

**Estabilidad de procesos de control de  
Markov para el caso descontado**

**Tesis que presenta  
Francisco Sergio Salem Silva  
Para obtener el grado de Doctor en Ciencias**

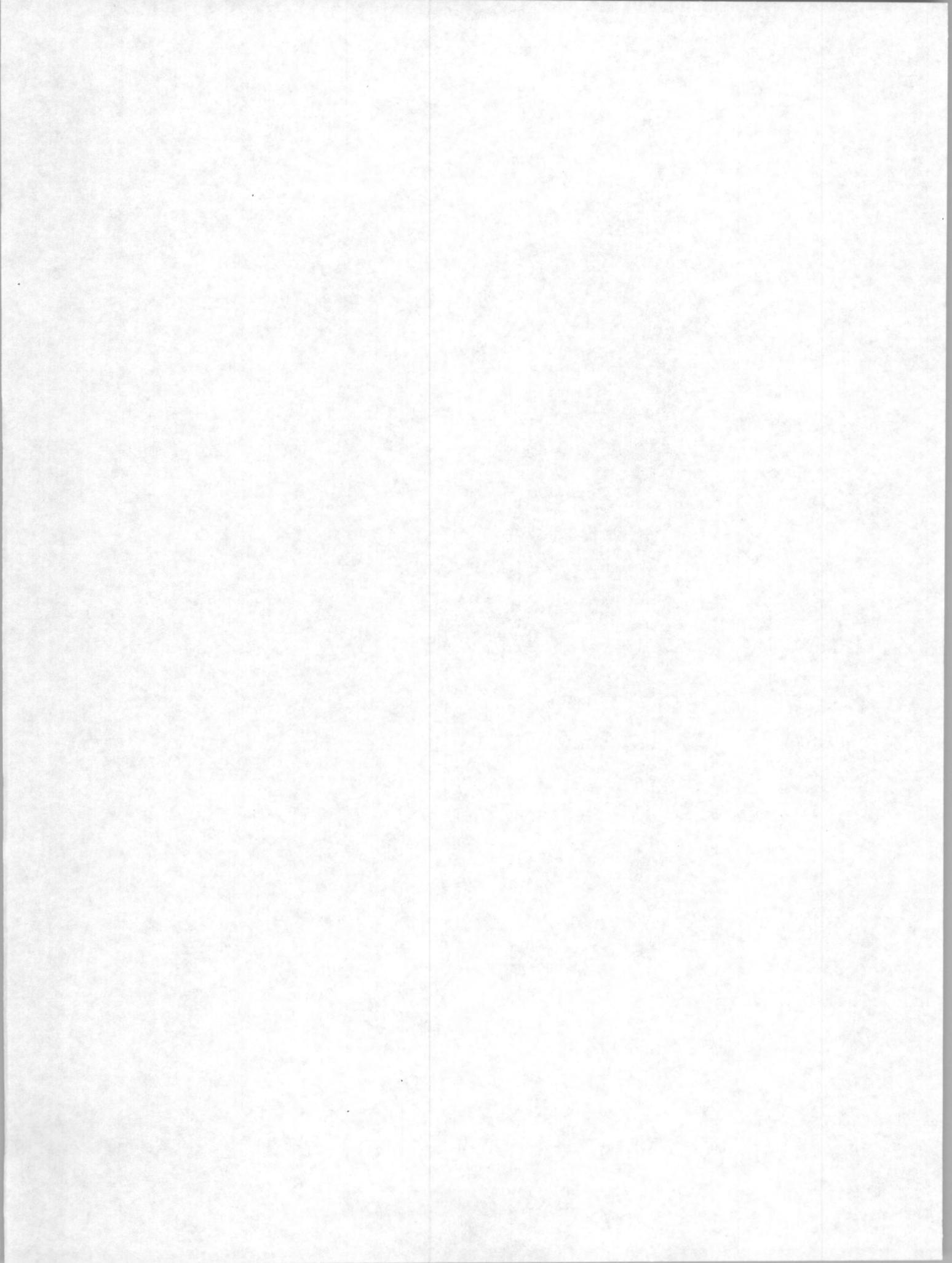
**Asesor  
Evgueni Gordienko**



**Casa abierta al tiempo**

Noviembre del-2000

**UNIVERSIDAD AUTÓNOMA METROPOLITANA**  
UNIDAD IZTAPALAPA División de Ciencias Básicas e Ingeniería



A la mujer que amo

## Agradecimientos:

La realización de este trabajo no hubiera sido posible sin el apoyo de mucha gente y las siguientes instituciones:

La Benemérita Universidad Autónoma de Puebla, mi centro de trabajo y estudio, donde he encontrado la libertad y los medios para mi desarrollo profesional.

El Conacyt de quien recibí beca por tres años

Industrias Negromex de quien recibí apoyo desinteresado durante tres años

La Universidad Autónoma Metropolitana que me abrió sus puertas desde hace ya más 10 años, siendo para mí una verdadera casa abierta al tiempo.

Entre la gente de quien recibí apoyo quiero destacar a Raúl quien siempre estuvo ahí en las buenas, en las malas y en las peores, para mí más que un amigo, un hermano.

A toda la Raza del departamento de Matemáticas de la UAMI., Marta , Mario, Juan , Felipe, los Julios, Manuel, Víctor, Dany boy ,las secres, etc

A mi ausente director de tesis, E.Gordienko, gran profesor del arte de la adivinación.

A mis padres ejemplo de constancia y solidaridad

A la Doctora, al flaquillo, al chato picudo, fuente inagotable de ternura

A Ursula y Bruno, siempre presentes

A la dueña de mis quincenas (no por eso menos amada) quien hace de mi vida un reto interesante y creador.

A todos ellos gracias

Estabilidad de procesos de control de Markov  
para el caso descontado

Francisco S. Salem Silva

26\09\00

# Índice General

<b>Introducción</b>	<b>iii</b>
<b>1 El modelo de control</b>	<b>1</b>
1.1 Introducción . . . . .	1
1.2 Procesos de control de Markov . . . . .	1
1.3 Políticas . . . . .	2
1.4 Planteamiento del problema . . . . .	5
<b>2 Estimación del índice de estabilidad</b>	<b>7</b>
2.1 Introducción . . . . .	7
2.2 Suposiciones . . . . .	8
2.3 El Problema de control óptimo. . . . .	9
2.4 Desigualdad de estabilidad del problema de control óptimo . .	14
2.5 Ejemplo I . . . . .	20
2.6 Ejemplo II . . . . .	23
<b>3 El índice de estabilidad relativo</b>	<b>27</b>
3.1 Introducción . . . . .	27
3.2 Suposiciones . . . . .	28
3.3 Comentarios sobre las suposiciones . . . . .	30
3.4 Desigualdades de Estabilidad . . . . .	31
3.5 Ejemplos . . . . .	37
<b>4 Estabilidad vía cadenas ordenadas</b>	<b>43</b>
4.1 Introducción . . . . .	43
4.2 Preliminares . . . . .	44
4.3 Estabilidad en el caso en que $0 \in \mathbb{X}$ se alcanza para cada política de control . . . . .	46

*ÍNDICE GENERAL*

ii

4.3.1	Tasas de convergencia para cadenas ordenadas no controladas . . . . .	46
4.3.2	Estabilidad para Procesos Controlados . . . . .	50
4.3.3	Ejemplo . . . . .	55
4.4	Procesos ordenados que no alcanzan el mínimo . . . . .	58
4.4.1	Cotas para modelos no controlados . . . . .	58
4.4.2	Modelos controlados . . . . .	61
4.4.3	Ejemplo . . . . .	63
5	<b>Conclusiones</b>	<b>68</b>

# Introducción

Una gran cantidad de problemas de optimización dinámica se caracterizan por el conocimiento inadecuado de sus parámetros, generalmente los parámetros de la ecuación que describe la dinámica del sistema. A menudo es conveniente capturar esta "inadecuación" incorporándola a una perturbación de los parámetros que describen el problema. En el caso de procesos de control de Markov a tiempo discreto, esta perturbación afecta a la ley de transición [3]. En esta situación la perturbación será típicamente pequeña y tendrá forma de parámetro, el cual será real, positivo mayor que 0.

Es por eso que es importante preguntarse, ¿cuál es el comportamiento del problema de optimización cuando el parámetro de perturbación tiende a 0 [3]?

El estudio de esta pregunta puede hacerse de varias maneras, por ejemplo Müller [28] utiliza una cierta relación de orden estocástica de los costos generados al aplicar diferentes políticas y se reduce al caso de horizonte finito.

A nosotros nos interesa el problema del comportamiento asintótico de un proceso de control de Markov cuando el parámetro de perturbación tiende a 0, generalmente expresado en términos de la distancia probabilística entre dos distribuciones.

Nosotros tomaremos el enfoque de [6] y [7] para resolver este problema. Consideraremos dos procesos de control de Markov a tiempo discreto  $\mathcal{P}$  y  $\tilde{\mathcal{P}}$  definidos en el mismo espacio de estados  $\mathbb{X}$  y con el mismo espacio de acciones  $A$ , donde tanto  $\mathbb{X}$  como  $A$  son espacios de Borel, con costos de un paso  $c$  y  $\tilde{c}$  posiblemente no acotados con probabilidades de transición  $p$  y  $\tilde{p}$ , respectivamente. Supondremos que  $\tilde{\mathcal{P}}$  es un proceso perturbado de  $\mathcal{P}$  del cual se conocen todos sus parámetros, y  $\tilde{p}$  esta cercano a  $p$  en términos de la métrica de la variación total, por este motivo lo llamaremos proceso aproximante de  $\mathcal{P}$ ,

Usaremos como criterio de desempeño la esperanza del costo total des-

contado y la denotaremos por  $V_\alpha(x, \pi)$  cuando nos estemos refiriendo a  $\mathcal{P}$ , y por  $\tilde{V}_\alpha(x, \tilde{\pi})$  cuando el proceso considerado sea  $\tilde{\mathcal{P}}$  donde  $x$  es el estado inicial,  $\alpha$  es un factor de descuento,  $\pi$  y  $\tilde{\pi}$  son las reglas de decisión (políticas) para tomar las acciones (estos conceptos serán definidos en el capítulo 1).

Denotaremos por  $\pi_\alpha^*$  y  $\tilde{\pi}_\alpha^*$  a las políticas que minimicen  $V_\alpha(x, \pi)$  y  $\tilde{V}_\alpha(x, \tilde{\pi})$ , respectivamente.

Estimaremos el costo adicional al usar  $\tilde{\pi}_\alpha^*$  en lugar de  $\pi_\alpha^*$ , por medio del índice de estabilidad descontado definido por:

$$\Delta_\alpha(x) = [V_\alpha(\tilde{\pi}_\alpha^*, x) - V_\alpha(\pi_\alpha^*, x)]$$

Si es más importante estimar la pérdida relativa, usaremos el índice de estabilidad relativo:

$$\bar{\Delta}_\alpha(x) = \frac{\Delta_\alpha(x)}{|V_\alpha(x, \pi_\alpha^*)|} \quad (1)$$

Diremos que el proceso  $\mathcal{P}$  es estable (estable relativo) respecto a una métrica probabilística  $m$ , si  $\Delta_\alpha(x)$  (o  $\bar{\Delta}_\alpha(x)$ ) tiende a 0, cuando la distancia entre las probabilidades de transición  $m(p, \tilde{p})$  tiende a 0.

Nuestro problema consiste en dar condiciones para garantizar la estabilidad del proceso respecto a la métrica de la variación total o variación total ponderada. Dicho de otra manera queremos encontrar cotas superiores para el índice de estabilidad y para el índice de estabilidad relativo.

Por otro lado, nos interesa estudiar el comportamiento asintótico de nuestras cotas de estabilidad, para el caso del índice relativo, cuando el factor de descuento  $\alpha$  tiende a 1, o más exactamente, nos gustaría tener cotas uniformes sobre  $\alpha \in (0, 1)$  de la siguiente forma:

$$\sup_{\alpha \in (0,1)} \bar{\Delta}_\alpha(x) \leq g_x^{(2)}(m(p, \tilde{p})) \quad (2)$$

donde  $g_x^{(2)}(z) \rightarrow 0$  cuando  $z \rightarrow 0$ , para lograr esta desigualdad será necesario pedir condiciones de ergodicidad en los procesos considerados, cuando aplicamos las políticas de control (véase [41], [6], [7], [8]).

En la sección de ejemplos del capítulo 3 daremos un ejemplo no estable en el caso descontado para el cual

$$\sup_{\alpha \in (0,1)} \bar{\Delta}_\alpha(x) = \infty$$

donde  $\bar{\Delta}_\alpha(x)$  está definida en términos de  $p(dy|x,a)$  y  $\tilde{p}(dy|x,a)$ , a pesar de que  $m(p,p) \rightarrow 0$ , donde  $m$  es la métrica de *Forét-Mourier*.

Se conocen algunas cotas superiores para el índice de estabilidad  $\Delta_\alpha(x)$  y para  $\bar{\Delta}_\alpha(x)$  (vea [40], [41], [11], [13], [45]) que tienen la siguiente estructura:

$$\Delta_\alpha(x) \leq M(\alpha)\psi(p,\tilde{p}) \quad (3)$$

donde  $\psi$  es alguna "medida" de la "diferencia" entre  $p$  y  $\tilde{p}$ , y  $M(\alpha)$  es una constante, que generalmente es de orden  $(1-\alpha)^{-2}$ . Bajo ciertas suposiciones adicionales (más o menos fuertes) las desigualdades en [40], [41] producen las cotas (3) con  $M(\alpha)$  de orden  $(1-\alpha)^{-1}$ . Pero en estas desigualdades el término  $\psi(p,\tilde{p})$  involucra cantidades que son muy difíciles de calcular porque dependen de los costos descontados  $V_\alpha$ . Más aún, los resultados mencionados en [40], [41] fueron obtenidos para procesos con espacios de estados numerable

También obtendremos cotas de la forma (2) uniformes en  $\alpha$ . En el trabajo [39] se obtienen cotas similares a (2) para una clase particular de procesos, aquellos que tienen un átomo. El método que se usa en ese trabajo consiste en reducir el problema, al estudio de la estabilidad de procesos regenerativos y se trabaja con funciones de costo de un paso acotada.

Nosotros obtendremos cotas de la forma (3) aplicando técnicas de operadores contractibles aplicados a la ecuación de optimalidad, y también obtenemos cotas de la forma (2) uniformes en  $\alpha$ , para procesos de control de Markov con funciones de costo no acotadas. Supondremos condiciones de ergodicidad en los procesos considerados, y luego reduciremos el problema, a uno de estabilidad de procesos no controlados, usando métodos de Zolotarev [46] y tasas de convergencia obtenidas en [20]. Después refinamos estas cotas usando tasas de convergencia de procesos estocásticamente ordenados [25], [37],[34] que acotan a los procesos de control cuando se aplican las políticas de control. La presente tesis se organiza de la manera siguiente:

- Capítulo 1: Se hace una descripción de los procesos de control de Markov; se definen las políticas que se van usar; el criterio de desempeño que usaremos y se plantea nuestro problema.
- Capítulo 2: Se obtienen cotas superiores de el índice de estabilidad usando técnicas de operadores de contracción aplicadas a la ecuación de optimalidad.
- Capítulo 3: Se utilizan propiedades de ergodicidad de los procesos en consideración, para encontrar tasas de convergencia a la distribución límite, usando

resultados de [20] se reduce al caso acotado para usar técnicas de Zolotarev [46] para el caso de procesos no controlados, para encontrar estimaciones uniformes del índice de estabilidad y el índice de estabilidad relativo.

Capítulo 4: Se aprovecha que algunos procesos controlados son acotados por arriba uniformemente sobre las políticas, por procesos estocásticamente ordenados, para usar las tasas de convergencia de estos, y en seguida se usan las mismas técnicas que en el capítulo 3 para encontrar cotas superiores de el índice de estabilidad.

Capítulo 5: Se dan las conclusiones y posibles extensiones de este trabajo.

# Capítulo 1

## El modelo de control

### 1.1 Introducción

Para poder plantear nuestro problema necesitamos describir (al menos brevemente) el problema de control óptimo de un proceso de Markov controlado a tiempo discreto. Esta descripción la tomaremos de [17]. Otras buenas referencias donde se puede revisar este modelo son [4], [5] y [19].

### 1.2 Procesos de control de Markov

**Definición 1.1** *Un modelo de control es una  $n$ -ada de tamaño 5*

$$(X, A, A(\cdot), c(\cdot, \cdot), Q(\cdot | \cdot, \cdot))$$

*que consiste de:*

- a)  $X$  un espacio de Borel, llamado el espacio de estados*
- b)  $A$  un espacio de Borel, llamado espacio de acciones*
- c) para cada  $x \in X$ , existe  $A(x)$  un subconjunto no-vacío de  $A$  (el conjunto de acciones admisibles).*
- d)  $Q(\cdot | x, a)$  es un kernel estocástico para  $x \in X$  y  $a \in A$  que también será llamado ley de transición*

e)  $c$  una función medible no negativa (posiblemente no acotada) que llamaremos costo por estado (costo de un paso).

Donde el conjunto

$$\mathbb{K} = \{(x, a) \mid x \in X, a \in A(x)\} \quad (1.1)$$

(conjunto de parejas admisibles) es un conjunto medible en  $X \times A$

**Nota 1** Si  $\mathcal{X}$  es un espacio topológico; y si existe un espacio separable métrico completo y un subconjunto  $\mathcal{B}$  Boreliano tal que  $\mathcal{X}$  es homeomorfo a  $\mathcal{B}$ , entonces decimos que  $\mathcal{X}$  es un espacio de Borel. Si  $X$  y  $Y$  son espacios de Borel; un kernel estocástico en  $X$  dado  $Y$  es una función  $P(\cdot|\cdot)$ , tal que  $P(\cdot|y)$  es una medida de probabilidad en  $X$  para  $y \in Y$  y  $P(B|\cdot)$  es una función medible en  $Y$  para cada  $B \in \mathbb{B}(X)$ . donde  $\mathbb{B}(X)$  es la  $\sigma$ -álgebra de Borel de  $X$ .

A la familia de todos los kernels de  $X$  dado  $Y$  la denotaremos por  $\mathcal{P}(X|Y)$

Además haremos la siguiente suposición:

**Suposición 1.1**  $\mathbb{K}$  contiene la gráfica de una función medible  $f : X \rightarrow A$ , tal que,  $f(x) \in A(x)$  para  $x \in X$ , a una función con estas propiedades se llamará selector medible, la familia de selectores medibles será denotada por  $\mathbb{F}$ .

Usaremos la definición usual de política de control [4, 14]

### 1.3 Políticas

Considere el modelo de control de la definición (1.1) y para cada  $t = 0, 1, 2, \dots$  defina el espacio  $H_t$  de historias admisibles hasta el tiempo  $t$  como  $H_0 \doteq X$  y

$$H_t \doteq \mathbb{K}^t \times H_{t-1} \text{ para } t = 1, 2, \dots$$

Un elemento genérico  $h_t \in H_t$ , que será llamado una  $t$ -historia admisible es un vector de la forma

$$h_t = (x_0, a_0, \dots, x_{t-1}, a_{t-1}, x_t)$$

con  $(x_i, a_i) \in \mathbb{K}$  para  $i = 0, \dots, t-1$  y  $x_t \in X$ . Observe que, para cada  $t$ ,  $H_t$  es un subespacio de

$$\bar{H}_t \doteq (X \times A)^t \times X, \quad t = 0, 1, 2, \dots$$

$$\text{y } \bar{H}_0 = H_0 = X$$

**Definición 1.2** Una política es una sucesión  $\pi = \{\pi_t; t = 0, 1, 2, \dots\}$  de kernels  $\pi_t$ , en el conjunto de controles  $A$  dado  $H_t$  que satisfacen que:

$$\pi_t(A(x_t) | h_t) = 1, \text{ para } h_t \in H_t \quad t = 0, 1, 2, \dots$$

El conjunto de políticas será denotado por  $\Pi$ .

Dada una política  $\pi = \{\pi_t; t = 0, 1, 2, \dots\}$  y  $\mu$  una medida de probabilidad en  $X$ , entonces por el teorema de Ionescou - Tulcea (vea el apéndice c de [17]) existe una medida de probabilidad  $P_\mu^\pi$  definida en  $(\Omega, \mathcal{F})$ , donde  $\Omega \doteq (X \times A)^\infty = \bar{H}_\infty$   $\mathcal{F}$  es la  $\sigma$ -álgebra producto y cumple lo siguiente:

- a)  $P_\mu^\pi(H_\infty) = 1$   
para  $B \in \mathbb{B}(X)$ ,  $C \in \mathbb{B}(A)$ ,  $h_t \in H_t$
- b)  $P_\mu^\pi(x_0 \in B) = \mu(B)$
- c)  $P_\mu^\pi(a_t \in C | h_t) = \pi_t(C | h_t)$
- d)  $P_\mu^\pi(x_{t+1} \in B | h_t, a_t) = Q(B | x_t, a_t)$

**Definición 1.3** El proceso estocástico  $(\Omega, \mathcal{F}, P_\mu^\pi, \{x_t\})$ , donde  $\pi$  es la política utilizada y  $\mu$  la medida inicial, será llamado proceso de control de Markov.

La mayor parte del tiempo nos restringiremos a las políticas que dependan de la historia solo a través de el último estado visitado por el proceso, y las llamaremos Markovianas.

**Definición 1.4** Una política  $\pi = \{\pi_t\} \subset \Pi$  se dice de Markov si existe una sucesión  $\{\phi_t\}$  de kernels estocásticos en  $A$  dado  $X$  tales que:

$$\phi_t(A(x)|x) = 1 \text{ y} \quad (1.2)$$

$$\pi_t(\cdot|h_t) = \phi_t(\cdot|x_t) \text{ para } h_t \in H_t, t = 0, 1, 2, \dots \quad (1.3)$$

Si en la ecuación 1.3  $\phi_t(\cdot|x_t) = \phi(\cdot|x_t)$ ,  $t = 0, 1, 2, \dots$  diremos que la política es estacionaria.

**Nota 2** Una función  $f \in \mathbb{F}$  se puede identificar con un kernel estocástico en  $A$  dado  $X$  con la condición (1.2) poniendo

$$\phi(C|x) = I_C[f(x)] \quad x \in X, C \in \mathbb{B}(A)$$

**Definición 1.5** Decimos que una política  $\pi = \{\pi_t\}$  es determinística estacionaria (D.E.) si existe una función  $f \in \mathbb{F}$  tal que:  $\pi_t(\cdot|h_t)$  está concentrada en  $f(x_t) \in A(x_t)$ ,  $h_t \in H_t$  y  $t = 0, 1, 2, \dots$  ie)  $\pi_t(C|h_t) = I_C[f(x)]$ .

Es por esto que a una política (DE)  $\pi$  se puede identificar con  $\{f, f, \dots\}$  y la denotaremos por  $\mathbf{f}$ . Al conjunto de políticas determinísticas estacionarias será denotaremos por  $\Pi_{de}$

**Notación 1** Sea  $\phi$  es una política Markoviana estacionaria; Sean  $c$  y  $Q$ , la función de costo de un paso y la ley de transición respectivamente, entonces definimos para  $x \in X$ ,

$$c(x, \phi) = \int_A c(x, a) \phi(da|x) \quad (1.4)$$

$$Q(\cdot|x, \phi) = \int_A Q(\cdot|x, a) \phi(da|x). \quad (1.5)$$

En particular para  $\mathbf{f} \in \Pi_{de}$  las ecuaciones (1.4) y (1.5) resultan

$$c(x, \mathbf{f}) = c(x, f(x)) \text{ y } Q(\cdot|x, \mathbf{f}) = Q(\cdot|x, f(x))$$

**Proposición 1.1** Sea  $\mu$  una distribución inicial arbitraria y  $f \in \Pi_{de}$  es una política (D.E.), entonces el proceso con kernel  $Q(\cdot|f)$  es un proceso de Markov -homogéneo [17]

## 1.4 Planteamiento del problema

Como criterio de desempeño usaremos *la esperanza del costo total  $\alpha$ -descontado* ( $\alpha \in (0, 1)$ ), definida para el proceso  $\mathcal{P}$  como:

**Definición 1.6** Cuando usamos la política  $\pi \in \Pi$ , el estado inicial del proceso es  $x \in X$ ,  $\alpha \in (0, 1)$  es un factor de descuento, y  $E_x^\pi$  representa la esperanza con respecto a la probabilidad  $P_x^\pi$ ;

$$V_\alpha(\pi, x) \doteq E_x^\pi \sum_{t=0}^{\infty} \alpha^t c(x_t, a_t) \quad (1.6)$$

y para el proceso  $\tilde{\mathcal{P}}$

$$\tilde{V}_\alpha(\pi, x) \doteq E_x^\pi \sum_{t=0}^{\infty} \alpha^t c(\tilde{x}_t, \tilde{a}_t) \quad (1.7)$$

**Definición 1.7** La función de costo óptimo se define por:

$$V_\alpha^*(x) \doteq \inf_{\pi \in \Pi} V_\alpha(\pi, x), \quad x \in X. \quad (1.8)$$

**Definición 1.8** Una política  $\pi_\alpha^*$  que satisfice

$$V_\alpha(\pi_\alpha^*, x) = \inf_{\pi \in \Pi} V_\alpha(\pi, x), \quad x \in X. \quad (1.9)$$

será llamada política  $\alpha$ -óptima descontada

La meta original de la optimización del control es encontrar una política  $\pi_\alpha^*$ ,  $\alpha$ -óptima para  $\mathcal{P}$ .

