UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

Eigenvalores y tasas de convergencia de algunas cadenas de Markov con aplicaciones.

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

Actuario

PRESENTA:

Martínez Hernández Israel

TUTORA

Dra. Ana Meda Guardiola

Ciudad Universitaria, Cd. Mx., 2022

Hoja de Datos del Jurado

1.Datos del alumno	1.Datos del alumno
Apellido paterno	Martínez
Apellido materno	Hernández
Nombre(s)	Israel
Teléfono	5520217320
Universidad Nacional Autónoma de México	Universidad Nacional Autónoma de México
Facultad de Ciencias	Facultad de Ciencias
Carrera	Actuaría
Número de cuenta	314133610
2. Datos del tutor	2. Datos del tutor
Grado	Dra
Nombre(s)	Ana
Apellido paterno	Meda
Apellido materno	Guardiola
3. Datos del sinodal 1	3. Datos del sinodal 1
Grado	Dr.
Nombre(s)	Yuri
Apellido paterno	Salazar
Apellido materno	Flores
4. Datos del sinodal 2	4. Datos del sinodal 2
Grado	Dr
Nombre(s)	Daniel
Apellido paterno	Labardini
Apellido materno	Fragoso
5. Datos del sinodal 3	5. Datos del sinodal 3
Grado	M. en C.
Nombre(s)	Clotilde
Apellido paterno	García
Apellido materno	Villa
6. Datos del sinodal 4	6. Datos del sinodal 5
Grado	Mat.
Nombre(s)	julio César
Apellido paterno	Guevara
Apellido materno	Bravo
7. Datos del trabajo escrito.	7.Datos del trabajo escrito.
Título	Eigenvalores y tasas de convergencia de algunas cadenas de Markov con aplicaciones.
Número de páginas	89
Año	2021

Índice general

Capítulos **Página**

Capítulo 1. Preliminares

En este capítulo se presentará una breve introducción al concepto de proceso estocástico, nos enfocaremos a procesos de tiempo discreto, específicamente a las cadenas de Markov, veremos propiedades y teoremas que serán de gran importancia para este trabajo. Cabe mencionar que no daremos la demostración de los teoremas y proposiciones que se encuentren en este capítulo, sin embargo daremos la referencia sobre dónde poder consultar la demostración completa.

Sección 1.1 Introducción a las cadenas de Markov

Consideremos un sistema que evoluciona en el tiempo entre diferentes estados (determinados previamente) este sistema se modelará por una ley de movimiento, con X_t representa el estado del sistema en el tiempo t, suponemos que la evolución de éste no es determinista, sino que está regida por algún componente aleatorio, resulta natural considerar a X_t como una variable aleatoria para cada valor t. Hasta aquí, tenemos una colección de variables aleatorias indexadas por el tiempo, a esta colección se le conoce como un proceso estocástico y sirve como modelo para representar la evolución aleatoria de un sistema a lo largo del tiempo. Resulta de gran relevancia estudiar las relaciones entre las variables aleatorias de dicha colección que permiten clasificar a los procesos estocásticos.
Daremos una definición formal de proceso estocástico.

Definición **1.1.**
Un proceso estocástico $X = \{X_t\}_{t\in T}$ es una colección de variables aleatorias, definidas en el mismo espacio de probabilidad, que toman valores en un conjunto E con parámetro T, es decir para toda $t \in T$, X_t es una variable aleatoria.

En la definición anterior, diremos que T es el espacio parametral y lo podemos interpretar como el tiempo, de esta manera X_t la entenderemos como el estado del proceso al tiempo t.

Continuando con la definición de proceso estocástico, lo podemos ver como una función de dos variables

$$X : T \times \Omega \to E$$

de tal forma que para cada pareja $(t,\omega) \in (T,\Omega)$ se le asocia $X(t,\omega) = X_t(\omega)$, por lo que para cada $t \in T$ la función

$$\omega \to X_t(\omega)$$

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

$$t \to X_t(\omega)$$

es una trayectoria o realización del proceso.
Dependiendo de las características del espacio parametral T, podemos hacer distinciones entre procesos estocásticos.

Definición **1.2.** *Proceso estocástico a tiempo discreto*
Sea $X = \{X_t\}_{t\in T}$ un proceso estocástico, si T es un conjunto a lo más numerable, se dice que X es un proceso estocástico a tiempo discreto*, nosotros ocuparemos $T \subset \mathbb{N}$ y denotaremos al proceso por $X = \{X_n\}_{n\geq 0}$.*

Intuitivamente, podemos definir *proceso estocástico a tiempo continuo*, pero como solo trabajaremos con procesos estocásticos a tiempo discreto no los definiremos.

El conjunto E donde las variables aleatorias toman valores, lo llamaremos *espacio de estados* del proceso y al igual que con T este puede ser discreto o continuo, por el momento no supondremos nada sobre si E es discreto o continuo.

Una propiedad importante que poseen algunos procesos estocásticos a tiempo discreto, es que el valor de la variable en el n-ésimo momento está completamente determinado por los valores en el n-1-ésimo momento. Esta propiedad es conocida como *propiedad de Markov*.

Ahora pasaremos definir las cadenas de Markov, junto con la *propiedad de Markov*.

Definición **1.3.** *Cadena de Markov*
Una cadena de Markov *es un proceso estocástico a tiempo discreto $\{X_n\}_{n\geq 0}$, con espacio de estados E discreto y que satisface la propiedad de Markov, esto es, para cualquier entero $n \geq 0$ y para cualquiera estados $x_0, \cdots, x_{n+1}$ se cumple que*

$$\mathbb{P}[X_{n+1} = x_{n+1} \mid X_0 = x_0, \cdots, X_n = x_n] = \mathbb{P}[X_{n+1} = x_{n+1} \mid X_n = x_n] \qquad (1.1)$$

A la distribución de X_0 le llamaremos distribución inicial. A la familia $\mathbb{P}[X_{n+1} = j \mid X_n = i]$ se le conoce como familia de probabilidades de transición de la cadena de Markov, esta familia describe la evolución de la cadena en el tiempo.

Si la familia de probabilidades de una cadena de Markov es independiente de n, es decir, $\mathbb{P}[X_{n+1} = j \mid X_n = i] = \mathbb{P}[X_1 = j \mid X_0 = i]$ para toda $n \in \mathbb{N}$, entonces diremos que es una *cadena de Markov homogénea*.

Cuando tenemos una cadena de Markov homogénea, usaremos una notación para la transición a un paso y a m pasos, denotaremos por $\mathbb{P}[X_{n+1} = j \mid X_n = i] = p_{ij}$, además

$\mathbb{P}[X_{n+m} = j \mid X_n = i] = p_{ij}^{(m)}$, para toda $i, j \in E$.

A partir de este punto trabajaremos con cadenas de Markov homogéneas con espacio de estados E finito.

Daremos la definición de matriz de transición.

Definición **1.4.** *Matriz de transición*
Sea $\{X_n\}_{n\geq 0}$ una cadena de Markov homogénea con espacio de estados finito $E = \{1, \ldots, r\}$, definimos la matriz de transición como la matriz cuadrada de orden r formada por las probabilidades de transición, esto es

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1r} \\ p_{21} & p_{22} & \cdots & p_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ p_{r1} & p_{r2} & \cdots & p_{rr} \end{bmatrix}$$

Notemos que la matriz de transición es una matriz cuadrada cuyas entradas son no negativas, además de que la suma de cada renglón es uno, a las matrices que cumplen estas dos condiciones las llamaremos *matrices estocásticas*.

Algo que es importante mencionar es que $p_{ij}^{(n)}$ es la entrada ij de la n-ésima potencia de P. En general, es difícil calcular explícitamente las potencias de nuestra matriz de transición, pero podemos ayudarnos de sistemas computacionales para resolver dicha tarea.

Otro resultado importante es la ecuación de Chapman-Kolmogorov, presentada a continuación y cuya demostración se puede encontrar en [Paul G. Hoel, Cap. 1].

Proposición **1.1.**
Sean $r,n,m \in \mathbb{N}$, $E = \{1,...,r\}$ y $X = \{X_n\}_{n\geq 0}$ una cadena de Markov con matriz de transición $P = (p_{ij})_{i,j\in E}$, entonces

$$p_{ij}^{(n+m)} = \sum_{k\in E} p_{ik}^{(n)} p_{kj}^{(m)}. \tag{1.2}$$

Sección 1.2 Clasificación de estados

Antes de clasificar nuestro espacio de estados es necesario dar algunas definiciones y algunos resultados, el primero de ellos es la definición de comunicación.

Definición **1.5.**
Sean $r \in \mathbb{N}$, $E = \{1,...,r\}$, además sean $x,y \in E$, decimos que y es accesible *desde x si existe $n \in \mathbb{N}$ de tal forma que $p_{xy}^{(n)} > 0$ denotado por $x \to y$.*
Además, si $y \to x$ y $x \to y$, entonces decimos que x e y se comunican y lo denotaremos por $x \leftrightarrow y$.

Para la definición anterior se tiene que la relación $x \leftrightarrow y$ que denominamos comunicación es una relación de equivalencia. La demostración de esta afirmación la podemos encontrar en [Rincón L., Cap. 3].
Además, como la comunicación es una relación de equivalencia podemos inducir una partición del espacio de estados E de una cadena de Markov, dada por subconjuntos de estados que se comunican entre sí, es decir, dos estados que pertenecen al mismo elemento de la partición si y sólo si estos se comunican, por lo que el espacio de estados está dividida en clases de comunicación. A la clase de comunicación del estado x la denotaremos por $C(x)$, por lo que $x \leftrightarrow y$ si y sólo si $C(x) = C(y)$.

Continuando con la partición inducida por la comunicación decimos que $C \subset E$ es un conjunto cerrado si para toda $y \in E \setminus C$ y para toda $x \in C$, x no puede acceder a y. Además, si C no es cerrado decimos que es abierto.

Definición **1.6.**
Sean $r, n \in \mathbb{N}$, $n \geq 0$ y $E = \{1, ..., r\}$, denotamos por $\mathfrak{f}_{ij}^{(n)}$ a la probabilidad de que una cadena que inicia en el estado $i \in E$, llegue al estado $j \in E$ por primera vez en exactamente n pasos, es decir,

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

Cada estado en una cadena de Markov tiene propiedades de acuerdo a su probabilidad de regresar, según la definición anterior, y con esto podemos clasificar a los estados de la siguiente manera

Definición **1.7.**
Sean $r \in \mathbb{N}$ y $E = \{1, ..., r\}$, decimos que el estado $x \in E$ es recurrente *si la probabilidad de regresar a x es uno, esto es*

$$\mathbb{P}[X_n = x \mid X_0 = i] = 1$$

para algún $n \geq 1$. Además, si un estado no es recurrente lo llamaremos transitorio.

Con esta clasificación que tenemos del espacio de estados y la partición del mismo inducida por la comunicación, tenemos el siguiente teorema llamado teorema de descomposición de las cadenas de Markov y cuya prueba la podemos encontrar en [Paul G. Hoel, Cap. 1].

Teorema **1.1.**
Sean $r \in \mathbb{N}$ y $E = \{1, ..., r\}$ el espacio de estados de una cadena de Markov, entonces E tiene la siguiente partición única

$$E = T \cup C_1 \cup C_2 \cup ...$$

donde T es el conjunto de estados transitorios y cada C_i son clases cerradas e irreducibles de estados recurrentes.

El teorema anterior nos muestra las opciones en que una cadena de Markov puede estar, esto es, si $x \in C_k$, entonces la cadena nunca abandonará a dicha clase por lo que

ahora podemos considerar un nuevo espacio de estados $E' = C_k$.
Por otro lado, si $x \in T$, entonces dependiendo si el espacio de estados es finito o no, la cadena permanece por siempre en T o se mueve a alguna clase C_k y permanecerá ahí eternamente. En el caso de que el espacio de estados sea finito, el primer caso nunca sucede, es decir la cadena se moverá a alguna clase C_k en tiempo finito casi seguramente.

Ahora, daremos varios resultados que derivan de las definiciones de comunicación, transitoriedad y recurrencia. La demostración de estos resultados podemos encontrarlos en [Paul G. Hoel, Cap. 1].

Proposición **1.2.**
Sean $r \in \mathbb{N}$ y $E = \{1,...,r\}$ el espacio de estados de una cadena de Markov, además sean $i, j \in E$ estados que se comunican, si i es transitorio entonces j también lo es.

La proposición anterior nos da como consecuencia directa que la transitoriedad o recurrencia son propiedades de clase.

Proposición **1.3.**
Sean $r \in \mathbb{N}$ y $E = \{1,...,r\}$ el espacio de estados de una cadena de Markov, entonces una clase es recurrente si y sólo si es cerrada.

Corolario **1.**
En una cadena de Markov irreducible con espacio de estados finito todos sus estados son recurrentes.

Corolario **2.** *Toda cadena de Markov con espacio de estados finito tiene por lo menos un estado recurrente.*

Ejemplo **1.1.**
Sea $\{X_n\}_{n\geq 0}$ una cadena de Markov con espacio de estados $E = \{0,1,\ldots,5\}$ y matriz de transición

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

Para clasificar nuestros estados es necesario saber cuales parejas $x, y \in E$ se comunican, según la definición 1.5*, para ello veremos a un paso quién es accesible desde quién, usando la notación de la definición* 1.5 *tenemos que* $0 \to 1$, $1 \to 0$, $2 \to 4$, $3 \to 0$, $3 \to 1$, $3 \to 4$, $4 \to 5$, $4 \to 2$, $5 \to 1$, $5 \to 3$ *y* $5 \to 4$.
Por lo que $0 \leftrightarrow 1$, $2 \leftrightarrow 4$ *y* $3 \leftrightarrow 5$. *Además, como* $0 \to x$ *si y sólo si* $x = 1$ *y* $1 \to y$ *si y sólo si* $y = 0$, *o bien* $0 \nrightarrow x$ *y* $1 \nrightarrow y$ *para toda* $x, y \in \{2,3,4,5\}$, *así* $C_1 = \{0,1\}$ *es un conjunto cerrado.*

Por otro lado, como $2 \to x$ si y sólo si $x = 4$ y $4 \to y$ si y sólo si $y = 2$, o bien $4 \nrightarrow x$ y $2 \nrightarrow y$ para toda $x, y \in \{0,1,3,5\}$, así $C_2 = \{2,4\}$ es otro conjunto cerrado.
Finalmente, $3 \to 0$ pero $0 \nrightarrow 3$, entonces $3 \notin C(0) = C(1)$, $3 \to 4$ pero $4 \nrightarrow 3$, entonces $3 \notin C(4) = C(2)$ y como $5 \to 1$ pero $1 \nrightarrow 5$, entonces $5 \notin C(1) = C(0)$, $5 \to 4$ pero $4 \nrightarrow 5$, entonces $5 \notin C(4) = C(2)$, por lo que $T = \{3,5\}$.
Así, podemos escribir a nuestro espacio de estados como $E = T \cup C_1 \cup C_2$, esta partició de E es para ejemplificar el teorema 1.1, además de que cada C_i es una clase cerrada e irreducible de estados recurrentes mientras que T es el conjunto de estados transitorios.

Sección 1.3 Distribución estacionaria

En esta sección hablaremos de un concepto importante en la teoría de las cadenas de Markov, estamos hablando de la distribución estacionaria la cual es muy importante para análisis de la cadena a largo plazo, presentaremos conceptos adicionales y propiedades que no serán demostradas, pero que pueden encontrarse en [Paul G. Hoel, Cap. 2].

Definición **1.8.**
Sean $r \in \mathbb{N}$, $E = \{1,...,r\}$ el espacio de estados de una cadena de Markov, $P = (p_{ij})_{i,j \in E}$ la matriz de transición de la cadena y $\pi \in \mathbb{R}^r$, decimos que $\pi = (\pi(1), \pi(2), ..., \pi(r))$ es una distribución estacionaria de la cadena de Markov, si satisface

1. $\pi(i) \geq 0$ *para toda* $i \in E$.
2. $\sum_{i \in E} \pi(i) = 1$.
3. $\sum_{i \in E} \pi(i) p_{ij} = \pi(j)$ *para toda* $j \in E$.

Este último punto puede ser reescrito es su forma matricial de la siguiente manera

$$\pi P = \pi$$

Por las condiciones que ponemos en la definición 1.8, es fácil ver que el encontrar una distribución estacionaria es equivalente a resolver un sistema de $r+1$ ecuaciones con r incógnitas, por lo que podemos no tener ninguna solución, tener una infinidad de soluciones o una única solución. En esta sección veremos resultados que serán de gran utilidad para la obtención de distribuciones estacionarias, así como condiciones necesarias y suficientes para garantizar la existencia de una única distribución estacionaria.

Proposición **1.4.**
Sean $r \in \mathbb{N}$ y $E = \{1,...,r\}$ el espacio de estados de una cadena de Markov, además sea π una distribución estacionaria de la cadena de Markov, si $i \in E$ es un estado transitorio entonces $\pi(i) = 0$.

Por definición, si la distribución inicial de una cadena de Markov es una distribución estacionaria entonces para toda $n \in \mathbb{N}$, X_n tiene la misma distribución.

Continuaremos con un ejemplo para ilustrar la definición 1.8 y la proposición 1.4.

Ejemplo **1.2.**
Continuando con el ejemplo 1.1*, veremos la o las distribuciones estacionarias de la cadena. Así, sea π una distribución de probabilidad sobre E tal que*

1. $\sum_{i\in E} \pi(i) = 1$ *para toda* $i \in E$.

2. $\sum_{i\in E} \pi(i)p_{ij} = \pi(j)$ *para toda* $j \in E$.

Las propiedades anteriores nos dan como resultado un sistema de 7 ecuaciones y 6 incognitas, las cuales podemos ver abajo

$$\left\{\begin{array}{lllllll}
\pi(0) & +\pi(1) & +\pi(2) & +\pi(3) & +\pi(4) & +\pi(5) & = 1 \\
\frac{1}{2}\pi(0) & +\frac{1}{3}\pi(1) & +0\pi(2) & +\frac{1}{4}\pi(3) & +0\pi(4) & +0\pi(5) & = \pi(0) \\
\frac{1}{2}\pi(0) & +\frac{2}{3}\pi(1) & +0\pi(2) & +\frac{1}{4}\pi(3) & +0\pi(4) & +\frac{1}{5}\pi(5) & = \pi(1) \\
0\pi(0) & +0\pi(1) & +\frac{1}{8}\pi(2) & +0\pi(3) & +\frac{3}{4}\pi(4) & +0\pi(5) & = \pi(2) \\
0\pi(0) & +0\pi(1) & +0\pi(2) & +0\pi(3) & +0\pi(4) & +\frac{1}{5}\pi(5) & = \pi(3) \\
0\pi(0) & +0\pi(1) & +\frac{7}{8}\pi(2) & +\frac{1}{4}\pi(3) & +\frac{1}{4}\pi(4) & +\frac{1}{5}\pi(5) & = \pi(4) \\
0\pi(0) & +0\pi(1) & +0\pi(2) & +\frac{1}{4}\pi(3) & +0\pi(4) & +\frac{2}{5}\pi(5) & = \pi(5)
\end{array}\right.$$

Pasando todas las variables del lado izquierdo de la igualdad y todas las constantes del lado derecho nuestro sistema de ecuaciones queda de la siguiente forma

$$\left\{\begin{array}{lllllll}
\pi(0) & +\pi(1) & +\pi(2) & +\pi(3) & +\pi(4) & +\pi(5) & = 1 \\
\frac{-1}{2}\pi(0) & +\frac{1}{3}\pi(1) & & +\frac{1}{4}\pi(3) & & & = 0 \\
\frac{1}{2}\pi(0) & -\frac{1}{3}\pi(1) & & +\frac{1}{4}\pi(3) & & +\frac{1}{5}\pi(5) & = 0 \\
 & & -\frac{7}{8}\pi(2) & & +\frac{3}{4}\pi(4) & & = 0 \\
 & & & -\pi(3) & & +\frac{1}{5}\pi(5) & = 0 \\
 & & +\frac{7}{8}\pi(2) & +\frac{1}{4}\pi(3) & -\frac{3}{4}\pi(4) & +\frac{1}{5}\pi(5) & = 0 \\
 & & & +\frac{1}{4}\pi(3) & & -\frac{3}{5}\pi(5) & = 0
\end{array}\right.$$

Sumando la ecuación (4) *con la ecuación* (6)*, tenemos como resultado que* $\frac{1}{4}\pi(3) + \frac{1}{5}\pi(5) = 0$*, pero* $\pi(i) \geq 0$ *para toda* $i \in E$*, por lo que* $\pi(3) = \pi(5) = 0$*. Así, nuestro sistema de ecuaciones queda de la siguiente forma*

$$\left\{\begin{array}{lllllll}
\pi(0) & +\pi(1) & +\pi(2) & & +\pi(4) & & = 1 \\
\frac{-1}{2}\pi(0) & +\frac{1}{3}\pi(1) & & & & & = 0 \\
\frac{1}{2}\pi(0) & -\frac{1}{3}\pi(1) & & & & & = 0 \\
 & & +\frac{1}{8}\pi(2) & & +\frac{3}{4}\pi(4) & & = 0 \\
 & & & \pi(3) & & & = 0 \\
 & & +\frac{7}{8}\pi(2) & & -\frac{3}{4}\pi(4) & & = 0 \\
 & & & & & \pi(5) & = 0
\end{array}\right.$$

En este punto podemos ver que nuestro sistema de ecuaciones no tiene solución única, debido a que la ecuación (2) *es igual a la ecuación* (3)*, así como la ecuación* (4) *es igual a la ecuación* (6)*, entonces podemos asignar un valor a* $\pi(1)$ *para determinar el* *valor de las demás variables en términos de esta. Supongamos que* $\pi(1) = \beta$*, entonces*

$$\begin{cases} \pi(0) + \pi(1) + \pi(2) + \pi(4) = 1 \\ \pi(0) = \frac{2}{3}\beta \\ \pi(1) = \beta \\ \pi(2) = \frac{6}{7}\pi(4) \\ \pi(3) = 0 \\ \pi(4) = \pi(4) \\ \pi(5) = 0 \end{cases}$$

Sustituyendo los valores de $\pi(i)$ *en la ecuación* (1) *tenemos que*

$$\frac{2}{3}\beta + \beta + \frac{6}{7}\pi(4) + \pi(4) = 1$$
$$\frac{5}{3}\beta + \frac{13}{7}\pi(4) = 1$$
$$\pi(4) = \frac{7}{13}(1 - \frac{5}{3}\beta)$$

Sea $\alpha = \frac{5}{3}\beta$*, entonces* $\pi(0) = \frac{2}{5}\alpha$ *y* $\pi(1) = \frac{3}{5}\alpha$*, mientras que* $\pi(2) = \frac{6}{13}(1-\alpha)$ *y* $\pi(4) = \frac{7}{13}(1-\alpha)$*. Para que la ecuación* (1) *se cumpla correctamente y que los* $\pi(i)$ *sean no negativos,* α *tiene que estar en el intervalo* $[0,1]$*.*
Además, Si definimos $\pi_{c_1} = (\frac{2}{5}, \frac{3}{5}, 0, 0, 0, 0)$ *y* $\pi_{c_2} = (0, 0, \frac{6}{13}, 0, \frac{7}{13}, 0)$*, entonces el conjunto* $\{\pi_\alpha = \alpha\pi_{c_1} + (1-\alpha)\pi_{c_2} | \alpha \in [0,1]\}$ *son todas las distribuciones estacionarias de nuestra cadena de Markov.*
Recordemos que en la clasificación de los estados de nuestra cadena de Markov, hecha en el ejemplo 1.1*, los estados 3 y 5 son transitorios y deacuerdo con la propoción* 1.4 *su valor en la distribución estaconaria debe ser cero, lo cual concuerda con lo obtenido en este ejemplo.*

Ahora hablaremos de un concepto importante para la obtención de una única distribución estacionaria, este concepto se relaciona con la posible periodicidad con la que se visita un mismo estado.

Definición **1.9.**
Sean $r \in \mathbb{N}$*,* $E = \{1, \ldots, r\}$ *el espacio de estados de una cadena de Markov y* $P = (p_{ij})_{i,j \in E}$ *la matriz de transición de la cadena, definimos al* periodo *de un estado* $x \in E$ *como el máximo común divisor del conjunto* $\{n \in \mathbb{N} \mid p_{ii}^{(n)} > 0\}$*, es decir,*

$$d(i) := MCD\{n \in \mathbb{N} \mid p_{ii}^{(n)} > 0\}.$$

Proposición **1.5.** *El periodo es una propiedad de clase.*

Definición **1.10.** *Periodo de una cadena de Markov*
Sea $\{X_n\}_{n \geq 0}$ *una cadena de Markov irreducible, decimos que la cadena tiene periodo* d *si el periodo de cada estado es* d*.*
Además, si $d = 1$ *diremos que la cadena es aperiódica.*

En las cadenas de Markov muchas veces nos interesa saber lo que pasa con la cadena a largo plazo, por lo que es importante saber lo que pasa con $p_{ij}^{(n)}$ cuando n es muy grande, la siguiente definición nos habla de una distribución límite.

Definición **1.11.**
Sean $r \in \mathbb{N}$, $E = \{1, ..., r\}$ el espacio de estados de una cadena de Markov y $P = (p_{ij})_{i,j \in E}$ la matriz de transición de la cadena, si para cada $i, j \in E$ el límite

$$\pi(j) = \lim_{n \to \infty} p_{ij}^{(n)}$$

existe, entonces llamaremos a π la distribución límite.

De la definición anterior tenemos como consecuencia directa que la distribución límite es inducida por la matriz de transición P y es independiente de la distribución inicial, además la distribución límite es una forma equivalente de definir a la distribución estacionaria.

El límite de las potencias de la matriz de transición P es una matriz estocástica Π en la cual cada renglón es la distribución límite π, y a esta matriz la llamaremos *matriz estacionaria*.

Es hora de presentar el teoremas con el cual podemos garantizar la existencia y unicidad de la distribución estacionaria de una cadena de Markov.

Teorema **1.2.**
Sean $r \in \mathbb{N}$, $E = \{1, ..., r\}$ el espacio de estados de una cadena de Markov irreducible, aperiódica y con matriz de transición $P = (p_{ij})_{i,j \in E}$, entonces

1. *Si $n \to \infty$, entonces las potencias P^n se aproximan a Π la matriz estacionaria.*
2. *π es un vector de probabilidad con entradas estrictamente positivas.*
3. *Cualquier vector μ tal que $\mu P = \mu$ es un escalar multiplicado por π (π es único como vector de probabilidad).*
4. $\lim_{n \to \infty} \mathbb{P}[X_n = i] = \pi(i)$ *para toda $i \in E$, además*

$$\lim_{n \to \infty} \mathbb{P}[X_n = i \mid X_0 = j] = \pi(i)$$

para toda $i, j \in E$.

y el teorema 1.2.

En general, obtener P^n para toda $n \in \mathbb{N}$ no es una tarea trivial por lo que dar un ejemplo del teorema 1.2, esto es, obtener la distribución estacionaria de una cadena de Markov irreducible y aperiódica por medio de la distribución límite, requiere de encontar P^n para toda $n \in \mathbb{N}$, actualmente no hemos hablado sobre cómo lograr eso de una manera sencilla, por lo que en el siguiente capítulo daremos un teorema para encontrar P^n para toda $n \in \mathbb{N}$.

Capítulo 2. Matrices de transición

En este capítulo vamos a trabajar con matrices de transición asociadas a cadenas de Markov con espacio de estados finito, esto es $E = \{1,\ldots,r\}$, $r \in \mathbb{N}$. Haremos uso de herramientas de álgebra lineal para obtener resultados que serán ocupados en capítulos posteriores. Nos apoyaremos en la teoría de eigenvalores y eigenvectores para hacer un análisis de la matriz de transición a n pasos, y luego analizaremos cuando n es arbitrariamente grande pues recordemos que el comportamiento asintótico de una cadena de Markov a tiempo n está completamente descrita por su matriz de transición a n pasos y en algunos otros casos por su distribución inicial. Además, definiremos las matrices de transición reversibles y un producto escalar, para crear una base ortonormal de $\mathbb{R}^r$ y con ello poder demostrar el teorema de Perron-Frobenius.

Sección 2.1 Teorema de Perron-Frobenius

Hay resultados importantes de álgebra lineal para matrices que podemos aplicar a las matrices de transición de una cadena de Markov cuando tienen un espacio de estados finito. El teorema 2.3 que afirma sobre eigenvalores y eigenvectores de matrices cuadradas con r eigenvalores diferentes, nos ayudará a dar pie al teorema de Perron-Frobenius. Pero antes daremos la definición de eigenvalor y eigenvector.

Definición **2.12.** *Eigenvalores y eigenvetores*
Sea $r \in \mathbb{N}$ y $A \in \mathscr{M}_{rxr}(\mathbb{C})$, una matriz cuadrada de dimensión r con coeficientes en el conjunto de los números complejos. Si existe un escalar $\lambda \in \mathbb{C}$ y un vector columna $v \in \mathbb{C}^r$, $v \neq 0$ ($u \in \mathbb{C}^r$, $u \neq 0$ respectivamente) de tal manera que

$$Av = \lambda v \ (u^t A = \lambda u^t resp.)$$

entonces v (u resp.) es llamado eigenvector derecho (izquierdo resp.) de A asociado al eigenvalor λ

Teorema **2.3.** *Descomposición espectral*
Sea $r \in \mathbb{N}$ y $A \in \mathscr{M}_{rxr}(\mathbb{R})$, una matriz de dimensión r, supongamos que $\lambda_1,\ldots,\lambda_r$ los eigenvalores de A son distintos, entonces existen $u_1,\ldots,u_r$ y $v_1,\ldots,v_r$ los eigenvectores izquierdos y derechos de A respectivamente, cada u_i y v_i asociados a λ_i, $i \in \{1,\ldots,r\}$,

de tal manera que:

a. Los eigenvalores cumplen que $u_i^t v_j =< u_i, v_j >= \begin{cases} 0 & si \quad i \neq j \\ 1 & si \quad i = j \end{cases}$ *para todo i,* $j \in \{1, ..., r\}$.

b. El conjunto $U = \{u_1, ..., u_r\}$ *es linealmente independiente, al igual que* $V = \{v_1, ..., v_r\}$.

c. $A^n = \sum_{i=1}^{r} \lambda_i^n v_i u_i^t$.

