994 & 0,006 \\ 0,014 & 0,986 \end{pmatrix}$ <p><math>\hat{\mu}_1 = 18,98 \quad \hat{\mu}_2 = 32,34</math>  <math>\hat{\sigma}_1 = 6,37 \quad \hat{\sigma}_2 = 6,01</math></p> <p>No. obs. en el estado 1=208            No. obs. en el estado 2=157</p> <p>Rango obs. estado 1: (3.95,35.20)            Rango obs. estado 2: (14.58,50.37)</p> <p><math>M = 25,03 \quad SD = 9,10</math></p>
<p>HMM- 3 estados- distribución gamma, año 2012  <math>\hat{\pi} = (1, 0, 0)</math></p> $\hat{A} = \begin{pmatrix} 0,801 & 0,199 & 0,000 \\ 0,082 & 0,87 & 0,048 \\ 0,000 & 0,082 & 0,918 \end{pmatrix}$ <p><math>\hat{\kappa}_1 = 11,56 \quad \hat{\kappa}_2 = 23,45 \quad \hat{\kappa}_3 = 36,46</math>  <math>\hat{\theta}_1 = 1,11 \quad \hat{\theta}_2 = 0,964 \quad \hat{\theta}_3 = 0,965</math></p> <p>No. obs. en el estado 1=74            No. obs. en el estado 2=191            No. obs. en el estado 3=101</p> <p>Rango obs. estado 1: (3.33, 25.75)            Rango obs. estado 2: (11.41, 39.41)            Rango obs. estado 3: (26.70, 52.37)</p> <p><math>M = 24,26 \quad SD = 9,37</math></p>	<p>HMM- 3 estados- distribución gamma, año 2015  <math>\hat{\pi} = (1, 0, 0)</math></p> $\hat{A} = \begin{pmatrix} 0,781 & 0,219 & 0,00 \\ 0,159 & 0,673 & 0,168 \\ 0,022 & 0,253 & 0,725 \end{pmatrix}$ <p><math>\hat{\kappa}_1 = 15,91 \quad \hat{\kappa}_2 = 31,40 \quad \hat{\kappa}_3 = 33,20</math>  <math>\hat{\theta}_1 = 0,919 \quad \hat{\theta}_2 = 0,745 \quad \hat{\theta}_3 = 1,01</math></p> <p>No. obs. en el estado 1= 123            No. obs. en el estado 2=148            No. obs. en el estado 3=94</p> <p>Rango obs. estado 1: (5.66, 24.37)            Rango obs. estado 2: (14.37, 32.79)            Rango obs. estado 3: (22.70, 50.04)</p> <p><math>M = 14,65 \quad SD = 9,86</math></p>

Cuadro 2.8: Parámetros de los modelos HMM ajustados en la zona centro en los años 2006, 2011, 2012 y 2015.

### Zona suroeste

Se presentan resultados para los años 1992, 2012 y 2014; debido a que para los años 2012 y 2014, el criterio de información bayesiana es muy cercano tanto para el HMM con observaciones gaussianas como para el HMM con observaciones gamma, se incluyó ambos modelos; esto permite evaluar aspectos adicionales.

HMM- 3 estados- distribución gaussiana, año 1992. En la parte superior izquierda del Cuadro 2.9 se puede observar el modelo HMM con 3 estados y distribución gaussiana ajustado en los niveles promedio diario de ozono del año 1992. A lo largo de este año se observaron los niveles más altos de ozono promedio diario en esta zona; así la media de la distribución asociada al estado 1 es igual a  $m_1 = 43,84$ ; la media de la distribución en el estado 2 es de  $m_2 = 73,49$  y por último la media en el estado 3 es de  $m_3 = 104,03$ . También se puede leer que ocurrieron valores extremos de niveles de ozono ya que el rango de valores que ocurrieron en el estado 3 va de 81.93 a 138.60. Por último ha de notarse el alto valor en la contaminación promedio anual de  $M = 73,54$ .

HMM- 3 estados- distribución gaussiana y gamma, año 2012. En el Cuadro 2.9 se observan los resultados de los modelos HMM con 3 estados y observaciones gamma y gaussiana, respectivamente. En ambos modelos, la matriz de transición señala la poca probabilidad de pasar del estado 3 al 1 y viceversa, además los valores de ambas matrices son parecidos, por lo que la secuencia de estados debe ser casi idéntica. En el modelo con observaciones de tipo gamma la media de la distribución asociada al estado 1 tiene un valor de 21.19 y en el modelo con observaciones de tipo gaussiana es de 22.82. Se puede ver que el rango de las observaciones que se agruparon en cada uno de los estados es casi idéntico en ambos modelos, por ejemplo en el estado 1, el rango bajo el HMM con observaciones gaussianas es de 4.54 a 35.79, mientras que bajo el HMM con observaciones gamma es de 4.54 a 33.75. Asimismo, ambos modelos estiman el mismo valor para la contaminación media ambiental anual de  $M = 34,5$  IMECAS. Lo anterior nos lleva a pensar que existe una consistencia en los resultados bajo ambos modelos.

HMM- 3 estados- distribución gaussiana y gamma, año 2014. Por último, en la zona noroeste, para el año 2014 se ajustaron los dos modelos; el HMM con 3 estados y observaciones gaussianas y gamma, ya que al igual que en el párrafo anterior, el criterio de información bayesiana muestra una diferencia mínima. En ambos modelos, es muy probable que la secuencia de estados de la cadena de Markov empiece en el estado 1; asimismo, como las matrices de probabilidad de transición tienen valores semejantes, se esperará que la secuencia de estados de la cadena de Markov sea muy parecida. Lo anterior, puede confirmarse con la Figura 2.8, donde aparece la secuencia de estados en ambos modelos, únicamente se observa una sutil diferencia, ya que en el HMM con observaciones gaussianas se visita más el estado 1, mientras que en el HMM con observaciones gamma se visita más el estado 3.

### Zona sureste

En este apartado se describe los modelos ajustados en los años 2006, 2007, 2011 y 2014 para la zona sureste.

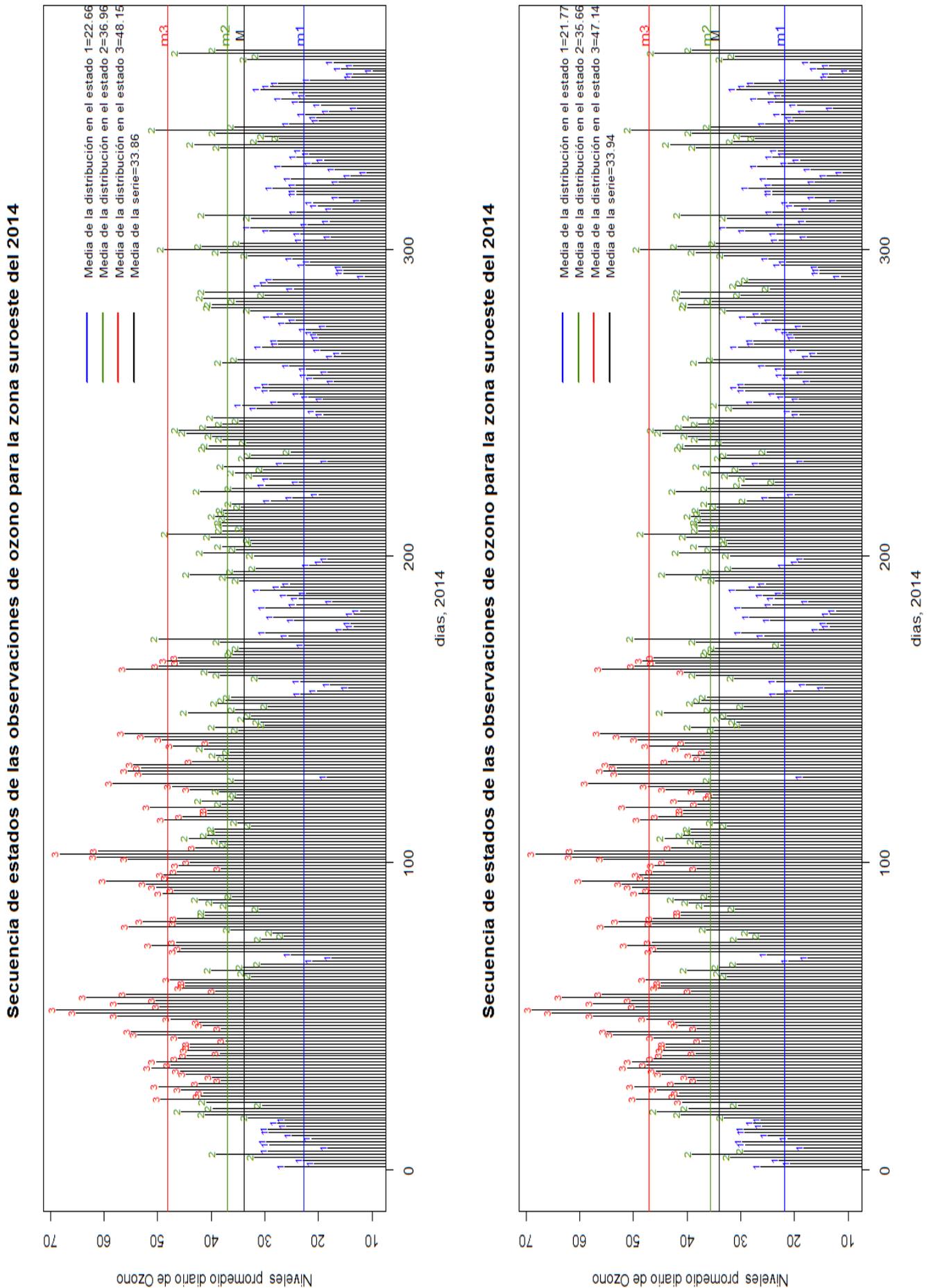
HMM- 3 estados- distribución gamma, año 2006. En la parte inferior izquierda del Cuadro 2.10 se puede ver los parámetros del modelo ajustado a los datos del año 2006. Si se observa la diagonal de la matriz de probabilidad de transición se puede decir que, si bien la cadena de Markov permanece

