



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

EL FENÓMENO DEL CUT-OFF

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIO

P R E S E N T A :

FRANCISCO JAVIER ZÁRATE ALMAGUER

TUTOR

DR. ADRIÁN GONZÁLEZ CASANOVA SOBERÓN



CIUDAD UNIVERSITARIA, CD. MX., 2020



Universidad Nacional
Autónoma de México



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

1. Introducción	1
1.1. Una herramienta importante	1
1.2. El Fenómeno del Cutoff	1
1.3. Visión Histórica del Fenómeno del Cutoff	2
2. Cadenas de Markov Finitas	3
2.1. Cadenas de Markov	3
2.2. Distribución Estacionaria	11
2.3. Mapeo Aleatorio	16
2.4. Irreducibilidad y Aperiodicidad	18
2.5. Caminatas Aleatorias en Gráficas	20
2.6. Tiempos de Llegada y Tiempos de Retorno	22
2.7. Existencia de una Distribución Estacionaria	23
2.8. Unicidad de la distribución estacionaria	24
2.9. Reversibilidad y reversiones de tiempo	26
2.10. Clasificación de los Estados de una Cadena de Markov	28
3. Cadenas Importantes; Urna de Ehrenfest y El Hipercubo	32
3.1. Ruina Del Jugador	32
3.2. Coleccionista de Cupones	38
3.3. El hipercubo y el Modelo de la Urna de Ehrenfest	42
4. Tiempo de Mezcla de Cadenas de Markov	45
4.1. Distancia de Variación Total	45
4.2. Introducción al Acoplamiento y su Relación con la Distancia de Variación Total	48
4.3. Teorema de Convergencia	51
4.4. Estandarización de Distancia Desde la Estacionariedad	52
4.5. Tiempo de Mezcla	55
5. Acoplamiento	56
5.1. Introducción al Acoplamiento	56
5.2. Distancia de Variación Total Delimitada	58
5.3. Acoplamiento en el Hipercubo	59

<i>ÍNDICE GENERAL</i>	III
6. Tiempos Estacionarios Fuertes	61
6.1. Top-to-Random Shuffle	61
6.2. Tiempos Estacionarios	64
7. El Fenómeno del Cutoff	68
7.1. Un Paso Antes del Cutoff	68
7.2. Definición del Cutoff	72
7.3. Teorema, El Hipercubo n-dimensional Tiene un Cutoff	80
8. Conclusiones	85
9. Apéndice: Prueba Alternativa del Cutoff en el Hipercubo	86

Índice de figuras

2.1. Descomposición de todos los estados, entre el estado x y el estado y	7
2.2. Diagrama de transición de una la cadena de dos estados.	8
2.3. 2 posibles trayectorias de la cadena de dos estados, donde $E = \{1,2\}$ con probabilidades de transición $p=1/2=q$	9
2.4. Diagrama de transición de la urna de Ehrenfest.	10
2.5. 2 posibles trayectorias de la cadena de Ehrenfest, iniciando en el estado 1, con probabilidades de transición dadas para $N=5$	10
2.6. Observemos que la caminata aleatoria izquierda está definida en \mathbb{Z}_{10} es periódica ya que a cada paso va de un estado par a un estado impar o viceversa. Mientras que la caminata aleatoria derecha está definida \mathbb{Z}_9 es aperiódica. Ver [13].	17
2.7. Ejemplo de una gráfica con un conjunto de vértices $\{1,2,3,4,5\}$ y 5 aristas.	20
2.8. Del lado izquierdo podemos ver un ejemplo donde el estado y es accesible desde x , mientras que del lado derecho y este comunicado con x	28
2.9. x es esencial ya que w, y y z son accesibles desde x	28
3.1. Simulación de 6 juegos, con $i=4$ y $N=5$	33
3.2. 9 posibles escenarios para un coleccionista de cupones, que busca coleccionar $N=5$ cupones.	39
3.3. Hipercubo en tres dimensiones.	42
3.4. Diagrama de transición de un hipercubo en tres dimensiones.	43
3.5. 2 posibles trayectorias de la caminata aleatoria en el cubo, iniciando en el estado $(1,1,1)$, con probabilidades de transición uniforme.	44
3.6. Ejemplo: Peso de Hamming, suponiendo que $W_{t_0} = j$, el peso puede incrementar en 1 unidad, cuando la caminata aleatoria en el hipercubo n -dimensional haya modificado una de las $n - j$ coordenadas con valor 0, por otro lado, si modifica alguna de las j coordenadas con valor 1, decrecerá el peso en 1 unidad.	44
4.1. Si definimos $B = \{x : \mu(x) \geq \nu(x)\}$, notamos que la región I tiene área $\mu(B) - \nu(B)$. La región II tiene área $\mu(B^c) - \nu(B^c)$, ahora dado que μ y ν son distribuciones de probabilidad el área que contienen debe ser 1 por lo cual la región I y II deben tener la misma área y esa área es $\ \mu - \nu\ _{TV}$. Ver [13].	47
5.1. Acoplamiento de caminatas aleatorias definidas en $\{0,1,\dots,n\}$	56

5.2. Del proceso 1 elegimos una coordenada uniformemente al azar, (en la imagen seleccionamos la sexta). Después con probabilidad $1/2$, cambiamos el bit de la coordenada elegida, pero en el proceso 2, para que el valor de las coordenadas en ambos procesos sea el mismo. Ver [13]	59
6.1. Método de barajar.	61
7.1. Las cartas estarán enumeradas de abajo hacia arriba, es decir, la carta de abajo será la 1 y la carta superior será la n , posteriormente la carta n bajará en alguna posición aleatoria.	70
7.2. Para una cadena de Markov con cutoff, la gráfica de $d_n(t)$ contra t , cuando se ve en la escala de tiempo $t_{mix}^{(n)}$, se aproxima a una función escalonada cuando $n \rightarrow \infty$. Ver [13].	73

Capítulo 1

Introducción

En este capítulo introductorio presentamos algunos antecedentes teóricos e históricos del *fenómeno del cutoff*. Daremos definiciones formales, pero también uno de los objetivos de esta tesis, es lograr que el lector tenga una visión intuitiva del *cutoff*, así como de las herramientas que ocuparemos para desarrollar dicho tema, por ejemplo, el *acoplamiento*, nos permite un enfoque más intuitivo y claro de lo que es el *cutoff*. En los capítulos 2 y 3, daremos una breve pero concisa introducción a cadenas finitas de Markov. El concepto de caminatas aleatorias en grupos, en los capítulos 4, 5 y 6, presentaremos algunos métodos para analizar tasas de convergencia y mezcla de cadenas, como también desarrollaremos el tema del *acoplamiento* y para finalizar este proceso constructivo se presentara un ejemplo relativamente simple pero de suma importancia el cual es la caminata aleatoria en el hipercubo, ya que este ejemplo permite entender de forma clara en qué momento sucede el *cutoff* y aparece en varias áreas de la investigación y por lo cual es el tema principal de análisis de esta tesis.

1.1. Una herramienta importante

El término acoplamiento, en el campo de la probabilidad, se refiere a la práctica de construir dos (o más) medidas de probabilidad en un solo espacio medible para así poder compararlas. Por lo general, esto se realiza para deducir las propiedades de las medidas individuales o marginales para investigar las similitudes entre las dos. La aplicación principal del acoplamiento que se explotará en esta tesis será estimar tasas de convergencia al equilibrio de las cadenas ergódicas de Markov.

1.2. El Fenómeno del Cutoff

Considere una familia de cadenas tales que $\{(\Omega_n, P_n, \pi_n) : n \in \mathbb{N}\}$, cada una con su distancia correspondiente a la estacionariedad $d_n(t)$ y su tiempo de mezcla $t_{mix}^{(n)}$. Hablando en términos generales, el fenómeno del cutoff, ocurre durante un período de tiempo insignificante, conocido como ventana de cutoff, la distancia de variación total, cae abruptamente de un valor cercano a 1 a un valor cercano de 0. En otras palabras,

uno debe llegar justo hasta el punto del cutoff para que las cadenas estén mezcladas en términos de la distancia de variación total.

Del mismo modo, consideraremos posteriormente las nociones correspondientes de los tiempos de mezcla.

1.3. Visión Histórica del Fenómeno del Cutoff

El fenómeno de cutoff se identificó por primera vez para transposiciones aleatorias en el ingenioso trabajo de Diaconis y Shahshahani [17] y más tarde por Aldous para caminatas aleatorias en el hipercubo [1]. Aldous y Diaconis le dieron su nombre al fenómeno del cutoff en su artículo [4] de 1986 en el que verificaron el cutoff para varios esquemas de barajado de cartas y sugirieron el siguiente problema abierto, al que se refieren como *el problema más interesante: encuentre condiciones abstractas que aseguren que ocurra el fenómeno del cutoff*.

Aunque se cree que muchas familias de cadenas exhiben un cutoff, demostrar la ocurrencia de este fenómeno es a menudo una tarea extremadamente desafiante. La verificación de la ocurrencia del fenómeno es un área importante de probabilidad moderna, y a pesar del progreso notable en las últimas tres décadas, todavía hay relativamente pocos ejemplos que se entiendan completamente. Por lo cual daremos una pequeña lista de cadenas que cumplen con el *fenómeno del cutoff*.

1. Dinámica glauber del modelo de Ising, Lubetzky y Sly. Incluyendo algunos resultados extremadamente precisos sobre el caso de la red, utilizando su enfoque de percolación de información [6], [14] y [8]. También demostraron que la caminata aleatoria simple en una secuencia de gráficos regulares aleatorios exhibe un cutoff (con alta probabilidad sobre la elección del gráfico) [7] .
2. Berestycki y Peres, luego generalizaron este resultado a la configuración de grafos aleatorios con una secuencia de grados dada [2] con un grado de al menos 3.
3. Bordenave [3] trató la configuración de dígrafo aleatorio.
4. El cutoff para la transposición adyacente aleatoria fue verificado por Lacoïn [10], quien recientemente verificó el fenómeno del cutoff para la configuración sesgada junto con Labbé [12].
5. Miller y Peres [11] verificaron el fenómeno del cutoff para una caminata aleatoria simple en gráficos de faroles cuyos gráficos base son localmente uniformemente transitorios.
6. En esta tesis demostraremos que la caminata aleatoria en el hipercubo n -dimensional, tiene un cutoff, para lo cual ocuparemos el acoplamiento de cadenas de Markov y una técnica de barajeo de cartas conocido como *Top-to-Random Shuffle*.

Capítulo 2

Cadenas de Markov Finitas

2.1. Cadenas de Markov

Definición 2.1 Proceso estocástico

Un proceso estocástico es una colección de variables aleatorias $\{X_t : t \in T\}$ parametrizada por un conjunto T , llamado espacio parametral, en donde las variables toman valores en un conjunto E llamado espacio de estados.

En esta tesis consideraremos como espacio parametral el conjunto discreto $T = \{0, 1, 2, \dots\}$ y estos números se interpretarán como tiempos. En este caso se dice que el proceso es a tiempo discreto, y en general este tipo de procesos se denotará por $\{X_t : t = 0, 1, 2, \dots\}$ o explícitamente,

$$X_0, X_1, X_2, \dots$$

así, para cada t , X_t es el valor del proceso o estado del sistema al tiempo t .

Recordatorio 2.1 Si $T = \mathbb{Z}^+$, entonces llamamos a X un proceso estocástico de tiempo discreto, y si $T = [0, \infty)$, entonces X es un proceso estocástico a tiempo continuo.

Los posibles espacios de estados que consideraremos son subconjuntos de \mathbb{Z} . En particular, para poder hablar de variables aleatorias con valores en el espacio de estados E , es necesario asociar a este conjunto una σ -álgebra. Consideremos que E es un subconjunto de \mathbb{R} , puede tomarse la σ -álgebra de Borel de \mathbb{R} restringida a E , es decir, $E \cap B(\mathbb{R})$. Un proceso estocástico, también llamado proceso aleatorio, puede considerarse como una función de dos variables

$$X : T \times \Omega \rightarrow E,$$

tal que la pareja (t, ω) se le asocia el valor o estado $X(t, \omega)$, lo cual también puede escribirse como $X_t(\omega)$. Para cada valor de $t \in T$, el mapeo $\omega \rightarrow X_t(\omega)$ es una variable aleatoria, mientras que para cada $\omega \in \Omega$ fijo, la función $t \rightarrow X_t(\omega)$ es llamada una trayectoria o realización del proceso.

Definición 2.2 Cadena de Markov

Una cadena de Markov es un proceso estocástico a tiempo discreto $\{X_t : t = 0, 1, 2, \dots\}$, con espacio de estados discreto, y que satisface la propiedad de Markov, esto es, para cualquier entero $t \geq 0$, y para cualesquiera estados $x_0, x_1, \dots, x_t, x_{t+1}$, se cumple

$$P(X_{t+1} = x_{t+1} | X_0 = x_0, \dots, X_{t-1} = x_{t-1}, X_t = x_t) = P(X_{t+1} = x_{t+1} | X_t = x_t). \quad (2.1)$$

La ecuación (2.1) establece que la distribución de probabilidad del estado del proceso al tiempo futuro $t + 1$ depende únicamente del estado del proceso al tiempo t , y no depende de los estados en los tiempos pasados $0, 1, \dots, t - 1$.

Definición 2.3 Función de Transición

Sea $(X_t)_{t \geq 0}$, una cadena de Markov con espacio de estados E . La función $P(x, y)$, donde $x, y \in E$ está definida por:

$$P(x, y) = P(X_1 = y | X_0 = x), \quad x, y \in E. \quad (2.2)$$

La ecuación (2.2) se conoce como función de transición de una cadena.

Proposición 2.1 La familia de probabilidades de transición $P = P(x, y)$ cumple las siguientes dos propiedades.

1. $P(x, y) \geq 0, \quad x, y \in E,$
2. $\sum_y P(x, y) = 1, \quad x \in E.$

Demostración

1. La primera condición es trivial, partiendo del hecho que son probabilidad.
2. Primero veamos que nuestro espacio de estados E cumple la descomposición disjunta

$$E = \bigcup_y (X_1 = y).$$

Por lo tanto, para cualesquiera estados x y y ,

$$1 = P\left(\bigcup_{y \in E} (X_1 = y) | X_0 = x\right) = \sum_{y \in E} P(X_1 = y | X_0 = x) = \sum_{y \in E} P(x, y).$$

■

Las probabilidades de transición no sólo dependen de los estados si no también del instante en el cual se efectúa la transición. Cuando estas probabilidades son independientes del tiempo t , decimos que la cadena tiene probabilidades de transición homogénea en el tiempo, si esto sucede entonces

$$P(X_{t+1} = y | X_t = x) = P(X_1 = y | X_0 = x) = P(x, y).$$

Definición 2.4 Distribución Inicial

La función $\pi_0(x)$, $x \in E$ está definida por

$$\pi_0(x) = P(X_0 = x), \quad x \in E. \tag{2.3}$$

La ecuación (2.3) es conocida como **distribución inicial** de una cadena.

Proposición 2.2 La distribución inicial cumple las siguientes propiedades

1. $\pi_0(x) \geq 0, \quad x \in E,$
2. $\sum_x \pi_0(x) = 1.$

Demostración

La demostración de esta Proposición es análoga a la demostración realizada en la Proposición 2.1.



En el caso que $E = k$ podemos pensar un arreglo matricial, el cual tendrá la siguiente estructura

$$P = \begin{matrix} & \begin{matrix} 0 & 1 & \cdots & k \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ \vdots \\ k \end{matrix} & \begin{pmatrix} P(0,0) & P(0,1) & \cdots & P(0,k) \\ P(1,0) & P(1,1) & \cdots & P(1,k) \\ \vdots & \vdots & \ddots & \vdots \\ P(k,0) & P(k,1) & \cdots & P(k,k) \end{pmatrix} \end{matrix}.$$

la matriz P se conoce como matriz de transición, si el número de estados es finito, entonces P será una matriz cuadrada cuya dimensión es $|E| \times |E|$.

Por la Proposición 2.1 tenemos

$$P(x, y) = P(X_{t+1} = y | X_t = x) \geq 0, \quad x, y \in E \tag{2.4}$$

y

$$\sum_{y \in E} P(x, y) = \sum_{y \in E} P(X_{t+1} = y | X_t = x) = 1, \quad x, y \in E. \tag{2.5}$$

Así cada renglón de la matriz P representa a las funciones de transición y además si una matriz cumple (2.4) y (2.5) se dice que es una matriz estocástica o de Markov.

Definición 2.5 Función Transición en t Pasos

Sea $(X_t)_{t \geq 0}$, una cadena de Markov con espacio de estados E . La función $P^t(x, y)$, donde $x, y \in E$ está definida por:

$$P^t(x, y) = P(X_t = y | X_0 = x), \quad x, y \in E. \quad (2.6)$$

La ecuación (2.6), corresponde a la probabilidad de pasar del estado x al estado y en t pasos. Al igual que en la matriz de transición a un paso, podemos realizar un arreglo matricial con las probabilidades de transición a t pasos

$$P^t = \begin{matrix} & \begin{matrix} 0 & 1 & \cdots & k \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ \vdots \\ k \end{matrix} & \begin{pmatrix} P^t(0,0) & P^t(0,1) & \cdots & P^t(0,k) \\ P^t(1,0) & P^t(1,1) & \cdots & P^t(1,k) \\ \vdots & \vdots & \ddots & \vdots \\ P^t(k,0) & P^t(k,1) & \cdots & P^t(k,k) \end{pmatrix} \end{matrix}.$$

Algo que nos gustaría identificar, son las probabilidades de transición pero a t pasos, es decir, estando en un estado x cual es la probabilidad de llegar al estado y en un número de pasos t , formalmente nos referimos, $P^t(x, y) = P(X_t = y | X_0 = x)$, para esto enunciaremos el siguiente teorema.

Teorema 2.1 Ecuaciones de Chapman-Kolmogorov

Si $P^t = (P^t(x, y))$ es la matriz de transición a t pasos de una cadena de Markov, entonces

$$P^t(x, y) = \sum_k P^r(x, k)P^s(k, y). \quad (2.7)$$

Para cualquier par fijo de enteros no-negativos r y s que satisfagan $r + s = t$, donde definimos

$$P^0(x, y) = \begin{cases} 1 & x = y, \\ 0 & x \neq y. \end{cases}$$

Demostración

Observemos la figura 2.1, buscaremos hacer la partición de todos los estados que existen entre el estado x y el estado y en el momento r .

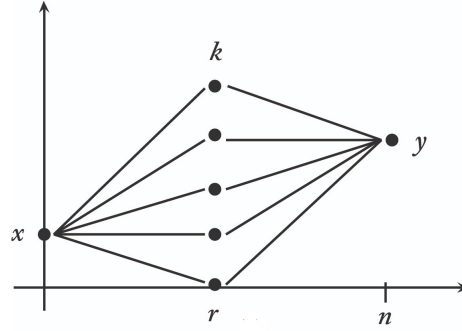


Figura 2.1: Descomposición de todos los estados, entre el estado x y el estado y .

$$\begin{aligned}
 P^t(x, y) &= P(X_t = y | X_0 = x) \\
 &= \sum_{k \in E} P(X_t = y, X_r = k | X_0 = x) \\
 &= \sum_{k \in E} \frac{P(X_t = y, X_r = k, X_0 = x)}{P(X_0 = x)} \\
 &= \sum_{k \in E} \frac{P(X_t = y, X_r = k, X_0 = x)}{P(X_r = k, X_0 = x)} \frac{P(X_r = k, X_0 = x)}{P(X_0 = x)} \\
 &= \sum_{k \in E} P(X_t = y | X_r = k, X_0 = x) P(X_r = k | X_0 = x) \\
 &= \sum_{k \in E} P(X_t = y | X_r = k) P(X_r = k | X_0 = x) \\
 &= \sum_{k \in E} P^r(x, k) P^s(k, y).
 \end{aligned}$$

■

Notemos que la ecuación (2.7), representa la multiplicación de las matrices P^r y P^s , de modo que P^t , es simplemente la t -ésima potencia de P .

$$P^{(t)} = \begin{pmatrix} P^t(0,0) & P^t(0,1) & \cdots & P^t(0,k) \\ P^t(1,0) & P^t(1,1) & \cdots & P^t(1,k) \\ \vdots & \vdots & \ddots & \vdots \\ P^t(k,0) & P^t(k,1) & \cdots & P^t(k,k) \end{pmatrix} = \begin{pmatrix} P(0,0) & P(0,1) & \cdots & P(0,k) \\ P(1,0) & P(1,1) & \cdots & P(1,k) \\ \vdots & \vdots & \ddots & \vdots \\ P(k,0) & P(k,1) & \cdots & P(k,k) \end{pmatrix}^t = P^t.$$

Ejemplo 2.1 Cadena de Dos Estados

Sea $E = \{x, y\}$ y $(X_t)_{t \geq 0}$, la cadena de Markov. Cuando el proceso se encuentra en el estado x , se mantendrá en x con probabilidad $1 - p$ y se moverá al estado y con probabilidad p , pero si se encuentra en el estado y se moverá al estado x con probabilidad q y con probabilidad $1 - q$ se mantendrá en el estado y esto es

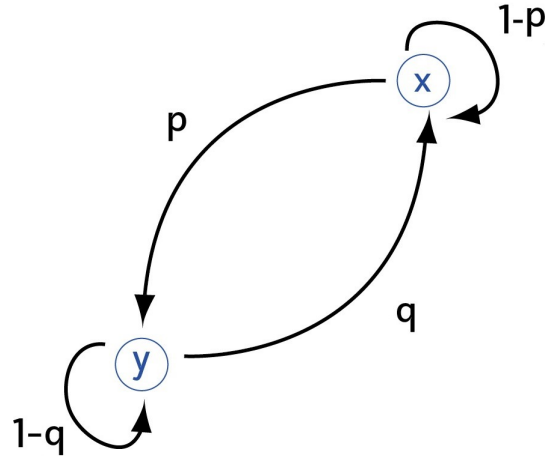


Figura 2.2: Diagrama de transición de una la cadena de dos estados.

$$P\{X_1 = x|X_0 = x\} = 1 - p, \quad P\{X_1 = y|X_0 = x\} = p \quad (2.8)$$

y

$$P\{X_1 = x|X_0 = y\} = q, \quad P\{X_1 = y|X_0 = y\} = 1 - q \quad (2.9)$$

por lo que tenemos el siguiente arreglo matricial

$$P = \begin{pmatrix} P(x,x) & P(x,y) \\ P(y,x) & P(y,y) \end{pmatrix} = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}.$$

Ahora, si queremos ver la probabilidad de estar en x y querer ir a y en dos pasos, o querer seguir estando en x después de dos pasos, esto lo podemos ver como

$$P(X_2 = x|X_0 = x) = (1-p)(1-p) + pq \quad (2.10)$$

y

$$P(X_2 = y|X_0 = x) = (1-p)p + p(1-q). \quad (2.11)$$

Continuando con este proceso podremos conocer las probabilidades de transición a t pasos, para los estados x y y , como se definió en la ecuación (2.6), las cuales podemos acomodar en un vector renglón de la siguiente manera

$$\mu_t := (P(X_t = x|X_0 = x), P(X_t = y|X_0 = x)). \quad (2.12)$$

Si asumimos que el proceso inicia en el estado x , podemos escribir $\mu_0 = (1, 0)$, y si queremos calcular las probabilidades de que se encuentre en algún otro estado, solo bastara calcular $\mu_1 = \mu_0 P$. Multiplicar por P a la derecha actualiza la distribución a un paso

$$\mu_t = \mu_{t-1} P \quad \text{para todo } t \geq 1. \quad (2.13)$$

De hecho, para cualquier distribución inicial μ_0

$$\mu_t = \mu_0 P^t \quad \text{para todo } t \geq 0. \quad (2.14)$$

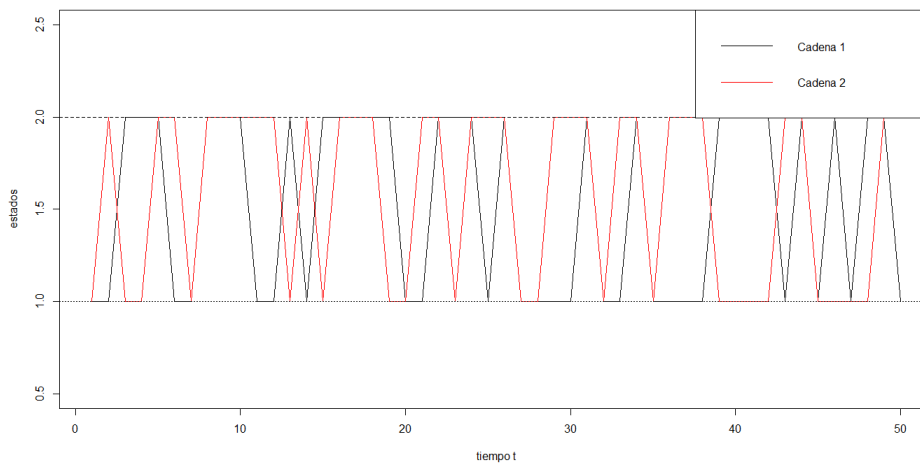


Figura 2.3: 2 posibles trayectorias de la cadena de dos estados, donde $E = \{1, 2\}$ con probabilidades de transición $p=1/2=q$.

Ejemplo 2.2 Cadena de Ehrenfest

Consideremos dos urnas A y B dentro de las cuales se encuentran distribuidas N bolas, podemos asumir que en la urna A se encuentran i bolas mientras que en la urna B hay $N - i$ bolas. En cada tiempo t se escoge una bola al azar y se cambia de urna, podemos suponer que cada bola esta numerada y seleccionamos un número al azar y buscamos la bola que tenga ese número y se cambia de urna.

Definamos los elementos de la cadena de Ehrenfest, sea X_t el número de bolas que hay en la urna A después de t extracciones. Por lo cual $\{X_t : t = 0, 1, \dots\}$ es una cadena de Markov con espacio de estados finito $E = \{0, 1, \dots, N\}$.

Supongamos que el proceso se encuentra en el estado i , la cadena de Markov solo tiene dos estados posibles, pasar al estado $i - 1$ ó $i + 1$, de tal manera que tenemos las siguientes probabilidades:

$$P\{X_1 = i - 1 | X_0 = i\} = \frac{i}{N}, \quad P\{X_1 = i + 1 | X_0 = i\} = \frac{N - i}{N}, \quad \text{para } i = 1, 2, \dots, N.$$

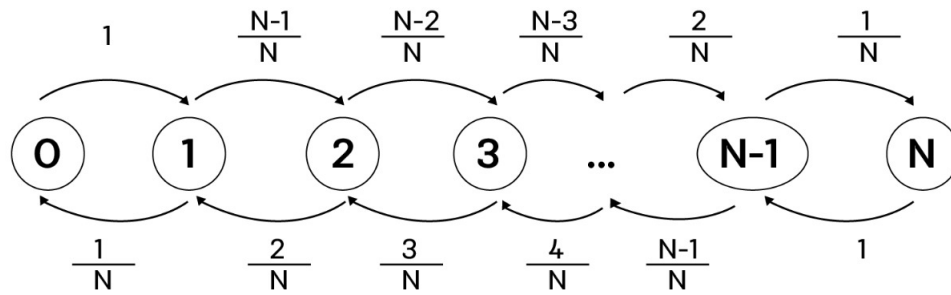


Figura 2.4: Diagrama de transición de la urna de Ehrenfest.

