



UNIVERSIDAD NACIONAL AUTÓNOMA DE  
MÉXICO

---

---

FACULTAD DE CIENCIAS

EL MÉTODO DE MONTE CARLO VÍA CADENAS DE MARKOV  
Y SU USO EN PROBLEMAS DE EPIDEMIOLOGÍA

T E S I S

QUE PARA OBTENER EL TÍTULO DE

MATEMÁTICA

PRESENTA

MARGARITA REYES FLORES

DIRECTORA DE TESIS:

DRA. ELIANE REGINA RODRIGUES



2008



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

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

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

## Hoja de Datos del Jurado

1. Datos del alumno  
Reyes  
Flores  
Margarita  
55 73 45 42  
Universidad Nacional Autónoma de México  
Facultad de Ciencias  
Matemáticas  
097145857
2. Datos del tutor  
Dra.  
Eliane Regina  
Rodrigues
3. Datos del sinodal 1  
Dr.  
Luis Antonio  
Rincón  
Solís
4. Datos del sinodal 2  
Act.  
Jaime  
Vázquez  
Alamilla
5. Datos del sinodal 3  
Mat.  
Margarita Elvira  
Chávez  
Cano
6. Datos del sinodal 4  
Dr.  
Juan  
González  
Hernández
7. Datos del trabajo escrito  
El Método de Monte Carlo vía Cadenas de Markov y su uso en problemas de epidemiología  
114 p  
2008

# Agradecimientos

Agradezco a mis sinodales Dr. Luis Antonio Rincón Solís, Act. Jaime Vázquez Alamilla, Mat. Margarita Elvira Chávez Cano y Dr. Juan González Hernández por el tiempo que dedicaron a corregir este trabajo y por sus buenos consejos. Agradezco muy especialmente a mi asesora Dra. Eliane Regina Rodrigues por su dirección, orientación y coordinación en la realización de esta tesis.

Gracias a todos mis amigos, compañeros, profesores y en general a todas las personas que de alguna u otra manera participaron para que este trabajo se llevara a cabo. Agradezco también a toda mi familia por su apoyo, cariño y fé en mí.

---

# Índice general

<b>Introducción</b>	<b>1</b>
<b>1. Cadenas de Markov</b>	<b>5</b>
1.1. Definiciones y resultados básicos . . . . .	5
1.2. Clasificación de estados . . . . .	13
1.3. Distribución límite y distribución estacionaria . . . . .	24
1.4. Cadenas de Markov reversibles en el tiempo . . . . .	32
<b>2. Ejemplos de Cadenas de Markov</b>	<b>39</b>
2.1. Modelos de Wright-Fisher . . . . .	39
2.1.1. Modelo haploide simple de reproducción aleatoria sin considerar la mutación ni la selección . . . . .	39
2.1.2. Modelo haploide simple de reproducción aleatoria considerando la mu- tación . . . . .	41
2.1.3. Modelo haploide simple de reproducción aleatoria considerando la se- lección . . . . .	43
2.2. Modelo de la reproducción celular . . . . .	45
2.3. Clasificación de estados . . . . .	47
2.3.1. Modelo haploide simple sin mutación ni selección . . . . .	49
2.3.2. Modelo haploide simple con mutación . . . . .	50

2.3.3. Modelo haploide simple con selección . . . . .	54
2.3.4. Modelo de la reproducción celular . . . . .	54
<b>3. El Algoritmo Metropolis-Hastings</b>	<b>57</b>
3.1. Método Monte Carlo . . . . .	57
3.2. Métodos Monte Carlo vía Cadenas de Markov . . . . .	60
3.3. Algoritmo Metropolis-Hastings . . . . .	62
3.3.1. Algoritmo Metropolis-Hastings . . . . .	63
<b>4. Estadística Bayesiana</b>	<b>67</b>
4.1. Función de verosimilitud . . . . .	67
4.2. Distribución a priori . . . . .	71
4.3. Distribución a posteriori . . . . .	73
4.4. Estimación puntual . . . . .	80
<b>5. Aplicación del algoritmo Metropolis-Hastings al estudio del VPH</b>	<b>87</b>
5.1. Modelo de la historia natural del VPH . . . . .	88
5.2. Modelo de la transición del VPH tomando en cuenta una vacuna . . . . .	92
5.3. Modelo de la progresión del VPH al CIS o al CCI . . . . .	95
5.4. Modelo Bayesiano . . . . .	97
5.4.1. Función de verosimilitud . . . . .	98
5.4.2. Distribuciones a priori . . . . .	99
5.4.3. Distribuciones a posteriori . . . . .	99
5.4.4. Otras distribuciones a priori . . . . .	100
<b>Discusión</b>	<b>102</b>
<b>A. Conceptos básicos de probabilidad</b>	<b>105</b>
A.1. Algunas definiciones . . . . .	105
A.2. Algunos resultados . . . . .	106

<b>B. Variables aleatorias</b>	<b>109</b>
B.1. Variables aleatorias discretas . . . . .	109
B.2. Variables aleatorias continuas . . . . .	110
B.3. Resultado complementario . . . . .	112
<b>Bibliografía</b>	<b>113</b>





# Introducción

Aunque las enfermedades de transmisión sexual producidas por bacterias como sífilis y gonorrea, o las producidas por organismos unicelulares más complejos como la tricomoniasis están en retroceso o por lo menos ya existen tratamientos accesibles y efectivos en su contra, mucho más preocupantes son las enfermedades de transmisión sexual provocadas por virus, tales como el sida causado por el Virus de la Inmunodeficiencia Humana (VIH).

Los virus son muy simples ya que constan de una porción de material genético, ya sea ácido ribonucleico (ARN) o desoxirribonucleico (ADN) envuelto en una o varias proteínas. Así, éste material solo incluye instrucciones para replicarse en miles de copias, sin que pueda efectuar algún otro proceso identificado con la actividad celular normal.

Los virus se pueden transmitir de varias maneras: a través de un estornudo, de la picadura de algún insecto, de una inyección con la aguja contaminada o por contacto directo entre las mucosas de personas infectadas. Esta última forma de transmisión corresponde al Virus del Papiloma Humano (VPH), que es el responsable de la infección venérea más frecuente en el mundo.

Existen más de cien variedades o cepas de VPH. Algunas de ellas (especialmente las cepas 16, 18, 31, 33, 35, 39, 45, 51, 52, 55, 58, 59 y 68) están relacionadas con lesiones epiteliales en el cuello uterino y con el cáncer cervicouterino, aunque la mayoría de las veces no produce

ningún síntoma. El virus solo se reproduce y se transmite sexualmente. Por lo general, quien está infectado no lo sabe y casi siempre la infección termina por curarse espontáneamente.

Por otro lado, algunos de los virus con ADN, como ciertas cepas del VPH logran mantenerse en el organismo por largo tiempo. Cuando esto sucede, este material genético extraño puede hacer que las células del organismo infectado funcionen mal y hasta pueden evadir la vigilancia del sistema inmune, entonces, se desarrolla el cáncer.

Según el Instituto de Estadística, Geografía e Informática, en México el cáncer cervicouterino es el más frecuente de todos los cánceres en mujeres y afecta principalmente a las que tienen entre 25 y 64 años. A escala mundial, este cáncer es el segundo más frecuente en mujeres y en países en desarrollo es el primero más frecuente.

Dos de los factores para el desarrollo del cáncer cervicouterino son que las mujeres inician su vida sexual a temprana edad y que a mayor número de compañeros sexuales aumenta el riesgo de contraer el virus causante del cáncer. Además, si las mujeres tienen un compañero sexual que hubiera tenido numerosas parejas, también aumenta el riesgo de contraer el virus.

Uno de los métodos utilizados para determinar la presencia del cáncer cervicouterino es el frotis del Papanicolau, que consiste en obtener una muestra de las células del cuello de la matriz, que al examinarla indica la presencia de cáncer incluso en etapas tempranas o de lesiones precancerosas del cáncer cervicouterino. Por medio de cirugía se pueden eliminar las lesiones cancerosas antes de que se diseminen. Así, cuando es bien aplicado, este método permite reducir las tasas del cáncer de matriz y salvar la vida de muchas mujeres.

A pesar de que existe el método de Papanicolau, el cáncer cervicouterino es una importante causa de mortalidad, por lo que es necesario prevenirlo. Ya que las proteínas de las cepas 16 y 18 del VPH (relacionadas con el cáncer de matriz) son estables, se ha intentado desarrollar

vacunas por medio de ellas. La idea consiste en copiar la proteína viral L1. Cuando está presente en grandes cantidades ésta se autoensambla formando la cubierta exterior del virus pero sin el ADN necesario para que el virus se replique. Así, estas partículas semejantes al virus activan el sistema inmune, el cual forma anticuerpos que actúan contra el verdadero virus cuando éste entra al organismo.

Por lo anterior, una vacuna es más eficaz cuando se aplica a un gran porcentaje de la población susceptible a la enfermedad, de modo que las personas no vacunadas tengan menos riesgos de infectarse. Además, se debe tener en cuenta que hombres y mujeres pueden contraer el VPH, pero sólo las mujeres pueden desarrollar el cáncer cervicouterino. Por esto, una estrategia global de vacunación debe incluir a ambos sexos y a todas las orientaciones sexuales.

Esta tesis se propone ejemplificar el uso de las cadenas de Markov para modelar la historia natural del VPH, la transición del VPH tomando en cuenta una vacuna y la progresión del VPH al cáncer cervicouterino, así como también ejemplificar el uso del algoritmo Metropolis-Hastings para estimar los parámetros de dichos modelos.

Para cumplir con los objetivos establecidos, este trabajo se estructuró de la siguiente manera. El Capítulo 1 presenta conceptos y resultados básicos de la teoría de probabilidad y de las cadenas de Markov que serán necesarios para comprender mejor los capítulos posteriores.

El Capítulo 2 contiene algunos ejemplos de cadenas de Markov relacionados con la genética. La importancia de este capítulo radica en que muestra cómo definir las cadenas de Markov y cómo encontrar sus probabilidades y matrices de transición, lo cual se hará para los modelos de la historia natural, la transición y la progresión del VPH.

El Capítulo 3 presenta una breve introducción a los métodos Monte Carlo vía Cadenas de

Markov y describe el algoritmo Metropolis-Hastings.

El Capítulo 4 contiene conceptos básicos y ejemplos relacionados con la inferencia Bayesiana.

En el Capítulo 5 se definen las cadenas de Markov que modelan la historia natural, la transición y la progresión del VPH. Además plantea un método Bayesiano para encontrar estimadores para las probabilidades de transición de dichas cadenas de Markov. Para el caso en que la distribución a priori no sea conjugada a la función de verosimilitud, se propone un método alternativo usando el algoritmo Metropolis-Hastings.

Al final se presenta una discusión respecto a los resultados presentados en la tesis y también dos apéndices con conceptos y resultados básicos directamente relacionados con los presentados en este trabajo.

# Capítulo 1

## Cadenas de Markov

Este capítulo contiene algunos resultados y conceptos relacionados con las cadenas de Markov. En la Sección 1.1 serán presentadas algunas definiciones y resultados básicos. Las secciones posteriores presentarán conceptos como clasificación de estados de una cadena de Markov, distribuciones límite y estacionaria y reversibilidad en el tiempo. Los resultados contenidos en este capítulo podrán ser encontrados en Ash (2000), Feller (1968), Grimmett y Stirzaker (1982), Karlin y Taylor (1975) y Ross (1996,1997,1998).

### 1.1. Definiciones y resultados básicos

Se comenzará con algunas definiciones básicas y a partir de ellas se pasará a definiciones más complejas, así como a la presentación de resultados que serán utilizados a lo largo de esta tesis.

**Definición 1.1** *Un experimento es la realización de una actividad que puede ser recreada bajo las mismas condiciones.*

**Definición 1.2** *Un experimento aleatorio es un experimento donde, al reproducirlo varias veces bajo las mismas condiciones, varios resultados son posibles. Se puede decir cuales son, pero no se puede asegurar cual de ellos ocurrirá.*

**Ejemplo 1.1** El lanzamiento, utilizando las manos, de una moneda con las dos caras distintas (águila y sol) es un experimento aleatorio ya que se sabe que sus resultados posibles pueden ser águila o sol, pero no se sabe con seguridad cual de los dos saldrá en cada lanzamiento. En cambio, el lanzamiento de una moneda que por los dos lados tenga sol (o águila) es un experimento determinista porque aunque se repita varias veces siempre se obtendrá el mismo resultado, es decir, su resultado ya está determinado.

**Definición 1.3** *Un espacio muestral es el conjunto de todos los posibles resultados de un experimento aleatorio. Se indicará un espacio muestral por  $\Omega$ . Un subconjunto del espacio muestral de un experimento aleatorio es llamado **evento**.*

**Ejemplo 1.2** Considerando el experimento aleatorio del Ejemplo 1.1, su espacio muestral es el conjunto  $\{\mathbf{a}, \mathbf{s}\}$  donde  $\mathbf{a}$  representa “águila”,  $\mathbf{s}$  representa “sol”, los cuales son todos los posibles resultados al lanzar una moneda.

**Ejemplo 1.3** El experimento aleatorio de lanzar tres monedas de una sola vez y anotar el resultado de cada una de ellas tiene el siguiente espacio muestral:  $\Omega = \{(a, a, a), (s, a, a), (a, s, a), (a, a, s), (s, s, a), (s, a, s), (a, s, s), (s, s, s)\}$ . En este caso los elementos de este conjunto son “ternas” porque cada entrada de ellas es el resultado de cada una de las monedas. Note que las siguientes características describen eventos relativos a este experimento:

1. En el primer lanzamiento se obtiene águila.
2. En el primer y tercer lanzamiento se obtienen resultados distintos.
3. No hay dos resultados consecutivos iguales.

Los elementos de  $\Omega$  que cumplen con cada una de estas descripciones son respectivamente  $A_1 = \{(a, a, a), (a, s, a), (a, a, s), (a, s, s)\}$ ,  $A_2 = \{(s, a, a), (a, a, s), (s, s, a), (a, s, s)\}$  y  $A_3 = \{(a, s, a), (s, a, s)\}$ .

**Definición 1.4** *Sea  $\mathcal{F}$  una colección de subconjuntos de un conjunto  $\Omega$ . A  $\mathcal{F}$  se le llama  $\sigma$ -álgebra si y solo si:*

a)  $\Omega \in \mathcal{F}$ .

b) Si  $A \in \mathcal{F}$ , entonces  $A^c \in \mathcal{F}$ .

c) Si  $A_1, A_2, \dots$  forman una colección numerable de subconjuntos de  $\mathcal{F}$ , entonces

$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}.$$

**Definición 1.5** Sea  $\mathcal{F}$  una  $\sigma$ -álgebra de subconjuntos de un conjunto  $\Omega$ . Una **medida**  $\mu$  sobre  $\Omega$  es una función  $\mu : \mathcal{F} \rightarrow [0, \infty]$  tal que:

a)  $\mu(\emptyset) = 0$ .

b)  $\mu(A) \geq 0$  para toda  $A \in \mathcal{F}$ .

c) Si  $A_1, A_2, \dots$  forman una colección finita o numerable de subconjuntos disjuntos de  $\mathcal{F}$ , entonces  $\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$ .

Si se tiene que  $\mu(\Omega) = 1$ , entonces  $\mu$  es llamada **medida de probabilidad** y se representa por  $P : \mathcal{F} \rightarrow [0, 1]$  o solamente por  $P$ .

**Definición 1.6** Un **espacio medible** es una terna  $(\Omega, \mathcal{F}, \mu)$  donde  $\Omega$  es un conjunto,  $\mathcal{F}$  es un  $\sigma$ -álgebra de subconjuntos de  $\Omega$  y  $\mu$  es una medida sobre  $\mathcal{F}$ . Si  $\mu = P$  es una medida de probabilidad, entonces  $(\Omega, \mathcal{F}, P)$  es llamado **espacio de probabilidad**.

**Observación. CÁLCULO EMPÍRICO DE UNA MEDIDA DE PROBABILIDAD.**

Suponiendo que todos los elementos del espacio muestral  $\Omega$  de un experimento aleatorio son igualmente probables, entonces se puede calcular empíricamente la probabilidad de un evento  $E$  relativo a este experimento. Una forma de hacer esto es usando la siguiente expresión:

$$P(E) = \frac{|E|}{|\Omega|} = \frac{\text{número de casos favorables relativos al evento}}{\text{número de casos posibles en el experimento}}.$$

**Ejemplo 1.4** Considere el Ejemplo 1.3. Suponga que todos los elementos de  $\Omega$  son igualmente probables. La probabilidad del evento  $A_1$  es  $\frac{4}{8} = \frac{1}{2}$ ; la del evento  $A_2$  es también  $\frac{4}{8} = \frac{1}{2}$  y la probabilidad del evento  $A_3$  es  $\frac{2}{8} = \frac{1}{4}$ .



**Definición 1.7** Sean  $(\Omega, \mathcal{F}, \mu)$  un espacio medible y  $h$  una función que lleva un elemento de  $\Omega$  a un elemento en  $\mathbb{R}$ . Se dice que  $h$  es una **función medible** con respecto a  $\mathcal{F}$  si  $h^{-1}(B) = \{\omega : h(\omega) \in B\} \in \mathcal{F}$  con  $B \in \mathfrak{B}(\mathbb{R})$  donde  $\mathfrak{B}(\mathbb{R})$  es la  $\sigma$ -álgebra de Borel<sup>1</sup> de  $\mathbb{R}$ .

**Definición 1.8** Una **variable aleatoria**  $X$  en un espacio de probabilidad  $(\Omega, \mathcal{F}, P)$  es una función medible con respecto a  $\mathcal{F}$  que asocia elementos de  $(\Omega, \mathcal{F}, P)$  a elementos de  $(\mathbb{R}, \mathfrak{B}(\mathbb{R}), P_X)$ , donde  $P_X$  es la función de probabilidad definida en  $\mathfrak{B}(\mathbb{R})$  inducida por la variable  $X$ , es decir, para  $x \in B \subset \mathfrak{B}(\mathbb{R})$

$$P_X(X = x) = P(\{\omega \in \Omega : X(\omega) = x\}).$$

**Ejemplo 1.5** Considere el espacio muestral del experimento descrito en el Ejemplo 1.1 y la siguiente variable aleatoria:

$$X_1(w) = \begin{cases} 1 & \text{si } w = a \\ 0 & \text{si } w = s. \end{cases}$$

$X_1$  es una variable aleatoria relativa al experimento del Ejemplo 1.1 y  $X_2 = \text{número de águilas en los dos últimos lanzamientos}$  es una variable aleatoria para el experimento descrito en el Ejemplo 1.3 porque a cada elemento del espacio muestral le asigna un valor real: a  $(a, a, a)$  le asigna 2 porque esta terna tiene 2 águilas en sus últimas dos entradas, a  $(a, s, a)$  le asigna 1 porque solo tiene un águila en sus últimas dos entradas, a  $(a, s, s)$  le asigna 0 porque no tiene águilas en sus dos últimas entradas, y así sucesivamente.

**Definición 1.9** Un **proceso estocástico**  $X = \{X_t : t \in T\}$  es una colección de variables aleatorias que están definidas sobre el mismo espacio muestral, están indexadas por el conjunto  $T$  y asumen valores en un mismo conjunto  $S \subseteq \mathbb{R}$ . El conjunto  $T$  es llamado **conjunto de índices** de  $X$  y la mayoría de las veces representa al tiempo. El conjunto  $S$  es llamado **espacio de estados** de  $X$ . Si  $T$  es un subconjunto de  $\mathbb{Z}$  entonces  $X$  es llamado **proceso estocástico de tiempo discreto** o simplemente **proceso estocástico discreto** y si  $T$  es un subintervalo de  $\mathbb{R}$  entonces  $X$  se llama **proceso estocástico de tiempo continuo**, o simplemente **proceso estocástico continuo**.

---

<sup>1</sup>Ver Apéndice A, Definición A.1.

**Observación.** Un proceso estocástico, en general, es utilizado para modelar un fenómeno aleatorio a lo largo del tiempo.

**Definición 1.10** Sea  $X = \{X_t : t \in T\}$  un proceso estocástico con espacio de estados  $S$  y donde  $T \subseteq \mathbb{Z}$  o  $T \subseteq \mathbb{R}$ . El proceso  $X$  es llamado un **proceso de Markov** si para  $t_0, t_1, \dots, t_{n-1}, t, s \in T$ ,  $x_0, x_1, \dots, x_{n-1}, x, y \in S$  con  $t_0 < t_1 < \dots < t_{n-1} < t$  lo siguiente es válido:

$$P\{X_{t+s} = y \mid X_{t_0} = x_0, X_{t_1} = x_1, \dots, X_{t_{n-1}} = x_{n-1}, X_t = x\} = P\{X_{t+s} = y \mid X_t = x\}.$$

**Observación.** La propiedad dada en la Definición 1.10, llamada “propiedad de Markov”, significa que la distribución condicional de cualquier estado futuro de  $X$  dado los estados pasados y el estado presente, depende únicamente de éste último y no de los estados pasados.

**Definición 1.11** Sea  $X = \{X_t : t \in T\}$  un proceso de Markov con espacio de estados  $S$ . Si  $S \subseteq \mathbb{Z}$ , entonces  $X$  es llamado **cadena de Markov**. Si  $T \subseteq \mathbb{Z}$ , entonces  $X$  es llamado **cadena de Markov discreta** y si  $T \subseteq \mathbb{R}$ , entonces  $X$  es llamado **cadena de Markov continua**.

**Definición 1.12** Sea  $X = \{X_n : n = 0, 1, \dots\}$  una cadena de Markov discreta con espacio de estados  $S$ . La **probabilidad de transición en  $m$  pasos** de  $X$  se define por

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

con  $i, j \in S$ ,  $n, m \geq 0$ . Cuando  $m = 1$  se tiene que  $P_{i,j}^{(n,n+1)} = P\{X_{n+1} = j \mid X_n = i\}$  y en este caso se llamará **probabilidad de transición** de  $X$ . Si  $P_{i,j}^{(n,n+m)}$  no depende del tiempo  $n$  el proceso es llamado **homogéneo en el tiempo** y se indica  $P_{i,j}^{(n,n+m)} = P_{i,j}^{(m)}$  y para  $m = 1$  se indica  $P_{i,j}^{(n,n+1)} = P_{i,j}^{(1)} = P_{i,j}$ . Además se tiene que

$$P_{i,j}^{(n,n)} = \begin{cases} 1 & \text{si } i = j, \\ 0 & \text{si } i \neq j. \end{cases}$$

**Observación.** La probabilidad de transición en  $m$  pasos de una cadena de Markov  $X$  es la probabilidad de que, estando  $X$  en el estado  $i$  al tiempo  $n$ , esté en el estado  $j$  al tiempo  $n+m$ .

De ahora en adelante se trabajará con cadenas de Markov discretas y homogéneas en el tiempo.

**Definición 1.13** Sea  $X = \{X_n : n = 0, 1, \dots\}$  una cadena de Markov con espacio de estados  $S$  y probabilidad de transición en  $m$  pasos  $P_{i,j}^{(m)}$ ,  $i, j \in S$ ,  $m \geq 0$ . La matriz  $P^{(m)} = \left( P_{i,j}^{(m)} \right)_{i,j \in S}$  es llamada **matriz de probabilidades de transición en  $m$  pasos de  $X$** . Cuando  $m = 1$  se dirá que  $P^{(1)} = P$  y se le llamará **matriz de transición de  $X$** .

**Observación.** Si  $|S| = n$  con  $n \in \{1, 2, \dots\}$  la matriz de transición  $P$  tiene la siguiente forma:

$$P = \begin{pmatrix} P_{1,1} & P_{1,2} & \dots & P_{1,n} \\ \vdots & \vdots & \vdots & \vdots \\ P_{n,1} & P_{n,2} & \dots & P_{n,n} \end{pmatrix}.$$

**Propiedad 1.1** Sea  $P^{(m)} = \left( P_{i,j}^{(m)} \right)_{i,j \in S}$  la matriz de transición en  $m$  pasos de una cadena de Markov  $X = \{X_n : n = 0, 1, \dots\}$  con espacio de estados  $S$ . Entonces para toda  $m \geq 0$ :

- a)  $P^{(m)}$  tiene entradas no negativas, es decir,  $P_{i,j}^{(m)} \geq 0$ .
- b)  $P^{(m)}$  es una matriz estocástica, es decir,  $\sum_{j \in S} P_{i,j}^{(m)} = 1$  para toda  $i \in S$ .

**Demostración.**

a) Por la Definición 1.12 se tiene que  $P_{i,j}^{(m)} = P\{X_{n+m} = j \mid X_n = i\}$ , y por definición de probabilidad condicional se sabe que esta última expresión es mayor o igual a 0 <sup>2</sup>.

b) Sea  $i \in S$  fijo. Entonces,

$$\begin{aligned} \sum_{j \in S} P_{i,j}^{(m)} &= \sum_{j \in S} P\{X_{n+m} = j \mid X_n = i\} \\ &= P \left\{ X_{n+m} \in \bigcup_{j \in S} \{j\} \mid X_n = i \right\} \\ &= P\{X_{n+m} \in S \mid X_n = i\} \\ &= 1. \end{aligned}$$

Lo mismo se tiene para toda  $i \in S$ . La primera igualdad se tiene por definición de probabilidad de transición en  $m$  pasos y la tercera igualdad vale dado que  $S = \bigcup_{j \in S} \{j\}$ .

**Teorema 1.1 (Ecuaciones de Chapman-Kolmogorov)**

Sea  $P^{(m)} = \left( P_{i,j}^{(m)} \right)_{i,j \in S}$  la matriz de transición en  $m$  pasos de una cadena de Markov  $X = \{X_n : n = 0, 1, \dots\}$  con espacio de estados  $S$ . Sean  $m, r \geq 0$  y tales que  $0 \leq r \leq m$ , entonces para toda  $i, j \in S$

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

---

<sup>2</sup>Ver Apéndice A, Definición A.4.

**Demostración.** Note que

$$\begin{aligned}
 P_{i,j}^{(m)} &= P\{X_{n+m} = j | X_n = i\} \\
 &= \frac{P\{X_{n+m} = j, X_n = i\}}{P\{X_n = i\}} \\
 &= \frac{\sum_{k \in S} P\{X_{n+m} = j, X_{n+r} = k, X_n = i\}}{P\{X_n = i\}} \\
 &= \sum_{k \in S} \frac{P\{X_{n+m} = j, X_{n+r} = k, X_n = i\}}{P\{X_n = i\}} \frac{P\{X_{n+r} = k, X_n = i\}}{P\{X_{n+r} = k, X_n = i\}} \\
 &= \sum_{k \in S} P\{X_{n+r} = k | X_n = i\} P\{X_{n+m} = j | X_{n+r} = k, X_n = i\} \\
 &= \sum_{k \in S} P\{X_{n+r} = k | X_n = i\} P\{X_{n+m} = j | X_{n+r} = k\} \\
 &= \sum_{k \in S} P_{i,k}^{(r)} P_{k,j}^{(m-r)}.
 \end{aligned}$$

La primera y última igualdad se tienen por la definición de probabilidad de transición en  $m$ ,  $r$  y  $m - r$  pasos, la sexta igualdad se tiene por la propiedad de Markov.

**Observación.** Note que para un espacio de estados finito se tiene que  $P^{(n+m)} = P^{(n)} \cdot P^{(m)}$  donde  $P^{(n+m)}$ ,  $P^{(n)}$  y  $P^{(m)}$  son matrices de transición en  $n + m$ ,  $n$  y  $m$  pasos respectivamente de una cadena de Markov. Para ver que esta afirmación es verdadera note que

$$\begin{aligned}
 P_{i,j}^{(n+m)} &= \sum_{k \in S} P_{i,k}^{(n)} P_{k,j}^{(m)} \\
 &= P^{(n)} \cdot P^{(m)}.
 \end{aligned}$$

La primera igualdad se tiene por las ecuaciones de Chapman-Kolmogorov y la segunda igualdad se tiene por la definición del producto entre dos matrices. Por lo tanto, conociendo  $P$  se puede calcular  $P^{(n)}$  dado que

$$P^{(n)} = \underbrace{P \cdot P \cdot P \cdots P}_{n \text{ factores}} = P^n.$$

## 1.2. Clasificación de estados

Esta sección proveerá al lector de la información necesaria para encontrar y clasificar los estados de una cadena de Markov por medio de algunas definiciones.

**Definición 1.14** Sean  $X = \{X_n : n = 0, 1, \dots\}$  una cadena de Markov,  $S$  su espacio de estados y  $P^{(n)} = (P_{i,j}^{(n)})_{i,j \in S}$  su matriz de transición en  $n \geq 0$  pasos. Tome  $i, j \in S$ . Se dice que el estado  $j$  es **accesible** desde el estado  $i$  si existe un entero  $m \geq 0$  tal que  $P_{i,j}^{(m)} > 0$  y se representa como  $\mathbf{i} \rightarrow \mathbf{j}$ . Se dice que  $i$  y  $j$  se **comunican** ( $\mathbf{i} \leftrightarrow \mathbf{j}$ ) si y sólo si  $i \rightarrow j$  y  $j \rightarrow i$ .