Zona suroeste

HMM- 3 estados- distribución gaussiana, año 1992	HMM- 3 estados- distribución gaussiana, año 2012	HMM-3 estados- distribución gaussiana, año 2014
$\hat{\pi} = (0, 0, 1)$  $\hat{A} = \begin{pmatrix} 0,822 & 0,153 & 0,025 \\ 0,081 & 0,755 & 0,164 \\ 0,035 & 0,333 & 0,632 \end{pmatrix}$  $\hat{\mu}_1 = 43,84 \quad \hat{\mu}_2 = 73,49 \quad \hat{\mu}_3 = 104,03$ $\hat{\sigma}_1 = 15,36 \quad \hat{\sigma}_2 = 10,96 \quad \hat{\sigma}_3 = 13,92$  No. obs. en el estado 1=70 No. obs. en el estado 2=147 No. obs. en el estado 3=64  Rango obs. estado 1: (17.46,76.00) Rango obs. estado 2: (50.00,97.71) Rango obs. estado 3:(81.93,138.60)  $M = 73,54 \quad SD = 24,80$	$\hat{\pi} = (1, 0, 0)$  $\hat{A} = \begin{pmatrix} 0,839 & 0,153 & 0,008 \\ 0,134 & 0,816 & 0,050 \\ 0,000 & 0,090 & 0,910 \end{pmatrix}$  $\hat{\mu}_1 = 22,82 \quad \hat{\mu}_2 = 35,17 \quad \hat{\mu}_3 = 49,09$ $\hat{\sigma}_1 = 5,66 \quad \hat{\sigma}_2 = 5,05 \quad \hat{\sigma}_3 = 7,43$  No. obs. en el estado 1=121 No. obs. en el estado 2=154 No. obs. en el estado 3=91  Rango obs. estado 1: (4.54,35.79) Rango obs. estado 2: (23.95,48.66) Rango obs. estado 3:(35.00,71.12)  $M = 34,50 \quad SD = 11,74$	$\hat{\pi} = (1, 0, 0)$  $\hat{A} = \begin{pmatrix} 0,876 & 0,111 & 0,013 \\ 0,137 & 0,782 & 0,082 \\ 0,000 & 0,134 & 0,866 \end{pmatrix}$  $\hat{\mu}_1 = 22,66 \quad \hat{\mu}_2 = 36,98 \quad \hat{\mu}_3 = 48,15$ $\hat{\sigma}_1 = 5,59 \quad \hat{\sigma}_2 = 4,96 \quad \hat{\sigma}_3 = 7,14$  No. obs. en el estado 1=145 No. obs. en el estado 2=133 No. obs. en el estado 3=87  Rango obs. estado 1: (9.83,34.20) Rango obs. estado 2: (25.16, 50.29) Rango obs. estado 3:(37.33,68.91)  $M = 33,86 \quad SD = 11,79$
HMM- 3 estados distribución gamma, año 2012 $\hat{\pi} = (1, 0, 0)$  $\hat{A} = \begin{pmatrix} 0,803 & 0,197 & 0,000 \\ 0,105 & 0,846 & 0,049 \\ 0,000 & 0,087 & 0,913 \end{pmatrix}$  $\hat{\kappa}_1 = 19,83 \quad \hat{\kappa}_2 = 39,30 \quad \hat{\kappa}_3 = 44,47$ $\hat{\theta}_1 = 1,069 \quad \hat{\theta}_2 = 0,862 \quad \hat{\theta}_3 = 1,096$  No. obs. en el estado 1=89 No. obs. en el estado 2=183 No. obs. en el estado 3=94  Rango obs. estado 1: (4.54, 33.75) Rango obs. estado 2: (22.04, 48.66) Rango obs. estado 3: (35.00, 71.12)  $M = 34,53 \quad SD = 11,66$	HMM-3 estados- distribución gamma, año 2014 $\hat{\pi} = (1, 0, 0)$  $\hat{A} = \begin{pmatrix} 0,851 & 0,135 & 0,014 \\ 0,138 & 0,800 & 0,062 \\ 0,000 & 0,096 & 0,904 \end{pmatrix}$  $\hat{\kappa}_1 = 17,81 \quad \hat{\kappa}_2 = 45,31 \quad \hat{\kappa}_3 = 43,08$ $\hat{\theta}_1 = 1,22 \quad \hat{\theta}_2 = 0,786 \quad \hat{\theta}_3 = 1,094$  No. obs. en el estado 1=133 No. obs. en el estado 2=132 No. obs. en el estado 3=100  Rango obs. estado 1: (9.83,32.62) Rango obs. estado 2: (22.62, 50.29) Rango obs. estado 3: (35.16,68.91)  $M = 33,94 \quad SD = 11,67$	

Cuadro 2.9: Parámetros obtenidos de algunos modelos ajustados en la zona suroeste en los años 1992, 2012 y 2014.

Figura 2.8: En la parte superior, secuencias de estados de la serie de 2014 en la zona suroeste, se obtuvo de un HMM con 3 estados-distribución gaussiana. En la parte inferior, la secuencia de estados de la serie del 2014, misma zona, se obtuvo con un HMM con 3 estados-distribución gamma.



varios días en cada estado, se espera que el número de días en el estado 1 sea menor que los días que permanece en los estados 2 y 3. Por otra parte, la media de la distribución en el estado 1, que corresponde a la contaminación de fondo, es igual a 20.40 IMECAS y la media de la contaminación ambiental promedio es de  $M = 32,67$  por lo que la contaminación de fondo representa un 62.4 por ciento con respecto a la contaminación ambiental.

HMM- 2 estados- distribución gaussiana, año 2007. En la parte superior izquierda del Cuadro 2.10 se encuentran los resultados del modelo HMM con 2 estados y distribución gaussiana ajustado a los niveles de ozono promedio diario del año 2007. Al igual que en la mayoría de los modelos ya presentados, en la diagonal de la matriz de probabilidades de transición hay valores cercanos a 1, por lo que es muy probable que las observaciones que suceden en el tiempo permanezcan en el mismo estado o régimen. En este modelo la media de la contaminación de fondo  $m_1 = 25,88$  mientras que la media de la contaminación ambiental es  $M = 33,12$ , el cociente entre ellas es  $\frac{m_1}{M} = 0,78$  por lo que la contaminación de fondo representa un 78 por ciento con respecto a la contaminación ambiental.

HMM- 3 estados- distribución gaussiana, año 2011. En la parte superior derecha del Cuadro 2.10 se pueden ver los parámetros del modelo ajustado a la zona sureste, en el año 2011. En este modelo, el mayor número de observaciones se agruparon en el primer régimen que tiene una media de  $m_1 = 26,39$ , en segundo lugar las observaciones se agrupan en el estado 2 con una media de  $m_2 = 36,29$ , valor cercano a la contaminación media anual de  $M = 35,79$ .

HMM- 3 estados- distribución gamma, año 2014. El último modelo que se presenta es un HMM con 3 estados y distribución gamma, parte inferior derecha del Cuadro 2.10. Nuevamente, los valores en la diagonal de la matriz de transiciones son cercanos a 1, siendo mayor para el estado 3, por lo que se espera que las observaciones a lo largo del tiempo permanezcan en este régimen, de hecho la mayoría de las observaciones, 151 de 365 se agruparon en este estado. Del rango de las observaciones en el estado 2 y 3 se puede ver que las funciones de densidad asociadas a cada uno de ellos se traslapan mucho más de lo que se ha observado en otros años y zonas.

### Relación entre la contaminación de fondo y contaminación ambiental

Uno de los aspectos que resultan de interés, es analizar la relación entre la contaminación de fondo y la contaminación ambiente; en particular, la razón entre la media de la contaminación ambiental  $M$  con respecto a la media del régimen de contaminación de fondo, o del estado 1,  $m_1$ ; en otras palabras  $\frac{M}{m_1}$ . Los valores de esta razón para la zona noroeste va de 0.99 a 1.90, en la zona noreste va de 1.19 a 1.89, en el centro de 1 a 1.90, en el suroeste de 1.24 a 2.07, y en el sureste de 0.85 a 2.15; cuando esta razón se acerca a 1 significa que la contaminación en ese año y zona se explica casi en su totalidad por la contaminación de fondo. En los Cuadros 2.11 y 2.12 se muestra, por zonas, el promedio de la contaminación media ambiental  $M$ , de la contaminación de fondo  $m_1$ , el promedio de la diferencia entre la contaminación ambiental y de fondo  $M - m_1$ ; así como el porcentaje que representa la magnitud de la media de contaminación de fondo con respecto a la contaminación ambiental. Cabe mencionar, que en la zona suroeste la contaminación de fondo tiene una menor relación con la contaminación ambiental, esto para los HMM con observaciones gaussiana y gamma.

