

# **Cadenas de Markov: Convergencia Geométrica de Probabilidades de Transición hacia Distribuciones Invariantes**

Gabriel Hugo Flores Gómez

4 de julio de 2007



# Agradecimientos

A mi asesor Dr. Raúl Montes de Oca, por su excelente dirección, por su calidez humana y su paciencia infinita que tuvo hacia mi persona.

Al Dr. Julio César García Corte, Dr. Juan González Hernández y al Dr. Evgueni Illich Gordienko, que con sus comentarios y correcciones me permitieron tener una mejor versión de este trabajo.

Al Dr. Evgueni Illich Gordienko por hacerme notar la existencia de los trabajos de N. V. Kartashov y su relación con esta tesis.

A Conacyt que me proporcionó beca durante dos años.

Con la UAM-I por las facilidades que me fueron otorgadas en la realización de mis estudios y en la realización de esta tesis.

Finalmente, con el Instituto de Educación Media Superior del D.F, que me brinda estabilidad y un espacio de desarrollo profesional y laboral.



# Índice general

<b>I Preliminares</b>	<b>9</b>
1. Introducción	11
2. Cadenas de Markov	15
2.1. Definición y propiedades . . . . .	15
2.2. Distribución invariante . . . . .	18
2.3. Análisis de la estructura en el espacio de estados . . . . .	19
2.3.1. Condición de minorización y de Foster-Lyapunov . . . . .	21
2.4. Cadenas de Markov particionantes . . . . .	23
2.4.1. Construcción de Nummelin . . . . .	23
2.4.2. Construcción de Athreya y Ney . . . . .	25
3. Acoplamiento y orden estocástico	29
3.1. Acoplamiento . . . . .	30
3.1.1. Definición y ejemplos . . . . .	30
3.1.2. Desigualdades de acoplamiento . . . . .	34
3.1.3. Acoplamientos particionantes . . . . .	37
3.2. Orden estocástico . . . . .	38

3.2.1. Acoplamiento ordenado en variables aleatorias . . . . .	40
3.2.2. Acoplamiento ordenado en elementos aleatorios y procesos estocásticos . . . . .	41

## II Planteamiento del problema 45

<b>4. Planteamiento y antecedentes</b>	<b>47</b>
4.1. Planteamiento . . . . .	47
4.2. Antecedentes . . . . .	49

## III Tasas de convergencia 51

<b>5. Método espectral</b>	<b>53</b>
<b>6. Acoplamiento</b>	<b>59</b>
6.1. Espacio de estados pequeño . . . . .	60
6.1.1. Caso finito . . . . .	60
6.1.2. Caso compacto . . . . .	61
6.2. Monotonicidad estocástica . . . . .	62
6.2.1. Desarrollo . . . . .	62
6.2.2. Ejemplos . . . . .	68
<b>7. Sucesiones aperiódicas y renovación</b>	<b>73</b>
7.1. Desarrollo . . . . .	73
7.1.1. Cotas . . . . .	75
7.1.2. Resultados Principales . . . . .	78
7.1.3. Ejemplos . . . . .	85

<i>ÍNDICE GENERAL</i>	7
-----------------------	---

7.2. Comentarios: teoría de Kartashov . . . . .	87
---	----

## **IV Conclusión y problemas abiertos** **89**

<b>8. Conclusión y problemas abiertos</b>	<b>91</b>
---	-----------

8.1. Conclusión . . . . .	91
---------------------------	----

8.2. Problemas abiertos . . . . .	95
-----------------------------------	----

## **V Apéndices** **97**

<b>A. Resultados básicos</b>	<b>99</b>
------------------------------	-----------

A.1. Existencia de conjuntos pequeños . . . . .	99
---	----

A.2. Convergencia a la distribución invariante . . . . .	102
--	-----

<b>B. Teorema de Strassen</b>	<b>107</b>
-------------------------------	------------

<b>C. Procesos de decisión de Markov</b>	<b>115</b>
--	------------

C.1. Problema de control óptimo (PCO) . . . . .	117
---	-----

C.2. Análisis del índice de estabilidad para procesos de decisión de Markov . . . . .	117
---	-----

C.2.1. Problema y motivación . . . . .	117
--	-----

C.2.2. Desarrollo . . . . .	118
-----------------------------	-----

C.2.3. Modelo autorregresivo con barrera reflejante . . . . .	120
---	-----

<b>D. Uso del teorema de Kendall y cotas para la V-norma</b>	<b>123</b>
--	------------

D.1. Resultados sobre convergencia . . . . .	123
--	-----

D.2. Desigualdades notables . . . . .	131
---------------------------------------	-----



**Parte I**

**Preliminares**



# Capítulo 1

## Introducción

Este trabajo está relacionado con Cadenas de Markov (CMs), con espacio de estados contenido en  $\mathbb{R}$ , las cuales suponemos que cuentan con distribuciones invariantes ([11], [33], [37], [40], [43]).

Para tales CMs se presenta un estudio acerca de las tasas de convergencia, en alguna métrica adecuada, de las probabilidades de transición hacia las distribuciones invariantes ([4],[23], [24], [26],[30], [31], [39], [41], [44], [45], [46], [47], [49], [51], [54]).

Cabe señalar que en algunas de las referencias mencionadas en el párrafo anterior, únicamente se presentan condiciones para garantizar la *existencia* de una tasa de convergencia de tipo geométrico, sin proponerla de forma explícita ([55]).

Esencialmente, en este trabajo nos concentraremos en el tratamiento *explícito* de las tasas de tipo geométrico.

Cuando se tienen tasas de convergencia de forma explícita, éstas pueden ser aplicadas, por ejemplo, en el análisis de la estabilidad de procesos de decisión de Markov (PDMs) ( véase [12], [35], [50]), y en los algoritmos relacionados con la Cadena de Markov de Montecarlo ([44]).

Específicamente, las tasas presentadas en este trabajo, corresponden a CMs con propiedades especiales, como son: la presencia de un espacio de estados finito, compacto o con elemento mínimo, la monotonicidad estocástica, y la renovación.

La teoría y los ejemplos presentados tienen como base, principalmente, los artículos [4], [30], [31], [35] y [47]. Esta selección de artículos, aunque no es exhaustiva, da un *panorama* de las distintas técnicas encontradas en la litera-

tura acerca de tasas de convergencia.

Para sustentar el desarrollo de los trabajos mencionados en el párrafo anterior, en la tesis se ofrece, de manera amplia, el tratamiento de los temas: desigualdades de acoplamiento, cadenas de Markov particionantes, y el Teorema de Strassen.

La teoría desarrollada en el trabajo está ilustrada con diversos ejemplos, entre los que se incluyen, caminatas aleatorias, colas, modelos autorregresivos lineales, y no lineales.

La tesis está organizada en cinco partes.

La primera parte es de Preliminares y consta de los capítulos 1, 2 y 3, denominados: Introducción, Cadenas de Markov, y Acoplamiento y orden estocástico, respectivamente.

La segunda parte está dedicada a establecer el problema principal de la tesis y consta de un sólo capítulo (capítulo 4).

La siguiente parte, llamada Tasas de convergencia, está constituida de tres capítulos. En ellos se tratan resultados sobre tasas de convergencia explícitas, de tipo de geométrico, para cadenas de Markov.

Comenzamos esta parte con el capítulo 5, mostrando una teoría que trata con CMs con espacios de estados finitos y que además explota las ideas de valores y vectores propios asociados a una matriz de transición.

Después, en el capítulo 6, se presentan dos teorías que tienen en común la técnica del acoplamiento. La primera trata el caso de espacios finitos o compactos, y usa principalmente la propiedad de aperiodicidad fuerte. Para ilustrar esta parte del capítulo se dan dos ejemplos.

En la segunda parte del capítulo 6, llamada monotonía estocástica, se tratan CMs con espacios de estados (posiblemente) no numerables que cuentan con un elemento mínimo. Las tasas presentadas en este capítulo requieren que las realizaciones de la cadena considerada, estén ordenadas en un sentido apropiado. En esta parte se presentan diversos ejemplos, incluyendo, en particular, la caminata aleatoria de Lindley, y una aplicación a PDMs. Esta última está desarrollada en el apéndice *C*.

Para finalizar la parte III de la tesis, presentamos un capítulo (capítulo 7) en el cual se analizan CMs que cumplen las propiedades de reversibilidad y/o positividad. Estas propiedades junto con algunos hechos de la teoría de procesos de renovación nos brindan las tasas de convergencia explícitas. Este capítulo está basado, principalmente, en [4]. En él también se incluyen comentarios acerca de la teoría desarrollada por Kartashov en [24], [25] y [26]. Finalmente el capítulo se ejemplifica con el algoritmo de Metropolis-Hastings.

La cuarta parte contiene un capítulo con la conclusión y los problemas abiertos de la tesis. En el capítulo ofrecemos, además, una tabla cuyo contenido es el conjunto de las condiciones requeridas en la parte III, con las cuales se obtiene convergencia de tipo geométrico a la distribución invariante. Esto último es con el propósito de que el lector tenga una guía rápida de consulta y acceso al material relacionado con las tasas de convergencia.

Finalmente, la última parte contiene cuatro apéndices. Los dos primeros muestran resultados auxiliares acerca de cadenas de Markov. El tercero presenta una introducción a los PDMs y a su estabilidad para el caso de costo promedio esperado, y en el cuarto se proporcionan resultados auxiliares para el capítulo 7.



## Capítulo 2

# Cadenas de Markov

Este capítulo está dedicado a establecer ciertos conceptos relacionados con cadenas de Markov.

Revisaremos las propiedades de CMs que se desprenden en relación a su construcción y estructura. Además, se da la teoría necesaria para construir realizaciones de Markov con distribuciones aleatorias (split chain), las cuales tendrán una propiedad de gran interés: **la regeneración**.

El presente capítulo está basado en las siguientes referencias: [2],[33], [36], [37], [43] y [56].

### 2.1. Definición y propiedades

A lo largo de este trabajo denotaremos :

- $(E, \xi)$  como un espacio medible, donde  $E$  es un conjunto no vacío y  $\xi$  es una  $\sigma$ -álgebra de subconjuntos de  $E$ , y
- $p$  como un kernel de transición, es decir,  $p : E \times \xi \rightarrow [0, 1]$  y cumple:
  - (i)  $p(x, \cdot)$  es medida de probabilidad en  $\xi$ , para todo  $x$  en  $E$ .
  - (ii)  $p(\cdot, A)$  es una función medible de  $\xi$  en  $\mathbb{B}[0, 1]$ , para toda  $A$  en  $\xi$ .

Considerando el kernel de transición  $p$  y una probabilidad  $\mu$  en  $\xi$ , podemos generar un proceso estocástico  $X = \{X_n\}_{n=0}^{\infty}$  (véase [1]), definido en cierto espacio de probabilidad  $(\Omega, \mathcal{F}, P_\mu)$ , donde:

- $\Omega = E^{N^*}$ , donde  $N^* = \mathbb{N} \cup \{0\}$  y  $\mathbb{N}$  es el conjunto de números naturales.
- $\mathcal{F}$  es la  $\sigma$ -álgebra producto de las  $\xi$ 's.
- $P_\mu$  es la probabilidad construida en el teorema de Ionescu-Tulcea (véase [1]).
- El proceso  $X$ , tiene la característica de que:

$$P_\mu(X_0 \in A) = \mu(A), \quad A \in \xi,$$

por eso diremos que  $X$  tiene **distribución inicial**  $\mu$ .

Con el contexto de la construcción anterior, definimos ahora un proceso estocástico que posee la propiedad de no depender de su pasado.

**Definición 2.1.1** Decimos que  $X = \{X_n\}_{n=0}^\infty$  es una **realización de Markov** con espacio base  $(\Omega, \mathcal{F}, P_\mu)$ , kernel de transición  $p$  y espacio de estados  $(E, \xi)$  si:

$$P_\nu(X_n \in A | X_{n-1}) = p(X_{n-1}, A), \quad c.s. \quad (2.1a)$$

$$P_\mu(X_n \in A | X_0, X_1, \dots, X_{n-1}) = P_\mu(X_n \in A | X_{n-1}), \quad c.s. \quad (2.1b)$$

para toda  $n \geq 1$ ,  $w \in \Omega$  donde  $p(X_n, A)(\omega) = p(X_n(\omega), A)$ . La ecuación 2.1b, es llamada **propiedad de Markov**.

Lo que indica la **propiedad de Markov** 2.1b, es que en cualquier tiempo fijo, el comportamiento del proceso sólo depende del tiempo inmediatamente anterior al fijado.

**Definición 2.1.2** Bajo el contexto de la definición 2.1.1, llamamos **cadena de Markov (CM)** con espacio de estados  $(E, \xi)$  a la familia:

$$\{X : X \text{ es una realización de Markov con distribución inicial } \lambda \in M_\xi\},$$

donde  $M_\xi = \{\lambda : \lambda \text{ es probabilidad en } (E, \xi)\}$ .

### Notación 2.1.1

- Denotaremos una cadena de Markov simplemente por  $X = \{X_n\}_{n=0}^\infty$ , es decir, igual a una realización de Markov.

- $P_x$  será la probabilidad obtenida con el teorema de Ionescu-Tulcea, cuando la distribución inicial es  $\delta_x$ , donde

$$\delta_x(A) := \begin{cases} 1, & \text{si } x \in A \\ 0, & \text{si } x \notin A \end{cases},$$

$A \in \xi$ .

- Si  $Y$  es una variable aleatoria definida en  $(\Omega, \mathcal{F}, P_\mu)$ , entenderemos por  $E_\mu[Y]$  la integral  $\int_\Omega Y dP_\mu$ . Cuando  $\mu = \delta_x$ , se denotará esta esperanza por  $E_x[Y]$ .

Ahora daremos algunas propiedades que tienen las realizaciones de Markov, refiriéndose a su respectiva ley de probabilidad (estas propiedades pueden ser consultadas en [33] y [43]).

**Propiedades 2.1.1** *Bajo el contexto de la definición 2.1.2, se cumplen las siguientes propiedades:*

- (a) Si  $n \geq 0$  y  $B_0, \dots, B_n \in \xi$ . Entonces

$$P_\mu(X_0 \in B_0, \dots, X_n \in B_n) = \int_{B_0} \mu(dx_0) \int_{B_1} p(x_0, dx_1) \dots \int_{B_n} p(x_{n-1}, dx_n).$$

- (b) Para  $n \geq 1$  y  $A \in \xi$  se define  $p^n(x, A) := \int_E p(x, dy) p^{n-1}(y, A)$ . Entonces se tiene:

$$P_x(X_n \in A) = p^n(x, A), \quad n \geq 1.$$

- (c) Sean  $A \in \mathcal{F}$ ,  $B \in \xi$ ,  $Y$  una variable aleatoria definida sobre  $\Omega$  y  $n \geq 1$ . Entonces se tiene:

(i)  $P_\mu(X_n \in B) = \mu p^n(B) := \int_E p^n(x, B) d\mu(x)$ ,

(ii)  $P_\mu(A) = \int_E P_x(A) d\mu(x)$ ,

(iii)  $E_\mu[Y] = \int_E E_x[Y] d\mu(x)$ ,

(iv)  $P_\mu((X_0, \dots, X_n, \dots) \in A | X_0 = x) = P_x((X_0, \dots, X_n, \dots) \in A)$ .

- (d) La ecuación de **Chapman-Kolmogorov**, es decir:

$$p^n(x, A) = \int_E p^m(x, dy) p^{n-m}(y, A), \quad m \leq n.$$

**Observación 2.1.1** *Existen varias caracterizaciones de la propiedad de Markov (2.1b). Los lectores interesados en consultarlas lo pueden hacer, por ejemplo, en [43].*

## 2.2. Distribución invariante

**Definición 2.2.1** Decimos que  $\pi$  es la **distribución invariante** ó **distribución estacionaria** para la realización  $X$ , si:

$$\pi(A) = \int_E p(x, A) d\pi(x), \quad A \in \xi.$$

o equivalentemente, para  $A \in \xi$

$$\pi(A) = P_\pi(X_1 \in A).$$

La explicación del nombre en la definición anterior, la podemos ver a partir de lo siguiente.

**Propiedades 2.2.1** Sea  $\pi$  una distribución invariante. Entonces

(a) Para  $n \geq 1$ , se cumple que:

$$\pi(A) = \int_E p^n(x, A) d\pi(x), \quad A \in \xi.$$

Por lo tanto, si  $\{X_n\}_{n=0}^\infty$  es una cadena de Markov con distribución inicial  $\pi$ , se tiene que:

$$P_\pi(X_n \in A) = P_\pi(X_m \in A).$$

para toda  $m$  y  $n$ .

(b) Si  $\{X_n\}_{n=0}^\infty$  es una cadena de Markov con distribución inicial  $\pi$ , entonces el proceso  $\{X_n\}_{n=0}^\infty$  es estacionario, es decir, los procesos  $\Theta(X)$  y  $X$  tienen la misma distribución, donde  $\Theta$  es el operador corrimiento (véase [43]).

### **Demostración.**

“ (a) ” A partir la definición 2.2.1, obtenemos la afirmación de manera directa.

“ (b) ” Resulta del hecho de que  $\pi p^n = \pi p^m$ , para toda  $m$  y  $n$  (véase inciso (c) de las propiedades 2.1.1). ■

La existencia de distribuciones invariantes se observa en el inciso (b) de la proposición 2.3.2, debido a su relación con el concepto de irreducibilidad.

## 2.3. Análisis de la estructura en el espacio de estados

Como motivación, recordemos que cuando consideramos CMs en espacios de estados numerables se tiene el concepto de irreducibilidad, el cual utiliza la siguiente noción de comunicación (véase [18]).

Para  $x$  y  $y$  estados, decimos que  $x \leftrightarrow y$  si y solo si,  $x \rightarrow y$  y  $y \rightarrow x$ , donde  $x \rightarrow y$  significa que, existe  $n \in \mathbb{N}$ , tal que  $p^n(x, y) > 0$ .

Además, cuando la cadena **no es irreducible**, podemos particionar el espacio de estados de la siguiente manera:

$$E = \bigcup_{x \in I} [x] \cup D,$$

donde  $D$ , es un conjunto de estados transitorios,  $I \subseteq E$  y  $[x]$  es la clase de equivalencia generada por la relación " $\leftrightarrow$ " tal que para todo  $x, y \in I$  se tiene  $[x] \cap [y] = \emptyset$  o  $[x] = [y]$ .

Con lo anterior podemos considerar que una cadena de Markov está compuesta de subcadenas de Markov (el número de éstas, depende de cuantos  $[x]$ 's tenemos), las cuales tienen como espacio de estados la clase  $[x]$  y kernel de transición igual al original, pero restringido a  $[x]$ .

El concepto de irreducibilidad para espacios de estados numerables, nos dice que la cadena no puede estar compuesta de subcadenas. De igual manera se tendrá esta propiedad para el caso de espacio de estados generales, pero con una definición diferente de irreducibilidad.

A continuación preparamos el camino para dar la generalización del concepto de **irreducibilidad en espacios de estados generales**, en dos definiciones. La primera nos indicará, que no habrá subcadenas; mientras que la segunda, garantizará la comunicación en ambos sentidos.

**Definición 2.3.1** Sea  $\varphi$  una **medida  $\sigma$ -finita** sobre  $(E, \xi)$  y sea  $B \in \xi$  tal que  $\varphi(B) > 0$ .

- (a) El **conjunto**  $B \in \xi$  es llamado  **$\varphi$ -comunicado** si para todo  $x \in B$ ,  $A \subseteq B$  medible con  $\varphi(A) > 0$  se cumple:

$$p^n(x, A) > 0,$$

para algún  $n \in \mathbb{N}$ .

- b)  $p$  es llamado  **$\varphi$ -irreducible**, si  $E$  es  $\varphi$ -comunicado. En este caso,  $p$  también será llamado **irreducible**, y  $\varphi$  medida irreducible para  $p$ .

**Definición 2.3.2** Sea  $M_{I_r} = \{\varphi : \varphi \text{ es medida de probabilidad irreducible para } p\}$  y  $\psi \in M_{I_r}$ .

Decimos que  $\psi$  es **medida irreducible maximal** si, para toda  $\lambda \in M_{I_r}$  se cumple que  $\lambda \ll \psi$ .

**Observación 2.3.1** Para CMs con espacios de estados numerables, se puede verificar que la medida irreducible maximal, es justamente la medida de conteo.

La siguiente proposición, muestra que la definición 2.3.2, no es vacía.

**Proposición 2.3.1** Suponga que  $p$  es  $\varphi$ -irreducible. Entonces existe una medida irreducible maximal.

**Demostración.** Véase [33, p.88], [37, p. 13] o [54, p. 841].

■

Así teniendo el kernel de transición  $p$  irreducible, llamaremos a la cadena  $X$   **$\psi$ -irreducible**; donde  $\psi$  es la medida irreducible maximal.

**Proposición 2.3.2**

- (a) Si una cadena de Markov es  $\varphi$ -irreducible y tiene distribución invariante  $\pi$ , entonces  $\varphi \ll \pi$ , es decir,  $\pi(A) > 0$  cuando  $\varphi(A) > 0$ ,  $A \in \xi$ .
- (b) Si  $p$  es irreducible para  $\varphi$  y  $\varphi$ -recurrente (ver def. 2.3.6 a continuación), entonces se tiene una única distribución invariante  $\pi$  la cual es  $\sigma$ -finita e irreducible maximal (la unicidad se entenderá, salvo múltiplos constantes).

**Demostración.**

“ (a) ”

Sea  $\varphi(A) > 0$ , entonces por la  $\varphi$ -irreducibilidad tenemos que para algún  $n \in \mathbb{N}$  y todo  $x \in E$  se establece,  $p^n(x, A) > 0$ . De aquí podemos ver :

$$E = \bigcup_{n,m \in \mathbb{N}} \{x \in E : p^n(x, A) \geq \frac{1}{m}\}.$$

Así, por numerabilidad aditiva, podemos encontrar  $n, m \in \mathbb{N}$  y  $B \in \xi$  tal que: si  $\pi(B) > 0$ , entonces

$$p^n(x, A) \geq \frac{1}{m},$$

para todo  $x \in B$ , pero entonces,

$$\begin{aligned}\pi(A) &= \int_E p^n(x, A)\pi(dx) \geq \int_B p^n(x, A)\pi(dx) \\ &\geq \int_B \frac{1}{m}\pi(dx) > 0.\end{aligned}$$

Por lo tanto  $\varphi \ll \pi$ .

“ (b) ”

Puede encontrarse en [37, p.32].

■

Finalizamos este apartado con la siguiente definición.

**Definición 2.3.3** Una cadena  $\psi$ -irreducible que cuenta con una distribución invariante, es llamada **positiva**; si no la tiene es llamada **nula**. Cuando la distribución invariante es única es llamada **ergódica**.

### 2.3.1. Condición de minorización y de Foster-Lyapunov

En este apartado definimos las propiedades importantes que tendrán las CMs a lo largo del trabajo.

**Definición 2.3.4** Sea  $p$  un kernel de transición. Decimos que  $p$  satisface la **condición de minorización**, si existen  $m_0 \in \mathbb{N}$ ,  $0 < \delta < 1$ , una función  $S$  que es medible de  $\xi$  en  $\mathbb{B}(\mathbb{R}^+)$  y una probabilidad  $\nu$  definida en  $\xi$ , tal que:

$$p^{m_0}(x, \cdot) \geq \delta S(x)\nu(\cdot), \quad x \in E. \quad (2.2)$$

A  $S$  y a  $\nu$  se le conocen como **función pequeña y medida pequeña**, respectivamente.

**Notación.** En el contexto de la definición 2.3.4, diremos que  $p$  satisface la condición  $M(m_0, \delta, S, \nu)$ .

**Observación 2.3.2** Meyn y Tweedie en [33], toman la definición 2.3.4 con  $m_0 = 1$ ; por generalidad nosotros no hacemos lo mismo.

**Definición 2.3.5** Un conjunto  $C \in \xi$  es llamado **pequeño** si existe  $n \in \mathbb{N}$  y una medida  $\tilde{\nu}$  definida sobre  $\xi$  que satisface la siguiente condición

$$p^n(x, A) \geq \tilde{\nu}(A), \quad x \in C \text{ y } A \in \xi.$$

**Observación 2.3.3** *El conjunto  $C \in \xi$  definido en la condición de minorización  $M(n, \delta, I_C, \nu)$ , es un conjunto pequeño, donde  $I_C$  es la función indicadora.*

Para que el kernel  $p$  satisfaga la condición  $M(m_0, \delta, S, \nu)$  necesitamos la existencia de la función  $S$ , la medida  $\nu$  y el conjunto  $C$ . En la sección A.1 de el **apéndice A**, se observa que si  $p$  es **irreducible** para  $\psi$ , entonces se tiene la existencia de la función  $S$ , la probabilidad  $\nu$  y el conjunto  $C$ .

Por otro lado, si un conjunto pequeño  $C$  tiene medida positiva con respecto a  $\nu$ , vemos de la definición 2.3.5, que  $\nu$  es medida irreducible para  $p$ ; así tenemos que la cadena  $X$  es  $\psi$ -irreducible. Con esto se tiene un método para verificar la  $\psi$ -irreducibilidad.

Ahora damos algunas definiciones relacionadas con los conceptos de conjunto pequeño y de recurrencia. Estas definiciones serán ocupadas en la construcción de cadenas particionantes (split chain).

**Definición 2.3.6** *En el contexto de la definición 2.3.5, diremos que :*

- (i)  $X$  es  $\varphi$ -**recurrente** ( o Harris recurrente), si para todo  $B \in \xi$  con  $\varphi(B) > 0$ , se tiene:

$$P_x(X_n \in B, \text{ para algún } n) = 1, \quad x \in B.$$

- (ii)  $X$  es llamada **recurrente en  $C$** , si:

$$P_x(X_n \in C, \text{ para algún } n) = 1, \quad x \in E.$$

- (iii)  $X$  es **fuertemente aperiódica**, si para todo  $A \subseteq C$  y todo  $x \in A$  se tiene:

$$p(x, A) \geq \delta \nu(A).$$

Finalmente terminamos este apartado con la siguiente definición, que será ocupada como hipótesis en el siguiente capítulo.

**Definición 2.3.7** *Sean  $C$  conjunto pequeño,  $0 < \lambda < 1$ ,  $0 < b < \infty$  y  $V : E \rightarrow [1, \infty]$ . Decimos que  $p$  satisface la **propiedad de Foster-Lyapunov** para  $V$ ,  $\lambda$ ,  $b$  y  $C$  si para  $x \in E$  se cumple:*

$$pV(x) := \int_E V(y)p(x, dy) \leq \lambda V(x) + bI_C(x).$$

## 2.4. Cadenas de Markov particionantes

En esta sección, construiremos realizaciones de Markov  $\{\tilde{X}_n\}_{n=0}^\infty$ , que tienen la propiedad de que su espacio de estados está formado por una partición de conjuntos y su kernel de transición está definido por partes. Estas cadenas las llamaremos **particionantes**.

Presentamos aquí, las dos construcciones existentes en la literatura, que generan este tipo de cadenas. El material aquí recopilado fue consultado en [2], [33] y [36].

Para el desarrollo del tema, damos el siguiente contexto.

Sea  $X = \{X_n\}$  una CM con kernel  $p$  y espacio de estados  $E$ . Además supóngase que  $X$  es Harris-recurrente y fuertemente aperiódica.

A partir de esta CM construiremos otra CM  $\tilde{X}$  con las propiedades mencionadas en el primer párrafo de esta sección.

Para lograr este fin, se construirá a  $\tilde{X}$  como una cadena bivariada, que tendrá en su primera componente a la CM  $X$  y en su segunda componente a una sucesión de variables aleatorias indicadoras.

$\tilde{X}$  tendrá la característica de que su espacio de estados será  $(\tilde{E}, \tilde{\xi})$ , donde

$$\begin{aligned}\tilde{E} &:= E \times \{0, 1\} \\ &= (E \times \{0\}) \cup (E \times \{1\})\end{aligned}$$

y  $\tilde{\xi}$  es la mínima  $\sigma$ -álgebra que contiene a los conjuntos de la forma  $A \times \{0\}$  o  $A \times \{1\}$ , con  $A \in \xi$ .

En las siguientes dos subsecciones damos construcciones que, de manera específica, generan cadenas de Markov particionantes.

### 2.4.1. Construcción de Nummelin

Por ser  $X$  fuertemente aperiódica se tiene:

$$p(x, B) \geq \delta \nu(B), \quad \forall x \in C \text{ y } B \in \xi.$$

Con lo cual podemos crear un kernel de transición:

$$Q(x, \cdot) := \frac{p(x, \cdot) - \delta \nu(\cdot)}{1 - \delta}, \quad \text{con } x \in C.$$

**Construcción 2.4.1 (Nummelin [36])** Sea  $\tilde{p}$  un kernel de transición definido en  $\tilde{E} \times \tilde{\xi}$ , donde  $\tilde{\xi}$  es la  $\sigma$ -álgebra de  $E \times \{0, 1\}$ .

$$\begin{aligned}\tilde{p}((x, 0), B \times \{0\}) &:= (1 - \delta) Q(x, B), \\ \tilde{p}((x, 1), B \times \{0\}) &:= (1 - \delta) \nu(B),\end{aligned}$$

$$\begin{aligned}\tilde{p}((x, 0), B \times \{1\}) &:= \delta Q(x, B), \\ \tilde{p}((x, 1), B \times \{1\}) &:= \delta \nu(B).\end{aligned}$$

donde  $x \in E$  y  $B \in \xi$ .

**Observación 2.4.1**

- (a) Con el kernel  $\tilde{p}$  y una medida de probabilidad  $\tilde{\lambda}$  en  $\tilde{\xi}$ , el teorema de Ionescu-Tulcea garantiza la existencia de una cadena de Markov  $\tilde{X} = \{\tilde{X}_n\}_{n=0}^{\infty}$  con espacio base  $(\tilde{\Omega}, \tilde{F}, \tilde{P}_{\tilde{\lambda}})$ , espacio de estados  $(\tilde{E}, \tilde{\xi})$ , y kernel  $\tilde{p}$ .
- (b) Se puede ver, a partir de la definición para el kernel  $\tilde{p}$ , que  $\tilde{X}_n = (X_n, I_n)$ , donde  $n = 0, 1, 2, \dots$  y  $\{I_n\}_{n=0}^{\infty}$  es una sucesión de variables aleatorias independientes con valores en 0 y 1, tal que  $\tilde{P}_{(x,a)}(I_n = 1) = \delta$  para todo  $(x, a) \in \tilde{E}$ .
- (c) Presentamos de manera gráfica la interpretación de la construcción anterior,

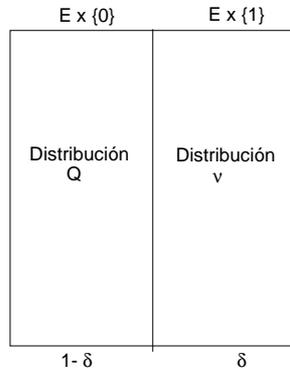


Figura 2.1: Descomposición generada por Nummelin

en donde podemos observar los siguientes hechos:

- La probabilidad de ir de  $E \times \{0\}$  a  $E \times \{1\}$  es  $\delta$  (la de ir de  $E \times \{1\}$  a  $E \times \{0\}$  es  $1 - \delta$ ).

- La distribución de  $\tilde{X}_n$  saliendo de  $E \times \{0\}$  será  $Q$ , y cuando sale de  $E \times \{1\}$  es  $\nu$ .

(d) Podemos observar que  $X$  es la marginal de  $\tilde{X}$ , debido a que se tiene el siguiente teorema.

**Teorema 2.4.1 ([33] p. 104)** Si  $\lambda$  es la distribución inicial de  $\{X_n\}$ , entonces se cumple que:

$$\int_E p^n(x, A) d\lambda(x) = \int_{E \times \{0,1\}} \tilde{p}^n((x, i), A \times \{0, 1\}) d\lambda^*(x, i),$$

donde  $\lambda^*$  es la probabilidad definida en  $(\tilde{E}, \tilde{\xi})$  por

$$\begin{aligned} \lambda^*(A \times \{0\}) &= \lambda(A \cap C)(1 - \delta) + \lambda(A \cap C^c), \\ \lambda^*(A \times \{1\}) &= \lambda(A \cap C)\delta, \end{aligned}$$

donde  $\delta$  y  $C$  son la constante y el conjunto pequeño definido por la propiedad de aperiodicidad fuerte de  $X$ .

Nótese que  $\lambda$  es la marginal de  $\lambda^*$ , es decir

$$\lambda^*(A \times \{0, 1\}) = \lambda(A),$$

$A \in \xi$ .

