



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

---

---

FACULTAD DE ESTUDIOS SUPERIORES ACATLÁN

CADENAS DE MARKOV Y SUS APLICACIONES

T E S I N A

QUE PARA OBTENER EL TÍTULO DE:

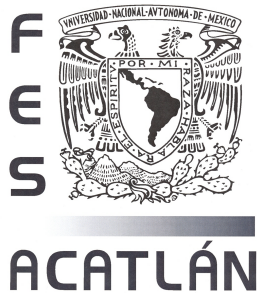
**ACTUARIO**

P R E S E N T A:

EDWIN JOSÉ LEÓN SÁNCHEZ

DIRECTOR DE TESINA:

M. en C. MIGUEL ÁNGEL SÁNCHEZ BARQUÍN



Naucalpan, Méx., marzo 2022



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



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

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

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

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

---

# ÍNDICE GENERAL

<b>Introducción</b>	<b>5</b>
<b>1. Teoría de la Probabilidad</b>	<b>8</b>
1.1. Conceptos básicos . . . . .	8
1.1.1. Eventos . . . . .	8
1.1.2. Variables Aleatorias . . . . .	9
1.1.3. Medidas de Probabilidad . . . . .	9
1.2. Probabilidad Condicional e Independencia . . . . .	12
1.2.1. Independencia de Eventos y de Variables Aleatorias . . . . .	12
1.2.2. Reglas de Bayes . . . . .	14
1.2.3. Propiedad de Markov . . . . .	16
1.3. Esperanza . . . . .	18
1.3.1. Función de Distribución . . . . .	18
1.3.2. Esperanza, media y varianza . . . . .	19
1.3.3. Variables Aleatorias famosas . . . . .	24
1.4. Vectores Aleatorios . . . . .	28
1.4.1. Vectores Aleatorios continuos . . . . .	28
1.4.2. Vectores Aleatorios Discretos . . . . .	30
1.5. Transformaciones de Distribuciones . . . . .	31
1.5.1. Funciones Generadoras . . . . .	31
1.5.2. Funciones Características . . . . .	34
1.6. Transformaciones de Vectores Aleatorios . . . . .	36
1.6.1. Cambios de variable . . . . .	36
1.7. Estadísticos de orden . . . . .	38
1.8. Esperanza Condicional de Variables Discretas . . . . .	39
1.8.1. Definición y Propiedades Básicas . . . . .	39
1.9. Ley Fuerte de los Grandes Números . . . . .	42
1.9.1. Lema de Borel-Cantelli . . . . .	42
1.9.2. Convergencia Casi Segura . . . . .	42
1.9.3. Desigualdad de Markov . . . . .	44
1.9.4. Prueba de la LFGN de Kolmogorov . . . . .	46

<b>2. Modelos de Markov a Tiempo Discreto</b>	<b>50</b>
2.1. La Matriz de Transición . . . . .	50
2.1.1. Propiedad de Markov . . . . .	50
2.1.2. Gráfica de Transición . . . . .	52
2.1.3. Distribución de una CMH . . . . .	53
2.2. Recurrencias de Markov . . . . .	54
2.2.1. Representación Canónica . . . . .	54
2.2.2. Ejemplos Famosos . . . . .	55
2.3. Análisis del primer paso . . . . .	60
2.3.1. Probabilidad de absorción . . . . .	60
2.3.2. Tiempo medio de absorción . . . . .	63
2.4. La Topología de la Matriz de Transición . . . . .	64
2.4.1. Comunicación . . . . .	64
2.4.2. Periodo . . . . .	65
2.5. Estados Estacionarios . . . . .	68
2.5.1. Propiedades Estacionarias . . . . .	68
2.5.2. Ejemplos . . . . .	68
2.6. Tiempo Reversible . . . . .	72
2.6.1. Cadena Reversible . . . . .	72
2.6.2. Reversibilidad del tiempo . . . . .	73
2.7. Regeneración . . . . .	74
2.7.1. Propiedad Fuerte de Markov . . . . .	74
2.7.2. Ciclos Regenerativos . . . . .	77
<b>3. Recurrencia y Ergodicidad</b>	<b>80</b>
3.1. El criterio de la Matriz Potencia . . . . .	80
3.1.1. Estados Recurrentes y Transitorios . . . . .	80
3.1.2. Matriz Potencia . . . . .	82
3.1.3. Estructura de la Matriz de Transición . . . . .	84
3.2. Recurrencia y Medidas Invariantes . . . . .	85
3.3. Recurrencia Positiva . . . . .	88
3.3.1. Criterio de la Distribución Estacionaria . . . . .	88
3.3.2. Ejemplos . . . . .	90
3.4. Promedios Empíricos . . . . .	94
3.4.1. Teorema Ergódico . . . . .	94
3.4.2. Ejemplos . . . . .	97
3.4.3. Teorema Elemental de Renovación . . . . .	101
<b>4. Comportamiento a Largo Plazo</b>	<b>104</b>
4.1. Acoplamiento . . . . .	104
4.1.1. Convergencia en Variación . . . . .	104
4.1.2. El método de Acoplamiento . . . . .	107
4.2. Convergencia al Estado de Equilibrio . . . . .	108
4.2.1. Caso Recurrente Positivo . . . . .	108
4.2.2. Caso Recurrente Nulo . . . . .	110
4.2.3. Tasas de convergencia mediante acoplamiento . . . . .	111

4.3.	Teoría de renovación a tiempo discreto . . . . .	112
4.3.1.	Ecuación de Renovación . . . . .	112
4.3.2.	Teorema de Renovación . . . . .	115
4.3.3.	Sucesiones de Renovación Defectuosas . . . . .	117
4.4.	Procesos Regenerativos . . . . .	120
4.4.1.	Ecuación de Renovación de un Proceso Regenerativo . . . . .	120
4.4.2.	Teorema Regenerativo . . . . .	121
4.5.	Vida antes de la absorción . . . . .	124
4.5.1.	Estancias infinitas . . . . .	124
4.5.2.	Tiempo de absorción . . . . .	128
4.6.	Absorción . . . . .	129
4.6.1.	Matriz Fundamental . . . . .	129
4.6.2.	Matriz de Absorción . . . . .	131
<b>5.</b>	<b>Aplicaciones</b>	<b>133</b>
5.1.	Campos Aleatorios de Markov . . . . .	133
5.1.1.	Vecindades y especificación local . . . . .	133
5.1.2.	Camarillas, Potencial, y Distribuciones de Gibbs . . . . .	136
5.2.	Equivalencia de Gibbs-Markov . . . . .	138
5.2.1.	Del potencial a la especificación local . . . . .	138
5.2.2.	De la especificación local al potencial . . . . .	140
5.3.	Muestreo de Gibbs . . . . .	146
5.3.1.	Simulación de campos aleatorios . . . . .	146
5.3.2.	Tasa de convergencia del muestreo de Gibbs . . . . .	149
5.4.	Simulación Monte Carlo . . . . .	151
5.4.1.	Principio General . . . . .	151
5.4.2.	Método de la inversa y aceptación-rechazo . . . . .	151
5.4.3.	MCMC . . . . .	153
	<b>Conclusiones</b>	<b>156</b>
	<b>Bibliografía</b>	<b>159</b>

---

# INTRODUCCIÓN

La lista de aplicaciones de las cadenas de Markov es virtualmente infinita, y uno tiene derecho a decir que es la clase individual más exitosa de procesos estocásticos, su éxito se debe a la relativa simplicidad de su teoría y al hecho de que los modelos simples de Markov pueden exhibir comportamientos extremadamente variados y complejos. El poder de modelado de las cadenas de Markov bien puede compararse con el de las ecuaciones diferenciales ordinarias.

La teoría de las cadenas de Markov con un espacio de estado contable es una introducción ideal a los procesos estocásticos; no está protegido por un muro de tecnicismos, por lo que se tiene acceso rápido a los principales resultados. De hecho, el equipo matemático necesario para un estudio gratificante del tema de modelos consiste únicamente en la noción de independencia condicional y la fuerte ley de los grandes números.

Otra característica agradable de la teoría de la cadena de Markov es que este tema clásico se puede presentar en términos de los elegantes conceptos de la teoría moderna de los procesos estocásticos, como la reversibilidad, los procesos martingalas y el acoplamiento.

Las cadenas de Markov llevan el nombre del Prof. Andrei A. Markov (1856-1922). Nació el 14 de junio de 1856 en Ryazan, Rusia y murió el 20 de julio de 1922 en San Petesburgo, Rusia. Markov se inscribió a la Universidad de San Petesburgo, donde consiguió el grado de maestro y el grado de doctor. Fue profesor en la misma universidad y miembro de la Academia de Ciencias Rusa. Se retiró en 1905, pero continuó sus enseñanzas en la universidad hasta su muerte. Markov es particularmente recordado por su estudio acerca de las cadenas de Markov. Sus investigaciones sobre cadenas de Markov son importantes en el estudio de los procesos estocásticos con muchas aplicaciones. [26]

Cuando Markov presentó su famoso modelo en 1906, no estaba preocupado por las aplicaciones. Solo quería mostrar que la independencia no es necesaria para la ley de los grandes números (la ley débil de los grandes números en ese momento, ya que Borel demostró la ley fuerte solamente para el experimento de cara y cruz en 1909). Un ejemplo que consideró fue la alternancia de consonantes y vocales en *Eugene Onegin* de Pushkin, que describió como una cadena de dos estados. Casi al mismo

tiempo, y fiel a la tradición francesa de los probabilistas del juego, Poincaré estudió las cadenas de Markov en grupos finitos, con aplicaciones al barajado de cartas. Los físicos austriacos Paul y Tatiana Ehrenfest propusieron en 1907 un modelo de cadena de Markov que ayudó mucho a aclarar la controvertida cuestión de la irreversibilidad termodinámica. Sir Francis Galton, un primo de Darwin, que estaba interesado en la probabilidad de supervivencia de la nobleza inglesa, fue el inventor del proceso de ramificación, otro famoso modelo de Markov con muchas aplicaciones además del original. Planteó el problema en *Educational Times* en 1873, y en el mismo año y en la misma revista, el reverendo Watson propuso el método de solución.

Las fechas mencionadas anteriormente muestran que los modelos de Markov ya existían incluso antes de que Markov comenzara el estudio sistemático de esta clase de sucesiones aleatorias. Sin embargo, el trabajo de Markov desafió a los mejores probabilistas, como Kolmogorov, Doebelin y Fréchet, solo por mencionar a las principales figuras pioneras. El resultado fue una teoría limpia y sólida lista para aplicaciones, y hoy, las cadenas de Markov son omnipresentes en las ciencias aplicadas. Por ejemplo, la biología es un consumidor importante de modelos de Markov, muchos de ellos relacionados con la genética y la teoría de poblaciones. En las ciencias sociales, la movilidad social se puede describir en términos de Markov. La psicología cuantitativa utiliza modelos de aprendizaje de Markov. La física es uno de los principales patrocinadores de la teoría de cadenas de Markov, y los modelos de Markov (por ejemplo, el modelo de difusión de Ehrenfest, el modelo de recocido y el modelo de transición de fase de Ising-Peierls) han sido muy útiles para comprender fenómenos cualitativamente complejos. Las cadenas de Markov han encontrado muchas aplicaciones en ingeniería eléctrica, por ejemplo, en el análisis de rendimiento de protocolos de comunicaciones de acceso múltiple y de redes de comunicaciones, en modulación codificada y en procesamiento de imágenes. La teoría de las cadenas de Markov han recibido un impulso adicional a partir del advenimiento de la simulación de la cadena de Markov de Monte Carlo. Las cadenas de Markov han encontrado un dominio privilegiado de aplicación en la investigación de operaciones, por ejemplo en la teoría de la confiabilidad y la teoría de colas.

## Objetivos

Este trabajo está dedicado al estudio de cadenas de Markov homogéneas (CMH) con un espacio de estados contable, en tiempo discreto. El trabajo refleja los avances recientes de la teoría y responde a la creciente necesidad de un tratamiento unificado de temas relacionados como campos de Gibbs finitos, procesos regenerativos a tiempo discreto y simulación Monte Carlo.

Una parte del trabajo está dedicada a la teoría básica (Capítulos 1,2). Esta parte del trabajo introduce a los *modelos de Markov a tiempo discreto* después se desarrollan las propiedades de *recurrencia y ergodicidad* (Capítulo 3) así como de su comportamiento a largo plazo (Capítulo 4). Los campos de Gibbs y la simulación de Monte Carlo (Capítulo 5) son temas muy importantes, de interés en física, procesamiento de imágenes y optimización. Hay bastantes buenas razones para incluir campos de Gibbs en un trabajo dedicado a las cadenas de Markov, además de la observación de que generalizan de forma natural la definición de Markov en la línea

real a conjuntos de índices discretos arbitrarios. Una de estas razones es que son un dominio privilegiado de aplicación de los algoritmos de simulación de cadena de Monte Carlo Markov. Este último es el último tema del Capítulo 5.

Aunado a esto, en cada capítulo se muestran ejemplos donde son aplicados los conceptos teóricos. Estos ejemplos muestran el alcance que pueden llegar a tener las cadenas de Markov vistas desde una perspectiva práctica, mostrando lo importante de las hipótesis y de los resultados de cada sección a través del desarrollo de ciertos ejemplos a lo largo del capítulo. Uno de estos es el de las *redes neuronales*, desarrolladas en el último capítulo, obteniendo un compendio de resultados importantes acerca del tema.

Claramente hay muchos temas de importancia en el estudio de cadenas de Markov, como los modelos de Markov homogéneos a tiempo continuo los cuales son esencialmente cadenas de Markov homogéneas a tiempo discreto con una escala de tiempo aleatoria. El tiempo que separa dos transiciones sucesivas no es una unidad, sino un tiempo exponencial que depende del estado actual. (Esta descripción rápida, por supuesto, salta los detalles técnicos finos). Otra afirmación informal es que las cadenas de Markov homogéneas a tiempo continuo son cadenas de Markov homogéneas a tiempo discreto con una pequeña dosis de procesos Poisson. El alcance práctico de temas desarrollados en el trabajo o como el mencionado anteriormente es basto y da pie a la investigación e implementación de estos.



---

---

# CAPÍTULO 1

---

## TEORÍA DE LA PROBABILIDAD

### 1.1. Conceptos básicos

El presente capítulo trata acerca de los conceptos esenciales de teoría de probabilidad directamente útiles para los capítulos posteriores y la información es suficiente para estudiar las Cadenas de Markov.

#### 1.1.1. Eventos

La teoría de la probabilidad proporciona un marco matemático para el estudio de fenómenos aleatorios. Requiere una descripción precisa del resultado de un experimento cuando tal fenómeno es observado. La colección de todos los resultados posibles  $\omega$  se denomina espacio muestral  $\Omega$ .

**Ejemplo 1.1.1.** Un posible experimento consiste en lanzar un dado una vez. Los posibles resultados son  $\omega = 1, 2, \dots, 6$  y el espacio muestral es el conjunto  $\Omega = \{1, 2, 3, 4, 5, 6\}$

**Ejemplo 1.1.2.** Otro posible experimento consiste en lanzar un dardo a una pared muy grande. El espacio muestral se puede ser el plano  $\mathbb{R}^2$ , una idealización de la pared, un resultado es la posición  $\omega = (x, y)$  golpeado por el dardo.

**Ejemplo 1.1.3.** También se puede considerar una sucesión infinita de lanzamientos de monedas. Se puede tomar como espacio muestral  $\Omega$  la colección de todas las sucesiones  $\omega = \{x_n\}_{n \geq 1}$ , donde  $x_n = 1$  ó  $0$ , dependiendo de si el  $n$ -ésimo lanzamiento da como resultado cara o cruz.

La teoría de la probabilidad tiene su propia interpretación de conjuntos. Se refleja en su terminología, que ahora se revisará. Cualquier subconjunto  $A$  del espacio muestral  $\Omega$  puede considerarse como una representación de algún evento. En el ejemplo 1.1.3,  $A = \{1, 3, 5\}$  es el evento «el resultado es impar». En el ejemplo 1.1.2, el subconjunto  $A = \{(x, y) : x^2 + y^2 \geq 1\}$  podría ser el evento «dar afuera de la diana» si la diana está centrada en  $0$  y de radio  $1$ . En el ejemplo 1.1.3, el

subconjunto  $A = \{\omega : x_k = 1 \text{ para } k \text{ de } 1 \text{ a } 1000\}$  es un evento muy afortunado si se apuesta siempre a cara.

La teoría de la probabilidad asigna a cada evento un número, la *probabilidad* de dicho evento. Por razones técnicas, la colección  $\mathcal{F}$  de eventos que son asignados a una probabilidad no siempre es idéntica a la colección de todos los subconjuntos de  $\Omega$ . Los requerimientos de  $\mathcal{F}$  son los siguientes:

1. El evento imposible  $\emptyset$  y el evento seguro  $\Omega$  están en  $\mathcal{F}$ .
2. Si  $A$  está en  $\mathcal{F}$ , entonces  $\Omega \setminus A = A^c$  también está.
3. Si  $A_1, A_2, \dots$  están en  $\mathcal{F}$ , entonces también lo está  $\bigcup_{k=1}^{\infty} A_k$ .

A la colección de subconjuntos  $\mathcal{F}$  se le conoce como *sigma-álgebra* de  $\Omega$ .

Si el espacio muestral  $\Omega$  es finito, usualmente se considera a cualquier subconjunto de  $\Omega$  un evento.

### 1.1.2. Variables Aleatorias

**Definición 1.1.1.** Una variable aleatoria es una función  $X : \Omega \rightarrow \overline{\mathbb{R}}$  tal que para cada  $a \in \mathbb{R}$ , a el evento  $\{\omega : X(\omega) \leq a\} = \{X \leq a\}$  se le puede asignar una probabilidad, esto es,  $\{X \leq a\} \in \mathcal{F}$ . [31]

Una función  $X : \Omega \rightarrow E$  donde  $E$  es un conjunto numerable es llamada *variable aleatoria discreta* si para toda  $x \in E$ ,  $\{X = x\} \in \mathcal{F}$ .

**Ejemplo 1.1.4.** Siguiendo el ejemplo 1.1.1, se puede definir a  $X$  como la función identidad  $X(\omega) = \omega$ . En el sentido de que  $X$  toma los valores que se obtengan de lanzar el dado.

**Ejemplo 1.1.5.** De acuerdo al ejemplo 1.1.2, aquí  $\omega = (x, y)$ , donde  $x, y \in \mathbb{R}$ . Se definen las variables aleatorias  $X, Y$  como  $X(\omega) = x, Y(\omega) = y$ .

**Ejemplo 1.1.6.** Continuando con el ejemplo 1.1.3, donde  $\omega = \{x_n\}_{n \geq 1}$ . Se define  $X_n$  como  $X_n(\omega) = x_n$ , el número obtenido en el  $n$ -ésimo lanzamiento.

### 1.1.3. Medidas de Probabilidad

La *probabilidad*  $P(A)$  de un evento  $A \in \mathcal{F}$  mide la posibilidad de que un evento ocurra. Como una función definida sobre  $\mathcal{F}$ , la medida de probabilidad  $P$  necesita satisfacer algunas propiedades, los *axiomas de probabilidad*.

**Definición 1.1.2.** [33] Una medida de probabilidad sobre el espacio  $(\Omega, \mathcal{F})$  es una función  $P : \mathcal{F} \rightarrow \mathbb{R}$  tal que:

1.  $0 \leq P(A) \leq 1$
2.  $P(\Omega) = 1$

$$3. P(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k)$$

Para  $A_i \neq A_j$  con  $i \neq j$

A la propiedad 3 se le conoce como *sigma-aditividad*. La tríada  $(\Omega, \mathcal{F}, P)$  es llamado *espacio de probabilidad*. Los axiomas de probabilidad están inspirados por la siguiente interpretación heurística de  $P(A)$  como la frecuencia empírica del evento  $A$ . Si  $n$  experimentos “independientes” son realizados, entre los cuales  $n_A$  experimentos resultan en la ocurrencia de  $A$ , entonces la frecuencia empírica

$$F(A) = \frac{n_A}{n}$$

debe estar próxima a  $P(A)$ , si  $n$  es lo “suficientemente grande”. Claramente, la función  $F$  satisface los axiomas.

No obstante, la presentación axiomática de la teoría de la probabilidad es lógicamente independiente de la interpretación de la frecuencia. De hecho, su éxito se debe a su aparente desconocimiento de la interpretación de la frecuencia, lo que desdibuja el panorama porque la frecuencia empírica  $F$  depende de demasiadas cosas: del número de experimentos y de los experimentos mismos. La teoría axiomática de la probabilidad se conecta a la interpretación de la frecuencia a posteriori: esta última aparece como un teorema, la famosa ley fuerte de los grandes números. Para obtenerla, todo lo que se necesita además de los axiomas de probabilidad y cálculos inteligentes es una buena definición de lo que se entiende por experimentos independientes.

**Ejemplo 1.1.7.** Continuando el ejemplo 1.1.1, para  $A \subset \Omega = \{1, 2, 3, 4, 5, 6\}$ , la fórmula  $P(A) = \frac{|A|}{6}$ , donde  $|A|$  es la cardinalidad de  $A$ , es decir, el número de elementos de  $A$ , define una medida de probabilidad  $P$ .

**Ejemplo 1.1.8.** Continuando el ejemplo 1.1.2, se puede tomar, por ejemplo,

$$P(A) = \frac{1}{2\pi\sigma^2} \iint_A e^{-\frac{x^2+y^2}{2\sigma^2}} dx dy$$

Puede demostrarse que  $P(\Omega) = 1$ . De hecho,

$$\begin{aligned} \frac{1}{2\pi\sigma^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-\frac{x^2+y^2}{2\sigma^2}} dx dy &= \frac{1}{\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-(x^2+y^2)} dx dy \\ &= \frac{1}{2\pi} \int_0^{2\pi} \int_0^{+\infty} e^{-\rho^2} 2\rho d\rho d\theta \\ &= \frac{1}{2\pi} \int_0^{2\pi} \int_0^{+\infty} e^{-u} du d\theta = 1 \end{aligned}$$

**Ejemplo 1.1.9.** Continuando el ejemplo 1.1.3, se puede escoger una medida de probabilidad  $P$  tal que para cualquier evento de la forma  $A = \{x_1 = a_1, \dots, x_k = a_k\}$ , donde  $a_1, \dots, a_k$  son arbitrarios en  $\{0, 1\}$ ,  $P(A) = \frac{1}{2^k}$ .

La elección de la medida de probabilidad  $P$  en estos ejemplos es arbitraria, y muchas otras opciones son posibles. En el ejemplo 1.1.7 se supone que el dado es

justo y cada número tiene la misma probabilidad en salir. La medida de probabilidad en el ejemplo 1.1.9 supone una moneda justa y lanzamientos independientes.

Para los axiomas de probabilidad, las siguientes propiedades son fáciles de demostrar [33]:

$$P(A^c) = 1 - P(A) \quad (1.1)$$

$$P(\emptyset) = 0 \quad (1.2)$$

$$\text{Si } A \subset B \implies P(A) \leq P(B), \quad (1.3)$$

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} P(A_k) \quad (1.4)$$

**Demostración.** Para 1.1,  $1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$ . Para 1.2  $P(\emptyset) = 1 - P(\Omega) = 0$ . Para 1.3, como  $B \subset A$ ,  $P(B) = P(A) + P(B \setminus A) \geq P(A)$ . Y finalmente para 1.4,

$$\bigcup_{k=1}^{\infty} A_k = \bigcup_{k=1}^{\infty} A_k \cap \left\{ \left( \bigcup_{i=1}^{k-1} A_i \right)^c \right\} = \bigcup_{k=1}^{\infty} B_k$$

Como cada elemento de la última unión es disjunto de cualquier otro:

$$\begin{aligned} P\left(\bigcup_{k=1}^{\infty} A_k\right) &= P\left(\bigcup_{k=1}^{\infty} A_k \cap \left\{ \left( \bigcup_{i=1}^{k-1} A_i \right)^c \right\}\right) \\ &= \sum_{k=1}^{\infty} P\left(A_k \cap \left\{ \left( \bigcup_{i=1}^{k-1} A_i \right)^c \right\}\right) \\ &= \sum_{k=1}^{\infty} P(B_k) \end{aligned}$$

Y como  $B_k \subset A_k$

$$\sum_{k=1}^{\infty} P(B_k) \leq \sum_{k=1}^{\infty} P(A_k)$$

**Teorema 1.1.1.** [12] Sea  $\{A_n\}_{n \geq 1}$  una sucesión no decreciente de eventos, es decir, para toda  $n \geq 1$ ,  $A_n \subset A_{n+1}$ . Entonces:

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \uparrow \infty} \uparrow P(A_n) \quad (1.5)$$

Similarmente, si  $\{B_n\}_{n \geq 1}$  es una sucesión no creciente de eventos, es decir, para toda  $n \geq 1$ ,  $B_{n+1} \subset B_n$ . Entonces:

$$P\left(\bigcap_{n=1}^{\infty} B_n\right) = \lim_{n \uparrow \infty} \downarrow P(B_n) \quad (1.6)$$

**Demostración.** Para 1.5 se puede escribir  $A_n$  como:

$$A_n = A_1 \cup (A_2 \setminus A_1) \cup \cdots \cup (A_n \setminus A_{n-1})$$

y como

$$\bigcup_{k=1}^{\infty} A_k = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \dots$$

por lo tanto:

$$\begin{aligned} P\left(\bigcup_{k=1}^{\infty} A_k\right) &= P(A_1) + \sum_{j=1}^{\infty} P(A_j \setminus A_{j-1}) \\ &= \lim_{n \uparrow \infty} \left\{ P(A_1) + \sum_{j=1}^n P(A_j \setminus A_{j-1}) \right\} \\ &= \lim_{n \uparrow \infty} P(A_n) \end{aligned}$$

Para 1.6 se escribe a  $P(\bigcap_{n=1}^{\infty} B_n)$  como:

$$P\left(\bigcap_{n=1}^{\infty} B_n\right) = 1 - P\left(\left\{\bigcap_{n=1}^{\infty} B_n\right\}^c\right) = 1 - P\left(\bigcup_{n=1}^{\infty} B_n^c\right)$$

y así usar 1.5 con  $A_n = B_n^c$

## 1.2. Probabilidad Condicional e Independencia

### 1.2.1. Independencia de Eventos y de Variables Aleatorias

**Definición 1.2.1.** [31] Dos eventos  $A$  y  $B$  son llamados *independientes* si:

$$P(A \cap B) = P(A)P(B) \quad (1.7)$$

Dos variables aleatorias  $X$  y  $Y$  son llamados independientes si para toda  $a, b \in \mathbb{R}$ :

$$P(X \leq a, Y \leq b) = P(X \leq a)P(Y \leq b) \quad (1.8)$$

En la definición 1.2.1 hay que recordar que  $P(X \leq a) = P(\{\omega : X(\omega) \leq a\})$  y  $P(X \leq a, Y \leq b) = P(\{\omega : X(\omega) \leq a\} \cap \{\omega : Y(\omega) \leq b\})$ .

**Ejemplo 1.2.1.** Este ejemplo es una continuación del ejemplo 1.1.8. Las variables aleatorias  $X$  y  $Y$  son independientes. Además,  $\{(x, y) \in \mathbb{R}^2 : x \leq a, y \leq b\} = \{X \leq a\} \cap \{Y \leq b\}$ , y por lo tanto:

$$\begin{aligned} P(\{X \leq a\} \cap \{Y \leq b\}) &= \frac{1}{2\pi\sigma^2} \int_{-\infty}^a \int_{-\infty}^b e^{-\frac{x^2+y^2}{2\sigma^2}} dx dy \\ &= \left( \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^a e^{-\frac{x^2}{2\sigma^2}} dx \right) \left( \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^b e^{-\frac{y^2}{2\sigma^2}} dy \right) \\ &= P(X \leq a)P(Y \leq b), \end{aligned}$$

Donde la última igualdad se da porque

$$P(X \leq a) = \lim_{n \uparrow \infty} P(X \leq a, Y \leq n)$$

ya que por el ejemplo 1.1.8

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2\sigma^2}} dy = 1$$

**Ejemplo 1.2.2.** Sea  $A, B$  dos conjuntos tales que  $A \cap B = \emptyset$ ,  $A, B$  son independientes si y sólo si al menos uno tiene probabilidad cero. Esto se da ya que si  $P(A \cap B) = P(\emptyset) = 0$ , la igualdad  $P(A \cap B) = P(A)P(B)$  se da si  $P(A)P(B) = 0$  y esto ocurre si y sólo si alguna de las dos probabilidades es cero. Este ejemplo demuestra que eventos disjuntos no significa eventos independientes.

**Definición 1.2.2.** [31] Una familia  $\{A_i\}_{i \in I}$  de eventos con índices en un conjunto arbitrario  $I$  es llamada independiente si para cualquier colección finita de eventos distintos  $A_{i_1}, \dots, A_{i_k}$  de esta familia,

$$P\left(\bigcap_{j=1}^k A_{i_j}\right) = \prod_{j=1}^k P(A_{i_j}) \quad (1.9)$$

Una familia  $\{X_i\}_{i \in I}$  de variables aleatorias reales con índices en un conjunto arbitrario  $I$  es llamado independiente si para cualquier colección finita de eventos distintos  $A_{i_1}, \dots, A_{i_k}$  de esta familia,

$$P\left(\bigcap_{j=1}^k \{X_{i_j} \leq a_j\}\right) = \prod_{j=1}^k P(X_{i_j} \leq a_j) \quad (1.10)$$

para los números reales  $a_1, \dots, a_k$ . La familia de variables aleatorias  $\{X_i\}_{i \in I}$  se dice que es independiente de la familia de variables aleatoria  $\{Y_j\}_{j \in J}$  si

$$P\left(\left(\bigcap_{l=1}^r \{X_{i_l} \leq a_l\}\right) \cap \left(\bigcap_{m=1}^s \{Y_{j_m} \leq b_m\}\right)\right) = P\left(\bigcap_{l=1}^r \{X_{i_l} \leq a_l\}\right) P\left(\bigcap_{m=1}^s \{Y_{j_m} \leq b_m\}\right) \quad (1.11)$$

para los índices  $i_1, \dots, i_r \in I$  y  $j_1, \dots, j_s \in J$ , y los números reales  $a_1, \dots, a_r, b_1, \dots, b_s$ . Aquí  $I, J$  son conjuntos de índices arbitrarios.

**Ejemplo 1.2.3.** Esta es una continuación del ejemplo 1.1.9. El evento  $\{X_k = a_k\}$  es la ocurrencia de los eventos  $\{X_1 = a_1, \dots, X_{k-1} = a_{k-1}, X_k = a_k\}$  con  $a_1, \dots, a_k \in \{0, 1\}$  para todos los posibles valores de  $(a_1, \dots, a_k)$ . Ya que hay  $2^k$  posibles valores para el vector anterior, cada posibilidad tiene probabilidad  $2^{-k}$ , si dejamos fijo  $a_k$  tenemos que  $P(X_k = a_k) = 2^{k-1}2^{-k} = \frac{1}{2}$ , es decir,  $P(X_k = 0) = P(X_k = 1) = \frac{1}{2}$  y por lo tanto,

$$P(X_1 = a_1, \dots, X_k = a_k) = \prod_{i=1}^k P(X_i = a_i) = \frac{1}{2^k}$$

para toda  $a_1, \dots, a_k \in \{0, 1\}$ , para los cuales se sigue por definición que  $X_1, \dots, X_k$  son variables aleatorias independientes, y  $\{X_n\}_{n \geq 1}$  es una familia de variables aleatorias independientes. Se puede decir que se tiene un modelo para los lanzamientos independientes de una moneda justa.

### 1.2.2. Reglas de Bayes

En la interpretación frecuentista, la definición de independencia 1.7, a groso modo, en términos imprecisos y usando la notación de 1.1.3,

$$\frac{n_{A \cap B}}{n} \approx \left(\frac{n_A}{n}\right)\left(\frac{n_B}{n}\right)$$

ó

$$\frac{n_{A \cap B}}{n_B} \approx \left(\frac{n_A}{n}\right)$$

es motivado por la idea de que las ocurrencias que refieren al evento  $A$  no varían cuando se realizan en una muestra neutral de población o en una muestra seleccionada de población caracterizada por la propiedad  $B$ . Por ejemplo, la proporción de personas con un apellido que empiece con  $H$  es similar en una población grande como lo es tomando en cuenta solo a los hombres. Esto es gran parte de la noción intuitiva de la independencia.

La dependencia entre  $A$  y  $B$  ocurre cuando  $P(A \cap B) \neq P(A)P(B)$ . en el caso de la frecuencia relativa  $\frac{n_{A \cap B}}{n_B} \approx \frac{P(A \cap B)}{P(B)}$  lo cual es distinto de  $\frac{n_A}{n}$ .

**Definición 1.2.3.** [33] La probabilidad condicional del evento  $A$  dado  $B$  es

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (1.12)$$

definida cuando  $P(B) > 0$ .

La cantidad  $P(A|B)$  representa la probabilidad de que  $A$  ocurra cuando la única información relevante disponible es que  $B$  ocurrió.

Una forma simétrica de 1.12 es

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A) \quad (1.13)$$

Claramente, cuando  $A$  y  $B$  son independientes, entonces

$$P(A|B) = P(A), \quad P(B|A) = P(B) \quad (1.14)$$

**Teorema 1.2.1.** [33] Sean  $A, B$  eventos de  $\Omega$ , con  $P(B) > 0$ , entonces

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1.15)$$

**Demostración.** Por 1.12 y luego 1.13

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)} \quad (1.16)$$

**Teorema 1.2.2.** [33] Para una partición de  $\Omega$ ,  $B_1, B_2, \dots$  con  $\cup_{i=1}^{\infty} B_i = \Omega$  y para toda  $A \subset \Omega$ , se cumple que

$$P(A) = \sum_{i=1}^{\infty} P(A|B_i)P(B_i) \quad (1.17)$$

**Demostración.** Solo hay que descomponer  $A$  como sigue

$$\begin{aligned} P(A) &= P\left(A \cap \left(\bigcup_{i=1}^{\infty} B_i\right)\right) = P\left(\bigcup_{i=1}^{\infty} (A \cap B_i)\right) \\ &= \sum_{i=1}^{\infty} P(A \cap B_i) = \sum_{i=1}^{\infty} P(A|B_i)P(B_i) \end{aligned}$$

**Teorema 1.2.3.** [33] Para cualquier sucesión de eventos  $A_1, \dots, A_k$  en  $\Omega$

$$P\left(\bigcap_{i=1}^k A_i\right) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P\left(A_k \mid \bigcap_{i=1}^{k-1} A_i\right) \quad (1.18)$$

**Demostración.** Se procederá por inducción, es claro que se cumple para  $k = 1$ , si se supone que se cumple para  $k$ .

$$P\left(\bigcap_{i=1}^{k+1} A_i\right) = P\left(\left(\bigcap_{i=1}^k A_i\right) \cap A_{k+1}\right) = P\left(A_{k+1} \mid \bigcap_{i=1}^k A_i\right)P\left(\bigcap_{i=1}^k A_i\right)$$

se sustituye a  $P(\bigcap_{i=1}^k A_i)$  por lo asumido y se obtiene 1.18.

Un ejemplo común para aplicar estos resultados es el de una prueba en un examen médico.

**Ejemplo 1.2.4.** Un doctor aplica una prueba a un paciente para saber si ha sido afectado por una enfermedad, la prueba detecta la enfermedad en el 99% de los casos. Sin embargo, pasa que en el 2% de los casos donde un paciente está sano, la prueba resulta positiva. También se conoce que una de cada mil personas tiene la enfermedad. La pregunta que se hace es ¿cuál es la probabilidad de que un paciente tenga la enfermedad dado que la prueba dio positiva?

Para responder a esta pregunta tomemos a  $M$  como el evento “el paciente está enfermo”, y  $+$ ,  $-$  los eventos “la prueba dio positivo” y “la prueba dio negativo”. Se conoce que

$$P(M) = 0.001, \quad P(+|M) = 0.99, \quad P(+|M^c) = 0.02$$

y que debemos encontrar  $P(M|+)$ . Por el teorema 1.2.1

$$P(M|+) = \frac{P(+|M)P(M)}{P(+)}$$

y por el teorema 1.2.2,

$$P(+) = P(+|M)P(M) + P(+|M^c)P(M^c)$$

por lo tanto,

$$P(M|+) = \frac{(0.99)(0.001)}{(0.99)(0.001) + (0.02)(0.999)} = 0.047$$

La probabilidad pequeña. Sin embargo, la situación no es tan mala como se ve, ya que en vista de los resultados ambiguos, todos los pacientes que den positivo se pueden hacer la prueba de nuevo, esta vez con un mejor resultado (podría resultar más caro). La ganancia está en que solo una pequeña porción de la población gastará en esa segunda prueba ya que  $P(+) = 0.003$ .



### 1.2.3. Propiedad de Markov

**Definición 1.2.4.** [22] Se dice que  $A$  y  $B$  son condicionalmente independientes dado  $C$  si

$$P(A \cap B|C) = P(A|C)P(B|C) \quad (1.19)$$

Sean  $X, Y, Z$  variables aleatorias discretas con dominios  $E, F, G$  respectivamente. Se dicen que  $X$  y  $Y$  son condicionalmente independientes dado  $Z$  si para toda  $x, y, z$  en  $E, F, G$ , respectivamente, los eventos  $\{X = x\}$  y  $\{Y = y\}$  son condicionalmente independientes dado  $\{Z = z\}$ .

Se puede notar que, para un conjunto  $C$  dado

$$P_C(A) = P(A|C) \quad (1.20)$$

define una medida de probabilidad  $P_C$ . La ecuación 1.19 expresa la independencia de  $A$  y  $B$  con respecto a la medida de probabilidad  $P_C$ .

**Teorema 1.2.4.** [22] Sean  $A_1, A_2, A_3$  tres eventos con probabilidad positiva. Los eventos  $A_1$  y  $A_3$  son condicionalmente independientes dado  $A_2$  si y sólo si la **Propiedad de Markov** se cumple, esto es,

$$P(A_3|A_1 \cap A_2) = P(A_3|A_2) \quad (1.21)$$

**Demostración.** Si se asume la independencia condicional. Entonces

$$\begin{aligned} P(A_3|A_1 \cap A_2) &= \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)} \\ &= \frac{P(A_1 \cap A_3|A_2)P(A_2)}{P(A_1 \cap A_2)} \\ &= \frac{P(A_1|A_2)P(A_3|A_2)P(A_2)}{P(A_1 \cap A_2)} \\ &= \frac{P(A_1|A_2)P(A_3|A_2)}{P(A_1|A_2)} \\ &= P(A_3|A_2) \end{aligned}$$

Si se asume la propiedad de Markov. Entonces

$$\begin{aligned} P(A_1 \cap A_3|A_2) &= \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_2)} \\ &= \frac{P(A_3|A_1 \cap A_2)P(A_1 \cap A_2)}{P(A_2)} \\ &= P(A_3|A_1 \cap A_2)P(A_1|A_2) \\ &= P(A_3|A_2)P(A_1|A_2) \end{aligned}$$

**Teorema 1.2.5.** [22] Sean  $X, Y, Z$  tres variables aleatorias discretas con dominios  $E, F, G$  respectivamente. Si para alguna función  $g : E \times F \rightarrow [0, 1]$  se cumple que

$$P(X = x|Y = y, Z = z) = g(x, y)$$

para toda  $x, y, z$ , entonces  $P(X = x|Y = y) = g(x, y)$  para toda  $x, y$ . También  $X, Y$  son condicionalmente independientes dado  $Z$ .

**Demostración.** Tenemos que

$$\begin{aligned}
 P(X = x, Y = y) &= \sum_z P(X = x, Y = y, Z = z) \\
 &= \sum_z P(X = x|Y = y, Z = z)P(Y = y, Z = z) \\
 &= g(x, y) \sum_z P(Y = y, Z = z) \\
 &= g(x, y)P(Y = y)
 \end{aligned}$$

Por lo tanto,

$$P(X = x|Y = y) = g(x, y) = P(X = x|Y = y, Z = z).$$

También  $X, Y$  son condicionalmente independientes por el teorema 1.2.4.

**Ejemplo 1.2.5.** Dos fábricas  $A$  y  $B$  manufacturan relojes. La fábrica  $A$  produce en promedio un artículo defectuoso de cada 100, y  $B$  uno de cada 200. Un comercio recibe un contenedor con relojes de una de las dos fábricas, pero no sabe de cuál.

Sea  $X_n$  el estado del  $n$ -ésimo reloj en el contenedor, con  $X_n = 1$  si funciona y  $X_n = 0$  si no funciona. Sea  $Y$  la fábrica de origen y supongamos que

$$P(Y = A) = P(Y = B) = \frac{1}{2}$$

y

$$P(X_i = 1, X_j = 0|Y = A) = P(X_i = 1|Y = A)P(X_j = 0|Y = A) \text{ para } i \neq j$$

Sabemos que

$$P(X_n = 0|Y = A) = 0.01, \quad P(X_n = 0|Y = B) = 0.005$$

El comercio revisa el primer reloj y funciona, si se quiere calcular la probabilidad de que el segundo reloj también funcione  $P(X_2 = 1|X_1 = 1)$ , ósea

$$\frac{P(X_1 = 1, X_2 = 1)}{P(X_1 = 1)}$$

por el teorema 1.2.2

$$\begin{aligned}
 P(X_1 = 1, X_2 = 1) &= P(X_1 = 1, X_2 = 1|Y = A)P(Y = A) \\
 &\quad + P(X_1 = 1, X_2 = 1|Y = B)P(Y = B) \\
 &= \frac{1}{2} \left(\frac{99}{100}\right)^2 + \frac{1}{2} \left(\frac{199}{200}\right)^2
 \end{aligned}$$

y aplicando lo mismo

$$\begin{aligned}
 P(X_1 = 1) &= P(X_1 = 1|Y = A)P(Y = A) \\
 &\quad + P(X_1 = 1|Y = B)P(Y = B) \\
 &= \frac{1}{2} \left(\frac{99}{100}\right) + \frac{1}{2} \left(\frac{199}{200}\right)
 \end{aligned}$$

Por lo tanto

$$P(X_2 = 1|X_1 = 1) = \frac{\left(\frac{99}{100}\right)^2 + \left(\frac{199}{200}\right)^2}{\left(\frac{99}{100}\right) + \left(\frac{199}{200}\right)}$$

Se puede observar que  $X_i$  no es independiente de  $X_j$  para  $i \neq j$ , de hecho si lo fueran

$$P(X_2 = 1|X_1 = 1) = P(X_2 = 1) = \frac{1}{2}\left(\frac{99}{100}\right) + \frac{1}{2}\left(\frac{199}{200}\right)$$

Un resultado distinto del anterior. Este ejemplo muestra como, incluso para algún evento  $C$ , dos eventos  $A$  y  $B$  pueden ser condicionalmente independientes dado  $C$ , condicionalmente independientes dado  $C^c$ , y aún así no necesitan ser independientes por si solos.

## 1.3. Esperanza

### 1.3.1. Función de Distribución

Una variable aleatoria  $X$  es descrita por su Función de Distribución

$$F(x) = P(X \leq x) \tag{1.22}$$

con  $x \in \mathbb{R}$ . La función de distribución tiene las siguientes propiedades:

- (i)  $F : \mathbb{R} \longrightarrow [0, 1]$
- (ii)  $F$  es no decreciente
- (iii)  $F$  es continua por la derecha
- (iv)  $F(+\infty) = \lim_{a \uparrow \infty} F(a) = P(X < \infty)$
- (v)  $F(-\infty) = \lim_{a \downarrow -\infty} F(a) = P(X = -\infty)$

**Demostración.** (i) Como  $x \in \mathbb{R}$  y  $P$  tiene como imagen  $[0, 1]$ ,  $F : \mathbb{R} \longrightarrow [0, 1]$

(ii) Si  $a \leq b$ , entonces  $\{X \leq a\} \subset \{X \leq b\}$ , y por 1.3 sucede que

$$P(X \leq a) \leq P(X \leq b)$$

- (iii) Se aplica 1.6 para  $B_n = \{X \leq a + \frac{1}{n}\}$ . Y como  $\cap_{n \geq 1} B_n = \{X \leq a\}$ , luego se tiene que  $\lim_{n \uparrow \infty} P(X \leq a + \frac{1}{n}) = P(X \leq a)$  y se llega a la conclusión,  $F$  es continua por la derecha.
- (iv) Se aplica 1.5 con  $A_n = \{X \leq n\}$ , así se tiene  $\cup_{n \geq 1} A_n = \{X < \infty\}$ , luego  $\lim_{n \uparrow \infty} P(X \leq n) = P(X < \infty)$  y se concluye que  $F(+\infty) = \lim_{a \uparrow \infty} F(a) = P(X < \infty)$

(v) Se aplica 1.6 con  $B_n = \{X \leq -n\}$ , así se tiene  $\cap_{n \geq 1} B_n = \{X < -\infty\}$ , luego  $\lim_{n \uparrow \infty} P(X \leq -n) = P(X < -\infty)$  y se concluye que  $F(-\infty) = \lim_{a \downarrow -\infty} F(a) = P(X < -\infty)$

Como  $F$  es no decreciente, se define lo siguiente

$$F(x-) = \lim_{h \downarrow 0} F(x - h) \quad (1.23)$$

el límite existe para toda  $x \in \mathbb{R}$ . Ahora, la sucesión  $B_n = \{a - \frac{1}{n} < X \leq a\}$  es decreciente y,  $\cap_{n \geq 1} B_n = \{X = a\}$ . Usando la propiedad 1.5 se obtiene lo siguiente:

$$P(X = a) = \lim_{n \uparrow \infty} P(a - \frac{1}{n} < X \leq a) = \lim_{n \uparrow \infty} \left( F(a) - F(a - \frac{1}{n}) \right) \quad (1.24)$$

que es lo mismo que

$$P(X = a) = F(a) - F(a-) \quad (1.25)$$

En particular,  $F$  es continua si y sólo si  $P(X = a) = 0$  para cualquier  $a \in \mathbb{R}$ . Por 1.25 y la definición de  $F$ , se tiene que

$$\begin{aligned} P(a < X \leq b) &= F(b) - F(a), \\ P(a \leq X < b) &= F(b-) - F(a-), \\ P(a \leq X \leq b) &= F(b) - F(a-), \\ P(a < X < b) &= F(b-) - F(a) \end{aligned} \quad (1.26)$$

Y también, con lo anterior

$$\begin{aligned} P(X = \infty) &= 1 - F(\infty) \\ P(X = -\infty) &= F(-\infty) \end{aligned} \quad (1.27)$$

Un caso especial de una función de distribución es

$$F(x) = \int_{-\infty}^x f(y) dy \quad (1.28)$$

para alguna función  $f \geq 0$  llamada *función de densidad* de  $X$ .

### 1.3.2. Esperanza, media y varianza

Los puntos donde  $F$  es discontinua son numerables, como es el caso de cualquier función acotada no decreciente. Llamando  $\{d_n\}$  a la sucesión de puntos de discontinuidad, se define

$$F_d(t) = \sum_{d_n \leq t} \{F(d_n) - F(d_n-)\} \quad (1.29)$$

y

$$F_c(t) = F(t) - F_d(t) \quad (1.30)$$

Los componentes discontinuos y continuos de  $F$ , respectivamente. Se usará el símbolo

$$\int_{-\infty}^{\infty} g(x) dF(x) \quad (1.31)$$

para representar la integral de Stieltjes-Lebesgue respecto de  $F$  de la función  $g$ . En el caso especial donde el componente continuo de la función de distribución es absolutamente continuo, es decir

$$F_c(x) = \int_{-\infty}^x f_c(y) dy \quad (1.32)$$

es suficiente para poder escribir 1.31 como

$$\int_{-\infty}^{\infty} g(x) dF(x) = \sum_{n=1}^{\infty} g(d_n)(F(d_n) - F(d_n-)) + \int_{-\infty}^{\infty} g(x) f_c(x) dx \quad (1.33)$$

**Ejemplo 1.3.1.** Para un punto aleatorio dentro del cuadrado unitario  $[0, 1]^2 = [0, 1] \times [0, 1]$ , el siguiente modelo es elegido:  $\Omega = [0, 1]^2$ ,  $P(A)$  = área de  $A$ . Sea  $(X, Y)$  las variables aleatorias que representan las coordenadas y sea  $Z = X + Y$ .

La función de distribución de  $Z$ ,  $P(Z < z)$ , se puede encontrar de la siguiente manera:

$$F_Z(z) = P(Z < z) = P(X + Y < z) = P(\{(\omega_1, \omega_2) : X(\omega_1) + Y(\omega_2) < z\})$$

con  $(\omega_1, \omega_2) \in [0, 1]^2$ . Por lo que la función de distribución sería el área del conjunto  $A = \{(x, y) : x + y < z\} \cap [0, 1]^2$ , es decir, si  $0 \leq z \leq 1$ :

$$\iint_A dx dy = \int_0^z \int_0^{z-y} dx dy = \int_0^z z - y dy = z^2 - \frac{z^2}{2} = \frac{z^2}{2}$$

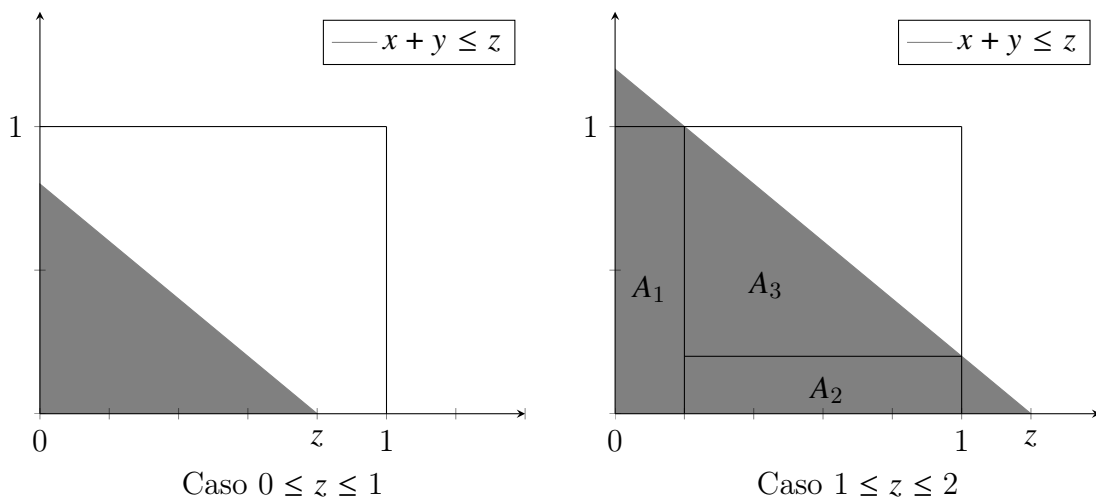
y si  $1 \leq z \leq 2$ :

$$\iint_{A_1} dx dy = \int_0^1 \int_0^{z-1} dx dy = \int_0^1 z - 1 dy = z - 1$$

$$\iint_{A_2} dx dy = \int_0^{z-1} \int_{z-1}^1 dx dy = \int_0^{z-1} 2 - z dy = (2 - z)(z - 1)$$

$$\iint_{A_3} dx dy = \int_{z-1}^1 \int_{z-1}^{z-y} dx dy = \int_{z-1}^1 1 - y dy = \frac{1}{2}(2 - z)^2$$

$$\iint_{A_1 \cup A_2 \cup A_3} dx dy = 1 - \frac{1}{2}(2 - z)^2$$



**Ejemplo 1.3.2.** Esta es una continuación del ejemplo 1.2.3. El número de ocurrencias donde la moneda cae cara en  $n$  lanzamientos es  $S_n = X_1 + \cdots + X_n$ . La variable aleatoria  $S_n$  toma valores enteros desde 0 a  $n$ . El evento  $\{S_n = k\}$  significa que  $k$  valores entre  $X_1 + \cdots + X_n$  son iguales a 1. Hay  $\binom{n}{k}$  distintas formas de que ocurran esos  $k$  eventos, todos con la misma probabilidad  $2^{-n}$ . Por lo tanto,

$$P(S_n = k) = \binom{n}{k} \frac{1}{2^n}$$

**Definición 1.3.1.** Sea  $X$  una variable aleatoria con función de distribución  $F(x) = P(X \leq x)$  y sea  $g : \overline{\mathbb{R}} \rightarrow \mathbb{R}$  tal que

$$\int_{-\infty}^{\infty} |g(x)| dF(x) < \infty \quad (1.34)$$

Entonces se define  $E[g(X)]$ , la esperanza de  $g(X)$ , con la formula

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) dF(x) \quad (1.35)$$

**Ejemplo 1.3.3.** Como continuación del ejemplo 1.3.1, la esperanza de  $Z^2$  es

$$E(Z^2) = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_0^1 x^3 dx + \int_1^2 x^2(2-x) dx \quad (1.36)$$

Y entonces  $E(Z^2) = \frac{7}{6}$

**Ejemplo 1.3.4.** Continuando el ejemplo 1.3.2, se considera  $S_n = X_1 + \cdots + X_n$ . Su

esperanza se puede calcular como sigue:

$$\begin{aligned}
E(S_n) &= \sum_{k=0}^n kP(S_n = k) \\
&= \frac{1}{2^n} \sum_{k=1}^n k \frac{n!}{k!(n-k)!} \\
&= \frac{n}{2^n} \sum_{k=1}^n \frac{(n-1)!}{(k-1)!((n-1)-(k-1))!} \\
&= \frac{n}{2^n} \sum_{j=0}^{n-1} \frac{(n-1)!}{j!(n-1-j)!} \\
&= \frac{n}{2^n} 2^{n-1} = \frac{n}{2}
\end{aligned}$$

La esperanza definida anteriormente hereda la propiedad de linealidad de la integral de Stieltjes - Lebesgue:

$$E(\lambda_1 g_1(X) + \lambda_2 g_2(X)) = \lambda_1 E(g_1(X)) + \lambda_2 E(g_2(X)) \quad (1.37)$$

donde  $\lambda_1, \lambda_2 \in \mathbb{R}$ , y  $g_1, g_2$  satisfacen la condición de integrabilidad 1.34. También la esperanza en monótona, en el sentido de que  $g_1(x) \leq g_2(x)$  para toda  $x$  implica:

$$E(g_1(X)) \leq E(g_2(X)) \quad (1.38)$$

Se define la función indicadora de  $A$  a continuación como:

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \notin A \end{cases} \quad (1.39)$$

La variable aleatoria  $X = \mathbb{1}_A$  toma el valor de 1 con probabilidad  $P(X = 1) = P(A)$  y valor de 0 con probabilidad  $P(X = 0) = P(A^c) = 1 - P(A)$ . Por lo tanto,

$$E(X) = 0 \cdot P(X = 0) + 1 \cdot P(X = 1) = P(A)$$

es decir,

$$E(\mathbb{1}_A) = P(A) \quad (1.40)$$

en particular  $E(1) = 1$ .

A veces, uno necesita definir  $E(g(X))$  para una función compleja, definida en [23],  $g : \mathbb{R} \rightarrow \mathbb{C}$ , esto es,  $g(x) = g_R(x) + ig_I(x)$  donde  $g_R$  y  $g_I$  toman valores reales. La definición de esperanza es ahora:

$$E(g(X)) = E(g_R(X)) + iE(g_I(X)) \quad (1.41)$$

siempre que las esperanzas del lado derecho sean finitas.

La desigualdad del triángulo

$$|E(g(X))| \leq E(|g(X)|) \quad (1.42)$$

es útil, y su demostración se sigue de la desigualdad análoga para la desigualdad de la integral de Stieltjes-Lebesgue.

$$\left| \int_{-\infty}^{\infty} g(x) dF(x) \right| \leq \int_{-\infty}^{\infty} |g(x)| dF(x)$$

**Definición 1.3.2.** La media  $\mu$  y varianza  $\sigma^2$  de una variable aleatoria  $X$  está definida por

$$\begin{aligned} \mu &= E(X) = \int_{-\infty}^{\infty} x dF(x) \\ \sigma^2 &= E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 dF(x) \end{aligned}$$

Claramente las integrales deben estar bien definidas. La varianza también se denota por  $\text{Var}(X)$ . De la linealidad de la esperanza, se sigue que  $E[(X - \mu)^2] = E[X^2] - 2\mu E[X] + \mu^2$ , esto es,

$$\text{Var}(X) = E[X^2] - \mu^2 \tag{1.43}$$

**Ejemplo 1.3.5.** Sean  $X_1, \dots, X_n$  variables aleatorias independientes distribuidas uniformemente sobre  $[0, 1]$ , es lo mismo que decir, con función de densidad  $f(x) = \mathbb{1}_{[0,1]}(x)$ . La esperanza de la variable aleatoria  $Z = \max(X_1, \dots, X_n)$  es,

$$P(\max(X_1, \dots, X_n) \leq z) = P(X_1 \leq z, \dots, X_n \leq z) = \prod_{k=1}^n P(X_k \leq z) = z^n$$

para  $z \in [0, 1]$ . Por lo tanto,  $Z$  tiene la función de densidad  $nz^{n-1}$  para  $z \in [0, 1]$ , y 0 en otro caso, y

$$E[Z] = \int_0^1 z(nz^{n-1}) dz = \frac{n}{n+1}$$

**Teorema 1.3.1.** Para una variable aleatoria  $X$  que toma valores en  $\mathbb{N}$ ,

$$E[X] = \sum_{n=1}^{\infty} P(X \geq n) \tag{1.44}$$

**Demostración.**

$$\begin{aligned} E[X] &= 1 \cdot P(X = 1) + 2 \cdot P(X = 2) + 3 \cdot P(X = 3) + \dots \\ &= P(X = 1) + P(X = 2) + P(X = 3) + \dots \\ &\quad + P(X = 2) + P(X = 3) + \dots \\ &\quad + P(X = 3) + \dots \\ &= P(X \geq 1) + P(X \geq 2) + P(X \geq 3) + \dots \\ &= \sum_{n=1}^{\infty} P(X \geq n) \end{aligned}$$



**Teorema 1.3.2.** Sea  $\{X_n\}_{n \geq 1}$  una sucesión de variables aleatorias tal que  $E[X_n] = E[X_1]$  para toda  $n \geq 1$ . Sea  $T$  una variable aleatoria que toma valores en  $\mathbb{N}$  tal que para toda  $n \geq 1$ , el evento  $\{T \geq n\}$  es independiente de  $X_n$ . Entonces

$$E\left[\sum_{n=1}^T X_n\right] = E[X_1]E[T] \quad (1.45)$$

**Demostración.** Sea  $S = \sum_{n=1}^T X_n$ . Entonces por la convergencia dominada [31]

$$E[S] = E\left[\sum_{n=1}^{\infty} X_n \mathbb{1}_{n \leq T}\right] = \sum_{n=1}^{\infty} E[X_n \mathbb{1}_{n \leq T}]$$

pero  $E[X_n \mathbb{1}_{n \leq T}] = E[X_n]E[\mathbb{1}_{n \leq T}] = E[X_1]P(\{T \geq n\})$ . El resultado se sigue del teorema 1.3.1.

