

INSTITUTO TECNOLÓGICO AUTÓNOMO DE MÉXICO



**Tiempos de mezcla en cadenas de
Markov: aplicación al problema del
barajeo de cartas**

TESIS

QUE PARA OBTENER EL TÍTULO DE

LICENCIADO EN ACTUARÍA

PRESENTA

ALBERTO SEBASTIÁN CHÁVEZ CANO

ASESORA: DRA. AIRAM ASERET BLANCAS BENÍTEZ

«Con fundamento en los artículos 21 y 27 de la Ley Federal del Derecho de Autor y como titular de los derechos moral y patrimonial de la obra titulada “**Tiempos de mezcla en cadenas de Markov: aplicación al problema del barajeo de cartas**”, otorgo de manera gratuita y permanente al Instituto Tecnológico Autónomo de México y a la Biblioteca Raúl Baillères Jr., la autorización para que fijen la obra en cualquier medio, incluido el electrónico, y la divulguen entre sus usuarios, profesores, estudiantes o terceras personas, sin que pueda percibir por tal divulgación una contraprestación.»

ALBERTO SEBASTIÁN CHÁVEZ CANO

FECHA

FIRMA

Agradecimientos

Me resulta grato reconocer el gran trabajo y la profunda influencia que han ejercido en mí los profesores de los que he aprendido matemáticas. Tan temprano como en la preparatoria el Ingeniero Eduardo Islas no solo me brindó un sólido cimiento sobre el cual desarrollar mis estudios universitarios, también, y aún más importante, me impulsó a descubrir las matemáticas como una de las pasiones de mi vida.

Agradezco a todos los profesores que he tenido a lo largo de mi estancia en el ITAM (sería imposible enlistarlos individualmente en esta sección) pues me han dado las herramientas necesarias para desarrollarme satisfactoriamente en el mundo laboral, así como en mis planes para estudiar un posgrado. Sin duda han sido cruciales en mi desarrollo personal.

Reconozco la valiosa guía que ha supuesto la Dra. Airam Blancas en esta última parte de mi carrera, sin ella esta tesis no hubiera sido posible.

A los amigos de siempre y a los que encontré durante la carrera les agradezco, han hecho de esta experiencia una más grata. Dedico esta última parte, al igual que todos mis logros, a mi familia que son el motor de mi vida.

A mi madre

Resumen

La presente tesis es una investigación en el área de procesos estocásticos a tiempo discreto. Más precisamente estudiamos el tiempo requerido para barajar adecuadamente una pila de n cartas, para ello caracterizamos la acción de barajar cartas como una caminata aleatoria en el grupo simétrico. Introducimos la distancia de variación total como una manera de medir la distancia entre dos distribuciones de probabilidad. Probamos propiedades de ella que permiten que sea razonable emplearla como métrica para cuantificar la discrepancia entre la distribución de la cadena de Markov ergódica y su distribución estacionaria a un tiempo t dado. Así mismo presentamos diversas técnicas y conceptos, con un énfasis especial en los tiempos fuertemente estacionarios, para investigar el tiempo de convergencia de la caminata aleatoria en el grupo simétrico. Con ello establecemos cotas al tiempo requerido para barajar correctamente si barajamos mediante transposición aleatoria y *riffle shuffle*.

Concluimos el trabajo complementando el análisis asintótico con un estudio exacto del tiempo de convergencia, posible en el caso del *riffle shuffle*, donde probamos la existencia del *cutoff phenomenon* y dejamos abierta la puerta para un estudio general de este fenómeno que supone actualmente una de las áreas más activas de investigación.

Índice general

Introducción	1
1. Preliminares de cadenas de Markov	4
1.1. Cadenas de Markov finitas	4
1.2. Representación canónica	13
1.3. Cadenas reversibles	16
1.4. Tiempos fuertemente estacionarios	19
1.5. Distancia de variación total	25
2. Barajar	50
2.1. Caminata aleatoria en grupos	50
2.2. Tiempo reverso y tiempos de mezcla	57
2.3. Caminata aleatoria en el grupo simétrico	60
2.3.1. Irreducibilidad	65
2.3.2. Periodicidad	69
2.3.3. Tiempo de mezcla	70
2.4. Barajar genes	83
3. Barajeo rápido	86
3.1. El modelo de barajeo rápido	86
3.2. Tiempo de mezcla	95

3.3. Resultados de Bayer y Diaconis 1992	99
Conclusiones	108
Bibliografía	112

Introducción

Markov (1906) y Poincaré (1912) en las primeras investigaciones sobre cadenas de Markov usaron como ejemplo principal barajar una pila de cartas. Entonces no debe sorprendernos que estudiar este proceso estocástico nos conduzca a un vasto campo de ideas matemáticas marcadas por el ingenio de quienes las desarrollaron.

En la presente tesis nos adentramos en este proceso motivados por dar respuesta a la pregunta *¿cuántas veces tenemos que barajar una pila de cartas para que esté bien aleatorizada?* Por supuesto, existen muchas maneras en las que podemos barajar una pila de cartas. Nosotros nos ocuparemos de dos: el barajeo por transposición aleatoria y el *riffle shuffle* que llamaremos barajeo rápido. El interés del primero radica en que es un primer acercamiento al tema, no obstante con algunas modificaciones sirve como modelo matemático para un problema en biología sobre la mutación del ADN. El segundo es un modelo adecuado para la forma de barajar cartas en la vida real.

Darnos a la tarea de resolver esta pregunta no se trata simplemente de un capricho matemático. Investigar esta cuestión impulsó el desarrollo de la teoría de los tiempos de mezcla dentro de las cadenas de Markov. Del mismo modo tiene implicaciones en varias áreas de las matemáticas que a primera vista parecen distantes a los procesos

estocásticos. A saber, Richard Stanley [24] expuso una íntima conexión con la teoría de funciones simétricas en el área de combinatoria. Por otro lado, Jason Fulman [13] descubrió una interesante intersección entre el barajeo de cartas y la teoría de grupos finitos de Lie. Así mismo se forma naturalmente un puente hacia la teoría de la representación del grupo simétrico. Por supuesto, existen más caminos que llevan a otras áreas de las matemáticas. Una lista exhaustiva con sus respectivas justificaciones se puede encontrar en [10].

Dando un paso fuera de las matemáticas, muchos trucos de magia son posibles gracias a que toman ventaja de una pila de cartas que no ha sido bien barajada. Quizá el más sorprendente de estos es “Premo” descubierto por Charles Jordan hace más de 100 años. Quién realiza el truco presenta una pila de cartas a los espectadores y se da la vuelta. Pide que hagan dos barajeos rápidos y tomen la carta de hasta arriba, anoten la carta y vuelvan a insertarla en el paquete. Un barajeo rápido final y entonces las cartas se extienden en forma de arco sobre la mesa. Finalmente quién hace el truco se da la vuelta, examina intensivamente y adivina la carta que los espectadores anotaron. Este hecho fantástico deja de manifiesto que tras tres barajeos existe una estructura explotable debido al orden que todavía no ha sido destruido. Aquí un motivo de sobra para que los casinos presten especial atención a este tema.

Existen distintas técnicas para abordar este problema, estas van desde el acoplamiento hasta el estudio de los eigenvalores de la matriz de transición. Nuestro enfoque principal serán los tiempos fuertemente estacionarios. La investigación la hemos dispuesto en 3 secciones correspondientes a los capítulos del presente trabajo. En el Capítulo 1 repasamos nociones básicas de las cadenas de Markov finitas e introducimos la idea que nos permitirá medir la distancia entre la cadena y su equilibrio. En el Capítulo 2 desarrollamos ideas

fundamentales de la teoría de grupos del álgebra abstracta con las cuales podemos pensar de forma matemática el proceso de barajar una pila de n cartas. Conectamos estas ideas con las desarrolladas en el Capítulo 1 y damos una respuesta en el modelo de transposición aleatoria. Finalmente en el Capítulo 3 estudiamos el tiempo de mezcla para el barajeo rápido y discutimos brevemente el transcendental artículo de Diaconis y Bayer “*Trailing the dovetail shuffle to its lair*” [5].

Capítulo 1

Preliminares de cadenas de Markov

1.1. Cadenas de Markov finitas

Una cadena de Markov a tiempo discreto es una sucesión de variables aleatorias $(X_t)_{t=0}^{\infty}$ que toman valores en un espacio de estados Ω (finito o contable) donde se cumple la propiedad de Markov. Esta establece heurísticamente que dado el presente, el futuro no depende del pasado. Formalmente para $t \geq 0$ se cumple lo siguiente:

$$\mathbb{P}(\underbrace{X_{t+1} = y}_{\text{futuro}} \mid \underbrace{X_t = x}_{\text{presente}}, \underbrace{X_{t-1} = x_{t-1}, \dots, X_0 = x_0}_{\text{pasado}}) = \mathbb{P}(X_{t+1} = y \mid X_t = x). \quad (1.1)$$

Observemos que la identidad es válida para cualquier t , la probabilidad es la misma independientemente del tiempo que consideremos. Esto motiva a que (1.1) sea conocida simplemente como la probabilidad de transición del estado x al y que denotaremos por $P(x, y)$. Los procesos que poseen esta característica son llamados

cadenas de Markov homogéneas. Todas las cadenas que consideremos, de ahora en adelante, son de este tipo y a menos que se especifique lo contrario el espacio de estados se supone finito.

El total de estados posibles del proceso se corresponde con la cardinalidad del conjunto Ω . Luego existe un total de $|\Omega|^2$ probabilidades de transición en un paso. Toda esta información se organiza en una matriz P de dimensión $|\Omega| \times |\Omega|$ tal que la entrada (x, y) es igual a $P(x, y)$, con $P(x, y)$ definido como en (1.1). De manera más general, dado que el estado actual es x el siguiente estado se elige acorde a la distribución $P(x, \cdot)$, es decir la fila x de la matriz P . Es natural llamar a P matriz de transición en un paso, y para ella se cumple

$$\sum_{y \in \Omega} P(x, y) = 1, \quad \text{para } x \in \Omega.$$

Las matrices que verifican esta propiedad son conocidas como matrices estocásticas.

La exposición de teoría siempre es más efectiva cuando va de la mano con ejemplos que ayuden a ilustrar los conceptos. Nosotros usaremos uno sencillo, pero efectivo, para acompañar la discusión.

Imaginemos una rana que vive felizmente en un estanque con dos lirios: el de la izquierda y el de la derecha que abreviaremos por i y d respectivamente. En cada uno de estos lirios se encuentra una moneda distinta. La actividad favorita de la rana es pasar todo el día en alguno de los lirios, ella decide al amanecer en cuál según la siguiente dinámica: lanza la moneda del lirio en el que se encuentra, si cae cara entonces saltará al otro lirio, si cae cruz entonces permanecerá donde está. La sucesión de lirios que la rana va ocupando según la dinámica descrita es una cadena de Markov (X_t) , con espacio de estados compuesto por dos elementos $\Omega = \{i, d\}$. Hemos advertido que las monedas sobre los lirios

son distintas, por lo que vamos a suponer que se tratan de monedas que no son justas. La moneda del lirio i tiene probabilidad p de caer cara mientras que la del lirio d tiene probabilidad q . Supongamos que p y q son positivas y ambas menor que 1. Así la matriz de transición P está descrita por

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

Supongamos que el proceso inicia en lunes y la rana pasa el día en el lirio i . La probabilidad de que la rana transcurra el martes en el otro lirio es p , por el contrario permanece en el lirio i con probabilidad $1 - p$. De manera más formal tenemos

$$\mathbb{P}(X_1 = i | X_0 = i) = 1 - p, \quad \mathbb{P}(X_1 = d | X_0 = i) = p. \quad (1.2)$$

Como habíamos señalado antes esto es la primera fila de P . La cuestión interesante es estudiar lo que sucede a medida que transcurren los días. La probabilidad de que la rana se encuentre en el lirio i el miércoles es

$$\begin{aligned} \mathbb{P}(X_2 = i | X_0 = i) &= \mathbb{P}(X_2 = i | X_1 = d) \mathbb{P}(X_1 = d | X_0 = i) \\ &\quad + \mathbb{P}(X_2 = i | X_1 = i) \mathbb{P}(X_1 = i | X_0 = i) \\ &= qp + (1 - p)(1 - p). \end{aligned} \quad (1.3)$$

Mientras que la probabilidad para el lirio d está dada por

$$\begin{aligned} \mathbb{P}(X_2 = d | X_0 = i) &= \mathbb{P}(X_2 = d | X_1 = d) \mathbb{P}(X_1 = d | X_0 = i) \\ &\quad + \mathbb{P}(X_2 = d | X_1 = i) \mathbb{P}(X_1 = i | X_0 = i) \\ &= (1 - q)p + p(1 - p). \end{aligned} \quad (1.4)$$

Es posible seguir avanzando en el tiempo de la manera en la que lo hemos hecho. Sin embargo, resulta en un trabajo largo y fatigoso que

podemos sortear mediante una aproximación sistemática apoyándonos en el álgebra matricial. En nuestro ejemplo, la suposición de que en el día lunes la rana está en el lirio i se puede expresar de la siguiente manera

$$\mu_0 = (1, 0).$$

Se trata de un vector fila con las probabilidades que rigen a X_0 . Esta función de distribución inicial (razón del subíndice empleado) asigna toda la probabilidad a i . De esta manera, la distribución condicional que se encontró en las ecuaciones (1.2) se resumen en la operación

$$\mu_1 = (\mathbb{P}(X_1 = i|X_0 = i), \mathbb{P}(X_1 = d|X_0 = i)) = \mu_0 P.$$

Mientras que las probabilidades en (1.3) y (1.4) se pueden resumir como

$$\mu_2 = \mu_0 P^2 = \mu_0 P P = \mu_1 P.$$

La utilidad de esta aproximación nos motiva a generalizar este método para cualquier cadena de Markov y para cualquier tiempo t . Con esta intención es conveniente definir el siguiente vector fila

$$\mu_t(x) = \mathbb{P}(X_t = x), \quad \text{para } x \in \Omega. \quad (1.5)$$

Derivamos las probabilidades para un tiempo adelante

$$\begin{aligned} \mu_{t+1}(y) &= \sum_{x \in \Omega} \mathbb{P}(X_{t+1} = y, X_t = x) \\ &= \sum_{x \in \Omega} \mathbb{P}(X_{t+1} = y|X_t = x) \mathbb{P}(X_t = x) \\ &= \sum_{x \in \Omega} \mathbb{P}(X_t = x) P(x, y) \quad \text{para } y \in \Omega. \end{aligned}$$

Observemos que esto es la multiplicación del vector fila μ_t con la matriz de transición P

$$\mu_{t+1} = \mu_t P \quad \text{para } t \geq 0, \quad (1.6)$$

y por lo tanto

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

Donde μ_0 es la distribución inicial del proceso estocástico y P^t es t veces el producto de P consigo mismo. Denotaremos por \mathbb{P}_μ y \mathbb{E}_μ a la probabilidad y esperanza suponiendo que $\mu_0 = \mu$. Lo más habitual es que las distribuciones iniciales estén concentradas en un estado x . Matemáticamente esta situación se representa con la distribución δ_x

$$\delta_x(y) = \begin{cases} 1 & \text{si } y = x \\ 0 & \text{si } y \neq x \end{cases}$$

Este desarrollo junto con la identidad (1.6) nos permiten establecer el siguiente hecho para cualquier $t \geq 1$

$$\mathbb{P}(X_t = y | X_0 = x) = \mathbb{P}_x(X_t = y) = (\delta_x P^t)(y) = P^t(x, y). \quad (1.7)$$

Por esta razón a la matriz P^t que contiene todas estas probabilidades se le llama matriz de transición en t pasos. Una noción fundamental que debemos concluir de esta exposición se deriva de la ecuación (1.6): multiplicar la distribución μ_t por la matriz de transición P por la derecha equivale a avanzar la distribución un paso.

Nos encontramos con uno de los temas medulares de la teoría de los procesos estocásticos: estudiar el comportamiento de μ_t en el largo plazo. Resulta natural pensar en primera instancia que μ_t va a ir cambiando conforme avanzamos en el tiempo, pero ¿esto es así siempre? Si la sucesión de las distribuciones (μ_t) converge a alguna distribución π , la identidad (1.6) nos dice que dicha distribución de equilibrio debe cumplir

$$\pi = \pi P. \quad (1.8)$$

Esta distribución recibe el nombre de **distribución estacionaria** y se denota usualmente por π .

Ilustremos este importante concepto en nuestro ejemplo. La igualdad que debe cumplir π se traduce en la siguiente ecuación

$$\pi = (a, b) = (a, b) \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}.$$

Debido a que p y q son mayores a 0 escribimos

$$a(1-p) + bq = a \iff a = \frac{bq}{p}.$$

No obstante a y b deben estar sujetos a que $a+b=1$ dado que π es una función de distribución. Combinando ambas ecuaciones, obtenemos

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

Esta es la distribución estacionaria para la cadena. Recordemos que en nuestro ejemplo $\mu_t = \mathbb{P}_i(X_t)$. Demos un paso más e investiguemos si $\mu_t \rightarrow \pi$. Para ello definamos Δ_t como una forma rudimentaria de “medir” la discrepancia entre $\mu_t(i)$ y $\pi(i)$ a un tiempo t

$$\Delta_t := \mu_t(i) - \frac{q}{p+q}, \quad \text{para } t \geq 0.$$

Usando la identidad (1.6) y que μ_t es una distribución notamos que Δ_t cumple la siguiente recursión

$$\Delta_{t+1} = \mu_t(i)(1-p) + (1-\mu_t(i))q - \frac{q}{p+q} = (1-p-q)\Delta_t,$$

luego tenemos

$$\Delta_t = (1-p-q)^t \Delta_0, \quad \text{para } t \geq 0. \quad (1.9)$$

En el contexto del problema p y q son probabilidades de modo que tenemos $-1 < 1-p-q < 1$. Esto significa que $\Delta_t \rightarrow 0$ cuando $t \rightarrow \infty$ y podemos concluir

$$\lim_{t \rightarrow \infty} \mu_t(i) = \frac{q}{p+q} \quad \text{y} \quad \lim_{t \rightarrow \infty} \mu_t(d) = \frac{p}{p+q}. \quad (1.10)$$

La igualdad del lado derecho se obtiene al considerar que μ_t es una distribución definida en un espacio con dos elementos. La misma conclusión se obtiene si por el contrario definimos $\Delta_t := \mu_t(d) - p/(p+q)$. Con esto hemos mostrado que la distribución de nuestra cadena converge a la distribución estacionaria π . Más aún, la convergencia observa un decrecimiento exponencial. Esto no es casualidad y se hablará de ello más adelante.

Observemos que (1.9) nos permite concluir $\mu_t \rightarrow \pi$ sin importar la distribución inicial μ_0 .

Los pilares de nuestra investigación versan sobre la naturaleza de la distribución estacionaria π : ¿Cuándo existe?; ¿es única?; ¿cuándo podemos decir que la cadena converge? Antes de adentrarnos en estas cuestiones recordemos dos nociones esenciales sin las cuales es imposible desarrollar una teoría sobre cadenas de Markov. Decimos que una matriz de transición P es *irreducible* si

$$\forall (x, y) \in \Omega^2, \quad \exists t \in \mathbb{N}, \quad \text{tal que } P^t(x, y) > 0. \quad (1.11)$$

Esto significa que podemos conectar cualesquiera dos estados mediante transiciones de probabilidad positiva.

Por otro lado, definimos el periodo de un estado x como el máximo común divisor (mcd) del conjunto $\mathcal{T}(x) = \{t \geq 1 : P^t(x, x) > 0\}$. En el caso que este conjunto sea vacío definimos el periodo igual a 0. A las cadenas que poseen periodo 1 en todos sus estados se les llama *aperiódicas*.

Con la finalidad de resolver las inquietudes antes expuestas sobre la naturaleza de π presentamos los siguientes resultados fundamentales en la teoría de los procesos estocásticos. Las demostraciones pueden encontrarse en cualquier libro introductorio a los procesos estocásticos como [21] o [22].

Lema 1.1. *Cualquier matriz de transición P con espacio de estados finito Ω tiene una distribución estacionaria π que satisface (1.8). Si además P es una matriz irreducible la distribución π es única y su soporte es Ω .*

Para una matriz irreducible podemos entonces referirnos a la distribución estacionaria asociada. Sin embargo, este lema no indica que para esta matriz tengamos $\mu_t \rightarrow \pi$, fenómeno que sí se presentó en nuestro ejemplo. Para poder concluir la convergencia necesitamos una condición para P estrictamente más fuerte que (1.11). Esto se consigue considerando que P sea irreducible y también aperiódica.

Teorema 1.2. *Sea P una matriz de transición irreducible y aperiódica con distribución estacionaria π . Entonces*

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

Pedir que una matriz P sea irreducible y aperiódica equivale a intercambiar el orden del cuantificador existencial y universal en (1.11). Decimos que una matriz P es ergódica si

$$\exists t \in \mathbb{N}, \quad \text{tal que } \forall (x, y) \in \Omega^2, \quad P^t(x, y) > 0.$$

Proposición 1.1. *Una matriz P es irreducible y aperiódica si, y solo si, P es ergódica.*

Notemos que esta condición significa que existe una potencia de P tal que todas las entradas son positivas. Esta condición se mantiene para todas las potencias siguientes.

Demostración. Para la prueba necesitamos hacer uso de un hecho de teoría de números: un conjunto de enteros no negativos, cerrado bajo la suma y cuyo mcd es 1, contiene a todos los enteros no negativos salvo

una cantidad finita de ellos. Esto quiere decir que existe un entero k tal que todo $n \geq k$ pertenece al conjunto. La demostración de este hecho puede consultarse en [16] (Lema 1.27).

Supongamos que P es irreducible y aperiódica. Como la cadena es aperiódica entonces el mcd de $\mathcal{T}(x)$ es 1 para todo $x \in \Omega$. Así mismo $\mathcal{T}(x)$ es cerrado bajo la suma. En efecto si $s, t \in \mathcal{T}(x)$ entonces

$$P^{s+t}(x, x) \geq P^s(x, x)P^t(x, x) > 0,$$

por lo tanto $s + t \in \mathcal{T}(x)$. Esto significa que existe $t(x)$ tal que $t \geq t(x)$ implica $t \in \mathcal{T}(x)$. Por ser P irreducible sabemos que para toda $y \in \Omega$ existe $r(x, y) = r$ tal que $P^r(x, y) > 0$. Para $t \geq t(x) + r$ tenemos

$$P^t(x, y) \geq P^{t-r}(x, x)P^r(x, y) > 0, \quad (1.12)$$

debido a que $t - r \geq t(x)$. Definimos $t'(x) := t(x) + \max_{y \in \Omega} r(x, y)$. Para $t \geq t'(x)$ tenemos

$$P^t(x, y) > 0 \quad \text{para } y \in \Omega,$$

en un argumento similar al verificarse

$$t - r(x, y) \geq t(x) + \max_{y \in \Omega} r(x, y) - r(x, y) \geq t(x)$$

para toda $y \in \Omega$. Finalmente para $t \geq \max_{x \in \Omega} t'(x)$ y nuevamente por el mismo razonamiento

$$P^t(x, y) > 0 \quad \text{para toda } (x, y) \in \Omega^2.$$

Supongamos que P es ergódica, entonces existe un r tal que $P^r(x, y) > 0$ para toda $(x, y) \in \Omega^2$. Es inmediato que P es irreducible. Por otro lado

$$P^{r+1} = P^r P$$

tiene todas las entradas positivas. Esto significa que $r + 1 \in \mathcal{T}(x)$ para todo $x \in \Omega$, y al ser r y $r + 1$ números coprimos tenemos que el mcd de $\mathcal{T}(x)$ es 1 sin importar el estado x . \square

Los cursos introductorios a la teoría de las cadenas de Markov están interesados comúnmente en resultados como el que vimos en (1.10). Nosotros estamos más interesados en la conclusión que derivamos de la expresión (1.9). Es decir, queremos saber **cómo** se da esta convergencia. Las siguientes subsecciones tienen como finalidad proveer las herramientas que nos permitan llevar acabo esta investigación.

1.2. Representación canónica

Muchas cadenas de Markov admiten una descripción natural en términos de una ecuación de recurrencia. Una **representación canónica**, de una matriz de transición P con espacio de estados Ω , es una función $f : \Omega \times \Lambda \rightarrow \Omega$ junto con una variable aleatoria Z definida en Λ , en donde se cumple la siguiente igualdad:

$$\mathbb{P}(f(x, Z) = y) = P(x, y). \quad (1.13)$$

Sean Z_1, Z_2, \dots una sucesión de variables aleatorias independientes e igualmente distribuidas (i.i.d) que Z . Sea X_0 con distribución μ independiente a la sucesión de variables aleatorias. Entonces la sucesión (X_t) definida por la ecuación de recurrencia

$$X_t := f(X_{t-1}, Z_t), \quad \text{para } t \geq 1, \quad (1.14)$$

es una cadena de Markov con matriz de transición P y distribución inicial μ . En efecto, iterar la ecuación (1.14) muestra que podemos

encontrar una función g_{t+1} tal que $X_{t+1} = g_{t+1}(X_0, Z_1, \dots, Z_{t+1})$ para $t \geq 1$. A partir de esto establecemos

$$\begin{aligned} & \mathbb{P}(X_{t+1} = y | X_t = x, \dots, X_0 = x_0) \\ &= \mathbb{P}(f(x, Z_{t+1}) = y | X_t = x, \dots, X_0 = x_0). \end{aligned}$$

El evento $\{X_t = x, X_{t-1} = x_{t-1}, \dots, X_0 = x_0\}$ puede ser escrito como $\{g_t(X_0, Z_1, \dots, Z_t) = x, \dots, X_0 = x_0\}$, es decir en términos de X_0, Z_1, \dots, Z_t . Por hipótesis esto es independiente de Z_{t+1} y finalmente para toda $x, y \in \Omega$ tenemos

$$\begin{aligned} \mathbb{P}(X_{t+1} = y | X_t = x, \dots, X_0 = x_0) &= \mathbb{P}(f(x, Z_{t+1}) = y) \\ &= \mathbb{P}(X_{t+1} = y | X_t = x) = P(x, y). \end{aligned}$$

De manera más general, si relajamos la igualdad (1.13), la sucesión (X_t) definida por la recurrencia (1.14) sigue siendo una cadena de Markov aunque ya no “forzamos” a que la matriz de transición sea específicamente P .