La optimalidad de  $\pi_*$  y  $\tilde{\pi}_*$  significa el cumplimiento de la relación (1.9) y la siguiente igualdad:

$$\tilde{V}_\alpha^*(x) = \tilde{V}_\alpha(\tilde{\pi}_*, x) \doteq \inf_{\pi \in \Pi} \tilde{V}_\alpha(\pi, x), \quad x \in X. \quad (1.10)$$

En el planteamiento del problema de la estimación de la estabilidad de un proceso de control de Markov a tiempo discreto (PCM) seguiremos el enfoque propuesto en [6], [7]. Consideraremos dos (PCM)  $\mathcal{P} = (\mathbf{X}, A, A(\cdot), c, p)$  y  $\tilde{\mathcal{P}} = (\mathbf{X}, A, A(\cdot), \tilde{c}, \tilde{p})$ . Supongamos que existen  $\pi_\alpha^*$  y  $\tilde{\pi}_\alpha^*$  políticas  $\alpha$ -óptimas

para  $\mathcal{P}$  y  $\tilde{\mathcal{P}}$  respectivamente; y que desconocemos tanto  $p$  como  $c$ . Por otro lado, supongamos que conocemos  $\tilde{p}$  y  $\tilde{c}$ . Interpretaremos a  $\tilde{p}$  como la aproximación (conocida) a una probabilidad de transición "verdadera"  $p$  (desconocida) de un proceso  $\mathcal{P}$  y que  $\tilde{c}$  esta cercano a  $c$ . En esta situación no es posible obtener  $\pi_\alpha^*$ , sin embargo podemos utilizar  $\tilde{\pi}_\alpha^*$  como una aproximación a  $\pi_\alpha^*$ .

Nuestro **problema** consiste en dar estimaciones del costo adicional al usar  $\tilde{\pi}_\alpha^*$ , como política de control de nuestro proceso en lugar de la  $\alpha$ -óptima, para estimar este costo necesitamos las siguientes definiciones.

**Definición 1.9** *El índice de estabilidad se define como:*

$$\Delta_\alpha(x) \doteq [V_\alpha(\tilde{\pi}_\alpha^*, x) - V_\alpha(\pi_\alpha^*, x)] \geq 0. \quad (1.11)$$

**Definición 1.10** *El índice de estabilidad relativo se define como:*

$$\bar{\Delta}_\alpha(x) \doteq \frac{V_\alpha(\tilde{\pi}_\alpha^*, x) - V_\alpha(\pi_\alpha^*, x)}{V_\alpha(\pi_\alpha^*, x)}. \quad (1.12)$$

Diremos que un (PCM)  $\mathcal{P}$  es estable (estable relativo) respecto a la métrica probabilística  $m$  si  $\Delta_\alpha(x) \rightarrow 0$  ( $\bar{\Delta}_\alpha(x) \rightarrow 0$ ) cuando  $m(p, \tilde{p}) \rightarrow 0$ . Así nuestro objetivo principal es, obtener condiciones que garanticen la estabilidad de nuestro proceso y obtener cotas superiores para las tasas de esta convergencia. Dicho de otra manera queremos encontrar desigualdades y funciones  $g_x, g_x^{(1)} : \mathbb{R} \rightarrow \mathbb{R}$  tales que:

$$\Delta_\alpha(x) \leq g_x(m(p, \tilde{p})); \quad g_x(s) \rightarrow 0, \text{ cuando } s \rightarrow 0 \quad (1.13)$$

o bien

$$\bar{\Delta}_\alpha(x) \leq g_x^{(1)}(m(p, \tilde{p})); \quad g_x^{(1)}(s) \rightarrow 0, \text{ cuando } s \rightarrow 0 \quad (1.14)$$

Cuando  $m$  es la métrica de la variación total, o bien una métrica ponderada.

Nos interesa también que las cotas en (1.14) sean uniformes sobre  $\alpha \in (0, 1)$  para saber el comportamiento  $\bar{\Delta}_\alpha(x)$  cuando el factor de descuento  $\alpha \rightarrow 1$ , i.e.

$$\sup_{\alpha \in (0, 1)} \bar{\Delta}_\alpha(x) \leq g_x^{(2)}(m(p, \tilde{p}))$$

donde  $g_x^{(2)} : \mathbb{R} \rightarrow \mathbb{R}$  y  $g_x^{(2)}(s) \rightarrow 0$ , cuando  $s \rightarrow 0$ .

Es importante observar que existen ejemplos para los cuales  $\bar{\Delta}_\alpha(x) \rightarrow \infty$  a pesar de que  $m(p, \tilde{p}) \rightarrow 0$ , cuando  $m$  es la métrica de Fortét-Mourier (vea [30] para la definición)

## Capítulo 2

# Estimación del índice de estabilidad

### 2.1 Introducción

Dados  $\mathcal{P} = (X, A, A(x), c, p)$  y  $\tilde{\mathcal{P}} = (X, A, A(x), \tilde{c}, \tilde{p})$ , dos procesos de control de Markov y considere los costos  $\alpha$ -descontados

$$V_{\alpha}(\pi, x) = E_x^{\pi} \left[ \sum_{t=0}^{\infty} \alpha^t c(x_t, a_t) \right] \text{ y}$$
$$\tilde{V}_{\alpha}(\pi, x) = E_x^{\pi} \left[ \sum_{t=0}^{\infty} \alpha^t \tilde{c}(\tilde{x}_t, a_t) \right],$$

donde  $\pi \in \Pi$ ,  $x \in X$ .

Nos interesa obtener cotas superiores para el índice de estabilidad  $\Delta_{\alpha}(x)$  definido en (1.11). Cotas para este índice se han encontrado en [7], para dos problemas de optimización; procesos de control de Markov para el caso promedio con recompensas acotadas, y sistemas de control a tiempo continuo definidos por ecuaciones diferenciales. Nosotros trabajaremos el caso descontado con costos no acotados.

Describiremos modelos con propiedades que garantizan la existencia de políticas óptimas, además, como utilizaremos la misma herramienta (proposición 3.2) que ha sido usada en [7], necesitamos demostrar la existencia de puntos fijos de algunos operadores relacionados con cada política admisible  $\pi \in \Pi$ .

Por otro lado, para poder manejar costos no acotados relacionaremos la tasa de crecimiento de la función de costo de un paso  $c(x, a)$  en  $x$ , con el máximo (sobre las políticas) del desplazamiento promedio del proceso en un paso. Para lograr estos objetivos hacemos las siguientes suposiciones.

## 2.2 Suposiciones

**Suposición 2.1** *Existe una función medible  $w : X \rightarrow (0, \infty)$  para la cual se cumple lo siguiente :*

a)  $c(x, a) \leq w(x)$  para cada  $(x, a) \in \mathbb{K}$

b) *Existe una constante  $\beta \in [\alpha, 1)$  tal que*

$$\int_X w(y) p(dy|x, a) \leq \frac{\beta}{\alpha} w(x) \text{ para toda } (x, a) \in \mathbb{K},$$

donde  $\alpha$  es el factor de descuento considerado en la definición 1.6.

**Nota 3** *La función  $w$  y la constante  $\beta$  en las Suposiciones 2.1 se consideran fijas de aquí en adelante. En Wessels [44] se introduce esta suposición, para tratar problemas relacionados con procesos de control de Markov, diferentes a los aquí tratados .*

**Definición 2.1** *Por  $M_w$  denotaremos el espacio lineal normado de funciones medibles  $u : X \rightarrow \mathbb{R}$ , con norma ponderada :*

$$\|u\|_w \doteq \sup_x \frac{|u(x)|}{w(x)} < \infty \quad (2.1)$$

Para asegurar la existencia de políticas óptimas, impondremos las siguientes condiciones .

**Suposición 2.2** *Para cada  $x \in X$  :*

a) *El conjunto  $A(x)$  es compacto ;*

- b)  $c(x, a)$  es semicontinua por abajo, como función de  $a$ , en  $A(x)$
- c)  $a \rightarrow \int_X u(y) p(dy|x, a)$  es una función continua para cada  $u \in B(X)$ , donde  $B(X)$  es el conjunto de funciones medibles y acotadas
- d) Para cada  $x \in X$ , la función  $\bar{w}(x, a) \doteq \int w(y) p(dy|x, a)$  es continua en  $a \in A(x)$
- e) para cada  $u \in M_w$  no-negativa y  $r \in \mathbb{R}$  el siguiente conjunto es de Borel en  $\mathbb{K}$

$$\left\{ (x, a) \in \mathbb{K} : \int_X u(y) p(dy|x, a) \leq r \right\}.$$

**Nota 4** De aquí en adelante la suposición (2.2) se cumple para cualquier (PCM) que se considere.

### 2.3 El Problema de control óptimo.

El problema de control óptimo consiste en encontrar una política  $\pi^*$   $\alpha$ -descontada óptima i.e.,  $\pi^*$  satisface

$$V_\alpha(\pi^*, x) = \inf_{A(x)} V_\alpha(\pi, x) = V_\alpha^*(x), \quad x \in X.$$

Para garantizar la existencia de políticas óptimas obtendremos la función  $V^*$  como la solución puntual de la ecuación de optimalidad  $\alpha$ -descontada, que a continuación establecemos,

$$v(x) = \inf_{A(x)} \left[ c(x, a) + \alpha \int v(y) p(dy|x, a) \right] \quad \text{para } x \in X. \quad (2.2)$$

El siguiente Teorema resuelve el problema de control óptimo

**Teorema 2.1** *Aceptemos que se cumplen las suposiciones 2.1 y 2.2*

- a) entonces, la función de valor satisface la ecuación 2.2 y existe una función  $f^* \in \mathbb{F}$  tal que  $f^*(x) \in A(x)$  alcanza el mínimo en la ecuación

2.2, para cualquier estado  $x \in X$ , dicho de otra manera (según notación 1)

$$V^*(x) = c(x, a) + \alpha \int V^*(y) p(dy | x, a) \quad \text{para } x \in X, \quad (2.3)$$

y la política  $f^* = \{f, f, f, \dots\}$  es  $\alpha$ -óptima; reciprocamente, si  $f \in \Pi_{de}$  es  $\alpha$ -descontada óptima entonces se satisface la ecuación 2.3.

- b) Una política  $\pi^*$  es  $\alpha$ -descontada óptima si sólo si la función de costo correspondiente  $V_\alpha(\pi^*, \cdot)$  satisface la ecuación de optimalidad (2.3)
- c) Si existe una política  $\alpha$ -óptima descontada, entonces existe una política determinística estacionaria que es  $\alpha$ -óptima descontada,

Antes de demostrar el Teorema 2.3, estableceremos algunos resultados e introduciremos el operador de programación dinámica  $T_\alpha$ .

**Definición 2.2** (Operador de programación dinámica) Dado una función  $u : X \rightarrow \mathbb{R}$  y  $0 < \alpha < 1$ , denotamos por  $T_\alpha u$  a la función dada por,

$$T_\alpha u(x) \doteq \inf_{A(x)} \left[ c(x, a) + \alpha \int_X u(y) p(dy | x, a) \right], \quad \text{para } x \in X. \quad (2.4)$$

Siempre que la integral esté bien definida.

Probaremos que  $T_\alpha$  es un operador de contracción de  $M_w$  en sí mismo.

**Lema 2.1** Sea

$$v : \mathbb{K} \rightarrow \mathbb{R}, \quad \text{donde } \mathbb{K} = \{(x, a) \mid x \in X, a \in A(x)\}$$

una función medible. Definimos

$$v^*(x) = \inf_{A(x)} v(x, a) \quad \text{para } x \in X, \quad (2.5)$$

donde  $A(x)$  es el conjunto de acciones admisibles y consideremos que se cumplen las suposiciones 2.1 y 2.2. Entonces, si  $v(x, \cdot)$  es semicontinua por abajo en  $A(x)$  para  $x \in X$  entonces, existe una función  $f \in \mathbb{F}$  tal que,  $f(x)$  alcanza el mínimo en (2.5), para cada  $x \in X$ , esto es:

$$v^*(x) = \inf_{A(x)} v(x, a)$$

y  $v^*$  es una función medible.

**Dem..** Vea proposición D.3 en el apéndice D del libro [14] ■

**Proposición 2.1** Consideremos que se cumplen las suposiciones 2.1 y 2.2, y sea  $T_\alpha$  el operador de programación dinámica definido en (2.4). Entonces,

(a)  $T_\alpha$  es un operador de contracción en  $M_w$ , con módulo de contracción  $\beta$  i.e.,  $T_\alpha$  mapea a  $M_w$  en sí mismo y

$$\|T_\alpha u - T_\alpha \acute{u}\|_w \leq \beta \|u - \acute{u}\|_w \text{ para } u, \acute{u} \in M_w; \quad (2.6)$$

(b) Para cualquier función  $u \in M_w$  existe un selector  $f_u \in \mathbb{F}$  tal que,

$$T_\alpha u(x) = c(x, f_u(x)) + \alpha \int u(y) p(dy|x, f(x)), \text{ para } x \in X, \quad (2.7)$$

para cualquier  $u \in M_w$ .

**Dem..** De la Suposición 2.2 b) y c), la función:

$$v(x, a) = c(x, a) + \alpha \int u(y) p(dy|x, a),$$

es semicontinua por abajo en  $a \in A(x)$ , para  $x \in X$ . Por lo tanto el Lema 2.1 nos asegura que  $T_\alpha u$  es una función medible, y que existe  $f_u \in \mathbb{F}$  que satisface la ecuación 2.7.

Veamos que  $\|T_\alpha u\|_w < \infty$ , si  $u \in M_w$ . Por la suposición 2.1

$$\begin{aligned} T_\alpha u(x) &\leq w(x) + \alpha \int w(y) u(y) w(y)^{-1} p(dy|x, a) \\ &\leq w(x) + \alpha \|u\|_w \int w(y) p(dy|x, a) \\ &\leq w(x) + \beta \|u\|_w w(x) = w(x) (1 + \beta \|u\|_w), \end{aligned} \quad (2.8)$$

entonces  $\|T_\alpha u\|_w < \infty$  si  $\|u\|_w < \infty$ . Por lo tanto,  $T_\alpha$  mapea  $M_w$  en sí mismo, más aún, sabiendo que  $|\inf f(y) - \inf g(y)| \leq \sup_y |f(y) - g(y)|$  (vea [19])

$$\begin{aligned} |T_\alpha v(x) - T_\alpha u(x)| &\leq \sup_{A(x)} \alpha \int |v(y) - u(y)| p(dy|x, a) \\ &\leq \alpha \|v - u\|_w \int w(y) p(dy|x, a) \\ &\leq \alpha \|v - u\|_w \frac{\beta}{\alpha} w(x). \end{aligned}$$

De donde se sigue el resultado (b). ■

También necesitaremos la siguiente proposición.

**Proposición 2.2** Para cada  $f \in \mathbb{F}$  el operador

$$T_f v(x) \doteq c(x, f(x)) + \alpha \int u(y) p(dy | x, f(x))$$

es una contracción en  $M_w$ , cuyo único punto fijo es  $V(f, x)$ .

**Dem..** La demostración de que es una contracción de  $M_w$ , se hace siguiendo los mismos argumentos que los usados para demostrar que el operador de programación dinámica es una contracción en  $M_w$ .

Sea

$$V(x, f) = E_x^f \left( \sum_{t=0}^{\infty} \alpha^t c(x_t, a_t) \right) = c(x, a) + \alpha E_x^f \left( \sum_{t=1}^{\infty} \alpha^{t-1} c(x_t, a_t) \right), \quad (2.9)$$

pero tenemos que,

$$\begin{aligned} E_x^f \left( \sum_{t=1}^{\infty} \alpha^{t-1} c(x_t, a_t) \right) &= E_x^f \left[ E_x^f \left( \sum_{t=1}^{\infty} \alpha^{t-1} c(x_t, a_t) \mid h_1 \right) \right] \\ &= E_x^f \left[ E_y^f \left( \sum_{t=1}^{\infty} \alpha^{t-1} c(x_t, a_t) \right) \right] \\ &= E_x^f (V(y, f)) = \int V(y, f) p(dy | x, f(x)), \end{aligned}$$

entonces la ecuación 2.9 nos queda,

$$\begin{aligned} V(x, f) &= c(x, f) + \alpha \int V(y, f) p(dy | x, f(x)) \\ &= T_f V(x, f), \end{aligned}$$

que es lo que queríamos probar ■

**Dem.. (del Teorema 2.1)**

Por la Proposición 2.1 y el Teorema del punto fijo de Banach,  $T_\alpha$  tiene un único punto fijo  $u^* \in M_w$ , i.e.

$$T_\alpha u^* = u^*.$$

Probaremos que  $u^* = V^*$ . Obsevemos que, para  $\pi \in \Pi$ ,  $x \in X$  y  $u \in M_w$

$$E_x^\pi (|u(x_t)|) \leq \|u\|_w E_x^\pi (w(x_t)) \leq \|u\|_w \frac{\beta^t}{\alpha^t} w(x),$$

por lo tanto

$$\lim_{t \rightarrow \infty} \alpha^t E_x^\pi u(x_t) = 0. \quad (2.10)$$

Consideremos la igualdad  $T_\alpha u^* = u^*$ , por la Proposición 2.1 existe un selector  $f^* \in \mathbb{F}$  tal que,

$$u^*(x) = c(x, f^*) + \alpha \int u^*(y) p(dy | x, f^*), \quad (2.11)$$

iterando 2.11 obtenemos,

$$u^*(x) = E_x^{f^*} \left( \sum_{t=0}^{n-1} \alpha^t c(x_t, f^*) \right) + \alpha^n E_x^{f^*} u^*(x_n),$$

tomando  $n \rightarrow \infty$  de (2.10) obtenemos,

$$u^*(x) = E_x^{f^*} \left( \sum_{t=0}^{\infty} \alpha^t c(x_t, f^*) \right) = V(x, f^*).$$

De la definición de  $V^*$  resulta  $u^*(x) \geq V^*(x)$ .

Para obtener la desigualdad reciproca note que,

$$u^*(x) \leq c(x, a) + \alpha \int u^*(y) p(dy | x, a), \text{ para } (x, a) \in \mathbb{K}, \quad (2.12)$$

entonces para cada política  $\pi \in \Pi$  y cada estado inicial  $x \in X$  la desigualdad 2.12 produce

$$\alpha^t E_x^\pi [c(x_t, a_t) + \alpha u^*(x_{t+1}) - u^*(x_t) | h_t, a_t] \geq 0, \quad (2.13)$$

esto produce,

$$u^*(x) \leq E_x^\pi \left( \sum_{t=0}^{n-1} \alpha^t c(x_t, a_t) \right) + \alpha^n E_x^\pi u^*(x).$$

Tomando  $n \rightarrow \infty$  se sigue que,

$$u^*(x) \leq V(x, \pi). \quad (2.14)$$

Por lo tanto  $u^*(x) = V^*(x, \pi)$ .

Dado  $u^*$  por la Proposición 2.1 existe  $f \in \mathbb{F}$  tal que,

$$T_\alpha u^*(x) = c(x, f^*) + \alpha \int u^*(y) p(dy | x, f^*) = u^*(x) = V^*.$$

Recíprocamente, para cualquier política  $f^* \in \Pi_{de}$   $\alpha$ -óptima descontada, (por el Lema 2.2), el costo  $\alpha$ -descontado satisface

$$V(x, f^*) \doteq c(x, f^*(x)) + \alpha \int u(y) p(dy | x, f^*(x)),$$

como  $f^*$  es  $\alpha$ -óptima entonces  $V(x, f^*) = V^*(x)$ .

Para la demostración de b) y c) vea [18] ■

## 2.4 Desigualdad de estabilidad del problema de control óptimo

Si la política  $\tilde{\pi}_*$ , que es óptima para  $\tilde{\mathcal{P}}$ , con las aproximaciones  $\tilde{p}$  y  $\tilde{c}$ , se usa para controlar  $\mathcal{P}$  (en lugar de  $\pi_*$ ), entonces, el deterioro de la calidad del control, en términos del índice de estabilidad queda:

$$\Delta_\alpha(x) \doteq V_\alpha(x, \tilde{\pi}_*) - V_\alpha(x, \pi_*) \quad (2.15)$$

donde  $x \in X$  es un estado inicial arbitrario, pero fijo de aquí en adelante.

Primero recordaremos la Proposición 3.2 de [7]. Sean  $X, U$  dos conjuntos y  $\rho$  una semimétrica definida en  $X$ .

Considere las siguientes funciones real valuadas:

$$F : U \times X \rightarrow \mathbb{R}$$

$$\tilde{F} : U \times X \rightarrow \mathbb{R},$$

definidas en el producto Cartesiano de  $U$  y  $X$ . Y para cada  $u \in U$  tenemos operadores:

$$\begin{aligned} T_u &: X \rightarrow X \\ \tilde{T}_u &: X \rightarrow X, \end{aligned}$$

tales que, tienen conjuntos no vacíos de puntos fijos  $\{x_u\}$  y  $\{\tilde{x}_u\}$  respectivamente.

Suponga que existe una regla para seleccionar un único punto fijo de  $\{x_u\}$  y un único punto fijo de  $\{\tilde{x}_u\}$ .

**Suposición 2.3 (S0)** Existen  $u_*$  y  $\tilde{u}_* \in U$  tal es que

$$\begin{aligned} F(u_*, x_{u_*}) &= \inf \{F(u, x_u) : u \in U\} < \infty, \\ \tilde{F}(\tilde{u}_*, x_{\tilde{u}_*}) &= \inf \{\tilde{F}(u, \tilde{x}_u) : u \in U\} < \infty. \end{aligned}$$

Para dos espacios métricos  $(X, \rho)$ ,  $(Y, r)$  y un número real positivo  $d$  denotamos

$$Lip_d = \{\phi : X \rightarrow Y \mid r(\phi(x_1), \phi(x_2)) \leq d\rho(x_1, x_2)\} \text{ para } x_1, x_2 \in X.$$

Sea  $X_0 \subset X$  un conjunto fijo para el cual  $\cup_{u \in U} \{x_u, \tilde{x}_u\} \subset X_0$ .

**Suposición 2.4 (S1)** Para toda  $x_1, x_2 \in X$

$$|F(u, x_1) - F(u, x_2)| \leq L[\rho(x_1, x_2)], \quad u \in U;$$

donde  $L(s)$ ,  $s \geq 0$  es una función continua no decreciente.

**Suposición 2.5 (S3)** Sea  $X_0 \subset X$  un conjunto tal que,  $\cup_{u \in U} \{x_u, \tilde{x}_u\} \in X_0$  entonces,

$$\tilde{T}_u \in Lip_d(X_0, X), \quad 0 < d < 1$$

Donde  $\tilde{T}_u$  es la restricción de  $T_u$  al conjunto  $X_0$ .

**Proposición 2.3 (3.2 de [7])** Considere ciertas las suposiciones S0, S1, S3 entonces,

$$\Delta \doteq F(u_*, x_{u_*}) - F(\tilde{u}_*, x_{\tilde{u}_*}) \leq 2(\varepsilon + L[\delta_0/(1-d)]),$$

donde

$$\begin{aligned} \varepsilon &= \sup \left\{ \left| \tilde{F}(u, x) - F(u, x) \right| : u \in U, x \in X_0 \right\} \\ \delta &= \sup \left\{ \rho(\tilde{T}_u, T_u) : u \in U, x \in X_0 \right\} \end{aligned}$$

El siguiente teorema nos da una estimación de  $\Delta_\alpha(x)$ .

**Teorema 2.2** *Si la suposición (2.1) y (2.2) se cumplen para dos (PCM)  $\mathcal{P}$  y  $\tilde{\mathcal{P}}$  entonces,*

$$\Delta_\alpha(x) \leq 2w(x)(1-\beta)^{-1} [\delta_1 + \delta_2\alpha(1-\beta)^{-1}] \quad (2.16)$$

donde

$$\delta_1 \doteq \sup_{(z,a) \in \mathbb{K}} [w(z)^{-1}] |c(z,a) - \tilde{c}(z,a)|$$

$$\delta_2 \doteq \sup_{(z,a) \in \mathbb{K}} [w(z)^{-1}] \times \int_{\mathbf{X}} w(y) \|p(dy|z,a) - \tilde{p}(dy|z,a)\| \quad (2.17)$$

y  $\|p(dy|z,a) - \tilde{p}(dy|z,a)\|$  es la variación de  $p - \tilde{p}$ .

