Universidad Nacional Autónoma de México

Facultad de Ciencias

EL TEOREMA DE PERRON-FROBENIUS: UNA APLICACIÓN EN CADENAS DE MARKOV

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

Matemático

PRESENTA:

Ricardo Roberto Bustos Ortiz

TUTORA

Ana Meda Guardiola

Cd. Mx. 2023

Hoja de Datos del Jurado

1. Datos del alumno	1. Datos del alumno
Apellido paterno	Bustos
Apellido materno	Ortiz
Nombre(s)	Ricardo Roberto
Teléfono	5520813892
Universidad Nacional Autónoma de México	
Facultad de Ciencias	
Carrera	Matemáticas
Número de cuenta	412006894
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	Dra.
Nombre(s)	Martha
Apellido paterno	Takane
Apellido materno	Imay
4. Datos del sinodal 2	4. Datos del sinodal 2
Grado	Dr.
Nombre(s)	Sergio Iván
Apellido paterno	López
Apellido materno	Ortega
5. Datos del sinodal 3	5. Datos del sinodal 3
Grado	Mat.
Nombre(s)	Ernesto
Apellido paterno	Mayorga
Apellido materno	Saucedo
6. Datos del sinodal 4	6. Datos del sinodal 4
Grado	Dra.
Nombre(s)	Ruth Selene
Apellido paterno	Fuentes
Apellido materno	García
7. Datos del trabajo escrito.	7.Datos del trabajo escrito.
Título	El Teorema de Perron-Frobenius: Una aplicación en cadenas de Markov
Número de páginas	70
Año	2020

A mis padres, hermana y abuelos.
Ustedes me han dado las lecciones más importantes.

Agradecimientos

En primer lugar quiero agradecer a mi familia. A mis padres por todo el sacrificio que han hecho estos años, no pude tener un mejor ejemplo y sé perfectamente que no podría haber llegado hasta aquí sin ustedes. Me dieron todo lo que tenían y siempre los llevaré en mi corazón.

A mi hermana por ser mi compañerita de vida, esa traviesa que me dio una de las misiones más importantes: ser hermano mayor. Espero siempre poder estar cerca para cuidarte y molestarte.

A mi abuelo Gonzalo porque todavía me duele su partida, pero estoy muy agradecido de haber podido crear tantos recuerdos con él. A mi abuela Ana por ser un gran ejemplo de vida y fortaleza, y a mi abuela Silvia por todo lo que me consintió y cuidó cuando era pequeño. No podría estar más orgulloso y agradecido de llevar su sangre en mis venas.

A mi tío Gonzalo por ser ese gran amigo que ya no está conmigo, pero que junto con mi abuelo extraño cada día de mi vida.

En segundo lugar quiero agradecer a esa otra familia que he ido formando poco a poco con los años: mis amigos. Carlos, Luis Enrique, Luis Fernando y Pedro, muchas gracias por su amistad incondicional. A Iván y Axel gracias por ser los mejores duelistamigos que cualquiera podría tener.

Gracias a todos mis amigos de la Facultad de Ciencias y de la casa de estudiantes: Allan, Ricardo, Alejandro, Giovanna, Mariana, Juan, Mitzi, David, Vianey, Pepe, Ana Luisa, Marco, Lorena y Jorge. La universidad no habría sido tan legendaria sin ustedes.

Muchas gracias a la Dra. Ana Meda Guardiola por su paciencia, por no rendirse conmigo y por su gran apoyo. Gracias por el gran impulso que dio en mi vida. En este mismo tenor, agradecer el apoyo de la Dra. Martha Takane, Dra. Ruth Fuentes, Dr. Sergio Iván López y Mat. Ernesto Mayorga para concluir este proyecto.

Finalmente, también quiero agradecer al Dr. Benito Nacif, a la Dra. Luz Mijangos, al Mtro. Jorge Egren, al Mtro. Alejandro Facio y a mis más cercanos compañeros y amigos de trabajo: Pablo, Lupita, Gabriel, Jimena y Erick, de todos ustedes aprendí muchísimo. Les agradezco su amistad, y les deseo que sigan cosechando éxitos a lo largo de su vida.

Por mi raza hablará el espíritu.

Índice general

Notación

(Ω, F, P)	Espacio de probabilidad
$A^{'}$	Transpuesta de la matriz A
$G_m(x,y)$	Suma de las probabilidades de transición de x a y desde 1 hasta m pasos
I	Matriz identidad
MCD	Máximo común divisor
$M_{m\times n}(F)$	Espacio de matrices $m \times n$ con valores en un campo F
$O(f(x))$	Función que es cota superior de la función $f(x)$ a partir de algún x_0
P	Medida de probabilidad
$P(E\|F)$	Probabilidad de que el evento E ocurra dado que sucede F
S_n	Media muestral
T_y	Primer momento positivo en el que la cadena de Markov $\{X_n\}$ se encuentra en el estado y
Λ	Función generadora de momentos logarítmica
Λ^*	Transformada de Fenchel-Legendre
∞	Infinito
lím	Límite
log	Logaritmo base e
$\mathbb{C}$	Conjunto de número complejos
$\mathbb{E}[\cdot]$	Esperanza
$\mathbb{E}[\cdot\|\cdot]$	Esperanza condicional
$\mathbb{N}$	Conjunto de números naturales
$\mathbb{R}$	Conjunto de número reales
$\mathbb{R}^+$	Conjunto de números reales positivos
$\mathbb{R}^n$	Producto cartesiano de n copias de $\mathbb{R}$
$\mathbb{R}^n_*$	Conjunto de elementos no nulos en $\mathbb{R}^n$ con entradas mayores o iguales a cero
máx	Máximo
mín	Mínimo
∇	Vector gradiente
$\bar{1}$	Vector columna con todas las entradas iguales a 1

π	Distribución estacionaria
π_0	Distribución inicial
ρ_{yy}	Probabilidad de que la cadena de Markov $\{X_n\}$ visite por primera vez el estado y dado que comienza en y
a_{ij}	Entrada (i,j) de la matriz A
$adj(A)$	Adjunta clásica de la matriz A
$det(A)$	Determinante de la matriz A
f^{-1}	Imagen inversa de la función f
m_x	Tiempo esperado de retorno al estado x cuando la cadena de Markov $\{X_n\}$ comienza en x
$p_{ij}^{(m)}$	Probabilidad de pasar del estado i al estado j en m pasos
$\lvert \cdot \rvert$	Valor absoluto
$\lvert\lvert \cdot \rvert\rvert$	Norma euclideana
P.G.D.	Principio de Grandes Desviaciones

Introducción

El Álgebra Lineal es una rama de las Matemáticas con una gran cantidad de resultados aplicables en múltiples áreas de las Matemáticas, Física y otras ciencias. En esta tesis tendremos como objeto de estudio uno de estos resultados: el Teorema de Perron-Frobenius.

El Teorema de Perron Frobenius prueba que una matriz con ciertas características cuenta siempre con un eigenvalor real de multiplicidad algebraica y geométrica 1, al que llamaremos eigenvalor de Perron-Frobenius, y cuya norma es estrictamente mayor a la norma compleja de cualquier otro de sus eigenvalores. Aunque al principio este resultado puede sonar simple, a lo largo de esta tesis resultará evidente la importancia de este eigenvalor y de los eigenvectores asociados al mismo, así como también algunas de sus propiedades.

El primer capítulo estará dedicado a enunciar y demostrar el teorema de Perron-Frobenius, además de probar también algunos resultados que se derivan del mismo y que serán útiles en capítulos posteriores. Cabe mencionar que la misma demostración de este teorema es bastante enriquecedora e ilustra muy bien cómo el Álgebra Lineal se interseca con múltiples áreas de las Matemáticas. Para este capítulo utilizaremos como principal referencia [9].

En el segundo capítulo nos adentraremos en el tema de cadenas de Markov, pues existe un tipo particular de estas cadenas que cuenta con una matriz de transición a la que podemos aplicar el Teorema de Perron-Frobenius. Aplicando el teorema de Perron-Frobenius a la matriz de transición de este tipo de cadenas, buscaremos expandir algunos resultados de Procesos Estocásticos. En este capítulo el material que utilizaremos como apoyo será, principalmente, [7].

Finalmente, concluimos en el tercer capítulo con una aplicación en teoría de Grandes Desviaciones de los resultados derivados del Teorema de Perron-Frobenius. El estudio de Grandes Desviaciones busca, a grandes rasgos, analizar la velocidad de convergencia a cero de probabilidades de “eventos raros” y, en el caso particular del tipo de cadenas introducido en el capítulo anterior, veremos cómo esta velocidad de convergencia está relacionada al eigenvalor de Perron-Frobenius de una matriz relacionada a la matriz de transición. En este capítulo utilizaremos como material principal a [5].

Capítulo 1

El Teorema de Perron-Frobenius

Este capítulo estará dedicado a demostrar el teorema de Perron-Frobenius, el cual se aplica a un tipo de matrices llamadas matrices regulares. En la primera sección definiremos este concepto junto con otros relacionados, y además adoptaremos algunas convenciones que serán útiles en la demostración del teorema de Perron-Frobenius.

En la segunda sección procederemos a enunciar y demostrar el teorema de Perron-Frobenius.

Definiciones

Definición 1. *Sea $A = (a_{ij}) \in M_{n\times n}(\mathbb{R})$. Decimos que A es una **matriz no negativa** si y sólo si $a_{ij} \geq 0$ para toda $i, j \in \{1,2,\dots,n\}$. En este caso, diremos que $A \geq 0$.*

*Si en particular se tiene que $a_{ij} > 0$ para toda $i, j \in \{1,2,\dots,n\}$, diremos que la matriz A es **estrictamente positiva** y lo denotaremos como $A > 0$.*

El teorema de Perron-Frobenius se aplica a las matrices no negativas con la propiedad de que existe una potencia para la cual la matriz elevada a dicha potencia sólo contiene entradas estrictamente positivas. Este tipo de matrices se define como sigue:

Definición 2. *Sea $A \in M_{n\times n}(\mathbb{R})$ una matriz no negativa. Decimos que A es una **matriz regular** si y sólo si existe $k \in \mathbb{N}$ tal que A^k es estrictamente positiva.*

Ejemplo 1. Sea $A = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}$. Podemos observar que A es no negativa y que

$$\begin{aligned} A^2 &= \begin{pmatrix} 1 & 2 & 0 \\ 0 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \\ 0 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix} \\ &= \begin{pmatrix} 1+0+0 & 2+4+0 & 0+6+0 \\ 0+0+3 & 0+4+6 & 0+6+9 \\ 1+0+3 & 2+4+6 & 0+6+9 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 6 & 6 \\ 3 & 10 & 15 \\ 4 & 12 & 15 \end{pmatrix}, \end{aligned}$$

por lo tanto la matriz A es regular.

Es importante observar que si una matriz A es regular, entonces A es distinta de cero (pues si $A = 0$, tendríamos que $A^m = 0$ para cualquier m). Más aún, se tiene lo siguiente:

Lema 1. *Sea $A \in M_{n\times n}(\mathbb{R})$ una matriz. Si A es una matriz regular, entonces cada fila de la matriz A contiene al menos un elemento distinto de cero.*

Demostración.

Sea $A \in M_{n\times n}(\mathbb{R})$ una matriz regular, $A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$, y supongamos que existe una fila en la matriz A donde todas las entradas son iguales a cero, es decir, existe $\alpha \in \{1,...,n\}$ tal que $a_{\alpha j} = 0$ para toda $j = 1,...,n$.

Si para cada $m \in \mathbb{N}$ denotamos como $a_{ij}^{(m)}$ a la ij-ésima entrada de la matriz A^m, se puede observar que

$$\textit{para toda } m \in \mathbb{N} \textit{ se tiene que } a_{\alpha j}^{(m)} = 0 \textit{ para cada } j = 1,...,n. \tag{1.1}$$

Realizaremos la prueba de (1.1) por inducción:

- $A^1 = A$, y como $a_{\alpha j} = 0$ para cada $j = 1,...,n$, la propiedad (1.1) queda demostrada para $m = 1$.
- Supongamos que existe alguna $k \in \mathbb{N}$ tal que $a_{\alpha j}^{(k)} = 0$ para cada $j = 1,...,n$. Como

$$\begin{aligned} A^{k+1} &= A^k \cdot A \\ &= \begin{pmatrix} a_{11}^{(k)} & a_{12}^{(k)} & \dots & a_{1n}^{(k)} \\ a_{21}^{(k)} & a_{22}^{(k)} & \dots & a_{2n}^{(k)} \\ \vdots & \vdots & & \vdots \\ a_{n1}^{(k)} & a_{n2}^{(k)} & \dots & a_{nn}^{(k)} \end{pmatrix} \cdot \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}, \end{aligned}$$

para cada $u \in \{1,...,n\}$ se tiene que

$$\begin{aligned} a^{k+1}_{\alpha u} &= a^{(k)}_{\alpha 1} a_{1u} + a^{(k)}_{\alpha 2} a_{2u} + ... + a^{(k)}_{\alpha n} a_{nu} \\ &= 0 \cdot a_{1u} + 0 \cdot a_{2u} + ... + 0 \cdot a_{nu} \\ &= 0, \end{aligned}$$

por lo que $a^{(k+1)}_{\alpha u} = 0$ para toda $u = 1,...,n$.

Esto concluye la demostración de la propiedad (1.1), es decir, para toda $m \in \mathbb{N}$ se tiene que $a^{(m)}_{\alpha j} = 0$ para cada $j = 1,...,n$. Esto es una contradicción, pues A es una matriz regular y por lo tanto debería existir una m para la cual A^m sea estrictamente positiva, contradicción que viene de suponer que la matriz A tiene una fila igual a cero. Por lo tanto, cada fila de la matriz A tiene al menos un elemento distinto de cero. □

Similarmente, si una matriz $A \in M_{n\times n}(\mathbb{R})$ es regular entonces cualquier columna de la matriz A contiene al menos una entrada positiva, lo cual se puede probar de forma análoga a la demostración del lema 1.

Por otro lado, para continuar debemos adoptar las siguientes convenciones:

- De manera natural utilizaremos la identificación $\mathbb{R}^n \cong M_{n\times 1}(\mathbb{R})$ para trasladar el orden de la definición 1 a $\mathbb{R}^n$.
- Denotaremos como $\mathbb{R}^n_*$ al conjunto de elementos no negativos en $\mathbb{R}^n$ distintos de cero. Es decir, $\mathbb{R}^n_* = \{x \in \mathbb{R}^n \mid x \geq 0 \text{ y } x \neq 0\}$.
- Sean A y B dos elementos de $M_{n\times m}(\mathbb{R})$, con $A = (a_{ij})$ y $B = (b_{ij})$. Diremos que $A \geq B$ si $a_{ij} \geq b_{ij}$ para cualesquiera i, j.
  Similarmente, diremos que $A > B$ si $a_{ij} > b_{ij}$ para cualesquiera i, j.
- Dada una matriz $A = (a_{ij}) \in M_{n\times m}(\mathbb{R})$, denotaremos como $A' = (a_{ji}) \in M_{m\times n}(\mathbb{R})$ a la matriz transpuesta de A.[1]

Finalmente, extenderemos de manera natural la definición de eigenvector como sigue:

Definición 3. *Sean F un campo, $A \in M_{n\times n}(F)$ una matriz y $\lambda \in F$.*

- *Decimos que $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in F^n$, $x \neq 0$, es un* ***eigenvector derecho*** *de A asociado a λ si $Ax = \lambda x$.*

[1] Usualmente se utiliza la notación A^t para la matriz transpuesta de A, pero a lo largo de este capítulo estaremos trabajando con matrices cuadradas elevadas a distintas potencias, por lo que utilizaremos la notación A' para evitar confusión.

- *Decimos que* $y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \in F^n$, $y \neq 0$, *es un* ***eigenvector izquierdo*** *de* A *asociado a* λ *si* $y'A = \lambda y'$.

Ejemplo 2. Sean

$$A = \begin{pmatrix} 2 & 3 \\ 0 & 2 \end{pmatrix}, \ x = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \ \text{y} \ y = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Podemos observar que

$$Ax = \begin{pmatrix} 2 & 3 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix} = 2x$$

y

$$y'A = \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 3 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 0 & 2 \end{pmatrix} = 2y',$$

por lo que x es un eigenvector derecho y y es un eigenvector izquierdo de la matriz A, ambos con eigenvalor asociado 2.

Teorema de Perron-Frobenius

Teorema 1 (Teorema de Perron-Frobenius)**.** *Sea* $T \in M_{n\times n}(\mathbb{R})$ *una matriz regular. Entonces existe un eigenvalor* r *de* T *tal que:*
(a) $r \in \mathbb{R}$ *y* $r > 0$.
(b) A r *se le pueden asociar eigenvectores izquierdos y derechos estrictamente positivos.*
(c) $r > |\lambda|$ *(en norma compleja) para todo* λ *eigenvalor de* T, $\lambda \neq r$.
(d) Los eigenvectores asociados con r *son únicos, salvo por múltiplos escalares.*
(e) Si $B \in M_{n\times n}(\mathbb{R})$ *es tal que* $0 \leq B \leq T$ *y* β *es un eigenvalor de* B*, entonces* $|\beta| \leq r$. *Más aún, si* $|\beta| = r$ *entonces* $B = T$.
(f) r *es una raíz simple del polinomio característico de* T.

Demostración.
Sea $T \in M_{n\times n}(\mathbb{R})$ una matriz regular, $T = \begin{pmatrix} t_{11} & t_{12} & \dots & t_{1n} \\ t_{21} & t_{22} & \dots & t_{2n} \\ \vdots & \vdots & & \vdots \\ t_{n1} & t_{n2} & \dots & t_{nn} \end{pmatrix}$.

(a) Realizaremos la demostración de este inciso en tres pasos: primero definiremos tres funciones en $\mathbb{R}^n_*$, luego las utilizaremos para demostrar la existencia del eigenvalor real r y finalmente demostraremos que $r > 0$.

Comenzando con la primera parte de la demostración, para cada $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n_*$ definimos la función $v_T : \mathbb{R}^n_* \to \mathbb{R}^n_*$ como

$$v_T(x) = x'T.$$

Es importante mencionar que v_T está bien definida, pues para cada $x \in \mathbb{R}^n_*$ se tiene lo siguiente:

$$\begin{aligned} v_T(x) &= x'T \\ &= \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} t_{11} & t_{12} & \dots & t_{1n} \\ t_{21} & t_{22} & \dots & t_{2n} \\ \vdots & \vdots & & \vdots \\ t_{n1} & t_{n2} & \dots & t_{nn} \end{pmatrix} \\ &= \begin{pmatrix} x_1t_{11} + x_2t_{21} + \dots + x_nt_{n1} & \dots & x_1t_{1n} + x_2t_{2n} + \dots + x_nt_{nn} \end{pmatrix} \\ &= \begin{pmatrix} \sum\limits_{i=1}^{n} x_it_{i1} & \dots & \sum\limits_{i=1}^{n} x_it_{in} \end{pmatrix} \in \mathbb{R}^n. \end{aligned}$$

Cada término de estas sumas es mayor o igual a cero ya que, al ser T una matriz regular y $x \in \mathbb{R}^n_*$, se tiene que $x_i \geq 0$ y $t_{ij} \geq 0$ para cualesquiera $i, j \in \{1, \dots, n\}$. Además, como $x \in \mathbb{R}^n_*$, existe una entrada x_ℓ del vector x que es estrictamente positiva, y por el lema 1 existe una entrada $t_{\ell m}$ de la fila ℓ que es estrictamente positiva, y por lo tanto

$$0 < x_\ell t_{\ell m} \leq \sum_{i=1}^{n} x_it_{im}.$$

Con esto verificamos que una de las entradas de $v_T(x)$ es estrictamente positiva, por lo que $v_T(x)$ está contenido en $\mathbb{R}^n_* = \{x \in \mathbb{R}^n \mid x \geq 0 \ y \ x \neq 0\}$ y la función v_T está bien definida.

Para cada $j \in \{1, \dots, n\}$ denotaremos a la j-ésima entrada de $v_T(x)$ como v_j^x, es decir,

$$v_j^x = \sum_{i=1}^{n} x_it_{ij} \text{ para toda } i = 1, 2, \dots, n.$$

Continuando el primer paso de la demostración, para cada $j = 1, 2, \dots, n$ definimos la función $f_j : \mathbb{R}^n_* \to [0, \infty]$ como sigue:

$$f_j(x) = \begin{cases} \frac{v_j^x}{x_j} & \text{si } x_j \neq 0 \\ \infty & \text{si } x_j = 0 \end{cases}.$$

Esta función está bien definida porque, en el caso de que x_j sea distinto de cero, se tiene que v_j^x es mayor o igual a cero y en ese caso $f_j(x) \in [0, \infty]$.

Finalmente, la tercera función que definimos es $R : \mathbb{R}^n_* \to [0, \infty]$, definida como

$$R(x) = \min_{1 \leq j \leq n} f_j(x).$$

Es importante observar dos cosas:

- R está bien definida, pues $f_j(x) \in [0, \infty]$ para cualquier j y por lo tanto $\min_{1 \leq j \leq n} f_j(x) = R(x) \in [0, \infty]$.
- Debido a que $x \in \mathbb{R}^n_*$, existe alguna ℓ para la cual $x_\ell > 0$, lo que implica que $f_\ell(x) < \infty$ y en consecuencia $R(x) \leq f_\ell(x) < \infty$.
- Como $R(x) = \min_{1 \leq j \leq n} f_j(x)$, existe una $k \in \{1, ..., n\}$ para la cual $R(x) = f_k(x)$, y como $R(x) < \infty$, entonces $f_k(x) < \infty$. En otras palabras,

$$\text{existe } k \in \{1, ..., n\} \text{ para la cual } R(x) = f_k(x) \text{ y } x_k \neq 0. \tag{1.2}$$

Una vez definidas estas funciones, el siguiente paso será probar la existencia del eigenvalor real r. Para ello, probaremos que la función R está acotada, luego probaremos que es una función continua y por lo tanto alcanza su máximo al restringir su dominio a un conjunto compacto. Posteriormente, probaremos que este máximo es en realidad el eigenvalor r.

Notemos que $R(x)x_j \leq v^x_j$ para cualquier j, pues

- Si $x_j = 0$ entonces $R(x)x_j = 0 \leq v^x_j$.
- Si $x_j > 0$ se tiene que $R(x) \leq f_j(x) = \frac{v^x_j}{x_j}$, por lo que al multiplicar por x_j se sigue que $R(x)x_j \leq v^x_j$.

Debido a esta desigualdad, se sigue que $R(x)x' \leq v_T(x) = x'T$.