El ejemplo clásico de una cadena de Markov que permite de forma natural una descripción de este tipo es la caminata aleatoria.

En el desarrollo anterior asumimos la existencia de una función f y una variable aleatoria Z que satisfacen ciertas propiedades respecto a una matriz P . Sin embargo, ¿para qué matrices P existe dicha dupla de f y Z ?

Proposición 1.2. *Toda matriz de transición en un espacio finito de estados tiene una representación canónica.*

Demostración. Sea P una matriz de transición de una cadena de Markov con espacio de estados $\Omega = \{x_1, x_2, \dots, x_n\}$. Designamos mediante Z, Z_1, Z_2, \dots , a la sucesión de variables aleatorias independientes con distribución uniforme en el intervalo $[0, 1]$.

Definimos la función $F_{j,k} = \sum_{i=1}^k P(x_j, x_i)$. Como se puede ver esto es la suma acumulada de la j -ésima fila de P hasta el k -ésimo elemento, y al ser esta fila una distribución entonces la suma se trata de su función de distribución acumulada. De aquí la razón de la notación elegida. Definimos a f como

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

De esta manera tenemos

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

y al ser Z uniforme

$$\begin{aligned} \mathbb{P}(F_{j,k-1} < Z \leq F_{j,k}) &= F_{j,k} - F_{j,k-1} \\ &= \sum_{i=1}^k P(x_j, x_i) - \sum_{i=1}^{k-1} P(x_j, x_i) \\ &= P(x_j, x_k). \end{aligned}$$

Por lo tanto

$$\mathbb{P}(f(x_j, Z) = x_k) = P(x_j, x_k).$$

□

Observemos en la demostración que debido a que Z es uniforme en $(0, 1)$ si hubiéramos cambiado a f por $f(x, 1 - z)$ llegaríamos a la misma conclusión. En otras palabras, la representación canónica no es única.

Retomemos nuestra matriz de ejemplo P . Supongamos que $p = q$ entonces

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

Por la Proposición 1.2 sabemos que existe una representación canónica. Sea Z una variable aleatoria con distribución Bernoulli con parámetro p . Sea la función $f : \Omega \times \{0, 1\} \rightarrow \Omega$ definida por

$$\begin{aligned} f(i, 0) &= i, & f(i, 1) &= d, \\ f(d, 0) &= d, & f(d, 1) &= i. \end{aligned}$$

Definimos una sucesión (Y_t) mediante la función f y una sucesión (Z_t) de variables aleatorias i.i.d Bernoulli con parámetro p de la siguiente manera

$$Y_t := f(Y_{t-1}, Z_t) \quad \text{para } t \geq 1.$$

Las probabilidades de transición de la cadena (Y_t) están dadas por

$$\begin{aligned} \mathbb{P}(f(i, Z) = i) &= 1 - p, & \mathbb{P}(f(i, Z) = d) &= p, \\ \mathbb{P}(f(d, Z) = i) &= p, & \mathbb{P}(f(d, Z) = d) &= 1 - p. \end{aligned}$$

De esta manera la función f definida junto con la variable aleatoria Z constituyen una representación canónica de la matriz de transición P .

1.3. Cadenas reversibles

Sea P una matriz de transición y supongamos que existe una distribución π que satisface las siguientes ecuaciones:

$$\pi(x)P(x, y) = \pi(y)P(y, x), \quad \text{para toda } x, y \in \Omega. \quad (1.15)$$

Estas igualdades reciben el nombre de **ecuaciones de balance**. El contenido de la siguiente proposición nos dice que entonces π debe ser una distribución estacionaria.

Proposición 1.3. *Sea P una matriz de transición en un espacio de estados Ω . Cualquier distribución π que satisface las ecuaciones de balance es una distribución estacionaria para P .*

Demostración. Como es usual mostraremos que $\pi = \pi P$. Sumando sobre todas las ecuaciones (1.15) tenemos

$$\sum_{y \in \Omega} \pi(y)P(y, x) = \sum_{y \in \Omega} \pi(x)P(x, y).$$

El lado derecho es $\pi(x)$ al ser P estocástica y por lo tanto

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

□

Supongamos que las ecuaciones (1.15) se cumplen para una cadena de Markov, entonces tenemos

$$\begin{aligned} & \pi(x_0)P(x_0, x_1) \dots \pi(x_{n-1})P(x_{n-1}, x_n) \\ &= \pi(x_n)P(x_n, x_{n-1}) \dots \pi(x_1)P(x_1, x_0). \end{aligned}$$

Eliminando los factores comunes $\pi(x_1), \pi(x_2), \dots, \pi(x_{n-1})$

$$\pi(x_0)P(x_0, x_1) \dots P(x_{n-1}, x_n) = \pi(x_n)P(x_n, x_{n-1}) \dots P(x_1, x_0).$$

Podemos reescribir esta igualdad de la siguiente manera

$$\mathbb{P}_\pi(X_0 = x_0, \dots, X_n = x_n) = \mathbb{P}_\pi(X_0 = x_n, \dots, X_n = x_0).$$

Esto quiere decir que si (X_t) satisface las ecuaciones de balance entonces la distribución de (X_0, \dots, X_t) es igual que la de (X_t, \dots, X_0) dado que la distribución inicial es π . Una manera de interpretar este resultado es diciendo que una transición del estado x al y hacia adelante en la cadena es equivalente a una transición de y a x hacia atrás en la cadena. Si $X_n = x$ y $X_{n+1} = y$ entonces tenemos una transición de x a y viendo la cadena hacia adelante, pero también tenemos una transición de y a

x viendo la cadena hacia atrás. Por esta razón a las cadenas de Markov que tienen esta característica se les conoce como cadenas reversibles.

El tiempo reverso de una matriz de transición irreducible P con distribución estacionaria π es la cadena con matriz de transición dada por

$$\hat{P}(x, y) = \frac{\pi(y)P(y, x)}{\pi(x)}, \quad \text{para toda } x, y \in \Omega. \quad (1.16)$$

Debido a que π es la distribución estacionaria para P la matriz \hat{P} es estocástica. En efecto, sumando sobre y tenemos

$$\sum_{y \in \Omega} \hat{P}(x, y) = \frac{1}{\pi(x)} \sum_{y \in \Omega} \pi(y)P(y, x) = 1, \quad \text{para cualquier } x \in \Omega.$$

Proposición 1.4. *Sea (X_t) una cadena de Markov irreducible con matriz de transición P y distribución estacionaria π . Sea (\hat{X}_t) el tiempo reverso de la cadena. Entonces π es estacionaria para \hat{P} , y para cualquier $x_0, \dots, x_t \in \Omega$*

$$\mathbb{P}_\pi(X_0 = x_0, \dots, X_t = x_t) = \mathbb{P}_\pi(\hat{X}_0 = x_t, \dots, \hat{X}_t = x_0).$$

Demostración. Primero mostraremos que π es estacionaria para \hat{P} .

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

Para demostrar la igualdad tenemos

$$\mathbb{P}_\pi(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) = \pi(x_0)P(x_0, x_1) \dots P(x_{n-1}, x_n).$$

El lado derecho de la igualdad se puede desarrollar de la siguiente manera usando (1.16)

$$\begin{aligned} \pi(x_1) \hat{P}(x_1, x_0) \dots \frac{\pi(x_n)}{\pi(x_{n-1})} \hat{P}(x_n, x_{n-1}) &= \pi(x_n) \hat{P}(x_n, x_{n-1}) \dots \hat{P}(x_1, x_0) \\ &= \mathbb{P}_\pi(\hat{X}_0 = x_n, \dots, \hat{X}_0 = x_0). \end{aligned}$$

□

Notemos que $\hat{P}(x, y) = P(x, y)$ para toda $x, y \in \Omega$ si y solo si la cadena con matriz de transición P es reversible.

1.4. Tiempos fuertemente estacionarios

Introducimos un concepto que resulta fundamental para la teoría de procesos estocásticos y también para los objetivos del presente texto.

Un **tiempo de paro** asociado a un proceso estocástico (X_t) en un espacio de estados Ω , es una variable aleatoria τ que toma valores en $\{0, 1, \dots, \infty\}$, si para cada $t \in \{0, 1, \dots\}$ el evento $\{\tau = t\}$ puede ser expresado en términos de X_0, \dots, X_t . Más formalmente un tiempo aleatorio τ es tiempo de paro si, y solo si

$$\mathbf{1}_{\{\tau=t\}} = \psi_t(X_0, \dots, X_t),$$

donde la función ψ_t es una función que solo toma valor 0 o 1. En términos simples, se puede pensar en un algoritmo que revisa la sucesión de estados y, acorde a una regla, “detiene” el proceso. Dicha regla tiene permitido observar lo que ocurre hasta el momento t pero no puede considerar momentos futuros.

Un ejemplo de tiempo de paro es el tiempo de llegada. Para una $x \in \Omega$ fija definimos

$$\tau_x = \min\{t \geq 0 : X_t = x\}.$$

Este tiempo de paro es la primera vez que la cadena de Markov (X_t) visita el estado x , entonces

$$\{\tau_x = t\} = \{X_0 \neq x, X_1 \neq x, \dots, X_{t-1} \neq x, X_t = x\}.$$

De manera general se puede definir τ_A , donde el ejemplo representa el caso $A = \{x\}$.

La Proposición 1.2 ha establecido que cualquier matriz P en un espacio de estados finito admite una representación canónica. Una variable aleatoria τ , análoga al tiempo de paro, se define utilizando la representación canónica. En este caso, este tiempo aleatorio recibe el nombre de **tiempo de paro aleatorizado** asociado a (X_t) , y debe cumplir que para cualquier tiempo $t \geq 0$ el evento $\{\tau = t\}$ pueda ser expresado en términos de X_0, Z_1, \dots, Z_t .

La presente investigación gira en torno a los tiempos requeridos por una cadena de Markov para estar cerca de su distribución estacionaria, resulta entonces inevitable que liguemos estos nuevos conceptos con la distribución π . Un tiempo de paro aleatorizado asociado a (X_t) es conocido como un tiempo estacionario si X_τ tiene distribución estacionaria

$$\mathbb{P}(X_\tau = x) = \pi(x) \quad \text{para toda } x \in \Omega.$$

Sin embargo, esta condición no es lo suficientemente fuerte para permitirnos estudiar los tiempos de mezcla. Si además X_τ es independiente de τ entonces a este tiempo de paro aleatorizado le llamamos **tiempo fuertemente estacionario**

$$\mathbb{P}_x(\tau = t, X_\tau = y) = \mathbb{P}_x(\tau = t)\pi(y). \quad (1.17)$$

Observemos que esta condición es equivalente a

$$\mathbb{P}(X_\tau = y | \tau = t) = \pi(y).$$

Para entender mejor la diferencia entre estos dos conceptos pensemos en la siguiente cadena de Markov definida en $\mathbb{Z}_4 = \{0, 1, 2, 3\}$. Una partícula se mueve sobre los puntos de \mathbb{Z}_4 dispuestos de forma circular (ver la Figura ??). Da un paso en el sentido del reloj al siguiente estado con probabilidad $1/4$, se mueve en

el otro sentido con probabilidad $1/4$, y se queda donde está con probabilidad $1/2$. Las probabilidades de transición se describen mediante la aritmética módulo 4

$$P(x, y) = \begin{cases} 1/4 & \text{si } y = x + 1 \pmod{4} \\ 1/4 & \text{si } y = x - 1 \pmod{4} \\ 1/2 & \text{si } y = x. \end{cases}$$

Esta cadena tiene una descripción natural en términos de una representación canónica. Sea Z una variable aleatoria definida en $\{-1, 0, 1\}$ tal que

$$\mathbb{P}(Z = k) = \begin{cases} 1/4 & \text{si } k = 1 \\ 1/4 & \text{si } k = -1 \\ 1/2 & \text{si } k = 0. \end{cases}$$

Sea Z_1, Z_2, \dots variables aleatorias i.i.d con la misma distribución que Z . Definimos (X_t) mediante la recurrencia

$$X_{t+1} = (X_t + Z_{t+1}) \pmod{4}, \quad \text{para } t \geq 0.$$

De esta manera para $x, y \in \mathbb{Z}_4$ tenemos

$$P(x, y) = \mathbb{P}((x + Z) \pmod{4} = y).$$

Este proceso se conoce como caminata aleatoria en los enteros \mathbb{Z}_4 y, como veremos más adelante, su distribución estacionaria es la distribución uniforme en \mathbb{Z}_4 .

Sea τ el primer momento donde la distribución de la cadena es estacionaria. Este tiempo de paro aleatorizado es estacionario (por definición) pero no fuertemente estacionario. En efecto, si $\{\tau = 1\}$ y el estado inicial es 0 entonces sabemos que X_τ no puede ser 2.

Vamos a presentar la construcción correcta de un tiempo fuertemente estacionario para esta cadena. Supongamos que el estado inicial es 0. Sea T_0 la primera vez que X_t cambia de estado, es decir $X_t = 3$ o $X_t = 1$. Sea $\tau = T_0 + 1$, un paso después de T_0 . Para mostrar que este tiempo de paro aleatorizado es un tiempo fuertemente estacionario debemos calcular la distribución de X_t dado $\{\tau = t\}$. Debido a las probabilidades de transición tenemos

$$\mathbb{P}(X_{\tau-1} = 1|\tau = t) = \mathbb{P}(X_{\tau-1} = 3|\tau = t) = \frac{1}{2}.$$

Luego

$$\begin{aligned}\mathbb{P}(X_\tau = 0|\tau = t) &= \left(\frac{1}{4}\right) \mathbb{P}(X_{\tau-1} = 1|\tau = t) \\ &\quad + \left(\frac{1}{4}\right) \mathbb{P}(X_{\tau-1} = 3|\tau = t) = \frac{1}{4}.\end{aligned}$$

El cálculo es idéntico para $\mathbb{P}(X_\tau = 2|\tau = t)$. Por otro lado

$$\mathbb{P}(X_\tau = 1|\tau = t) = \left(\frac{1}{2}\right) \mathbb{P}(X_{\tau-1} = 1|\tau = t) = \frac{1}{4}.$$

El resultado es el mismo para $\mathbb{P}(X_\tau = 3|\tau = t)$.

Abandonamos este ejemplo ilustrativo para continuar con una propiedad fundamental que nos permitirá usar este concepto para acotar los tiempos de mezcla.

Lema 1.3. *Sea (X_t) una cadena de Markov irreducible con distribución estacionaria π . Si τ es un tiempo fuertemente estacionario, entonces para toda $t \geq 0$*

$$\mathbb{P}_x(\tau \leq t, X_t = y) = \mathbb{P}(\tau \leq t)\pi(y).$$

Demostración. Sean Z_1, Z_2, \dots las variables aleatorias independientes e idénticamente distribuidas definidas en un espacio Λ usadas en la

representación canónica de (X_t) . Para cualquier $s \leq t$ podemos escribir

$$\begin{aligned}\mathbb{P}_x(\tau = s, X_t = y) &= \sum_{z \in \Omega} \mathbb{P}_x(\tau = s, X_t = y, X_s = z) \\ &= \sum_{z \in \Omega} \mathbb{P}_x(X_t = y | \tau = s, X_s = z) \mathbb{P}_x(\tau = s, X_s = z).\end{aligned}\tag{1.18}$$

Por otro lado, τ es también un tiempo de paro aleatorizado. Esto significa que el evento $\{\tau = s\}$ equivale a que existe una función $\psi_s(X_0, Z_1, \dots, Z_s)$. Si para (X_t) se usa la representación canónica, entonces X_t se define de manera recursiva $X_t = f(X_{t-1}, Z_t)$ para $t \geq 1$. Como f es la misma función para toda t , para expresar a X_t en términos de variables anteriores a las del tiempo $t - 1$, solo hace falta componer la función f consigo misma. Por ejemplo podemos escribir

$$X_t = f(f(X_{t-2}, Z_{t-1}), Z_t)$$

Naturalmente esto se puede extender a un número arbitrario de composiciones según queramos expresar a X_t en tiempos pasados. Estas iteraciones de la función f definen una nueva función \tilde{f}_n , donde n indica el número de composiciones. De esta manera, para enteros $r, s \geq 0$ existe entonces una función $\tilde{f}_r : \Omega \times \Lambda^r \rightarrow \Omega$ tal que

$$X_{s+r} = \tilde{f}_r(X_s, Z_{s+1}, \dots, Z_{s+r}).$$

Ahora podemos volver a expresar la siguiente probabilidad

$$\begin{aligned}\mathbb{P}_x(X_t = y | \tau = s, X_s = z) \\ &= \mathbb{P}_x(\tilde{f}_{t-s}(X_s, Z_{s+1}, \dots, Z_t) = y | \psi_s(X_0, Z_1, \dots, Z_s) = 1, X_s = z) \\ &= \mathbb{P}_x(\tilde{f}_{t-s}(z, Z_{s+1}, \dots, Z_t) = y | X_s = z) = P^{t-s}(z, y).\end{aligned}$$

La última igualdad se debe a que los vectores aleatorios (Z_1, \dots, Z_s) y (Z_{s+1}, \dots, Z_t) son independientes.

Reescribimos (1.18) usando la igualdad que obtuvimos

$$\mathbb{P}_x(\tau = s, X_t = y) = \sum_{z \in \Omega} P^{t-s}(z, y) \mathbb{P}_x(\tau = s, X_s = z),$$

y usando la definición (1.17) de tiempo fuertemente estacionario para τ resulta

$$\sum_{z \in \Omega} P^{t-s}(z, y) \mathbb{P}_x(\tau = s, X_s = z) = \mathbb{P}_x(\tau = s) \sum_{z \in \Omega} P^{t-s}(z, y) \pi(z).$$

Como π satisface $t - s$ veces la ecuación $\pi = \pi P$, entonces

$$\mathbb{P}_x(\tau = s) \sum_{z \in \Omega} P^{t-s}(z, y) \pi(z) = \mathbb{P}_x(\tau = s) \pi(y).$$

Por lo tanto para cualquier $s \leq t$ se cumple la siguiente igualdad

$$\mathbb{P}_x(\tau = s, X_t = y) = \mathbb{P}_x(\tau = s) \pi(y).$$

Esta equivalencia se deduce inmediatamente de la definición dada de tiempo fuertemente estacionario cuando $s = t$. Sin embargo, cuando la desigualdad es estricta era necesario proveer una demostración.

Para concluir la prueba sumamos para toda $s \leq t$

$$\sum_{s \leq t} \mathbb{P}_x(\tau = s, X_t = y) = \sum_{s \leq t} \mathbb{P}_x(\tau = s) \pi(y),$$

si y solo si,

$$\mathbb{P}_x(\tau \leq t, X_t = y) = \mathbb{P}_x(\tau \leq t) \pi(y).$$

□

Con este resultado tenemos la identidad

$$\mathbb{P}(X_t = y | \tau \leq t) = \pi(y).$$

El motivo por el cual hemos realizado este desarrollo sobre los tiempos aleatorios τ es porque son una herramienta que nos permitirán investigar los tiempos de convergencia. Retomaremos esta cuestión más adelante.

1.5. Distancia de variación total

El principal objetivo de esta primera parte de la tesis es investigar el tiempo que requiere una cadena de Markov irreducible y aperiódica en estar “cerca” de converger a su distribución estacionaria. Es decir, cuántas iteraciones hacen que el proceso esté bien mezclado. Es evidente que como primer paso para poder llevar acabo esta investigación debemos formalizar la noción de distancia existente entre la distribución de la cadena de Markov y la distribución estacionaria. Para este fin introducimos la siguiente métrica que mide la discrepancia entre dos distribuciones de probabilidad. Dadas dos distribuciones de probabilidad μ y ν en un mismo espacio Ω , la *distancia de variación total* se define como

$$\|\mu - \nu\|_{VT} := \max_{A \subset \Omega} |\mu(A) - \nu(A)|. \quad (1.19)$$

La distancia de variación total es la máxima diferencia, en términos de probabilidad, que las distribuciones le asignan a un mismo evento. Podemos advertir que se cumple

$$0 \leq \|\mu - \nu\|_{VT} \leq 1 \quad \text{para cualquier } \mu \text{ y } \nu. \quad (1.20)$$

Siendo la distancia nula si y solo si $\mu = \nu$. Por otro lado la distancia máxima ocurre si y solo si los soportes de las distribuciones son disjuntos.

Cómo es de esperarse esta no es la única manera de formalizar el concepto de distancia. Para $1 \leq p < \infty$ se puede definir la distancia $\ell^p(\pi)$ entre dos distribuciones μ y ν como

$$\|\mu - \nu\|_p = \left[\sum_{x \in \Omega} \left| \frac{\mu(x)}{\pi(x)} - \frac{\nu(x)}{\pi(x)} \right|^p \pi(x) \right]^{1/p}.$$

Así mismo podemos definir la distancia de Hellinger ampliamente usada

en estadística

$$d_H(\mu, \nu) = \sqrt{\sum_{x \in \Omega} \left(\sqrt{\mu(x)} - \sqrt{\nu(x)} \right)^2}.$$

Nosotros utilizaremos la distancia de variación total por ser más natural al contexto probabilístico de este trabajo.

Volvamos a nuestra cadena con dos estados para ilustrar este nuevo concepto calculando la distancia de variación total entre $\mu_t = (1, 0)P^t$ y la distribución estacionaria π . Encontramos la distribución estacionaria $\pi = \left(\frac{q}{p+q}, \frac{p}{p+q} \right)$ para nuestra matriz de transición

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

Gracias a que el espacio Ω consiste únicamente en dos estados existen cuatro eventos posibles de los cuales solo dos nos interesa estudiar. Excluimos los eventos Ω y \emptyset pues en ambos casos la distancia es nula como es fácil de comprobar. Por lo tanto nos centramos en $\{i\}$ y $\{d\}$. Sea

$$\Delta_t = \mu_t(i) - \pi(i) = \mathbb{P}(X_t = i | X_0 = i) - \pi(i) \quad \text{para } t \geq 0,$$

la diferencia para el primero de estos eventos entre las distribuciones μ_t y π . Notemos que ocurre $\mu_t(i) - \pi(i) = \pi(d) - \mu_t(d)$ al ser distribuciones definidas en un conjunto con dos elementos y con ello

$$|\mu_t(i) - \pi(i)| = |\mu_t(d) - \pi(d)|.$$

La definición de distancia de variación total nos requiere encontrar el evento con mayor diferencia de probabilidad. Sin embargo, ambos eventos presentan la misma diferencia en valor absoluto y podemos concluir

$$\|\mu_t - \pi\|_{VT} = |\Delta_t|, \quad \text{para } t \geq 0. \quad (1.21)$$

En nuestro ejemplo fue sencillo encontrar la distancia de variación total debido al tamaño del espacio Ω que nos hizo considerar solo dos eventos. En general, la definición de distancia de variación total nos obliga a considerar *todos* los subconjuntos del espacio de estados Ω a fin de encontrar aquel que maximice la diferencia. Como es de esperarse esto no es muy práctico cuando el espacio de estados es grande. Es conveniente encontrar otras caracterizaciones equivalentes.

Proposición 1.5. *Dadas μ y ν dos distribuciones en un mismo Ω , se tiene la siguiente igualdad*

$$\|\mu - \nu\|_{VT} = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|.$$

Demostración. Sea $B = \{x : \mu(x) \geq \nu(x)\}$ el subconjunto donde μ “domina” a ν , y sea $A \subset \Omega$ cualquier evento. Entonces el complemento está dado por $B^c = \{x : \mu(x) < \nu(x)\}$. Establecemos

$$\mu(A) - \nu(A) \leq \mu(A \cap B) - \nu(A \cap B). \quad (1.22)$$

En efecto, en el caso en el que A contenga elementos que están en B^c , en estos elementos se tiene $\mu(x) - \nu(x) < 0$. Al hacer $A \cap B$ se quitan estos elementos que disminuyen al resultado de $\mu - \nu$. En caso que $A \cap B^c = \emptyset$ se verifica la igualdad. Por otro lado

$$\mu(A \cap B) - \nu(A \cap B) \leq \mu(B) - \nu(B), \quad (1.23)$$

cuando A no abarca la totalidad de B existen elementos que se excluyen en los que $\mu(x) - \nu(x) \geq 0$.

