



"El saber de mis hijos  
hará mi grandeza"

---

# UNIVERSIDAD DE SONORA

DIVISIÓN DE CIENCIAS EXACTAS Y NATURALES

Programa de Posgrado en Matemáticas

Inferencia Estadística para Cadenas de Markov

## T E S I S

Que para obtener el grado académico de:

**Maestro en Ciencias**  
(Matemáticas)

Presenta:

Martha Cecilia Palafox Duarte

Director de Tesis: Dr. Miguel Nakamura Savoy

Hermosillo, Sonora, México, 29 de Junio, 2009.



## SINODALES

Dr. Jesús Adolfo Minjárez Sosa  
Departamento de Matemáticas, Universidad de Sonora

Dr. José Arturo Montoya Laos  
Departamento de Matemáticas, Universidad de Sonora

Dr. Miguel Nakamura Savoy  
Centro de Investigación en Matemáticas, Guanajuato, Gto.

M. C. María Teresa Robles Alcaraz  
Departamento de Matemáticas, Universidad de Sonora



# UNIVERSIDAD DE SONORA

## Acta de Examen de Grado



EL SABER DE MIS HIJOS  
PARA MI GRANDEZA

En la ciudad de Hermosillo, Sonora, México, siendo las 18:00 horas  
del día 29 de Junio del 2009, se reunieron en:

El Auditorio del Departamento de Matemáticas

de la Universidad de Sonora, los señores:

Dr. Adolfo Minjárez Sosa, Dr. José Arturo Montoya Laos

Dr. Miguel Nakamura Savoy

M.C. Ma. Teresa Robles Alcaraz

bajo la Presidencia del primero y fungiendo como Secretario el último, para efectuar el  
Examen de Grado a:

MARTHA CECILIA PALAFOX DUARTE

de Maestro en Ciencias (Matemáticas)

quien presentó el trabajo de tesis

"Inferencia Estadística para Cadenas de Markov"

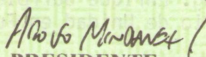
Acta No. 11  
Foja 11  
Libro 1  
Exp. No. 204290143

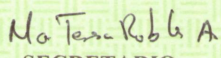
Los sinodales replicaron a la sustentante y después de debatir entre sí reservada y  
libremente, la declararon:

Aprobada por unanimidad

Acto continuo el Presidente del Jurado le hizo saber el resultado de su examen y para  
constancia se levanta la presente, firmando los que intervinieron.

  
Firma del sustentante

  
PRESIDENTE

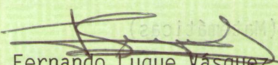
  
SECRETARIO

  
VOCAL



Dr. Fernando Luque Vásquez Coordinador  
del Programa de: Posgrado en Matemáticas  
de la Universidad de Sonora, hace constar que las firmas que anteceden corresponden a  
los sinodales que intervinieron en el examen que contiene la presente acta.

Hermosillo, Sonora, a 29 de Junio del 2009.

  
Dr. Fernando Luque Vásquez  
Coordinador del Programa

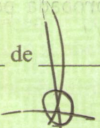


DIVISION DE CIENCIAS  
EXACTAS Y NATURALES  
COORDINACIÓN  
POSGRADO EN MATEMÁTICAS

Ing. Enrique Gurrola Mac, Secretario de Unidad en funciones de Vicerrector  
de la Unidad Regional Centro de la Universidad de Sonora, hace  
constar que: el Dr. Fernando Luque Vásquez  
es a la fecha de expedición de este documento, Coordinador del Programa arriba citado  
y suya la firma que aparece al calce del mismo.

Hermosillo, Sonora, a 29 de Junio del 2009.



  
Ing. Enrique Gurrola Mac  
Secretario de Unidad en Funciones de

EL SABER DE MIS HIJOS  
HARA MI GRANDEZ  
UNIDAD REGIONAL  
CENTRO  
VICERRECTORIA

**VICERRECTOR** de la  
**UNIDAD REGIONAL CENTRO**

# Agradecimientos

Primeramente agradezco a Dios por estar conmigo en cada paso que doy e iluminar mi mente y darme fuerzas para seguir adelante en cada tropiezo. Por permitirme haber llegado hasta este momento tan importante en mi vida y haber puesto en mi camino a aquellas personas que han sido mi soporte y compañía durante todo el período de estudio.

A mi familia que me acompañó de forma incondicional, entendieron mis ausencias y malos momentos, dándome ánimo para terminar este proyecto.

A mi amor Erik por su cariño, amor, comprensión y por brindarme su apoyo en todo momento.

Sin lugar a dudas, este trabajo no pudo haberse realizado sin la formación que recibí durante varios años en el Departamento de Matemáticas de la Universidad de Sonora. Por ello, mi agradecimiento a todos los maestros que contribuyeron en mi formación académica. Especialmente, al Dr. Miguel Nakamura Savoy por su paciencia y dirección de esta tesis; al Dr. José Arturo Montoya Laos por su asesoría, consejos y valiosas aportaciones; a la M.C. María Teresa Robles Alcaraz por sus enseñanzas y colaboración; a la M.C. Gudelia Figueroa Preciado por su apoyo y amistad brindada a mi persona.

A mis sinodales por el tiempo que dedicaron a revisar con paciencia este trabajo y que lo enriquecieron a través de sus comentarios.

A mi gran amiga y compañera María Elena Conde Hernández por compartir momentos de crisis y felicidad. A mis queridos compañeros que me apoyaron y permitieron entrar en sus vidas durante estos años y a todos mis amigos por el ánimo brindado en esta etapa.

Finalmente, hago extensivo mi más sincero agradecimiento a todas aquellas personas que de una forma u otra, colaboraron o participaron en la realización de este trabajo.

Sin todos ustedes, este trabajo no habría sido posible.





# Contenido

<b>Introducción</b>	<b>1</b>
<b>1 Teoría de Cadenas de Markov</b>	<b>3</b>
1.1 Conceptos básicos . . . . .	3
1.2 Ejemplos de cadenas de Markov . . . . .	8
1.3 Clasificación de estados de una Cadena de Markov . . . . .	10
1.4 Distribución límite de una cadena de Markov . . . . .	21
<b>2 Inferencia basada en Verosimilitud</b>	<b>39</b>
2.1 La función de verosimilitud . . . . .	39
2.2 La función de verosimilitud relativa . . . . .	41
2.3 Propiedades de la verosimilitud . . . . .	41
2.4 Intervalos de verosimilitud . . . . .	42
2.5 Intervalos de verosimilitud-confianza . . . . .	43
2.6 Ejemplo . . . . .	46
<b>3 Verosimilitud para Cadenas de Markov</b>	<b>49</b>
3.1 Caso paramétrico . . . . .	49
3.2 Caso no paramétrico . . . . .	55
3.3 Problema de aplicación real . . . . .	60
<b>4 Conclusiones</b>	<b>67</b>
<b>Bibliografía</b>	<b>67</b>



# Introducción

Los procesos estocásticos sirven para caracterizar y estudiar fenómenos aleatorios que evolucionan, generalmente, con el paso del tiempo. Una clase muy importante de este tipo de procesos la constituyen los procesos de Markov. Estos tienen la característica general que dado el valor actual del proceso  $X_t$ , los valores futuros  $X_s$  para  $s > t$  son independientes de los valores pasados  $X_u$  para  $u < t$ . Una cadena de Markov a tiempo discreto es un proceso de Markov cuyo espacio de estados  $T$  es un conjunto finito o numerable y el sistema es observado en tiempo discreto  $n = 0, 1, 2, \dots$ , Hoel (1972, pág. 1). Algunos ejemplos de cadenas de Markov ampliamente estudiados en la literatura de Probabilidad son: caminatas aleatorias unidimensional, la ruina del jugador, cadena de Ehrenfest, modelos de colas, entre otros.

Una cadena de Markov está completamente determinada si se especifican su matriz de transición,  $P$ , y la distribución inicial del proceso  $X_0$ ,  $\pi_0$ . Es decir, es posible calcular cualquier probabilidad asociada a la cadena usando la Ley de Probabilidad Total. En la teoría de probabilidad para cadenas de Markov se pone mucho énfasis en el estudio de las matrices de transición de orden  $n$ , clasificación de estados de la cadena, propiedades asintóticas, existencia y unicidad de distribuciones estacionarias, Feller(1950), Norris (1998).

Por otra parte, el tratamiento que le da la Estadística a cadenas de Markov es muy diferente al interés probabilista. Bajo el paradigma estadístico un problema relevante y poco estudiado es hacer inferencia sobre las probabilidades de transición basados en una muestra observada  $i_1, \dots, i_n$  de la cadena. Entiéndase por inferencia no sólo una estimación puntual del parámetro de interés  $P$  sino también estimación por intervalos de las probabilidades  $P_{ij}$  que la conforman.

El objetivo fundamental que persigue este trabajo es considerar el enfoque de verosimilitud para hacer inferencia estadística sobre las probabilidades de transición de una cadena de Markov paramétrica y no paramétrica. Se ejemplificará esta metodología con datos reales de una cadena de direcciones de viento. Además, se propondrá un método gráfico para validar la propiedad de Markov. Por último se aplicará el enfoque probabilista para describir las características que presenta la cadena y analizar el sistema con el paso del tiempo.

En el primer capítulo se expondrá la teoría de probabilidad relevante y necesaria sobre cadenas de Markov para tener los conceptos básicos acerca de este proceso que será retomado más adelante como ejemplo para ilustrar el enfoque de verosimilitud. También se describirán algunos ejemplos importantes de tal proceso.

En el Capítulo 2 se incluirán resultados teóricos importantes para hacer inferencia estadística utilizando la metodología de verosimilitud. Se mostrarán varios ejemplos donde se ilustra el proceso inferencial bajo este enfoque.

En el Capítulo 3 se realizarán los desarrollos correspondientes de dicha metodología

estadística para cadenas de Markov con el objetivo de hacer inferencia sobre las probabilidades de transición. Se considerará tanto el caso paramétrico, en el que las probabilidades de transición pueden ser indexadas con un vector de parámetros, como el no paramétrico. Además, se hará un estudio de simulación para verificar las propiedades estadísticas de la verosimilitud (cobertura de intervalos) bajo estos modelos de probabilidad. También, se proporcionará un método gráfico para validar la propiedad de Markov. Se aplicará este enfoque a datos reales de direcciones del viento que serán modelados con una cadena de Markov no paramétrica en donde se desean estimar las probabilidades de cambio de una dirección a otra. Finalmente, se analizarán y describirán las propiedades probabilísticas que presenta dicha cadena.

Las aportaciones principales de esta tesis son:

1. Realizar inferencia estadística sobre los parámetros de un proceso estocástico específico. En particular, sobre las probabilidades de transición de una cadena de Markov.
2. Dar elementos, en general, necesarios para aplicar el enfoque de verosimilitud.
3. Contribuir con material escrito sobre el método de verosimilitud para hacer inferencias sobre probabilidades de transición para una cadena de Markov tanto en el caso paramétrico como el no paramétrico.
4. Mostrar un proceso de simulación para generar observaciones de una cadena de Markov (paramétrica y no paramétrica) y verificar la cobertura de los intervalos obtenidos con el método de verosimilitud.
5. Proporcionar una validación gráfica de la propiedad de Markov para una muestra finita.
6. Mostrar una vinculación entre la Probabilidad y la Estadística.

Cabe señalar aquí que el software empleado para realizar las simulaciones y gráficas correspondientes mostradas en esta tesis se realizaron con el programa R versión 2.3.1 en una computadora Pentium 4.

# Capítulo 1

## Teoría de Cadenas de Markov

Una clase muy importante de procesos estocásticos la constituyen los procesos de Markov. Estos tienen la característica general de que las observaciones futuras sólo dependen del presente y no del pasado.

En esta parte se analizarán las cadenas de Markov desde el punto de vista probabilístico, en donde los resultados presentados fueron tomados de las referencias: Caballero (2004), Hoel (1972) y Karlin (1975).

Se iniciará en la Sección 1.1 definiendo tales procesos, así como algunas de sus características. Luego, en la segunda parte se introducirán algunos ejemplos importantes para ilustrar la teoría. En particular, se analizará una cadena de Markov conocida como caminata aleatoria.

En las secciones posteriores se hablará de la clasificación de los estados en una cadena de Markov ya que con ello se podrá analizar el comportamiento límite del sistema, y poder concluir características y propiedades del proceso.

### 1.1 Conceptos básicos

Sea  $(\Omega, \mathcal{A}, P)$  un espacio de probabilidad y  $T$  un conjunto no vacío, finito ó numerable. Considere una sucesión de variables aleatorias (v.a)  $X_n : \Omega \longrightarrow T$ . Esto es, para cada  $n \in \mathbb{N}$  y  $\omega \in \Omega$ , tenemos que  $X_n(\omega)$  es un elemento del conjunto  $T$ . Por lo que es usual decir que “el proceso se encuentra en el estado  $i$  al tiempo  $n$ ” cuando  $X_n(\omega) = i$ . El conjunto  $T$  es llamado el espacio de estados del proceso y cabe mencionar que en la mayoría de las situaciones se trabajará con  $T = \{0, 1, 2, \dots\}$ .

**Definición 1.1** *Un proceso estocástico  $\{X_n\}_{n \in \mathbb{N}}$  es una cadena de Markov si satisface la condición,*

$$\Pr\{X_{n+1} = j \mid X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i\} = \Pr\{X_{n+1} = j \mid X_n = i\} \quad (1.1)$$

para toda  $i_0, i_1, \dots, i, j \in T$  y  $n \in \mathbb{N}$ .

A la condición (1.1) se le conoce como la *propiedad de Markov*. Una cadena de Markov, entonces, es una sucesión de variables aleatorias tal que para cualquier  $n$ , la distribución de  $X_{n+1}$  es condicionalmente independiente de  $X_0, \dots, X_{n-1}$  dado  $X_n$ .

Las probabilidades  $\Pr\{X_{n+1} = j \mid X_n = i\}$  son llamadas *Probabilidades de transición* en un paso para la cadena de Markov  $\{X_n\}_{n \in \mathbb{N}}$ . Si las probabilidades de

transición son independientes de  $n$  se dice que la *cadena de Markov es homogénea* y en este caso se denota

$$P_{ij} = \Pr\{X_{n+1} = j | X_n = i\} \quad \text{para cada } i, j \in T \text{ y } n \in \mathbb{N}. \quad (1.2)$$

Las cantidades  $P_{ij}$  satisfacen las siguientes propiedades:

$$P_{ij} \geq 0, \quad \text{para toda } i, j \in T$$

y

$$\sum_{j \in T} P_{ij} = 1, \quad i \in T$$

En este trabajo se considerarán sólo cadenas de Markov homogéneas.

Si el espacio de estados  $T = \{0, 1, 2, \dots, n\}$  es finito, las probabilidades  $P_{ij}$  pueden ser representadas en una matriz cuadrada conocida como la *Matriz de Markov* o *Matriz de Probabilidad de Transición* del proceso y se denota como  $P = [P_{i,j}]$ , esto es:

$$P = \begin{bmatrix} P_{00} & P_{01} & \cdots & P_{0n} \\ P_{10} & P_{11} & \cdots & P_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{n0} & P_{n1} & \cdots & P_{nn} \end{bmatrix}$$

El  $i$  – ésimo renglón de  $P$ ,  $i = 0, 1, 2, \dots, n$ , es la distribución de probabilidad de  $X_{n+1}$  dado que  $X_n = i$ . Son precisamente, las probabilidades de transición (1.2) las que serán los parámetros de interés a estimar vía el enfoque de verosimilitud en el Capítulo 3.

La distribución de  $X_0$  se le llama la *distribución inicial* de la cadena y se denotará por  $\Pr\{X_0 = i\} = \pi_0(i)$ ,  $i \in T$ . Esta satisface las siguientes propiedades:

$$\pi_0(i) \geq 0, \quad i \in T$$

y

$$\sum_{i=0}^{\infty} \pi_0(i) = 1.$$

La distribución conjunta de las variables aleatorias  $X_0, \dots, X_n$  se puede expresar en términos de las probabilidades de transición y la distribución inicial, lo cual se muestra en la siguiente proposición.

**Proposición 1.1** Sea  $\{X_n\}_{n \in \mathbb{N}}$  una cadena de Markov con  $i_0, i_1, \dots, i_n \in T$  y  $\Pr\{X_0 = i_0\} = \pi_0(i_0)$ , entonces

$$\Pr\{X_0 = i_0, X_1 = i_1, \dots, X_n = i_n\} = \pi_0(i_0)P_{i_0, i_1} \cdots P_{i_{n-2}, i_{n-1}}P_{i_{n-1}, i_n} \quad (1.3)$$

**Demostración.** Se realizará por inducción. Nótese que

$$\begin{aligned} \Pr\{X_0 = i_0, X_1 = i_1\} &= \Pr\{X_0 = i_0\} \Pr\{X_1 = i_1 | X_0 = i_0\} \\ &= \pi_0(i_0)P_{i_0, i_1}. \end{aligned}$$



Supóngase que (1.3) se cumple para  $n - 1$ , es decir,

$$\Pr\{X_0 = i_0, \dots, X_{n-1} = i_{n-1}\} = \pi_0(i_0)P_{i_0, i_1} \cdots P_{i_{n-2}, i_{n-1}}.$$

Se demostrará que se cumple para  $n$ . Por definición de probabilidad condicional

$$\begin{aligned} \Pr\{X_0 = i_0, \dots, X_n = i_n\} &= \Pr\{X_0 = i_0, \dots, X_{n-1} = i_{n-1}\} \cdot \\ &\quad \Pr\{X_n = i_n | X_0 = i_0, \dots, X_{n-1} = i_{n-1}\}. \end{aligned}$$

Aplicando la propiedad de Markov dada en (1.1) en la expresión anterior, se tiene que

$$\Pr\{X_0 = i_0, \dots, X_n = i_n\} = \Pr\{X_0 = i_0, \dots, X_{n-1} = i_{n-1}\}P_{i_{n-1}, i_n}.$$

Luego, por hipótesis de inducción, se obtiene que

$$\Pr\{X_0 = i_0, X_1 = i_1, \dots, X_n = i_n\} = \pi_0(i_0)P_{i_0, i_1} \cdots P_{i_{n-2}, i_{n-1}}P_{i_{n-1}, i_n}.$$

■

Con lo anterior, se observa que la *distribución conjunta* está completamente definida una vez que se tiene la matriz de probabilidad de transición y la distribución inicial  $\pi_0$  del proceso.

**Corolario 1.2** Sea  $n \in \mathbb{N}$  y  $i_0, i_1, \dots, i_n \in T$ , entonces

$$\Pr\{X_1 = i_1, \dots, X_n = i_n | X_0 = i_0\} = P_{i_0, i_1} \cdots P_{i_{n-1}, i_n}.$$

**Demostración.** Por definición de probabilidad condicional se tiene que

$$\Pr\{X_1 = i_1, \dots, X_n = i_n | X_0 = i_0\} = \frac{\Pr\{X_0 = i_0, X_1 = i_1, \dots, X_n = i_n\}}{\Pr\{X_0 = i_0\}}.$$

Nótese que el numerador en la expresión anterior es la distribución conjunta. Por lo tanto, de la Proposición 1.1 se tiene que

$$\begin{aligned} \Pr\{X_1 = i_1, \dots, X_n = i_n | X_0 = i_0\} &= \frac{\pi_0(i_0) \cdot P_{i_0, i_1} \cdots P_{i_{n-2}, i_{n-1}}P_{i_{n-1}, i_n}}{\pi_0(i_0)} \\ &= P_{i_0, i_1} \cdots P_{i_{n-1}, i_n}. \end{aligned}$$

■

**Definición 1.2** Dada una cadena de Markov  $\{X_n\}_{n \in \mathbb{N}}$  y probabilidades de transición (función de transición) en un paso  $P_{ij}$ , se define la **probabilidad de transición de orden  $n$** ,  $n \in \mathbb{N}$ ,  $P_{ij}^{(n)}$ , como la probabilidad de que el proceso vaya del estado  $i$  al estado  $j$  en  $n$  transiciones:

$$P_{ij}^{(n)} = \Pr\{X_{m+n} = j | X_m = i\}, \forall i, j \in T \text{ y } m \in \mathbb{N} \quad (1.4)$$

y

$$P_{ij}^{(0)} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}. \quad (1.5)$$

Cuando el espacio de estados  $T$  es finito entonces se denotará  $P^{(n)} = [P_{ij}^{(n)}]$  como la matriz de transición en  $n - \text{pasos}$ . A continuación se verán algunas propiedades para las probabilidades de transición de orden  $n$ ,  $P_{ij}^{(n)}$ .

**Teorema 1.3** *Las probabilidades de transición de orden  $n$  satisfacen que*

$$P_{ij}^{(n)} = \sum_{k=0}^{\infty} P_{ik} P_{kj}^{(n-1)}.$$

**Demostración.** Se tiene la siguiente cadena de igualdades:

$$\begin{aligned} P_{ij}^{(n)} &= \Pr\{X_n = j \mid X_0 = i\} \\ &= \sum_{k=0}^{\infty} \Pr\{X_n = j, X_1 = k \mid X_0 = i\} \\ &= \sum_{k=0}^{\infty} \frac{\Pr\{X_n = j, X_1 = k, X_0 = i\}}{\Pr\{X_0 = i\}} \cdot \frac{\Pr\{X_1 = k, X_0 = i\}}{\Pr\{X_1 = k, X_0 = i\}} \\ &= \sum_{k=0}^{\infty} \Pr\{X_1 = k \mid X_0 = i\} \Pr\{X_n = j \mid X_1 = k, X_0 = i\} \\ &= \sum_{k=0}^{\infty} P_{i,k} P_{k,j}^{(n-1)}. \end{aligned}$$

■

Cuando el espacio de estados  $T$  es finito, se tiene que  $P$  es la matriz de las probabilidades de transición en un paso, mientras que  $P^{(n)}$  es la matriz de las probabilidades de transición de orden  $n$ . Se denotará por  $P^n$  al producto  $n - \text{veces}$  de  $P$ , que significa

$$P^n = \overbrace{PPP \cdots P}^{n - \text{factores}}$$

**Lema 1.4** *El producto  $n - \text{veces}$  de  $P$  contiene en sus entradas a la probabilidades de transición de orden  $n$ , es decir,  $P^{(n)} = P^n$ .*

**Demostración.** Se realizará por inducción matemática. Para  $n = 2$ ,

$$P^{(2)} = [P_{ij}^{(2)}],$$

donde  $P_{ij}^{(2)} = \sum_{k=0}^{\infty} P_{ik} P_{kj}$  con  $i, j \in T$ . De la definición de producto de matrices se tiene que

$$P^2 = PP = \left[ \sum_{k=0}^{\infty} P_{ik} P_{kj} \right].$$

Entonces  $P^{(2)} = P^2$ . Ahora, supongase que se cumple para  $n - 1$ ,

$$P^{(n-1)} = \left[ P_{ij}^{(n-1)} \right] = P^{n-1},$$

donde  $P_{ij}^{(n-1)} = \sum_{k=0}^{\infty} P_{ik} P_{kj}^{(n-2)}$ . Se demostrará que es válido para  $n - \text{pasos}$ , es decir que  $P^{(n)} = \left[ P_{ij}^{(n)} \right] = P^n$ . Por el Teorema 1.3 se cumple que

$$P_{ij}^{(n)} = \sum_{k=0}^{\infty} P_{ik} P_{kj}^{(n-1)}.$$

Así,

$$P^n = P P^{n-1} = P \left[ P_{ij}^{(n-1)} \right] = [P_{ij}] \left[ P_{ij}^{(n-1)} \right] = \left[ \sum_{k=0}^{\infty} P_{ik} P_{kj}^{(n-1)} \right] = P^{(n)}.$$

Por lo tanto, en la matriz  $P^n$  se encuentran las probabilidades de transición de orden  $n$ . ■

**Teorema 1.5** (*Chapman – Kolmogorov*): Para una cadena de Markov  $\{X_n\}_{n \in \mathbb{N}}$  con espacio de estados  $T$  y para todo  $n, m \in \mathbb{N}$  y toda pareja  $i, j \in T$  se cumple que

$$P_{ij}^{(n+m)} = \Pr\{X_{n+m} = j \mid X_0 = i\} = \sum_{k \in T} P_{ik}^{(n)} P_{kj}^{(m)}.$$

**Demostración.** Se tiene que

$$\begin{aligned} P_{ij}^{(n+m)} &= \Pr\{X_{n+m} = j \mid X_0 = i\} \\ &= \frac{\Pr\{X_{n+m} = j, X_0 = i\}}{\Pr\{X_0 = i\}} \\ &= P(X_{n+m} = j, \cup_{k \in T} \{X_n = k\}, X_0 = i) \cdot \frac{1}{\Pr\{X_0 = i\}} \\ &= \sum_{k \in T} \frac{\Pr\{X_{n+m} = j, X_n = k, X_0 = i\}}{\Pr\{X_0 = i\}} \\ &= \sum_{k \in T} \frac{\Pr\{X_{n+m} = j \mid X_n = k, X_0 = i\} \cdot \Pr\{X_n = k, X_0 = i\}}{\Pr\{X_0 = i\}} \\ &= \sum_{k \in T} \Pr\{X_{n+m} = j \mid X_n = k\} \Pr\{X_n = k \mid X_0 = i\} \\ &= \sum_{k \in T} P_{ik}^{(n)} P_{kj}^{(m)}. \end{aligned}$$

■

## 1.2 Ejemplos de cadenas de Markov

A continuación se presentan algunos ejemplos de cadenas de Markov homogéneas.