También resulta de interés conocer si existen diferencias significativas en la estimación de la razón  $\frac{M}{m_1}$  cuando se emplea un HMM con distribución gaussiana o bien un HMM con distribución gamma. Se ha

Zona sureste

<p>HMM-2 estados- distribución gaussiana, año 2007  <math>\hat{\pi} = (1, 0)</math></p> $\hat{A} = \begin{pmatrix} 0,925 & 0,075 \\ 0,046 & 0,954 \end{pmatrix}$ <p><math>\hat{\mu}_1 = 25,88 \quad \hat{\mu}_2 = 38,08</math>  <math>\hat{\sigma}_1 = 6,48 \quad \hat{\sigma}_2 = 6,65</math></p> <p>No. obs. en el estado 1=142            No. obs. en el estado 2=223</p> <p>Rango obs. estado 1: (10.20,40.91)            Rango obs. estado 2: (21.16,63.79)</p> <p><math>M = 33,12 \quad SD = 8,90</math></p>	<p>HMM- 3 estados- distribución gaussiana, año 2011.  <math>\hat{\pi} = (1, 0, 0)</math></p> $\hat{A} = \begin{pmatrix} 0,880 & 0,120 & 0,000 \\ 0,116 & 0,844 & 0,040 \\ 0,000 & 0,071 & 0,929 \end{pmatrix}$ <p><math>\hat{\mu}_1 = 26,39 \quad \hat{\mu}_2 = 36,29 \quad \hat{\mu}_3 = 51,84</math>  <math>\hat{\sigma}_1 = 5,71 \quad \hat{\sigma}_2 = 4,80 \quad \hat{\sigma}_3 = 6,23</math></p> <p>No. obs. en el estado 1=143            No. obs. en el estado 2=140            No. obs. en el estado 3=82</p> <p>Rango obs. estado 1: (9.04, 37.83)            Rango obs. estado 2: (28.20, 48.33)            Rango obs. estado 3: (35.20, 67.79)</p> <p><math>M = 35,79 \quad SD = 10,91</math></p>
<p>HMM-3 estados- distribución gamma, año 2006  <math>\hat{\pi} = (0, 0, 1)</math></p> $\hat{A} = \begin{pmatrix} 0,632 & 0,368 & 0,000 \\ 0,118 & 0,845 & 0,037 \\ 0,000 & 0,073 & 0,927 \end{pmatrix}$ <p><math>\hat{\kappa}_1 = 29,40 \quad \hat{\kappa}_2 = 40,12 \quad \hat{\kappa}_3 = 43,87</math>  <math>\hat{\theta}_1 = 0,694 \quad \hat{\theta}_2 = 0,774 \quad \hat{\theta}_3 = 0,957</math></p> <p>No. obs. en el estado 1=50            No. obs. en el estado 2=202            No. obs. en el estado 3=113</p> <p>Rango obs. estado 1: (9.79, 25.00)            Rango obs. estado 2: (19.29, 43.70)            Rango obs. estado 3: (28.83, 66.20)</p> <p><math>M = 32,67 \quad SD = 8,85</math></p>	<p>HMM- 3 estados- distribución gamma, año 2014  <math>\hat{\pi} = (1, 0, 0)</math></p> $\hat{A} = \begin{pmatrix} 0,809 & 0,191 & 0,000 \\ 0,140 & 0,769 & 0,091 \\ 0,000 & 0,073 & 0,927 \end{pmatrix}$ <p><math>\hat{\kappa}_1 = 28,05 \quad \hat{\kappa}_2 = 39,96 \quad \hat{\kappa}_3 = 43,56</math>  <math>\hat{\theta}_1 = 0,754 \quad \hat{\theta}_2 = 0,804 \quad \hat{\theta}_3 = 0,939</math></p> <p>No. obs. en el estado 1= 98            No. obs. en el estado 2= 116            No. obs. en el estado 3= 151</p> <p>Rango obs. estado 1: (9.91, 30.41)            Rango obs. estado 2: ( 21.54, 46.37)            Rango obs. estado 3: (22.54, 60.95)</p> <p><math>M = 32,87 \quad SD = 9,59</math></p>

Cuadro 2.10: Parámetros obtenidos de algunos modelos ajustados en la zona sureste, en los años 2006,2007, 2011 y 2014.

Estimación con HMM- distribución gamma.

Zona	M	$m_1$	$M - m_1$	$\frac{M}{m_1}$	%	Rango %
Noroeste	40.97	27.95	13.01	1.42	0.71	0.55-0.85
Noreste	38.44	26.28	12.16	1.41	0.72	0.54-0.84
Centro	37.45	23.83	13.62	1.49	0.69	0.53-1.00
Suroeste	47.80	28.88	18.93	1.59	0.64	0.52-0.80
Sureste	42.35	28.79	13.56	1.48	0.70	0.46-1.00

Cuadro 2.11: Relación de la contaminación de fondo y la contaminación ambiental, HMM- distribución gamma.

Estimación con HMM- distribución gaussiana. ]

Zona	$M$	$m_1$	$M - m_1$	$\frac{M}{m_1}$	%	Rango
Noroeste	40.98	30.20	10.78	1.37	0.73	0.56-0.88
Noreste	38.47	28.15	10.32	1.37	0.73	0.63-0.84
Centro	38.23	27.07	11.16	1.43	0.70	0.54-0.82
Suroeste	47.84	31.28	16.56	1.54	0.66	0.44-0.84
Sureste	42.93	30.82	12.11	1.42	0.71	0.50-0.93

Cuadro 2.12: Relación de la contaminación de fondo y la contaminación ambiental, HMM- distribución gaussiana.

evaluado esta idea a través de una prueba de medias pareadas; enseguida se muestran los resultados. Como puede observarse no existe diferencia significativa en la estimación de estas razones. En la parte A) del Cuadro 2.13 se muestra el estadístico de prueba para cada una de las zonas; la evidencia es más clara para las zonas noreste y suroeste.

Por último, se explora la idea de si la estimación de la contaminación ambiental promedio  $M$  es invariante ante el HMM con observaciones gaussianas o gamma. Mediante una prueba de medias pareadas se ha evaluado esta hipótesis. En la sección B) del Cuadro 2.13 se encuentran los resultados de la prueba por zonas. Como puede observarse, el modelo HMM estima de forma adecuada el valor medio de la contaminación ambiental aún cuando se cambie la función de densidad de las observaciones, esta evidencia es más clara para las zonas noroeste, suroeste y sureste.

A) Prueba sobre la contribución de la contaminación de fondo

Zona	Estadístico	P-value	Decisión
Noroeste	$t = 1.7664$	0.0906	No se rechaza $H_0$ .
Noreste	$t = 1.4809$	0.1522	No se rechaza $H_0$ .
Centro	$t = 1.8235$	0.0812	No se rechaza $H_0$ .
Suroeste	$t = 0.8196$	0.4208	No se rechaza $H_0$ .
Sureste	$t = 1.7845$	0.0875	No se rechaza $H_0$ .

B) Prueba sobre la contaminación ambiental media

Zona	Estadístico	P-Value	Decisión
Noroeste	$t = -0.2726$	0.7876	No se rechaza $H_0$ .
Noreste	$t = -1.9756$	0.0603	No se rechaza $H_0$ .
Centro	$t = -1.7426$	0.0947	No se rechaza $H_0$ .
Suroeste	$t = -1.0518$	0.3038	No se rechaza $H_0$ .
Sureste	$t = -1.0181$	0.3192	No se rechaza $H_0$ .

Cuadro 2.13: A) Prueba de medias pareadas con respecto a la contribución de la contaminación de fondo a la contaminación ambiental. Evalúa la hipótesis  $H_0$ : La contribución de la contaminación de fondo es la misma bajo el supuesto de un HMM con observaciones gamma y gaussianas.

B) Prueba de medias pareadas sobre la estimación de la contaminación media ambiental. Evalúa la hipótesis  $H_0$ : El nivel medio de contaminación ambiental es el mismo cuando se asume que los niveles promedio diario se distribuyen de acuerdo a un HMM con observaciones gaussianas y observaciones gamma.

En la parte superior de la Figura 2.9 se muestra la evolución de la razón  $\frac{M}{m_1}$  estimada con un HMM con observaciones gaussianas (en color rojo) y la razón estimada con un HMM con observaciones gamma (en color azul), para la zona noroeste. En la parte inferior de la Figura 2.9 se muestra la evolución de la razón  $\frac{M}{m_1}$  con los dos modelos, HMM con observaciones gaussianas y gamma, para la zona noreste.