Combinando las desigualdades (1.22) y (1.23) obtenemos

$$\mu(A) - \nu(A) \leq \mu(A \cap B) - \nu(A \cap B) \leq \mu(B) - \nu(B). \quad (1.24)$$

Mediante un razonamiento análogo

$$\nu(A) - \mu(A) \leq \nu(B^c) - \mu(B^c). \quad (1.25)$$

La cota superior anterior se puede operar de la siguiente manera

$$\nu(B^c) - \mu(B^c) = 1 - \nu(B) - (1 - \mu(B)) = \mu(B) - \nu(B), \quad (1.26)$$

por lo que es la misma cota superior que en (1.24), más aún

$$|\mu(B) - \nu(B)| = \mu(B) - \nu(B) = \nu(B^c) - \mu(B^c) = |\mu(B^c) - \nu(B^c)|. \quad (1.27)$$

Esta expresión cumple con la definición de distancia de variación total, pues $|\mu(A) - \nu(A)| \leq |\mu(B) - \nu(B)|$ para cualquier $A \subset \Omega$. Dependiendo del signo de $\mu(A) - \nu(A)$ esta desigualdad se verifica por (1.24) o (1.25). Por lo tanto

$$\|\mu - \nu\|_{VT} = |\mu(B) - \nu(B)| = |\mu(B^c) - \nu(B^c)|, \quad (1.28)$$

sumando ambas equivalencias y quitando los valores absolutos

$$\|\mu - \nu\|_{VT} = \frac{1}{2}(\mu(B) - \nu(B) + \nu(B^c) - \mu(B^c)).$$

Finalmente, reescribiendo la probabilidad de un evento como la suma de las probabilidades de los elementos que lo constituyen

$$\begin{aligned} \|\mu - \nu\|_{VT} &= \frac{1}{2} \left(\sum_{x \in B} [\mu(x) - \nu(x)] + \sum_{x \in B^c} [\nu(x) - \mu(x)] \right) \\ &= \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|. \end{aligned}$$

□

Este resultado reduce la distancia de variación total a una suma sobre todos los elementos de Ω y nos permite notar que la distancia de

variación total es la mitad de la distancia $\ell^1(\pi)$, de igual forma existe una desigualdad que relaciona la distancia de variación total con la distancia de Hellinger.

Otra ventaja de esta caracterización es que permite evidenciar más claramente que la distancia de variación total cumple con la desigualdad triangular:

$$\begin{aligned}\|\mu - \phi\|_{VT} + \|\phi - \nu\|_{VT} &= \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \phi(x)| + |\phi(x) - \nu(x)| \\ &\geq \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)| = \|\mu - \nu\|_{VT},\end{aligned}$$

para cualesquiera μ, ϕ y ν distribuciones en un mismo Ω . De esta manera la distancia de variación total es una métrica propiamente dicha, definida sobre la colección de todas las distribuciones de probabilidad en Ω que denotaremos por $\mathcal{P}(\Omega)$.

De la ecuación (1.28) también encontramos la siguiente equivalencia.

Corolario 1.

$$\|\mu - \nu\|_{VT} = \sum_{x \in B} (\mu(x) - \nu(x)) = \sum_{x \in B^c} (\nu(x) - \mu(x)).$$

Recordemos la definición $B = \{x : \mu(x) \geq \nu(x)\}$. Observemos que si $\mu \neq \nu$ entonces B es un subconjunto estricto del espacio Ω al ser ambas distribuciones de probabilidad. En nuestro ejemplo B consta de un elemento al igual que B^c .

La distancia de variación total entre dos distribuciones μ y ν también puede pensarse como la máxima área comprendida entre ambas distribuciones $\max_A |\mu(A) - \nu(A)|$. Observemos la Figura 1.1. En primera instancia resulta sorprendente el saber que las áreas I y II son iguales. El área I ocurre en los elementos de Ω donde la distribución μ está por encima de ν y de esta manera maximiza el área

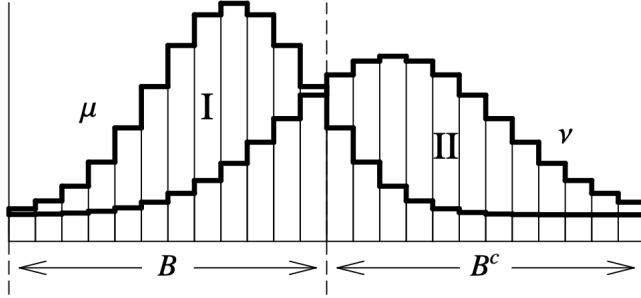


Figura 1.1. Recurso ilustrativo tomado de [16] donde se muestran dos distribuciones definidas en un mismo espacio.

resultando en $I = \mu(B) - \nu(B)$. El paso que falta para relacionar I y II nos lo da la ecuación (1.26), aquí se estableció

$$\mu(B) - \nu(B) = 1 - \nu(B^c) - (1 - \mu(B^c)).$$

El lado derecho de esta igualdad es precisamente $\nu(B^c) - \mu(B^c)$ que corresponde al área II. Así, las áreas I y II son iguales y, como se ha visto, ambos son la distancia de variación total.

Una vez introducido el concepto que nos permitirá medir la distancia que existe entre una cadena de Markov y su distribución estacionaria debemos asegurarnos que va a converger. Es decir, que la distancia de variación total entre ambos es 0 cuando t tiende a infinito. Esta conclusión la podemos derivar rápidamente del Teorema 1.2. Sin embargo, juzgamos conveniente presentar otra versión del teorema de la convergencia que involucra la nueva métrica que hemos presentado. A diferencia del Teorema 1.2 aquí sí desarrollaremos la demostración por ser pertinente al presente trabajo.

Teorema 1.4 (Teorema de la convergencia). *Sea P una matriz de transición irreducible y aperiódica con distribución estacionaria π .*

Entonces existen constantes $\alpha \in (0, 1)$ y $C > 0$ tal que

$$\max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{VT} \leq C \cdot \alpha^t.$$

Demostración. Como P es irreducible y aperiódica entonces por la Proposición 1.1 existe r tal que P^r tiene todas las entradas estrictamente positivas. Sea Π la matriz de misma dimensión que P definida por

$$\Pi(x, y) := \pi(y). \quad (1.29)$$

Podemos encontrar una $\delta > 0$ lo suficientemente pequeña de manera que

$$P^r(x, y) \geq \delta \pi(y) \quad \text{para todo } x, y \in \Omega.$$

Sea $\theta = 1 - \delta$, entonces la ecuación

$$P^r = (1 - \theta)\Pi + \theta Q \quad (1.30)$$

define una matriz Q que resulta ser estocástica. En efecto, para cada (x, y) se tiene

$$Q(x, y) = \frac{P^r(x, y) + (\theta - 1)\Pi(x, y)}{\theta},$$

fijando x y sumando sobre todo los elementos y

$$\begin{aligned} \sum_{y \in \Omega} Q(x, y) &= \sum_{y \in \Omega} \frac{P^r(x, y) + (\theta - 1)\Pi(x, y)}{\theta} \\ &= \frac{1}{\theta} \left(\sum_{y \in \Omega} P^r(x, y) + (\theta - 1) \sum_{y \in \Omega} \Pi(x, y) \right) \\ &= \frac{1}{\theta} (1 + (\theta - 1)) \\ &= 1. \end{aligned}$$

Para cualquier matriz estocástica M se cumple que $M\Pi = \Pi$, este hecho se debe a que cualquier entrada (x, y) de $M\Pi$ es el producto punto de la fila x de M con la columna y de Π , esto es

$$M\Pi(x, y) = \sum_{i \in \Omega} M(x, i)\Pi(i, y) = \pi(y) \sum_{i \in \Omega} M(x, i) = \pi(y) = \Pi(x, y). \quad (1.31)$$

Otra identidad que será de utilidad en el desarrollo subsecuente es que debido a que $\pi P^t = \pi$ para $t \geq 0$, entonces

$$\Pi P^t = \Pi \quad \text{para } t \geq 0. \quad (1.32)$$

Con estas propiedades en mente vamos a probar por inducción la siguiente igualdad para $k \geq 1$

$$P^{rk} = (1 - \theta^k) \Pi + \theta^k Q^k. \quad (1.33)$$

Para $k = 1$ se cumple debido a (1.30). Suponemos entonces que se cumple $P^{rn} = (1 - \theta^n)\Pi + \theta^n Q^n$. Multiplicamos por P^r ambos lados por la derecha, obteniendo

$$P^{r(n+1)} = (1 - \theta^n) \Pi P^r + \theta^n Q^n P^r.$$

Usamos (1.32) y expandimos P^r en el segundo término resultando

$$P^{r(n+1)} = (1 - \theta^n) \Pi + \theta^n (1 - \theta) Q^n \Pi + \theta^{n+1} Q^{n+1}.$$

Debido a (1.31) podemos simplificar la ecuación y reordenando términos

$$P^{r(n+1)} = (1 - \theta^{n+1}) \Pi + \theta^{n+1} Q^{n+1}.$$

Esto prueba que la identidad (1.33) es válida para $k \geq 1$.

Con las igualdades que necesitaremos ya establecidas podemos pasar a la demostración del teorema. Multiplicamos (1.33) por P^j por la derecha y operando con (1.32) resulta

$$P^{rk+j} = (1 - \theta^k) \Pi + \theta^k Q^k P^j = \theta^k (Q^k P^j - \Pi) + \Pi. \quad (1.34)$$

Observemos que $Q^k P^j$ es una matriz estocástica al verificarse

$$\sum_{y \in \Omega} Q^k P^j(x, y) = \frac{1}{\theta^k} \sum_{y \in \Omega} P^{rk+j}(x, y) + \left(1 - \frac{1}{\theta^k}\right) \sum_{y \in \Omega} \Pi(x, y) = 1,$$

para cualquier $x \in \Omega$. Fijamos algún renglón x en (1.34) y sumamos sobre los valores absolutos

$$\sum_{y \in \Omega} |P^{rk+j}(x, y) - \Pi(x, y)| = \theta^k \sum_{y \in \Omega} |Q^k P^j(x, y) - \Pi(x, y)|,$$

que es equivalente a

$$\sum_{y \in \Omega} |P^{rk+j}(x, y) - \pi(y)| = \theta^k \sum_{y \in \Omega} |Q^k P^j(x, y) - \pi(y)|.$$

Dividiendo entre 2 y en virtud de la Proposición 1.5

$$\left\| P^{rk+j}(x, \cdot) - \pi \right\|_{VT} = \theta^k \left\| Q^k P^j(x, \cdot) - \pi \right\|_{VT}.$$

Como hemos visto en (1.20) la distancia de variación total está acotada y a lo más puede ser igual a 1. Por lo tanto, para cualquier $x \in \Omega$ se tiene

$$\left\| P^{rk+j}(x, \cdot) - \pi \right\|_{VT} \leq \theta^k. \quad (1.35)$$

□

El teorema de la convergencia juega un papel fundamental para el desarrollo de la teoría. No obstante este no nos da indicio alguno sobre el comportamiento de las distancias respecto a la distribución estacionaria a medida que avanza el proceso. Una característica deseable sería que al avanzar en el tiempo la distancia al equilibrio no puede aumentar.

La siguiente proposición establece que una matriz de transición P “contrae” la distancia entre dos distribuciones.

Proposición 1.6. *Sea P una matriz de transición. Sean μ y ν dos distribuciones definidas en el mismo espacio Ω . Entonces*

$$\|\mu P - \nu P\|_{VT} \leq \|\mu - \nu\|_{VT}$$

Demostración. Empleando la caracterización alternativa de la distancia de variación total tenemos

$$\|\mu P - \nu P\|_{VT} = \frac{1}{2} \sum_{x \in \Omega} |\mu P(x) - \nu P(x)|. \quad (1.36)$$

Debido a que $\mu P(x)$ y $\nu P(x)$ son en realidad un producto punto, los reexpresamos de esta manera en (1.36) y factorizando obtenemos

$$\frac{1}{2} \sum_{x \in \Omega} \left| \sum_{y \in \Omega} \mu(y) P(y, x) - \sum_{y \in \Omega} \nu(y) P(y, x) \right| = \frac{1}{2} \sum_{x \in \Omega} \left| \sum_{y \in \Omega} (\mu(y) - \nu(y)) P(y, x) \right|.$$

Por la desigualdad del triángulo tenemos

$$\frac{1}{2} \sum_{x \in \Omega} \left| \sum_{y \in \Omega} (\mu(y) - \nu(y)) P(y, x) \right| \leq \frac{1}{2} \sum_{x \in \Omega} \sum_{y \in \Omega} |\mu(y) - \nu(y)| P(y, x).$$

Intercambiando el orden de las sumas del lado derecho de la desigualdad resulta

$$\frac{1}{2} \sum_{y \in \Omega} |\mu(y) - \nu(y)| \sum_{x \in \Omega} P(y, x) = \frac{1}{2} \sum_{y \in \Omega} |\mu(y) - \nu(y)| = \|\mu - \nu\|_{VT}.$$

Por lo tanto concluimos

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

□

Consideremos una cadena de Markov irreducible y aperiódica con distribución estacionaria π . Sea μ_t la distribución del proceso tras t pasos entonces por la proposición anterior tenemos:

$$\|\mu_{t+1} - \pi\|_{VT} = \|\mu_t P - \pi P\|_{VT} \leq \|\mu_t - \pi\|_{VT}.$$

Este resultado establece que la sucesión de distancias $\{\|\mu_0 P^t - \pi\|_{VT}\}_{t=1}^{\infty}$ es decreciente para μ_0 una distribución inicial arbitraria.

Uno de nuestros principales objetivos es poder acotar la distancia que existe entre una cadena de Markov y su distribución estacionaria. Vale la pena recordar que la distribución del proceso cambia según elijamos una distribución inicial μ_0 distinta. Esto quiere decir que para distintas μ_0 la distancia con respecto a la distribución estacionaria π va a ser distinta para un mismo tiempo t . Es entonces razonable tomar como referencia el peor caso, aquel donde la distancia respecto a π es mayor. De modo que si sabemos que bajo esta distribución, el proceso se encuentra arbitrariamente cercano al equilibrio, entonces el proceso sin importar que distribución inicial se especifique también lo va a estar.

Empecemos resolviendo este problema considerando únicamente las distribuciones del proceso que tienen distribuciones iniciales del tipo δ_x . Es decir, las que toda su masa de probabilidad se encuentra en un solo estado x . Dado que el espacio de estados Ω es finito, el número de distribuciones con este tipo de distribución inicial también es finito. De hecho, como hemos visto se tratan de las filas de la matriz de transición en t pasos. Por lo tanto para encontrar la distancia que acote a las demás basta encontrar el máximo de estas. Definimos entonces la siguiente función

$$d(t) := \max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{VT}. \quad (1.37)$$

Aunque es clara la razón por la que hemos definido a $d(t)$ aún no hemos justificado que esta distancia basta para acotar a *todas* las demás.

Las distribuciones de la cadena de Markov cuando μ_0 no es del tipo δ_x , al contrario que en el caso anterior, son infinitas puesto que son infinitas las maneras en las que podemos asignar probabilidad a los estados x de Ω . Recordemos que denotamos por $\mathcal{P}(\Omega)$ a la colección de todas las distribuciones de probabilidad definidas sobre un conjunto finito Ω . El conjunto que contiene la totalidad de distancias a un tiempo t con independencia de la naturaleza de μ_0 es $\{\|\mu_0 P^t - \pi\|_{VT} : \mu_0 \in \mathcal{P}(\Omega)\}$. Ante esta situación no basta tratar de encontrar en primera instancia un máximo de las distancias como lo hicimos con $d(t)$, ya que al ser conjunto infinito nada nos garantiza su existencia. Emplearemos la noción de supremo como herramienta para acotar a cualquier otra distancia respecto a la distribución estacionaria a un tiempo t .

Proposición 1.7. *Sea $\mathcal{P}(\Omega)$ la colección de todas las distribuciones de probabilidad definidas en un conjunto finito Ω . Entonces*

$$\sup_{\mu \in \mathcal{P}(\Omega)} \|\mu P^t - \pi\|_{VT} = d(t).$$

Demostración. En virtud de la Proposición 1.5 tenemos

$$\|\mu P^t - \pi\|_{VT} = \frac{1}{2} \sum_{y \in \Omega} |\mu P^t(y) - \pi(y)|.$$

Reescribiendo en términos del producto punto y usando la desigualdad

del triángulo

$$\begin{aligned}
\frac{1}{2} \sum_{y \in \Omega} \left| \sum_{x \in \Omega} \mu(x) P^t(x, y) - \sum_{x \in \Omega} \mu(x) \pi(y) \right| \\
= \frac{1}{2} \sum_{y \in \Omega} \left| \sum_{x \in \Omega} \mu(x) (P^t(x, y) - \pi(y)) \right| \\
\leq \frac{1}{2} \sum_{y \in \Omega} \sum_{x \in \Omega} \mu(x) |P^t(x, y) - \pi(y)|.
\end{aligned}$$

Intercambiando el orden de las sumas del lado derecho de la desigualdad obtenemos una distancia de variación total de la siguiente manera

$$\sum_{x \in \Omega} \mu(x) \frac{1}{2} \sum_{y \in \Omega} |P^t(x, y) - \pi(y)| = \sum_{x \in \Omega} \mu(x) \|P^t(x, \cdot) - \pi\|_{VT}.$$

Debido a que un promedio ponderado de números nunca es mayor que su elemento máximo, podemos acotar como sigue

$$\sum_{x \in \Omega} \mu(x) \|P^t(x, \cdot) - \pi\|_{VT} \leq \max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{VT},$$

y por lo tanto

$$\|\mu P^t - \pi\|_{VT} \leq \max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{VT} = d(t).$$

Es decir $d(t)$ es el máximo del conjunto de distancias $\{\|\mu_0 P^t - \pi\|_{VT} : \mu_0 \in \mathcal{P}(\Omega)\}$ y por lo tanto también es la menor cota superior del mismo. Con esto concluimos

$$\sup_{\mu \in \mathcal{P}(\Omega)} \|\mu P^t - \pi\|_{VT} = d(t).$$

□

Gracias a este resultado podemos simplificar el análisis, ya que basta considerar al conjunto $\{\|P^t(x, \cdot) - \pi\|_{VT} : x \in \Omega\}$ para acotar cualquier otra distancia al equilibrio a un tiempo t .

El Teorema de la Convergencia (Teorema 1.4) mostró que $\max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{VT} \rightarrow 0$ cuando $t \rightarrow \infty$. No obstante no se mostró explícitamente que este resultado se mantenía cuando la distribución inicial no es del tipo δ_x . Ahora sabemos que, sin importar la distribución inicial, la distribución del proceso va a converger a la distribución estacionaria π . Este hecho justifica que la distribución estacionaria también sea llamada *distribución de equilibrio*.

En la Proposición 1.6 vimos que avanzar un paso en la cadena de Markov no puede alejarnos de la distribución estacionaria π , combinando este hecho con la Proposición 1.7 obtenemos

$$\|P^{t+1}(x, \cdot) - \pi\|_{VT} \leq \|P^t(x, \cdot) - \pi\|_{VT} \leq d(t) \quad \text{para todo } x \in \Omega.$$

Esto muestra que $d(t+1) \leq d(t)$ para todo $t \in \mathbb{N}$.

Regresemos a nuestra bien conocida cadena de Markov para encontrar la función $d(t)$. Recordemos la matriz de transición

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

Así como la distribución estacionaria

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

En (1.21) habíamos concluido

$$\|\mu_t - \pi\|_{VT} = |\mu_t(i) - \pi(i)| = |\mu_t(d) - \pi(d)|,$$

donde $\mu_t = \mu_0 P^t$ con $\mu_0 = (1, 0)$.

Realicemos un análisis similar para $\mu'_t = \mu'_0 P^t$, donde $\mu'_0 = (0, 1)$. Notemos que $\mu_t = P^t(i, \cdot)$ y $\mu'_t = P^t(d, \cdot)$. Definimos

$$\Delta'_t = \mu'_t(i) - \frac{q}{p+q} \quad \text{para } t \geq 0.$$

Para Δ'_{t+1} tenemos

$$\Delta'_{t+1} = \mu'_{t+1}(i) - \frac{q}{q+p} = \mu'_t(i)(1-p) + \mu'_t(d)q - \frac{q}{q+p}.$$

Al ser μ'_t una distribución definida en un espacio con dos elementos llegamos a una conclusión semejante al caso de Δ_t ,

$$\Delta'_t = (1-p-q)^t \Delta'_0 \quad \text{para } t \geq 0.$$

Por otro lado, mediante un razonamiento análogo al hecho en (1.21) tenemos

$$\|\mu'_t - \pi\|_{VT} = |\mu'_t(i) - \pi(i)| = |\mu'_t(d) - \pi(d)|.$$

De esta manera determinar $d(t)$ se reduce a comparar

$$\begin{aligned} \|\mu_t - \pi\|_{VT} &= |(1-p-q)^t| \left| 1 - \frac{q}{p+q} \right| = |1-p-q|^t \left(\frac{p}{p+q} \right) \\ \|\mu'_t - \pi\|_{VT} &= |(1-p-q)^t| \left| -\frac{q}{p+q} \right| = |1-p-q|^t \left(\frac{q}{p+q} \right). \end{aligned}$$

En otras palabras $d(t)$ siempre es el cálculo de la distancia de variación total asociada al mismo renglón, obviamente el valor de este cambia al variar P^t . Podemos identificar fácilmente cuál renglón tiene una mayor distancia al equilibrio, si $q < p$ entonces para todo tiempo t

$$\|P^t(d, \cdot) - \pi\|_{VT} < \|P^t(i, \cdot) - \pi\|_{VT}.$$

El caso contrario ocurre si $p < q$. En resumen tenemos

$$d(t) = |1-p-q|^t \left(\frac{\max(p, q)}{p+q} \right) \quad \text{para } t \geq 0, \quad (1.38)$$

y la tasa de decaimiento está dada por $|1-q-p|$.

La pregunta fundamental de nuestro trabajo implica que consideremos la pregunta más general *¿cuándo un proceso está bien*

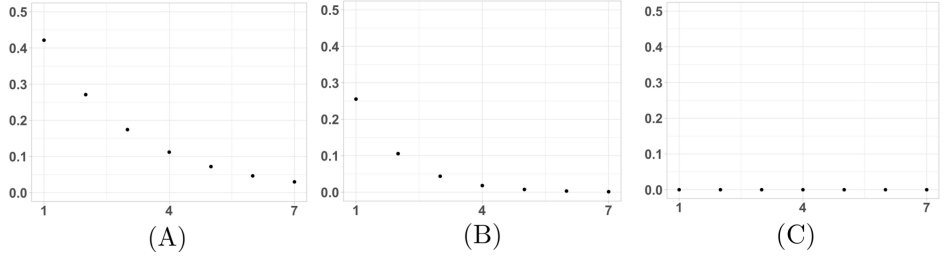


Figura 1.2. Gráfica de $d(t)$ contra el tiempo para (A) $p = 0.123$ y $q = 0.234$, (B) $p = 0.872$ y $q = 0.542$, y (C) $p = q = 0.5$.

mezclado? Ambas son preguntas todavía ambiguas y es necesario formalizarlas para poder dar una respuesta matemática rigurosa. La interpretación que hemos dado de la función $d(t)$ nos permite transformar esta cuestión general en la pregunta bien planteada *¿qué tan grande tiene que ser t para hacer $d(t) \leq \varepsilon$?* De ahora en adelante en las investigaciones sobre el tiempo de mezcla supondremos siempre una matriz P irreducible y aperiódica.

En la distancia inicial $d(0)$ estamos considerando únicamente las distribuciones iniciales del tipo δ_x . Por ser P irreducible la distribución estacionaria π tiene soporte completo. Gracias al Corolario 1 esto quiere decir que para estas distribuciones iniciales tenemos $\|\delta_x - \pi\|_{VT} = 1 - \pi(x)$ y por lo tanto

$$d(0) = 1 - \pi_\star \quad \text{donde } \pi_\star := \min_{x \in \Omega} \pi(x).$$

Nuevamente por ser π una función de distribución con soporte Ω se cumple $\pi_\star \leq |\Omega|^{-1}$ y con ello

$$\frac{|\Omega| - 1}{|\Omega|} \leq d(0).$$

De esta manera si el espacio de estados es grande entonces la distancia inicial es cercana a uno.

Una propiedad fundamental de $d(t)$ es la submultiplicatividad. Esta puede resumirse en que una vez que inicia la convergencia esta no puede ralentizarse.

