

# **Simulación e Inferencia de procesos de Markov con aplicaciones actuariales**

F. Baltazar-Larios

20 de febrero de 2023



# Índice general

<b>1. Conceptos básicos de Estadística</b>	<b>1</b>
1.1. Estimación máximo verosímil . . . . .	1
1.2. Propiedades asintóticas de los estimadores . . . . .	3
<b>2. Procesos estocásticos</b>	<b>5</b>
2.1. Definiciones elementales . . . . .	5
2.1.1. Clasificación general. . . . .	6
2.2. Cadenas de Markov a tiempo discreto . . . . .	6
2.2.1. Conceptos básicos . . . . .	6
2.2.2. Ejemplos . . . . .	7
2.2.3. Resultados importantes . . . . .	8
2.2.4. Clases de comunicación . . . . .	8
2.2.5. Periodo . . . . .	9
2.2.6. Propiedad Fuerte de Markov . . . . .	9
2.2.7. Tiempos de llegada y tiempos de absorción . . . . .	10
2.2.8. Clasificación de estados . . . . .	10
2.2.9. Distribuciones invariantes . . . . .	13
2.2.10. Cadenas regulares, absorbentes y convergentes. . . . .	14
2.2.11. Ejemplos . . . . .	15
2.3. Procesos de Saltos de Markov . . . . .	16
2.3.1. Definición . . . . .	16
2.3.2. Resultados elementales . . . . .	17

2.3.3. Generador infinitesimal . . . . .	20
2.3.4. Distribuciones estacionarias y distribuciones límite . . . . .	24
2.3.5. Ejemplos . . . . .	25
<b>3. Simulación de procesos de Markov</b>	<b>31</b>
3.1. Cadenas de Markov . . . . .	31
3.1.1. Algoritmo para cadenas de Markov . . . . .	31
3.1.2. Cálculo numérico de propiedades de cadenas de Markov . . . . .	33
3.2. Procesos de saltos de Markov . . . . .	34
3.2.1. Algoritmo para PSM . . . . .	34
3.2.2. Cálculo numérico de propiedades de PSM . . . . .	37
3.2.3. <b>Ejemplo. Modelo SIR-PSM</b> . . . . .	39
<b>4. Inferencia sobre procesos de Markov</b>	<b>41</b>
4.1. Cadenas de Markov . . . . .	41
4.1.1. Estimación máximo verosímil . . . . .	41
4.1.2. Propiedades asintóticas . . . . .	43
4.2. Procesos de saltos de Markov . . . . .	51
4.2.1. <b>Con información a tiempo continuo.</b> . . . . .	51
4.2.2. <b>Con información a tiempo discreto.</b> . . . . .	62
<b>5. Ecuaciones diferenciales estocásticas (EDE)</b>	<b>65</b>
5.1. Movimiento Browniano o proceso de Wiener . . . . .	65
5.1.1. Simulación de trayectorias del movimiento Browniano . . . . .	66
5.1.2. Aproximando al movimiento Browniano por una caminata aleatoria . . . . .	66
5.1.3. Puente Browniano . . . . .	68
5.2. Ecuaciones diferenciales estocásticas . . . . .	68
5.3. Procesos de difusión . . . . .	71
5.3.1. Proceso de difusión ergódicos . . . . .	72
5.3.2. Ecuaciones de Kolmogorov . . . . .	72
5.3.3. Puentes de difusión . . . . .	73

5.4. Fórmula de Itô . . . . .	73
5.4.1. Transformación de Lamperti . . . . .	77
5.4.2. Cociente de verosimilitud para procesos de difusión . . . . .	77
5.5. Ejercicios . . . . .	77
<b>6. Inferencia estadística para ecuaciones diferenciales estocásticas</b>	<b>79</b>
6.1. Simulación de procesos de difusión . . . . .	79
6.1.1. Método de Euler . . . . .	79
6.1.2. Método de Milstein . . . . .	81
6.1.3. Simulación de puentes de difusión . . . . .	82
6.1.4. Estimadores Monte-Carlo y técnicas de reducción de varianza . . . . .	82
6.2. Estimación por máxima verosimilitud . . . . .	82
6.2.1. Modelo Continuo . . . . .	83
6.3. Caso discreto . . . . .	84
6.3.1. Inferencia por verosimilitud exacta . . . . .	84
6.4. Ejercicios . . . . .	85
<b>A. Simulación de variables aleatorias</b>	<b>87</b>
<b>B. Convergencia de variables aleatorias y el Teorema central del límite.</b>	<b>89</b>
<b>C. Códigos en R</b>	<b>93</b>
<b>D. Códigos en Julia</b>	<b>95</b>
<b>Bibliografía</b>	<b>111</b>



# Capítulo 1

## Conceptos básicos de inferencia estadística por máxima verosimilitud

En este capítulo se estudian los conceptos básicos para realizar inferencia estadística mediante el método de máxima verosimilitud. La función de verosimilitud, el estimador de máxima verosimilitud y sus propiedades como: insesgamiento, consistencia y eficiencia asintótica son estudiados en este capítulo. Estos conceptos se utilizan en el Capítulo 4 para encontrar el estimador máximo verosímil de los procesos de Markov con espacio de estados finito y probar sus propiedades estadísticas. Para este capítulo se utilizó principalmente [? ].

### 1.1. Estimación máximo verosímil

Sea  $(\mathcal{X}, \mathcal{A}, \mathcal{Q})$  un modelo estadístico donde  $\mathcal{X}$  es el espacio muestral,  $\mathcal{A}$  una  $\sigma$ -álgebra de  $\mathcal{X}$  y  $\mathcal{Q} = \{\mathbb{Q}_\theta | \theta \in \Theta\}$  una clase parametrizada de medidas de probabilidad sobre  $(\mathcal{X}, \mathcal{A})$ . Se asume que el espacio parametral es un subconjunto del espacio Euclidiano  $d$ -dimensional  $\Theta \subseteq \mathbb{R}^n$ . Adicionalmente consideremos  $(\Omega, \mathcal{F})$  un espacio medible y  $X : \Omega \rightarrow \mathcal{X}$  una función medible y  $\mathcal{P} = \{\mathbb{P}_\theta | \theta \in \Theta\}$  medidas de probabilidad sobre  $(\Omega, \mathcal{F})$  tales que  $\mathbb{Q}_\theta$  es la medida de probabilidad inducida por  $X$ .

**Definición 1.1.** *Un estimador es una función  $\tilde{\theta} : \mathcal{X} \rightarrow \Theta$  y representa la forma de estimar el parámetro  $\theta$  de  $X$ , a partir de una muestra observada de  $X$ .*

**Definición 1.2.** *La función de verosimilitud  $L(\theta) = L(\theta; x)$  es la función de  $\theta$  que para cada  $x \in \mathcal{X}$  está dada por*

$$L(\theta) = L(\theta; x) = f_\theta(x), \quad \theta \in \Theta.$$

*La función de log-verosimilitud está dada por*

$$l(\theta) = l(\theta; x) = \log(L(\theta; x)).$$

*Donde  $f_\theta(x)$  es la función de densidad de  $x$  inducida por  $\mathbb{P}_\theta$ .*

Nuestro objetivo es encontrar un estimador de  $\theta$  a partir de una muestra observada, la función de verosimilitud es una herramienta muy importante para determinar buenos candidatos. Por ejemplo si tenemos que la razón de verosimilitud  $L(\theta_2)/L(\theta_1) = 10$  para  $\theta_1, \theta_2 \in \Theta$ , esto significa que bajo  $\mathbb{P}_{\theta_2}$  la observación  $x$  es 10 veces más probable que bajo la medida  $\mathbb{P}_{\theta_1}$ . En esta situación nosotros estamos más inclinados a

creer que  $\theta_2$  es mejor candidato que  $\theta_1$  para representar al parámetro desconocido  $\theta$ . El ejemplo antes mencionado nos da la motivación para dar la siguiente definición.

**Definición 1.3.** Si  $\hat{\theta} = \hat{\theta}(x)$  es tal que

$$l(\theta) \leq l(\hat{\theta}), \text{ para todo } \theta \in \Theta,$$

entonces  $\hat{\theta}(x)$  es llamado el estimador máximo verosímil.

Mientras que la ecuación

$$\frac{\partial l(\theta)}{\partial \theta} = 0$$

es llamada la ecuación de verosimilitud. Suponiendo claramente que la derivada de arriba exista  $\frac{\partial l(\theta)}{\partial \theta}$  denota el vector de derivadas parciales de  $l(\theta)$ .

Una observación que vale la pena mencionar es que cuando el estimador máximo verosímil  $\hat{\theta}(x)$  se encuentra en el interior de  $\Theta$ , éste resuelve la ecuación de verosimilitud.

Dado que ahora tenemos un posible candidato para estimar el verdadero parámetro  $\theta$ , una situación óptima sería cuando la distribución del estimador bajo la medida  $\mathbb{P}_\theta$  puede ser calculada. Esto permite hacer declaraciones de la cercanía con respecto al verdadero parámetro.

**Definición 1.4.** Se define el error cuadrático medio ( $ECM(\tilde{\theta})$ ) de un estimador  $\tilde{\theta}$  como

$$ECM(\tilde{\theta}) = \mathbb{E}_\theta[(\tilde{\theta}(X) - \theta)^2] = Var_\theta[\tilde{\theta}(X)] + \mathbb{E}_\theta[(\mathbb{E}[\tilde{\theta}(X)] - \theta)^2]$$

donde  $\mathbb{E}_\theta$  y  $Var_\theta$  son la esperanza y varianza respecto a la medida de probabilidad  $\mathbb{P}_\theta$  respectivamente.

El error cuadrático medio es una medida de la cercanía que tiene el estimador con respecto al verdadero parámetro.

**Definición 1.5.** Un estimador  $\tilde{\theta} : \mathcal{X} \rightarrow \Theta$  se dice insesgado si

$$\mathbb{E}_\theta[\tilde{\theta}(X)] = \theta, \text{ para cada } \theta \in \Theta.$$

Por otro lado si un estimador no es insesgado, diremos que es sesgado.

Los estimadores insesgados son preferidos respecto a los sesgados, pues su error cuadrático medio es menor.

**Ejemplo 1.1.** Sea  $X_1, \dots, X_n$  variables aleatorias i.i.d. (independientes e idénticamente distribuidas) con densidad

$$\beta \exp(-\beta x), \quad x > 0,$$

Donde  $\beta > 0$  es el parámetro. La función de log-verosimilitud es

$$l(\beta) = n(\log(\beta) - \beta \bar{x}),$$

aquí el estimador máximo verosímil es  $\hat{\beta} = \bar{X}^{-1}$ .

Ya que  $\bar{X}$  tiene una distribución gamma, la media del estimador es

$$\mathbb{E}_\beta[\hat{\beta}] = \frac{n}{n-1} \beta \neq \beta,$$

De este modo  $\hat{\beta}$  no es un estimador insesgado. Si en lugar del parámetro  $\beta$  nosotros consideramos el parámetro  $1/\beta$ , encontramos que el estimador máximo verosímil del nuevo parámetro es

$$1/\hat{\beta} = \bar{X}.$$

Claramente  $\mathbb{E}_{1/\beta}[1/\hat{\beta}] = 1/\beta$ , entonces  $\bar{X}$  es un estimador insesgado respecto al nuevo parámetro.

El ejemplo anterior muestra que el estimador máximo verosímil no es en general insesgado. También este ejemplo muestra que el insesgamiento se relaciona tanto con el estimador y la parametrización en cuestión, por tanto algunas parametrizaciones son más útiles que otras.

**Definición 1.6.** *La variable aleatoria*

$$U(\theta) = \frac{\partial l(\theta, X)}{\partial \theta},$$

*es llamada la función score. La matriz*

$$j(\theta) = -\frac{\partial^2 l(\theta, X)}{\partial \theta \partial \theta^T},$$

*es llamada la información observada, mientras que la esperanza de  $j(\theta)$*

$$i(\theta) = \mathbb{E}_\theta[j(\theta)],$$

*es llamada la información de Fisher o información esperada, esta matriz es importante porque bajo ciertas condiciones los estimadores cumplen*

$$(\tilde{\theta}(X) - \theta)i(\theta)^{1/2} \xrightarrow{d} N(\bar{0}, I_n)$$

*donde  $I_n$  denota la matriz identidad de dimensiones  $n \times n$ .*

## 1.2. Propiedades asintóticas de los estimadores

Intuitivamente mientras más grande sea la muestra, es posible tener una noción más precisa del valor del verdadero parámetro. Por esta razón es de vital importancia estudiar propiedades de los estimadores cuando el tamaño de la muestra crece.

Una propiedad deseable es que la esperanza del estimador converja al verdadero parámetro cuando el tamaño de la muestra tienda a infinito.

**Definición 1.7.** *Un estimador  $\tilde{\theta} : \mathcal{X} \rightarrow \Theta$  se dice asintóticamente insesgado si*

$$\mathbb{E}_\theta[\tilde{\theta}(X)] \rightarrow \theta, \text{ para cada } \theta \in \Theta.$$

La propiedad anterior es más débil que la propiedad de insesgamiento, esto nos permite estudiar una gran cantidad de estimadores, ya que algunos estimadores pueden ser asintóticamente insesgados pero no son insesgados (ver Ejemplo 1.1).

La siguiente propiedad es muy importante para el estudio de propiedades asintóticas de los estimadores.

**Definición 1.8.** *Un estimador  $\tilde{\theta} : \mathcal{X} \rightarrow \Theta$  se dice consistente si para cada  $\theta \in \Theta$*

$$\tilde{\theta}(X) \xrightarrow{c.s.} \theta.$$

Hasta ahora sólo hemos analizado propiedades de la media de los estimadores la cual nos permite medir el promedio del valor del estimador, sin embargo debemos analizar la varianza de este para poder determinar qué tan alejados en promedio nos encontramos del verdadero parámetro.

**Definición 1.9.** *Decimos que un estimador  $\tilde{\theta} : \mathcal{X} \rightarrow \Theta$  es asintóticamente eficiente si*

$$(\tilde{\theta}(X) - \theta)i(\theta)^{1/2} \xrightarrow{d} N(\hat{0}, I_n).$$

*Donde  $I_n$  denota la matriz identidad de dimensiones  $n \times n$*

Para una matriz semi-definida positiva  $B$ , nosotros denotamos por  $B^{1/2}$  su correspondiente raíz cuadrada. La propiedad anterior muestra que nosotros podemos aproximar la distribución del máximo verosímil mediante una normal multivariada con media  $\theta$  y matriz de covarianzas  $i(\theta)^{-1}$  cuando el tamaño de la muestra es grande. Un ejemplo de estimador que cumpla ser asintóticamente eficiente es el estimador de  $1/\beta$  del Ejemplo 1.1. La importancia de esta propiedad radica en que si la muestra es grande podemos calcular la probabilidad de que nuestro verdadero parámetro  $\theta$  se encuentre en cierta región de  $\mathbb{R}^d$ .

## Capítulo 2

# Procesos Estocásticos con espacio de estados a lo más numerable

En este capítulo se presenta una introducción breve al tema de los procesos estocásticos con espacio de estados a lo más numerable. Se explican algunos conceptos y propiedades generales de este tipo de procesos y se estudian brevemente algunos ejemplos particulares de procesos estocásticos. El contenido de este capítulo será usado en los capítulos tres y cuatro, en donde consideraremos procesos estocásticos con espacio de estados finito. La bibliografía básica para este capítulo es principalmente [? ], [? ] y [? ].

### 2.1. Definiciones elementales

En esta sección se presentan los resultados fundamentales de procesos estocásticos. Se describen sus características principales, se realiza una clasificación básica y se dan algunos ejemplos.

Comenzaremos dando una definición formal de un proceso estocástico.

**Definición 2.1.** *Un proceso estocástico  $X = \{X_t\}_{t \in \mathcal{T}}$  es una colección de variables aleatorias (definidas en el mismo espacio de probabilidad) que toman valores en un conjunto  $E$  con parámetro  $\mathcal{T}$ , es decir, para cada  $t \in \mathcal{T}$ ,  $X_t$  es una variable aleatoria.*

En esta definición a  $\mathcal{T}$  se le conoce como el espacio parametral y es interpretado como el tiempo, de esta forma a  $X_t$  se le entenderá como el estado del proceso al tiempo  $t$ . Además se considera que las variables aleatorias  $X_t$  para toda  $t \in \mathcal{T}$ , están definidas sobre el mismo espacio de probabilidad  $(\Omega, \mathcal{F}, \mathbb{P})$ . Un proceso estocástico puede ser interpretado como la evolución en el tiempo de algún fenómeno cuya dinámica se rige por el azar.

Siguiendo la definición de proceso estocástico, se le puede ver como una función de dos variables

$$X : \mathcal{T} \times \Omega \rightarrow E,$$

tal que a cada pareja  $(t, \omega) \in (\mathcal{T}, \Omega)$  se le asocia  $X(t, \omega) = X_t(\omega)$  y de esta manera para cada  $t \in \mathcal{T}$  la función de

$$\omega \rightarrow X_t(\omega)$$

es una variable aleatoria y para cada  $\omega \in \Omega$ , la función

$$t \rightarrow X_t(\omega)$$

es una trayectoria o realización del proceso.

### 2.1.1. Clasificación general.

Según las características del espacio parametral se puede hacer la primera distinción entre procesos estocásticos.

**Definición 2.2. Proceso estocástico a tiempo discreto** Si  $\mathcal{T}$  es un conjunto a lo más numerable, se dice que  $X$  es un proceso estocástico a tiempo discreto, para nosotros será  $\mathcal{T} = M$ , donde  $M \subset \mathbb{N}$  y en general se denota por  $X = \{X_n\}_{n \geq 0}$ .

**Definición 2.3. Proceso estocástico a tiempo continuo** Cuando  $\mathcal{T}$  sea un conjunto infinito no numerable, se dice que  $X$  es un proceso estocástico a tiempo continuo, para este caso  $\mathcal{T} = [0, s]$  ó  $[0, \infty)$ . y en será denotado por  $X = \{X_t\}_{t \geq 0}$ .

En cuanto al conjunto donde las variables aleatorias toman valores, se le conoce como espacio de estados del proceso y de manera análoga este también puede ser discreto o continuo.

## 2.2. Cadenas de Markov a tiempo discreto

Para esta sección de cadenas de Markov se utilizaron los libros [?] y [?], en donde se pueden consultar las demostraciones de los resultados aquí enunciados.

### 2.2.1. Conceptos básicos

**Definición 2.4.** Consideremos una colección de variables aleatorias  $X_0, X_1, \dots$ , interpretando a  $X_n$  como el estado de un sistema al tiempo  $n$  y supongamos que el conjunto de posibles valores de  $X_n$  es a lo más numerable, enumerándolos, tenemos que  $E = \{1, 2, \dots, N\}$  para el caso finito y  $E = \{1, 2, \dots\}$  para el caso numerable,  $E$  es llamado espacio de estados. Si este proceso satisface la **propiedad de Markov**

$$\mathbb{P}[X_{n+1} = x_{n+1} | X_n = x_n, \dots, X_0 = x_0] = \mathbb{P}[X_{n+1} = x_{n+1} | X_n = x_n]$$

con  $x_0, \dots, x_{n+1} \in E$ , decimos que  $\{X_n\}_{n \geq 0}$  es **una cadena de Markov**.

A la distribución de  $X_0$  se le llama **distribución inicial** y la denotaremos por  $\pi = (\pi_1, \dots, \pi_N)$ , es decir, si  $\pi_i = \mathbb{P}[X_0 = i]$  para toda  $i \in E$ .

Algunas observaciones importantes:

1. Si  $X_n = x_n, \dots, X_0 = x_0$  entonces decimos que se ha presentado la trayectoria  $x_0, \dots, x_n$ .

2. A la familia  $\mathbb{P}[X_{n+1} = j | X_n = i]$  se le conoce como familia de probabilidades de transición de la cadena de Markov y describe la evolución de la cadena en el tiempo.

**Definición 2.5.** Se dice que la cadena de Markov  $\{X_n\}_{n \geq 0}$  con espacio de estados  $E$  y distribución inicial  $\pi$  es homogénea si

$$\mathbb{P}[X_{n+1} = j | X_n = i] = \mathbb{P}[X_1 = j | X_0 = i]$$

para cada  $i, j \in E$  y para cada  $n \geq 0$ . En este caso definimos para cada  $i, j \in E$

$$p_{ij} = \mathbb{P}[X_1 = j | X_0 = i]$$

y para  $m \geq 1$

$$p_{ij}^{(m)} = \mathbb{P}[X_{m+n} = j | X_n = i],$$

que representa la probabilidad de pasar del estado  $i$  al estado  $j$  en  $m$  unidades de tiempo.

**Definición 2.6.** Matriz de transición:

Es la matriz cuadrada de orden  $N$  formada por las probabilidades de transición

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ p_{N1} & p_{N2} & \cdots & p_{NN} \end{pmatrix}$$

con entradas no negativas y la suma de cada renglón es uno (matriz estocástica).

### 2.2.2. Ejemplos

Veamos algunos ejemplos de cadenas de Markov

**Ejemplo 2.1.** Cadena de rachas.

Considere la observación de un experimento con probabilidad  $p$  de ocurrencia, definiendo como transición un ensayo y a  $X_n$  como los estados de estar en racha de éxitos o no, entonces  $X = \{X_n\}_{n \geq 0}$  define una cadena de Markov.

**Ejemplo 2.2.** Caminata aleatoria simple.

Un individuo está parado e inicia una caminata en el cual sólo se mueve hacia adelante con probabilidad  $p$  o hacia atrás con probabilidad  $1 - p$ . Definiendo a  $X_n$  como la posición del individuo después de  $n$ -pasos, se tiene entonces que  $E = \mathbb{Z}$  y las probabilidades de transición para cualquier  $n \geq 0$  e  $i, j \in E$  están dadas por:

$$p_{ij} = \begin{cases} p & j = i + 1 \\ 1 - p & j = i - 1 \\ 0 & e.o.c. \end{cases}$$

**Ejemplo 2.3.** *Problema del jugador.*

Suponga que un jugador  $A$  apuesta \$1 sucesivamente contra otro jugador  $B$ .  $A$  inicia con  $k$  unidades y  $B$  con  $N - k$ . En cada apuesta,  $A$  gana con probabilidad  $p$  y pierde con probabilidad  $1 - p$ . Si definimos  $X_n$  como la fortuna de  $A$  al tiempo  $n$ , entonces  $X = \{X_n\}_{n \geq 0}$  define una cadena de Markov.

**Ejemplo 2.4.** *Cadena de Ehrenfest.*

Se tienen dos urnas, entre ambas hay un total de  $k$  bolas. En cada paso se elige una de las bolas al azar y se cambia de urna. Si  $X_n =$  número de bolas en la urna  $A$  después de  $n$  ensayos, entonces  $X = \{X_n\}_{n \geq 0}$  define una cadena de Markov.

**2.2.3. Resultados importantes**

A continuación presentaremos algunos resultados importantes.

**Proposición 2.1.** *Ecuación de Chapman-Kolmogorov.*

Sea  $X = \{X_n\}_{n \geq 0}$  una cadena de Markov con matriz de transición  $P = \{p_{ij}\}_{ij \in E}$  entonces

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

Este resultado nos dice que  $p_{ij}^{(n)}$  es la entrada  $ij$  de la  $n$ -ésima potencia de  $P$ . En general, no es posible calcular explícitamente las potencias de la matriz de transición. Sin embargo, nos podemos auxiliar de paquetes computacionales para esta tarea.

**Proposición 2.2.** *Propiedad de corrimiento.*

Sea  $X = \{X_n\}_{n \geq 0}$  una cadena de Markov con espacio de estados  $E$  y con matriz de transición  $P$ . Sean  $n, k \in \mathbb{N}$  fijos y  $x_0, x_1, \dots, x_n, \dots, x_{n+k} \in E$  entonces

$$\mathbb{P}[X_{n+k} = x_{n+k}, \dots, X_{n+1} = x_{n+1} | X_n = x_n, \dots, X_0 = x_0] = \mathbb{P}[X_k = x_{n+k}, \dots, X_1 = x_{n+1} | X_0 = x_n].$$

Este resultado es una generalización de la propiedad de Markov.

**2.2.4. Clases de comunicación**

**Definición 2.7.** Sea  $X = \{X_n\}_{n \geq 0}$  una cadena de Markov con espacio de estados  $E$  y con matriz de transición  $P$ . Sean  $i$  e  $j$  dos estados de  $E$ , se dice que el estado  $j$  es accesible desde el estado  $i$  si existe  $n \in \mathbb{N}$ , tal que  $p_{ij}^{(n)} > 0$  ( $i \rightarrow j$ ). Si  $i \rightarrow j$  e  $j \rightarrow i$  entonces,  $i$  y  $j$  se dice que se **comunican**, y cuando esto ocurre escribimos  $i \leftrightarrow j$ .

El siguiente resultado nos muestra que la relación  $\leftrightarrow$  es de equivalencia.

**Proposición 2.3.** La relación  $x \leftrightarrow y$  (comunicación) es una relación de equivalencia, por lo tanto, induce a una partición en el espacio de estados.

A las clases de equivalencia inducidas por esta relación les llamaremos clases de comunicación.

**Definición 2.8.** Sea  $X = \{X_n\}_{n \geq 0}$  una cadena de Markov con espacio de estados  $E$  y con matriz de transición  $P$ , decimos que la cadena es irreducible si para cada  $i, j \in E$  tenemos que  $i \leftrightarrow j$ .

Sea  $i \in E$ , a su clase de comunicación la denotaremos por  $C_i$ , es decir

$$C_i = \{j \in E : i \leftrightarrow j\}.$$

En cuanto a la partición inducida por la comunicación de estados decimos que, tomando  $C \subset E$ , es un conjunto cerrado si para toda  $j \in E - C$ ,  $i \in C$  no puede acceder a  $j$ , en caso contrario decimos que es una clase abierta.

### 2.2.5. Periodo

**Definición 2.9.** El periodo de un estado  $i \in E$  es un número entero no negativo denotado por  $d(i)$ , y definido como sigue:

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

en donde m.c.d significa máximo común divisor. Cuando  $p_{ii}^{(n)} = 0$  para toda  $n \geq 1$ , se define  $d(i) = 0$ . En particular, se dice que un estado  $i \in E$  es aperiódico si  $d(i) = 1$ . Cuando  $d(i) = k \geq 2$  se dice que  $i$  es periódico de periodo  $k$ .

**Proposición 2.4.** Si los estados  $i, j \in E$  pertenecen a la misma clase de comunicación, entonces tienen el mismo periodo.

### 2.2.6. Propiedad Fuerte de Markov

Una extensión de la propiedad de Markov es la presentada a tiempos aleatorios conocida como propiedad fuerte de Markov. Comenzaremos esta sección dando la definición de tiempos aleatorios.

**Definición 2.10.** Un **tiempo aleatorio** es una variable aleatoria

$$\tau : \Omega \rightarrow \mathbb{N} \cup \{\infty\}.$$

**Definición 2.11.** Sea  $X = \{X_n\}_{n \geq 0}$  una cadena de Markov con espacio de estados  $E$  y con matriz de transición  $P$ . Un tiempo aleatorio  $\tau$  es un **tiempo de paro** respecto a la cadena  $X = \{X_n\}_{n \geq 0}$  si para  $n \in \mathbb{N}$  existe  $A_n \subset E^{n+1}$  tal que

$$\{\tau = n\} = \{(X_0, \dots, X_n) \in A_n\}.$$

A la variable aleatoria tiempo de paro la podemos interpretar como el tiempo que se obtiene de observar una trayectoria hasta que se cumpla cierta condición.

**Teorema 2.1. Propiedad Fuerte de Markov.** Sea  $\{X_n\}_{n \geq 0}$  una cadena de Markov con matriz de transición  $P$  y  $\tau$  un tiempo de paro. Entonces condicionado sobre  $\{\tau < \infty\} \cap \{X_\tau = i\}$  el proceso  $\{X_{\tau+m}\}_{m \geq 0}$  es una cadena de Markov que comienza en  $i$  con matriz de transición  $P$  y es independiente de las variables  $X_0, X_1, \dots, X_\tau$ .

### 2.2.7. Tiempos de llegada y tiempos de absorción

En el estudio de las cadenas de Markov resulta de particular interés el estudio de los tiempos aleatorios de la ocurrencia de eventos relacionados con la llegada a un estado o un conjunto de estados del espacio de la cadena. En esta sección estudiaremos este tipo de variables aleatorias.

**Definición 2.12.** Sea  $\tau_m^A$  la  $m$ -ésima vez que la cadena accede a  $A$  un subconjunto del espacio de estados, es decir

$$\tau_m^A = \begin{cases} \infty & X_n \in E - A \text{ para toda } n \\ \min\{n > \tau_{m-1}^A : X_n \in A\} & \text{e.o.c} \end{cases}$$

donde se define  $\tau_0^A = 0$

**Nota 2.1.** Bajo la hipótesis de que  $\tau_m^A < \infty$  para cada  $m$  y por el Teorema 2.1 tenemos que  $\{\tau_m^A\}_{m \geq 1}$  son i.i.d.

Veamos un ejemplo donde podemos ilustrar esta definición.

**Ejemplo 2.5. Racha de éxitos.** Sea  $E_1, E_2, \dots$ , una sucesión de ensayos Bernoulli con probabilidad de éxito  $p$  y probabilidad de fracaso  $q = 1 - p$ . Si  $X_n$  es el número de éxitos consecutivos previos al tiempo  $n$  (incluido el tiempo  $n$ ). Se dice que una cadena de  $r$  éxitos ocurre al tiempo  $n$  si en el ensayo  $n - r$  ocurre un fracaso y los resultados de los ensayos  $n - r + 1$  al  $n$  son éxitos.

**Definición 2.13.** Para cada  $n \geq 1$ , denotamos por  $f_{ij}^{(n)}$  a la probabilidad de que una cadena que inicia en el estado  $i$ , llegue al estado  $j$  por primera vez en exactamente  $n$  pasos, es decir,

$$f_{ij}^{(n)} = \mathbb{P}[X_n = j, X_{n-1} \neq j, \dots, X_1 \neq j, X_0 = i].$$

Además definimos

$$f_{ij}^{(0)} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{e.o.c} \end{cases}$$

Para la cadena de Markov de racha de éxitos tenemos (por ejemplo)

1.  $f_{01}^{(n)} = q^{n-1}p$ .
2.  $f_{00}^{(n)} = p^{n-1}q$ .

Un resultado importante es el que permite descomponer la probabilidad de visitar el estado  $j$  iniciando en  $i$  en términos de todas las posibilidades de tiempos que pueden ser la primera visita.

**Proposición 2.5.** Para cada  $n \geq 1$ ,

$$p_{ij}^{(n)} = \sum_{k=1}^n f_{ij}^{(k)} p_{ij}^{(n-k)}.$$

### 2.2.8. Clasificación de estados

De acuerdo a las propiedades que tiene cada estado en una cadena de Markov en cuanto a la probabilidad de regresar a él, los estados pueden ser clasificados de la siguiente manera.

**Definición 2.14.** Sea  $X = \{X_n\}_{n \geq 0}$  una cadena de Markov con espacio de estados  $E$  y con matriz de transición  $P$ . Se dice que un estado  $i$  es **recurrente** si la probabilidad de (eventualmente) regresar a  $i$ , es uno, es decir, si

$$\mathbb{P}[\tau_1^i < \infty | X_0 = i] = 1.$$

Un estado que no es recurrente se llama **transitorio**.

Tratar de clasificar los estado en recurrentes y transitorios partiendo de la definición no resulta, en general, tarea fácil. Una manera distinta de enunciar estos conceptos es la siguiente forma.

**Definición 2.15.** Definimos a

$$f_i = \sum_{n=1}^{\infty} f_{ii}^{(n)},$$

que es la probabilidad de un eventual regreso al estado  $i$ .

**Nota 2.2.** Decimos que:

- El estado  $i$  es **recurrente** si  $f_i = 1$ , en particular si  $f_{xx}^{(1)} = 1$  el estado se le llama **absorbente**.
- El estado  $x$  es **transitorio** si  $f_x < 1$ .

**Definición 2.16.** Definiendo al número de visitas al estado  $j$  hasta el tiempo  $n$  como la variable aleatoria

$$V_{ij}(n) = \sum_{k=1}^n \mathbb{I}_{\{X_k=j\}}, \text{ cuando } X_0 = i$$

**Proposición 2.6.**

1. Un estado  $i$  es recurrente si y sólo si  $\sum_n p_{ii}^{(n)} = \infty$ .
2. Un estado  $i$  es transitorio si y sólo si  $\sum_n p_{ii}^{(n)} < \infty$ .

Este resultado nos proporciona una equivalencia, en términos de las matrices de transición, para que un estado sea recurrente.

A continuación presentamos una serie de resultados respecto a la clasificación de estados y la comunicación.

**Proposición 2.7.** Sean  $i, j \in E$  tales que se comunican, si  $i$  es transitorio entonces  $j$  también es transitorio.

Este resultado tiene como consecuencia que la transitoriedad o recurrencia son propiedades de clase.

Un criterio fácil para saber si una clase es transitoria es el siguiente.

**Proposición 2.8.** Sea  $C \subset E$  una clase abierta. Entonces  $C$  es transitoria.

Usando estos resultados tenemos la siguiente conclusión.

**Proposición 2.9.** *Si el espacio de estados es finito, una clase es recurrente si y sólo si es cerrada.*

**Corolario 2.1.** *En una cadena de Markov irreducible con espacio de estados finito todos sus estados son recurrentes.*

**Corolario 2.2.** *Toda cadena de Markov con espacio de estados finito tiene por lo menos un estado recurrentes.*

**Nota 2.3.** *En general, el análisis de recurrencia y transitoriedad, en cadenas con espacio de estados infinito, es más complicado.*

Siguiendo la clasificación de estados, resulta importante conocer el número esperado de pasos para regresar a  $i$ , es decir,

$$\mu_i = \sum_{n=1}^{\infty} n f_{ii}^{(n)},$$

a la cual llamaremos **tiempo medio de recurrencia**.

**Nota 2.4.** *Puede darse el caso de que siendo  $i$  un estado recurrente el tiempo medio de recurrencia,  $\mu_i$ , sea infinito; siendo éste el caso en el que aunque el regreso a  $i$  es cierto se necesita un tiempo infinito. Así, realizamos la siguiente distinción entre los estados recurrentes obteniendo una subclasificación de estos.*

**Definición 2.17.** *Decimos que*

1. *El estado  $i$  es **recurrente positivo** si  $\mu_i < \infty$ .*
2. *El estado  $i$  es **recurrente nulo** si  $\mu_i = \infty$ .*

Se tiene directamente el siguiente resultado.

**Proposición 2.10.** *En una cadena irreducible con espacio de estados finito, todos los estados son recurrentes positivos.*

**Ejemplo 2.6.** *Veamos que la cadena de Markov de racha de éxitos es una cadena recurrente positiva.*

$$\mu_0 = \sum_{n=1}^{\infty} n f_{00}^{(n)} = \sum_{n=1}^{\infty} n(1-p)p^{n-1} = (1-p) \sum_{n=1}^{\infty} np^{n-1} = \frac{1}{1-p} < \infty.$$

Otro resultado de gran utilidad es el siguiente.

**Proposición 2.11.** *No existen estados recurrentes nulos en cadenas de Markov finitas.*

El siguiente resultado se sigue de la teoría que los promedios temporales son iguales a los promedios espaciales.

**Teorema 2.2. Teorema ergódico para cadenas de Markov** *Sea  $X = \{X_n\}_{n \geq 0}$  una cadena de Markov con espacio de estados  $E$  y con matriz de transición  $P$ . Sean  $i, j$  cualesquiera estados de una cadena de Markov irreducible se cumple que*

$$\lim_{n \rightarrow \infty} \frac{V_{ij}(n)}{n} = \frac{1}{\mu_j}$$

*casi seguramente.*

**Ejemplo 2.7.** Considere una cadena de Markov con espacio de estados  $\{0, \dots, 5\}$  y matriz de transición

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{3} & 0 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{8} & 0 & \frac{7}{8} & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & \frac{3}{4} & 0 & \frac{1}{4} & 0 \\ 0 & \frac{1}{5} & 0 & \frac{1}{5} & \frac{1}{5} & \frac{2}{5} \end{pmatrix}$$

Determine que estados son transitorios y cuales recurrentes. Calcular y estimar tiempos medios de recurrencia y número esperado de visitas a cada estado.

### 2.2.9. Distribuciones invariantes

**Definición 2.18.** Sea  $\nu \in R^{|E|}$ , decimos que  $\nu$  es una **distribución invariante o estacionaria** de la cadena de Markov con matriz de transición  $P$ , si satisface

1. Tiene todas sus entradas no negativas.
2.  $\sum_{i \in E} \nu_i = 1$ .
3.  $\sum_{j \in E} \nu_i p_{ij} = \nu_j$ .

En términos matriciales, la distribución de probabilidad  $\nu$  es invariante si

$$\nu P = \nu.$$

**Nota 2.5.** La interpretación de una distribución invariante es que una variable aleatoria con dicha distribución mantendrá su distribución en la evolución de la cadena de Markov en el tiempo.

Desde que encontrar una distribución estacionaria en una cadena de Markov consiste en resolver un sistema de  $|E| + 1$  ecuaciones y  $|E|$  incógnitas podemos no tener solución, una infinidad de soluciones o una única solución.

Los siguientes resultados son de gran utilidad en el cálculo de la distribución estacionaria de una cadena de Markov.

**Proposición 2.12.** Sea  $\nu$  una distribución estacionaria para una cadena de Markov. Si  $i$  es un estado transitorio o recurrente nulo, entonces  $\nu_i = 0$ .

Una propiedad importante de una distribución invariante es que si  $X_0$  tiene distribución  $\nu$ , entonces  $X_n$  tendrá la misma distribución para toda  $n$ .

