



CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS  
DEL INSTITUTO POLITÉCNICO NACIONAL

UNIDAD ZACATENCO  
DEPARTAMENTO DE MATEMÁTICAS

## **Algoritmo de Metropolis-Hastings y Recurrencia de Polya**

TESIS QUE PRESENTA

**Jesús Enrique Bejarano Ochoa**

Para obtener el grado de

Maestro en Ciencias

en la especialidad de Matemáticas

Director de Tesis: Dr. José Guadalupe Martínez Bernal

Ciudad de México

Febrero 2022



# Contenido

<b>Resumen</b>	<b>VIII</b>
<b>Introducción</b>	<b>XI</b>
<b>1. Potencias de matrices no-negativas</b>	<b>15</b>
1.1. Digráfica asociada a una matriz cuadrada . . . . .	15
1.2. Matrices reducibles e irreducibles . . . . .	16
1.3. Irreducibilidad, aperiodicidad y regularidad . . . . .	24
<b>2. Cadenas de Markov</b>	<b>31</b>
2.1. Clasificación de estados . . . . .	31
2.2. Distribuciones estacionarias . . . . .	35
2.3. Teorema de Perron-Frobenius . . . . .	41
2.4. Teorema ergódico de Markov . . . . .	43
<b>3. Algoritmo de Metropolis-Hastings</b>	<b>47</b>
3.1. Un enfoque probabilista . . . . .	47
3.2. Caminatas aleatorias . . . . .	51
3.3. Cadenas de Metropolis-Hastings . . . . .	53
3.4. Ejemplos . . . . .	57
<b>4. Teorema de Polya vía redes eléctricas</b>	<b>61</b>
4.1. Funciones armónicas . . . . .	63
4.2. Voltajes y corrientes . . . . .	65
4.3. Problema de Dirichlet discreto . . . . .	66

4.4. Principio de Rayleigh . . . . .	70
4.5. Resistencia efectiva y recurrencia . . . . .	75
4.6. Demostración del Teorema de Polya . . . . .	79
<b>5. Teorema de Polya vía funciones especiales</b>	<b>83</b>
5.1. Funciones generadoras exponenciales . . . . .	84
5.2. Funciones de Bessel modificadas . . . . .	86
5.3. Principio de Laplace . . . . .	87
5.4. Demostración del Teorema de Polya . . . . .	89
<b>Bibliografía</b>	<b>97</b>

# Agradecimientos

Agradezco al Consejo Nacional de Ciencia y Tecnología (Conacyt) por haber apoyado este trabajo de tesis con una beca de maestría.

Al Dr. José Martínez Bernal por asesorarme y apoyarme en la elaboración de ésta.

Especial agradecimiento a mi madre Leticia y a mi hermano Alan por su constante apoyo y comprensión durante este posgrado.

Dedicada a mi madre,  
con amor.

# Resumen

En este trabajo estudiamos dos resultados: el algoritmo de Metropolis-Hastings y el teorema de recurrencia de Polya. Empezamos por demostrar que una matriz cuadrada no negativa es regular si y sólo si es irreducible y aperiódica. Primero interpretamos este resultado tanto en términos matriciales como combinatorios, vía la digráfica asociada a la matriz, para después pasar directamente a una interpretación en términos de cadenas de Markov finitas sobre gráficas. En este contexto, la conclusión principal es que bajo ciertas condiciones, dada una distribución de probabilidad  $\pi \in \mathbb{R}^n$ , el algoritmo de Metropolis-Hastings nos permite encontrar una cadena de Markov cuya distribución estacionaria sea precisamente  $\pi$ .

Los conceptos de caminata aleatoria, recurrencia y transitoriedad introducidos en el caso de cadenas de Markov finitas se extienden al caso en el que ahora la gráfica es  $\mathbb{Z}^d$ , con cada arista teniendo peso  $1/2d$ . En este nuevo escenario, estudiamos el teorema de recurrencia de Polya, el cual afirma que toda caminata aleatoria simple en  $\mathbb{Z}^d$  es recurrente cuando  $d = 1, 2$  y transitoria para  $d \geq 3$ . Probamos dicho teorema de dos maneras totalmente distintas: vía la teoría de redes eléctricas en gráficas y vía la teoría de funciones especiales.

Para la primera prueba, desarrollamos la teoría de redes eléctricas en gráficas, la cual consiste en estudiar en abstracto la teoría de circuitos eléctricos siguiendo las leyes de Kirchhoff y la ley de Ohm. Cabe mencionar que definir una estructura de red eléctrica en una gráfica es equivalente a resolver un problema de Dirichlet discreto cuando la frontera consiste de únicamente dos vértices: en esta parte requerimos la noción de función armónica en una gráfica. El concepto de red eléctrica en una gráfica se conecta con la noción de caminata recurrente cuando expresamos la probabilidad

de que una caminata regrese a su vértice de partida en términos de la resistencia efectiva de la red.

Para la segunda prueba, utilizamos herramientas como funciones generadoras ordinarias y exponenciales, funciones de Bessel modificadas del primer tipo, la transformada de Borel y el principio de Laplace. La idea principal es llevar el problema al estudio del análisis asintótico de una integral. Esta integral viene de relacionar la función de Bessel modificada del primer tipo con la función generadora exponencial que resulta de contar bucles en la gráfica  $\mathbb{Z}$ . Finalmente, la prueba del teorema de Polya se reduce al estudio de la convergencia de una integral que depende únicamente de la dimensión de  $\mathbb{Z}^d$ .



# Summary

In this thesis we study two important results: the Metropolis-Hastings algorithm and Polya's recurrence theorem. We first prove that a non-negative square matrix is regular if and only if it is irreducible and aperiodic. This result is interpreted from a matrix and combinatoric theoretic perspective, via the digraph associated with a matrix, in order to then interpret it directly in terms of the theory of finite Markov chains over graphs. In this context, the main conclusion states that, under certain conditions, given a probability distribution  $\pi \in \mathbb{R}^n$ , we can construct a Markov chain with  $\pi$  as its stationary distribution through the Metropolis-Hastings algorithm.

The concepts of random walk, recurrence and transience introduced in the case of finite Markov chains are extended to the case where the graph is  $\mathbb{Z}^d$ , and each of its edges has a weight of  $1/2d$ . In this new scenario, we study Polya's recurrence theorem, which states that every simple random walk on  $\mathbb{Z}^d$  is recurrent when  $d = 1, 2$ , and transient for  $d \geq 3$ . We prove this theorem in two totally different ways: via the theory of electric circuits on graphs and via the theory of special functions.

In the first proof, we develop the theory of electric circuits on graphs, which studies the theory of electric circuits through the Kirchhoff and Ohm laws from an abstract viewpoint. It's important to mention that defining a structure of electric circuit on a graph is equivalent to solving a discrete Dirichlet problem when the boundary consists of only two vertices; The notion of harmonic function is needed for this part. The concept of electric circuit on a graph is connected to the idea of recurrent random walk when we express the probability that a random walk returns to its initial position in terms of the effective resistance of the circuit.

For the second approach, we use different tools like ordinary and exponential

generating functions, modified Bessel functions of the first kind, Borel's transform and Laplace's principle. The main idea is to transform the problem into the study of the asymptotic behaviour of a specific integral. This integral is a result of the relationship between the modified Bessel function of the first type and the exponential generating function associated with counting loops on the graph  $\mathbb{Z}$ . Finally, the proof of Polya's theorem is reduced to the convergence or divergence of an integral that depends only on the dimension of  $\mathbb{Z}^d$ .

# Introducción

En este trabajo estudiamos dos resultados importantes: el algoritmo de Metropolis-Hastings y el teorema de recurrencia de Polya.

Comenzamos con una introducción a la teoría de matrices (cuadradas) con entradas no negativas. En particular, estamos interesados en las propiedades de regularidad, irreducibilidad y aperiodicidad, y las relaciones que existen entre ellas. Básicamente, una matriz no negativa  $A$  es regular si existe un entero positivo  $N$  tal que la matriz  $A^N$  posee entradas estrictamente positivas. Por otro lado, decimos que  $A$  es irreducible si cualquiera dos vértices de su digráfica se comunican, pensando que esto se cumple cuando existe un camino dirigido de un vértice al otro. Finalmente,  $A$  es aperiódica si existe un entero positivo  $N$  tal que para todo  $k \geq N$ , y cada vértice  $v$  de su digráfica, podemos encontrar un ciclo que inicia y termina en  $v$  de longitud  $k$ . El teorema principal del primer capítulo concluye que una matriz no negativa es regular si y sólo si es irreducible y aperiódica. Este resultado nos permite pasar directamente a la teoría de cadenas de Markov, pues el comportamiento asintótico de tales cadenas depende de su matriz de transición y la propiedad de regularidad.

Una cadena de Markov se compone de una matriz estocástica  $P$ , llamada la matriz de transición, y una distribución (de probabilidad) inicial  $x_0$ . Con el fin de analizar el comportamiento asintótico de la sucesión de distribuciones  $x_0, Px_0, P^2x_0, \dots$ , en este trabajo demostramos que cualquier matriz estocástica y regular define una aplicación contractiva en el conjunto de todas las distribuciones de probabilidad. Entonces, una matriz no negativa, que además es estocástica y regular, posee una única distribución fija. De esto se sigue que cuando  $P$  es regular, no importa que distribución inicial tomemos, la sucesión  $(P^n x_0)$  siempre convergerá a una misma distribución

fija, llamada la distribución estacionaria de la cadena.

Ahora bien, surge de manera natural la siguiente pregunta: dada una distribución de probabilidad  $\pi$  ¿es posible construir una cadena de Markov que tenga a  $\pi$  como distribución estacionaria? La respuesta es afirmativa y aún más satisfactoria, podemos calcular tal cadena de Markov mediante el algoritmo de Metropolis-Hastings, el cual nos permite encontrar una caminata aleatoria cuya distribución estacionaria sea  $\pi$ .

El algoritmo de Metropolis fue desarrollado en el artículo de física estadística publicado por los autores Nicholas Metropolis, Arianna Rosenbluth, Marshall Rosenbluth, Augusta Teller y Edward Teller en los años 1950's [5]. Este algoritmo requiere una cadena de Markov finita, llamada cadena propuesta, con matriz de transición simétrica, sobre la cual se realizan una serie de cálculos para obtener la cadena de Markov con distribución estacionaria previamente dada. Posteriormente, en 1970, Wilfred Keith Hastings generalizó este algoritmo para admitir cadenas propuestas con matriz de transición no necesariamente simétrica. Esta generalización es el famoso algoritmo de Metropolis-Hastings.

A pesar de que el algoritmo de Metropolis-Hastings se utiliza de manera exhaustiva en una gran variedad de disciplinas, como la física estadística, el procesamiento de imágenes y la generación de números aleatorios, por mencionar algunas, las estimaciones para la convergencia de la cadena de Metropolis-Hastings a su distribución estacionaria dependen tanto de la cadena propuesta como de la distribución de probabilidad previamente escogida, y aún queda mucho por investigar en esa dirección [6]. Profundizar en la convergencia de la cadena de Metropolis-Hastings conlleva introducir enfoques demasiado técnicos, es por esta razón que hemos omitido este aspecto en el presente trabajo.

Los conceptos de caminata aleatoria, recurrencia y transitoriedad introducidos en el caso de cadenas de Markov finitas se pueden extender al caso en el que el espacio de estados es numerable, como por ejemplo, la malla  $\mathbb{Z}^d$ , donde cada arista tiene peso igual a  $1/2d$ . En este contexto, una caminata que comienza en un vértice dado es recurrente si la probabilidad de que regrese a éste es igual a 1, y es transitoria en otro caso. De esta manera, estudiamos el teorema de recurrencia de Polya, el cual afirma

que toda caminata aleatoria simple en  $\mathbb{Z}^d$  es recurrente cuando  $d = 1, 2$  y transitoria para  $d \geq 3$ .

El teorema de recurrencia de Polya se puede demostrar de diversas maneras, sin embargo, exponemos sólo dos enfoques: el primero vía la teoría de redes eléctricas en gráficas y el segundo vía la función de Bessel modificada de primer tipo y análisis asintótico. La teoría de redes eléctricas en gráficas consiste en estudiar matemáticamente el alcance de la teoría de circuitos eléctricos vía las leyes de Kirchhoff y la ley de Ohm, interpretando las conductancias como una función de pesos en la gráfica, y pensando en el voltaje y corriente eléctrica como funciones con valores reales definidas en los vértices y las aristas, respectivamente, que además obedezcan las leyes de Kirchhoff y la ley de Ohm. Estos conceptos se definen para gráficas finitas, y para pasar al caso numerable, truncamos la gráfica original para obtener subgráficas finitas en las que podemos definir una estructura de red eléctrica. Es importante mencionar que definir una estructura de red eléctrica en una gráfica es equivalente a resolver un problema de Dirichlet discreto cuando la frontera consiste únicamente de dos vértices. En el capítulo 4 desarrollamos este tema introduciendo el concepto de función armónica en una gráfica con pesos positivos y exponiendo el problema de Dirichlet discreto en el caso general cuando la frontera es cualquier subconjunto de vértices. Por último, el concepto de red eléctrica en una gráfica se conecta con la noción de caminata recurrente cuando expresamos la probabilidad de que una caminata regrese a su vértice de partida en términos de la resistencia efectiva de la red.

Para el segundo tratamiento, capítulo 5, utilizamos herramientas como funciones generadoras exponenciales, un uso particular de la función de Bessel modificada del primer tipo, la transformada de Borel para relacionar funciones generadoras ordinarias con funciones generadoras exponenciales de una misma sucesión, y el principio de Laplace. La idea principal es expresar nuestro problema como el análisis asintótico de una integral en infinito. Esta integral viene de relacionar la función de Bessel modificada del primer tipo con la función generadora exponencial que resulta de contar bucles en la malla  $\mathbb{Z}$ . Esto es suficiente, pues el análisis en la malla  $\mathbb{Z}^d$  se puede expresar en términos de la malla de dimensión 1. Finalmente, la prueba del teorema

de Polya se reduce al estudio de la convergencia de la integral

$$\int_1^{\infty} t^{-d/2} dt,$$

más específicamente, si  $d$  es tal que esta integral diverge, entonces toda caminata en  $\mathbb{Z}^d$  será recurrente, mientras que será transitoria en otro caso.

## CAPÍTULO 1

# Potencias de matrices no-negativas

La teoría de matrices con entradas no negativas es una área de las matemáticas con múltiples aplicaciones a diferentes ramas de las matemáticas aplicadas. Estamos particularmente interesados en estudiar las propiedades de irreducibilidad, aperiodicidad y regularidad de dichas matrices, ayudándonos de la teoría de digráficas para esclarecer conceptos, resultados e incluso procedimientos cuando sea posible.

### 1.1. Digráfica asociada a una matriz cuadrada

La clase de matrices no negativas puede estudiarse a través de la teoría de digráficas con pesos positivos en las aristas, esto se debe a la inmediata correspondencia entre ambas clases de objetos. Sea  $A$  una matriz cuadrada de orden  $n$ . A esta matriz le asociamos una digráfica  $\Gamma$  con  $n$  vértices, que estarán denotados por  $1, 2, \dots, n$ , y habrá una arista del vértice  $j$  al vértice  $i$  con peso igual a  $A_{ij}$  para cada par  $i, j = 1, 2, \dots, n$ . En particular,  $\Gamma$  tendrá un ciclo en el vértice  $j$  con peso  $A_{jj}$ . Adicionalmente, las aristas con peso cero que corresponden a las entradas nulas de  $A$  pueden omitirse de  $\Gamma$  sin ningún efecto en nuestros cálculos, de hecho omitir estas aristas hace más transparentes los procedimientos, pues se revela de manera más clara la estructura de la digráfica.

Más concretamente, si  $A$  es una matriz cuadrada no negativa, entonces los conjuntos  $V = \{j: \exists i \text{ tal que } A_{ij} > 0 \text{ o } A_{ji} > 0\}$  y  $E = V \times V$  dan lugar a una digráfica  $\Gamma = (V, E)$ . Al considerar la función de pesos positivos  $\omega: E \rightarrow \mathbb{R}$  dada por  $\omega(j, i) = A_{ij}$ , obtenemos la digráfica con pesos  $(\Gamma, \omega)$ , a la que llamaremos la **digráfica asociada** a la matriz  $A$ .

El siguiente teorema nos indica que podemos conocer las entradas de la  $k$ -ésima potencia de nuestra matriz  $A$  al fijarnos en los pesos de todos los caminos de longitud  $k$  de su digráfica, donde el **peso** de un **camino** en  $\Gamma$  se define como el producto de los pesos de todas las aristas de dicho camino.

**1.1.1 Teorema.** *Sea  $A$  una matriz cuadrada de orden  $n$ . Para cada entero positivo  $k$ , la  $ij$ -ésima entrada  $A_{ij}^{(k)}$  de la  $k$ -ésima potencia  $A^k$  de  $A$  es igual a la suma de todos los pesos de todas las caminatas en  $\Gamma$  de longitud  $k$  del vértice  $j$  al vértice  $i$ .*

Una aplicación muy sencilla de este teorema consiste en determinar si una matriz cuadrada no negativa, es nilpotente o no, en términos de su digráfica.

**1.1.2 Corolario.** *Sea  $A$  una matriz cuadrada de orden  $n$ . Entonces  $A$  es nilpotente si su digráfica  $\Gamma$  no posee ningún ciclo, en tal caso  $A^n = 0$ . Si además  $A$  es no negativa, entonces se cumple el recíproco.*

Como podemos ver, esta correspondencia permite estudiar la teoría de matrices no negativas desde un punto de vista combinatorio, en [15, Cap. 6] se puede consultar a detalle la correspondencia entre la teoría de matrices no negativas y la teoría de digráficas con pesos positivos.

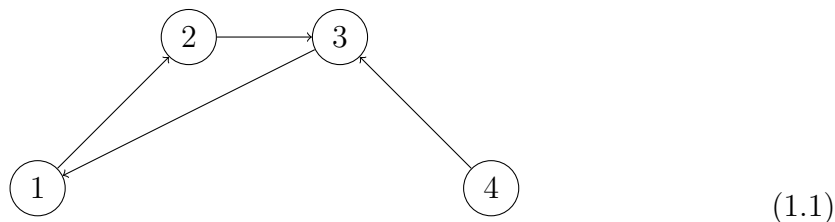
## 1.2. Matrices reducibles e irreducibles

Comenzamos por explorar el concepto de irreducibilidad para una matriz con entradas no negativas. Para esto, introducimos una relación en el conjunto de índices al cual llamaremos su **espacio de estados**.

Sea  $A$  una matriz cuadrada de orden  $N$  con entradas no negativas. Nos referimos a un elemento de  $S = \{1, \dots, N\}$  como un **estado** de  $A$ . Dados dos estados  $i$  y  $j$ , no necesariamente distintos, diremos que  $j$  **se comunica** con  $i$  si existe una potencia  $k \geq 0$  tal que  $(A^k)_{ij} > 0$ . Como estamos incluyendo el valor  $k = 0$  en esta definición, tenemos que cualquier estado  $j$  se comunica consigo mismo. En caso de que el estado  $j$  se comunique con el estado  $i$ , escribiremos  $j \rightarrow i$ . Cuando tengamos  $j \rightarrow i$  y  $i \rightarrow j$  simultáneamente, escribiremos  $i \leftrightarrow j$  y en este caso diremos que los estados  $i$  y  $j$  **se comunican**. En el contexto de la digráfica asociada a nuestra matriz no negativa



A, un vértice  $i$  se comunica con un vértice  $j$  si existe al menos un camino dirigido que comienza en  $i$  y termina en  $j$ . Esta relación es fácil de visualizar a través de la digráfica, por ejemplo, en la siguiente digráfica tenemos que  $1 \leftrightarrow 3$ , y al mismo tiempo  $4 \rightarrow 1$  pero 1 no se comunica con 4.



A continuación mostramos que  $\leftrightarrow$  es una relación de equivalencia.

**1.2.1 Teorema.** *La relación  $\leftrightarrow$  en el espacio de estados de una matriz no negativa  $A$  es una relación de equivalencia.*

*Demostración.* Como mencionamos anteriormente,  $i \leftrightarrow i$  para cualquier estado  $i$ , es decir, la relación  $\leftrightarrow$  es reflexiva. Es claro que la relación es simétrica, es decir,  $i \leftrightarrow j$  si y sólo si  $j \leftrightarrow i$ . Ahora veamos que la relación es transitiva: supongamos que  $i, j$  y  $\ell$  son estados tales que  $i \leftrightarrow j$  y  $j \leftrightarrow \ell$ . Como  $i \rightarrow j$  y  $j \rightarrow \ell$ , existen enteros no negativos  $k$  y  $k'$  tales que  $(A^k)_{ji} > 0$  y  $(A^{k'})_{\ell j} > 0$ , entonces

$$\begin{aligned} (A^{k+k'})_{\ell i} &= \sum_{m=1}^N (A^k)_{\ell m} (A^{k'})_{mi} \\ &\geq (A^k)_{\ell j} (A^{k'})_{ji} \\ &> 0, \end{aligned}$$

es decir,  $i \rightarrow \ell$ . De manera análoga podemos justificar que  $\ell \rightarrow i$ . Así, la relación es transitiva y hemos probado que  $\leftrightarrow$  es una relación de equivalencia.  $\square$

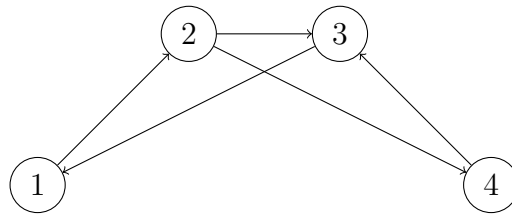
De hecho, esta relación de equivalencia se puede visualizar de la siguiente manera. Definamos el **soporte** de una matriz no negativa  $A$  como la matriz cuya  $ij$ -ésima entrada es 1 si  $A_{ij}$  es positiva y 0 en otro caso. Llamemos  $B$  al soporte de  $A$ , es fácil verificar que el soporte de  $B^k$  es el mismo que el soporte de  $A^k$  para cualquier entero positivo  $k$ . De esta manera, para verificar si dos vértices se comunican o no, podemos efectuar nuestros cálculos en el soporte de la matriz no negativa.

Más aún, notemos que en la demostración anterior se justificó que si los estados  $i, j$  y  $\ell$  verifican  $i \rightarrow j$  y  $j \rightarrow \ell$ , entonces  $i \rightarrow \ell$ . Esta observación es obvia cuando nos fijamos en la digráfica pues si  $i \rightarrow j$  y  $j \rightarrow k$ , entonces existen dos caminos dirigidos de  $i$  a  $j$  y de  $j$  a  $k$ , al concatenar estos dos caminos obtenemos un camino de  $i$  a  $k$ , es decir  $i \rightarrow k$ .

Antes de continuar con el estudio de esta relación de equivalencia, definimos la propiedad de irreducibilidad en términos de esta relación. La matriz  $A$  es **irreducible** si cualquiera dos de sus estados se comunican. Esto quiere decir que para una matriz irreducible su espacio de estados es la única clase de equivalencia respecto a la relación  $\leftrightarrow$ . Así, cualquiera dos vértices  $i$  y  $j$  de la digráfica de  $A$  se pueden unir con caminos dirigidos que parten tanto de  $i$  como de  $j$  si y sólo si  $A$  es una matriz irreducible. Por ejemplo, la matriz

$$\begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

es irreducible pues cualquiera dos vértices de su digráfica se pueden unir con caminos en ambas direcciones:



Pero dado que la matriz

$$\begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

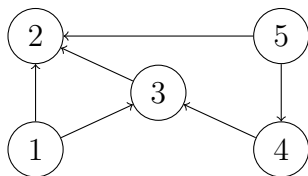
da lugar a la digráfica (1.1) en la cual 1 no se comunica con 4, esta matriz no puede ser irreducible.

Ahora es natural tomar una matriz arbitraria no negativa  $A$  y estudiar la partición que la relación  $\leftrightarrow$  induce en su espacio de estados. En esta dirección surge la idea de subconjunto cerrado del espacio de estados.

Un subconjunto no vacío  $C$  del espacio de estados  $S$  se dice que es **cerrado** si

$$A_{ij} = 0 \text{ cuando } j \in C \text{ pero } i \notin C.$$

Esta condición, básicamente expresa que no pueden salir aristas del conjunto  $C$ . En este contexto, si el estado  $i_0$  es tal que el conjunto  $\{i_0\}$  es cerrado, decimos que  $i_0$  es un **estado absorbente**. Por ejemplo, en la digráfica



de la matriz

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

notemos que el conjunto de estados  $C = \{1, 2, 3\}$  es cerrado, pues existen aristas que entran a  $C$  pero no al revés. Por otro lado, podemos ver que 2 es un estado absorbente pues sólo entran aristas a él.

**1.2.2 Proposición.** *Las siguientes afirmaciones se cumplen:*

- I. *Uniones arbitrarias e intersecciones no vacías de subconjuntos cerrados son subconjuntos cerrados.*
- II. *Si  $C_1$  y  $C_2$  son subconjuntos cerrados distintos que son minimales respecto a la inclusión, entonces  $C_1 \cap C_2 = \emptyset$ .*
- III. *Todo subconjunto cerrado contiene un subconjunto cerrado minimal respecto a la inclusión.*

*Demostración.* I. Esto es inmediato pues si  $C, D \subseteq \{1, \dots, N\}$  son subconjuntos cerrados, entonces  $A_{ij} = 0$  cuando  $j \in C \cup D$ ,  $i \notin C$  y  $i \notin D$ , es decir,  $C \cup D$  es cerrado. Si además  $C \cap D \neq \emptyset$ , entonces  $A_{ij} = 0$  cuando  $j \in C \cap D$  y se cumple  $i \notin C$  o  $i \notin D$ , es decir,  $C \cap D$  es un subconjunto cerrado.