### 1.3.3. Variables Aleatorias famosas

Sea  $\{X_n\}_{n \geq 1}$  una sucesión de variables aleatorias independientes que toman valores en  $\{0, 1\}$ , con la misma distribución de probabilidad

$$P(X_n = 1) = p$$

donde  $p \in (0, 1)$ . Como  $P(X_j = a_j) = p$  ó  $1 - p$  dependiendo si  $a_j = 1$  ó  $0$ , y ya que hay exactamente  $\sum_{j=1}^k a_j$  números entre  $a_1, \dots, a_k$  que son iguales a 1,

$$P(X_1 = a_1, \dots, X_k = a_k) = p^{\sum_{j=1}^k a_j} (1 - p)^{k - \sum_{j=1}^k a_j} \quad (1.46)$$

Se define ahora

$$S_n = X_1 + \dots + X_n$$

Esta variable aleatoria toma los valores  $0, 1, \dots, n$ . Para obtener el resultado  $S_n = i$  donde  $i \in [0, n]$ , tiene que suceder que  $X_1 = a_1, \dots, X_n = a_n$  con  $\sum_{j=1}^n a_j = i$ . Existen  $\binom{n}{i}$  distintas manera de que suceda, y cada una ocurre con probabilidad  $p^i (1 - p)^{n-i}$ . Por lo tanto para  $i \in [0, n]$ ,

$$P(S_n = i) = \binom{n}{i} p^i (1 - p)^{n-i} \quad (1.47)$$

Una variable aleatoria  $X$  con la distribución anterior es llamada *variable aleatoria binomial de tamaño  $n$  y probabilidad  $p \in (0, 1)$* . La media y la varianza son resultado de un cálculo directo y son las siguientes:

$$E[X] = np, \quad \text{Var}(X) = np(1 - p) \quad (1.48)$$

Ahora se define la variable aleatoria  $T$  como la primera ocurrencia del resultado 1, en la sucesión  $X_1, X_2, \dots$ , esto es,

$$T = \inf\{n \geq 1 : X_n = 1\}$$

con la convención de que si  $X_n = 0$  para toda  $n \geq 1$ , entonces  $T = \infty$ , aunque la probabilidad de este evento es 0. El evento  $\{T = k\}$  es exactamente  $\{X_1 = 0, \dots, X_{k-1} = 0, X_k = 1\}$ , y por lo tanto,  $P(T = k) = P(X_1 = 0) \cdots P(X_{k-1} = 0)P(X_k = 1)$ , esto es, para  $k \geq 1$ ,

$$P(T = k) = (1 - p)^{k-1}p \quad (1.49)$$

En particular,  $P(T < \infty) = \sum_{k=1}^{\infty} P(T = k) = \sum_{k=1}^{\infty} p(1 - p)^{k-1} = 1$  con  $p \in (0, 1)$ .

Una variable aleatoria con la distribución 1.49 es llamada *variable aleatoria geométrica con probabilidad p*. La esperanza de esta variable aleatoria es:

$$E[T] = \frac{1}{p} \quad (1.50)$$

Si se supone que se juega a lanzar la moneda para  $N$  número de turnos con una moneda tal que

$$P(X_n = 1) = \frac{\alpha}{N}$$

El número  $S_N$  de caras que se observan se distribuye como

$$P(S_N = k) = \binom{N}{k} \left(\frac{\alpha}{N}\right) \left(1 - \frac{\alpha}{N}\right)^{N-k}$$

De 1.48, el número promedio de veces que cae cara es  $N \cdot \frac{\alpha}{N} = \alpha$ , una constante. Es interesante observar el límite de la distribución  $S_N$  cuando  $N \rightarrow \infty$ . Denotando por  $p_N(k)$  la probabilidad de  $S_N = k$ , se tiene que

$$p_N(0) = \left(1 - \frac{\alpha}{N}\right)^N$$

se aproxima a  $e^{-\alpha}$  cuando  $N$  se aproxima a  $\infty$ . También,

$$\frac{p_N(k+1)}{p_N(k)} = \frac{\frac{N-k}{k+1} \frac{\alpha}{N}}{1 - \frac{\alpha}{N}}$$

tiende a  $\frac{\alpha}{k+1}$ . Por lo tanto, para toda  $k \geq 0$ ,

$$\lim_{n \uparrow \infty} p_n(k) = e^{-\alpha} \frac{\alpha^k}{k!}$$

con la convención de que  $0! = 1$ . Se define, una *variable aleatoria Poisson con parámetro  $\theta > 0$*  con la siguiente distribución para toda  $k \geq 0$ ,

$$P(X = k) = e^{-\theta} \frac{\theta^k}{k!} \quad (1.51)$$

Se puede observar que la distribución de  $S_N$  cuando  $N$  tiende a  $\infty$ , es la distribución de probabilidad de una variable aleatoria Poisson de parámetro  $\alpha$ .

La esperanza de esta variable aleatoria es

$$E[X] = \theta, \quad \text{Var}(X) = \theta \quad (1.52)$$

**Ejemplo 1.3.6.** Sean  $X_1, X_2$  variables aleatorias Poisson con parámetros  $\theta_1, \theta_2 > 0$  respectivamente. Para  $k \geq 0$ ,

$$\begin{aligned} P(X = k) &= P(X_1 + X_2 = k) = P\left(\bigcup_{i=0}^k \{X_1 = i, X_2 = k - i\}\right) \\ &= \sum_{i=0}^k P(X_1 = i, X_2 = k - i) = \sum_{i=0}^k P(X_1 = i)P(X_2 = k - i) \\ &= \sum_{i=0}^k e^{-\theta_1} \frac{\theta_1^i}{i!} e^{-\theta_2} \frac{\theta_2^{k-i}}{(k-i)!} = \frac{e^{-(\theta_1+\theta_2)}}{k!} (\theta_1 + \theta_2)^k \end{aligned}$$

entonces  $X = X_1 + X_2$  es una variable aleatoria Poisson con parámetro  $\theta = \theta_1 + \theta_2$ .

**Ejemplo 1.3.7.** Sea  $\{X_n\}_{n \geq 1}$  variables aleatorias independientes que toman valores en  $\{1, 0\}$  con probabilidad  $p$  y  $q = 1 - p$ , respectivamente, donde  $p \in (0, 1)$ . Sea  $T$  una variable aleatoria Poisson con parámetro  $\theta > 0$ , independiente de  $\{X_n\}_{n \geq 1}$ . Se define

$$S = X_1 + \dots + X_T$$

es decir,  $S(\omega) = X_1(\omega) + \dots + X_n(\omega)$  si  $T(\omega) = n$ . La distribución de probabilidad de  $S$  es

$$\begin{aligned} P(S = k) &= P(X_1 + \dots + X_T = k) \\ &= P\left(\bigcup_{n=k}^{\infty} \{X_1 + \dots + X_n = k, T = n\}\right) \\ &= \sum_{n=k}^{\infty} P(X_1 + \dots + X_n = k, T = n) \\ &= \sum_{n=k}^{\infty} P(X_1 + \dots + X_n = k)P(T = n) \\ &= \sum_{n=k}^{\infty} \frac{n!}{k!(n-k)!} p^k q^{n-k} e^{-\theta} \frac{\theta^n}{n!} = e^{-\theta} \frac{(p\theta)^k}{k!} \sum_{n=k}^{\infty} \frac{(q\theta)^{n-k}}{(n-k)!} \\ &= e^{-\theta} \frac{(p\theta)^k}{k!} \sum_{i=0}^{\infty} \frac{(q\theta)^i}{i!} = e^{-\theta} \frac{(p\theta)^k}{k!} e^{q\theta} = e^{-p\theta} \frac{(p\theta)^k}{k!} \end{aligned}$$

entonces  $S$  es una variable aleatoria Poisson con parámetro  $p\theta$ .

**Ejemplo 1.3.8.** Sea  $T$  la variable aleatoria que representa un número de pelotas, cada pelota es azul o rosa, con probabilidad  $p$  y  $q = 1 - p$  respectivamente. El ejercicio anterior muestra que si  $T$  es Poisson con parámetro  $\theta$ , entonces el número de pelotas azules es Poisson con parámetro  $\theta p$  y el número de pelotas rosas es Poisson con parámetro  $\theta q$ . Si  $S$  es el número de pelotas azules,  $T - S$  es el número de pelotas rosas. Entonces para cualquier entero  $k \geq 0, l \geq 0$ ,

$$\begin{aligned} P(S = k, T - S = l) &= P(S = k)P(T - S = l) \\ &= e^{-\theta p} \frac{(\theta p)^k}{k!} e^{-\theta q} \frac{(\theta q)^l}{l!} \end{aligned}$$

Ya que

$$\begin{aligned}
 P(S = k, T - S = l) &= P(S = k, T = k + l) \\
 &= P(X_1 + \dots + X_{k+l} = k, T = k + l) \\
 &= P(X_1 + \dots + X_{k+l} = k)P(T = k + l) \\
 &= \frac{(k+l)!}{k!l!} p^k q^l e^{-\theta} \frac{\theta^{k+l}}{(k+l)!} \\
 &= e^{-p\theta} \frac{(p\theta)^k}{k!} e^{-q\theta} \frac{(q\theta)^l}{l!}
 \end{aligned}$$

Lo que demuestra que el número de pelotas azules y el número de pelotas rosas es independiente.

Las siguientes variables aleatorias están definidas por su función de densidad, es decir

$$P(X \leq x) = \int_{-\infty}^x f(x) dx \quad (1.53)$$

donde  $f(x) \geq 0$ , y como  $X$  es real,

$$\int_{-\infty}^{\infty} f(x) dx = 1 \quad (1.54)$$

Una variable aleatoria  $X$  con función de densidad

$$f(x) = \frac{1}{b-a} \quad (1.55)$$

para  $x \in [a, b]$  y 0 en otro caso es llamada *variable aleatoria uniforme sobre  $[a, b]$* . Su media y varianza son:

$$E[X] = \frac{a+b}{2}, \quad \text{Var}(X) = \frac{(b-a)^2}{12} \quad (1.56)$$

Una variable aleatoria  $X$  con función de densidad

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}} \quad (1.57)$$

donde  $\mu \in \mathbb{R}$  y  $\sigma > 0$ , es llamada *variable aleatoria Gaussiana*. Su esperanza es  $\mu$  y varianza  $\sigma^2$ .

Una variable aleatoria  $X$  con función de densidad

$$f(x) = \lambda e^{-\lambda x} \mathbb{1}_{\{x \geq 0\}} \quad (1.58)$$

es llamada *variable aleatoria exponencial con parámetro  $\lambda$* . Su media es  $E[X] = \frac{1}{\lambda}$ .

## 1.4. Vectores Aleatorios

### 1.4.1. Vectores Aleatorios continuos

Un vector aleatorio de dimensión  $n$  es una colección de  $n$  variables aleatorias.

$$X = (X_1, \dots, X_n) \quad (1.59)$$

Cada una de las variables aleatorias  $X_1, \dots, X_n$  pueden ser caracterizadas desde el punto de vista probabilístico con su función de distribución.

Sin embargo, la función de distribución de cada una de las variables aleatorias no describe completamente el comportamiento de todo el vector.

La función de distribución del vector  $X = (X_1, \dots, X_n)$  es  $F : \mathbb{R}^n \rightarrow [0, 1]$  definida por [31]:

$$F(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n) \quad (1.60)$$

y esta contiene la información acerca del vector aleatorio  $X$ .

Un vector aleatorio continuo admite una función de densidad de probabilidad, esta es:

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(y_1, \dots, y_n) dy_1 \cdots dy_n \quad (1.61)$$

para una función no negativa  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  tal que

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(y_1, \dots, y_n) dy_1 \cdots dy_n = 1 \quad (1.62)$$

Para una función  $g : \mathbb{R}^n \rightarrow \mathbb{R}$ , la esperanza de  $g(X)$  es por definición:

$$E[g(X)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \dots, x_n) f(x_1, \dots, x_n) dx_1 \cdots dx_n \quad (1.63)$$

donde es necesario que  $E[|g(X)|] < \infty$  con el fin de hacer la integral 1.63 significativa.

Esta esperanza definida para vectores aleatorios también cumple las propiedades de linealidad, ser monótona, y la desigualdad del triángulo.

Si se considera el vector aleatorio  $X = (X_1, X_2)$  con función de densidad

$$f_{X_1, X_2}(x_1, x_2)$$

la función de densidad de  $X_1$  se puede obtener integrando sobre  $x_2$ :

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{X_1, X_2}(x_1, x_2) dx_2 \quad (1.64)$$

De hecho,

$$\begin{aligned} P(X_1 \leq a) &= P((X_1, X_2) \in (-\infty, a] \times \mathbb{R}) \\ &= \int_{-\infty}^a \int_{-\infty}^{\infty} f_{X_1, X_2}(x_1, x_2) dx_2 dx_1 \\ &= \int_{-\infty}^a f_{X_1}(x_1) dx_1 \end{aligned} \quad (1.65)$$

Esto se puede extender análogamente para vectores de tamaño  $n$ .

Si  $X_1, \dots, X_n$  son variables aleatorias independientes continuas con función de densidad  $f_1, \dots, f_n$ , respectivamente, entonces

$$\begin{aligned} P(X_1 \leq x_1, \dots, X_n \leq x_n) &= P(X_1 \leq x_1) \cdots P(X_n \leq x_n) \\ &= \left( \int_{-\infty}^{x_1} f_1(y_1) dy_1 \right) \cdots \left( \int_{-\infty}^{x_n} f_n(y_n) dy_n \right) \\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_1(y_1) \cdots f_n(y_n) dy_1 \cdots dy_n \end{aligned} \quad (1.66)$$

La función de distribución del vector  $(X_1, \dots, X_n)$  es por lo tanto el producto de la función de densidad de cada una de las coordenadas del vector aleatorio.

$$f(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n) \quad (1.67)$$

**Ejemplo 1.4.1.** Sean  $X, Y$  variables aleatorias continuas independientes, la probabilidad de que sean iguales es cero, ya que

$$P(X = Y) = E[\mathbb{1}_{\{X=Y\}}] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbb{1}_{\{X=Y\}} f_X(x) f_Y(y) dx dy$$

donde  $\mathbb{1}_{\{X=Y\}} f_X(x) f_Y(y)$  es nulo fuera de la diagonal. Y como la diagonal tiene medida cero, la integral es cero.

**Ejemplo 1.4.2.** Sea  $X_1, X_2$  variables aleatorias independientes con su respectiva función de distribución  $f_1, f_2$ . Se puede calcular

$$\begin{aligned} E[g(X_1, X_2)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x_1, x_2) f_1(x_1) f_2(x_2) dx_1 dx_2 \\ &= \int_{-\infty}^{\infty} f_1(x_1) \left\{ \int_{-\infty}^{\infty} g(x_1, x_2) f_2(x_2) dx_2 \right\} dx_1 \\ &= \int_{-\infty}^{\infty} f_1(x) E[g(x, X_2)] dx \end{aligned}$$

Ahora sea  $X_3$  otra variable aleatoria independiente de las dos anteriores y

$$g(X_1, X_2, X_3) = \mathbb{1}_{\{X_1 \leq X_2\}} \mathbb{1}_{\{X_1 \leq X_3\}} \mathbb{1}_{\{X_1 \leq x\}}$$

se puede observar que

$$\begin{aligned} E[g(y, X_2, X_3)] &= E[\mathbb{1}_{\{y \leq x\}} \mathbb{1}_{\{X_2 \geq y\}} \mathbb{1}_{\{X_3 \geq y\}}] \\ &= \mathbb{1}_{\{y \leq x\}} P(X_2 \geq y, X_3 \geq y) \\ &= \mathbb{1}_{\{y \leq x\}} P(X_2 \geq y) P(X_3 \geq y) \end{aligned}$$

entonces  $E[g(X_1, X_2, X_3)] = \int_{-\infty}^x P(X_2 \geq y) P(X_3 \geq y) f_1(y) dy$ .

**Ejemplo 1.4.3.** Sean  $X, Y$  variables aleatorias independientes con funciones de densidad  $f_X, f_Y$ , y sea  $Z = X + Y$ , la función de densidad de  $Z$  se puede encontrar de la siguiente manera:

$$\begin{aligned} P(Z \leq z) &= P(X + Y \leq z) = E[\mathbb{1}_{\{X+Y \leq z\}}] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbb{1}_{\{x+y \leq z\}} f_X(x) f_Y(y) dx dy \end{aligned}$$

La última integral puede ser escrita, por el teorema de Fubini,

$$\int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} \mathbb{1}_{\{y \leq z-x\}} f_Y(y) dy \right\} f_X(x) dx = \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{z-x} f_Y(y) dy \right\} f_X(x) dx$$

haciendo el cambio de variable  $y = u - x$

$$\int_{-\infty}^z \int_{-\infty}^{\infty} f_Y(u - x) f_X(x) dx du$$

## 1.4.2. Vectores Aleatorios Discretos

Ahora será considerado un vector  $X = (X_1, \dots, X_n)$  donde cada variable aleatoria  $X_i$  en el vector  $X$  toma sus valores en el mismo espacio numerable  $E$ .

La distribución  $(p(u) : u \in E^n)$  de  $X$  se define por [31]:

$$p(u) = P(X_1 = u_1, \dots, X_n = u_n)$$

y la esperanza de  $g(X)$  está definida por

$$E[g(X)] = \sum_{u \in E^n} g(u) p(u) \quad (1.68)$$

con  $g$  no negativa o la suma en 1.68 absolutamente convergente.

Aquí también, la esperanza es monótona, lineal y cumple la desigualdad del triángulo.

**Ejemplo 1.4.4.** Sean  $B_1, \dots, B_k$ ,  $k$  cajas y sean  $n$  pelotas colocadas en las cajas independientemente una de otra, con probabilidad  $p_i$  de que una pelota sea asignada en la caja  $B_i$ . Entonces

$$\sum_{i=1}^k p_i = 1 \quad (1.69)$$

Después de colocar todas las pelotas en las cajas, habrían  $X_i$  pelotas en la caja  $B_i$ , donde

$$\sum_{i=1}^k X_i = n \quad (1.70)$$

La distribución de probabilidad del vector aleatorio  $X = (X_1, \dots, X_n)$  es la *distribución multinomial* de tamaño  $(n, k)$  y parámetros  $p_1, \dots, p_k$

$$P(X_1 = m_1, \dots, X_k = m_k) = \frac{n!}{\prod_{i=1}^k (m_i)!} \prod_{i=1}^k p_i^{m_i} \quad (1.71)$$

donde  $m_1 + \dots + m_k = n$ .

Para probar esto, hay que observar que existen  $\frac{n!}{\prod_{i=1}^k (m_i)!}$  distintas maneras de colocar  $n$  pelotas en  $k$  cajas de tal manera que  $m_1$  pelotas queden en la caja  $B_1$ ,  $m_2$  estén en  $B_2$ , etc., y que cada una de estas formas tienen la misma probabilidad,  $\prod_{i=1}^k p_i^{m_i}$ .

Ahora si se juntan las cajas  $B_2, \dots, B_k$  tal que formen una sola. El proceso completo se convierte en colocar  $n$  pelotas en dos cajas, independientemente una de otra. Para una pelota, su probabilidad de ser puesta en  $B_1$  es  $p_1$ , y en la nueva caja la probabilidad sería  $p_2 + \dots + p_n = 1 - p_1$ . Se obtiene el mismo resultado cuando  $k = 2$ , con  $m_2 = n - m_1$ ,  $p_2 = 1 - p_1$ , y entonces 1.71 se reduce a

$$P(X_1 = m_1, X_2 = m_2) = P(X_1 = m_1) = \frac{n!}{m_1!(n - m_1)!} p_1^{m_1} (1 - p_1)^{n - m_1}$$

## 1.5. Transformaciones de Distribuciones

### 1.5.1. Funciones Generadoras

Sea  $X$  una variable aleatoria con valores en  $\mathbb{N}$ . Su función generadora es la función  $g_X : \overline{D}(0, 1) \rightarrow \mathbb{C}$  definida por [31]:

$$g_X(z) = E[z^X] = \sum_{k=0}^{\infty} P(X = k) z^k \quad (1.72)$$

donde  $\overline{D}(0, 1)$  es el disco cerrado con centro en cero y radio uno.

**Ejemplo 1.5.1.** Para la variable aleatoria binomial de tamaño  $n$  con parámetro  $p$ ,

$$\sum_{k=0}^n P(X = k) z^k = \sum_{k=0}^n \binom{n}{k} (zp)^k (1 - p)^{n - k}$$

y por lo tanto

$$g_X(z) = (1 - p + pz)^n$$

**Ejemplo 1.5.2.** Para la variable aleatoria Poisson de media  $\theta$ ,

$$\sum_{k=0}^{\infty} P(X = k) z^k = e^{-\theta} \sum_{k=0}^{\infty} \frac{(\theta z)^k}{k!}$$

y por lo tanto

$$g_X(z) = e^{\theta(z-1)}$$

**Ejemplo 1.5.3.** Sea  $\{Y_n\}_{n \geq 1}$  una sucesión de variables aleatorias con valores en  $\mathbb{N}$  y con función generadora  $g_Y$  para cada  $Y_i$ . Sea  $T$  otra variable aleatoria, independiente de la sucesión  $\{Y_n\}_{n \geq 1}$ , y sea  $g_T$  su función generadora. Por último, sea  $X$  otra variable aleatoria tal que

$$X = \sum_{n=1}^T Y_n$$



con la convención de que  $\sum_{n=1}^0 = 0$ . Entonces la función generadora de  $X$  puede ser calculada de la siguiente manera,

$$z^X = z^{\sum_{n=1}^T Y_n} = \sum_{k=0}^{\infty} \left\{ \left( z^{\sum_{n=1}^T Y_n} \mathbb{1}_{\{T=k\}} \right) \right\} = \sum_{k=0}^{\infty} \left( z^{\sum_{n=1}^k Y_n} \right) \mathbb{1}_{\{T=k\}}$$

Por lo tanto,

$$E[z^X] = \sum_{k=0}^{\infty} E \left[ \mathbb{1}_{\{T=k\}} \left( z^{\sum_{n=1}^k Y_n} \right) \right] = \sum_{k=0}^{\infty} E \left[ z^{\sum_{n=1}^k Y_n} \right] E \left[ \mathbb{1}_{\{T=k\}} \right]$$

donde se usa la independencia de  $\{Y_n\}_{n \geq 1}$  y  $T$ . Ahora,  $E[\mathbb{1}_{\{T=k\}}] = P(T = k)$ , y  $E[z^{\sum_{n=1}^k Y_n}] = g_Y(z)^k$ , y por lo tanto  $E[z^X] = \sum_{k=0}^{\infty} P(T = k)g_Y(z)^k$ . Esto es, finalmente,

$$g_X(z) = g_T(g_Y(z)) \quad (1.73)$$

Las series de potencia con la sucesión  $\{P(X = n)\}_n \geq 0$  tiene una radio de convergencia  $R \geq 1$ , ya que  $\sum_{n=0}^{\infty} P(X = n) = 1$ . Por lo tanto, el dominio en la definición de  $g_X$  contiene el disco unitario. Dentro de este disco unitario, la diferenciación término por término es posible, por ejemplo,

$$g'_X(z) = \sum_{n=1}^{\infty} nP(X = n)z^{n-1} \quad (1.74)$$

$$g''_X(z) = \sum_{n=2}^{\infty} n(n-1)P(X = n)z^{n-2} \quad (1.75)$$

El lado derecho de 1.74 está bien definido en  $z = 1$ , siendo igual a  $\sum_{n=1}^{\infty} nP(X = n)$ , una cantidad no negativa. Sin embargo el lado izquierdo de 1.74 no siempre está definido, por 1.72 para  $z = 1$ . Entonces se *define*  $g'(1) = \sum_{n=1}^{\infty} nP(X = n)$ , esto es,

$$g'(1) = E[X] \quad (1.76)$$

Por el Teorema de Abel, el límite como variable real  $x$  incrementando a 1 de  $\sum_{n=1}^{\infty} nP(X = n)x^{n-1}$  es  $\sum_{n=1}^{\infty} nP(X = n)$ , y por lo tanto  $g_X$ , como una función sobre el intervalo real  $[0, 1)$ , puede ser extendido a  $[0, 1]$  por 1.76, y esta extensión preserva la continuidad. Esto es usado como sigue: Suponer que  $E[X]$  no es conocida pero se tiene una expresión de  $g'_X(x)$  en  $[0, 1)$  para la cual se puede calcular  $\lim_{x \rightarrow 1} g'_X(x)$  así como  $x$  incrementa a 1. Entonces se sabe que el límite converge en  $E[X]$ .

Similarmente, empezando desde 1.75, y definiendo

$$g''_X(1) = E[X(X-1)] \quad (1.77)$$

se hace a  $g''_X(x)$  una función continua sobre  $[0, 1]$ .

**Ejemplo 1.5.4.** Continuando con el ejemplo anterior 1.5.3. La función generadora de  $X = \sum_{n=1}^T Y_n$  se había visto que era  $g_X(z) = g_T(g_Y(z))$ . Por lo tanto  $g'_X(x) = g'_T(g_Y(x))g'_Y(x)$  para toda  $x \in [0, 1)$ . Haciendo  $x \rightarrow 1$ ,  $g'_X(1) = g'_T(g_Y(1))g'_Y(1) = g'_T(1)g'_Y(1)$ , y por lo tanto  $E[X] = E[T]E[Y]$ .

La función generadora caracteriza a la distribución de una variable aleatoria. Esto significa lo siguiente. Suponer que, sin conocer la distribución de  $X$ , si se puede calcular su función  $g(z)$ , y también su expansión de series de potencia en una vecindad con centro en el origen:

$$g(z) = \sum_{n=0}^{\infty} a_n z^n$$

Como  $g(z)$  es la función generadora de  $Z$ ,

$$g(z) = \sum_{n=0}^{\infty} P(X = n) z^n$$

y ya que la serie de potencia alrededor del origen es única, la distribución de  $X$  se identifica como

$$P(X = n) = a_n$$

para toda  $n \geq 0$ . Similarmente, si dos variables aleatorias con dominio  $\mathbb{N}$  tienen la misma función generadora, entonces tienen la misma distribución.

**Teorema 1.5.1.** ( $\alpha$ ) Sea  $g : [0, 1] \rightarrow \mathbb{R}$  definida como  $g(x) = E[x^X]$ , donde  $X$  es una variable aleatoria con valores en  $\mathbb{N}$ . Entonces  $g$  es no decreciente y convexa. Más aún, si  $P(X = 0) < 1$ , entonces  $g$  es estrictamente creciente, y si  $P(X \leq 1) < 1$ , esta es estrictamente convexa.

( $\beta$ ) Suponiendo que  $P(X \leq 1) < 1$ . Si  $E[X] \leq 1$ , la ecuación  $x = g(x)$  tiene una única solución  $x \in [0, 1]$ , siendo esta  $x = 1$ . Si  $E[X] > 1$ , se tienen dos soluciones en  $[0, 1]$ ,  $x = 1$  y  $x = x_0 \in (0, 1)$ .

**Demostración.** Se observa que para  $x \in [0, 1]$ ,

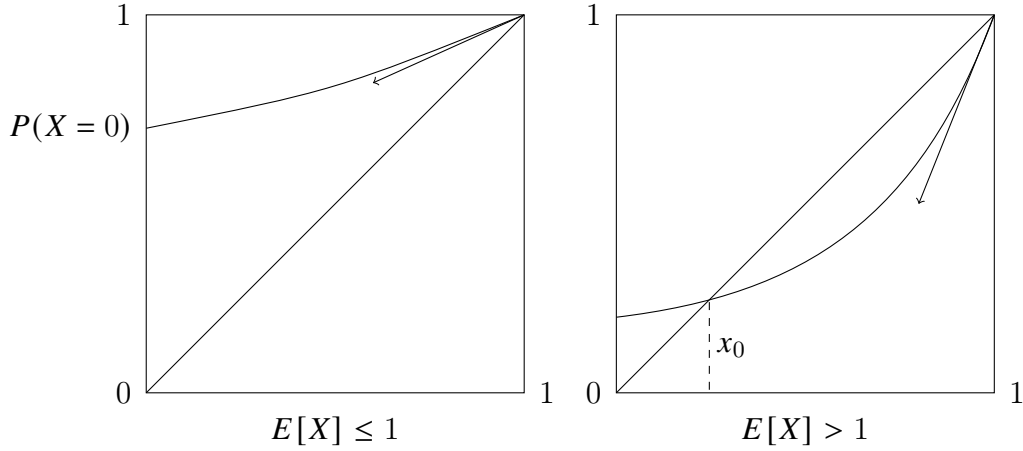
$$g'(x) = \sum_{n=1}^{\infty} nP(X = n)x^{n-1} \geq 0,$$

y por lo tanto  $g$  es no decreciente, y

$$g''(x) = \sum_{n=2}^{\infty} n(n-1)P(X = n)x^{n-2} \geq 0,$$

y por lo tanto  $g$  es convexa. Para que  $g'(x)$  sea nula para alguna  $x \in (0, 1)$ , esta necesita tener  $P(X = n) = 0$  para toda  $n \geq 1$ , y por lo tanto  $P(X = 0) = 1$ . Para que  $g''(x)$  sea nula para alguna  $x \in (0, 1)$ , se debe tener  $P(X = n) = 0$  para toda  $n \geq 2$ , y por lo tanto  $P(X = 0) + P(X = 1) = 1$ .

La gráfica de  $g : [0, 1] \rightarrow \mathbb{R}$  tiene, en el caso  $P(X \leq 1)$ , la siguiente forma, donde se distinguen dos casos:  $E[X] = g'(1) \leq 1$ , y  $E[X] = g'(1) > 1$ .



### 1.5.2. Funciones Características

**Definición 1.5.1.** La función característica  $\phi_X : \mathbb{R}^n \rightarrow \mathbb{C}$  de un vector aleatorio  $X = (X_1, \dots, X_n)$  está definida por [31]:

$$\psi_X(u) = E[e^{iu^t X}] \quad (1.78)$$

**Ejemplo 1.5.5.** Las funciones características de la distribución normal y exponencial son las siguientes:

- Distribución Normal

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}} \iff \psi(u) = e^{i\mu u - \frac{1}{2} \sigma^2 u^2}$$

- Distribución Exponencial

$$f(x) = \lambda e^{-\lambda x} \mathbb{1}_{x>0} \iff \psi(u) = \frac{\lambda}{\lambda - iu}$$

Las funciones características determinan la distribución de una vector aleatorio. Para vectores aleatorios continuos esto es una consecuencia de la teoría clásica de Fourier. De hecho, sean  $f_X$  y  $f_Y$  distribuciones de probabilidad de dos vectores aleatorios  $X$  y  $Y$  sobre  $\mathbb{R}^n$ , si sus funciones características son iguales. Entonces  $f_X - f_Y$  admite la transformación de Fourier  $\psi_X - \psi_Y \equiv 0$ , y por lo tanto, por el teorema de inversión de Fourier,  $\int_C f_X(x) dx = \int_C f_Y(y) dy$  para todo intervalo  $C \in \mathbb{R}^n$ .

La formula de inversión de Fourier da, cuando  $\psi_X(u)$  es integrable,

$$f_X(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} e^{-iu^t x} \psi_X(u) du$$

para la función de densidad de  $X$ .

**Ejemplo 1.5.6.** Para calcular la función característica de una variable aleatoria Cauchy, con la función de densidad

$$f(x) = \frac{1}{\pi} \frac{1}{1+x^2}$$

La transformada de Fourier para  $u \rightarrow e^{-|u|}$  es  $x \rightarrow \int_{-\infty}^{\infty} e^{-|u|} e^{-iux} du = \frac{2}{1+x^2} = g(x)$ . Y como  $g(x)$  es integrable, la formula de inversión de Fourier aplica, y se obtiene

$$e^{-|u|} = \frac{1}{2\pi} \int_{-\infty}^{\infty} g(x) e^{iux} dx = \int_{-\infty}^{\infty} \frac{1}{\pi} \frac{1}{1+x^2} e^{iux} dx$$

**Teorema 1.5.2.** Sean  $Y$  y  $Z$  dos vectores aleatorios de dimensiones  $p$  y  $q$  respectivamente, para toda  $v \in \mathbb{R}^p$  y  $w \in \mathbb{R}^q$ , si se cumple que

$$E[e^{i(v^t Y + w^t Z)}] = \psi_1(v) \psi_2(w) \quad (1.79)$$

donde  $\psi_1(v)$  y  $\psi_2(w)$  son las funciones características de los vectores aleatorios  $\tilde{Y}$  y  $\tilde{Z}$ . Entonces  $Y$  y  $Z$  son independientes y tienen la misma distribución que  $\tilde{Y}$  y  $\tilde{Z}$  respectivamente.

**Demostración.** Se define  $X = (Y, Z)$  y  $u = (v, w)$ , entonces 1.79 se lee

$$E[e^{iux^t X}] = \psi(u) = \psi_1(v) \psi_2(w)$$

Si se encuentra un vector  $\hat{X} = (\hat{Y}, \hat{Z})$  tal que  $\psi_{\hat{X}}(u) = \psi(u)$ , entonces, como la función característica caracteriza a la función de distribución,  $X = (Y, Z)$  tiene la misma función de distribución que  $\hat{X} = (\hat{Y}, \hat{Z})$ . Se toma a  $\hat{Y}$  distribuida como  $\tilde{Y}$ ,  $\hat{Z}$  distribuida como  $\tilde{Z}$ , con  $\hat{Y}$  y  $\hat{Z}$  independientes. Entonces, por el teorema de Fubini [35].

$$E[e^{iu^t \hat{X}}] = E[e^{iw^t \hat{Y}} e^{iv^t \hat{Z}}] = E[e^{iw^t \hat{Y}}] E[e^{iv^t \hat{Z}}] = \psi_1(v) \psi_2(w)$$

Por lo tanto,  $(Y, Z)$  tiene la misma distribución que  $(\hat{Y}, \hat{Z})$  y en particular,  $Y$  y  $Z$  son independientes.

**Teorema 1.5.3.** [29] Sea  $A$  un evento y  $X$  una variable aleatoria tal que para toda  $u \in \mathbb{R}$ ,

$$E[\mathbb{1}_A e^{iuX}] = P(A) E[e^{iuX}]$$

Entonces  $A$  y  $X$  son independientes, esto es,  $\mathbb{1}_A$  y  $X$  son independientes.

**Demostración.** Si  $\mathbb{1}_A$  y  $X$  son independientes, por Fubini [35], se puede observar que para toda  $u, v \in \mathbb{R}$ ,

$$E[e^{iuX} e^{iv\mathbb{1}_A}] = E[e^{iuX}] E[e^{iv\mathbb{1}_A}].$$

Ahora como

$$e^{iv\mathbb{1}_A} = 1 - \mathbb{1}_A + \mathbb{1}_A e^{iv} = 1 + \mathbb{1}_A (e^{iv} - 1),$$

por lo tanto

$$E[e^{iuX} e^{iv\mathbb{1}_A}] = E[e^{iuX}] + (e^{iv} - 1) E[\mathbb{1}_A e^{iuX}]$$

y

$$E[e^{iuX}] E[e^{iv\mathbb{1}_A}] = E[e^{iuX}] + (e^{iv} - 1) P(A) E[e^{iuX}]$$

lo que significa que  $\mathbb{1}_A$  y  $X$  son independientes si  $E[\mathbb{1}_A e^{iuX}] = P(A) E[e^{iuX}]$ .

Sea  $A$  un evento con probabilidad positiva, y sea  $P_A$  denotada como la probabilidad  $P$  condicionada por  $A$ , esto es,

$$P_A(\cdot) = P(\cdot|A)$$

**Definición 1.5.2.** Dos variables aleatorias  $X$  y  $Y$  se dice que son condicionalmente independientes dado un evento  $A$  si son independientes respecto a la medida de probabilidad  $P_A$ .

Por el teorema 1.5.3, una condición necesaria y suficiente para esto es que para toda  $u, v \in \mathbb{R}$ ,

$$E_A[e^{iuX} e^{ivY}] = E_A[e^{iuX}]E_A[e^{ivY}]$$

Se observa que para una variable aleatoria integrable no negativa  $Z$ ,

$$P(A)E_A[Z] = E[Z\mathbb{1}_A]$$

El siguiente resultado es inmediato. Y será recordado para futuras referencias.

**Teorema 1.5.4.** Sea  $A$  un evento con probabilidad positiva. Las variables aleatorias  $X$  y  $Y$  son condicionalmente independientes dado el evento  $A$  si y sólo si para toda  $u, v \in \mathbb{R}$ ,

$$P(A)E[e^{iuX} e^{ivY} \mathbb{1}_A] = E[e^{iuX} \mathbb{1}_A]E[e^{ivY} \mathbb{1}_A]$$

## 1.6. Transformaciones de Vectores Aleatorios

### 1.6.1. Cambios de variable

Sea  $X = (X_1, \dots, X_n)$  un vector aleatorio con función de densidad  $f_X$ , y se define el vector aleatorio

$$Y = g(X) \tag{1.80}$$

donde  $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$ . Explícitamente,

$$\begin{cases} Y_1 = g_1(X_1, \dots, X_n), \\ \vdots \\ Y_n = g_n(X_1, \dots, X_n). \end{cases} \tag{1.81}$$

Se asume que la función  $g$  es sobre un conjunto abierto  $U \subset \mathbb{R}^n$  a  $\mathbb{R}^n$ ,

$$g : U \longrightarrow \mathbb{R}^n \tag{1.82}$$

y las funciones coordenadas  $g_i$  son continuamente diferenciables sobre  $U$ . Se denota la matriz Jacobiana de la función  $g$  por

$$J_g(x_1, \dots, x_n) = \left\{ \frac{dg_i}{dx_j}(x_1, \dots, x_n) \right\}_{1 \leq i, j \leq n} \tag{1.83}$$

y sobre  $U$ ,

$$|\det J_g(x_1, \dots, x_n)| > 0 \quad (1.84)$$

Entonces  $V = g(U)$  es un conjunto abierto de  $\mathbb{R}^n$ , y existe una función inversa  $g^{-1} : V \rightarrow \mathbb{R}$  con las mismas propiedades de  $g$ . En particular, sobre  $V$ ,

$$|\det J_{g^{-1}}(y_1, \dots, y_n)| > 0$$

También,

$$J_{g^{-1}}(y) = J_g(g^{-1}(y))^{-1}$$

**Teorema 1.6.1.** Bajo las condiciones establecidas de  $X$ ,  $g$ , y  $U$ , y si además

$$P(X \in U) = 1 \quad (1.85)$$

entonces  $Y$  toma la función de densidad

$$f_Y(y) = f_X(g^{-1}(y)) |\det J_g(g^{-1}(y))|^{-1} \mathbb{1}_V(y) \quad (1.86)$$

**Demostración.** La prueba consiste en observar que para cualquier función acotada  $h : \mathbb{R} \rightarrow \mathbb{R}$ ,

$$E[h(Y)] = \int_{\mathbb{R}^n} h(y) \psi(y) dy \quad (1.87)$$

donde  $\psi$  es la función 1.86. De hecho, tomando  $h(y) = \mathbb{1}_{\{y \leq a\}} = \mathbb{1}_{\{y_1 \leq a_1\}} \cdots \mathbb{1}_{\{y_n \leq a_n\}}$ ,

$$P(Y_1 \leq a_1, \dots, Y_n \leq a_n) = \int_{-\infty}^{a_1} \cdots \int_{-\infty}^{a_n} \psi(y_1, \dots, y_n) dy_1 \cdots dy_n$$

Para probar que 1.87 se cumple para la  $\psi$  adecuada, se utilizará la regla del cambio de variable de cálculo [34]:

$$\int_U u(x) dx = \int_{g(U)} u(g^{-1}(y)) |\det J_{g^{-1}}(y)| dy$$

Usando esta regla

$$\begin{aligned} E[h(Y)] &= E[h(g(X))] \\ &= \int_U h(g(x)) f_X(x) dx \\ &= \int_V h(y) f_X(g^{-1}(y)) |\det J_{g^{-1}}(y)| dy. \end{aligned}$$

**Ejemplo 1.6.1.** Sea  $U = \mathbb{R}^n$  y  $g(x) = Ax + b$ , donde  $A$  es una matriz invertible de  $n \times n$  y  $b \in \mathbb{R}$ . Entonces  $|\det J_{g^{-1}}(y)| = \frac{1}{|\det A|}$ . Por lo tanto, el vector aleatorio

$$Y = AX + B \quad (1.88)$$

tiene como función de densidad a

$$f_Y(y) = f_X(A^{-1}(y - b)) \frac{1}{|\det A|} \quad (1.89)$$

## 1.7. Estadísticos de orden

Sean  $X_1, \dots, X_n$  variables aleatorias independientes con la misma función de densidad  $f_X(x)$ . En 1.4.1 se demostró que la probabilidad de que dos o más variables aleatorias tomen el mismo valor es cero. Entonces se pueden definir sin ambigüedad las variables aleatorias  $Z_1, \dots, Z_n$  obtenidas al reordenar  $X_1, \dots, X_n$  en orden ascendente:

$$\begin{cases} Z_i \in \{X_1, \dots, X_n\}, \\ Z_1 < Z_2 < \dots < Z_n. \end{cases} \quad (1.90)$$

En particular,  $Z_1 = \min(X_1, \dots, X_n)$  y  $Z_n = \max(X_1, \dots, X_n)$ .

**Teorema 1.7.1.** [8] La función de densidad de  $Z = (Z_1, \dots, Z_n)$  es

$$f_Z(z_1, \dots, z_n) = n! \left\{ \prod_{j=1}^n f_X(z_j) \right\} \mathbb{1}_C(z_1, \dots, z_n) \quad (1.91)$$

donde

$$C = \{(z_1, \dots, z_n) \in \mathbb{R}^n : z_1 < z_2 < \dots < z_n\} \quad (1.92)$$

**Demostración.** Sea  $\sigma$  la permutación de  $\{1, \dots, n\}$  que ordena  $X_1, \dots, X_n$  en orden ascendente, i.e.,

$$X_{\sigma(i)} = Z_i$$

Para cualquier conjunto  $A \subset \mathbb{R}^n$ ,

$$P(Z \in A) = P(Z \in A \cap C) = P(X_\sigma \in A \cap C) = \sum_{\sigma_0} P(X_{\sigma_0} \in A \cap C, \sigma = \sigma_0)$$

donde la suma es sobre todas las permutaciones de  $\{1, \dots, n\}$ . El hecho de que  $X_{\sigma_0} \in A \cap C$  implica que  $\sigma = \sigma_0$ ,

$$P(X_{\sigma_0} \in A \cap C, \sigma = \sigma_0) = P(X_{\sigma_0} \in A \cap C)$$

y como la distribución de probabilidad de  $X_{\sigma_0}$  no depende de una permutación  $\sigma_0$  por la independencia de cada  $X_i$ ,

$$P(X_{\sigma_0} \in A \cap C) = P(X \in A \cap C)$$

Por lo tanto,

$$\begin{aligned} P(Z \in A) &= \sum_{\sigma_0} P(X \in A \cap C) = n! P(X \in A \cap C) \\ &= n! \int_{A \cap C} \prod_{j=1}^n f_X(x_j) dx_1 \cdots dx_n \\ &= \int_A n! \prod_{j=1}^n f_X(x_j) \mathbb{1}_C(x_1, \dots, x_n) dx_1 \cdots dx_n \end{aligned}$$

**Ejemplo 1.7.1.** Aplicando la formula 1.91 a la situación donde las  $X_i$  se distribuyen uniformemente sobre el intervalo cerrado  $[a, b]$  da como resultado

$$f_Z(z_1, \dots, z_n) = \frac{n!}{(b-a)^n} \mathbb{1}_{[a,b]^n}(z_1, \dots, z_n) \mathbb{1}_C(z_1, \dots, z_n) \quad (1.93)$$

En particular, ya que  $\int_{\mathbb{R}^n} f_Z(z) dz = 1$ ,

$$\int_a^b \cdots \int_a^b \mathbb{1}_C(z_1, \dots, z_n) dz_1 \cdots dz_n = \frac{(b-a)^n}{n!} \quad (1.94)$$

## 1.8. Esperanza Condicional de Variables Discretas

### 1.8.1. Definición y Propiedades Básicas

El objetivo de esta sección es dar una introducción a la noción de la esperanza condicional para variables aleatorias discretas y se obtienen resultados que son necesarios en el estudio de las cadenas de Markov con espacio de estados numerable.

**Definición 1.8.1.** Sea  $X$  y  $Y$  dos variables aleatorias discretas que toman valores en los conjuntos numerables  $F$  y  $G$ , respectivamente, y sea  $g : F \times G \rightarrow \mathbb{R}$  una función no negativa. Se define la función  $\psi : G \rightarrow \mathbb{R}$  por

$$\psi(y) = \sum_{x \in F} g(x, y) P(X = x | Y = y). \quad (1.95)$$

Para cada  $y \in G$ ,  $\psi(y)$  es llamada esperanza condicional de  $g(X, Y)$  dado  $Y = y$ , y es denotada por  $E^{Y=y}[g(X, Y)]$ , o  $E[g(X, Y) | Y = y]$ :

$$E^{Y=y}[g(X, Y)] = \psi(y). \quad (1.96)$$

La variable aleatoria  $\psi(Y)$  es la esperanza condicional de  $g(X, Y)$  dado  $Y$ , y es denotada por  $E^Y[g(X, Y)]$  o  $E[g(X, Y) | Y]$ :

$$E^Y[g(X, Y)] = \psi(Y) \quad (1.97)$$

Se puede notar que si  $g \geq 0$  y  $E[g(X, Y)] < \infty$ , entonces  $\psi(y) < \infty$  para toda  $y \in G$  tal que  $P(Y = y) > 0$ ; que equivale a decir,  $\psi(Y) < \infty$  casi seguramente. Esto se sigue de

$$\sum_{y \in G} \psi(y) P(Y = y) = E[g(X, Y)] < \infty.$$

Por lo tanto, si  $g : F \times G \rightarrow \mathbb{R}$  es una función arbitraria tal que  $E[|g(X, Y)|] < \infty$ , se puede definir  $E^{Y=y}[g^+(X, Y)]$  y  $E^{Y=y}[g^-(X, Y)]$ . Como cada uno de los términos son finitos se puede definir entonces a

$$E^{Y=y}[g(X, Y)] = \sum_{x \in F} g(x, y) P(X = x | Y = y)$$

$$E^Y[g(X, Y)] = \sum_{x \in F} g(x, y) P(X = x | Y)$$

para el caso general.

Finalmente para un evento  $A$ , se tiene que  $P(A|Y) = P^Y(A)$ , la probabilidad de  $A$  dado  $Y$ , sería  $E^Y[\mathbb{1}_A]$ .



**Ejemplo 1.8.1.** Sea  $X_1$  y  $X_2$  dos variables aleatorias Poisson independientes con sus respectivos parámetros  $\theta_1 > 0$  y  $\theta_2 > 0$ . Con la definición dada se puede calcular  $E^Y[X]$  donde  $X = X_1$ ,  $Y = X_1 + X_2$ . Primero se tiene que calcular

$$\begin{aligned} P(X = x|Y = y) &= \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{P(X_1 = x, X_1 + X_2 = y)}{P(X_1 + X_2 = y)} \\ &= \binom{y}{x} \left( \frac{\theta_1}{\theta_1 + \theta_2} \right)^x \left( \frac{\theta_2}{\theta_1 + \theta_2} \right)^{y-x} \mathbb{1}_{\{y \geq x\}} \end{aligned}$$

Haciendo  $\alpha = \frac{\theta_1}{\theta_1 + \theta_2}$  y usando el procedimiento de 1.3.4,

$$\psi(y) = E^{Y=y}[X] = \sum_{x=0}^y x \binom{y}{x} \alpha^x (1 - \alpha)^{y-x} = \alpha y.$$

Finalmente,  $E^Y[X] = \psi(Y) = \alpha Y$ , esto es,

$$E^{X_1+X_2}[X_1] = \frac{\theta_1}{\theta_1 + \theta_2} (X_1 + X_2).$$

La linealidad es una consecuencia de la definición:

Para toda  $\lambda_1, \lambda_2 \in \mathbb{R}$ ,

$$E^Y[\lambda_1 g_1(X, Y) + \lambda_2 g_2(X, Y)] = \lambda_1 E^Y[g_1(X, Y)] + \lambda_2 E^Y[g_2(X, Y)] \quad (1.98)$$

siempre y cuando la esperanza condicional esté bien definida.

La esperanza condicional es monótona y también es consecuencia de la definición: si  $g_1(x, y) \leq g_2(x, y)$ , entonces

$$E^Y[g_1(X, Y)] \leq E^Y[g_2(X, Y)] \quad (1.99)$$

Ahora si se quiere calcular

$$\begin{aligned} E[E^Y[g(X, Y)]] &= \sum_{y \in G} \psi(y) P(Y = y) \\ &= \sum_{y \in G} \sum_{x \in F} g(x, y) P(X = x|Y = y) P(Y = y) \\ &= \sum_x \sum_y g(x, y) P(X = x, Y = y) \\ &= E[g(X, Y)] \end{aligned} \quad (1.100)$$

También,

$$E^Y[w(Y)] = w(Y), \quad (1.101)$$

y de manera general,

$$E^Y[w(Y)h(X, Y)] = w(Y)E^Y[h(X, Y)]. \quad (1.102)$$

siempre y cuando el lado derecho esté bien definido.

Si  $X$  y  $Y$  son independientes y si  $v : F \rightarrow \mathbb{R}$  es tal que  $E[|v(X)|] < \infty$ , entonces

$$E^Y[v(X)] = E[v(X)]. \quad (1.103)$$

De hecho,  $E^{Y=y}[v(X)] = \sum_{x \in F} v(x)P(X = x|Y = y) = \sum_{x \in F} v(x)P(X = x)$ .

En la definición de esperanza condicional, se asume solamente que  $G$  es un conjunto numerable. Se puede tomar a  $G$  como el producto de la forma  $G = G_1 \times G_2$ , es decir,  $Y = (Y_1, Y_2)$ , donde  $Y_1$  y  $Y_2$  toman sus valores en los conjuntos numerables  $G_1$  y  $G_2$ , respectivamente. En esta situación, se usará la siguiente notación

$$E^Y[g(X, Y)] = E^{Y_1, Y_2}[g(X, Y_1, Y_2)]. \quad (1.104)$$

**Teorema 1.8.1.** Bajo las siguientes condiciones:

- $g(X, Y) \geq 0$ , casi seguramente,
- $E[|g(X, Y)|] < \infty$ ,

se tiene que

$$E^{Y_2}[E^{Y_1, Y_2}[g(X, Y_1, Y_2)]] = E^{Y_2}[g(X, Y_1, Y_2)]. \quad (1.105)$$

**Demostración.** Sea

$$\psi(Y_1, Y_2) = E^{Y_1, Y_2}[g(X, Y_1, Y_2)].$$

Se debe mostrar que

$$E^{Y_2}[\psi(Y_1, Y_2)] = E^{Y_2}[g(X, Y_1, Y_2)]$$

Como

$$\psi(y_1, y_2) = \sum_x g(x, y_1, y_2)P(X = x|Y_1 = y_1, Y_2 = y_2)$$

y entonces

$$\begin{aligned} E^{Y_2=y_2}[\psi(Y_1, Y_2)] &= \sum_{y_1} \psi(y_1, y_2)P(Y_1 = y_1|Y_2 = y_2) \\ &= \sum_{y_1} \sum_x g(x, y_1, y_2)P(X = x|Y_1 = y_1, Y_2 = y_2)P(Y_1 = y_1|Y_2 = y_2) \\ &= \sum_{y_1} \sum_x g(x, y_1, y_2)P(X = x, Y_1 = y_1|Y_2 = y_2) \\ &= E^{Y_2=y_2}[g(X, Y_1, Y_2)] \end{aligned}$$

**Ejemplo 1.8.2.** Sea  $X_1$  y  $X_2$  dos variables aleatorias independientes idénticamente distribuidas (v.a.i.i.d.) con valores en el conjunto numerable  $E$ , con  $E[|X_1|] < \infty$ . Entonces

$$E^{X_1+X_2}[X_1] = \psi(X_1 + X_2)$$

y por simetría

$$E^{X_1+X_2}[X_2] = \psi(X_1 + X_2).$$

Por lo tanto

$$2\psi(X_1 + X_2) = E^{X_1+X_2}[X_1] + E^{X_1+X_2}[X_2] = E^{X_1+X_2}[X_1 + X_2] = X_1 + X_2$$

y finalmente

$$E^{X_1+X_2}[X_1] = E^{X_1+X_2}[X_2] = \frac{X_1 + X_2}{2}$$

## 1.9. Ley Fuerte de los Grandes Números

### 1.9.1. Lema de Borel-Cantelli

Se considera una sucesión de eventos  $\{A_n\}_{n \geq 1}$  donde el índice  $n$  puede ser, si uno lo desea, interpretado como el tiempo. Se define el siguiente evento:

$$\{\omega : \omega \in A_n \text{ para un número infinito de índices } n\} = \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k. \quad (1.106)$$

denotado por  $\{A_n \text{ i.s.}\}$ , donde i.s. significa infinitamente seguido, que significa que  $\omega \in A_n$  para toda  $n \geq 1$  en al menos uno de los conjuntos  $A_n, A_{n+1}, \dots$ .

**Lema 1.9.1.** Sea  $\{A_n\}_{n \geq 1}$  una sucesión de eventos tal que

$$\sum_{n=1}^{\infty} P(A_n) < \infty \quad (1.107)$$

Entonces

$$P\left(\bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k\right) = 0 \quad (1.108)$$

**Demostración.** Sea  $B_n = \bigcup_{k \geq n} A_k$ , donde  $B_n$  decrece cuando  $n$  crece, entonces por el teorema 1.1.1

$$P\left(\bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k\right) = \lim_{n \uparrow \infty} P\left(\bigcup_{k \geq n} A_k\right)$$

Pero por sub- $\sigma$ -aditividad 1.4,

$$P\left(\bigcup_{k \geq n} A_k\right) \leq \sum_{k \geq n} P(A_k),$$

y por hipótesis 1.107, el lado derecho de la desigualdad tiende a cero cuando  $n$  tiende a  $\infty$ .

### 1.9.2. Convergencia Casi Segura

**Definición 1.9.1.** [31] Una sucesión  $\{Z_n\}_{n \geq 1}$  de variables aleatorias se dice que converge  $P$ -casi seguramente ( $P$ -c.s.) a la variable aleatoria  $Z$  si

$$P\left(\lim_{n \uparrow \infty} Z_n = Z\right) = 1 \quad (1.109)$$

**Ejemplo 1.9.1.** Sea  $\{X_n\}_{n \geq 1}$  una familia de variables aleatorias con valores en  $\{0, 1\}$  tal que

$$\sum_{n=1}^{\infty} P(X_n = 1) < \infty \quad (1.110)$$

Entonces, en vista del Lema de Borel-Cantelli,  $P(\bigcap_{n=1}^{\infty} \bigcup_{k \geq n} \{X_k = 1\}) = 0$ . Pero el evento complementario  $\{\bigcap_{n=1}^{\infty} \bigcup_{k \geq n} \{X_k = 1\}\}^c = \bigcup_{n=1}^{\infty} \bigcap_{k \geq n} \{X_k = 1\}^c$  es

$$\bigcup_{n=1}^{\infty} \bigcap_{k \geq n} \{X_k = 0\} = A$$

lo que implica que  $\{X_n\}_{n \geq 1}$  converge  $P$ -c.s. a 0, ya que  $P(A) = 1$ .

Sea  $\{\varepsilon_n\}_{n \geq 1}$  una sucesión de números positivos que convergen a 0. Si para una  $\omega$  dada,  $|Z_n(\omega) - Z(\omega)| < \varepsilon_n$ , para todo pero un número finito de índices  $n$ , entonces  $\lim_{n \uparrow \infty} Z_n(\omega) = Z(\omega)$ . Por lo tanto, por el Lema de Borel-Cantelli, se tiene el siguiente teorema.

**Teorema 1.9.1.** Sea  $\{Z_n\}_{n \geq 1}$  y  $Z$  variables aleatorias. Si

$$\sum_{n \geq 1} P(|Z_n - Z| \geq \varepsilon_n) < \infty \quad (1.111)$$

para alguna sucesión de números positivos  $\{\varepsilon_n\}_{n \geq 1}$  que convergen a 0, entonces la sucesión  $\{Z_n\}_{n \geq 1}$  converge  $P$ -c.s. a  $Z$ ,

El siguiente resultado es similar, pero en esta ocasión se tiene una condición necesaria y suficiente.