**Proposición 2.13.** Toda cadena de Markov que es irreducible y recurrente positiva tiene una única distribución estacionaria dada por

$$\nu_i = \frac{1}{\mu_i} > 0.$$

**Teorema 2.3.** Para una cadena irreducible las siguientes condiciones son equivalentes.

1. Todos los estados son recurrentes positivos.
2. Algún estado es recurrente positivo.
3. La cadena tiene una única distribución invariante

**Ejemplo 2.8.**

$$P = \begin{pmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix}$$

$$\nu = (1/3, 2/3).$$

**Ejemplo 2.9. Cadena de Ehrenfest.** Como esta cadena de Markov es irreducible y con espacio de estados finito se tiene que es recurrente positiva y por los resultados anteriores tiene una única distribución invariante  $\nu$ .

### 2.2.10. Cadenas regulares, absorbentes y convergentes.

En esta sección presentaremos resultados del comportamiento de una cadena de Markov recurrente cuando la trayectoria es lo suficientemente grande. Otro concepto importante en el estudio de las cadenas de Markov es el relacionado con la posible periodicidad con la que se visita a cada estado.

**Definición 2.19.** Decimos que una cadena de Markov  $\{X_n\}$  con espacio de estados finito es **regular** si existe una potencia de la matriz de transición con todas las entradas positivas.

Ahora presentamos resultados fundamentales en el estudio de la convergencia de las cadenas de Markov.

**Definición 2.20.** Si los límites

$$\nu_j = \lim_{n \rightarrow \infty} p_{ij}^{(n)}$$

existen para cada estado  $j$  en el espacio de estados entonces  $\pi$ , es la distribución límite.

**Nota 2.6.** Las siguientes observaciones son consecuencia de la definición anterior.

1. La distribución límite es inducida por la matriz de transición  $P$ .
2. La distribución límite no depende de la distribución inicial.
3. El límite de las potencias de  $P$  es una matriz estocástica para la cual todos sus renglones son la distribución límite.
4. La distribución límite es una distribución estacionaria.

**Teorema 2.4.** (Existencia de la distribución límite.) Si la cadena de Markov es irreducible, aperiódica y con distribución invariante  $\nu$ , entonces

$$\nu_j = \lim_{n \rightarrow \infty} p_{ij}^{(n)}$$

**Teorema 2.5.** Si la cadena de Markov es irreducible, aperiódica y recurrente positiva existen las probabilidades límite y son la única distribución invariante.

## 2.2.11. Ejemplos

**Ejemplo 2.10.** Consideremos la cadena de Markov del Ejemplo 2.4, el espacio de estados es  $E = \{0, 1, 2, \dots, k\}$  y la matriz de transición asociada es

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 1/k & 0 & (k-1)/k & 0 & \dots & 0 & 0 \\ 0 & 2/k & 0 & (k-2)/k & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 & 1/k \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix}.$$

Es fácil mostrar que esta cadena es irreducible, por lo tanto existe una única clase de comunicación

$$\{0, 1, 2, \dots, k\}.$$

Por la Proposición 2.4 sabemos que el periodo para todos los estados es el mismo, además sabemos

$$\{n \geq 1 : p_{00}^{(n)} > 0\} = \{n \geq 1 : n \text{ es par}\},$$

es decir, tiene periodo 2.

Debido a que esta cadena es irreducible, por la Proposición 2.10 es recurrente positiva. Por lo tanto por la Proposición 2.13 tiene una única distribución estacionaria.

Desarrollando el sistema de ecuaciones  $\nu P = \nu$ , tenemos

$$\begin{aligned} \nu_0 &= \frac{1}{k}\nu_1 \\ \nu_1 &= \nu_0 + \frac{2}{k}\nu_2 \\ \nu_2 &= \frac{k-1}{k}\nu_1 + \frac{3}{k}\nu_3 \\ &\vdots \\ \nu_{k-1} &= \frac{2}{k}\nu_{k-2} + \nu_k \\ \nu_k &= \frac{1}{k}\nu_{k-1}. \end{aligned}$$

Reescribiendo cada ecuación en términos de  $\nu_0$  obtenemos

$$\nu_j = \binom{k}{j} \nu_0, \text{ para todo } j \in E.$$

Sumando todas estas ecuaciones

$$1 = \nu_0 \sum_{j=0}^k \binom{k}{j} = \nu_0 2^k,$$

por lo que concluimos que

$$\nu_j = \binom{k}{j} \frac{1}{2^k}, \text{ para todo } j \in E.$$

Por la Proposición 2.13 podemos calcular los tiempos medios de recurrencia mediante la distribución estacionaria

$$\mu_j = \frac{1}{\nu_j} = \frac{2^k}{\binom{k}{j}}, \text{ para todo } j \in E.$$

## 2.3. Procesos de Saltos de Markov

En esta sección consideraremos procesos de Markov a tiempo continuo, homogéneos y con espacio de estados numerable, describiremos características relevantes de dichos procesos que nos serán de utilidad para simular trayectorias de ellos.

### 2.3.1. Definición

**Definición 2.21.** *Un proceso de saltos de Markov (que abreviaremos como PSM) es un proceso de Markov en tiempo continuo  $\{X_t\}_{t \geq 0}$ , con espacio de estados a lo más numerable  $E$  y que cumple las siguientes condiciones:*

1. *Cada trayectoria  $t \rightarrow X_t$  es continua por la derecha con límite por la izquierda.*
2. *Para cualesquiera estados  $x_0, x_1, \dots, x_{n+1}$  y para cualesquiera  $t_i < t_{i+1} = s_{i+1} + t_i$  con  $i \in \{0, \dots, n\}$ , con  $t_0 = 0$  se cumple:*

$$\begin{aligned} \mathbb{P}[X_{t_{n+1}} = x_{n+1} | X_{t_n} = x_n, \dots, X_{t_0} = x_0] &= \mathbb{P}[X_{t_{n+1}} = x_{n+1} | X_{t_n} = x_n] \\ &= \mathbb{P}[X_{s_{n+1}} = x_{n+1} | X_0 = x_n]. \end{aligned}$$

**Nota 2.7.** *En lo sucesivo supondremos que el espacio de estados  $E = \{1, 2, \dots, N\}$ .*

**Proposición 2.14.** *Un PSM  $\{X_t\}_{t \geq 0}$  es constante por pedazos.*

*Demostración.* Sea  $s \geq 0$ , consideremos una trayectoria fija del proceso, supongamos que  $X_s = x_0$ , como el espacio de estado es finito, existe  $u > 0$  tal que ningún estado en  $E$  salvo  $x_0$  se encuentra en el intervalo  $(x_0 - u, x_0 + u)$ . Como el proceso es continuo por la derecha entonces existe  $\delta > 0$  tal que el proceso en  $[s, s + \delta)$  se encuentra en el intervalo  $(x_0 - u, x_0 + u)$ , pero como en este intervalo solo vive  $x_0$ , entonces se concluye que el proceso es constante en  $[s, s + \delta)$ .  $\square$

**Definición 2.22.** *La función definida por*

$$p_{xy}(t) = \mathbb{P}[X_t = y | X_0 = x]$$

*con  $t \geq 0$  y  $x, y \in E$  es llamada la función de transición.*

**Definición 2.23.** *Considere un PSM  $\{X_t\}_{t \geq 0}$  con espacio de estados  $E$ , se define la matriz de transición al tiempo  $t \geq 0$  como*

$$P(t) = \{p_{xy}(t)\}_{x, y \in E}.$$

**Definición 2.24.** *A la familia  $\{P(t)\}_{t \geq 0}$  se le llama semigrupo de transición del proceso de Markov.*

**Proposición 2.15.** *El semigrupo de transición  $\{P(t)\}_{t \geq 0}$  de un PSM  $\{X_t\}_{t \geq 0}$  es estándar, es decir*

$$\lim_{t \rightarrow 0^+} P(t) = I,$$

*donde  $I$  es la matriz identidad de dimensión  $N \times N$ .*

*Demostración.* Por la Proposición 2.14

$$\mathbb{I}_{\{X_0=x, X_t=x\}} \xrightarrow[t \rightarrow 0^+]{c.s.} \mathbb{I}_{\{X_0=x\}}$$

y por el Teorema de la convergencia dominada (ver [?] página 62)

$$\lim_{t \rightarrow 0^+} \mathbb{E}[\mathbb{I}_{\{X_0=x, X_t=x\}}] = \mathbb{E}[\mathbb{I}_{\{X_0=x\}}].$$

Concluimos que

$$\lim_{t \rightarrow 0^+} p_{xy}(t) = \begin{cases} 1 & \text{si } x = y, \\ 0 & \text{si } x \neq y. \end{cases}$$

□

### 2.3.2. Resultados elementales

**Proposición 2.16. (Ecuación de Chapman-Kolmogorov)** Considere un PSM  $\{X_t\}_{t \geq 0}$  con espacio de estados  $E$ , entonces para cualesquiera par de numeros  $t$  y  $s$  positivos y para todo  $x, y \in E$  tenemos

$$p_{xy}(t+s) = \sum_{z \in E} p_{xz}(t)p_{zy}(s).$$

Debido a que un PSM es constante por pedazos, ahora nos interesa con qué frecuencia este proceso cambia de estado.

**Definición 2.25.** Se define el tiempo de estancia en el estado  $X_t$  de un PSM  $\{X_t\}_{t \geq 0}$  como

$$\Delta(t) := \inf\{s > 0 : X_{t+s} \neq X_t\}$$

para todo  $t \geq 0$ .

**Nota 2.8.** Esta variable aleatoria está bien definida porque el proceso es constante por pedazos.

**Definición 2.26.** Se define el tiempo de primer salto de un PSM  $\{X_t\}_{t \geq 0}$  como

$$T_1 := \inf\{t > 0 : X_t \neq X_0\}$$

y el  $n$ -ésimo tiempo de salto como

$$T_n = \inf\{t > T_{n-1}; X_t \neq X_{T_{n-1}}\} \text{ para todo } n \geq 1.$$

Por convención definimos  $T_0 = 0$ .

**Teorema 2.6.** Considere un PSM  $\{X_t\}_{t \geq 0}$  con espacio de estados  $E$ , si en el tiempo  $t \geq 0$   $\Delta(t) < \infty$  entonces para todo  $x \in E$  existe  $\lambda_x > 0$ , independiente de  $t$ , tal que

$$\mathbb{P}[\Delta(t) > s | X_t = x] = \exp(-\lambda_x s)$$

para toda  $s > 0$ .

*Demostración.* Debido a que un PSM es homogéneo en el tiempo

$$\mathbb{P}[\Delta(t) > s | X_t = x] = \mathbb{P}[\Delta(0) > s | X_0 = x],$$

definimos  $h(s) = \mathbb{P}[\Delta(0) > s | X_0 = x]$ , entonces se tiene que

$$\begin{aligned} h(t+s) &= \mathbb{P}[\Delta(0) > t+s | X_0 = x] = \mathbb{P}[\Delta(0) > t, \Delta(t) > s | X_0 = x] \\ &= \mathbb{P}[\Delta(0) > t | X_0 = x] \mathbb{P}[\Delta(t) > s | \Delta(0) > t, X_0 = x] \\ &= h(t) \mathbb{P}[\Delta(t) > s | X_t = x] \\ &= h(t)h(s). \end{aligned}$$

La función  $h$  es continua por la derecha, ya que es función de sobrevivencia, además  $0 \leq h(s) \leq 1$  para toda  $s \geq 0$ , lo cual implica que la única solución debe ser

$$h(s) = \exp(-\lambda_x s),$$

donde  $\lambda_x > 0$ . □

Ahora demostraremos un importante teorema que nos muestra la manera de como simular procesos de tipo PSM.

**Teorema 2.7.** *Considere un PSM  $\{X_t\}_{t \geq 0}$  con  $T_1 < \infty$ , entonces para cualesquiera estados  $x_0, x_1, \dots, x_k$  y para cualesquiera tiempos  $t_i < t_{i+1} = s_{i+1} + t_i$  con  $i \in \{0, \dots, k-1\}$  y  $t_0 = 0, x_0 = y \neq x$  se tienen las siguientes dos afirmaciones*

1.

$$\mathbb{P}[X_{T_1} = y | T_1 \geq t, X_0 = x]$$

no depende de  $t$ , es decir la distribución de salto es independiente del tiempo de primer salto.

2.

$$\mathbb{P}[X_{T_1+t_i} = x_i, \forall i \leq k | T_1 \geq t, X_0 = x] = \mathbb{P}[X_{T_1} = y | T_1 \geq t, X_0 = x] \prod_{j=0}^{k-1} \mathbb{P}[X_{s_{j+1}} = x_{j+1} | X_0 = x_j],$$

es decir, el proceso después de  $T_1$  es independiente de  $T_1$  y se comporta en ley de la misma forma que el proceso original, con distribución inicial  $\mathbb{P}[X_{T_1} = y | T_1 \geq t, X_0 = x]$ .

*Demostración.* Considere el proceso estocástico definido por  $Y_n^m = X_{n/2^m}$ , con  $n \in \{0, 1, \dots\}$ , es fácil darse cuenta que es de Markov homogéneo en el tiempo.

Se define

$$\sigma_m = (1/2^m) \text{mín}\{n > 0 : Y_n^m \neq X_0\}.$$

Evidentemente se tiene que  $\sigma_m \geq \sigma_{m+1} \geq T_1$  para toda  $m \in \mathbb{N}$ , ya que

$$Y_{2^{m+1}\sigma_m}^{m+1} = X_{2^{m+1}\sigma_m/2^{m+1}} = X_{\sigma_m} = Y_{2^m\sigma_m}^m \neq X_0,$$

por lo que tenemos  $2^{m+1}\sigma_m \geq 2^{m+1}\sigma_{m+1}$  para toda  $m \in \mathbb{N}$ , además  $\sigma_m - T_1 < \frac{1}{2^m}$  para toda  $m \in \mathbb{N}$ . Demostraremos la segunda afirmación del teorema.

Denotemos a  $n_m$  como el menor entero  $n$  tal que  $\frac{n}{2^m} \geq t$ , debido a que  $\{X_t\}_{t \geq 0}$  es continuo por la derecha

$$\begin{aligned}
& \mathbb{P}[X_{T_1+t_i} = x_i, \forall i \leq k, T_1 \geq t, X_0 = x] \\
&= \lim_{m \rightarrow \infty} \mathbb{P}[X_{\sigma_m+t_i} = x_i, \forall i \leq k, \sigma_m \geq t, X_0 = x] \\
&= \lim_{m \rightarrow \infty} \mathbb{P}[X_{\sigma_m+t_i} = x_i, \forall i \leq k, \sigma_m \geq \frac{n_m}{2^m}, X_0 = x] \\
&= \lim_{m \rightarrow \infty} \sum_{n=n_m}^{\infty} \mathbb{P}[X_{\sigma_m+t_i} = x_i, \forall i \leq k, \sigma_m = \frac{n}{2^m}, X_0 = x] \\
&= \lim_{m \rightarrow \infty} \sum_{n=n_m}^{\infty} \mathbb{P}[X_{n/2^m+t_i} = x_i, \forall i \leq k, X_{0/2^m} = x, \dots, X_{(n-1)/2^m} = x]
\end{aligned}$$

y por la propiedad de Markov tenemos

$$\begin{aligned}
&= \lim_{m \rightarrow \infty} \sum_{n=n_m}^{\infty} \mathbb{P}[X_0 = x] \mathbb{P}[X_{1/2^m} = x | X_0 = x]^{n-1} \mathbb{P}[X_{1/2^m} = y | X_0 = x] \prod_{j=0}^{k-1} \mathbb{P}[X_{s_{j+1}} = x_{j+1} | X_0 = x] \\
&= \lim_{m \rightarrow \infty} \mathbb{P}[X_{1/2^m} = y | X_0 = x] \prod_{j=0}^{k-1} \mathbb{P}[X_{s_{j+1}} = x_{j+1} | X_0 = x] \sum_{n=n_m}^{\infty} \mathbb{P}[X_0 = x] \mathbb{P}[X_{1/2^m} = x | X_0 = x]^{n-1}
\end{aligned}$$

por lo anterior

$$\mathbb{P}[T_1 \geq t | X_0 = x] = \lim_{m \rightarrow \infty} (1 - \mathbb{P}[X_{1/2^m} = x | X_0 = x]) \sum_{n=n_m}^{\infty} \mathbb{P}[X_{1/2^m} = x | X_0 = x]^{n-1}$$

por lo tanto

$$\begin{aligned}
& \mathbb{P}[X_{T_1+t_i} = x_i, \forall i \leq k | T_1 \geq t, X_0 = x] \\
&= \prod_{j=0}^{k-1} \mathbb{P}[X_{s_{j+1}} = x_{j+1} | X_0 = x] \lim_{m \rightarrow \infty} \frac{\mathbb{P}[X_{1/2^m} = y | X_0 = x]}{1 - \mathbb{P}[X_{1/2^m} = x | X_0 = x]}
\end{aligned}$$

tomando  $k = 0$ , se concluye

$$\mathbb{P}[X_{T_1} = y | T_1 \geq t, X_0 = x] = \lim_{m \rightarrow \infty} \frac{\mathbb{P}[X_{1/2^m} = y | X_0 = x]}{1 - \mathbb{P}[X_{1/2^m} = x | X_0 = x]}$$

esto demuestra la primer y segunda afirmación. □

**Definición 2.27.** Dado un PSM  $\{X_t\}_{t \geq 0}$  con espacio de estados  $E$ , se define la matriz de transición asociada a la distribución de salto como:

$$P = \{p_{xy}\}_{x,y \in E} = \{\mathbb{P}[X_{T_1} = y | X_0 = x]\}_{x,y \in E}.$$

**Teorema 2.8.** Considere un PSM  $\{X_t\}_{t \geq 0}$  con espacio de estados  $E$ , supongamos que  $T_1 < \dots < T_n < \infty$ , entonces para cualesquiera tiempos  $t_1, \dots, t_{n-1}, t_n$ ,  $0 = s_0 < \dots < s_m$  y cualesquiera estados  $x_0, x_1, \dots, x_n = y_0, y_1, \dots, y_m \in E$ .

Bajo estas condiciones se tienen las siguientes dos afirmaciones:

1.  $\mathbb{P}[T_k - T_{k-1} \leq t_k, X_{T_k} = x_k, 1 \leq k \leq n, X_{T_n+s_j} = y_j, 0 \leq j \leq m | X_0 = x_0]$   
 $= \mathbb{P}[T_k - T_{k-1} \leq t_k, X_{T_k} = x_k, 1 \leq k \leq n | X_0 = x_0] \mathbb{P}[X_{s_k} = y_k, 0 \leq k \leq m | X_0 = y_0].$
2.  $\mathbb{P}[X_{T_n+s_j} = y_j, 0 \leq j \leq m | X_0 = x_0, T_k - T_{k-1} \leq t_k, X_{T_k} = x_k, 1 \leq k \leq n]$   
 $= \mathbb{P}[X_{s_k} = y_k, 0 \leq k \leq m | X_0 = y_0].$

*Demostración.* El teorema se sigue de aplicar el Teorema 2.7 multiples veces. □

**Nota 2.9.** Por el Teorema 2.6 y el Teorema 2.8 podemos concluir que para todo  $n \in \mathbb{N}$

1. El proceso después del tiempo  $T_n$  es condicionalmente independiente a la trayectoria antes de  $T_n$  dado  $X_{T_n}$  y se distribuye igual que el proceso original empezando en el estado  $X_{T_n}$ .
2.  $T_{n+1} - T_n$  condicionado a  $X_{T_n} = x$  se distribuye exponencial de parámetro  $\lambda_x$ , además es independiente a la distribución del siguiente salto.

**Corolario 2.3.** Considere un PSM  $\{X_t\}_{t \geq 0}$  con espacio de estados  $E$ , supongamos que  $T_1 < \dots < T_n < \infty$ , entonces para cualesquiera tiempos  $t_1, \dots, t_n$  y cualesquiera estados  $x_0, x_1, \dots, x_{n-1} \in E$

$$\mathbb{P}[\Delta(T_0) \leq t_1, \dots, \Delta(T_{n-1}) \leq t_n | X_{T_k} = x_k, 0 \leq k \leq n-1] = \prod_{k=0}^{n-1} \mathbb{P}[\Delta(0) \leq t_{k+1} | X_0 = x_k].$$

**Definición 2.28.** Dado un PSM  $\{X_t\}_{t \geq 0}$ , el Teorema 2.8 nos asegura que el proceso estocástico definido como:

$$\{Z_n = X_{T_n}\}_{n \geq 0}.$$

Conforma una cadena de Markov con matriz de transición dada por la Definición 2.27.

### 2.3.3. Generador infinitesimal

En esta parte del trabajo, nosotros caracterizaremos a un PSM mediante los cambios infinitesimales que ocurren en intervalos de tiempo muy pequeños.

Considere un PSM  $\{X_t\}_{t \geq 0}$  con espacio de estados  $E$ , con  $X_t = x$ . Ahora, estamos interesados en analizar que pasa en intervalos de tiempo infinitesimales  $(t, t+h)$ , tenemos los siguientes casos.

1. El proceso puede estar en el mismo estado al tiempo  $t+h$  y esto ocurre con probabilidad  $p_{xx}(h) + o(h)$ , donde  $o(h)$  representa la posibilidad que el proceso se salga y regrese al estado  $i$  en el intervalo de tiempo.
2. El proceso se mueva al estado  $y$  en  $t+h$  y esto ocurre con probabilidad  $p_{xy}(h) + o(h)$ , donde  $o(h)$  representa la posibilidad que el proceso realice dos o más transiciones en el intervalo de tiempo.

Ahora estamos interesados en el comportamiento de  $p_{xy}(h)$  para  $h$  pequeña, para esto consideremos la siguiente proposición.

**Proposición 2.17.** *Considere un PSM  $\{X_t\}_{t \geq 0}$  con espacio de estados  $E$ , entonces para cualesquiera  $x, y \in E$*

$$\lim_{h \rightarrow 0^+} \frac{p_{xy}(h) - \delta_{xy}}{h} = q_{xy}.$$

Donde

$$\delta_{xy} = \begin{cases} 1 & \text{para } x = y \\ 0 & \text{para } x \neq y. \end{cases}$$

y  $q_{xy} = \lambda_x p_{xy}$  si  $x \neq y$  y  $q_{xx} = -\lambda_x$ .

Para probar esta proposición necesitamos el siguiente lema.

**Lema 2.1.** *Considere un PSM  $\{X_t\}_{t \geq 0}$  con espacio de estados  $E$ , supongamos  $T_1 < \dots < T_n < \infty$ . Bajo estas condiciones*

$$\mathbb{P}[T_n \leq t] \leq \int_0^{\beta t} \frac{s^{n-1} \exp(-s)}{(n-1)!} ds.$$

Además si sumamos sobre  $n$  obtenemos

$$\sum_{n=1}^{\infty} \mathbb{P}[T_n \leq t] \leq \beta t.$$

Donde  $\beta = \max\{\lambda_x : x \in E\}$ .

*Demostración.* Sabemos que

$$\begin{aligned} \mathbb{P}[T_n \leq t | X_{T_k} = x_k, 0 \leq k \leq n-1] &= \mathbb{P}\left[\sum_{j=1}^n \frac{\lambda_{x_k}(T_j - T_{j-1})}{\lambda_{x_k}} \leq t | X_{T_k} = x_k, 0 \leq k \leq n-1\right] \\ &\leq \mathbb{P}\left[\sum_{j=1}^n \frac{\lambda_{x_k}(T_j - T_{j-1})}{\beta} \leq t | X_{T_k} = x_k, 0 \leq k \leq n-1\right] \\ &= \mathbb{P}\left[\sum_{j=1}^n \lambda_{x_k}(T_j - T_{j-1}) \leq \beta t | X_{T_k} = x_k, 0 \leq k \leq n-1\right]. \end{aligned}$$

Por el Corolario 2.3

$$\sum_{j=1}^n \lambda_{x_k}(T_j - T_{j-1}) | X_{T_k} = x_k, 0 \leq k \leq n-1$$

se distribuye Gamma de parámetros  $(n, 1)$ , por lo que se tiene

$$\mathbb{P}[T_n \leq t] \leq \int_0^{\beta t} \frac{s^{n-1} \exp(-s)}{(n-1)!} ds.$$

Sumando sobre  $n$  y aplicando Teorema de la convergencia dominada (ver [?] página 62)

$$\sum_{n=1}^{\infty} \mathbb{P}[T_n \leq t] \leq \int_0^{\beta t} \sum_{n=1}^{\infty} \frac{s^{n-1} \exp(-s)}{(n-1)!} ds = \int_0^{\beta t} ds = \beta t.$$

□

**Definición 2.29.** Dado un PSM  $\{X_t\}_{t \geq 0}$  se define la variable que cuanta los saltos hasta el instante  $t$  como

$$N_t = \text{máx}\{n \geq 1 : T_n \leq t\}.$$

Prueba de la Proposición 2.17

*Demostración.* La demostración se divide en dos casos.

Caso 1.  $x \neq y$ .

Por el Lema 2.1

$$\begin{aligned} \mathbb{P}[X_h = y | X_0 = x] &= \mathbb{P}[X_h = y, N_h = 1 | X_0 = x] + \mathbb{P}[X_h = y, N_h \geq 2 | X_0 = x] \\ &= \lambda_x p_{xy} \exp^{-\lambda_y h} [h \mathbb{I}_{\{\lambda_x = \lambda_y\}} + \left(\frac{\exp^{(\lambda_y - \lambda_x)h} - 1}{\lambda_y - \lambda_x}\right) \mathbb{I}_{\{\lambda_x \neq \lambda_y\}}] + o(h). \end{aligned}$$

Por lo anterior

$$\lim_{h \rightarrow 0^+} h^{-1} p_{xy}(h) = q_{xy}.$$

Caso 2.  $x = y$ .

Por el Caso 1 se tiene

$$\lim_{h \rightarrow 0^+} h^{-1} (p_{xx}(h) - 1) = -\lambda_x = q_{xx}.$$

□

**Definición 2.30.** El generador infinitesimal de un PSM, es la matriz  $Q = q_{xy, x, y \in E}$  con entradas  $q_{xy} = \lambda_x p_{xy}$  si  $x \neq y$  y  $q_{xx} = -\lambda_x$ .

**Nota 2.10.** Observemos que el generador infinitesimal contiene toda la información del proceso, tanto las intensidades de los tiempos de salto  $\lambda_x$  como la distribución de los saltos  $\exp(-\lambda_x t)$ .

El siguiente Teorema nos muestra que el generador infinitesimal de un PSM es suficiente para construir las probabilidades de transición  $\{p_{xy}(t)\}_{x, y \in E}$ .

**Teorema 2.9.** Dado un PSM  $\{X_t\}_{t \geq 0}$  con espacio de estados  $E$  y  $Q$  su correspondiente generador infinitesimal, entonces se satisface el siguiente sistema de ecuaciones diferenciales.

$$\frac{\partial P(t)}{\partial t} = QP(t).$$

*Demostración.* Sea  $\varepsilon > 0$

Por la igualdad de Chapman-Kolmogorov(2.16) tenemos

$$\frac{p_{xy}(t+h) - p_{xy}(t)}{h} = \sum_{z \in E} \frac{(p_{xz}(h) - \delta_{xz}) p_{zy}(t)}{h}.$$

Ahora basta mostrar que

$$\lim_{h \rightarrow 0^+} \frac{p_{xz}(h) - \delta_{xz}}{h} = q_{xz}.$$

Pero esto es resultado de la Proposición 2.17.

□

**Nota 2.11.** El teorema 2.9 muestra que  $P(t)$  cumple la ecuación diferencial

$$\frac{\partial P(t)}{\partial t} = QP(t),$$

si la matriz  $Q$  es diagonalizable, entonces  $P(t) = \exp(tQ) = \sum_{n=0}^{\infty} \frac{t^n Q^n}{n!}$ , es la única solución al sistema.

Ahora bien, como la tasa de saltos  $\lambda_x$  es independiente del tiempo  $t \geq 0$  y sólo depende del estado  $x \in E$ , clasificaremos a los estados.

**Definición 2.31.** Sea  $x \in E$  con tasa de salto asociada  $\lambda_x$ , entonces  $x$  se define como:

1. Permanente, si  $\lambda_x = 0$ .
2. Estable, si  $0 < \lambda_x < \infty$ .
3. Instantáneo, si  $\lambda_x = \infty$ .

Sea  $\{X_t\}_{t \geq 0}$  un PSM con  $X_t = x$  a tiempo  $t \geq 0$ , estudiaremos las consecuencias del proceso dependiendo cada tipo de estado.

1. Si  $x$  es permanente,  $P(\Delta(t) = \infty | X_t = x) = 1$  para toda  $s > t$ . Es decir, el PSM permanece siempre en el estado  $i \in E$ .
2. Si  $x$  es estable,  $P(0 < \Delta(t) < \infty | X_t = x) = 1$ , es decir el proceso saldrá del estado  $i$  en un tiempo de espera positivo y finito.
3. Si  $x$  es instantáneo,  $P(\Delta(t) = 0 | X_t = x) = 1$ , entonces el PSM existe en el estado tan pronto entra en él.

**Proposición 2.18. (Regularidad).** Considere un PSM  $\{X_t\}_{t \geq 0}$  y  $T_1 < \dots < T_n$  sus respectivos  $n$  tiempos de salto, asumamos que su correspondiente generador infinitesimal tiene entradas estrictamente negativas y finitas en la diagonal, bajo estas condiciones se tiene que

1.  $T_1 < \dots < T_n < \infty$  casi seguramente.
2.  $T_n \xrightarrow[n \rightarrow \infty]{c.s.} \infty$ .

*Demostración.* Por el Lema 2.1 podemos argumentar que para toda  $t \geq 0$

$$\mathbb{P}[T_n \leq t] \leq 1 - \exp(-\beta t) \sum_{j=0}^{n-1} (\beta t)^j / j!.$$

Donde  $\beta = \max\{\lambda_x : x \in E\}$ , así concluimos que

$$T_1 < \dots < T_n < \infty \text{ casi seguramente.}$$

La segunda es resultado de hacer tender  $n$  a infinito y notar que  $\mathbb{P}[\lim_{n \rightarrow \infty} T_n \leq t] \leq \lim_{n \rightarrow \infty} \mathbb{P}[T_n \leq t] = 0$  para cada  $t \geq 0$ .  $\square$

**Nota 2.12.** A partir de ahora solo consideraremos procesos cuyos generadores infinitesimales tengan entradas en la diagonal estrictamente negativas. La existencia de estos procesos se probará más adelante, específicamente en el Teorema 3.2.

### 2.3.4. Distribuciones estacionarias y distribuciones límite

**Definición 2.32.** Sea un PSM  $\{X_t\}_{t \geq 0}$  con espacio de estados  $E$ .

Se dice que  $\{X_t\}_{t \geq 0}$  es irreducible si su cadena de Markov asociada a los saltos  $\{Z_n\}_{n \geq 0}$  (ver Definición 2.28) es irreducible.

Por otro lado decimos que el proceso PSM es recurrente si  $\{Z_n\}_{n \geq 0}$  es recurrente.

**Definición 2.33.** Dado un PSM  $\{X_t\}_{t \geq 0}$  se define el tiempo de llegada hacia el estado  $x$  como:

$$\tau_x = \inf\{t > T_1 : X_t = x\}.$$

Un PSM se dice ser recurrente positivo si para todo  $x \in E$  se tiene que  $\mathbb{E}[\tau_x | X_0 = x] < \infty$ .

**Definición 2.34.** Un PSM  $\{X_t\}_{t \geq 0}$  se dice ergódico si es irreducible y recurrente positivo.

En lo que resta de este capítulo enunciaremos resultados que son análogos al caso de cadenas de Markov en tiempo discreto y sus demostraciones serán omitidas y podrán ser consultadas en [?].

**Proposición 2.19.** Considere un PSM  $\{X_t\}_{t \geq 0}$  que es recurrente e irreducible, entonces  $\{X_t\}_{t \geq 0}$  es ergódico.

**Definición 2.35.** Considere un PSM  $\{X_t\}_{t \geq 0}$ , una medida de probabilidad  $\nu = \{\nu_y\}_{y \in E}$  sobre  $E$  es una distribución estacionaria si para todo  $y \in E$  y  $t \geq 0$

$$\nu_y = \sum_{x \in E} \nu_x p_{xy}(t).$$

**Teorema 2.10.** Considere un PSM  $\{X_t\}_{t \geq 0}$  ergódico, entonces las siguientes afirmaciones son válidas.

1. Existe una única distribución estacionaria  $\nu$  dada por

$$\nu_y = \frac{\mathbb{E}[\int_0^{\tau_x} \mathbb{1}_{\{X_t=y\}} dt | X_0 = x]}{\mathbb{E}[\tau_x | X_0 = x]}, \text{ para todo } y \in E.$$

La cual es independiente del estado  $x$ .

2. La distribución  $\nu$  satisface la ecuación

$$\sum_{x \in E} \nu_x q_{xy} = 0, \text{ para todo } y \in E.$$

*Demostración.* Para su demostración ver ([?] página 261). □

**Teorema 2.11. (Ergódico).** Dado un PSM ergódico  $\{X_t\}_{t \geq 0}$  con distribución estacionaria  $\nu$  y una función  $f : E \rightarrow \mathbb{R}$ ,

$$\lim_{t \rightarrow \infty} t^{-1} \int_0^t f(X_s) ds = \sum_{x \in E} f(x) \nu_x \text{ casi seguramente.}$$

*Demostración.* Para su demostración ver ([?] página 265).  $\square$

**Definición 2.36.** Dado un PSM  $\{X_t\}_{t \geq 0}$  se define la distribución límite como

$$\lim_{t \rightarrow \infty} p_{xy}(t) \text{ para todo } y \in E.$$

Siempre que está exista y no dependa de  $x$ .

**Teorema 2.12. (Distribución límite)** Dado un PSM ergódico  $\{X_t\}_{t \geq 0}$ , existe la distribución límite y está dada por

$$\lim_{t \rightarrow \infty} p_{xy}(t) = \nu_y \text{ para todo } y \in E.$$

Donde  $\nu = \{\nu_y\}_{y \in E}$  está definida en el Teorema 2.10

Notemos que este límite no depende de  $x$ .

**Teorema 2.13.** Dado un PSM ergódico  $\{X_t\}_{t \geq 0}$  y  $A, B \subseteq E$  se tiene

$$\lim_{t \rightarrow \infty} t^{-1} \sum_{n=0}^{N_t} \mathbb{I}_{\{Z_n \in A, Z_{n+1} \in B\}} = \sum_{x \in A} \nu_x \sum_{y \in B} q_{xy}.$$

Este límite puede entenderse como el promedio de transiciones de  $B$  desde  $A$  por unidad de tiempo.

*Demostración.* La demostración se encuentra en ([?] página 269).  $\square$

### 2.3.5. Ejemplos

**Ejemplo 2.11. (Proceso de nacimiento y muerte)** Un proceso de nacimiento y muerte es un PSM con espacio de estados  $E = \{0, 1, \dots\}$  o  $E = \{0, 1, \dots, k\}$ , si  $E = \{0, 1, \dots\}$  el generador infinitesimal es

$$Q = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \dots \\ \alpha_1 & -(\lambda_1 + \alpha_1) & \lambda_1 & 0 & \dots \\ 0 & \alpha_2 & -(\lambda_2 + \alpha_2) & \lambda_2 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}.$$

Mientras que si  $E = \{0, 1, \dots, k\}$  el generador infinitesimal es

$$Q = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \dots & 0 \\ \alpha_1 & -(\lambda_1 + \alpha_1) & \lambda_1 & 0 & \dots & 0 \\ 0 & \alpha_2 & -(\lambda_2 + \alpha_2) & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \alpha_k & -\alpha_k \end{pmatrix}.$$

Donde  $\lambda_0, \lambda_1, \dots$  y  $\alpha_1, \alpha_2, \dots$  son constantes positivas conocidas como las tasas instantáneas de nacimiento y muerte, respectivamente. De acuerdo a lo desarrollado antes, a partir de esta matriz podemos concluir que el tiempo de estancia en cualquier estado  $x \in E$  tiene distribución  $\exp(\lambda_x + \alpha_x)$ , en donde se define  $\alpha_0 = 0$ . Las probabilidades de saltos de un estado a otro son

$$p_{xy} = \begin{cases} \frac{\lambda_x}{\lambda_x + \alpha_x} & \text{si } y = x + 1, \\ \frac{\alpha_x}{\lambda_x + \alpha_x} & \text{si } y = x - 1, \\ 0 & \text{otro caso.} \end{cases} \quad (2.1)$$

De este modo, un proceso de nacimiento y muerte permanece en cada uno de sus estados un tiempo exponencial, al cabo del cual salta una unidad hacia arriba o una unidad hacia abajo de acuerdo a las probabilidades de (2.1). Un salto hacia arriba se interpreta como un nacimiento, mientras que un salto hacia abajo representa una muerte. La variable  $X_t$  puede interpretarse como el número de individuos en una población al tiempo  $t$ , en donde pueden presentarse nacimientos y muertes de individuos, uno a la vez. Como  $\lambda_0 > 0$  y  $\alpha_0 = 0$ , la población puede crecer cuando se encuentre en el estado cero, pero nunca decrecer por debajo de ese nivel.

Nos concentraremos en el caso donde  $E = \{0, 1, \dots, k\}$  y las tasas sean estrictamente positivas, podemos notar que para este caso  $\lambda_k = 0$ , esto quiere decir que la población no puede ser mayor que  $k$ .