- II. Si  $C_1 \cap C_2 \neq \emptyset$ , entonces del inciso anterior,  $C_1 \cap C_2$  es un subconjunto cerrado, lo cual contradice la minimalidad de  $C_1$  y  $C_2$ . Por lo tanto  $C_1 \cap C_2 = \emptyset$ .
- III. Esto se verifica rápidamente porque el espacio de estados es finito.

□

Por otro lado, notemos que podemos describir la forma de nuestra matriz no negativa salvo una permutación de sus filas y columnas. Como ejemplo, supongamos que  $C = \{1, \dots, N'\}$  es un subconjunto cerrado en el espacio de estados  $\{1, \dots, N\}$  con  $N' \leq N$ , entonces, de acuerdo con la definición de subconjunto cerrado, la matriz  $A$  se puede escribir como

$$\begin{bmatrix} * & \dots & * & * & \dots & * \\ \vdots & & \vdots & \vdots & & \vdots \\ * & \dots & * & * & \dots & * \\ 0 & \dots & 0 & * & \dots & * \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & * & \dots & * \end{bmatrix}.$$

Si  $C_1, \dots, C_r$  son todos los subconjuntos cerrados minimales del espacio de estados, estos subconjuntos deben ser disjuntos a pares de acuerdo con la proposición anterior. Reetiquetar los estados de  $A$  corresponde a conjugar por una matriz adecuada  $T$ , más específicamente,  $T$  se obtiene de reordenar las filas de la correspondiente matriz identidad y la matriz resultante está dada por  $TAT^{-1}$ . Entonces, después de un reordenamiento de los estados de  $A$ , tendremos

$$C_1 = \{1, \dots, N_1\}, C_2 = \{N_1 + 1, \dots, N_2\}, C_r = \{N_{r-1} + 1, \dots, N_r\}.$$

Con ayuda de la Proposición 1.2.2 podemos escribir a  $A$ , después de reordenar sus filas y columnas, o bien, después de conjugar por una matriz adecuada  $T$ , como sigue:

$$\begin{bmatrix} A_1 & 0 & 0 & \dots & 0 & 0 & * \\ 0 & A_2 & 0 & \dots & 0 & 0 & * \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & A_r & * \\ 0 & 0 & 0 & \dots & 0 & 0 & * \end{bmatrix}, \quad (1.2)$$

donde  $A_{k+1}$  es la matriz cuadrada no negativa de orden  $N_{k+1} - N_k$  que corresponde a tomar las filas y columnas del subconjunto cerrado  $C_k$ . En algunos casos podemos omitir las matrices  $*$ .

Esta observación nos motiva a pensar que el concepto de subconjunto cerrado minimal coincide con nuestras clases de equivalencia respecto a  $\leftrightarrow$ , ya que las potencias de la matriz (1.2) están dadas por

$$\begin{bmatrix} A_1^k & 0 & 0 & \dots & 0 & 0 & * \\ 0 & A_2^k & 0 & \dots & 0 & 0 & * \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & A_r^k & * \\ 0 & 0 & 0 & \dots & 0 & 0 & * \end{bmatrix}. \quad (1.3)$$

Esto quiere decir que al tomar un estado  $j$  en  $C_k$ , cualquier potencia  $A^t$  de  $A$  satisfará  $(A^t)_{ij} > 0$  únicamente para estados  $i \in C_k$ . El siguiente teorema nos da una justificación más precisa de este hecho.

**1.2.3 Teorema.** *Todo subconjunto cerrado minimal es una clase de equivalencia respecto a  $\leftrightarrow$ .*

*Demostración.* Sea  $i \in C$  fijo. Definamos  $C_{[i]}$  como la colección de aquellos estados  $j \in C$  que no se comunican con  $i$ :

$$C_{[i]} = \{j \in C : (A^k)_{ij} = 0, \forall k \geq 0\}.$$

Mostremos que  $C_{[i]} = \emptyset$ , pero primero veamos que esto demuestra que cualquiera dos estados en  $C$  se comunican. Si  $C_{[i]} = \emptyset$ , entonces

$$C = C \setminus C_{[i]} = \{j \in C : (A^k)_{ij} > 0, \text{ para algún } k \geq 0\},$$

es decir,  $C = \{j \in C : j \rightarrow i\}$  y esto sucedería para cada estado  $i$ . Entonces cualquiera dos estados  $i$  y  $j$  de  $C$  satisfarán  $i \leftrightarrow j$ . Ahora pasemos a demostrar que  $C_{[i]} = \emptyset$ . Para esto, mostraremos que  $C_{[i]}$  es un subconjunto propio de  $C$  que también es cerrado; esto es suficiente gracias a la minimalidad de  $C$ . La primera afirmación es inmediata pues  $C_{[i]} \subseteq C$  y además  $i \in C \setminus C_{[i]}$ . Para la segunda afirmación, sean  $k \in C_{[i]}$  y  $\ell \notin C_{[i]}$  dados. Si  $\ell \notin C$ , entonces  $A_{\ell k} = 0$  porque  $k \in C_{[i]} \subseteq C$  y  $C$  es cerrado. Por

otro lado, supongamos que  $\ell \in C$ . Dado que  $\ell \notin C_{[i]}$ , tenemos que  $\ell \rightarrow i$ , entonces necesariamente  $A_{\ell k} = 0$ , de no ser así,  $A_{\ell k} > 0$  y luego  $k \rightarrow \ell$  implicaría que  $k \rightarrow i$  por transitividad, lo cual es absurdo pues  $k \in C_{[i]}$ . En cualquier caso se tiene  $A_{\ell k} = 0$  para  $k \in C_{[i]}$  y  $\ell \notin C_{[i]}$ , es decir,  $C_{[i]}$  es cerrado. Entonces la minimalidad de  $C$  implica que  $C_{[i]} = \emptyset$ , como se quería.

Ahora probemos que  $C$  es la clase de equivalencia de algún estado  $j_0 \in C$ . Sea  $j_0 \in C$  y denotemos por  $[j_0]$  a la clase de equivalencia de  $j_0$  respecto a  $\leftrightarrow$ . Del argumento anterior se sigue que  $C \subseteq [j_0]$ . Veamos que  $[j_0] \subseteq C$ . Sea  $i$  un estado tal que  $i \leftrightarrow j_0$  y supongamos que  $i \notin C$ . Dado que  $C$  es un conjunto cerrado, tenemos que  $A_{ij_0} = 0$ . Más aún, recordando que  $A$  es una matriz de orden  $N$ ,

$$\begin{aligned} (A^2)_{ij_0} &= \sum_{k=1}^N A_{ik}A_{kj_0} \\ &= \sum_{k \in C} A_{ik}A_{kj_0} + \sum_{k \notin C} A_{ik}A_{kj_0} \\ &= 0 \end{aligned}$$

pues  $A_{ik} = 0$  cuando  $k \in C$ ,  $i \notin C$  y por otro lado  $A_{kj_0} = 0$  cuando  $k \notin C$  pues  $j_0 \in C$ . Un argumento por inducción demuestra que  $(A^n)_{ij_0} = 0$  para cualquier entero no negativo  $n$ . Con esto concluimos que  $j_0$  no se comunica con  $i$ , lo cual es absurdo. Entonces  $C = [j_0]$  y el teorema queda demostrado.  $\square$

El siguiente corolario se desprende del teorema anterior y de nuestra definición de irreducibilidad para una matriz no negativa.

**1.2.4 Corolario.** *Una matriz no negativa  $A$  es irreducible si y sólo si  $S$  es su único subconjunto cerrado.*

De acuerdo con el corolario anterior, es natural pensar que las clases de equivalencia respecto a  $\leftrightarrow$  dan lugar a matrices irreducibles, y al considerar la digráfica asociada, esta conclusión es inmediata. Diremos que una digráfica es **fuertemente conexa** si para cualquiera dos vértices distintos  $u$  y  $v$  existen caminos dirigidos de  $u$  a  $v$  y de  $v$  a  $u$ , es decir, nuestra digráfica es fuertemente conexa si su matriz no negativa asociada es irreducible, pues su única clase de equivalencia sería su espacio de estados.

De esta manera, las clases de equivalencia con las que trabajamos arriba inducen una partición  $V_1, \dots, V_r$  en el conjunto de vértices. Las subdigráficas inducidas a partir de estas clases de equivalencia son llamadas las **componentes fuertes** de la digráfica. Es claro que nuestra digráfica es fuertemente conexa si y sólo si tiene una única componente fuerte.

**1.2.5 Teorema.** *Sea  $G$  una digráfica con conjunto de vértices  $V$ .*

1. *Entonces  $G$  es fuertemente conexa si y sólo si no es posible encontrar una partición de  $V$  de dos conjuntos no vacíos  $U$  y  $W$  tales que todas las aristas entre vértices de  $U$  y  $W$  tienen sus vértices iniciales en  $U$  y sus vértices terminales en  $W$ .*
2. *Las componentes fuertes de  $G$  se pueden enumerar  $G_1, \dots, G_r$  de manera que si  $(x, y)$  es una arista de  $G$ , donde  $x \in G_i$ ,  $y \in G_j$  con  $i \neq j$ , entonces  $i < j$  (de manera que todas las aristas entre componentes fuertes van de izquierda a derecha).*

Otra forma de llegar a la expresión (1.2) es aplicando estos resultados a la digráfica  $\Gamma$  de la matriz  $A$ . Observamos que las filas y columnas de  $A$  pueden ser simultáneamente permutadas para obtener una forma diagonal de bloques llamada la **Forma normal de Frobenius**, en otras palabras, existe una matriz de permutación  $Q$  tal que

$$QAQ^T = \begin{bmatrix} A_1 & A_{12} & A_{13} & \cdots & A_{1r} \\ 0 & A_2 & A_{23} & \cdots & A_{2r} \\ 0 & 0 & A_3 & \cdots & A_{3r} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & A_r \end{bmatrix},$$

donde  $A_1, \dots, A_r$  son matrices cuadradas irreducibles. Dado que  $A_1, \dots, A_r$  corresponden a las componentes fuertes de  $\Gamma$ , y como las componentes fuertes están únicamente determinadas por los vértices que conforman cada una de las clases de equivalencia, entonces las matrices  $A_1, \dots, A_r$  están únicamente determinadas salvo permutaciones simultáneas de sus filas y columnas, es decir, salvo el orden en el cual los vértices de las componentes fuertes están escritos. Las matrices  $A_1, \dots, A_r$  son las **componentes irreducibles** de  $A$ . Es obvio que  $A$  es irreducible si y sólo si tiene

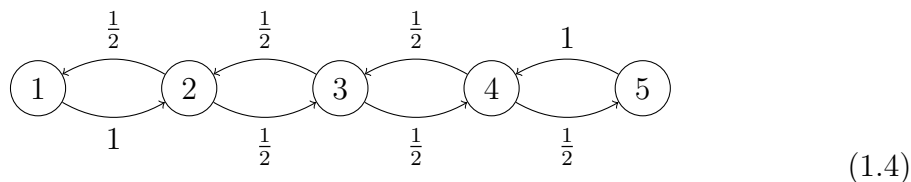
exactamente una componente irreducible. El orden en el cual aparecen las componentes irreducibles en la diagonal de la forma normal de Frobenius no es necesariamente único, depende de si las matrices  $A_{ij}$  son nulas o no.

### 1.3. Irreducibilidad, aperiodicidad y regularidad

Ahora presentamos el concepto de período de un estado de una matriz no negativa. Se define el **período** del estado  $j$ ,  $d_j$ , como

$$d_j = \gcd\{n \in \mathbb{N}: (A^n)_{jj} > 0\}.$$

Entonces un estado cuyo período es igual a uno es llamado **aperiódico**. Diremos que  $A$  tiene período  $d \geq 1$  si cada uno de sus estados tiene período igual a  $d$ . En caso de tener  $d = 1$ , decimos que  $A$  es **aperiódica**. Para entender este concepto, observemos la digráfica con pesos positivos



que viene de considerar a la matriz no negativa

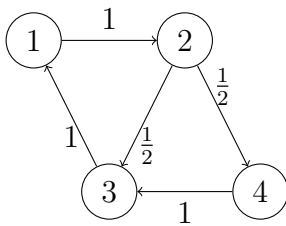
$$\begin{bmatrix} 0 & 1/2 & 0 & 0 & 0 \\ 1 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1 \\ 0 & 0 & 0 & 1/2 & 0 \end{bmatrix}. \quad (1.5)$$

Al pensar en las potencias de nuestra matriz como caminos de longitud fija en nuestra digráfica (1.4), es fácil verificar que el período de la matriz en (1.5) es 2 pues sólo es posible trazar caminos de orden par que comiencen en un vértice y terminen en ese mismo vértice. Por otro lado, la matriz

$$\begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1 \\ 0 & 1/2 & 0 & 0 \end{bmatrix} \quad (1.6)$$



da lugar a la digráfica



Entonces, dado que existen caminos de longitud 3 y 4 que parten y terminan en el vértice 1, ambos de probabilidad positiva, podemos concluir que el vértice 1 es aperiódico. Análogamente, podemos justificar que los vértices restantes son aperiódicos y por lo tanto la matriz en (1.6) es aperiódica. De alguna manera, es posible retornar al vértice de partida en caminos de longitud suficientemente grande. Esta última afirmación está justificada por el Lema 1.3.2 más adelante.

Hasta este momento, los conceptos de irreducibilidad y aperiodicidad parecen no tener relación entre ellos. Pronto veremos que estas dos nociones combinadas son equivalentes a lo que llamaremos regularidad. La siguiente observación nos coloca en esa dirección.

**1.3.1 Lema.** *Sea  $S \subseteq \mathbb{N}$  tal que*

- I.  $\gcd S = 1$  y también
- II.  $S$  es cerrado bajo adición, es decir, si  $a, a' \in S$ , entonces  $a + a' \in S$ .

*Entonces existe un entero positivo  $k$  tal que  $n \in S$  para cada  $n \geq k$ .*

*Demostración.* Primero veamos que existe un subconjunto finito  $F \subseteq S$  tal que  $\gcd F = 1$ . Supongamos que  $S = \{a_1, a_2, \dots\}$ . Notemos que la sucesión  $(d_n)$  dada por  $d_n = \gcd\{a_1, \dots, a_n\}$  es decreciente. Necesariamente la sucesión  $(d_n)$  es constante excepto por una cantidad finita de índices. Ahora es obvio que esta constante debe ser 1. La existencia de tal conjunto finito  $F$  se sigue.

Supongamos que  $F = \{a_1, \dots, a_n\}$ . Dado que  $\gcd F = 1$ , existen enteros  $r_i$  tales que

$$\sum_{i=1}^n r_i a_i = 1.$$

Esta suma podemos escribirla de la siguiente manera:

$$\sum_{r_i > 0} r_i a_i = 1 + \sum_{r_i < 0} (-r_i) a_i.$$

Dado que cada  $a_i \in F \subseteq S$  y  $S$  es cerrado bajo la adición, tenemos  $r_i a_i \in S$  cuando  $r_i > 0$  y  $(-r_i) a_i \in S$  cuando  $r_i < 0$ , de manera que al poner  $m = \sum_{r_i < 0} (-r_i) a_i$  se deduce que  $m \in S$  y  $m + 1 \in S$ .

Ahora veamos que las condiciones  $m, m + 1 \in S$  para algún  $m \in \mathbb{N}$  y que  $S$  un conjunto cerrado bajo la adición implican el resultado.

Observemos que  $m^2 + \ell \in S$ , para cada  $\ell = 0, 1, 2, \dots$ , pues para  $0 \leq \ell \leq m$  tenemos

$$m^2 + \ell = \underbrace{m + \dots + m}_{(m-\ell)\text{-veces}} + \underbrace{(m+1) + \dots + (m+1)}_{\ell\text{-veces}},$$

y si  $\ell > m$  pongamos  $\ell = mq + \ell'$ , para alguna  $q \in \mathbb{Z}$ , con  $0 \leq \ell' < m$  para obtener  $m^2 + \ell = m^2 + \ell' + mq$ . Del argumento anterior concluimos que  $m^2 + \ell' \in S$  y como  $mq \in S$ , se sigue que  $m^2 + \ell \in S$ . Entonces podemos tomar  $k = m^2$  en el enunciado.  $\square$

Este lema se relaciona con la aperiodicidad de la siguiente manera. Si  $i$  es un estado de una matriz aperiódica  $A$ , entonces cada uno de los conjuntos

$$S_i = \{n \geq 0 : (A^n)_{ii} > 0\}$$

satisface las condiciones del Lema 1.3.1 pues  $\gcd S_i = 1$  por definición, mientras que si  $a, b \in S_i$ , entonces  $(A^a)_{ii} > 0$  y también  $(A^b)_{ii} > 0$ , de manera que

$$\begin{aligned} (A^{a+b})_{ii} &= \sum_{\ell=1}^N (A^a)_{i\ell} (A^b)_{\ell i} \\ &\geq (A^a)_{ii} (A^b)_{ii} \\ &> 0, \end{aligned}$$

es decir  $a + b \in S_i$ .

Esta observación nos permite concluir el siguiente lema.

**1.3.2 Lema.** *Si  $A$  es una matriz aperiódica, entonces existe un entero positivo  $k$  tal que*

$$(A^n)_{ii} > 0$$

*para cada índice  $i$  y toda  $n \geq k$ .*

*Demostración.* Hemos visto que para cada estado  $i$ , el conjunto  $S_i$  satisface las condiciones del Lema 1.3.1. Entonces para cada estado  $i$  existe un entero positivo  $k_i$  tal que  $(A^n)_{ii} > 0$  para cada  $n \geq k_i$ . Pongamos  $k = 2 \max\{k_1, \dots, k_N\}$ , entonces para cualquier estado  $i$  y cualquier  $n \geq k$  se tiene que

$$\begin{aligned} (A^n)_{ii} &= \sum_{\ell=1}^N (A^{n-k_i})_{i\ell} (A^{k_i})_{\ell i} \\ &\geq (A^{n-k_i})_{ii} (A^{k_i})_{ii} \\ &> 0, \end{aligned}$$

precisamente porque  $n - k_i \geq k_i$  pues  $n \geq 2k_i$ . □

Decimos que una matriz no negativa  $A$  es **regular** si existe un entero positivo  $n$  tal que  $A^n$  es una matriz con entradas estrictamente positivas. Por conveniencia, introducimos la notación  $A > 0$  para indicar que  $A$  es una matriz no negativa con entradas estrictamente positivas, así, la matriz no negativa  $A$  es regular si  $A^n > 0$  para alguna  $n$ . Si  $\Gamma$  denota la digráfica de  $A$ , entonces  $A$  es regular si y sólo si existe un entero positivo  $n$  tal que dados cualquiera dos vértices  $i$  y  $j$  siempre es posible trazar caminos de  $i$  a  $j$  y de  $j$  a  $i$  de longitud  $n$  en  $\Gamma$ . Con estos resultados podemos probar que toda matriz irreducible y aperiódica es regular, e incluso un poco más.

**1.3.3 Teorema.** *Si  $A$  es una matriz no negativa, irreducible y aperiódica, entonces existe un entero positivo  $k$  tal que  $A^n > 0$  para todo  $n \geq k$ .*

*Demostración.* Dado que  $A$  es irreducible, para cada par de índices  $i$  y  $j$  existe un entero positivo  $n_{ij}$  tal que  $(A^{n_{ij}})_{ij} > 0$ . Por otro lado, como  $A$  es aperiódica, existe un entero positivo  $k'$  tal que  $(A^n)_{ii} > 0$  para cada índice  $i$  y toda  $n \geq k'$ , de acuerdo con el Lema 1.3.2.

Pongamos  $M_{ij} = k' + n_{ij}$  para cada par de índices  $i$  y  $j$ . Sea  $m \geq M_{ij}$ , entonces

$$\begin{aligned}(A^m)_{ij} &= \sum_{\ell=1}^N (A^{m-n_{ij}})_{i\ell} (A^{n_{ij}})_{\ell j} \\ &\geq (A^{m-n_{ij}})_{ii} (A^{n_{ij}})_{ij} \\ &> 0,\end{aligned}$$

pues  $m - n_{ij} \geq k'$  y luego  $(A^{m-n_{ij}})_{ii} > 0$  e igualmente  $(A^{n_{ij}})_{ij} > 0$  por la condición de regularidad en  $A$ . Al tomar  $k = 2 \max\{M_{ij} : i, j = 1, \dots, N\}$ , un argumento totalmente análogo al de la demostración del Lema 1.3.2 termina la prueba.  $\square$

Rápidamente verificamos que toda matriz regular necesariamente es irreducible, básicamente esto se cumple por definición. Sólo restaría justificar que toda matriz regular también es aperiódica. Esto es fácil y lo incluimos en el siguiente importante resultado.

**1.3.4 Corolario.** *Sea  $A$  una matriz no negativa. Entonces  $A$  es regular si y sólo si  $A$  es irreducible y aperiódica.*

*Demostración.* Si  $A$  es irreducible y aperiódica, entonces inmediatamente se sigue que  $A$  es regular gracias al Teorema 1.3.3.

Por otro lado, si  $A$  es regular, rápidamente concluimos que  $A$  es irreducible. Para ver que  $A$  necesariamente también es aperiódica, probaremos que  $A^n > 0$  y  $A^{n+1} > 0$  para  $n$  suficientemente grande. Sea  $n$  tal que  $A^n$  es una matriz estrictamente positiva, entonces

$$\begin{aligned}(A^{n+1})_{ij} &= \sum_{\ell=1}^N (A^n)_{i\ell} A_{\ell j} \\ &\geq (A^n)_{i\ell_0} A_{\ell_0 j}\end{aligned}$$

para cualquier estado  $\ell_0$ . Basta escoger una entrada no nula de la  $j$ -ésima columna de  $A$ . Observemos que

$$(A^2)_{ij} = \sum_{\ell=1}^N A_{i\ell} A_{\ell j} = 0$$

si la  $j$ -ésima columna de  $A$  es cero. De ser así, un argumento por inducción demuestra que  $(A^n)_{ij} = 0$  cualquiera que sea el índice  $i$ , pero esto es absurdo pues  $A^n > 0$ .

Entonces podemos escoger la entrada no nula de la  $j$ -ésima columna de  $A$  como decíamos para obtener que  $A^{n+1}$  también es una matriz estrictamente positiva.  $\square$

Cuando nuestra matriz  $A$  representa a una cadena de Markov, esta correspondencia nos permite introducir la noción de caminatas aleatorias en gráficas. En el capítulo 3 expondremos la teoría de probabilidad necesaria para este concepto.



## CAPÍTULO 2

# Cadenas de Markov

Una cadena de Markov puede describir el comportamiento de un sistema físico o matemático, indicando cómo pasa de un estado a otro, asumiendo que es posible representar su evolución en el tiempo como una sucesión de transiciones entre estados. Es entonces natural relacionar una cadena de Markov con una digráfica con pesos positivos, donde los vértices simbolizan los estados y los pesos de las aristas denotan la probabilidad de pasar de un estado a otro. Desde un punto de vista probabilista, requerimos que la transición de un estado a otro sólo dependa del estado actual y no de los estados previamente ocupados. Se pueden encontrar una gran cantidad de aplicaciones de cadenas de Markov, como por ejemplo en las ciencias de la computación, la teoría de la información y los motores de búsqueda en la web. En el artículo de Hilgers y Langville [9] se exponen estas aplicaciones y otras más con cierto detalle.

### 2.1. Clasificación de estados

Una **distribución (de probabilidad)** es cualquier vector  $\pi \in \mathbb{R}^N$  cuyas entradas son no negativas y suman 1. Escribiremos  $\pi(i)$  en lugar de  $\pi_i$  cuando sea conveniente. Diremos que una matriz no negativa  $A$  es una **matriz estocástica** si cada una de sus columnas es una distribución.

Una **cadena de Markov** es una sucesión de distribuciones  $x_0, Px_0, P^2x_0, \dots$ , donde  $P$  es una matriz estocástica de orden  $N$  y  $x_0$  es una distribución en  $\mathbb{R}^N$ .  $P$  se llama la **matriz de transición** y  $x_0$  la **distribución inicial** de la cadena. En tal caso, la cadena de Markov será denotada por la sucesión  $(P^n x_0)$  o por la pareja ordenada  $(P, x_0)$  indistintamente. Cuando la distribución inicial no sea de relevancia en los

argumentos, nos referiremos a la cadena simplemente por su matriz de transición  $P$ .

El propósito de esta definición es analizar el comportamiento asintótico de la sucesión de distribuciones  $(P^n x_0)_n$ . En particular, podemos preguntarnos sobre la convergencia de esta sucesión y la dependencia sobre la distribución inicial  $x_0$ . La rapidez con la que convergen cierto tipo de cadenas de Markov aún es objeto de investigación (ver el capítulo 3).

Como podemos ver, la matriz de transición de una cadena de Markov es simplemente una matriz estocástica, la cual es un caso particular de la clase de matrices no negativas. De esta manera, la teoría desarrollada en la sección anterior se puede utilizar para analizar nuestras cadenas de Markov. Entonces hablaremos del **espacio de estados** de una cadena de Markov pensando en el espacio de estados de su matriz de transición. La relación de **comunicación** entre estados, la noción del **período** de un estado y el concepto de **subconjunto cerrado** de una cadena de Markov se definen de manera similar. Igualmente hablaremos de cadenas **irreducibles**, **aperiódicas** o **regulares**, sin hacer distinción entre la cadena y su matriz de transición.