**Nota 5** (a) *Sea  $\mu$  una medida con signo, según la descomposición de Jordan-Hahn, existe un conjunto medible  $D$  tal que,  $\mu(\cdot \cap D)$  y  $-\mu(\cdot \cap D^c)$  son medidas positivas, denotadas por  $\mu^+$  y  $\mu^-$ , respectivamente. La variación total de  $\mu$  denotada por  $\|\mu\|$  se define como:*

$$\mu = \mu^+ + \mu^-$$

(b) *Cuando existen densidades  $\rho(\cdot|z,a)$ ,  $\tilde{\rho}(\cdot|z,a)$ , para  $p(\cdot|z,a)$  y  $\tilde{p}(\cdot|z,a)$  respectivamente, entonces*

$$\int_{\mathbf{X}} w(y) \|p(dy|z,a) - \tilde{p}(dy|z,a)\| = \int_{\mathbf{X}} w(y) \|\rho(dy|z,a) - \tilde{\rho}(dy|z,a)\|$$

*Esta última expresión nos permite calcular  $\delta_2$  (definida en la ecuación 2.17), en ejemplos particulares.*

**Dem..** Tomemos  $U = \Pi_{de}$  el conjunto de las políticas estacionarias y determinísticas y  $X = M_w$  como en la definición 2.1,  $\rho$  la métrica en  $M_w$  inducida por la norma ponderada  $\|\cdot\|_w$ ; y definimos

$$\begin{aligned} F &: \Pi_{de} \times M_w \rightarrow \mathbb{R}, \text{ por} \\ F(\mathbf{f}, u) &= u(x) \quad (x \in X \text{ fijo}) \\ \tilde{F} &= F. \end{aligned}$$

Para cada política  $\mathbf{f} \in \Pi_{de}$  definamos los operadores (recuerde que  $\mathbf{f} = \{f, f, \dots\}$  para algún  $f \in \mathbb{F}$ )

$$\begin{aligned} T_f u(x) &\doteq c(x, f(x)) + \alpha \int_X u(y) p(dy | x, f(x)), \\ \tilde{T}_f u(x) &\doteq \tilde{c}(x, f(x)) + \alpha \int_X u(y) \tilde{p}(dy | x, f(x)), \quad x \in X; u \in M_w. \end{aligned}$$

La suposición 2.1 implica que  $T_f, \tilde{T}_f$  son operadores que mapean  $M_w$  sobre sí mismo, más aún, estos operadores son contracciones (vea [23], [42]) con factor de contracción  $\beta$  ( $\beta$  igual que en la suposición 2.1) i.e.

$$\|T_f u_1 - T_f u_2\|_w \leq \beta \|u_1 - u_2\|_w \quad (2.18)$$

$$\|\tilde{T}_f u_1 - \tilde{T}_f u_2\|_w \leq \beta \|u_1 - u_2\|_w \quad (2.19)$$

para todo  $u_1, u_2 \in M_w$ . Por lo tanto existen puntos fijos únicos  $u_f, \tilde{u}_f \in M_w$  de  $T_f$  y  $\tilde{T}_f$  para  $f \in \mathbb{F}$ , respectivamente. Por el Lema 2.2 tenemos la siguiente igualdad

$$V_\alpha(x, \mathbf{f}) = c(x, f(x)) + \alpha \int_X V_\alpha(y, f) p(dy | x, f(x)), \quad x \in X, f \in \mathbb{F}.$$

Por otro lado, el Teorema 2.1 garantiza la existencia de  $\mathbf{f}^*$  y  $\tilde{\mathbf{f}}^* \in \Pi_{de}$   $\alpha$ -óptimas descontadas, por lo tanto  $F(\mathbf{f}^*, u_{f^*}) = V(x, \mathbf{f}^*) = \inf \{F(f, u_f)\}$ , donde tomamos  $u_f = V(x, \mathbf{f})$ , de la misma manera  $u_{\tilde{f}} = V(x, \tilde{\mathbf{f}})$  por lo tanto se cumple la suposición S0 de la Proposición 3.2 en [7].

Si tomamos  $L(s) = sw(x)$  (claramente continua y no-decreciente) tendremos

$$|F(f, u_1) - F(f, u_2)| \leq |u_1(x) - u_2(x)| \leq \|u_1 - u_2\|_w w(x)$$

entonces se cumple la Suposición S1.

Si tomamos  $X_0 = \cup_{f \in \Pi_{de}} \{u_f, u_{\tilde{f}}\}$  la Suposición 3 se cumple automáticamente.

Así, podemos aplicar la Proposición 3.2 de [7], que proporciona la siguiente estimación del índice de estabilidad

$$\Delta_\alpha(x) \leq \frac{2w(x)}{1-\beta} \sup_{f \in \mathbb{F}, u \in U_0} \|T_f u - \tilde{T}_f u\|_w, \quad (2.20)$$

donde  $U_0 = \cup_{f \in \Pi_{de}} \{\tilde{u}_f\}$ . Para hacer uso de (2.20) primero damos una cota para  $\|\tilde{u}_f\|_w$ .

Como

$$\tilde{u}_f(x) = \tilde{V}_\alpha(x, f(x)) = \tilde{E}_x^\pi \left[ \sum_{t=0}^{\infty} \alpha^t \tilde{c}(x_t, f(x_t)) \right],$$

de la suposición 2.1 (a) para el proceso  $\tilde{\mathcal{P}}$  obtenemos

$$\begin{aligned} |\tilde{u}_f|/w(x) &\leq [w(x)]^{-1} \tilde{E}_x^\pi \left[ \sum_{t=0}^{\infty} \alpha^t w(x_t) \right] \\ &\leq [w(x)]^{-1} \left[ \sum_{t=0}^{\infty} \alpha^t H^t w(x_t) \right], \quad x \in X; \end{aligned} \quad (2.21)$$

donde el operador  $H$  se define en  $M_w$  como

$$Hu(x) = \sup_{a \in A(x)} \int_X u(y) \tilde{p}(dy|x, a), \quad x \in X.$$

La última desigualdad en (2.21) fue probada en [16]. Por otra parte,  $H_0 w(x) \doteq w(x)$  y si  $H^{t-1} w(x) \leq \left(\frac{\beta}{\alpha}\right)^{t-1} w(x)$ ,  $t = 2, 3, \dots$  entonces,

$$H^t w(x) = \sup_{a \in A(x)} \int_X H^{t-1} w(y) \tilde{p}(dy|x, a) \leq \left(\frac{\beta}{\alpha}\right)^t w(x) \quad (2.22)$$

por la suposición 2.1 (b) para el proceso  $\tilde{\mathcal{P}}$ . Por lo tanto las ecuaciones (2.21) y (2.22) implican  $\|\tilde{u}_f\|_w \leq (1 - \beta)^{-1}$  para cualquier  $f \in \mathbb{F}$ . Por esta razón para cualquier  $u \in U_0$  obtenemos

$$\begin{aligned} \left\| T_f u - \tilde{T}_f u \right\|_w &\leq \sup_{z \in X} [w(z)]^{-1} |c(z, f(z)) - \tilde{c}(z, f(z))| + \\ &\alpha \sup_{z \in X} [w(z)]^{-1} \int_X |u(y)| |p(dy|z, f(z)) - \tilde{p}(dy|z, f(z))| \\ &\leq \sup_{(z,a) \in \mathbb{K}} [w(z)]^{-1} |c(z, a) - \tilde{c}(z, a)| + \\ &\alpha (1 - \beta)^{-1} \sup_{(z,a) \in \mathbb{K}} [w(z)]^{-1} \int |w(y)| |p(dy|z, a) - \tilde{p}(dy|z, a)|. \end{aligned} \quad (2.23)$$

Finalmente la desigualdad (2.16) se sigue de (2.20) y (2.23). ■

**Nota 6** Como la suposición (2.2) solo se necesita para garantizar la existencia de políticas óptimas, podemos obtener la siguiente versión de la desigualdad (2.16) :

$$\tilde{\Delta}_\varepsilon \leq \varepsilon + 2w(x) (1 - \beta)^{-1} [\delta_1 + \delta_2 \alpha (1 - \beta)^{-1}]. \quad (2.24)$$

Donde definimos el índice de estabilidad como sigue :

$$\tilde{\Delta}_\varepsilon = V_\alpha(x, \tilde{f}_\varepsilon) - \inf_{\pi \in \Pi} V_\alpha(x, \pi)$$

$\varepsilon > 0$  ;  $\tilde{f}_\varepsilon$  es una política  $\varepsilon$ -óptima para el proceso  $\tilde{\mathcal{P}}$  ,  $\varepsilon > 0$  ; para cada  $x \in X$  que el conjunto  $A(x)$  es  $\sigma$ -compacto, y que para cada  $x \in X$  ,  $r \in \mathbb{R}$  ,  $u \in M_w$  los conjuntos ,

$$\begin{aligned} \{(x, a) \in \mathbb{K} : \int_X u(y) p(dy|x, a) \leq r\}, \\ \{(x, a) \in \mathbb{K} : \int_X u(y) \tilde{p}(dy|x, a) \leq r\} \end{aligned}$$

son subconjuntos de Borel de  $\mathbb{K}$  ,

Para probar la desigualdad (2.24) usaremos los argumentos de la prueba del Teorema 2.2, y una ligera generalización del Lema 2 en [6] para políticas  $\varepsilon$ -óptimas. La único que hay que probar es que

$$\begin{aligned} \inf_{f \in \mathbb{F}} V_\alpha(x, f) &= \inf_{\pi \in \Pi} V_\alpha(x, \pi) \text{ y} \\ \inf_{f \in \mathbb{F}} \tilde{V}_\alpha(x, f) &= \inf_{\pi \in \Pi} \tilde{V}_\alpha(x, \pi) . \end{aligned}$$

Esto es para asegurar la existencia de políticas  $\varepsilon$ -óptimas estacionarias para los procesos  $\mathcal{P}$  y  $\tilde{\mathcal{P}}$ .

Bajo la suposición (2.1) y condiciones de medibilidad generales el Lema 2 en [42] prueba que el operador :

$$Tu(x) \doteq \inf_{a \in A(x)} \left\{ c(x, a) + \alpha \int_X u(y) p(dy | x, a) \right\}, x \in X; u \in M_w$$

es una contracción en  $M_w$ ; así, concluimos que existe una única solución  $V_\alpha^* \in M_w$  de la siguiente ecuación :

$$V_\alpha^*(x) = \inf_{a \in A(x)} \left\{ c(x, a) + \alpha \int_X V_\alpha^*(y) p(dy | x, a) \right\}. \quad (2.25a)$$

Esta solución es exactamente la función de valor definida en la ecuación (1.7). Usando 2.25a y una técnica estandar (vea, por ejemplo [2] cap. 9), encontramos que para probar la existencia de una política  $\varepsilon$ -estacionaria para el proceso  $\mathcal{P}$ , es suficiente probar que para cada  $\varepsilon > 0$  existe un  $\varepsilon$ -minimizador de la función  $v(a) = c(x, a) + \alpha \int_X V_\alpha^*(y) p(dy | x, a)$ , en el conjunto  $A(x)$ . Pero, la existencia de un tal  $\varepsilon$ -minimizador se puede derivar fácilmente de el Corolario 4.3 en [31] usando solo la  $\sigma$ -compacidad de  $A(x)$  y de las condiciones de medibilidad dadas en la nota 6. Argumentos similares se cumplen para el proceso  $\tilde{\mathcal{P}}$ .

## 2.5 Ejemplo I

En esta sección consideramos, un proceso de control de Markov particular, definido por las ecuaciones :

$$x_{t+1} = (x_t + a_t - \xi_t)^+, t = 0, 1, \dots; \quad (2.26)$$

donde  $x_t \in X \doteq [0, \infty)$ ,  $a_t \in A(x) = A$ , siendo  $A$  un subconjunto compacto de un intervalo finito de la forma  $(0, \theta]$ ,  $\theta \in A$ . En (2.26) las  $\{\xi_t\}$  forman una sucesión de variables aleatorias no-negativas independientes e idénticamente distribuidas. El proceso de control (2.26) aparece en teoría de inventarios([4], [14]) y en el sistema de colas del tipo  $GI | D | 1 | \infty$  con tasa de servicio controlable. En este último caso  $x_t$  y  $a_t$  denotan el tiempo de

espera y el tiempo de servicio del  $t$ -ésimo cliente ( $t = 0, 1, \dots$ ), mientras que  $\xi_t$  representa el tiempo entre la llegada del  $t$ -ésimo y el  $t+1$ -ésimo cliente.

Al proceso

$$\bar{x}_{t+1} = \left( \bar{x}_t + a_t - \tilde{\xi}_t \right)^+, t = 0, 1, \dots \quad (2.27)$$

que difiere del proceso (2.26) por la distribución de  $\tilde{\xi}_t$ , lo consideramos como una aproximación de éste.

Por  $\xi$  y  $\tilde{\xi}$  denotamos variables aleatorias genéricas distribuidas como  $\xi_t$  y  $\tilde{\xi}_t$  respectivamente.

**Suposición 2.6** (Condiciones de ergodicidad y continuidad).

- a)  $E\xi > \theta$  y  $E\tilde{\xi} > \theta$ ;
- b) las variables  $\xi$  y  $\tilde{\xi}$  tienen densidades acotadas y continuas en  $[0, \infty)$  y serán denotadas por  $\rho$  y  $\tilde{\rho}$ .

Como se demuestra en la proposición 5.4 en [10] la suposición 2.6 (a) produce las siguientes desigualdades:

$$\sup_{a \in A} Ee^{q(x+a-\xi)^+} \leq e^{qx} + 1;$$

$$\sup_{a \in A} Ee^{q(x+a-\tilde{\xi})^+} \leq e^{qx} + 1, x \in [0, \infty), \quad (2.28)$$

donde  $q = \min \left\{ 1, \sup (s : e^{s\theta} Ee^{-s\xi} \leq 1), \sup (s : e^{s\theta} Ee^{-s\tilde{\xi}} \leq 1) \right\} > 0$

**Suposición 2.7** Las funciones de costo son iguales ( $c = \tilde{c}$ ) no-negativas, medibles y existe una constante  $b \geq 1$  tal que:

$$c(x, a) \leq be^{qx}, x \in [0, \infty), a \in A \quad (2.29)$$

**Proposición 2.4** Dado que las suposiciones (2.6) y (2.7) se cumplen, entonces

$$\Delta_\alpha(x) \leq \frac{2b\alpha}{(1-\beta)^2} (e^{qx} + \gamma) e^{q\theta} \int_0^\infty |\rho(s) - \tilde{\rho}(s)| ds, \quad (2.30)$$

donde  $\beta$  es un número arbitrario en  $(\alpha, 1)$  y  $\gamma = \left(\frac{\beta}{\alpha} - 1\right)^{-1}$ .

**Dem..** Se prueba en [10] que la suposición 2.6 b) implica las suposiciones (2.2) para los procesos definidos por las ecuaciones ( 2.26 y 2.27). Por otro lado, denotando  $w(x) \doteq b(e^{qx} + \gamma)$ ,  $x \in [0, \infty)$  podemos ver de las ecuaciones (2.29) y (2.28) y de el Lema 1 de [42] que la suposición (2.6b)) se satisface con  $\beta$  seleccionada como arriba. Para completar la prueba es suficiente estimar  $\delta_2$  en la ecuación (2.17). El cálculo directo de  $p(\cdot |x, a)$  y  $\tilde{p}(\cdot |x, a)$  muestra que para cada  $z \in [0, \infty)$  y  $a \in A$

$$\begin{aligned}
 I(z, a) &\doteq \int_x w(y) |p(z|x, a) - \tilde{p}(z|x, a)| = \\
 &w(0) |p(\xi > x+a) - \tilde{p}(\xi > x+a)| + \\
 &+ \int_0^{z+a} w(y) |\rho(z+a-y) - \tilde{\rho}(z+a-y)| dy \\
 &\leq b(1+\gamma) \int_{z+a}^{\infty} |\rho(s) - \tilde{\rho}(s)| ds + b \int_0^{z+a} [e^{(z+a)q} e^{-qs} + \gamma] |\rho(s) - \tilde{\rho}(s)| ds
 \end{aligned}
 \tag{2.31}$$

es fácil ver que

$$(1+\gamma) / (e^{qz} + \gamma) \leq 1 \tag{2.32}$$

y

$$(e^{(z+a)q} e^{-qs} + \gamma) / ((e^{qz} + \gamma)) \leq e^{q\theta}, \quad s > 0. \tag{2.33}$$

Así, las ecuaciones (2.31) y (2.33) nos dan la siguiente desigualdad:

$$[w(z)]^{-1} I(z, a) \leq e^{q\theta} \int_0^{\infty} |\rho(s) - \tilde{\rho}(s)| ds,$$

de donde obtenemos la desigualdad (2.30). ■

En conclusión, examinemos el comportamiento del lado derecho de la desigualdad (2.30) cuando el sistema de control se acerca a una frontera de inestabilidad para el servicio más lento  $a_t = \theta$ ,  $t = 0, 1, \dots$  esto es, cuando el  $\min(E\xi, E\tilde{\xi}) \downarrow \theta$ . Si además suponemos que las variables  $\xi$  y  $\tilde{\xi}$  tienen distribución exponencial con parámetro  $\lambda$  y  $\tilde{\lambda}$  respectivamente;  $\frac{1}{\tilde{\lambda}} > \frac{1}{\lambda}$ ; y el valor  $\theta$  se incrementa de tal manera que  $\varepsilon \doteq (\frac{1}{\lambda} - \theta) \downarrow 0$ . También suponemos que  $\sup_{a \in A} c(x, a) \leq dx + r$  para algunas constantes  $d, r$  con  $d > 0$ .

Por cálculos simples obtenemos  $e^{s\theta} E e^{-s\varepsilon} = e^{s\theta} \lambda / (s + \lambda)$  y  $q = q(\varepsilon) = O(\varepsilon)$ , cuando  $\varepsilon \downarrow 0$ . Por otro lado, para que se satisfaga la ecuación (2.29) para  $\varepsilon > 0$  suficientemente pequeña debemos tomar en (2.29)  $b = b(q) = \left(\frac{d}{q}\right) e^{\frac{q}{d}-1} = O\left(\frac{1}{q}\right)$  cuando  $q \rightarrow 0$ . Así el lado derecho de la ecuación (2.30) es  $O(\varepsilon^{-1}) G(\varepsilon)$  cuando  $\varepsilon \downarrow 0$ , donde

$$\lim_{\varepsilon \rightarrow 0} G(\varepsilon) = \frac{2\alpha}{(1-\beta)^2} (1+\alpha) \int_0^\infty |\rho(s) - \tilde{\rho}(s)| ds.$$

La divergencia de la cota de la ecuación (2.30) cuando  $\varepsilon \downarrow 0$  no se debe a un defecto de la desigualdad. Experimentos computacionales muestran un crecimiento ilimitado en el índice de estabilidad  $\Delta(x)$  en el lado izquierdo de (2.30) para este ejemplo cuando  $\varepsilon \downarrow 0$  (a condición de que  $c(x, a)$  disminuya suficientemente rápido en  $a$ ).

Una manera obvia de evitar esta clase de inestabilidad es reduciendo el conjunto de acciones admisibles  $A$  a  $\hat{A} = A \cap (0, \theta']$  con  $\theta - \theta' > 0$ . El problema a resolver es encontrar un valor conveniente de  $\theta - \theta' > 0$  que dependa de la exactitud de la aproximación entre  $\rho$  y  $\tilde{\rho}$ .

## 2.6 Ejemplo II

A pesar de que la desigualdad (2.16) da una cota superior para  $\Delta_\alpha(x)$  y asegura que  $\Delta_\alpha(x) \rightarrow 0$  cuando la distancia ponderada entre  $p$  y  $\tilde{p}$  tiende a cero, la exactitud de la desigualdad (2.16) depende mucho de la manera de como se escoge la función de prueba  $w$  y el número  $\beta$ , que satisfacen la suposición (2.1). El siguiente ejemplo muestra que si  $w$  y  $\beta$  se escogen inapropiadamente, la desigualdad (2.16) puede dar en el lado derecho un valor mucho mayor que el valor real de  $\Delta_\alpha(x)$ .

### Ejemplo 2.1 Modelo lineal con costo cuadrático

Considere el sistema LQ que aparece en [15], con dinámica

$$x_{n+1} = \gamma x_n + \delta a_n + \xi_n \tag{2.34}$$

y función de costo de un paso

$$c(x_n, a_n) = qx_n^2 + ra_n^2, \tag{2.35}$$

donde, tanto el espacio de estados  $X$  como el de acciones  $A$  es todo  $\mathbb{R}$ , el factor de descuento como siempre lo denotamos por  $\alpha$ , las  $\xi_n$  son variables aleatorias, y  $\gamma, \delta, q$  y  $r$  son números reales.

Para que se cumplan las condiciones de el Teorema 2.2 suponemos lo siguiente:

### Suposición 2.8

- a)  $\gamma = \delta \neq 0, q = r = 1/50$
- b) para cada  $x \in X$ , el conjunto de acciones admisibles es el intervalo

$$A(x) = [-x, x]$$

- c) los ruidos  $\xi_n$  forman una sucesión de variables aleatorias independientes e idénticamente distribuidas, independientes de  $x_0$ , con densidad continua y acotada  $g$ ; con  $E[\xi_n] = 0$  y  $E[\xi_n^2] < \infty$ .

De tal suerte que el proceso (2.34) queda de la siguiente manera:

$$x_{n+1} = \gamma x_n + \gamma a_n + \xi_n \quad n = 0, 1.. \quad (2.36)$$

Si consideramos el proceso aproximante

$$\tilde{x}_{n+1} = \tilde{\gamma} \tilde{x}_n + \tilde{\gamma} \tilde{a}_n + \xi_n \quad t = 0, 1.. \quad (2.37)$$

que difiere del proceso (2.36) solo por el valor de  $\tilde{\gamma}$  y para el cual se cumple la siguiente suposición:

### Suposición 2.9 a) $q = r = 1/50$ ;

- b) para cada  $x \in X$ , el conjunto de acciones admisibles es el intervalo

$$A(x) = [-x, x];$$

- c) los ruidos  $\tilde{\xi}_n$  forman una sucesión de variables aleatorias independientes e idénticamente distribuidas, independientes de  $x_0$ , con densidad continua y acotada  $g$ , con  $E[\tilde{\xi}_n] = 0$  y  $E[\tilde{\xi}_n^2] < \infty$ .

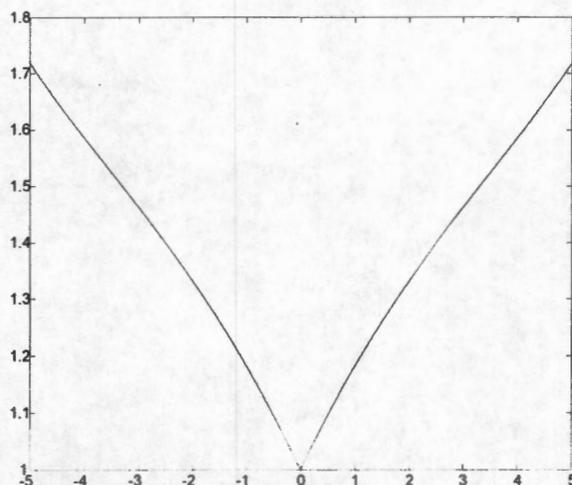


Figura 2.1: Gráfica de  $h(x) = e^{0.2|x|} - \frac{1}{2}(x^2)$

Daremos una estimación del índice de estabilidad para los procesos (2.36) y (2.37) que nos proporciona el Teorema 2.2.

En particular, si las sucesiones de variables aleatorias  $\{\xi_n\}$  y  $\{\tilde{\xi}_n\}$  tienen distribución uniforme en  $[-1, 1]$ , entonces estas suposiciones son suficientes para la hipótesis del Teorema 2.2. Los incisos a) y b) de las suposiciones 2.1 se cumplen trivialmente; los incisos c) y d) son consecuencia de que los  $\xi_n$  y  $\tilde{\xi}_n$  tienen función de densidad continua y acotada (vea [16]), si tomamos  $c(x, a) = [1/50](x^2 + a^2)$  y  $w(x) = e^{0.2|x|}$  vemos que:

a)  $1/50(x^2 + a^2) \leq e^{0.2|x|}$  para  $(x, a) \in \mathbb{K}$  ;

b)  $\int e^{0.2|y|} p(dy | x, a) \leq \frac{\beta}{\alpha} e^{0.2|x|}$   $\beta \in (0, 1)$ .

