Procesos de Control Semi-Markovianos con Distribución del Tiempo de Permanencia Desconocida

Luz del Carmen Rosas Rosas
Maestro en Ciencias (Matemáticas)

Programa de Posgrado en Matemáticas
Departamento de Matemáticas,
Universidad de Sonora.

Octubre de 2006.

Índice general

Introducción

El estudio de los procesos de control estocástico se clasifica como procesos en tiempo discreto o en tiempo continuo. Una clase importante de procesos de control en tiempo continuo la constituyen los Procesos de Control Semi-Markovianos (*PCSMs*), cuya evolución en el tiempo la podemos describir de la siguiente manera. Si en el tiempo de la n-ésima época de decisión el sistema se encuentra en el estado $x_n = x$, entonces el controlador elige una acción o control $a_n = a$ y sucede lo siguiente: 1) el sistema permanece en el estado x durante un tiempo aleatorio no negativo δ_{n+1} con distribución H; 2) se genera un costo c que depende del estado, el control y el tiempo de permanencia, 3) el sistema se mueve a un nuevo estado $x_{n+1} = y$ de acuerdo a una ley de transición. Una vez ocurrido lo anterior, el proceso se repite.

Los costos de operación se acumulan durante la evolución del sistema, y el objetivo es encontrar una política de control que minimice un índice de funcionamiento.

En muchos trabajos, usualmente, se supone que todas las componentes que intervienen en la descripción de los *PCSMs* son conocidas por el controlador; no obstante, existen situaciones en las cuales solamente se cuenta con estimaciones o aproximaciones de uno o más elementos del proceso.

Este trabajo de tesis se enfocará al estudio de los *PCSMs* con el criterio de costo total descontado y distribución del tiempo de permanencia H desconocida por el controlador. Además supondremos que los espacios de estado y control son de Borel y el costo por etapa es posiblemente no acotado.

El esquema previamente descrito produce algunas consecuencias en la evolución de un *PCSM* estándar. En principio, la función de costo por etapa c será también desconocida debido a que depende de los tiempos de permanencia. Por lo tanto, bajo este contexto, supondremos que antes de que el controlador seleccione el control a_n en el tiempo de la n-ésima época de decisión, se debe implementar un proceso de estimación de H, y por consiguiente del costo c, con el cual obtendremos los estimadores H_n y c_n. Luego, combinando el proceso de estimación con técnicas de control deberá entonces elegirse una acción $a = a_n (H_n)$.

Los *PCSMs* bajo el criterio de costo descontado, han sido estudiados por varios autores. Por ejemplo [12, 13, 23, 25] consideran el caso en que el espacio de estados es un conjunto numerable, mientras que en [2, 18] se considera espacio de estados de Borel. No obstante, tanto en los primeros como en los últimos trabajos aquí citados, todas las componentes que describen a los modelos en cuestión se suponen conocidas por el controlador, lo cual contrasta con la característica principal del presente trabajo que es suponer desconocida la distribución del tiempo de permanencia H, tal como es considerado por los autores en [17].

Este trabajo de tesis está estructurado en cinco capítulos diseñados de la siguiente manera.

En el Capítulo 1 se dan los elementos para definir los *PCSMs* así como el problema de control óptimo asociado. Además se introducen las hipótesis sobre el modelo en general.

En el Capítulo 2 se estudian los *PCSMs* estándar (todas sus componentes conocidas) bajo el criterio de costo descontado. Se establecen resultados sobre la existencia de políticas óptimas, así como una caracterización del costo óptimo. También se introduce el concepto de optimalidad asintótica, mismo que toma importancia al analizar la optimalidad de las políticas resultantes de la combinación de estimación y control antes mencionada.

El Capítulo 3 contiene la parte central de este trabajo. Aquí se estudian los *PCSMs* con criterio de costo descontado, cuya distribución del tiempo de permanencia H es deconocida. De hecho, se introduce un nuevo conjunto de hipótesis relacionadas con la distribución H, las cuales serán la base para establecer el método de estimación. Se construyen políticas asintóticamente óptimas, y se realiza un análisis del comportamiento límite de cierta clase de minimizadores.

Con la idea de ilustrar la teoría desarrollada en este trabajo, en el Capítulo 4 se incluye un ejemplo de un sistema de almacenamiento.

Finalmente, en el Capítulo 5 se presentan las conclusiones del trabajo, y además, se exponen algunas ideas sobre posibles extensiones de la teoría aquí desarrollada.

Capítulo 1

Procesos de Control Semi-Markovianos con Costos Descontados

1.1. Introducción

En este capítulo se introduce el problema de control óptimo semi-markoviano (PCO) con respecto al índice en costo descontado. Para la formulación del PCO se requieren tres elementos: un modelo de control, un conjunto de políticas admisibles y un índice de funcionamiento mediante el cual se evaluará el desempeño del sistema cuando se usan diferentes políticas. Lo anterior se desarrolla en las secciones 1.2, 1.3 y 1.4.

Por otra parte, para resolver el PCO como cualquiera otro problema de optimización, se requiere que el modelo de control satisfaga un conjunto de condiciones que se introducen en la sección 1.5. Las condiciones que consideraremos en el presente trabajo son de tres tipos: condiciones de continuidad y compacidad, que en términos generales permiten garantizar la existencia de minimizadores; condiciones sobre el crecimiento de las funciones de costo asociadas al problema de control. De hecho, estas condiciones nos permitirán analizar el PCO con costos no-acotados. Y una tercera clase de condición que es de tipo "probabilístico", la cual garantiza que los procesos semi-markovianos controlados son regulares, es decir, que experimentan un número finito de transiciones en intervalos finitos de tiempo. Además, esta misma condición garantiza que los "factores de descuento" son uniformemente acotados por una constante estrictamente menor que uno.

Notación. A lo largo de este capítulo asumiremos lo siguiente. Respectivamente, $\mathbb{N}$ y $\mathbb{N}_0$ representarán el conjunto de los números enteros positivos y el de los enteros no-negativos. Por otra parte, P_x^π denotará la medida de probabilidad cuando se usa la política π dado que el estado inicial es x; mientras que E_x^π indicará el operador esperanza con respecto a P_x^π. Recuérdese además que un *espacio de Borel* es un subconjunto de Borel de un espacio métrico, separable y completo.

1.2. Modelo de Control Semi-Markoviano

1.2.1. Descripción del Modelo

Definición 1.1 *Un modelo de control semi-Markoviano (MCSM)*

$$(\mathbb{X}, \mathbb{A}, \{A(x) : x \in \mathbb{X}\}, Q, H, D, d) \tag{1.1}$$

consiste en lo siguiente:

- $\mathbb{X}$ *es un espacio de Borel llamado espacio de estados*;
- $\mathbb{A}$ *es un espacio de Borel llamado espacio de controles o acciones*;
- *Para cada* $x \in \mathbb{X}$, $A(x) \subset \mathbb{A}$ *es un conjunto medible y no vacío, cuyos elementos representan las acciones admisibles cuando el sistema se encuentra en el estado* x.