**Ejemplo 1.2.1** Sean  $X_0, X_1, \dots, X_n$  variables aleatorias independientes con distribución común

$$\Pr\{X_n = i\} = \begin{cases} a_j, & i = 0, 1, \dots \\ 0, & i < 0 \end{cases}.$$

Como  $X_{n+1}$  es independiente de  $X_0, \dots, X_n$ , por definición de probabilidad condicional se tiene que:

$$\begin{aligned} \Pr\{X_{n+1} = j | X_0 = i_0, \dots, X_n = i\} &= \frac{\Pr\{X_{n+1} = j, X_0 = i_0, \dots, X_n = i\}}{\Pr\{X_0 = i_0, \dots, X_n = i\}} \\ &= \frac{\Pr\{X_{n+1} = j\} \Pr\{X_0 = i_0, \dots, X_n = i\}}{\Pr\{X_0 = i_0, \dots, X_n = i\}} \\ &= \Pr\{X_{n+1} = j\} = a_j = \Pr\{X_{n+1} = j | X_n = i\}. \end{aligned}$$

Por tanto,  $\{X_n\}_{n \in \mathbb{N}}$  es una cadena de Markov con espacio de estados  $T = \{0, 1, 2, \dots\}$  y probabilidad de transición  $P_{ij} = a_j$ , esto es:

$$P = \begin{bmatrix} a_0 & a_1 & a_2 & \cdots \\ a_0 & a_1 & a_2 & \cdots \\ a_0 & a_1 & a_2 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Recíprocamente, si  $\{X_n\}_{n \in \mathbb{N}}$  es una cadena de Markov cuya matriz de transición tiene todos sus renglones idénticos, es decir  $P_{ij} = a_j$ , entonces las variables aleatorias  $\{X_n\}_{n \in \mathbb{N}}$  son independientes e idénticamente distribuidas, con distribución común  $a_j, j \geq 0$ .

**Demostración.** Para  $n \geq 1$ ,

$$\begin{aligned} \Pr\{X_n = j\} &= \sum_{i \in T} \Pr\{X_n = j, X_{n-1} = i\} \\ &= \sum_{i \in T} \Pr\{X_n = j | X_{n-1} = i\} \Pr\{X_{n-1} = i\} \\ &= \sum_{i \in T} P_{ij} \Pr\{X_{n-1} = i\} \\ &= a_j \sum_{i \in T} \Pr\{X_{n-1} = i\} = a_j. \end{aligned}$$

De la Proposición 1.1 se tiene que

$$\Pr\{X_0 = i_0, X_1 = i_1, \dots, X_n = i_n\} = \pi_0(i_0) \cdot P_{i_0, i_1} \cdots P_{i_{n-2}, i_{n-1}} P_{i_{n-1}, i_n}.$$

Entonces,

$$\begin{aligned}\Pr\{X_0 = i_0, X_1 = i_1, \dots, X_n = i_n\} &= \Pr\{X_0 = i_0\}P_{i_0, i_1} \dots P_{i_{n-1}, i_n} \\ &= \Pr\{X_0 = i_0\}a_{i_1} \dots a_{i_n} \\ &= \Pr\{X_0 = i_0\} \Pr\{X_1 = i_1\} \dots \Pr\{X_n = i_n\}.\end{aligned}$$

Por lo tanto, las variables aleatorias  $X_0, X_1, \dots, X_n$  son independientes e idénticamente distribuidas. ■

**Ejemplo 1.2.2 (Caminatas Aleatorias)** Sean  $\xi_1, \xi_2, \dots, \xi_n$  variables aleatorias independientes con densidad común  $\Pr\{\xi_n = i\} = a_i$  y espacio de estados  $T = \{\dots, -2, -1, 0, 1, 2, \dots\}$ . Sea  $X_0$  una variable aleatoria independiente de las variables  $\xi_1, \xi_2, \dots, \xi_n$ . Considérese la sucesión de sumas parciales de  $\xi_n$ , es decir,

$$X_n = X_0 + \xi_1 + \dots + \xi_n, \quad n \in \mathbb{N}.$$

Al proceso  $\{X_n\}_{n \in \mathbb{N}}$  se le llama caminata aleatoria.

Se demostrará que  $\{X_n\}_{n \in \mathbb{N}}$  es una cadena de Markov con probabilidad de transición  $P_{ij} = a_{j-i}$  con  $i, j \in T$ .

Puesto que  $X_{n+1} = X_n + \xi_{n+1}$  y por independencia de  $X_k$ ,  $k = 0, \dots, n$  y  $\xi_{n+1}$  se tiene que:

$$\begin{aligned}\Pr\{X_{n+1} = j | X_0 = i_0, \dots, X_n = i\} &= \Pr\{\xi_{n+1} = j - X_n | X_0 = i_0, \dots, X_n = i\} \\ &= \Pr\{\xi_{n+1} = j - i | X_0 = i_0, \dots, X_n = i\} \\ &= \frac{\Pr\{\xi_{n+1} = j - i\} \Pr\{X_0 = i_0, \dots, X_n = i\}}{\Pr\{X_0 = i_0, \dots, X_n = i\}} \\ &= \Pr\{\xi_{n+1} = j - i\} = a_{j-i} \\ &= \Pr\{X_{n+1} = j | X_n = i\}\end{aligned}$$

Por lo tanto,  $P_{ij} = a_{j-i}$ ,  $i, j \in T$ .

Este ejemplo se puede interpretar de la siguiente manera. Supóngase que una partícula se mueve de acuerdo a esta cadena de Markov. Si la partícula inicialmente se encuentra en un estado  $i$ , esta salta al estado  $j$  con probabilidad  $a_{j-i}$ .

En el caso particular de que  $\Pr\{\xi_i = 1\} = p_i$ ,  $\Pr\{\xi_i = -1\} = q_i$  y  $\Pr\{\xi_i = 0\} = r_i$ , con  $p_i > 0$ ,  $q_i > 0$ ,  $r_i \geq 0$ , y  $p_i + q_i + r_i = 1$ ,  $\{X_n\}$  es una cadena de Markov con espacio de estados  $T = \{0, \pm 1, \pm 2, \dots\}$  y probabilidades de transición en un paso dadas por

$$P_{ij} = \begin{cases} p_i, & j = i + 1 \\ q_i, & j = i - 1 \\ r_i, & j = i \\ 0 & \text{otro caso.} \end{cases}$$

Por tanto, dado que la partícula esta en una posición inicial  $i$ , en una transición, sólo puede quedarse en el estado  $j = i$ , ó moverse hacia adelante o hacia atrás, es

decir llegar a los estados adyacentes de su posición original  $j = i + 1$  ó  $j = i - 1$ , respectivamente.

Un ejemplo muy conocido de una caminata aleatoria en una dimensión es la cadena de *Nacimiento y Muerte*. Esta cadena tiene espacio de estados finito  $T = \{0, 1, \dots, d\}$  y matriz de transición  $P$ :

$$P = \begin{bmatrix} r_0 & p_0 & 0 & 0 & \cdots & 0 \\ q_1 & r_1 & p_1 & 0 & \cdots & 0 \\ 0 & q_2 & r_2 & p_2 & \cdots & 0 \\ 0 & 0 & q_3 & r_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & 0 & \cdots & r_d \end{bmatrix}.$$

La frase “Nacimiento y Muerte”, se deriva de las aplicaciones en las cuales el estado es la población. El nacimiento corresponde pasar del estado  $i$  a la posición  $i + 1$ , mientras que si al encontrarse en  $i$  en una transición, alcanza al estado  $i - 1$ , se dice que corresponde a muerte.

### 1.3 Clasificación de estados de una Cadena de Markov

Se enunciarán algunas definiciones básicas para el estudio de cadenas de Markov como: estados comunicantes, estados recurrentes, transitorios, conjuntos irreducibles, entre otros. Estos conceptos serán de gran importancia en el estudio del comportamiento límite de una cadena de Markov, que también se abordará en este capítulo.

#### 1.3.1 Estados Recurrentes y Transitorios

**Definición 1.3** Sea  $\{X_n\}_{n \in \mathbb{N}}$  una cadena de Markov con espacio de estados  $T$  y probabilidades de transición  $P_{ij}$ ,  $i, j \in T$ . Se dice que

- (i) Un estado  $i$  se **comunica** al estado  $j$  si existe algún entero  $n \geq 0$ , tal que  $P_{ij}^n > 0$  y se denota  $i \rightarrow j$ .
- (ii) Los estados  $i$  y  $j$  son **comunicantes** entre sí, si  $i \rightarrow j$  y  $j \rightarrow i$ , lo cual se denota  $i \longleftrightarrow j$ .

Decir que un estado  $i$  se comunica con un estado  $j$ , indica que existe probabilidad positiva de pasar del estado  $i$  al  $j$  en un número finito de pasos.

**Definición 1.4** Una cadena de Markov  $\{X_n\}_{n \in \mathbb{N}}$  es irreducible si todos sus estados son comunicantes.

Uno de los conceptos principales que se ocupa al momento de estudiar cadenas de Markov es el primer tiempo de entrada que se introducirá a continuación.



**Definición 1.5** Sea  $\{X_n\}_{n \in \mathbb{N}}$  una cadena de Markov, con espacio de estados  $T$ . El primer tiempo de entrada al estado  $i$  se define como

$$R_i = \inf\{n \geq 1 : X_n = i\} \quad (1.6)$$

y

$$R_i = \infty \text{ si } X_n \neq i, \text{ para todo } n > 0. \quad (1.7)$$

En el siguiente resultado se verá una ecuación que relaciona el primer tiempo de entrada con las probabilidades de transición.

**Proposición 1.6** Sea  $\{X_n\}_{n \in \mathbb{N}}$  una cadena de Markov con espacio de estados  $T$  y probabilidad de transición  $P$ . Entonces se cumple que

$$P_{ij}^n = \sum_{m=1}^n \Pr\{R_j = m | X_0 = i\} P_{jj}^{n-m}$$

con  $n \geq 1$  y  $i, j \in T$ .

**Demostración.** Nótese que los eventos  $\{R_j = m, X_n = j\}$  con  $1 \leq m \leq n$  son disjuntos y que  $\{X_n = j\} = \cup_{m=1}^n \{R_j = m, X_n = j\}$ . Entonces

$$\begin{aligned} P_{ij}^n &= \sum_{m=1}^n \Pr\{R_j = m, X_n = j | X_0 = i\} \\ &= \sum_{m=1}^n \Pr\{R_j = m | X_0 = i\} \Pr\{X_n = j | X_0 = i, R_j = m\} \\ &= \sum_{m=1}^n \Pr\{R_j = m | X_0 = i\} \Pr\{X_n = j | X_0 = i, X_1 \neq j, \dots, X_{m-1} \neq j, X_m = j\} \\ &= \sum_{m=1}^n \Pr\{R_j = m | X_0 = i\} P_{jj}^{n-m}. \end{aligned}$$

■

**Definición 1.6** Sea  $\{X_n\}_{n \in \mathbb{N}}$  una cadena de Markov con espacio de estados  $T$  y probabilidades de transición  $P$ . Sea

$$f_{ij} = \Pr\{R_j < \infty | X_0 = i\} = \sum_{n=1}^{\infty} \Pr\{R_j = n | X_0 = i\}, \quad i, j \in T. \quad (1.8)$$

Entonces  $f_{ij}$  denota la probabilidad de que una cadena de Markov iniciando en  $i$  vaya a un estado  $j$  en un tiempo positivo. En particular,  $f_{ii}$  denota la probabilidad de regresar al estado  $i$ .

**Definición 1.7** Un estado  $i \in T$ , se le llama

- (i) Estado **recurrente** si  $f_{ii} = 1$ .
- (ii) Estado **transitorio** si  $f_{ii} < 1$ .
- (iii) Estado **absorbente** si  $P_{ii} = 1$ .

Se puede observar en esta definición que cuando el estado es recurrente, con probabilidad uno la cadena se regresa al mismo estado, mientras que si el estado es transitorio se tiene probabilidad positiva de no regresar. Es importante señalar que no sólo es de interés conocer si la cadena regresa o no al mismo estado sino también saber el número esperado de visitas, y se puede intuir que el número esperado de visitas para un estado transitorio es finito, en contraste cuando un estado es recurrente. Esto se mostrará formalmente más adelante.

**Definición 1.8** Sea una cadena de Markov  $\{X_n\}_{n \in \mathbb{N}}$ , con espacio de estados  $T$ . Se define

- (i) El número total de visitas al estado  $i$  como

$$N(i) = \sum_{n=1}^{\infty} \mathbb{I}\{X_n = i\}, \quad \text{donde } \mathbb{I}\{X_n = i\} = \begin{cases} 1, & X_n = i \\ 0, & \text{otro caso} \end{cases}. \quad (1.9)$$

- (ii) El número esperado de visitas a  $j$ , dado el estado inicial  $i$ , como

$$G_{ij} = E[N(j) | X_0 = i]. \quad (1.10)$$

Puesto que el evento  $\{N(j) \geq 1\}$  es equivalente a  $\{R_j < \infty\}$ , entonces

$$\Pr\{N(j) \geq 1 | X_0 = i\} = \Pr\{R_j < \infty | X_0 = i\} = f_{ij}. \quad (1.11)$$

**Lema 1.7** Sean  $i, j \in T$ , entonces se cumple que:

- (i)  $\Pr\{N(j) = 0 | X_0 = i\} = 1 - f_{ij}$ .
- (ii)  $\Pr\{N(j) \geq m | X_0 = i\} = f_{ij} f_{jj}^{m-1}$ .
- (iii)  $\Pr\{N(j) = m | X_0 = i\} = f_{ij} f_{jj}^{m-1} (1 - f_{jj})$ .

**Demostración.** Para demostrar (i) se utilizará la expresión dada en (1.11)

$$\Pr\{N(j) = 0 | X_0 = i\} = 1 - \Pr\{N(j) \geq 1 | X_0 = i\} = 1 - f_{ij}. \quad (1.12)$$

Ahora se demostrará (ii). La probabilidad de que una cadena de Markov que inició en un estado  $i$ , su primer visita al estado  $j$  sea en un tiempo  $m$  y la siguiente visita a  $j$  sea en un tiempo  $n$  más tarde, con  $1 \leq m < n$ , se puede denotar como

$$\Pr\{X_1 \neq j, \dots, X_{m-1} \neq j, X_m = j, X_{m+1} \neq j, \dots, X_{m+n} = j | X_0 = i\}$$

$$\begin{aligned}
&= \Pr\{X_{m+1} \neq j, \dots, X_{m+n-1} \neq j, X_{m+n} = j \mid X_0 = i, X_1 \neq j, \dots, X_{m-1} \neq j, X_m = j\} \\
&\quad \cdot \Pr\{X_1 \neq j, \dots, X_{m-1} \neq j, X_m = j \mid X_0 = i\} \\
&= \Pr\{X_1 \neq j, \dots, X_{n-1} \neq j, X_n = j \mid X_0 = j\} \Pr\{X_1 \neq j, \dots, X_{m-1} \neq j, X_m = j \mid X_0 = i\} \\
&= \Pr\{R_j = m \mid X_0 = i\} \Pr\{R_j = n \mid X_0 = j\}.
\end{aligned}$$

Por lo tanto,

$$\begin{aligned}
\Pr\{N(j) \geq 2 \mid X_0 = i\} &= \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \Pr\{R_j = m \mid X_0 = i\} \Pr\{R_j = n \mid X_0 = j\} \\
\sum_{m=1}^{\infty} \Pr\{R_j = m \mid X_0 = i\} \sum_{n=1}^{\infty} \Pr\{R_j = n \mid X_0 = j\} &= f_{ij} f_{jj}.
\end{aligned}$$

De manera similar se tiene que la probabilidad de que la cadena de Markov con estado inicial  $i$  visite al estado  $j$  por primera vez en un tiempo  $k_1$ , y la siguiente visita sea  $k_2$  unidades después, y la tercera visita sea  $k_3$  unidades después de la segunda, y así sucesivamente la  $m$ -ésima visita sea  $k_m$  unidades después de la  $(m-1)$ -ésima visita, está dada por:

$$\Pr\{R_j = k_1 \mid X_0 = i\} \Pr\{R_j = k_1 \mid X_0 = j\} \dots \Pr\{R_j = k_m \mid X_0 = j\}.$$

Entonces,

$$\begin{aligned}
\Pr\{N(j) \geq m \mid X_0 = i\} &= \sum_{k_1=1}^{\infty} \sum_{k_2=1}^{\infty} \dots \sum_{k_m=1}^{\infty} \Pr\{R_j = k_1 \mid X_0 = i\} \dots \Pr\{R_j = k_m \mid X_0 = j\} \\
&= \sum_{k_1=1}^{\infty} \Pr\{R_j = k_1 \mid X_0 = i\} \sum_{k_2=1}^{\infty} \Pr\{R_j = k_2 \mid X_0 = j\} \dots \sum_{k_m=1}^{\infty} \Pr\{R_j = k_m \mid X_0 = j\} \\
&= f_{ij} f_{jj} \dots f_{jj} = f_{ij} f_{jj}^{m-1}.
\end{aligned}$$

Así de lo anterior se cumple que

$$\Pr\{N(j) \geq m \mid X_0 = i\} = f_{ij} f_{jj}^{m-1}. \quad (1.13)$$

Para la demostración (iii), nótese que

$$\begin{aligned}
\Pr\{N(j) = m \mid X_0 = i\} &= \Pr\{N(j) \geq m \mid X_0 = i\} - \Pr\{N(j) \geq m+1 \mid X_0 = i\} \\
&= f_{ij} f_{jj}^{m-1} - f_{ij} f_{jj}^m = f_{ij} f_{jj}^{m-1} (1 - f_{jj}).
\end{aligned} \quad (1.14)$$

Con esto queda demostrado el Lema. ■

En el siguiente resultado se mostrará que el número esperado de visitas a un estado puede ser expresado en términos de las probabilidades de transición.

**Lema 1.8** *Sea una cadena de Markov  $\{X_n\}_{n \in \mathbb{N}}$ , con espacio de estados  $T$  y matriz de transición  $P$ . Se tiene que*

$$G_{ij} = E[N(j) \mid X_0 = i] = \sum_{n=1}^{\infty} P_{ij}^n.$$

**Demostración.** De la Definición 1.8 (i) y (ii) se tiene que

$$\begin{aligned}
 E[N(j) | X_0 = i] &= E \left[ \sum_{n=1}^{\infty} \mathbb{I}\{X_n = j\} | X_0 = i \right] \\
 &= \sum_{n=1}^{\infty} E [\mathbb{I}\{X_n = j\} | X_0 = i] \\
 &= \sum_{n=1}^{\infty} \Pr\{X_n = j | X_0 = i\} \\
 &= \sum_{n=1}^{\infty} P_{ij}^n.
 \end{aligned}$$

■

**Proposición 1.9** Sea una cadena de Markov  $\{X_n\}_{n \in \mathbb{N}}$ , con espacio de estados  $T$  y matriz de transición  $P$ .

(a) Sea  $j$  un estado transitorio. Entonces

(i)  $\Pr\{N(j) < \infty | X_0 = i\} = 1$  para todo  $i \in T$ .

(ii)  $G_{ij} = \frac{f_{ij}}{1 - f_{ii}} < \infty$  para toda  $i \in T$ .

(b) Sea  $j$  un estado recurrente. Entonces

(i)  $\Pr\{N(j) = \infty | X_0 = i\} = \Pr\{R_j < \infty | X_0 = i\} = f_{ij}, i \in T$ .

(ii)  $\Pr\{N(j) = \infty | X_0 = j\} = 1$  y  $G_{jj} = \infty$ .

(iii) Si  $f_{ij} = 0$ , entonces  $G_{ij} = 0$ , mientras que si  $f_{ij} > 0$ , entonces  $G_{ij} = \infty$ .

Si  $j$  es un estado transitorio, entonces no importa donde inicia la cadena, con probabilidad uno el número de visitas al estado  $j$  es finito, y el número esperado de visitas también es finito. Por el contrario, si  $j$  es recurrente con probabilidad uno la cadena regresa al estado  $j$  una infinidad de veces.