La demostración del teorema anterior se encuentra en [Friedberg, Cap. 7]. Ahora consideremos el siguiente ejemplo para ilustrar el teorema 2.3, este ejemplo puede ser engañoso pues a pesar de que la matriz es relativamente sencilla, encontrar algunos eigenvectores puede ser muy engorroso.

Ejemplo **2.3.**
Considere una cadena de Markov homogénea sobre $E = \{1, 2, 3\}$ *y una matriz de transición*

$$P = \begin{pmatrix} 1-\alpha & \alpha & 0 \\ 0 & 1-\beta & \beta \\ \gamma & 0 & 1-\gamma \end{pmatrix} \qquad donde\ \alpha, \beta, \gamma \in (0,1).$$

Primero hallemos los eigenvalores, para ello haremos uso del polinomio característico de P, el cual está definido como

$$CA_P(\lambda) = det(P - \lambda \mathbb{I}_{rxr}),$$

por lo que el polinomio característico de P es

$$CA_P(\lambda) = (1-\alpha-\lambda)(1-\beta-\lambda)(1-\gamma-\lambda) - \alpha\beta\gamma,$$

los eigenvalores de P que buscamos son las raíces del polinomio característico CA_P*, es decir necesitamos que*

$$CA_P(\lambda) = (1-\alpha-\lambda)(1-\beta-\lambda)(1-\gamma-\lambda) - \alpha\beta\gamma = 0. \tag{2.3}$$

Como P es una matriz estocástica entonces una raíz que tiene CA_P *es* $\lambda_1 = 1$*, es decir,* $\lambda = 1$ *es un eigenvalor de cualquier* matriz estocástica $P \in \mathscr{M}_{rxr}(\mathbb{R})$*, ya que*

$$P\mathbb{1}_r = 1\mathbb{1}_r. \tag{2.4}$$

Ahora, si desarrollamos la ecuación (2.3) y después factorizamos obtendremos que:

$$\begin{aligned} CA_P(\lambda) &= \lambda^3 - (3-\alpha-\beta-\gamma)\lambda^2 + ((2-\alpha-\beta)(1-\gamma) + (1-\alpha)(1-\beta))\lambda \\ &\quad - (1-\alpha)(1-\beta)(1-\gamma) - \alpha\beta\gamma \\ &= (\lambda-1)\left[\lambda^2 - (2-\alpha-\beta-\gamma)\lambda + (1-\alpha-\beta-\gamma+\alpha\beta+\alpha\gamma+\beta\gamma)\right] \\ &= (\lambda-1)\left[\lambda - \frac{(2-\alpha-\beta-\gamma)+h}{2}\right]\left[\lambda - \frac{(2-\alpha-\beta-\gamma)-h}{2}\right] \end{aligned}$$

donde $h = \sqrt{\alpha^2+\beta^2+\gamma^2-2(\alpha\beta+\alpha\gamma+\beta\gamma)}$.
Así los eigenvalores son: $\lambda_1 = 1, \lambda_2 = \frac{(2-\alpha-\beta-\gamma)+h}{2}$ *y* $\lambda_3 = \frac{(2-\alpha-\beta-\gamma)-h}{2}$.

Una vez que tenemos los eigenvalores, vamos a encontrar los eigenvectores, iniciaremos con el eigenvector izquierdo u_1 *y el derecho* v_1 *asociados a* $\lambda_1 = 1$*, como ya vimos el eigenvector derecho más sencillo de encontrar es* $v_1^* = \mathbb{1}_3$*, ahora encontremos a un eigenvector izquierdo* u_1^**, para ello debemos resolver lo siguiente:*

$$u^{*t}(P-\lambda_1\mathbb{I}_{3x3}) = 0$$

lo cual es equivalente a resolver, para $\lambda_1 = 1$

$$(u_x,u_y,u_z)\begin{pmatrix} -\alpha & \alpha & 0 \\ 0 & -\beta & \beta \\ \gamma & 0 & -\gamma \end{pmatrix} = (0,0,0)\,.$$

Resolviendo la ecuación anterior obtendremos que una solución es $u_1^* = \left(\frac{\gamma}{\alpha},\frac{\gamma}{\beta},1\right)^t$*, ahora solo tenemos que obligar a que* $<u_1^*,v_1^*>=1$*, para ello calcularemos*

$$\begin{aligned} <u_1^*,v_1^*> &=< \begin{pmatrix} \frac{\gamma}{\alpha} \\ \frac{\gamma}{\beta} \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} > \\ &= \frac{\gamma}{\alpha}+\frac{\gamma}{\beta}+1 \\ &= \frac{\alpha\gamma+\beta\gamma+\alpha\beta}{\alpha\beta} \end{aligned}$$

ahora podemos definir

$$u_1 = \frac{u_1^*}{<u_1^*,v_1^*>}, \qquad\qquad v_1 = \mathbb{1}_3$$

$$u_1 = \frac{1}{\alpha\gamma+\beta\gamma+\alpha\beta}\begin{pmatrix} \beta\gamma \\ \alpha\gamma \\ \alpha\beta \end{pmatrix} \qquad\qquad v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Con este procedimiento hemos asegurado que $<u_1,v_1>=1$*, con* u_1 *y* v_1 *un eigenvector izquierdo y derecho de P respectivamente, además estos tienen entradas estrictamente positivas. Podemos obtener los eigenvectores izquierdos* u_2,u_3 *y los eigenvectores derechos* v_2,v_3 *asociados a* λ_2 *y* λ_3 *respectivamente, de una manera análoga a como lo hicimos con* u_1 *y* v_1 *pero las cuentas son engorrosas cuando usamos los eigenvalores* λ_2 *y* λ_3*, debido a la forma que estos tienen, pero haciendo uso del teorema 2.3 tenemos que*

$$P^n = 1^n(v_1u_1^t)+\lambda_2^n(v_2u_2^t)+\lambda_3^n(v_3u_3^t)$$

$$P^n = \frac{1}{\alpha\gamma+\beta\gamma+\alpha\beta}\begin{pmatrix} \beta\gamma & \alpha\gamma & \alpha\beta \\ \beta\gamma & \alpha\gamma & \alpha\beta \\ \beta\gamma & \alpha\gamma & \alpha\beta \end{pmatrix} + \lambda_2^n(v_2u_2^t)+\lambda_3^n(v_3u_3^t). \tag{2.5}$$

Notemos que el primer sumando de la ecuación (2.5) *es la matriz estacionaria de la cadena de Markov, por lo que lo ideal es que los otros dos sumandos converjan a cero conforme n crece.*

Procederemos a presentar el teorema de Perron-Frobenius, pero antes daremos algunas definiciones que nos serán de utilidad. En el conjunto de las matrices no existe un orden como en los reales, pero es posible definir cuándo una matriz es no negativa o positiva, algunas de las definiciones pueden extenderse a matrices de dimensión *mxn* pero nosotros nos limitaremos a las matrices cuadradas, y más adelante solo a matrices estocásticas.

Definición **2.13.**
Sea $r \in \mathbb{N}$ *y* $A \in \mathscr{M}_{rxr}(\mathbb{R})$, *con* $A = (a_{ij})_{1 \leq i,j \leq r}$ *entonces:*

a. *A será llamada* matriz no negativa *(positiva respectivamente) si* $a_{ij} \geq 0$ *para todo i,* $j \in \{1,...,r\}$ *(*$a_{ij} > 0$ *para todo i,* $j \in \{1,...,r\}$ *respectivamente), y lo denotaremos por* $A \geq 0$ *(*$A > 0$ *respectivamente).*

b. *Diremos que A es* primitiva *si existe* $k \in \mathbb{N}$ *de tal forma que* $A^k > 0$.

Un resultado importante que podemos obtener de Perron-Frobenius es que la convergencia hacia la distribución estacionaria de una cadena de Markov con espacio de estados finito es geométrica, con velocidad de convergencia igual al módulo del segundo eigenvalor más grande, que llamaremos *SLEM* por sus siglas en Inglés.

Teorema **2.4.** *Perron-Frobenius*
Sean $r,n \in \mathbb{N}$ *y* $A \in \mathscr{M}_{rxr}(\mathbb{R})$, *no negativa y primitiva, además sean* $\lambda_1,\ldots,\lambda_r$ *los eigenvalores de A, ordenados de tal forma que* $|\lambda_i| \geq |\lambda_j|$, $1 \leq i < j \leq r$, *entonces* λ_1 *es un número real positivo de multiplicidad 1 tal que* $\lambda_1 > |\lambda_j|$ *para cualquier* $2 \leq j \leq r$, *además existen* u_1 *y* v_1 *un eigevector izquierdo y uno derecho respectivamente con entradas positivas, ambos asociados a* λ_1 *y que satisfacen que* $u_1^t v_1 = 1$, *además*

$$A^n = \lambda_1^n v_1 u_1^t + O(n^{m_2-1} \mid \lambda_2 \mid^n)$$

donde λ_2 *es el SLEM de A y* m_2 *es la multiplicidad de* λ_2.

La demostración del teorema 2.4 será mostrada al final de este capítulo.

Presentaremos un *lema* que afirma sobre el eigenvalor más grande en módulo de las matrices estocásticas y la multiplicidad de este, pues nos será de utilidad a lo largo de este capítulo y en la demostración del teorema de Perron-Frobenius.

Lema **1.**
Sean $r,n \in \mathbb{N}$ *y* $A = (a_{ij})_{1 \leq i,j \leq r}$ *una matriz de transición asociada a una cadena de Markov irreducible y aperiódica con espacio de estados* $E = \{1,\ldots,r\}$, *además sean* $\lambda_1,\ldots,\lambda_r$ *los eigenvalores de A, ordenados de tal forma que* $|\lambda_i| \geq |\lambda_j|$, $1 \leq i < j \leq r$, *entonces* λ_1 *es 1 y su multiplicidad es 1.*

Demostración.
Comencemos demostrando que $\lambda_1 = 1$.
Por la ecuación (2.4) sabemos que $\lambda = 1$ es un eigenvalor de A, por ser estocástica, entonces basta con probar que no existe ningún eigenvalor de A con norma mayor que uno.
Supongamos que existe $\lambda' \in \mathbb{C}$ un eigenvalor de A con $|\lambda'| > 1$ y existe $v \neq 0$ tal que $Av = \lambda' v$, es decir, para toda $i \in E$ tenemos que $\sum_{j\in E} a_{ij} v(j) = \lambda' v(i)$.
Sea $k \in E$ tal que $|v(k)| = \text{máx}_{\ell\in E}\{|v(\ell)|\}$, entonces

$$\begin{aligned}
|v(k)| &< |\lambda'||v(k)| \\
&= |\lambda' v(k)| \\
&= |\sum_{j\in E} a_{kj} v(j)| \\
&\leq \sum_{j\in E} |a_{kj}||v(j)| \\
&\leq \sum_{j\in E} a_{kj} |v(k)| \\
&= |v(k)| \sum_{j\in E} a_{kj} \\
&= |v(k)|
\end{aligned}$$

por ser A estocástica, lo cual es una contradicción.
La contradicción surge del hecho de suponer que existe un eigenvalor de A con norma mayor que uno.
Solo nos fata mostrar que λ_1 es de multiplicidad uno; recordemos que $v_1 = \mathbb{1}_r$ es un eigenvector derecho de A asociado a $\lambda_1 = 1$, entonces basta con mostrar que cualquier eigenvector derecho de A asociado a $\lambda_1 = 1$ es de la forma $\alpha\mathbb{1}_r$, $\alpha \in \mathbb{R}$.
Supongamos que $\lambda_1 = 1$ no es de multiplicidad uno, entonces existe
$v = (v(1), \ldots, v(r)) \in \mathbb{R}^r$ tal que $Av = v$, además de que $v \neq \alpha\mathbb{1}_r$, para todo $\alpha \in \mathbb{R}$, así existen $j, k \in E$ tal que $|v(k)| = \text{máx}_{\ell\in E}\{|v(\ell)|\}$ y $|v(j)| < |v(k)|$ por lo que

$$\begin{aligned}
|v(i)| &\leq |v(k)| && \text{para todo } i \in E, \\
a_{ki}|v(i)| &\leq a_{ki}|v(k)| && \text{para todo } i \in E,
\end{aligned}$$

entonces

$$\begin{aligned}
\sum_{i\in E} a_{ki}|v(i)| &< \sum_{i\in E} a_{ki}|v(k)| && \text{porque } |v(j)| < |v(k)| \\
&\leq |v(k)| && \text{por ser A estocástica} \\
&= |\sum_{i\in E} a_{ki} v(i)| \\
&\leq \sum_{i\in E} a_{ki}|v(i)|
\end{aligned}$$

Hemos caído en una contradicción, la cual deriva de la suposición de que $\lambda_1 = 1$ no es de multiplicidad uno. □

Otras consecuencias que obtenemos del teorema de Perron-Frobenius es que si A es estocástica pero no irreducible con m clases de equivalencia, entonces λ_1 es de multiplicidad m, pero si es irreducible con periodo $d > 1$, entonces existen d distintos eigenvalores λ_j de módulo 1, esto es $\mid \lambda_j \mid= 1$, llamados las d raíces unitarias y los otros eigenvalores tienen módulo estrictamente menor que 1, la demostración de estos resultados se encuentran [Seneta, Cap. 1].

Si regresamos al final del ejemplo 2.3, en la ecuación (2.5), podremos notar que P^n es la suma de la matriz estacionaria más otras dos matrices que convergen a la matriz cero conforme n crece, debido a que los eigenvalores λ_2 y λ_3 tienen módulo menor que 1, con ayuda de Perron-Frobenius podremos demostrar que la convergencia hacia la distribución estacionaria es geométrica con una velocidad de convergencia igual al *SLEM*.
Utilizaremos el siguiente ejemplo para ilustrar el teorema 2.3 y sus corolarios.

Ejemplo **2.4.**
Sea P la matriz de transición de una cadena de Markov homogénea sobre un espacio de estados $E = \{1,2,3\}$.

$$P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{4} & \frac{5}{12} \\ \frac{2}{3} & \frac{1}{4} & \frac{1}{12} \end{pmatrix}$$

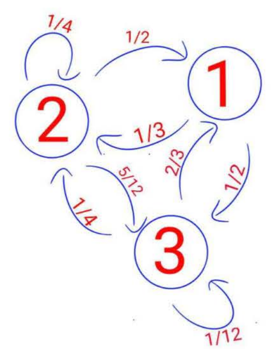

Figura 2.1: Cadena de Markov con 3 estados.

A base de la figura 2.1 podemos observar que P es irreducible y aperiódica. Así por el teorema de Perron-Frobenius y el lema 1 *tenemos que existe $\lambda_1 = 1$ de multiplicidad uno por ser P una matriz estocástica, aperiódica e irreducible, más aún $\mid \lambda_j \mid< 1$ con $j \in \{2,3\}$. Además podemos elegir u_1 y v_1, un eigenvector izquierdo y derecho respectivamente, asociado a λ_1, de tal manera que tanto u_1 como v_1 son positivos y satisfacen que $u_1^t v_1 = 1$, además*

$$P^n = \lambda_1^n v_1 u_1^t + O(\mid \lambda_2 \mid^n).$$

Verifiquemos lo anterior haciendo algunas cuentas. Tenemos que el polinomio característico de P es $CA_P(\lambda) = -\lambda^3 + \frac{1}{3}\lambda^2 + \frac{7}{12}\lambda + \frac{1}{12}$ *cuyas raíces son* $\lambda_1 = 1$, $\lambda_2 = \frac{-1}{2}$ *y* $\lambda_3 = \frac{-1}{6}$*, además con el método hecho en el ejemplo 2.3 para encontrar un eigenvector derecho* v_1 *y un eigenvector izquierdo* u_1 *asociados a* λ_1 *que satisfagan* $v_1^t v_1 = 1$*, tenemos que*

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \text{ y } u_1 = \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{pmatrix},$$

por lo que

$$P^n = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix} + O\left(\frac{1}{2}^n\right).$$

Cabe resaltar que la matriz $v_1 u_1^t$ *es la matriz estacionaria de nuestra cadena de Markov, por lo que la convergencia es geométrica con velocidad relativa de* $\frac{1}{2}$*, el cual es el* SLEM.

Como hemos recalcado en varias ociasiones la velocidad de convergencia de una cadena de Markov hacia su distribución estacionaria es geométrica con velocidad relativa igual al SLEM, por lo que es de interés acotar la distancia entre la matriz de transición a n pasos y la matriz estacionaria en función del SLEM, esto lo haremos en secciones posteriores.

Sección 2.2 Distribución Quasi-estacionaria

Como vimos en el capítulo 1, las cadenas de Markov cuentan con dos tipos de estados, los recurrentes y los transitorios. Durante este capítulo solo hemos trabajado con cadenas de Markov con estados recurrentes. Muchas veces no le damos importancia a los estados transitorios pues en la distribución estacionaria tienen probabilidad cero. En esta sección veremos lo que sucede con los estados transitorios antes de que la cadena llegue a un estado recurrente, veremos que existe una distribución asociada a estos estados a la que llamaremos *Distribución Quasi-estacionaria*.

Antes de empezar con todo el análisis, daremos una definición de matriz subestocástica

Definición **2.14.** *Matriz subestocástica*
Sea $r \in \mathbb{N}$ *y* $A \in \mathscr{M}_{rxr}(\mathbb{R})$*, con* $A = (a_{ij})_{1 \le i,j \le r}$*, no negativa. A es llamada* subestocástica *si existe* $k \in \{1, \ldots, r\}$ *de tal modo que* $\sum_{j=1}^{n} a_{ij} \le 1$*, para todo* $i \in \{1, ..., r\} \setminus \{k\}$ *y* $\sum_{j=1}^{n} a_{kj} < 1$.

Corolario **1.** *del Teorema de Perron-Frobenius*
Bajo las hipótesis de teorema 2.4*, si A es subestocástica, entonces* $\lambda_1 < 1$

Demostración.
Por el Teorema de Perron-Frobenius (2.4), tenemos que existe λ_1 un real positivo de multiplicidad uno tal que $\lambda_1 >| \lambda_j |$ para cualquier λ_j eigenvector derecho de A,
$2 \leq j \leq r$.
Sea $E = \{1, \ldots, r\}$ y $A = (a_{ij})_{i,j \in E}$.
Supongamos que $\lambda_1 > 1$, entonces existe $v \neq 0$ de tal forma que $Av = \lambda_1 v$, esto implica que

$$\sum_{j \in E} a_{ij} v(j) = \lambda_1 v(i) \ \text{ para todo } i \in E$$

entonces para toda $i \in E$

$$\begin{aligned} | v(i) | &< \lambda_1 | v(i) | \\ &= | \lambda_1 v(i) | \\ &= | \sum_{j \in E} a_{ij} v(j) | \\ &\leq \sum_{j \in E} a_{ij} | v(j) | \end{aligned}$$

por lo que

$$| v(i) | < \sum_{j \in E} a_{ij} | v(j) | \quad \text{para todo } i \in E. \tag{2.6}$$

Sea $k \in E$ tal que $| v(k) | = \text{máx}_{\ell \in \mathbb{E}} \{| v(\ell) |\}$, por lo que

$$\begin{aligned} | v(j) | &\leq | v(k) | && \text{para todo } j \in E \\ a_{kj} | v(j) | &\leq a_{kj} | v(k) | && \text{para todo } j \in E \\ \sum_{j \in E} a_{kj} | v(j) | &\leq \sum_{j \in E} a_{kj} | v(k) | \end{aligned}$$

como $| v(k) | < \sum_{j \in E} a_{kj} | v(j) |$ por la ecuación (2.6) y $\sum_{j \in E} a_{kj} | v(k) | \leq | v(k) |$ por ser A subestocástica, por lo que

$$| v(k) | < \sum_{j \in E} a_{kj} | v(j) | \leq | v(k) |$$

lo cual es una contradicción.
La contradicción surge del hecho de suponer que existe un eigenvalor de A mayor que 1.
Supongamos que $\lambda_1 = 1$, por definición de matriz subestocástica existe $i \in E$ de tal forma que $\sum_{j \in E} a_{ij} < 1$, por lo que $\mathbb{1}_r$ no es un eigenvector de A, entonces existe v un eigenvector de A tal que $v \neq \alpha \mathbb{1}_r$, $\alpha \in \mathbb{R}$.
Sean $i, k \in E$ de tal manera que $| v(k) | = \text{máx}_{\ell \in \mathbb{E}} \{| v(\ell) |\}$ y $| v(i) | < | v(k) |$, entonces

$$\begin{aligned} | v(j) | &\leq | v(k) | && \text{para todo } j \in E \\ a_{kj} | v(j) | &\leq a_{kj} | v(k) | && \text{para todo } j \in E \\ \sum_{j \in E} a_{kj} | v(j) | &< \sum_{j \in E} a_{kj} | v(k) | && \text{pues } \ | v(i) | < | v(k) | \end{aligned}$$

como $\mid v(k) \mid \leq \sum_{j\in E} a_{kj} \mid v(j) \mid$ porque v es eigenvector con eigenvalor igual a 1, además $\sum_{j\in E} a_{kj} \mid v(k) \mid \leq \mid v(k) \mid$ por ser A subestocástica, por lo que

$$\mid v(k) \mid \leq \sum_{j\in E} a_{kj} \mid v(j) \mid < \mid v(k) \mid$$

lo cual nos hace caer en una contradicción, esta deriva del suponer que $\lambda_1 = 1$. $\square$

Sea $\{X_n\}_{n\geq 0}$ una cadena de Markov con espacio de estados E finito y matriz de transición P. Supongamos que R es el espacio de estados recurrentes y T el de los transitorios, con T no vacío. Así, sin pérdida de generalidad, si es necesario reacomodamos los estados, podemos descomponer a P con respecto a la partición $R \cup T = E$ y obtener la siguiente matriz

$$P = \begin{pmatrix} D & 0 \\ B & Q \end{pmatrix} \tag{2.7}$$

donde

a. D es la matriz de transición asociada a los recurrentes.

b. 0 es la matriz cero.

c. Q es la matriz de transición asociada a los transitorios.

d. B es la matriz de transición de transitorio a recurrente.

Notemos que Q es una matriz subestocástica y $B \neq 0$, pues de serlo el conjunto T sería cerrado y por lo tanto contiene un estado recurrente ya que E es finito. Supongamos que Q es irreducible y aperiódica, entonces por el Teorema de Perron-Frobenius aplicado a Q (Teorema 2.4)

$$Q^n = \lambda_1^n v_1 u_1^t + O(n^{m_2-1} \mid \lambda_2 \mid^n) \tag{2.8}$$

Con λ_1, v_1, u_1, m_2 y λ_2 como en el Teorema. Como Q es subestocástica, $\lambda_1 \in (0,1)$ y $\mid \lambda_2 \mid < \lambda_1$.

Sea $w = inf\{n \geq 0 \mid X_n \in R\}$, aquí w es el tiempo de entrada a R, como T es finito demostraremos que w es finito casi seguramente, es decir $w < \infty \;\; c.s.$

Demostración.
w es finito casi seguramente si $\mathbb{P}[w < \infty] = 1$, en efecto, consideremos μ el vector de probabilidad sobre E que representa la distribución inicial de la cadena, entonces

$$\begin{aligned} \mathbb{P}[w < \infty] &= \sum_{i\in E} \mathbb{P}[w < \infty | X_0 = i] \mathbb{P}[X_0 = i] \\ &= \sum_{i\in E} \mathbb{P}_i[w < \infty] \mu(i) \end{aligned} \tag{2.9}$$

Para terminar el cálculo anterior, calcularemos por separado $\mathbb{P}_i[w < \infty]$, así

$$\begin{aligned} \mathbb{P}_i[w < \infty] &= 1 - \mathbb{P}_i[w = \infty] \\ &= 1 - \lim_{n\to\infty} \mathbb{P}_i[X_n \in T]. \end{aligned}$$

Por la continuidad de la probabilidad y por la definición de w tenemos que

$$\mathbb{P}_i[w=\infty] = \lim_{n\to\infty} \mathbb{P}_i[X_n \in T].$$

Calculemos este límite, como R es cerrado entonces,

$$\begin{aligned}\mathbb{P}_i[X_n \in T] &= \mathbb{P}_i[X_1 \in T, X_2 \in T, \ldots, X_n \in T] \\ &= \sum_{i_1\in T} \cdots \sum_{i_n \in T} p_{ii_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n}.\end{aligned}$$

Si lo vemos en su forma matricial, tenemos que

$$\mathbb{P}_i[X_n \in T] = \delta_i Q^n \mathbb{1}_{|T|}.$$

Ahora como Q es subestocástica, entonces por la ecuación (2.8) Q^n converge a la matriz cero conforme n tiende a infinito, esto es

$$\lim_{n\to\infty} \mathbb{P}_i[X_n \in T] = \lim_{n\to\infty} \delta_i Q^n \mathbb{1}_{|T|} = 0$$

entonces $\mathbb{P}_i[w<\infty] = 1$.
Por lo tanto, para toda $i \in E$ $\mathbb{P}_i[w<\infty] = 1$, ahora si nos regresamos a la ecuación (2.9) tenemos que

$$\begin{aligned}\mathbb{P}[w<\infty] &= \sum_{i\in E} \mathbb{P}_i[w<\infty]\mu(i) \\ &= \sum_{i\in E} \mu(i) \\ &= 1\end{aligned}$$

$\square$

Una vez que tenemos que el tiempo de entrada a los recurrentes es finito casi seguramente, calcularemos la probabilidad de que empezando en un estado transitorio, sigamos en un estado transitorio a tiempo n.
Sean $i, j \in T$

$$\begin{aligned}\mathbb{P}_i\left[X_n = j \mid w > n\right] &= \frac{\mathbb{P}_i[X_n = j, w > n]}{\mathbb{P}_i[w>n]} \\ &= \frac{\mathbb{P}_i[X_n = j]}{\mathbb{P}_i[X_n \in T]} \\ &= \frac{p_{ij}^n}{\sum_{k\in T} p_{ik}^n}\end{aligned}$$

Pero $p_{ik}^n = \lambda_1^n v_1(i)u_1(k) + O(n^{m_2-1} \mid \lambda_2 \mid^n)$, por lo que

$$\begin{aligned}
\mathbb{P}_i[X_n = j \mid w > n] &= \frac{\lambda_1^n v_1(i)u_1(j) + O(n^{m_2-1} \mid \lambda_2 \mid^n)}{\sum_{k\in T} \lambda_1^n v_1(i)u_1(k) + O(n^{m_2-1} \mid \lambda_2 \mid^n)} \\
&= \left(\frac{\lambda_1^n v_1(i)u_1(j) + O(n^{m_2-1} \mid \lambda_2 \mid^n)}{\sum_{k\in T} \lambda_1^n v_1(i)u_1(k) + O(n^{m_2-1} \mid \lambda_2 \mid^n)} \right) \left(\frac{\frac{1}{\lambda_1^n v_1(i)}}{\frac{1}{\lambda_1^n v_1(i)}} \right) \\
&= \frac{u_1(j) + O\left(n^{m_2-1} |\frac{\lambda_2}{\lambda_1}|^n \right)}{(\sum_{k\in T} u_1(k)) + O\left(n^{m_2-1} |\frac{\lambda_2}{\lambda_1}|^n \right)} \\
&= \frac{u_1(j)}{\sum_{k\in T} u_1(k)} + O\left(n^{m_2-1} \left(|\frac{\lambda_2}{\lambda_1}| \right)^n \right)
\end{aligned}$$

Así, tomando el límite cuando n tiende a infinito, tenemos que

$$\mathbb{P}_i[X_n = j \mid w > n] \xrightarrow{n\to\infty} \frac{u_1(j)}{\sum_{k\in T} u_1(k)}$$

así la distribución de probabilidad $\left\{ \frac{u_1(i)}{\sum_{k\in T} u_1(k)} \right\}_{i\in T}$ es llamada *distribución Quasi-estacionaria* de la cadena relativa a T.

Daremos dos ejemplos para ilustrar la distribución Quasi estacionaria, el primero será un ejemplo numérico mientras que en el otro hablaremos sobre el buscador de Google, sobre el funcionamiento del algoritmo *PageRank*, el algoritmo que ordena los resultados obtenidos en una búsqueda.

Ejemplo **2.5.**
Sea $E = \{0,1,2,3,4\}$ *con probabilidad de transición*

$$p_{xy} = \binom{4}{y} \left(1 - \frac{x}{4}\right)^y \left(\frac{x}{4}\right)^{4-y}, \quad x,y \in \{1,2,3\},$$

$$P_{0x} = \begin{cases} 1 & si \;\; x = 0 \\ 0 & en\ cualquier\ otro\ caso \end{cases}$$

y

$$P_{4x} = \begin{cases} 1 & si \;\; x = 4 \\ 0 & en\ cualquier\ otro\ caso \end{cases} .$$

Es fácil ver que 0 *y* 4 *son estados absorbentes mientras que los demás son transitorios, es decir* $R = \{0,4\}$ *y* $T = \{1,2,3\}$ *por lo que*

$$Q = \begin{pmatrix} \frac{27}{64} & \frac{27}{128} & \frac{3}{64} \\ \frac{1}{4} & \frac{3}{8} & \frac{1}{4} \\ \frac{3}{64} & \frac{27}{128} & \frac{27}{64} \end{pmatrix}$$

Así el polinomio característico de Q *es*

$$CA_P(\lambda) = \lambda^3 - \frac{39}{32}\lambda^2 + \frac{1584}{4096}\lambda - \frac{27}{1024}$$

Cuyas raíces son $\lambda_1 = \frac{3}{4}$, $\lambda_2 = \frac{3}{8}$ *y* $\lambda_3 = \frac{3}{32}$, *con esta información podemos encontrar un eigenvector izquierdo* u_1 *y un eigenvector derecho* v_1 *asociados a* λ_1, *los cuales son*

$$u_1 = \begin{pmatrix} 1 \\ \frac{9}{8} \\ 1 \end{pmatrix} \text{ y } v_1 = \begin{pmatrix} \frac{2}{7} \\ \frac{8}{21} \\ \frac{2}{7} \end{pmatrix},$$

por lo que

$$Q^n = \left(\frac{3}{4}\right)^n \begin{pmatrix} \frac{2}{7} & \frac{8}{21} & \frac{2}{7} \\ \frac{9}{28} & \frac{9}{21} & \frac{9}{28} \\ \frac{2}{7} & \frac{8}{21} & \frac{2}{7} \end{pmatrix} + \circ \left(\frac{3}{8}\right)^n$$

además $\{\frac{8}{25}, \frac{9}{25}, \frac{8}{25}\}$ *es la distribución Quasi-estacionaria relativa a* T.