En esta discusión sobre clasificación de estados de una cadena de Markov, no haremos mucho énfasis en la distribución inicial de la cadena. Estudiaremos ese concepto con mayor detalle en la siguiente sección.

Distinguiremos dos tipos de estados en una cadena de Markov: estados recurrentes y estados transitorios. Un estado  $i$  de una cadena de Markov con matriz de transición  $P$  es **recurrente** si se cumple que

$$\sum_{k=0}^{\infty} (P^k)_{ii} = \infty \quad (2.1)$$

y es **transitorio** en otro caso, es decir, si la suma en (2.1) es convergente.

Nuestra primera observación es que la relación de comunicación respeta la noción de recurrencia.

**2.1.1 Proposición.** *Sean  $i$  y  $j$  dos estados de la cadena tales que  $i$  es recurrente y  $i \leftrightarrow j$ . Entonces  $j$  también es un estado recurrente.*

*Demostración.* Como los estados  $i$  y  $j$  se comunican entre sí, podemos encontrar enteros positivos  $k'$  y  $k''$  tales que  $(P^{k'})_{ji} > 0$  y  $(P^{k''})_{ij} > 0$ . Dado  $n \in \mathbb{N}$  tenemos



que

$$\begin{aligned} (P^{n+k'+k''})_{jj} &\geq (P^{k'})_{ji}(P^{n+k''})_{ij} \\ &\geq (P^{k'})_{ji}(P^n)_{ii}(P^{k''})_{ij}, \end{aligned}$$

de donde concluimos que  $\sum_{n=1}^{\infty} (P^n)_{jj} = \infty$  pues  $(P^{k'})_{ji} > 0$ ,  $(P^{k''})_{ij} > 0$  y como  $i$  es recurrente se tiene que  $\sum_{n=1}^{\infty} (P^n)_{ii} = \infty$ .  $\square$

La discusión en la sección anterior sobre subconjuntos cerrados minimales del espacio de estados de una matriz no negativa nos ayudará a concentrarnos en cadenas irreducibles de ser necesario. Si nuestra cadena no es irreducible, siempre podemos encontrar un subconjunto cerrado minimal, y de acuerdo con el Corolario 1.3.4, la cadena que resulta al restringir nuestro espacio de estados a este subconjunto cerrado minimal será irreducible, sólo bastará verificar que nuestras afirmaciones sobre esta nueva cadena puedan trasladarse a la cadena original. La demostración del siguiente teorema hace uso de esta estrategia, combinada con el hecho de que nuestro espacio de estados siempre es finito.

**2.1.2 Teorema.** *Siempre existe al menos un estado recurrente.*

*Demostración.* Sea  $C$  un subconjunto cerrado minimal del espacio de estados de una cadena de Markov con matriz de transición  $P$ . Podemos suponer que  $C = \{1, \dots, r\}$ , con  $r \leq N$ , ya que siempre es posible renombrar los estados de  $P$  para poder escribir  $C$  de esta manera, además, la propiedad de recurrencia no se ve afectada por este reetiquetamiento. Consideremos la cadena restringida al conjunto  $C$ , es decir, consideremos la cadena  $Q$  que resulta de la matriz de transición  $P$  obtenida al tomar las filas y columnas indexadas por  $C$  de  $P$ : la matriz de transición  $Q$  se obtiene al tomar las primeras  $r$  filas y columnas de  $P$  en este caso. Entonces esta nueva cadena  $Q$  es irreducible por el Corolario 1.3.4, pues subconjuntos cerrados minimales de  $C$  respecto a la matriz  $Q$  son también cerrados minimales respecto a la cadena  $P$ . Probaremos que cualquier estado de esta nueva cadena debe ser recurrente, como consecuencia tendremos que  $P$  posee al menos un estado recurrente.

Sea  $j_0 \in C$ , mostremos que  $j_0$  es un estado recurrente para  $Q$ . Primero notemos que  $\sum_{i=1}^r (Q^k)_{ij_0} = 1$  para cada  $k \geq 1$ , pues cada  $Q^k$  es una matriz estocástica.

Entonces

$$\sum_{k=1}^{\infty} \sum_{i=1}^r (Q^k)_{ij_0} = \sum_{i=1}^r \sum_{k=1}^{\infty} (Q^k)_{ij_0} = \infty$$

y luego  $\sum_{k=1}^{\infty} (Q^k)_{i_0 j_0} = \infty$  para algún estado  $i_0 \in C$ . (Es importante notar que es aquí donde utilizamos el hecho de que nuestra cadena de Markov es finita: es decir, tenemos una cantidad finita de estados.) Por el Teorema 1.2.3, todos los estados en  $C$  se comunican, y por lo tanto, existe un entero no negativo  $k'$  tal que  $(Q^{k'})_{j_0 i_0} > 0$ .

Entonces

$$\begin{aligned} (Q^{k'+k})_{j_0 j_0} &= \sum_{\ell=1}^r (Q^{k'})_{j_0 \ell} (Q^k)_{\ell j_0} \\ &\geq (Q^{k'})_{j_0 i_0} (Q^k)_{i_0 j_0} \end{aligned}$$

se cumple para toda  $k \geq 1$ . Al tomar la suma respecto a  $k$ , obtenemos  $\sum_{k=1}^{\infty} (Q^k)_{j_0 j_0} = \infty$ , es decir, el estado  $j_0$  es recurrente respecto a la cadena  $Q$ .

Por último, veamos que esto implica que los estados de  $C$  son recurrentes para la cadena  $P$ . Esto es cierto porque si  $1 \leq i, j \leq r$ , se tiene que

$$\begin{aligned} (P^2)_{ij} &= \sum_{k=1}^N P_{ik} P_{kj} \\ &= \sum_{k \leq r} P_{ik} P_{kj} + \sum_{k > r} P_{ik} P_{kj} \\ &= \sum_{k \leq r} P_{ik} P_{kj} \\ &= \sum_{k=1}^r Q_{ik} Q_{kj} \\ &= (Q^2)_{ij}, \end{aligned}$$

pues como el conjunto  $C$  es cerrado,  $P_{k\ell} = 0$  cuando  $\ell \in C$  y  $k \notin C$ , en este caso  $P_{k\ell} = 0$  cuando  $1 \leq \ell \leq r$  y  $r < k \leq N$ . Es fácil verificar que  $(P^n)_{ij} = (Q^n)_{ij}$  para cualquier  $n$  y cualquiera índices  $1 \leq i, j \leq N$  a través de un argumento inductivo sobre  $n$ . Entonces, en particular, todo estado recurrente de  $Q$  también será un estado recurrente de  $P$ .  $\square$

Notemos que el argumento utilizado en la demostración del teorema anterior consiste, en esencia, en utilizar el hecho de que podemos escribir a la matriz  $P$  en la

forma (1.2) salvo un reordenamiento de los estados de  $P$  y por lo tanto las potencias de  $P$  tienen la forma (1.3).

Este último resultado nos permite pensar en la recurrencia de un estado en términos de su clase de equivalencia respecto a la relación de comunicación. En el momento en el que un estado es recurrente, de acuerdo con la Proposición 2.1.1, toda su clase de equivalencia consiste de estados recurrentes. Por ejemplo, si nuestra cadena es irreducible, el teorema anterior nos dice que al menos tenemos un estado recurrente y dado que cualquiera dos estados se comunican, gracias a que la cadena es irreducible, todo estado de esta cadena debe ser recurrente. En este sentido, podemos considerar a una cadena de Markov finita irreducible como una cadena *recurrente*.

Cuando intentamos describir el caso general, nos fijamos en la unión de los subconjuntos cerrados minimales. En la demostración del teorema anterior se verificó que si  $C_1, \dots, C_r$  denotan todos los subconjuntos cerrados minimales, entonces todo estado en  $C_1 \cup \dots \cup C_r$  es recurrente. El siguiente teorema nos dice que los estados restantes deben ser transitorios.

**2.1.3 Teorema.** *Si  $T = \{1, \dots, N\} \setminus (C_1 \cup \dots \cup C_r)$ , entonces todos los estados en  $T$  son transitorios.*

## 2.2. Distribuciones estacionarias

En la sección 1.2 estudiamos las propiedades de irreducibilidad, aperiodicidad y regularidad para matrices no negativas para concluir en el Corolario 1.3.4 que las primeras dos propiedades en conjunto son equivalentes a la tercera. En esta sección exploraremos la propiedad de regularidad en el contexto de las cadenas de Markov. Si nuestra cadena de Markov es regular, podremos garantizar la existencia y unicidad de una distribución estacionaria para la cadena. Esta distribución nos permitirá entender el comportamiento asintótico de nuestra cadena de Markov.

Una distribución  $\pi$  en  $\mathbb{R}^N$  es una **distribución estacionaria** para una cadena de Markov, con matriz de transición  $P$ , si  $P\pi = \pi$ . Notemos que para que una cadena de Markov tenga una distribución estacionaria, necesariamente debe de tener al número uno como valor propio. En estas condiciones, una distribución estacionaria debe ser

un vector propio asociado a este valor propio.

Recordemos el célebre teorema del punto fijo de Banach que establece que cualquier función contractiva definida en un espacio métrico completo debe tener un único punto fijo.

**2.2.1 Teorema** (del punto fijo de Banach). *Sean  $(X, d)$  un espacio métrico completo y  $f: X \rightarrow X$  una función para la cual existe  $0 \leq \alpha < 1$  tal que*

$$d(f(x), f(y)) \leq \alpha d(x, y).$$

*Entonces  $f$  posee un único punto fijo  $x_0$ , que también satisface  $f^n(x) \rightarrow x_0$  cuando  $n \rightarrow \infty$  para cada  $x \in X$ .*

Buscamos construir una aplicación contractiva a partir de la matriz de transición  $P$  suponiendo que la cadena es regular. El siguiente lema es un primer paso en esta dirección. Definimos la norma  $\|\cdot\|_1$  de manera usual, de modo que si  $x \in \mathbb{R}^N$ , entonces

$$\|x\|_1 = \sum_{k=1}^N |x_k|.$$

**2.2.2 Lema.** *Sea  $P$  una matriz estocástica de orden  $N$ . Denotemos por  $C^1, \dots, C^N$  a las columnas de  $P$  y sea  $K \subseteq \mathbb{R}^N$  su envoltura convexa. Pongamos*

$$C = \frac{1}{2} \max_{i,j} \|C^i - C^j\|_1.$$

*Entonces*

- I.  $C \leq 1$ .
- II. *Si  $P$  es estrictamente positiva, entonces  $C < 1$ .*
- III.  $\text{diam}_1(K) = \sup_{x,y \in K} \|x - y\|_1 = 2C$ .
- IV.  $\|Px - Py\|_1 \leq C\|x - y\|_1$  para cada par de distribuciones  $x, y$ .

*Demostración.* I. Dado que cada columna de  $P$  es una distribución, tenemos que

$$\|C^i - C^j\|_1 \leq \|C^i\|_1 + \|C^j\|_1 = 2 \text{ para cada par de índices } i \text{ y } j.$$

II. Como vimos en el inciso anterior,  $\|C^i - C^j\|_1 \leq 2$  para cualquier par de índices  $i$  y  $j$ , por lo que  $C \leq 1$ . Veamos que  $C < 1$  cuando  $P$  es estrictamente positiva. Para esto, probaremos que si  $x, y$  son dos distribuciones tales que  $\|x - y\|_1 = 2$ , entonces  $y$  posee al menos una entrada nula. Supongamos entonces que  $x, y$  satisfacen estas condiciones. Pongamos  $S = \{i: x_i > y_i\}$ ,  $T = \{1, \dots, N\} \setminus S$  y notemos que  $S \neq \emptyset$ , pues de no ser así tendríamos  $0 \leq x_i \leq y_i$  para cada índice  $i$ , y como  $x, y$  son distribuciones, necesariamente  $x = y$ , lo cual no es posible pues luego  $\|x - y\|_1 = 0$ . Por otro lado:

$$\begin{aligned} \sum_{i=1}^N x_i + \sum_{i=1}^N y_i &= \|x\|_1 + \|y\|_1 \\ &= \|x - y\|_1 \\ &= \sum_{i \in S} (x_i - y_i) + \sum_{i \in T} (y_i - x_i) \end{aligned}$$

y al descomponer las sumas en los conjuntos  $S$  y  $T$  obtenemos que

$$\sum_{i \in S} y_i + \sum_{i \in T} x_i = 0.$$

En particular,  $y_i = 0$  para cada  $i \in S \neq \emptyset$ . Entonces es claro que si  $P$  es positiva, tendremos  $\|C^i - C^j\| < 2$  para cualquier par de índices  $i$  y  $j$ , por lo que  $C < 1$ .

III. Pongamos  $\delta = \text{diam}_1 K$  y denotemos por  $B_1(x, r)$  a la bola cerrada con centro en  $x$  de radio  $r$  respecto a la norma  $\|\cdot\|_1$ . Por definición tenemos que  $2C \leq \delta$ , veamos que también se cumple que  $\delta \leq 2C$ . Fijemos un índice  $j$ . Por la manera en la que definimos  $C$ , tenemos que para cualquier índice  $i$  se cumple que  $C^i \in B_1(C^j, 2C)$ . Dado que  $B_1(C^j, 2C)$  es convexa, se sigue que  $K \subseteq B_1(C^j, 2C)$ . Entonces  $K \subseteq B_1(C^j, 2C)$  para cada índice  $j$  y luego  $C^j \in B_1(x, 2C)$  para cada  $x \in K$ , por lo que concluimos que  $K \subseteq B_1(x, 2C)$  para cualquier  $x \in K$ . De esta manera  $\|x - y\|_1 \leq 2C$  para cada par de elementos  $x, y \in K$ . Así  $\delta \leq 2C$ .

IV. Probemos que si  $a = (a_1, \dots, a_N) \neq 0$  es tal que  $\sum_{i=1}^N a_i = 0$ , entonces

$$\|Pa\|_1 \leq C\|a\|_1.$$

Primero observemos que

$$\|a\|_1 = \sum_{i=1}^N |a_i| = \sum_{a_i > 0} a_i - \sum_{a_i < 0} a_i.$$

Como  $\sum_{i=1}^N a_i = 0$ , tenemos que

$$\sum_{a_i > 0} a_i = - \sum_{a_i < 0} a_i = \frac{1}{2} \|a\|_1.$$

Entonces

$$x_0 = \frac{2}{\|a\|_1} \sum_{a_i > 0} a_i C^i \quad y \quad y_0 = - \frac{2}{\|a\|_1} \sum_{a_i < 0} a_i C^i$$

son combinaciones lineales convexas de las columnas de  $P$  y satisfacen  $\|x_0 - y_0\|_1 \leq 2C$  por el inciso anterior. Por lo tanto

$$\|Pa\|_1 = \left\| \sum_{i=1}^N a_i C^i \right\|_1 = \frac{\|a\|_1}{2} \|x_0 - y_0\|_1 \leq \frac{\|a\|_1}{2} 2C = C \|a\|_1.$$

Se tiene que la conclusión se sigue para vectores  $a$  tales que la suma de sus entradas es cero. Al tomar dos distribuciones  $x, y$ , el argumento anterior es válido para  $x - y$ , y el resultado queda demostrado. □

Notemos que los incisos (ii) y (iv) del lema anterior nos ayudan a concluir que si  $P$  es la matriz de transición de una cadena de Markov regular, entonces una potencia  $P^{k_0}$  de  $P$  tendrá entradas estrictamente positivas, necesariamente será una matriz estocástica y luego la aplicación  $x \mapsto P^{k_0}x$  es una función contractiva al escoger el dominio adecuado. El inciso (iv) sugiere considerar al conjunto de todas las distribuciones en  $\mathbb{R}^N$ , el cual es completo pues es un subconjunto cerrado del espacio métrico completo  $\mathbb{R}^N$ . Como consecuencia del teorema del punto fijo de Banach, tenemos que nuestra aplicación  $x \mapsto P^{k_0}x$  tiene una única distribución fija. Esta es la aplicación que buscábamos.

**2.2.3 Teorema.** *Sea  $P$  una matriz estocástica. Las siguientes condiciones son equivalentes*

- I.  $P$  es regular,

II.  $P$  es irreducible y existe una distribución  $\pi$  tal que  $P^n x \rightarrow \pi$  cuando  $n \rightarrow \infty$  para cualquier distribución  $x$ ,

III. existe un entero positivo  $k$  tal que para todo  $n \geq k$  se tiene que  $P^n$  es positiva.

Más aún, en tal caso,

IV.  $\pi$  es la única distribución fija de  $P$ .

V.  $\pi$  posee entradas estrictamente positivas.

*Demostración.* (i)  $\Rightarrow$  (ii). Sea  $k_0 \leq 1$  tal que  $P^{k_0}$  es positiva. Claramente  $P$  es irreducible. Por la proposición anterior, tenemos que la correspondencia  $E \rightarrow E$  dada por  $x \mapsto P^{k_0}x$  es una contracción, donde  $E$  denota el conjunto de las distribuciones en  $\mathbb{R}^N$ . Entonces, del teorema del punto fijo de Banach, se concluye el resultado.

(ii)  $\Rightarrow$  (iii). Primero veamos que  $P^n x \rightarrow \pi$  cuando  $n \rightarrow \infty$  para toda distribución  $x$  si y sólo si  $P^n e_i \rightarrow \pi$  para todo índice  $i$ , donde  $\{e_1, \dots, e_N\}$  denota la base canónica de  $\mathbb{R}^N$ . Entonces (ii) sucede si y sólo si las columnas de  $P^n$  convergen a  $\pi$  cuando  $n \rightarrow \infty$ . De esta manera, si  $\pi$  es positiva, entonces se concluye (iii). Veamos que  $\pi$  es positiva. Supongamos, sin pérdida de generalidad, que  $\pi_1 = 0$ . Dado que  $P$  es irreducible, para cada  $j$  sea  $n_{1j}$  tal que  $(P^{n_{1j}})_{1j}$  es positiva. Entonces, para  $k$  entero positivo tenemos

$$\begin{aligned} (P^{n_{1j}+k})_{1j} &= \sum_{\ell=1}^N (P^{n_{1j}})_{1\ell} (P^k)_{\ell j} \\ &\geq (P^{n_{1j}})_{1j} (P^k)_{jj}. \end{aligned}$$

Al tomar el límite cuando  $k \rightarrow \infty$  tenemos

$$\pi_1 = 0 \geq (P^{n_{1j}})_{1j} \pi_j \quad \forall j = 1, \dots, N$$

porque las columnas de  $P^n$  convergen, cada una, a la distribución  $\pi$  cuando  $n \rightarrow \infty$ . Entonces  $(P^{n_{1j}})_{1j} \pi_j = 0$  y dado que  $(P^{n_{1j}})_{1j} \neq 0$  para cada índice  $j$ , se sigue que  $\pi = 0$ , lo cual contradice el hecho de que  $\pi$  es una distribución. Por lo tanto  $\pi$  es positiva y la afirmación (iii) se sigue.

(iii)  $\Rightarrow$  (i). Esta implicación es inmediata. □

Con esto podemos concluir que cualquier cadena regular posee una y sólo una distribución estacionaria. Más aún, para una cadena de este tipo la noción de distribución inicial no es tan relevante pues cualquier distribución  $x$  satisfará  $P^n x \rightarrow \pi$  cuando  $n \rightarrow \infty$  de acuerdo con el teorema anterior. En resumen, una cadena de Markov regular determina una aplicación contractiva y la distribución de probabilidad fija de esta aplicación es la única distribución estacionaria de esta cadena. De acuerdo con el Corolario 1.3.4, lo mismo se cumple para cadenas irreducibles y aperiódicas.

Muchas veces es más fácil verificar que una cadena es reversible respecto a una distribución  $\pi$  dada y esta simple condición es suficiente para concluir que dicha distribución es estacionaria para la cadena de Markov. Dada una cadena de Markov con matriz de transición  $P$ , de orden  $N$ , una distribución  $\pi \in \mathbb{R}^N$  es **reversible** para esta cadena si para cada par de índices  $i$  y  $j$  se tiene

$$A_{ij}\pi_j = A_{ji}\pi_i \quad (2.2)$$

Al tomar la suma respecto al índice  $i$  en la ecuación anterior, tenemos el siguiente teorema.

**2.2.4 Teorema.** *Sea  $A$  una matriz estocástica. Si  $\pi$  es una distribución reversible para  $A$ , entonces  $\pi$  es una distribución estacionaria para  $A$ .*

*Demostración.* Esto es claro pues

$$\begin{aligned} \sum_{j=1}^N A_{ij}\pi_j &= \sum_{j=1}^N A_{ji}\pi_i \\ &= \left( \sum_{j=1}^N A_{ji} \right) \pi_i \\ &= \pi_i \end{aligned}$$

para cada par de índices  $i$  y  $j$ . Es decir,  $A\pi = \pi$ . □

A partir de los últimos dos resultados de arriba, podemos concluir fácilmente que cualquier cadena de Markov regular y reversible respecto a una distribución  $\pi$  debe tener a dicha distribución  $\pi$  como única distribución estacionaria. De hecho podemos conseguir a esta distribución a partir del límite

$$\lim_{n \rightarrow \infty} P^n x,$$



el cual no depende de la distribución inicial  $x$ .

## 2.3. Teorema de Perron-Frobenius

El teorema de Perron-Frobenius nos brinda aún más información espectral respecto a la distribución estacionaria.

Anteriormente escribíamos  $A > 0$  para indicar que  $A$  es una matriz con entradas positivas. Extenderemos esta notación escribiendo  $A \geq B$  para indicar que  $A$  y  $B$  son dos matrices de mismas dimensiones que además satisfacen  $A_{ij} \geq B_{ij}$  para cada par de índices  $i, j$ . Análogamente, se sobreentiende el significado de  $A > B$ .

En la prueba del teorema de Perron-Frobenius haremos uso del siguiente lema.

**2.3.1 Lema.** *Si  $A$  es una matriz cuadrada no negativa de orden  $N$ , entonces*

$$(I + A)^{N-1} > 0,$$

donde  $I$  es la correspondiente matriz identidad de orden  $N$ , es decir,  $(I + A)^{N-1}$  tiene entradas estrictamente positivas.

Intuitivamente, este lema nos dice que una matriz cuadrada no negativa está cerca de ser regular, pues  $I + A$  lo será y podemos decir con exactitud qué potencia de  $I + A$  es positiva.

El resultado de Perron originalmente consistía en describir el espectro de matrices estrictamente positivas en el mismo sentido del siguiente teorema. Más tarde el resultado de Perron fue generalizado para el caso de matrices no negativas por Frobenius con una condición un poco más débil: irreducibilidad.

**2.3.2 Teorema (Perron-Frobenius).** *Sea  $A$  una matriz cuadrada de orden  $n$  con entradas reales no negativas. Si  $A$  es irreducible, entonces  $A$  posee un valor propio  $r$  que satisface*

- I.  $r > 0$  y existe un vector propio  $z > 0$  asociado al valor propio  $r$ ;
- II.  $r \geq |\lambda|$  para cada valor propio  $\lambda$  de  $A$ ;
- III.  $r$  es raíz simple del polinomio característico de  $A$ .

*Demostración.* Demostremos la existencia del valor propio  $r > 0$ . Para esto, sea  $\rho: \mathbb{R}^N \rightarrow \mathbb{R}$  dada por

$$\rho(x) = \min \left\{ \frac{(Ax)_i}{x_i} : x_i \neq 0, 1 \leq i \leq N \right\}.$$

Notemos que si  $\lambda \in \mathbb{R}$  no es cero, entonces

$$\rho(\lambda x) = \rho(x),$$

por lo que podemos restringir el dominio de  $\rho$  al conjunto  $M = \{x \in \mathbb{S}^N : x \geq 0\}$  sin perder información (sobre  $\rho$ ).

Nuestro objetivo es encontrar un máximo para la función  $\rho$ . Dado que  $M$  es compacto, bastaría verificar que  $\rho$  es continua en  $M$ . Para evitar este inconveniente, definimos al conjunto  $M^* = \{(I + A)^{N-1}x : x \in M\}$ . Rápidamente se verifica que  $M^*$  es compacto y consiste únicamente de vectores positivos. Esto último gracias al Lema 2.3.1.

Por otro lado, dado que  $\rho(x)x_i \leq (Ax)_i$  para cada  $1 \leq i \leq n$ , tenemos

$$\rho(x)x \leq Ax$$

y por definición,  $\rho(x)$  es el máximo número con esta propiedad. Si  $B$  es una matriz positiva de orden  $N$ , entonces para  $y, z \in \mathbb{R}^N$  con  $0 \leq y \leq z$  se tiene  $By \leq Bz$ . Es por esto que, si  $y = (E + A)^{N-1}x$ , entonces  $\rho(x)x \leq Ax$  implica

$$\rho(x)(E + A)^{N-1}x \leq (E + A)^{N-1}Ax = A(E + A)^{N-1}x,$$

es decir,  $\rho(x)y \leq Ay$ . Como  $\rho(y)$  es el máximo con esta propiedad, obtenemos  $\rho(x) \leq \rho(y)$ , con lo cual, el máximo de  $\rho$  se encuentra en  $M^*$  y como  $M^*$  consiste sólo de vectores positivos,  $\rho$  es continua ahí.