Tomando $\overline{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$ y $K = \max\limits_{1 \leq i \leq n} \sum\limits_{j=1}^{n} t_{ij}$ podemos observar que

$$
\begin{aligned}
R(x) \cdot x' \cdot \overline{1} &= \begin{pmatrix} R(x)x_1 & R(x)x_2 & \dots & R(x)x_n \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \\
&= \left(\sum_{j=1}^{n} R(x)x_j \right) \\
&\leq \left(\sum_{j=1}^{n} v_j^x \right) \\
&= \left(\sum_{j=1}^{n} \left(\sum_{i=1}^{n} x_i t_{ij} \right) \right) \\
&= \left(\sum_{i=1}^{n} \left(\sum_{j=1}^{n} x_i t_{ij} \right) \right) \\
&= \left(\sum_{i=1}^{n} \left(x_i \sum_{j=1}^{n} t_{ij} \right) \right) \\
&\leq \left(\sum_{i=1}^{n} (x_i K) \right) \\
&= \left(K \sum_{i=1}^{n} x_i \right) \\
&= K \left(\sum_{i=1}^{n} x_i \right) \\
&= K \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \\
&= K \cdot x' \cdot \overline{1},
\end{aligned}
$$

esto es, $R(x) \cdot x' \cdot \overline{1} \leq K \cdot x' \cdot \overline{1}$.

De lo anterior se sigue que $R(x) \leq \frac{(K \cdot x' \cdot \overline{1})_{11}}{(x' \cdot \overline{1})_{11}} = K \frac{(x' \cdot \overline{1})_{11}}{(x' \cdot \overline{1})_{11}} = K$, por lo cual la función R está acotada por K y por lo tanto tiene un supremo en los reales. Sea $r = \sup\limits_{x \in \mathbb{R}_*^n} R(x)$.

Ahora buscaremos el máximo de R en un conjunto compacto y luego probaremos que también es máximo de la función R en todo $\mathbb{R}_*^n$. Debido a que f_j es continua para cada j y R es el mínimo de $\{f_1, f_2, ..., f_n\}$, se sigue que R también es continua. Restringiendo la función R al conjunto compacto $A = \{x \in \mathbb{R}_*^n : \|x\| = 1\}$, la función R sigue siendo continua y por lo tanto alcanza su máximo en el conjunto, es decir, existe $\hat{x} \in A$ tal que $R(x) \leq R(\hat{x})$ para cualquier $x \in A$. Lo siguiente será probar que $R(\hat{x}) = r$.

Sea $x \in \mathbb{R}^n_*$. Entonces

$$\begin{aligned} R(x) &= \min_{1\leq j\leq n} f_j(x) \\ &= \min_{\substack{1\leq j\leq n \\ x_j\neq 0}} \frac{v_j^x}{x_j} \\ &= \min_{\substack{1\leq j\leq n \\ x_j\neq 0}} \frac{\sum_{i=1}^{n} x_i t_{ij}}{x_j}. \end{aligned}$$

Como $x \neq 0$, $\frac{1}{\|x\|}$ está bien definido y se sigue que

$$\begin{aligned} R(x) &= \min_{\substack{1\leq j\leq n \\ x_j\neq 0}} \frac{\sum_{i=1}^{n} x_i t_{ij}}{x_j} \\ &= \min_{\substack{1\leq j\leq n \\ x_j\neq 0}} \frac{\frac{1}{\|x\|}\cdot\sum_{i=1}^{n} x_i t_{ij}}{\frac{1}{\|x\|}\cdot x_j} \\ &= \min_{\substack{1\leq j\leq n \\ x_j\neq 0}} \frac{\sum_{i=1}^{n} \frac{x_i}{\|x\|} t_{ij}}{\frac{x_j}{\|x\|}} \\ &= \min_{\substack{1\leq j\leq n \\ x_j\neq 0}} \frac{v_j^{\frac{x}{\|x\|}}}{\frac{x_j}{\|x\|}} \\ &= R\left(\frac{x}{\|x\|}\right) \leq R(\hat{x}) \quad \left(\text{ya que } \frac{x}{\|x\|} \in \{x \in \mathbb{R}^n_* : \|x\| = 1\} = A\right). \end{aligned}$$

Con esto se demuesta que para cualquier $x \in \mathbb{R}^n_*$ se tiene que $R(x) \leq R(\hat{x})$, por lo que $R(\hat{x})$ es una cota superior de la función R.

Observemos que

- Como r es el supremo de R, en particular es una cota superior de la función R y por lo tanto $R(\hat{x}) \leq r$.
- Como $R(\hat{x})$ es una cota superior de la función R y r es la mínima de las cotas superiores, entonces $r \leq R(\hat{x})$.

De esta manera, se tiene que $R(\hat{x}) = r$, es decir,

$$\text{existe una } \hat{x} \in \{x \in \mathbb{R}^n_* : \|x\| = 1\} \text{ tal que } R(\hat{x}) = r = \sup_{x\in\mathbb{R}^n_*} R(x). \tag{1.3}$$

A continuación nos enfocaremos en probar que $\hat{x}'T = r\hat{x}'$, lo que probaría que $\hat{x}$ es un eigenvector izquierdo de T y r un eigenvalor de T correspondiente al eigenvector $\hat{x}$. Definiendo $z' = \hat{x}'T - r\hat{x}'$, lo anterior es equivalente a probar que $z' = 0$.

En primer lugar, notemos que para cualquier $\ell = 1, ..., n$ se tiene lo siguiente:

- Si $\hat{x}_\ell = 0$, $r\hat{x}_\ell = 0 \leq \sum_{i=1}^{n} \hat{x}_i t_{i\ell} = v_\ell^{\hat{x}}$.

- Si $\hat{x}_\ell > 0$, $r = R(\hat{x}) = \min_{1\leq j\leq n} \frac{\sum_{i=1}^{n} \hat{x}_i t_{ij}}{\hat{x}_j} \leq \frac{\sum_{i=1}^{n} \hat{x}_i t_{i\ell}}{\hat{x}_\ell} = \frac{v_\ell^{\hat{x}}}{\hat{x}_\ell}$, lo que implica que $r\hat{x}_\ell \leq \sum_{i=1}^{n} \hat{x}_i t_{i\ell} = v_\ell^{\hat{x}}$.

En cualquier caso, $r\hat{x}_\ell \leq v_\ell^{\hat{x}}$, por lo que $r\hat{x}' \leq v_T(\hat{x}) = \hat{x}'T$. Debido a esto, $0 \leq \hat{x}'T - r\hat{x}' = z'$.

Supongamos que $0 < z'$, y sea $m \in \mathbb{N}$ tal que $0 < T^m$ (m existe porque T es regular). Entonces

$$\begin{aligned} 0 < z'T^m &= (\hat{x}'T - r\hat{x}')T^m \\ &= \hat{x}'T^{m+1} - r\hat{x}'T^m \\ &= (\hat{x}'T^m)T - r(\hat{x}'T^m). \end{aligned}$$

De lo anterior se sigue que $r(\hat{x}'T^m) < (\hat{x}'T^m)T$, y en particular $r((\hat{x}'T^m))_j < ((\hat{x}'T^m)T)_j$ para cualquier $j \in \{1, ..., n\}$. Entonces

$$r < \frac{((\hat{x}'T^m)T)_j}{(\hat{x}'T^m)_j} \text{ para cualquier } j = 1, ..., n \text{ donde ocurra que } (\hat{x}'T^m)_j \neq 0. \tag{1.4}$$

Notemos que, como $\hat{x} \in A \subseteq \mathbb{R}_*^n$ y T^m es una estrictamente positiva, existe $\alpha \in \{1, ..., n\}$ para el cual

$$(\hat{x}'T^m)_\alpha = \sum_{i=1}^{n} x_\alpha t_{i\alpha}^{(m)} > 0.$$

Esto demuestra que $\hat{x}'T^m \in \mathbb{R}_*^n$, por lo que la función R puede ser evaluada en $\hat{x}'T^m$. Utilizando (1.2), se tiene que existe $k \in \{1, ..., n\}$ para la cual $R(\hat{x}'T^m) = f_k(\hat{x}'T^m)$ y $(\hat{x}'T^m)_k \neq 0$, y por (1.4) se sigue que

$$r < \tfrac{((\hat{x}'T^m)T)_k}{(\hat{x}'T^m)_k} = f_k(\hat{x}'T^m) = R(\hat{x}'T^m),$$

lo cual es una contradicción ya que r era el supremo de la función R. Esta contradicción se sigue de suponer que $0 < z$, por lo que $0 = z$, es decir, $0 = \hat{x}'T - r\hat{x}'$ y por lo tanto $\hat{x}'T = r\hat{x}'$. Esto demuestra que r es un eigenvalor de T (y además r es real, pues es el supremo de la función R).

Finalmente, como T es una matriz regular cada columna de T contiene al menos una entrada positiva y entonces

$$0 < \sum_{i=1}^{n} t_{ij} \text{ para cada } j = 1,...,n,$$

por lo cual

$$\begin{aligned} 0 < \min_{1\leq j\leq n} \sum_{i=1}^{n} t_{ij} &= \min_{1\leq j\leq n} \frac{\sum_{i=1}^{n} 1 \cdot t_{ij}}{1} \\ &= \min_{1\leq j\leq n} \frac{v_j^{\bar{1}}}{1} \\ &= \min_{1\leq j\leq n} f_j(\bar{1}) \\ &= R(\bar{1}). \end{aligned}$$

Sabemos por (1.3) que $R(\hat{x}) = r = \sup_{x\in\mathbb{R}_*^n} R(x)$, y por lo anterior se sigue que $0 < R(\bar{1}) \leq R(\hat{x}) = r$. Esto demuestra que r es estrictamente positivo.

(b) En esta parte de la demostración buscaremos probar que el eigenvector izquierdo $\hat{x}$ encontrado en el inciso anterior es estrictamente positivo.

Comenzaremos recordando que $\hat{x} \in A = \{x \in \mathbb{R}_*^n : \|x\| = 1\}$, por lo que $0 \leq \hat{x}$ y $\|\hat{x}\| = 1$. Debido a esto, existe una $k \in \{1,...,n\}$ para la cual $0 < \hat{x}_k$. Además, como la matriz T es regular, existe una $m \in \mathbb{N}$ para la cual $T^m = (a_{ij})$ es estrictamente positiva, por lo que $0 < a_{ij}$ para cualesquiera $i, j = 1,...,n$. De esta manera,

$$0 < \hat{x}_k a_{kj} \leq \sum_{i=1}^{n} \hat{x}_i a_{ij} \text{ para cualquier } j = 1,...,n \tag{1.5}$$

Por otro lado, como r es un eigenvalor positivo de T, $\hat{x}'T^m = r^m\hat{x}'$ y por lo tanto $\hat{x}' = \frac{1}{r^m}\hat{x}'T^m$. Si $\hat{x} = (\hat{x}_1, \hat{x}_2, ..., \hat{x}_n)$, entonces

$$\begin{aligned} \hat{x}'T^m &= \begin{pmatrix} \hat{x}_1 a_{11} + \hat{x}_2 a_{21} + ... + \hat{x}_n a_{n1} & ... & \hat{x}_1 a_{1n} + \hat{x}_2 a_{2n} + ... + \hat{x}_n a_{nn} \end{pmatrix} \\ &= \begin{pmatrix} \sum_{i=1}^{n} \hat{x}_i a_{i1} & ... & \sum_{i=1}^{n} \hat{x}_i a_{in} \end{pmatrix}. \end{aligned}$$

Retomando (1.5), por la igualdad anterior se sigue que $0 < \hat{x}'T^m$. Finalmente, multiplicando por $\frac{1}{r^m}$ se tiene que $0 < \frac{1}{r^m}\hat{x}'T^m = \hat{x}'$.
Esto prueba que $\hat{x}$ es un eigenvector izquierdo estrictamente positivo.

(c) Sea α un eigenvalor de T distinto de r. Entonces existe un eigenvector izquierdo $y \in \mathbb{R}^n$, $y \neq 0$, tal que $\alpha y' = y'T$. Por lo tanto, para cualquier $j = 1, 2, ..., n$ se tiene que

$$\alpha y_j = \sum_{i=1}^{n} y_i t_{ij} \tag{1.6}$$

y

$$\alpha^m y_j = \sum_{i=1}^{n} y_i a_{ij}. \tag{1.7}$$

A continuación demostraremos que $|\alpha| \leq r$, y posteriormente que $|\alpha| < r$.

Tomando la norma en (1.6),

$$|\alpha y_j| = |\sum_{i=1}^{n} y_i t_{ij}| \leq \sum_{i=1}^{n} |y_i| t_{ij} \text{ para cualquier } j = 1, ..., n. \tag{1.8}$$

Definimos $y_+ = (|y_1|, |y_2|, ..., |y_n|)$. Como $y \neq 0$, entonces $y_+ \in \mathbb{R}^n_*$ y por lo tanto las funciones definidas en el inciso (a) se pueden aplicar a dicho vector. Por la desigualdad anterior

$$|\alpha| \leq \min_{\substack{1 \leq j \leq n \\ y_j \neq 0}} \frac{\sum_{i=1}^{n} |y_i| t_{ij}}{|y_j|} \leq \min_{1 \leq j \leq n} f_j(y_+) = R(y_+) \leq r.$$

Ahora demostraremos que $|\alpha| < r$ por reducción al absurdo.

Supongamos que $|\alpha| = r$. Por (1.8) se sigue que $r|y_j| \leq \sum_{i=1}^{n} |y_i| t_{ij}$, y mediante un proceso análogo al desarrollado en el inciso (a) se sigue que $r|y_j| = \sum_{i=1}^{n} |y_i| t_{ij}$, lo cual implica que y_+ es un eigenvector izquierdo de T.

Por lo tanto, $r^m |y_j| = \sum_{i=1}^{n} |y_i| a_{ij}$ y al tomar la norma en (1.7):

$$\begin{aligned}
|\sum_{i=1}^{n} y_i a_{ij}| &= |\alpha^m y_j| \\
&= |\alpha|^m |y_i| \\
&= r^m |y_j| \\
&= \sum_{i=1}^{n} |y_i a_{ij}| \\
&= \sum_{i=1}^{n} |y_i| a_{ij}.
\end{aligned}$$

Recordemos que, para cualquier par de números complejos, la norma de su suma es igual a la suma de sus normas si y solo si ambos se encuentran en la misma dirección[2], es decir, dados $z_1 = |z_1| e^{i\theta_1}$ y $z_2 = |z_2| e^{i\theta_2}$ números complejos, $|z_1 + z_2| = |z_1| + |z_2|$ si y solo si $\theta_1 = \theta_2 + 2k\pi$ para alguna $k \in \mathbb{N} \cup \{0\}$, y para más de dos sumandos el mismo resultado se sigue por inducción.

Si expresamos a y_j como $|y_j| e^{i\theta_j}$ para cada j, θ_j en realidad es la misma para cualquier j (salvo múltiplos de 2π), y entonces

$$y_j = |y_j| e^{i\theta}.$$

[2]Resultado contenido en [10, p. 17]

Sustituyendo este valor en (1.6),

$$\alpha|y_j|e^{i\theta} = \sum_{\ell=1}^{n} |y_\ell|e^{i\theta}t_{\ell j}.$$

Como $e^{i\theta}$ es una constante distinta de cero, se sigue que

$$\alpha|y_j| = \sum_{\ell=1}^{n} |y_\ell|t_{\ell j}.$$

Notemos que la parte derecha de esta última igualdad es real y mayor o igual a cero, por lo que sucede lo mismo con la parte izquierda. En particular, $0 \leq \alpha$.

Con esto se demuestra que $\alpha = |\alpha| = r$, lo cual es una contradicción ya que α y r eran distintos. Esto viene de suponer que $|\alpha| = r$, por lo cual $|\alpha| < r$.

(d) Sea $x \in \mathbb{C}^n \backslash \{0\}$ un eigenvector izquierdo de T asociado a r.

Como $x \neq 0$, existe una k para la cual $x_k \neq 0$. Definimos $c = \frac{\hat{x}_k}{x_k}$ (que es distinto de cero, pues en el inciso (b) se demostró que $\hat{x}$ es estrictamente positivo).

Notemos que para esa k particular se tiene que

$$\begin{aligned}(\hat{x} - cx)_k &= \hat{x}_k - cx_k \\ &= \hat{x}_k - \frac{\hat{x}_k}{x_k}x_k \\ &= 0.\end{aligned}$$

Supongamos que x no es múltiplo de $\hat{x}$, entonces $\hat{x} \neq cx$. Denotando como $z = \hat{x} - cx \neq 0$,

$$\begin{aligned}z'T = (\hat{x}' - cx')T &= \hat{x}'T - cx'T \\ &= r\hat{x}' - crx' \\ &= r(\hat{x}' - cx') \\ &= rz.\end{aligned}$$

Esto prueba que z también es un eigenvector izquierdo de T asociado al eigenvalor r.

Sea $z_+ = (|z_1|, |z_2|, ..., |z_n|)$. Con un procedimiento análogo al del inciso (c) se tiene que

$$\sum_{i=1}^{n} |z_i|t_{ij} = r|z_j|$$

Por lo tanto, z_+ es también un eigenvector izquierdo de T, y por el inciso (b) se sigue que $0 < z_+$. En particular, $0 < |(\hat{x} - cx)_k| = (z_+)_k$. Esto es una contradicción, pues $(\hat{x} - cx)_k = 0$, contradicción que viene de suponer que x no es múltiplo de $\hat{x}$. Por lo tanto, x es múltiplo de $\hat{x}$.

Eigenvectores derechos: La demostración para eigenvectores derechos en los incisos (a) al (d) es análoga a la realizada para eigenvectores izquierdos.

(e) Sea v un eigenvector derecho de B con eigenvalor asociado β, donde $B = (b_{ij})$.

El primer paso será demostrar que $|\beta| \leq r$ utilizando un procedimiento similar al del inciso (c), y después probaremos que si ambos son iguales entonces las matrices B y T son iguales.

Como v es un eigenvector derecho de B entonces $\beta v = Bv$, y por lo tanto para cualquier $j = 1, 2, ..., n$ se tiene que

$$\beta v_j = \sum_{i=1}^{n} b_{ji} v_i.$$

Tomando la norma, se sigue que

$$|\beta||v_j| = |\sum_{i=1}^{n} b_{ji} v_i| \leq \sum_{i=1}^{n} b_{ji} |v_i| \text{ para cualquier } j = 1, 2, ..., n.$$

Definiendo $v_+ = (|v_1|, |v_2|, ..., |v_n|) \in \mathbb{R}^n_*$, por esta última desigualdad y que $B \leq T$ se tiene que

$$|\beta| v_+ \leq B v_+ \leq T v_+. \tag{1.9}$$

Multiplicando esta desigualdad por $\hat{x}'$, donde $\hat{x}$ es el eigenvector izquierdo de T encontrado en el inciso (a),

$$|\beta| \hat{x}' v_+ \leq \hat{x}' T v_+ = r \hat{x}' v_+. \tag{1.10}$$

Notemos que, como $v \neq 0$, existe una $k \in \{1, 2, ..., n\}$ para la cual $v_k \neq 0$. También recordemos que en el inciso (b) se demostró que $0 < \hat{x}$, por lo que

$$0 < \hat{x}_k |v_k| \leq \sum_{i=1}^{n} \hat{x}_i |v_i| = \hat{x}' v_+.$$

Por esta razón, de (1.10) se sigue que $|\beta| \leq r$.

Ahora supongamos que $|\beta| = r$. De (1.9) se tiene que

$$r v_+ = |\beta| v_+ \leq T v_+.$$

Utilizando un procedimiento análogo al inciso (b) se sigue que $0 < r v_+ = T v_+$, y por la desigualdad anterior

$$r v_+ = B v_+ = T v_+. \tag{1.11}$$

Si $B \neq T$, como $B \leq T$ existen $i, j \in \{1, ..., n\}$ tales que $b_{ij} < t_{ij}$. Por otro lado, como v_+ es un eigenvector derecho de T con eigenvalor asociado r y $0 \leq v_+$, por los incisos (b) y (d) de este teorema se sigue que $0 < v_+$. Debido a esto,

$$\sum_{\ell=1}^{n} b_{i\ell} |v_\ell| < \sum_{\ell=1}^{n} t_{i\ell} |v_\ell|,$$

lo cual contradice (1.11). Por lo tanto, $B = T$.

(f) Para cualquier matriz $A \in M_{n\times n}(\mathbb{R})$ se tiene que

$$det(A)\cdot I = A\cdot adj(A) = adj(A)\cdot A, \tag{1.12}$$

donde I es la matriz identidad y $adj(A)$ es la adjunta clásica[3] de la matriz A.[4]

Haciendo $A = (xI - T)$, por (1.12) se tiene que

$$det(xI-T)I = (xI-T)\cdot adj(xI-T) \tag{1.13}$$

y

$$det(xI-T)I = adj(xI-T)\cdot(xI-T). \tag{1.14}$$

La demostración de este inciso se dividirá en tres partes: primero probaremos que todos los elementos de la matriz $adj(xI-T)$ son positivos, luego utilizaremos esto para demostrar que $\frac{d}{dx}det(xI-T)|_{x=r} > 0$ y finalmente veremos por qué esto implica que la multiplicidad del eigenvalor r es 1.

Haciendo $x = r$ en (1.13) se tiene que

$$det(rI-T)I = (rI-T)\cdot adj(rI-T).$$

Como r es un eigenvalor de T, $det(rI-T) = 0$. Sustituyendo este valor e la igualdad anterior,

$$\begin{aligned} 0 &= (rI-T)\cdot adj(rI-T) \\ &= r\cdot adj(rI-T) - T\cdot adj(rI-T), \end{aligned}$$

por lo cual

$$T\cdot adj(rI-T) = r\cdot adj(rI-T). \tag{1.15}$$

De esta igualdad se sigue que cada columna de la matriz $adj(rI-T)$ es cero o un eigenvector derecho de T con eigenvalor asociado r; en este último caso, por los incisos (b) y (d) de este teorema, todas las entradas de esa columna son estrictamente positivas.

Repitiendo el procedimiento anterior para (1.14) se sigue que cada renglón de $adj(rI-T)$ es cero o un eigenvector izquierdo estrictamente positivo de T con eigenvalor asociado r.

De estos dos últimos párrafos se desprende que la matriz $adj(rI-T)$ es igual a cero o es estrictamente positiva.

[3] La adjunta clásica de una matriz $A \in M_{n\times n}(\mathbb{R})$ se define como la matriz $adj(A) \in M_{n\times n}(\mathbb{R})$ en la cual la ij-ésima entrada es $(-1)^{i+j}det(\widetilde{A}_{ji})$, donde $\widetilde{A}_{ji} \in M_{(n-1)\times(n-1)}(\mathbb{R})$ es la matriz A' sin el renglón j ni la columna i. Esta definición puede ser consultada con más detalle en [6].

[4] Resultado contenido en [6, p. 219].

Supongamos que la matriz $adj(rI-T)$ es cero. Denotando como $I_{(n-1)}$ la matriz identidad sin la fila y columna n, y de igual manera a $T_{(n-1)}$ a la matriz T sin la fila y columna n, como la matriz $adj(rI-T)$ es cero se tiene que

$$\begin{aligned} 0 &= a_{nn} \\ &= (-1)^{n+n}\, det(rI_{(n-1)} - T_{(n-1)}) \\ &= det(rI_{(n-1)} - T_{(n-1)}), \end{aligned}$$

por lo que r es también un eigenvalor de la matriz $T_{(n-1)}$.

Sea $B = \begin{pmatrix} T_{(n-1)} & \mathbf{0} \\ \mathbf{0} & 0 \end{pmatrix}$. Notemos que $0 \leq B \leq T$, ya que T es una matriz con entradas no negativas, y además

$$\begin{aligned} det(rI-B) &= det \begin{pmatrix} rI_{(n-1)} - T_{(n-1)} & 0 \\ 0 & r \end{pmatrix} \\ &= r \cdot det(rI_{(n-1)} - T_{(n-1)}) \\ &= r \cdot 0 \\ &= 0, \end{aligned}$$

por lo cual r es un eigenvalor de B.