**Demostración.** (a) Suponga que  $j$  es un estado de transitorio. Entonces,  $0 \leq f_{jj} < 1$ . Se sigue de (1.13) que

$$\begin{aligned}
 \Pr\{N(j) = \infty | X_0 = i\} &= \lim_{m \rightarrow \infty} \Pr\{N(j) \geq m | X_0 = i\} \\
 &= \lim_{m \rightarrow \infty} f_{ij} f_{jj}^{m-1} = 0,
 \end{aligned}$$

lo cual implica que

$$\Pr\{N(j) < \infty | X_0 = i\} = 1.$$

De la Definición 1.8,  $G_{ij}$  representa el número esperado de visitas al estado  $j$  si  $X_0 = i$ ,

$$\begin{aligned} G_{ij} &= E[N(j) | X_0 = i] \\ &= \sum_{m=1}^{\infty} m \Pr\{N(j) = m | X_0 = i\} \\ &= \sum_{m=1}^{\infty} m f_{ij} f_{jj}^{m-1} (1 - f_{jj}) \\ &= f_{ij} (1 - f_{jj}) \sum_{m=1}^{\infty} m f_{jj}^{m-1}. \end{aligned}$$

Donde la tercera igualdad se sigue de (1.14). Luego sustituyendo  $t = f_{jj}$  en la serie de potencias

$$\sum_{m=1}^{\infty} m t^{m-1} = \frac{1}{(1-t)^2},$$

se tiene que,

$$\begin{aligned} G_{ij} &= f_{ij} (1 - f_{jj}) \frac{1}{(1 - f_{jj})^2} \\ &= \frac{f_{ij}}{1 - f_{jj}}. \end{aligned}$$

Por lo tanto,

$$G_{ij} = \frac{f_{ij}}{1 - f_{jj}} < \infty.$$

(b) Ahora sea  $j$  un estado recurrente. Entonces  $f_{jj} = 1$  y de (1.13) se obtiene que

$$\begin{aligned} \Pr\{N(j) = \infty | X_0 = i\} &= \lim_{m \rightarrow \infty} \Pr\{N(j) \geq m | X_0 = i\} \\ &= \lim_{m \rightarrow \infty} f_{ij} f_{jj}^{m-1} = f_{ij}. \end{aligned}$$

En particular si  $i = j$ ,  $\Pr\{N(j) = \infty | X_0 = j\} = 1$ . Si una variable aleatoria no negativa tiene probabilidad positiva de ser infinita, su esperanza es infinita. Entonces,

$$G_{jj} = E[N(j) | X_0 = j] = \infty.$$

Si  $f_{ij} = 0$  entonces  $\Pr\{R_j = m | X_0 = i\} = 0$  para todos los enteros positivos  $m$ . Así, por la Proposición 1.6 se tiene que

$$P_{ij}^n = \sum_{m=1}^n \Pr\{R_j = m | X_0 = i\} P_{jj}^{n-m} \text{ con } n \geq 1 \text{ y } i, j \in T.$$

Por tanto

$$P_{ij}^n = \sum_{m=1}^{\infty} 0 = 0.$$

Entonces por el Lema 1.8,

$$G_{ij} = \sum_{n=1}^{\infty} P_{ij}^n = 0.$$

Ahora, si  $f_{ij} > 0$  entonces  $\Pr\{N(j) = \infty | X_0 = i\} = f_{ij} > 0$  y por tanto  $G_{ij} = E[N(j) | X_0 = i] = \infty$ . ■

**Proposición 1.10** *Sea  $i$  un estado recurrente tal que  $i$  se comunica con  $j$ , entonces  $j$  es recurrente y  $f_{ij} = f_{ji} = 1$ .*

La proposición anterior implica que la propiedad de ser recurrente es una propiedad de clase, es decir, en una cadena irreducible, si un estado es recurrente entonces todos los estados son recurrentes.

**Demostración.** Supóngase que  $i \neq j$ , pues caso contrario no hay nada que demostrar. Puesto que  $i \rightarrow j$

$$\Pr\{R_j < \infty | X_0 = i\} = f_{ij} > 0,$$

entonces se verá que

$$\Pr\{R_j = n | X_0 = i\} > 0$$

para algún entero positivo  $n$ . Sea

$$n_0 = \min\{n \geq 1 : \Pr\{R_j = n | X_0 = i\} > 0\}.$$

Se sigue de lo anterior y de la Proposición 1.6 que  $P_{ij}^{n_0} > 0$  y

$$P_{ij}^m = 0, 1 \leq m \leq n_0. \quad (1.15)$$

Como  $P_{ij}^{n_0} > 0$ , entonces existen estados  $j_1, \dots, j_{n_0-1}$  tales que

$$\Pr\{X_1 = j_1, \dots, X_{n_0-1} = j_{n_0-1}, X_{n_0} = j | X_0 = i\} = P_{ij_1} \dots P_{j_{n_0-1}j} > 0.$$

Ninguno de los estados  $j_1, \dots, j_{n_0-1}$  son iguales a  $i$  o  $j$ , pues si fueran iguales existiría probabilidad positiva de ir de  $i$  a  $j$  en menos de  $n_0$  pasos, lo que contradice a (1.15). Se demostrará por contradicción que  $f_{ji} = 1$ , es decir suponemos que  $f_{ji} < 1$ . Entonces una cadena de Markov que comienza en  $j$  tiene probabilidad positiva  $1 - f_{ji}$  de nunca alcanzar el estado  $i$ . Más aún, una cadena de Markov que comienza en  $i$  tiene la probabilidad positiva de visitar los estados  $j_1, \dots, j_{n_0-1}, j$  sucesivamente en las primeras  $n_0$  transiciones y nunca regresar a  $i$  después de ese tiempo. Entonces

$$\begin{aligned} & \Pr\{X_1 = j_1, \dots, X_{n_0-1} = j_{n_0-1}, X_{n_0} = j, X_{n_0+k} \neq i \forall k \geq 1 | X_0 = i\} \\ &= \Pr\{X_{n_0+k} \neq i, \forall k \geq 1 | X_0 = i, X_1 = j_1, \dots, X_{n_0-1} = j_{n_0-1}, X_{n_0} = j\} \\ & \quad \Pr\{X_1 = j_1, \dots, X_{n_0-1} = j_{n_0-1}, X_{n_0} = j | X_0 = i\} \\ &= \Pr\{R_i = \infty | X_0 = j\} \cdot \Pr\{X_1 = j_1, \dots, X_{n_0-1} = j_{n_0-1}, X_{n_0} = j | X_0 = i\} \\ & \quad = P_{ij_1} \dots P_{j_{n_0-1}j} (1 - f_{ji}) > 0. \end{aligned}$$



Si lo anterior sucede, la cadena de Markov nunca regresa a  $i$  para cualquier  $n \geq 1$ , lo cual es una contradicción pues se supone que  $i$  es recurrente, por lo tanto  $f_{ji} = 1$ . Entonces, existe un entero positivo  $n_1$  tal que  $P_{ji}^{n_1} > 0$ .

Ahora,

$$\begin{aligned} P_{jj}^{n_1+n+n_0} &= \Pr\{X_{n_1+n+n_0} = j | X_0 = j\} \\ &\geq \Pr\{X_{n_1} = i, X_{n_1+n} = i, X_{n_1+n+n_0} = j | X_0 = j\} \\ &= P_{ji}^{n_1} P_{ii}^n P_{ij}^{n_0}. \end{aligned}$$

Por tanto,

$$\begin{aligned} G_{jj} &\geq \sum_{n=n_1+1+n_0}^{\infty} P_{jj}^n \\ &= \sum_{n=1}^{\infty} P_{jj}^{n_1+n+n_0} \\ &\geq P_{ji}^{n_1} P_{ij}^{n_0} \sum_{n=1}^{\infty} P_{ii}^n \\ &= P_{ji}^{n_1} P_{ij}^{n_0} G_{ii} = \infty, \end{aligned}$$

lo cual indica que  $j$  es un estado recurrente. Puesto que  $j$  es recurrente y como  $j$  se comunica con  $i$  entonces por la primera parte de la demostración se tiene que  $f_{ij} = 1$ . ■

**Definición 1.9** Una cadena de Markov  $\{X_n\}_{n \in \mathbb{N}}$  es llamada una **cadena de Tránsito** si todos sus estados son transitorios y una **cadena recurrente** si todos sus estados son recurrentes.

**Lema 1.11** En una cadena de Markov  $\{X_n\}_{n \in \mathbb{N}}$  con espacio de estados  $T$  finito existe al menos un estado recurrente y por lo tanto no puede ser una cadena de tránsito.

**Demostración.** Nótese que si  $j$  es un estado transitorio, del Lema 1.8 y de la Proposición 1.9a (ii), se tiene que

$$G_{ij} = \sum_{n=1}^{\infty} P_{ij}^n < \infty \text{ para } i \in T.$$

Entonces

$$\lim_{n \rightarrow \infty} P_{ij}^n = 0, i \in T. \quad (1.16)$$

Supóngase que todos los estados son transitorios. Puesto que  $T$  es finito y de (1.16) se tiene que

$$\begin{aligned} 0 &= \sum_{i \in T} \lim_{n \rightarrow \infty} P_{ij}^n = \lim_{n \rightarrow \infty} \sum_{i \in T} P_{ij}^n \\ &= \lim_{n \rightarrow \infty} \Pr\{X_n \in T | X_0 = i\} \\ &= \lim_{n \rightarrow \infty} 1 = 1, \end{aligned}$$

lo cual es una contradicción. Por lo tanto no todos los estados son transitorios. ■

**Corolario 1.12** *Si la cadena de Markov  $\{X_n\}_{n \in \mathbb{N}}$  es irreducible, entonces todos los estados son recurrentes o todos son transitorios.*

**Demostración.** Sea  $j$  un estado recurrente, entonces por ser la cadena irreducible los estados  $i$  y  $j$  son comunicables ( $i \longleftrightarrow j$ ). Luego por la Proposición 1.10 se tiene que  $i$  es un estado recurrente.

Ahora sea  $j$  un estado transitorio y suponga que  $i$  es un estado recurrente. Además  $i \longleftrightarrow j$ , de la Proposición 1.10, se tiene que  $j$  es un estado recurrente lo cual es una contradicción. Por lo tanto  $i$  es necesariamente un estado transitorio. ■

**Corolario 1.13** *Si  $\{X_n\}_{n \in \mathbb{N}}$  es una cadena de Markov irreducible con un número finito de estados, entonces cada estado es recurrente.*

**Demostración.** Del Lema 1.11 se tiene que existe al menos un estado  $i$  recurrente. Entonces por el corolario anterior todos los estados son recurrentes.

Sean  $i, j$  estados de la cadena, entonces por hipótesis todos los estados se comunican y, del Lema 1.11, existe al menos un estado  $i$  recurrente. Como  $i$  se comunica con  $j$ , por la Proposición 1.10 tenemos que  $j$  es recurrente. De manera similar se realiza para cada uno de los estado, por lo tanto cada estado es recurrente. ■

Entonces una cadena de Markov irreducible con un número finito de estados no puede ser transitoria. A continuación se verá un ejemplo que indica los estados recurrentes y transitorios en una cadena de Markov con espacio de estados finito.

**Ejemplo 1.1** *Sea una cadena de Markov con espacio de estados  $T = \{0, 1, 2, 3, 4, 5\}$  y matriz de transición*

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{5} & \frac{2}{5} & \frac{1}{5} & 0 & \frac{1}{5} \\ 0 & 0 & 0 & \frac{1}{6} & \frac{1}{3} & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{3}{4} \end{bmatrix}.$$

Se determinará cuáles estados son recurrentes o transitorios. Para ello se verificará cuáles estados se comunican con otros. De la matriz  $P$  se observa que  $P_{14} = 0$ , pero  $P_{14}^3 = P_{12}P_{23}P_{34} = (1/4)(1/5)(1/3) = 1/60 > 0$  lo que implica que  $f_{14} > 0$ , de manera similar se puede mostrar que  $f_{ij} > 0$  ó  $f_{ij} = 0$  para cada  $i, j \in T$ . La siguiente matriz muestra esto:

$$\begin{bmatrix} + & 0 & 0 & 0 & 0 & 0 \\ + & + & + & + & + & + \\ + & + & + & + & + & + \\ 0 & 0 & 0 & + & + & + \\ 0 & 0 & 0 & + & + & + \\ 0 & 0 & 0 & + & + & + \end{bmatrix}.$$

El signo  $+$  indica que un estado  $i$  se comunica con un estado  $j$ , es decir que  $f_{ij} > 0$ , y cero en caso contrario; es decir, en donde  $f_{ij} = 0$ . De esta última matriz se observa que el 0 es un estado absorbente y por tanto recurrente. Si se divide la matriz  $P$  como  $P_1$  y  $P_2$ , en donde  $P_1$  tiene a los estados  $\{1, 2\}$  y el resto,  $\{3, 4, 5\}$ , están en  $P_2$ . Los estados 1 y 2 se comunican con el estado 0, pero ninguno de los dos es alcanzado a partir del 0, por lo que de la Proposición 1.10 se tiene que estos estados son transitorios. En cambio, en  $P_2$  los tres estados son comunicantes, por lo que en  $P_2$  se tiene un espacio de estados finitos e irreducible. Entonces, por Corolario 1.13, se tiene que 3, 4 y 5 son estados recurrentes. Así, los estados transitorios son  $\{1, 2\}$  y los recurrentes  $\{0, 3, 4, 5\}$ .

En resumen, se puede comentar que un conjunto de estados recurrentes se puede ver como la unión finita o numerable de conjuntos disjuntos en donde todos los estados son comunicantes.

Con anterioridad se mencionó que si un estado  $i$  es recurrente significa que con probabilidad uno la cadena regresa al estado  $i$ . Una característica importante a estudiar en este caso es el tiempo que tardará en regresar, lo que se conoce como tiempo medio de retorno.

**Definición 1.10** Sea  $i$  un estado recurrente de una cadena de Markov. El **tiempo medio de retorno** se define como,

$$m_i = E[R_i | X_0 = i] = \sum_{n=1}^{\infty} n f_{ii}^n. \quad (1.17)$$

Es importante mencionar que el tiempo medio de retorno a un estado puede ser infinito, aún cuando este sea recurrente. Por lo que a partir de esto, los estados recurrentes son clasificados como se muestra a continuación.

**Definición 1.11** A un estado recurrente  $i$  se le llama:

- (i) **Recurrente nulo** si  $m_i = \infty$ .
- (ii) **Recurrente positivo** si  $m_i < \infty$ .

**Definición 1.12** Una cadena de Markov  $\{X_n\}_{n \in \mathbb{N}}$  es llamada:

- (i) **Cadena Recurrente Nula** si cada uno de los estados son recurrentes nulos.
- (ii) **Cadena Recurrente Positiva** si cada estado es recurrente positivo.

Hasta aquí se ha hablado de estados recurrentes y transitorios, del número de visitas esperado, del tiempo que tarda en regresar la cadena a un estado, entre otros. Ahora, se verá en cuantos pasos la cadena regresa a un estado dado que ya estuvo en él.

### 1.3.2 Periodicidad

Otra de las características importantes para el estudio de las cadenas de Markov es la periodicidad de los estados.

**Definición 1.13** Se define el **período**  $d(i)$  de un estado  $i$  como el máximo común divisor de todos los enteros  $n \geq 1$  tal que  $P_{ii}^n > 0$ , y se denota como

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

Si  $d(i) > 1$  se dice que  $i$  es un estado periódico con período  $d(i)$ . Mientras que si  $d(i) = 1$ , se dice que  $i$  es **Aperiódico**.

**Ejemplo 1.2** Considérese la caminata aleatoria del Ejemplo 1.2.2 con espacio de estados finito y matriz de transición

$$P = \begin{bmatrix} r & p & 0 & \cdots & 0 \\ q & r & p & \cdots & 0 \\ 0 & q & r & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & r \end{bmatrix}.$$

Se observa que si  $r = P_{ii} > 0$ , se tiene que estando en cualquier estado  $i$ , existe probabilidad positiva de permanecer en el mismo estado. De la matriz  $P$ , se obtiene que  $P_{11}^2 = \sum_{i \in T} P_{1i}P_{i1} > 0$ . En general  $P_{11}^n = \sum_{i \in T} P_{1i}P_{i1}^{n-1} > 0$ . De manera similar se tiene para cada  $i \in T$ . Por lo tanto  $P_{ii}^n > 0$  para toda  $n \geq 1$ , entonces cada estado tendrá período uno, es decir  $d(i) = 1$  con  $i \in T$ . Así, la cadena es Aperiódica.

Ahora, en el caso que  $r = P_{ii} = 0 \forall i \in T$ , la matriz de transición  $P$  tomará la siguiente forma

$$\begin{bmatrix} 0 & p & 0 & \cdots & 0 \\ q & 0 & p & \cdots & 0 \\ 0 & q & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Nótese que  $P_{ii}^2 = \sum_{j \in T} P_{ij}P_{ji} > 0$ , para cada  $i \in T$ . En general,  $P_{ii}^n > 0$  con  $n = 2k$ ,  $k \in \mathbb{N}$  lo que indica que saliendo de cualquier estado  $i$  sólo se puede regresar a él, en un número par de pasos, por lo que tomando en cuenta la definición de período, se tiene que  $d(i) = 2$  para cada  $i \in T$ .

**Proposición 1.14** Sea  $\{X_n\}_{n \in \mathbb{N}}$  una cadena de Markov con espacio de estados  $T$  y matriz de transición  $P$ . Si  $i, j \in T$  son comunicantes ( $i \leftrightarrow j$ ) entonces  $d(i) = d(j)$ .

**Demostración.** Si  $i \leftrightarrow j$ , existen  $n_1$  y  $n_2$  enteros positivos tales que

$$P_{ij}^{n_1} > 0 \quad y \quad P_{ji}^{n_2} > 0.$$

Entonces

$$P_{ii}^{n_1+n_2} \geq P_{ij}^{n_1} P_{ji}^{n_2} > 0.$$

Como  $d(i) = m.c.d\{n \geq 1 : P_{ii}^n > 0\}$ , se tiene que  $d(i)$  es un divisor de  $n_1 + n_2$ .

Si  $P_{jj}^n > 0$ , entonces

$$P_{ii}^{n_1+n+n_2} \geq P_{ij}^{n_1} P_{jj}^n P_{ji}^{n_2} > 0.$$

Así,  $d(i)$  es divisor de  $n_1 + n + n_2$ , y como  $d(i)$  también es un divisor de  $n_1 + n_2$  entonces  $d(i)$  es divisor de  $n$ . Por lo tanto  $d(i)$  es divisor de todos los números en el conjunto  $\{n \geq 1 : P_{jj}^n > 0\}$ . Pero como  $d(j)$  es el *m.c.d* del mismo conjunto, entonces se concluye que  $d(i) \leq d(j)$ . De manera análoga se cumple que

$$d(j) \leq d(i),$$

por tanto

$$d(i) = d(j).$$

■

Entonces el período es una propiedad de clase, es decir, en una cadena irreducible todos los estados tienen el mismo período.

## 1.4 Distribución límite de una cadena de Markov

La parte central de esta sección es el estudio del comportamiento límite de cadenas de Markov. Estos es, para cada par de estados  $i, j$  se desea encontrar condiciones para la existencia de límites de la forma

$$\lim_{n \rightarrow \infty} P_{ij}^n = \pi(j) \quad (1.18)$$

y

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n P_{ij}^k = \pi(j). \quad (1.19)$$

donde  $\pi$  es una distribución.

El límite (1.18) se cumple bajo condiciones muy restrictivas, en cambio, el límite (1.19) siempre existe como se verá en esta sección.

Para el estudio de estos límites se introducirán los siguientes conceptos.

**Definición 1.14** Sea  $\{X_n\}_{n \in \mathbb{N}}$  una cadena de Markov con espacio de estados  $T$  y probabilidades de transición  $P_{ij}$ ,  $i, j \in T$ .

(i) Una distribución de probabilidad  $\pi(i)$  es una **distribución estacionaria** o **distribución invariante** para la cadena de Markov si  $\pi P = \pi$ , es decir

$$\sum_i \pi(i) P_{ij} = \pi(j), \quad j \in T. \quad (1.20)$$

(ii) Una distribución de probabilidad  $\pi$  es una **distribución límite** si

$$\lim_{n \rightarrow \infty} P_{ij}^n = \pi(j), \quad \text{para todo } i, j \in T. \quad (1.21)$$

**Proposición 1.15** Si  $\pi$  es una distribución estacionaria entonces se cumple que  $\sum_i \pi(i) P_{ij}^{(n)} = \pi(j)$ ,  $j \in T$ .

**Demostración.** Se hará por inducción matemática. Sea  $\pi$  una distribución estacionaria. Entonces,

$$\begin{aligned} \sum_i \pi(i) P_{ij}^{(2)} &= \sum_i \pi(i) \sum_k P_{ik} P_{kj} \\ &= \sum_k \left( \sum_i \pi(i) P_{ik} \right) P_{kj} \\ &= \sum_k \pi(k) P_{kj} = \pi(j), \quad j \in T. \end{aligned}$$

Se supone para  $n$ ,

$$\sum_i \pi(i) P_{ij}^{(n)} = \pi(j).$$

y se demuestra para  $n + 1$ ,

$$\begin{aligned} \sum_i \pi(i) P_{ij}^{(n+1)} &= \sum_i \pi(i) \sum_k P_{ik}^{(n)} P_{kj} \\ &= \sum_k \left( \sum_i \pi(i) P_{ik}^{(n)} \right) P_{kj} \\ &= \sum_k \pi(k) P_{kj} \\ &= \pi(j), \quad j \in T. \end{aligned}$$

■

El nombre de distribución invariante se debe a que si la distribución inicial  $\pi_0$  de la cadena es invariante entonces la distribución de  $X_n$  es  $\pi_0$  para todo  $n = 0, 1, \dots$ , es decir  $\pi_0$  – *invariante* implica  $\Pr\{X_n = j\} = \pi_0(j)$  con  $j \in T$ , para toda  $n \geq 0$ .

**Proposición 1.16** Si la distribución inicial  $\pi_0$  es estacionaria entonces

$$\Pr\{X_n = j\} = \pi_0(j), \quad j \in T.$$

**Demostración.** Sea  $\pi_0$  una distribución inicial y estacionaria. Entonces,

$$\begin{aligned} \Pr\{X_n = j\} &= \sum_i \Pr\{X_0 = i, X_n = j\} \\ &= \sum_i \Pr\{X_0 = i\} \cdot \Pr\{X_n = j \mid X_0 = i\} \\ &= \sum_i \pi_0(i) P_{ij}^{(n)} = \pi_0(j). \end{aligned}$$

La última igualdad se sigue de la proposición anterior.

■



**Ejemplo 1.3** *Considérese una cadena de Markov con espacio de estados  $T = \{0, 1, 2\}$  y probabilidades de transición dadas en la siguiente matriz*

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}.$$

*Para encontrar la distribución estacionaria, se resuelve el sistema de ecuaciones dado en (1.20):*

$$\begin{aligned} \frac{\pi(0)}{2} + \frac{\pi(1)}{2} + \frac{\pi(2)}{4} &= \pi(0) \\ \frac{\pi(0)}{4} + \frac{\pi(2)}{4} &= \pi(1) \\ \frac{\pi(0)}{4} + \frac{\pi(1)}{2} + \frac{\pi(2)}{2} &= \pi(2). \end{aligned}$$

*Puesto que  $\pi$  es una distribución de probabilidad se cumple que*

$$\pi(0) + \pi(1) + \pi(2) = 1.$$

*Por lo tanto*

$$\pi = (0.4, 0.2, 0.4).$$

*Ahora se verá si la distribución estacionaria de esta cadena es una distribución límite. Nótese que*

$$\begin{aligned} P^2 &= \begin{bmatrix} 0.4375 & 0.1875 & 0.375 \\ 0.375 & 0.25 & 0.375 \\ 0.375 & 0.1875 & 0.4375 \end{bmatrix}, \\ P^3 &= \begin{bmatrix} 0.40625 & 0.20313 & 0.39063 \\ 0.40625 & 0.1875 & 0.40625 \\ 0.39063 & 0.20313 & 0.40625 \end{bmatrix} \\ P^6 &= \begin{bmatrix} 0.40015 & 0.19996 & 0.39991 \\ 0.3999 & 0.20020 & 0.3999 \\ 0.39991 & 0.19996 & 0.40015 \end{bmatrix} \\ P^{12} &= \begin{bmatrix} 0.40001 & 0.20001 & 0.40001 \\ 0.4 & 0.20001 & 0.4 \\ 0.40001 & 0.20001 & 0.40001 \end{bmatrix} \end{aligned}$$

*Así, computacionalmente se observa que  $\pi = (0.4, 0.2, 0.4)$  es también la distribución límite.*

**Ejemplo 1.4** *Considérese la cadena de Nacimiento y Muerte con espacio de estados  $T = \{0, 1, 2\}$  y matriz de transición igual a*

$$P = \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix}.$$

Para encontrar la distribución estacionaria se procede de manera similar al ejemplo anterior. El sistema de ecuaciones que resulta es el siguiente

$$\begin{aligned}\frac{\pi(1)}{2} &= \pi(0) \\ \pi(0) + \pi(2) &= \pi(1) \\ \frac{\pi(1)}{2} &= \pi(2) \\ \pi(0) + \pi(1) + \pi(2) &= 1.\end{aligned}$$

Resolviendo este sistema se tiene que la distribución estacionaria es

$$\pi = \left( \frac{1}{4}, \frac{1}{2}, \frac{1}{4} \right).$$

Ahora se verá si  $\pi$  es una distribución límite. De la matriz de transición  $P$  se puede obtener

$$\begin{aligned}P^2 &= \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}, \\ P^3 &= \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix} \\ P^4 &= \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}\end{aligned}$$

Así, se observa que las matrices de transición para potencias pares  $n = 2k$ , son de la forma

$$P^{2k} = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$$

y para potencias impares  $n = 2k + 1$ , se tiene que

$$P^{2k+1} = \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix}.$$

Por lo tanto  $\lim_{n \rightarrow \infty} P^n$  no existe.