Observemos que $P(0, 1) = 1$ y $P(N, N - 1) = 1$, además su matriz de transición es:

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ \frac{1}{N} & 0 & \frac{(N-1)}{N} & 0 & \dots & 0 & 0 \\ 0 & \frac{2}{N} & 0 & \frac{(N-2)}{N} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 & \frac{1}{N} \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix}.$$

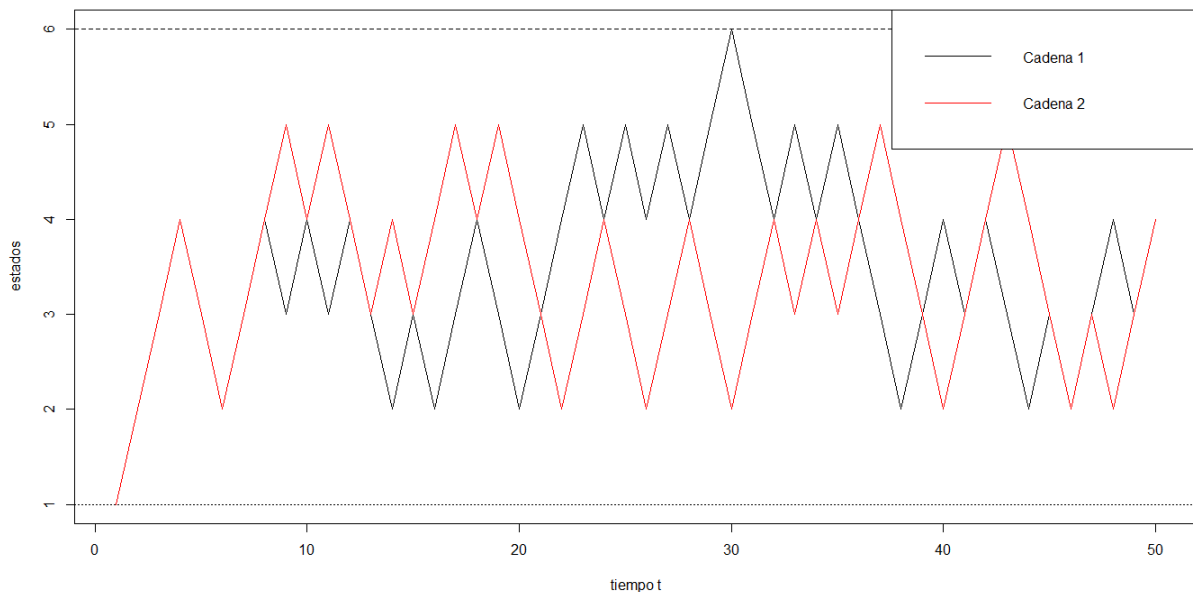


Figura 2.5: 2 posibles trayectorias de la cadena de Ehrenfest, iniciando en el estado 1, con probabilidades de transición dadas para $N=5$.

Una pregunta que nos surge, ¿qué sucede cuando t es muy grande?

2.2. Distribución Estacionaria

Definición 2.6 Distribución Estacionaria *Una distribución de probabilidad $\pi = (\pi(0), \pi(1), \dots)$ es estacionaria o invariante para una cadena de Markov con matriz de transición P si*

$$\sum_{x \in E} \pi(x) P(x, y) = \pi(y), \quad \forall y \in E. \quad (2.15)$$

Diremos que π es una distribución estacionaria para la cadena $(X_t)_{t \geq 0}$, con matriz de transición P .

Notemos que si tenemos una matriz de transición P y π es el vector de la distribución estacionaria, podemos reescribir matricialmente la ecuación (2.15) como $\pi' P = \pi$, donde π es un vector columna y π' es su transpuesto.

Proposición 2.3 *Sea $(X_t)_{t \geq 0}$, una cadena de Markov con distribución estacionaria π y matriz de transición P , entonces*

$$\sum_x \pi(x) P^n(x, y) = \pi(y), \quad \forall y \in E. \quad (2.16)$$

Demostración

Procederemos por inducción matemática

1) Base inductiva: Utilizando la ecuación Chapman Kolmogorov, (2.7), tenemos

$$\begin{aligned} \sum_x \pi(x) P^2(x, y) &= \sum_x \pi(x) \sum_k P(x, k) P(k, y) = \sum_k \left(\sum_x \pi(x) P(x, k) \right) P(k, y) \\ &= \sum_k \pi(k) P(k, y) = \pi(y). \end{aligned}$$

2) Paso inductivo: Supongamos que se cumple para $n-1$, por demostrar que también se cumple para n .

$$\begin{aligned} \sum_x \pi(x) P^n(x, y) &= \sum_x \pi(x) \sum_k P^{n-1}(x, k) P(k, y) \\ &= \sum_k \left(\sum_x \pi(x) P^{n-1}(x, k) \right) P(k, y) \\ &= \sum_k \pi(k) P^n(x, y) = \pi(y). \end{aligned}$$

Por lo tanto, si la distribución del estado X_0 es π , entonces la ecuación (2.16) implica que, para todo t ,

$$P(X_t = y) = \pi(y), \quad \forall y \in E.$$

Ahora supongamos que la distribución de X_t es constante para todo t . Esto quiere decir que la distribución estacionaria representa una distribución de equilibrio del proceso: si el proceso se inicia con una distribución estacionaria entonces es estrictamente estacionario.

Ahora supongamos que la distribución de X_t es independiente de t , entonces la distribución inicial π_0 debe satisfacer

$$\pi_0(y) = P(X_0 = y) = P(X_1 = y) = \sum_x \pi_0(x)P(x, y).$$

Por lo tanto π_0 satisface la ecuación (2.16) y además si la distribución de X_t es constante para todo t si y solo si la distribución inicial es una distribución estacionaria. ■

Teorema 2.2 *Sea $(X_t)_{t \geq 0}$, una cadena de Markov con espacio de estados finito y matriz de transición P . Supongamos que para algún estado $x \in E$ se cumple*

$$\lim_{t \rightarrow \infty} P^t(x, y) := \pi(y) \quad \text{para todo } y \in E.$$

Entonces el vector $\pi(y)$, $y \in E$ es un vector de probabilidad estacionario.

Demostración

Dado $\pi(y)$, sabemos que $0 \leq \pi(y) \leq 1$ para todo $y \in E$, además se cumple que para las potencias de la matriz de transición P , $0 \leq P^t(x, y) \leq 1$, para todo $t \geq 1$ y para $x, y \in E$ tenemos

$$\sum_{y \in E} \pi(y) = \sum_{y \in E} \lim_{t \rightarrow \infty} P^t(x, y) = \lim_{t \rightarrow \infty} \sum_{y \in E} P^t(x, y) = 1.$$

Podemos intercambiar la suma y el límite ya que E es finito. Solo nos falta ver qué π es un vector de probabilidad estacionario, pero por la ecuación de Chapman-Kolmogorov (2.7) tenemos que para todo $y \in E$.

$$\pi(y) = \lim_{t \rightarrow \infty} P^{t+1}(x, y) = \lim_{t \rightarrow \infty} \sum_{k \in E} P^t(x, k)P(k, y) = \sum_{k \in E} \lim_{t \rightarrow \infty} P^t(x, k)P(k, y) = \sum_{k \in E} \pi(k)P(k, y).$$

Ejemplo 2.3 *Retomando nuestro ejemplo de la cadena de dos estados 2.1, para encontrar la distribución estacionaria de esta cadena, enunciaremos 2 teoremas de álgebra lineal, los cuales no demostraremos ya que no aporta gran relevancia para el tema que estamos tratando, para leer una prueba ver [9] (Pag.271).* ■

Teorema 2.3 Sea A una matriz de transición de n con n valores propios distintos $\lambda_1, \lambda_2, \dots, \lambda_n$. Sean v_1, v_2, \dots, v_n vectores columna de \mathbb{R}^n tales que v_i es un vector propio correspondiente a λ_i para $i=1, 2, \dots, n$. Sea C la matriz n que tiene a v_i como i -ésimo vector columna. Entonces C es invertible y $C^{-1}AC = D$, con D la matriz cuyas entradas están dadas por $d_{i,i} = \lambda_i$ y $d_{i,j} = 0$ si $i \neq j$, $i, j \in 1, 2, \dots, n$. Además, la k -ésima potencia de A , está dada por:

$$A^{(k)} = CD^{(k)}C^{-1},$$

y las entradas D^k están dadas por $d_{i,j}^{(k)} = d_{i,j}^k$ para $i, j \in 1, 2, \dots, n$.

Para cualquier vector $x \in \mathbb{R}^n$ se tiene que:

$$A^{(k)}x = r_1\lambda_1^k v_1 + r_2\lambda_2^k v_2 + \dots + r_n\lambda_n^k v_n$$

donde $(r_i, i \in \{1, \dots, n\})$ están dados por:

$$C^{-1}x = (r_1, r_2, \dots, r_n)^t$$

Teorema 2.4 Sea $A \in M_{n \times n}(\mathbb{C})$ tales que las condiciones siguientes se satisfacen:

1. Si λ es un eigenvalor de A entonces $|\lambda| \leq 1$.
2. Si λ es un eigenvalor de A y si $|\lambda|=1$, entonces $\lambda = 1$.
3. A es diagonalizable.

Entonces existe $\lim_{m \rightarrow \infty} A^m$.

Para un mejor entendimiento del ejemplo y facilitar los cálculos consideremos

$$P = \begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

donde los eigenvalores de la matriz P están dados por $|P - \lambda I| = 0$, por lo que tenemos,

$$\begin{vmatrix} a-\lambda & b \\ c & d-\lambda \end{vmatrix} = (a-\lambda)(d-\lambda) - cd = 0.$$

Resolviendo esta ecuación, usando que $a + b = 1$ y $c + d = 1$ llegamos a que $\lambda_1 = 1$ y $\lambda_2 = a + d - 1$. Observemos que λ_2 está expresado en términos de a y b . Para calcular los eigenvectores correspondientes a los eigenvalores $\lambda_1 = 1$ resolvemos el sistema matricial

$$\begin{pmatrix} a-1 & b \\ c & d-1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

y usando el hecho de que $c = 1 - d$ y $b = 1 - a$, obtenemos el eigenvector

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Para calcular el eigenvector correspondiente al eigenvalor $\lambda_2 = a + d - 1$, resolvemos el sistema matricial

$$\begin{pmatrix} 1-d & 1-a \\ 1-d & 1-a \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

y obtenemos el eigenvector propio

$$\begin{pmatrix} -\frac{b}{c} \\ 1 \end{pmatrix}.$$

Por lo tanto P se puede ver como

$$P = \begin{pmatrix} 1 & -\frac{b}{c} \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & a+d-1 \end{pmatrix} \begin{pmatrix} 1 & \frac{b}{c} \\ -1 & 1 \end{pmatrix} \frac{1}{1+\frac{b}{c}}.$$

Por el Teorema 2.3 tenemos que la primera matriz está formada por los vectores propio encontrados, la segunda matriz es la matriz diagonal, cuya diagonal son los valores propios de P y la tercera matriz es la inversa de la primera matriz, es decir,

$$P = CDC^{-1},$$

lo que implica que $P^t = CD^t C^{-1}$, esto es,

$$P^t = \begin{pmatrix} 1 & -\frac{b}{c} \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & (a+d-1)^t \end{pmatrix} \begin{pmatrix} 1 & \frac{b}{c} \\ -1 & 1 \end{pmatrix} \frac{1}{1+\frac{b}{c}}.$$

Resolviendo este producto de matrices llegamos a que

$$P^t = \begin{pmatrix} \frac{c+b(a+d-1)^t}{c+b} & \frac{b-b(a+d-1)^t}{c+b} \\ \frac{c-c(a+d-1)^t}{c+b} & \frac{b+c(a+d-1)^t}{c+b} \end{pmatrix}.$$

Ahora por Teorema 2.4

$$\lim_{t \rightarrow \infty} P^{(t)} = \begin{pmatrix} \frac{c}{c+b} & \frac{b}{c+b} \\ \frac{c}{c+b} & \frac{b}{c+b} \end{pmatrix}.$$

Por lo tanto, la distribución estacionaria de la cadena de dos estados es

$$\pi = \left(\frac{q}{p+q}, \frac{p}{p+q} \right).$$

Ejemplo

Busquemos la distribución estacionaria de la cadena de Ehrenfest (Ejemplo 2.2). Para este caso resolveremos el sistema de ecuaciones $\pi P = \pi$ considerando que $\sum_y \pi(y) = 1$. En el Ejemplo 2.2 vimos que la matriz de transición está dada por

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ \frac{1}{N} & 0 & \frac{(N-1)}{N} & 0 & \cdots & 0 & 0 \\ 0 & \frac{2}{N} & 0 & \frac{(N-2)}{N} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & \frac{1}{N} \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}.$$

Notemos que $\pi P = \pi$ es equivalente al siguiente sistema de ecuaciones

$$\begin{cases} \pi(0) = \frac{1}{N}\pi(1), \\ \pi(1) = \pi(0) + \frac{2}{N}\pi(2), \\ \pi(2) = \frac{N-1}{N}\pi(1) + \frac{3}{N}\pi(3), \\ \vdots \\ \pi(N-1) = \frac{2}{N}\pi(N-1) + \pi(N), \\ \pi(N) = \frac{1}{N}\pi(N-1), \\ \sum_y \pi(y) = 1. \end{cases}$$

si y solo si

$$\begin{cases} \pi(0) = \frac{1}{N}\pi(1), \\ \pi(1) = \frac{2\pi(2)}{(N-1)}, \\ \pi(2) = \frac{3\pi(3)}{(N-2)}, \\ \vdots \\ \pi(N-1) = \frac{N\pi(N)}{(1)}, \\ \sum_y \pi(y) = 1. \end{cases}$$

Ahora probaremos por inducción que

$$\pi(k) = \binom{N}{k} \pi(0).$$

Para $k = 0, 1$ es cierto, dado que

$$\pi(0) = \binom{N}{0} \pi(0) = 1 \quad y \quad \pi(1) = \binom{N}{1} \pi(0) = \frac{N!}{1!(N-1)!} = N.$$

Supongamos que es cierto para $1 < k < N$. Por demostrar para $k + 1$, notemos que

$$\pi(k+1) = \frac{N}{k+1} \left(\pi(k) - \pi(k-1) \frac{N-k+1}{N} \right) = \frac{N}{k+1} \left(\binom{N}{k} - \binom{N}{k-1} \frac{N-k+1}{N} \right) \pi(0) = \binom{N}{k+1} \pi(0).$$

Finalmente,

$$\pi(N) = \frac{1}{N} \pi(N-1).$$

Ahora notemos que si sumamos todas las ecuaciones tenemos como resultado

$$1 = \pi(0) \sum_{k=0}^N \binom{N}{k} = \pi(0) 2^N$$

de donde se obtiene que $\pi(0) = \frac{1}{2^N}$. Notemos que no solo encontramos su distribución estacionaria, sino que además acabamos de demostrar que $\pi \sim Bin(N, p)$, con $p = \frac{1}{2}$.

2.3. Mapeo Aleatorio

Definición 2.7 Representación de un Mapeo Aleatorio

Una representación de mapeo aleatorio de una matriz de transición P en el espacio de estado E es una función determinista $f : E \times \Lambda \rightarrow E$, con una variable aleatoria Z con valores en Λ , que satisfacen

$$P \{ f(x, Z) = y \} = P(x, y), \tag{2.17}$$

donde $f(x, Z)$ es la función de mapeo aleatorio.

Si la sucesión Z_1, Z_2, \dots son variables aleatorias independientes donde cada uno tiene la misma distribución que Z , la variable aleatoria X_0 tiene distribución μ y es independiente de $(Z_t)_{t \geq 1}$, entonces la sucesión $(X_t)_{t \geq 0}$ definida por

$$X_t = f(X_{t-1}, Z_t) \quad \text{para } n \geq 1$$

es una cadena de Markov con matriz de transición P y distribución inicial μ .

Ejemplo 2.4 Caminata Aleatoria en Ciclos

Sea $E = \mathbb{Z}_n = \{0, 1, \dots, n-1\}$, el conjunto de residuos modulo n , con matriz de transición

$$\begin{cases} 1/2, & \text{si } k = j+1 \pmod{n}, \\ 1/2, & \text{si } k = j-1 \pmod{n}, \\ 0, & \text{en otro caso.} \end{cases}$$



Figura 2.6: Observemos que la caminata aleatoria izquierda está definida en \mathbb{Z}_{10} es periódica ya que a cada paso va de un estado par a un estado impar o viceversa. Mientras que la caminata aleatoria derecha está definida \mathbb{Z}_9 es aperiódica. Ver [13].

Es decir, suponga que Δ_i es una variable aleatoria la cual tomará valores -1 y $+1$, por ejemplo, si la cadena se encuentra en el estado $j \in \mathbb{Z}_n$, entonces el próximo estado es $j + \Delta_i \pmod{n}$.

$$P(f(j, \Delta_i) = k) = P\left\{\left(j + 1_{\Delta_i > \frac{1}{2}} - 1_{\Delta_i < \frac{1}{2}}\right) \pmod{n} = k\right\} = P(j, k). \quad (2.18)$$

Por lo tanto, la ecuación (2.18) genera una representación de mapeo aleatorio. Lo que nos lleva a la siguiente proposición.

Proposición 2.4 *Para cada cadena de Markov en E finito con matriz de transición P admite una función de mapeo aleatorio.*

Demostración

Sea $(X_t)_{t \geq 0}$ una cadena de Markov con matriz de transición P y espacio de estados $E = \{x_1, \dots, x_n\}$. Definamos el siguiente intervalo $\Lambda = [0, 1]$, en el cual Z_0, Z_1, \dots independientes tomarán valores uniformemente en Λ , definimos el conjunto $F_{j,k} = \sum_{i=1}^k P(x_j, x_i)$ y

$$f(x_j, z) := x_k \quad \text{donde} \quad F_{j,k-1} < z \leq F_{j,k}.$$

Por lo tanto, tenemos que

$$P\{f(x_j, Z) = x_k\} = P\{F_{j,k-1} < Z < F_{j,k}\} = P(x_j, x_k).$$

■

2.4. Irreducibilidad y Aperiodicidad

Definición 2.8 Cadena de Markov Irreducible

Se dice que una cadena de Markov es irreducible si para cualesquiera 2 estados $x, y \in E$ existe $t \in T$ tal que

$$P^t(x, y) > 0.$$

En otras palabras, lo que nos dice la Definición 2.8, es que es posible llegar de cualquier estado x a cualquier otro estado y .

Definición 2.9 Tiempo de Llegada

Para $x \in E$, se define el **tiempo de llegada** como

$$\tau_x := \min\{t \geq 0 : X_t = x\}. \quad (2.19)$$

Definición 2.10 Primer Tiempo de Retorno

Para $x \in E$, se define como **primer tiempo de retorno** para $X_0 = x$ como

$$\tau_x^+ := \min\{t \geq 1 : X_t = x\}. \quad (2.20)$$

Definición 2.11 Periodo

El periodo de un estado x es un número entero no negativo denotado por $d(x)$, y definido como sigue:

$$d(x) = m.c.d. \{t \geq 1 : P^t(x, x) > 0\},$$

en donde *m.c.d.* significa máximo común divisor.

Para alguna cadena donde $P^t(x, x) = 0$ se define $d(x) = 0$, pero si $d(x) = 1$ entonces diremos que el estado x es aperiódico.

Lema 2.1 Para una cadena de Markov irreducible, el período de cualesquiera dos estados es igual. Por esta razón, tiene sentido decir el período de la cadena de Markov.

Demostración

Sean x y y estados fijos en E . Existen enteros no negativos r y l tales que $P^r(x, y) > 0$ y $P^l(x, y) > 0$. Definamos $\Gamma(x) := \{t \geq 1 : P^t(x, x) > 0\}$ el conjunto de veces en que es posible que la cadena regrese a la posición inicial x . Sea $m = r + l$, tal que $m \in \Gamma(x) \cap \Gamma(y)$ y $\Gamma(x) \subset \Gamma(y) - m$. Donde el $m.c.d(\Gamma(y))$ divide a todos los elementos de $\Gamma(x)$. Por lo tanto $m.c.d(\Gamma(y))$. Por un argumento completamente análogo, $m.c.d(\Gamma(x)) \leq m.c.d(\Gamma(y))$.

La desigualdad opuesta es análoga.

■

Proposición 2.5 Si P es aperiódica e irreducible, entonces hay un número entero r_0 tal que $P^r(x, y) > 0 \forall x, y \in E$ y $r \geq r_0$.

Demostración

Para la demostración enunciaremos el Lema de Schur, para una prueba ver [13], Lema 1.30.

Lema 2.2 Si $S \subset \mathbb{Z}^+$ tiene $m.c.d.(S) = g_s$, entonces hay un número entero m_s tal que para todo $m \geq m_s$ el producto mg_s puede escribirse como una combinación lineal de elementos de S con coeficientes enteros no negativos.

Recordatorio 2.2 El número entero más grande que no puede representarse como una combinación entera no negativa de elementos de S se llama número de Frobenius.

Sea $x \in E$ y definamos $\Gamma(x) = \{t \geq 1 : P^t(x, x) > 0\}$. Como la cadena es aperiódica, el $m.c.d(\Gamma)$ es 1. El conjunto $\{\Gamma\}$ es cerrado bajo la suma: si $s, t \in \{\Gamma\}$, entonces $P^{t+s}(x, x)$ y ya que $s + t \in \{\Gamma\}$. Por lo tanto, existe un $t(x)$ tal que $t \geq t(x)$ implica que $t \in \{\Gamma\}$. Por irreducibilidad sabemos que para cualquier $y \in E$ existe un $r = r(x, y)$ tal que $P^r > 0$. Por lo tanto, para $t \geq t(x) + r$,

$$P^t(x, y) \geq P^{t-r}(x, x)P^r(x, y) > 0.$$

Para $t \geq t'(x) := t(x) + \max_{y \in E} r(x, y)$, tenemos $P^t(x, y) > 0$ para todo $y \in E$. Finalmente, si $t \geq \max_{x \in E} \{t'(x)\}$, entonces $P^t(x, y) > 0$ para todo $x, y \in E$.

■

Definición 2.12 Matriz de Transición Q

Sea $(X_t)_{t \geq 0}$, una cadena de Markov, con matriz de transición P entonces una versión simple (ó perezosa) de P es aquella en la que definimos

$$Q = \frac{(P + I)}{2},$$

de modo que con probabilidad $\frac{1}{2}$, la cadena se siente **perezosa** y permanece en el mismo estado.

La importancia de la Definición 2.12 es que las cadenas de Markov con matriz de transición Q son aperiódicas.

Ejemplo 2.5 Caminata Aleatoria en Ciclos con Matriz Q

Retomando el Ejemplo 2.4, para todo $n \geq 1$, la caminata aleatoria en **n-ciclos** es irreducible, entonces la caminata aleatoria en cualquier ciclo de longitud par es periódica, ya que $m.c.d \{t : P^t(x, x) > 0\} = 2$.

Nota 2.1 La caminata aleatoria en ciclos de longitud impar es aperiódica, ver la Figura 2.3.

Entonces para $n \geq 3$, la matriz de transición Q para la caminata aleatoria en **n-ciclos** es

$$Q_{jk} = \begin{cases} \frac{1}{4} & \text{si } k \equiv j+1 \pmod{n}, \\ \frac{1}{2} & \text{si } k \equiv j \pmod{n}, \\ \frac{1}{4} & \text{si } k \equiv j-1 \pmod{n}, \\ 0 & \text{en otro caso.} \end{cases}$$

2.5. Caminatas Aleatorias en Gráficas

Definición 2.13 Gráfica

Una gráfica $G = (V, E)$ consiste en un conjunto de vértices V y un conjunto de aristas E , donde los elementos de E son pares de vértices: $E \subset \{\{x, y\} : x, y \in V, x \neq y\}$.

Definición 2.14 Vecino

Para todo $\{x, y\} \in E$ si x y y están unidos por una arista, diremos que x y y son vecinos y lo denotaremos por $x \sim y$.

Definición 2.15 Grado

El **grado** de un vértice x es el número de vecinos de x y lo denotaremos por $\text{grad}(x)$.

Definición 2.16 Caminata Aleatoria en G

Dado una gráfica $G = (V, E)$, definimos la caminata aleatoria simple en G como la cadena de Markov con espacio de estados V y matriz de transición.

$$P(x, y) = \begin{cases} \frac{1}{\text{grad}(x)} & \text{si } y \sim x, \\ 0 & \text{en otro caso.} \end{cases} \quad (2.21)$$

En otras palabras, lo que nos dice la ecuación (2.21), cuando la cadena este en el vértice x , se moverá uniformemente al azar al siguiente vértice y .

Ejemplo 2.6 Caminata Aleatoria en una Gráfica

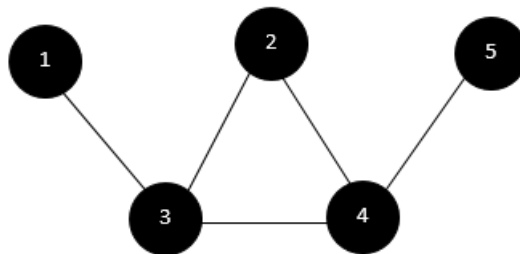


Figura 2.7: Ejemplo de una gráfica con un conjunto de vértices $\{1, 2, 3, 4, 5\}$ y 5 aristas.

Y su matriz de transición estará dada por

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

Ahora recordemos la ecuación 2.15

$$\pi(y) = \sum_{x \in E} \pi(x)P(x,y) \quad \text{para todo } y \in E.$$

Ejemplo 2.7

Consideremos una caminata simple en una gráfica $G = (V, E)$ para algún vértice $y \in E$.

$$\sum_{x \in V} grad(x)P(x,y) = \sum_{x \sim y} grad(x) \frac{1}{grad(x)} = \sum_{x \sim y} \frac{grad(x)}{grad(x)} = grad(y). \quad (2.22)$$

Lema 2.3 Sea $G = (V, E)$ entonces

$$\sum_{y \in V} grad(y) = 2|E| \quad (2.23)$$

Demostración

Sea $G = (V, E)$ una gráfica, definamos d_v como el número de aristas que tienen a v como uno de sus vértices, la suma cuenta el número de aristas de la gráfica, pero las aristas están contadas 2 veces (una vez por cada vértice de la arista), por lo cual la suma es el doble del número de aristas.

En otras palabras, lo que nos dice la ecuación (2.23), los vértices son personas que están en una fiesta y una arista corresponde cuando dos de las personas se estrechan las manos (un apretón de manos), la identidad ayuda a contar el número de apretones de manos.

Ahora por el Lema 2.3, podemos obtener una medida de probabilidad

$$\pi(y) = \frac{grad(y)}{2|E|} \quad \text{para todo } y \in E,$$

la cual es proporcional a los grados y siempre es una distribución estacionaria para la caminata. ■

Retomando el Ejemplo 2.6 y aplicando el Lema 2.3 tenemos

$$\pi = \left(\frac{1}{10}, \frac{2}{10}, \frac{3}{10}, \frac{3}{10}, \frac{1}{10} \right).$$

Definición 2.17 *G* *d*-Regular

Diremos que *G* es ***d*-regular** si tiene la propiedad de que cada vértice tiene el mismo grado *d*.

Lo que implicaría la Definición 2.17, es que si tenemos una gráfica *d*-regular entonces $2|E| = d|V|$ y la distribución uniforme $\pi(y) = \frac{1}{|V|}$ para todo $y \in V$ es estacionaria.

2.6. Tiempos de Llegada y Tiempos de Retorno

Nosotros ya definimos **Tiempo de Llegada** 2.9 y **Tiempo de Retorno** 2.10. Ahora veamos la importancia que tienen estas dos definiciones para nuestro tema a discutir. A partir de este momento consideraremos la siguiente notación

1. P_x es la probabilidad de transición considerando que el estado inicial es x .
2. \mathbb{E}_x es la esperanza considerando que el estado inicial es x .

Lema 2.4 Sea $\tau_x^+ := \min\{t \geq 1 : X_t = x\}$ el primer tiempo de regreso a x . Entonces para cualquier estado x y y de una cadena irreducible,

$$\mathbb{E}_x(\tau_y^+) < \infty. \quad (2.24)$$

Demostración

La definición de cadena irreducible implica que existe un número entero $r > 0$ y un real $\varepsilon > 0$ con la siguiente propiedad: para cualquier estado $z, w \in E$, existe $j \leq r$ con $P^j(z, w) > \varepsilon$. Por lo tanto, para cualquier valor de X_t , la probabilidad de llegar al estado y en un momento entre t y $t + r$ es al menos ε . Por lo tanto, para $k > 0$ tenemos

$$P_x\{\tau_y^+ < kr\} \leq (1 - \varepsilon)P_x\{\tau_y^+ < (k-1)r\}. \quad (2.25)$$

y a su vez

$$P_x\{\tau_y^+ > kr\} \leq (1 - \varepsilon)P_x\{\tau_y^+ > (k-1)r\} \leq (1 - \varepsilon)^2 P_x\{\tau_y^+ > (k-2)r\}.$$

Repitiendo este proceso, tenemos que

$$P_x\{\tau_y^+ > kr\} \leq (1 - \varepsilon)^k. \quad (2.26)$$

Recordatorio 2.3 Si Y es una variable aleatoria de valores enteros no negativos, tenemos que

$$\mathbb{E}(Y) = \sum_{t \geq 0} P\{T > t\}.$$

Observemos que la ecuación (2.26) es una función decreciente de t además

$$\mathbb{E}_x(\tau_y^+) = \sum_{t \geq 0} P_x \{ \tau_y^+ > t \} \leq \sum_{k \geq 0} r P_x \{ \tau_y^+ > kr \} \leq r \sum_{k \geq 0} (1 - \varepsilon)^k < \infty.$$

■

Sea $z \in E$ un estado arbitrario de la cadena de Markov. Examinaremos de cerca el tiempo promedio que la cadena pasa en cada estado entre visitas a z . Para lograr esto desarrollaremos el siguiente tema.