Para ver que se cumple a), observamos que la gráfica de  $h(x) = e^{0.2|x|} - \frac{1}{25}(x^2)$  en la figura 1 siempre está por encima del eje  $x$  ;

Para ver que se cumpla el inciso b) notamos que:

$$\begin{aligned} \int e^{0.2|y|} p(dy | x, a) &= \int_1^{-1} e^{0.2|\gamma x + \gamma a + s|} f(s) ds \\ &\leq e^{0.2|\gamma x + \gamma a|} \int_{-1}^1 \frac{1}{2} e^{0.2|s|} ds ; \text{ si tomamos } \gamma \leq .5 \\ &\leq e^{0.2|x|} \int_{-1}^1 \frac{1}{2} e^{0.2|s|} ds = e^{0.2|x|} 1.10701379 \end{aligned}$$

podemos tomar  $\beta = \alpha 1.10701379$ , siempre con tal que  $\alpha \leq \frac{1}{1.10701379} = .9$   
Ahora calculemos la cota. Como estamos considerando

$c(x, a) = \tilde{c}(x, a)$  la estimación de la cota queda:

$$\Delta(x) \leq \frac{2e^{0.2|x|}\alpha\delta_2}{(1-\beta)^2}. \quad (2.38)$$

Para calcular la parte derecha de (2.38) solo nos resta evaluar  $\delta_2$ .

Recordemos que  $\delta_2 = \sup_{(x,a)} \frac{1}{W(x)} \int_X W(y) |p(dy|x, a) - \tilde{p}(dy|x, a)|$  así  
 $\delta_2 = \sup_{(x,a)} \frac{1}{e^{0.2|x|}} \int_X e^{0.2|y|} |f(y - \gamma(x+a)) - f(y - \tilde{\gamma}(x+a))| dy$  donde

$$f(x) = \begin{cases} \frac{1}{2} & \text{si } t \in [-1, 1] \\ 0 & \text{si } t \notin [-1, 1] \end{cases}.$$

Como  $\sup_{(-1,1)} e^{0.2|x|} = e^{0.2}$ , entonces

$$\begin{aligned} \int_X e^{0.2|y|} |f(y - \gamma(x+a)) - f(y - \tilde{\gamma}(x+a))| dy &\leq \\ e^{0.2} \int_{-1}^1 |f(y - \gamma(x+a)) - f(y - \tilde{\gamma}(x+a))| dy &= \\ \frac{e^{0.2}}{2} \left\{ \begin{array}{l} |-1 - \gamma(x+a) - (-1 - \tilde{\gamma}(x+a))| + \\ |1 - \gamma(x+a) - (1 - \tilde{\gamma}(x+a))| \end{array} \right\} &= \\ \frac{e^{0.2}}{2} 2|\gamma - \tilde{\gamma}| |x+a| &\leq e^{0.2} |\gamma - \tilde{\gamma}| |x| \end{aligned}$$

entonces  $\delta_2 \leq \sup_{(x,a) \in \mathbb{K}} e^{0.2} |\gamma - \tilde{\gamma}| \frac{|x|}{e^{0.2|x|}} \approx (2.5) |\gamma - \tilde{\gamma}|$ , esto se debe a que  $\sup_{x \in \mathbb{R}} e^{0.2} \frac{|x|}{e^{0.2|x|}} \leq 2.5$ ; ahora si tomamos  $\alpha = .5$ , tendremos  $\beta \approx .55$ , entonces, en la parte derecha de la desigualdad (2.38)  $\frac{\alpha}{(1-\beta)^2}$  nos queda aproximadamente  $\frac{.5}{(1-.55)^2} = 2.4691$  de donde obtenemos

$$\Delta(x) \leq 12e^{0.2|x|} |\gamma - \tilde{\gamma}|. \quad (2.39)$$

Lo cual nos da cotas no muy buenas ya que aunque  $x \in [-1, 1]$ , por ejemplo  $x = 1$  y  $|\gamma - \tilde{\gamma}| \leq .01$  la parte derecha de la desigualdad (2.39) nos da  $12 * \exp(.2) * .01 = .14657$ . Por otro lado, en [15] se obtienen expresiones exactas para las políticas óptimas  $f^*$  y  $\tilde{f}^*$  de (2.36) y (2.37), aplicando estas políticas podemos obtener  $V(x, f^*(x))$  y  $V(x, \tilde{f}^*(x))$ , lo cual nos permite obtener el valor exacto de  $\Delta(x)$  que es .008.

## Capítulo 3

# El índice de estabilidad relativo

### 3.1 Introducción

Algunas veces es más conveniente considerar el *índice de estabilidad relativo*

$$\bar{\Delta}_\alpha(x) = \frac{V_\alpha(\tilde{\pi}_\alpha^*, x) - V_\alpha(\pi_\alpha^*, x)}{V_\alpha(\pi_\alpha^*, x)}.$$

La tasa de decrecimiento de  $\bar{\Delta}_\alpha(x)$  cuando  $\tilde{p}$  converge a  $p$  depende fuertemente de las propiedades del proceso en consideración, y más aún, de la manera en que  $\tilde{p}$  converge a  $p$ . En general el *índice de estabilidad relativo* puede no tender a cero cuando  $\tilde{p} \rightarrow p$ . En el ejemplo de la última sección de este capítulo  $\bar{\Delta}_\alpha(x)$  no va a cero y el  $\sup_{\alpha \in (0,1)} \bar{\Delta}_\alpha(x) = \infty$  a pesar de que  $\tilde{p}$  converge a  $p$  en la topología débil.

Como hemos visto en la introducción y en el capítulo II las cotas obtenidas para  $\Delta_\alpha(x)$  son de la forma

$$\Delta_\alpha(x) \leq M(\alpha) \psi(p, \tilde{p}) \quad (3.1)$$

donde  $M(\alpha)$  es del orden  $(1 - \alpha)^{-2}$ .

Es bien sabido que bajo condiciones generales se cumple la siguiente relación

$$0 < \limsup_{\alpha \rightarrow 1} (1 - \alpha) V_\alpha^*(\pi_\alpha) < \infty$$

(vea por ejemplo, [16]). Así, teniendo en mente encontrar cotas para  $\bar{\Delta}_\alpha(x)$  uniformes en  $\alpha$ , necesitamos algo como (3.1) con  $M(\alpha) = O((1 - \alpha)^{-1})$  cuando

$\alpha \rightarrow 1$ . Para conseguir esto, se impondrán *condiciones de ergodicidad*, en los procesos considerados. Suponiendo que las funciones de costo de ambos procesos son iguales ( $c = \tilde{c}$ ); obtendremos la siguiente desigualdad:

$$\sup_{\alpha \in (0,1)} \bar{\Delta}_\alpha(x) \leq K[\psi_\sigma(p, \tilde{p})]^r \max\{1, \log_\rho \psi_\sigma(p, \tilde{p})\}, \quad (3.2)$$

donde las constantes  $K < \infty$ ,  $\rho \in (0,1)$ ,  $r \in (0,1)$  son calculadas explícitamente, y

$$\psi_\sigma(p, \tilde{p}) \equiv \delta \triangleq \sup_{x \in X, a \in A(x)} \sigma(p(\cdot|x, a), \tilde{p}(\cdot|x, a))$$

donde  $A(x)$  es el conjunto de controles admisibles en el estado  $x$ , mientras  $\sigma$  es la métrica de *variación total* (vea, 3.4 para la definición). En algunos casos (vea un ejemplo en la última sección de este capítulo) la potencia de  $r$  en (3.2) puede ser elegida cercana a 1.

Las condiciones de ergodicidad son necesarias en la prueba de (3.2), debido a que la evaluación de  $\bar{\Delta}_\alpha(x)$  para cuando  $\alpha \rightarrow 1$  esta relacionada con la estabilidad respecto al criterio de costo promedio esperado asintótico.

$$\mathcal{J}(\pi, x) \triangleq \limsup_{n \rightarrow \infty} n^{-1} E_x^\pi \sum_{t=0}^{n-1} c(x_t, a_t). \quad (3.3)$$

En el capítulo anterior usamos las ecuaciones de optimalidad y técnicas de operadores contractibles y obtuvimos  $M(\alpha) = O((1-\alpha)^{-2})$  en (3.1), como se hizo en [11]. Para obtener una desigualdad como (3.1) con  $M(\alpha) = o((1-\alpha)^{-1})$ , en lugar de eso reducimos el problema al caso en que el costo de un paso es acotado, y después aplicamos el método de estimación de la estabilidad de Zolotarev, en procesos no controlados [46]. Para llevar a cabo esto, necesitamos usar resultados de Kartashov [20] sobre ergodicidad para obtener tasas de convergencia geométrica de los procesos a la distribución límite.

## 3.2 Suposiciones

Consideraremos un par de *modelos de control de Markov* a tiempo discreto ( $t = 0, 1, 2, \dots$ )  $\mathcal{P} = (X, A, A(x), c, p)$  y  $\tilde{\mathcal{P}} = (X, A, A(x), c, \tilde{p})$ . Los únicos componentes diferente en estos modelos son las *probabilidades de transición*  $p(B|x, a)$  y  $\tilde{p}(B|x, a)$ ,  $(x, a) \in \mathbb{K}$ ,  $B \in \mathbb{B}(X)$ .

Igual que en el capítulo anterior usaremos como criterio de desempeño para los procesos  $\mathcal{P}$  y  $\tilde{\mathcal{P}}$  la esperanza del costo total  $\alpha$ -descontado definidas por la ecuaciones (1.6) y (1.7).

Fijemos una función medible  $W : X \rightarrow [1, \infty)$  e introduzcamos las hipótesis necesarias para que se cumplan nuestros resultados.

**Suposición 3.1** (Condiciones de continuidad y acotamiento) se cumplen las condiciones 2.2 para los kernels  $p$  y  $\tilde{p}$  y además:

$\inf_{(x,a) \in \mathbb{K}} c(x, a) \geq \kappa > 0$  y  $\sup_{a \in A(x)} c(x, a) \leq [W(x)]^{1/s}$ ,  $x \in X$  donde  $\kappa$  y  $s > 1$  son constantes dadas;

**Suposición 3.2** (Condición de recurrencia)

Para cada política  $\mathbf{f} \in \Pi_{de}$  las cadenas de Markov con probabilidades de transición  $p(\cdot|x, f(x))$  y  $\tilde{p}(\cdot|x, f(x))$  son Harris-recurrente.

**Suposición 3.3** (Condiciones de ergodicidad) Existe una medida de probabilidad  $\nu$  en  $(X, \mathcal{B}_X)$  y un número  $\beta \in [0, 1)$  para los cuales se cumple lo siguiente:

Para cada política  $\mathbf{f} \in \Pi_{de}$  existe una función no-negativa medible  $h_f \leq 1$  en  $X$  tal que para todo  $x \in X$  y  $B \in \mathcal{B}_X$ :

a)  $p(B|x, f(x)) \geq h_f(x)\nu(B)$ ;

b)  $\int_X W(y)p(dy|x, f(x)) \leq$

$$\beta W(x) + h_f(x) \int_X W(y)\nu(dy) < \infty;$$

c)  $\inf_{f \in \mathbb{F}} \int_X h_f(x)\nu(dx) =: \gamma > 0$ .

**Suposición 3.4** La suposiciones (3.3) se cumple para la probabilidad de transición  $\tilde{p}$  con las mismas  $W, \nu, \beta, \gamma$  y algún  $\tilde{h}_f$  ( $f \in \mathbb{F}$ ).

### 3.3 Comentarios sobre las suposiciones

- (1) Al final del capítulo daremos un ejemplo de un sistema de colas  $GI|GI|1|\infty$  con tasas de servicio controladas. Para este ejemplo todas las suposiciones anteriores se cumplen .
- (2) Las Suposiciones (3.2) a (3.4) son las extensiones (dadas en [9]) de las hipótesis usadas en [20] para estudiar la ergodicidad y la estabilidad de procesos de Markov (no controlados)
- (3) Las Suposiciones (3.3 (b) y (c)) se pueden comprobar para el proceso  $\tilde{\mathcal{P}}$  con algún  $\tilde{\gamma} > 0$ ,  $\tilde{\beta} < 1$ ; entonces para que se cumpla la Suposición (3.4) podemos tomar  $\max(\beta, \tilde{\beta})$  y  $\min(\gamma, \tilde{\gamma})$ .
- (4) En vista de la Proposición (3.1) que se establece más abajo, las Suposiciones (3.1), (3.2), (3.3) y (3.4) aseguran la existencia de políticas estacionarias óptimas  $\mathbf{f}_\alpha, \tilde{\mathbf{f}}_\alpha$  para los procesos  $\mathcal{P}$  y  $\tilde{\mathcal{P}}$ . Estas políticas se usan para definir el índice de estabilidad relativo en (1.12). Por otro lado, las suposiciones (3.2), (3.3) y (3.4) que como se verá garantizan la ergodicidad del proceso, pueden no cumplirse para algunas políticas estacionarias (aún para muchas). (vea, por ejemplo [21]). Un análisis de la prueba del Teorema de la próxima sección muestra que se pueden debilitar las suposiciones (3.2), (3.3) y (3.4) si pedimos la existencia  $\mathbf{f}_\alpha$  y  $\tilde{\mathbf{f}}_\alpha$  en algún subconjunto  $S_0 \subset \Pi_{de}$  de políticas estacionarias y las Suposiciones (3.2), (3.3) y (3.4) sean satisfechas para cada  $\mathbf{f} \in S_0$  (pero probablemente no para cada  $\mathbf{f} \in \Pi_{de}$ ). La cota (3.7) continúa siendo cierta bajo dichas modificaciones de las hipótesis y, más aún, en esta situación no necesitamos la Suposición 3.1 (b), (c) y la suposición acerca de la compacidad de los conjuntos  $A(x)$  ( $x \in X$ ). Y no parece difícil modificar un poco la definición de  $\bar{\Delta}_\alpha(x)$ , la desigualdad (3.7) y su demostración para que sea suficiente solo suponer la existencia de políticas  $\epsilon$ -óptimas  $\mathbf{f}_{\epsilon,\alpha}, \tilde{\mathbf{f}}_{\epsilon,\alpha} \in S_0$  ( $\epsilon > 0$ ).

**Proposición 3.1** (a) *Considere que las Suposiciones (3.1), (3.2), (3.3) y (3.4) se cumplen y sea  $\alpha \in (0,1)$  un factor de descuento arbitrario, pero fijo. Entonces existen políticas estacionarias óptimas  $\mathbf{f}_\alpha, \tilde{\mathbf{f}}_\alpha \in \Pi_{de}$ , para los procesos  $\mathcal{P}$  y  $\tilde{\mathcal{P}}$  respectivamente. (véase [9] para la demostración).*

(b) El costo óptimo promedio asintótico  $J_*(x) = \inf_{\pi \in \Pi} J(x, \pi)$  (véase 3.3) no depende de  $x \in X$ .

**Nota 7** Usaremos la notación  $J_*$  para  $J_*(x)$ .

### 3.4 Desigualdades de Estabilidad

A lo largo de lo que resta de este capítulo fijamos una  $x \in X$  arbitraria, como el estado inicial de ambos procesos  $\mathcal{P}$  y  $\tilde{\mathcal{P}}$ , i.e.  $x_0 = x, \tilde{x}_0 = x$ . Debido a la Proposición 3.1 podemos reescribir el índice de estabilidad relativo en (1.12) como

$$\bar{\Delta}_\alpha(x) = [V_\alpha(x, \tilde{f}_\alpha) - V_\alpha(x, f_\alpha)]/V_\alpha^*(x).$$

Debido a la suposición (3.1 (a)) esta cantidad está bien definida.

La distancia de la variación total (ampliamente conocida) se define como:

$$\sigma(\xi, \zeta) \equiv \sigma(\mu_\xi, \mu_\zeta) := 2 \sup\{|P(\xi \in B) - P(\zeta \in B)| : B \text{ are Borel}\}. \quad (3.4)$$

Aquí,  $\mu_\xi$  y  $\mu_\zeta$  son las distribuciones de elementos aleatorios  $\xi$  y  $\zeta$  que toman valores en algún espacio de Borel.

Para escribir una cota para  $\bar{\Delta}_\alpha(x)$  introduciremos ciertas constantes que aparecen en la estimación de la tasa de convergencia en el Lema Ergodico 3.4 en ([9]). La apariencia de estas constantes es un poco complicada, pero es importante que éstas sean calculadas de manera precisa en términos de las cantidades involucradas en las suposiciones (3.1), (3.2), (3.3) y (3.4). También, todas las desigualdades dadas para las constantes se siguen de la prueba del Lema 3.4 en [9].

Fije cualquier número positivo  $\bar{\gamma}$  tal que  $\bar{\gamma} \leq \gamma \leq 1$  ( $\gamma$  viene de la Suposición 3.3 (c)) y, tomando como convención:  $(\log 1)/(1-1) := -1$ , definimos:  $\omega = 2 \exp\{[1 - \|\nu\|_W/(1-\beta)](\log \bar{\gamma})/(1-\bar{\gamma})\} - 1$ .

Definimos  $\|\nu\|_W := \int_X W(y)\nu(dy)$ , y escogemos:

$$\begin{aligned} d &= 1 - (1 - \beta^2)/[(1 - \beta) + \beta\omega\|\nu\|_W] < 1; \\ \rho &:= d + q < 1, \end{aligned} \quad (3.5)$$

donde  $q$  es número positivo arbitrario tal que  $q < 1 - d$ . Ahora sea  $t_* = [d/(1-d)]$  ( $[\cdot]$  es la parte entera),  $\tau = \max\{0, 1/\log(1+q/d) - 2\}$ , y

$$\begin{aligned} b_1 &= d^\tau(\tau + 2)/(d + q)^\tau & \text{si } \tau > t_*, \\ b_1 &= d^{t_*}(t_* + 2)/(d + q)^{t_*} & \text{si } \tau \leq t_*. \end{aligned}$$

Finalmente, definimos:

$$\bar{B} = \max \left\{ [1 + b_1 de/\beta][1 + \|\nu\|_W/(1 - \beta)], [\max \{1, (\|\nu\|_W + \beta/\rho)^{t^*} + \|\nu\|_W/[(1 - \beta)\rho^{t^*}]\} \right\}. \quad (3.6)$$

El resultado principal de este capítulo que fue obtenido en [12] es el siguiente teorema.

**Teorema 3.1** *Consideremos ciertas las Suposiciones (3.1), (3.2), (3.3) y (3.4) para dos (PCM)  $\mathcal{P}$  y  $\tilde{\mathcal{P}}$ . Entonces*

$$\bar{\Delta}_\alpha(x) \leq M(\alpha, x)\delta^{(s-1)/s} \max\{1, [\log_\rho \delta]\}, \quad (3.7)$$

donde:

$$\delta = \sup_{(x,a) \in \mathbb{K}} \sigma(p(\cdot|x, a), \tilde{p}(\cdot|x, a)), \quad (3.8)$$

$$M(\alpha, x) = B(x)/[(1 - \alpha)V_\alpha^*(x)] \leq B(x)/\kappa, \quad (3.9)$$

$$\lim_{\alpha \rightarrow 1} (1 - \alpha)V_\alpha^*(x) = \mathcal{J}_* \quad (3.10)$$

y, finalmente,

$$B(x) = 2\{2W(x)[1 + 2\bar{B}\rho^{-1}] + 2(1 - \beta)^{-1}\|\nu\|_W + 1\}.$$

**Corolario 3.1**

$$|V_\alpha^*(x) - \tilde{V}_\alpha^*(x)| \leq [2(1 - \alpha)]^{-1} B(x)\delta^{(s-1)/s} \max\{1, [\log_\rho \delta]\}.$$

**Corolario 3.2** *Sean los modelos  $\mathcal{P}$  y  $\tilde{\mathcal{P}}$  dados por las ecuaciones recurrentes:*

$$x_{t+1} = F(x_t, a_t, \xi_t), \quad t = 0, 1, \dots \quad (3.11)$$

$$\tilde{x}_{t+1} = F(\tilde{x}_t, a_t, \tilde{\xi}_t), \quad t = 0, 1, \dots \quad (3.12)$$

donde  $\{\xi_t\}$ ,  $\{\tilde{\xi}_t\}$  son sucesiones de elementos i.i.d. en algún espacio de Borel  $(Y, B_Y)$  con distribución común  $\mu_\xi$  y  $\mu_{\tilde{\xi}}$  respectivamente, entonces bajo la hipótesis del Teorema 3.1

$$\bar{\Delta}_\alpha(x) \leq M(\alpha, x) [\sigma(\mu_\xi, \mu_{\tilde{\xi}})]^{(s-1)/s}, \quad (3.13)$$

con la condición de que  $\sigma(\mu_\xi, \mu_{\tilde{\xi}}) \leq e^{-s/(s-1)}$ .

**Nota 8** Cuando la segunda desigualdad en la Suposición (3.1 (a)) se cumple para valores grandes de  $s$ , la potencia de  $\delta$  en (3.7) está cercana al mejor valor posible que es 1. En el ejemplo de la próxima sección  $X = [0, \infty)$  y  $W(x) = \bar{b}e^{hx}$ ,  $h > 0$  y  $\bar{b} > 0$  es arbitraria. Así, si  $\sup_{a \in A(x)} c(x, a)$  está acotado por algún polinomio entonces, para cada  $\epsilon > 0$  podemos escoger  $\bar{b} = \bar{b}(\epsilon)$  de tal manera que (3.7) se cumpla con  $\delta^{1-\epsilon}$ .

**Dem.**

En vista de la proposición (3.1) obtenemos (vea (1.9, 1.10))

$$\begin{aligned} & V_\alpha(x, \tilde{\mathbf{f}}_\alpha) - V_\alpha(x, \mathbf{f}_\alpha) \leq \\ & |V_\alpha(x, \tilde{\mathbf{f}}_\alpha) - \tilde{V}_\alpha(x, \tilde{\mathbf{f}}_\alpha)| + |\inf_{\mathbf{f} \in \Pi_{de}} \tilde{V}_\alpha(x, \mathbf{f}) - \inf_{\mathbf{f} \in \Pi_{de}} V_\alpha(x, \mathbf{f})| \leq \\ & 2 \sup_{\mathbf{f} \in \Pi_{de}} |\tilde{V}_\alpha(x, \mathbf{f}) - V_\alpha(x, \mathbf{f})| \leq \\ & 2 \sup_{\mathbf{f} \in \Pi_{de}} \sum_{t=0}^{\infty} \alpha^t |E_x^{\mathbf{f}} c(x_t, f(x_t)) - E_x^{\tilde{\mathbf{f}}} c(\tilde{x}_t, f(\tilde{x}_t))| \leq \\ & \frac{2}{1-\alpha} \sup_{\mathbf{f} \in \Pi_{de}} \sup_{t \geq 1} \epsilon_t(\mathbf{f}), \end{aligned} \quad (3.14)$$

donde

$$\epsilon_t(\mathbf{f}) := |E_x^{\mathbf{f}} c(x_t, f(x_t)) - E_x^{\tilde{\mathbf{f}}} c(\tilde{x}_t, f(\tilde{x}_t))|, \quad (3.15)$$

y  $x_t, \tilde{x}_t, t = 0, 1, \dots$  denotan las cadenas de los procesos  $\mathcal{P}$  y  $\tilde{\mathcal{P}}$  obtenidos por la aplicación de la política  $\mathbf{f} \in \Pi_{de}$ .

Así, para probar (3.7) es suficiente mostrar que para todo  $f \in \Pi_{de}$  y  $t$

$$\epsilon_t(f) \leq (B(x)/2)\delta^{(s-1)/s} \max\{1, [\log_\rho \delta]\}. \quad (3.16)$$

Para reducir la última desigualdad a estimaciones similares a aquellas usadas en (3.15) pero con funciones de costos acotadas, primero aseguramos que para cada política  $f \in \mathbb{F}$

$$\sup_{t \geq 0} E_x^f W(x_t) \leq C(x), \quad (3.17)$$

$$\sup_{t \geq 0} E_x^f W(\tilde{x}_t) \leq C(x) \quad (3.18)$$

donde  $C(x) = W(x) + \frac{\|\nu\|_W}{1-\beta}$ .

En efecto de la suposición 3.3 (a) y (b) tenemos que

$$\int_X W(y)p(dy|x_t, f(x_t)) \leq \beta W(x_t) + \|\nu\|_W$$

y por la propiedad de Markov de  $\{x_t\}$  obtenemos para cualquier historia fija  $h_{t-1}$  ( $t \geq 1$ ):

$$E_x^f[W(x_t)|h_{t-1}] = \int_X W(y)p(dy|x_{t-1}, f(x_{t-1})) \leq \beta W(x_{t-1}) + \|\nu\|_W$$

o bien,

$$E_x^f W(x_t) \leq \beta E_x^f W(x_{t-1}) + \|\nu\|_W.$$

Iterando la última desigualdad obtenemos

$$\begin{aligned} E_x^f W(x_t) &\leq \beta^t W(x) + \|\nu\|_W(1 + \beta + \dots + \beta^{t-1}) \\ &\leq W(x) + \|\nu\|_W/(1 - \beta) \text{ que implica (3.17).} \end{aligned}$$

La prueba de (3.18) es la misma.