**Teorema 1.9.2.** La sucesión  $\{Z_n\}_{n \geq 1}$  de variables aleatorias con valores en  $\mathbb{R}$  converge  $P$ -c.s. a la variable aleatoria  $Z$  si y sólo si para toda  $\varepsilon > 0$ ,

$$P\left(\bigcap_{n=1}^{\infty} \bigcup_{k \geq n} \{|Z_k - Z| \geq \varepsilon\}\right) = P(|Z_n - Z| \geq \varepsilon \text{ i.s.}) = 0 \quad (1.112)$$

**Demostración.** Se puede observar que

$$\{|Z_n - Z| \geq \varepsilon \text{ i.s.}\} \subset \{\omega : \lim_{n \uparrow \infty} Z_n(\omega) \neq Z(\omega)\}^c$$

y por lo tanto

$$P(|Z_n - Z| \geq \varepsilon \text{ i.s.}) \leq 1 - P(\lim_{n \uparrow \infty} Z_n = Z) = 0$$

Ahora, sea  $N_k$  el último índice  $n$  tal que  $|Z_n - Z| \geq \frac{1}{k}$  (se hace  $N_k = \infty$  si  $|Z_n - Z| \geq \frac{1}{k}$  para toda  $n \geq 1$ ). Por 1.112 con  $\varepsilon = \frac{1}{k}$ , se tiene que  $P(N_k < \infty) = 0$ . Por la propiedad de sub- $\sigma$ -aditividad 1.4,  $P(\bigcup_{k \geq 1} \{N_k < \infty\}) = 0$ . Equivalentemente,  $P(N_k = \infty, \text{ para toda } k \geq 1) = 1$ , lo cual implica que  $P(\lim_{n \uparrow \infty} Z_n = Z) = 1$ .

Una ley de números grandes es un enunciado acerca de la convergencia casi segura del promedio empírico

$$\frac{S_n}{n} = \frac{X_1 + \cdots + X_n}{n}$$

donde  $\{X_n\}_{n \geq 1}$  son v.a.i.i.d.

**Teorema 1.9.3.** Sea  $\{X_n\}_{n \geq 1}$  v.a.i.i.d. tal que

$$E[|X_1|] < \infty. \quad (1.113)$$

Entonces,  $P$ -c.s.

$$\lim_{n \uparrow \infty} \frac{S_n}{n} = E[X_1] \quad (1.114)$$

En otras palabras, existe un conjunto  $N$  de probabilidad 0, tal que si  $\omega \notin N$ , entonces

$$\lim_{n \uparrow \infty} \frac{S_n(\omega)}{n} = E[X_1]$$

El promedio empírico es asintóticamente igual al promedio probabilístico. Tal es el contenido «físico» de la ley fuerte de los grandes números de Kolmogorov (1993) [30]. Emile Borel probó la ley fuerte de los grandes números (LFGN) en 1909 en el caso especial donde  $X_n = 0$  ó  $1$ , con  $P(X_n = 1) = p$ , esto mostrando en particular que la fracción promedio de caras en un juego justo de lanzar monedas debería tender a  $\frac{1}{2}$  tanto como el número de lanzamientos incrementemente indefinidamente.

Existen numerosas versiones de la LFGN extendiendo el resultado de Kolmogorov a situaciones donde la independencia de la sucesión  $\{X_n\}_{n \geq 1}$  no es requerida. Como un hecho importante, el teorema ergódico para cadenas de Markov positivas irreducibles, es una de sus extensiones.

Se mostrará una prueba para la LFGN a continuación. Antes de eso se deberá probar la LFGN de Borel. Esto basado en el Lema de Borel-Cantelli y la desigualdad de Markov.

### 1.9.3. Desigualdad de Markov

**Teorema 1.9.4.** Sea  $X$  una variable aleatoria con valores en  $\mathbb{R}$ ,  $f : \mathbb{R} \rightarrow \mathbb{R}_+$ ,  $a > 0$ . Se tiene que

$$P(f(X) \geq a) \leq \frac{E[f(X)]}{a}. \quad (1.115)$$

**Demostración.** De la desigualdad

$$f(X) \geq a \mathbb{1}_{\{f(X) \geq a\}}$$

al tomar las esperanzas se sigue que

$$E[f(X)] \geq aE[\mathbb{1}_{\{f(X) \geq a\}}] = aP(f(X) \geq a).$$

Especialmente la desigualdad de Markov aplicada a  $f(x) = (x - \mu)^2$ ,  $a = \varepsilon > 0$ , se obtiene la *desigualdad de Chebyshev*.

$$P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2} \quad (1.116)$$

**Ejemplo 1.9.2** (LFGN de Borel). Esta es la LFGN de Kolmogorov cuando las  $X_n$  están acotadas. Se deberá acotar la probabilidad de que  $|\frac{S_n}{n} - m|$  exceda alguna  $\varepsilon > 0$  donde  $\mu = E[X_1]$ . Para esto se aplica la desigualdad de Markov

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) = P\left(\left(\frac{S_n}{n} - \mu\right)^4 \geq \varepsilon^4\right) \leq \frac{E\left[\left(\frac{S_n}{n} - \mu\right)^4\right]}{\varepsilon^4}.$$

Ahora,

$$\left(\frac{S_n}{n} - \mu\right)^4 = \frac{(\sum_{i=1}^n (X_i - \mu))^4}{n^4},$$

y por lo tanto

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \leq \frac{E[(\sum_{i=1}^n (X_i - \mu))^4]}{n^4 \varepsilon^4}.$$

Si se puede probar que

$$E\left[\left(\sum_{i=1}^n (X_i - \mu)\right)^4\right] \leq Kn^2 \tag{1.117}$$

para alguna  $K \in \mathbb{N}$ , entonces

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \leq \frac{K}{n^2 \varepsilon^4}$$

y en particular, con  $\varepsilon = n^{-\frac{1}{8}}$ ,

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq n^{-\frac{1}{8}}\right) \leq \frac{K}{n^{\frac{3}{2}}},$$

de lo cual se sigue que

$$\sum_{n=1}^{\infty} P\left(\left|\frac{S_n}{n} - \mu\right| \geq n^{-\frac{1}{8}}\right) < \infty$$

Por lo tanto, por el teorema 1.9.1,  $|\frac{S_n}{n} - \mu|$  converge  $P$ -c.s. a 0.

Ahora para probar 1.117, sea  $Y_i = X_i - \mu$ , recordando que  $E[Y_i] = 0$ . También, en vista de la independencia que se supuso,  $E[Y_1 Y_2 Y_3 Y_4] = E[Y_1]E[Y_2]E[Y_3]E[Y_4] = 0$ ,  $E[Y_1 Y_2^3] = E[Y_1]E[Y_2^3] = 0$ , y similares. Finalmente, en el siguiente desarrollo

$$E\left[\left(\sum_{i=1}^n Y_i\right)^4\right] = \sum_{i,j,k,l=1}^n E[Y_i Y_j Y_k Y_l],$$

solo los términos de la forma  $E[Y_i^4]$  y  $E[Y_i^2 Y_j^2]$  ( $i \neq j$ ) permanecen. Existen  $n$  términos del primer tipo y  $3n(n-1)$  términos del segundo tipo. Por lo tanto,

$$E\left[\left(\sum_{i=1}^n Y_i\right)^4\right] = nE[Y_1^4] + 3n(n-1)E[Y_1^2 Y_2^2],$$

que es más pequeño que  $Kn^2$  para alguna  $K \in \mathbb{N}$ .

**Ejemplo 1.9.3.** Sea  $X$  una variable aleatoria con valores en  $\mathbb{R}$  tal que para toda  $t \in \mathbb{R}$ ,  $E[e^{tX}] < \infty$ . Sea  $\psi(t) = \log E[e^{tX}]$ . Se observa que para  $t \geq 0$ ,  $a \in \mathbb{R}$ ,  $X > a$  es equivalente a  $e^{tX} > e^{ta}$ , y usando la desigualdad de Markov, se encuentra que

$$P(X > a) = P(e^{tX} > e^{ta}) \leq \frac{E[e^{tX}]}{e^{ta}} = \exp\{-(at - \psi(t))\}.$$

Esto es cierto para toda  $t \geq 0$ , y se sigue que

$$P(X > a) \leq e^{-h(a)}, \quad (1.118)$$

donde

$$h(a) = \sup_{t \geq 0} \{at - \psi(t)\}. \quad (1.119)$$

#### 1.9.4. Prueba de la LFGN de Kolmogorov

Se asumirá, sin pérdida de generalidad, que  $E[X_i] = 0$ .

**Lema 1.9.2.** Sea  $X_1, \dots, X_n$  variables aleatorias independientes tal que para toda  $i \in [1, n]$ ,

$$E[|X_i|^2] < \infty, \quad E[X_i] = 0. \quad (1.120)$$

Entonces para toda  $\lambda > 0$ ,

$$P\left(\max_{1 \leq k \leq n} |S_k| \geq \lambda\right) \leq \frac{E[S_n^2]}{\lambda^2}, \quad (1.121)$$

donde  $S_k = X_1 + \dots + X_k$ .

**Demostración.** Sea  $T$  una variable aleatoria que representa el primer índice  $k \in [1, n]$ , tal que  $|S_k| \geq \lambda$ , con  $T = \infty$  si es que  $\max_{1 \leq k \leq n} |S_k| < \lambda$ . Para  $k \leq n$ ,

$$\begin{aligned} E[S_n^2 \mathbb{1}_{\{T=k\}}] &= E[\mathbb{1}_{\{T=k\}} \{(S_n - S_k)^2 + 2S_k(S_n - S_k) + S_k^2\}] \\ &= E[\mathbb{1}_{\{T=k\}} \{(S_n - S_k)^2 + S_k^2\}] \geq E[\mathbb{1}_{\{T=k\}} S_k^2], \end{aligned}$$

donde se usó el hecho de que  $\mathbb{1}_{\{T=k\}} S_k$  es una función de  $X_1, \dots, X_k$  y por lo tanto es independiente de  $S_n - S_k$ , entonces  $E[\mathbb{1}_{\{T=k\}} S_k (S_n - S_k)] = E[\mathbb{1}_{\{T=k\}} S_k] E[S_n - S_k] = 0$ . Por lo tanto,

$$\begin{aligned} E[|S_n|^2] &\geq \sum_{k=1}^n E[\mathbb{1}_{\{T=k\}} S_k^2] \geq \sum_{k=1}^n E[\mathbb{1}_{\{T=k\}} \lambda^2] \\ &= \lambda^2 \sum_{k=1}^n P(T = k) = \lambda^2 P(T \leq n) \\ &= \lambda^2 P\left(\max_{1 \leq k \leq n} |S_k| \geq \lambda\right). \end{aligned}$$

El siguiente corolario contiene una demostración de la LFGN cuando  $E[|X_n|^2] < \infty$ .

**Corolario 1.9.4.1.** Sea  $\{X_n\}_{n \geq 1}$  una sucesión de variables aleatorias independientes tal que para toda  $n \geq 1$ ,

$$E[|X_n|^2] < \infty, \quad E[X_n] = 0. \quad (1.122)$$

Si

$$\sum_{n \geq 1} \frac{E[X_n^2]}{n^2} \leq \infty, \quad (1.123)$$

entonces

$$\lim_{n \uparrow \infty} \frac{1}{n} \sum_{k=1}^n X_k = 0, \quad P\text{-c.s.} \quad (1.124)$$

**Demostración.** Si  $2^{k-1} \leq n \leq 2^k$ , entonces  $\frac{|S_n|}{n} \geq \varepsilon$  implica que  $\frac{|S_n|}{2^{k-1}} \geq \varepsilon$ . Por lo tanto, para toda  $\varepsilon > 0$ , y para toda  $k \geq 1$ ,

$$\begin{aligned} & P\left(\frac{|S_n|}{n} \geq \varepsilon \text{ para alguna } n \in [2^{k-1}, 2^k]\right) \\ & \leq P(|S_n| \geq \varepsilon 2^{k-1} \text{ para alguna } n \in [2^{k-1}, 2^k]) \\ & \leq P(|S_n| \geq \varepsilon 2^{k-1} \text{ para alguna } n \in [1, 2^k]) \\ & = P\left(\max_{1 \leq n \leq 2^k} |S_n| \geq \varepsilon 2^{k-1}\right) \\ & \leq \frac{4}{\varepsilon^2} \frac{1}{(2^k)^2} \sum_{n=1}^{2^k} E[X_n^2], \end{aligned}$$

donde la última desigualdad se sigue de la desigualdad de Kolmogorov [7]. Definiendo  $m = m(n)$  por  $2^{m-1} \leq n < 2^m$ , se tiene que

$$\sum_{k=1}^{\infty} \frac{1}{(2^k)^2} \sum_{n=1}^{2^k} E[X_n^2] = \sum_{n=1}^{\infty} E[X_n^2] \sum_{j=m}^{\infty} \frac{1}{(2^j)^2},$$

que está acotado por

$$\sum_{n=1}^{\infty} E[X_n^2] \frac{K}{(2^m)^2} \geq \sum_{n=1}^{\infty} E[X_n^2] \frac{K}{n^2}$$

para alguna  $K \in \mathbb{N}$ . Por lo tanto por 1.123,

$$\sum_{k=1}^{\infty} P\left(\frac{|S_n|}{n} \geq \varepsilon \text{ para alguna } n : 2^{k-1} < n \leq 2^k\right) < \infty$$

y por el lema de Borel-Cantelli,

$$P\left(\frac{|S_n|}{n} \geq \varepsilon \text{ i.s.}\right) = 0$$

así, el resultado se sigue del teorema 1.9.2.



Habiendo probado el corolario, ahora solo queda deshacerse de la hipótesis de que  $E[|X_n|^2] < \infty$ , y para hacerlo se define

$$\tilde{X}_n = \begin{cases} X_n & \text{si } |X_n| \leq n, \\ 0 & \text{en otro caso.} \end{cases}$$

Ya que  $E[|X_1|] < \infty$ , se tiene que por la convergencia dominada [31]

$$\lim_{n \uparrow \infty} E[\tilde{X}_n \mathbb{1}_{\{|X_1| \geq n\}}] = E[X_1] = 0.$$

Y como  $X_n$  tiene la misma distribución que  $X_1$ ,

$$\lim_{n \uparrow \infty} E[\tilde{X}_n] = \lim_{n \uparrow \infty} E[\tilde{X}_n \mathbb{1}_{\{|X_1| \geq n\}}] = E[X_1] = 0.$$

En particular se cumple el *Lema de Cesaro* [37],

$$\lim_{n \uparrow \infty} \frac{1}{n} \sum_{k=1}^n E[\tilde{X}_k] = 0.$$

También,

$$\begin{aligned} E[|X_1|] &= \sum_{n=0}^{\infty} E[|X_1| \mathbb{1}_{(n, n+1]}(|X_1|)] \geq \sum_{n=0}^{\infty} E[n \mathbb{1}_{(n, n+1]}(|X_1|)] \\ &= \sum_{n=0}^{\infty} n(P(|X_1| > n) - P(|X_1| > n+1)) = \sum_{n=1}^{\infty} P(|X_1| > n). \end{aligned}$$

En particular,

$$\sum_{n=1}^{\infty} P(|X_n| > n) = \sum_{n=1}^{\infty} P(|X - 1| > n) \leq E[|X_1|] < \infty$$

y por lo tanto por el lema de Borel-Cantelli,

$$P(\tilde{X}_n \neq X_n \text{ i.s.}) = P(|X_n| > n \text{ i.s.}) = 0.$$

Por lo tanto, para probar la LFGN es suficiente mostrar que

$$\lim_{n \uparrow \infty} \frac{1}{n} \sum_{k=1}^n (\tilde{X}_k - E[\tilde{X}_k]) = 0.$$

En vista del último corolario, es suficiente probar que

$$\sum_{n=1}^{\infty} \frac{E[(\tilde{X}_n - E[\tilde{X}_n])^2]}{n^2} < \infty.$$

Pero

$$E[(\tilde{X}_n - E[\tilde{X}_n])^2] \leq E[\tilde{X}_n^2] = E[X_1^2 \mathbb{1}_{\{|X_1| \leq n\}}].$$

Es suficiente mostrar que

$$\sum_{n=1}^{\infty} \frac{E[X_1^2 \mathbb{1}_{\{|X_1| \leq n\}}]}{n^2} < \infty.$$

El lado izquierdo de la desigualdad es igual a

$$\sum_{n=1}^{\infty} \frac{1}{n^2} \sum_{k=1}^n E[X_1^2 \mathbb{1}_{\{k-1 < |X_1| \leq k\}}] = \sum_{n=1}^{\infty} \left( \sum_{n=k}^{\infty} \frac{1}{n^2} \right) E[X_1^2 \mathbb{1}_{\{k-1 < |X_1| \leq k\}}]. \quad (1.125)$$

Usando el hecho de que

$$\sum_{n=k}^{\infty} \frac{1}{n^2} \leq \frac{1}{k^2} + \int_k^{\infty} \frac{1}{x^2} dx = \frac{1}{k^2} + \frac{1}{k} \leq \frac{2}{k}$$

1.125 es menor o igual que

$$\begin{aligned} \sum_{k=1}^{\infty} \frac{2}{k} E[X_1^2 \mathbb{1}_{\{k-1 < |X_1| \leq k\}}] &= 2 \sum_{k=1}^{\infty} E\left[\frac{X_1^2}{k} \mathbb{1}_{\{k-1 < |X_1| \leq k\}}\right] \\ &\leq 2 \sum_{k=1}^{\infty} E[|X_1| \mathbb{1}_{\{k-1 < |X_1| \leq k\}}] \\ &= 2E[|X_1|] < \infty. \end{aligned}$$

y así se concluye con la demostración.

---

---

# CAPÍTULO 2

---

## MODELOS DE MARKOV A TIEMPO DISCRETO

### 2.1. La Matriz de Transición

#### 2.1.1. Propiedad de Markov

Las sucesiones de variables aleatorias independientes e idénticamente distribuidas son procesos estocásticos, pero estos no son siempre interesantes vistos como modelos estocásticos. Con el objetivo de introducir más variabilidad, se puede permitir alguna dependencia del pasado, de manera que se determinen ecuaciones de recurrencia. Las cadenas de Markov homogéneas poseen la característica requerida, ya que siempre pueden ser representadas al menos como una ecuación recurrente estocástica  $X_{n+1} = f(X_n, Z_{n+1})$ , donde  $\{Z_n\}_{n \geq 1}$  es una sucesión de v.a.i.i.d., independiente del estado inicial  $X_0$ .

La dependencia del pasado es solo a través del estados anterior, pero esta limitada cantidad de memoria es suficiente para producir una gran diversidad de comportamientos. Por esta razón las cadenas de Markov han encontrado aplicaciones en varias ramas de la ciencia, incluyendo biología, física, sociología, operaciones, e ingeniería, donde proveen respuestas tanto cualitativas como cuantitativas.

Una sucesión  $\{X_n\}_{n \geq 0}$  de variables aleatorias con valores en  $E$  es conocida como *proceso estocástico a tiempo discreto con espacio de estados  $E$* . Donde  $E$  es numerable.

**Definición 2.1.1.** Sea  $\{X_n\}_{n \geq 0}$  un proceso estocástico a tiempo discreto con espacio de estados numerable  $E$ . Si para todo entero  $n \geq 0$  y todos sus estados  $i_0, i_1, \dots, i_{n-1}, i, j$

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P(X_{n+1} = j | X_n = i) \quad (2.1)$$

donde ambos lados están bien definidos, a este proceso estocástico se le conoce como *Cadena de Markov*. Es llamada Cadena de Markov *homogénea* (CMH) si en adición, el lado derecho de 2.1 es independiente de  $n$ .

La propiedad 2.1 es conocida como *la propiedad de Markov*. La matriz  $\mathbf{P} = \{p_{ij}\}_{i,j \in E}$ , donde

$$p_{ij} = P(X_{n+1} = j | X_n = i), \quad (2.2)$$

es la *matriz de transición* de la CMH. Como cada una de sus entradas son probabilidades, y como una transición desde cualquier estado  $i$  debe de ser *algún* estado, se sigue que

$$p_{ij} \geq 0, \quad \sum_{k \in E} p_{ik} = 1$$

para todo estado  $i, j$ . Una matriz  $\mathbf{P}$  indexada por  $E$  y que satisface las propiedades anteriores es llamada *matriz estocástica*. El espacio de estados podría ser infinito, sin embargo, las operaciones de suma y multiplicación estarán definidas por las mismas reglas formales. Por ejemplo, con  $A = \{a_{ij}\}_{i,j \in E}$  y  $B = \{b_{ij}\}_{i,j \in E}$ , el producto  $C = AB$  es la matriz  $\{c_{ij}\}_{i,j \in E}$ , donde  $c_{ij} = \sum_{k \in E} a_{ik} b_{kj}$ . La notación  $x = \{x_i\}_{i \in E}$  representa formalmente a un vector columna, y  $x^t$  es un vector fila, la traspuesta de  $x$ . Por ejemplo,  $y = \{y_i\}_{i \in E}$  dado por  $y^t = x^t A$  está definido por  $y_i = \sum_{k \in E} x_k a_{ki}$ . Similarmente,  $z = \{z_i\}_{i \in E}$  dado por  $z = Ax$  está definido por  $z_i = \sum_{k \in E} a_{ik} z_k$ .

**Ejemplo 2.1.1.** Sea  $\{U_n\}_{n \geq 1}$  una sucesión de v.a.i.i.d. con valores en  $\{1, 2, \dots\}$ , con  $U_n$  representando el tiempo de vida de una máquina, el momento  $n + 1$  es el primer momento después de que la máquina falle. Así en un tiempo 0, la máquina 1 es puesta en servicio hasta que falle en el tiempo  $U_1$ , donde se reemplaza inmediatamente por la máquina 2, la cual deja de funcionar en el tiempo  $U_1 + U_2$ , así consecutivamente. El tiempo en servicio de la máquina actual en el tiempo  $n$  se representa por la variable aleatoria  $X_n$ . Así, el proceso  $\{X_n\}_{n \geq 0}$  que toma valores en  $E = \mathbb{N}$  e incrementa linealmente desde 0 hasta el tiempo  $R_k = \sum_{i=1}^k U_i$ .

La sucesión  $\{R_k\}_{k \geq 0}$  definida de la manera anterior, con  $R_0 = 0$ , es conocida como *sucesión de renovación*, y a  $X_n$  se le conoce como *recurrencia hacia atrás* en el tiempo  $n$ . Existe una amplia y útil teoría asociada con las sucesiones de renovación.

El proceso  $\{X_n\}_{n \geq 0}$  es una CMH con espacio de estados  $E = \mathbb{N}$ , y las entradas no nulas de su matriz de transición son de la forma  $p_{i,i+1}$  con  $p_{i,0} = 1 - p_{i,i+1}$ , donde

$$p_{i,i+1} = \frac{P(U_1 > i + 1)}{P(U_1 > i)}. \quad (2.3)$$

Para probar esto, se debe verificar 2.1.1, esto es, escribir

$$\begin{aligned} B &= \{X_0 = i_0, \dots, X_{n-1} = i_{n-1}\}, \\ \frac{P(X_{n+1} = j, X_n = i, B)}{P(X_n = i, B)} &= \frac{P(X_{n+1} = j, X_n = i)}{P(X_n = i)} \end{aligned}$$

para las sucesiones  $i_0, \dots, i_{n-1}, i, j$  tal que  $P(B, X_n = i, X_{n+1} = j) > 0$ . En particular,  $j = i + 1$  ó 0, e  $i_{n-1} = i - 1, \dots, i_{n-1} = 0$ .

Sea  $v(n)$  el número de tiempos de renovación  $R_k$  en el intervalo  $[1, n]$ . Tomando, por ejemplo,  $j = i + 1$ , y escribiendo  $D = \{X_{n-i-1} = i_{n-i-1}, \dots, X_0 = i_0\}$  se tiene que

$$\begin{aligned} &P(X_{n+1} = j, X_n = i, B) \\ &= P(X_{n+1} = i + 1, X_n = i, X_{n-1} = i - 1, \dots, X_{n-i} = 0, D) \\ &= \sum_{k=0}^{\infty} P(X_{n+1} = i + 1, X_n = i, X_{n-1} = i - 1, \dots, X_{n-i} = 0, D, v(n) = k). \end{aligned}$$

El término general en la última suma equivale a

$$\begin{aligned}
 & P(X_{n+1} = i + 1, X_n = i, X_{n-1} = i - 1, \dots, X_{n-i} = 0, D, v(n) = k) \\
 &= P(U_{k+1} > i, R_k = n - i, D) \\
 &= P(U_{k+1} > i + 1)P(R_k = n - i, D) \\
 &= P(U_1 > i + 1)P(R_k = n - i, D)
 \end{aligned}$$

La independencia de  $\{U_n\}_{n \geq 1}$  ha sido usada para la primera igualdad, y la identidad de las distribuciones de  $U_{k+1}$  y  $U_1$  en la segunda. Por lo tanto,

$$P(X_{n+1} = i + 1, X_n = i, B) = P(U_1 > i + 1) \left( \sum_{k=0}^{\infty} P(R_k = n - i, D) \right).$$

De cálculos similares se obtiene

$$P(X_n = i, B) = P(U_1 > i) \left( \sum_{k=0}^{\infty} P(R_k = n - i, D) \right).$$

entonces

$$P(X_{n+1} = i + 1 | X_n = i, B) = \frac{P(U_1 > i + 1)}{P(U_1 > i)}.$$

Los mismos cálculos dirigen al mismo resultado para  $P(X_{n+1} = i + 1 | X_n = i)$ .

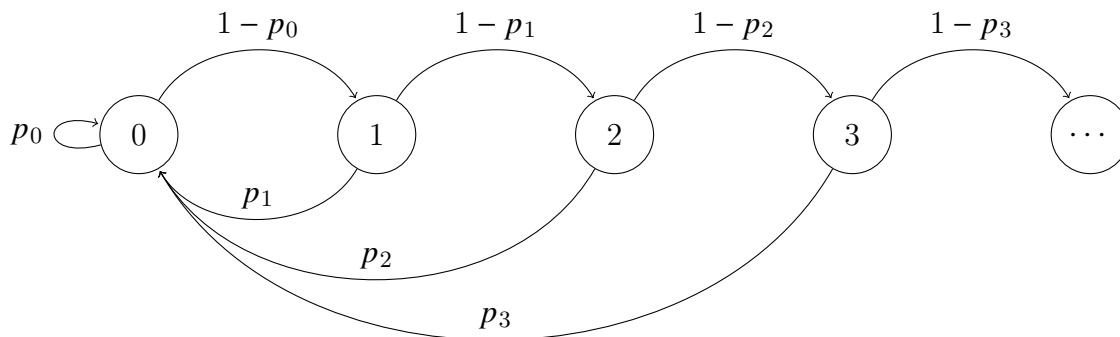
El ejemplo anterior es atípico. Probar la propiedad de Markov y calcular las probabilidades de transición son usualmente más fáciles. La mayor parte del tiempo, una representación del estado del proceso en términos de una ecuación de recurrencia hace las cosas fáciles. Sin embargo hay algunos casos difíciles.

## 2.1.2. Gráfica de Transición

Una matriz de transición  $\mathbf{P}$  a veces es representada por su *gráfica de transición*  $G$ , una gráfica con nodos (o vértices) que representan los estados de  $E$ . Esta gráfica tiene una línea orientada de  $i$  a  $j$  si y sólo si  $p_{ij} > 0$ , en este caso la arista está adornada con la etiqueta de  $p_{ij}$ .

A continuación se muestra la gráfica de transición de la cadena de Markov del ejemplo 2.1.1, donde

$$p_i = \frac{P(U_1 = i + 1)}{P(U_1 > i)}.$$



### 2.1.3. Distribución de una CMH

La variable aleatoria  $X_0$  es llamada *estado inicial*, y su distribución de probabilidad  $v$ ,

$$v(i) = P(X_0 = i), \quad (2.4)$$

es la *distribución inicial*. Por el teorema 1.2.3,

$$\begin{aligned} P(X_0 = i_0, X_1 = i_1, \dots, X_k = i_k) \\ = P(X_0 = i_0)P(X_1 = i_1|X_0 = i_0) \cdots P(X_k = i_k|X_{k-1} = i_{k-1}, \dots, X_0 = i_0), \end{aligned}$$

y por lo tanto, en vista de la propiedad homogénea y la definición de la matriz de transición, para toda  $k \geq 0$

$$P(X_0 = i_0, X_1 = i_1, \dots, X_k = i_k) = v(i_0)p_{i_0i_1} \cdots p_{i_{k-1}i_k} \quad (2.5)$$

esta es la *ley de probabilidad*, o la distribución de la CMH.

**Teorema 2.1.1.** La distribución de una CMH a tiempo discreto está determinada por su distribución inicial y su matriz de transición.

La distribución en tiempo  $n$  de la cadena es el vector  $v_n$ , donde

$$v_n(i) = P(X_n = i). \quad (2.6)$$

De el teorema 1.2.2,  $v_{n+1}(j) = \sum_{i \in E} v_n(i)p_{ij}$ , esto es, en forma matricial,  $v_{n+1}^t = v_n^t \mathbf{P}$ . La iteración de esta igualdad lleva a

$$v_n^t = v_0^t \mathbf{P}^n. \quad (2.7)$$

La matriz  $\mathbf{P}^n$  es llamada la *matriz de transición de  $n$ -pasos* porque su término general es

$$p_{ij}(m) = P(X_{n+m} = j|X_n = i). \quad (2.8)$$

De hecho, usando el teorema 1.2.3 y la propiedad de Markov, se puede encontrar para el lado derecho de la igualdad anterior que

$$\sum_{i_1, \dots, i_{m-1} \in E} p_{ii_1} p_{i_1i_2} \cdots p_{i_{m-1}j},$$

y esto es el término general de la  $m$ -ésima potencia de  $\mathbf{P}$ .

La propiedad de Markov en 2.1.1 se extiende a

$$\begin{aligned} P(X_{n+1} = j_1, \dots, X_{n+k} = j_k | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) \\ = P(X_{n+1} = j_1, \dots, X_{n+k} = j_k | X_n = i) \end{aligned} \quad (2.9)$$

para toda  $i_0, \dots, i_{n-1}, i, j_1, \dots, j_k$  tal que ambos lados de la igualdad están definidos. Escribiendo

$$A = \{X_{n+1} = j_1, \dots, X_{n+k} = j_k\}, \quad B = \{X_0 = i_0, \dots, X_{n-1} = i_{n-1}\}$$

la última ecuación 2.9 se lee entonces  $P(A|X_n = i, B) = P(A|X_n = i)$ , que es a su vez equivalente a

$$P(A \cap B|X_n = i) = P(A|X_n = i)P(B|X_n = i). \quad (2.10)$$

En palabras: El futuro a tiempo  $n$  y el pasado a tiempo  $n$  son condicionalmente independientes dado el estado actual  $X_n = i$ . Esto muestra en particular que la propiedad de Markov es independiente de la dirección del tiempo.

*Notación:* Se abreviará  $P(A|X_0 = i)$  como  $P_i(A)$ . Si  $\mu$  es una distribución de probabilidad sobre  $E$ , entonces  $P_\mu(A) = \sum_{i \in E} \mu(i)P_i(A)$  es la probabilidad de  $A$  dado que el estado inicial se distribuye de acuerdo a  $\mu$ .

## 2.2. Recurrencias de Markov

### 2.2.1. Representación Canónica

Muchas CMHs reciben una descripción natural en términos de una ecuación de recurrencia.

**Teorema 2.2.1.** Sea  $\{Z_n\}_{n \geq 1}$  una sucesión de v.a.i.i.d. con valores en un espacio numerable arbitrario  $F$ . Sea  $E$  un espacio numerable, y  $f : E \times F \rightarrow E$  una función. Sea  $X_0$  una variable aleatoria con valores en  $E$ , independiente de  $\{Z_n\}_{n \geq 1}$ . La ecuación de recurrencia

$$X_{n+1} = f(X_n, Z_{n+1}) \quad (2.11)$$

define una CMH.

**Demostración.** Iterar la ecuación 2.11 muestra que para toda  $n \geq 1$ , existe una función  $g_n$  tal que  $X_n = g_n(X_0, Z_1, \dots, Z_n)$ , y por lo tanto

$$\begin{aligned} P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) \\ &= P(f(i, Z_{n+1}) = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) \\ &= P(f(i, Z_{n+1}) = j), \end{aligned}$$

ya que el evento  $\{X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i\}$  se expresa en términos de  $X_0, Z_1, \dots, Z_n$  y es por lo tanto independiente de  $Z_{n+1}$ . Similarmente,

$$P(X_{n+1} = j | X_n = i) = P(f(i, Z_{n+1}) = j).$$

Se tiene por lo tanto una cadena de Markov, es homogénea, ya que el lado derecho de la última ecuación no depende de  $n$ . Explícitamente

$$p_{ij} = P(f(i, Z_1) = j). \quad (2.12)$$

No todas las CMH se describen naturalmente por el teorema 2.2.1. Una ligera modificación de este teorema, sin embargo, amplía este alcance. Su demostración es análoga.

**Teorema 2.2.2.** Sea  $\{Z_n\}_{n \geq 1}$  una sucesión de v.a.i.i.d. con valores en un espacio numerable arbitrario  $F$ . Sea  $E$  un espacio numerable, y  $f : E \times F \rightarrow E$  una función. Sea  $X_0$  una variable aleatoria con valores en  $E$ . Sea la ecuación de recurrencia

$$X_{n+1} = f(X_n, Z_{n+1})$$

Además,  $Z_{n+1}$  condicionalmente independiente de  $Z_n, \dots, Z_1, X_{n-1}, \dots, X_0$  dado  $X_n$ , esto es, para toda  $k, k_1, \dots, k_n \in F, i_0, i_1, \dots, i_{n-1}, i \in E$ ,

$$\begin{aligned} P(Z_{n+1} = k | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0, Z_n = k_n, \dots, Z_1 = k_1) \\ = P(Z_{n+1} = k | X_n = i), \end{aligned}$$

donde la última ecuación es independiente de  $n$ . Entonces  $\{X_n\}_{n \geq 0}$  es una CMH, con matriz de transición  $\mathbf{P}$  dada por

$$p_{ij} = P(f(i, Z_1) = j | X_0 = i).$$

Como observación, no todas las cadenas de Markov homogéneas reciben una descripción «natural» del tipo caracterizado en los últimos dos teoremas, como el primer ejemplo de este capítulo. Sin embargo, para cualquier matriz de transición  $\mathbf{P}$  sobre  $E$ , existe una cadena de Markov  $\{X_n\}_{n \geq 0}$  con esta matriz de transición y con una representación como en el teorema 2.2.1, esta es

$$X_{n+1} = j \quad \text{si} \quad Z_{n+1} \in \left[ \sum_{k=0}^{j-1} p_{X_n k}, \sum_{k=0}^j p_{X_n k} \right),$$

donde  $\{Z_n\}_{n \geq 1}$  es una sucesión de v.a.i.i.d., uniforme sobre  $[0, 1]$ . Aplicando el teorema 2.2.1, y revisando que es una CMH se tiene la matriz de transición anunciada. Esta representación artificial es útil para simular cadenas de Markov pequeñas y también puede ser útil para la teoría.

### 2.2.2. Ejemplos Famosos

Los ejemplos a continuación son usados usualmente para ilustrar la teoría.

**Ejemplo 2.2.1.** Sea  $X_0$  una variable aleatoria con valores en  $\mathbb{Z}$ . Sea  $\{Z_n\}_{n \geq 1}$  una sucesión de v.a.i.i.d. independientes de  $X_0$ , tomando los valores 1 ó  $-1$ , y con la distribución de probabilidad

$$P(Z_n = 1) = p,$$

donde  $p \in (0, 1)$ . El proceso  $\{X_n\}_{n \geq 1}$  definido por

$$X_{n+1} = X_n Z_{n+1}$$

es, en vista del teorema 2.2.1, una CMH, conocida como *caminata aleatoria* sobre  $\mathbb{Z}$ .

**Ejemplo 2.2.2.** Durante un día  $n$ ,  $Z_{n+1}$  máquinas se descomponen, y entran al taller de reparaciones al día  $n + 1$ . Todos los días una máquina entre estas es reparada. Así, denotando por  $X_n$  el número de máquinas en el taller al día  $n$ ,

$$X_{n+1} = (X_n - 1)^+ + Z_{n+1} \tag{2.13}$$

donde  $a^+ = \max(a, 0)$ . En particular, si  $\{Z_n\}_{n \geq 1}$  es una sucesión de v.a.i.i.d. con estado inicial  $X_0$ , entonces  $\{X_n\}_{n \geq 0}$  es una CMH. En términos de la distribución de probabilidad

$$P(Z_1 = k) = a_k, \quad k \geq 0 \tag{2.14}$$



su matriz de transición es

$$\mathbf{P} = \begin{pmatrix} a_0 & a_1 & a_2 & a_3 & \cdots \\ a_0 & a_1 & a_2 & a_3 & \cdots \\ 0 & a_0 & a_1 & a_2 & \cdots \\ 0 & 0 & a_0 & a_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad (2.15)$$

De hecho, por 2.12 y 2.14,

$$\begin{aligned} p_{ij} &= P((i-1)^+ + Z_1 = j) \\ &= P(Z_1 = j - (i-1)^+) \\ &= a_{j-(i-1)^+} \end{aligned}$$

**Ejemplo 2.2.3.** Se almacena un producto determinado para satisfacer una demanda continua. La demanda agregada entre un tiempo  $n$  y un tiempo  $n+1$  es  $Z_{n+1}$  unidades, y se asume que  $\{Z_n\}_{n \geq 1}$  es una sucesión de v.a.i.i.d. independientes del estado inicial  $X_0$  del almacén. El reabastecimiento de existencias se realiza inmediatamente después del tiempo  $n$  para toda  $n \geq 1$ .

Una estrategia popular para gestionar esto es la  $(s, S)$ -estrategia, donde  $s$  y  $S$  son enteros tales que  $0 < s < S$ . Bajo este manejo del inventario, si el nivel de almacenamiento en el tiempo  $n$  no es más grande que  $s$ , entonces se lleva al nivel  $S$  inmediatamente. En otro caso, no se hace nada. El almacenamiento inicial  $X_0$  se asume menor que  $S$ , y por lo tanto  $\{X_n\}_{n \geq 1}$  toma valores en  $E = \{S, S-1, S-2, \dots\}$ . Los valores negativos del almacén están permitidos, con la interpretación de que la demanda no cubierta se cubre inmediatamente hasta reabastecer. Con las reglas operativas descritas, la evolución del almacén está definida por la ecuación

$$X_{n+1} = \begin{cases} X_n - Z_{n+1} & \text{si } s < X_n < S \\ S - Z_{n+1} & \text{si } X_n \leq s \end{cases} \quad (2.16)$$

En vista de 2.16 y el teorema 2.2.1,  $\{X_n\}_{n \geq 1}$  es una CMH.

**Ejemplo 2.2.4.** En este modelo  $Z_n = (Z_n^{(1)}, Z_n^{(2)}, \dots)$ , donde las variables aleatorias  $\{Z_n^{(j)}\}_{n \geq 1, j \geq 1}$  son una sucesión de v.a.i.i.d. con valores en  $\mathbb{N}$ . La ecuación de recurrencia

$$X_{n+1} = \sum_{k=1}^{X_n} Z_{n+1}^{(k)}, \quad (2.17)$$

con la convención de que  $X_{n+1} = 0$  si  $X_n = 0$ , recibe la siguiente interpretación:  $X_n$  es el número de individuos de la  $n$ -ésima generación de una población dada (humanos, partículas, etc.). El individuo número  $k$  de la  $n$ -ésima generación da nacimiento a  $Z_{n+1}^{(k)}$  descendientes.

Si el número  $X_0$  de ancestros es independiente de  $\{Z_n\}_{n \geq 1}$ , entonces de acuerdo al teorema 2.2.1,  $\{X_n\}_{n \geq 0}$  es una CMH, conocida como proceso de ramificación por el árbol genealógico que genera. El proceso de ramificación es también conocido como *proceso Galton-Watson*.

Si se quiere calcular la probabilidad de extinción de la población se usará una función  $g$ , la función generadora en común de las variables  $Z_n^{(k)}$ . La ecuación de recurrencia 2.17 da como herramienta otra ecuación de recurrencia para la función generadora del número de individuos en la  $n$ -ésima generación,

$$\psi_n(z) = E[z^{X_n}]. \quad (2.18)$$

De hecho,

$$\psi_{n+1}(z) = E[z^{X_{n+1}}] = E[z^{\sum_{k=1}^{X_n} Z^{(k)}}],$$

donde  $Z_{n+1}^{(k)} = Z^{(k)}$  por simplicidad. Como  $X_n$  está en función de  $X_0, Z_1, \dots, Z_n$ , es independiente de  $Z_{n+1}$ , y por lo tanto, en la última ecuación,  $X_n$  es independiente de  $Z^{(k)}$ . Por el cálculo hecho en 1.5.3,

$$E\left[z^{\sum_{k=1}^{X_n} Z^{(k)}}\right] = \psi_n(g(z)).$$

Por lo tanto,

$$\psi_{n+1}(z) = \psi_n(g(z)).$$

Iterando la ecuación, se obtiene

$$\psi_{n+1}(z) = \psi_0(g^{(n+1)}(z)),$$

donde  $g^{(n)}$  es la  $n$ -ésima iteración de  $g$ . Si sólo hay *un ancestro*, ocurre que  $\psi_0(z) = z$ , y por lo tanto  $\psi_{n+1}(z) = g^{(n+1)}(z) = g(g^{(n)}(z))$ , esto es,

$$\psi_{n+1}(z) = g(\psi_n(z)). \quad (2.19)$$

En particular, como  $\psi_n(0) = P(X_n = 0)$ , se tiene que

$$P(X_{n+1} = 0) = g(P(X_n = 0)), \quad (2.20)$$

una ecuación que se usará para calcular la probabilidad de extinción. El evento  $\mathcal{E}$  = «la extinción ocurre» es equivalente a «al menos una generación está vacía», esto es,

$$\mathcal{E} = \bigcup_{n=1}^{\infty} \{X_n = 0\}.$$

También, como  $X_n = 0$  implica que  $X_{n+1} = 0$ , la familia  $\{X_n = 0\}$  es no decreciente, y por 1.1.1,

$$P(\mathcal{E}) = \lim_{n \uparrow \infty} P(X_n = 0). \quad (2.21)$$

La función generadora  $g$  es continua y por lo tanto de 2.20 y 2.21, la probabilidad de extinción tiene que satisfacer la ecuación

$$P(\mathcal{E}) = g(P(\mathcal{E})). \quad (2.22)$$

Sea  $Z$  cualquiera de las variables aleatorias  $Z_n^{(k)}$ . Si  $E[Z] \leq 1$ , la única solución para  $x = g(x)$  en  $[0, 1]$ , es 1. Y por lo tanto  $P(\mathcal{E}) = 1$ . El proceso de ramificación eventualmente se extinguirá. Si  $E[Z] > 1$ , hay dos posibles soluciones de  $x = g(x)$

en  $[0, 1]$ ,  $1$  y  $x_0 \in (0, 1)$ . De la convexidad estricta de  $g : [0, 1] \rightarrow [0, 1]$ , se sigue que la sucesión  $y_n = P(X_n = 0)$  que satisface  $y_0 = 0$  y  $y_{n+1} = g(y_n)$  converge a  $x_0$ . Por lo tanto, cuando el número promedio de descendientes  $E[Z]$  es estrictamente más grande que  $1$ , ocurre que  $P(\mathcal{E}) \in (0, 1)$ , y en particular, hay una probabilidad nula de extinción.

Como un pequeño ejemplo, supóngase que

$$g(z) = \frac{1}{4} + \frac{1}{4}z + \frac{1}{2}z^2,$$

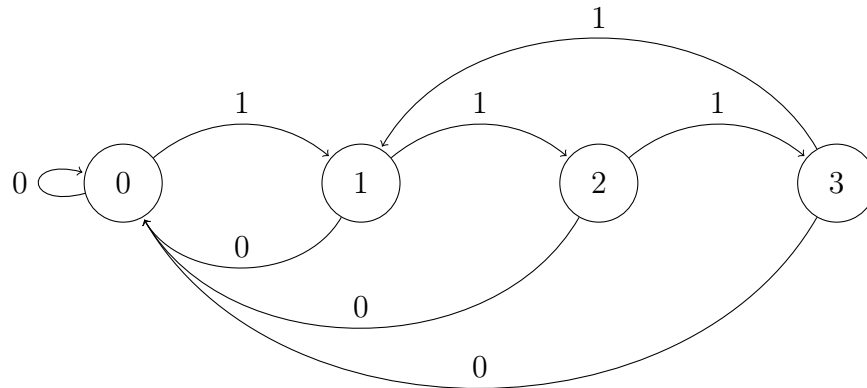
esto es, las probabilidades de tener  $0, 1$  ó  $2$  descendientes son  $\frac{1}{4}, \frac{1}{4}$  y  $\frac{1}{2}$  respectivamente. En particular  $E[Z] = 1.25 > 1$ , y  $P(\mathcal{E})$  es la solución a la siguiente ecuación

$$x = \frac{1}{4} + \frac{1}{4}x + \frac{1}{2}x^2,$$

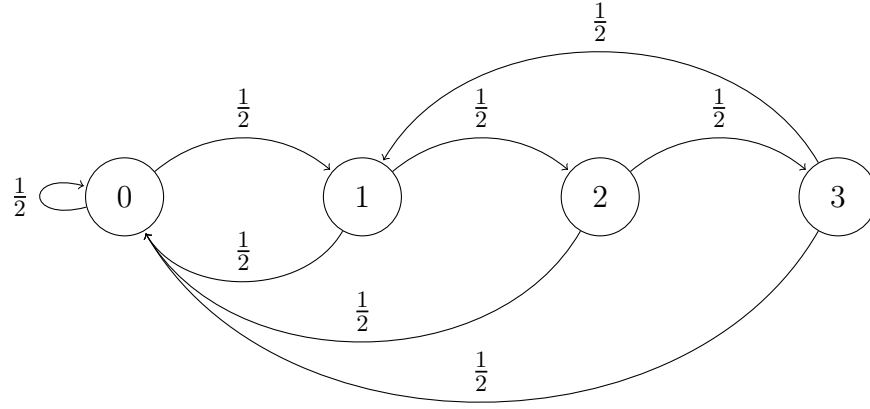
es decir,  $P(\mathcal{E}) = \frac{1}{2}$ .

**Ejemplo 2.2.5.** Un robot autómatata puede leer sucesiones de letras de un abecedario finito  $\mathcal{A}$  escritas en una cinta infinita. Este puede estar en cualquier estado de un conjunto finito  $E$ , y su evolución está definida por la función  $f : E \times \mathcal{A} \rightarrow E$ , como sigue. Cuando el robot está en el estado  $i \in E$  y lee la letra  $a \in \mathcal{A}$ , cambia del estado  $i$  al estado  $j = f(i, a)$  y entonces lee sobre la cinta la siguiente letra a la derecha.

Para ilustrar esto, sea  $\mathcal{A} = \{0, 1\}$  y  $E = \{1, 2, 3, 4\}$ , donde  $f$  queda definida por el siguiente diagrama donde cada nodo representa los valores de  $E$  y la etiqueta de las aristas los valores de  $\mathcal{A}$ , por ejemplo  $f(0, 1) = 1$  significa que se pasa del estado  $0$  al estado  $1$  si el robot lee  $1$ .



Su gráfica de transición para este proceso sería similar



La lectura de la siguiente cadena,

1 0 0 1 1 1 1 0 0 1 1 1 1 1 0 1 0

da como resultado la siguiente sucesión de estados,

0 1 0 0 1 2 3 1 0 0 1 2 3 1 2 3 0 1 0 .

En general, existe una arista que conecta al estado  $i$  con el estado  $j$  si y sólo si existe  $a \in \mathcal{A}$  tal que  $j = f(i, a)$ , y esta arista recibe la etiqueta de  $a$ . Si  $j = f(i, a_1) = f(i, a_2)$  para  $a_1 \neq a_2$ , entonces existen dos aristas de  $i$  a  $j$  con etiquetas  $a_1$  y  $a_2$ , o una arista con la etiqueta  $(a_1, a_2)$ . Una arista podría tener múltiples etiquetas de cualquier orden.

Si la sucesión de letras leídas por el robot es  $\{Z_n\}_{n \geq 1}$ , la sucesión de estados  $\{X_n\}_{n \geq 0}$  está dada por la ecuación de recurrencia  $X_{n+1} = f(X_n, Z_{n+1})$  y por lo tanto, si  $\{Z_n\}_{n \geq 1}$  es una sucesión de v.a.i.i.d. independiente del estado  $X_0$ , entonces  $\{X_n\}_{n \geq 0}$ , es, de acuerdo al teorema 2.2.1, una CMH.

**Ejemplo 2.2.6.** Este modelo simplificado de difusión a través de una membrana porosa fue propuesto en 1907 por los físicos Austríacos Tatiana y Paul Ehrenfest [27] para describir en términos de mecánica estadística el intercambio de calor entre dos sistemas a diferentes temperaturas. Su modelo también fue usado para entender la irreversibilidad termodinámica.

Sean  $N$  partículas distribuidas entre el compartimento  $A$  o en el compartimento  $B$ . Supóngase que en un tiempo  $n \geq 0$ ,  $X_n = i$  partículas se encuentran en  $A$ . Se elige una partícula de manera aleatoria, y esta partícula es movida en el tiempo  $n + 1$  de donde fue tomada al otro compartimento. Así, el siguiente estado  $X_{n+1}$  está entre  $i - 1$  (la partícula fue tomada de  $A$ ) con probabilidad  $\frac{i}{N}$ , o  $i + 1$  (fue tomada de  $B$ ) con probabilidad  $\frac{N-i}{N}$ .

Este modelo se puede describir con el teorema 2.2.2. Para toda  $n \geq 0$ ,

$$X_{n+1} = X_n + Z_{n+1} \tag{2.23}$$

donde  $Z_n \in \{-1, 1\}$  y  $P(Z_{n+1} = -1 | X_n = i) = \frac{i}{N}$ . Las entradas no nulas de la matriz de transición son por lo tanto

$$p_{i,i+1} = \frac{N-i}{N}, \quad p_{i,i-1} = \frac{i}{N}. \tag{2.24}$$

## 2.3. Análisis del primer paso

### 2.3.1. Probabilidad de absorción

Muchos funcionales de las cadenas de Markov Homogéneas, en particular las probabilidades de absorción por una clase absorbente ( $A$  es llamada clase absorbente si  $\sum_{j \in A} p_{ij} = 1$  para toda  $i \in A$ ) y tiempos promedio antes de la absorción, pueden ser evaluados a través de una técnica conocida como *análisis del primer paso*. Esta técnica, la cual es el motor de la mayoría de los cálculos en la teoría de las cadenas de Markov, se ilustra con el siguiente ejemplo.

**Ejemplo 2.3.1.** Dos jugadores  $A$  y  $B$  juegan «cara o cruz», donde la probabilidad de caer cara es  $p \in (0, 1)$ , y los resultados sucesivos forman una sucesión de v.a.i.i.d.. Sea  $X_n$  la fortuna del jugador  $A$  en el tiempo  $n$ , entonces  $X_{n+1} = X_n + Z_{n+1}$ , donde  $Z_{n+1} = 1$  ó  $-1$  con probabilidad  $p$  o  $q = 1 - p$ , y  $\{Z_n\}_{n \geq 1}$  es una sucesión de v.a.i.i.d.. En otras palabras,  $A$  gana \$1 por cada vez que caiga cara, y  $B$  gana \$1 cada vez que caiga cruz. Las fortunas iniciales de  $A$  y  $B$  son  $a$  y  $b$ . El juego termina cuando un jugador queda en la ruina, y por lo tanto el proceso  $\{X_n\}_{n \geq 1}$  es una caminata aleatoria, como en el ejemplo 2.2.1, excepto en que este está restringido a  $E = \{0, \dots, a, a+1, \dots, a+b = c\}$ . La duración del juego es  $T$ , el primer momento en el cual  $X_n = 0$  o  $c$ , y la probabilidad de ganar para  $A$  es  $u(a) = P(X_T = c | X_0 = a)$ .

En lugar de calcular  $u(a)$  solo, el análisis del primer paso calcula

$$u(i) = P(X_T = c | X_0 = i)$$

para todos los estados  $i \in [0, c]$ , y para esto, se genera primero una ecuación de recurrencia para las  $u(i)$ 's particionando el evento « $A$  gana» de acuerdo a lo que puede pasar después del primer paso (el primer lanzamiento). Si  $X_0 = i \in [1, c - 1]$ , entonces  $X_1 = i + 1$  (respectivamente,  $X_1 = i - 1$ ) con probabilidad  $p$  (respectivamente,  $q$ ), y la probabilidad de ruina de  $B$  empezando  $A$  con fortuna inicial  $i + 1$  (resp.,  $i - 1$ ) es  $u(i + 1)$  (resp.,  $u(i - 1)$ ). Por lo tanto para  $i \in [1, c - 1]$ ,

$$u(i) = pu(i + 1) + qu(i - 1), \quad (2.25)$$

con las condiciones

$$u(0) = 0, \quad u(c) = 1. \quad (2.26)$$

La ecuación característica asociada con esta ecuación de recurrencia lineal es  $pr^2 - r + q = 0$ . Este tiene dos raíces distintas,  $r_1 = 1$  y  $r_2 = \frac{q}{p}$  si  $p \neq q$ , y la multiplicidad  $r_1 = 1$  si  $p = q = \frac{1}{2}$ . Por lo tanto, la solución general es  $u(i) = \lambda r_1^i + \mu r_2^i = \lambda + \mu \left(\frac{p}{q}\right)^i$  cuando  $p \neq q$ , y  $u(i) = \lambda r_1^i + \mu i r_1^i = \lambda + \mu i$  cuando  $p = q = \frac{1}{2}$ . Tomando en cuenta las condiciones 2.26, se puede determinar los valores de  $\lambda$  y  $\mu$ . El resultado es, para  $p \neq q$ ,

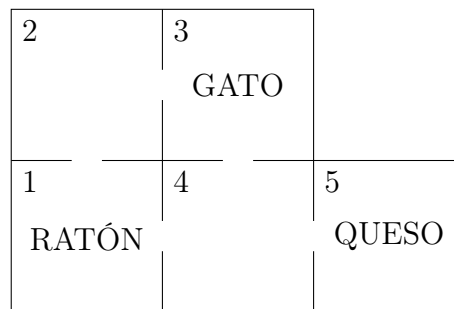
$$u(i) = \frac{1 - \left(\frac{q}{p}\right)^i}{1 - \left(\frac{q}{p}\right)^c}, \quad (2.27)$$

y para  $p = q = \frac{1}{2}$ ,

$$u(i) = \frac{i}{c}. \quad (2.28)$$

En el caso  $p = q = \frac{1}{2}$ , la probabilidad  $v(i)$  de que  $B$  gana cuando la fortuna inicial de  $B$  es  $c - i$  se obtiene reemplazando  $i$  por  $c - i$  en la expresión 2.28:  $v(i) = \frac{c-i}{c} = 1 - \frac{i}{c}$ . Se puede ver que  $u(i) + v(i) = 1$ , lo cual significa en particular que la probabilidad de que el juego duré para siempre es nula.

**Ejemplo 2.3.2.** Un ratón se mueve en un laberinto (ilustrado a continuación). Si este está en un tiempo  $n$  en un cuarto con  $k$  cuartos adyacentes, al tiempo  $n + 1$  este estará en uno de estos  $k$  cuartos, eligiendo uno aleatoriamente, cada uno con probabilidad  $\frac{1}{k}$ . Un gato se encuentra todo el tiempo en un cuarto dado, y una pieza de queso en otro cuarto diferente. Si el ratón entra en el cuarto del gato, este será comido. ¿Cuál será la probabilidad de que el ratón entre al cuarto del queso cuando empieza desde el cuarto 1, siendo que el gato y el queso están en los cuartos 3 y 5 respectivamente?