Ahora seguiremos con el ejemplo del buscador de Google.

Ejemplo **2.6.**
Con la llegada de Internet, se creó toda una revolución en el campo de la tecnología y de la información, la creación del buscador de Google representó una revolución equiparable a la anterior, al ser una herramienta que puso orden en ese mundo de información. El buscador fue diseñado por el matemático Sergei Brin y el informático Lawrence Page en 1998, antes de su diseño habían alrededor de 100 millones de páginas web y solo se atendían 20 millones de consultas al día, hoy en día se atienden más de 200 millones de consultas al día y cada búsqueda indexa varios millones de paginas web. La pregunta que surge es ¿Cómo ordenar los resultados en una búsqueda hecha en el buscador de Google? ¿Y cuál es su relación con las cadenas de Markov?
Primeramente supongamos que tenemos n páginas, digamos, $H_1, H_2, ..., H_n$ *asignaremos una importancia* x_i *a cada página* H_i, $i \in E = \{1, 2, ..., n\}$, *los cuales podrán ser vistos como una probabilidad. Ahora construiremos una matriz* P *la cual tendrá en la entrada* p_{ij} *un uno si existe un enlace de la página* i *a la página* j *y un cero en otro caso, con* $i, j \in E$. *La importancia que le daremos a cada página la tomaremos proporcional a la suma de importancias de las páginas que enlaza, i.e, si la página* H_1 *enlaza a las páginas* H_{12}, H_{24}, H_{200} *y la página* H_2 *con las páginas* H_1, H_3 *entonces*

$$\begin{aligned} x_1 &= \lambda(x_{12} + x_{24} + x_{200}) \\ x_2 &= \lambda(x_1 + x_3) \\ &\vdots \end{aligned}$$

donde λ *es una constante de proporcionalidad fija.*
Viendo lo anterior desde un punto de vista matricial, es equivalente a resolver:

$$Px = \lambda x, \tag{2.10}$$

es decir, nuestro problema se reduce a un problema de eigenvalores y eigenvectores, la solución que buscamos debe ser un vector de entradas no negativas.

Ahora para transformar a P, una matriz de ceros y unos, a una matriz de transición, usaremos una distribución uniforme (discreta) es decir definiremos $P' = (p'_{ij})_{i,j \in E}$ *donde*

$$p'_{ij} = \frac{p_{ij}}{\sum_{k \in E} p_{kj}}$$

Ahora, tenemos una cadena de Markov con espacio de estados $H_1, ..., H_n$ *y matriz de transición entre las páginas* P'*, con lo que la ecuación (*2.10*) se puede reformular de la siguiente forma:*

$$P'x = \lambda x, \tag{2.11}$$

en tal caso nos encontramos en una de las situaciones ya conocidas, pues queremos que x tenga entradas no negativas, entre 0 y 1.
La solución que resuelve ese problema es justo la distribución estacionaria.

Es reconfortante saber que el problema ya lo podemos resolver, pero hay tres cuestiones importantes la primera es ¿qué pasa si una página no enlaza con cualquier otra?, es decir si tenemos una fila de puros ceros, entonces nuestra matriz ya no sería estocástica. La segunda cuestión son los estados transitorios, pues nada de lo que hemos hecho hasta ahora nos dan la certeza de que todos los estados de la cadena de Markov sean recurrentes o si quiera que la matriz de transición P' *sea irreducible. La última cuestión es que si existe la distribución estacionaria, a los estados transitorios les asignará una importancia de cero, es decir, no nos mostrarán esas páginas.*

La resolución del primer problema es sencillo, solo ponemos una distribución uniforme en cada una de las entradas, es decir si nos encontramos en ese caso, colocaremos en todo el renglón correspondiente $\frac{1}{n}$*.*

La solución al segundo y tercer problema es usar un truco para que la matriz de transición se vuelva irreducible, definiremos:

$$P'' = cP' + (1-c)\begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix}(1,...,1) \tag{2.12}$$

donde $p_1, ..., p_n$ *es una distribución de probabilidad y c es una constante entre 0 y 1 (en los cálculos que hace Google,* $c \approx 0.85$*).*
Con ese procedimiento ya tendríamos que la matriz es irreducible, por lo que la estacionaria existe y es única.
Cabe resaltar que el buscador no descarta a los estados transitorios, pues si los descartara habría páginas en la red a las cuales jamás podríamos acceder, el algoritmo PageRank del buscador de Google le asigna una importancia uniforme para que se pueda acceder a estas páginas poco frecuentadas o sin enlace a otra página.

Sección 2.3 Matrices de transición reversibles

En esta sección vamos a introducir un concepto muy importante, que usaremos a lo largo de este trabajo, estamos hablando de las matrices de transición reversibles, definiremos un producto escalar y una norma, estas dependerán de un vector de probabilidad sobre nuestro espacio de estados $E = \{1, ..., r\}$, $r \in \mathbb{N}$, todo esto con el objetivo de poder formar una base ortonormal de $\mathbb{R}^r$ con eigenvectores de P, nuestra matriz de transición. Además veremos cómo transformar a una matriz de transición para que sea una matriz de transición reversible, finalmente construiremos propiedades importantes para la demostración del teorema de Perron-Frobenius que haremos al final de la sección.
Empezaremos definiendo a las matrices de transición reversibles.

Definición **2.15.** *Matriz de trasición reversible*
Sea P una matriz de transición de una cadena de Markov con espacio de estados finito E y π un vector de probabilidad estrictamente positivo sobre E. Decimos que la pareja (P, π) es reversible si para toda $i, j \in E$

$$\pi(i)p_{ij} = \pi(j)p_{ji}$$

En la definición anterior se habla de un vector de probabilidad estrictamente positivo sobre E, que en primera instancia podría ser cualquier vector de probabilidad pero la realidad es que solo puede ser una distribución estacionaria de P, la única en caso de que la cadena sea irreducible, la siguiente proposición refleja lo dicho aquí.

Proposición **2.6.**
Sea P una matriz de transición de una cadena de Markov con espacio de estados finito E y π un vector de probabilidad estrictamente positivo sobre E. Sea (P, π) reversible entonces π es una distribución estacionaria de P.

Demostración.
Recordemos que una distribución de probabilidad π es una distribución estacionaria de P si y sólo si $\pi^t = \pi^t P$ o bien si

$$\pi(i) = \sum_{j \in E} \pi(j)p_{ji} \qquad \text{para toda } i \in E.$$

Sea $i \in E$, entonces

$$\begin{aligned}
(\pi^t P)(i) &= \sum_{j \in E} \pi(j)p_{ji} \\
&= \sum_{j \in E} \pi(i)p_{ij} && \text{por ser } (P, \pi) \text{ reversible} \\
&= \pi(i) \sum_{j \in E} p_{ij} \\
&= \pi(i)
\end{aligned}$$

por lo tanto $\pi^t P = \pi^t$. $\square$

Procederemos a dar las condiciones necesarias y suficientes de reversibilidad, pero para ello es necesario definir dos productos escalares y su norma asociada.

Definición **2.16.** *Producto interior bajo π y norma π*
Sea $r \in \mathbb{N}$, $E = \{1,\dots,r\}$ el espacio de estados de una cadena de Markov y π una distribución de probabilidad estrictamente positiva sobre E. Sean $x,y \in \mathbb{R}^r$, entonces definimos producto interior bajo π y la norma π respectivamente como

$$\langle x,y\rangle_\pi := \sum_{i\in E} x(i)y(i)\pi(i)$$
$$\| x \|_\pi := \sum_{i\in E} x(i)^2\pi(i).$$

Además, definimos el producto interior bajo $\frac{1}{\pi}$ y la norma $\frac{1}{\pi}$ respectivamente como

$$\langle x,y\rangle_{\frac{1}{\pi}} := \sum_{i\in E} x(i)y(i)\frac{1}{\pi(i)}$$
$$\| x \|_{\frac{1}{\pi}} := \sum_{i\in E} x(i)^2\frac{1}{\pi(i)}.$$

Con la definición anterior, un hecho interesante es que π tiene norma $\frac{1}{\pi}$ igual a uno

$$\begin{aligned}\| \pi \|_{\frac{1}{\pi}} &= \sum_{i\in E} \pi(i)^2\frac{1}{\pi(i)} \\ &= \sum_{i\in E} \pi(i) \qquad\qquad \text{pues } \pi(i) > 0 \text{ para toda } i \in E \\ &= 1\end{aligned}$$

Solo nos falta un concepto más para dar las condiciones necesarias y suficientes de reversibilidad que es la definición de *adjunto de una matriz*, de hecho solo nos quedaremos con la noción de auto adjunto pues es una de las propiedades que tienen las matrices reversibles y ocuparemos mucho en esta sección.

Definición **2.17.** *Adjunto de una matriz*
Sea $r \in \mathbb{N}$ y V un espacio vectorial con producto interior de dimensión r. Sean $A,B \in \mathscr{M}_{rxr}$, decimos que B es el adjunto de A si satisface

$$\langle Ax,y\rangle = \langle x,By\rangle \qquad \textit{para toda } x,y \in V.$$

Si $A = B$, diremos que A es auto adjunta.

A partir de este punto supondremos que nuestra matriz de transición P es irreducible y que nuestro espacio de estados es $E = \{1,\dots,r\}$, $r \in \mathbb{N}$, es decir, un espacio de estados finito por lo que podemos garantizar la existencia de una única distribución estacionaria π con sus entradas estrictamente positivas y usarla para la definición 2.16.

Teorema **2.5.** *Condiciones necesarias y suficientes para reversibilidad*
Sea P una matriz de transición de una cadena de Markov irreducible con espacio de estados finito E y sea π su distribución estacionaria. Entonces (P,π) es reversible si y sólo si P es auto adjunta con el producto interior $\langle \cdot,\cdot \rangle_\pi$.

Demostración.
$\Rightarrow|$ Supongamos que la pareja (P,π) es reversible, entonces

$$\begin{aligned}
\langle Px,y\rangle_\pi &= \sum_{i\in E}\left(\sum_{j\in E} p_{ij}x(j)\right)y(i)\pi(i)\\
&= \sum_{i\in E}\left(\sum_{j\in E} p_{ij}\pi(i)x(j)\right)y(i)\\
&= \sum_{i\in E}\sum_{j\in E} p_{ji}\pi(j)x(j)y(i) && \text{usando reversibilidad}\\
&= \sum_{j\in E} x(j)\left(\sum_{i\in E} p_{ji}y(i)\right)\pi(j)\\
&= \langle x,Py\rangle_\pi
\end{aligned}$$

por lo tanto P es auto adjunta.

$\Leftarrow|$ Supongamos que P es auto adjunta, entonces como la propiedad de auto adjunta es válida para todo x en $\mathbb{R}^r$, en particular para δ_i (un vector en $\mathbb{R}^r$ con un 1 en la $i-$ésima entrada y ceros en todas las demás).
Sean $i,j\in\{1,\dots,r\}$, entonces

$$\begin{aligned}
\langle P\delta_i,\delta_j\rangle_\pi &= \sum_{k\in E}\left(\sum_{\ell\in E} p_{k\ell}\delta_i(\ell)\right)\delta_j(k)\pi(k)\\
&= \sum_{k\in E}\left(p_{ki}\delta_j(k)\pi(k)\right)\\
&= p_{ji}\pi(j)
\end{aligned}$$

por lo tanto para toda $i,j\in E$

$$\langle P\delta_i,\delta_j\rangle_\pi = p_{ji}\pi(j). \tag{2.13}$$

Así continuando con esta última ecuación

$$\begin{aligned}
p_{ji}\pi(j) &= \langle P\delta_i,\delta_j\rangle_\pi\\
&= \langle \delta_i,P\delta_j\rangle_\pi && \text{pues } P \text{ es auto adjunto}\\
&= \langle P\delta_j,\delta_i\rangle_\pi && \text{por simetría del producto escalar } \mathbb{R}^r\\
&= p_{ij}\pi(i) && \text{por la ecuación (2.13)}
\end{aligned}$$

por lo tanto (P,π) es reversible. □

En la matrices cuadradas, al multiplicar una matriz por sí misma usamos la notación de exponentes, por ejemplo $A^2 = AA$, estos exponentes usualmente son números naturales, aunque si la matriz A tiene inversa, denotada por A^{-1}, entonces los exponentes pueden ser números enteros, por ejemplo $A^{-2} = A^{-1}A^{-1}$, siguiendo con esta extensión podemos pasar del conjunto de los números enteros como exponentes al conjunto de los números racionales, haremos uso de esta notación y de algunos resultados importantes que serán expuestos en el apéndice A ubicada en la página 75.

Proposición **2.7.**
Sean $r \in \mathbb{N}$, P la matriz de transición de una cadena de Markov irreducible con espacio de estados finito $E = \{1, ..., r\}$ y sea π su distribución estacionaria, además sean $D = Diag(\pi(1), \ldots, \pi(r))$ y $P^* = D^{\frac{1}{2}} P D^{\frac{-1}{2}}$. Si P^* es simétrica entonces (P, π) es reversible.

Demostración.
Proponemos $D^{\frac{1}{2}} = Diag\left(\sqrt{\pi(1)}, \ldots, \sqrt{\pi(r)}\right)$, por lo que

$$P^* = D^{\frac{1}{2}} P D^{\frac{-1}{2}} = \left(\sqrt{\frac{\pi(i)}{\pi(j)}} p_{ij}\right)_{i,j \in E},$$

la matriz P^* está bien definida debido a que P es irreducible esto nos garantiza la existencia de una única distribución estacionaria π con todas sus entradas estrictamente positivas.
Supongamos que P^* es simétrica y que $P^* = (a_{ij})_{i,j \in E}$, entonces

$$\begin{aligned} a_{ij} &= a_{ji} \\ \sqrt{\frac{\pi(i)}{\pi(j)}} p_{ij} &= \sqrt{\frac{\pi(j)}{\pi(i)}} p_{ji} \\ \frac{\pi(i)}{\pi(j)} p_{ij}^2 &= \frac{\pi(j)}{\pi(i)} p_{ji}^2 \\ \pi^2(i) p_{ij}^2 &= \pi^2(j) p_{ji}^2 \\ \pi(i) p_{ij} &= \pi(j) p_{ji} \end{aligned}$$

esta última igualdad es cierta debido a que π es no negativa al igual que las probabilidades.

□

Cuando una matriz como P^* es simétrica sus eigenvalores son reales, además de que el conjunto de eigenvectores derechos es el mismo que el de eigenvectores izquierdos, más aún existe un conjunto de eigenvectores que forman una base ortonormal de $\mathbb{R}^r$ con respecto a la *norma euclideana*, la demostración de estás afirmaciones las podemos encontrar en el apéndice A de este trabajo.

Proposición **2.8.**
Bajo las hipótesis y la notación de la proposición 2.7, sean λ_i, $1 \leq i \leq r$, los eigenvalores de P^* y w_i eigenvectores de P^* asociados a λ_i respectivamente, $1 \leq i \leq r$, tal que $\{w_i\}_{i\in E}$ es una base ortonormal, definimos u_i, $v_i \in \mathbb{R}^r$ vectores que satisfacen respectivamente que $w_i = D^{\frac{-1}{2}} u_i$ y $w_i = D^{\frac{1}{2}} v_i$, entonces

a. P y P^* tienen los mismos eigenvalores.

b. u_i (v_i respectivamente) es un eigenvector izquierdo (derecho respectivamente) de P asociado al eigenvalor λ_i, $1 \leq i \leq r$.

Demostración.

a. $\underline{\Rightarrow|}$ Sea λ un eigenvalor de P, entonces existe $v \in \mathbb{R}^r$, $v \neq 0$ de tal forma que $Pv = \lambda v$, entonces como $D^{\frac{1}{2}}$ es una matriz diagonal distinta a la matriz cero y $v \neq 0$ entonces $D^{\frac{1}{2}} v \neq 0$ por lo que

$$\begin{aligned}
P^* &= D^{\frac{1}{2}} P D^{\frac{-1}{2}} && \\
P^* D^{\frac{1}{2}} &= D^{\frac{1}{2}} P && \text{multiplicando } D^{\frac{1}{2}} \text{ por la derecha} \\
P^* D^{\frac{1}{2}} v &= D^{\frac{1}{2}} P v && \text{multiplicando } v \text{ por la derecha} \\
P^* D^{\frac{1}{2}} v &= D^{\frac{1}{2}} \lambda v && v \text{ es eigenvector de } P \\
P^* \left(D^{\frac{1}{2}} v\right) &= \lambda \left(D^{\frac{1}{2}} v\right) && \\
P^* w &= \lambda w &&
\end{aligned}$$

por lo tanto λ es una eigenvalor de P^*.
$\underline{\Leftarrow|}$ Sea λ un eigenvalor de P^*, entonces existe $u \in \mathbb{R}^r$, $u \neq 0$ de tal forma que $P^* u = \lambda u$, entonces

$$\begin{aligned}
P^* &= D^{\frac{1}{2}} P D^{\frac{-1}{2}} && \\
D^{\frac{-1}{2}} P^* &= P D^{\frac{-1}{2}} && \text{multiplicando } D^{\frac{-1}{2}} \text{ por la izquierda} \\
D^{\frac{-1}{2}} P^* u &= P D^{\frac{-1}{2}} u && \text{multiplicando } u \text{ por la derecha} \\
D^{\frac{-1}{2}} \lambda u &= P D^{\frac{-1}{2}} u && u \text{ es eigenvector de } P^* \\
\lambda \left(D^{\frac{-1}{2}} u\right) &= P \left(D^{\frac{-1}{2}} u\right) && \\
\lambda w &= P w &&
\end{aligned}$$

así λ es una eigenvalor de P.

b. Sea $i \in E$ y λ_i un eigenvalor de P^* y sea w_i el eigenvector asociado a λ_i, es decir $P^* w_i = \lambda_i w_i$ y $w_i^t P^* = \lambda_i w_i^t$ debido a que w_i es tanto un eigenvector derecho como izquierdo, entonces

$$\begin{aligned}
P^*w_i &= D^{\frac{1}{2}}PD^{\frac{-1}{2}}w_i && \text{por definición de } P^* \\
P^*w_i &= D^{\frac{1}{2}}PD^{\frac{-1}{2}}D^{\frac{1}{2}}v_i && \text{por definición de } w_i \\
P^*w_i &= D^{\frac{1}{2}}Pv_i && \text{simplificando } D^{\frac{-1}{2}}D^{\frac{1}{2}} \\
\lambda_i w_i &= D^{\frac{1}{2}}Pv_i && w_i \text{ es eigenvector de } P^* \\
\lambda_i D^{\frac{1}{2}}v_i &= D^{\frac{1}{2}}Pv_i && \text{por definición de } w_i \\
D^{\frac{1}{2}}(\lambda_i v_i) &= D^{\frac{1}{2}}Pv_i && \\
D^{\frac{-1}{2}}D^{\frac{1}{2}}(\lambda_i v_i) &= D^{\frac{-1}{2}}D^{\frac{1}{2}}Pv_i && \text{multiplicando } D^{\frac{-1}{2}} \text{ por la izquierda} \\
\lambda_i v_i &= Pv_i
\end{aligned}$$

por lo que v_i es un eigenvector derecho de P asociado a λ_i.
Ahora veamos que u_i es eigenvector izquierdo de P asociado a λ_i, como

$$w_i^t = \left(D^{\frac{-1}{2}}u_i\right)^t = u_i^t D^{\frac{-1}{2}}$$

debido a que $D^{\frac{-1}{2}}$ es una matriz diagonal, entonces

$$\begin{aligned}
w_i^t P^* &= w_i^t D^{\frac{1}{2}}PD^{\frac{-1}{2}} && \text{por definición de } P^* \\
w_i^t P^* &= u_i^t D^{\frac{-1}{2}}D^{\frac{1}{2}}PD^{\frac{-1}{2}} && \text{por definición de } w_i \\
w_i^t P^* &= u_i^t PD^{\frac{-1}{2}} && \text{simplificando } D^{\frac{-1}{2}}D^{\frac{1}{2}} \\
\lambda_i w_i^t &= u_i^t PD^{\frac{-1}{2}} && w_i \text{ es eigenvector de } P^* \\
\lambda_i u_i^t D^{\frac{-1}{2}} &= u_i^t PD^{\frac{-1}{2}} && \text{por definición de } w_i \\
\lambda_i u_i^t D^{\frac{-1}{2}}D^{\frac{1}{2}} &= u_i^t PD^{\frac{-1}{2}}D^{\frac{1}{2}} && \text{multiplicando } D^{\frac{1}{2}} \text{ por la derecha} \\
\lambda_i u_i^t &= u_i^t P
\end{aligned}$$

por lo que u_i es un eigenvector izquierdo de P asociado a λ_i.

□

A partir de este punto trabajaremos bajo las hipótesis de la proposición 2.7 y con la notación de la proposición 2.8.
Una propiedad que tiene la matriz D es que podemos definir el producto interior bajo π en términos del producto matricial, la siguiente proposición nos será útil para construir otras propiedades con las que podremos demostrar el teorema de *Perron-Frobenius*.

Proposición **2.9.**
Bajo las hipótesis de la proposición 2.7 *y la notación de la proposición* 2.8*, sean* $x, y \in \mathbb{R}^r$*, entonces*

$$x^t Dy = \langle x, y \rangle_\pi.$$

Demostración.

$$\begin{aligned}\left(x^t D\right)(i) &= \sum_{j\in E} x(j)\pi(i)\delta_i(j)\\ &= x(i)\pi(i)\end{aligned}$$

Con lo que

$$\begin{aligned}x^t D y &= \sum_{i\in E}\left(x^t D\right)(i) y(i)\\ &= \sum_{i\in E} x(i)\pi(i)y(i)\\ &= \langle x, y\rangle_\pi\end{aligned}$$

□

Recordemos que la ortonormalidad de $\{w_i\}_{i\in E}$ es respecto a la norma euclideana pero el conjunto $\{v_i\}_{i\in E}$ y $\{u_i\}_{i\in E}$ también son ortonormales, pero respecto a la norma π y la norma $\frac{1}{\pi}$ respectivamente, es decir,

$$\begin{aligned}\langle v_i, v_j\rangle_\pi &= \delta_i(j) & \text{con } w_i &= D^{\frac{1}{2}} v_i\\ \langle u_i, u_j\rangle_{\frac{1}{\pi}} &= \delta_i(j) & \text{con } w_i &= D^{\frac{-1}{2}} u_i.\end{aligned}$$

Demostración.
Sean $i, j \in E$, entonces

$$\begin{aligned}\langle v_i, v_j\rangle_\pi &= \sum_{k\in E} v_i(k)v_j(k)\pi(k)\\ &= \sum_{k\in E}(\sqrt{\pi(k)}v_i(k))(\sqrt{\pi(k)}v_j(k)) && \text{debido a que } \pi(k) > 0\\ &= \sum_{k\in E} w_i(k)w_j(k) && \sqrt{\pi(k)}v_i(k) = \left(D^{\frac{1}{2}}v_i\right)(k)\\ &= w_i\cdot w_j\\ &= \delta_i(j)\end{aligned}$$

por lo tanto $\langle v_i, v_j\rangle_\pi = \delta_i(j)$. Análogamente $\langle u_i, u_j\rangle_{\frac{1}{\pi}} = \delta_i(j)$. □

Por la definición de u_i y de v_i tenemos que

$$u_i = Dv_i. \tag{2.14}$$

Como $V = \{v_i\}_{i\in E}$ es una base ortonormal de $\mathbb{R}^r$ con respecto a la norma π, entonces cualquier vector $x \in \mathbb{R}^r$ puede ser expresado como una combinación lineal de los elementos de V, es decir, existen $\alpha_1, \ldots, \alpha_r \in \mathbb{R}$ de tal forma que $x = \sum_{i\in E}\alpha_i v_i$, entonces sea $j \in E$ y calculamos

$$\begin{aligned}
\langle x, v_j \rangle_\pi &= \langle \sum_{i \in E} \alpha_i v_i, v_j \rangle_\pi \\
&= \sum_{i \in E} \alpha_i \langle v_i, v_j \rangle_\pi && \text{por la linalidad del producto interior} \\
&= \sum_{i \in E} \alpha_i \delta_i(j) && V \text{ es un conjunto ortonormal} \\
&= \alpha_j
\end{aligned}$$

por lo que $\langle x, v_j \rangle_\pi = \alpha_j$, entonces para toda $x \in \mathbb{R}^r$ tenemos que

$$x = \sum_{i \in E} \langle x, v_i \rangle_\pi v_i \tag{2.15}$$

de manera análoga,

$$x^t = \sum_{i \in E} \langle x, u_i \rangle_{\frac{1}{\pi}} u_i^t. \tag{2.16}$$

Como v_j es un eigenvector derecho de P, se cumple que $P^n v_j = \lambda_j^n v_j$ para toda $n \in \mathbb{N}$ y para toda $j \in E$, entonces para toda $x \in \mathbb{R}^r$ y para toda $n \in \mathbb{N}$

$$\begin{aligned}
P^n x &= P^n \left(\sum_{i \in E} \langle x, v_i \rangle_\pi v_i \right) && \text{por la ecuación (2.15)} \\
&= \sum_{i \in E} \langle x, v_i \rangle_\pi P^n v_i \\
&= \sum_{i \in E} \langle x, v_i \rangle_\pi \lambda_i^n v_i
\end{aligned}$$

por lo tanto, para toda $x \in \mathbb{R}^r$ y para toda $n \in \mathbb{N}$

$$P^n x = \sum_{i \in E} \lambda_i^n \langle x, v_i \rangle_\pi v_i \tag{2.17}$$

de manera análoga, para toda $x \in \mathbb{R}^r$ y para toda $n \in \mathbb{N}$

$$x^t P^n = \sum_{i \in E} \lambda_i^n \langle x, u_i \rangle_{\frac{1}{\pi}} u_i. \tag{2.18}$$

Daremos otra caracterización de la ecuación (2.17) en términos de los eigenvectores izquierdos y derechos de P

$$\begin{aligned}
P^n x &= \sum_{i \in E} \lambda_i^n \langle x, v_i \rangle_\pi v_i && \text{Por la ecuación (2.17)} \\
&= \sum_{i \in E} \lambda_i^n \left(x^t D v_i \right) v_i && \text{Por la proposición 2.9} \\
&= \sum_{i \in E} \lambda_i^n v_i \left((D v_i)^t x \right) \\
&= \sum_{i \in E} \lambda_i^n v_i u_i^t x && \text{Por la ecuación (2.14)}
\end{aligned}$$

Por lo tanto

$$P^n x = \sum_{i \in E} \lambda_i^n v_i u_i^t x. \tag{2.19}$$

Ahora daremos una pequeña demostración del Teorema de Perron-Frobenius, la demostración que se presenta es parecida a la hecha por Pierre Brémaud en [Bremaud, Cap. 6] la prueba se limitará a las matrices de transición reversibles, una demostración más general del teorema se encuentra en [Seneta, Cap. 1].