Por el inciso (e) de este teorema se sigue que $B = T$, es decir, $T = \begin{pmatrix} T_{(n-1)} & \mathbf{0} \\ \mathbf{0} & 0 \end{pmatrix}$. Como la última fila y el último renglón de la matriz T son cero, entonces $t_{nn}^{(m)} = 0$ para toda $m \in \mathbb{N}$, lo cual es una contradicción pues T es una matriz regular.

Por lo tanto, la matriz $adj(rI-T)$ es estrictamente positiva.

Definimos la función $f : \mathbb{C} \to \mathbb{C}$ como $f(x) = det(xI-T)$. Por (1.13) se tiene que

$$f(x)I = (xI-T) \cdot adj(xI-T).$$

Derivando esta expresión con respecto de x,

$$\begin{aligned} \frac{d}{dx} f(x)I &= \frac{d}{dx}[(xI-T) \cdot adj(xI-T)] \\ &= \left[\frac{d}{dx}(xI-T)\right] adj(xI-T) + (xI-T)\left[\frac{d}{dx} adj(xI-T)\right] \\ &= I \cdot adj(xI-T) + (xI-T)\left[\frac{d}{dx} adj(xI-T)\right]. \end{aligned}$$

Denotando $\frac{d}{dx} f(x) = f'(x)$, por lo anterior se tiene que

$$f'(x)I = adj(xI-T) + (xI-T)\left[\frac{d}{dx} adj(xI-T)\right]. \tag{1.16}$$

Haciendo $x = r$, de (1.16) se sigue que

$$f'(r)I = adj(rI - T) + (rI - T)\left[\frac{d}{dx}adj(rI - T)\right]. \tag{1.17}$$

Multiplicando por $\hat{x}'$, donde $\hat{x}$ es el eigenvector izquierdo de T asociado a r que encontramos en la demostración del inciso (a),

$$\begin{aligned}
f'(r)\hat{x}' &= \hat{x}'adj(rI - T) + \hat{x}'(rI - T)\left[\frac{d}{dx}adj(rI - T)\right] \\
&= \hat{x}'adj(rI - T) + (r\hat{x}'I - \hat{x}'T)\left[\frac{d}{dx}adj(rI - T)\right] \\
&= \hat{x}'adj(rI - T) + (r\hat{x}' - r\hat{x}')\left[\frac{d}{dx}adj(rI - T)\right] \\
&= \hat{x}'adj(rI - T) + (0)\left[\frac{d}{dx}adj(rI - T)\right] \\
&= \hat{x}'adj(rI - T).
\end{aligned}$$

Como la matriz $adj(rI - T)$ es estrictamente positiva y $\hat{x} > 0$, se sigue que $\hat{x}'adj(rI - T) > 0$ y por lo tanto $f'(r) > 0$.

Finalmente, notemos que

$$\begin{aligned}
f(x) &= det(xI - T) \\
&= (-1)^n det(T - xI).
\end{aligned}$$

Como $det(T - xI)$ es el polinomio característico de T, entonces

$$f(x) = (-1)^n(x - \theta_1)(x - \theta_2)(x - \theta_3)...(x - \theta_n), \tag{1.18}$$

donde $\theta_1, \theta_2, ..., \theta_n$ son los eigenvalores de T (uno de ellos es r).

Supongamos que la multiplicidad de r en el polinomio característico es mayor a 1, por lo cual existen al menos $i, j \in \{1, ..., n\}$ con $i \neq j$ tal que $\theta_i = \theta_j = r$. Sin pérdida de generalidad, supongamos que $\theta_1 = \theta_2 = r$.

De (1.18) se sigue que

$$\begin{aligned}
f(x) &= (-1)^n(x - r)^2(x - \theta_3)...(x - \theta_n) \\
&= (-1)^n(x - r)^2\prod_{i=3}^{n}(x - \theta_i).
\end{aligned}$$

Derivando con respecto de x,

$$f'(x) = (-1)^n \cdot 2(x - r) \cdot \prod_{i=3}^{n}(x - \theta_i) + \sum_{i=3}^{n}\left((-1)^n(x - r)^2\prod_{\substack{j=3\\ j\neq i}}^{n}(x - \theta_j)\right).$$

En particular, para $x = r$ se tiene que

$$\begin{aligned} f'(r) &= (-1)^n \cdot 2(r-r) \cdot \prod_{i=3}^{n}(r-\theta_i) + \sum_{i=3}^{n}\left((-1)^n (r-r)^2 \prod_{\substack{j=3\\ j\neq i}}^{n}(r-\theta_j)\right) \\ &= 0 + \sum_{i=3}^{n}(0) \\ &= 0, \end{aligned}$$

lo cual es una contradicción, pues ya hemos probado que $f'(r) > 0$. Por lo tanto, la multiplicidad de r en el polinimio característico de T es 1. $\square$

Al eigenvalor r obtenido en el teorema anterior se le llama eigenvalor de Perron-Frobenius.

Para terminar este capítulo, demostraremos un lema contenido en [5, p. 72], pues utilizaremos este resultado más adelante.

Lema 2. *Sean $T \in M_{n\times n}(\mathbb{R})$ una matriz regular, r el eigenvalor de Perron-Frobenius de la matriz T y $\phi \in \mathbb{R}^n$.*
Si $\phi > 0$, entonces para cada $i = 1, 2, \ldots, n$ se cumple que

$$\lim_{m\to\infty} \frac{1}{m} log\left[\sum_{j=1}^{n} a_{ij}^{(m)}\phi_j\right] = \lim_{m\to\infty} \frac{1}{m} log\left[\sum_{j=1}^{n} \phi_j a_{ji}^{(m)}\right] = log(r)$$

donde $a_{ij}^{(m)}$ son las entradas de la matriz T^m.

Demostración.
Sean $T \in M_{n\times n}(\mathbb{R})$ una matriz regular, r el eigenvalor de Perron-Frobenius de la matriz T, $\hat{x}$ el eigenvector derecho asociado a r, $\phi \in \mathbb{R}^n$ tal que $\phi > 0$ e $i \in \{1, 2, \ldots, n\}$.

Primero nos concentraremos en probar que

$$\lim_{m\to\infty} \frac{1}{m} log\left[\sum_{j=1}^{n} a_{ij}^{(m)}\phi_j\right] = log(r),$$

para lo cual buscaremos sucesiones $\{x_m\}$ y $\{z_m\}$ que acoten inferior y superiormente, respectivamente, a la sucesión $\{y_m\} = \left\{\frac{1}{m} log\left[\sum_{j=1}^{n} a_{ij}^{(m)}\phi_j\right]\right\}$, y que ambas converjan a $\log(r)$ cuando m tiende a infinito.

Sea $\hat{x}$ el eigenvector derecho asociado a r. Comenzaremos definiendo cuatro constantes, dos en función de $\hat{x}$ y dos en función de ϕ:

$$\beta = \min_{1\leq \ell\leq n} \hat{x}_\ell \;\text{ y }\; \alpha = \max_{1\leq \ell\leq n} \hat{x}_\ell$$

$$\delta = \min_{1\leq \ell \leq n} \phi_\ell \ \text{ y } \ \gamma = \max_{1\leq \ell \leq n} \phi_\ell.$$

Sea $j \in \{1, 2, ..., n\}$. Por como están definidas las constantes anteriores, se sigue que

$$\beta \leq \hat{x}_j \leq \alpha \tag{1.19}$$

y

$$\delta \leq \phi_j \leq \gamma\,. \tag{1.20}$$

Ahora buscaremos definir $\{x_m\}$. Primero notemos que $\hat{x}_j \leq \alpha$ por (1.19) y $\delta \leq \phi_j$ por (1.20), por lo cual

$$\hat{x}_j \cdot \delta \leq \alpha \cdot \phi_j\,.$$

Sea $m \in \mathbb{N}$. Ya que T es regular, todas sus entradas son mayores o iguales a cero, y por lo tanto $a_{ij}^{(m)} \geq 0$. Multiplicamos la desigualdad anterior por $a_{ij}^{(m)}$

$$\delta \cdot a_{ij}^{(m)} \cdot \hat{x}_j \leq \alpha \cdot a_{ij}^{(m)} \cdot \phi_j\,.$$

Como $\hat{x}$ es el eigenvalor derecho asociado a r, $\hat{x}$ es estrictamente positivo y por lo tanto $\max\limits_{1\leq i\leq n} \hat{x}_i = \alpha$ también es positivo. Ahora multiplicamos por $\frac{1}{\alpha}$

$$\tfrac{\delta}{\alpha} \cdot a_{ij}^{(m)} \cdot \hat{x}_j \leq a_{ij}^{(m)} \cdot \phi_j\,.$$

Ya que esto se cumple para cualquier $j = 1, ..., n$, sumando todos los términos se tiene que

$$\frac{\delta}{\alpha} \sum_{j=1}^{n} a_{ij}^{(m)} \cdot \hat{x}_j \leq \sum_{j=1}^{n} a_{ij}^{(m)} \cdot \phi_j. \tag{1.21}$$

Como T es regular, T^m es regular y por lo tanto no puede tener ningún renglón con todas sus entradas iguales a cero. En particular, existe una $a_{ik}^{(m)}$ estrictamente positiva. Además, como ϕ y $\hat{x}$ son estrictamente positivos, δ y $\hat{x}_k$ también son positivos. Debido a esto,

$$0 < \tfrac{\delta}{\alpha} a_{ik}^{(m)} \cdot \hat{x}_k \leq \tfrac{\delta}{\alpha} \sum_{j=1}^{n} a_{ij}^{(m)} \cdot \hat{x}_j \leq \sum_{j=1}^{n} a_{ij}^{(m)} \cdot \phi_j\,.$$

Acabamos de probar que ambos lados de (1.21) son estrictamente positivos, por lo cual podemos aplicar logaritmo en ambos lados de la desigualdad. Aplicando el logaritmo y multiplicando por $\frac{1}{m}$ se tiene que

$$\tfrac{1}{m} \log\left(\tfrac{\delta}{\alpha} \sum_{j=1}^{n} a_{ij}^{(m)} \cdot \hat{x}_j \right) \leq \tfrac{1}{m} \log\left(\sum_{j=1}^{n} a_{ij}^{(m)} \cdot \phi_j \right).$$

Notemos que el sentido de la desigualdad se conserva porque el logaritmo es una función creciente y $\frac{1}{m} > 0$.

Esto prueba que la sucesión $\left\{\frac{1}{m} log\left[\sum_{j=1}^{n} a_{ij}^{(m)} \phi_j\right]\right\} = \{y_m\}$ está acotada inferiormente por la sucesión $\left\{\frac{1}{m}\log\left(\frac{\delta}{\alpha}\sum_{j=1}^{n} a_{ij}^{(m)}\cdot \hat{x}_j\right)\right\}$.

Definimos $\{x_m\} = \left\{\frac{1}{m}\log\left(\frac{\delta}{\alpha}\sum_{j=1}^{n} a_{ij}^{(m)}\cdot \hat{x}_j\right)\right\}$.

Ahora seguiremos un procedimiento muy similar para definir $\{z_m\}$.

De (1.19) se tiene que $\beta \leq \hat{x}_j$ y de (1.20) que $\phi_j \leq \gamma$. Multiplicando estas desigualdades se sigue que

$$\phi_j \cdot \beta \leq \gamma \cdot \hat{x}_j \ .$$

Luego multiplicamos la desigualdad anterior por $a_{ij}^{(m)} \geq 0$

$$\phi_j \cdot a_{ij}^{(m)} \cdot \beta \leq \gamma \cdot a_{ij}^{(m)} \cdot \hat{x}_j \ .$$

Ya que $\hat{x}$ es estrictamente positivo, $\min\limits_{1\leq i\leq n} \hat{x}_i = \beta$ es positivo. Multiplicando la desigualdad anterior por $\frac{1}{\beta}$ se sigue que

$$a_{ij}^{(m)} \cdot \phi_j \leq \frac{\gamma}{\beta} \cdot a_{ij}^{(m)} \cdot \hat{x}_j.$$

Ya que esto se cumple para cualquier $j = 1, ..., n$,

$$\sum_{j=1}^{n} a_{ij}^{(m)} \phi_j \leq \frac{\gamma}{\beta} \sum_{j=1}^{n} a_{ij}^{(m)} \hat{x}_j. \tag{1.22}$$

Recordemos que $a_{ik}^{(m)} > 0$ y que ϕ es positivo, por lo que

$$0 < a_{ik}^{(m)} \phi_k \leq \sum_{j=1}^{n} a_{ij}^{(m)} \phi_j \leq \frac{\gamma}{\beta} \sum_{j=1}^{n} a_{ij}^{(m)} \hat{x}_j.$$

Esto prueba que ambos lados de (1.22) son positivos. Aplicando la función logaritmo y multiplicando por $\frac{1}{m}$ se tiene que

$$\frac{1}{m}\log\left(\sum_{j=1}^{n} a_{ij}^{(m)} \phi_j\right) \leq \frac{1}{m}\log\left(\frac{\gamma}{\beta}\sum_{j=1}^{n} a_{ij}^{(m)} \hat{x}_j\right).$$

Con esto, la sucesión $\left\{\frac{1}{m}\log\left(\frac{\gamma}{\beta}\sum_{j=1}^{n} a_{ij}^{(m)} \hat{x}_j\right)\right\}$ acota superiormente a $\{y_m\}$. Definimos $\{z_m\} = \left\{\frac{1}{m}\log\left(\frac{\gamma}{\beta}\sum_{j=1}^{n} a_{ij}^{(m)} \hat{x}_j\right)\right\}$.

Queda pendiente demostrar que $\{x_m\}$ y $\{z_m\}$ convergen a $\log(r)$. Para esto, consideremos una sucesión de la forma $\left\{\frac{1}{m}\log\left(c\cdot\sum_{j=1}^{n} a_{ij}^{(m)} \hat{x}_j\right)\right\}$, con c una constante positiva

($\{x_m\}$ y $\{z_m\}$ son de esta forma).

Como $\hat{x}$ es un eigenvalor derecho de T, para cada m se tiene que $T^m\hat{x} = r^m\hat{x}$, y por lo tanto:

$$\begin{aligned}\frac{1}{m}\log(c \cdot a_{ij}^{(m)}\hat{x}_j) &= \frac{1}{m}\log(c \cdot r^m\hat{x}_j)\\ &= \frac{1}{m}\log(c) + \frac{1}{m}\cdot m\log(r) + \frac{1}{m}\log(\hat{x}_j)\\ &= \frac{1}{m}\log(c) + \log(r) + \frac{1}{m}\log(\hat{x}_j)\end{aligned}$$

Ya que el límite cuando m tiende a infinito existe para cada uno de los sumandos en la parte derecha de la igualdad, se sigue que el límite de $\frac{1}{m}\log(c \cdot a_{ij}^{(m)}\hat{x}_j)$ también existe y además

$$\begin{aligned}\lim_{m\to\infty}\frac{1}{m}\log(c \cdot a_{ij}^{(m)}\hat{x}_j) &= \lim_{m\to\infty}\left(\frac{1}{m}\log(c) + \log(r) + \frac{1}{m}\log(\hat{x}_j)\right)\\ &= \lim_{m\to\infty}\frac{1}{m}\log(c) + \lim_{m\to\infty}\log(r) + \lim_{m\to\infty}\frac{1}{m}\log(\hat{x}_j)\\ &= 0 + \log(r) + 0\\ &= \log(r)\end{aligned}$$

Sustituyendo c por $\frac{\gamma}{\beta}$, esto prueba que $\{x_m\}$ converge a $\log(r)$. De igual forma, sustituyendo c por $\frac{\delta}{\alpha}$ se tiene que $\{z_m\}$ converge también a $\log(r)$.

Como $x_m \leq y_m \leq z_m$, y además $\{x_m\}$ y $\{z_m\}$ convergen a $\log(r)$, se sigue que

$$\lim_{m\to\infty}\tfrac{1}{m}log\left[\sum_{j=1}^{n} a_{ij}^{(m)}\phi_j\right] = log(r).$$

El procedimiento para el segundo límite es análogo. $\square$

Capítulo 2

El Teorema de Perron-Frobenius aplicado a matrices de transición

En este capítulo estudiaremos qué tipo de cadenas de Markov cuentan con una matriz de transición a la que podemos aplicar el Teorema de Perron-Frobenius, es decir, cuya matriz de transición sea regular, y en estos casos analizaremos la velocidad de convergencia de las probabilidades de transición a la distribución estacionaria.

Comenzaremos definiendo en la primera sección el concepto de cadena de Markov y matriz de transición, además de otras definiciones y resultados importantes en materia de Procesos Estocásticos. Este material puede ser consultado en [7].

Definiciones básicas

Definición 4 (Cadenas de Markov)**.** *Sean (Ω, F, P) un espacio de probabilidad, E un conjunto finito y $\{X_n\}$ una sucesión de variables aleatorias con valores en E (esto es, $X_n : \Omega \to E$ para cada $n \in \mathbb{N}$). Decimos que la sucesión $\{X_n\}$ es una **cadena de Markov** con espacio de estados E si*

$$P(X_{n+1} = x_{n+1} \mid X_0 = x_0, \ldots, X_n = x_n) = P(X_{n+1} = x_{n+1} \mid X_n = x_n)$$

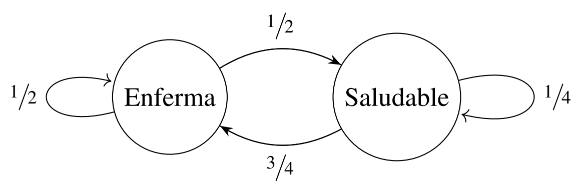
para cualesquiera $n \in \mathbb{N}$ y $x_0, x_1, \ldots, x_n \in E$. A esta propiedad se le llama propiedad de Markov.
*Decimos que una cadena de Markov $\{X_n\}$ es **homogénea** si $P(X_{n+1} = j \mid X_n = i)$ es independiente de n, es decir,*

$$P(X_{n+1} = j \mid X_n = i) = P(X_1 = j \mid X_0 = i) \text{ para cualesquiera } n \in \mathbb{N} \text{ e } i, j \in E.$$

Se puede definir el concepto de cadena de Markov en espacios de estados más generales, pero en esta tesis trabajaremos siempre con un espacio de estados finito y con cadenas de Markov homogéneas.

A continuación presentamos un ejemplo que nos permitirá ilustrar esta definición, y que retomaremos el resto de esta sección para las definiciones que posteriormente introduciremos.

Ejemplo 3. Sea X_n la variable aleatoria que describe si una persona está saludable o no en el día n: $X_n = 0$ si la persona está enferma o $X_n = 1$ si la persona está saludable. Supongamos que la probabilidad de que la persona pase de un estado a otro está dada de la siguiente manera:



Esta es una cadena de Markov homogénea con dos estados.

Definición 5. *Sea $\{X_n\}$ una cadena de Markov homogénea con espacio de estados E. Definimos la* ***distribución inicial*** *de $\{X_n\}$ como*

$$\pi_0(x) = P(X_0 = x).$$

También se define la ***matriz de transición*** *de la cadena como $P = (p_{ij})_{i,j \in E}$, donde*

$$p_{ij} = P(X_1 = j \mid X_0 = i).$$

Para simplifcar la notación escribiremos $P = (p_{ij})$, siempre bajo el entendido de que $i, j \in E$.

Retomando el ejemplo 3, la matriz de transición de la cadena es

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

Dada $\{X_n\}$ una cadena de Markov homogénea con espacio de estados E, si denotamos como $p_{ij}^{(m)}$ a la probabilidad de llegar del estado i al estado j en m pasos, es decir, $p_{ij}^{(m)} = P(X_m = j \mid X_0 = i)$, se puede demostrar que $P^m = (p_{ij}^{(m)})$. En otras palabras, la matriz P elevada a su m-ésima potencia contiene las probabilidades de transición en m pasos. La demostración de este resultado puede ser consultada en [7, p. 13].

Una vez introducidos los conceptos de cadena de Markov homogénea y matriz de transición, lo siguiente será introducir dos conceptos que nos permitirán establecer condiciones suficientes para asegurar que una cadena de Markov cuenta con una matriz de transición regular (requisito necesario para aplicar el teorema de Perron-Frobenius):

irreducibilidad y periodo de la cadena.

Abordemos primero el concepto de irreducibilidad de la cadena, para lo cual serán necesarias las siguientes definiciones:

Definición 6. *Sea $\{X_n\}$ una cadena de Markov homogénea con espacio de estados E, y sean $i, j \in E$. Decimos que* ***el estado*** *i* ***se comunica con el estado*** *j si existe $m \in \mathbb{N}$ para la cual $p_{ij}^{(m)} > 0$.*
Si el estado i se comunica con el estado j y el estado j se comunica con el estado i, diremos que ***los estados*** *i* ***y*** *j* ***se comunican entre sí****.*

Podemos notar en el ejemplo 3 que los estados 0 y 1 se comunican entre sí, pues $p_{01} = 1/2$ y $p_{10} = 1/4$.

Definición 7. *Sean $\{X_n\}$ una cadena de Markov homogénea con espacio de estados E y C un subconjunto de E no vacío. Decimos que C es un* ***subconjunto de estados cerrado*** *si ningún estado dentro de C se comunica con algún estado fuera de C, esto es,*

$$p_{ij}^{(m)} = 0 \text{ para cualesquiera } i \in C,\ j \notin C \text{ y } m \geq 1.$$

Con base en esta definición, se define la irreducibilidad de un subconjunto de estados y luego de toda la cadena $\{X_n\}$.

Definición 8. *Sean $\{X_n\}$ una cadena de Markov homogénea con espacio de estados E y C un subconjunto de estados cerrado. Decimos que C es* ***irreducible*** *si i se comunica con j para cualesquiera $i, j \in C$.*

Definición 9. *Sean $\{X_n\}$ una cadena de Markov homogénea con espacio de estados E. Decimos que la cadena $\{X_n\}$ es una* ***cadena de Markov irreducible*** *si el espacio de estados E es irreducible.*

Observemos que el conjunto de estados $\{0, 1\}$ del ejemplo 3 es irreducible, y más aún, por esta misma razón la cadena es irreducible.

El segundo concepto importante requiere, de igual manera, algunas definiciones previas.

Definición 10. *Sea $\{X_n\}$ una cadena de Markov homogénea irreducible con espacio de estados E, y sea $x \in E$ tal que $p_{xx}^{(m)} > 0$ para alguna $m \in \mathbb{N}$.*
Definimos el ***periodo*** *de x, d_x, como el máximo común divisor del conjunto de potencias para las cuales la probabilidad de regresar al estado x es positiva, es decir,*

$$d_x = MCD(\{m \in \mathbb{N} \mid p_{xx}^{(m)} > 0\}).$$

En [7, p. 72-73] se demuestra que si i y j son dos estados que se comunican entre sí, entonces $d_i = d_j$. Esto en particular demuestra que si la cadena $\{X_n\}$ es irreducible entonces todos sus estados tienen el mismo periodo d (como la cadena es irreducible todos sus estados se comunican entre sí). Este resultado se utiliza para definir el periodo de la cadena.

