



UNIVERSIDAD MICHOACANA DE SAN NICOLÁS
DE HIDALGO

Facultad de Ciencias Físico – Matemáticas

“Mat. Luis Manuel Rivera”

TESIS

Estimando el número de excedencias de ozono en la Ciudad de
México usando procesos de Poisson no-homogéneos y el
muestreador de Gibbs

Que para obtener el título de

LICENCIADO EN CIENCIAS FÍSICO-MATEMÁTICAS

Presenta:

JOSÉ JULIÁN JARA ETTINGER

Asesor:

DRA. ELIANE REGINA RODRIGUES

Morelia, Michoacán. Abril de 2011

Estimando el número de excedencias de ozono en la
Ciudad de México usando procesos de Poisson
no-homogéneos y el muestreador de Gibbs

José Julián Jara Ettinger

Agradecimientos

A mi asesora, la Dra. Eliane Rodrigues, por su paciencia y dedicación con este trabajo. Sus enseñanzas fueron mas allá de lo académico y ha sido un modelo a seguir. Agradezco a los revisores de esta tesis: El Dr. Francisco Dominguez-Mota, el Dr. Gerardo Tinoco, el Dr. Alberto Mendoza, el Dr. Armando Sepulveda y la doctora Karina Figueroa. También agradezco al Dr. Jorge Achcar de la Facultad de Medicina de Ribeirão Preto, Universidad de San Paulo, y a la Dra. Angela Yu del departamento de Ciencias Cognitivas, Universidad de California en San Diego por sus valiosos comentarios.

A mi familia por todo el apoyo que me han brindado y el entusiasmo con el cual siempre han apoyado mis metas.

Índice general

1. Introducción	5
2. Bases de probabilidad	7
2.1. Eventos y sus probabilidades	7
2.2. Variables aleatorias	10
3. Inferencia Bayesiana	15
3.1. Función de Verosimilitud	15
3.2. Distribución a priori	16
3.2.1. Distribuciones no-informativas	16
3.3. Distribución a posteriori	17
3.4. Distribuciones Conjugadas	18
3.5. Prueba de hipótesis Bayesiana	18
4. Cadenas de Markov	21
4.1. Cadenas de Markov y clasificación de estados	21
4.2. Distribuciones estacionarias	30
4.3. Teorema Fundamental de Convergencia	38
4.4. Cadenas Reversas y Cadenas Reversibles en el tiempo	42
5. Procesos de Poisson	45
5.1. El Proceso de Poisson	45
5.2. Procesos de Poisson No-homogéneos	51
6. Funciones de Riesgo y Supervivencia	53
6.1. Algunos modelos importantes	54
6.1.1. Weibull	54
6.1.2. Musa-Okumoto	55
6.1.3. Goel-Okumoto generalizada	55
7. Método Monte Carlo vía Cadenas de Markov	59
7.1. Método Monte Carlo	59
7.2. Muestreador de Gibbs	62
7.2.1. Implementación	63

7.3.	Algoritmo de Metropolis-Hastings	64
7.3.1.	Matrices de transición	66
7.3.2.	El muestreador de Gibbs como un caso especial del algoritmo Metropolis-Hastings	68
7.4.	Convergencia	70
7.4.1.	Diagnóstico informal	70
7.4.2.	Prueba de Gelman-Rubin	70
8.	Aplicación a los Niveles de Ozono del Distrito Federal	73
8.1.	Formulación del problema	73
8.2.	Acercamiento Bayesiano	74
8.3.	Solución computacional	76
8.4.	Modelos Utilizados	77
8.4.1.	Weibull	77
8.4.2.	Musa-Okumoto	81
8.4.3.	Goel-Okumoto Generalizada	84
8.5.	Comparación de modelos	88
8.6.	Obtención de las observaciones	89
8.7.	Detalles de implementación	90
8.7.1.	Ejemplo de implementación	90
9.	Resultados Finales y Conclusiones	95
9.1.	Conclusión	95
9.2.	Discusión	98
A.	Resultados Complementarios	99
B.	Distribuciones Utilizadas	103
B.1.	Distribución de Poisson	103
B.2.	Distribución Normal	103
B.3.	Distribución Beta	103
B.4.	Distribución Gamma	104
C.	Códigos Fuente	105
C.1.	Weibull	105
C.2.	Musa-Okumoto	107
C.3.	Goel-Okumoto Generalizada	108
D.	Resultados Completos	111
D.1.	Centro	111
D.2.	Noreste	113
D.3.	Noroeste	115
D.4.	Sureste	117
D.5.	Suroeste	119

Capítulo 1

Introducción

El ozono es uno de los compuestos químicos más conocidos. Es bien sabido que la capa de Ozono, ubicada entre 13 y 40 kilómetros sobre la superficie de la Tierra, absorbe alrededor del 98% de la luz ultravioleta emitida por el Sol y que llega a la Tierra. Sin la capa de Ozono la humanidad no podría sobrevivir.

A pesar de lo valioso que es el ozono en la estratósfera, dentro de la tropósfera es uno de los tóxicos más peligrosos. La exposición al ozono está ligada a la muerte prematura, asma, bronquitis y paros cardiacos entre otros problemas pulmonares. Varios estudios han demostrado que existe una correlación significativa entre los niveles de ozono y las enfermedades respiratorias (Por ejemplo Jerret *et al.* [20] y Wilson *et al.* [31]).

Cuando los niveles de Ozono en una ciudad superan la concentración de 0.11 partes por millón (0.11ppm) por un período mayor a una hora, estratos sensibles de la población (como ancianos y recién nacidos) pueden sufrir una deterioración de salud muy seria (ver Bell *et al.* [3], [4]). Por lo tanto, el problema de predecir cuando el límite de 0.11ppm, o cualquier límite mayor, es superado es un problema de gran importancia.

La concentración y el límite de ozono que se considera legalmente peligroso varía según cada país (ver reporte [8]). La reglamentación en México es que una persona, no debe estar expuesta a una concentración de 0.11ppm o mayor en media por una hora o más (NOM, [7]).

Existen varios acercamientos al problema de predecir situaciones en las cuales los niveles de ozono alcanzan niveles riesgosos en una ciudad; entre ellos están Horowitz [14] y Smith [28] empleando teoría de valores extremos, Flaum [9], Loomis [19], Zolghadri y Henry [32], y Pan y Chen [21] utilizando análisis de series de tiempo, Guardani [13] con redes neuronales, Guardani [12] con análisis multivariado, y Álvarez [1], Austin y Tran [2] y Larsen [17] utilizando modelos de cadenas de Markov.

Cuando el objetivo es estimar el número de veces que un estándar ambiental es violado en un intervalo de tiempo, Javits [15], Raftery [22] y Smith

[28] utilizan procesos de Poisson homogéneos para modelar el problema. Sin embargo, es bien sabido que la homogeneidad en el tiempo no es una propiedad de una sucesión de mediciones de Ozono. En la tarea de superar la hipótesis de la homogeneidad de tiempo, Achcar *et al.* ([36] y [16]) han estado utilizando procesos de Poisson no-homogéneos, que es tema central de esta tesis.

En el presente trabajo se utilizan las mediciones de ozono tomadas por la red de monitoreo de la Ciudad de México entre el 1 de enero del 2003 al 31 de diciembre del 2009, divididas en 5 zonas: Noreste, Noroeste, Sureste, Suroeste y Centro. A continuación se consideran tres modelos de procesos de Poisson no-homogéneos que podrían explicar el comportamiento de las mediciones observadas (definidas por medio de su función de riesgo). Utilizando el punto de vista Bayesiano y después algoritmos estocásticos, se estiman los parámetros del modelo que mejor representan las mediciones observadas. Por último, los tres modelos son comparados utilizando el criterio del factor de Bayes.

Esta tesis está organizada de la siguiente manera. El capítulo 2 presenta las bases de la teoría de probabilidad necesarios para poder comprender los siguientes capítulos. El capítulo 3 trata del acercamiento Bayesiano para estimar los parámetros de los cuales depende un modelo. Los capítulos 4, 5 y 6 contienen la teoría fundamental matemática necesaria para poder tratar el problema central. El capítulo 7 utiliza la teoría de los capítulos anteriores para el desarrollo de los algoritmos estocásticos que se utilizan en el capítulo 8, el cual contiene la formulación matemática del problema, la implementación computacional y los resultados obtenidos. Por último, el capítulo 9 contiene las conclusiones del trabajo y una discusión de problemas futuros.

Capítulo 2

Bases de probabilidad

En este capítulo se presentan algunos teoremas y definiciones que forman parte de la base de la teoría de probabilidad y que son utilizados a lo largo de esta tesis. Introducciones más detalladas con el mismo enfoque pueden ser consultadas en Alfaro [23], Grimmet y Stirzaker [33], Grinstead y Snell [39], Resnick [27] y Ross [26].

2.1. Eventos y sus probabilidades

Definición 2.1.1 Un **experimento** es cualquier procedimiento que puede ser repetido una cantidad infinita de veces bajo las mismas condiciones y tiene un conjunto de resultados bien definido. Un experimento es llamado **experimento aleatorio** si el resultado no se puede predecir con certeza.

Definición 2.1.2 Suponga que se efectúan N repeticiones de un experimento aleatorio. El conjunto $x = (x_1, x_2, \dots, x_N)$ de valores obtenidos es llamado **muestra aleatoria**.

Definición 2.1.3 Un **espacio muestral**, denotado por Ω , es el conjunto de todos los posibles resultados de un experimento. Un subconjunto de Ω es llamado un **evento**.

Ejemplo 2.1.4 La actividad que consiste en lanzar un dado y observar el número que muestra la cara superior es un experimento aleatorio con espacio muestral

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

Definición 2.1.5 Una colección \mathcal{F} de subconjuntos de Ω es una σ -álgebra si satisface

- a) $\emptyset \in \mathcal{F}$,
- b) si $A_1, A_2, \dots \in \mathcal{F}$ entonces $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$,

c) si $A \in \mathcal{F}$ entonces $\Omega \setminus A \in \mathcal{F}$.

Observación: Combinando las condiciones (a) y (c) de la Definición, se tiene que $\Omega \in \mathcal{F}$.

Definición 2.1.6 Una **medida de probabilidad \mathbf{P}** en (Ω, \mathcal{F}) es una función $\mathbf{P} : \mathcal{F} \rightarrow [0,1]$ que satisface:

a) $\mathbf{P}(\emptyset) = 0$,

b) $\mathbf{P}(\Omega) = 1$,

c) si A_1, A_2, \dots es una colección de elementos disjuntos dos a dos de \mathcal{F} entonces

$$\mathbf{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbf{P}(A_i).$$

Notación: Dado un conjunto C , se denota por 2^C al **conjunto potencia** de C , es decir, el conjunto que consta de todos los subconjuntos de C .

Definición 2.1.7 Un **espacio de probabilidad** es una 3-ada $(\Omega, \mathcal{F}, \mathbf{P})$ que consiste de un conjunto Ω , una σ -álgebra $\mathcal{F} \subseteq 2^\Omega$ y una medida de probabilidad \mathbf{P} .

Lema 2.1.8 Sea $(\Omega, \mathcal{F}, \mathbf{P})$ un espacio de probabilidad. Sean A y B elementos de \mathcal{F} , entonces

a) $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$,

b) si $A \subseteq B$ entonces $\mathbf{P}(B) = \mathbf{P}(A) + \mathbf{P}(B \setminus A)$,

c) $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$.

Demostración:

a) Note que por definición se tiene que $A \cup A^c = \Omega$ y que $A \cap A^c = \emptyset$. Por lo tanto, por la definición de σ -álgebra y medida de probabilidad se tiene que

$$\mathbf{P}(A \cup A^c) = \mathbf{P}(A) + \mathbf{P}(A^c) = 1.$$

b) Note que es posible escribir $B = A \cup (B \setminus A)$. Como A y $(B \setminus A)$ son disjuntos, por definición de probabilidad se tiene

$$\mathbf{P}(B) = \mathbf{P}(A) + \mathbf{P}(B \setminus A).$$

c) Por el inciso anterior se tiene que $\mathbf{P}(B) = \mathbf{P}(A \cap B) + \mathbf{P}(B \setminus A)$ y por lo tanto

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B \setminus A) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B).$$

■

Definición 2.1.9 Sean A y B eventos en \mathcal{F} . Si $\mathbf{P}(B) > 0$, la **probabilidad condicional** de A dado que el evento B sucedió se define como

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}. \quad (2.1)$$

Lema 2.1.10 Sea B_1, B_2, \dots, B_n una partición de Ω , entonces para cualquier evento A en \mathcal{F}

$$\mathbf{P}(A) = \sum_{i=1}^n \mathbf{P}(A|B_i) \mathbf{P}(B_i).$$

Demostración: Note que los $\{A \cap B_i\}$ son ajenos dos a dos y

$$\bigcup_i \{A \cap B_i\} = A,$$

por lo que

$$\mathbf{P}(A) = \sum_{i=1}^n \mathbf{P}(A \cap B_i) = \sum_{i=1}^n \mathbf{P}(A|B_i) \mathbf{P}(B_i).$$

■

Teorema 2.1.11 (Teorema de Bayes) Sean A y B eventos en \mathcal{F} tales que $\mathbf{P}(B) > 0$, entonces

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

Demostración: Por definición se tiene que

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)},$$

y adicionalmente, también por definición se puede escribir

$$\mathbf{P}(A \cap B) = \mathbf{P}(B|A) \mathbf{P}(A),$$

de donde se sigue el resultado. ■

Lema 2.1.12 Sean A , B , y C eventos en \mathcal{F} tales que $\mathbf{P}(B \cap C) > 0$ y $\mathbf{P}(C) > 0$, entonces

$$\mathbf{P}(A \cap B|C) = \mathbf{P}(A|B \cap C)\mathbf{P}(B|C).$$

Demostración: Utilizando la Definición 2.1.9 sucesivamente se tiene que

$$\begin{aligned} \mathbf{P}(A \cap B|C) &= \frac{\mathbf{P}(A \cap B \cap C)}{\mathbf{P}(C)} \\ &= \frac{\mathbf{P}(A|B \cap C)\mathbf{P}(B \cap C)}{\mathbf{P}(C)} \\ &= \mathbf{P}(A|B \cap C)\mathbf{P}(B|C). \end{aligned}$$

■

Definición 2.1.13 Dos eventos A y B en \mathcal{F} son **independientes** si

$$\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B).$$

De manera general, la familia $\{A_i | i \in I\}$ es independiente si

$$\mathbf{P}(\cap_{i \in I} A_i) = \prod_{i \in I} \mathbf{P}(A_i).$$

2.2. Variables aleatorias

Definición 2.2.1 Sea (Ω, \mathcal{F}) un espacio muestral. Una **variable aleatoria** es una función $X : \Omega \rightarrow \mathbb{R}$ Lebesgue-medible, es decir

$$\{\omega \in \Omega : \mathbf{X}(\omega) \leq x\} \in \mathcal{F}$$

para toda $x \in \mathbb{R}$.

Dada una variable aleatoria X , se puede construir un espacio muestral y una medida de probabilidad, llamados espacio muestral y medida de probabilidad inducidos.

Definición 2.2.2 Para una variable aleatoria X , el **espacio muestral inducido**, denotado por Ω_X , es el conjunto de todos los números reales x tales que existe un $\omega \in \Omega$ de tal forma que $X(\omega) = x$. En otras palabras, el espacio muestral inducido Ω_X es el rango de la función X .

Definición 2.2.3 Para cualquier subconjunto $A \in \mathbb{R}$, la **función de probabilidad inducida** es la función tal que $P_X(X \in A)$ es la probabilidad de que $X(\omega) \in A$. Es decir,

$$P_X(X \in A) = P(\{\omega \in \Omega | X(\omega) \in A\}).$$

Ejemplo 2.2.4 Suponga que se tienen 5 pelotas etiquetadas A, B, C, D y E . Las pelotas A y B son blancas y las pelotas C, D y E son negras. Considere el experimento de elegir 3 pelotas al azar sin repetición. Entonces el espacio muestral es

$$\Omega = \{ABC, ABD, ABE, ACD, ACE, ADE, BCD, BCE, BDE, CDE\}.$$

Sea X la función $X : \Omega \rightarrow \mathbb{R}$ que indica la cantidad de pelotas blancas que hay en un elemento de Ω . Claramente X es una variable aleatoria.

Como las pelotas son elegidas al azar, cada uno de los 10 elementos de Ω tiene $1/10$ de probabilidad de ocurrir. Adicionalmente, el espacio muestral inducido por la variable aleatoria X es $\Omega_X = \{0, 1, 2\}$ y la función de probabilidad inducida está determinada por

$$\begin{aligned} \mathbf{P}(X = 0) &= 1/10 && \{CDE\}, \\ \mathbf{P}(X = 1) &= 6/10 && \{ACD, ACE, ADE, BCD, BCE, BDE\}, \\ \mathbf{P}(X = 2) &= 2/10 && \{ABC, ABD\}. \end{aligned}$$

Definición 2.2.5 La **función de distribución** de una variable aleatoria X es la función $F : \mathbb{R} \rightarrow [0, 1]$ dada por

$$F(x) = \mathbf{P}(X \leq x). \quad (2.2)$$

Definición 2.2.6 Una variable aleatoria \mathbf{X} es **discreta** si toma valores en algún subconjunto numerable $\{x_1, x_2, \dots\} \in \mathbb{R}$.

Definición 2.2.7 La **función de probabilidad** de una variable aleatoria discreta X es la función $f : \mathbb{R} \rightarrow [0, 1]$ dada por

$$f(x) = \mathbf{P}(X = x).$$

Definición 2.2.8 Una variable aleatoria \mathbf{X} es **continua** si su función de distribución puede ser expresada como

$$F(x) = \int_{-\infty}^x f(u)du, x \in \mathbb{R} \quad (2.3)$$

para alguna función integrable $f : \mathbb{R} \rightarrow [0, \infty)$. La función f es llamada **función de densidad** de \mathbf{X} .

Un listado completo de las funciones de probabilidad utilizadas en el resto de los capítulos se puede consultar en el apéndice B .

A pesar de la existencia de variables aleatorias que no son ni continuas ni discretas (Ejemplo 2.2.9), éstas no son utilizadas en esta tesis por lo que este caso especial es ignorado en el resto de los capítulos.

Ejemplo 2.2.9 Una moneda es lanzada y se obtiene águila con probabilidad $p \in [0, 1]$ y sol con probabilidad $q = 1 - p$. Si el resultado es Sol, éste es guardado. Sin embargo, si el resultado es águila, una ruleta horizontal es girada y el ángulo formado con la dirección norte en sentido horario es medido. La formulación obvia es tomar

$$\Omega = \{S\} \cup \{(A, \theta), 0 \leq \theta < 2\pi\}.$$

Sea $\mathbf{X} : \Omega \rightarrow \mathbb{R}$ una variable aleatoria dada por

$$\mathbf{X}(S) = -1, \quad \mathbf{X}((A, \theta)) = \theta,$$

entonces \mathbf{X} toma valores en $\{-1\} \cup [0, 2\pi)$ y claramente no cumple con las definiciones de variable aleatoria discreta o continua.

Definición 2.2.10 La **esperanza** de una variable aleatoria X con función de densidad $f(x)$ se define como

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx. \quad (2.4)$$

Observación: Dada una variable aleatoria discreta X , su esperanza se puede escribir como

$$E(x) = \sum_{\{x:f(x)>0\}} xf(x). \quad (2.5)$$

Lema 2.2.11 Sea $X : \Omega \rightarrow \mathbb{N}$ una variable aleatoria, entonces

$$E(X) = \sum_{k \in \mathbb{N}} P(X \geq k).$$

Demostración: La demostración es directa notando que $E(X)$ se puede escribir como

$$E(X) = \sum_{k \in \mathbb{N}} \sum_{j=1}^k P(X = k),$$

y por lo tanto es equivalente

$$E(X) = \sum_{j \in \mathbb{N}} \sum_{k \geq j} P(X = k) = \sum_{j \in \mathbb{N}} P(X \geq j).$$

■

Definición 2.2.12 Sea $\mu = E[\mathbf{X}]$ el valor esperado de una variable aleatoria \mathbf{X} . La **variancia** de \mathbf{X} se define como

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

Teorema 2.2.13 Si \mathbf{X} es una variable aleatoria con media μ y tal que $E[\mathbf{X}^2] < \infty$, entonces se puede escribir

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

Demostración:

$$\begin{aligned} \text{Var}(\mathbf{X}) &= E[(\mathbf{X} - \mu)^2] \\ &= E[\mathbf{X}^2 - 2\mu\mathbf{X} + \mu^2] \\ &= E[\mathbf{X}^2] - 2\mu E[\mathbf{X}] + \mu^2 \\ &= E[\mathbf{X}^2] - 2\mu^2 + \mu^2 \\ &= E[\mathbf{X}^2] - \mu^2 \\ &= E[\mathbf{X}^2] - [E[\mathbf{X}]]^2. \end{aligned}$$

■

Definición 2.2.14 La **función de distribución conjunta** $F: \mathbb{R}^2 \rightarrow [0, 1]$ de dos variables aleatorias X y Y está dada por

$$F(x, y) = \mathbf{P}(X \leq x, Y \leq y).$$

Definición 2.2.15 La **función conjunta de probabilidad** de dos variables aleatorias discretas X y Y $f: \mathbb{R}^2 \rightarrow [0, 1]$ se define como

$$f(x, y) = \mathbf{P}(X = x, Y = y).$$

Capítulo 3

Inferencia Bayesiana

Este capítulo contiene los conceptos esenciales de la inferencia Bayesiana y algunos resultados que serán la base para la formulación del problema central de esta tesis. Introducciones más detalladas pueden ser consultadas en Bolstad [5], Gamerman y Lopes [37] y Reyes [10].

3.1. Función de Verosimilitud

Definición 3.1.1 Considere una variable aleatoria X tal que su función de probabilidad depende de un cierto valor θ . La variable θ que determina dicha función es llamada **parámetro**.

Definición 3.1.2 El conjunto de posibles valores que un parámetro puede tomar, denotado por Θ , es llamado **espacio parametral**.

Observación: El espacio parametral define una familia

$$F = \{f(x, \theta) | \theta \in \Theta\}$$

de funciones de densidad para la variable aleatoria \mathbf{X} .

Definición 3.1.3 Sea $x = (x_1, x_2, \dots, x_N)$ una muestra aleatoria y $\theta \in \Theta$ el parámetro del modelo que produjo la muestra x . La función que describe al modelo en términos de θ es llamada **función de verosimilitud**, denotada por $l(\theta|x)$. Si la muestra aleatoria es independiente entonces la función de verosimilitud es de la forma

$$l(\theta|x) = \prod_{i=1}^N f(x_i|\theta)$$

donde $f(x, \theta)$ es la función de densidad asociada al experimento.

3.2. Distribución a priori

Definición 3.2.1 Sea $x = \{x_1, x_2, x_3, \dots, x_N\}$ una muestra aleatoria y $\theta \in \Theta$ el parámetro que describe el experimento que produjo la muestra x . Llamamos **distribución a priori** a la distribución que se le da al parámetro θ sin tomar en cuenta la muestra aleatoria obtenida.

Notación: Dada una muestra aleatoria $x = (x_1, x_2, \dots, x_N)$ y un parámetro θ que describe el experimento que produjo x . Se denota por $p(\theta)$ a la distribución a priori del parámetro θ .

Observación: La distribución a priori de un parámetro θ es especificada por intuición o basada en estudios previos. Sin embargo, si no se tiene información alguna, se suele utilizar una distribución aleatoria uniforme sobre un espacio muestral lo suficientemente grande.

3.2.1. Distribuciones no-informativas

En casos en los cuales no existe información alguna para determinar la distribución a priori, lo más conveniente es elegir una distribución uniforme.

Definición 3.2.2 Se dice que una distribución a priori es **no-informativa** para un parámetro θ si la distribución asignada no favorece a ningún valor específico en Θ .

Si el parámetro toma valores en un espacio finito $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$, una distribución a priori no-informativa puede ser

$$p(\theta_i) = \frac{1}{n}$$

y si toma valores en un espacio continuo $[a, b]$ puede ser

$$p(\theta_i) = \frac{1}{b-a}.$$

Mas aún, si el espacio consta de toda la recta real $S = (-\infty, \infty)$ entonces se asigna la distribución

$$p(\theta) = c,$$

donde c es una constante positiva. Claramente la distribución elegida no es una distribución propia, pues

$$\int_{-\infty}^{\infty} p(\theta) d\theta = \infty.$$

Sin embargo, como se verá en la siguiente sección, está formulación puede ser útil en ciertos casos.

3.3. Distribución a posteriori

Definición 3.3.1 Sea $x = (x_1, x_2, \dots, x_N)$ una muestra aleatoria, $\theta \in \Theta$ el parámetro del modelo que produjo la muestra x , $p(\theta)$ la distribución a priori del parámetro θ y $l(\theta|x)$ la función de verosimilitud. La distribución del parámetro θ tomando en cuenta la muestra aleatoria x es llamada **distribución a posteriori**, denotada por $P(\theta|x)$.

En ocasiones, el parámetro θ de un modelo puede ser dependiente de otro parámetro aleatorio ν , sin embargo en muchos casos el valor de un hiperparámetro es de poca influencia y demasiado ambiguo como para ser de utilidad. Debido a que el trabajo final de esta tesis cae dentro de la situación descrita, los casos donde se utilizan hiperparámetros aleatorios para describir un modelo son ignorados.

Teorema 3.3.2 Sea $x = (x_1, x_2, \dots, x_N)$ una muestra aleatoria de un experimento, $\theta \in \Theta$ el parámetro del modelo y $p(\theta)$ la distribución a priori del parámetro, entonces

$$P(\theta|x) = \frac{l(x|\theta)p(\theta)}{\sum_{\theta \in \Theta} p(x, \theta)}.$$

Demostración: Por el Teorema de Bayes (Teorema 2.1.11) se tiene que primeramente

$$\mathbf{P}(\theta|x) = \frac{\mathbf{P}(x|\theta)p(\theta)}{\mathbf{P}(x)},$$

y por lo tanto, utilizando la notación definida en este capítulo,

$$\begin{aligned} P(\theta|x) &= \frac{l(x|\theta)p(\theta)}{P(x)} \\ &= \frac{l(x|\theta)p(\theta)}{\sum_{\theta \in \Theta} p(x, \theta)}. \end{aligned}$$

■

Observación: Una vez obtenida la muestra aleatoria x , $\sum_{\theta \in \Theta} P(x, \theta)$ es una constante y por lo tanto la distribución a posteriori se puede escribir de la forma más compacta

$$P(\theta|x) \propto l(x|\theta)p(\theta). \quad (3.1)$$

Observación: Cuando el espacio muestral consta de toda la recta real y se da la distribución a priori no-informativa, la distribución a posteriori puede ser correcta (salvo una constante de normalización) si $\sum_{\theta \in \Theta} p(x, \theta) < \infty$.

3.4. Distribuciones Conjugadas

Definición 3.4.1 Sea \mathbf{P} una familia de distribuciones, x una muestra aleatoria y θ un parámetro del cual depende el modelo que produjo x . Si tanto la distribución a priori $p(\theta)$ como la distribución a posteriori $P(\theta|x)$ son elementos de \mathbf{P} , las distribuciones son llamadas **distribuciones conjugadas**.

Debido a la definición, el uso de distribuciones conjugadas es útil únicamente si la familia \mathbf{P} es elegida de manera que una buena restricción acerca de las posibles formas de la distribución a posteriori sea inferible.

Ejemplo 3.4.2 Una función de densidad $f(x|y)$ pertenece a la familia de distribución exponencial uniparamétrica si es de la forma

$$f(x|\theta) = a(x)e^{\phi(\theta)t(x)+b(\theta)}. \quad (3.2)$$

La importancia de esta familia es debido a que contiene a varias de las distribuciones importantes como

a) Distribución Normal (Gaussiana) $N(\theta, \sigma^2)$

$$\phi(\theta) = \frac{\theta}{\sigma^2}, t(x) = x, b(\theta) = -\frac{\theta^2}{2\sigma^2}, a(x) = e^{-\frac{x^2}{2\sigma^2}}.$$

b) Distribución Binomial

$$\phi(\theta) = \log\left(\frac{\theta}{1-\theta}\right), t(x) = x, b(\theta) = n \log(1-\theta), a(x) = \binom{n}{x}.$$

c) Distribución exponencial

$$\phi(\theta) = -\theta, t(x) = x, b(\theta) = \log(\theta), a(x) = 1.$$

d) Distribución de Poisson

$$\phi(\theta) = \log(\theta), t(x) = x, b(\theta) = -\log(\theta), a(x) = 1.$$

Sin embargo no contiene a la distribución uniforme, a la distribución T, ni mezclas discretas de densidades.

3.5. Prueba de hipótesis Bayesiana

Suponga que se tiene un conjunto de hipótesis (llamadas modelos) que podrían explicar un fenómeno dado, si se da el caso en que no existe manera de saber que hipótesis es la correcta de manera definitiva, el acercamiento

más razonable es formalizar el modelo de manera que se pueda inferir que hipótesis es la más probable de ser la correcta.

El acercamiento Bayesiano consiste precisamente en calcular que tan probable es que un modelo sea más adecuado que otro, es decir, calcular las probabilidades posteriores de que cada modelo sea correcto. Suponga que se tienen dos modelos M_1, M_2 . Al observar el evento Y , se calcula el **factor de Bayes** dado por

$$\begin{aligned} BF_{M_1, M_2} &= \frac{\mathbf{P}(M_1|Y)/\mathbf{P}(M_2|Y)}{\mathbf{P}(M_1)\mathbf{P}(M_2)} \\ &= \frac{[\mathbf{P}(Y|M_1)\mathbf{P}(M_1)/\mathbf{P}(Y)][\mathbf{P}(Y|M_2)\mathbf{P}(M_2)/\mathbf{P}(Y)]}{\mathbf{P}(M_1)\mathbf{P}(M_2)} \\ &= \frac{\mathbf{P}(Y|M_1)}{\mathbf{P}(Y|M_2)}. \end{aligned}$$

Intuitivamente, el factor de Bayes BF_{M_1, M_2} es justamente la ventaja de M_1 sobre M_2 después de observar Y . Formalmente, si $BF_{M_1, M_2} > 1$, entonces el Modelo 1 es más probable de ser el correcto, y si $BF_{M_1, M_2} < 1$ entonces el Modelo 2 es el más probable de explicar el fenómeno.

Capítulo 4

Cadenas de Markov

En este capítulo se definen los conceptos esenciales y se demuestran los teoremas necesarios relacionados con las cadenas de Markov. Una vez que se ha desarrollado lo suficiente la teoría de las cadenas de Markov en las primeras dos secciones, la sección 4.3 presenta la demostración del Teorema Fundamental de Convergencia (Teorema 4.3.2), argumentablemente el Teorema de mayor importancia en el estudio de las cadenas de Markov, cuyo resultado es esencial para el problema central de esta tesis.