A continuación se introduce la notación para el número total de visitas y el número esperado de vistas de la cadena a un estado al tiempo  $n$ .

Sea  $\{X_n\}_{n \in \mathbb{N}}$  una cadena de Markov, para cada  $n = 1, 2, \dots$ . Sea  $N_n(j)$  el número total de visitas de la cadena al estado  $j$  hasta el tiempo  $n$ ,

$$N_n(j) = \sum_{m=1}^n \mathbb{I}\{X_m = j\}.$$

Nótese que

$$\lim_{n \rightarrow \infty} N_n(j) = N(j),$$

con  $N(j)$  como en la Definición 1.8 (i).

Sea  $G_{ij}^n$  el número esperado de visitas al estado  $j$  hasta el tiempo  $n$ , dado que  $X_0 = i$ ,

$$G_{ij}^n = E\{N_n(j) | X_0 = i\} = \sum_{m=1}^n P_{ij}^m.$$

**Proposición 1.17** *Sea  $j$  un estado transitorio. Entonces*

$$\lim_{n \rightarrow \infty} \frac{N_n(j)}{n} = 0 \text{ con probabilidad 1}$$

y

$$\lim_{n \rightarrow \infty} \frac{G_{ij}^n}{n} = 0, \quad i \in T.$$

Observe que  $N_n(j)/n$  es la proporción de veces que la cadena está en el estado  $j$  en las primeras  $n$  transiciones, y  $G_{ij}^n/n = \sum_{k=0}^n P_{ij}^k/n$  es el valor esperado de esta proporción cuando la cadena inicia en el estado  $i$ .

**Demostración.** Sea  $j$  un estado transitorio, entonces de la Proposición 1.9a (i) se tiene que

$$\lim_{n \rightarrow \infty} N_n(j) = N(j) < \infty \text{ con probabilidad 1.}$$

Así,

$$\lim_{n \rightarrow \infty} \frac{N_n(j)}{n} = 0 \text{ con probabilidad 1.}$$

De la Proposición 1.9a (ii),  $G_{ij} < \infty$  para todo  $i \in T$ , cuando  $j$  es transitorio. Entonces

$$\lim_{n \rightarrow \infty} G_{ij}^n = G_{ij} = \sum_{n=1}^{\infty} P_{ii}^n < \infty, \quad i \in T.$$

Por lo tanto,

$$\lim_{n \rightarrow \infty} \frac{G_{ij}^n}{n} = 0, \quad i \in T.$$

■

**Proposición 1.18** *Sea  $j$  un estado recurrente. Entonces,*

$$\lim_{n \rightarrow \infty} \frac{N_n(j)}{n} = \frac{\mathbb{I}_{\{R_j < \infty\}}}{m_j} \text{ con probabilidad 1} \quad (1.22)$$

y

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n P_{ij}^k = \lim_{n \rightarrow \infty} \frac{G_{ij}^n}{n} = \frac{f_{ij}}{m_j}, \quad i \in T. \quad (1.23)$$

Si  $m_j = \infty$  el lado derecho de las igualdades es igual a cero.

**Demostración.** Considérese una cadena de Markov que inicia en un estado recurrente  $j$ . Con probabilidad uno la cadena regresa a  $j$  una infinidad de veces. Esta propiedad nos garantiza que la variable aleatoria  $R_j^k$ , que representa el tiempo de la  $k$ -ésima visita al estado  $j$ , es finita con probabilidad uno. Es decir,

$$\Pr\{R_j^k < \infty\} = 1.$$

Ahora se definen las variables aleatorias

$$W_j^1 = R_j^1 \text{ y } W_j^k = R_j^k - R_j^{k-1}, \text{ para } k \geq 2.$$

Notése que  $W_j^k$  representa el tiempo de espera entre la  $(k-1)$  y  $k$ -ésima visita al estado  $j$ .

Entonces se cumple

$$R_j^k = W_j^1 + \dots + W_j^k.$$

Las variables aleatorias  $W_j^1, W_j^2, \dots$  son independientes e idénticamente distribuidas (véase en Hoel 1972, pág.59) y por tanto tienen media común

$$E\{W_j^1 | X_n = j\} = E\{R_j^1 | X_n = j\} = m_j.$$

Luego, por la Ley de los Grandes Números se tiene que

$$\lim_{n \rightarrow \infty} \frac{W_j^1 + \dots + W_j^n}{n} = m_j \text{ con probabilidad 1,}$$

esto es

$$\lim_{n \rightarrow \infty} \frac{R_j^n}{n} = m_j \text{ con probabilidad 1.} \quad (1.24)$$

Sea  $k = N_n(j)$ , el número de visitas al estado  $j$  en las primeras  $n$  transiciones. Entonces la  $k$ -ésima visita a  $j$  ocurre antes del tiempo  $n$  y la  $(k+1)$ -ésima visita a  $j$  ocurre después del tiempo  $n$ ; esto es,

$$R_j^{N_n(j)} \leq n < R_j^{N_n(j)+1},$$

por tanto

$$\frac{R_j^{N_n(j)}}{N_n(j)} \leq \frac{n}{N_n(j)} < \frac{R_j^{N_n(j)+1}}{N_n(j)}. \quad (1.25)$$

Puesto que  $j$  es un estado recurrente se cumple que

$$\lim_{n \rightarrow \infty} N_n(j) = \infty \text{ con probabilidad 1.}$$

Luego de (1.24) y de la desigualdad dada en (1.25) se obtiene que

$$\lim_{n \rightarrow \infty} \frac{n}{N_n(j)} = m_j \text{ con probabilidad 1.}$$

Equivalentemente

$$\lim_{n \rightarrow \infty} \frac{N_n(j)}{n} = \frac{1}{m_j} \text{ con probabilidad 1.}$$

Cuando  $X_0$  tiene distribución arbitraria, la cadena puede no alcanzar al estado  $j$ , pero si lo alcanza, el argumento anterior también es válido, y por lo tanto

$$\frac{N_n(j)}{n} \xrightarrow{n \rightarrow \infty} \frac{\mathbb{I}_{\{R_j < \infty\}}}{m_j}, \text{ con probabilidad 1,} \quad (1.26)$$

y con esto queda demostrado (1.22).

Por otra parte como  $0 \leq N_n(j) \leq n$ , se tiene que

$$0 \leq \frac{N_n(j)}{n} \leq 1. \quad (1.27)$$

El Teorema de Convergencia Dominada, permite concluir en (1.26) y (1.27) que

$$\begin{aligned} \lim_{n \rightarrow \infty} E \left[ \frac{N_n(j)}{n} \middle| X_0 = i \right] &= E \left[ \frac{\mathbb{I}_{\{R_j < \infty\}}}{m_j} \middle| X_0 = i \right] \\ &= \frac{\Pr\{R_j < \infty | X_0 = i\}}{m_j} \\ &= \frac{f_{ij}}{m_j}. \end{aligned}$$

Recuerde que

$$G_{ij}^n = E\{N_n(j) | X_0 = i\} = \sum_{m=1}^n P_{ij}^m$$

representa el número esperado de visitas al estado  $j$  hasta el tiempo  $n$ , dado  $X_0 = i$ . Por tanto

$$\begin{aligned} \lim_{n \rightarrow \infty} G_{ij}^n &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n P_{ij}^k \\ &= \lim_{n \rightarrow \infty} E \left[ \frac{G_{ij}^n}{n} \middle| X_0 = i \right] \\ &= \frac{f_{ij}}{m_j}. \end{aligned}$$

Con esto queda demostrado (1.23). ■

**Corolario 1.19** *De la proposición anterior puede observar que cuando un estado  $j$  es*

(i) *Recurrente Nulo, es decir  $1/m_j = 0$ , se obtiene*

$$\lim_{n \rightarrow \infty} \frac{G_{ij}^n}{n} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n P_{ij}^k = 0, \quad i \in T.$$

(ii) *Recurrente Positivo, es decir  $1/m_j > 0$ , entonces*

$$\lim_{n \rightarrow \infty} \frac{G_{jj}^n}{n} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n P_{jj}^k = \frac{1}{m_j} > 0.$$

De la Proposición 1.18 y Corolario 1.19, se sigue que  $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n P_{ij}^k$  siempre existe y satisface, para cada  $i \in T$ :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n P_{ij}^k = \begin{cases} 0, & \text{si } j \text{ transitorio o recurrente nulo,} \\ \frac{1}{m_j}, & \text{si } i = j, j \text{ recurrente positivo,} \\ \frac{f_{ij}}{m_j}, & \text{si } j \text{ es recurrente.} \end{cases}$$

**Proposición 1.20** *Sea  $i$  un estado recurrente positivo y suponga que  $i$  se comunica con  $j$ . Entonces  $j$  es un estado recurrente positivo.*

**Demostración.** Se sigue de la Proposición 1.10 que  $j$  se comunica con  $i$ , con  $j \in T$ . Entonces existen enteros positivos  $n_1$  y  $n_2$  tal que

$$P_{ij}^{n_1} > 0 \quad \text{y} \quad P_{ji}^{n_2} > 0.$$

Del Teorema de *Chapman – Kolmogorov* se obtiene que

$$P_{jj}^{n_1+m+n_2} \geq P_{ji}^{n_2} P_{ii}^m P_{ij}^{n_1}.$$

Sumando sobre  $m = 1, \dots, n$  y dividiendo por  $n$ , se tiene que

$$\begin{aligned} \frac{1}{n} \sum_{m=1}^n P_{jj}^{n_1+m+n_2} &\geq \frac{1}{n} \sum_{m=1}^n P_{ji}^{n_2} P_{ii}^m P_{ij}^{n_1} \\ \sum_{m=1}^{n_1+m+n_2} \frac{P_{jj}^m}{n} - \sum_{m=n_1+m+n_2}^{n_1+n_2} \frac{P_{jj}^m}{n} &\geq P_{ji}^{n_2} P_{ij}^{n_1} \sum_{m=1}^n \frac{P_{ii}^m}{n} \\ \frac{G_{jj}^{n_1+n+n_2}}{n} - \frac{G_{jj}^{n_1+n_2}}{n} &\geq P_{ji}^{n_2} P_{ij}^{n_1} \cdot \frac{1}{n} \sum_{m=1}^n P_{ii}^m \\ &= P_{ji}^{n_2} P_{ij}^{n_1} \frac{G_{ii}^n}{n}. \end{aligned}$$

Cuando  $n \rightarrow \infty$ , del Corolario 1.19 (ii) tenemos que la desigualdad anterior converge a

$$\frac{1}{m_j} \geq \frac{P_{ji}^{n_2} P_{ij}^{n_1}}{m_i} > 0,$$

y consecuentemente  $m_j < \infty$ . Esto demuestra que  $j$  es recurrente positivo. ■

De la Proposición 1.10 y del Corolario 1.12 se observa que si la cadena es irreducible, entonces cada estado se comunica, es decir, cada estado es recurrente nulo o recurrente positivo.

**Proposición 1.21** *Sea  $\{X_n\}_{n \in \mathbb{N}}$  una cadena de Markov con espacio de estados  $T$  finito, entonces se tiene al menos un estado recurrente positivo.*

**Demostración.** Notése que para cada  $m = 1, 2, \dots$ , se cumple que

$$\sum_{j \in T} P_{ij}^m = 1, i \in T.$$

Sumando sobre  $m = 1, \dots, n$  y dividiendo por  $n$  se tiene que

$$\sum_{m=1}^n \sum_{j \in T} \frac{1}{n} P_{ij}^m = \sum_{j \in C} \frac{G_{ij}^n}{n} = 1, i \in T.$$

Así,

$$\lim_{n \rightarrow \infty} \sum_{j \in T} \frac{G_{ij}^n}{n} = 1, i \in T.$$

Por otra parte si cada estado es transitorio o recurrente nulo entonces por el Corolario 1.19 (i) se tiene que

$$1 = \lim_{n \rightarrow \infty} \sum_{j \in T} \frac{G_{ij}^n}{n} = \sum_{j \in T} \lim_{n \rightarrow \infty} \frac{G_{ij}^n}{n} = 0, i \in T.$$

Esto es una contradicción, por lo tanto se tiene al menos un estado recurrente positivo. ■

**Corolario 1.22** *Sea  $\{X_n\}_{n \in \mathbb{N}}$  una cadena de Markov finita e irreducible. Entonces cada estado es recurrente positivo.*

**Demostración.** Puesto que la cadena es finita se tiene al menos un estado recurrente positivo, entonces existe al menos un estado recurrente positivo. Como  $\{X_n\}_{n \in \mathbb{N}}$  es irreducible, de la Proposición 1.20 tenemos que cada estado es recurrente positivo. ■

**Corolario 1.23** *Una cadena de Markov irreducible con un número finito de estados es recurrente positivo.*

**Corolario 1.24** *Una cadena de Markov con un número finito de estados no tiene estados recurrentes nulos.*

**Proposición 1.25** *Sea  $\pi$  una distribución estacionaria. Si  $j$  es un estado transitorio o recurrente nulo entonces  $\pi(j) = 0$ .*

**Demostración.** Si  $j$  es un estado transitorio o recurrente nulo. Entonces

$$\lim_{n \rightarrow \infty} \frac{G_{ij}^n}{n} = 0, \quad j \in T.$$

Sea  $\pi$  una distribución estacionaria. Por la Proposición 1.15 se tiene que

$$\sum_i \pi(i) P_{ij}^k = \pi(j), \quad k \geq 1. \quad (1.28)$$

Sumando  $k = 1, \dots, n$  y dividiendo por  $n$ , se obtiene que

$$\sum_i \pi(i) \frac{G_{ij}^n}{n} = \pi(j), \quad j \in T. \quad (1.29)$$

Se sigue de (1.28), (1.29) y del Teorema de Convergencia Acotada que

$$\pi(j) = \lim_{n \rightarrow \infty} \sum_i \pi(i) \frac{G_{ij}^n}{n} = 0.$$

Por lo tanto,  $\pi(j) = 0$ . ■

En consecuencia, una cadena de Markov que no tenga estado recurrentes positivos no tendrá distribución estacionaria.

**Proposición 1.26** *Una cadena de Markov  $X_n, n \geq 0$  irreducible y recurrente positiva tiene una única distribución estacionaria  $\pi$ , dada por*

$$\pi(j) = \frac{1}{m_j}, \quad j \in T.$$

**Demostración.** Se sigue de la Proposición 1.18 y de las hipótesis de esta proposición que

$$\lim_{n \rightarrow \infty} \frac{G_{ij}^n}{n} = \frac{1}{m_j}, \quad i, j \in T. \quad (1.30)$$

Suponga que  $\pi$  es una distribución estacionaria. De (1.29), (1.30) y del Teorema de Convergencia Acotada se tiene que

$$\begin{aligned} \pi(j) &= \lim_{n \rightarrow \infty} \sum_i \pi(i) \frac{G_{ij}^n}{n} \\ &= \frac{1}{m_j} \sum_i \pi(i) = \frac{1}{m_j}. \end{aligned}$$

Entonces si existe una distribución estacionaria debe estar dada por  $\pi(j) = 1/m_j, j \in T$ .

Para completar la demostración falta ver que  $\pi(j) = 1/m_j, j \in T$  es una distribución estacionaria. Puesto que  $j$  es recurrente positivo, se tiene que  $m_j < \infty$ , entonces  $\pi(j) > 0$ . Falta verificar que

$$\sum_j \frac{1}{m_j} = 1 \quad (1.31)$$

y que

$$\sum_j \frac{1}{m_j} P_{jk} = \frac{1}{m_k}, \quad k \in T. \quad (1.32)$$

Obsérvese que

$$\sum_j P_{ij}^m = 1.$$

Sumando  $m = 1, \dots, n$  y dividiendo entre  $n$  la expresión anterior, se puede concluir que

$$\sum_j \frac{G_{ij}^n}{n} = 1, \quad k \in T. \quad (1.33)$$



Por el Teorema de *Chapman – Kolmogorov* se obtiene que

$$\sum_j P_{ij}^m P_{jk} = P_{ik}^{m+1}.$$

Nuevamente, sumando  $m = 1, \dots, n$  y dividiendo por  $n$  la última expresión, se tiene que

$$\sum_j \frac{G_{ij}^n}{n} P_{jk} = \frac{G_{ik}^{n+1}}{n} - \frac{P_{ik}}{n}. \quad (1.34)$$

Si  $T$  es finito, se tiene de (1.30) y de (1.33) que

$$1 = \lim_{n \rightarrow \infty} \sum_j \frac{G_{ij}^n}{n} = \sum_j \frac{1}{m_j},$$

entonces se cumple (1.31). Ahora al tomar límite en (1.34) cuando  $n \rightarrow \infty$  se sigue que

$$\begin{aligned} \sum_{i \in T} \frac{1}{m_i} P_{ij} &= \lim_{n \rightarrow \infty} \sum_{i \in T} \frac{G_{ki}^n}{n} P_{ij} = \lim_{n \rightarrow \infty} \frac{G_{kj}^{n+1}}{n} \\ &= \lim_{n \rightarrow \infty} \frac{P_{kj}}{n} = \frac{1}{m_j}. \end{aligned}$$

Por lo tanto (1.32) se cumple. Entonces  $\pi$  es una distribución estacionaria cuando el espacio de estados es finito.

Ahora se verá el caso donde  $T$  es infinito ya que la argumentación anterior no es adecuada pues en general, el Teorema de la Convergencia Dominada no es aplicable. Sea  $T_1$  un subconjunto finito de  $T$ . De (1.33) se tiene que

$$\sum_{j \in T_1} \frac{G_{ij}^n}{n} \leq 1, \quad i \in T.$$

Puesto que  $T_1$  es finito, cuando  $n \rightarrow \infty$  en la última desigualdad y por (1.30), se tiene que

$$\sum_{j \in T_1} \frac{1}{m_j} \leq 1,$$

la cual se cumple para cualquier subconjunto finito  $T_1$  de  $T$ . Por tanto

$$\sum_{j \in T} \frac{1}{m_j} \leq 1, \quad (1.35)$$

ya que de lo contrario la suma sobre algún conjunto finito  $T_1$  de  $T$  también excederá a uno.

Similarmente se concluye de (1.35) que si  $T_1$  es un subconjunto finito de  $T$ , entonces

$$\sum_{j \in T_1} \frac{G_{ij}^n}{n} P_{jk} = \frac{G_{ik}^{n+1}}{n} - \frac{P_{ik}}{n}.$$

Al tomar límite cuando  $n \rightarrow \infty$  en esta desigualdad y usando (1.30), se obtiene que

$$\sum_{j \in T_1} \frac{1}{m_j} P_{jk} \leq \frac{1}{m_k}.$$

Al igual que en la demostración de (1.35), se concluye que

$$\sum_j \frac{1}{m_j} P_{jk} \leq \frac{1}{m_k}, \quad k \in T. \quad (1.36)$$

Observe que se cumple la igualdad en (1.36). Si no es así, entonces de (1.31) se cumple la suma sobre  $k$ ,

$$\begin{aligned} \sum_k \frac{1}{m_k} &> \sum_k \left( \sum_j \frac{1}{m_j} P_{jk} \right) \\ &= \sum_j \frac{1}{m_j} \left( \sum_k P_{jk} \right) \\ &= \sum_j \frac{1}{m_j}, \end{aligned}$$

lo cual es una contradicción. Por lo tanto la igualdad se cumple en (1.36), esto es

$$\sum_j \frac{1}{m_j} P_{jk} = \frac{1}{m_k}, \quad k \in T. \quad (1.37)$$

Sólo falta demostrar que

$$\sum_j \frac{1}{m_j} = 1.$$

Sea

$$c = \frac{1}{\sum_j \frac{1}{m_j}}.$$

Entonces por (1.37)

$$\pi(j) = \frac{c}{m_j}, \quad j \in T,$$

define una distribución estacionaria. Entonces por la primera parte de la demostración

$$\frac{c}{m_j} = \frac{1}{m_j},$$

y por tanto  $c = 1$ . Esto prueba que se cumple

$$\sum_j \frac{1}{m_j} = 1.$$

■

**Corolario 1.27** *Una cadena de Markov  $\{X_n\}_{n \in \mathbb{N}}$  irreducible es recurrente positivo si y sólo si tiene una distribución estacionaria.*

**Demostración.** Sea  $\{X_n\}_{n \in \mathbb{N}}$  una cadena de Markov irreducible y recurrente positivo. Entonces de la Proposición 1.26 se tiene una única distribución estacionaria dada por

$$\pi(j) = \frac{1}{m_j}, \quad j \in T.$$

Ahora suponga que  $\pi$  es una distribución estacionaria. Entonces si  $j$  es un estado transitorio o recurrente nulo, de la Proposición 1.25, tenemos que  $\pi(j) = 0$ .

Por lo tanto cada estado  $i \in T$  debe ser recurrente positivo y así la cadena será recurrente positivo. ■

**Proposición 1.28** *Si  $\{X_n\}_{n \in \mathbb{N}}$  es una cadena de Markov irreducible, recurrente positiva y con distribución estacionaria  $\pi$ . Entonces,*

$$\lim_{n \rightarrow \infty} \frac{N_n(j)}{n} = \pi(j), \quad j \in T.$$

Este resultado se conoce como la Ley Fuerte de los Grandes Números para cadenas de Markov.

**Demostración.** De la Proposición 1.18 se tiene que

$$\lim_{n \rightarrow \infty} \frac{N_n(j)}{n} = \frac{1}{m_j}, \quad j \in T.$$

De la Proposición 1.26, se tiene que una cadena de Markov irreducible y recurrente positiva tiene una única distribución dada por  $\pi(j) = 1/m_j$ .

Por lo tanto,

$$\lim_{n \rightarrow \infty} \frac{N_n(j)}{n} = \pi(j), \quad j \in T.$$

■

**Corolario 1.29** *Sea  $\{X_n\}_{n \in \mathbb{N}}$  una cadena de Markov irreducible. Si el espacio de estados  $T$  es finito entonces tiene una única distribución estacionaria  $\pi$ .*

**Demostración.** Del Corolario 1.23, se tiene que si una cadena de Markov es irreducible con espacio de estados finito entonces la cadena es recurrente positiva. Por lo tanto, de la Proposición 1.26 se tiene una única distribución estacionaria. ■

Si la cadena no tiene estados recurrentes positivos entonces no tiene distribución estacionaria, si tiene estados recurrentes positivos y son comunicantes entonces tiene una única distribución estacionaria, y si no son comunicantes entonces tiene infinidad de distribuciones estacionarias, Hoel (1972, pág 68).

**Teorema 1.30** (*básico de convergencia*) Sea  $\{X_n\}_{n \in \mathbb{N}}$  una cadena de Markov irreducible, recurrente positiva y con distribución estacionaria  $\pi$ . Si la cadena es aperiódica entonces

$$\lim_{n \rightarrow \infty} P_{ij}^n = \frac{1}{m_j} = \pi(j), \quad i, j \in T.$$

**Demostración.** Se considerará una cadena de Markov aperiódica, irreducible y recurrente positiva con función de transición  $P$ , espacio de estados  $T$  y distribución estacionaria  $\pi$ .