Por el Teorema 2.9 sabemos que  $P(t)$  cumple el siguiente sistema de ecuaciones diferenciales.

$$\frac{\partial P(t)}{\partial t} = QP(t).$$

Donde  $Q$  es el generador infinitesimal y  $P(t)$  es la matriz de transición al tiempo  $t$ . Para encontrar  $P(t)$  debemos probar primero que  $Q$  es diagonalizable, y por la teoría de ecuaciones diferenciales podemos concluir que

$$P(t) = \exp(Qt).$$

Sea  $D \in \mathcal{M}_{k+1 \times k+1}$  una matriz diagonal con entradas positivas en la diagonal

$$D = \begin{pmatrix} d_0 & 0 & 0 & 0 & \dots & 0 \\ 0 & d_1 & 0 & 0 & \dots & 0 \\ 0 & 0 & d_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & d_k \end{pmatrix},$$

la matriz resultante del producto  $D^{-1}QD$  esta dada por

$$D^{-1}QD = \begin{pmatrix} -\lambda_0 & \lambda_0 \frac{d_1}{d_0} & 0 & 0 & \dots & 0 \\ \alpha_1 \frac{d_0}{d_1} & -(\lambda_1 + \alpha_1) & \lambda_1 \frac{d_2}{d_1} & 0 & \dots & 0 \\ 0 & \alpha_2 \frac{d_1}{d_2} & -(\lambda_2 + \alpha_2) & \lambda_2 \frac{d_3}{d_2} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \alpha_k \frac{d_{k-1}}{d_k} & -\alpha_k \end{pmatrix}.$$

Nuestro objetivo es encontrar una matriz  $D$  diagonal tal que  $D^{-1}QD$  sea una matriz simétrica, de este modo se forma el siguiente sistema de ecuaciones

$$\begin{aligned} \alpha_1 \frac{d_0}{d_1} &= \lambda_0 \frac{d_1}{d_0} \\ \alpha_2 \frac{d_1}{d_2} &= \lambda_1 \frac{d_2}{d_1} \\ &\vdots \\ \alpha_k \frac{d_{k-1}}{d_k} &= \lambda_{k-1} \frac{d_k}{d_{k-1}}. \end{aligned}$$

Por lo anterior obtenemos

$$d_n = \sqrt{\frac{\alpha_n}{\lambda_{n-1}}} d_{n-1} \text{ para toda } 1 \leq n \leq k,$$

haciendo  $d_0 = 1$  obtenemos

$$d_n = \sqrt{\frac{M_n}{L_{n-1}}} \text{ para toda } 1 \leq n \leq k,$$

donde  $M_n = \prod_{i=1}^n \alpha_i$  y  $L_{n-1} = \prod_{i=0}^{n-1} \lambda_i$ .  
Sea  $U = D^{-1}QD$ , por lo anterior

$$U = \begin{pmatrix} -\lambda_0 & \sqrt{\lambda_0 \alpha_1} & 0 & 0 & \dots & 0 \\ \sqrt{\lambda_0 \alpha_1} & -(\lambda_1 + \alpha_1) & \sqrt{\lambda_1 \alpha_2} & 0 & \dots & 0 \\ 0 & \sqrt{\lambda_1 \alpha_2} & -(\lambda_2 + \alpha_2) & \sqrt{\lambda_2 \alpha_3} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \sqrt{\lambda_{k-1} \alpha_k} & -\alpha_k \end{pmatrix},$$

por el Teorema espectral ([?] página 481)  $U$  puede descomponerse en  $U = V^{-1}WV$ , con  $W$  una matriz diagonal con valores reales.

En conclusión  $Q$  es diagonalizable y se puede descomponer en

$$Q = DV^{-1}WVD^{-1} = \tilde{D}^{-1}W\tilde{D},$$

con  $\tilde{D} = VD^{-1}$ . Por lo tanto

$$P(t) = \tilde{D}^{-1} \exp(Wt) \tilde{D}.$$

El proceso es irreducible y recurrente, pues las probabilidades de (2.1) producen una cadena de Markov irreducible y recurrente. Por la Proposición 2.19 sabemos que también es ergódico, y por el Teorema 2.10 hay una única distribución estacionaria que cumple

$$\sum_{x \in E} \nu_x q_{xy} = 0, \text{ para todo } y \in E,$$

que se traduce en

$$\begin{aligned} -\lambda_0 \nu_0 + \alpha_1 \nu_1 &= 0 \\ \lambda_0 \nu_0 - (\lambda_1 + \alpha_1) \nu_1 + \alpha_2 \nu_2 &= 0 \\ \lambda_1 \nu_1 - (\lambda_2 + \alpha_2) \nu_2 + \alpha_3 \nu_3 &= 0 \\ &\vdots \\ \lambda_{k-2} \nu_{k-2} - (\lambda_{k-1} + \alpha_{k-1}) \nu_{k-1} + \alpha_k \nu_k &= 0 \\ -\alpha_k \nu_k + \lambda_{k-1} \nu_{k-1} &= 0 \end{aligned}$$

acomodando los términos obtenemos

$$\begin{aligned} \alpha_1 \nu_1 - \lambda_0 \nu_0 &= 0 \\ \alpha_2 \nu_2 - \lambda_1 \nu_1 &= \alpha_1 \nu_1 - \lambda_0 \nu_0 \\ \alpha_3 \nu_3 - \lambda_2 \nu_2 &= \alpha_2 \nu_2 - \lambda_1 \nu_1 \\ &\vdots \\ \alpha_k \nu_k - \lambda_{k-1} \nu_{k-1} &= \alpha_{k-1} \nu_{k-1} - \lambda_{k-2} \nu_{k-2} \\ \alpha_k \nu_k - \lambda_{k-1} \nu_{k-1} &= 0 \end{aligned}$$

por lo anterior  $\alpha_n \nu_n - \lambda_{n-1} \nu_{n-1} = 0$  para toda  $1 \leq n \leq k$ , es decir

$$\nu_n = \frac{\lambda_{n-1}}{\alpha_n} \nu_{n-1} = \dots = \frac{\lambda_0 \dots \lambda_{n-1}}{\alpha_1 \dots \alpha_n} \nu_0.$$

Sumando todo el espacio de estados  $E$  tenemos

$$\nu_0 = \frac{1}{1 + \sum_{i=1}^k \frac{\lambda_0 \dots \lambda_{i-1}}{\alpha_1 \dots \alpha_i}}.$$

Por el Teorema 2.10 sabemos que

$$\nu_y = \frac{1}{\lambda_y \mathbb{E}[\tau_y | X_0 = y]}, \text{ para todo } y \in E,$$

por tanto  $\mathbb{E}[\tau_y | X_0 = y] = \frac{1}{\nu_y \lambda_y}$ , para todo  $y \in E$ .

**Ejemplo 2.12. (Proceso de Poisson)** El proceso de Poisson homogéneo es un PSM que empieza en 0 es decir, la distribución inicial tiene el valor 1 en el estado 0. Los tiempos de estancia son exponenciales de parámetro  $\lambda = q_x$  y las probabilidades de salto de un estado a otro son

$$p_{xy} = \begin{cases} 1 & \text{si } y = x + 1, \\ 0 & \text{si } y \neq x + 1. \end{cases}$$

Así con nuestro generador infinitesimal podremos ir evolucionando el proceso de Poisson hasta un tiempo  $T$ , el cual es definido como:

$$q_{xy} = \begin{cases} \lambda & \text{si } y = x + 1, \\ 0 & \text{e.o.c.} \end{cases}$$

Considerando  $Q \in \mathbb{M}_{N \times N}$ :

$$Q(x, y) = \begin{cases} q_{xy} & \text{si } y \neq x, \\ -\lambda & \text{si } y = x. \end{cases}$$

Así tenemos:

$$Q = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & 0 & \dots \\ 0 & -\lambda & \lambda & 0 & 0 & \dots \\ 0 & 0 & -\lambda & \lambda & 0 & \dots \\ 0 & 0 & 0 & -\lambda & \lambda & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \dots \end{pmatrix}$$

Como el proceso siempre va creciendo, el proceso se degenera y no tiene solución. Así sugeriremos condiciones de frontera con espacio de estados  $E = \{0, 1, 2, 3, \dots, N\}$ , tomando el sistema de  $N + 1$  ecuaciones, así obtendríamos el siguiente generador infinitesimal:

$$Q = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & -\lambda & \lambda & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & -\lambda & \lambda & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \dots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 & -\lambda & \lambda \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \end{pmatrix}$$

Del cual obtenemos el siguiente sistema de ecuaciones:

$$\begin{cases} -\lambda \nu_0 = 0 \\ \lambda \nu_0 - \lambda \nu_1 = 0 \\ \lambda \nu_1 - \lambda \nu_2 = 0 \\ \vdots \\ \lambda \nu_{N-1} = 0 \\ \nu_1 + \nu_2 + \nu_3 + \dots + \nu_N = 1 \end{cases}$$

*Así  $\nu_N = 1$  y la distribución estacionaria del proceso de poisson con condiciones de frontera sería 1 en la última entrada:  $\nu = (0, 0, 0, \dots, 1)$*



# Capítulo 3

## Simulación de procesos de Markov

En este capítulo se pretende mostrar las principales técnicas de simulación de algunos procesos estocásticos, en particular nos concentraremos en los procesos estocásticos de Markov cuyas aplicaciones van desde las finanzas computacionales e investigación de operaciones hasta la genética y genómica humana. Este capítulo se centra en dos tipos de procesos de Markov, primero estudiaremos las técnicas de simulación sobre cadenas de Markov en tiempo discreto así como su respectiva justificación matemática, para posteriormente resolver algunos problemas analíticos mediante la simulación de cadenas de Markov. Posteriormente trabajaremos con procesos de saltos de Markov, mostraremos algunos resultados matemáticos que nos permitirán justificar la simulación de dichos procesos y finalmente resolveremos problemas analíticos mediante la simulación de dichos procesos. La bibliografía básica para este capítulo es [? ].

### 3.1. Cadenas de Markov

#### 3.1.1. Algoritmo para cadenas de Markov

Sabemos que dada una cadena de Markov homogénea podemos obtener su matriz de transición y su distribución inicial, ahora nos preguntamos si dada una matriz estocástica y una distribución inicial, es posible generar una cadena de Markov. El siguiente resultado nos asegura que sí se puede y además proporciona herramientas para generar trayectorias de cadenas de Markov.

**Teorema 3.1.** *Dada una matriz estocástica  $P = \{p_{ij}\}_{1 \leq i, j \leq N} \in \mathcal{M}_{N \times N}$ , una distribución inicial  $\pi$  sobre el espacio de estados  $E = \{1, 2, \dots, N\}$ , existe un espacio de probabilidad en el que están definidas una sucesión de variables aleatorias  $\{X_n\}_{n \in \mathbb{N}}$  en  $E$ , que conforman una cadena de Markov homogénea con espacio de estados  $E$ , matriz de transición  $P$  y distribución inicial  $\pi$ .*

*Demostración.* Consideremos un espacio de probabilidad  $(\Omega, \mathcal{F}, \mathbb{P})$  en el cual está definida una sucesión  $\{U_n\}_{n \in \mathbb{N}}$  de variables aleatorias independientes uniformes en el intervalo  $(0,1)$ . Definimos  $\phi_i(y)$  como la función de cuantiles (ver A.1) sobre  $E$  asociada a la distribución  $\{p_{ij}\}_{1 \leq j \leq N}$  y a  $\phi_0(y)$  como la función de cuantiles sobre  $E$  asociada a la distribución  $\pi$ , con  $y \in (0, 1)$ .

Sea  $X_0 = \phi_0(U_0)$  y  $X_{n+1} = \phi_{X_n}(U_{n+1})$  para  $n \geq 0$ .

Analicemos las probabilidades de transición

$$\mathbb{P}[X_n = x_n | X_{n-1} = x_{n-1}] = \mathbb{P}[\phi_{X_{n-1}}(U_n) = x_n | X_{n-1} = x_{n-1}]$$

Como  $X_{n-1}$  sólo depende de  $\{U_k\}_{k \leq n-1}$  y además las variables aleatorias  $\{U_k\}_{k \in \mathbb{N}}$  son independientes se tiene que

$$\mathbb{P}[\phi_{X_{n-1}}(U_n) = x_n | X_{n-1} = x_{n-1}] = \mathbb{P}[\phi_{x_{n-1}}(U_n) = x_n].$$

Por último  $\mathbb{P}[\phi_{x_{n-1}}(U_n) = x_n] = p_{x_{n-1}x_n}$  por definición de función de cuantiles.

Mostraremos que la sucesión de variables aleatorias  $\{X_n\}_{n \in \mathbb{N}}$ , cumple la propiedad de Markov.

$$\begin{aligned} \mathbb{P}[X_0 = x_0, X_1 = x_1, \dots, X_n = x_n] &= \mathbb{P}[\phi_0(U_0) = x_0, \phi_{X_0}(U_1) = x_1, \dots, \phi_{X_{n-1}}(U_n) = x_n] \\ &= \mathbb{P}[\phi_0(U_0) = x_0, \phi_{x_0}(U_1) = x_1, \dots, \phi_{x_{n-1}}(U_n) = x_n] \\ &\text{por la independencia de } \{U_k\}_{k \in \mathbb{N}} \\ &= \mathbb{P}[\phi_0(U_0) = x_0] \mathbb{P}[\phi_{x_0}(U_1) = x_1] \dots \mathbb{P}[\phi_{x_{n-1}}(U_n) = x_n] \\ &= \mathbb{P}[X_0 = x_0] \mathbb{P}[X_1 = x_1 | X_0 = x_0] \dots \mathbb{P}[X_n = x_n | X_{n-1} = x_{n-1}] \\ &= \mathbb{P}[X_0 = x_0] \prod_{i=0}^{n-1} p_{x_i x_{i+1}}. \end{aligned}$$

Por lo tanto  $\{X_n\}_{n \in \mathbb{N}}$  es una cadena de Markov, con distribución inicial  $\pi$  y matriz de transición  $P$ .  $\square$

El teorema anterior nos proporcionó la forma de generar trayectorias de una cadena de Markov homogénea sobre  $E$  con distribución inicial y matriz de transición dadas. A continuación daremos una construcción alternativa que nos servirá más adelante, sin embargo no se probará que esta construcción produce una cadena de Markov ya que es similar a la propuesta en el Teorema 3.1

**Definición 3.1. (Construcción 2 Cadenas de Markov).** Supongamos las mismas hipótesis que en el Teorema 3.1. Consideremos un espacio de probabilidad donde están definidas  $\{w_{i,k}\}_{i \in E, k \in \mathbb{N}}$  variables aleatorias independientes tales que

$$\mathbb{P}[w_{i,k} = j] = p_{ij} \text{ para todo } i, j \in E \text{ y } k \in \mathbb{N}. \quad (3.1)$$

Definimos a  $X_0 = \phi_0(U)$  y  $X_{k+1} = w_{Y_k, n_{Y_k}(k)+1}$  para cada  $k > 0$ .

Donde  $U$  es uniforme en  $(0,1)$  independiente de  $\{w_{i,k}\}_{i \in E, k \in \mathbb{N}}$ ,  $\phi_0$  la función de cuantiles correspondiente a la distribución inicial  $\pi = (\pi_1, \pi_2, \dots, \pi_N)$  y  $n_i(k)$  es el número de visitas al estado  $i$  hasta el tiempo  $k-1$ , definiendo  $n_i(0) = 0$  para todo  $i \in E$  y  $k \in \mathbb{N}$ .

Análogamente como se procedió en el Teorema 3.1 se demuestra que  $\{X_n\}_{n \in \mathbb{N}}$  es una cadena de Markov sobre  $E$  con matriz de transición  $P$  y distribución inicial  $\pi$ .

El resultado anterior nos proporciona una justificación de la existencia de cadenas de Markov, así como también nos señala una técnica para su simulación misma que se describe a continuación.

**Algoritmo 3.1. Cadenas de Markov a tiempo discreto.** Este algoritmo genera trayectorias de una cadena de Markov homogénea con distribución inicial  $\pi$  y matriz de transición  $P$  hasta el horizonte de tiempo  $m$ .

1. Generar  $X_0$  de la distribución inicial  $\pi$ .
2. Inicializamos  $n=1$ .
3. Generamos  $X_n$  de la distribución correspondiente al renglón de la matriz de transición asociado al estado  $X_{n-1}$ .
4. Si  $n$  es igual al horizonte de tiempo  $m$ , entonces detenemos el proceso, si no es así actualizamos  $n = n + 1$  y volvemos al paso 3.

**Ejemplo 3.1.** Se simuló una trayectoria de una cadena de Markov usando el Algoritmo 3.1 con la función del Código D.1 en el software Julia con las siguientes características:  
Matriz de transición

$$P = \begin{pmatrix} 0.2 & 0.1 & 0.5 & 0.2 \\ 0.5 & 0.2 & 0.2 & 0.1 \\ 0.2 & 0.5 & 0.1 & 0.2 \\ 0.3 & 0.5 & 0.1 & 0.1 \end{pmatrix}$$

vector de distribución inicial  $\pi$ :

$$(0.2 \quad 0.4 \quad 0.2 \quad 0.2)$$

vector de espacio de estados  $E$ :

$$(1 \quad 2 \quad 3 \quad 4).$$

Con horizonte de tiempo  $m=40$ .

En la figura ?? podemos observar la gráfica de una trayectoria simulada.

### 3.1.2. Cálculo numérico de propiedades de cadenas de Markov

A continuación aplicaremos técnicas de simulación desarrolladas anteriormente para realizar cálculos numéricos sobre propiedades de cadenas de Markov.

**Ejemplo 3.2.** Considere la cadena de Markov del Ejemplo 2.10 con  $k=3$ , por lo anterior el espacio de estados es  $E = \{0, 1, 2, 3\}$  y la matriz de transición es

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

En el Ejemplo 2.10 mostramos analíticamente que la cadena es irreducible. La función del Código D.7 determina cuales son las clases de comunicación de una cadena de Markov, como resultado nos muestra que para la cadena de Ehrenfest existe una única clase de comunicación

$$\{0, 1, 2, 3\},$$

con esto podemos concluir numéricamente que esta cadena es irreducible.

Con la función del Código D.8 es posible calcular el periodo de una cadena de Markov irreducible, como

resultado nos muestra que la cadena de Ehrenfest tiene periodo 2, este resultado coincide con las conclusiones del Ejemplo 2.10.

Siguiendo con nuestro análisis la función del Código D.9 nos muestra cuales son los estados recurrentes de una cadena de Markov dada, en el caso de la cadena de Ehrenfest nos muestra que todos los estados son recurrentes.

En el Ejemplo 2.10 calculamos analíticamente la única distribución estacionaria de la cadena de Ehrenfest, para el caso  $k = 3$  obtenemos

$$\nu = (1/8, 3/8, 3/8, 1/8),$$

mientras que con la función del Código D.10 con un horizonte de tiempo  $m = 10000$  produce una distribución estacionaria aproximada de

$$(0.1247, 0.3787, 0.3753, 0.1213).$$

Propiedad	Resultado	Código utilizado
Clases de comunicación	$\{0,1,2,3\}$	D.7
Irreducible	si	D.7
Periodo	2	D.8
Estados recurrentes	$\{0,1,2,3\}$	D.9
Estados transitorios	$\emptyset$	D.9
Ergódica	si	D.7
Distribución estacionaria	$(0.1247, 0.3787, 0.3753, 0.1213)$	D.10

Cuadro 3.1: Tabla resumen de resultados .

En la figura ?? podemos observar el comportamiento de la convergencia a la distribución estacionaria hasta el horizonte de tiempo  $m = 100000$ , para este análisis se implementó la función del Código D.11

## 3.2. Procesos de saltos de Markov

En esta sección daremos una justificación formal para la simulación de procesos de Markov en tiempo continuo con espacio de estados finito  $E = \{1, \dots, N\}$ .

Anteriormente estudiamos los procesos de saltos de Markov, ahora nos interesa simular trayectorias de dichos procesos, para ello debemos primero mostrar su existencia, debido a la Nota 2.9 podemos plantearnos como hacer la construcción de un PSM.

### 3.2.1. Algoritmo para PSM

**Definición 3.2.** Definamos  $S_0 = 0$  y  $\{S_n\}_{n \geq 1}$  variables aleatorias i.i.d. con distribución exponencial de parámetro 1 y  $\{\lambda_x\}_{x \in E}$  números estrictamente positivos.

Tomemos también  $\{Z_n\}_{n \in \mathbb{N}}$  una cadena de Markov con espacio de estados  $E$  independiente de la sucesión de variables  $\{S_n\}_{n \geq 0}$ , definimos el proceso en tiempo continuo dado por :

$$X_t = \sum_{n=0}^{\infty} Z_n \mathbb{I} \left\{ \sum_{j=0}^n \frac{S_j}{\lambda_{Z_{j-1}}} \leq t < \sum_{j=0}^{n+1} \frac{S_j}{\lambda_{Z_{j-1}}} \right\}. \quad (3.2)$$

Nos preguntamos si este nuevo proceso es un proceso PSM, la respuesta es afirmativa y es parte del siguiente Teorema.

**Teorema 3.2. (Existencia de PSM).**

Sea  $Q = (q_{xy})_{x,y \in E}$  una matriz con entradas  $q_{xy} = \lambda_x p_{xy}$  si  $x \neq y$  y  $q_{xx} = -\lambda_x$ , en donde  $\{\lambda_x\}_{x \in E}$  son números mayores que cero y para todo  $x \in E$   $\{p_{xy}\}_{y \in E}$  conforma una distribución sobre  $E$  tal que  $p_{xx} = 0$ , permitir además ser a  $\pi \in \mathbb{R}^N$  una distribución inicial en  $E$ .

Bajo estas condiciones existe un espacio de probabilidad, donde está definido un PSM  $\{X_t\}_{t \geq 0}$  con espacio de estados  $E$ , con distribución inicial  $\pi$  y cuyo generador infinitesimal está dado por  $Q$ .

*Demostración.* Definamos el proceso estocástico  $\{X_t\}_{t \geq 0}$  como en la Definición 3.2 en donde la cadena de Markov  $\{Z_n\}_{n \in \mathbb{N}}$  tiene matriz de transición  $P = \{p_{xy}\}_{x,y \in E}$  y distribución inicial  $\pi$ .

Para demostrar la afirmación basta con demostrar que para  $x_0, x_1, \dots, x_{n+1} \in E$  y para cualesquiera  $t_i < t_{i+1} = s_{i+1} + t_i$  con  $i \in \{0, \dots, n\}$ , con  $t_0 = 0$

$$\mathbb{P}[X_{t_{n+1}} = x_{n+1} | X_{t_n} = x_n, \dots, X_{t_0} = x_0] = \mathbb{P}[X_{t_{n+1}} = x_{n+1} | X_{t_n} = x_n].$$

Definamos  $N_t = \sum_{n=1}^{\infty} \mathbb{I}_{\{T_n \leq t\}}$  el número de saltos que el proceso ha realizado hasta el instante  $t$ , con  $T_n = \sum_{j=0}^n S_j / \lambda_{Z_{j-1}}$  si  $n \geq 0$ , entonces

$$\mathbb{P}[X_{t_{n+1}} = x_{n+1} | X_{t_n} = x_n, \dots, X_{t_0} = x_0] = \sum_{m=0}^{\infty} [a_m + b_m] \mathbb{P}[N_{t_n} = m | X_{t_n} = x_n, \dots, X_{t_0} = x_0]. \quad (3.3)$$

En donde

$$\begin{aligned} a_m &= \mathbb{P}[X_{t_n + s_{n+1}} = x_{n+1}, T_{m+1} > t_n + s_{n+1} | F_m] \\ b_m &= \mathbb{P}[X_{t_n + s_{n+1}} = x_{n+1}, T_{m+1} \leq t_n + s_{n+1} | F_m] \\ F_m &= \{N_{t_n} = m, X_{t_n} = x_n, \dots, X_{t_0} = x_0\}. \end{aligned}$$

Debido a que  $S_{m+1}$  es independiente a  $\{(Z_k, T_k)\}_{k \leq m}$  y por la propiedad de pérdida de memoria de la exponencial se tienen las siguientes igualdades.

$$\begin{aligned} a_m &= \mathbb{I}_{\{x_{n+1} = x_n\}} \mathbb{P}[T_{m+1} > t_n + s_{n+1} | N_{t_n} = m, X_{t_n} = x_n, \dots, X_{t_0} = x_0] \\ &= \mathbb{I}_{\{x_{n+1} = x_n\}} \mathbb{P}[T_{m+1} > t_n + s_{n+1} | T_{m+1} > t_n \geq T_m, Z_m = x_n, X_{t_{n-1}} = x_{n-1}, \dots, X_{t_0} = x_0] \\ &= \mathbb{I}_{\{x_{n+1} = x_n\}} \mathbb{P}[S_{m+1} / \lambda_{x_n} > (t_n - T_m) + s_{n+1} | S_{m+1} / \lambda_{x_n} > (t_n - T_m), (t_n - T_m) \geq 0] \\ &= \mathbb{I}_{\{x_{n+1} = x_n\}} \exp(-\lambda_{x_n} s_{n+1}). \end{aligned}$$

Por otro lado condicionando respecto a  $T_{m+1}$  y respecto a  $X_{T_{m+1}} = Z_{m+1}$  y haciendo un par de desarrollos se tiene:

$$\begin{aligned} b_m &= \sum_{k \neq x_n} \int_0^{s_{n+1}} f_{X_{t_n + s_{n+1}} | T_{m+1}, Z_{m+1}, N_{t_n}, X_{t_n}, \dots, X_{t_0}}(x_{n+1} | t_n + v, k, m, x_n, \dots, x_0) \\ &\quad \times f_{Z_{m+1}, T_{m+1} | N_{t_n}, X_{t_n}, \dots, X_{t_0}}(k, t_n + v | m, x_n, \dots, x_0) dv. \end{aligned}$$

Por construcción del proceso  $\{X_t\}_{t \geq 0}$ , se sabe que

$$f_{X_{t_n + s_{n+1}} | T_{m+1}, Z_{m+1}, N_{t_n}, X_{t_n}, \dots, X_{t_0}}(x_{n+1} | t_n + v, k, m, x_n, \dots, x_0) = \mathbb{P}[X_{s_{n+1} - v} = x_{n+1} | X_0 = k] = p_{kx_n}(s_{n+1} - v).$$

Por un argumento análogo al que se utilizó para calcular  $a_m$  y por la independencia que existe entre la cadena de Markov  $\{Z_n\}_{n \in \mathbb{N}}$  y la sucesión  $\{S_n\}_{n \in \mathbb{N}}$  obtenemos

$$f_{Z_{m+1}, T_{m+1} | N_{t_n}, X_{t_n}, \dots, X_{t_0}}(k, t_n + v | m, x_n, \dots, x_0) = \lambda_{x_n} \exp(-\lambda_{x_n} v) p_{x_n k}.$$

Uniendo estas igualdades podemos al fin calcular fácilmente  $b_m$ .

$$b_m = \sum_{k \neq x_n} \int_0^{s_{n+1}} p_{kx_n}(s_{n+1} - v) * \lambda_{x_n} \exp(-\lambda_{x_n} v) p_{x_n k} dv.$$

Uniendo los cálculos anteriormente hechos y notando que  $a_m$  y  $b_m$  son independientes de  $m$  y de  $t_n$  se tiene que (3.3) es igual a:

$$\mathbb{I}_{\{x_{n+1}=x_n\}} \exp(-\lambda_{x_n} s_{n+1}) + \sum_{k \neq x_n} \int_0^{s_{n+1}} p_{kx_n}(s_{n+1} - v) * \lambda_{x_n} \exp(-\lambda_{x_n} v) p_{x_n k} dv.$$

Cuya cantidad no depende del tiempo anterior a  $t_n$ , ni tampoco depende de  $t_n$  y haciendo un cálculo sencillo se tiene además que

$$\mathbb{P}[X_{s_{n+1}} = x_{n+1} | X_0 = x_n]$$

es igual a

$$\mathbb{I}_{\{x_{n+1}=x_n\}} \exp(-\lambda_{x_n} s_{n+1}) + \sum_{k \neq x_n} \int_0^{s_{n+1}} p_{kx_n}(s_{n+1} - v) * \lambda_{x_n} \exp(-\lambda_{x_n} v) p_{x_n k} dv.$$

Por lo que concluimos

$$\mathbb{P}[X_{t_{n+1}} = x_{n+1} | X_{t_n} = x_n, \dots, X_{t_0} = x_0] = \mathbb{P}[X_{s_{n+1}} = x_{n+1} | X_0 = x_n].$$

Por lo anterior afirmamos que es un proceso de Markov continuo por la derecha y por los Teoremas 2.6, 2.7 y 2.8 se sigue que el proceso anteriormente construido posee la distribución  $\pi$  y generador infinitesimal  $Q$ .  $\square$

**Definición 3.3. (Construcción 2 de un PSM).** Supongamos las mismas hipótesis que en el Teorema 3.2. Sean  $\{S_{i,0} = 0\}_{i \in E}$ ,  $\{S_{i,n}\}_{i \in E, n \geq 1}$  variables aleatorias i.i.d. con distribución exponencial de parámetro 1 y  $\{\lambda_x\}_{x \in E}$  números estrictamente positivos, consideremos también  $\{Z_n\}_{n \in \mathbb{N}}$  una cadena de Markov con espacio de estados  $E$ , matriz de transición  $P$  y distribución inicial  $\pi$ , dicha cadena como en la Definición 3.1, e independiente de la sucesión de variables  $\{S_{i,n}\}_{i \in E, n \geq 1}$ , definimos el proceso en tiempo continuo dado por

$$X_t = \sum_{n=1}^{\infty} Z_n \mathbb{I} \left\{ \sum_{j=1}^n \frac{S_{Z_{j-1}, n_{Z_{j-1}}(j)}}{\lambda_{Z_{j-1}}} \leq t < \sum_{j=1}^{n+1} \frac{S_{Z_{j-1}, n_{Z_{j-1}}(j)}}{\lambda_{Z_{j-1}}} \right\} + Z_0 \mathbb{I} \left\{ 0 \leq t < \frac{S_{Z_0, 1}}{\lambda_{Z_0}} \right\}. \quad (3.4)$$

Donde  $n_i(m)$  es la variable asociada al número de transiciones de  $i$  a cualquier estado hasta el tiempo  $m$  asociada a la cadena  $\{Z_n\}_{n \in \mathbb{N}}$ .

Análogamente como en el Teorema 3.2 se demuestra que  $\{X_t\}_{t \geq 0}$  es un PSM con los mismos parámetros dados en este teorema.

El Teorema 3.2 da lugar al siguiente algoritmo para simular un PSM.

**Algoritmo 3.2. PSM.** Este algoritmo genera trayectorias de un PSM con distribución inicial  $\pi$  y generador infinitesimal  $Q$ , hasta el horizonte de tiempo  $T$ .

1. Generar  $Z_0$  de la distribución inicial  $\pi$  y hacer  $i = 0$ ,  $T_0 = 0$ .
2. Dado que se está en el estado  $Z_i$ , simular  $T_{i+1}$  v.a. exponencial de parámetro  $\lambda_{Z_i}$ .

3. Hacer  $X_t = Z_i$  si  $T_i \leq t < T_{i+1}$ .
4. Simular una v.a.  $Z_{i+1}$  de la distribución de salto  $\{f_{X_{T_1}|X_0}(y|Z_i)\}_{y \in E}$  independientemente del tiempo en el que ocurrió.
5. Hacer  $i = i + 1$ .
6. Si  $T_i < T$  volver al paso 2, de lo contrario ir al paso 7.
7. Tomar la trayectoria del proceso solamente en el intervalo  $[0, T]$ , en donde  $T$  es el horizonte de tiempo.

**Ejemplo 3.3.** Se simuló una trayectoria de un PSM usando el Algoritmo 3.2 con la función del Código D.4 en el software Julia con las siguientes características:

Con generador infinitesimal:

$$Q = \begin{pmatrix} -.13 & .04 & .01 & .08 \\ .02 & -.10 & .04 & .04 \\ .2 & .1 & -.6 & .3 \\ .05 & .15 & .1 & -.3 \end{pmatrix}$$

vector de espacio de estados  $E$ :

$$(1 \ 2 \ 3 \ 4)$$

vector de distribución inicial  $\pi$ :

$$(.2 \ .2 \ .1 \ .5).$$

Con horizonte de tiempo  $T=50$ .

En la figura ?? podemos observar la gráfica de una trayectoria simulada.

### 3.2.2. Cálculo numérico de propiedades de PSM

A continuación aplicaremos técnicas de simulación desarrolladas anteriormente para realizar cálculos numéricos sobre propiedades de los procesos de saltos de Markov

**Ejemplo 3.4.** Considere la cadena de Markov del Ejemplo 2.11 con  $k=3$  y  $\lambda_0 = 1, \lambda_1 = 2, \lambda_2 = 1, \mu_1 = 2, \mu_2 = 1, \mu_3 = 2$  por lo anterior el espacio de estados es  $E = \{0, 1, 2, 3\}$  y el generador infinitesimal es

$$Q = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 2 & -4 & 2 & 0 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 2 & -2 \end{pmatrix},$$

mientras que la matriz de transición asociada a la cadena de saltos es

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Por otro lado implementando la función del Código [D.12](#)

$$P(t) = \tilde{D}^{-1} \exp(Wt) \tilde{D},$$

donde

$$\tilde{D}^{-1} = \begin{pmatrix} 0.210527 & 0.2 & 0.5 & -0.640487 \\ -0.891808 & -0.4 & 0.5 & -0.151198 \\ 0.34064 & -0.4 & 0.5 & 0.395843 \\ -0.210527 & 0.8 & 0.5 & 0.640487 \end{pmatrix},$$

$$\exp(Wt) = \begin{pmatrix} \exp(-5.23607t) & 0.0 & 0.0 & 0.0 \\ 0.0 & \exp(-3t) & 0.0 & 0.0 \\ 0.0 & 0.0 & \exp((1.53006e - 19)t) & 0.0 \\ 0.0 & 0.0 & 0.0 & \exp(-0.763932t) \end{pmatrix},$$

$$\tilde{D} = \begin{pmatrix} -0.362866 & -0.768563 & 0.58713 & -0.181433 \\ 0.333333 & -0.333333 & -0.666667 & 0.666667 \\ 0.666667 & 0.333333 & 0.666667 & 0.333333 \\ -0.817513 & -0.0964944 & 0.505251 & 0.408757 \end{pmatrix}.$$

La distribución estacionaria esta dada por

$$(1/3, 1/6, 1/3, 1/6),$$

mientras que la aproximación implementando la función del Código [D.15](#) con horizonte de tiempo  $T = 10000$  es

$$(0.332264, 0.16706, 0.332738, 0.16797).$$

Propiedad	Resultado	Código utilizado
Clases de comunicación	$\{0,1,2,3\}$	D.13
Irreducible	si	D.13
Estados recurrentes	$\{0,1,2,3\}$	D.14
Estados transitorios	$\emptyset$	D.14
Ergódica	si	D.13
Distribución estacionaria	(0.332264, 0.16706, 0.332738, 0.16797)	D.15

Cuadro 3.2: Tabla resumen de resultados .

En la figura ?? podemos observar el comportamiento de la convergencia a la distribución estacionaria hasta el horizonte de tiempo  $T = 100000$ , para este análisis se implementó la función del código D.16.

### 3.2.3. Ejemplo. Modelo SIR-PSM

El modelo SIR (Susceptibles-Infectedos-Recuperados) es un modelo de compartimentos donde la población bajo estudio se divide en clases epidemiológicas y se describe un flujo entre ellas.

Para poder modelar estocásticamente el fenómeno consideremos las variables aleatorias discretas:

$$S(t) = \text{número de suceptibles a tiempo } t$$

$$I(t) = \text{número de infectados}$$

Como el modelo no considera muertes consideraremos una población con  $N$  habitantes, es decir  $S(t), I(t) \in \{0, 1, 2, \dots, N\}$  con  $t \in [0, \infty)$ .

Para entender el modelo como uno de PSM, necesitamos encontrar las probabilidades de transición. Denotamos a las as probabilidades de transición asociadas al proceso estocástico como:

$$p_{(s,i),(s+k,i+j)}(\Delta t) = P(S(t+\Delta t), I(t+\Delta t) = (s+k, i+j) \mid S(t), I(t) = (s, i))$$

Las cuales dependen del tiempo entre los eventos, es decir de  $\Delta t$  y no específicamente del tiempo  $t$  así tenemos un proceso de tiempo homogéneo. Además dado el estado actual al tiempo  $t > 0$ , el estado futuro al tiempo  $t+\Delta t$  no depende de tiempos anteriores a  $t$ , por lo que cumple la propiedad de Markov; entonces consideraremos las probabilidades de transición definidas de la siguiente manera:

$$p_{(s,i),(s+k,i+j)}(\Delta t) = \begin{cases} \beta i \frac{S}{N} \Delta t + o(\Delta t) & \text{si } (k, j) = (-1, +1), \\ \gamma i \Delta t + o(\Delta t) & \text{si } (k, j) = (0, -1), \\ 1 - (\beta i \frac{S}{N} + \gamma i) \Delta t + o(\Delta t) & \text{si } (k, j) = (0, 0), \\ o(\Delta t) & \text{e.o.c.} \end{cases}$$

Donde  $\beta$  es la tasa de transmisión y  $\gamma$  la tasa de recuperación.

En resumen mostraremos los cambios asociados con los eventos de infección y recuperación en la siguiente tabla.