**Propiedad 1.2** El concepto de comunicación es una relación de equivalencia, es decir, para  $i, j, k \in S$  se cumple que:

1.  $i \leftrightarrow i$  (reflexividad).
2. Si  $i \leftrightarrow j$  entonces  $j \leftrightarrow i$  (simetría).
3. Si  $i \leftrightarrow j$ ,  $j \leftrightarrow k$  entonces  $i \leftrightarrow k$  (transitividad).

**Demostración.**

1. Por definición se sabe que

$$P_{i,j}^{(0)} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j. \end{cases}$$

Por lo tanto, existe un entero (que es 0) tal que  $P_{i,i}^{(0)} > 0$ . Entonces  $i \rightarrow i$  y por lo tanto  $i \leftrightarrow i$ .

2. Por hipótesis se sabe que  $i \leftrightarrow j$  y por la Definición 1.14 se tiene que esto pasa si y sólo si  $i \rightarrow j$  y  $j \rightarrow i$ . Por lo tanto se concluye que  $j \leftrightarrow i$ .
3. Por hipótesis se tiene que  $i \leftrightarrow j$  y  $j \leftrightarrow k$ . Por tanto, existen  $n, m \geq 0$  tales que  $P_{i,j}^{(n)} > 0$  y  $P_{j,k}^{(m)} > 0$ , y entonces  $P_{i,j}^{(n)} P_{j,k}^{(m)} > 0$ . Por otro lado se tiene que

$$P_{i,k}^{(n+m)} = \sum_{h \in S} P_{i,h}^{(n)} P_{h,k}^{(m)} \geq P_{i,j}^{(n)} P_{j,k}^{(m)} > 0.$$

## 1.2. Clasificación de estados

---

La primer igualdad de la expresión anterior se tiene por el Teorema 1.1 y la siguiente desigualdad se obtiene dado que todos los sumandos de la sumatoria son no negativos.

Haciendo  $r = n + m$  se tiene que existe  $r \geq 0$  tal que  $P_{i,k}^{(r)} > 0$ , por lo que  $i \rightarrow k$ . De forma análoga se puede demostrar que si  $k \rightarrow j$  y  $j \rightarrow i$ , entonces  $k \rightarrow i$ . Esto concluye la demostración.

**Nota.** Ya que el concepto de comunicación es una relación de equivalencia, se puede formar una partición en  $S$  donde los estados que forman una clase de equivalencia son los que se comunican entre sí.

**Definición 1.15** Se dice que una cadena de Markov es una **cadena irreducible** si tiene una única clase de equivalencia, es decir, si todos sus estados se comunican entre sí.

**Definición 1.16** Sean  $X = \{X_n : n = 0, 1, \dots\}$  una cadena de Markov,  $S$  su espacio de estados y  $P^{(n)} = \left(P_{i,j}^{(n)}\right)_{i,j \in S}$  su matriz de transición en  $n$  pasos. Tome  $i \in S$ . El **periodo** de  $i$  está definido por

$$d(i) = m.c.d.\{n \geq 1 \mid P_{i,i}^{(n)} > 0\}.$$

Si  $P_{i,i}^{(n)} = 0$  para toda  $n \geq 1$  entonces  $d(i) = 0$ . Si  $d(i) = 1$  entonces se dice que  $i$  es **aperiódico**.

**Definición 1.17** Se dice que una cadena de Markov es **aperiódica** si todos sus estados son aperiódicos.

**Propiedad 1.3** La periodicidad es una propiedad de clase, es decir, para  $i, j \in S$ ,  $S$  el espacio de estados de una cadena de Markov, si  $i \leftrightarrow j$ , entonces  $d(i) = d(j)$ .

**Demostración.**

Sea  $X = \{X_n : n = 0, 1, \dots\}$  una cadena de Markov con espacio de estados  $S$  y matriz de transición en  $m$  pasos  $P^{(m)} = \left(P_{i,j}^{(m)}\right)_{i,j \in S}$ . Por hipótesis tenemos que existen  $n_1, n_2 \geq 0$

tales que  $P_{i,j}^{(n_1)} > 0$  y  $P_{j,i}^{(n_2)} > 0$ . Usando las ecuaciones de Chapman-Kolmogorov (Teorema 1.1) se tiene que

$$P_{j,j}^{(n_1+n_2)} = \sum_{k \in S} P_{j,k}^{(n_2)} P_{k,j}^{(n_1)} \geq P_{j,i}^{(n_2)} P_{i,j}^{(n_1)} > 0.$$

La primera desigualdad es porque los elementos de la suma son todos positivos y la segunda desigualdad es por la hipótesis de que los dos términos son estrictamente mayores que cero, y por tanto, su producto también lo es.

Por otro lado, sea  $r \geq 1$  tal que  $P_{i,i}^{(r)} > 0$ , por lo que

$$P_{j,j}^{(n_2+r+n_1)} = \sum_{k \in S} P_{j,k}^{(n_2+r)} P_{k,j}^{(n_1)} = \sum_{k \in S} \sum_{h \in S} P_{j,h}^{(n_2)} P_{h,k}^{(r)} P_{k,j}^{(n_1)} \geq P_{j,i}^{(n_2)} P_{i,i}^{(r)} P_{i,j}^{(n_1)} > 0.$$

La primera igualdad es por el Teorema 1.1 aplicado a  $P_{j,j}^{(n_2+r+n_1)}$ , la segunda igualdad es por el mismo teorema pero aplicado a  $P_{j,k}^{(n_2+r)}$ . Las dos desigualdades siguientes son por el mismo argumento usado para demostrar que  $P_{j,j}^{(n_1+n_2)} > 0$ .

Como  $d(j) = m.c.d.\{n \geq 1 \mid P_{j,j}^{(n)} > 0\}$  entonces  $d(j)$  divide a  $n_1 + n_2$  y también divide a  $n_1 + r + n_2$ , por tanto,  $(n_1 + r + n_2) - (n_1 + n_2) = r$  también es dividido por  $d(j)$ . De esta manera  $d(j)$  divide a cualquier entero que sea elemento del siguiente conjunto:

$$D = \{s \geq 1 \mid P_{i,i}^{(s)} > 0\}.$$

Dado que  $d(i) \in D$  se tiene que  $d(j)$  divide a  $d(i)$ . Del mismo modo se puede ver que  $d(i)$  divide a  $d(j)$ . Por tanto  $d(j) = d(i)$ .

**Definición 1.18** Sea  $X = \{X_n : n = 0, 1, \dots\}$  una cadena de Markov con espacio de estados  $S$  y matriz de transición  $P = (P_{i,j})_{i,j \in S}$ . Tome  $i, j \in S$ . Se indicará por  $f_{i,j}^{(n)}$  a la probabilidad de que, partiendo del estado  $i$ , la **primer visita** al estado  $j$  sea en  $n$  pasos,  $n \geq 1$ , y se define por:

$$f_{i,j}^{(n)} = P\{X_{m+n} = j, X_{m+s} \neq j, s = 1, \dots, n-1 \mid X_m = i\}.$$



## 1.2. Clasificación de estados

---

En particular,  $f_{i,i}^{(n)}$  es la probabilidad de que, partiendo del estado  $i$ , el **primer retorno** a  $i$  es en  $n$  pasos. Para  $n = 0$  se define  $f_{i,i}^{(0)} = 0$  y para  $n = 1$  se tiene que  $f_{i,j}^{(1)} = P\{X_{m+1} = j | X_m = i\} = P_{i,j}$  donde  $i$  podría ser igual a  $j$ .

**Lema 1.1** Sea  $X = \{X_n : n = 0, 1, \dots\}$  una cadena de Markov con espacio de estados  $S$ . Para  $i, j \in S$  y  $n \geq 1$  lo siguiente es válido:

$$P_{i,j}^{(n)} = \sum_{k=0}^n f_{i,j}^{(k)} P_{j,j}^{(n-k)}.$$

### **Demostración.**

Esta igualdad se deriva de que el evento

{la cadena parte del estado  $i$  y debe llegar al estado  $j$  en  $n$  pasos}

es lo mismo que considerar todas las posibles realizaciones del siguiente evento:

$E_k =$  {la cadena parte del estado  $i$ , debe terminar en el estado  $j$  en  $n$  pasos  
y la primera llegada a  $j$  ocurre en  $k$  pasos}

(con  $k \leq n$ ). Por tanto,

$$\{X_{m+n} = j | X_m = i\} = \bigcup_{1 \leq k \leq n} E_k. \quad (1.1)$$

Además, para  $1 \leq k \leq n$  vale

$$\begin{aligned} P(E_k) &= P\{X_{m+k} = j, X_{m+s} \neq j, s = 1, \dots, k-1 | X_m = i\} \\ &= P\{X_{m+n} = j | X_{m+k} = j\} \\ &= f_{i,j}^{(k)} P_{j,j}^{(n-k)}. \end{aligned}$$

Así,

$$\begin{aligned}
P_{i,j}^{(n)} &= P\{X_{m+n} = j | X_m = i\} \\
&= P\left\{ \bigcup_{1 \leq k \leq n} E_k \right\} \\
&= \sum_{k=1}^n P[E_k] \\
&= \sum_{k=1}^n f_{i,j}^{(k)} P_{j,j}^{(n-k)} \\
&= f_{i,j}^{(0)} P_{j,j}^{(n)} + \sum_{k=1}^n f_{i,j}^{(k)} P_{j,j}^{(n-k)} \\
&= \sum_{k=0}^n f_{i,j}^{(k)} P_{j,j}^{(n-k)}.
\end{aligned}$$

La primer igualdad es por definición de probabilidad de transición en  $n$  pasos, la segunda igualdad es por la fórmula (1.1), la tercera igualdad es por definición de la probabilidad de la unión de eventos mutuamente excluyentes (los eventos  $E_k$  son mutuamente excluyentes porque estando en  $i$  el proceso puede ir por primera vez a  $j$  en  $k$  pasos donde  $k$  puede tomar *solo un* valor, de 1 hasta  $n$ , y después debe regresar a  $j$  en  $n - k$  pasos). La cuarta igualdad es solo la sustitución de  $P(E_k)$  y la quinta igualdad es porque  $f_{i,j}^{(0)} = 0$ .

**Lema 1.2** Sean  $X = \{X_n : n = 0, 1, \dots\}$  una cadena de Markov con espacio de estados  $S$  y  $i, j \in S$ . Entonces para  $n \geq 1$

$$f_{i,j}^{(n)} = P_{i,j}^{(n)} - \left[ \sum_{k=0}^{n-1} f_{i,j}^{(k)} P_{j,j}^{(n-k)} \right].$$

**Demostración.**

$$\begin{aligned}
 P_{i,j}^{(n)} &= \sum_{k=0}^n f_{i,j}^{(k)} P_{j,j}^{(n-k)} \\
 &= \sum_{k=0}^{n-1} f_{i,j}^{(k)} P_{j,j}^{(n-k)} + f_{i,j}^{(n)} P_{j,j}^{(0)} \\
 &= \sum_{k=0}^{n-1} f_{i,j}^{(k)} P_{j,j}^{(n-k)} + f_{i,j}^{(n)}.
 \end{aligned}$$

La primera igualdad se tiene por el Lema 1.1 y la última igualdad se tiene porque por definición  $P_{j,j}^{(0)} = 1$ . Por tanto el resultado se sigue.

**Definición 1.19** Sean  $X = \{X_n : n = 0, 1, \dots\}$  una cadena de Markov,  $S$  su espacio de estados y  $i, j \in S$ . La probabilidad de **alcanzar alguna vez al estado  $j$  partiendo del estado  $i$**  es

$$f_{i,j} = \sum_{n=1}^{\infty} f_{i,j}^{(n)}.$$

En particular,  $f_{i,i} = \sum_{n=1}^{\infty} f_{i,i}^{(n)}$  es la probabilidad de **regresar alguna vez al estado  $i$** .

**Lema 1.3** Sea  $X = \{X_n : n = 0, 1, \dots\}$  una cadena de Markov con espacio de estados  $S$  y sean  $i, j \in S$ . Entonces,  $f_{i,j} > 0$  si y sólo si  $i \rightarrow j$ .

**Demostración.**

(i) Por hipótesis se tiene que  $f_{i,j} > 0$ , es decir,

$$\sum_{n=1}^{\infty} f_{i,j}^{(n)} > 0. \tag{1.2}$$

También se sabe que  $f_{i,j}^{(n)} \geq 0$  con  $n = 1, 2, \dots$  porque es una probabilidad condicional. Así podemos decir que por lo menos un elemento de la suma (1.2) es mayor a cero, es decir, existe un entero, digamos  $n_1 \geq 1$  tal que  $f_{i,j}^{(n_1)} > 0$ .

Por el Lema 1.1 se tiene que

$$P_{i,j}^{(n_1)} = \sum_{k=0}^{n_1} f_{i,j}^{(k)} P_{j,j}^{(n_1-k)}.$$

Esto implica que  $P_{i,j}^{(n_1)} > 0$  ya que cuando  $k = n_1$  se tiene que el producto  $f_{i,j}^{(n_1)} P_{j,j}^{(n_1-n_1)}$  es estrictamente mayor que cero dado que  $P_{j,j}^{(0)} = 1$  y  $f_{i,j}^{(n_1)} > 0$ .

Por lo tanto se tiene que existe un entero  $n_1 \geq 0$  tal que  $P_{i,j}^{(n_1)} > 0$ , es decir,  $i \rightarrow j$ .

(ii) Por hipótesis se tiene que existe un entero  $m \geq 1$  tal que  $P_{i,j}^{(m)} > 0$  ya que  $i \rightarrow j$ .

Adicionalmente, por el Lema 1.1 se tiene que  $P_{i,j}^{(m)} = \sum_{k=0}^m f_{i,j}^{(k)} P_{j,j}^{(m-k)} > 0$ .

Tomando en cuenta que  $f_{i,j}^{(k)} \geq 0$  y  $P_{j,j}^{(m-k)} \geq 0$  con  $k = 0, \dots, m$  por ser unas probabilidades condicionales, se puede concluir que existe un entero, digamos  $m_1$ , tal que  $1 \leq m_1 \leq m$ ,  $f_{i,j}^{(m_1)} > 0$  y  $P_{j,j}^{(m-m_1)} > 0$ .

Así, por lo menos un sumando de  $f_{i,j} = \sum_{n=1}^{\infty} f_{i,j}^{(n)}$  es mayor a cero, por lo que  $f_{i,j}$  también es mayor a cero.

**Definición 1.20** Sea  $X = \{X_n : n = 0, 1, \dots\}$  una cadena de Markov con espacio de estados  $S$  y sea  $i \in S$ .

a) Se dice que  $i$  es **recurrente** si y sólo si  $f_{i,i} = 1$ .

b) Se dice que  $i$  es **transitorio** si y sólo si  $f_{i,i} < 1$ .

**Observación.** La Definición 1.20 se puede interpretar como sigue:

a)  $i$  es recurrente si es seguro regresar a él en un periodo finito de tiempo.

b)  $i$  es transitorio si hay una probabilidad positiva de no regresar a él.

**Proposición 1.1** *Al menos un estado de una cadena de Markov con espacio de estados finito es recurrente.*

**Demostración.**

Esta proposición se demostrará por contradicción. De esta forma, suponga que la hipótesis no es cierta, es decir que todos los estados son transitorios. Sea  $X = \{X_n : n = 0, 1, \dots\}$  una cadena de Markov y  $S = \{0, 1, \dots, N\}$  su espacio de estados donde todos son transitorios. Así, después de algún tiempo finito, digamos  $t_k$ , la cadena ya no regresará al estado  $k$  para toda  $k = 0, 1, \dots, N$  con probabilidad positiva. Entonces después del tiempo finito  $t = \max\{t_0, t_1, \dots, t_N\}$  la cadena no regresará a ningún estado con probabilidad positiva, lo cual es una contradicción, ya que ésta debe estar en algún estado después de este tiempo. Por lo tanto, al menos un estado de la cadena es recurrente.

**Teorema 1.2** *Sea  $X = \{X_n : n = 0, 1, \dots\}$  una cadena de Markov con espacio de estados  $S$  y matriz de transición en  $n$  pasos  $P^{(n)} = (P_{i,j}^{(n)})_{i,j \in S}$ . Tome  $i \in S$ .*

1. *El estado  $i$  es recurrente si y sólo si  $\sum_{n=1}^{\infty} P_{i,i}^{(n)} = \infty$ .*
2. *El estado  $i$  es transitorio si y sólo si  $\sum_{n=1}^{\infty} P_{i,i}^{(n)} < \infty$ .*

**Demostración.**

Considere la variable aleatoria  $N_i$  que cuenta el número de veces que la cadena  $X$  regresa al estado  $i$ . También considere la siguiente función indicadora:

$$I\{X_n = i\} = \begin{cases} 1 & \text{si } X_n = i \\ 0 & \text{si } X_n \neq i. \end{cases}$$

Note que  $N_i$  se puede expresar de la siguiente manera:

$$N_i = \sum_{n=0}^{\infty} I\{X_n = i\}.$$

El número esperado de veces que  $X$  regresa a  $i$  suponiendo que comienza de nuevo en éste estado es:

$$\begin{aligned}
 E[N_i|X_0 = i] &= E \left[ \sum_{n=0}^{\infty} I\{X_n = i\} | X_0 = i \right] \\
 &= \sum_{n=0}^{\infty} E[I\{X_n = i\} | X_0 = i] \\
 &= \sum_{n=0}^{\infty} \{1 \cdot P[X_n = i | X_0 = i] + 0 \cdot P[X_n \neq i | X_0 = i]\} \\
 &= \sum_{n=0}^{\infty} P[X_n = i | X_0 = i] \\
 &= \sum_{n=0}^{\infty} P_{i,i}^{(n)}. \tag{1.3}
 \end{aligned}$$

La primera igualdad es por definición de  $N_i$ , la segunda igualdad es una propiedad de la esperanza condicional, la tercera igualdad es por definición de esperanza condicional <sup>3</sup> y la última igualdad es por la definición de probabilidad de transición en  $n$  pasos.

1. (i) Suponga que  $i \in S$  es recurrente, es decir, la cadena regresará a él en un número finito de transiciones. Considerando que la cadena comienza en este estado (es posible hacerlo ya que  $X$  es una cadena de Markov, por lo que no importa que valores haya tomado antes de su primer visita a  $i$ ), entonces nuevamente regresará a este estado y así, regresará un número infinito de veces. De esta forma  $N_i$  tiene valor infinito y por lo tanto

$$E[N_i|X_0 = i] = \infty.$$

Consecuentemente, por la fórmula (1.3)

$$\sum_{n=0}^{\infty} P_{i,i}^{(n)} = \infty.$$

2. (i) Suponga que  $i \in S$  es transitorio, entonces por la Definición (1.20) se tiene que  $f_{i,i} < 1$ . También suponga que  $N_i = k$ . De esta forma

$$P[N_i = k | X_0 = i] = (f_{i,i})^k (1 - f_{i,i}) \tag{1.4}$$

---

<sup>3</sup>Ver Apéndice A, Definición A.5.

## 1.2. Clasificación de estados

---

para  $k = 0, 1, \dots$  ya que  $f_{i,i}$  es la probabilidad con la que  $X$  regresa cada vez a  $i$ , y  $1 - f_{i,i}$  es la probabilidad con la que no vuelve nunca más después de haber visitado a  $i$  las  $k$  veces.

Note que la fórmula (1.4) es una distribución Geométrica con parámetro  $1 - f_{i,i}$  y por tanto su valor esperado es  $\frac{1}{1-f_{i,i}}$ , es decir,

$$E[N_i | X_0 = i] = \frac{1}{1 - f_{i,i}}.$$

Finalmente, tomando en cuenta la fórmula (1.3) se concluye que

$$\sum_{n=0}^{\infty} P_{i,i}^{(n)} = \frac{1}{1 - f_{i,i}} < \infty.$$

Para demostrar los inversos del teorema se procederá por contradicción.

1. (ii) Por hipótesis se sabe que  $\sum_{n=0}^{\infty} P_{i,i}^{(n)} = \infty$ . Suponga que  $i \in S$  es transitorio. Por
2. (i) se tiene que

$$\sum_{n=0}^{\infty} P_{i,i}^{(n)} < \infty,$$

lo cual contradice a la hipótesis. Por lo tanto,  $i$  no puede ser transitorio, es recurrente.

2. (ii) Se tiene por un argumento similar al dado en 1. (ii).

**Proposición 1.2** Sea  $X = \{X_n : n = 0, 1, \dots\}$  una cadena de Markov con espacio de estados  $S$  y matriz de transición en  $n$  pasos  $P^{(n)} = (P_{i,j}^{(n)})_{i,j \in S}$ . Sean  $i, j \in S$ . Si  $i$  es recurrente y  $i \leftrightarrow j$  entonces  $j$  es recurrente.

### Demostración.

Por hipótesis se tiene que  $i \in S$  es recurrente, es decir  $\sum_{r=0}^{\infty} P_{i,i}^{(r)} = \infty$  y que existen  $n_1, n_2 \geq 0$  tales que  $P_{i,j}^{(n_1)} > 0$  y  $P_{j,i}^{(n_2)} > 0$ . Por un lado se tiene que para  $r \geq 0$

$$P_{j,j}^{(n_2+r+n_1)} = \sum_{k \in S} P_{j,k}^{(n_2+r)} P_{k,j}^{(n_1)} = \sum_{k \in S} \sum_{h \in S} P_{j,h}^{(n_2)} P_{h,k}^{(r)} P_{k,j}^{(n_1)} \geq P_{j,i}^{(n_2)} P_{i,i}^{(r)} P_{i,j}^{(n_1)}.$$

Las dos igualdades son por las ecuaciones de Chapman-Kolmogorov (Teorema 1.1) y la desigualdad es porque  $P_{j,i}^{(n_2)} P_{i,i}^{(r)} P_{i,j}^{(n_1)}$  es elemento de la suma doble de sumandos mayores o

iguales a cero.

Así,

$$\sum_{r=0}^{\infty} P_{j,j}^{(n_2+r+n_1)} \geq \sum_{r=0}^{\infty} P_{j,i}^{(n_2)} P_{i,i}^{(r)} P_{i,j}^{(n_1)} = P_{j,i}^{(n_2)} P_{i,j}^{(n_1)} \sum_{r=0}^{\infty} P_{i,i}^{(r)} = \infty.$$

La desigualdad se tiene porque cada término de la primer suma es mayor o igual a cada término de la segunda suma, la primer igualdad es porque la suma sólo afecta a  $r$  y la segunda igualdad es por hipótesis.

Por lo tanto, por el Teorema 1.2,  $j \in S$  es recurrente.

**Nota.** Este resultado muestra que ser recurrente o transitorio también es una propiedad de clase. Todos los estados en una misma clase son recurrentes o transitorios.

**Definición 1.21** Sea  $X = \{X_n : n = 0, 1, \dots\}$  una cadena de Markov con espacio de estados  $S$  y matriz de transición  $P = (P_{i,j})_{i,j \in S}$ . Se dice que  $i \in S$  es **absorbente** si y sólo si  $P_{i,i} = 1$ .

**Definición 1.22** Sea  $X = \{X_n : n = 0, 1, \dots\}$  una cadena de Markov con espacio de estados  $S$  y sea  $i \in S$ . El **tiempo del primer retorno** al estado  $i$  está definido por

$$T_i = \min\{n \geq 1 : X_{m+n} = i | X_m = i\}.$$

Por convención se dice que  $T_i = \infty$  si el primer retorno a  $i$  nunca ocurre.

**Observación.** Sean  $X$  una cadena de Markov,  $i \in S$  y  $T_i$  como en la Definición 1.22, entonces

$$P(T_i = n) = f_{i,i}^{(n)}.$$

**Definición 1.23** Sea  $X = \{X_n : n = 0, 1, \dots\}$  una cadena de Markov con espacio de estados  $S$  y sea  $i \in S$ . El **número esperado de transiciones necesarias para regresar por**



primera vez al estado  $i$  es:

$$\mu_i = E(T_i | X_m = i) = \begin{cases} \sum_{n=1}^{\infty} n f_{i,i}^{(n)} & \text{si } i \text{ es recurrente} \\ \infty & \text{si } i \text{ es transitorio.} \end{cases}$$

**Definición 1.24** Sea  $X = \{X_n : n = 0, 1, \dots\}$  una cadena de Markov con espacio de estados  $S$ . Tome  $i \in S$  recurrente.

1. Si  $\mu_i = \infty$  entonces  $i$  es llamado **recurrente nulo**.
2. Si  $\mu_i < \infty$  entonces  $i$  es llamado **recurrente positivo**.

**Definición 1.25** Sea  $X = \{X_n : n = 0, 1, \dots\}$  una cadena de Markov y  $S$  su espacio de estados.  $X$  es **irreducible** si  $i \leftrightarrow j$  para todo  $i, j \in S$ .

**Definición 1.26** Sea  $X = \{X_n : n = 0, 1, \dots\}$  una cadena de Markov.  $X$  es llamada cadena **ergódica** si es irreducible, aperiódica y si todos sus estados son recurrentes positivos. Si el espacio de estados  $S$  de  $X$  es finito entonces para que  $X$  sea ergódica basta que sea irreducible y aperiódica.

### 1.3. Distribución límite y distribución estacionaria

Esta sección presentará otros conceptos que también son básicos en la teoría de las cadenas de Markov y útiles para la comprensión de los capítulos posteriores de esta tesis.

**Definición 1.27** Sea  $X = \{X_n : n = 0, 1, \dots\}$  una cadena de Markov,  $S$  su espacio de estados y  $P^{(n)} = (P_{i,j}^{(n)})_{i,j \in S}$  su matriz de transición en  $n$  pasos. Se llama **distribución límite** de  $X$  al  $\lim_{n \rightarrow \infty} P_{i,j}^{(n)}$  cuando éste existe y no depende del estado  $i$ .

**Proposición 1.3** Sea  $X = \{X_n : n = 0, 1, \dots\}$  una cadena de Markov aperiódica e irreducible con espacio de estados  $S$ . Tome  $i, j \in S$ . Así,

$$\lim_{n \rightarrow \infty} P_{i,j}^{(n)} = \frac{1}{\mu_j}.$$

**Demostración.** Ver Grimmett y Stirzaker (1982).

**Proposición 1.4** Sea  $X = \{X_n : n = 0, 1, \dots\}$  una cadena de Markov aperiódica e irreducible con espacio de estados  $S$ .

1. Si  $j \in S$  es recurrente positivo entonces  $\lim_{n \rightarrow \infty} P_{i,j}^{(n)} > 0$ ,  $i \in S$ .
2. Si  $j \in S$  es transitorio o recurrente nulo entonces  $\lim_{n \rightarrow \infty} P_{i,j}^{(n)} = 0$ ,  $i \in S$ .

**Demostración.**

1. Como  $j \in S$  es recurrente positivo se tiene que  $\mu_j < \infty$ , es decir,  $\sum_{n=1}^{\infty} n f_{j,j}^{(n)}$  converge. Por otro lado, como  $X$  es irreducible y aperiódica, usando la Proposición 1.3 se tiene que

$$\lim_{n \rightarrow \infty} P_{i,j}^{(n)} = \frac{1}{\mu_j} > 0.$$

2. Si  $j \in S$  es transitorio o recurrente nulo entonces  $\mu_j = \infty$ . Entonces

$$\lim_{n \rightarrow \infty} P_{i,j}^{(n)} = \frac{1}{\mu_j} = 0.$$

**Definición 1.28** Sea  $X = \{X_n : n = 0, 1, \dots\}$  una cadena de Markov y  $S$  su espacio de estados. La **distribución inicial** de  $X$  está indicada por  $\pi_0(i)$ ,  $i \in S$  y está definida por

$$\pi_0(i) = P(X_0 = i).$$

**Definición 1.29** Sea  $X = \{X_n : n = 0, 1, \dots\}$  una cadena de Markov con espacio de estados  $S$  y matriz de transición  $P = (P_{i,j})_{i,j \in S}$ . Al conjunto  $\pi = \{\pi(j) : j \in S\}$  que satisface

- a)  $\pi(j) \geq 0$  para toda  $j \in S$ ,
- b)  $\sum_{j \in S} \pi(j) = 1$ ,
- c)  $\pi(j) = \sum_{i \in S} \pi(i) P_{i,j}$  para toda  $j \in S$ ,

se le llama **distribución estacionaria** de  $X$ .

**Definición 1.30** Sea  $X = \{X_n : n = 0, 1, \dots\}$  una cadena de Markov con espacio de estados  $S$ . La **distribución al tiempo  $n$**  de  $X$ ,  $n \geq 1$ , indicada por  $\pi_n(i)$ ,  $i \in S$ , está definida por

$$\pi_n(i) = P(X_n = i).$$

**Proposición 1.5** Sea  $X = \{X_n : n = 0, 1, \dots\}$  una cadena de Markov con espacio de estados  $S$ , con matriz de transición  $P = (P_{i,j})_{i,j \in S}$ , con distribución al tiempo  $n$   $\pi_n(i)$ ,  $i \in S$ ,  $n = 0, 1, \dots$  y  $\pi = \{\pi(i) : i \in S\}$  su distribución estacionaria. Si  $\pi_0(i) = \pi(i)$  para toda  $i \in S$  entonces  $\pi_n(i) = \pi(i)$  para toda  $i \in S$  y para toda  $n \geq 1$ .