Entonces  $\rho$  posee un máximo en  $M^*$ . Sea  $z \in M^*$  tal que

$$\rho(z) = r := \max_{x \neq 0} \rho(x).$$

Probaremos lo siguiente acerca de  $r$ :

- (a)  $r > 0$ . Para  $u = (1, \dots, 1) \in \mathbb{R}^N$  tenemos  $\rho(u) = \min_{1 \leq i \leq 1} \sum_{j=1}^N A_{ij}$ , por lo que  $\rho(u) > 0$  y entonces  $\rho(z) = r > 0$  pues  $\rho(z) \geq \rho(u)$ .

(b)  $Az = rz$  y  $z > 0$ . Pongamos  $y = (E + A)^{N-1}z$  y supongamos que  $Az - rz \neq 0$ . Entonces, por el Lema 2.3.1, tenemos  $y > 0$ . Por otro lado

$$rz = \rho(z)z \leq Az,$$

es decir,  $Az - rz \geq 0$ . Al multiplicar por  $(E + A)^{N-1}$  obtenemos

$$(E + A)^{N-1}(Az - rz) > 0,$$

es decir,  $ry < Ay$ . Dado que  $A$  es continua, para  $\varepsilon > 0$  suficientemente pequeño se cumple  $(r + \varepsilon)y < Ay$ , lo cual es absurdo ya que tendríamos  $r < r + \varepsilon \leq \rho(y)$ . Así,  $Az = rz$ . Finalmente,

$$\begin{aligned} 0 < y &= (E + A)^{N-1}z = \sum_{k=0}^{N-1} \binom{N-1}{k} A^k z \\ &= \left( \sum_{k=0}^{N-1} \binom{N-1}{k} r^k \right) z \\ &= (1 + r)^{N-1} z \end{aligned}$$

y luego  $z > 0$ .

Esto prueba (I). Para probar (II) sean  $\alpha$  un valor propio de  $A$  y  $y$  tal que  $Ay = \alpha y$ . Entonces

$$|\alpha| |y_i| \leq \sum_{j=1}^N A_{ij} |y_j|.$$

Si ponemos  $y^+ = (|y_1|, \dots, |y_n|)$ , entonces tenemos

$$|\alpha| y^+ \leq Ay^+,$$

por lo que  $|\alpha| \leq \rho(y^+) \leq r$ .

La última afirmación utiliza un caso particular de la fórmula de Jacobi. En particular, se puede encontrar en [7, Cap. 13] □

## 2.4. Teorema ergódico de Markov

Hemos estudiado qué condiciones son suficientes en una cadena de Markov para que tenga una distribución estacionaria desde los puntos de vista matricial y espectral. Ahora consideramos un enfoque geométrico, combinado con algunos resultados

topológicos. A partir de una cadena de Markov definimos una transformación lineal a la cual llamaremos transformación de Markov. Las condiciones de irreducibilidad, aperiodicidad y regularidad de nuestra cadena se traducen al lenguaje del álgebra lineal gracias a esta correspondencia.

Empecemos por definir qué es una transformación de Markov. Una transformación lineal  $T: \mathbb{R}^N \rightarrow \mathbb{R}^N$  es una **transformación de Markov** si  $T(x)$  es una distribución de probabilidad para toda distribución de probabilidad  $x$ . En otras palabras, si  $\Delta^N$  denota el **N-simplejo estocástico**, es decir,

$$\Delta^N = \left\{ (x_1, \dots, x_N) : x_i \geq 0, \sum_{i=1}^N x_i = 1 \right\},$$

entonces  $T$  es una transformación de Markov si  $T(\Delta^N) \subseteq \Delta^N$ . En lo que sigue omitiremos el superíndice  $N$  mientras no haya riesgo a confusiones.

Es inmediato verificar que una transformación lineal  $T$  es una transformación de Markov si y sólo si su representación matricial  $P$  respecto a la base canónica de  $\mathbb{R}^N$  es una matriz estocástica, o bien,  $P$  es la matriz de transición de una cadena de Markov. Ahora que hemos establecido una correspondencia entre cadenas de Markov y transformaciones lineales, procedemos a explorar qué tipo de transformaciones lineales obtenemos al fijarnos en cadenas irreducibles y aperiódicas.

Si  $T$  es una transformación de Markov, entonces la sucesión de conjuntos  $(T^k(\Delta))_k$  es decreciente respecto a la contención de conjuntos, es decir, se tiene

$$\Delta \supseteq T(\Delta) \supseteq T^2(\Delta) \supseteq T^3(\Delta) \supseteq \dots$$

Como cada elemento de esta sucesión es un conjunto compacto, pues  $T$  es una aplicación continua, se sigue que la intersección de todos estos conjuntos es no vacía. Definamos entonces

$$\sigma = \bigcap_{k=1}^{\infty} T^k(\Delta).$$

Si llamamos  $P$  a la representación matricial de  $T$  respecto a la base canónica, nos damos cuenta que cualquier distribución estacionaria  $\pi$  de  $P$  satisface  $\pi \in \sigma$ . Entonces tiene sentido decir que un punto  $x_0 \in \Delta$  es una **distribución estacionaria** para  $T$  si  $x_0 \in \sigma$ . Definimos anteriormente el período de un estado de una cadena de

Markov y hablábamos de *estados* aperiódicos. A diferencia de las cadenas de Markov, hablaremos de *distribuciones* periódicas de la transformación de Markov  $T$ .

De acuerdo con el teorema del punto fijo de Brouwer, como  $T$  es continua y  $T(\Delta) \subseteq \Delta$ ,  $T$  posee al menos un punto fijo. Denotemos por  $\tau$  a la colección de todos los puntos fijos de  $T$ , de manera que  $\tau \neq \emptyset$ . Claramente  $\tau \subseteq \sigma$ . Así, un punto  $x_0 \in \Delta$  es **periódico** si  $x_0 \in \sigma \setminus \tau$ . Por otro lado, nuestra transformación  $T$  se llamará **aperiódica** si  $T$  no posee distribuciones periódicas, es decir,  $T$  será aperiódica si  $\sigma = \tau$  y diremos que  $T$  es **periódica** en otro caso. Notemos que si  $T$  es una transformación de Markov aperiódica, entonces toda distribución estacionaria de  $T$  será una distribución estacionaria para  $P$  en el sentido usual. Es decir, para transformaciones aperiódicas, la noción de distribución estacionaria no cambia.

Al igual que en el caso de las cadenas Markov, también estamos interesados en distinguir cuándo una transformación posee una única distribución estacionaria. Este tipo de transformaciones se llamarán **estables**. Más específicamente, una transformación  $T$  es estable si  $\sigma = \tau$  consiste de un único punto. Si además este punto está en el interior topológico de  $\Delta$ , entonces diremos que  $T$  es **ergódicamente estable**. En otras palabras,  $T$  es ergódicamente estable si posee una única distribución estacionaria y es estrictamente positiva.

Para discutir el concepto análogo de irreducibilidad para transformaciones de Markov, primero definimos qué es un subsimplejo de  $\Delta^N$ . Dados un entero no negativo  $r \leq N - 1$  y una colección  $\{k_1, \dots, k_r\} \subseteq \{1, \dots, N\}$ , pongamos

$$\Delta_{k_1, \dots, k_r}^{N-r} = \{(x_1, \dots, x_N) \in \partial\Delta : x_{k_i} = 0, 1 \leq i \leq r\}.$$

Cualquiera de estos conjuntos será un **subsimplejo** de  $\Delta^N$ . De esta manera, diremos que una transformación  $T$  es **invariante** sobre un subsimplejo  $\Delta'$  si  $T(\Delta') \subseteq \Delta'$  y finalmente, llamaremos **irreducible** a  $T$  si el único subsimplejo sobre el cual  $T$  es invariante es el simplejo  $\Delta^N$  mismo. Esta vez, a diferencia del concepto de aperiodicidad, las nociones son equivalentes pues en el momento que una transformación  $T$  se reduce (es decir, no es irreducible), es claro que  $P$  también debe reducirse.

Con estos conceptos en mente, se pueden probar una serie de resultados que llevan a una demostración sencilla del teorema ergódico de Markov [16] que enunciamos a continuación.

**2.4.1 Teorema** (Ergódico de Markov). *Una transformación de Markov  $T$  es ergódicamente estable si y sólo si  $T^k(\Delta) \subseteq \Delta^\circ$  para algún entero positivo  $k$ .*

Es claro que la segunda condición es equivalente a pedir regularidad en la representación matricial  $P$ .

Para ver qué significa este teorema en términos de cadenas de Markov, exploremos qué quiere decir que nuestra transformación sea ergódicamente estable. Si nuestra transformación  $T$  es ergódicamente estable, entonces, tomando en cuenta que  $\sigma$  es un conjunto unipuntual, toda distribución  $x_0$  satisfará  $T^k(x_0) \in \Delta$  para cada  $k$  y luego, debido a la compacidad de  $\Delta$ , podemos suponer que la sucesión  $(T^k(x_0))_k$  es convergente a una distribución  $x \in \Delta$ , más aún,  $x \in \Delta \setminus \partial\Delta$ , es decir,  $x$  es una distribución estrictamente positiva. Utilizando la continuidad de  $T$ ,

$$T(x) = \lim_{k \rightarrow \infty} T^{k+1}(x_0) = x$$

y por lo tanto  $x \in \tau = \sigma$ . Finalmente, si  $T$  fuera una transformación reducible, entonces podríamos encontrar un subsimplejo  $\Delta'$  tal que  $T(\Delta') \subseteq \Delta'$  y entonces existe una distribución  $y \in \Delta'$  que satisface  $y \in \tau$  de acuerdo con un argumento anterior. Esto no tiene sentido pues  $\Delta' \subseteq \partial\Delta$  y dado que  $\tau = \sigma$  es un punto, tendríamos  $x = y$  con  $x$  en el interior de  $\Delta$  y  $y$  en su frontera.

Del inciso (ii) del Teorema 2.2.3 se sigue que una transformación de Markov ergódicamente estable debe tener su representación matricial  $P$  regular. Por otro lado, de acuerdo con el Corolario 1.3.4 y el resultado de arriba, una matriz irreducible y aperiódica debe inducir una transformación ergódicamente estable. El siguiente corolario resume esta discusión.

**2.4.2 Corolario.** *Una cadena de Markov es ergódicamente estable si y sólo si su matriz de transición  $P$  es tal que  $P^k$  posee todas sus entradas estrictamente positivas para algún  $k$ .*

Dicho de otra forma, una cadena de Markov es irreducible y aperiódica si y sólo si es regular.

## CAPÍTULO 3

# Algoritmo de Metropolis-Hastings

*The Metropolis algorithm is in the top 10 list of the computer algorithms with the greatest influence on the development and practice of science and engineering in the 20th century.*

J. Dongarra & F. Sullivan

Hemos visto que cierto tipo de cadenas de Markov admiten una distribución estacionaria, en particular aquellas que sean irreducibles, de acuerdo con el teorema de Perron-Frobenius. Ahora comenzamos con una distribución de probabilidad  $\pi$  y procedemos a construir una cadena de Markov irreducible y aperiódica que tenga a  $\pi$  como única distribución estacionaria, de acuerdo con el Corolario 1.3.4 y el Teorema 2.2.3, con ayuda del algoritmo de Metropolis-Hastings. El primer paso será construir lo que llamaremos la cadena de Metropolis-Hastings. Esta cadena se obtiene partiendo de una matriz previamente especificada, que llamaremos cadena propuesta, que es irreducible y aperiódica con el fin de que la cadena de Metropolis-Hastings herede estas propiedades. Entonces el algoritmo de Metropolis-Hastings viene de considerar la distribución estacionaria de una caminata aleatoria asociada a la cadena de Metropolis-Hastings. A continuación exploramos la íntima relación entre las cadenas de Markov y las caminatas aleatorias en digráficas con pesos.

### 3.1. Un enfoque probabilista

Como podemos ver, es posible estudiar la naturaleza de los estados recurrentes prescindiendo de la teoría de la probabilidad. Sin embargo, con sólo algunas nociones

básicas de probabilidad, podremos deducir un par de resultados que nos ayudarán a adquirir valiosa intuición sobre los estados recurrentes de una cadena de Markov. Un tratamiento detallado en este sentido se puede encontrar en [2, Cap. 4].

La manera usual en la que se define una cadena de Markov es como sigue. Si  $S$  es un conjunto finito de orden  $N$  y  $\nu$  es una distribución en  $\mathbb{R}^N$ , entonces una cadena de Markov, con espacio de estados  $S$ , distribución inicial  $\nu$  y matriz de transición  $P$ , es una sucesión finita de variables aleatorias  $\xi_0, \dots, \xi_n: T \rightarrow S$  que satisfacen  $\mathbb{P}(\xi_0 = x) = \nu(x)$  y

$$\mathbb{P}(\xi_{k+1} = x_{k+1} | \xi_0 = x_0, \xi_1 = x_1, \dots, \xi_k = x_k) = P_{k,k+1},$$

para toda  $k = 0, 1, \dots, n-1$  y cualquiera  $x, x_1, \dots, x_n \in S$  con  $\mathbb{P}(\xi_0 = x_0, \xi_1 = x_1, \dots, \xi_k = x_k) \neq 0$ . En esta definición asumimos que  $P$  es una matriz estocástica de orden  $N$  y que  $\mathbb{P}$  es la correspondiente medida de probabilidad en  $T$ . Si tomamos a la distribución de Dirac  $\delta_i$ , dada por  $\delta_i(i) = 1$  y  $\delta_i(j) = 0$  en otro caso, diremos que la cadena de Markov tiene **inicio determinista** en  $i$ .

En este sentido, podemos decir que una sucesión infinita  $(\xi_n)_n$  de variables aleatorias  $\xi_n: T \rightarrow S$  es una cadena de Markov si toda sucesión finita  $\xi_0, \dots, \xi_n$  es una cadena de Markov en el sentido de la definición anterior. En lo que sigue supondremos que  $(\xi_n)_n$  es una cadena de Markov con distribución inicial  $\nu$  y matriz de transición  $P$ .

La relación entre esta nueva definición de cadena de Markov y la anterior es que la probabilidad de pasar de un estado al siguiente está dada por las columnas de la matriz de transición. Más aún, uno puede probar que la probabilidad de pasar de un estado a otro después de  $k$  pasos está dada por las columnas de la  $k$ -ésima potencia  $P^k$  de  $P$  y entonces podemos ver cómo es posible abstraer esta definición de cadenas de Markov a través de las potencias de la matriz de transición.

Ahora estamos listos para regresar a la noción de recurrencia. Comenzamos por definir un par de probabilidades en las que se basará nuestra nueva definición de estado recurrente. Más precisamente, denotemos por  $f_{ij}^{(k)}$  a la probabilidad de que nuestra cadena visite al estado  $j$  partiendo del estado  $i$  por primera vez en el  $k$ -ésimo



paso, es decir, si nuestra cadena de Markov tiene inicio determinista en  $i$ ,

$$f_{ij}^{(k)} = \mathbb{P}(\xi_1 \neq j, \dots, \xi_{k-1} \neq j, \xi_k = j).$$

Con esta notación, el número  $f_{ij}^* = \sum_{k=1}^{\infty} f_{ij}^{(k)}$  es la probabilidad de que una caminata que comienza en  $i$  ocupará el estado  $j$  en algún paso en el futuro. Entonces un estado  $i$  se llamará **recurrente** si  $f_{ii}^* = 1$ . Si  $f_{ii}^* < 1$ , entonces diremos que el estado es **transitorio**.

Intuitivamente, un estado  $i$  será recurrente si podemos garantizar que la cadena visitará al estado  $i$  una infinidad de veces. Análogamente, este estado será transitorio si existe una probabilidad no nula de que la cadena nunca regresará al estado  $i$ , o en otras palabras,  $i$  será transitorio si la cadena de Markov sólo visitará a este estado una cantidad finita de veces, posiblemente cero. Nuestro primer resultado confirma nuestra intuición.

**3.1.1 Teorema.** *Un estado  $i$  es recurrente si y sólo si una caminata que comienza en  $i$  regresa una infinidad de veces al estado  $i$  con probabilidad igual a uno. El estado  $i$  es transitorio si y sólo si la probabilidad de que la caminata regrese al estado  $i$  una infinidad de veces es cero.*

*Demostración.* Notemos que la probabilidad de que una caminata ocupe el estado  $i$  después de exactamente  $k + k'$  pasos está dada por  $f_{ii}^{(k)} f_{ii}^{(k')}$ . Entonces la probabilidad de que la caminata regrese al estado  $i$  al menos dos veces está dada por  $\sum_{k, k' \geq 1} f_{ii}^{(k)} f_{ii}^{(k')} = (f_{ii}^*)^2$ . Procediendo de manera análoga, encontramos que  $\mathbb{P}(B_k) = (f_{ii}^*)^k$ , donde  $B_k$  denota el evento “la caminata regresa al estado  $i$  al menos  $k$  veces”. La sucesión de conjuntos  $(B_k)_k$  es decreciente y por lo tanto la probabilidad de su intersección está dada por

$$\mathbb{P}\left(\bigcap_{k=1}^{\infty} B_k\right) = \lim_{k \rightarrow \infty} (f_{ii}^*)^k = \begin{cases} 0 & \text{si } f_{ii}^* < 1 \\ 1 & \text{si } f_{ii}^* = 1 \end{cases}.$$

Es claro que la intersección  $\bigcap B_k$  denota el evento “la caminata regresa al estado  $i$  una infinidad de veces”. De esta manera, si  $i$  es un estado transitorio, la probabilidad de regresar a este estado una infinidad de veces es cero. Por otro lado, si el estado es recurrente, la probabilidad de regresar a  $i$  una infinidad de veces es exactamente igual a uno.  $\square$

Nuestro siguiente movimiento será verificar que nuestras dos nociones de recurrencia para un estado dado son en realidad equivalentes. Como preámbulo haremos un par de cálculos para conseguir una relación directa entre la suma parcial  $\sum_{t=1}^{k_0} (P^t)_{ii}$  y la probabilidad  $f_{ii}^*$ . Denotemos por  $E_k$  al evento “la caminata ocupa la posición  $i$  en el  $k$ -ésimo paso”. Sabemos que  $E_k$  tiene probabilidad  $(P^k)_{ii}$ . Primero notemos que  $E_k$  es la unión disjunta de los eventos

$$\begin{aligned} F_{1,k} &= \{X_1 = X_k = i\}, \\ F_{2,k} &= \{X_1 \neq i, X_2 = X_k = i\}, \\ &\vdots \\ F_{k,k} &= \{X_1 \neq i, \dots, X_{k-1} \neq i, X_k = i\}. \end{aligned}$$

Rápidamente se verifica que  $\mathbb{P}(F_{k-t,k}) = f_{ii}^{(k-t)}(P^t)_{ii}$  para  $t = 0, \dots, k-1$ , teniendo en mente que  $(P^0)_{ii} = 1$ . Entonces  $(P^k)_{ii} = \sum_{t=0}^{k-1} f_{ii}^{(k-t)}(P^t)_{ii}$ .

Por otro lado, para  $k_0$  arbitrario se tiene:

$$\begin{aligned} \sum_{k=1}^{k_0} (P^k)_{ii} &= \sum_{k=1}^{k_0} \sum_{t=0}^{k-1} f_{ii}^{(k-t)}(P^t)_{ii} \\ &= \sum_{t=0}^{k_0-1} (P^t)_{ii} \sum_{k=t+1}^{k_0} f_{ii}^{(k-t)} \\ &\leq \left( \sum_{t=0}^{k_0} (P^t)_{ii} \right) f_{ii}^* \\ &= \left( 1 + \sum_{t=1}^{k_0} (P^t)_{ii} \right) f_{ii}^* \end{aligned}$$

y por lo tanto

$$(1 - f_{ii}^*) \sum_{t=1}^{k_0} (P^t)_{ii} \leq f_{ii}^* \quad (3.1)$$

Por último, necesitaremos el célebre lema de Borel-Cantelli, el cual se puede encontrar en [10, p. 51].

**3.1.2 Lema** (Lema de Borel-Cantelli). *Sean  $E_1, E_2, \dots$  eventos y pongamos  $E^* = \limsup_{n \rightarrow \infty} E_n$ , es decir,  $E^* = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_k$ . Entonces si  $\sum_{n=1}^{\infty} \mathbb{P}(E_n) < \infty$ , se tiene que  $\mathbb{P}(E^*) = 0$ .*

Ahora estamos listos para demostrar la esperada equivalencia.

**3.1.3 Teorema.** *Para un estado  $i$ , las siguientes condiciones son equivalentes:*

- I. *La probabilidad  $f_{ii}^* = 1$ .*
- II. *La suma  $\sum_{n=1}^{\infty} (P^n)_{ii} = \infty$ .*

*De manera análoga,  $f_{ii}^* < 1$  si y sólo si  $\sum_{n=1}^{\infty} (P^n)_{ii}$  es convergente.*

*Demostración.* De la desigualdad (3.1) se deduce que si el estado  $i$  satisface  $f_{ii}^* < 1$ , entonces  $\sum_{t=0}^{\infty} (P^t)_{ii}$  es convergente. Por otro lado, si  $\sum_{t=0}^{\infty} (P^t)_{ii} = \infty$ , necesariamente  $f_{ii}^* = 1$ . Ahora supongamos que  $\sum_{t=0}^{\infty} (P^t)_{ii} < \infty$ . Entonces  $\sum_{k=0}^{\infty} \mathbb{P}(E_k) < \infty$  y por el lema de Borel-Cantelli concluimos que la caminata ocupará el estado  $i$  una infinidad de veces con probabilidad cero. Entonces, de acuerdo con la Proposición 3.1.1, el estado  $i$  satisface  $f_{ii}^* < 1$ . Así, es claro que si el estado  $i$  es recurrente respecto a la definición probabilista, entonces  $\sum_{t=0}^{\infty} \mathbb{P}(E_k) = \infty$  por simple contraposición de la afirmación anterior.  $\square$

## 3.2. Caminatas aleatorias

Como vimos al final de la sección 1.2, tenemos una correspondencia biunívoca entre digráficas con pesos positivos en las aristas y matrices con entradas no negativas. Al normalizar apropiadamente la matriz de incidencia, podemos construir una cadena de Markov a partir de una digráfica con pesos  $(\Gamma, \omega)$ . Por simplicidad asumiremos que el conjunto de vértices es  $\{1, \dots, N\}$ . Para cada vértice  $u$  consideremos la suma de los pesos que parten de  $u$ ,  $W_u$ , es decir

$$W_i = \sum_{(i,j) \in E} \omega_{ij},$$

la cual servirá de constante de normalización para la cadena de Markov. Por cada vértice  $(i, j) \in E$  pongamos entonces

$$P_{ji} = \frac{\omega_{ij}}{W_i}$$

y  $P_{ji} = 0$  si  $(i, j) \notin E$ . De esta manera  $P$  es una matriz estocástica, llamada la cadena de Markov inducida por  $(\Gamma, \omega)$ .

Ahora, sea  $(P, x_0)$  una cadena de Markov con distribución inicial  $x_0$ . Definimos una **caminata aleatoria** en la digráfica con pesos  $(\Gamma, \omega)$  asociada a  $P$  de longitud  $k$  como un elemento  $X$  de  $V^{k+1}$  tal que  $X_0$  fue elegido de  $V$  con probabilidad  $x_0$ ,  $X_1$  fue elegido de  $V$  con probabilidad  $\omega(X_0, \cdot)/W_0$  y así sucesivamente. Al omitir la longitud de la caminata nos referimos a un elemento  $X$  del conjunto de sucesiones con valores en  $V$ , denotado por  $V^\infty$ , tal que  $X_0$  fue elegido de  $V$  con probabilidad  $x_0$  y para  $k \geq 1$ ,  $X_{k+1}$  fue elegido de  $V$  con probabilidad  $\omega(X_k, \cdot)/W_k$ . En este sentido, hablaremos de una caminata aleatoria **simple** en una gráfica no dirigida  $\Gamma$  si nos referimos a la digráfica con pesos  $(\Gamma, \omega)$  donde  $\omega_{uv} = 1/d_u$  con  $d_u$  igual al grado del vértice  $u$ , de manera que si la caminata se encuentra en el estado  $i$ , todos los estados que se pueden ocupar tienen la misma probabilidad de ocurrir.

De esta manera, podemos estudiar gráficas y caminatas aleatorias en gráficas con pesos a través del lenguaje de las cadenas de Markov. Por ejemplo, es posible definir un vértice recurrente de  $(\Gamma, \omega)$  como aquel vértice cuyo estado asociado a él sea recurrente respecto a la cadena de Markov  $P$ . Entonces, para una caminata aleatoria en  $(\Gamma, \omega)$  asociada a  $P$ , un estado es recurrente si la caminata visita a este estado una infinidad de veces de acuerdo con el Teorema 3.1.1. Análogamente, la relación de comunicación entre estados se puede interpretar como una relación de equivalencia en el conjunto de vértices de  $\Gamma$ : un vértice  $u$  se comunicará con otro vértice  $v$  si existe un camino dirigido de aristas que comienza en  $u$  y termina en  $v$ . Exposiciones detalladas sobre estas relaciones se pueden encontrar en [15, Cap. 6] y [14, Cap. 5].