La demostración de este hecho se puede consultar en la referencia citada en el encabezado del teorema.

### 2.4.2. Construcción de Athreya y Ney

La construcción mostrada a continuación, fue publicada en [2] y genera realizaciones markovianas particionantes que tienen la propiedad de ser regenerativas. La idea básica de esta construcción es generar un kernel de transición que parta de puntos que se encuentran dentro del conjunto pequeño  $C$ , definido en la propiedad de aperiodicidad fuerte.

**Construcción 2.4.2 (Athreya y Ney [2])** *Bajo el mismo contexto de la construcción de Nummelin, definimos las transiciones que generan la cadena de interés.*

Sea  $B \in \xi$ .

Cuando  $x \in C$ , se tienen las siguientes transiciones:

$$\begin{aligned}\tilde{p}((x, 1), B \times \{0\}) &= (1 - \delta) \nu(C \cap B), \\ \tilde{p}((x, 1), B \times \{1\}) &= \delta \nu(C \cap B), \\ \tilde{p}((x, 0), B \times \{0\}) &= (1 - \delta) Q(x, B), \\ \tilde{p}((x, 0), B \times \{1\}) &= \delta Q(x, B).\end{aligned}$$

Cuando  $x \notin C$ ,

$$\begin{aligned}\tilde{p}((x, 1), B \times \{0\}) &= (1 - \delta) p(x, B), \\ \tilde{p}((x, 1), B \times \{1\}) &= \delta p(x, B), \\ \tilde{p}((x, 0), B \times \{0\}) &= (1 - \delta) p(x, B), \\ \tilde{p}((x, 0), B \times \{1\}) &= \delta p(x, B).\end{aligned}$$

**Observación 2.4.2** *Nótese que podemos interpretar la construcción anterior como en la siguiente figura, de la cual podemos apreciar los siguientes hechos:*

Ex{0}	Ex{1}
Distribución $\mu p^n$	Distribución $\mu p^n$
Cx{0} Distribución Q	Cx{1} Distribución v

Figura 2.2: Descomposición generada por Athreya y Ney

- La probabilidad de ir de  $E \times \{0\}$  a  $E \times \{1\}$  es  $\delta$  (la de ir de  $E \times \{1\}$  a  $E \times \{0\}$  es  $1 - \delta$ ),
- La distribución de  $\tilde{X}_n$  saliendo de  $C \times \{0\}$  será  $Q$ , cuando sale de  $(E \times \{0\}) \setminus (C \times \{0\})$  o  $(E \times \{1\}) \setminus (C \times \{1\})$  es  $\mu p^n$ , y saliendo de  $C \times \{1\}$  es  $\nu$ .

Verifiquemos ahora, que la cadena  $X$  es un proceso regenerativo, es decir, existe un tiempo aleatorio donde el proceso después de ese tiempo, comienza como nuevo independientemente de su pasado.

**Lema 2.4.1 (Propiedad de regeneración)** *Sea  $X = \{X_n\}_{n=0}^\infty$  una cadena de Markov recurrente y fuertemente aperiódica (véase definición 2.3.6). Entonces,  $X$  tiene la propiedad de regeneración, es decir, existe un tiempo aleatorio  $N$  definido en  $(\tilde{\Omega}, \tilde{F})$  tal que:  $P_x(N < \infty) = 1$  y*

$$P_x(X_{n+1} \in B, N = n) = \nu(B)P_x(N = n),$$

$x \in E$ ,  $B \in \xi$  y  $n \in \mathbb{N}$ .

**Demostración.** Definimos  $N(\omega) = \min\{n : n \in \mathbb{N}, \tilde{X}_n(\omega) \in C \times \{1\}\}$ , con  $w \in \Omega$  y  $L_i$  el  $i$ -ésimo tiempo de alcance a  $C$  (obtenido por la propiedad de recurrencia).

Sea  $x \in E$ , verifiquemos que  $P_x(N < \infty) = 1$ .

Sea  $a \in \{0, 1\}$  y  $\{L_i\}$  una sucesión de los tiempos sucesivos de alcance al conjunto  $C$  por parte de  $X$ . Entonces

$$\begin{aligned} P_x(N = \infty) &= \tilde{P}_{(x,a)}(N = \infty) \\ &= \tilde{P}_{(x,a)}(\tilde{X}_{L_i} \in C \times \{0\}, i \in \mathbb{N}) \\ &= \tilde{P}_{(x,a)}(I_{L_i} = 0, i \in \mathbb{N}) \\ &= \lim_{k \rightarrow \infty} (1 - \delta)^k \\ &= 0. \end{aligned}$$

Por lo tanto tenemos  $P_x(N < \infty) = 1$ .

Ahora hagamos notar el resultado principal. Sean  $B \in \xi$  y  $n \in \mathbb{N}$ . Entonces

$$\begin{aligned} P_x(X_{n+1} \in B, N = n) &= \tilde{P}_{(x,a)}(X_{n+1} \in B, I_n = 0, N = n) + \\ &\quad \tilde{P}_{(x,a)}(X_{n+1} \in B, I_n = 1, N = n) \\ &= \tilde{P}_{(x,a)}(\tilde{X}_{n+1} \in B \times \{0\} | N = n) \tilde{P}_{(x,a)}(N = n) + \\ &\quad \tilde{P}_{(x,a)}(\tilde{X}_{n+1} \in B \times \{1\} | N = n) \tilde{P}_{(x,a)}(N = n) \\ &= (1 - \delta) \nu(B) P_x(N = n) + \\ &\quad \delta \nu(B) \tilde{P}_{(x,a)}(N = n) \\ &= \nu(B) P_x(N = n), \end{aligned}$$

donde la tercera igualdad es debida a la propiedad de Markov. ■

**Corolario 2.4.1** *Bajo el contexto del lema anterior, existe una sucesión de tiempos aleatorios  $\{N_k\}$ , para el proceso  $\tilde{X}$ , que cumple para toda  $k \geq 1$  y  $x \in E$  que  $P_x(N_k < \infty) = 1$  y*

$$P_x(X_{n+1} \in B, N_k = n) = \nu(B)P_x(N_k = n), \quad B \in \xi \text{ y } n \in \mathbb{B}.$$

## Capítulo 3

# Acoplamiento y orden estocástico

En este capítulo, presentamos las técnicas de acoplamiento y orden estocástico. Usamos estas técnicas como herramientas para obtener tasas de convergencia de tipo geométrico (véase capítulo 6).

La técnica del acoplamiento se describe en la sección 3.1 y la del ordenamiento estocástico en la sección 3.2

En la sección de acoplamiento (en inglés *coupling*), damos el concepto y una caracterización para los casos de variables aleatorias, elementos aleatorios y procesos estocásticos, respectivamente. Posteriormente, desarrollamos la desigualdad de tiempo de acoplamiento, la cual tiene gran impacto en este trabajo. Finalmente, construimos un acoplamiento de realizaciones markovianas particionantes que cuenta con la propiedad de regeneración.

Es importante señalar que el material presentado, en esta sección, es una pequeña parte de la amplia teoría que se conoce hoy en día; el lector interesado en consultar más sobre este tema, lo puede hacer en [28] y [53].

En la sección de orden estocástico, presentamos una versión del teorema de Strassen, resultado pilar sobre el cual descansan los conceptos de monotonicidad estocástica y ordenamiento por realizaciones (véase los ejemplos 3.2.1 y 3.2.2).

El capítulo está basado en las siguientes referencias: [22], [28], [29] y [53].

### 3.1. Acoplamiento

La idea general de la técnica de acoplamiento es simple. Consiste en proporcionar una familia de variables aleatorias definidas sobre un mismo espacio base. La característica primordial de esta familia, es que cada uno de sus integrantes tiene la misma distribución a la de un miembro de otra familia de variables aleatorias (esta última familia esta dada desde un inicio y tiene las propiedades de tener la misma cardinalidad de la familia proporcionada y el espacio base de sus integrantes no es necesariamente el mismo).

Usaremos el acoplamiento para acotar la diferencia de la transición en  $n$  pasos con la distribución invariante de una cadena de Markov (véase la subsección 3.1.2.).

Para comenzar a describir este concepto, consideraremos a  $J$  como un conjunto de índices no vacío.

#### 3.1.1. Definición y ejemplos

**Definición 3.1.1** Sean  $(A_j, \mathbb{A}_j, P_j)$  espacios de probabilidad, para toda  $j \in J$  y  $\mathbb{P}$  una medida de probabilidad definida sobre  $\left( \prod_{j \in J} A_j, \prod_{j \in J} \mathbb{A}_j \right)$ . Llamaremos a  $\mathbb{P}$  **medida de acoplamiento** para la familia  $\{P_j, j \in J\}$ , si  $P_j$  es la  $j$ -ésima marginal de  $\mathbb{P}$ , es decir,

$$\mathbb{P}(y : y_j \in A) = P_j(A), \quad (3.1)$$

donde  $y \in \prod_{j \in J} A_j$  y  $A \in \mathbb{A}_j$ .

Obsérvese que se puede reescribir la igualdad (3.1) de la siguiente manera,

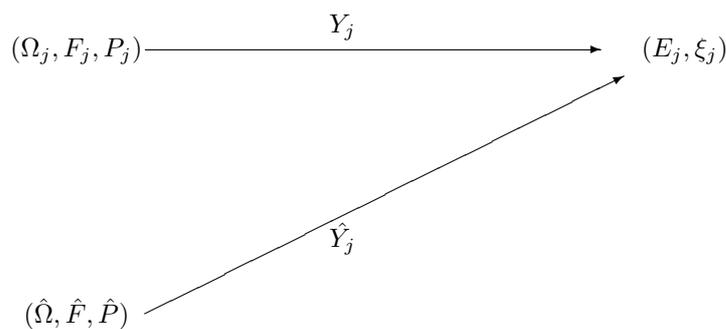
$$\mathbb{P}_{\pi_j}(A) = P_j(A),$$

donde  $\pi_j$  es la función proyección en la  $j$ -ésima coordenada y  $A \in \mathbb{A}_j$ .

**Definición 3.1.2** Considérese que para todo  $j \in J$   $(\Omega_j, F_j, P_j)$  son espacios de probabilidad y  $(E_j, \xi_j)$  son espacios medibles; también considérese a  $(\hat{\Omega}, \hat{F}, \hat{P})$  como un espacio de probabilidad y a  $(Y_j, j \in J)$  como una familia de elementos aleatorios en donde cada uno de los integrantes esta definido de  $\Omega_j$  en  $E_j$ . Una familia de elementos aleatorios  $(\hat{Y}_j, j \in J)$ ,  $\hat{Y}_j$  definido de  $\hat{\Omega}$  en  $E_j$  para todo  $j \in J$ , es llamada **Acoplamiento para la familia**  $(Y_j, j \in J)$ , si  $\hat{Y}_j$  es **copia** de  $Y_j$ , es decir:

$$Y_j \stackrel{D}{=} \hat{Y}_j, \quad j \in J.$$

**Observación 3.1.1** *En la definición 3.1.2 no se necesita que  $\hat{Y}_j$  y  $Y_j$  estén definidos en el mismo espacio dominio, pero sí en el mismo espacio codominio. Además todos los  $\hat{Y}_j$  deben estar definidos en UN MISMO espacio de probabilidad. Estos hechos son mostrados en el siguiente diagrama.*



**Diagrama de Acoplamiento**

La definición anterior tiene sentido debido a que se tiene el siguiente resultado.

**Proposición 3.1.1** *Toda familia de elementos aleatorios tiene un acoplamiento.*

**Demostración.** Sea  $Y_j$  elemento aleatorio definido de  $(\Omega_j, \mathcal{F}_j, P_j)$  en  $(E_j, \xi_j)$ ,  $j \in J$ .

Definimos:

- $(E_j, \xi_j, \mathcal{P}_j)$  un espacio de probabilidad, donde  $\mathcal{P}_j(\cdot) = P_j(Y_j \in \cdot)$ ;
- $\hat{\Omega}$  es el espacio producto de los  $E_j$ 's ;
- $\hat{\mathcal{F}}$  es la sigma-álgebra producto de los  $\xi_j$ 's; y
- $\hat{P}$  es la medida producto de los  $\mathcal{P}_j$ .

Con lo anterior, podemos construir elementos aleatorios  $\hat{Y}_j$  definidos de  $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{P})$  en  $(E_j, \xi_j)$  por  $\hat{Y}_j(\omega) = \omega_j$ , con  $\omega \in \hat{\Omega}$ .

Entonces para  $A_j \in \xi$  se tiene:

$$\begin{aligned}
 P_j(Y_j \in A_j) &= \mathcal{P}_j(A_j) \\
 &= \hat{P}(E_1 \times E_2 \times \dots \times E_{j-1} \times A_j \times E_{j+1} \times \dots) \\
 &= \hat{P}(\{\omega | \omega_j \in A_j\}) \\
 &= \hat{P}(\{\omega | \hat{Y}_j(\omega) \in A_j\}) \\
 &= \hat{P}(\hat{Y}_j \in A_j).
 \end{aligned}$$

Así tenemos que  $Y_j \stackrel{D}{=} \hat{Y}_j$  para  $j \in J$  y con este hecho mostramos que  $(\hat{Y}_j : j \in J)$  es un acoplamiento para la familia de elementos aleatorios  $\{Y_j : j \in J\}$ . ■

### Observación 3.1.2

- El acoplamiento construido en el cuerpo de la demostración anterior, lo llamaremos **acoplamiento canónico**.
- Las definiciones 3.1.1 y 3.1.2 están relacionadas estrechamente debido al acoplamiento canónico. Este hecho lo mostramos en el siguiente diagrama.

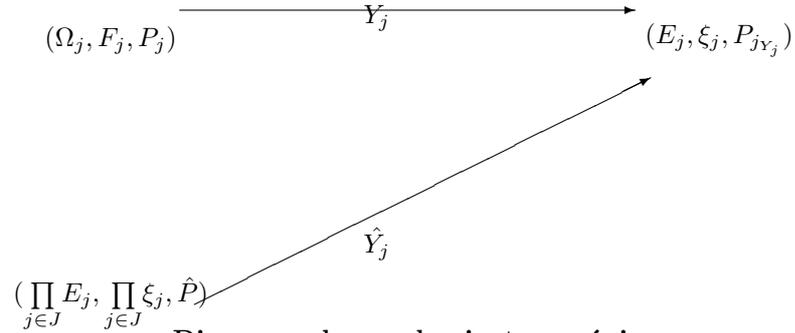


Diagrama de acoplamiento canónico

Los acoplamientos canónicos no son los únicos que se pueden dar para cierto conjunto de elementos aleatorios. A continuación mostramos otras formas de generar acoplamientos.

La manera de crear un acoplamiento para dos variables aleatorias  $X$  y  $Y$ , la damos en el siguiente lema.

**Lema 3.1.1** Sea  $U$  una variable aleatoria con distribución uniforme en  $[0, 1]$  y  $Z$  una variable aleatoria con función de distribución  $F$ . Si definimos una variable aleatoria  $\hat{Z}$  de  $(\mathbb{R}, \mathbb{B}(\mathbb{R}), P')$  en  $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$  como:

$$\hat{Z} := F^{-1} \circ U,$$

donde  $F^{-1}$  es la inversa generalizada de  $F$ . Entonces tenemos que  $Z \stackrel{D}{=} \hat{Z}$ .

**Demostración.** Sea  $z \in \mathbb{R}$ . Entonces,  $P'(\hat{Z} \leq z) = P'(F^{-1} \circ U \leq z) = P'(U \leq F(z)) = F(z)$  esto implica  $P'_{\hat{Z}}((-\infty, z]) = P_Z((-\infty, z])$ . Aplicando el teorema de clases monótonas obtenemos  $P'_{\hat{Z}} = P_Z$  ■

**Ejemplo 3.1.1 (Acoplamiento de variables aleatorias.)** Aplicando el lema anterior a cualesquiera dos variables aleatorias  $X_1$  y  $X_2$ , con distribuciones  $F_1$  y  $F_2$ ; obtenemos variables aleatorias  $\hat{X}_1$  y  $\hat{X}_2$ , con la propiedad de que:

$$X_1 \stackrel{D}{=} \hat{X}_1 \text{ y } X_2 \stackrel{D}{=} \hat{X}_2.$$

Así podemos decir que  $(\hat{X}_1, \hat{X}_2)$  es acoplamiento para  $X_1$  y  $X_2$ .

**Ejemplo 3.1.2 (Acoplamiento clásico o de Doeblin)** Sean  $X$  y  $X'$  dos realizaciones de Markov con transición  $p$  y distribuciones iniciales  $\mu$  y  $\nu$ , respectivamente. Definimos:

$$T = \inf \{n \in \mathbb{N} : X_n = X'_n\} \quad \text{y} \quad X''_n = \begin{cases} X'_n & \text{si } n < T \\ X_n & \text{si } n \geq T \end{cases}.$$

Con esto podemos ver que en el tiempo  $T$ , los procesos  $X$  y  $X'$  están en un mismo estado y continuarán su propio trayecto como si hubieran empezado desde ese estado, es decir,  $(X_n, X_{n+1}, \dots) \stackrel{D}{=} (X'_n, X'_{n+1}, \dots)$ , para  $n \geq T$ . Así la realización  $X''$ , es una modificación de la realización  $X'$  después del tiempo  $T$  y tiene la misma distribución que  $X'$ , es decir,  $X'' \stackrel{D}{=} X'$ . Con esto tenemos que  $(X, X'')$  es acoplamiento para  $X$  y  $X'$ .

**Ejemplo 3.1.3 (Acoplamiento para realizaciones markovianas)** Sea  $X = \{X_n\}_{n=0}^{\infty}$  cadena de Markov con espacio de estados  $(E, \xi)$ , espacio base  $(\Omega, \mathcal{F})$  y kernel de transición  $p$ . Sean  $\mu$  y  $\nu$  medidas de probabilidad definidas en  $(E, \xi)$ . A partir de  $p$ , podemos definir otro kernel  $\hat{p}$  en  $(E^2, \xi^2)$ , de la siguiente manera:

$$\begin{aligned} \hat{p}((x, y), A \times E) &= p(x, A), \\ \hat{p}((x, y), E \times A) &= p(y, A), \\ \hat{p}((x, y), A \times B) &= p(x, A \cap B), \end{aligned}$$

donde  $x, y \in E$  y  $A, B \in \xi$ .

Considerando la probabilidad  $\mu \times \nu$  definida en  $(E^2, \xi^2)$ , el teorema de Ionescu-Tulcea garantiza la existencia de una realización markoviana

$\hat{X} = \left( \{\hat{X}_n^1\}_{n=0}^\infty, \{\hat{X}_n^2\}_{n=0}^\infty \right)$  definida en  $((E^2)^\mathbb{N}, (\xi^2)^\mathbb{N}, P_{\mu \times \nu})$ , y con espacio de estados  $(E^2, \xi^2)$ ; donde  $\hat{X}^1 := \{\hat{X}_n^1\}_{n=0}^\infty$ ,  $\hat{X}^2 := \{\hat{X}_n^2\}_{n=0}^\infty$ , son realizaciones markovianas con kernel  $p$  y distribución inicial  $\mu$  y  $\nu$ , respectivamente.

**Definición 3.1.3 (Acoplamiento markoviano)** Sean  $\hat{X}$  y  $X$  como en el ejemplo anterior. Decimos que  $\hat{X}$  es **acoplamiento** de  $X$  si:

$$P_{\mu \times \nu}(A \times \Omega) = P_\mu(A), \quad (3.2)$$

$$P_{\mu \times \nu}(\Omega \times A) = P_\nu(A), \quad y \quad (3.3)$$

$$P_{\mu \times \nu}(\Omega^o) = 0, \quad (3.4)$$

donde  $A \in \xi^\mathbb{N}$ ,  $\Omega^o = \{\omega \in (E^2)^\mathbb{N} : \tau_\Delta < \infty \text{ y } \omega_n \notin \Delta \text{ para algún } n > \tau_\Delta\}$ ,  $\Delta = \{(x, x) : x \in E\}$  y  $\tau_\Delta = \inf\{n \in \mathbb{R} : \omega_n \in \Delta\}$ .

Denotaremos en este caso la medida  $\hat{P}$  por  $P_{\mu, \nu}$ , para resaltar la dependencia de las medidas iniciales de las trayectorias markovianas tomadas.

**Observación 3.1.3** Bajo el contexto del ejemplo anterior, si definimos un kernel de transición de la siguiente manera

$$\hat{p}((x, y), A \times B) = p(x, A)p(y, B),$$

donde  $x, y \in E$  y  $A, B \in \xi$ , y tomamos una medida inicial, obtenemos que  $\hat{X}$  es acoplamiento de  $X$ . Además suponiendo la existencia de  $\tau_\Delta$ , con la propiedad de que  $\tau_\Delta < \infty$ , entonces  $(X, X')$  es el acoplamiento descrito por Doeblin (véase [15]). En la literatura, este acoplamiento es llamado acoplamiento clásico o de Doeblin.

Encontraremos otros ejemplos de acoplamiento, en los apartados 3.1.3 y 3.2, debido a su relación con los conceptos de CMs particionantes y de orden estocástico, respectivamente.

A continuación expondremos, el uso que le daremos al acoplamiento.

### 3.1.2. Desigualdades de acoplamiento

Una aplicación que se le ha dado al acoplamiento consiste en acotar la diferencia entre dos distribuciones pertenecientes a variables aleatorias, elementos aleatorios o procesos estocásticos, respectivamente. Para mostrar este hecho damos dos desigualdades notables. La primera está relacionada con variables y elementos aleatorios, y la segunda con procesos estocásticos.

Comenzaremos con la primera de ellas.

**Desigualdad del evento de acoplamiento.**

**Definición 3.1.4** Consideremos el contexto de la definición 3.1.2. Llamaremos **evento de acoplamiento** de la familia  $(Y_i, i \in J)$  a un evento  $C \in \hat{F}$  que cumple  $\hat{Y}_i(\omega) = \hat{Y}_j(\omega)$ , para todo  $\omega \in C$  y  $i, j \in J$ .

**Lema 3.1.2 (Desigualdad del evento de acoplamiento)** Para cualesquiera dos elementos aleatorios  $Y_i, Y_j$ , se tiene:

$$\left\| P_{i_{Y_i}}(\cdot) - P_{j_{Y_j}}(\cdot) \right\|_{VT} \leq \hat{P}(C^c), \quad (3.5)$$

donde  $C$  es el evento de acoplamiento y  $\|\cdot\|_{VT}$ , es la métrica de la variación total definida por:

$$\left\| P_{i_{Y_i}}(\cdot) - P_{j_{Y_j}}(\cdot) \right\|_{VT} = 2 \sup_{A \in \xi} \left| P_{i_{Y_i}}(A) - P_{j_{Y_j}}(A) \right| \quad (3.6)$$

**Demostración.** Sea  $A \in \xi$ . Entonces,

$$\begin{aligned} P_{i_{Y_i}}(A) - P_{j_{Y_j}}(A) &= \hat{P}_{\hat{Y}_i}(A) - \hat{P}_{\hat{Y}_j}(A) \\ &= \hat{P}(\hat{Y}_i^{-1}(A)) - \hat{P}(\hat{Y}_j^{-1}(A)) \\ &= \hat{P}(\hat{Y}_i^{-1}(A) \cap C^c) - \hat{P}(\hat{Y}_j^{-1}(A) \cap C^c) + \\ &\quad \hat{P}(\hat{Y}_i^{-1}(A) \cap C) - \hat{P}(\hat{Y}_j^{-1}(A) \cap C) \\ &= \hat{P}(\hat{Y}_i^{-1}(A) \cap C^c) - \hat{P}(\hat{Y}_j^{-1}(A) \cap C^c) \\ &\leq \hat{P}(\hat{Y}_i^{-1}(A) \cap C^c) \\ &\leq \hat{P}(C^c). \end{aligned}$$

Por lo tanto, de la desigualdad anterior, y por un argumento simétrico para obtener que  $P_{j_{Y_j}}(A) - P_{i_{Y_i}}(A) \leq \hat{P}(C^c)$ , se sigue la desigualdad (3.5). ■

**Desigualdad del tiempo de acoplamiento**

Para establecer esta desigualdad necesitamos los siguientes conceptos.

**Definición 3.1.5** Sean  $X = \{X_n\}_{n=0}^{\infty}$  y  $Y = \{Y_n\}_{n=0}^{\infty}$  dos procesos estocásticos definidos de  $(\Omega_i, F_i, P_i)$  en  $(E_i, \xi_i)$ ,  $i = 1, 2$ , y  $\hat{Z} = \left( \{\hat{X}_n\}_{n=0}^{\infty}, \{\hat{Y}_n\}_{n=0}^{\infty} \right)$  un acoplamiento para  $\{X_n\}_{n=0}^{\infty}$  y  $\{Y_n\}_{n=0}^{\infty}$ .

(a) Llamaremos **tiempo de acoplamiento**, para los procesos  $\hat{X}$  y  $\hat{Y}$ , a una variable aleatoria  $T$  definida en  $(\hat{\Omega}, \hat{F}, \hat{P})$  por:

$$T = \inf \left\{ n \in \mathbb{N} : \hat{X}_n = \hat{Y}_n \right\},$$

tal que

$$\hat{X}_n(\omega) = \hat{Y}_n(\omega), \quad n \geq T, \quad \omega \in \{T < \infty\}.$$

(b) El proceso  $\hat{Z}$  es llamado **acoplamiento exacto**, si tiene tiempo de acoplamiento.

(c)  $\hat{Z}$  es llamado **acoplamiento exitoso**, si su tiempo de acoplamiento cumple  $\hat{P}(T < \infty) = 1$ .

**Lema 3.1.3 (Desigualdad del tiempo de Acoplamiento)** Sean  $\{X_n^1\}_{n=0}^\infty$  y  $\{X_n^2\}_{n=0}^\infty$  realizaciones markovianas cuyo kernel de transición es  $p$  y sus medidas iniciales son  $\mu$  y  $\nu$ , respectivamente. Si  $\left( \left\{ \hat{X}_n^1 \right\}_{n=0}^\infty, \left\{ \hat{X}_n^2 \right\}_{n=0}^\infty \right)$  es un acoplamiento exacto para  $\{X_n^1\}_{n=0}^\infty$  y  $\{X_n^2\}_{n=0}^\infty$ , entonces se tiene que:

$$\begin{aligned} \|\mu p^n(\cdot) - \nu p^n(\cdot)\|_{VT} &\leq \hat{P}(\hat{X}_n^1 \neq \hat{X}_n^2) \\ &\leq \hat{P}(T > n), \end{aligned} \quad (3.7)$$

donde  $\lambda p^n(A) := \int_E p^n(x, A) d\lambda(x) = \int_E P_x(X_n \in A) d\lambda(x) = P_\lambda(X_n \in A)$ ,  $A \in \xi$ ,  $\lambda$  es una probabilidad definida en  $\xi$  y  $\|\cdot\|_{VT}$  es la métrica de la variación total definida previamente en la ecuación 3.6.

**Demostración.** Sea  $A \in \xi$

$$\begin{aligned} \mu p^n(A) - \nu p^n(A) &= P_\mu(X_n^1 \in A) - P_\nu(X_n^2 \in A) \\ &= \hat{P}(\hat{X}_n^1 \in A) - \hat{P}(\hat{X}_n^2 \in A) \\ &\leq \hat{P}(\hat{X}_n^1 \in A, \hat{X}_n^1 \neq \hat{X}_n^2) - \hat{P}(\hat{X}_n^2 \in A, \hat{X}_n^1 \neq \hat{X}_n^2) \\ &\quad + \hat{P}(\hat{X}_n^1 \in A, \hat{X}_n^1 = \hat{X}_n^2) - \hat{P}(\hat{X}_n^2 \in A, \hat{X}_n^1 = \hat{X}_n^2) \\ &\leq \hat{P}(\hat{X}_n^1 \in A, \hat{X}_n^1 \neq \hat{X}_n^2) - \hat{P}(\hat{X}_n^2 \in A, \hat{X}_n^1 \neq \hat{X}_n^2) \\ &\leq \hat{P}(\hat{X}_n^1 \in A, \hat{X}_n^1 \neq \hat{X}_n^2) \\ &\leq \hat{P}(\hat{X}_n^1 \neq \hat{X}_n^2) \\ &\leq \hat{P}(T > n), \end{aligned}$$

la última desigualdad, se obtiene por la definición de  $T$ . De la desigualdad anterior y por simetría se obtiene la desigualdad (3.7). ■

### 3.1.3. Acoplamientos particionantes

Finalizamos esta sección con la construcción de *acoplamientos exitosos* para realizaciones markovianas que son *particionantes* (véase secciones 2.4 y 3.1.2).

Para la construcción, consideremos la siguiente hipótesis.

**Hipótesis 3.1.1** Sean  $\{X_n\}_{n=0}^\infty$  y  $\{Y_n\}_{n=0}^\infty$  dos realizaciones markovianas con distribuciones iniciales  $\mu$  y  $\pi$ , respectivamente y cuyo kernel de transición es  $p$ . Además supóngase que son recurrentes y fuertemente aperiódicas.

**Observaciones 3.1.1** Debido a que  $p$  es fuertemente aperiódica, podemos considerar  $\delta$ ,  $\nu$ ,  $C$  y  $Q$  como fueron tomadas en la subsección 2.4.

A continuación damos la construcción mencionada, mediante el siguiente algoritmo expresado en pseudocódigo.

Sea  $\{I_n\}_{n=0}^\infty$  una sucesión de variables aleatorias independientes que toman valores en  $\{0, 1\}$ , con probabilidad  $\delta$  de que  $I_n = 1$ , para todo  $n$  (consúltese el contexto de esta sucesión en la sección 2.4).

**Algoritmo 3.1.1** ([47] pág. 401) Consideremos las hipótesis 3.1.1 y definamos:

$$\hat{X}_0 \stackrel{D}{=} \mu, \hat{Y}_0 \stackrel{D}{=} \pi, n = 0, \text{Indicador} = 0 \text{ y } \omega \in \Omega$$

**Mientras** ( $\text{Indicador} = 0$ )

{

**Si** ( $X_n(\omega) \in C, Y_n(\omega) \in C$ )

**Si** ( $I_n(\omega) = 1$ )

$$\begin{aligned} \hat{X}_k(\omega) &= \hat{Y}_k(\omega), \quad \forall k \geq n+1, \\ \text{Indicador} &= 1. \end{aligned}$$

**Otro.**

$$\begin{aligned} \hat{X}_{n+1} &\stackrel{D}{=} Q(X_n(\omega), \cdot), \\ \hat{Y}_{n+1} &\stackrel{D}{=} Q(Y_n(\omega), \cdot) \text{ y} \\ n &:= n+1. \end{aligned}$$

*Otro*

$$\begin{aligned}\hat{X}_{n+1} &\stackrel{D}{=} p(X_n(\omega), \cdot), \\ \hat{Y}_{n+1} &\stackrel{D}{=} p(Y_n(\omega), \cdot) \text{ y} \\ n &:= n + 1\end{aligned}$$

}

#### Observación 3.1.4

- Con este algoritmo se encuentra una cadena de Markov  $(\hat{X}, \hat{Y})$ , donde a la vez  $\hat{X}$  y  $\hat{Y}$  son cadenas de Markov con distribuciones iniciales  $\mu$  y  $\pi$ , respectivamente.
- Cuando ocurre en el algoritmo  $I_n(\omega) = 1$ , se tiene como consecuencia la siguiente interpretación: con probabilidad  $\delta$ , se tiene  $\hat{X}_k = \hat{Y}_k$  para todo  $k \geq n + 1$ , y con probabilidad  $1 - \delta$  los procesos aún no se han acoplado.
- La construcción realizada, cumple con la desigualdad de acoplamiento:

$$\|P_\mu(\hat{X}_n \in \cdot) - \pi(\cdot)\|_{VT} \leq \hat{P}(T > n), \quad n \in \mathbb{N}.$$

- Cuando el espacio de estados es igual al conjunto pequeño (es decir,  $E = C$ ), se tiene que:

$$\hat{P}(T > n) \leq (1 - \delta)^n \quad n \in \mathbb{N}.$$

En general, cuando no se tiene que  $E = C$ , se tendrá que  $\hat{P}(T > n)$  está acotado por la suma de  $(1 - \delta)^n$  y un término adicional. Estos resultados serán revisados en el capítulo 5.

## 3.2. Orden estocástico

El concepto de orden estocástico se creó en la década de los 70's y en este trabajo lo utilizamos para ordenar acoplamientos.

Esta sección muestra, a través del teorema de Strassen<sup>1</sup>, el ordenamiento de dichos acoplamientos.