Sea  $a \in T$ . Se define el conjunto  $I$  por:

$$I = \{n > 0 : P_{aa}^n > 0\}.$$

Entonces

(i) *m.c.d.*  $I = 1$

(ii) Si  $m \in I$  y  $n \in I$  entonces  $m + n \in I$ .

La propiedad (ii) se sigue de la desigualdad

$$P_{aa}^{m+n} \geq P_{aa}^m P_{aa}^n.$$

Las propiedades (i) y (ii) implican que existe un entero positivo  $n_1$  tal que  $n \in I$  para toda  $n \geq n_1$ . Para detalles de la demostración véase en Hoel (1972, págs. 79-80). Por lo que se cumple que

$$P_{aa}^n > 0 \text{ para } n \geq n_1.$$

Sean  $i, j \in T$ . Como la cadena es irreducible, existen enteros positivos  $k, l$  tal que

$$P_{ia}^k > 0 \quad \text{y} \quad P_{aj}^l > 0.$$

Entonces para  $n \geq n_1$  se tiene que

$$P_{ij}^{k+n+l} \geq P_{ia}^k P_{aa}^n P_{aj}^l > 0.$$

Así, para cada  $i, j \in T$ , existe un entero positivo  $n_0 = k + l + n_1$  tal que

$$P_{ij}^n > 0 \text{ para } n \geq n_0. \tag{1.38}$$

Ahora, sea

$$T^2 = \{(i, j) : i, j \in T\}.$$

Entonces  $T^2$  es el conjunto de las parejas ordenadas de elementos en  $T$ . Se considerará una cadena de Markov  $(X_n, Y_n)$  con espacio de estados  $T^2$  y función de transición  $P_2$  definida por

$$P_2((i_0, j_0), (i, j)) = P(i_0, i)P(j_0, j).$$

Se sigue que  $\{X_n, n \geq 0\}$  y  $\{Y_n, n \geq 0\}$  son cadenas de Markov con función de transición  $P$  y las transiciones de  $X_n$  y  $Y_n$  son independientemente una de otra.

Ahora se demostrará que la cadena de Markov  $(X_n, Y_n)$  es aperiódica, irreducible y recurrente positiva.

Sean  $(i_0, j_0)$  y  $(i, j) \in T^2$ . Por (1.38) existe  $n_0$  tal que

$$P_{i_0 i}^n > 0 \quad \text{y} \quad P_{j_0 j}^n > 0, \quad n \geq n_0.$$

Entonces

$$P_2^n((i_0, j_0), (i, j)) = P^n(i_0, i)P^n(j_0, j) > 0, \quad n \geq n_0.$$

Por tanto, la cadena  $(X_n, Y_n)$  es irreducible y aperiódica.

La distribución  $\pi_2$  en  $T^2$  definida por  $\pi_2(i_0, j_0) = \pi(i_0)\pi(j_0)$  es una distribución estacionaria,

$$\begin{aligned} & \sum_{(i_0, j_0) \in T^2} \pi_2(i_0, j_0) P_2((i_0, j_0), (i, j)) \\ &= \sum_{i_0 \in T} \sum_{j_0 \in T} \pi(i_0)\pi(j_0) P_{i_0 i} P_{j_0 j} \\ &= \left( \sum_{i_0 \in T} \pi(i_0) P_{i_0 i} \right) \left( \sum_{j_0 \in T} \pi(j_0) P_{j_0 j} \right) \\ &= \pi(i)\pi(j) = \pi_2(i, j). \end{aligned}$$

Por lo tanto  $\pi_2$  es una distribución estacionaria para la cadena  $(X_n, Y_n)$ . Luego del Corolario 1.27 se sigue que es recurrente positiva. En particular, es recurrente.

Sea

$$L = \min\{n > 0 : X_n = Y_n\}.$$

Sea  $a \in T$ . Puesto que la cadena  $(X_n, Y_n)$  es recurrente,

$$L_{aa} = \min\{n > 0 : (X_n, Y_n) = (a, a)\}$$

es finito con probabilidad uno y además  $L \leq L_{aa}$ . Por lo tanto  $L$  es finito con probabilidad uno.

Se demostrará que para algún  $n \geq 1$ ,

$$\Pr\{X_n = j, L \leq n\} = \Pr\{Y_n = j, L \leq n\}, \quad j \in T. \quad (1.39)$$

Esta fórmula es intuitivamente razonable, ya que las cadenas son indistinguibles para  $n \geq L$ .

Para demostrar (1.39), se toma  $1 \leq m \leq n$ . Entonces para un estado  $z \in T$  se tiene que

$$\begin{aligned} \Pr\{X_n = j | L = m, X_m = Y_m = z\} \\ = \Pr\{Y_n = j | L = m, X_m = Y_m = z\}, \end{aligned} \quad (1.40)$$

ya que ambas probabilidades condicionales son iguales a  $P_{zj}^{n-m}$ . Ahora el evento  $\{L \leq n\}$  es la unión de los eventos disjuntos

$$\{L = m, X_m = Y_m = z\}, 1 \leq m \leq n \text{ y } z \in T.$$

Así se sigue de (1.40) que

$$\Pr\{X_n = j | L \leq n\} = \Pr\{Y_n = j | L \leq n\}, \quad j \in T.$$

Lo anterior implica que

$$\begin{aligned} P\{X_n = j\} &= \Pr\{X_n = j, L \leq n\} + \Pr\{X_n = j, L > n\} \\ &= \Pr\{Y_n = j, L \leq n\} + \Pr\{X_n = j, L > n\} \\ &\leq \Pr\{Y_n = j\} + \Pr\{L > n\}, \end{aligned}$$

y de manera similar

$$P\{Y_n = j\} \leq \Pr\{X_n = j\} + \Pr\{L > n\}.$$

De esta manera, para  $n \geq 1$  se cumple la siguiente desigualdad

$$|\Pr\{X_n = j\} - \Pr\{Y_n = j\}| \leq \Pr\{L > n\}, \quad j \in T. \quad (1.41)$$

Por otra parte, como  $L$  es finito con probabilidad uno, entonces

$$\lim_{n \rightarrow \infty} \Pr\{L > n\} = 0. \quad (1.42)$$

De (1.41) y (1.42) que

$$\lim_{n \rightarrow \infty} (\Pr\{X_n = j\} - \Pr\{Y_n = j\}) = 0, \quad j \in T. \quad (1.43)$$

Sea  $i \in T$  y sea la distribución inicial de  $(X_n, Y_n)$  tal que

$$\Pr\{X_0 = i\} = 1 \text{ y } \Pr\{Y_0 = j_0\} = \pi_0(j_0), \quad j_0 \in T.$$

Puesto que  $X_n$  y  $Y_n$  son cadenas de Markov con función de transición  $P$ , se verá que

$$\Pr\{X_n = j\} = P_{ij}^n, \quad j \in T, \quad (1.44)$$

y

$$\Pr\{Y_n = j\} = \pi(j), \quad j \in T. \quad (1.45)$$

Entonces por (1.43), (1.44) y (1.45) se cumple la siguiente cadena de igualdades,

$$\lim_{n \rightarrow \infty} (P_{ij}^n - \pi(j)) = \lim_{n \rightarrow \infty} (\Pr\{X_n = j\} - \Pr\{Y_n = j\}) = 0.$$

Por tanto

$$\lim_{n \rightarrow \infty} P_{ij}^n = \pi(j).$$

■

En este capítulo se describió detalladamente el estudio de cadenas de Markov desde el punto de vista probabilístico. Se expuso teoría relevante correspondiente a conceptos y propiedades básicas de estos procesos. Además, se analizó el comportamiento asintótico de estas cadenas. Todo esto con el objetivo de mostrar el énfasis de la teoría básica de Probabilidad para cadenas de Markov.

El tratamiento que le da el enfoque estadístico a cadenas de Markov es muy diferente al enfoque probabilista. Bajo el paradigma estadístico un problema relevante y poco estudiado es hacer inferencia sobre las probabilidades de transición con base en una muestra observada de la cadena. En el siguiente capítulo se presenta el método de verosimilitud para Inferencia Estadística, en particular se utilizará esta metodología para hacer inferencia sobre las probabilidades de transición de una cadena de Markov.





## Capítulo 2

### Inferencia basada en Verosimilitud

En este capítulo se darán los conceptos básicos para realizar inferencia estadística a través del enfoque de verosimilitud. Por ejemplo: la función de verosimilitud, el estimador de máxima verosimilitud (*EMV*), la verosimilitud relativa, intervalos de verosimilitud, entre otros. Esta metodología será ejemplificada ampliamente en el siguiente capítulo al abordar el problema de estimación sobre las probabilidades de transición en una cadena de Markov. Cabe señalar que dicho problema es la parte central de esta tesis.

#### 2.1 La función de verosimilitud

Considérese  $X = (X_1, \dots, X_n)$  un vector de variables aleatorias discretas con función de probabilidad  $\Pr\{X = x; \theta\}$  que depende de un número finito de parámetros reales desconocidos  $\theta = (\theta_1, \dots, \theta_k)$  que están dentro de un espacio parametral  $\Theta \subset \mathbb{R}^k$ . El espacio parametral  $\Theta$  es la región de valores posibles que puede tomar el vector de parámetros  $\theta$ .

**Definición 2.1** *La función de verosimilitud de  $\theta$  se define como proporcional a la probabilidad de la muestra observada pero como función del parámetro,*

$$L(\theta; x) \propto \Pr\{X = x; \theta\}, \quad (2.1)$$

donde  $x$  es la muestra observada, Fisher (1921).

El rol de la verosimilitud es inferir sobre los parámetros de la distribución que haya sido elegido para describir algún fenómeno aleatorio de interés a partir de una muestra observada. Nótese que esto es relevante una vez realizado el experimento. También proporciona un orden de preferencia entre los posibles valores del parámetro  $\theta$  basada en la muestra observada  $X = x$ . Si  $\Pr\{X = x; \theta_1\} > \Pr\{X = x; \theta_2\}$  entonces  $\theta = \theta_1$  es más plausible que  $\theta = \theta_2$  dada la muestra observada  $X = x$  ( $\theta_1$  hace más probable a la muestra observada de lo que la hace  $\theta_2$ ). En consecuencia,  $L(\theta_1; x) > L(\theta_2; x)$ .

En el caso de que  $X_i$  sea una variable aleatoria continua, se tiene que  $\Pr\{X_i = x_i; \theta\} = 0$  para todo  $\theta$  y toda  $x_i$ . Sin embargo, observar  $X_i = x_i$  se interpreta como  $x_i - \frac{\epsilon}{2} \leq X_i \leq x_i + \frac{\epsilon}{2}$ , donde  $\epsilon > 0$  es un número fijo que representa la precisión del instrumento de medición. En la realidad, los datos observados siempre son discretos ya que los instrumentos, con los cuales se obtienen estos, tienen precisión

finita pues sólo pueden registrar mediciones con un número finito de decimales. Entonces, sin pérdida de generalidad, para una muestra de variables independientes e idénticamente distribuidas  $X = (X_1, \dots, X_n)$  con función de densidad  $f(x; \theta)$ , la función de verosimilitud de  $\theta$  es proporcional a la probabilidad conjunta de la muestra observada,

$$L(\theta; x) \propto \prod_{i=1}^n \Pr\{x_i - \frac{\epsilon}{2} \leq X_i \leq x_i + \frac{\epsilon}{2}; \theta\} = \prod_{i=1}^n \int_{x_i - \frac{\epsilon}{2}}^{x_i + \frac{\epsilon}{2}} f(x_i; \theta) dx. \quad (2.2)$$

A la función de verosimilitud definida en (2.2) se le llama *verosimilitud exacta* de  $\theta$ , Montoya (2008).

Nótese que la función de verosimilitud exacta puede tener una expresión matemática inconveniente al momento de realizar cálculos. A continuación se muestra una aproximación continua a la función de verosimilitud exacta.

El Teorema de valor medio para integrales definidas, establece que si  $f(\cdot)$  es una función continua en el intervalo  $[a, b]$  entonces existe  $c \in [a, b]$  tal que

$$\int_a^b f(x) dx = (b - a) f(c).$$

Por lo tanto, en el caso donde la función de densidad  $f(x; \theta)$  es continua en el intervalo  $[x_i - \frac{\epsilon}{2}, x_i + \frac{\epsilon}{2}]$ , se tiene que la  $i$ -ésima integral en (2.2) es igual a  $\epsilon f(x'; \theta)$  para algún  $x' \in [x_i - \frac{\epsilon}{2}, x_i + \frac{\epsilon}{2}]$ . Si  $f(x; \theta)$  es aproximadamente constante en este intervalo para todo  $\theta$  se tiene que  $f(x; \theta) \approx f(x'; \theta)$  en ese mismo intervalo. Si esta aproximación es adecuada para algunos o todos los  $i$ , con  $i = 1, \dots, n$ , y si  $\epsilon$  no depende de  $\theta$  entonces las correspondientes probabilidades en (2.2) se pueden reemplazar por sus respectivas densidades evaluadas en los datos observados. Esto se conoce como la aproximación continua a la función de verosimilitud exacta,

$$L(\theta; x) \propto \prod_{i=1}^n \int_{x_i - \frac{\epsilon}{2}}^{x_i + \frac{\epsilon}{2}} f(x_i; \theta) dx \propto \prod_{i=1}^n \epsilon(x_i; \theta). \quad (2.3)$$

Cabe señalar aquí, que en mucha bibliografía estadística definen a la función de verosimilitud como la aproximación continua dada en (2.3), Berger *et al.* (1999), Cox y Hinkley (1974), Casella y Berger (2002). Se debe tener algún cuidado con esto puesto las funciones de densidad puede presentar problemas de singularidad. Montoya *et al* (2009).

Obsérvese que la función de verosimilitud  $L(\theta; x)$  se define como proporcional, y no igual, a la función de probabilidad  $\Pr\{X = x; \theta\}$ . Por esta razón es importante señalar que el cociente de verosimilitudes

$$\frac{L(\theta_1; x)}{L(\theta_2; x)} = \frac{\Pr\{X = x; \theta_1\}}{\Pr\{X = x; \theta_2\}}$$

tiene significado en el sentido que es una medida de la plausibilidad de  $\theta_1$  relativa a  $\theta_2$  basada en la muestra observada  $X = x$ . La razón

$$\frac{L(\theta_1; x)}{L(\theta_2; x)} = k$$

significa que  $\theta_1$  es  $k$  – veces más plausible que  $\theta_2$  en el sentido de que  $\theta_1$  hace a la muestra observada  $k$  – veces más probable que  $\theta_2$ . Por esto, se sugiere utilizar la función de verosimilitud relativa para hacer inferencias sobre  $\theta$ . Esta función se define a continuación.

## 2.2 La función de verosimilitud relativa

Es conveniente estandarizar la función de verosimilitud respecto a su máximo para obtener una única representación que no involucre una constante arbitraria. El resultado de esto es la función de verosimilitud relativa, Sprott (2000).

**Definición 2.2** *La función de verosimilitud relativa de  $\theta$  se define como*

$$R(\theta; x) = \frac{L(\theta; x)}{\max_{\theta \in \Theta} L(\theta; x)} = \frac{L(\theta; x)}{L(\hat{\theta}; x)}, \quad (2.4)$$

donde  $\hat{\theta} = \hat{\theta}(x)$  es el valor que maximiza a  $L(\theta; x)$  y es llamado el **estimador de máxima verosimilitud** (EMV) de  $\theta$ .

El EMV de  $\theta$ ,  $\hat{\theta}$ , es el valor más plausible de  $\theta$  a la luz de la muestra observada. Es decir,  $\hat{\theta}$  es el valor del parámetro  $\theta$  que hace más probable a la muestra observada.

Puesto que  $\Pr\{X = x; \theta\}$  es una función de probabilidad, tiene que ser acotada entre 0 y 1. Por tanto, el denominador en (2.4) existe y es finito. Así, la función de verosimilitud relativa se encuentra entre cero y uno,  $0 \leq R(\theta; x) \leq 1$ , para todo valor de  $\theta \in \Theta$ . La función de verosimilitud relativa mide la plausibilidad de algún valor específico  $\theta$  relativo al de  $\hat{\theta}$ . Valores del parámetro  $\theta$  que tienen una verosimilitud relativa  $R(\theta; x)$  cercana a uno indican que son muy plausibles dada la muestra observada. En contraste, valores de  $\theta$  que tienen una verosimilitud relativa cercana a cero son implausibles.

## 2.3 Propiedades de la verosimilitud

### 2.3.1 Funciones de verosimilitud no son funciones de probabilidad

Es importante enfatizar que la función de verosimilitud no es una función de probabilidad. La verosimilitud es una función puntual que va del espacio parametral  $\Theta$  hacia los reales  $\mathbb{R}$ . En contraste, la probabilidad es una función de conjuntos que tiene como dominio una sigma-álgebra y como contradominio el intervalo  $[0,1]$ . Por ejemplo, una función de verosimilitud no tiene porque integrar a uno sobre su dominio, en cambio toda función de probabilidad integra a uno.

### 2.3.2 Combinación de observaciones

Una propiedad de la verosimilitud es que puede combinar datos de diferentes experimentos de una manera muy simple. Puesto que la probabilidad conjunta de

eventos independientes es el producto de sus probabilidades, la verosimilitud de  $\theta$ , definida en (2.1), basada en conjunto de datos independientes es el producto de la verosimilitud individual basada en cada uno de estos conjuntos de datos.

Puesto que en algunas situaciones es más fácil trabajar con la función logaritmo natural en procesos de optimización, es común tomar el logaritmo de la función de verosimilitud,  $\log L(\theta; x)$  y utilizar esta función en lugar de la propia verosimilitud para calcular el *EMV* de  $\theta$ . Esto es posible ya que el valor de  $\theta$  que maximiza a la  $L(\theta; x)$  es el mismo que maximiza a  $\log L(\theta; x)$  pues la función logaritmo natural es estrictamente creciente. Al logaritmo de la verosimilitud se le conoce como la función de *logverosimilitud*. Nótese que la logverosimilitud basada en una muestra independiente resulta ser la suma de los logaritmos de verosimilitud individuales basada en cada uno de estos conjuntos de datos. Es importante resaltar que este enfoque nos permite de una manera relativamente simple combinar información de diferentes experimentos para hacer inferencia sobre un mismo parámetro de interés.

### 2.3.3 Invarianza funcional

Otra de las propiedades que tiene la verosimilitud es la invarianza funcional. Esto quiere decir que para cualquier declaración cuantitativa acerca del parámetro  $\theta$  implica una declaración cuantitativa correspondiente a cualquier función uno a uno del parámetro  $\theta$ . En otras palabras, esta propiedad es de gran importancia ya que existen situaciones en donde el interés principal no radica en el parámetro  $\theta$  mismo sino en otro parámetro que se encuentra en función de  $\theta$ .

Si  $R_\theta(\theta; x)$  es la función de verosimilitud relativa de  $\theta$ , entonces la verosimilitud relativa de  $\delta = \delta(\theta)$  es  $R_\delta(\delta; x) = R_\theta[\theta(\delta); x]$  y  $\hat{\delta} = \delta(\hat{\theta})$ . Por ejemplo, si  $\theta > 0$  y  $\delta = \log \theta$ , entonces el *EMV* para  $\delta$  es  $\hat{\delta} = \log \hat{\theta}$  y su función de verosimilitud relativa es  $R_\delta(\delta; x) = R_\theta(\exp(\delta); x)$ . También  $a \leq \theta \leq b \iff \log a \leq \delta \leq \log b$ , Sprott (2000).

## 2.4 Intervalos de verosimilitud

Un estimador puntual sobre un parámetro  $\theta$  no especifica ni da idea acerca de la incertidumbre de esta declaración. Otra forma de estimación es por intervalos. A continuación se definen los intervalos de verosimilitud para un parámetro  $\theta$

**Definición 2.3** *Un intervalo o región de verosimilitud de nivel  $p$  para  $\theta$ , se define como*

$$IV(p) = \{\theta : R(\theta; x) \geq p\}, \quad (2.5)$$

donde  $0 \leq p \leq 1$ .

Cada valor específico de  $\theta$  dentro de la región de verosimilitud  $IV(p)$  tiene  $R(\theta) \geq p$  y todo valor de  $\theta$  fuera del intervalo  $IV(p)$  tiene verosimilitud relativa  $R(\theta) < p$ . De esta manera, los intervalos o regiones de verosimilitud separan los valores plausibles del parámetro  $\theta$  de los no plausibles a un nivel  $p$ . En el caso de que  $\theta$  sea un escalar

las regiones de verosimilitud a cierto nivel de plausibilidad  $p$  pueden ser un intervalo o la unión de intervalos disjuntos, esto va a depender de la forma que presente la gráfica de la función de verosimilitud, Sprott (2000).

Cuando se considera al parámetro  $\theta$  unidimensional, los intervalos se pueden obtener trazando una línea recta horizontal en la gráfica de  $R(\theta; x)$  a un nivel  $p$  paralela al eje cartesiano  $\theta$ . Variando el nivel de plausibilidad  $p$  entre 0 y 1 se produce un conjunto de intervalos anidados los cuales convergen al  $EMV$ ,  $\hat{\theta}$ , cuando  $p$  tiende a 1. El  $EMV$ ,  $\hat{\theta}$ , está contenido en todos los intervalos ya que  $R(\hat{\theta}; x) = 1$ . Así, este conjunto de intervalos anidados es equivalente a la función de verosimilitud y reproduce la gráfica de  $R(\theta; x)$ .

Por otra parte, un sólo intervalo de verosimilitud no proporciona información acerca de la simetría de la función de verosimilitud respecto a  $\hat{\theta}$  ni de como cambia la plausibilidad de los valores de  $\theta$  dentro del intervalo. Por ello, se sugiere dar el  $EMV$  de  $\theta$  junto con los intervalos de nivel  $p = 0.036, 0.147$  y  $0.25$ . En la siguiente sección se verá que estos niveles de verosimilitud están asociadas a confianzas aproximadas del 99%, 95% y 90%, respectivamente.

## 2.5 Intervalos de verosimilitud-confianza

En esta sección se muestra que es posible asociar un nivel de confianza a los intervalos de verosimilitud bajo ciertas condiciones. Sea  $X = (X_1, \dots, X_n)$  un vector de variables aleatorias con función de probabilidad  $\Pr\{X = x; \theta\}$ , donde  $\theta = \theta_0$  es un escalar fijo y desconocido. A partir de esta muestra observada se puede calcular un intervalo  $[A, B]$  para el valor verdadero  $\theta_0$ . Cabe mencionar que cada vez que cambia la muestra, los valores de los extremos del intervalo  $[A, B]$  también pueden cambiar. La distribución de probabilidad de estas variables aleatorias  $A$  y  $B$  se puede calcular a partir de la distribución del vector aleatorio  $X$  y generalmente depende de  $\theta_0$ . Obsérvese que como el intervalo  $[A, B]$  puede cambiar a medida que cambia la muestra entonces algunas veces puede o no incluir al valor verdadero del parámetro,  $\theta_0$ .

**Definición 2.4** La *probabilidad de cobertura* de un intervalo aleatorio  $[A, B]$  es la probabilidad de que el intervalo  $[A, B]$  incluya o cubra el valor verdadero del parámetro  $\theta_0$ ,

$$PC(\theta_0) = \Pr\{A \leq \theta_0 \leq B; \theta = \theta_0\} \quad (2.6)$$

Podemos decir que la probabilidad de cobertura es la fracción de veces que el intervalo  $[A, B]$  contiene al valor verdadero del parámetro  $\theta_0$  en un número muy grande de repeticiones de la muestra con  $\theta = \theta_0$  fijo. Aunque existen definiciones más generales de probabilidad de cobertura, se adoptará esta definición porque es suficiente y clara para los objetivos de este trabajo.