Una introducción más detallada y un estudio más profundo al tema está disponible en Grimmett y Stirzaker [33], Karlin y Taylor [38], Ross [26], Stroock [29] y Tudor [30].

4.1. Cadenas de Markov y clasificación de estados

Definición 4.1.1 Un **proceso estocástico** $\mathbf{X} = \{X_t : t \in T\}$ es una colección de variables aleatorias indexadas por un conjunto T llamado conjunto de índices (en general indica el tiempo). Al conjunto de valores que pueden tomar las variables aleatorias se le llama **espacio de estados** y se denota por E .

Definición 4.1.2 Un proceso estocástico \mathbf{X} con espacio de estados E es una **cadena de Markov** si satisface la **condición de Markov** :

$$\mathbf{P}(X_n = e | X_0 = x_0, X_1 = x_1, X_2 = x_2, \dots, X_{n-1} = x_{n-1}) = \mathbf{P}(X_n = e | X_{n-1} = x_{n-1})$$

para toda $n \geq 1$ y toda $e, x_0, x_1, \dots, x_{n-1} \in \mathbf{E}$.

Intuitivamente, la condición de Markov dice que la cadena cambiará de estado acorde únicamente a su estado actual, sin memoria alguna de los estados anteriores.

Ejemplo 4.1.3 (Caminata aleatoria unidimensional) Imagine una partícula situada inicialmente en un punto $z \in \mathbb{Z}$. En cada paso, la partícula se mueve un punto a la izquierda con probabilidad $p \in (0, 1)$ o un punto a la derecha con probabilidad $q = 1 - p$. Sea $X = \{X_t \mid t \in \mathbb{N}\}$ un proceso estocástico donde la variable aleatoria X_t indica la posición de la partícula después de t saltos. Suponga que después de n saltos la partícula ha tomado la trayectoria $(z, x_2, x_3, \dots, x_n)$. Entonces

$$\mathbf{P}(X_{n+1} = s \mid X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_1 = z) =$$

$$\mathbf{P}(X_{n+1} = s \mid X_n = x_n) = \begin{cases} p & \text{si } s = x_n + 1 \\ q & \text{si } s = x_n - 1 \\ 0 & \text{en cualquier otro caso} \end{cases}$$

Por lo tanto X es una cadena de Markov.

Definición 4.1.4 Una cadena de Markov X es **homogénea** si

$$\mathbf{P}(X_{n+1} = j \mid X_n = i) = P(X_1 = j \mid X_0 = i),$$

es decir, la cadena es independiente del tiempo.

Debido a que las cadenas de Markov no-homogéneas no son utilizadas durante el resto de esta tesis, todas las cadenas de Markov utilizadas a partir de este punto son homogéneas a menos que sea explícitamente aclarado.

Definición 4.1.5 Sea X una cadena de Markov con espacio de estados E . La **matriz de transición** P de X es la matriz de tamaño $|E| \times |E|$ tal que la entrada (i, j) es la probabilidad de que la cadena, estando en el estado i , llegue al estado j en el siguiente salto. Es decir,

$$\begin{aligned} p_{ij} &= \mathbf{P}(X_{n+1} = j \mid X_n = i) \\ &= \mathbf{P}(X_1 = j \mid X_0 = i). \end{aligned}$$

Definición 4.1.6 La **matriz de transición del n-ésimo salto** $P_n = (p_{ij}(n))$ es la matriz de tamaño $|E| \times |E|$ con las probabilidades de transición n-ésima, es decir,

$$\begin{aligned} p_{ij}(n) &= \mathbf{P}(X_{m+n} = j \mid X_m = i) \\ &= \mathbf{P}(X_n = j \mid X_0 = i). \end{aligned}$$

El siguiente teorema es de gran utilidad para comprender como una cadena de Markov evoluciona con el tiempo. Nos dice que la matriz de salto $m + n$, en el caso que E sea finito, es justamente el producto de la matriz de salto m por la matriz de salto n .

Teorema 4.1.7 (Chapman-Kolmogorov) Sea X una cadena de Markov con espacio de estados E , entonces para $i, j \in E, m, n \geq 0$, se tiene que

$$p_{ij}(m+n) = \sum_{k \in E} p_{ik}(m)p_{kj}(n).$$

Demostración: Utilizando el Teorema 2.1.12 y la ley de probabilidad total, se tiene,

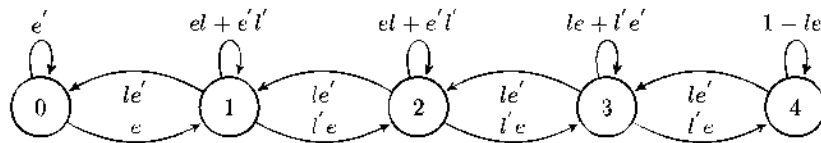
$$\begin{aligned} p_{ij}(m+n) &= \mathbf{P}(X_{m+n} = j | X_0 = i) \\ &= \sum_{k \in E} \mathbf{P}(X_{m+n} = j, X_m = k | X_0 = i) \\ &= \sum_{k \in E} \mathbf{P}(X_{m+n} = j | X_m = k, X_0 = i) \mathbf{P}(X_m = k | X_0 = i) \\ &= \sum_{k \in E} \mathbf{P}(X_{m+n} = j | X_m = k) \mathbf{P}(X_m = k | X_0 = i), \end{aligned}$$

y el resultado sigue. ■

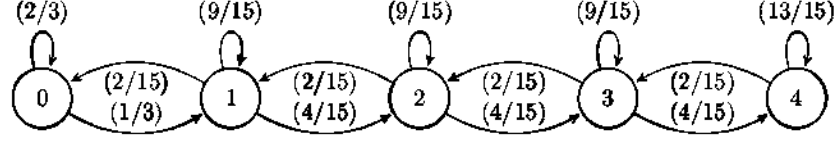
Corolario 4.1.8 Sea X una cadena de Markov con espacio de estados E finito, entonces la matriz de transición n -ésima es la n -ésima potencia de la matriz de transición, es decir, $P_n = P^n$.

En el siguiente Ejemplo (Ejemplo 4.1.9) se muestra la aplicación del Corolario 4.1.8 aplicado a un modelado de líneas telefónicas en un centro de atención a clientes.

Ejemplo 4.1.9 En un centro de atención a clientes se tienen 4 líneas telefónicas y en cada intervalo $[n, n+10]$ de 10 minutos entra una llamada con probabilidad $e = 1/3$ (y ninguna llamada entra con probabilidad $e' = 2/3$), a menos que las 4 líneas estén ocupadas, en cuyo caso las llamadas entrantes son rechazadas. Si hay llamadas en transcurso, una llamada es liberada con probabilidad $l = 1/5$ (y no es liberada con probabilidad $l' = 4/5$). Sea $X = \{X_n : n \in \mathbb{N}\}$ una cadena de Markov tal que X_n es la variable aleatoria que indica la cantidad de líneas ocupadas en el intervalo de tiempo $[10n, 10(n+1)]$. Entonces la cadena puede ser visualizada de la siguiente manera:



O sustituyendo los valores,



La matriz de transición de la cadena será

$$P = \begin{pmatrix} 2/3 & 1/3 & 0 & 0 & 0 \\ 2/15 & 9/15 & 4/15 & 0 & 0 \\ 0 & 2/15 & 9/15 & 4/15 & 0 \\ 0 & 0 & 2/15 & 9/15 & 4/15 \\ 0 & 0 & 0 & 2/15 & 13/15 \end{pmatrix}.$$

Mas aún, como E es finito, por el corolario 4.1.8 se tiene que las matrices saltos 2 y 3 son:

$$P_2 = \begin{pmatrix} \frac{22}{45} & \frac{19}{45} & \frac{4}{45} & 0 & 0 \\ \frac{38}{225} & \frac{11}{25} & \frac{8}{25} & \frac{16}{225} & 0 \\ \frac{4}{225} & \frac{4}{25} & \frac{97}{25} & \frac{8}{225} & \frac{16}{225} \\ 0 & \frac{4}{225} & \frac{4}{25} & \frac{97}{225} & \frac{88}{225} \\ 0 & 0 & \frac{4}{225} & \frac{44}{225} & \frac{59}{75} \end{pmatrix}, \quad P_3 = \begin{pmatrix} \frac{86}{225} & \frac{289}{675} & \frac{112}{675} & \frac{16}{675} & 0 \\ \frac{578}{3375} & \frac{49}{135} & \frac{1076}{3375} & \frac{16}{125} & \frac{64}{3375} \\ \frac{112}{3375} & \frac{538}{3375} & \frac{43}{125} & \frac{356}{1125} & \frac{496}{3375} \\ \frac{8}{3375} & \frac{4}{3375} & \frac{178}{1125} & \frac{1125}{1193} & \frac{1532}{3375} \\ 0 & \frac{8}{3375} & \frac{124}{3375} & \frac{766}{3375} & \frac{2477}{3375} \end{pmatrix}.$$

Teorema 4.1.10 Sean X y Y dos cadenas de Markov independientes con matrices de transición P y P' respectivamente. Sea $Z = \{Z_n = (X_n, Y_n) | n \in \mathbb{N}\}$, entonces Z es una cadena de Markov con matriz de transición dada por

$$Q_{(x_1, y_1), (x_2, y_2)} = P_{x_1, x_2} P'_{y_1, y_2}.$$

Demostración: El hecho de que Z es una cadena de Markov es inmediato a partir de que X y Y son ambas cadenas de Markov independientes. La matriz de transición de Z entonces es

$$\begin{aligned} Q_{(x_1, y_1), (x_2, y_2)} &= \mathbf{P}(Z_{n+1} = (x_2, y_2) | Z_n = (x_1, y_1)) \\ &= \mathbf{P}(X_{n+1} = x_2, Y_{n+1} = y_2 | X_n = x_1, Y_n = y_1) \\ &= \frac{\mathbf{P}(X_{n+1} = x_2, Y_{n+1} = y_2 | X_n = x_1, Y_n = y_1)}{\mathbf{P}(X_n = x_1, Y_n = y_1)} \\ &= \frac{\mathbf{P}(X_{n+1} = x_2, X_n = x_1) \mathbf{P}(Y_{n+1} = y_2, Y_n = y_1)}{\mathbf{P}(X_n = x_1) \mathbf{P}(Y_n = y_1)} \\ &= \mathbf{P}(X_{n+1} = x_2 | X_n = x_1) \mathbf{P}(Y_{n+1} = y_2 | Y_n = y_1) \\ &= P_{x_1, x_2} P'_{y_1, y_2}, \end{aligned}$$

donde las igualdades 3 y 5 son por la definición de probabilidad condicional y la cuarta igualdad es gracias a la independencia de X y Y . ■

Definición 4.1.11 Un estado i **accede** a j ($i \rightarrow j$) si la probabilidad de que la cadena visite el estado j , dado que comenzó en el estado i , es positiva, es decir, existe $m \geq 0$ tal que $p_{ij}(m) > 0$.

Notación: Dos estados i y j se **comunican** entre sí cuando $i \rightarrow j$ y $j \rightarrow i$, y se denota por $i \leftrightarrow j$.

Teorema 4.1.12 La relación de comunicación (\leftrightarrow) es una relación de equivalencia.

Demostración: Para demostrar que la relación es de equivalencia se necesitan demostrar las siguientes tres propiedades:

- a) reflexividad: es trivial, pues $p_{ii}(0) = 1$ y por lo tanto $i \leftrightarrow i$.
- b) simetría: La relación es simétrica por definición. Si $i \leftrightarrow j$ entonces $j \leftrightarrow i$.
- c) transitividad: Sean i, j, k estados tales que $i \leftrightarrow j$ y $j \leftrightarrow k$, entonces existen $m, n \in \mathbb{N}$ tales que $p_{ij}(m) > 0$ y $p_{jk}(n) > 0$. Por el Teorema 4.1.7,

$$p_{ik}(m+n) = \sum_{w \in E} p_{iw}(m)p_{wk}(n) \geq p_{ij}(m)p_{jk}(n) > 0$$

y por lo tanto $i \rightarrow k$. Similarmente se muestra que $k \rightarrow i$ y se sigue que la relación \leftrightarrow es transitiva. ■

Definición 4.1.13 El **período** $d(i)$ de un estado i es el máximo común divisor de los tiempos en los cuales la probabilidad de un regreso al estado de origen i es positiva, es decir,

$$d(i) = \text{mcd}\{n : p_{ii}(n) > 0\}.$$

Un estado i es **periódico** si $d(i) > 1$ y **aperiódico** si $d(i) = 1$. Si todos los estados de una cadena son aperiódicos, se dice que la cadena X es una cadena aperiódica.

Lema 4.1.14 Si $i \leftrightarrow j$ entonces existe $n \geq 0$ tal que $p_{ii}(n) > 0$.

Demostración: La demostración es idéntica a la parte (c) del Teorema 4.1.12 tomando $k = i$. ■

Teorema 4.1.15 Si $i \leftrightarrow j$ entonces $d(i) = d(j)$.

Demostración: Como $i \leftrightarrow j$, por definición existen enteros n y m tales que $p_{ij}(n) > 0$ y $p_{ji}(m) > 0$. Sea s tal que $p_{ii}(s) > 0$. Por el Teorema 4.1.7,

$$p_{jj}(n+m) \geq p_{ji}(m)p_{ij}(n) > 0,$$

y

$$p_{jj}(n+s+m) \geq p_{ji}(m+s)p_{ij}(n) \geq p_{ji}(m)p_{ii}(s)p_{ij}(n) > 0.$$

Por lo tanto, dada una cadena que comienza en el estado j , su regreso al estado original tiene probabilidad mayor que 0 en los tiempos $n+m$ y $n+m+s$ y consecuentemente $d(j)$ divide tanto a $n+m$ como a $n+m+s$. Como s es la diferencia de dos números divisibles entre $d(j)$ ($n+m$ y $n+m+s$), se sigue que $d(j)$ también divide a s . Como este argumento fue construido para un tiempo s arbitrario tal que $p_{ii}(s) > 0$, se sigue que $d(j)$ divide a todos los $e \geq 0$ tales que $p_{ii}(e) > 0$ y por lo tanto $d(j)$ divide a $d(i)$.

Un argumento simétrico intercambiando los estados i y j demuestra que $d(i)$ divide a $d(j)$. Como $d(i)|d(j)$ y $d(j)|d(i)$, necesariamente $d(j) = d(i)$ y el teorema queda demostrado. ■

Teorema 4.1.16 *Sea X una cadena de Markov con espacio de estados E , entonces para todo estado $i \in E$ existe $N \in \mathbb{N}$ tal que para todo $n \geq N$*

$$p_{ii}(nd(i)) > 0. \quad (4.1)$$

Demostración: Sea $A = \{a_k, k \in \mathbb{N} | p_{ii}(k) > 0\} \subseteq \mathbb{N}$ de tal forma que si $u < v$ entonces $a_u < a_v$, en otras palabras A es el conjunto ordenado que consta de todos los tiempos en los cuales el regreso de la cadena a su estado original i es probable. Se quiere demostrar entonces que para todo natural n mayor que un determinado N , $nd(i) \in A$. Claramente A es un conjunto infinito numerable, pues por tratarse de una cadena homogénea, si $p_{ii}(n) > 0$ entonces $p_{ii}(kn) > 0$ para todo $k \in \mathbb{N}$.

Por el Corolario A.0.2 existe un subconjunto finito

$$M = \{b_1, b_2, \dots, b_m\} \subset A$$

tal que $\text{mcd}(M) = d(i)$. Utilizando ahora el Corolario A.0.4 existe un $N \in \mathbb{N}$ tal que todo $nd(i) \geq N$ puede ser expresado como combinación lineal sobre \mathbb{N} de los a_i . Por lo tanto para todo $nd(i) \geq N$ existe un conjunto $\{k_1, k_2, \dots, k_m\}$ tal que $nd(i) = k_1b_1 + k_2b_2 + \dots + k_mb_m$ lo que implica que $nd(i) \in A$. ■

Observación: Si un estado i es aperiódico entonces el Teorema anterior dice que existe un natural N tal que para toda $n \geq N$, $p_{ii}(n) > 0$.

Corolario 4.1.17 Si $p_{ij}(m) > 0$ entonces $p_{ij}(m + nd(i)) > 0$ para todo n suficientemente grande.

Demostración: Se sigue directamente del Teorema 4.1.16 debido a que la cadena es homogénea y del hecho que $i \rightarrow j$. ■

Definición 4.1.18 llamamos

$$f_{ij}(n) = \mathbf{P}(X_n = j, X_{n-1} \neq j, \dots, X_1 \neq j | X_0 = i) \quad (4.2)$$

a la probabilidad de que la cadena llegue al estado j por primera vez en n pasos comenzando en el estado i .

Observación: Es claro que $f_{ii}(1) = p_{ii}$ y recursivamente

$$p_{ii}(n) = \sum_{k=0}^{n-1} f_{ii}(k)p_{ii}(n-k). \quad (4.3)$$

Definición 4.1.19 Un estado i es **recurrente** si una vez que la cadena se encuentra en este estado, su regreso es seguro. Si la probabilidad de que el regreso al estado i no suceda es positiva, el estado es llamado **transitorio**. Formalmente un estado i es recurrente si

$$\mathbf{P}(X_n = i \text{ para algún } n \in \mathbb{N} | X_0 = i) = 1$$

y transitorio si

$$\mathbf{P}(X_n = i \text{ para algún } n \in \mathbb{N} | X_0 = i) < 1.$$

Lema 4.1.20 Sean $x, y \in E$ estados tales que $x \leftrightarrow y$, entonces:

a) x es recurrente si y sólo si y es recurrente

b) x es transitorio si y sólo si y es transitorio

Demostración: Como $x \leftrightarrow y$, entonces existen $n, m \in \mathbb{N}$ tales que $k = p_{xy}(n)p_{yx}(m) > 0$. Por el Teorema 4.1.7, para todo $r \in \mathbb{N}$

$$p_{yy}(n+m+r) \geq p_{yx}(m)p_{xx}(r)p_{xy}(n) = kp_{xx}(r) \quad (4.4)$$

y

$$p_{xx}(n+m+r) \geq p_{xy}(n)p_{yy}(r)p_{yx}(m) = kp_{yy}(r). \quad (4.5)$$

Para ver ahora que la afirmación (a) es verdadera, es suficiente notar que si x es un estado recurrente entonces

$$\sum_{n \in \mathbb{N}} p_{xx}(n) = \infty,$$

y por lo tanto, utilizando la desigualdad (4.4),

$$\sum_{n \in \mathbb{N}} p_{yy}(n+r+m) = \infty$$

por lo que y será recurrente. Para la parte (b) suponga que x es transitorio, esto implica que

$$\sum_{n \in \mathbb{N}} p_{xx}(n+m+r) < \infty$$

y por lo tanto, utilizando la desigualdad (4.5),

$$\sum_{n \in \mathbb{N}} p_{xx}(n) < \infty,$$

lo cual significa que y es un estado transitorio. ■

Definición 4.1.21 Dado $A \subset E$ un subconjunto propio del espacio de estados, el primer tiempo de entrada a A es la variable aleatoria

$$T_A = \begin{cases} \min\{n \geq 1 : X_n \in A\} & \text{si } \{n \geq 1 : X_n \in A\} \neq \emptyset \\ +\infty & \text{si } \{n \geq 1 : X_n \in A\} = \emptyset \end{cases}$$

Notación: Si el conjunto A consta de un solo punto la notación utilizada es $T_{\{x\}} = T_x$.

Teorema 4.1.22 Sea X una Cadena de Markov con espacio de estados E , entonces

- a) $x \in E$ es un estado recurrente si y sólo si $P(T_x < \infty | X_0 = x) = 1$
- b) $x \in E$ es un estado transitorio si y sólo si $P(T_x < \infty | X_0 = x) < 1$

Demostración: Las afirmaciones son triviales notando que por definición $T_x < \infty$ si y sólo si $X_n = x$ para algún $n \in \mathbb{N}$. ■

Definición 4.1.23 Dada una cadena de Markov X con espacio de estados E , definimos μ_i como el tiempo esperado de que la cadena regrese por primera vez al estado de origen i . Es decir,

$$\mu_i = \begin{cases} \infty & \text{si } i \text{ es transitorio} \\ \sum_{n=1}^{\infty} n f_{ii}(n) & \text{si } i \text{ es recurrente} \end{cases}$$

Incluso cuando i es recurrente, μ_i puede ser infinito y por lo tanto se introduce la siguiente definición.

Definición 4.1.24 El estado i es **positivo recurrente** si $\mu_i < \infty$ y **nulo recurrente** si $\mu_i = \infty$.

Definición 4.1.25 Un conjunto C de estados es

- **cerrado** si $p_{ij} = 0$ para todo $i \in C, j \notin C$.
- **irreducible** si $i \leftrightarrow j$ para todo $i, j \in C$.

Teorema 4.1.26 (Teorema de la Descomposición) *Un espacio S de estados se puede particionar de manera única como*

$$S = T \cup C_1 \cup C_2 \cup \dots \quad (4.6)$$

donde T es el conjunto de estados transitorios y los C_i son conjuntos cerrados irreducibles de estados recurrentes.

Demostración: Sean C_i las clases de equivalencia construidas por la relación \leftrightarrow (Teorema 4.1.12) y $T = \{\cup_r C_r | C_r \text{ es de estados transitorios}\}$ la unión de todos los conjuntos C_r que constan de estados transitorios. Con esta construcción el conjunto T contiene a todos los estados transitorios y los conjuntos C_i que no forman parte del conjunto T son conjuntos irreducibles (por la definición de la relación \leftrightarrow) que constan de estados recurrentes (pues no forman parte de T). Para concluir con la demostración basta únicamente mostrar que si C_r es tal que $C_r \cap T = \emptyset$ entonces C_r es un conjunto cerrado. Suponga que C_r no es cerrado, entonces existen $i \in C_r, j \notin C_r$ y $k \in \mathbb{N}$, tales que $p_{ij}(k) > 0$, y por lo tanto

$$\begin{aligned} \mathbf{P}(X_n \neq i, \text{ para toda } n \geq 1 | X_0 = i) &\geq \mathbf{P}(X_k \neq i | X_0 = i) \\ &\geq \mathbf{P}(X_k = j | X_0 = i) \\ &> 0, \end{aligned}$$

es decir, la probabilidad de que la cadena no regrese al estado originario i es mayor que cero, lo cual contradice el hecho que i es recurrente. ■

Ejemplo 4.1.27 (Caminata aleatoria con barreras absorbentes)

Considere la cadena de Markov descrita en el ejemplo 4.1.3 con la diferencia de que ahora se limita el espacio de estados al conjunto $\{0, 1, \dots, N\}$. Si la cadena llega al estado 0 o al estado N , la cadena se queda en el mismo estado por el resto del tiempo. La cadena de Markov se puede imaginar como una persona apostando una moneda en cada paso y jugando hasta perder todo su dinero (en cuyo caso llegaría al estado 0) o hasta obtener una cantidad buscada N . La evolución de la cadena de Markov está dada por

$$\mathbf{P}(X_{n+1} = j | X_n = i) = \begin{cases} p & \text{si } j = i + 1, \\ q & \text{si } j = i - 1, \\ 0 & \text{en cualquier otro caso.} \end{cases}$$

Con matriz de transición dada por

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & \dots & \dots & 0 \\ q & 0 & p & 0 & \dots & \dots & \dots & 0 \\ 0 & q & 0 & p & \dots & \dots & \dots & 0 \\ \cdot & & & \cdot & & & & \\ & \cdot & & \cdot & & & & \\ & & \cdot & & \cdot & & & \\ \cdot & \cdot & \cdot & \cdot & 0 & 0 & 1 \end{pmatrix}$$

La descomposición según el Teorema 4.1.26 es

$$S = T \cup C_1 \cup C_2,$$

donde $C_1 = \{0\}$, $C_2 = \{n\}$ y $T = \{1, 2, \dots, N-1\}$. Además, todos los estados son transitorios con periodo $d(i) = 2$ cuando $i \in T$ y aperiódicos cuando $i \in C_1$ o C_2 .

Definición 4.1.28 Un estado aperiódico positivo recurrente i es llamado **ergódico**.

4.2. Distribuciones estacionarias

Definición 4.2.1 Sea X una cadena de Markov con espacio de estados E . Un vector π es una **distribución estacionaria** de la cadena X si π tiene entradas $(\pi_j : j \in E)$ tales que

- $\pi_j \geq 0$, para toda $j \in E$,
- $\sum_{j \in E} \pi_j = 1$,
- $\pi = \pi \mathbf{P}$, es decir $\pi_j = \sum_{i \in E} \pi_i p_{ij}$.

La distribución es llamada estacionaria debido a que utilizando (c) inductivamente se sigue que para toda $n \geq 0$,

$$\pi \mathbf{P}^n = \pi.$$

Definición 4.2.2 Una cadena de Markov \mathbf{X} es llamada **estacionaria** si su distribución inicial es una distribución estacionaria.

Ejemplo 4.2.3 Sea X una cadena de Markov con espacio de estados $E = \{1, 2, 3\}$ y matriz de transición

$$P = \begin{pmatrix} 0.5 & 0.25 & 0.25 \\ 0.5 & 0 & 0.5 \\ 0.25 & 0.25 & 0.5 \end{pmatrix}.$$

Como

$$\begin{pmatrix} 0.5 & 0.25 & 0.25 \\ 0.5 & 0 & 0.5 \\ 0.25 & 0.25 & 0.5 \end{pmatrix} \begin{pmatrix} 0.4 & 0.2 & 0.4 \\ 0.4 & 0.2 & 0.4 \\ 0.4 & 0.2 & 0.4 \end{pmatrix} = \begin{pmatrix} 0.4 & 0.2 & 0.4 \\ 0.4 & 0.2 & 0.4 \\ 0.4 & 0.2 & 0.4 \end{pmatrix}$$

entonces el vector $\pi = (0.4, 0.2, 0.4)$ es un vector de distribución estacionaria de la cadena X .

Teorema 4.2.4 Sea X una cadena de Markov con vector de distribución estacionaria π . Si a X_0 se le asigna la distribución π , es decir,

$$P(X_0 = y) = \pi_y,$$

entonces

a) para todo $m \in \mathbb{N}$, $P(X_m = y) = \pi_y$

b) para todo $k \in \mathbb{N}$ y para todo estado $y \in E$ se tiene que

$$P(X_n \neq y, n = 1, 2, 3, \dots, k) = P(X_n \neq y, n = 0, 1, \dots, k-1). \quad (4.7)$$

Demostración:

a) La demostración de la primera parte se hará por un inducción. Suponga que $n = 0$. Como la distribución inicial de la cadena es la distribución estacionaria π entonces

$$P(X_0 = y) = \pi_y.$$

Suponga ahora que distribución hasta el salto $n - 1$ es la distribución estacionaria. Utilizando la parte c) de la Definición 4.2.1

$$P(X_n = y) = \sum_i p_{iy} P(X_{n-1} = i) = \sum_i p_{iy} \pi_y = \pi_y.$$

b) La ecuación 4.7 se sigue directamente del inciso anterior, pues para todo estado i

$$P(X_1 = i) = P(X_0 = i) = \pi_i.$$

■

A continuación se presentará una serie de resultados que serán necesarios para poder demostrar el Teorema Fundamental de Convergencia, uno de los teoremas más importantes de esta tesis.

Lema 4.2.5 *Sea X una cadena de Markov irreducible con espacio de estados E y vector de probabilidad estacionario π , entonces*

a) $\pi_e > 0$ para todo $e \in E$

b) la cadena X es recurrente

Demostración: Note primero que por ser π una distribución estacionaria entonces

$$\sum_{e \in E} \pi_e = 1,$$

y por lo tanto existe $x \in E$ tal que $\pi_x > 0$.

Ahora, por la Definición 4.2.1

$$\pi_y = \sum_{e \in E} \pi_e p_{ey}(n).$$

Como X es una cadena irreducible, para cualquier pareja de estados $x, y \in E$ existe $n \in \mathbb{N}$ tal que $p_{xy}(n) > 0$ y por lo tanto

$$\pi_y = \sum_{e \in E} \pi_e p_{ey}(n) \geq \pi_x p_{xy}(n) > 0.$$

Con esto queda demostrada la parte (a).

Para ver que la parte (b) se cumple note primero que X es irreducible y entonces por el Lema 4.1.20 si existe un estado transitorio, entonces todos los estados serían transitorios, y por lo tanto

$$\lim_{n \rightarrow \infty} p_{xy}(n) = 0$$

para cualesquiera $x, y \in E$. Ahora, utilizando el Teorema A.0.5, se tiene que para toda $y \in E$

$$\pi_y = \lim_{n \rightarrow \infty} \sum_{e \in E} \pi_e p_{ey}(n) = \sum_{e \in E} \lim_{n \rightarrow \infty} \pi_e p_{ey}(n) = 0.$$

Esto contradice el hecho de que π es un vector de probabilidad estacionaria, y por lo tanto no existen estados transitorios. Con esto el Lema queda demostrado. ■

Notación: Dado un estado y , se denota por $a_m(y)$ a la probabilidad de que el tiempo de entrada sea mayor que m , es decir

$$a_m(y) = \mathbf{P}(T_y > m).$$

Teorema 4.2.6 *Sea X una cadena irreducible. Si existe un vector de probabilidad estacionaria π , entonces la cadena es positivo-recurrente y*

$$\pi_x \mu_x = 1$$

para cualquier $x \in E$.