Supondremos que el conjunto $\mathbb{K} := \{(x, a) : x \in \mathbb{X},\ a \in A(x)\}$ *de pares estado-acción admisibles, es un subconjunto de Borel del espacio* $\mathbb{X} \times \mathbb{A}$. *También suponemos que* $\mathbb{K}$ *contiene la gráfica de una función medible* $f : \mathbb{X} \longrightarrow \mathbb{A}$, *de manera que* $f(x) \in A(x)$, *para todo* $x \in \mathbb{X}$.

- *La ley de transición* $Q(\cdot \mid \cdot)$ *es un kernel estocástico (véase Apéndice B, Definición B.1) sobre* $\mathbb{X}$ *dado* $\mathbb{K}$.
- $H(\cdot \mid x, a)$ *es la función de distribución para cada* $(x, a) \in \mathbb{K}$, *y suponemos que es medible en las variables* $(x, a) \in \mathbb{K}$.
- D *y* d *son las funciones de costo, las cuales son reales medibles no negativas sobre* $\mathbb{K}$.

1.2.2. Interpretación

Un *MCSM* representa un sistema dinámico que evoluciona de la siguiente manera. En el tiempo de la n-ésima época de decisión T_n, el sistema se encuentra en el estado $x_n = x$ y el controlador elige una acción $a_n = a \in A(x)$, generándose con ello lo que se describe a continuación:

1. Se incurre en un costo inmediato $D(x, a)$.
2. El sistema permanece en dicho estado $x_n = x$ durante un tiempo aleatorio no-negativo δ_{n+1} con distribución $H(\cdot \mid x, a)$.
3. Luego, en el tiempo $T_{n+1} := T_n + \delta_{n+1}$ ($n \in \mathbb{N}, T_0 := 0$), el sistema transita a un nuevo estado $x_{n+1} = y$ de acuerdo a la distribución $Q(\cdot \mid x, a)$.

4. Se produce un costo debido al tiempo de permanencia en el estado x, cuya razón de costo es $d(x, a)$.

5. Finalmente, una vez en el estado y el proceso se repite.

1.3. Políticas de Control Admisibles

Definición 1.2 *Dado un MCSM, se define el espacio de historias admisibles hasta la n-ésima época de decisión mediante $\mathbb{H}_0 := \mathbb{X}$ y $\mathbb{H}_n := (\mathbb{K} \times \mathbb{R}_+)^n \times \mathbb{X}$ para $n \in \mathbb{N}$.*

De lo anterior se desprende que un elemento $h_n \in \mathbb{H}_n$, llamado en lo sucesivo n-historia, es un vector de la forma

$$h_n := (x_0, a_0, \delta_1, x_1, a_1, \delta_2, ..., x_{n-1}, a_{n-1}, \delta_n, x_n),$$

donde $(x_k, a_k, \delta_{k+1}) \in \mathbb{K} \times \mathbb{R}_+$ para $k = 0, 1, ..., n-1$, y $x_n \in \mathbb{X}$.

Definición 1.3 *Una política de control admisible o simplemente una política es una sucesión $\pi := \{\pi_n\}$, donde cada π_n es un kernel estocástico sobre $\mathbb{A}$ dado $\mathbb{H}_n$, tales que satisfacen la restricción $\pi_n(A(x_n) \mid h_n) = 1$ para todo $h_n \in \mathbb{H}_n$ y $n \in \mathbb{N}_0$.*

Se denotará por Π al conjunto de todas las políticas.

Sea $\mathbb{F}$ el conjunto de funciones medibles $f : \mathbb{X} \to A$ tal que $f(x) \in A(x)$ para todo $x \in \mathbb{X}$.

Definición 1.4 *Diremos que una política $\pi := \{\pi_n\} \in \Pi$ es:*

(a) Determinística si existe una sucesión $\{g_n\}$ de funciones medibles $g_n : \mathbb{H}_n \to \mathbb{A}$ tales que $\pi_n(\cdot \mid h_n)$ está concentrada en $g_n(h_n)$ para todo $h_n \in \mathbb{H}_n$ y $n \in \mathbb{N}_0$; es decir, $\pi_n(B \mid h_n) := 1_B[g_n(h_n)]$, $B \in \mathfrak{B}(\mathbb{A})$.

(b) Markoviana si existe una sucesión $\{f_n\}$ de funciones $f_n \in \mathbb{F}$ tales que $\pi_n(\cdot \mid h_n)$ está concentrada en $f_n(x_n) \in A(x_n)$ para todo $h_n \in \mathbb{H}_n$ y $n \in \mathbb{N}_0$.

(c) Estacionaria si existe una función $f \in \mathbb{F}$ tal que $\pi_n(\cdot \mid h_n)$ está concentrada en $f(x_n) \in A(x_n)$ para todo $h_n \in \mathbb{H}_n$ y $n \in \mathbb{N}_0$.

Debido a la definición previa, sin pérdida de generalidad, en adelante identificaremos al conjunto de todas las políticas estacionarias con $\mathbb{F}$.

Sea $(\Omega, \mathcal{A})$ el espacio medible canónico en el cual Ω es el espacio producto $(\mathbb{K} \times \mathbb{R}_+)^\infty$ y $\mathcal{A}$ la σ-álgebra producto correspondiente. Puede observarse que un elemento $\omega \in \Omega$ tendrá la forma $\omega := (x_0, a_0, \delta_1, x_1, a_1, \delta_2, ...)$. A las variables coordenadas $x_k \in \mathbb{X}$, $a_k \in A(x_k)$ y $\delta_{k+1} \in \mathbb{R}_+$, se les llamará variables de estado, control y tiempo de transición, respectivamente. De acuerdo al Teorema de C. Ionescu Tulcea (véase Apéndice C), para cada estado inicial $x \in \mathbb{X}$ y cada $\pi \in \Pi$, existe una medida de probabilidad P_x^π tal que para cada $B \in \mathfrak{B}(\mathbb{A})$, $C \in \mathfrak{B}(\mathbb{X})$ y $h_n \in \mathbb{H}_n$, $n \in \mathbb{N}_0$, se tiene:

$$P_x^\pi(x_0 = x) = 1, \tag{1.2}$$

$$P_x^\pi(a_n \in B \mid h_n) = \pi_n(B \mid h_n), \tag{1.3}$$

$$P_x^\pi(x_{n+1} \in C \mid h_n, a_n, \delta_{n+1}) = Q(C \mid x_n, a_n) \tag{1.4}$$

$$P_x^\pi(\delta_{n+1} \leq t \mid h_n, a_n) = H(t \mid x_n, a_n). \tag{1.5}$$

Observación 1.1 *De (1.5) se tiene que las variables aleatorias $\delta_1, \delta_2, ...,$ son condicionalmente independientes dado el proceso $x_0, a_0, x_1, a_1, ...$*

Para una política arbitraria $\pi \in \Pi$, la variable x_n describe el estado del sistema en el tiempo de la n-ésima transición (o época de decisión) T_n y a_n representa el control elegido de acuerdo a la política π. Nótese que en general, el estado x_n depende de la evolución del sistema en las primeras $n-1$ transiciones, no obstante en el caso de una política estacionaria f, $\{x_n\}_{n=0}^\infty$ es una cadena de Markov con probabilidad de transición $Q(\cdot \mid x, f(x))$, lo cual es una consecuencia de las propiedades de la esperanza condicional, así como de (1.4) que en lo sucesivo se identificará como la *propiedad de Markov*.