En la parte superior de la Figura 2.10 se muestra la evolución de la razón  $\frac{M}{m_1}$  en la zona centro, entre los años 1992 a 2015, considerando los HMM con observaciones gaussianas y gamma. En la parte inferior de la Figura 2.10 se muestra la evolución de la razón  $\frac{M}{m_1}$  para la zona suroeste. Por último, en la parte superior de la Figura 2.11 se observa la evolución de la razón  $\frac{M}{m_1}$  para la zona sureste.

En la parte inferior de la Figura 2.11 se encuentra la evolución de la media de la contaminación ambiental  $M$ , estimada para la zona sureste, bajo un HMM con observaciones gaussianas y gamma. Mientras que en la Figura 2.12 se muestra la evolución de la contaminación ambiental media para las zonas noroeste y noreste. Finalmente, en la Figura 2.13 se encuentra la evolución de la contaminación ambiental media en las zonas centro y suroeste.

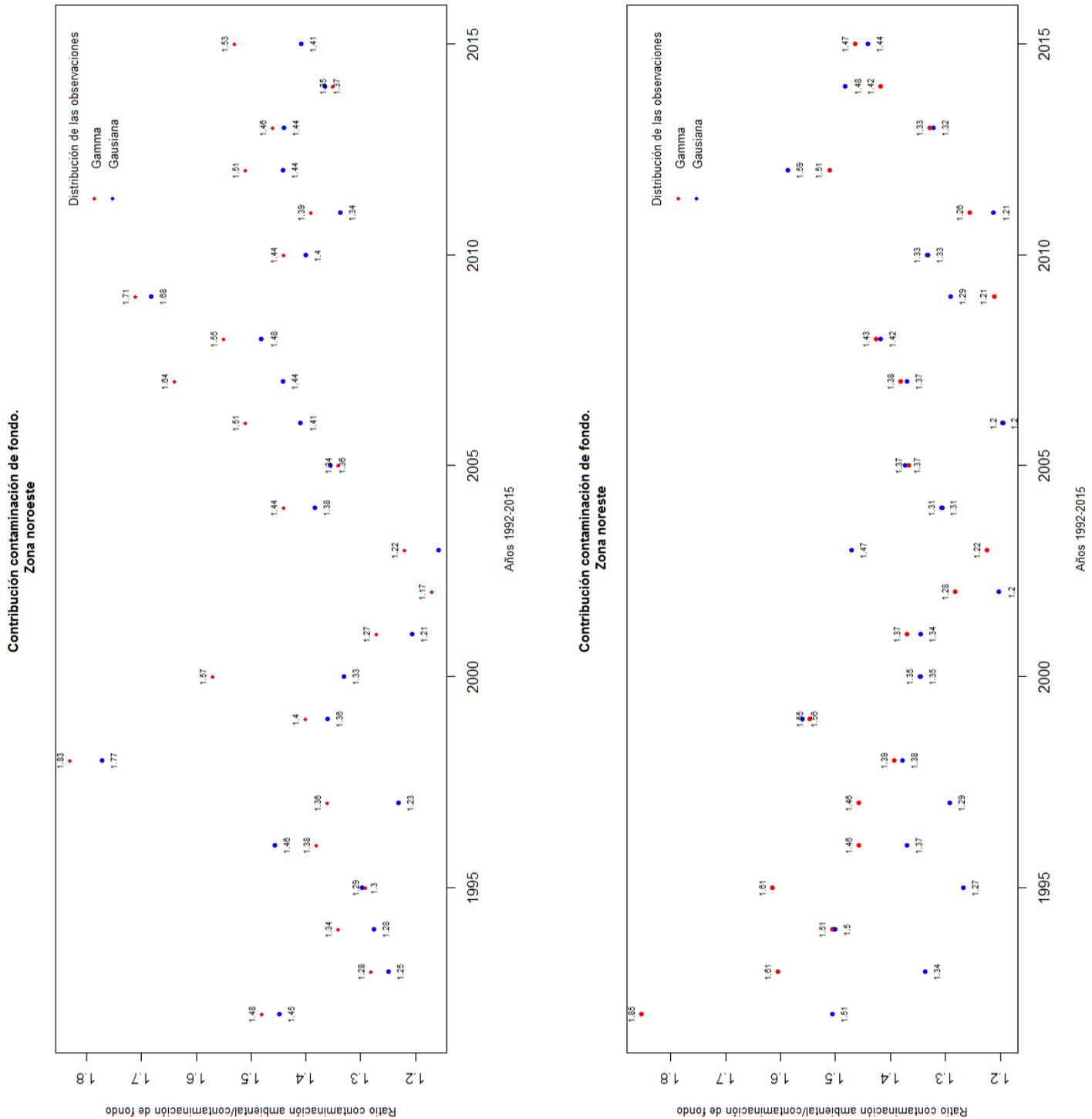


Figura 2.9: Contribución de la contaminación de fondo a la contaminación ambiental bajo los supuestos de un HMM con observaciones gaussianas y gamma. Evolución de 1992 a 2015 de las zonas noroeste y noreste.

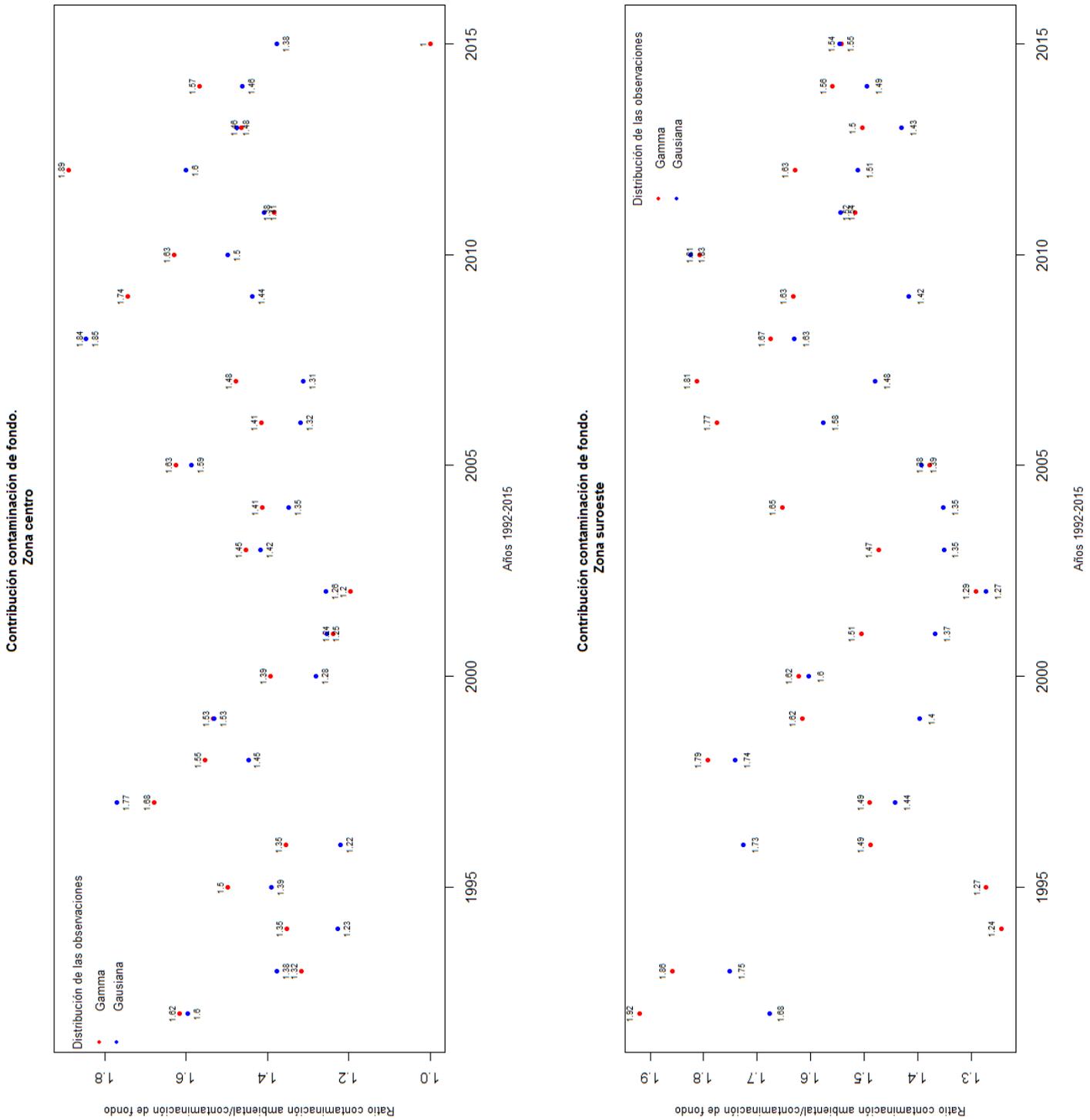


Figura 2.10: Contribución de la contaminación de fondo a la contaminación ambiental bajo los supuestos de un HMM con observaciones gaussianas y gamma. Evolución de 1992 a 2015 de las zonas centro y suroeste.