Supuestos del modelo SIR		
Evento	Cambio ( $\Delta S, \Delta T$ )	Probabilidad
Infección	$(-1, +1)$	$\beta i \frac{S}{N} \Delta t + o(\Delta t)$
Recuperación	$(0, -1)$	$\gamma i \Delta t + o(\Delta t)$

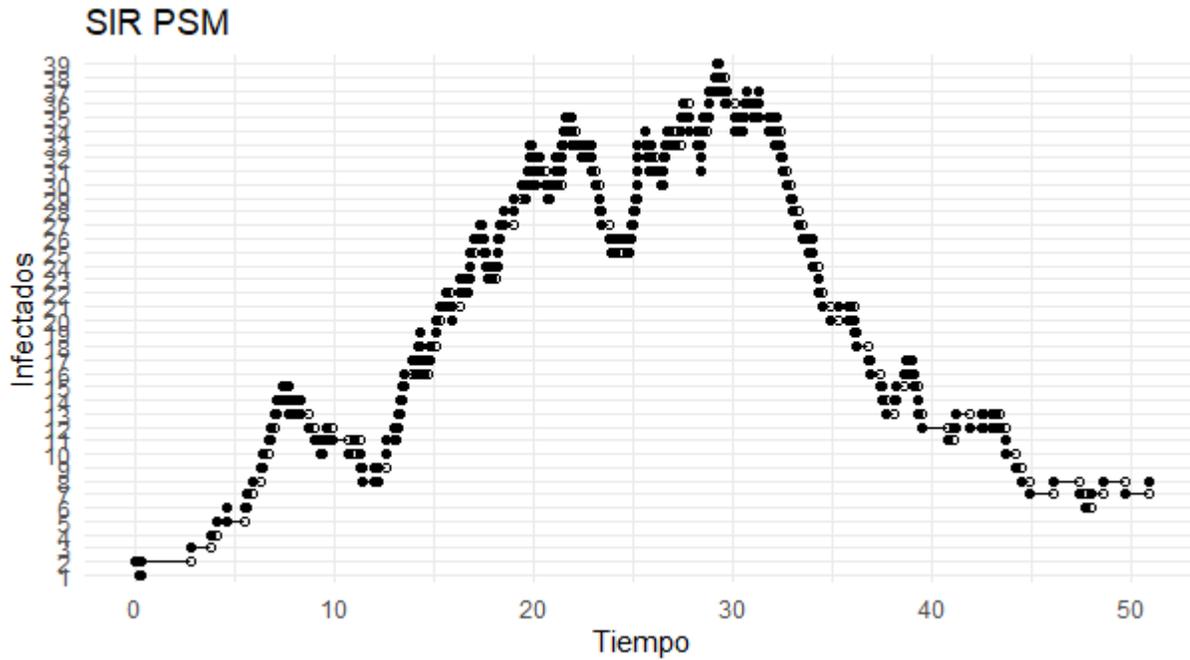


Figura 3.1: Gráfica de la norma entre  $P$  y  $\hat{P}(n)$  respecto al tiempo

Tamaño de la población  $N=200$  Tiempo de observación  $t=50$  tasa de infección  $\beta=0.30$  tasa de recuperación  $\gamma=0.15$  condiciones iniciales  $S_0 = N-2$   $I_0 = 2$   $R_0 = 0$

# Capítulo 4

## Inferencia sobre procesos de Markov

A continuación estudiaremos la estimación de los parámetros de trayectorias de cadenas de Markov, así como también demostraremos algunas características estadísticas importantes sobre los estimadores e implementaremos ejemplos numéricos donde se ilustren los principales resultados. La bibliografía básica para este capítulo es principalmente [? ].

### 4.1. Cadenas de Markov

#### 4.1.1. Estimación máximo verosímil

Supóngase que se tiene observada una trayectoria de una cadena de Markov con espacio de estados  $E$ , nuestro problema ahora es tratar de estimar las probabilidades de transición de la cadena mediante el método de máxima verosimilitud. La siguiente proposición nos proporciona las condiciones para la existencia y unicidad del estimador máximo verosímil para la matriz de transición de una cadena de Markov homogénea con espacio de estados finito.

**Proposición 4.1.** *Dada una trayectoria observada  $\{x_n\}_{n \in \{0,1,\dots,m\}}$  de una cadena de Markov homogénea e irreducible  $X = \{X_n\}_{n \in \mathbb{N}}$  con espacio de estados  $E = \{1, 2, \dots, N\}$  y matriz de transición  $P$ , el estimador máximo verosímil de la matriz de transición está dado por*

$$\hat{P}(m) = \{\hat{p}_{ij}(m) = \frac{n_{ij}(m)}{n_i(m)}\}_{i,j \in E}.$$

Donde  $m$  denota el horizonte de tiempo,  $n_{ij}(m)$  es el número de veces en el que la trayectoria pasó del estado  $i$  al  $j$  en un paso y  $n_i(m)$  es el número de visitas al estado  $i$  hasta el tiempo  $m-1$ , y está dado por  $\sum_{j=1}^N n_{ij}(m)$  para todo  $i \in E$ .

**Nota 4.1.** *Una observación que vale la pena mencionar es que la existencia de este estimador está condicionada a que  $n_i(m) > 0$  para todo  $i \in E$ , que es equivalente a pedir que hasta el tiempo  $m-1$  el proceso haya visitado todos los estados, lo cual se logra debido a la irreducibilidad.*

*Demostración.* Supongamos que tenemos observada la trayectoria  $\{X_0 = x_0, X_1 = x_1, \dots, X_m = x_m\}$

entonces la función de verosimilitud está dada por:

$$L(P; \{x_n\}_{n=0}^m) = \mathbb{P}[X_0 = x_0, X_1 = x_1, \dots, X_m = x_m] = \mathbb{P}[X_0 = x_0] \prod_{i=1}^m \prod_{j=1}^N p_{ij}^{n_{ij}(m)}.$$

Entonces la correspondiente log-verosimilitud es:

$$L(P; \{x_n\}_{n=0}^m) = \log(\mathbb{P}[X_0 = x_0]) + \sum_{i=1}^N \sum_{j=1}^N n_{ij}(m) \log(p_{ij}).$$

El espacio parametral está dado por

$$\mathcal{P} = \{P \in \mathcal{M}_{N \times N} : p_{ij} \geq 0 \text{ y } \sum_{j=1}^N p_{ij} = 1 \text{ para todo } i \in E\}.$$

Ahora maximizaremos la log-verosimilitud en función de las probabilidades de transición mediante multiplicadores de Lagrange, con las restricciones :  $\sum_{j=1}^N p_{i,j} = 1$  para todo  $i \in E$ .

Así planteamos el problema de maximización

$$H(P) = \log(\mathbb{P}[X_0 = y_0]) + \sum_{i=1}^N \sum_{j=1}^N n_{ij}(m) \log(p_{ij}) + \sum_{i=1}^N \lambda_i (1 - \sum_{j=1}^N p_{ij}).$$

Derivando con respecto a  $p_{ij}$  e igualando a cero tenemos las siguientes ecuaciones:

$$\frac{n_{ij}(m)}{p_{ij}} - \lambda_i = 0 \text{ para todo } i, j \in E. \quad (4.1)$$

Entonces

$$p_{ij} = \frac{n_{ij}(m)}{\lambda_i} \text{ para todo } i, j \in E.$$

Sumando sobre  $j$  para todo  $i \in E$  se tiene

$$1 = \sum_{j=1}^N p_{ij} = \frac{n_i(m)}{\lambda_i} \text{ para todo } i \in E,$$

por tanto

$$\lambda_i = n_i(m) \text{ y } \hat{p}_{ij}(m) = \frac{n_{ij}(m)}{n_i(m)} \text{ para todo } i, j \in E.$$

Hasta aquí se vió que  $\{\hat{p}_{ij}(m) = n_{ij}(m)/n_i(m)\}_{i,j \in E}$  es un punto crítico de  $H$  en  $\mathcal{P}$ . Ahora queremos ver que en este punto hay un máximo. Sabemos que  $L(P; \{x_n\}_{n=0}^m)$  es una función continua y el espacio parametral anteriormente definido es un conjunto compacto en el espacio euclidiano, por esto existe  $\hat{P}(m)$  tal que  $L(P; \{x_n\}_{n=0}^m) \leq L(\hat{P}(m); \{x_n\}_{n=0}^m)$  y  $L(\hat{P}(m); \{x_n\}_{n=0}^m) > 0$ , entonces se tiene para cada  $P$  en el espacio parametral y donde esté bien definida la log-verosimilitud

$$l(\hat{P}(m); \{x_n\}_{n=0}^m) \leq l(\hat{P}(m); \{x_n\}_{n=0}^m).$$

Por el Teorema del multiplicador de Lagrange (ver [?] página 414)  $\hat{P}(m)$  satisface las ecuaciones (4.1) y por lo anteriormente mostrado se tiene

$$\hat{P}(m) = \{\hat{p}_{ij}(m) = \frac{n_{ij}(m)}{n_i(m)}\}_{i,j \in E},$$

es el estimador máximo verosímil. □

**Ejemplo 4.1.** *Estimación por máxima verosimilitud de la matriz de transición correspondiente a la trayectoria simulada en el Ejemplo 3.1, la función que se utilizó se encuentra en el Código D.2.*

$$\hat{P}(m) = \begin{pmatrix} 0.181818 & 0.0909091 & 0.636364 & 0.0909091 \\ 0.454545 & 0.181818 & 0.181818 & 0.181818 \\ 0.166667 & 0.416667 & 0.166667 & 0.25 \\ 0.333333 & 0.5 & 0.166667 & 0.0 \end{pmatrix}.$$

Ahora mostraremos que el estimador máximo verosímil de la matriz de transición cumple propiedades estadísticas muy importantes como son consistencia y eficiencia asintótica.

Denotemos por  $\nu = (\nu_1, \dots, \nu_N)$  a la distribución estacionaria (si existe) de la cadena de Markov  $\{X_n\}_{n \geq 0}$ .

**Nota 4.2.** *Una observación que vale la pena mencionar es que si la cadena  $\{X_n\}_{n \in \mathbb{N}}$  es irreducible y recurrente positiva entonces  $\nu$  existe y  $\nu_i$  es estrictamente positiva para todo  $i \in E$ .*

#### 4.1.2. Propiedades asintóticas

**Teorema 4.1. (consistencia).** *Dada una Cadena de Markov irreducible y recurrente positiva con espacio de estados  $E$  y matriz de transición  $P$ , los estimadores  $\hat{p}_{ij}(m) = \frac{n_{ij}(m)}{n_i(m)}$  convergen casi seguramente al verdadero parámetro  $p_{ij}$  cuando  $m$  tiende a infinito para todo  $i, j \in E$ . Más aún*

$$\|P - \hat{P}(m)\| \xrightarrow[m \rightarrow \infty]{c.s.} 0.$$

Donde  $\|*\|$  denota la norma usual en el espacio euclidiano. En otras palabras el estimador  $\hat{P}(m)$  es consistente.

*Demostración.* Podemos escribir a nuestro estimador de la siguiente forma:

$$\hat{p}_{ij}(m) = \frac{n_{ij}(m)}{n_i(m)} = \frac{\sum_{k=1}^{n_i(m)} \mathbb{I}_{\{X_{1+\tau_k^i}=j\}} + \mathbb{I}_{\{X_0=i, X_1=j\}}}{n_i(m)}.$$

Por el Teorema 2.1 sabemos que  $\{\mathbb{I}_{\{X_{1+\tau_k^i}=j\}}\}_{k \geq 0}$  son i.i.d. Sabemos que la cadena  $\{X_n\}_{n \geq 0}$  es recurrente positiva, lo que implica que  $\lim_{m \rightarrow \infty} n_i(m) = \infty$  y aplicando la ley fuerte de los grandes números obtenemos

$$\lim_{m \rightarrow \infty} \hat{p}_{ij}(m) = \mathbb{P}[X_{1+\tau_1^i} = j].$$

Por otro lado sabemos que  $\{X_n\}_{n \geq 0}$  es recurrente positiva, lo que implica  $\mathbb{P}[\tau_1^i < \infty] = 1$ . Por la propiedad de Markov y debido a que  $\tau_1^i$  es un tiempo de paro tenemos:

$$\begin{aligned} \mathbb{P}[X_{1+\tau_1^i} = j] &= \sum_{k=1}^{\infty} \mathbb{P}[X_{1+\tau_1^i} = j, \tau_1^i = k] \\ &= \sum_{k=1}^{\infty} \mathbb{P}[X_{1+k} = j, \tau_1^i = k] \\ &= p_{ij} \sum_{k=1}^{\infty} \mathbb{P}[\tau_1^i = k] \\ &= p_{ij}. \end{aligned}$$

Por lo tanto

$$\|P - \hat{P}(m)\| \xrightarrow[m \rightarrow \infty]{c.s.} 0.$$

□

**Ejemplo 4.2.** Se simuló una trayectoria de la cadena de Markov del Ejemplo 3.1 con horizonte  $m=10000$ . En la figura 4.1 se muestra  $\|\hat{P}(n) - P\|$  para cada  $n \leq 10000$ , para este análisis se utilizó el Código D.3 .

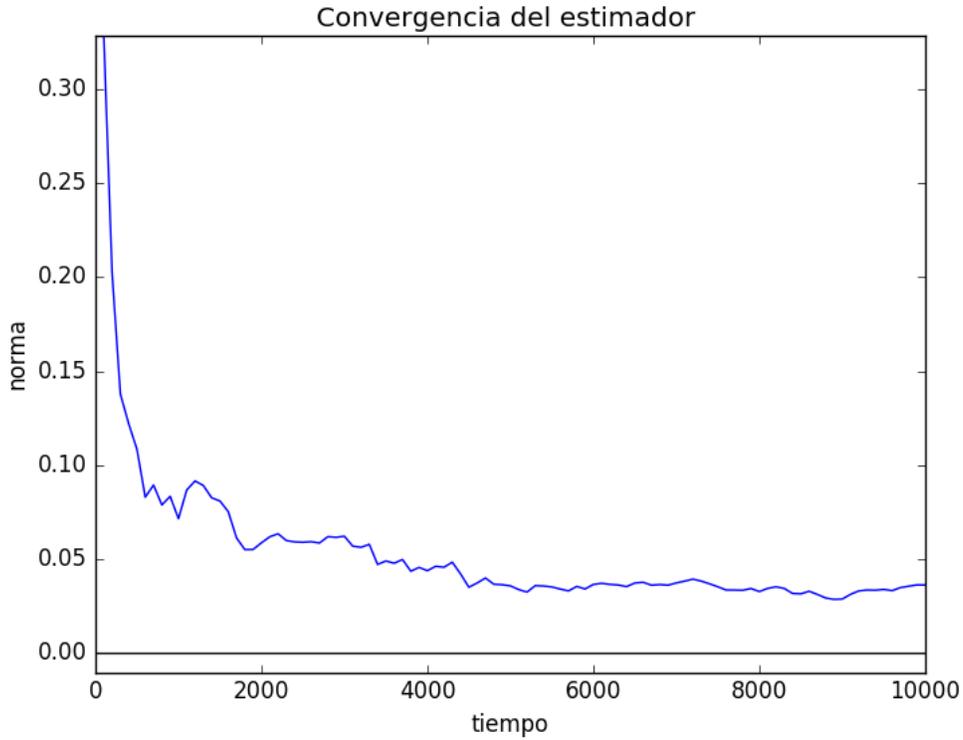


Figura 4.1: Gráfica de la norma entre  $P$  y  $\hat{P}(n)$  respecto al tiempo

Ahora nos concentraremos en demostrar que el estimador máximo verosímil de la matriz de transición  $\hat{P}(m)$  tiene la propiedad de eficiencia asintótica, esta propiedad nos ayudará a hacer estimación por intervalo de la matriz de transición de una cadena de Markov.

Para lograr este objetivo primero debemos demostrar el siguiente lema.

**Lema 4.1.** Sean  $\{(g_{i1}(m), \dots, g_{iN}(m))\}_{1 \leq i \leq N}$  una familia independiente de vectores aleatorios sobre  $\mathbb{R}^N$  tal que para cada  $i \in E$

$$(g_{i1}(m), \dots, g_{iN}(m)) \sim \text{multinomial}(m, p_{i1}, \dots, p_{iN}).$$

Entonces

$$G(m) \xrightarrow[m \rightarrow \infty]{d} N(\bar{0}, \Gamma).$$

En donde

$$G(m) = \left\{ \frac{g_{ij}(m) - mp_{ij}}{\sqrt{m}} \right\}_{i,j \in E}$$

y  $\bar{0}$  denota la matriz de tamaño  $N \times N$  con todas las entradas iguales a cero, mientras que la matriz de varianzas y covarianzas tiene dimensión  $N^2 \times N^2$  y está dada por

$$\Gamma = \{\delta_{ik}(\delta_{jl}p_{ij} - p_{ij}p_{il})\}_{i,j,l,k \in E}.$$

En donde

$$\delta_{ik} = \begin{cases} 1 & \text{si } i = k, \\ 0 & \text{si } i \neq k. \end{cases}$$

*Demostración.* Denotemos por  $G_i(m)$  el vector cuyas entradas están dadas por  $G_i(m)_j = \frac{g_{ij}(m) - mp_{ij}}{\sqrt{m}}$  para todo  $i, j \in E$ .

A continuación probaremos que la función generadora de momentos de  $G_i(m)$  converge a la generadora de momentos de un vector aleatorio que se distribuye normal, y así concluir lo que se desea demostrar. Sea  $\bar{t} = (t_1, \dots, t_N) \in \mathbb{R}^N$  y  $\langle *, * \rangle$  que denota el producto interno usual en el espacio euclidiano de dos matrices de igual dimensión.

Si  $Z = (z_1, \dots, z_N) \sim \text{multinomial}(m, p_1, p_2, \dots, p_N)$  su función generadora de momentos está dada por

$$M_Z(\bar{t}) = \left[ \sum_{k=1}^N p_k \exp(t_k) \right]^m.$$

Por tanto

$$\mathbb{E}[\exp(\langle \bar{t}, G_i(m) \rangle)] = \left( \sum_{j=1}^N p_{ij} \exp \left[ \frac{1}{\sqrt{m}} \left( t_j - \sum_{k=1}^N t_k p_{ik} \right) \right] \right)^m.$$

Esta última igualdad se deduce fácilmente de la función generadora de momentos de una multinomial de parámetros  $(m, p_{i1}, \dots, p_{iN})$ .

Desarrollando la serie de Taylor de la función generadora de momentos  $\mathbb{E}[\exp(\langle \bar{t}, G_i(m) \rangle)]$  en el punto  $(0, \dots, 0) \in \mathbb{R}^N$  tenemos que

$$\begin{aligned} \sum_{j=1}^N p_{ij} \exp \left[ \frac{1}{\sqrt{m}} \left( t_j - \sum_{k=1}^N t_k p_{ik} \right) \right] &= 1 + \sum_{k_1=1}^N \sum_{k_2=1}^N t_{k_1} t_{k_2} \sum_{j=1}^N \frac{p_{ij}}{2m} (\delta_{k_1 j} - p_{ik_1})(\delta_{k_2 j} - p_{ik_2}) \\ + \sum_{k_1=1}^N \sum_{k_2=1}^N \sum_{k_3=1}^N t_{k_1} t_{k_2} t_{k_3} \sum_{j=1}^N \frac{p_{ij}}{6m^{3/2}} \exp \left[ \frac{1}{\sqrt{m}} \left( t_{m,j} - \sum_{k=1}^N t_{m,k} p_{ik} \right) \right] & (\delta_{k_1 j} - p_{ik_1})(\delta_{k_2 j} - p_{ik_2})(\delta_{k_3 j} - p_{ik_3}). \end{aligned}$$

Donde  $(t_{m,1}, \dots, t_{m,N}) = \alpha_m \bar{t}$  para alguna  $\alpha_m \in (0,1)$  y las derivadas parciales de  $\mathbb{E}[\exp(\langle \bar{t}, G_i(m) \rangle)]$  están dadas por

$$\begin{aligned} \frac{\partial}{\partial t_{k_1}} \sum_{j=1}^N p_{ij} \exp \left[ \frac{1}{\sqrt{m}} \left( t_j - \sum_{k=1}^N t_k p_{ik} \right) \right] &= \sum_{j=1}^N p_{ij} \exp \left[ \frac{1}{\sqrt{m}} \left( t_j - \sum_{k=1}^N t_k p_{ik} \right) \right] \frac{1}{\sqrt{m}} (\delta_{jk_1} - p_{ik_1}) = 0. \\ \frac{\partial^2}{\partial t_{k_2} \partial t_{k_1}} \sum_{j=1}^N p_{ij} \exp \left[ \frac{1}{\sqrt{m}} \left( t_j - \sum_{k=1}^N t_k p_{ik} \right) \right] & \end{aligned}$$

$$= \sum_{j=1}^N p_{ij} \exp \left[ \frac{1}{\sqrt{m}} \left( t_j - \sum_{k=1}^N t_k p_{ik} \right) \right] \frac{1}{\sqrt{m}} (\delta_{jk_1} - p_{ik_1}) \frac{1}{\sqrt{m}} (\delta_{jk_2} - p_{ik_2}).$$

y por último

$$\begin{aligned} & \frac{\partial^3}{\partial t_{k_3} \partial t_{k_2} \partial t_{k_1}} \sum_{j=1}^N p_{ij} \exp \left[ \frac{1}{\sqrt{m}} \left( t_j - \sum_{k=1}^N t_k p_{ik} \right) \right] \\ &= \sum_{j=1}^N p_{ij} \exp \left[ \frac{1}{\sqrt{m}} \left( t_j - \sum_{k=1}^N t_k p_{ik} \right) \right] \frac{1}{\sqrt{m}} (\delta_{jk_1} - p_{ik_1}) \frac{1}{\sqrt{m}} (\delta_{jk_2} - p_{ik_2}) \frac{1}{\sqrt{m}} (\delta_{jk_3} - p_{ik_3}). \end{aligned}$$

Es fácil darse cuenta que el último sumando de la serie de Taylor de la función  $\mathbb{E}[\exp(\langle \bar{t}, G_i(m) \rangle)]$  en valor absoluto puede ser acotado por  $\frac{M}{m^{3/2}}$  para alguna  $M > 0$ , por lo anterior concluimos que para toda  $m > 0$  suficientemente grande se tiene

$$\begin{aligned} & \left( 1 + \sum_{k_1=1}^N \sum_{k_2=1}^N t_{k_1} t_{k_2} \sum_{j=1}^N \frac{p_{ij}}{2m} (\delta_{k_1 j} - p_{ik_1}) (\delta_{k_2 j} - p_{ik_2}) - \frac{M}{m^{3/2}} \right)^m \leq \mathbb{E}[\exp(\langle \bar{t}, G_i(m) \rangle)] \\ & \leq \left( 1 + \sum_{k_1=1}^N \sum_{k_2=1}^N t_{k_1} t_{k_2} \sum_{j=1}^N \frac{p_{ij}}{2m} (\delta_{k_1 j} - p_{ik_1}) (\delta_{k_2 j} - p_{ik_2}) + \frac{M}{m^{3/2}} \right)^m. \end{aligned}$$

Haciendo  $m \rightarrow \infty$  se tiene que

$$\mathbb{E}[\exp(\langle \bar{t}, G_i(m) \rangle)] \rightarrow \exp \left[ \sum_{k_1=1}^N \sum_{k_2=1}^N t_{k_1} t_{k_2} \sum_{j=1}^N \frac{p_{ij}}{2} (\delta_{k_1 j} - p_{ik_1}) (\delta_{k_2 j} - p_{ik_2}) \right].$$

Este último término se simplifica en

$$\exp \left[ \sum_{k_1=1}^N \sum_{k_2=1}^N t_{k_1} t_{k_2} \frac{p_{ik_1}}{2} (\delta_{k_2 k_1} - p_{ik_2}) \right].$$

De donde obtenemos

$$\mathbb{E}[\exp(\langle \bar{t}, G_i(m) \rangle)] \xrightarrow{m \rightarrow \infty} \exp \left[ \frac{1}{2} \bar{t}^T \Sigma^i \bar{t} \right].$$

Donde  $\Sigma^i = \{p_{ik_1} (\delta_{k_2 k_1} - p_{ik_2})\}_{k_1, k_2 \in E}$  y  $\bar{t}^T$  es el vector transpuesto asociado a  $\bar{t}$ , por lo anterior y por la independencia de los vectores  $\{G_i(m)\}_{i \in E}$  podemos afirmar que la matriz  $G(m)$  cuyo  $i$ -ésimo renglón es  $G_i(m)$  tiende en distribución a una normal con media cero y matriz de varianzas y covarianzas

$$\Gamma = \{\delta_{ik} (\delta_{jl} p_{ij} - p_{ij} p_{il})\}_{i, j, l, k \in E}.$$

□

**Teorema 4.2.** Sea  $X = \{X_n\}_{n \in \mathbb{N}}$  una cadena de Markov con matriz de transición  $P = \{p_{ij}\}_{i, j \in E}$  y con espacio de estados  $E$ , suponemos además que es irreducible y recurrente positiva, con distribución estacionaria  $\nu = (\nu_1, \dots, \nu_N)$ , entonces

$$\sqrt{m}(\hat{P}(m) - P) \xrightarrow{m \rightarrow \infty} N(\bar{0}, \Sigma).$$

Donde  $\Sigma$  es una matriz de dimensión  $N^2 \times N^2$  con entradas  $\Sigma_{ij,kl} = \frac{1}{\nu_i} \delta_{ik} (\delta_{jl} p_{ij} - p_{ij} p_{il})$  para todo  $i, j, l, k \in E$ .

*Demostración.* Sin pérdida de generalidad podemos suponer que la cadena  $\{X_n\}_{n \in \mathbb{N}}$  está definida como en la Definición 3.1, pues como se vio anteriormente esta construcción conserva la misma ley.

Se sabe que  $(n_{i1}(m), \dots, n_{iN}(m))$  es el vector de frecuencias de  $(w_{i,1}, \dots, w_{i,n_i(m)})$ . Ahora denotemos por  $(h_{i1}(m), \dots, h_{iN}(m))$  el correspondiente vector de frecuencias de  $(w_{i,1}, \dots, w_{i, \lceil m\nu_i \rceil})$ , es facil darse cuenta que sigue una distribución multinomial de parámetros  $(\lceil m\nu_i \rceil, p_{i1}, \dots, p_{iN})$ .

Además  $(h_{i1}(m), \dots, h_{iN}(m))$  es independiente al vector  $(h_{j1}(m), \dots, h_{jN}(m))$  si  $i \neq j$ .

Por el Lema 4.1 podemos afirmar que

$$H(m) \xrightarrow[m \rightarrow \infty]{d} N(0, \Gamma). \quad (4.2)$$

En donde

$$H(m) = \left\{ \frac{h_{ij}(m) - \lceil m\nu_i \rceil p_{ij}}{\sqrt{\lceil m\nu_i \rceil}} \right\}_{i,j \in E}$$

y la matriz de varianzas y covarianzas está dada por

$$\Gamma = \{\delta_{ik}(\delta_{jl}p_{ij} - p_{ij}p_{il})\}_{i,j,l,k \in E}.$$

Sabemos que

$$\begin{aligned} \sqrt{m}(\hat{P}(m) - P) &= \left\{ \left( \frac{n_{ij}(m) - p_{ij}n_i(m)}{\sqrt{n_i(m)}} \right) \left( \frac{\sqrt{m}}{\sqrt{n_i(m)}} \right) \right\}_{i,j \in E} \\ &= \left\{ \left( \frac{n_{ij}(m) - p_{ij}n_i(m)}{\sqrt{m}} \right) \left( \frac{m}{n_i(m)} \right) \right\}_{i,j \in E} \\ &= \left\{ \left( \frac{n_{ij}(m) - p_{ij}n_i(m)}{\sqrt{m}} - \tilde{H}(m)_{i,j} + \tilde{H}(m)_{i,j} \right) \left( \frac{m}{n_i(m)} \right) \right\}_{i,j \in E}. \end{aligned}$$

Donde

$$\tilde{H}(m) = \left\{ \frac{h_{ij}(m) - \lceil m\nu_i \rceil p_{ij}}{\sqrt{m}} \right\}_{1 \leq i,j \leq N}.$$

Supongamos por el momento que

$$\left( \frac{n_{ij}(m) - p_{ij}n_i(m)}{\sqrt{m}} - \tilde{H}(m)_{i,j} \right) \left( \frac{m}{n_i(m)} \right) \xrightarrow[m \rightarrow \infty]{p} 0 \text{ para todo } i, j \in E.$$

Por lo anterior y por la Proposición B.9 basta con mostrar la convergencia para

$$\left\{ \tilde{H}(m)_{i,j} \left( \frac{m}{n_i(m)} \right) \right\}_{i,j \in E}.$$

Por otro lado

$$\left\{ \tilde{H}(m)_{i,j} \left( \frac{m}{n_i(m)} \right) \right\}_{i,j \in E} = \left\{ H(m)_{i,j} \left( \frac{\sqrt{\lceil m\nu_i \rceil} \sqrt{m}}{\sqrt{n_i(m)} \sqrt{n_i(m)}} \right) \right\}_{i,j \in E},$$

por el Teorema ergódico (Teorema 2.2)

$$\frac{\sqrt{\lceil m\nu_i \rceil}}{\sqrt{n_i(m)}} \xrightarrow[m \rightarrow \infty]{c.s.} 1 \text{ y } \frac{\sqrt{m}}{\sqrt{n_i(m)}} \xrightarrow[m \rightarrow \infty]{c.s.} \frac{1}{\sqrt{\nu_i}},$$

por lo anterior, por (4.2) y por la Proposición B.7 concluimos que basta con mostrar la convergencia para el vector

$$\left\{ H(m)_{i,j} \left( \frac{1}{\sqrt{\nu_i}} \right) \right\}_{i,j \in E}.$$

Debido a que  $H(m)$  tiende en distribución a un vector aleatorio normal, entonces se deduce que

$$\sqrt{m}(\hat{P}(m) - P) \xrightarrow[m \rightarrow \infty]{d} N(\bar{0}, \Sigma).$$

Donde  $\Sigma$  es una matriz de dimensiones  $N^2 \times N^2$ , con  $\Sigma$  la matriz con entradas  $\Sigma_{ij,kl} = \frac{1}{\nu_i} \delta_{ik} (\delta_{jl} p_{ij} - p_{ij} p_{il})$  para todo  $i, j, l, k \in E$ .

Ahora mostraremos que

$$\left( \frac{n_{ij}(m) - p_{ij}n_i(m)}{\sqrt{m}} - \tilde{H}(m)_{i,j} \right) \left( \frac{m}{n_i(m)} \right) \xrightarrow[m \rightarrow \infty]{p} 0.$$

Por la Proposición B.7 y el Teorema ergódico (2.2) basta con probar la convergencia

$$\left( \frac{n_{ij}(m) - p_{ij}n_i(m)}{\sqrt{m}} - \tilde{H}(m)_{i,j} \right) \xrightarrow[m \rightarrow \infty]{p} 0.$$

Sea  $\varepsilon > 0$ .

Por el Teorema ergódico (2.2) podemos afirmar que existe  $k$  tal que para todo  $m > k$  se cumple

$$\mathbb{P}[|n_i(m) - \lceil m\nu_i \rceil| > m\varepsilon^3] < \varepsilon.$$

Denotemos a

$$S_n^{ij} = \sum_{r=1}^n \mathbb{I}_{\{w_{i,r}=j\}} - p_{ij}n.$$

Por lo anterior

$$\begin{aligned} \mathbb{P} \left[ \left| \frac{n_{ij}(m) - p_{ij}n_i(m)}{\sqrt{m}} - \tilde{H}(m)_{i,j} \right| > \varepsilon \right] &\leq \varepsilon + \mathbb{P} \left[ \max\{|S_n^{ij} - S_{\lceil m\nu_i \rceil}^{ij}| : |n - \lceil m\nu_i \rceil| \leq m\varepsilon^3\} > \varepsilon\sqrt{m} \right] \\ &\leq \varepsilon + 2\mathbb{P} \left[ \max\{|S_n^{ij}| : 1 \leq n \leq m\varepsilon^3\} > \frac{\varepsilon\sqrt{m}}{2} \right]. \end{aligned}$$

Esta última desigualdad se debe a que  $\{w_{i,r}\}_{i \geq 1}$  son i.i.d.

Además por la desigualdad de Kolmogorov (ver [?] página 350)

$$\mathbb{P} \left[ \left| \frac{n_{ij}(m) - p_{ij}n_i(m)}{\sqrt{m}} - \tilde{H}(m)_{i,j} \right| > \varepsilon \right] \leq \varepsilon(1 + 8p_{ij}(1 - p_{ij})).$$

Lo anterior se traduce en

$$\left( \frac{n_{ij}(m) - p_{ij}n_i(m)}{\sqrt{m}} - \tilde{H}(m)_{i,j} \right) \left( \frac{m}{n_i(m)} \right) \xrightarrow[m \rightarrow \infty]{p} 0.$$

□

Sea  $X = \{X_n\}_{n \in \mathbb{N}}$  una cadena de Markov con espacio de estados  $E$  y con matriz de transición  $P = \{p_{ij}\}_{i,j \in E}$ . Supongamos que todas las probabilidades de transición son estrictamente positivas, esto con

el objeto de tener un espacio parametral abierto y así poder definir la matriz de información de Fisher. Definimos a nuestro espacio parametral como

$$\Theta = \{\tilde{P} \in \mathcal{M}_{N \times N-1} : p_{ij} > 0 \text{ para todo } i \in E, 1 \leq j \leq N-1 \text{ y } \sum_{j=1}^{N-1} p_{ij} < 1 \text{ para todo } i \in E\}.$$

El espacio parametral anterior ahora es un conjunto abierto con la norma usual, además nos permitió quitar las dependencias entre las probabilidades de transición correspondientes a cada renglón.

**Nota 4.3.** Por simplicidad adoptamos la notación  $L(P; \{x_n\}_{n=0}^m)$ , para referirnos a la función de verosimilitud de la muestra  $\{x_n\}_{n=0}^m$  de la cadena de Markov  $\{X_t\}_{t \geq 0}$  con parámetro  $\tilde{P}$ .

En lo que resta de esta sección vamos a trabajar específicamente con el espacio parametral anteriormente definido.

**Proposición 4.2.** Sea  $\{X_n\}_{n \in \mathbb{N}}$  una Cadena de Markov homogénea con espacio de estados  $E = \{1, 2, \dots, N\}$  y  $P$  su correspondiente matriz de transición con todas sus entradas estrictamente positivas. Entonces la matriz de información de Fisher está dada por

$$(i_m(\tilde{P}))_{i_1 j_1, i_2 j_2} = \delta_{i_1 i_2} \mathbb{E}[n_{i_1}(m)] \left[ \delta_{j_1 j_2} \frac{1}{p_{i_1 j_1}} + \frac{1}{p_{i_1 N}} \right].$$

Donde  $i_m(\tilde{P}) \in \mathcal{M}_{N^2-N \times N^2-N}$  y los índices corren en  $i_1, i_2 \in E$  y  $1 \leq j_1, j_2 \leq N-1$ .

Además si suponemos que la cadena es irreducible y recurrente positiva, la matriz de información es invertible y su inversa está dada por

$$(i_m(\tilde{P}))_{i_1 j_1, i_2 j_2}^{-1} = \frac{\delta_{i_1 i_2} (\delta_{j_1 j_2} p_{i_1 j_1} - p_{i_1 j_1} p_{i_1 j_2})}{\mathbb{E}[n_{i_1}(m)]}$$

con  $i_1, i_2 \in E$  y  $1 \leq j_1, j_2 \leq N-1$ .

*Demostración.* Se sabe que el espacio  $\Theta$  es abierto y no vacío, además se tiene que por la definición de  $\Theta$  El conjunto

$$A = \{(x_0, x_1, \dots, x_m) | L(P; \{x_n\}_{n=0}^m) > 0\}$$

es independiente del parámetro en el espacio  $\tilde{P} \in \Theta$ , más aún resulta ser  $E^m$ .

Debido a que  $L(P; \{x_n\}_{n=0}^m)$  es en realidad un polinomio que depende solamente de potencias del conjunto  $\{p_{ij}\}_{i \in E, 1 \leq j \leq N-1}$  se tiene que

$$\frac{\partial^2 L(P; \{x_n\}_{n=0}^m)}{\partial p_{i_1 j_1} \partial p_{i_2 j_2}}$$

existe y es finita para todo  $i_1, i_2 \in E$  y  $1 \leq j_1, j_2 \leq N-1$ .

Por estas últimas observaciones se sabe que la matriz de información de Fisher también puede ser calculada como

$$i_m(\tilde{P})_{i_1 j_1, i_2 j_2} = \mathbb{E} \left[ - \frac{\partial^2 \log(L(P; \{X_n\}_{n=0}^m))}{\partial p_{i_1 j_1} \partial p_{i_2 j_2}} \right]$$

que resulta ser igual a

$$\delta_{i_1 i_2} \mathbb{E} \left[ \delta_{j_1 j_2} \frac{n_{i_1 j_1}(m)}{p_{i_1 j_1}^2} + \frac{n_{i_1 N}(m)}{(1 - \sum_{j=1}^{N-1} p_{i_1 j})^2} \right]$$

Por otra parte ya que

$$\sum_{x_0, x_1, \dots, x_m \in E} L(P; \{x_n\}_{n=0}^m) = 1$$

la derivada puede introducirse a la suma y en consecuencia

$$\mathbb{E} \left[ \frac{\partial \log(L(P; \{X_n\}_{n=0}^m))}{\partial p_{i_1 j_1}} \right] = 0$$

para todo  $i_1 \in E$  y  $1 \leq j_1 \leq N-1$  Lo que significa que  $i_m(\tilde{P})_{i_1 j_1, i_2 j_2}$  es igual a la covarianza entre los siguientes términos

$$\frac{\partial \log(L(P; \{X_n\}_{n=0}^m))}{\partial p_{i_1 j_1}}$$

y

$$\frac{\partial \log(L(P; \{X_n\}_{n=0}^m))}{\partial p_{i_2 j_2}}.$$

Por lo que se concluye que la matriz de información de Fisher es una matriz semi-definida positiva. Además

$$\begin{aligned} i_m(\tilde{P})_{i_1 j_1, i_2 j_2} &= \delta_{i_1 i_2} \mathbb{E} \left[ \delta_{j_1 j_2} \frac{n_{i_1 j_1}(m)}{p_{i_1 j_1}^2} + \frac{n_{i_1 N}(m)}{(1 - \sum_{j=1}^{N-1} p_{i_1 j})^2} \right] \\ &= \delta_{i_1 i_2} \mathbb{E}[n_{i_1}(m)] \left[ \delta_{j_1 j_2} \frac{1}{p_{i_1 j_1}} + \frac{1}{(1 - \sum_{j=1}^{N-1} p_{i_1 j})} \right]. \end{aligned}$$

Sean  $\{A^i\}_{i \in E}$  matrices de dimensión  $N-1 \times N-1$  tales que  $A^i_{j_1 j_2} = \left[ \delta_{j_1 j_2} \frac{1}{p_{i j_1}} + \frac{1}{p_{i N}} \right]$  para todo  $i \in E$  y  $0 < j_1, j_2 < N$ , entonces se tiene

$$i_m(\tilde{P}) = \begin{pmatrix} \mathbb{E}[n_1(m)]A^1 & 0 & \cdots & 0 \\ 0 & \mathbb{E}[n_2(m)]A^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbb{E}[n_N(m)]A^N \end{pmatrix}.$$

Ahora queremos probar que esta matriz es invertible, basta con probar que las matrices  $\{A^i\}_{i \in E}$  lo son. Realizando operaciones elementales se muestra que la matrices  $\{A^i\}_{i \in E}$  son invertibles y además

$$(A^i)_{j_1 j_2}^{-1} = (\delta_{j_1 j_2} p_{i j_1} - p_{i j_1} p_{i j_2})$$

para todo  $i \in E$  y  $0 < j_1, j_2 < N$

Por lo anterior la matriz de información de Fisher es definida positiva ya que es semi-definida positiva e invertible. Por último concluimos que

$$(i_m(\tilde{P}))_{i_1 j_1, i_2 j_2}^{-1} = \frac{\delta_{i_1 i_2} (\delta_{j_1 j_2} p_{i_1 j_1} - p_{i_1 j_1} p_{i_1 j_2})}{\mathbb{E}[n_{i_1}(m)]}$$

para todo  $i_1, i_2 \in E$  y  $1 \leq j_1, j_2 \leq N-1$ .