Demostración. Teorema 2.4 Perron-Frobenius
Sean $r, n \in \mathbb{N}$, $r > 0$ y sea $A \in \mathscr{M}_{rxr}(\mathbb{R})$, $A = (a_{ij})_{1 \leq i,j \leq r}$, la matriz de transición reversible (Definición 2.15) asociada a una cadena de Markov irreducible y aperiódica sobre un espacio de estados $E = \{1, \ldots, r\}$.
Como A es irreducible sobre un espacio de estados finito entonces existe una única distribución estacionaria π con entradas estrictamente positivas, así usando la distribución estacionaria de A podemos definir la norma π (Definición 2.16).
Sean $\lambda_1, \ldots, \lambda_r$ los eigenvalores de A ordenados de tal que $\mid \lambda_i \mid \geq \mid \lambda_j \mid$, $1 \leq i < j \leq r$ y $V = \{v_1, \ldots, v_r\}$ una base ortonormal de eigenvectores derechos de A (ortonormales con respecto a la norma π). Entonces por la propiedad 2.9 y la ecuación (2.17), para todo $x \in \mathbb{R}^r$

$$A^n x = \sum_{\ell=1}^{r} \lambda_\ell^n \langle x, v_\ell \rangle_\pi v_\ell$$

$$A^n x - \langle x, v_1 \rangle_\pi v_1 = \sum_{\ell=2}^{r} \lambda_\ell^n \langle x, v_\ell \rangle_\pi v_\ell \qquad \text{pues } \lambda_1 = 1.$$

Esta última igualdad es válida en particular para $x = \delta_k$, con $1 \leq k \leq r$, por lo que al sustituir $x = \delta_k$ obtenemos el siguiente vector

$$A^n \delta_k - [v_1(k)\pi(k)]\, v_1 = \sum_{\ell=2}^{r} [\lambda_\ell^n v_\ell(k)\pi(k)]\, v_\ell \quad \text{para todo } k \in E. \tag{2.20}$$

Así, sea $i \in E$ y $a_{ik}^{(n)}$ la entrada (i,k) de la matriz A^n, entonces al obtener la entrada i del vector de la ecuación (2.20) tenemos que

$$a_{ik}^{(n)} - v_1(k)\pi(k)v_1(i) = \sum_{\ell=2}^{r} \lambda_\ell^n v_\ell(k)\pi(k)v_\ell(i) \quad \text{para todo } k \in E. \tag{2.21}$$

además por el Lema 1, sabemos que los eigenvectores derechos de A asociados a λ_1 son de la forma $\alpha \mathbb{1}_r$, $\alpha \in \mathbb{R}$, pero el único que tienen norma π igual a uno y entradas positivas es $\mathbb{1}_r$, por lo que $v_1(k)v_1(i) = 1$, para todo $k \in E$, entonces la ecuación (2.21) es equivalente a

$$a_{ik}^{(n)} - \pi(k) = \sum_{\ell=2}^{r} \lambda_\ell^n v_\ell(k)v_\ell(i)\pi(k) \quad \text{para todo } k \in E. \tag{2.22}$$

Ahora, denotaremos por A_ℓ^n al $\ell-$ésimo renglón de la matriz A^n, entonces

$$\begin{aligned}
\| A_i^n - \pi \|_1 &= \sum_{k\in E} | a_{ik}^{(n)} - \pi(k) | \\
&= \sum_{k\in E} | \sum_{j=2}^{r} \lambda_j^n v_j(k) v_j(i) \pi(k) | && \text{Por la ecuación (2.22)} \\
&\leq \sum_{k\in E} \sum_{j=2}^{r} | \lambda_j |^n | v_j(k) || v_j(i) || \pi(k) | && \text{por desigualdad del triángulo} \\
&= \sum_{j=2}^{r} \sum_{k\in E} | \lambda_j |^n | v_j(k) || v_j(i) || \pi(k) | && \text{intercambiando las sumas} \\
&= \sum_{j=2}^{r} \left[| \lambda_j |^n | v_j(i) | \left(\sum_{k\in E} | v_j(k) || \pi(k) | \right) \right]
\end{aligned}$$

antes de seguir acotaremos al término que tenemos entre paréntesis

$$\begin{aligned}
\sum_{k\in E} | v_j(k) || \pi(k) | &= \sum_{k\in E} | v_j(k) | \pi(k) \\
&= \sum_{k\in E} \left(| v_j(k) | \sqrt{\pi(k)} \right) \left(\sqrt{\pi(k)} \right) \\
&\leq \left(\sum_{k\in E} v_j^2(k) \pi(k) \right)^{\frac{1}{2}} \left(\sum_{k\in E} \pi(k) \right)^{\frac{1}{2}} && \text{por la desigualdad de Cauchy-Schwarz} \\
&= 1
\end{aligned}$$

por lo que

$$\sum_{k\in E} | v_j(k) || \pi(k) | \leq 1 \tag{2.23}$$

entonces

$$\begin{aligned}
\| A_i^n - \pi \|_1 &\leq \sum_{j=2}^{r} | \lambda_j |^n | v_j(i) | && \text{sustituyendo la ecuación (2.23)} \\
&\leq \sum_{j=2}^{r} | \lambda_2 |^n | v_j(i) | \\
&\leq r | \lambda_2 |^n \left(\sup_{2\leq j\leq r} | v_j(i) | \right)
\end{aligned}$$

por lo que

$$\| A_i^n - \pi \|_1 \leq r \left(\sup_{2\leq j\leq r} | v_j(i) | \right) | \lambda_2 |^n . \tag{2.24}$$

En esta última ecuación podemos ver la importancia que tiene el *SLEM* para la velocidad de convergencia de A_i^n, $n \in \mathbb{N}$, a su estacionaria. En los siguientes capítulos nos enfocaremos en encontrar distintas cotas para $\| A_i^n - \pi \|_1$ en la ecuación (2.24) y después acotar el valor de λ_2, nuestro *SLEM*.

Ahora, sea u_1 un eigenvector izquierdo de A asociado a $\lambda_1 = 1$ con norma $\frac{1}{\pi}$ igual a uno; como λ_1 es de multiplicidad uno entonces el único eigenvector izquierdo de A con norma $\frac{1}{\pi}$ igual a uno y entradas positivas es π, además $v_1 u_1^t = \Pi$, la matriz estacionaria.

Además, como $|\lambda_2| < 1$ entonces la ecuación (2.24) tiende a cero conforme n tiende a infinito, es decir tenemos que $\| A_i^n - \pi \|_1$ está acotado por una función que depende de $|\lambda_2|^n$ y que además tiende a cero, entonces

$$\begin{aligned} \| A^n - \Pi \| &= O(|\lambda_2|^n) \\ \| A^n - \lambda_1 v_1 u_1^t \| &= O(|\lambda_2|^n) \\ A^n &= \lambda_1 v_1 u_1^t + \varepsilon \end{aligned}$$

donde $\| \varepsilon \| = O(|\lambda_2|^n)$. □

Capítulo 3. Velocidad de convergencia

En el capítulo anterior encontramos una velocidad de convergencia de la matriz de transición a n pasos hacia su estacionaria en función de una constante y su *SLEM*. En este capítulo nos enfocaremos en encontrar otras cotas, construiremos cotas relativamente más sencillas, primero sobre las matrices de transición reversibles y luego para las matrices de transición no reversibles.
Se demostrarán tres teoremas, dos para matrices de transición reversibles, los teoremas 3.6 y 3.7, y un teorema para las matrices de transición *no* reversibles, el teorema 3.8, al final haremos una comparación entre el teorema 3.7 y el teorema 3.8.

Sección 3.1 Caso reversible

Para esta primera sección supondremos que la matriz de transición P es reversible, crearemos dos cotas nuevas una por cada norma definida en el capítulo anterior, así tendremos el teorema (3.6) al cual llamaremos la cota π y el teorema 3.7 denominado la cota $\frac{1}{\pi}$.
Para lograr esto tenemos que definir una distancia entre dos distribuciones de probabilidad sobre el mismo espacio de estados, además de algunas propiedades que nos serán de utilidad para la construcción de las cotas.

Cota π

En este caso, vamos a definir la distancia en variación entre dos distribuciones de probabilidad α y β sobre un mismo conjunto finito E, que denotaremos por $d_v(\alpha,\beta)$.

Definición **3.18.** *Distancia en variación*
Sean $r \in \mathbb{N}$ y α, β dos distribuciones de probabilidad sobre un mismo conjunto finito $E = \{1,\dots,r\}$. Definimos la distancia en variación entre α y β como

$$d_v(\alpha,\beta) := \frac{1}{2}\sum_{i\in E} \mid \alpha(i) - \beta(i) \mid \tag{3.25}$$

Las propiedades que derivan de la definición anterior son mostradas en el *apéndice B* ubicada en la página 79 de este trabajo.
Vamos a definir la varianza de un punto $x \in \mathbb{R}^r$ y después mostraremos dos propiedades que nos serán de utilidad.

Definición **3.19.** *Varianza*
Sea $r \in \mathbb{N}$, $E = \{1, 2, \ldots, r\}$ el espacio de estados de una cadena de Markov y π una distribución de probabilidad estrictamente positiva sobre E. Sea $x \in \mathbb{R}^r$, definimos la varianza de x bajo pi como

$$Var_\pi(x) = \| x \|_\pi^2 - \langle x, \mathbb{1}_r \rangle_\pi^2$$

Sean $r \in \mathbb{N}$, $x \in \mathbb{R}^r$ y π una distribución de probabilidad estrictamente positiva, además sea X una variable aleatoria tal que $\mathbb{P}[X = x(i)] = \pi(i)$. Calcularemos la esperanza, el segundo momento y la varianza de X y compararemos los resultados con los términos que definen la varianza en la definición 3.19.

Primero calcularemos la esperanza de X

$$\begin{aligned} \mathbb{E}[X] &= \sum_{i \in E} x(i)\pi(i) \\ &= \sum_{i \in E} x(i)1\pi(i) \\ &= \langle x, \mathbb{1}_r \rangle_\pi. \end{aligned}$$

Calculando el segundo momento de X obtenemos

$$\begin{aligned} \mathbb{E}[X^2] &= \sum_{i \in E} x(i)^2\pi(i) \\ &= \langle x, x \rangle_\pi \\ &= \| x \|_\pi^2, \end{aligned}$$

por lo que la varianza de X es

$$\begin{aligned} Var(X) &= \mathbb{E}[X^2] - \mathbb{E}^2[X] \\ &= \| x \|_\pi^2 - \langle x, \mathbb{1}_r \rangle_\pi^2 \\ &= Var_\pi(x). \end{aligned}$$

Proposición **3.10.**
Sean $r \in \mathbb{N}$ y $A \in \mathscr{M}_{rxr}(\mathbb{R})$ la matriz de transición reversible (Definición 2.15) asociada a una cadena de Markov irreducible y aperiódica sobre un espacio de estados $E = \{1, \ldots, r\}$, además sea π su distribución estacionaria. Sea $x \in \mathbb{R}^r$, entonces

$$Var_\pi(x) = \sum_{j=2}^{r} | \langle x, v_j \rangle_\pi |^2$$

Demostración.
Como A es irreducible sobre un espacio de estados finito, entonces existe una única

distribución estacionaria π con entradas estrictamente positivas, así usando la distribución estacionaria de A podemos definir la norma π (Definición 2.16).
Sea $V = \{v_1, \ldots, v_r\}$ una base ortonormal de eigenvectores derechos de A (ortonormal con respecto a la norma π) y sea $x \in \mathbb{R}^r$, entonces

$$\begin{aligned}
\| x \|_\pi^2 &= \langle x, x \rangle_\pi \\
\| x \|_\pi^2 &= \langle \sum_{i \in E} \langle x, v_i \rangle_\pi v_i, \sum_{j \in E} \langle x, v_j \rangle_\pi v_j \rangle_\pi && \text{por la ecuación (2.15)} \\
\| x \|_\pi^2 &= \sum_{i \in E} \sum_{j \in E} \langle x, v_i \rangle_\pi \langle x, v_j \rangle_\pi \langle v_i, v_j \rangle_\pi \\
\| x \|_\pi^2 &= \sum_{i \in E} | \langle x, v_i \rangle_\pi |^2 \\
\| x \|_\pi^2 &= \langle x, \mathbb{1}_r \rangle_\pi^2 + \sum_{i=2}^{r} | \langle x, v_i \rangle_\pi |^2 \\
\| x \|_\pi^2 - \langle x, \mathbb{1}_r \rangle_\pi^2 &= \sum_{i=2}^{r} | \langle x, v_i \rangle_\pi |^2 \\
Var_\pi(x) &= \sum_{i=2}^{r} | \langle x, v_i \rangle_\pi |^2
\end{aligned}$$

□

Proposición **3.11.**
Bajo las hipótesis de la proposición 3.10*. Sea* $x \in \mathbb{R}^r$*, si* $\sup_{i \in E}\{| x(i) |\} \leq 1$ *entonces* $Var_\pi(x) \leq 1$.

Demostración.
Por definición sabemos que la varianza de x es

$$\begin{aligned}
Var_\pi(x) &= \| x \|_\pi^2 - \langle x, \mathbb{1}_r \rangle_\pi^2 \\
&\leq \| x \|_\pi^2 \\
&= \sum_{k \in E} | x(k) |^2 \pi(k) \\
&\leq \sum_{k \in E} \pi(k) && \text{porque } \sup_{i \in E}\{| x(i) |\} \leq 1. \\
&= 1
\end{aligned}$$

□

Teorema **3.6.** *Cota* π
Sean $r \in \mathbb{N}$*,* $r > 0$*, sea* $P \in \mathscr{M}_{rxr}(\mathbb{R})$*,* $P = (p_{ij})_{1 \leq i,j \leq r}$*, la matriz de transición reversible (Definición* 2.15*) asociada a una cadena de Markov irreducible y aperiódica sobre un espacio de estados* $E = \{1, \ldots, r\}$ *y sea* π *su distribución estacionaria. Entonces para todo* $n \in \mathbb{N}$ *con* $n \geq 1$ *y para todo* $i \in E$

$$d_v^2 \left(\delta_i^t P^n, \pi \right) \leq \left(\frac{p_{ii}^{(2)}}{\pi(i)} \right) \rho^{2n-2} \tag{3.26}$$

con ρ el SLEM de P y $p_{ij}^{(2)}$ la entrada (i,j), $i,j \in E$, de la matriz P^2.

Demostración.
Como P es irreducible sobre un espacio de estados finito entonces existe una única distribución estacionaria π con entradas estrictamente positivas, así usando la distribución estacionaria de P podemos definir la norma π (Definición 2.16).
Sean $\lambda_1,\dots,\lambda_r$ los eigenvalores de P ordenados de modo que $\mid \lambda_i \mid \geq \mid \lambda_j \mid$ para todo $1 \leq i < j \leq r$ y $V = \{v_1,\dots,v_r\}$ una base ortonormal de eigenvectores derechos de A (ortonormal con respecto a la norma π).
Sea $x \in \mathbb{R}^r$, entonces por la ecuación (2.17) tenemos que

$$
\begin{aligned}
P^n x &= \sum_{j\in E} \lambda_j^n \langle x, v_j\rangle_\pi v_j \\
P^n x &= \lambda_1^n \langle x, v_1\rangle_\pi v_1 + \sum_{j=2}^{r} \lambda_j^n \langle x, v_j\rangle_\pi v_j \\
P^n x - \langle x, \mathbb{1}_r\rangle_\pi \mathbb{1}_r &= \sum_{j=2}^{r} \lambda_j^n \langle x, v_j\rangle_\pi v_j \\
\| P^n x - \langle x, \mathbb{1}\rangle_\pi \mathbb{1} \|_\pi^2 &= \| \sum_{j=2}^{r} \lambda_j^n \langle x, v_j\rangle_\pi v_j \|_\pi^2 && \text{tomando norma al cuadrado en ambos lados} \\
&= \langle \sum_{j=2}^{r} \lambda_j^n \langle x, v_j\rangle_\pi v_j, \sum_{k=2}^{r} \lambda_k^n \langle x, v_k\rangle_\pi v_k \rangle_\pi \\
&= \sum_{j=2}^{r}\sum_{k=2}^{r} \lambda_j^n \lambda_k^n \langle x, v_j\rangle_\pi \langle x, v_k\rangle_\pi \langle v_j, v_k\rangle_\pi \\
&= \sum_{j=2}^{r} \mid \lambda_j \mid^{2n} \mid \langle x, v_j\rangle_\pi \mid^2 \\
&\leq \rho^{2n} \sum_{j=2}^{r} \mid \langle x, v_j\rangle_\pi \mid^2 \\
&= \rho^{2n} Var_\pi(x) && \text{por la proposición 3.10}
\end{aligned}
$$

entonces

$$
\| P^n x - \langle x, \mathbb{1}\rangle_\pi \mathbb{1} \|_\pi^2 \leq \rho^{2n} Var_\pi(x) \tag{3.27}
$$

por otro lado, sea $i \in E$, entonces

$$\begin{aligned}
\mid \sum_{j\in E} p_{ij}x(j) \mid^2 &= \mid \sum_{j\in E} p_{ji}\frac{\pi(j)}{\pi(i)}x(j) \mid^2 && \text{ya que } P \text{ es reversible} \\
&\leq \left(\sum_{j\in E} x^2(j)\pi(j)\right)\left(\sum_{j\in E}\left(\frac{p_{ji}}{\pi(i)}\right)^2\pi(j)\right) && \text{por la desigualdad de Cauchy-Schwarz} \\
&= \| x \|_\pi^2 \left(\sum_{j\in E}\frac{p_{ji}}{\pi(i)}\frac{p_{ji}\pi(j)}{\pi(i)}\right) \\
&= \| x \|_\pi^2 \left(\sum_{j\in E}\frac{p_{ij}p_{ji}}{\pi(i)}\right) && \text{usando reversibilidad} \\
&= \| x \|_\pi^2 \frac{1}{\pi(i)} p_{ii}^{(2)}
\end{aligned}$$

por lo que para todo $x \in \mathbb{R}^r$

$$\mid \sum_{j\in E} p_{ij}x(j) \mid^2 \leq \| x \|_\pi^2 \frac{1}{\pi(i)} p_{ii}^{(2)} \tag{3.28}$$

Tomaremos una x particular la cual es $x = P^{n-1}y - \langle y, \mathbb{1}_r\rangle_\pi \mathbb{1}_r$, para alguna y que definiremos más adelante, entonces la entrada ℓ de este vector es $x(\ell) = \sum_{k\in E} p_{\ell k}^{n-1}y(k) - \pi(k)y(k)$, entonces por la propiedad de Chapman-Kolmogorov

$$\mid \sum_{k\in E}(p_{ik}^n y(k) - \pi(k)y(k)) \mid^2 = \mid \sum_{k\in E}\left(\sum_{j\in E} p_{ij}p_{jk}^{n-1}y(k)\right) - \sum_{k\in E}(\pi(k)y(k)) \mid^2$$

ahora sí, multiplicando por uno

$$\begin{aligned}
\mid \sum_{k\in E}(p_{ik}^n y(k) - \pi(k)y(k)) \mid^2 &= \mid \sum_{k\in E}\left(\sum_{j\in E} p_{ij}p_{jk}^{n-1}y(k)\right) - \sum_{k\in E}\left(\sum_{j\in E}(p_{ij}\pi(k)y(k))\right) \mid^2 \\
&= \mid \sum_{k\in E}\left(\sum_{j\in E} p_{ij}p_{jk}^{n-1}y(k) - p_{ij}\pi(k)y(k)\right) \mid^2 \\
&= \mid \sum_{j\in E}\sum_{k\in E}\left(p_{ij}p_{jk}^{n-1}y(k) - p_{ij}\pi(k)y(k)\right) \mid^2 \\
&= \mid \sum_{j\in E} p_{ij}\left(\sum_{k\in E}(p_{jk}^{n-1}y(k) - \pi(k)y(k))\right) \mid^2 \\
&= \mid \sum_{j\in E} p_{ij}x(j) \mid^2
\end{aligned}$$

usando la ecuación (3.28) tenemos que

$$\mid \sum_{k\in E}(p_{ik}^n y(k) - \pi(k)y(k)) \mid^2 \leq \frac{p_{ii}^{(2)}}{\pi(i)} \| x \|_\pi^2 \tag{3.29}$$

además

$$\begin{aligned} \| x \|_\pi^2 &= \| P^{n-1}y - \langle y, \mathbb{1}_r\rangle_\pi \mathbb{1}_r \|_\pi^2 \\ &\leq \rho^{2n-2} Var_\pi(y) && \text{usando la ecuación (3.27)} \end{aligned}$$

por lo que la ecuación (3.29) lo podemos reescribir como

$$| \sum_{k\in E} (p_{ik}^n y(k) - \pi(k)y(k)) |^2 \leq \frac{p_{ii}^{(2)}}{\pi(i)} Var_\pi(y)\rho^{2n-2} \tag{3.30}$$

por otro lado, como $\sup_{i\in E}(| y(i) |) = 1$ entonces

$$d_v(\delta_i^t P^n, \pi) = \frac{1}{2}\sup\left(| \sum_{k\in E}(p_{ik}^n y(k) - \pi(k)y(k)) | \right)$$

por lo que

$$\begin{aligned} d_v^2(\delta_i^t P^n, \pi) &= \frac{1}{4}\sup{}^2\left(| \sum_{k\in E}(p_{ik}^n y(k) - \pi(k)y(k)) | \ ; \ \sup_{i\in E}(| y(i) |) = 1\right) \\ &= \frac{1}{4}\sup\left(| \sum_{k\in E}(p_{ik}^n y(k) - \pi(k)y(k)) |^2 \ ; \ \sup_{i\in E}(| y(i) |) = 1\right) \\ &\leq \frac{1}{4}\sup\left(\frac{p_{ii}^{(2)}}{\pi(i)} Var_\pi(y)\rho^{2n-2} \ ; \ \sup_{i\in E}(| y(i) |) = 1\right) \\ &\leq \frac{1}{4}\frac{p_{ii}^{(2)}}{\pi(i)}\rho^{2n-2}\sup\left(Var_\pi(y) \ ; \ \sup_{i\in E}(| y(i) |) = 1\right) \\ &\leq \frac{p_{ii}^{(2)}}{\pi(i)}\rho^{2n-2} && \text{por la proposición 3.11} \end{aligned}$$

□

Cota $\frac{1}{\pi}$

Definición **3.20.** *El contraste χ^2*
Sean α y β dos distribuciones de probabilidad sobre un mismo conjunto finito E, además sea $A = \{i \in E \mid \beta(i) \neq 0\}$, entonces definimos el contraste χ^2 *de α respecto a β como*

$$\chi^2(\alpha;\beta) := \sum_{i\in A}\frac{(\alpha(i) - \beta(i))^2}{\beta(i)} \tag{3.31}$$

Cuando tenemos una cadena de Markov irreducible y aperiódica podemos obtener una única distribución estacionaria π con sus entradas estrictamente positivas, así podemos expresar al contraste χ^2 de α con respecto a π por definición en términos de la norma $\frac{1}{\pi}$ de la siguiente manera

$$\chi^2(\alpha;\pi) = \| \alpha - \pi \|_{\frac{1}{\pi}}^2 \tag{3.32}$$

Proposición **3.12.**
Sean $r \in \mathbb{N}$, α y β dos distribuciones de probabilidad sobre un mismo conjunto finito $E = \{1, ..., r\}$, entonces

$$4d_v^2(\alpha,\beta) \leq \chi^2(\alpha;\beta) \tag{3.33}$$

Demostración.
Por definición de distancia en variación entre α y β tenemos que

$$d_v(\alpha,\beta) = \frac{1}{2}\sum_{i\in E} \mid \alpha(i) - \beta(i) \mid$$

por lo que

$$\begin{aligned} 4d_v^2(\alpha,\beta) &= \left(\sum_{i\in E} \mid \alpha(i) - \beta(i) \mid\right)^2 \\ &= \left(\sum_{i\in E} \mid \frac{\alpha(i)}{\beta(i)} - 1 \mid \beta(i)\right)^2 \end{aligned}$$

así usando la desigualdad de Cauchy-Schwarz

$$\begin{aligned} 4d_v^2(\alpha,\beta) &\leq \left(\sum_{i\in E} \mid \frac{\alpha(i)}{\beta(i)} - 1 \mid^2 \beta(i)\right)\left(\cancelto{1}{\sum_{i\in E}\beta(i)}\right) \\ &= \left(\sum_{i\in E} \frac{(\alpha(i)-\beta(i))^2}{\beta(i)}\right) \\ &= \chi^2(\alpha;\beta) \end{aligned}$$

□

A continuación enunciaremos el segundo teorema que nos da una cota para la velocidad de convergencia de la matriz de transición a n pasos hacia su distribución estacionaria utilizando la norma $\frac{1}{\pi}$.

Teorema **3.7.** *Cota* $\frac{1}{\pi}$
Sean $r \in \mathbb{N}$ y P la matriz de transición de una cadena de Markov irredducible y aperiódica, sobre un espacio de estados finito $E = \{1, 2, \cdots, r\}$ y sea π su distribución estacionaria. Sean ρ el SLEM de P y $\cdot(A) = \sum_{i\in A} \cdot(i)$. Además, supongamos que (P, π) es reversible, entonces para cualquier distribución inicial μ sobre E y para cualquier número natural $n \geq 1$

$$\parallel \mu^t P^n - \pi^t \parallel_{\frac{1}{\pi}} \leq \rho^n \parallel \mu^t - \pi^t \parallel_{\frac{1}{\pi}} \tag{3.34}$$

además, para cualquier estado inicial $i \in E$ y para cualquier subconjunto A de E

$$\mid \delta_i^t P^n(A) - \pi^t(A) \mid \leq \left(\frac{1-\pi(i)}{\pi(i)}\right)^{\frac{1}{2}} \text{mín}\left(\pi(A)^{\frac{1}{2}}, \frac{1}{2}\right)\rho^n \tag{3.35}$$

particularmente por el Lema 3 *tenemos que*

$$4d_v^2(\delta_i^t P^n, \pi) \leq \left(\frac{1-\pi(i)}{\pi(i)} \right) \rho^{2n}. \tag{3.36}$$

Esta última ecuación aparece de esta forma particular para poder relacionarlo con el teorema 3.8.

Demostración.
Como P es irreducible sobre un espacio de estados finito entonces existe una única distribución estacionaria π con entradas estrictamente positivas, así usando la distribución estacionaria de P podemos definir la norma $\frac{1}{\pi}$ (Definición 2.16).
Sean $\lambda_1, \ldots, \lambda_r$ los eigenvalores de P ordenados de modo que $\mid \lambda_i \mid \geq \mid \lambda_j \mid$ para todo $1 \leq i < j \leq r$ y $U = \{u_1, \ldots, u_r\}$ una base ortonormal de eigenvectores izquierdos de P (ortonormal con respecto a la norma $\frac{1}{\pi}$).
Recordemos que $u_1 = \pi$, así calculando

$$\begin{aligned}
\langle \mu - \pi, u_1 \rangle_{\frac{1}{\pi}} &= \langle \mu - \pi, \pi \rangle_{\frac{1}{\pi}} \\
&= \sum_{i \in E} \frac{(\mu(i) - \pi(i))(\cancel{\pi(i)})}{\cancel{\pi(i)}} \\
&= \sum_{i \in E} \mu(i) - \sum_{i \in E} \pi(i) \\
&= 0
\end{aligned}$$

Ahora, comenzaremos con la demostración de la ecuación (3.34), para ello calcularemos la siguiente norma

$$\| (\mu - \pi)^t P^n \|_{\frac{1}{\pi}}^2 = \langle (\mu - \pi)^t P^n, (\mu - \pi)^t P^n \rangle_{\frac{1}{\pi}}$$

entonces usando la ecuación (2.18) que nos permite representar el producto entre un vector $x \in \mathbb{R}^r$ y P^n, en términos del producto interior $\frac{1}{\pi}$, los eigenvalores y los eigenvectores izquierdos de P, así $(\mu - \pi)^t P^n = \sum_{i \in E} \lambda_i^n \langle \mu - \pi, u_i \rangle_{\frac{1}{\pi}} u_i$ por lo que

$$\begin{aligned}
\| (\mu - \pi)^t P^n \|_{\frac{1}{\pi}}^2 &= \langle \sum_{i \in E} \lambda_i^n \langle \mu - \pi, u_i \rangle_{\frac{1}{\pi}} u_i, \sum_{j \in E} \lambda_j^n \langle \mu - \pi, u_j \rangle_{\frac{1}{\pi}} u_j \rangle_{\frac{1}{\pi}} \\
&= \sum_{i \in E} \sum_{j \in E} \lambda_i^n \lambda_j^n \langle \mu - \pi, u_i \rangle_{\frac{1}{\pi}} \langle \mu - \pi, u_j \rangle_{\frac{1}{\pi}} \langle u_i, u_j \rangle_{\frac{1}{\pi}} \\
&= \sum_{i \in E} \lambda_i^{2n} \langle \mu - \pi, u_i \rangle_{\frac{1}{\pi}}^2 \\
&\leq \rho^{2n} \sum_{i \in E} \langle \mu - \pi, u_i \rangle_{\frac{1}{\pi}}^2 \\
&= \rho^{2n} \sum_{i \in E} \sum_{j \in E} \langle \mu - \pi, u_i \rangle_{\frac{1}{\pi}} \langle \mu - \pi, u_j \rangle_{\frac{1}{\pi}} \langle u_i, u_j \rangle_{\frac{1}{\pi}} \\
&= \rho^{2n} \langle \sum_{i \in E} \langle \mu - \pi, u_i \rangle_{\frac{1}{\pi}} u_i, \sum_{j \in E} \langle \mu - \pi, u_j \rangle_{\frac{1}{\pi}} u_j \rangle_{\frac{1}{\pi}}
\end{aligned}$$

así usando la ecuación (2.16) que nos permite representar a un vector $x \in \mathbb{R}^r$ en términos del producto interior $\frac{1}{\pi}$ y los eigenvectores izquierdos de P, podemos saber que $\mu - \pi = \sum_{i \in E} \langle \mu - \pi, u_i \rangle_{\frac{1}{\pi}} u_i$ entonces

$$\begin{aligned} \| (\mu - \pi)^t P^n \|^2_{\frac{1}{\pi}} &= \rho^{2n} \langle \mu - \pi, \mu - \pi \rangle_{\frac{1}{\pi}} \\ &= \rho^{2n} \| \mu - \pi \|^2_{\frac{1}{\pi}} \end{aligned}$$

por lo que

$$\| (\mu - \pi)^t P^n \|^2_{\frac{1}{\pi}} \leq \rho^{2n} \| \mu - \pi \|^2_{\frac{1}{\pi}} \tag{3.37}$$

por otro lado

$$\begin{aligned} \| (\mu - \pi)^t P^n \|_{\frac{1}{\pi}} &= \| \mu^t P^n - \pi^t P^n \|_{\frac{1}{\pi}} \\ &= \| \mu^t P^n - \pi^t \|_{\frac{1}{\pi}} \qquad \text{porque } \pi \text{ es la distribución estacionaria de P} \end{aligned}$$

así de la ecuación (3.37) obtenemos

$$\| \mu^t P^n - \pi^t \|^2_{\frac{1}{\pi}} \leq \rho^{2n} \| \mu - \pi \|^2_{\frac{1}{\pi}}$$

por lo tanto

$$\| \mu^t P^n - \pi^t \|_{\frac{1}{\pi}} \leq \rho^{n} \| \mu - \pi \|_{\frac{1}{\pi}}$$

Con esto hemos terminado de verificar la primera afirmación del teorema.
Ahora probaremos la ecuación (3.35), para simplificar las cuentas denotaremos por

$\delta_i^t P^n = \mu_n^t$. Sea $A \subset E$, entonces

$$
\begin{aligned}
\mid \mu_n^t(A) - \pi^t(A) \mid^2 &= \mid \sum_{i\in A} \left(\frac{\mu_n(i)}{\pi(i)} - 1\right)\pi(i) \mid^2 \\
&\leq \left(\sum_{i\in A}\left(\frac{\mu_n(i)}{\pi(i)} - 1\right)^2 \pi(i)\right)\pi(A) \quad \text{por la desigualdad de Cauchy-Schwarz} \\
&\leq \left(\sum_{i\in E}\left(\frac{\mu_n(i)}{\pi(i)} - 1\right)^2 \pi(i)\right)\pi(A) \\
&= \left(\sum_{i\in E}\left(\frac{(\mu_n(i)-\pi(i))^2}{\pi(i)}\right)\right)\pi(A) \\
&= \| \mu_n^t - \pi^t \|_{\frac{1}{\pi}}^2 \ \pi(A) \\
&= \| \delta_i^t P^n - \pi^t \|_{\frac{1}{\pi}}^2 \ \pi(A) \\
&\leq \| \delta_i - \pi \|_{\frac{1}{\pi}}^2 \ \rho^{2n}\pi(A) \quad \text{usando la ecuación (3.34)} \\
&= \frac{(1-\pi(i))^2}{\pi(i)}\rho^{2n}\pi(A) \\
&\leq \frac{1-\pi(i)}{\pi(i)}\rho^{2n}\pi(A)
\end{aligned}
$$

por lo tanto

$$
\mid \mu_n^t(A) - \pi^t(A) \mid \leq \left(\frac{1-\pi(i)}{\pi(i)}\right)^{\frac{1}{2}} \rho^n \pi^{\frac{1}{2}}(A) \tag{3.38}
$$

por otro lado, por el Lema 3 tenemos que

$$
\begin{aligned}
\mid \mu_n^t(A) - \pi^t(A) \mid^2 &\leq d_v^2(\mu_n, \pi) \\
&\leq \frac{1}{4}\chi^2(\mu_n; \pi) \quad \text{por la ecuación (3.33)}
\end{aligned}
$$

además por la ecuación (3.32) tenemos que

$$
\begin{aligned}
\chi^2(\mu_n; \pi) &= \| \delta_i^t P^n - \pi^t \|_{\frac{1}{\pi}}^2 \\
&\leq \rho^{2n} \| \delta_i - \pi \|_{\frac{1}{\pi}}^2 \quad \text{por la ecuación (3.37)} \\
&\leq \rho^{2n}\left(\frac{1-\pi(i)}{\pi(i)}\right)
\end{aligned}
$$

por lo que

$$
\mid \mu_n^t(A) - \pi^t(A) \mid \leq \frac{1}{2}\rho^n \left(\frac{1-\pi(i)}{\pi(i)}\right)^{\frac{1}{2}}. \tag{3.39}
$$

Así usando las ecuaciones (3.38) y (3.39) podemos concluir

$$| \mu_n^t(A) - \pi^t(A) | \leq \rho^n \left(\frac{1-\pi(i)}{\pi(i)} \right)^{\frac{1}{2}} \text{mín}\{\pi^{\frac{1}{2}}(A), \frac{1}{2}\}. \tag{3.40}$$

Con esto hemos terminado de verificar la segunda afirmación del teorema.
Finalmente tomando supremos sobre todos los subconjuntos A de E y elevando al cuadrado de ambos lados de la ecuación (3.40) obtenemos

$$4d_v^2(\delta_i^t P^n, \pi) \leq \left(\frac{1-\pi(i)}{\pi(i)} \right) \rho^{2n}. \tag{3.41}$$

□

Sección 3.2 Caso no reversible

En la sección anterior encontramos distintas cotas para la convergencia de la matriz de transición P a n pasos hacia su distribución estacionaria π, pero asumimos que P era una matriz reversible y esta propiedad de reversibilidad no la poseen todas la matrices de transición por lo que en esta sección nos enfocaremos a encontrar cotas similares pero para las matrices de transición **no reversibles**.