De aquí en adelante E_x^π denotará al operador esperanza correspondiente a P_x^π. Además, para una política estacionaria $f \in \mathbb{F}$, se usará la notación (x, f) en lugar de $(x, f(x))$.

1.4. Índice de Funcionamiento y Problema de Control Óptimo en Costo Descontado

Iniciaremos la presente sección señalando que en lo sucesivo, para un *MCSM* dado, estaremos considerando los costos descontados en forma continua. Esto es, para un factor de descuento $\alpha > 0$ un costo k generado al tiempo t equivale a un costo $ke^{-\alpha t}$ en tiempo 0, de manera que de acuerdo a la interpretación del *MCSM*, la función de costo por etapa, cuando el proceso se encuentra en el estado x y se elige la acción a, tomará la forma

$$c(x,a) := D(x,a) + d(x,a) \int_0^\infty \int_0^t e^{-\alpha s} ds H(dt \mid x,a). \tag{1.6}$$

Con el fin de medir el comportamiento de los costos por etapa usaremos como índice de funcionamiento el costo total esperado α-descontado.

Definición 1.5 *Sean $\pi \in \Pi$, $x_0 = x \in \mathbb{X}$ y $\alpha > 0$. Se define el costo total esperado α-descontado como*

$$V(\pi, x) := E_x^\pi \left[\sum_{n=0}^{\infty} e^{-\alpha T_n} c(x_n, a_n) \right],$$

donde $c(\cdot, \cdot)$ es la función introducida en (1.6).

Por lo tanto, dado el *MCSM* (1.1), el *problema de control óptimo* consiste en encontrar una política $\pi^* \in \Pi$ tal que

$$V(\pi^*, x) = \inf_{\pi \in \Pi} V(\pi, x) \qquad \forall x \in \mathbb{X}.$$

A la función

$$V_\alpha^*(x) := \inf_{\pi \in \Pi} V(\pi, x) \qquad x \in \mathbb{X}, \tag{1.7}$$

le llamaremos *función de valor óptimo*, y en este caso decimos que la política π^* es *óptima*.

1.5. Hipótesis sobre el Modelo de Control

En esta sección introduciremos tres clases de condiciones, las cuales serán usadas a lo largo del trabajo.

Hipótesis 1.1 *Condiciones de continuidad y compacidad.*

(a) Para cada $x \in \mathbb{X}$, $A(x)$ es un conjunto compacto.

(b) Para cada $x \in \mathbb{X}$, las funciones de costo $D(x, a)$ y $d(x, a)$ son reales medibles no-negativas y semi-continuas inferiormente (s.c.i.) en $a \in A(x)$ (véase Apéndice A, Definición A.2(a)).

(c) La ley de transición $Q(\cdot \mid x, a)$ es fuertemente continua en $a \in A(x)$, es decir, para cada función medible y acotada v sobre $\mathbb{X}$, se tiene que

$$a \mapsto \int_{\mathbb{X}} v(y) Q(dy \mid x, a)$$

es una función continua y acotada sobre $A(x)$.

(d) La función $H(t \mid x, a)$ es continua en $a \in A(x)$ para cada $x \in \mathbb{X}$ y $t \in \mathbb{R}$.

A continuación presentamos algunas hipótesis de crecimiento. Sea $W : \mathbb{X} \to [1, \infty)$ una función medible. Denotemos por $\mathbb{B}_W(\mathbb{X})$ el espacio lineal normado consistente de todas las funciones medibles $u : \mathbb{X} \to \mathbb{R}$, tales que satisfacen la condición

$$\|u\|_W := \sup_{x\in\mathbb{X}} \frac{|u(x)|}{W(x)} < \infty. \tag{1.8}$$

Hipótesis 1.2 *Condiciones de crecimiento.*

(a) Existe una función medible $W : \mathbb{X} \to [1,\infty)$, *así como constantes* $\bar{c}_1, \bar{c}_2 > 0$, $p > 1$, $b_0 > 0$ *y* $0 < \beta_0 < 1$, *tales que*

$$\sup_{a\in A(x)} |D(x,a)| \le \bar{c}_1 W(x), \qquad \sup_{a\in A(x)} |d(x,a)| \le \bar{c}_2 W(x), \tag{1.9}$$

y

$$\int_{\mathbb{X}} W^p(y) Q(dy \mid x,a) \le \beta_0 W^p(x) + b_0, \quad x \in \mathbb{X},\ a \in A(x). \tag{1.10}$$

(b) Para cada $x \in \mathbb{X}$, *la función*

$$a \mapsto \int_{\mathbb{X}} W(y) Q(dy \mid x,a)$$

es continua en $A(x)$.

Una consecuencia de la relación (1.10) es la siguiente desigualdad

$$\int_{\mathbb{X}} W(y) Q(dy \mid x,a) \le \beta W(x) + b, \quad \forall\ (x,a) \in \mathbb{K}, \tag{1.11}$$

donde $\beta = \beta_0^{1/p}$ y $b = b_0^{1/p}$. En efecto, como $p > 1$, de la Desigualdad de Jensen se tiene

$$\left[\int_{\mathbb{X}} W(y) Q(dy \mid x,a)\right]^p \le \int_{\mathbb{X}} W^p(y) Q(dy \mid x,a),$$

es decir

$$\int_{\mathbb{X}} W(y) Q(dy \mid x,a) \le \left[\int_{\mathbb{X}} W^p(y) Q(dy \mid x,a)\right]^{\frac{1}{p}}.$$

Luego, de (1.10) tenemos

$$\left[\int_{\mathbb{X}} W^p(y) Q(dy \mid x,a)\right]^{\frac{1}{p}} \le \left[\beta_0 W^p(x) + b_0\right]^{\frac{1}{p}}.$$

Finalmente, por el Teorema del Binomio se tiene que

$$\left[\beta_0 W^p(x) + b_0\right]^{\frac{1}{p}} \le \left[\beta_0 W^p(x)\right]^{\frac{1}{p}} + \left[b_0\right]^{\frac{1}{p}},$$

lo cual implica (1.11).