Nótese que para que la inversa este bien definida necesitamos que

$$\mathbb{E}[n_{i_1}(m)] > 0 \text{ para todo } i_1 \in E.$$

Esto se obtiene por las hipótesis de recurrencia positiva e irreducibilidad cuando  $m$  es grande.  $\square$

**Teorema 4.3. (Eficiencia Asintótica)** Sea  $\{X_n\}_{n \in \mathbb{N}}$  una Cadena de Markov homogénea con espacio de estados  $E = \{1, 2, \dots, N\}$  suponemos además que es irreducible y recurrente positiva entonces

$$\sqrt[2]{m}(\hat{P}(m) - \tilde{P}) \xrightarrow[m \rightarrow \infty]{d} N(\tilde{0}, \tilde{\Sigma})$$

Donde

$$\tilde{\Sigma} = \lim_{m \rightarrow \infty} m(i_m(\tilde{P}))^{-1},$$

$\hat{P}(m)$  es la matriz  $\hat{P}(m)$  quitándole la última columna y  $\tilde{0}$  es la matriz de dimensión  $N \times (N - 1)$  con todas sus entradas iguales a cero.

*Demostración.* La prueba de este Teorema se sigue de los Teoremas 2.2, 4.2 y la Proposición 4.2.  $\square$

**Ejemplo 4.3.** Se simuló una cadena de Markov con espacio de estados  $E = \{1, 2\}$ , distribución inicial

$$\pi = (0.5, 0.5)$$

y matriz de transición

$$P = \begin{pmatrix} 0.2 & 0.8 \\ 0.8 & 0.2 \end{pmatrix},$$

con horizonte  $m = 100000$ .

Es fácil mostrar que  $(0.5, 0.5)$  es la única distribución estacionaria, por lo tanto

$$\tilde{\Sigma} = \begin{pmatrix} 0.32 & 0.0 \\ 0.0 & 0.32 \end{pmatrix}.$$

Mientras que la inversa de la matriz de información de Fisher estimada esta dada por

$$m \left[ \widehat{i_m(\tilde{P})}^{-1} \right] = \begin{pmatrix} 0.32075 & 0.0 \\ 0.0 & 0.323248 \end{pmatrix}.$$

Para la estimación anterior se utilizó la función del Código D.17

## 4.2. Procesos de saltos de Markov

### 4.2.1. Con información a tiempo continuo.

Nuestro objetivo en esta sección es estimar el generador infinitesimal dada una trayectoria de un PSM, así como también demostrar algunas propiedades estadísticas del estimador de máxima verosimilitud.

En secciones pasadas estudiamos la importancia del generador infinitesimal, por lo cual nuestro objetivo será estimar sus parámetros vía máxima verosimilitud.

A partir de ahora sólo consideraremos procesos cuyos generadores infinitesimales tengan entradas en la diagonal estrictamente negativas, esto implica que los tiempos de saltos sean finitos casi seguramente. La existencia de estos procesos se probó en el Teorema 3.2.

### Estimación máximo verosímil

La siguiente proposición demuestra la existencia del estimador máximo verosímil correspondiente a una observación en tiempo continuo  $\{x_t\}_{0 \leq t \leq T}$  de un PSM.

**Proposición 4.3.** *Supongamos que tenemos una trayectoria observada en tiempo continuo  $\{x_t\}_{0 \leq t \leq T}$  de un PSM  $\{X_t\}_{t \geq 0}$  con espacio de estados  $E$  y generador infinitesimal  $Q$ , entonces el estimador máximo verosímil del generador infinitesimal dado que el proceso inicio en el estado  $x_0$  está dado por :*

$$\hat{Q}(T) = \{\hat{q}_{xy}(T)\}_{x,y \in E}$$

donde

$$\hat{q}_{xy}(T) = \begin{cases} N_{xy}(T)/R_x(T) & \text{para } x \neq y \\ -N_x(T)/R_x(T) & \text{para } x = y. \end{cases}$$

Donde  $T$  es el horizonte de tiempo,  $\hat{Q}(T)$  denota el estimador máximo verosímil del generador infinitesimal,  $N_{xy}(T)$  es el número de veces observadas que el proceso saltó del estado  $x$  al estado  $y$  en un tiempo de salto, mientras que  $R_x(T) = \int_0^T \mathbb{I}_{\{X_t=x\}} dt$ , es el tiempo total observado en el cual el proceso se mantuvo en el estado  $x$ , denotemos por  $N_x(T) = \sum_{y \neq x} N_{xy}(T)$ .

**Nota 4.4.** *Una observación que vale la pena mencionar es que la existencia de este estimador está condicionada a que  $N_x(T) > 0$  para todo  $x \in E$ .*

*Demostración.* Asumimos que el proceso tuvo  $n$  saltos, sean  $0 = t_0 < t_1 < \dots < t_n$  los tiempos observados de salto hasta el tiempo  $T$ . Por los Teoremas 2.6, 2.7 y 2.8 la función de verosimilitud es igual a

$$\begin{aligned} L(Q; \{x_t\}_{0 \leq t \leq T}) &= f_{T_1, X_{T_1}, T_2, X_{T_2}, \dots, T_n, X_{T_n}, N_T | X_0}(t_1, x_{t_1}, t_2, x_{t_2}, \dots, t_n, x_{t_n}, n | x_0) \\ &= f_{T_1, X_{T_1}, T_2, X_{T_2}, \dots, T_n, X_{T_n} | X_0}(t_1, x_{t_1}, t_2, x_{t_2}, \dots, t_n, x_{t_n} | x_0) \mathbb{P}[T_{n+1} - T_n > T - t_n | X_{T_n} = x_{t_n}] \\ &= (\lambda_{x_0} e^{-\lambda_{x_0} t_1} p_{x_0 x_{t_1}}) f_{T_1, X_{T_1}, T_2, X_{T_2}, \dots, T_{n-1}, X_{T_{n-1}} | X_0}(t_2, x_{t_2}, \dots, t_n, x_{t_n} | x_{t_1}) \exp(-\lambda_{x_{t_n}}(T - t_n)). \end{aligned}$$

Por lo que obtenemos

$$\begin{aligned} L(Q; \{x_t\}_{0 \leq t \leq T}) &= \left[ \prod_{i=1}^n (\lambda_{x_{t_{i-1}}} \exp(-\lambda_{x_{t_{i-1}}}(t_i - t_{i-1})) p_{x_{t_{i-1}} x_{t_i}}) \right] \exp(-\lambda_{x_{t_n}}(T - t_n)) \\ &= \prod_{x \in E} \lambda_x^{N_x(T)} \exp(-\lambda_x R_x(T)) \prod_{y \neq x} p_{xy}^{N_{xy}(T)} \\ &= \prod_{x \in E} \prod_{y \neq x} (\lambda_x p_{xy})^{N_{xy}(T)} \exp(-\lambda_x p_{xy} R_x(T)) \\ &= \prod_{x \in E} \prod_{y \neq x} q_{xy}^{N_{xy}(T)} \exp(-q_{xy} R_x(T)). \end{aligned}$$

El espacio parametral está dado por

$$\mathcal{Q} = \{Q \in \mathcal{M}_{N \times N} : q_{xy} \geq 0 \text{ para todo } y \neq x \text{ y } \sum_{y \neq x} q_{xy} = -q_{xx} > 0 \text{ para todo } x \in E\}.$$

Por otro lado la función de log-verosimilitud es

$$l(Q; \{x_t\}_{0 \leq t \leq T}) = \sum_{x \in E} \sum_{y \neq x} N_{xy}(T) \log(q_{xy}) - q_{xy} R_x(T).$$

La última expresión depende sólo de las entradas fuera de la diagonal de la matriz  $Q$ , esto quiere decir que no tenemos ninguna restricción adicional a la no negatividad sobre las  $\{q_{xy}\}_{x,y \in E, x \neq y}$ , por lo cual derivando la log-verosimilitud e igualando a cero tenemos como punto crítico

$$\hat{q}_{xy}(T) = \begin{cases} N_{xy}(T)/R_x(T) & \text{para } x \neq y \\ -N_x(T)/R_x(T) & \text{para } x = y. \end{cases} \quad (4.3)$$

Hasta aquí se mostró que

$$\hat{Q}(T) = \{\hat{q}_{xy}(T)\}_{x,y \in E} \quad (4.4)$$

es un punto crítico de  $l(Q; \{x_t\}_{0 \leq t \leq T})$  sobre  $\mathcal{Q}$ , ahora sólo falta probar que es un punto máximo. Nótese que

$$l(Q; \{x_t\}_{0 \leq t \leq T}) = \left( \sum_{x \in E} \sum_{y \in E_x} N_{xy}(T) \log(q_{xy}) \right) + \left( \sum_{x \in E} q_{xx} R_x(T) \right), \quad (4.5)$$

donde  $E_x = \{y \neq x : N_{xy}(T) > 0\}$ .

Por (4.5) sabemos que si  $Q$  es tal que  $q_{xy} > 0$  para algún  $y \in E - E_x$ , el valor de  $\sum_{x \in E} q_{xx} R_x(T)$  disminuye ya que  $q_{xx}$  se hace más pequeño.

Por lo anterior  $q_{xy} = 0$  para todo  $x, y \in E$  tal que  $y \in E - E_x$ , en resumen maximizar  $l(Q; \{x_t\}_{0 \leq t \leq T})$  se reduce a maximizar en el conjunto  $\{q_{xy} : x, y \in E, y \in E_x\}$ .

Por otro lado para mostrar que (4.4) es un punto máximo de  $l(Q; \{x_t\}_{0 \leq t \leq T})$ , por el Teorema 6.9.4 (página 364 de [? ]) es suficiente mostrar que

$$\left\{ \frac{\partial^2 l(\hat{Q}(T); \{x_t\}_{0 \leq t \leq T})}{\partial q_{x_1 y_1} \partial q_{x_2 y_2}} \right\}_{x_1, x_2 \in E, y_1 \in E_{x_1}, y_2 \in E_{x_2}},$$

es una matriz diagonal y que en la diagonal sus elementos son estrictamente negativos, desarrollando las derivadas de segundo orden obtenemos

$$\frac{\partial^2 l(\hat{Q}(T); \{x_t\}_{0 \leq t \leq T})}{\partial q_{x_1 y_1} \partial q_{x_2 y_2}} = \delta_{x_1 x_2} \delta_{y_1 y_2} \frac{-N_{x_1 y_1}(T)}{(\hat{q}_{x_1 y_1}(T))^2}, \text{ para todo } x_1, x_2 \in E, y_1 \in E_{x_1}, y_2 \in E_{x_2}.$$

Lo anterior prueba que  $\hat{Q}(T)$  de la expresión (4.4) es el estimador máximo verosímil.  $\square$

**Ejemplo 4.4.** Estimación por máxima verosimilitud del generador infinitesimal correspondiente a la trayectoria simulada en el Ejemplo 3.3, la función que se utilizó se encuentra en el Código D.5.

$$\hat{Q}(T) = \begin{pmatrix} -0.144964 & 0.0362409 & 0.0181205 & 0.0906023 \\ 0.0304693 & -0.182816 & 0.0761733 & 0.0761733 \\ 0.187188 & 0.140391 & -0.655159 & 0.32758 \\ 0.0172995 & 0.121096 & 0.138396 & -0.276792 \end{pmatrix}.$$

A continuación mostraremos algunas propiedades del estimador de máxima verosimilitud del generador infinitesimal de un PSM dado, dichas características prueban la importancia de utilizar este estimador. El siguiente teorema prueba que nuestro estimador converge al verdadero parámetro conforme la muestra crece.

### Propiedades asintóticas

**Teorema 4.4. (Consistencia)** *Sea un PSM  $\{X_t\}_{t \geq 0}$  ergódico con generador infinitesimal  $Q$  y con espacio de estados  $E$ , el estimador máximo verosímil del generador infinitesimal  $\hat{Q}(T)$  dado por  $\hat{Q}(T) = \{\hat{q}_{xy}(T)\}_{x,y \in E}$  converge casi seguramente al verdadero parámetro  $Q$  cuando  $T$  (horizonte) tiende a infinito. En otras palabras el estimador es consistente.*

*Demostración.* Por los Teoremas 2.11 y 2.13 concluimos

$$\frac{N_{xy}(T)}{R_x(T)} = \frac{T}{R_x(T)} \frac{N_{xy}(T)}{T} \xrightarrow[T \rightarrow \infty]{c.s.} q_{xy}.$$

Ya que el espacio de estados es finito se puede afirmar lo siguiente

$$\|Q - \hat{Q}(T)\| \xrightarrow[T \rightarrow \infty]{c.s.} 0 \text{ cuando } T \rightarrow \infty.$$

□

**Ejemplo 4.5.** *Se simuló una trayectoria del PSM del Ejemplo 3.3 con horizonte  $T=10000$ . En la figura 4.2 se muestra  $\|\hat{Q}(t) - Q\|$  para cada  $t \leq 10000$ , para este análisis se utilizó el Código D.6.*

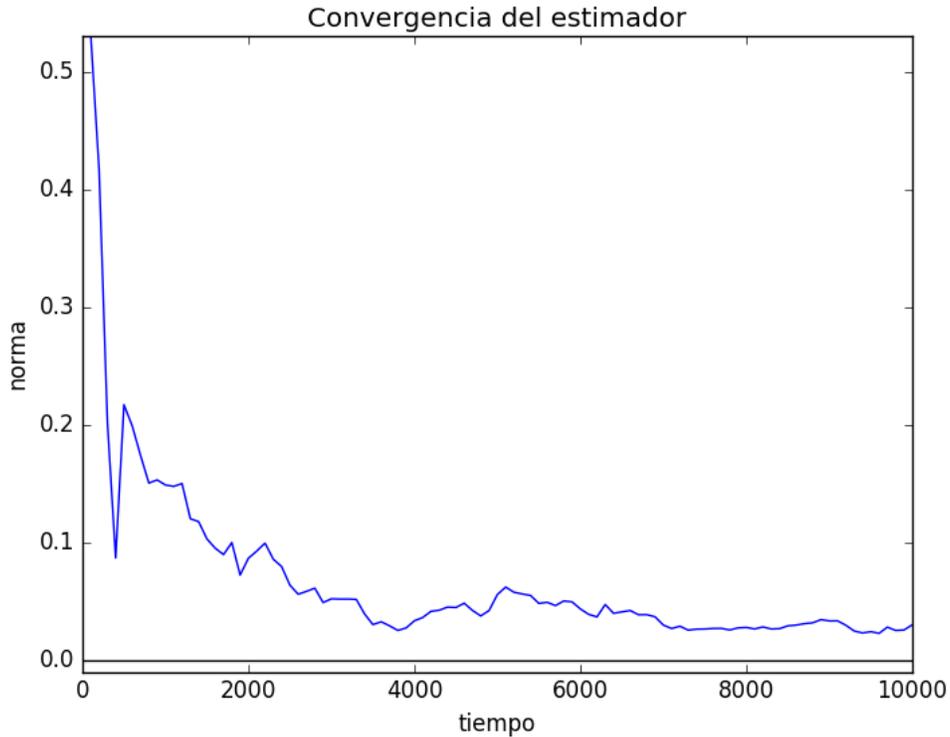


Figura 4.2: Gráfica de la norma entre  $Q$  y  $\hat{Q}(T)$  respecto al tiempo

Al igual que en el caso de las cadenas de Markov a tiempo discreto, es posible probar en el caso de PSM que su estimador máximo verosímil asociado al generador infinitesimal, es asintóticamente eficiente. Para mostrar esta propiedad, requerimos el siguiente lema.

**Lema 4.2.** Sean  $0 = T_0 < T_1 < \dots < T_n < \infty$  los tiempos de salto de un PSM  $\{X_t\}_{t \geq 0}$  con espacio de estados  $E$  y generador infinitesimal  $Q$ , entonces

$$\frac{\mathbb{E}[T - T_{N_T}]}{\sqrt{T}} \xrightarrow{T \rightarrow \infty} 0.$$

En particular por la Proposición B.3 tenemos convergencia en probabilidad.

*Demostración.*

$$\begin{aligned} \mathbb{E}[T - T_{N_T}] &= T\mathbb{P}[N_T = 0] + \sum_{n=1}^{\infty} \mathbb{E}[(T - T_{N_T})\mathbb{I}_{\{N_T=n\}}] \\ &= T\mathbb{P}[N_T = 0] + \sum_{n=1}^{\infty} \mathbb{E}[(T - T_n)\mathbb{I}_{\{T_n \leq T < T_{n+1}\}}] \\ &= T\mathbb{P}[N_T = 0] + \sum_{n=1}^{\infty} \int_0^T \int_T^{\infty} (T - u) f_{T_n, T_{n+1}}(u, v) dv du \\ &= T\mathbb{P}[N_T = 0] + \sum_{n=1}^{\infty} \sum_{x \in E} \int_0^T \int_T^{\infty} (T - u) f_{T_n, X_{T_n}}(u, x) f_{T_0|X_0=x}(v - u) dv du. \end{aligned}$$

La igualdad anterior se debe al Teorema 2.8, dicho lo anterior integrando y sumando sobre  $x \in E$  obtenemos

$$\begin{aligned} \sum_{x \in E} \int_0^T \int_T^{\infty} (T - u) f_{T_n, X_{T_n}}(u, x) f_{T_0|X_0=x}(v - u) dv du &= \sum_{x \in E} \int_0^T (T - u) f_{T_n, X_{T_n}}(u, x) \exp^{-\lambda_x(T-u)} du \\ &\leq \int_0^T (T - u) f_{T_n}(u) \exp^{-\alpha(T-u)} du. \end{aligned}$$

Donde  $\alpha = \min\{\lambda_x : x \in E\}$ , ahora integramos por partes el término anterior y obtenemos

$$\begin{aligned} &\int_0^T (T - u) f_{T_n}(u) \exp^{-\alpha(T-u)} du = \\ &\int_0^{T-\frac{1}{\alpha}} -F_{T_n}(u) \exp^{-\alpha(T-u)} [\alpha(T-u) - 1] du + \int_{T-\frac{1}{\alpha}}^T -F_{T_n}(u) \exp^{-\alpha(T-u)} [\alpha(T-u) - 1] du. \end{aligned} \quad (4.6)$$

El primer término de (4.6) es negativo mientras que el segundo es positivo, por esta razón se sigue del Lema 2.1 la siguiente desigualdad, donde  $\beta = \max\{\lambda_x : x \in E\}$

$$\sum_{n=1}^{\infty} \int_0^T (T - u) f_{T_n}(u) \exp^{-\alpha(T-u)} du \leq \int_{T-\frac{1}{\alpha}}^T \beta u \exp^{-\alpha(T-u)} [1 - \alpha(T-u)] du.$$

Por esta desigualdad

$$\mathbb{E}[T - T_{N_T}] \leq T\mathbb{P}[N_T = 0] + \int_{T-\frac{1}{\alpha}}^T \beta u \exp^{-\alpha(T-u)} [1 - \alpha(T-u)] du. \quad (4.7)$$

Integrando por partes el segundo miembro de la desigualdad (4.7) tenemos

$$\begin{aligned} \int_{T-\frac{1}{\alpha}}^T \beta u \exp^{-\alpha(T-u)} [1 - \alpha(T-u)] du &= \frac{\exp(-1)}{\alpha} + \int_{T-\frac{1}{\alpha}}^T \beta(T-u) \exp^{-\alpha(T-u)} du \\ &= \frac{\exp(-1)}{\alpha} + \int_0^{\frac{1}{\alpha}} \beta v \exp^{-\alpha v} dv. \end{aligned}$$

El segundo sumando de la última igualdad es menor o igual que la esperanza de una exponencial de parámetro  $\alpha$  multiplicada por  $\frac{\beta}{\alpha}$ , por esta razón tenemos

$$\mathbb{E}[T - T_{N_T}] \leq \sum_{x \in E} T \mathbb{P}[X_0 = 0] \exp^{-\lambda_x T} + \frac{\exp(-1) + \beta/\alpha}{\alpha}.$$

Concluimos la convergencia

$$\frac{\mathbb{E}[T - T_{N_T}]}{\sqrt{T}} \xrightarrow{T \rightarrow \infty} 0.$$

Por último debido a la Proposición B.2 tenemos

$$\frac{T - T_{N_T}}{\sqrt{T}} \xrightarrow{T \rightarrow \infty} 0.$$

□

**Teorema 4.5.** *Considere un PSM  $\{X_t\}_{t \geq 0}$  ergódico con espacio de estados  $E$  y generador infinitesimal  $Q$ , entonces*

$$\sqrt{T}(\hat{Q}(T) - \tilde{Q}) \xrightarrow{T \rightarrow \infty} N(\tilde{0}, \Lambda).$$

Donde  $\hat{Q}(T)$  y  $\tilde{Q}$  son el estimador y el generador infinitesimal sin la diagonal, es decir matrices de dimensión  $N \times (N-1)$  y  $\Lambda \in \mathcal{M}_{(N^2-N) \times (N^2-N)}$  es la matriz con entradas  $\Lambda_{x_1 y_1, x_2 y_2} = \frac{1}{\nu_{x_1}} \delta_{x_1 x_2} \delta_{y_1 y_2} q_{x_1 y_1}$  para todo  $x_1, y_1, x_2, y_2 \in E$  con  $x_1 \neq y_1$  y  $x_2 \neq y_2$ .

*Demostración.* Sin pérdida de generalidad suponemos que el proceso  $\{X_t\}_{t \geq 0}$  está construido como en la Definición 3.3.

$$\begin{aligned} \sqrt{T}(\hat{Q}(T) - \tilde{Q}) &= \left\{ \sqrt{T} \left( \frac{N_{xy}(T)}{R_x(T)} - q_{xy} \right) \right\}_{x \in E, y \neq x} \\ &= \left\{ \sqrt{T} \left( \frac{N_{xy}(T)}{R_x(T)} - \lambda_x \frac{N_{xy}(T)}{N_x(T)} + \lambda_x \frac{N_{xy}(T)}{N_x(T)} - q_{xy} \right) \right\}_{x \in E, y \neq x} \\ &= \left\{ \sqrt{T} \left( \frac{N_{xy}(T)}{N_x(T)} \left( \frac{N_x(T)}{R_x(T)} - \lambda_x \right) + \lambda_x \left( \frac{N_{xy}(T)}{N_x(T)} - p_{xy} \right) \right) \right\}_{x \in E, y \neq x} \\ &= \left\{ \left( \frac{TN_{xy}(T)}{N_x(T)R_x(T)} \left( \frac{N_x(T) - \lambda_x R_x(T)}{\sqrt{T}} \right) + \frac{\lambda_x T}{N_x(T)} \left( \frac{N_{xy}(T) - p_{xy} N_x(T)}{\sqrt{T}} \right) \right) \right\}_{x \in E, y \neq x}. \end{aligned}$$

Supongamos por el momento las siguientes afirmaciones:

$$\left| \frac{N_x(T) - \lambda_x R_x(T)}{\sqrt{T}} - \frac{[T\nu_x \lambda_x] - \lambda_x \sum_{r=0}^{[T\nu_x \lambda_x]} \frac{S_{x,r}}{\lambda_x}}{\sqrt{T}} \right| \xrightarrow{T \rightarrow \infty} 0. \quad (4.8)$$

$$\left| \frac{N_{xy}(T) - p_{xy}N_x(T)}{\sqrt{T}} - \frac{\sum_{r=1}^{\lceil T\nu_x\lambda_x \rceil} \mathbb{I}_{\{w_{x,r}=y\}} - p_{xy} \lceil T\nu_x\lambda_x \rceil}{\sqrt{T}} \right| \xrightarrow[T \rightarrow \infty]{p} 0. \quad (4.9)$$

Definimos  $\xi^x(T) = \frac{\lceil T\nu_x\lambda_x \rceil - \lambda_x \sum_{r=0}^{\lceil T\nu_x\lambda_x \rceil} \frac{S_{x,r}}{\lambda_x}}{\sqrt{T}}$ ,  $\rho^{xy}(T) = \frac{\sum_{r=1}^{\lceil T\nu_x\lambda_x \rceil} \mathbb{I}_{\{w_{x,r}=y\}} - p_{xy} \lceil T\nu_x\lambda_x \rceil}{\sqrt{T}}$ , también

$$Y^T = \left\{ \frac{TN_{xy}(T)}{N_x(T)R_x(T)} \left( \frac{N_x(T) - \lambda_x R_x(T)}{\sqrt{T}} - \xi^x(T) \right) + \frac{\lambda_x T}{N_x(T)} \left( \frac{N_{xy}(T) - p_{xy}N_x(T)}{\sqrt{T}} - \rho^{xy}(T) \right) \right\}_{x \in E, y \neq x}$$

y

$$Z^T = \left\{ \left( \frac{TN_{xy}(T)}{N_x(T)R_x(T)} (\xi^x(T)) + \frac{\lambda_x T}{N_x(T)} (\rho^{xy}(T)) \right) \right\}_{x \in E, y \neq x}.$$

Bajo la notación anterior tenemos que  $\sqrt{T}(\hat{Q}(T) - \tilde{Q}) = Y^T + Z^T$ .

Por otro lado por los Teoremas 2.11 y 2.13 sabemos que

$$\frac{N_{xy}(T)}{N_x(T)} \xrightarrow[T \rightarrow \infty]{c.s.} p_{xy}, \quad \frac{T}{R_x(T)} \xrightarrow[T \rightarrow \infty]{c.s.} \frac{1}{\nu_x}, \quad \frac{T}{N_x(T)} \xrightarrow[T \rightarrow \infty]{c.s.} \frac{1}{\nu_x \lambda_x},$$

y debido al Lema B.1 y las afirmaciones (4.8) y (4.9) tenemos que  $Y^T \xrightarrow[T \rightarrow \infty]{p} 0$ . y por la Proposición B.9 sabemos que es suficiente verificar la convergencia para  $Z^T$ .

Establezcamos la siguiente notación

$$Z_1^T = \left\{ \left( \frac{TN_{xy}(T)}{N_x(T)R_x(T)} - \frac{p_{xy}}{\nu_x} \right) \xi^x(T) + \left( \frac{\lambda_x T}{N_x(T)} - \frac{1}{\nu_x} \right) \rho^{xy}(T) \right\}_{x \in E, y \neq x}$$

y

$$Z_2^T = \left\{ \left( \frac{p_{xy}}{\nu_x} \right) \xi^x(T) + \left( \frac{1}{\nu_x} \right) \rho^{xy}(T) \right\}_{x \in E, y \neq x}.$$

Bajo la notación anterior tenemos que  $Z^T = Z_1^T + Z_2^T$ , a continuación mostraremos que  $Z_1^T \xrightarrow[T \rightarrow \infty]{p} 0$  y por la Proposición B.9 bastaría con mostrar la convergencia para  $Z_2^T$ .

Por la Definición 3.3 sabemos que  $\{S_{x,r}\}_{r \geq 1}$  y  $\{w_{i,r}\}_{i \geq 1}$  son independientes y por el Teorema central del límite tenemos que

$$\xi^x(T) \xrightarrow[T \rightarrow \infty]{d} N(0, \nu_x \lambda_x)$$

y

$$\rho^{xy}(T) \xrightarrow[T \rightarrow \infty]{d} N(0, \nu_x \lambda_x p_{xy}).$$

Por otro lado por los Teoremas 2.11 y 2.13 y por la Proposición B.7 tenemos que

$$Z_1(T) \xrightarrow[T \rightarrow \infty]{p} 0.$$

Por lo anterior basta ver la convergencia de  $Z_2(T)$  y por la Definición 3.3 sabemos que  $\{S_{x,r}\}_{r \geq 1}$  y  $\{w_{i,r}\}_{i \geq 1}$  son independientes, por tanto  $\xi^x(T)$  es independiente a  $\rho^{xy}(T)$ . Por otro lado por el Teorema de Curtiss (B.3) basta analizar la convergencia para los siguientes dos términos

$$\mathbb{E} \left[ \exp \left( \sum_{x \in E} \sum_{y \neq x}^N u_{xy} \frac{p_{xy}}{\nu_x} \left( \frac{\lceil T\nu_x\lambda_x \rceil - \sum_{r=0}^{\lceil T\nu_x\lambda_x \rceil} S_{x,r}}{\sqrt{T}} \right) \right) \right]$$

y

$$\mathbb{E} \left[ \exp \left( \sum_{x \in E} \sum_{y \neq x}^N u_{xy} \frac{1}{\nu_x} \left( \frac{\sum_{r=1}^{\lceil T\nu_x\lambda_x \rceil} \mathbb{I}_{\{w_{x,r}=y\}} - p_{xy} \lceil T\nu_x\lambda_x \rceil}{\sqrt{T}} \right) \right) \right].$$

Donde  $U = \{u_{xy}\}_{x,y \in E, x \neq y}$  es cualquier matriz de dimensión  $N \times (N-1)$ . El primer término se puede escribir como

$$\prod_{x \in E} \mathbb{E} \left[ \exp \left( \left( \frac{\lceil T\nu_x\lambda_x \rceil - \sum_{r=0}^{\lceil T\nu_x\lambda_x \rceil} S_{x,r}}{\sqrt{\lceil T\nu_x\lambda_x \rceil}} \right) \sum_{y \neq x} u_{xy} \frac{p_{xy} \sqrt{\lceil T\nu_x\lambda_x \rceil}}{\nu_x \sqrt{T}} \right) \right]$$

el cual por la Proposición B.9 y el Teorema B.2 converge a

$$\prod_{x \in E} \exp \left( \frac{\frac{\lambda_x}{\nu_x} (\sum_{y \neq x} u_{xy} p_{xy})^2}{2} \right). \quad (4.10)$$

Mientras que el segundo término puede ser escrito como

$$\mathbb{E} \left[ \exp \left( \sum_{x \in E} \sum_{y \neq x}^N u_{xy} \sqrt{\frac{\lambda_x}{\nu_x}} \left( \frac{\sum_{r=1}^{\lceil T\nu_x\lambda_x \rceil} \mathbb{I}_{\{w_{x,r}=y\}} - p_{xy} \lceil T\nu_x\lambda_x \rceil}{\sqrt{\lceil T\nu_x\lambda_x \rceil}} \right) \right) \right].$$

Por el Lema 4.1 converge al número

$$\prod_{x \in E} \exp \left[ \sum_{k_1 \neq x} \sum_{k_2 \neq x} \frac{\lambda_x}{\nu_x} u_{xk_1} u_{xk_2} \frac{p_{xk_1}}{2} (\delta_{k_2k_1} - p_{xk_2}) \right]. \quad (4.11)$$

Uniendo (4.10) y (4.11) se concluye que

$$\sqrt{T}(\hat{Q}(T) - \tilde{Q}) \xrightarrow[T \rightarrow \infty]{d} N(\tilde{0}, \Lambda).$$

Donde  $\Lambda_{x_1y_1, x_2y_2} = \frac{1}{\nu_{x_1}} \delta_{x_1x_2} \delta_{y_1y_2} q_{x_1y_1}$  para todo  $x_1, y_1, x_2, y_2 \in E$  con  $x_1 \neq y_1$  y  $x_2 \neq y_2$ .

Ahora sólo resta probar

1.  $\left| \frac{N_x(T) - \lambda_x R_x(T)}{\sqrt{T}} - \frac{\lceil T\nu_x\lambda_x \rceil - \lambda_x \sum_{r=0}^{\lceil T\nu_x\lambda_x \rceil} \frac{S_{x,r}}{\lambda_x}}{\sqrt{T}} \right| \xrightarrow[T \rightarrow \infty]{p} 0.$
2.  $\left| \frac{N_{xy}(T) - p_{xy} N_x(T)}{\sqrt{T}} - \frac{\sum_{r=1}^{\lceil T\nu_x\lambda_x \rceil} \mathbb{I}_{\{w_{x,r}=y\}} - p_{xy} \lceil T\nu_x\lambda_x \rceil}{\sqrt{T}} \right| \xrightarrow[T \rightarrow \infty]{p} 0.$

Mostraremos la primer afirmación. Sabemos que

$$R_x(T) = \sum_{r=0}^{N_x(T)} \frac{S_{x,r}}{\lambda_x} + (T - T_{N_T}) \mathbb{I}_{\{X_{T_{N_T}}=x\}}.$$

Por el Lema 4.2 tenemos

$$\left| \frac{R_x(T) - \sum_{r=0}^{N_x(T)} \frac{S_{x,r}}{\lambda_x}}{\sqrt{T}} \right| \leq \frac{T - T_{N_T}}{\sqrt{T}} \xrightarrow[T \rightarrow \infty]{p} 0.$$

Por esto último basta probar

$$\left| \frac{N_x(T) - \lambda_x \sum_{r=0}^{N_x(T)} \frac{S_{x,r}}{\lambda_x}}{\sqrt{T}} - \frac{\lceil T\nu_x\lambda_x \rceil - \lambda_x \sum_{r=0}^{\lceil T\nu_x\lambda_x \rceil} \frac{S_{x,r}}{\lambda_x}}{\sqrt{T}} \right| \xrightarrow[T \rightarrow \infty]{p} 0.$$

Sea  $\varepsilon > 0$ .

Por el Teorema 2.13 podemos afirmar que existe  $k$  tal que para todo  $T > k$  se cumple

$$\mathbb{P}[|N_x(T) - \lceil T\nu_x\lambda_x \rceil| > T\varepsilon^3] < \varepsilon.$$

Denotemos a

$$V_n^{xy} = n - \sum_{r=0}^n S_{x,r}.$$

Por lo anterior

$$\mathbb{P}\left[ \left| \frac{N_x(T) - \sum_{r=0}^{N_x(T)} S_{x,r}}{\sqrt{T}} - \frac{\lceil T\nu_x\lambda_x \rceil - \sum_{r=0}^{\lceil T\nu_x\lambda_x \rceil} S_{x,r}}{\sqrt{T}} \right| > \varepsilon \right] \leq$$

$$\varepsilon + \mathbb{P}[\max\{|V_n^{xy} - V_{\lceil T\nu_x\lambda_x \rceil}^{xy}| : |n - \lceil T\nu_x\lambda_x \rceil| \leq T\varepsilon^3\} > \varepsilon\sqrt{T}]$$

que a su vez debido a que  $\{S_{x,r}\}_{r \geq 1}$  son i.i.d.

$$\leq \varepsilon + 2\mathbb{P}\left[ \max\{|V_n^{xy}| : 1 \leq n \leq T\varepsilon^3\} > \frac{\varepsilon\sqrt{T}}{2} \right].$$

Además por la desigualdad de Kolmogorov (ver [?] página 350)

$$\mathbb{P}\left[ \left| \frac{N_x(T) - \sum_{r=0}^{N_x(T)} S_{x,r}}{\sqrt{T}} - \frac{\lceil T\nu_x\lambda_x \rceil - \sum_{r=0}^{\lceil T\nu_x\lambda_x \rceil} S_{x,r}}{\sqrt{T}} \right| > \varepsilon \right] \leq 9\varepsilon.$$

Lo anterior se traduce en

$$\left| \frac{N_x(T) - \lambda_x R_x(T)}{\sqrt{T}} - \frac{\lceil T\nu_x\lambda_x \rceil - \lambda_x \sum_{r=0}^{\lceil T\nu_x\lambda_x \rceil} \frac{S_{x,r}}{\lambda_x}}{\sqrt{T}} \right| \xrightarrow[T \rightarrow \infty]{p} 0.$$

Ahora sólo resta probar la segunda afirmación.

Sea  $\varepsilon > 0$ . Por el Teorema 2.13 podemos afirmar que existe  $k$  tal que para todo  $T > k$  se cumple

$$\mathbb{P}[|N_x(T) - \lceil T\nu_x\lambda_x \rceil| > T\varepsilon^3] < \varepsilon.$$

Denotemos a

$$S_n^{xy} = \sum_{r=1}^n \mathbb{I}_{\{w_{x,r}=y\}} - p_{xy}n.$$

Por lo anterior

$$\mathbb{P}\left[ \left| \frac{N_{xy}(T) - p_{xy}N_x(T)}{\sqrt{T}} - \frac{\sum_{r=1}^{\lceil T\nu_x\lambda_x \rceil} \mathbb{I}_{\{w_{x,r}=y\}} - p_{xy} \lceil T\nu_x\lambda_x \rceil}{\sqrt{T}} \right| > \varepsilon \right]$$

es menor o igual que

$$\varepsilon + \mathbb{P}[\max\{|S_n^{xy} - S_{\lceil T\nu_x\lambda_x \rceil}^{xy}| : |n - \lceil T\nu_x\lambda_x \rceil| \leq T\varepsilon^3\} > \varepsilon\sqrt{T}]$$

a su vez dedido a que  $\{w_{x,r}\}_{r \geq 1}$  es i.i.d. es menor o igual que

$$\varepsilon + 2\mathbb{P}[\max\{|S_n| : 1 \leq n \leq T\varepsilon^3\} > \frac{\varepsilon\sqrt{T}}{2}].$$

Esta última desigualdad se debe a que  $\{w_{i,r}\}_{i \geq 1}$  son i.i.d.