**Demostración.**

Haciendo esta demostración por inducción sobre  $n$  se tiene lo siguiente:

1. Tome  $n = 1$ . Así

$$\begin{aligned} \pi_1(i) &= P\{X_1 = i\} \\ &= \sum_{k \in S} P\{X_1 = i | X_0 = k\} P\{X_0 = k\} \\ &= \sum_{k \in S} P_{k,i} \pi(k) \\ &= \pi(i). \end{aligned}$$

La primera igualdad es por la Definición 1.30, la tercera igualdad es por definición de probabilidad de transición y por hipótesis, y la última igualdad es por la Definición 1.29, c).

2. Suponga que la igualdad es cierta para  $n = m$ , es decir,  $\pi_m(i) = \pi(i)$  para toda  $i \in S$ .
3. Demostrando para  $n = m + 1$  se tiene:

$$\begin{aligned} \pi_{m+1}(i) &= P\{X_{m+1} = i\} \\ &= \sum_{k \in S} P\{X_{m+1} = i | X_m = k\} P\{X_m = k\} \\ &= \sum_{k \in S} P_{k,i} \pi(k) \\ &= \pi(i). \end{aligned}$$

**Definición 1.31** Sea  $X = \{X_n : 0, 1, \dots\}$  una cadena de Markov. Se dice que  $X$  es una **cadena estacionaria** si su distribución inicial es igual a su distribución estacionaria.

**Observación.** Sea  $X = \{X_n : 0, 1, \dots\}$  una cadena de Markov con distribución estacionaria  $\pi = \{\pi(j) : j \in S\}$ . Se dice que  $X$  ha alcanzado su **estado estacionario al tiempo  $n$**  cuando  $P(X_n = j) = \pi(j)$  para toda  $j \in S$ .

**Teorema 1.3** Sea  $X = \{X_n : 0, 1, \dots\}$  una cadena de Markov ergódica con espacio de estados  $S$  y matriz de transición en  $n$  pasos  $P^{(n)} = \left( P_{i,j}^{(n)} \right)_{i,j \in S}$ . El conjunto

$$\pi = \left\{ \pi(j) = \lim_{n \rightarrow \infty} P_{i,j}^{(n)} > 0, i \in S, j \in S \right\}$$

es la única distribución estacionaria de  $X$ .

**Demostración.**

1. Por ser  $X$  una cadena ergódica todos sus estados son recurrentes positivos, por lo que  $\lim_{n \rightarrow \infty} P_{i,j}^{(n)}$  es mayor a cero y por tanto se cumple que  $\pi(j) \geq 0$ .
2. Se debe demostrar que  $\pi(j) = \sum_{i \in S} \pi(i) P_{i,j}$  con  $j \in S$ . Primero se demostrará que

$$\pi(j) = \sum_{i \in S} \pi(i) P_{i,j}^{(n)} \tag{1.5}$$

para toda  $n \geq 0$  y para toda  $j \in S$ .

- a) Suponga que  $S$  es finito. Por el Teorema 1.1 se sabe que para  $n, m \geq 0$

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

Haciendo tender  $m$  a infinito se tiene que

$$\lim_{m \rightarrow \infty} P_{h,j}^{(m+n)} = \sum_{i \in S} \left( \lim_{m \rightarrow \infty} P_{h,i}^{(m)} \right) P_{i,j}^{(n)}.$$

Finalmente,

$$\pi(j) = \sum_{i \in S} \pi(i) P_{i,j}^{(n)}.$$

b) Suponga que  $S$  es numerable. Sabemos que

$$P_{h,j}^{(m+n)} = \sum_{i \in S} P_{h,i}^{(m)} P_{i,j}^{(n)} \geq \sum_{i=0}^M P_{h,i}^{(m)} P_{i,j}^{(n)}$$

para toda  $M \in S$  y  $m, n \geq 0$ . La igualdad de arriba es por el Teorema 1.1 y la desigualdad es porque la segunda suma es una parte de la primer suma.

Haciendo tender  $m$  a infinito se tiene que

$$\lim_{m \rightarrow \infty} P_{h,j}^{(m+n)} \geq \sum_{i=0}^M \left( \lim_{m \rightarrow \infty} P_{h,i}^{(m)} \right) P_{i,j}^{(n)}.$$

Así,

$$\pi(j) \geq \sum_{i=0}^M \pi(i) P_{i,j}^{(n)}$$

para toda  $M \in S$ . Cuando  $M$  tiende a infinito se tiene que

$$\lim_{M \rightarrow \infty} \pi(j) = \pi(j) \geq \lim_{M \rightarrow \infty} \sum_{i=0}^M \pi(i) P_{i,j}^{(n)} = \sum_{i=0}^{\infty} \pi(i) P_{i,j}^{(n)}.$$

Hasta aquí se tiene entonces que  $\pi(j) \geq \sum_{i \in S} \pi(i) P_{i,j}^{(n)}$  con  $j \in S$  y  $n \geq 0$ , pero suponiendo que

$$\pi(j) > \sum_{i \in S} \pi(i) P_{i,j}^{(n)} \tag{1.6}$$

con  $j \in S$  y  $n \geq 0$  se llegará a una contradicción: al considerar la suma sobre  $j$  en (1.6) se tiene que

$$\sum_{j \in S} \pi(j) > \sum_{j \in S} \sum_{i \in S} \pi(i) P_{i,j}^{(n)} = \sum_{i \in S} \pi(i) \sum_{j \in S} P_{i,j}^{(n)} = \sum_{i \in S} \pi(i)$$

donde la última igualdad ocurre ya que, por la Propiedad 1.1,  $\sum_{j \in S} P_{i,j}^{(n)} = 1$  para toda  $i \in S$  y  $n \geq 0$ . Por lo tanto

$$\pi(j) = \sum_{i \in S} \pi(i) P_{i,j}^{(n)}.$$

En particular, cuando  $n = 1$  se tiene que

$$\pi(j) = \sum_{i \in S} \pi(i) P_{i,j}$$

para toda  $j \in S$ .

3. Ahora se debe demostrar que  $\sum_{j \in S} \pi(j) = 1$  con  $\pi(j) = \lim_{n \rightarrow \infty} P_{i,j}^{(n)}$ .

a) Suponga que  $S$  es finito. Por el inciso b) de la Propiedad 1.1 se tiene que

$$\sum_{j \in S} P_{i,j}^{(n)} = 1$$

para toda  $i \in S$  y  $n \geq 0$ . Haciendo tender  $n$  a infinito entonces

$$\sum_{j \in S} \left( \lim_{n \rightarrow \infty} P_{i,j}^{(n)} \right) = 1$$

por lo que

$$\sum_{j \in S} \pi(j) = 1.$$

b) Suponga que  $S$  es numerable. Por un lado se tiene que

$$1 = \sum_{j \in S} P_{i,j}^{(n)} \geq \sum_{j=0}^M P_{i,j}^{(n)}$$

para toda  $M \in S$  y para toda  $n \geq 0$ . Si  $n$  tiende a infinito entonces

$$1 \geq \sum_{j=0}^M \left( \lim_{n \rightarrow \infty} P_{i,j}^{(n)} \right) = \sum_{j=0}^M \pi(j)$$

para toda  $M \in S$ . Ahora, cuando  $M$  tiende a infinito

$$1 \geq \lim_{M \rightarrow \infty} \sum_{j=0}^M \pi(j) = \sum_{j \in S} \pi(j),$$

lo cual indica que  $\sum_{j \in S} \pi(j)$  converge.

Además note que

$$P_{i,j}^{(n)} \leq 1$$

para toda  $j \in S$  y para toda  $n \geq 0$  por ser una probabilidad condicional, por lo que  $P_{i,j}^{(n)}$  es uniformemente acotado.

Por lo tanto, si  $n$  tiende a infinito en la expresión (1.5) se tiene que

$$\lim_{n \rightarrow \infty} \pi(j) = \pi(j) = \sum_{i \in S} \pi(i) \left( \lim_{n \rightarrow \infty} P_{i,j}^{(n)} \right) = \sum_{i \in S} \pi(i) \pi(j) = \pi(j) \sum_{i \in S} \pi(i)$$

y así

$$1 = \sum_{i \in S} \pi(i)$$

ya que por hipótesis  $\pi(j) > 0$ .

4. Ahora se debe demostrar la unicidad de esta distribución. Suponga que existe  $\nu_j$  con  $j \in S$  tal que el conjunto  $\{\nu_j : j \in S\}$  sea una distribución estacionaria de  $X$ , es decir, cumple con lo siguiente:

- a)  $\nu_j \geq 0$  para toda  $j \in S$ ,
- b)  $\sum_{j \in S} \nu_j = 1$ ,
- c)  $\nu_j = \sum_{i \in S} \nu_i P_{i,j}$  para toda  $j \in S$ .

Se demostrará que  $\nu_j = \pi(j)$  para toda  $j \in S$ . Para esto primero note que por hipótesis la fórmula (1.5) también es válida para el conjunto  $\{\nu_j : j \in S\}$ :

$$\nu_j = \sum_{i \in S} \nu_i P_{i,j}^{(n)} \tag{1.7}$$

para toda  $n \geq 0$  y para toda  $j \in S$ . A partir de aquí se dividirá esta demostración en dos casos.

a) Suponga que  $S$  es finito. Haciendo que  $n$  tienda a infinito en la expresión (1.7) se tiene

$$\nu_j = \sum_{i \in S} \nu_i \left( \lim_{n \rightarrow \infty} P_{i,j}^{(n)} \right) = \sum_{i \in S} \nu_i \pi(j) = \pi(j)$$

donde la última igualdad es válida ya que por el inciso b) de la hipótesis a respecto de  $\nu(\cdot)$ ,  $\sum_{j \in S} \nu_j = 1$ .

Así,

$$\nu_j = \pi(j)$$

para toda  $j \in S$ .

b) Suponga que  $S$  es numerable. De la expresión (1.7) se tiene que

$$\nu_j \geq \sum_{i=0}^M \nu_i P_{i,j}^{(n)}$$

para toda  $M \in S$ . Tomando el límite cuando  $n$  tiende a infinito se tiene

$$\lim_{n \rightarrow \infty} \nu_j = \nu_j \geq \lim_{n \rightarrow \infty} \sum_{i=0}^M \nu_i P_{i,j}^{(n)} = \sum_{i=0}^M \nu_i \left( \lim_{n \rightarrow \infty} P_{i,j}^{(n)} \right) = \sum_{i=0}^M \nu_i \pi(j) \quad (1.8)$$

para todo  $M \in S$ . Si  $M$  tiende a infinito en la expresión (1.8) entonces

$$\lim_{M \rightarrow \infty} \nu_j = \nu_j \geq \lim_{M \rightarrow \infty} \sum_{i=0}^M \nu_i \pi(j) = \sum_{i \in S} \nu_i \pi(j) = \pi(j). \quad (1.9)$$

La última igualdad se tiene por el inciso b) de la hipótesis a respecto de  $\nu(\cdot)$ .

Por otro lado, también de la expresión (1.7) se deduce que

$$\nu_j = \sum_{i=0}^M \nu_i P_{i,j}^{(n)} + \sum_{i=M+1}^{\infty} \nu_i P_{i,j}^{(n)} \quad (1.10)$$

para toda  $n \geq 0$ .

Dado que  $P_{i,j}^{(n)} \leq 1$  para toda  $n \geq 0$  por ser una probabilidad condicional, se tiene de la expresión (1.10) que

$$\nu_j \leq \sum_{i=0}^M \nu_i P_{i,j}^{(n)} + \sum_{i=M+1}^{\infty} \nu_i$$

para toda  $n$ . Cuando  $n$  tiende a infinito en la expresión anterior se tiene

$$\lim_{n \rightarrow \infty} \nu_j = \nu_j \leq \sum_{i=0}^M \nu_i \left( \lim_{n \rightarrow \infty} P_{i,j}^{(n)} \right) + \sum_{i=M+1}^{\infty} \nu_i = \sum_{i=0}^M \nu_i \pi(j) + \sum_{i=M+1}^{\infty} \nu_i. \quad (1.11)$$

Ya que  $\sum_{i \in S} \nu_i = 1$  se tiene

$$\sum_{i=0}^M \nu_i + \sum_{i=M+1}^{\infty} \nu_i = 1$$



y la expresión (1.11) se convierte en

$$\nu_j \leq \sum_{i=0}^M \nu_i \pi(j) + \left(1 - \sum_{i=0}^M \nu_i\right).$$

Si  $M$  tiende a infinito en la expresión anterior entonces

$$\lim_{M \rightarrow \infty} \nu_j = \nu_j \leq \sum_{i \in S} \nu_i \pi(j) + \left(1 - \sum_{i \in S} \nu_i\right) = \pi(j) \sum_{i \in S} \nu_i = \pi(j). \quad (1.12)$$

Finalmente, de las expresiones (1.9) y (1.12) podemos concluir que

$$\nu_j = \pi(j)$$

para toda  $j \in S$ .

**Observación.**

- a) Sea  $X = \{X_n : 0, 1, \dots\}$  una cadena de Markov aperiódica e irreducible con espacio de estados  $S$ . Si todos sus estados son recurrentes nulos o transitorios entonces no existe la distribución límite de  $X$ .
- b) La distribución estacionaria  $\pi(\cdot)$  de una cadena de Markov  $X = \{X_n : 0, 1, \dots\}$  con espacio de estados  $S$  se puede interpretar como la probabilidad de que la cadena esté en el estado  $j \in S$  dado que suficiente tiempo ha transcurrido desde el momento en que se empieza a observar la sucesión.

## 1.4. Cadenas de Markov reversibles en el tiempo

Esta sección hablará de una propiedad con respecto al tiempo que tienen algunas cadenas de Markov y que será útil en los siguientes capítulos de esta tesis.

**Definición 1.32** Sea  $X = \{X_n : n = \dots, -1, 0, 1, \dots\}$  una cadena de Markov estacionaria con espacio de estados  $S$  y matriz de transición  $P = (P_{i,j})_{i,j \in S}$ . La **cadena reversa en el**

**tiempo** con respecto a  $P_{i,j}$  de  $X$  es el proceso estocástico  $X^* = \{X_n^* : n = \dots, -1, 0, 1, \dots\}$  definido por el siguiente condicional:

$$\text{Si } X_n^* = X_n \text{ entonces } X_{n+1}^* = X_{n-1} \text{ para toda } n = 0, 1, \dots$$

Es decir,  $X^*$  es la cadena  $X$  pero observada del presente al pasado: si  $X_n^*$  es el estado presente de la cadena  $X^*$ , sus estados pasados son  $X_{n+1}, X_{n+2}, \dots$  y sus estados futuros son  $X_{n-1}, X_{n-2}, \dots$

**Observación.** En el caso de una cadena de Markov  $X = \{X_n : n = 0, 1, \dots\}$ , ser estacionaria es equivalente a imaginar que la cadena empezó en el tiempo  $-\infty$  y por tanto en el presente ya está en el estado estacionario, por lo tanto se puede escribir  $X = \{X_n : n = \dots, -1, 0, 1, \dots\}$  y por tanto  $X^*$  también puede extenderse para  $n \in \mathbb{Z}$ .

**Proposición 1.6** Sean  $X = \{X_n : n = 0, 1, \dots\}$  una cadena de Markov estacionaria,  $S$  su espacio de estados,  $P = (P_{i,j})_{i,j \in S}$  su matriz de transición,  $\pi = \{\pi(i) : i \in S\}$  su distribución estacionaria y  $X^*$  su cadena reversa en el tiempo. Entonces,  $X^*$  también es una cadena de Markov con espacio de estados  $S$ . Su probabilidad de transición  $P_{i,j}^*$ ,  $i, j \in S$  está definida por

$$P_{i,j}^* = P\{X_{n+1}^* = j | X_n^* = i\} = \frac{\pi(j)}{\pi(i)} P_{j,i}$$

para toda  $i, j \in S$ .

**Demostración.**

Para que  $X^*$  sea una cadena de Markov debe cumplir que

$$P\{X_{n+m}^* = i_{n+m} | X_n^* = i_n, \dots, X_{n+m-1}^* = i_{n+m-1}\} = P\{X_{n+m}^* = i_{n+m} | X_{n+m-1}^* = i_{n+m-1}\}$$

donde  $n, m \geq 0$ ,  $i_n, \dots, i_{n+m-1}, i_{n+m} \in S$ .

$$\begin{aligned}
& P\{X_{n+m}^* = i_{n+m} | X_n^* = i_n, \dots, X_{n+m-1}^* = i_{n+m-1}\} = \\
& P\{X_{n-m} = i_{n+m} | X_n = i_n, \dots, X_{n-m+1} = i_{n+m-1}\} = \\
& \frac{P\{X_{n-m} = i_{n+m}, X_{n-m+1} = i_{n+m-1}, \dots, X_n = i_n\}}{P\{X_{n-m+1} = i_{n+m-1}, \dots, X_n = i_n\}} = \\
& \frac{P\{X_{n-m} = i_{n+m}, X_{n-m+2} = i_{n+m-2}, \dots, X_n = i_n | X_{n-m+1} = i_{n+m-1}\}}{P\{X_{n-m+2} = i_{n+m-2}, \dots, X_n = i_n | X_{n-m+1} = i_{n+m-1}\}} \times \\
& \frac{P\{X_{n-m+1} = i_{n+m-1}\}}{P\{X_{n-m+1} = i_{n+m-1}\}} = \\
& P\{X_{n-m} = i_{n+m} | X_{n-m+1} = i_{n+m-1}\} \times \\
& \frac{P\{X_{n-m+2} = i_{n+m-2}, \dots, X_n = i_n | X_{n-m+1} = i_{n+m-1}\}}{P\{X_{n-m+2} = i_{n+m-2}, \dots, X_n = i_n | X_{n-m+1} = i_{n+m-1}\}} = \\
& P\{X_{n-m} = i_{n+m} | X_{n-m+1} = i_{n+m-1}\} = \\
& P\{X_{n+m}^* = i_{n+m} | X_{n+m-1}^* = i_{n+m-1}\}.
\end{aligned}$$

La primera y última igualdad se tienen por la Definición 1.32 y la cuarta igualdad se tiene porque la cadena  $X$  satisface la propiedad de Markov, la cual es equivalente a la siguiente igualdad <sup>4</sup>:

$$\begin{aligned}
& P\{X_{n+m} = i_{n+m}, X_{n+m-2} = i_{n+m-2}, \dots, X_0 = i_0 | X_{n+m-1} = i_{n+m-1}\} = \\
& P\{X_{n+m} = i_{n+m} | X_{n+m-1} = i_{n+m-1}\} P\{X_{n+m-2} = i_{n+m-2}, \dots, X_0 = i_0 | X_{n+m-1} = i_{n+m-1}\}.
\end{aligned}$$

Por otro lado se tiene que

$$\begin{aligned}
P_{i,j}^* &= P\{X_{n+1}^* = j | X_n^* = i\} \\
&= \frac{P\{X_{n-1} = j, X_n = i\}}{P\{X_n = i\}} \frac{P\{X_{n-1} = j\}}{P\{X_{n-1} = j\}} \\
&= \frac{P\{X_n = i | X_{n-1} = j\} P\{X_{n-1} = j\}}{P\{X_n = i\}} \\
&= \frac{P_{j,i} \pi(j)}{\pi(i)}.
\end{aligned}$$

---

<sup>4</sup>Ver Apéndice A, Lema A.1.

La primera igualdad es por definición de probabilidad de transición y la última igualdad es también por la definición de probabilidad de transición y porque se está suponiendo que  $X$  es estacionaria (Proposición 1.5 y Definición 1.32).

**Propiedad 1.4** Sean  $X = \{X_n : n = \dots, -1, 0, 1, \dots\}$  una cadena de Markov estacionaria,  $S$  su espacio de estados,  $P = (P_{i,j})_{i,j \in S}$  su matriz de transición,  $\pi = \{\pi(i) : i \in S\}$  su distribución estacionaria y  $X^* = \{X_n^* : n = \dots, -1, 0, 1, \dots\}$  su cadena reversa en el tiempo. La distribución estacionaria de  $X^*$  también es  $\pi$ .

**Demostración.**

Por ser  $\pi = \{\pi(j) : j \in S\}$  distribución estacionaria de  $X$  cumple que

1.  $\pi(j) \geq 0$  para toda  $j \in S$ ,
2.  $\sum_{j \in S} \pi(j) = 1$ ,
3.  $\pi(j) = \sum_{i \in S} \pi(i)P_{i,j}$  para toda  $j \in S$ .

Como  $X^*$  está definida también en  $S$  entonces  $\pi(j)$  cumple con las primeras dos propiedades anteriores, por lo que solo debe demostrarse que  $\pi(j) = \sum_{i \in S} \pi(i)P_{i,j}^*$  para toda  $j \in S$ .

$$\begin{aligned} \sum_{i \in S} \pi(i)P_{i,j}^* &= \sum_{i \in S} \pi(i) \left( \frac{\pi(j)}{\pi(i)} \right) P_{j,i} \\ &= \sum_{i \in S} \pi(j)P_{j,i} \\ &= \pi(j) \sum_{i \in S} P_{j,i} \\ &= \pi(j) \end{aligned}$$

La primera igualdad se obtiene sustituyendo la expresión para  $P_{i,j}^*$  y la última igualdad es por la Propiedad 1.1.

**Definición 1.33** Sean  $X = \{X_n : n = \dots, -1, 0, 1, \dots\}$  una cadena de Markov estacionaria,  $S$  su espacio de estados,  $P = (P_{i,j})_{i,j \in S}$  su matriz de transición,  $\pi = \{\pi(i) : i \in S\}$  su

#### 1.4. Cadenas de Markov reversibles en el tiempo

---

distribución estacionaria y  $X^* = \{X_n^* : n = \dots, -1, 0, 1, \dots\}$  su cadena reversa en el tiempo con matriz de transición  $P^* = \left(P_{i,j}^*\right)_{i,j \in S}$ . Si  $P_{i,j}^* = P_{i,j}$  para toda  $i, j \in S$  se dice que  $X$  es una **cadena reversible en el tiempo** con respecto a  $\pi$ .

**Propiedad 1.5** Sean  $X = \{X_n : n = 0, 1, \dots\}$  una cadena de Markov estacionaria,  $S$  su espacio de estados,  $P = \left(P_{i,j}\right)_{i,j \in S}$  su matriz de transición y  $\pi = \{\pi(i) : i \in S\}$  su distribución estacionaria.  $X$  es reversible en el tiempo con respecto a  $\pi$  si y sólo si

$$\pi(i)P_{i,j} = \pi(j)P_{j,i} \quad (1.13)$$

para toda  $i, j \in S$ .

**Demostración.**

- (i) Suponga que  $X$  es reversible en el tiempo, es decir que  $P_{i,j}^* = P_{i,j}$ . De aquí  $\frac{\pi(j)}{\pi(i)}P_{j,i} = P_{i,j}$ , por lo que  $\pi(i)P_{i,j} = \pi(j)P_{j,i}$ .
- (ii) Ahora suponga que  $\pi(i)P_{i,j} = \pi(j)P_{j,i}$ . En consecuencia  $P_{i,j} = \frac{\pi(j)}{\pi(i)}P_{j,i}$  y por la Proposición 1.6 el lado derecho de la igualdad es  $P_{i,j}^*$ .

Así  $X$  es reversible en el tiempo.

**Nota.** La igualdad (1.13) es llamada **ecuación detallada de balance**.

**Propiedad 1.6** Sean  $X = \{X_n : n = 0, 1, \dots\}$  una cadena de Markov estacionaria,  $S$  su espacio de estados,  $P = \left(P_{i,j}\right)_{i,j \in S}$  su matriz de transición. Supongamos que  $X$  satisface la ecuación

$$f(i)P_{i,j} = f(j)P_{j,i} \quad (1.14)$$

para toda  $i, j \in S$ , donde  $f$  es una función de probabilidad definida sobre  $S$ .

- a) La función de probabilidad  $f$  es la distribución estacionaria de la cadena  $X$ .
- b) La cadena  $X$  es reversible en el tiempo con respecto a  $f$ .

**Demostración.**

a) Para asegurar esto se debe ver que:

- i)  $f(j) \geq 0$  para toda  $j \in S$ ,
- ii)  $\sum_{j \in S} f(j) = 1$ ,
- iii)  $f(j) = \sum_{i \in S} f(i)P_{i,j}$  para toda  $j \in S$ .

Por ser  $f$  una función de probabilidad sobre  $S$  se cumplen los incisos i) y ii) <sup>5</sup>. Falta verificar iii). Tome  $j \in S$ , entonces

$$\begin{aligned} \sum_{i \in S} f(i)P_{i,j} &= \sum_{i \in S} f(j)P_{j,i} \\ &= f(j) \sum_{i \in S} P_{j,i} \\ &= f(j). \end{aligned}$$

La primera igualdad se tiene por la expresión (1.14) y la última igualdad se tiene ya que  $\sum_{i \in S} P_{j,i} = 1$  según la Propiedad 1.1.

b) Por a) se sabe que  $f$  es la distribución estacionaria de  $X$ , por lo que se cumple la ecuación (1.13) y por la Propiedad 1.5 la cadena es reversible en el tiempo con respecto a  $f$ .

---

<sup>5</sup>Ver Apéndice A, Definición A.2.



## Capítulo 2

# Ejemplos de Cadenas de Markov

En este capítulo se presentarán algunos ejemplos de cadenas de Markov relacionados con la genética. Se encontrarán las formas generales de sus probabilidades de transición y se clasificarán los estados de sus matrices de transición, mostrando así la utilidad de las cadenas de Markov en esta área de la ciencia.

### 2.1. Modelos de Wright-Fisher

En esta sección se obtendrán las probabilidades de transición de las cadenas de Markov que modelan la reproducción aleatoria de una población con determinadas características.

#### 2.1.1. Modelo haploide simple de reproducción aleatoria sin considerar la mutación ni la selección

Considere una población de tamaño fijo  $2N$  compuesta por dos tipos de individuos: aquellos que tienen genes tipo  $A$  y los que tienen genes tipo  $B$ . Suponga que la reproducción es de tipo haploide (de un solo progenitor). Además, es igualmente probable que cualquiera de los individuos se reproduzca y también suponga que cada individuo puede tener varios



hijos. Cada nacimiento es independiente de los otros y no hay mutaciones (si el progenitor es de tipo  $A$ , el descendiente será también de tipo  $A$ ) ni selección (la probabilidad de que un individuo del tipo  $A$  se reproduzca es la misma que la probabilidad del de tipo  $B$ ). Suponga que hay  $i$  individuos del tipo  $A$ . Dado que el tamaño de la población es fijo, se tiene que existen  $2N - i$  individuos del tipo  $B$ . La probabilidad de que un descendiente tipo  $A$  sea generado es  $p_i = \frac{i}{2N}$  ya que  $i$  es el número de individuos tipo  $A$  y no hay mutación, por tanto, los  $i$  individuos de este tipo son los posibles progenitores (casos favorables) y  $2N$  es el número de individuos en la población (casos totales).

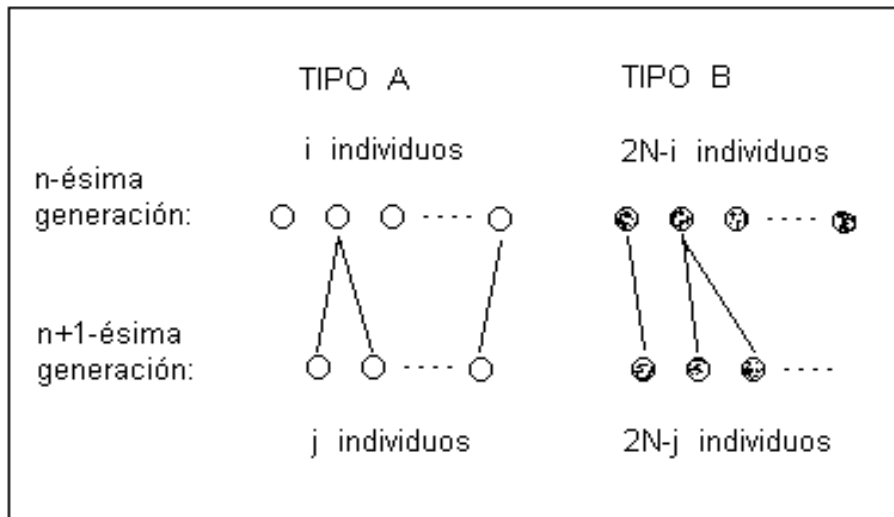


Figura 2.1: Diagrama de la reproducción aleatoria sin mutación ni selección.