Demostración: Suponga que X tiene como distribución inicial a π . Por el Lema 4.2.5, para todo $e \in E$, $\pi_e > 0$. Ahora, utilizando el Lema 2.2.11 ,

$$\begin{aligned} \pi_y \mu_y &= \left(\sum_{n=1}^{\infty} n f_{yy}(n) \right) \pi_y \\ &= E(T_y | X_0 = y) \mathbf{P}(X_0 = y) \\ &= \sum_{n=1}^{\infty} \mathbf{P}(T_y \geq n | X_0 = y) \mathbf{P}(X_0 = y) \\ &= \sum_{n=1}^{\infty} \mathbf{P}(T_y \geq n, X_0 = y). \end{aligned}$$

Note ahora que para cada $n \geq 2$, la última ecuación se puede escribir como

$$\begin{aligned} \mathbf{P}(T_y \geq n, X_0 = y) &= \mathbf{P}(X_0 = y, X_m \neq y, m = 1, \dots, n-1) \\ &= \mathbf{P}(X_m \neq y, m = 1, \dots, n-1) \\ &\quad - \mathbf{P}(X_m \neq y, m = 0, 1, \dots, n-1), \end{aligned}$$

que se puede reescribir por el Teorema 4.2.4 como

$$\begin{aligned} &= \mathbf{P}(X_m \neq y, m = 0, 1, \dots, n-2) \\ &\quad - \mathbf{P}(X_m \neq y, m = 0, \dots, n-1) \\ &= a_{n-2}(y) - a_{n-1}(y). \end{aligned}$$

Ahora, cuando $n = 1$, se tiene

$$\mathbf{P}(T_y \geq 1, X_0 = y) = \mathbf{P}(X_0 = y),$$

y por lo tanto se puede escribir

$$\begin{aligned} \pi_y \mu_y &= \mathbf{P}(X_0 = y) + \sum_{n=2}^{\infty} (a_{n-2}(y) - a_{n-1}(y)) \\ &= \mathbf{P}(X_0 = y) + \lim_{m \rightarrow \infty} \sum_{n=2}^m (a_{n-2}(y) - a_{n-1}(y)). \end{aligned}$$

Note ahora que $\sum_{n=2}^m (a_{n-2}(y) - a_{n-1}(y))$ es una suma telescópica en donde los términos se van eliminando y resulta en $a_0(y) - a_{m-1}(y)$. O sea,

$$\begin{aligned}\pi_y \mu_y &= \mathbf{P}(X_0 = y) + a_0(y) + \lim_{m \rightarrow \infty} (-a_{m-1}(y)) \\ &= 1 - \lim_{m \rightarrow \infty} a_{m-1}(y) \\ &= 1 - \lim_{m \rightarrow \infty} a_m(y),\end{aligned}$$

donde se utilizo el hecho de que

$$\mathbf{P}(X_0 = y) + a_0(y) = \mathbf{P}(X_0 = y) + \mathbf{P}(X_0 \neq y) = 1.$$

Para demostrar el Teorema solo falta ver que $\lim_{m \rightarrow \infty} a_m = 0$.

Por el Lema 4.2.5 se sabe que y es un estado recurrente y por lo tanto

$$\lim_{m \rightarrow \infty} \mathbf{P}(T_y \geq m, X_0 = y) = 0,$$

y

$$\lim_{m \rightarrow \infty} a_m(y) = \lim_{m \rightarrow \infty} \sum_{x \in E: \mathbf{P}(X_0=x)>0} \mathbf{P}(T_y > m, X_0 = x) = 0$$

donde nuevamente la última igualdad es debido al Lema A.0.5. Por lo tanto

$$\pi_y \mu_y = 1 - \lim_{m \rightarrow \infty} a_m(y) = 1,$$

y como $\pi_y > 0$ y $\mu_y < \infty$ para todo $y \in E$, se sigue que todos los estados son positivo-recurrentes. \blacksquare

Lema 4.2.7 *Sea X una cadena de Markov irreducible y recurrente, entonces para $x, y \in E$ se tiene*

a) *Existe $n \in \mathbb{N}$ tal que $\mathbf{P}(T_y = n | X_0 = x) > 0$.*

b) *$\mathbf{P}(T_x < \infty | X_0 = y) = 1$.*

Demostración: Como X es una cadena recurrente, entonces

$$\mathbf{P}(T_x < \infty | X_0 = x) = 1$$

para todo $x \in E$. Dados $x, y \in E$, sea

$$S = \{n \in \mathbb{N} : p_{xy}(n) > 0\}$$

el conjunto de todos los tiempos en los cuales la cadena, comenzando en el estado x , puede llegar al estado y . Como X es una cadena irreducible, se sigue que S es un conjunto no vacío. Sea $k_0 = \min(S)$, entonces

$$c = \mathbf{P}(T_y = k_0 | X_0 = x) > 0$$

y por lo tanto, para todo $n > k_0$,

$$\begin{aligned}
\mathbf{P}(T_x \geq n | X_0 = x) &\geq \mathbf{P}(T_x \geq n, T_y = k_0 | X_0 = x) \\
&= \mathbf{P}(T_x \geq n | T_y = k_0, X_0 = x) \mathbf{P}(T_y = k_0 | X_0 = x) \\
&= c \mathbf{P}(T_x \geq n | X_0 = x, X_1 \neq y, \dots, X_{k_0-1} \neq y, X_{k_0} = y) \\
&= c \mathbf{P}(T_x \geq n - k_0 | X_0 = y),
\end{aligned}$$

donde la última igualdad es gracias a que la cadena es homogénea.

Por ser X una cadena irreducible y recurrente, note que $\mathbf{P}(T_x = \infty | X_0 = x) = 0$. Tomando ahora el límite de $\mathbf{P}(T_x \geq n | X_0 = x)$ cuando n se acerca a infinito, se tiene que

$$\lim_{n \rightarrow \infty} \mathbf{P}(T_x \geq n | X_0 = x) = \mathbf{P}(T_x = \infty | X_0 = x) = 0$$

y como

$$\mathbf{P}(T_x \geq n | X_0 = x) \geq c \mathbf{P}(T_x \geq n - k_0 | X_0 = y)$$

se sigue que

$$\mathbf{P}(T_x = \infty | X_0 = y) = 0.$$

■

Notación: Para facilitar la lectura del inciso (c) de la demostración del Lema 4.2.8 se denota por $\lambda_{e,x}(n)$ a la probabilidad de que la cadena que comenzó en e , llegue al estado x en n saltos, con la condición adicional de que el tiempo de entrada a e sea mayor que n , es decir,

$$\lambda_{e,x}(n) = \mathbf{P}(X_n = x, T_e \geq n | X_0 = e).$$

Lema 4.2.8 *Sea X una cadena de Markov con espacio de estados E , entonces para cualesquiera estados $x, z \in E$,*

a) $\lambda_{z,x}(1) = p_{zx}$

b) $\lambda_{z,x}(n) = \sum_{y \neq z} \lambda_{z,y}(n-1) p_{yx}, \quad n \geq 2.$

Demostración: Para la parte (a) es suficiente notar que

$$\lambda_{z,x}(1) = \mathbf{P}(X_1 = x, T_z \geq 1 | X_0 = z) = p_{zx}.$$

Suponga ahora que $n \geq 2$, entonces

$$\begin{aligned}
\lambda_{z,x}(n) &= \mathbf{P}(X_n = x, T_z \geq n | X_0 = z) \\
&= \mathbf{P}(X_n = x, X_{n-1} \neq z, T_z \geq n - 1 | X_0 = z) \\
&= \sum_{y \neq z} \mathbf{P}(X_n = x, X_{n-1} = y, T_z \geq n - 1 | X_0 = z) \\
&= \sum_{y \neq z} \mathbf{P}(X_n = x | X_{n-1} = y, T_z \geq n - 1, X_0 = z) \times \\
&\quad \mathbf{P}(X_{n-1} = y, T_z \geq n - 1, X_0 = z) \\
&= \sum_{y \neq z} \mathbf{P}(X_n = x | X_{n-1} = y) \times \\
&\quad \mathbf{P}(X_{n-1} = y, T_z \geq n - 1 | X_0 = z) \\
&= \sum_{y \neq z} p_{yx} \lambda_{z,y}(n - 1).
\end{aligned}$$

Lo cual demuestra la parte (b) del Lema. ■

Notación: Se denota por $\mathbf{1}_p$ a la función $f : \{V, F\} \rightarrow \{0, 1\}$ que devuelve 1 si la proposición p es verdadera y 0 si la proposición p es falsa.

Lema 4.2.9 *Sea X una cadena de Markov irreducible con espacio de estados E y matriz de transición P , entonces para cualquier estado positivo-recurrente $e \in E$ existe un vector $\gamma(e) = (\gamma_x(e))_{x \in E}$ el cual satisface:*

a) $\gamma_x(e) < \infty$ para todo $x \in E$

b) $\sum_{x \in E} \gamma_x(e) = E(T_e | X_0 = e)$

c) $\gamma(e) = \gamma(e)P$

Demostración: Sea $e \in E$ un estado positivo-recurrente arbitrario pero fijo. Sea $\gamma(e) = (\gamma_x(e))_{x \in E}$ tal que

$$\begin{aligned}
\gamma_x(e) &= \sum_{n=1}^{\infty} E(\mathbf{1}_{(X_n=x) \cap (T_e \geq n)} | X_0 = e) \\
&= E\left(\sum_{n=1}^{T_e} \mathbf{1}_{(X_n=x)} | X_0 = e\right).
\end{aligned}$$

Es decir, la entrada x -ésima del vector $\gamma(e)$ es el valor esperado del número de visitas a x antes de regresar a su estado inicial e . Se demostrará ahora que el vector construido cumple las tres propiedades.

b) Note que

$$\begin{aligned}
E(T_e|X_0 = e) &= \sum_{n=1}^{\infty} \mathbf{P}(T_e \geq n|X_0 = e) \\
&= \sum_{n=1}^{\infty} \mathbf{P}(\cup_{x \in E} (X_n = x, T_e \geq n)|X_0 = e) \\
&= \sum_{n=1}^{\infty} \sum_{x \in E} \mathbf{P}(X_n = x, T_e \geq n|X_0 = e) \\
&= \sum_{x \in E} \sum_{n=1}^{\infty} \mathbf{P}(X_n = x, T_e \geq n|X_0 = e) \\
&= \sum_{x \in E} \gamma_x(e).
\end{aligned}$$

Donde la primera igualdad es por el Lema 2.2.11 y la cuarta igualdad es por el Teorema A.0.5.

a) La igualdad demostrada nos dice que $E(T_e|X_0 = e)$ es la suma del tiempo promedio en que la cadena estará en un estado distinto del estado e antes de su primer regreso a e . Utilizando la parte (b) y sabiendo que la cadena es positivo recurrente, se sigue directamente que para todo $x, e \in E$

$$\gamma_x(e) < \infty.$$

c) Por el inciso b),

$$\begin{aligned}
\gamma_x(e) &= \sum_{n=1}^{\infty} \lambda_{e,x}(n) \\
&= \lambda_{e,x}(1) + \sum_{n=2}^{\infty} \lambda_{e,x}(n).
\end{aligned}$$

Utilizando ahora el Lema 4.2.8 en la igualdad anterior, y después reescri-

biendo con el Teorema A.0.5, se tiene que

$$\begin{aligned}
 \gamma_x(e) &= p_{ex} + \sum_{n=2}^{\infty} \sum_{y \neq e} p_{yx} \lambda_{e,y}(n-1) \\
 &= p_{ex} + \sum_{y \neq e} p_{yx} \left(\sum_{n=2}^{\infty} \lambda_{e,y}(n-1) \right) \\
 &= p_{ex} + \sum_{y \neq e} p_{yx} \left(\sum_{n=1}^{\infty} \lambda_{e,y}(n) \right) \\
 &= \gamma_e(e) p_{ex} + \sum_{y \neq e} p_{yx} \gamma_y(e) \\
 &= \sum_{y \in E} \gamma_y(e) p_{yx}.
 \end{aligned}$$

Note que la penúltima igualdad fue debido a que $\gamma_e(e) = 1$ y $\sum_{n=1}^{\infty} \lambda_{e,y}(n) = \gamma_y(e)$. Por lo tanto se tiene que $\gamma(e) = \gamma(e)P$ y por lo tanto el vector $(\gamma_x(e))_{e \in E}$ es estacionario.

■

Corolario 4.2.10 *Sea X una cadena de Markov irreducible con espacio de estados E finito y un estado positivo recurrente, entonces para X*

- a) *existe un único vector de probabilidad estacionaria π*
- b) $\pi_x = \gamma_x(e)/\mu_e$ y $\gamma_x(e) = \mu_e/\mu_x$
- c) *Todos los estados son positivo-recurrentes.*

4.3. Teorema Fundamental de Convergencia

Teorema 4.3.1 *Sea X una cadena de Markov irreducible con espacio de estados E , entonces las siguientes condiciones son equivalentes:*

- a) *Existe un único vector de probabilidad estacionaria $\pi = (\pi_e)_{e \in E}$*
- b) *La cadena es positivo-recurrente*
- c) *Existe un estado positivo recurrente*

Además, si cualquiera de las condiciones se cumple, entonces para todo estado $x \in E$,

$$\pi_x \mu_x = 1. \tag{4.8}$$

Demostración: La demostración se hará por cuatro partes, a) \Rightarrow b), b) \Rightarrow c), c) \Rightarrow a), y por último que la ecuación 4.8 se sigue el inciso a).

- i) El Lema 4.2.6 nos dice justamente que si existe un vector de probabilidad estacionaria, entonces la cadena es positivo-recurrente. Por lo tanto (a) implica (b).
- ii) Por definición, una cadena positivo-recurrente es una cadena tal que todos sus estados son positivo-recurrentes y por lo tanto (b) implica (c).
- iii) Por el Lema 4.2.9 se tiene que si la cadena tiene un estado positivo-recurrente entonces existe un vector de distribución estacionaria y claramente es único debido al Lema 4.2.6 y al hecho de que los estados son positivo-recurrentes.
- iv) Nuevamente por el Lema 4.2.6, sabemos que si existe un vector de probabilidad estacionaria, entonces la ecuación 4.8 se cumple.

Con esto último el teorema queda demostrado. ■

La teoría desarrollada en este capítulo es suficiente para poder demostrar el Teorema fundamental de convergencia. A pesar de la naturaleza aleatoria de una cadena de Markov, el Teorema fundamental de convergencia permite encontrar cierto orden dentro de la evolución de una cadena ergódica. El siguiente teorema afirma que si se deja correr una cadena de Markov ergódica por un tiempo muy largo, entonces la cadena se asentará en una distribución (su distribución estacionaria) y esto será independiente del estado en el que ésta comenzó.

Teorema 4.3.2 (Teorema Fundamental de Convergencia) *Sea X una cadena de Markov irreducible, aperiódica y positivo-recurrente (ergódica) con espacio de estados E , entonces para todo $x, y \in E$*

$$\lim_{n \rightarrow \infty} p_{xy}(n) = \frac{1}{\mu_y}. \quad (4.9)$$

Demostración: Para demostrar el teorema se construirá primero una cadena auxiliar de la siguiente manera:

Sea Y una cadena de Markov independiente de X con el mismo espacio de estados E y la misma matriz de transición. Se define ahora

$$Z = \{Z_n = (X_n, Y_n) | n \in \mathbb{N}\}.$$

Por el Lema 4.1.10, Z es una cadena de Markov con matriz de transición

$$q_{(x_1, y_1)(x_2, y_2)} = p_{x_1 x_2} p_{y_1 y_2}.$$

Como X y Y son aperiódicas, se sigue que Z es aperiódica. Utilizando ahora el Corolario 4.1.17, existe un $N \in \mathbb{N}$ tal que para todo $n \geq N$

$$p_{x_1 x_2}(n) > 0, \quad p_{y_1 y_2}(n) > 0,$$

y por lo tanto

$$q_{(x_1, y_1), (x_2, y_2)}(n) = p_{x_1, x_2}(n)p_{y_1, y_2}(n) > 0$$

por lo que Z es una cadena irreducible.

Como X es positivo-recurrente, entonces por el Teorema 4.3.1, existe un único vector de distribución estacionaria π para X y para Y , y por lo tanto Z también tiene una distribución estacionaria

$$\nu = (\nu_{(x, y)} = \pi_x \pi_y : (x, y) \in E \times E).$$

Mas aún, como $\pi_x, \pi_y > 0$, entonces $\pi_x \pi_y > 0$ para todo $x, y \in E$ y por lo tanto utilizando nuevamente el Teorema 4.3.1, Z será una cadena positivo-recurrente.

Se demostrará ahora que las respectivas distribuciones estacionarias de X y Y no dependen del estado inicial.

Sea $s \in E$ un estado arbitrario y

$$T = \min\{n | Z_n = (s, s)\}$$

el menor tiempo en el cual la cadena Z llega al estado (s, s) .

Como Z es irreducible y positivo-recurrente, por el Lema 4.2.7, se tiene que

$$\mathbf{P}(T < \infty | Z_0 = (x, y)) = 1.$$

Una vez que Z llega al estado (s, s) , las cadenas X y Y se comportarán igual por tener la misma distribución, es decir

$$\mathbf{P}(X_n = y, T \leq n) = \mathbf{P}(Y_n = y, T \leq n). \quad (4.10)$$

La demostración de la ecuación (4.10) se hará por inducción. El caso en que $n = 1$ es claro. Suponga ahora que (4.10) se cumple para n , entonces para $n + 1$ se tiene

$$\begin{aligned} \mathbf{P}(X_{n+1} = y, T \leq n + 1) &= \mathbf{P}(X_{n+1} = y, T = n + 1) \\ &\quad + \sum_{e \in E} \mathbf{P}(X_{n+1} = y, X_n = e, T \leq n) \\ &= \mathbf{P}(X_{n+1} = y, T = n + 1) \\ &\quad + \sum_{e \in E} p_{e, y} \mathbf{P}(X_n = e, T \leq n) \\ &= \mathbf{P}(Y_{n+1} = y, T = n + 1) \\ &\quad + \sum_{e \in E} p_{e, y} \mathbf{P}(Y_n = e, T \leq n) \\ &= \mathbf{P}(Y_{n+1} = y, T \leq n + 1) \end{aligned}$$

donde la tercera igualdad fue debido a que por hipótesis de inducción

$$\mathbf{P}(X_n = e, T \leq n) = \mathbf{P}(Y_n = e, T \leq n)$$

y además

$$\mathbf{P}(X_{n+1} = y, T = n + 1) = \mathbf{P}(Y_{n+1} = y, T = n + 1)$$

ya que si $T = n + 1$ entonces necesariamente $X_{n+1} = Y_{n+1} = s$. Con esto se concluye que la ecuación (4.10) es verdadera.

Suponga ahora que la cadena Y tiene distribución inicial igual a la distribución estacionaria π y suponga que la cadena X comienza en el estado x . Por el Teorema 4.3.1

$$\mathbf{P}(Y_n = y) = \pi_y = \frac{1}{\mu_y}.$$

Por otra parte, note que

$$\mathbf{P}(X_n = y | X_0 = x) = \mathbf{P}(X_n = y, T \leq n | X_0 = x) + \mathbf{P}(X_n = y, T > n | X_0 = x)$$

y además

$$\pi_y = \frac{1}{\mu_y} = \mathbf{P}(Y_n = y, T \leq n) + \mathbf{P}(Y_n = y, T > n).$$

Ahora, utilizando las dos últimas ecuaciones obtenidas y la ecuación (4.10) se sigue que

$$p_{xy}(n) - \pi_y \leq \mathbf{P}(X_n = y, T > n) + \mathbf{P}(Y_n = y, T > n).$$

Análogamente se obtiene

$$\pi_y - p_{xy}(n) \leq \mathbf{P}(X_n = y, T > n) + \mathbf{P}(Y_n = y, T > n)$$

y por lo tanto

$$|p_{xy}(n) - \pi_y| \leq \mathbf{P}(X_n = y, T > n) + \mathbf{P}(Y_n = y, T > n) \leq 2\mathbf{P}(T > n).$$

Como $\lim_{n \rightarrow \infty} \mathbf{P}(T > n | X_0 = x) = 0$ se concluye que

$$\lim_{n \rightarrow \infty} \mathbf{P}(X_n = y | X_0 = x) = \pi_y.$$

■

4.4. Cadenas Reversas y Cadenas Reversibles en el tiempo

Considere una cadena de Markov \mathbf{X} irreducible, aperiódica y positivo-recurrente. Por el Teorema 4.3.2, la cadena \mathbf{X} tiene una distribución estacionaria única π y aun más,

$$\lim_{n \rightarrow \infty} \mathbf{P}(X_n = y) = \pi_y.$$

Si se le da la distribución estacionaria π como distribución inicial a la cadena \mathbf{X} , entonces ésta se puede idear como un proceso que comenzó en el tiempo $t = -\infty$.

Definición 4.4.1 Sea \mathbf{X} una cadena de Markov estacionaria con espacio de estados E , probabilidad de transición p_{ij} y distribución estacionaria π . Se define la cadena reversa \mathbf{X}^* como la cadena de estados

$$\dots X_{n+2}, X_{n+1}, X_n, X_{n-1}, X_{n-2} \dots$$

Observación: Para poder definir correctamente la cadena reversa, es necesario que la cadena haya comenzado en el tiempo $t = -\infty$, o que la cadena cumpla con las premisas del Teorema 4.3.2 y tenga distribución inicial estacionaria.

Teorema 4.4.2 *La cadena reversa de una cadena de Markov es una cadena de Markov*

Demostración: Se quiere demostrar que

$$\mathbf{P}(X_n = x_n | X_{n+1} = x_{n+1}, \dots) = \mathbf{P}(X_n = x_n | X_{n+1} = x_{n+1}). \quad (4.11)$$

Como $\dots, X_n, X_{n+1}, \dots$ es una cadena de Markov, X_n es independiente de X_m siempre que $n > m + 1$, y esto es justamente lo que afirma la ecuación 4.11. ■

Teorema 4.4.3 *la probabilidad de que la cadena reversa visite el estado j dado que comenzo en el estado i esta dada por*

$$p_{ji}^* = \frac{p_{ji}\pi_j}{\pi_i}$$

Demostración:

$$\begin{aligned} p_{ij}^* &= \mathbf{P}(X_n = j | X_{n+1} = i) \\ &= \frac{\mathbf{P}(X_{n+1} = i | X_n = j) \mathbf{P}(X_n = j)}{\mathbf{P}(X_{n+1} = i)} \\ &= \frac{p_{ji}\pi_j}{\pi_i} \end{aligned}$$

donde el Teorema de Bayes (Teorema 2.1.11) fue utilizado. ■

Definición 4.4.4 Una cadena de Markov \mathbf{X} con espacio de estados E es **reversible en el tiempo** si $p_{ij}^* = p_{ij}$ para toda $i, j \in E$.

Corolario 4.4.5 Una cadena es reversible si y solamente si $\pi_i p_{ij} = \pi_j p_{ji}$ para todos los estados $i, j \in E$.

Teorema 4.4.6 Una cadena de Markov estacionaria con espacio de estados E es reversible en el tiempo si y sólo si cualquier camino de regreso al estado de origen i tiene la misma probabilidad que el camino reverso. Es decir,

$$p_{i_1} p_{i_1 i_2} \cdots p_{i_k i} = p_{i_1} p_{i_1 i_2} \cdots p_{i_k i}, \quad \text{para toda } \{i, i_1, \dots, i_k\} \subseteq E. \quad (4.12)$$

Demostración: Si la cadena es reversible, entonces la ecuación (4.12) se cumple directamente por el Corolario 4.4.5.

Suponga ahora que la ecuación (4.12) se cumple. Sean i y j estados arbitrarios pero fijos, entonces se tiene que

$$p_{i_1} p_{i_1 i_2} \cdots p_{i_k j} p_{ji} = p_{ij} p_{ji} p_{i_k i_{k-1}} \cdots p_{i_1 i}.$$

Sumando ahora sobre todos los estados para cada i_n con $n \in \{1, 2, \dots, k\}$,

$$p_{ij}(k+1)p_{ji} = p_{ij}p_{ji}(k+1).$$

Es decir, la suma sobre todas las trayectorias posibles de n saltos desde el estado i al estado j es justamente $p_{ij}(n)$. Sumando ahora sobre k ,

$$p_{ji} \sum_{k=1}^n p_{ij}(k+1) = p_{ij} \sum_{k=1}^n p_{ji}(k+1)$$

y tomando el límite cuando $n \rightarrow \infty$, por el Teorema 4.3.2, se obtiene

$$p_{ji}\pi_j = p_{ij}\pi_i. \quad \blacksquare$$

Teorema 4.4.7 Dada una cadena irreducible \mathbf{X} con espacio de estados E , si existe π tal que

a) $0 \leq \pi_i \leq 1$

b) $\sum_i \pi_i = 1$

c) $\pi_i p_{ij} = \pi_j p_{ji}$

para todo $i, j \in E$, entonces la cadena es reversible en el tiempo y es positivo-recurrente con distribución estacionaria π .

Demostración: Sumando sobre todos los estados, se tiene que

$$\sum_{i \in E} \pi_i p_{ij} = \sum_{i \in E} \pi_j p_{ji} = \pi_j \sum_{i \in E} p_{ji} = \pi_j. \quad (4.13)$$

Note ahora que $a)$, $b)$ y la ecuación (4.13) cumplen con la Definición 4.2.1 y por lo tanto π es una distribución estacionaria. Por el Teorema 4.3.2, se tiene que la cadena es positivo-recurrente y es reversible en el tiempo gracias al Corolario 4.4.5. ■

Ejemplo 4.4.8 (Modelo de difusión de Ehrenfest) Dos contenedores A y B se unen por medio de una pequeña manguera por la cual se permite que gas pase de un contenedor a otro. En total, hay m moléculas de gas distribuidas entre los dos contenedores y en cada tiempo, una de las m moléculas, con probabilidad uniforme de ser elegida, se cambia de contenedor. Sea X_n la cantidad de moléculas en el contenedor A en el tiempo n . Claramente $X = \{X_0, X_1, \dots\}$ es una cadena de Markov homogénea irreducible ($i \leftrightarrow j$ para todos $0 \leq i, j \leq m$) con matriz de transición dada por

$$p_{i,j} = \begin{cases} \frac{i}{m} & \text{si } j = i - 1, \\ 1 - \frac{i}{m} & \text{si } j = i + 1. \end{cases}$$

Es fácil verificar que

$$\pi_i = \frac{\binom{m}{i}}{2^m}$$

es una solución para la ecuación

$$\pi_i p_{ij} = \pi_j p_{ji},$$

por lo que el Teorema 4.4.7 afirma que el modelo de difusión de Ehrenfest es una cadena de Markov reversible en el tiempo con distribución estacionaria π .

Capítulo 5

Procesos de Poisson

Informalmente, un proceso de Poisson, llamado así en honor al matemático francés Siméon-Denis Poisson, es un proceso estocástico que cuenta eventos que suceden de manera independiente. En este capítulo se presenta la definición formal de un proceso de Poisson junto con algunos teoremas útiles y una extensión a los procesos de Poisson no-homogéneos.

Un acercamiento más profundo al tema puede ser encontrado en Karlin y Taylor [38] y en Ross [26].

5.1. El Proceso de Poisson

Definición 5.1.1 Un proceso $\{N(t), t \geq 0\}$ es un **proceso de conteo** si $N(t)$ representa la cantidad total de eventos ocurridos hasta un tiempo t , es decir, satisface:

- a) $N(t) \in \mathbb{N}$,
- b) Si $s < t$ entonces $N(s) \leq N(t)$,
- c) Si $s < t$, entonces $N(t) - N(s)$ es la cantidad de eventos que han ocurrido en el intervalo $(s, t]$.

Definición 5.1.2 Un proceso de conteo $\{N(t), t \geq 0\}$ es de **incrementos independientes** si el número de eventos que ocurre en intervalos de tiempo disjuntos son independientes, es decir, para $s < t < r < u$, $N_t - N_s$ y $N_r - N_s$ son variables aleatorias independientes.

Definición 5.1.3 Un proceso de conteo $\{N(t), t \geq 0\}$ es de **incrementos estacionarios** si la distribución de la cantidad de eventos en un intervalo depende únicamente de la longitud del intervalo, es decir, la distribución de $N_t - N_s$ depende solamente de $t - s$.

Definición 5.1.4 Un proceso de conteo $\{N(t) : t \geq 0\}$ es un **proceso de Poisson** con tasa $\lambda > 0$ si

- a) $N(0) = 0$.
- b) El proceso tiene incrementos independientes y estacionarios.
- c) El número de eventos en cualquier intervalo de longitud t tiene distribución de Poisson con parámetro λ , es decir,

$$\mathbf{P}(N(t+s) - N(s) = n) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}, \quad n = 0, 1, 2, \dots$$

Observación: Las condiciones de incrementos independientes y estacionarios se pueden interpretar como el proceso reiniciándose cada vez que un evento sucede sin mantener memoria de la ocurrencia de los eventos anteriores.

Debido a la dificultad para mostrar la parte (c) de la definición 5.1.4 la siguiente equivalencia es de gran utilidad.

Definición 5.1.5 Una función f es $o(h)$ si

$$\lim_{h \rightarrow 0} \frac{f(h)}{h} = 0.$$

Teorema 5.1.6 Un proceso de conteo $\{N(t), t \geq 0\}$ es un proceso de Poisson con tasa $\lambda > 0$ si y sólo si

- a) $N(0) = 0$.
- b) El proceso es de incrementos estacionarios e independientes.
- c) $\mathbf{P}(N(h) = 1) = \lambda h + o(h)$.
- d) $\mathbf{P}(N(h) \geq 2) = o(h)$.

Demostración:

- a) Se demostrará primero que si N es un proceso de Poisson, entonces (a), (b), (c), y (d) son satisfechas.

Si N es un proceso de Poisson, por la definición 5.1.4 se tiene que (a) y (c) son inmediatos. Nuevamente, utilizando la definición 5.1.4 se tiene que

$$\mathbf{P}(N(h) = 1) = e^{-\lambda h} \frac{(\lambda h)^1}{1!} = \lambda h e^{-\lambda h} = \lambda h + \lambda h \left(\frac{1}{e^{\lambda h}} - 1 \right).$$

Note que

$$\lim_{h \rightarrow 0} \frac{\lambda h (\frac{1}{e^{\lambda h}} - 1)}{h} = \lim_{h \rightarrow 0} \lambda (\frac{1}{e^{\lambda h}} - 1) = 0,$$

y por lo tanto $\mathbf{P}(N(h) = 1) = \lambda h + o(h)$.

Ahora, también por la definición 5.1.4 se tiene que

$$\mathbf{P}(N(h) \geq 2) = \sum_{i=2}^{\infty} e^{-\lambda h} \frac{(\lambda h)^i}{i!} = e^{-\lambda h} \sum_{i=2}^{\infty} \frac{(\lambda h)^i}{i!}.$$

Para ver que en efecto $\mathbf{P}(N(h) \geq 2)$ es $o(h)$ note primero que

$$\sum_{i=1}^{\infty} \frac{(\lambda h)^i}{i!} = e^{\lambda h} - 1,$$

por lo tanto

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{e^{-\lambda h} \sum_{i=2}^{\infty} \frac{(\lambda h)^i}{i!}}{h} &= \lim_{h \rightarrow 0} \frac{e^{-\lambda h} (\sum_{i=1}^{\infty} \frac{(\lambda h)^i}{i!} - \lambda h)}{h} \\ &= \lim_{h \rightarrow 0} \frac{e^{-\lambda h} (e^{\lambda h} - 1 - \lambda h)}{h} \\ &= \lim_{h \rightarrow 0} \frac{1 - e^{-\lambda h} - \lambda h e^{-\lambda h}}{h}. \end{aligned}$$

Como el último límite es la forma $\frac{0}{0}$ entonces se puede utilizar la regla de L'Hopital y por lo tanto

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{1 - e^{-\lambda h} - \lambda h e^{-\lambda h}}{h} &= \lim_{h \rightarrow 0} \frac{d(1 - e^{-\lambda h} - \lambda h e^{-\lambda h})/dh}{d(h)/dh} \\ &= \lim_{h \rightarrow 0} \frac{\lambda e^{-\lambda h} - \lambda e^{-\lambda h} + \lambda^2 h e^{-\lambda h}}{1} \\ &= 0. \end{aligned}$$

- b) Considere ahora que (a), (b), (c) y (d) son verdaderas, se demostrará que N es un proceso de Poisson. La demostración es por inducción. Primero obtendremos algunas ecuaciones que serán de utilidad.