En particular, podemos interpretar la noción de distribución estacionaria para cadenas de Markov en términos de sus caminatas aleatorias asociadas. Consideremos una cadena de Markov  $(P, x_0)$  y sea  $X$  una caminata aleatoria asociada a  $P$ . Para cada  $k \geq 1$ , sea  $X^{(k)}$  la caminata aleatoria de longitud  $k$  que resulta de tomar las primeras  $k$  entradas de  $X$ . Denotemos por  $Y^{(k)}$  al vector en  $\mathbb{R}^N$  cuya  $j$ -ésima entrada  $Y_j^{(k)}$  es igual a la cantidad de veces que aparece el estado  $j$  en la caminata  $X^{(k)}$  para cada  $j = 1, 2, \dots, N$ , de manera que las entradas del vector  $Z^{(k)} = \frac{1}{k}Y^{(k)}$  se pueden interpretar como la cantidad promedio de veces que la caminata  $X$  visita a cada uno de los estados después de  $k$  pasos. Con la teoría que hemos desarrollado hasta ahora, no es de sorprenderse que la sucesión  $(Z^{(k)})_k$  converge cuando  $P$  es una

cadena regular, es decir, cuando  $P$  es irreducible y aperiódica. Además, si cualquier sucesión  $(Z^{(k)})_k$  proveniente de una caminata  $X$  asociada a  $P$  converge a un límite que no depende de la caminata  $X$ , entonces necesariamente  $P$  debe ser regular. Bajo estas condiciones, decimos que cualquier límite  $\lim Z^{(k)}$ , cuando  $k \rightarrow \infty$ , es la **distribución estacionaria** de la caminata  $X$ .

El siguiente argumento sirve para convencernos de que este hecho es verdad. Supongamos que  $P$  es regular y sea  $X$  una caminata aleatoria asociada a  $P$ . De acuerdo con el Teorema 2.2.3, las columnas de  $P^k$  convergen a una distribución  $\pi$  cuando  $k \rightarrow \infty$ . Por otro lado, dado que  $P^k x \rightarrow \pi$  para cualquier distribución  $x$  cuando  $k \rightarrow \infty$ , no importa que distribución inicial se haya tomado para  $X$ , para  $n$  suficientemente grande se tiene que la probabilidad de que  $X_n = j$  es aproximadamente  $\pi_j$ . ¿Cuál es el significado de la sucesión  $(Z^{(k)})_k$ ? Después de  $n$  pasos, al tomar un estado al azar de manera uniforme, la probabilidad de que este estado sea  $j$  será  $Z_j^{(n)}$ , es decir, la probabilidad de que  $X_n = j$  será aproximadamente  $Z_j^{(n)}$ . Entonces

$$\lim_{k \rightarrow \infty} Z^{(k)} = \pi$$

y este límite no depende de la caminata  $X$ . Podemos ver que si esto se cumple para toda caminata aleatoria  $X$ , entonces se satisfará la condición (ii) del Teorema 2.2.3 y por lo tanto la cadena  $P$  será regular.

### 3.3. Cadenas de Metropolis-Hastings

Nos concentraremos en el caso en el que la distribución  $\pi$  es estrictamente positiva. Sea entonces  $Q$  la matriz de transición de una cadena irreducible y aperiódica. A esta matriz  $Q$  la llamaremos **matriz propuesta** (o **cadena propuesta**). Pasamos a definir la matriz de transición de la **cadena de Metropolis-Hastings**, para  $i, j \in \{1, 2, \dots, N\}$  distintos pongamos

$$M_{ij} = \frac{1}{\pi_j} \min\{\pi_i Q_{ji}, \pi_j Q_{ij}\}$$

de manera que

$$M_{ij}\pi_j = M_{ji}\pi_i \tag{3.2}$$

cuando  $i \neq j$ . Notemos que  $0 \leq M_{ij} \leq 1$ , pues  $M_{ij} \leq Q_{ij} \leq 1$ . Por esta misma razón

$$\sum_{j, j \neq i} M_{ij} \leq 1,$$

ya que  $Q$  es una matriz estocástica.

Por último, ponemos

$$M_{jj} = 1 - \sum_{j, j \neq i} M_{ij}$$

y llamemos  $M$  a la matriz con entradas  $M_{ij}$ , donde  $1, \leq i, j \leq N$ . De acuerdo con la manera en la que construimos a  $M$ , es claro que  $M$  es una matriz estocástica. También se cumple la ecuación (3.2) para cualquier par de índices  $i$  y  $j$ , en otras palabras,  $M$  es reversible respecto a la distribución  $\pi$ . Entonces  $\pi$  es una distribución estacionaria para la cadena de Metropolis-Hastings por el Teorema 2.2.4.

Para poder expresar a la cadena de Metropolis-Hastings más fácilmente en términos de una caminata aleatoria, supondremos que  $Q_{ij} > 0$  si y sólo si  $Q_{ji} > 0$  con el fin de evitar la división por cero. En este caso, observemos que cuando  $M_{ij} > 0$ ,

$$\begin{aligned} M_{ij} &= \frac{1}{\pi_j} \min\{\pi_i Q_{ji}, \pi_j Q_{ij}\} \\ &= \min\left\{\frac{\pi_i}{\pi_j} Q_{ji}, Q_{ij}\right\} \\ &= \min\left\{\frac{\pi_i Q_{ji}}{\pi_j Q_{ij}}, 1\right\} Q_{ij}. \end{aligned} \tag{3.3}$$

Sólo faltaría verificar que la cadena de Metropolis-Hastings hereda las propiedades de irreducibilidad y aperiodicidad de la matriz propuesta [6].

**3.3.1 Teorema.** *Supongamos que la matriz propuesta  $Q$  es irreducible, aperiódica y satisface  $Q_{ij} > 0$  si y sólo si  $Q_{ji} > 0$  para cada par de índices  $i$  y  $j$ . Si además la distribución de probabilidad  $\pi$  es estrictamente positiva, entonces la cadena de Metropolis-Hastings es una cadena de Markov irreducible y aperiódica que además es reversible respecto a la distribución  $\pi$ .*

*Demostración.* De acuerdo con la igualdad (3.3), si  $Q_{ij} > 0$ , necesariamente  $M_{ij} > 0$  pues

$$M_{ij} = \begin{cases} \frac{\pi_i}{\pi_j} Q_{ji} & \text{si } \pi_i Q_{ji} \leq \pi_j Q_{ij} \\ M_{ij} & \text{otro caso} \end{cases}.$$

Es por esto que  $M$  hereda las propiedades de irreducibilidad y aperiodicidad de  $Q$ .  $\square$

De hecho, si buscamos que la cadena de Metropolis-Hastings sea irreducible y aperiódica, sólo puede suceder que su distribución estacionaria sea positiva, de manera que la restricción  $\pi > 0$  en realidad es una condición necesaria.

**3.3.2 Teorema** (Algoritmo de Metropolis-Hastings). *Sea  $Q$  una cadena propuesta irreducible y aperiódica que además satisface  $Q_{ij} > 0$  si y sólo si  $Q_{ji} > 0$ . Sea  $\pi$  una distribución positiva. Entonces el algoritmo siguiente describe una caminata aleatoria asociada a la cadena de Metropolis-Hastings. De acuerdo con el Teorema 3.3.1, la distribución  $\pi$  es la única distribución estacionaria de esta caminata.*

- (1) *Supongamos que la caminata está ubicada en la posición  $j$  en este momento. Escogemos a  $i$  de acuerdo con la matriz  $Q$ , es decir, con probabilidad  $Q_{ij}$ .*
- (2) *Si  $\pi_j Q_{ij} \leq \pi_i Q_{ji}$ , entonces la siguiente posición de la caminata será  $i$ .*
- (3) *Si, por el contrario,  $\pi_i Q_{ji} < \pi_j Q_{ij}$ , la probabilidad de que la siguiente posición de la caminata sea  $i$  será  $\pi_i Q_{ji} / \pi_j Q_{ij}$  y lo opuesto para  $j$ , es decir, la probabilidad de que la siguiente posición de la caminata sea  $j$  será  $1 - \pi_i Q_{ji} / \pi_j Q_{ij}$ .*

Observemos que de acuerdo con la caminata de arriba, no existen abigüedades si alguna entrada de la distribución  $\pi$  es cero. De igual manera, podemos escoger una matriz propuesta cualquiera y el algoritmo de Metropolis-Hastings sigue funcionando, salvo un par de detalles. Si la cadena propuesta no es regular, existe la posibilidad de que la cadena de Metropolis-Hastings tampoco lo sea y entonces no podemos garantizar la convergencia del algoritmo. También podría suceder que la cadena de Metropolis-Hastings tenga más de una distribución estacionaria y que la convergencia a distintas distribuciones estacionarias dependa directamente de la distribución inicial elegida.

El algoritmo de Metropolis-Hastings surge de generalizar el algoritmo de Metropolis, el cual asume que la matriz propuesta es simétrica. Entonces la **cadena de Metropolis** es simplemente la cadena de Metropolis-Hastings asociada a una matriz

propuesta  $Q$  simétrica [11, Cap. 5], de manera que de la ecuación (3.3) se sigue que para cada par de índices  $i$  y  $j$ ,

$$M_{ij} = \min \left\{ \frac{\pi_i}{\pi_j}, 1 \right\} Q_{ij}$$

y podemos escribir explícitamente a la matriz  $M$  como sigue:

$$M_{ij} = \begin{cases} Q_{ij} & \text{si } \pi_j \leq \pi_i \\ \frac{\pi_i}{\pi_j} Q_{ij} & \text{si } \pi_i < \pi_j \end{cases}.$$

**3.3.3 Teorema** (Algoritmo de Metropolis). *Sea  $Q$  una matriz propuesta irreducible, aperiódica y simétrica. Si  $\pi$  es una distribución positiva, el siguiente algoritmo describe una caminata aleatoria asociada a la cadena de Metropolis. Al igual que en el algoritmo anterior, la distribución  $\pi$  es la única distribución estacionaria de esta caminata.*

- (1) *Supongamos que la caminata está ubicada en la posición  $j$  en este momento. Escogemos a  $i$  de acuerdo con la matriz  $Q$ , es decir, con probabilidad  $Q_{ij}$ .*
- (2) *Si  $\pi_j \leq \pi_i$ , entonces la siguiente posición de la caminata será  $j$ .*
- (3) *Si, por el contrario,  $\pi_i < \pi_j$ , la probabilidad de que la siguiente posición de la caminata sea  $i$  será  $\pi_i/\pi_j$  y lo opuesto para  $j$ , es decir, la probabilidad de que la siguiente posición de la caminata sea  $j$  será  $1 - \pi_i/\pi_j$ .*

Un nuevo estado  $i$  menos favorable que el estado anterior  $j$  no es rechazado inmediatamente, en lugar de esto, es aceptado con una probabilidad que decrece con el incremento de la diferencia  $\pi_j - \pi_i$ . Esto permitirá que el algoritmo no se quede estancado en mínimos locales de la distribución  $\pi$ .

Como habíamos indicado anteriormente, las condiciones de irreducibilidad y aperiocidad sólo se utilizan para que la cadena de Metropolis pueda heredarlas de la matriz propuesta. A diferencia del algoritmo de Metropolis-Hastings, en el algoritmo de Metropolis requerimos que nuestra cadena propuesta sea simétrica para que el algoritmo tenga sentido.

Como ejemplo, consideremos  $\Gamma = (V, E)$  una gráfica conexa finita y una caminata aleatoria simple en  $\Gamma$ , es decir, la matriz de incidencia asociada a la caminata aleatoria



está dada por  $Q_{uv} = 1/d_u$ , donde  $d_u$  es el grado del vértice  $u \in V$ . Si tomamos a esta matriz  $Q$  como matriz propuesta, entonces la cadena de Metropolis-Hastings está dada por

$$M_{ij} = \min \left\{ \frac{\pi_i d_i}{\pi_j d_j}, 1 \right\} \frac{1}{d_i},$$

siguiendo a la ecuación (3.3). La matriz propuesta necesariamente es irreducible debido a la conexidad de  $\Gamma$  y del argumento de la demostración del Teorema 3.3.1 concluimos que la cadena de Metropolis-Hastings asociada también es irreducible.

Por último, si  $M$  denota la cadena de Metropolis-Hastings, es decir,  $M$  proviene de una cadena propuesta  $Q$  irreducible y aperiódica que además satisface  $Q_{ij} > 0$  si y sólo si  $Q_{ji} > 0$ , entonces

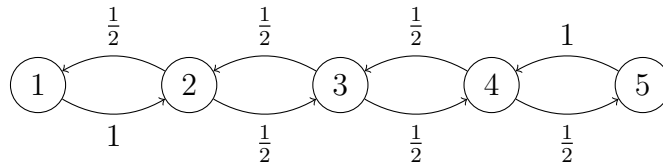
$$\lim_{n \rightarrow \infty} (M^n)_{ij} = \pi_i,$$

donde  $\pi$  es la única distribución estacionaria de  $M$ , de acuerdo con los Teoremas 3.3.1 y 2.2.3. A pesar de tener este importante resultado, en la práctica es necesario saber qué tan grande debería ser el exponente  $n$  para tener  $M_{ij}^n$  suficientemente cercano a  $\pi_i$ . En el artículo [6] se dan estimaciones de la convergencia de la cadena de Metropolis-Hastings para una colección de ejemplos de naturaleza simétrica. El análisis de estos ejemplos está en términos de la geometría de la digráfica con pesos asociada a la cadena de Metropolis-Hastings.

### 3.4. Ejemplos

La eficiencia del algoritmo de Metropolis-Hastings depende de la matriz propuesta  $Q$ . En los siguientes ejemplos comparamos la rapidez con la que la caminata aleatoria se aproxima a la distribución estacionaria al considerar la caminata sobre gráficas lineales, cíclicas y completas. En estos casos, la matriz propuesta será la matriz de transición asociada a la caminata aleatoria simple en estas gráficas.

Fijemos  $N = 5$  y consideremos una gráfica lineal de orden cinco. Nos concentraremos en una caminata aleatoria simple en la que la probabilidad de pasar del estado uno al estado dos será uno y lo mismo al pasar del estado cinco al cuatro. La gráfica con pesos es entonces:



Esta caminata aleatoria simple da lugar a nuestra cadena de Markov propuesta con matriz transición de orden cinco dada por

$$Q = \begin{bmatrix} 0 & 1/2 & 0 & 0 & 0 \\ 1 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1 \\ 0 & 0 & 0 & 1/2 & 0 \end{bmatrix}.$$

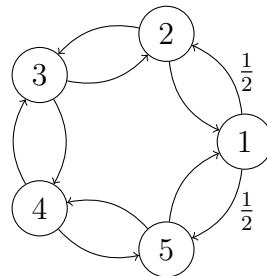
Pongamos

$$\pi = (0,22580569, 0,25035357, 0,25089408, 0,24560642, 0,02734024)$$

y utilicemos el algoritmo de Metropolis-Hastings (Teorema 3.3.2) para encontrar una caminata aleatoria  $(X_n)$  con distribución asociada que se aproxima a  $\pi$ . Las aproximaciones a la distribuciones estacionaria partiendo de una caminata aleatoria de longitud 30000 se muestran truncando esta caminata en los pasos 10000, 20000 y 30000 en la siguiente tabla.

$\pi$	0.22580569	0.25035357	0.25089408	0.24560642	0.02734024
$n = 10000$	0.2109	0.2432	0.2568	0.2614	0.0277
$n = 20000$	0.2123	0.24525	0.25485	0.25785	0.02975
$n = 30000$	0.21696667	0.24673333	0.254	0.2542	0.0281

Para un 5-ciclo, como se muestra en la siguiente figura, donde cada una de las aristas dirigidas tiene peso  $1/2$ .



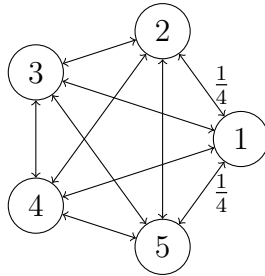
Nuestra cadena propuesta tiene a

$$Q = \begin{bmatrix} 0 & 1/2 & 0 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 0 & 1/2 & 0 \end{bmatrix}$$

como matriz de transición. Al igual que en el ejemplo anterior, a través del algoritmo de Metropolis-Hastings construimos una caminata aleatoria de longitud 30000 y mostramos las aproximaciones que resultan de truncar esta caminata en los pasos 10000, 20000 y 30000 en la siguiente tabla.

$\pi$	0.22580569	0.25035357	0.25089408	0.24560642	0.02734024
$n = 10000$	0.219	0.2497	0.2555	0.2486	0.0272
$n = 20000$	0.2315	0.24925	0.2476	0.2436	0.02805
$n = 30000$	0.23243333	0.25116667	0.2484	0.24046667	0.02753333

Finalmente, veamos el comportamiento del algoritmo de Metropolis-Hastings para una gráfica completa de orden cinco. Las flechas  $\leftrightarrow$  representan dos aristas dirigidas: una de ida y otra de regreso, cada una de ellas con peso igual a  $1/4$ .



En este caso, la matriz propuesta está dada por

$$\begin{bmatrix} 0 & 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 0 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 0 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 0 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 & 0 \end{bmatrix}.$$

Tenemos la siguiente tabla.

$\pi$	0.22580569	0.25035357	0.25089408	0.24560642	0.02734024
$n = 10000$	0.2245	0.2503	0.2574	0.24	0.0278
$n = 20000$	0.224	0.24825	0.2579	0.2434	0.02645
$n = 30000$	0.2256	0.24773333	0.25326667	0.24726667	0.02613333



## CAPÍTULO 4

# Teorema de Polya vía redes eléctricas

*A drunk man will eventually find  
his way home, but a drunk bird may  
get lost forever*

S. Kakutani

Hemos tratado con cadenas de Markov con espacio de estados finito, comúnmente llamadas cadenas de Markov finitas y ahora nos concentramos en cadenas de Markov cuyo espacio de estados puede ser numerable. En particular, la noción de caminata aleatoria con una cantidad numerable de estados es totalmente análoga a su contraparte finita. El teorema de recurrencia de Polya es un resultado fundamental en la teoría de caminatas aleatorias y es por eso que en la segunda parte de este trabajo damos dos demostraciones totalmente distintas de dicho teorema. La primera se apoya en la teoría de redes eléctricas en gráficas, para lo cual introduciremos las ideas básicas de esta teoría: en particular, las ideas de voltaje, flujo de corriente, las leyes de Kirchoff, la ley de Ohm y la noción de resistencia efectiva, para después exponer la relación entre esta última y el concepto de recurrencia de una caminata aleatoria en una gráfica con pesos [8, Cap. 1]. La segunda consiste en utilizar argumentos combinatorios con un poco de teoría de funciones especiales, concluyendo con el principio de Laplace, un razonamiento empleado en el análisis asintótico [13].

Dada una gráfica y un vértice inicial, seleccionar aleatoriamente un vecino de manera uniforme y moverse a dicho vecino, entonces repetir el proceso y así continuar sucesivamente. La sucesión de vértices obtenida de esta manera se llama una

caminata aleatoria simple en la gráfica. El teorema de recurrencia de Polya dice lo siguiente.

**4.0.1 Teorema (Polya).** *Toda caminata aleatoria simple en  $\mathbb{Z}^d$  es recurrente para  $d \in \{1, 2\}$  y es transitoria para  $d \geq 3$ .*

La teoría de redes eléctricas en gráficas es el resultado de generalizar la teoría de circuitos eléctricos vía las leyes de Kirchoff y la ley de Ohm. Más precisamente, pensamos en un circuito eléctrico como una gráfica con pesos, interpretando a los pesos como resistencias o conductancias. Las nociones de voltaje y corriente eléctrica se tratarán simplemente como funciones con valores reales definidas en los vértices y las aristas, respectivamente, que además satisfagan las leyes de voltaje y de corriente de Kirchoff, en conjunto con la ley de Ohm.

Veremos que establecer un voltaje y una corriente en una gráfica con pesos es equivalente a definir una función en los vértices con valores reales que sea armónica con valores específicos en la fuente y el sumidero de nuestra gráfica, en otras palabras, estamos hablando del problema de Dirichlet discreto cuando el conjunto frontera consta de únicamente dos vértices. Como podemos ver, el problema de Dirichlet discreto está íntimamente relacionado con la teoría de redes eléctricas en gráficas y sirve como generalización de esta teoría cuando el conjunto frontera contiene más de dos vértices.

Para demostrar el teorema de recurrencia de Polya en dimensiones 1 y 2 con este enfoque, es necesario introducir el concepto de resistencia efectiva, ya que se cumple que una caminata es recurrente si y sólo si la resistencia efectiva es infinita. Entonces en dimensiones bajas debemos demostrar que la resistencia efectiva es infinita mientras que en dimensiones mayores o iguales a 3 operamos de manera distinta: en este caso debemos definir qué entenderemos por un flujo y su energía disipada. Básicamente, un flujo será una generalización de una corriente eléctrica. Por otro lado, su energía disipada será su energía pensando en el correspondiente circuito eléctrico. Entonces, cuando la dimensión es mayor o igual a 3, lo que haremos es encontrar un flujo con energía disipada finita para demostrar que toda caminata será transitoria.

Para poder enunciar el problema de Dirichlet discreto, necesitamos definir qué entendemos por una función armónica en una gráfica con pesos. De hecho, veremos que

la función de voltaje en una red eléctrica es necesariamente armónica y que además es solución a un problema de Dirichlet muy particular, uno en el que la frontera consta de dos elementos: la fuente y el sumidero de nuestra red eléctrica. Finalmente, este problema de Dirichlet en particular sirve como introducción al problema de Dirichlet más general en el que el conjunto frontera es cualquier subconjunto de nuestro conjunto de vértices.

## 4.1. Funciones armónicas

Consideremos una gráfica conexa finita no dirigida  $\Gamma$  con conjunto de vértices  $V$  y conjunto de aristas  $E$ , y sea  $w: E \rightarrow \mathbb{R}^+$  una función de pesos positivos sobre el conjunto de aristas. Por conveniencia escribiremos  $w_{uv}$  en lugar de  $w(\{u, v\})$  en caso de que  $\{u, v\} \in E$ , de manera que  $w_{uv} = w_{vu}$ . Diremos que la arista  $e \in E$  tiene **conductancia**  $w_e$  y análogamente diremos que  $e$  tiene **resistencia**  $1/w_e$ , sin embargo algunas veces es útil interpretar nuestra función de pesos directamente como resistencias en lugar de conductancias para simplificar cálculos.

Una vez especificadas la gráfica  $\Gamma$  y su colección de conductancias  $w$ , nos concentramos en el par  $(\Gamma, w)$ . Recordemos que si  $A \subseteq \mathbb{C}$  es un conjunto abierto, entonces una función continua  $h: A \rightarrow \mathbb{R}$  es armónica si y sólo si para cualquier  $r > 0$  suficientemente pequeño y cualquier punto  $a \in A$  se cumple la propiedad del valor medio dada por

$$h(a) = \frac{1}{2\pi} \int_0^{2\pi} h(a + re^{i\theta}) d\theta.$$

Con el fin de imitar este comportamiento en gráficas con pesos, consideremos a la cadena de Markov en  $(\Gamma, w)$  que tiene como matriz de transición a la matriz  $P$  dada por

$$P_{uv} = \frac{w_{uv}}{W_v},$$

donde

$$W_v = \sum_{u \in V} w_{vu}.$$

Antes de continuar, notemos que la manera en la que definimos una cadena de Markov en la gráfica  $(\Gamma, w)$  nos garantiza que estamos tratando con una cadena

reversible respecto a la distribución  $\pi$  dada por  $\pi_v = W_v/W^*$ , donde  $W^* = \sum_{u \in V} W_u$ :

$$P_{uv}\pi_v = \frac{w_{uv}}{W^*} = P_{vu}\pi_u.$$

Más aún, dado que  $\Gamma$  es conexa, cualquiera dos estados distintos de la cadena de Markov asociada se comunican, es decir, la cadena es irreducible. Entonces la cadena de Markov asociada a una gráfica con pesos es irreducible y reversible respecto a la distribución definida arriba.

Entonces decimos que la función  $f: V \rightarrow \mathbb{R}$  es **armónica** en  $U$  si satisface una clase de propiedad de valor medio como en el caso complejo, más precisamente, si

$$f(u) = \sum_{v \in V} P_{vu}f(v), \quad u \in U.$$

Las funciones de probabilidad de llegada son ejemplos básicos de funciones armónicas. Estas funciones calculan la probabilidad de que una caminata aleatoria que comienza en un punto dado llegue a otro punto dado antes de pasar por un cierto conjunto de vértices. Por ejemplo, dados  $W \subseteq V$ ,  $U = V \setminus W$  y  $s \notin W$ , definamos a  $g: U \rightarrow \mathbb{R}$  de manera que si  $u \in U$ ,  $g(u)$  es la probabilidad de que una caminata aleatoria  $Z$  en  $(\Gamma, w)$  que parte de  $u$ , llegue al estado  $s$  antes que a  $W$ , en otras palabras,

$$g: U \rightarrow \mathbb{R}, \quad g(u) = \mathbb{P}(Z_n = s \text{ para alguna } n < T_W | Z_0 = u),$$

donde

$$T_W = \min\{n \geq 0: Z_n \in W\}.$$

Es importante notar que esta función no es trivial pues a pesar de que  $g(v) = 0$  para cualquier  $v \in W$ , se tiene que claramente  $g(s) = 1$ .