Se ha organizado esta sección en tres partes. En cada una de ellas se muestra el teorema de Strassen para variables aleatorias, elementos aleatorios y procesos estocásticos, respectivamente.

<sup>1</sup>Lo que aquí llamamos teorema de Strassen, es una consecuencia del resultado 11 obtenido por Strassen en [52] (ver apéndice B).

Para mostrar estos resultados, daremos a continuación el concepto de orden estocástico y su caracterización. Para este fin, consideramos un espacio medible  $(E, \xi)$ , dotado de un orden parcial  $\preceq$  en  $E$ .

**Definición 3.2.1 (Orden estocástico)** Sean  $\lambda$  y  $\mu$  probabilidades en  $(E, \xi)$  y  $X, X'$  elementos aleatorios definidos de  $(\Omega_i, F_i, P_i)$  en  $(E, \xi)$  con  $i = \{1, 2\}$ , respectivamente.

a) Decimos que  $\lambda$  es estocásticamente menor o igual a  $\mu$  ( $\lambda \stackrel{est.}{\leq} \mu$ ), si y solo si:

$$\lambda(A) \leq \mu(A), \quad A \in B,$$

donde  $B = \{A \in \xi : I_A(\cdot) \text{ es una función no decreciente}\}$ .

b) Decimos que  $X$  es estocásticamente menor o igual que  $X'$ , ( $X \stackrel{est.}{\leq} X'$ ) si  $\lambda \stackrel{est.}{\leq} \mu$ , donde:

$$\begin{aligned} \lambda &= (P_1)_X, \quad y \\ \mu &= (P_2)_{X'}. \end{aligned}$$

**Lema 3.2.1 (Caracterización de orden estocástico)** Sean  $P$  y  $P'$  dos probabilidades definidas en  $(E, \xi)$ . Entonces  $P \stackrel{est.}{\leq} P'$  si y solo si, para toda  $f \in A_\xi$  se cumple:

$$\int_E f dP \leq \int_E f dP',$$

donde

$A_\xi := \{f \mid f \text{ es una función medible de } \xi \text{ en } B(\mathbb{R}), \text{ no decreciente y acotada}\}$

**Demostración.**

“ $\Leftarrow$ ”

Tomando  $f = I_A$ ,  $A \in B$ , se obtiene  $P \stackrel{est.}{\leq} P'$ .

“ $\Rightarrow$ ”

Obsérvese que  $\int f dP \leq \int f dP'$ , es válida usando funciones indicadoras. Posteriormente considerando funciones simples, no negativas e integrables, se obtiene el resultado requerido. ■

**Observación 3.2.1** Cuando las medidas  $\mu$  y  $\lambda$  son distribuciones en  $(\mathbb{R}, \mathbb{B}(\mathbb{R}))$ , la definición 3.2.1 queda descrita de la siguiente manera:

decimos que  $\lambda$  es estocásticamente menor o igual que  $\mu$  ( $\lambda \stackrel{est.}{\leq} \mu$ ), si:

$$\lambda((x, \infty)) \leq \mu((x, \infty)), \quad \text{para todo } x \in \mathbb{R}.$$

A continuación mostramos el teorema de Strassen en su versión para variables aleatorias, elementos aleatorios y procesos estocásticos, respectivamente.

### 3.2.1. Acoplamiento ordenado en variables aleatorias

**Teorema 3.2.1 (Strassen, [28])** Sean  $X$  y  $X'$  variables aleatorias definidas en

$(\Omega, \mathcal{F}, P)$ . Entonces  $X \stackrel{\text{est.}}{\leq} X'$  si y solo si existe un acoplamiento  $(\hat{X}, \hat{X}')$  de  $X$  y  $X'$ , ordenado de manera puntual, es decir,  $\hat{X}(\omega) \leq \hat{X}'(\omega)$ , para todo  $\omega \in \Omega$ .

**Demostración.**

“ $\implies$ ”

Sean  $F$  y  $F'$  las funciones de distribución para  $X$  y  $X'$ , respectivamente. Definiendo  $(\hat{X}, \hat{X}')$  como en el ejemplo 3.1.1, tenemos que  $(\hat{X}, \hat{X}')$  es acoplamiento de  $X$  y  $X'$ , donde

$$\begin{aligned}\hat{X} &:= F^{-1} \circ U, \\ \hat{X}' &:= F'^{-1} \circ U,\end{aligned}$$

y  $U$  es una variable aleatoria uniforme en  $[0, 1]$ .

De la hipótesis, podemos deducir que  $F \geq F'$ . Aplicando este hecho obtenemos la siguiente desigualdad para las funciones inversa generalizadas de  $F$  y  $F'$ :

$$F'^{-1} \geq F^{-1}.$$

Por lo tanto, haciendo la composición con  $U$  en cada lado, concluimos:

$$\hat{X} \leq \hat{X}'.$$

“ $\impliedby$ ”

Sea  $(\hat{X}, \hat{X}')$  acoplamiento ordenado de manera puntual, para las variables aleatorias  $X$  y  $X'$  definidas en  $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{P})$ .

Por lo tanto tenemos para todo  $y \in \mathbb{R}$ :

$$\hat{P}(\hat{X} > y) \leq \hat{P}(\hat{X}' > y) \quad \Leftrightarrow \quad P(X > y) \leq P(X' > y),$$

es decir  $X \stackrel{\text{est.}}{\leq} X'$ . ■

**Lema 3.2.2 (Caracterización de Orden estocástico)** Sean  $X$  y  $X'$  variables aleatorias definidas en  $(\Omega, \mathcal{F}, P)$ . La siguientes afirmaciones son equivalentes:

(a)  $X \stackrel{est.}{\leq} X'$ .

(b) Para toda función  $f : \mathbb{R} \rightarrow \mathbb{R}$ , no decreciente y acotada, se tiene:

$$\int_{\mathbb{R}} f dP_X \leq \int_{\mathbb{R}} f dP_{X'}.$$

**Demostración.**

“ $a \implies b$ ”

Por el teorema anterior, existe un acoplamiento  $(\widehat{X}, \widehat{X}')$  de  $X$  y  $X'$  definido en  $(\mathbb{R}, \mathbb{B}(\mathbb{R}), \widehat{P})$ , tal que  $\widehat{X} \leq \widehat{X}'$ , lo cual implica que  $f(\widehat{X}) \leq f(\widehat{X}')$ , para toda función  $f$  no decreciente. Ahora

$$\begin{aligned} \int_{\mathbb{R}} f dP_X &= \int_{\mathbb{R}} f d\widehat{P}_{\widehat{X}} \\ &= \int_{\mathbb{R}} f \circ \widehat{X} d\widehat{P} && \text{(Por el teorema de cambio de variable)} \\ &\leq \int_{\mathbb{R}} f \circ \widehat{X}' d\widehat{P} \\ &= \int_{\mathbb{R}} f d\widehat{P}_{\widehat{X}'} \\ &= \int_{\mathbb{R}} f dP_{X'}. \end{aligned}$$

“ $b \longleftarrow a$ ”

Con  $f = I_{(x, \infty)}$ , se tiene lo requerido. ■

### 3.2.2. Acoplamiento ordenado en elementos aleatorios y procesos estocásticos

Los resultados que presentamos en esta sección, son versiones del teorema de Strassen para elementos aleatorios y procesos estocásticos, respectivamente. Debido a la extensión en sus demostraciones, referimos éstas al apéndice B del presente trabajo. También establecemos en esta sección, un ejemplo obtenido del teorema de Strassen, versión procesos estocásticos, el cual será utilizado en el capítulo 6.

A continuación enunciamos el teorema de Strassen para elementos aleatorios.

**Teorema 3.2.2 (Strassen, [28])** Sean  $P$  y  $P'$  medidas de probabilidad en  $(E, \xi)$ .

Entonces  $P \stackrel{est.}{\leq} P'$  si y solo si, existen elementos aleatorios  $Z$  y  $Z'$  con espacio imagen  $(E, \xi)$ , además  $Z \stackrel{D}{=} P$ ,  $Z \stackrel{D}{=} P'$  y  $Z \leq Z'$ .

Para enunciar el resultado de Strassen para procesos estocásticos, necesitamos antes una definición.

**Definición 3.2.2** Sean  $E_i$  espacios con un orden parcial  $\preceq_i$ ,  $i = 1, 2$ ,  $K$  y  $K'$  kérneles de transición en  $E_1 \times E_2$ . Decimos que  $K$  es **dominado estocásticamente** por  $K'$  si:

$$K(x, \cdot) \stackrel{est.}{\preceq}_2 K'(x', \cdot) \text{ para todo } x \preceq_1 x'.$$

**Teorema 3.2.3 (Strassen,[28])** Sean  $(E_i, \xi_i)$  espacios medibles dotados con un orden parcial  $\preceq_i$ ,  $i \in \mathbb{N}$ ;  $\lambda$  y  $\mu$  medidas de probabilidad en  $E$ ,  $K_i$ ,  $K'_i$  kérneles de transición en  $\prod_{j=0}^{i-1} E_j \times \xi_i$ ,  $i \in \mathbb{N}$ .

Si  $\lambda \stackrel{est.}{\leq} \mu$  y  $K_i \stackrel{est.}{\leq} K'_i$  para toda  $i$  entonces existen procesos estocásticos  $Z = \{Z_n\}_{n=0}^\infty$  y  $Z' = \{Z'_n\}_{n=0}^\infty$  en  $(\prod_{i=0}^\infty E_i, \prod_{i=0}^\infty \xi_i)$  tales que:

$$(a) (Z_0, \dots, Z_n) \stackrel{D}{=} P_\lambda^n, (Z'_0, \dots, Z'_n) \stackrel{D}{=} P_\mu^n, \text{ para toda } n.$$

$$(b) Z_n \preceq_n Z'_n, \text{ para toda } n.$$

donde  $P_\lambda^n$  y  $P_\mu^n$  son las medidas de probabilidad definidas en  $\prod_{i=0}^n \xi_i$ , obtenidas por el teorema de la medida producto, cuando las medidas iniciales son  $\lambda$  y  $\mu$ , respectivamente.

**Demostración.** La prueba completa puede consultarse en el apéndice B.

Debido a la riqueza de ideas desarrolladas en la demostración y a las conclusiones importantes que podemos sacar de éstas, presentamos a continuación un esbozo de la prueba.

A partir de que  $\lambda \stackrel{est.}{\leq} \mu$  y que  $k_i \stackrel{est.}{\leq} k'_i$ , para toda  $i \in \mathbb{N}$ ; se tiene,

$$P_\lambda^n \stackrel{est.}{\leq} P_\mu^n,$$

para toda  $n \in \mathbb{N}$ .  $P_\lambda^n$  y  $P_\mu^n$  son las probabilidades mencionadas en la construcción de Ionescu-Tulcea.

Ahora, del hecho  $P_\lambda^n \stackrel{est.}{\leq} P_\mu^n$ , para toda  $n \in \mathbb{N}$ , se tiene,

$$\hat{P} \stackrel{est.}{\leq} \hat{P}',$$

donde  $\hat{P}$  y  $\hat{P}'$  son probabilidades definidas en  $(\prod_{i=0}^\infty E_i, \prod_{i=0}^\infty \xi_i)$ , tales que:  $\hat{P}_{\pi_n} = P_\lambda^n$ ,  $\hat{P}'_{\pi_n} = P_\mu^n$ , donde  $\pi_n$  es la proyección de  $\prod_{i=0}^\infty E_i$  a  $\prod_{i=0}^{n-1} E_i$ .

De este último hecho se concluye que,

$$Z_n \preceq Z'_n, \quad n \geq 0.$$

■

**Ejemplo 3.2.1 (Cadena de Markov estocásticamente monótona)**

Considerando el contexto y las hipótesis del Teorema 3.2.3. Llamaremos cadena de Markov **estocásticamente monótona** a la cadena  $X = \{X_n\}$  que cuenta con **kérel de transición monótono**  $p$ , es decir,  $K_i = K'_i = p$  para toda  $i \geq 0$ .

A continuación mostramos dos propiedades que tienen este tipo de cadenas.

**Corolario 3.2.1** Sea  $X$  una cadena estocásticamente monótona y  $X'$  y  $X''$  dos de sus realizaciones, las cuales tienen distribuciones iniciales  $\lambda$  y  $\mu$ , respectivamente, con  $\lambda \stackrel{est.}{\leq} \mu$ . Entonces se cumple lo siguiente,

$$(i) P_\lambda(X'_n \in \cdot) \stackrel{est.}{\leq} P_\mu(X''_n \in \cdot),$$

$$(ii) X'_n \leq X''_n, \text{ para todo } n \geq 0.$$

**Demostración.** Véase la demostración del Teorema 3.2.3. ■

**Observaciones 3.2.1** Cuando  $X$  cumple la condición (i) del Corolario 3.2.1, los autores de [30] llaman a esta cadena de Markov **ordenada estocásticamente**, y cuando satisface la condición (ii) la nombran cadena de Markov **ordenada por realizaciones**.

**Ejemplo 3.2.2 (Cadena de Markov ordenada por realizaciones)** Este tipo cadenas resultan directamente del Teorema 3.2.3, como se mencionó en la observación 3.2.1.

Mostramos a continuación dos propiedades importantes que también se desprenden del Teorema 3.2.3 y que serán utilizados en el capítulo 7.

**Corolario 3.2.2** Sea  $X$  una cadena de Markov estocásticamente monótona,  $X'$  y  $X''$  dos de sus realizaciones. Si las distribuciones iniciales de  $X'$  y  $X''$  son:

1.  $\delta_x$  y  $\delta_{x'}$ , respectivamente con  $x \leq x'$ . Entonces

$$X'_n \leq X''_n,$$

para toda  $n \geq 0$ .

2.  $\delta_a$  y  $\mu$ , respectivamente, donde  $a \preceq x$  para todo  $x \in E$ . Entonces

$$X'_n \leq X''_n,$$

para toda  $n \geq 0$ .

**Demostración.** Aplicando el Lema 3.2.3 en el Teorema 3.2.3, se obtiene los resultados requeridos. ■

**Lema 3.2.3** Sea  $(E, \preceq)$ , un espacio de estados parcialmente ordenado.

(a) Para  $x \preceq x'$ , se cumple lo siguiente:

$$\delta_x \stackrel{est}{\preceq} \delta_{x'}.$$

(b) Si existe un elemento mínimo  $a \in E$ , es decir,  $a \preceq x, \forall x \in E$ . Entonces

$$\delta_a \stackrel{est.}{\preceq} \mu, \text{ para toda } \mu \text{ y todo } x \in E.$$

**Demostración.**

Consideremos  $A \in B$  (véase el inciso “a” de la definición 3.2.1).

“a” Vamos a mostrar que  $\delta_x(A) \leq \delta_{x'}(A)$ . Para esto, tomemos  $y \in \mathbb{R}$  y analicemos los siguientes casos:

- (i) Si  $x \in A$ , entonces  $x' \in A$ , debido a que  $I_A$  es no decreciente. Por lo tanto  $\delta_x(A) \leq \delta_{x'}(A)$ .
- (ii) Si  $x \notin A$  y  $x' \in A$ , ó  $x \notin A, x' \notin A$ , claramente obtenemos:  $\delta_x(A) \leq \delta_{x'}(A)$ .

Con esto podemos concluir que  $\delta_x \stackrel{est.}{\leq} \delta_{x'}$

“b” Como  $a \preceq x$  entonces se cumple  $\delta_a \stackrel{est.}{\leq} \delta_x$ , con lo cual se da la siguiente desigualdad:

$$\begin{aligned} \mu(A) &= P_\mu(X_0''' \in A) \\ &= \int_E P_x(X_0 \in A) d\mu(x) \\ &= \int_E \delta_x(A) d\mu(x) \\ &\geq \int_E \delta_a(A) d\mu(x) \\ &= \delta_a(A). \end{aligned}$$

Por lo tanto tenemos  $\delta_a \stackrel{est.}{\leq} \mu$  ■

## Parte II

### Planteamiento del problema



# Capítulo 4

## Planteamiento y antecedentes

### 4.1. Planteamiento

Es conocido, en la literatura de CMs, que bajo ciertas condiciones, se tiene

$$\lim_{n \rightarrow \infty} \|p^n(x, \cdot) - \pi(\cdot)\| = 0,$$

donde  $p^n$  denota la función de transición en  $n$ -pasos,  $\pi$  la distribución invariante, y  $\|\cdot\|$  es una métrica apropiada (por ejemplo, ésta puede ser la métrica de la variación total o de la variación ponderada) (véase [2], [7], [20], [23] y [37]).

Una pregunta importante relacionada con este límite es **¿qué tan rápido tiende a cero?**, es decir, es importante conocer la velocidad de convergencia de las transiciones a la medida invariante; la cual llamaremos desde este momento, **tasa de convergencia**.

Matemáticamente hablando, en el contexto de cadenas de Markov, entenderemos por **tasa de convergencia** a una función  $\phi$  que cumple para toda  $n \in \mathbb{N}$  y  $x \in E$ ,

$$\|p^n(x, \cdot) - \pi(\cdot)\| = O(\phi(n)), \quad \text{o} \quad (4.1)$$

$$\|p^n(x, \cdot) - \pi(\cdot)\| = o(\phi(n)), \quad (4.2)$$

donde  $\|p^n(x, \cdot) - \pi(\cdot)\| = O(\phi(n))$  significa que existe  $M_x > 0$  tal que

$$\|p^n(x, \cdot) - \pi(\cdot)\| \leq M_x |\phi(n)| \quad \text{y}$$

por  $\|p^n(x, \cdot) - \pi(\cdot)\| = o(\phi(n))$  entenderemos que

$$\lim_{n \rightarrow \infty} \frac{\|p^n(x, \cdot) - \pi(\cdot)\|}{|\phi(n)|} = 0.$$

En el contexto del párrafo anterior, diremos que  $\phi$  es:

- (a) **ergódica** si ocurre  $\|p^n(x, \cdot) - \pi(\cdot)\| = O(\phi(n))$ ,
- (b) **uniformemente ergódica**, si existe  $M > 0$  tal que

$$\|p^n(x, \cdot) - \pi(\cdot)\| \leq M|\phi(n)|, \quad x \in E.$$

Si  $\phi$  es ergódica ( o uniformemente ergódica) y tiene la siguiente forma  $\phi(n) = r^n$ , con  $0 < r < 1$  la tasa es llamada **geoméricamente ergódica** ( o uniforme y geoméricamente ergódica).

Cabe mencionar que hay trabajos en la literatura de cadenas de Markov en las cuales se garantiza la **existencia** de tasas de convergencia, pero **no** se proveen éstas de manera explícita (véase [27],[38],[55] y [57]).

La tasa de convergencia tiene aplicación, principalmente, en los algoritmos de simulaciones estocásticas que se hacen para ciertos modelos que pueden ser descritos como cadenas de Markov, tales como: MCMC, PDMs, etc (véase [44], [48] y [35]). De aquí, la importancia para su estudio.

A continuación describiremos el problema relacionado con las tasas de convergencia que estudiaremos en el presente trabajo.

**PROBLEMA.** *Considérese cadenas de Markov con los siguientes requerimientos:*

- a) *tienen espacios de estados contenidos en  $\mathbb{R}$  ; y*
- b) *cuentan con una medida invariante. Además, se supondrá que la función de transición en  $n$ -pasos converge a la medida invariante (en alguna métrica adecuada), y que para dicha convergencia **existe** una tasa.*

*En este contexto, el problema de la tesis consiste en:*

*Analizar distintas condiciones que permitan expresar las tasas de convergencia de **forma explícita y de tipo geométrico**. Dicho análisis se realizará para cada uno de los siguientes casos:*

- *CMs con espacios finitos o compactos,*

- *CMs con kernel estocásticamente monótono, o*
- *CMs con la propiedad de reversibilidad y/o positividad.*

*Además, buscaremos ejemplificar la teoría desarrollada en cada uno de los casos anteriores.*

## 4.2. Antecedentes

A finales de los 50's e inicio de la década de los 60's, se realizaron trabajos relacionados con la *existencia* de tasas de convergencia en espacio de estados discretos (véase [27] y [57]). Posteriormente, hasta la década de los 70's, se publicaron trabajos relacionados con la existencia de tasas de convergencia para cadenas de Markov con espacio de estados generales (véase [55] y [38]).

Aquellos trabajos, basaban sus resultados básicamente en el uso de las condiciones de minorización, Foster-Lyapunov y  $V$ -uniformidad. Como veremos a continuación, dichos criterios están fuertemente relacionados.

Para obtener la condición de Foster-Lyapunov se necesita el establecimiento de la condición de minorización. Popov en [42] y Tweedie en [55], probaron, para el caso de espacio de estados numerable y general, respectivamente, que la propiedad de Foster-Lyapunov es la condición necesaria para la obtención de tasas de convergencia de tipo geométrico.

Hordijk y Spieksma en [19], Meyn y Tweedie en [54], mostraron respectivamente, que la clase de funciones que satisfacen la condición de Foster-Lyapunov es equivalente a la clase de funciones que hacen que una cadena de Markov sea ergódica y  $V$ -uniforme.

Por otro lado, a partir de la década de los 80's, se comenzaron a desarrollar resultados que permitieron obtener de forma explícita, tasas de convergencia ([4], [25], [24], [26] [30], [34], [46] y [48]).

Estos nuevos resultados, tuvieron como base para su desarrollo, las mismas condiciones que permitieron conocer la existencia de tasas, como a su vez, la incorporación de técnicas probabilísticas. Las técnicas fueron; teoría de renovación, acoplamientos bivariados, monotonicidad estocástica y el teorema de Kendall.

De manera cronológica, estos resultados se presentaron en [25], [24], [26], [34], [48], [30] [46] y [4].

Esencialmente hoy en día, existen diversas clasificaciones para caracterizar a las

tasas de convergencia. En este trabajo nos concentraremos en tasas de convergencia ergódicas, no ergódica, y de orden de convergencia geométrico.

## Parte III

### Tasas de convergencia



## Capítulo 5

# Método espectral

El resultado presentado en este capítulo fue consultado en [47], y en él se observa una tasa de convergencia de tipo geométrico con respecto a la métrica de la variación total para cadenas de Markov con espacio de estado  $E$  finito.

Para este caso, la tasa de convergencia está relacionada con los valores y vectores propios de la matriz de transición.

A continuación proporcionamos el contexto, y un resultado previo que nos servirá para presentar el resultado principal del capítulo.

Sean  $X$  una CM con  $n$  estados y  $\lambda_0, \lambda_1, \dots$  y  $\lambda_{n-1}$  los valores propios asociados a la matriz de transición  $p$  (incluidos con multiplicidad algebraica). Además

$$\begin{aligned}\lambda_0 &= 1 \quad \text{y} \\ \lambda^* &= \max_{1 \leq j \leq n-1} |\lambda_j|,\end{aligned}$$

Nótese que el valor propio  $\lambda_0 = 1$  tiene sentido, porque toda matriz estocástica tiene como valor propio a 1 (véase [47]). Denotemos por  $v_0$ , a su vector propio.

### Lema 5.0.1

- (a)  $\lambda^* \leq 1$ ; y
- (b) Si  $p(x, y) > 0$  para todo  $x, y \in E$ , se tiene  $\lambda^* < 1$ .

### Demostración.

- (a)

Sea  $\lambda$  un valor propio asociado al vector  $v$ . Escogemos  $x \in E$  de tal manera que se cumpla,

$$|v(y)| \leq |v(x)|, \quad \text{para todo } y \in E.$$

De este hecho podemos obtener

$$\begin{aligned} |\lambda v(x)| &= |pv(x)| \\ &= \left| \sum_{y \in E} v(y)p(x, y) \right| \\ &\leq \sum_{y \in E} |v(y)|p(x, y) \\ &\leq \sum_{y \in E} |v(x)|p(x, y) \\ &= |v(x)|, \end{aligned}$$

es decir,  $|\lambda v(x)| \leq |v(x)|$ . Con lo cual se tiene que  $|\lambda| \leq 1$ , y así concluimos que  $|\lambda^*| \leq 1$ .

(b)

Supóngase que  $pv = \lambda v$ , de aquí si  $|v(x)| = |v(y)|$  para todo  $x, y \in E$ , se debe tener que  $v(x) = v(y)$  para  $x, y \in E$ . Con esto se observa que  $v$  es un vector constante, con lo cual se nota que su único valor propio asociado es  $\lambda_0 = 1$ . De aquí se obtiene que  $\lambda^* = 0 < 1$ , y que  $p$  es diagonalizable.

Si  $p$  no es diagonalizable, se necesita probar que el valor propio  $\lambda_0 = 1$  no forma parte del bloque de Jordan más grande. Probemos este hecho por contradicción.

Supongamos que  $\lambda_0$  pertenece al bloque de Jordan más grande, entonces para algún vector  $v$  se tiene

$$pv = v + u,$$

donde  $u = (1, 1, \dots, 1)^t$ . Por lo tanto podemos escoger  $x \in E$  de tal manera que

$$\operatorname{Re} v(y) \leq \operatorname{Re} v(x), \quad \text{para todo } y \in E.$$

Con este hecho obtenemos:

$$\begin{aligned} 1 + \operatorname{Re} v(x) &= \operatorname{Re} (pv(x)) \\ &= \operatorname{Re} \sum_{y \in E} v(y)p(x, y) \\ &\leq \sum_{y \in E} |v(y)|p(x, y) \\ &\leq \operatorname{Re} \sum_{y \in E} v(x)p(x, y) \\ &= \operatorname{Re} v(x), \end{aligned}$$

lo cual resulta una contradicción. Así,  $\lambda_0$  no pertenece al bloque de Jordan más grande, con lo cual se tiene que  $\lambda^* < 1$ . ■

Con lo anterior, el resultado principal del capítulo se enuncia de la siguiente manera.

**Teorema 5.0.1** *Sea  $\{X_n\}_{n=0}^\infty$  una realización de Markov finita con matriz de transición  $p$ , la cual es diagonalizable, y tiene distribución inicial  $\mu_0$ . Supóngase que  $p(x, y) > 0$  para todo  $x, y \in E$  (obsérvese que en este caso,  $\lambda^* < 1$  debido al lema anterior). Entonces se cumple lo siguiente.*

- a) *Existe una única distribución invariante  $\pi$ .*  
 b) *Para cada  $x \in E$ , existe una constante  $C_x > 0$  tal que:*

$$|P_{\mu_0}(X_k = x) - \pi(x)| \leq C_x(\lambda^*)^k,$$

donde  $C$  es

$$C_x = \sum_{m=1}^{n-1} |a_m v_m(x)|,$$

y

- $v_0, v_1, \dots, v_{n-1}$  forman la base de vectores propios izquierdos correspondientes a  $\lambda_0, \lambda_1, \dots, \lambda_{n-1}$ , respectivamente; y
- $a_0, a_1, \dots, a_{n-1}$  son los escalares de la combinación lineal que generan a  $\mu_0$ , es decir

$$\mu_0 = a_0 v_0 + a_1 v_1 + \dots + a_{n-1} v_{n-1}. \quad (5.1)$$

### Demostración.

Nótese que para cierto vector propio  $v$ , asociado al valor propio  $\lambda$ , la expresión  $vp = \lambda v$  implica que  $vp^k = \lambda^k v$ , para toda  $k \geq 1$ . Por lo tanto aplicando esto a la ecuación (5.1), se tiene, para  $k \geq 1$  que

$$\begin{aligned} \mu_0 p^k &:= a_0 v_0 p^k + a_1 v_1 p^k + \dots + a_{n-1} v_{n-1} p^k \\ &= a_0 \lambda_0^k v_0 + a_1 \lambda_1^k v_1 + \dots + a_{n-1} \lambda_{n-1}^k v_{n-1} \\ &= a_0 v_0 + a_1 \lambda_1^k v_1 + \dots + a_{n-1} \lambda_{n-1}^k v_{n-1}. \end{aligned}$$

Definimos  $\pi = a_0 v_0$  como la distribución invariante, debido a que cumple,

$$\lim_{k \rightarrow \infty} \mu_0 p^k = a_0 v_0,$$

y satisface la siguiente igualdad  $\pi p = \pi$  (véase la definición 2.2.1). En efecto,

$$\begin{aligned}\pi p(x) &= \sum_{y \in E} \pi(y) p(x, y) \\ &= \sum_{y \in E} a_0 v_0(y) p(x, y) \\ &= a_0 v_0(x) \\ &= \pi(x).\end{aligned}$$

Supongamos ahora que  $\pi'$  es otra medida invariante, por lo cual cumple:

$$\begin{aligned}\sum_{x \in E} \pi'(x) &= 1; \quad y \\ \pi'(z) &= \sum_{x \in E} \pi'(x) p(z, x).\end{aligned}$$

Por ser  $p(x, y) > 0$  para todo  $x, y \in E$  se tiene que

$$\lim_{n \rightarrow \infty} p^n(x, y) = \pi(y).$$

Entonces para todo  $y \in E$  se cumple

$$\pi'(y) = \sum_{x \in E} \pi'(x) \lim_{n \rightarrow \infty} p^n(x, y) = \pi(y) \sum_{x \in E} \pi'(x) = \pi(y).$$

Por lo tanto la medida invariante es única.

Por otro lado, para todo  $x \in E$  y  $k \geq 1$ , tenemos:

$$\begin{aligned}|\mu_0 p^k(x) - \pi(x)| &= |a_1 \lambda_1^k v_1(x) + \dots + a_{n-1} \lambda_{n-1}^k v_{n-1}(x)| \\ &\leq \sum_{m=1}^{n-1} |a_m v_m(x)| |\lambda_m^k| \\ &\leq \left( \sum_{m=1}^{n-1} |a_m v_m(x)| \right) |\lambda^*|^k.\end{aligned}$$

Así que:

$$|\mu_0 p^k(x) - \pi(x)| \leq \left( \sum_{m=1}^{n-1} |a_m v_m| \right) \lambda^{*k}.$$

Por lo tanto se cumple lo requerido.

■

**Observación 5.0.1** Del inciso (b) del Teorema 5.0.1 se obtiene :

$$\|\mu_0 p^k(\cdot) - \pi(\cdot)\|_{VT} \leq C(\lambda^*)^k,$$

donde  $C = \frac{1}{2} \sum_{x \in E} C_x$ .

**Ejemplo 5.0.1** Consideremos una cadena de Markov sobre el espacio de estados  $E = \{1, 2, 3, 4\}$  con probabilidades de transición:

$$p = \begin{pmatrix} 0,4 & 0,2 & 0,3 & 0,1 \\ 0,4 & 0,4 & 0,2 & 0 \\ 0,6 & 0,2 & 0,1 & 0,1 \\ 0,7 & 0,1 & 0 & 0,2 \end{pmatrix}.$$

Supongamos que la cadena comienza en 1, así tenemos que la distribución inicial es  $\mu_0 = (1, 0, 0, 0)$ .

Calculando los valores propios observamos que:  $\lambda_0 = 1$ ,  $\lambda_1 = 0,2618$ ,  $\lambda_2 = 0,0382$  y  $\lambda_3 = 0,2$ . Además los vectores propios correspondientes a cada uno de ellos son:

$$\begin{aligned} v_0 &= (0,4671, 0,2394, 0,2089, 0,0846), \\ v_1 &= (-0,4263, 0, 0,4263, 0), \\ v_2 &= (-0,0369, 0,2301, -0,556, 0,3721), \\ v_3 &= (-0,2752, 0,4854, 0,0898, -0,3). \end{aligned}$$

En términos de estos vectores propios, el vector inicial queda descrito:

$$\mu_0 = v_0 - 1,031v_1 - 0,4518v_2 - 0,2791v_3.$$

Así tenemos que  $\pi(\cdot) = v_0$  y por lo tanto, para  $k \geq 1$ :

$$\begin{aligned} \mu_0 p^k(3) &= v_0(3) - 1,031(\lambda_1^k)v_1(3) - 0,4518(\lambda_2^k)v_2(3) - 0,2791(\lambda_3^k)v_3(3) \\ &= 0,2089 - 1,031(0,2618)^k(0,4263) - 0,4518(0,0382)^k(-0,556) \\ &\quad - 0,2791(-0,2)^k(0,0898). \end{aligned}$$

Por lo tanto, para  $k \geq 1$ :

$$|\mu_0 p^k(3) - \pi(3)| < 0,8(0,2618)^k.$$

El cálculo para los demás estados, se puede obtener de manera similar al realizado para el estado 3.