Defina

$$P_n(t) = \mathbf{P}(N(t) = n),$$

entonces notamos

$$\begin{aligned}
 P_0(t+h) &= \mathbf{P}(N(t+h) = 0) \\
 &= \mathbf{P}(N(t) = 0, N(t+h) - N(t) = 0) \\
 &= \mathbf{P}(N(t) = 0)\mathbf{P}(N(t+h) - N(t) = 0) \\
 &= \mathbf{P}(N(t) = 0)\mathbf{P}(N(h) = 0) \\
 &= \mathbf{P}_0(t)\mathbf{P}_0(h),
 \end{aligned}$$

donde en la tercera y cuarta igualdad se utiliza la afirmación (b).

Ahora

$$\mathbf{P}(N(h) \geq 2) = 1 - \mathbf{P}(N(h) = 1) - \mathbf{P}(N(h) = 0),$$

utilizando (c) y (d) obtenemos

$$o(h) = 1 - \lambda h - o(h) - \mathbf{P}(N(h) = 0).$$

Como el conjunto de las funciones $o(h)$ es cerrado bajo sumas entonces

$$P_0(h) = 1 - \lambda h + o(h)$$

y por lo tanto

$$P_0(t+h) = P_0(t)[1 - \lambda h + o(h)]$$

de donde se deduce

$$\begin{aligned}
 \frac{P_0(t+h) - P_0(t)}{h} &= \frac{P_0(t)[1 - \lambda h + o(h)] - P_0(t)}{h} \\
 &= \frac{P_0(t)[- \lambda h + o(h)]}{h} \\
 &= -\lambda P_0(t) + \frac{P_0(t)o(h)}{h} \\
 &= -\lambda P_0(t) + \frac{o(h)}{h}.
 \end{aligned}$$

Haciendo $h \rightarrow 0$ se tiene que

$$P_0' = -\lambda P_0(t),$$

o sea,

$$\frac{P_0'(t)}{P_0(t)} = -\lambda. \quad (5.1)$$

Integrando (5.1) respecto a t obtenemos

$$\log P_0(t) = -\lambda t + c,$$

o equivalentemente

$$P_0(t) = e^{-\lambda t + c} = e^{-\lambda t} e^c = ce^{-\lambda t}.$$

Por la afirmación (a), se tiene que $P_0(0) = 1$, y se sigue que $c = 1$ y así

$$P_0(t) = e^{-\lambda t}. \quad (5.2)$$

Para el caso en que $n \geq 1$, utilizando un razonamiento análogo se tiene

$$\begin{aligned} P_n(t+h) &= \mathbf{P}(N(t+h) = n) \\ &= \mathbf{P}(N(t) = n, N(t+h) - N(t) = 0) \\ &\quad + \mathbf{P}(N(t) = n-1, N(t+h) - N(t) = 1) \\ &\quad + \mathbf{P}(N(t+h) = n, N(t+h) - N(t) \geq 2). \end{aligned}$$

Por (d), el último término es $o(h)$ y así

$$\begin{aligned} P_n(t+h) &= P_n(t)P_0(h) + P_{n-1}(t)P_1(h) + o(h) \\ &= (1 - \lambda h)P_n(t) + \lambda h P_{n-1}(t) + o(h). \end{aligned}$$

Por lo tanto

$$\frac{P_n(t+h) - P_n(t)}{h} = -\lambda P_n(t) + \lambda P_{n-1}(t) + \frac{o(h)}{h}.$$

Tomando el límite cuando $h \rightarrow 0$

$$P_n'(t) = -\lambda P_n(t) + \lambda P_{n-1}(t).$$

Reacomodando los términos y multiplicando por $e^{\lambda t}$ se obtiene que

$$e^{\lambda t}[P_n'(t) + \lambda P_n(t)] = \lambda e^{\lambda t} P_{n-1}(t),$$

y por lo tanto

$$\frac{d}{dt}(e^{\lambda t} P_n(t)) = \lambda e^{\lambda t} P_{n-1}(t). \quad (5.3)$$

Como base de inducción, cuando $n = 1$, por (5.2)

$$\frac{d}{dt}(e^{\lambda t} P_1(t)) = \lambda,$$

o equivalentemente,

$$P_1(t) = (\lambda t + c)e^{-\lambda t}.$$

Como $P_1(0) = 0$, entonces $c = 0$ y así

$$P_1(t) = \lambda t e^{-\lambda t}.$$

Suponga ahora que $P_{n-1}(t) = e^{-\lambda t}(\lambda t)^{n-1}/(n-1)!$ es verdadero, entonces por (5.3),

$$\frac{d}{dt}(e^{\lambda t} P_n(t)) = \frac{\lambda(\lambda t)^{n-1}}{(n-1)!},$$

o equivalentemente,

$$e^{\lambda t} P_n(t) = \frac{(\lambda t)^n}{n!} + c.$$

Como $P_n(0) = 0$ entonces $c = 0$ y así

$$P_n(t) = e^{-\lambda t} \frac{(\lambda t)^n}{n!},$$

para $n = 1, 2, 3, \dots$. Por lo tanto, $N = \{N(t), t \geq 0\}$ que satisface (a), (b), (c), y (d) es un proceso de Poisson. ■

Notación: Sea $N = \{N(t), t \geq 0\}$ un proceso de Poisson, llamamos X_1 al tiempo en el que ocurre el primer evento, similarmente llamamos $X_n, n \geq 2$ al tiempo entre el $(n-1)$ -ésimo evento y el n -ésimo.

Teorema 5.1.7 $\{X_n\}$ es un conjunto de variables aleatorias exponenciales independientes idénticamente distribuidas con tasa λ

Demostración: Por tratarse de un proceso de Poisson las variables son de incrementos estacionarios e independientes. Ahora

$$\mathbf{P}(X_1 > t) = \mathbf{P}(N(t) = 0) = e^{-\lambda t}.$$

Utilizando inducción y el hecho de que los procesos de Poisson tienen incrementos independientes y estacionarios

$$\begin{aligned} \mathbf{P}(X_n > t | X_{n-1} = s) &= \mathbf{P}(N(s+t) - N(s) = 0 | X_{n-1} = s) \\ &= \mathbf{P}(N(s+t) - N(s) = 0) \\ &= \mathbf{P}(N(t) = 0) = e^{-\lambda t}. \end{aligned}$$
■

5.2. Procesos de Poisson No-homogéneos

Definición 5.2.1 El proceso de conteo $\{N(t), t \geq 0\}$ es un proceso de Poisson **no-homogéneo** con función de intensidad $\lambda(t) > 0, t \geq 0$ si

- a) $N(0) = 0$.
- b) $\{N(t), t \geq 0\}$ es de incrementos independientes.
- c) $\mathbf{P}(N(t+h) - N(t) = 1) = \lambda(t)h + o(h)$.
- d) $\mathbf{P}(N(t+h) - N(t) \geq 2) = o(h)$.

Teorema 5.2.2 Sea

$$m(t) = \int_0^t \lambda(s) ds,$$

entonces

$$\mathbf{P}(N(t+s) - N(t) = n) = e^{-[m(t+s)-m(t)]} \frac{[m(t+s) - m(t)]^n}{n!}. \quad (5.4)$$

Demostración: El esquema general de la demostración es muy similar a la del Teorema 5.1.6.

Debido a que la aportación es puramente técnica, la demostración se omite, salvo su primera parte la cual contiene cambios sustanciales. Dado t , sea

$$P_n(s) = \mathbf{P}(N(t+s) - N(t) = n).$$

de donde deducimos

$$\begin{aligned} P_0(s+h) &= \mathbf{P}(N(t+s+h) - N(t) = 0) \\ &= \mathbf{P}(0 \text{ eventos en } [t, t+s] \text{ y en } [t+s, t+s+h]) \\ &= \mathbf{P}(0 \text{ eventos en } [t, t+s]) \mathbf{P}(0 \text{ eventos en } [t+s, t+s+h]) \\ &= P_0(s) \mathbf{P}(0 \text{ eventos en } [t+s, t+s+h]) \\ &= P_0(s) [1 - \lambda(t+s)h + o(h)], \end{aligned}$$

y por lo tanto

$$\frac{P_0(s+h) - P_0(s)}{h} = -\lambda(t+s)P_0(s) + \frac{o(h)}{h},$$

tomando el límite cuando $h \rightarrow 0$

$$P_0'(s) = -\lambda(t+s)P_0(s).$$

Reacomodando los términos e integrando respecto a s se obtiene

$$\log P_0(s) = - \int_0^s \lambda(t+u) du,$$

y así

$$P_0(s) = e^{-[m(t+s)-m(t)]}.$$

A partir de aquí los pasos son análogos a la demostración del Teorema 5.1.6. ■

Capítulo 6

Funciones de Riesgo y Supervivencia

En este capítulo se presenta una breve introducción a algunos conceptos básicos de donde surge la elección de las funciones de intensidad que serán utilizadas más tarde. Una introducción más extensa está disponible en [18].

Considere una variable aleatoria no-negativa T que representa el tiempo de vida de individuos en una cierta población. Suponga además que T es continua con función de densidad $f(t)$.

Definición 6.0.3 La función de **supervivencia** de la variable aleatoria T es la probabilidad de que un individuo sobreviva hasta el tiempo t , es decir

$$S(t) = \mathbf{P}(T \geq t) = \int_t^{\infty} f(x)dx. \quad (6.1)$$

Claramente $S(t)$ es una función continua monótona decreciente tal que $S(0) = 1$ y $S(\infty) = \lim_{t \rightarrow \infty} S(t) = 0$.

Definición 6.0.4 La función de **riesgo** $h(t)$ se define como

$$\begin{aligned} h(t) &= \lim_{\Delta t \rightarrow 0} \frac{\mathbf{P}(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{\mathbf{P}(t \leq T < t + \Delta t)}{(\Delta t)\mathbf{P}(T \geq t)} \\ &= \lim_{\Delta t \rightarrow 0} \frac{S(t) - S(t + \Delta t)}{(\Delta t)S(t)} \\ &= \frac{f(t)}{S(t)}. \end{aligned}$$

Observación: La función de riesgo es la tasa instantánea de muerte en el tiempo t dado que el individuo sobrevivió hasta el tiempo t . Más aún $h(t)\Delta t$

es la probabilidad aproximada de muerte en el intervalo $[t, t + \Delta t)$ una vez garantizada la supervivencia hasta el tiempo t .

Observación: La función de probabilidad $f(t)$, la función de densidad $F(t)$, la función de supervivencia $S(t)$ y la función de riesgo $h(t)$ dan descripciones equivalentes acerca de la distribución de la variable aleatoria en cuestión. En otras palabras, a partir de cualquiera de las funciones anteriores se pueden deducir las demás.

6.1. Algunos modelos importantes

La motivación para utilizar un modelo sobre otro está comúnmente fundamentada de manera empírica utilizando información sobre el proceso de las fallas reduciendo así las opciones. Algunos modelos comunes, los cuales serán utilizados más tarde, se presentan a continuación.

Para facilitar la comprensión de las funciones de riesgo, las figuras 6.1, 6.2 y 6.3 muestran el comportamiento de éstas con distintos valores de parámetros. Como estas funciones son las utilizadas para modelar los niveles de ozono en la Ciudad de México, es importante notar que bajo ciertos parámetros es posible obtener formas decrecientes. Esta cualidad es de suma importancia debido a que la cantidad de excedencias ozono ha ido disminuyendo estrictamente durante los últimos 10 años.

6.1.1. Weibull

La distribución de Weibull es probablemente el modelo de distribución de vida más utilizado. Su función de riesgo es de la forma

$$h(t) = \frac{\alpha}{\sigma} \left(\frac{t}{\sigma} \right)^{(\alpha-1)} \quad (6.2)$$

donde $\alpha, \sigma > 0$ son parámetros reales.

Sus funciones de densidad y de supervivencia son

$$f(t) = \frac{\alpha}{\sigma} \left(\frac{t}{\sigma} \right)^{(\alpha-1)} \exp\left[- \left(\frac{t}{\sigma} \right)^\alpha \right]$$

y

$$S(t) = e^{-(t/\sigma)^\alpha}$$

respectivamente.

La función de riesgo de Weibull es monótona creciente si $\alpha > 1$, decreciente si $\alpha < 1$, y constante si $\alpha = 1$. En la figura 6.1 se presentan algunos ejemplos del comportamiento de la función de riesgo Weibull.

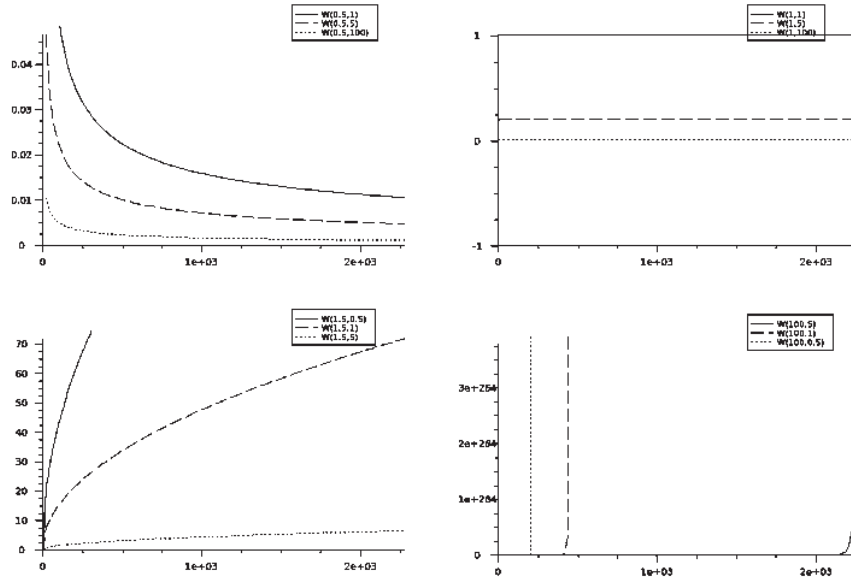


Figura 6.1: Ejemplos del comportamiento de $h(t)$ para diferentes valores de α y σ

6.1.2. Musa-Okumoto

La forma de Musa-Okumoto está dada por la función de riesgo

$$h(t) = \frac{\beta}{t + \alpha}, \alpha, \beta > 0.$$

Su función de distribución es

$$f(t) = \frac{\beta}{(t + \alpha)(t/\alpha + 1)^\beta},$$

y su función de supervivencia

$$S(t) = \frac{1}{(t/\alpha + 1)^\beta}.$$

En la figura 6.2 se presentan algunos ejemplos gráficos del comportamiento de $h(t)$ Musa-Okumoto para diferentes valores de α y β .

6.1.3. Goel-Okumoto generalizada

La forma de Goel-Okumoto generalizada es una función dependiente de tres parámetros reales $\alpha, \beta, \gamma > 0$. Su función de riesgo es

$$h(t) = \alpha\beta\gamma t^{\gamma-1} e^{-\beta t^\gamma}.$$

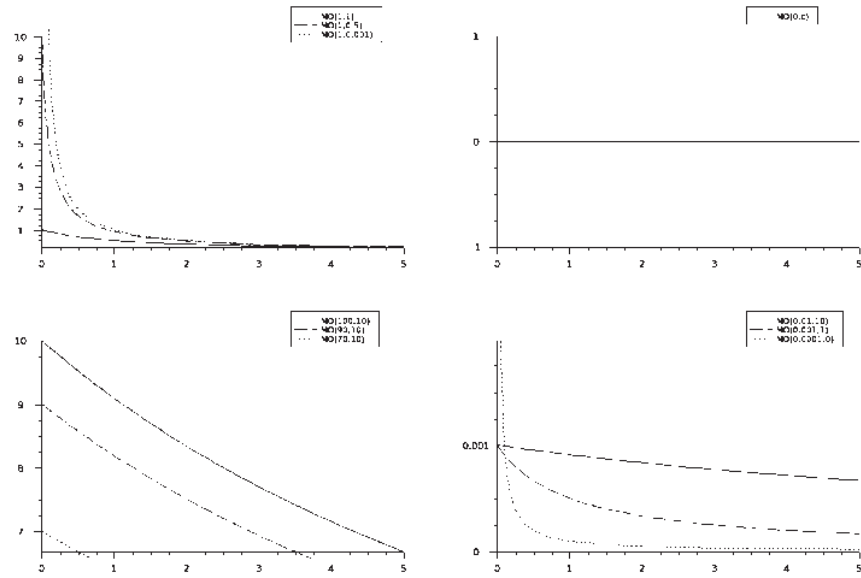


Figura 6.2: gráficas de la función de riesgo Musa-Okumoto para diferentes valores de α y β .

Sus funciones de distribución y de supervivencia son

$$f(t) = \alpha\beta\gamma t^{\gamma-1} e^{-\beta t^\gamma + \alpha(1 - \exp(-\beta t^\gamma))}.$$

y

$$S(t) = e^{\alpha(1 - \exp(-\beta t^\gamma))}$$

respectivamente. En la figura 6.3 se tienen ejemplos del comportamiento de la función de riesgo Goel-Okumoto generalizada para diferentes valores de α , β , y σ .

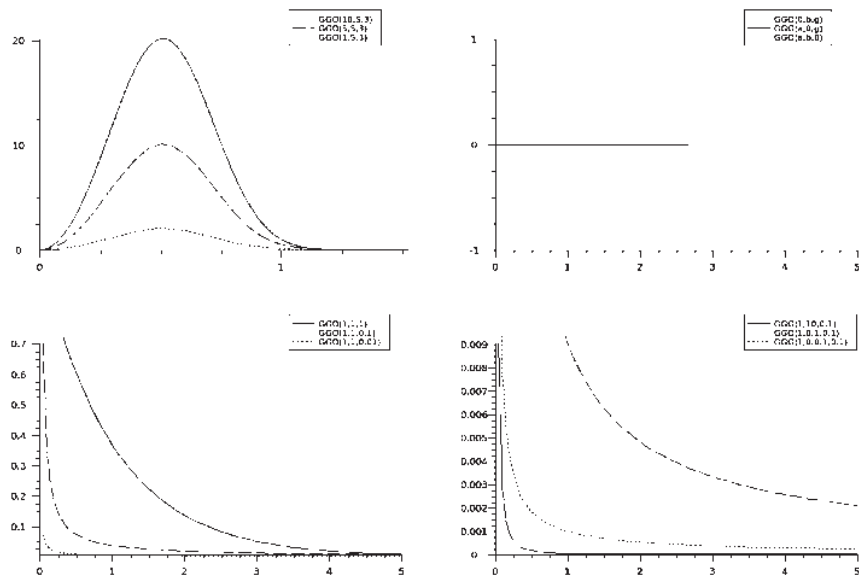


Figura 6.3: Gráficas de la función de riesgo Goel-Okumoto generalizada para diferentes valores de α , β , y σ .

Capítulo 7

Método Monte Carlo vía Cadenas de Markov

Este capítulo, el cual se basa en los resultados del Capítulo 4, contiene la descripción de una clase de algoritmos probabilísticos llamados métodos de Monte Carlo vía cadenas de Markov (MCMC, por sus siglas en inglés) junto con una serie de resultados que garantizan su funcionamiento.

Antes del uso de los MCMC, tomar un acercamiento Bayesiano (capítulo 3) muchas veces resultaba de poco uso debido a que las distribuciones a posteriori no siempre toman formas cerradas y obtener estadísticas del modelo puede ser muy complicado analíticamente. Los algoritmos MCMC resuelven este problema construyendo una cadena de Markov con distribución estacionaria igual a la distribución a posteriori, lo cual permite interpretar los estados de la cadena como muestras de la distribución una vez que la convergencia se alcanzó.

Los algoritmos descritos y variaciones de éstos pueden ser encontrados en Gamerman y Lopes [37], Kroese y Rubinstein [35], Richardson y Spiegelhater [34], Rosenthal [25] y Romero-Hidalgo [24].

7.1. Método Monte Carlo

Teorema 7.1.1 (Desigualdad de Chebyshev) *Sea X una variable aleatoria discreta con esperanza μ y variancia σ^2 . Entonces para toda $\epsilon > 0$,*

$$\mathbf{P}(|x - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}.$$

Demostración: Sea $\epsilon > 0$, entonces por definición

$$\mathbf{P}(|x - \mu| \geq \epsilon) = \sum_{\{x: |x - \mu| \geq \epsilon\}} f(x),$$

donde f es la función de probabilidad de \mathbf{X} . Por lo tanto

$$\begin{aligned} \epsilon^2 \mathbf{P}(|x - \mu| \geq \epsilon) &= \sum_{|x - \mu| \geq \epsilon} \epsilon^2 f(x) \\ &\leq \sum_{\{x: |x - \mu| \geq \epsilon\}} (x - \mu)^2 f(x) \\ &\leq \sum_{x \in \mathbb{R}} (x - \mu)^2 f(x) \\ &= \sigma^2. \end{aligned}$$

■

Lema 7.1.2 Sea $X = \{X_t : t \in \mathbb{N}\}$ un proceso numerable de variables independientes con esperanza μ y variancia σ^2 cada uno. Sea $S_n = X_1 + X_2 + \dots + X_n$ y $A_n = \frac{S_n}{n}$, entonces:

a) $\text{Var}(S_n) = n\sigma^2$,

b) $\text{Var}(A_n) = \frac{\sigma^2}{n}$.

Demostración:

a) $\text{Var}(S_n) = \text{Var}(X_1 + X_2 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n) = n\sigma^2$

b) $\text{Var}\left(\frac{S_n}{n}\right) = \frac{\text{Var}(S_n)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$

■

Teorema 7.1.3 (Ley de los Grandes Números) Sea $X = \{X_t : t \in \mathbb{N}\}$ un proceso numerable de variables independientes con esperanza μ y variancia σ^2 . Sea S_n igual al lema anterior. Entonces para todo $\epsilon > 0$,

$$\mathbf{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) \rightarrow 0$$

o equivalentemente

$$\mathbf{P}\left(\left|\frac{S_n}{n} - \mu\right| < \epsilon\right) \rightarrow 1$$

conforme $n \rightarrow \infty$.

Demostración: Sea $\epsilon > 0$, utilizando la desigualdad de Chebyshev (Lema 7.1.1) y la parte (b) del Lema 7.1.2, se obtiene

$$\mathbf{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2},$$

pues la variable aleatoria S_n tiene variancia $\frac{\sigma^2}{n}$. Por lo tanto cuando $n \rightarrow \infty$,

$$\mathbf{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) \rightarrow 0.$$

■

La ley de los grandes números afirma que dada una variable aleatoria X con función de probabilidad f , el valor esperado de X puede ser aproximado por medio de la media empírica de muestras obtenidas a partir de f , es decir,

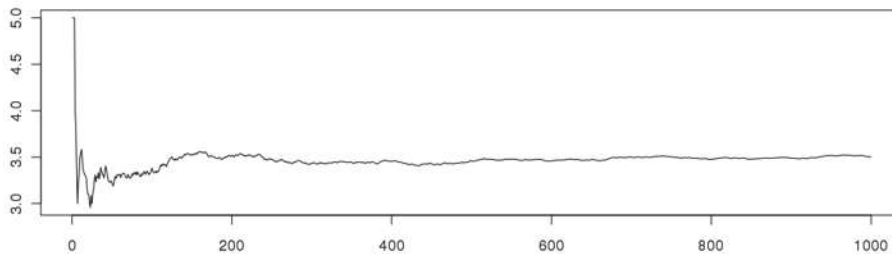
$$E[X] \approx \frac{1}{N} \sum_{i=1}^N f(x_i). \quad (7.1)$$

Definición 7.1.4 Sea X una variable aleatoria con función de probabilidad f . La aproximación de la esperanza de X a partir de una muestra aleatoria $x = (x_1, x_2, \dots, x_N)$ por medio de

$$\frac{1}{N} \sum_{i=1}^N f(x_i)$$

es llamado una **integración Monte Carlo**.

Ejemplo 7.1.5 Suponga que se tiene un dado y se quiere saber el valor esperado de la cara superior al lanzarlo. Si no se sabe con certeza la probabilidad de que cada cara aparezca, entonces la esperanza no puede ser determinada analíticamente. La solución utilizando la Ley de los Grandes números consiste en lanzar el dado continuamente e ir calculando el valor medio empírico de los resultados observados. A continuación se muestra el resultado de una simulación del experimento de lanzar 1000 veces un dado e ir calculando la media observada, la cual finalmente fue 3.502.



Gracias a la Ley de los Grandes números, si obtener la media o cualquier otra estadística dependiente de la media es demasiado complicado, una buena aproximación puede ser obtenida por medio de una muestra lo suficientemente grande utilizando métodos Monte Carlo. De esta manera, el problema de estimar las estadísticas de una variable aleatoria, es reducible a encontrar un método para obtener una muestra aleatoria de la variable en cuestión.

7.2. Muestreador de Gibbs

Suponga que se quiere generar una muestra aleatoria con distribución $\pi(\theta)$, donde $\theta = (\theta_1, \theta_2, \dots, \theta_n) \in \mathbb{R}^n$ es un vector de parámetros. Adicionalmente, suponga que se conocen todas las distribuciones condicionales

$$\mathbf{P}(\theta_i | \theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_n), \quad i \in \{1, \dots, n\}.$$

En muchas ocasiones la generación directa de muestras de una distribución π es complicada, computacionalmente costosa o imposible. Si la generación de muestras de las distribuciones condicionales es viable, el muestreador de Gibbs soluciona el problema con un esquema de muestreo indirecto. El algoritmo se describe a continuación:

- 1) Generar un vector aleatorio $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_n^{(0)})$ llamado vector inicial.
- 2) Dado el vector $\theta^{(j-1)}$ en la $(j-1)$ -ésima iteración del algoritmo, generar n muestras aleatorias de la siguiente manera:

$$\begin{aligned} \theta_1^{(j)} &\leftarrow \pi(\theta_1 | \theta_2^{(j-1)}, \theta_3^{(j-1)}, \dots, \theta_n^{(j-1)}), \\ \theta_2^{(j)} &\leftarrow \pi(\theta_2 | \theta_1^{(j)}, \theta_3^{(j-1)}, \dots, \theta_n^{(j-1)}), \\ &\vdots \\ \theta_i^{(j)} &\leftarrow \pi(\theta_i | \theta_1^{(j)}, \theta_2^{(j)}, \dots, \theta_{i-1}^{(j)}, \theta_{i+1}^{(j-1)}, \dots, \theta_n^{(j-1)}), \\ &\vdots \\ \theta_n^{(j)} &\leftarrow \pi(\theta_n | \theta_1^{(j)}, \theta_2^{(j)}, \dots, \theta_{n-1}^{(j)}). \end{aligned}$$

- 3) Tomar $\theta^{(j)} = (\theta_1^{(j)}, \theta_2^{(j)}, \dots, \theta_n^{(j)})$ y volver al paso 2.

En otras palabras, el muestreador de Gibbs comienza a generar muestras a partir de las distribuciones condicionales, cambiando el valor de cada parámetro a partir de su distribución condicional dado el valor actual del resto de los parámetros.

Claramente el muestreador de Gibbs define una cadena de Markov homogénea con matriz de transición

$$p_{\theta\theta'} = p(\theta, \theta') = \prod_{k=1}^n \pi(\theta_k | \theta_1, \theta_2, \dots, \theta_{k-1}, \theta'_{k+1}, \dots, \theta'_n).$$

Como se verá en la sección 7.3.2, el muestreador de Gibbs es un caso particular del algoritmo de Metropolis-Hastings y por lo tanto si el muestreador de Gibbs es irreducible entonces la prueba de que el algoritmo de Metropolis-Hastings converge a su distribución estacionaria será suficiente para asegurar la convergencia del muestreador de Gibbs.

Si el muestreador de Gibbs converge a alguna distribución, entonces la distribución de convergencia será precisamente la distribución estacionaria $\pi(\theta)$. Sin embargo, la cadena construida por el muestreador de Gibbs puede no ser irreducible en cuyo caso, como se muestra en el siguiente ejemplo, no convergerá.

Ejemplo 7.2.1 En un centro meteorológico el clima es descrito por medio del vector de parámetros $\theta = (s, n) \in \{0, 1\}^2$, donde s y n son valores booleanos que indican si hay sol y si está nublado respectivamente. Suponga que la función de distribución f está dada por

$$\begin{aligned} f(s = 1, n = 0) &= \frac{1}{2}, & f(s = 1, n = 1) &= 0, \\ f(s = 0, n = 0) &= 0, & f(s = 0, n = 1) &= \frac{1}{2}. \end{aligned}$$

Si se quisieran tomar muestras de esta distribución entonces el muestreador de Gibbs sería inútil. Para ver esto solo hay que notar que si el vector inicial es $\theta = (s = 1, n = 0)$ entonces las distribuciones condicionales serían

$$\begin{aligned} f(s = 0 | n = 0) &= 0, & f(s = 1 | n = 1) &= 1, \\ f(n = 0 | s = 1) &= 1, & f(n = 1 | s = 1) &= 0, \end{aligned}$$

por lo que el muestreador de Gibbs se quedaría siempre en el vector inicial. Similarmente, si el vector inicial se tomara como $\pi = (s = 0, n = 1)$, el muestreador sería incapaz de producir muestras distintas.

7.2.1. Implementación

La implementación directa para obtener n muestras de la distribución π consiste en generar n cadenas de Markov a partir del muestreador de Gibbs y dejarlas correr por un tiempo de calentamiento m (el tiempo que tardan en alcanzar la distribución estacionaria), teniendo un costo total de mn muestreos.

De manera alterna, se puede generar una sola cadena y después del tiempo de calentamiento tomar los siguientes n valores, reduciendo así el costo a $m + n$, sin embargo, los elementos de la muestra aleatoria no serían independientes, pues la cadena de Markov utilizada no es independiente.

Para formar una muestra semi-independiente se pueden tomar las muestras con una distancia k , con un costo de $m + nk$, sin embargo si k es muy grande ($k > m(n - 1)/n$), utilizar una sola cadena deja de ser útil.