**4.1.1 Proposición.** *La función  $g$  definida arriba es armónica en  $U \setminus \{s\}$ .*

*Demostración.* Esto es inmediato, pues

$$\begin{aligned} g(v) &= \sum_{u \in V} P_{uv} \mathbb{P}(Z_n = s \text{ para alguna } n < T_W | Z_1 = v, Z_0 = u) \\ &= \sum_{u \in V} P_{uv} g(v), \end{aligned}$$

como se quería. □



## 4.2. Voltajes y corrientes

Ahora pasemos a definir qué entenderemos por las leyes de voltaje y de corriente de Kirchhoff, así como la ley de Ohm. Para esto, fijemos dos vértices distinguidos  $s, t \in V$  distintos y consideremos un par de funciones  $\varphi: V \rightarrow \mathbb{R}$ ,  $i: V \times V \rightarrow \mathbb{R}$ . Por simplicidad escribiremos  $i_{uv}$  en lugar de  $i(u, v)$ . Decimos que  $\varphi$  satisface la **ley de voltajes de Kirchhoff** si para cualquier camino cerrado  $v_1, \dots, v_k, v_{k+1} = v_1$  en  $\Gamma$  se cumple

$$\sum_{j=1}^k \varphi(v_j) - \varphi(v_{j+1}) = 0. \quad (4.1)$$

Por otro lado,  $i$  satisface la **ley de corrientes de Kirchhoff** si la corriente total en cualquier vértice  $u \in V \setminus \{s, t\}$  es cero, es decir,

$$\sum_{v \in V} i_{uv} = 0. \quad (4.2)$$

Por último, la relación entre las resistencias,  $\varphi$  e  $i$  está dada por la ley de Ohm. Diremos que las funciones  $\varphi$  e  $i$  satisfacen la **ley de Ohm** si para cualquier par de vértices  $(u, v) \in V \times V$ ,

$$i_{uv} = w_{uv}(\varphi(v) - \varphi(u)). \quad (4.3)$$

Dado que el valor de la corriente está en función de la conductancia, si no hay conductancia entre dos vértices dados, necesariamente  $i_{uv} = 0$  si  $u$  y  $v$  no son vecinos, es decir,  $i_{uv} = 0$  cuando  $\{u, v\} \notin E$ .

Una vez que hayamos especificado un par de vértices distintos  $s, t \in V$  y dos funciones  $\varphi$  e  $i$  que satisfagan las leyes de Kirchhoff y la ley de Ohm, pensaremos en la tupla ordenada  $(\Gamma, w, s, t, \varphi, i)$  como una **red eléctrica**. Siempre que no exista el riesgo de confusión, pensaremos una red eléctrica como una gráfica con pesos con estructura adicional, con el fin de evitar nombrar a la tupla explícitamente. En caso de tener una red eléctrica, al vértice  $s$  le llamaremos la **fuerza** y a  $t$  el **sumidero** de la red. Por otro lado, llamaremos a  $\varphi$  el **voltaje** y a  $i$  la **corriente** de la red eléctrica.

Veamos que el voltaje de una red eléctrica es una función armónica en el conjunto  $V \setminus \{s, t\}$ . Sea  $u \in V \setminus \{s, t\}$ , al tomar la suma con respecto a  $v$  sobre todo el conjunto

de vértices en (4.3) obtenemos

$$\sum_{v \in V} i_{uv} = \sum_{v \in V} w_{uv}(\varphi(v) - \varphi(u)).$$

De acuerdo con la ley de corrientes de Kirchhoff (4.2), el lado izquierdo es cero y entonces

$$W_u \varphi(u) = \left( \sum_{v \in V} w_{uv} \right) \varphi(u) = \sum_{v \in V} w_{uv} \varphi(v).$$

**4.2.1 Teorema.** *El voltaje de cualquier red eléctrica es una función armónica en el conjunto de vértices, exceptuando la fuente y el sumidero.*

También es válido definir un voltaje en una gráfica con pesos como cualquier función armónica en el conjunto de vértices, salvo un par de vértices distintos. Con esto, podemos definir su corriente asociada como aquella que proviene de la ley de Ohm [12, Cap. 1]. En este sentido, si  $\varphi$  es una función armónica en  $V \setminus \{s, t\}$ , definamos a  $i: E \rightarrow \mathbb{R}$  a través de la ley de Ohm, es decir, para cada arista  $\{u, v\} \in E$  sigamos a la ecuación (4.3). Veamos que estas dos funciones definen una red eléctrica en  $(\Gamma, w)$ . Dado que la ley de Ohm se satisface por definición, debemos probar únicamente las dos leyes de Kirchhoff. La ley de voltajes de Kirchhoff se satisface trivialmente para cualquier función en el conjunto de vértices. Por otro lado, sea  $u \in V \setminus \{s, t\}$ , entonces

$$\begin{aligned} \sum_{v \in V} i_{uv} &= \sum_{v \in V} w_{uv}(\varphi(v) - \varphi(u)) \\ &= \sum_{v \in V} w_{uv} \varphi(v) - \sum_{v \in V} w_{uv} \varphi(u) \\ &= W_u \varphi(u) - W_u \varphi(u) = 0 \end{aligned}$$

debido a que  $\varphi$  es armónica. Con esto,  $i$  satisface la ley de corrientes de Kirchhoff. Además, trivialmente  $i_{uv} = -i_{vu}$  para cualquiera  $u, v \in V$ .

### 4.3. Problema de Dirichlet discreto

De acuerdo con la discusión anterior, estudiar redes eléctricas en el sentido de un voltaje y una corriente que satisfagan las leyes de Kirchhoff y la ley de Ohm es

equivalente a estudiar funciones armónicas en  $V \setminus \{s, t\}$  y sus respectivas corrientes asociadas. De esta manera la teoría de redes eléctricas se puede generalizar a través de la teoría de funciones armónicas en gráficas con pesos. A continuación veremos que, a diferencia del problema de Dirichlet en el plano, el problema de Dirichlet en el contexto de nuestras redes eléctricas siempre tiene solución, y como esperaríamos, esta solución es única. Dada una gráfica  $(\Gamma, w)$  con dos vértices distinguidos  $s$  y  $t$  que jugarán el papel de la fuente y el sumidero, respectivamente, y un par de números reales  $\alpha, \beta$ , el problema de encontrar una función armónica  $\psi: V \rightarrow \mathbb{R}$  en  $V \setminus \{s, t\}$  tal que  $\psi(s) = \alpha$  y  $\psi(t) = \beta$  es llamado el **problema de Dirichlet** en  $(\Gamma, w)$  con fuente  $s$  y sumidero  $t$ . A este problema de Dirichlet, con el único objetivo de escribir menos, lo llamaremos **problema de Dirichlet discreto**.

En lo que sigue asumiremos que  $\psi: V \rightarrow \mathbb{R}$  es una función armónica en  $V \setminus \{s, t\}$ . Siguiendo la teoría del problema de Dirichlet en el plano, comenzamos por probar lo siguiente.

**4.3.1 Proposición.** *Una combinación lineal de funciones armónicas también es armónica.*

*Demostración.* Esto se verifica rápidamente: si  $\psi_1$  y  $\psi_2$  son armónicas, entonces

$$\begin{aligned} (\lambda_1\psi_1 + \lambda_2\psi_2)(v) &= \lambda_1 \sum_{u \in V} P_{uv}\psi_1(u) + \lambda_2 \sum_{u \in V} P_{uv}\psi_2(u) \\ &= \sum_{u \in V} P_{uv}(\lambda_1\psi_1 + \lambda_2\psi_2)(u) \end{aligned}$$

para cualquier par de números reales  $\lambda_1$  y  $\lambda_2$ . □

Ahora pasamos a considerar condiciones triviales en la frontera. El resultado análogo al problema en el plano, siguiendo nuestra teoría de redes eléctricas, también se cumple.

**4.3.2 Proposición.** *Si  $\psi$  es una función armónica que satisface  $\psi(s) = \psi(t) = 0$ , entonces  $\psi$  es idénticamente cero.*

La unicidad de la solución al problema de Dirichlet discreto se puede deducir fácilmente de los dos resultados anteriores. Si tenemos dos soluciones al problema, su

diferencia también lo será de acuerdo con la Proposición 4.3.1. Esta diferencia será cero en la fuente  $s$  y el sumidero  $t$ . Por lo tanto esta diferencia es idénticamente cero y ambas soluciones deben ser iguales según la Proposición 4.3.2.

**4.3.3 Teorema.** *Cualquiera dos soluciones al problema de Dirichlet discreto son necesariamente iguales.*

Finalmente, establecemos la existencia de una única solución al problema de Dirichlet discreto sin importar la forma de nuestra gráfica  $(\Gamma, w)$ .

**4.3.4 Teorema.** *Siempre existe una solución al problema de Dirichlet discreto y ésta es única.*

*Demostración.* Pongamos  $k = |V|$  y notemos que una función armónica  $\psi$  en  $V \setminus \{s, t\}$  con  $\psi(s) = \alpha$  y  $\psi(t) = \beta$  debe satisfacer el sistema de  $k - 2$  ecuaciones lineales en  $k - 2$  incógnitas dado por

$$\psi(v) = \sum_{u \in V} P_{uv} \psi(u)$$

para cada  $u, v \in V \setminus \{s, t\}$ . Este sistema de ecuaciones lineales posee una única solución, pues, de acuerdo con el Teorema 4.3.3, la matriz de coeficientes asociada al sistema es no singular y por lo tanto invertible.  $\square$

El problema de Dirichlet con fuente  $s$  y sumidero  $t$  que describimos arriba es un caso particular del problema de Dirichlet en  $(\Gamma, w)$  cuando el conjunto frontera consta de únicamente dos elementos; en este caso la frontera es el conjunto formado por la fuente y el sumidero. A continuación presentamos otra forma de expresar las leyes de Kirchhoff y la ley de Ohm con ayuda del lenguaje del álgebra lineal con el objetivo de generalizar nuestro problema de Dirichlet discreto. En particular, estamos interesados en extender la noción de frontera a conjuntos más grandes.

Consideremos una gráfica no dirigida con pesos  $(\Gamma, w)$  y orientemos cada arista de  $\Gamma$  de manera arbitraria reemplazando cada arista por una arista dirigida. La **matriz de incidencia signada** de  $\Gamma$ , respecto a esta orientación, es la matriz  $D$  indexada  $V$ -por- $E$  con entradas

$$D_{ve} = \begin{cases} 1 & \text{si para alguna } u \neq v, e \text{ parte de } u \text{ hacia } v \\ -1 & \text{si para alguna } u \neq v, e \text{ parte de } v \text{ hacia } u \\ 0 & \text{otro caso.} \end{cases} \quad (4.4)$$

Por otro lado, sea  $Y$  la matriz diagonal indexada por  $E$  con entradas  $Y_{ee} = w_e$ . A pesar de que necesitamos una orientación para reescribir la ley de Ohm y las leyes de Kirchhoff, veremos que el resultado final no depende de la orientación elegida.

Supongamos que  $\varphi$  e  $i$  son un voltaje y una corriente, respectivamente, que hacen de  $(\Gamma, w)$  una red eléctrica con fuente  $s$  y sumidero  $t$ . Sea  $\vec{\varphi} = (\varphi(v): v \in V)$  el vector indexado por  $V$  inducido por  $\varphi$  y análogamente sea  $\vec{i} = (i_e: e \in E)$  el vector indexado por  $E$  inducido por la corriente  $i$ . Observemos que

$$\begin{aligned} (YD^T)_{ev} &= \sum_{\ell \in E} Y_{e\ell} (D^T)_{\ell v} \\ &= Y_{ee} D_{ve} = \begin{cases} w_e & \\ -w_e & \\ 0 & \end{cases} \end{aligned}$$

de acuerdo con la definición (4.4) de  $D$ . Por lo tanto, las filas de  $YD^T$  consisten de pares 1 y  $-1$  de acuerdo con el inicio y término de las correspondientes aristas dirigidas. Entonces, dependiendo de la orientación elegida, tendremos

$$[(YD^T)\vec{\varphi}]_{u,v} = \pm w_{uv}(\varphi(u) - \varphi(v)).$$

De esta manera, la ley de Ohm se puede escribir como

$$\vec{i} = -(YD^T)\vec{\varphi}.$$

Por otro lado, es fácil verificar que la ley de corrientes de Kirchhoff se puede reescribir como  $D\vec{i} = \delta_s - \delta_t$ , donde  $\delta_v$  denota la distribución de Dirac en  $V$  centrada en  $v \in V$ , es decir,  $\delta_v(u) = 1$  siempre que  $u \neq v$  y  $\delta_v(v) = 1$ . Por lo tanto, una función  $\varphi: V \rightarrow \mathbb{R}$  es una solución a la ley de Ohm y las leyes de Kirchhoff si y sólo si

$$DYD^T\varphi = \delta_t - \delta_s.$$

Observemos que la matriz  $DYD^T$  es independiente de la orientación elegida en  $\Gamma$  pues

$$\begin{aligned} [D(YD^T)]_{uv} &= \sum_{e \in E} D_{ue} (YD^T)_{ev} \\ &= \sum_{e \in E} D_{ue} Y_{ee} D_{ve}, \end{aligned}$$

de manera que si  $e = \{u, v\}$  parte de  $u$  hacia  $v$ , entonces  $D_{ue} = -1$  y  $D_{ve} = 1$ . Análogamente, el signo del término  $D_{ue}Y_{ee}D_{ve}$  se preserva si  $e$  parte de  $v$  hacia  $u$ . A la matriz  $DYD^T$  la llamamos la **matriz Laplaciana** de  $(\Gamma, w)$ .

**4.3.5 Teorema.** *Sea  $(\Gamma, w)$  una gráfica con pesos y denotemos por  $L$  a su matriz Laplaciana. Sean  $U \subseteq V$  no vacío y  $f: U \rightarrow \mathbb{R}$ . Entonces existe un único par de funciones  $\varphi, h: V \rightarrow \mathbb{R}$  tales que  $\varphi = f$  en  $U$ ,  $h = 0$  en  $V \setminus U$ , y además  $L\varphi = h$ .*

Al problema resuelto por este teorema se le conoce como el problema de Dirichlet en  $(\Gamma, w)$  con **frontera**  $U$  y **condiciones de frontera** dadas por  $\{f(u): u \in U\}$ , donde el conjunto  $V \setminus U$  es llamado el **interior** de  $V$ . Con ayuda de la matriz Laplaciana podemos decir que  $\varphi$  es armónica en un vértice  $v \in V$  si  $(L\varphi)(v) = 0$ . Entonces el Teorema 4.3.5 establece que para cualquier red  $(\Gamma, w)$ , cualquier conjunto frontera  $U \subseteq V$  y cualesquiera condiciones de frontera, existirá una única función  $\varphi$  que satisfaga las condiciones de frontera, que además sea armónica en cada uno de los vértices interiores.

## 4.4. Principio de Rayleigh

El principio de Rayleigh establece que la resistencia efectiva de una red eléctrica (en el sentido de gráficas, claro) aumenta si las resistencias en los vértices aumentan, o dicho de otra forma, la resistencia efectiva aumenta si las conductancias en los vértices disminuyen, hecho que es inmediato en la teoría de circuitos eléctricos. Por ejemplo, si identificamos dos vértices de nuestra gráfica, es decir, si disminuimos la resistencia de una arista a cero, entonces la resistencia efectiva debe disminuir o quedarse igual. Este uso del principio de Rayleigh será particularmente importante en la prueba del teorema de Polya.

Por otro lado, una manera de entender la importancia de la corriente eléctrica en una red eléctrica dada es a través del principio de Thomson, el cual caracteriza a la corriente eléctrica de forma única al pensar en flujos y calculando lo que llamaremos la energía disipada de dicho flujo. Comencemos entonces definiendo qué es un flujo en este contexto.

Sean  $s, t \in V$  distintos, un  $(s, t)$ -**flujo**  $j$  es una función  $j: V \times V \rightarrow \mathbb{R}$  tal que

$$(I) \quad j_{uv} = -j_{vu},$$

$$(II) \quad j_{uv} = 0 \text{ cuando } \{u, v\} \notin V,$$

$$(III) \quad \text{si } u \neq s, t, \text{ entonces tendremos } \sum_{v \in V} j_{uv} = 0.$$

Siguiendo la terminología de redes eléctricas, los vértices  $s$  y  $t$  son llamados la fuente y el sumidero de un  $(s, t)$ -flujo y usualmente diremos flujo en lugar de  $(s, t)$ -flujo. Observemos que las condiciones (ii) y (iii) anteriores hacen referencia a la ley de Ohm y la ley de corrientes de Kirchhoff, respectivamente. De hecho, nuestra noción de corriente en el sentido de redes eléctricas es un caso particular de un  $(s, t)$ -flujo. Para tener la ley de Ohm asociamos un voltaje al flujo, al igual que asociábamos una corriente (o flujo) a un voltaje (o función armónica) dado: a través de la ecuación (4.3). Así, a partir de un flujo podemos conseguir una red eléctrica al especificar condiciones  $\varphi(s) = \alpha$  y  $\varphi(t) = \beta$  para nuestro voltaje.

Para un  $(s, t)$ -flujo, escribiremos

$$J_u = \sum_{v \in V} j_{uv}, \quad u \in V.$$

Notemos que para  $u \in V \setminus \{s, t\}$ ,  $J_u = 0$ , por lo que

$$J_s + J_t = \sum_{u \in V} J_u = \sum_{u, v \in V} j_{uv} = \frac{1}{2} \sum_{u, v \in V} (j_{uv} + j_{vu}) = 0.$$

Entonces  $J_s = -J_t$  y definimos el **tamaño** del flujo  $j$  como  $|j| = |J_s|$ . Si  $|J_s| = 1$ , decimos que  $j$  es un **flujo unitario**. Por lo general se toma  $J_s > 0$ , pensando en  $s$  como la fuente y en  $t$  como el sumidero.

Ahora introducimos el concepto de energía disipada de un  $(s, t)$ -flujo. Consideremos  $j$  un  $(s, t)$ -flujo y recordemos que estamos denotando por  $w_e$  a la conductancia de una arista  $e$ . En este contexto, la **energía disipada** en  $j$  se define por

$$E(j) = \sum_{e \in E} j_e^2 / w_e = \frac{1}{2} \sum_{u, v \in V} j_{uv}^2 / w_{uv}.$$

Con el siguiente resultado podremos introducir el concepto de resistencia efectiva de una red eléctrica.

**4.4.1 Proposición.** Sea  $\psi: V \rightarrow \mathbb{R}$  y sea  $j$  un  $(s, t)$ -flujo. Entonces

$$(\psi(t) - \psi(s))J_s = \frac{1}{2} \sum_{u,v \in V} (\psi(v) - \psi(u))j_{uv}.$$

*Demostración.* A partir de las propiedades de un flujo obtenemos

$$\begin{aligned} \sum_{u,v \in V} (\psi(v) - \psi(u))j_{uv} &= \sum_{u,v \in V} \psi(v)j_{uv} - \sum_{u,v \in V} \psi(u)j_{uv} \\ &= \sum_{v \in V} \psi(v) \sum_{u \in V} j_{uv} - \sum_{u \in V} \psi(u) \sum_{v \in V} j_{uv} \\ &= -\sum_{v \in V} \psi(v)J_v - \sum_{u \in V} \psi(u)J_u \\ &= -2(\psi(s)J_s + \psi(t)J_t) \\ &= 2(\psi(t) - \psi(s))J_s, \end{aligned}$$

como se quería. □

Supongamos que  $\varphi$  e  $i$  son un voltaje y una corriente en la gráfica  $(\Gamma, w)$ , respectivamente, con fuente y sumidero dados por  $s, t$ , respectivamente. Entonces, al aplicar la ley de Ohm y la proposición anterior obtenemos que

$$\begin{aligned} (\varphi(t) - \varphi(s))\mathcal{I}_s &= \frac{1}{2} \sum_{u,v \in V} (\varphi(v) - \varphi(u))i_{uv} \\ &= \frac{1}{2} \sum_{u,v \in V} w_{uv}(\varphi(v) - \varphi(u))^2 \\ &= \frac{1}{2} \sum_{u,v \in V} \frac{i_{uv}^2}{w_{uv}} \\ &= E(i). \end{aligned}$$

Definimos entonces a la **conductancia efectiva** como

$$W_{\text{eff}} = \frac{\mathcal{I}_s}{\varphi(t) - \varphi(s)},$$

de manera que  $W_{\text{eff}}$  está de acuerdo con la ley de Ohm, pues de haber una arista que conecte a la fuente y el sumidero, la ley de Ohm sería satisfecha. De manera análoga se define la **resistencia efectiva** como

$$R_{\text{eff}} = \frac{1}{W_{\text{eff}}}.$$



Suponiendo que  $W_{\text{eff}} \neq 0$ , se tiene que

$$E(i) = (\varphi(t) - \varphi(s))\mathcal{I}_s = (\varphi(t) - \varphi(s))^2 W_{\text{eff}}.$$

Despejemos a la conductancia efectiva y multipliquemos arriba y abajo por  $W_{\text{eff}}^2$  en el lado derecho para obtener

$$W_{\text{eff}} = \frac{E(i)W_{\text{eff}}^2}{\mathcal{I}_s^2},$$

de donde concluimos que

$$W_{\text{eff}} = \frac{\mathcal{I}_s^2}{E(i)} \quad \text{y} \quad R_{\text{eff}} = \frac{E(i)}{\mathcal{I}_s^2}. \quad (4.5)$$

De acuerdo con la discusión anterior, concluimos el siguiente lema.

**4.4.2 Lema.** *La resistencia efectiva  $R_{\text{eff}}$  de una red eléctrica entre los vértices  $s$  y  $t$  es igual a la energía disipada cuando un flujo unitario pasa por  $s$  y  $t$ . Es decir, si  $i$  es un  $(s, t)$ -flujo unitario que satisface las leyes de Kirchhoff, entonces  $R_{\text{eff}} = E(i)$ .*

En la teoría de circuitos eléctricos se pueden demostrar las leyes de resistencias en serie y en paralelo con ayuda de la ley de Ohm, las leyes de Kirchhoff y la leyes de conservación de la carga eléctrica y la energía. Las leyes de resistencias en serie y en paralelo básicamente nos muestran cómo manipular una red eléctrica, de manera sencilla, sin alterar su resistencia efectiva. Podemos probar estas leyes apoyándonos de la Proposición 4.4.1, más aún, esta proposición tiene una estrecha relación con la ley de la conservación de la energía de la física. Es por estas razones que es usual referirse al resultado de la Proposición 4.4.1 como la ley de la conservación de la energía para redes eléctricas.

Finalmente, enunciamos el principio de Thomson para después concluir esta sección con el esperado principio de Rayleigh.

**4.4.3 Teorema** (Principio de Thomson). *Sea  $(\Gamma, w)$  una gráfica con pesos y sean  $s, t \in V$  distintos. Entonces el  $(s, t)$ -flujo que satisface las leyes de Kirchhoff es el único flujo  $i$  que minimiza a la energía disipada, es decir,*

$$E(i) = \inf\{E(j) : j \text{ es un flujo unitario de } s \text{ a } t\}.$$

*Demostración.* Sea  $j$  un  $(s, t)$ -flujo unitario y pongamos  $k = j - i$ , donde  $i$  es el único flujo unitario que es solución a las leyes de Kirchhoff. Si  $r_{uv} = 1/w_{uv}$ , entonces

$$\begin{aligned} 2E(j) &= \sum_{u,v \in V} j_{uv}^2 r_e = \sum_{u,v \in V} (k_{uv} + i_{uv})^2 r_e \\ &= \sum_{u,v \in V} k_{uv}^2 r_e + \sum_{u,v \in V} i_{uv}^2 r_e + 2 \sum_{u,v \in V} i_{uv} k_{uv} r_e \end{aligned}$$

y luego

$$E(j) = E(k) + E(i) + \sum_{u,v \in V} i_{uv} k_{uv} r_e.$$

Denotemos por  $\varphi$  al voltaje asociado a  $i$ . A partir de la ley de Ohm y la Proposición 4.4.1,

$$\begin{aligned} \sum_{u,v \in V} i_{uv} k_{uv} r_e &= \sum_{u,v \in V} (\varphi(v) - \varphi(u)) k_{uv} \\ &= 2(\varphi(v) - \varphi(u)) K_s = 0, \end{aligned}$$

porque  $k$  es un flujo de tamaño cero, pues  $k = j - i$  y ambos flujos son unitarios. Por lo tanto  $E(j) \geq E(i)$  y la igualdad se cumple si y sólo si  $j = i$ .  $\square$

El principio de Thomson nos ayuda a entender cuál es el significado de tener una función de voltaje y una corriente eléctrica en una gráfica que además satisfagan las leyes de Kirchhoff y la ley de Ohm. El principio de Thomson nos dice que tal flujo debe ser aquel que posea la menor energía disipada posible. Finalmente, al considerar dos funciones de resistencias distintas en una red eléctrica y comparar sus flujos eléctricos respectivos, podemos utilizar el principio de Thomson para demostrar el principio de Rayleigh.

**4.4.4 Teorema** (Principio de Rayleigh). *La resistencia efectiva  $R_{\text{eff}}$  de una red eléctrica es una función no decreciente de las resistencias de las aristas  $r_e$  para  $e \in E$ .*

*Demostración.* Sean  $r$  y  $r'$  dos funciones de resistencias en una misma gráfica  $\Gamma$  tales que  $r_e \leq r'_e$  para cada  $e \in E$ . Supongamos que  $i$  e  $i'$  son los correspondientes flujos

unitarios que satisfacen las leyes de Kirchhoff. Entonces, de acuerdo con la ecuación (4.5) y el principio de Thomson,

$$R_{\text{eff}} = \frac{1}{2} \sum_{u,v \in V} i_{uv}^2 r_e \leq \frac{1}{2} \sum_{u,v \in V} (i'_{uv})^2 r_e.$$

Por otro lado, cada  $r_e \leq r'_e$  y tenemos que

$$\begin{aligned} R_{\text{eff}} &\leq \frac{1}{2} \sum_{u,v \in V} (i'_{uv})^2 r_e \\ &\leq \frac{1}{2} \sum_{u,v \in V} (i'_{uv})^2 r'_e \\ &= R'_{\text{eff}}. \end{aligned}$$

□

A pesar de que utilizamos el principio de Thomson para demostrar el principio de Rayleigh, éste último es válido aún cuando la corriente eléctrica de la red no es unitaria. Esto porque

$$R_{\text{eff}} = \frac{E(i)}{|i|^2} = E\left(\frac{i}{|i|}\right)$$

y podemos aplicar el resultado anterior a la red con corriente  $i/|i|$ .