Antes de pasar a discutir el anterior punto recordemos la norma infinita de matrices que denotaremos por $\| \cdot \|_\infty$, en este caso estamos interesados en matrices cuadradas de dimensión $|\Omega| \times |\Omega|$. Para ella tenemos la siguiente caracterización:

$$\|A\|_\infty = \max_{x \in \Omega} \sum_{y \in \Omega} |A(x, y)|.$$

Es decir, se trata de la máxima suma de los renglones de A en valor absoluto. Una propiedad fundamental de esta norma es que es submultiplicativa, esto es:

$$\|AB\|_\infty \leq \|A\|_\infty \|B\|_\infty \quad \text{para cualquier } A, B.$$

Lema 1.5. *Para toda $t, s \in \mathbb{N}$ tenemos*

$$d(t + s) \leq 2d(t)d(s).$$

Demostración. Sea P la matriz de transición de una cadena ergódica con distribución estacionaria π . Sea Π la matriz definida en (1.29) con renglón común π . Debido a las propiedades de Π se cumple

$$\Pi^2 = P\Pi = \Pi P = \Pi. \tag{1.39}$$

Esto significa que las matrices Π y P conmutan con lo que $P^k \Pi^s = \Pi^s P^k$ y es posible expandir $(P - \Pi)^t$ usando el teorema del binomio,

$$(P - \Pi)^t = \sum_{k=0}^t \binom{t}{k} (-1)^k \Pi^k P^{t-k}.$$

Gracias a (1.39) podemos descomponer la suma de la siguiente forma

$$\begin{aligned}\sum_{k=0}^t \binom{t}{k} (-1)^k \Pi^k P^{t-k} &= P^t + \sum_{k=1}^t \binom{t}{k} (-1)^k \Pi^k P^{t-k} \\ &= P^t + \Pi \sum_{k=1}^t \binom{t}{k} (-1)^k.\end{aligned}$$

Usando la ecuación $(1 - 1)^t = \sum_{k=0}^t \binom{t}{k} (-1)^k = 0$ finalmente tenemos la igualdad

$$(P - \Pi)^t = P^t - \Pi.$$

Consideremos la norma infinita de matrices introducida anteriormente. Observemos que se cumple para toda $t \geq 1$

$$\|(P - \Pi)^t\|_\infty = \max_{x \in \Omega} \sum_{y \in \Omega} |P^t(x, y) - \pi(y)|,$$

y en virtud de la Proposición 1.5

$$2d(t) = \|(P - \Pi)^t\|_\infty.$$

Por la propiedad submultiplicativa obtenemos

$$\|(P - \Pi)^{t+s}\|_\infty \leq \|(P - \Pi)^t\|_\infty \|(P - \Pi)^s\|_\infty,$$

y por lo tanto

$$d(t + s) \leq 2d(t)d(s).$$

□

A partir de este resultado podemos derivar para k y t enteros positivos

$$\begin{aligned}d(kt) &= d((k-1)t + t) \leq 2d(t)d((k-1)t) \leq 2^2 d(t)^2 d((k-2)t) \\ &\leq \dots \leq 2^{k-1} d(t)^k \leq (2d(t))^k.\end{aligned}\tag{1.40}$$

Luego si t iteraciones fueron suficientes para hacer $d(t) \leq 1/4$ entonces kt son suficientes para reducir la distancia al equilibrio a 2^{-k} .

Del álgebra lineal sabemos que una matriz de transición P tiene $|\Omega|$ eigenvalores (contando las multiplicidades algebraicas). Denotemos por $S_p(P)$ al conjunto de los eigenvalores de la matriz P , a este conjunto se le conoce como el espectro de P . Una propiedad fundamental para el espectro de una matriz de transición P (ver Capítulo 3 de [23]) establece que $S_p(P) \subset \{\lambda \in \mathbb{C} : |\lambda| \leq 1\}$.

Por otro lado denotemos por $\rho(P)$ al máximo módulo de los eigenvalores de P , este número es llamado el **radio espectral** de P .

En virtud de la fórmula de Gelfand [15], para una matriz P ergódica tenemos

$$\rho(P - \Pi) = \lim_{t \rightarrow \infty} \|(P - \Pi)^t\|_{\infty}^{\frac{1}{t}},$$

o de manera equivalente

$$\rho(P - \Pi) = \lim_{t \rightarrow \infty} (2d(t))^{\frac{1}{t}}.$$

Más aún, $\rho(P - \Pi) \leq \|(P - \Pi)^t\|_{\infty}^{\frac{1}{t}}$ para toda $t \in \mathbb{N}$ lo que significa que el límite se aproxima desde arriba.

No es complicado mostrar la siguiente elegante caracterización del radio espectral de $P - \Pi$ (pág. 32 de [23]):

$$\rho(P - \Pi) = \max\{|\lambda| : \lambda \text{ es eigenvalor de } P \text{ y } |\lambda| \neq 1\}.$$

Esto nos permite establecer que las cadenas de Markov ergódicas convergen exponencialmente rápido.

Corolario 1.1 (Decrecimiento geométrico). *Sea P una matriz irreducible y aperiódica, entonces existe $\lambda_{\star} < 1$ tal que*

$$\lim_{t \rightarrow \infty} d(t)^{1/t} = \lambda_{\star}, \tag{1.41}$$

donde

$$\lambda_{\star} = \max\{|\lambda| : \lambda \text{ es eigenvalor de } P \text{ y } |\lambda| \neq 1\}.$$

Observemos que esto es un refinamiento del clásico teorema sobre la convergencia de una matriz irreducible y aperiódica (Teorema 1.2), pues además nos dice que asintóticamente $d(t)$ decae como λ_{\star}^t . Este fenómeno se presentó en nuestra cadena de Markov ilustrativa con dos estados, donde la tasa de decaimiento encontrada $|1 - p - q|$ es precisamente el valor absoluto de un eigenvalor de la matriz P .

Este importante resultado explica el comportamiento asintótico de la convergencia de una cadena de Markov. Sin embargo es de poco interés práctico pues justamente describe la distancia al equilibrio para tiempos en los que dicha distancia ya es infinitesimalmente pequeña. En términos de nuestro problema central: estamos interesados en saber que tan bien está barajada una pila de cartas tras unos cuantos barajeos no después de 1 millón de barajeos.

Los desarrollos que hemos expuesto nos abren las puertas a una interesante teoría donde se utilizan los eigenvalores y eigenfunciones, asociados a una matriz irreducible y aperiódica P , para investigar los tiempos de mezcla. Dicha teoría es especialmente fructífera para cadenas que además son reversibles. No ahondaremos más en este tema ya que está fuera de los objetivos de esta tesis. Referimos a los Capítulos 12 y 13 de [16] para un estudio formal de estas cuestiones.

De ahora en adelante vamos a dirigir la investigación al comportamiento no asintótico de la convergencia de una cadena de Markov. Para ello es conveniente definir un parámetro que contabilice el tiempo necesario para que una cadena de Markov se encuentre arbitrariamente cerca del equilibrio, el *tiempo de mezcla* cumplirá

esta función y se define como:

$$t_{\text{mezcla}}(\varepsilon) := \min \{t : d(t) \leq \varepsilon\}. \quad (1.42)$$

Así mismo definimos

$$t_{\text{mezcla}} := t_{\text{mezcla}}\left(\frac{1}{4}\right). \quad (1.43)$$

En primera instancia puede parecer arbitraria la elección $\varepsilon = 1/4$. Sin embargo es suficiente, cualquier valor más preciso puede ser acotado gracias a la siguiente desigualdad

$$t_{\text{mezcla}} \leq t_{\text{mezcla}}(\varepsilon) \leq t_{\text{mezcla}} \lceil \log_2 \varepsilon^{-1} \rceil, \quad \text{para } \varepsilon \in (0, 1/4),$$

donde $\lceil \cdot \rceil$ denota a la función techo. La cota inferior es inmediata. Para la cota superior consideremos (1.40) de la siguiente manera

$$d(kt_{\text{mezcla}}(\varepsilon)) \leq (2\varepsilon)^k.$$

Evaluando en $\varepsilon = 1/4$ tenemos

$$d(kt_{\text{mezcla}}) \leq 2^{-k}.$$

Notemos que $k \geq \log_2 \varepsilon^{-1}$ hace que $2^{-k} \leq \varepsilon$. La función $d(t)$ está definida solo en los enteros positivos por lo que tomaremos $k = \lceil \log_2 \varepsilon^{-1} \rceil$ y por lo tanto

$$t_{\text{mezcla}}(\varepsilon) \leq t_{\text{mezcla}} \lceil \log_2 \varepsilon^{-1} \rceil.$$

Juzgamos conveniente explicar ahora la manera en la que se usan los tiempos fuertemente estacionarios τ , ya introducidos, para proporcionar una cota superior a $d(t)$. Recordemos que un tiempo de paro aleatorizado τ es un tiempo fuertemente estacionario si satisface que

$$\mathbb{P}_x(\tau = t, X_\tau = y) = \mathbb{P}_x(\tau = t)\pi(y).$$

Así mismo resulta útil introducir la siguiente función $s_x(t)$. Esta recibe el nombre de ***distancia de separación*** y se define como

$$s_x(t) := \max_{y \in \Omega} \left[1 - \frac{P^t(x, y)}{\pi(y)} \right]. \quad (1.44)$$

También definimos

$$s(t) := \max_{x \in \Omega} s_x(t). \quad (1.45)$$

La primera función puede pensarse como un parámetro que compara cada entrada de una fila determinada x de la matriz P^t , con las entradas de la distribución estacionaria π . Buscamos aquella que maximice esta “distancia”. Mientras que la segunda tiene la misma finalidad, pero aplicado sobre la totalidad de las entradas de la matriz P^t . Escribimos distancia entre comillas pues no es una distancia propiamente dicho.

La siguiente proposición describe la relación entre los tiempos fuertemente estacionarios y los tiempos de mezcla.

Proposición 1.8. *Si τ es un tiempo fuertemente estacionario, entonces*

$$d(t) = \max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{VT} \leq \max_{x \in \Omega} \mathbb{P}_x(\tau > t). \quad (1.46)$$

La demostración se divide en dos lemas. Este primer lema destaca la importancia de un tiempo fuertemente estacionario sobre un tiempo solamente estacionario, y muestra cómo se relaciona con la distancia de separación antes definida.

Lema 1.6. *Si τ es un tiempo fuertemente estacionario, entonces*

$$s_x(t) \leq \mathbb{P}_x(\tau > t).$$

Demostración. Para una $x \in \Omega$ fija y cualquier $y \in \Omega$ ocurre

$$1 - \frac{P^t(x, y)}{\pi(y)} = 1 - \frac{\mathbb{P}_x(X_t = y)}{\pi(y)} \leq 1 - \frac{\mathbb{P}_x(X_t = y, \tau \leq t)}{\pi(y)}. \quad (1.47)$$

Esta desigualdad es cierta porque $\{X_t = y, \tau \leq t\} \subseteq \{X_t = y\}$. La igualdad se verifica cuando $\{\tau \leq t\} \subseteq \{X_t = y\}$. Al estado $y \in \Omega$ que cumple esta propiedad se le llama **estado de parada**.

En virtud del Lema 1.3 el lado derecho de la desigualdad (1.47) equivale a

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

Debido a que la desigualdad (1.47) sucede para cualquier $y \in \Omega$ también sucede para el máximo, entonces concluimos

$$s_x(t) \leq \mathbb{P}_x(\tau > t).$$

□

A los tiempos fuertemente estacionarios que satisfacen la igualdad se les conocen como *perfectos*. Para más detalles se puede consultar [20].

El siguiente lema nos dice que la distancia de separación domina a la distancia de variación total y por lo tanto proporciona una cota superior.

Lema 1.7. *La distancia de separación satisface la siguiente desigualdad*

$$\|P^t(x, \cdot) - \pi\|_{VT} \leq s_x(t).$$

Demostración. Empleando una de las equivalencias de la distancia de variación total

$$\begin{aligned} \|P^t(x, \cdot) - \pi\|_{VT} &= \sum_{\substack{y \in \Omega \\ P^t(x, y) < \pi(y)}} (\pi(y) - P^t(x, y)) \\ &= \sum_{\substack{y \in \Omega \\ P^t(x, y) < \pi(y)}} \pi(y) \left(1 - \frac{P^t(x, y)}{\pi(y)}\right). \end{aligned}$$

Debido a que un promedio ponderado nunca es mayor que su elemento máximo, podemos acotar de la siguiente manera

$$\|P^t(x, \cdot) - \pi\|_{VT} \leq \max_{y \in \Omega} \left[1 - \frac{P^t(x, y)}{\pi(y)} \right] = s_x(t). \quad (1.48)$$

□

Retomemos el ejemplo de la caminata aleatoria en \mathbb{Z}_4 . Nuestra cadena (X_t) tiene estado inicial 0. Como veremos más adelante en (2.8) es suficiente con analizar esta distribución para poder acotar a $d(t)$. Encontramos que τ , definido como un paso después de la primera vez que X_t toma el valor 1 o 3, es un tiempo fuertemente estacionario. Siguiendo la Proposición 1.8 para acotar el tiempo de mezcla es necesario calcular la probabilidad de la cola derecha de τ ,

$$\mathbb{P}_0(\tau > t) = \mathbb{P}_0(X_1 = 0, X_2 = 0, \dots, X_{t-1} = 0) = \frac{1}{2^{t-1}}$$

De este modo, para $t = 3$ tenemos

$$d(3) \leq \mathbb{P}_0(\tau > 3) = \frac{1}{4}.$$

Debido a la monotonidad de $d(t)$

$$t_{\text{mezcla}} \leq 3.$$

Evidentemente en este caso es fácil obtener los valores exactos de $d(t)$. Sin embargo, a medida que el tamaño de espacios crece se vuelve cada vez menos práctico proceder mediante fuerza bruta hasta el punto que se hace inviable el cálculo de las distancias. Nuestro problema central sobre el barajeo de cartas es un ejemplo perfecto, pues en este caso $|\Omega| = 52!$. En estas situaciones sale a relucir el verdadero poder de esta técnica.

Hasta ahora hemos hablado únicamente sobre cotas superiores al tiempo de mezcla. Pasemos a discutir brevemente las cotas inferiores.

Existen varias técnicas que nos permiten acotar inferiormente a la función $d(t)$. La que es pertinente para nuestro problema de barajar cartas es la siguiente. Sea (X_t) una cadena de Markov irreducible y aperiódica con distribución estacionaria π uniforme. Definimos a $\deg(x)$ como el total de estados que son accesibles desde x en un paso

$$\deg(x) := |\{y : P(x, y) > 0\}|.$$

Así mismo definimos el máximo número de estados accesibles por la cadena en un paso

$$\Delta := \max_{x \in \Omega} \deg(x). \quad (1.49)$$

Escribimos Ω_t^x para denotar al conjunto de estados que son accesibles desde x en t pasos. Es fácil ver que se cumple $|\Omega_t^x| \leq \Delta^t$. Notemos que tenemos las siguientes relaciones

$$P^t(x, \Omega_t^x) = 1 \quad \text{y} \quad \pi(\Omega_t^x) = \frac{|\Omega_t^x|}{|\Omega|} \leq \frac{\Delta^t}{|\Omega|}. \quad (1.50)$$

La segunda igualdad ocurre al considerar que π es la distribución uniforme en Ω . Si $\Delta^t < (1 - \varepsilon)|\Omega|$ entonces gracias a la definición de distancia de variación total y a (1.50) tenemos

$$\|P^t(x, \cdot) - \pi\|_{VT} \geq P^t(x, \Omega_t^x) - \pi(\Omega_t^x) \geq 1 - \frac{\Delta^t}{|\Omega|} > \varepsilon.$$

Por lo tanto si $t < \frac{\log((1-\varepsilon)|\Omega|)}{\log \Delta}$ entonces tenemos $d(t) > \varepsilon$. De aquí concluimos

$$t_{\text{mezcla}}(\varepsilon) \geq \frac{\log((1 - \varepsilon)|\Omega|)}{\log \Delta}. \quad (1.51)$$

Observemos que esta técnica se basa en la idea intuitiva de que si tras t pasos los posibles estados de la cadena no forman una parte significativa del espacio Ω , entonces la cadena no puede estar cerca de parecerse a la distribución uniforme.

Capítulo 2

Barajar

En el capítulo anterior se han discutido diversas cuestiones generales para procesos estocásticos con espacio de estados finito. La finalidad de este capítulo es desarrollar los conceptos que nos permitan analizar el barajar cartas y con ellos estudiar el tiempo de mezcla de un primer modelo conocido como transposición aleatoria. Este consiste en seleccionar dos cartas de la pila al azar e intercambiar sus posiciones en cada iteración. Con este objetivo en mente introducimos la cadena de Markov que será fundamental para el modelado matemático de esta cuestión.

2.1. Caminata aleatoria en grupos

Tomemos una sucesión de variables aleatorias (Y_t) independientes e idénticamente distribuidas definidas en \mathbb{Z} , tal que $\mathbb{E}[|Y_i|] < \infty$. Sea $x_0 \in \mathbb{Z}$, escribimos

$$X_0 = x_0 \quad \text{y} \quad X_t = \sum_{i=1}^t Y_i \quad \text{para } t \geq 1. \quad (2.1)$$

La sucesión (X_t) define una cadena de Markov en $\Omega = \mathbb{Z}$ y es conocida como ***caminata aleatoria***. Las probabilidades de transición están dadas por la siguiente igualdad

$$P(x, y) = \mathbb{P}(Y_i = y - x) \quad \text{para toda } x, y \in \Omega.$$

Estos procesos estocásticos son de gran utilidad para modelar diversos fenómenos. El ejemplo más sencillo es cuando las variables aleatorias Y_i solo pueden tomar el valor 1 o -1 . Así Y_i se puede pensar como una apuesta donde se puede ganar o perder una unidad, y X_t representa el capital de una persona después de realizar la apuesta t . Este proceso es llamado caminata aleatoria simple y tenemos

$$Y_i = \begin{cases} 1 & \text{con probabilidad } p \\ -1 & \text{con probabilidad } 1 - p \end{cases} \quad \text{para } i \geq 1. \quad (2.2)$$

Con probabilidades de transición dadas por

$$P(x, y) = \begin{cases} p & \text{si } y = x + 1 \\ 1 - p & \text{si } y = x - 1 \\ 0 & \text{cualquier otro caso.} \end{cases}$$

Evidentemente el dominio de Y_i puede extenderse a cualquier subconjunto de los números enteros. En cuyo caso la caminata aleatoria es el movimiento en una gráfica de una partícula que se mueve sobre los enteros del eje y . Recordemos que \mathbb{Z} con la operación suma forma un grupo. Al igual que en este clásico ejemplo es posible extender la noción de caminata aleatoria a un grupo arbitrario.

Presentamos a continuación las nociones más importantes de teoría de grupos por considerarlas esenciales para los objetivos de este trabajo. Usaremos como referencia el clásico libro de Herstein [14] al que referimos para un estudio más profundo de estos temas.

Un conjunto no vacío G se dice que forma un grupo si en él se define una operación binaria, llamada producto y denotada por “ \cdot ”, tal que cumpla las siguientes condiciones:

1. Sean $a, b \in G$ entonces $a \cdot b \in G$ (cerradura).
2. Para $a, b, c \in G$ se tiene que $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ (ley asociativa).
3. Existe un único elemento $\text{id} \in G$ tal que $a \cdot \text{id} = \text{id} \cdot a = a$, para todo $a \in G$ (elemento identidad).
4. Para cada $a \in G$ existe un único elemento $a^{-1} \in G$ tal que $a \cdot a^{-1} = a^{-1} \cdot a = \text{id}$ (elemento inverso).

Con esta definición en mente podemos a su vez dar la definición de una caminata aleatoria en un grupo.

Sea μ una distribución de probabilidad definida sobre los elementos del grupo (G, \cdot) . Una ***caminata aleatoria en un grupo G con distribución de incremento μ*** es el proceso con espacio de estados $\Omega = G$, el cual transcurre mediante el producto (por la izquierda) del estado actual con un elemento aleatorio, escogido según la distribución μ . Más precisamente, la matriz de transición en un paso P tiene entradas dadas por la siguiente ecuación

$$P(g, h) = \mu(h \cdot g^{-1}) \quad \text{para todo } g, h \in G. \quad (2.3)$$

En otras palabras, sean g_1, g_2, \dots , elementos aleatorios de G escogidos según μ de forma independiente. Escribimos

$$\begin{aligned} X_0 &= g_0 \\ X_1 &= g_1 \cdot g_0 \\ &\vdots \\ X_t &= g_t X_{t-1} = g_t \cdots g_1 \cdot g_0. \end{aligned}$$

La sucesión (X_t) es una caminata aleatoria en el grupo G con distribución de incremento μ , partiendo de un estado inicial g_0 escogido según la distribución inicial. Nosotros elegimos la convención de operar por la izquierda, ya que muchos grupos no son conmutativos (abelianos).

Observemos que en la definición de la caminata aleatoria clásica (2.1) estamos considerando al grupo abeliano infinito $(\mathbb{Z}, +)$. Más precisamente, en la caminata aleatoria simple la distribución de incremento está descrita por (2.2).

Las caminatas aleatorias en un mismo grupo (G, \cdot) varían para elecciones distintas de distribuciones de incrementos. A pesar de ello, un hecho sorprendente es que todas ellas comparten una misma distribución estacionaria.

Proposición 2.1. *Sea P la matriz de transición de una caminata aleatoria en un grupo finito G con distribución de incremento μ . Entonces la distribución uniforme U definida en G es una distribución estacionaria.*

Demostración. Vamos a probar que la distribución uniforme U satisface la ecuación $U = UP$. Dada la distribución μ , para cualquier $g \in G$ se tiene

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

puesto que al ser G un grupo sabemos que para todo $h \in G$ existe un elemento $k \in G$ tal que $k \cdot h = g$, y también $h = k^{-1} \cdot g$. Es válido volver a indexar la suma mediante k , debido a que para cada h el elemento k es único. Empleando la ecuación (2.3) obtenemos la siguiente igualdad

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

Por lo que concluimos

$$\sum_{h \in G} U(h)P(h, g) = U(g) \quad \text{para toda } g \in G.$$

□

Observemos que para dar la prueba no ha sido necesario especificar condiciones para la distribución de incremento μ .

Gracias a esta proposición ahora sabemos que U es *una* distribución estacionaria, pero aún no hemos determinado para qué procesos es única y podemos hablar de convergencia. Es decir, necesitamos investigar bajo qué condiciones la caminata aleatoria en un grupo G con distribución de incremento μ es irreducible y aperiódica. Nos ocuparemos primero de estudiar la naturaleza irreducible de estos procesos. Para ello necesitamos considerar algunas ideas adicionales de teoría de grupos.

Con la finalidad de simplificar la notación entenderemos ab como el producto $a \cdot b$. Un subconjunto no vacío H de un grupo G se dice que es un *subgrupo* de G si, y solo si,

1. Si $a, b \in H$ entonces $ab \in H$.
2. Si $a \in H$ entonces $a^{-1} \in H$.

De aquí es inmediato que $\text{id} \in H$. Este concepto nos permite establecer el siguiente hecho. Sea W un subconjunto de G . Denotamos por $\langle W \rangle$ al subgrupo *más pequeño* que contiene a W . Entonces $\langle W \rangle$ contiene obviamente a los elementos de W , pero también a sus inversas y todos los elementos de G que puedan ser escritos como combinaciones de ellos. Decimos que $\langle W \rangle$ es generado por W .

Regresamos a la caminata aleatoria para especificar las condiciones bajo las cuales la cadena de Markov es irreducible.

Proposición 2.2. *Sea μ una distribución de probabilidad definida en G . La caminata aleatoria en el grupo G con distribución de incremento μ es irreducible si, y solo si, el soporte $S = \{g \in G : \mu(g) > 0\}$ de μ genera a G .*

Demostración. Primero supondremos que la cadena es irreducible. Dado un elemento $a \in G$, como la cadena es irreducible existe un $r > 0$ que satisface $P^r(\text{id}, a) > 0$. Esto es equivalente a que exista una sucesión de elementos $s_1, s_2, \dots, s_r \in G$ tal que

$$s_r \dots s_1 \text{id} = s_r \dots s_1 = a \quad \text{con } s_i \in S \text{ para } i = 1, \dots, r.$$

Entonces $a \in \langle S \rangle$ y en consecuencia $G \subset \langle S \rangle$, pero como $\langle S \rangle$ no puede tener más elementos que G concluimos que $\langle S \rangle = G$.

Para probar la implicación contraria supongamos que S genera a G . Consideremos $a, b \in G$, por hipótesis podemos escribir ba^{-1} como un producto de elementos de S y sus inversas. Sea entonces

$$ba^{-1} = u_m u_{m-1} \dots u_1, \tag{2.4}$$

donde algunos u_i pertenecen a S mientras que algunos otros pueden ser elementos inversos. Si esto ocurre nos apoyaremos en el hecho de que todos los elementos de G son de orden finito, esto quiere decir que para cualquier $x \in G$ existe un entero n (que depende de x) tal que $x^n = \text{id}$. Digamos que dentro de la expresión para ba^{-1} , el elemento u_i tiene como inversa a $s_i \in S$. Entonces podemos establecer

$$s_i^{n_i} = \text{id} \iff s_i^{n_i-1} = u_i \quad \text{para alguna } n_i \in \mathbb{Z}^+.$$

Así, cualquier elemento inverso u_i en la expresión de ba^{-1} puede ser sustituido por $s_i^{n_i}$, para alguna n_i positiva, y de esta manera lograr

que ba^{-1} pueda ser escrito como una combinación exclusivamente de elementos de S . Sea la expresión resultante

$$ba^{-1} = s_r s_{r-1} \cdots s_1 \quad \text{con } s_j \in S \text{ para } j = 1, \dots, r.$$

Por lo tanto,

$$\begin{aligned} & P(a, s_1 a) P(s_1 a, s_2 s_1 a) \cdots P(s_{r-1} \cdots s_1 a, (ba^{-1})a) \\ &= \mu(s_1) \mu(s_2) \cdots \mu(s_r) > 0. \end{aligned}$$

□

Nombramos a un conjunto S que genera a G *simétrico* si $s \in S$ implica que $s^{-1} \in S$. Esta propiedad tiene consecuencias interesantes en la caminata aleatoria sobre G . De modo similar, llamamos a la distribución μ en el grupo G *simétrica* si cumple

$$\mu(g) = \mu(g^{-1}) \quad \text{para toda } g \in G. \quad (2.5)$$

Proposición 2.3. *La caminata aleatoria en G con distribución de incremento μ es reversible si μ es simétrica.*

Demostración. Sea U la distribución uniforme definida en G . Para cualquier $g, h \in G$, obtenemos las siguientes ecuaciones

$$\begin{aligned} U(g)P(g, h) &= \frac{1}{|G|} \mu(hg^{-1}) \\ U(h)P(h, g) &= \frac{1}{|G|} \mu(gh^{-1}). \end{aligned}$$

Como vimos en el capítulo anterior, una cadena de Markov es reversible si satisface las ecuaciones de balance. En este caso son

$$U(g)P(g, h) = U(h)P(h, g), \quad \text{para toda } g, h \in \Omega.$$

pero esto sucede si, y solo si,

$$\mu(hg^{-1}) = \mu((hg^{-1})^{-1}) = \mu(gh^{-1}).$$

□

2.2. Tiempo reverso y tiempos de mezcla

Dada una distribución μ en un grupo G la *distribución inversa* $\hat{\mu}$ se define como

$$\hat{\mu}(g) := \mu(g^{-1}), \quad \text{para toda } g \in G. \quad (2.6)$$

Observemos que si μ es simétrica entonces $\hat{\mu} = \mu$.