7.3. Algoritmo de Metropolis-Hastings

Sea X una variable aleatoria con función de probabilidad π . Suponga nuevamente que la tarea de muestrear directamente a partir de la distribución es complicada, costosa o imposible. Si existe una cadena de Markov Y con espacio de estados E igual al dominio de X , es decir $\{s : X(s) > 0\} = E$, el algoritmo de Metropolis-Hastings propone una solución indirecta. El algoritmo de Metropolis-Hastings contruye una muestra aleatoria de X a partir del comportamiento de Y por medio de una cadena de Markov y un método de aceptación y rechazo. El algoritmo de describe a continuación:

Sea $q_{ij} = \mathbf{P}(Y_1 = j | Y_0 = i)$ la probabilidad de transición de Y a seguir de un estado i a un estado j y Z la cadena de Markov que será construida.

- 1) Elegir un estado inicial aleatorio, denotado por $Z_1 = z_0$.
- 2) Dado el estado actual $Z_n = \zeta$, genere un estado ζ' de la distribución $\mathbf{P}(Y_{n+1} = \zeta' | Y_n = \zeta)$.
- 3) Tomar

$$Z_{n+1} = \begin{cases} \zeta' & \text{con probabilidad } \alpha(\zeta, \zeta') = \min \left\{ \frac{\pi(\zeta')q_{\zeta'\zeta}}{\pi(\zeta)q_{\zeta\zeta'}}, 1 \right\}, \\ \zeta & \text{con probabilidad } 1 - \alpha(\zeta, \zeta'). \end{cases}$$

- 4) Repetir paso 2)
-

En otras palabras, para cada tiempo de la cadena Y , la cadena de Markov construida Z , copia el estado que Y tomó, o se mantiene en su mismo estado. La probabilidad con la cual Z copiará el estado de Y es

$$\min \left\{ \frac{\pi(\zeta')q_{\zeta'\zeta}}{\pi(\zeta)q_{\zeta\zeta'}}, 1 \right\}$$

es llamada probabilidad de aceptación.

Observación: El algoritmo de Metropolis-Hastings permite construir la muestra aleatoria a partir de una función de probabilidad más sencilla. Si se quiere generar una muestra aleatoria de la función de probabilidad π a

partir de la función de probabilidad f , es suficiente utilizar la cadena de Markov Y con probabilidad de transición $q_{ij} = f(j)$.

Como se verá en el siguiente Teorema, el algoritmo de Metropolis-Hastings construye una cadena de Markov con distribución estacionaria π y garantiza la convergencia a la distribución. Por lo tanto, si se quiere obtener una muestra aleatoria de π , es suficiente correr el algoritmo de Metropolis-Hastings por un tiempo de calentamiento (para que alcance la convergencia a su distribución estacionaria) y después tomar los estados como la muestra aleatoria buscada.

Teorema 7.3.1 *La cadena de Markov Z construida por el algoritmo de Metropolis-Hastings converge a su distribución estacionaria π .*

Demostración: Claramente, la cadena formada por Metropolis-Hastings es una cadena de Markov con matriz de transición

$$p_{ij} = \begin{cases} q_{ij}\alpha(i, j) & \text{si } i \neq j, \\ 1 - \sum_{k \neq i} q_{ik}\alpha(ik) & \text{si } i = j. \end{cases}$$

Se demostrará ahora que la cadena es reversible. Por ser π la función de probabilidad de la variable aleatoria X , entonces

a) $0 \leq \pi_i \leq 1$,

b) $\sum_i \pi_i = 1$.

Para ver que la última propiedad

c) $\pi_i q_{ij} = \pi_j q_{ji}$

se cumple, suponga que

$$\frac{\pi_j q_{ji}}{\pi_i q_{ij}} < 1.$$

Entonces,

$$\begin{aligned} \pi_i p_{ij} &= (\pi_i q_{ij})\alpha(i, j) \\ &= \pi_i q_{ij} \frac{\pi_j q_{ji}}{\pi_i q_{ij}} \\ &= \pi_j q_{ji} \\ &= \pi_j q_{ji} \min \left\{ \frac{\pi_i q_{ij}}{\pi_j q_{ji}}, 1 \right\} \\ &= \pi_j p_{ji}. \end{aligned}$$

Si

$$\frac{\pi_j q_{ji}}{\pi_i q_{ij}} > 1, \quad \text{entonces} \quad \frac{\pi_i q_{ij}}{\pi_j q_{ji}} < 1,$$

y por un razonamiento idéntico al anterior, se sigue nuevamente que $\pi_j p_{ji} = \pi_i p_{ij}$.

Por último, el caso en que

$$\frac{\pi_j q_{ji}}{\pi_i q_{ij}} = 1$$

es directo. Con esto, por el Teorema 4.4.7 se concluye que la cadena es positivo-recurrente y reversible con distribución estacionaria π .

Por último, como Y es una cadena irreducible, entonces para cualquier pareja de estados i, j ,

$$\pi_i q_{ij} > 0,$$

y por lo tanto

$$\min \left\{ 1, \frac{\pi_i q_{ij}}{\pi_j q_{ji}} \right\} > 0.$$

Con esto, se tiene que la cadena que construye el algoritmo de Metropolis-Hastings es reversible en el tiempo, aperiódica y positivo-recurrente con distribución estacionaria π por lo que el Teorema 4.3.2 asegurará la convergencia de la cadena a su distribución estacionaria. ■

7.3.1. Matrices de transición

La elección de la matriz de transición de Y es arbitraria, sin embargo, es importante que la elección sea de tal forma que simular la ejecución de la cadena Y sea sencillo para así evadir la dificultad del muestreo directo de π . También se debe notar que la velocidad con la que la cadena generada convergerá a la distribución buscada es dependiente de la matriz de transición de la cadena Y .

A continuación se presentan las elecciones más comunes para el comportamiento de Y :

a) Cadena Simétrica

Una cadena es simétrica si $q_{ij} = q_{ji}$, $i, j \in S$. La ventaja de tener una matriz de transición Y simétrica se debe a que la probabilidad de aceptación se reduce a

$$\alpha(i, j) = \min \left\{ 1, \frac{\pi_j}{\pi_i} \right\}.$$

b) Cadena de Caminata aleatoria

Formalmente, una caminata aleatoria es una cadena determinada por

$$X_t = X_{t-1} + w_t,$$

donde w_t es una variable aleatoria con distribución f_w independiente de la cadena. La propuesta de Y acorde a la caminata aleatoria está dada por

$$q_{ij} = f_w(j - i).$$

Las elecciones más comunes para f_w son la distribución Normal y la distribución T, sin embargo, esto da raíz al problema de la elección de la variancia σ^2 ; valores altos permitirán movimientos distantes en la cadena con probabilidades de aceptación muy bajas mientras que valores bajos permiten cambios cercanos con probabilidades de aceptación muy altos.

c) Cadenas Independientes

La idea consiste en elegir Y de modo que sea independiente del estado actual, es decir,

$$q_{ij} = f(j),$$

y así

$$\frac{\pi_j q_{ji}}{\pi_i q_{ij}} = \frac{(\pi_j / q_{ij})}{(\pi_i / q_{ji})} = \frac{(\pi_j / f(j))}{(\pi_i / f(i))}.$$

Observación: Cualquier muestreo de π a partir de una función de probabilidad distinta se hace a partir de una cadena independiente.

Ejemplo 7.3.2 Suponga que se quiere generar una muestra aleatoria de una distribución Normal con media 0 y variancia 1, entonces la función de probabilidad de la cual se quiere generar una muestra es

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Sea Y la cadena de Markov a partir de la cual se construirá la muestra buscada. Para la matriz de transición de Y considere una distribución uniforme en el intervalo $[-5, 5]$. En otras palabras, la muestra de la distribución normal será producida a partir de una muestra de una distribución uniforme.

Note que $q_{ij} = 1/10$ para cualesquiera estados y por lo tanto, dado el estado $Y_{n-1} = \zeta_1$ actual y el estado propuesto $Y_n = \zeta_2$, la probabilidad de aceptación del algoritmo de Metropolis-Hastings será

$$\begin{aligned} \alpha(\zeta_1, \zeta_2) &= \min \left\{ 1, \frac{\pi(\zeta_2) q_{\zeta_2 \zeta_1}}{\pi(\zeta_1) q_{\zeta_1 \zeta_2}} \right\} \\ &= \min \left\{ 1, \frac{\frac{1}{\sqrt{2\pi}} e^{-\frac{\zeta_2^2}{2}}}{\frac{1}{\sqrt{2\pi}} e^{-\frac{\zeta_1^2}{2}}} \right\} \\ &= \min \left\{ 1, \frac{e^{-\frac{\zeta_2^2}{2}}}{e^{-\frac{\zeta_1^2}{2}}} \right\} \\ &= \min \left\{ 1, \exp \left(\frac{\zeta_1^2 - \zeta_2^2}{2} \right) \right\}. \end{aligned}$$

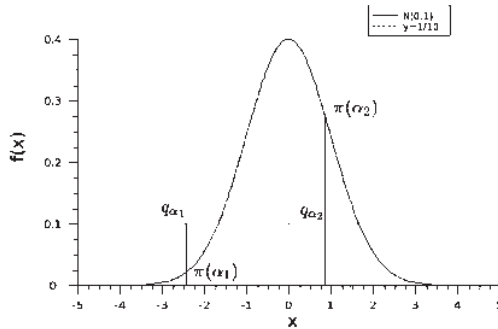


Figura 7.1: Representación gráfica del ejemplo 7.3.2, los valores propuestos son generados a partir de la distribución uniforme y la probabilidad de aceptación es proporcional a la probabilidad de la distribución Normal buscada.

Note que el algoritmo aceptará un valor propuesto si éste es más probable que el estado actual. Si el valor propuesto es menos probable que el estado actual, su probabilidad de aceptación será justamente la proporción de las probabilidades.

En la figura 7.2 se muestra la frecuencia de los valores generados por el algoritmo de Metropolis-Hastings después de 100, 1000, 10000 y 100000 iteraciones, descartando la primera mitad de las iteraciones como el tiempo de calentamiento.

7.3.2. El muestreador de Gibbs como un caso especial del algoritmo Metropolis-Hastings

Suponga que se quiere obtener una muestra aleatoria de una distribución $\pi(\theta)$ donde $\theta = (\theta_1, \theta_2, \dots, \theta_n)$ es el vector de parámetros. Para obtener la muestra se utilizará el algoritmo de Metropolis-Hastings. Suponga que es sencillo obtener muestras aleatorias de las distribuciones condicionales completas del muestreador de Gibbs

$$\pi_i = \pi(\theta_i | \theta_1, \theta_2, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_n),$$

entonces la mejor opción para la matriz de transición Q es tomar $q_{i,j} = \pi_j$. De esta manera la probabilidad de aceptación será

$$\min \left\{ 1, \frac{\pi_{\alpha'} q_{\alpha' \alpha}}{\pi_{\alpha} q_{\alpha \alpha'}} \right\} = \min \left\{ 1, \frac{\pi_{\alpha'} \pi_{\alpha}}{\pi_{\alpha} \pi_{\alpha'}} \right\} = 1.$$

Por lo tanto, el algoritmo estará generando valores a partir de la distribución condicional y aceptando con probabilidad 1 todas las propuestas de Q . En otras palabras, el algoritmo generado es el muestreador de Gibbs.

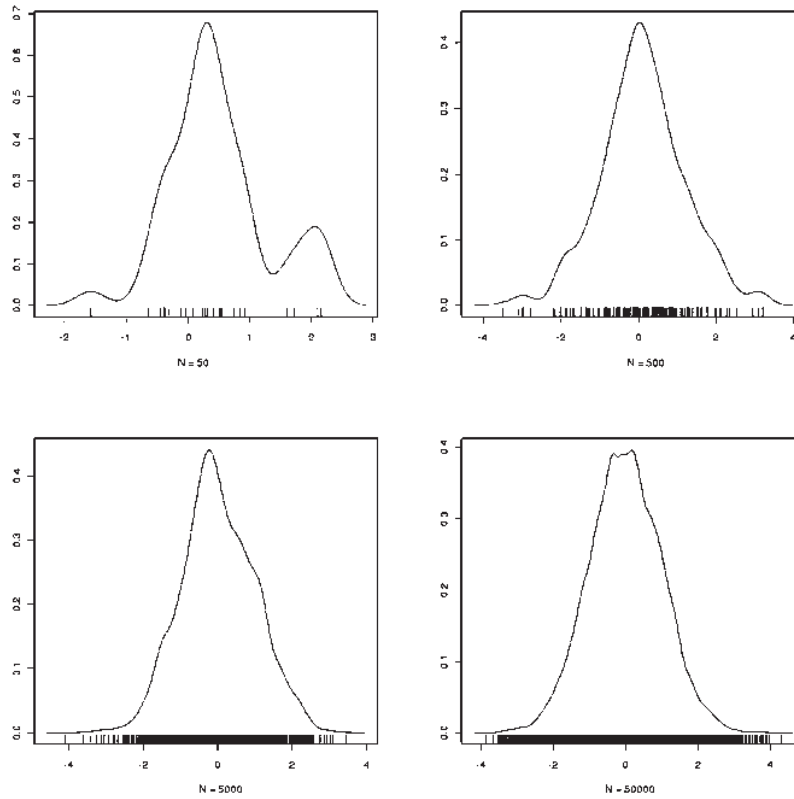


Figura 7.2: Generación de una muestra aleatoria de la distribución normal $N(0,1)$ a partir de una distribución uniforme por medio del algoritmo Metropolis-Hastings. (a) 100 iteraciones con los primeros 50 estados descartados, (b) 1000 iteraciones con los primeros 500 estados descartados, (c) 10000 iteraciones con los primeros 5000 estados descartados y (d) 100000 iteraciones con los primeros 50000 estados descartados.

7.4. Convergencia

Tanto el muestreador de Gibbs como el algoritmo de Metropolis-Hastings dependen crucialmente de un diagnóstico adecuado de convergencia para poder extraer valores de la distribución buscada. Si el tiempo de calentamiento es demasiado pequeño, la muestra producida no será una muestra correcta de la distribución. Por otro lado, si el tiempo de calentamiento tomado es demasiado grande, tomar la muestra puede resultar ser demasiado costoso computacionalmente.

7.4.1. Diagnóstico informal

Una manera de obtener una noción acerca de la convergencia de los algoritmos consiste en correr varias veces el algoritmo de forma paralela comenzando en estados distintos; para cada cadena se puede ir calculando el promedio ergódico de m iteraciones dado por

$$f_m = \sum_{i=0}^m \frac{x_i}{m},$$

donde x_i , $i = 1, 2, \dots, m$ es la muestra generada por el algoritmo MCMC. El acercamiento de los promedios de las distintas cadenas a un mismo punto sugiere que la convergencia ha sido alcanzada.

Observación: Al correr las K cadenas y calcular los promedios ergódicos, si los promedios son muy distintos, entonces se puede asegurar que la convergencia no fue alcanzada por al menos una de las cadenas. Sin embargo, si todos los promedios ergódicos convergen a un mismo número, esto no garantiza con seguridad la convergencia de las cadenas a la distribución estacionaria.

7.4.2. Prueba de Gelman-Rubin

Un método más riguroso para estimar la convergencia fué desarrollado por Gelman y Rubin (1992). Para utilizar la prueba de convergencia de Gelman-Rubin es necesario ejecutar al menos dos cadenas con estados iniciales distintos. La idea esencial detrás de la prueba es ver si la dispersión dentro de cada cadena es mayor a la dispersión entre cada cadena. Equivalentemente, se busca que el histograma de cada cadena sea similar al histograma de las cadenas combinadas.

Definición 7.4.1 Suponga que se tienen m cadenas paralelas de longitud n y sea $\psi_i^{(j)}$ el valor j -ésimo de la cadena i . La variancia entre cadenas B y

la variancia dentro de las cadenas W están dadas por

$$B = \frac{n}{m-1} \sum_{i=1}^m (\bar{\psi}_i - \bar{\psi})^2, \quad W = \frac{1}{m(n-1)} \sum_{i=1}^m \sum_{j=1}^n (\psi_i^{(j)} - \bar{\psi}_i)^2,$$

donde $\bar{\psi}_i$ es el promedio de las observaciones de la cadena i , $i = 1, \dots, m$ y $\bar{\psi}$ es el promedio de estos promedios. Es decir,

$$\bar{\psi}_i = \frac{1}{n} \sum_{j=1}^n \psi_i^{(j)} \quad \text{y} \quad \bar{\psi} = \frac{1}{m} \sum_{i=1}^m \bar{\psi}_i.$$

Observación: A partir de las estadísticas definidas, se pueden obtener tres estimaciones de la variancia de ψ . Específicamente B , W y el promedio

$$\hat{\sigma}_\psi^2 = \frac{n-1}{n}W + \frac{1}{n}B.$$

Más aún, por la definición de W se sigue claramente que

$$\hat{\sigma}_\psi^2 \geq W.$$

Si la convergencia de todas las cadenas ha sido alcanzada, entonces los mn valores obtenidos forman una muestra aleatoria de la distribución estacionaria y por lo tanto la variancia σ_ψ^2 de la cadena ψ puede ser estimada de manera correcta por medio de W , B y $\hat{\sigma}_\psi^2$.

Definición 7.4.2 Suponga que se tienen m cadenas paralelas de longitud n y sea $\psi_i^{(j)}$ el valor j -ésimo de la cadena i . Sea además B la variancia entre cadenas, W la variancia dentro de las cadenas y $\hat{\sigma}_\psi^2$ el promedio. El indicador de convergencia de Gelman-Rubin \hat{R} está dado por

$$\hat{R} = \sqrt{\hat{\sigma}_\psi^2 / W},$$

Observación: El indicador de convergencia \hat{R} siempre es mayor o igual a 1. Además note que

$$\lim_{n \rightarrow \infty} \hat{R} = 1,$$

pues ambos estimados convergerán a σ_ψ^2 por el Teorema fundamental de convergencia y así, la convergencia de las cadenas puede ser evaluada por medio de la proximidad de \hat{R} a 1.

La distancia mínima a 1 que debe tomar \hat{R} no está fija, sin embargo, Gelman [11] sugiere aceptar la convergencia cuando \hat{R} esté debajo de 1.2.

Capítulo 8

Aplicación a los Niveles de Ozono del Distrito Federal

En este capítulo se utilizan los resultados derivados en los capítulos anteriores para poder dar una formulación matemática adecuada al problema de estimar la cantidad de veces que un estándar ambiental es violado en un intervalo de tiempo.

El ozono en la tropósfera se forma cuando monóxido de carbono, óxidos nitrosos y compuestos orgánicos volátiles (como los isoprenos y terpenos) reaccionan en la atmósfera con la luz del sol. Estos compuestos tienen una gran cantidad de orígenes desde emisiones vehiculares e industriales y solventes químicos hasta emisiones de plantas en el caso de los isoprenos. Más aún, el ozono producido comúnmente es desplazado por corrientes de viento por cientos de kilómetros de su lugar de origen.

Las reacciones químicas involucradas en la formación de ozono son una serie de ciclos complejos los cuales van más allá del alcance de esta tesis, sin embargo, es importante notar que la formación de ozono depende de una gran cantidad de factores altamente volátiles y difíciles de comprender por lo que un acercamiento determinístico para entender el comportamiento del ozono sería muy poco fructífero.

Una discusión más profunda acerca del acercamiento Bayesiano para modelar los procesos de Poisson puede ser encontrada en [6].

El acercamiento para resolver la formulación es por medio de métodos Monte Carlo vía cadenas de Markov, utilizando un algoritmo construido a partir del muestreador de Gibbs y del algoritmo Metropolis-Hastings.

8.1. Formulación del problema

El problema en cuestión se describe de la siguiente manera: Sea $N_t \geq 0, t \geq 0$ la cantidad de veces que un estándar ambiental de un contaminante

dado es violado en el intervalo de tiempo $[0, t)$. Para describir el comportamiento observado se utilizará un proceso de Poisson no-homogéneo.

Como se vió en el capítulo 5, un proceso de Poisson no-homogéneo requiere de una función de intensidad $\lambda(t) > 0, t \geq 0$. Una vez que se conoce la función $\lambda(t)$, entonces la probabilidad de que el estándar ambiental sea violado en el intervalo de tiempo $[t_1, t_2)$ podrá ser calculado con la ecuación 5.4, es decir

$$\mathbf{P}(N(t_2) - N(t_1) = n) = e^{-[m(t_2) - m(t_1)]} \frac{[m(t_2) - m(t_1)]^n}{n!},$$

donde $m(t)$ es la función la media dada por

$$m(t) = \int_0^t \lambda(s) ds.$$

Observación: Dada una función de intensidad $\lambda(t) > 0, t \geq 0$, la cual depende de un vector de parámetros $\theta = (\theta_1, \theta_2, \dots, \theta_J)$. El problema descrito se reduce a estimar el vector de parámetros θ que mejor describa el comportamiento de los datos observados. Es decir, el vector θ que produzca al proceso de Poisson no-homogéneo que mejor represente el conjunto de observaciones $N_t, t \in \mathbb{R}^+$.

8.2. Acercamiento Bayesiano

Sea $D = \{d_1, d_2, \dots, d_K\}$ el conjunto de días en los cuales un estándar ambiental de un contaminante ha sido violado. Sea $\lambda(t), t \geq 0$ la función de intensidad de la cual depende el proceso de Poisson no-homogéneo el cual se utilizará para describir la muestra observada D . Suponga además que la función $\lambda(t)$ depende de un vector de parámetros $\theta = (\theta_1, \theta_2, \dots, \theta_J) \in \mathbb{R}^J$. Entonces el problema se reduce a encontrar el vector $\theta \in \mathbb{R}^J$ que mejor explique los datos D .

Utilizando el punto de vista Bayesiano, existe una relación entre la distribución a priori del vector θ con la distribución a posteriori al tomar en cuenta la información observada D . Es decir,

$$p(\theta|D) \propto L(D|\theta)p(\theta).$$

Teorema 8.2.1 *Sea $D = \{d_1, d_2, \dots, d_K\}$ una muestra aleatoria obtenida de un proceso de Poisson no-homogéneo con función de intensidad $\lambda(t)$, $t \in \mathbb{N}$ dependiente de un vector de parámetros $\theta = (\theta_1, \theta_2, \dots, \theta_J) \in \mathbb{R}^J$, entonces, bajo el modelo de Poisson no-homogéneo, la función de verosimilitud $L(D|\theta)$ es de la forma*

$$L(D|\theta) = \left[\prod_{i=1}^n \lambda(d_i) \right] e^{-m(d_n)}.$$

Demostración: Sea $\epsilon > 0$. Por tratarse de un proceso de Poisson no-homogéneo, es posible descomponer el tiempo en intervalos independientes y analizar cada uno por separado. Considere la descomposición en los intervalos

$$(d_i, d_{i+1} - \epsilon] \cup (d_{i+1} - \epsilon, d_{i+1}], \quad i = 1, 2, \dots, n,$$

es decir,

$$(0, d_n] = (0, d_1 - \epsilon] \cup (d_1 - \epsilon, d_1] \cup (d_1, d_2 - \epsilon] \cup \dots \cup (d_n - \epsilon, d_n].$$

La probabilidad de que eventos del proceso de Poisson no-homogéneo hayan sucedido en los intervalos de tiempo $(d_i - \epsilon, d_i]$ y que ningún evento haya sucedido en los intervalos de la forma $(d_i, d_{i+1} - \epsilon]$ es el producto de las probabilidades (gracias a la independencia de eventos del proceso). Formalmente, la probabilidad de observar eventos únicamente en los intervalos de distancia ϵ por la izquierda al día de la observación es el producto de las observaciones de cada intervalo dadas por la ecuación (5.4)

$$\begin{aligned} P(N(d_1 - \epsilon) - N(0) = 0) &= e^{-m(d_1 - \epsilon)} \\ P(N(d_1) - N(d_1 - \epsilon) = 1) &= e^{-m(d_1) + m(d_1 - \epsilon)} [m(d_1) - m(d_1 - \epsilon)] \\ P(N(d_2 - \epsilon) - N(d_1) = 0) &= e^{-m(d_2 - \epsilon) + m(d_1)} \\ P(N(d_2) - N(d_2 - \epsilon) = 1) &= e^{-m(d_2) + m(d_2 - \epsilon)} [m(d_2) - m(d_2 - \epsilon)] \\ &\vdots \\ P(N(d_n - \epsilon) - N(d_{n-1}) = 0) &= e^{-m(d_n - \epsilon) + m(d_{n-1})} \\ P(N(d_n) - N(d_n - \epsilon) = 1) &= e^{-m(d_n) + m(d_n - \epsilon)} [m(d_n) - m(d_n - \epsilon)]. \end{aligned}$$

Así, la probabilidad conjunta, siendo el producto de las probabilidades será

$$e^{-m(d_n)} \prod_{i=1}^n [m(d_i) - m(d_i - \epsilon)]. \quad (8.1)$$

Tomando ahora $\lim_{\epsilon \rightarrow 0}$, la ecuación 8.1 describe justamente la probabilidad de que se observen eventos únicamente en el conjunto de tiempos $\{d_1, d_2, \dots, d_n\}$. Mas aún, recordando que

$$m(t) = \int_0^t \lambda(s) ds,$$

se sigue que conforme ϵ tiende a 0, $[m(d_i) - m(d_i - \epsilon)] = \lambda(d_i)$ resultando en

$$\left[\prod_{i=1}^n \lambda(d_i) \right] e^{-m(d_n)}.$$

■

Corolario 8.2.2 Sea $D = \{d_1, d_2, \dots, d_K\}$ el conjunto de días en los cuales un estándar ambiental de un contaminante dado ha sido violado. Sea $\lambda(t)$ la función de intensidad del proceso de Poisson no-homogéneo que registra los días en los cuales el estándar fue violado dependiente del vector de parámetros $\theta = (\theta_1, \theta_2, \dots, \theta_K) \in \mathbb{R}^K$, entonces

$$p(\theta|D) \propto \left[\prod_{i=1}^K \lambda(d_i) \right] e^{-m(d_K)} p(\theta). \quad (8.2)$$

Observación: Dada la función de intensidad $\lambda(t)$ de un proceso de Poisson no-homogéneo que produjo una muestra aleatoria $D = \{d_1, d_2, \dots, d_n\}$, la ecuación (8.2) proporciona una distribución de los posibles vectores de parámetros que determinan un proceso de Poisson no-homogéneo que produce a D . Por lo tanto, el vector de parámetros θ que hace que el ajuste a los datos observados sea el mejor es el que se considera como parámetro del modelo.

8.3. Solución computacional

Debido a que las funciones de densidad a posteriori de los parámetros de la función de intensidad $\lambda(t)$ pueden resultar en distribuciones en las cuales es muy complicado calcular la media y otras características de la distribución $f(\theta|D)$, se utiliza un método Monte Carlo vía cadenas de Markov para obtener muestras de la distribución a posteriori y así aproximar la media. En otras palabras, dada una función de intensidad $\lambda(t)$ dependiente del vector de parámetros θ , se construirá una cadena de Markov tal que su distribución estacionaria sea $p(\theta|D)$ y así, una vez que la cadena de Markov alcance convergencia, los estados en los que se sitúe la cadena de Markov pueden ser interpretados como muestras de la distribución a posteriori.

Sea $D = \{d_1, d_2, \dots, d_K\}$ una muestra aleatoria producida por un proceso de Poisson no-homogéneo con función de intensidad $\lambda(t)$. Si $\lambda(t)$ depende de un vector de parámetros $\theta = (\theta_1, \theta_2, \dots, \theta_J) \in \mathbb{R}^J$ entonces se utiliza un muestreador de Gibbs para ir tomando muestras de cada parámetro θ_i . Es decir, iterativamente, se irán tomando muestras de la distribución

$$p(\theta_i|\theta_j, j \neq i, D). \quad (8.3)$$

Para tomar muestras de la distribución condicional de la ecuación (8.3) se utiliza un algoritmo de Metropolis-Hastings que converge a la distribución marginal.

En otras palabras, si el vector de parámetros θ es un elemento del espacio \mathbb{R}^J , entonces se construyen J algoritmos de Metropolis-Hastings tal que la distribución estacionaria del i -ésimo algoritmo de Metropolis-Hastings

es la ecuación (8.3). Una vez que estas J cadenas alcanzan sus distribuciones estacionarias, entonces el muestreador de Gibbs que extrae muestras aleatorias de las distribuciones marginales puede obtener las muestras de los parámetros, y así estimar la media.

A continuación se describe el algoritmo de manera más formal:

-
- 1) Generar un vector aleatorio $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_J^{(0)})$ donde el elemento $\theta_i^{(0)}$ fue obtenido de la distribución a priori $p(\theta_i)$.
 - 2) Generar un vector aleatorio $\theta' = (\theta'_1, \theta'_2, \dots, \theta'_J)$ de la misma manera que en el paso anterior.
 - 3) Dado el vector $\theta^{(j-1)}$ en la $(j-1)$ -ésima iteración del algoritmo, generar el vector $\theta^{(j)}$, donde la entrada $\theta_i^{(j)}$ está dada por

$$\theta_i^{(j)} = \left\{ \begin{array}{ll} \theta'_i & \text{con probabilidad } a \\ \theta_i^{(j-1)} & \text{con probabilidad } 1 - a \end{array} \right\}$$

donde

$$a = \min \left\{ 1, \frac{p(\theta'_i | \theta_1^{(j)}, \dots, \theta_{i-1}^{(j)}, \theta_{i+1}^{(j-1)}, \dots, \theta_K^{(j)}) p(\theta_i^{(j-1)})}{p(\theta_i^{(j-1)} | \theta_1^{(j)}, \dots, \theta_{i-1}^{(j)}, \theta_{i+1}^{(j-1)}, \dots, \theta_K^{(j)}) p(\theta'_i)} \right\}.$$

- 4) Volver al paso 2.
-

8.4. Modelos Utilizados

Debido a que la frecuencia con la cual los niveles de ozono superan el umbral ambiental establecido por la legislación de la ciudad de México ha ido disminuyendo desde el año 2001, es claro que los modelos propuestos deben aceptar parámetros que den una forma decreciente. Los modelos utilizados fueron elegidos debido a que la forma de la función de intensidad permita la posibilidad de admitir una forma decreciente.

8.4.1. Weibull

La función de intensidad de la forma de Weibull está dada por

$$\lambda(t) = \left(\frac{\alpha}{\sigma}\right) \left(\frac{t}{\sigma}\right)^{\alpha-1}, \quad (8.4)$$

por lo tanto, el vector de parámetros que se quiere estimar es $\theta = (\alpha, \sigma) \in \mathbb{R}^+ \times \mathbb{R}^+$.