**Definición 2.5** El intervalo  $[A, B]$  es un *intervalo de confianza* para  $\theta$  si la probabilidad de cobertura no depende del parámetro  $\theta_0$ , Kalbfleisch (1985, pág 113).

La probabilidad de cobertura de un  $IV(p)$  para  $\theta = \theta_0$ , se puede calcular de la siguiente manera. Nótese que

$$\theta_0 \in IV(p) = \{\theta : R(\theta; x) \geq p\} \text{ si y sólo si } R(\theta_0; x) \geq p.$$

Así,

$$\theta_0 \in IV(p) \text{ si y sólo si } D_n \leq -2 \log(p)$$

donde  $D_n \equiv -2 \log R(\theta_0; x)$ . Por lo tanto, la probabilidad de cobertura del  $IV(p)$  es

$$\begin{aligned} PC(\theta_0) &= \Pr\{\theta_0 \in IV(p); \theta = \theta_0\} \\ &= \Pr\{D_n \leq -2 \log(p); \theta = \theta_0\}. \end{aligned} \quad (2.7)$$

En muchas ocasiones es difícil encontrar la distribución de probabilidad exacta de  $D_n$ . Sin embargo, bajo ciertas condiciones de regularidad, se tiene el siguiente Teorema para la distribución límite de  $D_n$  para todo  $\theta_0 \in \Theta$ . Tales condiciones se enuncian a continuación.

**Condiciones de Regularidad.** Sea  $\Theta$  un intervalo abierto, no necesariamente finito, (Serfling, 1980).

**R1** Para  $\theta \in \Theta$ , las derivadas

$$\frac{\partial \log f(x; \theta)}{\partial \theta}, \frac{\partial^2 \log f(x; \theta)}{\partial \theta^2} \text{ y } \frac{\partial^3 \log f(x; \theta)}{\partial \theta^3}$$

existen para toda  $x$ .

**R2** Para  $\theta_0 \in \Theta$  existen funciones  $g(x)$ ,  $h(x)$  y  $H(x)$ , tal que para  $\theta$  en una vecindad de  $\theta_0$  se cumple que

$$\left| \frac{\partial \log f(x; \theta)}{\partial \theta} \right| \leq g(x), \left| \frac{\partial^2 \log f(x; \theta)}{\partial \theta^2} \right| \leq h(x) \text{ y } \left| \frac{\partial^3 \log f(x; \theta)}{\partial \theta^3} \right| \leq H(x)$$

para toda  $x$ . Además,  $\int g(x)dx < \infty$ ,  $\int h(x)dx < \infty$  y  $E_\theta[H(x)] < \infty$ .

**R3** Para cada  $\theta \in \Theta$ , se satisface que

$$0 < E_\theta \left[ \frac{\partial \log f(x; \theta)}{\partial \theta} \right]^2 < \infty.$$

La condición R1 asegura que la función  $\partial \log f(x; \theta) / \partial \theta$  tiene, para cada  $x$ , una expansión de Taylor como una función de  $\theta$ . La condición R2 asegura que las funciones  $\int f(x; \theta)dx$  y  $\int [\partial \log f(x; \theta) / \partial \theta] dx$  pueden ser diferenciables con respecto a  $\theta$  bajo la integral. La condición R3 establece que las variables aleatorias  $\partial \log f(x; \theta) / \partial \theta$  tienen varianza finita positiva.

**Teorema 2.1** Bajo  $H_0 : \theta = \theta_0$ , la estadística  $D_n$  converge en distribución a una Ji-cuadrada con un grado de libertad,  $\chi_{(1)}^2$ .

**Demostración.** Véase Serfling (1980, pág. 155). ■

Las ideas principales de la demostración son las siguientes. Se realiza la expansión en serie de Taylor de la función logverosimilitud, que se denotará como  $\ell_n(\theta) = \log L(\theta; x)$ , alrededor de  $\theta = \hat{\theta}_n$ ,

$$\ell_n(\theta_0) = \ell_n(\hat{\theta}_n) + (\theta_0 - \hat{\theta}_n)\ell'_n(\hat{\theta}_n) + \frac{(\hat{\theta}_n - \theta_0)^2}{2!}\ell''_n(\hat{\theta}_n) + R_1(\theta_0, \hat{\theta}_n).$$

Nótese que  $\ell'_n(\hat{\theta}_n) = 0$  puesto que  $\hat{\theta}$  es el *EMV* de  $\theta$ . Ahora al pasar  $\ell_n(\hat{\theta}_n)$  al lado izquierdo de la igualdad anterior se obtiene que

$$\ell_n(\theta_0) - \ell_n(\hat{\theta}_n) = \frac{(\hat{\theta}_n - \theta_0)^2}{2!}\ell''_n(\hat{\theta}_n) + R_1(\theta_0, \hat{\theta}_n).$$

Así, multiplicando por  $-2nI(\theta_0)/nI(\theta_0)$  se tiene que

$$D_n = -2 \left[ \ell_n(\theta_0) - \ell(\hat{\theta}_n) \right] = n(\hat{\theta}_n - \theta_0)^2 I(\theta_0) \left[ \frac{\ell''_n(\hat{\theta}_n)}{nI(\theta_0)} + R_2(\theta_0, \hat{\theta}_n) \right].$$

Por asintoticidad normal de  $\hat{\theta}$  se cumple que

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \sqrt{I(\theta_0)} \xrightarrow{d} N(0, 1) \implies n(\hat{\theta}_n - \theta_0)^2 I(\theta_0) \xrightarrow{d} \chi^2_{(1)},$$

(Serfling, 1980, pág 145, Teorema 4.4.2). Por otro lado, por la Ley Fuerte de los Grandes Números (Hoel, pág 58) se tiene la siguiente convergencia casi segura (*c.s.*)

$$-\frac{\ell''_n(\theta_0)}{n} \xrightarrow{c.s.} I(\theta_0) \implies -\frac{\ell''_n(\hat{\theta}_n)}{nI(\theta_0)} \xrightarrow{P} 1.$$

Además, por consistencia del *EMV* de  $\theta$  y por las condiciones de regularidad antes mencionadas, se obtiene que el residuo  $R_2(\theta_0, \hat{\theta}_n)$  converge en probabilidad a cero,

$$R_2(\theta_0, \hat{\theta}_n) \xrightarrow{P} 0.$$

Entonces se sigue que

$$\left[ \frac{\ell''_n(\hat{\theta}_n)}{nI(\theta_0)} + R_2(\theta_0, \hat{\theta}_n) \right] \xrightarrow{P} 1.$$

Por último, aplicando el Teorema de Slutsky (Casella y Berger, 2002)

$$D_n \xrightarrow{d} \chi^2_{(1)}.$$

Del teorema anterior, se tiene que

$$\lim_{n \rightarrow \infty} \Pr\{D_n \leq d; \theta = \theta_0\} = \Pr\{\chi^2_{(1)} \leq d\}, \text{ para todo } d > 0. \quad (2.8)$$

Así, la probabilidad de cobertura del *IV*( $p$ ) dada en (2.8) es aproximadamente  $\Pr\{D_n \leq d; \theta = \theta_0\}$ , donde  $d = -2 \log(p)$ . Despejando el nivel de verosimilitud en

términos de  $d$  se obtiene que  $p = \exp(-d/2)$ . Así, tomando  $d = q_{(\alpha,1)}$ , donde  $q(\alpha,1)$  es el cuantil  $(1 - \alpha)$  de una distribución Ji-cuadrada con un grado de libertad, el intervalo de verosimilitud  $IV(p)$  hereda una probabilidad de cobertura aproximada del  $100(1 - \alpha)\%$  donde  $p = \exp(-q_{(\alpha,1)}/2)$ . Por tanto, el  $IV(p)$  también es un intervalo o región de confianza para  $\theta$ , al cual se le nombra **intervalo o región de verosimilitud-confianza**.

Los valores 2.706, 3.841 y 6.635 son los cuantiles 0.90, 0.95 y 0.99 de una distribución Ji-cuadrada con un grado de libertad, respectivamente. Sustituyendo estos valores en  $p = \exp(-q_{(\alpha,1)}/2)$ , se obtiene que los  $IV(p)$  para  $p = 0.258, 0.147$  y  $0.036$  tienen una probabilidad de cobertura aproximada del 90%, 95% y 99% respectivamente.

## 2.6 Ejemplo

Para ejemplificar el uso de la función de verosimilitud relativa para hacer inferencias sobre un parámetro escalar, se considerará el siguiente ejemplo. Sea  $X$  una *v.a.* binomial con la función de probabilidad

$$\Pr\{X = k; n, \theta\} = \binom{n}{k} \theta^k (1 - \theta)^{n-k}.$$

En este caso se considera a  $n$  fijo y conocido, y el parámetro a estimar es  $\theta$ . Entonces, para una muestra observada  $X = x$ , la función de verosimilitud de  $\theta$  dada en (2.1) se define como

$$L(\theta; x) \propto \theta^x (1 - \theta)^{n-x}.$$

Para encontrar el *EMV* de  $\theta$  se procederá de la siguiente manera. Tomando logaritmo natural se obtiene la función logverosimilitud de  $\theta$ ,

$$\log L(\theta; x) = x \log \theta + (n - x) \log (1 - \theta).$$

Derivando e igualando a cero, se obtiene un punto crítico

$$\theta = \frac{x}{n}.$$

Utilizando el criterio de la segunda derivada se verificará si este punto crítico maximiza a la función de verosimilitud de  $\theta$ . Para  $1 \leq x \leq n - 1$  se tiene que

$$\left. \frac{\partial \log L(\theta;)}{\partial \theta^2} \right|_{\theta = \frac{x}{n}} = \frac{-n^2}{x} - \frac{n^2}{n - x} < 0.$$

Entonces la función de verosimilitud de  $\theta$  tiene un máximo local en  $\theta = x/n$ . Además, puesto que  $L(\theta) = 0$  para  $\theta = 0$  y  $\theta = 1$  se tiene que el *EMV* de  $\theta$  es  $\hat{\theta} = x/n$ .

Utilizando la definición verosimilitud relativa dada en (2.4), se tiene que

$$R(\theta; x) = \left( \frac{\theta}{\hat{\theta}} \right)^x \left( \frac{1 - \theta}{1 - \hat{\theta}} \right)^{n-x}.$$



sustituyendo el  $EMV$  en la última expresión, la función de verosimilitud queda expresada como

$$R(\theta; x) = n^n \left( \frac{\theta}{k} \right)^k \left( \frac{1 - \theta}{n - k} \right)^{n-k}.$$

La Figura 2.1 muestra la gráfica de la función de verosimilitud relativa de  $\theta$ , los intervalos de verosimilitud a los niveles de  $p = 0.036, 0.15$  y  $0.25$  y al  $EMV$  de  $\theta$ ,  $\hat{\theta} = 0.6$ . Se observa que la función de verosimilitud relativa de  $\theta$  se encuentra un poco sesgada a la izquierda. Esto también se puede ver al notar la ubicación de  $\hat{\theta}$  con respecto a los límites del intervalo.

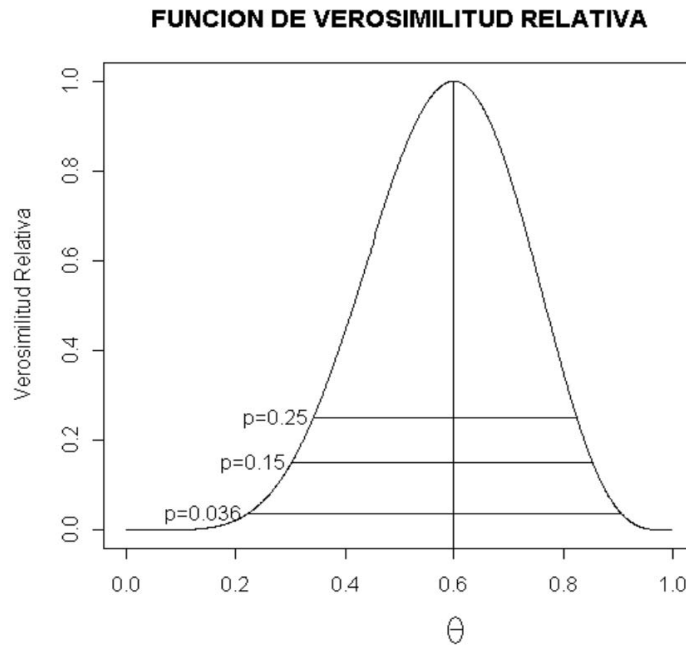


Fig. 2.1 Binomial  $n = 10, x = 6$ .

p	Inferior	EMV	Superior
0.036	0.2211018	0.6	0.906116
0.15	0.3016381	0.6	0.8531966
0.25	0.3416063	0.6	0.824376

Tabla 2.1 Resumen de inferencias para el ejemplo Binomial.

Un resumen de las inferencias para  $\theta$  bajo esta metodología son las que se muestran en la Tabla 2.1. Se considerarán los niveles de verosimilitud para  $p = 0.036, 0.15$  y  $0.25$  puesto que están relacionados a niveles de confianza aproximados del 99%, 95% y 90%. Nótese, que valores de  $\theta < 0.22$  y valores de  $\theta > 0.9$  son muy implausibles.



## Capítulo 3

### Verosimilitud para Cadenas de Markov

En los capítulos previos se habló en general de los conceptos básicos de la teoría de probabilidad para cadenas de Markov, así como de la teoría de verosimilitud para inferencia estadística. En esta parte se verá como el enfoque de verosimilitud se aplica a cadenas de Markov. En particular, se desea mostrar el proceso inferencial sobre las probabilidades de transición de una cadena de Markov paramétrica y no paramétrica. Dicho proceso se ejemplificará con datos reales de direcciones del viento que serán modelados con una cadena de Markov no paramétrica.

Sea  $\{X_n\}_{n \in \mathbb{N}}$  una Cadena de Markov irreducible (ergódica) con espacio de estados  $T = \{1, 2, \dots, m\}$ , probabilidades de transición  $P_{ij}$ , distribución inicial  $\pi_0(i)$  y distribución estacionaria (límite)  $\pi(j)$ . Cabe señalar aquí que en esta tesis los parámetros a estimar son precisamente estas probabilidades de transición. Se considerarán dos casos: **Paramétrico y No – Paramétrico**.

#### 3.1 Caso paramétrico

Aquí las probabilidades de transición pueden ser indexadas por medio de un vector de parámetros. Supóngase  $P_{ij} = P_{ij}(\theta)$ , donde  $\theta = (\theta_1, \theta_2, \dots, \theta_w)$  es un vector de parámetros desconocidos que toma valores en  $\mathbb{R}^w$ .

Sea  $x = \{x_1, \dots, x_n\}$  una realización de una cadena de Markov de longitud  $n$ . Recuerdese que la función de verosimilitud dada en (2.1) se define como proporcional a la probabilidad de ocurrencia de la muestra. Entonces utilizando la probabilidad conjunta de la muestra dada en (1.3) se tiene que

$$L(P_{ij}(\theta); x) \propto \pi_0(i_0) \prod_{k=1}^n P_{x_{k+1}x_k} = \pi_0(i_0) \prod_{i,j=1}^m [P_{ij}(\theta)]^{n_{ij}}, \quad (3.1)$$

donde  $n_{ij}$  son las frecuencias de las transiciones en un paso de ir de un estado  $i$  a un estado  $j$ . Tomando logaritmo natural se tiene la siguiente expresión para la logverosimilitud

$$\log L(P_{ij}(\theta); x) = \log \pi_0(i_0) + \sum_{i,j} n_{ij} \log P_{ij}(\theta). \quad (3.2)$$

Aquí se supondrá que  $\pi_0(i_0)$  no tiene información de las  $P_{ij}(\theta)$ . En el caso de que  $n$  sea grande el efecto de  $\pi_0(i_0)$  puede ser ignorado, Basawa (1980, pág.53). Así,  $\log \pi_0(i_0)$  puede ser considerado una constante en la Ecuación (3.2).

Los estimadores de máxima verosimilitud de los parámetros  $\theta_1, \theta_2, \dots, \theta_w$  pueden obtener resolviendo el sistema de ecuaciones

$$\frac{\partial \log L(P_{ij}(\theta); x)}{\partial \theta_r} = 0, \quad r = 1, 2, 3, \dots, w.$$

Es decir,

$$\sum_{i,j=1}^m \left[ n_{ij} \frac{\partial P_{ij}(\theta)}{\partial \theta_r} \cdot \frac{1}{P_{ij}(\theta)} \right] = 0, \quad r = 1, 2, 3, \dots, w, \quad (3.3)$$

simultáneamente para  $\theta_1, \theta_2, \dots, \theta_w$ .

**Ejemplo 3.1** *Considérese una cadena de Markov paramétrica con espacio de estados  $T = \{1, 2\}$  y matriz de transición*

$$P = \begin{bmatrix} \theta & 1 - \theta \\ 1 - \theta^\delta & \theta^\delta \end{bmatrix}, \quad (3.4)$$

donde  $\delta = 6$ . Cabe señalar que esta estructura de dependencia paramétrica de las probabilidades de transición puede ser un poco artificial pero se utilizará sólo para fines de ilustrar las inferencias a través del enfoque de verosimilitud para esta situación. De (3.1), se tiene que la función de verosimilitud para  $\theta$  es

$$L(\theta; x) \propto \theta^{(n_{11} + 6n_{22})} (1 - \theta)^{n_{12}} (1 - \theta^6)^{n_{21}}.$$

Así, usando la definición de la función de verosimilitud relativa dada en (2.4) se tiene que

$$R(\theta; x) = \left( \frac{\theta}{\hat{\theta}} \right)^{(n_{11} + 6n_{22})} \left( \frac{1 - \theta}{1 - \hat{\theta}} \right)^{n_{12}} \left( \frac{1 - \theta^6}{1 - \hat{\theta}^6} \right)^{n_{21}}.$$

En la Figura 3.1 se muestra la gráfica de verosimilitud relativa de  $\theta$ , considerando  $n_{11} = 1527$ ,  $n_{12} = n_{21} = 243$  y  $n_{22} = 170$ . Estas frecuencias se obtuvieron de datos reales de direcciones del viento discretizados o clasificados en dos estados, norte (N) y sur (S). Mayor información acerca del registro de estos datos se presenta más adelante en la Sección 3.3. Además, en la Figura 3.1 también se presenta el EMV para  $\theta$ ,  $\hat{\theta} = 0.863$  (este valor se calculó utilizando métodos numéricos) y los intervalos

de verosimilitud-confianza a los niveles de  $p=0.036, 0.15$  y  $0.25$ .

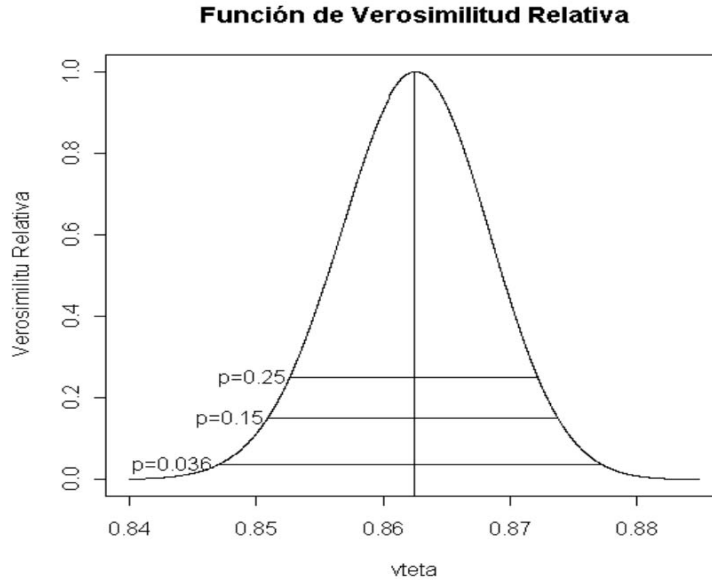


Fig. 3.1 Caso Paramétrico.

En la Tabla 3.1 se muestra el resumen de inferencias de verosimilitud para  $\theta$ . Nótese que valores de  $\theta$  menores que 0.847 y mayores que 0.88 son implausibles.

p	Inferior	EMV	Superior
0.036	0.8469	0.863	0.8773
0.15	0.8508	0.863	0.8738
0.25	0.8526	0.863	0.8722

Tabla 3.1 Resumen de inferencias para el ejemplo paramétrico.

La estimación puntual de la matriz de transición  $P_{ij}(\theta)$  se obtiene sustituyendo en (3.4) con  $\delta = 6$ ,

$$\widehat{P}_{ij} = \begin{bmatrix} 0.863 & 0.137 \\ 0.60 & 0.40 \end{bmatrix} \quad (3.5)$$

Esta matriz se puede interpretar de la siguiente manera. Si el viento se encuentra en dirección norte (N), es más probable (0.863) permanecer en el mismo estado norte (N) que cambiar o pasar a dirección sur (S). En cambio, si la dirección del viento se encuentra en sur (S), es más probable (0.6) cambiar a norte que permanecer en la misma dirección (0.4).

A continuación se verá como simular una cadena de Markov paramétrica con el objetivo de utilizar este proceso para verificar las probabilidades de cobertura asociadas a los intervalos de verosimilitud.

### 3.1.1 Simulación de una cadena de Markov paramétrica

En esta sección se verificarán las probabilidades de coberturas asociadas a los intervalos de verosimilitud a través de un estudio de simulación. Aquí se trabajará con simulaciones de una cadena de Markov paramétrica y en la Sección 3.2.1 se realizarán simulaciones en el caso no paramétrico. También se incluirán los algoritmos de los programas utilizados para obtener las simulaciones de una cadena de Markov finita.

Para realizar la simulación de una cadena de Markov finita se utilizó el siguiente diagrama de flujo

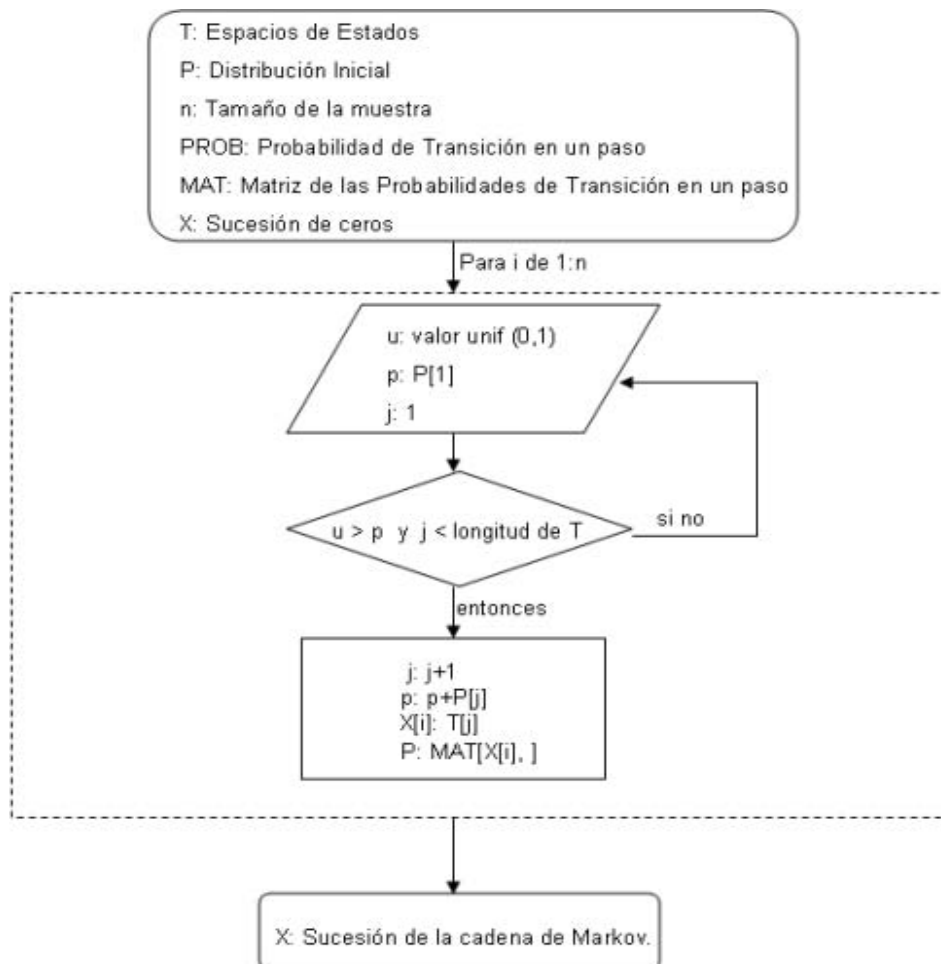


Fig. 3.2 Sucesión de una cadena de Markov finita.