Para aplicar el análisis del primer paso, sea  $u(i)$  la probabilidad de que el ratón inicia en el cuarto  $i$  llega al queso sin ser comido por el gato. Las condiciones

$$u(3) = 0, \quad u(5) = 1$$

son claras. Si el ratón está en el cuarto 1, su primer movimiento lo llevará al cuarto 2 (resp. 4) con probabilidad  $\frac{1}{2}$ , y la probabilidad de comerse el queso será entonces  $u(2)$  (resp.  $u(4)$ ). Por lo tanto,

$$u(1) = \frac{1}{2}u(2) + \frac{1}{2}u(4)$$

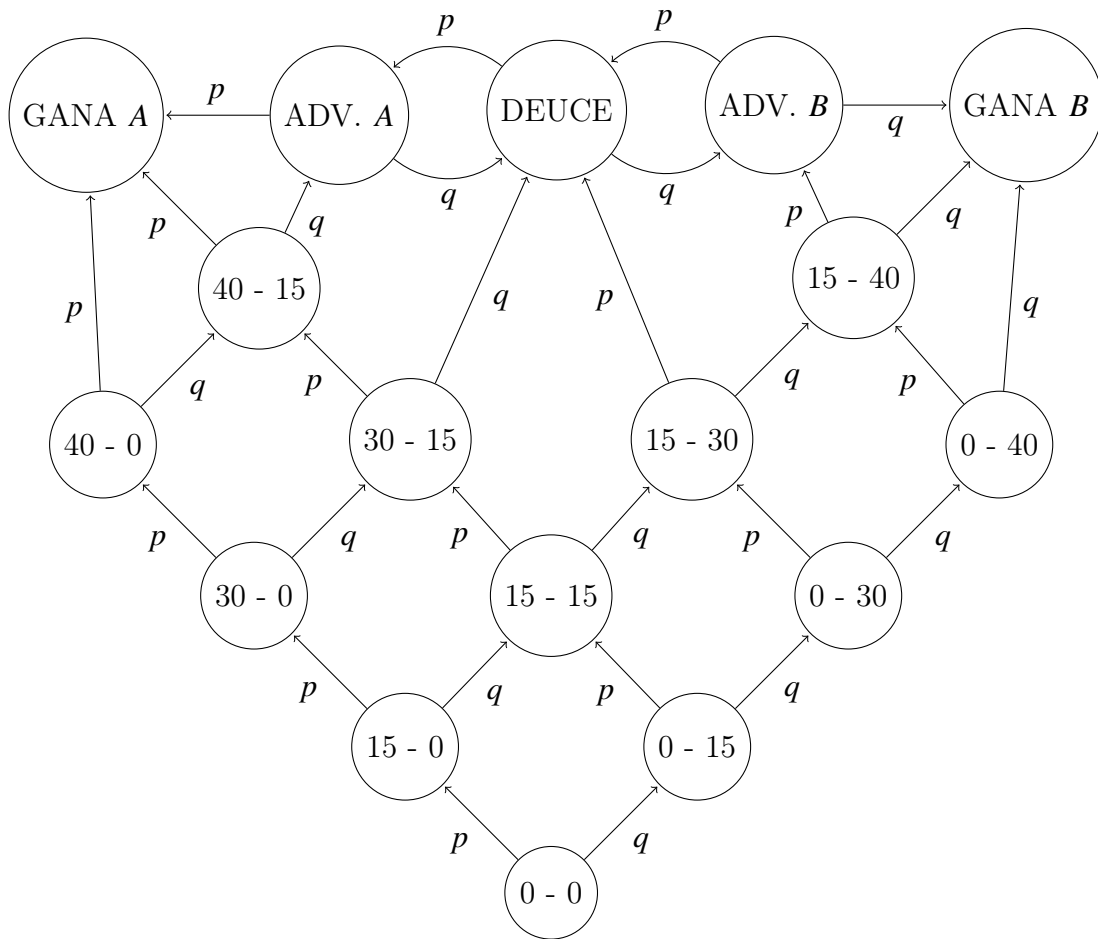
Similarmente,

$$u(2) = \frac{1}{2}u(1) + \frac{1}{2}u(3) = \frac{1}{2}u(1),$$

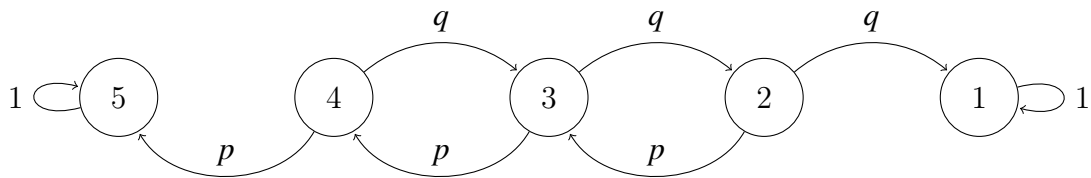
$$u(4) = \frac{1}{3}u(1) + \frac{1}{3}u(3) + \frac{1}{3}u(5) = \frac{1}{3}u(1) + \frac{1}{3}.$$

La solución es  $u(1) = \frac{2}{7}$ .

**Ejemplo 2.3.3.** Ignorando los desempates, un juego de tenis puede ser modelado como una cadena de Markov con la gráfica de transición dibujada a continuación, donde  $p$  es la probabilidad de que el jugador  $A$  gane el punto, y  $q = 1 - p$ .



Si se quiere calcular la probabilidad de que  $B$  gane el juego, se puede observar que el diagrama anterior se divide en dos partes: primero se recorren los estados en los primeros 5 niveles hasta llegar a los estados en el nivel superior donde evoluciona hasta la absorción en «GANA A» o «GANA B». Convirtiendo este nivel superior en estados del 5 al 1, se obtiene la siguiente gráfica de transición.



Esta gráfica de transición es similar a la que se generaría en el ejemplo de la ruina 2.3.1. Para un análisis de primer paso, sea  $b_i$ , la probabilidad de que  $B$  gane dado que el juego empieza en el estado superior  $i$ , con

$$b_1 = 1, \quad b_5 = 0$$

y

$$\begin{aligned} b_2 &= q + pb_3, \\ b_3 &= qb_2 + pb_4, \\ b_4 &= qb_3. \end{aligned}$$

Para  $p \neq q$ ,

$$(b_1, b_2, b_3, b_4, b_5) = \left(1, q \frac{1-pq}{1-2pq}, \frac{q^2}{1-2pq}, \frac{q^3}{1-2pq}, 0\right)$$

Empezando del estado 0 - 0, la probabilidad de que  $B$  gane es  $\sum_{i=1}^5 p(i)b_i$ , donde  $p(i)$  es la probabilidad de que el primer estado superior alcanzado sea  $i$ . Una simple enumeración de lo posibles caminos desde el 0 - 0 al estado 1 da  $p(1) = q^4 + q^3pq + q^2pq^2 + qpq^3 + pq^4$ , esto es,

$$p(1) = q^4(1 + 4p).$$

Cálculos similares dan como resultado a

$$p(2) = 4q^3p^2, \quad p(3) = 6p^2q^2, \quad p(4) = 4p^3q^2, \quad p(5) = p^4(1 + 4q).$$

Uniendo todo lo anterior, se encuentra que la probabilidad para que  $B$  gane es

$$\frac{q^4(1 + 2p)(1 + 4p^2)}{1 - 2pq}$$

### 2.3.2. Tiempo medio de absorción

Los ejemplos 2.3.1, 2.3.2 y 2.3.3 lidian básicamente con el mismo problema, el de calcular la probabilidad de llegar a un estado antes de que otro estado sea visitado. El análisis del primer paso también puede ser usado para calcular tiempos promedio antes de la absorción.

**Ejemplo 2.3.4.** Continuando con el ejemplo 2.3.1. La duración promedio  $m(i) = E[T|X_0 = i]$  del juego cuando la fortuna inicial del jugador  $A$  es  $i$  satisface la siguiente ecuación de recurrencia

$$m(i) = 1 + pm(i + 1) + qm(i - 1) \quad (2.29)$$

para  $i \in [1, c - 1]$ . Si la moneda es lanzada al menos una vez, entonces con probabilidad  $p$  (resp.  $q$ ) la fortuna del jugador  $A$  será  $i + 1$  (resp.  $i - 1$ ), y por lo tanto  $m(i + 1)$  (resp.  $m(i - 1)$ ) más lanzamientos serán necesarios en promedio antes de que uno de los jugadores se arruine. Con las condiciones

$$m(0) = 0, \quad m(c) = 0. \quad (2.30)$$

Para resolver 2.29 con las condiciones 2.30, se escribe 2.29 de la forma

$$-1 = p(m(i + 1) - m(i)) - q(m(i) - m(i - 1)).$$

Definiendo

$$y_i = m(i) - m(i - 1),$$

se tiene, para  $i \in [1, c - 1]$ ,

$$-1 = py_{i+1} - qy_i \quad (2.31)$$



y

$$m(i) = y_1 + y_2 + \cdots + y_i. \quad (2.32)$$

Si se resuelve 2.31 para  $p = q = \frac{1}{2}$

$$\begin{aligned} -1 &= \frac{1}{2}y_2 - \frac{1}{2}y_1, \\ -1 &= \frac{1}{2}y_3 - \frac{1}{2}y_2, \\ &\vdots \\ -1 &= \frac{1}{2}y_i - \frac{1}{2}y_{i-1}, \end{aligned}$$

y por lo tanto, sumando,

$$-(i-1) = \frac{1}{2}y_i - \frac{1}{2}y_1,$$

esto es, para  $i \in [1, c]$ ,

$$y_i = y_1 - 2(i-1).$$

Usando esta ecuación en 2.32, y observando que  $y_1 = m(1)$ , se obtiene

$$m(i) = im(1) - 2[1 + 2 + \cdots + (i-1)] = im(1) - i(i-1).$$

La condición  $m(c) = 0$  lleva a  $cm(1) = c(c-1)$  y por lo tanto, finalmente,

$$m(i) = i(c-i).$$

El análisis del primer paso guía a las condiciones necesarias en forma de un sistema de ecuaciones lineales. En los ejemplos anteriores, el sistema en cuestión tiene una solución única, una situación que prevalece cuando el espacio de estados es finito pero que no es el caso general con un espacio de estados infinito. La cuestión de la unicidad, y de qué solución elegir en caso de no singularidad, se aborda en el Capítulo 4, donde se estudia la absorción con más detalle.

## 2.4. La Topología de la Matriz de Transición

### 2.4.1. Comunicación

Todas las propiedades definidas en esta sección son *topológicas* en el sentido de que se refieren sólo al gráfico de transición sin las etiquetas.

**Definición 2.4.1.** Un estado  $j$  se dice que es *accesible* desde el estado  $i$  si existe  $M \geq 0$  tal que  $p_{ij}(M) > 0$ . En particular, un estado  $i$  siempre es accesible desde él mismo, como  $p_{ij}(0) = 1$ . Los estados  $i$  y  $j$  se dicen que están *comunicados* si  $i$  es accesible desde  $j$  y  $j$  es accesible desde  $i$ , y esto se denota por  $i \leftrightarrow j$ .

Para  $M \geq 1$ ,  $p_{ij}(M) = \sum_{i_1, \dots, i_{M-1}} p_{ii_1} p_{i_1 i_2} \cdots p_{i_{M-1} j}$ , y por lo tanto  $p_{ij}(M) > 0$  si y sólo si existe al menos un camino  $i, i_1, \dots, i_{M-1}, j$  desde  $i$  a  $j$  tal que

$$p_{ii_1} p_{i_1 i_2} \cdots p_{i_{M-1} j} > 0,$$

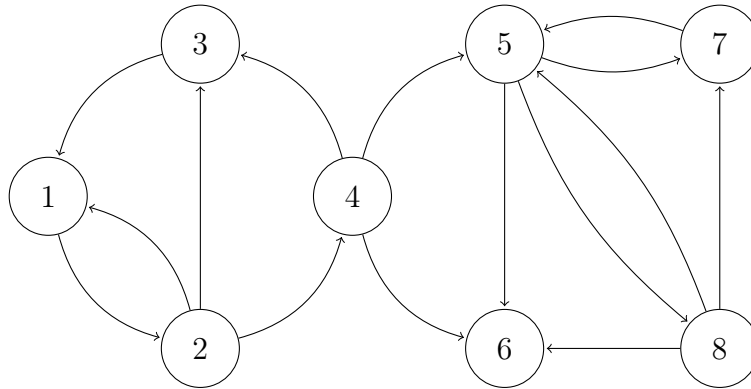
si existe un camino orientado desde  $i$  a  $j$  en la gráfica de transición  $G$ . Se puede notar que,

$$\begin{aligned} i \leftrightarrow j & \text{ reflexividad,} \\ i \leftrightarrow j & \Rightarrow j \leftrightarrow i \text{ simetría,} \\ i \leftrightarrow j, j \leftrightarrow k & \Rightarrow i \leftrightarrow k \text{ transitividad} \end{aligned}$$

Por lo tanto, la relación de comunicación ( $\leftrightarrow$ ) es una relación de equivalencia, y genera una partición en el espacio de estados  $E$  generando clases de equivalencia llamadas *clases de comunicación*.

**Definición 2.4.2.** Un estado  $i$  tal que  $p_{ii} = 1$  se conoce como *absorbente*. En general, un conjunto  $C$  de estados tal que para toda  $i \in C, \sum_{j \in C} p_{ij} = 1$  es llamado *absorbente*.

**Ejemplo 2.4.1.** El siguiente diagrama tiene 3 clases de comunicación:  $\{1, 2, 3, 4\}$ ,  $\{5, 7, 8\}$  y  $\{6\}$ . Donde el estado 6 es absorbente. La clase de comunicación  $\{5, 7, 8\}$  no es absorbente, pero el conjunto  $\{5, 6, 7, 8\}$  si lo es.



**Definición 2.4.3.** Si existe solo una clase de comunicación, entonces la cadena, su matriz de transición, y su gráfica de transición, se dice que son *irreducibles*.

## 2.4.2. Periodo

Sea la caminata aleatoria del ejemplo 2.2.1. Como  $p \in (0, 1)$ , esta es irreducible. Se puede observar que  $E = C_0 + C_1$ , donde  $C_0$  y  $C_1$ , el conjunto de enteros pares e impares respectivamente, tiene la siguiente propiedad. Si se comienza desde  $i \in C_0$  (resp.  $C_1$ ), entonces en un paso solo se puede ir a un estado  $j \in C_1$  (resp.  $C_0$ ). La cadena  $\{X_n\}$  se recorre alternando desde una clase a la otra. En este sentido, la cadena tiene un comportamiento periódico, cuyo periodo es 2. En general, se tiene lo siguiente.

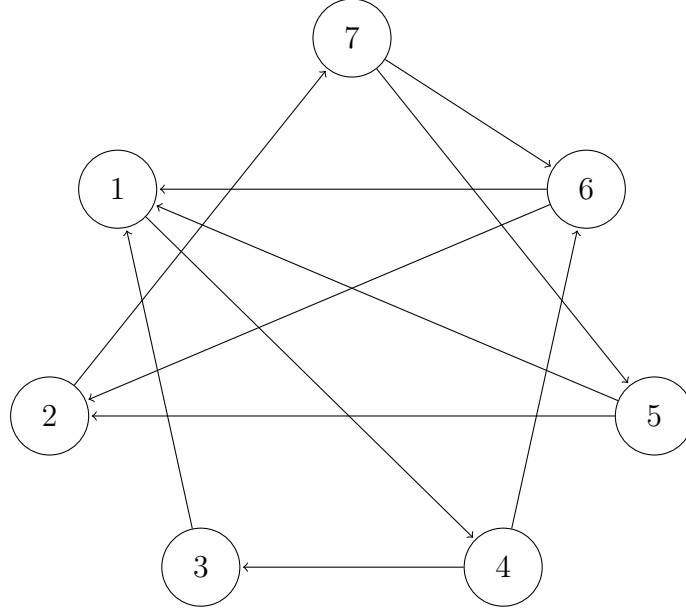
**Teorema 2.4.1.** Para cualquier cadena de Markov irreducible, se puede encontrar una única partición de  $E$  en  $d$  clases  $C_0, C_1, \dots, C_{d-1}$  tal que para toda  $k$ , y para toda  $i \in C_k$ ,

$$\sum_{j \in C_k} p_{ij} = 1,$$

con la convención de  $C_0 = C_d$ , y donde  $d$  es maximal, esto es, no existe otra partición  $C'_0, C'_1, \dots, C'_{d'-1}$  con  $d' > d$ .

El número  $d \geq 1$  es conocido como *periodo* de la cadena. Las clases de esta  $C_0, C_1, \dots, C_{d-1}$  son conocidas como *clases cíclicas*. La cadena se traslada de una clase a otra, y esto es cíclico.

**Ejemplo 2.4.2.** La cadena con la siguiente gráfica de transición es irreducible y tiene periodo  $d = 3$ , con las clases cíclicas  $C_0 = \{1, 2\}$ ,  $C_1 = \{4, 7\}$ ,  $C_2 = \{3, 5, 7\}$ .



Se considera ahora una cadena irreducible de periodo  $d$  con las clases cíclicas  $C_0, C_1, \dots, C_d$ . Enumerando de nuevo los estados de  $E$  si es necesario, la matriz de transición tiene la estructura siguiente, donde  $d = 4$ ,

$$\mathbf{P} = \begin{matrix} & C_0 & C_1 & C_2 & C_3 \\ \begin{matrix} C_0 \\ C_1 \\ C_2 \\ C_3 \end{matrix} & \begin{pmatrix} 0 & A_0 & 0 & 0 \\ 0 & 0 & A_1 & 0 \\ 0 & 0 & 0 & A_2 \\ A_3 & 0 & 0 & 0 \end{pmatrix} \end{matrix}$$

y por lo tanto  $\mathbf{P}^2$ ,  $\mathbf{P}^3$  y  $\mathbf{P}^4$  también tienen una estructura similar correspondiente a  $C_0, C_1, C_2, C_3$ :

$$\mathbf{P}^2 = \begin{pmatrix} 0 & 0 & B_0 & 0 \\ 0 & 0 & 0 & B_1 \\ B_2 & 0 & 0 & 0 \\ 0 & B_3 & 0 & 0 \end{pmatrix}, \mathbf{P}^3 = \begin{pmatrix} 0 & 0 & 0 & D_0 \\ D_1 & 0 & 0 & 0 \\ 0 & D_2 & 0 & 0 \\ 0 & 0 & D_3 & 0 \end{pmatrix}, \mathbf{P}^4 = \begin{pmatrix} E_0 & 0 & 0 & 0 \\ 0 & E_1 & 0 & 0 \\ 0 & 0 & E_2 & 0 \\ 0 & 0 & 0 & E_3 \end{pmatrix}$$

Se observan dos fenómenos: los cambios de bloques y la diagonal de bloques que forman  $\mathbf{P}^4$ . En general:  $\mathbf{P}^d$  tiene una diagonal de bloques correspondiente a las clases cíclicas  $C_0, C_1, \dots, C_{d-1}$ :

$$\mathbf{P}^d = \begin{matrix} & C_0 & C_1 & \cdots & C_{d-1} \\ \begin{matrix} C_0 \\ C_1 \\ \vdots \\ C_{d-1} \end{matrix} & \begin{pmatrix} E_0 & 0 & \cdots & 0 \\ 0 & E_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & E_{d-1} \end{pmatrix} \end{matrix} \quad (2.33)$$

La matriz de transición en  $d$  pasos  $\mathbf{P}^d$  es también una matriz estocástica. Además, sea  $i, j \in C_0$ , como estos estados están comunicados, existe  $m > 0$  y  $n > 0$  tal que  $p_{ij}(m) > 0$  y  $p_{ij}(n) > 0$ . Pero como  $\mathbf{P}$  tiene periodo  $d$ , necesariamente  $m = Md, n = Nd$  para alguna  $M > 0$  y  $N > 0$ . Por lo tanto,  $p_{ij}(Md) > 0$  y  $p_{ij}(Nd) > 0$ . Pero  $p_{ij}(Md)$  es el término  $(i, j)$  de  $(\mathbf{P}^d)^M$ , y similarmente para  $p_{ij}(Nd)$ . Se dice que  $i$  y  $j$  están  $\mathbf{P}^d$ -comunicados.

Para una matriz de transición arbitraria, no necesariamente irreducible, la noción formal de periodo es la siguiente.

**Definición 2.4.4.** El periodo  $d_i$  del estado  $i \in E$  es, por definición,

$$d_i = \text{mcd}\{n \geq 1 : p_{ii}(n) > 0\}, \quad (2.34)$$

con la convención de  $d_i = \infty$  si no existe  $n \geq 1$  con  $p_{ii}(n) > 0$ . Si  $d_i = 1$ , el estado  $i$  es llamado *aperiódico*.

**Teorema 2.4.2.** Si los estados  $i$  y  $j$  se comunican entonces tienen el mismo periodo.

**Demostración.** Como  $i$  y  $j$  están comunicados, existen enteros  $N$  y  $M$  tal que  $p_{ij}(M) > 0$  y  $p_{ji}(N) > 0$ . Para cualquier  $k \geq 1$ ,

$$p_{ii}(M + nk + N) \geq p_{ij}(M)(p_{jj}(k))^n p_{ji}(N)$$

(de hecho, el camino  $X_0 = i, X_M = j, X_{M+k} = j, \dots, X_{M+nk} = j, X_{M+nk+N} = i$  es solo un camino de ir desde  $i$  a  $i$  en  $M + nk + N$  pasos).

Por lo tanto, para cualquier  $k \geq 1$  tal que  $p_{jj}(k) > 0$ , se tiene que  $p_{ii}(M+nk+N) > 0$  para toda  $n \geq 1$ . Entonces,  $d_i$  divide a  $M+nk+N$  para toda  $n \geq 1$ , y en particular,  $d_i$  divide a  $k$ . Se tiene por lo tanto que  $d_i$  divide a  $k$  con  $p_{jj}(k) > 0$ , y en particular,  $d_i$  divide a  $d_j$ . Por simetría,  $d_j$  divide a  $d_i$ , y finalmente  $d_i = d_j$ .

También se puede hablar del periodo de una clase de comunicación o de una cadena irreducible.

**Teorema 2.4.3.** Sea  $\mathbf{P}$  una matriz estocástica irreducible con periodo  $d$ . Entonces para todos los estados  $i, j$  existe  $m \geq 0$  y  $n_0 \geq 0$  tal que

$$p_{ij}(m + nd) > 0, \quad \forall n \geq n_0 \quad (2.35)$$

**Demostración.** Como la cadena es irreducible,  $j$  es accesible desde  $i$ , esto es, existe  $m$  tal que  $p_{ij}(m) > 0$  y por lo tanto para alguna  $n_0 \geq 0$  se tiene  $p_{jj}(nd) > 0$  para toda  $n \geq n_0$ , entonces  $p_{ij}(m + nd) > p_{ij}(m)p_{jj}(nd) > 0$  para toda  $n \geq n_0$ . El máximo común divisor del conjunto  $A = \{k \geq 1 : p_{jj}(k) > 0\}$  es  $d$ , y  $A$  es cerrado bajo la suma. El conjunto  $A$  por lo tanto contiene todos pero un número finito de los múltiplos positivos de  $d$  [19]. En otras palabras, existe  $n_0$  tal que  $n > n_0$  implica que  $p_{jj}(nd) > 0$ .

## 2.5. Estados Estacionarios

### 2.5.1. Propiedades Estacionarias

**Definición 2.5.1.** Una distribución  $\pi$  que satisface

$$\pi^t = \pi^t \mathbf{P} \quad (2.36)$$

es llamada *distribución estacionaria* de la matriz de transición  $\mathbf{P}$ , de la correspondiente CMH.

La ecuación de balance global 2.36 dice que para todos los estados  $i$ ,

$$\pi(i) = \sum_{j \in E} \pi(j) p_{ji} \quad (2.37)$$

Iterando 2.37 se tiene que  $\pi^t = \pi^t \mathbf{P}^n$  para toda  $n \geq 0$ , y por lo tanto, usando 2.7, si la distribución inicial  $\nu = \pi$ , entonces  $\nu_n = \pi$  para toda  $n \geq 0$ . Así, si una cadena es iniciada con una distribución estacionaria, se conserva la misma distribución para siempre. Y también,

$$P(X_n = i_0, X_{n+1} = i_1, \dots, X_{n+k} = i_k) = \pi(i_0) p_{i_0 i_1} \cdots p_{i_{k-1} i_k} \quad (2.38)$$

no depende de  $n$ . En este sentido la cadena es *estacionaria*.

**Teorema 2.5.1.** Una cadena que inicia con una distribución estacionaria es estacionaria.

La ecuación  $\pi^t = \pi^t \mathbf{P}$ , junto con el requerimiento de que  $\pi$  sea un vector de probabilidad, i.e.,  $\pi^t \mathbf{1} = 1$  (donde  $\mathbf{1}$  es un vector columna con todas sus entradas igual a 1), constituye cuando  $E$  es finito,  $|E|+1$  ecuaciones para  $|E|$  variables desconocidas. Se puede observar que sumando todas las ecuaciones de  $\pi^t = \pi^t \mathbf{P}$  se obtiene la ecuación  $\pi^t \mathbf{1} = \pi^t \mathbf{P} \mathbf{1}$ , esto es,  $\pi^t \mathbf{1} = 1$ .

### 2.5.2. Ejemplos

**Ejemplo 2.5.1.** Tomando  $E = \{1, 2\}$  y definiendo la matriz de transición

$$\mathbf{P} = \begin{matrix} & \begin{matrix} 1 & 2 \end{matrix} \\ \begin{matrix} 1 \\ 2 \end{matrix} & \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix} \end{matrix}$$

donde  $\alpha, \beta \in (0, 1)$ . Las ecuaciones de balance global son

$$\begin{aligned} \pi(1) &= \pi(1)(1 - \alpha) + \pi(2)\beta, \\ \pi(2) &= \pi(1)\alpha + \pi(2)(1 - \beta). \end{aligned}$$

Este es un sistema dependiente el cual se reduce a la ecuación  $\pi(1)\alpha = \pi(2)\beta$ , usando  $\pi(1) + \pi(2) = 1$ . Se obtiene

$$\pi(1) = \frac{\beta}{\alpha + \beta}, \quad \pi(2) = \frac{\alpha}{\alpha + \beta}$$

**Ejemplo 2.5.2.** La CMH descrita en el ejemplo 2.2.6 tiene como ecuaciones de balance global para  $i \in [1, N - 1]$  a

$$\pi(i) = \pi(i - 1) \left(1 - \frac{i - 1}{N}\right) + \pi(i + 1) \frac{i + 1}{N}$$

y, para los estados de los extremos,

$$\pi(0) = \pi(1) \frac{1}{N}, \quad \pi(N) = \pi(N - 1) \frac{1}{N}$$

Dejando  $\pi(0)$  indeterminado, se pueden resolver las ecuaciones de balance para  $i = 0, 1, \dots, N$  sucesivamente, para obtener

$$\pi(i) = \pi(0) \binom{N}{i}$$

El valor de  $\pi(0)$  está entonces determinado usando el hecho de que  $\pi$  es un vector de probabilidad:

$$1 = \sum_{i=0}^N \pi(i) = \pi(0) \sum_{i=0}^N \binom{N}{i} = \pi(0) 2^N$$

Esto significa que  $\pi$  es la distribución binomial de tamaño  $N$  y parámetro  $\frac{1}{2}$ :

$$\pi(i) = \frac{1}{2^N} \binom{N}{i} \tag{2.39}$$

Esta es la distribución que se obtendría colocando independientemente cada partícula en los compartimentos, con probabilidad  $\frac{1}{2}$  para cada compartimento.

**Ejemplo 2.5.3.** Una caminata aleatoria simétrica sobre  $\mathbb{Z}$  no puede tener una distribución estacionaria. De hecho, la solución de la ecuación de balance

$$\pi(i) = \frac{1}{2} \pi(i - 1) + \frac{1}{2} \pi(i + 1)$$

para  $i \geq 0$ , con la información inicial  $\pi(0)$  y  $\pi(1)$ , es

$$\pi(i) = \pi(0) + (\pi(1) - \pi(0))i$$

Como  $\pi(i) \in [0, 1]$ , necesariamente  $\pi(1) - \pi(0) = 0$ . Por lo tanto,  $\pi(i)$  es una constante, necesariamente 0 ya que la suma de todos los valores  $\pi(i)$  tiene que ser finita. Así para toda  $i \geq 0$ , y por lo tanto, en vista de la ecuación de balance global, para toda  $i$ ,  $\pi(i) = 0$ , una contradicción si se quiere pensar a  $\pi$  como una distribución de probabilidad.

**Ejemplo 2.5.4.** Si se observa a la matriz identidad como matriz de transición. Entonces cualquier distribución de probabilidad en el espacio de estados es una distribución estacionaria.

Las ecuaciones de recurrencia pueden ser usadas para obtener la distribución estacionaria cuando la última existe y es única. Las funciones generadoras algunas veces también son de ayuda.

**Ejemplo 2.5.5.** Continuando con el ejemplo 2.2.2. Para cualquier número complejo  $z$  con módulo no mayor a 1, se sigue de la ecuación de recurrencia 2.13 que

$$\begin{aligned} z^{X_{n+1}+1} &= (z^{(X_n-1)^++1})z^{Z_{n+1}} \\ &= (z^{X_n} \mathbb{1}_{\{X_n>0\}} + z \mathbb{1}_{\{X_n=0\}})z^{Z_{n+1}} \\ &= (z^{X_n} - \mathbb{1}_{\{X_n=0\}} + z \mathbb{1}_{\{X_n=0\}})z^{Z_{n+1}} \end{aligned}$$

y por lo tanto

$$z z^{X_{n+1}} - z^{X_n} z^{Z_{n+1}} = (z-1) \mathbb{1}_{\{X_n=0\}} z^{Z_{n+1}}$$

De la independencia de  $X_n$  y  $Z_{n+1}$ ,  $E[z^{X_n} z^{Z_{n+1}}] = E[z^{X_n}] g_Z(z)$ , donde  $g_Z(z)$  es la función generadora de  $Z_{n+1}$ , y  $E[\mathbb{1}_{\{X_n=0\}} z^{Z_{n+1}}] = \pi(0) g_Z(z)$ , con  $\pi(0) = P(X_n = 0)$ .

Por lo tanto,

$$z E[z^{X_{n+1}}] - g_Z(z) E[z^{X_n}] = (z-1) \pi(0) g_Z(z)$$

En estado estable,  $E[z^{X_{n+1}}] = E[z^{X_n}] = g_X(z)$ , y entonces

$$g_X(z)(z - g_Z(z)) = \pi(0)(z-1)g_Z(z) \quad (2.40)$$

Esto da como resultado la función generadora  $g_X(z) = \sum_{i=0}^{\infty} \pi(i) z^i$ , siempre que  $\pi(0)$  esté definida. Para obtener  $\pi(0)$ , se diferencia 2.40:

$$g'_X(z)(z - g_Z(z)) + g_X(z)(1 - g'_Z(z)) = \pi(0)(g_Z(z) + (z-1)g'_Z(z)),$$

y sea  $z = 1$ , para obtener, tomando en cuenta las ecuaciones  $g_X(1) = g_Z(1) = 1$  y  $g'_Z(1) = E[Z]$ ,

$$\pi(0) = 1 - E[Z]. \quad (2.41)$$

Ya que  $\pi(0)$  debe ser no negativa, esto genera inmediatamente la condición necesaria  $E[Z] \leq 1$ . De hecho, se tiene, si el caso trivial  $Z_{n+1} \equiv 1$  se excluye,

$$E[Z] < 1. \quad (2.42)$$

Ahora, si  $E[Z] = 1$ , implica que  $\pi(0) = 0$ , se sigue de 2.40 que

$$g_X(x)(x - g_Z(x)) = 0$$

para toda  $x \in [0, 1]$ . Pero si se excluye el caso  $Z_{n+1} \equiv 1$  (esto es,  $g_Z(x) \equiv x$ ), la ecuación  $x - g_Z(x) = 0$  tiene solo a  $x = 1$  como solución cuando  $g'_Z(1) = E[Z] \leq 1$ . Por lo tanto,  $g_X(x) \equiv 0$  para toda  $x \in [0, 1)$ , y consecuentemente  $g_X(z) \equiv 0$  sobre  $\{|z| < 1\}$ . Esto nos guía a una contradicción, ya que la función generadora de una variable aleatoria con valores enteros no puede ser nula.

Siendo por el momento, de 2.40 y 2.41 se sigue que, si la distribución estacionaria existe, entonces su función generadora está dada por la formula

$$\sum_{i=0}^{\infty} \pi(i) z^i = (1 - E[Z]) \frac{(z-1)g_Z(z)}{z - g_Z(z)} \quad (2.43)$$





**Ejemplo 2.5.7.** Si el modelo anterior se considera para el espacio de estados  $E = \mathbb{N}$ . Los mismos cálculos como arriba, guían a la expresión 2.44 para la solución general de  $\pi^t \mathbf{P} = \pi^t$ , la cual depende de la condición inicial  $\pi(0)$ . Para que esta solución sea una medida de probabilidad, se necesita tener  $\pi(0) > 0$ . También, escribiendo  $\sum_{i=1}^{\infty} \pi(i) = 1$ ,

$$\pi(0) \left\{ 1 + \frac{1}{q_1} + \sum_{j=1}^{\infty} \frac{p_1 p_2 \cdots p_j}{q_1 q_2 \cdots q_{j+1}} \right\} = 1 \quad (2.46)$$

Así una distribución estacionaria existe si y sólo si

$$\sum_{j=1}^{\infty} \frac{p_1 p_2 \cdots p_j}{q_1 q_2 \cdots q_{j+1}} < \infty. \quad (2.47)$$

En este caso  $\pi(i)$  está dada por la expresión 2.44, donde  $\pi(0)$  está determinada por 2.46.

## 2.6. Tiempo Reversible

### 2.6.1. Cadena Reversible

Sea  $\{X_n\}_{n \geq 0}$  una CMH con matriz de transición  $\mathbf{P}$  y que admite una distribución estacionaria  $\pi$  tal que

$$\pi(i) > 0 \quad (2.48)$$

para todos los estados  $i$ . Se define la matriz  $\mathbf{Q}$ , indexada por  $E$ , por

$$\pi(i)q_{ij} = \pi(j)p_{ji}. \quad (2.49)$$

Esta matriz es estocástica, ya que

$$\sum_{j \in E} q_{ij} = \sum_{j \in E} \frac{\pi(j)}{\pi(i)} p_{ji} = \frac{1}{\pi(i)} \sum_{j \in E} \pi(j) p_{ji} = \frac{\pi(i)}{\pi(i)} = 1,$$

donde en la tercer igualdad se usan las ecuaciones de balance. Su interpretación es la siguiente: Suponga que la distribución inicial de  $\{X_n\}$  es  $\pi$ , en tal caso para toda  $n \geq 0$ , toda  $i \in E$ ,

$$P(X_n = i) = \pi(i). \quad (2.50)$$

Entonces, de 1.15,

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

esto es, en vista de 2.49 y 2.50,

$$P(X_n = j | X_{n+1} = i) = q_{ji}. \quad (2.51)$$

Se puede observar que  $\mathbf{Q}$  es la matriz de transición de la cadena inicial cuando el tiempo se revierte.

La siguiente es una observación muy simple que se promoverá al rango de teorema en vista de su utilidad y también para facilitar la referencia.

**Teorema 2.6.1.** Sea  $\mathbf{P}$  una matriz estocástica indexada por un conjunto numerable  $E$ , y sea  $\pi$  una distribución de probabilidad sobre  $E$ . Sea  $\mathbf{Q}$  una matriz estocástica indexada por  $E$  tal que para toda  $i, j \in E$ ,

$$\pi(i)q_{ij} = \pi(j)p_{ji}. \quad (2.52)$$

Entonces  $\pi$  es una distribución estacionaria de  $\mathbf{P}$ .

**Demostración.** Para una  $i \in E$  fija, se suman las ecuaciones 2.52 respecto de  $j \in E$  para obtener

$$\sum_{j \in E} \pi(i)q_{ij} = \sum_{j \in E} \pi(j)p_{ji}.$$

Como el lado izquierdo es igual a  $\pi(i) \sum_{j \in E} q_{ij} = \pi(i)$ , por lo tanto, para toda  $i \in E$ ,

$$\pi(i) = \sum_{j \in E} \pi(j)p_{ji}.$$

La reversión del tiempo en las cadenas también puede ser usada para extender a tiempos negativos una cadena  $\{X_n\}_{n \geq 0}$  con su distribución estacionaria  $\pi$  tal que  $\pi(i) > 0$  para toda  $i \in E$ .

## 2.6.2. Reversibilidad del tiempo

**Definición 2.6.1.** Se le llama *reversible* a una cadena de Markov estacionaria con distribución inicial  $\pi$  (una distribución estacionaria) asumida positiva si para toda  $i, j \in E$ ,

$$\pi(i)p_{ij} = \pi(j)p_{ji} \quad (2.53)$$

En este caso,  $q_{ij} = P_{ji}$ , y por lo tanto la cadena y la cadena a tiempo revertido son en cierto modo la misma, ya que la distribución de una CMH está enteramente determinada por su distribución inicial y su matriz de transición 2.1.1. Las ecuaciones 2.53 son llamadas ecuaciones de balance *detalladas*.

**Corolario 2.6.1.1.** Sea  $\mathbf{P}$  una matriz de transición sobre el espacio numerable  $E$ , y sea  $\pi$  alguna distribución de probabilidad sobre  $E$ . Si para toda  $i, j \in E$ , las ecuaciones 2.53 se satisfacen, entonces  $\pi$  es una distribución estacionaria de  $\mathbf{P}$ .

**Ejemplo 2.6.1.** Continuando los ejemplos 2.2.6 y 2.5.6. Se había obtenido la expresión

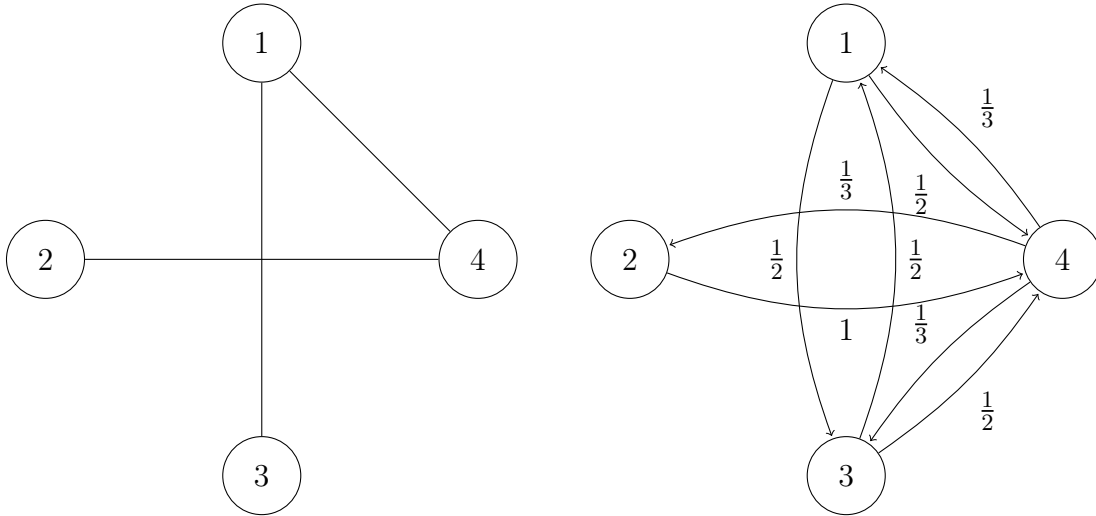
$$\pi(i) = \frac{1}{2^N} \binom{N}{i}$$

para la distribución estacionaria. Sus ecuaciones de balance detalladas serían entonces:

$$\pi(i)p_{i,i+1} = \pi(i+1)p_{i+1,i}.$$

**Ejemplo 2.6.2.** En una forma abstracta, el ejemplo 2.3.2 se puede generalizar. Este ejemplo se conoce como *caminata aleatoria sobre una gráfica*. Se considera una gráfica no orientada y se tomará a  $E$  como el conjunto de vértices, o nodos, de esta

gráfica. Sea  $d_i$  el número de aristas «adyacentes» al nodo  $i$ . Transformando esta gráfica en una gráfica orientada dividiendo cada arista en dos aristas orientadas en opuestas direcciones, y convirtiendo esta en una gráfica de transición asociando a la arista orientada de  $i$  a  $j$  la probabilidad de transición  $\frac{1}{d_i}$ .



Se asumirá que  $d_i > 0$  para todos los estados  $i$ . Una distribución estacionaria (de hecho, la distribución estacionaria) está dada por

$$\pi(i) = \frac{d_i}{\sum_{j \in E} d_j}$$

Para esto, se puede usar el corolario 2.6.1.1, haciendo la suposición de que la cadena es reversible. Se puede observar que

$$\pi(i) \frac{1}{d_i} = \pi(j) \frac{1}{d_j}.$$

## 2.7. Regeneración

### 2.7.1. Propiedad Fuerte de Markov

**Definición 2.7.1.** Un *tiempo de paro* respecto de un proceso estocástico  $\{X_n\}_{n \geq 0}$  es, por definición, una variable aleatoria  $\tau$  que toma sus valores en  $\mathbb{N} \cup \{\infty\}$  y tal que para todos los enteros  $m \geq 0$ , el evento  $\{\tau = m\}$  puede ser expresado en términos de  $X_0, X_1, \dots, X_m$ .

La última propiedad se simboliza con la notación

$$\{\tau = m\} \in X_0^m. \quad (2.54)$$

Cuando el espacio es numerable, 2.54 significa que

$$\mathbb{1}_{\{\tau=m\}} = \psi_m(X_0, \dots, X_m),$$

para alguna función  $\psi_m$  con valores en  $\{0, 1\}$ .

**Ejemplo 2.7.1.** En la teoría de las cadenas de Markov, un típico y más importante tiempo de paro es el *tiempo de retorno* al estado  $i \in E$ ,

$$T_i = \inf\{n \geq 1 : X_n = i\}, \quad (2.55)$$

donde  $T_i = \infty$  si  $X_n \neq i$  para toda  $n \geq 1$ . Este es un tiempo de paro.

Se puede observar que  $T_i \geq 1$ , y en particular,  $X_0 = i$  no implica que  $T_i = 0$ . Esto es por lo que  $T_i$  es llamado tiempo de retorno a  $i$ , y no el tiempo de llegada a  $i$ . Por último sea  $S_i = T_i$  si  $X_0 \neq i$ , y  $S_i = 0$  si  $X_0 = i$ . Este también es un tiempo de paro.

**Ejemplo 2.7.2.** Una constante es un tiempo de paro.

**Ejemplo 2.7.3.** Si  $\tau$  es un tiempo de paro y  $n_0$  un tiempo determinista no negativo, entonces  $\tau + n_0$  es un tiempo de paro. De hecho,  $\{\tau + n_0 = m\} \equiv \{\tau = m - n_0\}$  es expresable en términos de  $X_0, X_1, \dots, X_{m-n_0}$ .

Para un tiempo de paro  $\tau$  dado, se puede ver si  $\tau = m$  solo observando  $X_0, X_1, \dots, X_m$ . Esto es por lo que los tiempos de paro se dice que son no anticipados.

**Ejemplo 2.7.4.** El tiempo aleatorio

$$\tau = \inf\{n \geq 0 : X_{n+1} = i\},$$

donde  $\tau = \infty$  si  $X_{n+1} \neq i$  para toda  $n \geq 0$ , es anticipado porque

$$\{\tau = m\} = \{X_1 \neq i, \dots, X_m \neq i, X_{m+1} = i\}$$

para toda  $m \geq 0$ . El conocimiento de este tiempo provee información acerca del valor del proceso solo después de este. Este no es un tiempo de paro.

**Ejemplo 2.7.5.** Sean  $\tau_1 = T_i, \tau_2, \dots$  tiempos de retorno sucesivos al estado  $i$ . Si existen solo  $r$  retornos al estado  $i$ , sea  $\tau_{r+1} = \tau_{r+2} = \dots = \infty$ . Estos tiempos aleatorios son tiempos de paro con respecto de  $\{X_n\}_{n \geq 0}$ , ya que para cualquier  $m \geq 1$ ,

$$\{\tau_k = m\} = \left\{ \sum_{n=1}^m \mathbb{1}_{\{X_n=i\}} = k, X_m = i \right\}$$

es expresable en términos de  $X_0, \dots, X_m$ .

Sea  $\tau$  un tiempo aleatorio que toma valores en  $\mathbb{N} \cup \{\infty\}$ , y sea  $\{X_n\}_{n \geq 0}$  un proceso estocástico con valores en el conjunto numerable  $E$ . Para definir  $X_\tau$  cuando  $\tau = \infty$ , se tiene que decidir como definir  $X_\infty$ . Esto se hace tomando algún elemento arbitrario  $\Delta$  que no esté en  $E$  y haciendo

$$X_\infty = \Delta.$$

Por definición, «el proceso  $\{X_n\}$  después de  $\tau$ » es el proceso estocástico

$$\{X_{n+\tau}\}_{n \geq 0}.$$

El «proceso  $\{X_n\}$  después de  $\tau$ » es el proceso estocástico

$$\{X_{n \wedge \tau}\}_{n \geq 0}$$

el cual se congela al tiempo  $\tau$  en el valor  $X_\tau$ .

El resultado principal de esta sección a grandes rasgos dice que la propiedad de Markov, es decir, la independencia del pasado y futuro dado el estado presente, se extiende a la situación donde el tiempo actual es un tiempo de paro. Precisamente:

**Teorema 2.7.1** (Propiedad Fuerte de Markov). Sea  $\{X_n\}_{n \geq 0}$  una CMH con espacio de estados numerable  $E$  y matriz de transición  $\mathbf{P}$ . Sea  $\tau$  un tiempo de paro con respecto de esta cadena. Entonces para cualquier estado  $i \in E$ , dado que  $X_\tau = i$  (en particular,  $\tau < \infty$ , ya que  $i \neq \Delta$ ), lo siguiente se cumple:

( $\alpha$ ) El proceso después de  $\tau$  y el proceso antes de  $\tau$  son independientes.

( $\beta$ ) El proceso después de  $\tau$  es una CMH con matriz de transición  $\mathbf{P}$ .

**Demostración.** ( $\alpha$ ) Se debe mostrar que para todos los tiempos  $k \geq 1, n \geq 0$ , y todos los estados  $i_0, \dots, i_n, i, j_1, \dots, j_k$ ,

$$\begin{aligned} P(X_{\tau+1} = j_1, \dots, X_{\tau+k} = j_k | X_\tau = i, X_{\tau \wedge 0} = i_0, \dots, X_{\tau \wedge n} = i_n) \\ = P(X_{\tau+1} = j_1, \dots, X_{\tau+k} = j_k | X_\tau = i). \end{aligned} \quad (2.56)$$

Una versión simplificada de la ecuación anterior es

$$P(X_{\tau+k} = j | X_\tau = i, X_{\tau \wedge n} = i_n) = P(X_{\tau+k} = j | X_\tau = i). \quad (2.57)$$

El caso general se obtiene con argumentos similares. El lado izquierdo de la ecuación 2.57 es igual a

$$\frac{P(X_{\tau+k} = j, X_\tau = i, X_{\tau \wedge n} = i_n)}{P(X_\tau = i, X_{\tau \wedge n} = i_n)}.$$

El numerador puede ser desarrollado como

$$\sum_{r \geq 0} P(\tau = r, X_{r+k} = j, X_r = i, X_{r \wedge n} = i_n). \quad (2.58)$$

Y cada elemento de la suma se puede ver como

$$\begin{aligned} P(\tau = r, X_{r+k} = j, X_r = i, X_{r \wedge n} = i_n) = \\ P(X_{r+k} = j | X_r = i, X_{r \wedge n} = i_n, \tau = r) P(\tau = r, X_{r \wedge n} = i_n, X_r = i), \end{aligned}$$

y ya que  $r \wedge n \leq r$  y  $\{\tau = r\} \in X_0^r$ , el evento  $B = \{X_{r \wedge n} = i_n, \tau = r\}$  está en  $X_0^r$ . Por lo tanto, por la propiedad de Markov,

$$P(X_{r+k} = j | X_r = i, X_{r \wedge n} = i_n, \tau = r) = P(X_{r+k} = j | X_r = i) = p_{ij}(k).$$

Finalmente, la expresión 2.58 se reduce a

$$\sum_{r \geq 0} p_{ij}(k) P(\tau = r, X_{r \wedge n} = i_n, X_r = i) = p_{ij}(k) P(X_{\tau=i}, X_{\tau \wedge n} = i_n).$$

Por lo tanto, el lado izquierdo de 2.57 es  $p_{ij}(k)$ . Cálculos similares muestran que el lado derecho de 2.57 es también  $p_{ij}(k)$ , y así  $(\alpha)$  queda probado.

$(\beta)$  Se debe mostrar que para todos los estados  $i, j, k, i_{n-1}, \dots, i_1$ ,

$$\begin{aligned} P(X_{\tau+n+1} = k | X_{\tau+n} = j, X_{\tau+n-1} = i_{n-1}, \dots, X_{\tau} = i) \\ = P(X_{\tau+n+1} = k | X_{\tau+n} = j) = p_{jk}. \end{aligned}$$

Pero la primera parte de la ecuación se sigue de  $(\alpha)$  para el tiempo de paro  $\tau' = \tau + n$ , el proceso antes y después de  $\tau'$  son independientes dado  $X_{\tau'} = j$ . La segunda parte de la ecuación se sigue de cálculos similares tal y como en  $(\alpha)$ .

## 2.7.2. Ciclos Regenerativos

Sea el número de visitas al estado  $i$  estrictamente después del tiempo 0 denotado por

$$N_i = \sum_{n \geq 1} \mathbb{1}_{\{X_n = i\}}. \quad (2.59)$$

**Teorema 2.7.2.** La distribución de  $N_i$  dado  $X_0 = j$  es

$$P_j(N_i = r) = \begin{cases} f_{ji} f_{ii}^{r-1} (1 - f_{ii}) & \text{para } r \geq 1, \\ 1 - f_{ji} & \text{para } r = 0 \end{cases} \quad (2.60)$$

donde

$$f_{ji} = P_j(T_i < \infty) \quad (2.61)$$

y  $T_i$  es el tiempo de retorno a  $i$ .

**Demostración.** Para  $r = 0$ , es solo la definición de  $f_{ji}$ . Ahora sea  $r \geq 1$ , y se asume que 2.60 es cierto para toda  $k \in [1, r]$ . En particular,

$$P_j(N_i > r) = 1 - \sum_{k=0}^r P_j(N_i = k) = f_{ji} f_{ii}^r.$$

Denotando por  $\tau_r$  el  $r$ -ésimo tiempo de retorno al estado  $i$ ,

$$\begin{aligned} P_j(N_i = r + 1) &= P_j(N_i = r + 1, X_{\tau_{r+1}} = i) \\ &= P_j(\tau_{r+2} - \tau_{r+1} = \infty, X_{\tau_{r+1}} = i) \\ &= P_j(\tau_{r+2} - \tau_{r+1} = \infty | X_{\tau_{r+1}} = i) P(X_{\tau_{r+1}} = i) \end{aligned}$$

Pero

$$\begin{aligned} P_j(\tau_{r+2} - \tau_{r+1} = \infty | X_{\tau_{r+1}} = i) &= P(\tau_{r+2} - \tau_{r+1} = \infty | X_{\tau_{r+1}} = i, X_0 = j) \\ &= P(\tau_{r+2} - \tau_{r+1} = \infty | X_{\tau_{r+1}} = i) \end{aligned}$$

por la propiedad fuerte de Markov. Ya que  $\tau_{r+2} - \tau_{r+1}$  es el tiempo de retorno a  $i$  del proceso después de  $\tau_{r+1}$ , la propiedad fuerte de Markov da como resultado

$$P(\tau_{r+2} - \tau_{r+1} = \infty | X_{\tau_{r+1}} = i) = P(T_i = \infty | X_0 = i).$$

También,

$$P_j(X_{\tau_{r+1}} = i) = P_j(N_i > r)$$

(si  $N_i \leq r$ , entonces  $X_{\tau_{r+1}} = X_\infty = \Delta \notin E$ ). Por lo tanto,

$$P_j(N_i = r + 1) = P_i(T_i = \infty)P_j(N_i > r) = (1 - f_{ii})f_{ji}f_{ii}^r.$$

El resultado entonces está probado por inducción.

La distribución de  $N_i$  dado  $X_0 = j$  y dado  $N_i \geq 1$  es geométrica. Esto tiene dos consecuencias principales. Primeramente,

$$P_i(T_i < \infty) = 1 \iff P_i(N_i = \infty) = 1$$

En palabras: si se comienza desde el estado  $i$  es casi seguro que se retorne a  $i$ , entonces se visitará  $i$  infinitamente seguido. Segundo, se tiene

$$E_i[N_i] = \sum_{r=1}^{\infty} r P_i(N_i = r) = \sum_{r=1}^{\infty} r f_{ii}^r (1 - f_{ii}) = \frac{f_{ii}}{1 - f_{ii}}.$$

En particular,

$$P_i(T_i < \infty) < 1 \iff E_i[N_i] < \infty.$$

**Teorema 2.7.3.** Para cualquier estado  $i \in E$ ,

$$P_i(T_i < \infty) = 1 \iff P_i(N_i = \infty) = 1, \quad (2.62)$$

y

$$P_i(T_i < \infty) < 1 \iff P_i(N_i = \infty) = 0 \iff E_i[N_i] < \infty. \quad (2.63)$$

En particular, el evento  $\{N_i = \infty\}$  tiene  $P_i$ -probabilidad 0 ó 1.

Si se considera una cadena de Markov con un estado convencional denotado por 0 tal que  $P_0(T_0 < \infty) = 1$ . En vista del teorema anterior, la cadena empezando del estado 0 retornará infinitamente seguido a este estado. Sea  $\tau_1 = T_0, \tau_2, \dots$  los tiempos de retorno sucesivos a 0, y sea  $\tau_0 \equiv 0$ .

Por la propiedad fuerte de Markov, para cualquier  $k \geq 1$ , el proceso después de  $\tau_k$  es independiente del proceso antes de  $\tau_k$  (observar que la condición  $X_{\tau_k} = 0$  siempre se satisface), y el proceso después de  $\tau_k$  es una cadena de Markov con la misma matriz de transición que la matriz original, y con estado inicial 0, por construcción. Por lo tanto, se tiene lo siguiente

**Teorema 2.7.4** (Teorema del ciclo regenerativo). Sea  $\{X_n\}_{n \geq 0}$  una CMH con estado inicial 0 que es casi seguramente visitado infinitamente seguido. Denotando por  $\tau_0, \tau_1, \tau_2, \dots$  los tiempos de visita sucesivos a 0, las piezas de la trayectoria

$$\{X_{\tau_k}, X_{\tau_k+1}, \dots, X_{\tau_{k+1}-1}\}, \quad k \geq 0,$$

son independientes e idénticamente distribuidas.

Tales piezas son conocidas como *ciclos regenerativos* de la cadena entre las visitas al estado 0. Cada tiempo aleatorio  $\tau_k$  es un *tiempo regenerativo*, en el sentido de que  $\{X_{\tau_k+n}\}_{n \geq 0}$  es independiente del pasado  $X_0, \dots, X_{\tau_k-1}$  y tiene la misma distribución que  $\{X_n\}_{n \geq 0}$ . En particular, la sucesión  $\{\tau_k - \tau_{k-1}\}_{k \geq 1}$  es i.i.d.

**Ejemplo 2.7.6.** Sea  $\tau = T_0, \tau_2, \dots$  los tiempos de retorno sucesivos al estado 0 de la caminata aleatoria sobre  $\mathbb{Z}$  del ejemplo 2.2.1 con  $p = \frac{1}{2}$ . Se admitirá que  $P_0(T_0 < \infty) = 1$ , un hecho que se probará después, y se obtendrá la distribución de probabilidad de  $T_0$  dado  $X_0 = 0$ .

Se observa que para  $n \geq 1$ ,

$$P_0(X_{2n} = 0) = \sum_{k \geq 1} P_0(\tau_k = 2n),$$

y por lo tanto, para toda  $z \in \mathbb{C}$  tal que  $|z| < 1$ ,

$$\sum_{n \geq 1} P_0(X_{2n} = 0)z^{2n} = \sum_{k \geq 1} \sum_{n \geq 1} P_0(\tau_k = 2n)z^{2n} = \sum_{k \geq 1} E_0[z^{\tau_k}].$$

Como  $\tau_k = \tau_1 + (\tau_2 - \tau_1) + \dots + (\tau_k - \tau_{k-1})$  y por lo tanto, en vista del teorema 2.7.4 y ya que  $\tau_1 = T_0$ ,

$$E_0[z^{\tau_k}] = (E_0[z^{T_0}])^k.$$

En particular,

$$\sum_{n \geq 0} P_0(X_{2n} = 0)z^{2n} = \frac{1}{1 - E_0[z^{T_0}]}$$

(notar que la última suma incluye el término para  $n = 0$ , esto es, 1). Una evaluación directa del lado izquierdo da como resultado

$$\sum_{n \geq 0} \frac{1}{2^{2n}} \frac{(2n)!}{n!n!} z^{2n} = \frac{1}{\sqrt{1 - z^2}}.$$

Por lo tanto, la función generadora del tiempo de retorno a 0 dado  $X_0 = 0$  es

$$E_0[z^{T_0}] = 1 - \sqrt{1 - z^2}.$$

Su primera derivada

$$\frac{z}{\sqrt{1 - z^2}}$$

tiende a  $\infty$  como  $z \rightarrow 1$  por debajo a través de valores reales. Por lo tanto, por el teorema de Abel [1],

$$E_0[T_0] = \infty.$$

Se puede observar que a pesar de que dado  $X_0 = 0$  y el tiempo de retorno es casi seguramente finito, tiene esperanza infinita.



---

---

# CAPÍTULO 3

---

## RECURRENCIA Y ERGODICIDAD

### 3.1. El criterio de la Matriz Potencia

#### 3.1.1. Estados Recurrentes y Transitorios

Se considera una cadena de Markov con valores en  $E = \mathbb{N}$ . Existe la posibilidad que para cualquier estado inicial  $i \in \mathbb{N}$  la cadena no visitará el estado  $i$  después de algún tiempo aleatorio finito. Esto es a menudo una característica indeseable. Por ejemplo, si la cadena cuenta el número de consumidores esperando en la fila de un mostrador, tal comportamiento implica que la fila de espera eventualmente irá más allá de los límites de instalación de espera. En este sentido, el sistema correspondiente es inestable.

La noción aceptable de estabilidad para una CMH irreducible es el de *recurrencia positiva*, cuando algún estado dado es visitado infinitamente seguido y cuando el tiempo promedio entre dos visitas sucesivas a este estado es finito. El problema principal es encontrar las suficientes, y tal vez necesarias, condiciones que garanticen la estabilidad. Se comienza con el criterio de la *matriz potencia* (condición necesaria y suficiente), la cual es de interés principalmente teórico. Condiciones adicionales, como el *teorema de Foster* será dado en el capítulo 5.

Por el momento, se presentan definiciones relevantes. Primero hay que recordar que  $T_i$  denota el tiempo de retorno al estado  $i$ .

**Definición 3.1.1.** Un estado  $i \in E$  se llama *recurrente* si

$$P_i(T_i < \infty) = 1, \quad (3.1)$$

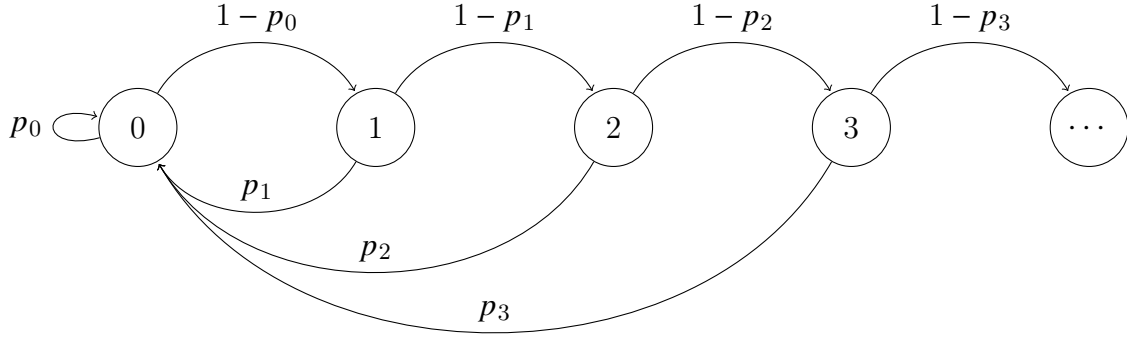
y en otro caso este estado es llamado *transitorio*. Un estado recurrente  $i \in E$  es llamado *positivo recurrente* si

$$E_i[T_i] < \infty \quad (3.2)$$

de otra manera se le conoce como recurrente *nulo*.