**Observación 5.0.2** En [9] se exhibe un resultado que presenta el mismo contexto del Teorema 5.0.1, pero su tasa es distinta, a saber, para todo  $x, y \in E$  se cumple

$$|p^n(x, y) - \pi(y)| < \alpha\beta^n,$$

donde

$$\begin{aligned}\beta &= \text{máx}\{|\lambda_1|, |\lambda_2|, \dots, |\lambda_{n-1}|\}, \\ \alpha &= \sup_{(x,y)} \{|B_1(x,y)| + |B_2(x,y)| + \dots + |B_{n-1}(x,y)|\}, \quad y\end{aligned}$$

$|B_k(x,y)|$  es el determinante de la matriz espectral asociada a  $\lambda_k$  (véase [9]).

## Capítulo 6

# Acoplamiento

Considerando cadenas de Markov con espacio de estados  $(E, \xi)$  y kernel de transición  $p$ , en este capítulo damos dos métodos que auxilian a la técnica de acoplamiento para encontrar tasas de convergencia de tipo geométrico con respecto a la métrica de la variación total.

En el primer método requerimos que el espacio de estados sea pequeño, es decir, se necesita que la CM  $X$  cumpla la propiedad  $M(m_0, \delta, I_E, \nu)$  (véase la definición 2.3.4). También presentamos dos resultados; en uno de ellos se considera que el espacio de estados es finito y en el otro, se considera a  $E$  como un subconjunto compacto de  $\mathbb{R}$ .

Para el segundo método, se tiene como base teórica el teorema de Strassen (véase sección 3.2), el cual nos garantiza la existencia de acoplamientos que tienen la propiedad de estar ordenados por realizaciones (o trayectorias). En esta parte ofrecemos dos resultados que dependerán de la forma del espacio de estados.

En ambos métodos, el acoplamiento nos servirá para obtener la desigualdad de acoplamiento vista en la subsección 3.1.2.

Los resultados que se presentarán más adelante fueron consultados en [30], [31], [35], [46], [47] y [56].

## 6.1. Espacio de estados pequeño

### 6.1.1. Caso finito

Para el resultado que se presenta a continuación, supóngase que el espacio de estados es un conjunto **finito y pequeño**.

**Teorema 6.1.1** *Sea  $X = \{X_n\}$  una CM con kernel de transición  $p$ , la cual es recurrente y fuertemente aperiódica (véase inciso (iii) de la definición 2.3.6). Supóngase que  $\mu_0$  es la distribución inicial de  $X$  y  $\pi$  su medida invariante. Entonces*

$$\|P_{\mu_0}(X_n \in \cdot) - \pi(\cdot)\|_{VT} \leq (1 - \delta)^n, \quad n \geq 1, \quad (6.1)$$

donde,

$$\delta = \sum_{y \in E} \min_{x \in E} p(x, y). \quad (6.2)$$

**Demostración.** La recurrencia y la aperiodicidad fuerte de  $X$ , permite la creación de un acoplamiento para  $X$  con distribuciones iniciales  $\mu_0$  y  $\pi$ , respectivamente (véase el algoritmo 3.1.1). Con lo que se tiene la desigualdad (6.1). Por otro lado, nótese que (6.2) se obtiene directamente de la aperiodicidad fuerte de  $p$ .

■

**Ejemplo 6.1.1** *Considérese el espacio de estados  $E = \{0, 1, 2, 3, 4\}$  y el siguiente kernel de transición,*

$$p = \begin{pmatrix} 0,2 & 0,2 & 0,3 & 0,3 & 0 \\ 0,4 & 0 & 0,3 & 0,3 & 0 \\ 0,2 & 0,2 & 0,4 & 0,1 & 0,1 \\ 0,2 & 0,1 & 0,3 & 0,1 & 0,3 \\ 0,2 & 0 & 0,5 & 0,3 & 0 \end{pmatrix}.$$

Para mostrar la aperiodicidad fuerte, definimos para  $\bar{y} \in E$ :

$$\begin{aligned} q(\bar{y}) &= \min_{x \in E} p(x, \bar{y}), \\ \nu(\bar{y}) &= \frac{q(\bar{y})}{\delta}, \quad \bar{y} \in E \quad y \\ \delta &= \sum_{y \in E} \min_{x \in E} p(x, y). \end{aligned}$$

De aquí tenemos que  $\delta = 0,2 + 0,3 + 0,1 = 0,6$ ,  $q(0) = 0,2$ ,  $q(1) = 0$ ,  $q(2) = 0,3$ ,  $q(3) = 0,1$  y  $q(4) = 0,6$

$$\|\mu_0 p^k(\cdot) - \pi(\cdot)\|_{VT} \leq (1 - \delta)^k = (0,4)^k, \quad k \geq 1.$$

### 6.1.2. Caso compacto

El resultado que presentamos en este apartado, ofrece una tasa ergódica de tipo geométrico para CMs que tienen espacios de estados **compactos y pequeños**, pero no necesariamente finitos.

Cabe señalar que en la siguiente sección, se presenta una teoría con la cual se cubre el caso de espacio de estados compactos. No obstante, podemos observar que en ambas teorías existen diferencias debidas a sus hipótesis.

**Teorema 6.1.2** Sea  $X = \{X_n\}$  una CM con kernel de transición  $p$ , la cual es recurrente y fuertemente aperiódica (véase inciso (iii) de la definición 2.3.6). Supóngase que  $\mu_0$  es la distribución inicial de  $X$  y  $\pi$  su medida invariante. Entonces

$$\|P_{\mu_0}(X_n \in \cdot) - \pi(\cdot)\|_{VT} \leq (1 - \delta)^n, \quad n \geq 1, \quad (6.3)$$

donde,

$$\delta = \int_E \inf_{x \in E} p(x, dy), \quad y$$

para cada  $x \in E$ ,  $p(x, \cdot)$  es absolutamente continua con respecto a la medida de Lebesgue.

**Demostración.** Es similar a la del Teorema 6.1.1. ■

**Ejemplo 6.1.2** Sea  $E = [0, 1]$ . Para cada  $x \in E$  y  $A \in \mathbb{B}[0, 1]$ , definimos

$$p(x, A) = \int_A \frac{1 + x + y}{\frac{3}{2} + x} dy,$$

donde para cada  $x$ ,  $g_x(y) = \frac{1+x+y}{\frac{3}{2}+x}$ ,  $y \in E$ , es la densidad condicional asociada a la medida  $p(x, \cdot)$ . Para mostrar la aperiodicidad fuerte en  $X$ , definimos para  $A \in \xi$ :

$$\begin{aligned} q(A) &= \int_A \inf_{x \in E} p(x, dy), \\ \nu(A) &= \frac{q(A)}{\delta}, \quad y \\ \delta &= \int_E \inf_{x \in E} p(x, dy). \end{aligned}$$

Así tenemos:

$$\begin{aligned}\delta &= \int_0^1 \inf_{x \in E} \left( \frac{1+x+y}{\frac{3}{2}+x} \right) dy \\ &= \int_0^{\frac{1}{2}} \frac{2}{3}(1+y)dy + \int_{\frac{1}{2}}^1 \frac{2}{5}(2+y)dy \\ &= \frac{29}{30}.\end{aligned}$$

Por lo tanto obtenemos:

$$\|P_{\mu_0}(X_n \in \cdot) - \pi(\cdot)\|_{VT} \leq \left(\frac{1}{30}\right)^n,$$

$n \geq 1$ .

## 6.2. Monotonicidad estocástica

En esta sección presentamos una teoría que exhibe tasa de convergencia de tipo geométrico para cadenas de Markov que son estocásticamente monótonas y cuentan con espacio de estados que contienen un elemento mínimo.

En la teoría presentada, damos un ejemplo que tiene aplicabilidad a procesos de almacenamiento, como son, colas, inventarios y presas. También en el apéndice C damos una aplicación a control estocástico.

**Observación 6.2.1** *El resultado que se muestra en esta sección, contiene el caso de espacio de estado compacto visto en la subsección 6.1.2. Esto se debe a que los espacios de estados compactos contienen un elemento mínimo, y estos últimos pueden ser no acotados por arriba.*

*Además, podemos resaltar que la compacidad en esta teoría no es un hecho importante, si no la presencia en el espacio de estados de un elemento mínimo.*

La fuente bibliográfica para este capítulo es esencialmente: [30], [31], [35], y [47].

### 6.2.1. Desarrollo

En esta sección consideramos por comodidad, CMs con espacios de estados  $E = [0, \infty)$ . En general, podríamos pensar en tener un espacio de la forma  $E = [a, \infty)$ , donde  $a \in \mathbb{R}$ . El hecho importante es que el espacio de estados tenga un elemento mínimo.

La teoría que presentamos a continuación fue desarrollada por Lund y Tweedie en [30].

**Lema 6.2.1 ([30] p. 185)** *Sea  $X = \{X_n\}_{n=0}^\infty$  una cadena de Markov que satisface la condición de Foster-Lyapunov para  $V$ ,  $\lambda$ ,  $b$  y  $C = \{0\}$ , con  $V(0) = 1$  (véase 2.3.7). Entonces para  $\tau_0 := \inf\{n \in \mathbb{N} : X_n = 0\}$  se cumple:*

$$(a) \quad E_x[r^{\tau_0}] \leq V(x), \text{ con } r \leq \frac{1}{\lambda} \text{ y } x > 0.$$

$$(b) \quad E_0[r^{\tau_0}] \leq r(\lambda + b), \text{ con } r \leq \frac{1}{\lambda}.$$

**Demostración.** Observemos que para cada  $x \in E$ , la desigualdad de Foster-Lyapunov:

$$pV(x) \leq \lambda V(x) + bI_{\{0\}}(x), \quad (6.4)$$

cambia de forma al considerar estados positivos o al estado cero (esto se debe a la función indicadora). A continuación analizaremos las consecuencias de estos cambios.

Si  $x > 0$ , entonces la desigualdad queda de la siguiente forma:

$$pV(x) \leq \lambda V(x).$$

Dividiendo el espacio de estados en  $\{0\} \cup (0, \infty)$  e iterando la desigualdad anterior obtenemos,

$$\begin{aligned} V(x) &\geq \lambda^{-1} pV(x) \\ &= \lambda^{-1} \left[ \int_{\{0\}} V(z)p(x, dz) + \int_{(0, \infty)} V(z)p(x, dz) \right] \\ &= \lambda^{-1} \left[ V(0)p(x, 0) + \int_{(0, \infty)} V(z)p(x, dz) \right], \end{aligned}$$

es decir,

$$V(x) \geq \lambda^{-1} \left[ P_x(\tau_0 = 1) + \int_{(0, \infty)} V(z)p(x, dz) \right].$$

Realizando este procedimiento una vez más obtenemos:

$$V(x) \geq \lambda^{-2} P_x(\tau_0 = 2) + \lambda^{-1} P_x(\tau_0 = 1) + \int_{(0, \infty)} \int_{(0, \infty)} V(z_2)p(z, dz_2)p(x, dz).$$

Como podemos observar, continuando de forma sucesiva este procedimiento, obtenemos la serie  $\sum_{n=1}^{\infty} \lambda^{-n} P_x(\tau_0 = n)$ , y una integral múltiple cuyo valor es positivo. De este hecho ver que

$$\begin{aligned} V(x) &\geq \sum_{n=1}^{\infty} \lambda^{-n} P_x(\tau_0 = n) \\ &= E_x[(\lambda^{-1})^{\tau_0}]. \end{aligned} \quad (6.5)$$

Para obtener la consecuencia descrita en el inciso (a) del lema, tenemos que aplicar  $\lambda^{-1} \geq r$  a la desigualdad (A.3).

Por otro lado, si  $x = 0$ , la desigualdad (4.1) queda descrita de la siguiente forma:

$$PV(0) \leq \lambda + b.$$

Ahora, un cálculo similar al realizado para el caso  $x > 0$ , permite obtener que

$$\begin{aligned} E_0[r^{\tau_0}] &= rp(0, 0) + r \int_{(0, \infty)} E_z[r^{\tau_0}] p(0, dz) \\ &\leq rp(0, 0) + \int_{(0, \infty)} rV(z) P(0, dz) \\ &= rPV(0) \\ &\leq r(\lambda + b). \end{aligned}$$

Por lo tanto

$$E_0[r^{\tau_0}] \leq r(\lambda + b), \quad r \leq \lambda^{-1},$$

con lo cual se concluye la prueba. ■

**Teorema 6.2.1 ([30] p. 184)** *Suponga que  $X = \{X_n\}_{n=0}^{\infty}$  es una cadena de Markov estocásticamente monótona (véase el ejemplo 3.2.1) que satisface la propiedad de Foster-Lyapunov para  $V$ ,  $\lambda$ ,  $b$  y  $C = \{0\}$ , con  $V(0) = 1$ . Entonces:*

$$\lim_{n \rightarrow \infty} r^n \|p^n(x, \cdot) - \pi(\cdot)\|_{VT} = 0 \quad \text{para toda } r \leq \lambda^{-1}, \quad x \geq 0,$$

donde  $\pi$  es la distribución invariante.

**Demostración.** Sean  $x \in E$ ,  $\{X'_n\}_{n=0}^{\infty}$ ,  $\{X''_n\}_{n=0}^{\infty}$  y  $\{X'''_n\}_{n=0}^{\infty}$  tres realizaciones de Markov, con distribuciones iniciales  $\delta_x$ ,  $\pi$ , y  $\nu$ , respectivamente, donde  $\delta_x$  es la probabilidad concentrada en el estado  $x$ ,  $\nu = P(\text{máx}(Y, x))$  y  $Y$  es una variable aleatoria con distribución  $\pi$ .

Denotemos la probabilidad de acoplamiento para las probabilidades  $P_x$ ,  $P_\pi$  y  $P_\nu$  (véase definición 3.1.1) por  $\hat{P}$  y con  $T$  a la variable aleatoria definida por:

$$T = \inf\{n \geq 0 : X'_n = X''_n\},$$

es decir,  $T$  es el tiempo de acoplamiento para las realizaciones  $\{X'_n\}_{n=0}^\infty$  y  $\{X''_n\}_{n=0}^\infty$ .

Observemos que para cada  $n \in \mathbb{N}$ , y  $A \in \xi$  se cumple,

$$P_\pi(X''_n \in A) = \pi(A),$$

por ser  $\pi$  la distribución invariante. Aplicando este hecho y el Lema (3.1.3) a las realizaciones  $\{X'_n\}_{n=0}^\infty$  y  $\{X''_n\}_{n=0}^\infty$  se tiene que para  $A \in \xi$  y  $n \geq 1$ ,

$$\|P_x(X'_n \in A) - P_\pi(X''_n \in A)\|_{VT} \leq \hat{P}(T > n). \quad (6.6)$$

Por otro lado, usando la hipótesis de que la cadena es estocásticamente monótona, se tiene que la cadena es ordenada por trayectorias (véase el ejemplo 3.2.2). Con este resultado se deduce que cuando haya alcanzado al estado cero la realización de Markov con estado inicial más grande, entonces la realización con estado inicial más pequeño también lo habrá alcanzado; es decir, en términos de el tiempo de acoplamiento para  $\{X'_n\}_{n=0}^\infty$  y  $\{X''_n\}_{n=0}^\infty$ , y el tiempo de alcance al estado 0 por parte de la realización  $\{X'''_n\}_{n=0}^\infty$ , se tiene:

$$\{\tau_0 \leq n\} \subseteq \{T \leq n\}, \quad (6.7)$$

para toda  $n \geq 1$ . Aplicando (4.4) en (4.3), para  $0 < r < 1$  y  $n \geq 1$  se tiene que

$$\begin{aligned} r^n \|P_x(X'_n \in \cdot) - P_\pi(X''_n \in \cdot)\|_{VT} &\leq r^n \hat{P}(T > n) \\ &\leq r^n \hat{P}(\tau_0 > n) \\ &= r^n P_\nu(\tau_0 > n). \end{aligned}$$

Por otro lado, observemos que para tener la convergencia,

$$\lim_{n \rightarrow \infty} r^n \|P_x(X'_n \in \cdot) - P_\pi(X''_n \in \cdot)\|_{VT} = 0, \quad (6.8)$$

se debe de cumplir que

$$\lim_{n \rightarrow \infty} r^n P_\nu(\tau_0 > n) = 0. \quad (6.9)$$

El Lema 6.2.3, nos indica a  $E_\nu[r^{\tau_0}] < \infty$  como condición necesaria para obtener (6.9). Por lo tanto, enfocaremos nuestro esfuerzo a corroborar esta condición.

Observemos que

$$\begin{aligned}
E_\nu[r^{\tau_0}] &= \int_{[0,\infty)} E_z[r^{\tau_0}] d\nu(z) \\
&= \int_{[0,x]} E_z[r^{\tau_0}] d\nu(z) + \int_{(x,\infty)} E_z[r^{\tau_0}] d\nu(z) \\
&= P(X \leq x) E_x[r^{\tau_0}] + \int_{(x,\infty)} E_z[r^{\tau_0}] d\pi(z) \\
&\leq E_x[r^{\tau_0}] + \int_{(x,\infty)} V(z) d\pi(z) < \infty,
\end{aligned}$$

debido a que  $\int_{(x,\infty)} V(z) d\pi(z) \leq \frac{b}{1-\lambda} < \infty$  (véase [33, p. 345]). Con lo cual se finaliza la demostración. ■

De hecho, se puede reformular el resultado anterior de las siguientes maneras:

**Teorema 6.2.2 ([30] p. 186)** *Suponga que  $X = \{X_n\}_{n=0}^\infty$  es una cadena de Markov estocásticamente monótona que satisface la propiedad de Foster- Lyapunov para  $V$ ,  $\lambda$ ,  $b$  y  $C = \{0\}$ , con  $V(0) = 1$ . Entonces para  $r > 1$  y  $n \in \mathbb{N}$  se tiene que*

$$\|p^n(x, \cdot) - \pi(\cdot)\|_{VT} \leq C_x(r) r^{-n},$$

donde  $C_x(r) = E_\nu[r^{\tau_0}] \leq E_x[r^{\tau_0}] + \frac{b}{1-\lambda}$ .

**Demostración.** El resultado buscado, forma parte del cuerpo de la prueba del teorema anterior. ■

**Teorema 6.2.3 ([30] p. 189)** *Suponga que  $X = \{X_n\}_{n=0}^\infty$  es una cadena de Markov estocásticamente monótona tal que  $E_0[r^{\tau_0}] < \infty$  para algún  $r > 1$ . Entonces:*

$$\lim_{n \rightarrow \infty} r^n \|\delta_x p^n(\cdot) - \pi(\cdot)\|_{VT} = 0 \quad \text{para toda } r \leq \lambda^{-1}, \quad x \geq 0.$$

**Demostración.** Este resultado se obtiene directamente de los siguientes lemas. ■

**Lema 6.2.2 ([30] p. 184)** *Suponga que  $X = \{X_n\}_{n=0}^\infty$  es una cadena de Markov estocásticamente monótona. Entonces se tiene lo siguiente: para todo  $R > 1$  y  $x > 0$ , son equivalentes  $E_0[R^{\tau_0}] < \infty$  y  $E_x[R^{\tau_0}] < \infty$ .*

**Demostración.** Primero probaremos que  $E_x[R^{\tau_0}] < \infty$  implica  $E_0[R^{\tau_0}] < \infty$  para  $x > 0$  y  $R > 1$ .

Obsérvese que para  $x \in E$  se tiene que  $P_0 \stackrel{\text{est.}}{\leq} P_x$ , y  $g(t) := R^t$ ,  $t \geq 1$  es una función no decreciente. Aplicando estos dos hechos en el Lema 3.2.1, obtenemos:

$$\begin{aligned} E_0[R^{\tau_0}] &= \int R^{\tau_0} dP_0 \\ &\leq \int R^{\tau_0} dP_x \\ &= E_x[r^{\tau_0}] < \infty. \end{aligned}$$

Ahora supongamos que  $E_0[R^{\tau_0}] < \infty$  y usemos la hipótesis de irreducibilidad, seleccionando un  $m$  tal que  ${}_0p^m(0, [x, \infty)) > 0$ . La monotonicidad de  $E_x[R^{\tau_0}]$  en  $x$ , y la propiedad de Markov implican que

$$E_0[R^{\tau_0}] \geq {}_0p^m(0, [x, \infty)) R^m E_x[R^{\tau_0}].$$

De aquí concluimos que  $E_x[R^{\tau_0}] < \infty$ . ■

**Lema 6.2.3 ([30] p. 189)** *Suponga que cadena  $\{X_n\}_{n=0}^{\infty}$  es una cadena de Markov ordenada estocásticamente tal que  $E_0[r^{\tau_0}] < \infty$  para algún  $r > 1$  fijo. Entonces la función  $V$  definida por  $V(0) = 1$  y  $V(x) = E_x[r^{\tau_0}]$  para  $x > 0$ , constantes  $\lambda = \frac{1}{r}$  y  $b = \frac{1}{r}[E_0[r^{\tau_0}] - 1]$  satisface la condición de Foster- Lyapunov, con igualdad.*

**Demostración.** Observemos que para  $z \in E$  se cumple,

$$\begin{aligned} V(z) &= E_z[r^{\tau_0}] \\ &= \sum_{n=0}^{\infty} r^n P_z(\tau_0 = n) \\ &= \sum_{n=0}^{\infty} r^n p^n(z, 0). \end{aligned}$$

Entonces para  $x > 0$ , se tiene:

$$\begin{aligned}
 pV(x) &= \int_{[0, \infty)} \left\{ \sum_{n=0}^{\infty} r^n p^n(z, 0) \right\} p(x, dz) \\
 &= \int_{\{0\}} \left\{ \sum_{n=0}^{\infty} r^n p^n(z, 0) \right\} p(x, dz) + \int_{(0, \infty)} \left\{ \sum_{n=0}^{\infty} r^n p^n(z, 0) \right\} p(x, dz) \\
 &= p(x, 0) + \int_{(0, \infty)} \left\{ \sum_{n=0}^{\infty} r^n P^n(z, 0) \right\} p(x, dz) \\
 &= \sum_{n=0}^{\infty} r^n p^{n+1}(x, 0) \quad [\text{por el teorema de Fubini}] \\
 &= \frac{r}{r} \sum_{n=0}^{\infty} r^n p^{n+1}(x, 0) \\
 &= \frac{1}{r} \sum_{n=0}^{\infty} r^{n+1} P_z(\tau_0 = n + 1) \\
 &= \frac{1}{r} E_x[r^{\tau_0}].
 \end{aligned}$$

De la última igualdad, se obtiene lo buscado.

Cálculos similares a los anteriores son realizados para el caso  $x = 0$ . ■

### 6.2.2. Ejemplos

#### La caminata aleatoria Bernoulli

Considérese la caminata aleatoria  $X = \{X_n\}$  sobre el espacio de estados  $\{0, 1, 2, \dots\}$  tal que  $p + q = 1$  y:

$$p(i, j) = \begin{cases} q, & \text{si } i \geq 0 \text{ y } j = i + 1, \\ p, & \text{si } i \geq 1 \text{ y } j = i - 1, \\ p, & \text{si } i = j = 0 \end{cases} .$$

Notemos que para que  $X$  sea recurrente tenemos que pedir que  $p > \frac{1}{2}$ .

Se puede comprobar que  $X$ :

- es estocásticamente monótona,
- la distribución invariante es:  $\pi(\{k\}) = \left(1 - \frac{p}{q}\right) \left(\frac{q}{p}\right)^k$ ,  $k \in \mathbb{Z}^+$ ,

$$\begin{aligned} \blacksquare E_k[r^{\tau_0}] &= \left[ \frac{1 - (1 - 4pqr^2)^{\frac{1}{2}}}{2qr} \right]^k, \text{ con } k \geq 1 \text{ y} \\ \blacksquare E_0[r^{\tau_0}] &= \left[ \frac{2pr + 1 - (1 - 4pqr^2)^{\frac{1}{2}}}{2qr} \right]^k, \text{ con } k \geq 1 \end{aligned}$$

Nótese que  $E_0[r^{\tau_0}] < \infty$  si y sólo si  $r \leq \left(\frac{1}{4pq}\right)^{\frac{1}{2}}$ . Por lo tanto de acuerdo al Teorema 6.2.3 obtenemos:

$$\lim_{n \rightarrow \infty} r^n \|\delta_x p^n(\cdot) - \pi(\cdot)\|_{VT} = 0 \text{ para } r \leq \left(\frac{1}{4pq}\right)^{\frac{1}{2}}, \quad x > 0.$$

### Una cola M/GI/1

Sea  $X = \{X_n\}$  el número de clientes en la cola  $M/GI/1$ ,  $\beta$  denota la tasa de llegada exponencial de los clientes y  $H$  es la distribución del tiempo de servicio de los clientes en la cola. De aquí podemos determinar que:

$$p(i, j) = \begin{cases} q_j & \text{si } i = 0 \text{ y } j \geq 0, \\ q_{j-i+1} & \text{si } i \geq 1 \text{ y } j \geq i - 1, \\ 0 & \text{en otro caso.} \end{cases}$$

$$\text{donde } q_k = \int_0^\infty [e^{-\beta t} (\beta t)^k] H(dt).$$

De esta CM obtenemos las siguientes propiedades:

- es estocásticamente monótona,
- tiene distribución invariante  $\pi$ . Este hecho se prueba suponiendo que la intensidad de tráfico de la cola  $\rho := \beta\mu_H$ , es menor que uno y que  $H^*(s) = \int_{[0, \infty)} e^{st} H(dt) < \infty$ , para algún  $s > 0$ ; donde  $\mu_H = \int_{[0, \infty)} t H(dt)$ .
- satisface la propiedad de Foster-Lyapunov. Nótese que para  $V(x) = r^x$ ,  $r > 1$  y  $x > 0$ , se obtiene:

$$\begin{aligned} pV(x) &= E_x [r^{X_1}] = \sum_{n=0}^{\infty} r^{x+n-1} q_n \\ &= V(x)\psi(r), \end{aligned}$$

donde

$$\psi = r^{-1} \sum_{n=0}^{\infty} r^n q_n = r^{-1} H^*(\beta[r-1]).$$

De aquí, podemos tomar  $\lambda = \inf\{\psi(r) : r \geq 1\}$

Por lo tanto, de acuerdo al Teorema 6.2.1 se tiene para todo  $r \leq \lambda^{-1}$  y  $x \geq 0$  que:

$$\lim_{n \rightarrow \infty} r^n \|\delta_x p^n(\cdot) - \pi(\cdot)\|_{VT} = 0.$$

### La caminata de Lindley

En este apartado mostraremos un ejemplo de una cadena de Markov que tiene la tasa descrita en el Teorema 6.2.3. Esta aplicación fue consultada en [31].

Sean  $\{A_n\}$  una sucesión de variables aleatorias independientes e idénticamente distribuidas, y  $F$  su función de distribución. Supóngase que  $E[A_1] < 0$ .

Llamaremos **caminata de Lindley** a la sucesión de variables aleatorias  $X = \{X_n\}_{n=0}^\infty$ , descritas de la forma siguiente:

$$X_n = \max [X_{n-1} + A_n, 0],$$

donde  $n \in \mathbb{N}$  y  $X_0 \geq 0$ .

Con la definición anterior, podemos verificar las siguientes observaciones.

#### Observaciones 6.2.1

- *El espacio de estados para la caminata de Lindley es  $E = [0, \infty)$ .*
- *Su kernel de transición es,  $p(x, (-\infty, y]) = \begin{cases} 0 & \text{si } y < 0, \\ F(y - x) & \text{si } y \geq 0 \end{cases}$ .*
- *$X$  está ordenada estocásticamente. (Nótese que para  $x, y \in E$  tal que  $x < y$  y todo  $z \in E$  se tiene*

$$P(x, [z, \infty)) = 1 - F(z - x) \leq 1 - F(z - y) = P(y, [z, \infty)).$$

- *La caminata de Lindley tiene distribución invariante  $\pi$  que satisface:*

$$\pi((0, z]) = \int_E F(z - y) \pi(dy),$$

$$z \in E.$$

Con las observaciones descritas previamente, podemos ver que la caminata de Lindley cumple la mayor parte de las hipótesis del Teorema 6.2.1, con las cuales se obtiene convergencia de tipo geométrico. Sólo faltaría mostrar, que la caminata cumple con la propiedad de Foster-Lyapunov. Para lograr este fin, como se vio en los Lemas 6.2.2 y 6.2.3, respectivamente, veremos que  $E_0[r^{\tau_0}] < \infty$ .

**Hipótesis 6.2.1**

- Supóngase que  $E[A_1] < 0$ ,
- Existe  $r_0 > 1$  tal que  $\phi(r_0) =: E[r_0^{A_1}] < \infty$  y  $\phi'(r_0) = 0$ ,
- Si  $A_1$  es una retícula (en inglés lattice), entonces  $P(A_1 = 0) > 0$ .

**Lema 6.2.4** ([31], p. 808) *Bajo las hipótesis 6.2.1 se cumple que  $E_0[r^{\tau_0}] < \infty$ .*

**Demostración.** Se puede ver en [10] que

$$E_0(r^{\tau_0}) = 1 + (r - 1) \exp \left[ \sum_{k=1}^{\infty} \frac{r^k}{k} P(S_k > 0) \right],$$

donde  $S_k = \sum_{j=1}^k A_j$ . De aquí,  $E_0[r^{\tau_0}] < \infty$  si y sólo si

$$H(r) := \sum_{k=1}^{\infty} k^{-1} r^k P(S_k > 0) < \infty.$$

El teorema 1 de [8], indica que el radio de convergencia de  $H(r)$  es  $(\phi(r_0))^{-1}$ . Para ver que  $H(r) < \infty$  cuando  $r = (\phi(r_0))^{-1}$  aplicamos la expansión  $P(S_k > 0) \sim M(\phi(r_0))^k k^{-1/2}$  de [3]. Con esto se obtiene el resultado requerido.

■

**Teorema 6.2.4** ([31] p. 807) *Bajo las hipótesis 6.2.1 y si  $r$  es tal que  $E_0[r^{\tau_0}] < \infty$ , entonces para todo  $x \in E$  y  $s \leq \phi(r_0)^{-1}$  se cumple:*

$$\lim_{n \rightarrow \infty} s^n \|p^n(x, \cdot) - \pi(\cdot)\|_{VT} = 0. \quad (6.10)$$

**Demostración.** Es consecuencia directa del Teorema 6.2.3. ■

**Una cola GI/G/1**

Consideremos una cola  $GI/G/1$ , donde  $S_n$  es el tiempo de servicio del  $n$ -ésimo cliente y  $I_n$  es el tiempo entre el arribo del cliente  $n$  y el  $n+1$ . Aquí  $\{S_n\}$  y  $\{I_n\}$  son sucesiones de variables aleatorias no negativas, independientes e idénticamente distribuidas. Sea  $Q_n$  el tiempo que pasa el  $n$ -ésimo cliente esperando para recibir su servicio. Entonces podemos ver que  $Q_n = \max(Q_{n-1} + S_{n-1} - I_n, 0)$ ,

para  $n \geq 1$  y  $Q_0 = x$ . Si tomamos  $A_n = S_n - I_n$ , podemos ver que  $\{Q_n\}$  es una caminata de Lindley, la cual tiene distribución invariante  $\pi$ , cuando  $E[S_0] < E[I_1]$ . Definiendo  $\phi(r) = E[r^{S_0 - I_1}]$ , suponiendo que existe  $r_0 > 1$  tal que  $\phi(r_0) < \infty$  y  $\phi'(r_0) = 0$ , y  $P(S_0 - I_1 = 0) > 0$ , entonces tenemos que  $\{Q_n\}$  tiene tasa de convergencia  $\phi(r_0)^{-1}$ .

Específicamente, considerando que  $S_n$  tiene distribución exponencial con parámetro 3 y  $I_n \equiv \frac{2}{3}$ . Tenemos que  $E[S_0 - I_1] = -\frac{1}{3} < 0$ . Asumimos que  $r < e^3$ , entonces  $\phi(r_0) = 3r^{-\frac{2}{3}}(3 - \ln(r)) < \infty$  y existe  $r_0 = e^{\frac{3}{2}} > 1$  tal que  $\phi'(r_0) = 0$ . Por lo tanto la tasa de convergencia geométrica es  $\phi(r_0^{-1}) = \frac{e}{2}$ .

## Capítulo 7

# Sucesiones aperiódicas y renovación

En este último capítulo, dedicado a tasas de convergencia, se analizarán los resultados obtenidos por Baxendale en [4], en los cuales muestra convergencia hacia la distribución invariante de tipo geométrico y ergódico, bajo la métrica  $V$ -norma.

La teoría mostrada en estos resultados, necesita la construcción de un proceso de renovación a través de una sucesión aperiódica formada por distribuciones de variables aleatorias.