Definición 11. *Sea $\{X_n\}$ una cadena de Markov homogénea irreducible con espacio de estados E. Definimos el* ***periodo de la cadena*** *d como el periodo de cualquiera de sus estados.*

Decimos que la cadena $\{X_n\}$ es ***aperiódica*** *si $d = 1$ y* ***periódica*** *si $d > 1$, en cuyo caso se dice que d es el periodo de la cadena.*

Notemos que la cadena de Markov definida en el ejemplo 3 es aperiódica, pues

$$d_0 = MCD(\{m \in \mathbb{N} \mid p_{00}^{(m)} > 0\}) = 1,$$

ya que $p_{00} = 1/2 > 0$, y como la cadena es irreducible entonces $d_0 = 1$ es justamente el periodo de la cadena.

Para concluir esta sección introduciremos el concepto de distribución estacionaria y enunciaremos condiciones suficientes para que una cadena de Markov cuente con una distribución estacionaria, resultado que utilizaremos más adelante.

Definición 12. *Sea $\{X_n\}$ una cadena de Markov homogénea con espacio de estados E y P su matriz de transición.*
Si π_i, $i \in E$, son números no negativos tales que $\sum_{i\in E} \pi_i = 1$ y

$$\sum_{i\in E} \pi_i p_{ij} = \pi_j \text{ para toda } j \in E,$$

se dice que $\pi = (\pi_i)_{i\in E}$ es una ***distribución estacionaria*** *de la cadena.*

Para el ejemplo 3, podemos notar que $\pi = \begin{pmatrix} 1/3 \\ 2/3 \end{pmatrix}$ es una distribución estacionaria de la cadena, pues

$$\pi_0 p_{00} + \pi_1 p_{10} = \frac{1}{3} \cdot \frac{1}{2} + \frac{2}{3} \cdot \frac{1}{4} = \frac{1}{6} + \frac{2}{12} = \frac{1}{3} = \pi_0$$

y

$$\pi_0 p_{01} + \pi_1 p_{11} = \frac{1}{3} \cdot \frac{1}{2} + \frac{2}{3} \cdot \frac{3}{4} = \frac{2}{12} + \frac{6}{12} = \frac{2}{3} = \pi_1.$$

Teorema 2. *Sea $\{X_n\}$ una cadena de Markov homogénea con espacio de estados E. Si la cadena $\{X_n\}$ es irreducible, entonces la cadena tiene una única distribución estacionaria π.*

La demostración de este teorema es consecuencia de los resultados contenidos en [7, p. 62] y [7, p. 64]. [1]

Cadenas de Markov regulares

Una vez introducidos los conceptos y resultados anteriores, es importante notar que no necesariamente todas las cadenas de Markov homogéneas cuentan con una matriz

[1] Es importante mencionar que el teorema de Perron-Frobenius y este resultado pueden extenderse al caso de matrices infinitas, pero este caso no será objeto de estudio en esta tesis.

de transición regular. Como el teorema de Perron-Frobenius sólo puede ser aplicado a matrices regulares, comenzaremos estudiando qué condiciones son suficientes para asegurar que la matriz de transición de una cadena de Markov homogénea sea regular, y veremos qué ocurre si no lo es.

Definición 13. *Sea $\{X_n\}$ una cadena de Markov homogénea con espacio de estados E y P su matriz de transición. Decimos que la cadena $\{X_n\}$ es una **cadena de Markov regular** si la matriz P es regular, es decir, si existe $k \in \mathbb{N}$ tal que $P^k > 0$.*

Nuevamente el ejemplo 3 ilustra nuestra definición, pues la matriz de transición P de la cadena es positiva y por lo tanto la cadena de Markov es regular.

Teorema 3. *Sea $\{X_n\}$ una cadena de Markov homogénea con espacio de estados E. Si $\{X_n\}$ es irreducible y aperiódica, entonces $\{X_n\}$ es una cadena de Markov regular.*

La demostración de este teorema se dividirá en dos partes: primero nos concentraremos en probar que para cualesquiera dos estados x, y existe una n_{xy} para la cual la probabilidad de llegar en m pasos de x a y es positiva si m es mayor o igual a n_{xy} (es decir, $p_{xy}^{(m)} > 0$ para cualquier $m \geq n_{xy}$), y en el segundo paso utilizaremos esto para definir una k para la cual P^k sea estrictamente positiva.

Debemos mencionar que en la primera parte será necesario el siguiente resultado:

Lema 3. *Sea I un conjunto no vacío de enteros positivos tales que*

- *$MCD(I) = 1$;*
- *si $a, b \in I$ entonces $a + b \in I$.*

Entonces existe $n_0 \in I$ para el cual $m \in I$ para cualquiera $m \geq n_0$.

La demostración de este lema puede ser encontrada en [7, p. 79-80].

Demostración del teorema 3.
Sea $\ell \in E$. Para esta ℓ arbitraria y fija, definimos el conjunto $I_\ell = \{m \geq 1 \mid p_{\ell\ell}^{(m)} > 0\}$. Notemos dos cosas:

- Como la cadena es aperiódica, $1 = d = MCD(\{m \geq 1 \mid p_{\ell\ell}^{(m)} > 0\}) = MCD(I_\ell)$.
- Sean $a, b \in I_\ell$. Ya que $p_{\ell\ell}^{(a)} > 0$ y $p_{\ell\ell}^{(b)} > 0$, se sigue que

$$p_{\ell\ell}^{(a+b)} = \sum_{i=1}^{|E|} p_{\ell i}^{(a)} p_{i\ell}^{(a)} \geq p_{\ell\ell}^{(a)} p_{\ell\ell}^{(b)} > 0,$$

  y por lo tanto $a + b \in I_\ell$.

Como se cumplen las condiciones del lema 3 para I_ℓ, existe un $n_\ell \in I_\ell$ para el cual $m \in I_\ell$ para cualquiera $m \geq n_\ell$, es decir, $p_{\ell\ell}^{(m)} > 0$ para cualquier $m \geq n_\ell$.

Sean $x, y \in E$. Como la cadena es irreducible, existen $n_{x\ell}$ y $n_{\ell y}$ enteros positivos tales que $p_{x\ell}^{(n_{x\ell})} > 0$ y $p_{\ell y}^{(n_{\ell y})} > 0$. Definimos $n_{xy} = n_{x\ell} + n_{\ell\ell} + n_{\ell j}$.

Sea $m \geq n_{xy}$. Notemos que $m = n_{x\ell} + (n_{\ell\ell} + r) + n_{\ell j}$ para alguna $r \geq 0$, y ya que $p_{x\ell}^{(n_{x\ell})} > 0$, ocurre que $p_{\ell y}^{(n_{\ell y})} > 0$ y $p_{\ell\ell}^{(n_{\ell\ell}+r)} > 0$ (esto porque $n_{\ell\ell} + r \geq n_{\ell\ell}$), se sigue que

$$p_{xy}^{(m)} = p_{xy}^{(n_{x\ell}+(n_{\ell\ell}+r)+n_{\ell j})} \geq p_{x\ell}^{(n_{x\ell})} p_{\ell\ell}^{(n_{\ell\ell}+r)} p_{xy}^{(n_{\ell j})} > 0.$$

Esto prueba que, para cualesquiera $x, y \in E$, existe n_{xy} tal que $p_{xy}^{(m)} > 0$ para toda $m \geq n_{xy}$.

Ahora, para la segunda parte de la demostración, definimos $k = \text{máx}\{n_{xy} \mid x, y \in E\}$. Sean $i, j \in E$. Como $k \geq n_{ij}$, entonces $p_{ij}^{(k)} > 0$ y por lo tanto $P^k > 0$. Esto demuestra que la matriz P es regular, y por lo tanto la cadena $\{X_n\}$ es regular. □

Gracias al teorema 3, sabemos que las cadenas irreducibles y aperiódicas cuentan con una matriz de transición regular. A continuación aplicaremos el teorema de Perron-Frobenius a la matriz de transición de una cadena con estas características.

Teorema 4. *Sea $\{X_n\}$ una cadena de Markov homogénea irreducible y aperiódica con espacio de estados E, $|E| = N$, y $P \in M_{N\times N}(\mathbb{R})$ la matriz de transición de la cadena. Entonces:*
(a) Existe una única distribución estacionaria π para $\{X_n\}$.
(b) 1 es el eigenvalor de Perron-Frobenius de P asociado a los eigenvectores derecho $\overline{1}$ e izquierdo π, donde $\overline{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^N$ y π es la distribución estacionaria de la cadena. Además, para cualquier eigenvalor $\theta_i \in \mathbb{C}$ de P se tiene que $|\theta_i| < 1$, con $i \in \{2, 3, \ldots, N\}$.

Demostración.
Sean $\{X_n\}$ una cadena de Markov homogénea irreducible y aperiódica con espacio de estados E, y $P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \vdots & \vdots & & \vdots \\ p_{N1} & p_{N2} & \cdots & p_{NN} \end{pmatrix}$ su matriz de transición.

(a) Por el teorema 3, $\{X_n\}$ es una cadena de Markov regular y por lo tanto existe $k \in \mathbb{N}$ para la cual $P^k > 0$. Debido a esto, $p_{ij}^{(k)} > 0$ para cualesquiera estados i, j, lo que implica que la cadena es irreducible, y por el teorema 2, $\{X_n\}$ tiene una única distribución estacionaria π.

(b) Por el inciso anterior sabemos que la matriz P es regular. Utilizando el teorema de Perron-Frobenius, existe un eigenvalor r de P tal que r es un número real y además es el eigenvalor más grande en norma compleja. Para demostrar este inciso, primero comprobaremos que 1 es un eigenvalor de la matriz P y posteriormente que es el eigenvalor más grande, lo que prueba que es el eigenvalor de Perron-Frobenius.

Consideremos el vector $\overline{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^N$. Notemos que

$$\begin{aligned} P\overline{1} &= \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1N} \\ p_{21} & p_{22} & \dots & p_{2N} \\ \vdots & \vdots & & \vdots \\ p_{N1} & p_{N2} & \dots & p_{NN} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} p_{11} + p_{12} + \dots + p_{1N} \\ p_{21} + p_{22} + \dots + p_{2N} \\ \vdots \\ p_{N1} + p_{N2} + \dots + p_{NN} \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \\ &= \overline{1}, \end{aligned}$$

por lo que

$$P\overline{1} = 1 \cdot \overline{1} \tag{2.1}$$

por lo cual 1 es un eigenvalor de P con eigenvector derecho $\overline{1}$.

Por otro lado, por (a) se tiene que $\sum_{i=1}^{N} \pi_i p_{ij} = \pi_j$ para toda $j \in E$, por lo que

$$\begin{aligned} \pi' P &= (\pi_1 \ \pi_2 \ \dots \ \pi_N) \begin{pmatrix} p_{11} + p_{12} + \dots + p_{1N} \\ p_{21} + p_{22} + \dots + p_{2N} \\ \vdots \\ p_{N1} + p_{N2} + \dots + p_{NN} \end{pmatrix} \\ &= (\sum_{i=1}^{N} \pi_i p_{i1} \ \sum_{i=1}^{N} \pi_i p_{i2} \ \dots \ \sum_{i=1}^{N} \pi_i p_{i2}) \\ &= (\pi_1 \ \pi_2 \ \dots \ \pi_N) \\ &= \pi', \end{aligned}$$

y por lo tanto π es un eigenvector izquierdo asociado a 1.

Ahora, para demostrar que 1 es el eigenvalor de Perron-Frobenius, tomaremos otro eigenvalor de la matriz P y probaremos que su norma compleja es menor o igual a 1, lo que prueba que 1 es el eigenvalor máximo en norma compleja.

Sean $\lambda \in \mathbb{C}$ un eigenvalor de la matriz P y $\phi = \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_N \end{pmatrix}$ el eigenvector derecho asociado a ese eigenvalor, es decir, $\lambda\phi = P\phi$. Desarrollando esta igualdad se tiene que

$$\begin{pmatrix} \lambda\phi_1 \\ \lambda\phi_2 \\ \vdots \\ \lambda\phi_N \end{pmatrix} = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1N} \\ p_{21} & p_{22} & \dots & p_{2N} \\ \vdots & \vdots & & \vdots \\ p_{N1} & p_{N2} & \dots & p_{NN} \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_N \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{N} p_{1i}\phi_i \\ \sum_{i=1}^{N} p_{2i}\phi_i \\ \vdots \\ \sum_{i=1}^{N} p_{Ni}\phi_i \end{pmatrix}.$$

Esto muestra que

$$\lambda\phi_j = \sum_{i=1}^{N} p_{ji}\phi_i \text{ para cualquier } j = 1,2,...,N. \tag{2.2}$$

Ya que ϕ tiene un número finito de entradas, existe $k \in \{1,2,...,N\}$ para la cual $|\phi_k| = \max_{1\le i\le N} |\phi_i|$. Tomando la norma en (2.2) para esta k particular se tiene que

$$\begin{aligned} |\lambda\phi_k| = |\sum_{i=1}^{N} p_{ki}\phi_i| &\le \sum_{i=1}^{N} p_{ki}|\phi_i| \\ &\le \sum_{i=1}^{N} p_{ki}|\phi_k| = |\phi_k| \sum_{i=1}^{N} p_{ki}. \end{aligned}$$

Como P es una matriz de transición, $\sum_{i=1}^{N} p_{ki} = 1$, por lo que hemos visto que

$$|\lambda\phi_k| \le |\phi_k|. \tag{2.3}$$

Como ϕ es un eigenvector derecho de P, entonces ϕ es distinto de cero y por lo tanto $\max_{1\le i\le N} |\phi_i| = |\phi_k| > 0$. Considerando esto, por (2.3) se sigue que

$$|\lambda| \le 1. \tag{2.4}$$

Finalmente, sea r el eigenvalor de Perron-Frobenius de la matriz P. Hemos probado en (2.1) que 1 es un eigenvalor de la matriz P, por lo que el inciso (c) del Teorema de Perron-Frobenius implica que $1 \le r$. Por otro lado, por (2.4) se tiene que $r \le 1$. Esto demuestra que $r = 1$, es decir, el eigenvalor de Perron-Frobenius de la matriz P es 1. $\square$

El teorema anterior demuestra que 1 es el eigenvalor de Perron-Frobenius de la matriz P. Sin embargo, es importante notar que esto no sólo nos brinda información sobre dicho eigenvalor, sino también de todos los demás eigenvalores de la matriz. Si $\theta \in \mathbb{C}$ es un eigenvalor de la matriz P distinto del eigenvalor de Perron-Frobenius, por el inciso (c) del teorema de Perron-Frobenius se tiene que $|\theta|$ es estrictamente menor que el eigenvalor de Perron-Frobenius, es decir, $|\theta| < 1$.

Terminaremos este capítulo utilizando el resultado anterior para estudiar la velocidad de convergencia de la sucesión $\{p_{ij}^{(m)}\}$ a la distribución estacionaria π. Esta convergencia se da por el siguiente teorema:

Teorema 5. *Sea $\{X_n\}$ una cadena de Markov homogénea irreducible con espacio de estados E y π la distribución estacionaria de la cadena. Si la cadena es aperiódica, entonces*

$$\lim_{m\to\infty} p_{ij}^{(m)} = \pi_j \text{ para cualesquiera } i, j \in E.$$

La demostración de este teorema puede ser consultada en [7, p. 73].

Para estudiar la velocidad de convergencia de $\{p_{ij}^{(m)}\}$ a π_j cuando m tiende a infinito utilizaremos la siguiente definición:

Definición 14. *Sean f una función con valores en $\mathbb{R}$ ó $\mathbb{C}$, y g una función que toma sólo valores positivos. Decimos que*

$$f(m) = O(g(m)) \text{ si } \overline{\lim_{m\to\infty}} \frac{|f(m)|}{g(m)} < \infty.$$

Ejemplo 4. Sean $f, g : \mathbb{R} \to \mathbb{R}$ dos funciones definidas como $f(x) = x$ y $g(x) = x^2 + 1$. Notemos que

$$\overline{\lim_{m\to\infty}} \frac{|f(m)|}{g(m)} = \overline{\lim_{m\to\infty}} \frac{|m|}{(m^2+1)} \le \overline{\lim_{m\to\infty}} \frac{|m|}{m^2} = \overline{\lim_{m\to\infty}} \frac{1}{|m|} = 0 < \infty,$$

por lo que $f(m) = O(g(m))$.

Teorema 6. *Sea $\{X_n\}$ una cadena de Markov homogénea irreducible y aperiódica con espacio de estados E y $P \in M_{N\times N}(\mathbb{R})$ su matriz de transición. Si $\theta_1, \theta_2, ..., \theta_N \in \mathbb{C}$ son los eigenvalores de P y $|\theta_1| > |\theta_2| > ... > |\theta_N|$, entonces*

$$|p_{ij}^{(m)} - \pi_j| = O(|\theta_2|^m) \text{ cuando } m \to \infty,$$

donde $\pi = (\pi_i)_{i\in E}$ es la distribución estacionaria de la cadena.

Este teorema está contenido en [8, p. 284] y para demostrarlo construiremos primero una expresión de la matriz P^m en términos de sus eigenvalores, eigenvectores derechos y eigenvectores izquierdos. El procedimiento que a continuación realizaremos tendrá por objeto encontrar dicha expresión.

Para cada θ_i consideremos su eigenvector derecho asociado $\alpha_i = \begin{pmatrix} \phi_1^i \\ \phi_2^i \\ \vdots \\ \phi_N^i \end{pmatrix}$.

Definimos $\Phi = (\alpha_1\ \alpha_2\ \dots\ \alpha_N) = \begin{pmatrix} \phi_1^1 & \phi_1^2 & \dots & \phi_1^N \\ \phi_2^1 & \phi_2^2 & \dots & \phi_2^N \\ \vdots & \vdots & & \vdots \\ \phi_N^1 & \phi_N^2 & \dots & \phi_N^N \end{pmatrix}$.

Como $P\alpha_i = \theta_i \alpha_i$ para cada $i = 1, 2, \dots, N$,

$$\begin{pmatrix} p_{11} & p_{12} & \dots & p_{1N} \\ p_{21} & p_{22} & \dots & p_{2N} \\ \vdots & \vdots & & \vdots \\ p_{N1} & p_{N2} & \dots & p_{NN} \end{pmatrix} \begin{pmatrix} \phi_1^i \\ \phi_2^i \\ \vdots \\ \phi_N^i \end{pmatrix} = \begin{pmatrix} \theta_i\phi_1^i \\ \theta_i\phi_2^i \\ \vdots \\ \theta_i\phi_N^i \end{pmatrix}.$$

Desarrollando la multiplicación del lado izquierdo en esta igualdad se tiene que

$$\begin{pmatrix} \sum_{j=1}^{N} p_{1j}\phi_j \\ \sum_{j=1}^{N} p_{2j}\phi_j \\ \vdots \\ \sum_{j=1}^{N} p_{Nj}\phi_j \end{pmatrix} = \begin{pmatrix} \theta_i\phi_1^i \\ \theta_i\phi_2^i \\ \vdots \\ \theta_i\phi_N^i \end{pmatrix}. \tag{2.5}$$

Considerando que

$$\begin{aligned} P\Phi &= \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1N} \\ p_{21} & p_{22} & \dots & p_{2N} \\ \vdots & \vdots & & \vdots \\ p_{N1} & p_{N2} & \dots & p_{NN} \end{pmatrix} \begin{pmatrix} \phi_1^1 & \phi_1^2 & \dots & \phi_1^N \\ \phi_2^1 & \phi_2^2 & \dots & \phi_2^N \\ \vdots & \vdots & & \vdots \\ \phi_N^1 & \phi_N^2 & \dots & \phi_N^N \end{pmatrix} \\ &= \begin{pmatrix} \sum_{j=1}^{N} p_{1j}\phi_j^1 & \sum_{j=1}^{N} p_{1j}\phi_j^2 & \dots & \sum_{j=1}^{N} p_{1j}\phi_j^N \\ \sum_{j=1}^{N} p_{2j}\phi_j^1 & \sum_{j=1}^{N} p_{2j}\phi_j^2 & \dots & \sum_{j=1}^{N} p_{2j}\phi_j^N \\ \vdots & \vdots & & \vdots \\ \sum_{j=1}^{N} p_{Nj}\phi_j^1 & \sum_{j=1}^{N} p_{Nj}\phi_j^2 & \dots & \sum_{j=1}^{N} p_{Nj}\phi_j^N \end{pmatrix}, \end{aligned}$$

por (2.5) se tiene que

$$
\begin{aligned}
P\Phi &= \begin{pmatrix} \theta_1\phi_1^1 & \theta_2\phi_1^2 & \dots & \theta_N\phi_1^N \\ \theta_1\phi_2^1 & \theta_2\phi_2^2 & \dots & \theta_N\phi_2^N \\ \vdots & \vdots & & \vdots \\ \theta_1\phi_N^1 & \theta_2\phi_N^2 & \dots & \theta_N\phi_N^N \end{pmatrix} \\
&= \begin{pmatrix} \phi_1^1 & \phi_1^2 & \dots & \phi_1^N \\ \phi_2^1 & \phi_2^2 & \dots & \phi_2^N \\ \vdots & \vdots & & \vdots \\ \phi_N^1 & \phi_N^2 & \dots & \phi_N^N \end{pmatrix} \begin{pmatrix} \theta_1 & 0 & \dots & 0 \\ 0 & \theta_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \theta_N \end{pmatrix}.
\end{aligned}
$$

Si denotamos como $diag(\theta) = \begin{pmatrix} \theta_1 & 0 & \dots & 0 \\ 0 & \theta_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \theta_N \end{pmatrix}$, por el procedimiento anterior se tiene que

$$P\Phi = \Phi diag(\theta). \tag{2.6}$$

El siguiente paso involucrará a la matriz Φ^{-1}. Sin embargo, antes debemos comprobar que la matriz Φ es invertible. Para esto, utilizaremos el siguiente lema:

Lema 4. *Si los eigenvalores $\theta_1, \theta_2, ..., \theta_N$ son distintos, entonces sus respectivos vectores derechos asociados $\alpha_1, \alpha_2, ..., \alpha_N$ son linealmente independientes.*

La demostración de este lema puede ser consultada en [8, p. 274-275].

Debido al lema 4, el conjunto $\{\alpha_1, \alpha_2, ..., \alpha_N\}$ es una base del espacio $\mathbb{R}^N$ y por lo tanto la matriz $\Phi = (\alpha_1\ \alpha_2\ ...\ \alpha_N)$ es invertible.