De la misma manera, la probabilidad de generar un descendiente de tipo  $B$  es  $q_i = \frac{2N-i}{2N} = 1 - \frac{i}{2N}$  porque los posibles progenitores son los individuos tipo  $B$  (pues no hay mutación), y existen  $2N - i$  de tales individuos. Los casos totales son los  $2N$  individuos de la población.

Considere la cadena de Markov  $X = \{X_n : n = 0, 1, \dots\}$  donde

$X_n = \text{número de } A\text{-genes en la } n\text{-ésima generación.}$

De esta forma,  $X$  tiene espacio de estados  $S = \{0, \dots, 2N\}$  dado que en cualquier generación el número de individuos de tipo  $A$  podría ser cero, en cuyo caso no habría individuos de este tipo, o podría ser  $2N$  y entonces no habría individuos tipo  $B$ , o podría ser igual a cualquier número entre 0 y  $2N$ .

Así, la probabilidad de transición de  $X$  es la siguiente:

$$P_{i,j} = P\{X_{n+1} = j \mid X_n = i\} = \binom{2N}{j} \left(\frac{i}{2N}\right)^j \left(\frac{2N-i}{2N}\right)^{2N-j}$$

para todo  $i, j \in S$ . Esto se debe a que de los  $2N$  individuos de la generación  $n+1$ , se quiere que  $j$  individuos sean de tipo  $A$ , así que ¿de cuántas maneras podemos elegir, de toda la generación, a estos  $j$  individuos? La respuesta es de  $\binom{2N}{j}$  maneras. Considerando que en la  $n$ -ésima generación aparece un descendiente tipo  $A$  con probabilidad  $\frac{i}{2N}$  y que se quieren  $j$  individuos de este tipo en la  $(n+1)$ -ésima generación entonces de aquí se tiene  $\left(\frac{i}{2N}\right)^j$ . Adicionalmente se quiere que los otros  $2N-j$  individuos sean de tipo  $B$ , cada uno con probabilidad  $\frac{2N-i}{2N}$  de ser procreados en la generación  $n$ -ésima, de aquí se tiene  $\left(\frac{2N-i}{2N}\right)^{2N-j}$ .

### 2.1.2. Modelo haploide simple de reproducción aleatoria considerando la mutación

Considere el modelo de reproducción dado en la subsección anterior, pero ahora suponga que existe la posibilidad de reproducir individuos mutantes. Por tanto, si el progenitor es de tipo  $A$ , el descendiente es de tipo  $B$  con probabilidad  $\alpha_1$  y, si el progenitor es de tipo  $B$ , el descendiente es de tipo  $A$  con probabilidad  $\alpha_2$ . De igual manera, el tamaño de la población en cada generación se mantendrá fijo, es decir, su tamaño será  $2N$ .

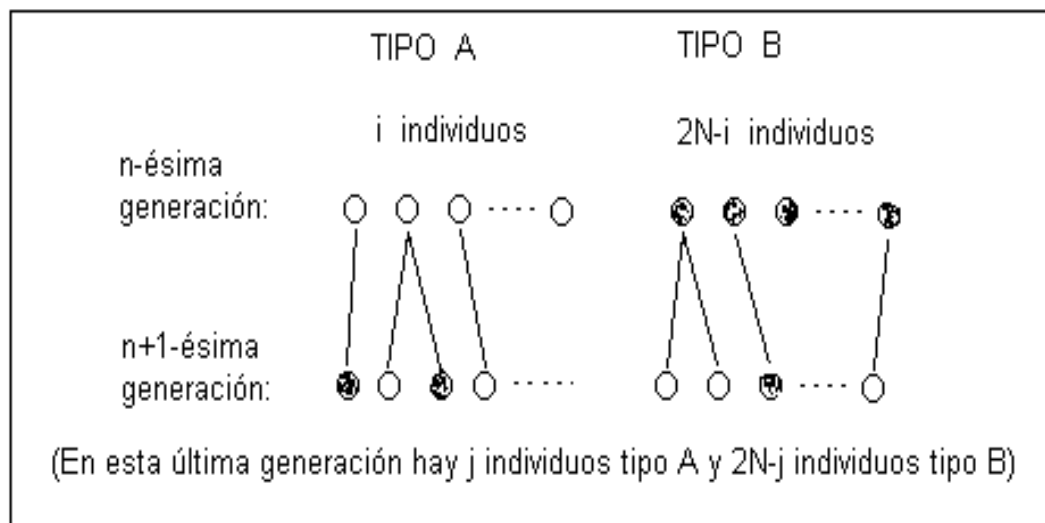


Figura 2.2: Diagrama de la reproducción aleatoria con mutación.

Suponga que existen  $i$  individuos de tipo  $A$  en una generación dada. La probabilidad de que sea procreado un individuo tipo  $A$  en la siguiente generación es

$$p_i = \left( \frac{i}{2N} \right) (1 - \alpha_1) + \left( \frac{2N - i}{2N} \right) (\alpha_2).$$

Esto se debe a que en cada evento de reproducción puede pasar sólo uno de los siguientes dos casos: que el progenitor sea tipo  $A$  y este no produzca un mutante (de donde se tiene el primer sumando), ó que sea tipo  $B$ , y sí produzca un mutante (segundo sumando).

Analizando el primer caso se tiene que  $\frac{i}{2N}$  es la probabilidad de que este individuo sea procreado por alguien de tipo  $A$  (por haber en total  $i$  individuos de este tipo en la generación reproductora) y  $1 - \alpha_1$  es la probabilidad de la *no mutación* cuando el progenitor es tipo  $A$ .

En el segundo caso,  $2N - i$  es el número de individuos tipo  $B$  en la generación reproductora, por lo que  $\frac{2N-i}{2N}$  es la probabilidad de que un individuo en la generación producida sea procreado por alguien de tipo  $B$ , y  $\alpha_2$  es la probabilidad de producir un mutante cuando el progenitor es de tipo  $B$ .

Por otro lado,

$$q_i = \binom{i}{2N} (\alpha_1) + \binom{2N-i}{2N} (1 - \alpha_2)$$

es la probabilidad de que se genere un individuo de tipo  $B$ . Como arriba, tenemos dos posibles casos: puede ser que el progenitor sea de tipo  $A$  y produzca un mutante, de donde tenemos  $\binom{i}{2N} \alpha_1$ , ó puede que sea de tipo  $B$  y produzca un no mutante, de donde se obtiene el otro sumando.

De todo lo anterior se tiene que en este caso la probabilidad de transición de  $X = \{X_n : n = 0, 1, \dots\}$  definido en el modelo anterior es:

$$\begin{aligned} P_{i,j} &= P\{X_{n+1} = j \mid X_n = i\} \\ &= \binom{2N}{j} \left( \binom{i}{2N} (1 - \alpha_1) + \binom{2N-i}{2N} \alpha_2 \right)^j \left( \binom{i}{2N} \alpha_1 + \binom{2N-i}{2N} (1 - \alpha_2) \right)^{2N-j} \\ &= \binom{2N}{j} (p_i)^j (q_i)^{2N-j} \end{aligned}$$

para todo  $i, j \in S$ . Esta expresión se obtiene ya que hay  $\binom{2N}{j}$  maneras de elegir de un grupo de  $2N$  individuos (de la  $(n+1)$ -ésima generación) a un grupo de  $j$  individuos tipo  $A$ , pero la probabilidad para cada uno de ellos de ser de este tipo es  $p_i$ , entonces la probabilidad de que todos los  $j$  individuos sean de tipo  $A$  es  $(p_i)^j$ . Lo mismo se tiene para el grupo de los  $2N - j$  individuos restantes, los cuales deben ser de tipo  $B$ , y esto ocurre con probabilidad  $(q_i)^{2N-j}$ .

### 2.1.3. Modelo haploide simple de reproducción aleatoria considerando la selección

Suponga ahora que la fuerza de selección operará a favor de los individuos tipo  $A$  y que el número esperado de sus descendientes es *proporcional* a  $1 + s$ , con  $0 < s < 1$ . Por otro lado, para los individuos tipo  $B$  se espera que sea proporcional a 1. Se supone que cada generación mantiene su tamaño en  $2N$  donde si  $i$  de los individuos en una generación dada son del tipo

A, el restante, es decir,  $2N - i$  individuos, son del tipo B.

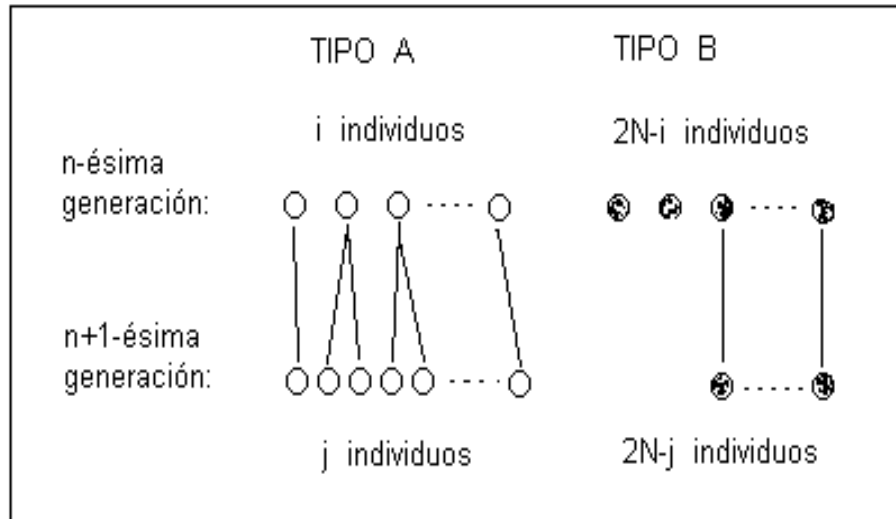


Figura 2.3: Diagrama de la reproducción aleatoria con selección.

La probabilidad de que sea procreado un individuo tipo A está dada por  $p_i = \frac{(1+s)i}{2N+si}$  dado que por la fuerza de selección  $(1+s)i$  es el número esperado de individuos de este tipo en la generación reproductora (es el número de posibles progenitores). Sin embargo se debe cumplir que los  $(2N - i)$  individuos restantes en esta misma generación, los de tipo B, mantengan su proporción unitaria. Entonces el total esperado de individuos en la generación progenitora es  $(1+s)i + 2N - i = 2N + si$ , el cual es el denominador en  $p_i$ .<sup>1</sup>

La probabilidad de que sea generado un individuo tipo B es  $q_i = \frac{2N-i}{2N+si}$  porque, igual que para  $p_i$ , en la generación reproductora se espera que hayan  $2N - i$  posibles progenitores y se tienen  $2N + si$  individuos en total.

<sup>1</sup>Note que si  $i = 2N$  y en lugar de  $2N + si$  se tuviera  $2N$  en el denominador de  $p_i$  se tendría que esta probabilidad sería igual a  $\frac{(1+s)2N}{2N} = 1 + s$ , lo cual es una contradicción ya que por ser una probabilidad,  $p_i$  debe ser menor o igual a 1. Esto muestra que se debe considerar la proporción de el número esperado de individuos tipo A para el total esperado de individuos en cada generación.

Finalmente, la probabilidad de transición de  $X = \{X_n : n = 0, 1, \dots\}$  definido en la primera subsección es:

$$P_{i,j} = P\{X_{n+1} = j \mid X_n = i\} = \binom{2N}{j} \left(\frac{i + si}{2N + si}\right)^j \left(\frac{2N - i}{2N + si}\right)^{2N-j}$$

para todo  $i, j \in S$ . Esta expresión se obtiene porque  $\binom{2N}{j}$  es el número de todas las formas en las que podemos elegir a un grupo de  $j$  individuos de los  $2N$  que son en total, el segundo factor es la probabilidad para este grupo de ser tipo  $A$  y el tercer factor es la probabilidad para el grupo restante de ser tipo  $B$ .

## 2.2. Modelo de la reproducción celular

En esta sección se obtendrán las probabilidades de transición de la cadena de Markov que modela la reproducción celular.

Todas las *células* tienen una subunidad llamada *núcleo* donde están localizados los *cromosomas* que son las unidades donde se localizan los *genes*. Los genes a su vez están formados de subunidades llamadas *nucleótidos*, que en el proceso de reproducción celular pueden producir subunidades mutantes. Cuando los cromosomas se reproducen se dividen en dos (siempre conservando su condición mutante o no mutante) logrando que cada parte de la célula también se divida en dos y, así, ella se reproduzca. Suponga que un gen tiene  $N$  nucleótidos. Al duplicarse se tendrán  $2N$  nucleótidos. En la etapa reproductiva éstas  $2N$  subunidades se repartirán aleatoriamente entre los dos nuevos genes, de tal manera que cada gen volverá a tener  $N$  nucleótidos. Este fenómeno se puede comparar con las extracciones de bolas de una urna.

Sea  $X = \{X_n : n = 0, \dots\}$  una cadena de Markov donde  $X_n = i$  si un gen, en la  $n$ -ésima generación, tiene una composición de  $i$  subunidades mutantes y por tanto  $N - i$  subunidades no mutantes. Por lo tanto  $X$  tiene espacio de estados  $S = \{0, \dots, N\}$  ya que en alguna

## 2.2. Modelo de la reproducción celular

---

generación un gen podría tener puras subunidades no mutantes ( $i = 0$ ), ó podría tener todas sus subunidades mutantes ( $i = N$ ), ó podría tener cualquier cantidad, entre cero y  $N$ , de subunidades mutantes.

Así, la probabilidad de transición de  $X$  es:

$$P_{i,j} = P\{X_{n+1} = j \mid X_n = i\} = \frac{\binom{2i}{j} \binom{2N-2i}{N-j}}{\binom{2N}{N}}$$

para todo  $i, j \in S$ . Esta expresión se obtiene por la siguiente razón: suponga que un gen tiene  $N$  nucleótidos con  $i$  de ellos mutantes. Después de duplicarse, habrán  $2N$  nucleótidos con  $2i$  de ellos mutantes y por tanto  $2N - 2i$  no mutantes que deben repartirse entre los dos nuevos genes, cada uno con un total de  $N$  subunidades. Sin embargo, esto debe ocurrir de tal manera que uno de los genes tenga exactamente  $j$  nucleótidos mutantes y por tanto  $N - j$  no mutantes. Cabe preguntarse ¿de cuántas maneras se puede elegir a un grupo de  $j$  subunidades mutantes de las  $2i$  que son en total? La respuesta es  $\binom{2i}{j}$  maneras, pero por cada una de ellas hay otras  $\binom{2N-2i}{N-j}$  maneras en las que podemos elegir al grupo de los  $N - j$  nucleótidos no mutantes del total de ellos que es  $2N - 2i$ . Por esto,  $\binom{2i}{j} \binom{2N-2i}{N-j}$  es el total de maneras en las que podemos formar al nuevo gen con  $j$  nucleótidos no mutantes (casos favorables). Por último,  $\binom{2N}{N}$  es el total de formas en las que se pueden elegir de  $2N$  subunidades a un grupo de  $N$  de ellas (casos totales).

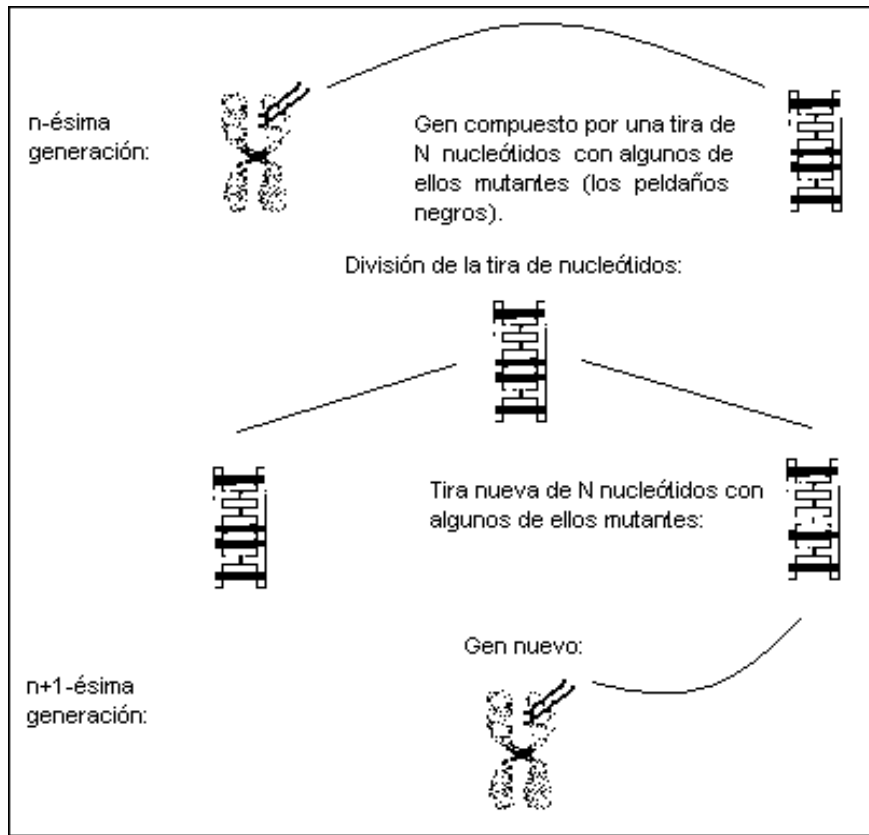


Figura 2.4: Diagrama de la reproducción celular.

## 2.3. Clasificación de estados

En esta sección se mostrará un ejemplo de la clasificación de los estados de una matriz de transición de una cadena de Markov para después mostrar las matrices de transición de las cadenas de Markov de los modelos anteriores y clasificar sus estados.

Considere la siguiente matriz de transición correspondiente a una cadena de Markov con espacio de estados  $S = \{0, 1, 2, 3, 4, 5\}$ :



$$P = \begin{pmatrix} \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ \frac{3}{4} & 0 & 0 & \frac{1}{4} & 0 & 0 \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & 0 \\ 0 & \frac{1}{10} & 0 & \frac{1}{10} & \frac{1}{10} & \frac{7}{10} \end{pmatrix}.$$

En esta matriz existen tres clases, a saber,  $\mathcal{C}_1 = \{0, 3\}$ ,  $\mathcal{C}_2 = \{1, 2, 4\}$  y  $\mathcal{C}_3 = \{5\}$ . Esto es así dado que el estado 0 sólo se comunica con el estado 3, es decir, existen enteros  $m_1, m_2 \geq 0$  tales que  $P_{0,3}^{(m_1)} > 0$  y  $P_{3,0}^{(m_2)} > 0$ , los cuales son  $m_1 = m_2 = 1$ ,  $P_{0,3} = \frac{1}{2}$  y  $P_{3,0} = \frac{3}{4}$ . Además los estados 1, 2, 4 y 5 no son accesibles desde los estados 0 y 3 ya que  $P_{0,j}^{(m)} = P_{3,j}^{(m)} = 0$  para todo entero  $m \geq 0$  y  $j = 1, 2, 4, 5$ , lo cual hace que estos dos estados no se comuniquen con los estados restantes y por eso los dos forman una sola clase.

Los estados de  $\mathcal{C}_2$  se comunican entre sí ya que 1 y 2 son los enteros mayores a cero tales que  $P_{1,2} > 0$ ,  $P_{1,4}^{(2)} > 0$ ,  $P_{2,1}^{(2)} > 0$ ,  $P_{2,4} > 0$ ,  $P_{4,1} > 0$  y  $P_{4,2} > 0$  donde  $P_{1,2} = P_{1,4}^{(2)} = \frac{1}{4}$ ,  $P_{2,1}^{(2)} = P_{4,1} = P_{4,2} = \frac{1}{5}$  y  $P_{2,4} = 1$ . Para que los estados 0, 3 y 5 puedan estar en  $\mathcal{C}_2$  debe ocurrir que ellos se comuniquen con los estados 1, 2 y 4, es decir, que los estados 0, 3 y 5 sean accesibles desde 1, 2 y 4 y que a su vez 1, 2 y 4 sean accesibles desde 0, 3 y 5, pero ya se vió que esta última condición no se cumple, por lo que 0, 3 y 5 no están comunicados con los estados de  $\mathcal{C}_2$ .

Finalmente, note que el estado 5 no es accesible desde los demás estados ya que  $P_{i,5}^{(m)} = 0$  para todo entero  $m \geq 0$  con  $i = 0, 1, 2, 3, 4$ , por lo que no se comunica con algún otro estado y por eso él forma su propia clase.

La clase  $\mathcal{C}_1 = \{0, 3\}$  es recurrente porque es finita y una vez que se entra a ella nunca se podrá salir. La clase  $\mathcal{C}_2 = \{1, 2, 4\}$  es transitoria porque hay una probabilidad positiva de pasar a la clase  $\mathcal{C}_1$  y no salir de ella, es decir, hay una probabilidad positiva de salir de la clase  $\mathcal{C}_2$  y jamás regresar. La clase  $\mathcal{C}_3 = \{5\}$  también es transitoria porque existe una probabilidad positiva de salir de ella, entrar a cualquiera de las otras dos clases y no volver a ella.

Para las siguientes subsecciones considere un caso particular de los ejemplos de las Secciones 2.1 y 2.2 donde se toma  $N = 2$ .

### 2.3.1. Modelo haploide simple sin mutación ni selección

En este caso el proceso estóastico considerado es  $X = \{X_n : n = 0, 1, \dots\}$  donde

$X_n =$  número de  $A$ -genes en la  $n$ -ésima generación

y  $S = \{0, 1, 2, 3, 4\}$ . Así, la matriz de transición de este proceso es:

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.277 & 0.421 & 0.210 & 0.046 & 0.003 \\ 0.062 & 0.250 & 0.375 & 0.250 & 0.062 \\ 0.003 & 0.046 & 0.210 & 0.421 & 0.277 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Esta matriz tiene tres clases:  $\mathcal{C}_1 = \{0\}$ ,  $\mathcal{C}_2 = \{4\}$  y  $\mathcal{C}_3 = \{1, 2, 3\}$ . Note que los estados 0 y 4 no se comunican entre ellos ni con los demás estados porque no existen enteros positivos  $n$  y  $m$  tales que  $P_{0,k}^{(n)} > 0$  y  $P_{4,j}^{(m)} > 0$  donde  $k = 1, 2, 3, 4$  y  $j = 0, 1, 2, 3$ . Por esto los estados 0 y 4 forman cada uno su propia clase.

### 2.3. Clasificación de estados

---

Los estados de  $\mathcal{C}_3$  forman esta clase porque se comunican entre ellos ya que 1 es el número de pasos en los que, con probabilidad positiva, se puede llegar de el estado 1 a los estados 2 y 3, del estado 2 a los estados 1 y 3 y del estado 3 a los estados 1 y 2, pero no se comunican con los estados 0 y 4.

Las clases  $\mathcal{C}_1$  y  $\mathcal{C}_2$  son recurrentes porque no se puede salir de ellas. También son absorbentes porque hay una probabilidad positiva de que, desde cualquier estado de  $\mathcal{C}_3$ , se pueda llegar a ellas pero no se pueda salir. La clase  $\mathcal{C}_3$  es transitoria dado que existe una probabilidad positiva de salir de esta clase y jamás regresar.

**Observación.** Los resultados dados en esta subsección reflejan lo que pasa en la población, dado que si en alguna generación no hay individuos de tipo  $A$  entonces, como no hay mutación, en la siguiente generación tampoco habrá individuos de este tipo. Esto significa que el estado 0 es absorbente. Pero si en alguna generación todos los individuos son de tipo  $A$  entonces también en la siguiente generación todos los individuos serán de este tipo y por esto el estado 4 también es absorbente.

#### 2.3.2. Modelo haploide simple con mutación

Para este ejemplo se tiene el mismo proceso estocástico considerado en el ejemplo anterior, con el mismo espacio de estados. Ahora para fines de ilustración se darán valores fijos a las probabilidades de mutación (de individuos tipo  $A$  a individuos tipo  $B$  y viceversa) para que se pueda analizar una matriz de transición en particular. Tome  $\alpha_1 = \frac{1}{4}$  y  $\alpha_2 = \frac{2}{3}$ . La matriz de transición en este caso es

$$P = \begin{pmatrix} 0.012 & 0.098 & 0.296 & 0.395 & 0.197 \\ 0.009 & 0.083 & 0.276 & 0.406 & 0.223 \\ 0.007 & 0.070 & 0.256 & 0.414 & 0.251 \\ 0.005 & 0.057 & 0.233 & 0.419 & 0.282 \\ 0.003 & 0.046 & 0.210 & 0.421 & 0.316 \end{pmatrix}.$$

En esta matriz todos los estados se comunican entre sí ya que 1 es el entero positivo tal que  $P_{i,j}^{(1)} > 0$  donde  $i$  y  $j$  pueden ser cualquiera de los cinco estados, es decir, un estado es accesible desde cualquier otro estado, por lo que sólo hay una clase que es  $\mathcal{C}_1 = \{0, 1, 2, 3, 4\}$ . Esta clase es recurrente porque al menos un estado de esta matriz es recurrente (por la Proposición 1.1), pero este estado pertenece a la misma clase que los demás estados (con quienes está comunicado), por lo que todos los otros estados también son recurrentes (por la Proposición 1.2).

Note que en este caso no se puede asegurar, como en el ejemplo anterior, que si no hay individuos tipo  $A$  en alguna generación entonces en la siguiente generación tampoco los habrá. Esto es porque en este caso sí hay mutación. Si no hubiera individuos de tipo  $A$  entonces todos ellos serían del tipo  $B$  y sí podrían engendrar individuos de tipo  $A$ . Lo mismo pasaría si todos los individuos fueran de tipo  $A$ , ellos también podrían tener hijos tipo  $B$ .

Por otro lado, considerando que  $\alpha_1$  y  $\alpha_2$  toman valores entre cero y uno (por ser probabilidades) se pueden encontrar las formas generales que las matrices de transición tienen según los valores de éstas probabilidades (ver Cuadro 2.1), esto a través de las probabilidades de transición del proceso (ver la Subsección 2.1.2).

2.3. Clasificación de estados

	$\alpha_2 = 0$	$0 < \alpha_2 < 1$	$\alpha_2 = 1$
$\alpha_1 = 0$	$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ p_{1,0} & p_{1,1} & p_{1,2} & p_{1,3} & p_{1,4} \\ p_{2,0} & p_{2,1} & p_{2,2} & p_{2,3} & p_{2,4} \\ p_{3,0} & p_{3,1} & p_{3,2} & p_{3,3} & p_{3,4} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} p_{0,0} & p_{0,1} & p_{0,2} & p_{0,3} & p_{0,4} \\ p_{1,0} & p_{1,1} & p_{1,2} & p_{1,3} & p_{1,4} \\ p_{2,0} & p_{2,1} & p_{2,2} & p_{2,3} & p_{2,4} \\ p_{3,0} & p_{3,1} & p_{3,2} & p_{3,3} & p_{3,4} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$
$0 < \alpha_1 < 1$	$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ p_{1,0} & p_{1,1} & p_{1,2} & p_{1,3} & p_{1,4} \\ p_{2,0} & p_{2,1} & p_{2,2} & p_{2,3} & p_{2,4} \\ p_{3,0} & p_{3,1} & p_{3,2} & p_{3,3} & p_{3,4} \\ p_{4,0} & p_{4,1} & p_{4,2} & p_{4,3} & p_{4,4} \end{pmatrix}$	$\begin{pmatrix} p_{0,0} & p_{0,1} & p_{0,2} & p_{0,3} & p_{0,4} \\ p_{1,0} & p_{1,1} & p_{1,2} & p_{1,3} & p_{1,4} \\ p_{2,0} & p_{2,1} & p_{2,2} & p_{2,3} & p_{2,4} \\ p_{3,0} & p_{3,1} & p_{3,2} & p_{3,3} & p_{3,4} \\ p_{4,0} & p_{4,1} & p_{4,2} & p_{4,3} & p_{4,4} \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ p_{1,0} & p_{1,1} & p_{1,2} & p_{1,3} & p_{1,4} \\ p_{2,0} & p_{2,1} & p_{2,2} & p_{2,3} & p_{2,4} \\ p_{3,0} & p_{3,1} & p_{3,2} & p_{3,3} & p_{3,4} \\ p_{4,0} & p_{4,1} & p_{4,2} & p_{4,3} & p_{4,4} \end{pmatrix}$
$\alpha_1 = 1$	$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} p_{0,0} & p_{0,1} & p_{0,2} & p_{0,3} & p_{0,4} \\ p_{1,0} & p_{1,1} & p_{1,2} & p_{1,3} & p_{1,4} \\ p_{2,0} & p_{2,1} & p_{2,2} & p_{2,3} & p_{2,4} \\ p_{3,0} & p_{3,1} & p_{3,2} & p_{3,3} & p_{3,4} \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ p_{1,0} & p_{1,1} & p_{1,2} & p_{1,3} & p_{1,4} \\ p_{2,0} & p_{2,1} & p_{2,2} & p_{2,3} & p_{2,4} \\ p_{3,0} & p_{3,1} & p_{3,2} & p_{3,3} & p_{3,4} \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$

Cuadro 2.1: Matrices de transición para los varios valores de las probabilidades de mutación

En las matrices del Cuadro 2.1 las probabilidades  $P_{i,j}$  con  $i, j \in \{0, 1, 2, 3, 4\}$  toman valores estrictamente mayores a cero. A continuación se analizarán dichas matrices.