### 7.1. Desarrollo

En este capítulo consideraremos CMs  $X = \{X_n\}$  con espacio de estados contenidos en  $\mathbb{R}$ .

Presentamos tres resultados que exhiben tasas de convergencia uniformemente ergódicas, de tipo geométrico, para la medida invariante  $\pi$ , en la métrica  $V$ -norma. Es decir,

$$\|p^n(x, \cdot) - \pi(\cdot)\|_V \leq MV(x)\gamma^n, \quad (7.1)$$

donde

- $\|p^n(x, \cdot) - \pi(\cdot)\|_V := \sup_{|g| \leq V} \left| \int g(y) p^n(x, dy) - \int g(y) d\pi(y) \right|$ ,
- $V$  es una función medible de  $E$  en  $[c, \infty)$  con  $c > 0$ ,

- $g$  es una función medible de  $E$  en  $\mathbb{R}$  y
- $M$  y  $\gamma$  son constantes tales que  $M > 0$  y  $0 < \gamma < 1$ .

Cabe señalar, que la métrica de la variación total se obtiene de la  $V$ -norma considerándose funciones  $g$  tal que  $|g(x)| \leq 1$

Básicamente, los tres resultados que se presentan en el capítulo, nos ofrecen cotas para las constantes  $M$  y  $\gamma$  de la desigualdad (7.1), respectivamente.

El primero de ellos es el caso más general en el siguiente sentido, **no pide** que la CM  $X$  cuente con las propiedades de reversibilidad y positividad (definidas en los siguientes párrafos).

El segundo resultado pide que la CM  $X$  tenga como hipótesis principal la propiedad de **reversibilidad** con respecto a su medida invariante  $\pi$ , es decir, para toda  $f$  y  $g$  definidas de  $E$  en  $\mathbb{R}$  se cumple

$$\int_E (pf)(x) g(x) d\pi(x) = \int_E f(x) (pg)(x) d\pi(x). \quad (7.2)$$

En teoría de operadores la igualdad 7.2 corresponde al operador  $(p)$  auto-adjunto.

Finalmente en el tercer resultado se requiere que la CM sea reversible y positiva, donde entenderemos que una cadena de Markov es **positiva** si:

$$\int_E (pf)(x) f(x) d\pi(x) \geq 0,$$

par toda  $f \in L^2(\pi)$ .

Cabe señalar que el uso de las hipótesis extras descritas anteriormente, refinarán las cotas para  $M$  y  $\gamma$ .

Las hipótesis comunes para los resultados descritos previamente son:

### Hipótesis 7.1.1

- Se cumple la condición de minorización  $M(1, \tilde{\delta}, I_C, \nu)$ , es decir

$$p(x, \cdot) \geq \tilde{\delta}\nu(\cdot), \quad x \in C. \quad (7.3)$$

- Existe una función medible  $V : E \rightarrow [1, \infty)$  y constantes  $0 < \lambda < 1$  y  $0 < K < \infty$  que satisfacen:

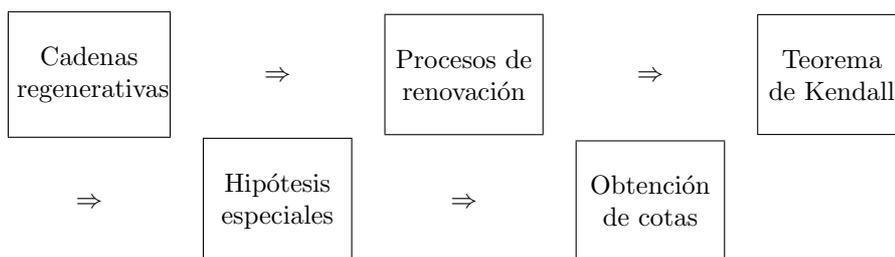
$$pV(x) \leq \lambda V(x) I_{C^c}(x) + K I_C(x), \quad x \in E. \quad (7.4)$$

- Se cumple la condición de aperiodicidad fuerte, es decir, existe  $\delta > 0$  tal que:

$$\tilde{\delta\nu}(C) \geq \delta. \quad (7.5)$$

El fundamento teórico para cada uno de los resultados de este capítulo, es el uso del teorema obtenido por Kendall en la teoría de procesos de renovación.

Para facilitar la lectura del presente material ofrecemos un esquema que muestra las ideas generales utilizadas en esta teoría.



Cabe señalar que en esta metodología se tendrá en cuenta dos casos. El primero de ellos está relacionado con la suposición de que en el espacio de estados exista un átomo ( $\alpha \subseteq E$  es llamado **átomo** si para todo  $a, y \in \alpha$  se tiene  $p(x, \cdot) = p(y, \cdot)$ ). El segundo caso no supone la existencia de un átomo en el espacio de estados.

Además resaltamos el hecho de que la teoría de sucesiones aperiódicas aplicada a procesos de renovación, será utilizada para acotar al siguiente término,

$$\left| \sum_{n=0}^{\infty} (u_n - u_{\infty}) z^n \right|,$$

donde  $\{u_n\}$  será la sucesión de renovación generada a partir de la sucesión aperiódica, y  $u_{\infty}$  será el límite de  $\{u_n\}$  (véase apéndice D).

Dicha serie potencias, se localizará en las desigualdades D.2.1 y D.2.2 del apéndice D.

### 7.1.1. Cotas

Con el propósito de simplificar la redacción para los tres resultados que se presentan en el presente capítulo, damos a continuación la definición de las cotas  $M, \gamma$  y  $\rho$  mencionadas en 7.2.

Para este fin, consideremos las hipótesis 7.1.1 y una sucesión,  $b = \{b_n\}_{n=1}^{\infty}$ , de variables aleatorias que cumple las siguientes dos propiedades:

- es **aperiódica**, es decir, el máximo común divisor del conjunto  $I(b) := \{n \mid b_n > 0\}$  es 1 y
- $\sum_{n=1}^{\infty} b_n R^n \leq L$ , para  $R > 1$  y  $L < \infty$ .

**Definición 7.1.1**

- $\alpha_1 = 1 + \log \left( \frac{K - \tilde{\delta}}{1 - \tilde{\delta}} \right) / \log(\lambda^{-1})$ .
- $\alpha_2 = 1 + \frac{\log \left( \frac{K}{\delta} \right)}{\log(\lambda^{-1})}$ .
- $R_0 = \min \left( \lambda^{-1}, (1 - \tilde{\delta})^{-1/\alpha_1} \right)$
- $1 < R \leq R_0$
- $\tilde{R} = \arg \max_{1 < R \leq R_0} R_1(\delta, R, L(R))$
- $L(R) = \frac{\tilde{\delta} R^{\alpha_2}}{1 - (1 - \tilde{\delta})R^{\alpha_1}}$
- $R_1 = R_1(\delta, R, L)$  es la única solución para la ecuación

$$\frac{r - 1}{r(\log(\frac{R}{r}))^2} = \frac{e^2 \delta (R - 1)}{8(L - 1)},$$

en el intervalo  $(1, R)$ , considerando como incógnita a  $x$  y  $L > 1$ .

- para algún  $r \in (1, R_1)$  definimos  $K_1 = K_1(r, \delta, R, L)$  de la siguiente manera,

$$K_1 = \left( \frac{2\delta + 2 \log(N) (\log(\frac{R}{r}))^{-1} - 8Ne^{-2}(r-1)r^{-1}(\log(R/r))^{-2}}{[(r-1)\beta - 8Ne^{-2}(r-1)r^{-1}(\log(\frac{R}{r}))^{-2}]} \right),$$

donde  $N = \frac{L - 1}{R - 1}$ .

- $D = \gamma - \lambda$ ,  $H = K\gamma - \lambda$ ,  $\gamma \in (\rho, 1)$  y  $\rho$  véase en las definiciones para primera, segunda y tercera, respectivamente (varía su definición).

**Primera cota****Si C es un átomo**

Se establecen  $\rho = \frac{1}{R_1}$ , donde  $R_1 = R_1(\delta, \lambda^{-1}, \lambda^{-1}K)$ .

$$M = \frac{\max\left(\lambda, \left(K - \frac{\lambda}{\gamma}\right)\right)}{D} + \frac{K\left(K - \frac{\lambda}{\gamma}\right)}{\gamma D} K_1 + \frac{\left(K - \frac{\lambda}{\gamma}\right) \max(\lambda, K - \lambda)}{D(1 - \lambda)} + \frac{\lambda(K - 1)}{D(1 - \lambda)}.$$

**Si C no es un átomo**

Se establecen  $\rho = \frac{1}{R_1}$ , donde  $R_1 = R_1(\delta, \tilde{R}, L(\tilde{R}))$ . Entonces

$$M = \frac{\max\left(\lambda, \left(K - \frac{\lambda}{\gamma}\right)\right)}{D} + \frac{K[H - \tilde{\delta}D]}{\gamma^2 D[1 - (1 - \delta)\gamma^{-\alpha_1}]} + \frac{\tilde{\delta}\gamma^{-\alpha_2 - 2}KH}{D[1 - (1 - \tilde{\delta})\gamma^{-\alpha_1}]^2} K_1 + \frac{\gamma^{-\alpha_2 - 1}H}{D[1 - (1 - \tilde{\delta})\gamma^{-\alpha_1}]^2} \times \left( \frac{\tilde{\delta} \max(\lambda, K - \lambda)}{1 - \lambda} + \frac{(1 - \tilde{\delta})(\gamma^{-\alpha_1} - 1)}{\gamma^{-1} - 1} \right) + \frac{\gamma^{-\alpha_2} \lambda(K - 1)}{(1 - \lambda)D[1 - (1 - \tilde{\delta})\gamma^{-\alpha_1}]} + \frac{K - \lambda - \tilde{\delta}(1 - \lambda)}{(1 - \lambda)(1 - \gamma)} \times \left( (\gamma^{\alpha_2} - 1) + \frac{(1 - \tilde{\delta})(\gamma^{-\alpha_1} - 1)}{\tilde{\delta}} \right).$$

**Segunda cota**

Suponemos que la cadena de Markov es reversible.

**Si C es un átomo**

$\rho = \frac{1}{R_2}$ , donde

$$R_2 = \begin{cases} \sup \left\{ r < \lambda : 1 + 2\delta r > r^{1 + \frac{\log(K)}{\log(\lambda^{-1})}} \right\} & \text{si } K > \lambda + 2\delta, \\ \lambda^{-1} & \text{si } K \leq \lambda + 2\delta. \end{cases}$$

La definición de  $M$  se da de forma similar a la del caso anterior, sólo hay que reemplazar  $K_1$  por  $K_2 := 1 + \frac{1}{\gamma - \rho}$ .

**Si  $C$  no es un átomo**

$\rho = \frac{1}{R_2^*}$ , donde

$$R_2^* = \begin{cases} \sup\{r < R_0 : 1 + 2\beta r > L(r)\} & \text{si } L(R_0) > 1 + 2\beta R_0, \\ R_0 & \text{si } L(R_0) \leq 1 + 2\beta R_0. \end{cases}$$

La definición de  $M$  se da de forma similar al de la primera cota, sólo hay que reemplazar  $K_1$  por  $K_2 := 1 + \frac{\sqrt{\delta}}{\gamma - \rho}$ .

**Tercera cota**

Suponemos que la cadena de Markov es reversible y positiva. Definimos

$$\rho = \begin{cases} \lambda, & \text{si } C \text{ es un átomo,} \\ \frac{1}{R_0}, & \text{si } C \text{ no es un átomo.} \end{cases},$$

y definición de  $M$  es igual como para el caso de Segunda cota.

## 7.1.2. Resultados Principales

**Caso: general**

Sea  $\{X_n\}$  cadena de Markov homogénea con espacio de estados  $(E, \xi)$  y kernel de transición  $p$ .

**Teorema 7.1.1 ([4] p. 701)** *Bajo las hipótesis 7.1.1, la cadena  $\{X_n\}$ :*

a) *tiene una única distribución invariante  $\pi$  que cumple*

$$\int V d\pi < \infty, \quad y$$

b) es  $V$ -uniformemente ergódica, es decir, se cumple la siguiente desigualdad,

$$\|p^n(x, \cdot) - \pi(\cdot)\|_V \leq MV(x)(\gamma)^n,$$

donde  $\rho$  y  $M$  son las constantes definidas en la subsección 7.1.1.

**Demostración.** Sea  $C$  el conjunto pequeño involucrado en la propiedad de minorización. La prueba que se muestra a continuación se realizará en dos casos:

**Caso 1.**  $C$  es un átomo.

Con esta suposición, en la propiedad de minorización podemos tomar  $\tilde{\delta} = 1$  y  $\nu = p(a, \cdot)$ , para algún punto  $a \in C$  fijo. Sea  $\tau$  el tiempo de regeneración cuando la cadena alcanza a  $C$ , es decir,

$$\tau = \inf\{n \geq 1 : X_n \in C\},$$

y definimos

$$u_n = P_a(X_n \in C),$$

para  $n \geq 0$ . Así  $u_n$  es una sucesión de renovación cuya sucesión de incrementos es  $b_n = P_a(\tau = n)$ .

Definimos funciones  $G(r, x)$  y  $H(r, x)$  por:

$$\begin{aligned} G(r, x) &= E_x[r^\tau], \\ H(r, x) &= E_x \left[ \sum_{n=1}^{\tau} r^n V(X_n) \right], \end{aligned}$$

para todo  $x \in E$  y todo  $r > 0$ .

Observemos que, aplicando el resultado D.2.1, obtenemos:

$$\begin{aligned} \sum_{n=1}^{\infty} b_n \lambda^{-n} &= E_a [\lambda^{-\tau}] \\ &= G(\lambda^{-1}, a) \\ &\leq \lambda^{-1} K. \end{aligned}$$

Aplicando la propiedad de aperiodicidad, tenemos que  $b_1 = p(a, C) \geq \delta$ .

Así aplicando el Teorema D.1.2, obtenemos:

$$\sup_{|z| \leq r} \left| \sum_{n=0}^{\infty} (u_n - u_{\infty}) z^n \right| \leq K_1,$$

donde  $1 < r < R_1$ ,  $R_1$  y  $K_1$  son las contantes descritas en la definición 7.1.1.

Sustituyendo este resultado en el de inciso (iv) del Teorema D.2.1 y después aplicando el resultado auxiliar D.2.1:

$$\frac{G(r, x) - 1}{r - 1} \leq \frac{G(\lambda^{-1}, x) - 1}{\lambda^{-1} - 1} \leq \frac{\max(\lambda, K - \lambda)}{1 - \lambda} V(x),$$

Obtenemos

$$\sup_{|z| \leq r} \left| \sum_{n=0}^{\infty} \left( p^n g(x) - \int g d\pi \right) z^n \right| \leq MV(x),$$

y así

$$\left| p^n g(x) - \int g d\pi \right| \leq MV(x) r^{-n},$$

donde

$$\begin{aligned} M &= \frac{\max(\lambda, (K - \frac{\lambda}{\gamma}))}{D} + \frac{K(K - \frac{\lambda}{\gamma})}{\gamma DK_1} + \frac{(K - \frac{\lambda}{\gamma}) \max(\lambda, K - \lambda)}{D(1 - \lambda)} \\ &+ \frac{\lambda(K - 1)}{D(1 - \lambda)}. \end{aligned}$$

Por lo tanto tomando  $\rho = \frac{1}{R_1}$  y la fórmula para  $M$  es obtenida haciendo  $r = \frac{1}{\gamma}$

**Caso 2.** *C no es un átomo*

Como  $C$  no es un átomo, en la condición de minorización se tiene  $\tilde{\delta} < 1$ . Por lo tanto utilizando la construcción del algoritmo 3.1.1, podemos construir una cadena de Markov  $(X_n, I_n)$  con espacio de estados  $E \times \{0, 1\}$ , la cual cuenta con un átomo:  $E \times \{1\}$ .

Aplicando las ideas realizadas en el caso anterior a la cadena  $(X_n, I_n)$  con:

- Tiempo de paro

$$T = \min\{n \geq 1 : I_n = 1\}$$

- Sucesión de renovación  $\bar{u}_n = P_{(x,1)}(Y_n = 1)$ , con sucesión de incrementos  $\bar{b}_n = P_{(a,1)}(T = n)$ ,  $n \geq 1$  y  $a \in E$  fijo.
- y funciones

$$\begin{aligned} \bar{G}(r, x, i) &= E_{(x,i)} [r^T]. \\ \bar{H}(r, x, i) &= \sum_{n=1}^T r^n V(X_n), \end{aligned}$$

para todo  $x \in E$ ,  $i = 0, 1$  y  $r > 0$ .

Además, definimos para toda  $\zeta$ :

$$\bar{E}_x[\zeta] = \left[1 - \tilde{\delta}I_C(x)\right] E_{x,0}[\zeta] + \tilde{\delta}I_C(x)E_{x,1}[\zeta],$$

entonces se puede ver que  $\bar{E}_x$  coincide con  $E_x$  sobre  $\sigma\{X_n : n \geq 0\}$ . Ahora definimos

$$\begin{aligned}\bar{G}(r, x) &= \bar{E}_x[r^T]. \\ \bar{H}(r, x) &= \bar{E}_x\left[\sum_{n=1}^T r^n V(X(n))\right],\end{aligned}$$

Aplicando las técnicas usadas en el resultado auxiliar (parte 2), obtenemos

$$\begin{aligned}\sup_{|z| \leq r} \left| \sum_{n=1}^{\infty} \left( p^n g(x) - \int g d\pi \right) z^n \right| &\leq \bar{H}(r, x) \\ &+ \bar{G}(r, x) \bar{H}(r, a, 1) \sup_{|z| \leq r} \left| \sum_{n=0}^{\infty} (\bar{u}_n - \bar{u}_{\infty}) z^n \right| \\ &+ \bar{H}(r, a, 1) \frac{\bar{G}(r, x) - 1}{r - 1} \\ &+ \frac{\bar{H}(r, a, 1) - r \bar{H}(1, a, 1)}{r - 1} < \infty,\end{aligned}$$

para todo  $r > 1$ .

Ahora se extenderá las estimaciones realizadas para  $G(r, x)$  y  $H(r, x)$  del caso anterior, pero para este caso, las funciones correspondientes serán  $\bar{G}(r, x, i)$  y  $\bar{H}(r, x, i)$ .

Definimos

$$\bar{G}(r) = \sup\{\bar{E}_{x,0}[r^T] : x \in C\}.$$

Observemos que la condición inicial  $(x, 0)$  para  $x \in C$ , representa que la cadena no se regenera.

Así  $\bar{G}(r)$  representa la contribución extra a  $\bar{G}(r, x, i)$  y  $\bar{H}(r, x, i)$ , las cuales ocurren cada vez que  $X_n \in C$ , pero falla cuando se tiene  $Y_n = 1$  ■

### Caso: reversible

**Teorema 7.1.2 ([4] p. 702)** *Bajo las hipótesis 7.1.1 y suponiendo que la cadena  $X = \{X_n\}$  es reversible, se cumple que:*

a) existe una única distribución invariante  $\pi$  que cumple

$$\int V d\pi < \infty, \quad y$$

b)  $X$  es  $V$ -uniformemente ergódica, es decir, se cumple la siguiente desigualdad,

$$\|p^n(x, \cdot) - \pi(\cdot)\|_V \leq MV(x)(\gamma)^n,$$

donde  $\rho$  y  $M$  son las constantes definidas en la subsección 7.1.1.

**Demostración.** Sea  $C$  el conjunto pequeño involucrado en la propiedad de minorización. La prueba que se muestra a continuación se realizará en dos casos:

**Caso 1.**  $C$  es un átomo

Con esta suposición, en la propiedad de minorización podemos tomar  $\tilde{\delta} = 1$  y  $\nu(\cdot) = p(a, \cdot)$ , para algún punto  $a \in C$  fijo. Sea  $\tau$  el tiempo de regeneración cuando la cadena alcanza a  $C$ , es decir,

$$\tau = \inf\{n \geq 1 : X_n \in C\},$$

y definimos

$$u_n = P_a(X_n \in C),$$

para  $n \geq 0$ . Así  $u_n$  es una sucesión de renovación cuya sucesión de incrementos es  $b_n = P_a(\tau = n)$ .

Definimos funciones  $G(r, x)$  y  $H(r, x)$  por:

$$\begin{aligned} G(r, x) &= E_x[r^\tau], \\ H(r, x) &= E_x \left[ \sum_{n=1}^{\tau} r^n V(X_n) \right], \end{aligned}$$

para todo  $x \in E$  y todo  $r > 0$ .

Observemos que aplicando el resultado D.2.1, obtenemos:

$$\begin{aligned} \sum_{n=1}^{\infty} b_n \lambda^{-n} &= E_a [\lambda^{-\tau}] \\ &= G(\lambda^{-1}, a) \\ &\leq \lambda^{-1} K. \end{aligned}$$

Aplicando la propiedad de aperiodicidad, tenemos que  $b_1 = p(a, C) \geq \delta$ .

Así, aplicando el Teorema D.1.2, obtenemos:

$$\sup_{|z| \leq r} \left| \sum_{n=0}^{\infty} (u_n - u_{\infty}) z^n \right| \leq K_1,$$

donde  $1 < r < R_1$ ,  $R_1$  y  $K_1$  son las constantes descritas en la definición 7.1.1.

Sustituyendo este resultado en el de inciso (iv) del Teorema D.2.1 y después aplicando el resultado auxiliar D.2.1:

$$\frac{G(r, x) - 1}{r - 1} \leq \frac{G(\lambda^{-1}, x) - 1}{\lambda^{-1} - 1} \leq \frac{\text{máx}(\lambda, K - \lambda)}{1 - \lambda} V(x),$$

obtenemos

$$\sup_{|z| \leq r} \left| \sum_{n=0}^{\infty} \left( p^n g(x) - \int g d\pi \right) z^n \right| \leq MV(x),$$

y así

$$\left| p^n g(x) - \int g d\pi \right| \leq MV(x) r^{-n},$$

donde

$$\begin{aligned} M &= \frac{\text{máx}(\lambda, (K - \frac{\lambda}{\gamma}))}{D} + \frac{K(K - \frac{\lambda}{\gamma})}{\gamma DK_1} + \frac{(K - \frac{\lambda}{\gamma}) \text{máx}(\lambda, K - \lambda)}{D(1 - \lambda)} \\ &+ \frac{\lambda(K - 1)}{D(1 - \lambda)}. \end{aligned}$$

Por lo tanto tomando  $\rho = \frac{1}{R_1}$  y la fórmula para  $M$  es obtenida haciendo  $r = \frac{1}{\gamma}$ .

**Caso 2** *C no es átomo* Tenemos

$$\begin{aligned} \sum_{n=1}^{\infty} \bar{b}_n R^n &= \bar{G}(R, a, 1) \\ &\leq L(R), \end{aligned}$$

válida para todo  $1 < R < R_0$ , y podemos escoger  $R$  de tal manera que podamos aplicar el Teorema D.1.3. Si  $1 + 2\delta R_0 \geq L(R_0)$ , entonces  $L(R_0) < \infty$  y

$$\begin{aligned} \sum_{n=1}^{\infty} \bar{b}_n R_0^n &= \bar{G}(R_0, a, 1) \\ &\leq L(R_0). \end{aligned}$$

Este caso puede ocurrir sólo cuando  $R_0 = \frac{1}{\lambda} < (1 - \tilde{\delta})^{-1\alpha_1}$ . En otro caso, tomamos  $\tilde{R}$  como la única solución en el intervalo  $(1, R_0)$  de la ecuación

$$1 + 2\delta R = L(R).$$

Aplicamos el Teorema D.1.3, con  $R = \tilde{R}_0$ , para obtener  $R_2 = \tilde{R}$ . Así por la primera parte de el Teorema D.1.3, para  $1 < r < R_2$ , tenemos

$$\sup_{|z| \leq r} \left| \sum_{n=0}^{\infty} (\bar{u}_n - \bar{u}_{\infty}) z^n \right| \leq K_2,$$

para algún  $K_2 < \infty$ . Inicialmente no tenemos una estimación para  $K_2$ , pero el mismo método permite usar la segunda parte del Teorema D.1.3 y asegurar que  $K_2 = 1 + \frac{\sqrt{\tilde{\delta}r}}{(1 - \frac{r}{R_2})}$ . El resto de la prueba es similar a la presentada en la sección no atómico del teorema anterior.

Tomamos  $\rho = \frac{1}{R_2}$ , donde  $R_2 = \sup\{r < R_0 : 1 + 2r\delta \geq L(r)\}$  y

$$\begin{aligned} M &= \frac{\max\left(\lambda, \left(K - \frac{\lambda}{\gamma}\right)\right)}{D} + \frac{K\left(K - \frac{\lambda}{\gamma}\right)}{\gamma DK_1} + \frac{\left(K - \frac{\lambda}{\gamma}\right) \max(\lambda, K - \lambda)}{D(1 - \lambda)} \\ &+ \frac{\lambda(K - 1)}{D(1 - \lambda)}. \end{aligned}$$

■

### Caso: reversible y positiva

**Teorema 7.1.3 ([4] p. 702)** *Bajo las hipótesis 7.1.1 y suponiendo que la cadena  $X = \{X_n\}$  es reversible y positiva, se cumple que:*

a) *existe una única distribución invariante  $\pi$  que cumple*

$$\int V d\pi < \infty, \quad y$$

b)  *$X$  es  $V$ -uniformemente ergódica, es decir, se cumple la siguiente desigualdad,*

$$\|p^n(x, \cdot) - \pi(\cdot)\|_V \leq MV(x)(\gamma)^n,$$

*donde  $\rho$  y  $M$  son las constantes definidas en la subsección 7.1.1.*

**Demostración.** Es similar a la prueba anterior, sólo hay que cambiar el uso del Teorema D.1.3 por el Teorema D.1.4. ■

### 7.1.3. Ejemplos

#### Caminata aleatoria reflejada

Con este modelo ilustramos los teoremas 7.1.1 y 7.1.2.

Considérese la caminata aleatoria Bernoulli definida en  $\mathbb{Z}^+$ , con probabilidades de transición  $p(i, i-1) = p$ ,  $p(i, i+1) = q := 1-p$ , tal que  $p(0,0) = p$  y  $p(0,1) = q$ , con  $p > \frac{1}{2}$  y  $i \geq 1$ .

Tomando  $C = \{0\}$  y  $V(i) = \left(\frac{p}{q}\right)^{i/2}$ , obtenemos:

$$\begin{aligned}\lambda &= 2\sqrt{pq}, \\ K &= p + \sqrt{pq}, \\ \delta &= p \text{ y}, \\ \rho &= \frac{1}{R_1}, \text{ véase el Teorema 7.1.1 o,} \\ \rho &= \frac{1}{R_2}, \text{ véase el Teorema 7.1.2,}\end{aligned}$$

donde  $R_1$  y  $R_2$  dependen únicamente de  $\lambda$ ,  $\delta$  y  $K$ , respectivamente. El segundo valor de  $\rho$  se cumple, debido a que la caminata aleatoria es reversible.

En particular si  $p = \frac{2}{3}$ , tenemos por el Teorema 7.1.1 que  $\rho = 0,9994$ , y por el Teorema 7.1.2 que  $\rho = 0,9428$ .

#### Algoritmo Metropolis-Hastings

El propósito de este algoritmo es crear una cadena de Markov con distribución invariante  $\pi$ , y cuyo kernel de transición  $p$  sea reversible ( $\pi(y)p(y,x) = \pi(x)p(x,y)$ ) y simétrico, es decir,  $p(x,y) = p(y,x)$ .

El mérito de este algoritmo radica en que se inicia desconociendo la ley de distribución para  $\pi$ , sólo se conoce el comportamiento de las razones  $\frac{\pi(x)}{\pi(y)}$ , para todo  $x, y \in E$ .

La idea original de este algoritmo fue de Metropolis en [32]. Hastings en [16], generalizó dicho algoritmo y lo extendió a espacios de estados generales.

#### Descripción 7.1.1

- Se comienza con el kernel de transición  $Q$  de alguna cadena de Markov,

de tal manera que para todo  $x \in E$ ,  $Q(x, \cdot)$  es absolutamente continuo con respecto a la medida de Lebesgue. Denótese por  $q(x, y)$  a la derivada de Radon-Nikodym o la densidad de  $Q(x, \cdot)$  y la medida de Lebesgue.

- Definimos una probabilidad  $\alpha$ , con la cual estableceremos el kernel de transición  $p$ , de la siguiente manera:

$$\alpha(x, y) = \begin{cases} \min \left\{ \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)}, 1 \right\}, & \text{si } \pi(x)q(x, y) > 0, \\ 1, & \text{si } \pi(x)q(x, y) = 0. \end{cases}$$

$$p(x, y) = \begin{cases} q(x, y)\alpha(x, y), & \text{si } y \neq x, \\ \int q(x, y)[1 - \alpha(x, y)]dy, & \text{si } y = x. \end{cases}$$

Con lo cual obtenemos para  $A \in \xi$ ,

$$p(x, A) = \int_A p(x, y)dy + p(x, x)I_A(x).$$

En este caso, simularemos a  $\pi$  como normal estándar. Para este fin tomemos  $q = (x, \cdot) = N(x, 1)$  y

$$\alpha(x, y) = \min\{e^{-(y^2-x^2)/2}, 1\}.$$

$$\text{Así obtenemos } p(x, y) = \begin{cases} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y-x)^2}{2}\right), & \text{si } |x| \geq |y|, \\ \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y-x)^2 + y^2 - x^2}{2}\right), & \text{si } |x| < |y|. \end{cases}$$

Por la forma en que se construyó esta cadena, podemos ver que tiene la propiedad de reversibilidad y positividad.

Para comenzar a satisfacer las condiciones descritas en 7.1.1 usamos la función  $V(x) = e^{s|x|}$  y el conjunto pequeño  $C[-d, d]$ , con  $x, s, d \geq 0$ . Consideramos ahora

$$\begin{aligned}
\lambda(x, s) &:= \frac{pV(x)}{V(x)} \\
&= \exp\left(\frac{s^2}{2}\right) [\Phi(-s) - \Phi(-x - s)] + \exp\left(\frac{s^2}{2} - 2sx\right) \times \\
&\quad [\Phi(-x + s) - \Phi(-2x + s)] + \frac{1}{\sqrt{2}} \exp\left(\frac{(x-s)^2}{4}\right) \Phi\left(\frac{s-x}{\sqrt{2}}\right) + \\
&\quad \frac{1}{\sqrt{2}} \exp\left(\frac{x^2 - 6xs + s^2}{4}\right) \Phi\left(\frac{s-3x}{\sqrt{2}}\right) + \Phi(0) + \Phi(-2x) - \\
&\quad \frac{1}{\sqrt{2}} \exp\left(\frac{x^2}{4}\right) \left[ \Phi\left(\frac{-x}{\sqrt{2}}\right) + \Phi\left(\frac{-3x}{\sqrt{2}}\right) \right],
\end{aligned}$$

donde  $\Phi$  denota la función de distribución de la normal estándar. Entonces generamos:

$$\begin{aligned}
\lambda &= \min_{|x| \geq d} \lambda(x, s) = \lambda(d, s), \\
K &= \max_{|x| \leq d} pV(x) = PV(d) = e^{sd} \lambda(d, s) \text{ y} \\
b &= \max_{|x| \leq d} pV(x) - \lambda V(x) = PV(0) - \lambda V(0) = \lambda(0, s) - \lambda,
\end{aligned}$$

de donde se puede comprobar que:

$$pV(x) \leq \lambda V(x) + bI_C.$$

Por otro lado se debe de cumplir la condición de aperiodicidad fuerte, para esto definimos para  $A \in \xi$ :

$$\begin{aligned}
\delta &= \tilde{\delta} = \sqrt{2}e^{-d^2} \left[ \Phi(\sqrt{2}d) - \frac{1}{2} \right] \text{ y} \\
\nu(A) &= \int_{A \cap C} c e^{-x^2} dx.
\end{aligned}$$

Así tenemos que  $\tilde{\delta} > \delta$ .

Finalmente, definiendo como  $d = 1$  y  $s = 0,07$  se obtiene por el Teorema 7.1.2 que  $1 - \rho = 0,0091$ , y por el Teorema 7.1.3, tomando como  $d = 1,1$  y  $s = 0,16$  se tiene que  $1 - \rho = 0,0253$ .

## 7.2. Comentarios: teoría de Kartashov

Existe teoría desarrollada por Kartashov ([24] y [26]), en la cual exhibe una tasa de convergencia explícita utilizando una metodología similar y anterior a la expuesta en este capítulo (de hecho la teoría de Kartashov, data de 1985).

Kartashov generó esta idea a partir de Kalashnikov [21].