Es importante señalar que con este algoritmo se obtienen sucesiones de cadenas de Markov tanto en el caso paramétrico como en el no paramétrico y con las características que se deseen, las cuales deben ser proporcionadas al inicio del programa.

Para realizar la simulación se consideró el Ejemplo 3.1 en donde se tiene una cadena de Markov paramétrica con espacio de estados  $T = \{1,2\}$  y matriz de transición

estimada dada en (3.5). Se considerará como distribución inicial a la distribución estacionaria dada por

$$\pi_0(1) = \Pr\{X_0 = 1\} = 0.81 \text{ y } \pi_0(2) = \Pr\{X_0 = 2\} = 0.19;$$

Sin embargo, es necesario mencionar que se probaron otras distribuciones iniciales sin alterar las conclusiones que se verán más adelante. Se tomó una muestra de tamaño  $n = 2184$  datos provenientes de una cadena de Markov paramétrica simulada con las características antes mencionadas y se realizará el proceso inferencial vía verosimilitud. Nótese que las probabilidades de transición en este ejemplo se encuentra en términos del parámetro  $\theta$ .

En la función de verosimilitud definida en (3.1) aparece el término  $n_{ij}$  que indica las frecuencias de las transiciones en un paso de ir de un estado  $i$  a un estado  $j$  en toda la muestra observada. Por ello, se realizó otro programa que proporcionará y colocará estos valores en una matriz que se le conoce como matriz de frecuencias  $F$ . En el siguiente diagrama de flujo se observa como se obtiene esta matriz para una sucesión  $X_n$  de una cadena de Markov finita.

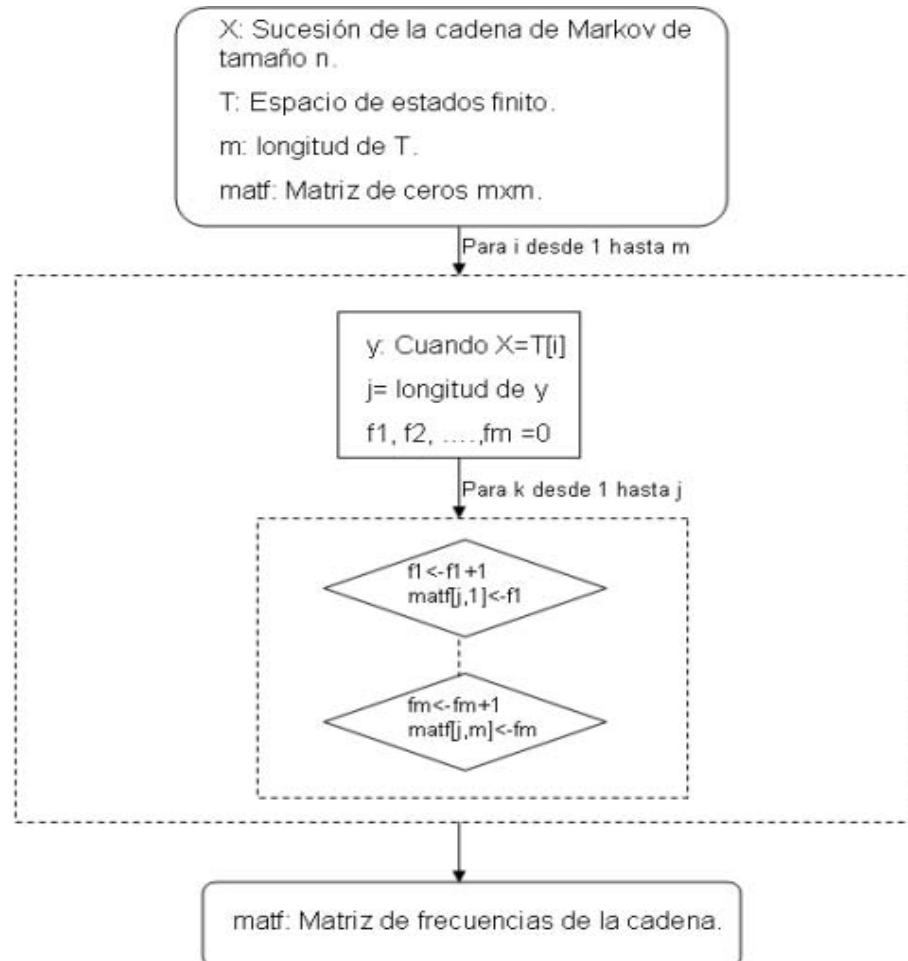


Fig. 3.3 Matriz de frecuencias de  $i$  a  $j$ .

Nótese que esto es de gran utilidad al momento de hacer los cálculos para realizar la inferencia. Una vez proporcionada la matriz de frecuencias  $F$  se continúa con el proceso inferencial.

Aplicando toda la teoría de la Sección 3.1, se tiene los resultados de la inferencia y se presentan a continuación

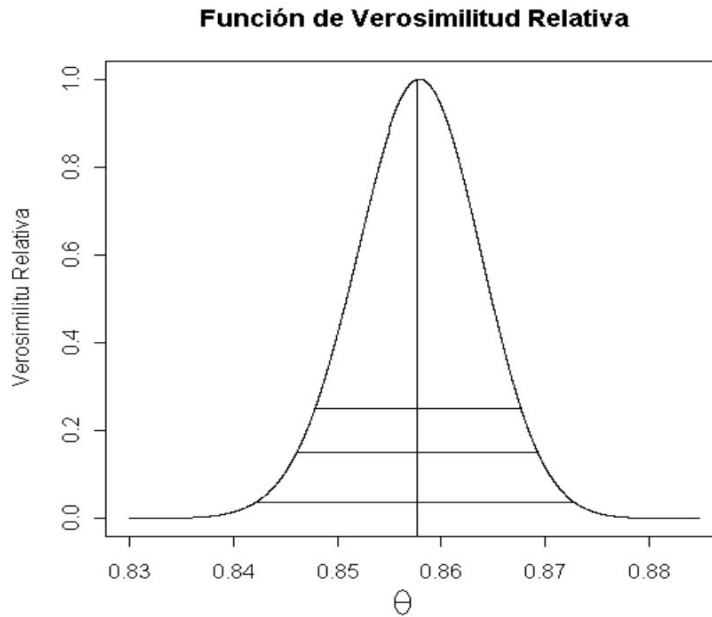


Fig. 3.4 Simulación particular de una cadena de Markov Paramétrica.

En la Figura 3.4 se tiene la gráfica de la función de verosimilitud relativa en la cual se observa el  $EMV$  de  $\theta$  y los intervalos de verosimilitud a los diferentes niveles.

p	Inferior	EMV	Superior
0.036	0.84211	0.858	0.8729
0.147	0.84608	0.858	0.86935
0.25	0.84784	0.858	0.86773

Tabla 3.2 Resumen de inferencias para simulación paramétrica.



En la Tabla 3.2 se muestra el resumen de las inferencias de verosimilitud para  $\theta$ , en donde se observa que valores para  $\theta$  mayores a 0.873 y menores a 0.842 son poco plausibles. Por otra parte, las probabilidades de transición estimadas, en términos de  $\theta$  son

$$\begin{bmatrix} 0.858 & 0.142 \\ 0.60 & 0.399 \end{bmatrix},$$

lo que indica que al estar en el estado uno es altamente probable permanecer en el mismo y que es poco probable cambiar al estado dos. Sin embargo, estando en el estado dos, es más probable pasar al estado uno, que permanecer en el mismo estado 2.

Recuerdese que el interés aquí es verificar la confianza asociada a los intervalos de verosimilitud. Así, una vez que se calcularon los intervalos de verosimilitud a los diferentes niveles de plausibilidad  $p = 0.036, 0.147, 0.25$  se verificó si el intervalo cubre o no el verdadero valor del parámetro  $\theta$ . Se realizó este proceso 10000 veces y se estimaron las probabilidades de cobertura con las frecuencias relativas observadas. Los resultados fueron los siguientes

Nivel $p$	$1 - \alpha$	$PC(\theta)$
0.036	0.99	0.994
0.147	0.95	0.964
0.25	0.90	0.933

Tabla 3.3 Resultados de Probabilidades de Coberturas.

En la Tabla 3.3 se observan las probabilidades de coberturas estimadas y se puede observar que son similares a lo teóricamente indicado. Por lo que podemos concluir que a estos intervalos de verosimilitud se les puede asociar una confianza del 99%, 95% y 90%. Por otra parte, los resultados de este estudio de simulación no se ven afectados si se da como distribución inicial  $\pi_0$  a la distribución estacionaria o cualquier otra distribución en el sentido de que las probabilidades de coberturas estimadas son muy parecidas.

Es importante mencionar que muchas veces puede ser complicado obtener un modelo paramétrico para un fenómeno que se desea estudiar, por eso otra opción que se tiene es trabajar con modelos no paramétricos, lo cual se verá a continuación.

## 3.2 Caso no paramétrico

Este caso, como su nombre lo dice, es aquel en donde las  $P_{ij}$  no pueden ser indexadas a través de un vector de parámetros. Nótese que  $P$  es la matriz de las probabilidades de transición, es decir,  $P = [P_{ij}]$ .

Sea  $x = \{x_1, \dots, x_n\}$  una realización de una cadena de Markov de longitud  $n$ . Recuerdese que la función de verosimilitud dada en (2.1) se define como proporcional

a la probabilidad de ocurrencia de la muestra pero como función del parámetro. Por tanto, utilizando la probabilidad conjunta dada en (1.3) se tiene que

$$L(P_{ij}; x) \propto \Pr\{x; P_{ij}\} = \pi_0(i_0) \prod_{k=1}^n P_{x_{k+1}x_k} = \pi_0(i_0) \prod_{i,j=1}^m P_{ij}^{n_{ij}} \quad (3.6)$$

donde  $n_{ij}$  son las frecuencias de las transiciones de ir del estado  $i$  al  $j$  en un paso en la muestra  $x$  y estas frecuencias se colocarán en una matriz  $F = [n_{ij}]$  llamada matriz de frecuencias. Para encontrar el *EMV* de  $P_{ij}$  se necesita maximizar a la función de verosimilitud y se procederá de la siguiente manera.

Tomando logaritmo natural de la verosimilitud dada en (3.6) se obtiene

$$\log L(P_{ij}; x) = \log \pi_0(i_0) + \sum_{i,j} n_{ij} \log P_{ij}, \quad (3.7)$$

sujeto a la restricción  $\sum_{j=1}^m P_{ij} = 1$  y utilizando multiplicadores de Lagrange se tiene la siguiente expresión

$$l(P_{ij}, x)^* = \log \pi_0(i_0) + \sum_{i,j} n_{ij} \log P_{ij} + \lambda(1 - \sum_{j=1}^m P_{ij}). \quad (3.8)$$

Derivando  $l^*$  respecto a las  $P_{ij}$  y a  $\lambda$ , se tiene

$$\frac{\partial l^*}{\partial P_{ij}} = \frac{n_{ij}}{P_{ij}} - \lambda \quad y \quad \frac{\partial l^*}{\partial \lambda} = 1 - \sum_{j=1}^m P_{ij}$$

igualando ambas a cero, se obtiene

$$\lambda = \frac{n_{ij}}{P_{ij}} \implies P_{ij} = \frac{n_{ij}}{\lambda}.$$

Sumando y usando que  $\sum_{j=1}^m P_{ij} = 1$ , se deduce que

$$\sum_{j=1}^m P_{ij} = \frac{\sum_{j=1}^m n_{ij}}{\lambda} \implies 1 = \frac{\sum_{j=1}^m n_{ij}}{\lambda} \implies \sum_{j=1}^m n_{ij} = \lambda.$$

Tomando  $\lambda = n_i$ , se tiene que el *EMV* es

$$\widehat{P}_{ij} = \frac{n_{ij}}{n_i}, \quad (3.9)$$

donde  $n_i = \sum_{j=1}^m n_{ij}$ .

Nótese que al momento de derivar (3.7) respecto a las  $P_{ij}$  se supondrá que  $\pi_0(i)$  no tiene información sobre de las probabilidades de transición. En el caso de que  $n$

sea grande el efecto del primer término en (3.7) puede ser ignorado, Basawa (1980, pág.53).

Para hacer inferencias sobre las probabilidades de transición  $P_{ij}$  se considerará la distribución marginal de  $n_{ij}$ . Si se fija  $i$  en la matriz de frecuencias  $F$ , se puede suponer que el vector  $(n_{i1}, n_{i2}, \dots, n_{im})$  sigue una distribución multinomial con parámetros  $(n_i, P_{i1}, P_{i2}, \dots, P_{im})$ . Ahora, al fijar  $j$  se tiene que  $n_{ij} \sim \text{binomial}(n_i, P_{ij})$ . Por lo que la función de verosimilitud de  $P_{ij}$  basada en la distribución marginal de  $n_{ij}$  es

$$L(P_{ij}) \propto P_{ij}^{n_{ij}} (1 - P_{ij})^{n_i - n_{ij}}. \quad (3.10)$$

Nótese que  $P_{ij}$  es una entrada de la matriz de transición  $P$ , las  $n_{ij}$  son una entrada de la matriz de frecuencias  $F$  y la suma de las frecuencias en el renglón  $i$  está denotada por  $n_i$ . Esta función de verosimilitud de  $P_{ij}$  basada en la distribución marginal de  $n_{ij}$  sirve para encontrar la estimación de un parámetro específico cuando se tienen varios y en este caso como ya se había dicho sólo se desea dar los intervalos de verosimilitud para cada una de las probabilidades de transición  $P_{ij}$ . Para esto, se necesita la función de verosimilitud relativa  $R(P_{ij})$ :

$$R(P_{ij}; x) = \frac{L(P_{ij}; x)}{L(\widehat{P}_{ij}; x)} = \left( \frac{P_{ij}}{\widehat{P}_{ij}} \right)^{n_{im}} \left( \frac{1 - P_{ij}}{1 - \widehat{P}_{ij}} \right)^{n_i - n_{ij}}. \quad (3.11)$$

Ya teniendo (3.11), se pueden encontrar los intervalos de verosimilitud a los diversos niveles de  $p$  para cada una de las probabilidades de transición. En la siguiente sección se simulará una cadena de Markov no paramétrica y se aplicará la teoría mencionada en esta parte.

### 3.2.1 Simulación de una cadena de Markov no paramétrica

En esta sección se realizará la simulación de una cadena de Markov no paramétrica finita. Con ello, se ejemplificará el proceso inferencial vía verosimilitud en este caso y se verificarán las probabilidades de coberturas asociadas a los intervalos de verosimilitud. Tales simulaciones se realizarán utilizando los mismos algoritmos computacionales mostrados en el caso paramétrico en la Sección 3.1.1.

Para llevar a cabo la simulación de una cadena de Markov no paramétrica se

consideró el espacio de estados  $T = \{1, \dots, 8\}$ , matriz de transición dada por

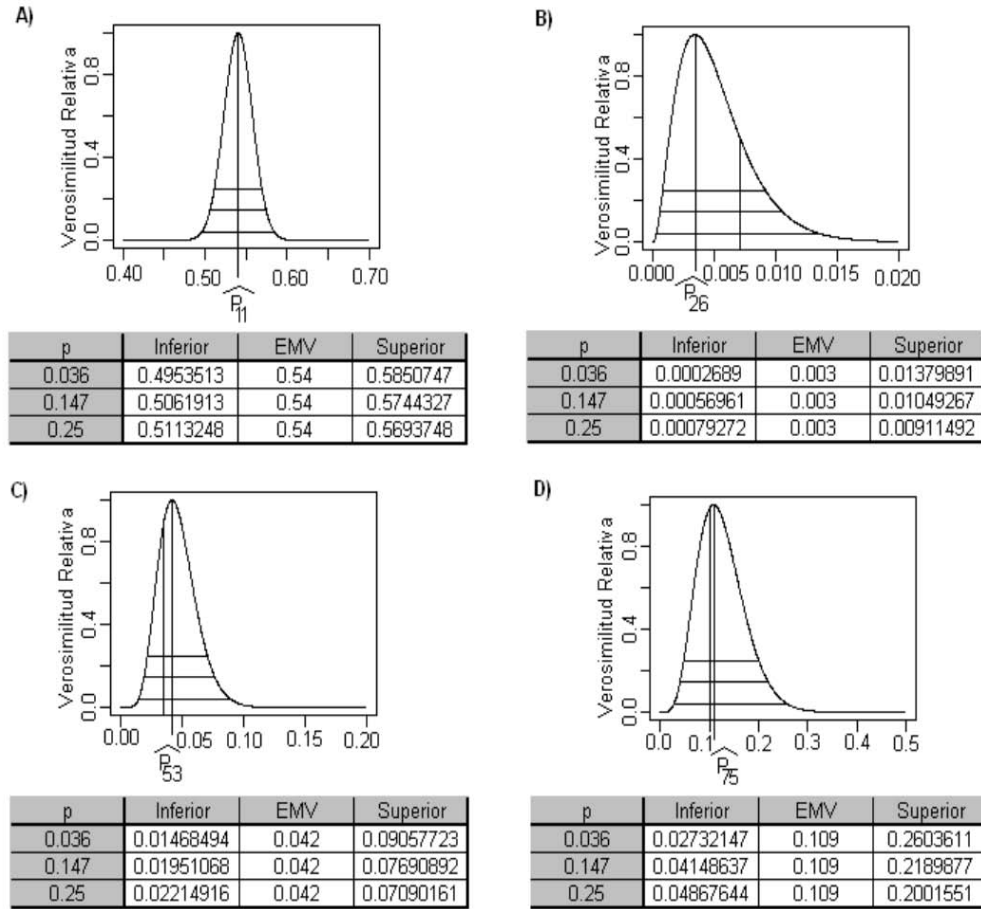
$$P = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{matrix} & \begin{bmatrix} 0.54 & 0.31 & 0.06 & 0.033 & 0.02 & 0.01 & 0.007 & 0.012 \\ 0.36 & 0.44 & 0.1 & 0.04 & 0.026 & 0.007 & 0.0062 & 0.009 \\ 0.25 & 0.28 & 0.27 & 0.083 & 0.06 & 0.03 & 0.01 & 0.005 \\ 0.18 & 0.11 & 0.1 & 0.31 & 0.18 & 0.08 & 0.035 & 0 \\ 0.12 & 0.075 & 0.031 & 0.22 & 0.32 & 0.16 & 0.07 & 0.006 \\ 0.16 & 0.057 & 0.023 & 0.034 & 0.25 & 0.27 & 0.12 & 0.08 \\ 0.37 & 0.1 & 0.06 & 0.02 & 0.1 & 0.12 & 0.14 & 0.06 \\ 0.36 & 0.2 & 0.033 & 0.23 & 0.066 & 0 & 0.033 & 0.066 \end{bmatrix} \end{matrix} \quad (3.12)$$

y distribución inicial

$$\begin{aligned} \pi_0(1) &= \Pr\{X_0 = 1\} = 0.3, & \pi_0(2) &= \Pr\{X_0 = 2\} = 0.21, \\ \pi_0(3) &= \Pr\{X_0 = 3\} = 0.15, & \pi_0(4) &= \Pr\{X_0 = 4\} = 0.12, \\ \pi_0(5) &= \Pr\{X_0 = 5\} = 0.13, & \pi_0(6) &= \Pr\{X_0 = 6\} = 0.04, \\ \pi_0(7) &= \Pr\{X_0 = 7\} = 0.027, & \pi_0(8) &= \Pr\{X_0 = 8\} = 0.023. \end{aligned}$$

A partir de una muestra de la cadena de tamaño  $n = 2184$  se estimaron cada una de las probabilidades de transición utilizando la expresión para  $\widehat{P}_{ij}$  dada en (3.9), de las cuales sólo se mostrarán las inferencias para las probabilidades encerradas en

la matriz de transición dada en (3.12)



3.5 Funciones de verosimilitud relativa en la simulación del caso no paramétrico

En la Figura 3.5 se tienen las gráficas de las funciones de verosimilitud relativa para las probabilidades de transición  $P_{11}$ ,  $P_{26}$ ,  $P_{53}$  y  $P_{75}$ , en donde se indican en cada una de estas el  $EMV$  y el valor verdadero del parámetro, así como los intervalos de verosimilitud en los niveles de plausibilidad  $p = 0.036$ ,  $0.147$  y  $0.25$ . También, en las tablas se presenta el resumen de las inferencias de verosimilitud, en donde se pueden observar los valores plausibles para cada probabilidad de transición. En particular, se tiene que  $EMV$  para  $P_{11}$  es  $0.54$  y que valores del parámetro menores que  $0.48$  y mayores que  $0.59$  son poco plausibles.

Para verificar las probabilidades de coberturas asociadas a los intervalos de verosimilitud, se calcularon los intervalos de verosimilitud a los cuales se les asocia una confianza y se verificó si el intervalo cubre o no el valor verdadero del parámetro verdadero  $P_{ij}$ . Se realizó el proceso 10000 veces y se calcularon las probabilidades

de cobertura con las frecuencias relativas observadas.

Nivel $p$	$1 - \alpha$	$PC(P_{11})$	$PC(P_{26})$	$PC(P_{53})$	$PC(P_{75})$
0.036	0.99	0.988	0.982	0.982	0.977
0.147	0.95	0.946	0.93	0.944	0.942
0.25	0.90	0.893	0.905	0.911	0.89

Tabla 3.4 Probabilidades de Cobertura Estimadas

En la Tabla 3.4 se presenta las probabilidades de cobertura estimadas para algunas  $P_{ij}$  observando que son similares a lo teóricamente indicado, por lo que se puede resumir que a los intervalos de verosimilitud de plausibilidad 0.036, 0.147 y 0.25 se les puede asociar una confianza del 99%, 95% y 90% respectivamente.

### 3.3 Problema de aplicación real

En esta sección se trabajará con datos reales provenientes de una cadena de Markov no paramétrica, con los cuales se ejemplificará la metodología de verosimilitud. El objetivo primordial no es sólo proporcionar la estimación puntual del parámetro de interés  $P$  sino también la estimación por intervalos de las probabilidades de transición de esta cadena. Además, se analizarán las propiedades probabilísticas que se presentan en el sistema y con ello ver el comportamiento asintótico de este.

Los datos que se utilizarán son direcciones del viento cuyas mediciones se encuentran en grados y fueron tomadas en la estación meteorológica del Dr. Miguel Nakamura ubicada en las instalaciones del Centro de Investigación en Matemáticas (CIMAT) en la Ciudad de Guanajuato, Guanajuato. Los registros de estos se realizaron cada hora durante el verano 2006 en dicha localidad.