Ahora, para  $b > 0$  arbitrario, pero fijo, definimos

$$c_b(x, a) := \begin{cases} c(x, a) & \text{si } c(x, a) \leq b \\ 0 & \text{en cualquier otro caso} \end{cases} \quad (3.19)$$

Aplicando la suposición 3.1 (a), las desigualdades de Hölder y Chebyshev obtenemos que para toda  $t \geq 0$ ,  $\mathbf{f} \in \Pi_{de}$  (abajo:  $1/s + 1/\ell = 1$ ):

$$\begin{aligned}
 & |E_x^{\mathbf{f}}c(x_t, f(x_t)) - E_x^{\mathbf{f}}c_b(x_t, f(x_t))| = \\
 & E_x^{\mathbf{f}}\{c(x_t, f(x_t)); c(x_t, f(x_t)) > b\} \leq \\
 & E_x^{\mathbf{f}}[W^{1/s}(x_t)I_{\{W(x_t) > b^s\}}] \leq \\
 & \{E_x^{\mathbf{f}}W(x_t)\}^{1/s} \{P(W(x_t) > b^s)\}^{1/\ell} \leq \\
 & \{E_x^{\mathbf{f}}W(x_t)\}^{1/s} \{E_x^{\mathbf{f}}W(x_t)\}^{1/\ell} b^{-s/\ell} \leq \mathbb{C}(x) b^{1-s}.
 \end{aligned} \tag{3.20}$$

De manera similar obtenemos:

$$|E_x^{\mathbf{f}}c(\tilde{x}_t, f(\tilde{x}_t)) - E_x^{\mathbf{f}}c_b(\tilde{x}_t, f(\tilde{x}_t))| \leq \mathbb{C}(x) b^{1-s},$$

$f \in \mathbb{F}$ ,  $t = 0, 1, \dots$

Por consiguiente (vea (3.15)).

$$\begin{aligned}
 \epsilon_t(\mathbf{f}) & \leq |E_x^{\mathbf{f}}c(x_t, f(x_t)) - E_x^{\mathbf{f}}c_b(x_t, f(x_t))| + \\
 & |E_x^{\mathbf{f}}c_b(x_t, f(x_t)) - E_x^{\mathbf{f}}c_b(\tilde{x}_t, f(\tilde{x}_t))| + \\
 & |E_x^{\mathbf{f}}c_b(\tilde{x}_t, f(\tilde{x}_t)) - E_x^{\mathbf{f}}c(\tilde{x}_t)| \leq \\
 & 2\mathbb{C}(x) b^{1-s} + \epsilon_{b,t}(\mathbf{f}),
 \end{aligned}$$

donde  $\epsilon_{b,t}(\mathbf{f})$  representa el penúltimo término de la primera desigualdad en (3.20).

Tomemos una política arbitraria  $\mathbf{f} \in \Pi_{de}$ , y sean  $q_f, \tilde{q}_f$  las medidas invariantes para los procesos de Markov con probabilidades de transición,  $p(\cdot|x, f(x))$  y  $\tilde{p}(\cdot|x, f(x))$  respectivamente. La suposición (3.2) garantiza la existencia y unicidad de  $q_f$  y  $\tilde{q}_f$  (vea, por ejemplo [29]). Para aplicar el enfoque de Zolotarev (vea [46]) en la estimación  $\sup_{\mathbf{f} \in \Pi_{de}} \epsilon_{b,t}(\mathbf{f})$  necesitamos que las tasas de convergencia de  $x_t$  y  $\tilde{x}_t$  hacia las probabilidades invariantes sea uniforme sobre  $\mathbf{f} \in \Pi_{de}$ . Como  $W \geq 1$  podemos usar el lema 3.4 en [9] para obtener las siguientes desigualdades:

$$\sup_{f \in \mathbb{F}} \sigma(x_t, x_\infty^{(f)}) \leq \bar{B}W(x)\rho^t \tag{3.21}$$

$$\sup_{f \in \mathbb{F}} \sigma(\tilde{x}_t, \tilde{x}_\infty^{(f)}) \leq \bar{B}W(x)\rho^t, \quad t = 1, 2, \dots \quad (3.22)$$

donde los elementos aleatorios  $x_\infty^{(f)}$  y  $\tilde{x}_\infty^{(f)}$  (con valores en  $X$ ) tienen, distribuciones  $q_f, \tilde{q}_f$ , respectivamente, y las constantes  $\bar{B} < \infty$  y  $\rho < 1$  fueron introducidas en (3.5), (3.6).

Sea  $n \geq 1$  un entero (fijo).

Entonces

$$\epsilon_{b,t}(\mathbf{f}) \leq b\sigma(x_t, \tilde{x}_t) \leq b \max_{t \leq n} \sigma(x_t, \tilde{x}_t) \quad \text{si } t \leq n \quad (3.23)$$

y para  $t > n$  por (3.21), (3.22) siguiendo [46] obtenemos

$$\begin{aligned} \epsilon_{b,t}(\mathbf{f}) &\leq b[\sigma(x_t, x_\infty^{(f)}) + \sigma(x_\infty^{(f)}, \tilde{x}_\infty^{(f)}) + \sigma(\tilde{x}_t, \tilde{x}_\infty^{(f)})] \leq \\ &b[2\bar{B}W(x)\rho^n + \sigma(x_\infty^{(f)}, \tilde{x}_\infty^{(f)})]. \end{aligned} \quad (3.24)$$

Por otro lado,

$$\begin{aligned} \sigma(x_\infty^{(f)}, \tilde{x}_\infty^{(f)}) &\leq \sigma(x_\infty^{(f)}, x_n) + \sigma(x_n, \tilde{x}_n) + \sigma(\tilde{x}_n, \tilde{x}_\infty^{(f)}) \leq \\ &2\bar{B}W(x)\rho^n + \max_{t \leq n} \sigma(x_t, \tilde{x}_t). \end{aligned} \quad (3.25)$$

Combinando (3.23)-(3.25) vemos que

$$\epsilon_{b,t}(f) \leq b[4\bar{B}W(x)\rho^n + \max_{t \leq n} \sigma(x_t, \tilde{x}_t)], \quad t = 1, 2, \dots \quad (3.26)$$

Explotando la siguiente representación dual

$$\sigma(\mu_\xi, \mu_\zeta) = \sup_{\varphi: \|\varphi\|_\infty \leq 1} \left| \int_X \varphi(x) d\mu_\xi - \int_X \varphi(x) d\mu_\zeta \right|$$

de métrica de la variación total en (3.4), usando la propiedad de Markov de  $x_t, \tilde{x}_t$ , la suposición  $x_0 = \tilde{x}_0$  e inducción podemos mostrar que para cualquier  $f \in \mathbb{F}$

$$\max_{t \leq n} \sigma(x_t, \tilde{x}_t) \leq n \sup_{x \in X} \sigma(p(\cdot|x, f(x)), \tilde{p}(\cdot|x, f(x))) \leq n\delta, \quad (3.27)$$

(vea [39]).

De (3.20), (3.26) y (3.27) obtenemos que para cualesquiera  $n \geq 1$ ,  $b > 0$

$$\epsilon_t(\mathbf{f}) \leq 2C(x)b^{1-s} + b(4\bar{B}W(x)\rho^n + n\delta) \quad (3.28)$$

donde la parte derecha es independiente de la política  $\mathbf{f}$ . Escoja en (3.28)  $n = \max\{1, [\log_\rho \delta]\}$ ,  $b = \delta^{-1/s}$ ; aquí  $[z]$  significa el mayor entero menor o igual que  $z$ . Entonces, por cálculos elementales obtenemos

$$\begin{aligned} \epsilon_t(\mathbf{f}) &\leq 2C(x)\delta^{(s-1)/s} + \delta^{-1/s}(4\bar{B}W(x)\rho^{-1}\delta + \max\{1, \log_\rho \delta\}\delta) \leq \\ &\{2C(x) + 4\bar{B}W(x)\rho^{-1} + 1\}\delta^{(s-1)/s} \max\{1, \log_\rho \delta\}. \end{aligned} \quad (3.29)$$

La última desigualdad implica (3.16), por lo tanto en vista de (3.14) obtenemos la desigualdad requerida (3.7).

La desigualdad (3.9) es una consecuencia de la definición (1.9) y la suposición (3.1 (a)).

Finalmente, la existencia y el valor de límite en (3.10) puede ser establecida haciendo pequeños cambios en la prueba de el teorema 4.2 en [16]; (vea también la prueba del teorema 2.6 en [9]).

La desigualdad en el Corolario (3.1) se sigue de (3.14) y (3.29).

Para obtener la desigualdad (3.13) en el corolario 3.2 es suficiente observar que la función  $z^{(s-1)/s} \max\{1, [\log_\rho z]\}$  es creciente para  $z \leq e^{-s/(s-1)}$  junto con la siguiente desigualdad para los procesos (3.11) y (3.12):

$$\sigma(p(\cdot|x, a), \tilde{p}(\cdot|x, a)) = 2 \sup_{B \in \mathcal{B}_X} |P(F(x, a, \xi_0) \in B) - P(F(x, a, \tilde{\xi}_0) \in B)| =$$

$$2 \sup_{B \in \mathcal{B}_X} |P(\xi_0 \in G_{x,a}^{-1}(B)) - P(\tilde{\xi}_0 \in G_{x,a}^{-1}(B))| \leq \sigma(\xi_0, \tilde{\xi}_0),$$

$$\text{donde } G_{x,a}(\cdot) \doteq F(x, a, \cdot).$$

### 3.5 Ejemplos

Empezaremos con un ejemplo de un modelo *inestable* de optimización de costo descontado. Este ejemplo ilustra también que la convergencia de las funciones de valor  $\tilde{V}_\alpha \rightarrow V_\alpha$  (como en el Corolario 1) no puede ser visto como un problema de estabilidad si, estamos interesados en la magnitud de  $\Delta_\alpha$ .

Considere los PCM dados por las ecuaciones:

$$x_{t+1} = x_t a_t \xi_t, y \quad (3.30)$$

$$\tilde{x}_{t+1} = \epsilon + \tilde{x}_t \alpha_t \tilde{\xi}_t; \quad t = 0, 1, \dots \quad (3.31)$$

en los espacios  $X = [0, \infty)$   $A(x) = A := \{0, 1\}$ ,  $x \in X$  con los mismos estados iniciales  $x_0 = \tilde{x}_0 = x = 1$ . En (3.30) y (3.31),  $\epsilon$  es algún número en  $(0, 1)$ ;  $\{\xi_t\}$ ,  $\{\tilde{\xi}_t\}$  son sucesiones de variables aleatorias i.i.d. (v.a. por brevedad) y donde  $\xi_0$  tiene la distribución uniforme en  $[1/2, 1]$ , y

$$P(\tilde{\xi}_0 \in (y, y + dy)) = 2(1 - \epsilon)dy, \quad y \in [1/2, 1],$$

$P(\tilde{\xi}_0 = b_\epsilon) = \epsilon$  donde  $b_\epsilon \geq 1$  que será escogido más adelante.

Sea  $\alpha \in (0, 1)$  un factor de descuento y definamos la función de costo como sigue:

$$c(x, 0) = \begin{cases} 0 & \text{si } x = 0; \\ 2 - x & \text{si } x \in (0, 1]; \\ 1 & \text{de otra forma;} \end{cases} \quad (3.32)$$

$$c(x, 1) = 2 - x, \quad x \in [0, \infty). \quad (3.33)$$

Para cada política  $\pi$  es claro que  $x_t \in [0, 1]$ ,  $t = 0, 1, \dots$ , por lo tanto la política estacionaria  $f_*(x) = 0$ ,  $x \in X$  es óptima para el proceso (3.30), además  $V_\alpha^*(1) = V_\alpha(1, f_*) = 1$ . Por otro lado, si aplicamos cualquier política  $\pi$  al proceso (3.31) obtenemos

$$\tilde{x}_t \in [\epsilon, \infty), \quad \tilde{x}_t^* \geq \tilde{x}_t, \quad t = 0, 1, \dots, \quad (3.34)$$

donde  $\{\tilde{x}_t^*\}$  es la cadena que obtenemos al usar la política estacionaria  $\tilde{f}_*(x) = 1$ ,  $x \in X$ . De (3.32)-(3.34) tenemos que la política  $\tilde{f}_*$  es óptima para el proceso (3.31) y, tomando  $b_\epsilon$  de tal manera que:  $\mu = \mu(b_\epsilon) := E\tilde{\xi}_0 \in (1, 1/\alpha)$  obtenemos

$$\tilde{V}_\alpha^*(1) = \tilde{V}_\alpha(1, \tilde{f}_*) = \sum_{t=0}^{\infty} \alpha^t E_1^{f_*} C(\tilde{x}_t^*, 1) = \quad (3.35)$$

$$\sum_{t=0}^{\infty} \alpha^t E(2 - \tilde{x}_t^*) = \frac{1}{1-\alpha} \left(2 - \frac{\epsilon}{1-\mu}\right) + \frac{1}{1-\alpha\mu} \left(\frac{\epsilon}{1-\mu} - 1\right).$$

La última igualdad en (3.35) se debe a que

$$\tilde{x}_i^* = \tilde{\xi}_0 \tilde{\xi}_2 \dots \tilde{\xi}_{t-1} + \epsilon(\tilde{\xi}_1 \dots \tilde{\xi}_{t-1} + \tilde{\xi}_2 \dots \tilde{\xi}_{t-1} + \dots + \tilde{\xi}_{t-1} + 1),$$

por lo tanto,

$$E(2 - \tilde{x}_i^*) = 2 - \frac{\epsilon}{1 - \mu} + \mu^t \left( \frac{\epsilon}{1 - \mu} - 1 \right).$$

Es fácil calcular que el límite de el lado derecho (3.35) cuando  $\mu \rightarrow 1$  es  $(1 + \epsilon)(1 - \alpha)^{-1} > 1$  y por otra parte es igual a  $-\infty$  cuando  $\mu \rightarrow 1/\alpha$ . Entonces, por continuidad, existe  $b_\epsilon = b_\epsilon(\alpha)$  tal que

$$\tilde{V}_\alpha^*(1) = 1 = V_\alpha^*(1) \tag{3.36}$$

para cada  $\epsilon$ . Al mismo tiempo, el índice de estabilidad relativo en esta situación es:

$$\bar{\Delta}_\alpha(1) = \left[ \sum_{t=0}^{\infty} \alpha^t E_1^{\tilde{x}_i^*} C(x_t, a_t) - 1 \right] / 1 = \tag{3.37}$$

$$\sum_{t=1}^{\infty} \alpha^t E(2 - \xi_0 \xi_1 \dots \xi_{t-1}) = \frac{2\alpha}{1-\alpha} - \frac{3\alpha}{4-3\alpha} > 0.$$

Ahora, sea  $\bar{\nu}$  la métrica de probabilidad de Fortét-Mourier (correspondiente a la convergencia débil de variables aleatorias, vea [30]). De (3.30), (3.31) y de propiedades bien conocidas de  $\bar{\nu}$  obtenemos fácilmente que en este ejemplo

$$\sup_{(x,a)} \bar{\nu}(p(\cdot|x,a), \tilde{p}(\cdot|x,a)) \rightarrow 0 \text{ cuando } \epsilon \rightarrow 0$$

independientemente de la elección de  $b_\epsilon$ .

Por otro lado, para un factor de descuento fijo  $\alpha$ , usando una  $b_\epsilon$  que garantice (3.36), tenemos que las funciones de valor son iguales y el  $\inf_{\epsilon \in (0,1)} \bar{\Delta}_\alpha(1) > 0$ . Más aún, de (3.37) se tiene que  $\sup_{\alpha \in (0,1)} \Delta_\alpha(1) = \infty$  para cada  $\epsilon \in (0, 1)$ .

El próximo ejemplo presenta un modelo de control de Markov estable, para el que se cumplen todas las suposiciones de la Sección 2, y lo que es más importante, todas estas suposiciones y constantes involucradas en estas suposiciones, se expresan en términos simples (hasta cierto punto) de la existencia de momentos exponenciales. Además este ejemplo resulta una generalización del ejemplo I en el capítulo II, este ejemplo se usa en [10] para otro propósito.

**Ejemplo 3.1** Sea  $X = [0, \infty)$ ,  $A(x) = A$  para todo  $x \in X$  donde  $A$  es un conjunto compacto de algún intervalo  $(0, \theta]$  (con  $\theta \in A$ ). Defina

$$x_{t+1} = (x_t + a_t \eta_t - \xi_t)^+, \quad (3.38)$$

$$\tilde{x}_{t+1} = (\tilde{x}_t + a_t \tilde{\eta}_t - \tilde{\xi}_t)^+, \quad t = 0, 1, \dots, \quad (3.39)$$

$x_0 = \tilde{x}_0 = x$  dado, donde  $z^+ = \max(0, z)$ ;  $\{\eta_t\}$ ,  $\{\xi_t\}$ ,  $\{\tilde{\eta}_t\}$  y  $\{\tilde{\xi}_t\}$  son sucesiones i.i.d de v.a.'s no-negativas tales que las  $\{\eta_t\}$  son independientes de las  $\{\xi_t\}$  y las  $\{\tilde{\eta}_t\}$  son independientes de las  $\{\tilde{\xi}_t\}$ .

Las ecuaciones (3.38) y (3.39) representan una versión controlada de la caminata aleatoria en  $[0, \infty)$ , la cual aparece en varios modelos aplicados, por ejemplo, en manejo de recursos acuíferos o en producción-inventarios (vea [17]). Otras aplicaciones importantes del proceso dado en (3.38) es un modelo de control de las tasas de servicio  $a_t$  en un sistema de colas del tipo  $GI|GI|1|\infty$ . En este caso  $x_t$  se interpreta como el tiempo de espera de el  $t$ -ésimo cliente, mientras  $\xi_t$  es el tiempo entre la llegada del  $t$ -ésimo cliente y la del  $(t+1)$ -ésimo cliente. Las variables aleatorias  $\eta_t$  ( $t = 0, 1, \dots$ ) describen diferencias entre las variables designadas  $a_t$  y las tasas de servicio real. Las variables en el proceso aproximante (3.39) se interpretan de la misma manera. En [43] Se pueden encontrar comentarios de las aplicaciones de este modelo.

Probaremos que este ejemplo cumple las hipótesis del Teorema 3.1, si se cumplen las suposiciones 3.5 y 3.6. Denotaremos por  $\eta, \xi, \tilde{\eta}, \tilde{\xi}$  a las v.a. genericas r.v.'s distribuidas, como  $\eta_0, \xi_0, \tilde{\eta}_0, \tilde{\xi}_0$ , y definimos

$$\zeta = \theta\eta - \xi, \quad \tilde{\zeta} = \theta\tilde{\eta} - \tilde{\xi}. \quad (3.40)$$

### Suposición 3.5

(a) Las v.a.  $\eta, \xi, \tilde{\eta}, \tilde{\xi}$  tienen densidades continuas y acotadas en  $[0, \infty)$

b)

$$E\zeta < 0, \quad E\tilde{\zeta} < 0; \quad (3.41)$$

c) existen constantes positivas  $h', \tilde{h}'$  tales que

$$E \exp(h'\zeta) < \infty, \quad E \exp(\tilde{h}'\tilde{\zeta}) < \infty. \quad (3.42)$$

Como se observa en [10] de (3.41) y (3.42) se deduce la existencia de  $h > 0$  tal que

$$\beta \doteq \max\{E \exp(h\zeta), E \exp(h\tilde{\zeta})\} < 1. \quad (3.43)$$

**Suposición 3.6** *La función de costo de un paso  $c$  es medible y estrictamente positiva tal que, para toda  $x \in [0, \infty)$ ,  $c(x, \cdot)$  es semicontinua por abajo en  $A$ , y*

$$\sup_{a \in A} c(x, a) \leq (\bar{b})^{1/s} \exp(xh/s),$$

donde  $s > 1$  y  $\bar{b}$  son constantes positivas arbitrarias.

**Nota.** La expresión  $E \exp(h\zeta)$  se puede calcular explícitamente para algunos casos específicos. Por ejemplo en [10], si  $\eta$  y  $\xi$  tienen distribución exponencial con medias  $E(\eta) = \frac{1}{w_1}$  y  $E\left(\frac{1}{w_2}\right)$  entonces

$$E \exp(h\zeta) = \left[ \frac{w_1}{w_1 - h\theta} \right] \left[ \frac{\xi}{\xi + h} \right]$$

y  $E \exp(h\zeta) < 1$  si  $h < \frac{w_1}{\theta} - w_2$ .

Bajo las Suposiciones 3.5 y 3.6, el trabajo de verificar las Suposiciones 3.1, 3.2, 3.3 y 3.4 de la Sección 2, escogiendo  $W(x) = \bar{b} \exp(hx)$  fue hecho en [10], usando las siguientes funciones

$$h_f(x) = P(x + f(x)\eta - \xi \leq 0),$$

$$\tilde{h}_f(x) = P(x + f(x)\tilde{\eta} - \tilde{\xi} \leq 0), \quad x \in [0, \infty);$$

$$\nu = \delta_0 \quad \text{the Dirac distribution}.$$

Entonces las cotas (3.7), (3.13) y el Corolario (3.1) se cumplen para los procesos (3.38) y (3.39). Más aún, la distancia de la variación total en (3.13) se puede calcular (véase [39]) en términos de las densidades de  $\eta, \xi, \tilde{\eta}, \tilde{\xi}$ ; y las potencias de  $(s-1)/s$  pueden ser escogidas tan cercanas a 1 como

se quiera si el  $\sup_{a \in A} c(x, a)$  está dominado por alguna función polinomial. Desafortunadamente, en este ejemplo no es posible simplificar las expresiones para las constantes involucradas en  $B(x)$  en (3.7) comparado con (3.5), (3.6). La única observación es que  $\|\nu\|_W = \bar{b}$ . Es por esto que en el siguiente capítulo usaremos los mismos argumentos del Teorema (3.1) para encontrar una cota como la de (3.7) reemplazando  $B(x)$  en (3.9) por otra constante  $D(x)$  más sencilla. El punto es reemplazar en las pruebas (3.21), (3.22) por otras estimaciones de las tasas de convergencias para procesos de Markov estocásticamente ordenados en  $X \subset \mathbb{R}$  encontradas en [37], [24] y en [34]. Las cotas para las tasas convergencias en estos Trabajos, tienen expresiones más simples, comparadas con  $\bar{B}$  en (3.21), (3.22) (vea los ejemplos en [37]). Esperamos que el uso de estas cotas para este caso particular nos proporcione mejores cotas para  $\bar{\Delta}_\alpha(x)$  en (3.7).

# Capítulo 4

## Estabilidad vía cadenas ordenadas

### 4.1 Introducción

En este capítulo reemplazaremos las tasas de convergencia a las distribuciones límite de Kartashov [20], que usamos en las desigualdades (3.21 y 3.22), por tasas que son más fáciles de calcular y nos dan mejores cotas de estabilidad. Se ha desarrollado un trabajo considerable (en los últimos años) sobre el problema de determinar cuantitativamente tasas de convergencia entre la probabilidad de transición de  $n$ -pasos  $P^n$  y la medida invariante  $m$  de una cadena para cadenas de Markov no controladas.

En los últimos años, se ha desarrollado un trabajo considerable sobre el problema de determinar cuantitativamente tasas de convergencia entre la probabilidad de transición de  $n$ -pasos  $P^n$  y la medida invariante  $m$ , para cadenas de Markov no controladas. En [27],[35] y [36] se han dado cotas geométricas para dicha convergencia, de la siguiente forma:

$$\sigma(P^n(\cdot|x), m(\cdot)) \leq v(x)\rho^n, \quad \rho < 1, \quad (4.1)$$

donde  $\sigma$  es la métrica de la variación total (definida en el capítulo 3) y  $\rho$  nos da la tasa de convergencia.