Si  $\alpha_1 = \alpha_2 = 0$  la matriz tiene tres clases que son  $\mathcal{C}_1 = \{0\}$  (recurrente y absorbente),  $\mathcal{C}_2 = \{1, 2, 3\}$  (transitoria) y  $\mathcal{C}_3 = \{4\}$  (recurrente y absorbente), coincidiendo con el caso de la matriz de la Subsección 2.3.1.

Si  $\alpha_1 = 0$  y  $0 < \alpha_2 < 1$  se tienen las clases  $\mathcal{C}_1 = \{0, 1, 2, 3\}$  y  $\mathcal{C}_2 = \{4\}$ . La primer clase es transitoria y la segunda es recurrente y absorbente ya que existe una probabilidad positiva de pasar de  $\mathcal{C}_1$  a  $\mathcal{C}_2$  y nunca salir de allí.

Cuando  $\alpha_1 = 0$  y  $\alpha_2 = 1$  las clases son  $\mathcal{C}_1 = \{0\}$ ,  $\mathcal{C}_2 = \{1\}$ ,  $\mathcal{C}_3 = \{2\}$ ,  $\mathcal{C}_4 = \{3\}$  y  $\mathcal{C}_5 = \{4\}$ ; las primeras cuatro son transitorias y la quinta es recurrente y absorbente ya que desde

cualquier estado es seguro pasar al estado 4 sin salir de él.

Si se tiene que  $0 < \alpha_1 < 1$  y  $\alpha_2 = 0$  las clases son  $\mathcal{C}_1 = \{0\}$  (recurrente y absorbente) y  $\mathcal{C}_2 = \{1, 2, 3, 4\}$  (transitoria) ya que existe una probabilidad positiva de pasar de  $\mathcal{C}_2$  a  $\mathcal{C}_1$  y no salir de allí.

Si  $0 < \alpha_1 < 1$  y  $0 < \alpha_2 < 1$  solo hay una clase:  $\mathcal{C}_1 = \{0, 1, 2, 3, 4\}$  que es recurrente, como en la matriz del ejemplo de ésta subsección.

Cuando  $0 < \alpha_1 < 1$  y  $\alpha_2 = 1$  también hay una sola clase que es recurrente ya que todos los estados de la matriz se comunican entre sí. Note que para ir desde el estado 0 a los demás estados se necesitan dos pasos (se debe pasar primero por el estado 4), mientras que para ir desde los otros estados a cualquier estado (incluso al estado 0) se necesita solo un paso.

Si  $\alpha_1 = 1$  y  $\alpha_2 = 0$  las clases son  $\mathcal{C}_1 = \{0\}$ ,  $\mathcal{C}_2 = \{1\}$ ,  $\mathcal{C}_3 = \{2\}$ ,  $\mathcal{C}_4 = \{3\}$  y  $\mathcal{C}_5 = \{4\}$ . La primera es recurrente y absorbente, todas las demás son transitorias. Note que esta matriz es parecida a la matriz donde  $\alpha_1 = 0$  y  $\alpha_2 = 1$ , con la diferencia de que en este caso es seguro pasar al estado 0 (y no al estado 4) desde cualquier estado.

Si  $\alpha_1 = 1$  y  $0 < \alpha_2 < 1$  entonces hay solo una clase, la cual es recurrente. En este caso, para ir desde el estado 4 a los otros estados se debe pasar primero por el estado 0.

Cuando  $\alpha_1 = \alpha_2 = 1$  las clases son  $\mathcal{C}_1 = \{0, 4\}$  (recurrente) y  $\mathcal{C}_2 = \{1, 2, 3\}$  (transitoria) ya que existe una probabilidad positiva de salir de  $\mathcal{C}_2$  y nunca más regresar.

### 2.3.3. Modelo haploide simple con selección

En este ejemplo se considera el mismo proceso estocástico del ejemplo de la Subsección 2.1.3 con el mismo espacio de estados pero ahora tomando en cuenta que el número esperado de los descendientes tipo  $A$  es proporcional a  $1 + s$  con  $s = \frac{1}{3}$  y el número esperado de descendientes tipo  $B$  es proporcional a 1. La matriz de transición en este caso es:

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.229 & 0.408 & 0.272 & 0.080 & 0.008 \\ 0.033 & 0.179 & 0.359 & 0.319 & 0.106 \\ 0.001 & 0.025 & 0.153 & 0.409 & 0.409 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

La clasificación de los estados de esta matriz es exactamente como en el ejemplo de la Subsección 2.3.1 porque las entradas que son mayores a cero y las que son iguales a cero ocupan los mismos lugares que en la otra matriz. Así,  $\mathcal{C}_1 = \{0\}$  y  $\mathcal{C}_2 = \{4\}$  son las clases recurrentes, y  $\mathcal{C}_3 = \{1, 2, 3\}$  es la clase transitoria.

Por no haber mutación, si hubiera solo individuos tipo  $B$  (o de tipo  $A$ ) en alguna generación, en la siguiente generación también habría solo individuos de este tipo, lo que comprueba que los estados 0 y 4 son absorbentes.

### 2.3.4. Modelo de la reproducción celular

En este caso el proceso estocástico es  $X = \{X_n : n = 0, 1, \dots\}$  donde

$X_n =$  número de nucleótidos mutantes de un gen en la  $n$ -ésima generación

y  $S = \{0, 1, 2\}$ . La matriz de transición es:

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0.17 & 0.66 & 0.17 \\ 0 & 0 & 1 \end{pmatrix}.$$

Las clases de esta matriz son  $\mathcal{C}_1 = \{0\}$ ,  $\mathcal{C}_2 = \{1\}$  y  $\mathcal{C}_3 = \{2\}$  ya que el estado 0 y el estado 2 no se comunican entre ellos ni con el estado 1. Además existen probabilidades positivas de salir de  $\mathcal{C}_2$ , entrar a  $\mathcal{C}_1$  o  $\mathcal{C}_3$  y no salir de ellas, por lo que las clases  $\mathcal{C}_1$  y  $\mathcal{C}_3$  son recurrentes y absorbentes, y  $\mathcal{C}_2$  es la clase transitoria.

En este ejemplo de reproducción celular, cuando los nucleótidos mutantes se dividen en dos partes éstas se mantienen mutantes. Por esto, si un gen en alguna generación no tiene algún nucleótido mutante, al repartirse los nucleótidos divididos, el nuevo gen tampoco tendrá algún nucleótido mutante, lo cual concuerda con que la cadena no pueda salir del estado 0. Igualmente se puede verificar que la cadena no puede salir del estado 2.





## Capítulo 3

# El Algoritmo

# Metropolis-Hastings

En este capítulo se presentarán algunos métodos para estimar integrales, los cuales estarán basados en la simulación de variables aleatorias. El uso de estos resultados básicos será ilustrado en los siguientes capítulos de esta tesis.

### 3.1. Método Monte Carlo

En esta sección se explicará la utilidad del Método Monte Carlo y se presentarán algunos ejemplos de algoritmos que utilizan este método para obtener aproximaciones de integrales y series. Los aspectos teóricos presentados aquí podrán ser encontrados en Carlin y Louis (2000), Evans y Swartz (2000), Gutiérrez-Peña (1997), Robert y Casella (1999) y Ross (1997).

Considere una variable aleatoria  $X$  definida sobre un espacio de probabilidad  $(\Omega, \mathcal{F}, \mathcal{P})$  que toma valores en  $S \subseteq \mathbb{R}$  con función de densidad  $f$  y sea  $h : S \rightarrow \mathbb{R}$  una función tal que  $h^{-1}(B) \in \mathfrak{B}(\mathbb{R})$  para todo  $B \in \mathbb{R}$ , es decir, tal que  $h(X)$  también sea una variable aleatoria

real. Note que se puede escribir

$$E[h(X)] = \int_S h(x)f(x)dx. \quad (3.1)$$

Cuando la expresión (3.1) es difícil de calcular directamente, ya sea por la complejidad de  $S$  o de las funciones involucradas, el método Monte Carlo permite aproximar esta expresión a través de la simulación de variables aleatorias con función de densidad  $f$  y del uso de la Ley Fuerte de los Números Grandes <sup>1</sup>. Para mostrar cómo se puede hacer esto considere  $X_1, X_2, X_3, \dots$  un conjunto de variables aleatorias independientes e idénticamente distribuidas con función de densidad  $f$ . La expresión

$$\frac{1}{N} \sum_{k=1}^N h(X_k)$$

converge casi seguramente a la fórmula (3.1) cuando  $N$  tiende a infinito, es decir,

$$P \left( \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N h(X_k) = E[h(X)] \right) = 1,$$

ver, por ejemplo, Feller (1968). Así (3.1) se puede aproximar generando una cantidad suficientemente grande de valores  $x_1, x_2, x_3, \dots, x_N$  a partir de la función  $f$ , evaluándolos en la función  $h$  y calculando su promedio muestral.

Incluso, este método no solo ayuda a calcular la integral (3.1). También se puede usar para encontrar otras integrales, por ejemplo

$$\int_S g(x)dx \quad (3.2)$$

donde  $g : S \rightarrow \mathbb{R}$ . Sea  $f$  una función de densidad definida sobre  $S$  (Figura 3.1) tal que

- a)  $f(x) > 0$  para toda  $x \in S$ ,
- b)  $f$  tiene mayor densidad en los intervalos de  $S$  donde la función  $g$  tiene mayor área,

---

<sup>1</sup>Ver Apéndice A, Teorema A.2.

c) es fácil simular valores a partir de ella.

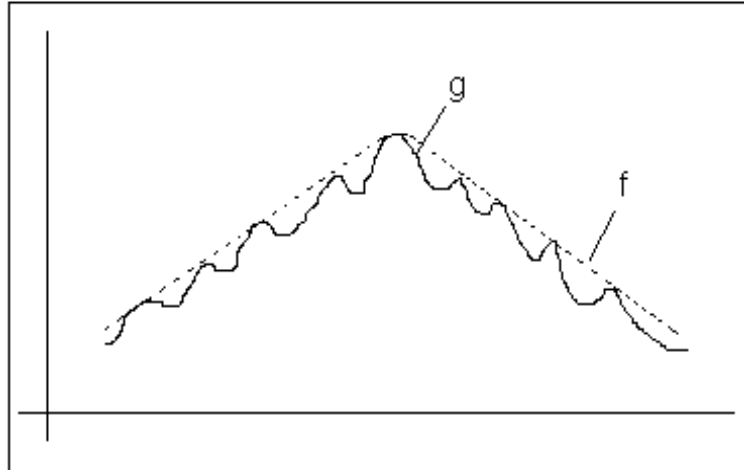


Figura 3.1: Gráficas de las funciones  $f$  y  $g$ .

Note que la integral

$$\int_S \frac{g(x)}{f(x)} f(x) dx \quad (3.3)$$

es igual a (3.1) donde ahora  $h(x) = \frac{g(x)}{f(x)}$ . Por esta razón, para aproximar esta integral se deben obtener valores  $x_1, x_2, \dots, x_N$  simulados a partir de  $f$  con  $N$  suficientemente grande, evaluarlos en  $h$  y calcular su promedio muestral, es decir, la fórmula (3.2) puede ser aproximada por

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

para  $N$  suficientemente grande.

Un caso particular del uso de este método es el siguiente: suponga que se quiere calcular la integral

$$\int_a^b g(x) dx \quad (3.4)$$

donde  $g : \mathbb{R} \rightarrow \mathbb{R}$ . Suponga también que existe  $c > 0$  tal que  $0 \leq g(x) \leq c$  para toda  $x \in [a, b]$ . Esta integral se puede aproximar al obtenerse valores muestreados a partir de la

función de densidad uniforme siguiente:

$$f(x, y) = \begin{cases} \frac{1}{c(b-a)} & \text{si } (x, y) \in L \\ 0 & \text{e.o.c,} \end{cases}$$

donde  $y = g(x)$  y  $L = [a, b] \times [0, c]$ , y así obtener una muestra suficientemente grande de valores de  $f$  para después contar cuantos de ellos caen bajo la curva que forma la función  $g$  en el intervalo  $[a, b]$  (Figura 3.2).

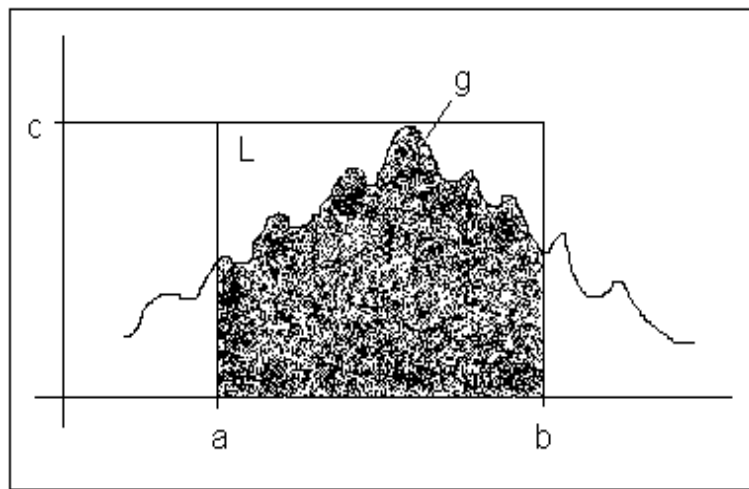


Figura 3.2: Gráfica de la función  $g$ .

Así la expresión (3.4) se puede aproximar por medio de

$$\sum_{i=1}^N I_B(x, y)$$

donde  $B = \{(x, y) \in L : a \leq x \leq b, 0 \leq g(x) \leq c\}$ .

### 3.2. Métodos Monte Carlo vía Cadenas de Markov

Esta sección dará una descripción rápida de los métodos de Monte Carlo basados en cadenas de Markov utilizados para obtener muestras a partir de una distribución deseada y donde

es muy difícil obtener esta muestra directamente de ella.

Algunas veces no es tan fácil obtener directamente muestras de algunas distribuciones, como por ejemplo de distribuciones con una forma algebraica compleja o las que son multivariadas. En estos casos se pueden usar métodos basados en una cadena de Markov ergódica cuya distribución estacionaria es precisamente la distribución de interés.

Un método de Monte Carlo vía cadenas de Markov es cualquier método que produce una cadena de Markov ergódica con distribución estacionaria  $f$ , donde  $f$  es la distribución a partir de la cuál se quiere simular valores.

Estos métodos funcionan en general de la siguiente manera: tomando cualquier valor inicial se genera una cadena de Markov  $X = \{X_n : n = 0, 1, 2, \dots\}$  con distribución estacionaria  $f$ . Así, para un tiempo  $T_0$  suficientemente grande se puede considerar que  $X_{T_0}$  tiene distribución  $f$  y las variables aleatorias  $X_{T_0}, X_{T_0+1}, X_{T_0+2}, \dots$  forman una muestra generada por  $f$ . Note que por ser  $X$  una cadena de Markov las variables  $X_{T_0}, X_{T_0+1}, X_{T_0+2}, \dots$  son dependientes. Para que se pueda obtener una muestra aproximadamente independiente se tiene que estimar el número de pasos en que se debe tomar los valores simulados. Una sugerencia es tomar el siguiente camino. Se puede tomar un valor  $n_0$  suficientemente grande de forma que  $X_{T_0+n_0}, X_{T_0+2n_0}, \dots$  sean casi independientes. El valor  $n_0$  puede ser obtenido de la siguiente forma. Sea  $\lambda$  el mayor eigenvalor distinto de uno de la matriz de transición  $P = (P_{i,j})_{i,j \in S}$  de la cadena  $X$ , entonces se tiene que para  $n \geq 1$  vale

$$d(P_{i_0, \cdot}^{(n)}, f(\cdot)) \leq \lambda^n$$

donde  $d(\cdot, \cdot)$  es una distancia entre funciones de distribución. Sea  $\epsilon > 0$  un valor muy pequeño. Entonces  $n_0$  es el valor tal que  $\lambda^{n_0} \approx \epsilon$ .

Dos de los métodos de Monte Carlo vía Cadenas de Markov son el algoritmo Metropolis-Hastings y el muestreo de Gibbs. Para el primero se necesitan muy pocos requerimientos

sobre la distribución que se quiere simular y se puede obtener con él una muestra dependiente de variables aleatorias cuya distribución siempre converge para la distribución de interés. Para el segundo se necesita cierto conocimiento de esta distribución para realizar algunos cálculos y la muestra de variables aleatorias obtenida mediante él no siempre convergerá.

### 3.3. Algoritmo Metropolis-Hastings

Esta sección explicará la utilidad y base teórica del algoritmo Metropolis-Hastings.

Suponga que se quiere obtener valores a través de una simulación de una sucesión de variables aleatorias independientes con función de probabilidad  $f(i) = \frac{a_i}{\sum_{i=1}^m a_i}$  donde  $a_i > 0$ ,  $i = 1, \dots, m$  con  $m$  grande y  $\sum_{i=1}^m a_i$  difícil de calcular. Una manera de hacer esto es construir, por medio del algoritmo Metropolis-Hastings, una cadena de Markov que tenga como distribución estacionaria al conjunto  $\{\pi(i) = f(i) : i = 1, \dots, m\}$  y que se puedan simular fácilmente valores de esta cadena de Markov. El algoritmo Metropolis-Hastings puede ser descrito de la siguiente forma.

Sea  $Q = (Q_{i,j})_{i,j \in S}$  una matriz de probabilidades de transición sobre  $S = \{1, \dots, m\}$ , y tal que  $Q_{i,j} > 0$  para todo  $i, j \in S$  de forma que sean fácil de simular valores con esta distribución. Con esta matriz se debe generar una cadena de Markov  $X = \{X_k : k = 0, 1, 2, \dots\}$  definida de la siguiente manera:

1. Si  $X_n = i$ , genere una variable aleatoria  $Y$  tal que  $P(Y = \cdot | X_n = i) = Q_{i,\cdot}$ , donde  $i = 1, \dots, m$ .

2. Si  $Y = j$  entonces

$$\begin{cases} P(X_{n+1} = j) = \rho(i, j) \\ P(X_{n+1} = i) = 1 - \rho(i, j) \end{cases} \quad (3.5)$$

donde

$$\rho(i, j) = \min \left\{ \frac{f(j)Q_{j,i}}{f(i)Q_{i,j}}, 1 \right\} = \min \left\{ \frac{a_j Q_{j,i}}{a_i Q_{i,j}}, 1 \right\}.$$

### 3.3.1. Algoritmo Metropolis-Hastings

Considerando lo anterior, los pasos para iterar el algoritmo son los siguientes:

1. Elegir una matriz de transición  $Q = (Q_{i,j})_{i,j \in S}$  a partir de la cual sea fácil simular valores. Sea  $X_n = k$ ,  $k \in S$ . Al comenzar el algoritmo se tiene  $n = 0$ .
2. Generar una variable aleatoria  $Y$  tal que  $P(Y = \cdot | X_n = k) = Q_{k,\cdot}$ . donde  $k \in S$  y generar también una variable aleatoria  $U$  que se distribuya Uniforme sobre el intervalo  $(0,1)$ .
3. Suponga que  $Y = j$  donde  $j \in S$ . Si  $U = u < \rho(k, j)$  hacer  $X_{n+1} = j$ . Si  $u \geq \rho(k, j)$  hacer  $X_{n+1} = k$ .
4. Volver al paso 2 donde ahora  $n = n + 1$ .

**Teorema 3.1** Sea  $X = \{X_k : k = 0, 1, 2, \dots\}$  la cadena de Markov con espacio de estados  $S = \{1, \dots, m\}$  generada por el algoritmo Metropolis-Hastings. Entonces,

a) La matriz de transición de  $X$  es  $P = (P_{i,j})_{i,j \in S}$  donde

$$P_{i,j} = \begin{cases} Q_{i,j}\rho(i, j) & \text{si } j \neq i \\ Q_{i,i} + \sum_{j \neq i} Q_{i,j}(1 - \rho(i, j)) & \text{si } j = i. \end{cases} \quad (3.6)$$

b) La distribución estacionaria de  $X$  es el conjunto  $\{\pi(i) = f(i) : i = 1, 2, \dots, m\}$ .

c)  $X$  es ergódica.

**Demostración.**

- a) Note que del algoritmo y de la ecuación (3.5) se tiene que para que la variable aleatoria  $X_{n+1}$  sea igual a  $j$  dado que  $X_n = i$ , la variable  $Y$  debió tomar el valor  $j$  (esto



ocurrió con probabilidad  $Q_{i,j}$  y la variable  $X_{n+1}$  también debió tomar el valor  $j$  (esto con probabilidad  $\rho(i,j)$ ). Como estos dos eventos son independientes se multiplican sus probabilidades. Así  $P(X_{n+1} = j|X_n = i) = Q_{i,j}\rho(i,j)$  con  $j \neq i$ . Sin embargo, dado que  $X_n = i$ ,  $X_{n+1}$  pudo haber tomado el valor  $i$  de dos maneras distintas y ajenas: que al simular la variable  $Y$ , ésta hubiera tomado el mismo valor que  $X_n$  (esto con probabilidad  $Q_{i,i}$ ) y  $X_{n+1}$  hubiera tomado el mismo valor  $i$  (con probabilidad  $\rho(i,i) = 1$ ) ó que al simular la variable  $Y$  ésta hubiera tomado cualquier valor distinto a  $i$ , mismo que después fue rechazado por el algoritmo y entonces  $X_{n+1}$  tomó el valor  $i$  (esto con probabilidad  $1 - \rho(i,j)$ ). Finalmente,  $P(X_{n+1} = i|X_n = i) = Q_{i,i} + \sum_{j \neq i} Q_{i,j}(1 - \rho(i,j))$  con  $j = i$ . Por tanto, la probabilidad de transición de la nueva cadena  $X$  es la expresión (3.6).

- b) Para probar que la nueva cadena  $X$  con matriz de transición  $P = \left( P_{i,j} \right)_{i,j \in S}$  donde  $P_{i,j}$  es como en (3.6) tiene como distribución estacionaria al conjunto  $\{\pi(i) = f(i) : i = 1, 2, \dots, m\}$ , primero se verificará que

$$\pi(i)P_{i,j} = \pi(j)P_{j,i} \quad (3.7)$$

para todo estado  $i, j \in S$ . Note que esto se cumple inmediatamente cuando  $i = j$  y para cuando  $i \neq j$  entonces

$$\begin{aligned} f(i)P_{i,j} &= f(i)Q_{i,j} \min \left\{ \frac{f(j)Q_{j,i}}{f(i)Q_{i,j}}, 1 \right\} \\ &= \min \{ f(j)Q_{j,i}, f(i)Q_{i,j} \} \\ &= \min \left\{ f(j)Q_{j,i}, \frac{f(i)Q_{i,j}}{f(j)Q_{j,i}} f(j)Q_{j,i} \right\} \\ &= f(j)Q_{j,i} \min \left\{ \frac{f(i)Q_{i,j}}{f(j)Q_{j,i}}, 1 \right\} \\ &= f(j)P_{j,i}. \end{aligned}$$

Ya que (3.7) se satisface, entonces por la Propiedad 1.6 se tiene que el conjunto  $\{\pi(i) = f(i) : i \in S\}$  es la distribución estacionaria de la cadena  $X$ .

c) Finalmente, para que la nueva cadena  $X$  con espacio de estados finito  $S$  y matriz de transición  $P = \left( P_{i,j} \right)_{i,j \in S}$  sea ergódica deben cumplirse las condiciones siguientes:

1. Que  $X$  sea aperiódica, lo cual se tiene ya que  $P_{i,i} = Q_{i,i} + \sum_{j \neq i} Q_{i,j}(1 - \rho(i,j)) > 0$  para toda  $i \in S$  porque  $Q$  es positiva. Esto implica que  $d(i) = \text{m.c.d.}\{n \geq 1 : P_{i,i}^{(n)} > 0\} = 1$  para todo  $i \in S$ .
2. Que  $X$  sea irreducible, esto es que todos sus estados se comuniquen entre sí. Para que esto ocurra deben existir enteros  $n, m > 0$  tales que  $P_{i,j}^{(n)} > 0$  y  $P_{j,i}^{(m)} > 0$  para todo  $i, j \in S$ . Por la demostración anterior ya se tiene que  $P_{i,i} > 0$  para toda  $i \in S$  y para cuando  $i \neq j$  también  $P_{i,j} = Q_{i,j}\rho(i,j) > 0$  porque  $Q$  es positiva y  $\rho(i,j) > 0$ . Esto muestra que  $P_{i,j} > 0$  para toda  $i, j \in S$  y haciendo  $n = m = 1$  se tiene la irreducibilidad.

**Observación.**

Ya que el espacio de estados  $S$  de  $X$  es finito, para ver que  $X$  sea ergódica basta ver que sea irreducible y aperiódica (ver la Definición 1.26).

Por otro lado, note que cuando  $Q$  es simétrica (es decir, cuando  $Q_{i,j} = Q_{j,i}$ )  $\rho(i,j)$  sólo dependerá de  $\frac{a_j}{a_i}$ . Además, en todos los casos este algoritmo sólo depende de este último cociente y de  $\frac{Q_{j,i}}{Q_{i,j}}$  sin importar la constante normalizadora  $\sum_{i=1}^m a_i$  que a veces es difícil de calcular.



## Capítulo 4

# Estadística Bayesiana

En este capítulo se presentarán algunas definiciones, teoremas y ejemplos relacionados con la inferencia Bayesiana. Estos resultados serán utilizados en el próximo capítulo de esta tesis. La información presentada aquí fue obtenida a partir de Carlin y Louis (2000), Lee (1989), Leonard y Hsu (1999) y López Ramírez (2005).

### 4.1. Función de verosimilitud

En esta sección se comenzará con algunas definiciones básicas útiles para la comprensión del resto de este capítulo.

**Definición 4.1** *Considere un experimento aleatorio que puede ser descrito mediante un modelo matemático (estocástico o no). La variable que determina dicho modelo se llama **parámetro** y se denotará por  $\theta$ .*

**Definición 4.2** *El conjunto de posibles valores que el parámetro puede tomar se llama **espacio parametral** y se denotará por  $\Theta$ . El espacio parametral especifica una familia de funciones de densidad, es decir, especifica*

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

#### 4.1. Función de verosimilitud

---

donde  $f(x, \theta)$  es una función de densidad con parámetro  $\theta \in \Theta$  y  $x$  es un elemento del espacio donde  $f$  toma valores.

En esta tesis se considerará que  $\theta$  es una variable aleatoria a la que se le puede asignar una función de probabilidad.

**Definición 4.3** Suponga que se efectúan  $N$  repeticiones de un experimento aleatorio. Al conjunto de valores  $\mathbf{x} = (x_1, \dots, x_N)$  obtenidos a partir de las variables aleatorias idénticamente distribuidas  $\mathbf{X} = \{X_1, X_2, \dots, X_N\}$ , donde  $X_i$  con  $i = 1, 2, \dots, N$  registra el resultado de la  $i$ -ésima repetición de un mismo experimento aleatorio, se le llama **muestra aleatoria**.

En esta tesis se considerará que los experimentos son realizados independientemente, es decir, las muestras aleatorias mencionadas estarán compuestas de valores obtenidos independientemente.

**Definición 4.4** Sea  $\mathbf{x} = (x_1, x_2, \dots, x_N)$  una muestra aleatoria y  $\theta \in \Theta$  el parámetro que describe el modelo del experimento aleatorio que produjo esta muestra. Una **estadística** es cualquier función de la muestra que no depende del parámetro desconocido  $\theta$ .

#### Observación.

Algunos ejemplos de estadísticas son:

- a) La media muestral:  $T(\mathbf{x}) = \frac{\sum_{i=1}^N x_i}{N}$ ,
- b) La varianza muestral:  $T(\mathbf{x}) = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N-1}$ ,
- c) La mínima estadística de orden:  $T(\mathbf{x}) = \min\{x_1, x_2, \dots, x_n\}$ ,
- d) La máxima estadística de orden:  $T(\mathbf{x}) = \max\{x_1, x_2, \dots, x_n\}$ ,
- e) El  $r$ -ésimo momento muestral:  $T(\mathbf{x}) = \frac{\sum_{i=1}^N x_i^r}{N}$ .

**Definición 4.5** Sea  $\mathbf{x} = (x_1, x_2, \dots, x_N)$  una muestra aleatoria y  $\theta \in \Theta$  el parámetro desconocido del modelo que produjo  $\mathbf{x}$ . La función que describe el modelo que produjo  $\mathbf{x}$  vista

como función de  $\theta$  es llamada la **función de verosimilitud** que se denotará por  $l(\theta|\mathbf{x})$ . La función de verosimilitud es tal que

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

donde  $f(x|\theta)$  es la función de densidad asociada al experimento que produjo  $\mathbf{x}$ .