Lema 8.4.1 *La media de un proceso de Poisson no-homogéneo con función de riesgo que toma la forma de Weibull es*

$$m(t) = \left(\frac{t}{\sigma}\right)^\alpha.$$

Demostración: Por definición

$$m(t) = \int_0^t \lambda(s) ds$$

y así

$$\begin{aligned} m(t) &= \int_0^t \left(\frac{\alpha}{\sigma}\right) \left(\frac{s}{\sigma}\right)^{(\alpha-1)} ds \\ &= \frac{\alpha}{\sigma^\alpha} \int_0^t s^{\alpha-1} ds \\ &= \frac{\alpha s^\alpha}{\sigma^\alpha \alpha} \Big|_0^t \\ &= \left(\frac{t}{\sigma}\right)^\alpha. \end{aligned}$$

■

Teorema 8.4.2 *Considere la función de tasa Weibull, si asumimos que los parámetros son a priori independientes, entonces la distribución a posteriori de los parámetros está dada por*

$$p(\alpha, \sigma | D) = \left(\frac{\alpha}{\sigma^\alpha}\right)^K \left[\prod_{i=1}^K d_i^{\alpha-1} \right] e^{-(d_K/\sigma)^\alpha} p(\alpha) p(\sigma).$$

Demostración: Sustituyendo la forma de la función de tasa y la media en (8.2) se tiene que

$$\begin{aligned} p(\alpha, \sigma | D) &\propto L(D | \alpha, \sigma) p(\alpha, \sigma) \\ &= L(D | \alpha, \sigma) p(\alpha) p(\sigma) \\ &= \left[\prod_{i=1}^K \lambda(d_i) \right] e^{-m(d_K)} p(\alpha) p(\sigma) \\ &= \left[\prod_{i=1}^K \left(\frac{\alpha}{\sigma}\right) \left(\frac{d_i}{\sigma}\right)^{\alpha-1} \right] e^{-(d_K/\sigma)^\alpha} p(\alpha) p(\sigma) \\ &= \left(\frac{\alpha}{\sigma^\alpha}\right)^K \left[\prod_{i=1}^K d_i^{\alpha-1} \right] e^{-(d_K/\sigma)^\alpha} p(\alpha) p(\sigma). \end{aligned}$$

■

Corolario 8.4.3 *La probabilidad condicional del parámetro α está dada por*

$$p(\alpha|\sigma, D) \propto \left(\frac{\alpha}{\sigma}\right)^K \left[\prod_{i=1}^K d_i^{\alpha-1} \right] e^{-(d_K/\sigma)^\alpha} p(\alpha). \quad (8.5)$$

Demostración: *Se sigue directamente del hecho de que si σ está fijo entonces $p(\sigma)$ es constante y utilizando la distribución posterior dada por el Teorema 8.4.2* ■

Corolario 8.4.4 *La probabilidad condicional del parámetro σ está dada por*

$$p(\sigma|\alpha, D) \propto \frac{e^{-(d_K/\sigma)^\alpha} p(\sigma)}{\sigma^{\alpha K}}. \quad (8.6)$$

Demostración: Es suficiente notar que si α está fijo entonces

$$\alpha^K \left[\prod_{i=1}^K d_i^{\alpha-1} \right] p(\alpha)$$

es constante. ■

Como se describió anteriormente, la estimación de los parámetros se hará a partir de las distribuciones condicionales (8.5) y (8.6) por medio de algoritmos Metropolis-Hastings. Para las cadenas Q se tomarán cadenas independientes iguales a distribución a priori.

Teorema 8.4.5 *Sea $D = \{d_1, d_2, \dots, d_K\}$ los días en que un estándar ambiental ha sido violado. Suponga que D obedece un proceso de Poisson no-homogéneo con función de intensidad de Weibull. Si se quieren obtener muestras de las distribuciones condicionales de la distribución a posteriori de los parámetros α y σ por medio de algoritmos Metropolis-Hastings donde la distribución de Y es la distribución a priori del parámetro, entonces las probabilidades de aceptación serán*

a) *en el caso del parámetro α*

$$\min \left\{ 1, \left(\frac{\alpha'}{\alpha}\right)^K \sigma^{(\alpha-\alpha')K} \exp \left[(\alpha' - \alpha) \sum_{i=1}^K \log(d_i) + (d_K/\sigma)^\alpha - (d_K/\sigma)^{\alpha'} \right] \right\}$$

b) *en el caso del parámetro σ*

$$\min \left\{ 1, \exp \left(\alpha^K \log \left(\frac{\sigma}{\sigma'} \right) + (d_K/\sigma)^\alpha - (d_K/\sigma')^\alpha \right) \right\}$$

Demostración:

- a) Dado el estado actual α de la cadena construida por el algoritmo y un estado propuesto α' por Y , entonces la probabilidad de aceptación de α' está dada por

$$\text{mín} \{1, a\}$$

donde

$$\begin{aligned} a &= \frac{p(\alpha'|\sigma, D)q_{\alpha'\alpha}}{p(\alpha|\sigma, D)q_{\alpha\alpha'}} \\ &= \frac{p(\alpha'|\sigma, D)p(\alpha)}{p(\alpha|\sigma, D)p(\alpha')} \\ &= \frac{\left(\frac{\alpha'}{\sigma\alpha'}\right)^K \left[\prod_{i=1}^K d_i^{\alpha'-1}\right] e^{-(d_K/\sigma)^{\alpha'}} p(\alpha')p(\alpha)}{\left(\frac{\alpha}{\sigma\alpha}\right)^K \left[\prod_{i=1}^K d_i^{\alpha-1}\right] e^{-(d_K/\sigma)^\alpha} p(\alpha)p(\alpha')} \\ &= \frac{\left(\frac{\alpha'}{\sigma\alpha'}\right)^K e^{(\alpha'-1)\sum_{i=1}^K \log(d_i)} e^{-(d_K/\sigma)^{\alpha'}}}{\left(\frac{\alpha}{\sigma\alpha}\right)^K e^{(\alpha-1)\sum_{i=1}^K \log(d_i)} e^{-(d_K/\sigma)^\alpha}} \\ &= \left(\frac{\alpha'}{\alpha}\right)^K \sigma^{(\alpha-\alpha')K} \frac{e^{(\alpha'-1)\sum_{i=1}^K \log(d_i)} e^{-(d_K/\sigma)^{\alpha'}}}{e^{(\alpha-1)\sum_{i=1}^K \log(d_i)} e^{-(d_K/\sigma)^\alpha}} \\ &= \left(\frac{\alpha'}{\alpha}\right)^K \sigma^{(\alpha-\alpha')K} \frac{\exp((\alpha'-1)\sum_{i=1}^K \log(d_i) - (d_K/\sigma)^{\alpha'})}{\exp((\alpha-1)\sum_{i=1}^K \log(d_i) - (d_K/\sigma)^\alpha)}. \end{aligned}$$

Restando los exponentes de e resulta finalmente en

$$\left(\frac{\alpha'}{\alpha}\right)^K \sigma^{(\alpha-\alpha')K} \exp((\alpha'-\alpha)\sum_{i=1}^K \log(d_i) + (d_K/\sigma)^\alpha - (d_K/\sigma)^{\alpha'}).$$

- b) Para la distribución condicional del parámetro σ , si la cadena se encuentra en el valor σ y tiene propuesto el valor σ' , la probabilidad de

aceptación estará dada por

$$\begin{aligned}
a &= \frac{p(\sigma'|\alpha, D)q_{\sigma'\sigma}}{p(\sigma|\alpha, D)q_{\sigma\sigma'}} \\
&= \frac{p(\sigma'|\alpha, D)p(\sigma)}{p(\sigma|\alpha, D)p(\sigma')} \\
&= \frac{e^{-(d_K/\sigma')^\alpha} p(\sigma')p(\sigma)/\sigma'^{\alpha K}}{e^{-(d_K/\sigma)^\alpha} p(\sigma)p(\sigma')/\sigma^{\alpha K}} \\
&= \frac{e^{-(d_K/\sigma')^\alpha} \sigma^{\alpha K}}{e^{-(d_K/\sigma)^\alpha} \sigma'^{\alpha K}} \\
&= \frac{e^{-(d_K/\sigma')^\alpha} e^{\alpha K \log(\sigma)}}{e^{-(d_K/\sigma)^\alpha} e^{\alpha K \log(\sigma')}} \\
&= \frac{e^{\alpha K \log(\sigma) - (d_K/\sigma')^\alpha}}{e^{\alpha K \log(\sigma') - (d_K/\sigma)^\alpha}} \\
&= \exp\left(\alpha K \log\left(\frac{\sigma}{\sigma'}\right) + (d_K/\sigma)^\alpha - (d_K/\sigma')^\alpha\right).
\end{aligned}$$

■

8.4.2. Musa-Okumoto

La función de intensidad de la forma de Musa-Okumoto está dada por

$$\lambda(t) = \frac{\beta}{t + \alpha}. \quad (8.7)$$

Por lo tanto, el vector de parámetros que se quiere estimar es $\theta = (\alpha, \beta) \in \mathbb{R}_+ \times \mathbb{R}_+$.

Lema 8.4.6 *La media de un proceso de Poisson no-homogéneo con función de riesgo que toma la forma de Musa-Okumoto es*

$$m(t) = \beta \left(\log \left(\frac{t}{\alpha} + 1 \right) \right).$$

Demostración: Como

$$m(t) = \int_0^t \lambda(s) ds,$$

entonces

$$\begin{aligned}
 m(t) &= \int_0^t \frac{\beta ds}{s + \alpha} \\
 &= \beta \int_0^t \frac{ds}{s + \alpha} \\
 &= \beta \log(s + \alpha) \Big|_0^t \\
 &= \beta \log(t + \alpha) - \beta \log(\alpha) \\
 &= \beta \left(\log \left(\frac{t + \alpha}{\alpha} \right) \right) \\
 &= \beta \left(\log \left(\frac{t}{\alpha} + 1 \right) \right).
 \end{aligned}$$

■

Teorema 8.4.7 *En la forma de Musa-Okumoto, si la distribución a priori de los parámetros es independiente, entonces la distribución a posteriori de los parámetros está dada por*

$$p(\alpha, \beta | D) \propto \beta^K \left[\prod_{i=1}^K \left(\frac{1}{d_i + \alpha} \right) \right] e^{-\beta \left(\log \left(\frac{d_K}{\alpha} + 1 \right) \right)} p(\alpha) p(\beta).$$

Demostración: Sustituyendo,

$$\begin{aligned}
 p(\alpha, \beta | D) &\propto L(D | \alpha, \beta) p(\alpha, \beta) \\
 &= L(D | \alpha, \beta) p(\alpha) p(\beta) \\
 &= \left[\prod_{i=1}^K \lambda(d_i) \right] e^{-m(d_K)} p(\alpha) p(\beta) \\
 &= \left[\prod_{i=1}^K \left(\frac{\beta}{d_i + \alpha} \right) \right] e^{-\beta \left(\log \left(\frac{d_K}{\alpha} + 1 \right) \right)} p(\alpha) p(\beta) \\
 &= \beta^K \left[\prod_{i=1}^K \left(\frac{1}{d_i + \alpha} \right) \right] e^{-\beta \left(\log \left(\frac{d_K}{\alpha} + 1 \right) \right)} p(\alpha) p(\beta).
 \end{aligned}$$

■

Corolario 8.4.8 *La probabilidad condicional del parámetro α está dada por*

$$p(\alpha | \beta, D) \propto \left[\prod_{i=1}^K \left(\frac{1}{d_i + \alpha} \right) \right] e^{-\beta \left(\log \left(\frac{d_K}{\alpha} + 1 \right) \right)} p(\alpha). \quad (8.8)$$

Corolario 8.4.9 *La probabilidad condicional del parámetro β está dada por*

$$p(\beta|\alpha, D) \propto \beta^K e^{-\beta \left(\log \left(\frac{d_K}{\alpha} + 1 \right) \right)} p(\beta). \quad (8.9)$$

Observación: si la distribución a priori del parámetro β es una distribución uniforme, entonces

$$p(\beta|\alpha, D) \propto \beta^K e^{-\beta \left(\log \left(\frac{d_K}{\alpha} + 1 \right) \right)},$$

la cual es justamente una distribución Gamma $G(n+1, \log(d_n/\alpha+1))$ (ver apéndice B). Similarmente, si la distribución a priori de β es una distribución Gamma $G(a_1, b_1)$, entonces la distribución a posteriori sería nuevamente una distribución Gamma $G(n+a_1+1, b_1+\log(d_n/\alpha+1))$.

Debido a que las distribuciones a priori utilizadas para el parámetro β fueron únicamente estas dos, el muestreo puede ser directo. La generación de una muestra aleatoria con distribución Gamma es una tarea común para la cual se han desarrollado varios métodos los cuales no dependen de un algoritmo de Metropolis-Hastings. Debido a que R, el lenguaje en el cual fueron implementados los algoritmos, ya cuenta con un método efectivo para generar muestras de una distribución Gamma, para este paso no se requerirá construir un algoritmo de Metropolis-Hastings y se asumirá que las muestras aleatorias necesarias están disponibles.

Al igual que en la forma de Weibull, la estimación de los parámetros se hará a partir de las distribuciones condicionales (8.8) y (8.9) por medio de un algoritmo Metropolis-Hastings para el parámetro α y un muestreo directo para el parámetro β . Para la cadena de transición Y para α se tomará una cadena independiente igual a la distribución a priori de la variable.

Teorema 8.4.10 *Sea $D = \{d_1, d_2, \dots, d_K\}$ los días en que un estándar ambiental ha sido violado. Suponga que D obedece un proceso de Poisson no-homogéneo con función de intensidad de Musa-Okumoto. Si se quieren obtener muestras de las distribuciones condicionales de la distribución a posteriori del parámetro α por medio de algoritmos Metropolis-Hastings donde la distribución de Y es la distribución a prior del parámetro, entonces la probabilidad de aceptación de α será*

$$\min \left\{ 1, \exp \left(\left[\sum_{i=1}^K \log \left(\frac{d_i + \alpha}{d_i + \alpha'} \right) \right] + \beta \log \left(\frac{\alpha' (d_K + \alpha)}{\alpha (d_K + \alpha')} \right) \right) \right\}.$$

Demostración: La probabilidad de aceptación es

$$\min \{1, a\}$$

donde

$$\begin{aligned}
a &= \frac{p(\alpha'|\beta, D)q_{\alpha'\alpha}}{p(\alpha|\beta, D)q_{\alpha\alpha'}} \\
&= \frac{\left[\prod_{i=1}^K \frac{1}{d_i + \alpha'} \right] e^{-\beta \log(d_K/\alpha' + 1)} p(\alpha') p(\alpha)}{\left[\prod_{i=1}^K \frac{1}{d_i + \alpha} \right] e^{-\beta \log(d_K/\alpha + 1)} p(\alpha) p(\alpha')} \\
&= \frac{\exp\left[\left[\sum_{i=1}^K -\log(d_i + \alpha') \right] - \beta \log(d_K/\alpha' + 1) \right]}{\exp\left[\left[\sum_{i=1}^K -\log(d_i + \alpha) \right] - \beta \log(d_K/\alpha + 1) \right]} \\
&= \exp\left(\left[\sum_{i=1}^K \log\left(\frac{d_i + \alpha}{d_i + \alpha'} \right) \right] + \beta \left(\log\left(\frac{d_K + \alpha}{\alpha} \right) - \log\left(\frac{d_K + \alpha'}{\alpha'} \right) \right) \right) \\
&= \exp\left(\left[\sum_{i=1}^K \log\left(\frac{d_i + \alpha}{d_i + \alpha'} \right) \right] + \beta \log\left(\frac{\alpha' (d_K + \alpha)}{\alpha (d_K + \alpha')} \right) \right).
\end{aligned}$$

■

8.4.3. Goel-Okumoto Generalizada

La función de intensidad con la forma de Goel-Okumoto generalizada está dada por

$$\lambda(t) = \alpha\beta\gamma t^{\gamma-1} e^{-\beta t^\gamma}. \quad (8.10)$$

Por lo tanto, el vector de parámetros que se quiere estimar es $\theta = (\alpha, \beta, \gamma) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+$.

Lema 8.4.11 *La media de un proceso de Poisson no-homogéneo con función de riesgo que toma la forma de Goel-Okumoto generalizada es*

$$m(t) = \alpha(1 - e^{-\beta t^\gamma}).$$

Para calcular la media anterior solo se debe integrar la ecuación (8.10) respecto al tiempo. Debido a la longitud del desarrollo de esta integral, la demostración se omite.

Teorema 8.4.12 *En la forma de Goel-Okumoto generalizada, si la distribución a priori de los parámetros es independiente, entonces la distribución a posteriori de los parámetros está dada por*

$$(\alpha\beta\gamma)^K \left[\prod_{i=1}^K d_i^{\gamma-1} \right] e^{-\alpha \log(1 - e^{-\beta d_K^\gamma}) - \beta \sum_{i=1}^K \log(d_i^\gamma)} p(\alpha) p(\beta) p(\gamma).$$

Demostración: Sustituyendo,

$$\begin{aligned}
p(\alpha, \beta, \gamma|D) &\propto L(D|\alpha, \beta, \gamma)p(\alpha, \beta, \gamma) \\
&= L(D|\alpha, \beta, \gamma)p(\alpha)p(\beta)p(\gamma) \\
&= \left[\prod_{i=1}^K \lambda(d_i) \right] e^{-m(d_K)} p(\alpha)p(\beta)p(\gamma) \\
&= \left[\prod_{i=1}^K \alpha\beta\gamma d_i^{\gamma-1} e^{-\beta d_i^\gamma} \right] e^{-\alpha \log(1-e^{-\beta d_K^\gamma})} p(\alpha)p(\beta)p(\gamma) \\
&= (\alpha\beta\gamma)^K \left[\prod_{i=1}^K d_i^{\gamma-1} e^{-\beta d_i^\gamma} \right] e^{-\alpha \log(1-e^{-\beta d_K^\gamma})} p(\alpha)p(\beta)p(\gamma).
\end{aligned}$$

Utilizando el hecho de que $x^y = e^{y \log(x)}$ se obtiene

$$(\alpha\beta\gamma)^K \left[\prod_{i=1}^K d_i^{\gamma-1} \right] e^{-\beta \sum_{i=1}^K \log(d_i^\gamma)} e^{-\alpha \log(1-e^{-\beta d_K^\gamma})} p(\alpha)p(\beta)p(\gamma),$$

o de forma equivalente

$$(\alpha\beta\gamma)^K \left[\prod_{i=1}^K d_i^{\gamma-1} \right] e^{-\alpha \log(1-e^{-\beta d_K^\gamma}) - \beta \sum_{i=1}^K \log(d_i^\gamma)} p(\alpha)p(\beta)p(\gamma).$$

■

Corolario 8.4.13 *La distribución posterior condicional del parámetro α está dada por*

$$p(\alpha|\beta, \gamma, D) \propto \alpha^K e^{-\alpha \log(1-\exp(-\beta d_K^\gamma))} p(\alpha). \quad (8.11)$$

Corolario 8.4.14 *La distribución posterior condicional del parámetro β está dada por*

$$p(\beta|\alpha, \gamma, D) \propto (\beta)^K e^{-\alpha \log(1-e^{-\beta d_K^\gamma}) - \beta \sum_{i=1}^K \log(d_i^\gamma)} p(\beta). \quad (8.12)$$

Corolario 8.4.15 *La distribución condicional del parámetro γ está dada por*

$$p(\gamma|\alpha, \beta, D) \propto (\gamma)^K \left[\prod_{i=1}^K d_i^{\gamma-1} \right] e^{-\alpha \log(1-e^{-\beta d_K^\gamma}) - \beta \sum_{i=1}^K \log(d_i^\gamma)} p(\gamma). \quad (8.13)$$

Como se describió en la sección anterior, el muestreo se hará a partir de las distribuciones condicionales (8.11), (8.12) y (8.13) por medio de un muestreador de Gibbs. Para obtener muestras de las distribuciones condicionales se utilizarán tres algoritmos de Metropolis-Hastings cada uno convergiendo a cada una de las distribuciones posteriores condicionales.

Teorema 8.4.16 Sea $D = \{d_1, d_2, \dots, d_K\}$ los días en que un estándar ambiental ha sido violado. Suponga que D obedece un proceso de Poisson no-homogéneo con función de intensidad de Goel-Okumoto generalizada. Si se quieren obtener muestras de la distribución condicional de la distribución a posteriori del parámetro α por medio de algoritmos Metropolis-Hastings donde la distribución de Y es la distribución a priori del parámetro, entonces las probabilidades de aceptación de α , β y γ serán

a)

$$\min \left\{ 1, \exp \left(\left[\sum_{i=1}^K \log \left(\frac{d_i + \alpha}{d_i + \alpha'} \right) \right] + \beta \log \left(\frac{\alpha' (d_K + \alpha)}{\alpha (d_K + \alpha')} \right) \right) \right\}.$$

b)

$$\min \left\{ 1, \left(\frac{\beta'}{\beta} \right)^K \exp(\alpha(\log(1 - e^{-\beta' d_K^\gamma}) - \log(1 - e^{-\beta d_K^\gamma})) + (\beta - \beta') \sum_{i=1}^K \log(d_i^\gamma)) \right\}$$

c)

$$\min \left\{ 1, \left(\frac{\gamma'}{\gamma} \right)^K \left[\prod_{i=1}^K d_i^{(\gamma' - \gamma)} \right] e^{\alpha \log \left(\frac{1 - \exp(-\beta d_K^\gamma)}{1 - \exp(-\beta d_K^{\gamma'})} \right) + \beta \sum_{i=1}^K \log(d_i^{\gamma' - \gamma})} \right\}$$

respectivamente.

Demostración:

a) Dado el estado actual α y el estado propuesto α' obtenido de la distribución a priori, la probabilidad de aceptación es

$$\min \{1, a\}$$

donde

$$\begin{aligned}
a &= \frac{p(\alpha' | \beta, D) q_{\alpha' \alpha}}{p(\alpha | \beta, D) q_{\alpha \alpha'}} \\
&= \frac{\left[\prod_{i=1}^K \frac{1}{d_i + \alpha'} \right] e^{-\beta \log(d_K / \alpha' + 1)} p(\alpha') p(\alpha)}{\left[\prod_{i=1}^K \frac{1}{d_i + \alpha} \right] e^{-\beta \log(d_K / \alpha + 1)} p(\alpha) p(\alpha')} \\
&= \frac{\exp\left(\left[\sum_{i=1}^K -\log(d_i + \alpha')\right] - \beta \log(d_K / \alpha' + 1)\right)}{\exp\left(\left[\sum_{i=1}^K -\log(d_i + \alpha)\right] - \beta \log(d_K / \alpha + 1)\right)} \\
&= \exp\left(\left[\sum_{i=1}^K \log\left(\frac{d_i + \alpha}{d_i + \alpha'}\right)\right] + \beta \left(\log\left(\frac{d_K + \alpha}{\alpha}\right) - \log\left(\frac{d_K + \alpha'}{\alpha'}\right)\right)\right) \\
&= \exp\left(\left[\sum_{i=1}^K \log\left(\frac{d_i + \alpha}{d_i + \alpha'}\right)\right] + \beta \log\left(\frac{\alpha' (d_K + \alpha)}{\alpha (d_K + \alpha')}\right)\right).
\end{aligned}$$

- b) Dado el estado actual β y el estado propuesto β' obtenido de la distribución a priori, la probabilidad de aceptación es

$$\min\{1, a\}$$

donde

$$\begin{aligned}
a &= \frac{p(\beta' | \alpha, \gamma, D) p(\beta)}{p(\beta | \alpha, \gamma, D) p(\beta')} \\
&= \frac{(\beta')^K e^{-\alpha \log(1 - e^{-\beta' d_K^\gamma}) - \beta' \sum_{i=1}^K \log(d_i^\gamma)} p(\beta') p(\beta)}{(\beta)^K e^{-\alpha \log(1 - e^{-\beta d_K^\gamma}) - \beta \sum_{i=1}^K \log(d_i^\gamma)} p(\beta) p(\beta')} \\
&= \left(\frac{\beta'}{\beta}\right)^K e^{\alpha(\log(1 - e^{-\beta' d_K^\gamma}) - \log(1 - e^{-\beta d_K^\gamma})) + (\beta - \beta') \sum_{i=1}^K \log(d_i^\gamma)}.
\end{aligned}$$

- c) Dado el estado actual γ y el estado propuesto γ' , la probabilidad de que la cadena acepte el valor propuesto está dada por

$$\min\{1, a\}$$

donde

$$\begin{aligned}
a &= \frac{p(\gamma'|\alpha, \beta, D)p(\gamma)}{p(\gamma|\alpha, \beta, D)p(\gamma')} \\
&= \frac{(\gamma')^K \left[\prod_{i=1}^K d_i^{\gamma'-1} \right] e^{-\alpha \log(1-e^{-\beta d_K^{\gamma'}}) - \beta \sum_{i=1}^K \log(d_i^{\gamma'})} p(\gamma') p(\gamma)}{(\gamma)^K \left[\prod_{i=1}^K d_i^{\gamma-1} \right] e^{-\alpha \log(1-e^{-\beta d_K^{\gamma}}) - \beta \sum_{i=1}^K \log(d_i^{\gamma})} p(\gamma) p(\gamma')} \\
&= \left(\frac{\gamma'}{\gamma} \right)^K \left[\prod_{i=1}^K d_i^{\gamma'-\gamma} \right] \frac{e^{-\alpha \log(1-e^{-\beta d_K^{\gamma'}}) - \beta \sum_{i=1}^K \log(d_i^{\gamma'})}}{e^{-\alpha \log(1-e^{-\beta d_K^{\gamma}}) - \beta \sum_{i=1}^K \log(d_i^{\gamma})}} \\
&= \left(\frac{\gamma'}{\gamma} \right)^K \left[\prod_{i=1}^K d_i^{\gamma'-\gamma} \right] e^{\alpha \log \left(\frac{1-\exp(-\beta d_K^{\gamma'})}{1-\exp(-\beta d_K^{\gamma})} \right) + \beta \sum_{i=1}^K \log(d_i^{\gamma'-\gamma})}.
\end{aligned}$$

■

8.5. Comparación de modelos

Una vez que se han estimado los parámetros de cada función de intensidad, resta la tarea de decidir cual de las funciones propuestas es la mejor para explicar el conjunto de observaciones D .

El acercamiento para resolver este problema es el descrito en la sección 3.5. Formalmente, dados dos modelos (en este caso un modelo es una función de intensidad con determinados parámetros) M_1 y M_2 que pudieron dar origen a la muestra observada D , se calculará el factor de Bayes

$$BF_{M_1, M_2} = \frac{P(D|M_1)}{P(D|M_2)}$$

para determinar que modelos es más probable de haber producido D . Como en este caso hay tres modelos propuestos (Weibull (W), Musa-Okumoto (MO) y Goel-Okumoto generalizada (GGO)) se calcularon los factores $BF_{W,MO}$ y $BF_{MO,GGO}$ inicialmente. Así se tienen las siguientes opciones:

- Si $BF_{W,MO} > 1$ y $BF_{MO,GGO} > 1$ entonces el Modelo que mejor representa las observaciones es el de Weibull.
- Si $BF_{W,MO} < 1$ y $BF_{MO,GGO} > 1$ entonces el Modelo que mejor representa las observaciones es el de Musa-Okumoto.
- Si $BF_{W,MO} < 1$ y $BF_{MO,GGO} < 1$ entonces el Modelo que mejor representa las observaciones es el de Goel-Okumoto generalizado.

- d) Si $BF_{W,MO} > 1$ y $BF_{MO,GGO} < 1$ entonces se calcula el factor adicional $BF_{GGO,W}$. Si $BF_{GGO,W} > 1$ entonces el Modelo que mejor representa las observaciones es el de Goel-Okumoto generalizado y si $BF_{GGO,W} < 1$ entonces el modelo preferencial es el de Weibull.

Observación: La probabilidad de que un modelo M_k haya producido las observaciones D puede ser reescrito como

$$P(D|M_k) = \int P(\theta_k, M_k)P(\theta_k|M_k)d\theta_k, \quad (8.14)$$

donde θ_k es el vector de parámetros del cual depende el modelo M_k .

Como en este caso el Modelo consta únicamente de la función de intensidad y su vector de parámetros, la ecuación (8.14) se puede escribir como

$$P(D|M_k) = \int P(D|\theta_k)P(\theta_k)d\theta_k. \quad (8.15)$$

Si se tiene una muestra aleatoria de tamaño J de la distribución a posteriori $f(\theta|D)$ de los parámetros, entonces la ecuación (8.15) puede ser aproximada por medio de

$$\hat{V}_{M_k} = \frac{1}{J} \sum_{i=1}^J L(D|\theta_i^{(k)})$$

donde $\theta^{(k)}$ es la muestra aleatoria que fué utilizada para estimar el vector de parámetros θ del modelo k .

Observación: La ley de los grandes números (Teorema 7.1.3) asegura la convergencia de \hat{V}_{M_k} a $P(D|M_k)$. Por lo tanto, la muestra aleatoria obtenida con los algoritmos descritos en este capítulo puede ser utilizada adicionalmente para estimar el factor de Bayes y así poder comparar los distintos modelos propuestos.

8.6. Obtención de las observaciones

Para poder aplicar la solución propuesta en este capítulo, solo falta obtener un conjunto de observaciones de la violación de un estándar ambiental.

Las observaciones de los días en los cuales el Ozono superó el límite de 0.17ppm en la Ciudad de México fue obtenido a partir de la Red Automática de Monitoreo Atmosférico (RAMA). La RAMA tiene 36 estaciones de monitoreo en la Ciudad de México. Estas estaciones reportan cada 5 minutos el promedio de las mediciones de Ozono obtenidas en dicho intervalo.

Para el modelado de los niveles de Ozono utilizando los procesos de Poisson no-homogéneos descritos anteriormente, se dividieron las estaciones en cinco zonas: Noreste (NE), Noroeste (NO), Centro (CE), Sureste (SE) y Suroeste (SO). Los datos utilizados conforman las mediciones tomadas desde el 1 de Enero del 2003 hasta el 31 de Diciembre del 2009. Para el conjunto de estaciones de cada una de las zonas se utilizó el máximo de las mediciones de las estaciones del conjunto. Si por un tiempo mayor a una hora (12 mediciones consecutivas) los niveles de Ozono permanecen por encima de 0.17 ppm entonces el día de la observación es agregado al conjunto D .

Para más detalles de las estaciones y una visualización de la división en las cinco zonas mencionadas se puede consultar en:

<http://www.sma.df.gob.mx/simat/pnrama2.htm>.

8.7. Detalles de implementación

La implementación computacional de los algoritmos, disponibles en el Apéndice C, fue hecha en el lenguaje R (<http://www.r-project.org>). La elección de este lenguaje sobre otros es debido a su orientación especial a la probabilidad y estadística. Las librerías disponibles del lenguaje R contienen grandes tablas de muestras pseudoaleatorias de distribuciones comunes, lo cual facilitó la obtención de las muestras de las distribuciones a priori utilizadas.