2.7. Existencia de una Distribución Estacionaria

Daremos una construcción explícita de la distribución estacionaria π , que en el caso irreducible da una identidad útil $\pi(x) = \left[\mathbb{E}_x(\tau_y^+) \right]^{-1}$.

$$\begin{aligned} \bar{\pi}(y) &:= \mathbb{E}_z(\text{número de visitas a } y \text{ antes de volver a } z) \\ &= \sum_{t=0}^{\infty} P_z \{ X_t = y, \tau_z^+ > t \} \end{aligned} \quad (2.27)$$

Proposición 2.6 *Sea $\bar{\pi}$ la medida en E definida por 2.27.*

1. Si $P_z \{ \tau_z^+ < \infty \} = 1$, entonces $\bar{\pi}$ satisface $\bar{\pi}P = \bar{\pi}$.
2. Si $\mathbb{E}_z(\tau_z^+) < \infty$, entonces $\pi(x) := \frac{\bar{\pi}(x)}{\mathbb{E}_z(\tau_z^+)}$, $\forall x \in E$ es una distribución estacionaria.

Recordatorio 2.4 *El Lema 2.4 muestra que si P es irreducible, entonces $\mathbb{E}_z(\tau_z^+) < \infty$.*

Demostración

Para algún estado y , tenemos $\bar{\pi}(y) \leq \mathbb{E}_z(\tau_z^+)$. Por el Lema 2.4 asegura que $\bar{\pi}(y) \leq \infty$ para todo $y \in E$. Comprobemos que $\bar{\pi}$ es estacionaria, a partir de la definición:

$$\sum_{x \in E} \bar{\pi}(x) P(x, y) = \sum_{x \in E} \sum_{t=0}^{\infty} P_z \{ X_t = x, \tau_z^+ > t \} P(x, y). \quad (2.28)$$

Porque el evento $\{ \tau_z^+ > t+1 \} = \{ \tau_z^+ > t \}$ está determinado por X_0, \dots, X_t ,

$$P_z \{ X_t = x, X_{t+1} = y, \tau_z^+ > t+1 \} = P_z \{ X_t = x, \tau_z^+ > t+1 \} P(x, y). \quad (2.29)$$

Invirtiendo el orden de suma en (2.28) y usando la identidad (2.29) muestra que

$$\begin{aligned} \sum_{x \in E} \bar{\pi}(x) P(x, y) &= \sum_{t=0}^{\infty} P_z \{ X_{t+1} = y, \tau_z^+ \geq t+1 \} \\ &= \sum_{t=0}^{\infty} P_z \{ X_t = y, \tau_z^+ \geq t \}. \end{aligned} \quad (2.30)$$

La expresión en (2.30) es muy similar a (2.27), por lo que casi hemos terminado. De hecho,

$$\begin{aligned} & \sum_{t=0}^{\infty} P_z \{X_t = y, \tau_z^+ \geq t\} \\ &= \bar{\pi}(y) - P_z \{X_0 = y, \tau_z^+ \geq 0\} + \sum_{t=1}^{\infty} P_z \{X_t = y, \tau_z^+ \geq t\} \\ &= \bar{\pi}(y) - P_z \{X_0 = y\} + P_z \{X_{\tau_z^+} = y\} \end{aligned} \quad (2.31)$$

$$= \bar{\pi}(y) - P_z \{X_0 = y\} + P_z \{X_{\tau_z^+} = y\}. \quad (2.32)$$

La igualdad (2.32) se sigue al considerar dos casos:

$y = z$: Ya que $X_0 = z$ y $X_{\tau_z^+} = z$, los dos últimos términos de (2.31) son ambos 1, y se cancelan mutuamente.

$y \neq z$: Aquí ambos términos de (2.31) son 0.

Por lo tanto, combinando (2.30) con (2.32) muestra que $\bar{\pi} = \bar{\pi}P$. Finalmente, para obtener una medida de probabilidad, normalizamos por $\sum_{x \in E} \bar{\pi}(x) = \mathbb{E}_z(\tau_z^+)$:

$$\pi(x) = \frac{\bar{\pi}(x)}{\mathbb{E}_z(\tau_z^+)} \quad \text{satisface} \quad \pi = \pi P. \quad (2.33)$$

■

2.8. Unicidad de la distribución estacionaria

Definición 2.18 Función Armónica

Nombraremos una función $h : E \rightarrow \mathbb{R}$ como **armónica** en x si

$$h(x) = \sum_{y \in E} P(x, y)h(y). \quad (2.34)$$

Notemos que si una función h es armónica en $D \subset E$, entonces es armónica para todo estado $x \in D$. Si consideramos h como un vector columna, entonces una función que es armónica en todo E satisface la ecuación matricial $Ph = h$.

Lema 2.5 Sea $(X_t)_{t \geq 0}$, una cadena de Markov con matriz de transición P irreducible. Si una función h que es armónica en cada estado de E entonces es constante.

Demostración

Como E es finito, debe existir un estado x_0 tal que $h(x_0) = M$ es maximal. Si para algún estado z tal que $P(x_0, z) > 0$ tenemos $h(z) < M$, entonces

$$h(x_0) = P(x_0, z)h(z) + \sum_{y \neq z} P(x_0, y)h(y) < M, \quad (2.35)$$

lo que sería una contradicción. se deduce que $h(z) = M$ para todos los estados z tal que $P(x_0, z) > 0$.

Para cualquier $y \in E$, irreducible implica que hay una sucesión $x_0, x_1, \dots, x_n = y$ con $P(x_i, x_{i+1}) > 0$. Repetir el argumento anterior nos dice que $h(y) = h(x_{n-1}) = \dots = h(x_0) = M$. Por lo tanto, h es constante. ■

Corolario 2.1 *Sea P una matriz de transición de una cadena de Markov irreducible. Entonces existe una única distribución de probabilidad π que satisface $\pi = \pi P$.*

Demostración

Por la Proposición 2.6 existe al menos una medida de probabilidad π y por el Lema 2.5 implica que el núcleo de $P - I$ tiene dimensión 1, entonces el rango de columna de $P - I$ es $|E| - 1$.

Dado que el rango de fila de cualquier matriz es igual a su rango de columna, la ecuación de fila-vector $v = vP$ también tiene un espacio unidimensional de soluciones. Este espacio contiene solo un vector cuyas entradas suman 1. ■

Proposición 2.7 *Si P es una matriz de transición irreducible y π es la única solución de distribución de probabilidad tal que $\pi = \pi P$, entonces para todo estado z ,*

$$\pi(z) = \frac{1}{\mathbb{E}_z(\tau_z^+)}. \quad (2.36)$$

Demostración

Sea $\tilde{\pi}_z(y)$ igual a $\tilde{\pi}$ como se definió en la ecuación 2.27, y escribimos

$$\pi_z(y) = \frac{\tilde{\pi}_z(y)}{\mathbb{E}_z(\tau_z^+)}.$$

La Proposición 2.6 implica que z es una distribución estacionaria, entonces $\pi_z = \pi$. Por lo tanto,

$$\pi(z) = \pi_z(z) = \frac{\tilde{\pi}_z(z)}{\mathbb{E}_z(\tau_z^+)} = \frac{1}{\mathbb{E}_z(\tau_z^+)}.$$

■

2.9. Reversibilidad y reversiones de tiempo

Definición 2.19 Ecuaciones de Balance Detalladas

Decimos que π satisface la ecuación de balance si

$$\pi(x)P(x, y) = \pi(y)P(y, x) \quad \text{para todo } x, y \in E. \quad (2.37)$$

Proposición 2.8 Sea P la matriz de transición de una cadena de Markov con espacio de estado E . Para cualquier distribución π que satisfaga la ecuación (2.37) es estacionaria para P .

Demostración

Sumando en ambos lados de la ecuación (2.37)

$$\sum_{y \in E} \pi(y)P(y, x) = \sum_{y \in E} \pi(x)P(x, y) = \pi(x) \sum_{y \in E} P(x, y) = \pi(x) \times 1 = \pi(x).$$

■

Definición 2.20 Cadenas de Markov Reversible

Sea $(X_t)_{t \geq 0}$, una cadena de Markov con distribución estacionaria π , si X_t satisface (2.37) entonces la distribución (X_0, X_1, \dots, X_t) es igual a la distribución de $(X_t, X_{t-1}, \dots, X_0)$.

En otras palabras, lo que nos dice la Definición 2.20 se deriva de la ecuación (2.37), es decir, veamos que

$$\pi(x_0)P(x_0, x_1) \cdots P(x_{t-1}, X_t) = \pi(x_t)P(x_t, x_{t-1}) \cdots P(x_1, X_0).$$

lo que podemos reescribir de la siguiente manera

$$P_\pi \{X_0 = x_0, \dots, X_t = x_t\} = P_\pi \{X_0 = x_t, \dots, X_t = x_0\}.$$

Ejemplo 2.8 Caminata Aleatoria Simple en una Gráfica

Consideremos la caminata aleatoria simple en una gráfica G . Nosotros vimos que la distribución estacionaria $\pi(x) = \frac{\text{grad}(x)}{2|E|}$ ahora veamos que es una cadena reversible

$$\pi(x)P(x, y) = \frac{\text{grad}(x)}{2|E|} \frac{1_{\{x \sim y\}}}{\text{grad}(x)} = \frac{1_{\{x \sim y\}}}{2|E|} = \pi(y)P(y, x).$$

Por lo tanto, es una cadena reversible.

Ejemplo 2.9 Considere la caminata aleatoria sesgada en el ciclo n : una partícula se mueve en sentido horario con probabilidad p y se mueve en sentido antihorario con probabilidad $q = 1 - p$. La distribución estacionaria permanece uniforme: $\pi(k) = \frac{1}{n}$, entonces

$$\sum_{j \in Z_n} \pi(j)P(j, k) = \pi(k-1)p + \pi(k+1)q = \frac{1}{n},$$

de donde π es la distribución estacionaria. Sin embargo, si $\pi \neq \frac{1}{2}$, entonces

$$\pi(k)P(k, k+1) = \frac{p}{n} \neq \frac{q}{n} = \pi(k+1)P(k+1, k).$$

Definición 2.21 Inversión del Tiempo

La **Inversión del tiempo** de una cadena de Markov con matriz de transición P y distribución estacionaria π es la cadena con matriz

$$\hat{P}(x, y) := \frac{\pi(y)P(y, x)}{\pi(x)}. \quad (2.38)$$

Ahora por la ecuación estacionaria $\pi = \pi P$ implica que \hat{P} es una matriz estocástica.

Proposición 2.9 Sea $(X_t)_{t \geq 0}$, una cadena de Markov irreducible con matriz de transición P y distribución estacionaria π . Definimos $(\hat{X}_t)_{t \geq 0}$, para la cadena de tiempo invertido con la matriz de transición \hat{P} . Entonces π es estacionaria para \hat{P} , y para cualquier $x_0, \dots, x_t \in E$ tenemos

$$P_\pi \{X_0 = x_0, \dots, X_t = x_t\} = P_\pi \{\hat{X}_0 = x_t, \dots, \hat{X}_t = x_0\}.$$

Demostración

Primero veamos que π es estacionaria para \hat{P}

$$\sum_{y \in E} \pi(y) \hat{P}(y, x) = \sum_{y \in E} \pi(y) \frac{\pi(x)P(x, y)}{\pi(y)} = \sum_{y \in E} \pi(x)P(x, y) = \pi(x).$$

Ahora veamos que las probabilidades de las dos trayectorias son iguales, para lo cual veamos

$$\begin{aligned} P_\pi \{X_0 = x_0, \dots, X_t = x_t\} &= \pi(x_0)P(x_0, x_1)P(x_1, x_2) \cdots P(x_{t-1}, x_t) \\ &= \pi(x_t)\hat{P}(x_t, x_{t-1})\hat{P}(x_{t-1}, x_{t-2}) \cdots \hat{P}(x_1, x_0) \\ &= P_\pi \{\hat{X}_0 = x_t, \dots, \hat{X}_t = x_0\}, \end{aligned}$$

ya que $P(x_{i-1}, x_i) = \frac{\pi(x_i)\hat{P}(x_i, x_{i-1})}{\pi(x_{i-1})}$ para cada i .

■

Algo importante a resaltar es que si una cadena con matriz de transición P es reversible, entonces $\hat{P} = P$.

2.10. Clasificación de los Estados de una Cadena de Markov

Definición 2.22 Estado Accesible Diremos que el estado y es accesible desde el estado x si existe un número $t \geq 0$ tal que $P^t(x, y) > 0$, esto lo denotaremos por $x \rightarrow y$.

Definición 2.23 Estados Comunicados Diremos que los estados x, y están comunicados si se cumple que $x \rightarrow y$ y $y \rightarrow x$ y lo denotaremos por $x \leftrightarrow y$.



Figura 2.8: Del lado izquierdo podemos ver un ejemplo donde el estado y es accesible desde x , mientras que del lado derecho y está comunicado con x .

Definición 2.24 Estados Esencial Un estado $x \in E$ es llamado **esencial** si para todo y se cumple que $x \rightarrow y$ también es cierto que $y \rightarrow x$.

Definición 2.25 Estados Inesencial Diremos que un estado x es **Inesencial** si no es **Esencial**.

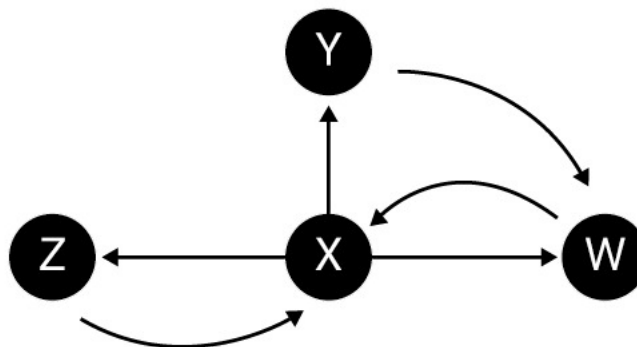


Figura 2.9: x es esencial ya que w, y y z son accesibles desde x .

Definición 2.26 Estado recurrente *Se dice que un estado x es recurrente si la probabilidad de eventualmente regresar a x , partiendo de x , es uno, es decir, si*

$$P(X_t = x \text{ para algún } t \geq 1 | X_0 = x) = 1. \quad (2.39)$$

Definición 2.27 Estado transitorio *Se dice que un estado x es transitorio si la probabilidad de eventualmente regresar a x , partiendo de x , es menor que uno, es decir, si*

$$P(X_t = x \text{ para algún } t \geq 1 | X_0 = x) < 1. \quad (2.40)$$

Teorema 2.5 *La comunicación es una relación de equivalencia, es decir, es*

- *Reflexiva: $x \leftrightarrow x$.*
- *Simétrica: si $x \leftrightarrow y$ entonces $y \leftrightarrow x$.*
- *Transitiva: Si $x \leftrightarrow y$ y $y \leftrightarrow k$ entonces $x \leftrightarrow k$.*

Demostración

La propiedad 1 y 2 son triviales por lo cual solo daremos una justificación a la propiedad 3. Supongamos que existen enteros n, m tales que $P^n(x, k) > 0$ definamos $l = n + m$, y por la ecuación de Chapman Kolmogorov (2.7)

$$P^l(x, y) \geq P^n(x, k)P^m(k, y) > 0.$$

■

En consecuencia, por el Teorema 2.5, la **comunicación** induce una partición del espacio de estados de una cadena de Markov dada por los subconjuntos de estados comunicantes, es decir, dos estados pertenecen al mismo elemento de la partición si, y sólo si, estos se comunican.

Definición 2.28 Clase de Comunicación *A la clase de comunicación de un estado x se le denotará por $[x]$.*

Por lo tanto, $x \leftrightarrow y$ si y sólo si, $[x] = [y]$.

Lema 2.6 *Si x es un estado esencial y $x \leftrightarrow y$, entonces y es esencial.*

Demostración

Si $y \rightarrow z$, entonces $x \rightarrow z$. Por lo tanto, x al ser esencial, $z \rightarrow x$, donde $z \rightarrow y$.

Del Lema anterior deducimos que los estados en una sola clase comunicante son todos esenciales o todos no son esenciales. Por lo cual podemos clasificar las clases comunicantes como esenciales o inesenciales.

■

Definición 2.29 Estado absorbente *Para un estado x de una cadena de Markov se llama absorbente si $P^1(x, x) = 1$. O Si $[x] = \{x\}$ y x es esencial; la cadena nunca saldrá del estado x una vez que visite el estado x .*

Por otro lado, si $[x] = \{x\}$ y x es inessential, entonces una vez que la cadena deje x , nunca regresará.

Lema 2.7 *Toda cadena con espacio de estados finito tiene al menos una clase esencial.*

Demostración

Definamos inductivamente la siguiente sucesión $(y_0, y_1, \dots) \in E$. Fijemos arbitrariamente un estado inicial y_0 . Para $k \geq 1$, dada la sucesión (y_0, \dots, y_{k-1}) , si y_{k-1} es esencial entonces se cumple el Lema 2.7. Por otro lado, si y_k es tal que $y_{k-1} \rightarrow y_k$ pero $y_k \not\rightarrow y_{k-1}$. Pero no podría haber estados repetidos en la sucesión, porque si $j < k$ y $y_k \rightarrow y_j$, entonces $y_k \rightarrow y_{k-1}$, lo que es una contradicción. Por lo cual, si seguimos repitiendo el proceso, de buscar otro estado y_k y como la sucesión no repite elementos, en algún momento debe de terminar en algún estado esencial.

■

Definición 2.30 Matriz de Transición Restringida a una Clase Esencial *Sea $P_C = P_{C \times C}$ la restricción de la matriz P al conjunto de estados $C \subset E$.*

Por la Definición 2.30, si $C = [x]$ es una clase esencial, entonces P_C es una matriz estocástica, es decir, $\sum_{y \in [x]} P(x, y) = 1$, esto es porque $P(x, z) = 0$ para estado $z \notin [x]$. Además P_C es irreducible por definición de clases de comunicación 2.28.

Corolario 2.2 *Si $P_z \{ \tau_z^+ < \infty \} = 1$ entonces $\mathbb{E}_z \{ \tau_z^+ \} < \infty$, para E finito.*

Demostración

Sea z un estado tal que $P_z \{ \tau_z^+ < \infty \} = 1$, entonces z es un estado esencial. Es decir, $C = [z]$, por lo cual P_C es irreducible y además aplicando el Lema 2.4 a P_C muestra que $\mathbb{E}_z \{ \tau_z^+ \} < \infty$.

■

Proposición 2.10 *Si π es la medida estacionaria para la matriz de transición finita P , entonces $\pi(y_0) = 0$ para todo estado inessential y_0 .*

Demostración

Sea C una clase de comunicación esencial. Entonces

$$\pi(C)P(C) = \sum_{z \in C} (\pi P)(z) = \sum_{z \in C} \left[\sum_{y \in C} \pi(y)P(x, z) + \sum_{y \notin C} \pi(y)P(x, z) \right]$$

Intercambiando el orden de la suma, tenemos

$$\pi(C)P(C) = \sum_{y \in C} \pi(y) \sum_{z \in C} P(y, z) + \sum_{z \in C} \sum_{y \notin C} \pi(y)P(y, z).$$

Al ser P una matriz estocástica implica que $\sum_{z \in C} P(y, z) = 1$, para $y \in C$ entonces

$$\pi(C)P(C) = \pi(C) + \sum_{z \in C} \sum_{y \notin C} \pi(y)P(y, z). \quad (2.41)$$

Notemos que de la ecuación (2.41), π es invariante entonces $\pi P(C) = \pi(C)$. Por lo cual $\sum_{z \in C} \sum_{y \notin C} \pi(y)P(y, z) = 0$ para todo $y \notin C$ y $z \in C$. Para esto supongamos que y_0 es inesencial. Ahora por Lema 2.7, tenemos que hay una sucesión de estados y_0, y_1, \dots, y_r que satisface $P(y_{i-1}, y_i) > 0$, donde los estados y_0, y_1, \dots, y_{r-1} son inesenciales, y $y_r \in D$, donde D es una clase de comunicación esencial. Como $P(y_{r-1}, y_r) > 0$ por lo que acabamos de probar que $\pi(y_{r-1})P(y_{r-1}, y_r) = 0$ esto se debe a que $\pi(y_{r-1}) = 0$. Pero si $\pi(y_k) = 0$, entonces

$$\pi(y_k) = 0 = \sum_{y \in E} \pi(y)P(y, y_k).$$

Esto implicaría que $\pi(y)P(y, y_k) = 0$ para todo y . En particular, para $\pi(y_{k-1}) = 0$. Por lo tanto, si seguimos con este proceso inductivo hacia atrás, recorriendo toda la sucesión, encontramos que $\pi(y_0) = 0$.

■

Ahora nuestro propósito será desarrollar cadenas de Markov que serán muy útiles para el tema central de este trabajo.

Capítulo 3

Cadenas Importantes; Urna de Ehrenfest y El Hipercubo

3.1. Ruina Del Jugador

Ejemplo 3.1 Ruina del Jugador *Considere un jugador que comienza con una fortuna inicial de \$1, posteriormente, en casa apuesta, gana \$1 o pierde \$1 independiente del pasado con probabilidad p y $q = 1 - p$ respectivamente. Sea $(X_t)_{t \geq 0}$, la fortuna total después de la t -ésima apuesta. El objetivo del jugador es alcanzar una fortuna total N , sin arruinarse (quedarse sin dinero). Si el jugador tiene éxito, se dice que el jugador gana el juego. En cualquier caso, el jugador deja de jugar después de ganar o arruinarse, lo que ocurra primero. No hay nada especial en comenzar con \$1, más generalmente el jugador comienza con \$ i donde $0 < i < N$.*

Notemos que X_t genera una cadena de Markov en el espacio de estados $E = \{0, 1, \dots, N\}$. Las probabilidades de transición están dadas por $P(i, i + 1) = p$, $P(i, i - 1) = q$, $0 < i < N$. Observemos que los estados 0 y N son estados absorbentes, es decir, $P(0, 0) = 1 = P(N, N)$.

Por ejemplo, para $N = 4$, la matriz de transición está dadas por

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

Mientras el juego avanza forma una caminata aleatoria simple

$$X_t = i + \sum_{k=1}^t \Delta_k, \quad t \geq 1, \quad X_0 = i,$$

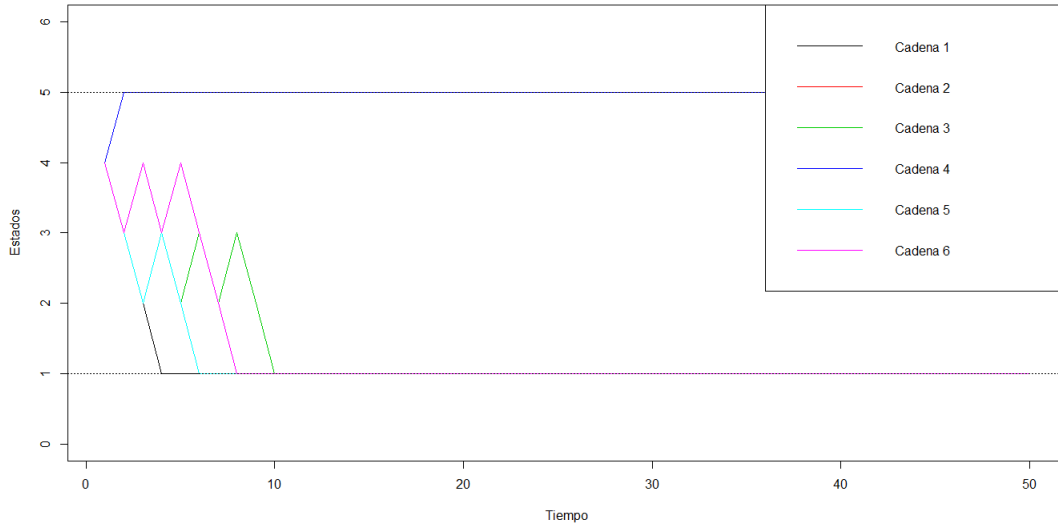


Figura 3.1: Simulación de 6 juegos, con $i=4$ y $N=5$.

donde $\{\Delta_k\}$, es una sucesión de variables aleatorias independientes e idénticamente distribuidas con distribución

$$P(X_{t+1} = y | X_t = x) \begin{cases} p & \text{si } \Delta = 1, \\ q & \text{si } \Delta = -1, \\ 0 & \text{en otro caso.} \end{cases}$$

El juego termina cuando $X_t=0$ o $X_t=N$, definimos el siguiente tiempo de paro.

$$\tau_i = \text{mín} \{t \geq 0 : X_t \in \{0, N\} | X_0 = i\}.$$

Denota el momento en que el juego se detiene, si $X_{\tau_i} = N$ o $X_{\tau_i} = 0$ dado que inicio con $\$i$.

$P_i(N)$ denota la probabilidad de que el jugador comenzando inicialmente con $\$i$, alcance una fortuna total de N , antes de la ruina, $1-P_i(N)$ corresponde a la probabilidad de ruina. Es claro que $P(0, N) = 0$ y $P(N, N) = 1$ por definición y luego podemos calcular $P_i(N)$, $1 \leq i \leq N - 1$.

Recordatorio 3.1 Teorema 3.1 Teorema de Probabilidad Total Para un espacio de probabilidad dado (E, \mathfrak{S}, P) , si B_1, B_2, \dots, B_n , es una colección de elementos mutuamente disjuntos en \mathfrak{S} que satisfacen $E = \bigcup_{j=1}^n B_j$ y $P[B_j] > 0$ para $j = 1, \dots, n$, entonces para todo evento en \mathfrak{S} , $P(A) = \sum_{j=1}^n P(A|B_j) P(B_j)$. Ver [15] pág.(50).

Proposición 3.1

$$P_i(N) = \begin{cases} \frac{1 - \left(\frac{q}{p}\right)^i}{1 - \left(\frac{q}{p}\right)^N}, & \text{si } p \neq q, \\ \frac{i}{N}, & \text{si } p = q = \frac{1}{2}. \end{cases}$$

Demostración

Para esta prueba definimos $P_i = P(i, N)$, es decir, suprimimos la dependencia de N para facilitar la notación. Por el Teorema de Probabilidad Total 3.1, condicionando el resultado de la primera apuesta, $\Delta_1 = 1$ o $\Delta_1 = -1$, en consecuencia, tenemos

$$P_i = pP_{i+1} + qP_{i-1}. \quad (3.1)$$

De la ecuación (3.1), podemos deducir lo siguiente, si $\Delta_1 = 1$, entonces la fortuna total del jugador aumenta a $X_1 = i + 1$ y, por lo tanto, la probabilidad de ganar por la propiedad de Markov, el jugador ahora ganará con probabilidad P_{i+1} . Del mismo modo, si $\Delta_1 = -1$, entonces la fortuna del jugador disminuye a $X_1 = i - 1$ y, por lo tanto, por la propiedad de Markov, el jugador ahora ganará con probabilidad P_{i-1} . Dado que las probabilidades correspondientes a los dos resultados son p y q , dando (3.1). Y dado que $p + q = 1$, la ecuación (3.1) la podemos reescribir como

$$P_i = pP_{i+1} + qP_{i-1} = pP_i + qP_i = pP_{i+1} + qP_{i-1},$$

o bien

$$P_{i+1} - P_i = \frac{q}{p}(P_i - P_{i-1}), \quad i = \{1, 2, \dots, N-1\}.$$

Evaluando

$$P_2 - P_1 = \frac{q}{p}(P_1 - P_0) = \frac{q}{p}P_1$$

$$P_3 - P_2 = \frac{q}{p}(P_2 - P_1) = \left(\frac{q}{p}\right)^2 P_1$$

⋮

$$P_N - P_{N-1} = \frac{q}{p}(P_{N-1} - P_{N-2}) = \left(\frac{q}{p}\right)^{N-1} P_1.$$

Sumando las primeras $i - 1$ ecuaciones tenemos

$$P_i - P_1 = P_1 \left[\frac{p}{q} + \left(\frac{q}{p}\right)^2 + \dots + \left(\frac{q}{p}\right)^{i-1} \right].$$

Lo que implica

$$P_i = P_1 \left[1 + \frac{p}{q} + \left(\frac{q}{p}\right)^2 + \dots + \left(\frac{q}{p}\right)^{i-1} \right]. \quad (3.2)$$

Notemos que la ecuación (3.2), es una serie geométrica $\left(\sum_{n=0}^i a^n = \frac{1-a^{i+1}}{1-a}\right)$ ahora notemos que tenemos 2 casos,

- $\frac{q}{p} = 1$, es decir $p = \frac{1}{2}$.
- $\frac{q}{p} \neq 1$, es decir $p \neq \frac{1}{2}$.