Lema 1.1 *Bajo la Hipótesis 1.2, para cada* $\pi \in \Pi$ *y* $x \in \mathbb{X}$ *se satisfacen las siguientes desigualdades*

$$\sup_{n\in\mathbb{N}_0} E_x^\pi \left[W^p(x_n)\right] < \infty$$

y

$$\sup_{n\in\mathbb{N}_0} E_x^\pi \left[W(x_n)\right] < \infty.$$

Demostración. Para demostrar que $\sup_{n\in\mathbb{N}_0} E_x^\pi[W^p(x_n)] < \infty$ para cada $\pi \in \Pi$ y $x \in \mathbb{X}$, obsérvese que de (1.4) y (1.10) se tiene

$$E_x^\pi[W^p(x_n) \mid h_{n-1}, a_{n-1}, \delta_n] = \int_{\mathbb{X}} W^p(y)Q(dy \mid x_{n-1}, a_{n-1}) \leq \beta_0 W^p(x_{n-1}) + b_0.$$

Luego, tomando esperanza en ambos lados de esta expresión se obtiene la desigualdad

$$E_x^\pi \left[W^p(x_n)\right] \leq \beta_0 E_x^\pi \left[W^p(x_{n-1})\right] + b_0.$$

Iterando esta desigualdad se obtiene

$$E_x^\pi[W^p(x_n)] \leq \beta_0^n E_x^\pi \left[W^p(x_0)\right] + b_0 \left(1 + \beta_0 + \beta_0^2 + ... + \beta_0^{n-1}\right).$$

Puesto que $\beta_0 \in (0,1)$, resulta

$$E_x^\pi[W^p(x_n)] \leq W^p(x) + \frac{b_0}{1-\beta_0} < \infty \quad \forall n \in \mathbb{N}.$$

Además, como $W(\cdot) \geq 1$,

$$E_x^\pi[W^p(x_n)] \;\leq \left(1 + \frac{b_0}{1-\beta_0}\right) W^p(x) < \infty \quad \forall n \in \mathbb{N}. \tag{1.12}$$

Por consiguiente,

$$\sup_{n\in\mathbb{N}_0} E_x^\pi[W^p(x_n)] < \infty \quad .$$

La demostración de la segunda desigualdad sigue un procedimiento similar, excepto que aquí se aplica (1.11), y en ese caso obtenemos

$$E_x^\pi[W(x_n)] \;\leq \left(1 + \frac{b}{1-\beta}\right) W(x) < \infty \quad \forall n \in \mathbb{N}. \tag{1.13}$$

■

Observación 1.2 *Bajo la Hipótesis 1.1(c), para cada* $u \in \mathbb{B}_W(\mathbb{X})$ *y* $x \in \mathbb{X}$,

$$a \mapsto \int_{\mathbb{X}} u(y) Q(dy \mid x, a)$$

es una función s.c.i. en $A(x)$.

Hipótesis 1.3 *Condición de regularidad.*

Existen reales $\varepsilon > 0$ *y* $\theta > 0$ *tales que*

$$H(\theta \mid x, a) \leq 1 - \varepsilon, \qquad \forall (x, a) \in \mathbb{K}.$$

Tal como se menciona al inicio de este capítulo, la condición precedente garantiza que en intervalos de tiempo acotados, los procesos controlados experimentan un número finito de transiciones, lo cual se demostrará en la Proposición 1.1, (véase pág. 9).

Las funciones definidas a continuación se usarán en los resultados subsecuentes.

Definición 1.6 *Para cada* $(x, a) \in \mathbb{K}$ *definimos*

$$\Delta_\alpha(x, a) := \int_0^\infty e^{-\alpha t} H(dt \mid x, a)$$

y

$$\tau_\alpha(x, a) := \frac{1 - \Delta_\alpha(x, a)}{\alpha}.$$

Nótese que de la relación (1.6):

$$\begin{aligned}
c(x,a) &= D(x,a) + d(x,a) \int_0^\infty \left[\int_0^t e^{-\alpha s} ds \right] H(dt \mid x, a) \\
&= D(x,a) + d(x,a) \int_0^\infty \left\{ -\frac{1}{\alpha} \left[e^{-\alpha t} - 1 \right] \right\} H(dt \mid x, a) \\
&= D(x,a) + d(x,a) \frac{1}{\alpha} \left\{ \int_0^\infty H(dt \mid x, a) - \int_0^\infty e^{-\alpha t} H(dt \mid x, a) \right\} \\
&= D(x,a) + d(x,a) \frac{1}{\alpha} \left\{ 1 - \int_0^\infty e^{-\alpha t} H(dt \mid x, a) \right\}.
\end{aligned}$$

Es decir,

$$c(x,a) = D(x,a) + \tau_\alpha(x,a) d(x,a), \quad (x,a) \in \mathbb{K}. \tag{1.14}$$

Observación 1.3 *Bajo la Hipótesis 1.1 se tiene que las funciones* $\Delta_\alpha(x,\cdot)$ *y* $\tau_\alpha(x,\cdot)$ *son continuas.*

Proposición 1.1 *Si se cumple la Hipótesis 1.3 , entonces:*

(a) $\rho_\alpha := \sup_{\mathbb{K}} \Delta_\alpha(x,a) < 1,$
(b) $P_x^\pi\left[\sum_{n=1}^{\infty}\delta_n = \infty\right] = 1 \quad \forall x \in \mathbb{X},\, \pi \in \Pi.$

Demostración. *(a)* Usando la fórmula de integración por partes para integrales de Riemann-Stieltjes

$$\int_\varsigma^\nu f(t)\,dg(t) + \int_\varsigma^\nu g(t)\,df(t) = f(\nu)\,g(\nu) - f(\varsigma)\,g(\varsigma),$$

y considerando en particular $\varsigma = 0$, $\nu = \infty$, $f(t) = e^{-\alpha t}$, $g(t) = H(t \mid x,a)$, se tiene

$$\int_0^\infty e^{-\alpha t} H(dt \mid x,a) + \int_0^\infty H(t \mid x,a)\,d\left(e^{-\alpha t}\right) = 0.$$

La igualdad a cero en esta expresión se deduce de las siguientes relaciones:

$$e^{-\alpha t} \underset{t\to\infty}{\rightarrow} 0 \qquad y \qquad H(t \mid x,a) \underset{t\to\infty}{\rightarrow} 1;$$

mientras que para $t = 0$:

$$e^{-\alpha t} = 1 \qquad y \qquad H(t \mid x,a) = 0.$$

Por lo tanto,

$$\begin{aligned}\Delta_\alpha(x,a) &= \int_0^\infty e^{-\alpha t} H(dt \mid x,a) = \alpha \int_0^\infty e^{-\alpha t} H(t \mid x,a)\,dt \\ &= \alpha\left[\int_0^\theta e^{-\alpha t} H(t \mid x,a)\,dt + \int_\theta^\infty e^{-\alpha t} H(t \mid x,a)\,dt\right].\end{aligned}$$