**Ejemplo 3.1.1.** La regla del juego es la siguiente: Una moneda se tira repetidamente, y cuando el resultado es cruz (probabilidad  $q = 1 - p$ ), se sube un escalón, pero

si el resultado es cara, caes hasta la parte más baja. Si  $X_n$  representa la posición al tiempo  $n$ ,  $\{X_n\}_{n \geq 0}$  forma una CMH con espacio de estados  $E = \mathbb{N}$  y en un caso especial de la cadena con la gráfica de transición siguiente



Se usará la condición de que

$$p_i \in (0, 1) \text{ para toda } i \in E,$$

para garantizar que la cadena sea irreducible. Se calculará la probabilidad de retorno al estado 0 y el tiempo medio de retorno al estado 0, y de esta expresión obtenida, se deducirá la naturaleza del estado 0.

Existe solo una manera de ir del estado 0 al estado 0 en exactamente  $n$  pasos. El camino correspondiente es  $0, 1, 2, \dots, n-1, 0$ , y por lo tanto  $P_0(T_0 = 1) = p_0$ , y para  $n \geq 1$ ,

$$P_0(T_0 = n) = (1 - p_0) \cdots (1 - p_{n-2}) p_{n-1}.$$

Definiendo  $u_0 = 1$ , y para  $n \geq 1$ ,

$$u_n = (1 - p_0) \cdots (1 - p_{n-1}),$$

se observa que

$$P_0(T_0 = n) = u_{n-1} - u_n.$$

Ya que

$$P_0(T_0 < \infty) = \sum_{n=1}^{\infty} P_0(T_0 = n) = \lim_{m \uparrow \infty} \sum_{n=1}^m P_0(T_0 = n) = \lim_{m \uparrow \infty} (1 - u_m),$$

se tiene

$$P_0(T_0 < \infty) = 1 - \lim_{m \uparrow \infty} \prod_{i=0}^{m-1} (1 - p_i).$$

Por lo tanto, en vista del clásico resultado sobre productos infinitos. [34]

$$P_0(T_0 < \infty) = 1 \iff \prod_{i=0}^{\infty} (1 - p_i) = 0 \iff \sum_{i=0}^{\infty} p_i = \infty.$$

En general, no es sencillo revisar si un estado es transitorio o recurrente. Uno de las metas de la teoría en cadenas de Markov es proveer un criterio de recurrencia.

El problema de encontrar condiciones de recurrencia útiles (fáciles de verificar) es un área activa de investigación. Sin embargo, la teoría tiene algunas condiciones que califican como útiles y son aplicables a muchas situaciones prácticas. Aunque el siguiente criterio es de interés teórico más que práctico, puede ser útil en algunas situaciones, por ejemplo, en el estudio de la recurrencia de caminatas aleatorias.

### 3.1.2. Matriz Potencia

La *matriz potencia*  $\mathbf{G}$  asociada con la matriz  $\mathbf{P}$  está definida por

$$\mathbf{G} = \sum_{n \geq 0} \mathbf{P}^n.$$

Su término general

$$g_{ij} = \sum_{n=0}^{\infty} p_{ij}(n) = \sum_{n=0}^{\infty} P_i(X_n = j) = \sum_{n=0}^{\infty} E_i[\mathbb{1}_{\{X_n=j\}}] = E_i \left[ \sum_{n=0}^{\infty} \mathbb{1}_{\{X_n=j\}} \right]$$

es el número promedio de visitas al estado  $j$ , dado que la cadena inicia en el estado  $i$ .

**Teorema 3.1.1** (Criterio de la Matriz Potencia). El estado  $i \in E$  es recurrente si y sólo si

$$\sum_{n=0}^{\infty} p_{ii}(n) = \infty. \quad (3.3)$$

**Demostración.** Se sigue del teorema 2.7.3.

**Ejemplo 3.1.2.** Se considera la cadena de Markov descrita en el ejemplo 2.2.1. Las entradas no nulas de su matriz de transición son

$$p_{i,i+1} = p, \quad p_{i,i-1} = 1 - p,$$

donde  $p \in (0, 1)$ . Se observará si el estado 0 es recurrente o transitorio. Se tiene que  $p_{00}(2n+1) = 0$  y

$$p_{00}(2n) = \frac{(2n)!}{n!n!} p^n (1-p)^n.$$

Por la fórmula de equivalencia de Stirling [11]  $n! \sim (n/e)^n \sqrt{2\pi n}$ , la cantidad de arriba es similar a

$$\frac{[4p(1-p)]^n}{\sqrt{\pi n}}, \quad (3.4)$$

y la naturaleza de la serie  $\sum_{n=0}^{\infty} p_{00}(n)$  (convergente o divergente) es el de la serie con término general 3.4. Si  $p \neq \frac{1}{2}$ , en tal caso  $4p(1-p) < 1$ , la última serie converge, y si  $p = \frac{1}{2}$ , en tal caso  $4p(1-p) = 1$ , esta diverge. En resumen, los estados de la caminata aleatoria son transitorios si  $p \neq \frac{1}{2}$ , y recurrentes si  $p = \frac{1}{2}$ .

El ejemplo 2.7.6 muestra que para la caminata aleatoria con  $p = \frac{1}{2}$ , en efecto los estados son recurrentes nulos.

**Ejemplo 3.1.3.** Ahora se considera una CMH con espacio de estados  $E = \mathbb{Z}$ . Denotando por  $e_1, e_2$  y  $e_3$  los vectores base canónicos de  $\mathbb{R}^3$ , y sean los términos no nulos de la matriz de transición

$$p_{x, x \pm e_i} = \frac{1}{6}$$

Denotando al vector  $(0, 0, 0)$  con simplemente 0 se cumple que  $p_{00}(2n+1) = 0$  para toda  $n \geq 0$ , y

$$p_{00}(2n) = \sum_{0 \leq i+j \leq n} \frac{(2n)!}{(i!j!(n-i-j)!)^2} \left(\frac{1}{6}\right)^{2n}$$

Esto puede reescribirse como

$$p_{00}(2n) = \sum_{0 \leq i+j \leq n} \frac{1}{2^{2n}} \binom{2n}{n} \left( \frac{n!}{i!j!(n-i-j)!} \right)^2 \left(\frac{1}{3}\right)^{2n}$$

Usando la formula trinomial

$$\sum_{0 \leq i+j \leq n} \frac{n!}{i!j!(n-i-j)!} \left(\frac{1}{3}\right)^n = 1,$$

se obtiene la cota

$$p_{00}(2n) \leq K_n \frac{1}{2^{2n}} \binom{2n}{n} \left(\frac{1}{3}\right)^n,$$

donde

$$K_n = \max_{0 \leq i+j \leq n} \frac{n!}{i!j!(n-i-j)!}$$

Para valores grandes de  $n$ ,  $K_n$  se acota como sigue. Sea  $i_0$  y  $j_0$  los valores de  $i, j$  que maximizan  $n!/(i!j!(n-i-j)!)$  en el dominio de interés  $0 \leq i+j \leq n$ . De la definición de  $i_0$  y  $j_0$ , las cantidades

$$\frac{n!}{(i_0-1)!j_0!(n-i_0-j_0+1)!}$$

$$\frac{n!}{(i_0+1)!j_0!(n-i_0-j_0+1)!}$$

$$\frac{n!}{i_0!(j_0-1)!(n-i_0-j_0+1)!}$$

$$\frac{n!}{i_0!(j_0+1)!(n-i_0-j_0+1)!}$$

están acotadas por

$$\frac{n!}{i_0!j_0!(n-i_0-j_0+1)!}$$

Las desigualdades correspondientes se reducen a

$$\begin{aligned} n - i_0 - 1 &\leq 2j_0 \leq n - i_0 + 1 \\ n - j_0 - 1 &\leq 2i_0 \leq n - j_0 + 1, \end{aligned}$$

y esto muestra que para  $n$  grande,  $i_0 \sim n/3$  y  $j_0 \sim n/3$ . Por lo tanto, para  $n$  grande,

$$p_{00}(2n) \sim \frac{n!}{(n/3)!(n/3)!2^{2n}e^n} \binom{2n}{n}.$$

Por la formula de equivalencia de Stirling [11], el lado derecho de la última equivalencia es similar a

$$\frac{3\sqrt{3}}{2(\pi n)^{3/2}},$$

el término general de una serie divergente. El estado 0 es por lo tanto transitorio.

Si el estado  $i \in E$  es recurrente, y accesible del estado  $j \in E$ . Esto es, empezando desde  $j$ , la probabilidad de visitar  $i$  al menos una vez es positiva (accesibilidad de  $i$  desde  $j$ ), y empezando desde  $i$ , el número promedio de visitas a  $i$  es infinito (recurrencia de  $i$ ). Por lo tanto, empezando desde  $j$  el número promedio de visitas a  $i$  es infinito:

$$E_j[N_i] = \sum_{n \geq 1} p_{ji}(n) = \infty.$$

Similarmente, si  $i$  es transitorio, entonces para cualquier estado  $j \in E$ ,

$$E_j[N_i] = \sum_{n \geq 1} p_{ji}(n) < \infty.$$

### 3.1.3. Estructura de la Matriz de Transición

Una aplicación teórica del criterio de la matriz potencia es la prueba de que la recurrencia es una propiedad de clase (comunicación).

**Teorema 3.1.2.** Si  $i$  y  $j$  se comunican, entonces ambos estados son transitorios o recurrentes.

**Demostración.** Por definición,  $i$  y  $j$  se comunican si y sólo si existen enteros  $M$  y  $N$  tales que  $p_{ij}(M) > 0$ ,  $p_{ji}(N) > 0$ . Ir de  $i$  a  $j$  en  $M$  pasos, de  $j$  a  $j$  en  $n$  pasos, y luego de  $j$  a  $i$  en  $N$  pasos, es solo una manera de empezar en  $i$  y volver a  $i$  en  $M + n + N$  pasos. Por lo tanto,

$$p_{ii}(M + n + N) \geq p_{ij}(M)p_{jj}(n)p_{ji}(N)$$

Similarmente

$$p_{jj}(N + n + M) \geq p_{ji}(N)p_{ii}(n)p_{ij}(M).$$

Por lo tanto, escribiendo  $\alpha = p_{ij}(M)p_{ji}(N)$  (una cantidad estrictamente positiva), se tiene que

$$p_{ii}(M + N + n) \geq \alpha p_{jj}(n) \text{ y } p_{jj}(M + N + n) \geq \alpha p_{ii}(n).$$

Esto implica que las series

$$\sum_{n=0}^{\infty} p_{ii}(n) \text{ y } \sum_{n=0}^{\infty} p_{jj}(n)$$

ambas convergen o ambas divergen. El teorema 3.1.1 concluye la demostración.

Será probado en un futuro que la recurrencia positiva (resp. recurrencia nula) es también una propiedad de clase, en el sentido que si los estados  $i$  y  $j$  se comunican y si uno de ellos es positivo recurrente (resp. recurrente nulo), entonces también el otro lo es.

Una cadena de Markov irreducible tiene por lo tanto todos sus estados de la misma naturaleza: transitorios, positivos recurrentes, o recurrentes nulos. Se podrá decir entonces que es una cadena transitoria, una cadena positiva recurrente, o una cadena recurrente nula, y para determinar a qué categoría pertenece es suficiente estudiar *un* estado, seleccionando el estado para el cual los cálculos lucen más sencillos (como el estado 0 en el ejemplo 3.1.1).

Se sigue de lo anterior que existen dos tipos de clases de comunicación: las *clases transitorias* y las *clases recurrentes*. Se llamará  $T$  al conjunto de los estados transitorios y  $R$  al conjunto de los estados recurrentes. El conjunto  $R$  podría estar compuesto de otras clases de comunicación disjuntas  $R_1, R_2$ , etc. Cualquier clase de comunicación recurrente,  $R_1$  por ejemplo, es cerrada. De hecho, si la cadena va desde  $i \in R_1$  a alguna  $j \in E$ , tendrá que volver a  $i$ , ya que  $i$  es recurrente, y por lo tanto  $i$  y  $j$  deben estar comunicados, entonces  $j$  debe estar en  $R_1$ .

## 3.2. Recurrencia y Medidas Invariantes

La noción de medida invariante juega un importante rol técnico en la teoría de la recurrencia de cadenas de Markov. Esta extiende la noción de distribución estacionaria.

**Definición 3.2.1.** Un vector no trivial (esto es, no nulo)  $x = \{x_i\}_{i \in E}$  es llamado *medida invariante* de la matriz estocástica  $\mathbf{P} = \{p_{ij}\}_{i,j \in E}$  si para toda  $i \in E$ ,

$$x_i \in [0, \infty) \quad (3.5)$$

y

$$x_i = \sum_{j \in E} x_j p_{ji}. \quad (3.6)$$

(En notación abreviada,  $0 \leq x < \infty$  y  $x^t \mathbf{P} = x^t$ .)

**Teorema 3.2.1.** Sea  $\mathbf{P}$  una matriz de transición de una CMH irreducible recurrente  $\{X_n\}_{n \geq 0}$ . Sea 0 un estado arbitrario y sea  $T_0$  el tiempo de retorno a 0. Se define para toda  $i \in E$

$$x_i = E_0 \left[ \sum_{n \geq 1} \mathbb{1}_{\{X_n=i\}} \mathbb{1}_{\{n \leq T_0\}} \right] \quad (3.7)$$

(Para  $i \neq 0$ ,  $x_i$  es el número esperado de visitas al estado  $i$  antes de retornar a 0). Entonces, para toda  $i \in E$ ,

$$x_i \in (0, \infty), \quad (3.8)$$

y  $x$  es una medida invariante de  $\mathbf{P}$ .

Observar que para  $n \in [1, T_0]$ ,  $X_n = 0$  si y sólo si  $n = T_0$ . Por lo tanto,

$$x_0 = 1 \quad (3.9)$$

También,

$$\begin{aligned} \sum_{i \in E} \sum_{n \geq 1} \mathbb{1}_{\{X_n=i\}} \mathbb{1}_{\{n \leq T_0\}} &= \sum_{n \geq 1} \left\{ \sum_{i \in E} \mathbb{1}_{\{X_n=i\}} \right\} \mathbb{1}_{\{n \leq T_0\}} \\ &= \sum_{n \geq 1} \mathbb{1}_{\{n \leq T_0\}} = T_0 \end{aligned}$$

y por lo tanto

$$\sum_{i \in E} x_i = E_0[T_0]. \quad (3.10)$$

Para la prueba del teorema, se define la cantidad

$$\begin{aligned} {}_0p_{0i}(n) &= E_0[\mathbb{1}_{\{X_n=i\}} \mathbb{1}_{\{n \leq T_0\}}] \\ &= P_0(X_1 \neq 0, \dots, X_{n-1} \neq 0, X_n = i). \end{aligned} \quad (3.11)$$

Esta es la probabilidad, empezando desde 0, de visitar  $i$  en el tiempo  $n$  antes de regresar a 0. De la definición de  $x$ ,

$$x_i = \sum_{n \geq 1} {}_0p_{0i}(n). \quad (3.12)$$

**Demostración.** Primero se observa que

$${}_0p_{0i}(1) = p_{0i} \quad (3.13)$$

y, usando el análisis del primer paso, para toda  $n \geq 2$ ,

$${}_0p_{0i}(n) = \sum_{j \neq 0} {}_0p_{0j}(n-1) p_{ji}. \quad (3.14)$$

Sumando para cada  $n \geq 1$ , y tomando en cuenta 3.12, se obtiene

$$x_i = p_{0i} + \sum_{j \neq 0} x_j p_{ji},$$

esto es, 3.6, ya que  $x_0 = 1$  por 3.9.

Ahora se mostrará que  $x_i > 0$  para toda  $i \in E$ . De hecho, iterando 3.6, se encuentra que  $x^t = x^t \mathbf{P}^n$ , y ya que  $x_0 = 1$ ,

$$x_i = \sum_{j \in E} x_j p_{ji}(n) = p_{0i}(n) + \sum_{j \neq 0} x_j p_{ji}(n).$$

Si  $x_i$  fuera nulo para alguna  $i \in E$ ,  $i \neq 0$ , la última ecuación podría implicar que  $p_{0i}(n) = 0$  para toda  $n \geq 0$ , lo cual significa que 0 e  $i$  no están comunicados, contradiciendo la suposición de que la cadena es irreducible.

Queda mostrar que  $x_i < \infty$  para toda  $i \in E$ . Como antes, se observa que

$$1 = x_0 = \sum_{j \in E} x_j p_{j0}(n)$$

para toda  $n \geq 1$ , y por lo tanto si  $x_i = \infty$  para alguna  $i$ , necesariamente  $p_{i0}(n) = 0$  para toda  $n \geq 1$ , y esto también contradice que la cadena sea irreducible.

**Teorema 3.2.2.** La medida invariante de una matriz estocástica irreducible recurrente es única.

En la prueba del teorema 3.2.1 se mostró que para una medida invariante  $y$  de una cadena irreducible,  $y_i > 0$  para toda  $i \in E$ , y por lo tanto, uno puede definir, para toda  $i, j \in E$ , la matriz  $\mathbf{Q}$  por

$$q_{ij} = \frac{y_i}{y_j} p_{ij}. \quad (3.15)$$

Esta es una matriz de transición, ya que

$$\sum_{i \in E} q_{ji} = \frac{1}{y_j} \sum_{i \in E} y_i p_{ij} = \frac{y_j}{y_j} = 1.$$

El término general de  $\mathbf{Q}^n$  es

$$q_{ji}(n) = \frac{y_i}{y_j} p_{ij}(n). \quad (3.16)$$

De hecho, suponiendo que 3.16 es cierta para  $n$ ,

$$\begin{aligned} q_{ji}(n+1) &= \sum_{k \in E} q_{jk} q_{ki}(n) = \sum_{k \in E} \frac{y_k}{y_j} p_{kj} \frac{y_i}{y_k} p_{ik}(n) \\ &= \frac{y_i}{y_j} \sum_{k \in E} p_{ik}(n) p_{kj} = \frac{y_i}{y_j} p_{ij}(n+1), \end{aligned}$$

y 3.16 se cumple, por inducción, para toda  $n \geq 1$ .

Claramente  $\mathbf{Q}$  es irreducible, ya que  $\mathbf{P}$  es irreducible (sólo basta observar que  $q_{ji}(n) > 0$  si y sólo si  $p_{ij}(n) > 0$  en vista de 3.16). También,  $p_{ii}(n) = q_{ii}(n)$ , y por lo tanto  $\sum_{n \geq 0} q_{ii}(n) = \sum_{n \geq 0} p_{ii}(n)$ , y esto asegura que  $\mathbf{Q}$  es recurrente por el criterio de la matriz potencia. Sea  $g_{ji}(n)$  la probabilidad, relativa a la cadena gobernada por la matriz de transición  $\mathbf{Q}$ , de retornar al estado  $i$  por primera vez en  $n$  pasos cuando se empieza en  $j$ . El análisis de primer paso da como resultado que  $g_{i0}(n+1) = \sum_{j \neq 0} q_{ij} g_{j0}(n)$ , esto es, usando 3.15,

$$y_i g_{i0}(n+1) = \sum_{j \neq 0} (y_j g_{j0}(n)) p_{ji}.$$

Renombrando a  ${}_0 p_{0i}(n+1) = \sum_{j \neq 0} {}_0 p_{0j}(n) p_{ji}$ , ó, equivalentemente,

$$y_0 {}_0 p_{0i}(n+1) = \sum_{j \neq 0} (y_0 {}_0 p_{0j}(n)) p_{ji}.$$

Por lo tanto se puede ver que las sucesiones  $\{y_0 {}_0 p_{0i}\}$  y  $\{y_i g_{i0}(n)\}$  satisfacen la misma ecuación de recurrencia. Sus primeros términos ( $n = 1$ ), respectivamente  $y_0 {}_0 p_{0i}(1) = y_0 p_{0i}$  y  $y_i g_{i0}(1) = y_i q_{i0}$ , son iguales en vista de 3.15. Por lo tanto, para toda  $n \geq 1$ ,

$${}_0 p_{0i}(n) = \frac{y_i}{y_0} g_{i0}(n).$$

Sumando con respecto de  $n \geq 1$  y usando  $\sum_{n \geq 1} g_{i0}(n) = 1$  ( $\mathbf{Q}$  es recurrente), se obtiene el resultado anunciado  $x_i = \frac{y_i}{y_0}$ .

La ecuación 3.10 y la definición de recurrencia positiva dan como resultado lo siguiente.



**Teorema 3.2.3.** Una CMH irreducible recurrente es positiva recurrente si y sólo si sus medidas invariantes  $x$  satisfacen que

$$\sum_{i \in E} x_i < \infty \quad (3.17)$$

Una CMH podría bien ser irreducible y poseer una medida invariante, y aún no ser recurrente. El ejemplo que muestra esto es la caminata aleatoria simétrica, la cual se mostró que es transitoria (ejemplo 2.7.6 y la cual admite  $x_i \equiv 1$  como medida invariante.)

### 3.3. Recurrencia Positiva

#### 3.3.1. Criterio de la Distribución Estacionaria

En la sección anterior, una cadena de Markov irreducible fue asumida recurrente, y se mostró que este tiene una distribución estacionaria única si esta es positiva recurrente. También se observó que la existencia de una medida invariante no es suficiente para la recurrencia. Resulta que la existencia de una distribución de probabilidad estacionaria es condición necesaria y suficiente para una cadena irreducible ser recurrente positiva.

**Teorema 3.3.1.** Una CMH irreducible es recurrente positiva si y sólo si existe una distribución estacionaria. Además, la distribución estacionaria  $\pi$  es, cuando esta existe, única, y  $\pi > 0$ .

**Demostración.** La parte directa se sigue de los teoremas 3.2.1 y 3.2.3. Para el regreso, se asume la existencia de una distribución estacionaria  $\pi$ . Iterando  $\pi^t = \pi^t \mathbf{P}$ , se obtiene  $\pi^t = \pi^t \mathbf{P}^n$ , esto es, para toda  $i \in E$ ,

$$\pi(i) = \sum_{j \in E} \pi(j) p_{ji}(n).$$

Si la cadena fuera transitoria, entonces, en vista del criterio de la matriz potencia 3.1.1 y la discusión siguiente, para todos los estados  $i, j$ ,

$$\lim_{n \uparrow \infty} p_{ji}(n) = 0,$$

y ya que  $p_{ji}(n)$  está acotada por 1 uniformemente en  $j$  y  $n$ , por el teorema de la convergencia dominada para series [31]

$$\pi(i) = \lim_{n \uparrow \infty} \sum_{j \in E} \pi(j) p_{ji}(n) = \sum_{j \in E} \pi(j) \left( \lim_{n \uparrow \infty} p_{ji}(n) \right) = 0.$$

Esto contradice la suposición de que  $\pi$  es una distribución estacionaria (en particular,  $\sum_{i \in E} \pi(i) = 1$ ). La cadena debe ser por lo tanto recurrente, y por el teorema 3.2.3, esta es positiva recurrente.

La distribución estacionaria  $\pi$  de una cadena irreducible recurrente positiva es única (aquí se usa el teorema 3.2.2). También recordando que  $\pi(i) > 0$  para toda  $i \in E$  (3.2.1).

**Teorema 3.3.2.** Sea  $\pi$  la única distribución estacionaria de una cadena irreducible positiva recurrente, y sea  $T_i$  el tiempo de retorno al estado  $i$ . Entonces

$$\pi(i)E_i[T_i] = 1. \quad (3.18)$$

**Demostración.** Esta ecuación es una consecuencia directa de la expresión 3.7 para la medida invariante. De hecho,  $\pi$  es obtenida por la normalización de  $x$ : para toda  $i \in E$ ,

$$\pi(i) = \frac{x_i}{\sum_{j \in E} x_j},$$

y en particular, para  $i = 0$ , usando 3.6 y 3.10,

$$\pi(0) = \frac{x_0}{\sum_{j \in E} x_j} = \frac{1}{E_0[T_0]}.$$

Ya que el estado 0 no juega un rol especial en el análisis, 3.18 es cierta para toda  $i \in E$ .

La situación es simple cuando el espacio de estados es finito.

**Teorema 3.3.3.** Una CMH irreducible con espacio de estados finito es recurrente positiva.

**Demostración.** Primero se probará la recurrencia. Si la cadena fuera transitoria, entonces, por el criterio de la matriz potencia y las observaciones siguientes, para toda  $i, j \in E$ ,

$$\sum_{n \geq 0} p_{ij}(n) < \infty,$$

y por lo tanto, ya que el espacio de estados es finito

$$\sum_{j \in E} \sum_{n \geq 0} p_{ij}(n) < \infty.$$

La última suma es igual a

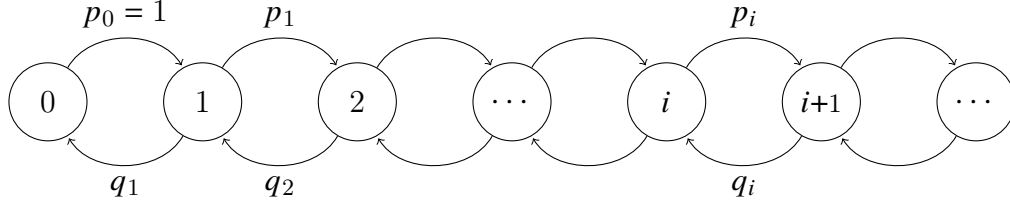
$$\sum_{n \geq 0} \sum_{j \in E} p_{ij}(n) = \sum_{n \geq 0} 1 = \infty,$$

una contradicción. Por lo tanto, la cadena es recurrente. Por el teorema 3.2.1 esta tiene una medida invariante  $x$ . Ya que  $E$  es finito,  $\sum_{i \in E} x_i < \infty$ , y por lo tanto la cadena es recurrente positiva por el teorema 3.3.1.

Una «prueba hablada» de la recurrencia en el teorema 3.3.3 es la siguiente: Los estados no pueden ser todos visitados sólo un número finito de veces; de lo contrario, podría existir un tiempo aleatorio finito después del cual no se visita ningún estado.

### 3.3.2. Ejemplos

**Ejemplo 3.3.1.** Se considera una cadena con espacio de estados  $E = \mathbb{N}$ . Se asume que  $p_i$  (y por lo tanto  $q_i = 1 - p_i$ ) está en el intervalo abierto  $(0, 1)$  para toda  $i \in E$ , y se considera la siguiente gráfica de transición para que entonces la cadena sea irreducible.



La ecuación para la medida invariante  $x^t = x^t \mathbf{P}$  toma valores de la forma

$$\begin{aligned} x_0 &= x_1 q_1, \\ x_i &= x_{i-1} p_{i-1} + x_{i+1} q_{i+1}, \quad i \geq 1, \end{aligned}$$

con  $p_0 = 1$ . La solución general es, para  $i \geq 1$ ,

$$x_i = x_0 \frac{p_0 \cdots p_{i-1}}{q_1 \cdots q_i}.$$

La condición para que la cadena sea recurrente positiva  $\sum_{i \in E} x_i < \infty$  es

$$1 + \sum_{i \geq 1} \frac{p_0 \cdots p_{i-1}}{q_1 \cdots q_i} < \infty, \quad (3.19)$$

y si esta se satisface, la distribución estacionaria  $\pi$  se obtiene normalizando la solución general. Esto da

$$\pi(0) = \left( 1 + \sum_{i \geq 1} \frac{p_0 \cdots p_{i-1}}{q_1 \cdots q_i} \right)^{-1} \quad (3.20)$$

y para  $i \geq 1$ ,

$$\pi(i) = \pi(0) \frac{p_0 \cdots p_{i-1}}{q_1 \cdots q_i}. \quad (3.21)$$

En el caso especial donde  $p_i = p, q_i = q = 1 - p$ , la condición para que la cadena sea recurrente positiva se convierte en

$$1 + \frac{1}{q} \sum_{j \geq 0} \left( \frac{p}{q} \right)^j < \infty,$$

es decir  $p < q$ , o equivalentemente,

$$p < \frac{1}{2}.$$

**Ejemplo 3.3.2.** El resultado del ejemplo 2.1.1 será utilizado una vez más, esta vez tomando en cuenta el criterio de la distribución estacionaria. Se hará  $q_i = 1 - p_i$ . La ecuación  $x^t = x^t \mathbb{P}$  toma la forma

$$x_0 = p_0 x_0 + p_1 x_1 + p_2 x_2 + \cdots$$

y para  $i \geq 1$ ,

$$x_i = q_{i-1} x_{i-1}.$$

Por lo tanto, dejando de lado la primer ecuación, para  $i \geq 1$ ,

$$x_i = (q_0 q_1 \cdots q_{i-1}) x_0.$$

Descartando la posibilidad de que  $x_0 < 0$ , la cual podría implicar que  $x$  es negativa o nula, la primera ecuación se satisface si y sólo si

$$1 = p_0 + q_0 p_1 + q_0 q_1 p_2 + \cdots$$

esto es, ya que  $q_0 q_1 \cdots q_{n-1} p_n = q_0 q_1 \cdots q_{n-1} - q_0 q_1 \cdots q_n$ ,

$$\prod_{i=0}^{\infty} q_i = 0. \quad (3.22)$$

Ya que  $q_i = 1 - p_i$  y  $p_i \in (0, 1)$ , el criterio de convergencia para productos infinitos [34] dice que esto es equivalente a

$$\sum_{i=0}^{\infty} p_i = \infty.$$

La divergencia de las series  $\sum_{i=0}^{\infty} p_i$  es por lo tanto una condición necesaria y suficiente de la existencia de una medida invariante.

Sin embargo, la existencia de una medida invariante no implica recurrencia. Pero la existencia de una distribución estacionaria si implica la recurrencia (y de hecho recurrencia positiva) por el criterio de la distribución estacionaria.

Bajo las condiciones 3.22, existe una medida invariante, y esta medida tiene masa finita  $\sum_{i=0}^{\infty} x_i < \infty$  si y sólo si

$$1 + \sum_{n=1}^{\infty} \left( \prod_{i=0}^{n-1} q_i \right) < \infty. \quad (3.23)$$

La distribución estacionaria está entonces dada por

$$\pi(0) = \left( 1 + \sum_{n=1}^{\infty} \left( \prod_{i=0}^{n-1} q_i \right) \right)^{-1} \quad (3.24)$$

y para  $i \geq 1$ ,

$$\pi(i) = \left( \prod_{j=0}^{i-1} q_j \right) \pi_0. \quad (3.25)$$

Si se tiene que

$$p_i = \frac{P(U = i + 1)}{P(U > i)}.$$

La desigualdad 3.23 implica que  $E[U] < \infty$ , y 3.3.2, 3.25 dan como resultado

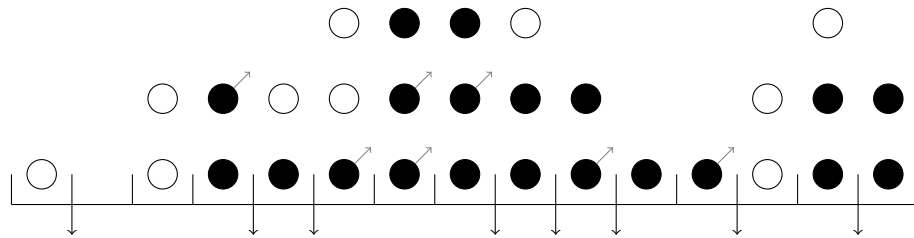
$$\pi(i) = \frac{P(U > i)}{E[U]}. \quad (3.26)$$

El criterio de la distribución estacionaria también puede ser usada para probar inestabilidad.

**Ejemplo 3.3.3** (Inestabilidad de ALOHA). Una situación típica en un sistema de comunicaciones por satélite de acceso múltiple es la siguiente. Los usuarios, cada uno identificado con un mensaje, compiten por el acceso a un enlace de comunicaciones por satélite de un solo canal con el fin de transmitir mensajes. Dos o más mensajes en el aire al mismo tiempo se atascan entre sí y no se transmiten correctamente. Los usuarios de alguna manera son capaces de detectar una colisión de este tipo e intentarán retransmitir más tarde el mensaje involucrado en una colisión. La dificultad de estos sistemas de comunicación reside principalmente en la falta de cooperación entre los usuarios, que desconocen la intención de transmisión de los usuarios competidores.

El protocolo ALOHA por slots impone a los usuarios las siguientes reglas:

- (i) Las transmisiones y retransmisiones de mensajes solo pueden comenzar en momentos igualmente espaciados; el intervalo entre dos tiempos de (re) transmisión consecutivos se denomina *slot*; la duración de un slot es siempre mayor que la de cualquier mensaje.
- (ii) Todos los mensajes atrasados, es decir, aquellos mensajes que ya han intentado sin éxito (quizás más de una vez) atravesar el enlace, requieren retransmisión con independencia entre sí con probabilidad  $\nu \in (0, 1)$  en cada slot. Esta es conocida como política de retransmisión de Bernoulli.
- (iii) Los mensajes nuevos (los que se presentan por primera vez) intentan comunicarse de inmediato.



- Mensaje nuevo
- Mensaje atrasado, no autorizado para intentar retransmitir
- <sup>↗</sup> Mensaje atrasado, autorizado para intentar retransmitir
- ↓ Transmisión exitosa (o retransmisión)

Sea  $X_n$  el número de mensajes atrasados al comienzo del slot  $n$ . Los mensajes atrasados se comportan independientemente, y cada uno tiene probabilidad  $v$  de intentar la retransmisión en el slot  $n$ . En particular, si existe  $X_n = k$  mensajes atrasados, la probabilidad de que  $i$  entre ellos intenten retransmitirse en el slot  $n$  es

$$b_i(k) = \binom{k}{i} v^i (1-v)^{k-i}. \quad (3.27)$$

Sea  $A_n$  el número de peticiones nuevas para transmisión en el slot  $n$ . La sucesión  $\{A_n\}_{n \geq 0}$  se asume i.i.d. con la distribución

$$P(A_n = j) = a_j. \quad (3.28)$$

La cantidad

$$\lambda = E[A_n] = \sum_{i=1}^{\infty} i a_i \quad (3.29)$$

es la *intensidad del tráfico*. Si se supone que  $a_0 + a_1 \in (0, 1)$ , entonces  $\{X_n\}_{n \geq 0}$  es una CMH irreducible. Su matriz de transición es

$$p_{ij} = \begin{cases} b_1(i)a_0 & \text{si } j = i - 1, \\ [1 - b_1(i)]a_0 + b_0(i)a_1 & \text{si } j = i, \\ [1 - b_0(i)]a_1 & \text{si } j = i + 1, \\ a_{j-1} & \text{si } j \geq i + 2. \end{cases} \quad (3.30)$$

Esto último se obtiene por conteo. Por ejemplo, la primera línea corresponde a uno entre los  $i$  mensajes atrasados teniendo éxito en su retransmisión, y para esto no debería haber ningún mensaje nuevo (probabilidad  $a_0$ ) y solo uno de los  $i$  mensajes atrasados autorizado para su retransmisión (probabilidad  $b_1(i)$ ). De la segunda línea, cada término corresponde respectivamente a estos dos eventos «ningún mensaje nuevo y cero o estrictamente mayor a dos peticiones de retransmisión de los mensajes atrasados» y «cero peticiones de retransmisión de los mensajes atrasados y un mensaje nuevo».

El objetivo del ejemplo es mostrar que el sistema usando la política de retransmisión de Bernoulli *no es estable*, en el sentido de que la cadena  $\{X_n\}_{n \geq 0}$  *no es recurrente positiva*. Para probar la inestabilidad, se debe, en vista del teorema 3.3.1, contradecir la existencia de una distribución estacionaria  $\pi$ .

Si tal distribución estacionaria existe, esta debería satisfacer las ecuaciones de balance

$$\begin{aligned} \pi(i) = & \pi(i) \{ [1 - b_1(i)]a_0 + b_0(i)a_1 \} + \pi(i-1) [1 - b_0(i-1)]a_1 \\ & + \pi(i+1) b_1(i+1)a_0 + \sum_{l=2}^{\infty} \pi(i-l)a_l \end{aligned}$$

donde  $\pi(j) = 0$  si  $j < 0$ . Escribiendo

$$P_N = \sum_{i=0}^N \pi(i)$$

y sumando las ecuaciones de balance de 1 a  $N$ , se obtiene

$$P_N = \pi(N)b_0(N)a_1 + \pi(N+1)b_1(N+1)a_0 + \sum_{l=0}^N a_l P_{N-l}.$$

Esto se convierte en

$$P_N(1 - a_0) = \pi(N)b_0(N)a_1 + \pi(N+1)b_1(N+1)a_0 + \sum_{l=0}^N a_l P_{N-l}.$$

Pero como  $P_N$  incrementa con  $N$  y  $\sum_{l=1}^N a_l \leq \sum_{l=0}^{\infty} a_l = 1 - a_0$ , se tiene

$$\sum_{l=0}^N a_l P_{N-l} \leq P_{N-1}(1 - a_0),$$

y por lo tanto

$$P_N(1 - a_0) \leq \pi(N)b_0(N)a_1 + \pi(N+1)b_1(N+1)a_0 + P_{N-1}(1 - a_0),$$

para lo cual se sigue que

$$\frac{\pi(N+1)}{\pi(N)} \geq \frac{1 - a_0 - b_0(N)a_1}{b_1(N+1)a_0}$$

Usando la expresión 3.27, se obtiene

$$\frac{\pi(N+1)}{\pi(N)} \geq \frac{(1 - a_0) - (1 - v)^N a_1}{(N+1)v(1 - v)^N a_0}.$$

Para todos los valores de  $v \in (0, 1)$ , el lado derecho de esta desigualdad eventualmente se vuelve infinito, y esto contradice la ecuación  $\sum_{N=1}^{\infty} \pi(N) = 1$  y las desigualdades  $\pi(N) > 0$  que  $\pi$  debería satisfacer como la distribución estacionaria de una cadena de Markov irreducible.

## 3.4. Promedios Empíricos

### 3.4.1. Teorema Ergódico

Esta subsección está dedicada al teorema ergódico de las cadenas de Markov. Da condiciones que garantizan que los promedios empíricos del tipo

$$\frac{1}{N} \sum_{k=1}^N f(X_k, \dots, X_{k+L})$$

converge a los promedios probabilísticos.

**Proposición 3.4.1.** Sea  $\{X_n\}_{n \geq 0}$  una CMH irreducible recurrente, y sea  $x$  la medida invariante canónica asociada con el estado  $0 \in E$ ,

$$x_i = E_0 \left[ \sum_{n \geq 1} \mathbb{1}_{\{X_n=i\}} \mathbb{1}_{\{n \leq T_0\}} \right], \quad (3.31)$$

donde  $T_0$  es el tiempo de retorno a 0. Definido para  $n \geq 1$

$$v(n) = \sum_{k=1}^n \mathbb{1}_{\{X_k=0\}}. \quad (3.32)$$

Sea  $f : E \rightarrow \mathbb{R}$  tal que

$$\sum_{i \in E} |f(i)| x_i < \infty. \quad (3.33)$$

Entonces, para cualquier distribución inicial  $\mu$ ,  $P_\mu$ -c.s.,

$$\lim_{N \uparrow \infty} \frac{1}{v(N)} \sum_{k=1}^N f(X_k) = \sum_{i \in E} f(i) x_i. \quad (3.34)$$

**Demostración.** Sea  $T_0 = \tau_1, \tau_2, \tau_3, \dots$  los tiempos de retornos sucesivos al estado 0, y se define

$$U_p = \sum_{n=\tau_p+1}^{\tau_{p+1}} f(X_n).$$

En vista del teorema del ciclo regenerativo 2.7.4,  $\{U_p\}_{p \geq 1}$  es una sucesión de v.a.i.i.d. Asumiendo  $f \geq 0$  y usando la propiedad fuerte de Markov,

$$\begin{aligned} E[U_1] &= E_0 \left[ \sum_{n=1}^{T_0} f(X_n) \right] \\ &= E_0 \left[ \sum_{n=1}^{T_0} \sum_{i \in E} f(i) \mathbb{1}_{\{X_n=i\}} \right] \\ &= \sum_{i \in E} f(i) \left[ \sum_{n=1}^{T_0} \mathbb{1}_{\{X_n=i\}} \right] \\ &= \sum_{i \in E} f(i) x_i. \end{aligned}$$

Por hipótesis, esta cantidad es finita, y por lo tanto la ley fuerte de los grandes números aplica, esto da

$$\lim_{n \uparrow \infty} \frac{1}{n} \sum_{p=1}^n U_p = \sum_{i \in E} f(i) x_i,$$

esto es,

$$\lim_{n \uparrow \infty} \frac{1}{n} \sum_{k=T_0+1}^{\tau_{n+1}} f(X_k) = \sum_{i \in E} f(i) x_i. \quad (3.35)$$



Observando que

$$\tau_{v(n)} \leq n < \tau_{v(n)+1},$$

se tiene

$$\frac{\sum_{k=1}^{\tau_{v(n)}} f(X_k)}{v(n)} \leq \frac{\sum_{k=1}^{\tau_n} f(X_k)}{v(n)} \leq \frac{\sum_{k=1}^{\tau_{v(n)+1}} f(X_k)}{v(n)}$$

Ya que la cadena es recurrente,  $\lim_{n \uparrow \infty} v(n) = \infty$ , y por lo tanto, de 3.35, los términos extremos de la desigualdad anterior tienden a  $\sum_{i \in E} f(i)x_i$  cuando  $n$  tiende a  $\infty$ , y esto implica 3.34. El caso de una función con signo arbitrario se obtiene al considerar 3.34 separando por  $f^+ = \max(0, f)$  y  $f^- = \max(0, -f)$ , y luego tomando la diferencia de estas. La diferencia no toma una forma indeterminada por la hipótesis 3.33.

**Teorema 3.4.1** (Teorema Ergódico). Sea  $\{X_n\}_{n \geq 0}$  una cadena de Markov irreducible recurrente positiva con distribución estacionaria  $\pi$ , y sea  $f : E \rightarrow \mathbb{R}$  tal que

$$\sum_{i \in E} |f(i)|\pi(i) < \infty. \quad (3.36)$$

Entonces para cualquier distribución inicial  $\mu$ ,  $P_\mu$ -c.s.,

$$\lim_{n \uparrow \infty} \frac{1}{N} \sum_{k=1}^N f(X_k) = \sum_{i \in E} f(i)\pi(i). \quad (3.37)$$

**Demostración.** Se aplica la proposición 3.4.1 a  $f \equiv 1$ . La condición 3.33 se satisface, ya que en el caso de ser recurrente positiva,  $\sum_{i \in E} x_i < \infty$ . Por lo tanto,  $P_\mu$ -c.s.,

$$\lim_{N \uparrow \infty} \frac{N}{v(N)} = \sum_{j \in E} x_j.$$

Ahora,  $f$  satisfaciendo 3.36 también satisface 3.33, ya que  $x$  y  $\pi$  son proporcionales, y por lo tanto,  $P_\mu$ -c.s.,

$$\lim_{N \uparrow \infty} \frac{1}{v(N)} \sum_{k=1}^N f(X_k) = \sum_{i \in E} f(i)x_i.$$

La combinación de las ecuaciones anteriores da como resultado,  $P_\mu$ -c.s.,

$$\lim_{N \uparrow \infty} \frac{1}{N} \sum_{k=1}^N f(X_k) = \lim_{N \rightarrow \infty} \frac{v(N)}{N} \frac{1}{v(N)} \sum_{k=1}^N f(X_k) = \frac{\sum_{i \in E} f(i)x_i}{\sum_{j \in E} x_j},$$

de lo cual 3.37 se sigue, ya que  $\pi$  es obtenida de la normalización de  $x$ .

**Corolario 3.4.1.1.** Sea  $\{X_n\}_{n \geq 1}$  una cadena de Markov irreducible positiva recurrente con distribución estacionaria  $\pi$ , y sea  $g : E^{L+1} \rightarrow \mathbb{R}$  tal que

$$\sum_{i_0, \dots, i_L} |g(i_0, \dots, i_L)|\pi(i_0)p_{i_0 i_1} \cdots p_{i_{L-1} i_L} < \infty. \quad (3.38)$$

Entonces para toda distribución inicial  $\mu$ ,  $P_\mu$ -c.s.,

$$\lim_{N \uparrow \infty} \frac{1}{N} \sum_{k=1}^N g(X_k, X_{k+1}, \dots, X_{k+L}) = \sum_{i_0, i_1, \dots, i_L} g(i_0, i_1, \dots, i_L)\pi(i_0)p_{i_0 i_1} \cdots p_{i_{L-1} i_L}.$$

**Demostración.** Aplicando el teorema ergódico a la cadena

$$\{(X_n, X_{n+1}, \dots, X_{n+L})\}_{n \geq 0}$$

, la cual es irreducible recurrente y admite la distribución estacionaria

$$\pi(i_0)p_{i_0i_1} \cdots p_{i_{L-1}i_L}.$$

Notando que

$$\sum_{i_0, i_1, \dots, i_L} g(i_0, i_1, \dots, i_L) \pi(i_0)p_{i_0i_1} \cdots p_{i_{L-1}i_L} = E_\pi[g(X_0, \dots, X_L)]$$

La versión del teorema ergódico para cadenas de Markov presentada en el teorema 3.4.1 es una especie de ley fuerte de los grande números y puede usarse en simulaciones para calcular, cuando se desconoce  $\pi$ , cantidades del tipo  $E_\pi[f(X_0)]$ .

### 3.4.2. Ejemplos

**Ejemplo 3.4.1.** Se adopta la interpretación del ejemplo del reemplazo de una máquina 2.1.1 asumiendo recurrencia positiva. Una visita de la cadena al estado 0 corresponde a la avería de una máquina, y por lo tanto, en vista del teorema ergódico,

$$\pi(0) = \lim_{N \uparrow \infty} \frac{1}{N} \sum_{k=1}^N \mathbb{1}_{\{X_k=0\}}$$

es la frecuencia empírica de averías. Renombrada por

$$\pi(0) = E_0[T_0]^{-1},$$

donde  $T_0$  es el tiempo de retorno a 0. Aquí,

$$E_0[T_0] = E[U],$$

y por lo tanto

$$\lim_{N \uparrow \infty} \frac{1}{N} \sum_{k=1}^N \mathbb{1}_{\{X_k=0\}} = \frac{1}{E[U]}. \quad (3.39)$$

Se usa el supuesto de que el costo de una avería es tan importante que es mejor reemplazar una máquina en funcionamiento durante su vida útil (la avería implica reparaciones costosas, mientras que el reemplazo solo implica costos de mantenimiento moderados). La *política de jubilación a una edad fija* asigna un número entero  $T \geq 1$  y requiere que una máquina que haya alcanzado la edad  $T$  sea reemplazada inmediatamente. El interés recae en calcular la frecuencia empírica de averías (no sustituciones).

La cadena correspondiente a esta situación es la misma que antes, excepto que los tiempos  $U_n$  son reemplazados por  $V_n = U_n \wedge T$ . También, un reemplazo (no avería)

ocurre en el tiempo  $n$  si y sólo si  $X_n = 0$  y  $X_{n-1} = T - 1$ . Como  $X_{n-1} = T - 1$  y esto implica que  $X_n = 0$ , por lo tanto un reemplazo ocurre en el tiempo  $n$  si y sólo si

$$X_{n-1} = T - 1$$

La frecuencia empírica de reemplazos es, por lo tanto, en vista del teorema ergódico,

$$\lim_{N \uparrow \infty} \frac{1}{N} \sum_{k=1}^N \mathbb{1}_{\{X_k=T-1\}} = \pi(T-1).$$

La formula 3.26 da como resultado

$$\pi(T-1) = \frac{P(V \geq T)}{E[V]},$$

y por lo tanto, ya que  $V = U \wedge T$ ,

$$\pi(T-1) = \frac{P(U \geq T)}{E[U \wedge T]}.$$

La frecuencia empírica de visitas al estado 0 es, por 3.39,

$$\frac{1}{E[U \wedge T]}.$$

La frecuencia empírica de averías es por lo tanto

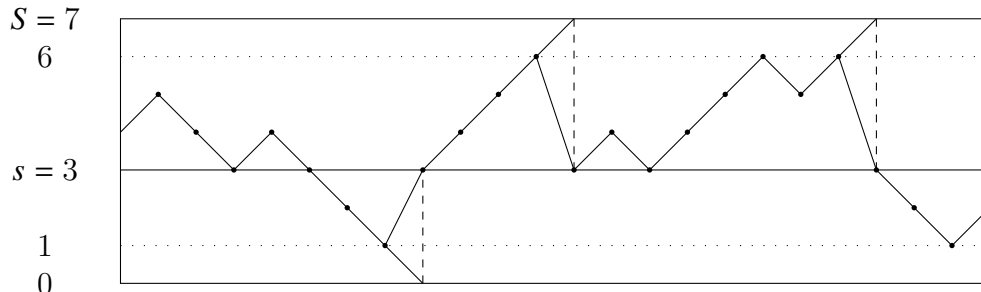
$$\frac{1}{E[U \wedge T]} - \frac{P(U \geq T)}{E[U \wedge T]} = \frac{P(U < T)}{E[U \wedge T]}.$$

**Ejemplo 3.4.2.** El nivel de efectivo en un banco al comienzo del día  $n$  es  $X_n$ , el número de unidades inactivas en la caja fuerte. Por cada día que una unidad pasa inactiva en la caja fuerte, el banco incurre en una pérdida de  $r$  unidades. En la ausencia de control, el nivel de efectivo fluctúa como una caminata aleatoria simétrica.

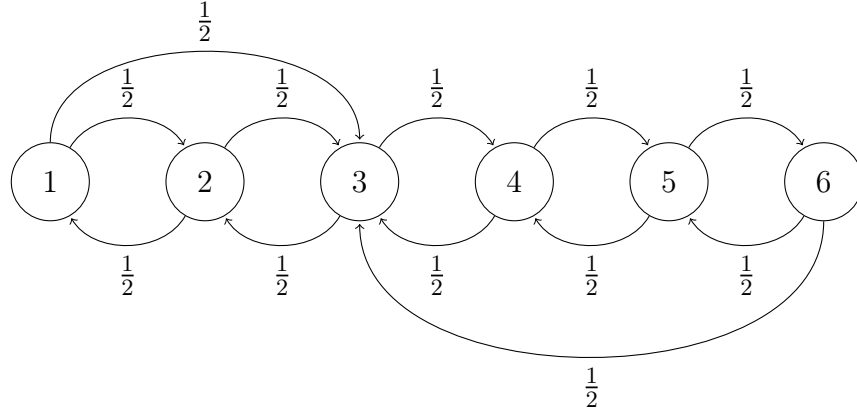
El modelo usa dos enteros  $s$  y  $S$ , donde  $0 < s < S$ . El número  $S$  es una barrera superior: Si al inicio del día  $n$ , el nivel de efectivo es  $X_n = S - 1$  y si durante el día  $n$  una unidad más entra al banco, entonces durante la noche desde el día  $n$  al día  $n + 1$ ,  $S - s$  unidades se transfieren a una ocupación más rentable (bonos, acciones, etc.), y el nivel de efectivo al inicio del día  $n + 1$  es por lo tanto  $X_{n+1} = s$ .

También, si  $X_n = 1$  y se retira una unidad más de la caja fuerte durante el día  $n$ , la caja fuerte es repuesta al nivel  $s$  durante la noche, entonces  $X_{n+1} = s$ .

Por lo tanto, el proceso  $\{X_n\}_{n \geq 0}$  toma valores en  $\{1, \dots, S - 1\}$  si se asume que  $X_0 \in [1, S - 1]$ . Una trayectoria típica del nivel de efectivo se muestra a continuación:



El proceso  $\{X_n\}_{n \geq 0}$  es una CMH con la gráfica de transición siguiente, donde todas las probabilidades de transición son iguales a  $\frac{1}{2}$ .



Además del costo de almacenamiento  $r$  por día por unidad, el banco incurre en un costo de transacción de  $\alpha$  cada vez que se mueve una unidad dentro o fuera de la caja fuerte. Por lo tanto, el costo por día a largo plazo del modelo  $(s, S)$  es

$$\lim_{N \uparrow \infty} \frac{1}{N} \sum_{n=1}^N \left[ \left( \sum_{k=1}^{S-1} rk \mathbb{1}_{\{X_n=k\}} \right) + \alpha(S-s) \mathbb{1}_{\{X_n=S-1, X_{n+1}=s\}} + \alpha s \mathbb{1}_{\{X_n=1, X_{n+1}=s\}} \right].$$

La cadena  $\{X_n\}_{n \geq 0}$  es irreducible, y ya que el espacio de estados es finito, es recurrente positiva. Sea  $\pi$  su distribución estacionaria.

Por el teorema ergódico,

$$\lim_{N \uparrow \infty} \frac{1}{N} \sum_{n=1}^N \left( \sum_{k=1}^{S-1} rk \mathbb{1}_{\{X_n=k\}} \right) = r \sum_{k=1}^{S-1} k\pi(k)$$

y

$$\lim_{N \uparrow \infty} \frac{1}{N} \sum_{n=1}^N \alpha(S-s) \mathbb{1}_{\{X_n=S-1, X_{n+1}=s\}} + \alpha s \mathbb{1}_{\{X_n=1, X_{n+1}=s\}}$$

es igual a

$$\alpha(S-s)\pi(S-1)\frac{1}{2} + \alpha s\pi(1)\frac{1}{2},$$

así que el costo a la larga por día es

$$C = r \sum_{k=1}^{S-1} k\pi(k) + \frac{1}{2}\alpha(S-s)\pi(S-1) + \frac{1}{2}\alpha s\pi(1).$$

La distribución estacionaria  $\pi$  está determinada por las ecuaciones de balance

global

$$\begin{aligned}
\pi(1) &= \frac{1}{2}\pi(2), \\
\pi(2) &= \frac{1}{2}\pi(1) + \frac{1}{2}\pi(3), \\
&\vdots \\
\pi(s-1) &= \frac{1}{2}\pi(s-2) + \frac{1}{2}\pi(s), \\
\pi(s) &= \frac{1}{2}\pi(s-1) + \frac{1}{2}\pi(s+1) + \frac{1}{2}\pi(S-1) + \frac{1}{2}\pi(1), \\
\pi(s+1) &= \frac{1}{2}\pi(s) + \frac{1}{2}\pi(s+2), \\
&\vdots \\
\pi(S-2) &= \frac{1}{2}\pi(S-3) + \frac{1}{2}\pi(S-1), \\
\pi(S-1) &= \frac{1}{2}\pi(S-2)
\end{aligned}$$

junto con la condición de normalización  $\sum_{i=1}^{S-1} \pi(i) = 1$ . Como el rango de la matriz de transición  $\mathbf{P}$  es  $|E|-1$ , una de las  $S-1$  ecuaciones de balance puede ser eliminada, y puede ser elegida arbitrariamente. Se elige la ecuación relativa a  $s$ .

De las ecuaciones 1 a  $s-1$ , se obtiene para  $i \in [1, s]$ ,

$$\pi(i) = i\pi(1).$$

Las ecuaciones de  $s+1$  a  $S$  dan como resultado para  $j \in [1, S-s]$ ,

$$\pi(S-j) = j\pi(S-1).$$

Tomando  $i = s$  y  $j = S-s$ , se tiene  $\pi(s) = s\pi(1) = (S-s)\pi(S-1)$ , entonces

$$\pi(S-1) = \frac{s}{S-s}\pi(1).$$

De la ecuación  $\sum_{i=1}^{S-1} \pi(i) = 1$  se tiene que

$$\pi(1)(1 + \dots + s) + \pi(S-1)(1 + \dots + (S-s-1)) = 1,$$

esto es,

$$\pi(1) \frac{s(s+1)}{2} + \pi(S-1) \frac{(S-s-1)(S-s)}{2} = 1,$$

y tomando en cuenta la expresión de  $\pi(S-1)$  en términos de  $\pi(1)$ , se tiene

$$\pi(1) \left[ \frac{s(s+1)}{2} + \frac{(S-s-1)s}{2} \right] = 1,$$

esto es,

$$\pi(1) = \frac{2}{sS}.$$

Por lo tanto, finalmente,

$$\pi(i) = \begin{cases} \frac{2i}{sS} & \text{si } i \in [1, s] \\ \frac{2(S-i)}{(S-s)S} & \text{si } i \in [s+1, S]. \end{cases}$$

El costo  $C$  por lo tanto puede ser calculado en términos de  $s$  y  $S$  y esto puede ser explotado para propósitos de optimización [36].

### 3.4.3. Teorema Elemental de Renovación

**Teorema 3.4.2** (Teorema Elemental de Renovación). Sea  $\{S_n\}_{n \geq 1}$  una sucesión de v.a.i.i.d. tal que  $E[S_1] < \infty$ , y sea  $R_0$  una variable aleatoria no negativa independiente de esta sucesión. Se define para toda  $n \geq 0$ ,  $R_{n+1} = R_n + S_{n+1}$  y para  $t \geq 0$ ,  $N(t) = \sum_{n=1}^{\infty} \mathbb{1}_{\{R_n \leq t\}}$ .