Es decir

$$P_{i+1} = \begin{cases} (i+1)P_1 & \text{si } \frac{q}{p} = 1, \\ \frac{1-\left(\frac{q}{p}\right)^{i+1}}{1-\frac{q}{p}} P_1 & \frac{q}{p} \neq 1. \end{cases} \quad (3.3)$$

Para $i = N-1$ y usando $A_N = 1$

$$1 = P_N = \begin{cases} NP_1 & \text{si } \frac{q}{p} = 1, \\ \frac{1-\left(\frac{q}{p}\right)^N}{1-\frac{q}{p}} P_1 & \frac{q}{p} \neq 1. \end{cases} \quad (3.4)$$

De donde se sigue

$$P_1 = \begin{cases} \frac{1}{N} & \text{si } p = \frac{1}{2}, \\ \frac{1-\frac{q}{p}}{1-\left(\frac{q}{p}\right)^N} & p \neq \frac{1}{2}. \end{cases}$$

de las ecuaciones (3.3) y (3.4). Por lo tanto

$$P_i = \begin{cases} \frac{i}{N} & \text{si } p = \frac{1}{2}, \\ \frac{1-\left(\frac{q}{p}\right)^i}{1-\left(\frac{q}{p}\right)^N} & p \neq \frac{1}{2}. \end{cases}$$

■

Proposición 3.2 *El número esperado de apuestas antes de la ruina es*

$$\mathbb{E}_i(\tau) = \begin{cases} i(N-i) & \text{si } p = \frac{1}{2}, \\ \frac{1}{q-p} \left(i - N \frac{1-\left(\frac{q}{p}\right)^i}{1-\left(\frac{q}{p}\right)^N} \right) & p \neq \frac{1}{2}. \end{cases}$$

Recordatorio 3.2 Ley de Probabilidad Total para Esperanzas

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} E(X|Y = k) P(Y = k). \quad (3.5)$$

Demostración

Se sigue de la demostración de la Proposición 3.1. Sin pérdida de generalidad denotemos por B_1 y B_2 los eventos en los cuales el jugador A gana el primer juego y el jugador B pierde el primer juego. Sea τ el tiempo requerido para que la cadena sea absorbida en el estado 0 o N . Por el teorema de la Probabilidad Total para Esperanzas (3.5)

$$\mathbb{E}[A_i] = \mathbb{E}[A_i|B_1] P(B_1) + \mathbb{E}[A_i|B_2] P(B_2). \quad (3.6)$$

Si gana el primer juego, tiene $i + 1$ unidades, por lo que la distribución de P_i dado B_1 tiene la misma distribución que $1 + P_{i+1}$, donde P_{i+1} mide la duración del juego a partir de $i + 1$ unidades. Por lo tanto $\mathbb{E}[P_i|B_1] = 1 + \mathbb{E}(P_{i+1})$. Análogo para $\mathbb{E}[P_i|B_2] = 1 + \mathbb{E}(P_{i-1})$. Entonces

$$\mathbb{E}(P_i) = p(1 + \mathbb{E}(P_{i+1})) + q(1 + \mathbb{E}(P_{i-1})). \quad (3.7)$$

Dado que $p + q = 1$ tenemos

$$\mathbb{E}(P_i) = 1 + p\mathbb{E}(P_{i+1}) + q\mathbb{E}(P_{i-1}). \quad (3.8)$$

Para facilitar el entendimiento de la demostración, reescribiremos la ecuación (3.8), como

$$\mathbb{E}_{i+1} = 1 + p\mathbb{E}_{i+1} + q\mathbb{E}_{i-1}. \quad (3.9)$$

Al igual que la demostración pasada, sustituimos $(p + q)\mathbb{E}_i$ por \mathbb{E}_i y agrupamos términos

$$\mathbb{E}_{i+1} - \mathbb{E}_i = \frac{q}{p} (\mathbb{E}_{i+1} + \mathbb{E}_{i-1}) - \frac{1}{p}. \quad (3.10)$$

Iterando las primeras $i - 1$ ecuaciones

$$\begin{aligned} \mathbb{E}_2 - \mathbb{E}_1 &= \frac{q}{p} (\mathbb{E}_1 + \mathbb{E}_0) - \frac{1}{p} = \frac{q}{p} (\mathbb{E}_1) - \frac{1}{p}, \\ \mathbb{E}_2 - \mathbb{E}_1 &= \frac{q}{p} (\mathbb{E}_1 + \mathbb{E}_0) - \frac{1}{p} = \frac{q}{p} \left(\frac{q}{p} (\mathbb{E}_1) - \frac{1}{p} \right) - \frac{1}{p} = \left(\frac{q}{p} \right)^2 \mathbb{E}_1 - \frac{1}{p} \left[1 + \frac{q}{p} \right], \end{aligned}$$

⋮

$$\mathbb{E}_i - \mathbb{E}_{i-1} = \frac{q}{p} (\mathbb{E}_{i-1} + \mathbb{E}_{i-2}) - \frac{1}{p} = \frac{q}{p} \left(\left(\frac{q}{p} \right)^{i-2} (\mathbb{E}_1) - \frac{1}{p} \right) - \frac{1}{p} = \left(\frac{q}{p} \right)^{i-1} \mathbb{E}_1 - \frac{1}{p} \left[1 + \frac{q}{p} + \dots + \left(\frac{q}{p} \right)^{i-2} \right].$$

Sea

$$S_i = 1 + \left(\frac{q}{p} \right) + \left(\frac{q}{p} \right)^2 + \dots + \left(\frac{q}{p} \right)^i$$

entonces

$$\mathbb{E}_i - \mathbb{E}_{i-1} = \left(\frac{q}{p} \right)^{i-1} \mathbb{E}_1 - \frac{1}{p} S_{i-2}.$$

Sumando todas las ecuaciones, se sigue que

$$\mathbb{E}_i - \mathbb{E}_1 = \mathbb{E}_1 (S_{i-1} - 1) - \frac{1}{p} \sum_{k=0}^{i-2} S_k. \quad (3.11)$$

Es decir,

$$\mathbb{E}_i = \mathbb{E}_1 (S_{i-1}) - \frac{1}{p} \sum_{k=0}^{i-2} S_k. \quad (3.12)$$

Para $i = N$

$$\mathbb{E}_N = \mathbb{E}_1 (S_{N-1}) - \frac{1}{p} \sum_{k=0}^{i-2} S_k.$$

y usando que $\mathbb{E}_0 = 0$

$$\mathbb{E}_1 = \frac{1}{S_{N-1}} \frac{1}{p} \sum_{k=0}^{i-2} S_k.$$

Sustituyendo en la ecuación (3.12)

$$\mathbb{E}_i = \frac{S_{i-1}}{S_{N-1}} \frac{1}{p} \sum_{k=0}^{N-2} S_k - \frac{1}{p} \sum_{k=0}^{i-2} S_k. \quad (3.13)$$

Al igual que en la demostración anterior, tenemos dos casos,

$$S_i = 1 + \left(\frac{q}{p}\right) + \cdots + \left(\frac{q}{p}\right)^i = \begin{cases} i+1 & \text{si } p = \frac{1}{2}, \\ \frac{1 - \left(\frac{q}{p}\right)^{i+1}}{1 - \left(\frac{q}{p}\right)} & \text{si } p \neq \frac{1}{2}. \end{cases}$$

Sustituyendo en la ecuación (3.13), llegamos a que si $p = 1/2$, haciendo uso de la suma de los n primeros números

$$\mathbb{E}_i = \frac{k}{N} \frac{1}{p} \left[\frac{(N-1)(N)}{2} \right] - \frac{1}{p} \left[\frac{(k-1)(k)}{2} \right].$$

dado que $p = 1/2$

$$= k(N-1) - (k-1)(k) = kN - k - k^2 + k = kN - k^2 = k(N - K).$$

Para el caso en que $p \neq \frac{1}{2}$. Lo omitiremos porque no es del interés de este trabajo. ■

3.2. Coleccionista de Cupones

Ejemplo 3.2 Coleccionista de Cupones

Una compañía emite diferentes tipos de cupones. Un coleccionista desea tener la colección completa. Supongamos que hay N cupones diferentes y que son igualmente probables, la probabilidad de comprar cualquier tipo en cualquier momento es $\frac{1}{N}$. Cuando el coleccionista tiene k cupones de diferentes tipos, entonces solo le faltan $N - k$ tipos.

¿ Cuántos cupones debe obtener para que su colección contenga todos los tipos N ?

Sea $(X_t)_{t \geq 0}$, el número de diferentes tipos de cupones coleccionados después de t unidades de tiempo y asumiremos nuevamente que la probabilidad de encontrar un cupón de cualquier tipo en cualquier momento es $p = \frac{1}{N}$. Es claro que $(X_t)_{t \geq 0}$, es una cadena de Markov en el espacio de estados $E = \{0, 1, \dots, N\}$ con función de transición dada por,

$$P\{X_{t+1} = k+1 | X_t = k\} = \frac{N-k}{N}$$

y

$$P\{X_{t+1} = k | X_t = k\} = \frac{k}{N}.$$

Y su matriz de transición es

$$P = \begin{pmatrix} 0 & 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \frac{1}{N} & \frac{N-1}{N} & 0 & \cdots & \cdots & 0 \\ 0 & 0 & \frac{2}{N} & \frac{N-2}{N} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & \frac{N-1}{N} & \frac{1}{N} \\ 0 & \cdots & \cdots & \cdots & 0 & 0 & 1 \end{pmatrix}.$$

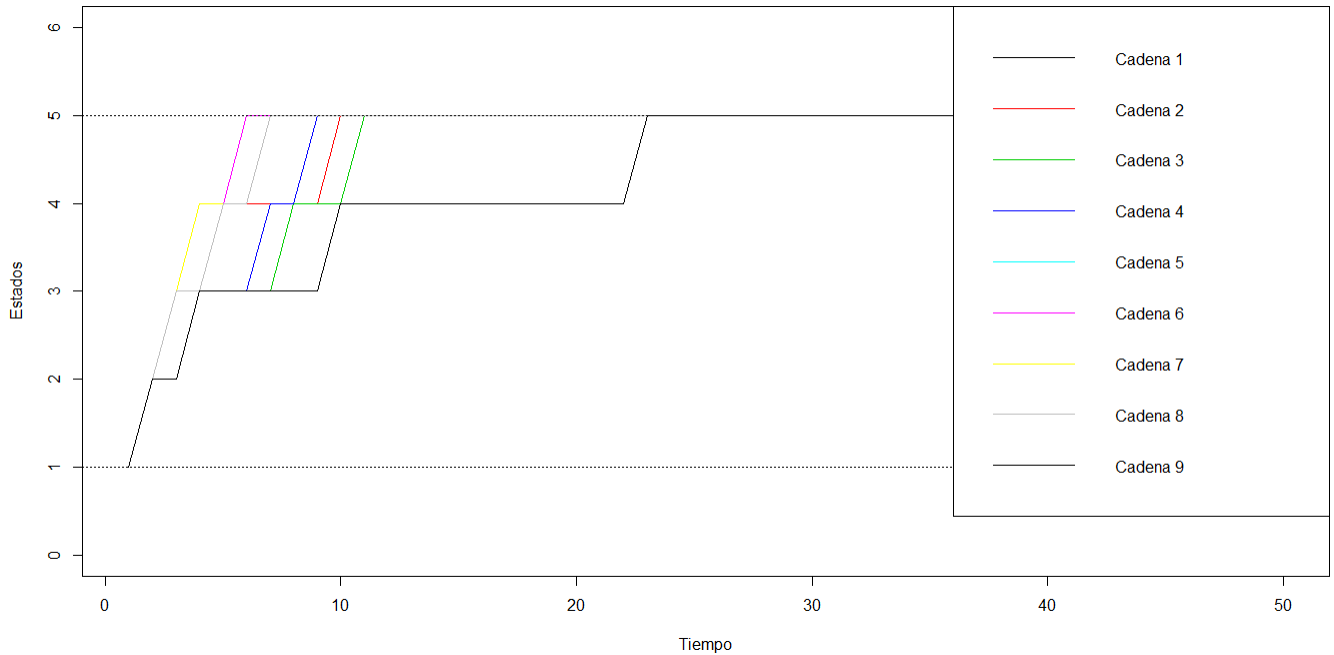


Figura 3.2: 9 posibles escenarios para un coleccionista de cupones, que busca coleccionar $N=5$ cupones.

Notemos que las trayectorias de esta cadena son no decrecientes, Figura 3.2, cuando la cadena llega al estado N (correspondiente a tener una colección completa, se observe en ese estado, por lo tanto, el estado N es un estado absorbente. Por lo cual una pregunta importante sería, ¿ Cuántos pasos son necesarios para llegar al estado N ?

Proposición 3.3 *Considere un coleccionista que intenta recolectar un conjunto completo de cupones. Suponga que cada nuevo cupón se elige de manera uniforme e independiente del conjunto de N tipos posibles, y sea τ el número (aleatorio) de cupones recolectados cuando el primer conjunto contiene cada tipo. Entonces*

$$\mathbb{E}(\tau) = N \sum_{k=1}^N \frac{1}{k}.$$

Demostración

X denota el número (aleatorio) de cupones que necesitamos comprar para completar nuestra colección, podemos escribir $X = X_1 + X_2 + \cdots + X_N$ donde $i = \{1, 2, \dots, N\}$, X_i

denota la cantidad adicional de cupones que necesitamos comprar para pasar de $i - 1$ a i diferentes tipos de cupones en nuestra colección. Es claro que $X_1 = 1$, y dado que estamos considerando el caso de una distribución uniforme, se deduce que cuando se han recolectado diferentes tipos de cupones, un nuevo cupón comprado será de un tipo distinto con probabilidad igual a $\frac{N-i}{N}$. Por el supuesto de independencia, obtenemos que la variable aleatoria X_i para $i \in \{2, 3, \dots, N\}$ es independiente de la otra variable y tiene una ley geométrica con parámetro $\frac{N-i+1}{N}$. Por lo tanto el número esperado de cupones que tenemos que comprar para completar la colección será

$$\mathbb{E}[X] = \mathbb{E}[X_1] + \dots + \mathbb{E}[X_N].$$

Recordemos que la esperanza de una variable aleatorio geométrica está dada por $\mathbb{E}(x) = \frac{1}{p}$, donde $p = \frac{N-i+1}{N}$. Entonces

$$\begin{aligned} &= 1 + \frac{1}{\frac{N-2+1}{N}} + \frac{1}{\frac{N-3+1}{N}} + \dots + \frac{1}{\frac{N-(N-1)+1}{N}} + \frac{1}{\frac{N-N+1}{N}} = 1 + \frac{N}{N-1} + \frac{N}{N-2} + \dots + \frac{N}{2} + N \\ &= N \sum_{i=1}^N \frac{1}{i}. \end{aligned}$$

■

Lema 3.1 Para $n \in \mathbb{N}$,

$$\log(n+1) \leq \sum_{k=1}^n k^{-1} \leq 1 + \log(n). \quad (3.14)$$

Demostración

Para la demostración de este lema ocuparemos la siguiente ecuación

$$\frac{1}{(k+1)} \leq \int_k^{k+1} \frac{dx}{x} \leq \frac{1}{k} \quad (3.15)$$

Por la ecuación (3.15)

$$\frac{1}{2} \leq \int_1^2 \frac{dx}{x} \leq 1,$$

$$\frac{1}{3} \leq \int_2^3 \frac{dx}{x} \leq \frac{1}{2},$$

⋮

$$\frac{1}{n+1} \leq \int_n^{n+1} \frac{dx}{x} \leq \frac{1}{n-1},$$

Sumando

$$\frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots + \frac{1}{n+1} \leq \int_1^{n+1} \frac{dx}{x} \leq 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots + \frac{1}{n}.$$

Finalmente

$$1 + \int_1^n \frac{dx}{x} \geq 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots + \frac{1}{n} \geq \int_1^{n+1} \frac{dx}{x}.$$

Es decir,

$$1 + \log(n) \geq \sum_{i=1}^n \frac{1}{n} \geq \log(n+1).$$

Por el Lema 3.1 y por la Proposición 3.3, tenemos

$$\left| \sum_{k=1}^n \frac{1}{k} - \log(n) \right| \leq 1 \quad \Leftrightarrow \quad \left| n \sum_{k=1}^n \frac{1}{k} - n \log(n) \right| \leq n.$$

■

Corolario 3.1

$$\lim_{N \rightarrow \infty} \frac{\mathbb{E}[N]}{N \log N} \rightarrow 1.$$

Lo que motiva la siguiente proposición.

Proposición 3.4 *Sea τ una variable aleatoria de un coleccionista de cupones como se definió en la Proposición 3.3. Para cualquier $c > 0$*

$$P\{\tau > \lceil n \log(n) + cn \rceil\} \leq e^{-c}. \quad (3.16)$$

Nota Usaremos \log para denotar el logaritmo natural. *Demostración*

Sea A_i el evento en que el i -ésimo tipo no aparezca entre los primeros $\lceil n \log(n) + cn \rceil$ cupones sorteados. Primero observemos

$$P\{\tau > \lceil n \log(n) + cn \rceil\} = P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i).$$

Recordemos que nuestros eventos tienen una distribución geométrica con parámetro $P = \frac{n-1}{n}$, por lo que tenemos

$$\sum_{i=1}^n \left(1 - \frac{1}{n}\right)^{\lceil n \log(n) + cn \rceil} = n * e^{\log(1 - \frac{1}{n}) \lceil n \log(n) + cn \rceil}.$$

Usando la expansión de $\log\left(1 - \frac{1}{n}\right)$

$$n * e^{\lceil n \log(n) + cn \rceil * \log\left(1 - \frac{1}{n}\right)} \approx n * e^{\lceil n \log(n) + cn \rceil * \left(-\frac{1}{n}\right)}$$

⇒

$$n * e^{\lceil n \log(n) + cn \rceil * \left(-\frac{1}{n}\right)} = n * e^{\left(-\frac{n \log(n) + cn}{n}\right)} = n * e^{\left(\frac{-n \log(n)}{n}\right)} e^{\left(\frac{-cn}{n}\right)} = e^{-c}.$$

■

3.3. El hipercubo y el Modelo de la Urna de Ehrenfest

Ejemplo 3.3 Caminata Aleatoria en el Hipercubo

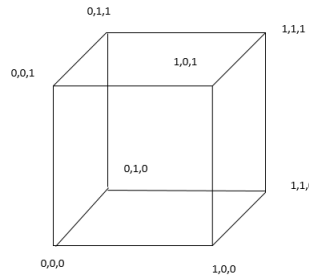


Figura 3.3: Hipercubo en tres dimensiones.

El **Hipercubo n-dimensional** es una gráfica cuyos vértices son **n-tuplas** binarias $\{0, 1\}^n$. Dos vértices están conectados por un un borde cuando difieren exactamente en una coordenada (Figura 3.3). La caminata aleatoria simple en el hipercubo se mueve desde un vértice (x^1, x^2, \dots, x^n) eligiendo una coordenada $j \in \{1, 2, \dots, n\}$ de manera uniforme al azar y establecer el nuevo estado igual a $(x^1, x^2, \dots, x^{j-1}, 1 - x^j, x^{j+1}, \dots, x^n)$. Es decir, el bit en la coordenada elegida de la caminata se voltea. Es un caso especial de una caminata aleatoria en una gráfica.

Notemos que la caminata aleatoria en el hipercubo es periódica, ya que cada movimiento invierte la paridad del número en 1, figura 3.3.

Pero recordemos que definimos la caminata aleatoria perezosa, la cual es aperiódica ya que permanece en su posición con probabilidad $\frac{1}{2}$ y se mueve con probabilidad $\frac{1}{2}$. Esta cadena se puede realizar eligiendo una coordenada uniforme al azar y actualizando el bit en esta reemplazándola con un bit aleatorio imparcial e independiente del tiempo, el estado actual y la coordenada elegida. Como ya dijimos la caminata aleatoria en el hipercubo es una caminata en una gráfica y dado que el hipercubo es una gráfica n-regular, por el Ejemplo 2.7 implica que la distribución estacionaria de las caminatas aleatorias simples y perezosas es uniforme en $\{0, 1\}^n$.

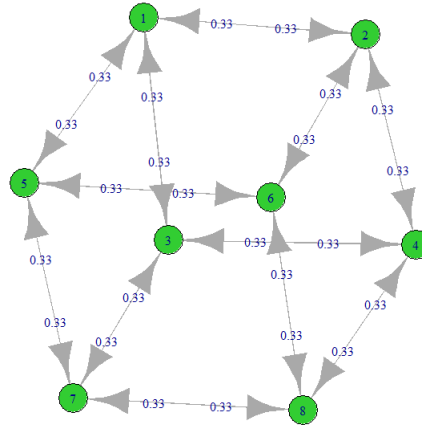


Figura 3.4: Diagrama de transición de un hipercubo en tres dimensiones.

Ahora consideremos el proceso de la urna de Ehrenfest 2.2. Supongamos que, en cada movimiento, una bola se selecciona uniformemente al azar y se transfiere desde su urna actual a la otra urna. Si X_t es el número de bolas en la urna A en el tiempo t , entonces la matriz de transición para X_t es

$$P(j, k) = \begin{cases} \frac{n-j}{n} & \text{si } k = j + 1, \\ \frac{j}{n} & \text{si } k = j - 1, \\ 0 & \text{en otro caso.} \end{cases} \quad (3.17)$$

Definición 3.1 Peso de Hamming

Se define el peso de Hamming $W(x)$ de un vector $x := (x^1, \dots, x^n) \in \{0, 1\}^n$ como su número de coordenadas con valor 1:

$$W(x) = \sum_{j=1}^n x^j. \quad (3.18)$$

Consideremos $(X_t)_{t \geq 0}$, una caminata aleatoria simple en el hipercubo n -dimensional, y sea $W_t = W(X_t)$ el peso de Hamming de la posición de la caminata en el tiempo t . Es claro ver que $(W_t)_{t \geq 0}$, es una cadena de Markov con probabilidades de transición dadas por (3.17).

Para este momento hemos desarrollado el tema de cadena de Markov, en el cual definimos y demostramos las principales propiedades que ocuparemos más adelante, además de mostrar ciertos ejemplos que son cruciales en el desarrollo del tema principal de esta tesis.

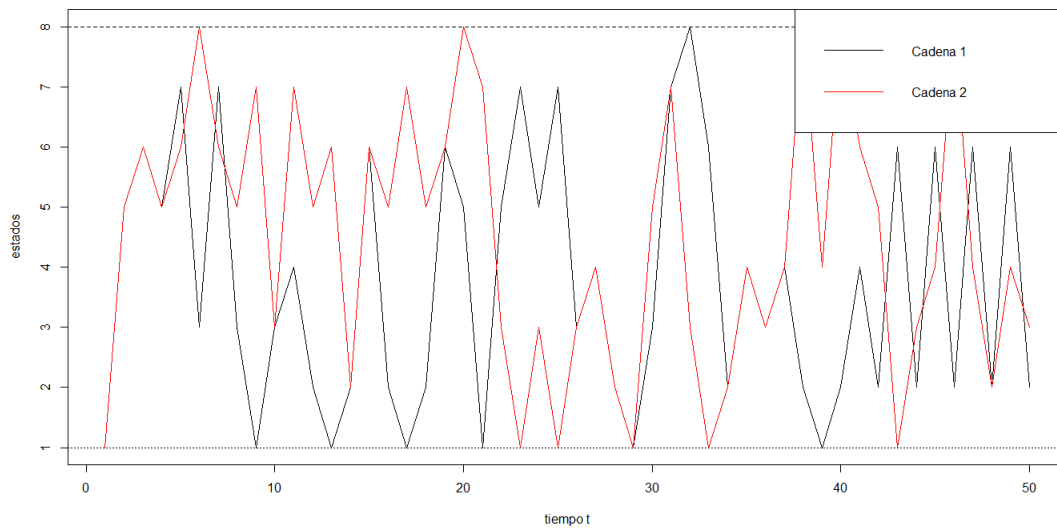


Figura 3.5: 2 posibles trayectorias de la caminata aleatoria en el cubo, iniciando en el estado (1, 1, 1), con probabilidades de transición uniforme.

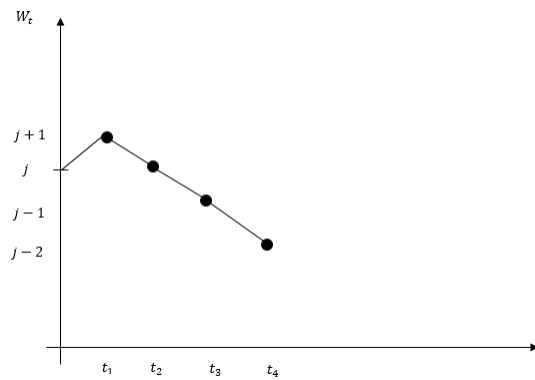


Figura 3.6: Ejemplo: Peso de Hamming, suponiendo que $W_{t_0} = j$, el peso puede incrementar en 1 unidad, cuando la caminata aleatoria en el hipercubo n -dimensional haya modificado una de las $n - j$ coordenadas con valor 0, por otro lado, si modifica alguna de las j coordenadas con valor 1, decrecerá el peso en 1 unidad.

Capítulo 4

Tiempo de Mezcla de Cadenas de Markov

Ahora discutiremos uno de los principales temas de este trabajo, el cual es analizar el comportamiento a largo plazo de las cadenas finitas de Markov, cuantificando la velocidad de convergencia de familias de las cadenas de Markov, para lo cual definiremos una métrica apropiada para medir la distancia entre las distribuciones.

4.1. Distancia de Variación Total

Definición 4.1 La *distancia de variación total* entre dos distribuciones μ y ν en un espacio medible (Ω, \mathfrak{S}) se define por:

$$\|\mu - \nu\|_{TV} = \max_{A \in \mathfrak{S}} |\mu(A) - \nu(A)|. \quad (4.1)$$

En otras palabras, la Definición 4.1, es la distancia entre μ y ν es la diferencia máxima entre las probabilidades asignadas a un evento A .

Ejemplo

Retomando la cadena de dos estados, Ejemplo 2.1, la matriz de transición es

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

y su distribución estacionaria es $\pi = \left(\frac{q}{p+q}, \frac{p}{p+q}\right)$.

Supongamos que la cadena inicia en $\mu_0 = (1, 0)$, y definimos

$$\Delta_t = \mu_t(x) - \pi(x).$$

Dado que $|E| = 2$, solo hay cuatro eventos posibles $A \subseteq E$. Recordemos que $\mu_t := (P\{X_t = x|X_0 = x\}, P\{X_t = y|X_0 = x\}) = (P^t(x, x), P^t(x, y))$, entonces podemos ver que

$$\|\mu_t - \pi\|_{TV} = |\Delta_t| = |P^t(x, x) - \pi(x)| = |\pi(y) - P^t(x, y)|. \quad (4.2)$$

Proposición 4.1 Sean μ_n y μ medidas, si $\mu_n \rightarrow \mu$ en el sentido de variación total, es decir, $\|\mu_n - \mu\|_{TV} \rightarrow 0$, entonces, $\mu_n \Rightarrow \mu$, es decir, convergencia débil.

Demostración

Sea μ una medida continua definida en $B \subseteq A \subseteq E$ (de hecho, para cualquier conjunto A)

$$\begin{aligned} |\mu_n(B) - \mu(B)| &\leq \max_{A \subseteq E} |\mu_n(A) - \mu(A)| \\ &= \|\mu_n - \mu\|_{TV} \rightarrow 0. \end{aligned}$$

■

La ecuación (4.1) no siempre es la forma más conveniente de estimar la distancia, debido a que esta definición considera el máximo en todos los subconjuntos de E . Por lo cual daremos algunas reformulaciones útiles de esta distancia.