Multiplicando a cada lado de la igualdad en (2.6) por Φ^{-1}, tanto por la izquierda y como por la derecha,

$$\Phi^{-1}P\Phi\Phi^{-1} = \Phi^{-1}\Phi diag(\theta)\Phi^{-1},$$

y desarrollando ambos lados de la igualdad se tiene que

$$\Phi^{-1}P = diag(\theta)\Phi^{-1}. \tag{2.7}$$

Si $\Phi^{-1} = \begin{pmatrix} \psi_1^1 & \psi_2^1 & \dots & \psi_N^1 \\ \psi_1^2 & \psi_2^2 & \dots & \psi_N^2 \\ \vdots & \vdots & & \vdots \\ \psi_1^N & \psi_2^N & \dots & \psi_N^N \end{pmatrix}$, de la parte izquierda de (2.7) se sigue que

$$
\begin{aligned}
\Phi^{-1}P &= \begin{pmatrix} \psi_1^1 & \psi_2^1 & \dots & \psi_N^1 \\ \psi_1^2 & \psi_2^2 & \dots & \psi_N^2 \\ \vdots & \vdots & & \vdots \\ \psi_1^N & \psi_2^N & \dots & \psi_N^N \end{pmatrix} \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1N} \\ p_{21} & p_{22} & \dots & p_{2N} \\ \vdots & \vdots & & \vdots \\ p_{N1} & p_{N2} & \dots & p_{NN} \end{pmatrix} \\
&= \begin{pmatrix} \sum\limits_{j=1}^{N} \psi_j^1 p_{j1} & \sum\limits_{j=1}^{N} \psi_j^1 p_{j2} & \dots & \sum\limits_{j=1}^{N} \psi_j^1 p_{jN} \\ \sum\limits_{j=1}^{N} \psi_j^2 p_{j1} & \sum\limits_{j=1}^{N} \psi_j^2 p_{j2} & \dots & \sum\limits_{j=1}^{N} \psi_j^2 p_{jN} \\ \vdots & \vdots & & \vdots \\ \sum\limits_{j=1}^{N} \psi_j^N p_{j1} & \sum\limits_{j=1}^{N} \psi_j^N p_{j2} & \dots & \sum\limits_{j=1}^{N} \psi_j^N p_{jN} \end{pmatrix},
\end{aligned}
$$

mientras que en la parte derecha de (2.7) se tiene que

$$
\begin{aligned}
diag(\theta)\Phi &= \begin{pmatrix} \theta_1 & 0 & \dots & 0 \\ 0 & \theta_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \theta_N \end{pmatrix} \begin{pmatrix} \psi_1^1 & \psi_2^1 & \dots & \psi_N^1 \\ \psi_1^2 & \psi_2^2 & \dots & \psi_N^2 \\ \vdots & \vdots & & \vdots \\ \psi_1^N & \psi_2^N & \dots & \psi_N^N \end{pmatrix} \\
&= \begin{pmatrix} \theta_1 \psi_1^1 & \theta_1 \psi_2^1 & \dots & \theta_1 \psi_N^1 \\ \theta_2 \psi_1^2 & \theta_2 \psi_2^2 & \dots & \theta_2 \psi_N^2 \\ \vdots & \vdots & & \vdots \\ \theta_N \psi_1^N & \theta_N \psi_2^N & \dots & \theta_N \psi_N^N \end{pmatrix}.
\end{aligned}
$$

Si $\Phi^{-1} = \begin{pmatrix} \beta_1' \\ \beta_2' \\ \vdots \\ \beta_N' \end{pmatrix}$, donde $\beta_i' = (\psi_1^i \ \psi_2^i \ ... \ \psi_N^i)$ para cada $i = 1,2,...,N$, entonces

$$\begin{aligned}
\beta_i' P &= (\psi_1^i \ \psi_2^i \ ... \ \psi_N^i) \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \vdots & \vdots & & \vdots \\ p_{N1} & p_{N2} & \cdots & p_{NN} \end{pmatrix} \\
&= (\sum_{j=1}^{N} \psi_j^i p_{j1} \ \sum_{j=1}^{N} \psi_j^i p_{j2} \ ... \ \sum_{j=1}^{N} \psi_j^i p_{jN}) \\
&= (\theta_i \psi_1^i \ \theta_i \psi_2^i \ ... \ \theta_i \psi_N^i) \\
&= \theta_i (\psi_1^i \ \psi_2^i \ ... \ \psi_N^i) \\
&= \theta_i \beta_i'
\end{aligned}$$

Esto prueba que cada β_i es un eigenvector izquierdo de P con eigenvalor asociado θ_i.

Retomando (2.6),

$$\begin{aligned}
P^m &= \Phi diag(\theta)^m \Phi^{-1} \\
&= (\alpha_1 \ \alpha_2 \ ... \ \alpha_N) diag(\theta)^m \begin{pmatrix} \beta_1' \\ \beta_2' \\ \vdots \\ \beta_N' \end{pmatrix} \\
&= (\theta_1^m \alpha_1 \ \theta_2^m \alpha_2 \ ... \ \theta_N^m \alpha_N) \begin{pmatrix} \beta_1' \\ \beta_2' \\ \vdots \\ \beta_N' \end{pmatrix} \\
&= \sum_{i=1}^{N} \theta_i^m \alpha_i \beta_i'
\end{aligned}$$

para cualquier $m \in \mathbb{N}$. A esta expresión se le llama **representación espectral** de P^m.

Demostración del teorema 6.
Sea $m \in E$. La representación espectral de P^m es

$$P^m = \sum_{\ell=1}^{N} \theta_\ell^m \alpha_\ell \beta_\ell', \tag{2.8}$$

donde cada $\alpha_i = (\phi_j^i)_{j\in E}$ es el eigenvector derecho asociado a θ_i y cada $\beta_i = (\psi_j^i)_{j\in E}$ es el eigenvector izquierdo asociado a θ_i.

El primer paso será buscar una expresión de $p_{ij}^{(m)}$ para cada $i,j\in E$. Desarrollando (2.8) se tiene que

$$\begin{aligned}
\begin{pmatrix} p_{11}^{(m)} & p_{12}^{(m)} & \dots & p_{1N}^{(m)} \\ p_{21}^{(m)} & p_{22}^{(m)} & \dots & p_{2N}^{(m)} \\ \vdots & \vdots & \dots & \vdots \\ p_{N1}^{(m)} & p_{N2}^{(m)} & \dots & p_{NN}^{(m)} \end{pmatrix} &= \sum_{\ell=1}^{N} \theta_\ell^m \begin{pmatrix} \phi_1^\ell \\ \phi_2^\ell \\ \vdots \\ \phi_N^\ell \end{pmatrix} (\psi_1^\ell \ \psi_2^\ell \ \dots \ \psi_N^\ell) \\
&= \sum_{\ell=1}^{N} \theta_\ell^m \begin{pmatrix} \phi_1^\ell\psi_1^\ell & \phi_1^\ell\psi_2^\ell & \dots & \phi_1^\ell\psi_N^\ell \\ \phi_2^\ell\psi_1^\ell & \phi_2^\ell\psi_2^\ell & \dots & \phi_2^\ell\psi_N^\ell \\ \vdots & \vdots & \dots & \vdots \\ \phi_N^\ell\psi_1^\ell & \phi_N^\ell\psi_2^\ell & \dots & \phi_N^\ell\psi_N^\ell \end{pmatrix} \\
&= \sum_{\ell=1}^{N} \begin{pmatrix} \theta_\ell^m\phi_1^\ell\psi_1^\ell & \theta_\ell^m\phi_1^\ell\psi_2^\ell & \dots & \theta_\ell^m\phi_1^\ell\psi_N^\ell \\ \theta_\ell^m\phi_2^\ell\psi_1^\ell & \theta_\ell^m\phi_2^\ell\psi_2^\ell & \dots & \theta_\ell^m\phi_2^\ell\psi_N^\ell \\ \vdots & \vdots & \dots & \vdots \\ \theta_\ell^m\phi_N^\ell\psi_1^\ell & \theta_\ell^m\phi_N^\ell\psi_2^\ell & \dots & \theta_\ell^m\phi_N^\ell\psi_N^\ell \end{pmatrix},
\end{aligned}$$

por lo que

$$p_{ij}^{(m)} = \sum_{\ell=1}^{N} \theta_\ell^m \phi_i^\ell \psi_j^\ell \text{ para cada } i,j\in E. \tag{2.9}$$

Como la cadena es irreducible y aperiódica, entonces la cadena es regular y por lo tanto el teorema de Perron-Frobenius puede ser aplicado a la matriz de transición. Como $|\theta_1| > |\theta_\ell|$ para cualquier $\ell \in \{2,\dots,N\}$, por el inciso (c) del teorema de Perron-Frobenius se tiene que θ_1 es el eigenvalor de Perron-Frobenius de la matriz P, y por el teorema 4 se sigue que $\theta_1 = 1$, $\alpha_1 = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$ y $\beta_1 = \pi$. Utilizando esto en (2.9),

$$\begin{aligned}
p_{ij}^{(m)} &= \theta_1^m \phi_i^1 \psi_j^1 + \sum_{\ell=2}^{N} \theta_\ell^m \phi_i^\ell \psi_j^\ell \\
&= \pi_j + \sum_{\ell=2}^{N} \theta_\ell^m \phi_i^\ell \psi_j^\ell \text{ para cada } i,j\in E.
\end{aligned}$$

Sean $i,j\in E$. Por la igualdad anterior,

$$p_{ij}^{(m)} - \pi_j = \sum_{\ell=2}^{N} \theta_\ell^m \phi_i^\ell \psi_j^\ell,$$

y tomando la norma se sigue que

$$|p_{ij}^{(m)} - \pi_j| = |\sum_{\ell=2}^{N} \theta_\ell^m \phi_i^\ell \psi_j^\ell| \leq \sum_{\ell=2}^{N} |\theta_\ell|^m |\phi_i^\ell||\psi_j^\ell|. \tag{2.10}$$

Como $|\theta_N| < \dots < |\theta_2|$, por (2.10) se tiene que

$$|p_{ij}^{(m)} - \pi_j| \leq \sum_{\ell=2}^{N} |\theta_2|^m |\phi_i^\ell||\psi_j^\ell| = |\theta_2|^m \sum_{\ell=2}^{N} |\phi_i^\ell||\psi_j^\ell|.$$

Diviendo entre $|\theta_2|^m$,

$$\left| \frac{p_{ij}^{(m)} - \pi_j}{\theta_2^m} \right| \leq \sum_{\ell=2}^{N} |\phi_i^\ell||\psi_j^\ell|. \tag{2.11}$$

Hemos probado que la sucesión $\{|\frac{p_{ij}^{(m)} - \pi_j}{\theta_2^m}|\}$ está acotada, por lo que su límite superior existe y, por lo tanto, $|p_{ij}^{(m)} - \pi_j| = O(|\theta_2|^m)$ cuando $m \to \infty$.

$\square$

Capítulo 3

Grandes desviaciones para cadenas de Markov regulares

Consideremos una sucesión de variables aleatorias $\{X_n\}$ independientes e idénticamente distribuidas (i.i.d.), cada una con esperanza $\mu = \mathbb{E}[X_1] \in \mathbb{R}$. Por la Ley Débil de los Grandes Números sabemos que, para $a > 0$ fija, la probabilidad del evento

$$\left| \frac{\sum_{i=1}^{n} X_i}{n} - \mu \right| > a$$

converge a cero cuando n tiende a infinito. Un Principio de Grandes Desviaciones es un criterio que busca estudiar la velocidad de convergencia a cero de la probabilidad de este evento, y en este capítulo trataremos con el caso de una funcional de una cadena de Markov (notemos que una cadena de Markov es una sucesión de variables aleatorias que no necesariamente son i.i.d.).

El principal material utilizado para este capítulo es [5] y [4].

Principio de Grandes Desviaciones (P.G.D.)

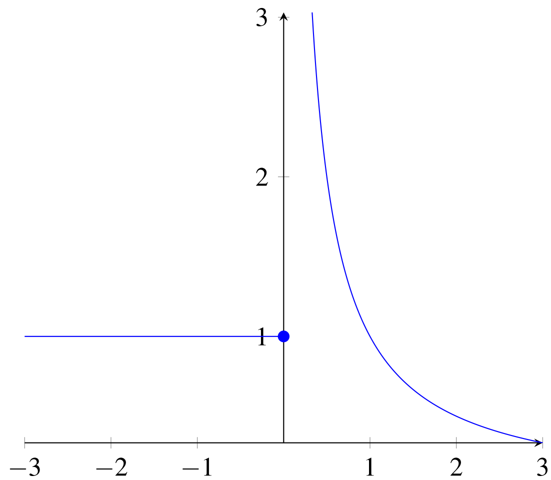
Para introducir el Principio de Grandes Desviaciones que utilizaremos en este capítulo son necesarios dos conceptos: semicontinuidad inferior y función de grandes desviaciones, por lo que comenzaremos esta sección introduciendo dichos conceptos.

Definición 15. *Sea (X, τ) un espacio topológico. Una función $f : X \to \mathbb{R} \cup \{-\infty, \infty\}$ es **semicontinua inferiormente (s.c.i.)** si $f^{-1}((a, \infty))$ es un conjunto abierto para cualquier $a \in \mathbb{R} \cup \{-\infty, \infty\}$.*

Ejemplo 5. Definimos $f : \mathbb{R} \to \mathbb{R} \cup \{-\infty, \infty\}$ como

$$f(x) = \begin{cases} 1 & \text{si } x \leq 0 \\ \frac{1}{x} & \text{si } x > 0 \end{cases}$$

A continuación presentamos la gráfica de esta función:



La función f es un ejemplo de una función semicontinua inferiormente (s.c.i.), pues dada $a \in \mathbb{R} \cup \{-\infty, \infty\}$,

- si $a > 1$, $f^{-1}((a, \infty)) = (0, \frac{1}{a})$ es abierto;
- si $a \in (0, 1]$, $f^{-1}((a, \infty)) = (-\infty, \frac{1}{a})$ es abierto;
- y si $a \leq 0$, $f^{-1}((a, \infty)) = \mathbb{R}$ es abierto.

Es importante mencionar que cualquier función continua es s.c.i. Además, dados (X, τ) un espacio topológico y $f : X \to \mathbb{R} \cup \{-\infty, \infty\}$ una función, pedir que $f^{-1}((a, \infty)) = \{x \in X \mid f(x) \in (-\infty, \infty)\}$ sea abierto es equivalente a pedir que el conjunto $\{x \in X \mid f(x) \leq a\}$ sea cerrado. Es decir, f es s.c.i. si y sólo si

$$\psi_1(a) = \{x \in X \mid f(x) \leq a\}$$

es un conjunto cerrado para cualquier $a \in \mathbb{R} \cup \{-\infty, \infty\}$.

Utilizando este concepto, definimos función de grandes desviaciones como sigue:

Definición 16. *Sea (X, τ) un espacio topológico e $I : X \to [0, \infty]$ una función.*

- *Decimos que I es una* ***función de grandes desviaciones*** *si I es s.c.i.*
- *Decimos que I es una* ***buena función de grandes desviaciones*** *si $\psi_1(a) = \{x \in X \mid f(x) \leq a\}$ es compacto para cualquier $a \in \mathbb{R} \cup \{-\infty, \infty\}$.*

Notemos que cualquier buena función de grandes desviaciones es una función de grandes desviaciones. Ahora procederemos a definir el Principio de Grandes Desviaciones.

Definición 17 (Principio de Grandes Desviaciones)**.** *Sea $\{\mu_n\}$ una familia de medidas de probabilidad en $\mathbb{R}^d$. Se dice que $\{\mu_n\}$ satisface el* ***Principio de Grandes Desviaciones (P.G.D.)*** *si existen una sucesión $\{x_n\} \subseteq \mathbb{R}$ que converge a 0 cuando n tiende a infinito y una función de grandes desviaciones $I : \mathbb{R}^d \to [0, \infty]$ tales que*

$$-\inf_{x\in int(A)} I(x) \leq \underline{\lim}_{n\to\infty} x_n \; log\; \mu_n(A) \leq \overline{\lim}_{n\to\infty} x_n \; log\; \mu_n(A) \leq -\inf_{x\in \bar{A}} I(x)$$

para cualquier $A \subseteq \mathbb{R}^d$.

Es importante mencionar que esta definición puede ser extendida a medidas de probabilidad en espacios más generales, pero en este trabajo nos concentraremos en el caso donde el espacio muestral es $\mathbb{R}^d$.

El Teorema de Gärtner-Ellis

El teorema de Gärtner-Ellis establece condiciones bajo las cuales una familia de medidas de probabilidad cumple el P.G.D. con una buena función de grandes desviaciones Λ^*. Antes de presentar dicho teorema es necesario introducir los conceptos de función generadora de momentos logarítmica y transformada de Fenchel-Legendre, pues son indispensables para definir la función Λ^*.

Definición 18. *Sea X una variable aleatoria d-dimensional. Definimos la* ***función generadora de momentos logarítmica*** $\Lambda : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty, \infty\}$ *como*

$$\Lambda(\lambda) = log\; \mathbb{E}[e^{\langle \lambda, X\rangle}],$$

donde $\langle \lambda, X\rangle$ *es el producto interno usual.*

Es importante notar que

- $\Lambda(\lambda) \in (-\infty, \infty]$ para toda $\lambda \in \mathbb{R}^d$,
- $\Lambda(0) = log\; \mathbb{E}[e^{\langle 0, X\rangle}] = 0$ y
- si $d = 1$, entonces $\Lambda(\lambda) = log\; \mathbb{E}[e^{\lambda X}]$.

Ejemplo 6. Sea X una variable aleatoria Bernoulli con parámetro p. Su función generadora de momentos es

$$\begin{aligned}
\Lambda(\lambda) &= log\; \mathbb{E}[e^{\lambda X}] \\
&= log\Big(e^{\lambda\cdot 0}(1-p) + e^{\lambda \cdot 1}(p)\Big) \\
&= log\big(1 - p + e^{\lambda} p\big) \\
&= log\big(p\; [e^{\lambda} - 1] + 1\big).
\end{aligned}$$

Definición 19. *Sea* $f : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty, \infty\}$ *una función. Definimos la* ***transformada de Fenchel-Legendre*** *de* f *como*

$$f^*(x) = \sup_{\lambda \in \mathbb{R}^d} \{\langle \lambda, x\rangle - f(\lambda)\}.$$

Ejemplo 7. Sea $f : \mathbb{R} \to \mathbb{R}$ la función definida como $f(x) = x^2$, y para cada $x \in \mathbb{R}$ fija definimos

$$g(\lambda) = \lambda x - \lambda^2.$$

Podemos notar que $g'(\lambda) = x - 2\lambda$ es igual a cero cuando $\lambda = \frac{x}{2}$, y que $g''(\lambda) = -2$, por lo que la función g alcanza su máximo cuando $\lambda = \frac{x}{2}$. Por lo tanto, la transformada de Fenchel-Legendre de la función f es

$$f^*(x) = \sup_{\lambda \in \mathbb{R}^d} \{\langle \lambda, x\rangle - f(\lambda)\} = \frac{x^2}{2} - \frac{x^4}{4} = \frac{x^2}{2}\left(1 - \frac{x^2}{2}\right).$$

Más adelante definiremos una función utilizando el límite de una sucesión de funciones generadoras de momentos logarítmicas, y será la transformada de Fenchel-Legendre de esa función la que utilizaremos como buena función de grandes desviaciones.

El siguiente paso es enunciar un teorema que utilizaremos en la demostración del teorema de Gärtner-Ellis, el cual hace uso de dos definiciones: función esencialmente suave e hiperplano expuesto. A continuación introducimos estas dos definiciones, cada una con su respectivo ejemplo.

Definición 20. *Sean $f : \mathbb{R}^d \to (-\infty, \infty]$ una función convexa*[1] *y $D_f = \{x \in \mathbb{R}^d \mid f(x) < \infty\}$ su dominio esencial. Decimos que f es* ***esencialmente suave*** *si*

(i) *$int(D_f)$ es no vacío,*

(ii) *f es derivable en $int(D_f)$ y*

(iii) *si $\{\lambda_n\}$ es una sucesión en $int(D_f)$ que converge a un punto frontera de $int(D_f)$, entonces $\lim_{n\to\infty} |\nabla f(\lambda_n)| = \infty$.*[2]

Ejemplo 8. Sea $f : \mathbb{R} \to (-\infty, \infty]$ la función $f(x) = x^2$, función que es convexa y $D_f = \{x \in \mathbb{R}^d \mid f(x) < \infty\} = \mathbb{R}$. Las condiciones (i) y (ii) de la definición se cumplen porque $int(D_f) = int(\mathbb{R}) = \mathbb{R}$ y f es derivable en ese conjunto; por otro lado, $int(D_f)$ no tiene puntos frontera, por lo que la condición (iii) también se cumple. Esto prueba que la función $f(x) = x^2$ es esencialmente suave.

Definición 21. *Sean X una variable aleatoria d-dimensional, Λ su función generadora de momentos logarítmica, Λ^* la transformada de Fenchel-Legendre de Λ y $x \in \mathbb{R}^d$. Decimos que x es un* ***punto expuesto*** *de Λ^* si para alguna $\lambda \in \mathbb{R}^d$ y para toda $y \in \mathbb{R}^d$ distinta de x se cumple que*

$$\langle \lambda, y\rangle - \Lambda^*(y) < \langle \lambda, x\rangle - \Lambda^*(x). \tag{3.1}$$

En este caso, a λ se le llama un ***hiperplano expuesto*** *de Λ^*.*

[1] Dados X un conjunto convexo en un espacio vectorial y $f : X \to \mathbb{R}$ una función, f es una **función convexa** si para cualesquiera $x_1, x_2 \in X$ y $t \in [0,1]$ se tiene que

$$f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2).$$

[2] $\nabla f(\lambda_n) = \left(\frac{\partial f(\lambda_n)}{\partial x_1}, \ldots, \frac{\partial f(\lambda_n)}{\partial x_d}\right)$.

Ejemplo 9. Consideremos una variable aleatoria Bernoulli de parámetro $p = \frac{1}{2}$. Por el ejemplo 6 sabemos que

$$\begin{aligned}\Lambda(\lambda) &= log\big(p\ [e^{\lambda} - 1] + 1\big) \\ &= log\Big(\frac{1}{2}\ [e^{\lambda} - 1] + 1\Big),\end{aligned}$$

y utilizando un procedimiento similar al procedimiento del ejemplo 7 se puede ver que

$$\Lambda^*(x) = \begin{cases} xlog\big(2x\big) + log\big(2(1-x)\big)[1-x] & \text{si } x \in (0,1) \\ \infty & \text{en otro caso} \end{cases}$$

Tomando $x = \frac{1}{2}$, podemos ver que para $\lambda = 0$ la desigualdad (3.1) se reduce a

$$\begin{aligned}-\Lambda^*(y) < -\Lambda^*\Big(\frac{1}{2}\Big) &= \Big(\frac{1}{2}\Big)log\left(2\Big(\frac{1}{2}\Big)\right) + log\left(2\Big(1-\frac{1}{2}\Big)\right)\left[1-\Big(\frac{1}{2}\Big)\right] \\ &= \Big(\frac{1}{2}\Big)log(1) - log(1)\left[1-\Big(\frac{1}{2}\Big)\right] \\ &= 0,\end{aligned}$$

desigualdad que se cumple para cualquier $y \in \mathbb{R}$ distinta de x, pues

- si $y \in (-\infty, 0] \cup [1, \infty)$ entonces $-\Lambda^*(y) = -\infty$,
- y si $y \in (0,1)$, definiendo la función $g : (0,1) \to \mathbb{R}$ como

  $$g(h) = -\Lambda^*(h) = -\Big[hlog\big(2h\big) + log\big(2(1-h)\big)[h-1]\Big],$$

  se tiene que $g'(h) = log(2-2h) - log(2h)$ es igual a cero solo cuando $h = \frac{1}{2}$ y

  $$g''\big(\tfrac{1}{2}\big) = \frac{1}{(\frac{1}{2})^2 - \frac{1}{2}} = -4 < 0,$$

  por lo que la función g alcanza su máximo en $h = \frac{1}{2}$; y como y es distinta de $x = \frac{1}{2}$ entonces

  $$-\Lambda^*(y) = g(y) < g(x) = g\Big(\tfrac{1}{2}\Big) = -\Lambda^*\Big(\tfrac{1}{2}\Big) = 0.$$

Esto prueba que $x = \frac{1}{2}$ es un punto expuesto de Λ^* y $\lambda = 0$ es un hiperplano expuesto de Λ^*.