Ahora, por la Hipótesis 1.3

$$\begin{aligned}\alpha\int_0^\theta e^{-\alpha t} H(t \mid x,a)\,dt &\le \alpha(1-\varepsilon)\int_0^\theta e^{-\alpha t}dt \quad \forall (x,a) \in \mathbb{K} \\ &\le (1-\varepsilon)\left(1-e^{-\alpha\theta}\right);\end{aligned}$$

y además, dado que $H(t \mid x,a) \le 1$, entonces

$$\begin{aligned}\alpha\int_\theta^\infty e^{-\alpha t} H(t \mid x,a)\,dt &\le \alpha\int_\theta^\infty e^{-\alpha t}dt \quad \forall (x,a) \in \mathbb{K} \\ &\le e^{-\alpha\theta}.\end{aligned}$$

Esto es,

$$\begin{aligned}\Delta_\alpha(x,a) &\leq (1-\varepsilon)\left(1-e^{-\alpha\theta}\right)+e^{-\alpha\theta} \\ &= 1-\varepsilon\left(1-e^{-\alpha\theta}\right)<1 \quad \forall (x,a)\in\mathbb{K}. \end{aligned} \tag{1.15}$$

Por consiguiente

$$\rho_\alpha := \sup_{\mathbb{K}} \Delta_\alpha(x,a) < 1.$$

(b) Supongamos por el momento que para cada $x\in\mathbb{X}$, $\pi\in\Pi$:

$$E_x^\pi\left[\prod_{n=0}^{\infty}\Delta_\alpha(x_n,a_n)\right] = E_x^\pi\left[\exp\left(-\alpha\sum_{n=1}^{\infty}\delta_n\right) \mid x_0,a_0,x_1,a_1,...\right], \tag{1.16}$$

y

$$E_x^\pi\left[\prod_{n=0}^{\infty}\Delta_\alpha(x_n,a_n)\right]=0. \tag{1.17}$$

Partiendo de aquí se deduce que

$$E_x^\pi\left[\exp\left(-\alpha\sum_{n=1}^{\infty}\delta_n\right) \mid x_0,a_0,x_1,a_1,...\right]=0;$$

luego, debido a la positividad de la función exponencial se tiene

$$P_x^\pi\left[\sum_{n=1}^{\infty}\delta_n=\infty\right]=1 \quad \forall x\in\mathbb{X},\pi\in\Pi.$$

Esto demuestra la parte *(b)*, suponiendo que las igualdades (1.16) y (1.17) se cumplen. Ahora demostraremos (1.16) y (1.17). Para verificar la igualdad (1.16) obsérvese que

$$\begin{aligned}E_x^\pi\left[\prod_{n=0}^{\infty}\Delta_\alpha(x_n,a_n)\right] &= E_x^\pi\left[\prod_{n=0}^{\infty}\int_0^\infty e^{-\alpha t}H(dt\mid x_n,a_n)\right] \\ &= E_x^\pi\left[\prod_{n=0}^{\infty}E_x^\pi\left[e^{-\alpha\delta_{n+1}}\mid x_n,a_n\right]\right].\end{aligned}$$

Luego, de la Observación 1.1 se tiene

$$\begin{aligned}&E_x^\pi\left[\prod_{n=0}^{\infty}E_x^\pi\left[e^{-\alpha\delta_{n+1}}\mid x_n,a_n\right]\right] \\ &\qquad = E_x^\pi\left[E_x^\pi\left[\exp\left(-\alpha\sum_{n=1}^{\infty}\delta_n\right) \mid x_0,a_0,x_1,a_1,...\right]\right].\end{aligned}$$

por lo que, debido a propiedades de la esperanza condicional, lo anterior implica

$$E_x^\pi \left[\prod_{n=0}^{\infty} \Delta_\alpha (x_n, a_n) \right] = E_x^\pi \left[\exp \left(-\alpha \sum_{n=1}^{\infty} \delta_n \right) \mid x_0, a_0, x_1, a_1, ... \right],$$

lo cual demuestra (1.16).

Para probar la relación (1.17) nótese que de la parte *(a)* se tiene que $\Delta_\alpha (x, a) < 1$ para todo $(x, a) \in \mathbb{K}$, lo cual implica que

$$\prod_{k=0}^{n} \Delta_\alpha (x_k, a_k) \leq \rho_\alpha^{n+1} \underset{n \to \infty}{\to} 0.$$

Esto es, $\prod_{n=0}^{\infty} \Delta_\alpha (x_n, a_n)$ "diverge a cero", de donde

$$E_x^\pi \left[\prod_{n=0}^{\infty} \Delta_\alpha (x_n, a_n) \right] = 0.$$

■

Para concluir el capítulo presentamos el siguiente lema que incluye dos resultados importantes. El primero de ellos es uno de los "Teoremas de Selección Medible", mientras que el segundo proporciona condiciones que permiten el intercambio de límite e ínfimo.

Lema 1.2 *Supongamos que $A(x)$, $x \in \mathbb{X}$, es compacto (véase Hipótesis 1.1(*a*)) y sea $v : \mathbb{K} \to \mathbb{R}$ una función medible.*

(a) Si $v(x, a)$ es s.c.i. en $a \in A(x)$ para cada $x \in \mathbb{X}$, entonces

$$v^*(x) := \inf_{a \in A(x)} v(x, a) \tag{1.18}$$

es medible, y además existe $f^* \in \mathbb{F}$ tal que

$$v^*(x) = v(x, f^*) = \min_{a \in A(x)} v(x, a) \quad \forall x \in \mathbb{X}. \tag{1.19}$$

(b) Sean v y v_n $(n \in \mathbb{N})$ funciones medibles en $\mathbb{K}$ y s.c.i. sobre $A(x)$, tales que $v_n \uparrow v$. Entonces para cada $x \in \mathbb{X}$,

$$\min_{a \in A(x)} v_n(x, a) \underset{n \to \infty}{\to} \min_{a \in A(x)} v(x, a).$$

Demostración. La parte *(a)* es consecuencia del Corolario 4.3 en [24]. La demostración de la parte *(b)* es la siguiente. Dado que

$$v_n \uparrow v, \tag{1.20}$$

se tiene que para cada $x \in \mathbb{X}$:

$$v_n(x,a) \leq v(x,a) \quad \forall a \in A(x), \quad n \in \mathbb{N},$$

de modo que

$$\min_{a \in A(x)} v_n(x,a) \leq \min_{a \in A(x)} v(x,a).$$

Entonces

$$l(x) := \lim_{n \to \infty} \min_{a \in A(x)} v_n(x,a) \leq \min_{a \in A(x)} v(x,a) =: v\prime(x). \tag{1.21}$$