Ahora sea  $\{Y_n\}_{n \geq 1}$  una sucesión de v.a.i.i.d. tal que  $E[|Y_1|] < \infty$ . Entonces

$$\lim_{t \uparrow \infty} \frac{N(t)}{t} = \frac{1}{E[S_1]} \quad (3.40)$$

y

$$\lim_{t \uparrow \infty} \frac{\sum_{n=1}^{N(t)} Y_n}{t} = \frac{E[Y_1]}{E[S_1]}. \quad (3.41)$$

**Demostración.** Como  $R_{N(t)} \leq t < R_{N(t)+1}$ , se tiene

$$\frac{N(t)}{R_{N(t)+1}} < \frac{N(t)}{t} \leq \frac{N(t)}{R_{N(t)}}.$$

Pero el término de la derecha es el inverso de

$$\frac{R_{N(t)}}{N(t)} = \frac{R_0 + \sum_{n=1}^{N(t)} S_n}{N(t)}.$$

Por la ley fuerte de los grandes números y el hecho de que  $\lim_{t \uparrow \infty} N(t) = \infty$  (las  $S_n$ 's son finitas), esta cantidad tiende a  $E[S_1]$ , y similarmente,

$$\frac{R_{N(t)+1}}{N(t)} = \frac{R_{N(t)+1}}{N(t)+1} \cdot \frac{N(t)+1}{N(t)}$$

tiende a  $E[S_1]$  cuando  $t \rightarrow \infty$ .

La demostración de 3.41 se sigue de la LFNN y 3.40, ya que

$$\frac{\sum_{n=1}^{N(t)} Y_n}{t} = \frac{\sum_{n=1}^{N(t)} Y_n}{N(t)} \cdot \frac{N(t)}{t}.$$

Se nota que el resultado de arriba cubre el caso a tiempo discreto (las  $S_n$ 's y  $t$  son enteros) tanto como el caso a tiempo continuo (las  $S_n$ 's y  $t$  son reales).

El siguiente resultado extiende la fórmula 3.7.

**Ejemplo 3.4.3.** Sea  $\{X_n\}_{n \geq 0}$  una CMH recurrente positiva con espacio de estados  $E$  y distribución estacionaria  $\pi$ . Sea  $S$  un tiempo de paro de esta cadena y sea  $i$  un estado, tal que,  $P_i$ -c.s.,  $S \in (0, \infty)$  y  $X_S = i$ . Entonces para toda  $j \in E$ ,

$$E_i \left[ \sum_{k=0}^{S-1} \mathbb{1}_{\{X_k=j\}} \right] = E_i[S] \pi(j). \quad (3.42)$$

Para probar esto, se observa que  $S$ , siendo un tiempo de paro de la cadena, puede ser escrito como

$$S = \phi(X_0, X_1, \dots)$$

para algún funcional  $\phi$ , es decir

$$S = \sum_{m=0}^{\infty} m \psi_m(X_0, \dots, X_m),$$

donde

$$\mathbb{1}_{\{S=m\}} = \psi_m(X_0, \dots, X_m).$$

Se define la sucesión  $\{S_n\}_{n \geq 1}$  por  $S_1 = S$ , y para  $k \geq 0$ ,

$$S_{k+1} = \phi(X_{R_k}, X_{R_k+1}, \dots),$$

donde  $R_0 = 0$  y, para  $k \geq 1$ ,  $R_k = S_1 + \dots + S_k$ . Como un hecho de importancia, ya que  $S$  es un tiempo de paro,  $S_{k+1}$  depende solo sobre el ciclo

$$C_k = (X_{R_k}, \dots, X_{R_k+S_{k+1}-1})$$

(recordando que  $X_{R_k+S_{k+1}} = i$ . Por la propiedad fuerte de Markov, los ciclos  $\{C_k\}_{k \geq 0}$  son i.i.d. bajo  $P_i$ , y en particular,  $\{S_n\}_{n \geq 1}$  es i.i.d.

Si  $E_i[S] < \infty$ . El teorema elemental de renovación con

$$Y_n = \sum_{k=R_{n-1}}^{R_n-1} \mathbb{1}_{\{X_k=j\}}$$

(se observa que  $E_i[|Y_1|] \leq E_i[S_1] < \infty$ ) da como resultado

$$\lim_{t \uparrow \infty} \frac{\sum_{n=1}^{N(t)} Y_n}{t} = \frac{E_i[\sum_{k=0}^{S-1} \mathbb{1}_{\{X_k=j\}}]}{E_i[S]}.$$

Como

$$\frac{\sum_{n=1}^{N(t)} Y_n}{t} = \frac{\sum_{k=R_0}^{R_{N(t)}-1} \mathbb{1}_{\{X_k=j\}}}{t} = \frac{\sum_{k=R_0}^{R_{N(t)}-1} \mathbb{1}_{\{X_k=j\}}}{R_{N(t)}} \cdot \frac{R_{N(t)}}{t}.$$

Ahora,  $\lim_{t \uparrow \infty} \frac{R_{N(t)}}{t} = 1$ , y por lo tanto, por el teorema ergódico,

$$\lim_{t \uparrow \infty} \frac{\sum_{n=1}^{N(t)} Y_n}{t} = \pi(j).$$

Esto suponiendo que  $E_i[S] < \infty$ . Si  $E_i[S] = \infty$ , se considera  $S^{(n)} = S \wedge \tau_n(i)$ , donde  $\tau_n(i)$  es el tiempo de retorno  $n$  a  $i$ , escribiendo 3.42 para  $S^{(n)}$ , y sea  $n \rightarrow \infty$ , se obtiene que

$$E_i\left[\sum_{k=0}^{S-1} \mathbb{1}_{\{X_k=j\}}\right] = \infty.$$

**Ejemplo 3.4.4.** [2] Sea  $i$  y  $j$  dos estados distintos y sea  $S$  el primer tiempo de retorno a  $i$  después de la primera visita a  $j$ . Entonces  $E_i[S] = E_i[T_j] + E_j[T_i]$  (se usa la propiedad fuerte de Markov en  $T_j$ ). También,

$$E_i\left[\sum_{n=0}^{S-1} \mathbb{1}_{\{X_n=j\}}\right] = E_i\left[\sum_{n=T_j}^{S-1} \mathbb{1}_{\{X_n=j\}}\right] = E_j\left[\sum_{n=0}^{T_i-1} \mathbb{1}_{\{X_n=j\}}\right],$$

donde la última ecuación es la propiedad fuerte de Markov. Por lo tanto, por 3.42,

$$E_j\left[\sum_{n=0}^{T_i-1} \mathbb{1}_{\{X_n=j\}}\right] = \pi(j)(E_i[T_j] + E_j[T_i]). \quad (3.43)$$

Usando palabras, el lado izquierdo de esta ecuación es

$$E_j[\text{número de visitas a } j \text{ antes de } i].$$

Ahora, la probabilidad de que  $j$  no sea visitada entre dos visitas sucesivas a  $i$  es  $P_i(T_j > T_i)$ . Por lo tanto, el número de visitas a  $i$  (incluyendo el tiempo 0) antes de  $T_j$  tiene una distribución geométrica con parámetro  $p = P_i(T_j > T_i)$ , y el número promedio de tales visitas es

$$\frac{1}{P_i(T_j < T_i)}.$$

Por lo tanto, por 3.43, después de intercambiar los roles de  $i$  y  $j$ ,

$$P_i(T_j < T_i) = \frac{1}{\pi(i)(E_i[T_j] + E_j[T_i])}.$$



---

---

# CAPÍTULO 4

---

## COMPORTAMIENTO A LARGO PLAZO

### 4.1. Acoplamiento

#### 4.1.1. Convergencia en Variación

Se considera una CMH irreducible recurrente positiva. En particular, si su distribución inicial es la distribución estacionaria, mantiene la misma distribución en todo momento. Entonces se dice que la cadena está en *régimen estacionario*, en *equilibrio*, o en *estado estacionario*.

Surge naturalmente una pregunta: ¿Cuál es el comportamiento a largo plazo de la cadena cuando la distribución inicial  $\mu$  es arbitraria? Por ejemplo, ¿convergerá al equilibrio y en qué sentido?

Cuando la CMH es reducible, otro tipo de problema es de interés. Se usa la suposición, por ejemplo, que el conjunto de estados transitorios no es vacío y que cada estado restante es absorbente. Es posible calcular la probabilidad de alcanzar un estado absorbente dado cuando el estado inicial es transitorio, o la probabilidad de permanecer para siempre en el conjunto transitorio. En este caso especial, donde todos los estados recurrentes están absorbiendo, la probabilidad de dejar el conjunto transitorio es exactamente la probabilidad de converger. Se trata aquí de una convergencia casi segura.

Para una CMH ergódica, el tipo de convergencia de interés no es una convergencia casi segura, si no una convergencia en la variación de la distribución en el tiempo  $n$  a la distribución estacionaria. Este tipo de convergencia es relativo a una estructura métrica que se procede a definir.

**Definición 4.1.1.** Sea  $E$  un espacio numerable y sea  $\alpha$  y  $\beta$  distribuciones de probabilidad sobre  $E$ . La *distancia en variación*  $d_V(\alpha, \beta)$  está definida por

$$d_V(\alpha, \beta) = \frac{1}{2}|\alpha - \beta| = \frac{1}{2} \sum_{i \in E} |\alpha(i) - \beta(i)|. \quad (4.1)$$

La distancia en variación entre dos variables aleatorias  $X$  y  $Y$  con valores en  $E$  y respectivas distribuciones  $\mathcal{L}(X)$  y  $\mathcal{L}(Y)$  es  $d_V(\mathcal{L}(X), \mathcal{L}(Y))$ , y esta es denotada con un pequeño abuso de notación por  $d_V(X, Y)$ .

**Lema 4.1.1.** Sea  $X$  y  $Y$  dos variables aleatorias con valores en el mismo espacio numerable  $E$ . Entonces

$$\sup_{A \subset E} |P(X \in A) - P(Y \in A)| = \sup_{A \subset E} \{P(X \in A) - P(Y \in A)\} = d_V(X, Y). \quad (4.2)$$

**Demostración.** Para la primera ecuación se observa que para cada  $A$  existe una  $B$  tal que

$$|P(X \in A) - P(Y \in A)| = P(X \in E) - P(Y \in B)$$

(se toma  $B = A$  ó  $A^c$ ). Para la segunda ecuación, se escribe

$$P(X \in A) - P(Y \in A) = \sum_{i \in E} \mathbb{1}_A(i) \{P(X = i) - P(Y = i)\}$$

y se observa que el lado derecho es maximal para

$$A = \{i \in E : P(X = i) > P(Y = i)\}$$

También, para cualquier  $A \subset E$ ,

$$\sum_{i \in E} \mathbb{1}_A(i) \{P(X = i) - P(Y = i)\} + \sum_{i \in E} \mathbb{1}_{A^c}(i) \{P(X = i) - P(Y = i)\} = 0$$

porque  $\sum_{i \in E} \{P(X = i) - P(Y = i)\} = 0$ . Para el conjunto específico  $A$  de arriba,

$$P(X = i) - P(Y = i) = |P(X = i) - P(Y = i)|$$

sobre  $A$ , y

$$P(X = i) - P(Y = i) = -|P(X = i) - P(Y = i)|$$

sobre  $A^c$ . Por lo tanto, para este conjunto  $A$  en particular,

$$\begin{aligned} \sum_{i \in E} \mathbb{1}_A(i) \{P(X = i) - P(Y = i)\} &= \sum_{i \in E} \mathbb{1}_A(i) |P(X = i) - P(Y = i)| \\ &= \sum_{i \in E} \mathbb{1}_{A^c}(i) |P(X = i) - P(Y = i)| \\ &= \frac{1}{2} \sum_{i \in E} |P(X = i) - P(Y = i)|. \end{aligned}$$

Para dos distribuciones de probabilidad  $\alpha$  y  $\beta$  sobre el conjunto numerable  $E$ , sea  $\mathcal{D}(\alpha, \beta)$  la colección de vectores aleatorios  $(X, Y)$  que toman valores en  $E \times E$ , y con distribuciones marginales  $\alpha$  y  $\beta$ , esto es,

$$\alpha = \mathcal{L}(X), \quad \beta = \mathcal{L}(Y). \quad (4.3)$$

**Teorema 4.1.1.** Para cualquier  $(X, Y) \in \mathcal{D}(\alpha, \beta)$ ,

$$P(X = Y) \leq 1 - d_V(\alpha, \beta), \quad (4.4)$$

y la igualdad se cumple para alguna pareja  $(X, Y) \in \mathcal{D}(\alpha, \beta)$ , a la cual se le dice que realiza la *coincidencia maximal*.

**Demostración.** Para una  $A \subset E$  arbitraria,

$$\begin{aligned} P(X \neq Y) &\geq P(X \in A, Y \in A^c) = P(X \in A) - P(X \in A, Y \in A) \\ &\geq P(X \in A) - P(Y \in A), \end{aligned}$$

y por lo tanto

$$P(X \neq Y) \geq \sup_{A \subset E} \{P(X \in A) - P(Y \in A)\} = d_V(\alpha, \beta).$$

Para finalizar la demostración, es suficiente construir  $(X, Y) \in \mathcal{D}(\alpha, \beta)$  que cumpla la igualdad. Se necesita la siguiente observación:

$$\frac{1}{2}|\alpha - \beta| = \sum_{i \in E} (\alpha(i) - \beta(i))^+ = \sum_{i \in E} (\beta(i) - \alpha(i))^+ = \sum_{i \in E} \min(\alpha(i), \beta(i)). \quad (4.5)$$

Sea  $U, Z, V, W$  variables aleatorias independientes;  $U$  con valores en  $\{0, 1\}$ , y  $Z, V, W$  con valores en  $E$ . Las distribuciones de estas variables aleatorias están dadas por:

$$\begin{aligned} P(U = 1) &= 1 - d_V(\alpha, \beta) \\ P(Z = i) &= \min(\alpha(i), \beta(i)) / (1 - d_V(\alpha, \beta)), \\ P(V = i) &= (\alpha(i) - \beta(i))^+ / d_V(\alpha, \beta), \\ P(W = i) &= (\beta(i) - \alpha(i))^+ / d_V(\alpha, \beta) \end{aligned}$$

Definiendo

$$X = UZ + (1 - U)V, \quad Y = UZ + (1 - U)W,$$

se tiene que

$$\begin{aligned} P(X = i) &= P(U = 1, Z = i) + P(U = 0, V = i) \\ &= P(U = 1)P(Z = i) + P(U = 0)P(V = i) \\ &= \min(\alpha(i), \beta(i)) + (\alpha(i) - \beta(i))^+ = \alpha(i), \end{aligned}$$

y similarmente,  $P(Y = i) = \beta(i)$ . Por lo tanto,  $(X, Y) \in \mathcal{D}(\alpha, \beta)$ . También,

$$P(X = Y) = P(U = 1) = 1 - d_V(\alpha, \beta).$$

**Ejemplo 4.1.1.** Sean  $X, Y$  variables aleatorias con valores en  $\{0, 1\}$  con

$$P(X = 1) = a, \quad P(Y = 1) = b,$$

donde  $a, b \in (0, 1)$ , y tal que  $P(X = Y)$  es maximal. En la notación de la teoría de arriba,

$$\alpha = (1 - a, a), \quad \beta = (1 - b, b),$$

y por lo tanto

$$d_V(\alpha, \beta) = |a - b|.$$

Si  $a \geq b$ . Las variables aleatorias de la demostración del teorema 4.1.1 tienen las siguientes distribuciones.

$$\begin{aligned} P(U = 1) &= 1 - a + b, \\ P(Z = 1) &= \frac{b}{1 - a + b}, \quad P(Z = 0) = \frac{1 - a}{1 - a + b}, \\ V &= 1, \\ W &= 0. \end{aligned}$$

Aquí  $X = UZ + 1 - U$ ,  $Y = UZ$ .

**Definición 4.1.2** (Convergencia en Variación). Sea  $\{\alpha\}_{n \geq 0}$  y  $\beta$  distribuciones de probabilidad con espacio de estados numerable  $E$ . Si

$$\lim_{n \uparrow \infty} d_V(\alpha_n, \beta) = 0,$$

la sucesión  $\{\alpha\}_{n \geq 0}$  se dice que *converge en variación* a la distribución de probabilidad  $\beta$ .

Sea  $\{X_n\}_{n \geq 0}$  un proceso estocástico con valores en  $E$ . Si para alguna distribución de probabilidad  $\pi$  sobre  $E$ , la distribución  $\mathcal{L}(X_n)$  de la variable aleatoria  $X_n$  converge en variación a  $\pi$ , i.e., si

$$\lim_{n \uparrow \infty} \sum_{i \in E} |P(X_n = i) - \pi(i)| = 0, \quad (4.6)$$

entonces  $\{X_n\}_{n \geq 0}$  se dice que *converge en variación* a  $\pi$ , entonces

$$\lim_{n \uparrow \infty} E[f(X_n)] = \pi(f) \quad (4.7)$$

para todas las funciones acotadas  $f : E \rightarrow \mathbb{R}$ , donde

$$\pi(f) = \sum_{i \in E} \pi(i) f(i). \quad (4.8)$$

Además, si  $M$  es una cota superior de  $|f|$ , entonces

$$|E[f(X_n)] - \pi(f)| = \left| \sum_{i \in E} f(i)(P(X_n = i) - \pi(i)) \right| \leq M \sum_{i \in E} |P(X_n = i) - \pi(i)|.$$

### 4.1.2. El método de Acoplamiento

El acoplamiento es una vieja idea de Doeblin (1938) [10], revivió en la teoría de las cadenas de Markov por influencia del trabajo de Griffeath (1975)[15] y Pitman (1974)[28], y fue traído a la fama por Lindvall (1977) quien dio una demostración puramente probabilística del teorema de renovación. El método de acoplamiento tiene un rango muy amplio de aplicaciones [21][3].

Se observa que la definición 4.1.2 considera solamente las distribuciones marginales del proceso, no el proceso en sí. Por lo tanto, si existe otro proceso  $\{X'_n\}_{n \geq 0}$  con  $\mathcal{L}(X'_n) = \mathcal{L}(X_n)$  para toda  $n \geq 0$ , y si existe un tercer proceso  $\{X''_n\}_{n \geq 0}$  tal que  $\mathcal{L}(X''_n) = \pi$  para toda  $n \geq 0$ , entonces de 4.6 se sigue que

$$\lim_{n \uparrow \infty} d_V(X'_n, X''_n) = 0. \quad (4.9)$$

Esta observación es útil debido a la libertad resultante de la elección de  $\{X'_n\}_{n \geq 0}$  y  $\{X''_n\}_{n \geq 0}$ . En particular, se pueden usar versiones dependientes, y el caso más interesante ocurre cuando existe un tiempo aleatorio finito  $\tau$  tal que  $X'_n = X''_n$  para toda  $n \geq \tau$ . Se sigue entonces, como se probó en el teorema 4.1.1, que

$$d_V(X'_n, X''_n) \leq P(\tau > n). \quad (4.10)$$

El hecho de que  $\tau$  sea finito es equivalente a  $\lim_{n \uparrow \infty} P(\tau > n) = 0$ , y por lo tanto 4.9 es una consecuencia de 4.10.

**Definición 4.1.3** (Acoplamiento). Dos procesos estocásticos  $\{X'_n\}_{n \geq 0}$  y  $\{X''_n\}_{n \geq 0}$  que toman valores en un conjunto numerable  $E$  se dice que se *acoplan* si existe casi seguramente un tiempo aleatorio *finito*  $\tau$  tal que

$$n \geq \tau \implies X'_n = X''_n. \quad (4.11)$$

La variable aleatoria  $\tau$  se llama *tiempo de acoplamiento* de los dos procesos.

**Teorema 4.1.2.** La desigualdad

$$d_V(X'_n, X''_n) \leq P(\tau > n). \quad (4.12)$$

se cumple para cualquier tiempo de acoplamiento  $\tau$  de  $\{X'_n\}_{n \geq 0}$  y  $\{X''_n\}_{n \geq 0}$ .

**Demostración.** Para toda  $A \subset E$ ,

$$\begin{aligned} P(X'_n \in A) - P(X''_n \in A) &= P(X'_n \in A, \tau \leq n) + P(X'_n \in A, \tau > n) \\ &\quad - P(X''_n \in A, \tau \leq n) - P(X''_n \in A, \tau > n) \\ &= P(X'_n \in A, \tau > n) - P(X''_n \in A, \tau > n) \\ &\leq P(X'_n \in A, \tau > n) \\ &\leq P(\tau > n). \end{aligned}$$

La desigualdad 4.12 entonces se sigue del lema 4.1.1.

## 4.2. Convergencia al Estado de Equilibrio

### 4.2.1. Caso Recurrente Positivo

**Teorema 4.2.1.** Sea  $\mathbf{P}$  una matriz de transición ergódica sobre el espacio de estados  $E$ . Para toda distribución de probabilidad  $\mu$  y  $\nu$  sobre  $E$ ,

$$\lim_{n \uparrow \infty} d_V(\mu^t \mathbf{P}^n, \nu^t \mathbf{P}^n) = 0. \quad (4.13)$$

En particular, si  $\nu$  es la distribución estacionaria  $\pi$ ,

$$\lim_{n \uparrow \infty} |\mu^t \mathbf{P}^n - \pi^t| = 0,$$

y con  $\mu = \delta_j$ , la distribución de probabilidad es

$$\lim_{n \uparrow \infty} \sum_{i \in E} |p_{ji}(n) - \pi(i)| = 0.$$

De la discusión que precede a la definición 4.1.3, es suficiente construir dos cadenas de acoplamiento con distribuciones iniciales  $\mu$  y  $\nu$ , respectivamente.

**Teorema 4.2.2.** Sea  $\{X_n^{(1)}\}_{n \geq 0}$  y  $\{X_n^{(2)}\}_{n \geq 0}$  CMH independientes ergódicas con la misma matriz de transición  $\mathbf{P}$  y distribuciones iniciales  $\mu$  y  $\nu$ , respectivamente. Sea  $\tau = \inf\{n \geq 0 : X_n^{(1)} = X_n^{(2)}\}$ , con  $\tau = \infty$  si las cadenas nunca se cruzan. Entonces  $\tau$  es finita casi seguramente. Más aún, el proceso  $\{X'_n\}_{n \geq 0}$  definido por

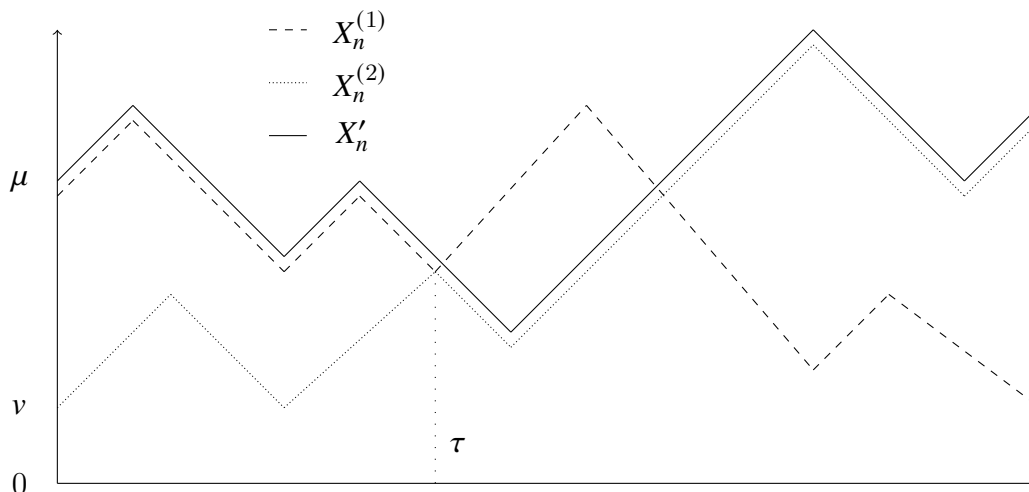
$$X'_n = \begin{cases} X_n^{(1)} & \text{si } n \leq \tau, \\ X_n^{(2)} & \text{si } n \geq \tau \end{cases} \quad (4.14)$$

es una CMH con matriz de transición  $\mathbf{P}$ .

**Demostración.** Para esta prueba se considera la CMH producto  $\{Z_n\}_{n \geq 0}$  definida por  $Z_n = (X_n^{(1)}, X_n^{(2)})$ . Esta toma valores en  $E \times E$ , y la probabilidad de transición de  $(i, k)$  a  $(j, l)$  en  $n$  pasos es  $p_{ij}(n)p_{kl}(n)$ . Esta cadena es irreducible. Como  $\mathbf{P}$  es irreducible y *aperiódica*, por el teorema 2.4.3, existe  $m$  tal que para todas las parejas  $(i, j)$  y  $(k, l)$ ,  $n \geq m$  implica  $p_{ij}(n)p_{kl}(n) > 0$ . Esto implica que el periodo de la cadena producto es 1, de nuevo por el teorema 2.4.3.

Se puede observar que  $\{\pi(i)\pi(j)\}_{(i,j) \in E^2}$  es una distribución estacionaria para la cadena producto, donde  $\pi$  es la distribución estacionaria de  $\mathbf{P}$ . Por lo tanto, por el criterio de la distribución estacionaria, la cadena producto es recurrente positiva. En particular, se llega a la diagonal de  $E \times E$  en un tiempo finito, y consecuentemente,  $P(\tau < \infty) = 1$ .

Queda demostrar que  $\{X'_n\}_{n \geq 0}$  es una CMH con matriz de transición  $\mathbf{P}$ . Esto es una consecuencia de la propiedad fuerte de Markov aplicada a la cadena producto.



**Teorema 4.2.3** (Caso Periódico). Sea  $\mathbf{P}$  una matriz de transición irreducible recurrente positiva sobre el espacio de estados numerable  $E$ , con periodo  $d$ . Sea  $\pi$  su distribución estacionaria. Si  $\mu$  es una distribución de probabilidad tal que  $\mu(C_0) = 1$  para alguna clase cíclica  $C_0$ , entonces

$$\lim_{n \uparrow \infty} \sum_{i \in C_0} |(\mu^t \mathbf{P}^{nd})_i - d\pi(i)| = 0. \quad (4.15)$$

**Demostración.** Se considera la restricción de  $\mathbf{P}^d$  a  $C_0$ , la cual es irreducible y aperiódica. Esta es positiva recurrente, ya que tiene una medida invariante con masa finita  $\pi$  restringida a  $C_0$ . Queda demostrar que  $d\pi$  restringida a  $C_0$  es una distribución de probabilidad, esto es,  $\pi(C_0) = 1/d$ . Por el teorema ergódico,

$$\lim_{N \uparrow \infty} \frac{1}{N} \sum_{n=1}^N \mathbb{1}_{\{X_n \in C_0\}} = \pi(C_0),$$

y como  $X_n \in C_0$  una vez cada  $d$  pasos, el lado izquierdo es igual a  $1/d$ .

### 4.2.2. Caso Recurrente Nulo

**Teorema 4.2.4** (Teorema de Orey). Sea  $\mathbf{P}$  una matriz de transición irreducible recurrente nula sobre  $E$ . Entonces para toda  $i, j \in E$ ,

$$\lim_{n \uparrow \infty} p_{ij}(n) = 0. \quad (4.16)$$

**Demostración.** El caso periódico se sigue del caso aperiódico al considerar la restricción de  $\mathbf{P}^d$  a  $C_0$ , una clase cíclica arbitraria, y observar que esta restricción también es nula recurrente. Por lo tanto,  $\mathbf{P}$  será asumida aperiódica.

En este caso la CMH producto  $\{Z_n\}_{n \geq 0} = \{(X_n^{(1)}, X_n^{(2)})\}_{n \geq 0}$  definida en la demostración del teorema 4.2.2 es irreducible y aperiódica. Sin embargo, no se puede argumentar que esta es recurrente, a pesar de que sus componentes son recurrentes. Se deberá separar en dos casos posibles.

Si la cadena producto es transitoria, su probabilidad de transición en  $n$  pasos de  $(i, i)$  a  $(j, j)$  es  $[p_{ij}(n)]^2$ , y en vista de 3.1.2, esta tiende a 0 cuando  $n \rightarrow \infty$ , y el resultado se prueba.

Si la cadena producto es recurrente, el argumento de acoplamiento usado en la demostración del teorema 4.2.2 aplica y sostiene que

$$\lim_{n \uparrow \infty} |\mu^t \mathbf{P}^n - \nu^t \mathbf{P}^n| = 0 \quad (4.17)$$

para distribuciones iniciales arbitrarias  $\mu$  y  $\nu$ . Si se supone ahora que 4.16 no es cierto para alguna  $i, j \in E$ . Se puede encontrar entonces una sucesión  $\{n_k\}_{k \geq 0}$  de enteros estrictamente creciente a  $\infty$  tal que

$$\lim_{k \uparrow \infty} p_{ij}(n_k) = \alpha > 0.$$

Para alguna  $i \in E$  fija escogida como arriba, la sucesión  $(\{p_{is}(n_k), s \in E\})_{k \geq 0}$  de vectores de  $[0, 1]^E$  es compacta en la topología de convergencia puntual [25]. Por lo

tanto, existe una subsucesión  $\{m_l\}_{l \geq 0}$  de enteros estrictamente creciente a  $\infty$  y un vector  $\{x_s, s \in E\} \in [0, 1]^E$  tal que para toda  $s \in E$ ,

$$\lim_{l \uparrow \infty} p_{is}(m_l) = x_s.$$

Ahora,  $x_j = \alpha > 0$ , y por lo tanto  $\{x_s, s \in E\}$  no es vacío. Ya que  $\sum_{s \in E} p_{is}(m_l) = 1$ , se sigue del lema de Fatou [5], que

$$\sum_{s \in E} x_s \leq 1.$$

En 4.17, se toma  $\mu = \delta_i$  y  $\nu^t = \delta_i^t \mathbf{P}$  para obtener, para toda  $s \in E$ ,

$$\lim_{l \uparrow \infty} |p_{is}(m_l) - p_{is}(m_l + 1)| = 0. \quad (4.18)$$

Como  $p_{is}(m_l + 1) = \sum_{k \in E} p_{ik}(m_l) p_{ks}$ . Por lo tanto, por la convergencia dominada en 4.18, se obtiene para toda  $s \in E$ ,

$$x_s = \sum_{k \in E} x_k p_{ks}.$$

En otras palabras,  $\{x_s, s \in E\}$  es una medida invariante de  $\mathbf{P}$  con masa finita, lo cual implica que  $\mathbf{P}$  es recurrente positiva, una contradicción. Por lo tanto, 4.16 se cumple.

### 4.2.3. Tasas de convergencia mediante acoplamiento

Sabiendo que una cadena de Markov ergódica converge al equilibrio, la siguiente pregunta es: ¿Qué tan rápido? El primer resultado a continuación no es explícito, pero se puede utilizar en principio para cadenas con un número infinito de estados.

Se utilizará la función  $o$ . Esta representa una función definida en una vecindad de cero y tal que  $\lim_{t \rightarrow 0} \frac{|o(t)|}{t} = 0$ .

**Teorema 4.2.5** (Tasa de Convergencia). Sea  $\tau$  un tiempo de acoplamiento con las características del 4.2.2, este satisface

$$E[\psi(\tau)] < \infty \quad (4.19)$$

para alguna función no decreciente  $\psi : \mathbb{N} \rightarrow \mathbb{R}_+$  tal que  $\lim_{n \uparrow \infty} \psi(n) = \infty$ . Entonces para cualesquiera distribuciones iniciales  $\mu$  y  $\nu$

$$|\mu^t \mathbf{P}^n - \nu^t \mathbf{P}^n| = o\left(\frac{1}{\psi(n)}\right). \quad (4.20)$$

**Demostración.** Como  $\psi$  es no decreciente,  $\psi(\tau) \mathbb{1}_{\{\tau > n\}} \geq \psi(n) \mathbb{1}_{\{\tau > n\}}$ , y por lo tanto

$$\psi(n) P(\tau > n) \leq E[\psi(\tau) \mathbb{1}_{\{\tau > n\}}].$$

Ahora,

$$\lim_{n \uparrow \infty} E[\psi(\tau) \mathbb{1}_{\{\tau > n\}}] = 0$$



por la convergencia dominada [31], ya que

$$\lim_{n \uparrow \infty} \psi(\tau) \mathbb{1}_{\{\tau > n\}} = 0,$$

dado que  $\tau$  es finito, y  $\psi(\tau) \mathbb{1}_{\{\tau > n\}}$  está acotado por la variable aleatoria integrable  $\psi(\tau)$ .

El tiempo  $\tau$  en el 4.2.2 es el tiempo de entrada de la cadena producto en el conjunto diagonal de  $E \times E$ . En principio, la distribución de  $\tau$  se puede calcular explícitamente. Sin embargo, los cálculos reales suelen ser difíciles cuando el espacio de estados es infinito. Cuando el espacio de estados es finito, la convergencia es exponencial (geométrica). Se pueden encontrar aplicaciones de el teorema 4.2.5 en [21].

**Teorema 4.2.6.** Sea  $\mathbf{P}$  una matriz de transición ergódica sobre el espacio de estados finito  $E$ . Entonces para cualesquiera distribuciones iniciales  $\mu$  y  $\nu$ , se pueden construir dos CMH  $\{X_n\}_{n \geq 0}$  y  $\{Y_n\}_{n \geq 0}$  sobre  $E$  con la misma matriz de transición  $\mathbf{P}$ , y sus respectivas distribuciones iniciales  $\mu$  y  $\nu$ , de tal manera que se acoplan en un tiempo finito  $\tau$  tal que  $E[e^{\alpha\tau}] < \infty$  para alguna  $\alpha > 0$ .

## 4.3. Teoría de renovación a tiempo discreto

### 4.3.1. Ecuación de Renovación

En el enfoque analítico de las cadenas de Markov, la prueba de convergencia al estado estacionario de una CMH ergódica es una consecuencia de un resultado en series de potencia llamado teorema de renovación. Este resultado constituye el tema de la presente sección. Sin embargo, el teorema de renovación no se utilizará como paso esencial hacia el teorema de convergencia, sino que, por el contrario, se obtendrá como corolario de este último.

Se comienza con definiciones básicas. Sea  $\{S_n\}_{n \geq 1}$  una sucesión de v.a.i.i.d. con valores en  $\{1, 2, \dots, \infty\}$  y con distribución de probabilidad

$$P(S_1 = k) = f_k. \quad (4.21)$$

Se define para  $n \geq 0$ ,

$$R_{n+1} = R_n + S_{n+1}, \quad (4.22)$$

donde  $R_0$  es una variable aleatoria arbitraria con valores en  $\mathbb{N}$ .

**Definición 4.3.1.** La sucesión  $\{R_n\}_{n \geq 0}$  es llamada *sucesión de renovación retrasada* (por  $R_0$ ) con *distribución de renovación*  $\{f_k\}_{k \geq 1}$ . Si  $R_0 \equiv 0$ , se habla de una sucesión de renovación no retrasada, o simplemente, una sucesión de renovación. Si  $P(S_1 = \infty) = 0$ , la sucesión de renovación es llamada *sucesión de renovación propia*, y  $\{f_k\}_{k \geq 1}$  es llamada *distribución de renovación propia*. En otro caso, se habla de una *sucesión de renovación defectuosa* y de una *distribución de renovación defectuosa*.

La cantidad

$$\alpha = P(S_1 = \infty)$$

es el *defecto* de la distribución de renovación. El tiempo aleatorio  $R_k$  es el  $k$ -ésimo tiempo de renovación, y la sucesión  $\{S_n\}_{n \geq 1}$  es la *sucesión inter-renovación*.

**Definición 4.3.2.** Con la distribución de renovación  $\{f_k\}_{k \geq 1}$  está asociada la *ecuación de renovación*

$$u_n = v_n + \sum_{k=1}^n f_k u_{n-k} \quad (4.23)$$

(para  $n = 0$ , esto se reduce a  $u_0 = v_0$ ). La sucesión  $\{u_n\}_{n \geq 0}$  es la *sucesión desconocida*, y  $\{v_n\}_{n \geq 0}$  son los *datos*, una sucesión de números reales tal que

$$\sum_{k=0}^{\infty} |v_k| < \infty. \quad (4.24)$$

Ya que  $u_n$  puede ser calculada recursivamente como una función de  $u_0, \dots, u_{n-1}$ ,  $v_0, \dots, v_n$ , una solución de la ecuación de renovación siempre existe y es única.

**Ejemplo 4.3.1.** Se define el tiempo de vida  $L$  de una sucesión de renovación defectuosa por

$$L = \inf\{R_k : k \geq 0, S_{k+1} = \infty\}.$$

Este es el último tiempo de renovación en una distancia finita. Se mostrará que  $u_n = P(L > n)$  satisface una ecuación de renovación. Para esto, se escribe

$$\mathbb{1}_{\{L > n\}} = \mathbb{1}_{\{L > n\}} \mathbb{1}_{\{S_1 > n\}} + \mathbb{1}_{\{L > n\}} \mathbb{1}_{\{S_1 \leq n\}}.$$

Se observa que  $\{L > n, S_1 > n\} = \{n < S_1 < \infty\}$ . También, denotando por  $\bar{L}$  al tiempo de vida asociado con la sucesión de renovación  $\{R_{n+1} - R_1\}_{n \geq 0}$ , se tiene el conjunto identidad

$$\{L > n, S_1 \leq n\} = \{\bar{L} > n - S_1, S_1 > n\}.$$

Por lo tanto,

$$P(L > n) = P(n < S_1 < \infty) + P(\bar{L} > n - S_1, S_1 \leq n).$$

Ahora,  $L$  y  $\bar{L}$  tienen la misma distribución, y  $\bar{L}$  es independiente de  $S_1$ . Por lo tanto,

$$P(\bar{L} > n - S_1, S_1 \leq n) = \sum_{k=1}^n P(\bar{L} > n - k) P(S_1 = k) = \sum_{k=1}^n P(L > n - k) P(S_1 = k).$$

Esto muestra que  $u_n$  satisface la ecuación de renovación con  $v_n = P(n < S_1 < \infty)$ .

**Definición 4.3.3.** Se define la *sucesión de Dirac*  $\{\delta_n\}_{n \geq 0}$  por  $\delta_0 = 1$ ,  $\delta_n = 0$  para  $n \geq 1$ . Cuando los datos  $\{v_n\}_{n \geq 0}$  es la sucesión de Dirac, la ecuación de renovación es llamada la *ecuación de renovación básica*, y su solución la *solución fundamental*.

La solución fundamental será denotada por  $\{h_n\}_{n \geq 0}$ , y por lo tanto  $h_0 = 1$ , y para  $n \geq 1$ ,

$$h_n = \sum_{k=1}^n f_k h_{n-k}. \quad (4.25)$$

La solución fundamental tiene una interpretación simple.  $h_n$  es la probabilidad de que  $n$  sea un tiempo de renovación (se dice entonces, para resumir, « $n$  es una renovación»). Esto es suficiente para mostrar que  $u_n = P(n \text{ es una renovación})$  es la única solución de la ecuación de renovación básica. Claramente,  $u_0 = 1$ . También, haciendo  $A =$  la última renovación estrictamente antes de  $n$  es  $k$ .

$$\begin{aligned}
P(n \text{ es una renovación}) &= \sum_{k=0}^{n-1} P(n \text{ es una renovación}, A) \\
&= \sum_{i=0}^{\infty} \sum_{k=0}^{n-1} P(S_{i+1} = n - k, k = R_i) \\
&= \sum_{i=0}^{\infty} \sum_{k=0}^{n-1} P(S_{i+1} = n - k) P(k = R_i) \\
&= \sum_{k=0}^{n-1} P(S_1 = n - k) \left( \sum_{i=0}^{\infty} P(k = R_i) \right) \\
&= \sum_{k=0}^{n-1} P(S_1 = n - k) P(k \text{ es una renovación}) \\
&= \sum_{k=0}^{n-1} u_k f_{n-k} = \sum_{k=0}^n f_k u_{n-k}.
\end{aligned}$$

Por lo tanto,

$$h_k = P(k \text{ es una renovación}). \quad (4.26)$$

En particular, si  $v_n$  es el número de tiempos de renovación  $R_k$  en el intervalo  $[0, n]$ , entonces

$$v_n = \sum_{k=0}^n h_k.$$

Se define ahora la *convolución* de dos sucesiones reales  $\{x_n\}_{n \geq 0}$  y  $\{y_n\}_{n \geq 0}$  como la sucesión real  $\{z_n\}_{n \geq 0}$  definida por

$$z_n = \sum_{k=0}^n x_k y_{n-k}$$

Usando una notación conveniente, se puede escribir como  $z = x \cdot y$ .

**Teorema 4.3.1.** La ecuación de renovación

$$u_n = v_n + \sum_{k=1}^n f_k u_{n-k} \quad (4.27)$$

tiene como única solución a

$$u = h \cdot v. \quad (4.28)$$

**Demostración.** La existencia y unicidad ya han sido observadas. Para verificar que la solución anunciada es la correcta, se escribe la ecuación de renovación como  $u = v + f \cdot u$  (con  $f_0 = 0$ ) y la ecuación fundamental como  $h = \delta + f \cdot h$ . Usando 4.28 en la ecuación de renovación se obtiene

$$h \cdot v = v + f \cdot (h \cdot v)$$

lo cual es cierto, ya que el lado derecho es

$$v + (f \cdot h) \cdot v = v + (h - \delta) \cdot v = v + h \cdot v - \delta \cdot v,$$

esto es,  $h \cdot v$ , porque  $\delta \cdot v = v$ .

**Ejemplo 4.3.2.** Cuando la distribución de la inter-renovación típica es geométrica, i.e., para  $k \geq 1$ ,

$$P(S_1 = k) = p(1 - p)^{k-1},$$

la solución fundamental está dada por  $h_0 = 1$ , y

$$h_n = p,$$

para  $n \geq 1$ . La solución de la ecuación general de renovación es entonces

$$u_n = v_n + p(v_0 + \cdots + v_{n-1}).$$

Se observa que en este caso particular que como  $\lim_{n \uparrow \infty} v_n = 0$  en vista de 4.27,

$$\lim_{n \uparrow \infty} u_n = p \sum_{k=0}^{\infty} v_k = \frac{\sum_{k \geq 0} v_k}{\sum_{k \geq 1} k f_k}.$$

Este resultado puede ser generalizado con el teorema de renovación.

### 4.3.2. Teorema de Renovación

**Definición 4.3.4.** La distribución de renovación  $\{f_k\}_{k \geq 1}$  es llamada *reticular* (resp. *no reticular*) si

$$d = \text{mcd}\{k : k \geq 1, f_k > 0\} > 1$$

(resp.  $d = 1$ ). El entero  $d$  es llamado el *lapso* de la distribución de renovación.

**Teorema 4.3.2** (Teorema de Renovación). Sea  $\{f_k\}_{k \geq 1}$  una distribución de renovación propia y no reticular. Para la solución única de la ecuación de renovación que satisface 4.27, se cumple que

$$\lim_{n \uparrow \infty} u_n = \frac{\sum_{k \geq 0} v_k}{\sum_{k \geq 1} k f_k} \quad (4.29)$$

donde el radio de convergencia del lado derecho es 0 si  $\sum_{k \geq 1} k f_k = \infty$ .

**Demostración.** La demostración será en dos partes.

1.- Se asume que el resultado es cierto para la solución fundamental, esto es,

$$\lim_{n \uparrow \infty} h_n = \frac{1}{\sum_{k \geq 1} k f_k} = h_\infty. \quad (4.30)$$

De la expresión 4.28 de la solución en términos de la solución fundamental, se obtiene

$$\sum_{k=0}^n (h_{n-k} - h_\infty) v_k = u_n - h_\infty \sum_{k=0}^n v_k.$$

El resultado se sigue si se puede probar que el lado izquierdo de la ecuación de arriba converge a 0 cuando  $n \rightarrow \infty$ . En efecto, con

$$g(n, k) = (h_{n-k} - h_\infty) v_k \mathbb{1}_{\{k \leq n\}},$$

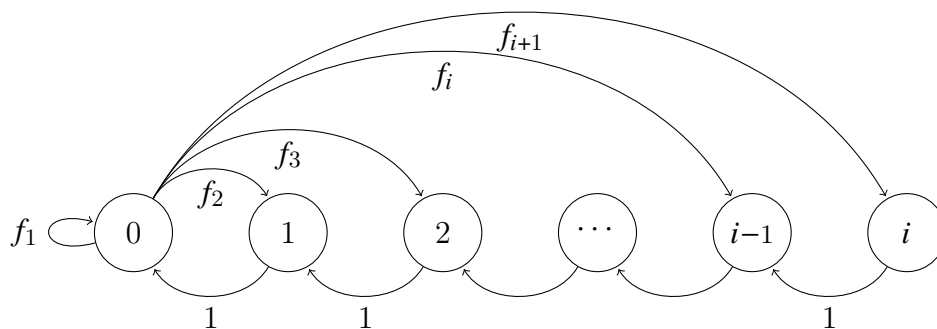
se tiene para una  $k$  fija,  $\lim_{n \uparrow \infty} g(n, k) = 0$ , y  $|g(n, k)| \leq |v_k|$ , donde  $\sum_{k \geq 0} |v_k| < \infty$ . Por lo tanto, por la convergencia dominada para series,

$$\lim_{n \uparrow \infty} \sum_{k \geq 0} g(n, k) = \sum_{k \geq 0} \lim_{n \uparrow \infty} g(n, k) = 0.$$

2.- Queda probar 4.29. Para esto, se usa una cadena de Markov con espacio de estados  $E = \mathbb{N}$  si el soporte de  $\{f_k\}_{k \geq 1}$  no es acotado, y el espacio de estados  $E = \{0, \dots, M-1\}$  si  $M < \infty$  es el valor más grande de  $S_1$ . Las entradas no nulas de la matriz de transición son

$$\begin{aligned} p_{i, i-1} &= 1, \quad i \geq 1, \\ p_{0, i} &= f_{i+1}, \quad i \geq 0. \end{aligned}$$

Con gráfica de transición



Se puede notar que esta es la gráfica de transición de la CMH  $\{X_n\}_{n \geq 0}$  definida por

$$X_n = \inf\{R_k : R_k \geq n\} - n.$$

Esta cadena es irreducible. La distribución del tiempo de retorno al estado 0 es

$$P_0(T_0 = n) = f_n.$$

El evento  $\{T_0 = n\}$  implica el evento  $\{X_n = 0\}$ , y por lo tanto  $P_0(X_n = 0) \geq P_0(T_0 = n)$ . Consecuentemente, el conjunto  $A = \{n \geq 1 : p_{00}(n) > 0\}$  contiene el conjunto  $B = \{n \geq 1 : f_n > 0\}$ , y entonces el m.c.d. de  $A$  es menor o igual que el m.c.d. de  $B$ . Por lo tanto, el m.c.d. de  $A$  es igual a 1, esto es, la cadena es aperiódica.

Como la distribución de renovación es asumida propia, se tiene que

$$P_0(T_0 < \infty) = \sum_{n \geq 1} f_n = 1,$$

y por lo tanto la cadena es recurrente. Si  $E_0[T_0] < \infty$ , esta es ergódica, y entonces

$$\lim_{n \uparrow \infty} p_{00}(n) = \pi_0 = \frac{1}{E_0[T_0]}.$$

Si la cadena no es ergódica pero es recurrente nula, entonces  $\lim_{n \uparrow \infty} p_{00}(n) = 0$  por el teorema de Orey 4.2.4. En ambos casos, como  $E_0[T_0] = \sum_{k \geq 1} k f_k$ ,

$$\lim_{n \uparrow \infty} p_{00}(n) = \frac{1}{\sum_{k \geq 1} k f_k}.$$

Y como  $p_{00}(n) = P(n \text{ es una renovación}) = h_n$ , la demostración se concluye.

**Teorema 4.3.3.** Bajo las mismas condiciones que el teorema 4.3.2, excepto que el lapso  $d$  de la distribución de renovación es ahora estrictamente más grande que 1, la solución de la ecuación de renovación 4.23 satisfaciendo 4.24 cumple, para toda  $r \in [0, d - 1]$ ,

$$\lim_{N \uparrow \infty} u_{r+Nd} = d \frac{\sum_{k \geq 0} v_{r+kd}}{\sum_{k \geq 1} k f_k}. \quad (4.31)$$

**Demostración.** Se observa que  $\{f_k\}_{k \geq 1}$  es propia y reticular con lapso  $d$ , la distribución  $\{f_{Nd}\}_{N \geq 1}$  es propia y no reticular. Por otro lado, la ecuación de renovación 4.23 se divide en  $d$  ecuaciones de renovación. La  $r$ -ésima ( $r \in [0, d - 1]$ ) es

$$u_{r+Nd} = v_{r+Nd} + \sum_{l=1}^N f_l d u_{r+Nd-l},$$

donde  $N$  es la variable tiempo. El teorema de renovación 4.3.2 puede ser aplicado a cada una, y se obtiene 4.31 después de observar que

$$\sum_{N=1}^{\infty} N f_{Nd} = \frac{1}{d} \sum_{N=1}^{\infty} N d f_{Nd} = \frac{1}{d} \sum_{k=1}^{\infty} k f_k.$$

### 4.3.3. Sucesiones de Renovación Defectuosas

**Teorema 4.3.4** (Teorema de Renovación Defectuoso). Sea la distribución de renovación defectuosa, y la sucesión  $\{v_n\}_{n \geq 1}$  que satisface

$$\lim_{n \uparrow \infty} v_n = v_{\infty} < \infty. \quad (4.32)$$

La solución de la ecuación de renovación entonces satisface

$$\lim_{n \uparrow \infty} u_n = \frac{v_{\infty}}{\alpha}, \quad (4.33)$$

donde  $\alpha = P(S_1 = \infty)$  es el defecto de la distribución de renovación.

**Demostración.** La CMH usada en la segunda parte de la demostración del teorema 4.3.2 ahora con espacio de estados  $E = \mathbb{N} \cup \{\infty\}$  y con  $\infty$  como un estado cerrado. Todos los estados además de  $\infty$  son transitorios. En particular, el número promedio de visitas a 0 es finito:

$$v_\infty = \sum_{k=0}^{\infty} h_k < \infty.$$

De la expresión de la solución

$$u_n = \sum_{k=0}^n h_k v_{n-k},$$

se obtiene por la convergencia dominada para series [31],

$$\lim_{n \uparrow \infty} u_n = \left( \sum_{k=0}^{\infty} h_k \right) v_\infty = v_\infty v_\infty.$$

Ahora, la probabilidad de  $n$  visitas a 0 es  $(1 - \alpha)^{n-1} \alpha$ , y por lo tanto el número promedio de visitas a 0 es  $v_\infty = \frac{1}{\alpha}$ .

**Teorema 4.3.5.** Sea una distribución de renovación no reticular y defectuosa, existe  $\gamma > 1$  tal que

$$\sum_{k=0}^{\infty} \gamma^k f_k = 1, \quad (4.34)$$

y la sucesión  $\{v_n\}_{n \geq 1}$  satisface

$$\sum_{k=0}^{\infty} \gamma^k |v_k| < \infty. \quad (4.35)$$

La solución de la ecuación de renovación entonces satisface

$$\lim_{n \uparrow \infty} \gamma^n u_n = \frac{\sum_{k=0}^{\infty} \gamma^k v_k}{\sum_{k=0}^{\infty} k \gamma^k f_k}. \quad (4.36)$$

**Demostración.** Se observa que si se define  $\tilde{f}_n = \gamma^n f_n$ ,  $\tilde{v}_n = \gamma^n v_n$ ,  $\tilde{u}_n = \gamma^n u_n$ , entonces

$$\tilde{u}_n = \tilde{v}_n + \sum_{k=1}^n \tilde{f}_k \tilde{u}_{n-k}.$$

Esta ecuación de renovación es no reticular y propia, y por lo tanto el resultado anunciado se sigue del teorema de renovación.

Una consecuencia de 4.34 es la caída exponencial de la distribución de renovación. Esto muestra en particular que 4.34 es una suposición que no siempre se cumple.

**Ejemplo 4.3.3.** Se toma la situación como la del teorema 4.3.4, donde además se supone que la distribución de renovación no es reticular y, además, 4.3.4 es verdadera para algunas  $\gamma > 1$ . Se busca comprender cómo  $u_n$  tiende a  $u_\infty$ . Para esto se define  $\hat{u}_n = u_n - u_\infty$ . Reescribiendo la ecuación de renovación para  $u_n$  como

$$u_n - u_\infty = v_n - u_\infty + \sum_{k=1}^n f_k(u_{n-k} - u_\infty) + u_\infty \sum_{k=1}^n f_k,$$

se observa que  $\hat{u}_n$  satisface la ecuación de renovación con

$$\hat{v}_n = v_n - u_\infty P(S_1 > n).$$

Se puede por lo tanto aplicar el teorema 4.3.5 para obtener, después de reordenar,

$$\lim_{n \uparrow \infty} \gamma^n \left( u_n - \frac{v_\infty}{\alpha} \right) = \frac{1}{\gamma} \left\{ \frac{\sum_{k=0}^{\infty} \gamma^k v_k}{\sum_{k=0}^{\infty} \gamma^k P(S_1 > k)} - \frac{v_\infty}{P(S_1 = \infty)} \right\}$$

Una ecuación de renovación *excesiva* es aquella en la que  $\sum_{k=0}^{\infty} f_n > 1$ . El teorema 4.3.5 tiene entonces una contra parte obvia. Sin embargo, se tiene en cuenta que en el caso 4.34 siempre se tiene una solución  $\gamma$ , y está en  $(0, 1)$ .

**Ejemplo 4.3.4.** En cada momento  $n \in \mathbb{Z}$ , nace un número promedio  $u_n$  de hijas. Cada una de ellas da a luz independientemente de las demás mujeres. El número promedio de hijas de cualquier mujer en el  $k$ -ésimo año de su vida,  $k \geq 1$ , es  $f_k$ . En el tiempo 0 la población tiene  $\alpha(i)$  mujeres de edad  $i$ . Expresando que  $u_n$  es la suma de  $v_n$ , el número promedio de hijas nacidas en el momento  $n$  de madres nacidas en el momento 0 o antes, y de  $r_n$ , el número promedio de hijas nacidas en el momento  $n$  de madres nacidas estrictamente después del tiempo 0 y hasta el tiempo  $n$ , se obtiene la ecuación de renovación con sucesión de datos

$$v_n = \sum_{i=0}^{\infty} \alpha(i) f_{n+1}.$$

En este contexto, la ecuación de renovación es conocida como la *ecuación de Lotka-Volterra*. Denotando por

$$\rho = \sum_{k=1}^{\infty} f_n$$

al número promedio de hijas de una mujer determinada, y suponiendo que este número es positivo y finito. Suponiendo también que es diferente de 1. Se asume que existe  $\gamma$  definida por 4.34 y que la distribución de renovación no es reticular. Denotando con  $C$  el lado derecho de 4.36,

$$\lim_{n \uparrow \infty} \gamma^n u_n = C.$$

Se nota que  $\gamma < 1$  si  $\rho > 1$ , y  $\gamma > 1$  si  $\rho < 1$ . El primer caso corresponde a la explosión exponencial, mientras que el segundo es la extinción exponencial.



## 4.4. Procesos Regenerativos

### 4.4.1. Ecuación de Renovación de un Proceso Regenerativo

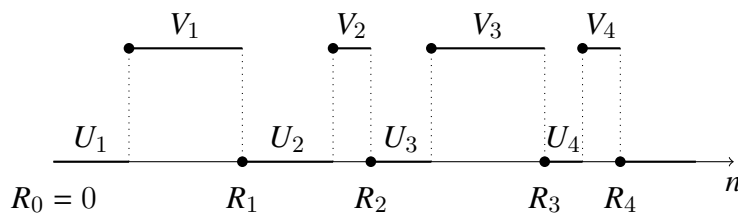
En las líneas introductorias de la sección anterior, se mencionó que el teorema de renovación puede probarse independientemente de los resultados básicos de la convergencia al estado estacionario, y que incluso podría usarse para probar tales resultados de convergencia. Por lo tanto, parece que el enfoque probabilístico de la convergencia otorga un estatus insignificante al teorema de la renovación, que disfruta de una posición central en el enfoque analítico. Sin embargo, el teorema de la renovación sigue siendo indispensable para demostrar la convergencia al equilibrio de los procesos estocásticos de naturaleza más general que las cadenas de Markov homogéneas, es decir, los procesos regenerativos. La característica común que comparten estos procesos con las cadenas de Markov homogéneas recurrentes es la existencia de ciclos regenerativos.

**Definición 4.4.1.** Sea  $\{Z_n\}_{n \geq 0}$  un proceso estocástico con valores en un espacio arbitrario  $E$  y sea  $\{R_n\}_{n \geq 0}$  una sucesión de renovación retrasada. El proceso  $\{Z_n\}_{n \geq 0}$  se dice que es regenerativo respecto de la sucesión de renovación  $\{R_n\}_{n \geq 0}$  si para toda  $k \geq 0$ ,  $\{Z_{n+R_k}\}_{n \geq 0}$  es independiente de  $R_0, S_1, \dots, S_k$  y tiene la misma distribución que  $\{Z_{n+R_0}\}_{n \geq 0}$ .

Se hace notar que la definición no requiere que  $\{Z_{n+R_k}\}_{n \geq 0}$  sea independiente de  $\{Z_n, n \in [0, R_k - 1]\}$ , a pesar de que en varios ejemplos esto se satisface. La libertad resultante de las condiciones de la definición 4.4.1 puede ser bastante útil.

**Ejemplo 4.4.1.** Sea  $\{X_n\}_{n \geq 0}$  una CMH irreducible recurrente, con distribución inicial arbitraria. Sea  $\{R_n\}_{n \geq 0}$  la sucesión de tiempos de *golpe* al estado 0. El teorema del ciclo regenerativo 2.7.4 dice que  $\{X_n\}_{n \geq 0}$  es regenerativo respecto de  $\{R_n\}_{n \geq 0}$ .