Sea P la matriz de transición de una caminata aleatoria irreducible en un grupo G con distribución de incremento μ , entonces la caminata aleatoria en G con distribución de incremento $\hat{\mu}$ define el tiempo reverso y por lo tanto tiene matriz \hat{P} . Siguiendo la definición de tiempo reverso dada en (1.16) y considerando que la distribución estacionaria π es la distribución uniforme, para toda $g, j \in G$ se cumple

$$\hat{P}(g, j) = \hat{\mu}(jg^{-1}) = \mu((jg^{-1})^{-1}) = \frac{\pi(j)P(j, g)}{\pi(g)} = P(j, g). \quad (2.7)$$

Cuando la distribución de incremento μ es simétrica la Proposición 2.3 nos dice que la caminata es reversible como puede verse fácilmente al obtener $\hat{P}(g, j) = P(g, j)$ en la ecuación (2.7). Sin embargo, aún cuando la caminata aleatoria no sea reversible la distancia con la distribución estacionaria es la misma para P y \hat{P} .

Lema 2.1. *Sea P la matriz de transición de una caminata aleatoria en un grupo G con distribución de incremento μ y sea \hat{P} el tiempo reverso de esta caminata. Sea π la distribución uniforme en G . Para cualquier $t \geq 0$*

$$\|P^t(\text{id}, \cdot) - \pi\|_{VT} = \|\hat{P}^t(\text{id}, \cdot) - \pi\|_{VT}.$$

Demostración. Escribimos $(X_t) = (\text{id}, X_1, \dots)$ para la cadena de Markov con matriz P y estado inicial id . Podemos expresar

$$X_k = g_k \cdots g_1,$$

donde los elementos aleatorios $g_1, \dots, g_k \in G$ son escogidos independientemente y según la ley μ . Así mismo, sea (\hat{X}_t) un proceso con matriz \hat{P} , con incrementos aleatorios independientes h_1, h_2, \dots escogidos según $\hat{\mu}$. Para cualquier $a_1, a_2, \dots, a_t \in G$ fijos

$$\begin{aligned} \mathbb{P}(X_0 = \text{id}, X_1 = a_1, X_2 = a_2 a_1, \dots, X_t = a_t a_{t-1} \cdots a_1) \\ &= \mathbb{P}(g_1 = a_1, g_2 = a_2, \dots, g_t = a_t) \\ &= \mu(a_1) \mu(a_2) \cdots \mu(a_t) \\ &= \hat{\mu}(a_t^{-1}) \hat{\mu}(a_{t-1}^{-1}) \cdots \hat{\mu}(a_1^{-1}) \\ &= \mathbb{P}(h_1 = a_t^{-1}, h_2 = a_{t-1}^{-1}, \dots, h_t = a_1^{-1}). \end{aligned}$$

Por lo tanto,

$$\mathbb{P}(X_0 = \text{id}, \dots, X_t = a_t \cdots a_1) = \mathbb{P}(\hat{X}_0 = \text{id}, \dots, \hat{X}_t = a_1^{-1} \cdots a_t^{-1}).$$

Notemos que $a = a_t \cdots a_1 \iff a^{-1} = a_1^{-1} \cdots a_t^{-1}$. Puede haber más formas de pasar de la id al elemento a en t pasos. Sumamos sobre todas las cadenas tal que $a_t \cdots a_1 = a$ para obtener

$$P^t(\text{id}, a) = \hat{P}^t(\text{id}, a^{-1}).$$

Por lo tanto

$$\begin{aligned} \sum_{a \in G} |P^t(\text{id}, a) - |G|^{-1}| &= \sum_{a \in G} |\hat{P}^t(\text{id}, a^{-1}) - |G|^{-1}| \\ &= \sum_{a \in G} |\hat{P}^t(\text{id}, a) - |G|^{-1}|, \end{aligned}$$

y en virtud de la Proposición 1.5

$$\|P^t(\text{id}, \cdot) - \pi\|_{VT} = \|\hat{P}^t(\text{id}, \cdot) - \pi\|_{VT}.$$

□

El razonamiento usado en la demostración del lema nos permite exponer el siguiente hecho. Sea $(X_t) = (\text{id}, X_1, \dots)$ una caminata aleatoria en un grupo G con distribución de incremento μ y estado inicial id . Sea $(X'_t) = (h, X'_1, \dots)$ la cadena con estado inicial $h \in G$. Para cualquier g_1, \dots, g_t con $g_i \in G$ tenemos

$$\begin{aligned} \mathbb{P}(X'_0 = h, X'_1 = g_1 h, \dots, X'_t = g_t \dots g_1 h) \\ = \mu(g_1) \mu(g_2) \dots \mu(g_t) \\ = \mathbb{P}(X_0 = \text{id}, X_1 = g_1, \dots, X_t = g_t \dots g_1). \end{aligned}$$

Sea $a = g_t \dots g_1 h$, esto equivale a $ah^{-1} = g_t \dots g_1$. Sumando sobre todas las cadenas $g_t \dots g_1$ que permitan obtener ah^{-1} resulta

$$P^t(\text{id}, ah^{-1}) = P^t(h, a).$$

Lo importante de este razonamiento es notar que todos los renglones de P^t tienen los mismos elementos, aunque evidentemente no dispuestos de la misma manera sobre las entradas. En efecto, expresemos el producto ah^{-1} mediante $k = ah^{-1}$. Observemos que al variar a , k debe variar también pues

$$a_1 h^{-1} = a_2 h^{-1} \implies a_1 = a_2.$$

La validez de este hecho se mantiene para toda $h \in G$. Usando la caracterización de la distancia de variación total vista en (1.5) podemos concluir

$$\|P^t(\text{id}, \cdot) - \pi\|_{VT} = \|P^t(h, \cdot) - \pi\|_{VT} \quad \text{para toda } h \in G, \quad (2.8)$$

donde π es la distribución uniforme en G .

Denotemos por $d(t)$ a la distancia asociada a la caminata aleatoria en G con distribución de incremento μ , de la misma manera escribimos

$d'(t)$ para la distancia asociada a la caminata inversa. En virtud de lo expuesto en las líneas anteriores y del Lema 2.1 tenemos

$$d(t) = d'(t) \quad (2.9)$$

y con ello podemos establecer el siguiente corolario.

Corolario 2.1. *Si t_{mezcla} es el tiempo de mezcla de una caminata aleatoria en un grupo y \hat{t}_{mezcla} es el tiempo de mezcla de la caminata inversa, entonces $t_{mezcla} = \hat{t}_{mezcla}$.*

2.3. Caminata aleatoria en el grupo simétrico

Un primer acercamiento al modelado matemático del problema que es objeto de esta tesis es estudiar el proceso de barajar un mazo de n cartas mediante la transposición aleatoria. Con esa finalidad aprovecharemos la estructura algebraica de los grupos, enfocándonos particularmente en uno de ellos.

Denotamos por \mathcal{S}_n al conjunto que contiene *todas* las funciones uno-a-uno de un conjunto $S = \{x_1, \dots, x_n\}$ sobre sí mismo. Este conjunto \mathcal{S}_n satisface las cuatro propiedades de un grupo bajo la operación de composición de funciones que denotamos como es usual por “ \circ ”. El grupo (\mathcal{S}_n, \circ) recibe el nombre de **grupo simétrico** de grado n . Es importante mencionar que este grupo no es abeliano.

Una manera muy útil de pensar a \mathcal{S}_n es considerarla como el grupo constituido por todas las permutaciones de un conjunto S con n elementos. De esta manera $|S| = n$ y por lo tanto $|\mathcal{S}_n| = n!$. Supongamos $\phi \in \mathcal{S}_n$, entonces ϕ es una función uno-a-uno de S sobre sí mismo y podemos escribirla especificando qué es lo que hace a cada elemento de S . Por ejemplo, si $S = \{x_1, x_2, x_3, x_4\}$ escribimos $\phi : x_1 \rightarrow x_2, x_2 \rightarrow x_4, x_3 \rightarrow x_1$ y $x_4 \rightarrow x_3$. Es evidente que esta manera

de escribir funciones no es muy práctica, por ello recurriremos a otra expresión. Retomando el ejemplo, vamos a representar a ϕ de la siguiente manera

$$\begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ x_2 & x_4 & x_1 & x_3 \end{pmatrix}.$$

Establecimos que \mathcal{S}_n forma un grupo bajo la composición de funciones. Para dos funciones $\theta, \psi \in \mathcal{S}_n$ adoptaremos la convención de entender $\psi \circ \theta$ como primero aplicar la función θ y después ψ . Sin embargo, por simplicidad denotaremos a $\psi \circ \theta$ como $\psi\theta$. Así supongamos que

$$\theta = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ x_3 & x_1 & x_2 & x_4 \end{pmatrix} \quad \text{y} \quad \psi = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ x_1 & x_3 & x_2 & x_4 \end{pmatrix}.$$

Entonces la composición $\psi\theta$ está dada por

$$\psi\theta = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ x_3 & x_1 & x_2 & x_4 \end{pmatrix} \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ x_1 & x_3 & x_2 & x_4 \end{pmatrix} = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ x_2 & x_1 & x_3 & x_4 \end{pmatrix}.$$

Esta es la manera en la que vamos a multiplicar estas matrices. Con estas ideas en mente estamos en condiciones de trasladar el barajar cartas a operaciones en el grupo simétrico.

Una pila de n cartas puede ser visto como un elemento del grupo simétrico \mathcal{S}_n del conjunto $S = [n]$, donde $[n] = \{1, \dots, n\}$. Cada permutación actúa sobre la posición de las cartas y no sobre sus valores. La identidad del grupo es aquella función que deja las cartas en su lugar, $\text{id}(k) = k$ para $k = 1, \dots, n$. Mientras que para cada $\sigma \in \mathcal{S}_n$ existe una única $\sigma^{-1} \in \mathcal{S}_n$, tal que $\sigma(j) = i$ si, y solo si, $\sigma^{-1}(i) = j$ para cualquier $i, j \in S$. Es decir, existe una permutación inversa.

Pensemos que para cada tiempo t aplicamos a la pila de n cartas una permutación σ , elegida según una distribución μ definida en \mathcal{S}_n .

De este modo definimos una *caminata aleatoria en el grupo simétrico* S_n *con distribución de incremento* μ . En otras palabras, este proceso modela barajar una pila de n cartas paso a paso. Por ejemplo, consideremos aplicar a una pila de 4 cartas la permutación $\sigma \in S_4$ dada por

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 2 & 4 \end{pmatrix}.$$

Esto es mover la primera carta a la tercera posición con lo que se recorre el orden de la segunda y tercera carta hacia arriba, mientras que la última carta permanece en su lugar.

Otra manera útil de escribir permutaciones es utilizando la *notación por ciclos*. Mediante ella, (abc) se refiere a la permutación que efectúa $\sigma(a) = b$, $\sigma(b) = c$, $\sigma(c) = a$, dejando todas las demás cartas en la misma posición. Cuando varios ciclos se realizan de forma consecutiva estos deben realizarse uno a la vez de derecha a izquierda, para ser consistentes con la forma en la que hemos entendido la composición de funciones. De este modo

$$(13)(12) = (123) = \begin{pmatrix} 1 & 2 & 3 & 4 & \dots & n \\ 2 & 3 & 1 & 4 & \dots & n \end{pmatrix}. \quad (2.10)$$

Primero intercambiamos la primera y segunda carta, luego la primera y tercera carta. La pila resultante es aquella donde la carta superior toma la tercera posición empujando a la segunda hasta arriba y la tercera a la segunda posición. En este ejemplo (123) es un 3-ciclo, mientras que (12) y (13) son 2-ciclos también llamado *transposición*. Notemos la siguiente cuestión. Un m -ciclo puede ser escrito como un producto de $m - 1$ transposiciones

$$(a_1 a_2 \dots a_m) = (a_1 a_m) \dots (a_1 a_3) (a_1 a_2). \quad (2.11)$$

En particular tenemos

$$(123 \dots m) = (1m) \dots (13)(12),$$

como se pudo observar en (2.10).

Un resultado muy importante en teoría de grupos (Lema 2.10.1 en [14]) enuncia que cualquier permutación σ puede ser expresada de forma única como producto de ciclos disjuntos. Más aún, hemos visto que cualquier m -ciclo admite una expresión como producto de 2-ciclos disjuntos. Por lo tanto cualquier permutación σ puede ser escrita mediante un producto únicamente de transposiciones, aunque en este caso la caracterización puede ya no ser única. No obstante, no puede ocurrir que una permutación se pueda expresar como producto de un número par de transposiciones y que también admita una expresión usando un número impar de estos.

Aprovechándonos de este hecho entonces tiene sentido clasificar las permutaciones según esta característica. Si σ se puede expresar usando un número par de transposiciones recibe el nombre de ***permutación par***, en caso contrario es una ***permutación impar***. Esta manera de caracterizar permutaciones crea una partición del conjunto \mathcal{S}_n . Observemos que una permutación (ab) es una *única* transposición. Por lo tanto cualquier transposición es una permutación impar. Ilustremos este concepto tomando $S = \{1, \dots, 6\}$ y la permutación

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 5 & 4 & 3 & 1 & 2 \end{pmatrix}.$$

Los ciclos de esta permutación son (1625) y (34) . Expresando a σ

mediante el producto de estos ciclos resulta

$$\begin{aligned}(34)(1625) &= \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 5 & 3 & 4 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 2 & 4 & 3 & 5 & 6 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 5 & 4 & 3 & 1 & 2 \end{pmatrix}.\end{aligned}$$

Podemos hacer uso de la descomposición para un m -ciclo (2.11) para expresar $(1625) = (15)(12)(16)$ y con ello escribir a σ como producto de transposiciones

$$\sigma = (34)(1625) = (34)(15)(12)(16).$$

De esta manera hemos encontrado una expresión para la permutación σ como un producto de 4 transposiciones por lo que entonces es una permutación par.

A partir de estas ideas establecemos lo siguiente

1. El producto de dos permutaciones par es una permutación par.
2. El producto de una permutación par con una impar resulta en una permutación impar. Sin importar el orden de las permutaciones.
3. El producto de dos permutaciones impar es una permutación par.

Como podemos ver, esta característica de las permutaciones bajo la composición de funciones se comporta de la misma manera que los números pares e impares bajo la adición.

Designemos a A_n como el conjunto que contiene todas las permutaciones pares. En virtud de que la composición de permutaciones pares es par este conjunto es un subgrupo de \mathcal{S}_n y es llamado el **grupo alterno** A_n . Mientras que el conjunto de todas las

permutaciones impares no es un subgrupo, ya que componer dos permutaciones impares nos da una permutación en A_n .

Se han discutido diversas nociones fundamentales sobre \mathcal{S}_n y hemos establecido a la caminata aleatoria en el grupo simétrico con distribución de incremento μ , como la cadena de Markov que modela el proceso de barajar una pila de n cartas. Sin embargo, aún no hemos precisado *cómo* iremos barajando las cartas en cada paso. En otras palabras, no hemos descrito a la distribución de incremento μ .

Este capítulo está dedicado a estudiar el modelo de la transposición aleatoria que como su nombre lo indica consiste en elegir dos cartas al azar e intercambiar su posición. Más precisamente, para cada tiempo t elegimos independientemente dos cartas distintas de la pila, de manera aleatoria y uniforme sobre todas las cartas, e intercambiamos sus posiciones dejando todas las demás cartas en el mismo lugar. Usaremos la notación L_t y R_t para designar a la primera y segunda carta seleccionada, esto viene de pensar que la primera carta se toma con la mano izquierda y la segunda con la derecha. Bajo estas condiciones, sea $T \subseteq \mathcal{S}_n$ el conjunto que contiene a todas las transposiciones. Entonces la distribución de incremento μ es la distribución uniforme definida en T . Con ello la caminata aleatoria se encuentra totalmente descrita. Ahora nos interesa estudiar el número de clases y el periodo.

2.3.1. Irreducibilidad

Aunque rápidamente se puede dar una respuesta a partir del Lema 2.10.2 de [14], pues cualquier permutación se puede escribir como un producto de transposiciones con lo que $\langle T \rangle = \mathcal{S}_n$, creemos que es conveniente dar otro desarrollo. Esto es porque nos servirá para ilustrar técnicas que se usarán más adelante y nos permitirá familiarizarnos más con el lenguaje de cartas.

Consideremos el siguiente algoritmo. Sea σ_0 la permutación identidad. Para cada $k = 1, \dots, n - 1$ vamos a construir σ_k a partir de σ_{k-1} intercambiando las posiciones de las cartas k y J_k , donde J_k es un número aleatorio uniforme en $\{k, \dots, n\}$ y es independiente de $\{J_1, \dots, J_{k-1}\}$. El agudo lector notará que si $\{J_k = k\}$ se estaría efectuando la permutación identidad al no cambiar las posiciones de las cartas, pero id no está propiamente en T al no ser una transposición. Este detalle no es realmente importante al no estar interesados en el tiempo requerido para formar una permutación particular σ_{k-1} tanto como en la posibilidad de construirla. Además, obtener id a partir de transposiciones solo requiere un tiempo adicional puesto que $(ab)(ab) = \text{id}$. Matemáticamente este procedimiento se describe por

$$\sigma_k(i) = \begin{cases} \sigma_{k-1}(i) & \text{si } i \neq J_k, i \neq k \\ \sigma_{k-1}(J_k) & \text{si } i = k \\ \sigma_{k-1}(k) & \text{si } i = J_k \end{cases}. \quad (2.12)$$

Ilustraremos este algoritmo mediante un ejemplo tomando $n = 4$. Iniciar el proceso en σ_0 significa acomodar las cartas en orden descendente, con la carta número 1 hasta arriba mientras la carta número 4 se coloca hasta abajo. Para $k = 1$ supongamos que $J_1 = 3$. Tenemos $\sigma_1(1) = 3$, $\sigma_1(3) = 1$, $\sigma_1(2) = 2$ y $\sigma_1(4) = 4$.

$$\sigma_1 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 4 \end{pmatrix} = (13) \quad (2.13)$$

Supongamos que $J_2 = 3$ cuando $k = 2$. Entonces $\sigma_2(2) = \sigma_1(3)$, $\sigma_2(3) = \sigma_1(2)$, $\sigma_2(1) = \sigma_1(1)$ y $\sigma_2(4) = \sigma_1(4)$.

$$\sigma_2 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 2 & 4 \end{pmatrix} = (132)$$

Finalmente, para $k = 3$ si $J_3 = 4$ tenemos $\sigma_3(3) = \sigma_2(4)$, $\sigma_3(4) = \sigma_2(3)$ y las demás cartas permanecen en su lugar.

$$\sigma_3 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{pmatrix} = (1342)$$

Antes de mostrar que este esquema construye una permutación aleatoria uniforme es conveniente precisar este y otro concepto. Una **permutación aleatoria uniforme** es aquella en la que cada una de las $n!$ posibles permutaciones tienen la misma probabilidad. Por otro lado, una **k -permutación** es una sucesión que contiene k de los n elementos. En total tenemos el siguiente número de k -permutaciones posibles

$$\frac{n!}{(n-k)!} = n(n-1) \cdots (n-k+1).$$

Proposición 2.4. *La permutación σ_{n-1} , obtenida según el procedimiento recursivo (2.12), es una permutación aleatoria uniforme en \mathcal{S}_n .*

Demostración. Para probar la proposición primero demostraremos el siguiente hecho mediante inducción: después de realizar la t transposición todas las t -permutaciones posibles en $\sigma_t(1), \sigma_t(2), \dots, \sigma_t(t)$ son equiprobables.

Para $t = 1$, la primera permutación σ_1 le asigna una posición a la carta 1 de entre las posiciones $\{1, 2, 3, \dots, n\}$ posibles (incluyendo la posibilidad de que permanezca en la primera posición). Esta elección corresponde a la variable J_1 , que como se detalló anteriormente, tiene distribución uniforme. Así pues, la probabilidad de que $\sigma_1(1)$ contenga una 1-permutación dada es

$$\frac{1}{n} = \frac{(n-1)!}{n!}.$$

Supongamos ahora que es válido después de realizar la $t - 1$ transposición. Es decir, cada $(t - 1)$ -permutación posible aparece en $\sigma_{t-1}(1), \dots, \sigma_{t-1}(t - 1)$ con probabilidad $(n - t + 1)!/n!$.

Sea $\{x_1, x_2, \dots, x_{t-1}, x_t\}$ una t -permutación particular. Esta permutación consiste en la $(t - 1)$ -permutación $\{x_1, \dots, x_{t-1}\}$ seguido del elemento x_t colocado en $\sigma_t(t)$. Escribimos E_1 para el evento en el que las primeras $t - 1$ transposiciones dieron lugar a la permutación $\{x_1, \dots, x_{t-1}\}$. Por hipótesis de inducción

$$\mathbb{P}(E_1) = \frac{(n - t + 1)!}{n!}.$$

Sea E_2 el evento en el que la t transposición coloca a x_t en $\sigma_t(t)$. Obtener la permutación $\{x_1, x_2, \dots, x_t\}$ después de la t transposición ocurre cuando sucede tanto E_1 como E_2 . Por lo tanto

$$\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_2|E_1)\mathbb{P}(E_1) = \frac{(n - t + 1)!}{n!} \frac{1}{n - t + 1} = \frac{(n - t)!}{n!}. \quad (2.14)$$

Esto concluye la prueba por inducción.

Finalmente, cuando $t = n - 1$ la probabilidad de que σ_{n-1} sea una permutación $\sigma \in \mathcal{S}_n$ es

$$\frac{(n - n + 1)!}{n!} = \frac{1}{n!}.$$

□

Esta proposición mostró que podemos construir cualquier $\sigma \in \mathcal{S}_n$ a partir de la identidad usando solo transposiciones, luego $\langle T \rangle = \mathcal{S}_n$ y en virtud de la Proposición 2.2 esta caminata aleatoria es irreducible.

2.3.2. Periodicidad

Notemos que el grupo simétrico puede descomponerse de la siguiente manera

$$\mathcal{S}_n = A_n \cup B_n.$$

Donde A_n es el conjunto que contiene todas las permutaciones pares, mientras que B_n denota al conjunto que contiene a todas las impares. El soporte de μ es el conjunto de todas las transposiciones T y sabemos que todas ellas son permutaciones impares con lo que $T \subseteq B_n$. Si al tiempo t estamos en el estado $\sigma_t \in A_n$ entonces al tiempo $t + 1$ solo podemos ir a un estado $\sigma_{t+1} \in B_n$. Más precisamente, iniciando en la identidad σ_0 después de un número par (impar) de pasos estaremos en una permutación par (impar). El proceso (X_t) va a ir alternando sucesivamente entre ambos conjuntos. En este sentido decimos que la cadena de Markov tiene un comportamiento con periodo 2.

Para zanjar este problema tenemos que asegurarnos que $\mu(\text{id}) > 0$, ya que en general para una caminata aleatoria irreducible en G con distribución de incremento μ se cumple lo siguiente para cualquier $g \in G$

$$\mu(\text{id}) = P(g, \text{id}g) = P(g, g) > 0.$$

Por lo tanto $1 \in \{t : P^t(g, g) > 0\}$ y la cadena es entonces aperiódica.

Incorporar esta idea para asegurar que nuestro proceso de barajar sea aperiódico requiere que hagamos una ligera modificación. Al tiempo t elegimos dos cartas aleatoriamente, de manera independiente y uniforme sobre todas las cartas, para la mano izquierda y derecha L_t y R_t . Si las cartas seleccionadas son distintas intercambiamos sus posiciones, de otra

forma no hacemos nada. La distribución de μ entonces está descrita por

$$\mu(\sigma) = \begin{cases} \frac{1}{n} & \text{si } \sigma = \text{id} \\ \frac{2}{n^2} & \text{si } \sigma = (ij) \\ 0 & \text{de otro modo.} \end{cases} \quad (2.15)$$

La segunda probabilidad de (2.15) corresponde a la probabilidad de elegir cartas distintas y que estas efectúen la transposición (ij) . Sea E_1 el evento de elegir dos cartas distintas, esto es realizar cualquier transposición. La probabilidad de este evento es

$$\mathbb{P}(E_1) = 1 - \frac{1}{n} = \frac{n-1}{n}.$$

Sea E_2 el evento que corresponde a elegir la transposición (ij) sobre todas las transposiciones posibles. Observemos que la transposición (ij) es la misma que (ji) , por lo tanto para calcular el total de transposiciones usamos las combinaciones de n en 2

$$|T| = \binom{n}{2} = \frac{n(n-1)}{2}.$$

De esta manera, la probabilidad $\mu((ij))$ equivale a la probabilidad de $\{E_1 \cap E_2\}$ y resulta

$$\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_2|E_1)\mathbb{P}(E_1) = \frac{2}{n(n-1)} \frac{n-1}{n} = \frac{2}{n^2}.$$

2.3.3. Tiempo de mezcla

Acotar el tiempo requerido para que una pila de n cartas esté bien barajada significa acotar el tiempo de mezcla. Cuando la cadena de Markov está cerca de su equilibrio, la distribución del acomodo de las cartas es aproximadamente uniforme sobre todas las permutaciones \mathcal{S}_n . Es decir, las cartas están bien aleatorizadas. En virtud de este fin

emplearemos las ideas que se desarrollaron en el primer capítulo. Iniciaremos dando una cota superior mediante un tiempo fuertemente estacionario acreditado a Andre Broder [2].