Por lo tanto, basta con demostrar que $l(x) \geq v\prime(x)$. Para ello considérese $x \in \mathbb{X}$ fijo y defínanse los conjuntos:

$$\begin{aligned} A_0 \quad &: \quad = \{a \in A(x) \ : \ v(x,a) = v\prime(x)\} \\ A_n \quad &: \quad = \{a \in A(x) \ : \ v_n(x,a) \leq v\prime(x)\}, \ n \in \mathbb{N}. \end{aligned}$$

Ahora obsérvese lo siguiente:

(b_1) Dada la compacidad de $A(x)$ y, debido a la semicontinuidad inferior de $v(x,\cdot)$, $v_n(x,\cdot)$, $x \in \mathbb{X}$, se tiene que A_0 y A_n son compactos no vacíos.

(b_2) De (1.20) se deduce que $A_n \downarrow A_0$.

Elíjase $a_n \in A_n$ (para cada $n \in \mathbb{N}$) de tal forma que

$$v_n(x,a_n) = \min_{a \in A(x)} v_n(x,a). \tag{1.22}$$

Nótese que (b_1) y (b_2) implican que existen $a_0 \in A_0$ y $\{a_{n_i}\} \subset \{a_n\}$ tales que $a_{n_i} \underset{i}{\to} a_0$. Mientras que nuevamente por (1.20), para cada $n \in \mathbb{N}$ y toda $n_i \geq n$ se observa que

$$v_{n_i}(x,a_{n_i}) \geq v_n(x,a_{n_i}),$$

razón por la cual

$$\lim_{i \to \infty} v_{n_i}(x,a_{n_i}) \quad \geq \quad \lim_{i \to \infty} v_n(x,a_{n_i}) \geq \lim_{i \to \infty} \inf v_n(x,a_{n_i}) \geq v_n(x,a_0). \tag{1.23}$$

La última desigualdad se debe a que v_n es s.c.i. sobre $A(x)$ y a la Proposición A.1 (véase Apéndice A).

Obsérvese ahora de (1.22) que

$$\lim_{i \to \infty} v_{n_i}(x,a_{n_i}) = \lim_{i \to \infty} \min_{a \in A(x)} v_{n_i}(x,a) =: l(x). \tag{1.24}$$

Luego, por (1.23) y (1.24) se tiene

$$l(x) \geq v_n(x,a_0) \quad \forall \ n \in \mathbb{N};$$

lo cual, debido a (1.20), produce

$$l(x) \geq v(x, a_0). \tag{1.25}$$

Mientras que de la definición de mínimo

$$v(x, a_0) \geq \min_{a \in A(x)} v(x, a) =: v\prime(x)\,; \tag{1.26}$$

de modo que (1.25) y (1.26) implican

$$l(x) \geq v(x, a_0). \tag{1.27}$$

Por consiguiente, de (1.21) y (1.27) se obtiene la igualdad $l(x) = v\prime(x)$, es decir

$$\lim_{n \to \infty} \min_{a \in A(x)} v_n(x, a) = \min_{a \in A(x)} v(x, a).$$

■

Observación 1.4 *Del Lema 1.2 y de las Hipótesis 1.1 y 1.2 se tiene que para cada $x \in \mathbb{X}$ y $u \in \mathbb{B}_W(\mathbb{X})$ la función*

$$a \longmapsto c\,(x, a) + \Delta_\alpha\,(x, a) \int_{\mathbb{X}} u(y) Q(dy \mid x, a)$$

es s.c.i. en $A\,(x)$. Además, existe $f^* \in \mathbb{F}$ tal que

$$c\,(x, f^*) + \Delta_\alpha\,(x, f^*) \int_{\mathbb{X}} u(y) Q(dy \mid x, f^*)$$

$$= \inf_{a \in A(x)} \left\{ c\,(x, a) + \Delta_\alpha\,(x, a) \int_{\mathbb{X}} u(y) Q(dy \mid x, a) \right\}.$$

Capítulo 2

Existencia de Políticas Óptimas y Optimalidad Asintótica

2.1. Introducción

En el presente capítulo se demuestra la existencia de políticas óptimas estacionarias bajo las hipótesis presentadas en el capítulo precedente. Lo anterior se consigue (en el Teorema 2.1) mostrando que la función de valor óptimo satisface la ecuación de optimalidad. Además, en general no existe una política óptima, ya que el índice en costo descontado se caracteriza por una fuerte dependencia de los controles aplicados durante las primeras etapas, que es precisamente donde el método de estimación que será introducido en el siguiente capítulo, no aporta suficiente información acerca de la densidad requerida. Tales circunstancias nos obligan a considerar un criterio más débil llamado optimalidad asintótica, el cual es introducido en la Sección 2.4.

2.2. Criterio de Optimalidad de Costo Descontado

Recordemos de la Definición 1.5 que, dados $x \in \mathbb{X}$ y $\pi \in \Pi$,

$$V(\pi, x) := E_x^\pi \left[\sum_{n=0}^{\infty} e^{-\alpha T_n} c(x_n, a_n) \right]$$

define el *costo total esperado α-descontado* al usar la política π dado el estado inicial $x_0 = x$. Además el *costo α-descontado óptimo* (función de valor óptimo) cuando el estado inicial es $x_0 = x$, se define como

$$V_\alpha^*(x) := \inf_{\pi\in\Pi} V(\pi, x); \tag{2.1}$$

y diremos que $\pi^* \in \Pi$ es una *política óptima* (α-óptima) si

$$V(\pi^*, x) = \inf_{\pi\in\Pi} V(\pi, x) \qquad \forall x \in \mathbb{X}. \tag{2.2}$$

En el siguiente resultado se introduce una forma alternativa de expresar el costo total esperado α-descontado.

Proposición 2.1 *Para todo $x \in \mathbb{X}$ y $\pi \in \Pi$ se tiene que*

$$V(\pi, x) = E_x^\pi \left[c(x_0, a_0) + \sum_{n=1}^{\infty} \prod_{j=0}^{n-1} \Delta_\alpha(x_j, a_j)\, c(x_n, a_n) \right], \tag{2.3}$$

donde $\Delta_\alpha(\cdot,\cdot)$ se introdujo en la Definición 1.6.