En general, usando teoría de funcionales de potencial, coeficientes ergódicos y kérneles particionantes (véase [25], [24] y [26]), Kartashov obtiene cuatro resultados de tasas de convergencia explícitas, en los cuales se puede notar que, sólo en uno de ellos presenta convergencia de tipo geométrico (uniforme y geométricamente ergódica, véase [25]).

Por otro lado, a partir de la teoría desarrollada por Kartashov en [24], bajo la métrica  $V$ -norma, Gordienko y Hernández-Lerma en [14], obtuvieron una tasa de convergencia explícita de tipo geométricamente ergódico. Con esto podemos pensar que en dos de los cuatro resultados obtenidos por Kartashov se obtiene orden de convergencia de tipo geométrico.

Terminamos esta sección de comentarios, presentando algunas diferencias observadas entre la metodología de Baxendale y los trabajos de Kartashov [24] y [26].

- Después de haber construido el proceso de renovación, Baxendale utiliza el teorema de Kendall para encontrar la tasa de convergencia, en cambio Kartashov usa teoría propia desarrollada en [26].
- Las tasas obtenidas por Kartashov son aplicables a una familia de métricas probabilísticas, en cambio los de Baxendale sólo se aplican a la  $V$ -norma.
- El orden de convergencia es geométrico en los resultados de Baxendale y de orden logarítmico y geométrico para los de Kartashov.

## Parte IV

### Conclusión y problemas abiertos



## Capítulo 8

# Conclusión y problemas abiertos

### 8.1. Conclusión

En esta tesis hemos considerado cadenas de Markov que tienen espacios de estados contenidos en  $\mathbb{R}$ . Además, se supuso que las cadenas cuentan con una distribución invariante, a la cual converge la función de transición en  $n$ -pasos (esta convergencia es en la métrica de la variación total o en la V-norma).

La idea general del trabajo consistió en estudiar distintas condiciones que permitieran establecer, de forma explícita, la tasa de convergencia de tipo geométrico, de la función de transición en  $n$ -pasos hacia la distribución invariante.

El estudio fue realizado considerando:

- *CMs con espacios finitos o compactos,*
- *CMs con kernel estocásticamente monótono, o*
- *CMs con la propiedad de reversibilidad y/o positividad.*

Los casos anteriores fueron analizados por medio de la aplicación de las siguientes tres técnicas:

#### **Método Espectral** ([47])

Este método fue utilizado en CMs con espacios de estados finitos, y utiliza las propiedades de los valores y vectores propios asociados la matriz de transición.

La tasa fue obtenida considerando la métrica de la variación total, además ésta es geoméricamente ergódica y no depende del estado inicial (en este sentido diremos que la tasa obtenida es **uniforme**). El método es ilustrado con un ejemplo.

**Acoplamiento**([30], [31], [35], [46], [47] y [56])

De manera general, este método consiste en encontrar cotas adecuadas para el lado derecho de la desigualdad de acoplamiento (véase Lema 3.1.3).

Utilizando este método se presentaron tres resultados referentes a tasas geométricas relacionadas con la métrica de la variación total. En principio, se consideraron dos resultados para CMs con espacios de estados finitos o compactos, y se obtuvieron tasas uniformemente ergódicas. Posteriormente, para el tercero, se consideraron CMs con espacios de estados de la forma  $[0, \infty)$ , y se obtuvieron tasas de convergencia de ambos tipos: ergódicas y no ergódicas.

Esta metodología fue ilustrada en el trabajo con cinco ejemplos. También fue incluida, en un apéndice (véase apéndice C), una aplicación de las tasas obtenidas a PDMs (en esta aplicación se consideró la teoría desarrollada para CMs con espacio de estados de la forma  $[0, \infty)$ ).

**Sucesiones Aperiódicas y Renovación** ([4])

Para este método se consideran CMs con espacios de estados contenidos en  $\mathbb{R}$ . La tasa fue obtenida mediante propiedades estructurales de las CMs involucradas (es decir, propiedades como: una condición tipo Foster-Lyapunov, la recurrencia y la existencia de conjuntos pequeños o de átomos), y mediante el acotamiento de una serie de potencias compleja, obtenida a partir de una sucesión de renovación aperiódica (el acotamiento fue hecho usando el teorema de Kendall).

Principalmente, en este método, se han obtenido tasas de tipo geoméricamente ergódicas con respecto a la  $V$ -norma.

El método se ejemplificó por medio de la caminata aleatoria y del algoritmo Metropolis-Hastings.

Para finalizar esta sección, presentamos una tabla que proporciona un panorama general del conjunto de hipótesis que permiten obtener tasas de convergencia de tipo geométrico.

También para facilitar la lectura de la tabla, después de ésta, ofrecemos un resumen de las principales condiciones utilizadas en el trabajo.

Sean  $\lambda^*$  el máximo de los valores propios de cierta la matriz de transición  $p$ ,  $C$  es un conjunto pequeño,  $c \in \mathbb{R}$ ,  $x \in E$  y  $n \in \mathbb{N}$ .

Ubicación	Forma del espacio		Irreducibilidad	Aperioidicidad	Positiva	Harris Recurrente	Minorización	Fuertemente aperiódica.	Foster-Lyapunov	Monotonicidad estocástica.	Reversibilidad	Positividad	Condición particular
	Finito	General											
§ 6.1.1	✓							✓					$\lambda^* < 1$
§ 6.1.2		Compacto						✓					
§ 6.2.1		Con mínimo					$C = \{0\}$	✓	✓	✓			
§ 6.2.2		Con mínimo	✓	✓			$C = \{0\}$	✓	✓	✓			$p^n(x, 0) > 0$
§ 7.2.2	✓	✓					✓	✓	✓				
§ 7.2.2	✓	✓					✓	✓	✓		✓		
§ 7.2.2	✓	✓					✓	✓	✓		✓	✓	

Cuadro 8.1: Condiciones para tener convergencia de tipo geométrico.

**Resumen de condiciones**

I Decimos que una cadena de Markov es  $\Psi$ -irreducible si existe una medida  $\phi$  tal que  $\phi \ll \Psi$ , y para todo  $A \in \mathcal{E}$  y  $x \in \xi$  se tiene  $p^n(x, A) > 0$ , para algún  $n \in \mathbb{N}$ .

II Una cadena  $\psi$ -irreducible que cuenta con una distribución invariante, es llamada **positiva**; si no la tiene es llamada **nula**. Cuando la distribución invariante es única es llamada **ergódica**.

III Sea  $p$  un kernel de transición. Decimos que  $p$  satisface la **condición de minorización**, si existen  $m_0 \in \mathbb{N}$ ,  $\delta > 0$ , una función  $S$  que es  $\xi/\mathbb{B}(\mathbb{R}^+)$  medible y una probabilidad  $\nu$  definida en  $\xi$ , tal que:

$$p^{m_0}(x, \cdot) \geq \delta S(x)\nu(\cdot), \quad x \in E. \quad (8.1)$$

IV En el contexto de la definición anterior,  $X$  es **fuertemente aperiódica**, si para todo  $A \subseteq C$  y todo  $x \in A$  se tiene:

$$p(x, A) \geq \delta\nu(A).$$

V  $X$  es  $\varphi$ -**Harris recurrente**, si para todo  $B \in \xi$  con  $\varphi(B) > 0$ , se tiene:

$$P_x(X_n \in B, \text{ para algún } n) = 1, \quad x \in B.$$

VI Sea  $C$  conjunto pequeño,  $\lambda < 1$ ,  $b < \infty$  y  $V : E \rightarrow \mathbb{R}$ . Decimos que  $p$  satisface la **propiedad de Foster-Lyapunov** para  $V$ ,  $\lambda$ ,  $b$  y  $C$  si para  $x \in E$  se cumple:

$$pV(x) := \int_E V(y)p(x, dy) \leq \lambda V(x) + bI_C(x).$$

VII Diremos que una cadena de Markov es estocásticamente monótona, si su kernel de transición  $p$  cumple para todo  $z \in \mathbb{R}$ ,  $x, y \in E$  y  $x \leq y$  que  $p(x, (z, \infty)) \leq p(y, (z, \infty))$ .

VIII Una cadena de Markov es **reversible** con respecto a su medida invariante  $\pi$ , si

$$\int_E pf(x)g(x)d\pi(x) = \int_E f(x)pg(x)d\pi(x).$$

IX Una cadena de Markov es **positiva** si

$$\int_E (pf)(x)f(x) d\pi(x) \geq 0,$$

par toda  $f \in L^2(\pi)$ .

## 8.2. Problemas abiertos

Con base a la teoría estudiada se han detectado los siguientes problemas:

- I.- Realizar un análisis comparativo sobre las técnicas y/o condiciones que exhiben tasas de convergencia de tipo geométrico.
- II.- Para la familia de CMs proporcionados en [6], establecer condiciones que permitan obtener tasas explícitas de convergencia de tipo geométrico.
- III.- Analizar la estabilidad de los PDMs con espacios de estados  $E \subseteq \mathbb{R}$ , sin elemento mínimo, y con costo promedio esperado mediante la teoría de tasas de convergencia correspondiente.

**Observación 8.2.1** *Hasta donde sabemos no hay nada hecho con relación a III. A la fecha, únicamente se encuentra desarrollado el caso con elemento mínimo ([35]). Cabe señalar que III es un problema interesante debido a que en la teoría de PDMs existen ejemplos clásicos con  $E = \mathbb{R}$ , como el modelo lineal con costo cuadrático en una dimensión (véase [5]).*



**Parte V**

**Apéndices**



# Apéndice A

## Resultados básicos

Considérese CMs con espacio de estados  $(E, \xi)$ , kernel de transición  $p$  y distribución invariante  $\pi$ .

Este apéndice está dedicado a establecer la demostración de dos suposiciones que realizamos en la parte de tasas de convergencia, los cuales son:

- la existencia de conjuntos pequeños, y
- $\lim_{n \rightarrow \infty} \|p^n(x, \cdot) - \pi(\cdot)\|_{VT} = 0$ .

Los resultados, que aquí se presentan, fueron consultados de [1], [20] y [40].

### A.1. Existencia de conjuntos pequeños

Como se definió en 2.3.5, un conjunto  $C$  es llamado pequeño si el kernel de transición  $p$  satisface la condición de minorización  $M(m_0, \delta, I_C, \nu)$  (véase la definición 2.3.4). Por lo tanto, el objetivo de esta sección es exhibir a las constantes  $m_0, \delta$ , a la función indicadora  $I_C$  y a la probabilidad  $\nu$  que satisfagan  $M(m_0, \delta, I_C, \nu)$ . La prueba de este hecho, necesita los siguientes resultados.

Sea  $(\Omega, \mathcal{F})$  un espacio medible.

- (a) **Teorema A.1.1 (Radon-Nikodym, [1] p. 63)** Sean  $\mu$  una medida  $\sigma$ -finita y  $\lambda$  una medida con signo definidas en  $\mathcal{F}$ . Supóngase que  $\lambda \ll \mu$ ,

entonces existe una función  $f$  medible de  $\Omega$  en  $\mathbb{R}$  tal que

$$\lambda(B) = \int_B f(y)\mu(dy),$$

para todo  $B \in \mathcal{F}$ .

- (b) **Teorema A.1.2 (Descomposición de Lebesgue, [1] p. 68)** Sean  $\mu$  y  $\lambda$  dos medidas definidas en  $\mathcal{F}$ . Si  $\mu$  es  $\sigma$ -finita y  $\lambda$  es una medida con signo, entonces existen dos medidas con signo  $\lambda_{a.c.}$  y  $\lambda_{m.s.}$  definidas también en  $\mathcal{F}$  tal que

$$\begin{aligned} \lambda &= \lambda_{a.c.} + \lambda_{m.s.}, \\ \lambda_{a.c.} &\ll \mu \text{ y} \\ \lambda_{m.s.} &\perp \mu. \end{aligned}$$

**Corolario A.1.1** Si es  $\varphi$  una medida  $\sigma$ -finita en  $\xi$  y  $p^n(x, \cdot)$  es la medida con signo mencionada en el teorema anterior se tiene que para cada  $x \in E$  y  $n \in \mathbb{N}$  existe una medida  $\mathbf{p}^n(x, \cdot)$  y otra  $p_{m.s.}^n(x, \cdot)$  tal que  $p^n(x, \cdot) = \mathbf{p}^n(x, \cdot) + p_{m.s.}^n(x, \cdot)$ .

- (c) **Lema A.1.1 ([40] p. 5)** Bajo el contexto del corolario anterior se establece la siguiente desigualdad

$$\begin{aligned} \mathbf{p}^{m+n}(x, \cdot) &\geq \int p^m(x, dy)\mathbf{p}^n(y, z) \\ &\geq \int \mathbf{p}^n(x, y)\mathbf{p}^n(x, y)\varphi(dy). \end{aligned}$$

- (d) **Lema A.1.2 ([40] p. 5)** Bajo el contexto del corolario anterior y considerando  $A, B \in \xi^2$  y  $\varphi^3(A \circ B) > 0$  donde  $A \circ B := \{(x, y, z) \in E^3 : (x, y) \in A \text{ y } (y, z) \in B\}$ ,  $\varphi^3 := \varphi \times \varphi \times \varphi$ . Entonces, existen conjuntos  $C$  y  $D$   $\xi$ -medibles, tales que:

$$\inf_{x \in C, z \in D} \varphi(A_1(x) \cap B_2(z)) > 0.$$

$A_1(x), B_2(z)$  son la primera y segunda sección de los conjuntos  $A$  y  $B$ , respectivamente.

- (e) **Teorema A.1.3 (Diferenciación de medidas, [11] p. 344)** Sean  $\lambda$  y  $\mu$  dos medidas finitas sobre  $(E, \xi)$  tal que  $\lambda = \lambda_{a.c.} + \lambda_{m.s.}$  donde  $\lambda_{a.c.} \ll \mu$  y  $\lambda_{m.s.} \perp \mu$ . Entonces

$$\lim_{ni \rightarrow \infty} I_{\{\varphi(E_x^i)\}} \frac{\lambda(E_x^i)}{\mu(E_x^i)} = \frac{d}{d\mu} \lambda_{a.c.}$$

donde  $E_x^i$  es el único miembro de  $E_i$  que contiene a  $x$ .

Además utilizaremos la definición de la siguiente propiedad.

**Propiedad de separabilidad** Sea  $\{E_i\}_{i \in I}$  una partición finita de  $E$  con la propiedad de que si  $A \subseteq E_i$  entonces  $A$  es unión finita de elementos de  $E_{i+1}$ .

Ahora presentamos el resultado principal de la sección.

**Teorema A.1.4 (Jain y Jamison)** *Sea  $X$  cadena de Markov  $\varphi$ -irreducible con espacio de estados  $(E, \xi)$  y kernel de transición  $p$ . Entonces existe un conjunto pequeño  $C$  con la propiedad de  $\varphi(C) > 0$ .*

**Demostración.**

Obsérvese que para todo  $n \in \mathbb{N}$  y  $x \in E$ ,  $p^n(x, \cdot)$  es una medida de probabilidad finita (por lo tanto con signo). Entonces aplicando el Corolario A.1.1 y el Lema A.1.1 obtenemos,

$$\begin{aligned} p^n(x, B) &= \int_B \mathbf{p}^n(x, y) \varphi(dy) + p_{m.s.}^n(x, B) \\ &\geq \int_B \mathbf{p}^n(x, y) \varphi(dy). \end{aligned} \quad (\text{A.1})$$

Utilizaremos la desigualdad (A.1) para mostrar que para algunos conjuntos medibles  $C$  y  $D$  y  $\delta \in [0, 1]$  se tiene,

$$\mathbf{p}^n(x, y) \geq \delta, \quad x \in C, \quad y \in D, \quad (\text{A.2})$$

De este hecho tendremos

$$\begin{aligned} p^n(x, B) &\geq \int_B \mathbf{p}^n(x, y) \varphi(dy) \\ &\geq \delta \varphi(B), \end{aligned}$$

con lo cual se finalizaría la prueba, ya que estaríamos probando que  $p$  satisface  $M(n, \delta, I_C, I_D \varphi_n)$ .

Para establecer la desigualdad (A.2) se necesitan establecer los siguientes hechos:

- (a)  $\mathbf{p}^n(x, y)$  es una función medible de  $E \times E$  en  $\mathbb{R}$ .
- (b) Exhibir a  $C$  y  $D$ .

“(a)” Definimos

$$f_n(x, y) = \begin{cases} \frac{p(x, E_y^n)}{\varphi_n(E_y^n)} & \text{si } \varphi_n(E_y^n) > 0 \\ 0 & \text{si } \varphi_n(E_y^n) = 0 \end{cases}$$

Se puede verificar que  $f_n$  es  $\xi \times \xi$  medible, y al tomar el límite a las funciones  $f(n)$  cuando  $n$  tiende a infinito, se obtiene una función  $\xi \times \xi$  medible (Este límite existe, por el Lema A.1.3), la cuál es una versión de  $\mathbf{p}^n(x, y)$ .

“(b) ”La existencia de  $C$  y  $D$  la da el Lema A.1.2.

Ahora mostramos la desigualdad A.2.

Por la  $\varphi$ -irreducibilidad tenemos que, para algún  $x \in E$

$$\sum_n^{\infty} \mathbf{p}^n(x, y) > 0,$$

para casi todo  $y$  (con respecto a  $\pi$ ). Se sigue entonces que existen enteros  $m_1, m_2$  mayores o iguales a 1 tal que:

$$\int \int \int \mathbf{p}^n(x, y) \mathbf{p}^n(x, y) \varphi(dx) \varphi(dy) \varphi(dz) > 0,$$

la cual implica, que para  $\delta \in [0, 1]$  y  $\varphi^3(A \circ B) > 0$ , donde

$$\begin{aligned} A &= \{(x, y) : \mathbf{p}^{m_1}(x, y) \geq \delta\}, \\ B &= \{(y, z) : \mathbf{p}^{m_2}(x, y) \geq \delta\}. \end{aligned}$$

Considerando  $C, D$  como en el Lema A.1.2 y ocupando el Lema A.1.1 tenemos que para todo  $x \in C$  y  $z \in D$

$$\begin{aligned} \mathbf{p}^{m_1+m_2}(x, z) &\geq \int_{A_1(x) \cap B_2(z)} \mathbf{p}^{m_1}(x, y) \mathbf{p}^{m_2}(y, z) \varphi(dy) \\ &\geq \delta^2 \inf_{x \in C, z \in D} \varphi(A_1(x) \cap B_2(z)). \end{aligned}$$

Así, el kernel  $p$  satisface la propiedad de minorización

$$M \left( m_1 + m_2, \delta^2 \inf_{x \in C, z \in D} \varphi(A_1(x) \cap B_2(z)), I_C, I_D \cdot \varphi \right).$$

■

## A.2. Convergencia a la distribución invariante

A continuación mostramos la prueba que garantiza que una CM converge a la distribución invariante bajo la métrica de la variación total.

Cabe señalar que la metodología utilizada en esta prueba, es el acoplamiento y la regeneración (véase la sección 3.1.3).

**Teorema A.2.1** *Sea  $p$  kernel de transición de una cadena de Markov  $\{X_n\}_{n=0}^\infty$   $\varphi$ -irreducible, aperiódica y con **distribución invariante**  $\pi$ . Entonces:*

$$\lim_{n \rightarrow \infty} \|p^n(x, \cdot) - \pi(\cdot)\|_{VT} = 0, \quad \pi - c.s.$$

**Demostración.** Observemos que:

$$\begin{aligned} \lim_{n \rightarrow \infty} \|p^n(x, \cdot) - \pi(\cdot)\|_{VT} &= \lim_{n \rightarrow \infty} \left\| \int_E p^n(x, \cdot) \pi(dy) - \int_E p^n(y, \cdot) \pi(dy) \right\|_{VT} \\ &= \lim_{n \rightarrow \infty} \left\| \int_E [p^n(x, \cdot) - p^n(y, \cdot)] \pi(dy) \right\|_{VT} \\ &\leq \lim_{n \rightarrow \infty} \int_E \|p^n(x, \cdot) - p^n(y, \cdot)\|_{VT} \pi(dy). \end{aligned}$$

Por otro lado, por ser la cadena  $\{X_n\}_{n=0}^\infty$   $\varphi$ -irreducible, existe un conjunto pequeño  $C$ , tal que  $\varphi(C) > 0$  (véase Teorema A.1.4). Ocupando este conjunto y el algoritmo 3.1.1 generamos:

- un acoplamiento exacto  $(\{X'_n\}_{n=0}^\infty, \{X''_n\}_{n=0}^\infty)$  de  $\{X_n\}_{n=0}^\infty$ , donde podemos pensar que  $X'_0 \stackrel{D}{=} \delta_x$ ,  $X''_0 \stackrel{D}{=} \delta_y$  y
- una ley de probabilidad  $P_{(x,y)}$ .

Sea  $T$ , el tiempo de acoplamiento de  $\{X'_n\}_{n=0}^\infty$  y  $\{X''_n\}_{n=0}^\infty$ . Por lo tanto, usando las desigualdades A.3 y la de acoplamiento (véase Lema 3.1.3), tenemos:

$$\lim_{n \rightarrow \infty} \|p^n(x, \cdot) - \pi(\cdot)\|_{VT} \leq \lim_{n \rightarrow \infty} \int_E P_{(x,y)}(T > n) \pi(dy).$$

Consideremos

$$\begin{aligned} G &= \{(x, y) : P_{(x,y)}((X'_n, X''_n) \in C \times C \text{ i.o.}) = 1\}, \\ G_x &= \{y \in E : (x, y) \in G\} \\ \bar{G} &= \{x \in E : \pi(G_x) = 1\}, \end{aligned}$$

Con lo anterior, afirmamos lo siguiente:

- (a)  $\lim_{n \rightarrow \infty} P_{(x,y)}(T > n) = 0$ ,  $(x, y) \in G$  y
- (b)  $\pi(\bar{G}) = 1$ .

Con las suposiciones realizadas y el teorema de convergencia dominada de Lebesgue obtenemos lo requerido.

Ahora probemos las suposiciones hechas.

“(a)” Cuando  $(X_n, Y_n) \in C \times C$ , del algoritmo 3.1.1 tenemos que con probabilidad  $\delta$ ,  $X_m = Y_m$  para todo  $m \geq n + k_0$ . Así podemos definir  $T = n + k_0$  (tiempo de acoplamiento). Ahora si la cadena parte de  $(x, y) \in G$ ,

$$P_{(x,y)}((X_n, X'_n) \in C \times C \text{ i.o.}) = 1$$

esto implica que  $T$  será finito. Por lo tanto

$$\lim_{n \rightarrow \infty} P_{(x,y)}(T > n) = 0, \quad (x, y) \in G$$

“(b)” Supongamos lo contrario, es decir  $\pi(\overline{G}) < 1$ . Entonces podemos afirmar que  $(\pi \times \pi)(G) = 1$ . A continuación probamos este hecho:

Debido a que  $\varphi(C) > 0$ , tenemos que  $P_x(T_C < \infty) > 0$  y  $P_y(T_C < \infty) > 0$ , esto implica que existen  $n_x$  y  $n_y$  enteros positivos tales que:

$$\begin{aligned} p^{n_x}(x, C) &> 0, \\ p^{n_y}(y, C) &> 0. \end{aligned}$$

Sea  $n_*$  como en el Lema 3, con esto tenemos que

$$\begin{aligned} p^{n_x+m_0+n}(x, C) &> 0 \\ p^{n_y+m_0+n}(y, C) &> 0 \end{aligned}$$

cuando  $n \geq n_*$ . Sea  $l \geq \max(n_x, n_y) + m_0 + n$ , así se cumple:

$$P_{(x,y)}(T_{C \times C} \leq l) > 0$$

ya que  $P_{(x,y)}(T \leq l) > p^l(x, C)p^l(y, C) > 0$ . Este resultado es debido a que antes del acoplamiento las cadenas se comportan de manera independiente. En particular tenemos:  $P_{(x,y)}(T < \infty) > 0$ , y así tenemos

$$P_{(x,y)}(T_{C \times C} < \infty) = 1 \quad \text{p.c.s.,}$$

con  $(x, y) \notin C \times C$ . Por ser  $(\{X'_n\}_{n=0}^\infty, \{X''_n\}_{n=0}^\infty)$  cadena de Markov y como se cumple  $P_{(x,y)}(T_{C \times C} < \infty) = 1$ , tenemos que  $(\{X'_n\}_{n=0}^\infty, \{X''_n\}_{n=0}^\infty)$  es recurrente i.e.  $(\pi \times \pi)(G) = 1$ .

Por lo tanto tenemos:

$$\begin{aligned}\pi \times \pi(G^c) &= \int_E \pi(G^c)_x \pi(dx) \\ &= \int_E [1 - \pi(G_x)] \pi(dx) \\ &= \int_{\bar{G}} [1 - \pi(G_x)] \pi(dx) + \int_{\bar{G}^c} [1 - \pi(G_x)] \pi(dx) \\ &> 0.\end{aligned}$$

La última línea es debida a la definición de  $\bar{G}$ . Esto contradice la suposición, así  $\pi(\bar{G}) < 1$ . Por lo tanto  $\pi(\bar{G}) < 1$ . ■



## Apéndice B

# Teorema de Strassen

Este apéndice tiene como objetivo establecer la relación existente entre el ordenamiento estocástico (véase la definición 3.2.1 y el Lema 3.2.1 ) y el ordenamiento puntual de elementos aleatorios y procesos estocásticos. Para este fin, en el desarrollo de este apartado presentamos siete resultados.

El material que aquí se presenta fue consultado de [22], [29] y [52].

De inicio, presentamos el resultado que fundamenta el concepto de ordenamiento estocástico.<sup>1</sup>

**Teorema B.0.2 ([52] p.436)** Sean  $S, T$  dos espacios métricos separables,  $\mu$  y  $\nu$  probabilidades definidas en  $S$  y  $T$ , respectivamente.

Existe una medida de probabilidad  $\lambda$  en  $S \times T$ , con marginales  $\mu$  y  $\nu$  respectivamente, tal que  $\lambda(\omega) \geq 1 - \epsilon$ , con  $\omega \in S \times T$  si y solo si para todo abierto  $U$  contenido en  $T$ , se cumple:

$$\nu(U) \leq \mu(\pi_s(\omega \cap (S \times U))) + \epsilon,$$

donde  $\pi_s$ , es la proyección a  $S$ .

**Demostración.** Consultar Teorema 7 y 11 de [52]. ■

A continuación reescribiremos el teorema anterior, utilizando  $\epsilon = 0$ ,  $S = T = E$  y  $M := \{(x, x') \in E^2 : x \preceq x'\}$ .

---

<sup>1</sup>Cabe mencionar que este resultado no presenta la definición actual de ordenamiento estocástico, más bien, de este resultado se obtiene.

**Corolario B.0.1** Sean  $\mu$  y  $\nu$  medidas de probabilidad sobre  $(E, \xi)$ . Entonces existe un acoplamiento  $\lambda$  para las medias  $\mu$  y  $\nu$ , respectivamente con la propiedad:  $\lambda(M) = 1$  si y solo si:

$$\nu \stackrel{est.}{\leq} \mu.$$

De hecho, ampliaremos la relación existente entre el ordenamiento estocástico y el ordenamiento puntual en el siguiente teorema.

**Teorema B.0.3 ([22], p. 900)** Sean  $\nu$  y  $\mu$  medidas de probabilidad en  $(E, \xi)$ . Las siguientes condiciones son equivalentes.

- (i)  $\nu \stackrel{est.}{\leq} \mu$ .
- (ii) Existe una medida  $\lambda \in \xi^2$  con marginales  $\nu$  y  $\mu$  respectivamente, con la propiedad de que  $\lambda(M) = 1$ .
- (iii) Existe una variable aleatoria definida en  $M$  y dos funciones medibles  $f$  y  $g$ , definidas de  $\mathbb{R}$  a  $E$ , tal que  $f \leq g$ ,  $\lambda_{f \circ Z} \stackrel{D}{=} \nu$  y  $\lambda_{g \circ Z} \stackrel{D}{=} \mu$ .
- (iv) Existen elementos aleatorios  $X_1, X_2$ , definidos de  $M$  en  $E$  tal que  $X_1 \leq X_2$ ,  $X_1 \stackrel{D}{=} \nu$  y  $X_2 \stackrel{D}{=} \mu$ .
- (v) Existe un kernel  $k$  sobre  $E \times \xi$  tal que:  $k(x, \{y : x \preceq y\}) = 1$  y  $\mu(\cdot) = \int k(x, \cdot) \nu(dx)$ .
- (vi)  $\nu(B) \leq \mu(B)$  para todo cerrado  $B \in C_{ind}$ .

**Demostración.** Mostraremos que  $i \Rightarrow ii \Rightarrow iii \Rightarrow iv \Rightarrow v \Rightarrow vi \Rightarrow 1$ .

$i \Rightarrow ii$  Es clara del corolario anterior.

$ii \Rightarrow iii$  Se puede ver que  $(M, \mathbb{B}(M), \lambda)$  y  $(B, \mathbb{B}, P)$  son isomorfos; donde  $B$  es un subconjunto de Borel de  $\mathbb{R}$ ,  $\mathbb{B}$  son los borelianos de  $B$  y  $P$  es una medida definida en  $(B, \mathbb{B})$ . Sea  $Z$  el isomorfismo,  $\pi_i$  la proyección en la  $i$ -ésima coordenada y definimos  $f = \pi_1 \circ Z^{-1}$ ,  $g = \pi_2 \circ Z^{-1}$ . Así obtenemos:

$$\begin{aligned} \lambda_{f \circ Z} &\stackrel{D}{=} \nu, \\ \lambda_{g \circ Z} &\stackrel{D}{=} \mu. \end{aligned}$$

Sea  $x \in B$ ,  $Z^{-1}(x) = (x_0, x'_0)$  donde  $x_0 \preceq x'_0$ . Así vemos que  $f(x) \leq g(x)$ .

$iii \Rightarrow iv$  Definiendo  $X_1 = f \circ Z$ ,  $X_2 = g \circ Z$ , se obtiene lo requerido.

$iv \Rightarrow v$  Ver Lindvall.

$v \Rightarrow vi$  Sea  $B \in C_{ind}$  cerrado, si  $x \in B$  entonces  $k(x, B) = 1$ . Así

$$\begin{aligned}\mu(B) &= \int_E k(x, B)\nu(dx) \\ &\geq \int_B k(x, B)\nu(dx) \\ &= \nu(B).\end{aligned}$$

$vi \Rightarrow i$ . Sea  $\epsilon > 0$ ,  $A \in C_{ind}$ , entonces existe un subconjunto compacto  $K_\epsilon$  de  $A$  tal que,  $\nu(A) - \nu(K_\epsilon) < \epsilon$ .  $H = \{y \in E : x \preceq y \text{ para algún } x \in K_\epsilon\}$ . Se puede ver que  $H$  es cerrado y pertenece a  $C_{ind}$ , así  $\nu(H) - \mu(H) < \epsilon$ . Como  $K_\epsilon \subseteq H \subseteq A$ , por lo tanto:

$$\nu(A) \leq \nu(H) + \epsilon \leq \mu(H) + \epsilon \leq \mu(A) + \epsilon.$$

■

El siguiente teorema es el que nosotros llamamos de Strassen para elementos aleatorios.

**Teorema B.0.4 (Strassen)** Sean  $\nu$  y  $\mu$  medidas de probabilidad en  $(E, \xi)$ .  $\nu \stackrel{est.}{\leq} \mu$  si y solo si existen elementos aleatorios  $Z$  y  $Z'$  cuyo espacio imagen es  $(E, \xi)$  tal que  $Z \stackrel{D}{=} \nu$ ,  $Z \stackrel{D}{=} \mu$  y  $Z \leq Z'$ .

**Demostración.**  $\Rightarrow$ ) Es consecuencia de B.0.1.

$\Leftarrow$  Es consecuencia directa de los resultados anteriores. ■

A continuación presentamos lo necesario para establecer el teorema de Strassen para procesos estocásticos.