Proposición 4.2 Sea μ y ν dos distribuciones de probabilidad en E . Entonces

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sum_{x \in E} |\mu(x) - \nu(x)|, \quad (4.3)$$

y

$$\|\mu - \nu\|_{TV} = \sum_{x \in E, \mu(x) \geq \nu(x)} [\mu(x) - \nu(x)]. \quad (4.4)$$

Demostración

Definimos $B = \{x : \mu(x) - \nu(x) \geq 0\}$ y sea $A \subseteq E$ cualquier evento. Entonces

$$\begin{aligned} \mu(A) - \nu(A) &= (\mu(A \cap B) - \nu(A \cap B)) + (\mu(A \cap B^c) - \nu(A \cap B^c)) \leq \\ &\leq \mu(A \cap B) - \nu(A \cap B) \\ &= (\mu(B) - \nu(B)) - (\mu(B \cap A^c) - \nu(B \cap A^c)) \\ &\leq \mu(B) - \nu(B). \end{aligned}$$

Análogamente podemos ver que $\nu(A) - \mu(A) \leq \nu(B^c) - \mu(B^c)$. Por lo tanto, sumando las dos desigualdades tenemos

$$2|\mu(A) - \nu(A)| \leq (\mu(B) - \nu(B) + \nu(B^c) - \mu(B^c))$$

si y solo si

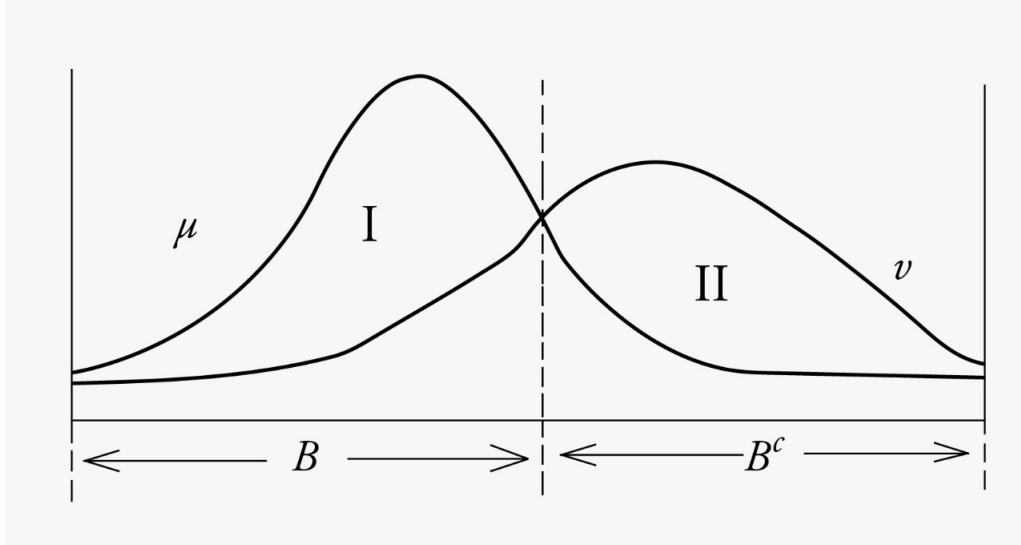


Figura 4.1: Si definimos $B = \{x : \mu(x) \geq \nu(x)\}$, notamos que la región I tiene área $\mu(B) - \nu(B)$. La región II tiene área $\mu(B^c) - \nu(B^c)$, ahora dado que μ y ν son distribuciones de probabilidad el área que contienen debe ser 1 por lo cual la región I y II deben tener la misma área y esa área es $\|\mu - \nu\|_{TV}$. Ver [13].

$$\begin{aligned} |\mu(A) - \nu(A)| &\leq \frac{1}{2} (\mu(B) - \nu(B) + \nu(B^c) - \mu(B^c)) \\ &= \frac{1}{2} \sum_{x \in B} |\mu(x) - \nu(x)| + \frac{1}{2} \sum_{x \in B^c} |\mu(x) - \nu(x)| \\ &= \frac{1}{2} \sum_{x \in B} |\mu(x) - \nu(x)|. \end{aligned}$$

Ahora para finalizar, observemos que las desigualdades anteriores se igualan cuando $A = B$, por lo que este conjunto es el conjunto que alcanza el máximo en la Definición de 4.1. ■

Por la Proposición 4.2 y la desigualdad para números reales, es fácil ver que la distancia de variación total satisface la desigualdad del triángulo: para distribuciones de probabilidad μ , ν y η

$$\|\mu - \nu\|_{TV} \leq \|\mu - \eta\|_{TV} + \|\eta - \nu\|_{TV}. \quad (4.5)$$

Proposición 4.3 *Sea μ y ν dos distribuciones de probabilidad en E . Entonces la distancia de variación total satisface*

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sup \{ \mathbb{E}_\mu(f) - \mathbb{E}_\nu(f) : |f| \leq 1 \}. \quad (4.6)$$

Si E es discreto, entonces

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sup \left\{ \sum_{x \in E} f(x)\mu(x) - \sum_{x \in E} f(x)\nu(x) : |f| \leq 1 \right\}. \quad (4.7)$$

Demostración

Si $\max_{x \in E} |f(x)| \leq 1$, entonces, por un lado

$$\frac{1}{2} \left| \sum_{x \in E} f(x)\mu(x) - \sum_{x \in E} f(x)\nu(x) \right| \leq \frac{1}{2} \sum_{x \in E} |\mu(x) - \nu(x)| = \|\mu - \nu\|_{TV}.$$

Por lo tanto, el lado derecho de (4.6) es a lo mas $\|\mu - \nu\|_{TV}$.

Por otro lado, definimos

$$f^* = \begin{cases} 1 & \text{si } \mu(x) \geq \nu(x), \\ -1 & \text{si } \mu(x) < \nu(x). \end{cases}$$

Entonces usando la ecuación (4.4)

$$\begin{aligned} \frac{1}{2} \left[\sum_{x \in E} f^*(x)\mu(x) - \sum_{x \in E} f^*(x)\nu(x) \right] &= \frac{1}{2} \sum_{x \in E} f^*(x) [\mu(x) - \nu(x)] \\ &= \frac{1}{2} \left[\sum_{x \in E, \mu(x) \geq \nu(x)} [\mu(x) - \nu(x)] + \sum_{x \in E, \mu(x) < \nu(x)} [\mu(x) - \nu(x)] \right]. \end{aligned}$$

Por lo tanto, el lado derecho de (4.6) es al menos $\|\mu - \nu\|_{TV}$. ■

4.2. Introducción al Acoplamiento y su Relación con la Distancia de Variación Total

Definición 4.2 Acoplamiento

Un acoplamiento de dos distribuciones de probabilidad μ y ν es un par de variables aleatorias X y Y definidas en un único espacio de probabilidad tal que la distribución marginal de X es μ y la distribución marginal de Y es ν , es decir,

$$P\{X = x\} = \mu(x) \quad \text{y} \quad P\{Y = y\} = \nu(y) \quad \forall x, y \in E. \quad (4.8)$$

La importancia de introducir la Definición 4.2, es que el **acoplamiento** es una técnica general y poderosa, ya que es aplicable de diversas maneras. Por lo cual daremos una introducción a la relación que existe entre el **Acoplamiento y la Distancia de Variación Total**.

Ejemplo 4.1 Moneda justa.

Consideremos una moneda justa y sean μ y ν las medidas de probabilidad de la moneda, dando peso $\frac{1}{2}$ a los elementos $\{0, 1\}$.

- Si buscamos acoplar μ y ν es definir (X, Y) para un par de monedas independientes, tal que $P\{X = x, Y = y\} = 1/4$ para todo $x, y \in \{0, 1\}$.
- Otra forma de hacer un acoplamiento de μ y ν es dejar que X sea un lanzamiento de moneda justo y $Y = X$. En este caso, $P\{X = Y = 0\} = 1/2$, $P\{X = Y = 1\} = 1/2$ y $P\{X \neq Y\} = 0$.

Definición 4.3 Distribución Conjunta

Dado un acoplamiento de (X, Y) de μ y ν , si q es la distribución conjunta de (X, Y) en $E \times E$, lo que significa que $q(x, y) = P\{X = x, Y = y\}$, entonces q satisface

$$\sum_{y \in E} q(x, y) = \sum_{y \in E} P\{X = x, Y = y\} = P\{X = x\} = \mu(x), \quad x \in E$$

y

$$\sum_{x \in E} q(x, y) = \sum_{x \in E} P\{X = x, Y = y\} = P\{Y = y\} = \nu(y), \quad y \in E.$$

Por otro lado, notemos que si tenemos una distribución q en el espacio producto $E \times E$ que satisface

$$\sum_{y \in E} q(x, y) = \mu(x) \quad y \quad \sum_{x \in E} q(x, y) = \nu(y),$$

entonces hay un par de variables aleatorias (X, Y) que tienen a q como distribución conjunta y, en consecuencia, este par (X, Y) es un acoplamiento de μ y ν . Por lo tanto, podemos escribir un acoplamiento como un par de variables aleatorias (X, Y) definidas en un espacio de probabilidad común o mediante una distribución q en $E \times E$.

Ahora retomando el Ejemplo 4.1, lo especificaremos de manera equivalente por la distribución de probabilidad $q_1(x, y)$ en $\{0, 1\}^2$ dada por

$$q_1(x, y) = \frac{1}{4} \quad \text{para todo } (x, y) \in \{0, 1\}^2.$$

Y de la misma manera podemos ver el punto dos como la distribución de probabilidad $q_2(x, y)$ dada por

$$q_2(x, y) = \begin{cases} \frac{1}{2} & \text{si } (x, y) = (0, 0), \quad (x, y) = (1, 1), \\ 0 & \text{si } (x, y) = (0, 1), \quad (x, y) = (1, 0). \end{cases}$$

Supongamos que tenemos dos distribuciones μ y ν con un acoplamiento independiente. Sin embargo, cuando μ y ν no son idénticos, no será posible que X y Y tengan siempre el mismo valor. Por lo cual una pregunta a responder es ¿Qué tan cerca puede llegar un acoplamiento a tener X y Y idénticos?.

Proposición 4.4 Acoplamiento Óptimo

Sea μ y ν dos distribuciones de probabilidad en E . Entonces

$$\|\mu - \nu\|_{TV} = \inf \{P\{X \neq Y\} : (X, Y) \text{ es un acoplamiento de } \mu \text{ y } \nu\}. \quad (4.9)$$

Este ínfimo se alcanza para algún acoplamiento, que llamaremos el **acoplamiento óptimo**.

Demostración

Para cualquier evento A tenemos

$$\begin{aligned} \sum_{x \in E} \mu(x) \wedge \nu(x) &= \sum_{x \in E, \mu(x) \leq \nu(x)} \mu(x) + \sum_{x: \mu(x) > \nu(x)} \mu(x) - \sum_{x: \mu(x) > \nu(x)} \mu(x) + \sum_{x \in E, \mu(x) > \nu(x)} \nu(x) \\ \sum_{x \in E} \mu(x) \wedge \nu(x) &= 1 - \sum_{x \in E, \mu(x) > \nu(x)} [\mu(x) - \nu(x)]. \end{aligned}$$

Ahora por la ecuación (4.4) tenemos

$$\sum_{x \in E} \mu(x) \wedge \nu(x) = 1 - \|\mu - \nu\|_{TV} = p. \quad (4.10)$$

Supongamos que lanzamos una moneda con una probabilidad igual a p .

1. Si la moneda sale cara, elija un valor Z de acuerdo con la distribución de probabilidad,

$$\gamma_{III}(x) = \frac{\mu(x) \wedge \nu(x)}{p},$$

y establecemos $X = Y = Z$.

2. Si la moneda sale cruz, elija X de acuerdo con la distribución de probabilidad

$$\gamma_I(x) = \begin{cases} \frac{\mu(x) - \nu(x)}{\|\mu - \nu\|_{TV}} \text{ si } \mu(x) > \nu(x), \\ 0 \text{ en otro caso,} \end{cases}$$

e independientemente elija Y de acuerdo con la distribución de probabilidad

$$\gamma_{II}(x) = \begin{cases} \frac{v(x) - \mu(x)}{\|\mu - v\|_{TV}} & \text{si } v(x) > \mu(x), \\ 0 & \text{en otro caso,} \end{cases}$$

y por (4.4) aseguramos que γ_I y γ_{II} son distribuciones de probabilidad. Es claro que

$$p\gamma_{III} + (1 - p)\gamma_I = \mu$$

y

$$p\gamma_{III} + (1 - p)\gamma_{II} = v.$$

Por lo tanto, la distribución de X es μ y la distribución de Y es v . Notemos que en el caso de que la moneda caiga hacia arriba, $X \neq Y$, ya que γ_I y γ_{II} son positivos en subconjuntos disjuntos de X . Por lo tanto, $X = Y$ si y solo si el lanzamiento de la moneda es cara. Por lo cual concluimos que

$$P\{X \neq Y\} = \|\mu - v\|_{TV}.$$

4.3. Teorema de Convergencia

Teorema 4.1 Convergencia

Si P es irreducible y aperiódica, con distribución estacionaria π . Entonces existe constantes $\alpha \in (0,1)$ y $C > 0$ tal que

$$\max_{x \in E} \|P^t(x, \cdot) - \pi(\cdot)\|_{TV} \leq C\alpha^t \rightarrow 0 \quad \text{cuando } t \rightarrow \infty. \quad (4.11)$$

Demostración

La demostración de este teorema la dejaremos para más adelante, ya que daremos una prueba agradable y bonita usando acoplamiento.

4.4. Estandarización de Distancia Desde la Estacionariedad

Definición 4.4 Sea $d(t)$ la distancia desde la estacionariedad en el sentido variación total:

$$d(t) = \max_{x \in E} \|P^t(x, \cdot) - \pi\|_{TV} \quad (4.12)$$

Por el Teorema 4.1, sabemos que si $t \rightarrow \infty$ entonces $d(t) \rightarrow 0$ por lo cual, la siguiente definición será útil.

$$\bar{d}(t) = \max_{x, y \in E} \|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV} \quad (4.13)$$

Ahora veamos la relación que existe entre las ecuaciones (4.12) y (4.13).

Proposición 4.5 La siguiente desigualdad relaciona $d(t)$ y $\bar{d}(t)$:

$$d(t) \leq \bar{d}(t) \leq 2d(t) \quad \forall t. \quad (4.14)$$

Demostración

Por un lado $\bar{d}(t) \leq 2d(t)$ se sigue de la desigualdad del triángulo para distancia de variación total 4.5, insertando π como punto intermedio.

Por otro lado, para $d(t) \leq \bar{d}(t)$ usaremos la identidad $\pi(A) = \sum_{y \in E} \pi(y) P^t(y, A)$:

$$\begin{aligned} \|P^t(x, \cdot) - \pi(\cdot)\|_{TV} &= \max_{A \subset E} |P^t(x, A) - \pi(A)| \\ &= \max_{A \subset E} \left| \left(\sum_{y \in E} \pi(y) \right) P^t(x, A) - \sum_{y \in E} \pi(y) P^t(y, A) \right| \\ &= \max_{A \subset E} \left| \sum_{y \in E} \pi(y) (P^t(x, A) - P^t(y, A)) \right| \\ &\leq \max_{A \subset E} \sum_{y \in E} \pi(y) |P^t(x, A) - P^t(y, A)| \\ &\leq \sum_{y \in E} \pi(y) \max_{A \subset E} |P^t(x, A) - P^t(y, A)| \\ &= \sum_{y \in E} \pi(y) |P^t(x, A) - P^t(y, A)| \\ &\leq \max_{y \in E} |P^t(x, A) - P^t(y, A)|. \end{aligned}$$

Tomando el máximo sobre todo $x \in E$

$$\leq \max_{x, y \in E} |P^t(x, A) - P^t(y, A)|.$$

■

Proposición 4.6 *La función \bar{d} es submultiplicativa:*

$$\bar{d}(t+s) \leq \bar{d}(s)\bar{d}(t) \quad \forall t. \quad (4.15)$$

Demostración

Fijemos x, y en E , y sea (X_s, Y_s) el acoplamiento óptimo de $P^s(x, \cdot)$ y $P^s(y, \cdot)$ cuya existencia está garantizada por la Proposición 4.4. Por lo tanto

$$\|P^s(x, \cdot) - P^s(y, \cdot)\|_{TV} = P\{X_s \neq Y_s\}. \quad (4.16)$$

Tenemos

$$P^{s+t}(x, w) = \sum_{z \in E} P\{X_s = z\} P^t(z, w) = \mathbb{E}(P^t(X_s, w)). \quad (4.17)$$

Del mismo modo tenemos

$$P^{s+t}(y, w) = \sum_{z \in E} P\{Y_s = z\} P^t(z, w) = \mathbb{E}(P^t(Y_s, w)). \quad (4.18)$$

Para un conjunto $A \subset E$, sumando sobre $w \in A$, muestra que

$$\begin{aligned} \|P^{s+t}(x, \cdot) - P^{s+t}(y, \cdot)\|_{TV} &= \frac{1}{2} \sum_{w \in E} |\mathbb{E}(P^t(X_s, w) - P^t(Y_s, w))| \\ &\leq \mathbb{E} \left(\frac{1}{2} \sum_{w \in E} |P^t(X_s, w) - P^t(Y_s, w)| \right) \\ &= \mathbb{E} (\|P^t(X_s, \cdot) - P^t(Y_s, \cdot)\|_{TV}) \\ &= \mathbb{E} (\|P^t(X_s, \cdot) - P^t(Y_s, \cdot)\|_{TV} 1_{\{X_s \neq Y_s\}}) \\ &\leq \mathbb{E} (\bar{d}(t) 1_{\{X_s \neq Y_s\}}) \\ &= \bar{d}(t) P\{X_s \neq Y_s\} \\ &= \bar{d}(t)\bar{d}(s). \end{aligned}$$

Tomando el máximo sobre x y y , completa la demostración.

■

Corolario 4.1 *Para enteros no negativos c ,*

$$d(ct) \leq \bar{d}(ct) = \bar{d}(t+t+\cdots+t) \leq \bar{d}(t)^c \leq 2^c d(t)^c$$

Y aquí es donde retomamos la demostración del Teorema 4.1.

Demostración Teorema 4.1

$$\begin{aligned} \max_{x \in E} \|P^t(x, \cdot) - \pi(\cdot)\|_{TV} &= \max_{x \in E} \frac{1}{2} \sum_{y \in E} |P^t(x, y) - \pi(y)| \\ &\leq \max_{x \in E} \frac{1}{2} \sum_{y \in E} \max_{x, y \in E} (|P^t(x, y) - \pi(y)|) \\ &= \frac{n}{2} \max_{x, y \in E} (|P^t(x, y) - \pi(y)|), \end{aligned}$$

donde $n := |E|$. De la ecuación (5.7) sabemos que existe una $\delta > 0$ tal que para alguna $y \in E$, $t \geq 0$,

$$\max_{x, y \in E} (|P^t(x, y) - \pi(y)|) \leq \frac{1}{\delta} e^{-\delta t},$$

entonces bajo este escenario $C := \frac{n}{2\delta}$ y $\alpha := e^{-\delta}$ tenemos (5.7). ■

Proposición 4.7 Sean μ y ν distribuciones de probabilidad en una cadena de Markov P , entonces

$$\|\mu P - \nu P\|_{TV} \leq \|\mu - \nu\|_{TV}.$$

Demostración

Sea X_0, Y_0 el acoplamiento óptimo para μ y ν . Usando una representación de mapeo aleatorio como se vio en la Sección 2.3 y $Z \sim \text{Unif}(0, 1)$ para definir:

$$X_1 = f(X_0, Z)$$

y

$$Y_1 = f(Y_0, Z).$$

Por definición de la representación de mapeo aleatorio X_1, Y_1 se distribuyen como μP y νP respectivamente y también son un acoplamiento de estas distribuciones. Además si $X_0 = Y_0$, entonces $X_1 = Y_1$, es decir, $\{X_0 = Y_0\} \subset \{X_1 = Y_1\}$. Por lo tanto $\{X_0 \neq Y_0\} \supset \{X_1 \neq Y_1\}$.

$$\begin{aligned} \|\mu - \nu\|_{TV} &= P\{X_0 \neq Y_0\} \\ P\{X_0 \neq Y_0\} &\geq P\{X_1 \neq Y_1\} \\ &\geq \inf\{P(X \neq Y) : X, Y \text{ un acoplamiento de } \mu P, \nu P\} \\ &= \|\mu P - \nu P\|_{TV}. \end{aligned}$$

En consecuencia, esto muestra que $d(s)$ y $\bar{d}(s)$ son funciones no crecientes de s . ■

4.5. Tiempo de Mezcla

Esta sección será corta pero muy importante, ya que introduciremos un parámetro que mida el tiempo necesario por una cadena de Markov para que la distancia a la estacionaria sea pequeña.

Definición 4.5 Para una cadena de Markov dada, sea $\epsilon \in (0, 1)$ y definimos

$$t_{mix}(\epsilon) = \min\{t : d(t) \leq \epsilon\} \quad (4.19)$$

y

$$t_{mix} = t_{mix}\left(\frac{1}{4}\right). \quad (4.20)$$

Ahora por la Proposición 4.5 y por el Corolario 4.1 muestra que cuando c es un entero positivo,

$$d(ct_{mix}(\epsilon)) \leq \bar{d}(t_{mix}(\epsilon))^c \leq (2\epsilon)^c. \quad (4.21)$$

Si tomamos $\epsilon = \frac{1}{4}$

$$d(ct_{mix}) \leq 2^{-c} \quad (4.22)$$

Haciendo un cambio de variable $t = ct_{mix}$, $c = \frac{t}{t_{mix}}$, entonces tenemos

$$d(t) \leq \left(2^{\frac{1}{t_{mix}}}\right)^{-t} \quad (4.23)$$

y

$$t_{mix}(\epsilon) \leq \lceil \log_2 \epsilon^{-1} \rceil t_{mix}. \quad (4.24)$$

Capítulo 5

Acoplamiento

5.1. Introducción al Acoplamiento

Este capítulo es vital para el desarrollo de este trabajo, ya que una comparación entre distribuciones se reduce a una comparación entre variables aleatorias. A partir de la Proposición 4.4 caracteriza $\|\mu - \nu\|_{TV}$ como el mínimo sobre todos los acoplamientos (X, Y) de μ y ν , de que las probabilidades de que X y Y sean diferentes. En consecuencia, podemos encontrar cotas superiores en la distancia de variación total.

Ejemplo 5.1 Caminata Aleatoria Simple

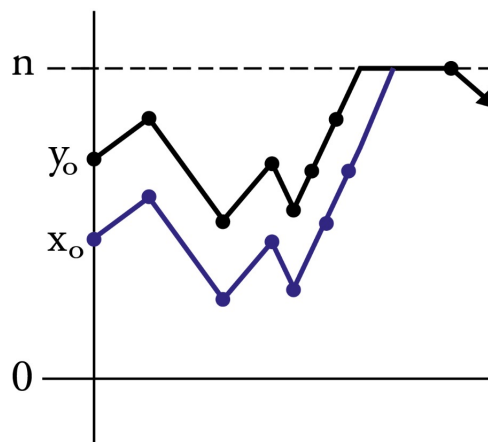


Figura 5.1: Acoplamiento de caminatas aleatorias definidas en $\{0, 1, \dots, n\}$.

Consideremos dos caminatas aleatorias simples en el segmento $\{0, 1, \dots, n\}$. Al momento en que la caminata llega al estado 0 no podrá bajar más o si la caminata llega al estado n no podrá subir más. En consecuencia si tenemos dos caminatas aleatorias X_t y Y_t , $t \geq 0$ para estados $x_0 \leq y_0$ entonces $P^t(x_0, n) \leq P^t(y_0, n)$.

Sea $\Delta_1, \Delta_2, \dots$ una sucesión de variables aleatorias independientes e idénticamente distribuidas que toman valores en $\{-1, 1\}$ y con media cero, entonces cada Δ_i es igualmente probable que sean $+1$ como -1 . Ambas cadenas X_t y Y_t se mueven con la misma regla: si $\Delta_t = +1$, ambas cadenas suben una unidad, siempre y cuando sea posible, por otro lado, si $\Delta_t = -1$, ambas cadenas bajan en una unidad, siempre y cuando sea posible. Las cadenas solamente pueden encontrarse en 0 o n , después de ese momento se mantendrán juntas.

Por lo cual X_t tiene función de transición $P^t(x, \cdot)$ y Y_t tiene función de transición $P^t(y, \cdot)$. Supongamos que $X_t = n$, entonces significa que $Y_t = n$, de donde concluimos que

$$P^t(x, n) = P\{X_t = n\} \leq P\{Y_t = n\} = P^t(y, n). \quad (5.1)$$

Y aquí es donde notamos la magia del acoplamiento, logramos juntar dos cadenas de tal manera que siempre $X_t \leq Y_t$, además, toda la información de las variables aleatorias, es fácil de leer.

El ejemplo anterior nos muestra un método de hacer un acoplamiento muy, pero muy útil, ya que nos permite construir dos copias de una cadena de Markov usando una fuente común de aleatoriedad.

Definición 5.1 Acoplamiento de Cadenas de Markov

Definimos un acoplamiento de cadenas de Markov con matriz de transición P para que sea un proceso $(X_s, Y_s)_{s=0}^{\infty}$ con la propiedad de que tanto (X_t) como (Y_t) son cadenas de Markov con la matriz de transición P , aunque las dos cadenas posiblemente tengan distribuciones iniciales diferentes.

Al igual que en el ejemplo anterior, la Definición 5.1, permite que las dos cadenas puedan tener diferentes distribuciones iniciales.

Definición 5.2 Acoplamiento Markoviano

Dada una cadena de Markov en E con matriz de transición P , un acoplamiento Markoviano de dos cadenas, es una cadena de Markov $\{(X_t, Y_t)\}_{t \geq 0}$ con espacio de estados $E \times E$ que satisface, para todo x, y, x', y'

$$P\{X_{t+1} = x' | X_t = x, Y_t = y\} = P(x, x')$$

$$P\{Y_{t+1} = y' | X_t = x, Y_t = y\} = P(y, y')$$

Definición 5.3 Acoplamiento Pegajoso

Dado un acoplamiento Markoviano, diremos que es un acoplamiento pegajoso si

$$X_s = Y_s, \quad \text{entonces} \quad X_t = Y_t \quad \text{para} \quad t \geq s. \quad (5.2)$$

A partir de este momento si X_t y Y_t son cadenas de Markov acopladas con $X_0 = x$ y $Y_0 = y$, entonces escribiremos $P_{x,y}$ para la función de probabilidad en el espacio donde X_t y Y_t están definidas.

5.2. Distancia de Variación Total Delimitada

Teorema 5.1 Sea (X_t, Y_t) un acoplamiento pegajoso como se definió en (5.2) con $X_0 = x$ y $Y_0 = y$. Sea τ_{acop} la primera vez que las cadenas se encuentran

$$\tau_{acop} = \min \{t : X_s = Y_s \text{ para todo } s \geq t\}. \quad (5.3)$$

Entonces

$$\|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV} \leq P_{x,y} \{\tau_{acop} > t\}. \quad (5.4)$$

Demostración

Primero veamos que $P^t(x, z) = P_{x,y} \{X_t = z\}$ y $P^t(y, z) = P_{x,y} \{Y_t = z\}$. Y dado que (X_t, Y_t) es un acoplamiento de $P^t(x, \cdot)$ y $P^t(y, \cdot)$ y por la Proposición 4.4

$$\|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV} \leq P_{x,y} \{X_t \neq Y_t\} \leq P_{x,y} \{\tau_{acop} > t\}.$$

La última desigualdad se sigue de $\{X_t \neq Y_t\} \subset \{\tau_{acop} > t\}$ por definición de τ .

■

Corolario 5.1 Supongamos que para cada par de estados $x, y \in E$ hay un acoplamiento (X_t, Y_t) con $X_0 = x$ y $Y_0 = y$. Para cada uno de estos acoplamientos, sea τ_{acop} el tiempo de coalescencia de las cadenas, como se define en (5.3).

Entonces

$$d(t) \leq \max_{x,y \in E} P_{x,y} \{\tau_{acop} > t\}.$$

Y por lo tanto $t_{mix} \leq 4 \max_{x,y} \mathbb{E}_{x,y}(\tau_{acop})$.

Demostración

Por la Proposición 4.5 tenemos que

$$d(t) \leq \bar{d}(t).$$

Por el Teorema 5.1

$$d(t) \leq \max_{x,y \in E} P_{x,y} \{\tau_{acop} > t\}$$

y

$$d(t) \leq \bar{d}(t) \leq \max_{x,y \in E} P_{x,y} \{\tau_{acop} > t\}.$$

■

5.3. Acoplamiento en el Hipercubo

La caminata aleatoria en el hipercubo $\{0, 1\}^n$, la definimos en el Ejemplo 3.3. Además, esta cadena es periódica, por lo estudiamos la cadena perezosa, en cada paso, el proceso permanece en su posición actual con probabilidad $\frac{1}{2}$ y con probabilidad $\frac{1}{2}$ se mueve una posición elegida uniformemente al azar entre todos los vértices vecinos. Por lo cual decidimos que una forma conveniente de generar la caminata perezosa es la siguiente

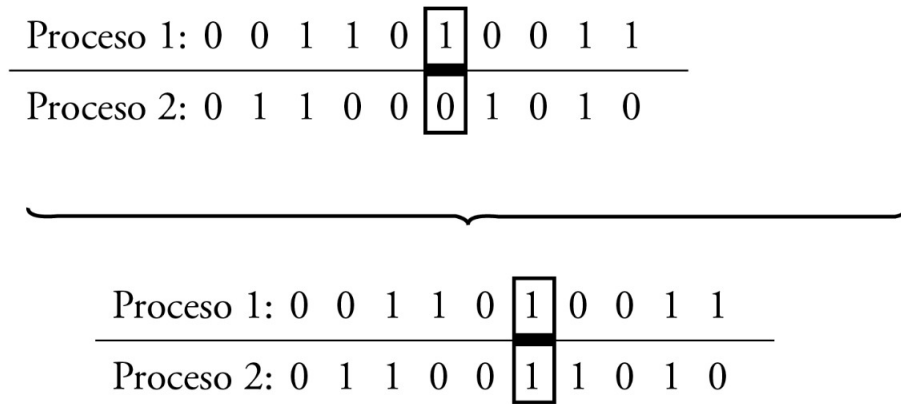


Figura 5.2: Del proceso 1 elegimos una coordenada uniformemente al azar, (en la imagen seleccionamos la sexta). Después con probabilidad $1/2$, cambiamos el bit de la coordenada elegida, pero en el proceso 2, para que el valor de las coordenadas en ambos procesos sea el mismo. Ver [13]

1. Elegir uniformemente una coordenada de las n -coordenadas.
2. Actualizar el bit de la coordenada seleccionada con un bit aleatorio en $\{0, 1\}$ con probabilidad $\frac{1}{2}$.