Proposición 2.5. *En el proceso de barajar descrito por (2.15) supongamos que al tiempo $t = 0$ ninguna carta ha sido marcada. Al tiempo t marcamos la carta R_t si no ha sido marcada y ocurre:*

$$\text{La carta } L_t \text{ está marcada o } L_t = R_t.$$

Sea τ el tiempo en el cual todas las cartas han sido marcadas. Entonces τ es un tiempo fuertemente estacionario para esta cadena de Markov.

Conviene adelantar una explicación heurística sobre por qué este esquema es un tiempo fuertemente estacionario antes de pasar a la demostración formal.

Una forma de construir una permutación σ aleatoria uniforme en \mathcal{S}_n es armar un mazo de n cartas paso a paso, colocando cada carta en una posición aleatoria uniforme respecto a las demás cartas que ya forman parte del mazo. Este razonamiento no debe resultar extraño, pues es el mismo que se siguió en (2.12) para mostrar que existe una serie de transposiciones para obtener σ aleatoria uniforme. Sin embargo, en general no podemos esperar que se replique la serie de transposiciones usadas en (2.12). El esquema que se describe en esta proposición construye una permutación σ aleatoria uniforme sin requerir una sucesión de transposiciones específica.

Demostración. La prueba se realiza mediante inducción sobre t . Adoptamos la siguiente notación en favor de la claridad de la prueba. Escribimos K_t para denotar el número de cartas que han sido marcadas al tiempo t . Designamos con $\mathcal{V}_t = \{V_1, \dots, V_k\}$ al conjunto de las cartas marcadas al tiempo t , así mismo \mathcal{U}_t es el conjunto de las

posiciones de estas cartas. Sea $\Pi_t : \mathcal{V}_t \rightarrow \mathcal{U}_t$ la biyección del conjunto \mathcal{V}_t en \mathcal{U}_t . Es decir, el ordenamiento de la cartas marcadas dentro de la pila después de la t transposición.

El paso esencial en la demostración es probar que dado K_t, \mathcal{V}_t y \mathcal{U}_t , el orden $\Pi_t : \mathcal{V}_t \rightarrow \mathcal{U}_t$ se distribuye uniforme sobre todas las permutaciones de \mathcal{V}_t en \mathcal{U}_t .

La afirmación es válida por vacuidad para $t = 1$ y los primeros tiempos en los que aún no ha sido marcada ninguna carta. Continúa siendo trivialmente válida mientras se haya marcada solo una carta. Esta situación ocurre necesariamente la primera vez que $L_t = R_t$ y entonces \mathcal{V}_t y \mathcal{U}_t contienen un solo elemento. Notemos que bajo este esquema la primera carta marcada (la primera carta que forma la permutación σ) tiene solo una posición posible, a diferencia del proceso (2.12) donde la primera carta de la permutación σ tenía n posibles posiciones. Por lo tanto, podemos pensar que este esquema construye σ al revés que (2.12).

Supongamos que la afirmación es válida para t . En $t+1$ puede ocurrir que se marque una nueva carta o no. Supongamos que sí. Condicionamos en los eventos $K_{t+1} = k + 1, \mathcal{V}_{t+1}$ y \mathcal{U}_{t+1} . Como se ha marcado una nueva carta esto significa que L_{t+1} tiene la misma probabilidad de ser cualquiera de las cartas en $\mathcal{V}_{t+1} = \mathcal{V}_t \cup \{R_{t+1}\}$, y tenemos $\mathcal{U}_{t+1} = \mathcal{U}_t \cup \{p\}$ donde $p = \sigma_t(R_{t+1})$ es la posición al tiempo t de la nueva carta marcada. Dado R_{t+1} , la permutación Π_{t+1} está determinada unívocamente por la elección de Π_t y de L_{t+1} . Esto es más evidente si analizamos un ejemplo. Tenemos $\mathcal{V}_{t+1} = \{V_1, \dots, V_k, R_{t+1}\}$, $\mathcal{U}_{t+1} = \{U_1, \dots, U_k, p\}$. Sea la permutación σ^* dada por

$$\sigma^* = \begin{pmatrix} V_1 & V_2 & \dots & V_k & R_{t+1} \\ U_1 & U_2 & \dots & U_k & p \end{pmatrix}.$$

La probabilidad del evento $\{\Pi_{t+1} = \sigma^*\}$ es el producto de la probabilidades de los eventos $\{\Pi_t = \sigma'\}$ y $\{L_{t+1} = R_{t+1}\}$, donde la permutación σ' es la misma que σ^* sin la última columna. Por hipótesis esto es

$$\mathbb{P}(\Pi_{t+1} = \sigma^*) = \frac{1}{k!} \frac{1}{k+1} = \frac{1}{(k+1)!}. \quad (2.16)$$

Es claro que el mismo razonamiento se sigue sin importar cuál sea la permutación σ^* . En otras palabras, si se marca una nueva carta la permutación Π_{t+1} se distribuye uniforme sobre de todas las permutaciones de \mathcal{V}_{t+1} en \mathcal{U}_{t+1} . Esto sigue siendo cierto si relajamos el condicionamiento en R_{t+1} y en su posición p .

Por otro lado, supongamos que en $t+1$ no se marca una nueva carta, es decir $\mathcal{V}_{t+1} = \mathcal{V}_t$. Existen tres maneras en las que esto puede suceder:

- Si L_{t+1} y R_{t+1} son diferentes y no han sido marcadas entonces, por hipótesis, la permutación Π_{t+1} se mantiene con distribución uniforme.
- Si tanto L_{t+1} y R_{t+1} ya han sido marcadas previamente entonces estamos aplicando una transposición aleatoria sobre el conjunto de cartas marcadas, que por hipótesis ya están en un orden perfectamente aleatorio. Por lo tanto la uniformidad de Π_{t+1} no se altera.
- De otra manera L_{t+1} no ha sido marcada y R_{t+1} sí. En este caso Π_{t+1} sigue siendo uniforme después de sustituir $\sigma_t(R_{t+1})$ con $\sigma_t(L_{t+1})$ en el conjunto \mathcal{U}_t y con ello obtener \mathcal{U}_{t+1} .

En definitiva, independientemente de que en $t+1$ se marque una nueva carta o no, la permutación Π_{t+1} se distribuye de manera uniforme con lo que la afirmación está probada.

Siguiendo la definición del tiempo aleatorio τ este es el momento en el que las n cartas han sido marcadas. Es el tiempo t en el que $\mathcal{V}_t = \mathcal{V}_{t-1} \cup V_n = [n]$. Según lo visto en (2.16) para cualquier permutación $\sigma \in \mathcal{S}_n$ tenemos

$$\mathbb{P}(\sigma_\tau = \sigma | \tau = t) = \mathbb{P}(X_\tau = \sigma | \tau = t) = \frac{1}{n} \frac{1}{(n-1)!} = \frac{1}{n!}.$$

Esto prueba que τ es un tiempo fuertemente estacionario. □

La manera precisa en la que usaremos este tiempo τ para acotar el tiempo de mezcla la dan los siguientes resultados.

Lema 2.2. *El tiempo de paro τ definido en la Proposición 2.5 satisface*

$$\mathbb{E}[\tau] = 2n(\log(n) + O(1))$$

y

$$\text{Var}(\tau) = O(n^2).$$

Demostración. Observemos que podemos descomponer a τ de la siguiente manera

$$\tau = \tau_0 + \tau_1 + \dots + \tau_{n-1}.$$

Donde τ_k es el número de transposiciones después de marcar la carta V_k hasta e incluyendo la transposición en la que se marca la carta V_{k+1} , adoptando la convención de denotar por V_k a la k -ésima carta en ser marcada. Debido al esquema descrito en la Proposición 2.5 las variables τ_k son independientes y podemos afirmar lo siguiente sobre ellas

$$\tau_k \sim \text{Geo} \left(\frac{(k+1)(n-k)}{n^2} \right).$$

En virtud de que $\tau_k \in \{1, 2, \dots\}$ es la cantidad de transposiciones que se necesitaron después de marcar a V_k para marcar la siguiente carta.

Bajo este escenario ya se han marcado k cartas. Para marcar una nueva carta debe suceder que R_t sea alguna de las $n - k$ cartas que no han sido marcadas y que L_t sea alguna de las k cartas marcadas o que sea igual a R_t . Así, la probabilidad de marcar una nueva carta en cada tiempo t (probabilidad de éxito) es

$$p = \frac{n - k}{n} \frac{k + 1}{n} = \frac{(k + 1)(n - k)}{n^2}. \quad (2.17)$$

Encontrando la esperanza de τ

$$\mathbb{E}[\tau] = \mathbb{E} \left[\sum_{k=0}^{n-1} \tau_k \right] = \sum_{k=0}^{n-1} \mathbb{E}[\tau_k] = \sum_{k=0}^{n-1} \frac{n^2}{(k + 1)(n - k)}. \quad (2.18)$$

Usando la descomposición en fracciones parciales

$$\frac{1}{(k + 1)(n - k)} = \frac{1}{n + 1} \left(\frac{1}{k + 1} + \frac{1}{n - k} \right)$$

en (2.18), resulta

$$\begin{aligned} \mathbb{E}[\tau] &= \frac{n^2}{n + 1} \sum_{k=0}^{n-1} \left(\frac{1}{k + 1} + \frac{1}{n - k} \right) \\ &= \frac{n^2}{n + 1} 2 \sum_{k=1}^n \frac{1}{k} \\ &= 2n \frac{n}{n + 1} \sum_{k=1}^n \frac{1}{k}. \end{aligned} \quad (2.19)$$

De la comparación de la integral de $1/x$ con sus sumas de Riemann se puede mostrar la desigualdad $\log(n) \leq \sum_{k=1}^n k^{-1} \leq \log(n) + 1$, equivalente a $0 \leq \sum_{k=1}^n k^{-1} - \log(n) \leq 1$. Esto nos permite establecer

$$\sum_{k=1}^n \frac{1}{k} = \log(n) + O(1).$$

Usando este hecho en (2.19) obtenemos la igualdad buscada.

Por otro lado, calculando la varianza y debido a la independencia de las variables escribimos

$$\text{Var}(\tau) = \sum_{k=0}^{n-1} \text{Var}(\tau_k) = \sum_{k=0}^{n-1} \frac{1-p}{p^2} = \sum_{k=0}^{n-1} \frac{n^2(n^2 - (k+1)(n-k))}{((k+1)(n-k))^2}, \quad (2.20)$$

donde p es el parámetro de la distribución geométrica (2.17). Observemos que $(n^4 - n^2(k+1)(n-k)) < n^4$ para $k = 0, 1, \dots, n-1$. Luego acotamos de la siguiente manera

$$\sum_{k=0}^{n-1} \frac{n^2(n^2 - (k+1)(n-k))}{((k+1)(n-k))^2} < \sum_{k=0}^{n-1} \frac{n^4}{(k+1)^2(n-k)^2}. \quad (2.21)$$

Dividimos la suma del lado derecho de (2.21) en las siguientes dos sumas

$$\begin{aligned} & \sum_{k=0}^{n-1} \frac{n^4}{(k+1)^2(n-k)^2} = \\ & \sum_{0 \leq k < n/2} \frac{n^4}{(k+1)^2(n-k)^2} + \sum_{n/2 \leq k < n} \frac{n^4}{(k+1)^2(n-k)^2}. \end{aligned} \quad (2.22)$$

Cuando n es un número par entonces ambas sumas tienen la misma cantidad de términos. Si n es impar entonces la primera suma tiene un término más. Notemos que cuando $k = 0$ en la primera suma y $k = n-1$ en la segunda los términos son los mismos. Los términos duplicados continúan correspondiéndose hasta $k = \lfloor \frac{n}{2} - 1 \rfloor$ en la primera suma y $k = \lceil \frac{n}{2} \rceil$. De modo que establecemos la siguiente desigualdad para (2.22)

$$\begin{aligned} & \sum_{0 \leq k < n/2} \frac{n^4}{(k+1)^2(n-k)^2} + \sum_{n/2 \leq k < n} \frac{n^4}{(k+1)^2(n-k)^2} \\ & \leq 2 \sum_{0 \leq k < n/2} \frac{n^4}{(k+1)^2(n-k)^2}. \end{aligned}$$

La igualdad se verifica cuando n es par. Por lo tanto

$$\text{Var}(\tau) < 2 \sum_{0 \leq k \leq n/2} \frac{n^4}{(k+1)^2(n-k)^2}.$$

Debido a que $(n-k)^2$ se encuentra en el denominador de cada término estos son mayores cuando $(n-k)^2$ es menor. El mínimo lo alcanza cuando $k = n/2$ y toma el valor de $(n/2)^2$. De modo que podemos acotar de la siguiente manera

$$\text{Var}(\tau) < 2 \sum_{0 \leq k \leq n/2} \frac{n^4}{(k+1)^2(n-k)^2} < 2 \frac{n^4}{(n/2)^2} \sum_{0 \leq k \leq n/2} \frac{1}{(k+1)^2}. \quad (2.23)$$

Es muy conocido la convergencia de la serie de los cuadrados inversos $\sum_{k=1}^{\infty} k^{-2}$. La suma del lado derecho de (2.23) se trata de una suma parcial de esta serie por lo que entonces es $O(1)$. El término restante es $O(n^2)$ y finalmente tenemos

$$\text{Var}(\tau) = O(n^2).$$

□

Proposición 2.6. *Para el proceso de barajar una pila de n cartas descrito en (2.15), tenemos*

$$t_{mezcla} \leq (2 + o(1))n \log(n).$$

Demostración. Sea τ el tiempo fuertemente estacionario definido en la Proposición 2.5. Sea $t_0 = \mathbb{E}[\tau] + 2\sqrt{\text{Var}(\tau)}$. De la desigualdad de Chebyshev tenemos

$$\mathbb{P}\left(\tau - \mathbb{E}[\tau] \geq 2\sqrt{\text{Var}(\tau)}\right) + \mathbb{P}\left(\tau - \mathbb{E}[\tau] \leq -2\sqrt{\text{Var}(\tau)}\right) \leq \frac{1}{4} \quad (2.24)$$

si, y solo si,

$$\mathbb{P}(\tau \geq t_0) \leq \frac{1}{4}. \quad (2.25)$$

Por otro lado, a partir del Lema 2.2 tenemos

$$t_0 = \mathbb{E}[\tau] + 2\sqrt{\text{Var}(\tau)} = 2n \log(n) + O(n).$$

La definición de una función $O(n)$ nos dice que n es una cota superior para el comportamiento asintótico al infinito. Formalmente se tiene que la función es menor o igual que Mn a partir de algún entero n_0 donde M es un número real positivo. Por esta razón el límite al infinito del cociente de la función $O(n)$ entre $n \log(n)$ es 0 y podemos escribir

$$t_0 = (2 + o(1))n \log(n),$$

donde o es la notación *pequeña o* propuesta por Landau.

La Proposición 1.8 evaluada en $t = t_0$ junto con la desigualdad (2.25) resulta en

$$d(t_0) \leq \mathbb{P}(\tau > t_0) \leq \frac{1}{4}.$$

En el primer capítulo mostramos que $d(t)$ es una función decreciente. Por lo tanto t_{mezcla} es a lo más igual a t_0 , esto es

$$t_{mezcla} \leq (2 + o(1))n \log(n).$$

□

La objeción que se puede hacer al esquema de Broder es el marcar cartas únicamente con la mano derecha. Existe un algoritmo que incorpora ciertas modificaciones llamado *Matthew's scheme* [19] que solventa este problema. Al marcar cartas con ambas manos esperaríamos que la cota superior se ajustara a la mitad,

$$t_{mezcla} \leq (1 + o(1))n \log(n).$$

Resulta que el esquema de Broder también es considerablemente lento al inicio debido a la naturaleza misma del mecanismo. El Teorema 1.4 en [6]

prueba que el tiempo fuertemente estacionario propuesto por Matthew da la siguiente cota superior para n suficientemente grande

$$t_{mezcla} \leq (1 + o(1)) \frac{1}{2} n \log(n). \quad (2.26)$$

Ahora nos ocuparemos de encontrar una cota inferior para t_{mezcla} , o de manera más general para $t_{mezcla}(\varepsilon)$.

Proposición 2.7. *Sea $0 < \varepsilon < 1$. Para el proceso de barajar una pila de n cartas descrito en (2.15), tenemos*

$$t_{mezcla}(\varepsilon) \geq \frac{n-1}{2} \log \left(\frac{1-\varepsilon}{6} n \right).$$

Demostración. Sea $F(\sigma)$ el número de puntos fijos de la permutación σ respecto a id. Si σ es una permutación aleatoria uniforme en \mathcal{S}_n entonces el número esperado de puntos fijos es

$$\mathbb{E}[F(\sigma)] = \mathbb{E} \left[\sum_{i=1}^n \mathbf{1}_{\{\sigma(i)=i\}} \right] = \sum_{i=1}^n \mathbb{P}(\sigma(i) = i) = \sum_{i=1}^n \frac{1}{n} = 1. \quad (2.27)$$

Donde $\mathbf{1}_{\{\sigma(i)=i\}}$ es la indicadora del evento donde la permutación σ tiene fija la carta i .

Obtener σ después de aplicar t transposiciones partiendo de la permutación identidad significa que $F(\sigma)$ es la suma de dos variables: el número de cartas que no han sido tocadas por alguna transposición que denotaremos por N_t , y el número de cartas que fueron movidas pero que eventualmente regresaron a su posición original. Estudiemos más a detalle la primera de estas variables aleatorias.

Para el número de cartas que no han sido tocadas después de t transposiciones tenemos

$$N_t = \sum_{j=1}^n \mathbf{1}_j(t).$$

Donde $\mathbf{1}_j(t)$ es la indicadora del evento donde la carta j no ha sido tocada después de t transposiciones. La probabilidad de este evento está dada por

$$p = \left(\left(\frac{n-1}{n} \right)^2 \right)^t = \left(\frac{n-1}{n} \right)^{2t}.$$

Que corresponde a la probabilidad de no escoger la carta j para R_t y L_t en cada uno de los t tiempos.

Al calcular la esperanza resulta

$$\mathbb{E}[N_t] = \mathbb{E} \left[\sum_{j=1}^n \mathbf{1}_j(t) \right] = \sum_{j=1}^n \mathbb{E} [\mathbf{1}_j(t)] = \sum_{j=1}^n p = np = \tilde{\mu}.$$

Mientras que para la varianza tenemos

$$\text{Var}(N_t) = \sum_{j=1}^n \text{Var}(\mathbf{1}_j(t)) + \sum_{i \neq j} \text{Cov}(\mathbf{1}_i(t), \mathbf{1}_j(t)).$$

Las indicadoras tienen correlación negativa. Esto se debe a que

$$\mathbb{E} [\mathbf{1}_i(t) \mathbf{1}_j(t)] = \left(\frac{n-2}{n} \right)^{2t},$$

luego

$$\text{Cov}(\mathbf{1}_i(t), \mathbf{1}_j(t)) = \left(\frac{n-2}{n} \right)^{2t} - \left(\frac{n-1}{n} \right)^{4t} \leq 0.$$

Por lo tanto

$$\text{Var}(N_t) = np(1-p) + \sum_{i \neq j} \text{Cov}(\mathbf{1}_i(t), \mathbf{1}_j(t)) \leq np(1-p). \quad (2.28)$$

Una vez visto algunas características de $F(\sigma)$ cuando σ tiene distintas distribuciones, sea $A = \{\sigma : F(\sigma) \geq \frac{\tilde{\mu}}{2}\}$. Compararemos la probabilidad de A bajo la distribución uniforme π y $P^t(\text{id}, \cdot)$. Es decir, cuando σ es una permutación aleatoria uniforme y cuando σ se obtiene

después de t transposiciones. Si σ es una permutación aleatoria uniforme en \mathcal{S}_n y aplicamos la desigualdad de Markov obtenemos

$$\pi(A) = \mathbb{P}(F(\sigma) \geq \frac{\tilde{\mu}}{2}) \leq \frac{\mathbb{E}[F(\sigma)]}{\tilde{\mu}/2}.$$

Sustituyendo el valor de la esperanza obtenido en (2.27)

$$\pi(A) \leq \frac{2}{\tilde{\mu}}. \quad (2.29)$$

Por otro lado, aplicando a N_t la desigualdad de Chebyshev en un argumento similar a (2.24) resulta

$$\mathbb{P}\left(N_t \leq \tilde{\mu} - k\sqrt{\text{Var}(N_t)}\right) \leq \frac{1}{k^2},$$

y evaluando en $k = \mu/(2\sqrt{\text{Var}(N_t)})$

$$\mathbb{P}\left(N_t \leq \frac{\tilde{\mu}}{2}\right) \leq \frac{4\text{Var}(N_t)}{\tilde{\mu}^2} \leq \frac{4}{\tilde{\mu}}. \quad (2.30)$$

La última cota resulta de considerar la cota superior para $\text{Var}(N_t)$ obtenida en (2.28). Así mismo tenemos

$$P^t(\text{id}, A^c) \leq \mathbb{P}\left(N_t \leq \frac{\tilde{\mu}}{2}\right). \quad (2.31)$$

Esta afirmación es cierta debido a que $P^t(\text{id}, A^c) = \mathbb{P}(F(\sigma) < \frac{\tilde{\mu}}{2})$, donde σ resulta de haber realizado t transposiciones. Pero $F(\sigma)$ es la suma de N_t y el número de cartas que regresaron a su posición original digamos N'_t . Por consiguiente

$$\mathbb{P}\left(F(\sigma) < \frac{\tilde{\mu}}{2}\right) = \mathbb{P}\left(N_t + N'_t < \frac{\tilde{\mu}}{2}\right) = \mathbb{P}\left(N_t < \frac{\tilde{\mu}}{2}, N'_t < \frac{\tilde{\mu}}{2} - N_t\right).$$

Combinando las desigualdades (2.30) y (2.31)

$$P^t(\text{id}, A^c) \leq \frac{4}{\tilde{\mu}}. \quad (2.32)$$

Sumando las desigualdades (2.32) y (2.29) resulta

$$1 - \frac{4}{\tilde{\mu}} + \pi(A) \leq P^t(\text{id}, A) + \frac{2}{\tilde{\mu}},$$

si y solo si,

$$1 - \frac{6}{\tilde{\mu}} \leq P^t(\text{id}, A) - \pi(A).$$

Por la definición de distancia de variación total (1.19), esta es mayor o igual que cualquier diferencia en probabilidad que las distribuciones asignan a un mismo evento

$$P^t(\text{id}, A) - \pi(A) \leq \|P^t(\text{id}, \cdot) - \pi\|_{VT}.$$

Por lo tanto

$$1 - \frac{6}{\tilde{\mu}} \leq \|P^t(\text{id}, \cdot) - \pi\|_{VT}.$$

Así mismo, por definición $\|P^t(\text{id}, \cdot) - \pi\|_{VT} \leq d(t)$ con lo que tenemos

$$1 - \frac{6}{\tilde{\mu}} \leq d(t).$$

Queremos encontrar qué tan pequeño tienen que ser los valores de t para que tengamos $\varepsilon \leq 1 - \frac{6}{\tilde{\mu}}$ y con ello $\varepsilon \leq d(t)$. Sustituyendo el valor de $\tilde{\mu}$

$$\frac{6}{1 - \varepsilon} \leq \tilde{\mu} = n \left(\frac{n-1}{n} \right)^{2t} \iff 2t \log \left(\frac{n}{n-1} \right) \leq \log \left(\frac{n(1-\varepsilon)}{6} \right). \quad (2.33)$$

Usando la desigualdad $\log(1+x) \leq x$ con $x = 1/(n-1)$ resulta

$$2t \log \left(\frac{n}{n-1} \right) \leq \frac{2t}{n-1}.$$

La desigualdad (2.33) siempre se cumple cuando

$$\frac{2t}{n-1} \leq \log \left(\frac{n(1-\varepsilon)}{6} \right),$$

y esto ocurre si y solo si,

$$t \leq \frac{n-1}{2} \log \left(\frac{n(1-\varepsilon)}{6} \right).$$

Por lo tanto si $t \leq \frac{n-1}{2} \log \left(\frac{n(1-\varepsilon)}{6} \right)$ entonces $\varepsilon \leq d(t)$. De manera que al ser $d(t)$ una función decreciente

$$\frac{n-1}{2} \log \left(\frac{n(1-\varepsilon)}{6} \right) \leq t_{mezcla}(\varepsilon).$$

□

2.4. Barajar genes

El orden de los genes dentro de los cromosomas puede cambiar debido a mutaciones que invierten el orden de segmentos enteros de estos genes. Por ejemplo, si representamos los genes de un cromosoma por el ordenamiento

ABCDEFGHIJKLMN**O**PQRSTUVWXYZ,

una mutación puede alterar el orden de los genes en

ABCDEFGHI**N**MLKJIHPQRSTUVWXYZ.