La función  $v$  y la constante  $\rho$  se calculan en términos de conjuntos pequeños y condiciones de sesgo de tipo Lyapunov-Foster. Sin embargo estas cotas salvo en casos particulares, no son muy "exactas", por ejemplo en ([35]) para obtener una cota como (4.1) se requiere de un acoplamiento ("coupling", véase [22]) de una cadena de Markov bivariada en muchos casos estas

cotas no son muy exactas [34]. Esto para cadenas no controladas, si además pedimos que las condiciones de ergodicidad se cumplan uniformemente para todas las políticas estacionarias, no podemos esperar que las cotas sean buenas, sin embargo en [24] se ha mostrado que si pedimos que la cadena tenga monotonicidad estocástica, se pueden obtener mejores cotas. Estos resultados nos permiten establecer velocidad de convergencia geométrica para procesos controlados por políticas estacionarias y determinísticas ( $f \in \Pi_{de}$ ), y calcular las constantes involucradas en las cotas de las tasas de convergencia en términos de los momentos exponenciales de tiempo de regreso a un estado o estado fijo.

Para algunos ejemplos podemos "heredar" las condiciones de ergodicidad simplemente, si para cada política  $f \in \Pi_{de}$ , acotamos el proceso controlado por una cadena estocásticamente ordenada. Luego como en el Teorema 3.1, reduciremos el problema al caso acotado y aplicaremos la técnica de Zolotarev [46] para establecer cotas para el índice de estabilidad para cadenas no controladas de la siguiente forma:

$$\Delta_{\alpha}(x) \leq M(\alpha, x) \delta^{(s-1)/s} \max\{1, [\log_{\rho} \delta]\} \text{ con } s > 1 \text{ y}$$

$$M(\alpha, x) = O\left(\frac{1}{1-\alpha}\right) \text{ cuando } \alpha \rightarrow 1.$$

Este enfoque nos permite obtener de manera más sencilla las constantes que aparecen en las cotas para el índice de estabilidad  $\Delta_{\alpha}(x)$ , además, esperamos que las cotas obtenidas sean más exactas comparadas con la desigualdad (3.7) Sin embargo este enfoque, se aplica solo a una clase restringida de procesos controlables, aquellos que pueden ser acotados por procesos estocásticamente ordenados.

## 4.2 Preliminares

**Definición 4.1** Sean  $\lambda$  y  $\mu$  dos distribuciones en  $\mathbb{R}$ , decimos que  $\lambda$  es estocásticamente menor que  $\mu$  si,  $\lambda([y, \infty)) \leq \mu([y, \infty))$  para  $y \in \mathbb{R}$  y lo denotaremos por  $\lambda \leq^s \mu$ .

**Definición 4.2** Decimos que un kernel  $P(\cdot|\cdot)$  en  $\mathbb{R}$  es monótono si, para  $x, \hat{x} \in \mathbb{R}$  con  $x \leq \hat{x}$   $P(\cdot|x) \leq^s P(\cdot|\hat{x})$ .

**Definición 4.3** Decimos que una cadena de Markov es estocásticamente ordenada si está gobernada por un kernel monótono

**Ejemplo 4.1** La caminata aleatoria en  $[0, \infty)$ , definida por  $Y_0 = x$ ,  $x \geq 0$  y  $Y_{n+1} = (Y_n + \xi_{n+1})^+$  para  $n \geq 0$  (donde las  $\xi_n$  forman una sucesión de variables aleatorias iid), es un ejemplo de una cadena de Markov estocásticamente ordenada [24].

Se pueden encontrar otros ejemplos de cadenas estocásticamente ordenadas en teoría de inventarios [38],[25].

En [22] se demuestra que si  $P$  es un kernel monótono y  $\lambda, \mu$  son distribuciones de  $\mathbb{R}$  donde  $\lambda \leq^s \mu$ , entonces se pueden construir cadenas  $\{Y_n\}$  y  $\{Y'_n\}$  gobernadas por  $P$  con distribución inicial  $\lambda$  y  $\mu$  respectivamente tales que,

$$Y_n \leq Y'_n, \quad n \geq 0, \text{ con probabilidad } 1 \quad (4.2)$$

por esta razón cuando consideremos dos cadenas  $\{Y_n\}$  y  $\{Y'_n\}$ , gobernadas por el mismo kernel monótono  $P$ , con distribución inicial  $\lambda$  y  $\mu$  respectivamente donde  $\lambda \leq^s \mu$  daremos por supuesto que se cumple (4.2).

En este capítulo estudiamos en particular tasas de convergencia de cadenas de Markov (no controladas) de la probabilidad de  $n$  pasos a la medida de probabilidad invariante, en términos de la métrica de la variación total. Para manejar estos conceptos necesitamos, las siguientes definiciones. Consideremos una cadena Markov  $\{Y_n\}$  con espacio de estados  $X \subset \mathbb{R}$  con probabilidad de transición  $P$ . Denotaremos por  $P^n$  a la probabilidad de  $n$  pasos, i.e. para  $A \in \mathbb{B}(X)$ ,  $x \in X$  defina

$$P^n(A|x) \doteq P(Y_n \in A|x), \quad n = 1, 2, \dots$$

**Definición 4.4** Dada  $f : X \rightarrow X$ , defina:  $Pf(x) = \int_X f(z) P(dz|x)$ ,  $x \in X$

**Definición 4.5** El primer tiempo de alcance a  $A \in \mathbb{B}(X)$  se define como :

$$\tau_A = \inf \{n > 0; Y_n \in A\}. \quad (4.3)$$

### 4.3 Estabilidad en el caso en que $0 \in \mathbb{X}$ se alcanza para cada política de control

#### 4.3.1 Tasas de convergencia para cadenas ordenadas no controladas

Sea  $\{Y_n, n \geq 0\}$  una cadena de Markov a tiempo discreto en el espacio  $(\Omega, \mathcal{F}, P)$  que toma valores en el espacio de estados  $\mathbb{X} \subset \mathbb{R}$  y cuya probabilidad de transición es  $P$ . En esta sección supondremos que  $0 \in \mathbb{X}$ .

**Suposición 4.1** *Nuestra probabilidad de transición cumple que para toda  $x \in \mathbb{X}$ ,  $P^n(0|x) > 0$  para algún  $n$ ,  $n = 1, 2, \dots$  (condición de irreducibilidad).*

Parte de los resultados serán expresados en términos de la siguiente función generadora del tiempo de regreso al estado 0.

Como en 4.3 definimos  $\tau_0 = \inf \{n > 0; Y_n \in 0\}$

**Definición 4.6** *La función generadora de  $\tau_0$  se define por*

$$G_x(r) = E_x[r^{\tau_0}], \quad r > 0, \quad (4.4)$$

donde  $x \in \mathbb{X}$  es un estado fijo.

En [24] se demuestra el siguiente resultado acerca de la convergencia a la distribución límite y sus tasas de convergencia para procesos estocásticamente ordenados.

**Teorema 4.1** *Suponga que  $\{Y_n\}$  es una cadena con espacios de estados  $[0, \infty)$ , estocásticamente ordenada para la cual existe una distribución límite  $m_Y$  y una función  $v : [0, \infty) \rightarrow [1, \infty)$  con  $v(0) = 1$  tal que*

$$Pv(x) \leq \lambda v(x) + b\mathbb{I}_{\{0\}}(x) \quad \text{donde } 0 \leq \lambda < 1 \text{ y } 0 \leq b < \infty \quad (4.5)$$

( $\mathbb{I}_A$  es la función indicadora del conjunto  $A$ ).

Entonces

$$\sup_{A \in \mathcal{B}(X)} |P^n(x, \cdot) - m_Y| \leq r^{-n} C_r(x) \quad \text{para todo } r < \lambda^{-1} \text{ y } x \geq 0, \quad (4.6)$$

donde

$$C_x(r) \leq E_x[r^{\tau_0}] + \frac{b}{1-\lambda} < \infty \quad (4.7)$$

Antes de probar este Teorema, enunciaremos dos Lemas (también se demuestran en [24]) uno de ellos nos permiten relacionar las condiciones de sesgo de tipo Foster-Lyapunov (4.5) con la función  $G_x(r)$  definida en (4.4), y el otro garantiza la existencia de una función  $v$  que satisface la desigualdad (4.5), cuando  $G_0(r) < \infty$ . En algunos casos, como en el ejemplo que presentamos al final de esta sección, el lema proporciona una manera de calcular las constantes involucradas, en la desigualdad (4.6).

**Lema 4.1** *Dada una cadena  $\{Y_n\}$  de Markov estocásticamente ordenada que cumple las desigualdad (4.5) entonces  $G_x(r) \leq v(x)$  para  $r \leq \lambda^{-1}$  si  $x > 0$  y  $G_0(r) \leq r(\lambda + b)$  para  $r \leq \lambda^{-1}$ .*

**Lema 4.2** *Dada una cadena  $\{Y_n\}$  una de Markov estocásticamente ordenada, con  $G_0(r_1) < \infty$ , para algún  $r_1 > 1$ . Entonces la función definida por  $v(0) = 1$ ,  $v(x) = G_x(r_1)$ , para  $x > 0$ , junto con las constantes  $\lambda = r_1^{-1}$  y  $b = r_1^{-1}[G_0(r_1) - 1]$ , satisfacen las condiciones de sesgo (4.5) con igualdad.*

**Dem.. (esquema de la demostración del Teorema 4.1)**

La idea para obtener (4.6) es acoplar la cadena estacionaria, con la cadena que tiene como estado inicial  $x$ . Sea  $\{Y_n\}$  la cadena cuando  $Y_0 = x$  y  $\{Y'_n\}$  denota la cadena cuando  $Y'_0 = X$ . Aquí  $X$  es una variable aleatoria en  $(\Omega, \mathcal{F}, P)$  con distribución igual a la distribución límite  $m_Y$ . Como  $\{Y'_n\}$  es estacionaria;  $P_{m_Y}(Y'_n \in A) = m_Y(A)$  para toda  $A \in \mathbb{B}(X)$ .

Defina el tiempo de acople como

$$T = \inf \left\{ n \geq 0; Y_n = Y'_n \right\}.$$

Notemos que se cumple la siguiente desigualdad:

$$\begin{aligned}
P(Y_n \in A) - P(Y'_n \in A) &= P(Y_n \in A, T \leq n) + P(Y_n \in A, T \geq n) \\
&\quad - P(Y' \in A, T \leq n) - P(Y'_n \in A, T \geq n) \\
&= P(Y_n \in A, T \leq n) - P(Y'_n \in A, T \geq n) \\
&\quad + P(Y_n \in A, T \geq n) - P(Y' \in A, T \leq n) \\
&\leq P(Y'_n \in A, T \geq n) \leq P(T \geq n),
\end{aligned}$$

de aquí obtenemos:

$$\sup_{A \in \mathbb{B}(X)} |P^n(A|x) - m_Y(A)| \leq P_{x m_Y}(T > n). \quad (4.8)$$

Los dos subíndices de  $P$  en (4.8) indican la dependencia de  $T$  en las distribuciones iniciales de  $\{Y_n\}$  y  $\{Y'_n\}$

Sin embargo, como las cadenas están ordenadas, una vez que la cadena con valor inicial mayor alcance el estado 0, la cadena con menor valor inicial deberá estar en 0 también, por lo tanto  $P_{x m_Y}(T > n) \leq P_\nu(\tau_0 > n)$ , donde  $\nu(A) = P(\max(X, x) \in A)$  para toda  $A \in \mathbb{B}(X)$ , de donde obtenemos

$$\sup_{A \in \mathbb{B}(X)} |P^n(A|x) - m_Y(A)| \leq P_\nu(\tau_0 > n), \quad (4.9)$$

y usando la desigualdad de Markov

$$r^n P_\nu(\tau_0 > n) \leq E_\nu[r^{\tau_0}], \quad (4.10)$$

solo nos faltaría estimar  $E_\nu[r^{\tau_0}]$

$$\begin{aligned}
E_\nu[r^{\tau_0}] &= \int_{[0, \infty)} G_z(r) \nu(dz) \\
&= P(X \leq x) G_x(r) + \int_{(x, \infty)} G_z(r) m_Y(dz) \\
&\leq G_x(r) + \int_{(x, \infty)} G_z(r) m_Y(dz) \gamma.
\end{aligned} \quad (4.11)$$

Como se cumplen las condiciones 4.5, el Lema 4.1 nos asegura que

$$G_x(r) \leq v(x) \text{ , para } r \leq \lambda^{-1}, x > 0$$

entonces:

$$\int_{(x, \infty)} G_z(r) m_Y(dz) \leq \int_{[0, \infty)} v(z) m_Y(dz) \text{ ,}$$

así, aplicando (14.3.7) de [27]) con  $f(x) = (1 - \lambda)v(x)$  y  $s(x) = b\mathbb{I}_{\{0\}}(x)$ , obtenemos

$$\int_{[0, \infty)} v(z) m_Y(dz) \leq \frac{b}{1 - \lambda} \text{ ,} \quad (4.12)$$

combinando las desigualdades (4.9),(4.11) y (4.12) obtenemos

$$\sup_{A \in \mathcal{B}(X)} |P^n(x, A) - m_Y(A)| \leq C_x(r) r^{-n} \quad (4.13)$$

que es lo que queríamos probar. ■

Por otro lado, se pueden obtener tasas de convergencia a la distribución límite para cadenas no-ordenadas, a partir de cadenas estocásticamente ordenadas. Más específicamente si  $\{X_n\}$  es una cadena de Markov con espacios de estados  $[0, \infty)$  y  $\{Y_n\}$  una cadena de Markov (también con espacio de estados  $[0, \infty)$ ) estocásticamente ordenada en  $(\Omega, \mathcal{F}, P)$  con distribuciones límite  $m_X$  y  $m_Y$  respectivamente, con la propiedad de que:  $X_n(\omega) \leq Y_n(\omega)$  para toda  $n \geq 0$  y toda  $\omega \in \Omega$ . Sean  $X$  y  $Y$  variables aleatorias en  $(\Omega, \mathcal{F}, P)$  con distribuciones  $m_X$  y  $m_Y$  respectivamente. Como  $X_n(\omega) \leq Y_n(\omega)$  para toda  $n$  y  $\omega \in \Omega$ , tenemos  $m_Y([0, a]) \leq m_X([0, a])$  para toda  $a \geq 0$ , por lo tanto  $Y$  es estocásticamente mayor que  $X$ . De [22], podemos suponer que  $X(\omega) \leq Y(\omega)$  para toda  $\omega \in \Omega$ .

Ahora considere las cadenas  $\{X_n\}$  y  $\{X'_n\}$  para la cadena general, cuando las distribuciones iniciales son  $X_0 = x_0$  y  $X'_0 = X$  respectivamente, también consideremos  $\{Y_n\}$  y  $\{Y'_n\}$  para la cadena estocásticamente ordenada cuando las distribuciones iniciales son  $Y_0 = x_0$  y  $Y'_0 = Y$  respectivamente. Si  $\{Y_n\}$  y  $\{Y'_n\}$  están en el estado 0 en el tiempo  $k$ , luego por los ordenamientos  $\{X_n\}$  y  $\{X'_n\}$ , también estarán en el estado 0 en el tiempo  $k$ , entonces  $\{X_n\}$  y  $\{X'_n\}$  se han acoplado en el tiempo  $k$  y

$$\tilde{P}_{x m_X}(\tilde{T} > n) \leq \tilde{P}_v(\tilde{\tau}_0 > n) \text{ ,} \quad (4.14)$$

donde  $\tilde{T} = \inf \{n > 0 : X_n = X'_n\}$  y  $\nu$  es la medida definida por  $\nu(A) = P(\max(X, x) \in A)$  y  $\tilde{\tau}_0 = \inf \{n > 0; Y_n = 0\}$ , combinando (4.9), (4.10) y (4.14) y los argumentos del Teorema 4.1 se obtiene el siguiente resultado que aparece en [24].

**Teorema 4.2** *Suponga que  $\{X_n\}$  es una cadena de Markov, con espacios de estados  $[0, \infty)$ , acotada por una cadena estocásticamente ordenada, digamos  $\{Y_n\}$  (con el mismo espacio de estados), esto es,  $X_n(\omega) \leq Y_n(\omega)$  para  $n \geq 0$  y  $\omega \in \Omega$ , más aún suponga que  $\{X_n\}$  y  $\{Y_n\}$  tienen distribuciones límite  $m_X$  y  $m_Y$  respectivamente, además supongamos que  $\{Y_n\}$  cumple las condiciones (4.5) para  $\lambda$  y  $b$ ,  $0 \leq \lambda < 1$ ,  $0 \leq b < \infty$ .*

Entonces,

$$\sup_{A \in \mathbb{B}(X)} \left| \tilde{P}^n(x, A) - m_X(A) \right| \leq C_x(r) r^{-n},$$

donde  $\tilde{P}$  es la probabilidad de transición de la cadena  $\{X_n\}$ .

### 4.3.2 Estabilidad para Procesos Controlados

Como siempre consideraremos un par de *modelos de control de Markov* a tiempo discreto  $\mathcal{P} = (X, A, A(x), c, p)$  y  $\tilde{\mathcal{P}} = (X, A, A(x), c, \tilde{p})$ , con espacio de estados  $\mathbb{X} = [0, \infty)$ , de acciones  $A \subset [0, \infty)$ , donde  $A$  es compacto.

Los únicos componentes diferente en estos modelos, son las *probabilidades de transición*  $p(B|x, a)$  y  $\tilde{p}(B|x, a)$ ,  $(x, a) \in \mathbb{K}$   $B \in \mathbb{B}(X)$ .

Igual que en los capítulos anteriores, usaremos como criterio de desempeño para los procesos  $\mathcal{P}$  y  $\tilde{\mathcal{P}}$ , *las esperanzas del costo total  $\alpha$ -descontado*  $V_\alpha(\pi, x)$  y  $\tilde{V}_\alpha(\pi, x)$ , definidas por las ecuaciones (1.6) y (1.7), y como antes denotaremos por:  $\pi_\alpha^*$  y  $\tilde{\pi}_\alpha^*$  a las políticas  $\alpha$ -óptimas, de  $\mathcal{P}$  y  $\tilde{\mathcal{P}}$  respectivamente. Entonces tenemos el siguiente lema que nos permite calcular tasas de convergencia para procesos controlados,

**Lema 4.3** *Sea  $\mathcal{P} = (X, A, A(x), c, p)$  un proceso de control de Markov, para el que existe  $\{Y_n\}$  una cadena de Markov (eo), con distribución límite  $m_Y$  y cumple las condiciones (4.5) para  $\lambda$  y  $b$ ;  $0 \leq \lambda \leq 1$ ,  $0 \leq b < \infty$ , tal que, para  $f \in \mathbb{F}$   $x_n^f(w) \leq Y_n(w)$  para  $n \in \mathbb{N}$  y para  $w \in \Omega$  (donde  $x_n^f$  es el proceso que se obtiene al aplicar la política  $f$  al proceso  $\mathcal{P}$ ), si para*

cualquier política  $f \in \mathbb{F}$  el proceso  $\{x_f\}$  tienen distribución límite  $q_f$ , entonces

$$\sup_{A \in \mathcal{B}(X)} |P(x_n^f \in A | x) - q_f(A)| \leq C_x(r) r^{-n} \text{ para } f \in \mathbb{F},$$

para  $r \leq \lambda^{-1}$  donde  $C_x(r)$  cumple la desigualdad (4.7).

**Dem..** Aplicando el teorema 4.2 al proceso  $\{x_t^f\}$  el resultado se sigue. ■

Pasaremos ahora a la estimación del índice de estabilidad

$$\Delta_\alpha(x) = [V_\alpha(\tilde{\pi}_\alpha^*, x) - V_\alpha(\pi_\alpha^*, x)]$$

para los dos procesos  $\mathcal{P}$  y  $\tilde{\mathcal{P}}$ . La idea principal en el enfoque de este capítulo es encontrar, cadenas de Markov no controlados que tienen buenas propiedades de convergencia geométrica a sus respectivas distribuciones invariantes y que acoten a los procesos  $\mathcal{P}$  y  $\tilde{\mathcal{P}}$ , Para conseguir estos objetivos utilizaremos las siguientes hipótesis.

**Suposición 4.2** Existen cadenas de Markov (no controladas)  $\{Y_n\}$  y  $\{\tilde{Y}_n\}$  estocásticamente ordenadas ambas con espacios de estado  $[0, \infty)$ , con distribuciones invariantes  $m_Y$  y  $m_{\tilde{Y}}$ , tales que, para cada política estacionaria  $f \in \mathbb{F}$  y para todo valor inicial  $x \in [0, \infty)$ , se cumple

$$\begin{aligned} x_n^f(\omega) &\leq Y_n(\omega) \\ \tilde{x}_n^f(\omega) &\leq \tilde{Y}_n(\omega) \end{aligned}$$

para  $\omega \in \Omega$ ,  $n = 1, 2, \dots$

Aquí  $\{x_n^f\}$  y  $\{\tilde{x}_n^f\}$  son las cadenas de los procesos  $\mathcal{P}$  y  $\tilde{\mathcal{P}}$  al aplicar la política estacionaria  $f$ .

**Suposición 4.3** Existe una función  $W : [0, \infty) \rightarrow [1, \infty)$  medible, no decreciente, con  $W(0) = 1$  y constantes  $\lambda, b$ ,  $0 \leq \lambda < 1$ ,  $0 \leq b < \infty$ , tales que para todo  $x \geq 0$

$$\begin{aligned} P^Y W(x) &\leq \lambda W(x) + b \mathbb{1}_{\{0\}} \text{ y} \\ P^{\tilde{Y}} W(x) &\leq \lambda W(x) + b \mathbb{1}_{\{0\}}. \end{aligned} \tag{4.15}$$

Donde  $P^Y$  es la probabilidad de transición de  $\{Y_n\}$  y  $P^{\tilde{Y}}$  es la probabilidad de transición de  $\{\tilde{Y}_n\}$ .

**Suposición 4.4** Para ambos procesos  $\mathcal{P}$  y  $\tilde{\mathcal{P}}$  la Suposición 3.1 del capítulo 3 se cumplen con la función  $W$  introducida en la suposición 4.3.

En particular ,

$$\sup_{a \in A(x)} c(x, a) \leq [W(x)]^{\frac{1}{s}}, \quad x \in [0, \infty)$$

para algún  $s > 1$ .

**Nota 9** El Lema 4.2 nos permite calcular la función  $W$  y las constantes  $\lambda$  y  $b$  de la suposición 4.3 en términos de los momentos exponenciales de los tiempos de regreso al estado 0 de los procesos  $\{Y_n\}$  y  $\{\tilde{Y}_n\}$ .

Para establecer el Teorema Principal ( de estabilidad) necesitamos introducir las siguientes constantes .

Fijamos una  $r$  tal que  $1 < r < \lambda^{-1}$  y denotamos  $\rho = 1/r$ .

Denotaremos  $d_x = \max \{E_x [r^{\tau_0}], E_x [r^{\tilde{\tau}_0}] \} + \frac{b}{1-\lambda}$ ,

donde

$$\begin{aligned} \tau_0 &\doteq \min \{n > 0 : Y_n = 0\} \\ \tilde{\tau}_0 &\doteq \min \{n > 0 : \tilde{Y}_n = 0\} \end{aligned}$$

**Nota 10** El lema 4.1 garantiza que  $d_x < \infty$  y da cotas superiores para  $d_x$  , en términos de  $W(x)$ ,  $r$ ,  $\lambda$  y  $b$ .

**Teorema 4.3** Sean dos procesos de control de Markov  $\mathcal{P} = (X, A, A(x), c, p)$  y  $\tilde{\mathcal{P}} = (X, A, A(x), c, \tilde{p})$  con  $X = [0, \infty)$  ,  $A \subset [0, \infty)$  y  $A(x)$  compacto, para los cuales se cumplen las suposiciones (4.2) , (4.3) y (4.4), entonces

(a) existen  $f^*$  y  $\tilde{f}^* \in \mathbb{F}$  políticas óptimas para  $\mathcal{P}$  y  $\tilde{\mathcal{P}}$  respectivamente.

(b) se obtiene la siguiente desigualdad para el índice de estabilidad,

$$\begin{aligned} \Delta(x) &= V_\alpha(x, \tilde{f}^*) - V_\alpha(x, f^*) \leq \\ &\frac{2}{1-\alpha} \{2(W(x) + b(1-\lambda)^{-1}) + 4d_x r + 1\} \\ &\times \delta^{\frac{(s-1)}{s}} \max\{1, [\log_\rho \delta]\}. \end{aligned} \quad (4.16)$$

donde  $s > 1$  es la constante de la suposición 4.4, y

$$\delta = \sup_{(x,a) \in \mathbb{K}} \sigma(p(\cdot | x, a), \tilde{p}(\cdot | x, a)),$$

**Dem..** De la suposición 4.3 y el Teorema 4.1 se sigue que,

$$\begin{aligned} \sup_{A \in \mathbb{B}(X)} |P(Y_n \in A | Y_0 = x) - m_Y(A)| &\leq d_x(r) \rho^n \text{ y} \\ \sup_{A \in \mathbb{B}(X)} |P(\tilde{Y}_n \in A | \tilde{Y}_0 = x) - m_{\tilde{Y}}(A)| &\leq d_x(r) \rho^n \end{aligned}$$

Aplicando la Suposición 4.2 y el Lema 4.3 se obtienen las siguientes desigualdades

$$\begin{aligned} \sup_{A \in \mathbb{B}(X)} |P^f(x_n^f \in A | x_0 = x) - q_f(A)| &\leq d_x(r) \rho^n \\ \sup_{A \in \mathbb{B}(X)} |\tilde{P}^f(\tilde{x}_n^f \in A | \tilde{x}_0 = x) - \tilde{q}_f(A)| &\leq d_x(r) \rho^n \end{aligned} \quad (4.17)$$

$x \in [0, \infty)$ ,  $n = 1, 2, \dots$ ,  $f \in \mathbb{F}$ ,

donde  $P^f$  y  $\tilde{P}^f$  son las probabilidades de transición de las cadenas de Markov que se obtienen al aplicar la política  $f$  a los procesos  $\mathcal{P}$  y  $\tilde{\mathcal{P}}$ , respectivamente. Las desigualdades (4.17) juegan el mismo papel de las desigualdades (3.21) y (3.22) en la demostración del Teorema 3.1 del capítulo 3. La demostración se obtiene repitiendo los argumentos del Teorema 3.1.