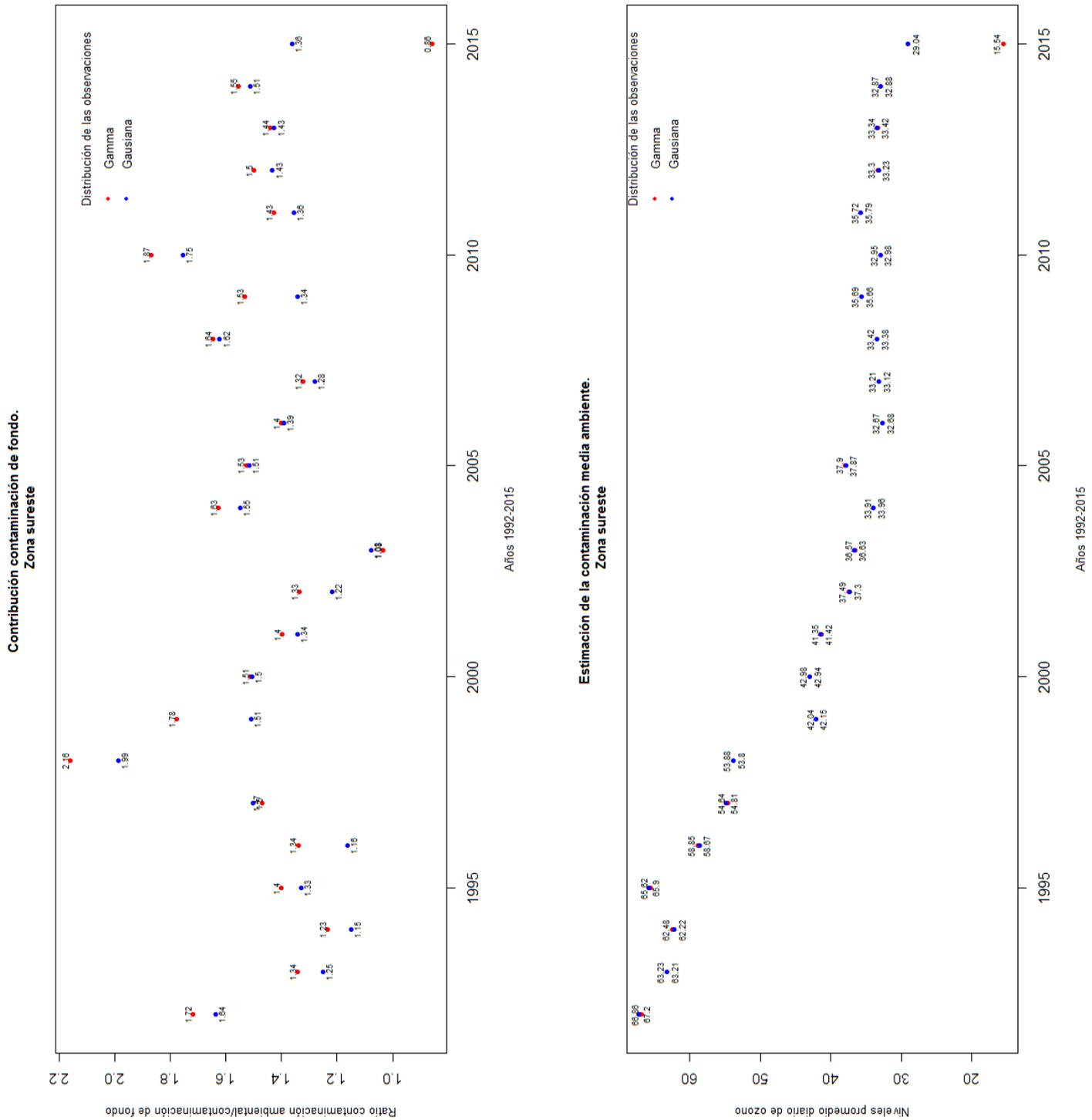


Figura 2.11: Contribución de la contaminación de fondo a la contaminación ambiental bajo los supuestos de un HMM con observaciones gaussianas y gamma. Evolución de 1992 a 2015 de la zona sureste. Evolución de la contaminación ambiental media bajo un HMM con observaciones gamma y gaussianas, zona sureste.

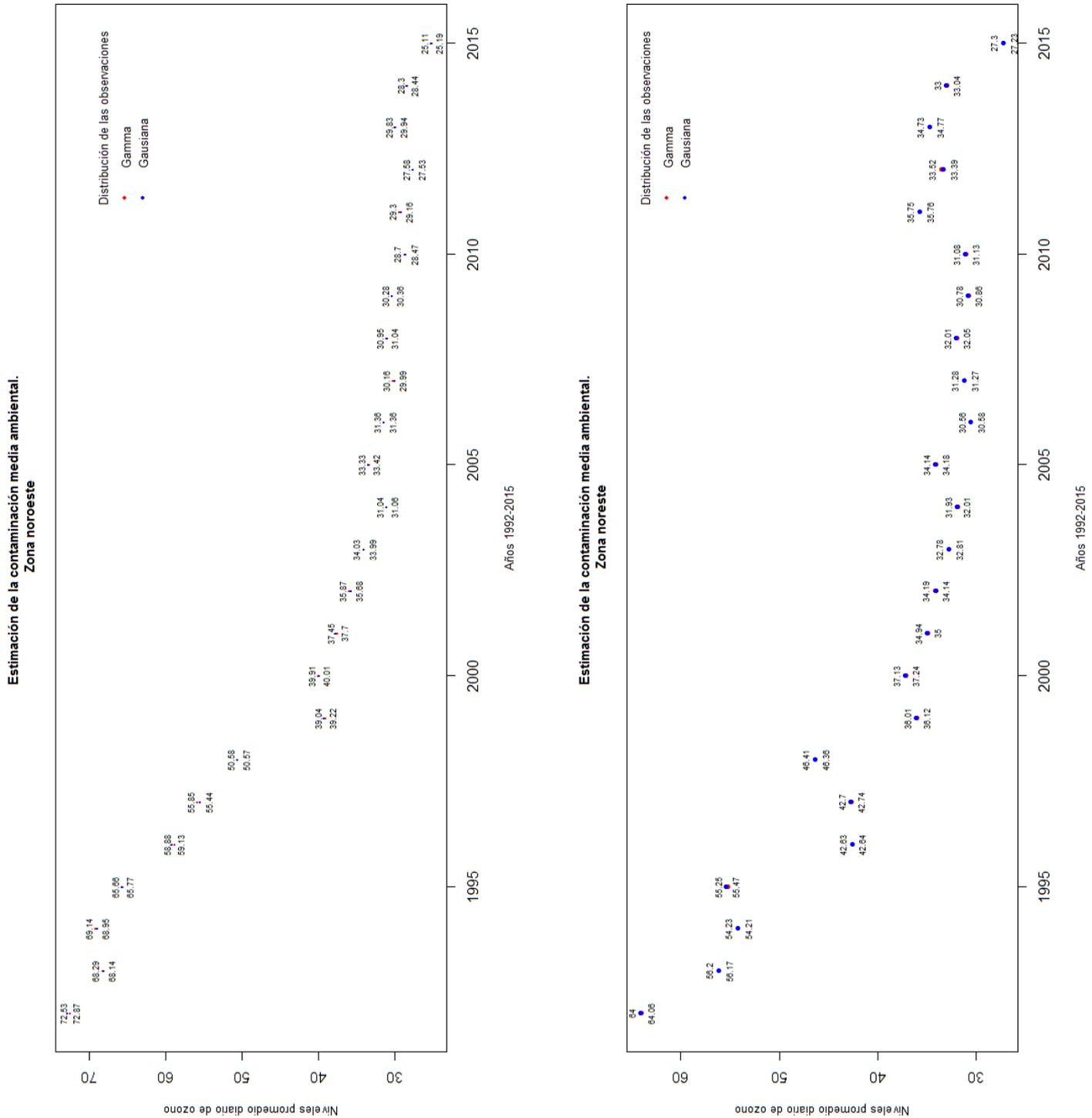


Figura 2.12: Evolución de la contaminación ambiental media bajo un HMM con observaciones gamma y gaussiana, zona noroeste y noreste.