Definamos τ como la primera vez que todas las coordenadas se han seleccionado al menos una vez, entonces los dos procesos se mueven igual desde el momento τ en adelante. En consecuencia, tenemos que la distribución de τ es exactamente la misma que la variable aleatoria del Coleccionista de cupones 3.2, por el Corolario 5.1 y por la Proposición 3.4 se tiene que

$$d(n \log n + cn) \leq P\{\tau \leq n \log n + cn\} \leq e^{-c}. \tag{5.5}$$

Entonces

$$t_{mix}(\epsilon) \leq n \log n + \log \left(\frac{1}{\epsilon} \right) n. \quad (5.6)$$

De donde tenemos $t_{mix} = O(n \log n)$ cuando $n \rightarrow \infty$.

En este punto hemos realizado un trabajo preliminar y hemos establecido alguna notación, por lo que ahora daremos una buena prueba del teorema de convergencia usando el acoplamiento y así demostraremos lo útil y poderoso que es.

Teorema 5.2 Convergencia

Si P es irreducible y aperiódica, con distribución estacionaria π . Entonces existe constantes $\alpha \in (0,1)$ tal que

$$d(t) \leq (1 - \alpha)^t. \quad (5.7)$$

Demostración

Sea (X_t, Y_t) un acoplamiento de μ y ν . Recordemos que, por definición, X_0 tiene distribución μ y Y_0 tiene distribución ν . Denotemos esto por $X_0 \sim \mu$ y $Y_0 \sim \nu$. Como las distribuciones marginales de X_t y Y_t son μP^t y νP^t respectivamente, por el Teorema 5.1 y la Proposición 4.4,

$$\|\mu P^t - \nu P^t\|_{TV} \leq P_{x,y} \{X_t \neq Y_t\} = P \{\tau_{acop} > t\}.$$

Tengamos en cuenta que podemos tomar $\pi = \nu$, de modo que la desigualdad anterior acota la distancia desde la distribución estacionaria. Solo tenemos que demostrar que hay un acoplamiento (X_t, Y_t) , donde X_t y Y_t tiene transiciones independientes entre sí definidas por P , pero por la propiedad pegajosa, después de τ_{acop} se moverán del mismo modo.

Ya que P es irreducible y por la Proposición 2.5, existe un entero positivo r y un real positivo ϵ tal que para todo $x, y \in \Omega$, $P^r(x, y) > \epsilon > 0$. Ahora considere la probabilidad de que para algunos x_0 , $X_t \neq x_0$ y $Y_t \neq y_0$, lo que es equivalente a $P^t(X_t \neq Y_t)$. Por probabilidades complementarias e independencia, sabemos

$$P^t(X_t \neq Y_t) = 1 - P^t(X_t = x_0, Y_t = y_0) = 1 - P^t(X_t = x_0) P^t(Y_t = y_0) \leq 1 - \epsilon^2.$$

Ahora tengamos en cuenta que $X_{kr} \neq Y_{kr}$, es aproximadamente igual al momento inicial k , repitiendo este proceso el acoplamiento al momento r desde estados arbitrarios, encontramos que $X_r \neq Y_r$. Se sigue que $P^{kr}(X_t \neq Y_t) \leq (1 - \epsilon^2)^k$. Así, para nuestro acoplamiento independiente (X_t, Y_t) , con $t = kr > 0$, cuando $k \rightarrow \infty$,

$$\|\mu P^t - \pi\|_{TV} \leq P \{\tau_{acop} > t\} = P_{x,y} \{X_t \neq Y_t\} \leq (1 - \epsilon^2)^k \rightarrow 0.$$

■

Capítulo 6

Tiempos Estacionarios Fuertes

6.1. Top-to-Random Shuffle

Ejemplo 6.1 Top-to-Random Shuffle (Barajeo aleatorio la carta de arriba pasa a un lugar aleatorio)

Consideremos un mazo de n cartas:

1. Tomaremos la carta superior.
2. La insertaremos de manera uniforme al azar en el mazo.

Si repetimos este proceso, en algún momento, se mezclará el mazo. Teniendo que los arreglos sucesivos del mazo son una caminata aleatoria en el grupo simétrico S_n en $n!$ posibles permutaciones de las tarjetas. Lo que nos lleva a hacernos la siguiente pregunta, ¿ Existe una distribución estacionaria? y si es así, ¿ Cómo se comporta?.

Definición 6.1 Grupo

Un grupo es un conjunto G dotado de una operación asociativa, $\cdot : G \times G \rightarrow G$, si cumple

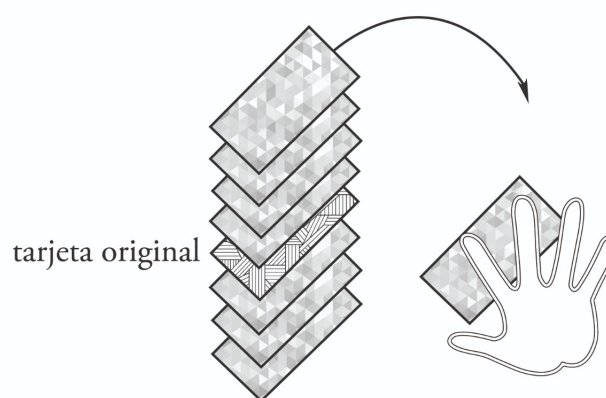


Figura 6.1: Método de barajar.

1. (G, \cdot) tiene un elemento neutro id para todo $g \in G$, $id \cdot g = g$ y $g \cdot id = g$.
2. Existe un elemento inverso $g^{-1} \in G$ para cada $g \in G$, $g \cdot g^{-1} = g^{-1} \cdot g = id$.

Definición 6.2 Caminata Aleatoria por la Izquierda

Dada una distribución de probabilidad μ en un grupo (G, \cdot) , definimos la **caminata aleatoria por la izquierda** en G con distribución de incremento μ de la siguiente manera: es una cadena de Markov con espacio de estado G y que se mueve multiplicando el estado actual a la izquierda por un elemento aleatorio de G seleccionado de acuerdo con μ .

En consecuencia, la matriz de transición P de esta cadena tiene entradas

$$P(g, hg) = \mu(h)$$

para todo $g, h \in G$.

Recordatorio 6.1 Como definimos la caminata aleatoria en grupos, multiplicamos el estado actual por el incremento a la izquierda. Una forma alternativa, uno puede considerar la caminata aleatoria por la derecha, donde

$$P(g, gh) = \mu(h) \quad \forall g, h \in G.$$

Ejemplo 6.2 El Hipercubo

Un ejemplo de una caminata aleatoria en un grupo G , es la caminata aleatoria en el hipercubo, definidos en el grupo \mathbb{Z}_2^n , que es el producto directo de n copias del grupo de dos elementos, $\mathbb{Z}_2 = \{0, 1\}$. Para la caminata aleatoria simple, la distribución es uniforme en el conjunto $\{e_i : 1 \leq i \leq n\}$, donde el vector e_i tiene un 1 en la posición i y 0 en todas las demás entradas. En el caso de la caminata en el hipercubo perezosa, la distribución μ da al vector 0 con peso $\frac{1}{2}$ y cada e_i con peso $\frac{1}{2n}$.

Proposición 6.1 Sea P la matriz de transición de una caminata aleatoria en un grupo finito G y sea U la distribución de la probabilidad uniforme en G . Entonces U es una distribución estacionaria para P .

Demostración

Sea μ los incrementos de la distribución de la caminata aleatoria. Entonces para cada $g \in G$

$$\sum_{h \in G} U(h)P(h, g) = \frac{1}{|G|} \sum_{k \in G} P(k^{-1}g, g) = \frac{1}{|G|} \sum_{k \in G} \mu(k) = \frac{1}{|G|} = U(g),$$

donde $k = gh^{-1}$.

■

Retomando el Ejemplo 6.1 y por la Proposición 6.1 tiene una distribución estacionaria uniforme.

Ahora una buena pregunta que nos interesa responder es, ¿Cuánto tiempo debemos barajar usando este método hasta que la disposición de la baraja sea casi aleatoria?.

Para lo cual definimos lo siguiente, sea τ_{top} la primera ocasión en que la carta inferior original se haya movido a la parte superior del mazo, ver 6.1.

Veamos que la disposición de las tarjetas en el momento τ_{top} se distribuye uniformemente en el conjunto S_n de todas las permutaciones de $\{1, \dots, n\}$ y, además, este elemento aleatorio de S_n es independiente del tiempo τ_{top} . Lo que nos lleva a la siguiente proposición.

Proposición 6.2 *Sea $(X_t)_{t \geq 0}$, la caminata aleatoria en S_n correspondiente a top-to-random shuffle 6.1 en n cartas. Dado el momento t que hay k cartas debajo de la carta inferior original, cada una de las $k!$ posibles pedidos de estas tarjetas son igualmente probables. Por lo tanto, si τ_{top} es un barajeo después de la primera vez que la carta inferior original se mueve a la parte superior del mazo, la distribución de $X_{\tau_{top}}$ es uniforme sobre S_n , y el tiempo τ_{top} es independiente de $X_{\tau_{top}}$*

Demostración

1. $t = 0$, no hay tarjetas debajo de la tarjeta inferior original y el argumento es trivialmente válido.
2. $t > 0$, Ahora supongamos que el argumento se mantiene. Por lo cual tenemos dos posibilidades en el tiempo $t + 1$: una tarjeta se coloca debajo de la tarjeta inferior original o no. En el segundo caso, las cartas debajo de la carta inferior original permanecen en orden aleatorio. En el primer caso, dado que la tarjeta se coloca debajo de la tarjeta inferior original, cada una de las ubicaciones posibles de $k + 1$ para la tarjeta es igualmente probable, por lo que cada uno de los pedidos $(k + 1)$ es equiprobable.

En otras palabras, la Proposición 6.2, es que para cualquier t , dado que $\tau_{top} = t$, la distribución de X_t es uniforme. ■

Ejemplo 6.3 Caminata Aleatoria en el Hipercubo

Retomando la caminata aleatoria perezosa $(X_t)_{t \geq 0}$, en el hipercubo $\{0, 1\}^n$, utilizamos el acoplamiento para limitar el tiempo de mezcla. Si utilizamos la representación de mapeo aleatorio, un elemento (j, B) de $\{1, 2, \dots, n\} \times \{0, 1\}$ se selecciona de manera uniformemente al azar, y la coordenada j del estado actual se actualiza con el bit B .

En consecuencia, acabamos de construir una sucesión (Z_t) independientes e idénticamente distribuidas, donde $Z_t = (j_t, B_t)$ es la coordenada y el bit utilizado para actualizar el paso t . Definimos

$$\tau_i := \min \{t \geq 0 : j_t = i, \quad t_i \leq t \quad \forall i \in \{1, \dots, n\}\}, \quad (6.1)$$

es decir, la primera vez cuando todas las coordenadas se han seleccionado al menos una vez para actualizar. Por lo tanto, en el momento $\tau_{refresh}$ todas las coordenadas han sido remplazadas por bits justos e independientes, la distribución de la cadena en este momento es uniforme en $\{0, 1\}^n$. Es decir, $X_{\tau_{refresh}}$, es una muestra exacta de la distribución estacionaria π .

6.2. Tiempos Estacionarios

Definición 6.3 Tiempo Estacionario

Un **tiempo estacionario** τ es un tiempo de paro aleatorio con la propiedad de que en el tiempo τ , la cadena de Markov ha alcanzado su distribución estacionaria:

$$P_x \{X_\tau = y\} = \pi(y) \quad \text{para todo } y. \quad (6.2)$$

Definición 6.4 Tiempo Estacionario Fuerte

Un tiempo estacionario fuerte para una cadena de Markov (X_t) con distribución estacionaria π es un tiempo de paro aleatorio τ , posiblemente dependiendo de la posición inicial x , es un tiempo estacionario para τ y X_τ son independientes, es decir :

$$P_x \{\tau = t, X_\tau = y\} = P_x \{\tau = t\} \pi(y) \quad \forall y. \quad (6.3)$$

Definición 6.5 Tiempo de Paro Aleatorio

Sean Z_1, Z_2, \dots, Z_n variables aleatorias que determinan X_1, X_2, \dots, X_n a través de la representación de mapeo aleatorio $X_n = f(X_{n-1}, Z_n)$, τ es llamado un tiempo de paro aleatorio si $\{\tau \leq n\} \subset \sigma(Z_1, \dots, Z_n)$, donde $\sigma(Z_1, \dots, Z_n)$ es la σ -álgebra generada por Z_1, \dots, Z_n .

Lema 6.1 Sea $(X_t)_{t \geq 0}$, una cadena de Markov irreducible con distribución estacionaria π . Si τ es un tiempo estacionario fuerte para X_t , entonces para todo $t \geq 0$ y

$$P_x \{\tau \leq t, X_\tau = y\} = P_x \{\tau \leq t\} \pi(y). \quad (6.4)$$

En palabras X_τ tiene distribución π y es independiente de τ .

Demostración

Sean Z_1, Z_2, \dots , valores aleatorios de la cadena de Markov, ahora utilizando la representación de mapeo aleatorio. Sea $s \leq t$ y escribimos

$$P_x \{\tau = s, X_t = y\} = \sum_{z \in E} P_x \{X_t = y | \tau = s, X_s = z\} P_x \{\tau = s, X_s = z\}$$

Ahora, $\{\tau = s, X_s = z\} \subset \sigma(Z_1, \dots, Z_n)$, mientras $\{X_t = y\} = \{f(X_s, (Z_{s+1}, Z_{s+2}, \dots, Z_t)) = y\}$, depende solo de X_s y $Z_{s+1}, Z_{s+2}, \dots, Z_t$. Dado que los (Z_t) son independientes, entonces

$$\begin{aligned}
 P_x \{X_t = y | \tau = s, X_s = z\} &= P_x \{f(X_s, (Z_{s+1}, Z_{s+2}, \dots, Z_t)) = y | \tau = s, X_s = z\} \\
 &= P_x \{X_t = y | \tau = s, X_s = z\} \\
 &= P_x \{f(X_s, (Z_{s+1}, Z_{s+2}, \dots, Z_t)) = y | \tau = s, X_s = z\} \\
 &= P_x \{f(z, (Z_{s+1}, Z_{s+2}, \dots, Z_t)) = y | \tau = s, X_s = z\} \\
 &= P_x \{f(z, (Z_{s+1}, Z_{s+2}, \dots, Z_t)) = y\} \\
 &= P^{t-s}(z, y).
 \end{aligned}$$

Por la definición de tiempo estacionario fuerte 6.4, tenemos

$$\begin{aligned}
 P_x \{\tau = s, X_t = y\} &= \sum_{z \in E} P_x \{X_t = y | \tau = s, X_s = z\} P_x \{\tau = s, X_s = z\} \\
 &= \left(\sum_{z \in E} P^{t-s}(z, y) \pi(z) \right) P_x \{\tau = t\} \\
 &= \pi(y) P_x \{\tau = t\}.
 \end{aligned}$$

La última igualdad es consistente ya que π es la distribución estacionaria. Sumando z da el resultado final. ■

Proposición 6.3 *Si τ es un tiempo estacionario fuerte, entonces*

$$d(t) = \max_{x \in E} \|P^t(x, \cdot) - \pi\|_{TV} \leq \max_{x \in E} P_x \{\tau > t\}. \quad (6.5)$$

La demostración de la Proposición 6.3 la desarrollaremos por medio de dos lemas. Para la cual definimos

$$s_x(t) := \max_{y \in E} \left[1 - \frac{P^t(x, y)}{\pi(y)} \right] \quad (6.6)$$

Además, mostraremos que $d_x(t) \leq s_x(t)$ y $s_x(t) \leq P_x \{\tau > t\}$. La combinación de estas dos desigualdades y tomando el máximo sobre x demostrarán la Proposición 6.3. Recordemos que $s_x(t)$ se llama la distancia de separación entre π y $P^t(x, \cdot)$.

Lema 6.2 *Si τ es un tiempo estacionario fuerte, entonces*

$$s_x(t) \leq P_x \{\tau > t\}. \quad (6.7)$$

Demostración

Sea $x \in E$, notemos que para cualquier $y \in E$

$$1 - \frac{P^t(x, y)}{\pi(y)} = 1 - \frac{P_x \{X_t = y\}}{\pi(y)} \leq 1 - \frac{P_x \{X_t = y, \tau \leq t\}}{\pi(y)}.$$

Por el Lema 6.1

$$1 - \frac{\pi(y)P_x \{\tau \leq t\}}{\pi(y)} = P_x \{\tau > t\}.$$

■

Nota 6.1 La desigualdad anterior es una igualdad si y solo si hay un nodo y para el cual $P_x \{X_t = y\} = P_x \{X_t = y, \tau \leq t\}$. Esto sucede si y solo si $\{X_t = y\} \subset \{\tau \leq t\}$. El nodo y se llama **estado de detención**, es un nodo para el que cuando llegamos allí sabemos que la cadena de Markov ha alcanzado su distribución de equilibrio.

Ejemplo 6.4 Caminata Aleatoria en el Hipercubo

Consideremos el estado $x = (x_1, \dots, x_n)$ donde $x_i \in \{0, 1\}$ en la caminata aleatoria del hipercubo n -dimensional, y definamos \bar{x}_i como el cambio de bit de la i -ésima coordenada. El estado $\bar{x} = (\bar{x}_1, \dots, \bar{x}_n)$ se alcanza cuando se ha actualizado el bit de todas las coordenadas. El estado \bar{x} es un estado de detención, ya que solo se alcanza cuando se haya actualizado cada bit. Si hay un estado de detención que comienza en x , entonces tendremos la siguiente igualdad $s_x(t) = P_x \{\tau > t\}$.

Lema 6.3 La **Distancia de Separación** acota la **Distancia de Variación Total** $\forall x, t$

$$\|P^t(x, \cdot) - \pi\|_{TV} \leq s_x(t), \quad (6.8)$$

donde π es la medida invariante.

Demostración

$$\begin{aligned} \|P^t(x, \cdot) - \pi\|_{TV} &= \sum_{y: P^t(x, y) < \pi(y)} \pi(y) - P^t(x, y) \\ &= \sum_{y: P^t(x, y) < \pi(y)} \pi(y) \left(1 - \frac{P^t(x, y)}{\pi(y)}\right) \\ &\leq \sum_{y: P^t(x, y) < \pi(y)} \pi(y) s_x(t) = s_x(t). \end{aligned}$$

■

Lema 6.4 Para una cadena reversible, las distancias de separación y variación total satisfacen

$$s(2t) \leq 1 - (1 - \bar{d}(t))^2 \leq 2\bar{d}(t) \leq 4d(t) \quad \forall t. \quad (6.9)$$

Demostración

La desigualdad de en medio se deduce de la expansión del cuadrado, y la última desigualdad se debe a la Proposición 4.5, por lo cual solo demostraremos la primera desigualdad.

Por reversibilidad, $\frac{P^t(z,y)}{\pi(y)} = \frac{P^t(y,z)}{\pi(z)}$ y Chapman-Kolmogorov obtenemos

$$\frac{P^{2t}(x,y)}{\pi(y)} = \sum_{z \in E} \frac{P^t(x,z)P^t(z,y)}{\pi(y)} = \sum_{z \in E} \pi(z) \frac{P^t(x,z)P^t(z,y)}{\pi(y)^2}.$$

Aplicando la desigualdad de Cauchy-Schwarz al lado derecho de arriba, tenemos

$$\begin{aligned} \frac{P^{2t}(x,y)}{\pi(y)} &\geq \left(\sum_{z \in E} \sqrt{P^t(x,z)P^t(y,z)} \right)^2 \\ &\geq \left(\sum_{z \in E} P^t(x,z) \wedge P^t(y,z) \right)^2. \end{aligned}$$

Por la ecuación (4.10)

$$\frac{P^{2t}(x,y)}{\pi(y)} \geq (1 - \|P^t(x, \cdot) - P^t(y, \cdot)\|_{TV})^2 \geq (1 - \bar{d}(t))^2$$

sí y solo si

$$\begin{aligned} \max_{x,y \in E} \left(1 - \frac{P^{2t}(x,y)}{\pi(y)} \right) &\leq 1 - (1 - \bar{d}(t))^2 \\ s(2t) &\leq 1 - (1 - \bar{d}(t))^2. \end{aligned}$$

■

Ejemplo 6.5 Hipercubo.

Como ya hemos visto, la variable aleatoria τ_i , (6.1), es el tiempo cuando cada coordenada ha sido seleccionada al menos una vez, es un tiempo estacionario fuerte. El tiempo τ_i ya ocurrió en el acoplamiento de coordenadas por coordenadas, y también vimos que es idéntico al tiempo del coleccionista de cupones. En consecuencia, obtenemos exactamente el mismo límite superior para t_{mix} que encontramos usando el método de acoplamiento. Por lo cual si combinamos la Proposición 3.4 y por el Lema 6.2, muestran que la distancia de separación satisface, para cada x ,

$$s_x(n \log n + cn) \leq e^{-c}. \quad (6.10)$$

Por el Lema 6.3,

$$t_{mix}(\epsilon) \leq n \log n + \log(\epsilon^{-1})n. \quad (6.11)$$

Capítulo 7

El Fenómeno del Cutoff

Comenzamos este capítulo observando una baraja de n cartas, en particular aplicaremos el barajeo *top-to-random shuffle*, que se discutió en el Ejemplo 6.1. Mediante el acoplamiento, podemos demostrar que para una baraja de n cartas, se aproxima a una configuración uniforme de la baraja en el límite a medida que $n \rightarrow \infty$ después de aproximadamente $n \log n + cn$ movimientos, donde c es una constante, es decir, después de $n \log n + cn$ la baraja estará totalmente barajada.

También demostraremos que si solo aplicamos $n \log n + cn$ el método *top-to-random shuffle* al mazo de n cartas, entonces las cartas barajadas aún se parecerán mucho a la configuración original. Por lo tanto, la configuración cae de *estar cerca de la configuración original* a *uniforme* en aproximadamente $n \log n$ movimientos de grado n .

Esta caída se llama cutoff y constituye el tema principal de esta tesis.

7.1. Un Paso Antes del Cutoff

Proposición 7.1 *Sea (X_t) la cadena de top-to-random 6.1 en n cartas. Para cualquier $\varepsilon > 0$, existe una constante $\alpha(\varepsilon)$ tal que $\alpha > \alpha(\varepsilon)$ implica que para todo n suficientemente grande,*

$$d_n(n \log n - \alpha n) \geq 1 - \varepsilon. \quad (7.1)$$

Es decir,

$$t_{mix}(1 - \varepsilon) \leq n \log n - \alpha n. \quad (7.2)$$

Demostración

Sea τ_1 la primera vez que se inserta una tarjeta debajo de la tarjeta inferior original. Dado que la probabilidad de que una tarjeta se ponga en la parte inferior es igual a $\frac{1}{n}$, esto sucederá después de aproximadamente n barajeos. Luego, sea τ_2 la primera vez que se coloca una tarjeta debajo de la tarjeta inferior original, es decir, después de τ_1 . Esto sucede con probabilidad $\frac{2}{n}$, esto sucederá después de aproximadamente $\frac{2}{n}$ barajeos. Las dos cartas debajo de la carta inferior original tienen la misma probabilidad de estar en cualquier orden. De manera similar, podemos definir $n - 1$ veces $\tau_1 < \tau_2 < \dots < \tau_{n-1}$.

En el momento τ_{n-1} , la tarjeta que originalmente estaba abajo, aparecerá en la parte superior. Por lo que se deduce que las $n - 1$ tarjetas inferiores tienen la misma probabilidad de estar en cualquiera de los $(n - 1)!$. En el momento $\tau_n = \tau_{n-1} + 1$, se vuelve a colocar la tarjeta inferior original en el mazo en una de las n posición. Por lo tanto, el mazo es igualmente probable que esté en cualquiera de los $n!$ arreglos en S_n al tiempo τ_n .

Cuando la tarjeta inferior original está en la posición i , se necesita aproximadamente $\frac{n}{i}$ barajeos para colocar una tarjeta debajo de ella. Por lo tanto tenemos

$$\mathbb{E}(\tau_i - \tau_{i-1}) = \frac{n}{i}$$

ya que $\tau_n = \sum_{i=1}^n (\tau_i - \tau_{i-1})$ con $\tau_0 = 0$, tenemos

$$\begin{aligned} \mathbb{E}(\tau_n) &= \mathbb{E}\left(\sum_{i=1}^n (\tau_i - \tau_{i-1})\right) \\ &= \sum_{i=1}^n \mathbb{E}(\tau_i - \tau_{i-1}) \\ &= \sum_{i=1}^n \frac{n}{i} \sim n \log n \quad \text{cuando } n \rightarrow \infty. \end{aligned}$$

Observa que τ_n es un tiempo fuerte y uniforme.

Para la demostración del Teorema, compararemos el top-to-random shuffle con una variable auxiliar aleatoria V_n , que representa el número de sorteos necesarios para sacar todas las n bolas, con reemplazo, al menos una vez, de una urna que contiene n bolas. Para $i = \{0, 1, \dots, n\}$, sea V_i que denota el número de sorteos necesarios, hasta que se hayan sacado i bolas distintas. Observemos que

$$\tau_i - \tau_{i-1} \stackrel{D}{=} V_{n-(i-1)} - V_{n-i} \stackrel{D}{=} \text{Geo}\left(\frac{i}{n}\right), \quad i = 1, \dots, n.$$

son independientes y que

$$\tau_n = \sum_{i=1}^n (\tau_i - \tau_{i-1}) \stackrel{D}{=} \sum_{i=1}^n (V_{n-(i-1)} - V_{n-i}) = V_n.$$

Sea id la permutación identidad y consideremos los siguientes eventos. Sea un mazo compuesta con n tarjetas, las cuales estarán indexadas de la siguiente manera, ver la Figura, 7.1, entonces definamos

$$B_j := \{\text{las cartas iniciales } j \text{ están en su orden original}\}, \quad j = 1, \dots, n. \quad (7.3)$$

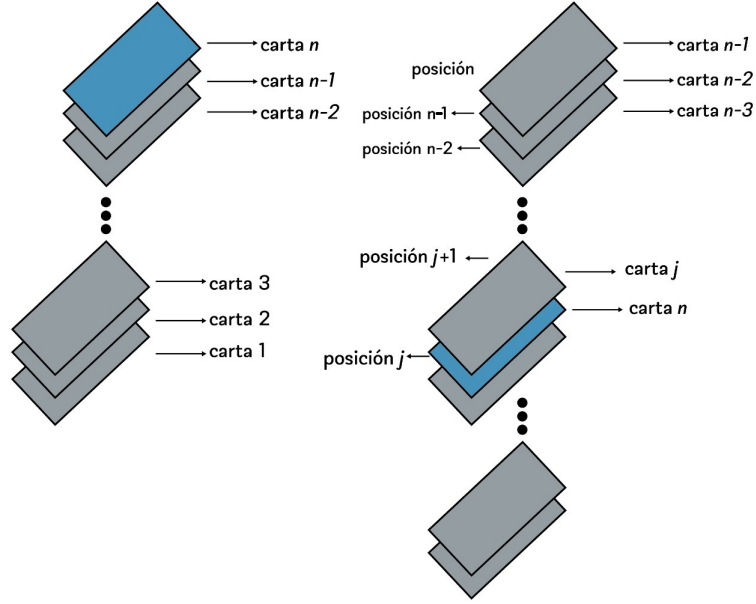


Figura 7.1: Las cartas estarán enumeradas de abajo hacia arriba, es decir, la carta de abajo será la 1 y la carta superior será la n , posteriormente la carta n bajará en alguna posición aleatoria.