**Observación.**

- a) La función de verosimilitud resume toda la información del modelo contenida en los datos y toda inferencia acerca del parámetro debe hacerse tomando en cuenta dicha función.
- b) El principio de verosimilitud establece que si para dos experimentos distintos se obtienen funciones de verosimilitud proporcionales, entonces las inferencias obtenidas de ellas son las mismas.

**Definición 4.6** Sean  $\mathbf{x} = (x_1, x_2, \dots, x_N)$  una muestra aleatoria,  $\theta \in \Theta$  el parámetro del modelo que describe el experimento que produjo la muestra  $\mathbf{x}$  y  $l(\theta|\mathbf{x})$  la función de verosimilitud asociada al modelo. La **función log-verosimilitud** está dada por

$$L(\theta|\mathbf{x}) = \log l(\theta|\mathbf{x}).$$

**Definición 4.7** Sean  $\mathbf{x} = (x_1, x_2, \dots, x_N)$  una muestra aleatoria y  $\theta \in \Theta$  el parámetro desconocido del modelo que describe el experimento que produjo la muestra  $\mathbf{x}$ . Un **estimador** es una estadística que sirve para aproximar el valor de alguna función del parámetro del modelo.

**Definición 4.8** Sean  $\mathbf{x} = (x_1, x_2, \dots, x_N)$  una muestra aleatoria y  $\theta \in \Theta$  el parámetro desconocido del modelo que describe el experimento que produjo la muestra  $\mathbf{x}$ . El **estimador máximo verosimil** de  $\theta$  es cualquier  $\hat{\theta}$  tal que maximice la función de verosimilitud, es decir

$$l(\hat{\theta}|\mathbf{x}) \geq l(\theta|\mathbf{x})$$

para todo  $\theta \in \Theta$ .

4.1. Función de verosimilitud

---

**Ejemplo 4.1** Para ejemplificar las definiciones anteriores considere lo siguiente. Sea  $\mathbf{x} = (x_1, x_2, \dots, x_N)$  una muestra aleatoria producida por variables aleatorias  $X_i$  con  $i = 1, 2, \dots, N$  que se distribuyen como sigue:

$$P(x_i|\lambda, \theta) = \lambda \exp(-\lambda(x_i - \theta)), \quad \theta < x_i < \infty$$

donde  $0 < \lambda < \infty$  y  $-\infty < \theta < \infty$ . La función de verosimilitud para los dos parámetros de interés es

$$\begin{aligned} l(\lambda, \theta|\mathbf{x}) &\propto \prod_{i=1}^N P(x_i|\lambda, \theta) \\ &= \prod_{i=1}^N \lambda \exp(-\lambda(x_i - \theta)) \prod_{i=1}^N I_{(\theta, \infty)}(x_i) \\ &= \lambda^N \exp\left(-\lambda\left(\sum_{i=1}^N x_i - N\theta\right)\right) \prod_{i=1}^N I_{(\theta, \infty)}(x_i) \\ &= \lambda^N \exp(-N\lambda(\bar{x} - \theta)) I_{(\theta, y_1)}(x_i) \end{aligned} \quad (4.1)$$

donde  $y_1 = \min\{x_1, x_2, \dots, x_N\}$  y  $\bar{x} = \frac{\sum_{i=1}^N x_i}{N}$ . Tomando en cuenta a la fórmula (4.1) como función de  $\theta$  tenemos que

$$l(\lambda, \theta|\mathbf{x}) \propto \exp(N\lambda\theta) I_{(-\infty, y_1)}(\theta).$$

Como la función exponencial es creciente tenemos que  $\hat{\theta} = y_1$  es el valor máximo para  $l(\lambda, \theta|\mathbf{x})$  y por tanto es el estimador máximo verosimil para  $\theta$ .

Por otro lado, tomando en cuenta de nuevo la fórmula (4.1) como función de  $\lambda$ , la función de log-verosimilitud es

$$L(\lambda, \theta|\mathbf{x}) = N \log(\lambda) - N\lambda(\bar{x} - \theta). \quad (4.2)$$

El valor máximo local  $\hat{\lambda}$  de esta función satisface que

$$\left[ \frac{\partial L(\lambda, \theta|\mathbf{x})}{\partial \lambda} \right]_{\lambda=\hat{\lambda}} = \frac{N}{\hat{\lambda}} - N(\bar{x} - \theta) = 0. \quad (4.3)$$

De la ecuación (4.3) se tiene que

$$\hat{\lambda} = \frac{N}{N(\bar{x} - \theta)} = \frac{1}{\bar{x} - \theta}.$$

Para que  $\hat{\lambda}$  sea el máximo global de la función (4.2) debe cumplir que

$$\left[ \frac{\partial^2 L(\lambda, \theta | \mathbf{x})}{\partial \lambda^2} \right]_{\lambda=\hat{\lambda}} = -N(\bar{x} - \theta)^2$$

sea negativo, lo cual siempre es cierto. Por otra parte, suponiendo que los valores  $x_1, x_2, \dots, x_N$  no son todos iguales se tiene que

$$y_1 < x_i$$

para toda  $i \in \{1, 2, \dots, N\}$ . Entonces

$$Ny_1 < \sum_{i=1}^N x_i$$

y consecuentemente

$$y_1 < \bar{x}.$$

Así,

$$\hat{\theta} = y_1 \quad \text{y} \quad \hat{\lambda} = \frac{1}{\bar{x} - y_1}$$

son los estimadores máximo verosímil para  $\theta$  y  $\lambda$  respectivamente.

## 4.2. Distribución a priori

En esta sección se presentará la definición de la distribución a priori y se explicará porqué es necesario considerar dichas distribuciones al hacer inferencias.

**Definición 4.9** Sean  $\mathbf{x} = (x_1, x_2, \dots, x_N)$  una muestra aleatoria,  $\theta \in \Theta$  el parámetro desconocido del modelo que describe el experimento que produjo la muestra  $\mathbf{x}$ . La **distribución a priori** de  $\theta$  es la distribución del parámetro  $\theta$  que se le asigna sin tener en cuenta la muestra obtenida del experimento y se denotará por  $P_{\text{priori}}(\theta)$ .



**Observación.** Las distribuciones a priori son especificadas por intuición, por medio de estudios previos o de opiniones de expertos en el área, pero si no se tiene intuición o información alguna generalmente se usa una distribución aleatoria uniforme sobre un espacio muestral suficientemente grande.

Para ejemplificar la necesidad de tomar en cuenta la información previa a la hora de hacer inferencias, considere lo siguiente:

- a) Una mujer ebria afirma que es capaz de saber, con un sólo trago, la manera en que un café con leche fue preparado: si la leche fue agregada antes o después del café. Imagine que la mujer acertó en diez tazas de dicha bebida.
- b) Un músico afirma que puede saber si una página de una obra musical pertenece a una obra de Hayden o a una de Mozart. Imagine que el músico clasificó correctamente diez páginas.

Considerando las variables aleatorias

$X = \text{número de aciertos de la mujer y}$

$Y = \text{número de páginas clasificadas correctamente por el músico}$

se tiene que

$$P(X|10, \theta) \quad \text{y} \quad P(Y|10, \theta)$$

se distribuyen Binomial(10,  $\theta$ ) donde  $\theta$  es la probabilidad de éxito en cada experimento. Como la mujer y el músico acertaron las diez veces que intentaron sus experimentos se infiere que

$$\theta > \frac{1}{2}$$

para los dos casos, aunque el grado de creencia para la mujer ebria es muy bajo y para el músico es medianamente alto. Así, tomando en cuenta la información personal y los datos observados a la hora de inferir se evitarán tales subjetividades.

A pesar de lo anterior, en algunas ocasiones será difícil encontrar una distribución a priori para el parámetro desconocido basándose solamente en la muestra, ya sea porque la información sobre él no es muy confiable o porque no es suficiente. En otras ocasiones se deseará hacer inferencias sólo tomando en cuenta la muestra, para “no contaminar” los datos. En estos casos es necesario encontrar una distribución a priori que aporte poco, es decir, que sea no informativa.

**Definición 4.10** Sean  $\mathbf{x} = (x_1, x_2, \dots, x_N)$  una muestra aleatoria y  $\theta \in \Theta$  el parámetro del modelo que produjo la muestra  $\mathbf{x}$ . Se dice que una distribución a priori es **no informativa** para  $\theta$  si no favorece a algún valor de  $\theta$  sobre otros.

**Observación.** En general, las distribuciones no informativas son **impropias**, es decir,

$$\sum_{\theta \in \Theta} P_{\text{priori}}(\theta) = \infty$$

cuando  $\Theta$  es discreto, o

$$\int_{\Theta} P_{\text{priori}}(\theta) d\theta = \infty$$

cuando  $\Theta$  es continuo. Sin embargo, muchas veces son tomadas distribuciones uniformes definidas en un conjunto suficientemente grande.

### 4.3. Distribución a posteriori

En esta sección se definirán, entre otras cosas, las distribuciones a posteriori, se explicará en qué consiste el método de inferencia Bayesiano y se presentarán algunos ejemplos acerca de dicho método.

**Definición 4.11** Sean  $\mathbf{x} = (x_1, x_2, \dots, x_N)$  una muestra aleatoria,  $\theta \in \Theta$  el parámetro del modelo que produjo la muestra  $\mathbf{x}$ ,  $P_{\text{priori}}(\theta)$  la función a priori de  $\theta$  y  $l(\theta|\mathbf{x})$  la función de verosimilitud. La **distribución a posteriori** es la distribución del parámetro  $\theta$  cuando se toma en cuenta la información producida por la muestra obtenida, es decir, es  $P(\theta|\mathbf{x})$ .

**Observación.** A veces las distribuciones a priori dependen de otros parámetros que pueden ser conocidos o desconocidos y que son llamados **hiperparámetros**. Cuando son desconocidos pueden tratarse de dos formas:

- i) El método propiamente Bayesiano consiste en asignarle al hiperparámetro una distribución a priori (también llamada **hiperpriori**) y a partir de ella encontrar la distribución a posteriori. Por otro lado, la distribución hiperpriori podría depender de una colección de parámetros desconocidos. A la especificación de un modelo sobre varios niveles se le llama **modelación jerárquica**, en donde cada nueva distribución forma un nivel de la jerarquía.
- ii) El segundo método, llamado **análisis empírico de Bayes**, consiste en encontrar un valor de  $\eta$  que maximice la distribución marginal de  $P(\mathbf{x}|\eta)$  pero vista como función de  $\eta$ . Si  $\eta = \hat{\eta}$  fuera dicho valor entonces la función a posteriori que se trabajará será  $P(\theta|\mathbf{x}, \hat{\eta})$ . La expresión para  $P(\theta|\mathbf{x}, \eta)$  puede obtenerse sustituyendo  $\hat{\eta}$  en la expresión (4.4) abajo.

Esto es visto más formalmente en el siguiente teorema.

**Teorema 4.1** Sean  $\mathbf{x} = (x_1, x_2, \dots, x_N)$  una muestra de un experimento aleatorio,  $\theta \in \Theta$  el parámetro del modelo que produjo  $\mathbf{x}$  y  $P_{\text{priori}}(\theta, \eta)$  la distribución a priori que depende de un hiperparámetro  $\eta$ .

a) Si el valor de  $\eta$  es conocido la distribución a posteriori es

$$P(\theta|\mathbf{x}, \eta) = \frac{P(\mathbf{x}|\theta)P_{\text{priori}}(\theta|\eta)}{\sum_{\gamma \in \Theta} P(\mathbf{x}|\gamma)P_{\text{priori}}(\gamma|\eta)}. \quad (4.4)$$

b) Si el valor de  $\eta$  es desconocido la distribución a posteriori es

$$P(\theta|\mathbf{x}) = \frac{\sum_{\eta} P(\mathbf{x}|\theta)P_{\text{priori}}(\theta|\eta)P_{\text{priori}}(\eta)}{\sum_{\eta} \sum_{\gamma \in \Theta} P(\mathbf{x}|\gamma)P_{\text{priori}}(\gamma|\eta)P_{\text{priori}}(\eta)}. \quad (4.5)$$

**Demostración.**

a) Note que

$$\begin{aligned}
P(\theta|\mathbf{x}, \eta) &= \frac{P(\theta, \mathbf{x}, \eta)}{P(\mathbf{x}, \eta)} \\
&= \frac{P(\mathbf{x}|\theta, \eta)P(\theta, \eta)}{P(\mathbf{x}, \eta)} \frac{P(\eta)}{P(\eta)} \\
&= \frac{P(\mathbf{x}|\theta, \eta)P(\theta|\eta)}{P(\mathbf{x}|\eta)} \\
&= \frac{P(\mathbf{x}|\theta, \eta)P(\theta|\eta)}{\sum_{\gamma \in \Theta} P(\mathbf{x}, \gamma|\eta)} \\
&= \frac{P(\mathbf{x}|\theta)P(\theta|\eta)}{\sum_{\gamma \in \Theta} P(\mathbf{x}|\gamma, \eta)P(\gamma|\eta)} \\
&= \frac{P(\mathbf{x}|\theta)P_{\text{priori}}(\theta|\eta)}{\sum_{\gamma \in \Theta} P(\mathbf{x}|\gamma)P_{\text{priori}}(\gamma|\eta)}.
\end{aligned}$$

La quinta y sexta proposición se tienen porque la muestra depende del hiperparámetro  $\eta$  sólo a través del parámetro  $\theta$ .

b) Ya que  $\eta$  es desconocido, también se le asignará una distribución a priori  $P_{\text{priori}}(\eta)$ .

$$\begin{aligned}
P(\theta|\mathbf{x}) &= \frac{P(\theta, \mathbf{x})}{P(\mathbf{x})} \\
&= \frac{\sum_{\eta} P(\theta, \mathbf{x}, \eta)}{\sum_{\gamma \in \Theta} P(\mathbf{x}, \gamma)} \\
&= \frac{\sum_{\eta} P(\mathbf{x}|\theta, \eta)P(\theta, \eta)}{\sum_{\eta} \sum_{\gamma \in \Theta} P(\mathbf{x}, \gamma, \eta)} \\
&= \frac{\sum_{\eta} P(\mathbf{x}|\theta, \eta)P(\theta|\eta)P(\eta)}{\sum_{\eta} \sum_{\gamma \in \Theta} P(\mathbf{x}|\gamma, \eta)P(\gamma, \eta)} \\
&= \frac{\sum_{\eta} P(\mathbf{x}|\theta)P_{\text{priori}}(\theta|\eta)P_{\text{priori}}(\eta)}{\sum_{\eta} \sum_{\gamma \in \Theta} P(\mathbf{x}|\gamma)P_{\text{priori}}(\gamma|\eta)P_{\text{priori}}(\eta)}.
\end{aligned}$$

La quinta igualdad se tiene porque  $\mathbf{x}$  sólo depende de  $\eta$  a través de  $\theta$ .

**Observación.** Cuando  $\eta$  es conocido no es necesario escribirlo en la distribución a posteriori, y como el denominador en la expresión (4.4) ya no depende de  $\theta$  entonces ésta se puede simplificar y obtener la expresión

$$P(\theta|\mathbf{x}) \propto l(\theta|\mathbf{x})P_{\text{priori}}(\theta) \quad (4.6)$$

### 4.3. Distribución a posteriori

---

donde  $l(\theta|\mathbf{x})$  es la función de verosimilitud del modelo y la constante de proporcionalidad, que sólo depende de  $\mathbf{x}$ , hace que  $P(\theta|\mathbf{x})$  integre o sume uno.

Note entonces que en la estadística Bayesiana se trabaja con la probabilidad de una hipótesis con respecto al parámetro que se desea estimar tomando en cuenta:

- i) las creencias que se tienen acerca del parámetro sin tomar en cuenta los datos obtenidos de una muestra (distribución a priori); y
- ii) la función de verosimilitud para obtener la distribución a posteriori, que es una actualización, hecha con base en el Teorema de Bayes <sup>1</sup>, de la información aportada por la distribución a priori.

En cambio, en la estadística clásica se trabaja directamente con los datos obtenidos de una muestra y las probabilidades son usadas en el sentido de frecuencias de ocurrencia de un evento determinado.

Por otro lado, aunque las distribuciones a priori sean impropias, al combinarlas con la información muestral producen distribuciones a posteriori propias, incluso su uso como representación de ignorancia o poca información es bien aceptado actualmente.

**Ejemplo 4.2** Sean  $\mathbf{x} = (x_1, x_2, \dots, x_N)$  una muestra aleatoria y  $\theta \in \Theta$  el parámetro del modelo que describe el experimento que produjo  $\mathbf{x}$ , donde  $\Theta = (0, \infty)$  y sea  $l(\theta|\mathbf{x})$  la función de verosimilitud. Una de las funciones que no favorece a algún valor de  $\theta$  sobre otro es una función constante. Así, suponga que

$$P_{priori}(\theta) = c$$

donde  $c > 0$ , aunque ésta es impropia ya que

$$\int_0^{\infty} P_{priori}(\theta) d\theta = \infty.$$

---

<sup>1</sup>Ver Apéndice A, Teorema A.1.

De aquí se tiene que la función a posteriori es

$$\begin{aligned}
 P(\theta|\mathbf{x}) &= \frac{cl(\theta|\mathbf{x})}{\int cl(\theta|\mathbf{x})d\theta} \\
 &= \frac{l(\theta|\mathbf{x})}{\int l(\theta|\mathbf{x})d\theta} \\
 &= \frac{l(\theta|\mathbf{x})}{K}
 \end{aligned} \tag{4.7}$$

donde  $K$  es el valor (finito) de la integral con respecto a  $\theta$  de  $l(\theta|\mathbf{x})$ . Note que la integral con respecto  $\theta$  de la fórmula (4.7) es igual a uno, por lo que sí se puede considerar como una función de distribución.

**Ejemplo 4.3** Sea  $\mathbf{x} = (x_1, x_2, \dots, x_N)$  una muestra aleatoria donde cada  $x_i$  es producida por la variable aleatoria  $X_i$  con  $i = 1, 2, \dots, N$  que tiene distribución Normal( $\theta, \sigma^2$ ) con media  $\theta \in \mathbb{R}$  desconocida y varianza  $\sigma^2 > 0$  conocida. En este caso, la función de verosimilitud es

$$\begin{aligned}
 l(\theta|\mathbf{x}) &= \prod_{i=1}^N \left[ (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \left(\frac{x_i - \theta}{\sigma}\right)^2\right) \right] \\
 &= (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left(-\frac{1}{2} \frac{\sum_{i=1}^N (x_i - \theta)^2}{\sigma^2}\right).
 \end{aligned} \tag{4.8}$$

Suponga que la distribución a priori de  $\theta$  es

$$P_{priori}(\theta|\mu, \tau) = (2\pi\tau^2)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \left(\frac{\theta - \mu}{\tau}\right)^2\right) \tag{4.9}$$

donde  $\mu \in \mathbb{R}$  y  $\tau > 0$  son hiperparámetros conocidos. Tome  $\eta = (\mu, \tau)$ . Finalmente, susti-

tuyendo las fórmulas (4.8) y (4.9) en la expresión (4.6) se obtiene

$$\begin{aligned}
 P(\theta|\mathbf{x}, \eta) &\propto \exp\left(-\frac{1}{2} \frac{\sum_{i=1}^N (x_i - \theta)^2}{\sigma^2}\right) \exp\left(-\frac{1}{2} \left(\frac{\theta - \mu}{\tau}\right)^2\right) \\
 &= \exp\left\{-\frac{1}{2} \left(\frac{\sum_{i=1}^N x_i^2 - 2N\bar{x}\theta + N\theta^2}{\sigma^2} + \frac{\theta^2 - 2\theta\mu + \mu^2}{\tau^2}\right)\right\} \\
 &\propto \exp\left\{-\frac{1}{2} \left(\frac{-2N\bar{x}\theta + N\theta^2}{\sigma^2} + \frac{\theta^2 - 2\theta\mu}{\tau^2}\right)\right\} \\
 &= \exp\left\{-\frac{1}{2} \left(-2\theta \left(\frac{N\bar{x}}{\sigma^2} + \frac{\mu}{\tau^2}\right) + \theta^2 \left(\frac{N}{\sigma^2} + \frac{1}{\tau^2}\right)\right)\right\} \\
 &= \exp\left\{-\frac{1}{2} \left(\frac{N}{\sigma^2} + \frac{1}{\tau^2}\right) \left(\theta^2 - 2\theta \frac{\frac{N\bar{x}}{\sigma^2} + \frac{\mu}{\tau^2}}{\frac{N}{\sigma^2} + \frac{1}{\tau^2}}\right)\right\} \\
 &\propto \exp\left\{-\frac{1}{2} \left(\frac{N}{\sigma^2} + \frac{1}{\tau^2}\right) \left(\theta - \frac{\frac{N\bar{x}}{\sigma^2} + \frac{\mu}{\tau^2}}{\frac{N}{\sigma^2} + \frac{1}{\tau^2}}\right)^2\right\}
 \end{aligned}$$

por lo que  $\theta$  se distribuye a posteriori  $\text{Normal}(\alpha, \beta)$  donde

$$\alpha = \frac{\frac{N\bar{x}}{\sigma^2} + \frac{\mu}{\tau^2}}{\frac{N}{\sigma^2} + \frac{1}{\tau^2}} = \frac{\tau^2 N\bar{x} + \sigma^2 \mu}{\tau^2 N + \sigma^2} \quad \text{y} \quad \beta = \frac{1}{\frac{N}{\sigma^2} + \frac{1}{\tau^2}} = \frac{\sigma^2 \tau^2}{\tau^2 N + \sigma^2}.$$

Note que, en efecto, la muestra depende de  $\eta$  sólo a través de  $\theta$ , ya que en este caso  $X$  se distribuye  $\text{Normal}(\theta, \sigma^2)$  y  $\theta$  se distribuye  $\text{Normal}(\alpha, \beta)$  donde  $\alpha$  y  $\beta$  son funciones de  $\eta = (\mu, \tau)$ .

**Definición 4.12** Sean  $\mathbf{x} = (x_1, x_2, \dots, x_N)$  una muestra aleatoria,  $\theta \in \Theta$  el parámetro del modelo que produjo  $\mathbf{x}$ ,  $l(\theta|\mathbf{x})$  la función de verosimilitud y  $\mathcal{F}_{\text{priori}}$  una familia de distribuciones a priori. Se dice que  $\mathcal{F}_{\text{priori}}$  es una **familia conjugada** para una función de verosimilitud si es tal que la distribución a posteriori

$$P(\theta|\mathbf{x}) \propto l(\theta|\mathbf{x})P_{\text{priori}}(\theta)$$

pertenece a  $\mathcal{F}_{\text{priori}}$  para todo valor de  $\mathbf{x}$  cuando  $P_{\text{priori}}(\theta)$  también pertenece a  $\mathcal{F}_{\text{priori}}$ .

**Proposición 4.1** Sean  $\mathbf{x} = (x_1, x_2, \dots, x_N)$  una muestra aleatoria,  $\theta \in \Theta$  el parámetro del modelo que produjo  $\mathbf{x}$ ,  $P_{\text{priori}}^{(1)}(\theta)$ ,  $P_{\text{priori}}^{(2)}(\theta)$  funciones a priori de una familia conjugada y

$P^{(1)}(\theta|\mathbf{x})$ ,  $P^{(2)}(\theta|\mathbf{x})$  funciones a posteriori obtenidas de las respectivas funciones a priori.

Si

$$P_{\text{priori}}(\theta) = a P_{\text{priori}}^{(1)}(\theta) + b P_{\text{priori}}^{(2)}(\theta)$$

donde  $a + b = 1$  entonces

$$P(\theta|\mathbf{x}) \propto a P^{(1)}(\theta|\mathbf{x}) + b P^{(2)}(\theta|\mathbf{x}).$$

**Demostración.**

$$\begin{aligned} P(\theta|\mathbf{x}) &\propto l(\theta|\mathbf{x})P_{\text{priori}}(\theta) \\ &= l(\theta|\mathbf{x}) \left( a P_{\text{priori}}^{(1)}(\theta) + b P_{\text{priori}}^{(2)}(\theta) \right) \\ &= a l(\theta|\mathbf{x})P_{\text{priori}}^{(1)}(\theta) + b l(\theta|\mathbf{x})P_{\text{priori}}^{(2)}(\theta) \\ &\propto a P^{(1)}(\theta|\mathbf{x}) + b P^{(2)}(\theta|\mathbf{x}). \end{aligned}$$

La primera y última proposición se tienen por la observación del Teorema 4.1 y la segunda igualdad se tiene por hipótesis.

**Ejemplo 4.4** Sea  $\mathbf{x} = (x_1, x_2, \dots, x_N)$  una muestra aleatoria producida por variables aleatorias  $X_i$  con  $i = 1, 2, \dots, N$  que tienen distribución Poisson con parámetro  $\theta \in \mathbb{R}$  desconocido.

La función de verosimilitud para éste caso es

$$\begin{aligned} l(\theta|\mathbf{x}) &= \prod_{i=1}^N \frac{\exp(-\theta)\theta^{x_i}}{x_i!} \\ &= \frac{\exp(-N\theta)\theta^{\sum_{i=1}^N x_i}}{\prod_{i=1}^N x_i!}. \end{aligned} \tag{4.10}$$

Suponiendo que la distribución a priori es

$$P_{\text{priori}}(\theta) \propto \exp\left(-\frac{\beta}{2}\theta\right) \theta^{\frac{\alpha}{2}-1} \tag{4.11}$$

donde  $\alpha$  y  $\beta$  son constantes adecuadas y conocidas se tiene que  $\theta$  se distribuye a priori Gama( $\alpha_1, \beta_1$ ) donde  $\alpha_1 = \frac{\alpha}{2}$  y  $\beta_1 = -\frac{\beta}{2}$ . Haciendo  $T = \sum_{i=1}^N x_i$  en la función de verosimi-



litud (4.10), la función a posteriori resulta ser

$$\begin{aligned}
 P(\theta|\mathbf{x}) &\propto l(\theta|\mathbf{x})P_{\text{priori}}(\theta) \\
 &\propto (\exp(-N\theta)\theta^T) \left( \exp\left(-\frac{\beta}{2}\theta\right) \theta^{\frac{\alpha}{2}-1} \right) \\
 &\propto \exp\left(-\frac{\theta(2N+\beta)}{2}\right) \theta^{\frac{2T+\alpha}{2}-1}.
 \end{aligned} \tag{4.12}$$

De aquí tenemos que  $\theta$  se distribuye a posteriori  $\text{Gama}(\alpha_2, \beta_2)$  donde  $\alpha_2 = \frac{2T+\alpha}{2}$  y  $\beta_2 = -\frac{2N+\beta}{2}$ . Sea  $\mathcal{F}_{\text{gama}}$  la familia de distribuciones a priori de tipo  $\text{Gama}(\alpha, \beta)$  con  $\alpha, \beta > 0$ . Como las distribuciones (4.11) y (4.12) pertenecen a  $\mathcal{F}_{\text{gama}}$ , ésta es una familia conjugada para la función de verosimilitud (4.10). Note además que aunque las observaciones son discretas ellas dependen de un parámetro continuo, por lo que los elementos de la familia  $\mathcal{F}_{\text{gama}}$  son distribuciones continuas.

## 4.4. Estimación puntual

En esta sección se presentarán algunos resultados relacionados con las estadísticas suficientes. También se mostrará la utilidad de las familias exponencial y conjugada en el método de inferencia Bayesiano.

**Ejemplo 4.5** Considere  $\mathbf{x} = (x_1, x_2, \dots, x_N)$  una muestra aleatoria distribuida  $\text{Normal}(\theta, \sigma)$ , pero donde  $\theta \in \mathbb{R}$  es conocida y  $\sigma > 0$  es desconocida. Para este caso la función de verosimilitud  $l(\sigma|\mathbf{x})$  es la misma que la fórmula (4.8) pero vista como función de  $\sigma$ . Haciendo  $T = \sum_{i=1}^N (x_i - \theta)^2$  se tiene que

$$l(\sigma|\mathbf{x}) = (2\pi\sigma)^{-\frac{N}{2}} \exp\left(-\frac{1}{2} \frac{T}{\sigma}\right).$$

Ahora se debe proponer una distribución a priori para  $\sigma$ . Sea

$$P_{\text{priori}}(\sigma) = (2\pi\sigma)^{-\frac{m}{2}} \exp\left(-\frac{1}{2} \frac{\mu}{\sigma}\right)$$

donde  $m$  y  $\mu$  son constantes adecuadas. Conviene que la distribución a priori sea de esta manera para que la distribución a posteriori sea fácil de manejar, e incluso haciendo  $m = r-2$

las cosas se facilitarán más. Así, la distribución a posteriori queda como

$$\begin{aligned} P(\sigma|\mathbf{x}) &\propto l(\sigma|\mathbf{x})P_{\text{priori}}(\sigma) \\ &\propto \left( (2\pi\sigma)^{-\frac{N}{2}} \exp\left(-\frac{1}{2}\frac{T}{\sigma}\right) \right) \left( (2\pi\sigma)^{-\frac{(r-2)}{2}} \exp\left(-\frac{1}{2}\frac{\mu}{\sigma}\right) \right) \\ &\propto (\sigma)^{-\frac{N+r}{2}+1} \exp\left(-\frac{1}{2}\frac{T+\mu}{\sigma}\right) \end{aligned}$$

pero haciendo el cambio de variable  $\lambda = \frac{1}{\sigma}$  se tiene que  $\sigma = \lambda^{-1}$  y entonces

$$\begin{aligned} P(\lambda|\mathbf{x}) &\propto P(\sigma|\mathbf{x}) \left| \frac{\partial\sigma}{\partial\lambda} \right| \\ &\propto \left( \lambda^{\frac{N+r}{2}-1} \exp\left(-\frac{\lambda}{2}(T+\mu)\right) \right) (\lambda^{-2}) \\ &\propto \lambda^{\frac{N+r}{2}-1} \exp\left(-\frac{\lambda(T+\mu)}{2}\right) \end{aligned}$$

por lo que  $(T+\mu)\lambda$  se distribuye  $\chi_{(N+r)}^2$  y  $1/\sigma$  se distribuye  $(T+\mu)\chi_{(N+r)}^2$ . Note que el estimador encontrado depende de la muestra sólo a través de  $T$ . Igual que en este caso, a veces es posible resumir los datos de la muestra en una o dos cantidades, lo cual justifica la siguiente definición.