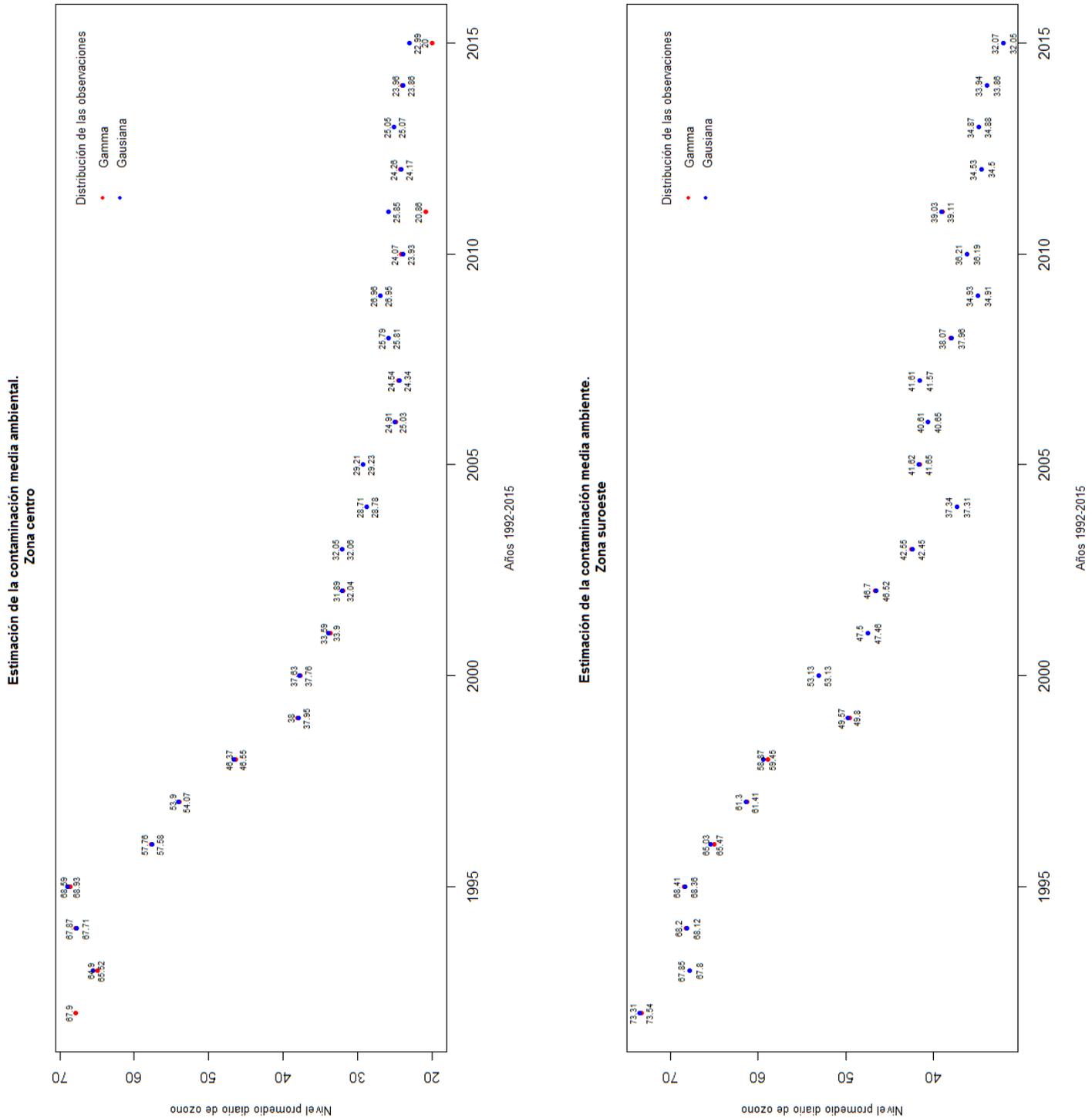


Figura 2.13: Evolución de la contaminación ambiental media bajo un HMM con observaciones gamma y gausiana, zona centro y suroeste.

# Capítulo 3

## Conclusiones

A lo largo de este documento se ha descrito la estructura matemática de los modelos de Markov oculto y se mostró el procedimiento teórico para obtener los estimadores del modelo. Asimismo, se ilustra la aplicación de los HMM para describir aspectos sobre los niveles promedio diario de ozono en cinco zonas de la Ciudad de México. Dadas las repercusiones de este contaminante en la salud humana, toda información adicional al respecto debe ser considerada al diseñar y evaluar políticas públicas con el objeto de mantener una buena calidad del aire.

Vale la pena comentar algunos de los principales aspectos observados en la aplicación de los HMM a la serie de ozono en la Ciudad de México:

- 1) Primero, la concentración de fondo representa en promedio entre un 64 y 72 por ciento de la contaminación del aire; por lo que es un indicador importante de la exposición real de las personas al ozono a lo largo de un año.
- 2) La estimación de la contaminación ambiente promedio anual,  $M$ , en las cinco zonas de la Ciudad de México, muestra que desde 1992 a la fecha, ésta presenta una reducción de cerca del 66.6 por ciento, por lo que los objetivos referentes a los límites permisibles de concentración de ozono en el aire ambiente así como los criterios para su evaluación, establecidos en la NORMA Oficial Mexicana NOM-020-SSA1-2014, Salud ambiental, han tenido resultados positivos en el largo plazo.
- 3) Si bien el criterio de información bayesiana nos permitió seleccionar el número de estados necesarios en el HMM, cabe mencionar que fueron necesarios dos y tres estados cuando los niveles promedio diario de ozono se mantuvieron en su mayoría entre (0, 63.63), es decir, en un nivel bueno; no obstante, fueron necesarios cuatro estados cuando a lo largo del año se reportaron, durante varios días, niveles de ozono que superaron este límite.
- 4) Por otro lado, la matriz de probabilidades de transición, que representa la parte dinámica del HMM, permitió conocer qué tan probable es que, tras haber observado una concentración de ozono asignada por el modelo a un estado o régimen en particular, la siguiente observación en la serie sea similar (permanecer en el mismo régimen de concentración) o no (pasar a otro estado). En el análisis presentado, a lo largo de los años y zonas, se puede ver que la matriz de probabilidades de transición estimadas, se caracterizan por tener valores cercanos a 1 en la diagonal, y valores cercanos a cero en las entradas por arriba y abajo de la diagonal. Lo anterior, significa que hay una estabilidad en la serie de tiempo, por lo que las observaciones consecutivas pueden permanecer en el mismo régimen por varios días y que es poco probable pasar de un régimen que contienen registros altos de ozono a uno con registros bajos en un sólo día.
- 5) En cuanto a la selección del mejor modelo, el criterio de información bayesiana, mostró que el HMM

con distribución gaussiana fué un mejor modelo en los primeros años, de 1992 al 2000; no obstante, en años posteriores se encontraron mejores resultados bajo un HMM con distribución gamma.

Por último, la naturaleza de los HMM permite considerarlos como una herramienta en el análisis de series de tiempo de contaminantes; las principales razones en las que se sustenta son:

- 1) Permite establecer una relación entre los conceptos de estados de una cadena de Markov y regímenes de contaminación (intervalos de niveles del contaminante en cuestión), por lo que pueden ser considerados una herramientas para el análisis de contribución de fuentes de contaminantes.
- 2) En especial, permite caracterizar a la contaminación de fondo asociada al primer estado de la cadena de Markov. La relevancia de la contaminación de fondo es que representa la concentración subyacente o umbral en la que no se esperan grandes cambios a lo largo del tiempo si las condiciones atmosféricas permanecen aproximadamente constantes, siendo una concentración característica del área de estudio; por consiguiente, puede considerarse un indicador de la exposición real promedio anual a la que se enfrenta la población.
- 3) Es posible conocer la contribución de cada régimen de contaminación respecto a la contaminación ambiente, esto a partir de la exploración de la relación entre la media de cada régimen  $m_i$  y la media anual,  $M$ , de la serie de datos del contaminante.
- 4) Se puede establecer la probabilidad de cambio entre los diferentes estados de la cadena de Markov, por consiguiente, la probabilidad de cambio entre los diferentes regímenes de contaminación.
- 5) Representan un análisis estadístico individual, que destaca por su sustento teórico sólido, que además cuenta con la facilidad de su interpretación y la reproducibilidad de sus resultados; así como el hecho de que la modelización puede llevarse a cabo empleando software gratuito como R.

Con respecto a los cuatro objetivos planteados en este documento; en el capítulo 1 se presentó de forma detallada y amigable la teoría concerniente a los HMM, mientras que en el capítulo 2 se describen los detalles en el uso de los HMM como una herramienta complementaria para el análisis de contribución de fuentes de contaminantes, en particular se analizaron aspectos relevantes de la contaminación de fondo del ozono en la Ciudad de México; cabe mencionar que para el cálculo de los resultados fue necesario implementar la solución de los tres problemas concernientes a un HMM en software R; por lo anterior, se considera que los objetivos propuestos se cumplieron satisfactoriamente.

Como trabajo futuro; por un lado, sería importante incorporar otras variables al modelo como temperatura, velocidad del viento, otras partículas, etc, y modelar a través de un HMM con distribución multivariable en las observaciones; esto podría ofrecer un enfoque diferente sobre las variables que influyen en la formación de ozono.

Aunado a lo anterior, este análisis puede realizarse para otros contaminantes en el aire como son dióxido de azufre, monóxido de carbono, dióxido de nitrógeno y partículas suspendidas,  $PM_{10}$ ,  $PM_{2.5}$ , lo que podría proporcionar más información sobre la contaminación de fondo en conjunto y sobre la relevancia de cada uno en el índice metropolitano de calidad del aire.