La elección de las distribuciones a priori fue basada en la intuición del comportamiento de las distribuciones a posteriori para intentar acelerar la convergencia de los algoritmos a sus distribuciones estacionarias. En algunos casos (como el descrito en la subsección 8.7.1), un primer estimado de los parámetros fue hecha a partir de distribuciones no-informativas a priori. Con la información obtenida, se corrió un segundo algoritmo modificando la distribución a priori con la intención de obtener un estimado más preciso.

Los resultados completos de todas las zonas y los resultados de cada función de intensidad están disponibles en el Apéndice D.

8.7.1. Ejemplo de implementación

Como ejemplo representativo se describe la obtención de los parámetros para la forma de Weibull en la zona Noroeste.

En la figura 8.1 se muestra la gráfica de la cantidad de días transcurridos desde el 1 de enero del 2001 hasta el 31 de diciembre del 2009 contra la cantidad de días en que se ha superado el límite de 0.17ppm de Ozono en zona noroeste de la Ciudad de México.

Debido a que los niveles de ozono han ido disminuyendo consistentemente, las distribuciones a priori de los parámetros α y σ de la forma de

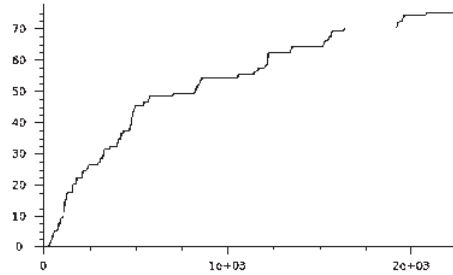


Figura 8.1: Días transcurridos desde el 01-Enero-2001 hasta 31-Diciembre-2009 contra la cantidad de días en las cuales los niveles de ozono se mantuvieron por encima de 0.17ppm por mas de una hora.

Weibull son restringidos de tal forma que la función de intensidad siempre es decreciente.

Inicialmente se corrió el algoritmo tomando para el parámetro α una distribución a priori no-informativa (uniforme) en el intervalo $(0, 1)$ y para el parámetro σ una distribución a priori no-informativa en el intervalo $(0, 100)$. Se corrió el algoritmo tres veces con los estados iniciales 0.8, 0.6 y 0.9 para α y 10, 5 y 20 para σ . Cada una de las tres cadenas fue corrida por 100,000 iteraciones (produciendo así una muestra de 300,000 elementos para α y σ).

Como resultado de la prueba de convergencia de Gelman-Rubin se descartaron las primeras 30,000 iteraciones de cada cadena (iteraciones en las cuales el factor de estimación \hat{R} se encontró por encima de 1.05). De los 70,000 elementos restantes, se utilizó únicamente un valor de cada 20 para así reducir la dependencia de los elementos de la muestra. Con esto, la muestra aleatoria final de cada cadena fue de 3,500 elementos, dando 10,500 como total de las tres cadenas.

El estimado de la media y la variancia utilizando dicha muestra se obtuvieron los siguientes resultados:

	Media empírica	Variancia empírica
α	0.6534	0.004074
σ	3.47	5.631053

Las figuras 8.2a y 8.2b muestran las medias estimadas durante las 100,000 iteraciones de las tres cadenas. Note que la escala en 8.2a es mucho más pequeña por lo cual las medias estimadas están aparentemente distanciadas. Una muestra de las pruebas de convergencia se puede ver en la figura 8.3.

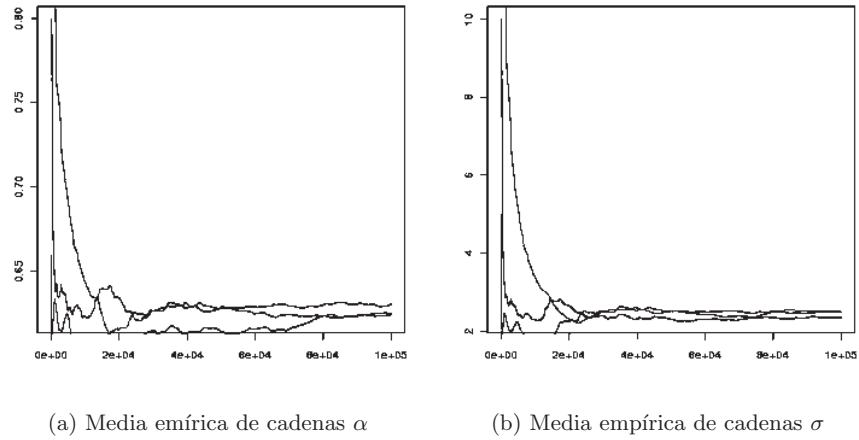


Figura 8.2: Cantidad de iteraciones de las tres cadenas contra la media empírica para los parámetros α y σ .

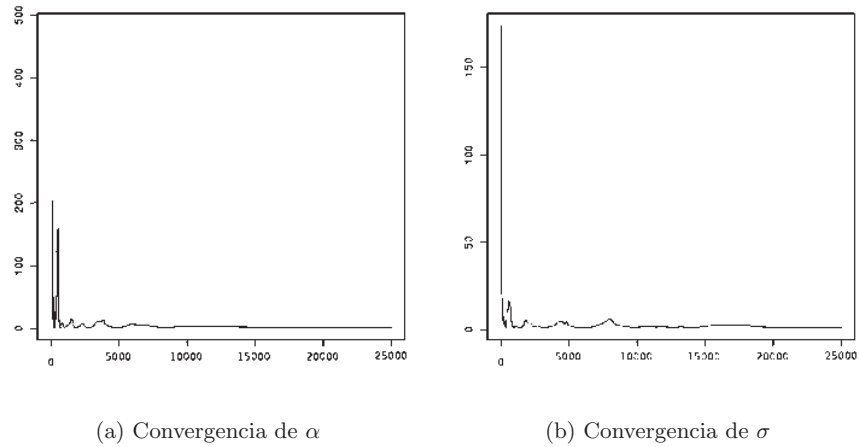


Figura 8.3: Prueba de Gelman-Rubin para las cadenas 1 y 2 (Los puntos iniciales están disponibles en el apéndice D) durante las primeras 25000 iteraciones de los algoritmos.

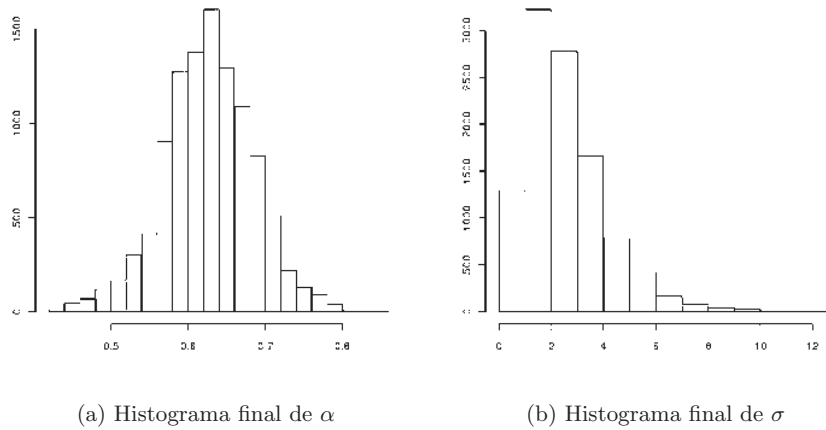


Figura 8.4: Histogramas finales una vez descartados los tiempos de calentamiento e iteraciones intermedias.

Observando la forma de la distribución a posteriori de la estimación, se decidió utilizar una distribución Beta para la distribución a priori del parámetro α y una distribución Gamma para la distribución a priori del parámetro σ (ver Apéndice B para las definiciones y sus hiperparámetros). Los hiperparámetros que definen la forma de las distribuciones Beta ($\alpha = 3/25$, $\beta = 2/25$) y Gamma ($k = 3/2$, $\theta = 1/2$) fueron elegidos de tal forma que la distribución a priori tuviera media y variancia similar a las obtenidas de la muestra aleatoria anterior.

Con estas nuevas distribuciones a priori, se corrieron nuevamente tres cadenas con los mismos puntos iniciales por 100,000 iteraciones. A pesar de que la convergencia según la prueba de Gelman-Rubin se alcanzó más rápidamente que en la ejecución anterior. Se descartaron nuevamente las primeras 30,000 iteraciones de cada cadena como tiempo de calentamiento y de la muestra restante se utilizaron los valores obtenidos en las iteraciones múltiplos de 20 para eliminar la dependencia. Así, con esta nueva muestra aleatoria de 10,500 elementos, se obtuvieron como parámetros

	Media empírica	Variancia empírica
α	0.6257	0.002977
σ	2.482	2.322984

En la figura 8.4 se presentan las gráficas de las funciones de densidad estimadas para α y σ .

Capítulo 9

Resultados Finales y Conclusiones

9.1. Conclusión

En esta tesis, se describió una metodología para explicar de manera probabilística el comportamiento de las mediciones de ozono obtenidas por la red de monitoreo de la Ciudad de México. La descripción de los eventos observados se hizo por medio de procesos de Poisson no-homogéneos (Capítulo 5), definidos a partir de su función de intensidad (Capítulo 6). La estimación de los parámetros de los cuales dependen las funciones de intensidad se hizo utilizando un acercamiento Bayesiano (Capítulo 3). Para poder obtener información de la distribución a posteriori de los parámetros, se utilizó una versión del muestreador de Gibbs (Capítulo 7). El funcionamiento de los algoritmos estocásticos utilizados está garantizado por los resultados demostrados relacionados con las cadenas de Markov (Capítulo 4). La metodología desarrollada se aplicó a modelos utilizando tres funciones de intensidad: La funciones Weibull, Musa-Okumoto y Goel-Okumoto generalizada. El criterio para elegir el modelo que mejor se ajusta a los datos observados fué por medio de factores de Bayes.

En la figura 9.1 se muestra la media acumulada de las observaciones de cada zona junto con la media acumulada estimada de los cinco modelos que mejor se apegan a los datos: Weibull con distribución a priori no-informativa (\blacktriangle), Weibull con distribuciones a priori Beta y Gamma (\blacktriangledown), Goel-Okumoto generalizada con distribución a priori no-informativa (\blacksquare) y Musa-Okumoto con distribución a priori no-informativa (\blacklozenge). Los detalles correspondientes a cada modelo pueden ser consultados en el Apéndice D (Simulaciones 1, 2, 3 y 5 respectivamente).

Observación: La figura 9.1 muestra la coherencia de los resultados obtenidos. A pesar de que algunas de las medias obtenidas de los modelos parecen no

ajustarse adecuadamente a la curva, se debe notar la escala utilizada en el eje y , donde se puede ver que la media estimada predice una cantidad muy cercana de días esperados de excedencia de ozono a los observados empíricamente.

Observando los factores de Bayes disponibles en el Apéndice D, se puede ver que la función Weibull con distribución a priori no informativa es la que produce el factor de Bayes más grande para las zonas Noreste y Suroeste y la función Weibull con distribución a priori informativa es la que produce el factor de Bayes más grande para las zonas Centro, Noroeste y Sureste. Por lo tanto, estos modelos deben ser utilizados acorde a este criterio.

Con los resultados obtenidos, es posible calcular la probabilidad de que los niveles de ozono superen las 0.17ppm una cierta cantidad de veces en un intervalo de tiempo dado. Utilizando los parámetros estimados, la probabilidad de que los niveles de ozono sean superados n veces entre los días $(k, k + d]$, donde $k = 0$ para el 01 de Enero del año 2001, se puede calcular utilizando la fórmula 5.4, es decir:

CE)

$$P(N(k+d)-N(k) = n) = e^{-[(k+d)^{0.568} - k^{0.568}]/1.678} \frac{[(k+d)^{0.568} - k^{0.568}]^n}{1.678^n n!}.$$

NE)

$$P(N(k+d)-N(k) = n) = e^{-[(k+d)^{0.704} - k^{0.704}]/17.35} \frac{[(k+d)^{0.704} - k^{0.704}]^n}{17.35^n n!}.$$

NO)

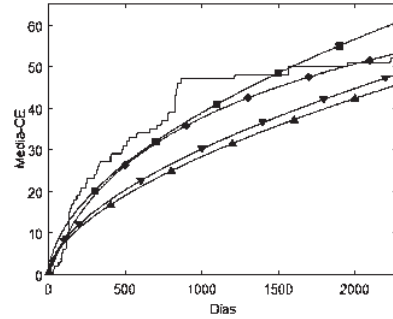
$$P(N(k+d)-N(k) = n) = e^{-[(k+d)^{0.625} - k^{0.625}]/1.766} \frac{[(k+d)^{0.625} - k^{0.625}]^n}{1.766^n n!}.$$

SE)

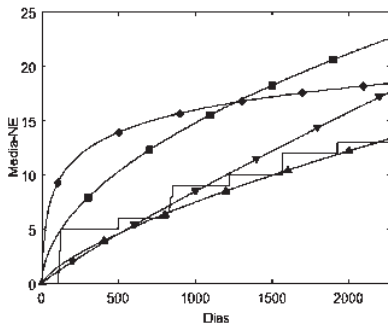
$$P(N(k+d)-N(k) = n) = e^{-[(k+d)^{0.659} - k^{0.659}]/4.582} \frac{[(k+d)^{0.659} - k^{0.659}]^n}{4.582^n n!}.$$

SO)

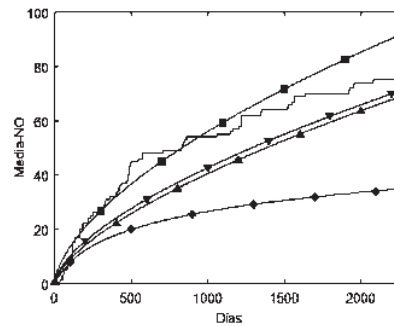
$$P(N(k+d)-N(k) = n) = e^{-[(k+d)^{0.617} - k^{0.617}]/0.861} \frac{[(k+d)^{0.617} - k^{0.617}]^n}{0.861^n n!}.$$



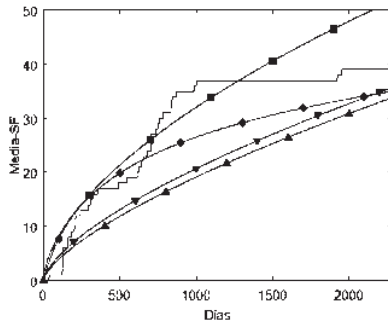
(a) Centro



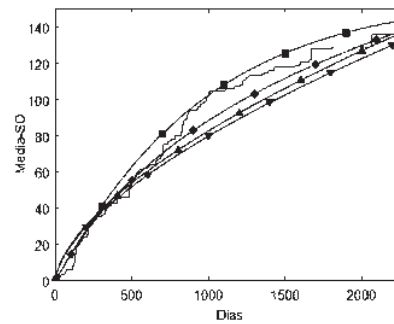
(b) Noreste



(c) Noroeste



(d) Sureste



(e) Suroeste

Figura 9.1: Gráfica de la media de las observaciones de ozono tomadas por la RAMA comparada contra la media de los modelos propuestos. En cada zona ▲ identifica al modelo de Weibull con distribución a priori no-informativa, ▼ identifica al modelo de Weibull con distribución a priori informativa, ■ identifica al modelo de Goel-Okumoto generalizado con distribución a priori no-informativa y ◆ al modelo de Musa-Okumoto.

9.2. Discusión

Comparado con trabajos anteriores ([36], [1]), en este trabajo se programaron directamente los algoritmos utilizando el lenguaje de programación R (<http://www.r-project.org>) en lugar de utilizar paquetes disponibles como WinBugs (<http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml>) el cual a pesar de ser una gran aplicación para implementar algoritmos MCMC, tiene una ejecución notablemente mas lenta comparado con la implementación directa. A pesar de que la velocidad de convergencia a las distribuciones estacionarias fue notablemente más tardada que la requerida usualmente (ver [11] para ejemplos), la implementación directa en R permitió poder ejecutar las cadenas de Markov por una larga cantidad de iteraciones en una cantidad de tiempo razonable.

Por último, cabe notar que aunque en este trabajo el modelo que utilizó la función de riesgo Weibull es la que mejor se ajusta a los datos observados, la metodología explicada en capítulo 8 puede ser utilizada con cualquier función de riesgo propuesta para luego ser comparada con los modelos desarrollados previamente por medio de factores de Bayes. Como se puede notar en la figura 9.1, algunos de los modelos logran un buen ajuste a la media observada por un intervalo de tiempo inicial para después alejarse de las observaciones. Aunque es posible la existencia de una función de riesgo que pueda representar mejor los datos, un acercamiento razonable sería dividir el intervalo de tiempo en dos partes y estimar parámetros distintos para cada una de las dos secciones. Este acercamiento ha sido utilizado previamente en problemas similares, sin embargo, su implementación utilizando las mediciones tomadas en los años que comprende este estudio, no ha sido hecho aún.

Apéndice A

Resultados Complementarios

Lema A.0.1 Sea $A = \{a_1, a_2, \dots\} \subset \mathbb{N}$, tal que $\text{mcd}(A) = 1$, entonces existe $B \subset A$ finito tal que $\text{mcd}(B) = 1$.

Demostración: Sea $b_0 \in A$, si $b_0 = 1$, $B = \{b_0\}$ es el conjunto buscado.

Supongamos que $b_0 \neq 1$, entonces podemos descomponer $b_0 = p_1 p_2 \dots p_n$ como producto único de primos salvo el orden. Para cada p_i , existe un $b_i \in A$ tal que $p_i \nmid b_i$, pues si todos los $b \in B$ fuesen divisibles entre p_i , entonces p_i sería el máximo común divisor de A .

El conjunto $B = \{b_0, b_1, \dots, b_n\}$ es por construcción un subconjunto finito de A tal que $\text{mcd}(B) = 1$, pues si $a \mid b_0$ entonces necesariamente no divide a alguno de los $b_i, i \neq 0$. ■

Corolario A.0.2 Sea $A = \{a_1, a_2, \dots\} \subset \mathbb{N}$ tal que $\text{mcd}(A) = k$, entonces existe $B \subset A$ finito tal que $\text{mcd}(B) = k$.

Demostración: Sea $T = \{t_1, t_2, \dots\}$ el conjunto tal que $t_i = a_i/k$. Entonces $\text{mcd}(T) = 1$ y entonces por el Lema A.0.1 existe $B' \subset T$ finito tal que $\text{mcd}(B') = 1$. Sea $B = kB'$, es decir, el conjunto que consta de todos los elementos de B' multiplicados por la constante k . Entonces B es un conjunto finito tal que $\text{mcd}(B) = k$. ■

Lema A.0.3 Dado un conjunto $A = \{a_1, a_2, \dots, a_m\} \subset \mathbb{N}$ tal que $\text{mcd}(A) = 1$, existe un $N \in \mathbb{N}$ tal que todo $n \geq N$ se puede escribir como combinación lineal positiva de los a_i .

Demostración: Supongamos sin pérdida de generalidad que $a_1 < a_2 < \dots < a_m$. Como $\text{mcd}(A) = 1$ entonces existen $c_1, c_2, \dots, c_m \in \mathbb{Z}$ tales que

$$c_1 a_1 + c_2 a_2 + \dots + c_m a_m = 1.$$

Sea $C = \min\{c_i : 1 \leq i \leq m\}$, definimos ahora $C_i = (a_1 - 1)|C|$, $i \in \{1, 2, \dots, m\}$. Entonces $N = \sum_{i=1}^m C_i a_i$ es el número buscado.

Para ver esto es suficiente notar que cualquier natural dentro del intervalo N y $N+a_1-1$ puede ser escrito como la combinación lineal que resulta en N mas $(n-N) \sum_{i=1}^m c_i a_i$, que por la construcción seguirá teniendo coeficientes positivos. Para $x = N + a_1$ la combinación lineal será simplemente $(c_1 + 1)a_1 + \sum_{i=2}^m a_i c_i$ y así sucesivamente. ■

Corolario A.0.4 *Dado un conjunto $A = \{a_1, a_2, \dots, a_m\} \subset \mathbb{N}$ tal que $\text{mcd}(A) = k$, existe un $N \in \mathbb{N}$ tal que todo $kn \geq N$ se puede escribir como combinación lineal positiva de los a_i*

Demostración: Sea $B = \{b_1, b_2, \dots, b_m\}$ el conjunto tal que $b_i = a_i/k$. Por su construcción, B es un conjunto tal que $\text{mcd}(B) = 1$. Por el Lema A.0.3, existe un $N \in \mathbb{N}$ tal que todo $n \geq N$ puede ser escrito como combinación lineal positiva de los b_i . Esto implica directamente que todo $kn \geq kN$ puede ser escrito como combinación lineal positiva de los elementos de A . ■

Teorema A.0.5 (Teorema de convergencia dominada) *Sea $(a_x^n)_{(x,n) \in \mathbb{N} \times \mathbb{N}}$ una sucesión de números reales con índices en $\mathbb{N} \times \mathbb{N}$ tal que*

- a) $\lim_{n \rightarrow \infty} a_x^n = 0$ para todo $x \in \mathbb{N}$,
 b) existe una sucesión $(b_x)_{x \in \mathbb{N}}$ tal que $\sum_{x \in \mathbb{N}} b_x < \infty$ y $|a_x^n| \leq b_x$ para todo $x, n \in \mathbb{N}$.

Entonces

$$\lim_{n \rightarrow \infty} \sum_{x \in \mathbb{N}} a_x^n = \sum_{x \in \mathbb{N}} \lim_{n \rightarrow \infty} a_x^n = 0.$$

Demostración: Sean F y E conjuntos tales que $F \subset E \subseteq \mathbb{N}$ y F es finito. Entonces,

$$\sum_{x \in E} |a_x^n| = \sum_{x \in F} |a_x^n| + \sum_{x \in (E \setminus F)} |a_x^n|.$$

Por ser una suma finita, el término $\sum_{x \in F} |a_x^n|$ tiene un límite converge, el cual es precisamente 0 por la hipótesis (a). Ahora, por la hipótesis (b), la suma $\sum_{x \in E} |a_x^n|$ tiene un límite finito y así

$$\begin{aligned} \lim_{n \rightarrow \infty} \sum_{x \in E} |a_x^n| &\leq \lim_{n \rightarrow \infty} \sum_{x \in F} |a_x^n| + \limsup_{n \rightarrow \infty} \sum_{x \in (E \setminus F)} |a_x^n| \\ &= \sum_{x \in F} \lim_{n \rightarrow \infty} |a_x^n| + \limsup_{n \rightarrow \infty} \sum_{x \in (E \setminus F)} |a_x^n| \\ &= \limsup_{n \rightarrow \infty} \sum_{x \in (E \setminus F)} |a_x^n| \\ &\leq \limsup_{n \rightarrow \infty} \sum_{x \in (E \setminus F)} b_x \\ &= \sum_{x \in (E \setminus F)} b_x. \end{aligned}$$

Como la desigualdad anterior es válida para cualquier subconjunto F finito de E y la sucesión $(b_x)_{x \in E}$ converge, entonces se sigue que

$$\lim_{n \rightarrow \infty} \sum_{x \in E} |a_x^n| = 0.$$

■

Apéndice B

Distribuciones Utilizadas

B.1. Distribución de Poisson

Notación	$\text{Pois}(\lambda)$
Parámetros	$\lambda > 0 \in \mathbb{R}$
función de masa	$f(k) = \frac{\lambda^k}{k!} e^{-\lambda}$
media	λ
variancia	λ

B.2. Distribución Normal

Notación	$N(\mu, \sigma^2)$
Parámetros	$\mu \in \mathbb{R}, \sigma^2 \geq 0$
función de masa	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
media	μ
variancia	σ^2

B.3. Distribución Beta

Notación	$B(\alpha, \beta)$
Parámetros	$\alpha, \beta > 0$
función de masa	$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 t^{\alpha-1}(1-t)^{\beta-1} dt}$
media	$\frac{\alpha}{\alpha+\beta}$
variancia	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

B.4. Distribución Gamma

Notación	$G(k, \theta)$
Parámetros	$k, \theta > 0$
función de masa	$f(x) = x^{k-1} \frac{\theta^k e^{-x\theta}}{\Gamma(k)}$
media	k/θ
variancia	k/θ^2

Apéndice C

Códigos Fuente

C.1. Weibull

Las entradas de la función son como siguen: zona de la cual se utilizan los datos, cantidad de iteraciones del algoritmo, estado inicial del parámetro alfa, estado inicial del parámetro sigma, hiperparámetros de la distribución beta a priori de alfa e hiperparámetros de la distribución gamma a priori de beta.

```
w<-function(zona,cant_iteraciones,alfa_inicial,
            sigma_inicial,a1,b1,a2,b2){
  #crea las cadenas y guarda el estado inicial
  alfa<-vector("numeric",cant_iteraciones)
  sigma<-vector("numeric",cant_iteraciones)
  alfa[1]<-alfa_inicial
  sigma[1]<-sigma_inicial
  #carga el vector de dias en los cuales el estandar fue excedido
  datos_observados<-scan(file = zona, what = double(0), nmax = -1,
    sep = " ", quote = "\"", dec = ".",
    skip = 0, nlines = 0, na.strings = "NA",
    flush = FALSE, fill = FALSE, strip.white = FALSE,
    quiet = FALSE, blank.lines.skip = TRUE, multi.line = TRUE,
    comment.char = "", allowEscapes = TRUE)
  for(i in 2:cant_iteraciones){
    #obtener valores propuestos a partir de las distribuciones
    #a priori beta y gamma.
    alfa_propuesta<-rbeta(1,a1,b1)
    sigma_propuesta<-rgamma(1,a2,b2)
    #dejar a sigma fijo y ver si
    #el valor propuesto de alfa es aceptado
    aceptacion<-min(1,aceptar_alfa(sigma[i-1],alfa[i-1],
      alfa_propuesta,datos_observados))
```

```

a<-runif(1)
if(a<aceptacion)
  alfa[i]=alfa_propuesta #aceptado
else
  alfa[i]=alfa[i-1] #rechazado
#dejar a alfa fijo y ver si el valor propuesto
#de sigma es aceptado
aceptacion<-min(1,aceptar_sigma(tail(datos_observados,1),
                                length(datos_observados),alfa[i],
                                sigma[i-1],sigma_propuesta))

a<-runif(1)
if(a<aceptacion)
  sigma[i]=sigma_propuesta #aceptado
else
  sigma[i]=sigma[i-1] #rechazado
}
}
aceptar_alfa<-function(sigma_actual,alfa_actual,
                      alfa_propuesta,datos_observados){
  n<-length(datos_observados)
  dn<-tail(datos_observados,1)
  logsum<-0
  for(k in 1:n){
    logsum<-logsum+log(datos_observados[k])
  }
  term<-n*log(alfa_propuesta/alfa_actual)+n*log(sigma_actual)
    *(alfa_actual-alfa_propuesta)+logsum
    *(alfa_propuesta-alfa_actual)
    +(dn/sigma_actual)^(alfa_actual)
    -(dn/sigma_actual)^(alfa_propuesta)
  exp(term)
}
aceptar_sigma<-function(dn,n,alfa_actual,sigma_actual,
                      sigma_propuesta){
  term<-n*alfa_actual*log(sigma_actual/sigma_propuesta)
    +(dn/sigma_actual)^(alfa_actual)
    -(dn/sigma_propuesta)^(alfa_actual)
  exp(term)
}

```

C.2. Musa-Okumoto

Las entradas de la función son como siguen: zona de la cual se utilizan los datos, cantidad de iteraciones del algoritmo, estado inicial del parámetro alfa, estado inicial del parámetro beta, hiperparámetros de la distribución gamma a priori de alfa e hiperparámetros de la distribución gamma a priori de beta.

```
mo<-function(zona, cant_iteraciones, alfa_inicial,
             beta_inicial, a3, b3, a4, b4){
  #creacion de las cadenas y la insercion del estado inicial
  alfa<-vector("numeric", cant_iteraciones)
  beta<-vector("numeric", cant_iteraciones)
  alfa[1]<-alfa_inicial
  beta[1]<-beta_inicial
  #cargar el vector de dias en los cuales el estandar fue excedido
  datos_observados<-scan(file = zona, what = double(0), nmax = -1,
    sep = " ", quote = "", dec = ".",
    skip = 0, nlines = 0, na.strings = "NA",
    flush = FALSE, fill = FALSE, strip.white = FALSE,
    quiet = FALSE, blank.lines.skip = TRUE, multi.line = TRUE,
    comment.char = "", allowEscapes = TRUE)
  n<-length(datos_observados)
  dn<-tail(datos_observados, 1)
  for(i in 2:cant_iteraciones){
    #obtener un valor de propuesta de la distribucion a priori
    alfa_propuesta<-rgamma(1, a3, b3)
    aceptacion<-aceptar_alfa(beta[i-1], alfa[i-1], alfa_propuesta,
                             datos_observados)

    a<-runif(1)
    if(a<aceptacion)
      alfa[i]=alfa_propuesta #aceptado
    else
      alfa[i]=alfa[i-1] #rechazado
    #se obtiene una muestra aleatoria de la distribucion gamma.
    beta[i]<-rgamma(1, n+1, log(1+(dn/alfa[i])))
  }
}
aceptar_alfa<-function(current_beta, alfa_actual, alfa_propuesta,
                      datos_observados){
  #extraer longitud y el ultimo dia de las observaciones
  n<-length(datos_observados)
  dn<-tail(datos_observados, 1)
  #calcular la suma de log(di)
```

```

logsum<-0
for(i in 1:n){
  logsum<-logsum+log((datos_observados[i]+alfa_actual)
                    /(datos_observados[i]+alfa_propuesta))
}
term<-logsum+current_beta*log((alfa_actual*(alfa_propuesta+dn))
                              /(alfa_propuesta*(alfa_actual+dn)))
exp(term)
}

```

C.3. Goel-Okumoto Generalizada

Las entradas de la función son como siguen: zona de la cual se utilizan los datos, cantidad de iteraciones del algoritmo, estado inicial del parámetro alfa, estado inicial del parámetro beta, hiperparámetros de la distribución uniforme a priori de alfa, hiperparámetros de la distribución uniforme a priori de beta e hiperparámetros de la distribución gamma a priori de gamma.

```

ggo<-function(zona, cant_iteraciones, alfa_inicial,
             beta_inicial, gamma_inicial, a5, b5,
             a6, b6, a7, b7){
#create chains 1,2 and 3 and store initial values
alfa<-vector("numeric", cant_iteraciones)
beta<-vector("numeric", cant_iteraciones)
gamma<-vector("numeric", cant_iteraciones)
alfa[1]<-alfa_inicial
beta[1]<-beta_inicial
gamma[1]<-gamma_inicial
#carga el vector de dias en los cuales el estandar
#fue excedido
datos_observados<-scan(file = zona, what = double(0), nmax = -1,
  sep = " ", quote = "\"", dec = ".",
  skip = 0, nlines = 0, na.strings = "NA",
  flush = FALSE, fill = FALSE, strip.white = FALSE,
  quiet = FALSE, blank.lines.skip = TRUE, multi.line = TRUE,
  comment.char = "\"", allowEscapes = TRUE)
logsum<-0
for(i in 1:length(datos_observados)){
  logsum<-logsum+log(datos_observados[i])
}
for(i in 2:cant_iteraciones){
#obtener valores propuestos a partir de las distribuciones
#a priori uniformes y gamma.
alfa_propuesto<-runif(1, a5, b5)