Haciendo uso de las definiciones anteriores, procedemos a enunciar el teorema de Gärtner-Ellis:

Teorema 7 (Teorema de Gärtner-Ellis)**.** *Sean $\{S_n\}$ una sucesión de vectores aleatorios en $\mathbb{R}^d$, donde cada S_n tiene ley μ_n y función generadora de momentos logarítmica $\Lambda_n(\lambda) = log\ \mathbb{E}[e^{\langle \lambda, S_n \rangle}]$. Supongamos que* *(i)* $\lim_{n\to\infty} \frac{1}{n}\Lambda_n(n\lambda)$ *existe en* $\mathbb{R}\cup\{-\infty,\infty\}$ *y,*

(ii) definiendo la función $\Lambda(\lambda) = \lim_{n\to\infty} \frac{1}{n}\Lambda_n(n\lambda)$*, se tiene que* $0 \in int(D_\Lambda)$ *(recordemos que* $D_\Lambda = \{\lambda \in \mathbb{R}^d \mid \Lambda(\lambda) < \infty\}$ *es el dominio esencial de* Λ*).*

Entonces

(I) $\overline{\lim_{n\to\infty}} \frac{1}{n} log\left(\mu_n(F)\right) \leq -\inf_{x\in F} \Lambda^*(x)$ *para cualquier conjunto cerrado F y*

(II) $-\inf_{x\in G\cap\mathscr{F}} \Lambda^*(x) \leq \underline{\lim_{n\to\infty}} \frac{1}{n} log\left(\mu_n(G)\right)$ *para cualquier conjunto abierto G, donde* $\mathscr{F}$ *es el conjunto de puntos expuestos de* Λ^* *cuyo hiperplano expuesto pertenece a* $int(D_\Lambda)$.

(III) Si, además de los supuestos (i) y (ii), se tiene que

(a) Λ *es finita en todo* $\mathbb{R}^d$*,*

(b) Λ *es esencialmente suave y*

(c) Λ *es s.c.i.;*

entonces el P.G.D. se cumple para la buena función de grandes desviaciones Λ^**, donde* Λ^* *es la transformada de Fenchel-Legendre de la función* Λ*.*

Este teorema es una versión más general del teorema de Cramer, el cual supone también que las variables son i.i.d. (por lo que (i) se cumple trivialmente). Nos interesa el teorema de Gärtner-Ellis porque trabajaremos con cadenas de Markov y en ese caso no necesariamente contamos con el supuesto de variables aleatorias i.i.d..

La demostración del teorema de Gärtner-Ellis se descompone en dos partes: la prueba de (I) y (II) está contenida en [5, p. 48-50], mientras que la demostración de (III) es un ejercicio en [5, p. 52], el cual haremos a continuación en nueve pasos.

Demostración.
Sea $\delta > 0$. Para cada $n \in \mathbb{N}$ definimos $S_{n,\delta} = S_n + \sqrt{\frac{\delta}{n}}V$, donde V es una variable aleatoria normal multivariada estándar independiente de S_n. Denotaremos como $\Lambda_{n,\delta}$ a la función generadora de momentos logarítmica de $S_{n,\delta}$.

Los pasos a seguir son los siguientes:

(1) La sucesión $\{S_{n,\delta}\}$ cumple con los supuestos (i) y (ii) del teorema. Además, dada $\lambda \in \mathbb{R}^d$ y denotando $\Lambda_\delta(\lambda) = \lim_{n\to\infty} \frac{1}{n}\Lambda_{n,\delta}(n\lambda)$, se tiene que

$$\Lambda_\delta(\lambda) = \Lambda(\lambda) + \frac{\delta|\lambda|^2}{2}.$$

(2) Para toda $x \in \mathbb{R}^d$ se tiene que $\Lambda_\delta^*(x) \leq \Lambda^*(x)$.

(3) Utilizando el inciso (I) del teorema para $F = \mathbb{R}^d$ se tiene que D_{Λ^*} es no vacío. Además, cada $x \in \mathbb{R}^d$ es un punto expuesto de Λ_δ^*.

(4) Utilizando el inciso (II) del teorema y los primeros tres pasos de esta demostración, probar que para cualquier $x \in \mathbb{R}^d$ y cualquier $\varepsilon > 0$ se tiene que

$$-\Lambda^*(x) \leq \underline{\lim}_{n\to\infty} \frac{1}{n} log\, P\left(S_{n,\delta} \in B\left(x, \frac{\varepsilon}{2}\right)\right). \tag{3.2}$$

(5) Probar que

$$P\left(\sqrt{\frac{1}{n}}|V| \geq \frac{\varepsilon}{2}\right) \leq 2e^{\frac{-1}{\delta}\left(\frac{\varepsilon^2}{8}n\right)} \tag{3.3}$$

para toda $\varepsilon > 0$.

(6) Demostrar que

$$P\left(S_n + \sqrt{\frac{1}{n}}V \in B\left(x, \frac{\varepsilon}{2}\right)\right) - P\left(\sqrt{\frac{1}{n}}|V| \geq \frac{\varepsilon}{2}\right) \leq P\left(S_n \in B(x, \varepsilon)\right) \tag{3.4}$$

para toda $x \in \mathbb{R}^d$ y $\varepsilon > 0$.

(7) Utilizando (3.4), (3.3) y (3.2), probar que

$$-\Lambda^*(x) \leq \underline{\lim}_{n\to\infty} \frac{1}{n} log P\left(S_n \in B(x, \varepsilon)\right) \tag{3.5}$$

para toda $x \in \mathbb{R}^d$ y $\varepsilon > 0$.

(8) Usar (3.5) para probar que la cota inferior del P.G.D. se cumple para las leyes $\{\mu_n\}$ correspondientes a la sucesión $\{S_n\}$.

(9) Utilizar el inciso (I) del teorema para concluir que $\{\mu_n\}$ satisface el P.G.D. con la buena función de grandes desviaciones Λ^*.

A continuación realizaremos la prueba de cada uno de estos pasos:

(1) Sea $\lambda \in \mathbb{R}^d$. Comencemos observando que

$$\begin{aligned}
\frac{1}{n}\Lambda_{n,\delta}(n\lambda) &= \frac{1}{n} log \mathbb{E}\left[e^{\langle n\lambda, S_{n,\delta}\rangle}\right] \\
&= \frac{1}{n} log \mathbb{E}\left[e^{\left\langle n\lambda, S_n + \sqrt{\frac{\delta}{n}}V\right\rangle}\right] \\
&= \frac{1}{n} log \mathbb{E}\left[e^{\langle n\lambda, S_n\rangle + \left\langle n\lambda, \sqrt{\frac{\delta}{n}}V\right\rangle}\right] \\
&= \frac{1}{n} log \mathbb{E}\left[e^{\langle n\lambda, S_n\rangle} e^{\left\langle n\lambda, \sqrt{\frac{\delta}{n}}V\right\rangle}\right].
\end{aligned}$$

Como V es independiente de S_n, de la igualdad anterior se desprende que

$$\begin{aligned}\frac{1}{n}\Lambda_{n,\delta}(n\lambda) &= \frac{1}{n}log\left(\mathbb{E}\left[e^{\langle n\lambda,S_n\rangle}\right]\mathbb{E}\left[e^{\left\langle n\lambda,\sqrt{\frac{\delta}{n}}V\right\rangle}\right]\right)\\ &= \frac{1}{n}log\left(\mathbb{E}\left[e^{\langle n\lambda,S_n\rangle}\right]\right)+\frac{1}{n}log\left(\mathbb{E}\left[e^{\left\langle n\lambda,\sqrt{\frac{\delta}{n}}V\right\rangle}\right]\right)\\ &= \frac{1}{n}log\left(\mathbb{E}\left[e^{\langle n\lambda,S_n\rangle}\right]\right)+\frac{1}{n}log\left(\mathbb{E}\left[e^{\left\langle \sqrt{\frac{\delta}{n}}n\lambda,V\right\rangle}\right]\right)\\ &= \frac{1}{n}log\left(\mathbb{E}\left[e^{\langle n\lambda,S_n\rangle}\right]\right)+\frac{1}{n}log\left(\mathbb{E}\left[e^{\langle \sqrt{n\delta}\lambda,V\rangle}\right]\right)\\ &= \frac{1}{n}log\left(\mathbb{E}\left[e^{\langle n\lambda,S_n\rangle}\right]\right)+\frac{1}{n}log\left(M_V(\sqrt{n\delta}\lambda)\right),\end{aligned}$$

donde M_V es la función generadora de momentos de V.

Recordemos que $\Lambda_n(n\lambda) = log\mathbb{E}\left[e^{\langle n\lambda,S_n\rangle}\right]$ y que $M_V(x) = e^{\frac{|x|^2}{2}}$[3], por lo que

$$\begin{aligned}\frac{1}{n}\Lambda_{n,\delta}(n\lambda) &= \frac{1}{n}\Lambda_n(n\lambda)+\frac{1}{n}log\left(e^{\frac{n\delta|\lambda|^2}{2}}\right)\\ &= \frac{1}{n}\Lambda_n(n\lambda)+\frac{1}{n}\frac{n\delta|\lambda|^2}{2}\\ &= \frac{1}{n}\Lambda_n(n\lambda)+\frac{\delta|\lambda|^2}{2}.\end{aligned}$$

Tomando el límite cuando n tiende a infinito,

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\Lambda_{n,\delta}(n\lambda) &= \lim_{n\to\infty}\frac{1}{n}\Lambda_n(n\lambda)+\lim_{n\to\infty}\frac{\delta|\lambda|^2}{2}\\ &= \Lambda(\lambda)+\frac{\delta|\lambda|^2}{2}.\end{aligned}$$

Esto demuestra que $\lim_{n\to\infty}\frac{1}{n}\Lambda_{n,\delta}(n\lambda)$ existe, por lo que el supuesto (i) del teorema se cumple para la sucesión $S_{n,\delta}$. Además, denotando $\Lambda_\delta(\lambda) = \lim_{n\to\infty}\frac{1}{n}\Lambda_{n,\delta}(n\lambda)$, se tiene que $\Lambda_\delta(\lambda) = \Lambda(\lambda)+\frac{\delta|\lambda|^2}{2}$.

[3] Si V es una variable aleatoria normal multivariada con parámetros μ y matriz de covarianza Σ, su función generadora de momentos es

$$M_V(x) = e^{\mu' x+\frac{x'\Sigma x}{2}}. \tag{3.6}$$

Cuando V es estándar, $\mu = 0$ y Σ es la matriz identidad, por lo que (3.6) queda reducida a

$$M_V(x) = e^{\frac{|x|^2}{2}}.$$

Por otro lado, de (b) se desprende que $\Lambda(\lambda) < \infty$, lo que implica que $\Lambda_\delta(\lambda) = \Lambda(\lambda) + \frac{\delta|\lambda|^2}{2} < \infty$. Esto demuestra que Λ_δ es finita en todo $\mathbb{R}^d$, por lo que

$$D_{\Lambda_\delta} = \{\lambda \in \mathbb{R}^d \mid \Lambda_\delta(\lambda) < \infty\} = \mathbb{R}^d.$$

En particular, $0 \in \mathbb{R}^d = int(\mathbb{R}^d) = int(D_{\Lambda_\delta})$ y por lo tanto el supuesto (ii) del teorema se cumple para la sucesión $S_{n,\delta}$.

(2) Sea $x \in \mathbb{R}^d$. Utilizando la igualdad para Λ_δ obtenida en el paso (1) de esta demostración,

$$\begin{aligned}\Lambda_\delta^*(x) &= \sup_{\lambda\in\mathbb{R}^d} \{\langle\lambda,x\rangle - \Lambda_\delta(\lambda)\} \\ &= \sup_{\lambda\in\mathbb{R}^d} \left\{\langle\lambda,x\rangle - \left(\Lambda(\lambda) + \frac{\delta|\lambda|^2}{2}\right)\right\},\end{aligned}$$

por lo que

$$\Lambda_\delta^*(x) = \sup_{\lambda\in\mathbb{R}^d} \left\{\left(\langle\lambda,x\rangle - \Lambda(\lambda)\right) - \frac{\delta|\lambda|^2}{2}\right\} \tag{3.7}$$

Dada $\lambda \in \mathbb{R}^d$, como $-\frac{\delta|\lambda|^2}{2} \leq 0$ entonces

$$\left(\langle\lambda,x\rangle - \Lambda(\lambda)\right) - \tfrac{\delta|\lambda|^2}{2} \leq \langle\lambda,x\rangle - \Lambda(\lambda).$$

Tomando el supremo en ambos lados de la desigualdad,

$$\sup_{\lambda\in\mathbb{R}^d} \left\{\left(\langle\lambda,x\rangle - \Lambda(\lambda)\right) - \tfrac{\delta|\lambda|^2}{2}\right\} \leq \sup_{\lambda\in\mathbb{R}^d} \{\langle\lambda,x\rangle - \Lambda(\lambda)\} = \Lambda^*(x).$$

Retomando (3.7), por la desigualdad anterior se sigue que

$$\Lambda_\delta^*(x) = \sup_{\lambda\in\mathbb{R}^d} \left\{\left(\langle\lambda,x\rangle - \Lambda(\lambda)\right) - \tfrac{\delta|\lambda|^2}{2}\right\} \leq \Lambda^*(x).$$

(3) Ya que la sucesión $\{S_n\}$ cumple con los supuestos (i) y (ii), por el inciso (I) se tiene que

$$\overline{\lim_{n\to\infty}} \frac{1}{n} log\left(\mu_n(\mathbb{R}^d)\right) \leq - \inf_{x\in\mathbb{R}^d} \Lambda^*(x). \tag{3.8}$$

Podemos observar que

$$\begin{aligned}\overline{\lim_{n\to\infty}} \frac{1}{n} log\left(\mu_n(\mathbb{R}^d)\right) &= \overline{\lim_{n\to\infty}} \frac{1}{n} log(1) \\ &= \overline{\lim_{n\to\infty}} \frac{1}{n} \cdot 0 \\ &= 0.\end{aligned}$$

Sustituyendo este valor en (3.8),

$$0 \leq - \inf_{x \in \mathbb{R}^d} \Lambda^*(x),$$

y al multiplicar por -1 se sigue que

$$0 \geq \inf_{x \in \mathbb{R}^d} \Lambda^*(x). \tag{3.9}$$

Supongamos que $D_{\Lambda^*} = \{x \in \mathbb{R}^d \mid \Lambda^*(x) < \infty\}$ es vacío, entonces $\Lambda^*(x) = \infty$ para cualquier $x \in \mathbb{R}^d$ y por lo tanto $\inf_{x \in \mathbb{R}^d} \Lambda^*(x) = \infty$, lo cual contradice (3.9). Esto demuestra que D_{Λ^*} no es vacío.

(4) Sean $x \in \mathbb{R}^d$ y $\varepsilon > 0$.

Ya verificamos que la sucesión $S_{n,\delta}$ cumple las condiciones (i) y (ii), por el inciso (II) se tiene que

$$-\inf_{z \in B(x,\frac{\varepsilon}{2}) \cap \mathscr{F}} \Lambda^*(z) \leq \varliminf_{n \to \infty} \frac{1}{n} logP\left(S_{n,\delta} \in \left(B(x, \frac{\varepsilon}{2})\right)\right), \tag{3.10}$$

donde $\mathscr{F}$ es el conjunto de puntos expuestos de Λ^* cuyo hiperplano expuesto pertenece a $int(D_{\Lambda_\delta})$. Del paso anterior se sigue que $\mathscr{F} = \mathbb{R}^d$, por lo que (3.10) se simplifica a

$$-\inf_{z \in B(x,\frac{\varepsilon}{2})} \Lambda^*(z) \leq \varliminf_{n \to \infty} \frac{1}{n} logP\left(S_{n,\delta} \in \left(B(x, \frac{\varepsilon}{2})\right)\right). \tag{3.11}$$

Como $\Lambda^*(x) \geq \inf_{z \in B(x,\frac{\varepsilon}{2})} \Lambda^*(z)$, entonces $-\Lambda^*(x) \leq -\inf_{z \in B(x,\frac{\varepsilon}{2})} \Lambda^*(z)$. Aplicando esto a (3.11), se sigue que

$$-\Lambda^*(x) \leq \varliminf_{n \to \infty} \frac{1}{n} logP\left(S_{n,\delta} \in \left(B(x, \frac{\varepsilon}{2})\right)\right). \tag{3.12}$$

(5) Sea $\varepsilon > 0$.

Recordemos que cualquier $x \in \mathbb{R}$ se tiene que

$$P(|V| \geq x) \leq 2e^{-\frac{x^2}{2}}.$$[4]

Haciendo $x = \frac{\varepsilon}{2}\sqrt{\frac{n}{\delta}}$,

$$P\left(|V| \geq \frac{\varepsilon}{2}\sqrt{\frac{n}{\delta}}\right) \leq 2e^{\frac{-\left(\frac{\varepsilon}{2}\sqrt{\frac{n}{\delta}}\right)^2}{2}}. \tag{3.13}$$

Para la parte izquierda de (3.13) podemos notar que

$$P\left(|V| \geq \frac{\varepsilon}{2}\sqrt{\frac{n}{\delta}}\right) = P\left(\sqrt{\frac{\delta}{n}}|V| \geq \frac{\varepsilon}{2}\right), \tag{3.14}$$

[4]Resultado contenido en [1, p. 455].

mientras que para la parte derecha de (3.13) se tiene que

$$\begin{aligned} 2e^{\frac{-\left(\frac{\varepsilon}{2}\sqrt{\frac{n}{\delta}}\right)^2}{2}} &= 2e^{\frac{-\frac{\varepsilon^2}{4}\frac{n}{\delta}}{2}} \\ &= 2e^{\frac{-\varepsilon^2}{8\delta}n}. \end{aligned}$$

Sustituyendo estos valores en (3.13), se sigue que

$$P\left(\sqrt{\tfrac{\delta}{n}}|V| \geq \tfrac{\varepsilon}{2}\right) \leq 2e^{\frac{-\varepsilon^2}{8\delta}n}.$$

Esto prueba que

$$P\left(\sqrt{\tfrac{1}{n}}|V| \geq \tfrac{\varepsilon}{2}\right) \leq 2e^{\frac{-1}{\delta}\left(\frac{\varepsilon^2}{8}n\right)}.$$

(6) Sean $x \in \mathbb{R}^d$ y $\varepsilon > 0$. Para demostrar este inciso primero demostraremos que

$$P\left(\left\{S_n + \sqrt{\tfrac{1}{n}}V \in B\left(x, \tfrac{\varepsilon}{2}\right)\right\} \cap \left\{\sqrt{\tfrac{1}{n}}|V| < \tfrac{\varepsilon}{2}\right\}\right) \leq P(S_n \in B(x, \varepsilon)),$$

y posteriormente veremos por qué esto implica (3.4).

Comencemos notando que

$$\left\{S_n + \sqrt{\tfrac{1}{n}}V \in B\left(x, \tfrac{\varepsilon}{2}\right)\right\} = \left\{\left|S_n + \sqrt{\tfrac{1}{n}}V - x\right| < \tfrac{\varepsilon}{2}\right\}.$$

Si $\left|S_n + \sqrt{\tfrac{1}{n}}V - x\right| < \tfrac{\varepsilon}{2}$ y $\sqrt{\tfrac{1}{n}}|V| < \tfrac{\varepsilon}{2}$, sumando estas desigualdades se sigue que

$$\begin{aligned} \left|S_n + \sqrt{\frac{1}{n}}V - x\right| + \left|-\sqrt{\frac{1}{n}}V\right| &= \left|S_n + \sqrt{\frac{1}{n}}V - x\right| + \left|\sqrt{\frac{1}{n}}V\right| \\ &= \left|S_n + \sqrt{\frac{1}{n}}V - x\right| + \sqrt{\frac{1}{n}}|V| < \varepsilon, \end{aligned}$$

y por la desigualdad del triángulo se tiene que

$$|S_n - x| = \left|S_n + \sqrt{\tfrac{1}{n}}V - x + -\sqrt{\tfrac{1}{n}}V\right| \leq \left|S_n + \sqrt{\tfrac{1}{n}}V - x\right| + \left|-\sqrt{\tfrac{1}{n}}V\right|,$$

lo que implica que

$$|S_n - x| < \varepsilon.$$

Esto demuestra que

$$\left\{S_n+\sqrt{\tfrac{1}{n}}V\in B\left(x,\tfrac{\varepsilon}{2}\right)\right\}\cap\left\{\sqrt{\tfrac{1}{n}}|V|<\tfrac{\varepsilon}{2}\right\}\subseteq\{S_n\in B(x,\varepsilon)\},$$

lo que implica que

$$P\left(\left\{S_n+\sqrt{\frac{1}{n}}V\in B\left(x,\frac{\varepsilon}{2}\right)\right\}\cap\left\{\sqrt{\frac{1}{n}}|V|<\frac{\varepsilon}{2}\right\}\right)\leq P(S_n\in B(x,\varepsilon)). \quad (3.15)$$

Ahora procederemos a demostrar (3.4). Definiendo los eventos

$$A=\left\{S_n+\sqrt{\tfrac{1}{n}}V\in B\left(x,\tfrac{\varepsilon}{2}\right)\right\},$$

$$B=\left\{\sqrt{\tfrac{1}{n}}|V|<\tfrac{\varepsilon}{2}\right\} \text{ y}$$

$$C=\{S_n\in B(x,\varepsilon)\},$$

podemos observar que (3.15) queda expresado como

$$P(A\cap B)\leq P(C). \quad (3.16)$$

Por otro lado, considerando la resta de $P(A)$ menos $P(B^C)$ se puede ver que

$$\begin{aligned} P(A)-P(B^C) &= P(A)-[1-P(B)] \\ &= P(A)+P(B)-1. \end{aligned}$$

Como $P(A\cup B)\leq 1$, entonces $-1\leq -P(A\cup B)$ y de lo anterior se sigue que

$$P(A)-P(B^C)\leq P(A)+P(B)-P(A\cup B)=P(A\cap B).$$

Retomando (3.16),

$$P(A)-P(B^C)\leq P(A\cap B)\leq P(C),$$

en particular,

$$P(A)-P(B^C)\leq P(C).$$

Retomando las definiciones de los eventos A, B y C, la última desigualdad queda expresada como

$$P\left(S_n+\sqrt{\tfrac{1}{n}}V\in B\left(x,\tfrac{\varepsilon}{2}\right)\right)-P\left(\sqrt{\tfrac{1}{n}}|V|\geq\tfrac{\varepsilon}{2}\right)\leq P\left(S_n\in B(x,\varepsilon)\right).$$

Esto concluye la demostración de (3.4).