Demostración. De la definición de $\Delta_\alpha(x, a)$ puede observarse que

$$\begin{aligned}
& E_x^\pi \left[c(x_0, a_0) + \sum_{n=1}^{\infty} \prod_{k=0}^{n-1} \Delta_\alpha(x_k, a_k)\, c(x_n, a_n) \right] \\
&\quad = E_x^\pi \left[c(x_0, a_0) + \sum_{n=1}^{\infty} \prod_{k=0}^{n-1} \int_0^\infty e^{-\alpha t} H(dt \mid x_k, a_k)\, c(x_n, a_n) \right] \\
&\quad = E_x^\pi \left[c(x_0, a_0) + \sum_{n=1}^{\infty} \prod_{k=0}^{n-1} E_x^\pi \left[e^{-\alpha \delta_{k+1}} \mid x_k, a_k \right] c(x_n, a_n) \right].
\end{aligned}$$

Luego, de la Observación 1.1 note que

$$\begin{aligned}
\prod_{k=0}^{n-1} E_x^\pi \left[e^{-\alpha\delta_{k+1}} \mid x_k, a_k \right] &= E_x^\pi \left[\prod_{k=0}^{n-1} e^{-\alpha\delta_{k+1}} \mid x_k, a_k \right] \\
&= E_x^\pi \left[e^{-\alpha(\delta_1+\cdots+\delta_n)} \mid x_0, a_0,\ x_1, a_1, ... \right];
\end{aligned}$$

entonces

$$\begin{aligned}
& E_x^\pi \left[c(x_0, a_0) + \sum_{n=1}^{\infty} \prod_{k=0}^{n-1} \Delta_\alpha(x_k, a_k)\, c(x_n, a_n) \right] \\
&\qquad = E_x^\pi \left[\sum_{n=0}^{\infty} E_x^\pi \left[e^{-\alpha T_n} \mid x_0, a_0,\ x_1, a_1, ... \right] c(x_n, a_n) \right],
\end{aligned}$$

donde T_n $(n \in \mathbb{N}_0)$ es precisamente la variable aleatoria introducida en la Sección 1.2.2, la cual representa la n-ésima época de decisión y, misma que admite también la expresión: $T_0 = 0$ y $T_n = \sum_{k=1}^{n} \delta_k$, para $n \in \mathbb{N}$. Ahora bien, por las propiedades de esperanza condicional se obtiene

$$E_x^\pi \left[c(x_0, a_0) + \sum_{n=1}^{\infty} \prod_{k=0}^{n-1} \Delta_\alpha (x_k, a_k)\, c(x_n, a_n) \right]$$
$$= \sum_{n=0}^{\infty} E_x^\pi \left[e^{-\alpha T_n} c(x_n, a_n) \right].$$

En consecuencia,

$$E_x^\pi \left[c(x_0, a_0) + \sum_{n=1}^{\infty} \prod_{k=0}^{n-1} \Delta_\alpha (x_k, a_k)\, c(x_n, a_n) \right] = V(\pi, x).$$

■

Para $x \in \mathbb{X}$, $\pi \in \Pi$ y $n \in \mathbb{N}$, denotemos por $V_n(\pi, x)$ el costo esperado α-descontado hasta la n-ésima transición, es decir,

$$V_n(\pi, x) = E_x^\pi \left[\sum_{k=0}^{n-1} e^{-\alpha T_k} c(x_k, a_k) \right].$$

Obsérvese que

$$V_n(\pi, x) \uparrow V(\pi, x).$$

Por otra parte, con argumentos similares a los de la demostración correspondiente a la Proposición 2.1, $V_n(\pi, x)$ puede ser expresado como

$$V_1(\pi, x) = E_x^\pi \left[c(x_0, a_0) \right]$$
$$V_n(\pi, x) = E_x^\pi \left[c(x_0, a_0) + \sum_{k=1}^{n-1} \prod_{j=0}^{k-1} \Delta_\alpha (x_j, a_j)\, c(x_k, a_k) \right], \ n = 2, 3, ...$$

Luego, introduciendo la notación

$$\begin{aligned} \Lambda_\alpha^0 &: = 1 \\ \Lambda_\alpha^k &: = \prod_{j=0}^{k-1} \Delta_\alpha (x_j, a_j), \quad k \in \mathbb{N}, \end{aligned} \tag{2.4}$$

se tiene

$$V_n(\pi, x) = E_x^\pi \left[\sum_{k=0}^{n-1} \Lambda_\alpha^k c(x_k, a_k) \right] \quad para \ \ n \in \mathbb{N}; \tag{2.5}$$

y

$$V(\pi, x) = \lim_{n \to \infty} V_n(\pi, x) := E_x^\pi \left[\sum_{k=0}^{\infty} \Lambda_\alpha^k c(x_k, a_k) \right]. \tag{2.6}$$

2.3. Ecuación de Optimalidad

Sean $c(x,a)$ como en (1.14), $\Delta_\alpha(x,a)$ como en la Definición 1.6 y $\alpha > 0$. Diremos que una función medible $u : \mathbb{X} \to \mathbb{R}$ es una solución de la *ecuación de optimalidad en costo α-descontado* (*EOCD*) si satisface

$$u(x) = \min_{a \in A(x)} \left\{ c(x,a) + \Delta_\alpha(x,a) \int_{\mathbb{X}} u(y) Q(dy \mid x,a) \right\} \quad \forall x \in \mathbb{X}.$$

El objetivo de esta sección es demostrar que la función de valor óptimo satisface la *EOCD*. Esto nos permitirá mostrar la existencia de políticas óptimas. La idea general para hacer lo anterior consiste en lo siguiente.

Denotemos por v_n la función de valor óptimo hasta la n-ésima transición. Es decir,

$$\begin{aligned} v_0(x) &: = 0 \\ v_n(x) &: = \inf_{\pi \in \Pi} V_n(\pi, x), \quad x \in \mathbb{X}. \end{aligned}$$

Del Algoritmo de Programación Dinámica (véase [8] y [14]) observamos que las funciones v_n pueden ser obtenidas de forma iterada, ésto es, para $x \in \mathbb{X}$:

$$\begin{aligned} v_0(x) &= 0 \quad \textit{y para} \quad n \geq 1, \\ v_n(x) &= \min_{a \in A(x)} \left\{ c(x,a) + \Delta_\alpha(x,a) \int_{\mathbb{X}} v_{n-1}(y) Q(dy \mid x,a) \right\}. \end{aligned} \tag{2.7}$$

Por lo tanto, si mostramos que $v_n \to V_\alpha^*$ podemos esperar que V_α^* satisfaga la *EOCD*. Formalizaremos ésto más delante en el Teorema 2.1.