Lo único que faltaría establecer, como en (3.17) y en (3.18), son las siguientes cotas:

$$\begin{aligned} \sup_{t \geq 0} E_x^f W(x_t^f) &\leq W(x) + b/(1-\lambda) \\ \sup_{t \geq 0} E_x^f W(\tilde{x}_t^f) &\leq W(x) + b/(1-\lambda) \end{aligned} \quad (4.18)$$

$f \in \mathbb{F}$ .

En cada caso, es suficiente mostrar que (vea demostración del Teorema 3.1)

$$\sup_{f \in \mathbb{F}} P^f W(x) \leq \lambda W(x) + b \quad (4.19)$$

$$\sup_{f \in \mathbb{F}} \tilde{P}^f W(x) \leq \lambda W(x) + b \quad (4.20)$$

Las últimas desigualdades se siguen de la desigualdad de la suposición 4.3 y de la observación de que:

$$x_n^f(\omega) \leq Y_n(\omega), \quad \omega \in \Omega, \quad n = 1, 2, \dots$$

(suposición 4.2) y la monotonía de  $W(x)$  implica que, para ( $f \in \mathbb{F}$ )

$$\begin{aligned} P^f W(x) &= \int W(y) P(dy | x, f(x)) \\ &= E_x W(x_1^f) \leq E_x W(Y_1) \\ &= \int W(y) P^Y(dy | x) \\ &= P^Y W(x) \leq \lambda W(x) + b \mathbb{1}_{\{0\}} \\ &= \lambda W(x) + b. \end{aligned}$$

Esto implica la desigualdad (4.19).

La desigualdad (4.20) se demuestra de manera similar. En particular, las hipótesis del Teorema 8.3.6 de [18] se cumplen. Entonces, podemos concluir la existencia de las políticas óptimas estacionarias  $f^*$  y  $\tilde{f}^*$ , para los procesos  $\mathcal{P}$  y  $\tilde{\mathcal{P}}$ . ■

También obtenemos el equivalente del Corolario 3.2;

**Corolario 4.1** Sean los modelos  $\mathcal{P}$  y  $\tilde{\mathcal{P}}$  que cumplen las hipótesis del Teorema 4.6 y que están definidos por las ecuaciones recurrentes

$$\begin{aligned} x_{t+1} &= F(x_t, a_t, \xi_t) \\ \tilde{x}_{t+1} &= F(\tilde{x}_t, \tilde{a}_t, \tilde{\xi}_t) \end{aligned} \quad t = 0, 1, \dots$$

obtenemos la siguiente desigualdad para el índice de estabilidad

$$\Delta_\alpha(x) \leq \frac{2}{1-\alpha} \{2(W(x) + b(1-\lambda)^{-1}) + 4d_x r + 1\} [\sigma(\mu_\xi, \mu_{\tilde{\xi}})]^{\frac{\alpha-1}{\alpha}},$$

Siempre que  $\sigma(\mu_\xi, \mu_{\tilde{\xi}}) \leq e^{-\frac{\alpha}{(\alpha-1)}}$  donde  $\mu_\xi, \mu_{\tilde{\xi}}$  son las distribuciones de  $\xi$  y  $\tilde{\xi}$  respectivamente.

### 4.3.3 Ejemplo

**Ejemplo 4.2** Retomaremos el ejemplo 3.1 del Capítulo 3, consideremos los procesos

$$\begin{aligned} x_{n+1} &= (x_n + a_n \eta_n - \xi_n)^+ \\ \tilde{x}_{n+1} &= (\tilde{x}_n + a_n \tilde{\eta}_n - \tilde{\xi}_n)^+, \quad n = 1, 2, \dots \end{aligned} \quad (4.21)$$

$A(x) = A$  para todo  $x \in X$  donde  $A$  es un conjunto compacto de algún intervalo dado  $(0, \theta]$  (con  $\theta \in A$ ), con  $x_0 = \tilde{x}_0 = x$ , con  $x \in [0, \infty)$  y  $\{\eta_n\}$ ,  $\{\xi_n\}$ ,  $\{\tilde{\eta}_n\}$  y  $\{\tilde{\xi}_n\}$  son sucesiones de variables aleatorias independientes e idénticamente distribuidas. Además las  $\{\eta_n\}$  son independientes de las  $\{\xi_n\}$  y las  $\{\tilde{\eta}_n\}$  son independientes de las  $\{\tilde{\xi}_n\}$ . Denotamos por  $\eta, \xi, \tilde{\eta}, \tilde{\xi}$ , a las va genéricas distribuidas como  $\eta_0, \xi_0, \tilde{\eta}_0, \tilde{\xi}_0$ , definimos

$$\rho = \theta\eta - \xi, \quad \tilde{\rho} = \theta\tilde{\eta} - \tilde{\xi}$$

y suponemos que se cumple la siguiente ,

#### Suposición 4.5

- a)  $E\rho < 0$ ,  $E\tilde{\rho} < 0$  ;  
 b) sea  $\Phi(r) = E(r^\zeta)$  suponga que :

existe una  $r_0 > 1$  con  $\Phi(r_0) < \infty$  y que satisface  $\Phi'(r_0) = 0$

- c) La condición b) se cumple para  $\tilde{\zeta}$  con  $\tilde{\Phi}(r) = E(r^{\tilde{\zeta}})$  y algún  $\tilde{r}_0 > 1$ .

d) La función de costo de un paso cumplen condiciones a) y b) de las suposiciones (3.1) del capítulo 3 con

$$W(x) = \begin{cases} \max \{E_x[r^{\tau_0}], E_x[r^{\tilde{\tau}_0}]\} & \text{si } x > 0, \\ 1 & \text{si } x = 0. \end{cases}$$

Donde

$$\begin{aligned} \tau_0 &\doteq \min \{n > 0 : Y_n = 0\} \\ \tilde{\tau}_0 &\doteq \min \{n > 0 : \tilde{Y}_n = 0\}. \end{aligned}$$

Probaremos que este ejemplo cumple las condiciones del Teorema 4.3 y por lo tanto el Corolario 4.1.

En primer lugar, observamos que los procesos no controlados  $x_{n+1} = (x_n + \theta\eta_n - \xi_n)^+$ , y  $\tilde{x}_{n+1} = (\tilde{x}_n + \theta\tilde{\eta}_n - \tilde{\xi}_n)^+$  cumplen

$$(x_n + f(x_n)\eta_n - \xi_n)^+ \leq (x_n + \theta\eta_n - \xi_n)^+ \quad (4.22)$$

$$(\tilde{x}_n + f(\tilde{x}_n)\tilde{\eta}_n - \tilde{\xi}_n)^+ \leq (\tilde{x}_n + \theta\tilde{\eta}_n - \tilde{\xi}_n)^+, \quad (4.23)$$

para toda política  $f \in \mathbb{F}$  (debido a la forma del conjunto de acciones  $A$ ), y de la Suposición 4.5 (a) podemos asegurar que son caminatas aleatorias irreducibles, para las que existen medidas de probabilidad invariantes  $m_x^\theta$  y  $m_{\tilde{x}}^\theta$  (vea [29] ejemplo 5.2) y son ordenados [24]. Con lo cual se cumple la suposición 4.2.

En la demostración del Corolario 2 de [26] se prueba que si se cumple la condición b) de la suposición 4.5 entonces  $E_0(r^{\tau_0}) < \infty$  si  $r \leq \Phi(r_0)^{-1}$  y si se cumple la condición c), entonces  $E_0(r^{\tilde{\tau}_0}) < \infty$ , cuando  $r \leq \tilde{\Phi}(\tilde{r}_0)^{-1}$  por lo tanto, podemos aplicar el Lema 4.2 a ambos procesos para asegurar que la

funciones  $v_1(x) = \begin{cases} G_x(r) & \text{si } x > 0 \\ 1 & \text{si } x = 0 \end{cases}$  y  $v_2(x) = \begin{cases} \tilde{G}_x(r) & \text{si } x > 0 \\ 1 & \text{si } x = 0 \end{cases}$  cumplen

$$Pv_1(x) \leq \lambda_1 v_1(x) + b_1 \mathbb{I}_{\{0\}}$$

$$\tilde{P}v_2(x) \leq \lambda_2 v_1(x) + b_2 \mathbb{I}_{\{0\}}$$

donde

$$\begin{aligned} \lambda_1 &= r_0^{-1}, \quad b_1 = r_0^{-1} [G_0(r) - 1] \\ \lambda_2 &= \tilde{r}_0^{-1}, \quad b_2 = \tilde{r}_0^{-1} [\tilde{G}_0(r) - 1]. \end{aligned}$$

Tomando  $\lambda = \max\{\lambda_1, \lambda_2\}$ ,  $b = \max\{b_1, b_2\}$  y  $W(x) = \max\{v_1, v_2\}$ , entonces tenemos

$$\begin{aligned} PW(x) &\leq \lambda W(x) + b \mathbb{I}_{\{0\}} \\ \tilde{P}W(x) &\leq \lambda W(x) + b \mathbb{I}_{\{0\}}, \end{aligned} \quad (4.24)$$

La suposición 3.1 a) y b) se cumplen debido a que la suposición 4.5 (d) también se cumple. La suposición 4.5 (d) y la hipótesis c) se siguen de que las variables aleatorias  $\eta, \xi, \tilde{\eta}, \tilde{\xi}$  tienen funciones de densidad acotada (vea [17]). Por lo tanto tenemos Suposición 4.4.

Aplicando el corolario (4.1) obtenemos la siguiente desigualdad para el índice de estabilidad

$$\Delta(x) \leq \frac{2}{1-\alpha} \left[ (2+4r) \left( W(x) + \frac{b}{1-\lambda} \right) \right] [\sigma(\mu_\xi, \mu_{\tilde{\xi}})]^{\frac{s-1}{s}}. \quad (4.25)$$

siempre que  $\sigma(\mu_\xi, \mu_{\tilde{\xi}}) \leq e^{-s/(s-1)}$ . donde  $\mu_\xi, \mu_{\tilde{\xi}}$  son las distribuciones de  $\xi$  y  $\tilde{\xi}$  respectivamente.

Para hacer una estimación de las constantes necesitamos calcular  $\Phi(r)$ ,  $r_0, \tilde{\Phi}(r), \tilde{r}_0$ , (mencionadas en la Suposición 4.5).

Realizaremos estos cálculos para cuando  $\eta, \xi, \tilde{\eta}, \tilde{\xi}$  tienen distribución exponencial con parámetros  $w_1, w_2, \tilde{w}_1, \tilde{w}_2$ , respectivamente.

Por la Suposición 4.5 (a) tenemos que:

$$E[\theta\eta - \xi] = \frac{\theta}{w_1} - \frac{1}{w_2} < 0 \quad \text{y} \quad E[\theta\tilde{\eta} - \tilde{\xi}] = \frac{\theta}{\tilde{w}_1} - \frac{1}{\tilde{w}_2} < 0.$$

**Solución 1** La función generadora de momentos,  $M_X(t) = E[e^{Xt}]$  de una variable aleatoria  $X$ , con distribución exponencial con parámetro  $w$  es  $M_X(t) = \frac{w}{w-t}$ , por otro lado, dadas  $X$  y  $Y$  independientes con funciones generadoras de momentos  $M_X(t)$  y  $M_Y(t)$ , entonces  $Z = X + Y$  tiene función

generadora de momentos  $M_X(t)$   $M_Y(t)$  y dada  $\beta$  una constante, entonces  $\beta X$  tiene función generadora de momentos  $M_{\beta X}(t) = M_X(\beta t)$ . De estas propiedades obtenemos que  $\zeta = \theta\eta - \xi$  tiene función generadora de momentos  $M_\zeta(t) = E[e^{\zeta t}] = \left[ \frac{w_1}{w_1 - \theta t} \right] \left[ \frac{w_2}{w_2 + t} \right]$  así que;

$$\Phi(r) = E[r^x] = M_\zeta(\ln r) = \left[ \frac{w_1}{w_1 - \theta \ln r} \right] \left[ \frac{w_2}{w_2 + \ln r} \right]$$

y

$$\Phi'(r_0) = \left[ \frac{w_1}{w_1 - \theta \ln r} \right] \left[ \frac{-w_2}{(w_2 + \ln r)^2} \right] \frac{1}{r} + \left[ \frac{w_1}{(w_1 - \theta \ln r)^2} \right] \left[ \frac{w_2}{w_2 + \ln r} \right] \frac{\theta}{r}.$$

Si igualamos  $\Phi'(r_0)$  a 0 y despejamos, resulta que  $r_0 = \exp\left(\frac{1}{2}\left[\frac{w_1}{\theta} - w_2\right]\right)$ , de (4.3.3) obtenemos  $\frac{w_1}{\theta} - w_2 > 0$  luego  $r_0 > 1$ .

Si tomamos el proceso aproximante, procediendo de manera semejante obtenemos  $\tilde{\Phi}(r) = \left[ \frac{\tilde{w}_1}{\tilde{w}_1 - \theta \ln r} \right] \left[ \frac{\tilde{w}_2}{\tilde{w}_2 + \ln r} \right]$  y  $\tilde{r}_0 = \exp\left(\frac{1}{2}\left[\frac{\tilde{w}_1}{\theta} - \tilde{w}_2\right]\right) > 1$ .

De la demostración del Teorema 3 de [26] se puede concluir que:

$$E_x[r^{r_0}] \leq k_1 r_0^x, \text{ para algún } r < \Phi^{-1}(r_0),$$

donde  $k_1 = r[1 - \Phi^{-1}(r_0)]^{-1}$  y

$$E_x[r^{\tilde{r}_0}] \leq k_2 \tilde{r}_0^x, \text{ para algún } r < \tilde{\Phi}^{-1}(\tilde{r}_0).$$

Donde  $k_2 = r[1 - \tilde{\Phi}^{-1}(\tilde{r}_0)]^{-1}$ .

De todo esto se pueden hacer estimaciones de las constantes del lado derecho de la desigualdad (4.25).

## 4.4 Procesos ordenados que no alcanzan el mínimo

### 4.4.1 Cotas para modelos no controlados

En esta sección el espacio de estados será  $\mathbb{X} = (-\infty, \infty)$  ó  $(0, \infty)$ .

El Teorema 2.2 de la referencia [34] nos da tasas de convergencia a la distribución límite para cadenas estocásticamente ordenadas y lo presentamos sin demostración. Para que se cumpla dicho Teorema es necesario que se cumpla, una condición de conjunto pequeño y una condición de sesgo de Lyapunov -Foster, más explícitamente, haremos las siguientes suposiciones.

**Suposición 4.6** *Existe  $C \subseteq \mathbb{X}$  que satisface la condición de minorización*

$$P(\cdot|x) \geq \varepsilon \phi(\cdot), \quad x \in C \quad (4.26)$$

para algún  $\varepsilon > 0$  y una medida de probabilidad  $\phi(\cdot)$  en  $X$ , en este caso diremos que  $C$  es un conjunto pequeño.

**Suposición 4.7** *Existen una función  $V : \mathbb{X} \rightarrow \mathbb{R}$ ,  $V(x) < \infty$ , y constantes  $b < \infty$  y  $\lambda < 1$  que cumple con la siguiente desigualdad*

$$PV(x) \doteq \int_X P(dy|x) V(y) \leq \lambda V(x) + b \mathbb{1}_C(x), \quad x \in X \quad (4.27)$$

En adelante diremos que la función  $V$  es una función de sesgo geométrico, si cumple con la suposición (4.7).

**Teorema 4.4** *Suponga que  $\{x_n\}_{n=1}^{\infty}$  es un cadena estocásticamente ordenada con distribución límite  $m$  que satisface las suposiciones 4.6 y 4.7, para algún conjunto pequeño  $C$ , de la forma  $(-\infty, c]$ ,  $c \in \mathbb{X}$ , alguna función  $V$ , constantes  $\lambda$ ,  $b$  y con distribución inicial  $\mu = \delta_x$  (la delta de Dirac concentrada en  $x$ ). Entonces, tenemos*

$$\sup_{A \in \mathcal{B}(X)} |P_\mu(x_n \in A) - m_x(A)| \leq r(n, \varepsilon, \lambda, b, V, d) \quad (4.28)$$

para dar una expresión de  $r(n, \varepsilon, \lambda, b, V, d)$  necesitamos las siguientes constantes que se obtienen de las desigualdades (4.26) y (4.27)

$$\begin{aligned} d &= \sup_{x \in C} V(x) \\ J &= d + \lambda^{-1}(b - \varepsilon) \\ \eta &= \log\left(\frac{J}{1 - \varepsilon}\right) / \log \lambda^{-1}. \end{aligned} \quad (4.29)$$

Si  $X$  es una variable aleatoria cuya distribución es  $m$  definimos  $\tilde{\mu} \doteq P(\max(X, x) \in A)$  para toda  $A \in \mathbb{B}(X)$ , también definimos la siguiente constante:

$$\varpi = \frac{\log \tilde{\mu}(V)}{\log \lambda^{-1}} \leq \frac{\log \left( \frac{b}{1-\lambda} + V(x) \right)}{\log \lambda^{-1}}, \quad (4.30)$$

donde  $\tilde{\mu}(V) = \int_X V(y) \tilde{\mu}(dy)$ . La desigualdad se obtiene ya que  $m(V) \leq \frac{b}{1-\lambda}$ , como en [24] y [27].

Aquí se pueden identificar dos contextos diferentes :

a) Si  $J \geq 1$  entonces podemos tomar

$$\rho = (1 - \varepsilon)^{\eta-1} > \lambda \quad (4.31)$$

y entonces para  $n > \varpi$

$$r(n, \varepsilon, \lambda, b, V, d) \doteq [\tilde{\mu}(V)]^{\frac{\log \rho}{\log \lambda}} \frac{e\varepsilon(n - \varpi + \eta)}{\eta} \rho^n \quad (4.32)$$

b) Si  $J < 1$  entonces podemos tomar  $\rho = \lambda$  y para  $n > \varpi + \frac{[\eta(1-\varepsilon)]}{\lambda^n - (1-\varepsilon)}$

$$r(n, \varepsilon, \lambda, b, V, d) \doteq [\tilde{\mu}(V)] \frac{\varepsilon}{1-J} \rho^n. \quad (4.33)$$

Como en la sección 4.3 se pueden obtener tasas de convergencia a la distribución límite para cadenas no-ordenadas a partir de cadenas monótona (que no alcanza el mínimo), siguiendo los mismos argumentos que se usaron para demostrar el teorema 4.2 y donde el papel del 0 lo desempeña  $C$  obtenemos el siguiente resultado:

**Teorema 4.5** *Suponga que  $\{X_n\}$  es una cadena de Markov que cumple lo siguiente :*

a) *Existe una cadena estocásticamente ordenada por digamos,  $\{Y_n\}$  tal que*

$$X_n(\omega) \leq Y_n(\omega) \text{ para } n \geq 0 \text{ y } \omega \in \Omega.$$

- b) Las condiciones 4.26 y 4.27 se cumplen para  $\{X_n\}$  y  $\{Y_n\}$  con el mismo conjunto pequeño  $C$ , la misma función de sesgo  $V$  y las mismas constantes  $\lambda$  y  $b$
- c)  $\{X_n\}$  y  $\{Y_n\}$  tienen distribuciones límite  $m_x$  y  $m_y$  respectivamente, entonces,

$$\sup_{A \in \mathbb{B}(X)} |P_x(X_n \in A) - m_x(A)| \leq r(n, \varepsilon, \lambda, b, V, d) \quad (4.34)$$

Aquí  $r(n, \varepsilon, \lambda, b, V, d)$  es la misma que aparece en el Teorema 4.4

#### 4.4.2 Modelos controlados

Consideraremos un par de *modelos de control de Markov* a tiempo discreto ( $t = 0, 1, 2, \dots$ ) (vea capítulo 2)  $\mathcal{P} = (X, A, A(x), c, p)$  y  $\tilde{\mathcal{P}} = (X, A, A(x), c, \tilde{p})$  con espacio de estados  $\mathbb{X}$ .

Los únicos componentes diferente en estos modelos son las *probabilidades de transición*  $p(B|x, a)$  y  $\tilde{p}(B|x, a)$ ,  $(x, a) \in \mathbb{K}$   $B \in \mathbb{B}(X)$ .

Tenemos el equivalente al Lema 4.3 que es como sigue:

**Lema 4.4** Sea  $\mathcal{P} = (X, A, A(x), c, p)$  un *procesos de control de Markov* para el que existe  $\{Y_n\}$  una *cadena de Markov ordenada* tal que  $x_n^f(w) \leq Y_n(w)$  para toda  $n$  y para toda  $w \in \Omega$  (donde  $x_n^f$  es el proceso que se obtienen al aplicar la política  $f$  al procesos  $\mathcal{P}$ ) donde  $\{Y_n\}$  tiene *distribución límite*  $m_y$  y cumple 4.26 y 4.27 entonces, podemos asegurar que para cualquier política  $f \in \mathbb{F}$  el procesos  $\{x_n^f\}$  tienen *distribución límite* que llamaremos  $q_f$  y

$$\sup_{A \in \mathbb{B}(X)} |P_x(x_n^f \in A) - q_f(A)| \leq r(n, \varepsilon, \lambda, b, V, d) \quad \text{para } f \in \mathbb{F},$$

aquí  $r(n, \varepsilon, \lambda, b, V, d)$  es la misma que aparece en el teorema 4.4 y tiene las mismas restricciones

**Dem..** aplicando el teorema 4.5 al proceso  $\{x_i^f\}$  el resultado es inmediato. ■

**Teorema 4.6** Sean dos procesos de control de Markov  $\mathcal{P} = (X, A, A(x), c, p)$  y  $\tilde{\mathcal{P}} = (X, A, A(x), c, \tilde{p})$  con  $\mathbb{X}$ , totalmente ordenado para los cuales se cumplen las hipótesis del lema 4.3 ie) se cumplen las siguientes hipótesis :

- a) existen cadenas de Markov  $\{Y_n\}$  y  $\{\tilde{Y}_n\}$  irreducibles para los que:
- i) se cumplen las condiciones 4.26 y 4.27 para  $\{X_n\}$   $\{\tilde{X}_n\}$  para un conjunto pequeño de la forma  $C = (0, c]$  o  $C = (-\infty, c]$
  - ii) se cumplen las condiciones 4.26 y 4.27 para  $\{Y_n\}$   $\{\tilde{Y}_n\}$  con el mismo conjunto pequeño  $C$  y la misma función de sesgo  $V$ .
  - iii) si  $n \in \mathbb{Z}$ ,  $w \in \Omega$ ,  $f \in \mathbb{F} \Rightarrow x_n^f(w) \leq Y_n(w)$  y  $\tilde{x}_n^f(w) \leq \tilde{Y}_n(w)$
- b) las suposiciones 3.1 se cumplen para  $\mathcal{P}$  y  $\tilde{\mathcal{P}}$  con  $W = V$ ,

entonces, existen  $f^*$  y  $\tilde{f}^* \in \mathbb{F}$  políticas óptimas para  $\mathcal{P}$  y  $\tilde{\mathcal{P}}$  respectivamente, además obtenemos la siguiente desigualdad para el índice de estabilidad

$$\begin{aligned} \Delta(x) = V_\alpha(x, \tilde{f}^*) - V_\alpha(x, f^*) \leq \\ \frac{2}{1-\alpha} \{2[V(x) + b/(1-\lambda)] + 4g_x(r)r^{-1} + 1\} \\ \times \delta^{\frac{(1-s)}{s}} \max\{1, [\log_r \delta]\} \end{aligned} \quad (4.35)$$

Aquí  $s > 1$ ,  $\delta = \sup_{(x,a) \in \mathbb{K}} \sigma(p(\cdot | x, a), \tilde{p}(\cdot | x, a))$ , donde se pueden identificar dos contextos

I) Si  $J \geq 1$  entonces podemos tomar  $g_x(r) = [\tilde{\mu}(V)]^{\frac{\log \rho}{\log \lambda} \frac{\epsilon \epsilon(n-\varpi+\eta)}{\eta}}$  con  $n > \varpi$  y  $r = (1-\epsilon)^{\eta-1} > \lambda$ .