**Hipótesis B.0.1** Sean  $(E_i, \xi_i)$  espacios medibles tales que  $(E_i, \preceq_i)$  es espacio con orden parcial,  $i \in \mathbb{N}$ .  $\preceq_k^*$  es un orden parcial en  $\prod_{j=0}^k$  definido de la siguiente manera:

$$\begin{aligned}(x_0, \dots, x_k) &\preceq_k^* (x'_0, \dots, x'_k) \text{ si y solo si} \\ x_i &\preceq_i x'_i, \text{ para todo } i \in \{1, 2, \dots, k\}.\end{aligned}$$

**Lema B.0.1** Bajo la hipótesis B.0.1, consideremos dos medidas  $\lambda, \mu$  definidas en  $(E_0, \xi_0)$  con la propiedad de que  $\lambda \stackrel{est.}{\preceq} \mu$ ; kernels de transición  $k_i, k'_i$  definidos en  $\prod_{j=0}^{i-1} E_j \times \xi_i$  de tal manera que  $k_i \stackrel{est.}{\preceq}_i k'_i$ , para todo  $i \in \{1, \dots, n\}$ . Entonces

$$P_\lambda^n \stackrel{est.}{\preceq}_n^* P_\mu^n,$$

donde  $P_\lambda^n$  y  $P_\mu^n$  son las medidas definidas en  $\prod_{i=0}^n \xi_i$  obtenidas con el teorema de la medida producto, considerando como medidas iniciales a  $\lambda$  y  $\mu$ , respectivamente.

**Demostración.** Este resultado es probado por inducción.

Para  $n = 1$ , consideremos una función  $f$ , creciente, acotada y definida en  $E_0 \times E_1$ .

$$\int_{E_0 \times E_1} f(x_0, x_1) P_\lambda^1(d(x_0, x_1)) = \int_{E_0} \left[ \int_{E_1} f(x_0, x_1) k_1(x_0, dx_1) \right] \lambda(dx_0)$$

definiendo como:

$$\begin{aligned} g(x_0) &= \int_{E_1} f(x_0, x_1) k_1(x_0, dx_1), \\ g'(x_0) &= \int_{E_1} f(x_0, x_1) k'_1(x_0, dx_1), \end{aligned}$$

podemos ver que  $g$  y  $g'$  son funciones crecientes, acotadas en  $E_0$  y  $g \preceq_0 g'$ , debido al orden estocástico entre  $k_1$  y  $k'_1$ . Además por el orden estocástico entre las medidas iniciales tenemos:

$$\begin{aligned} \int_{E_0 \times E_1} f(x_0, x_1) P_\lambda(d(x_0, x_1)) &\preceq_2^* \int_{E_0} \left[ \int_{E_1} f(x_0, x_1) k'_1(x_0, dx_1) \right] \mu(dx_0) \\ &= \int_{E_0 \times E_1} f(x_0, x_1) P_\mu(d(x_0, x_1)). \end{aligned}$$

Ahora supongamos válida la propiedad para  $n - 1$ , es decir,  $P_\lambda^{n-1} \preceq_{n-1}^{* \text{ est.}} P_\mu^{n-1}$ . Consideremos  $f$  creciente, acotada y definida en  $\prod_{j=0}^n E_j$

$$\begin{aligned} \int_{\prod_{j=0}^n E_j} f(x_0, \dots, x_n) P_\lambda^n(d(x_0, \dots, x_n)) &= \\ \int_{\prod_{j=0}^{n-1} E_j} \int_{E_n} [f(x_0, \dots, x_n) k_n((x_0, \dots, x_{n-1}), dx_n)] P_\lambda^{n-1}(d(x_0, \dots, x_n)) \end{aligned}$$

podemos observar que al definir

$$\begin{aligned} g(x_0, \dots, x_{n-1}) &= \int_{E_n} f(x_0, \dots, x_n) k_n((x_0, \dots, x_{n-1}), dx_n) \\ g'(x_0, \dots, x_{n-1}) &= \int_{E_n} f(x_0, \dots, x_n) k'_n((x_0, \dots, x_{n-1}), dx_n), \end{aligned}$$

obtenemos funciones crecientes y acotadas en  $\prod_{j=0}^{n-1} E_j$  tales que  $g \preceq_{n-1}^* g'$  (por el orden estocástico entre  $k_n$  y  $k'_n$ ). Así, aplicando la hipótesis de inducción

obtenemos:

$$\begin{aligned} & \int_{\prod_{j=0}^n E_j} f(x_0, \dots, x_n) P_\lambda^n(d(x_0, \dots, x_n)) \preceq_n^* \\ & \int_{\prod_{j=0}^{n-1} E_j} \left[ \int_{E_n} f(x_0, \dots, x_n) k'_n((x_0, \dots, x_{n-1}), dx_n) \right] P_\mu^{n-1}(d(x_0, \dots, x_n)) \\ & = \int_{\prod_{j=0}^n E_j} f(x_0, \dots, x_n) P_\mu^n(d(x_0, \dots, x_n)) \end{aligned}$$

Por lo tanto tenemos  $P_\lambda^n \stackrel{\text{est.}}{\preceq_n^*} P_\mu^n$

■

Para el siguiente resultado consideremos  $\pi_n : \prod_{j=0}^\infty E_j \rightarrow \prod_{j=0}^n E_j$  como  $\pi_n(x_0, x_1, \dots) = (x_0, \dots, x_n)$ .

**Lema B.0.2** *Bajo la hipótesis B.0.1 y considerando medidas de probabilidad  $P$  y  $P'$  sobre  $(\prod_{j=0}^\infty E_j, \prod_{j=0}^\infty \xi_j)$  tal que  $P_{\pi_n} \stackrel{\text{est.}}{\preceq_n^*} P'_{\pi_n}$  entonces*

$$P \stackrel{\text{est.}}{\preceq_\infty^*} P'.$$

**Demostración.** Sea  $f$  una función creciente y acotada definida sobre  $\prod_{j=0}^\infty E_j$ ,  $z := (z_1, z_2, \dots) \in \prod_{j=0}^\infty E_j$ .

Definamos  $f_n$  y  $\pi_n^*$  de la siguiente manera:

$$\begin{aligned} \pi_n^*(x_0, \dots, x_n) &= (x_0, \dots, x_n, z_{n+1}, \dots), \\ f_n &\equiv f \circ \pi_n^*. \end{aligned}$$

Con esto, podemos darnos cuenta que:

- (i)  $f_n$  es creciente y acotada.
- (ii)  $f_n \circ \pi_n \uparrow f$  c.s., con respecto a  $P$  y  $P'$ , respectivamente.

Entonces por el orden estocástico que se tiene para  $P \stackrel{\text{est.}}{\preceq_\infty^*} P'$  y los teoremas de

cambio de variable y convergencia monótona, tenemos:

$$\begin{aligned} \int f dP &= \lim_{n \rightarrow \infty} \int f_n \circ \pi_n dP \\ &\leq \lim_{n \rightarrow \infty} \int f_n \circ \pi_n dP' \\ &= \int f dP', \end{aligned}$$

así el tenemos el resultado requerido. ■

**Teorema B.0.5 (Strassen)** Sean  $(E_i, \xi_i)$  espacios medibles dotados con un orden parcial  $\preceq_i$ ,  $i \in \mathbb{N}$ ;  $\lambda$  y  $\mu$  medidas de probabilidad en  $E_0$ ,  $k_i, k'_i$  kernels de transición en  $\prod_{j=0}^{i-1} E_j \times \xi_i$ ,  $i \in \mathbb{N}$ .

Si  $\lambda \stackrel{est.}{\leq} \mu$  y  $K_i \stackrel{est.}{\leq} K'_i$  para toda  $i$  entonces existen procesos estocásticos  $Z = \{Z_n\}_{n=0}^\infty$ ,  $Z' = \{Z'_n\}_{n=0}^\infty$  en  $(\prod_{i=0}^\infty E_i, \prod_{i=0}^\infty \xi_i)$  tal que:

- (a)  $(Z_0, \dots, Z_n) \stackrel{D}{=} P_\lambda^n$ ,  $(Z_0, \dots, Z_n) \stackrel{D}{=} P_\mu^n$  para toda  $n$ .  
 (b)  $Z_n \preceq_n Z'_n$  para toda  $n$ .

donde  $P_\lambda^n$  y  $P_\mu^n$  son las medidas de probabilidad definidas en  $\prod_{i=0}^n \xi_i$ , obtenidas por el teorema de la medida producto, cuando las medidas iniciales son  $\lambda$  y  $\mu$ , respectivamente.

**Demostración.** Al considerar los kernels de transición  $\{k_i\}_{i \in \mathbb{N}}$ ,  $\{k'_i\}_{i \in \mathbb{N}}$ , las medidas iniciales  $\lambda$  y  $\mu$  como en los lemas anteriores, el teorema de Ionescu-Tulcea garantiza la existencia de procesos estocásticos  $\{Z_n\}_{n=0}^\infty$ ,  $\{Z'_n\}_{n=0}^\infty$  definidos en  $(\prod_{j=0}^\infty E_j, \prod_{j=0}^\infty \xi_j, P_\lambda)$  tal que:

$$\begin{aligned} (Z_0, \dots, Z_n) &\stackrel{D}{=} P_\lambda^n \stackrel{D}{=} P_{\lambda \pi_n} \\ (Z'_0, \dots, Z'_n) &\stackrel{D}{=} P_\mu^n \stackrel{D}{=} P_{\mu \pi_n}. \end{aligned}$$

Por otro lado, por ser  $k_i \stackrel{est.}{\preceq}_i k'_i$ , el Lema B.0.1 garantiza

$$P_{\lambda \pi_n} \stackrel{est.}{\preceq}_n^* P_{\mu \pi_n},$$

así, por el Lema B.0.2 se tiene:

$$P_\lambda \stackrel{est.}{\preceq}_\infty^* P_\mu.$$

Por lo tanto la versión de Strassen para elementos aleatorios, garantiza la existencia de procesos estocásticos  $X = \{X_n\}_{n=0}^{\infty}$  y  $X' = \{X'_n\}_{n=0}^{\infty}$  tales que

$$X_n \preceq X'_n, \quad \forall n \in \mathbb{N},$$

$$X \stackrel{D}{=} P_{\lambda}, \quad X' \stackrel{D}{=} P_{\mu}.$$

■



## Apéndice C

# Procesos de decisión de Markov

Este apéndice está basado en [35], y en él se trabajan con los procesos estocásticos que modelan la siguiente situación.

Supóngase que en el tiempo  $n$  ( $n = 1, 2, 3, \dots$ ), el proceso se encuentra en el estado  $x$ , y se va a tomar una acción que va a repercutir en la visita posterior del proceso a los estados existentes, generando un costo por tomar dicha acción. Además la transición de ir de un estado  $x$  a un conjunto de estados, dependerá de la acción tomada.

De la descripción anterior podemos mencionar que el proceso necesita los siguientes elementos:

Un espacio de estados  $E$ , un conjunto de acciones  $A$ , las acciones disponibles para el estado  $s$  denotadas por  $A(s)$ , un kernel de transición  $Q$  y una función de costo, denotada por  $c$ .

**Definición C.0.1** *A la quintupla  $(E, A, \{A(s) : s \in E\}, Q, c)$  se le conoce como **modelo de control de Markov (MCM)** donde  $E$  y  $A$  son espacios de Borel (es decir, subconjuntos medibles de espacios métricos, separables y completos). Para todo  $s \in E$ ,  $A(s)$  es un conjunto medible y no vacío, contenido en  $A$ ,  $Q$  es un kernel de transición sobre  $E$  dado  $\mathbb{K}$ , donde:*

$$\mathbb{K} = \{(x, a) : x \in E, a \in A(x)\},$$

*y  $c : \mathbb{K} \rightarrow \mathbb{R}$  es una función medible, llamada función de costo.*

Dado un MCM se necesita una regla ó una estrategia que indique la manera de

seleccionar las acciones en cada tiempo.

A continuación mostramos el contexto necesario para definir una política de control, la cual realizará la selección previamente mencionada.

Sea  $\mathbb{F} = \{f : E \rightarrow A \mid f \text{ es medible, y } f(x) \in A(x)\}$ . Las historias admisibles hasta el tiempo  $t$  quedan definidas de la siguiente manera.  $H_0 = E$  y para  $n \geq 1$   $H_n = \mathbb{K}^n \times E$ .

### Definición C.0.2

- (a) Una **política de control** es una sucesión  $\pi = \{\pi_n\}_{n=0}^\infty$  de kérneles estocásticos definidos sobre  $E$  dado  $H_n$  tal que:

$$\pi_t(h_n, A(x_n)) = 1,$$

para todo  $n \geq 0$  y toda  $h_n \in H_n$ . La clase de toda las políticas de control será denotada por  $\Pi$ .

- (b) Una política de control  $\pi = \{\pi_n\}$  es **estacionaria** si existe  $f \in \mathbb{F}$  tal que

$$\pi_n(h_n, \{f(x_n)\}) = 1,$$

para todo  $n \geq 0$  y toda  $h_n \in H_n$ . En este caso se identifica a  $\pi$  con  $f$ , y la clase de las políticas estacionarias coincide con  $\mathbb{F}$ .

**Observación C.0.1** Obsérvese que  $\mathbb{F} \subset \Pi$ .

Con una política de control  $\pi$  y una medida inicial  $\nu$ , el teorema de Ionescu-Tulcea garantiza la existencia de un espacio de probabilidad  $((E \times A)^\infty, \sigma(\prod_{i=1}^\infty \sigma(E \times A)), P_\nu^\pi)$  y un proceso estocástico  $\{X_n^{\nu, \pi}\}_{n=0}^\infty$ . Este proceso es llamado **proceso de decisión de Markov** (PDM) (ó proceso de control de Markov).

Obsérvese que un PDM, modela lo descrito anteriormente, y además **no es necesariamente un cadena de Markov**, ya que la política en un tiempo  $n$  puede depender de  $h_n$ .

**Proposición C.0.1** Si  $(E, A, \{A(s) : s \in E\}, Q, c)$  es M.C.M.,  $\nu$  medida inicial y  $\pi$  política estacionaria entonces  $\{X_t^\pi\}_{t=0}^\infty$  es un proceso de Markov con kérnel de transición estacionario definido por

$$Q_f(x, B) := Q((x, f(x)), B),$$

para toda  $x \in E$  y  $B \subset E$  medible.

**Demostración.** Véase [17], página 19. ■

## C.1. Problema de control óptimo (PCO)

Sea  $\mathbb{P} = (E, A, \{A(s) \mid s \in E\}, Q, c)$ , un modelo de control de Markov fijo.

### Definición C.1.1

(a) Sean  $\pi$  una política y  $\delta_x$  una medida inicial. Llamamos **el costo promedio esperado** a

$$J(\pi, x) := \limsup_{n \rightarrow \infty} \frac{E_x \left[ \sum_{j=0}^{n-1} c(x_j, a_j) \right]}{n}.$$

(b) Una política  $\pi^*$  es llamada **óptima** si

$$J(\pi^*, x) = \inf_{\pi \in \Pi} J(\pi, x),$$

para toda  $x \in E$ .

### Observaciones C.1.1

- La existencia de la política óptima es garantizada en [13]. Supondremos como hipótesis este hecho, en el desarrollo del trabajo.
- También en [13], se señala una caracterización para el costo promedio de un política estacionaria  $f \in \mathbb{F}$ , cuando la cadena de Markov  $\{X_n^f\}$  tiene una medida invariante  $m_f$ . Esta caracterización es la siguiente,

$$J(f, x) = J(f) = \int c(y, f(y)) m_f(dy),$$

donde  $x \in E$ .

## C.2. Análisis del índice de estabilidad para procesos de decisión de Markov

### C.2.1. Problema y motivación

Consideremos  $E = [0, \infty)$  y  $\mathbb{P} = (E, A, \{A(s) \mid s \in E\}, Q, c)$  un M.C.M (véase la definición C.0.1), donde el kernel de transición  $Q$  no es conocido completamente.

Pediremos que este modelo  $\mathbb{P}$  satisfaga el **PCO** (ver C.1). Dadas las características del modelo planteado, la política óptima  $f^*$ , no se tiene de forma específica, pero existe. Una referencia para verificar este hecho la podemos encontrar en [13].

Para dar solución a este problema, aproximaremos  $f^*$  con la política óptima  $\tilde{f}^*$  del modelo  $\tilde{\mathbb{P}} = (E, A, \{A(s)|s \in E\}, \tilde{Q}, c)$ . Aquí  $\tilde{Q}$  si esta completamente determinado, y de alguna manera es una aproximación a  $Q$ .

Con esta tipo de solución se querrá saber que tan buena es  $\tilde{f}^*$  en el modelo  $\mathbb{P}$ . Matemáticamente hablando, se buscará que

$$\Delta(x) := J(f^*, x) - J(\tilde{f}^*, x) < \epsilon,$$

para todo  $\epsilon > 0$ , con  $x \in E$ .

A continuación presentamos un desarrollo donde mostramos que:

$$\Delta(x) \leq g_x(m(Q, \tilde{Q})) \text{ y}$$

$g_x(s) \rightarrow 0$ , cuando  $s \rightarrow 0$  y  $m$  es la métrica de la variación total.

### C.2.2. Desarrollo

Debido a que los modelos  $\mathbb{P}$  y  $\tilde{\mathbb{P}}$  tienen los mismos espacios de estados y acciones, se tiene como consecuencia que ambos poseen el mismo conjunto de políticas. Denotemos por  $\mathbb{F}$  el conjunto de políticas invariantes (obsérvese que  $\mathbb{F}$  coincide con el conjunto de políticas estacionarias).

Tomando dos políticas invariantes, digamos  $f$  y  $g$ , se sabe que se crean dos Cadenas de Markov  $X^f$  y  $X^g$ , que están definidas en los espacios  $(\Omega, F, P_\nu^f)$  y  $(\Omega, F, P_\lambda^g)$  respectivamente, donde  $\nu$  y  $\lambda$  son las distribuciones iniciales de cada cadena. A este tipo de P.D.Ms, se les impondrán condiciones para que cumplan las hipótesis del resultado de tasas de convergencia de la sección 6.2. A continuación se muestran.

**Hipótesis C.2.1** *Existe una política invariante  $g$ , para la cual se cumple lo siguiente:*

(i) *Se cumple la desigualdad de Foster-Lyapunov para las constantes  $\lambda$  y  $b$ , con  $0 < \lambda < 1$ ,  $0 \leq b < \infty$ .  $V$  es una función creciente*

(ii) *Si  $X_0^f = X_0^g$ , entonces*

$$X_n^f(\omega) \leq X_n^g(\omega),$$

*para toda  $f \in \mathbb{F}$ ,  $\omega \in \Omega$  y  $n = 1, 2, 3, \dots$*

(iii) Existe una constante  $s \geq 1$  tal que

$$\sup_{a \in A(x)} |c(x, a)| \leq [V(x)]^{\frac{1}{s}}, \quad x \in E.$$

### Hipótesis C.2.2

(i) Existe una política invariante  $\hat{f}$  la cual es óptima.

(ii) Para cada política invariante  $f \in \mathbb{F}$ , el costo promedio  $J(f, \cdot)$  es igual a la constante  $J(f)$ , donde

$$J(f) = \int c(y, f(y)) \pi_f(dy),$$

y  $\pi_f$  es la probabilidad invariante correspondiente.

**Lema C.2.1** Bajo las hipótesis C.2.1 y  $f \in \mathbb{F}$  se tiene:

(i) Existe una medida invariante  $\pi_f$  correspondiente al kernel  $Q$ .

(ii)  $\lim_{n \rightarrow \infty} \|X_n^f - \pi_f\|_{VT} = 0$ .

**Demostración.** Fíjese  $f \in \mathbb{F}$ , y sea  $g$  la política distinguida (véase la hipótesis C.2.1). Denotamos como  $\tau^f$  y  $\tau^g$ , los tiempos de retorno al estado  $x = 0$  de las realizaciones  $\{X_n^f\}$  y  $\{X_n^g\}$ , ambas con distribución inicial  $\delta_0$ .

Por el orden de las trayectorias, se tiene:

$$E_0[\tau^f] \leq E_0[\tau^g] < \infty,$$

para todo  $n = 1, 2, 3, \dots$ . Así por el Corolario 5.3 de [37],  $\{X_n^f\}$  es positiva y Harris recurrente. De aquí se sigue la existencia la medida invariante  $\pi_f$ . De lo realizado en la sección 6.2, es claro que  $\pi_f$  es el límite de las distribuciones de  $X_n^f$  en sentido de la métrica de la variación total. ■

**Teorema C.2.1** Supóngase que se establecen las hipótesis C.2.1, C.2.2. Sean  $f^*$  y  $\tilde{f}^*$  las políticas óptimas para los modelos  $\mathbb{P}$  y  $\tilde{\mathbb{P}}$ , respectivamente. Entonces

$$\Delta(x) \leq 2 \left( \frac{2b}{1-\lambda} + 2rh_x(r) + 1 \right) \delta^{\frac{s-1}{s}} \max\{1, \log_\rho(\delta)\},$$

donde  $\delta = \sup_{x \in E} \sup_{a \in A(x)} \|\tilde{Q}(\cdot|x, a) - Q(\cdot|x, a)\|$  y  $\rho = \frac{1}{r}$ .

**Demostración.** Los lectores interesados en revisar la prueba pueden hacerlo en [35]. ■

### C.2.3. Modelo autorregresivo con barrera reflejante

Sea  $E = [0, \infty)$  y  $A$  un subconjunto compacto del intervalo  $(0, \Theta]$ , con  $\Theta \in A$ . Definimos el conjunto de acciones disponibles como  $A(x) = A$ , para todo  $x \in E$  y los procesos de decisión de Markov para los modelos  $\mathbb{P}$  y  $\tilde{\mathbb{P}}$  de la siguiente manera:

$$\begin{aligned} X_0 &= \tilde{X}_0 \\ X_{n+1} &= (X_n + a_n \eta_n - \varepsilon_n)^+, \end{aligned} \quad (\text{C.1})$$

$$\tilde{X}_{n+1} = \left( \tilde{X}_n + a_n \tilde{\eta}_n - \tilde{\varepsilon}_n \right)^+, \quad (\text{C.2})$$

donde  $\{\eta_n\}$ ,  $\{\tilde{\eta}_n\}$ ,  $\{\varepsilon_n\}$  y  $\{\tilde{\varepsilon}_n\}$  son sucesiones de variables aleatorias independientes idénticamente distribuidas. Sus distribuciones son  $\eta$ ,  $\tilde{\eta}$ ,  $\varepsilon$  y  $\tilde{\varepsilon}$ , respectivamente.

Para obtener las hipótesis del Teorema C.2.1, se imponen las siguientes condiciones al modelo.

#### Hipótesis C.2.3

- (i)  $\eta$ ,  $\tilde{\eta}$ ,  $\varepsilon$  y  $\tilde{\varepsilon}$  tienen densidades continuas y acotadas, respectivamente.
- (ii) Para todo  $n = 0, 1, 2, \dots$  se tiene que  $\varepsilon_n$  y  $\eta_n$  son independientes, también lo son  $\tilde{\varepsilon}_n$  y  $\tilde{\eta}_n$ .
- (iii) Para  $\xi := \Theta\eta - \varepsilon$  y  $\tilde{\xi} := \Theta\tilde{\eta} - \tilde{\varepsilon}$ , se cumple:  $E[\xi] < 0$  y  $E[\tilde{\xi}] < 0$ .
- (iv) Existe  $r_0 > 1$  y  $\tilde{r}_0 > 0$  tal que:
  - a)  $E[r_0^\xi] < \infty$  y  $E[\tilde{r}_0^{\tilde{\xi}}] < \infty$
  - b)  $\frac{d}{dr_0} E[r_0^\xi] = \frac{d}{d\tilde{r}_0} E[\tilde{r}_0^{\tilde{\xi}}] = 0$ .
- (v) Existe una constante  $s \geq 1$  tal que:

$$\sup_{a \in A(x)} |c(x, a)| \leq [V(x)]^{\frac{1}{s}}, \quad x \in E.$$

donde:

- $V(x) = \begin{cases} \text{máx}\{E_x[r^{\tau_0}], E_x[r^{\tilde{\tau}_0}]\} & \text{si } x > 0, \\ 1 & \text{si } x = 0. \end{cases}$
- $\tau = \text{mín}\{n > 0 : X_n^g = 0\}$  y  $\tilde{\tau} = \text{mín}\{n > 0 : \tilde{X}_n^{g_1} = 0\}$
- $g$  y  $g_1$  políticas invariantes definidas para todo  $x \in E$  como  $g(x) = \Theta$  y  $g_1(x) = \Theta$ .

En [13] se establece que con las hipótesis (i), (ii) y (iii) de C.2.3 se cumplen las hipótesis C.2.1 del Teorema C.2.1.

Sustituyendo las políticas invariantes  $g$  y  $g_1$ , definidas previamente, en las ecuaciones C.1 y C.2, respectivamente tenemos:

$$\begin{aligned} X_{n+1} &= (X_n + \Theta\eta_n - \varepsilon_n)^+, \\ \tilde{X}_{n+1} &= (\tilde{X}_n + \Theta\tilde{\eta}_n - \tilde{\varepsilon}_n)^+, \end{aligned}$$

son cadenas de Markov Ordenadas por trayectorias. Por otro lado, para cada política  $f \in \mathbb{F}$ , se puede obtener:

$$\begin{aligned} (X_n + f(X_n)\eta_n - \varepsilon_n)^+ &\leq (X_n + \Theta\eta_n - \varepsilon_n)^+, \\ (\tilde{X}_n + f(\tilde{X}_n)\tilde{\eta}_n - \tilde{\varepsilon}_n)^+ \tilde{X}_{n+1} &= (\tilde{X}_n + \Theta\tilde{\eta}_n - \tilde{\varepsilon}_n)^+. \end{aligned}$$

De aquí la hipótesis (ii) de C.2.2 se establece.

Ahora usando la hipótesis (iv) de C.2.3 se obtiene que:

$$\begin{aligned} E_0[r^{\tau_0}] &< \infty, \quad \text{para } 1 < r < \frac{1}{E[r_0^\xi]} \quad \text{y} \\ E_0[\tilde{r}^{\tilde{\tau}_0}] &< \infty, \quad \text{para } 1 < r < E[\tilde{r}_0^{\tilde{\xi}}]. \end{aligned}$$

Tomando como  $M = \min \left\{ \frac{1}{E[r_0^\xi]}, \frac{1}{E[\tilde{r}_0^{\tilde{\xi}}]} \right\}$ , tenemos que  $E_0[r^{\tau_0}] < \infty$  y

$E_0[\tilde{r}^{\tilde{\tau}_0}] < \infty$ , para alguna  $r$  tal que  $1 < r < M$ . utilizando la condición (iv) de las hipótesis C.2.3 se generan funciones  $V_1(x) = E_0[r^{\tau_0}]$ ,  $V_2(x) = E_0[\tilde{r}^{\tilde{\tau}_0}]$  y escalares  $\lambda$  y  $b$  que cumple la condición de Foster-Lyapunov. Definiendo

$$V(x) = \begin{cases} \max\{E_x[r^{\tau_0}], E_x[\tilde{r}^{\tilde{\tau}_0}]\} & \text{si } x > 0, \\ 1 & \text{si } x = 0. \end{cases}$$

se satisface la hipótesis C.2.2. Debido a que se cumplen todas las hipótesis del Teorema C.2.1, se obtiene que:

$$\Delta(x) \leq 2 \left( \frac{2b}{1-\lambda} + 2rh_x(r) + 1 \right) \left( \|P(\xi \in \cdot) - P(\tilde{\xi} \in \cdot)\|_{VT} \right)^{\frac{s-1}{s}}$$



## Apéndice D

# Uso del teorema de Kendall y cotas para la V-norma

Este apartado sirve como material de apoyo para el capítulo 8 y fue compilado de [4].

Los temas aquí tratados están relacionados con la teoría de procesos de renovación, particularmente con el teorema de Kendall. Presentamos esencialmente dos tipos de cotas para resultados en:

- convergencia de procesos de renovación, y
- desigualdades notables utilizadas en los Teoremas 7.1.1, 7.1.2 y 7.1.3.

### D.1. Resultados sobre convergencia

Sean  $V_1, V_2, \dots$ , sucesión de variables aleatorias definidas sobre  $\mathbb{Z}^+$ , independientes e idénticamente distribuidas. Definimos:

- $b_n := P(V_1 = n)$ , para todo  $n \geq 1$ ,
- $T_k := \begin{cases} 0 & \text{si } k = 0, \\ \sum_{i=1}^k V_i & \text{si } k \geq 1 \end{cases} \quad y$
- $u_n := P\left(\bigcup_{k=1}^{\infty} \{T_k = n\}\right)$

Se obtendrán tres diferentes cotas para el radio de convergencia de  $\sum_{n=0}^{\infty} (u_n - u_{\infty})z^n$ , basadas principalmente en el siguiente resultado.

**Teorema D.1.1 (Kendall)** *Supóngase que la sucesión  $\{b_n\}$  es aperiódica y cumple que  $\sum_{n=1}^{\infty} b_n R^n < \infty$  para algún  $R > 1$ . Entonces*

- i)  $u_{\infty} := \lim_{n \rightarrow \infty} u_n$  existe, y
- ii)  $\sum_{n=0}^{\infty} (u_n - u_{\infty})z^n$  tiene radio de convergencia mayor que 1.

Ahora definimos las siguientes constantes:

**Definición D.1.1**

- Sean  $R > 1$  y  $L < \infty$  de tal manera que  $\sum_{n=1}^{\infty} b_n R^n \leq L$
- $\beta > 0$ , de tal manera que  $\beta \leq b_1$
- $N = \frac{L-1}{R-1}$  (la cual es mayor o igual a 1),
- $R_1$  es la única solución en el intervalo  $(1, R)$ , para la ecuación

$$\frac{x-1}{x(\log(\frac{R}{x}))^2} = \frac{e^2 \beta}{8N},$$

considerando como incógnita a  $x$ .

- para algún  $r \in (1, R_1)$  se tiene que

$$K_1 = \frac{1}{r-1} \left( 1 + \frac{\beta + 2 \log(N) (\log(\frac{R}{r}))^{-1}}{\beta - 8N e^{-2} (r-1) r^{-1} (\log(\frac{R}{r}))^{-2}} \right)$$

**Teorema D.1.2** *Considerando la constantes definidas en D.1.1, se cumple:*

- i)  $\sum_{n=0}^{\infty} (u_n - u_{\infty})z^n$  tiene radio de convergencia mayor o igual a  $R_1$
- ii)  $\left| \sum_{n=0}^{\infty} (u_n - u_{\infty})z^n \right| \leq K_1$ , para todo  $|z| \leq r$ .

**Demostración.** Sea  $|z| < 1$  y  $j \geq 0$ . Definimos

$$c_j = \sum_{k=j+1}^{\infty} b_k, \quad b(z) = \sum_{n=1}^{\infty} b_n z^n$$

$$c(z) = \sum_{n=0}^{\infty} c_n z^n, \quad y \quad u(z) = \sum_{n=0}^{\infty} u_n z^n.$$

La ecuación de renovación nos indica que:

$$c(z) = \frac{1 - b(z)}{1 - z} = \frac{1}{(1 - z)u(z)} = \frac{1}{1 - \sum_{n=1}^{\infty} (u_{n-1} - u_n)z^n} \quad (\text{D.1})$$

Debido a que la serie de potencias de  $c(z)$ , tiene coeficientes no negativos, podemos ver que para  $|z| < R$  se cumple:

$$|c(z)| \leq c(R) = \frac{b(R) - 1}{R - 1} \leq \frac{L - 1}{R - 1} = N.$$

Así podemos ver que  $c(z)$  es holomórfica sobre  $|z| < R$ .

Por otro lado,

$$\begin{aligned} \Re((1 - z)c(z)) &= \Re(1 - b(z)) \\ &= \sum_{n=1}^{\infty} b_n \Re(1 - z^n) \\ &\geq \delta \Re(1 - z), \end{aligned}$$

para todo  $z \leq 1$ . Por lo tanto:

$$|c(re^{i\theta})| \geq \delta \frac{\Re(1 - re^{i\theta})}{1 - re^{i\theta}} \geq \delta \left| \operatorname{sen}\left(\frac{\theta}{2}\right) \right|,$$

para toda  $r \leq 1$ . En particular, debido a que  $c(r) > 0$ , para todo  $r \geq 0$ , observamos que  $c(z) \neq 0$ , cuando  $|z| < 1$ .

Si  $1 \leq r < R$ ,

$$\begin{aligned} |c(re^{i\theta})| &\geq \delta \left| \operatorname{sen}\left(\frac{\theta}{2}\right) \right| - |c(re^{i\theta}) - c(e^{i\theta})| \\ &\geq \delta \left| \operatorname{sen}\left(\frac{\theta}{2}\right) \right| - c(r) + c(1). \end{aligned}$$

Más aún, para  $1 \leq r < R$

$$\begin{aligned} |c(re^{i\theta})| &\geq c(r) - |re^{i\theta-r}| \sup\{|c'(z)| : z \in [r, re^{i\theta}]\} \\ &\geq c(r) - |re^{i\theta-r}| c'(r) \\ &= c(r) - 2r \left| \operatorname{sen}\left(\frac{\theta}{2}\right) \right| c'(r). \end{aligned}$$

Combinando estas dos estimaciones, obtenemos:

$$|c(re^{i\theta})| \geq \frac{\beta - A(r)}{\frac{\delta}{c(r)} + B(r)},$$

donde  $A(r) = \frac{2rc'(r)|c(r) - c(1)|}{c(r)}$  y  $B(r) = \frac{2rc'(r)}{c(r)}$ .