# Bibliografía

- [1] Altman, R.M (2004). “Assessing the Goodness-of-Fit of Hidden Markov Models”. *Biometrics*, Vol. 60(2), pp. 444-450.
- [2] Ash, R.B y Gardner,M.F. (1975) “Topics in Stochastic Processes”, Academic Press. pp. 1-49.
- [3] Aas, K., Eikvil, L., Huseby, R.B. (1999) “Applications of hidden Markov chains in imagen analysis”, *Pattern Recognition*, Vol. 32(4), pp.703-713.
- [4] Aspirot, L. y Bermole,P. (2004) “Ancho de banda efectivo para cadenas de Markov escondidas”, *Curso Performance de Redes de Telecomunicaciones*, dictado por el Instituto de Ingeniería Eléctrica de la Fac de Ing. de la Universidad de la República Uruguay. pp. 1-16.
- [5] Belis, C.A., Karagulian, F., Larsen, B.R., et al. (2013) “Critical review and meta-analysis of ambient particulate matter source apportionment using receptor models in Europe”. *Atmospheric Environment*, Vol. 69, pp. 94-108.
- [6] Belis, C.A., Larsen, B.R., Amato, F., et al. (2014) “European guide on air pollution source apportionment with receptor models”. Luxembourg. Joint Research Centre Reference Reports. EUR 26080 EN. pp.1-87.
- [7] Bilmes, J.A. (1998) “A gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models”, *International Computer Science Institute Berkeley CA*, pp. 1-13.
- [8] Cappé, O., Moulines, E. y Rydén, T. (2005) “Inference in hidden Markov models”. Springer, New York.
- [9] Chang T.Y Rudy S.J. (1993) “Ozone -precursor relationships: a modeling study of semiempirical relationships”, *Enviromental science and technology*. Vol. 27(10), pp. 2213-2219.
- [10] Cohen, A. (1998) “Hidden Markov Models in Biomedical Signal Processing”, *Processings of the 20th Annual International Conference of the Engineeering in Medicine and Biology Society*, Vol. 20(3), pp. 1145-1150 .
- [11] Comisión para la Cooperación Ambiental, CEC. (2008) “ Aire y Atmósfera. Ozono troposférico. El Mosaico de América del Norte panorama de los problemas ambientales más relevantes”. Pag. 1-4. <http://www3.cec.org/islandora/es/item/986-north-american-mosaic-overview-key-environmental-issues-es.pdf>

- [12] Carol J, Fiona E. Matthews. (2009) “Estimating Life Expectancy in Health and Health by Using a Hidden Markov Model”, *Journal of the Royal Statistical Society*, Vol 58(4), pp. 449-465.
- [13] Davis, D.D., y Orendavici, T. (2006) “Incidence of ozono symptoms on vegetation within National Wildlife Refuge in New Jersey, USA”. *Environmental Pollution*, Vol. 143(3), pp. 555-564.
- [14] Dempster, A.P., Laird, N.M., y Rubin, D.B. (1977) “Maximum-likelihood from incomplete data via the EM algorithm”, *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 39(1), pp. 1-38.
- [15] Durbin, R., Eddy, S., Krogh, A. y Mitchison, G. (1998) “Biological sequence analysis”, *Probabilistic models of proteins and nucleic acids*, Cambridge University Press: Cambridge.
- [16] Escobedo, F., y Chacalo, A. (2008) “Estimación preliminar de la descontaminación atmosférica por el arbolado urbano de la Ciudad de México”. *Interciencia*, Vol 33(1). pp 29.
- [17] Forney D. Jr. (1973) “The Viterbi algorithm”, *Proceedings of The IEEE*, Vol. 61(3), pp. 268-278.
- [18] Filleul, L., Cassadou, S., Médina, S., Fabres, P., Lefranc, A., Eilstein, D., Le Tertre, A., y Ledrans, M. (2006). “The relation between temperature, ozone, and mortality in nine French cities during the heat wave of 2003”. *Environmental Health Perspectives*, 114 (9), pp. 1344-1347.
- [19] Ghahramani, Z., y Jordan, M. (1995) “Learning from incomplete data”, *Technical Report AI Lab Memo No. 1509, CBCL paper No. 108, MIT AI Lab*.
- [20] Gomez-Losada, A., Pires, J.C.M., Pino-Mejías, R. (2015) “Time series clustering for estimating particulate matter contributions and its use in quantifying impacts from deserts”. *Atmospheric Environment*, Vol. 117, pp. 217-281.
- [21] Gomez-Losada, A., Pires, J.C.M., Pino-Mejías, R. (2016) “Of background air pollution exposure in urban environments using a metric based on Hidden Markov Models”. *Atmospheric environment*, Vol. 127, pp. 255-261.
- [22] Hernandez-Lerma, O. (1981) “Procesos estocásticos: Introducción a la Teoría de Colas”, 2do. Coloquio del Departamento de Matemáticas. Centro de Investigación y Estudios Avanzados del IPN.
- [23] Jordan, M., y Jacobs, R. (1994) “Hierarchical mixtures of experts and the EM algorithm”, *Neural Computation*, Vol. 6, pp. 181-214.
- [24] Laverty, W.H., Milet, M.J., y Kelly, W. (2002) “Simulation of hidden Markov models with EXCEL”, *The Statistician*, Vol 51, pp. 31-40.
- [25] Lenschow, P., Abraham, H.J., Kutzner, K., Lutz, M., Preub, J.D., Reichenbacher, W. (2001) “Some ideas about the sources of PM10. *Atmospheric Environmental*”, Vol. 35, pp. S23-S33.
- [26] Mohamed, M.A and Garder, P. (2000) “Generalized Hidden Markov Models Part I: Theoretical Frameworks”, *IEEE Transactions on Fuzzy Systems*, Vol 8, pp. 67-81.

- [27] Newman, K.B et all. (2006) "Hidden Process Models for Animal Population Dynamic", Ecological Applications, Vol 16(1), pp.74-86.
- [28] Norris, J.R. (2004) "Markov Chain", Cambridge Series in Statistical and Probabilistic Mathematics.
- [29] Petrushin A. Valery. (2000) "Hidden Markov Models: Fundamentals and Applications", Center for Strategic Technology Research.
- [30] Psychogios, D.C Ungar, L.H. (1992) "A hybrid neural network -first principles approach to process modeling", AIChE Journal. Vol 38(10), pp.1499-1511.
- [31] Rabiner, L.R y Juang, B.H. (1986) "An Introduction to Hidden Markov Models", IEEE ASSP MAGAZINE, pp. 4-16.
- [32] Rabiner,L.R.(1989) "A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", IEEE ASSP MAGAZINE, Vol 77(2). pp. 257-286.
- [33] Rodriguez, F., y Bautista, S. (2000) "Modelos ocultos de Markov para el análisis de patrones espaciales", Ecosistemas, Asociación Española de Ecología Terrestre, Alicante España. Vol XV(3), pp. 69-76.
- [34] Sansom, J. (1998) "A hidden Markov model for rainfall using breakpoint data", Journal of Climate. Vol 11, pp. 42-53.
- [35] Scebba, F., Canaccini, F., Castagna, A., Bender, J., Weigel, H., y Ranieri, A. (2006) "Physiological and biochemical stress responses in grassland species are influenced by both early-season ozone exposure and interspecific competition". Environmental Pollution, 142 (3), pp 540-548.
- [36] Sousa, S.I.V., Martins, F.G., Alvim-Ferraz, M.C.M., y Pereira, M.C. (2007) "Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations". Environmental modelling and software, 22(1), pp. 97-103.
- [37] Velázquez, F. C., y Jimenez, A. S. (2007) "La contaminación por ozono troposférico. El caso de Motril Granada". Observatorio Mediambiental. Vol 10, pp. 265-280.
- [38] Viana, M., Kuhlbusch, T.A.J., Querol, X., et al. (2008) "Source apportionment of particulate matter in Europe: A review of method and results". Journal of Aerosol Science, Vol. 39, pp. 827-49.
- [39] Wu, C.F.J. (1984) "On the convergence properties of EM algorithm". The Annals of Statistics, Vol 11, pp. 95-103.
- [40] Xu, L y Jordan, M.I. (1996) "On convergence properties of the EM algorithm for Gaussian mixture", Neural computation, Vol 8, pp. 129-151.
- [41] Zucchini, W. y MacDonald, I.L.(2009) "Hidden Markov Models for Time Series. An Introduction Using R". Monographs on Statistics and Applied Probability 110. CRC Press. Florida, pp. 50-103.
- [42] [WWW.stanfordphd.com/Statisticalsoftware.html](http://WWW.stanfordphd.com/Statisticalsoftware.html).

# Apéndice

## Código del programa

El código empleado en el software R para estimar los modelos es el siguiente.

Para obtener los parámetros del modelo HMM con observaciones gaussianas, se puede utilizar el siguiente código en cada una de las series.