**Ejemplo 4.4.2.** Sea  $\{U_n\}_{n \geq 1}$  y  $\{V_n\}_{n \geq 1}$  dos sucesiones independientes de v.a.i.i.d. con valores en  $\mathbb{N}$ . Se define la sucesión  $\{S_n\}_{n \geq 1}$  por  $S_n = U_n + V_n$ , y sea  $\{R_n\}_{n \geq 0}$  la sucesión de renovación no retrasada asociada ( $R_0 \equiv 0$ ). Se define un proceso  $\{Z_n\}_{n \geq 0}$  con valores en  $\{0, 1\}$  como



Entonces  $\{Z_n\}_{n \geq 0}$  es un proceso regenerativo con respecto de  $\{R_n\}_{n \geq 0}$ .

Los procesos regenerativos generan ecuaciones de renovación y son la principal motivación para el estudio de dichas ecuaciones. Por ejemplo, si  $f : E \rightarrow \mathbb{R}$  es una función no negativa, y si  $\{Z_n\}_{n \geq 0}$  es un proceso regenerativo con valores en  $E$  con respecto a la sucesión de renovación no retrasada  $\{R_n\}_{n \geq 0}$ , entonces la sucesión  $\{u_n\}_{n \geq 0}$ , donde  $u_n = E[f(Z_n)]$ , satisface una ecuación de renovación. En efecto,

$$E[f(Z_n)] = E[f(Z_n) \mathbb{1}_{\{n < S_1\}}] + E[f(Z_n) \mathbb{1}_{\{n \geq S_1\}}],$$

y haciendo,  $\tilde{Z}_n = Z_{n+S_1}$ , se tiene

$$\begin{aligned}
E[f(Z_n)\mathbb{1}_{\{n \geq S_1\}}] &= E[f(\tilde{Z}_{n-S_1})\mathbb{1}_{\{n \geq S_1\}}] \\
&= \sum_{k=1}^{\infty} E[f(\tilde{Z}_{n-S_1})\mathbb{1}_{\{n \geq S_1\}}\mathbb{1}_{\{S_1=k\}}] \\
&= \sum_{k=1}^n E[f(\tilde{Z}_{n-k})\mathbb{1}_{\{S_1=k\}}] \\
&= \sum_{k=1}^n E[f(\tilde{Z}_{n-k})]P(S_1 = k) \\
&= \sum_{k=1}^n E[f(Z_{n-k})]P(S_1 = k)
\end{aligned}$$

donde la independencia de  $S_1$  y  $\{\tilde{Z}_n\}_{n \geq 0}$ , así como el supuesto de igual distribución de  $\{\tilde{Z}_n\}_{n \geq 0}$  y  $\{Z_n\}_{n \geq 0}$ , han sido tomados en cuenta. Por lo tanto,

$$E[f(Z_n)] = E[f(Z_n)\mathbb{1}_{\{n < S_1\}}] + \sum_{k=1}^n E[f(Z_{n-k})]P(S_1 = k),$$

la cual es precisamente la ecuación de renovación con sucesión de datos

$$v_n = E[f(Z_n)\mathbb{1}_{\{n < S_1\}}].$$

#### 4.4.2. Teorema Regenerativo

Se observa que

$$\sum_{n=0}^{\infty} |v_n| = \sum_{n=0}^{\infty} |E[f(Z_n)\mathbb{1}_{\{n < S_1\}}]| \leq E\left[\sum_{n=0}^{\infty} |f(Z_n)|\mathbb{1}_{\{n < S_1\}}\right] = E\left[\sum_{n=0}^{S_1-1} |f(Z_n)|\right].$$

Por lo tanto, por el teorema de renovación, se tiene lo siguiente:

**Teorema 4.4.1** (Teorema Regenerativo). Sea  $\{Z_n\}_{n \geq 0}$  un proceso regenerativo no retrasado ( $R_0 = 0$ ) y sea  $f : E \rightarrow \mathbb{R}$  tal que

$$E\left[\sum_{n=0}^{S_1-1} |f(Z_n)|\right] < \infty. \quad (4.37)$$

Si la distribución de  $S_1$  es propia y no reticular, entonces

$$\lim_{n \uparrow \infty} E[f(Z_n)] = \frac{E[\sum_{n=0}^{S_1-1} f(Z_n)]}{E[S_1]}. \quad (4.38)$$

La última formula es conocida como *formula de Smith*.

**Ejemplo 4.4.3.** Esta es una continuación del ejemplo 4.4.2. Se asume que  $S_1 = U_1 + V_1$  es propia y no reticular. Aplicando el teorema regenerativo con  $f(z) = \mathbb{1}_{\{0\}}(z)$ , y asumiendo  $E[U_1] < \infty$ , se encuentra que

$$\lim_{n \uparrow \infty} P(Z_n = 0) = \frac{E[U_1]}{E[U_1] + E[V_1]},$$

ya que

$$E[f(Z_n)] = E[\mathbb{1}_{\{0\}}(Z_n)] = P(Z_n = 0),$$

y

$$\sum_{n=0}^{S_1-1} \mathbb{1}_{\{0\}}(Z_n) = U_1.$$

**Ejemplo 4.4.4.** Se considera la sucesión de renovación de la definición 4.3.1, con  $R_0 = 0$ . Se define para cada  $n \geq 0$  el tiempo de recurrencia hacia atrás  $B_n$  y el tiempo de recurrencia hacia adelante  $F_n$  por

$$B_n = n - L_n, \quad F_n = N_n - n,$$

donde

$$L_n = \sup\{R_k : k \geq 0, R_k \leq n\}$$

y

$$N_n = \inf\{R_k : k > 0, R_k > n\}.$$

En particular, si  $n = R_m$  para alguna  $m$ , entonces  $B_n = 0$  y  $F_n = R_{m+1} - R_m = S_{m+1}$ . Se observa que  $F_n \geq 1$  para toda  $n \geq 0$ . También, si  $n \in [R_m, R_{m+1})$ , entonces  $B_n + F_n = S_{m+1}$ .



El teorema regenerativo con  $Z_n = (B_n, F_n)$  y  $f(Z_n) = \mathbb{1}_{\{B_n=i\}} \mathbb{1}_{\{F_n=j\}}$  de como resultado, debido a que la distribución de  $S_1$  es propia y no reticular,

$$\lim_{n \uparrow \infty} P(B_n = i, F_n = j) = \frac{P(S_1 = i + j)}{E[S_1]}. \quad (4.39)$$

En efecto la suma  $\sum_{n=0}^{S_1-1} \mathbb{1}_{\{B_n=i, F_n=j\}}$  tiene a lo más un término distinto de cero, en tal caso esta es igual a 1. Para este término, correspondiente al índice  $n = n_0$ ,  $B_{n_0} + F_{n_0} = S_1 = i + j$ . Por lo tanto la suma es igual a  $\mathbb{1}_{\{S_1=i+j\}}$ .

Sumando 4.39 de  $j = 1$  a  $\infty$ , y recordando que  $F_n \geq 1$ , se obtiene

$$\lim_{n \uparrow \infty} P(B_n = i) = \frac{P(S_1 > i)}{E[S_1]}. \quad (4.40)$$

Similarmente, para el tiempo de recurrencia hacia adelante,

$$\lim_{n \uparrow \infty} P(F_n = j) = \frac{P(S_1 \geq j)}{E[S_1]}. \quad (4.41)$$

Los roles de  $B_n$  y  $F_n$  no son simétricos. Para tener simetría, se debe considerar  $B_n$  y  $F'_n = F_n - 1$ . Entonces

$$\lim_{n \uparrow \infty} P(F'_n = j) = \frac{P(S_1 > j)}{E[S_1]}. \quad (4.42)$$

Como  $B_n + F_n = S_m$  para alguna  $m$  determinada por la condición  $n \in [R_m, R_{m+1})$ , se podría esperar que  $P(B_n + F_n = k) = P(S_1 = k)$ . Pero esto es en general falso, y constituye la aparente *paradoja de tiempos recurrentes*. Es cierto que  $P(B_n + F_n = k) = P(S_m = k)$ , pero  $m$  es aleatoria, y por lo tanto no existe razón de por qué  $S_m$  debería tener la misma distribución que  $S_1$ . Como un hecho de importancia,

$$\begin{aligned} \lim_{n \uparrow \infty} P(B_n + F_n = k) &= \lim_{n \uparrow \infty} \sum_{\substack{i,j \\ i+j=k}} P(B_n = i, F_n = j) \\ &= \sum_{\substack{i,j \\ i+j=k}} \frac{P(S_1 = i + j)}{E[S_1]} = \frac{kP(S_1 = k)}{E[S_1]}. \end{aligned}$$

**Teorema 4.4.2.** Sea  $\{Z_n\}_{n \geq 0}$  un proceso regenerativo posiblemente retrasado (recordar, sin embargo, que  $R_0 < \infty$ ). Sea  $f : E \rightarrow \mathbb{R}$  tal que

$$\lim_{n \uparrow \infty} E[f(Z_n) \mathbb{1}_{\{n < R_0\}}] = 0 \quad (4.43)$$

y

$$E \left[ \sum_{k=R_0}^{R_1-1} |f(Z_k)| \right] < \infty. \quad (4.44)$$

Entonces, si la distribución de renovación es propia y no reticular,

$$\lim_{n \uparrow \infty} E[f(Z_n)] = \frac{E[\sum_{k=R_0}^{R_1-1} f(Z_k)]}{E[S_1]}. \quad (4.45)$$

**Demostración.** Es suficiente mostrar que el límite de  $E[f(Z_n) \mathbb{1}_{\{n \geq R_0\}}]$  es igual al lado derecho de 4.45. Se define  $\{\tilde{Z}_n\}_{n \geq 0} = \{Z_{n+R_0}\}_{n \geq 0}$ , y se observa que este proceso es un proceso regenerativo no retrasado con respecto de  $\{R_n - R_0\}_{n \geq 1}$  que es propio y no reticular. Se tiene que

$$E[f(Z_n) \mathbb{1}_{\{n \geq R_0\}}] = E[f(\tilde{Z}_{n-R_0}) \mathbb{1}_{\{n \geq R_0\}}] = \sum_{k=0}^n E[f(\tilde{Z}_{n-k})] P(R_0 = k).$$

Por el teorema 4.4.1, se tiene que

$$\lim_{n \uparrow \infty} E[f(\tilde{Z}_n)] = \frac{E[\sum_{k=R_0}^{R_1-1} f(Z_k)]}{E[R_1 - R_0]},$$

y por lo tanto, por el teorema de convergencia dominada para series [31],

$$\lim_{n \uparrow \infty} E[f(\tilde{Z}_{n-k})]P(R_0 = k) = \frac{E[\sum_{k=R_0}^{R_1-1} f(Z_k)]}{E[R_1 - R_0]}.$$

Un caso útil cuando se satisface 4.43 es el de  $f$  acotada usando la convergencia dominada.

## 4.5. Vida antes de la absorción

### 4.5.1. Estancias infinitas

Se mencionó en 3.1.3 que el espacio de estados  $E$  puede ser descompuesto como

$$E = T \cup \left( \bigcup_j R_j \right)$$

donde  $R_1, R_2, \dots$  son las clases recurrentes disjuntas y  $T$  es la colección de estados transitorios. (Se tiene en cuenta que el número de clases recurrentes así como el número de estados transitorios puede ser infinito.) De la discusión en 3.1.3, la matriz de transición puede dividirse en bloques como

$$\mathbf{P} = \begin{matrix} & \begin{matrix} R_1 & R_2 & \dots & T \end{matrix} \\ \begin{matrix} R_1 \\ R_2 \\ \vdots \\ T \end{matrix} & \begin{pmatrix} \mathbf{P}_1 & 0 & 0 & 0 \\ 0 & \mathbf{P}_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ B(1) & B(2) & \dots & \mathbf{Q} \end{pmatrix} \end{matrix} \quad (4.46)$$

o en una notación condensada,

$$\mathbf{P} = \begin{pmatrix} D & 0 \\ B & Q \end{pmatrix}. \quad (4.47)$$

La forma 4.46 explica el hecho de que no se puede pasar de un estado en una clase recurrente dada a ningún estado que no pertenezca a esta clase recurrente. En otras palabras, una clase recurrente es cerrada.

Una pregunta típica es: ¿Cuál es la probabilidad de ser absorbido por una clase recurrente dada cuando se parte de un estado transitorio dado? Este tipo de problema ya se abordó en el capítulo 2 para introducir el análisis del primer paso. Condujo a sistemas de ecuaciones lineales con condiciones de fronteras, cuya solución fue única, debido a la finitud del espacio de estados. Con un espacio de estados infinito, el problema de la unicidad no puede pasarse por alto, y el problema de la absorción se reconsiderará con esto en mente, y también con la intención de encontrar expresiones algebraicas matriciales generales para las soluciones. Otro fenómeno que no se manifiesta en el caso finito es la posibilidad, cuando el conjunto de estados transitorios es infinito, de nunca ser absorbido por el conjunto recurrente. Consideraremos este problema primero, y luego procederemos a derivar la distribución del tiempo hasta

la absorción por el conjunto recurrente y la probabilidad de ser absorbido por una clase recurrente dada.

Sea  $A$  un subconjunto del espacio de estados  $E$  (es típico que se use como el conjunto de estados transitorios, pero no necesariamente). El objetivo es calcular para cualquier estado inicial  $i \in A$  la probabilidad de permanecer para siempre en  $A$ ,

$$v(i) = P_i(X_r \in A : r \geq 0). \quad (4.48)$$

Definiendo

$$v_n(i) = P_i(X_1 \in A, \dots, X_n \in A),$$

se tiene que, por el teorema 1.1.1,

$$\lim_{n \uparrow \infty} \downarrow v_n(i) = v(i).$$

Como para toda  $j \in A$

$$P_i(X_1 \in A, \dots, X_{n-1} \in A, X_n = j) = \sum_{i_1 \in A} \cdots \sum_{i_{n-1} \in A} p_{ii_1} \cdots p_{i_{n-1}j}$$

es el término general  $q_{ij}(n)$  de la  $n$ -ésima iteración de  $\mathbf{Q}$ , la restricción de  $\mathbf{P}$  al conjunto  $A$ . Por lo tanto,

$$v_n(i) = \sum_{j \in A} q_{ij}(n).$$

Esta ecuación se lee como

$$v_n = \mathbf{Q}^n \mathbf{1}_A, \quad (4.49)$$

donde  $\mathbf{1}_A$  es el vector columna indexado por  $A$  con todas sus entradas igual a 1. De esta ecuación se obtiene

$$v_{n+1} = \mathbf{Q}v_n, \quad (4.50)$$

y por la convergencia dominada  $v = \mathbf{Q}v$ . Más aún,  $\mathbf{0}_A \leq v \leq \mathbf{1}_A$  donde  $\mathbf{0}_A$  es el vector columna indexado por  $A$  con todas sus entradas igual a 0. El resultado anterior puede ser refinado.

**Teorema 4.5.1** (Estancia Infinita). El vector  $v$  es la solución maximal de

$$v = \mathbf{Q}v, \quad (4.51)$$

$$\mathbf{0}_A \leq v \leq \mathbf{1}_A. \quad (4.52)$$

Además, ya sea que  $v = \mathbf{0}_A$  ó  $\sup_{i \in A} v(i) = 1$ .

**Demostración.** Solo queda demostrar que  $v$  es maximal y la última afirmación. Se considera un vector  $u$  indexado por  $A$  tal que  $u = \mathbf{Q}u$  y  $\mathbf{0}_A \leq u \leq \mathbf{1}_A$ . Iterar  $u = \mathbf{Q}u$  produce  $u = \mathbf{Q}^n u$ , y  $u \leq \mathbf{1}_A$  implica que  $\mathbf{Q}^n u \leq \mathbf{Q}^n \mathbf{1}_A = v_n$ . Por lo tanto,  $u \leq v_n$ , lo cual da  $u \leq v$  al pasar por el límite.

Para probar la última afirmación del teorema, sea  $c = \sup_{i \in A} v(i)$ . De  $v \leq c \mathbf{1}_A$ , se obtiene  $v \leq cv_n$  como arriba, y por lo tanto, en el límite,  $v \leq cv$ . Esto implica que  $v = \mathbf{0}_A$  ó  $c = 1$ .

La ecuación  $v = \mathbf{Q}v$  en 4.51 se lee como

$$v(i) = \sum_{j \in A} p_{ij}v(j) \quad (i \in A). \quad (4.53)$$

El análisis del primer paso hace a 4.52 una condición necesaria. Sin embargo, esto no ayuda a determinar qué solución para 4.52 elegir, en el caso de que existan varias.

**Ejemplo 4.5.1.** La matriz de transición de la caminata aleatoria sobre  $\mathbb{N}$  del ejemplo 3.3.1,

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

donde  $p \in (0, 1)$ , es irreducible. Intuitivamente, si  $p > q$ , se da un paso a la derecha, y se esperaría que la cadena fuera transitoria. Esto se puede probar formalmente mostrando que la probabilidad  $v(i)$  de nunca visitar el estado 0 cuando se inicia del estado  $i \geq 1$  es estrictamente positiva. Para poder aplicar el teorema 4.5.1 con  $A = \mathbb{N} - \{0\}$ , se debe encontrar la solución general de  $u = \mathbf{Q}u$ . Esta ecuación es

$$\begin{aligned} u(1) &= pu(2), \\ u(2) &= qu(1) + pu(3), \\ u(3) &= qu(2) + pu(4), \\ &\dots \end{aligned}$$

y su solución general es

$$u(i) = u(1) \sum_{j=0}^{i-1} \left(\frac{q}{p}\right)^j.$$

El valor más grande de  $u(1)$  respetando la restricción  $u(i) \in [0, 1]$  es  $u(1) = 1 - \left(\frac{q}{p}\right)$ . La solución  $v(i)$  se obtiene asignando este valor, y por lo tanto

$$v(i) = 1 - \left(\frac{q}{p}\right)^i. \quad (4.54)$$

Si el conjunto de estados transitorios  $T$  es finito, la probabilidad de estancia infinita en  $T$  es nula. Esto se deduce de la aplicación de la ecuación 4.49 con  $A = T$ . De hecho, con la notación de 4.47,

$$\mathbf{P}^n = \begin{pmatrix} D^n & 0 \\ B_n & \mathbf{Q}^n \end{pmatrix}, \quad (4.55)$$

y por lo tanto el término general de  $\mathbf{Q}^n$  es  $p_{ij}(n)$ . Pero se sabe que para cualquier estado transitorio  $j$ ,  $\lim_{n \uparrow \infty} p_{ij}(n) = 0$  para toda  $i \in E$ , y por lo tanto si  $T$  es finita,

$$v(i) = \lim_{n \uparrow \infty} \sum_{j \in T} p_{ij}(n) = 0.$$

**Ejemplo 4.5.2.** Esta CMH se presentó en el ejemplo 2.2.2. Su espacio de estados  $E = \mathbb{N}$ , y su matriz de transición es

$$\mathbf{P} = \begin{pmatrix} a_0 & a_1 & a_2 & a_3 & \cdots \\ a_0 & a_1 & a_2 & a_3 & \cdots \\ 0 & a_0 & a_1 & a_2 & \cdots \\ 0 & 0 & a_0 & a_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

donde  $\{a_i : i \geq 0\}$  es una distribución de probabilidad sobre  $\mathbb{N}$  de media

$$\rho = \sum_{k=0}^{\infty} k a_k.$$

Se asume que  $a_0 > 0$  y  $a_0 + a_1 < 1$ , entonces la cadena es irreducible. Se puede usar el teorema 4.5.1 para probar que si  $\rho \leq 1$ , la cadena es recurrente. Para esto se considera la matriz

$$\mathbf{Q} = \begin{pmatrix} a_1 & a_2 & a_3 & \cdots \\ a_1 & a_2 & a_3 & \cdots \\ a_0 & a_1 & a_2 & \cdots \\ 0 & a_0 & a_1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

la cual es la restricción de  $\mathbf{P}$  a  $A_i = \{i+1, i+2, \dots\}$  para cualquier  $i \geq 0$ . Así, la solución maximal de  $v = \mathbf{Q}v$ ,  $\mathbf{0}_A \leq v \leq \mathbf{1}_A$  (donde  $A = A_i$ ), tiene en vista del teorema 4.5.1 las siguientes dos interpretaciones. Primeramente, para  $i \geq 1$ ,

$$v(i) = P_i(X_n \geq 1, n \geq 0), \quad (4.56)$$

esto es,  $v(i)$  es la probabilidad de nunca visitar 0 cuando se inicia desde  $i \geq 1$ . En segundo lugar,  $v(1)$  es la probabilidad de nunca visitar  $\{0, 1, \dots, i\}$  cuando se comienza desde  $i+1$ . Ya que cuando se empieza desde  $i+1$ , la cadena visita  $\{0, 1, \dots, i\}$  si y sólo si esta visita  $i$ ,  $v(1)$  es la probabilidad de nunca visitar  $i$  cuando se empieza desde  $i+1$ .

Se puede expresar la probabilidad de visitar 0 cuando se comienza desde  $i+1$  como

$$1 - v(i+1) = (1 - v(1))(1 - v(i)),$$

ya que para poder ir desde  $i+1$  a 0 se tiene que llegar primero a  $i$ , y entonces ir a 0. Por lo tanto, con

$$v(1) = 1 - \beta,$$

para toda  $i \geq 1$ ,

$$v(i) = 1 - \beta^i. \quad (4.57)$$

Esta solución depende de un parámetro  $\beta \in [0, 1]$ . Para determinar  $\beta$ , se escribe la primera ecuación de  $v = \mathbf{Q}v$ :

$$v(1) = a_1 v(1) + a_2 v(2) + \cdots,$$



esto es,

$$(1 - \beta) = a_1(1 - \beta) + a_2(1 - \beta^2) + \dots .$$

Tomando  $\sum_{i \geq 0} a_i = 1$  en cuenta, esto se reduce a

$$\beta = g(\beta) \tag{4.58}$$

donde  $g(z)$  es la función generadora de la distribución  $(a_k, k \geq 0)$ . También, cualquier otra ecuación de  $v = \mathbf{Q}v$  se reduce a 4.58.

Recordando el teorema 1.5.1, el cual dice que bajo las condiciones de una cadena irreducible  $a_0 > 0$ ,  $a_0 + a_1 < 1$ , 4.58 tiene sólo una solución en  $[0, 1]$ , la cual es  $\beta = 1$  si  $\rho \leq 1$ , mientras que si  $\rho > 1$ , se tienen dos soluciones en  $[0, 1]$ , esta probabilidad es  $\beta = 1$  y  $\beta = \beta_0 \in (0, 1)$ .

El teorema 4.5.1 dice que se debe de tomar la  $\beta$  más pequeña que satisface 4.58. Por lo tanto, si  $\rho > 1$ , la probabilidad de visitar el estado 0 cuando se empieza desde el estado  $i \geq 1$  es  $1 - v(i) = \beta_0^i < 1$ , y por lo tanto la cadena es transitoria. Si  $\rho \leq 1$ , la última probabilidad es  $1 - v(i) = 1$ , y por lo tanto la cadena es recurrente.

Se conoce de 2.2.2, que para  $\rho = 1$ , no existe una distribución estacionaria. Por lo tanto,

Si  $\rho = 1$ , la cadena es recurrente nula.

Si  $\rho > 1$ , la cadena es transitoria.

El caso  $\rho < 1$  queda pendiente, donde la cadena es recurrente positiva.

## 4.5.2. Tiempo de absorción

Se continúa ahora con la determinación de la distribución de  $\tau$ , el tiempo de salir del conjunto transitorio  $T$ . El teorema 4.5.1 dice que  $v = \{v(i)\}_{i \in T}$ , donde  $v(i) = P_i(\tau = \infty)$ , es la solución más grande de  $v = \mathbf{Q}v$  sujeto a las limitaciones  $\mathbf{0}_T \leq v \leq \mathbf{1}_T$ , donde  $\mathbf{Q}$  es la restricción de  $\mathbf{P}$  al conjunto transitorio  $T$ .

La distribución de probabilidad de  $\tau$  cuando el estado inicial es  $i \in T$  se calcula a partir de la identidad

$$P_i(\tau = n) = P_i(\tau \geq n) - P_i(\tau \geq n+1)$$

y la observación  $\{\tau \geq n\} = \{X_{n-1} \in T\}$  para  $n \geq 1$ , de lo cual se obtiene, para  $n \geq 1$ ,

$$P_i(\tau = n) = P_i(X_{n-1} \in T) - P(X_n \in T) = \sum_{j \in T} (p_{ij}(n-1) - p_{ij}(n)).$$

Ahora,  $p_{ij}(n)$  es para  $i, j \in T$  el término general de  $\mathbf{Q}^n$ , y por lo tanto se tiene el siguiente resultado.

**Teorema 4.5.2** (Distribución del Tiempo de Absorción).

$$P_i(\tau = n) = \{(\mathbf{Q}^{n-1} - \mathbf{Q}^n)\mathbf{1}_T\}_i. \tag{4.59}$$

En particular, si  $P_i(\tau = \infty) = 0$ ,

$$P_i(\tau > n) = \{\mathbf{Q}^n \mathbf{1}_T\}_i. \tag{4.60}$$

En particular, cuando  $T$  es finito, para cualquier distribución  $v$  tal que  $v(T) = 1$ , se tiene

$$P_v(\tau > n) = v^t \mathbf{Q}^n \mathbf{1}_T.$$

## 4.6. Absorción

### 4.6.1. Matriz Fundamental

Los resultados de la sección anterior se refieren a la probabilidad de permanecer para siempre en el conjunto transitorio o, desde el punto de vista dual, a la probabilidad de nunca ser absorbido por el conjunto recurrente. Queda por calcular la probabilidad de ser absorbido por una determinada clase recurrente  $R_j$  al empezar de un estado inicial  $i \in T$ . Para ello, se necesita la noción de matriz fundamental, que está ligada a la de matriz potencia, y por tanto al número de visitas a un estado particular  $j$  cuando se parte de un estado  $i$ .

Como se verá, solo el caso  $i \in T, j \in T$  requiere análisis, y en este caso se busca el número de visitas a  $j \in T$  antes de la absorción por  $R = R_1 \cup R_2 \cup \dots$  cuando se empieza de  $i \in T$ .

Cuando se comienza de cualquier estado  $i$ , el número promedio de visitas a un estado arbitrario  $j$  es

$$E_i \left[ \sum_{n=0}^{\infty} \mathbb{1}_{\{X_n=j\}} \right] = \sum_{n=0}^{\infty} p_{ij}(n), \quad (4.61)$$

y esto es el elemento  $(i, j)$  de la matriz potencia

$$\mathbf{G} = \sum_{n=0}^{\infty} \mathbf{P}^n. \quad (4.62)$$

Solo las parejas  $(i, j)$  donde  $i$  y  $j$  son transitorias requieren análisis, ya que en cualquier otra situación la cantidad nula o infinita 4.61. Por ejemplo, si  $i$  y  $j$  están en la misma clase recurrente, el número de visitas a  $j$  empezando de  $i$  es infinito, mientras que es cero si  $i$  y  $j$  pertenecen a diferentes clases recurrentes, o si  $i$  es un estado recurrente y  $j$  un estado transitorio. Si  $i$  es transitorio y  $j$  es recurrente, el término correspondiente es infinito si  $j$  es accesible desde  $i$ , y este es cero si  $j$  no es accesible desde  $j$ .

Para evaluar 4.62 para  $i \in T, j \in T$ , se usa la representación 4.47 de la matriz de transición, para obtener

$$\mathbf{G} = \begin{pmatrix} E & 0 \\ F & S \end{pmatrix} \quad (4.63)$$

donde

$$S = \sum_{n=0}^{\infty} \mathbf{Q}^n. \quad (4.64)$$

**Definición 4.6.1.** Una CMH con al menos un estado transitorio y un estado recurrente es llamada *absorbente*, y la matriz  $S$  es su *matriz fundamental*.

Se hace notar que existen dos tipos de matrices fundamentales, una para cadenas ergódicas y una para cadenas absorbentes.

Denotando por  $\mathbf{I}$  a la matriz identidad indexada por  $T$ , y observando que

$$(\mathbf{I} - \mathbf{Q}) \left( \sum_{n=0}^N \mathbf{Q}^n \right) = \mathbf{I} - \mathbf{Q}^{N+1},$$



### 4.6.2. Matriz de Absorción

Se busca calcular la probabilidad de absorción para una clase recurrente dada cuando se parte de un estado transitorio dado. Como se verá más adelante, basta con que se trate el caso en el que las clases recurrentes son instancias únicas. Por tanto, suponemos que la matriz de transición tiene la forma

$$\mathbf{P} = \begin{pmatrix} \mathbf{I} & 0 \\ \mathbf{B} & \mathbf{Q} \end{pmatrix}. \quad (4.67)$$

Sea  $f_{ij}$  la probabilidad de absorción de la clase recurrente  $R_j = \{j\}$  cuando se empieza del estado transitorio  $i$ . De 4.67,

$$\mathbf{P}^n = \begin{pmatrix} \mathbf{I} & 0 \\ L_n & \mathbf{Q}^n \end{pmatrix},$$

donde  $L_n = (\mathbf{I} + \mathbf{Q} + \cdots + \mathbf{Q}^n)\mathbf{B}$ . Por lo tanto,  $\lim_{n \uparrow \infty} L_n = \mathbf{SB}$ . Para  $i \in T$ , el término  $(i, j)$  de  $L_n$  es

$$L_n(i, j) = P(X_n = j | X_0 = i).$$

Ahora, si  $T_j$  es el primer tiempo de visita a  $R_j = \{j\}$  después del tiempo 0, entonces

$$L_n(i, j) = P_i(T_j \leq n),$$

ya que  $R_j = \{j\}$  es un estado cerrado. Haciendo a  $n$  tender a  $\infty$  se obtiene lo siguiente

**Teorema 4.6.2.** Se considera una CMH con matriz de transición  $\mathbf{P}$  como en 4.67. La probabilidad de absorción de una clase recurrente  $R_j = \{j\}$  empezando del estado transitorio  $i$  es

$$P_i(T_{R_j} < \infty) = (\mathbf{SB})_{ij}. \quad (4.68)$$

El caso general, donde las clases de recurrencia no son necesariamente instancias únicas, puede ser reducido al caso particular como sigue. Sea  $\mathbf{P}^*$  la matriz obtenida de la matriz de transición  $\mathbf{P}$  dada por 4.67, agrupando para cada  $j$  los estados de la clase recurrente  $R_j$  en un solo estado  $\hat{j}$ :

$$\mathbf{P}^* = \begin{matrix} & \hat{1} & \hat{2} & \cdots & T \\ \hat{1} & \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ b_1 & b_2 & \cdots & \mathbf{Q} \end{pmatrix} & & & \end{matrix} \quad (4.69)$$

donde  $b_{\hat{j}} = \mathbf{B}(j)\mathbf{1}_T$  es obtenida por la suma de las columnas de  $\mathbf{B}(j)$ , la matriz que consta de las columnas  $i \in R_j$  de  $\mathbf{B}$ .

La probabilidad  $f_{iR_j}$  de absorción por la clase  $R_j$  cuando se empieza de  $i \in T$  es igual a  $\hat{f}_{i\hat{j}}$ , la probabilidad de siempre visitar  $\hat{j}$  cuando se comienza de  $i$ , calculado para la matriz de transición  $\mathbf{P}^*$ .

**Ejemplo 4.6.2.** Se considera la cadena con espacio de estados  $E = \{1, 2, 3, 4, 5, 6, 7\}$  y matriz de transición

$$\mathbf{P} = \begin{pmatrix} 0.5 & 0.5 & & & & & \\ 0.8 & 0.2 & & & & & \\ & & 0 & 0.4 & 0.6 & & \\ & & 1 & 0 & 0 & & \\ & & 1 & 0 & 0 & & \\ 0.1 & 0 & 0.2 & 0.1 & 0.2 & 0.3 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0 & 0.1 & 0.2 & 0.4 \end{pmatrix}.$$

Esta tiene dos clases recurrentes  $R_1 = \{1, 2\}$ ,  $R_2 = \{3, 4, 5\}$  y una clase transitoria  $T = \{6, 7\}$ . Bajo las notaciones usadas en esta sección,

$$B_1 = \begin{pmatrix} 0.1 & 0 \\ 0.1 & 0.1 \end{pmatrix}, B_2 = \begin{pmatrix} 0.2 & 0.1 & 0.2 \\ 0.1 & 0 & 0.1 \end{pmatrix}, \mathbf{Q} = \begin{pmatrix} 0.3 & 0.1 \\ 0.2 & 0.4 \end{pmatrix},$$

y por lo tanto

$$b_1 = \begin{pmatrix} 0.1 \\ 0.2 \end{pmatrix}, b_2 = \begin{pmatrix} 0.5 \\ 0.2 \end{pmatrix}, \hat{B} = \begin{pmatrix} 0.1 & 0.5 \\ 0.2 & 0.2 \end{pmatrix}.$$

También,

$$\mathbf{P}^* = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0.1 & 0.5 & 0.3 & 0.1 \\ 0.2 & 0.2 & 0.2 & 0.4 \end{pmatrix}.$$

Los cálculos llevan a

$$S = (\mathbf{I} - \mathbf{Q})^{-1} = \begin{pmatrix} 1.5 & 0.25 \\ 0.5 & 1.75 \end{pmatrix}, S\hat{B} = \begin{pmatrix} 0.2 & 0.8 \\ 0.4 & 0.6 \end{pmatrix}.$$

En particular, la probabilidad de absorción para el estado transitorio 6 a la clase  $\{3, 4, 5\}$  es 0.8.

---

---

# CAPÍTULO 5

---

## APLICACIONES

### 5.1. Campos Aleatorios de Markov

#### 5.1.1. Vecindades y especificación local

La propiedad de Markov de una sucesión estocástica  $\{X_n\}_{n \geq 0}$  implica que para toda  $n \geq 1$ ,  $X_n$  es independiente de  $(X_k, k \notin \{n-1, n, n+1\})$  dado  $(X_{n-1}, X_{n+1})$ .

Si  $n$  es llamado sitio,  $X_n$  el valor del proceso en el sitio  $n$ , y el conjunto  $\{n-1, n, n+1\}$  la vecindad del sitio  $n$ , la propiedad anterior puede ser descrita como: Para toda  $n \geq 1$ , el valor en el sitio  $n$  es independiente de los valores en los sitios  $k \notin \{n-1, n, n+1\}$  dados los valores en la vecindad del sitio  $n$ . Esto es el punto de partida para una generalización natural de la propiedad de Markov.

**Definición 5.1.1** (Campo Aleatorio). Sea  $S$  un conjunto finito, con elementos denotados por  $s$  y llamados sitios, y sea  $\Lambda$  un conjunto finito llamado espacio de fases. Un campo aleatorio sobre  $S$  con fases en  $\Lambda$  es una colección  $X = \{X(s)\}_{s \in S}$  de variables aleatorias con valores en  $\Lambda$ .

Un campo aleatorio puede ser considerado como una variable aleatoria que toma valores en el *espacio de configuraciones*  $\Lambda^S$ . Una configuración  $x \in \Lambda^S$  es de la forma  $x = (x(s), s \in S)$ , donde  $x(s) \in \Lambda$  para toda  $s \in S$ . Para una configuración dada  $x$  y un subconjunto dado  $A \subset S$ , se define

$$x(A) = (x(s), s \in A),$$

la restricción de  $x$  a  $A$ . Si  $S \setminus A$  denota el complemento de  $A$  en  $S$ , se puede expresar  $x = (x(A), x(S \setminus A))$ . En particular, para una  $s \in S$  fija,  $x = (x(s), x(S \setminus s))$ , donde  $S \setminus s$  es una manera abreviada de escribir  $S \setminus \{s\}$ .

Los campos de Markov caracterizados por *interacciones locales* son de especial interés. Para establecer la propiedad de Markov para campos aleatorios, primero se necesita introducir una *topología* sobre los sitios.

**Definición 5.1.2** (Vecindades). Un sistema de vecindades sobre  $S$  es una familia  $N = \{\mathcal{N}_s\}_{s \in S}$  de subconjuntos de  $S$  tal que para toda  $s \in S$ ,

- (i)  $s \notin \mathcal{N}_s$ ,
- (ii)  $t \in \mathcal{N}_s \implies s \in \mathcal{N}_t$ .

El subconjunto  $\mathcal{N}_s$  es llamada la vecindad del sitio  $s$ . La pareja  $(S, N)$  es una *topología*. La *frontera* de  $A \subset S$  es, por definición, el conjunto  $\partial A = (\cup_{s \in A} \mathcal{N}_s) \setminus A$ .

A continuación,  $\tilde{\mathcal{N}}_s$  denota el conjunto  $\mathcal{N}_s \cup \{s\}$ .

**Definición 5.1.3** (Campo aleatorio de Markov). El campo aleatorio  $X$  es llamado un *campo aleatorio de Markov* respecto del sistema de vecindades  $N$  si para todos los sitios  $s \in S$  las variables aleatorias  $X(s)$  y  $X(S \setminus \tilde{\mathcal{N}}_s)$  son independientes dada  $X(\mathcal{N}_s)$ .

En otros términos:

$$P(X(s) = x(s) | X(S \setminus s) = x(S \setminus s)) = P(X(s) = x(s) | X(\mathcal{N}_s) = x(\mathcal{N}_s)) \quad (5.1)$$

para toda  $s \in S, x \in \Lambda^S$ . La propiedad 5.1 es claramente del tipo Markov: la distribución de la fase en el sitio está directamente influenciada solo por las fases de los sitios vecinos.

Se nota que cualquier campo aleatorio es de Markov respecto de la topología trivial, donde la vecindad de cualquier sitio es la totalidad del conjunto  $S$ . Sin embargo, como se hará evidente, los campos de Markov de interés (desde el punto de vista de modelar, simular, y optimizar) son aquellos con vecindades relativamente pequeñas.

**Definición 5.1.4** (Especificación local). La *característica local* del campo aleatorio de Markov en el sitio  $s$  es la función  $\pi^s = \Lambda^S \rightarrow [0, 1]$  definida por

$$\pi^s(x) = P(X(s) = x(s) | X(\mathcal{N}_s) = x(\mathcal{N}_s)). \quad (5.2)$$

La familia  $\{\pi^s\}_{s \in S}$  es llamada *especificación local* del campo aleatorio de Markov.

Esto se suele escribir como

$$\pi^s(x) = \pi(x(s) | x(\mathcal{N}_s))$$

con el fin de enfatizar el papel del sistema de vecindades.

La finitud del espacio de fases  $\Lambda$  no es hasta ahora esencial. Una extensión inmediata que no requiere un cambio en la notación es a un espacio de fases numerable. Los espacios de fases arbitrarios no presentan dificultad excepto para la notación, ya que  $\pi^s(x) = \pi(x(s) | x(\mathcal{N}_s))$ , considerado como una función de  $x(s)$  para una  $x(\mathcal{N}_s)$  fija, ahora se interpreta como una función de densidad con respecto a alguna medida. Se describirá brevemente un caso sencillo que contiene todo lo que se necesita en las aplicaciones.

Aquí el espacio de fases es  $\Lambda = F \times \mathbb{R}^k$ , donde  $F$  es un conjunto numerable. Así  $x = (u, v)$ , donde  $u = (u(s), s \in S)$ ,  $v = (v(s), s \in S)$ ,  $u(s) \in F$ ,  $v(s) \in \mathbb{R}^k$ .

En particular,  $X = (U, V) = \{(U(s), V(s))\}_{s \in S}$ . La característica local  $\pi^s$  es ahora interpretada de acuerdo a

$$P(U(s) = \alpha, V(s) \leq \beta | U(\mathcal{N}_s) = u(\mathcal{N}_s), V(\mathcal{N}_s) = v(\mathcal{N}_s)) = \int_{-\infty}^{\beta} \pi(\alpha, \gamma | u(\mathcal{N}_s), v(\mathcal{N}_s)) d\gamma,$$

donde la integral es una integral múltiple de orden  $k$ . Así  $\pi(\alpha, \gamma | u(\mathcal{N}_s), v(\mathcal{N}_s))$  es una función de densidad respecto de  $\gamma$  para una  $\alpha$  fija. La propiedad de Markov ahora se lee

$$P(U(s) = \alpha, V(s) \leq \beta | U(S \setminus s), V(S \setminus s)) = P(U(s) = \alpha, V(s) \leq \beta | U(\mathcal{N}_s), V(\mathcal{N}_s))$$

para toda  $\alpha \in F$ ,  $\beta \in \mathbb{R}^k$ ,  $s \in S$ .

**Definición 5.1.5.** La distribución de probabilidad  $\pi$  sobre el espacio de configuraciones finito  $\Lambda^S$ , donde  $S = \{1, 2, \dots, K\}$ , se dice que satisface la *condición de positividad* si para toda  $j \in S$ ,  $x_j \in \Lambda$ ,

$$(\pi_j(x_j) = 0) \implies (\pi(y_1, \dots, y_{j-1}, x_j, y_{j+1}, \dots, y_K) = 0) \quad (5.3)$$

para toda  $y_1, \dots, y_{j-1}, y_{j+1}, \dots, y_K \in \Lambda$ , donde  $\pi_j$  es la distribución marginal sobre el sitio  $j$ .

**Teorema 5.1.1.** Dos distribuciones de un campo aleatorio de Markov con un espacio de configuración definido  $\Lambda^S$  que satisfacen la condición de positividad y tienen la misma especificación local son idénticas.

**Demostración.** Enumerando  $S$  como en la definición 5.1.5. El resultado se sigue de la identidad

$$\pi(x) = \prod_{i=1}^K \frac{\pi(x_i | x_1, \dots, x_{i-1}, y_{i+1}, \dots, y_K)}{\pi(y_i | x_1, \dots, x_{i-1}, y_{i+1}, \dots, y_K)} \pi(y) \quad (5.4)$$

para cualquier  $x, y \in \Lambda^S$  con probabilidad positiva. De hecho, si la especificación local es especificada y  $y$  es fija, entonces  $\pi$  está determinada por el factor multiplicativo  $\pi(y)$ . Pero esto último está a su vez determinado por normalización.

Para la demostración de la identidad de arriba, se comprueba que

$$\pi(x) = \frac{\pi(x_K | x_1, \dots, x_{K-1})}{\pi(y_K | x_1, \dots, x_{K-1})} \pi(x_1, \dots, x_{K-1}, y_K)$$

escribiendo estas probabilidades usando la definición de Bayes. Similarmente,

$$\pi(x_1, \dots, x_{K-1}, y_K) = \frac{\pi(x_{K-1} | x_1, \dots, x_{K-2}, y_K)}{\pi(y_{K-1} | x_1, \dots, x_{K-2}, y_K)} \pi(x_1, \dots, x_{K-2}, y_{K-1}, y_K)$$

etcétera. Los cálculos anteriores tienen sentido debido a que la condición de positividad y el hecho de que  $\pi(x)$  y  $\pi(y)$  son positivas implican que para toda  $j \in [1, K]$ ,

$$\pi(x_1, \dots, x_j, y_{j+1}, \dots, y_K) > 0$$

Se tiene en cuenta que el teorema 5.1.1 no es cierto en general para un número infinito de sitios, y es por eso que se recordó el supuesto básico sobre la finitud de  $S$ . En el caso de que  $S$  sea infinito, puede haber varias distribuciones correspondientes a una especificación local dada.



### 5.1.2. Camarillas, Potencial, y Distribuciones de Gibbs

La noción básica en esta sección viene de la física, donde esta fue introducida por Gibbs [14]. Se considera la distribución de probabilidad

$$\pi_T(x) = \frac{1}{Z_T} e^{-\frac{1}{T}\mathcal{E}(x)} \quad (5.5)$$

sobre el espacio de configuraciones  $\Lambda^S$ , donde  $T > 0$  es la *temperatura*,  $\mathcal{E}(x)$  es la *energía* de la configuración  $x$ , y  $Z_T$  es la constante de normalización, llamada *función de partición*. Ya que  $\pi(T)$  toma sus valores en  $[0, 1]$ , necesariamente  $-\infty < \mathcal{E}(x) \leq \infty$ . Notando que  $\mathcal{E}(x) < \infty$  si y sólo si  $\pi_T(x) > 0$ .

Estas distribuciones son interesantes para los físicos cuando la energía se expresa en términos de una función potencial que describe las interacciones locales. La noción de camarilla juega entonces un papel central. Recordando que, en el lenguaje corriente, una camarilla es un grupo de personas que se conocen (y se favorecen) unas a otras.

**Definición 5.1.6** (Camarillas). Cualquier instancia única  $\{s\}$  es una camarilla. Un subconjunto  $C \subset S$  con más de un elemento es llamado camarilla de la topología  $(S, N)$  si y sólo si cualesquiera dos sitios distintos de  $C$  son vecinos mutuos. Una camarilla  $C$  es llamada *maximal* si para cualquier sitio  $s$ ,  $C \cup \{s\}$  no es una camarilla.

**Ejemplo 5.1.1.** En física estadística, el siguiente modelo es considerado como una idealización cualitativamente correcta de una pieza de material ferromagnético. Aquí  $S = \mathbb{Z}_m^2 = \{(i, j) \in \mathbb{Z}^2, i, j \in [1, m]\}$  y  $\Lambda = \{1, -1\}$ , donde  $\pm 1$  es la orientación del giro magnético en un sitio dado.

**Definición 5.1.7.** Un *potencial de Gibbs* sobre  $\Lambda^S$  relativo al sistema de vecindades  $N$  es una colección  $\{V_C\}_{C \subset S}$  de funciones  $V_C : \Lambda^S \rightarrow \mathbb{R} \cup \{\infty\}$  tal que

- (i)  $V_C \equiv 0$  si  $C$  no es una camarilla,
- (ii) para toda  $x, x' \in \Lambda^S$  y toda  $C \subset S$ ,

$$(x(C) = x'(C)) \implies (V_C(x) = V_C(x')). \quad (5.6)$$

La función de energía  $\mathcal{E} : \Lambda^S \rightarrow \mathbb{R} \cup \{\infty\}$  se dice que *se deriva del potencial*  $\{V_C\}_{C \subset S}$  si

$$\mathcal{E}(x) = \sum_C V_C(x). \quad (5.7)$$

La función  $V_C$  depende solo de las fases en los sitios dentro del subconjunto  $C$ . Se podría escribir de manera explícita  $V_C(x(C))$  en lugar de  $V_C(x)$ , pero esta notación no será usada.

En este contexto, la distribución en 5.5 se conoce como *distribución de Gibbs*.

**Ejemplo 5.1.2** (Modelo de Ising). Este modelo fue introducido por Ising [18] para comprender cualitativamente el fenómeno de transición de fase en materiales ferromagnéticos. En el modelo de Ising finito,  $S = \mathbb{Z}_m^2$ ,  $\Lambda = \{1, -1\}$ . El potencial de Gibbs es

$$V_{\{s\}}(x) = -\frac{H}{k}x(s),$$

$$V_{\langle s,t \rangle}(x) = -\frac{J}{k}x(t)x(s),$$

donde  $\langle s,t \rangle$  es la camarilla de 2 elementos  $\{s,t\}$ , donde  $t \in \mathcal{N}_s$ . Aquí,  $k$  es la constante de Boltzmann,  $H$  es el campo magnético externo, y  $J$  es la energía interna de un dipolo magnético elemental. La función de energía correspondiente a este potencial es por lo tanto

$$\mathcal{E}(x) = -\frac{J}{k} \sum_{\langle s,t \rangle} x(s)x(t) - \frac{H}{k} \sum_{s \in S} x(s).$$

**Ejemplo 5.1.3** (Red Neuronal). En este ejemplo, la topología  $(S, N)$  es como en el modelo de Ising, pero ahora el espacio de fases es  $\Lambda = \{0, 1\}$ . Un sitio  $s$  se interpreta como una *neurona* que se *excita* si  $x(s) = 1$  e *inhibida* si  $x(s) = 0$ . Si  $t \in \mathcal{N}_s$ , se dice que existe una *sinapsis* de  $s$  a  $t$ , y tal sinapsis tiene una *fuerza*  $w_{st}$ . Si  $w_{st} > 0$ , se dice que la sinapsis es *excitatoria*; en otro caso esta es llamada *inhibitoria*. La función de energía es

$$\mathcal{E}(x) = \sum_{s \in S} \sum_{t \in \mathcal{N}_s} w_{ts}x(t)x(s) - \sum_{s \in S} h_s x(s),$$

donde  $h_s$  es llamado el *umbral* de la neurona  $s$  (esto se entenderá después). De la expresión

$$\mathcal{E}(x) = \sum_{\langle s,t \rangle} (w_{ts} + w_{st})x(t)x(s) - \sum_{s \in S} h_s x(s),$$

se observa que la energía se deriva del potencial de Gibbs

$$V_{\{s\}}(x) = -h_s x(s),$$

$$V_{\langle s,t \rangle}(x) = (w_{ts} + w_{st})x(t)x(s).$$

Desde un punto de vista biológico, el modelo anterior no es interesante porque el coeficiente de  $x(t)x(s)$  es simétrico en  $t, s$ , y esto implica una simetría entre dos neuronas que interactúan que no se encuentra en la naturaleza. Sin embargo, este modelo es de interés para la inteligencia artificial y es entonces denominada red neuronal artificial.

Uno de los desafíos asociados con los modelos de Gibbs es obtener fórmulas explícitas para promedios, considerando que generalmente es difícil calcular la función de partición. Sin embargo, los físicos han adquirido cierta experiencia en este dominio con modelos relativamente simples; ver, por ejemplo, (Baxter, 1982)[6]. Aquí está el más simple, que Ising resolvió en 1925.

**Ejemplo 5.1.4.** Tomando el modelo de Ising del ejemplo 5.1.2, excepto que el espacio de sitios  $S = \{1, 2, \dots, N\}$  consiste de  $N$  puntos arreglados en este orden sobre un círculo. En particular, los sitios 1 y  $N$  son vecinos. Los vecinos del sitio  $i$  son  $i + 1$  e  $i - 1$ , con la convención de que el sitio  $N + 1$  es el sitio 1. El espacio de fases es  $\Lambda = \{1, -1\}$ . Una configuración  $x \in E = \Lambda^S$  es denotada por  $(x_1, x_2, \dots, x_N)$ . Por lo tanto, con  $K = \frac{J}{kT}$  y  $h = \frac{H}{kT}$ ,

$$\pi_T(x) = \frac{1}{Z_N} e^{K \sum_{i=1}^N x_i x_{i+1} + h \sum_{i=1}^N x_i},$$

donde la función de partición es

$$Z_N = \sum_{x \in E} e^{K \sum_{i=1}^N x_i x_{i+1} + h \sum_{i=1}^N x_i}.$$

Esto toma la forma conveniente de

$$Z_N = \sum_{x \in E} R(x_1, x_2) R(x_2, x_3) \cdots R(x_N, x_1), \quad (5.8)$$

donde

$$R(a, b) = e^{Kab + \frac{1}{2}h(a+b)}.$$

Definiendo la *matriz de transferencia*

$$R = \begin{pmatrix} R(1, 1) & R(1, -1) \\ R(-1, 1) & R(-1, -1) \end{pmatrix} = \begin{pmatrix} e^{K+h} & e^{-K} \\ e^{-K} & e^{K-h} \end{pmatrix},$$

se observa que el término general de la suma en 5.8 es igual a

$$\sum_{x_1 \in \Lambda} \sum_{x_1, \dots, x_N \in \Lambda} R(x_1, x_2) R(x_2, x_3) \cdots R(x_N, x_1) = \sum_{x_1 \in \Lambda} R^N(x_1, x_1).$$

Por lo tanto,  $Z_N$  es la traza de  $R^N$ . En particular, si se denota por  $\lambda_+$  y  $\lambda_-$  los eigenvalores de  $R$ , entonces

$$Z_N = (\lambda_+)^N + (\lambda_-)^N. \quad (5.9)$$

El método de la matriz de transferencia se puede aplicar en principio a otros modelos de Ising traslacionalmente invariantes. Para obtener más ejemplos de modelos resueltos con exactitud, todos relacionados con el modelo básico de Ising, véase (Baxter, 1982)[6].

## 5.2. Equivalencia de Gibbs-Markov

### 5.2.1. Del potencial a la especificación local

Las distribuciones de Gibbs con una energía derivada de un potencial de Gibbs en relación con un sistema de vecindades son distribuciones de campos de Markov en relación con el mismo sistema de vecindades. Este resultado admite un recíproco (parcial), dado en la siguiente sección.

**Teorema 5.2.1.** Si  $X$  es un campo aleatorio con una distribución  $\pi$ , de la forma 5.5, donde la energía  $\mathcal{E}(x)$  se deriva de un potencial de Gibbs  $\{V_C\}_{C \in S}$  relativo al sistema de vecindades  $N$ , entonces  $X$  es de Markov relativo al mismo sistema de vecindades  $N$ . Más aún, su especificación local está dada por la formula

$$\pi^s(x) = \frac{e^{-\sum_{C \ni s} V_C(x)}}{\sum_{\lambda \in \Lambda} e^{-\sum_{C \ni s} V_C(\lambda, x(S \setminus s))}}, \quad (5.10)$$

donde la notación  $\sum_{C \ni s}$  significa que la suma se extiende sobre los conjuntos  $C$  que contienen el sitio  $s$ .

**Demostración.** Primero se observa que el lado derecho de 5.10 depende de  $x$  solo a través de  $x(s)$  y  $x(N_s)$ . De hecho,  $V_C(x)$  depende solo de  $(x(t), t \in C)$ , y para una camarilla  $C$ , si  $t \in C$  y  $s \in C$ , entonces ocurre  $t = s$  ó  $t \in N_s$ .

Por lo tanto, si se puede mostrar que  $P(X(s) = x(s) | X(S \setminus s) = x(S \setminus s))$  es igual al lado derecho de 5.10, y en particular es una función de  $x(s)$  y  $x(N_s)$  solamente, entonces por el teorema 1.2.5, se probará la propiedad de Markov 5.1 y la ecuación 5.10. Por definición de probabilidad condicional,

$$P(X(s) = x(s) | X(S \setminus s) = x(S \setminus s)) = \frac{\pi(x)}{\sum_{\lambda \in \Lambda} \pi(\lambda, x(S \setminus s))}. \quad (5.11)$$

Como

$$\pi(x) = \frac{1}{Z} e^{-\sum_{C \ni s} V_C(x) + \sum_{C \not\ni s} V_C(x)},$$

y similarmente,

$$\pi(\lambda, x(S \setminus s)) = \frac{1}{Z} e^{-\sum_{C \ni s} V_C(\lambda, x(S \setminus s)) - \sum_{C \not\ni s} V_C(\lambda, x(S \setminus s))}.$$

Si  $C$  es una camarilla y  $s$  no está en  $C$ , entonces  $V_C(\lambda, x(S \setminus s)) = V_C(x)$  y es por lo tanto independiente de  $\lambda \in \Lambda$ . Por lo tanto, el lado derecho de 5.11 se encuentra, después de factorizar  $\exp\{-\sum_{C \not\ni s} V_C(x)\}$ , para ser igual al lado derecho de 5.10.

La *energía local* en el sitio  $s$  de la configuración  $x$  es

$$\mathcal{E}_s(x) = \sum_{C \ni s} V_C(x).$$

Con esta notación, 5.10 se convierte en

$$\pi^s(x) = \frac{e^{-\mathcal{E}_s(x)}}{\sum_{\lambda \in \Lambda} e^{-\mathcal{E}_s(\lambda, x(S \setminus s))}}.$$

**Ejemplo 5.2.1** (Red Neuronal). Continuando con el ejemplo 5.1.3. Se encuentra por la aplicación de 5.10, y recordando que  $\Lambda = \{0, 1\}$ , que

$$\pi_T^s(x) = \frac{e^{-\frac{1}{T} \{\sum_{t \in N_s} (w_{ts} + w_{st})x(t) - h_s\}x(s)}}{1 + e^{-\frac{1}{T} \{\sum_{t \in N_s} (w_{ts} + w_{st})x(t) - h_s\}}}. \quad (5.12)$$

**Ejemplo 5.2.2** (Modelo de Ising). Continuando con el ejemplo 5.1.2. Las características locales en el Modelo de Ising son

$$\pi_T^s(x) = \frac{e^{\frac{1}{kT}\{J \sum_{t:|t-s|=1} x(t)+H\}x(s)}}{e^{\frac{1}{kT}\{J \sum_{t:|t-s|=1} x(t)+H\}} + e^{-\frac{1}{kT}\{J \sum_{t:|t-s|=1} x(t)+H\}}}. \quad (5.13)$$

La energía local en  $s$  es

$$\mathcal{E}_s(x) = \frac{1}{k} \left\{ J \sum_{t:|t-s|=1} x(t) + H \right\} x(s).$$

### 5.2.2. De la especificación local al potencial

El teorema 5.2.1 anterior es la parte directa del teorema de equivalencia de Gibbs-Markov: una distribución de Gibbs relativa a un sistema de vecindades es la distribución de un campo de Markov con respecto al mismo sistema de vecindades. La parte inversa, debida a Hammersley y Clifford (1968)[20], es importante desde un punto de vista teórico, ya que junto con la parte directa se concluye que las distribuciones de Gibbs y los campos aleatorios de Markov son esencialmente los mismos objetos, con una provisión explicada en el siguiente enunciado.

**Teorema 5.2.2** (Teorema de Hammersley-Clifford). Sea  $\pi$  la distribución de un campo aleatorio de Markov con respecto de la topología  $(S, N)$  que satisface la condición de positividad. Entonces

$$\pi(x) = \frac{1}{Z} e^{-\mathcal{E}(x)}$$

para alguna función de energía  $\mathcal{E}(x)$  que se deriva del potencial de Gibbs  $\{V_C\}_{C \subset S}$  asociado con la topología  $(S, N)$ .

**Demostración.** La siguiente demostración, debida a Grimmet (1973) [16], está basada en la formula de Möbius.

**Lema 5.2.1** (Formula de Möbius). Sea  $\Phi$  y  $\Psi$  dos funciones definidas sobre  $\mathbf{P}(E)$ , la colección de subconjuntos de un conjunto finito  $E$ . Las dos afirmaciones a continuación son equivalentes

$$\Phi(A) = \sum_{B \subset A} (-1)^{|A \setminus B|} \Psi(B), \quad \text{para toda } A \subset E, \quad (5.14)$$

$$\Psi(A) = \sum_{B \subset A} \Phi(B), \quad \text{para toda } A \subset E, \quad (5.15)$$

donde  $|C|$  es el número de elementos del conjunto  $C$ .