Debido a que la serie de potencias para  $c$  tiene coeficientes no negativos, podemos aplicar la desigualdad de Hölder y obtener

$$c(s) \leq c(r) \left(\frac{s}{r}\right)^{\frac{\log \frac{c(R)}{c(r)}}{\log \frac{R}{r}}},$$

para  $0 < r < s < R$ . Tomando  $s \downarrow r$ . Obtenemos:

$$c'(r) \leq \frac{c(r) \log \frac{c(R)}{c(r)}}{r \log \frac{R}{r}},$$

y consecuentemente,

$$c(r) - c(1) \leq \frac{(r-1)c(r) \log \frac{c(R)}{c(r)}}{r \log \frac{R}{r}}.$$

para  $1 \leq r < R$ . Así obtenemos las estimaciones

$$A(r) \leq \frac{2(r-1)c(r)}{r} \left[ \log \frac{N}{c(r)} \right]^2 \left[ \log \frac{R}{r} \right]^{-2},$$

y

$$B(r) \leq 2 \left[ \log \frac{N}{c(r)} \right] \left[ \log \frac{R}{r} \right]^{-1}.$$

Usando la desigualdad

$$x \left[ \log \frac{N}{c(r)} \right]^2 \leq 4Ne^{-2},$$

para  $0 < x < N$  en  $A(r)$  y la desigualdad  $c(r) \geq 1 \in B(r)$ , obtenemos:

$$A(r) \geq \frac{8Ne^{-2}(r-1)}{r} \left[ \log \frac{R}{r} \right]^{-2}$$

y

$$B(r) \leq 2 \log(N) \left[ \log \frac{R}{r} \right]^{-1}.$$

Así para  $1 < r < R_1$ , tenemos:

$$|c(re^{i\theta})| \geq \left| \frac{\delta - 8Ne^{-2}(r-1)r^{-1}(\log \frac{R}{r})^{-2}}{\delta + 2(\log N)(\log \frac{R}{r})^{-1}} \right| > 0. \quad (\text{D.2})$$

Por otro lado, retomando la ecuación (D.1), y considerando  $c(z) \neq 0$  para todo  $|z| < R_1$ . observamos que  $\sum_{n=1}^{\infty} (u_{n-1} - u_n)z^n$  es holomórfica sobre  $|z| < R_1$  y así

$$\lim_{n \rightarrow \infty} r^n |u_{n-1} - u_n| = 0,$$

para cada  $r < R_1$ . De aquí se sigue que  $u_\infty = \lim_{n \rightarrow \infty} u_n$  existe y  $\lim_{n \rightarrow \infty} r^n |u_{n-1} - u_n| = 0$ , para todo  $r < R_1$ .

Además usando el hecho de que:

$$u_n - u_\infty = \sum_{m=n+1}^{\infty} (u_{m-1} - u_m),$$

obtenemos:

$$\sum_{n=0}^{\infty} (u_n - u_\infty)z^n = \frac{1}{z-1} \left( \sum_{m=1}^{\infty} (u_{m-1} - u_{m-1})z^m - (1 - u_\infty) \right),$$

cuando  $1 < |z| < R_1$ , así usando (D.1) otra vez, para  $1 < r < R_1$ , obtenemos:

$$\sup_{|z| \leq r} \left| \sum_{n=0}^{\infty} (u_n - u_\infty)z^n \right| = \sup_{|z|=r} \left| \sum_{n=0}^{\infty} (u_n - u_\infty)z^n \right| = \frac{1}{r-1} \left( 1 + \sup_{|z|=r} \frac{1}{c(z)} \right),$$

así (ii) se sigue de (D.2) ■

**Teorema D.1.3** Sea  $\{X_n\}$  una cadena de Markov la cual es **reversible** con respecto a la medida invariante  $\pi$  y satisface la condición de minorización  $M(\delta, I_c, \nu)$ . Sea  $\{u_n\}$  la sucesión de renovación descrita por:

$$u_n = \begin{cases} 1, & \text{si } n = 0, \\ \tilde{\delta}P_\nu(X_{n-1} \in C), & \text{para } n \geq 1 \end{cases},$$

y suponga que la correspondiente sucesión de incrementos  $\{b_n\}$  satisface que

$$\sum_{n=1}^{\infty} b_n R^n \leq L, \text{ con } b_1 \geq \delta > 0, R > 1 \text{ y } L < \infty.$$

Si  $L > 1 + 2\delta R$ , definimos  $R_2$  como la única solución  $r \in (1, R)$  de la ecuación:

$$1 + 2\delta r = r^{\frac{\log L}{\log R}},$$

y  $R_2 = R$  en otro caso.

Entonces

- (i)  $\sum_{n=0}^{\infty} (u_n - u_{\infty})z^n$  tiene radio de convergencia mayor o igual que  $R_2$ .
- (ii) Si  $\lim_{n \rightarrow \infty} \left| \int_C p^n I_C(x) d\pi(x) - (\pi(C))^2 \right| r^n < \infty$ , para todo  $r < R_2$ , entonces para  $1 < r < R_2$  se cumple:

$$\sum_{n=1}^{\infty} |u_n - u_{\infty}| r^n \leq \frac{r\sqrt{\tilde{\delta}}}{1 - \frac{r}{R_2}}.$$

**Demostración.** Aplicando el algoritmo 3.1.1 se obtiene que  $\{u_n\}$  es una sucesión de renovación. La propiedad de reversibilidad implica que el operador de transición  $p$ , actúa como una contracción autoadjunta sobre el espacio de Hilbert  $L^2(\pi)$ . Denotaremos  $\langle \cdot, \cdot \rangle$  el producto interno en  $L^2(\pi)$  y  $\|\cdot\|$  la norma correspondiente.

Para  $A \subset E$  tenemos:

$$\pi(A) = \int p(x, A) d\pi(x) \geq \tilde{\delta} \nu(A) \pi(C),$$

así observamos que  $\nu$  es absolutamente continua con respecto a  $\pi$  y tiene derivada de Radon-Nikodym  $\frac{d\nu}{d\pi} \leq \frac{1}{\tilde{\delta}\pi(C)}$ . A lo largo de esta prueba escribiremos

$$f = I_C \text{ y } g = \frac{d\nu}{d\pi}. \text{ Por tenemos que } \|f\|^2 = \pi(C) \text{ y } g \leq \frac{1}{\tilde{\delta}\pi(C)}$$

Sea  $|z| < 1$

$$\begin{aligned} (1-z)u(z) &= (1-z) + \tilde{\delta}(1-z) \sum_{n=1}^{\infty} \langle p^{n-1} f, g \rangle z^n \\ &= (1-z) + \tilde{\delta}z(1-z) \langle (I-zp)^{-1} f, g \rangle. \end{aligned}$$

Como  $p$  es una contracción auto-adjunta sobre  $L^2(\pi)$ , su espectro es un subconjunto de  $[-1, 1]$  y tenemos la resolución espectral,

$$p = \int \lambda dE(\lambda),$$

donde  $E(1) = I$  y  $\lim_{\lambda \uparrow -1} E(\lambda) = 0$ . Escribimos  $F(\lambda) = \langle E(\lambda)f, g \rangle$ . La función  $F$  es de variación acotada y la correspondiente medida con signo  $\mu_{f,g}$  es soportada sobre  $[-1, 1]$  y tiene masa total

$$|\mu_{f,g}|([-1, 1]) \leq \|f\| \cdot \|g\| \leq \tilde{\delta}^{-\frac{1}{2}}.$$

Obtenemos para  $|z| < 1$ ,

$$(1-z)u(z) = (1-z) + \tilde{\delta}z(1-z) \int_{[-1,1]} (1-z\lambda^{-1})\mu_{f,g}d\lambda, g.$$

y así la función  $(1-z)u(z)$  tiene una extensión holomórfica al menos en  $\{z \in \mathbb{C} : z^{-1} \notin [-1, 1]\}$ . La ecuación de renovación es:

$$(1-z)u(z) = \frac{1-z}{1-b(z)},$$

para  $|z| < 1$ , y la función  $b$  es holomórfica en  $B(0, R)$ . Se sigue que las únicas soluciones en  $B(0, R)$  de la ecuación  $b(z) = 1$  pertenecen a uno de los siguientes intervalos  $(-R, -1]$  y  $[1, R)$ . Debido a que  $b'(z) > 0$ , el cero de  $b(z) - 1$  en  $z = 1$  es el cero. Para  $1 < r \leq R$  tenemos que  $b(r) > b(1) = 1$ . Para  $1 < r < R$ , tenemos  $b(-r) \leq -2b_1r + b(r)$ . Usando la estimación

$$b(r) \leq [b(R)]^{(\log r)/\log R} = r^{(\log L)/\log R},$$

se sigue que para  $1 < r < R_2$ , tenemos  $b(-r) < 1$ , donde  $R_2$  es la constante mencionada en el teorema. Así  $(1-z)u(z)$  tiene una extensión holomórfica a  $B(0, R_2)$ , y la primera parte del teorema se sigue.

Ahora supongamos la parte (ii). Dado  $r < R_2$ , tenemos

$$|\langle p^n f, f \rangle - (\pi(C))^2| \leq Mr^n, \quad (\text{D.3})$$

para alguna  $M$ . Renombrando la resolución espectral, tenemos:

$$\lim_{n \rightarrow \infty} \langle p^n f, f \rangle = \int_1 d \langle E(\lambda), f \rangle,$$

y así podemos reescribir (D.3)

$$\left| \int_{[-1, 1]} \lambda^n d \langle E(\lambda), f \rangle \right|$$

Ahora  $\lambda \rightarrow \langle E(\lambda)f, f \rangle$  es una función creciente, de aquí le corresponde una medida positiva  $\mu_f([-\frac{1}{r}, \frac{1}{r})) = \mu_f((\frac{1}{r}, 1])$ , así se cumple que para toda  $r < R_2$ , se tiene que  $\langle E(\lambda)f, f \rangle$  es constante sobre  $[-1, \frac{1}{R_2}]$  y sobre  $(\frac{1}{R_2}, 1]$ . Se aquí se sigue  $F(\lambda) = \langle E(\lambda)f, f \rangle$ , es constante sobre estos mismos intervalos y el soporte de  $|\mu_{f,g}|$  está contenida en  $[\frac{-1}{R_2}, \frac{1}{R_2}]$ . Nota que

$$\begin{aligned} u_\infty &= \tilde{\delta} \lim_{n \rightarrow \infty} \langle p^n f, f \rangle \\ &= \tilde{\delta} \lim_{n \rightarrow \infty} \int \lambda^{n-1} \mu_{f,g} d\lambda \\ &= \tilde{\delta} \mu_{f,g}(\{1\}) \end{aligned}$$

Obtenemos, para  $n \geq 1$ ,

$$\begin{aligned} |u_n - u_\infty| &= \tilde{\delta} \left| \int_{\left[\frac{-1}{R_2}, \frac{1}{R_2}\right]} \lambda^{n-1} \mu_{f,g} d\lambda \right| \\ &\leq \tilde{\delta} \left( \frac{1}{R_2} \right)^{n-1} |\mu_{f,g}| \left( \left[ \frac{-1}{R_2}, \frac{1}{R_2} \right] \right) \\ &\leq \sqrt{\tilde{\delta}} \left( \frac{1}{R_2} \right)^{n-1}. \end{aligned}$$

Así para  $r < R_2$ , obtenemos

$$\sum_{n=1}^{\infty} |u_n - u_\infty| r^n \leq \frac{r \sqrt{\tilde{\delta}}}{1 - \frac{r}{R_2}}$$

como lo fue requerido.

■

**Teorema D.1.4** Sea  $\{X_n\}$  una cadena de Markov la cual es **reversible y positiva** con respecto a la medida invariante  $\pi$  y satisface la condición de minorización  $M(\delta, I_C, \nu)$ . Sea  $\{u_n\}$  la sucesión de renovación descrita por:

$$u_n = \begin{cases} 1, & \text{si } n = 0, \\ \tilde{\delta} P_\nu(X_{n-1} \in C), & \text{para } n \geq 1 \end{cases},$$

y suponga que la correspondiente sucesión de incrementos  $\{b_n\}$  satisface que  $\sum_{n=1}^{\infty} b_n R^n \leq L$ , con  $b_1 \geq \delta > 0$ ,  $R > 1$  y  $L < \infty$ .

Si  $L > 1 + 2\delta R$ , definimos  $R_2$  como la única solución  $r \in (1, R)$  de la ecuación:

$$1 + 2\delta r = r^{\frac{\log L}{\log R}},$$

y  $R_2 = R$  en otro caso.

Entonces

- (i)  $\sum_{n=0}^{\infty} (u_n - u_\infty) z^n$  tiene radio de convergencia mayor o igual que  $R_2$ .
- (ii) Si  $\lim_{n \rightarrow \infty} \left| \int_C p^n I_C(x) d\pi(x) - (\pi(C))^2 \right| r^n < \infty$ , para todo  $r < R_2$ , entonces para  $1 < r < R_2$  se cumple:

$$\sum_{n=1}^{\infty} |u_n - u_\infty| r^n \leq \frac{r \sqrt{\tilde{\delta}}}{1 - \frac{r}{R_2}}.$$

**Demostración.** La hipótesis adicional implica que el espectro de  $p$  está contenido en  $[0, 1]$ . Siguiendo el argumento de la prueba anterior, obtenemos para  $|z| < 1$ :

$$(1 - z)u(z) = (1 - z) + \tilde{\delta}z(1 - z) \int_{[0,1]} (1 - z\lambda)^{-1} \mu_{f,g} d\lambda,$$

y así, la función  $(1 - z)u(z)$  tiene una extensión holomorfa al menos en  $\{z \in \mathbb{C} : z^{-1} \notin [0, 1]\}$ . Se sigue que la ecuación  $b(z) = 1$  no tiene solución en  $(-R, -1]$  y así  $(1 - z)u(z)$  es holomorfa sobre  $B(0, R)$ . El resto de la prueba es similar a la realizada en el teorema anterior. ■

## D.2. Desigualdades notables

Los siguientes resultados forman se utilizarán en el capítulo 7.

**Teorema D.2.1** *Suponga que se establece la condición*

$$pV \leq \begin{cases} \lambda V(x), & \text{si } x \notin C \\ K & \text{si } x \in C. \end{cases} ,$$

para alguna función medible  $V : E \rightarrow [1, \infty)$  y constantes  $\lambda < 1$  y  $K < \infty$ .

Entonces:

(i) Para todo  $x \in E$ ,  $P_x(\tau < \infty) = 1$ .

(ii) Para  $1 \leq r \leq \lambda^{-1}$

$$G(r, x) \leq \begin{cases} V(x), & \text{si } x \notin C \\ rK & \text{si } x \in C. \end{cases} .$$

(iii) Para  $0 < r < \lambda^{-1}$

$$H(r, x) \leq \begin{cases} \frac{r\lambda V(x)}{1 - r\lambda}, & \text{si } x \notin C \\ \frac{r(K - r\lambda)}{1 - r\lambda} & \text{si } x \in C. \end{cases} .$$

(iv) Para  $0 < r < \lambda^{-1}$  y  $x \in C$

$$\frac{H(r, x) - rH(1, x)}{r - 1} \leq \frac{\lambda r(K - 1)}{(1 - \lambda)(1 - r\lambda)}.$$

**Proposición D.2.1** *Asuma que la cadena de Markov es geoméricamente ergódica con medida invariante  $\pi$ ,  $C$  es un átomo y que  $V$  es una función o negativa. Suponga que  $g : E \rightarrow \mathbb{R}$  y satisface  $\|g\|_V \leq 1$ . Entonces*

$$\begin{aligned} & \sup_{|z| \leq r} \left| \sum_{n=1}^{\infty} \left( p^n g(x) - \int g d\pi \right) z^n \right| \\ & \leq H(r, x) + G(r, x) H(r, a) \sup_{|z| \leq r} \left| \sum_{n=0}^{\infty} (u_n - u_\infty) z^n \right| \\ & + H(r, a) \frac{G(r, x) - 1}{r - 1} + \frac{H(r, a) - rH(1, a)}{r - 1} < \infty, \end{aligned}$$

para  $r > 1$ .

**Proposición D.2.2** *Asuma que la cadena de Markov es geoméricamente ergódica con medida invariante  $\pi$ ,  $C$  es un átomo y que  $V$  es una función o negativa. Suponga que  $g : E \rightarrow \mathbb{R}$  y satisface  $\|g\|_V \leq 1$ . Entonces*

$$\begin{aligned} & \sup_{|z| \leq r} \left| \sum_{n=1}^{\infty} \left( p^n g(x) - \int g d\pi \right) z^n \right| \\ & \leq \bar{H}(r, x) + \bar{G}(r, x) \bar{H}(r, a, 1) \sup_{|z| \leq r} \left| \sum_{n=0}^{\infty} (\bar{u}_n - \bar{u}_\infty) z^n \right| \\ & + \bar{H}(r, a, 1) \frac{\bar{G}(r, x) - 1}{r - 1} + \bar{H}(r, a, 1) \frac{\bar{G}(r, x) - 1}{r - 1} \\ & + \frac{\bar{H}(r, a, 1) - r\bar{H}(1, a, 1)}{r - 1}. \end{aligned}$$

**Proposición D.2.3** *Considere la condición de Minorización y la condición Drift. Definiendo*

$$\alpha_1 = 1 + \left( \log \frac{K - \tilde{\delta}}{1 - \tilde{\delta}} \right) \div (\log \lambda^{-1}),$$

$$\alpha_2 = 1 + \left( \log \frac{K}{\tilde{\delta}} \right) \div \log(\lambda^{-1}) \text{ y}$$

$$R_0 = \min \left( \lambda^{-1}, (1 - \tilde{\delta})^{-1 + \alpha_1} \right) \text{ se tiene:}$$

$$\bar{G}(r) \leq r^{\alpha_1}, \quad \text{para } 1 \leq r \leq \lambda^{-1} \quad (\text{D.4})$$

$$\bar{G}(r, x) \leq \frac{\tilde{\delta}G(r, x)}{1 - (1 - \tilde{\delta}r^{\alpha_1})}, \quad (\text{D.5})$$

$$\bar{G}(r, a, 1) \leq \frac{\tilde{\delta}r^{\alpha_2}}{1 - (1 - \tilde{\delta}r^{\alpha_1})} \equiv L(r), \quad (\text{D.6})$$

$$\bar{H}(r, x) \leq H(r, x) + \frac{r[K - r\lambda - \tilde{\delta}(1 - r\lambda)]}{(1 - r\lambda)[1 - (1 - \tilde{\delta})r^{\alpha_1}]} G(r, x). \quad (\text{D.7})$$

$$\bar{H}(r, a, 1) \leq \frac{r^{\alpha_2+1}(K - r\lambda)}{(1 - r\lambda)[1 - (1 - \tilde{\delta})r^{\alpha_1}]}, \quad (\text{D.8})$$

$$\begin{aligned} \frac{\bar{H}(r, a, 1) - r\bar{H}1, a, 1}{r - 1} &\leq \frac{r^{\alpha_2+1}\lambda(K - 1)}{(1 - \lambda)(1 - r\lambda)[1 - (1 - \tilde{\delta})r^{\alpha_1}]} \\ &+ \frac{r[K - \lambda - \tilde{\delta}(1 - \lambda)]}{(1 - \lambda)[1 - (1 - \tilde{\delta})r^{\alpha_1}]} \left( \frac{r^{\alpha_2-1}}{r - 1} + \frac{(1 - \tilde{\delta})(r^{\alpha_1} - 1)}{\tilde{\delta}(r - 1)} \right). \end{aligned} \quad (\text{D.9})$$



# Bibliografía

- [1] R. B. Ash. *Real Analysis and Probability*. Academic Press, 1972.
- [2] K. B. Athreya and P. Neyn. A new approach to the limit theory of recurrent Markov chains. *Amer.Math. Soc.*, 245:493–501, 1978.
- [3] R. R. Bahadur and R. R. Rao. On deviations of the sample mean. *Ann. Math. Statist*, 31:1015–1027, 1960.
- [4] P. H. Baxendale. Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Probab.*, 15(1B):700–738, 2005.
- [5] D. P. Bertsekas. *Dynamic Programming: Deterministic and Stochastic Models*. Prentice-Hall, 1987.
- [6] R. N . Bhattacharya and C. Lee. Ergodicity of nonlinear first order autoregressive models. *J. Theoret. Probab.*, 8(1):207–219, 1995.
- [7] P. Billingsley. *Probability and Measure*. John Wiley and Sons, 1979.
- [8] H. Chernoff. A measure of asymptotic efficiency for test of a hypothesis based on the sum of observations. *Ann. Math. Statist*, 23:493–507, 1952.
- [9] E. Cinlar. *Introduction to Stochastic Processes*. Prentice-Hall, Inc, 1975.
- [10] J. W. Cohen. *The Simple Server Queue*. North-Holland, 1982.
- [11] J. L. Doob. *Stochastic Processes*. John Wiley and Sons Inc, 1953.
- [12] E. I. Gordienko. An estimate of stability of optimal control of certain stochastic and deterministic systems. *J. Soviet. Math.*, 59:891–899, 1992.
- [13] E. I. Gordienko and O. Hernández-Lerma. Average cost Markov control processes with weight norms: value iteration. *Appl. Math. (Warsaw)*, 23:219–237, 1995.
- [14] E. I. Gordienko and O. Hernández-Lerma. Average cost Markov control processes with weighted norms: existence of canonical policies. *Appl. Math. (Warsaw)*, 23(2):199–218, 1995.

- [15] D. Griffeath. *Coupling Methods for Markov Processes*, volume 2 of *Studies in probability and ergodic theory*. Academic Press, Inc, 1978.
- [16] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.
- [17] O. Hernández-Lerma and J. B. Lasserre. *Discrete-Time Markov Control Processes: Basic Optimality Criteria*. Springer-Verlag, 1996.
- [18] P. G. Hoel, S. C. Port, and C. J. Stone. *Introduction to Stochastic Processes*. Waveland Press, inc., 1972.
- [19] A. Hordijk and F. M. Spieksma. On ergodicity and recurrence properties of a Markov chain with application. *Adv. Appl. Probab.*, 24:343–376, 1992.
- [20] N. Jain and B. Jamison. Contributions to Doeblin’s theory of Markov processes. *Z. Wahrsch. Verw. Gebiete*, 8:19–40, 1967.
- [21] V. V. Kalashnikov. Qualitative analysis of the behavior of complex systems by the method of test functions. *Series in Theory and Methods of Systems Analysis “Nauka”*, 1978. En Ruso.
- [22] T. Kamae, U. Krengel, and G. L. O’Brien. Stochastic inequalities on partially ordered spaces. *Ann. Probability*, 5(6):899–912, 1977.
- [23] N. V. Kartashov. Criteria for uniform ergodicity and strong stability of Markov chains with common phase space. *Teor. Veroyatnost. i Mat. Stat. Vyp.*, 30:65–81, 1984. Versión en inglés *Theor. Probab. Math. Statistics*, 30: 71-89, 1985.
- [24] N. V. Kartashov. Inequalities in theorems of ergodicity and stability for Markov chains with common phase space. II. *Teor. Veroyatnost. i Primenen.*, 30(3):478–485, 1985. Versión en inglés *Theory Probab. Appl.*, 30(3):507-515, 1985.
- [25] N. V. Kartashov. Inequalities in theorems of ergodicity and stability for Markov chains with common phase space. I. *Teor. Veroyatnost. i Primenen.*, 30(2):230–240, 1985. Versión en inglés *Theory Probab. Appl.*, 30(2):247-259, 1985.
- [26] N. V. Kartashov. A refinement ergodicity estimates of renewal processes and Markov chains. *Teor. Veroyatnost. i Mat. Stat. Vyp.*, (32):27–33, 133, 1985. Versión en inglés *Theor. Probab. Math. Statistics*, No. 32: 27-34, 1986.
- [27] D. G. Kendall. *Unitary dilatations of Markov transition operators and the corresponding integral representation for transition-probability matrices*. In: *Probability and Statistics: The Harald Cramér volume* (edited by U. Grenander), 1959.
- [28] T. Lindvall. *Lectures on the Coupling Method*. Wiley series in probability and mathematical. John Wiley and Sons, Inc, 1992.

- [29] T. Lindvall. On Strassen theorem on stochastic domination. *Electron. Comm. Probab.*, 4:51–59, 1999.
- [30] R. B. Lund and R. L. Tweedie. Geometric convergence rates for stochastically ordered Markov chains. *Math. Oper. Res.*, 20(1):182–194, 1996.
- [31] R.B. Lund. The geometric convergence rate of a Lindley random walk. *J. Appl. Prob.*, 34:806–811, 1997.
- [32] N. Metropolis, A. W. Rosenbluth, M. N . Rosenbluth, A. H. Teller, and E. Teller. Equations os state calculations by fast computing machines. *J. Chem. Phys*, 21:1087–1092, 1953.
- [33] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag, 1993.
- [34] S. P. Meyn and R. L. Tweedie. Computable bounds for convergence rates of Markov chains. *Ann. Appl. Probab.*, 4:981–1011, 1994.
- [35] R. Montes-de-Oca and F. Salem-Silva. Estimates for perturbations of average Markov decision processes with a minimal state and upper bounded by stochastically ordered Markov chains. *Kybernetika (Prague)*, 41(6):757–772, 2005.
- [36] E. Nummelin. A splitting technique for Harris recurrent Markov chains. *Z. Wahrsch. Verw. Gebiete*, 43:309–318, 1978.
- [37] E. Nummelin. *General Irreducible Markov Chains and Non-negative Operators*. Cambridge University Press, 1984.
- [38] E. Nummelin and P. Tuominen. Geometric ergodicity of Harris recurrent Markov chains with applications to renewal theory. *Stochastic Process. Appl.*, 12:187–202, 1982.
- [39] E. Nummelin and R. L. Tweedie. Geometric ergodicity and R.positivity for geometric Markov chains. *Ann. Probability*, 6(3):404–420, 1978.
- [40] S. Orey. *Limit theorems for Markov chain Transition Probabilities*. Van Nostrand Reinhold Company, 1971.
- [41] J. W. Pitman. Uniform rates of convergence for Markov chain transition probabilities. *Z. Wahrsch. Verw. Gebiete*, 29:193–227, 1974.
- [42] N. Popov. Conditions for geometric ergodicity of countable Markov chain. *Sov. Math. Dokl.*, 18, 1977.
- [43] D. Revuz. *Markov Chains*. North-Holland, 1975.
- [44] G. O. Roberts and J. S. Rosenthal. General state space Markov chains and MCMC algorithms. *Probab. Surv.*, pages 20–71, 2004.

- [45] G. O. Roberts and R. L. Tweedie. Bounds on regeneration times and convergence rates for Markov chains. *Stochastic Process. Appl.*, 80:211–229, 1999.
- [46] G. O. Roberts and R. L. Tweedie. Rates of convergence of stochastically monotone and continuous Markov models. *J. Appl. Prob.*, 37:359–373, 2000.
- [47] J. S. Rosenthal. Convergence rates for Markov chains. *SIAM Rev.*, 37(3):387–405, 1995.
- [48] J. S. Rosenthal. Minorization conditions and convergence rates for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.*, 90:558–566, 1995.
- [49] J. S. Rosenthal. A review of asymptotic convergence for general state space Markov chains. *Far East J. Theor. Stat.*, 5:37–50, 2001.
- [50] F. Salem-Silva. *Estabilidad de Procesos de Control de Markov para el Caso Descontado*. Tesis de doctorado, UAM-Iztapalapa, 2000.
- [51] D. J. Scott and R. L. Tweedie. Explicit Rates of Convergence of Stochastically Ordered Markov Chains. Proc. Athens Conference on Applied Probability and Time Series Analysis: Papers in Honour of J. M., 1995.
- [52] V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Statist.*, 36:423–439, 1965.
- [53] H. Thorisson. *Coupling, Stationarity and Regeneration*. Series in Probability and its Applications. Springer-Verlag, 2000.
- [54] R. L. Tweedie. R-theory for Markov chains on a general state space I: solidarity properties and R-recurrent chains. *Ann. Probability*, 2(5):840–864, 1974.
- [55] R. L. Tweedie. *Criteria for rates of convergence of Markov chains, with applications to queueing and storage theory*. In: Papers in Probability Statistics and Analysis (edited by J.F.C Kingman and G.E.H Reuter), 1983.
- [56] R. L. Tweedie. *Markov chains: structure and applications, (edited by D. N. Shanbhay and C.R.Rao)*, volume 19 of *Handbook of Statistics*. Elsevier, Vol. 19, 817-851, 1998.
- [57] D. Vere-Jones. Geometric ergodicity in denumerable Markov chains. *Quart. J. Math.*, 13:7–28, 1962.

# Índice alfabético

- V-norma, 73
- $\varphi$ -recurrente, 22
- Átomo, 75
- Acoplamiento
  - clásico, 34
  - de Doeblin, 34
  - de elementos aleatorios, 30
  - de medidas, 30
  - evento de, 35
  - exitoso, 36
  - markoviano, 34
  - ordenado, 40
  - tiempo de, 35
- acoplamiento
  - exacto, 36
- Algoritmo
  - acoplamiento particionante, 37
  - Metropolis-Hastings, 85
- Cadena
  - estocásticamente monótona, 43
  - de Markov
    - positiva, 74
    - reversible, 74
  - ordenada estocásticamente, 43
  - ordenada por realizaciones, 43
- Cadena de Markov
  - particionante, 23
- Condición de
  - Foster-Lyapunov, 22
- Condición de
  - minorización, 21
- Conjunto
  - $\varphi$ -comunicado, 19
  - pequeño, 21
- Copias de elementos aleatorios, 30
- Desigualdad
  - de tiempo de acoplamiento, 36
- Distribución
  - inicial, 16
  - invariante, 18
- Función
  - pequeña, 21
- Harris recurrente, 22
- Kérnel
  - $\varphi$ -irreducible, 19
  - de transición, 15
  - estocásticamente dominado, 42
  - irreducible, 19
  - monótono, 43
- Lebesgue
  - descomposición de, 100
- Lindley
  - caminata, 70
- Métrica de la variación total, 35
- Markov
  - cadena
    - recurrente, 22
  - cadena de, 16
  - Cadena ergódica de, 21
  - cadena fuertemente aperiódica,
    - 22
  - cadena nula de, 21
  - cadena positiva de, 21
  - propiedad de, 16
  - realización de, 16
- Medida
  - invariante, 18

- irreducible, 19
- irreducible maximal, 20
- pequeña, 21
- Modelo de control de Markov, 115
  
- Orden estocástico, 39
  - para medidas, 39
  
- P.D.M., 116
- Política, 116
- Política de control., 116
- Proceso
  - regenerativo, 27
- Proceso de decisión de Markov, 116
- Propiedad
  - de separabilidad, 101
  
- Regeneración
  - propiedad de, 27
  
- Sucesión
  - aperiódica, 76
  
- Tasa de Convergencia, 47
  - ergódica, 48
  - geométrica, 48
  - uniformemente ergódica, 48

# Notación

## Cadenas de Markov

$(E, \xi)$  espacio de estados

$p$  kernel de transición

$p^n$  kernel de transición en  $n$ -pasos

$(\Omega, \mathcal{F}, P_\mu)$  espacio base para la cadena con distribución inicial  $\mu$

$\delta_x$  distribución concentrada en  $x$

$\pi$  distribución invariante

$\varphi$  medida irreducible

$\Psi$  medida irreducible maximal

$M(n, \delta, S, \nu)$  condición de minorización

$C$  conjunto pequeño

$\tilde{X} = (X_n, I_n)$  cadena particionante

$\hat{X}$  acoplamiento para  $X$

$(\hat{\Omega}, \hat{\mathcal{F}}, \hat{P})$  espacio base de un acoplamiento

$T$  tiempo de acoplamiento

$X \stackrel{D}{=} Y$   $X$  y  $y$  tienen la misma distribución

$\stackrel{est.}{\leq}$  orden estocástico

## Convergencia

$\| \cdot \|$  métrica probabilística

$\| \cdot \|_{VT}$  métrica de la variación total

$\| \cdot \|_V$  métrica de la variación ponderada

$\phi$  Tasa de convergencia