**Definición 4.13** Sean  $\mathbf{x} = (x_1, x_2, \dots, x_N)$  una muestra aleatoria y  $\theta \in \Theta$  el parámetro del modelo que generó la muestra  $\mathbf{x}$ . La estadística  $T(\mathbf{x})$  es **suficiente** para  $\theta$  dado  $\mathbf{x}$  si y sólo si la densidad condicional de la muestra dado la estadística no depende del parámetro, es decir, si

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

donde  $t = T(\mathbf{x})$ .

#### **Teorema 4.2 (Factorización de Neyman)**

Sean  $\mathbf{x} = (x_1, x_2, \dots, x_N)$  una muestra aleatoria y  $\theta \in \Theta$  el parámetro del modelo que generó la muestra  $\mathbf{x}$ . La estadística  $T(\mathbf{x})$  es suficiente para  $\theta$  dado  $\mathbf{x}$  si y sólo si existen funciones  $g$  y  $h$  tales que

$$P(\mathbf{x}|\theta) = g(t, \theta)h(\mathbf{x})$$

---

<sup>2</sup>Ver Apéndice B, Definición B.8.

donde  $t = T(\mathbf{x})$ ,  $g$  es no negativa y sólo depende de la muestra a través de la estadística, y  $h$  es también no negativa y sólo depende de la muestra.

**Demostración.** Ver Lee (1989).

**Proposición 4.2 (Principio de Suficiencia)**

Sean  $\mathbf{x} = (x_1, x_2, \dots, x_N)$  una muestra aleatoria y  $\theta \in \Theta$  el parámetro del modelo que generó la muestra  $\mathbf{x}$ . La estadística  $T(\mathbf{x})$  es suficiente para  $\theta$  dado  $\mathbf{x}$  si y sólo si

$$l(\theta|\mathbf{x}) \propto l(\theta|t)$$

donde  $t = T(\mathbf{x})$  y la constante de proporcionalidad no depende de  $\theta$ .

**Demostración.** Ver Lee (1989).

**Observación.** En otras palabras, el Principio de Suficiencia asegura que si existe una estadística suficiente para  $\theta$ , entonces en ella está contenida toda la información sobre  $\theta$  proporcionada por la muestra.

**Proposición 4.3** Sean  $\mathbf{x} = (x_1, x_2, \dots, x_N)$  una muestra aleatoria y  $\theta \in \Theta$  el parámetro del modelo que generó la muestra  $\mathbf{x}$ . Para cualquier distribución a priori, la distribución a posteriori de  $\theta$  dado  $\mathbf{x}$  es la misma que la distribución a posteriori de  $\theta$  dado una estadística suficiente, es decir,

$$P(\theta|\mathbf{x}) = P(\theta|t)$$

donde  $t = T(\mathbf{x})$  es una estadística suficiente.

**Demostración.** Ver Carlin y Louis (2000).

**Ejemplo 4.6** Para ejemplificar el uso del Teorema 4.2, considere la función de verosimilitud (4.10) del Ejemplo 4.4. Haciendo

$$g(T, \theta) = \exp(-N\theta)\theta^T$$

donde  $T = \sum_{i=1}^N x_i$  y

$$h(\mathbf{x}) = \left( \prod_{i=1}^N x_i! \right)^{-1}$$

se tiene que  $T$  es una estadística suficiente para  $\theta$  dado  $\mathbf{x}$ .

Por otro lado, para encontrar el estimador máximo verosímil para  $\theta$  se debe considerar la función de log-verosimilitud de (4.10) que es

$$\begin{aligned} L(\theta|\mathbf{x}) &= \log l(\theta|\mathbf{x}) \\ &\propto \log(\exp(-N\theta)\theta^{\sum_{i=1}^N x_i}) \\ &\propto -N\theta - \sum_{i=1}^N x_i \log(\theta). \end{aligned} \quad (4.13)$$

Para que  $\theta = \hat{\theta}$  sea un máximo local de la función (4.13) debe satisfacer que

$$\left[ \frac{\partial \log(l(\theta|\mathbf{x}))}{\partial \theta} \right]_{\theta=\hat{\theta}} = 0$$

es decir,

$$-N + \frac{\sum_{i=1}^N x_i}{\hat{\theta}} = 0.$$

De aquí se tiene que

$$\hat{\theta} = \frac{\sum_{i=1}^N x_i}{N} = \bar{x}.$$

Para que este valor sea un máximo global de la función (4.13) debe satisfacer que

$$\left[ \frac{\partial^2 \log(l(\theta|\mathbf{x}))}{\partial \theta^2} \right]_{\theta=\hat{\theta}} < 0 \quad (4.14)$$

para todo  $\mathbf{x}$ . El lado izquierdo de la expresión (4.14) es igual a

$$-\frac{\sum_{i=1}^N x_i}{\hat{\theta}^2}$$

y siempre es negativo ya que  $x_i \in \mathbb{N}$ . Por tanto, según la Definición 4.8,

$$\hat{\theta} = \bar{x}$$

es el estimador máximo verosímil para  $\theta$ .

**Definición 4.14** Sean  $\mathbf{x} = (x_1, x_2, \dots, x_N)$  una muestra aleatoria.  $T(\mathbf{x})$  es una estadística minimal y suficiente si puede escribirse como una función de cualquier otra estadística suficiente.

**Definición 4.15** Sean  $\mathbf{x} = (x_1, x_2, \dots, x_N)$  una muestra aleatoria y  $\theta \in \Theta$  el parámetro del modelo que generó la muestra  $\mathbf{x}$ . Una función de densidad es miembro de la **familia exponencial** de  $\theta$  si puede escribirse como

$$P(\mathbf{x}|\theta) = a(\mathbf{x})b(\theta) \exp [c(\theta)d(\mathbf{x})].$$

**Observación.** Sean  $\mathbf{x} = (x_1, x_2, \dots, x_N)$  una muestra aleatoria,  $\theta \in \Theta$  el parámetro del modelo que generó la muestra  $\mathbf{x}$  y supongamos que la función de verosimilitud  $l(\theta|\mathbf{x})$  pertenece a la familia exponencial de  $\theta$ , es decir

$$\begin{aligned} l(\theta|\mathbf{x}) &\propto \prod_{i=1}^N f(x_i|\theta) \\ &= \prod_{i=1}^N a(x_i)b(\theta) \exp [c(\theta)d(x_i)] \\ &= (b(\theta))^N \left[ \prod_{i=1}^N a(x_i) \right] \exp \left( c(\theta) \sum_{i=1}^N d(x_i) \right), \end{aligned}$$

o equivalentemente

$$l(\theta|\mathbf{x}) \propto (b(\theta))^N \exp \left( c(\theta) \sum_{i=1}^N d(x_i) \right) \quad (4.15)$$

y aplicando el Teorema de Factorización de Neyman (Teorema 4.2) se tiene que  $T(\mathbf{x}) = \sum_{i=1}^N d(x_i)$  es una estadística suficiente para  $\theta$  dado  $\mathbf{x}$ . Algunas de las distribuciones que pertenecen a la familia exponencial son la binomial, la geométrica, la Poisson, la Gama y la Beta.

Por otro lado, el uso de familias conjugadas facilita los cálculos para encontrar la distribución a posteriori, suponiendo que se pueden ajustar a la información inicial. Si la función de verosimilitud pertenece a la familia exponencial entonces se puede construir fácilmente una distribución a posteriori que pertenezca a una familia conjugada. Considere un modelo con función de verosimilitud (4.15) y con distribución a priori

$$P_{\text{priori}}(\theta) \propto (b(\theta))^{N_0} \exp[d_0 c(\theta)].$$

De aquí la distribución a posteriori resulta

$$P(\theta|\mathbf{x}) \propto (b(\theta))^{N+N_0} \exp \left[ c(\theta) \left( d_0 + \sum_{i=1}^N d(x_i) \right) \right]$$

por lo que ésta distribución pertenece a la misma familia que la distribución a priori.

Finalmente, en el siguiente cuadro se presentan algunas distribuciones a priori y posteriori pertenecientes a una familia conjugada para una función de verosimilitud.

Función de verosimilitud	Distribución a priori	Distribución a posteriori
Binomial( $\mathbf{x} n, \theta$ )	Beta( $\theta \alpha, \beta$ )	Beta( $\theta \alpha + \mathbf{x}, \beta + n - \mathbf{x}$ )
Poisson( $\mathbf{x} \theta$ )	Gama( $\theta \alpha, \beta$ )	Gama( $\theta \alpha + \sum_i x_i, \beta + n$ )
Normal( $\mathbf{x} \mu, \sigma$ )*	Normal( $\mu \theta, \eta$ )	Normal( $\mu \frac{\eta\theta+n\sigma\bar{x}}{\eta+n\sigma}, \eta + n\sigma$ )
Gama( $\mathbf{x} \alpha, \beta$ )**	Gama( $\beta \theta, \eta$ )	Gama( $\beta \theta + n\alpha, \eta + \sum_i x_i$ )

\*  $\sigma$  conocido, \*\*  $\alpha$  conocido



## Capítulo 5

# Aplicación del algoritmo

# Metropolis-Hastings al estudio del VPH

Existen más de 100 tipos diferentes de cepas del Virus del Papiloma Humano (VPH) y aproximadamente 40 de ellas son transmitidas sexualmente. La mayoría de las infecciones con VPH avanzan y se resuelven sin producir síntomas, aunque algunos tipos del virus sí los producen. Como ejemplo se pueden citar las cepas tipo 6 y 11 que son causantes de las verrugas genitales, y las cepas tipo 16 y 18 que son asociadas con ciertos tipos de cáncer. De hecho, las infecciones con VPH son reconocidas como la mayor causa del cáncer cérvico ya que más del 90 por ciento de las mujeres que han tenido este cáncer también han sido infectadas con VPH, especialmente con las cepas 16 y 18 (ver Sanders *et al.* (2003), Myers *et al.* (2000), Gálvez-Correa *et al.* (2006)). A pesar de que no existen métodos comunes para eliminar la infección del VPH, las lesiones precancerosas y las verrugas causadas por estos virus sí pueden ser tratadas.



A continuación se mostrará el uso de las cadenas de Markov para modelar la historia natural del VPH, su transmisión y su evolución hasta el cáncer cérvico. Se presentará solamente la parte teórica del modelo, sin incluir simulación y análisis con datos reales.

## 5.1. Modelo de la historia natural del VPH

En esta sección se presentará un modelo que simula la historia natural de la infección del virus del papiloma humano y el cáncer cervicouterino basado en el estudio simulado realizado por Goldie *et al.* (2003) donde una población de 100,000 niñas de 13 años entraron al modelo. En este estudio se simuló el comportamiento de esta población con respecto a la infección, así como el posible desarrollo del cáncer cérvico.

El tiempo total que se consideró en la simulación fue la vida de una mujer, desde los 12 hasta los 65 años. Los ciclos para la cadena de Markov fueron periodos de 6 meses, es decir, un paso de la cadena de Markov corresponde a 6 meses de la vida de una mujer. Así, en cada periodo de tiempo cada niña tenía el riesgo de contagiarse de algún tipo de VPH (transitorio; tipo 16 y 18; diferente al 16 o 18; de bajo riesgo respecto al cáncer cérvico).

Las niñas infectadas con VPH podrían tratarse para combatir la infección causada por el virus, o podrían progresar en la infección desarrollando lesiones epiteliales como la Neoplasia Cérvica Intraepitelial tipo 1 (CIN 1) <sup>1</sup> y que a su vez podría progresar a Neoplasia Cérvica Intraepitelial tipo 2-3 (CIN 2-3) o podría regresar a la pura infección. Finalmente, de CIN 2-3 se puede regresar a CIN 1 o progresar al *Carcinoma in Situ* (CIS), del cual se puede progresar al cáncer invasivo <sup>2</sup> y de aquí se pueden desarrollar síntomas o progresar a la siguiente

---

<sup>1</sup>La Neoplasia Cérvica Intraepitelial se refiere a la presencia de células anormales en la superficie del cuello del útero. Comúnmente no existen síntomas visibles de neoplasia, los cambios celulares se detectan mediante pruebas de Papanicolau o colposcopia.

<sup>2</sup>El cáncer cérvico invasivo es causado por el crecimiento de células cancerosas en el cuello uterino (también llamado cervix) el cual conecta al útero con la vagina. Una masa de tejido extra llamada tumor crece cuando las células del cervix se reproducen sin control aunque no se necesiten células nuevas. El término cáncer

etapa del cáncer (se consideran 4 etapas). También se supuso que las mujeres con cáncer invasivo recibieron tratamiento específico a la etapa de la enfermedad en la que se encontraran.

De esta forma, la historia natural de la infección con VPH puede modelarse a través de una cadena de Markov  $X = \{X_n : 1, 2, \dots\}$  donde

$X_n =$  estado de salud al tiempo  $n$  de una determinada mujer en la población.

Cada  $n$  representa periodos de 6 meses, durante los cuales las mujeres pueden pasar de un estado a otro.  $X$  tiene espacio de estados  $S = \{1, \dots, 15\}$  en donde cada número representa un estado diferente de salud:

- 1: estado normal de salud,
- 2: contagio por cepas de VPH transitorio,
- 3: contagio por las cepas 16 o 18 de VPH,
- 4: contagio por cepas distintas a la 16 o 18 de VPH, tasa
- 5: contagio por cepas de VPH de bajo riesgo,
- 6: CIN 1,
- 7: CIN 2-3,
- 8: CIS,
- 9: cáncer invasivo, etapa I,
- 10: cáncer invasivo, etapa II,
- 11: cáncer invasivo, etapa III,
- 12: cáncer invasivo, etapa IV,

---

se refiere a tumores malignos, los cuales pueden invadir los tejidos cercanos y esparcirse a otras partes del cuerpo. El término *Carcinoma in Situ* se refiere al cáncer que todavía no ha invadido algún tejido cercano.

13: aparición de síntomas del cáncer invasivo,

14: muerte y

15: eliminación de la presencia del VPH.

Las probabilidades de transición de  $X$  pueden ser representadas de la siguiente forma:

$$P_{i,j} = P\{X_{n+1} = j | X_n = i\} = \left\{ \begin{array}{ll} \frac{a_1}{s_1} \text{ si } i = 1, j = 2; & \frac{e_3}{s_7} \text{ si } i = 7, j = 8 \\ \frac{b_1}{s_1} \text{ si } i = 1, j = 3; & 1 \text{ si } i = 8, j = 9 \\ \frac{c_1}{s_1} \text{ si } i = 1, j = 4; & \frac{e_4}{s_8} \text{ si } i = 9, j = 10 \\ \frac{d_1}{s_1} \text{ si } i = 1, j = 5; & \frac{e_5}{s_8} \text{ si } i = 9, j = 13 \\ \frac{a_2}{s_2} \text{ si } i = 2, j = 6; & \frac{e_6}{s_8} \text{ si } i = 9, j = 14 \\ \frac{a_3}{s_2} \text{ si } i = 2, j = 15; & \frac{e_7}{s_9} \text{ si } i = 10, j = 11 \\ \frac{b_2}{s_3} \text{ si } i = 3, j = 6; & \frac{e_8}{s_9} \text{ si } i = 10, j = 13 \\ \frac{b_3}{s_3} \text{ si } i = 3, j = 15; & \frac{e_9}{s_9} \text{ si } i = 10, j = 14 \\ \frac{c_2}{s_4} \text{ si } i = 4, j = 6; & \frac{e_{10}}{s_{10}} \text{ si } i = 11, j = 12 \\ \frac{c_3}{s_4} \text{ si } i = 4, j = 15; & \frac{e_{11}}{s_{10}} \text{ si } i = 11, j = 13 \\ \frac{d_2}{s_5} \text{ si } i = 5, j = 6; & \frac{e_{12}}{s_{10}} \text{ si } i = 11, j = 14 \\ \frac{d_3}{s_5} \text{ si } i = 5, j = 15; & \frac{e_{13}}{s_{11}} \text{ si } i = 12, j = 13 \\ \frac{a_4}{s_6} \text{ si } i = 6, j = 2; & \frac{e_{14}}{s_{11}} \text{ si } i = 12, j = 14 \\ \frac{b_4}{s_6} \text{ si } i = 6, j = 3; & 1 \text{ si } i = 13, j = 1 \\ \frac{c_4}{s_6} \text{ si } i = 6, j = 4; & 1 \text{ si } i = 14, j = 14 \\ \frac{d_4}{s_6} \text{ si } i = 6, j = 5; & 1 \text{ si } i = 15, j = 15 \\ \frac{e_1}{s_6} \text{ si } i = 6, j = 7; & 0 \text{ e. o. c.,} \\ \frac{e_2}{s_7} \text{ si } i = 7, j = 6 & \end{array} \right.$$

donde,

$$\begin{aligned}
 s_1 &= a_1 + b_1 + c_1 + d_1 & s_7 &= e_2 + e_3 \\
 s_2 &= a_2 + a_3 & s_8 &= e_4 + e_5 + e_6 \\
 s_3 &= b_2 + b_3 & s_9 &= e_7 + e_8 + e_9 \\
 s_4 &= c_2 + c_3 & s_{10} &= e_{10} + e_{11} + e_{12} \\
 s_5 &= d_2 + d_3 & s_{11} &= e_{13} + e_{14}. \\
 s_6 &= a_4 + b_4 + c_4 + d_4 + e_1
 \end{aligned}$$

Estas expresiones se obtuvieron del trabajo de Goldie *et al.* (2003) y son resultados de las hipótesis hechas sobre el modelo de evolución de la infección. La matriz de transición  $P = (P_{i,j})_{i,j \in S}$  de  $X$  que se utilizó en el estudio de Goldie *et al.* (2003) para simular el comportamiento de la población fue la siguiente:

$$P = \begin{pmatrix}
 0 & 0.510 & 0.163 & 0.163 & 0.163 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0.166 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.833 \\
 0 & 0 & 0 & 0 & 0 & 0.257 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.742 \\
 0 & 0 & 0 & 0 & 0 & 0.257 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.742 \\
 0 & 0 & 0 & 0 & 0 & 0.164 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.835 \\
 0 & 0.220 & 0.173 & 0.173 & 0.123 & 0 & 0.309 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0.512 & 0 & 0.487 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.669 & 0 & 0 & 0.267 & 0.062 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.512 & 0 & 0.384 & 0.102 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.331 & 0.544 & 0.123 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.757 & 0.242 & 0 \\
 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
 \end{pmatrix}.$$

Los valores presentados en  $P$  fueron obtenidos utilizando datos reales de mujeres en Estados Unidos. Las probabilidades de transición fueron estimadas empíricamente, es decir, se contaron las transiciones de un estado a otro y se dividió cada resultado entre el número posible de transiciones. Estos valores fueron introducidos en un árbol Markoviano implementado utilizando el software TreeAge. Al final de la simulación se tiene el número de mujeres entre las 100,000 que estarán en cada uno de los estados de  $X$ .

## 5.2. Modelo de la transición del VPH tomando en cuenta una vacuna

Una cadena de Markov también puede modelar la transmisión del VPH teniendo en cuenta la introducción de una vacuna para un tipo específico de virus. Este modelo está desarrollado en el trabajo de Hughes *et al.* (2002), en el que una población (considerando hombres y mujeres) que se dividió en tres grupos según su actividad sexual (alta, media y baja actividad sexual), medida por la tasa de cambio de parejas, fue vacunada. Esta vacuna tenía las siguientes propiedades:

- i) reducía la susceptibilidad a la infección,
- ii) reducía la transmisibilidad del virus y
- iii) reducía la duración de la infección,

pero podía fallar de las siguientes maneras:

- a) no teniendo efecto en algunas personas,
- b) reduciendo, pero no eliminando totalmente, la susceptibilidad a la infección o
- c) perdiendo inmunidad a lo largo del tiempo.

Se consideró entonces que una parte de la población, aunque fue vacunada, podría haber quedado susceptible por cualquiera de las fallas de la vacuna. Así, las personas vacunadas

eficientemente y las susceptibles podrían infectarse y luego recuperarse. Tomando en cuenta lo anterior, la cadena de Markov se puede definir también como  $X = \{X_n : 1, 2, \dots\}$  donde

$X_n = \text{estado de salud al tiempo } n \text{ de un individuo específico en la población.}$

En este caso, el espacio de estados de  $X$  es  $S = \{1, 2, 3, 4, 5\}$  en donde cada número representa un estado diferente de salud:

1: estar sano,

2: estar vacunado,

3: estar propenso al contagio del VPH,

4: estar infectado por el virus y

5: haber eliminado el virus del organismo.

Las posibles transiciones de  $X$  son: dado que inicialmente todos los individuos son vacunados se tiene que un individuo vacunado puede permanecer sano o puede pasar a la calidad de propenso al contagio de VPH, y después de eso podría pasar al estado infectado o regresar al estado sano. Una vez infectado un individuo puede eliminar el virus del organismo o permanecer infectado.

5.2. Modelo de la transición del VPH tomando en cuenta una vacuna

---

Las probabilidades de transición de  $X$  son (ver Hughes *et al.* (2002))

$$P_{i,j} = P\{X_{n+1} = j | X_n = i\} = \begin{cases} \phi & \text{si} & i = 1, j = 2 \\ 1 - \phi & \text{si} & i = 1, j = 3 \\ \mu & \text{si} & i = 2, j = 1 \\ \sigma & \text{si} & i = 2, j = 3 \\ \varphi & \text{si} & i = 2, j = 4 \\ \frac{\mu}{s_1} & \text{si} & i = 3, j = 1 \\ \frac{\lambda}{s_1} & \text{si} & i = 3, j = 4 \\ \frac{\mu}{s_2} & \text{si} & i = 4, j = 1 \\ \frac{\gamma}{s_2} & \text{si} & i = 4, j = 5 \\ 1 & \text{si} & i = 5, j = 1 \\ 0 & \text{en otro caso,} \end{cases}$$

donde  $s_1 = \mu + \lambda$  y  $s_2 = \mu + \gamma$ . Estas probabilidades de transición son resultados de las hipótesis del modelo considerado por Hughes *et al.* (2002). La matriz de transición de  $X$  utilizada para simular el comportamiento de la población fue

$$P = \begin{pmatrix} 0 & 0.729 & 0.271 & 0 & 0 \\ 0.066 & 0 & 0.100 & 0.833 & 0 \\ 0.019 & 0 & 0 & 0.980 & 0 \\ 0.090 & 0 & 0 & 0 & 0.909 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Estos valores fueron obtenidos y utilizados de la misma forma que los obtenidos para el modelo presentado en la Sección 5.1. En este caso el resultado arrojado también fue el número de mujeres en cada estado de  $X$ .

### 5.3. Modelo de la progresión del VPH al CIS o al CCI

En este caso se quiere modelar la progresión de la infección de VPH al *Cáncer in Situ* (CIS) o al Cáncer Cérvico Invasivo (CCI) tomando en cuenta la histerectomía y la prueba del Papanicolau. Para esto una muestra de mujeres se dividió en cuatro grupos según su actividad sexual. Además se consideró que empezaron a ser sexualmente activas a los 16 años de edad por lo que se supuso que a esta edad comenzaron a ser susceptibles a la infección. Así, el riesgo de la infección en las mujeres susceptibles dependía de su edad y de el grupo de actividad sexual al que pertenecieran.

Ya infectadas, las mujeres podrían desarrollar CIS o CCI dependiendo de cuanto tiempo la infección ha estado presente en ellas. Además, cuando un cáncer no detectado está en su etapa final, en algunos casos podría detectarse alguna lesión que indicaría la existencia del cáncer y podría tratarse.

Por otro lado la prueba del Papanicolau puede reducir el riesgo del CIS, dependiendo de qué tan agresivamente se traten las lesiones de bajo grado. Se hizo la suposición de que algunas mujeres dieron positivo para displasia y fueron tratadas, y algunas otras tuvieron histerectomías, por lo que todas ellas no tuvieron riesgo de desarrollar CIS o CCI.

Las mujeres susceptibles pudieron haberse infectado y pudieron haberse practicado una histerectomía (esto sin haber tenido cáncer). Las mujeres infectadas pudieron haber desarrollado CIS o CCI, o pudieron haberse hecho una histerectomía y las que desarrollaron el cáncer debieron haberse hecho la histerectomía. En el modelo se consideró también la posibilidad de muerte.

La cadena de Markov que modela lo anterior es  $X = \{X_n : 1, 2, \dots\}$  donde

$X_n =$  estado de salud al tiempo  $n$  de una determinada mujer en la población.



### 5.3. Modelo de la progresión del VPH al CIS o al CCI

---

En este caso el espacio de estados de la cadena es  $S = \{1, \dots, 5\}$  donde cada número representa un estado diferente de salud:

- 1: ser susceptible a la infección,
- 2: haberse infectado del VPH,
- 3: haber desarrollado CIS o CCI,
- 4: haberse practicado la histerectomía y
- 5: haber muerto por otra causa que no fuera cáncer cervicouterino.

Las posibles transiciones de  $X$  son: de ser susceptibles a la infección, las mujeres podrían infectarse y después podrían desarrollar CIS o CCI. Una vez teniendo alguna de estas lesiones el tratamiento podría ser una histerectomía. Note que la muerte es otra posibilidad en cualquier transición porque se está considerando que es causada por algo distinto al cáncer cervicouterino.

Con base en las hipótesis de transición consideradas en el modelo estudiado por Hughes *et al.* (2002), las probabilidades de transición de  $X$  son

$$P_{i,j} = P\{X_{n+1} = j | X_n = i\} = \begin{cases} \frac{\xi}{s_1} & \text{si} & i = 1, j = 2 \\ \frac{\nu}{s_1} & \text{si} & i = 1, j = 4 \\ \frac{\eta}{s_1} & \text{si} & i = 1, j = 5 \\ \frac{\beta}{s_2} & \text{si} & i = 2, j = 3 \\ \frac{\theta}{s_2} & \text{si} & i = 2, j = 4 \\ \frac{\eta}{s_2} & \text{si} & i = 2, j = 5 \\ \frac{\nu}{s_3} & \text{si} & i = 3, j = 4 \\ \frac{\alpha}{s_3} & \text{si} & i = 3, j = 5 \\ 1 & \text{si} & i = 4, 5, j = 5 \\ 0 & \text{en otro caso,} \end{cases}$$

donde  $s_1 = \xi + \nu + \eta$ ,  $s_2 = \beta + \theta + \eta$  y  $s_3 = \nu + \alpha$ . Considerando esta formulación obtenida del estudio de Hughes *et al.* (2002) se obtiene la siguiente matriz de transición para  $X$  que fue utilizada para simular el comportamiento de la población:

$$P = \begin{pmatrix} 0 & 0.300 & 0 & 0.600 & 0.100 \\ 0 & 0 & 0.055 & 0.833 & 0.111 \\ 0 & 0 & 0 & 0.050 & 0.949 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Nuevamente, los valores fueron obtenidos y utilizados de la misma forma que para los modelos presentados en las Secciones 5.1 y 5.2.

## 5.4. Modelo Bayesiano

Como ya se mencionó, las matrices de transición dadas en las Secciones 5.1, 5.2 y 5.3 fueron obtenidas utilizando métodos empíricos. Las probabilidades de transición fueron calculadas utilizando el número de veces que los pacientes pasaron de un estado a otro de la enfermedad, dividido por el número total de transiciones (número total de pacientes que entran a determinado estudio). Por lo tanto, las probabilidades de transición están totalmente basadas en la información que un conjunto específico de datos proporcionan, a saber, datos proporcionados por clínicas ginecológicas.