(7) Sean $x \in \mathbb{R}^d$ y $\varepsilon > 0$. Recordemos que en el inciso (6) probamos que

$$P\left(S_n + \sqrt{\tfrac{1}{n}}V \in B\left(x, \tfrac{\varepsilon}{2}\right)\right) - P\left(\sqrt{\tfrac{1}{n}}|V| \geq \tfrac{\varepsilon}{2}\right) \leq P\left(S_n \in B(x,\varepsilon)\right).$$

De esta desigualdad se sigue que

$$P\left(S_n + \sqrt{\tfrac{1}{n}}V \in B\left(x, \tfrac{\varepsilon}{2}\right)\right) \leq P\left(S_n \in B(x,\varepsilon)\right) + P\left(\sqrt{\tfrac{1}{n}}|V| \geq \tfrac{\varepsilon}{2}\right),$$

y retomando (3.3),

$$P\left(S_n + \sqrt{\tfrac{1}{n}}V \in B\left(x, \tfrac{\varepsilon}{2}\right)\right) \leq P\left(S_n \in B(x,\varepsilon)\right) + 2e^{\frac{-1}{\delta}\left(\frac{\varepsilon^2}{8}n\right)}.$$

Aplicando el logaritmo a la desigualdad anterior, multiplicando por $\frac{1}{n}$ y luego considerando el límite inferior cuando n tiende a infinito tenemos que

$$\underline{\lim}_{n\to\infty} \tfrac{1}{n} log P\left(S_n + \sqrt{\tfrac{1}{n}}V \in B\left(x, \tfrac{\varepsilon}{2}\right)\right) \leq$$
$$\underline{\lim}_{n\to\infty} \tfrac{1}{n} log \left[P\left(S_n \in B(x,\varepsilon)\right) + 2e^{\frac{-1}{\delta}\left(\frac{\varepsilon^2}{8}n\right)}\right],$$

mientras que en (3.2) probamos que

$$-\Lambda^*(x) \leq \underline{\lim}_{n\to\infty} \tfrac{1}{n} log P\left(S_n + \sqrt{\tfrac{1}{n}}V \in B\left(x, \tfrac{\varepsilon}{2}\right)\right).$$

De estas dos desigualdades se desprende que

$$-\Lambda^*(x) \leq \underline{\lim}_{n\to\infty} \frac{1}{n} log \left[P\left(S_n \in B(x,\varepsilon)\right) + 2e^{\frac{-1}{\delta}\left(\frac{\varepsilon^2}{8}n\right)}\right] \text{ para toda } \delta > 0 \quad (3.17)$$

(esto porque la δ que tomamos al inicio del ejercicio era arbitraria).

Ahora probaremos que

$$-\Lambda^*(x) \leq \underline{\lim}_{n\to\infty} \tfrac{1}{n} log P\left(S_n \in B(x,\varepsilon)\right)$$

por reducción al absurdo: supondremos lo contrario y llegaremos a una contradicción con (3.17).

Supongamos que

$$-\Lambda^*(x) > \underline{\lim}_{n\to\infty} \frac{1}{n} log P\left(S_n \in B(x,\varepsilon)\right). \quad (3.18)$$

Tomando

$$M = \frac{-\Lambda^*(x) - \lim\limits_{n\to\infty} \frac{1}{n} log P(S_n \in B(x,\varepsilon))}{2} > 0$$

se tiene que $e^M > 1$ y entonces $\left(e^M - 1\right) > 0$. Definimos

$$c = \left(e^M - 1\right) P\left(S_n \in B(x,\varepsilon)\right) > 0. \tag{3.19}$$

Observemos que $\lim\limits_{\delta\to 0^+} \frac{-1}{\delta} = -\infty$ y $\frac{\varepsilon^2}{8} n > 0$, por lo que

$$\lim_{\delta\to 0^+} \frac{-1}{\delta}\left(\frac{\varepsilon^2}{8} n\right) = -\infty,$$

lo que implica que

$$\lim_{\delta\to 0^+} e^{\frac{-1}{\delta}\left(\frac{\varepsilon^2}{8} n\right)} = 0.$$

Por este último límite sabemos que para cualquier constante positiva existe una $\delta > 0$ tal que $e^{\frac{-1}{\delta}\left(\frac{\varepsilon^2}{8} n\right)}$ es estrictamente menor a esa constante. En particular, para la c definida en (3.19) existe una $\delta_0 > 0$ para la cual

$$e^{\frac{-1}{\delta_0}\left(\frac{\varepsilon^2}{8} n\right)} < \frac{c}{2}.$$

Esta última desigualdad implica que

$$P\left(S_n \in B(x,\varepsilon)\right) + 2e^{\frac{-1}{\delta_0}\left(\frac{\varepsilon^2}{8} n\right)} < P\left(S_n \in B(x,\varepsilon)\right) + c,$$

y por lo tanto

$$log\left[P\left(S_n \in B(x,\varepsilon)\right) + 2e^{\frac{-1}{\delta_0}\left(\frac{\varepsilon^2}{8} n\right)}\right] < log\left[P\left(S_n \in B(x,\varepsilon)\right) + c\right]. \tag{3.20}$$

Notemos que

$$\begin{aligned}
log\left[P\left(S_n \in B(x,\varepsilon)\right) + c\right] &= log\left[P\left(S_n \in B(x,\varepsilon)\right) + \left(e^M - 1\right) P\left(S_n \in B(x,\varepsilon)\right)\right] \\
&= log\left[P\left(S_n \in B(x,\varepsilon)\right)\left(1 + e^M - 1\right)\right] \\
&= log\left[P\left(S_n \in B(x,\varepsilon)\right) e^M\right] \\
&= log P\left(S_n \in B(x,\varepsilon)\right) + log\left(e^M\right) \\
&= log P\left(S_n \in B(x,\varepsilon)\right) + M,
\end{aligned}$$

de manera que (3.20) puede ser simplificada a

$$log\left[P\left(S_n \in B(x,\varepsilon)\right) + 2e^{\frac{-1}{\delta_0}\left(\frac{\varepsilon^2}{8} n\right)}\right] < log P\left(S_n \in B(x,\varepsilon)\right) + M.$$

Multiplicando esta desigualdad por $\frac{1}{n}$ tenemos que

$$\tfrac{1}{n}log\left[P\left(S_n\in B(x,\varepsilon)\right)+2e^{\frac{-1}{\delta_0}\left(\frac{\varepsilon^2}{8}n\right)}\right]<\tfrac{1}{n}logP\left(S_n\in B(x,\varepsilon)\right)+\tfrac{M}{n},$$

y considerando que $\frac{M}{n}\leq M$ entonces

$$\frac{1}{n}log\left[P\left(S_n\in B(x,\varepsilon)\right)+2e^{\frac{-1}{\delta_0}\left(\frac{\varepsilon^2}{8}n\right)}\right]<\frac{1}{n}logP\left(S_n\in B(x,\varepsilon)\right)+M. \tag{3.21}$$

Retomando la definición de M en la parte derecha de (3.21),

$$\begin{aligned}\frac{1}{n}logP\left(S_n\in B(x,\varepsilon)\right)+M &= \frac{1}{n}logP\left(S_n\in B(x,\varepsilon)\right)+\frac{-\Lambda^*(x)-\varliminf_{n\to\infty}\frac{1}{n}logP\left(S_n\in B(x,\varepsilon)\right)}{2}\\ &= \frac{2\frac{1}{n}logP\left(S_n\in B(x,\varepsilon)\right)}{2}+\frac{-\Lambda^*(x)-\varliminf_{n\to\infty}\frac{1}{n}logP\left(S_n\in B(x,\varepsilon)\right)}{2}\\ &= \frac{-\Lambda^*(x)+\varliminf_{n\to\infty}\frac{1}{n}logP\left(S_n\in B(x,\varepsilon)\right)}{2},\end{aligned}$$

y por (3.18) se tiene que

$$\tfrac{-\Lambda^*(x)+\varliminf_{n\to\infty}\frac{1}{n}logP(S_n\in B(x,\varepsilon))}{2}<\tfrac{-\Lambda^*(x)-\Lambda^*(x)}{2}=-\Lambda^*(x).$$

Esto implica que

$$\tfrac{1}{n}logP\left(S_n\in B(x,\varepsilon)\right)+M<-\Lambda^*(x),$$

y por (3.21) se sigue que

$$\tfrac{1}{n}log\left[P\left(S_n\in B(x,\varepsilon)\right)+2e^{\frac{-1}{\delta_0}\left(\frac{\varepsilon^2}{8}n\right)}\right]<\tfrac{1}{n}logP\left(S_n\in B(x,\varepsilon)\right)+M<-\Lambda^*(x).$$

Tomando el límite inferior cuando n tiende a infinito en ambos extremos de esta desigualdad tenemos que

$$\varliminf_{n\to\infty}\tfrac{1}{n}log\left[P\left(S_n\in B(x,\varepsilon)\right)+2e^{\frac{-1}{\delta_0}\left(\frac{\varepsilon^2}{8}n\right)}\right]<\varliminf_{n\to\infty}-\Lambda^*(x)=-\Lambda^*(x).$$

Hemos encontrado entonces una $\delta_0>0$ para la cual

$$\varliminf_{n\to\infty}\tfrac{1}{n}log\left[P\left(S_n\in B(x,\varepsilon)\right)+2e^{\frac{-1}{\delta_0}\left(\frac{\varepsilon^2}{8}n\right)}\right]<-\Lambda^*(x),$$

lo cual contradice (3.17).

Por lo tanto,

$$-\Lambda^*(x) \leq \underline{\lim}_{n\to\infty} \tfrac{1}{n} logP(S_n \in B(x,\varepsilon)).$$

(8) Sea $A \subseteq \mathbb{R}^d$. Recordemos que en este paso debemos probar que se cumple la cota inferior del P.G.D. para la sucesión S_n, es decir, debemos probar que

$$-\inf_{x\in int(A)} \Lambda^*(x) \leq \underline{\lim}_{n\to\infty} \frac{1}{n}\ log\ P(S_n \in A). \tag{3.22}$$

Si $int(A) = \emptyset$, (3.22) se cumple trivialmente. Si $int(A) \neq \emptyset$, por el paso anterior sabemos que

$$-\Lambda^*(x) \leq \underline{\lim}_{n\to\infty} \tfrac{1}{n} logP(S_n \in B(x,\varepsilon))$$

para cualquier $x \in \mathbb{R}^d$ y $\varepsilon > 0$. En particular, tomando $x_0 \in int(A)$ y $\varepsilon_0 > 0$ tal que $B(x_0,\varepsilon_0) \subseteq A$ se tiene que

$$-\Lambda^*(x_0) \leq \underline{\lim}_{n\to\infty} \tfrac{1}{n} logP(S_n \in B(x_0,\varepsilon_0)) \leq \underline{\lim}_{n\to\infty} \tfrac{1}{n} logP(S_n \in A)$$

Esto prueba que, para toda $x \in int(A)$,

$$-\Lambda^*(x) \leq \underline{\lim}_{n\to\infty} \tfrac{1}{n} logP(S_n \in A),$$

por lo que

$$\Lambda^*(x) \geq -\underline{\lim}_{n\to\infty} \tfrac{1}{n} logP(S_n \in A),$$

y por lo tanto la parte derecha de esta desigualdad es una cota inferior del conjunto $\{\Lambda^*(x) \mid x \in int(A)\}$. De la definición de ínfimo se sigue que

$$\inf_{x\in int(A)} \Lambda^*(x) \geq -\underline{\lim}_{n\to\infty} \tfrac{1}{n}\ log\ P(S_n \in A),$$

y al multiplicar esta desigualdad por -1 se sigue (3.22).

(9) Sea $A \subseteq \mathbb{R}^d$. Por el inciso (I) tenemos que

$$\overline{\lim}_{n\to\infty} \tfrac{1}{n}\ log\ P(S_n \in A) \leq -\inf_{x\in\bar{A}} \Lambda^*(x),$$

mientras que en el paso anterior probamos que

$$-\inf_{x\in int(A)} \Lambda^*(x) \leq \underline{\lim}_{n\to\infty} \tfrac{1}{n}\ log\ P(S_n \in A).$$

Juntando ambas desigualdades tenemos que

$$-\inf_{x\in int(A)} \Lambda^*(x) \leq \underline{\lim}_{n\to\infty} \tfrac{1}{n}\ log\ P(S_n \in A) \leq \overline{\lim}_{n\to\infty} \tfrac{1}{n}\ log\ P(S_n \in A) \leq -\inf_{x\in\bar{A}} \Lambda^*(x),$$

lo que demuestra que la sucesión $\{S_n\}$ satisface el P.G.D. con la buena función de grandes desviaciones Λ^*.

$\square$

La demostración anterior nos permite comprender un poco más el Principio de Grandes Desviaciones (P.G.D.). Después de añadir una variable aleatoria normal multivariada estándar, del paso (5) se sigue que la sucesión $\{S_n\}$ sólo fue perturbada por muy poco y, además, como la sucesión $\{S_{n,\delta}\}$ cumple con el P.G.D., resulta natural que la sucesión original $\{S_n\}$ también cumpliera con dicho principio. En la teoría de grandes desviaciones existen muchos ejemplos similares donde, para probar que una familia cumple con un P.G.D., se añade otra variable aleatoria que perturba marginalmente al proceso original.

Principio de Grandes Desviaciones para cadenas de Markov regulares

Para concluir este capítulo utilizaremos el teorema de Gärtner-Ellis para demostrar que el P.G.D. se cumple para una funcional de cadenas de Markov regulares. Comenzaremos definiendo y estudiando dicha funcional.

Definición 22. *Sean $P = (p_{ij}) \in M_{N\times N}(\mathbb{R})$ una matriz, E un conjunto de cardinalidad N, $f : E \to \mathbb{R}^d$ una función y $\lambda \in \mathbb{R}^d$. Definimos la matriz $P_\lambda \in M_{N\times N}(\mathbb{R})$ como*

$$P_\lambda = (p_{ij} \cdot e^{\langle \lambda, f(j)\rangle})_{i,j\in E}.$$

Teorema 8. *Sean $\{X_n\}$ una cadena de Markov con espacio de estados finito E y matriz de transición P, $f : E \to \mathbb{R}^d$ una función y $\lambda \in \mathbb{R}^d$. Si la cadena $\{X_n\}$ es regular entonces la matriz P_λ es regular.*

Demostración.
Para simplificar la notación, denotaremos como q_{ij} a la ij-ésima entrada de la matriz P_λ, y para cada $m \in \mathbb{N}$ denotaremos como $q_{ij}^{(m)}$ a la ij-ésima entrada de la matriz $(P_\lambda)^m$.

La demostración de este teorema se dividirá en dos partes: primero probaremos que, para cualquier $m \in \mathbb{N}$, $q_{ij}^{(m)}$ puede ser expresada en términos de la entrada ij-ésima entrada de la matriz P^m (y por lo tanto, de la probabilidad de llegar del estado i al estado j en m pasos). Posteriormente utilizaremos esto para probar que existe una potencia k para la cual $(P_\lambda)^k$ es estrictamente positiva.

Sean $i, j \in E$. Buscamos probar la siguiente propiedad:

$$\textit{Para toda } m \in \mathbb{N} \textit{ existe una } c > 0 \textit{ tal que } q_{ij}^{(m)} = c p_{ij}^{(m)}. \tag{3.23}$$

Realizaremos la prueba de (3.23) por inducción:

- Por definición, $q_{ij}^{(1)} = q_{ij} = p_{ij} \cdot e^{\langle \lambda, f(j)\rangle}$. Haciendo $c = e^{\langle \lambda, f(j)\rangle} > 0$, la propiedad (3.23) queda demostrada para $m = 1$.

- Supongamos que existe una $c > 0$ tal que $q_{ij}^{(\ell)} = cp_{ij}^{(\ell)}$ para alguna $\ell \in \mathbb{N}$; ahora debemos demostrar que existe $c' > 0$ tal que $q_{ij}^{(\ell+1)} = c' p_{ij}^{(\ell+1)}$.

  Recordemos que si $A = (a_{vw})$ y $B = (b_{vw})$ son dos matrices en $M_{N\times N}(\mathbb{R})$, entonces $AB = \left(\sum_{u=1}^{N} a_{vu}b_{uw}\right)$. Considerando que $(P_\lambda)^{\ell+1} = (P_\lambda)^{\ell} \cdot P_\lambda$, entonces

$$\begin{aligned} q_{ij}^{(\ell+1)} &= \sum_{u=1}^{N} q_{iu}^{(\ell)} q_{uj} \\ &= \sum_{u=1}^{N} q_{iu}^{(\ell)} \cdot p_{uj} \cdot e^{\langle\lambda, f(j)\rangle} \\ &= e^{\langle\lambda, f(j)\rangle} \sum_{u=1}^{N} q_{iu}^{(\ell)} p_{uj} \end{aligned}$$

  Como (3.23) se cumple para ℓ, entonces para cada $u \in \{1, \ldots, N\}$ existe $c_u > 0$ tal que $q_{iu}^{(\ell)} = c_u p_{iu}^{(\ell)}$. Sustituyendo este valor en la ecuación anterior,

$$q_{ij}^{(\ell+1)} = e^{\langle\lambda, f(j)\rangle} \sum_{u=1}^{N} c_u p_{iu}^{(\ell)} p_{uj}. \tag{3.24}$$

  Sean $a = \min_{1\le u\le N} c_u$ y $b = \max_{1\le u\le N} c_u$. Podemos notar que

$$\sum_{u=1}^{N} a p_{iu}^{(\ell)} p_{uj} \le \sum_{u=1}^{N} c_u p_{iu}^{(\ell)} p_{uj} \le \sum_{u=1}^{N} b p_{iu}^{(\ell)} p_{uj},$$

  por lo que

$$a \sum_{u=1}^{N} p_{iu}^{(\ell)} p_{uj} \le \sum_{u=1}^{N} c_u p_{iu}^{(\ell)} p_{uj} \le b \sum_{u=1}^{N} p_{iu}^{(\ell)} p_{uj}.$$

  Como $\sum_{u=1}^{N} p_{iu}^{(\ell)} p_{uj} = p_{ij}^{(\ell+1)}$, por la desigualdad anterior se tiene que

$$a p_{ij}^{(\ell+1)} \le \sum_{u=1}^{N} c_u p_{iu}^{(\ell)} p_{uj} \le b p_{ij}^{(\ell+1)}. \tag{3.25}$$

  Definimos $g : [a,b] \to \mathbb{R}$ como $g(x) = x\ p_{ij}^{(\ell+1)}$. La función g es una función continua, y además por (3.25) se tiene que

$$g(a) \le \sum_{u=1}^{N} c_u p_{iu}^{(\ell)} p_{uj} \le g(b).$$

  Por el teorema del valor intermedio, existe una $\xi \in [a,b]$ tal que $g(\xi_i) = \sum_{u=1}^{N} c_u p_{iu}^{(\ell)} p_{uj}$, es decir,

$$\xi p_{ij}^{(\ell+1)} = \sum_{u=1}^{N} c_u p_{iu}^{(\ell)} p_{uj}.$$

Sustituyendo este valor en (3.24), se sigue que

$$p_{ij}^{(\ell+1)} = e^{\langle \lambda, f(j) \rangle} \xi p_{ij}^{(\ell+1)}$$

Notemos que $e^{\langle \lambda, f(j) \rangle} > 0$ y $\xi \geq a = \min_{1 \leq u \leq N} c_u > 0$.

Definiendo $c' = e^{\langle \lambda, f(j) \rangle} \cdot \xi > 0$, se sigue que

$$p_{ij}^{(\ell+1)} = c' p_{ij}^{(\ell+1)}.$$

Esto concluye la demostración de la propiedad (3.23).

Ahora procederemos a demostrar que P_λ es una matriz regular.

Como la cadena es regular, P es regular y por lo tanto existe una $k \in \mathbb{N}$ tal que $P^k > 0$, es decir,

$$p_{ij}^{(k)} > 0 \text{ para cualesquiera } i, j \in E. \tag{3.26}$$

Por la propiedad (3.23), para cada $i, j \in E$ existe $c_{ij} > 0$ tal que $q_{ij}^{(k)} = c_{ij} p_{ij}^{(k)}$, y por (3.26) también se tiene que $p_{ij}^{(k)} > 0$. Esto implica que

$$q_{ij}^{(k)} = c_{ij} p_{ij}^{(k)} > 0,$$

lo que demuestra que cada entrada de la matriz $(P_\lambda)^k$ es estrictamente positiva. Por lo tanto, la matriz P_λ es regular. □

Denotaremos al eigenvalor de Perron-Frobenius de la matriz P_λ como r_λ. Es importante observar que, aunque las matrices P y P_λ son regulares, no necesariamente sucede que sus eigenvalores de Perron-Frobenius son iguales.

En el caso de una cadena de Markov irreducible y aperiódica con matriz de transición P, por el teorema 4 sabemos que P es regular y por lo tanto la matriz P_λ también será regular, lo cual nos permite aplicar el teorema de Perron-Frobenius a la matriz P_λ. Con esto estamos en condiciones de enunciar y probar el principal resultado de este capítulo:

Teorema 9. *Sean $\{X_n\}$ una cadena de Markov irreducible y aperiódica con espacio de estados finito E, P su matriz de transición y $f : E \to \mathbb{R}^d$ una función. Para cada $\lambda \in \mathbb{R}^d$ definimos la matriz*

$$P_\lambda = (p_{ij} \cdot e^{\langle \lambda, f(j) \rangle})_{i,j \in E}$$

y denotamos como r_λ al eigenvalor de Perron-Frobenius de la matriz P_λ.
Posteriormente, para cada $z \in \mathbb{R}^d$ se define

$$I(z) = \sup_{\lambda \in \mathbb{R}^d} \{\langle \lambda, z \rangle - log(r_\lambda)\}.$$

Entonces I es una buena función de grandes desviaciones convexa y la media muestral $S_n = \frac{1}{n}\sum_{i=1}^{n} f(X_i)$ *satisface el P.G.D. con la función I.*

Para demostrar este teorema utilizaremos el inciso (III) del teorema de Gärtner-Ellis, por lo que debemos verificar que todos los supuestos de dicho teorema se cumplen, es decir, debemos probar lo siguiente:

Dada $\Lambda_n(\lambda) = log\ \mathbb{E}[e^{\langle \lambda, S_n \rangle}]$ la función generadora de momentos logarítmica de S_n,

(a) $\lim_{n\to\infty} \frac{1}{n}\Lambda_n(n\lambda)$ existe en $\mathbb{R} \cup \{-\infty\} \cup \{\infty\}$;

y definiendo $\Lambda(\lambda) = \lim_{n\to\infty} \frac{1}{n}\Lambda_n(\lambda)$,

(b) Λ es finita en todo $\mathbb{R}^d$,

(c) Λ es esencialmente suave y

(d) Λ es semicontinua inferiomente.

Además también debemos verificar que

(e) $\Lambda^* = I$ y

(f) Λ^* es convexa.

A continuación realizaremos la demostración de cada uno de estos incisos en el orden mencionado.

Demostración.