Para el caso de las matrices no reversibles podemos obtener cotas en función del *SLEM* de una matriz reversible que está relacionada con la matriz original, así que primero construiremos esta matriz reversible, para ello será necesario la siguiente definición.

Definición **3.21.**
Sea $r \in \mathbb{N}$ y $P \in \mathscr{M}_{rxr}(\mathbb{R})$ una matriz de transición irreducible y aperiódica sobre un espacio de estados finito E con distribución estacionaria π. Definimos una nueva matriz $\widehat{P} = (\widehat{p}_{ij})_{i,j \in E}$ de tal forma que

$$\widehat{p}_{ij} = \frac{\pi(j)p_{ji}}{\pi(i)}$$

o en su forma matricial

$$\widehat{P} = D^{-1}P^tD$$

donde $D = Diag(\pi)$

Daremos una propiedad de $\widehat{P}$ que nos será de gran utilidad para la demostración del teorema 3.8.

Lema **2.**
Sea $r \in \mathbb{N}$, P una matriz de transición irreducible y aperiódica sobre un espacio de estados finito $E = \{1, ..., r\}$ y π la distribución estacionaria de P. Además, sea $\widehat{P}$ como en la definición 3.21*, entonces $\widehat{P}$ es el adjunto de P, es decir, para cualesquiera $x, y \in \mathbb{R}^r$*

$$\langle \widehat{P}x, y \rangle_\pi = \langle x, Py \rangle_\pi \tag{3.42}$$

Demostración.
Sean $x, y \in \mathbb{R}^r$, entonces

$$\begin{aligned}
\langle \widehat{P}x, y\rangle_\pi &= \sum_{i\in E}\left(\sum_{j\in E}\widehat{p}_{ij}x(j)\right)y(i)\pi(i)\\
&= \sum_{i\in E}\sum_{j\in E}\frac{\pi(j)p_{ji}}{\cancel{\pi(i)}}x(j)y(i)\cancel{\pi(i)}\\
&= \sum_{j\in E}x(j)\left(\sum_{i\in E}p_{ji}y(i)\right)\pi(j)\\
&= \langle x, Py\rangle_\pi
\end{aligned}$$

$\square$

Ahora mostraremos una forma de reversibilización para nuestra matriz de transición P, con M una matriz reversible en función de P, podremos crear una velocidad de convergencia para matrices no reversibles.

Proposición **3.13.**
Sea P una matriz de transición irreducible y aperiódica sobre un espacio de estados finito E con distribución estacionaria π, si definimos $M = M(P) = P\widehat{P}$, entonces (M, π) es reversible.

Demostración.
Como P es una matriz irreducible entonces podemos garantizar la existencia de una única distribución estacionaria π cuyas entradas son estrictamente positivas, por lo que solo nos falta verificar que M es una matriz estocástica que cumple la condición de reversibilidad respecto a π, así primero mostraremos que M es estocástica, sea $i \in E$, entonces

$$\begin{aligned}
\sum_{j\in E} m_{ij} &= \sum_{j\in E}\sum_{k\in E} p_{ik}\widehat{p}_{kj}\\
&= \sum_{j\in E}\sum_{k\in E}\frac{p_{ik}p_{jk}\pi(j)}{\pi(k)}\\
&= \sum_{k\in E}\sum_{j\in E}\frac{p_{ik}p_{jk}\pi(j)}{\pi(k)}\\
&= \sum_{k\in E}\left(\frac{p_{ik}}{\pi(k)}\sum_{j\in E}p_{jk}\pi(j)\right)\\
&= \sum_{k\in E}\frac{p_{ik}}{\cancel{\pi(k)}}\cancel{\pi(k)}\\
&= 1
\end{aligned}$$

por lo que M es una matriz estocástica, únicamente nos falta verificar la condición de reversibilidad, sea $i,j \in E$, entonces

$$\begin{aligned}
\pi(i)m_{ij} &= \pi(i)\sum_{k\in E} p_{ik}\widehat{p}_{kj} \\
&= \pi(i)\sum_{k\in E} \frac{p_{ik}p_{jk}\pi(j)}{\pi(k)} \\
&= \pi(j)\sum_{k\in E} \frac{p_{jk}p_{ik}\pi(i)}{\pi(k)} \\
&= \pi(j)\sum_{k\in E} p_{jk}\widehat{p}_{ki} \\
&= \pi(j)m_{ji}
\end{aligned}$$

por lo tanto (M,π) es reversible. □

Ahora que sabemos que (M,π) es reversible, podemos hacer un análisis sobre M y sus eigenvalores puesto que hasta este punto no sabemos cómo son estos, por el capítulo anterior vimos que si $M^* = D^{\frac{1}{2}}MD^{\frac{-1}{2}}$ es una matriz simétrica entonces sus eigenvalores son reales, además estos coinciden con los eigenvalores de M, entonces demostremos que efectivamente M^* es simétrica.

Demostración.
Para probar que M^* es simétrica solo es necesario verificar que existe una matriz A tal que $M^* = AA^t$, en efecto

$$\begin{aligned}
M^* &= D^{\frac{1}{2}}MD^{\frac{-1}{2}} \\
&= D^{\frac{1}{2}}P\widehat{P}D^{\frac{-1}{2}} \\
&= D^{\frac{1}{2}}PD^{-1}P^tDD^{\frac{-1}{2}} \\
&= D^{\frac{1}{2}}PD^{-1}P^tD^{\frac{1}{2}} \\
&= D^{\frac{1}{2}}PD^{\frac{-1}{2}}D^{\frac{-1}{2}}P^tD^{\frac{1}{2}} \\
&= \left(D^{\frac{1}{2}}PD^{\frac{-1}{2}}\right)\left(D^{\frac{1}{2}}PD^{\frac{-1}{2}}\right)^t
\end{aligned}$$

□

Como M es una matriz estocástica entonces sus eigenvalores están contenidos en el intervalo $[-1,1]$, de hecho todos ellos se encuentran en el intervalo $[0,1]$ esto debido a que M^* es una matriz no negativa definida, recordemos que una matriz $A \in \mathscr{M}_{rxr}(\mathbb{R})$ es una matriz no negativa definida si y sólo si para cualquier vector $x \in \mathbb{R}^r$, $x^tAx \geq 0$.

Demostraremos que M^* es una matriz no negativa definida

Demostración.
Sea $x \in \mathbb{R}^r$, si definimos $\overline{P} = D^{\frac{1}{2}} P D^{\frac{-1}{2}}$, entonces $M^* = \overline{P}\overline{P}^t$

$$\begin{aligned} x^t M^* x &= x^t \overline{P}\overline{P}^t x \\ &= \left(x^t \overline{P}\right)\left(x^t \overline{P}\right)^t \\ &= \| \left(x^t \overline{P}\right) \| \geq 0 \end{aligned}$$

□

La matriz M es una forma multiplicativa de reversibilización, existen otro tipo de reversibilización como por ejemplo la aditiva. La reversibilización aditiva de P es definiendo la matriz reversible $A = A(P) = \frac{1}{2}(P + \widehat{P})$, aunque es una forma más sencilla de obtener una matriz reversible la rersibilización multiplicativa es más útil para el teorema cota de contraste.

Antes de enunciar el teorema 3.8 cota de contraste, daremos una definición, que ocuparemos en la demostración del teorema, las propiedades que derivan de esta definición las podemos encontrar en el Apéndice C ubicado en la página 83 de este trabajo.

Definición **3.22.** *La forma de Dirichlet*
Sea $r \in \mathbb{N}$, P la matriz de transición de una cadena de Markov sobre un espacio de estados finito $E = \{1, \ldots, r\}$ y π una distribución estacionaria de P, además sea (P, π) reversible, entonces definimos la forma de Dirichlet *como:*

$$\varepsilon_\pi(x, x) := \langle (I - P)x, x \rangle_\pi$$

Teorema **3.8.** *Cota de contraste*
Sea $r \in \mathbb{N}$ y P la matriz de transición de una cadena de Markov irreducible y aperiódica, sobre un espacio de estados finito $E = \{1, ..., r\}$ y sea π su distribución estacionaria. Además, sea $M(P)$ la matriz reversible asociada a P, sea $\gamma = \gamma(M)$ el SLEM *de M, entonces para cualquier distribución inicial μ sobre E y para todo $n \in \mathbb{N}$ tenemos que*

$$4 d_v^2(\mu^t P^n, \pi^t) \leq \gamma^n \chi^2(\mu; \pi) \tag{3.43}$$

Demostración.
Sea $x \in \mathbb{R}^r$, además definimos $\widehat{x} = x - \langle x, \mathbb{1}_r \rangle_\pi \mathbb{1}_r$, entonces

$$\begin{aligned} \| \widehat{x} \|_\pi^2 &= \langle \widehat{x}, \widehat{x} \rangle_\pi \\ &= \langle x - \langle x, \mathbb{1}_r \rangle_\pi \mathbb{1}_r, x - \langle x, \mathbb{1}_r \rangle_\pi \mathbb{1}_r \rangle_\pi \\ &= \langle x, x \rangle_\pi - 2 \langle x, \mathbb{1}_r \rangle_\pi^2 + \langle x, \mathbb{1}_r \rangle_\pi^2 \cancelto{1}{\langle \mathbb{1}_r, \mathbb{1}_r \rangle_\pi} \\ &= \| x \|_\pi^2 - \langle x, \mathbb{1}_r \rangle_\pi^2 \\ &= Var_\pi(x) \end{aligned}$$

por otro lado, sea $z \in \mathbb{R}^r$ y $\varepsilon_\pi(z,z)$ como en la definición 3.22 aplicado a la matriz reversible M, así

$$\begin{aligned}\varepsilon_\pi(z,z) &= \langle(\mathbb{I}_{rxr} - M)z, z\rangle_\pi\\ &= \langle z - Mz, z\rangle_\pi\\ &= \langle z,z\rangle_\pi - \langle Mz,z\rangle_\pi\\ &= \| z \|_\pi^2 - \langle Mz,z\rangle_\pi\end{aligned}$$

por lo que para toda $z \in \mathbb{R}^r$

$$\varepsilon_\pi(z,z) = \| z \|_\pi^2 - \langle Mz,z\rangle_\pi \tag{3.44}$$

así, calculando

$$\begin{aligned}\langle M\widehat{x},\widehat{x}\rangle_\pi &= \| \widehat{x} \|_\pi^2 - \varepsilon_\pi(\widehat{x},\widehat{x}) && \text{por la ecuación (3.44)}\\ &= \| \widehat{x} \|_\pi^2 - \varepsilon_\pi(x,x) && \text{por la proposición 6.20}\\ &= \| \widehat{x} \|_\pi^2 - \| x \|_\pi^2 + \langle Mx,x\rangle_\pi\\ &= \langle P\widehat{P}x,x\rangle_\pi + Var_\pi(x) - \| x \|_\pi^2\\ &= \langle \widehat{P}x,\widehat{P}x\rangle_\pi - (\langle x,\mathbb{1}_r\rangle_\pi)^2 && P \text{ es adjunta de } \widehat{P}\\ &= \| \widehat{P}x \|_\pi^2 - (\langle x,P\mathbb{1}_r\rangle_\pi)^2\\ &= \| \widehat{P}x \|_\pi^2 - \left(\langle \widehat{P}x,\mathbb{1}_r\rangle_\pi\right)^2\\ &= Var_\pi(\widehat{P}x)\end{aligned}$$

por lo que

$$\begin{aligned}\varepsilon_\pi(x,x) &= \varepsilon_\pi(\widehat{x},\widehat{x}) && \text{por la proposición 6.20}\\ &= \| \widehat{x} \|_\pi^2 - \langle M\widehat{x},\widehat{x}\rangle_\pi && \text{por la ecuación (3.44)}\\ &= Var_\pi(x) - Var_\pi(\widehat{P}x)\end{aligned}$$

por lo tanto

$$Var_\pi(x) = Var_\pi(\widehat{P}x) - \varepsilon_\pi(x,x) \tag{3.45}$$

Ahora denotaremos por $\chi_n^2 = \chi^2(\mu^t P^n;\pi)$ y por $\rho_n(i) = \frac{(\mu^t P^n)(i)}{\pi(i)}$, aunque a ρ_n lo podemos representar en su forma matricial de la siguiente forma $\rho_n = (\mu^t P^n D^{-1})^t$, ahora calculemos su varianza π, para ello primero calcularemos su norma al cuadrado y su producto interno π con el vector de unos, así

$$\begin{aligned}\| \rho_n \|_\pi^2 &= \sum_{i\in E} \rho(i)^2\pi(i)\\ &= \sum_{i\in E}\left(\frac{(\mu^t P^n)(i)}{\pi(i)}\right)^2 \pi(i)\\ &= \sum_{i\in E}\frac{(\mu^t P^n)^2(i)}{\pi(i)}\end{aligned}$$

además,

$$\begin{aligned}\langle \rho_n, \mathbb{1}_r \rangle_\pi &= \sum_{i\in E} \rho(i)\pi(i) \\ &= \sum_{i\in E} (\mu^t P^n)(i) \\ &= 1\end{aligned}$$

por lo que

$$\begin{aligned}Var_\pi(\rho_n) &= \sum_{i\in E} \frac{(\mu^t P^n)^2(i)}{\pi(i)} - 1 \\ &= \sum_{i\in E} \frac{(\mu^t P^n)^2(i)}{\pi(i)} - 2 + 1 \\ &= \sum_{i\in E} \frac{(\mu^t P^n)^2(i)}{\pi(i)} - 2\sum_{i\in E}(\mu^t P^n)(i) + \sum_{i\in E}\pi(i) \\ &= \sum_{i\in E} \frac{(\mu^t P^n)^2(i) - 2(\mu^t P^n)(i)\pi(i) + \pi(i)^2}{\pi(i)} \\ &= \sum_{i\in E} \frac{((\mu^t P^n)(i) - \pi(i))^2}{\pi(i)} \\ &= \chi^2(\mu^t P^n; \pi)\end{aligned}$$

$$Var_\pi(\rho_n) = \chi_n^2 \tag{3.46}$$

por otro lado veamos lo que vale $\widehat{P}\rho_n$

$$\begin{aligned}\widehat{P}\rho_n &= \widehat{P}\left(\mu^t P^n D^{-1}\right)^t \\ &= \widehat{P} D^{-1} (P^t)^n \mu \\ &= D^{-1} P^t D D^{-1} (P^t)^n \mu \\ &= D^{-1} (P^t)^{n+1} \mu \\ &= \left(\mu^t P^{n+1} D^{-1}\right)^t. \\ &= \rho_{n+1}\end{aligned}$$

Así usando la ecuación (3.45) para ρ_n obtenemos

$$\begin{aligned}Var_\pi(\rho_n) &= Var_\pi(\widehat{P}\rho_n) + \varepsilon_\pi(\rho_n, \rho_n) \\ Var_\pi(\rho_n) &= Var_\pi(\rho_{n+1}) + \varepsilon_\pi(\rho_n, \rho_n) \\ \chi_n^2 &= \chi_{n+1}^2 + \varepsilon_\pi(\rho_n, \rho_n) && \text{por la ecuación (3.46).}\end{aligned}$$

Ahora, ρ_n es un vector constante cuando para algún $n \in \mathbb{N}$, $\mu^t P^n = \pi^t$, con lo que para toda $m \geq n$, $\chi_m^2 = 0$ al igual que $\varepsilon_\pi(\rho_m, \rho_m)$, por lo que nos enfocaremos en los $n \in \mathbb{N}$

tales que ρ_n no es constante.
Usaremos el lema 4, ubicado en el apéndice C en la página 84, este lema nos da una cota inferior de la forma de Dirichlet de x en términos del SLEM y la varianza de x.
Como ρ_n no es constante, entonces por el lema 4 tenemos que

$$
\begin{aligned}
\varepsilon_\pi(\rho_n,\rho_n) &\geq (1-\gamma)Var_\pi(\rho_n) \\
\varepsilon_\pi(\rho_n,\rho_n) &\geq (1-\gamma)\chi_n^2 && \text{por la ecuación (3.46)} \\
\varepsilon_\pi(\rho_n,\rho_n) &\geq \chi_n^2 - \gamma\chi_n^2 \\
\gamma\chi_n^2 &\geq \chi_n^2 - \varepsilon_\pi(\rho_n,\rho_n) \\
\gamma\chi_n^2 &\geq \chi_{n+1}^2
\end{aligned}
$$

por lo que $\chi_n^2 \leq \gamma\chi_{n-1}^2 \leq \ldots \leq \gamma^n\chi_0^2$.
Usando la proposición 3.33, que nos da una relación entre la distancia en variación d_v y el contraste χ^2, esto es

$$4d_v^2(\mu^t P^n,\pi) \leq \chi_n^2$$

por lo tanto

$$4d_v^2(\mu^t P^n,\pi) \leq \gamma^n\chi^2(\mu;\pi)$$

□

Recordemos que en esta sección estamos trabajando con matrices de transición no reversibles, por lo que una pregunta válida es ¿Qué pasa si usamos el teorema 3.8 cota de contraste con una matriz reversible?
Lo primero que debemos ver, es lo que pasa con $\widehat{P}$, por la definición de $\widehat{P}$ (definición 3.21) y por la definición de matriz de transición reversible (definición 2.15), podemos notar que $\widehat{P} = P$.
Es fácil ver que $M(P) = P\widehat{P} = P^2$, así si γ es el SLEM de M y ρ el de P tenemos que $\gamma = \rho^2$. Además, sean $i \in E$ y $\mu = \delta_i$ una distribución inicial de la cadena de Markov, entonces por el teorema 3.8 cota de contraste tenemos que

$$
\begin{aligned}
4d_v^2(\delta_i^t P^n,\pi) &\leq \gamma^n\chi^2(\delta_i;\pi) \\
&\leq \rho^{2n}\chi^2(\delta_i;\pi) \\
&\leq \rho^{2n} \parallel \delta_i - \pi \parallel_{\frac{1}{\pi}}^2 \\
&\leq \left(\frac{1-\pi(i)}{\pi(i)}\right)\rho^{2n}
\end{aligned}
$$

por lo que

$$4d_v^2(\delta_i^t P^n,\pi) \leq \left(\frac{1-\pi(i)}{\pi(i)}\right)\rho^{2n}.$$

Esta última desigualdad es la obtenida en la ecuación (3.36) del teorema 3.7 cota $\frac{1}{\pi}$, es decir, si usamos el teorema 3.8 cota de contraste con una matriz reversible recuperaremos uno de los resultados del teorema 3.7 cota $\frac{1}{\pi}$.

Capítulo 4. Cotas para el SLEM

En el capítulo anterior encontramos distintas velocidades de convergencia de P^n, $n \in \mathbb{N}$, a su distribución estacionaria. Todas estas cotas encontradas están en función del *SLEM* de P, pero conforme el espacio de estados de la cadena de Markov crece, encontrar el valor exacto del *SLEM* es cada vez más costoso, computacionalmente hablando.

En este capitulo nos enfocaremos en encontrar cotas para el SLEM, veremos dos maneras distintas para lograrlo, la primera será por medio de trayectorias pesadas y la segunda será por medio de la conductancia.

Para este capítulo, trabajaremos solo con matrices de transición reversibles pues en el capitulo anterior vimos que los eigenvalores de estas matrices son reales y contenidas en el intervalo $[-1,1]$, así que vamos a reordenarlos con respecto al orden de los reales, es decir $1 = \lambda_1 > \lambda_2 \geq \ldots \geq \lambda_r \geq -1$. Cabe destacar que este orden es diferente al establecido en el Teorema de *Perron-Frobenius*, en este caso λ_2 es el segundo eigenvalor más grande, SLE por sus siglas en inglés, mientras que en el Teorema de Perron-Frobenius λ_2 es el segundo eigenvalor con módulo más grande, SLEM por sus siglas en inglés.

Sección 4.1 Trayectorias pesadas

En esta sección daremos dos teoremas que nos ayudarán a acotar al SELM, estos teoremas nos darán cotas superiores de λ_2 e inferiores para λ_r en términos de la geometría de la gráfica de transición.

Antes de enunciar el teorema 4.9, es necesario definir el coeficiente de Poincaré, para ello sea $r \in \mathbb{N}$ y P la matriz de transición de una cadena de Markov irreducible y reversible sobre un espacio de estados finito $E = \{1,...,r\}$, además sea π su distribución estacionaria. En la gráfica de transición asociada a P, denotaremos a la arista orientada $i \Longrightarrow j$ por e, además denotaremos al vértice inicial por $e^- = i$ y al vértice final por $e^+ = j$.
Ahora para cualquier arista orientada e definamos el peso de la arista como

$$Q(e) := \pi(i)p_{ij}$$

así, para cada pareja ordenada (i, j), seleccionaremos una y solo una trayectoria $i, i_1, ..., i_m, j$ de tal manera que $p_{ii_1}p_{i_1i_2}...p_{i_mj} > 0$, además no usaremos la misma arista dos veces,

sea Γ el conjunto de las trayectorias seleccionadas.
Ahora, para cada trayectoria $\gamma_{ij} \in \Gamma$ definimos el peso de una trayectoria como

$$| \gamma_{ij} |_Q := \sum_{e \in \gamma_{ij}} \frac{1}{Q(e)}$$

finalmente definiremos el coeficiente de Poincaré como

$$K = K(\Gamma) = \max_{e \in E^2} \left\{ \sum_{\gamma_{ij} \ni e} | \gamma_{ij} |_Q \, \pi(i)\pi(j) \right\}$$

Teorema **4.9.** *Cota superior de una trayectoria pesada*
Sea $r \in \mathbb{N}$, P la matriz de transición de una cadena de Markov irreducible y reversible sobre un espacio de estados finito $E = \{1, ..., r\}$ y π su distribución estacionaria. Sean λ_2 es el segundo eigenvalor más grande de P, SLE por sus siglas en inglés, Γ un conjunto de trayectorias seleccionadas como en la definición del coeficiente de Poincaré y $K = K(\Gamma)$ el coeficiente de Poincaré, entonces

$$\lambda_2 \leq 1 - \frac{1}{K} \tag{4.47}$$

Demostración.
Por la conclusión de este teorema, basta con demostrar que K satisface las hipótesis del colorario 1 del teorema de Rayleigh, localizado en la página 89 de este trabajo, es decir, solo debemos demostrar que K es positivo y que $Var_\pi(x) \leq K\varepsilon_\pi(x,x)$ para todo $x \in \mathbb{R}^r$.
Primero mostremos que K es positivo.
Como por definición $Q(e) = \pi(i)p_{ij} \geq 0$, para toda $i, j \in E$, entonces

$$| \gamma_{ij} |_Q = \sum_{e \in \gamma_{ij}} \frac{1}{Q(e)} > 0$$

por lo que

$$K = \max_{e \in E^2} \left\{ \sum_{\gamma_{ij} \ni e} | \gamma_{ij} |_Q \, \pi(i)\pi(j) \right\} > 0$$

Ahora veamos que $Var_\pi(x) \leq K\varepsilon_\pi(x,x)$ para todo $x \in \mathbb{R}^r$.
Sea $x \in \mathbb{R}^r$, entonces

$$\begin{aligned}
Var_\pi(x) &= \| x \|_\pi^2 - \langle x, \mathbb{1} \rangle_\pi^2 \\
&= \left(\sum_{i \in E} x^2(i)\pi(i) \right) - \left(\sum_{i \in E} x(i)\pi(i) \right)^2 \\
&= \sum_{i \in E} x^2(i)\pi(i) - \sum_{i \in E} \sum_{j \in E} x(i)x(j)\pi(i)\pi(j) \\
&= \sum_{i \in E} \sum_{j \in E} x^2(i)\pi(i)\pi(j) - \sum_{i \in E} \sum_{j \in E} x(i)x(j)\pi(i)\pi(j) \\
&= \sum_{i \in E} \sum_{j \in E} (x(i) - x(j))\, x(i)\pi(i)\pi(j)
\end{aligned}$$

Como la suma es independiente del índice que ocupemos, entonces

$$\begin{aligned} Var_\pi(x) &= \sum_{j\in E}\sum_{i\in E}(x(j)-x(i))\,x(j)\pi(j)\pi(i) \\ &= \sum_{j\in E}\sum_{i\in E}(x(i)-x(j))\,(-x(j))\,\pi(j)\pi(i) \end{aligned}$$

por lo que

$$\begin{aligned} Var_\pi(x) &= \frac{1}{2}\sum_{i\in E}\sum_{j\in E}(x(i)-x(j))\,x(i)\pi(i)\pi(j) + \frac{1}{2}\sum_{j\in E}\sum_{i\in E}(x(i)-x(j))\,(-x(j))\,\pi(j)\pi(i) \\ &= \frac{1}{2}\sum_{i,j\in E}(x(i)-x(j))^2\,\pi(i)\pi(j) \end{aligned}$$

además, sea $\gamma_{ij}\in\Gamma$, notemos que

$$\sum_{e\in\gamma_{ij}}\left(x(e^-)-x(e^+)\right) = x(i)-x(j)$$

por ser telescópica, entonces continuando con la varianza de x

$$\begin{aligned} Var_\pi(x) &= \frac{1}{2}\sum_{i,j\in E}\left(\sum_{e\in\gamma_{ij}}\left(x(e^-)-x(e^+)\right)\right)^2\pi(i)\pi(j) \\ &= \frac{1}{2}\sum_{i,j\in E}\left(\sum_{e\in\gamma_{ij}}\frac{1}{Q^{\frac{1}{2}}(e)}Q^{\frac{1}{2}}(e)\left(x(e^-)-x(e^+)\right)\right)^2\pi(i)\pi(j) \end{aligned}$$

Así usando la desigualdad de Cauchy-Schwarz

$$\begin{aligned} Var_\pi(x) &\le \frac{1}{2}\sum_{i,j\in E}\left(\sum_{e\in\gamma_{ij}}\frac{1}{Q(e)}\right)\left(\sum_{e\in\gamma_{ij}}Q(e)\left(x(e^-)-x(e^+)\right)^2\right)\pi(i)\pi(j) \\ &= \frac{1}{2}\sum_{i,j\in E}\mid\gamma_{ij}\mid_Q\left(\sum_{e\in\gamma_{ij}}Q(e)\left(x(e^-)-x(e^+)\right)^2\right)\pi(i)\pi(j) \\ &= \frac{1}{2}\sum_{i,j\in E}\sum_{e\in\gamma_{ij}}\mid\gamma_{ij}\mid_Q Q(e)\left(x(e^-)-x(e^+)\right)^2\pi(i)\pi(j) \\ &= \frac{1}{2}\sum_{e\in E^2}\sum_{\gamma_{i,j}\ni e}Q(e)\left(x(e^-)-x(e^+)\right)^2\mid\gamma_{ij}\mid_Q\pi(i)\pi(j) \\ &= \frac{1}{2}\sum_{e\in E^2}Q(e)\left(x(e^-)-x(e^+)\right)^2\left(\sum_{\gamma_{i,j}\ni e}\mid\gamma_{ij}\mid_Q\pi(i)\pi(j)\right) \\ &\le \frac{1}{2}\sum_{e\in E^2}Q(e)\left(x(e^-)-x(e^+)\right)^2K(\Gamma). \end{aligned}$$