Los biólogos pueden comparar cadenas de genes homogéneos entre dos organismos para estimar la distancia evolutiva. Un ejemplo de ello es el análisis llevado a cabo por Ranz, Sagarra y Ruiz [18] donde se comparó los genes del cromosoma 2 de *Drosophila repleta* con los del cromosoma 3R de *D. melanogaster*. Ambas son especies de mosca ampliamente utilizadas en experimentos genéticos debido al bajo número de cromosomas que poseen. Este estudio motivó al tratamiento matemático de esta cuestión.

Etiquetamos una cadena de genes mediante $[n]$ y suponemos que cualquiera de los $n!$ ordenamientos es posible. La primera modificación al proceso de transposición aleatoria es restringir las transposiciones posibles en cada tiempo, únicamente consideramos las transposiciones adyacentes. Es decir, intercambiamos las posiciones de cartas contiguas. El modelo de transposición adyacente aleatoria ralentiza el tiempo de mezcla y modifica lo suficiente la estructura para requerir otras herramientas con el fin de estudiarlo. Sin embargo, este modelo no es lo suficientemente poderoso.

En el contexto de las mutaciones no todas las inversiones tienen la misma probabilidad, podemos incluir este hecho restringiendo la longitud máxima de los segmentos invertibles a $L(n)$ tal que $1 \leq L(n) \leq n$. Para simplificar tratamos la cadena como un arreglo dispuesto de forma circular de manera que una inversión puede abarcar el final y el principio de la cadena. En cada tiempo elegimos i

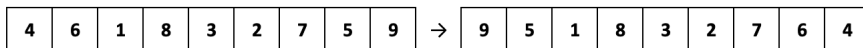


Figura 2.1. Inversión de las cartas de la posición 8 a la 2.

distribuido uniformemente en $[n]$. Así mismo elegimos j distribuido de manera uniforme en $\{0, \dots, L - 1\}$ e invertimos el segmento de la posición i a $i + j$, donde la aritmética se realiza modulo n . La caminata aleatoria descrita es conocida como **cadena L-inversa**.

Nuevamente, para el estudio de los tiempos de mezcla es necesario desarrollar otras técnicas como el método de Wilson y la comparación de trayectorias. El motivo de presentar esta breve exposición es subrayar la importancia de la transposición aleatoria, un primer modelo de barajar

que a primera vista puede parecer simple. El estudio detallado sobre la cadena L-inversa y su tiempo de mezcla puede encontrarse en el artículo de Durrett [\[12\]](#).

Capítulo 3

Barajeo rápido

A lo largo de los capítulos precedentes hemos desarrollado la teoría necesaria para dar respuesta a la principal interrogante de esta investigación: ¿Cuántos barajeos tenemos que realizar para que una pila de cartas esté bien aleatorizada? En el capítulo anterior presentamos la transposición aleatoria. Sin embargo, es evidente que las personas no barajan de esta manera. Este capítulo estará dedicado a desarrollar un modelo matemático que refleje, de manera adecuada, el proceso real de barajar una pila de n cartas y encontrar cotas al tiempo requerido para estar bien barajada.

3.1. El modelo de barajeo rápido

El procedimiento más popular de barajar se conoce como *riffle shuffle* que nosotros llamamos barajeo rápido. Consiste en cortar la pila de cartas en dos subpilas aproximadamente del mismo tamaño y formar una nueva pila con las cartas superiores de ambas subpilas.

Hay dos cuestiones que faltan precisar: el número de cartas en cada

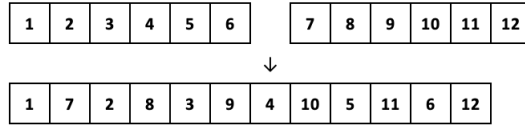


Figura 3.1. Barajeo rápido perfectamente intercalado en 12 cartas

subpila y cómo se van a intercalar las cartas superiores de las subpilas en el nuevo ordenamiento. Existe un modelo matemático que podemos analizar y que considera estas cuestiones. Presentamos tres modelos que permiten barajar cartas a propósito de esta discusión.

Modelo 1. Cortamos la pila de n cartas en dos subpilas. La primera tiene las primeras M cartas de la pila y la segunda las cartas restantes, siendo M una variable aleatoria con distribución binomial($n, 1/2$). La probabilidad refleja el hecho de que ambas subpilas son aproximadamente del mismo tamaño. Al momento de obtener la nueva pila el orden relativo de las dos subpilas se mantiene. Es decir, en la nueva pila la primera carta que encontramos de la primera subpila es la carta superior de esta, la segunda carta que encontramos de esta subpila es precisamente la segunda carta que la constituía, y así sucesivamente. Por supuesto, no esperamos que todas las cartas de la primera subpila se encuentren juntas una tras otra. Debido a la naturaleza del barajeo rápido lo más natural es que las cartas de la primera subpila estén intercaladas con las de la segunda subpila para las cuales también se aplica este mismo principio del orden relativo. Dejar esta explicación únicamente en palabras no es suficiente por lo que vamos a formalizarla. La primera subpila esta compuesta por las cartas con posición $\{1, 2, \dots, M\}$ a las

que se les va a asignar un nuevo ordenamiento conservando el orden relativo. Sean estas nuevas posiciones $\{k_1, \dots, k_M\}$ con $k_i \in [n]$, entonces el barajeo rápido σ se describe por

$$\sigma = \begin{pmatrix} 1 & 2 & \dots & M & M+1 & M+2 & \dots & n \\ k_1 & k_2 & \dots & k_M & j_1 & j_2 & \dots & j_{n-M} \end{pmatrix}.$$

Tal que $\{k_1, \dots, k_M\} \cup \{j_1, \dots, j_{n-M}\} = [n]$, $k_1 < k_2 < \dots < k_M$, y además $j_1 < \dots < j_{n-M}$. Solo queda otra cuestión a resolver dentro de este modelo: dado M ¿cómo intercalamos las cartas de la primera y segunda pila? Notemos que la nueva pila queda totalmente determinada por la manera en la que ordenamos la primera subpila. Por lo tanto, dado M , tenemos un total de $\binom{n}{M}$ de ordenamientos producto de las distintas maneras en las que podemos intercalar las cartas de ambas pilas. Elegimos una de estas de manera aleatoria uniforme.

Modelo 2. Sea M una variable aleatoria binomial($n, 1/2$). La primera subpila tiene las primeras M cartas y la segunda las restantes $n - M$ cartas. La manera en que se forma la nueva pila es distinta a la del primer modelo. Si la primera subpila tiene a cartas y la segunda b cartas, entonces la próxima carta que formará parte de la nueva pila es la carta superior de la primera subpila con probabilidad $a/(a+b)$ y será la superior de la segunda con probabilidad $b/(a+b)$. Repetimos este proceso hasta que todas las cartas han sido colocadas.

Modelo 3. Asignamos a cada una de las n cartas un bit, es decir 0 o 1 de forma aleatoria con la misma probabilidad e independientemente. Colocamos todas las cartas marcadas con 0 hasta arriba de la pila preservando el orden relativo, enseguida ordenamos a las cartas con 1 de la misma manera.

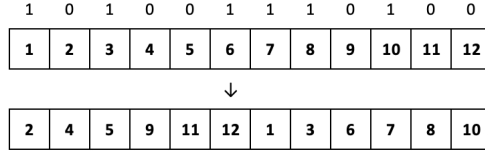


Figura 3.2. Tercer modelo aplicado sobre una pila de 12 cartas.

Antes de estudiar las distribuciones de incremento correspondientes a cada modelo es necesario introducir la siguiente idea central. Una ***sucesión creciente*** de una permutación σ es el conjunto máximo de cartas que se encuentran en un orden relativo correcto. Por ejemplo, en la Figura 3.1 las sucesiones crecientes están dadas por $\{1, 2, 3, 4, 5, 6\}$ y $\{7, 8, 9, 10, 11, 12, 13\}$.

Regresando al análisis de los modelos, afirmamos que el Modelo 1 y 2 generan la misma distribución de incremento Q sobre las permutaciones,

$$Q(\sigma) = \begin{cases} (n+1)/2^n & \text{si } \sigma = \text{id}, \\ 1/2^n & \text{si } \sigma \text{ tiene exactamente dos sucesiones crecientes,} \\ 0 & \text{cualquier otro caso.} \end{cases} \quad (3.1)$$

Explicaremos primero el Modelo 1. Observemos que si aplicamos la permutación identidad todas las primeras M cartas conservan su posición (del mismo modo las $n - M$ cartas inferiores). Si esto ocurre solo tenemos una sucesión creciente debido a que la pila es exactamente la misma que antes de ser barajada. Por otro lado, basta que ocurra al menos un intercalado de una carta de la primera subpila con una de la segunda para romper la sucesión creciente y que tengamos dos de estas, como es claro en la Figura 3.1. Notemos que no puede haber más de dos sucesiones crecientes tras aplicar un barajeo rápido. No olvidemos que hasta ahora nuestra investigación ha

supuesto siempre M dada. Extendamos ahora nuestro análisis. Es tentador querer afirmar que hay un total de 2^n permutaciones distintas posibles tras un barajeo al considerar todas las elecciones posibles de M

$$\sum_{k=0}^n \binom{n}{k} = 2^n. \quad (3.2)$$

No obstante, hay distintas maneras de obtener el elemento id . Para ver esto tomemos por ejemplo $M = 3$. Si la elección de las posiciones de las cartas de la primera subpila es $\{1, 2, 3\}$ entonces estamos aplicando la identidad a las cartas. Para $M = 4$ esta elección es $\{1, 2, 3, 4\}$, y en general para cualquier valor de M siempre existe la posibilidad de conservar el orden de las cartas. Se incluyen también los casos cuando $M = 0$ y $M = n$ para las cuales esto resulta obvio. Por lo tanto la probabilidad del elemento id bajo Q resulta

$$Q(\text{id}) = \sum_{k=0}^n \mathbb{P}(\text{id}|M = k)\mathbb{P}(M = k) = \sum_{k=0}^n \binom{n}{k}^{-1} \binom{n}{k} \left(\frac{1}{2}\right)^n = \frac{n+1}{2^n}.$$

Cuando σ es una permutación específica con dos secuencias crecientes solo existe un valor M y una elección de entre los $\binom{n}{M}$ ordenamientos posibles que nos permiten obtenerla. Luego,

$$Q(\sigma) = \mathbb{P}(\sigma|M = k)\mathbb{P}(M = k) = \binom{n}{k}^{-1} \binom{n}{k} \left(\frac{1}{2}\right)^n = \frac{1}{2^n}.$$

Hemos evidenciado el siguiente hecho: a pesar de que existen 2^n maneras distintas de barajar no todas ellas producen permutaciones distintas. De hecho el total correcto de permutaciones diferentes tras un barajeo rápido es $2^n - n$. La distribución de probabilidad Q definida en \mathcal{S}_n resultante es conocida en la literatura matemática como el Modelo de Gilbert-Shannon-Reeds.

Por otro lado, para ver que el Modelo 2 genera la misma distribución de incremento Q basta verificar que, dado M , la probabilidad de un

ordenamiento determinado es $\binom{n}{M}^{-1}$. Debido al algoritmo por el cual formamos la nueva pila de cartas, al principio $a = M$ y $b = n - M$ con lo que $a + b = n$. Después de elegir la primera carta que formará la nueva pila tenemos $a + b = n - 1$. Así sucesivamente hasta que hemos colocado $n - 1$ cartas momento en el cual $a + b = 1$. De esta manera la probabilidad bajo este modelo de un barajeo rápido σ que intercala perfectamente las primeras dos cartas de las dos subpilas está dada por

$$\frac{M}{n} \frac{n - M}{n - 1} \frac{M - 1}{n - 2} \frac{n - M - 1}{n - 3} \dots$$

En general, ya que todas las cartas van a ser colocadas eventualmente, los factores de la probabilidad de un barajeo rápido σ en específico dentro de los $\binom{n}{M}$ posibles siempre pueden reorganizarse del siguiente modo

$$\mathbb{P}(\sigma|M) = \frac{M(M-1) \dots 1(n-M)(n-M-1) \dots 1}{n(n-1) \dots 1} = \frac{M!(n-M)!}{n!}.$$

Al igual que en el Modelo 1 el elemento id siempre es una opción sin importar el valor M . Este se obtiene al colocar consecutivamente las M cartas de la primera subpila.

La cadena de Markov consiste en iteraciones sucesivas de barajeos rápidos que pueden ser expresadas usando convoluciones

$$Q^*Q(\sigma_2) = \sum_{\sigma \in \mathcal{S}_n} Q(\sigma_2\sigma^{-1})Q(\sigma),$$

donde $*$ denota la convolución. Así la probabilidad que después de 2 barajeos rápidos el orden de la pila sea σ_2 se calcula primero con la probabilidad de elegir σ y después elegir el elemento que resulte en σ_2 . De manera análoga $Q^{*t}(\sigma_2) = \sum_{\sigma} Q^{*(t-1)}(\sigma_2\sigma^{-1})Q(\sigma)$. De esta manera podemos expresar la probabilidad del estado de la cadena después de t pasos, usando la notación (1.5) tenemos

$$Q^{*t}(\sigma) = \mathbb{P}(X_t = \sigma | X_0 = \text{id}) = \mu_t(\sigma). \quad (3.3)$$

Recordemos la definición de distribución de incremento inversa dada en (2.6). Una distribución inversa \hat{Q} es aquella que satisface $\hat{Q}(\sigma) = Q(\sigma^{-1})$ para toda $\sigma \in \mathcal{S}_n$. El Modelo 3 genera \hat{Q} con lo que la caminata aleatoria asociada es la inversa a la del Modelo 1 y 2. Podemos interpretar este mecanismo como un barajeo al revés, iniciando donde termina el algoritmo del barajeo rápido. Al asignar 0 o 1 a cada una de las cartas estamos eligiendo la subpila de la que provienen, indicando con el 0 la pertenencia de la carta a la primera subpila. Regresamos al orden inicial del barajeo rápido juntando la subpila 1 con la subpila 2 al reordenar las cartas colocando aquellas marcadas con 0 por encima de las marcadas con 1, respetando el orden relativo.

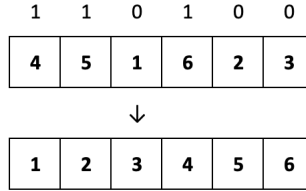


Figura 3.3. Inversa de un barajeo rápido sobre 6 cartas mediante el Modelo 3.

Es claro que existen 2^n maneras en las que repartir los bits a las n cartas, cada una de estas con probabilidad $1/2^n$. De manera similar a lo sucedido con la distribución Q el elemento id es de interés debido a que existen diferentes maneras de obtenerlo. Asignaciones de bits de las formas $01 \dots 1, 001 \dots 1, 0001 \dots 1$, etc., preservan el orden de las cartas. Así, la probabilidad del elemento id bajo \hat{Q} resulta

$$\hat{Q}(\text{id}) = \frac{n+1}{2^n}.$$

Por otro lado, dado un barajeo rápido σ el barajeo inverso σ^{-1} significa deshacer el orden creado por σ , que como hemos visto consiste en asignar cero a las cartas provenientes de la primera subpila. Entonces tenemos

$$\hat{Q}(\sigma^{-1}) = \frac{1}{2^n} = Q(\sigma), \quad \text{para } \sigma \text{ con dos sucesiones crecientes.}$$

Cabe mencionar que la distribución Q no es simétrica según la definición dada en (2.5) y por lo tanto la caminata con distribución de incremento Q no es reversible. Para ver esto retomemos el barajeo rápido σ sobre seis cartas ilustrado en la Figura 3.3

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 5 & 6 & 1 & 2 & 4 \end{pmatrix}.$$

El barajeo inverso σ^{-1} no es un barajeo rápido al tener tres sucesiones crecientes

$$\sigma^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 5 & 1 & 6 & 2 & 3 \end{pmatrix}.$$

Esto es $Q(\sigma) = 1/2^n$ y $Q(\sigma^{-1}) = 0$. De hecho, un resultado probado por Mike Atkinson [4] muestra que el número de barajeos rápidos que son al mismo tiempo inversa de algún otro es $\binom{n+1}{3} + 1$.

David Aldous [1] se dio cuenta que es más sencillo analizar la caminata aleatoria inversa. En este punto es crucial el resultado derivado a partir del Lema 2.1, donde establecimos que

$$d(t) = d'(t).$$

Es decir, la distancia al equilibrio para la caminata aleatoria es la misma que para la caminata aleatoria inversa. Por este motivo es suficiente analizar el Modelo 3.

Pensemos en el proceso de aplicar el algoritmo del Modelo 3 reiteradamente. Después de t tiempos cada una de las cartas posee una

cadena de bits de longitud t , consecuencia de elegir 0 o 1 en cada momento. Una cadena de bits puede ser expresada mediante $a_t a_{t-1} \cdots a_1$ con $a_i \in \{0, 1\}$. La lectura de las cadenas la realizamos de izquierda a derecha para iniciar con el último bit en ser añadido. Es

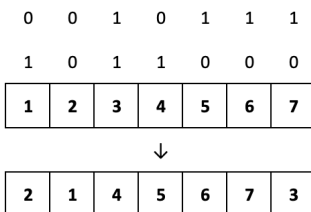


Figura 3.4. Ordenamiento de 6 cartas después de aplicar dos veces el barajeo rápido inverso.

claro que en el orden final las cartas cuyas cadenas inician con 0 están por delante de las que inician con 1, pero ¿qué ocurre entre aquellas con el mismo bit inicial? Para precisar el orden necesitamos retroceder un bit y aplicar la misma lógica, de esta manera las cartas que inician con 00 están por delante de aquellas que empiezan con 01.

De manera más general, observemos que tras t iteraciones existen 2^t cadenas de bits posibles. Estas cadenas las podemos ordenar siguiendo el *orden lexicográfico*. Dadas dos cadenas $a = a_t a_{t-1} \cdots a_1$ y $b = b_t b_{t-1} \cdots b_1$, decimos que $a < b$ si y solo si $a_i < b_i$ donde i es la primera posición tal que $a_i \neq b_i$. Así podemos determinar la permutación resultante tras t barajeos inversos examinando las cadenas de bits en cada carta. Si las cartas a y b tienen distintas cadenas, entonces en el orden final a y b siguen el orden lexicográfico ($a < b$ significa que la carta a esta por encima de la carta b). Por otro lado si dos cartas tienen la misma cadena de bits se sigue el orden original: si $a < b$ en el orden inicial entonces $a < b$ en el orden final.

3.2. Tiempo de mezcla

Seguir el orden de las cartas a través de los barajeos rápidos inversos mediante las cadenas de bits nos permite encontrar un tiempo fuertemente estacionario y con ello una cota superior al tiempo de mezcla.

Proposición 3.1. *Sea τ el número de barajeos rápidos inversos necesarios para que todas las cartas tengan cadenas de bits distintos. Entonces τ es un tiempo fuertemente estacionario.*

Demostración. Iniciemos observando que para una carta k su cadena de bits se distribuye uniforme entre las 2^t cadenas posibles. Supongamos que $\tau = t$, entonces todas las cadenas de longitud t de las n cartas son distintas. Por lo tanto la permutación final sigue el orden lexicográfico descrito anteriormente, notemos que de esta manera eliminamos la dependencia con la permutación original al no existir dos cadenas iguales.

Analicemos que sucede para n cadenas distintas determinadas. La manera en la que asignemos las cadenas a las cartas determina la permutación de las cartas, así una permutación específica tiene probabilidad

$$\frac{1}{2^t} \frac{1}{2^t - 1} \cdots \frac{1}{2^t - n + 1}.$$

Esto mismo ocurre para cada una de las $\binom{2^t}{n}$ n cadenas distintas. Por lo tanto

$$\mathbb{P}(X_\tau = \sigma | \tau = t) = \binom{2^t}{n} \frac{1}{2^t} \cdots \frac{1}{2^t - n + 1} = \frac{1}{n!}.$$

□

De manera similar a lo hecho en el modelo de transposiciones aleatorias necesitamos investigar la cola derecha de la distribución de τ para poder establecer una cota superior.

Proposición 3.2. *Para el proceso de barajeo rápido en una pila de n cartas tenemos*

$$t_{mezcla} \leq 2 \log_2 \left(\frac{4n}{3} \right).$$

Para n suficientemente grande.

Demostración. En el modelo de barajeo rápido inverso sea τ el tiempo fuertemente estacionario definido en la Proposición 3.1. Si $\tau \leq t$ entonces después de realizar t barajeos inversos todas las cartas poseen cadenas de bits diferentes. Observemos que una vez que las cadenas son distintas estas permanecerán de este modo en los tiempos siguientes. La probabilidad de este evento resulta

$$\mathbb{P}(\tau \leq t) = \prod_{k=0}^{n-1} \left(1 - \frac{k}{2^t} \right) = \left(\frac{2^t}{2^t} \right) \left(\frac{2^t - 1}{2^t} \right) \cdots \left(\frac{2^t - n + 1}{2^t} \right).$$

Sea $t = 2 \log_2(n/c)$, entonces $2^t = n^2/c^2$ y tenemos

$$\log \prod_{k=0}^{n-1} \left(1 - \frac{k}{2^t} \right) = \sum_{k=0}^{n-1} \log \left(1 - \frac{k}{2^t} \right) = \sum_{k=0}^{n-1} \log \left(1 - \frac{c^2 k}{n^2} \right). \quad (3.4)$$

De la serie de Taylor de $\log(x)$ alrededor de $a = 1$ establecemos

$$\log \left(1 - \frac{c^2 k}{n^2} \right) = -\frac{c^2 k}{n^2} - \frac{1}{2} O \left(\frac{k}{n^2} \right)^2. \quad (3.5)$$

Sustituyendo (3.5) en (3.4)

$$\begin{aligned}\log \prod_{k=0}^{n-1} \left(1 - \frac{k}{2^t}\right) &= - \sum_{k=0}^{n-1} \left(\frac{c^2 k}{n^2} + O\left(\frac{k}{n^2}\right)^2 \right) \\ &= - \frac{c^2(n-1)n}{2n^2} + O\left(\frac{n^3}{n^4}\right) \\ &= - \frac{c^2}{2} + O\left(\frac{1}{n}\right).\end{aligned}$$

Por lo tanto

$$\mathbb{P}(\tau \leq t) = e^{-c^2/2 + O(1/n)},$$

y de aquí que

$$\lim_{n \rightarrow \infty} \frac{\mathbb{P}(\tau \leq t)}{e^{-c^2/2}} = 1.$$

Nosotros estamos interesados en hacer la probabilidad de la cola derecha de τ pequeña

$$\mathbb{P}(t < \tau) \leq 1 - e^{-c^2/2} < \frac{1}{4}.$$

Resolviendo esta desigualdad para c resulta

$$c < \sqrt{2 \log(4/3)} \approx 0.75853.$$

Esto significa que para cualquier valor de c en este rango y en virtud de la Proposición 1.8 se satisface $d(t) \leq \mathbb{P}(t < \tau) < 1/4$. Tomemos $c = 3/4$, resolviendo para t tenemos

$$t = 2 \log_2 \left(\frac{4n}{3} \right),$$

y así

$$t_{mezcla} \leq 2 \log_2 \left(\frac{4n}{3} \right).$$

□

Observemos que este problema es una variante del problema del cumpleaños. El evento $\{\tau \leq t\}$ es equivalente al evento donde ninguna persona comparte cumpleaños. Por el contrario, el evento $\{\tau > t\}$ es análogo al evento donde al menos dos personas comparten fecha de cumpleaños. El giro de nuestro problema radica en que el número de personas (cartas) es fijo, a saber n . Estamos interesados en qué tan grande tiene que ser el “año” 2^t para que la probabilidad del evento $\{\tau \leq t\}$ (las n personas tienen distintos cumpleaños) sea muy probable, y con esto $\mathbb{P}(\tau < t)$ sea pequeño.

Para encontrar una cota inferior en el tiempo de mezcla haremos uso del concepto introducido al final del segundo capítulo aplicable cuando la distribución estacionaria es uniforme en el espacio de estados.

Proposición 3.3. *Consideremos el proceso de barajar una pila de n cartas mediante barajeo rápido. Para $0 < \varepsilon, 0 < \delta < 1$ fijos y n lo suficientemente grande tenemos*

$$t_{mezcla}(\varepsilon) \geq (1 - \delta) \log_2(n).$$

Demostración. Pensemos en el proceso de barajeo inverso. La asignación de bits a las cartas genera una transición en la cadena. Hemos visto que existen 2^n maneras de asignar los bits en cada iteración. Por lo tanto puede haber a lo más 2^n estados accesibles en un paso de la cadena. Esto es $\Delta \leq 2^n$, donde Δ es el máximo número de estados accesibles en un paso definido en (1.49). El espacio de estados tiene tamaño $|\mathcal{S}_n| = n!$. Por la aproximación de Stirling tenemos

$$\log(n!) = n \log(n) - n + O(\log(n)).$$

Dividiendo entre $\log(2)$ obtenemos la fórmula que nos será útil

$$\log_2(n!) = n \log_2(n) - n \log_2(e) + O(\log_2(n)).$$

Observemos que el límite del cociente entre $n \log_2(n)$ de los últimos dos términos cuando $n \rightarrow \infty$ es 0. Por lo tanto podemos resumirlos en la expresión $o(n \log_2(n))$ y tenemos

$$\log_2(n!) = n \log_2(n) (1 + o(1)).$$

Usando esta aproximación junto con $\log_2 \Delta \leq n$ en la cota encontrada en (1.51) resulta

$$\begin{aligned} t_{\text{mezcla}}(\varepsilon) &\geq \frac{\log((1-\varepsilon)|\mathcal{S}_n|)}{\log \Delta} \\ &= \frac{\log_2(1-\varepsilon) + \log_2(n!)}{\log_2 \Delta} \\ &\geq \frac{\log_2(1-\varepsilon)}{n} + \frac{\log_2(n!)}{n} \\ &= \frac{\log_2(1-\varepsilon)}{n} + \log_2(n) (1 + o(1)). \end{aligned}$$

El primer término rápidamente se hace insignificante para precisiones razonables. Por otro lado, una función que es $o(1)$ significa que tiende 0 cuando $n \rightarrow \infty$. Estos razonamientos prueban la desigualdad buscada.