II) Si  $J < 1$  entonces podemos tomar  $g_x(r) = [\tilde{\mu}(V)]^{\frac{\epsilon}{1-J}}$  con  $n > \varpi + \frac{[\eta(1-\epsilon)]}{\lambda^n - (1-\epsilon)}$  y  $r = \lambda$ .

Tanto en I) como en II)  $e$  es la base de los logaritmos neperianos

Dem.. De Lema 4.4 y el Teorema 4.4 y se sigue que,

$$\begin{aligned} \sup_{A \in \mathbb{B}(X)} |P_x(x_n^f \in A) - q_f(A)| \leq g_x(r) r^n \text{ para } f \in \mathbb{F} \\ \sup_{A \in \mathbb{B}(X)} |\tilde{P}_x(\tilde{x}_n^f \in A) - \tilde{q}_f(A)| \leq g_x(r) r^n \text{ para } f \in \mathbb{F} \end{aligned} \quad (4.36)$$

de las suposiciones 3.1 del capítulo 3, la desigualdades 4.27 y 4.36 se deduce la existencia de las políticas óptimas  $f^*$  y  $\tilde{f}^* \in F$  (vea el Teorema 8.2 de [18]). También obtenemos

$$\sup_{t \geq 0} E_x^f W(\tilde{x}_t) \leq C_0,$$

donde  $C_0 = V(x) + b/(1 - \lambda)$  (vea la demostración de la desigualdad 3.18) substituyendo  $C$  por  $C_0$  y  $\bar{B}W(x)$  por  $g_x(r)$  en (3.21) y (3.22) del capítulo 3. Se puede seguir exactamente los mismos argumentos usados en la demostración del teorema 3.1 para obtener el resultado. ■

También obtenemos el equivalente del Corolario 3.2

**Corolario 4.2** Sean los modelos  $\mathcal{P}$  y  $\tilde{\mathcal{P}}$  que cumplen las hipótesis del teorema 4.6 y que están definidos por las ecuaciones recurrentes

$$\begin{aligned} x_{t+1} &= F(x_t, a_t, \xi_t) \\ \tilde{x}_{t+1} &= F(\tilde{x}_t, \tilde{a}_t, \tilde{\xi}_t) \end{aligned} \quad t = 0, 1, \dots$$

entonces, obtenemos la siguiente desigualdad para el índice de estabilidad

$$\Delta(x) \leq \frac{2}{1 - \alpha} \{2V(x) + b/(1 - \lambda) + 4g_x(r)r^{-1} + 1\} [\sigma(\mu_\xi, \mu_{\tilde{\xi}})]^{\frac{\alpha-1}{\alpha}}$$

Siempre que  $\sigma(\mu_\xi, \mu_{\tilde{\xi}}) \leq e^{-\frac{\alpha}{(\alpha-1)}}$  donde  $\mu_\xi, \mu_{\tilde{\xi}}$  son las distribuciones de  $\xi$  y  $\tilde{\xi}$  respectivamente

### 4.4.3 Ejemplo

Construiremos un ejemplo de dos procesos de control de Markov que cumplen con las hipótesis del Teorema 4.6 y daremos una estimación de el índice de estabilidad.

**Ejemplo 4.3** Sea  $\mathbb{X} = (0, \infty)$  y  $A = A(x) = [\delta, 1]$  donde  $\delta > 0$  defina

$$x_{n+1} = \xi_n a_n x_n^{\alpha_1} \tag{4.37}$$

$$\tilde{x}_{n+1} = \tilde{\xi}_n a_n \tilde{x}_n^{\alpha_1} \quad (4.38)$$

$n = 0, 1, \dots$  y  $0 < \alpha_1 < 1$  y que cumplen con lo siguiente:

**Suposición 4.8**

- a)  $x_0 = \tilde{x}_0 = x$
- b)  $\{\xi_n\}$  y  $\{\tilde{\xi}_n\}$  son sucesiones de variables aleatorias i.i.d que toman valores en  $(0, \infty)$
- c) Si  $\xi$  y  $\tilde{\xi}$  representan variables genéricas de las sucesiones  $\{\xi_n\}$  y  $\{\tilde{\xi}_n\}$  respectivamente, entonces, se cumple lo siguiente
  - i)  $E[\ln |\xi|] < \infty$ ,  $E[\ln \xi] < 0$
  - ii)  $E[\ln |\tilde{\xi}|] < \infty$ ,  $E[\ln \tilde{\xi}] < 0$
  - iii)  $\xi$  y  $\tilde{\xi}$  tienen funciones de densidad continuas y acotadas en  $(0, \infty)$ .

Los procesos

$$\begin{aligned} y_{n+1} &= \xi_n y_n^{\alpha_1}, y \\ \tilde{y}_{n+1} &= \tilde{\xi}_n \tilde{y}_n^{\alpha_1} \end{aligned}$$

que se obtienen cuando tomamos la política  $f(x) = 1$  para toda  $x \in \mathbb{X}$ , aparece en [37], en esta referencia se comenta que es un proceso estocásticamente ordenado por ser  $\alpha < 1$ , y se calcula la tasa de convergencia a la distribución límite para una modificación del espacio. Nosotros usaremos los mismos argumentos ([37]) para probar que este proceso cumple (4.26) y (4.27) para todo el espacio  $(0, \infty)$ .

**Suposición 4.9** La función de costo es una función medible  $c(x, a)$  que cumple las condiciones a) y b) de las suposiciones 3.1 con  $W(x) = \ln(x + e)$ .

$$\begin{aligned}
P(W(x)) &= P(\ln(x+e)) = \int_{\mathcal{X}} \ln(y+e) p(dy|x) \\
&= \int_0^\infty \ln(sx^{\alpha_1} + e) f(s) ds = \int_0^\infty \ln \left[ (sx^{\alpha_1} + e) \frac{(x+e)^{\alpha_1}}{(x+e)^{\alpha_1}} \right] f(s) ds \\
&= \alpha_1 \ln(x+e) + \int_0^\infty \ln \left( \frac{(sx)^{\alpha_1}}{(x+e)^{\alpha_1}} + \frac{(e)^{\alpha_1}}{(x+e)^{\alpha_1}} \right) f(s) ds,
\end{aligned} \tag{4.39}$$

sea  $\xi$  la variable genérica de la sucesión  $\{\xi_n\}$  y como tiene función de densidad entonces,

$$E[\ln(\xi + d)] \rightarrow E[\ln(\xi)] < 0 \tag{4.40}$$

cuando  $d \rightarrow 0$ ; entonces podemos encontrar  $d_0 > 0$  tal que:

$$E[\ln(\xi + d_0)] < 0,$$

ahora escogemos  $c \in \mathbb{R}^+$  tal que si  $x > c$  implica  $\frac{e}{(x+e)^{\alpha_1}} < d_0$ , así que tomemos  $C = (0, c]$

Si tomamos  $\lambda_C = \alpha_1$ , para  $x > c$  tendríamos

$$P(W(x)) \leq \alpha_1 W(x)$$

y para  $x \in (0, c]$

$$P(W(x)) \leq \alpha_1 W(x) + b$$

donde,

$$b = \int_0^\infty \ln(s + e^{1-\alpha_1}) f(s) ds < \infty.$$

Esto debido a que

$$\begin{aligned}
\left( \frac{(sx)^{\alpha_1}}{(x+e)^{\alpha_1}} + \frac{(e)^{\alpha_1}}{(x+e)^{\alpha_1}} \right) &= \left( \frac{(s)^{\alpha_1}}{\left(1 + \frac{e}{x}\right)^{\alpha_1}} + \frac{(e)^{\alpha_1}}{(x+e)^{\alpha_1}} \right) \\
&\leq \frac{(s)^{\alpha_1}}{(1)^{\alpha_1}} + \frac{(e)^{\alpha_1}}{(x+e)^{\alpha_1}} \\
&\leq s + \frac{(e)^{\alpha_1}}{(e)^{\alpha_1}}
\end{aligned}$$

de donde obtenemos 4.27.

Falta probar que  $C$  es un conjunto pequeño, para probar esto consideremos  $\mu$  la medida de Lebesgue en  $(0, c]$  y  $\phi(A) = \mu(A \cap (0, c])$   $A \in \mathbb{B}(\mathbb{R}^+)$ . Si definimos  $\varepsilon = \inf_{x \in C} f(x)$  (función de densidad acotada) tenemos,

$$P(A|x) = \int_A p(dy|x) = \int_A f\left(\frac{s}{x^{\alpha_1}}\right) ds > \varepsilon \phi(A).$$

Usando estos mismos argumentos se puede demostrar que el proceso  $\{x_n^f\}$  cumple con la condición de *minorización* (4.26) con el mismo conjunto pequeño  $C = (0, c]$ .

Por otro lado como el proceso aproximante 4.38 cumple la condición c) ii) de las suposiciones 4.8 debe existir un  $\tilde{d}_0$  para la que se cumpla

$$E \left[ \ln \left( \tilde{\xi} + \tilde{d}_0 \right) \right] < 0, \tag{4.41}$$

por lo que si tomamos  $d_m = \min \{d_0, \tilde{d}_0\}$ , entonces si escogemos  $c_1$  tal que si  $x > c_1$  implica  $\frac{e}{(x+e)^\alpha} < d_m$ , así que, se cumplen las desigualdades (4.26) y (4.27) para ambos procesos, podemos tomar  $C_1 = (0, c_1]$ , y el mismo  $b$ , por esta razón se cumple la hipótesis a) ii) del teorema 4.6; la condición a) iii) es inmediata de la definición de el conjunto de controles ( $A(x) = (0, 1]$ ),

La condición b) se sigue de que las  $\xi_n$  y  $\tilde{\xi}_n$  tienen funciones de densidad continuas y acotadas (vea [10]).

Así, tenemos que nuestro ejemplo cumple con las hipótesis del Teorema 4.6 y por lo tanto podemos usar el corolario 4.2.

Para obtener una estimación del índice de estabilidad  $\Delta_\alpha(x)$  necesitamos evaluar las constantes que aparecen en el corolario 4.2, para ello evaluamos lo siguientes:

I)  $d = \sup_{(0, c_1]} \{\ln(x+e)\} = \ln(c_1+e)$

II)  $J = \ln(c_1+e) + \alpha_1(b-\varepsilon)$

III)  $\eta = \frac{\ln\left(\frac{\ln(c_1+e) + \alpha_1(b-\varepsilon)}{1-e}\right)}{\ln \alpha_1^{-1}}$  (recuerde que  $\lambda = \alpha_1$ )

IV)  $\varpi \leq \frac{\log\left(\frac{b}{1-\alpha_1} + \log(x+e)\right)}{\log \alpha_1^{-1}}$  esto debido a que la distribución inicial  $\mu(A) = \delta_x(A)$  y la distribución límite cumple que  $m(W) \leq \frac{b}{1-\alpha_1}$ .

como  $J > 1$  se puede tomar  $r = (1 - \varepsilon)^{\eta^{-1}} > \alpha_1$ , para  $n > \varpi$ , de aquí obtenemos que la  $g_x(r)$  del corolario 4.2 se puede tomar como:

$$g_x(r) = \left[ \ln(x+e) + \frac{b}{1-\alpha_1} \right] \frac{\ln r}{\ln \alpha_1} \frac{[\varepsilon \varepsilon (n - \varpi + \eta)]}{\eta} \quad (4.42)$$

y  $C_0 = \left[ \ln(x+e) + \frac{\int_0^\infty \ln(s+e^{1-\alpha_1}) f(s) ds}{1-\alpha_1} \right]$ , así que tenemos la siguiente estimación del índice de estabilidad

$$\Delta(x) \leq \frac{2}{1-\alpha} \left\{ \begin{array}{l} 2 \left[ \ln(x+e) + \frac{\int_0^\infty \ln(s+e^{1-\alpha_1}) f(s) ds}{1-\alpha_1} \right] + \\ 4 \left[ \ln(x+e) + \frac{b}{1-\alpha_1} \right] \frac{\ln r}{\ln \alpha_1} \frac{[\varepsilon \varepsilon (n - \varpi + \eta)]}{\eta} r^{-1} + 1 \end{array} \right\} \\ \times [\sigma(\mu_\xi, \mu_{\tilde{\xi}})]^{\frac{s-1}{s}}$$

siempre que  $\sigma(\mu_\xi, \mu_{\tilde{\xi}}) \leq e^{-s/(s-1)}$ , donde  $\mu_\xi, \mu_{\tilde{\xi}}$  son las distribuciones de  $\xi$  y  $\tilde{\xi}$  respectivamente.

## Capítulo 5

### Conclusiones

En este trabajo hemos obtenido estimaciones de los índices de estabilidad definidos en (1.11) y en (1.12) para procesos de control de Markov con espacio de estado generales y función de costo de un paso posiblemente no acotada. Las desigualdades (2.16), (3.7), (4.16) y (4.35) aunque se obtienen para funciones de costo no acotada, se pide que la función de costo de un paso cumpla  $c(x, a) \leq w(x)$  para cierta función  $w(x) : X \rightarrow \mathbb{R}$  y  $a \in A(x)$ . Esta restricción parece ser una condición un poco restrictiva. Ya que se conocen casos de modelos usados en economía (por ejemplo) que no cumplen esta condición, sin embargo, el operador de programación dinámica tiene puntos fijos por lo que, se puede aplicar la misma técnica que usamos aquí, para encontrar una estimación del índice de estabilidad (1.11) al menos para el caso de la desigualdad (2.16) y así ampliar el marco de las aplicaciones.

La desigualdad (3.7) se obtiene bajo condiciones de ergodicidad bastante fuertes y las desigualdades (4.16) y (4.35) se obtienen bajo la hipótesis de que las cadenas obtenidas al aplicar una cierta política están acotadas por una cadena de Markov ordenada con buenas propiedades de convergencia a la distribución límite, aunque estas condiciones son restrictivas, encontramos que se cumplen para diversos procesos de control, algunos modelos de crecimiento económico, inventarios, presas, algunos modelos de control lineal con costo cuadrática, etc.

Por otro lado combinando (3.14) y (3.16) se obtiene una cota  $a$  para el índice de estabilidad descontado de la forma  $\Delta_\alpha(x) \leq M(\alpha)\psi(p, \bar{p})$  donde  $M(\alpha)$  es de orden  $(1 - \alpha)^{-1}$ . Por esta razón la estimación del índice de estabilidad relativo  $\bar{\Delta}_\alpha(x)$  (3.7) tiene buenas propiedades asintóticas cuando el factor de descuento tiende a 1. Esta característica nos permite obtener la

cota (3.7) que es uniforme respecto al factor de descuento.

También es importante que las cotas obtenidas se expresen en términos de la métrica de la variación total o una métrica ponderada, por eso en las desigualdades (2.16), (3.7), (4.16) y (4.35) las constantes que aparecen se expresen en términos de los modelos y las suposiciones correspondientes (compare con [40] y [41]).

Una extensión de este trabajo que podría ser importante, consistiría en establecer la estabilidad de nuestros procesos respecto a la métrica de Prokhorov.

Sobre esta posibilidad, en [33] motivados por problemas de precisión computacional, estudian la siguiente situación: dada  $p(\cdot|x)$  la ley de transición de una cadena de Markov general (no controlada) con espacio de estados  $\mathbb{X}$ , donde  $\mathbb{X}$  es métrico separable, con métrica  $d(\cdot, \cdot)$ . y  $h: \mathbb{X} \rightarrow \mathbb{X}$ , una función de redondeo, de tal manera que  $h(x)$  está cercana a  $x$  para cada  $x \in \mathbb{X}$ . Definimos un "redondeo de la cadena" de la manera siguiente:  $\tilde{p}(A|x) = p(h^{-1}(A)|x)$ . Por ejemplo cuando  $\mathbb{X} = \mathbb{R}^n$  y donde  $h$  es la función que redondea cada componente de  $x \in \mathbb{R}^n$  al múltiplo más cercano de  $\delta$ . Esto es  $h_i(x) = \delta \lceil \frac{x_i}{\delta} \rceil$ , donde  $\delta > 0$ . En [33] se demuestra que si  $\tilde{p}$  está cerca de  $p$  y se cumple una cierta condición de tipo Foster para alguna función  $V$  tal que,  $\log V$  es uniformemente continua; entonces  $\tilde{p}$  mantendrá las propiedades de aperiodicidad, ergodicidad geométrica, irreducibilidad y pequeñez de los conjuntos.

Por otro lado se pueden modificar las cotas de la distancia de la variación total entre la transición de  $k$ -pasos y la medida estacionaria, en términos de la métrica de Prokhorov usando condiciones de sesgo y minorización.

Creemos que estos resultados nos darán tasas de convergencia a la distribución límite en términos de la métrica de Prokhorov y usando los mismos argumentos de los capítulos 3 y 4 obtendremos expresiones numéricas para cotas superiores del índice relativo  $\Delta(x)$  y cotas uniformes en  $\alpha$  para el índice de estabilidad relativo  $\tilde{\Delta}(x)$ .

## Bibliografía

- [1] Abbad, J.A.Filar, *Perturbation and Stability theory for Markov control processes*, IEEE Trans. Automat.Control **37** (1992) 1415-1420.
- [2] D.P. Bertsekas, S.E.Shreve, *Stochastic Optimal Control: The Discrete Time case*, Academic Press, New York 1978.
- [3] T. Bielecki, L.Stettner *Ergodic Control of as Singularity perturbed Markov Process in Discrete Time with General State and Compact Action Spaces*, Applied Mathematics and Optimization(1998) **38**:261-281
- [4] E.B. Dynkin, A.A. Yuskevich, *Controlled Markov Processes*, Springer Verlag , New York, 1979.
- [5] I.I.Gihman, A.V. Skorohod, *Controlled Stochastic Processes*, Springer Verlag, New York, 1977
- [6] E.I. Gordienko, *An estimate of stability of optimal control of certain stochastic and deterministic systems*, J.Soviet. Math. **59** (1992) 891-899
- [7] E.I. Gordienko, *Lectures Notes on Stability Estimation in Markov Decision Processes*, Universidad Autónoma Metropolitana, México D.F.,1994.
- [8] E.I. Gordienko,(1988) , *Stability estimates of optimal control chains with minorant*. J. Soviet Math. **40** ,481-486.
- [9] E.I. Gordienko, O. Hernández-Lerma, *Average cost Markov control processes with weighted norms : existence of canonical policies*, appl. Math. **23** (1995) 199-218.

- [10] E.I. Gordienko, O. Hernández-Lerma, *Average cost Markov control processes with weighted norms : value iteration*, appl. Math. 23 (1995) 219-237.
- [11] E.I. Gordienko, F.S. Salem *Robustness inequality for Markov control processes with unbounded costs*, Syst. control Lett. 33(1998) 125-130.
- [12] E.I. Gordienko, F.S. Salem, *Estimates of Stability of Markov Control Process With Unbounded Costs*, Kybernetika-vol 36 (2000), Number 2, Pages 195-210
- [13] Gordienko, E.I., Isauro-Martínez, M.E. and Marcos Carrillo, R.M. *Estimation of stability in controlled storage systems*. Research report. 04.0405.I.01.001.97, Dep. de Matemáticas, Universidad Autónoma Metropolitana, México, 1997.
- [14] O. Hernández-Lerma, *Adaptive Markov Control Processes*, Springer, New York, 1989.
- [15] O. Hernández-Lerma, *Lectures Notes on Discrete-Time Markov Control Processes*, CINVESTAV mex,1990.
- [16] O. Hernández-Lerma, J.B. Lasserre, *Value iteration and rolling plans for Markov control processes with unbounded rewards*, Math.Anal. Appl.177 (1993) 38-55.
- [17] O. Hernández-Lerma, J.B. Lasserre, *Discrete-Time Markov Control Processes*, Springer, New York, 1996.
- [18] O. Hernández-Lerma, J.B. *Further Topics on Discrete-Time Markov Control Processes*, Springer, New York, 1999.
- [19] K. Hinderer, *Foundations of Non-Stationary Dynamic Programming with Discrete Time Parameter*, Lectures Notes, Oper. Res. No. 33 Springer, New York, 1970.
- [20] Kartashov, N.V. *Inequalities in theorems of ergodicity and stability for Markov chains with common phase space II*. Theory Probab. Appl. 30 (1985), 507-515.

- [21] Kumar, P.R. and Varaiya, P. *Stochastic Systems: Estimation, Identification and Adaptive control*. Prentice-Hall, Englewood Cliffs, N.J., 1986.
- [22] Lindvall, T. *Lectures on the Coupling Method*, Wiley interscience New York (1992).
- [23] S.A. Lippman, *On dynamics programming with unbounded rewards*, Manage. Sci. **21** (1975) 1225-1233.
- [24] Lund, R.B and R.L. Tweedie, R.L. *Geometric convergence rates for stochastically ordered Markov chains*. Math. Operations Res., **21**:182-194, 1996.
- [25] R.B.Lund (1994) *A dam with seasonal input*. J. Appl. Prob. **31** 526-541.
- [26] R.B.Lund *The Geometric Convergence Rate of a Lindley Random Walk* J. Appl. Prob. **34**, 806-811 (1997).
- [27] Meyn, S.P. and Tweedie, R.L. *Markov Chains and Stochastic Stability* Springer-Verlag (1993).
- [28] A. Muller, *How does the value function of a Markov decision process depend on the transition probabilities ?*, (1997) MOR **4** 872-884.
- [29] Nummelin, E. *General Irreducible Markov Chains and Non-Negative Operators*. Cambridge University Press, Cambridge, 1984.
- [30] Rachev, S.T. *Probability Metrics and the Stability of Stochastic Models*. John Wiley & Sons, New York, 1991.
- [31] U. Rieder, *Measurable selection theorems for optimization problems*, Manuscripta Math. **24** (1978) 115-131.
- [32] Roberts, G.O. and N.G. Polson (1994). *A note on geometric convergence rates of the Gibbs sampler, to appear*, the J. Royal Statistical Society, series B.
- [33] Roberts, G.O., Rosenthal, J.S. and P. O. Schwartz *Convergence properties of perturbed Markov chains* J. Appl. Prob. **35-1**(1998) 1-11

- [34] Roberts, G.O. Tweedie, R.L. (1998) *Rates of convergence of stochastically monotone and continuous time models* preprint.
- [35] Roberts, G.O. Tweedie, R.L. (1999) *Bounds on regeneration times and convergence rates for Markov chains*. *Stoch. Proc. Appl.* **80**, 211-229.
- [36] Rosenthal, J.S. (1995) *Minorization conditions and convergence rates for Markov Chain Montecarlo*. *J. Amer. Stat. Assoc.* **90**, 558-566
- [37] Scott, D.J. and Tweedie, R.L. *Explicit rates of convergence of stochastically ordered Markov chains*, *Proc. Athens Conference of Applied Probability and Time Series Analysis: Papers in Honour of J.M. Gani and E.J. Hannan*. Edd. C.C. Heyde, Yu. V. Prohorov, R. Pyke and S.T. Rachev, Springer-Verlag, New York, (1995), 176-191.
- [38] Stoyan, D. (1983) *Comparison Methods for Queues and other Stochastic Models*, John Wiley and sons, New York (1995), 176-191.
- [39] N. Torres *El método de regeneración para la estabilidad de procesos controlables con costo descontado* tesis de maestría, UAMI (1998).
- [40] N.M. Van Dijk, *Perturbation theory for unbounded Markov rewards processes With applications to queueing*, *Adv. Appl. Probab.* **20** (1988) 99-111.
- [41] N.M. Van Dijk, M. L Puterman, *Perturbation theory for Markov reward processes with applications to queueing systems*, *Adv. Appl. Probab.* **20** (1998) 79-98.
- [42] J.A.E.E. Van Numen, J. Wessels, *A note on dynamics programming with unbounded rewards*, *Manage. Sci.* **24** (1978) 576-580.
- [43] Weber, R.R. and Stidham, S, jr. *Optimal control of service rates in networks of queues*. *Adv. Appl. Prob.* **19** (1987), 202-218.
- [44] J. Wessels, *Markov Programming by successive approximations with respect to weight supremum norms*, *J. Math. Anal. Appl.* **58** (1977) 326-335.
- [45] W. Whitt, *Approximations of dynamic programs*, *J. Math. Oper. Res.* **3** (1978) 231-243.

- [46] Zolotarev, V.M. *On stochastic continuity of queueing systems of type  $G|G|1$* . *Theory Probab. Appl.* **21** (1976), 250-269.