Por otro lado, recordemos la proposición 6.19, ubicado en el apéndice C en la página 85, esta proposición nos da una equivalencia de la forma de Dirichlet en términos de x, π y P. Ahora daremos una ecuación equivalente en términos del peso de las aristas.

$$\begin{aligned}\varepsilon_\pi(x,x) &= \frac{1}{2}\sum_{i,j\in E}\pi(i)p_{ij}\,(x(i)-x(j))^2 \\ &= \frac{1}{2}\sum_{e\in E^2}Q(e)\left(x(e^-)-x(e^+)\right)^2\end{aligned}$$

por lo que para toda $x\in\mathbb{R}^r$

$$var_\pi(x)\leq \varepsilon_\pi(x,x)K(\Gamma)$$

Hasta aquí se satisfacen las hipótesis del primer colorario del teorema de Rayleigh, por lo que $\lambda_2\leq 1-\frac{1}{K}$ $\square$

Usando el coeficiente de Poincaré no es la única forma de acotar a λ_2, sino que también podemos acotarlo si variamos el coeficiente, esto es, definiendo

$$K=\max_{e\in E^2}\frac{1}{Q(e)}\left(\sum_{\gamma_{ij}\ni e}\mid\gamma_{ij}\mid\pi(i)\pi(j)\right)$$

en donde $\mid\gamma_{ij}\mid$ es la longitud de la trayectoria γ_{ij}.
Una vez que conocemos una cota de λ_2, nos interesa también acotar a λ_r pero inferiormente, para ello haremos lo siguiente, para cada estado $i\in E$ seleccionaremos exactamente una trayectoria cerrada σ_i de i a i de tal forma que no pase dos veces por una misma arista y con un número impar de pasos, para que esto sea posible asumiremos que P es aperiódica. Sea Σ el conjunto de estas trayectorias, así para cada $\sigma\in\Sigma$ sea

$$\mid\sigma_i\mid_Q=\sum_{e\in\sigma_i}\frac{1}{Q(e)},$$

además

$$\alpha=\alpha(\Sigma)=\max_{e\in E^2}\left(\sum_{\sigma_i\ni e}\mid\sigma_i\mid_Q\pi(i)\right).$$

Teorema **4.10.** *Cota inferior de una trayectoria pesada*
Sea $r\in\mathbb{N}$, P la matriz de transición de una cadena de Markov irreducible, aperiódica y reversible sobre un espacio de estados finito $E=\{1,\ldots,r\}$ y π su distribución estacionaria. Sean λ_r es el eigenvalor más pequeño de P, SE por sus siglas en inglés, Σ un conjunto de trayectorias y $\alpha=\alpha(\Sigma)$ definidas como en el párrafo anterior, entonces

$$\lambda_r\geq -1+\frac{2}{\alpha}. \tag{4.48}$$

Demostración.
Por la conclusión de este teorema, basta con demostrar que $\frac{2}{\alpha}$ satisface las hipótesis del colorario 2 del teorema de Rayleigh, localizado en la página 90 de este trabajo, es

decir, solo debemos mostrar que $\frac{2}{\alpha} > 0$ y que para cualquier $x \in \mathbb{R}^r$,
$\langle Px, x\rangle_\pi + \| x \|_\pi^2 \geq \frac{2}{\alpha} \| x \|_\pi^2$.
Primero probaremos que $\frac{2}{\alpha} > 0$, como por definición $Q(e) = \pi(i)p_{ij} \geq 0$, para toda $i, j \in E$, entonces

$$| \sigma_i |_Q = \sum_{e \in \sigma_i} \frac{1}{Q(e)} > 0$$

por lo que

$$\alpha = \max_{e \in E^2} \left\{ \sum_{\sigma_i \ni e} | \sigma_i |_Q \, \pi(i) \right\} > 0.$$

Ahora, probemos que para cualquier $x \in \mathbb{R}^r$, $\langle Px, x\rangle_\pi + \| x \|_\pi^2 \geq \frac{2}{\alpha} \| x \|_\pi^2$.
Sea $x \in \mathbb{R}^r$, entonces

$$\begin{aligned}
\| x \|_\pi^2 + \langle Px, x\rangle_\pi &= \sum_{i \in E} x(i)^2 \pi(i) + \sum_{i \in E} \sum_{j \in E} p_{ij} x(i) x(j) \pi(i) \\
&= \sum_{i \in E} \sum_{j \in E} p_{ij} x(i)^2 \pi(i) + \sum_{i \in E} \sum_{j \in E} p_{ij} x(i) x(j) \pi(i) \\
&= \sum_{i \in E} \sum_{j \in E} p_{ij} x(i)^2 \pi(i) + p_{ij} x(i) x(j) \pi(i) \\
&= \sum_{i \in E} \sum_{j \in E} (x(i) + x(j)) x(i) p_{ij} \pi(i).
\end{aligned}$$

Como la suma de la ecuación anterior es independiente del índice, podemos hacer lo siguiente

$$\begin{aligned}
\| x \|_\pi^2 + \langle Px, x\rangle_\pi &= \sum_{j \in E} \sum_{i \in E} (x(j) + x(i)) x(j) p_{ji} \pi(j) \\
&= \sum_{j \in E} \sum_{i \in E} (x(i) + x(j)) x(j) p_{ij} \pi(i) \qquad \text{porque P es reversible}
\end{aligned}$$

por lo que

$$\| x \|_\pi^2 + \langle Px, x\rangle_\pi = \frac{1}{2} \sum_{i \in E} \sum_{j \in E} (x(i) + x(j)) x(i) p_{ij} \pi(i) + \frac{1}{2} \sum_{j \in E} \sum_{i \in E} (x(i) + x(j)) x(j) p_{ij} \pi(i)$$

$$\| x \|_\pi^2 + \langle Px, x\rangle_\pi = \frac{1}{2} \sum_{i, j \in E} (x(i) + x(j))^2 p_{ij} \pi(i). \tag{4.49}$$

Por otro lado, sea $i \in E$, como σ_i es una trayectoria que va del estado i al estado i en un número impar de pasos, esto es $\sigma_i = (i = i_0, i_1, ..., i_{2m-1}, i_{2m}, i_0)$, para algún $m \in \mathbb{N}$, entonces

$$\begin{aligned}
x(i) &= \frac{1}{2} \{ [x(i_0) + x(i_1)] - [x(i_1) + x(i_2)] + ... - [x(i_{2m-1}) + x(i_{2m})] + [x(i_{2m}) + x(i_0)] \} \\
&= \frac{1}{2} \sum_{e \in \sigma_i} (-1)^{n(e)} (x(e^+) + x(e^-))
\end{aligned}$$

donde si $e=(i_k,i_{k+1})\in\sigma_i$ entonces $n(e)=k$.
Ahora veremos una cota para la norma π de x

$$
\begin{aligned}
\| x \|_\pi^2 &= \sum_{i\in E} x(i)^2\pi(i)\\
&= \sum_{i\in E}\left(\frac{1}{2}\sum_{e\in\sigma_i}(-1)^{n(e)}(x(e^+)+x(e^-))\right)^2\pi(i)\\
&= \sum_{i\in E}\frac{\pi(i)}{4}\left(\sum_{e\in\sigma_i}(-1)^{n(e)}(x(e^+)+x(e^-))\right)^2\\
&= \sum_{i\in E}\frac{\pi(i)}{4}\left(\sum_{e\in\sigma_i}\frac{1}{Q^{\frac{1}{2}}(e)}Q^{\frac{1}{2}}(e)(-1)^{n(e)}(x(e^+)+x(e^-))\right)^2
\end{aligned}
$$

así, usando la desigualdad de Cauchy-Schwarz tenemos que

$$
\begin{aligned}
\| x \|_\pi^2 &\leq \sum_{i\in E}\frac{\pi(i)}{4}\left(\sum_{e\in\sigma_i}\frac{1}{Q(e)}\right)\left(\sum_{e\in\sigma_i}Q(e)(x(e^+)+x(e^-))^2\right)\\
&= \sum_{i\in E}\sum_{e\in\sigma_i}\frac{\pi(i)}{4} \mid \sigma_i \mid_Q Q(e)(x(e^+)+x(e^-))^2\\
&= \frac{1}{4}\sum_{e\in E^2}\sum_{\sigma_i\ni e}\pi(i) \mid \sigma_i \mid_Q Q(e)(x(e^+)+x(e^-))^2\\
&= \frac{1}{4}\sum_{e\in E^2}\sum_{\sigma_i\ni e}(x(e^+)+x(e^-))^2 Q(e) \mid \sigma_i \mid_Q \pi(i)\\
&= \frac{1}{4}\sum_{e\in E^2}(x(e^+)+x(e^-))^2 Q(e)\left(\sum_{\sigma_i\ni e}\mid \sigma_i \mid_Q \pi(i)\right)\\
&\leq \frac{\alpha}{4}\sum_{e\in E^2}(x(e^+)+x(e^-))^2 Q(e)
\end{aligned}
$$

si la ecuación anterior lo ponemos en términos de $i,j\in E$ tenemos que

$$
\begin{aligned}
\| x \|_\pi^2 &\leq \frac{\alpha}{4}\sum_{i,j\in E}(x(i)+x(j))^2 pij\pi(i)\\
&= \frac{\alpha}{2}\left(\frac{1}{2}\sum_{i,j\in E}(x(i)+x(j))^2 pij\pi(i)\right)\\
&= \frac{\alpha}{2}(\| x \|_\pi^2 + \langle Px,x\rangle_\pi) && \text{por la eciación (4.49)}
\end{aligned}
$$

□

Con el teorema 4.9 pudimos encontrar una cota superior para λ_2, mientras que con el teorema 4.10 encontramos una cota inferior para λ_r, con esto ya es posible acotar a

nuestro SLEM superiormente y estará acotado de la siguiente manera
$\rho \leq \text{máx}(1 - \frac{1}{K}, |-1 + \frac{2}{\alpha}|)$, con K como en el teorema 4.9 y α como en el teorema 4.10.
Para encontrar la cota del SLEM como en el párrafo anterior es necesario encontrar K y α, lo que también puede llegar a ser costoso, computacionalmente hablando, pero no tanto como el calcular cada eigenvalor de P si este tiene un espacio de estados grande.

Sección 4.2 Conductancia

En la sección anterior encontramos una cota superior para λ_2 (el SLE) y una cota inferior de λ_r con lo que pudimos definir una cota superior para el SLEM ρ, además recordemos que $\lambda_2 \leq \rho$, así si sobrestimamos el valor del SLEM podríamos tener una cota de la convergencia de P^n, $n \in \mathbb{N}$, a su estacionaria mucho más grande de lo que realmente es, por lo que puede ser de nuestro interés también encontrar una cota inferior para λ_2, así podríamos tener tanto una cota inferior del SLEM como una superior.
En esta sección acotaremos a λ_2, tanto inferiormente como superiormente, haciendo uso de la conductancia de la cadena.
Antes de dar la definición de *conductancia* daremos otras definiciones y propiedades que nos serán de utilidad para demostrar el teorema 4.11

Definición **4.23.** *Capacidad*
Sea $r \in \mathbb{N}$ y $E = \{1, ..., r\}$ el espacio de estados de una cadena de Markov irreducible y reversible, además sea π su distribución estacionaria. Sea $B \subseteq E$, $B \neq \emptyset$, definimos la capacidad de B como

$$\pi(B) := \sum_{i \in B} \pi(i)$$

Definición **4.24.** *Flujo ergódico fuera de B*
Sea $r \in \mathbb{N}$ y $E = \{1, ..., r\}$ el espacio de estados de una cadena de Markov irreducible y reversible, además sea π su distribución estacionaria. Sea $B \subseteq E$, $B \neq \emptyset$, definimos el flujo ergódico fuera de B como

$$F(B) := \sum_{i \in B, j \in B^c} \pi(i) p_{ij}$$

Proposición **4.14.**
Sea $r \in \mathbb{N}$ y $E = \{1, ..., r\}$ el espacio de estados de una cadena de Markov irreducible y reversible, además sea π su distribución estacionaria, entonces para toda $B \subseteq E$, $B \neq \emptyset$

$$0 < F(B) \leq \pi(B) \leq 1$$

Demostración.

Sea $B \subseteq E$, $B \neq \emptyset$, entonces

$$\begin{aligned} F(B) &= \sum_{i\in B, j\in B^c} \pi(i)p_{ij} \\ &= \sum_{i\in B} \pi(i) \left(\sum_{j\in B^c} p_{ij} \right) \\ &\leq \sum_{i\in B} \pi(i) \\ &= \pi(B) \leq 1 \end{aligned}$$

además como B es no vacío entonces $F(B) > 0$ □

Ahora procederemos a dar la definición de la conductancia de una cadena de Markov

Definición **4.25.** *Conductancia*
Sea $r \in \mathbb{N}$, P la matriz de transición de una cadena de Markov irreducible y reversible sobre un espacio de estados finito $E = \{1,...,r\}$ y π su distribución estacionaria, definimos la conductacia de una cadena de Markov como

$$\varphi(P) := \inf_{B\subseteq E, B\neq\emptyset, \pi(B)\leq\frac{1}{2}} \left\{ \frac{F(B)}{\pi(B)} \right\}$$

Teorema **4.11.** *Cota de conductancia*
Sea $r \in \mathbb{N}$, P la matriz de transición de una cadena de Markov irreducible y reversible sobre un espacio de estados finito $E = \{1,...,r\}$ y π su distribución estacionaria. Además sea λ_2 el segundo eigenvalor más grande de P, SLE por sus siglas en inglés, entonces

$$1 - 2\varphi(P) \leq \lambda_2 \leq 1 - \frac{\varphi^2(P)}{2}$$

Demostración.
Empezaremos con la cota superior.
Sean $\lambda \neq 1$ un eigenvalor de P y u un eigenvector izquierdo de P asociado a λ, tal que u y π sean ortogonales respecto a la norma $\frac{1}{\pi}$, esto es

$$\begin{aligned} \langle u, \pi \rangle_{\frac{1}{\pi}} &= \sum_{i\in E} u(i)\pi(i)\frac{1}{\pi(i)} \\ &= \sum_{i\in E} u(i) = 0. \end{aligned}$$

Como u y π son ortogonales, entonces u tiene entradas tanto positivas como negativas, pues $u \neq 0$, así sin pérdida de generalidad existe $k \in \{1,...,r\}$ de tal forma que $u(1) \geq ... \geq u(k) \geq 0 \geq u(k+1) \geq ... \geq u(r)$ y $\pi(S) \leq \frac{1}{2}$ donde $S = \{1,...,k\}$, de ser necesario reordenaremos los estados y/o cambiamos u por $-u$.
Definamos un nuevo vector de la siguiente manera

$$y(i) = \begin{cases} \frac{u(i)}{\pi(i)} & \text{si } u(i) > 0 \\ 0 & \text{en cualquier otro caso} \end{cases}$$

Como u es un eigenvector izquierdo de P asociado a λ, entonces
$u^t(\mathbb{I}_{rxr} - P) = (1-\lambda)u^t$ por lo que

$$\begin{aligned}
u^t(\mathbb{I}_{rxr} - P)y &= (1-\lambda)u^t y \\
&= (1-\lambda)\sum_{i\in E} u(i)y(i) \\
&= (1-\lambda)\sum_{i\in S} u(i)y(i) \\
&= (1-\lambda)\sum_{i\in S} \pi(i)\frac{u(i)}{\pi(i)}y(i) \\
&= (1-\lambda)\sum_{i\in S} \pi(i)y^2(i)
\end{aligned}$$

así,

$$u^t(\mathbb{I}_{rxr} - P)y = (1-\lambda)\sum_{i\in S} \pi(i)y^2(i). \tag{4.50}$$

Por otro lado,

$$\begin{aligned}
u^t(\mathbb{I}_{rxr} - P)y &= \sum_{i\in S}\sum_{j\in E} u(j)(\delta_j(i) - p_{ji})y(i) \\
&\geq \sum_{i,j\in S} u(j)(\delta_j(i) - p_{ji})y(i),
\end{aligned}$$

la desigualdad anterior es cierta debido a que para cualquier $i \in S$ y $j \notin S$, $u(j) \leq 0$ y $(\delta_j(i) - p_{ji})y(i) = -p_{ji}y(i) \leq 0$ por lo que $u(j)(\delta_j(i) - p_{ji})y(i) \geq 0$, entonces

$$\begin{aligned}
u^t(\mathbb{I}_{rxr} - P)y &\geq \sum_{i,j\in S} \frac{u(j)}{\pi(j)}(\delta_j(i) - p_{ji})y(i)\pi(j) \\
&= \sum_{i,j\in S} y(j)(\delta_j(i) - p_{ji})y(i)\pi(j) \\
&= \sum_{i,j\in E} y(j)(\delta_j(i) - p_{ji})y(i)\pi(j) \qquad \text{pues } y(i) = 0 \text{ para toda } i \notin S \\
&= \sum_{j\in E} y(j)\left(\sum_{i\in E}(\delta_j(i) - p_{ji})y(i)\right)\pi(j) \\
&= \langle y, (\mathbb{I}_{rxr} - P)y\rangle_\pi \\
&\geq \varepsilon_\pi(y,y).
\end{aligned}$$

Utilizaremos la proposición 6.19 del apéndice C, el cual nos da una definición equivalente a
$\varepsilon_\pi(x,x)$ en términos de P, π y las entradas de x, esto es $\varepsilon_\pi(x,x) = \frac{1}{2}\sum_{i,j\in E} \pi(i)p_{ij}(x(i) - x(j))^2$.

Por lo que,

$$\begin{aligned} u^t(\mathbb{I}_{rxr}-P)y &\geq \varepsilon_\pi(y,y) \\ &= \frac{1}{2}\sum_{i,j\in E}\pi(i)p_{ij}\left(y(i)-y(j)\right)^2 \\ &= \sum_{i<j}\pi(i)p_{ij}\left(y(i)-y(j)\right)^2 \end{aligned}$$

así, usando la ecuación (4.50) tenemos que

$$\begin{aligned} (1-\lambda)\sum_{i\in S}\pi(i)y^2(i) &= u^t(\mathbb{I}_{rxr}-P)y \\ &\geq \sum_{i<j}\pi(i)p_{ij}\left(y(i)-y(j)\right)^2, \end{aligned}$$

entonces

$$1-\lambda \geq \frac{\sum_{i<j}\pi(i)p_{ij}\left(y(i)-y(j)\right)^2}{\sum_{i\in S}\pi(i)y^2(i)}. \tag{4.51}$$

Trabajaremos con un término similar al del numerador de la inecuación (4.51), para ello usaremos la siguiente propiedad $(a+b)^2\leq 2(a^2+b^2)$ para todo $a,b\in\mathbb{R}$, entonces

$$\begin{aligned} \sum_{i<j}\pi(i)p_{ij}\left(y(i)+y(j)\right)^2 &\leq 2\sum_{i<j}\pi(i)p_{ij}\left(y^2(i)+y^2(j)\right) \\ &= 2\left(\sum_{i<j}\pi(i)p_{ij}y^2(i)+\sum_{i<j}\pi(i)p_{ij}y^2(j)\right) \\ &= 2\left(\sum_{i<j}\pi(i)p_{ij}y^2(i)+\sum_{i<j}\pi(j)p_{ji}y^2(j)\right) && \text{por reversibilidad} \\ &= 2\left(\sum_{i<j}\pi(i)p_{ij}y^2(i)+\sum_{i>j}\pi(i)p_{ij}y^2(i)\right) && \text{intercambiado los índices} \\ &= 2\left(\sum_{i\neq j}\pi(i)p_{ij}y^2(i)\right) \\ &\leq 2\sum_{i\in E}\pi(i)y^2(i) \\ &= 2\sum_{i\in S}\pi(i)y^2(i) \end{aligned}$$

por lo que

$$\frac{\sum_{i<j}\pi(i)p_{ij}\left(y(i)+y(j)\right)^2}{2\sum_{i\in S}\pi(i)y^2(i)} \leq 1, \tag{4.52}$$

así, continuando con la inecuación (4.51) y usando la inecuación (4.52) tenemos

$$
\begin{aligned}
1-\lambda &\geq \frac{\sum_{i<j}\pi(i)p_{ij}\,(y(i)-y(j))^2}{\sum_{i\in S}\pi(i)y^2(i)} \\
&\geq \left(\frac{\sum_{i<j}\pi(i)p_{ij}\,(y(i)-y(j))^2}{\sum_{i\in S}\pi(i)y^2(i)}\right)\left(\frac{\sum_{i<j}\pi(i)p_{ij}\,(y(i)+y(j))^2}{2\sum_{i\in S}\pi(i)y^2(i)}\right) \\
&= \frac{1}{2}\left(\frac{\left(\sum_{i<j}\pi(i)p_{ij}\,(y(i)-y(j))^2\right)\left(\sum_{i<j}\pi(i)p_{ij}\,(y(i)+y(j))^2\right)}{\left(\sum_{i\in S}\pi(i)y^2(i)\right)^2}\right)
\end{aligned}
$$

y usando la desigualdad de Cauchy-Schwarz tenemos que

$$
1-\lambda \geq \frac{1}{2}\left(\frac{\left(\sum_{i<j}\pi(i)p_{ij}(y(i)-y(j))(y(i)+y(j))\right)^2}{\left(\sum_{i\in S}\pi(i)y^2(i)\right)^2}\right)
$$

entonces

$$
1-\lambda \geq \frac{1}{2}\left(\frac{\sum_{i<j}\pi(i)p_{ij}\left(y^2(i)-y^2(j)\right)}{\sum_{i\in S}\pi(i)y^2(i)}\right)^2. \tag{4.53}
$$

Ahora, definamos a $S_\ell = \{1,...,\ell\}$ para algún $\ell \in E$, entonces

$$
\begin{aligned}
\sum_{i<j}\pi(i)p_{ij}\left(y^2(i)-y^2(j)\right) &= \sum_{i<j}\pi(i)p_{ij}\left(\sum_{i\leq\ell<j} y^2(\ell)-y^2(\ell+1)\right) \\
&= \sum_{i<j}\sum_{i\leq\ell<j}\pi(i)p_{ij}\left(y^2(\ell)-y^2(\ell+1)\right) \\
&= \sum_{\ell=1}^{k}\sum_{i\in S_\ell, j\notin S_\ell}\pi(i)p_{ij}\left(y^2(\ell)-y^2(\ell+1)\right) \\
&= \sum_{\ell=1}^{k}\left(y^2(\ell)-y^2(\ell+1)\right)\left(\sum_{i\in S_\ell, j\notin S_\ell}\pi(i)p_{ij}\right) \\
&= \sum_{\ell=1}^{k}\left(y^2(\ell)-y^2(\ell+1)\right)F(S_\ell)
\end{aligned}
$$

pero como para cada $\ell \in \{1,...,k\}$, $\pi(S_\ell) \leq \pi(S) \leq \frac{1}{2}$, entonces por la definición de conductancia (Definición 4.25) tenemos que $\frac{F(S_\ell)}{\pi(S_\ell)} \geq \varphi(P)$, esto es $F(S_\ell) \geq \varphi(P)\pi(S_\ell)$, por lo que continuando con la ecuación anterior

$$\begin{aligned}
\sum_{i<j}\pi(i)p_{ij}\left(y^2(i)-y^2(j)\right) &\geq \varphi(P)\sum_{\ell=1}^{k}\left(y^2(\ell)-y^2(\ell+1)\right)\pi(S_\ell)\\
&= \varphi(P)\sum_{\ell=1}^{k}\left(y^2(\ell)-y^2(\ell+1)\right)\left(\sum_{i=1}^{\ell}\pi(i)\right)\\
&= \varphi(P)\sum_{\ell=1}^{k}\sum_{i=1}^{\ell}\left(y^2(\ell)-y^2(\ell+1)\right)\pi(i)\\
&= \varphi(P)\sum_{i=1}^{k}\sum_{\ell=i}^{k}\left(y^2(\ell)-y^2(\ell+1)\right)\pi(i)\\
&= \varphi(P)\sum_{i\in S}\pi(i)\left(\sum_{\ell=i}^{k}\left(y^2(\ell)-y^2(\ell+1)\right)\right)\\
&= \varphi(P)\sum_{i\in S}\pi(i)\left(y^2(i)-y^2(k+1)\right)\\
&= \varphi(P)\sum_{i\in S}\pi(i)y^2(i)
\end{aligned}$$

por lo tanto

$$\frac{\sum_{i<j}\pi(i)p_{ij}\left(y^2(i)-y^2(j)\right)}{\sum_{i\in S}\pi(i)y^2(i)} \geq \varphi(P) \tag{4.54}$$

Así, usando las ecuaciones(4.53) y (4.54), tenemos que para cualquier eigenvalor $\lambda \neq 1$ de P, $1-\lambda \geq \frac{1}{2}\varphi^2(P)$ en particular para λ_2, es decir, $1-\lambda_2 \geq \frac{1}{2}\varphi^2(P)$, por lo que $1-\frac{1}{2}\varphi^2(P) \geq \lambda_2$.

Hasta aquí hemos terminado de demostrar la cota superior del SLE, solo nos falta ver la cota inferior.

Utilizaremos el Teorema de Rayleigh (Teorema 6.13) del apéndice C, el cuál nos dice que para cualquier $j \geq 2$,

$$1-\lambda_j = inf\left\{\frac{\varepsilon_\pi(x,x)}{Var_\pi(x)} \mid \langle x, v_i\rangle_\pi = 0\ \forall i \in \{1,\ldots,j-1\}, x \neq 0\right\}.$$

En particular para λ_2 tenemos que

$$1-\lambda_2 = inf\left\{\frac{\varepsilon_\pi(x,x)}{Var_\pi(x)} \mid \langle x, \mathbb{1}_r\rangle_\pi = 0, x \neq 0\right\}. \tag{4.55}$$

Sea $B \subset E$, $B \neq \emptyset$, de tal forma que $\pi(B) \leq \frac{1}{2}$ y defnimos

$$x(i) = \begin{cases} 1-\pi(B) & \text{si } i \in B \\ -\pi(B) & \text{si } i \notin B \end{cases}$$

Mostremos que x como lo definimos no es el vector cero y su producto interno π entre él y el vector de unos es cero.

En efecto, es fácil ver que x no es el vector cero debido a que $B \neq \emptyset$ y que para cada $i \in B$, $x(i) = 1 - \pi(B) \geq \frac{1}{2}$, por otro lado,

$$\begin{aligned}
\langle x, \mathbb{1}_r \rangle_\pi &= \sum_{i \in E} x(i)\pi(i) \\
&= \sum_{i \in B} x(i)\pi(i) + \sum_{i \notin B} x(i)\pi(i) \\
&= \sum_{i \in B} (1 - \pi(B))\pi(i) - \sum_{i \notin B} \pi(B)\pi(i) \\
&= (1 - \pi(B)) \sum_{i \in B} \pi(i) - \pi(B) \sum_{i \notin B} \pi(i) \\
&= (1 - \pi(B))\pi(B) - \pi(B)(1 - \pi(B)) = 0,
\end{aligned}$$

por lo que $1 - \lambda_2 \leq \frac{\varepsilon_\pi(x,x)}{Var_\pi(x)}$, así solo veremos cuanto valen $\varepsilon_\pi(x,x)$ y $Var_\pi(x)$. Comenzaremos calculando la varianza de x, como $\langle x, \mathbb{1}_r \rangle_\pi = 0$, entonces $Var_\pi(x) = \| x \|_\pi^2$, así calculando

$$\begin{aligned}
\| x \|_\pi^2 &= \sum_{i \in E} x^2(i)\pi(i) \\
&= \sum_{i \in B} x^2(i)\pi(i) + \sum_{i \notin B} x^2(i)\pi(i) \\
&= \sum_{i \in B} (1 - \pi(B))^2\pi(i) + \sum_{i \notin B} \pi^2(B)\pi(i) \\
&= (1 - \pi(B))^2\pi(B) + \pi^2(B)(1 - \pi(B)) \\
&= \pi(B)(1 - \pi(B))\left((1 - \pi(B)) + \pi(B)\right)
\end{aligned}$$

por lo que

$$\| x \|_\pi^2 = \pi(B)(1 - \pi(B)) \tag{4.56}$$

ahora calculemos $\varepsilon_\pi(x,x)$

$$\begin{aligned}
\varepsilon_\pi(x,x) &= \frac{1}{2} \sum_{i,j \in E} \pi(i)p_{ij}\,(x(i) - x(j))^2 \\
&= \frac{1}{2}\left(\sum_{i \in B}\sum_{j \in E} \pi(i)p_{ij}\,(x(i) - x(j))^2 + \sum_{i \notin B}\sum_{j \in E} \pi(i)p_{ij}\,(x(i) - x(j))^2 \right) \\
&= \frac{1}{2}\left(\sum_{i \in B}\sum_{j \notin B} \pi(i)p_{ij}\,(x(i) - x(j))^2 + \sum_{i \notin B}\sum_{j \in B} \pi(i)p_{ij}\,(x(i) - x(j))^2 \right) \\
&= \frac{1}{2}\left(\sum_{i \in B}\sum_{j \notin B} \pi(i)p_{ij}\,[(1 - \pi(B)) + \pi(B)]^2 + \sum_{i \notin B}\sum_{j \in B} \pi(i)p_{ij}\,[-\pi(B) - (1 - \pi(B))]^2 \right) \\
&= \frac{1}{2}\left(\sum_{i \in B}\sum_{j \notin B} \pi(i)p_{ij} + \sum_{i \notin B}\sum_{j \in B} \pi(i)p_{ij} \right)
\end{aligned}$$

así, usando la propiedad de reversibilidad

$$\begin{aligned}
\varepsilon_\pi(x,x) &= \frac{1}{2}\left(\sum_{i\in B}\sum_{j\notin B}\pi(i)p_{ij} + \sum_{i\notin B}\sum_{j\in B}\pi(j)p_{ji}\right) \\
&= \frac{1}{2}\left(\sum_{i\in B}\sum_{j\notin B}\pi(i)p_{ij} + \sum_{j\in B}\sum_{i\notin B}\pi(j)p_{ji}\right) \\
&= \sum_{i\in B}\sum_{j\notin B}\pi(i)p_{ij} \\
&= F(B).
\end{aligned}$$

Entonces, con la ecuación anterior y las ecuaciones (4.55) y (4.56) tenemos que

$$\begin{aligned}
1-\lambda_2 &\leq \frac{\varepsilon_\pi(x,x)}{\| x \|_\pi^2} \\
&= \frac{F(B)}{\pi(B)(1-\pi(B))}
\end{aligned}$$

además como $\pi(B) \leq \frac{1}{2}$, tenemos que $(1-\pi(B)) \geq \frac{1}{2}$, por lo que

$$\begin{aligned}
1-\lambda_2 &\leq \frac{F(B)}{\pi(B)\frac{1}{2}} \\
&= 2\frac{F(B)}{\pi(B)}.
\end{aligned}$$

Así, como para cualquier $B \subset E, B \neq \emptyset$ tal que $\pi(B) \leq \frac{1}{2}$, se satisface que $1-\lambda_2 \leq 2\frac{F(B)}{\pi(B)}$ entonces también es válido para el ínfimo, esto es, $1-\lambda_2 \leq 2\varphi(P)$. □

Con el teorema anterior pudimos encontrar una cota inferior para λ_2, con esto ya es posible acotar a nuestro SLEM inferiormente y estará acotado de la siguiente manera $\rho \geq 1-2\varphi(P)$.
Más aún, en la sección anterior vimos que $\lambda_2 \leq 1-\frac{1}{K}$, con K el coeficiente de Poincaré y en esta sección encontramos que $\lambda_2 \leq 1-\frac{\varphi^2(P)}{2}$, por lo que $\lambda_2 \leq \text{mín}\left\{1-\frac{1}{K}, 1-\frac{\varphi^2(P)}{2}\right\}$ con esto podemos mejorar las cotas de nuestro SLEM

$$1-2\varphi(P) \leq \rho \leq \text{máx}\left\{\left| \frac{2}{\alpha}-1 \right|, \text{mín}\left\{1-\frac{1}{K}, 1-\frac{\varphi^2(P)}{2}\right\}\right\} \tag{4.57}$$

con $\varphi^2(P)$ la conductancia de P, K como en el teorema 4.9 y α como en el teorema 4.10.