Ahora suponga que se quiere encontrar las probabilidades de transición de las matrices de transición de los modelos presentados en las Secciones 5.1, 5.2 y 5.3, tomando en cuenta no sólo la información producida por los datos; sino también permitiendo que cierta variación pueda ocurrir. Esto se puede hacer a través de un modelo Bayesiano. Para esto, considere una cadena de Markov  $X = \{X_n : 1, 2, \dots\}$  con espacio de estados  $S = \{1, 2, \dots, M\}$ , en la que sus estados representan los diferentes estados de salud de los individuos en la población que se quiere estudiar, donde  $M > 0$  representa el número de posibles estados que  $X$  puede

tomar (por ejemplo, en el modelo dado en la Sección 5.1 se tiene  $M = 15$ ).

Suponga que  $\mathbf{x} = (x_1, \dots, x_N)$  es una secuencia de valores obtenidos a partir de las primeras  $N$  variables aleatorias que conforman la cadena de Markov, es decir, las primeras  $N$  transiciones en el estado de salud de un individuo específico en la población. Sea  $\mathbf{P} = (P_{i,j})_{i,j \in S}$  la matriz de transición de  $X$ , entonces se tiene que

$$\begin{aligned} P(X_n = x_n | X_1 = x_1, \dots, X_{n-1} = x_{n-1}, \mathbf{P}) &= P(X_n = x_n | X_{n-1} = x_{n-1}, \mathbf{P}) \\ &= P_{x_{n-1}, x_n}. \end{aligned}$$

#### 5.4.1. Función de verosimilitud

Dado que se asume un modelo Markoviano para describir la evolución y/o la transmisión del virus así como la evolución de la infección causada por el VPH, la función de verosimilitud para  $\mathbf{x} = (x_1, \dots, x_N)$  dado el parámetro  $\mathbf{P}$  es

$$\begin{aligned} l(\mathbf{P}|\mathbf{x}) &\propto f(\mathbf{x}|\mathbf{P}) \\ &\propto P(X = \mathbf{x}|\mathbf{P}) \\ &= P(X_1 = x_1|\mathbf{P})P(X_2 = x_2|X_1 = x_1, \mathbf{P})P(X_3 = x_3|X_2 = x_2, \mathbf{P}) \cdots \\ &\quad P(X_{N-1} = x_{N-1}|X_{N-2} = x_{N-2}, \mathbf{P})P(X_N = x_N|X_{N-1} = x_{N-1}, \mathbf{P}) \\ &= P(X_1 = x_1|\mathbf{P}) \prod_{k=2}^N P(X_k = x_k|X_{k-1} = x_{k-1}, \mathbf{P}) \\ &= P(X_1 = x_1|\mathbf{P}) \prod_{k=2}^N P_{x_{k-1}, x_k}. \end{aligned} \tag{5.1}$$

Haciendo la suposición de que  $X_1$  se distribuye uniformemente, es decir que

$$P(X_1 = x_1) = \frac{1}{M}$$

se tiene que la fórmula (5.1) es

$$l(\mathbf{P}|\mathbf{x}) \propto \prod_{k=2}^N P_{x_{k-1}, x_k}.$$

Note que la función de verosimilitud es proporcional a una distribución multinomial con parámetros  $\{P_i, i \in S\}$  donde

$$P_i = (P_{i,j})_{j \in S},$$

los cuales son los renglones de la matriz de transiciones de  $X$ .

### 5.4.2. Distribuciones a priori

Ya que la función de verosimilitud es proporcional a una distribución multinomial, conviene elegir distribuciones a priori Dirichlet para los renglones de la matriz, ya que éstas son conjugadas para las funciones multinomiales <sup>3</sup>.

Haciendo el supuesto de que los renglones de la matriz son independientes se tiene que

$$P_{\text{priori}}(P_i) = \frac{\Gamma(\alpha_{i,1} + \dots + \alpha_{i,M})}{\Gamma(\alpha_{i,1}) \dots \Gamma(\alpha_{i,M})} \prod_{j=1}^M (P_{i,j})^{\alpha_{i,j}-1}, \quad i = 1, 2, \dots, M$$

donde  $\alpha_i = (\alpha_{i,1}, \dots, \alpha_{i,M})$ ,  $i = 1, 2, \dots, M$  son los parámetros conocidos de las distribuciones.

Una forma de asociar valores a los parámetros  $\alpha_i$ ,  $i = 1, 2, \dots, M$  es relacionarlos directamente con el número de transiciones de un estado a otro. Por ejemplo, si en el renglón  $P_1$  la probabilidad de transición de mayor valor es  $P_{1,k_1}$ , entonces  $\alpha_{1,k_1}$  tendrá un valor más alto asociado a él, si  $P_{1,k_2}$  es el segundo mayor valor, entonces  $\alpha_{1,k_2}$  tendrá el segundo mayor valor y así en adelante.

### 5.4.3. Distribuciones a posteriori

Para obtener las distribuciones a posteriori defina  $n_{i,j}$ ,  $i, j \in S$  como el número de veces que la cadena pasa del estado  $i$  al estado  $j$  en  $\mathbf{x}$ , es decir,

$$n_{i,j} = \sum_{k=2}^N I(X_{k-1} = i, X_k = j) \quad (5.2)$$

<sup>3</sup>Ver Apéndice B, Definiciones B.4 y B.10.

donde  $i, j = 1, 2, \dots, M$ . Para cada  $i = 1, 2, \dots, M$  considere el siguiente vector:

$$n_i = (n_{i,1}, n_{i,2}, \dots, n_{i,M}).$$

Dado que la función de verosimilitud es proporcional a una distribución multinomial y las distribuciones a priori para los renglones de  $\mathbf{P}$  son proporcionales a distribuciones Dirichlet, entonces sus distribuciones a posteriori también son proporcionales a distribuciones Dirichlet con parámetros  $\alpha_i + n_i$  con  $i = 1, 2, \dots, M$ <sup>4</sup>.

**Observación.**

Para los modelos de la transmisión del virus considerando una vacuna y del progreso del virus al cáncer se tiene que la función de verosimilitud es proporcional a una distribución multinomial con parámetros  $\{P_i, i \in S\}$  y las distribuciones a priori y a posteriori son proporcionales a una distribución Dirichlet con parámetros  $\alpha_i = (\alpha_{i,1}, \dots, \alpha_{i,M})$  y  $\alpha_i + n_i$  con  $i = 1, 2, \dots, M$  respectivamente, donde  $n_i = (n_{i,1}, n_{i,2}, \dots, n_{i,M})$  con  $n_{i,j}$  definida en la fórmula (5.2).

Para estimar  $\mathbf{P}$  se puede tomar para cada renglón el vector que maximiza la distribución a posteriori marginal para aquél renglón. De Evans y Swartz (2000) se tiene que la moda de una distribución de Dirichlet con parámetros  $(\alpha_1, \alpha_2, \dots, \alpha_M)$  es

$$P_{i,j} = \frac{n_{i,j} + \alpha_{i,j}}{\sum_{k=1}^M [n_{i,k} + \alpha_{i,k}]}, \quad i, j = 1, 2, \dots, M.$$

**5.4.4. Otras distribuciones a priori**

Note que la distribución de Dirichlet no es la única distribución que puede ser utilizada como distribución a priori para los renglones de  $\mathbf{P}$ . También se puede utilizar cualquier otra distribución definida en el conjunto

$$\Delta_M = \left\{ x = (x_1, x_2, \dots, x_M) : x_i \geq 0, \sum_{i=1}^M x_i = 1 \right\},$$

---

<sup>4</sup>Ver Apéndice B, Proposición B.1.

donde  $M$  es el número de estados de la cadena de Markov describiendo el modelo. En algunos casos es posible que la distribución a priori no sea conjugada a la función de verosimilitud. En estos casos se podría utilizar algún método de Monte Carlo vía Cadenas de Markov para obtener los valores de  $\mathbf{P}$ . Una opción es considerar un algoritmo del tipo Metropolis-Hastings. En este caso la cadena de Markov que tendría como su distribución estacionaria la distribución a posteriori, que en este caso está dada por

$$P(\mathbf{P}|\mathbf{x}) \propto \left[ \prod_{k=2}^N P_{x_{k-1}, x_k} \right] \left[ \prod_{i=1}^M P_{priori}(P_i) \right], \quad (5.3)$$

puede ser descrita como sigue.

Suponga que el valor actual de un renglón  $i$  de  $\mathbf{P}$  es  $P_i^{(n)} = (P_{i,1}^{(n)}, P_{i,2}^{(n)}, \dots, P_{i,M}^{(n)})$  y que el candidato a ser el nuevo renglón es muestreado a partir de la distribución a priori de  $P_i$  para cada uno de los renglones  $P_i$ . Sea  $P_i^{(n+1)}$  el candidato a este renglón. La probabilidad de aceptar el nuevo renglón y consecuentemente la nueva matriz es

$$\begin{aligned} \rho(P^{(n)}, P^{(n+1)}) &= \min \left\{ 1, \frac{l(\mathbf{x}|P^{(n+1)}) \left[ \prod_{i=1}^M P_{priori}(P_i^{(n+1)}) \right]}{l(\mathbf{x}|P^{(n)}) \left[ \prod_{i=1}^M P_{priori}(P_i^{(n)}) \right]} \frac{\prod_{i=1}^M P_{priori}(P_i^{(n)})}{\prod_{i=1}^M P_{priori}(P_i^{(n+1)})} \right\} \\ &= \min \left\{ 1, \frac{l(\mathbf{x}|P^{(n+1)})}{l(\mathbf{x}|P^{(n)})} \right\}. \end{aligned}$$

Si se espera un tiempo suficientemente largo, los valores muestreados por este algoritmo tendrán como distribución la distribución (5.3).



# Discusión

A pesar de existir datos que apoyan la idea de que la infección del VPH recurrente está relacionada con el desarrollo del cáncer cérvico, la verdad es que el curso completo de la enfermedad, desde la adquisición del VPH al desarrollo del cáncer cérvico invasivo, nunca ha sido y nunca será completamente observado. Es por esto que son importantes los modelos estadísticos y probabilísticos en el entendimiento de la biología, epidemiología e implicaciones políticas de la infección del VPH y el cáncer cérvico. Por otro lado, aunque cualquier modelo matemático es una simplificación y representación de la realidad, esto es quizás especialmente cierto en modelos de infecciones transmitidas sexualmente, en los que es necesario conocer la dinámica de los cambios de pareja en la población para el entendimiento de la enfermedad.

Por otro lado, el desarrollo de una vacuna contra la infección del VPH será de gran utilidad, en adición a las tecnologías ya disponibles en el control del cáncer cérvico. Por esto, un modelo matemático que explore los impactos de la vacunación tomando en cuenta los datos epidemiológicos, clínicos y virológicos existentes proveerá una importante ayuda al desarrollo de la política de salud.

Así, los modelos probabilísticos y estadísticos ayudan a los investigadores clínicos, a los analistas y tomadores de decisiones en salud pública para determinar acciones apropiadas basadas en los datos clínicos y epidemiológicos disponibles. En esta tesis, las cadenas de Markov son de gran utilidad para modelar la historia natural del VPH, la transición del



VPH tomando en cuenta una vacuna y la progresión del VPH al cáncer cérvico, mientras que el modelo Bayesiano y el algoritmo Metropolis-Hastings proporcionan una herramienta útil para encontrar estimadores para las probabilidades de transición de las cadenas de Markov.

Dado que no se propone hacer un estudio aplicado, solo se describió el algoritmo Metropolis-Hastings, sin la programación del mismo. La programación así como la aplicación a datos reales podría ser el tema de estudios posteriores.

Con la formulación Bayesiana es posible obtener no sólo estimadores para las probabilidades de transición utilizadas para simular el comportamiento de una población bajo los supuestos de cualquiera de los modelos presentados en el Capítulo 5 (por ejemplo, la moda de la distribución a posteriori marginal), sino también proporciona una idea de como estas probabilidades se comportan en el momento en que proporcionan información sobre su función de distribución.

Por otro lado, otros métodos de Monte Carlo vía Cadenas de Markov podrían ser utilizados en algunos modelos epidemiológicos como los métodos de muestreo de Gibbs y métodos de muestreo de Gibbs por bloques. El algoritmo que se utiliza depende del problema que está siendo tratado, aunque la ventaja del algoritmo Metropolis-Hastings es que con él se puede obtener una muestra dependiente de variables aleatorias cuya distribución siempre converge para la distribución de interés.

## Apéndice A

# Conceptos básicos de probabilidad

El propósito de este apéndice es recordar algunos conceptos y resultados de la teoría de probabilidad. Para mayor información se recomienda consultar Ash (2000), Grimmet y Stirzaker (1982), Ross (1998) y Ruíz-Maya y Martín Pliego (1995).

### A.1. Algunas definiciones

**Definición A.1** La  $\sigma$ -álgebra de Borel, denotada por  $\mathfrak{B}(\mathbb{R})$ , es la mínima  $\sigma$ -álgebra de subconjuntos de  $\mathbb{R}$  que contiene a todos los intervalos  $(a, b]$ ,  $a, b \in \mathbb{R}$ .

**Definición A.2** Sea  $X$  una variable aleatoria sobre un espacio de probabilidad  $(\Omega, \mathcal{F}, \mathcal{P})$ . La función  $f : \mathbb{R} \rightarrow [0, \infty)$  es una **función de densidad** de  $X$  si y sólo si

a)  $f(x) \geq 0$  para toda  $x \in \mathbb{R}$ ,

b)

$$\int_{\mathbb{R}} f(x) dx = 1. \tag{A.1}$$

## A.2. Algunos resultados

---

Si  $X$  es una variable aleatoria discreta,  $f$  es llamada **función de probabilidad** y la fórmula (A.1) se convierte en

$$\sum_x f(x) = 1.$$

**Definición A.3** Sea  $X$  una variable aleatoria sobre un espacio de probabilidad  $(\Omega, \mathcal{F}, \mathcal{P})$ .

Si  $X$  es continua su **esperanza** se define como

$$E[X] = \int_{\mathbb{R}} xf(x)dx,$$

donde  $f$  es la función de densidad de  $X$ . Si  $X$  es discreta, su esperanza es

$$E[X] = \sum_x xf(x).$$

**Definición A.4** Sea  $(\Omega, \mathcal{F}, \mathcal{P})$  un espacio de probabilidad. Tome  $A, B \in \mathcal{F}$ . Si  $P(B) > 0$ , la **probabilidad condicional** de que ocurra  $A$  dado que ha ocurrido  $B$  está dada por la expresión

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

**Definición A.5** Sean  $X$  y  $Y$  variables aleatorias sobre un espacio de probabilidad  $(\Omega, \mathcal{F}, \mathcal{P})$ .

La **esperanza condicional** de  $X$  dado  $Y$  está dada por

$$E[X|Y = y] = \int xf(x|y)dx \quad (\text{A.2})$$

donde  $f(x|y)$  es la función de densidad condicional de  $x$  dado  $y$ . Si  $X$  es discreta, la expresión (A.2) se convierte en

$$\sum_x xf(x|y).$$

## A.2. Algunos resultados

**Teorema A.1 (Teorema de Bayes)** Sea  $(\Omega, \mathcal{F}, \mathcal{P})$  un espacio de probabilidad. Considere la colección  $\{A_k\}_{k=1}^n \subset \mathcal{F}$  tal que  $\cup_{k=1}^n A_k = \Omega$  y  $A_i \cap A_j = \phi$  para toda  $i, j = 1, 2, \dots, n$  con  $i \neq j$ . Entonces

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{k=1}^n P(A_k)P(B|A_k)}, \quad i = 1, 2, \dots, n$$

para todo  $B \in \mathcal{F}$ .

**Demostración.** Ver Ross (1998).

**Teorema A.2 (Ley Fuerte de los Números Grandes)** Sea  $X_1, X_2, \dots$  una sucesión de variables aleatorias independientes e idénticamente distribuídas, cada una con esperanza finita  $E[X_k] = \mu$  para toda  $k = 1, 2, \dots$  Entonces

$$P\left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_k = \mu\right) = 1.$$

**Demostración.** Ver Ross (1998).

**Lema A.1** Sea  $X = \{X_n : n = 0, 1, \dots\}$  una cadena de Markov con espacio de estados  $S$ .

Para  $i_0, i_1, \dots, i_n \in S$  la igualdad

$$\begin{aligned} P\{X_n = i_n, X_{n-2} = i_{n-2}, \dots, X_0 = i_0 | X_{n-1} = i_{n-1}\} = \\ P\{X_n = i_n | X_{n-1} = i_{n-1}\} P\{X_{n-2} = i_{n-2}, X_{n-3} = i_{n-3}, \dots, X_0 = i_0 | X_{n-1} = i_{n-1}\} \end{aligned}$$

es equivalente a

$$P\{X_n = i_n | X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}\} = P\{X_n = i_n | X_{n-1} = i_{n-1}\}.$$

**Demostración.**

(i) Por hipótesis se tiene que

$$\begin{aligned} P\{X_n = i_n | X_{n-1} = i_{n-1}\} = \\ \frac{P\{X_n = i_n, X_{n-2} = i_{n-2}, \dots, X_0 = i_0 | X_{n-1} = i_{n-1}\}}{P\{X_{n-2} = i_{n-2}, X_{n-3} = i_{n-3}, \dots, X_0 = i_0 | X_{n-1} = i_{n-1}\}}. \end{aligned} \tag{A.3}$$

Note que el lado derecho de la igualdad es equivalente a la expresión

$$\frac{P\{X_0 = i_0, \dots, X_{n-2} = i_{n-2}, X_{n-1} = i_{n-1}, X_n = i_n\}}{P\{X_0 = i_0, \dots, X_{n-3} = i_{n-3}, X_{n-2} = i_{n-2}, X_{n-1} = i_{n-1}\}}, \tag{A.4}$$

y esta expresión es igual a

$$P\{X_n = i_n | X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}\}.$$

(ii) Por la demostración del inciso anterior de este lema se tiene que

$$P\{X_n = i_n | X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}\} = \frac{P\{X_n = i_n, X_{n-2} = i_{n-2}, \dots, X_0 = i_0 | X_{n-1} = i_{n-1}\}}{P\{X_{n-2} = i_{n-2}, X_{n-3} = i_{n-3}, \dots, X_0 = i_0 | X_{n-1} = i_{n-1}\}}. \quad (\text{A.5})$$

Por hipótesis se sabe que el lado izquierdo de la fórmula (A.5) es igual a

$$P\{X_n = i_n | X_{n-1} = i_{n-1}\}.$$

De aquí se deduce que la expresión (A.3) es cierta, y así se tiene lo que se quería demostrar.

## Apéndice B

# VARIABLES ALEATORIAS

El propósito de este apéndice es recordar las variables aleatorias usadas en esta tesis y presentar un resultado sobre inferencia Bayesiana usado en el Capítulo 5.

### B.1. Variables aleatorias discretas

**Definición B.1** Sea  $X$  una variable aleatoria discreta que toma valores no negativos. Se dice que  $X$  es una **variable aleatoria Poisson** con parámetro  $\lambda > 0$  si tiene función de probabilidad

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \dots$$

**Definición B.2** Suponga que se realizan repeticiones independientes de un experimento aleatorio que tiene dos posibles resultados, los cuales se pueden clasificar como éxito y fracaso, y donde la probabilidad de éxito es  $p \in [0, 1]$ . Sea  $X$  el número de fracasos antes del primer éxito.  $X$  es llamada **variable aleatoria Geométrica** con parámetro  $p$  y su función de probabilidad es

$$P(X = k) = p(1 - p)^k, \quad k = 0, 1, 2, \dots$$

**Definición B.3** Considere  $n$  repeticiones independientes de un experimento aleatorio que tiene dos posibles resultados, los cuales se pueden clasificar como éxito y fracaso, y donde la

probabilidad de éxito es  $p \in [0, 1]$ . Sea  $X$  el número de éxitos ocurridos en las  $n$  repeticiones. Así,  $X$  es llamada **variable aleatoria Binomial** con parámetros  $(n, p)$ , y su función de probabilidad está dada por

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n.$$

**Definición B.4** Considere  $n$  repeticiones independientes de un experimento aleatorio que tiene  $r$  posibles resultados con respectivas probabilidades  $p_1, p_2, \dots, p_r$ ,  $\sum_{i=1}^r p_i = 1$ . También considere el vector aleatorio  $X = (X_1, X_2, \dots, X_r)$  donde  $X_k$  es el número de repeticiones con resultado tipo  $k$ . Así,  $X$  es llamada **variable aleatoria multinomial** con parámetros  $n$  y  $p = (p_1, p_2, \dots, p_r)$  y con función de probabilidad

$$P(X = x) = \frac{n!}{x_1! x_2! \dots x_r!} p_1^{x_1} p_2^{x_2} \dots p_r^{x_r}$$

donde  $\sum_{i=1}^r x_i = n$  y  $x_k \geq 0$ .

**Observación.** Si  $X = (X_1, X_2, \dots, X_r)$  tiene distribución multinomial con parámetros  $n$  y  $p$ , entonces cada  $X_k$  tiene distribución binomial con parámetros  $(n, p_k)$  para toda  $k = 1, 2, \dots, r$ .

## B.2. Variables aleatorias continuas

**Definición B.5** Sea  $X$  una variable aleatoria que toma valores en el intervalo  $(a, b)$  donde  $a, b \in \mathbb{R}$ .  $X$  es llamada **variable aleatoria Uniforme** sobre el intervalo  $(a, b)$  si tiene función de densidad

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } a < x < b, \\ 0 & \text{e.o.c.} \end{cases}$$

**Definición B.6** Se dice que  $X$  es una **variable aleatoria Normal** con parámetros  $(\mu, \sigma^2)$  si su función de densidad es de la forma

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \quad \text{si } -\infty < x < \infty.$$

**Definición B.7** Considere una variable aleatoria continua  $X$  que toma valores no negativos.  $X$  es llamada **variable aleatoria Gama** con parámetros  $(\alpha, \beta)$  donde  $\alpha, \beta > 0$  si su función de densidad está dada por

$$f(x) = \begin{cases} \frac{\beta}{\Gamma(\alpha)} (\beta x)^{\alpha-1} e^{-\beta x} & \text{si } x \geq 0, \\ 0 & \text{e.o.c.} \end{cases}$$

donde, para  $\alpha > 0$ ,

$$\Gamma(\alpha) = \begin{cases} (\alpha - 1)! & \text{si } \alpha \text{ es entero,} \\ \int_0^\infty x^{\alpha-1} e^{-x} dx & \text{si } \alpha \text{ no es entero.} \end{cases}$$

**Definición B.8** Sea  $X$  una variable aleatoria continua que toma valores no negativos.  $X$  es llamada **variable aleatoria Ji-Cuadrada** con  $\nu$  grados de libertad, denotada por  $\chi_{(\nu)}^2$ , si su función de densidad es

$$f(x) = \begin{cases} \frac{1}{\Gamma(\nu/2)2^{\nu/2}} x^{\nu/2-1} e^{-x/2} & \text{si } x > 0, \\ 0 & \text{e.o.c.} \end{cases}$$

**Definición B.9** Se dice que  $X$  es una **variable aleatoria Beta** con parámetros  $(a, b)$  si tiene función de densidad

$$f(x) = \begin{cases} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1} & \text{si } 0 < x < 1, \\ 0 & \text{e.o.c.} \end{cases}$$

**Definición B.10** Sea  $X = (X_1, X_2, \dots, X_r)$  un vector aleatorio. Se dice que  $X$  tiene **distribución de Dirichlet** con parámetro  $a = (a_1, a_2, \dots, a_r)$ ,  $a_k > 0$ ,  $k = 1, 2, \dots, r$  si tiene función de densidad

$$f(x) = \frac{\Gamma(a_1 + a_2 + \dots + a_r)}{\Gamma(a_1)\Gamma(a_2)\dots\Gamma(a_r)} x_1^{a_1-1} \dots x_r^{a_r-1}$$

donde  $x_k \geq 0$ ,  $k = 1, 2, \dots, r$ .

**Observación.** Si  $X = (X_1, X_2, \dots, X_r)$  tiene distribución de Dirichlet con parámetro  $a = (a_1, a_2, \dots, a_r)$ , entonces cada  $X_k$  se distribuye Beta con parámetros  $(a_k, \sum_{i=1}^r a_i - a_k)$  para toda  $k = 1, 2, \dots, r$ .



### B.3. Resultado complementario

**Proposición B.1** *Considere un modelo Bayesiano en el que la función de verosimilitud  $l(p|\mathbf{x})$  es proporcional a una función multinomial con parámetros  $n$  y  $p = (p_1, p_2, \dots, p_r)$ , y en el que la distribución a priori para el vector aleatorio  $p$  es una distribución de Dirichlet con parámetro  $a = (a_1, a_2, \dots, a_r)$ ,  $a_k > 0$ ,  $k = 1, 2, \dots, r$ . Entonces, la distribución a posteriori para  $p$  es proporcional a una distribución de Dirichlet con parámetro  $\bar{a} = (a_1 + x_1, a_2 + x_2, \dots, a_r + x_r)$ .*

**Demostración.**

Por hipótesis se tiene que la función de verosimilitud y la distribución a priori satisfacen, respectivamente, que

$$l(p|\mathbf{x}) \propto \prod_{k=1}^r p_k^{x_k},$$

y que

$$P_{\text{priori}}(p) \propto \prod_{k=1}^r p_k^{a_k-1}.$$

Por la fórmula (4.6) del Capítulo 4 se sabe que la distribución a posteriori satisface que

$$\begin{aligned} P(p|\mathbf{x}) &\propto l(p|\mathbf{x})P_{\text{priori}}(p) \\ &\propto \left( \prod_{k=1}^r p_k^{x_k} \right) \left( \prod_{k=1}^r p_k^{a_k-1} \right) \\ &= \prod_{k=1}^r p_k^{x_k+a_k-1}, \end{aligned}$$

por lo que se puede concluir que la distribución a posteriori es proporcional a una distribución de Dirichlet con parámetro  $\bar{a} = (a_1 + x_1, a_2 + x_2, \dots, a_r + x_r)$ .

# Bibliografía

- [1] Ash, R. B. (2000). *Probability and measure theory*. Academic Press, New York.
- [2] Carlin, B. P. y Louis, T. A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*. 2nd. ed. Chapman and Hall, London.
- [3] Evans, M. y Swartz, T. (2000). *Aproximating integrals via Monte Carlo and deterministic methods*. Oxford Statistical Sciences Series 20, UK:Oxford University Press.
- [4] Feller, W. (1968). *An introduction to probability theory and its applications*, Vol. I. 3rd. ed. John Wiley and Sons, New York.
- [5] Gálvez-Correa, G. (2006). El virus del papiloma humano. *¿Cómo ves?*, **94**, 10-14.
- [6] Goldie, S. J., Grima, D., Kohli, M., Wright, T. C., Whinstin, M. y Franco, E. (2003). A comprehensive natural history model of HPV infection and cervical cancer to estimate the clinical impact of a prophylactic HPV-16/18 vaccine. *Int. J. Cancer*, **106**, 896-904.
- [7] Grimmett, G. R. y Stirzaker, D. R. (1982). *Probability and random processes*. Oxford Science Publications.
- [8] Gutiérrez-Peña, E. (1997). *Métodos computacionales en la inferencia Bayesiana*. Monografías. Vol. 6, No. 15. I.I.M.A.S., México.
- [9] Hughes, J. P., Garnett, G. P. y Koutsky, L. (2002). The Theoretical Population-Level Impact of a Prophylactic Human Papilloma Virus Vaccine. *Epidemiology*, **13**, 631-639.

- [10] Karlin, S. y Taylor, H. (1975). *A first course in stochastic processes*. 2nd. ed. Academic Press.
- [11] Lee, P. M. (1989). *Bayesian Statistics. An Introduction*. 2nd. ed. New York: John Wiley and Sons.
- [12] Leonard, T. y Hsu, J. S. (1999). *Bayesian Methods. An Analysis for Statisticians and Interdisciplinary Researchers*. Cambridge University Press.
- [13] López Ramírez, Jorge (2005). El uso del muestreo de Gibbs para identificar segmentos homogéneos en una secuencia de ADN heterogénea. *Tesis-Licenciatura*. U.N.A.M.
- [14] Myers, E. R., McCrory, D. C., Nanda, K., Bastian, L. y Matchar, D. B. (2000). Mathematical Model for the Natural History of Human Papillomavirus Infection and Cervical Carcinogenesis. *American Journal of Epidemiology*, **151**, 1158-1171.
- [15] Robert, C. y Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer Verlag, New York.
- [16] Ross, Sheldon M. (1996). *Stochastic Processes*. 2nd. ed. New York: John Wiley and Sons.
- [17] Ross, Sheldon M. (1997). *Introduction to probability models*. 6th. ed. Academic Press, San Diego, California.
- [18] Ross, Sheldon M. (1998). *A First Course in Probability*. 5th. ed. Prentice-Hall, Inc.
- [19] Ross, Sheldon M. (1997). *Simulation*. 2nd. ed. Academic Press, San Diego, California.
- [20] Ruíz-Maya Pérez, L. y Martín Pliego, J. (1995). *Estadística I*. 2da. ed. Madrid: Universidad Autónoma de Madrid.
- [21] Sanders, G. D. y Taira, A. V. (2003). Cost Effectiveness of a Potential Vaccine for Human papillomavirus. *Emerging Infectious Diseases*, **9**, 37-48.