Sea τ'_j que la cantidad de barajeos requeridos para mover la carta j de abajo hacia arriba, y sea $\tau_{j,i}$ denota el numero de barajeos necesarios para colocar la tarjeta i debajo de la tarjeta j . Ya que

$$\tau_{j,(i+1)} - \tau_{j,i} \stackrel{D}{=} \text{Geo}\left(\frac{i+j}{n}\right)$$

tenemos

$$\tau'_j = \sum_{i=0}^{n-j-1} (\tau_{j(i+1)} - \tau_{ji}) \stackrel{D}{=} \sum_{i=j}^{n-1} (V_{n-(i-1)} - V_{n-i}).$$

Ya que $\mathbb{E}(V_{n-(i-1)} - V_{n-i}) = \frac{n}{i}$ y $\text{Var}(V_{n-(i-1)} - V_{n-i}) = \frac{n^2(1-\frac{i}{n})}{i^2} < \frac{n^2}{i^2}$, tenemos

$$\begin{aligned} \mathbb{E}(\tau'_j) &= \mathbb{E}\left(\sum_{i=j}^{n-1} (V_{n-i} - V_{n-(i+1)})\right) \\ &= \sum_{i=j}^{n-1} \mathbb{E}(V_{n-i} - V_{n-(i+1)}) \\ &= \sum_{i=j}^{n-1} \frac{n}{i} \geq n(\log n - \log(j-1)) \end{aligned}$$

y

$$\text{Var}(\tau'_j) = \sum_{i=j}^{n-1} \frac{n^2}{i^2} < \frac{n^2}{j-1}, \quad j \geq 2.$$

Combinando estos límites con la desigualdad de Chebyshev obtenemos

$$\begin{aligned} P(\tau'_j < n \log n - \alpha n) &\leq P(\tau'_j - E(\tau'_j) < -n(\alpha - \log(j-1))) \\ &\leq P(|\tau'_j - E(\tau'_j)| > n(\alpha - \log(j-1))) \\ &\leq \frac{\text{Var}(\tau'_j)}{n^2(\alpha - \log(j-1))^2} \\ &\leq \frac{1}{j-1}, \quad \alpha \geq \log j + 2, \end{aligned}$$

la última desigualdad se cumple para $\alpha \geq \log j + 1$. Tomando $t_n(\alpha) = n \log n - \alpha n$. Entonces

$$P(X_{t_n(\alpha)} \in B_j) \geq P(\tau'_j < t_n(\alpha)) > 1 - \frac{1}{j-1}.$$

Junto con la observación de que $\pi(B_j) = \frac{1}{j!} \leq \frac{1}{(j-1)}$, y para $\alpha \geq \log j + 1$, tenemos

$$d_n(t_n(\alpha)) = \|P(X_{t_n(\alpha)} \in \cdot) - \pi(\cdot)\|_{TV} \geq P(X_{t_n(\alpha)} \in B_j) - \pi(B_j) > 1 - \frac{2}{j-1}.$$

Ahora tomemos $j = e^{\alpha-2}$, así que $n \geq e^{\alpha-2}$, y definimos

$$g(\alpha) := \frac{2}{e^{\alpha-2} - 1}.$$

Entonces para todo $\epsilon > 0$ existe un $\alpha_0 := \alpha_0(\epsilon)$ con $g(\alpha_0) = 1 - \epsilon$. Por lo tanto para todo $\alpha \geq \alpha_0$ tenemos

$$d_n(n \log n - \alpha n) \geq 1 - \epsilon.$$

■

Lo cual prueba la Proposición 7.1. Por lo tanto,

$$\liminf_{n \rightarrow \infty} d_n(t_n(\alpha)) > g(\alpha)$$

donde $g(\alpha) \rightarrow 1$ cuando $\alpha \rightarrow \infty$.

Recordemos que para el *top-to-random shuffle* 6.1 en n cartas, obtuvimos 5.5

$$d_n(n \log n + \alpha n) \leq e^{-\alpha}, \quad (7.4)$$

y por la Proposición 7.1

$$\liminf_{n \rightarrow \infty} d_n(n \log n + \alpha n) \geq 1 - 2e^{2-\alpha}. \quad (7.5)$$

En consecuencia, la cota superior en (7.4) tiende a 0 cuando $\alpha \rightarrow \infty$ y la cota inferior en (7.5) tiende a 1 cuando $\alpha \rightarrow \infty$.

Notemos que esta estimación es mucho más precisa que el hecho de que el tiempo de mezcla es del orden $n \log n$.

7.2. Definición del Cutoff

Con todo el trabajo desarrollado podemos proceder a definir el fenómeno del Cutoff.

Definición 7.1 Fenómeno del Cutoff

Para una sucesión de cadenas de Markov indexadas por $n \in \mathbb{N}$ con los tiempos de mezcla correspondientes $t_{mix}^{(n)}(\varepsilon)$. Esta sucesión tiene un cutoff si para todo $\varepsilon > 0$ tenemos

$$\lim_{n \rightarrow \infty} \frac{t_{mix}^{(n)}(\varepsilon)}{t_{mix}^{(n)}(1-\varepsilon)} = 1. \quad (7.6)$$

Por las ecuaciones (7.4) y (7.5) para la cadena *top-to-random* muestra que la distancia de variación total d_n para la cadena de n tarjetas decrece muy pero muy rápido a $t_{mix}^{(n)}$. En otra palabras, si nosotros hacemos un reescalamiento del tiempo t por $n \log n$, como $n \rightarrow \infty$ la función d_n se aproxima a una función indicadora:

$$\lim_{n \rightarrow \infty} d_n(cn \log n) = \begin{cases} 1 & \text{si } c < 1, \\ 0 & \text{si } c > 1. \end{cases} \quad (7.7)$$

Por lo tanto, esta propiedad es una caracterización de una sucesión de cadenas cuando tiene un cutoff.

Otra caracterización estaría dada por el Lema 18.1 que se encuentra en [13].

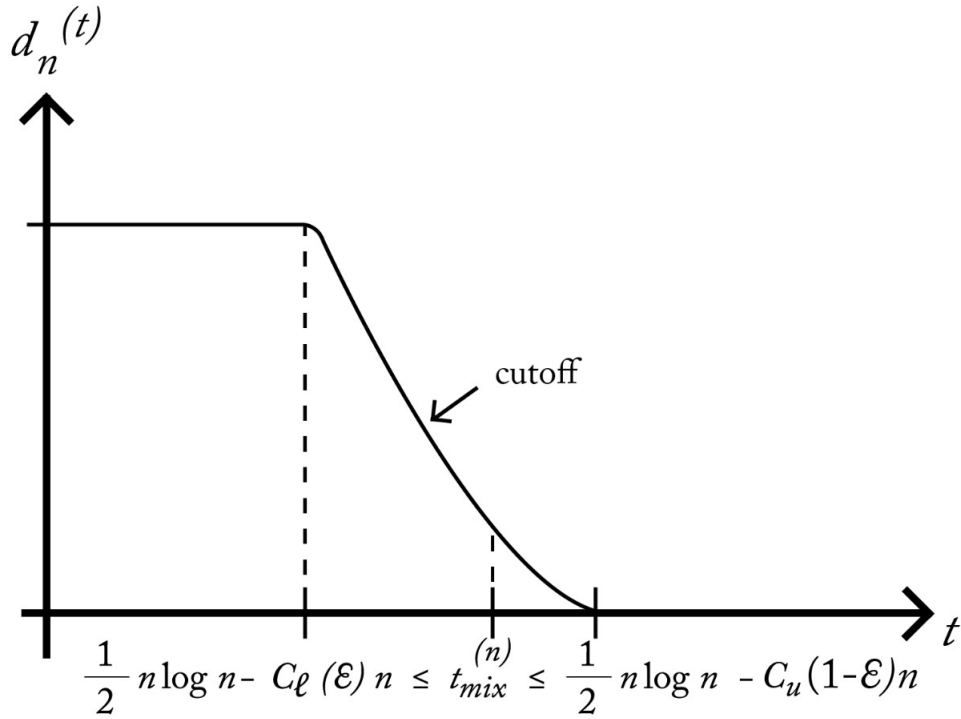


Figura 7.2: Para una cadena de Markov con cutoff, la gráfica de $d_n(t)$ contra t , cuando se ve en la escala de tiempo $t_{mix}^{(n)}$, se aproxima a una función escalonada cuando $n \rightarrow \infty$. Ver [13].

Lema 7.1 Sean $t_{mix}^{(n)}$ y d_n el tiempo de mezcla y la distancia a la distribución estacionaria, respectivamente, para la n -ésima cadena de una sucesión de cadenas de Markov. La sucesión tiene un límite si y solo si

$$\lim_{n \rightarrow \infty} d_n(ct_{mix}^{(n)}) = \begin{cases} 1 & \text{si } c < 1, \\ 0 & \text{si } c > 1. \end{cases} \quad (7.8)$$

Lema 7.2 Una sucesión de cadenas de Markov exhibe un cutoff si y solo si

$$d_n(ct_{mix}^{(n)}) \xrightarrow{n \rightarrow \infty} \begin{cases} 1 & \text{si } c < 1, \\ 0 & \text{si } c > 1. \end{cases}$$

Demostración \Leftarrow) De la definición de límite, para cualquier $\gamma > 0$ podemos elegir n lo suficientemente grande para hacer $d_n(ct_{mix}^{(n)})$ arbitrariamente cercano a 1. De lo anterior se deduce que para cualquier $0 < \epsilon < \frac{1}{2}$ podemos elegir n de tal manera que $d_n((1-\gamma)t_{mix}^{(n)}) > (1-\epsilon)$. Por tanto, para tal n ,

$$t_{mix}^{(n)}(1-\epsilon) \geq t_{mix}^{(n)}(1-\gamma), \quad (7.9)$$

de la misma manera podemos elegir n lo suficientemente grande para que

$$t_{mix}^{(n)}(\varepsilon) \geq t_{mix}^{(n)}(1 + \gamma). \quad (7.10)$$

Por lo tanto existe n tal que

$$1 \leq \frac{t_{mix}^{(n)}(\varepsilon)}{t_{mix}^{(n)}(1 - \varepsilon)} \leq \frac{1 + \gamma}{1 - \gamma}, \quad (7.11)$$

y tomando el límite cuando $\gamma \rightarrow 0$ (para que $n \rightarrow \infty$) obtenemos la definición de corte.

\Rightarrow) De la definición de límite, fijando γ , existe \bar{n} tal que $\forall n > \bar{n}$

$$\frac{t_{mix}^{(n)}(\varepsilon)}{t_{mix}^{(n)}(1 - \varepsilon)} \leq (1 + \gamma) = c \quad (7.12)$$

así que

$$t_{mix}^{(n)}(\varepsilon) \leq c t_{mix}^{(n)}(1 - \varepsilon) \leq c t_{mix}^{(n)}.$$

Por tanto, a partir de la definición de $d^{(n)}(\cdot)$ y tomando el límite de n para cubrir $0 < \varepsilon < \frac{1}{2}$,

$$\lim_{n \rightarrow \infty} d^{(n)}(c t_{mix}^{(n)}) \leq \varepsilon. \quad (7.13)$$

tomando $\varepsilon \rightarrow 0$ tenemos el caso $c > 1$. El caso $c < 1$ es completamente análogo. ■

Retomemos el ejemplo *top-to-random shuffle* 6.1, por construcción, las ecuaciones (7.4) y (7.5), son intervalos de longitud αn centrado en $n \log n$, por lo que la distancia de variación total, inicia es 1 y decrece hasta llegar a cero, lo que nos motiva a realizar la siguiente definición.

Definición 7.2 Ventana

Una sucesión de cadenas de Markov tiene un **cutoff** con una **ventana** de tamaño $O(w_n)$ si $w_n = o\left(t_{mix}^{(n)}\right)$ y

$$\lim_{\alpha \rightarrow -\infty} \liminf_{n \rightarrow \infty} d_n\left(t_{mix}^{(n)} + \alpha w_n\right) = 1, \quad (7.14)$$

y

$$\lim_{\alpha \rightarrow -\infty} \limsup_{n \rightarrow \infty} d_n \left(t_{mix}^{(n)} + \alpha w_n \right) = 0. \quad (7.15)$$

Definición 7.3 Pre-Cutoff

Diremos que una familia de cadenas tiene un **pre-cutoff** si satisface

$$\sup_{0 < \varepsilon < 1/2} \limsup_{n \rightarrow \infty} \frac{t_{mix}^{(n)}(\varepsilon)}{t_{mix}^{(n)}(1 - \varepsilon)} < \infty. \quad (7.16)$$

Definición 7.4 Dada μ una distribución de probabilidad en E y $f : E \rightarrow \Lambda$, escribimos μf^{-1} para la distribución de probabilidad definida por

$$(\mu f^{-1})(A) := \mu(f^{-1}(A))$$

para $A \subseteq \Lambda$.

Notemos que la definición anterior nos motiva a pensar que cuando X es una variable aleatoria con valores en E y distribución μ , entonces $f(X)$ tiene distribución μf^{-1} en Λ , por lo cual enunciamos el siguiente lema.

Lema 7.3 Sea μ y ν distribuciones de probabilidad en E , y sea $f : E \rightarrow \Lambda$ una función en E , donde Λ es un conjunto finito. Entonces

$$\|\mu - \nu\|_{TV} \geq \|\mu f^{-1} - \nu f^{-1}\|_{TV}.$$

Demostración

Sea $B \subseteq A \subseteq \Lambda$. Dado que

$$|\mu f^{-1}(B) - \nu f^{-1}(B)| = |\mu(f^{-1}(B)) - \nu(f^{-1}(B))|.$$

Entonces

$$\max_{B \subseteq \Lambda} |\mu f^{-1}(B) - \nu f^{-1}(B)| \leq \max_{A \subseteq E} |\mu(A) - \nu(A)|.$$

■

Proposición 7.2 Sea $f : E \rightarrow \Lambda$, con Λ un conjunto finito, y sean μ y ν dos medidas de probabilidad en E tal que

$$|\mathbb{E}_\mu(f) - \mathbb{E}_\nu(f)| \geq r \sigma_*,$$

Para algún $r > 0$, donde

$$\mathbb{E}_\mu(f) = \sum_{x \in E} \mu(x) f(x)$$

y

$$\sigma_* := \sqrt{\max\{Var_\mu(f), Var_\nu(f)\}}.$$

Entonces

$$\|\mu - \nu\|_{TV} \geq 1 - \frac{8}{r^2}.$$

Demostración Sea

$$A := \left(\mathbb{E}_\mu(f) + \frac{r\sigma_*}{2}, \infty \right)$$

y aplicando la desigualdad de Chebyshev tenemos

$$\begin{aligned} \mu f^{-1}(A) &= \mu \left(\left\{ x : f(x) \geq \mathbb{E}_\mu(f) + \frac{r\sigma_*}{2} \right\} \right) \\ &\leq \mu \left(\left\{ x : |f(x) - \mathbb{E}_\mu(f)| \geq \frac{r\sigma_*}{2} \right\} \right) \\ &\leq \frac{4}{r^2}. \end{aligned}$$

$$\begin{aligned} \nu f^{-1}(A) &= \nu \left(\left\{ x : f(x) \geq \mathbb{E}_\nu(f) + \frac{r\sigma_*}{2} \right\} \right) \\ &\leq \nu \left(\left\{ x : |f(x) - \mathbb{E}_\nu(f)| \geq \frac{r\sigma_*}{2} \right\} \right) \\ &\leq \frac{4}{r^2}. \end{aligned}$$

Por el Lema 7.3 se tiene

$$\begin{aligned} \|\mu - \nu\|_{TV} &\geq \|\mu f^{-1} - \nu f^{-1}\|_{TV} = \sup_{B \subset \Lambda} |\mu f^{-1}(B) - \nu f^{-1}(B)| \\ &\geq \mu f^{-1}(A) - \nu f^{-1}(A) \geq 1 - \frac{4}{r^2} - \frac{4}{r^2} \\ &= 1 - \frac{8}{r^2}. \end{aligned}$$

■

Otra proposición similar a la proposición anterior, es la siguiente, la cual la podemos encontrar en [13] como la Proposición 7.12.

Proposición 7.3 *Sea μ y ν dos distribuciones de probabilidad en E , y sea f una función de valores reales en E . Si*

$$|\mathbb{E}_\mu(f) - \mathbb{E}_\nu(f)| \geq r\sigma, \quad (7.17)$$

donde $\sigma^2 = (\text{Var}_\mu(f) - \text{Var}_\nu(f))/2$, entonces

$$\|\mu - \nu\|_{TV} \geq 1 - \frac{4}{4 + r^2}. \quad (7.18)$$

El propósito de la Proposición 7.2 es encontrar una cota inferior al tiempo de mezcla para la caminata aleatoria en el hipercubo.

Lema 7.4 *Consideremos el problema del coleccionista de cupones con n tipos de cupones distintos, y sea $I_j(t)$ la indicadora del evento en que el j -ésimo cupón no ha sido recolectado al tiempo t . Sea $R_t = \sum_{j=1}^n I_j(t)$ el número de cupones que aún no se han adquirido al tiempo t . Las variables aleatorias $(I_j(t))$ están correlacionadas negativamente y sea $p = (1 - \frac{1}{n})^t$, para $t \geq 0$ tenemos*

$$\mathbb{E}(R_t) = np, \quad (7.19)$$

y

$$\text{Var}(R_t) \leq np(1-p) \leq \frac{n}{4}. \quad (7.20)$$

Demostración

Notemos que si $I_j(t) = 1$ si y solo si los primeros cupones al tiempo t no son del tipo j , entonces

$$\mathbb{E}(I_j(t)) = \left(1 - \frac{1}{n}\right)^t = p \quad \text{y} \quad \text{Var}(I_j(t)) = p(1-p).$$

Similarmente, para $j \neq k$,

$$\mathbb{E}(I_j(t)I_k(t)) = \left(1 - \frac{2}{n}\right)^t.$$

De donde, obtenemos que

$$\text{Cov}(I_j(t), I_k(t)) = \left(1 - \frac{2}{n}\right)^t - \left(1 - \frac{1}{n}\right)^{2t} < 0.$$

■

Recordemos que demostramos la ecuación (5.6)

$$t_{mix}(\epsilon) \leq n \log n + \log \left(\frac{1}{\epsilon} \right) n. \quad (7.21)$$

En resumen, tenemos que

$$t_{mix}(1 - \epsilon) \geq \frac{1}{2} n \log n - c(\epsilon) n. \quad (7.22)$$

y

$$t_{mix}(\epsilon) \leq \frac{1}{2} n \log n - c(\epsilon) n. \quad (7.23)$$

Para la cadena *top-to-random shuffle* de n cartas, obtuvimos las ecuaciones (7.22) y (7.23) las cuales demuestran que hay un **cutoff** en $\frac{1}{2} n \log n$ de orden n .

Teorema 7.1 *Sea X_t , $t \geq 0$, la caminata aleatoria en el hipercubo n -dimensional con distribución estacionaria π y sea $(W_t)_{t \in \mathbb{N}}$ la caminata en la urna de Ehrenfest con distribución estacionaria π_W . Entonces*

$$\|P_1(X_t \in \cdot) - \pi(\cdot)\|_{TV} = \|P_n(X_t \in \cdot) - \pi_W(\cdot)\|_{TV}. \quad (7.24)$$

Denotaremos por P_1 a la probabilidad inicial del vértice con todas las coordenadas de 1's y denotaremos por P_n la probabilidad inicial de que las n bolas estén en la urna A.

Demostración

Sea W la función de peso del Hamming como la definimos, y para $w \in \{0, 1, \dots, n\}$ sea $E_w := \{x \in \mathbb{Z}_2^n : W(x) = w\}$ el conjunto para todo $x \in \mathbb{Z}_2^n$ con w bits. Recordemos que para todo $x, y \in E_w$, $P_0(X_t = x) = P_0(X_t = y)$ y $\pi(x) = \pi(y)$, de modo que

$$\sum_{x \in E_w} |P_0(X_t = x) - \pi(x)| = \left| \sum_{x \in E_w} (P_0(X_t = x) - \pi(x)) \right| = |P_0(W_t = w) - \pi_w(w)|.$$

Sumando sobre $w \in \{0, 1, \dots, n\}$ y dividiendo entre 2, obtenemos

$$\|X_t - \pi\|_{TV} = \|W_t - \pi_W\|_{TV}.$$

■

Teorema 7.2 *Sea (Δ_i) valores enteros de variables aleatorias con media cero y varianza σ^2 . Sea $X_t = \sum_{i=1}^t \Delta_i$. Entonces*

$$P\{X_t \neq 0 \text{ para } 1 \leq t \leq r\} \leq \frac{4\sigma}{\sqrt{r}}. \quad (7.25)$$

Demostración

Para $I \subset \mathbb{Z}$ sea

$$L_r(I) := \{t \in \{0, 1, \dots, r\} : X_t \in I\}$$

el conjunto de tiempos hasta r , en el que $X_t \in I$. Entonces sea

$$A_r := \{t \in L_r(0) : X_{t+u} \neq 0 \text{ para } 1 \leq u \leq r\}$$

el conjunto de tiempos t en $\{0\}$ después de lo cual la caminata no toca 0 para otros r pasos (claramente $|A_r| \leq 1$).

Dado que el futuro de la caminata después de visitar 0 no depende de lo que sucedió antes,

$$P(t \in A_r) = P(t \in L_r(0)) \alpha_r$$

donde

$$\alpha_r := P_0(X_t \neq 0, t = 1, 2, \dots, r).$$

Sumando sobre t , tenemos

$$\begin{aligned} 1 &\geq \mathbb{E}[|A_r|] = \mathbb{E}\left[\sum_{t=0}^r \mathbf{1}_{\{t \in A_r\}}\right] = \sum_{t=0}^r P(t \in A_r) \\ &= \sum_{t=0}^r P(t \in L_r(0)) \alpha_r = \mathbb{E}\left[\sum_{t=0}^r \mathbf{1}_{\{t \in L_r(0)\}}\right] \alpha_r \\ &= \mathbb{E}[|L_r(0)|] \alpha_r. \end{aligned}$$

Solo queda estimar una cota inferior para $E[|L_r(0)|]$. Por la desigualdad de Chebyshev (teniendo en cuenta que $\text{Var}(X_t) = t\sigma^2$)

$$P(|X_t| \geq \sigma\sqrt{r}) = P(|X_t - 0| \geq \sigma\sqrt{t} \frac{\sqrt{r}}{\sqrt{t}}) \leq \frac{t}{r}.$$

Tomando $I := (-\sigma\sqrt{r}, \sigma\sqrt{r})$,

$$\mathbb{E}[|L_r(I^c)|] = \mathbb{E}\left[\sum_{t=0}^r \mathbf{1}_{\{t \in L_r(I^c)\}}\right] \leq \sum_{t=0}^r \frac{t}{r} = \frac{r+1}{2},$$

De dónde

$$\mathbb{E}[|L_r(I)|] = r + 1 - \mathbb{E}[|L_r(I^c)|] > \frac{r}{2}.$$

Además, para cualquier $v \neq 0$,

$$\begin{aligned} \mathbb{E}[|L_r(v)|] &= \mathbb{E}\left[\sum_{t=0}^r 1_{\{X_t=v\}}\right] = \mathbb{E}\left[\sum_{t=\tau_v}^r 1_{\{X_t=v\}}\right] \leq \mathbb{E}_v\left[\sum_{t=0}^r 1_{\{X_t=v\}}\right] \\ &= \mathbb{E}_0\left[\sum_{t=0}^r 1_{\{X_t=0\}}\right], \end{aligned}$$

donde para la desigualdad hemos utilizado la propiedad de Markov, que dice que la cadena después de τ_v tiene la misma distribución de la cadena comenzada desde v .

Así

$$\frac{r}{2} \leq \mathbb{E}[|L_r(I)|] \leq 2\sigma\sqrt{r}\mathbb{E}[|L_r(0)|],$$

que junto con $\mathbb{E}[|L_r(0)|] \alpha_r$ prueba el teorema. ■

7.3. Teorema, El Hipercubo n-dimensional Tiene un Cutoff

Proposición 7.4 *Para la caminata aleatoria en el hipercubo n-dimensional, se tiene que*

$$d\left(\frac{1}{2}n \log n - \alpha n\right) \geq 1 - 4e^{-2\alpha+1}. \quad (7.26)$$

Demostración

Sea $(X_t)_{t \geq 0}$, la caminata aleatoria en el hipercubo n-dimensional y sea $W(x) = \sum_{i=1}^n x^i$ la función de peso de Hamming. Dado que $(X_t)_{t \geq 0}$ es una caminata aleatoria reversible y transitiva, sabemos que π es uniforme en \mathbb{Z}_2^n . Bajo π , la variable aleatoria $W_t := W(X_t)$ es binomial con parámetros n y $p = \frac{1}{2}$ y por lo tanto

$$\mathbb{E}_\pi(W_t) = \frac{n}{2} \quad y \quad Var_\pi(W_t) = \frac{n}{4}. \quad (7.27)$$

Ahora sea R_t el número de coordenadas que no se actualizaron en el tiempo t . Al comenzar desde 0, la distribución condicional de W_t dada R_t es binomial con parámetros $n - R_t$ y $p = 1/2$. Entonces

$$\mathbb{E}_0(W_t|R_t) = \frac{n - R_t}{2} \quad y \quad Var_0(W_t|R_t) = \frac{n - R_t}{4}.$$

Usando el Lema 7.4, obtenemos

$$\mathbb{E}_0(W_t) = \mathbb{E}(\mathbb{E}_0(W|R_t)) = \frac{n - \mathbb{E}_0(R_t)}{2} = \frac{n}{2} \left(1 - \left(1 - \frac{1}{n}\right)^t\right). \quad (7.28)$$

Usando la identidad

$$\text{Var}(Y) = \text{Var}(\mathbb{E}(Y|Z)) + \mathbb{E}(\text{Var}(Y|Z)),$$

obtenemos

$$\begin{aligned} \text{Var}_0(W_t) &= \text{Var}_0(\mathbb{E}_0(W_t|R_t)) + \mathbb{E}_0(\text{Var}_0(W_t|R_t)) \\ &= \text{Var}_0\left(\frac{n - R_t}{2}\right) + \mathbb{E}_0\left(\frac{n - R_t}{4}\right) \\ &= \frac{n}{4} + \frac{1}{4}(\text{Var}_0(R_t) - \mathbb{E}_0(R_t)) \leq \frac{n}{4}, \end{aligned}$$

Donde la última desigualdad se debe al hecho de que I_j están negativamente relacionadas de modo que $\mathbb{E}(R_t) \geq \text{Var}(R_t)$. Tomando

$$\sigma = \sqrt{\text{máx}\{\text{Var}_\pi(W_t), \text{Var}_0(W_t)\}} = \frac{\sqrt{n}}{2}.$$

Entonces, por las ecuaciones (7.27) y (7.28), tenemos

$$\begin{aligned} |\mathbb{E}_0(W_t) - \mathbb{E}_\pi(W_t)| &= \left| \frac{n}{2} \left(1 - \left(1 - \frac{1}{n}\right)^t\right) - \frac{n}{2} \right| \\ &= \frac{n}{2} \left(1 - \frac{1}{n}\right)^t = \sigma \sqrt{n} \left(1 - \frac{1}{n}\right)^t \\ &= \sigma e^{t \log(1 - \frac{1}{n}) + \frac{1}{2} \log n}. \end{aligned}$$

Usando el hecho de que $\log(1 - x) \leq -x - x^2$ para $0 \leq x \leq 1/2$, tenemos

$$\sigma e^{t \log(1 - \frac{1}{n}) + \frac{1}{2} \log n} \geq \sigma e^{-\frac{t}{n} \left(\frac{n+1}{n}\right) + \frac{1}{2} \log n}.$$

Aplicando la Proposición 7.3 con $r = e^{-\frac{t}{n} \log(\frac{n+1}{n}) + \frac{1}{2} \log n}$, obtenemos

$$\begin{aligned} \|P_0(X_t \in \cdot) - \pi(\cdot)\|_{TV} &\geq 1 - \frac{4}{4 + e^{-\frac{2t}{n} \left(\frac{n+1}{n}\right) + \frac{1}{2} \log n}} \\ &\geq 1 - 4e^{-\frac{2t}{n} \left(\frac{n+1}{n}\right) - \log n}. \end{aligned}$$

Para $\alpha \leq n \log n$ tenemos

$$(n+1) \left(\frac{1}{2} n \log n - \alpha n \right) \leq n \left(\frac{1}{2} n \log n - \alpha n \right) + \frac{1}{2} n^2.$$