Leer la base de datos y llamar las librerías necesarias.

```
data = read.table(“unidad:/carpeta/datos.txt”, header = T)
sample <- data.frame(y=data)
library(depMixS4)
```

Para cada modelo con 1, 2,.. 7 estados ajustar:

```
m1 <- depMix(y.Noroeste ~ 1, data=sample, ns=1, ntimes=nrow(sample))
fm1 <- fit(m1, em=em.control(maxit=2000, tol=0.00000001, crit=relative))
m2 <- depMix(y.Noroeste ~ 1, data=sample, ns=2, ntimes=nrow(sample))
fm2 <- fit(m2, em=em.control(maxit=2000, tol=0.00000001, crit=relative))
m3 <- depMix(y.Noroeste ~ 1, data=sample, ns=3, ntimes=nrow(sample))
fm3 <- fit(m3, em=em.control(maxit=2000, tol=0.00000001, crit=relative))
m4 <- depMix(y.Noroeste ~ 1, data=sample, ns=4, ntimes=nrow(sample))
fm4 <- fit(m4, em=em.control(maxit=2000, tol=0.00000001, crit=relative))
m5 <- depMix(y.Noroeste ~ 1, data=sample, ns=5, ntimes=nrow(sample))
fm5 <- fit(m5, em=em.control(maxit=2000, tol=0.00000001, crit=relative))
m6 <- depMix(y.Noroeste ~ 1, data=sample, ns=6, ntimes=nrow(sample))
fm6 <- fit(m6, em=em.control(maxit=2000, tol=0.00000001, crit=relative))
m7 <- depMix(y.Noroeste ~ 1, data=sample, ns=7, ntimes=nrow(sample))
fm7 <- fit(m7, em=em.control(maxit=2000, tol=0.00000001, crit=relative))
```

Determinar cuál es el mejor modelo de acuerdo al criterio BIC y AIC

```
bic <- c(BIC(fm2), BIC(fm3), BIC(fm4), BIC(fm5), BIC(fm6), BIC(fm7))
which.min(bic)
aic <- c(AIC(fm1), AIC(fm2), AIC(fm3), AIC(fm4), AIC(fm5), AIC(fm6), AIC(fm7))
which.min(aic)
```

Obtener los parámetros del modelo seleccionado

```
summary(Modelo seleccionado)
```

Se determina la secuencia de estados y el rango de valores en cada estado:

```
probs <- posterior(modelo seleccionado)
table(probs[,1])
probs <- cbind(probs,sample)
first <- which(probsstate == 1)
second <- which(probsstate == 2)
range(probsy.Noroeste[first],na.rm = TRUE)
range(probsy.Noroeste[second],na.rm = TRUE)
apply(probs[,2:4],2,mean)
```

Ejemplo de la elaboración del gráfico de ajuste de los modelos:

*Elaborar el histograma de los datos:*

```
hist(samplesy.Noreste, nclass=17, freq=FALSE, ylim=c(0,0.06),col="light grey",lwd=0.001,
pch=0.1, bg=5, lty=3, fg=1, main="HMM- 3 estados- distribución gaussiana, xlab=Niveles pro-
medio diario de Ozono, ylab=Densidad)
x <- seq(-20,20,by=.1)
```

*Dibujar la curva asociada a cada estado, así como la leyenda correspondiente:*

```
curve(0.36*dnorm(x, mean=18.89, sd=4.32), col ="blue", lty = 1, lwd = 1,add=TRUE)
points(18.89, 0, type= "p", col="blue", lwd=4)
curve(0.43*dnorm(x, mean=28, sd=4.91), col =çchartreuse4", lty = 1, lwd = 1,add=TRUE)
points(28, 0, type= "p", col="blue", lwd=4)
curve(0.21*dnorm(x, mean=40.22, sd=6.34), col =red", lty = 1, lwd = 1,add=TRUE)
points(40.22, 0, type= "p", col="blue", lwd=4)
legend("topright",col=-c("blue","green","red"),lty = 1,lwd = 2,legend =
c(Primer estado, Segundo estado, Tercer estado),bty = "n")
```

*Dibujar la curva definida por el HMM así como sus leyendas:*

```
curve(0.36*dnorm(x, mean=18.89, sd=4.32)+ 0.43*dnorm(x, mean=28, sd=4.91)+
0.21*dnorm(x, mean=40.22, sd=6.34),col ="black", lty = 1, lwd = 1,add=TRUE)
points(27.23,0, type= "p",pch=5, col="black", lwd=2)
text(18.89,0.002, "m1",cex=0.8,col="blue")
text(28,0.002, "m2",cex=0.8,col=çchartreuse4")
text(40.22,0.002, "m3",cex=0.8,col=red")
text(27.23,0.003, "M",cex=0.8,col="black")
```

Gráfico de la secuencia de estados

*Gráfico de las observaciones*

```
plot(sampley.Suroeste, type="h", main="Secuencia de estados de las observaciones de ozono
para la zona suroeste del 2014 ", xlab="dias, 2014", ylab="Niveles promedio diario de Ozono",
xlim=c(1,365))
```

*Escribir en cada observación el estado correspondiente:*

```
for(i in 1:nrow(sample))
if(probsstate[i]==3)
text(i, sampley.Suroeste[i]+1, probsstate[i],cex=0.7, col=red")
if(probsstate[i]==2)
text(i, sampley.Suroeste[i]+1, probsstate[i],cex=0.7, col=chartreuse4")
if(probsstate[i]==1)
text(i, sampley.Suroeste[i]+1, probsstate[i],cex=0.7, col="blue")
```

*Dibujar las líneas correspondientes a las medias de cada distribución asociada a cada estado:*

```
abline(h=22.66, col="blue", lty=1, lwd=1)
text(370, 23.6, "m1",cex=0.9, col="blue")
abline(h=36.96, col=chartreuse4", lty=1, lwd=1)
text(370, 37.6, "m2",cex=0.9, col=chartreuse4")
abline(h=48.15, col=red", lty=1, lwd=1)
text(370, 49, "m3",cex=0.9, col=red")
abline(h=33.86, col="black", lty=1, lwd=1)
text(370, 34.8, "M",cex=0.9, col="black")

legend(245,66,col=c("blue",chartreuse4", red", "black"),lty=1,lwd=2,legend=c("Media de la
distribución en el estado 1=22.66", "Media de la distribución en el estado 2=36.96", "Media de
la distribución en el estado 3=48.15", "Media de la serie=33.86"), bty="n",cex=0.8)
```

Para obtener los parámetros de un HMM con observaciones gamma se puede emplear el siguiente código:

Leer los datos y llamar a las librerías necesarias:

```
data = read.table(Ünidad:/carpeta/datos.txt", header = T)
sample< -data.frame(y=data)
library(depmixS4)
```

Ajustar los HMM con 1, 2, ..., 7 estados:

```

m1 <- depmix(y.Noroeste ~ 1, data=sample, ns=1, family=Gamma("identity"),
respstar=c(0.6), ntimes=nrow(sample))
fm1 <- fit(m1, em=em.control(maxit=2000, tol=0.00000001, crit="relative"))
m2 <- depmix(y.Noroeste ~ 1, data=sample, ns=2, family=Gamma("identity"),
respstar=c(0.6,0.8), ntimes=nrow(sample))
fm2 <- fit(m2, em=em.control(maxit=2000, tol=0.00000001, crit="relative"))
m3 <- depmix(y.Noroeste ~ 1, data=sample, ns=3, family=Gamma("identity"),
respstar=c(0.6,0.8, 1), ntimes=nrow(sample))
fm3 <- fit(m3, em=em.control(maxit=2000, tol=0.00000001, crit="relative"))
m4 <- depmix(y.Noroeste ~ 1, data=sample, ns=4, family=Gamma("identity"),
respstar=c(0.6,0.8, 1,1), ntimes=nrow(sample))
fm4 <- fit(m4, em=em.control(maxit=2000, tol=0.00000001, crit="relative"))
m5 <- depmix(y.Noroeste ~ 1, data=sample, ns=5, family=Gamma("identity"),
respstar=c(0.6,0.8, 1,1,1), ntimes=nrow(sample))
fm5 <- fit(m5, em=em.control(maxit=2000, tol=0.00000001, crit="relative"))
m6 <- depmix(y.Noroeste ~ 1, data=sample, ns=6, family=Gamma("identity"),
respstar=c(0.6,0.8, 1,1,1,1), ntimes=nrow(sample))
fm6 <- fit(m6, em=em.control(maxit=2000, tol=0.00000001, crit="relative"))
m7 <- depmix(y.Noroeste ~ 1, data=sample, ns=7, family=Gamma("identity"),
respstar=c(0.6,0.8, 1,1,1,1,1), ntimes=nrow(sample))
fm7 <- fit(m7, em=em.control(maxit=2000, tol=0.00000001, crit="relative"))

```

Determinar el modelo con mejor ajuste:

```

bic <- -c(BIC(fm1), BIC(fm2), BIC(fm3), BIC(fm4), BIC(fm5), BIC(fm6),
BIC(fm7))
which.min(bic)

```

Obtener los estimadores de los parámetros de la cadena de Markov:

```

summary(fm3)

```

Determinar la secuencia de estados del modelo y el rango de observaciones asociadas a cada estado:

```
probs <- posterior(modelo seleccionado)
table(probs[,1])
probs <- cbind(probs,sample)
first <- which(probs state==1)
second <- which(probs state==2)
third <- which(probs state==3)
range(probs y.Noroeste [first],na.rm=TRUE)
range(probs y.Noroeste [second],na.rm=TRUE)
range(probs y.Noroeste [third],na.rm=TRUE)
apply(probs[,2:4],2,mean)
```

Calcular los estimadores de los parámetros de escala y de forma:

```
varianza1 <- var(probs y.Noroeste [first])
desviacionesta1 <- sqrt(var(probs y.Noroeste [first]))
escala1 <- varianza1/media1
forma1 <- media1/escala1
varianza2 <- var(probs y.Noroeste [second])
desviacionesta2 <- sqrt(var(probs y.Noroeste [second]))
escala2 <- varianza2/media2
forma2 <- media2/escala2
varianza3 <- var(probs y.Noroeste [third])
desviacionesta3 <- sqrt(var(probs y.Noroeste [third]))
escala3 <- varianza3/media3
forma3 <- media3/escala3
```