La dirección del viento se define como la orientación del vector del viento en la horizontal. Para propósitos meteorológicos, la dirección del viento se define como la dirección desde la cual sopla el viento. Los vientos predominantes en la ciudad de Guanajuato, gracias a su latitud  $21^{\circ}00'59''\text{N}$  (distancia angular entre el ecuador y un punto determinado del planeta), son los alisios, es decir aquellos que provienen de dirección noreste (NE). Los vientos alisios constituyen el segmento final de uno de los regímenes de circulación global de aire (llamados celdas): el aire en el Ecuador tiende a ascender como resultado del fuerte calentamiento solar; una vez que llega a la parte alta de la atmósfera fluye en dirección NE, perdiendo gradualmente gran parte de su humedad y enfriándose; aproximadamente a latitud  $30^{\circ}\text{N}$ , el mismo aire desciende cálido y seco (produciendo desiertos en muchas partes del mundo en esta latitud); el aire ahora en la capa inferior de la atmósfera que regresa hacia el sur se desvía hacia el oeste debido a la rotación de la Tierra, convirtiéndose así en los predominantes vientos alisios propios de esta latitud. Para conocer un poco de donde se tomaron las observaciones mostramos una imagen de la ciudad de Guanajuato.



El problema estadístico que se desea abordar es el de hacer inferencia sobre las probabilidades de cambio de una dirección a otra en base un muestra observada de direcciones del viento. Por ello, se analizaron estos datos con la finalidad de modelarlos como una cadena de Markov. Se categorizaron según los puntos cardinales,



asociándose de la siguiente manera:

Puntos Cardinales	Intervalo Asociado (Grados)
N	[337.5,22.5)
NE	[22.5,67.5)
E	[67.5,112.5)
SE	[112.5,157.5)
S	[157.5,202.5)
SO	[202.5,247.5)
O	[247.5,292.5)
NO	[292.5,337.5)

y definiendo el espacio de estados como  $T = \{N, NE, E, SE, S, SO, O, NO\}$ . Luego, se realizó la gráfica del comportamiento de la cadena de direcciones y en la Figura 3.6

se muestra con los primeros cien datos.

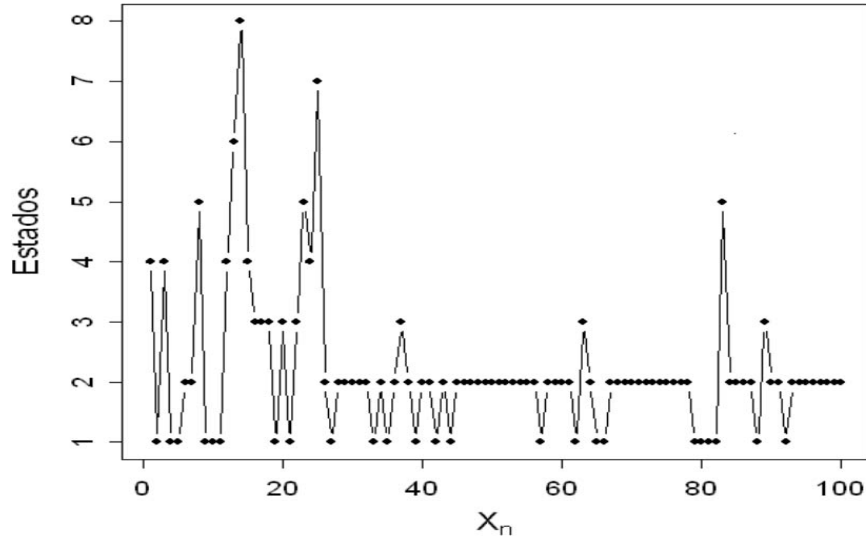


Fig. 3.6 Cadena de direcciones del viento.

Enseguida, se deseaba dar evidencia para suponer que estos datos podían ser modelados como una cadena de Markov por lo que se debía validar la propiedad de Markov definida en (1.1). Para ello, por cuestiones prácticas, se propone dar evidencia que se cumple la siguiente igualdad:

$$\Pr\{X_2 = i_2 | X_0 = i_0, X_1 = i_1\} = \Pr\{X_2 = i_2 | X_1 = i_1\}, \quad (3.13)$$

donde  $i_0, i_1, i_2 \in T$ . Para cada una de estas ternas, la probabilidad a la izquierda en (3.13) será estimada con

$$Y(i_0, i_1, i_2) = \frac{\text{No. de veces que } X_n \text{ pasa de } i_0, i_1, i_2}{\text{No. de veces que } X_n \text{ pasa de } i_0, i_1}. \quad (3.14)$$

Por otro lado, la probabilidad a la derecha en (3.13) será estimada con

$$X(i_1, i_2) = \frac{\text{No. de veces que } X_n \text{ pasa de } i_1, i_2}{\text{No. de veces que } X_n \text{ sale de } i_1}. \quad (3.15)$$

La propuesta gráfica de validación de la propiedad de Markov consiste en mostrar que para cada terna  $i_0, i_1, i_2 \in T$ , los puntos en el plano cartesiano conformados por las frecuencias observadas  $X$  y  $Y$  en (3.15) y (3.14) se encuentran alrededor de una línea recta de 45 grados de inclinación. Cualquier desviación de esto es considerado evidencia en contra de que la cadena cumple con la propiedad de Markov. Además, se recomienda realizar un histograma del cociente de estas dos probabilidades y observar que la mayor frecuencia esta concentrada alrededor del valor uno.



En la Figura 3.7 A y 3.7 B se muestran las gráficas de validación con los datos de direcciones del viento y se observa que la gran mayoría de puntos caen alrededor de la recta. Nótese que algunas probabilidades estimadas son iguales a cero, esto se debe a que dichas trayectorias, pasar de  $i_0$ ,  $i_1$  a  $i_2$ , no ocurrieron dentro de la muestra observada de la cadena de tamaño  $n = 2184$ . Además, se observa en el histograma que la frecuencia está concentrada en el número uno.

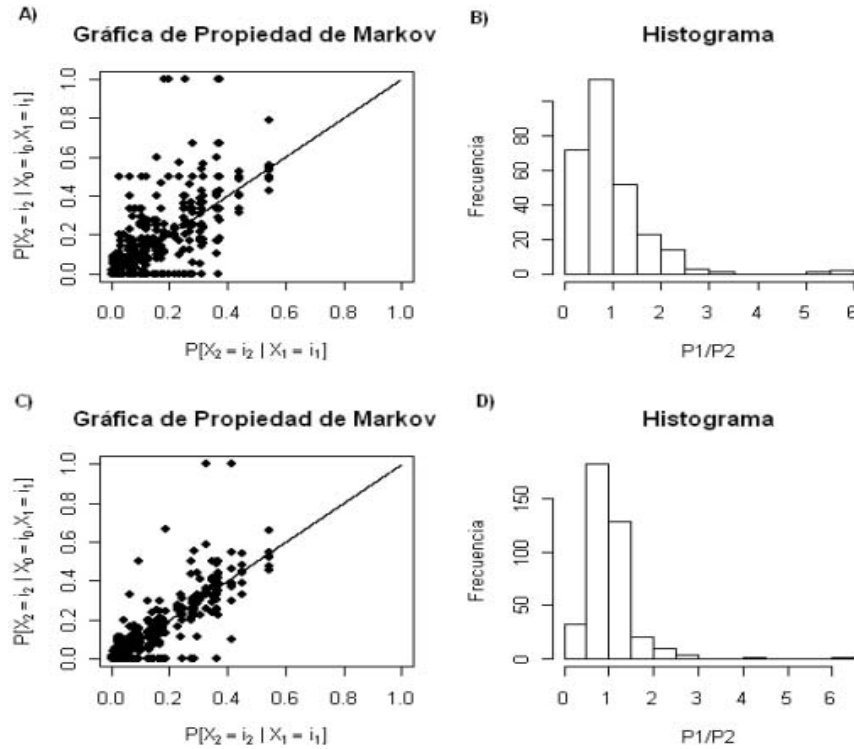


Fig. 3.7 Validación de la propiedad de Markov.

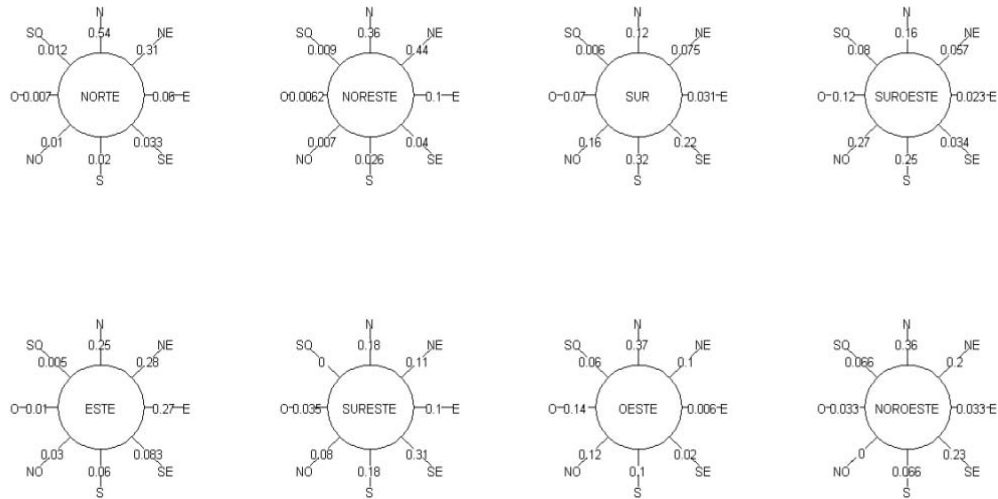
Por otra parte, en las Figura 3.7 C y 3.7 D se presentan las mismas gráficas pero con datos simulados de una cadena de Markov legítima. Se observa una gran concentración de puntos alrededor de la recta y se siguen presentando probabilidades estimadas iguales a cero con el tamaño de muestra  $n = 2184$ . En el histograma se tiene la mayor frecuencia en 1. El objetivo principal de mostrar estas gráficas con datos simulados es el de exhibir el comportamiento del método gráfico propuesto para validar la propiedad de Markov.

Con lo anterior se tiene evidencia para suponer que los datos de direcciones del viento pueden ser modelados como una cadena de Markov y puesto que las probabilidades de cambio entre una dirección y otra no se pudieron indexar por medio de un vector de parámetros, se consideró el caso no paramétrico. Aplicando los resultados de la Sección 3.2 se puede realizar el proceso inferencial para cada una de las probabilidades de transición. La matriz de probabilidades de cambio estimadas fueron obtenidas utilizando (3.9) y se presentan en (3.16)

$$\hat{P} = \begin{matrix} & \begin{matrix} N & NE & E & SE & S & SO & O & NO \end{matrix} \\ \begin{matrix} N \\ NE \\ E \\ SE \\ S \\ SO \\ O \\ NO \end{matrix} & \begin{bmatrix} 0.54 & 0.31 & 0.06 & 0.033 & 0.02 & 0.01 & 0.007 & 0.012 \\ 0.36 & 0.44 & 0.1 & 0.04 & 0.026 & 0.007 & 0.0062 & 0.009 \\ 0.25 & 0.28 & 0.27 & 0.083 & 0.06 & 0.03 & 0.01 & 0.005 \\ 0.18 & 0.11 & 0.1 & 0.31 & 0.18 & 0.08 & 0.035 & 0 \\ 0.12 & 0.075 & 0.031 & 0.22 & 0.32 & 0.16 & 0.07 & 0.006 \\ 0.16 & 0.057 & 0.023 & 0.034 & 0.25 & 0.27 & 0.12 & 0.08 \\ 0.37 & 0.1 & 0.06 & 0.02 & 0.1 & 0.12 & 0.14 & 0.06 \\ 0.36 & 0.2 & 0.033 & 0.23 & 0.066 & 0 & 0.033 & 0.066 \end{bmatrix} \end{matrix} \quad (3.16)$$

En esta matriz estimada se indican las probabilidades de cambio de una dirección a otra, como por ejemplo, la probabilidad de que el viento permanezca en el norte (N) dado que se encuentra en esa misma dirección es de 0.54. En cambio, estando en esta misma dirección (N) se observa que es poco probable pasar a otros estados. En general, es más probable permanecer en la misma dirección si el viento se encuentra en alguna dirección específica.

Otra manera más ilustrativa para representar estas probabilidades puede ser a través de los siguientes diagramas



Aquí se coloca en el centro la dirección actual del viento y en cada una de las ramas se indica la probabilidad de cambio a las distintas direcciones. La interpretación es la misma que en el caso de la matriz de transición puesto que se tiene la misma información pero mostrada de diferente forma.

Nótese que hasta ahora sólo se ha dado la estimación puntual de las probabilidades de cambio de una dirección a otra, pero esto no es todo el proceso que se realiza al momento de hacer la inferencia, pues ya mostró que también se proporcionan las gráficas de las funciones de verosimilitud e intervalos de verosimilitud a diferentes niveles. En particular se utilizarán los niveles de plausibilidad de 0.036, 0.147 y 0.25 pues cumplen, teóricamente, con cierta confianza asociada.

En la Figura 3.8 se muestran las funciones de verosimilitud relativa y el resumen de las inferencias para algunas de las probabilidades de transición.

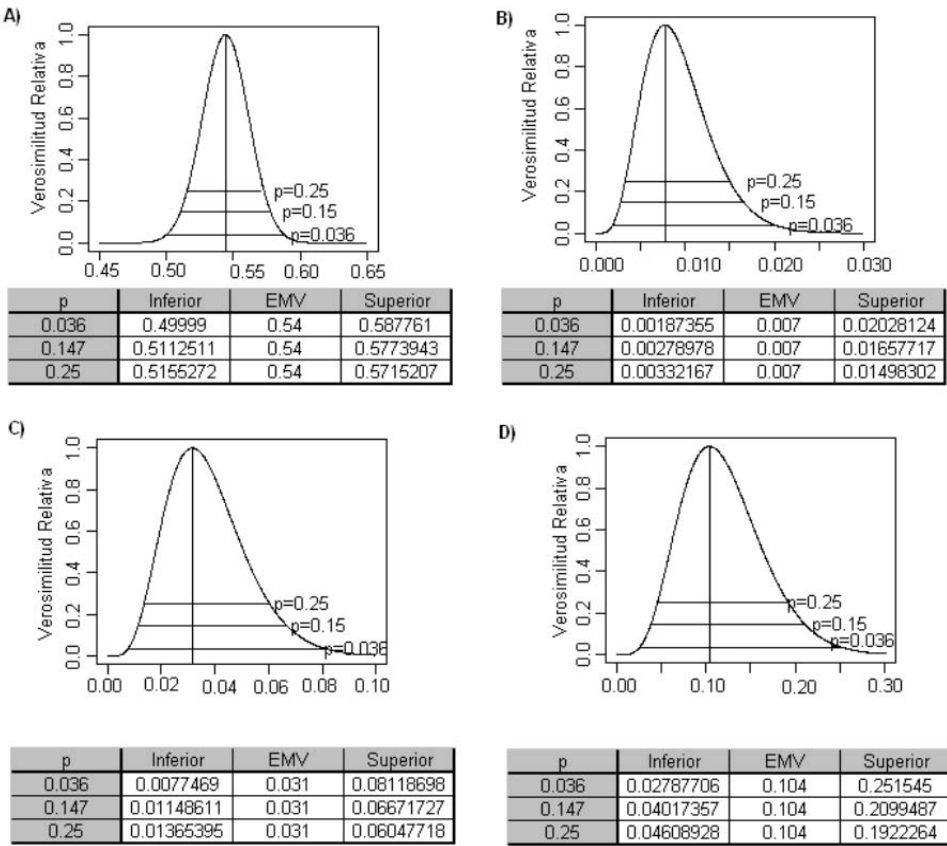


Fig. 3.8 Funciones de verosimilitud relativa.

En estas gráficas de las funciones de verosimilitud realtiva para  $P_{NN}$ ,  $P_{NOSO}$ ,  $P_{SE}$ ,  $P_{OS}$  se indican los  $EMV$  para cada una de ellas y los intervalos respectivos a lo niveles de verosimilitud  $p = 0.036$ , 0.147, 0.25 pues como ya se mencionó tienen asociada un 99%, 95% y 90% de confianza. En la Figura 3.8 A se puede observar que la probabilidad de que el viento permanezca en dirección norte (N) dado que se encontraba en esa misma dirección es de 0.54 y que valores para este parámetro,  $P_{NN}$ , menores a 0.49 y mayores a 0.58 son poco plausibles. En la Figura 3.8 B se puede comentar que la probabilidad de cambio del viento de dirección noreste (NE)

a dirección suroeste (SO) es muy pequeña y que valores para este parámetro poco plausibles son  $0.0018 \leq \widehat{P_{NESO}} \leq 0.020$ . De manera similar se pueden interpretar el resto de las gráficas.

Ahora se analizará la cadena descrita en (3.16) desde el punto de vista probabilístico. Se observa que es finita puesto que el espacio de estados son los puntos cardinales. Por otra parte, se tiene que todas sus entradas son positivas, excepto para  $P_{SENO} = 0$  y  $P_{NOSO} = 0$ . Pero  $P_{SENO}^2 = P_{SEN}P_{NNO} = 0.002 > 0$  y de manera análoga se tiene que  $P_{NOSO}^2 > 0$ . Con lo anterior se puede concluir que todos sus estados son comunicantes, por lo que en término probabilista se dice que la cadena es irreducible. En un contexto real esto quiere decir que el viento puede pasar de cualquier dirección a otra.

Otra de las propiedades importantes al momento de estudiar cadenas de Markov, desde el enfoque probabilista, es la recurrencia pues si algún estado es recurrente indica que con probabilidad uno la cadena regresará a ese mismo estado. Más interesante aún es conocer el tiempo medio que tardará en regresar a ese estado, si el tiempo medio de regreso es finito se tiene que ese estado será recurrente positivo y si cada estado de la cadena es recurrente positivo entonces la cadena de Markov es recurrente positiva. De la Proposición 1.20, la Proposición 1.21 y del Corolario 1.22, se tiene que la cadena de direcciones del viento es recurrente positiva.

Cuando una cadena de Markov es finita entonces existe al menos un estado recurrente positivo y si la cadena es irreducible entonces todos sus estados se comunican y se tiene que cada estado es recurrente positivo, lo que nos lleva a tener una cadena de Markov recurrente positiva.

Parte importante al estudiar cadenas de Markov es conocer en cuántos pasos un estado regresa a el mismo. En la matriz estimada (3.16) de las probabilidades de cambio de una dirección a otra, se tiene para la dirección norte (N) una probabilidad positiva de permanecer en esa dirección,  $P_{NN} > 0$ , de manera similar  $P_{NN}^2 > 0$  y en general  $P_{NN}^n > 0$ . Por tanto, el período para esta dirección del viento es uno y de la Proposición 1.14 se tiene que la cadena de Markov es aperiódica.

En resumen, las características que presenta esta cadena son: finita, irreducible, recurrente positiva y aperiódica. Para encontrar la distribución estacionaria se utiliza el sistema de ecuaciones definido en (1.20). Del Teorema 1.30 se tiene que la distribución límite existe y coincide con la distribución estacionaria, teniendo que

$$\pi = (0.38, 0.29, 0.094, 0.079, 0.072, 0.04, 0.022, 0.014).$$

En esta distribución se observa que las probabilidades más altas se encuentran en los estados  $N$  y  $NE$ . Esto parece congruente ya que los vientos predominantes en Guanajuato son los alisios.

## Capítulo 4

### Conclusiones

Lo que se puede concluir con este trabajo es que aún cuando la Probabilidad y la Estadística tienen diferentes objetivos e intereses ante una misma situación en el mundo real, aquí se logró ilustrar una conexión entre estas dos grandes ramas y bien establecidas de las matemáticas pues se conjuntaron las herramientas de ambas disciplinas en un mismo problema de aplicación real.

Además, se logró mostrar en general la metodología de verosimilitud para hacer inferencia estadística sobre algún parámetro de interés. Realizando los desarrollos correspondientes se logró hacer inferencia sobre un parámetro de un proceso estocástico específico. En particular, se estimaron las probabilidades de transición de una cadena de Markov paramétrica y no paramétrica mediante dicho enfoque.

Cabe mencionar que no sólo se dio la estimación puntual de la probabilidades de transición sino también se obtuvieron los intervalos de verosimilitud, que para ciertos niveles de verosimilitud, teóricamente, estos intervalos tienen asociados niveles de confianza. Estas propiedades estadísticas de la verosimilitud bajo este modelo de probabilidad (cadenas de Markov) se lograron verificar a través de simulaciones de dicho proceso.

Finalmente, se aplicó el enfoque de verosimilitud a datos reales de direcciones del viento modelados como una cadena de Markov no paramétrica en donde se estimaron las probabilidades de cambio de una dirección a otra. Además, se analizaron las propiedades probabilísticas que presenta esta cadena y se dio la distribución límite.



## Bibliografía

- [1] Basawa, I.V. y Prakasa Rao, B.L.S. (1980). *Statistical Inference for Stochastic Process*. Academic Press Inc.
- [2] Berger, J.O., Liseo, B. y Wolper, R.L. (1999). Integrated Likelihood Methods for Eliminating Nuisance Parameters. *Statistical Science*. V.14, pág. 1-28.
- [3] Caballero, M. E., Rivero, V. M., Uribe, G. y Velarde, C. (2004). *Cadenas de Markov: un enfoque elemental*. Sociedad Matemática Mexicana.
- [4] Casella, G. y Berger, R. (2002). *Statistical Inference*. segunda edición. Pacific Grove, Ca: Duxbury Press.
- [5] Cinlar, E. (1975). *Introduction to stochastic processes*. Prentice-Hall, Inc.
- [6] Cox, D. R. y Hinkley, D.V. (1974). *Theoretical Statistics*. London: Chapman and Hall.
- [7] Feller, W. (1950). *An Introduction to Probability Theory and Its Applications*. V. I. Princenton University.
- [8] Hoel, P. G., Port, S.C. y Stone, C. J. (1972). *Introduction to Stochastic Processes*. University of California, Los Angeles: Houghton Mifflin Company.
- [9] Johnson, R. (1995). *Just the Essentials of Elementary Statistics*. Duxbury Press.
- [10] Kalbfleisch, J. G. (1985). *Probability and Statistical Inference*. V. 2: Statistical Inference. segunda edición. Nueva York: Springer-Verlag.
- [11] Karlin, S. (1975). *A First Course in Stochastic Processes*. Academic, Press Inc.
- [12] Karlin, S. y Taylor, H. M. (1998). *An Introduction to Stochastic Modeling*. Academic, Press Inc.
- [13] Mendenhall, W. (1982). *Introducción a la Probabilidad y la Estadística*. Grupo Editorial Iberoamérica.
- [14] Montoya, J. A., Díaz-Francés, E. y Sprott, D. (2009). On a criticism of the profile likelihood function. *Statistical Papers*. V. 50, pág. 195-202.
- [15] Nakamura-Savoy, M. (2000). *Notas curso Elementos de Estadística y Probabilidad*. Guanajuato, Gto.
- [16] Norris, J.R. (1998). Markov Chains. *Cambridge Series in Statistical and Probabilistic Mathematics* (No. 2). University of Cambridge.

- [17] Ross, S. M. (1999). *Simulación*. Prentice Hall, Hispanoamericana.
- [18] Serfling, R. J. (1980). *Aproximation Theorems of Mathematical Statistics*. John Wiley & Sons.
- [19] Sprott, D. A. (2000). *Statistical Inference in Science*. Nueva York: Springer.