**Demostración.** Primero se mostrará que 5.14 implica 5.15. Escribiendo el lado derecho de 5.15 usando 5.14:

$$\begin{aligned} \sum_{B \subset A} \Phi(B) &= \sum_{B \subset A} \sum_{D \subset B} (-1)^{|B \setminus D|} \Psi(D) \\ &= \sum_{D \subset A} \sum_{C \subset A \setminus B} (-1)^{|C|} \Psi(D) \\ &= \sum_{D \subset A} \Psi(D) \sum_{C \subset A \setminus D} (-1)^{|C|}. \end{aligned}$$

Pero si  $A \setminus D = \emptyset$ ,

$$\sum_{C \subset A \setminus D} (-1)^{|C|} = (-1)^{|\emptyset|} = (-1)^0 = 1,$$

mientras que si  $A \setminus D \neq \emptyset$ ,

$$\begin{aligned} \sum_{C \subset A \setminus D} (-1)^{|C|} &= \sum_{k=0}^{|A \setminus D|} (-1)^k |\{C : |C| = k, C \subset A \setminus D\}| \\ &= \sum_{k=0}^{|A \setminus D|} (-1)^k \binom{|A \setminus D|}{k} = (1 - 1)^{|A \setminus D|} = 0 \end{aligned}$$

y por lo tanto

$$\sum_{D \subset A} \Psi(D) \sum_{C \subset A \setminus D} (-1)^{|C|} = \Psi(A).$$

Ahora se mostrará que 5.15 implica 5.14. Escribiendo el lado derecho de 5.14 usando 5.15:

$$\sum_{B \subset A} (-1)^{|A \setminus B|} \Psi(B) = \sum_{D \subset B \subset A} (-1)^{|A \setminus B|} \Phi(D) = \sum_{D \subset A} \Phi(D) \sum_{C \subset A \setminus D} (-1)^{|C|}$$

Por el mismo argumento de arriba, lo último equivale a  $\Phi(A)$ .

Ahora se probará el teorema 5.2.2. Sea  $0$  un elemento fijo del espacio de estados  $\Lambda$ . También, sea  $0$  la configuración con todas sus fases igual a  $0$ . El contexto va a prevenir la confusión entre  $0 \in \Lambda$  y  $0 \in \Lambda^S$ . Se puede suponer en vista de la condición de positividad que  $\pi(0) > 0$ . Sea  $x$  una configuración, y sea  $A$  un subconjunto de  $S$ . La notación  $x^A$  representa una configuración de  $\Lambda^S$  coincidiendo con  $x$  sobre  $A$ , y con la fase  $0$  fuera de  $A$ .

Se define para  $A \subset S$ ,  $x \in \Lambda^S$ ,

$$V_A(x) = \sum_{B \subset A} (-1)^{|A \setminus B|} \ln \frac{\pi(0)}{\pi(x^B)}. \quad (5.16)$$

De la formula de Möbius,

$$\ln \frac{\pi(0)}{\pi(x^A)} = \sum_{B \subset A} V_B(x).$$

Tomando  $A = S$  se obtiene  $\ln \frac{\pi(0)}{\pi(x)} = \sum_{B \subset S} V_B(x)$ ; esto es,

$$\pi(x) = \pi(0) e^{-\sum_{A \subset S} V_A(x)}.$$

Queda mostrar que  $V_A$  depende solo de las fases sobre  $A$ , y que  $V_A \equiv 0$  si  $A$  no es una camarilla de  $(S, N)$ .

Si  $x, y \in \Lambda^S$  son tal que  $x(A) = y(A)$ , entonces para cualquier  $B \subset A$ ,  $x^B = y^B$ , y por lo tanto, por 5.16,  $V_A(x) = V_A(y)$ .

Usando un sitio arbitrario  $t \in A$ , se escribe 5.16 como sigue:

$$\begin{aligned} V_A(x) &= \sum_{B \subset A, B \not\ni t} (-1)^{|A \setminus B|} \ln \frac{\pi(0)}{\pi(x^B)} \sum_{B \subset A, B \ni t} (-1)^{|A \setminus B|} \ln \frac{\pi(0)}{\pi(x^B)} \\ &= \sum_{B \subset A \setminus \{t\}} (-1)^{|A \setminus B|} \left\{ \ln \frac{\pi(0)}{\pi(x^B)} - \ln \frac{\pi(0)}{\pi(x^{B \cup \{t\}})} \right\}. \end{aligned}$$

Esto es,

$$V_A(x) = \sum_{B \subset A \setminus \{t\}} (-1)^{|A \setminus B|} \ln \frac{\pi(x^{B \cup \{t\}})}{\pi(x^B)}. \quad (5.17)$$

Ahora, si  $t$  no está en  $B \subset A$ , entonces

$$\frac{\pi(x^{B \cup \{t\}})}{\pi(x^B)} = \frac{\pi^t(x^{B \cup \{t\}})}{\pi^t(x^B)},$$

y por lo tanto

$$V_A(x) = \sum_{B \subset A \setminus \{t\}} (-1)^{|A \setminus B|} = \frac{\pi^t(x^{B \cup \{t\}})}{\pi^t(x^B)},$$

y, por los mismos cálculos que llevaron a 5.17,

$$V_A(x) = - \sum_{B \subset A} (-1)^{|A \setminus B|} \ln \pi^t(x^B). \quad (5.18)$$

Recordando que  $t \in A$ , y por lo tanto, si  $A$  no es una camarilla, se puede encontrar  $s \in A$  tal que  $s$  no es un vecino de  $t$ . Fijando tal  $s$ , y dividiendo la suma en 5.18 como sigue:

$$\begin{aligned} V_A(x) &= - \sum_{B \subset A \setminus \{s,t\}} (-1)^{|A \setminus B|} \ln \pi^t(x^B) - \sum_{B \subset A \setminus \{s,t\}} (-1)^{|A \setminus (B \cup \{t\})|} \ln \pi^t(x^{B \cup \{t\}}) \\ &\quad - \sum_{B \subset A \setminus \{s,t\}} (-1)^{|A \setminus (B \cup \{s\})|} \ln \pi^t(x^{B \cup \{s\}}) - \sum_{B \subset A \setminus \{s,t\}} (-1)^{|A \setminus (B \cup \{s,t\})|} \ln \pi^t(x^{B \cup \{s,t\}}) \\ &= - \sum_{B \subset A \setminus \{s,t\}} (-1)^{|A \setminus B|} \ln \frac{\pi^t(x^B) \pi^t(x^{B \cup \{s,t\}})}{\pi^t(x^{B \cup \{s\}}) \pi^t(x^{B \cup \{t\}})}. \end{aligned}$$

Pero ya que  $s \neq t$  y  $s \notin \mathcal{N}_t$ , se tiene que  $\pi^t(x^B) = \pi^t(x^{B \cup \{s\}})$  y  $\pi^t(x^{B \cup \{t\}}) = \pi^t(x^{B \cup \{s,t\}})$ , y por lo tanto  $V_A(x) = 0$ .

Debe notarse que la función de energía  $\mathcal{E}$  y la función de partición no son únicas, ya que se puede agregar una constante a la función de energía y luego multiplicar el factor de normalización por una constante apropiada. Asimismo, y lo que es más importante, el potencial de Gibbs asociado con  $\pi$  no es único. Sin embargo, la unicidad se puede forzar en el resultado si se impone una determinada propiedad al potencial, a saber, la normalización con respecto a un valor de fase fijo.

En muchas situaciones, tanto el potencial como la topología de  $S$  se pueden obtener directamente de la expresión de la energía.

**Definición 5.2.1.** Un potencial de Gibbs  $\{V_C\}_{C \subset S}$  se dice que se normaliza respecto de una fase dada en  $\Lambda$ , convencionalmente denotada por 0, si  $V_C(x) = 0$  siempre que exista  $t \in C$  tal que  $x(t) = 0$ .

**Teorema 5.2.3.** Existe uno y solo un potencial normalizado respecto de una fase dada  $0 \in \Lambda$  correspondiente a la distribución de Gibbs.

**Demostración.** La formula 5.17 produce por 5.16 la expresión

$$V_C(x) = \sum_{B \subset C \setminus \{t\}} (-1)^{|A \setminus B|} \ln \frac{\pi^{B \cup \{t\}}}{\pi(x^B)}$$

independiente de  $t$  en la camarilla  $C$ . En particular, eligiendo cualquier  $t \in C$  tal que  $x(t) = 0$ ,  $x^{B \cup \{t\}} = x^B$  para toda  $B \subset C \setminus \{t\}$ , y por lo tanto  $V_C(x) = 0$ . Esto muestra que 5.16 da como resultado un potencial normalizado respecto de  $0 \in \Lambda$ . Para probar la unicidad, se toma como supuesto que

$$\pi(x) = \frac{1}{Z_U} e^{-\mathcal{E}_U(x)} = \frac{1}{Z_W} e^{-\mathcal{E}_W(x)}$$

para dos funciones de energía  $\mathcal{E}_U$  y  $\mathcal{E}_W$  derivadas de los potenciales  $U$  y  $W$ , respectivamente, ambas normalizadas respecto de  $0 \in \Lambda$ . Ya que  $\mathcal{E}_U(0) = \sum_{C \in \mathcal{C}} U_C(0) = 0$ , y similarmente  $\mathcal{E}_W(0) = 0$ , se sigue que  $Z_U = Z_W = \pi(0)^{-1}$ , y por lo tanto  $\mathcal{E}_U \equiv \mathcal{E}_W$ . Si  $U_A = W_A$  para toda  $A \in \mathcal{C}$  tal que  $|A| \leq k$  (propiedad  $\mathcal{P}_k$ ). Queda demostrar por inducción que  $\mathcal{P}_k \implies \mathcal{P}_{k+1}$  y que  $\mathcal{P}_1$  se cumple.

Para probar que  $\mathcal{P}_k \implies \mathcal{P}_{k+1}$ , se fija  $A \subset S$  con  $|A| = k + 1$ . Para probar que  $U_A \equiv W_A$  es suficiente mostrar que  $U_A(x) = W_A(x)$  para toda  $x \in \Lambda^S$  tal que  $x = x^A$ . Fijando esta  $x$ . Entonces

$$\mathcal{E}_U(x) = \sum_C U_C(x) = \sum_{C \subset A} U_C(x),$$

ya que  $x$  tiene fase 0 fuera de  $A$  y  $U$  está normalizada con respecto de 0. También,

$$\mathcal{E}_U(x) = \sum_{C \subset A} U_C(x) = U_A(x) + \sum_{C \subset A, |C| \leq k} U_C(x), \quad (5.19)$$

con una ecuación similar para  $\mathcal{E}_W(x)$ . Por lo tanto como  $\mathcal{E}_U(x) = \mathcal{E}_W(x)$ , se obtiene  $U_A(x) = W_A(x)$  en vista de la hipótesis de inducción. La raíz  $\mathcal{P}_1$  de la hipótesis de inducción es cierta, ya que cuando  $|A| = 1$ , 5.19 se convierte en  $\mathcal{E}_U(x) = W_A(x)$ , así que  $U_A(x) = W_A(x)$  es consecuencia de  $\mathcal{E}_U(x) = \mathcal{E}_W(x)$ .

**Ejemplo 5.2.3.** Sea  $S = \{0, 1, \dots, N\}$  y  $\Lambda = E$ , un espacio finito. Un campo aleatorio  $X$  sobre  $S$  con espacio de fases  $\Lambda$  es por lo tanto un vector  $X$  con valores en  $E^{N+1}$ . Suponiendo que  $X_0, \dots, X_N$  es una cadena de Markov homogénea con matriz de transición  $\mathbf{P} = \{p_{ij}\}_{i,j \in E}$  y distribución inicial  $\nu = \{\nu_i\}_{i \in E}$ . En particular, con  $x = (x_0, \dots, x_N)$ ,

$$\pi(x) = \nu_{x_0} p_{x_0 x_1} \cdots p_{x_{N-1} x_N},$$



esto es,

$$\pi(x) = e^{-\mathcal{E}(x)},$$

donde

$$\mathcal{E}(x) = -\ln v_{x_0} - \sum_{n=0}^{N-1} (\ln p_{x_n x_{n+1}}).$$

Esta energía se deriva del potencial de Gibbs asociado con la topología del vecino más cercano para la cual las camarillas son, además de las instancias únicas, las parejas de sitios adyacentes. Las funciones potencial no nulas pueden escogerse de la siguiente manera:

$$V_{\{0\}}(x) = -\ln v_{x_0}, \quad V_{\{n, n+1\}}(x) = -\ln p_{x_n x_{n+1}}.$$

La característica local en el sitio  $s = n \in [2, N-1]$  está dada por la formula 5.10,

$$\pi^n(x) = \frac{\exp(\ln p_{x_{n-1}x_n} + \ln p_{x_n x_{n+1}})}{\sum_{y \in E} \exp(\ln p_{x_{n-1}y} + \ln p_{yx_{n+1}})};$$

esto es,

$$\pi^n = \frac{P_{x_{n-1}x_n} P_{x_n x_{n+1}}}{P_{x_{n-1}x_{n+1}}^{(2)}},$$

donde  $p_{ij}^{(2)}$  es el término general de la matriz en dos pasos  $\mathbf{P}^2$ . Cálculos similares dan  $\pi^0(x)$  y  $\pi^N(x)$ .

En vista de la estructura de vecindades, para  $n \in [2, N-1]$ ,  $X_n$  es independiente de  $X_0, \dots, X_{n-2}, X_{n+2}, \dots, X_N$  dado  $X_{n-1}$  y  $X_{n+1}$ .

**Ejemplo 5.2.4.** El ejemplo anterior puede ser particularizado a una cadena de Markov con espacio de estados  $E = \{-1, 1\}$  y matriz de transición

$$\mathbf{P} = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix}$$

donde  $\alpha, \beta \in (0, 1)$ , y con la distribución estacionaria inicial

$$(v_0, v_1) = \frac{1}{\alpha + \beta}(\beta, \alpha).$$

Observando que si  $x \in \{-1, 1\}$ , entonces  $\frac{1-x}{2} = \mathbb{1}_{\{x=-1\}}$  y  $\frac{1+x}{2} = \mathbb{1}_{\{x=1\}}$ , se tiene

$$\begin{aligned} 4 \ln p_{xy} &= (1-x)(1+y) \ln(\alpha) + (1-x)(1-y) \ln(1-\alpha) \\ &\quad + (1+x)(1-y) \ln(\beta) + (1+x)(1+y) \ln(1-\beta) \end{aligned}$$

y

$$2 \ln v(x) = (1-x) \ln \left( \frac{\beta}{\alpha + \beta} \right) + (1+x) \ln \left( \frac{\alpha}{\alpha + \beta} \right).$$

Por lo tanto, particularizando la expresión para la energía en el ejemplo anterior, se obtiene

$$\begin{aligned}
-\mathcal{E}(x) &= \frac{1}{4}N \ln(\beta\alpha(1-\beta)(1-\alpha)) + \frac{1}{2} \ln\left(\frac{\beta\alpha}{(\alpha+\beta)^2}\right) \\
&+ \frac{1}{2} \ln\left(\frac{1-\beta}{1-\alpha}\right) \sum_{i=0}^N x_i + \frac{1}{4}(x_0+x_N) \ln\left(\frac{\alpha(1-\beta)}{\beta(1-\alpha)}\right) \\
&\frac{1}{4} \left( \sum_{n=0}^{N-1} x_n x_{n+1} \right) \ln\left(\frac{(1-\alpha)(1-\beta)}{\alpha\beta}\right).
\end{aligned}$$

Aquí el potencial normalizado respecto de 0 aparece naturalmente en los términos no constantes.

**Ejemplo 5.2.5.** El ejemplo 5.2.3 muestra que las cadenas de Markov pueden pensarse de dos maneras distintas, siendo la elección entre una descripción unilateral en términos de la matriz de transición y una descripción bilateral en términos de las características locales. La noción de campo aleatorio de Markov unilateral generaliza esta situación.

Un campo aleatorio de Markov unilateral recibe una descripción análoga a la de una cadena de Markov, excepto que la sucesión de tiempos  $1, 2, 3, \dots$  es ahora una sucesión de sitios  $s_1, s_2, \dots, s_N$  que ocupa todos los sitios de  $S$  sin repetición. El tiempo está indirectamente involucrado en que al menos para algunas aplicaciones en el procesamiento de imágenes, los sitios de  $S$  se escanean secuencialmente en el orden anterior. Por esta razón, dicha enumeración se denomina *escaneo* de  $S$ .

De 1.2.3,

$$\pi(x) = P(X(s_1) = x(s_1)) \prod_{j=2}^N P(X(s_j) = x(s_j) | X(s_1) = x(s_1), \dots, X(s_{j-1}) = x(s_{j-1})).$$

Sea  $X$  un campo aleatorio sobre  $S$  con espacio de fases  $\Lambda$ , y sea  $s_1, \dots, s_N$  un escaneo de  $S$ . Sea  $\mathcal{P}_1 = \mathcal{P}(s_1) = \phi, \mathcal{P}_2 = \mathcal{P}(s_2), \dots, \mathcal{P}_N = \mathcal{P}(s_N)$  una sucesión de subconjuntos de  $S$  tal que para toda  $j \geq 1$ ,

$$\mathcal{P}(s_j) \subset \{s_1, \dots, s_{j-1}\} = S_{j-1}.$$

Suponiendo que para toda  $j \geq 1$ ,

$$P(X(s_j) = x(s_j) | X(S_{j-1}) = x(S_{j-1})) = P(X(s_j) = x(s_j) | X(\mathcal{P}(s_j)) = x(\mathcal{P}(s_j))). \tag{5.20}$$

Entonces  $\{\mathcal{P}(s)\}_{s \in S}$  es llamada *estructura de relevancia pasada* respecto del escaneo  $s_1, \dots, s_N$ ;  $\mathcal{P}(s)$  es llamado *pasado relevante* en el sitio  $s$ ; y  $X$  es llamada *campo aleatorio de Markov unilateral* respecto de  $\{\mathcal{P}(s)\}_{s \in S}$  y el escaneo  $s_1, \dots, s_N$ .

La *función de transición* de  $X$  en el sitio  $s \in S$  es la siguiente función de  $x(s)$  y  $x(\mathcal{P}(s))$ :

$$p(x(s) | x(\mathcal{P}(s))) = P(X(s) = x(s) | X(\mathcal{P}(s)) = x(\mathcal{P}(s))), \tag{5.21}$$

y por lo tanto, para cualquier configuración  $x$ ,

$$\pi(x) = \prod_{j=1}^N p(x(s_j)|x(\mathcal{P}(s_j))) \quad (5.22)$$

Se considera el siguiente sistema de vecindades

$$\mathcal{N}_s = \left( \bigcup_{t:\mathcal{P}(t)\ni s} \{\mathcal{P}(t) \cup \{t\}\} \setminus \{s\} \right) \cup \mathcal{P}(s) \quad (5.23)$$

se mostrará que  $P(X(s) = x(s)|X(S \setminus s) = x(S \setminus s))$  depende solo sobre  $x(\mathcal{N}_s \cup \{s\})$ . La regla de Bayes da como resultado

$$\pi^s(x) = \frac{\pi(x(s), x(S \setminus s))}{\sum_{\lambda \in \Lambda} \pi(\lambda, x(S \setminus s))}. \quad (5.24)$$

de 5.22 se tiene que  $\pi(\lambda, x(S \setminus s)) = A \times B$ , donde  $A$  es independiente de  $\lambda$  y por lo tanto se factoriza en la última fracción, y

$$B = p(\lambda|x(\mathcal{P}(s))) \times \prod_{t:s \in \mathcal{P}(t)} p(x(t)|\lambda, x(\mathcal{P}(t) \setminus \{s\})).$$

Por lo tanto,

$$\pi^s(x) = \frac{p(x(s)|x(\mathcal{P}(s))) \times \prod_{t:s \in \mathcal{P}(t)} p(x(t)|x(\mathcal{P}(t)))}{p(\lambda|x(\mathcal{P}(s))) \times \prod_{t:s \in \mathcal{P}(t)} p(x(t)|\lambda, x(\mathcal{P}(t) \setminus \{s\}))}, \quad (5.25)$$

una cantidad que depende solo sobre  $x(\mathcal{N}_s \cup \{s\})$ .

## 5.3. Muestreo de Gibbs

### 5.3.1. Simulación de campos aleatorios

Se considera un campo aleatorio que cambia aleatoriamente con el tiempo. Entonces se tiene un proceso estocástico  $\{X_n\}_{n \geq 0}$ , donde

$$X_n = (X_n(s), s \in S)$$

y  $X_n(s) \in \Lambda$ . El estado en el tiempo  $n$  de este proceso es un campo aleatorio en  $S$  con fases en  $\Lambda$ , o equivalentemente, una variable aleatoria con valores en el espacio de estados  $E = \Lambda^S$ , que por simplicidad se asume finito. El proceso estocástico  $\{X_n\}_{n \geq 0}$  se denominará *campo aleatorio dinámico*.

El propósito de la sección actual es mostrar cómo un campo aleatorio dado con distribución de probabilidad

$$\pi(x) = \frac{1}{Z} e^{-\mathcal{E}(x)} \quad (5.26)$$

puede surgir como la distribución estacionaria de una cadena de Markov con valores en un campo. Este problema es interesante en el siguiente contexto de simulación.

Suponiendo que existe una cadena de Markov aperiódica irreducible  $\{X_n\}_{n \geq 0}$  con espacio de estados  $E = \Lambda^S$  y distribución estacionaria 5.26. Por el resultado fundamental de la convergencia al estado estacionario (Teorema 4.2.1), para cualquier distribución inicial,

$$\lim_{n \uparrow \infty} d_V(P(X_n = \cdot), \pi) = 0. \quad (5.27)$$

Si se es capaz de generar una realización de la CMH anterior, su distribución en un tiempo grande  $n$  será cercana a  $\pi$ , y por lo tanto se habrá simulado  $\pi$ . En este contexto, el mecanismo de transición de la cadena se denominará algoritmo de simulación, o algoritmo de muestreo, y la distribución asintótica  $\pi$  la distribución objetivo, o distribución simulada, o distribución muestreada.

El primer problema es el de identificar una cadena  $\{X_n\}_{n \geq 0}$  con las propiedades de arriba. El *muestreo de Gibbs* usa una distribución estrictamente positiva  $(q_s, s \in S)$  sobre  $S$ , y la transición de  $X_n = x$  a  $X_{n+1} = y$  está hecha de acuerdo a la siguiente regla.

El nuevo estado  $y$  es obtenido del viejo estado  $x$  cambiando (o no) el valor de la fase en *un solo sitio*. El sitio  $s$  que se va a cambiar (o no) en el tiempo  $n$  es elegido independientemente del pasado con probabilidad  $q_s$ . Cuando el sitio  $s$  ha sido elegido, la configuración actual  $x$  se cambia a  $y$  como sigue:  $y(S \setminus s) = x(S \setminus s)$ , y la nueva fase  $y(s)$  en el sitio  $s$  es seleccionada con probabilidad  $\pi(y(s)|x(S \setminus s))$ . Así, la configuración  $x$  se cambia a  $y = (y(s), x(S \setminus s))$  con probabilidad  $\pi(y(s)|x(S \setminus s))$ , de acuerdo a la especificación local en el sitio  $s$ . Esto da para las entradas no nulas de la matriz de transición

$$P(X_{n+1} = y | X_n = x) = q_s \pi(y(s)|x(S \setminus s)) \mathbb{1}_{\{y(S \setminus s) = x(S \setminus s)\}}. \quad (5.28)$$

La cadena correspondiente es irreducible y aperiódica. Para probar que  $\pi$  es la distribución estacionaria, se usa el teorema 2.6.1. Se tiene que revisar que para toda  $x, y \in \Lambda^S$ ,

$$\pi(x)P(X_{n+1} = y | X_n = x) = \pi(y)P(X_{n+1} = x | X_n = y),$$

esto es, en vista de 5.28, para toda  $s \in S$ ,

$$\pi(x)q_s \pi(y(s)|x(S \setminus s)) = \pi(y)q_s \pi(x(s)|y(S \setminus s)).$$

Pero la última ecuación es justo

$$\pi(x)q_s \frac{\pi(y(s)|x(S \setminus s))}{P(X(S \setminus s) = x(S \setminus s))} = \pi(y)q_s \frac{\pi(x(s))}{P(X(S \setminus s) = x(S \setminus s))}.$$

**Ejemplo 5.3.1** (Simulación del Modelo de Ising). La especificación local en el sitio  $s$  depende solo de la configuración local  $x(\mathcal{N}_s)$ . Teniendo en cuenta que las vecindades pequeñas aceleran los cálculos. Teniendo en cuenta también que el muestreo de Gibbs es un muestreo natural, en el sentido de que en una pieza de material ferromagnético, por ejemplo, los espines se cambian aleatoriamente de acuerdo con la especificación local. Cuando la naturaleza decide actualizar la orientación de un dipolo, lo hace de acuerdo con la ley de la mecánica estadística. Calcula la energía local para cada uno de los dos posibles giros,  $E_+ = E(1, x(\mathcal{N}_s))$  y  $E_- = E(-1, x(\mathcal{N}_s))$ , y toma la orientación correspondiente con una probabilidad proporcional a  $e^{E_+}$  y  $e^{E_-}$ , respectivamente.

**Ejemplo 5.3.2** (Muestreo de Gibbs para redes neuronales). El marco y la notación son los de los ejemplos 5.1.3 y 5.2.1, donde se obtuvieron las características locales

$$\pi_T(x(s)|x(\mathcal{N}_s)) = \frac{e^{-\frac{1}{T}\{\sum_{t \in \mathcal{N}_s} (w_{ts} + w_{st})x(t) - h_s\}x(s)}}{1 + e^{-\frac{1}{T}\{\sum_{t \in \mathcal{N}_s} (w_{ts} + w_{st})x(t) - h_s\}}}. \quad (5.29)$$

Por lo tanto, en un tiempo dado  $n$ , el muestreo de Gibbs examina el sitio  $s$  (con probabilidad  $q_s$ ) y actualiza su fase. Este elige la fase 0 con probabilidad

$$p_0 = \frac{1}{1 + e^{-\frac{a}{T}}},$$

donde

$$a = \sum_{t \in \mathcal{N}_s} (w_{ts} + w_{st})x(t) - h_s, \quad (5.30)$$

y este elige la fase 1 con probabilidad  $p_1 = 1 - p_0$ .

Este procedimiento también puede ser descrito en términos de un *jitter de umbral* aleatorio  $\sigma$  con la función de distribución

$$P(\sigma \leq a) = \frac{e^{-\frac{a}{T}}}{1 + e^{-\frac{a}{T}}}. \quad (5.31)$$

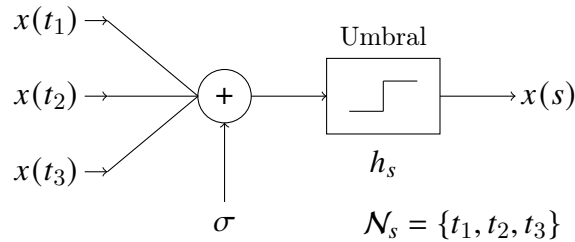
El muestreo de Gibbs selecciona la fase 0 si

$$\sum_{t \in \mathcal{N}_s} (w_{ts} + w_{st})x(t) < h_s + \sigma, \quad (5.32)$$

y 1 en otro caso. De hecho,

$$P(\text{fase 0}) = P\left(\sigma > \sum_{t \in \mathcal{N}_s} (w_{ts} + w_{st})x(t) - h_s\right) = \frac{1}{1 + e^{-\frac{1}{T}\{\sum_{t \in \mathcal{N}_s} (w_{ts} + w_{st})x(t) - h_s\}}}.$$

Se puede interpretar  $h_s$  como *umbral nominal* en el sitio  $s$  y  $h_s + \sigma$  como el umbral real (aleatorio). También la cantidad  $\sum_{t \in \mathcal{N}_s} (w_{ts} + w_{st})x(t)$  es la entrada a la neurona  $s$ . Así la excitación de la neurona  $s$  se obtiene comparando su entrada con un umbral aleatorio.



**Ejemplo 5.3.3** (Muestreo de Gibbs para vectores aleatorios). El muestreo de Gibbs se aplica a cualquier distribución de probabilidad multivariante

$$\pi(x(1), \dots, x(N))$$

sobre un conjunto  $E = \Lambda^N$ , donde  $\Lambda$  es numerable (pero esta restricción no es esencial, como se mencionó anteriormente). Esta observación tiene la intención de recordar que hay muchas aplicaciones del muestreo de Gibbs, y más generalmente de la simulación de la cadena de Monte Carlo Markov, fuera de la física o el procesamiento de imágenes, y especialmente en estadística.

El paso básico del muestreo de Gibbs para la distribución multivariada  $\pi$  consiste en seleccionar un número coordenada  $i \in [1, N]$ , al azar, y eligiendo el nuevo valor  $y(i)$  de la coordenada correspondiente, dados los valores presentes  $x(1), \dots, x(i-1), x(i+1), \dots, x(N)$  de las otras coordenadas, con probabilidad

$$\pi(y(i)|x(1), \dots, x(i-1), x(i+1), \dots, x(N)).$$

Como se indicó anteriormente, se verifica que  $\pi$  es la distribución estacionaria de la cadena correspondiente.

**Ejemplo 5.3.4.** En la práctica, los sitios actualizados no se eligen al azar, sino en un orden bien determinado  $s(1), s(2), \dots, s(N)$ , donde  $\{s(i)\}_{1 \leq i \leq N}$  es una enumeración de todos los sitios de  $S$ , denominada *política de escaneo*. Los sitios se visitan periódicamente en este orden. El estado del campo aleatorio después del  $n$ -ésimo barrido es  $Z_n = X_{nN}$ , donde  $X_k$  denota la imagen antes del  $k$ -ésimo tiempo de actualización. En el momento  $k$ , el sitio  $s(k \bmod N)$  se actualiza para producir la nueva imagen  $X_{k+1}$ . Si  $X_k = x$  y  $s(k \bmod N)$ , entonces  $X_{k+1} = (y(s), x(S \setminus s))$  con probabilidad  $\pi(y(s)|x(S \setminus s))$ . La distribución de Gibbs  $\pi$  es estacionaria para  $\{X_k\}_{k \geq 0}$ , en el sentido de que si  $P(X_k = \cdot) = \pi$ , entonces  $P(X_{k+1} = \cdot) = \pi$ . En particular,  $\pi$  es una distribución estacionaria de la cadena de Markov aperiódica irreducible  $\{Z_n\}_{n \geq 0}$ , y  $\lim_{n \uparrow \infty} P(Z_n = \cdot) = \pi$ .

La matriz de transición  $\mathbf{P}$  de  $\{Z_n\}_{n \geq 0}$  es

$$\mathbf{P} = \prod_{k=1}^N \mathbf{P}_{s(k)}, \quad (5.33)$$

donde  $\mathbf{P}_s = \{p_{xy}^s\}_{x, y \in \Lambda^s}$ , y la entrada  $p_{xy}^s$  de  $\mathbf{P}_s$  es no nula si y solo si  $y(S \setminus s) = x(S \setminus s)$ , y entonces

$$p_{xy}^s = \frac{e^{-\mathcal{E}(y(s), x(S \setminus s))}}{\sum_{\lambda \in \Lambda} e^{-\mathcal{E}(\lambda, x(S \setminus s))}}. \quad (5.34)$$

### 5.3.2. Tasa de convergencia del muestreo de Gibbs

**Ejemplo 5.3.5.** La expresión 5.34 se utilizará para producir una tasa geométrica de convergencia del muestreo de Gibbs periódico,

$$|\mu \mathbf{P}^n - \pi| \leq \frac{1}{2} |\mu - \pi| (1 - e^{-N\Delta})^n \quad (5.35)$$

donde

$$\Delta = \sup_{s \in S} \delta_s \quad (5.36)$$

y

$$\delta_s = \sup\{|\mathcal{E}(x) - \mathcal{E}(y)| : x(S \setminus s) = y(S \setminus s)\}. \quad (5.37)$$

La prueba de 5.35 usa la desigualdad de Dobrushin [9] la cual da como resultado

$$|\mu \mathbf{P}^n - \pi| \leq \frac{1}{2} |\mu - \pi| \delta(\mathbf{P})^n.$$

Usando el coeficiente ergódico [9] se sigue que para cualquier matriz de transición  $\mathbf{P}$  sobre un espacio finito  $E$ ,

$$\delta(\mathbf{P}) = 1 - \inf_{i,j \in E} \sum_{k \in E} p_{ik} \wedge p_{jk} \leq 1 - |E| \left( \inf_{i,j \in E} p_{ij} \right). \quad (5.38)$$

Si se define  $m_s(x) = \inf\{\mathcal{E}(y) : y(S \setminus s) = x(S \setminus s)\}$ , se sigue de 5.34 que

$$p_{xy}^s = \frac{\exp\{-(\mathcal{E}(y(s), x(S \setminus s)) - m_s(x))\}}{\sum_{z(s) \in \Lambda} \exp\{-(\mathcal{E}(z(s), x(S \setminus s)) - m_s(x))\}} \geq \frac{e^{\delta_s}}{|\Lambda|},$$

y por lo tanto, de 5.33,

$$\min_{x,y \in \Lambda^S} p_{xy} \geq \prod_{k=1}^N \frac{e^{\delta_s(k)}}{|\Lambda|} \geq \frac{e^{-N\Delta}}{|\Lambda|^N}.$$

Usando 5.38,

$$\delta(\mathbf{P}) \leq 1 - |\Lambda|^N \frac{e^{-N\Delta}}{|\Lambda|^N} = 1 - e^{-N\Delta},$$

y 5.35 se sigue.

La convergencia en la variación del muestreo de Gibbs a la distribución objetivo tiene lugar con velocidad geométrica. Ésta es una característica común de todos los algoritmos de simulación de cadena de Monte Carlo Markov y, de manera más general, de todas las CMH ergódicas finitas. Sin embargo, el muestreo de Gibbs tiene una característica especial. Resulta que la distancia en variación a la distribución objetivo disminuye a 0, como muestra el siguiente resultado.

**Teorema 5.3.1.** Sea  $\mu$  una medida de probabilidad arbitraria sobre  $\Lambda^S$  y sea  $\nu$  la medida de probabilidad obtenida aplicando el muestreo de Gibbs a un sitio arbitrario  $s \in S$ . Entonces  $d_V(\nu, \pi) \leq d_V(\mu, \pi)$ .

**Demostración.** Se toma  $s = 1$  sin pérdida de generalidad. Se tiene que

$$\begin{aligned} \sum_x |\nu(x) - \pi(x)| &= \sum_x |\pi(x(1)|x(S \setminus 1))\mu(x(S \setminus 1)) - \pi(x)| \\ &= \sum_x |\pi(x(1)|x(S \setminus 1))[\mu(x(S \setminus 1)) - \pi(x(S \setminus 1))]| \\ &= \sum_{x(S \setminus 1)} \left\{ \sum_{x(1)} \pi(x(1)|x(S \setminus 1)) \right\} |\mu(x(S \setminus 1)) - \pi(x(S \setminus 1))| \\ &= \sum_{x(S \setminus 1)} |\mu(x(S \setminus 1)) - \pi(x(S \setminus 1))| \\ &= \sum_{x(S \setminus 1)} \left| \sum_{x(1)} (\mu(x) - \pi(x)) \right| \\ &\leq \sum_x |\mu(x) - \pi(x)|. \end{aligned}$$

Es bastante improbable que un buen algoritmo de simulación elija una estrategia miope en la que la distancia en la variación se reduce en cada paso.

## 5.4. Simulación Monte Carlo

### 5.4.1. Principio General

Se adoptará un enfoque más general de la simulación.

En vista de evaluar la esperanza  $E[\phi(Z)]$  de un vector aleatorio  $Z$  de dimensión  $k$  con una función de densidad  $f(x)$ , donde  $\phi : \mathbb{R}^k \rightarrow \mathbb{R}$  y  $E[|\phi(Z)|] < \infty$ , la formula

$$E[\phi(Z)] = \int_{\mathbb{R}^n} \phi(x) f(x) dx$$

puede ser usada si se es capaz de calcular la integral de manera analítica, que rara vez es el caso. Una primera alternativa es la integración numérica. Una segunda alternativa es generar una sucesión de v.a.i.i.d.  $\{Z_n\}_{n \geq 1}$  con la misma distribución que  $Z$ , e invocar la ley fuerte de los grandes números

$$E[\phi(Z)] = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \phi(Z_i)$$

para obtener una estimación basada en una muestra finita  $(Z_1, \dots, Z_n)$ . La elección de esta muestra de tamaño  $n$  puede ser hecha a través de intervalos de confianza para poder alcanzar una precisión dada; ver, por ejemplo, (Ross, 1987)[33], (Ripley, 1987)[32], (Fishman, 1996)[13].

Este método, conocido como *Monte Carlo*, debe compararse con la integración numérica. Es un hecho que la simulación Monte Carlo es competitiva con la integración numérica para una dimensión grande  $k$ . En algunas aplicaciones, el método de Monte Carlo es inevitable. Uno de ellos ocurre cuando  $Z_n$  es un funcional complicado del  $n$ -ésimo ciclo de un proceso regenerativo.

Otra situación es aquella en la que la función de densidad  $f(x)$  se conoce solo hasta un factor de normalización, es decir,  $f(x) = K\tilde{f}(x)$ , y cuando, por supuesto, la integral de  $\tilde{f}(x)$  que da el factor de normalización es difícil o imposible de calcular. En física, este es un caso frecuente: la función de partición de una distribución de Gibbs suele ser incalculable de manera cerrada, debido al tamaño del espacio de estados en los modelos correspondientes.

Una tercera situación es aquella en la que los métodos estándar de generación de variables aleatorias que se presentan a continuación no son aplicables, debido a errores de redondeo o problemas de codificación.

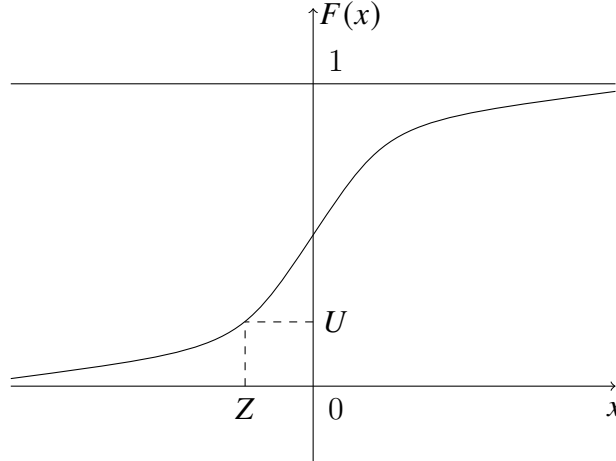
### 5.4.2. Método de la inversa y aceptación-rechazo

Una vez que se decide recurrir a la simulación de Monte Carlo, el problema es generar réplicas de  $Z$ . El método de *distribución inversa* utiliza una secuencia de v.a.i.i.d.  $\{U_n\}_{n \geq 1}$  distribuidas uniformemente en  $[0, 1]$ , y generar  $\{Z_n\}_{n \geq 1}$  con

$$Z_n = F^{-1}(U_n),$$



donde  $F^{-1}(u)$  es la inversa de la función de  $u = F(x) = \int_{-\infty}^x f(y)dy$  (se considera el caso unidimensional, pero no es esencial en el tema). El alcance de este método está limitado por el hecho de que  $F^{-1}(u)$  a menudo es difícil de obtener y, en cualquier caso, el método de la función inversa requiere un conocimiento completo de la función de densidad de probabilidad  $f(x)$ .



**Ejemplo 5.4.1** (Variable Aleatoria Exponencial). La función inversa de  $F(x) = 1 - e^{-\theta x}$  es  $F^{-1}(y) = -\frac{1}{\theta} \ln(1 - y)$ . Por lo tanto, si  $U$  se distribuye uniformemente sobre  $[0, 1]$ , entonces  $-\frac{1}{\theta} \ln(1 - U)$  es una variable aleatoria exponencial con media  $\theta^{-1}$ . Esta es

$$Z = -\frac{1}{\theta} \ln(1 - U),$$

ya que  $U$  y  $1 - U$  se distribuyen de manera idéntica.

Otro método, llamado *aceptación-rechazo* (A-R para abreviar), se acerca a una solución eficiente. De hecho, es el antepasado del método de la cadena de Monte Carlo Markov (MCMC por sus siglas en inglés).

Suponiendo que se puede generar una sucesión de v.a.i.i.d.  $\{Y_n\}_{n \geq 1}$  con función de densidad  $g(x)$  satisfaciendo, para toda  $x$ ,

$$\frac{f(x)}{g(x)} \leq c$$

para una constante finita  $c$ . Sea  $\{U_n\}_{n \geq 1}$  una sucesión de v.a.i.i.d. que se distribuyen uniformemente sobre  $[0, 1]$ . Si se define  $\tau$  como el primer índice  $n \geq 1$  para el cual

$$U_n \leq \frac{f(Y_n)}{cg(Y_n)}$$

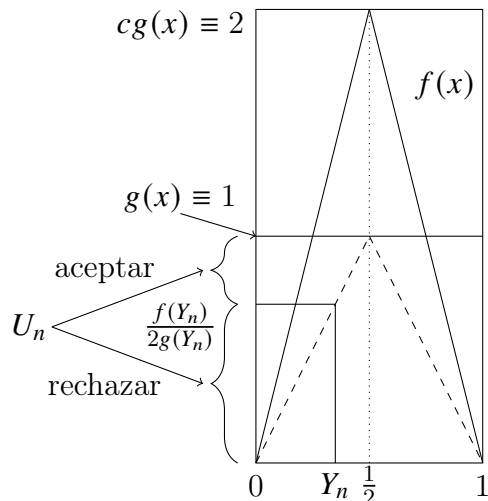
y sea

$$Z = Y_\tau$$

entonces  $Z$  tiene función de densidad  $f(x)$  y

$$E[\tau] = c.$$

**Ejemplo 5.4.2.** La siguiente gráfica resume las operaciones en el método A-R en un ejemplo artificial (donde el método A-R definitivamente no se necesita). Aquí la densidad objetivo  $f(x)$  es un triángulo de base  $[0, 1]$  y punta en  $(\frac{1}{2}, 2)$ , y  $g(x)$  es la distribución uniforme sobre  $[0, 1]$ . Se toma la  $c$  más pequeña posible, aquí  $c = 2$ . Por tanto, el número medio de pruebas para obtener una muestra es  $c = 2$ .



Tanto el método inverso como el método de aceptación-rechazo tienen contrapartes obvias cuando  $Z$  es una variable aleatoria discreta con valores en un espacio finito  $E = \{1, 2, \dots, r\}$ . Entonces se denota por  $\pi$  a la distribución de  $Z$ . El método inverso es en este caso teóricamente factible: Consiste en generar una variable aleatoria  $U$  distribuida uniformemente en  $[0, 1]$  y dejando  $Z = i$  si y solo si  $\sum_{l=1}^{i-1} \pi(l) \leq U < \sum_{l=1}^i \pi(l)$ . Cuando el tamaño  $r$  del espacio de estados  $E$  es grande, surgen problemas que se deben al tamaño pequeño de los intervalos de partición  $[0, 1]$  y al costo de precisión al calcularlo. En la simulación de un campo aleatorio, otra razón, quizás más importante, es la necesidad de enumerar las configuraciones, lo que implica codificar y decodificar un mapeo de los enteros al espacio de configuración.

El método A-R usa una distribución  $p$  sobre  $E$  tal que  $\frac{\pi(i)}{p(i)} \leq c < \infty$  para toda  $i \in E$ . El problema principal es que  $\pi$  a menudo se conoce sólo hasta un factor normalizador (el método inverso sufre de la misma dolencia).

### 5.4.3. MCMC

En una serie de aplicaciones, y especialmente en la simulación de grandes campos aleatorios que ocurren en la práctica, surgen todos los problemas anteriores, y la búsqueda de un generador de variables aleatorias sin estos problemas está en el origen del método de la cadena de Monte Carlo Markov.

Se construye una CMH irreducible y aperiódica  $\{X_n\}_{n \geq 0}$  con espacio de estados  $E$  con la distribución estacionaria  $\pi$ . Ya que  $E$  es finito, la cadena es ergódica [9], y para cualquier distribución inicial  $\mu$  y toda  $i \in E$ ,

$$\lim_{n \rightarrow \infty} P_\mu(X_n = i) = \pi(i) \quad (5.39)$$

y

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \phi(X_n) = E_{\pi}[\phi(X)]. \quad (5.40)$$

Por 5.39, cuando  $n$  es «grande», se puede considerar que  $X_n$  tiene una distribución cercana a  $\pi$ . Por supuesto, se quisiera saber con qué precisión  $X_n$  imita una variable aleatoria  $Z$  con valores en  $E$  y con distribución  $\pi$ . Para ello se buscará obtener estimaciones de la forma

$$|\mu \mathbf{P}^n - \pi| \leq A\alpha^n,$$

donde  $\alpha < 1$ . El teorema de Perron-Frobenius [9] da una respuesta teórica, en términos del segundo módulo de valor propio más grande (SLEM por sus siglas en inglés). Para que sea aplicable, es necesario encontrar buenos límites superiores del SLEM, ya que la estructura propia de una matriz de transición es a menudo muy difícil de obtener con exactitud.

Por el momento, el problema básico es el de diseñar el algoritmo MCMC. Se debe encontrar una matriz de transición ergódica  $\mathbf{P}$  sobre  $E$ , cuya distribución estacionaria  $\pi$  es la distribución objetivo. Hay infinitas matrices de transición de este tipo, y entre ellas hay infinitas que corresponden a una cadena reversible, es decir, tales que

$$\pi(i)p_{ij} = \pi(j)p_{ji}. \quad (5.41)$$

Se buscan soluciones de la forma

$$p_{ij} = q_{ij}\alpha_{ij} \quad (5.42)$$

para  $j \neq i$ , donde  $\mathbf{Q} = \{q_{ij}\}_{i,j \in E}$  es una matriz de transición arbitraria irreducible sobre  $E$ , llamada *matriz generadora de candidatos*: cuando el estado actual es  $i$ , el siguiente estado tentativo  $j$  se elige con probabilidad  $q_{ij}$ . Cuando  $j \neq i$ , este nuevo estado se acepta con probabilidad  $\alpha_{ij}$ . De lo contrario, el siguiente estado es el mismo estado  $i$ . Por tanto, la probabilidad resultante de pasar de  $i$  a  $j$  cuando  $i \neq j$  está dado por 5.42. Queda por seleccionar las probabilidades de aceptación  $\alpha_{ij}$ . Una forma bastante general de las probabilidades de aceptación que cubren los algoritmos MCMC más conocidos se debe a Hastings (1970)[17] ; ver (Fishman, 1996) [13] para bibliografía adicional.

$$\alpha_{ij} = \frac{s_{ij}}{1 + t_{ij}}, \quad (5.43)$$

donde  $\Sigma = \{s_{ij}\}_{i,j \in E}$  es una matriz simétrica y

$$t_{ij} = \frac{\pi(i)q_{ij}}{\pi(j)q_{ji}}. \quad (5.44)$$

Por supuesto,  $\Sigma$  debe seleccionarse de tal manera que se respete la restricción  $\alpha_{ij} \in [0, 1]$ . Se puede comprobar que se cumple la condición de reversibilidad 5.41 y, por tanto,  $\pi$  es la distribución estacionaria, mediante 2.6.1.1.

Las formas particulares del algoritmo de Hastings que siguen se deben a Metropolis et alii (1953) [24] y Barker (1965) [4], respectivamente.

**Ejemplo 5.4.3** (Algoritmo de Metropolis). Para satisfacer la restricción  $\alpha_{ij} \in [0, 1]$ , se debe tener

$$s_{ij} \leq 1 + \min(t_{ij}, t_{ji})$$

(recordando que  $\Sigma$  se asumió simétrica). La igualdad corresponde al algoritmo de Metropolis

$$\alpha_{ij} = \min\left(1, \frac{\pi(j)q_{ji}}{\pi(i)q_{ij}}\right). \quad (5.45)$$

Un caso especial es aquel para el que el mecanismo de generación de candidatos es puramente aleatorio, es decir,  $q_{ij}$  constante. Entonces

$$\alpha_{ij} = \min\left(1, \frac{\pi(j)}{\pi(i)}\right).$$

**Ejemplo 5.4.4** (Algoritmo de Barker). Con la elección  $s_{ij} = 1$ , se tiene el algoritmo de Barker, para el cual

$$\alpha_{ij} = \frac{\pi(j)q_{ji}}{\pi(j)q_{ji} + \pi(i)q_{ij}}. \quad (5.46)$$

En el caso especial de una selección puramente aleatoria del candidato,

$$\alpha_{ij} = \frac{\pi(j)}{\pi(j) + \pi(i)}.$$

Es importante observar que los algoritmos de Hastings requieren el conocimiento de la distribución objetivo  $\pi$  solo hasta una constante de normalización, ya que depende solo de las relaciones  $\pi(j)/\pi(i)$ . La última afirmación es cierta solo mientras se conozca la matriz  $\mathbf{Q}$  generadora de candidatos. Este no es el caso en el siguiente ejemplo.

**Ejemplo 5.4.5.** Con  $q_{ij} = \pi(j)$  y  $\alpha_{ij} = 1$ , para toda  $i, j \in E$  se tiene un muestreo independiente, esto es,  $p_{ij} = \pi(j)$  para toda  $i, j \in E$ .

---

# CONCLUSIONES

La naturaleza multidisciplinaria de esta investigación se presta a muchas preguntas, como se mencionó en los objetivos del presente trabajo, esta investigación responde a la creciente necesidad de un tratamiento unificado de temas relacionados con cadenas de Markov como campos de Gibbs finitos, procesos regenerativos a tiempo discreto y simulación Monte Carlo. La mayoría de estos son tratados en el último capítulo junto con algunos más, igualmente de gran importancia, como redes neuronales, algoritmos de simulación, modelos como el de Ising y el de acoplamiento. Todo esto a través de una línea teórica con los fundamentos necesarios para tratar cada tema con el cuidado adecuado incluyendo demostraciones clave y ejemplos donde se da otra perspectiva del tema.

En el Capítulo 1 se incluyeron los temas esenciales de la teoría de probabilidad para el estudio de las cadenas de Markov esto para evitar un alargamiento innecesario del tema, también se incluyó bibliografía que trata los fundamentos de probabilidad de manera amplia y profunda. Empezando con los conceptos básicos y concluyendo con la prueba de la Ley Fuerte de los Grandes Números de Kolmogorov. Una vez dejadas claras las bases el camino hacia las cadenas de Markov a tiempo discreto queda libre, empezando con la parte esencial de esta que es la matriz de transición, hablando de las recurrencias, análisis de primer paso para la resolución de problemas, la topología de la matriz de transición para la clasificación y análisis adecuado de los estados, teniendo los imprescindibles estados estacionarios, viendo las cadenas desde otra perspectiva cuando el tiempo se revierte, llegando al auge del capítulo 2 con la propiedad fuerte de Markov y ciclos regenerativos.

Todo esto suscita al estudio del comportamiento a largo plazo hablando de la convergencia en variación a través de las preguntas hechas en la sección: ¿Cuál es el comportamiento a largo plazo de la cadena cuando la distribución inicial  $\mu$  es arbitraria? Por ejemplo, ¿convergerá al equilibrio y en qué sentido?, dando una respuesta clara a través de una definición para la convergencia en variación. Usando este resultado se puede comenzar a hablar de un tema con un rango muy amplio de aplicaciones el cual es el método de acoplamiento y de la teoría de renovación, demostrando el teorema de Orey, definiendo la ecuación de renovación y finalmente demostrando el teorema de renovación. Se vio que este último teorema podría ser

usado para demostrar resultados en temas de convergencia, se comienza a hablar de procesos regenerativos y demostrando el teorema regenerativo. Por último para este capítulo se hace un estudio de la absorción o la probabilidad de permanecer para siempre en un conjunto transitorio, revisando los tiempos para que suceda esto y la matriz de absorción.

Con todas las herramientas necesarias se pasa a lo que son enteramente aplicaciones de las cadenas de Markov siguiendo esta línea particular de investigación en este trabajo. Primero hablando de el concepto de vecindades y especificación local para cadenas de Markov, definiendo los campos aleatorios a través de la idea de *sitios* y *espacios de fases*, una idea análoga a las variables aleatorias y, dada la analogía, pudiendo hablar de campos aleatorios de Markov.

Toda esta construcción da paso a las camarillas, potencial, y distribuciones de Gibbs, aquí se logra concretar la idea de unificar las cadenas de Markov con su aplicación en campos de Gibbs, y además se habla del Modelo de Ising y la Red Neuronal vistos desde esta aproximación markoviana. Para finalizar con esta relación Gibbs-Markov se logra ver que las distribuciones de Gibbs con una energía derivada de un potencial de Gibbs en relación con un sistema de vecindades son distribuciones de campos de Markov en relación con el mismo sistema de vecindades. Adicional a este resultado no se puede dejar de lado mencionar el muestro de Gibbs y la simulación Monte Carlo que son aplicaciones de gran importancia en la actualidad, aproximadas a través de la línea de investigación seguida en este trabajo.

---

# BIBLIOGRAFÍA

- [1] Lars Valerians Ahlfors. *Complex Analysis (Third ed.)* McGraw Hill Higher Education, 1980. ISBN: 0-07-085008-9.
- [2] David Aldous y James Allen Fill. «Reversible Markov Chains and Random Walks on Graphs». En: (2002).
- [3] Tom Andersson Håkanand Britton. *Coupling methods*. Springer New York, 2000. ISBN: 978-1-4612-1158-7.
- [4] Randolph Barker. *Estimation and Aggregation of Firm Supply Functions*. Wiley, 1965.
- [5] Robert G Bartle. *The Elements of Integration and Lebesgue Measure*. Wiley, 1995. ISBN: 978-0471042228.
- [6] Laurence A. Baxter. *Exactly Solved Models in Statistical Mechanics*. Academic Press, 1982. ISBN: 0-12-083180-5.
- [7] Patrick Billingsley. *Probability and Measure*. John Wiley & Sons, 1995. ISBN: 0-471-00710-2.
- [8] George Casella. *Statistical Inference*. Duxbury, 2001. ISBN: 0-534-24312-6.
- [9] M. F. Chen. *Eigenvalues, Inequalities, and Ergodic Theory*. Springer, 2005.
- [10] W. Doeblin. «Exposé sur la théorie des chaînes simples constantes de Markoff à un nombre fini d'états.» En: (1938).
- [11] William Feller. «A Direct Proof of Stirling's Formula». En: (1967). DOI: <https://doi.org/10.2307/2315671>.
- [12] William Feller. *An Introduction to Probability Theory and Its Applications*. John Wiley & Sons, 1967. ISBN: 978-0-321-79477-2.
- [13] George Fishman. *Monte Carlo*. Springer, 1996. ISBN: 978-0-387-94527-9.
- [14] Josiah Willard Gibbs. *Elementary principles in statistical mechanics*. Charles Scribner's Sons, 1902.
- [15] D Griffeath. «Coupling methods for nonhomogeneous Markov chains.» En: (1975).

- [16] G. R. Grimmett. *A Theorem about Random Fields*. London Mathematical Society, 1973.
- [17] W. K. Hastings. *Monte Carlo sampling methods using Markov chains and their applications*. Oxford University Press, 1970.
- [18] Ernst Ising. *Contribution to the Theory of Ferromagnetism*. Zeitschrift für Physik, 1925.
- [19] Gareth A Jones. *Elementary Number Theory*. Springer, 1998. ISBN: 978-3540761976.
- [20] John Lafferty y Andrew McCallum. *Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data*. University of Pennsylvania, 2001.
- [21] Torgny Lindvall. *Lectures on the Coupling method*. Wiley New York, 1992.
- [22] A. A. Markov. *Theory of Algorithms*. Academy of Sciences of the USSR, 1954.
- [23] Jerrold E. Marsden. *Basic Complex Analysis*. W. H. Freeman, 1999. ISBN: 0-7167-2877-X.
- [24] Nicholas Metropolis. *Equation of state calculations by fast computing machines*. The Journal of Chemical Physics, 1953.
- [25] James Munkres. *Topology*. Prentice Hall, 2000. ISBN: 84-205-3180-4.
- [26] J J O'Connor y E F Robertson. *Andrei Andreyevich Markov*. 2006. URL: <https://mathshistory.st-andrews.ac.uk/Biographies/Markov/>.
- [27] T. Ehrenfest P. Ehrenfest. «Physiit. Z.» En: (1907). DOI: <https://doi.org/10.1103/PhysRev.103.17>.
- [28] J. W Pitman. «Uniform rates of convergence for Markov chain transition probabilities.» En: (1974).
- [29] Pierre Priouret. *Markov Chains*. Springer, 2018. ISBN: 978-3-319-97703-4.
- [30] Yuri Prokhorov. *Strong law of large numbers*. Encyclopedia of Mathematics, 1933.
- [31] Jean Jacod & Philip Protter. *Probability Essentials*. Springer, 2004. ISBN: 978-3-540-43871-7.
- [32] Brian D. Ripley. *Stochastic Simulation*. Wiley, 1997. ISBN: 978-0-471-81884-7.
- [33] Sheldon Ross. *A First Course In Probability*. Prentice Hall, 2006. ISBN: 978-0-321-79477-2.
- [34] Walter Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, 1976. ISBN: 0-07-054235-X.
- [35] Michael Spivak. *Calculus*. Reverté Ediciones, 1996. ISBN: 968-6708-18-9.
- [36] Taylor y Karlin. *An Introduction to Stochastic Modeling*. Academic Press, 1998. ISBN: 978-0-12-684887-8.
- [37] David Williams. *Probability with Martingales*. Cambridge University, 1991. ISBN: 978-0-511-81365-8.