```

```

beta_propuesto<-runif(1,a6,b6)
gamma_propuesto<-rgamma(1,a7,b7)
#dejar beta y gamma fijos y ver si alfa es aceptado
aceptacion<-min(1,aceptar_alfa(beta[i-1],alfa[i-1],
                             alfa_propuesto ,datos_observados ,gamma[i-1]))
a<-runif(1)
if(a<aceptacion)
  alfa[i]=alfa_propuesto #aceptado
else
  alfa[i]=alfa[i-1] #rechazado
#dejar alfa y gamma fijos y ver si alfa es aceptado
aceptacion<-min(1,aceptar_beta(tail(datos_observados,1),
                              length(datos_observados),alfa[i],
                              beta[i-1],beta_propuesto ,
                              datos_observados ,gamma[i-1]))
a<-runif(1)
if(a<aceptacion)
  beta[i]=beta_propuesto #aceptado
else
  beta[i]=beta[i-1] #rechazado
#dejar alfa y beta fijos y ver si alfa es aceptado
aceptacion<-min(1,aceptar_gamma(datos_observados ,logsum ,
                                gamma_propuesto ,gamma[i-1],beta[i-1],alfa[i-1]))
a<-runif(1)
if(a<aceptacion)
  gamma[i]=gamma_propuesto #aceptado
else
  gamma[i]=gamma[i-1] #rechazado
}
}
aceptar_alfa<-function(beta_actual , alfa_actual ,
                      alfa_propuesto , datos_observados ,
                      gamma_actual){
  n<-length(datos_observados)
  dn<-tail(datos_observados,1)
  res<-(alfa_propuesto/alfa_actual)^n
  *exp(( alfa_actual-alfa_propuesto)
  *(1-exp(-beta_actual*dn^gamma_actual)))
}
aceptar_beta<-function(dn,n, alfa_actual , beta_actual ,
                      beta_propuesto , datos_observados ,
                      gamma_actual){
  disum<-0
  for(i in 1:n){

```

```

    disum<-disum+datos_observados [ i ] ^ gamma_actual
  }
  term<-disum*(beta_actual-beta_propuesto)
    +alfa_actual*(exp(-beta_propuesto*dn^gamma_actual)
    -exp(-beta_actual*dn^gamma_actual))
  res<-((beta_propuesto/beta_actual)^n)*exp(term)
  res
}
aceptar_gamma<-function ( datos_observados , logsum ,
                          gamma_propuesto , gamma_actual ,
                          beta_actual , alfa_actual ){
  n<-length ( datos_observados )
  dn<-tail ( datos_observados , 1 )
  expsum<-0
  for ( i in 1:n ){
    expsum<-expsum+(datos_observados [ i ] ^ gamma_actual
    -datos_observados [ i ] ^ gamma_propuesto)
  }
  term<-logsum*(gamma_propuesto-gamma_actual)
    +beta_actual*expsum+alfa_actual
    *(exp(-beta_actual*dn^gamma_propuesto)
    -exp(-beta_actual*dn^gamma_actual))
  result <-((gamma_propuesto/gamma_actual)^n)*exp(term)
  result
}

```

Apéndice D

Resultados Completos

D.1. Centro

Simulación 1:

Forma:	Weibull
$p(\alpha)$:	$U(0, 1)$
$p(\sigma)$:	$U(0, 10)$
Cantidad de cadenas ejecutadas:	3
Estados iniciales de alfa:	0.8, 0.6, 0.9
Estados iniciales de sigma:	4, 1, 7
Tamaño por cadena:	100,000 iteraciones
Tiempo de calentamiento:	30,000 iteraciones
Tamaño de muestra aleatoria final:	10,500
Factor de Bayes:	8.2569e-26

	Media	Variancia	95 % Confianza
α	0.5781	0.063	0.46224-0.69967
σ	3.075	2.136	0.43255-8.492

Simulación 2:

Forma:	Weibull
$p(\alpha)$:	$B(1/8, 1/8)$
$p(\sigma)$:	$G(9/4, 3/4)$
Cantidad de cadenas ejecutadas:	3
Estados iniciales de alfa:	0.8, 0.6, 0.9
Estados iniciales de sigma:	4, 1, 7
Tamaño por cadena:	200,000 iteraciones
Tiempo de calentamiento:	130,000 iteraciones
Tamaño de muestra aleatoria final:	10,500
Factor de Bayes:	5.6684e-29

	Media	Variancia	95 % Confianza
α	0.568	0.052	0.4712537-6.743113
σ	2.488	1.502	0.5245473-6.228336

Simulación 3:

Forma:	Goel-Okumoto generalizada
$p(\alpha)$:	$U(300, 2000)$
$p(\beta)$:	$U(0.0000001, 0.001)$
$p(\gamma)$:	$G(0.5, 1.5)$
Cantidad de cadenas ejecutadas:	3
Estados iniciales de alfa:	1000, 1300, 1400
Estados iniciales de beta:	0.0003, 0.0006, 0.0008
Estados iniciales de gamma:	1, 1.1, 1.2
Tamaño por cadena:	100,000 iteraciones
Tiempo de calentamiento:	30,000 iteraciones
Tamaño de muestra aleatoria final:	10,500
Factor de Bayes:	1.52702e-104

	Media	Variancia	95 % Confianza
α	1309	424.0873	486.9847-1964.205
β	0.000675	0.0002	0.0002656413-0.0009851874
γ	0.55	0.0573	0.4494016-0.6712511

Simulación 4:

Forma:	Goel-Okumoto generalizada
$p(\alpha)$:	$U(200, 3000)$
$p(\beta)$:	$U(0.0000001, 0.01)$
$p(\gamma)$:	$G(0.5, 1.5)$
Cantidad de cadenas ejecutadas:	3
Estados iniciales de alfa:	1000, 1300, 1400
Estados iniciales de beta:	0.0003, 0.0006, 0.0008
Estados iniciales de gamma:	1, 1.1, 1.2
Tamaño por cadena:	200,000 iteraciones
Tiempo de calentamiento:	130,000 iteraciones
Tamaño de muestra aleatoria final:	10,500
Factor de Bayes:	2.379792e-105

	Media	Variancia	95 % Confianza
α	980.6	632628.3	210.435-2817.569
β	0.002343	4.305797e-06	210.435-2817.569
γ	0.5135	0.00547119	0.3729599-0.6658699

Simulación 5:

Forma:	Musa-Okumoto
$p(\alpha)$:	$U(0, 700)$
$p(\sigma)$:	$U(1, 2500)$
Cantidad de cadenas ejecutadas:	3
Estados iniciales de alfa:	1, 7, 350
Estados iniciales de beta:	10, 20, 50
Tamaño por cadena:	100,000 iteraciones
Tiempo de calentamiento:	30,000 iteraciones
Tamaño de muestra aleatoria final:	10,500
Factor de Bayes:	2.673235e-108

	Media	Variancia	95 % Confianza
α	209.7355	102.5246	73.91646-476.3019
β	21.49208	4.840973	13.94708-32.71157

D.2. Noreste

Simulación 1:

Forma:	Weibull
$p(\alpha)$:	$U(0, 1)$
$p(\sigma)$:	$U(0, 100)$
Cantidad de cadenas ejecutadas:	3
Estados iniciales de alfa:	0.8, 0.6, 0.9
Estados iniciales de sigma:	53, 70, 30
Tamaño por cadena:	100,000 iteraciones
Tiempo de calentamiento:	30,000 iteraciones
Tamaño de muestra aleatoria final:	10,500
Factor de Bayes:	3.499e-16

	Media	Variancia	95 % Confianza
α	0.7	0.11	0.476-0.917
σ	3.47	2.37	0.7635-9.652

Simulación 2:

Forma:	Weibull
$p(\alpha)$:	$B(7/200, 3/200)$
$p(\sigma)$:	$G(3249/640, 57/640)$
Cantidad de cadenas ejecutadas:	3
Estados iniciales de alfa:	0.8, 0.6, 0.9
Estados iniciales de sigma:	53, 70, 30
Tamaño por cadena:	200,000 iteraciones
Tiempo de calentamiento:	130,000 iteraciones
Tamaño de muestra aleatoria final:	10,500
Factor de Bayes:	2.0712e-19

	Media	Variación	95 % Confianza
α	0.891	0.14	0.575-1
σ	91.05	33.404	28.909-152.84

Simulación 3:

Forma:	Goel-Okumoto generalizada
$p(\alpha)$:	$U(300, 2000)$
$p(\beta)$:	$U(0.0000001, 0.001)$
$p(\gamma)$:	$G(0.5, 1.5)$
Cantidad de cadenas ejecutadas:	3
Estados iniciales de alfa:	1000, 1300, 1400
Estados iniciales de beta:	0.0003, 0.0006, 0.0008
Estados iniciales de gamma:	1, 1.1, 1.2
Tamaño por cadena:	100,000 iteraciones
Tiempo de calentamiento:	30,000 iteraciones
Tamaño de muestra aleatoria final:	10,500
Factor de Bayes:	9.081666e-35

	Media	Variación	95 % Confianza
α	1024	483.15	330.6505-1938.551
β	3.946e-04	3.946e-04	2.339055e-05-0.0009471859
γ	0.5222	0.126	0.3216789-0.8165164

Simulación 4:

Forma:	Goel-Okumoto generalizada
$p(\alpha)$:	$U(100, 2500)$
$p(\beta)$:	$U(0.0000001, 0.01)$
$p(\gamma)$:	$G(0.5, 1.5)$
Cantidad de cadenas ejecutadas:	3
Estados iniciales de alfa:	1000, 1300, 1400
Estados iniciales de beta:	0.0003, 0.0006, 0.0008
Estados iniciales de gamma:	1, 1.1, 1.2
Tamaño por cadena:	300,000 iteraciones
Tiempo de calentamiento:	230,000 iteraciones
Tamaño de muestra aleatoria final:	10,500
Factor de Bayes:	1.052917e-35

	Media	Variación	95 % Confianza
α	807.6	424377	109.6125-2330.337
β	1.516e-03	3.65897e-06	6.588586e-05-0.007344133
γ	0.4671	0.01527799	0.2554738-0.7210053

Simulación 5:

Forma:	Musa-Okumoto
$p(\alpha)$:	$U(0, 700)$
$p(\sigma)$:	$U(1, 2500)$
Cantidad de cadenas ejecutadas:	3
Estados iniciales de alfa:	1, 7, 350
Estados iniciales de beta:	10, 20, 50
Tamaño por cadena:	100,000 iteraciones
Tiempo de calentamiento:	30,000 iteraciones
Tamaño de muestra aleatoria final:	10,500
Factor de Bayes:	3.564126e-39

	Media	Variancia	95 % Confianza
α	199.178	127346.4	0-1115
β	2.981576	26.99194	0-16.09559

D.3. Noroeste

Simulación 1:

Forma:	Weibull
$p(\alpha)$:	$U(0, 1)$
$p(\sigma)$:	$U(0, 100)$
Cantidad de cadenas ejecutadas:	3
Estados iniciales de alfa:	0.8, 0.6, 0.9
Estados iniciales de sigma:	10, 5, 20
Tamaño por cadena:	100,000 iteraciones
Tiempo de calentamiento:	30,000 iteraciones
Tamaño de muestra aleatoria final:	10,500
Factor de Bayes:	1.684e-44

	Media	Variancia	95 % Confianza
α	0.6534	0.063	0.543-0.792
σ	3.47	2.37	0.763-9.561

Simulación 2:

Forma:	Weibull
$p(\alpha)$:	$B(3/25, 2/25)$
$p(\sigma)$:	$G(3/2, 1/2)$
Cantidad de cadenas ejecutadas:	3
Estados iniciales de alfa:	0.8, 0.6, 0.9
Estados iniciales de sigma:	10, 5, 20
Tamaño por cadena:	100,000 iteraciones
Tiempo de calentamiento:	30,000 iteraciones
Tamaño de muestra aleatoria final:	10,500
Factor de Bayes:	6.4851e-35

	Media	Variancia	95 % Confianza
α	0.6257	0.054	0.5262227-0.7268434
σ	2.482	1.524	0.6420253-5.884047

Simulación 3:

Forma:	Goel-Okumoto generalizada
$p(\alpha)$:	$U(300, 2000)$
$p(\beta)$:	$U(0.0000001, 0.001)$
$p(\gamma)$:	$G(0.5, 1.5)$
Cantidad de cadenas ejecutadas:	3
Estados iniciales de alfa:	1000, 1300, 1400
Estados iniciales de beta:	0.0003, 0.0006, 0.0008
Estados iniciales de gamma:	1, 1.1, 1.2
Tamaño por cadena:	100,000 iteraciones
Tiempo de calentamiento:	30,000 iteraciones
Tamaño de muestra aleatoria final:	10,500
Factor de Bayes:	7.395323e-144

	Media	Variancia	95 % Confianza
α	1200	454.73	389.1736-1957.981
β	0.0006416	0.00022	0.0002121242-0-0009803474
γ	0.6241	0.0624	0.5173508-0.7563098

Simulación 4:

Forma:	Goel-Okumoto generalizada
$p(\alpha)$:	$U(100, 3000)$
$p(\beta)$:	$U(0.0000001, 0.01)$
$p(\gamma)$:	$G(0.5, 1.5)$
Cantidad de cadenas ejecutadas:	3
Estados iniciales de alfa:	1000, 1300, 1400
Estados iniciales de beta:	0.0003, 0.0006, 0.0008
Estados iniciales de gamma:	1, 1.1, 1.2
Tamaño por cadena:	200,000 iteraciones
Tiempo de calentamiento:	130,000 iteraciones
Tamaño de muestra aleatoria final:	10,500
Factor de Bayes:	2.102242e-143

	Media	Variancia	95 % Confianza
α	283.3	164149.9	101.0952-1702.55
β	0.00361	4.347517e-06	0.000499058-0.00846268s
γ	0.6959	0.01005014	0.5146901-0.9015067

Simulación 5:

Forma:	Musa-Okumoto
$p(\alpha)$:	$U(0, 700)$
$p(\sigma)$:	$U(1, 2500)$
Cantidad de cadenas ejecutadas:	3
Estados iniciales de alfa:	1, 7, 350
Estados iniciales de beta:	10, 20, 50
Tamaño por cadena:	100,000 iteraciones
Tiempo de calentamiento:	30,000 iteraciones
Tamaño de muestra aleatoria final:	10,500
Factor de Bayes:	1.996936e-146

	Media	Variancia	95 % Confianza
α	96.21955	28191.97	0-536.9866
β	10.90067	318.4107	0-49.76161

D.4. Sureste

Simulación 1:

Forma:	Weibull
$p(\alpha)$:	$U(0, 1)$
$p(\sigma)$:	$U(0, 100)$
Cantidad de cadenas ejecutadas:	3
Estados iniciales de alfa:	0.8, 0.6, 0.9
Estados iniciales de sigma:	18, 30, 10
Tamaño por cadena:	100,000 iteraciones
Tiempo de calentamiento:	30,000 iteraciones
Tamaño de muestra aleatoria final:	10,500
Factor de Bayes:	2.6252e-29

	Media	Variancia	95 % Confianza
α	0.6896	0.093	0.5255692-0.8845968
σ	13.83	10	2.129357-39.86523

Simulación 2:

Forma:	Weibull
$p(\alpha)$:	$B(3/25, 2/25)$
$p(\sigma)$:	$G(169/100, 13/100)$
Cantidad de cadenas ejecutadas:	3
Estados iniciales de alfa:	0.8, 0.6, 0.9
Estados iniciales de sigma:	18, 30, 10
Tamaño por cadena:	100,000 iteraciones
Tiempo de calentamiento:	30,000 iteraciones
Tamaño de muestra aleatoria final:	10,500
Factor de Bayes:	1.00755e-28

	Media	Variación	95 % Confianza
α	0.6591	0.075	0.5171135-0.8112807
σ	10.07	6.373	1.984287-25.83128

Simulación 3:

Forma:	Goel-Okumoto generalizada
$p(\alpha)$:	$U(300, 2000)$
$p(\beta)$:	$U(0.0000001, 0.001)$
$p(\gamma)$:	$G(0.5, 1.5)$
Cantidad de cadenas ejecutadas:	3
Estados iniciales de alfa:	1000, 1300, 1400
Estados iniciales de beta:	0.0003, 0.0006, 0.0008
Estados iniciales de gamma:	1, 1.1, 1.2
Tamaño por cadena:	100,000 iteraciones
Tiempo de calentamiento:	30,000 iteraciones
Tamaño de muestra aleatoria final:	10,500
Factor de Bayes:	4.714356e-86

	Media	Variación	95 % Confianza
α	1059	480.94	332.1221-1942.537
β	0.000516	0.0025	0.0001180893-0.0009694225
γ	0.5917	0.076	0.4511069-0.7464183

Simulación 4:

Forma:	Goel-Okumoto generalizada
$p(\alpha)$:	$U(0, 3000)$
$p(\beta)$:	$U(0.0000001, 0.01)$
$p(\gamma)$:	$G(0.5, 1.5)$
Cantidad de cadenas ejecutadas:	3
Estados iniciales de alfa:	1000, 1300, 1400
Estados iniciales de beta:	0.0003, 0.0006, 0.0008
Estados iniciales de gamma:	1, 1.1, 1.2
Tamaño por cadena:	200,000 iteraciones
Tiempo de calentamiento:	130,000 iteraciones
Tamaño de muestra aleatoria final:	10,500
Factor de Bayes:	1.125576e-83

	Media	Variación	95 % Confianza
α	196.1	190671.7	32.1058-1726.108
β	0.00231	4.510331e-06	0.000227808-0.008254914
γ	0.8815	0.05321338	0.4666955-1.268913

Simulación 5:

Forma:	Musa-Okumoto
$p(\alpha)$:	$U(0, 700)$
$p(\sigma)$:	$U(1, 2500)$
Cantidad de cadenas ejecutadas:	3
Estados iniciales de alfa:	1, 7, 350
Estados iniciales de beta:	10, 20, 50
Tamaño por cadena:	100,000 iteraciones
Tiempo de calentamiento:	30,000 iteraciones
Tamaño de muestra aleatoria final:	10,500
Factor de Bayes:	4.402195e-89

	Media	Variancia	95 % Confianza
α	307.7262	148.8562	134.278-672.9357
β	20.83115	4.98808	12.37612-31.46646

D.5. Suroeste

Simulación 1:

Forma:	Weibull
$p(\alpha)$:	$U(0, 1)$
$p(\sigma)$:	$U(0, 100)$
Cantidad de cadenas ejecutadas:	3
Estados iniciales de alfa:	0.8, 0.6, 0.7
Estados iniciales de sigma:	0.7, 0.6, 0.8
Tamaño por cadena:	100,000 iteraciones
Tiempo de calentamiento:	30,000 iteraciones
Tamaño de muestra aleatoria final:	10,500
Factor de Bayes:	2.7104e-69

	Media	Variancia	95 % Confianza
α	0.6176	0.013	0.5918602-0.642426
σ	0.7854	0.087	0.6260739-0.9625298

Simulación 2:

Forma:	Weibull
$p(\alpha)$:	$B(3/25, 2/25)$
$p(\sigma)$:	$G(49/20, 7/20)$
Cantidad de cadenas ejecutadas:	3
Estados iniciales de alfa:	0.8, 0.6, 0.7
Estados iniciales de sigma:	0.7, 0.6, 0.8
Tamaño por cadena:	200,000 iteraciones
Tiempo de calentamiento:	130,000 iteraciones
Tamaño de muestra aleatoria final:	10,500
Factor de Bayes:	7.1844e-57

	Media	Variación	95 % Confianza
α	0.6193	0.029	0.5692607-0.6758804
σ	0.8487	0.31	0.3876562-1.541083

Simulación 3:

Forma:	Goel-Okumoto generalizada
$p(\alpha)$:	$U(300, 2000)$
$p(\beta)$:	$U(0.0000001, 0.001)$
$p(\gamma)$:	$G(0.5, 1.5)$
Cantidad de cadenas ejecutadas:	3
Estados iniciales de alfa:	1000, 1300, 1400
Estados iniciales de beta:	0.0003, 0.0006, 0.0008
Estados iniciales de gamma:	1, 1.1, 1.2
Tamaño por cadena:	180,000 iteraciones
Tiempo de calentamiento:	115,000 iteraciones
Tamaño de muestra aleatoria final:	9,750
Factor de Bayes:	7.838641e-233

	Media	Variación	95 % Confianza
α	710	450.3	303.1205-1837.182
β	0.0007243	0.000724	0.0003053257-0.0009878492
γ	0.7728	0.0758	0.583-0.9612

Simulación 4:

Forma:	Goel-Okumoto generalizada
$p(\alpha)$:	$U(0, 3000)$
$p(\beta)$:	$U(0.0000001, 0.01)$
$p(\gamma)$:	$G(0.5, 1.5)$
Cantidad de cadenas ejecutadas:	3
Estados iniciales de alfa:	1000, 1300, 1400
Estados iniciales de beta:	0.0003, 0.0006, 0.0008
Estados iniciales de gamma:	1, 1.1, 1.2
Tamaño por cadena:	200,000 iteraciones
Tiempo de calentamiento:	130,000 iteraciones
Tamaño de muestra aleatoria final:	10,500
Factor de Bayes:	2.076314e-229

	Media	Variación	95 % Confianza
α	157.3	331.4336	136.7958-181.0212
β	0.0008039	2.033309e-07	0.0002530337-0.001957276
γ	1.039	0.006723714	0.873712-1.196391

Simulación 5:

Forma:	Musa-Okumoto		
$p(\alpha)$:	$U(0, 700)$		
$p(\sigma)$:	$U(1, 2500)$		
Cantidad de cadenas ejecutadas:	3		
Estados iniciales de alfa:	1, 7, 350		
Estados iniciales de beta:	10, 20, 50		
Tamaño por cadena:	100,000 iteraciones		
Tiempo de calentamiento:	30,000 iteraciones		
Tamaño de muestra aleatoria final:	10,500		
Factor de Bayes:	4.037815e-250		
	Media	Variación	95 % Confianza
α	507.8983	109.1856	292.9967-687.4573
β	81.2965	10.885889	60.82191-102.7176

Índice alfabético

- σ -álgebra, 7
- Cadena
 - de Markov, 21
 - estacionaria, 30
 - homogénea, 22
 - reversible, 43
- Condición de Markov, 21
- Conjunto potencia, 8
- Distribución
 - a posteriori, 17
 - a priori, 16
 - conjugada, 18
 - estacionaria, 30
 - no-informativa, 16
- Espacio
 - de estados, 21
 - muestral, 7
 - parametral, 15
- Esperanza, 12
- Estado
 - accesible, 25
 - aperiódico, 25
 - cerrado, 29
 - comunicado, 25
 - ergódico, 30
 - irreducible, 29
 - nulo recurrente, 29
 - periódico, 25
 - positivo recurrente, 29
 - recurrente, 27
 - transitorio, 27
- Evento, 7
 - independiente, 10
- Experimento, 7
 - aleatorio, 7
- Factor de Bayes, 19
- Función
 - conjunta de probabilidad, 13
 - de densidad, 11
 - de distribución, 11
 - de distribución conjunta, 13
 - de probabilidad, 11
 - de riesgo, 53
 - de supervivencia, 53
 - de verosimilitud, 15
- Ley de los grandes números, 60
- Método Monte Carlo, 61
- Matriz de transición, 22
- Muestra
 - aleatoria, 7
- Parámetro de una variable aleatoria, 15
- Periodo de un estado, 25
- Probabilidad
 - condicional, 9
 - espacio de, 8
 - medida de, 8
- Proceso
 - de Poisson homogéneo, 46
 - de Poisson no-homogéneo, 51
- Proceso estocástico, 21
- Teorema
 - Chapman-Kolmogorov, 23
 - de Bayes, 9
 - de la Descomposición, 29

Fundamental de Convergencia, 39

Variable aleatoria, 10

continua, 11

discreta, 11

Variancia, 12

Bibliografía

- [1] L.J. Álvarez, A.A. Fernández-Bremauntz, E.R. Rodrigues, and G. Tzintzun. Maximum a posteriori estimation of the daily ozone peaks in mexico city. *Journal of Agricultural, Biological, and Environmental Statistics*, 10:276–290, 2005.
- [2] J. Austin and H. Tran. A characterization of the weekday-weekend behavior of ambient ozone concentrations in california. *Air Pollution*, VII:645–661, 1999.
- [3] M.L Bell, R. Goldberg, C. Rogrefe, P.L. Kinney, K. Knowlton, B. Lynn, J. Rosenthal, C. Rosenzweig, and J.A. Patz. Climate change, ambient ozone, and health in 50 us cities. *Climate Change*, pages 61–75, 2007.
- [4] M.L. Bell, A. McDermontt, S.L. Zeger, J.M. Samet, and F. Dominici. Ozone and short-term mortality in 95 us urban communities. *Journal of the American Medical society*, pages 2372–2378, 2004.
- [5] William M. Bolstad. *Introduction to Bayesian Statistics*. 2007.
- [6] David R. Cox. *Statistical Analysis of Series of Events*. Springer, 1966.
- [7] Diario Oficial de la Federación. Modificación a la norma oficial mexicana. Technical Report NOM-020-SSA1, NOM, October 1993.
- [8] EPA. Guidelines for the interpretation of ozone air quality standards. Technical Report EPA-450/4-79-003, US Environmental Protection Agency, 1979.
- [9] J.B. Flaum, S.T. Rao, and I.G. Zurbenko. Moderating influence of meterological conditions on ambient ozone centrations. *Air and Waste Management Assoc.*, (46):33–46, 1996.
- [10] Margarita Reyes Flores. El método de monte carlo vía cadenas de markov y su uso en problemas de epidemiología. Master’s thesis, Universidad Nacional Autónoma de México, 2008.
- [11] Andrew Gelman. *Inference and monitoring convergence*. Chapman and Hall, 1996.

- [12] R. Guardani, C.A.O. Nascimento, J.L. Aguilar, C.I.V. Lacava, and Y. Yanagi. Ground-level ozone mapping in large urban areas using multivariate analysis: application to the são paulo metropolitan area. *Air and Waste Management Assoc*, (20):733–748, 2003.
- [13] R. Guardani, C.A.O. Nascimento, M.L.G. Guardani, M.H.R.B. Martins, and J. Romando. Study of atmospheric ozone formation by means of a neural network based model. *Air and Waste Management Assoc*, (49):316–323, 1999.
- [14] J. Horowitz. Extreme values from a nonstationary stochastic process: an application to air quality analysis. *Technometrics*, (22):469–482, 1980.
- [15] J.S. Javits. Statistical interdependencies in the ozone national ambient air quality standard. *Air Poll. control Assoc.*, (30):469–482, 1980.
- [16] Julian Jara-Ettinger Jorge A. Achcar, Eliane R. Rodrigues. *A Gibbs Sampling Algorithm to Estimate the Occurrence of Ozone Exceedances in Mexico City*. InTech, 2011.
- [17] L.C. Larsen, R.A. Bradley, and G.L. Honcoop. A new method of characterizing the variability of air quality-related indicators. *Air and Waste Management Association's International Speciality Conference of Tropospheric Ozone and the Environment, Los Angeles, USA.*, 1990.
- [18] J. F. Lawless. *Statistical Models and Methods for Lifetime Data*. John Wiley and Sons, 1982.
- [19] D. Loomis, V.H. Borja-Arbutto, S.I. Bangdiwala, and C.M. Shy. Ozone exposure and daily mortality in mexico city: a time series analysis. *Health Effects Institute Research Report*, 75:1–46, 1996.
- [20] Richard T. Burnett y C. Arden Pope Michael Jerrett. Long-term ozone exposure and mortality. *N. Engl. J. Med*, 2009.
- [21] J. N. Pan and S.T. Chen. Monitoring long-memory air quality data using afirma model. *Environmetrics*, 2007.
- [22] A.E. Raftery. Are ozone exceedance rate decreasing?, comment of the paper "extreme value analysis of environmental time series: an application to trend detection in ground-level ozone" by r. l. smith. *Statistical Sciences*, (4):378–381, 1989.
- [23] Yesenia Alfaro Ramírez. Estimación de líneas ancestrales en genética utilizando el coalescente de kingman y métodos de monte carlo. Master's thesis, Universidad Nacional Autónoma de México, 2008.

- [24] Sandra Romero-Hidalgo. *Análisis de las herramientas metodológicas para evidencias ligamiento en enfermedades genéticas comunes*. PhD thesis, Posgrado en Ciencias Biomédicas, UNAM, 2006.
- [25] Jeffrey Rosenthal. Understanding the metropolis-hastings algorithm. *American Statistician*, 49:327–335, 1995.
- [26] Sheldon M. Ross. *Stochastic Processes*. Wiley, 1995.
- [27] Resnick. Sidney. *A Probability Path*. Birkhauser Boston, 1999.
- [28] R.L. Smith. Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statistical Sciences*, 4:367–393, 1989.
- [29] Daniel W. Stroock. *An Introduction to Markov Processes*. Springer, 1995.
- [30] Constantin Tudor. *Procesos Estocásticos*. Sociedad Matemática Mexicana, 2002.
- [31] Elizabeth K. Wilson. Ozone’s health impact. *American Chemical Society Publications*, 2009.
- [32] A. Zolghadri y D. Henry. Application to ozone time series data measured in bordeaux. *Minimax statistical models for air pollution time series*, 2004.
- [33] Geoffrey R. Grimmett y David R. Stirzaker. *Probability and Random Processes*. Oxford University Press, 2001.
- [34] S. Richardson y David Spiegelhalter W.R. Gilks. *Markov Chain Monte Carlo in Practice*. Chapman and Hall, 1995.
- [35] Reuven Y. Rubinstein y Drik P. Kroese. *Simulation and the Monte Carlo Method*. 2007.
- [36] Adrián A. Fernández-Bremauntz Eliane R. Rodrigues y Guadalupe Tzintzun Jorge A. Achcar. Estimating the number of ozone peaks in mexico city using a non-homogeneous poisson model. *Wiley InterScience*, 2007.
- [37] Dani Gamerman y Hedibert F. Lopes. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman and Hall, 2006.
- [38] Samuel Karlin y Howard M. Taylor. *A First Course in Stochastic Processes*. Academic Press, 1975.
- [39] Charles M. Grinstead y J. Laurie Snell. *Introduction to Probability*. 1997.