Además por la desigualdad de Kolmogorov (ver [?] página 350)

$$\mathbb{P} \left[ \left| \frac{N_{xy}(T) - p_{xy}N_x(T)}{\sqrt{T}} - \frac{\sum_{r=1}^{\lceil T\nu_x\lambda_x \rceil} \mathbb{I}_{\{w_{x,r}=y\}} - p_{xy} \lceil T\nu_x\lambda_x \rceil}{\sqrt{T}} \right| > \varepsilon \right] \leq \varepsilon(1 + 8p_{ij}(1 - p_{ij})).$$

Lo anterior se traduce en

$$\left| \frac{N_{xy}(T) - p_{xy}N_x(T)}{\sqrt{T}} - \frac{\sum_{r=1}^{\lceil T\nu_x\lambda_x \rceil} \mathbb{I}_{\{w_{x,r}=y\}} - p_{xy} \lceil T\nu_x\lambda_x \rceil}{\sqrt{T}} \right| \xrightarrow{T \rightarrow \infty} 0.$$

□

Sea un PSM  $X = \{X_t\}_{t \geq 0}$  con espacio de estados  $E$  y con generador infinitesimal  $Q = \{q_{xy}\}_{x,y \in E}$ .

Supongamos que todas las probabilidades de transición  $\{p_{xy}\}_{x \in E, y \in E - \{x\}}$  son estrictamente positivas, esto con el objeto de tener un espacio parametral abierto y así poder definir la matriz de información de Fisher.

Definimos a nuestro espacio parametral como

$$\Theta = \{\tilde{Q} \in \mathcal{M}_{N \times N-1} : q_{xy} > 0 \text{ para todo } x \in E, y \in E - \{x\}\}.$$

El espacio parametral anterior ahora es un conjunto abierto con la norma usual, además nos permitió quitar las dependencias entre la diagonal del generador infinitesimal y las otras entradas de la matriz.

**Nota 4.5.** Por simplicidad adoptamos la notación  $L(Q; \{x_t\}_{0 \leq t \leq T})$ , para referirnos a la función de verosimilitud de la muestra  $\{x_t\}_{0 \leq t \leq T}$  del PSM  $\{X_t\}_{t \geq 0}$  con parámetro  $\tilde{Q}$ .

En lo que resta de esta sección vamos a trabajar específicamente con el espacio parametral anteriormente definido.

**Proposición 4.4.** Sea un PSM  $\{X_t\}_{t \geq 0}$  con espacio de estados  $E = \{1, 2, \dots, N\}$  y  $Q$  su correspondiente generador infinitesimal, entonces la matriz de información de Fisher está dada por

$$(i_T(\tilde{Q}))_{x_1y_1, x_2y_2} = \delta_{x_1x_2} \delta_{y_1y_2} \frac{\mathbb{E}[N_{x_1y_1}(T)]}{q_{x_1, y_1}^2}.$$

Donde  $i_T(\tilde{Q}) \in \mathcal{M}_{N^2 - N \times N^2 - N}$  y los índices corren en  $x_1, x_2 \in E$  y  $y_1 \in E - \{x_1\}, y_2 \in E - \{x_2\}$ .

Además si suponemos que el proceso es ergódico, la matriz de información es invertible y su inversa está dada por

$$(i_T(\tilde{Q}))_{x_1y_1, x_2y_2}^{-1} = \delta_{x_1x_2} \delta_{y_1y_2} \frac{q_{x_1, y_1}^2}{\mathbb{E}[N_{x_1y_1}(T)]}.$$

Donde  $x_1, x_2 \in E$  y  $y_1 \in E - \{x_1\}, y_2 \in E - \{x_2\}$ .

*Demostración.* Se sabe que el espacio  $\Theta$  es abierto y no vacío, además se tiene que por la definición de  $\Theta$  el conjunto

$$A = \{\{x_t\}_{0 \leq t \leq T} | L(Q; \{x_t\}_{0 \leq t \leq T}) > 0\}$$

es independiente del parámetro  $\tilde{Q} \in \Theta$ .

Debido a que

$$L(Q; \{x_t\}_{0 \leq t \leq T}) = \prod_{x \in E} \prod_{y \neq x} q_{xy}^{N_{xy}(T)} \exp(-q_{xy}R_x(T))$$

se tiene que

$$\frac{\partial^2 L(Q; \{x_t\}_{0 \leq t \leq T})}{\partial q_{x_1 y_1} \partial q_{x_2 y_2}}$$

existe y es finita para todo  $x_1, x_2 \in E$ ,  $y_1 \in E - \{x_1\}$ ,  $y_2 \in E - \{x_2\}$ .

Por estas últimas observaciones se sabe que la matriz de información de Fisher también puede ser calculada como

$$(i_T(\tilde{Q}))_{x_1 y_1, x_2 y_2} = \mathbb{E} \left[ -\frac{\partial^2 \log(L(Q; \{X_t\}_{0 \leq t \leq T}))}{\partial q_{x_1 y_1} \partial q_{x_2 y_2}} \right]$$

por la demostración de la Proposición 4.3

$$(i_T(\tilde{Q}))_{x_1 y_1, x_2 y_2} = \mathbb{E} \left[ \delta_{x_1 x_2} \delta_{y_1 y_2} \frac{N_{x_1 y_1}(T)}{q_{x_1 y_1}^2} \right] = \delta_{x_1 x_2} \delta_{y_1 y_2} \frac{\mathbb{E}[N_{x_1 y_1}(T)]}{q_{x_1 y_1}^2}.$$

Por lo que se concluye que la matriz de información de Fisher es semi-definida positiva, pues es una matriz diagonal y en la diagonal sus entradas son no negativas.

Gracias a la forma de la matriz de información de Fisher tenemos que

$$(i_T(\tilde{Q}))_{x_1 y_1, x_2 y_2}^{-1} = \delta_{x_1 x_2} \delta_{y_1 y_2} \frac{q_{x_1 y_1}^2}{\mathbb{E}[N_{x_1 y_1}(T)]}.$$

Nótese que para que la inversa este bien definida necesitamos que

$$\mathbb{E}[N_{x_1 y_1}(T)] > 0 \text{ para todo } x_1 \in E, y_1 \in E - \{x_1\}.$$

Esto se obtiene por las hipótesis de ergodicidad cuando  $T$  es grande.  $\square$

**Teorema 4.6. (Eficiencia Asintótica)** Sea un PSM  $\{X_t\}_{t \geq 0}$  con espacio de estados  $E = \{1, 2, \dots, N\}$  y  $Q$  su correspondiente generador infinitesimal, suponemos además que es ergódico entonces

$$\sqrt[2]{T}(\hat{Q}(T) - \tilde{Q}) \xrightarrow[T \rightarrow \infty]{d} N(\tilde{0}, \Lambda).$$

Donde

$$\Lambda = \lim_{T \rightarrow \infty} T(i_T(\tilde{Q}))^{-1}.$$

*Demostración.* La prueba de este teorema se obtiene de aplicar el Teorema 4.5, la Proposición 4.4 y el Teorema 2.13.  $\square$

**Ejemplo 4.6.** Se simuló un PSM con espacio de estados  $E = \{1, 2, 3\}$ , distribución inicial

$$\pi = (1/3, 1/3, 1/3)$$

y generador infinitesimal

$$Q = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix},$$

con horizonte  $T = 10000$ .

Es fácil mostrar que  $(1/3, 1/3, 1/3)$  es la única distribución estacionaria, por lo tanto

$$\Lambda = \begin{pmatrix} 3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 \end{pmatrix}.$$

Mientras que la inversa de la matriz de información de Fisher estimada esta dada por

$$T \left[ \widehat{i_T(\tilde{Q})}^{-1} \right] = \begin{pmatrix} 3.13155 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2.96953 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3.00593 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2.92098 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3.09605 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3.04055 \end{pmatrix}.$$

Para la estimación anterior se utilizó la función del Código [D.18](#)

#### 4.2.2. Con información a tiempo discreto.

En esta sección estudiaremos un método para generar trayectorias de un PSM condicionado a puntos finales. Además analizaremos la rapidez y precisión del método. Usaremos el método para estimar el generador infinitesimal de un PSM a través del método Monte Carlo de una Cadena de Markov y la Estimación de Máxima Verosimilitud.

#### Método Rejection Sampling REJ

Empezaremos con la definición de Puente de Markov necesaria para la explicación de esta sección.

**Definición 4.1.** *Un puente de Markov con parámetros  $s, a, b$  es un proceso estocástico con parámetro  $t \in [0, s]$  y con la misma distribución de un Proceso de Saltos de Markov  $\{X_t\}_{t=0}^s$  condicionado a  $X_0 = a$  y  $X_s = b$ , con  $a, b \in E$  y  $s > 0$ .*

Específicamente trayectorias que no cumplieron la condición este algoritmo simula una trayectoria del PSM condicionando la posición inicial  $X_0 = a$  y termina hasta generar una trayectoria con posición final  $X_s = b$ , rechazando todas aquellas.

Puentes de Markov con parámetros  $s, a, b$ .

**Algoritmo 4.1.** 1. Hacemos  $X_0 = a$ .

2. Simulamos una trayectoria  $X_s \sim PSM(s)$  (PSM con tiempo de horizonte  $s$ , generador infinitesimal  $Q$  y distribución inicial  $\alpha$ ).

3. Si  $X_s = b$  se termina, de lo contrario pasamos a 2.

Ya que el algoritmo rechaza trayectorias hasta encontrar una trayectoria que cumpla las condiciones mencionadas anteriormente, entonces es importante mencionar la siguiente proposición:

**Proposición 4.5.** *En el método REJ, la probabilidad de alcanzar el estado final observado es  $p_{ab}(T)$ .*

1. Si  $T$  es grande entonces  $p_{ab}(T) \approx \pi_b$ .

2. Si  $T$  es pequeño entonces  $p_{ab}(T) \approx q_{ab}T$ .

Tendremos en consideración los tiempos de ejecución excesivos, ya que muchas rutas se rechazan si:

- $T$  es grande y  $\pi_b$  es pequeño, o,
- $T$  es pequeño y  $a \neq b$ .

Así, junto al código [D.1](#) visualizaremos estos casos con los siguientes ejemplos. En el primer caso consideramos una trayectoria con  $T = 1000$  y para que  $\pi_1$  sea pequeño bastara hacer el generador infinitesimal

$$Q = \begin{pmatrix} -56.00 & 5 & 3 & 2 & 3 & 8 & 7 & 9 & 10 & 9.00 \\ 0.00 & -43 & 9 & 7 & 2 & 7 & 4 & 4 & 2 & 8.00 \\ 0.00 & 10 & -56 & 8 & 5 & 7 & 9 & 8 & 8 & 1.00 \\ 0.00 & 4 & 10 & -55 & 10 & 6 & 6 & 9 & 1 & 9.00 \\ 0.00 & 1 & 8 & 8 & -41 & 1 & 6 & 7 & 8 & 2.00 \\ 0.00 & 1 & 4 & 9 & 4 & -42 & 2 & 7 & 6 & 9.00 \\ 0.00 & 2 & 1 & 1 & 7 & 5 & -40 & 9 & 7 & 8.00 \\ 0.00 & 3 & 7 & 3 & 7 & 5 & 5 & -41 & 2 & 9.00 \\ 0.00 & 3 & 7 & 5 & 6 & 10 & 1 & 2 & -42 & 8.00 \\ 0.05 & 3 & 3 & 3 & 7 & 8 & 9 & 7 & 5 & -45.05 \end{pmatrix}$$

y  $\pi = (1/10, 1/10, 1/10, 1/10, 1/10, 1/10, 1/10, 1/10, 1/10, 1/10)$ . Para el segundo caso donde el tiempo es pequeño y  $a$  distinto de  $b$ , consideremos un puente de Markov con  $T = 0.0005$ ,  $a = 1$  y  $b = 2$ .



## Capítulo 5

# Ecuaciones diferenciales estocásticas (EDE)

### 5.1. Movimiento Browniano o proceso de Wiener

En 1828 el botánico Robert Brown observó que los granos de polen suspendidos en una cierta substancia presentan movimientos irregulares (véase [5]). Con la contribución del trabajo presentado por Albert Einstein en 1905 (véase [8]), se explicó que este movimiento es producto de las múltiples colisiones aleatorias de las moléculas de la substancia con los granos de polen. Se puede decir que un movimiento de este tipo tiene las siguientes características:

1. Es un movimiento continuo.
2. Tiene desplazamientos independientes en intervalos de tiempo disjuntos.
3. Los desplazamientos pueden modelarse por variables aleatorias gaussianas (teorema central del límite, véase [9]).

En 1923 el matemático Norbert Wiener demostró que existe un proceso estocástico con estas propiedades.

**Definición 5.1.** *Un proceso de Wiener (Movimiento Browniano) estándar unidimensional es un proceso estocástico  $W = \{W_t\}_{t \geq 0}$  tal que*

1.  $W_0 = 0$  casi seguramente.
2. Tiene trayectorias continuas.
3. Tiene incrementos independientes.
4. La variable aleatoria  $W_t - W_s$  tiene distribución  $N(0, t - s)$  para  $0 \leq s < t$ .

Pruebas a detalle de la existencia un proceso con estas características se pueden encontrar en [4].

### 5.1.1. Simulación de trayectorias del movimiento Browniano

Desde que los incrementos de un movimiento Browniano siguen una distribución Normal, son independientes y las propiedades de la distribución Normal, podemos simular una trayectoria de un movimiento Browniano en un intervalo de tiempo  $[0, T]$  con  $T > 0$  usando el siguiente algoritmo

**Algoritmo 5.1.** Dividimos al intervalo  $[0, T]$  en  $n$  subintervalos de la misma longitud  $\Delta = t_i - t_{i-1}$  para  $i = 1, \dots, n$  donde  $0 = t_0 < t_1 < \dots < t_n = T$ . Hacemos  $i = 0$ ,  $W_i = 0$  y luego

1. Generamos  $x \sim N(0, 1)$ .
2.  $i = i + 1$ .
3. Hacemos  $W_i = W_{i-1} + x\sqrt{\Delta}$ .
4. Si  $i = n$  paramos, en otro caso regresamos al paso 1.

El Código 5.1 genera trayectorias de un movimiento Browniano.

#### Código 5.1.

```

1 #####
2 #Drawn Paths of Brownian Motion
3 #####
4 BM=function(n,Delta)
5
6     trayMB=zeros(Float64,(n+1))
7     trayMB[1]=0
8     times=[0:Delta:T]
9
10    for i=1:n
11        trayMB[i+1]=trayMB[i]+rand(Normal(0,1))*Delta
12    end
13
14    plot(times, trayMB)
15    ylabel("W");
16    xlabel("Tiempo")
17
18    return(trayMB)
19
20 end

```

En la Figura ?? podemos ver una trayectoria de un movimiento Browniano con  $T = 1$  y  $n = 1,000$ .

### 5.1.2. Aproximando al movimiento Browniano por una caminata aleatoria

En la Figura ?? podemos ver una trayectoria de un caminata aleatoria con  $T = 1$ , y cuando  $n = 10$ ,  $n = 100$  y  $n = 1,000$ . Podemos ver la convergencia conforme  $n$  crece.

Consideremos una sucesión  $\{X_i\}_{i=1}^n$  de variables aleatorias independientes tales que

$$\Pr[X_i = 1] = \Pr[X_i = -1] = 0.5.$$

Si definimos

$$S_n = \sum_{i=1}^n X_i,$$

entonces si  $n$  crece a  $\infty$  tenemos por el Teorema del Límite Central tenemos que

$$\Pr\left[\frac{S_{[nt]}}{\sqrt{n}} < x\right] \longrightarrow \Pr[W_t < x],$$

es decir

$$\frac{S_n}{\sqrt{n}} \xrightarrow{d} N(0, 1).$$

La aproximación de la Figura ?? fue generada usando el Código 5.2 para diferentes valores de  $n$ .

### Código 5.2.

```

1
2 #####
3 #Approx MOVIMIENTO BROWNNIANO CAMINATA ALEATORIA
4 #####
5 AMBCA=function(n,T)
6
7     Delta=T/n
8
9     E=[-1,1]
10    p=[0.5,0.5]
11    p=WeightVec(p)
12    S=cumsum(sample(E,n))
13    times=[0:Delta:T]
14
15
16
17    labels=n*times
18    tl=length(labels)-count(c -> (0<c),labels)
19
20
21    W=zeros(length(labels))
22    W[tl+1:end]=S[labels[tl+1]:labels[end]]
23    W=W/sqrt(n)
24    plot(times,W)
25    ylabel("CA");
26    xlabel("Tiempo")
27
28    return(W)
29
30
31 end

```

**Nota 5.1.** Es importante notar que es remondable usar el Algoritmo 5.1 cuando sólo se necesita simular una posición específica de un movimiento Browniano, por ejemplo, en la evaluación del *payoff* de un subyacente en una opción Europea. Por otro lado, si se requiere conocer información sobre la evolución de la trayectoria del proceso, se pueden utilizar cualquiera de las formas de simulación presentadas, por ejemplo, en la valuación de opciones Americanas, Asiáticas y exóticas.

### 5.1.3. Puente Browniano

**Definición 5.2.** Se define como puente Browniano al movimiento Browniano  $\{W_t^p, t \in [t_0, T] : W_{t_0} = x, W_T = y\}$  que empieza en  $x$  al tiempo  $t_0$  y termina en  $y$  al tiempo  $T$  y está dado por

$$W_t^p = x - W_{t-t_0} - \frac{t-t_0}{T-t_0}(W_{T-t_0} - y + x).$$

El Código 5.3 es una implementación para simular trayectorias de un puente Browniano usando la Definición 5.2.

#### Código 5.3.

```

1 #####
2 #Puente Browniano de t0 a T
3 #####
4 BBridge=function(x,y,n,t0,T)
5
6     Delta=(T-t0)/n
7     t0=t0/n
8
9     bridge=zeros(n+1)
10
11
12     W=BM(n,Delta)
13
14     for i=1:n+1
15         bridge[i]=x+W[i]-((i-1)/n)*(W[n+1]-y+x)
16     end
17
18     times=[0:Delta:T]
19     plot(times,bridge)
20     ylabel("Puente Browniano ");
21     xlabel("Tiempo")
22
23     return(bridge)
24
25
26 end

```

En la Figura 5.1 presentamos algunas trayectorias de un puente Browniano  $\{W_t^p, t \in [0, 1] : W_{t_0} = 0, W_T = 0\}$ .

## 5.2. Ecuaciones diferenciales estocásticas

Es común modelar la evolución en el tiempo de un sistema por ecuaciones diferenciales, las cuales usualmente describen la razón de cambio del sistema. Si denotamos por  $X_t$  el estado del sistema al tiempo  $t$  para  $t \geq 0$ , podemos caracterizar la evolución del sistema en el tiempo por la siguiente ecuación,

$$\frac{d(X_t)}{dt} = a(t)X_t, \quad (5.1)$$

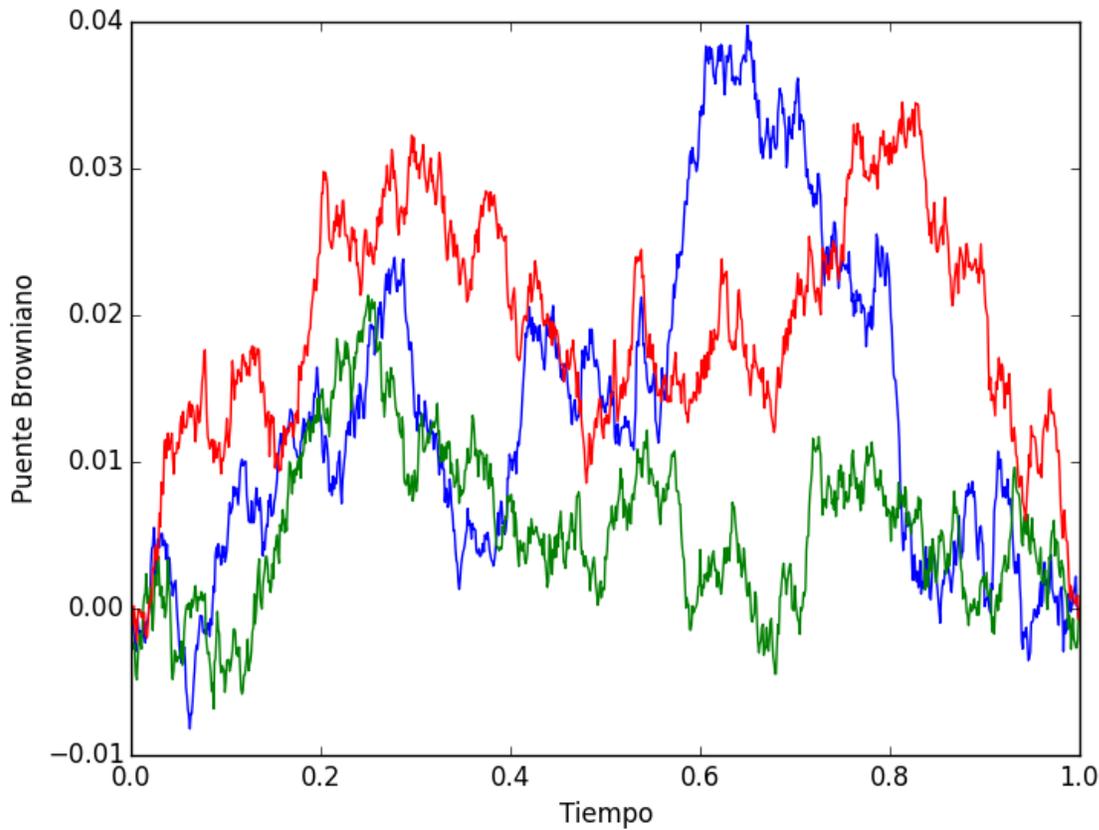


Figura 5.1: Trayectorias de puente Browniano de 0 a 0

donde  $X_0 = x_0$  es el estado inicial del sistema y  $a(t)$  representa la función de cambio al tiempo  $t$ . Este modelo es adecuado cuando la función de cambio es completamente conocida. En muchos casos esta función no es completamente conocida, pero es razonable asumir que está sujeta a algún tipo de efecto aleatorio, entonces tenemos que

$$a(t) = b(t) + \text{ruido}, \quad (5.2)$$

donde no es totalmente conocido el comportamiento del término ruido, pero sabemos que está regido por una distribución de probabilidad. Entonces una forma (heurística) más general de escribir a la ecuación (5.1) es:

$$\frac{d(X_t)}{dt} = b(t, X_t) + \sigma(t, X_t) \times \text{ruido} \quad (5.3)$$

donde  $b, \sigma$  son funciones reales conocidas. Para modelar el término de ruido se utilizan los incrementos de un proceso de Wiener. Si la evolución del sistema es observado en el intervalo de tiempo  $[0, T]$  y consideramos una partición de este intervalo,  $0 = t_0 < t_1 < \dots < t_n = T$ , podemos escribir a la ecuación (5.3) en forma de incrementos como sigue

$$X_{i+1} - X_i = b(t_i, X_i)\Delta t_i + \sigma(t_i, X_i)\Delta W_i, \quad (5.4)$$

donde  $X_i = X_{t_i}$ ,  $\Delta W_i = W_{t_{i+1}} - W_{t_i}$  y  $\Delta t_i = t_{i+1} - t_i$ . De esta forma, utilizando la ecuación (5.3) podemos describir la evolución del proceso  $\mathbf{X} = \{X_t\}_{t=0}^T$  por:

$$X_t = X_0 + \sum_{i=1}^{n-1} b(t_i, X_i)(\Delta t_i) + \sum_{i=1}^{n-1} \sigma(t_i, X_i) \times (\Delta W_i). \quad (5.5)$$

Dado que esta ecuación depende de la partición del intervalo, el modelo resulta ser más adecuado en el límite  $\Delta t_i \rightarrow 0$ . Esto nos lleva a la siguiente ecuación

$$X_t = X_0 + \int_0^t b(t_i, X_i) ds + \int_0^t \sigma(t_i, X_i) dW_s, \quad (5.6)$$

donde el último término es conocido como integral estocástica. Podemos escribir la ecuación anterior en forma diferencial

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t \quad (5.7)$$

para toda  $t \geq 0$  y  $X_0 = x_0$ , a ecuaciones de este tipo se les conoce como ecuaciones diferenciales estocásticas.

La ecuación (5.6) involucra dos tipos de integrales:

$$\int_0^t f(t_i, X_i) ds,$$

que es una integral de Riemann y

$$\int_0^t f(t_i, X_i) dW_s$$

llamada **integral de Itô** o integral estocástica respecto al proceso de Wiener y puede ser interpretada de la siguiente forma.

**Definición 5.3.** Sea  $f(t, x)$  una función continua en  $t$  y  $x$ , tal que

$$\int_0^T \mathbb{E}(f^2(t, X_t)) dt < \infty,$$

entonces la integral de Itô de  $f(t, X_t)$  con respecto a  $W_t$  está definida como

$$\int_0^T f(t_i, X_i) dW_t = \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} f(t_i, X_i) \Delta W_i, \quad (5.8)$$

donde  $0 = t_0 < t_1 < \dots < t_n = T$  entendiendo el límite en el sentido que la distancia máxima entre dos puntos de la partición tiende a cero.

Para que la definición anterior tenga sentido se necesita que el proceso  $\mathbf{X}$  sea adaptado a la filtración natural del proceso de Wiener.

**Proposición 5.1. Propiedades de la integral de Itô** Si  $\mathbf{X}$  es Itô integrable entonces:

1.  $\mathbb{E}[\int_0^T X_s dW_s] = 0$ .
2.  $Var[\int_0^T X_s dW_s] = \int_0^T \mathbb{E}[X_s^2] dW_s$ , **isometría de Itô**.

3. Si  $\mathbf{X}$  y  $\mathbf{Y}$  son Itô integrables y  $a, b$  son dos constantes, entonces

$$\int_0^T (aX_s + bY_s)dW_s = a \int_0^T X_s dW_s + b \int_0^T Y_s dW_s,$$

*linealidad.*

4.  $\int_0^T aW_s dW_s = aW(T).$

5.  $\int_0^T W_s dW_s = \frac{W_T^2}{2} - \frac{T}{2}.$

6. El proceso  $M_t = M_0 + \int_0^T X_s dW_s$  es una martingala.

Una prueba de estas propiedades se puede consultar en [15].

### 5.3. Procesos de difusión

**Definición 5.4.** Un *proceso de difusión*  $X = \{X_t\}_{t \geq 0}$  es la solución a una ecuación de la forma (5.6) o (5.7), donde los coeficientes  $b(t, x)$  y  $\sigma(t, x)$  son funciones reales de  $t$  y  $x$ , y se les conoce como *coeficiente de deriva* y *difusión* respectivamente. Denotemos por  $I = [r, s]$  al espacio de estados.

En la definición anterior  $b(t, x)$  y  $\sigma(t, x)$  pueden ser interpretados como la media y varianza infinitesimal de los incrementos del proceso  $X$ , es decir, si  $\Delta h X_t = X_{t+h} - X_t$  y  $x \in I$ , entonces

$$\lim_{h \downarrow 0} \frac{1}{h} \mathbb{E}[\Delta h X_t | X_t = x] = b(t, x)$$

y

$$\lim_{h \downarrow 0} \frac{1}{h} \mathbb{P}[(\Delta h X_t)^2 | X_t = x] = \sigma(t, x)$$

Al proceso de difusión se le interpreta como el estado del sistema que evoluciona de manera determinista gobernado por la parte no aleatoria de la ecuación pero perturbado por un ruido aditivo dado por la integral estocástica.

Para que una ecuación de la forma (5.7) tenga alguna solución se deben imponer condiciones a sus coeficientes. Existen resultados de existencia y unicidad para ecuaciones diferenciales estocásticas que establecen condiciones de regularidad para sus coeficientes. El siguiente resultado es básico para dichas condiciones.

**Teorema 5.1.** Si los coeficientes  $b(t, x)$  y  $\sigma(t, x)$  de la ecuación (5.7) satisfacen la condición de Lipschitz para toda  $x \in \mathbb{R}$ ,

$$|b(t, x) - b(t, y)|^2 + |\sigma(t, x) - \sigma(t, y)|^2 \leq K|x - y|^2,$$

y la condición de crecimiento en  $x$

$$|b(t, x)|^2 + |\sigma(t, x)|^2 \leq K(1 + |x|^2),$$

con  $K > 0$  una constante, entonces existe un proceso de difusión  $\mathbf{X} = \{X_t\}_{t \geq 0}$  solución de (5.7) que es adaptado, continuo, uniformemente acotado en  $L^2$  y es único en el sentido de indistinguibilidad<sup>1</sup>.

A esta solución se le conoce como solución fuerte. Una prueba de este resultado se puede encontrar en [15].

<sup>1</sup> $\{X_t\}$  y  $\{Y_t\}$  son indistinguibles si  $P(X_t = Y_t) = 1$  para cada  $t \geq 0$

### 5.3.1. Proceso de difusión ergódicos

Si un proceso de difusión es ergódico entonces, para cualquier función medible  $f$ , se tiene que con probabilidad 1:

$$\frac{1}{T} \int_0^T f(X_s) ds \rightarrow \int_{\mathbb{R}} f(x) \pi(x) dx = \mathbb{E}[h(U)],$$

donde  $\pi$  es la densidad estacionaria del proceso de difusión y  $U$  es una varibale aleatoria con densidad  $\pi$ .

**Nota 5.2.** Si un proceso de difusión  $\mathbf{X}$  tiene distribución estacionaria, ésta está dada por

$$\pi(x) = \frac{m(x)}{M}, \quad (5.9)$$

donde

$$m(x) = \frac{1}{\sigma^2(x)s(x)}, \quad (5.10)$$

es la medida de velocidad,

$$s(x) = \exp \left[ -2 \int_{x_0}^x \frac{b(y)}{\sigma^2(y)} dy \right], \quad (5.11)$$

es la medida de escala y  $M$  la constante de normalización.

### 5.3.2. Ecuaciones de Kolmogorov

Dado que los procesos de difusión tienen la propiedad de Markov, tiene sentido definir las probabilidades de transición del estado  $x$  al tiempo  $s$  al estado  $y$  en un tiempo posterior  $t$  ( $0 \leq s < t$ ). Denotamos esas propabilidades por

$$p(t-s, y|x) := \mathbb{P}[X_t = y | X_s = x].$$

Las probabilidades de transición satisfacen las ecuaciones de Kolmogorov (ver [15]), **forward**,

$$\frac{\partial p(t-s, y|x)}{\partial t} = -b(x) \frac{\partial p(t-s, y|x)}{\partial x} + \frac{1}{2} \sigma(x) \frac{\partial^2 p(t-s, y|x)}{\partial x^2} \quad (5.12)$$

y **backward**

$$\frac{\partial p(t-s, y|x)}{\partial s} = -b(x) \frac{\partial p(t-s, y|x)}{\partial x} - \frac{1}{2} \sigma(x) \frac{\partial^2 p(t-s, y|x)}{\partial x^2}. \quad (5.13)$$

Una relación muy importante para la simulación de trayectorias de procesos de difusión y a su vez para hacer inferencia, es obtenida de la ecuación 5.12 cuando  $t \rightarrow -\infty$ , entonces tenemos que

$$\frac{d^2}{dx^2} (\sigma^2(x) \pi(x)) = 2 \frac{d}{dx} (b(x) \pi(x)).$$

De esta ecuación es fácil obtener

$$b(x) = \frac{1}{dx^2} (2\pi(x)) \frac{d}{dx} (\sigma^2(x) \pi(x)),$$

e integrando tenemos que

$$\sigma^2(x) = \frac{2}{\pi(x)} \int_0^x b(y) \pi(y) dy.$$

### Generador infinitesimal

Si el proceso  $\mathbf{X}$  es solución a la ecuación (5.5) con  $X_0 = x$ , al operador diferencial  $\mathcal{L}$  definido por

$$(\mathcal{L}f)(x) := \frac{\sigma^2(x)}{2} f''(x) + b(x) f'(x) \quad (5.14)$$

para  $f$  dos veces diferenciable, es llamado el generador infinitesimal del proceso de difusión  $\mathbf{X}$ .

#### 5.3.3. Puentes de difusión

**Definición 5.5.** *A una solución de la ecuación 6.2 en el intervalo  $[t_1, t_2]$  ( $0 \leq t_1 < t_2 \leq \infty$ ) tal que  $X_{t_1} = a$  y  $X_{t_2} = b$ , con  $a, b \in I$ , se le llama puente de difusión **puente**  $(t_1, a, t_2, b)$ .*

**Ejemplo 5.1.** *Consideremos el proceso Ornstein-Uhlenbeck que es solución a la ecuación*

$$dX_t = -\alpha X_t dt + \sigma dW_t,$$

*condicionado a  $X_0 = a$  y  $X_1 = b$  para  $a, b \in \mathbb{R}$ . Desde que las densidades de transición de este proceso siguen una ley Gaussiana es posible calcular las densidades de transición del puente de difusión  $(0, a, 1, b)$  (ver [1]).*

Por otro lado, en [3] proponen un algoritmo para generar el puente de difusión  $(0, a, 1, b)$  de forma exacta pero que resulta numericamente más estable. El algoritmo se presenta en el siguiente lema.

**Lema 5.1.** *Consideremos los momentos en el tiempo  $0 = t_0 < t_1 \leq t_2 < \dots < t_n < t_{n+1} = 1$  y sea  $X_0 = a$ . Si*

$$X_{t_i} = e^{-\alpha(t_i - t_{i-1})} X_{t_{i-1}} + W_i$$

*para  $i = 1, \dots, n$  y  $W_i \sim N(0, \sigma^2(1 - e^{-2\alpha(t_i - t_{i-1})})/(2\alpha))$  y definimos*

$$Z_{t_i} = X_{t_i} + (b - X_{t_{n+1}}) \frac{e^{\alpha t_i} - e^{-\alpha t_i}}{e^{\alpha t_{n+1}} - e^{-\alpha t_{n+1}}}$$

*para  $i = 0, \dots, n+1$ . Entonces  $\{Z_{t_i}\}_{i=0}^{n+1}$  tiene la misma distribución que un puente de difusión Ornstein-Uhlenbeck  $(0, a, 1, b)$ .*

## 5.4. Fórmula de Itô

Otro resultado importante en teoría de procesos de difusión es la fórmula de Itô, que ayuda a encontrar soluciones a ecuaciones diferenciales estocásticas.

**Teorema 5.2. (Fórmula de Itô)** *Si  $X = \{X_t\}_{t \geq 0}$  es un proceso de difusión dado por la ecuación (5.7) y  $f(t, x)$  una función de clase  $C^1$  como función de  $t$  y de clase  $C^2$  como función de  $x$ , entonces el proceso  $Y_t = f(t, X_t)$  es también un proceso de difusión y es solución a la ecuación*

$$dY_t = f_t(t, X_t) dt + f_x(t, X_t) dX_t + \frac{1}{2} f_{xx}(t, X_t) (dX_t)^2. \quad (5.15)$$

En la ecuación (5.15) los subíndices denotan las derivadas, los detalles de las demostraciones se pueden ver en [15].

**Ejemplo 5.2. Movimiento Browniano Geométrico o modelo Black-Scholes-Merton.** Sea

$$S_t = S_0 \exp \left( (r - \sigma^2/2)t + \sigma W_t \right) \quad (5.16)$$

y si hacemos

$$f(t, x) = S_0 \exp \left( (r - \sigma^2/2)t + \sigma x \right),$$

entonces  $f(t, W_t) = S_t$  y por lo tanto

$$dS_t = rS_t + \sigma S_t dW_t.$$

Este modelo fue utilizado por Black-Scholes-Merton en el contexto financiero para modelar precios de acciones (ver [13]) donde  $r$  es interpretado como la tasa de interés y  $\sigma$  como la volatilidad de los activos riesgosos. Si  $S_0 = s_0$  y hacemos

$$\eta = \log(s_0) + (r - \sigma^2/2)t$$

y

$$\nu = \sigma^2 t,$$

de 5.16 tenemos que para este proceso las densidades de transición tienen distribución log-normal de parámetros  $\eta$  y  $\nu$ , es decir,

$$p(t - s, y | s_0) = \frac{1}{y\nu\sqrt{2\pi}} \exp \left\{ -\frac{(\log(y) - \eta)^2}{2\nu^2} \right\}.$$

El Código 5.4 simula trayectorias de una difusión determinada por la ecuación 5.2 basado en las probabilidades de transición obtenidas.

**Código 5.4.**

```

1 #####
2 # BG-BSM
3 #####
4 BGBSM=function(x,r,sigma,n,T)
5
6     Delta=T/n
7
8     path=zeros(Float32,n+1)
9     path[1]=x
10
11     for i=2:n+1
12         etha=log(path[i-1])+(r-(sigma^2)/2)*Delta
13         nu=(sigma^2)*Delta
14         path[i]=rand(LogNormal(etha,nu))
15     end
16
17     times=[0:Delta:T]
18     plot(times,path)
19     ylabel("BG-BSM");
20     xlabel("Tiempo")
21
22     return(path)
23
24
25     end

```

La Figura ?? se presentan 3 trayectorias de un movimiento Browniano Geométrico con parámetros  $r = 0.4$  (azul),  $r = 0.5$  (verde) y  $r = 0.6$  (rojo), todos con  $\sigma = 1$ .