□

3.3. Resultados de Bayer y Diaconis 1992

Concluiremos el capítulo con una breve discusión sobre los resultados principales del artículo de Bayer y Diaconis [5]. Por supuesto, la exposición no pretende ser detallada, sin embargo es imposible pasar por alto los resultados publicados en este artículo siendo estos definitivos sobre el tiempo de mezcla del barajeo rápido.

Pensemos en una generalización del barajeo rápido donde la pila se divide en a subpilas, es decir se hacen $a - 1$ cortes a la pila, acorde a

una distribución multinomial

$$\binom{n}{p_1, \dots, p_a} \left(\frac{1}{a}\right)^{p_1} \dots \left(\frac{1}{a}\right)^{p_a} = \binom{n}{p_1, \dots, p_a} \frac{1}{a^n},$$

donde p_i es no negativo representando el número de cartas en la subpila i tal que $p_1 + \dots + p_a = n$. Las cartas se colocan con probabilidad proporcional al tamaño de la subpila, al igual que en el barajeo rápido. De esta manera, dado las subpilas, todas las maneras de intercalar las cartas son equiprobables y el total de formas de hacerlo está dado por el coeficiente multinomial

$$\binom{n}{p_1, \dots, p_a},$$

correspondiente a las distintas formas de asignar los n espacios disponibles a las cartas de la primera subpila, a las cartas de la segunda subpila, etc. Siempre conservando el orden relativo dentro de cada subpila. Así la probabilidad de cortar la pila de una forma específica e intercalarla también de una manera determinada, es decir, la probabilidad de un barajeo específico es

$$\binom{n}{p_1, \dots, p_a}^{-1} \binom{n}{p_1, \dots, p_a} \frac{1}{a^n} = \frac{1}{a^n}, \quad (3.6)$$

resultado análogo a la distribución Q .

El proceso descrito recibe el nombre de a -barajeo y define una distribución de incremento Q_a en \mathcal{S}_n . En este contexto para la distribución de incremento Q del barajeo rápido tenemos $Q_2(\sigma) = Q(\sigma)$, nosotros nos referiremos a esta distribución simplemente como Q . Es fácil ver que tras un a -barajeo la permutación resultante σ cumple $1 \leq r(\sigma) \leq a$, donde $r(\sigma)$ denota el número de sucesiones crecientes en σ .

Conviene recordar que barajeos distintos no significa necesariamente obtener diferentes permutaciones, esta observación la hicimos también

en el barajeo rápido donde la permutación id puede ser obtenida por distintos barajeos. La validez de este hecho se repite en el a -barajeo y cobra nueva relevancia pues la mayoría de permutaciones posibles pueden obtenerse por diversos barajeos como veremos adelante.

Las analogías con la distribución Q no terminan aquí. Suponer dadas las a subpilas nos permite asociar distintas permutaciones a los $\binom{n}{p_1, \dots, p_a}$ barajeos posibles. Por esta razón y gracias a (3.6) la probabilidad de una permutación σ específica bajo Q_a es el cociente entre las distintas maneras de cortar la pila que permiten obtener a σ y todas las maneras de realizar un barajeo, a saber a^n .

Ahondaremos más en este razonamiento y para ello seguiremos la explicación propuesta por Mann [17]. Consideremos una permutación σ determinada y con ello el número de sucesiones crecientes $r(\sigma) = r$. Con esta información tenemos la certeza de dónde ocurrieron $r - 1$ cortes en la pila, esto es entre la última carta de una sucesión creciente y la primera de la siguiente sucesión. Restan $a - 1 - (r - 1) = a - r$ cortes sin especificar pudiendo haber ocurrido literalmente en cualquier lugar. Pensemos en estas subpilas “escondidas” a plena vista debido a que siguieron correctamente el orden relativo respecto a la subpila contigua. Calcular las distintas maneras en las que es posible distribuir los $a - r$ cortes restantes equivale a conocer $Q_a(\sigma)$. Esto se traduce en el clásico problema de combinatoria donde tenemos $n + a - r$ espacios en blanco de los cuales n queremos que sean cartas. Notemos que consideramos la posibilidad de que las subpilas puedan estar vacías. De esta manera

$$Q_a(\sigma) = \frac{\binom{n+a-r}{n}}{a^n}.$$

Es claro que el resultado se mantiene válido si consideramos una permutación distinta con la misma cantidad de sucesiones crecientes. En otras palabras, la probabilidad de una permutación σ particular bajo

Q_a no esta determinada por la totalidad de información sobre σ , solo es necesario conocer el número de sucesiones crecientes $r(\sigma)$. Por lo tanto bajo Q_a todas las permutaciones con el mismo número de sucesiones tienen la misma probabilidad

$$Q_a(\sigma) = \frac{\binom{n+a-r(\sigma)}{n}}{a^n} \quad \text{para } \sigma \in \mathcal{S}_n, \quad (3.7)$$

donde $1 \leq r(\sigma) \leq a$. Observemos que al sustituir $a = 2$ y considerar 1 o 2 sucesiones crecientes obtenemos la distribución Q (3.1).

A pesar de lo asombroso que es el resultado que se ha desarrollado, este no parece tener una implicación directa en el problema del barajeo rápido y solo parece ser relevante para seres con a manos capaces de mezclar a subpilas al mismo tiempo. Bayer y Diaconis llenan este espacio mediante un teorema que afirma lo siguiente: Realizar un a -barajeo seguido de un b -barajeo es equivalente a realizar un solo ab -barajeo, en términos de las distribuciones de incremento resultantes para ambos procesos. Matemáticamente esto es

$$Q_a^* Q_b = Q_{ab}.$$

La demostración es esencialmente geométrica y puede consultarse en la referencia [5] (Lema 1), nosotros la omitiremos por no ser fundamental para los presentes objetivos.

Regresemos al barajeo rápido. En la ecuación (3.3) establecimos que Q^{*t} es la distribución de probabilidad de la cadena después de t pasos iniciando en la permutación identidad. En el lenguaje del a -barajeo esto se traduce en

$$Q^{*t} = Q_{2^t}$$

y por lo tanto

$$\mu_t(\sigma) = Q_{2^t}(\sigma) = \frac{\binom{n+2^t-r(\sigma)}{n}}{2^{tn}}, \quad \text{para } \sigma \in \mathcal{S}_n. \quad (3.8)$$

A partir de esta expresión es posible calcular la distancia de variación total

$$\|P^t(\text{id}, \cdot) - \pi\|_{VT} = \frac{1}{2} \sum_{\sigma \in \mathcal{S}_n} \left| Q_{2^t}(\sigma) - \frac{1}{n!} \right| = \frac{1}{2} \sum_{\sigma \in \mathcal{S}_n} \left| \frac{\binom{n+2^t-r(\sigma)}{n}}{2^{tn}} - \frac{1}{n!} \right|.$$

Recordemos que en (2.8) vimos que

$$d(t) = \|P^t(\text{id}, \cdot) - \pi\|_{VT}.$$

Así mismo, agrupando a las permutaciones según el número de sucesiones crecientes que poseen, somos capaces de simplificar la expresión para la distancia al equilibrio de la cadena

$$d(t) = \frac{1}{2} \sum_{j=1}^n R_{nj} \left| \frac{\binom{n+2^t-j}{n}}{2^{tn}} - \frac{1}{n!} \right|. \quad (3.9)$$

La expresión R_{nj} denota el número de permutaciones con j sucesiones crecientes que existen en \mathcal{S}_n . Afortunadamente estos números son bien conocidos en el análisis combinatorio y corresponden a los números eulerianos. Existen diversas fórmulas para calcular estos números, estas se pueden consultar en [25].

Finalmente estos resultados nos permiten calcular la distancia de variación total para el barajeo rápido de una pila de 52 cartas de manera exacta que aquí presentamos con 3 decimales de precisión.

t	1	2	3	4	5	6	7	8	9
$d(t)$	1.000	1.000	1.000	1.000	.924	.614	.334	.167	.085

El último de los teoremas del artículo de Bayer y Diaconis, y el más importante, es un análisis asintótico para $d(t)$ a medida que el número de cartas va incrementando. Para $t = \frac{3}{2} \log_2(n) + \theta$ tenemos

$$d(t) = 1 - 2\Phi\left(\frac{-2^{-\theta}}{4\sqrt{3}}\right) + O\left(\frac{1}{n^{1/4}}\right), \quad (3.10)$$

donde $\Phi(x)$ es la función de distribución acumulada de una variable aleatoria normal estandar

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy.$$

Esta expresión nos permite estudiar el comportamiento de la distancia con la distribución estacionaria cerca de $\frac{3}{2} \log_2(n)$ barajeos al considerar n grande.

Explicaremos los razonamientos modulares en la demostración de la expresión (3.10), omitiendo en la medida de lo posible, manipulaciones algebraicas que pueden consultarse en el artículo original.

Retomemos la distribución $Q_a(\sigma)$. Ahora sabemos que la probabilidad de una permutación solo depende del número de sucesiones crecientes, por lo que es conveniente escribir simplemente $Q_a(r)$.

Un hecho importante sobre esta distribución es que es estrictamente decreciente en r . Esto significa que si $1 \leq r_1 < r_2 \leq n$ entonces resulta $Q_a(r_1) > Q_a(r_2)$. Para convencernos de este hecho solo debemos comparar los coeficientes binomiales $\binom{a+n-r_1}{n}$ y $\binom{a+n-r_2}{n}$, que corresponden a las expresiones para $Q_a(r_1)$ y $Q_a(r_2)$. La comparación se puede simplificar más al considerar únicamente

$$\begin{aligned} \frac{(a+n-r_1)!}{(a-r_1)!} &= (a+n-r_1) \dots (a-r_1+1), \\ \frac{(a+n-r_2)!}{(a-r_2)!} &= (a+n-r_2) \dots (a-r_2+1). \end{aligned}$$

Ambas expresiones tienen n términos y es evidente que la de arriba, correspondiente a $Q_a(r_1)$, es mayor.

Por otro lado, investiguemos los valores de la distribución en los extremos. Es decir, cuando $r = 1$ y $r = n$. Tomemos el primer caso

$$Q_a(1) = \frac{\binom{a+n-1}{n}}{a^n} = \frac{(a+n-1) \dots (a)}{a^n n!} = \frac{a+n-1}{a} \dots \frac{a}{a} \frac{1}{n!},$$

y por lo tanto $Q_a(1) > 1/n!$. De manera similar

$$Q_a(n) = \frac{\binom{a}{n}}{a^n} = \frac{a(a-1)\dots(a-n+1)}{a^n n!} = \frac{a}{a} \dots \frac{a-n+1}{a} \frac{1}{n!},$$

resultando $Q_a(n) < 1/n!$. Esto significa que existe un máximo r en el que se cumple $Q_a(r) \geq 1/n!$.

El análisis asintótico usa la caracterización de la distancia de variación total vista en el Corolario 1. Recordemos que dadas dos distribuciones μ y ν tenemos

$$\|\mu - \nu\|_{VT} = \sum_{x \in B} (\mu(x) - \nu(x)),$$

donde B es el subconjunto donde $\mu(x) \geq \nu(x)$.

Con estas nociones podemos pasar a esbozar la demostración del resultado (3.10). Por conveniencia volvamos a expresar el parámetro del tiempo como $t = \log_2(n^{3/2}c)$, donde $c = 2^\theta$ y de esta manera $0 < c < \infty$. Así mismo, sea el número de sucesiones crecientes $r = n/2 + h$ con $-n/2 + 1 \leq h \leq n/2$. Entonces una versión equivalente a (3.9) puede ser escrita como

$$d(t) = \sum_{h \in B} R_{nh} \left(Q^{*t} \left(\frac{n}{2} + h \right) - \frac{1}{n!} \right). \quad (3.11)$$

Donde R_{nh} es el número de permutaciones en \mathcal{S}_n con $n/2 + h$ sucesiones crecientes, y B es el conjunto donde Q^{*t} domina a la distribución uniforme

$$B = \left\{ h : Q^{*t} \left(\frac{n}{2} + h \right) \geq \frac{1}{n!} \right\}.$$

Para el análisis debemos investigar el comportamiento asintótico de Q^{*t} , R_{nh} , y del conjunto B . Empecemos con el primero. Usando la fórmula

(3.8) tenemos

$$\begin{aligned}
n!Q^{*t}(r) &= n! \frac{1}{2^{tn}} \frac{(n+2^t-r)!}{(2^t-r)!n!} \\
&= \frac{2^t+n-r}{2^t} \cdots \frac{2^t-r+1}{2^t} \\
&= \exp \left(\sum_{i=0}^{n-1} \log \left(1 + \frac{n-r-i}{2^t} \right) \right) \\
&= \exp \left(\sum_{i=0}^{n-1} \log \left(1 + \frac{n/2-h-i}{n^{3/2}c} \right) \right)
\end{aligned}$$

Expandiendo $\log(1+x)$ hasta el 4to término y aplicando las correspondientes fórmulas de suma resulta

$$\begin{aligned}
n!Q^{*t} \left(\frac{n}{2} + h \right) &= \exp \left(-\frac{h}{c\sqrt{n}} - \frac{1}{24c^2} - \frac{1}{2} \left(\frac{h}{cn} \right)^2 \right. \\
&\quad \left. + O \left(\frac{1}{n} \right) + O \left(\frac{h}{n} \right) \right). \tag{3.12}
\end{aligned}$$

Pasemos a investigar al conjunto B . Debido a que Q^{*t} es estrictamente decreciente en el número de sucesiones crecientes, definimos el entero h^* como el punto donde la distribución cruza por debajo de la distribución uniforme $1/n!$

$$Q^{*t} \left(\frac{n}{2} + h \right) \geq \frac{1}{n!} \iff h \leq h^*.$$

El lado izquierdo ocurre si, y solo si, $n!Q^{*t}(n/2+h) \geq 1$. Por lo tanto al igualar a 0 el exponente en (3.12) resulta para n suficientemente grande

$$h^* = -\frac{\sqrt{n}}{24c} + \frac{1}{12c^3} + B + O \left(\frac{1}{\sqrt{n}} \right),$$

donde $|B| \leq 1$. De esta manera (3.11) equivale a

$$d(t) = \sum_{-n/2 < h \leq h^*} \frac{R_{nh}}{n!} \left(n!Q^{*t} \left(\frac{n}{2} + h \right) - 1 \right).$$

Por último, Tanny [25] mostró que si X_1, \dots, X_n son variables aleatorias i.i.d uniformes $(0, 1)$ entonces

$$\frac{R_{nj}}{n!} = \mathbb{P}(j \leq X_1 + \dots + X_n \leq j + 1).$$

Esto justifica el hecho de usar el teorema del límite central en el análisis asintótico. Los tres resultados que hemos obtenido se combinan en una exhibición magistral de manipulación algebraica que resulta en la expresión (3.10). Los detalles se pueden consultar en [5] (pág. 308).

Conclusiones

En ambos modelos estudiados el análisis ha sido principalmente asintótico y en función del parámetro n (número de cartas). Haber realizado dicho análisis limitándonos a una cadena de Markov con n fijo, y centrar nuestros esfuerzos en tratar de determinar la función asociada $d(t)$ hubiera supuesto un esfuerzo fútil por dos motivos concretos. El primero se debe a que no existe garantía que de ser capaces de conocer a $d(t)$ de manera detallada. La mayoría de las veces esto no es posible y resultados como los expuestos por Bayer y Diaconis en el barajeo rápido son excepcionales, en este caso suscitado por propiedades algebraicas específicas. Subrayamos la importancia de desarrollar métodos potentes como los tiempos fuertemente estacionarios que no son particulares a un proceso. Por el contrario Aldous y Diaconis [3] mostraron que un tiempo óptimo siempre existe para cualquier cadena ergódica, la dificultad reside entonces en encontrarlo y estudiarlo. Por otro lado, no es suficiente considerar una cadena de Markov con un tamaño de espacio de estados fijo como en la teoría clásica. Actualmente el interés radica en cadenas con grandes espacios de estados, pensemos rápidamente en los métodos MCMC.

Retomando las estimaciones sobre el tiempo de mezcla del barajeo

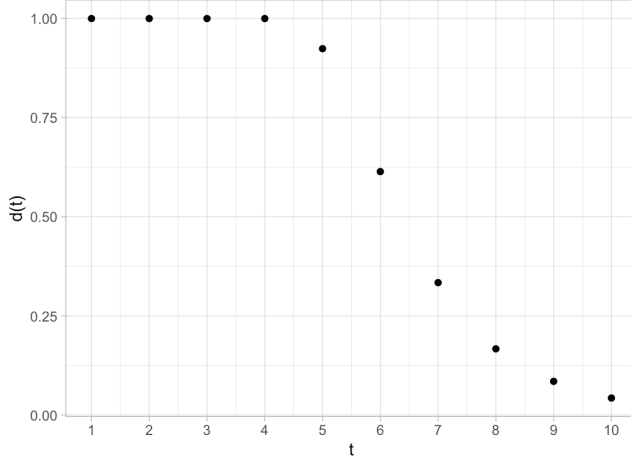


Figura 3.5. Gráfica de las distancias contra el tiempo.

rápido de 52 cartas con una precisión de $\varepsilon = 1/4$ tenemos

$$6 \leq t_{\text{mezcla}}(\varepsilon) \leq 12.$$

Notemos que el proceso mezcla rápido, especialmente comparando la estimación con el tamaño del espacio de estados $|\mathcal{S}_{52}|$ (un número de 60 dígitos). Gracias a la fórmula cerrada podemos corroborar este resultado y de manera exacta $t_{\text{mezcla}} = 8$. Además es posible proveer un análisis más detallado del comportamiento de las distancias cerca de este punto.

Finalmente podemos dar una respuesta contundente a la pregunta que motivó esta investigación. Se requieren no menos de 7 barajeos para eliminar, razonablemente bien, cualquier tipo de orden subyacente en la pila de cartas. Este hecho se conoce popularmente como “7 barajeos son suficientes en la práctica”.

Observando la Figura 3.5 advertimos un comportamiento sorprendente. Durante los primeros 5 barajeos la distancia con la distribución uniforme es prácticamente máxima. Pasado este punto las

distancias registran una abrupta caída a través de un decrecimiento geométrico con factor $1/2$ que continua por siempre. Esto significa que la convergencia toma lugar en una pequeña ventana de tiempo.

El barajeo rápido y el barajeo por transposición fueron los primeros procesos en los que se descubrió este interesante fenómeno y motivó la siguiente importante noción. Consideremos una familia de cadenas de Markov $(X_t^{(n)})$. Es decir, definimos una sucesión de matrices de transición (P_n) cuyas dimensiones crecen según un parámetro n , comúnmente el tamaño del espacio de estados (el número de cartas en el barajeo de cartas). Así mismo escribimos $t_{\text{mezcla}}^{(n)}(\varepsilon)$ para el correspondiente tiempo de mezcla y $d_n(t)$ para la distancia, en este contexto es útil pensar a esta función definida sobre los reales positivos. Se dice que existe (asintóticamente) un *cutoff phenomenon* si para todo $\varepsilon > 0$ tenemos que

$$d_n((1 - \varepsilon)t_n) \rightarrow 1,$$

pero

$$d_n((1 + \varepsilon)t_n) \rightarrow 0$$

cuando $n \rightarrow \infty$. En términos sencillos, asintóticamente existe un punto en el tiempo tal que al superarlo la distancia decae exponencialmente a 0, no obstante antes de este punto la distancia se mantiene esencialmente en su valor máximo y la función se aproxima a una función escalonada alrededor de t_n . Notemos que la siguiente condición es equivalente

$$t_{\text{mezcla}}^{(n)}(1 - \varepsilon) \sim t_{\text{mezcla}}^{(n)}(\varepsilon), \quad \text{para } 0 < \varepsilon < 1,$$

donde $a_n \sim b_n$ significa $a_n/b_n \rightarrow 1$ cuando $n \rightarrow \infty$, estandarizando tenemos $t_n = t_{\text{mezcla}}^{(n)}$. Esto significa que el tiempo requerido para mezclar ligeramente (digamos $\varepsilon = 0.99$) es asintóticamente equivalente al necesario para mezclar completamente (digamos $\varepsilon = 0.01$).

En el barajeo rápido de cartas la existencia de este fenómeno se demuestra por la identidad establecida en (3.10). En este resultado se observa que $t_n = 3/2 \log_2(n)$ y la distancia tiende a 0 exponencialmente rápido cuando $c \rightarrow \infty$ mientras que tiende doblemente exponencial a 1 cuando $c \rightarrow -\infty$. Para el barajeo por transposición Diaconis [11] encontró que $t_n = 1/2n \log_2(n)$, esto puede ser probado auxiliándonos en la cota superior proporcionada por el tiempo fuertemente estacionario (2.26).

El *cutoff phenomenon* supone actualmente una de las áreas más importantes de investigación en la probabilidad moderna. Se sabe que ocurre en varias cadenas (por ejemplo en la urna de Ehrenfest), también se sabe que en otras no (por ejemplo en la caminata aleatoria en los enteros mod n). No obstante, el entendimiento de este fenómeno sigue siendo limitado y las demostraciones son complicadas y específicas para cada cadena. La búsqueda de una explicación general sigue siendo el problema abierto más relevante en esta área.

Bibliografía

- [1] David Aldous. Random walks on finite groups and rapidly mixing Markov chains. In *Séminaire de Probabilités XVII 1981/82: Proceedings*, pages 243–297. Springer, 1983.
- [2] David Aldous and Persi Diaconis. Shuffling cards and stopping times. *The American Mathematical Monthly*, 93(5):333–348, 1986.
- [3] David Aldous and Persi Diaconis. Strong uniform times and finite random walks. *Advances in Applied Mathematics*, 8(1):69–97, 1987.
- [4] Mike D Atkinson. Restricted permutations. *Discrete Mathematics*, 195(1-3):27–38, 1999.
- [5] Dave Bayer and Persi Diaconis. Trailing the dovetail shuffle to its lair. *The Annals of Applied Probability*, pages 294–313, 1992.
- [6] Nathanaël Berestycki. Mixing times of Markov chains: Techniques and examples. *Alea-Latin American Journal of Probability and Mathematical Statistics*, 2016.
- [7] Pierre Brémaud. *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*, volume 31. Springer Science & Business Media, 2001.

- [8] Persi Diaconis. Group representations in probability and statistics. *Lecture notes-monograph series*, 11:i–192, 1988.
- [9] Persi Diaconis. The cutoff phenomenon in finite Markov chains. *Proceedings of the National Academy of Sciences*, 93(4):1659–1664, 1996.
- [10] Persi Diaconis. Mathematical developments from the analysis of riffle shuffling. *Groups, combinatorics & geometry (Durham, 2001)*, pages 73–97, 2003.
- [11] Persi Diaconis and Mehrdad Shahshahani. Generating a random permutation with random transpositions. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57(2):159–179, 1981.
- [12] Rick Durrett. Shuffling chromosomes. *Journal of Theoretical Probability*, 16:725–750, 2003.
- [13] Jason Fulman. Semisimple orbits of lie algebras and card-shuffling measures on coxeter groups. *Journal of Algebra*, 224(1):151–165, 2000.
- [14] Israel N Herstein. *Topics in algebra*. John Wiley & Sons, 1991.
- [15] Victor Kozyakin. On accuracy of approximation of the spectral radius by the Gelfand formula. *Linear Algebra and its Applications*, 431(11):2134–2141, 2009.
- [16] David A Levin and Yuval Peres. *Markov chains and mixing times*, volume 107. American Mathematical Soc., 2017.
- [17] Brad Mann. How many times should you shuffle a deck of cards. *UMAP J*, 15(4):303–332, 1994.

- [18] José María Ranz, Carmen Segarra, and Alfredo Ruiz. Chromosomal homology and molecular organization of Muller’s elements d and e in the *Drosophila repleta* species group. *Genetics*, 145(2):281–295, 1997.
- [19] Peter Matthews. A strong uniform time for random transpositions. *Journal of Theoretical Probability*, 1:411–423, 1988.
- [20] Igor Pak. *Random walks on groups: strong uniform time approach*. Harvard University, 1997.
- [21] Mark Pinsky and Samuel Karlin. *An introduction to stochastic modeling*. Academic press, 2010.
- [22] Sheldon M Ross. *Stochastic processes*. John Wiley & Sons, 1995.
- [23] Justin Salez. Mixing times of Markov chains, <https://www.ceremade.dauphine.fr/~salez/mix.pdf>, 2022.
- [24] Richard P Stanley. Generalized riffle shuffles and quasisymmetric functions. *Annals of Combinatorics*, 5(3):479–491, 2001.
- [25] Stephen Tanny. A probabilistic interpretation of eulerian numbers. 1973.