Si tomamos

$$t_n = \frac{n}{n+1} \left(\frac{1}{2} n \log n - \left(\alpha - \frac{1}{2} \right) n \right),$$

entonces

$$t_n > \frac{1}{2} n \log n - \alpha n.$$

Por lo tanto,

$$d(t_n) = \|P_0(X_{t_n} \in \cdot) - \pi(\cdot)\|_{TV} \geq 1 - 4e^{-2\alpha+1}. \quad (7.29)$$

■

Proposición 7.5 *La caminata aleatoria en el hipercubo n -dimensional existe una constante $c > 0$ tal que*

$$d\left(\frac{1}{2} n \log n + \alpha n\right) \leq \frac{c}{\sqrt{\alpha}}. \quad (7.30)$$

Demostración Dado que la cadena es transitiva, tenemos,

$$d(t) = \|P_1(X_t \in \cdot) - \pi(\cdot)\|_{TV} = \|P_n(W_t \in \cdot) - \pi_W(\cdot)\|_{TV},$$

Sea (W_t, Z_t) un acoplamiento de dos cadenas de Ehrenfest, donde W_t y Z_t inician en w y z respectivamente. Siempre que las dos partículas aún no hayan chocado, en cada movimiento, se lanza una moneda justa para determinar cuál de las dos partículas se mueve; la partícula elegida realiza una transición de acuerdo con la siguiente matriz de transición

$$P(j, k) = \begin{cases} \frac{n-j}{n} & \text{si } k = j+1, \\ \frac{j}{n} & \text{si } k = j-1, \\ 0 & \text{en otro caso,} \end{cases} \quad (7.31)$$

Mientras que la otra partícula permanece en su posición actual. Las partículas se mueven juntas una vez que se encuentran por primera vez. Al igual como lo hemos hecho, sin pérdida de generalidad, que $z \geq w$. Como las partículas nunca se cruzan,

$Z_t \geq W_t$ para todo t . Sea $D_t = |Z_t - W_t|$, entonces $D_t = Z_t - W_t \geq 0$. Definamos $\tau := \min\{t \geq 0 : Z_t = W_t\}$. Si condicionamos que $(Z_t, W_t) = (z_t, w_t)$ donde $z_t \neq w_t$

$$D_{t+1} - D_t = \begin{cases} 1 & \text{con probabilidad } \left(\frac{1}{2}\right)\left(1 - \frac{z_t}{n}\right) + \left(\frac{1}{2}\right)\left(\frac{w_t}{n}\right), \\ -1 & \text{con probabilidad } \left(\frac{1}{2}\right)\left(\frac{z_t}{n}\right) + \left(\frac{1}{2}\right)\left(1 - \frac{w_t}{n}\right). \end{cases} \quad (7.32)$$

De (7.32) podemos ver que para los eventos $\{\tau > t\}$,

$$\mathbb{E}_{z,w}[D_{t+1} - D_t | Z_t = z_t, W_t = w_t] = -\frac{(z_t - w_t)}{n} = -\frac{D_t}{n}. \quad (7.33)$$

Ahora dado que $\mathbf{1}_{\{\tau > t\}} \leq \mathbf{1}_{\{Z_t \neq W_t\}}$, depende solo de la historia de la cadena hasta el tiempo t y de la propiedad de Markov, tenemos

$$\mathbb{E}_{z,w}[\mathbf{1}_{\{Z_t \neq W_t\}}(D_{t+1}) | Z_0, \dots, Z_t, W_0, \dots, W_t] = \left(1 - \frac{1}{n}\right) \mathbf{1}_{\{Z_t \neq W_t\}}(D_{t+1}) \quad (7.34)$$

Tomando esperanza en ambos lados, tenemos

$$\mathbb{E}_{z,w}[\mathbf{1}_{\{Z_t \neq W_t\}} D_{t+1}] = \left(1 - \frac{1}{n}\right) \mathbb{E}_{z,w}[\mathbf{1}_{\{Z_t \neq W_t\}} D_{t+1}],$$

para todo t , y ya que $\mathbf{1}_{\{\tau > t+1\}} \leq \mathbf{1}_{\{\tau > t\}}$, tenemos

$$\mathbb{E}_{z,w}[\mathbf{1}_{\{\tau > t+1\}} D_{t+1}] \leq \left(1 - \frac{1}{n}\right) \mathbb{E}_{z,w}[\mathbf{1}_{\{\tau > t\}} D_{t+1}].$$

Definamos $a_t := \mathbb{E}[\mathbf{1}_{\{\tau > t\}} D_t]$, entonces para $t = 0$ $a_0 = \mathbb{E}[\mathbf{1}_{\{\tau > 0\}} D_0] = z - w$. Supongamos que $a_0 = \mathbb{E}[\mathbf{1}_{\{\tau > 0\}} D_0] = z - w$. Entonces tenemos que

$$a_t = \left(1 - \frac{1}{n}\right) a_{t-1} \leq \left(1 - \frac{1}{n}\right)^t z - w.$$

Veamos que

$$e^{\log(1-\frac{1}{n})^t} = e^{t \log(1-\frac{1}{n})} \approx e^{-\frac{t}{n}}.$$

Por lo tanto, tenemos que

$$\mathbb{E}_{z,w} [D_t 1_{\{\tau > t\}}] \leq \left(1 - \frac{1}{n}\right)^t (z - w) \leq n e^{-\frac{t}{n}}. \quad (7.35)$$

Ahora observemos que si $\tau > t$ los incrementos $D_{t+1} - D_t$ tienden a ser negativos; de hecho, las probabilidades en (7.32) y (7.33) dicen que en cada paso D_t aumenta en una unidad con probabilidad no mayor que $\frac{1}{2} - \frac{1}{2n}$ y disminuye en una unidad con probabilidad al menos $\frac{1}{2} + \frac{1}{2n}$. Por lo tanto, es posible acoplar D_t con un paseo aleatorio simétrico S_t en el espacio de estados $E_S = \mathbb{N}$ que tiene probabilidad $\frac{1}{2} - \frac{1}{2n}$ de ir hacia la arriba o hacia la abajo y que es un poco perezoso (se queda quieto con probabilidad $\frac{1}{2}$). Podemos suponer que $S_0 = z - w$ y obligar a S_t dominar a D_t en el sentido de que siempre que S_t da un paso hacia arriba, obligamos a D_t a hacer lo mismo, de esta manera, $\forall t \leq \tau, \forall t \leq \tau$.

Definamos $\tilde{\tau} := \min\{t \geq 0 : S_t = 0\}$, entonces $\tau \leq \tilde{\tau}$. Y por el Teorema 7.2, existe una constante c tal que para todo $k \geq 0$,

$$P_{z,w} \{\tau > u\} \leq P_{z-w} \{\tilde{\tau} > u\} \leq \frac{c(z-w)}{\sqrt{u}}. \quad (7.36)$$

En consecuencia, tenemos que

$$P_{z,w} \{\tau > s + u | D_0, \dots, D_s\} = 1_{\{\tau > s\}} P_{D_s} \{\tau > u\} \leq 1_{\{\tau > s\}} \frac{cD_s}{\sqrt{u}}.$$

Tomando esperanza y aplicando (7.35) tenemos

$$P_{z,w} \{\tau > s + u\} \leq \frac{c n e^{-\frac{s}{n}}}{\sqrt{u}}, \quad (7.37)$$

donde $u = \alpha n$ y $s = \left(\frac{1}{2}\right) n \log n$ y por el Corolario 5.1

$$P_{z,w} \left(\tau > \left(\frac{1}{2}\right) n \log n + \alpha n \right) \leq \frac{c}{\sqrt{\alpha}}. \quad (7.38)$$

■

Capítulo 8

Conclusiones

Teorema 8.1 *La perezosa caminata aleatoria en el hipercubo n -dimensional tiene un cutoff en $\frac{1}{2}n \log n$ con una ventana de tamaño n .*

Demostración

Por lo tanto, por la Proposición 7.5

$$\lim_{\alpha \rightarrow \infty} \limsup_{n \rightarrow \infty} \left[d_n \left(\left(\frac{1}{2} \right) n \log n + \alpha n \right) \right] = 0. \quad (8.1)$$

Por la Proposición 7.4

$$\lim_{\alpha \rightarrow -\infty} \limsup_{n \rightarrow \infty} \left[d_n \left(\left(\frac{1}{2} \right) n \log n - \alpha n \right) \right] = 1. \quad (8.2)$$

Por lo tanto, el hipercubo n -dimensional tiene un cutoff de ventana.

■

Finalmente, utilizando muchas técnicas interesantes, logramos probar la existencia del límite para nuestra caminata aleatoria. Hemos visto que la distribución estacionaria de la cadena es la medida uniforme sobre todos los vértices del hipercubo y que para estar lo suficientemente cerca de esta medida tenemos que correr la cadena durante aproximadamente $\left(\frac{1}{2}\right) n \log n$ pasos, ese es el orden correcto para t_{mix} . Si esperamos menos de este tiempo, todavía somos capaces de reconocer de alguna manera desde qué posición partimos y la distribución uniforme no está bien aproximada, mientras que esperar más es bastante inútil. En particular, estos 'menos' y 'más' son cuantificables: están descritos exactamente por el tamaño de la ventana $O(n)$.

Capítulo 9

Apéndice: Prueba Alternativa del Cutoff en el Hipercubo

Por último, daremos una prueba del *cutoff* en el hipercubo n -dimensional, pero desde un punto de vista muy distinto, ahora ocuparemos una de las ramas importantes de las matemáticas, álgebra lineal, la cual nos ayudará a explicar cuándo ocurre el *cutoff*, por lo cual analizaremos los valores propios y las funciones propias (o vectores propios) del núcleo de transición de la cadena de Markov, así podremos acotar la distancia de variación total y los tiempos de mezcla.

Retomando la definición de distancia de separación (6.6)

$$s(t) := \max_{x \in E} s_x(t), \quad \text{donde} \quad s_x(t) := \max_{y \in E} \left[1 - \frac{P_x(X_t = y)}{\pi(y)} \right]. \quad (9.1)$$

Por los Lemas 6.2 y 6.3 tenemos

$$\|P_x(X_t \in \cdot) - \pi(\cdot)\|_{TV} \leq s_x(t) \leq P_x(\tau > t). \quad (9.2)$$

Sea $\langle \cdot, \cdot \rangle$ que denota el producto interno en \mathbb{R}^E , el cual está dado por $\langle f, g \rangle = \sum_{x \in E} f(x)g(x)$. Al estar trabajando con cadenas de Markov con distribución estacionaria, es necesario considerar un producto interno ponderado en $\mathbb{R}^{|E|}$ dado por

$$\langle f, g \rangle_\pi := \sum_{x \in E} f(x)g(x)\pi(x). \quad (9.3)$$

Lema 9.1 *Dada $(X_t)_{t \geq 0}$ una cadena de Markov irreducible y reversible con el núcleo de transición P , se cumple:*

1. *El espacio con producto interior $(\mathbb{R}^{|E|}, \langle \cdot, \cdot \rangle_\pi)$ tiene una base ortonormal de funciones propias de valor real $\{f_j\}_{j=1}^n$ correspondientes a valores propios reales $\{\lambda_j\}_{j=1}^n$, con $n = |E|$, ordenados como $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$.*

2. La matriz P puede descomponerse como

$$\frac{P^t(x, y)}{\pi(y)} = \frac{P_x(X_t = y)}{\pi(y)} = \sum_{j=1}^n f_j(x) f_j(y) \lambda_j^t.$$

3. La función propia f_1 correspondiente al valor propio $\lambda_1 = 1$ puede tomarse como el vector constante 1, de modo que

$$\frac{P^t(x, y)}{\pi(y)} = \frac{P_x(X_t = y)}{\pi(y)} = 1 + \sum_{j=2}^n f_j(x) f_j(y) \lambda_j^t.$$

Demostración

1. Dada una cadena reversible de Markov X_t , $t \geq 0$ reversible con núcleo de transición P . Sea D_π una matriz diagonal tal que $D_\pi(x, x) = \pi(x)$ y $D_\pi(x, y) = 0$ si $x \neq y$, y tomando $A = D_\pi^{\frac{1}{2}} P D_\pi^{-\frac{1}{2}}$, tal que $A(x, y) = \sqrt{\pi(x)} \sqrt{\pi(y)} P(x, y)$ para todo $x, y \in E$. Entonces (X_t) al ser reversible tenemos

$$\begin{aligned} A(x, y) &= \frac{\sqrt{\pi(x)}}{\sqrt{\pi(y)}} P(x, y) = \frac{\pi(x)}{\sqrt{\pi(x)} \pi(y)} P(x, y) \\ &= \frac{\pi(y)}{\sqrt{\pi(x)} \pi(y)} P(y, x) = \frac{\sqrt{\pi(y)}}{\sqrt{\pi(x)}} P(y, x) = A(y, x). \end{aligned}$$

Dado que A es una matriz simétrica y por la teoría espectral para matrices simétricas sabemos que A tiene funciones propias de valor real, digamos $\{\varphi_j\}_{j=1}^n$, correspondiente a valores propios de valor real $\{\lambda_j\}_{j=1}^n$, que forman una base ortonormal, sea B el espacio con productora interna $(R^{|E|}, \langle \cdot, \cdot \rangle)$. Tomando $f_j = D_\pi^{-\frac{1}{2}} \varphi_j$ para $j=1, \dots, n$. Entonces

$$P f_j = P D_\pi^{-\frac{1}{2}} \varphi_j = D_\pi^{-\frac{1}{2}} A D_\pi^{\frac{1}{2}} D_\pi^{-\frac{1}{2}} \varphi_j = D_\pi^{-\frac{1}{2}} A \varphi_j = \lambda_j D_\pi^{-\frac{1}{2}} \varphi_j = \lambda_j f_j.$$

Sea $\delta_{i,j} : B \times B \rightarrow \{0, 1\}$ que denota la función delta de Dirac, es decir, $\delta_{i,j}(\varphi_i, \varphi_j) = 1$ si y solo si $i=j$. Dado que las funciones propias originales de A son ortonormales en $(R^{|E|}, \langle \cdot, \cdot \rangle)$, tenemos

$$\begin{aligned} \delta_{i,j}(\varphi_i, \varphi_j) &= \langle \varphi_i, \varphi_j \rangle = \langle D_\pi^{\frac{1}{2}} f_i, D_\pi^{\frac{1}{2}} f_j \rangle \\ &= \sum_{x \in E} \left(\sqrt{\pi(x)} f_i(x) \sqrt{\pi(x)} f_j(x) \right) \\ &= \sum_{x \in E} (f_i(x) f_j(x) \pi(x)) = \langle f_i, f_j \rangle_\pi. \end{aligned}$$

Así, el espacio con producto interior $(R^{|E|}, \langle \cdot, \cdot \rangle_\pi)$ tiene una base ortonormal de valor real $\{f_j\}_{j=1}^n$ con valores propios correspondientes $\{\lambda_j\}_{j=1}^n$.

2. Observemos que $\sqrt{\pi}$ es una función propia de A con el valor propio correspondiente $\lambda_1 = 1$, ya que

$$\begin{aligned} (\sqrt{\pi}A)(x) &= \sum_{y \in E} \sqrt{\pi(y)} \frac{\sqrt{\pi(y)}}{\sqrt{\pi(x)}} P(y, x) = \sum_{y \in E} \frac{\pi(y)}{\sqrt{\pi(y)}} P(y, x) \\ &= \sum_{y \in E} \sqrt{\pi(x)} P(x, y) = \sqrt{\pi(x)}. \end{aligned}$$

Esto implica que $D_\pi^{-\frac{1}{2}} \sqrt{\pi} = 1$ es un vector propio de P . Sea δ_y la función definida como

$$\delta_y(x) = \begin{cases} 1, & \text{si } x = y, \\ 0, & \text{si } x \neq y. \end{cases}$$

En el espacio con producto interior, $(R^{|E|}, \langle \cdot, \cdot \rangle_\pi)$ con bases ortonormales $\{f_j\}_{j=1}^n$, podemos descomponer esta función de la siguiente manera:

$$\delta_y = \sum_{j=1}^n (\langle \delta_y, f_j \rangle_\pi f_j) = \sum_{j=1}^n \left[\sum_{x \in E} (\delta_y f_j \pi(x)) \right] f_j = \sum_{j=1}^n \pi(y) f_j(y) f_j.$$

Usando este resultado, obtenemos

$$P^t(x, y) = (P^t \delta_y)(x) = \left(\sum_{j=1}^n \pi(y) f_j(y) P^t f_j \right)(x) = \sum_{j=1}^n \pi(y) f_j(y) \lambda_j^t f_j(x),$$

de donde obtenemos

$$\frac{P^t(x, y)}{\pi(y)} = \sum_{j=1}^n f_j(x) f_j(y) \lambda_j^t.$$

3. Dado que $f_1 = 1$, tenemos

$$\frac{P^t(x, y)}{\pi(y)} = 1 + \sum_{j=2}^{|E|} f_j(x) f_j(y) \lambda_j^t.$$

■

Definición 9.1 Para $p \geq 0$, la norma $l^p(\pi)$ en $R^{|E|}$ se define como

$$\|f\|_{\pi, p} := \left[\sum_{x \in E} |f(x)|^p \right]^{\frac{1}{p}}.$$

Para $p = 2$, $\|f\|_2 = \sqrt{\langle f, f \rangle_\pi}$.

Teorema 9.1 Sea $(X_t)_{t \geq 0}$ con un núcleo de transición reversible P con valores propios $\lambda_1 = 1 \geq \lambda_2 \geq \dots \geq -1$ correspondiente a las funciones propias $\{f_j\}_{j=1}^n$ y ortonormal con respecto al producto interno $\langle \cdot, \cdot \rangle_\pi$. Entonces

$$4\|P_x(X_t \in \cdot) - \pi(\cdot)\|_{TV}^2 \leq \left\| \frac{P_x(X_t \in \cdot)}{\pi(\cdot)} - 1 \right\|_2^2 = \sum_{j=2}^{|E|} f_j(x)^2 \lambda_j^{2t}.$$

Además, si la cadena (X_t) es transitiva, entonces tenemos

$$4\|P_x(X_t \in \cdot) - \pi(\cdot)\|_{TV}^2 \leq \sum_{j=2}^{|E|} \lambda_j^{2t}.$$

Demostración

Usando el Lema 9.1 tenemos

$$\begin{aligned} \left\| \frac{P_x(X_t \in \cdot)}{\pi(\cdot)} - 1 \right\|_2^2 &= \left\| \sum_{j=2}^{|E|} (\lambda_j^t f_j(x) f_j) \right\|_2^2 \\ &= \left\langle \sum_{j=2}^{|E|} \lambda_j^t f_j(x) f_j, \sum_{j=2}^{|E|} \lambda_j^t f_j(x) f_j \right\rangle_\pi \\ &= \sum_{j=2}^{|E|} f_j(x)^2 \lambda_j^{2t}. \end{aligned} \tag{9.4}$$

Además, usando el hecho de que $\|\mu - \nu\|_{TV} = \frac{1}{2} \sum_{x \in E} |\mu(x) - \nu(x)|$ tenemos

$$\begin{aligned} \|P_x(X_t \in \cdot) - \pi(x)\|_{TV} &= \frac{1}{2} \sum_{x \in E} \left(\left| \frac{P_x(X_t = y)}{\pi(y)} - 1 \right| \pi(y) \right) \\ &= \frac{1}{2} \left\| \frac{P_x(X_t \in \cdot)}{\pi(\cdot)} - 1 \right\|_1. \end{aligned}$$

Como $\|p\|_f$ es no decreciente, obtenemos

$$4\|P_x(X_t \in \cdot) - \pi(x)\|_{TV}^2 = \left\| \frac{P_x(X_t \in \cdot)}{\pi(\cdot)} - 1 \right\|_1^2 \leq \left\| \frac{P_x(X_t \in \cdot)}{\pi(\cdot)} - 1 \right\|_2^2$$

Si la cadena también es transitiva, entonces para cada par $(x, y) \in E^2$ existe una biyección $\varphi_{(x,y)} : E \rightarrow E$, tal que $\varphi(x) = y$ y $P(z, w) = P(\varphi(z), \varphi(w))$ para todo $z, w \in E$. Como la cadena también es reversible, tenemos que π es uniforme en E . Por lo tanto, el lado izquierdo de (9.4) no depende de x . Entonces para cualquier $x_0 \in E$ tenemos

$$\left\| \frac{P_{x_0}(X_t \in \cdot)}{\pi(\cdot)} - 1 \right\|_2^2 = \sum_{j=2}^{|E|} f_j(x)^2 \lambda_j^{2t}.$$

Sumando sobre x en ambos lados de la igualdad tenemos

$$\begin{aligned} |E| \left\| \frac{P_{x_0}(X_t \in \cdot)}{\pi(\cdot)} - 1 \right\|_2^2 &= |E| \sum_{x \in E} \left(\pi(x) \sum_{j=2}^{|E|} f_j(x)^2 \lambda_j^{2t} \right) \\ &= |E| \sum_{j=2}^{|E|} \left(\sum_{x \in E} f_j(x)^2 \pi(x) \right) \lambda_j^{2t} = |E| \sum_{j=2}^{|E|} \lambda_j^{2t}. \end{aligned}$$

La última igualdad se cumple dividiendo por $|E|$. ■

Teorema 9.2 *Para la caminata aleatoria simple en el hipercubo n -dimensional*

$$1 - 4e^{-2\alpha+1} \leq d \left(\frac{1}{2} n \log n + \alpha n \right) \leq \frac{1}{\sqrt{2}} e^{-\alpha}. \quad (9.5)$$

Demostración

El lado inferior se cumple por la Proposición 7.4. Por otro lado, para la cota superior, veremos los valores y funciones propias de la caminata aleatoria en el hipercubo n -dimensional. Para este propósito, veremos el producto de n cadenas de Markov irreducibles, para $j = 1, 2, \dots, n$. Sea P_j una matriz de transición irreducible en el espacio de estado \mathbb{E}_j , con distribución estacionaria π_j . La cadena selecciona una coordenada i de acuerdo con alguna distribución w en $\{1, \dots, n\}$ y luego solo se mueve de acuerdo a P_i en la coordenada i . Sea $X = (x_1, x_2, \dots, x_n) \in E$. Entonces la matriz de transición \tilde{P} de esta cadena se define como

$$\tilde{P}(x, y) := \sum_{j=1}^n P_j(x_j, y_j) \prod_{i=1, i \neq j}^n 1_{\{x_i = y_i\}}. \quad (9.6)$$

Para lo cual enunciaremos el siguiente lema.

Lema 9.2 *Para $j = 1, \dots, n$, sea P_j una matriz de transición irreducible en el espacio de estado \mathbb{E}_j , con distribución estacionaria π_j . Sea \tilde{P} el kernel de transición en $\tilde{E} := E_1 \times E_2 \times \dots \times E_n$. Entonces*

1. La función $\varphi := \varphi^1 \otimes \dots \otimes \varphi^n$ es una función propia de la matriz de transición \tilde{P} , con valores propios $\sum_{j=1}^n w_j \lambda^{(j)}$.
2. Supongamos que para cada j , B_j es una base ortogonal en $l^2(\pi_j)$. Entonces la colección

$$\tilde{B} := \left\{ \bigotimes_{j=1}^n \varphi^{(j)} : \varphi^{(j)} \in B_j; j = 1, \dots, n \right\}, \quad (9.7)$$

es una base para $l^2(\pi_1 \times \dots \times \pi_n)$.

Ahora comenzaremos a observar la caminata aleatoria perezosa en el hipercubo n -dimensional y utilizaremos Lema 9.2 para determinar sus funciones propias y sus valores propios. Para cada $j = 1, \dots, n$ tenemos que el *kernel* de transición esta dado por

$$P_j = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Realizando los cálculos para valores propios tenemos $\lambda_0^{(j)} = 1$ y $\lambda_1^{(j)} = 0$ con las funciones propias correspondientes $\lambda_0^{(j)} = \left(\frac{1}{\sqrt{2}}\right)(1, 1)^t$ y $\lambda_1^{(j)} = \left(\frac{1}{\sqrt{2}}\right)(1, -1)^t$. Por el Lema 9.2 Está claro que la caminata aleatoria en el hipercubo n -dimensional tiene 2^n valores propios y funciones propias. Que puede ser etiquetado por $J \in \mathbb{Z}_2^n$ como

$$\lambda_J = \sum_{i=1}^n \pi(i) \lambda_{J_i}^{(i)} = \frac{1}{n} \sum_{i=1}^n 1_{\{J_i=0\}} = \frac{n - W(J)}{n} \quad y \quad \varphi_J = \bigotimes_{i=1}^n \varphi_{J_i}^{(i)}.$$

Notemos que las funciones propias φ_j corresponden a las columnas J de la matriz Hadamard H_{2^n} y así formar una base ortonormal para \tilde{P} . Retomando la demostración, la caminata aleatoria en el hipercubo n -dimensional es transitiva. Usando el Teorema 9.1 con los valores propios que hemos encontrado anteriormente, por lo tanto, tenemos

$$4 \|P_0(X_t \in \cdot) - \pi(x)\|_{TV}^2 \leq \sum_{J \in \mathbb{Z}_2^n \setminus \{0\}} \lambda_J^2 = \sum_{j=1}^n \left(1 - \frac{j}{n}\right)^{2t} \binom{n}{j} \leq \sum_{j=1}^n e^{-2jtn} \binom{n}{j}$$

Tomando $t = \frac{1}{2} n \log n + \alpha n$, tenemos

$$4 \|P_0(X_t \in \cdot) - \pi(x)\|_{TV}^2 \leq \left(1 - \frac{1}{n} e^{-2\alpha}\right) - 1 \leq e^{e^{-2\alpha}} - 1 \leq 2e^{-2\alpha}.$$

Por lo tanto, dividiendo por 4 y tomando la raíz cuadrada en ambos lados, obtenemos

$$\|P_0(X_t \in \cdot) - \pi(x)\|_{TV} \leq \frac{1}{\sqrt{2}} e^{-\alpha}.$$

■

Bibliografía

- [1] David Aldous. Random walks on finite groups and rapidly mixing markov chains. *In Séminaire de Probabilités XVII 1981/82.*, 1983.
- [2] Nathanaël Berestycki, Eyal Lubetzky, Yuval Peres, and Allan Sly. Random walks on the random graph. *The Annals of Probability*, 46:456–490, 01 2018.
- [3] Caputo P. Salez J. Bordenave, C. Random walk on sparse random digraphs. *Probability Theory and Related Fields*, 2018.
- [4] Aldous David and Diaconis Persi. Shuffling cards and stopping times. *The American Mathematical Monthly.*, 1986.
- [5] P Diaconis. The cutoff phenomenon in finite markov chains. *Proceedings of the National Academy of Sciences*, 93(4):1659–1664, 1996.
- [6] Lubetzky Eyal and Sly Allan. Information percolation for the ising model: cutoff in three dimensions up to criticality. Technical report, Technical report, Citeseer.
- [7] Lubetzky Eyal and Sly Allan. Cutoff phenomena for random walks on random regular graphs. *Duke Mathematical Journal*, 2010.
- [8] Allan Sly. Eyal Lubetzky. Information percolation and cutoff for the stochastic ising model. *Journal of the American Mathematical Society*, 2016.
- [9] S.H. Friedberg, A.J. Insel, and L.E. Spence. *Linear Algebra*. Featured Titles for Linear Algebra (Advanced) Series. Pearson Education, 2003.
- [10] Lacoïn Hubert. Mixing time and cutoff for the adjacent transposition shuffle and the simple exclusion. *The Annals of Probability*, 2016.
- [11] Miller Jason and Peres Yuval. Uniformity of the uncovered set of random walk and cutoff for lamplighter chains. *The Annals of Probability*, 2012.
- [12] Cyril Labbé and Hubert Lacoïn. Cutoff phenomenon for the asymmetric simple exclusion process and the biased card shuffling. *The Annals of Probability*, 47, 10 2016.
- [13] D.A. Levin, Y. Peres, and E.L. Wilmer. *Markov Chains and Mixing Times*. American Mathematical Soc.

- [14] Eyal Lubetzky and Allan Sly. Universality of cutoff for the ising model. *The Annals of Probability*, 45, 07 2014.
- [15] A.M.F. Mood and F.A. Graybill. *Introduction to the Theory of Statistics*. International Student Edition: McGraw-Hill Series in Probability and Statistics. McGraw-Hill, 1963.
- [16] Harm Nooitgedagt. Two convergence limits of markov chains: Cutoff and metastability. Master's thesis, Mathematisch Instituut, Universiteit Leiden, 2010.
- [17] Diaconis Persi and Shahshahani Mehrdad. Generating a random permutation with random transpositions. *Probability Theory and Related Fields*, 1981.