**Ejemplo 5.3. Proceso Ornstein-Uhlenbeck o proceso de Vasicek.** Un importante proceso de difusión  $X = \{X_t\}_{t \geq 0}$  que será de gran utilidad en el desarrollo de este trabajo es la solución a la ecuación diferencial estocástica

$$dX_t = (\lambda - \alpha X_t)dt + \sigma dW_t, \quad (5.17)$$

donde  $\alpha, \sigma > 0$  son constantes y  $X_0 = x$ . Utilizando la fórmula de Itô se llega a la siguiente solución explícita de esta ecuación:

$$X_t = \frac{\lambda}{\alpha} + e^{-\alpha t} \left( x - \frac{\lambda}{\alpha} \right) + \sigma \int_0^t e^{\alpha(s-t)} dW_s. \quad (5.18)$$

De la solución tenemos que las probabilidades de transición son Gaussianas de parámetros

$$\eta = \frac{\lambda}{\alpha} + e^{-\alpha t} \left( x - \frac{\lambda}{\alpha} \right)$$

y

$$\nu = \frac{\sigma^2(1 - e^{-2\alpha t})}{2\alpha}.$$

Con base a las densidades de transición podemos simular trayectorias de este proceso de difusión. Con el Código 5.5 de podemos generar trayectorias.

#### Código 5.5.

```

1 #####
2 # OU-Vasicek
3 #####
4 Vasicek=function(x,alpha,lambda,sigma,n,T)
5
6     Delta=T/n
7
8     vas=zeros(Float32,n+1)
9     vas[1]=x
10
11
12     for i=2:n+1
13         vas[i]=lambda/alpha+exp(-alpha*Delta)*(vas[i-1]-lambda/alpha)+sigma^2*(1-
14             exp(-2*alpha*Delta))/(2*alpha)*rand(Normal())
15     end
16
17     times=[0:Delta:T]
18     plot(times,vas)
19     ylabel("OU-Vasicek");
20     xlabel("Tiempo")
21
22     return(vas)
23
24 end

```

En la Figura ?? se presenta la simulación de la solución a la ecuación 5.17 con  $\lambda = \alpha = 1$  y tres diferentes valores para  $\sigma = 0.5$  (rojo),  $\sigma = 1.0$  (azul) y  $\sigma = 2.0$  (verde).

**Nota 5.3.** Este proceso es ergódico y tiene distribución invariante Gaussiana de parámetros  $\frac{\lambda}{\alpha}$  y  $\frac{\sigma^2}{2\alpha}$ .

**Ejemplo 5.4. Cox-Ingersoll-Ross.** Este proceso de difusión es usado en el ambiente financiero para modelar tasas de interés a corto plazo (ver [6].) y es la solución a la ecuación diferencial estocástica

$$dX_t = (\lambda - \alpha X_t)dt + \sigma\sqrt{X_t}dW_t, \quad (5.19)$$

donde  $X_0 = x_0 > 0$  y  $\alpha, \lambda, \sigma > 0$ . Si  $2\lambda > \sigma^2$  el proceso es estacionario.

La solución a la ecuación 5.19 es:

$$X_t = \left(X_0 - \frac{\lambda}{\alpha}\right)e^{-\alpha t} + \frac{\lambda}{\alpha} + \sigma e^{-\alpha t} \int_0^t e^{\alpha s} \sqrt{X_s} dW_s.$$

### Difusiones en sistemas dinámicos.

Una de las principales características estudiadas en sistemas dinámicos es la estacionalidad de los procesos que describen el equilibrio del sistema. Aquí enunciaremos los procesos de difusión más populares en esta área.

**Ejemplo 5.5. Modelo tipo N.** Este proceso es solución a la siguiente ecuación diferencial estocástica

$$dX_t = r(\theta - X_t)dt + \sqrt{\epsilon}dW_t,$$

con  $r, \epsilon > 0$ .

**Ejemplo 5.6. Modelo tipo N.** Este proceso es solución a la siguiente ecuación diferencial estocástica

$$dX_t = r(\theta - X_t)dt + \sqrt{\epsilon X_t(1 - X_t)}dW_t,$$

con  $r, \epsilon > 0$ .

**Ejemplo 5.7. Modelo tipo N.** Este proceso es solución a la siguiente ecuación diferencial estocástica

$$dX_t = r(\theta - X_t)dt + \sqrt{\epsilon}dW_t,$$

con  $r, \epsilon > 0$ .

### Difusiones en epidemiología

Para tener un panorama general de la teoría matemática en modelos de epidemiología se puede consultar [2]. Supongamos que  $X_t$  representa la fracción de población infectada al tiempo  $t$ , en general, la tasa de cambio es modelada con el proceso de difusión que es la solución a:

$$dX_t = (aX_t(1 - X_t) - bX_t + c(1 - X_t))dt + \sqrt{\epsilon X_t(1 - X_t)}dW_t, \quad (5.20)$$

donde  $a > 0$  es la tasa de transmisión (persona a persona),  $b > 0$  representa la tasa de recuperación y  $c > 0$  es la tasa de transmisión desde una fuente externa y  $\epsilon > 0$ .

Los modelos con Arno

### 5.4.1. Transformación de Lamperti

Una aplicación de la fórmula de Itô muy importante en las áreas de simulación y estimación de parámetros para ecuaciones diferenciales estocásticas es transformación estándar conocida como transformada de Lamperti dada por transformation

$$h(x) = \int_{x^*}^x \frac{1}{\sigma(y)} dy, \quad (5.21)$$

donde  $x^*$  es un elemento arbitrario en el espacio de estados de  $\mathbf{X}$ . Entonces, por la fórmula de Itô

$$Y_t = h_\beta(X_t)$$

satisface la ecuación diferencial estocástica

$$dY_t = \mu(Y_t)dt + dW_t, \quad (5.22)$$

donde

$$\mu(y) = \frac{b(h^{-1}(y))}{\sigma(h^{-1}(y))} - \frac{1}{2}\sigma'(h^{-1}(y)),$$

donde  $\sigma'(x)$  denota la derivada de  $\sigma$  con respecto a  $x$ .

### 5.4.2. Cociente de verosimilitud para procesos de difusión

En el contexto de inferencia estadística para procesos de difusión el **Teorema de Girsanov** es usado para obtener el cociente de verosimilitud para hacer inferencia por máxima verosimilitud.

**Teorema 5.3. Teorema de Girsanov.** Sea  $\mathbb{P}$  la medida de probabilidad inducida en  $([0, T], \mathcal{B}_{\mathbb{R}}([0, T]))$  por  $\mathbf{Y}$ , la solución de la ecuación (5.22) donde  $\mu$  satisface las condiciones de la fórmula de Itô y sea  $\mathbb{Q}$  la medida de Wiener, entonces las medidas son equivalentes y su correspondiente derivada de Radon-Nikodín es

$$\frac{d\mathbb{P}}{d\mathbb{Q}} = \exp \left( \int_0^T \mu(Y_s) dY_s - \frac{1}{2} \int_0^T \mu^2(Y_s) ds \right)$$

Para una prueba de este resultado ver [11].

Sea  $\mathcal{S}$

## 5.5. Ejercicios

1. Encontrar las probabilidades de transición del CIR e implementar un algoritmo para simular trayectorias de este proceso de difusión.
2. Encontrar solución, la distribución estacionaria, las densidades de transición, la media, la moda y la varianza de la difusión del Ejemplo 5.5.
3. Encontrar solución, la distribución estacionaria, las densidades de transición, la media, la moda y la varianza de la difusión del Ejemplo 5.6.
4. Encontrar solución, la distribución estacionaria, las densidades de transición, la media, la moda y la varianza de la difusión del Ejemplo 5.7.



# Capítulo 6

## Inferencia estadística para ecuaciones diferenciales estocásticas

### 6.1. Simulación de procesos de difusión

En esta sección presentaremos los dos métodos más importantes para la simulación de trayectorias a soluciones de ecuaciones diferenciales estocásticas. Estos métodos son utilizados cuando no se conoce explícitamente las probabilidades de transición o aún cuando se conocen resulta difícil de simular. Debido a que estos métodos están basados en aproximaciones a tiempo discreto a los procesos en tiempo continuo es importante empezar definiendo el tipo de convergencia de los procesos discretos a los continuos.

**Definición 6.1. Convergencia Fuerte.** Se dice que una aproximación a tiempo discreto  $\mathbf{X}^\delta$  de un proceso a tiempo continuo  $\mathbf{X}$ , donde  $\delta$  denota el incremento de tiempo máximo en la discretización, se dice que es de nivel fuerte de convergencia  $\tau$  si

$$E[|X_t^\delta - X_t|] \leq C\delta^\tau$$

para todo momento  $t \geq 0$ ,  $\delta < \delta_0$  con  $\delta_0 > 0$  y  $C$  una constante independiente de  $\delta$ .

**Definición 6.2. Convergencia Débil.** Considerando otra vez la aproximación a tiempo discreto  $\mathbf{X}^\delta$  de un proceso a tiempo continuo  $\mathbf{X}$ , se dice que es de nivel débil de convergencia  $\tau$  si

$$|E[g(X_t^\delta)] - E[g(X_t)]| \leq C\delta^\tau$$

para todo  $t \geq 0$ ,  $\delta < \delta_0$  con  $\delta_0 > 0$ ,  $C$  una constante independiente de  $\delta$ , cualquier función polinómica creciente continua y diferenciable de orden  $2(\tau + 1)$ .

#### 6.1.1. Método de Euler

Consideremos un proceso de difusión  $\mathbf{X} = \{X_t\}_{t \geq 0}$  que es solución a la ecuación diferencial estocástica

$$dX_t = b(X_t, t)dt + \sigma(X_t, t)dW_t$$

con valor inicial determinista  $X_{t_0} = x_0$  y un nivel de discretización

$$0 = t_0 < t_1 < \dots < t_n = T$$

en el intervalo  $[0, T]$ . La aproximación de Euler de  $X$  es un proceso estocástico continuo  $Y$  que satisface el siguiente sistema iterativo:

$$Y_{t_{i+1}} = Y_{t_i} + b(Y_{t_i}, t_i)\Delta_i + \sigma(Y_{t_i}, t_i)\Delta W_i$$

para  $i = 0, 1, \dots, n-1$ , con  $Y_0 = x_0$  y

$$\begin{aligned}\Delta_i &= t_{i+1} - t_i, \\ \Delta W_i &= W_{t_{i+1}} - W_{t_i}\end{aligned}$$

entonces  $\Delta W_i \sim N(0, \Delta_i)$ .

Por otro lado como entre cualesquiera dos momentos de observación  $t_i$  y  $t_{i+1}$ , el proceso puede ser definido diferenciable. Una propuesta natural es considerar una interpolación lineal para  $Y_t$  definida por

$$Y_t = Y_{t_i} + \frac{t - t_i}{t_{i+1} - t_i}(Y_{t_{i+1}} - Y_{t_i})$$

para  $t \in [t_i, t_{i+1})$ .

### Código 6.1.

```

1  #
   #####
2  #Drawn Paths of diffusion processes by Euler Scheme
3  #
   #####
4
5  function Euler(theta, init, delta, n)
6
7      brow_inc=rand(Normal(0, sqrt(delta)), n)
8      tray=zeros(Float64, (n+1))
9      tray[1]=init
10
11     for i=1:n
12         tray[i+1]=tray[i]+b(tray[i], theta)*delta+sigma(tray[i], theta)*brow_inc[
           i]
13     end
14
15     tray
16
17 end

```

**Ejemplo 6.1.** Aquí, se presenta una simulación de un proceso de difusión con el esquema de Euler, consideramos el proceso de Ornstein-Uhlenbeck, que es una solución de la ecuación diferencial estocástica

$$dX_t = -\alpha X_t dt + \sigma dW_t,$$

donde  $\alpha > 0$  y  $\sigma > 0$  son los parámetros del proceso  $W$  es un proceso de Wiener estándar. Si suponemos que  $X_0 = 0$  y consideremos un intervalo de realización  $[0, 10]$  con un nivel de discretización  $0 = t_0 < t_1 < \dots < t_n = 10$ , donde  $n = 100$  y  $\Delta = \Delta_i = 0.1$ , para todo  $i = 1, 2, \dots, n - 1$ , entonces el método de Euler para este proceso es:

$$Y_{t_{i+1}} = Y_{t_i} - \alpha Y_{t_i} \Delta + \sigma \Delta W.$$

En la Figura ?? podemos ver trayectorias características de un Ornstein-Uhlenbeck con diferentes parámetros  $\alpha = 1, \sigma = 1$  (azul),  $\alpha = 1.5, \sigma = 0.3$  (rojo) y  $\alpha = 3.5, \sigma = 0.3$  (verde). [Simular de las transiciones y comparar los métodos qqplots](#)

### 6.1.2. Método de Milstein

En la aproximación de Milstein se utiliza el lema de Itô para aumentar la precisión en la aproximación mediante la incorporación del término de segundo orden, podemos escribir a esta aproximación por:

$$Y_{t_{i+1}} = Y_{t_i} + b(Y_{t_i}, t_i) \Delta_i + \sigma(Y_{t_i}, t_i) \Delta W_i + \frac{1}{2} \sigma(Y_{t_i}, t_i) \sigma'(Y_{t_i}, t_i) t_i [(\Delta W_i)^2 - \Delta_i],$$

**Código 6.2.**

```

1  #
   #####
2  #Drawn Paths of diffusion processes by Milstein Scheme
3  #
   #####
4
5  function Milstein(theta, init, delta, n)
6
7      brow_inc=rand(Normal(0, sqrt(delta)), n)
8      tray=zeros(Float64, (n+1))
9      tray[1]=init
10
11     for i=1:n
12         tray[i+1]=tray[i]+b(tray[i], theta)*delta+sigma(tray[i], theta)*brow_inc[
           i]+(1/2)*sigma(tray[i], theta)*sigma_x(tray[i], theta)*(brow_inc[i]^2-
           delta)
13     end
14
15     tray
16
17 end

```

**Ejemplo 6.2.** Para ejemplificar el uso de este método de simulación consideremos el movimiento Browniano geométrico, el cual es solución a la ecuación diferencial estocástica

$$dX_t = \theta_1 X_t dt + \theta_2 X_t dW_t,$$

para este proceso, tenemos que  $b(x, t) = \theta_1 x$ ,  $\sigma(x, t) = \theta_2 x$  y  $\sigma_x(x, t) = \theta_2$  entonces el esquema de Milstein para este proceso es

$$\begin{aligned} Y_{t_{i+1}} &= Y_{t_i} + \theta_1 Y_{t_i} \Delta_i + \theta_2 \Delta Y_{t_i} W_i + \frac{1}{2} \theta_2 Y_{t_i} [(\Delta W_i)^2 - \Delta_i] \\ &= Y_{t_i} \left(1 + \Delta\theta_1 - \frac{\theta_2^2}{2}\right) + \theta_2 \Delta W_i + \frac{\theta_2^2}{2} Y_{t_i} (\Delta W_i)^2 \end{aligned}$$

En la Figura ?? podemos ver trayectorias características de un Movimiento Browniano Geométrico con diferentes parámetros  $\theta_1 = 1, \theta_2 = 0.1$  (azul),  $\theta_1 = 1, \theta_2 = 1$  (rojo) y  $\theta_1 = 0.1, \theta_2 = 0.1$  (verde).

Simular de las transiciones y comparar los métodos qqplots

### 6.1.3. Simulación de puentes de difusión

En esta sección presentaremos un algoritmo para simular puentes de difusión presentado en [3].

Sea  $\mathbf{X}$  un proceso de difusión solución de 6.2, con espacio de estados  $I = (r, s)$  y que los coeficientes satisfacen las condiciones del teorema 5.1. Sea  $a, b \in I$  y  $W^1$  y  $W^2$  dos procesos de Wiener independientes y  $\mathbf{X}^1$  y  $\mathbf{X}^2$  soluciones de

$$dX_t^i = b(X_t^i)dt + \sigma(X_t^i)dW_t^i \quad (6.1)$$

$i = 1, 2$ ,  $X_0^1 = a$  y  $X_0^2 = b$ . En el siguiente teorema utilizaremos estos dos procesos para aproximar el puente  $(0, a, \Delta, b)$

**Teorema 6.1.** Sea  $\tau = \inf\{0 \leq t \leq \Delta \mid X_t^1 = X_{\Delta-t}^2\}$  y

$$Z_t = \begin{cases} X_t^1 & \text{si } 0 \leq t \leq \tau \\ X_{\Delta-t}^2 & \text{si } \tau < t \leq \Delta \end{cases}$$

Entonces la distribución de  $Z = \{Z_t\}_{0 \leq t \leq \Delta}$ , condicionado a  $\{\tau \leq \Delta\}$ , es igual a la distribución de un puente  $(0, a, \Delta, b)$ , condicionado al evento que el puente es alcanzado por una difusión (independiente) con EDE (6.2) y distribución inicial con densidad  $p_\Delta(b, *)$ .

La prueba de este resultado se puede consultar a detalle en [3].

OU con este algoritmo y comparar con el exacto

### 6.1.4. Estimadores Monte-Carloy técnicas de reducción de varianza

Usando los métodos de Monte-Carlo se puede estimar algunas cosas, poner ejemplos financieros (precios de opciones). Revisar las técnicas de Malliavin en [14] y el método presentado en [12].

## 6.2. Estimación por máxima verosimilitud

Abordaremos el problema de estimación de parámetros en procesos de difusión cuando éstos han sido observados a tiempo discreto. Consideremos que las difusiones son observadas en el intervalo de tiempo  $[0, T]$  en  $n$  momentos  $0 = t_1 < t_2 < \dots < t_n = T$  y el tiempo entre cada observación es igual entre cualesquiera dos observaciones consecutivas, es decir,  $\Delta = t_i - t_{i-1}$  para  $i = 2, \dots, n$  con los siguientes tipos de información:

1. **Muestra grande:** Se mantiene fijo a  $\Delta = \Delta_n$  y se aumenta el número de observaciones, es decir,  $n$  crece. En este caso, aumenta el intervalo de observación  $[0, T]$ . Para este tipo de información se asumirá que los procesos de difusión son estacionarios y/o ergódicos.
2. **Alta frecuencia:** Se hace tender a cero a  $\Delta$  conforme aumenta  $n$  y el intervalo de observación  $[0, T]$  se mantiene fijo.
3. **Crecimiento rápido:**  $\Delta = \Delta_n$  decrece a cero cuando  $n$  crece y además el intervalo de observación  $[0, T]$  crece como función de  $n$ , es decir,  $n\Delta_n \rightarrow \infty$ . Para este caso también se asuma que los procesos de difusión son estacionarios y/o ergódicos.

### 6.2.1. Modelo Continuo

Consideremos un proceso de difusión  $X = \{X_t\}_{t \geq 0}$  que es solución a la ecuación diferencial estocástica

$$dX_t = b(X_t, \theta)dt + \sigma(X_t, \theta)dW_t, \quad (6.2)$$

donde  $\theta \in \Theta \subset \mathbb{R}$  es un parámetro multidimensional y  $\theta_0$  es el verdadero valor del parámetro. Las funciones

$$b : \mathbb{R} \times \Theta \rightarrow \mathbb{R}$$

y

$$\sigma : \mathbb{R} \times \Theta \rightarrow (0, \infty)$$

son conocidas y cumplen con condiciones para la existencia de una solución de (6.2) (ver Teorema 5.1). Denotamos por  $I = (l, r)$  con  $-\infty \leq l < r < \infty$  al espacio de estados y supongamos que es el mismo para toda  $\theta$ . Supongamos la solución  $\mathbf{X}$  de (6.2) tiene valor inicial determinista  $X_{t_0} = x_0$  y que se cumplen las condiciones para la existencia de la distribución estacionaria  $\pi_\theta$ .

Cuando se tiene el proceso observado a tiempo continuo, la estimación del parámetro resulta eficiente. En particular, desde que la variación cuadrática de un proceso de difusión que es solución de (6.2) está dada por (ver [15])

$$\langle X, X \rangle_T = \int_0^T \sigma^2(X_s, \theta) ds, \quad (6.3)$$

entonces,

$$\langle X, X \rangle_T = \int_0^T \sigma^2(X_s, \theta) ds = \lim_{n \rightarrow \infty} \sum_{k=1}^{2^n} (X_{T \wedge k/2^n} - X_{T \wedge (k-1)/2^n})^2$$

en probabilidad bajo la medida  $P_\theta$  (La distribución de  $X$ ). Si asumimos que  $\sigma(x, \theta) = \sigma(x)$  y que el resto de los parámetros están sólo en la deriva, entonces, éstos pueden ser estimados por máxima verosimilitud.

Para el intervalo de tiempo  $[0, T]$  la función de verosimilitud está dada por

$$L_T(\theta) = \exp \left( \int_0^T \frac{\mu(X_s, \theta)}{\sigma^2(X_s)} dX_s - \frac{1}{2} \int_0^T \frac{\mu^2(X_s, \theta)}{\sigma^2(X_s)} ds \right). \quad (6.4)$$

escribir como llevar a esta

**Nota 6.1.** Podemos afirmar que esta situación ocurre con probabilidad cero.

### 6.3. Caso discreto

Suponemos que tenemos el proceso observado a tiempo discreto en el intervalo  $[0, T]$  a los tiempos  $\{0 = t_1 < t_2 < \dots < t_n = T\}$ , es decir, la información que se tiene es  $X_{obs} = \{X_{t_1} = x_1, \dots, X_{t_n} = x_n\}$ . Denotamos por  $\mathcal{F}_n = \sigma\{x_i; 1 \leq i \leq n\}$  a la  $\sigma$ -álgebra generada por las  $n$  observaciones y por  $\mathcal{F}_0$  a la  $\sigma$ -álgebra trivial. Si denotamos a las densidades de transición (entre las observaciones) por  $p_\theta(\Delta, x_i|x_{i-1})$ , por la propiedad de Markov de los procesos de difusión podemos escribir la función de verosimilitud basada en  $X_{obs}$  como

$$L_n(\theta) = \prod_{i=1}^n p_\theta(\Delta, x_i|x_{i-1})p_\theta(x_0) \quad (6.5)$$

y la log-verosimilitud es

$$\ell_n(\theta) = \sum_{i=1}^n \log p_\theta(\Delta, x_i|x_{i-1}) + \log p_\theta(x_0).$$

**Nota 6.2.** *Bajo condiciones de regularidad las probabilidades de transición  $p_\theta(\Delta, x_i|x_{i-1})$  satisfacen las ecuación de Kolmogorov 5.12 y 5.13 pero unicamente en casos muy sencillos se puede resolver estas ecuaciones diferenciales parciales.*

#### 6.3.1. Inferencia por verosimilitud exacta

Consideremos el siguiente conjunto de hipótesis necesarias para asegurar suficiencia y consistencia asintótica de los estimadores máximo verosimiles que presentaremos, para ver los detalles ver [7] y [10].

1. **Creciento lineal.** Existe una constante  $C$  (independiente de  $\theta$ ) tal que, para todo  $x \in I$

$$|\mu(x, \theta)| + |\sigma(x, \theta)| \leq (1 + |x|).$$

2. **Lipschitz.** Existe una constante  $C$  (independiente de  $\theta$ ) tal que, para todo  $x, y \in I$

$$|\mu(x, \theta) - \mu(y, \theta)| + |\sigma(x, \theta) - \sigma(y, \theta)| \leq (|x - y|).$$

3. **Coefficiente de difusión positivo.**

$$\inf_{x \in I} \sigma^2(x, \theta) > 0.$$

4. **Momentos acotados** Para todo  $k > 0$ , todos los momentos de orden  $k$  existen y

$$\sup_{t \in T} \text{Exp}|X_t|^k < \infty.$$

**Nota 6.3.** *Si los parámetros de los coeficientes de deriva y difusión son separables, la tasa de convergencia es más rápida en el parámetro de difusión y es posible utilizar la propuesta presentada en [16] para encontrar los estimadores. Generalmente la tasa de convergencia del coeficiente de difusión es  $\sqrt{n}$  y la de la deriva  $\sqrt{n}\Delta_n$  (condicionado a  $n\Delta_n^3 \rightarrow 0$ ).*

**Ejemplo 6.3.** *Ornstein-Uhlenbeck o Vasicek*

Large sample, $\Delta = 1$						
n	$\alpha = 1$	2.5 % - 97.5 %	$\beta = 3$	2.5 % - 97.5 %	$\sigma = 2$	2.5 % - 97.5 %
100	1.2016	(0.71783,2.1760)	3.5340	(2.05939,6.4328)	2.0797	(1.68729,2.8068)
200	0.99640	(0.69757,1.4247)	3.19439	(2.19813,4.6007)	2.05284	(1.779,2.4380)
500	1.0033	(0.80208,1.2556)	3.057849	(2.42023,3.848)	2.038970	(1.85789,2.2646)
1000	1.0111	(0.86386,1.1839)	3.1094	(2.63914,3.6573)	2.0143	(1.88388,2.1667)

Cuadro 6.1:

High-frequency, $T = 100$						
$\Delta$	$\alpha = 1$	2.5 % - 97.5 %	$\beta = 3$	2.5 % - 97.5 %	$\sigma = 2$	2.5 % - 97.5 %
1	0.9326945	(0.55658,1.5430)	2.91223	(1.68594,4.8569)	1.99326	(1.4897,2.548636)
1/2	0.8664383	(0.56497,1.2216)	2.74615	(1.71730,3.9431)	2.01957	(1.7898,2.2935)
1/5	0.9653574	(0.67449,1.2742)	2.99912	(2.02426,4.0291)	1.97410	(1.8987,2.1169)
1/10	1.0138	(0.72397,1.3124)	2.7642	(1.88794,3.6641)	1.94155	(1.85554,2.0347)

Cuadro 6.2:

Consideremos la difusión  $\mathbf{X}$  solución de 5.17 wcon  $X_0 = x_0$ ,  $\lambda, \alpha \in \mathbb{R}$ ,  $\sigma \in R^+$  y es ergódico si  $\alpha > 0$ . En este caso, la densidad condicional  $p_\theta(\Delta, x_i | x_{i-1})$  tiene distribución Gaussiana de parámetros

$$X_{t+\Delta} | X_t \sim N \left( \frac{\beta}{\alpha} + (x_t - \frac{\beta}{\alpha}) e^{-\alpha\Delta}, \sigma^2 \frac{1 - e^{-2\alpha\Delta}}{2\alpha} \right).$$

Entonces, tenemos una expresión explícita para (6.5) y podemos encontrar los estimadores máximo verosímiles.

El proceso de Ornstein-Uhlenbeck, en la versión de dos parámetros, definido por la ecuación diferencial estocástica

$$dX_t = -\alpha X_t dt + \sigma dW_t,$$

tiene una forma explícita para los estimadores máximo verosímil dado por :

$$\hat{\alpha} = -\frac{1}{\Delta} \log \left( \frac{\sum_{i=1}^n X_{i-1} X_i}{\sum_{i=1}^n X_{i-1}^2} \right)$$

definido siempre que ocurra  $\sum_{i=1}^n X_{i-1} X_i > 0$ , y

$$\hat{\sigma}^2 = \frac{2\hat{\alpha}}{n(1 - e^{-2\Delta\hat{\alpha}})} \sum_{i=1}^n (X_i - X_{i-1} e^{-\Delta\hat{\alpha}})^2.$$

## 6.4. Ejercicios

1. Mostrar que el método de Euler tiene orden fuerte de convergencia con  $\tau = 1/2$ .
2. Mostrar que esta aproximación tiene convergencia débil y fuerte de orden uno.

High-frequency							
$\Delta$	$n$	$\alpha = 1$	2.5 % - 97.5 %	$\beta = 3$	2.5 % - 97.5 %	$\sigma = 2$	2.5 % - 97.5 %
1/1000	1000	3.9832	(-1.2793, 9.2737)	8.0231	(-1.9707, 18.0656)	2.0301	(1.9442, 2.1227)

Cuadro 6.3:



## Apéndice A

# Simulación de variables aleatorias

**Definición A.1.** Sea  $X$  variable aleatoria con soporte en  $E = \{x_1, x_2, \dots, x_n\}$  con  $x_1 < x_2 < \dots < x_n$  y  $f$  su función de densidad asociada  
Se define

$$\phi(y) = x_{\min\{i; \sum_{k=1}^i f(x_k) \geq y\}}.$$

Se dice que  $\phi$  es la función de cuantiles asociada a la variable aleatoria  $X$ .

**Lema A.1.** Considérese  $U$  una v.a. uniforme en  $(0, 1)$  y  $X$  una v.a. con soporte en  $E = \{x_1, x_2, \dots, x_n\}$  con  $x_1 < x_2 < \dots < x_n$ , con función de distribución  $f$  y con función de cuantiles  $\phi$ , entonces se tiene que  $\phi(U)$  tiene la misma distribución que  $X$ .

*Demostración.*

$$\begin{aligned} \mathbb{P}[\phi(U) \leq x_j] &= \mathbb{P}[\min\{i; \sum_{k=1}^i f(x_k) \geq U\} \leq j]. \\ &= \mathbb{P}[U \leq \sum_{k=1}^j f(x_k)]. \\ &= \sum_{k=1}^j f(x_k). \end{aligned}$$

por lo tanto  $\phi(U)$  tiene distribución  $f$ . □



## Apéndice B

# Convergencia de variables aleatorias y el Teorema central del límite.

En este apéndice se presentan los modos de convergencia de vectores aleatorios, así como también sus respectivas relaciones y los principales resultados que se utilizan a lo largo del trabajo. Este apéndice está basado principalmente en [? ], [? ] y [? ].

### Convergencia casi segura

En algunas situaciones la convergencia puntual resulta ser una condición muy fuerte pues se pide la convergencia de la sucesión evaluada en todos y cada uno de los elementos del espacio muestral. Se puede ser menos estricto y pedir, por ejemplo, que la convergencia se verifique en todo el espacio  $\Omega$  excepto en un subconjunto de probabilidad cero.

**Definición B.1.** (*Convergencia casi segura*). La sucesión de variables aleatorias  $\xi_1, \xi_2, \dots$  converge casi seguramente a la variable  $\xi$ , si  $P\{\omega \in \Omega : \lim_{n \rightarrow \infty} \xi_n(\omega) = \xi(\omega)\} = 1$ .

Es decir, en la convergencia casi segura se permite que para algunos valores de  $\omega$ , la sucesión numérica  $\xi_1(\omega), \xi_2(\omega), \dots$  pueda no converger, sin embargo el subconjunto de  $\Omega$  en donde esto suceda debe tener probabilidad cero. Para indicar la convergencia casi segura se escribe  $\xi_n \xrightarrow{c.s.} \xi$ , o bien  $\lim_{n \rightarrow \infty} \xi_n = \xi$  c.s. A menudo se utiliza el término convergencia *casi donde quiera*, o bien convergencia *casi siempre* para denotar este tipo de convergencia. Observe que omitiendo el argumento  $\omega$ , la condición para la convergencia casi segura se escribe en la forma más corta:  $\mathbb{P}[\lim_{n \rightarrow \infty} \xi_n = \xi] = 1$ , o simplemente  $\mathbb{P}[\xi_n \rightarrow \xi] = 1$ . Es posible demostrar que el conjunto  $\{\omega \in \Omega : \xi_n(\omega) \rightarrow \xi(\omega)\}$  es medible de modo que tiene sentido aplicar la probabilidad. Puede también demostrarse que bajo este tipo de convergencia, el límite es único casi seguramente, es decir, si  $\xi_n$  converge a  $\xi$  c.s. y también converge a  $\eta$  c.s., entonces  $\xi = \eta$  casi seguramente.

### Convergencia en probabilidad

Un tipo de convergencia aún menos restrictiva que la convergencia casi segura es la convergencia en probabilidad la cual se define a continuación.

**Definición B.2.** (*Convergencia en probabilidad*). La sucesión de variables aleatorias  $\xi_1, \xi_2, \dots$  converge en probabilidad a  $\xi$ , si para cada  $\epsilon > 0$ ,  
$$\lim_{n \rightarrow \infty} P\{\omega \in \Omega : |\xi_n(\omega) - \xi(\omega)| > \epsilon\} = 0.$$

Cuando hablamos de vectores solo cambiamos el valor absoluto por la norma euclidiana.

Para denotar la convergencia en probabilidad se escribe  $\xi_n \xrightarrow{p} \xi$ , y omitiendo el argumento  $\omega$  la condición se escribe  $\lim_{n \rightarrow \infty} \mathbb{P}[|\xi_n - \xi| > \epsilon] = 0$ . Nuevamente puede comprobarse que el límite es único casi seguramente.

### Convergencia en media

En este tipo de convergencia se usa la esperanza para determinar la cercanía entre dos variables aleatorias.

**Definición B.3.** (CONVERGENCIA EN MEDIA). La sucesión de variables aleatorias integrables  $\xi_1, \xi_2, \dots$  converge en media a la variable aleatoria integrable  $\xi$  si

$$\lim_{n \rightarrow \infty} \mathbb{E}[|\xi_n - \xi|] = 0.$$

A este tipo de convergencia también se le llama convergencia en  $L^1$  y se le denota por  $\xi_n \xrightarrow{m} \xi$ , o  $\xi_n \xrightarrow{L^1} \xi$ . A partir de la definición de convergencia en media es inmediato preguntarse si de allí se sigue la convergencia de la sucesión de medias. La respuesta es afirmativa.

### Convergencia en media cuadrática

Nuevamente usando el concepto de esperanza pero ahora aplicado al segundo momento se tiene la convergencia en media cuadrática. Más adelante demostraremos que la convergencia en media cuadrática implica la convergencia en media.

**Definición B.4.** (CONVERGENCIA EN MEDIA CUADRÁTICA). La sucesión de variables aleatorias  $\xi_1, \xi_2, \dots$  converge en media cuadrática a  $\xi$ , si

$$\lim_{n \rightarrow \infty} \mathbb{E}[|\xi_n - \xi|^2] = 0.$$

En este tipo de convergencia se presupone que tanto los elementos de la sucesión como el límite mismo son variables aleatorias con segundo momento finito. A este tipo de convergencia también se le llama convergencia en  $L^2$ , y se le denota por  $\xi_n \xrightarrow{m.c.} \xi$ , o  $\xi_n \xrightarrow{L^2} \xi$ .

En general puede definirse la convergencia en  $L^k$ , para cada entero  $k \geq 1$ , cuando se cumple la condición  $\mathbb{E}[|\xi_n - \xi|^k] \rightarrow 0$ . Resulta que mientras mayor es el valor de  $k$ , más restrictiva es la condición de convergencia.

### Convergencia en distribución

Este es el tipo de convergencia menos restrictiva de todas las mencionadas. En contextos más generales se le llama también convergencia débil.

**Definición B.5.** (Convergencia en distribución). La sucesión de variables aleatorias  $\xi_1, \xi_2, \dots$  converge en distribución a  $\xi$ , si para todo punto  $x$  en donde la función  $F_\xi(x)$  es continua, se cumple que

$$\lim_{n \rightarrow \infty} F_{\xi_n}(x) = F_\xi(x).$$

Cuando esto ocurre escribimos  $\xi_n \xrightarrow{d} \xi$  o bien  $\xi_n \xrightarrow{ley} \xi$ .

## Resultados más importantes de los tipos de convergencia

**Proposición B.1.** *Convergencia c.s.  $\implies$  convergencia en probabilidad.*

**Proposición B.2.** *Convergencia en  $L_p \implies$  convergencia en media.*

**Proposición B.3.** *Convergencia en media  $\implies$  convergencia en probabilidad.*

**Proposición B.4.** *Convergencia en probabilidad.  $\implies$  convergencia en distribución.*

Todos los resultados anteriores de esta sección pueden ser consultados en [?] en la parte de Convergencia Capítulo 7.

**Definición B.6.** *Sean  $\xi, \xi_1, \xi_2, \dots$  vectores aleatorios, entonces la sucesión  $\{\xi_n\}_{n \in \mathbb{N}}$  converge casi seguramente (en probabilidad) a  $\xi$  si la sucesión  $\|\xi - \xi_n\|$  converge c.s. (en probabilidad), en tal caso escribimos  $\xi_n \xrightarrow{c.s.} \xi$  ( $\xi_n \xrightarrow{p} \xi$ ).*

**Lema B.1.** *(Criterio de subsucesión) Sean  $\xi, \xi_1, \xi_2, \dots$  vectores aleatorios. Entonces  $\xi_n \xrightarrow{p} \xi$  si y sólo si cada subsucesión  $N' \subset \mathbb{N}$  tiene una subsucesión  $N'' \subset N'$  de tal manera que  $\xi_n \rightarrow \xi$  c.s. a través de la subsucesión  $N''$ . En particular,  $\xi_n \rightarrow \xi$  c.s. implica  $\xi_n \xrightarrow{p} \xi$ .*

Para su demostración ver [?] página 40.

**Proposición B.5.** *(aplicaciones continuas) Sean  $\xi, \xi_1, \xi_2, \dots$  variables aleatorias en  $\Omega$  con  $\xi_n \xrightarrow{p} \xi$  y considera una función  $f : \Omega \rightarrow \mathbb{R}$  de tal manera que  $f$  es c.s. continua. Entonces  $f(\xi_n) \xrightarrow{p} f(\xi)$ .*

*Demostración.* Se deduce fácilmente del Lema B.1. □

**Proposición B.6.** *(Operaciones elementales) Sean  $\xi, \xi_1, \xi_2, \dots$  y  $\eta, \eta_1, \eta_2, \dots$  variables aleatorias con  $\xi_n \xrightarrow{p} \xi$  y  $\eta_n \xrightarrow{p} \eta$ . Entonces  $a\xi_n + b\eta_n \xrightarrow{p} a\xi + b\eta$ , para todo  $a, b \in \mathbb{R}$ , y  $\xi_n \eta_n \xrightarrow{p} \xi\eta$ . Además  $\xi_n/\eta_n \xrightarrow{p} \xi/\eta$  cuando c.s.  $\eta \neq 0$  y  $\eta_n \neq 0$  para toda  $n$ .*

*Demostración.* Se deduce del Lema B.1. □

**Proposición B.7.** *Sean  $\xi, \xi_1, \xi_2, \dots$  y  $\eta_1, \eta_2, \dots$  variables aleatorias con  $\xi_n \xrightarrow{d} \xi$  y  $\eta_n \xrightarrow{c.s.} 0$ . Entonces*

$$\xi_n \eta_n \xrightarrow{p} 0.$$

*Demostración.* Sea  $\epsilon > 0$  y  $\delta > 0$ , sabemos que existe  $N \in \mathbb{N}$  tal que para toda  $n \geq N$  se tiene que  $\mathbb{P}(|\eta_n| > \delta\epsilon) < \delta$  y por tanto

$$\mathbb{P}(|\xi_n \eta_n| > \epsilon) \leq \mathbb{P}(|\xi_n \eta_n| > \epsilon, |\eta_n| \leq \delta\epsilon) + \delta \tag{B.1}$$

$$= \delta + \mathbb{P}(|\xi_n| > \frac{1}{\delta}). \tag{B.2}$$

Sabemos que  $\mathbb{P}(|\xi_n| > \frac{1}{\delta}) \xrightarrow[n \rightarrow \infty]{} \mathbb{P}(|\xi| > \frac{1}{\delta})$ , entonces

$$\limsup_{n \rightarrow \infty} \mathbb{P}(|\xi_n \eta_n| > \epsilon) \leq \delta + \mathbb{P}(|\xi| > \frac{1}{\delta}).$$

Por lo tanto  $\mathbb{P}(|\xi_n \eta_n| > \epsilon) \xrightarrow[n \rightarrow \infty]{} 0$ . □

**Proposición B.8.** Sean  $\hat{X}, \hat{X}_1, \hat{X}_2, \dots$  vectores aleatorios, las siguientes afirmaciones son equivalentes

1.  $F_{\hat{X}_n}(\hat{t}) \rightarrow F_{\hat{X}}(\hat{t})$  para todo punto  $\hat{t}$  de continuidad.
2. Para toda función de valores reales  $f$  uniformemente continua y acotada  $\mathbb{E}[f(\hat{X}_n)] \rightarrow \mathbb{E}[f(\hat{X})]$ .
3. Para toda función de valores reales  $f$  continua y acotada  $\mathbb{E}[f(\hat{X}_n)] \rightarrow \mathbb{E}[f(\hat{X})]$ .