Capítulo 5. Aplicaciones

Hasta este punto hemos encontrado que podemos acotar la convergencia de la matriz de transición de una cadena de Markov irreducible y aperiódica con espacio de estados finito, con su distribución estacionaria en términos del SLEM de la matriz de transición. En el capítulo 3, encontramos diferentes cotas para un tipo especial de matrices, las reversibles, esto con ayuda de las normas que definimos, la norma $_{\pi}$ y la norma $\frac{1}{\pi}$ (definición 2.16), también encontramos cotas para matrices no reversibles. En el capítulo 4, nos dedicamos a encontrar cotas para el *SLEM* de nuestra matriz de transición, solo nos enfocamos para matrices reversibles puesto que también dimos un método para hacer que una matriz sea reversible.
Una pregunta natural que surgen en este tipo de trabajos es ¿qué aplicaciones tiene toda esta teoría? En este capítulo veremos una aplicación de lo que hemos trabajado hasta ahora, el cual es las redes neuronales de Hopfield, cuya aplicación es la de servir como una red de memoria asociativa, la definiremos en términos de un campo aleatorio de Markov, pero este último podremos verlo en términos de una cadena de Markov con el método de muestreo de Gibbs, por lo que primero daremos una breve introducción de Campos de Markov.

Sección 5.1 Campos de Markov

Dada una cadena de Markov, $\{X_n\}_{n\leq 0}$, la propiedad de Markov implica que para $n \geq 1$, X_n es independiente de X_k, para todo $k \notin \{n-1, n, n+1\}$ dado (X_{n-1}, X_{n+1}).
Así, si llamamos a n un sitio, X_n el valor del proceso en el sitio n y el conjunto $\{n-1, n+1\}$ los vecinos del sitio n, la propiedad de Markov puede ser reescrita como: "Para toda $n \geq 1$, el valor en el sitio n es independiente del valor en el sitio $k \notin \{n-1, n, n+1\}$ dado los valores de los vecinos del sitio n".
Este es el punto de partida para poder generalizar la propiedad de Markov y pasamos de tener cadenas de Markov a tener campos de Markov.

Antes de pasar a la definición formal de un campo de Markov, daremos la definición de Campo aleatorio y la de sistema de vecindades.

Definición **5.26.** *Campo aleatorio*
Sea S un conjunto finito con s elementos a los que llamaremos sitios. Además, sea Λ

un conjunto finito llamado espacio de fases. Definimos un campo aleatorio *sobre S con fases en Λ como una colección $X = \{X(s)\}_{s\in S}$ de tal forma que para cada $s \in S$, $X(s)$ es una variable aleatoria que toma valores en Λ.*

Un campo aleatorio X también puede ser considerado como un vector aleatorio con espacio de configuraciones Λ^S y una configuración $x \in \Lambda^S$ es de la forma $x = (x(s), s \in S)$ en donde $x(s) \in \Lambda$ y $s \in S$.

Definición **5.27.** *Sistema de vecindades*
Un sistema de vecindades *sobre S, es una familia $N = \{\mathcal{N}_s\}_{s\in S}$ de subconjuntos de S de tal forma que para toda $s \in S$*

1. $s \notin \mathcal{N}_s$
2. *si $t \in \mathcal{N}_s$, entonces $s \in \mathcal{N}_t$*

El subconjunto $\mathcal{N}_s$ es llamado los vecinos del sitio s, además en la interpretación gráfica S es el conjunto de vértices mientras que N es el conjunto de aristas, así s y t están unidos por una arista si y sólo si $t \in \mathcal{N}_s$.

Definición **5.28.** *Campo aleatorio de Markov*
Sean S, Λ conjuntos finitos, N un sistema de vecindades y X un campo aleatorio sobre S con fases en Λ, además denotaremos por $\widehat{\mathcal{N}_s} = \mathcal{N}_s \cup \{s\}$.
Diremos que X es un campo aleatorio de markov *con respecto al sistema de vecindades N si para todo sitio $s \in S$ la variable aleatoria $X(s)$ y $X(S \setminus \widehat{\mathcal{N}_s})$ son independientes dado $X(\mathcal{N}_s)$, es decir para cualesquiera $s \in S$ y $x(s) \in \Lambda^S$*

$$\mathbb{P}[X(s) = x(s) \mid X(S \setminus s) = x(S \setminus s)] = \mathbb{P}[X(s) = x(s) \mid X(\mathcal{N}_s) = x(\mathcal{N}_s)] \tag{5.58}$$

En los campos de Markov podemos notar que la distribución de la fase en un sitio está completamente influenciado por las fases de los sitios vecinos.

Definición **5.29.** *Especificación local*
Sean S, Λ conjuntos finitos, N un sistema de vecindades y X un campo aleatorio de Markov con respecto al sistema de vecindades N. La característica local del campo en un sitio s es una función $\pi^s : \Lambda^S \to [0,1]$ definida por la correspondencia

$$\pi^s(x) = \mathbb{P}[X(s) = x(s) \mid X(\mathcal{N}_s) = x(\mathcal{N}_s)].$$

A la familia de funciones $\{\pi^s\}_{s\in S}$ es llamada la especificación local del campo de Markov.

Como π^s depende solo de s y de sus vecinos entonces denotaremos por $\pi^s(x) = \pi(x(s) \mid x(\mathcal{N}_s))$ esto para hacer énfasis al sistema de vecindades del campo de Markov.

La especificación local de un campo de Markov es un conjunto de funciones pero estas pueden ser condensadas en una única función llamada distribución de Gibbs la cual se ve de la siguiente manera

$$\pi(x) = \frac{1}{Z} e^{-\varepsilon(x)} \tag{5.59}$$

donde $x \in \Lambda^S$, Z es una constante de normalización y $\varepsilon(x)$ es la energía de la configuración x.
Daremos primero un par de definiciones para dar una definición exacta de ε.

Definición **5.30.** *Cliques*
Sean S un conjunto finito y N un sistema de vecindades, entonces para cualquier $s \in S$ el conjunto $\{s\}$ es un clique*, además un subconjunto $C \subset S$ con más de un elemento es llamado* clique *de la gráfica (S,N) si y sólo si para cuales quiera dos sitios distintos $s,t \in C$, $s \in \mathcal{N}_t$.*

Los cliques de una gráfica (S,N) dependen tanto de S como del sistema de vecindades, así si elegimos un sistema diferente entonces cambiarán los cliques definidos.

Definición **5.31.** *Potencial de Gibbs*
Sean S, Λ conjuntos finitos y N un sistema de vecindades. Un potencial de Gibbs sobre Λ^S relativo al sistema de vecindades N es una colección $\{V_C\}_{C\subset S}$ de funciones $V_C : \Lambda^S \to \mathbb{R} \cup \{\infty\}$ tales que

1. *$V_C = 0$ si C no es un clique.*
2. *Para toda $x, y \in \Lambda^S$ y para todo $C \subset S$, si $x(C) = y(C)$, entonces $V_C(x) = V_C(y)$*

Así, la función de energía $\varepsilon : \Lambda^S \to \mathbb{R} \cup \{\infty\}$ se dice que deriva del potencial $\{V_C\}_{C\subset S}$ si

$$\varepsilon(x) = \sum_{C\subset S} V_C(x)$$

bajo este contexto π de la ecuación (5.59) es llamado distribución de Gibbs.

Sección 5.2 Redes Neuronales de Hopfield

En el año de 1982, John Hopfield introdujo un tipo nuevo de redes neuronales, que después fue conocido como *redes neuronales de Hopfield*. La red de Hopfield básica tiene una sola capa de neuronas, con la diferencia de que todas están conectadas entre sí mediante arcos dobles ponderados bidireccionales. El peso con el que se pondera el valor de cada neurona puede ser tanto positivo como negativo. Esta topología doble de interconexión convierte a la red en una red retroalimentada o recursiva, por lo que es un buen candidato para modelar como un campo de Markov.

Para este modelo consideremos al conjunto de sitio $S = \mathbb{Z}_m^2$ y al espacio de fases $\Lambda = \{0,1\}$, además el sistema de vecindades pueden ser como se muestra en la figura 5.2, pero nosotros solo nos enfocaremos en el sistema de vecindades como en (α).

En este contexto un sitio s puede ser interpretado como una neurona excitada si $x(s) = 1$ e inhibida si $x(s) = 0$. Si $t \in \mathcal{N}_s$, podemos decir que hay una sinapsis de s a t con fuerza w_{st}, si $w_{st} > 0$ decimos que la sinapsis es excitatoria mientras que $w_{st} < 0$ la llamaremos inhibidora.

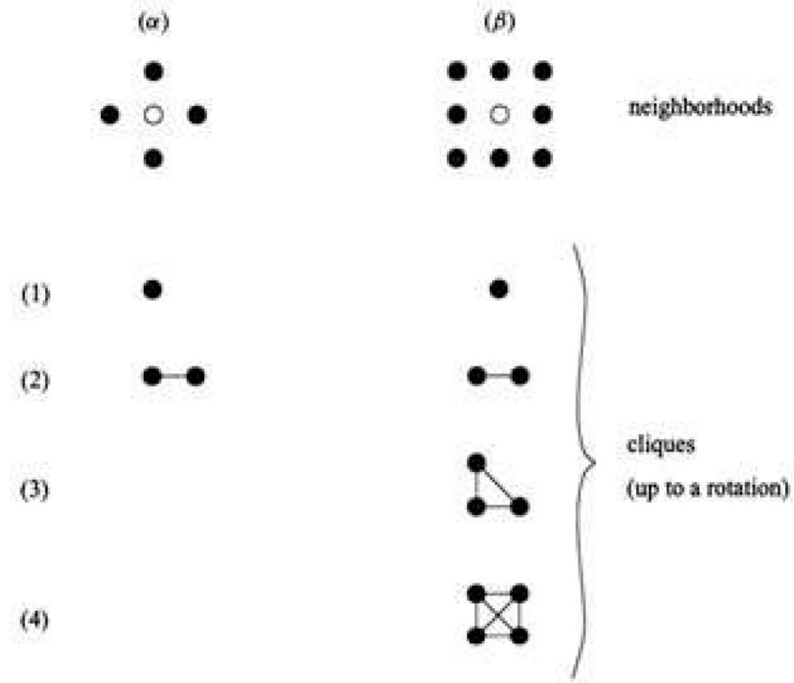

Figura 5.2: Dos ejemplos de vecindades y sus cliques.

Una de las aplicaciones de la red de Hopfield es la de servir como memoria asociativa. Una memoria asociativa es aquella que una fracción de una señal es capaz de reproducir la totalidad de la señal, por ejemplo, con el inicio de una canción el cerebro humano es capaz de reconocer y reproducir la melodía completa. Las redes de memoria asociativa son de gran utilidad para la reconstrucción de señales incompletas o deterioradas por el ruido.

Para fungir como memoria asociativa, la matriz W de pesos de la red debe ser igual a

$$W = \sum_i Y_i^t Y_i$$

donde Y_i es el i-ésimo vector a reconocer, lo que se «almacenará» en la memoria, expresado como vector renglón, por lo que $Y_i^t Y_i$ es la matriz de pesos correspondiente al i-ésimo vector a reconocer y la suma abarca todas las matrices de pesos correspondiente a todos los vectores a reconocer.

Ahora, una vez que tenemos los pesos o fuerza de las neuronas podemos dar su función de energía

$$\varepsilon(x) = \sum_{s \in S} \sum_{t \in \mathcal{N}_s} w_{ts} x(t) x(s) - \sum_{s \in S} h_s x(s)$$

donde h_s es llamado el umbral de la neurona s. De manera equivalente y recordando que los cliques en nuestro sistema de vecindades son dos neuronas tenemos que

$$\varepsilon(x) = \sum_{(s,t)} (w_{ts} + w_{st}) x(t) x(s) - \sum_{s \in S} h_s x(s).$$

Así, podemos ver que la función de energía deriva del potencial de Gibbs

$$\begin{aligned} V_{\{s\}}(x) &= -h_s x(s) \\ V_{(s,t)}(x) &= (w_{ts} + w_{st}) x(t) x(s). \end{aligned}$$

Una vez que tenemos todo lo necesario para nuestra red, podemos utilizarla para producir el vector de salida deseado, aunque el vector de entrada esté incompleto, para ello debemos dar como entrada a la red el vector de entrada después se deja iterando a la red hasta que esta alcance el estado estable más cercano, que será el vector reconocido.

Si se requiere más información sobre las redes de Hopfiel pueden consultar [Hopfield]

Sección 5.3 Método de muestreo de Gibbs

Una vez que sabemos lo esencial de los campos de Markov y de las redes neuronales de Hopfield, lo relacionaremos con las cadenas de Markov pues nuestras aplicaciones están en términos de campos de Markov mientras que los resultados obtenidos en los capítulos anteriores están en términos de cadenas de Markov, por lo que es importante saber cómo pasar de un campo de Markov a una cadena de Markov, más aún si esta puede ser irreducible y/o aperiódica.

Sean S y Λ conjuntos finitos, consideremos un campo aleatorio que cambia aleatoriamente con el tiempo, entonces tendremos un proceso estocástico $\{X_n\}_{n \geq 0}$, donde $X_n = (X_n(s), s \in S)$ y $X_n(s) \in \Lambda$. El estado a tiempo n de este proceso es un campo aleatorio sobre S con fases en Λ, o equivalentemente un vector aleatorio con valores en el espacio de estados $E = \Lambda^S$, el cual es finito debido a que tanto S como Λ lo son.

Veamos como un campo aleatorio de Markov con función de distribución de Gibbs

$$\pi(x) = \frac{1}{Z} e^{-\varepsilon(x)}, \qquad x \in \Lambda^S \tag{5.60}$$

puede verse como una cadena de Markov.

Supongamos que dado un campo aleatorio de Markov, existe $\{X_n\}_{n \geq 0}$ una cadena de Markov irreducible y aperiódica sobre un espacio de estados $E = \Lambda^S$ y con distribución estacionaria como en la ecuación (5.60), entonces podremos usar todos los resultados que hemos obtenido en los anteriores capítulos, desde calcular la velocidad de convergencia hasta acotar al *SLEM*.

La dificultad radica en encontrar una cadena de Markov irreducible y aperíodica, el método de muestreo de Gibbs nos ayudará a resolver esto de la siguiente manera:

Definimos una distribución de probabilidad estrictamente positiva sobre S, digamos $(q_s, s \in S)$ *y la transición de* $X_n = x$ *a* $X_{n+1} = y$ *se hace bajo la siguiente regla "Se actualiza (o no) la fase del sitio s de x dejando todo exactamente igual, esto es* $x(S \setminus s) = y(S \setminus s)$ *y esto lo hará con probabilidad* q_s*, entonces* $y(s)$ *es seleccionado con probabilidad* $\pi(y(s) \mid x(S \setminus s))$*"*
con esto la probabilidad de transición de $X_n = x$ *a* $X_{n+1} = y$ *está dada por*

$$\mathbb{P}[X_{n+1} = y \mid X_n = x] = q_s \pi(y(s) \mid x(S \setminus s)) \mathbb{1}_{x(S\setminus s)=y(S\setminus s)}.$$

Con este procedimiento aseguramos que $\{X_n\}_{n\geq 0}$ es una cadena de Markov irreducible, aperiodica con distribución estacionaria π definida como en la ecuación (5.60).

En la practica, los sitios no son actualizados de manera aleatoria sino que como S es finito podemos definir un 'orden' por ejemplo $S = \{s_1, s_2, ..., s_N\}$, y actualizarlos en ese 'orden', así cada sitio es visitado periódicamente. Denotaremos por Z_n al estado del campo aleatorio después del n-ésimo ciclo, así $Z_n = X_{nN}$ donde X_k denota el estado antes de la k-ésima actualización.
Al tiempo k, el sitio s $(s = kmodN)$ es actualizado, con lo que si $X_k = x$ entonces $X_{k+1} = (y(s), x(S \setminus s))$ con probabilidad $\pi(y(s) \mid x(S \setminus s))$, con esto la distribución π es estacionaria para la cadena $\{X_n\}_{n\geq 0}$.

Así, si tomamos $\{Z_n\}_{n\geq 0}$, este será una cadena de Markov irreducible y aperiodica con distribución estacionaria π, la cual en cada transición se actualizan (o no) todos los sitios de S, además la matriz de transición de nuestra cadena está dada por

$$P = \Pi_{k=1}^{N} P_{s_k}$$

donde la entrada (x, y) de la matriz $P_s = \{p_{xy}^s\}_{x,y\in\Lambda^S}$ está dada por

$$p_{xy}^s = \frac{e^{-\varepsilon(y(s),x(S\setminus s))}}{\sum_{\alpha\in\Lambda} e^{-\varepsilon(\alpha,x(S\setminus s))}}.$$

Con esto ya podremos pasar un campo de Markov a una cadena de Markov irreducible y aperiódica, además de conocer su matriz de transición y su distribución estacionaria, por lo que ya podemos aplicar todos los resultados de los capítulos anteriores.
Ahora veremos el método de muestreo de Gibbs, aplicado a las redes neuronales de Hopfield para la actualización de cada uno de sus sitios.

En la sección anterior obtuvimos la función de energía para las redes, con lo que podemos deducir que la distribución de Gibbs, como en la ecuación 5.60, es

$$\pi(x) = \frac{1}{Z} e^{-\sum_{(s,t)}(w_{ts}+w_{st})x(t)x(s)+\sum_{s\in S} h_s x(s)}, \qquad x \in \Lambda^S$$

o bien la especificación local del sitio s

$$\pi(x(s) \mid x(\mathcal{N}_s)) = \frac{1}{Z} e^{-\left(\sum_{t \in \mathcal{N}_s} (w_{ts} + w_{st}) x(t) - h_s\right) x(s)}, \qquad x \in \Lambda^S$$

Así, a tiempo n, el método de muestreo de Gibbs se sitúa en el sitio $s \in S$, con probabilidad q_s, y actualiza su fase. Se actualizará la fase por 0 con probabilidad

$$p_0 = \frac{1}{1 + e^{-\left(\sum_{t \in \mathcal{N}_s} (w_{ts} + w_{st}) x(t) - h_s\right)}}$$

y por consiguiente se actualizará a 1 con probabilidad $1 - p_0$.

Recordemos que la red de memoria asociativa requiere de un valor inicial y bajo el método de muestreo de Gibbs se itera la red hasta obtener un estado estable, es decir que el valor de entrada sea el mismo que el de salida, la cuestión es que desconocemos el número de iteraciones que se deben hacer para recuperar la información que deseamos es por eso que con ayuda del método de muestreo de Gibbs podemos ver a la red de Hopfield como una cadena de Markov irreducible y aperiódica en donde podemos iterar el valor de la cadena, además el método también nos da a conocer la matriz de transición de la cadena con lo que podemos aplicar todo lo desarrollado en los capítulos anteriores, esto es podemos usar a P para producir una cota para la convergencia como la encontrada en el teorema de Perron-Frobenius (ecuación (2.24))

$$\| \delta_i^t P^n - \pi \|_1 \leq \mid \Lambda^S \mid \rho^n \left(\sup_{2 \leq j \leq |\Lambda^S|} \mid v_j(i) \mid \right)$$

e incluso podríamos encontrar mejores cotas como los mostrados en el teorema 3.6 al cual llamamos la cota π,

$$d_v(\delta_i^t P^n, \pi) \leq \frac{p_{ii}^{(2)}}{\pi(i)} \rho^{2n-2}$$

y el teorema 3.7 denominado la cota $\frac{1}{\pi}$

$$d_v^2(\delta_i^t P^n, \pi) \leq \frac{1}{4} \left(\frac{1 - \pi(i)}{\pi(i)} \right) \rho^{2n}$$

estas últimas cotas solo con válidas para matrices reversibles, pero en caso de que nuestra matriz no sea reversible podemos ocupar el teorema 3.8, que llamamos cota de contraste, construyendo una matriz auxiliar $M = PD^{-1}P^tD$, donde $D = diag(\pi)$ y cuyo SLEM es denotado por γ

$$d_v^2(\delta_i^t P^n, \pi) \leq \frac{1}{4} \gamma^n \chi^2(\delta_i; \pi)$$

donde $\chi^2(\cdot;\cdot)$ es el contraste χ^2 (definición 3.20).

Además, en caso de que se dificulte encontrar al *SLEM* de P podemos acotarlo como se hizo en el capítulo 4,

$$1 - 2\varphi(P) \leq \rho \leq \text{máx} \left\{ \mid \frac{2}{\alpha} - 1 \mid, \text{mín} \left\{ 1 - \frac{1}{K}, 1 - \frac{\varphi^2(P)}{2} \right\} \right\}$$

con $\varphi^2(P)$ la conductancia de P, K como en el teorema 4.9 y α como en el teorema 4.10.

Una vez que hemos conseguido nuestra cota y la velocidad de convergencia, podemos despejar a n para poder conocer un número de iteraciones aproximado que debemos dejar a nuestra red para conseguir el estado estable y recuperar la información que requerimos.

Apéndice A: Matrices

Matriz elevada a un número racional

En la matrices cuadradas, al multiplicar una matriz por sí misma solemos usar la notación de exponentes, por ejemplo $A^2 = AA$, estos exponentes usualmente son números naturales, aunque si la matriz A tiene inversa, denotada por A^{-1}, entonces los exponentes pueden ser números enteros, por ejemplo $A^{-2} = A^{-1}A^{-1}$, siguiendo esta lógica podemos seguir extendiendo el conjunto de exponentes y pasar del conjunto de los números enteros al de los números racionales.
En la siguiente definición se refleja lo dicho en el párrafo anterior, solo definiremos una matriz elevada a un número racional de la forma $\frac{1}{n}$, $n \in \mathbb{N}$.

Definición **6.32.** *Matriz elevada a un número racional ($\frac{1}{n}$)*
Sean $r, n \in \mathbb{N}$ y $D \in \mathscr{M}_{rxr}(\mathbb{R})$, si existe $B \in \mathscr{M}_{rxr}(\mathbb{R})$ de tal forma que

$$B^n = D$$

entonces decimos que B es la matriz D elevada a la $\frac{1}{n}$ y lo denotaremos por $D^{\frac{1}{n}}$.

Un resultado importante es la existencia de las matrices inversas, la siguiente proposición nos da la idea de quién es la matriz inversa de una matriz elevada a un número racional de la forma $\frac{1}{n}$, $n \in \mathbb{N}$.

Proposición **6.15.**
Sean $r, n \in \mathbb{N}$, $n > 0$ y $D \in \mathscr{M}_{rxr}(\mathbb{R})$, supongamos que D es invertible y que existe $B \in \mathscr{M}_{rxr}(\mathbb{R})$ tal que $B^n = D$, es decir, existe $D^{\frac{1}{n}}$, entonces B también es invertible y además $(B^{-1})^n = D^{-1}$, es decir existe $D^{\frac{-1}{n}}$.

Demostración.
Primero mostraremos que B también es invertible, para ello recordemos que una matriz cuadrada es invertible si y sólo si su determinante es diferente de cero, además de que para cuales quiera dos matrices cuadradas A, C el determinante del producto matricial es el producto de sus determinantes, esto es $det(AC) = det(A)det(C)$.
Con estos dos conceptos es fácil ver que

$$\begin{aligned} (det(B))^n &= det(B^n) \\ &= det(D) \\ &\neq 0 \end{aligned}$$

además como $n > 0$ entonces $det(B) \neq 0$, por lo tanto B es una matriz invertible. Sigamos con la demostración de que $(B^{-1})^n = D^{-1}$, para ello recordemos que

$$\begin{aligned} D^{-1}B^n &= D^{-1}(D) \\ &= \mathbb{I}_{rxr} \end{aligned}$$

por otro lado

$$\begin{aligned} (B^{-1})^n B^n &= (B^{-1})^{n-1} B^{-1} B B^{n-1} \\ &= (B^{-1})^{n-1} \mathbb{I}_{rxr} B^{n-1} \\ &= (B^{-1})^{n-1} B^{n-1} \\ &= (B^{-1})^{n-2} B^{n-2} \\ &\vdots \\ &= B^{-1}B \\ &= \mathbb{I}_{rxr} \end{aligned}$$

con esto tenemos que

$$D^{-1}B^n = (B^{-1})^n B^n$$

además, como B es invertible podemos cancelar por la derecha, por lo que $D^{-1} = (B^{-1})^n$ y entonces podemos decir que existe $D^{\frac{-1}{n}}$. □

Matrices simétricas

Existe un tipo especial de matrices con lo que podemos tener resultados interesantes dentro de la teoría de eigenvalores y eigenvectores, estas matrices son las simétricas, pues con ellas podemos garantizar que los eigenvalores son reales y además podemos encontrar una base ortonormal de $\mathbb{R}^r$, con r la dimensión de la matriz.

Los resultados mostrados en esta sección, son usados en el capítulo 2 para crear una base ortonormal de eigenvectores izquierdos de la matriz de transición y otra de eigenvectores derechos, estas bases son usadas para la demostración de los teoremas 3.6 y 3.7.

Primeramente definiremos las matrices simétricas.

Definición **6.33.** *Matrices simétricas*
Sean $r \in \mathbb{N}$ y $A \in \mathscr{M}_{rxr}(\mathbb{R})$ una matriz cuadrada con entradas reales, decimos que A es simétrica si es igual a su transpuesta o bien si $A = (a_{ij})_{i,j\in\{1,\dots,r\}}$ entonces para cualesquiera $i, j \in \{1,\dots,r\}$ se cumple que $a_{ij} = a_{ji}$.

Ahora enunciaremos el teorema para las matrices simétricas, cabe mencionar que no daremos la demostración del teorema pero podemos encontrarla en [David C. Lay, Cap. 7].

Teorema **6.12.** *Sean $r \in \mathbb{N}$ y $A \in \mathscr{M}_{rxr}(\mathbb{R})$, supongamos que A es simétrica, entonces*

1. *A tiene r eigenvalores reales contando multiplicidades.*
2. *La dimensión de los eigenespacios de cada eigenvalor λ es igual a la multiplicidad de λ como raíz del polinomio característico.*
3. *Los eigenespacios son mutuamente ortogonales, en el sentido de que eigenvectores correspondientes a diferentes eigenvalores son ortogonales.*
4. *A es ortogonalmente diagonalizable.*

Apéndice B: Espacios métricos

A lo largo de este trabajo usamos la definición (3.18), distancia en variación, para encontrar cotas para la velocidad de convergencia de la matriz de transición a n pasos, $n \in \mathbb{N}$, de una cadena de Markov irreducible y apériodica con espacio de estados finito hacia su distribución estacionaria, en este apartado demostraremos algunas de las propiedades usadas.

Comenzaremos con la definición de un espacio métrico

Definición **6.34.** *Espacio métrico*
Sea X un conjunto no vacío, una métrica para X es una función d que va de $X \times X$ en $\mathbb{R}$ de tal forma que para cualquier $x, y, z \in X$,

a. $d(x,y) \geq 0$

b. $d(x,y) = 0$ *si sólo si* $x = y$

c. $d(x,y) = d(y,x)$

d. $d(x,z) \leq d(x,y) + d(y,z)$

entonces decimos que (X,d) es un espacio métrico.

Proposición **6.16.**
Sea E un conjunto numerable y X el conjunto de las distribuciones de probabilidad sobre E, además sea d_v la distancia en variación definida como en la definición (3.18), entonces (X, d_v) es un espacio métrico.