(a) Sea $\lambda \in \mathbb{R}^d$. Comencemos observando que

$$\begin{aligned}
\Lambda_n(n\lambda) &= log\ \mathbb{E}[e^{\langle n\lambda, S_n\rangle}] \\
&= log\ \mathbb{E}[e^{\langle \lambda, nS_n\rangle}] \\
&= log\ \mathbb{E}\left[e^{\left\langle \lambda, \sum_{i=1}^{n} f(X_i)\right\rangle}\right] \\
&= log\left[\sum_{x_1,\ldots,x_n}\left(e^{\left\langle \lambda, \sum_{i=1}^{n} f(x_i)\right\rangle}\right)P_\sigma(X_1 = x_1, \ldots, X_n = x_n)\right] \\
&= log\left[\sum_{x_1,\ldots,x_n}\left(\prod_{i=1}^{n} e^{\langle \lambda, f(x_i)\rangle}\right)P_\sigma(X_1 = x_1, \ldots, X_n = x_n)\right] \\
&= log\left[\sum_{x_1,\ldots,x_n}\left(\prod_{i=1}^{n} e^{\langle \lambda, f(x_i)\rangle}\right)p_{\sigma x_1}p_{x_1x_2}\cdots p_{x_{n-1}x_n}\right] \\
&= log\left[\sum_{x_1,\ldots,x_n}\left(p_{\sigma x_1}e^{\langle \lambda, f(x_1)\rangle}\right)\left(p_{x_1x_2}e^{\langle \lambda, f(x_2)\rangle}\right)\ldots\left(p_{x_{n-1}x_n}e^{\langle \lambda, f(x_n)\rangle}\right)\right]. \\
&= log\left[\sum_{x_1,\ldots,x_n} q_{\sigma x_1}q_{x_1x_2}\ldots q_{x_{n-1}x_n}\right],
\end{aligned}$$

donde q_{ij} es la entrada ij-ésima de la matriz P_λ. Retomando la notación $q_{ij}^{(n)}$ para la ij-ésima de la matriz $(P_\lambda)^n$, de la igualdad anterior se sique que

$$\Lambda_n(n\lambda) = log\left[\sum_{x_n} q_{\sigma x_n}^{(n)}\right]. \tag{3.27}$$

Como la cadena es regular, por el teorema 8 se tiene que P_λ es una matriz regular. Sea r_λ el eigenvalor de Perron-Frobenius de la matriz P_λ.

Utilizando el lema 2 para la matriz P_λ y $\phi = \overline{1}' = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$,

$$log(r_\lambda) = \lim_{n\to\infty} \frac{1}{n} log\left[\sum_j q_{\sigma j}^{(n)}\right],$$

y por (3.27) se sigue que

$$log(r_\lambda) = \lim_{n\to\infty} \frac{1}{n}\Lambda_n(n\lambda).$$

Por lo tanto, $\lim\limits_{n\to\infty} \frac{1}{n}\Lambda_n(n\lambda)$ existe y además $\lim\limits_{n\to\infty} \frac{1}{n}\Lambda_n(n\lambda) = log(r_\lambda) \in \mathbb{R}$.

Definimos $\Lambda(\lambda) = \lim\limits_{n\to\infty} \frac{1}{n}\Lambda_n(n\lambda)$. Es importante observar que, por lo anterior,

$$\Lambda(\lambda) = \lim_{n\to\infty} \frac{1}{n}\Lambda_n(n\lambda) = log(r_\lambda). \tag{3.28}$$

(b) Sea $\lambda \in \mathbb{R}^d$. Por el inciso (a) de esta demostración, sabemos que $\Lambda(\lambda) = log(r_\lambda) < \infty$, por lo cual la función Λ es finita en todo $\mathbb{R}^d$.

(c) Para demostrar que la función Λ es esencialmente suave, debemos probar que Λ es convexa y luego que cumple los incisos (i), (ii) y (iii) de la definición 20.

Sean $\lambda_1, \lambda_2 \in \mathbb{R}^d$ y $t \in [0,1]$.

- Si $t = 0$,

$$\begin{aligned} \Lambda(t\lambda_1 + (1-t)\lambda_2) &= \Lambda(0\cdot\lambda_1 + 1\cdot\lambda_2) \\ &= \Lambda(\lambda_2) \\ &= 0\cdot\Lambda(\lambda_1) + (1-0)\Lambda(\lambda_2) \\ &= t\Lambda(\lambda 1) + (1-t)\Lambda(\lambda_2). \end{aligned}$$

- Si $t = 1$,

$$\begin{aligned} \Lambda(t\lambda_1 + (1-t)\lambda_2) &= \Lambda(1\cdot\lambda_1 + 0\cdot\lambda_2) \\ &= \Lambda(\lambda_1) \\ &= 1\cdot\Lambda(\lambda_1) + (1-1)\Lambda(\lambda_2) \\ &= t\Lambda(\lambda_1) + (1-t)\Lambda(\lambda_2). \end{aligned}$$

- Supongamos que $t \in (0,1)$.
  Comencemos observando que

$$\begin{aligned} \Lambda_n(n[t\lambda_1 + (1-t)\lambda_2]) &= log\ \mathbb{E}\left[e^{\langle n[t\lambda_1+(1-t)\lambda_2], S_n\rangle}\right] \\ &= log\ \mathbb{E}\left[e^{\langle t\lambda_1+(1-t)\lambda_2, nS_n\rangle}\right] \\ &= log\ \mathbb{E}\left[\left(e^{\langle\lambda_1, nS_n\rangle}\right)^t \left(e^{\langle\lambda_2, nS_n\rangle}\right)^{(1-t)}\right]. \end{aligned}$$

  Haciendo $X = \left(e^{\langle\lambda_1, nS_n\rangle}\right)^t$ y $Y = \left(e^{\langle\lambda_2, nS_n\rangle}\right)^{(1-t)}$, por la igualdad anterior se tiene que

$$\Lambda_n(n[t\lambda_1 + (1-t)\lambda_2]) = log\ \mathbb{E}[XY]. \tag{3.29}$$

  Por otro lado, como $t \in (0,1)$ entonces $(1-t) \in (0,1)$. Definiendo $p = \frac{1}{t}$ y $q = \frac{1}{1-t}$, tenemos que $p, q \in (0,\infty)$ y $\frac{1}{p} + \frac{1}{q} = t + 1 - t = 1$.

Como X y Y son siempre positivos, aplicando la desigualdad de Hölder[5] a (3.29) se sigue que

$$\Lambda_n(n[t\lambda_1+(1-t)\lambda_2]) = log\ \mathbb{E}[XY] \leq log\left(\mathbb{E}[X^p]^{\frac{1}{p}}\mathbb{E}[Y^q]^{\frac{1}{q}}\right). \quad (3.30)$$

Para el caso de X podemos notar que

$$\begin{aligned}\mathbb{E}[X^p]^{\frac{1}{p}} &= \mathbb{E}\left[\left(\left(e^{\langle\lambda_1,nS_n\rangle}\right)^t\right)^{\frac{1}{t}}\right]^t \\ &= \mathbb{E}\big[e^{\langle\lambda_1,nS_n\rangle}\big]^t,\end{aligned}$$

y similarmente para Y se tiene que

$$\begin{aligned}\mathbb{E}[Y^q]^{\frac{1}{q}} &= \mathbb{E}\left[\left(\left(e^{\langle\lambda_2,nS_n\rangle}\right)^{1-t}\right)^{\frac{1}{1-t}}\right]^{1-t} \\ &= \mathbb{E}\big[e^{\langle\lambda_2,nS_n\rangle}\big]^{1-t}.\end{aligned}$$

Sustituyendo estos valores en (3.30) se sigue que

$$\Lambda_n(n[t\lambda_1+(1-t)\lambda_2]) \leq log\left(\mathbb{E}\big[e^{\langle\lambda_1,nS_n\rangle}\big]^t\ \mathbb{E}\big[e^{\langle\lambda_2,nS_n\rangle}\big]^{(1-t)}\right). \quad (3.31)$$

Como

$$\begin{aligned}log\left(\mathbb{E}\big[e^{\langle\lambda_1,nS_n\rangle}\big]^t\ \mathbb{E}\big[e^{\langle\lambda_2,nS_n\rangle}\big]^{(1-t)}\right) &= t log\mathbb{E}\big[e^{\langle\lambda_1,nS_n\rangle}\big]+(1-t)log\mathbb{E}\big[e^{\langle\lambda_2,nS_n\rangle}\big] \\ &= t log\mathbb{E}\big[e^{\langle n\lambda_1,S_n\rangle}\big]+(1-t)log\mathbb{E}\big[e^{\langle n\lambda_2,S_n\rangle}\big] \\ &= t\Lambda_n(n\lambda_1)+(1-t)\Lambda(n\lambda_2),\end{aligned}$$

por (3.31) tenemos que

$$\Lambda_n(n[t\lambda_1+(1-t)\lambda_2]) \leq t\Lambda_n(n\lambda_1)+(1-t)\Lambda(n\lambda_2).$$

Finalmente, multiplicando ambos lados de esta desigualdad por $\frac{1}{n}$ y tomando el límite cuando n tiende a infinito se sigue que

$$\Lambda(t\lambda_1+(1-t)\lambda_2) \leq t\Lambda(\lambda_1)+(1-t)\Lambda(\lambda_2).$$

En cualquier caso se tiene que $\Lambda(t\lambda_1+(1-t)\lambda_2) \leq t\Lambda(\lambda_1)+(1-t)\Lambda(\lambda_2)$, por lo que la función Λ es convexa.

Ahora probaremos que la función Λ cumple los incisos (i), (ii) y (iii) de la definición 20:

[5]Sean X y Y variables aleatorias d-dimensionales y $p,q\in[1,\infty)$ tales que $\frac{1}{p}+\frac{1}{q}=1$. Entonces

$$\mathbb{E}[|XY|] \leq \mathbb{E}[|X|^p]^{\frac{1}{p}}\mathbb{E}[|Y|^q]^{\frac{1}{q}}.$$

Este resultado está contenido en [3, p. 75].

(i) Por el inciso (b) de este teorema, Λ es finita en todo $\mathbb{R}^d$ y entonces $D_\Lambda = \{\lambda \in \mathbb{R}^d \mid \Lambda(\lambda) < \infty\} = \mathbb{R}^d$. Esto implica que $int(D_\Lambda) = \mathbb{R}^d$ y por lo tanto $int(D_\Lambda)$ es no vacío.

(ii) Sea $\lambda \in \mathbb{R}^d$. Hemos probado que $\Lambda(\lambda) = log(r_\lambda)$, donde r_λ es el eigenvalor de Perron-Frobenius de la matriz P_λ.
Definiendo la función $F:\ \mathbb{R}^d \to \{x \in \mathbb{R} | x > 0\}$ como

$$F(\lambda) = r_\lambda,$$

en el apéndice A probamos que la función F es derivable y al componerla con la función logaritmo sigue siendo derivable, por lo que $\Lambda(\lambda) = log(r_\lambda)$ es derivable.

(iii) $D_\Lambda = \mathbb{R}^d$, por lo que $int(D_\Lambda) = \mathbb{R}^d$. Como $\mathbb{R}^d$ no tiene puntos frontera, este inciso se cumple por vacuidad.

(d) Como la función Λ es derivable, es continua y en particular es semicontinua inferiormente (s.c.i.).

(e) Sea $z \in \mathbb{R}^d$. Utilizando la definición de la transformada de Fenchel-Legendre y el inciso (a) de este teorema, se sigue que

$$\begin{aligned}\Lambda^*(z) &= \sup_{\lambda \in \mathbb{R}^d} \{\langle \lambda, z \rangle - \Lambda(\lambda)\} \\ &= \sup_{\lambda \in \mathbb{R}^d} \{\langle \lambda, z \rangle - log(r_\lambda)\} \\ &= I(z),\end{aligned}$$

por lo que $\Lambda^* = I$.

(f) Sean $x, y \in \mathbb{R}^d$ y $t \in [0,1]$. Notemos que

$$\begin{aligned}\Lambda^*(tx + (1-t)y) &= \sup_{\lambda \in \mathbb{R}^d} \{\langle \lambda, tx + (1-t)y \rangle - log(r_\lambda)\} \\ &= \sup_{\lambda \in \mathbb{R}^d} \{t\langle \lambda, x \rangle + (1-t)\langle \lambda, y \rangle - log(r_\lambda)\} \\ &= \sup_{\lambda \in \mathbb{R}^d} \{t\langle \lambda, x \rangle + (1-t)\langle \lambda, y \rangle - log(r_\lambda) + tlog(r_\lambda) - tlog(r_\lambda)\},\end{aligned}$$

por lo que

$$\Lambda^*(tx + (1-t)y) = \sup_{\lambda \in \mathbb{R}^d} \{t[\langle \lambda, x \rangle - log(r_\lambda)] + (1-t)[\langle \lambda, y \rangle - log(r_\lambda)]\}. \quad (3.32)$$

Como el supremo de la suma es menor o igual que la suma de los supremos, de (3.32) se sigue que

$$\Lambda^*(tx + (1-t)y) \leq \sup_{\lambda \in \mathbb{R}^d} \{t[\langle \lambda, x \rangle - log(r_\lambda)]\} + \sup_{\lambda \in \mathbb{R}^d} \{(1-t)[\langle \lambda, y \rangle - log(r_\lambda)]\}. \quad (3.33)$$

Observemos que $t \geq 0$, por lo que

$$\sup_{\lambda \in \mathbb{R}^d} \{t[\ \langle \lambda, x \rangle - log(r_\lambda)\ ]\} = t \sup_{\lambda \in \mathbb{R}^d} \{\langle \lambda, x \rangle - log(r_\lambda)\} = t\Lambda^*(x); \tag{3.34}$$

similarmente, $(1-t) \geq 0$, por lo cual

$$\sup_{\lambda \in \mathbb{R}^d} \{(1-t)[\langle \lambda, y \rangle - log(r_\lambda)]\} = (1-t) \sup_{\lambda \in \mathbb{R}^d} \{\langle \lambda, y \rangle - log(r_\lambda)\} = (1-t)\Lambda^*(y). \tag{3.35}$$

Sustituyendo en (3.33) los valores encontrados en (3.34) y (3.35),

$$\Lambda^*(tx + (1-t)y) \leq t\Lambda^*(x) + (1-t)\Lambda^*(y).$$

Esto demuestra que la función Λ^* es convexa.

Por el teorema de Gärtner-Ellis, la familia $S_n = \frac{1}{n}\sum_{i=1}^{n} f(X_i)$ satisface el P.G.D. con la buena función de grandes desviaciones $\Lambda^* = I$, y además por el inciso (f) de esta demostración se tiene que I es también una función convexa. $\square$

Capítulo 4

Conclusiones

En el primer capítulo introdujimos el concepto de matriz regular. Posteriormente, enunciamos y demostramos el Teorema de Perron-Frobenius (el cual se aplica a matrices regulares) y, naturalmente, nombramos al eigenvalor real que se desprende de dicho teorema como el eigenvalor de Perron-Frobenius. Concluimos el capítulo estudiando si un conjunto particular de sucesiones que dependen de las entradas de una matriz regular y de un vector positivo arbitrario convergen, y fue en este punto donde comenzamos a vislumbrar algunas consecuencias del teorema de Perron-Frobenius, pues probamos que dichas sucesiones sí convergen y además el término al que convergen es el logaritmo del eigenvalor de Perron-Frobenius.

En el segundo capítulo demostramos que, si una cadena de Markov es irreducible y aperiódica, entonces dicha cadena cuenta con una matriz regular, es decir, a partir de un número finito de pasos todas las entradas de la matriz de transición son positivas y por lo tanto podemos aplicar el teorema de Perron-Frobenius a dicha matriz. Estudiando más a fondo este tipo de cadenas, al que llamamos cadenas regulares, pudimos observar que el eigenvalor de Perron-Frobenius de la matriz de transición resulta ser 1, lo cual implica que el resto de los eigenvalores de la matriz de transición cuentan con norma compleja estrictamente menor a 1.

Un resultado que se puede ver en un curso de Procesos Estocásticos es que las probabilidades de transición en m pasos de una cadena irreducible y aperiódica convergen a la probabilidad estacionaria, pero usando el teorema de Perron-Frobenius pudimos ir más allá: en el caso de este tipo de cadenas, la velocidad de convergencia de dichas probabilidades de transición es del orden de la norma del segundo eigenvalor más grande (el cual sabemos por el resultado anterior que es estrictamente menor a 1).

Finalmente, en el tercer capítulo introdujimos y probamos una versión del teorema de Gärtner-Ellis. Este teorema nos permitió obtener un resultado aún más fuerte que los anteriores: la media muestral de una funcional de cadenas de Markov irreducibles y aperiódicas satisface el Principio de Grandes Desviaciones definido en 17 con la buena función de grandes desviaciones

$$I(z) = \sup_{\lambda \in \mathbb{R}^d} \{\langle \lambda, z \rangle - log(r_\lambda)\},$$

donde r_λ es el eigenvalor de Perron-Frobenius de una matriz regular que depende de la matriz de transición original. Esta expresión resulta relevante en nuestro estudio, pues nuevamente nos encontramos con el eigenvalor de Perron-Frobenius de una matriz regular.

Bibliografía

[1] BARANOSKI, G. V., ROKNE, J. G., AND XU, G. Applying the exponential chebyshev inequality to the nondeterministic computation of form factors. *Journal of Quantitative Spectroscopy and Radiative Transfer 69*, 4 (2001), 447–467.

[2] BIGGINS, J., AND SANI, A. R. *Extended Perron-Frobenius Results*. Citeseer, 2004.

[3] BILLINGSLEY, P. *Probability and measure*. John Wiley & Sons, 2008.

[4] BODINEAU, T. Nonequilibrium statistical mechanics & large deviation theory. `https://youtu.be/s4pznYlf0aM`. Fecha de consulta: 2020-02-24.

[5] DEMBO, A., AND ZEITOUNI, O. *Large Deviations Techniques and Applications*, vol. 38. 1998. 2da. Edición.

[6] FRIEDBERG, S. H., INSEL, A. J., AND SPENCE, L. E. *Algebra lineal*. Publicaciones Cultural, 1982.

[7] HOEL, P. G., PORT, S. C., AND STONE, C. J. *Introduction to stochastic processes*. Waveland Press, 1986.

[8] ROLSKI, T., SCHMIDLI, H., SCHMIDT, V., AND TEUGELS, J. L. *Stochastic processes for insurance and finance*, vol. 505. John Wiley & Sons, 2009.

[9] SENETA, E. *Non-negative matrices and Markov chains*. Springer Science & Business Media, 2006.

[10] SHANKAR, R. *Principles of quantum mechanics*. Springer Science & Business Media, 2012.

[11] TAYLOR, A. E., AND MANN, W. R. *Advanced calculus*. 1983.

Apéndice A

Demostración de la derivabilidad de la función F

A continuación enunciaremos y demostraremos un teorema que nos permite asegurar la derivabilidad de la función F definida en la demostración del teorema 9. Esta demostración estará basada en el ejercicio contenido en [2, p. 3].

Teorema 10. *Sean $L \subseteq \mathbb{R}^d$ un conjunto abierto (no vacío) con respecto a la topología usual y $m_{ij} : L \to \mathbb{R}$ una función para cada $i, j \in \{1, ..., n\}$, $n \in \mathbb{N}$.*
Para cada $\lambda \in L$ se definen $M(\lambda) = (m_{ij}(\lambda)) \in M_{n\times n}(\mathbb{R})$ y $L' = \{\lambda \in L \mid m_{ij}(\lambda) \geq 0$ para toda $i, j\}$.
Supongamos que para cada $i, j \in \{1, ..., n\}$ la función m_{ij} es q veces derivable, $q \in \mathbb{N}$, y para cualquier $\theta \in L'$ la matriz $M(\theta)$ es regular.
Entonces para cada $\theta \in L'$ existe $\delta_\theta > 0$ tal que $\lambda \in B(\theta, \delta_\theta) \subseteq L$ implica que la función $\varphi(\lambda) = \rho(\lambda)$ es q veces derivable, donde $\rho(\lambda)$ es el eigenvalor más grande (en norma compleja) de la matriz $M(\lambda)$.

Demostración.
Para cada $\lambda \in L$ definimos $f(z, \lambda)$ como el polinomio característico de la matriz $M(\lambda)$, es decir, $f(z, \lambda) = det(M(\lambda) - zI)$, donde I es la matriz identidad. Observemos que

$$M(\lambda) - zI = \begin{pmatrix} m_{11}(\lambda) - z & m_{12}(\lambda) & ... & m_{1n}(\lambda) \\ m_{21}(\lambda) & m_{22}(\lambda) - z & ... & m_{2n}(\lambda) \\ \vdots & \vdots & & \vdots \\ m_{n1}(\lambda) & m_{n2}(\lambda) & ... & m_{nn}(\lambda) - z \end{pmatrix},$$

por lo que $M(\lambda) - zI$ tiene, en cada una de sus entradas, una función q-veces derivable en λ y al ser $det(M(\lambda) - zI) = f(z, \lambda)$ una suma finita de productos finitos de esas funciones, $f(z, \lambda)$ también resulta ser una función q-veces derivable en λ. En particular, $f(z, \theta)$ es q-veces derivable para cualquier $\theta \in L' \subseteq L$.

Sea $\theta \in L'$. Como la matriz $M(\theta)$ es regular, por el teorema de Perron-Frobenius esta matriz tiene un eigenvalor real positivo $\rho(\theta)$ de multiplicidad 1 y que es mayor que la norma compleja del resto de sus eigenvalores $\rho_1(\theta)$, $\rho_2(\theta)$, ..., $\rho_s(\theta)$, donde $1 \leq s \leq n-1$.

Como $\rho(\theta)$ es una raíz simple del polinomio característico $det(M(\theta) - zI) = f(z, \theta)$, entonces f satisface las condiciones del teorema de la función implícita en $(\rho(\theta), \theta)$[1].

Por el teorema de la función implícita existe $\delta_1 > 0$ tal que $B(\theta, \delta_1) \subseteq L$ y hay una única función $r(\lambda)$ q veces derivable tal que $r(\theta) = \rho(\theta)$ y que satisface que $f(r(\lambda), \lambda) = 0$ para toda $\lambda \in B(\theta, \delta_1)$. Observemos que, dada $\lambda \in B(\theta, \delta_1)$, como $f(r(\lambda), \lambda) = 0$ entonces $r(\lambda)$ es un eigenvalor de la matriz $M(\lambda)$.

Queda probar que, para una δ lo suficientemente pequeña, $r(\lambda)$ es el eigenvalor de Perron-Frobenius de la matriz $M(\lambda)$.

Tomemos $\delta_2 > 0$ tal que las bolas de radio δ_2 centradas en los distintos eigenvalores de $M(\theta)$ sean disjuntas dos a dos, y definimos $\delta_\theta = \text{mín}\{\delta_1, \delta_2\}$.

Entonces para $\lambda \in B(\theta, \delta_\theta)$ se tiene que $B(\theta, \delta_\theta) \subseteq L$ y la función $r(\lambda)$ es derivable. Además, como $r(\lambda)$ es un eigenvalor de la matriz $M(\lambda)$ y el único eigenvalor en $B(\theta, \delta_\theta)$ es el eigenvalor de Perron-Frobenius, se sigue que $r(\lambda)$ es el eigenvalor de Perron-Frobenius. Finalmente, como la función $r(\lambda)$ es la única tal que $r(\theta) = \phi(\theta)$, entonces $r(\lambda) = \varphi(\theta) = \rho(\theta)$. $\square$

[1]Teorema contenido en [11, p. 228].