Si alguna de estas condiciones se cumple se dice que  $\hat{X}_n$  converge en distribución o converge débilmente a  $\hat{X}$  que se escribe  $\hat{X}_n \xrightarrow{d} \hat{X}$

La demostración se encuentra en [?] página 16.

**Proposición B.9.** Sean  $\hat{X}, \hat{X}_1, \hat{X}_2, \dots$  y  $\hat{Y}, \hat{Y}_1, \hat{Y}_2, \dots$  vectores aleatorios sobre  $\mathbb{R}^k$  y  $\mathbb{R}^p$  respectivamente, tales que  $\hat{X}_n \xrightarrow{d} \hat{X}$  y  $\hat{Y}_n \xrightarrow{p} \hat{Y}$  entonces

$$(\hat{X}_n, \hat{Y}_n) \xrightarrow{d} (\hat{X}, \hat{Y})$$

Donde  $\hat{Y}$  es un vector constante

*Demostración.* Sea  $H : \mathbb{R}^{k+p} \rightarrow \mathbb{R}$  una función uniformemente continua y acotada, entonces

$$| \mathbb{E}[H(\hat{X}_n, \hat{Y}_n)] - \mathbb{E}[H(\hat{X}, \hat{Y})] | \leq \mathbb{E}[| H(\hat{X}_n, \hat{Y}_n) - H(\hat{X}_n, \hat{Y}) |] + | \mathbb{E}[H(\hat{X}_n, \hat{Y})] - \mathbb{E}[H(\hat{X}, \hat{Y})] |$$

El primer sumando tiende a cero por que H es acotada y uniformemente continua, mientras que el segundo tiende a cero por que  $\hat{Y}$  es constante y por la Proposición anterior  $\square$

Los siguientes resultados pueden ser consultados en [?]

**Teorema B.1. (Ley fuerte de los grandes números).** Sea  $\xi_1, \xi_2, \dots$  una sucesión de variables aleatorias independientes e idénticamente distribuidas con media  $\mu$ . Entonces

$$\frac{1}{n} \sum_{i=1}^n \xi_i \xrightarrow{c.s.} \mu.$$

**Teorema B.2. (Teorema Central del Límite).** Sea  $\xi_1, \xi_2, \dots$  una sucesión de variables aleatorias independientes e idénticamente distribuidas con media  $\mu$  y varianza finita  $\sigma$ . Entonces

$$\sum_{i=1}^n \frac{\xi_i - \mu}{\sqrt{n}\sigma} \xrightarrow{d} N(0, 1).$$

Para la demostración del siguiente teorema ver [?].

**Teorema B.3. (Teorema de Curtiss).** Sea  $\xi, \xi_1, \xi_2, \dots$  una sucesión de vectores aleatorios con funciones generadoras de momentos  $M, M_1, M_2, \dots$  finitas en algún abierto  $D$  que contenga al cero.

Si  $M_n(t) \xrightarrow{n \rightarrow \infty} M(t)$  para  $t \in D$  se tiene que  $\xi_n \xrightarrow{n \rightarrow \infty} \xi$ .

# Apéndice C

## Códigos en R

En el código C.1 se realizan trayectorias que reciben como parámetros:

1.  $N$ ; tamaño de la población
2.  $t$ ; tiempo de observación / tiempo de horizonte
3.  $\beta$ ; tasa de infección
4.  $\gamma$ ; tasa de recuperación
5.  $S_0, I_0, R_0$ ; condiciones iniciales.

### Código C.1.

```
1      sir_gillespie <- function(S0,I0,R0,beta,gamma,t){
2          S=S0
3          I=I0
4          R=R0
5          N=S+I+R
6          t_ev=0
7          times <- c(0)
8          St <- c(S)
9          It <- c(I)
10         proba <- numeric(2)
11
12         while(t_ev < t && I > 0){
13
14             lambda= ((beta*S*I)/N) + (gamma*I)
15             t_ev <- t_ev + rexp(1,rate=lambda)
16             times <- append(times,t_ev)
17             pi = c((beta*S*I/N)/lambda , (gamma*I)/lambda)
18             eventos <- sample(c(1,-1),1, prob = pi)
19
20             if(eventos==1){
21                 I <- I + eventos
22                 S <- S - 1
```

```
23         R <- R
24     } else {
25         I <- I + eventos
26         S <- S
27         R <- R + 1
28     }
29     It <- append(It,I)
30     St <- append(St,S)
31 }
32 return( data.frame(It=It, St=St, times=times))
33 }
```

# Apéndice D

## Códigos en Julia

El código [D.1](#) crea un método llamado *trayectoria* que recibe como parámetros:

1. *matrixtrans*; es la matriz de transición de la cadena de Markov a simular, es de tamaño  $N \times N$
2. *distinicial*; es el vector de transición inicial de la cadena de Markov, es de tamaño  $N$
3. *tiempo*; es el tiempo hasta el cual se quiere simular la cadena de Markov
4. *E*; es el espacio de estados de la cadena de Markov, es de tamaño  $N$ .

El método nos proporciona una trayectoria de la cadena de Markov, es decir un vector y nos muestra la gráfica de la trayectoria simulada.

### Código D.1.

```
1  Pkg.add("QuantEcon")
2  Pkg.add("StatsBase")
3  Pkg.add("Distributions")
4  Pkg.add("PyPlot")
5  using PyPlot
6  using StatsBase
7  using Distributions
8  using QuantEcon
9
10
11 function trayectoria(matrixtrans, distinicial, tiempo, E)
12     dim=length(E)
13     posicion=collect(1:1:dim)
14     estado=0
15     simulaciones=zeros(tiempo+1)
16     estado=usample(posicion, distinicial, 1)[1]
17     simulaciones[1]=E[estado]
18
19     k=2
20     sim=zeros(dim)
21     while k<=tiempo+1
22         h=1
```

```

23 while h<=dim
24   sim[h]=matrixtrans[estado,h]
25   h=h+1
26 end
27 estado=wsample(posicion,sim,1)[1]
28 simulaciones[k]=E[estado]
29 k=k+1
30 end
31 x=collect(0:1:tiempo)
32
33
34 return hcat(x,simulaciones)
35 end
36
37 matrixtrans=[.2 .1 .5 .2;.5 .2 .2 .1;.2 .5 .1 .2;.3 .5 .1 .1]
38 distinicial=[.2,.4,.2,.2]
39 E=[1,2,3,4]
40 tiempo=30
41 trayecto=trayectoria(matrixtrans,distinicial,tiempo,E)
42 plot(trayecto[1:end,1],trayecto[1:end,2],".")
43 axis([0,tiempo,E[1]-1,E[end]+1])
44 title("Trayectoria de una Cadena de Markov")

```

El código [D.2](#) crea un método llamado *inferencia* que calcula los estimadores máximo verosímiles de las probabilidades de transición teniendo como información una trayectoria observada de una cadena de Markov con espacio de estados  $E$ , recibe como parámetros:

1. *trayecto*; es una trayectoria de la cadena de Markov, es un vector
2.  $E$ ; es el espacio de estados en el que vive la cadena de Markov, es de tamaño  $N$ .

El método nos proporciona como salida una matriz de tamaño  $N \times N$ .

### Código D.2.

```

1  Pkg.add("QuantEcon")
2  Pkg.add("StatsBase")
3  Pkg.add("Distributions")
4  Pkg.add("PyPlot")
5  using PyPlot
6  using StatsBase
7  using Distributions
8  using QuantEcon
9
10 function inferencia(trayecto,E)
11   N=length(E)
12   P=zeros(N,N)
13   n=zeros(N)
14   t=length(trayecto)
15   k=1
16   l=2
17   u=1
18   while(trayecto[1]!=E[k])
19     k=k+1

```

```

20  end
21  while (l<=t)
22  u=1
23  while (trayecto[l]!=E[u])
24  u=u+1
25  end
26  P[k,u]=1+P[k,u]
27  n[k]=1+n[k]
28  k=u
29  l=l+1
30  end
31  i=1
32  j=1
33  while (i<=N)
34  j=1
35  while (j<=N)
36  P[i,j]=P[i,j]/n[i]
37  j=j+1
38  end
39  i=i+1
40  end
41  return P
42
43  end
44  estimador=inferencia(trayecto[1:end,2],E)

```

El código [D.3](#) crea un método para analizar la convergencia de la distancia entre la matriz de transición real y la estimada conforme el horizonte de tiempo converge a infinito, recibe los siguientes parámetros:

1. *matrixtrans*; es la matriz de transición de la cadena de Markov a simular, es de tamaño  $N \times N$
2. *distinicial*; es el vector de distribución inicial de la cadena de Markov, es de tamaño  $N$
3. *tiempo*; es el tiempo hasta el cual se quiere analizar la distancia
4. *E*; es el espacio de estados de la cadena de Markov, es de tamaño  $N$ .

El método nos proporciona la gráfica de la convergencia de la distancia.

### Código D.3.

```

1  function convergencia(matrixtrans, distinicial, tiempo, E)
2  distancia=zeros(1)
3  vectortiempo=zeros(1)
4  t=100
5  simulaciones=trayectoria(matrixtrans, distinicial, max(tiempo, 100), E)
6  k=1
7  i=1
8  while (t<=max(tiempo, 100))
9  i=1

```

```

10 while (t>simulaciones[i,1])
11     i=i+1
12 end
13 estimador=inferencia(simulaciones[1:i,2],E)
14 distancia[k]=norm(estimador-matrixtrans,2)
15 vectortiempo[k]=t
16 distancia=vcat(distancia,0)
17 vectortiempo=vcat(vectortiempo,0)
18 k=k+1
19 t=t+100
20 end
21
22 plot(vectortiempo[1:end-1],distancia[1:end-1])
23 hlines(xmin=0,y=0,xmax=vectortiempo[end-1])
24 axis([0,vectortiempo[end-1],-.01,maximum(distancia[1:end-1])])
25 xlabel("tiempo")
26 ylabel("norma")
27 title("Convergencia del estimador ")
28 return hcat(vectortiempo[1:end-1],distancia[1:end-1])
29 end

```

El código D.4 crea un método llamado *PSM* que simula una cadena de Markov en tiempo continuo (PSM), recibe como parámetros :

1. *generador*; es la matriz de intensidades (generador infinitesimal), es de tamaño  $N \times N$
2. *distinicial*; es el vector de distribución inicial del proceso, es de tamaño  $N$
3. *tiempo*; es el tiempo hasta el cual se quiere simular el proceso
4. *E*; es el espacio de estados del proceso, es de tamaño  $N$ .

El método nos proporciona la trayectoria simulada.

#### Código D.4.

```

1
2
3 function PSM(generator, distinicial, tiempo, E)
4     dim=length(E)
5     posicion=collect(1:1:dim)
6     estado=0
7     saltos=zeros(1)
8     tiempos=zeros(1)
9     estado=usample(posicion, distinicial, 1)[1]
10    saltos[1]=E[estado]
11    tiempos[1]=log(rand())/generator[estado, estado]
12    k=tiempos[1]
13    sim=zeros(dim)
14    cont=2
15    while k<=tiempo
16        h=1

```

```

17 while h<=dim
18   sim[h]=-generador[estado,h]/generador[estado,estado]
19   h=h+1
20   sim[estado]=0
21 end
22 estado=wsample(posicion,sim,1)[1]
23 saltos=vcatsaltos,0)
24 saltos[cont]=E[estado]
25 tiempos=vcattiempos,0)
26 tiempos[cont]=(log(rand())/generador[estado,estado])+k
27 k=tiempos[cont]
28 cont=cont+1
29 end
30 ##Aquí definimos 2 nuevos arreglos llamados tiempoinicial y tiempofinal,
31 ##estos representan en su i-esima entrada el tiempo inicial y tiempo final
32 ##en el que el proceso se mantuvo constante en el (i-1)-esimo salto.
33 tiempoinicial=zeros(length(saltos))
34 tiempofinal=zeros(length(saltos))
35 tiempofinal=tiempos[1:(length(saltos))]
36 tiempoinicial[2:(length(saltos))]=tiempos[1:(length(saltos)-1)]
37 tiempofinal[length(saltos)]=tiempos
38 tiempoinicial[1]=0
39 return hcat(saltos,tiempoinicial,tiempofinal)
40 end
41 generador=[-.13 .04 .01 .08;.02 -.10 .04 .04;.2 .1 -.6 .3;.05 .15 .1 -.3]
42 distinicial=[.2,.2,.1,.5]
43 E=[1 ,2,3,4]
44 tiempo=50
45 simulaciones=PSM(generador,distinicial,tiempo,E)
46 n=div(length(simulaciones),3)
47
48 hlines(y=simulaciones[1:n,1],xmin=simulaciones[1:n,2],xmax=simulaciones[1:n,3],
         color="blue",linewidth=3)
49 plot(simulaciones[1:n,2],simulaciones[1:n,1],"bo",color="blue")
50 plot(simulaciones[1:(n-1),3],simulaciones[1:(n-1),1],"o",color="white")
51 axis([0,tiempo,E[1]-1,E[end]+1])
52 title("Trayectoria del PSM")

```

El código D.5 crea un método llamado *inferenciaPSM*, calcula el estimador máximo verosímil del generador infinitesimal de una trayectoria dada de un PSM, los parámetros del método son:

1. *tiempos*; es un vector que almacena los tiempos entre saltos, es decir los tiempos de inter arribo del proceso.
2. *estados*; es un vector que almacena los estados visitados del proceso en cada transición
3. *E*; es el espacio de estados del proceso, es de tamaño N.

El método nos proporciona una matriz de  $N \times N$ , la cual representa el estimador máximo verosímil del generador infinitesimal de la trayectoria dada.

#### Código D.5.

```

1
2

```

```

3  function inferenciaPSM(tiempos, estados, E)
4  N=length(E)
5  Q=zeros(N,N)
6  R=zeros(N)
7  t=length(estados)
8  k=1
9  l=2
10 u=1
11 while(estados[1]!=E[k])
12 k=k+1
13 end
14
15 while(l<=t)
16 u=1
17 while(estados[l]!=E[u])
18 u=u+1
19 end
20 Q[k,u]=1+Q[k,u]
21 Q[k,k]=1+Q[k,k]
22 R[k]=R[k]+tiempos[l-1]
23 k=u
24 l=l+1
25 end
26 R[u]=R[u]+tiempos[t]
27 i=1
28 while(i<=N)
29 j=1
30 while(j<=N)
31 Q[i,j]=Q[i,j]/R[i]
32
33 j=j+1
34 end
35 Q[i,i]=-Q[i,i]
36 i=i+1
37 end
38 return Q
39
40 end
41
42 generadorestimado=inferenciaPSM(simulaciones[1:end,3]-simulaciones[1:end,2],
    simulaciones[1:end,1],E)

```

El código D.6 crea un método llamado *convergencia* que sirve para analizar la convergencia de la distancia entre el generador infinitesimal real y el estimado conforme el horizonte de tiempo tiende a infinito, recibe los siguientes parámetros:

1. *generador*; es la matriz de intensidades (generador infinitesimal) del proceso, es de tamaño  $N \times N$
2. *distinicial*; es el vector de distribución inicial del proceso, es de tamaño  $N$
3. *tiempo*; es el tiempo hasta el cual se quiere analizar la convergencia, al menos debe ser 100

4.  $E$ ; es el espacio de estados del proceso, es de tamaño  $N$ .

El método nos proporciona la gráfica de la convergencia de la distancia.

#### Código D.6.

```

1  function convergencia(generator, distinicial, tiempo, E)
2  distancia=zeros(1)
3  vectortiempo=zeros(1)
4  t=100
5  d=0
6  simulaciones=PSM(generator, distinicial, max(tiempo, 100), E)
7  k=1
8  i=1
9  while(t<=max(tiempo, 100))
10 i=1
11 while(t>simulaciones[i, 3])
12 i=i+1
13 end
14 d=simulaciones[i, 3]-t
15 simulaciones[i, 3]=t
16 distancia[k]=norm(inferenciaPSM(simulaciones[1:i, 3]-simulaciones[1:i, 2],
17     simulaciones[1:i, 1], E)-generator, 2)
18 vectortiempo[k]=t
19 simulaciones[i, 3]=t+d
20 distancia=vcat(distancia, 0)
21 vectortiempo=vcat(vectortiempo, 0)
22 k=k+1
23 t=t+100
24 end
25
26 plot(vectortiempo[1:end-1], distancia[1:end-1])
27 hlines(xmin=0, y=0, xmax=vectortiempo[end-1])
28 axis([0, vectortiempo[end-1], -.01, maximum(distancia[1:end-1])])
29 xlabel("tiempo")
30 ylabel("norma")
31 title("Convergencia del estimador")
32 return hcat(vectortiempo[1:end-1], distancia[1:end-1])
33 end

```

El código D.7 crea un método llamado *clasescomunicacion* recibe como parámetros:

1. *matrixtrans*; es la matriz de transición de una cadena de Markov
2.  $E$ ; es el espacio de estados de la cadena de Markov.

Como resultado nos proporciona todas las clases de comunicación de la cadena de Markov

#### Código D.7.

```

1  function clasescomunicacion(matrixtrans, E)
2  mc=MarkovChain(matrixtrans)
3  clases=communication_classes(mc)

```

```

4  tam1=length(communication_classes(mc))
5  for i in 1:tam1
6  tam2=length(communication_classes(mc)[i])
7  for j in 1:tam2
8  clases[i][j]=E[clases[i][j]]
9  end
10 end
11
12 return clases
13 end
14
15 matrixtrans =[0 1 0 0;1/3 0 2/3 0;0 2/3 0 1/3;0 0 1 0]
16 E=[0,1,2,3]
17 clasescomunicacion(matrixtrans,E)

```

El código D.8 crea un método llamado *periodo* recibe un sólo parámetro:

1. *matrixtrans*; es la matriz de transición de una cadena de Markov.

Como resultado nos muestra el periodo asociado a la cadena de Markov, suponiendo que dicha cadena es irreducible.

#### Código D.8.

```

1  function periodo(matrixtrans)
2  mc=MarkovChain(matrixtrans)
3  return period(mc)
4  end
5
6  matrixtrans =[0 1 0 0;1/3 0 2/3 0;0 2/3 0 1/3;0 0 1 0]
7  periodo(matrixtrans)

```

El código D.9 crea un método llamado *estadosrecurrentes* recibe como parámetros:

1. *matrixtrans*; es la matriz de transición de una cadena de Markov
2. *E*; es el espacio de estados de la cadena de Markov.

Como resultado nos muestra el conjunto de estados de la cadena que son recurrentes.

#### Código D.9.

```

1  function estadosrecurrentes(matrixtrans,E)
2  mc=MarkovChain(matrixtrans)
3  recurrentes=recurrent_classes(mc)[1][1:end]
4  for i in 1:length(recurrentes)
5  recurrentes[i]=E[recurrentes[i]]
6  end
7
8  return recurrentes
9  end
10
11 matrixtrans =[0 1 0 0;1/3 0 2/3 0;0 2/3 0 1/3;0 0 1 0]
12 E=[0,1,2,3]
13 estadosrecurrentes(matrixtrans,E)

```

El código D.10 crea un método llamado *estacionaria* que sirve para dar una aproximación de la distribución estacionaria, recibe como parámetros:

1. *matrixtrans*; es la matriz de transición de una cadena de Markov
2. *E*; es el espacio de estados de la cadena de Markov
3. *tiempo*; es el tiempo hasta el cual se quiere analizar la convergencia a la distribución estacionaria.

Como resultado nos muestra la aproximación del vector estacionario.

#### Código D.10.

```

1  function estacionaria(matrixtrans,E,tiempo)
2  dim=length(E)
3  nu=zeros(dim)
4  distinicial=zeros(dim)
5  distinicial[1]=1
6  trayecto=trayectoria(matrixtrans,distinicial,tiempo,E)[1:end,2]
7  for i in 2:tiempo+1
8  for j in 1:dim
9  if trayecto[i]==E[j]
10 nu[j]=nu[j]+1
11 end
12
13 end
14
15 end
16 return nu/tiempo
17 end
18
19 tiempo=10000
20 matrixtrans =[0 1 0 0;1/3 0 2/3 0;0 2/3 0 1/3;0 0 1 0]
21 E=[0,1,2,3]
22 estacionaria(matrixtrans,E,tiempo)

```

El código D.11 crea un método llamado *convergenciaestacionaria* recibe como parámetros:

1. *matrixtrans*; es la matriz de transición de una cadena de Markov
2. *E*; es el espacio de estados de la cadena de Markov
3. tiempo; es el tiempo hasta el cual se quiere analizar la convergencia de la distribución estacionaria
4. vectorinvariante; es la distribución estacionaria de la cadena de Markov.

Como resultado nos muestra la gráfica de la convergencia a la distribución estacionaria.

**Código D.11.**

```

1  function convergenciaestacionaria(matrixtrans,E,tiempo,vectorinvariante)
2  dim=length(E)
3  nu=zeros(dim)
4  d=div(tiempo,1000)
5  convergencia=zeros(d)
6
7
8  distinicial=zeros(dim)
9  distinicial[1]=1
10
11 trayecto=trayectoria(matrixtrans,distinicial,tiempo,E)[1:end,2]
12 for k in 1:d
13   for i in 1000*(k-1)+2:1000*k+1
14     for j in 1:dim
15       if trayecto[i]==E[j]
16         nu[j]=nu[j]+1
17       end
18     end
19
20   end
21   convergencia[k]=norm((nu/(1000*k))-vectorinvariante,2)
22 end
23
24 plot(collect(1000:1000:tiempo),convergencia[1:end])
25 hlines(xmin=0,y=0,xmax=d*1000)
26 axis([0,d*1000,-.01,maximum(convergencia[1:end])])
27 xlabel("tiempo")
28 ylabel("norma")
29 title("Convergencia al vector estacionario")
30 return convergencia
31 end
32
33 tiempo=100000
34 matrixtrans =[0 1 0 0;1/3 0 2/3 0;0 2/3 0 1/3;0 0 1 0]
35 E=[0,1,2,3]
36 vectorinvariante=[1/8,3/8,3/8,1/8]
37 convergenciaestacionaria(matrixtrans,E,tiempo,vectorinvariante)

```

El código [D.12](#) llamado *probatransicion* sirve para encontrar la matriz de probabilidades de transición al tiempo t del PSM, recibe un sólo parámetro:

1. *generador*; es la matriz de intensidades (generador infinitesimal) del proceso, es de tamaño  $N \times N$ .

La función muestra la diagonalización del generador infinitesimal, para después expresar a la matriz de probabilidades de transición al tiempo t del PSM.

**Código D.12.**

```

1  function probatransicion(generador)
2  dim=length(generador[1:end,1])
3  I=zeros(dim,dim)
4  for i in 1:dim
5    I[i,i]=eig(generador)[1][i]
6  end

```

```

7  return Array[(eig(generator)[2]),I,inv(eig(generator)[2])]
8  end
9  generador=[-1 1 0 0;2 -4 2 0;0 1 -2 1;0 0 2 -2]
10 probatransicion(generador)

```

El código D.13 crea un método llamado *clasescomunicacionPSM* recibe como parámetros:

1. *generador*; es la matriz de intensidades (generador infinitesimal) del proceso
2. *E*; es el espacio de estados del PSM.

Como resultado nos proporciona todas las clases de comunicación de la cadena de saltos asociada al PSM.

### Código D.13.

```

1  function clasescomunicacionPSM(generador,E)
2
3  d=length(E)
4  matrixtrans=zeros(d,d)
5  k=0
6  for i in 1:d
7  k=generador[i,i]
8  for j in 1:d
9  matrixtrans[i,j]=-generador[i,j]/k
10 end
11 matrixtrans[i,i]=0
12 end
13
14 mc=MarkovChain(matrixtrans)
15 clases=communication_classes(mc)
16 tam1=length(communication_classes(mc))
17 for i in 1:tam1
18 tam2=length(communication_classes(mc)[i])
19 for j in 1:tam2
20 clases[i][j]=E[clases[i][j]]
21 end
22 end
23
24 return clases
25 end
26
27 generador=[-1 1 0 0;2 -4 2 0;0 1 -2 1;0 0 2 -2]
28 E=[0,1,2,3]
29 clasescomunicacionPSM(generador,E)

```

El código D.14 crea un método llamado *estadosrecurrentesPSM* recibe como parámetros:

1. *generador*; es la matriz de intensidades (generador infinitesimal) del PSM
2. *E*; es el espacio de estados del PSM.

Como resultado nos muestra el conjunto de estados del PSM que son recurrentes.

**Código D.14.**

```

1  function estadosrecurrentesPSM(generator,E)
2  d=length(E)
3  matrixtrans=zeros(d,d)
4  k=0
5  for i in 1:d
6  k=generator[i,i]
7  for j in 1:d
8  matrixtrans[i,j]=-generator[i,j]/k
9  end
10 matrixtrans[i,i]=0
11 end
12 mc=MarkovChain(matrixtrans)
13 recurrentes=recurrent_classes(mc)[1][1:end]
14 for i in 1:length(recurrentes)
15 recurrentes[i]=E[recurrentes[i]]
16 end
17
18 return recurrentes
19 end
20
21 generator=[-1 1 0 0;2 -4 2 0;0 1 -2 1;0 0 2 -2]
22 E=[0,1,2,3]
23 estadosrecurrentesPSM(generator,E)

```

El código [D.15](#) crea un método llamado *estacionariaPSM* recibe como parámetros:

1. *generator*; es la matriz de intensidades (generador infinitesimal) del proceso, es de tamaño  $N \times N$
2. *tiempo*; es el tiempo hasta el cual se desea calcular la aproximación de la distribución estacionaria
3. *E*; es el espacio de estados del PSM, es un vector de tamaño  $N$ .

Como resultado nos muestra la aproximación de la distribución estacionaria que es un vector de la misma longitud que *E*.

**Código D.15.**

```

1  function estacionariaPSM(generator,tiempo,E)
2  dim=length(E)
3  distinicial=zeros(dim)
4  distinicial[1]=1
5  posicion=collect(1:1:dim)
6  tiempoestancia=zeros(dim)
7  estado=0
8  estado=usample(posicion,distinicial,1)[1]
9  tiempoestancia[estado]=log(rand())/generator[estado,estado]
10 k=tiempoestancia[estado]
11 s=k
12 sim=zeros(dim)
13
14
15 while k<=tiempo
16 h=1
17 while h<=dim

```

```

18  sim[h]=-generador[estado,h]/generador[estado,estado]
19  h=h+1
20  sim[estado]=0
21  end
22  estado=wsample(posicion,sim,1)[1]
23  s=(log(rand())/generador[estado,estado])
24  k=s+k
25  tiempoestancia[estado]=s+tiempoestancia[estado]
26  end
27  tiempoestancia[estado]=tiempoestancia[estado]-s+k-tiempo
28  return tiempoestancia/tiempo
29  end
30  generador=[-1 1 0 0;2 -4 2 0;0 1 -2 1;0 0 2 -2]
31  E=[0,1,2,3]
32  tiempo=10000
33  estacionariaPSM(generador,tiempo,E)

```

El código [D.16](#) llamado *convergenciaestacionariaPSM* recibe como parámetros:

1. *generador*; es la matriz de intensidades (generador infinitesimal) del proceso, es de tamaño  $N \times N$
2. *tiempo*; es el tiempo hasta el cual se desea analizar la convergencia a la distribución estacionaria
3. *E*; es el espacio de estados del PSM, es un vector de tamaño  $N$
4. *vectorinvariante*; es la distribución estacionaria del PSM, es un vector de tamaño  $N$ .

Como resultado nos muestra la gráfica de la convergencia hacia la distribución estacionaria.

#### Código D.16.

```

1  function convergenciaestacionariaPSM(generador,tiempo,E,vectorinvariante)
2  dim=length(E)
3  d=div(tiempo,1000)
4  convergencia=zeros(d)
5  distinicial=zeros(dim)
6  distinicial[1]=1
7  posicion=collect(1:1:dim)
8  tiempoestancia=zeros(dim)
9  estado=0
10 estado=wsample(posicion,distinicial,1)[1]
11 tiempoestancia[estado]=log(rand())/generador[estado,estado]
12 k=tiempoestancia[estado]
13 s=k
14 sim=zeros(dim)
15 for l in 1:d
16 while k<=l*1000
17 h=1
18 while h<=dim
19 sim[h]=-generador[estado,h]/generador[estado,estado]
20 h=h+1
21 sim[estado]=0

```

```

22  end
23  estado=wsample(posicion,sim,1)[1]
24  s=(log(rand())/generador[estado,estado])
25  k=s+k
26  tiempoestancia[estado]=s+tiempoestancia[estado]
27  end
28  tiempoestancia[estado]=tiempoestancia[estado]-s+k-l*1000
29  convergencia[l]=norm(vectorinvariante-(tiempoestancia/(l*1000)),2)
30  tiempoestancia[estado]=tiempoestancia[estado]+s-k+l*1000
31  end
32
33  plot(collect(1000:1000:tiempo),convergencia[1:end])
34  hlines(xmin=0,y=0,xmax=d*1000)
35  axis([0,d*1000,-.01,maximum(convergencia[1:end])])
36  xlabel("tiempo")
37  ylabel("norma")
38  title("Convergencia al vector estacionario")
39  return convergencia
40
41  return convergencia
42  end
43
44
45  generador=[-1 1 0 0;2 -4 2 0;0 1 -2 1;0 0 2 -2]
46  E=[0,1,2,3]
47  tiempo=100000
48  vectorinvariante=[1/3,1/6,1/3,1/6]
49  convergenciaestacionariaPSM(generador,tiempo,E,vectorinvariante)

```

El código [D.17](#) crea un método llamado *asintotica* que recibe los siguientes parámetros:

1. *matrixtrans*; es la matriz de transición de la Cadena de Markov, es de tamaño  $N \times N$
2. *distinicial*; es el vector de la distribución inicial de la cadena de Markov, es de tamaño  $N$
3. *tiempo*; es el tiempo hasta el cual se quiere aproximar a la matriz  $\tilde{\Sigma}$
4. *E*; es el espacio de estados de la cadena de Markov, es de tamaño  $N$ .

El método nos proporciona la estimación de  $\tilde{\Sigma}$  (ver Teorema [4.3](#)).

#### Código D.17.

```

1  function asymptotica(matrixtrans, distinicial, tiempo, E)
2  trayecto=trayectoria(matrixtrans, distinicial, tiempo, E)[1:end,2]
3  N=length(E)
4  matrixFisher=zeros(N^2-N,N^2-N)
5  P=zeros(N,N)
6  n=zeros(N)
7  t=length(trayecto)
8  k=1
9  l=2

```

```

10  u=1
11  while(trayecto[1]!=E[k])
12  k=k+1
13  end
14  while(l<=t)
15  u=1
16  while(trayecto[l]!=E[u])
17  u=u+1
18  end
19  P[k,u]=1+P[k,u]
20  n[k]=1+n[k]
21  k=u
22  l=l+1
23  end
24
25  for i in 1:N
26  for j1 in 1:N-1
27  for j2 in 1:N-1
28  if j1==j2
29  matrixFisher[((i-1)*(N-1))+j1,((i-1)*(N-1))+j2]=((P[i,j1]/n[i])-(P[i,j1]/n[i])
      ^2)/(n[i]/tiempo)
30  else
31  matrixFisher[((i-1)*(N-1))+j1,((i-1)*(N-1))+j2]=(-(P[i,j1]/n[i])^2)/(n[i]/
      tiempo)
32  end
33  end
34  end
35  end
36  return matrixFisher
37
38  end
39  E=[1,2]
40  matrixtrans=[.2 .8 ; .8 .2]
41  distinicial=[.5,.5]
42  tiempo=100000
43  asintotica(matrixtrans,distinicial,tiempo,E)

```

El código [D.18](#) crea un método llamado *asintoticaPSM* recibe como parámetros:

1. *generador*; es la matriz de intensidades (generador infinitesimal) del proceso, es de tamaño  $N \times N$
2. *distinicial*; es la distribución inicial del PSM, es de tamaño  $N$
3. *E*; es el espacio de estados del PSM, es un vector de tamaño  $N$
4. *tiempo*; es el tiempo hasta el cual se quiere aproximar a  $\Lambda$ .

Como resultado nos muestra la estimación de  $\Lambda$  (ver Teorema [4.5](#)).

**Código D.18.**

```

2
3  function asintoticaPSM(generator, distinicial, tiempo, E)
4  simulaciones=PSM(generator, distinicial, tiempo, E)
5  tiempos=simulaciones[1:end,3]-simulaciones[1:end,2]
6  estados=simulaciones[1:end,1]
7  N=length(E)
8  matrixFisher=zeros(N^2-N, N^2-N)
9  Q=zeros(N, N)
10 R=zeros(N)
11 t=length(estados)
12 k=1
13 l=2
14 u=1
15 while(estados[1]!=E[k])
16 k=k+1
17 end
18
19 while(l<=t)
20 u=1
21 while(estados[l]!=E[u])
22 u=u+1
23 end
24 Q[k, u]=1+Q[k, u]
25 Q[k, k]=1+Q[k, k]
26 R[k]=R[k]+tiempos[l-1]
27 k=u
28 l=l+1
29 end
30 R[u]=R[u]+tiempos[t]
31
32
33 for i in 1:N
34
35 j1=1
36 l=0
37 while j1<=N
38 if i==j1
39 l=1
40 j1=j1+1
41 else
42 matrixFisher[((i-1)*(N-1))+j1-l, ((i-1)*(N-1))+j1-l]=(Q[i, j1]/R[i])^2/(Q[i, j1]/
    tiempo)
43 j1=j1+1
44 end
45
46 end
47 end
48 return matrixFisher
49 end
50
51
52 E=[1, 2, 3]
53 distinicial=[1/3, 1/3, 1/3]
54 generator=[-2 1 1 ; 1 -2 1; 1 1 -2 ]
55 tiempo=10000
56 asintoticaPSM(generator, distinicial, tiempo, E)

```

Listing D.1: Puente Markoviano

```
1  puente_rej <- function(t,a,b,Q){
2      n <- nrow(Q)
3      mu <- rep(0,n)
4      mu[a] <- 1
5      k <- 0
6      while(T){
7          k <- k+1
8          trayectoria <- ctmc(Q,mu,t)
9          if(tail(trayectoria$Xn,1) == b){
10             break
11         }
12     }
13     print(k)
14     return(trayectoria)
15 }
```



# Bibliografía

- [1] S. Asmussen and P. W. Glynn. *Stochastic Simulation: Algorithms and Analysis*. Springer, Stochastic Modelling and Applied Probability, 2007.
- [2] N. Bailey. *The Mathematical Theory of Epidemics*. Grin, London, 1957.
- [3] M. Bladt and M. Sørensen. Simple simulation of diffusion bridges with application to likelihood inference for diffusions. *Bernoulli*, 2(20):645–675, 2014.
- [4] G. U. Bravo. *Procesos Estocásticos*.
- [5] R. Brown. A brief account of microscopical observations made in the months of june, july, and august, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. *Philosophical Magazine N. S.*, (4):161–173, 1828.
- [6] I. J. R. S. Cox, J.C. A theory of the term structure of interest rates,. *Econometrica*, (53):385–408, 1985.
- [7] D. Dacunha-Castelle and D. Florens-Zmirou. Estimation of the coefficients of a diffusion from discrete observations. *Stochastics*,, 19:263–284, 1986.
- [8] A. Einstein. *Investigations on the theory of the Brownian movement*. Dover, 1956.
- [9] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume 1. Wiley Series in Probability and Mathematical Statistics, 3 edition, 196.
- [10] S. Iacus. *Simulation and Inference for Stochastic Differential Equations with R Examples*. Springer Series in Statistics, 2008.
- [11] Y. Kutoyants. *Statistical Inference for Ergodic Diffusion Processes*. Springer-Verlag, London, 2004.
- [12] P. É. Lapayre, B. and R. Sentis. *Introduction to Monte-Carlo Methods for Transportation and Diffusion Equations*. Oxford University Press, New York, 2003.
- [13] R. Merton. Theory of rational option pricing. *Bell J. Econ. Manage. Sci.*, 4(1):141–183, 1973.
- [14] N. Newton. Variance reduction for simulated diffusions. *SIAMJ. Appl. Math*, 54(6):1780–1805, 1994.
- [15] B. Øksendal. *Stochastic Differential Equations*. 1998.
- [16] N. Yoshida. Estimation for diffusion processes from discrete observation. *J. Multivar. Anal.*, 41(2):220–242, 1992.