## 4.5. Resistencia efectiva y recurrencia

En esta sección veremos que uno puede expresar la probabilidad de que una caminata en la “gráfica”  $\mathbb{Z}^d$  regrese a su estado inicial en términos de la resistencia efectiva de gráficas que resultan de “truncar” nuestra gráfica original. Para esto, cambiemos un poco de dirección y consideremos una gráfica *infinita* conexa tal que el grado de cada uno de sus vértices es finito, llamémosle  $\Gamma = (V, E)$  a esta gráfica. Sea  $w: E \rightarrow \mathbb{R}^+$  una función de conductancias y nos concentramos en el par  $(\Gamma, w)$ . Es importante notar que sólo hemos definido redes eléctricas en gráficas *finitas* y esto es debido a la naturaleza del Teorema 4.3.4; en cuanto la gráfica es infinita, ya no es posible garantizar la existencia de un voltaje y una corriente que satisfagan las leyes de Kirchhoff y la ley de Ohm. Pero gracias al hecho de que los vértices de  $\Gamma$  poseen

grado finito, aún podemos considerar caminatas aleatorias en  $(\Gamma, w)$  en el sentido de la sección 3.2.

El motivo de introducir gráficas infinitas es para poder estudiar caminatas aleatorias simples en la malla  $\mathbb{Z}^d$  y así acercarnos al teorema de recurrencia de Polya. Nuestra estrategia para examinar gráficas infinitas en el contexto de la teoría de redes eléctricas que hemos desarrollado hasta ahora será analizar subgráficas que resultan de truncar la gráfica original. En particular, esto nos ayudará a definir qué entenderemos por la resistencia efectiva de una gráfica infinita.

Sea  $0$  un vértice distinguido de  $\Gamma$ , que llamaremos el “origen” de  $\Gamma$ . La distancia entre dos vértices  $u, v$ , en el sentido de la teoría de gráficas, y denotada por  $d(u, v)$ , es el número de aristas en el camino más corto entre  $u$  y  $v$ . Definamos

$$\begin{aligned}\Lambda_n &= \{v \in V : d(0, v) \leq n\}, \\ \partial\Lambda_n &= \Lambda_n \setminus \Lambda_{n-1} = \{v \in V : d(0, v) = n\}.\end{aligned}$$

Para  $n \geq 1$  definimos a  $\Gamma_n = (V_n, E_n)$  como la gráfica conexa finita obtenida a partir de  $\Gamma$  al identificar todos los vértices de  $V \setminus \Lambda_{n-1}$ , y denotemos al “vértice identificado” como  $I_n$ . La función de conductancias  $w_n$  en  $E_n$  será simplemente la restricción de  $w$  a  $E_n \subseteq E$ . En la siguiente discusión trabajaremos con la red eléctrica en  $(\Gamma_n, w_n)$  que resulta de considerar a la única solución  $g_n$  del problema de Dirichlet discreto en  $(\Gamma_n, w_n)$  con condiciones de frontera dadas por  $g_n(0) = 0$  y  $g_n(I_n) = 1$ . De acuerdo con el Teorema 4.3.4 y la Proposición 4.1.1, la función  $g_n : V \rightarrow \mathbb{R}$  está dada por

$$g_n(v) = \mathbb{P}(Z_k = I_n \text{ para alguna } k < T_0 | Z_0 = v), \quad v \in \Lambda_n,$$

con  $g_n(0) = 0$ ,  $g_n(I_n) = 1$  y  $Z$  una caminata aleatoria en  $(\Gamma_n, w_n)$  respecto a la correspondiente cadena de Markov.

Denotemos por  $R_{\text{eff}}(n)$  a la resistencia efectiva de esta nueva red eléctrica en  $(\Gamma_n, w_n)$ . Dado que la gráfica  $\Gamma_n$  puede obtenerse a partir de  $\Gamma_{n+1}$  identificando los vértices de  $\partial\Lambda_n \cup \partial\Lambda_{n+1}$  (i.e. disminuyendo algunas resistencias a cero), entonces por el principio de Rayleigh,  $R_{\text{eff}}(n)$  es creciente para  $n \in \mathbb{N}$ . Es por esto que el límite,  $\lim R_{\text{eff}}(n)$ , existe cuando  $n \rightarrow \infty$ . Definimos la resistencia efectiva de la gráfica  $(\Gamma, w)$  como

$$R_{\text{eff}} = \lim_{n \rightarrow \infty} R_{\text{eff}}(n).$$

**4.5.1 Teorema.** *Sea  $Z$  una caminata aleatoria en la gráfica conexa infinita con pesos positivos en las aristas  $(\Gamma, w)$  y con inicio determinista en el origen. La probabilidad de retorno de la caminata  $Z$  a 0 está dada por*

$$\mathbb{P}(Z_n = 0 \text{ para algún } n \geq 1 | Z_0 = 0) = 1 - \frac{1}{W_0 R_{\text{eff}}},$$

donde  $W_0 = \sum_{v \sim 0} w_{0v}$ .

*Demostración.* Consideremos a la gráfica  $(\Gamma_n, w_n)$  como una red eléctrica con fuente 0, sumidero  $I_n$  y voltaje  $g_n$ . Nuestra corriente será la inducida por  $g_n$  a través de la ley de Ohm, llamémosla  $i(n)$ . De manera que, en conjunto con la ley de Ohm,

$$\begin{aligned} \mathbb{P}(Z_k = 0 \text{ para alguna } k < T_{\partial\Lambda_n} | Z_0 = 0) &= 1 - \sum_{v \sim 0} P_{0v} \mathbb{P}(Z_k \in \partial\Delta_n \\ &\quad \text{para alguna } k < T_0 | Z_0 = v) \\ &= 1 - \sum_{v \sim 0} \frac{w_{0v}}{W_0} g_n(v) \\ &= 1 - \sum_{v \sim 0} \frac{w_{0v}}{W_0} (g_n(v) - g_n(0)) \\ &= 1 - \frac{|i(n)|}{W_0}, \end{aligned}$$

donde  $|i(n)|$  es el tamaño del flujo, es decir,  $|i(n)| = \mathcal{I}(n)_s$ . De acuerdo con la definición de resistencia efectiva,

$$\mathcal{I}(n)_s = \frac{1}{R_{\text{eff}}(n)} (g_n(I_n) - g_n(0)) = \frac{1}{R_{\text{eff}}(n)}.$$

El teorema queda probado una vez que verificamos que

$$\mathbb{P}(Z_k \in \partial\Lambda_n \text{ para alguna } k < T_0 | Z_0 = v) \rightarrow \mathbb{P}(Z_n = 0 \text{ para alguna } n \geq 1)$$

cuando  $n \rightarrow \infty$ , lo cual sí se tiene gracias a la continuidad de medidas de probabilidad.  $\square$

Del Teorema 4.5.1 concluimos rápidamente que si la resistencia efectiva de la gráfica infinita  $(\Gamma, w)$  es infinita, entonces cualquier caminata aleatoria en  $(\Gamma, w)$  es recurrente.

Notemos que podemos definir un flujo  $j$  en nuestra gráfica infinita  $(\Gamma, w)$ , pensando en el origen como una fuente sin necesidad de exhibir un sumidero, simplemente pidiendo que  $j$  satisfaga la ley de corrientes de Kirchhoff, es decir,

$$\sum_{v \in V} j_{uv} = 0 \text{ si } u \neq 0,$$

que también satisfaga  $j_{uv} = -j_{vu}$  para cada par de vértices  $u, v$  de  $\Gamma$  y que  $j_{uv} = 0$  cuando  $\{u, v\}$  no sea una arista de  $\Gamma$ . En tal caso, llamaremos a  $j$  un  $(0, \infty)$ -flujo. A continuación veremos cómo expresar la condición de transitoriedad en términos de flujos y energía disipada.

**4.5.2 Corolario.** *En las condiciones del teorema anterior:*

- I. *La caminata  $Z$  es recurrente si y sólo si  $R_{\text{eff}} = \infty$ .*
- II. *La caminata  $Z$  es transitoria si y sólo si existe un  $(0, \infty)$ -flujo  $j$  en  $\Gamma$  cuya energía disipada  $E(j) = \sum_{e \in E} j_e^2/w_e$  satisfaga  $E(j) < \infty$ .*

*Demostración.* La afirmación (I) es consecuencia inmediata del Teorema 4.5.1. Pasemos a demostrar la afirmación restante. Veamos que si la cadena es recurrente entonces la energía disipada de cualquier  $(0, \infty)$ -flujo no nulo es infinita. Sea  $i(n)$  un flujo unitario en  $\Gamma_n$  que satisfaga las leyes de Kirchhoff. Del Lema 4.4.2 se sigue que  $E(i(n)) = R_{\text{eff}}(n)$ . Ahora sea  $i$  un  $(0, \infty)$ -flujo no nulo unitario (de no ser así simplemente lo normalizamos). Notemos que al restringir  $i$  a  $E_n$ , el conjunto de aristas de  $\Gamma_n$ ,  $i$  es un flujo unitario de 0 a  $I_n$ . Al aplicar el Teorema 4.4.3 (el principio de Thomson) obtenemos que

$$E(i(n)) \leq \sum_{e \in E_n} i_e^2/w_e = \text{energía de } i \text{ en } \Gamma_n \leq E(i),$$

donde  $E(i(n))$  denota la energía disipada del flujo  $i(n)$  en la gráfica  $\Gamma_n$ . Entonces

$$\lim_{n \rightarrow \infty} E(i(n)) = R_{\text{eff}} \leq E(i).$$

Ahora es claro que si  $R_{\text{eff}} = \infty$ , entonces  $E(i) = \infty$ .

Supongamos que la cadena es transitoria. Entonces del inciso (I) se concluye que  $R_{\text{eff}} < \infty$ . Continuemos con la notación anterior. Dado que  $E(i(n)) = R_{\text{eff}}(n)$  y

$\lim R_{\text{eff}}(n)$  es finito, existe una sucesión estrictamente creciente de enteros positivos  $(n_k)_k$  tal que  $i(n_k)$  converge, digamos, al límite  $i$ . Es claro que  $i$  es un  $(0, \infty)$ -flujo unitario pues cada uno de los flujos  $i(n_k)$  es unitario. Entonces

$$\begin{aligned} E(i(n_k)) &= \sum_{e \in E_{n_k}} i(n_k)_e^2 / w_e \\ &\geq \sum_{e \in E_m} i(n_k)_e^2 / w_e \end{aligned}$$

para cada  $m \leq n_k$  y como  $\lim_{k \rightarrow \infty} i(n_k) = i$  obtenemos

$$\lim_{k \rightarrow \infty} \sum_{e \in E_m} i(n_k)_e^2 / w_e = \sum_{e \in E_m} i_e^2 / w_e.$$

Por otro lado,

$$\lim_{m \rightarrow \infty} \sum_{e \in E_m} i_e^2 / w_e = E(i).$$

Por lo tanto, al tomar primero el límite con respecto a  $k$  y después con respecto a  $m$ , obtenemos

$$E(i) \leq \lim_{k \rightarrow \infty} E(i(n_k)) = \lim_{k \rightarrow \infty} R_{\text{eff}}(n_k) = R_{\text{eff}} < \infty.$$

Así,  $i$  es el flujo que buscábamos y la afirmación queda demostrada.  $\square$

## 4.6. Demostración del Teorema de Polya

Ahora que tenemos clara la relación entre la resistencia efectiva y la recurrencia de caminatas aleatorias en una gráfica con vértices de grado finito, podemos pasar a la demostración del teorema de Polya, el cual establece que toda caminata aleatoria en la malla  $\mathbb{Z}^d$  es recurrente para dimensiones 1 y 2, mientras que será transitoria para dimensiones mayores. Para dimensiones 1 y 2 calcularemos explícitamente la resistencia efectiva de nuestra malla  $\mathbb{Z}^d$ . En dimensiones mayores construiremos un  $(0, \infty)$ -flujo con energía disipada finita. En ambos casos seguiremos el Corolario 4.5.2.

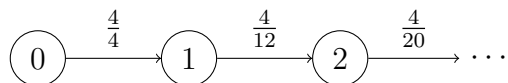
De acuerdo con el Corolario 4.5.2, para probar que una caminata en  $\mathbb{Z}^d$  para  $d = 1, 2$  es recurrente, debemos verificar que la resistencia efectiva es infinita. El caso unidimensional es inmediato ya que una caminata aleatoria simple en  $\mathbb{Z}$  se representa

por una gráfica lineal con peso  $1/2$  en cada una de sus aristas. Es fácil ver que la resistencia efectiva debe ser infinita después de aplicar la ley de resistencias en serie.

Para el caso  $d = 2$ , lo que haremos es reducir nuestro problema a analizar una red eléctrica de una sólo dimensión. Lograremos esto acotando nuestra resistencia efectiva por abajo por la resistencia efectiva de la red eléctrica que resulta de identificar los conjuntos  $\partial\Lambda_k = \{v \in \mathbb{Z}^2 : d(0, v) = k\}$ . Esto es posible porque de acuerdo con el principio de Rayleigh la identificación de dos vértices en una red eléctrica es equivalente a disminuir una arista (o resistor) a cero y entonces la resistencia efectiva de la nueva red (que resulta de identificar dos vértices) debe disminuir. De esta manera,  $\mathbb{Z}^2$  se transforma en la siguiente gráfica con pesos positivos.



Es fácil verificar que el vértice  $n - 1$  se conecta con el vértice  $n$  a través de  $8n - 4$  aristas, cada una de ellas con un mismo peso. La gráfica anterior tiene la misma resistencia efectiva que la siguiente gráfica lineal después de utilizar la ley de series en paralelo.



Entonces, de acuerdo con el principio de Rayleigh, se tiene que

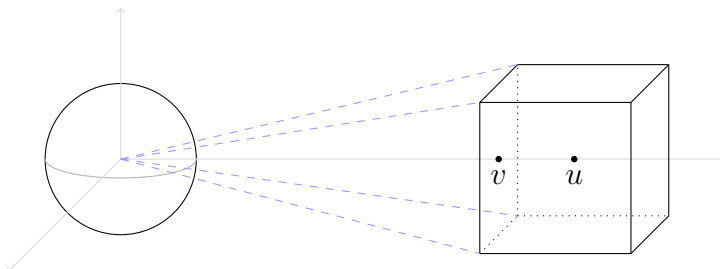
$$R_{\text{eff}} \geq \sum_{n=1}^{\infty} \frac{4}{8n - 4} = \infty$$

y por lo tanto toda caminata aleatoria simple en  $\mathbb{Z}^2$  debe ser recurrente.

Para dimensiones mayores, basta fijarnos en la malla  $\mathbb{Z}^3$ . De acuerdo con el Corolario 4.5.2, es suficiente construir un  $(0, \infty)$ -flujo no nulo con energía disipada finita. El valor absoluto de este flujo  $j$  evaluado en una pareja ordenada de vértices  $(u, v)$  que son vecinos en nuestra malla será el área que resulta de proyectar en la esfera unitaria de  $\mathbb{R}^3$  la cara del cubo de lado 2 centrado en  $u$  (de manera que  $v$  está en alguno de los lados de este cubo) en la que  $v$  se encuentra, donde los lados de este cubo son paralelos a los ejes coordenados. Por ejemplo, en la siguiente figura se



muestra cual sería la cara que debemos proyectar en la esfera unitaria para después calcular su respectiva área.



El signo de  $j_{uv}$  será el mismo signo del producto interno  $\langle u + v, u - v \rangle$ . Cuando  $u$  y  $v$  no sean vecinos, pondremos  $j_{uv} = 0$ . Entonces, cuando  $u$  y  $v$  son vecinos,  $u - v$  es un vector con entradas todas cero excepto por una de ellas que será igual a 1 ó  $-1$ , de esta manera, el producto interno  $\langle u + v, u - v \rangle$  tiene la forma  $\mp(2u_\ell \pm 1)$  para algún subíndice  $\ell = 1, 2, 3$ . Así pues, si por ejemplo, cada  $u_i \geq 1$ , entonces  $u_\ell \geq 1$  y luego  $2u_\ell \pm 1 \geq 1$ , con lo que concluimos que el signo de  $\langle u + v, u - v \rangle$  cambia de acuerdo a las caras del cubo, en este caso en particular el cambio es por caras opuestas: a caras opuestas corresponderán signos opuestos.

Por otro lado, necesitamos verificar que  $j$  es en realidad un flujo. Rápidamente notamos que  $j$  satisface dos de las tres propiedades de un  $(0, \infty)$ -flujo. Más precisamente,  $j_{uv} = -j_{vu}$  y  $j_{uv} = 0$  cuando  $u$  y  $v$  no son vecinos. Sólo nos resta comprobar que cuando  $u \neq 0$ ,

$$J_u = \sum_{v \in \mathbb{Z}^3} j_{uv} = 0.$$

Esto no es muy difícil de ver ya que de acuerdo a nuestra observación anterior, las caras proyectadas en la esfera se cancelarán a pares cuando el cubo se encuentre en el primer octante (donde cada una de las coordenadas son positivas). En cualquier caso se verifica que la suma de flujos al rededor de un punto resulta cero.

Finalmente, este flujo no es nulo, pues al hacer los cálculos se tiene que  $J_0 \neq 0$ . Ahora procedemos a estimar la energía disipada de este flujo. A partir de consideraciones geométricas, es fácil demostrar que existe una constante  $A > 0$  tal que  $|j_{uv}| \leq A/d(0, u)^2$  para toda  $u \in \mathbb{Z}^3$  y de igual manera que existe una constante  $B > 0$  tal que la cantidad de vértices  $v$  con  $d(0, v) = n$  es menor o igual que  $Bn^2$ ,

recordando que  $d(u, v)$  denota la cantidad de aristas en el recorrido más corto de  $u$  a  $v$  en  $\mathbb{Z}^3$ . Entonces

$$E(j) = \sum_{u \neq 0} \sum_{v \sim u} 6j_{uv}^2,$$

pues cada  $w_{uv} = 1/6$  y luego

$$\begin{aligned} E(j) &\leq \sum_{d(0,u) \geq 1} \sum_{v \sim u} 6 \left( \frac{A}{d(0,u)^2} \right)^2 \\ &= \sum_{n=1}^{\infty} \sum_{v \sim u} 6n^2 \left( \frac{A}{n^2} \right)^2 \\ &\leq \sum_{n=1}^{\infty} 6Bn^2 \left( \frac{A}{n^2} \right)^2 < \infty. \end{aligned}$$

Es decir,  $j$  es un flujo no nulo con energía disipada finita. Del Corolario 4.5.2 se sigue que toda caminata aleatoria en  $\mathbb{Z}^3$  es transitoria. Observemos que este mismo argumento funciona para construir un flujo no nulo en  $\mathbb{Z}^d$  con energía disipada finita cuando  $d \geq 4$ . Esto termina la demostración del teorema de recurrencia de Polya.

## CAPÍTULO 5

# Teorema de Polya vía funciones especiales

Esta demostración consiste en utilizar funciones generadoras exponenciales, la función de Bessel modificada del primer tipo, la transformada de Borel y el principio de Laplace. La idea principal de este enfoque es reducir el problema al análisis asintótico de una integral en infinito. Lograremos esto introduciendo las funciones generadoras, ordinarias y exponenciales, que resultan de contar bucles arbitrarios y bucles irreducibles, pensando en un bucle como una caminata que empieza y termina en el origen, llamándole irreducible si no es posible expresarla como concatenación de otros bucles. A grandes rasgos, hacer esto transforma nuestro problema al de calcular un límite particular. El siguiente paso será expresar el límite de esta función como la transformada de Borel de la función de Bessel modificada de orden cero. Con esto, podemos utilizar la expresión integral de la función de Bessel modificada para estimar dicho límite con ayuda del principio de Laplace. Como último comentario, el principio de Laplace reduce el teorema de recurrencia de Polya a la siguiente afirmación: Si  $d$  es tal que la integral

$$\int_1^{\infty} t^{-d/2} dt$$

diverge, entonces una caminata aleatoria simple en  $\mathbb{Z}^d$  será recurrente, y será transitoria en otro caso.

## 5.1. Funciones generadoras exponenciales

A cada sucesión  $(a_n)$  le podemos asociar una función generadora. Existen distintos tipos de ellas, pero consideraremos funciones generadoras ordinarias y exponenciales. Las funciones generadoras son una herramienta natural y poderosa en combinatoria, su importancia radica en la cantidad de información que podemos concluir acerca de la sucesión asociada. Por ejemplo, si la sucesión  $(a_n)$  está descrita de manera recursiva, en algunos casos es posible expresarla de manera explícita con la ayuda de su función generadora ordinaria o exponencial, dependiendo del crecimiento de los  $a_n$ .

Dada una sucesión de números reales  $(a_n)$ , la serie de potencias formales  $f(x) = \sum_{n \geq 0} a_n x^n$  es llamada la **función generadora ordinaria** o simplemente la función generadora de la sucesión  $(a_n)$ . Entre otras utilidades de la función generadora, está la de contar la cantidad total de formas posibles de definir un tipo de estructura en un conjunto finito [4, Cap. 8].

**5.1.1 Teorema.** *Dada dos estructuras combinatorias  $A$  y  $B$ , sean  $a_n$  y  $b_n$  el número total de formas posibles de dotar a un conjunto de  $n$  elementos con las estructuras  $A$  y  $B$ , respectivamente. Por otro lado, sea  $c_n$  el número de formas de partir al conjunto  $\{1, \dots, n\}$  en dos intervalos  $S = \{1, \dots, i\}$ ,  $T = \{i+1, \dots, n\}$ , posiblemente vacíos, y dotarlos de las estructuras  $A$  a  $S$  y  $B$  a  $T$ . Si  $A(x)$ ,  $B(x)$  y  $C(x)$  denotan las funciones generadoras ordinarias de las sucesiones  $(a_n)$ ,  $(b_n)$  y  $(c_n)$ , respectivamente, entonces*

$$A(x)B(x) = C(x).$$

Es importante notar que, de acuerdo con el teorema anterior, el producto de dos funciones generadoras ordinarias nos ayuda a contar el número de maneras distintas de expresar un conjunto de  $n$  elementos como unión de dos *intervalos* disjuntos, donde uno de ellos está dotado de una estructura y el otro con la estructura restante. Al cambiar funciones generadoras ordinarias en el teorema anterior por funciones generadoras exponenciales, estaremos contando la manera de expresar a un conjunto de  $n$  elementos como unión de dos *conjuntos* disjuntos, donde cada uno de ellos está dotado de una de las dos estructuras. Antes de enunciar este teorema, pasemos

a definir la función generadora exponencial de una sucesión. Para una sucesión de números reales  $(a_n)$ , la serie de potencias formales

$$f(x) = \sum_{n=0}^{\infty} \frac{a_n}{n!} x^n$$

es llamada su **función generadora exponencial**.

**5.1.2 Teorema.** *Sean  $A$  y  $B$  dos estructuras combinatorias. Sean  $a_n$  y  $b_n$  el número total de formas posibles de dotar a un conjunto de  $n$  elementos con las estructuras  $A$  y  $B$ , respectivamente. Por otro lado, sea  $c_n$  el número de formas de expresar al conjunto  $\{1, \dots, n\}$  como la unión de dos conjuntos  $S, T$  disjuntos y dotarlos de las estructuras  $A$  a  $S$  y  $B$  a  $T$ . Si  $A(x)$ ,  $B(x)$  y  $C(x)$  denotan las funciones generadoras exponenciales de las sucesiones  $(a_n)$ ,  $(b_n)$  y  $(c_n)$ , respectivamente, entonces*

$$A(x)B(x) = C(x).$$

Este teorema, al igual que el anterior, se demuestra de manera sencilla: calculando la multiplicación de las series en cuestión y justificando la conclusión del teorema de manera puramente combinatoria. Por ejemplo, por definición

$$A(x) = \sum_{k=0}^{\infty} \frac{a_k}{k!} x^k, \quad B(x) = \sum_{k=0}^{\infty} \frac{b_k}{k!} x^k$$

luego,

$$(AB)(x) = \sum_{k=0}^{\infty} \frac{c_k}{k!} x^k, \quad \text{donde} \quad c_n = \sum_{k=0}^n \binom{n}{k} a_k b_{n-k}, \quad (5.1)$$

pues

$$\begin{aligned} \frac{c_n}{n!} &= \sum_{k=0}^n \frac{a_k}{k!} \frac{b_{n-k}}{(n-k)!} \\ &= \left( \sum_{k=0}^n \binom{n}{k} a_k b_{n-k} \right) \frac{1}{n!}. \end{aligned}$$

Que  $C(x) = (AB)(x)$  se sigue fácilmente.

Para terminar esta sección, discutiremos la transformada integral de Borel. Nuestro objetivo es mostrar que si  $f$  es la función generadora exponencial a una sucesión

$(a_n)$ , entonces la transformada de Borel de  $f$  es igual a la función generadora ordinaria de la misma sucesión.