Demostración.
Sea $\alpha, \beta, \gamma \in X$, entonces

a. Como $\mid \alpha(i) - \beta(i) \mid \geq 0$ para toda $i \in E$, entonces $\frac{1}{2}\sum_{i \in E} \mid \alpha(i) - \beta(i) \mid \geq 0$, por lo que $d_v(\alpha, \beta) = \frac{1}{2}\sum_{i \in E} \mid \alpha(i) - \beta(i) \mid \geq 0$.

b. Observemos que $d_v(\alpha, \beta) = 0$ si y sólo si $\frac{1}{2}\sum_{i \in E} \mid \alpha(i) - \beta(i) \mid = 0$, pero esta última igualdad es cierta si y solo si para todo $i \in E$ $\mid \alpha(i) - \beta(i) \mid = 0$, el cual es verdadero si y sólo si para todo $i \in E$ $\alpha(i) = \beta(i)$, por lo que $d_v(\alpha, \beta) = 0$ si y sólo si $\alpha = \beta$.

c. Como $\mid \alpha(i)-\beta(i) \mid = \mid \beta(i)-\alpha(i) \mid$, entonces

$$\begin{aligned} d_v(\alpha,\beta) &= \frac{1}{2}\sum_{i\in E} \mid \alpha(i)-\beta(i) \mid \\ &= \frac{1}{2}\sum_{i\in E} \mid \beta(i)-\alpha(i) \mid \\ &= d_v(\beta,\alpha) \end{aligned}$$

d. Solo nos falta demostrar la desigualdad triangular, así

$$\begin{aligned} d_v(\alpha,\beta) &= \frac{1}{2}\sum_{i\in E} \mid \alpha(i)-\beta(i) \mid \\ &= \frac{1}{2}\sum_{i\in E} \mid (\alpha(i)-\gamma(i))+(\gamma(i)-\beta(i)) \mid \\ &\le \frac{1}{2}\left(\sum_{i\in E} \mid \alpha(i)-\gamma(i) \mid + \sum_{i\in E} \mid \gamma(i)-\beta(i) \mid\right) \\ &= \frac{1}{2}\sum_{i\in E} \mid \alpha(i)-\gamma(i) \mid + \frac{1}{2}\sum_{i\in E} \mid \gamma(i)-\beta(i) \mid \\ &= d_v(\alpha,\gamma)+d_v(\gamma,\beta) \\ \text{por lo que } \; d_v(\alpha,\beta) &\le d_v(\alpha,\gamma)+d_v(\gamma,\beta) \end{aligned}$$

por lo tanto (X,d_v) es un espacio métrico.

$\square$

Probaremos otras propiedades que la distancia en variación tiene, estas propiedades son formas equivalentes de definir esta distancia.

Proposición **6.17.**
Sea E un conjunto finito y $y: E \to \mathbb{R}$ una función tal que $\sup_{i\in E} \mid y(i) \mid = 1$, además sean α y β dos distribuciones de probabilidad sobre E, entonces

$$d_v(\alpha,\beta) = \frac{1}{2}\sup\left(\sum_{i\in E}(\alpha(i)y(i)) - \sum_{i\in E}(\beta(i)y(i))\right)$$

Demostración.
Sea $A = \{x \in E \mid \alpha(x) > \beta(x)\}$ y definimos

$$y(x) = \begin{cases} 1 & si \quad x \in A \\ -1 & si \quad x \notin A \end{cases}$$

por la definición del conjunto A y la de la función y, tenemos que para toda $x \in E$

$$\mid \alpha(x)-\beta(x) \mid = (\alpha(x)-\beta(x))y(x)$$

por lo que

$$\sum_{j\in E}(\alpha(j)-\beta(j))y(j) = \sum_{j\in A} \mid \alpha(j)-\beta(j) \mid$$
$$= 2d_v(\alpha,\beta)$$

por lo tanto

$$d_v(\alpha,\beta) = \frac{1}{2}\sup\left(\sum_{i\in E}(\alpha(i)y(i)) - \sum_{i\in E}(\beta(i)y(i))\right)$$

□

Lema **3.**
Sean α y β dos distribuciones de probabilidad sobre un mismo conjunto numerable E, entonces

$$d_v(\alpha,\beta) = \sup_{A\subset E}(\mid \alpha(A)-\beta(A) \mid)$$

donde $\alpha(A) := \sum_{i\in A}\alpha(i)$

Demostración.
Sea $A \subset E$, entonces existe $B \subset E$ de tal forma que

$$\mid \alpha(A)-\beta(A) \mid = \alpha(B)-\beta(B) \tag{6.61}$$

La B de la que nos habla puede ser A o $A^c = E\setminus A$, según nos convenga.

Ahora notemos que podemos escribir la diferencia entre $\alpha(A)$ y $\beta(A)$ como sigue

$$\alpha(A)-\beta(A) = \sum_{i\in E}\mathbb{1}_A(i)\,(\alpha(i)-\beta(i)) \tag{6.62}$$

Además notemos que para cualquier $A \subset E$, tenemos que

$$\sum_{i\in E}\mathbb{1}_A(i)\,(\alpha(i)-\beta(i)) + \sum_{i\in E}\mathbb{1}_{A^c}(i)\,(\alpha(i)-\beta(i)) = \sum_{i\in E}(\alpha(i)-\beta(i)) = 0 \tag{6.63}$$

Sea $A = \{i\in E; \alpha(i) > \beta(i)\}$, entonces el lado derecho de la ecuación (6.62) se maximiza, además usando las ecuaciones (6.62) y (6.63) tenemos que

$$\alpha(i)-\beta(i) = \begin{cases} \mid \alpha(i)-\beta(i) \mid & si \quad i\in A \\ -\mid \alpha(i)-\beta(i) \mid & si \quad i\notin A \end{cases} \tag{6.64}$$

por lo que

$$\sum_{i\in E}\mathbb{1}_A(i)\,(\alpha(i)-\beta(i)) = \sum_{i\in E}\mathbb{1}_A(i) \mid \alpha(i)-\beta(i) \mid \tag{6.65}$$

por otro lado

$$
\begin{aligned}
\sum_{i\in E} \mathbb{1}_A(i)\,(\alpha(i)-\beta(i)) &= -\sum_{i\in E} \mathbb{1}_{A^c}(i)\,(\alpha(i)-\beta(i)) && \text{por la ecuación (6.63)}\\
&= -\sum_{i\in E} \mathbb{1}_{A^c}(i)(-\mid \alpha(i)-\beta(i)\mid)\\
&= \sum_{i\in E} \mathbb{1}_{A^c}(i)\mid \alpha(i)-\beta(i)\mid\\
&= \frac{1}{2}\left(\sum_{i\in E} \mathbb{1}_{A}(i)\mid \alpha(i)-\beta(i)\mid + \sum_{i\in E} \mathbb{1}_{A^c}(i)\mid \alpha(i)-\beta(i)\mid\right)\\
&= \frac{1}{2}\sum_{i\in E} \mid \alpha(i)-\beta(i)\mid\\
&= d_v(\alpha,\beta)
\end{aligned}
$$

por lo tanto $d_v(\alpha,\beta) = \sup_{A\subset E}(\mid \alpha(A)-\beta(A)\mid)$ □

Apéndice C: Teorema de Rayleigh

Para este apéndice trabajaremos con matrices de transición reversibles, además ordenaremos a los eigenvalores de nuestra matriz de transición P con respecto al orden de los reales, es decir $1 = \lambda_1 > \lambda_2 \geq \ldots \geq \lambda_r \geq -1$, a menos que se indique lo contrario. Cabe destacar que este orden es el establecido en el capítulo 4, en donde λ_2 es el segundo eigenvalor más grande, SLE por sus siglas en inglés, mientras que en los capítulos anteriores a este λ_2 representa al segundo eigenvalor con módulo más grande, SLEM por sus siglas en inglés.

Antes de enunciar el Teorema de Rayleigh y sus corolarios, daremos algunas definiciones y probaremos algunos resultados que nos serán de utilidad.

Definición **6.35.** *Laplaciano de P*
Sea P la matriz de transición de una cadena de Markov con espacio de estados finito, definimos el Laplaciano *de P como la matriz $\mathbb{I} - P$.*

Proposición **6.18.**
Sean $r \in \mathbb{N}$ y P la matriz de transición de una cadena de Markov sobre un espacio de estados finito $E = \{1, \ldots, r\}$, además sean λ_i, los eigenvalores de P y v_i un eigenvector de P asociado a λ_i, $i \in E$. Sean β_i, $i \in E$, los eigenvalores del Laplaciano de P, entonces para toda $i \in E$ $\beta_i = 1 - \lambda_i$ y el eigenvector asociado a β_i es el mismo que el asociado a λ_i, es decir, v_i un eigenvector del Laplaciano de P asociado a β_i.

Demostración.
Sea $i \in E$, entonces

$$\begin{aligned}(\mathbb{I}_{rxr} - P)\, v_i &= v_i - P v_i \\ &= v_i - \lambda_i v_i \\ &= (1 - \lambda_i) v_i \\ &= \beta_i v_i\end{aligned}$$

por lo que para toda $i \in E$, v_i es un eigenvector del Laplaciano de P asociado al eigenvalor β_i. $\square$

Antes de enunciar el siguiente lema tenemos que recordar la definición 3.22, la forma de Dirichlet, del capítulo 3 de este trabajo, pues es un elemento importante tanto para el lema como para las siguientes propiedades.

Lema **4.**
Sean $r \in \mathbb{N}$ y P la matriz de transición de una cadena de Markov aperiódica e irreducible, sobre un espacio de estados finito $E = \{1,...,r\}$ y sea π su distribución estacionaria. Sean $\lambda_1,...,\lambda_r$ los eigenvalores de P ordenados de modo que $|\lambda_i| \geq |\lambda_j|$ para todo $1 \leq i < j \leq r$.
Supongamos que (P,π) es reversible, entonces para cualquier $x \in \mathbb{R}^r$, x no constante

$$\frac{\varepsilon_\pi(x,x)}{Var_\pi(x)} \geq 1 - |\lambda_2| \tag{6.66}$$

Demostración.
Como P es irreducible sobre un espacio de estados finito, entonces existe una única distribución estacionaria π con entradas estrictamente positivas, así usando la distribución estacionaria de P podemos definir la norma π (Definición 2.16).
Sea $V = \{v_1 = \mathbb{1}_r, \ldots, v_r\}$ una base ortonormal de eigenvectores derechos de P (ortonormal con respecto a la norma π).
Haremos uso de la ecuación (2.15) que nos permite representar a un vector $z \in \mathbb{R}^r$ en términos del producto interior π y los elementos de V, también usaremos la ecuación (2.17) que nos permite representar el producto de un vector $z \in \mathbb{R}^r$ con P, en términos del producto interior π, los eigenvalores de P y los elementos de V.
Sea $x \in \mathbb{R}^r$, x no constante, entonces por la ecuación (2.15)

$$x = \sum_{i \in E} \langle x, v_i \rangle_\pi v_i \tag{6.67}$$

y por la ecuación (2.17)

$$(\mathbb{I}_{rxr} - P)x = \sum_{i \in E} (1 - \lambda_i) \langle x, v_i \rangle_\pi v_i. \tag{6.68}$$

Ahora podemos obtener una cota inferior de la forma de Dirichlet de x

$$\begin{aligned}
\varepsilon_\pi(x,x) &= \langle (\mathbb{I}_{rxr} - P)x, x \rangle_\pi \\
&= \langle \sum_{i \in E} (1 - \lambda_i) \langle x, v_i \rangle_\pi v_i, x \rangle_\pi && \text{por la ecuación (6.68)} \\
&= \sum_{i \in E} (1 - \lambda_i) \langle x, v_i \rangle_\pi^2 \\
&= \sum_{i=2}^{r} (1 - \lambda_i) \langle x, v_i \rangle_\pi^2 && \text{porque } \lambda_1 = 1
\end{aligned}$$

$$\varepsilon_\pi(x,x) \geq \sum_{i=2}^{r} (1 - |\lambda_i|) \langle x, v_i \rangle_\pi^2 \tag{6.69}$$

Además, podemos obtener una ecuación equivalente de la norma π de x

$$\begin{aligned}
\| x \|_\pi^2 &= \langle x,x\rangle_\pi \\
&= \langle \sum_{i\in E} \langle x,v_i\rangle_\pi v_i, \sum_{j\in E} \langle x,v_j\rangle_\pi v_j\rangle_\pi && \text{por la ecuación (6.67)} \\
&= \sum_{i\in E}\sum_{j\in E} \langle x,v_i\rangle_\pi \langle x,v_j\rangle_\pi \langle v_i,v_j\rangle_\pi \\
&= \sum_{i\in E} \langle x,v_i\rangle_\pi^2,
\end{aligned}$$

con lo que la varianza π de x es

$$\begin{aligned}
Var_\pi(x) &= \| x \|_\pi^2 - \langle x, \mathbb{1}_r\rangle_\pi^2 \\
&= \sum_{i\in E} \langle x,v_i\rangle_\pi^2 - \langle x, \mathbb{1}_r\rangle_\pi^2 \\
&= \sum_{i=2}^{r} \langle x,v_i\rangle_\pi^2 && \text{debido a que } v_1 = \mathbb{1}_r
\end{aligned}$$

Como $Var_\pi(x) \neq 0$, debido a que x no es constante, entonces

$$\begin{aligned}
\frac{\varepsilon_\pi(x,x)}{Var_\pi(x)} &\geq \frac{\sum_{i=2}^{r}(1-\mid \lambda_i \mid)\langle x,v_i\rangle_\pi^2}{\sum_{i=2}^{r}\langle x,v_i\rangle_\pi^2} \\
&= \frac{\sum_{i=2}^{r}(1-\mid \lambda_i \mid)\langle x,v_i\rangle_\pi^2}{\sum_{i=2}^{r}\langle x,v_i\rangle_\pi^2} + (1-\mid \lambda_2 \mid) - (1-\mid \lambda_2 \mid) \\
&= (1-\mid \lambda_2 \mid) + \frac{\sum_{i=2}^{r}(1-\mid \lambda_i \mid -1+\mid \lambda_2 \mid)\langle x,v_i\rangle_\pi^2}{\sum_{i=2}^{r}\langle x,v_i\rangle_\pi^2} \\
&= (1-\mid \lambda_2 \mid) + \frac{\sum_{i=2}^{r}(\mid \lambda_2 \mid - \mid \lambda_i \mid)\langle x,v_i\rangle_\pi^2}{\sum_{i=2}^{r}\langle x,v_i\rangle_\pi^2} \\
&\geq (1-\mid \lambda_2 \mid)
\end{aligned}$$

□

Proposición **6.19.**
Sea P la matriz de transición de una cadena de Markov irredducible y aperiódica, sobre un espacio de estados finito E y sea π su distribución estacionaria. Además, supongamos que (P,π) es reversible, entonces

$$\varepsilon_\pi(x,x) = \frac{1}{2}\sum_{i,j\in E} \pi(i) p_{ij} \left(x(j) - x(i)\right)^2$$

<u>*Demostracion*</u>

$$\begin{aligned}
\varepsilon_\pi(x,x) &= \langle (I-P)x, x\rangle_\pi \\
&= \langle x,x\rangle_\pi - \langle Px,x\rangle_\pi \\
&= \sum_{i\in E}\left(\pi(i)x^2(i) - \sum_{j\in E} p_{ij}x(j)x(i)\pi(i)\right) \\
&= \sum_{i\in E}\left(\sum_{j\in E} p_{ij}\pi(i)x^2(i) - \sum_{j\in E} p_{ij}x(j)x(i)\pi(i)\right) \\
&= \sum_{i,j\in E}\left(p_{ij}\pi(i)x^2(i) - p_{ij}x(j)x(i)\pi(i)\right) \\
&= \sum_{i,j\in E}\pi(i)p_{ij}x(i)\left(x(i)-x(j)\right) \\
&= \sum_{i,j\in E}\pi(j)p_{ji}x(i)\left(x(i)-x(j)\right) && \text{Por } (P,\pi) \text{ reversible.} \\
&= \sum_{i,j\in E}\pi(i)p_{ij}x(j)\left(x(j)-x(i)\right) && \text{Invirtiendo los indices.}
\end{aligned}$$

Por lo que $\varepsilon_\pi(x,x) = \sum_{i,j\in E}\pi(i)p_{ij}x(i)\left(x(i)-x(j)\right) = \sum_{i,j\in E}\pi(i)p_{ij}x(j)\left(x(j)-x(i)\right)$, entonces

$$\begin{aligned}
\varepsilon_\pi(x,x) &= \frac{1}{2}\sum_{i,j\in E}\pi(i)p_{ij}x(i)\left(x(i)-x(j)\right) + \frac{1}{2}\sum_{i,j\in E}\pi(i)p_{ij}x(j)\left(x(j)-x(i)\right) \\
&= \frac{1}{2}\sum_{i,j\in E}\pi(i)p_{ij}x(i)\left(x(i)-x(j)\right) - \frac{1}{2}\sum_{i,j\in E}\pi(i)p_{ij}x(j)\left(x(i)-x(j)\right) \\
&= \frac{1}{2}\sum_{i,j\in E}\pi(i)p_{ij}\left(x(j)-x(i)\right)^2
\end{aligned}$$

$\therefore\ \varepsilon_\pi(x,x) = \frac{1}{2}\sum_{i,j\in E}\pi(i)p_{ij}\left(x(j)-x(i)\right)^2$ ⊠

Proposición **6.20.**
Una de las propiedades de la forma de Dirichlet es la siguiente:
$\varepsilon_\pi(x,x) = \varepsilon_\pi(x-c\mathbb{1}, x-c\mathbb{1})$ *para toda* $c\in\mathbb{R}$.

<u>*Demostracion*</u>

Sea $c \in \mathbb{R}$ y $x \in \mathbb{R}^r$

$$\begin{aligned}
\Rightarrow \varepsilon_\pi(x - c\mathbb{1}, x - c\mathbb{1}) &= \frac{1}{2}\sum_{i,j\in E} \pi(i)p_{ij}\left((x(j) - c) - (x(i) - c)\right)^2 \quad \text{Por la proposición anterior} \\
&= \frac{1}{2}\sum_{i,j\in E} \pi(i)p_{ij}\left(x(j) - \cancel{c} - x(i) + \cancel{c}\right)^2 \\
&= \frac{1}{2}\sum_{i,j\in E} \pi(i)p_{ij}\left(x(j) - X(i)\right)^2 \\
&= \varepsilon_\pi(x, x)
\end{aligned}$$

$\therefore \varepsilon_\pi(x,x) = \varepsilon_\pi(x - c\mathbb{1}, x - c\mathbb{1})\ \forall c \in \mathbb{R}$ $\boxtimes$

Uno de los propósitos es acotar al *SLEM*, el siguiente Teorema nos dará una caracterización de los eigenvalores del Laplaciano de P, es decir de β_i. Con ello podremos dar una cota para λ_2 y λ_r, así podremos conseguir una buena cota del *SLEM*.

Teorema **6.13.** *Rayleigh*
Sea P la matriz de transición de una cadena de Markov irredducible, sobre un espacio de estados finito E y sea π su distribución estacionaria. Además, supongamos que (P,π) es reversible, entonces

$$\beta_j = inf\left\{\frac{\varepsilon_\pi(x,x)}{Var_\pi(x)} \mid \langle x, v_i\rangle_\pi = 0\ \forall i \in \{1,\ldots,j-1\}, x \neq 0\right\}\ \forall j \geq 2$$

$\underline{Demestracion|}$
Recordemos que podemos escoger al conjunto de eigenvectores derechos, $\{v_1,\ldots,v_r\}$ como una base ortonormal de $\mathbb{R}^r$, con lo que podemos usar las ecuaciones (2.17) y (2.15).

Sea $x \in \mathbb{R}^r$
$\Rightarrow (\mathbb{I} - P) = \sum_{i\in E}\beta_i\langle x, v_j\rangle_\pi v_j$
Con lo que podemos dar una ecuación equivalente de la forma de Dirichlet y de la norma de x bajo π

$$\begin{aligned}
\varepsilon_\pi(x,x) &= \langle(\mathbb{I} - P)x, x\rangle_\pi \\
&= \langle\sum_{i\in E}\beta_i\langle x, v_j\rangle_\pi v_j, x\rangle_\pi \\
&= \sum_{i\in E}\beta_i\langle x, v_j\rangle_\pi^2
\end{aligned}$$

$$\varepsilon_\pi(x,x) = \sum_{i\in E}\beta_i\langle x, v_j\rangle_\pi^2 \tag{6.70}$$

$$\begin{aligned}\| x \|^2 &= \langle x,x\rangle_\pi \\ &= \langle \sum_{i\in E} \langle x,v_i\rangle_\pi v_j, x\rangle_\pi \\ &= \sum_{i\in E} \langle x,v_i\rangle_\pi^2\end{aligned}$$

$$\langle x,x\rangle_\pi = \sum_{i\in E} \langle x,v_i\rangle_\pi^2 \tag{6.71}$$

Sea $j \in \{2,\dots,r\}$, supongamos que $\langle x,v_k\rangle_\pi = 0$ para toda $k \in \{1,\dots,j-1\}$, entonces en particular

$$\begin{aligned}\langle x\rangle_\pi &= \langle x,\mathbb{1}\rangle_\pi \\ &= \langle x,v_1\rangle_\pi \\ &= 0\end{aligned}$$

Como $\langle x\rangle_\pi = 0$ entonces $Var_\pi(x) = \| x \|^2 = \langle x,x\rangle_\pi$

$$\begin{aligned}\Rightarrow \frac{\varepsilon_\pi(x,x)}{Var_\pi(x)} &= \frac{\sum_{i\in E} \beta_i \langle x,v_j\rangle_\pi^2}{\sum_{i\in E} \langle x,v_i\rangle_\pi^2} \\ &= \frac{\sum_{i=j}^r \beta_i \langle x,v_j\rangle_\pi^2}{\sum_{i=j}^r \langle x,v_i\rangle_\pi^2} \\ &= \beta_j - \beta_j + \frac{\sum_{i=j}^r \beta_i \langle x,v_j\rangle_\pi^2}{\sum_{i=j}^r \langle x,v_i\rangle_\pi^2} \\ &= \beta_j + \frac{\sum_{i=j+1}^r (\beta_i - \beta_j) \langle x,v_j\rangle_\pi^2}{\sum_{i=j}^r \langle x,v_i\rangle_\pi^2} \\ &\geq \beta_j \qquad \text{Debido a que } (\beta_i - \beta_j) \geq 0 \; \forall i \geq j\end{aligned}$$

Como x es arbitrario, entonces

$$\beta_j \leq inf \left\{ \frac{\varepsilon_\pi(x,x)}{Var_\pi(x)} \mid \langle x,v_i\rangle_\pi = 0 \; \forall i \in \{1,\dots,j-1\}, x \neq 0 \right\} \tag{6.72}$$

Una observación importante es que el vector que hace que $\frac{\varepsilon_\pi(x,x)}{Var_\pi(x)}$ alcance su ínfimo en β_j es el eigenvector asociado a este, es decir, v_j. Como $v_j \neq 0$ y además $\langle v_j, v_i\rangle_\pi = 0 \; \forall i \in \{1,\dots,j-1\}$, entonces

$$inf \left\{ \frac{\varepsilon_\pi(x,x)}{Var_\pi(x)} \mid \langle x,v_i\rangle_\pi = 0 \; \forall i \in \{1,\dots,j-1\}, x \neq 0 \right\} \leq \frac{\varepsilon_\pi(v_j,v_j)}{Var_\pi(v_j)} \tag{6.73}$$

Ahora calculemos

$$\begin{aligned}\varepsilon_\pi(v_j,v_j) &= \sum_{i\in E} \beta_j \langle v_j,v_i\rangle_\pi^2 && \text{Por la ecuación (6.70)} \\ &= \beta_j \langle v_j,v_j\rangle_\pi^2 && \text{Por la orgonalidad de las } v_i \\ &= \beta_j && \text{Ya que } \langle v_j,v_j\rangle_\pi = 1\end{aligned}$$

$$\therefore \varepsilon_\pi(v_j, v_j) = \beta_j \tag{6.74}$$

Además

$$\begin{aligned} Var_\pi(v_j) &= \sum_{i\in E} \langle v_j, v_j \rangle_\pi^2 && \text{Por la eciación (6.71)} \\ &= 1 && \text{Por la ortonormalidad de las } v_i \end{aligned}$$

$$\therefore Var_\pi(v_j) = 1 \tag{6.75}$$

Por lo que usando las ecuaciones (6.74) y (6.75) tenemos que

$$\frac{\varepsilon_\pi(v_j, v_j)}{Var_\pi(v_j)} = \frac{\beta_j}{1} = \beta_j$$

$$\Rightarrow inf \left\{ \frac{\varepsilon_\pi(x,x)}{Var_\pi(x)} \mid \langle x, v_i \rangle_\pi = 0 \ \forall i \in \{1, \ldots, j-1\}, x \neq 0 \right\} \leq \frac{\varepsilon_\pi(v_j, v_j)}{Var_\pi(v_j)} = \beta_j$$

$$\therefore \qquad \beta_j = inf \left\{ \frac{\varepsilon_\pi(x,x)}{Var_\pi(x)} \mid \langle x, v_i \rangle_\pi = 0 \ \forall i \in \{1, \ldots, j-1\}, x \neq 0 \right\} \ \forall j \geq 2$$

$\boxtimes$

Usando el Teorema de *Rayleigh* para β_2 tenemos que

$$\begin{aligned} \beta_2 &= inf \left\{ \frac{\varepsilon_\pi(x,x)}{Var_\pi(x)} \mid \langle x \rangle_\pi = 0, \ x \neq 0 \right\} \\ &= inf \left\{ \frac{\varepsilon_\pi(x,x)}{Var_\pi(x)} \mid x \text{ es no constante} \right\} \end{aligned}$$

Es fácil de ver la igualdad, teniendo en cuenta que el vector que hace que $\frac{\varepsilon_\pi(x,x)}{Var_\pi(x)}$ alcance su ínfimo en β_2 es el eigenvector asociado a este, es decir, v_2. Y este eigenvector, por tener media respecto a π igual a cero, no puede ser constante.

La fuerza que tiene el Teorema de Rayleigh, radica en los siguientes dos corolarios, los cuales nos ayudarán a encontrar una cota superior para λ_2 y una cota inferior para λ_r, esto con el objetivo de acotar al *SLEM*.

Corolario **1.**
Sea P la matriz de transición de una cadena de Markov irredducible y aperiódica, sobre un espacio de estados finito E y sea π su distribución estacionaria. Además, supongamos que (P, π) es reversible. Sea $A \in \mathbb{R}$ no negativo de tal forma que $Var_\pi(x) \leq A\varepsilon_\pi(x,x)$ para toda $x \in \mathbb{R}^r$, entonces

$$\lambda_2 \leq 1 - \frac{1}{A} \tag{6.76}$$

con λ_2 el SLE

<u>*Demostracion*</u>|
Como $Var_\pi(x) \leq A\varepsilon_\pi(x,x)$ para toda $x \in \mathbb{R}^r$, en particular para $x \in \mathbb{R}^r$ que tienen media respecto a π igual a cero y no son el vector cero.
$\Rightarrow \frac{1}{A} \leq \frac{\varepsilon_\pi(x,x)}{Var_\pi(x)}$ $\forall x \in \mathbb{R}^r$ tal que $\langle x \rangle_\pi = 0$ y $x \neq 0$
$\Rightarrow \frac{1}{A} \leq inf \left\{ \frac{\varepsilon_\pi(x,x)}{Var_\pi(x)} \mid \langle x \rangle_\pi = 0,\ x \neq 0 \right\} = \beta_2$
Pero recordemos que $\beta_2 = 1 - \lambda_2$, entonces $\frac{1}{A} \leq 1 - \lambda_2$

$$\therefore \qquad \lambda_2 \leq 1 - \frac{1}{A}$$

$\boxtimes$

Corolario **2.**
Sea P la matriz de transición de una cadena de Markov irredducible y aperiódica, sobre un espacio de estados finito E y sea π su distribución estacionaria. Además, supongamos que (P,π) es reversible. Sea $B \in \mathbb{R}$ no negativo de tal forma que $\langle Px,x \rangle_\pi + \parallel x \parallel_\pi^2 \geq B \parallel x \parallel_\pi^2$ para toda $x \in \mathbb{R}^r$, entonces

$$\lambda_r \geq -1 + B \tag{6.77}$$

Con λ_r el eigenvalor más pequeño

<u>*Demostracion*</u>
Como $\langle Px,x \rangle_\pi + \parallel x \parallel_\pi^2 \geq B \parallel x \parallel_\pi^2$ para toda $x \in \mathbb{R}^r$, en particular para $x = v_r$

$$\begin{aligned} &\Rightarrow \langle Pv_j, v_j \rangle_\pi + \parallel v_j \parallel_\pi^2 && \geq B \parallel v_j \parallel_\pi^2 \\ &\Rightarrow \lambda_j \langle v_j, v_j \rangle_\pi + 1 && \geq B \end{aligned}$$

$$\therefore \lambda_j \geq -1 + B$$

$\boxtimes$

Bibliografía

[Bremaud] BRÉMAUD, P. (1999). "*Eigenvalues and Nohomogeneous Markov Chains*". IN P. BRÉMAUD, MARKOV CHAINS GIBBS FIELDS, MONTE CARLO SIMULATION AND QUEUES (p. 444). NEW YORK: SPRINGER-VERLAG.

[David C. Lay] DAVID C. LAY, S. R. (2016). "*Linear Algebra and its applications*". WASHINGTON: PEARSON.

[Friedberg] FRIEDBERG, S. H. (1982). "*Álgebra lineal*". MÉXICO: PUBLICACIONES CULTURAL.

[Paul G. Hoel] HOEL, P. G. (1972). "*Introduction to stochastic processes*". USA: HOUGHTON MIFFLIN.

[Hopfield] HOPFIELD J J. (1985) *"Neuronal networks and physical systems with emergent collective computational abilities"*. BIOL. CYBERNETICS.

[Iosifescu] IOSIFESCU, M. (1980). In M. Iosifescu, "*Finnite markov processes and their applications*". ROMANIA.

[Rincón L.] RINCÓN, L. (2011). Capítulo 3. Cadenas de Markov. In L. Rincón, " *Introducción a los procesos estocásticos*".

[Seneta] SENETA, E. (1973). "*Non-negative matrices and Markov chains*". LONDRES: SPRINGER.