La transformada integral

$$(\mathcal{B}f)(z) = \int_0^{\infty} f(tz)e^{-t} dt,$$

es llamada la **transformada de Borel**  $f$ . Si, por ejemplo,  $f$  es la función generadora exponencial de  $(a_n)$ , observemos que al escribir la serie de Maclaurin de  $f(tz)$  e intercambiar los símbolos de suma e integración, obtenemos

$$\begin{aligned} (\mathcal{B}f)(z) &= \int_0^{\infty} f(tz)e^{-t} dt \\ &= \int_0^{\infty} \left( \sum_{n=0}^{\infty} \frac{a_n}{n!} t^n z^n e^{-t} \right) dt \\ &= \sum_{n=0}^{\infty} \frac{a_n}{n!} z^n \left( \int_0^{\infty} t^n e^{-t} dt \right) \\ &= \sum_{n=0}^{\infty} a_n z^n, \end{aligned}$$

donde hemos utilizado el hecho de que

$$\int_0^{\infty} t^n e^{-t} dt = n!,$$

con lo cual hemos regresado a la función generadora ordinaria de la sucesión, como se quería.

## 5.2. Funciones de Bessel modificadas

La **ecuación de Bessel** de orden  $\alpha$  está dada por

$$\frac{d^2 y}{dx^2} + \frac{1}{x} \frac{dy}{dx} + \left(1 - \frac{\alpha^2}{x^2}\right) y = 0.$$

Cuando  $\operatorname{Re} \alpha > -1/2$  se puede probar que la solución  $J_\alpha$  a la ecuación de Bessel de orden  $\alpha$  posee la representación en serie de potencias

$$J_\alpha(x) = \sum_{k=0}^{\infty} \frac{(-1)^k (x/2)^{2k+\alpha}}{k! \Gamma(k + \alpha + 1)}$$

y la representación integral

$$J_\alpha(x) = \frac{1}{\sqrt{\pi}\Gamma(\alpha + 1/2)}(x/2)^\alpha \int_0^\pi e^{ix \cos \theta} (\sin \theta)^{2\alpha} d\theta.$$

La ecuación diferencial que hemos introducido,

$$\frac{d^2y}{dx^2} + \frac{1}{x} \frac{dy}{dx} - \left(1 + \frac{\alpha^2}{x^2}\right)y = 0,$$

con  $x$  real, surge de manera frecuente en la física-matemática. Es posible verificar que, para  $x$  real, la correspondencia  $x \mapsto e^{-\alpha\pi i/2} J_\alpha(xe^{\pi i/2})$  es una función con valores reales [1, Cap. 4]. Definimos entonces la **función de Bessel modificada** del primer tipo, de orden  $\alpha$ , como

$$I_\alpha(x) = e^{-\alpha\pi i/2} J_\alpha(xe^{\pi i/2}).$$

Por otro lado, con ayuda de las expresiones anteriores, se prueba que  $I_\alpha$  posee la siguiente representación en serie de potencias

$$I_\alpha(x) = \sum_{k=0}^{\infty} \frac{(x/2)^{2k+\alpha}}{k!\Gamma(k + \alpha + 1)}. \quad (5.2)$$

También es posible expresar a  $I_\alpha$  de manera integral como:

$$I_\alpha(x) = \frac{(x/2)^\alpha}{\sqrt{\pi}\Gamma(\alpha + 1/2)} \int_0^\pi e^{x \cos \theta} (\sin \theta)^{2\alpha} d\theta. \quad (5.3)$$

Estas expresiones, en conjunto con el Principio de Laplace (ver abajo), nos servirán para pasar de la representación en serie de potencias (5.2) a la expresión integral en (5.3) y estimar el comportamiento en infinito de la función de Bessel modificada, en particular, estamos interesados en la función de Bessel modificada de orden cero en la prueba del teorema de recurrencia de Polya.

### 5.3. Principio de Laplace

En la prueba del teorema de recurrencia de Polya, nos concentraremos en analizar la función de Bessel modificada de orden cero introducida en la sección anterior. Esta función es un caso particular del principio de Laplace. Dicho principio nos brinda una estimación en infinito para integrales de la forma

$$I(t) = \int_a^b e^{-t\varphi(x)} f(x) dx, \quad t \in (0, \infty),$$

donde  $\varphi$  y  $f$  son funciones suaves en el intervalo  $[a, b]$ , y además  $\varphi$  posee un mínimo global en  $(a, b)$ . Más específicamente [3, Cap. 5]:

**5.3.1 Teorema** (Principio de Laplace). *Supongamos que  $a < b$  y  $\varphi, f: [a, b] \rightarrow \mathbb{R}$  son suaves en  $[a, b]$ . Si además  $\varphi$  posee un mínimo global en  $x_0 \in (a, b)$ , entonces la función  $I$  definida arriba satisface*

$$I(t) = \sqrt{\frac{2\pi}{t\varphi''(x_0)}} f(x_0) e^{-t\varphi(x_0)} + O\left(\frac{e^{-t\varphi(x_0)}}{t^{3/2}}\right)$$

cuando  $t \rightarrow \infty$ . En particular, concluimos que

$$I(t) \sim \text{cte} \cdot t^{-1/2} e^{-t\varphi(x_0)}.$$

Podemos darnos una idea de por qué este resultado es cierto, definamos la integral  $I_0$  como:

$$I_0(t) = e^{-t\varphi(x_0)} \int_{x_0-\varepsilon}^{x_0+\varepsilon} e^{-t(\varphi(x)-\varphi(x_0))} dx,$$

donde  $\varepsilon > 0$  es pequeño y fijo. Supongamos que se tiene

$$\lim_{t \rightarrow \infty} \frac{I_0(t)}{I(t)} = 1. \quad (5.4)$$

Dado que  $\varepsilon > 0$  es pequeño, tiene sentido aproximar a  $\varphi(x) - \varphi(x_0)$  y a  $f$  en  $[x_0 - \varepsilon, x_0 + \varepsilon]$  por el primer término no nulo de su serie de Taylor alrededor de  $t = x_0$  (por simplicidad asumiremos que  $f(x_0) \neq 0$ ). Entonces

$$e^{t\varphi(x_0)} I_0(t) \approx \int_{x_0-\varepsilon}^{x_0+\varepsilon} f(x_0) e^{-\frac{t}{2}\varphi''(x_0)(x-x_0)^2} dx.$$

De manera que, mientras más pequeño sea  $\varepsilon$ , mejor aproximará el lado derecho al lado izquierdo de la ecuación anterior. Como se cumple (5.4), tenemos que para  $t$  suficientemente grande,

$$I(t) \approx e^{-t\varphi(x_0)} f(x_0) \int_{x_0-\varepsilon}^{x_0+\varepsilon} e^{-\frac{t}{2}\varphi''(x_0)(x-x_0)^2} dx.$$

Finalmente, hagamos el cambio de variable

$$u = \sqrt{\frac{t}{2}\varphi''(x_0)}(x - x_0)$$



para obtener

$$I(t) \approx e^{-t\varphi(x_0)} f(x_0) \sqrt{\frac{2}{t\varphi''(x_0)}} \int_{-c_0\varepsilon}^{c_0\varepsilon} e^{-u^2} du,$$

donde

$$c_0 = \sqrt{\frac{t}{2}\varphi''(x_0)}.$$

Notemos que no importa que tan pequeño  $\varepsilon$  sea, podemos tomar  $t$  tan grande como sea necesario, de manera que el valor de la integral sólo se alterará un poco si cambiamos los límites de integración de  $-\infty$  a  $\infty$ :

$$\begin{aligned} I(t) &\approx e^{-t\varphi(x_0)} f(x_0) \sqrt{\frac{2}{t\varphi''(x_0)}} \int_{-\infty}^{\infty} e^{-u^2} du \\ &= e^{-t\varphi(x_0)} f(x_0) \sqrt{\frac{2\pi}{t\varphi''(x_0)}}. \end{aligned}$$

## 5.4. Demostración del Teorema de Polya

Con esta herramienta a nuestra disposición, ahora procedemos a demostrar el teorema de recurrencia de Polya. Recordemos que nuestro objetivo es demostrar que la probabilidad  $p$ , de que una caminata (aleatoria simple) en la malla  $\mathbb{Z}^d$  que comienza en el origen visitará nuevamente el origen, es igual a 1 cuando  $d = 1, 2$  y que  $p < 1$  cuando  $d \geq 3$ . Para esto, denotemos por  $E$  al evento de que una caminata en  $\mathbb{Z}^d$  que comienza en el origen regresa a su posición inicial, de manera que  $p = \mathbb{P}(E)$ . Lo que queremos es exhibir la dependencia de  $E$  respecto a la dimensión  $d$  de la malla  $\mathbb{Z}^d$  y probar que sólo cuando  $d = 1, 2$  se tiene que  $p = 1$ . Comencemos por expresar a  $E$  como una unión disjunta de eventos más sencillos. Para  $n \geq 1$ , consideremos al evento  $E_n$  que consiste de todas las caminatas que regresan a su posición inicial por primera vez en el paso  $n$ . Convenimos en poner  $E_0 = \emptyset$ , que corresponde al hecho de que la posición inicial de la caminata aleatoria no cuenta como retorno. Entonces los eventos  $E_n$  son mutuamente disjuntos para distintos valores de  $n$  y

$$E = \bigsqcup_{n \geq 0} E_n,$$

de manera que

$$p = \sum_{n \geq 0} p_n,$$

donde  $p_n = \mathbb{P}(E_n)$ , y en particular  $p_0 = 0$ .

Notemos que la función generadora ordinaria  $P(x) = \sum_0^\infty p_n x^n$ , satisface  $P(1) = p$ . Lo que haremos es expresar a  $P$  en términos de la función generadora ordinaria de la sucesión de probabilidades  $(q_n)$ , donde  $q_n$  denota la probabilidad de que una caminata aleatoria regrese a su posición inicial en el paso  $n$ . Lograremos obtener esta relación al contar bucles en la malla  $\mathbb{Z}^d$ , pensando en un **bucle** como una caminata que inicia y termina en el origen. En caso de que un bucle tenga longitud cero, diremos que se trata de un bucle **trivial**. Por otro lado, un bucle no trivial que **no se descompone** será aquel que no es posible expresar como la concatenación de dos bucles no triviales. En otras palabras, un bucle se descompone si pasa por el origen más de dos veces, pues todo bucle siempre comienza y termina en el origen. Podemos asumir que toda caminata comienza en el origen pues nuestros cálculos no dependerán de este hecho.

Sean  $\ell_n$  el número de bucles de longitud  $n$  y  $r_n$  la cantidad de bucles que no se descomponen. Es claro que  $\ell_0 = 1$ , mientras que  $r_0 = 0$  pues el único bucle de longitud cero es el trivial. Observemos que al considerar las funciones generadoras ordinarias asociadas a las sucesiones  $(r_n)$ ,  $(\ell_n)$ , y aplicar el Teorema 5.1.1, se sigue que

$$\ell_n = \sum_{k=0}^n r_k \ell_{n-k}$$

para  $n \geq 1$ , pues cada bucle no trivial se puede expresar como la concatenación de un bucle que no se descompone, seguido por un bucle cualquiera de longitud  $n - k$  posiblemente trivial. Al dividir ambos lados de esta ecuación por  $(2d)^n$ , el número total de caminatas de longitud  $n$  en  $\mathbb{Z}^d$ , obtenemos que

$$q_n = \sum_{k=0}^n p_k q_{n-k} \tag{5.5}$$

para cada  $n \geq 1$  con  $q_0 = 1$ , donde, al igual que antes,  $p_n$  denota la probabilidad de que la caminata regrese al origen por primera vez en el paso  $n$ , mientras que  $q_n$  es la probabilidad de que la caminata se encuentre en el origen después de  $n$  pasos.

Consideremos ahora a las funciones generadoras ordinarias:

$$P(z) = \sum_{n=0}^{\infty} p_n z^n \quad \text{y} \quad Q(z) = 1 + \sum_{n=1}^{\infty} q_n z^n,$$

teniendo en cuenta que  $P(1) = p$ . De acuerdo con la relación entre  $(p_n)$  y  $(q_n)$  en (5.5), se tiene que

$$P(z)Q(z) = Q(z) - 1$$

en el álgebra de series de potencias formales  $\mathbb{R}[[z]]$ .

Hasta ahora, hemos conseguido una fuerte relación entre  $P$  y  $Q$ , pero esto no es suficiente, pues la evaluación en series de potencias formales no tiene mucha relevancia. Es por eso que buscamos que esta relación se cumpla como funciones analíticas en el disco unitario complejo, ya que así podríamos considerar el límite cuando  $z \rightarrow 1$ . Dado que  $p_n \leq q_n \leq 1$ , la idea anterior se puede llevar a cabo pues podemos concluir que cada una de estas series tiene radio de convergencia mayor o igual que 1, de manera que esta identidad puede considerarse válida en el álgebra de las funciones analíticas en el disco abierto complejo de radio 1. Dado que cada  $q_n > 0$ , tenemos que  $Q(z) > 0$  cuando  $z \in [0, 1)$ , y entonces

$$P(z) = 1 - \frac{1}{Q(z)}, \quad z \in [0, 1).$$

Nuestra tarea ahora es calcular el límite de  $Q$  cuando  $z \rightarrow 1$ , ya que al aplicar el teorema de series de potencias de Abel tendremos

$$p = \lim_{\substack{z \rightarrow 1 \\ z \in [0,1)}} P(z) = 1 - \frac{1}{\lim_{\substack{z \rightarrow 1 \\ z \in [0,1)}} Q(z)}. \quad (5.6)$$

Es claro que el límite en el denominador es  $\infty$  o un número real positivo mayor que 1. En el primer caso tendremos  $p = 1$  (recurrencia), y en el otro  $p < 1$  (transitoriedad). Nuestro siguiente paso es mostrar que todos los cálculos dependen de la dimensión  $d$  de la malla  $\mathbb{Z}^d$ . De esta manera, debemos probar que

$$\lim_{\substack{z \rightarrow 1 \\ z \in [0,1)}} Q(z) = \infty$$

para dimensiones 1 y 2, y que el límite es finito para dimensiones mayores o iguales a 3.

Para analizar este límite necesitamos una representación manejable de  $Q$ , veremos que esto se reduce a encontrar una expresión para la función generadora de bucles

$$L(z) = \sum_{n=0}^{\infty} \ell_n z^n,$$

teniendo en mente que  $Q(z) = L(z/2d)$ .

Mientras que  $L(z)$  es difícil de analizar directamente, la función generadora exponencial de bucles

$$E(z) = \sum_{n=0}^{\infty} \ell_n \frac{z^n}{n!} \quad (5.7)$$

es bastante más accesible, ya que de acuerdo con el Teorema 5.1.2, multiplicación de funciones generadoras exponenciales corresponde a contar el número total de formas posibles en las que podemos mezclar dos estructuras distintas en un conjunto dado, en este caso las estructuras corresponden a escoger un bucle u otro, contando la cantidad de formas posibles en las que podemos mezclar bucles distintos.

Para mostrar la dependencia del límite en (5.6) de la dimensión  $d$  de la malla  $\mathbb{Z}^d$ , comencemos por denotar a la función  $E$  en (5.7), de dimensión  $d$ , por  $E_d$  y veamos que es posible expresar a  $E_2$  en términos de  $E_1$ . De igual manera, denotemos por  $\ell_n^{(d)}$  al número de bucles en la malla  $\mathbb{Z}^d$  de longitud  $n$ . Cuando  $d = 2$ , un bucle en  $\mathbb{Z}^2$  es una caminata cerrada que toma pasos unitarios en únicamente dos direcciones: horizontal y vertical. Un bucle de longitud  $n$  en  $\mathbb{Z}^2$  consiste de un cierto número  $k$  de pasos horizontales y  $n - k$  pasos verticales, de manera que los  $k$  pasos horizontales forman un bucle de longitud  $k$  en  $\mathbb{Z}^1$  y los  $n - k$  pasos verticales constituyen un bucle de longitud  $n - k$  en  $\mathbb{Z}^1$ . Entonces la cantidad de bucles en  $\mathbb{Z}^2$  de longitud  $n$  proviene de una mezcla de bucles (que pueden o no descomponerse) en  $\mathbb{Z}^1$ , en las direcciones vertical y horizontal, luego, como consecuencia del Teorema 5.1.2, resulta que

$$\ell_n^{(2)} = \sum_{k=0}^n \binom{n}{k} \ell_k^{(1)} \ell_{n-k}^{(1)}.$$

Entonces, de la expresión anterior se deduce que

$$E_2(z) = E_1(z)^2.$$

Observemos que este mismo razonamiento es válido para cualquier dimensión  $d$ , pues para  $d = 3$ , podemos descomponer una caminata en tres componentes, y así de manera sucesiva. De esta manera, concluimos que

$$E_d(z) = E_1(z)^d.$$

Con esto, todo lo que necesitamos hacer ahora es calcular la función  $E_1(z)$ . Afortunadamente contar bucles en dimensión uno es fácil ya que si  $n = 2k$  es par, por cada colección de  $k$  elementos de un conjunto de  $2k$  elementos podemos construir un bucle de longitud  $2k$ , pues tal bucle debe consistir en  $k$  pasos a la izquierda y  $k$  pasos a la derecha. Puesto que no existen bucles de longitud impar en  $\mathbb{Z}^1$ , tenemos

$$\ell_n^{(1)} = \begin{cases} \binom{2k}{k} & \text{si } n = 2k \text{ es par} \\ 0 & \text{si } n \text{ es impar} \end{cases},$$

y entonces

$$E_1(z) = \sum_{k=0}^{\infty} \binom{2k}{k} \frac{z^{2k}}{(2k)!} = \sum_{k=0}^{\infty} \frac{z^{2k}}{k!k!}.$$

Es aquí cuando obtenemos una conexión con las soluciones de la ecuación de Bessel, pues nuestra función generadora exponencial para caminatas en la malla de dimensión uno es una función de Bessel modificada del primer tipo. En efecto, a partir de la representación en serie de potencias de  $I_0$ , observamos que

$$I_0(2z) = \sum_{k=0}^{\infty} \frac{z^{2k}}{k!k!} = E_1(z),$$

de modo que

$$E_d(z) = I_0(2z)^d.$$

Esto completa nuestro análisis de la función  $E$ .

Recordemos que en realidad estamos buscando una representación útil de nuestra función  $L$ . También es necesario tener en mente que la relación entre  $E$  y  $L$  es que ambas son funciones generadoras de la misma sucesión, sólo que  $E$  es exponencial mientras que  $L$  es ordinaria. Así pues, dado que la transformada de Borel convierte funciones generadoras exponenciales en funciones generadoras ordinarias, como vimos en la sección 5.1, se sigue que

$$L_d(z) = \mathcal{B}E_d(z) = \mathcal{B}[I_0(2z)^d],$$

donde  $L_d$  denota la correspondiente función generadora ordinaria de la sucesión  $(\ell_n^{(d)})$  que cuenta bucles de longitud  $n$  en  $\mathbb{Z}^d$ .

Entonces

$$Q(z) = L\left(\frac{z}{2d}\right) = \int_0^\infty I_0\left(\frac{tz}{d}\right)^d e^{-t} dt. \quad (5.8)$$

Ahora estamos listos para estimar el límite en (5.6). Utilizando esta representación integral, es suficiente estimar la siguiente cola cuando  $N$  es suficientemente grande:

$$\int_N^\infty I_0\left(\frac{tz}{d}\right)^d e^{-t} dt.$$

Para  $N$  suficientemente grande, el comportamiento de la integral está determinado por el comportamiento del integrando cuando  $t \rightarrow \infty$ . Para estimar el integrando, utilizaremos la representación integral de  $I_0$ ,

$$I_0\left(\frac{tz}{d}\right) = \frac{1}{\pi} \int_0^\pi e^{tf(\theta)} d\theta,$$

donde  $f(\theta) = \frac{z}{d} \cos(\theta)$ .

El siguiente paso es utilizar el principio de Laplace, Teorema 5.3.1, para obtener una estimación de la representación integral de  $Q$  dada en (5.8). Para esto, observemos que  $f$ , como función de  $\theta$ , es suave con mínimo global en  $[0, \pi]$  dado por  $\cos(\pi) = -1$ . A pesar de que este mínimo está en el extremo derecho y no en el interior del intervalo, podemos aplicar el principio de Laplace en el intervalo  $[0, \pi + \varepsilon]$  y después tomar el límite cuando  $\varepsilon \rightarrow 0$ . Entonces, dado que  $f''(\pi) = z/d$ , concluimos que

$$\begin{aligned} I_0\left(\frac{tz}{d}\right) &\sim \frac{1}{\pi} \sqrt{\frac{2\pi d}{tz}} e^{tz/d} \\ &= \text{cte} \cdot (tz)^{-1/2} e^{tz/d} \end{aligned}$$

pensando en  $\varphi(x) = -\frac{z}{d} \cos(x)$  y  $x_0 = \pi$ , de acuerdo a la notación del Teorema 5.3.1. Al elevar a la potencia  $d$  obtenemos

$$I_0\left(\frac{tz}{d}\right)^d \sim \text{cte} \cdot e^{tz} (tz)^{-d/2}$$

y finalmente resulta que

$$I_0\left(\frac{tz}{d}\right)^d e^{-t} \sim \text{cte} \cdot e^{t(z-1)} (tz)^{-d/2},$$

donde cada una de estas estimaciones se cumplen cuando  $t \rightarrow \infty$ .

Al tomar la integral en la última estimación obtenemos que

$$\int_N^\infty I_0\left(\frac{tz}{d}\right)^d e^{-t} dt \sim \int_N^\infty \text{cte} \cdot e^{t(z-1)}(tz)^{-d/2} dt,$$

cuando  $t \rightarrow \infty$ . Ahora el único paso restante es tomar el límite cuando  $z \rightarrow 1$  con  $z \in [0, 1)$ . Esto se logra aplicando el teorema de la convergencia monótona de Lebesgue, pues la colección de funciones  $t \mapsto e^{t(z-1)}(tz)^{-d/2}$  es creciente precisamente cuando  $z \rightarrow 1$  con  $z$  real. Entonces se sigue que

$$\lim_{\substack{z \rightarrow 1 \\ z \in [0, 1)}} \int_N^\infty e^{t(z-1)}(tz)^{-d/2} dt = \int_N^\infty \lim_{\substack{z \rightarrow 1 \\ z \in [0, 1)}} e^{t(z-1)}(tz)^{-d/2} dt = \int_N^\infty t^{-d/2} dt.$$

Así pues, para  $N$  suficientemente grande,

$$\lim_{\substack{z \rightarrow 1 \\ z \in [0, 1)}} \int_N^\infty I_0\left(\frac{tz}{d}\right)^d e^{-t} dt \sim \text{cte} \cdot \int_N^\infty t^{-d/2} dt$$

y por lo tanto, partiendo de la ecuación (5.8) y la estimación de arriba, se concluye que la recurrencia o transitoriedad de la caminata aleatoria simple en  $\mathbb{Z}^d$  es equivalente a la divergencia o convergencia de la integral

$$\int_N^\infty t^{-d/2} dt,$$

respectivamente. Dado que esta integral diverge para  $d = 1, 2$  y converge para  $d \geq 3$ , el teorema de Polya queda demostrado.





# Bibliografía

- [1] G. E. Andrews, R. Askey y R. Roy. *Special Functions*. Cambridge University Press, 2000. ISBN: 0-521-78988-5.
- [2] E. Behrends. *Introduction to Markov Chains with Special Emphasis on Rapid Mixing*. Vieweg, 2000. ISBN: 978-3-528-06986-5.
- [3] N. Bleistein y R. A. Handelsman. *Asymptotic Expansions of Integrals*. Holt, Rineheart y Winston, New York, 1975. ISBN: 0-486-65082-0.
- [4] M. Bóna. *A Walk Through Combinatorics: An Introduction to Enumeration and Graph Theory*. Second. World Scientific, 2006. ISBN: 981-256-885-9.
- [5] S. Chibb y E. Greenberg. “Understanding the Metropolis-Hastings Algorithm”. *The American Statistician* (1995).
- [6] P. Diaconis y L. Saloff-Coste. “What Do We Know about the Metropolis Algorithm?” *Journal of Computer and System Sciences* (1998).
- [7] F. R. Gantmacher. *The Theory of Matrices, Volume Two*. AMS Chelsea Publishing, 2000. ISBN: 0-8218-2664-6.
- [8] G. Grimmet. *Probability on graphs*. Cambridge, 2010.
- [9] P. Von Hilgers y A. N. Langville. “The Five Greatest Applications of Markov Chains”. *American Mathematical Society* (2006).
- [10] A. Klenke. *Probability Theory: A Comprehensive Course*. Second. Springer, 2013. ISBN: 978-1-4471-5360-3.
- [11] G. Modica y L. Poggiolini. *A First Course in Probability and Markov Chains*. Wiley, 2013. ISBN: 978-1-119-94487-4.
- [12] A. Nachmias. *Planar Maps, Random Walks and Circle Packing*. Springer, 2010. ISBN: 978-3-030-27967-7.
- [13] J. Novak. “Pólya’s Random Walk Theorem”. *The American Mathematical Monthly* (2013).
- [14] N. Privault. *Understanding Markov Chains*. Springer, 2013. ISBN: 978-981-4451-50-5.

- [15] V. N. Sachkov y V. E. Tarakanov. *Combinatorics of Nonnegative Matrices*. American Mathematical Society, 2002. ISBN: 0-8218-2788-X.
- [16] R. Z. Yeh. "A Geometric Proof of Markov Ergodic Theorem". *American Mathematical Society* (1969).