Proposición 2.2 *Bajo la Hipótesis 1.2(a), $V_\alpha^* \in \mathbb{B}_W(\mathbb{X})$.*

Demostración. Nótese que de la Hipótesis 1.2*(a)* y por (1.14)

$$\begin{aligned} E_x^\pi[c(x_n,a_n)] &= E_x^\pi[D(x_n,a_n) + \tau_\alpha(x_n,a_n) d(x_n,a_n)] \\ &\leq E_x^\pi\left[\bar{c}_1 W(x_n) + \frac{\bar{c}_2}{\alpha} W(x_n)\right] = \bar{c} E_x^\pi[W(x_n)], \end{aligned}$$

donde $\bar{c} = \bar{c}_1 + \frac{\bar{c}_2}{\alpha}$.

Luego, considerando (1.13), de lo anterior se obtiene

$$E_x^\pi[c(x_n,a_n)] \leq \left(1 + \frac{b}{1-\beta}\right) \bar{c} W(x). \tag{2.8}$$

Ahora, por (2.3) y la Proposición 1.1*(a)*

$$\begin{aligned} V(\pi, x) &\leq \left(1 + \frac{b}{1-\beta}\right) \bar{c} W(x) \sum_{n=0}^{\infty} (\rho_\alpha)^n \\ &\leq \bar{c} \left(1 + \frac{b}{1-\beta}\right) \left(\frac{1}{1-\rho_\alpha}\right) W(x). \end{aligned}$$

Por (2.1) y tomando $M := \bar{c}\left(1 + \frac{b}{1-\beta}\right)\left(\frac{1}{1-\rho_\alpha}\right)$, de lo anterior se tiene

$$V_\alpha^*(x) \leq MW(x). \tag{2.9}$$

En consecuencia, V_α^* pertenece a $\mathbb{B}_W(\mathbb{X})$.

■

Observación 2.1 *De forma similar que en la proposición previa se demuestra que para cada $n \in \mathbb{N}$,*

$$v_n(x) \leq MW(x), \qquad x \in \mathbb{X}.$$

Teorema 2.1 *Supóngase que se cumplen las Hipótesis 1.1, 1.2 y 1.3. Entonces*

(a) $\|v_n - V_\alpha^*\|_W \to 0$ *cuando* $n \to \infty$.

(b) V_α^* *satisface la EOCD, es decir,*

$$V_\alpha^*(x) = \min_{a \in A(x)} \left\{ c(x,a) + \Delta_\alpha(x,a) \int_{\mathbb{X}} V_\alpha^*(y) Q(dy \mid x,a) \right\} \qquad x \in \mathbb{X}. \tag{2.10}$$

(c) *Existe* $f^* \in \mathbb{F}$ *que minimiza el lado derecho en* (2.10), *es decir*

$$V_\alpha^*(x) = c(x,f^*) + \Delta_\alpha(x,f^*) \int_{\mathbb{X}} V_\alpha^*(y) Q(dy \mid x,f^*) \qquad \forall x \in \mathbb{X}, \tag{2.11}$$

y la política estacionaria $\pi = \{f^*, f^*, ...\}$ *es óptima.*

Demostración. *(a)* Primero obsérvese (véase (2.5) y (2.6)) que para cada $x \in \mathbb{X}$ y $\pi \in \Pi$,

$$\begin{aligned} V(\pi, x) &= E_x^\pi \left[\sum_{k=0}^{\infty} \Lambda_\alpha^k c(x_k, a_k) \right] \\ &= E_x^\pi \left[\sum_{k=0}^{n-1} \Lambda_\alpha^k c(x_k, a_k) \right] + E_x^\pi \left[\sum_{k=n}^{\infty} \Lambda_\alpha^k c(x_k, a_k) \right] \\ &= V_n(\pi, x) + E_x^\pi \left[\Lambda_\alpha^n \sum_{k=n}^{\infty} \Lambda_\alpha^{k-n} c(x_k, a_k) \right]. \end{aligned} \tag{2.12}$$

Ahora, para cada $x \in \mathbb{X}$,

$$\begin{aligned} |v_n(x) - V_\alpha^*(x)| &= \left| \inf_{\pi\in\Pi} V_n(\pi, x) - \inf_{\pi\in\Pi} V(\pi, x) \right| \\ &\leq \sup_{\pi\in\Pi} \left| V_n(\pi, x) - V(\pi, x) \right|, \end{aligned}$$

de donde, usando las relaciones (2.12) y (2.8), así como de la Proposición 1.1*(b)* se tiene lo siguiente

$$\begin{aligned} |v_n(x) - V_\alpha^*(x)| &\leq \rho_\alpha^n \sup_{\pi\in\Pi} \sum_{k=n}^{\infty} \rho_\alpha^{k-n} E_x^\pi \left[c(x_k, a_k) \right] \\ &\leq \rho_\alpha^n \sup_{\pi\in\Pi} \sum_{k=n}^{\infty} \left(1 - \frac{b}{1-\beta} \right) \bar{c} W(x) \rho_\alpha^{k-n} \\ &\leq \rho_\alpha^n \left(1 - \frac{b}{1-\beta} \right) \bar{c} W(x) \frac{1}{1-\rho_\alpha} = M \rho_\alpha^n W(x). \end{aligned}$$

Por consiguiente, de la definición de la norma $\|\cdot\|_W$ (véase (1.8)) lo anterior implica que

$$\|v_n - V_\alpha^*\|_W \leq M \rho_\alpha^n.$$

Haciendo tender n a infinito, y usando de la Proposición 1.1 el hecho de que $\rho_\alpha < 1$, concluimos la demostración de la parte *(a)*.

(b) Tomando límite inferior en ambos lados de (2.7), por el Lema de Fatou "Generalizado"(véase Apéndice D, Proposición D.1) se obtiene que para cada $x \in \mathbb{X}$,

$$\begin{aligned} \liminf_{n\to\infty} v_n(x) &= \min_{a\in A(x)} \left\{ c(x, a) + \right. \\ &\quad \left. + \Delta_\alpha(x, a) \liminf_{n\to\infty} \left(\int_{\mathbb{X}} v_{n-1}(y) Q(dy \mid x, a) \right) \right\} \\ &\geq \min_{a\in A(x)} \left\{ c(x, a) + \right. \\ &\quad \left. + \Delta_\alpha(x, a) \int_{\mathbb{X}} \liminf_{n\to\infty} v_{n-1}(y) Q(dy \mid x, a) \right\}; \end{aligned}$$

luego, aplicando lo demostrado en la parte *(a)*, lo anterior implica que

$$V_\alpha^*(x) \geq \min_{a\in A(x)} \left\{ c(x, a) + \Delta_\alpha(x, a) \int_{\mathbb{X}} V_\alpha^*(y) Q(dy \mid x, a) \right\}. \tag{2.13}$$

Ahora obsérvese que (2.7) implica

$$v_n(x) \leq c(x,a) + \Delta_\alpha(x,a)\int_{\mathbb{X}} v_{n-1}(y)Q(dy \mid x,a), \quad x \in \mathbb{X},\ a \in A(x). \tag{2.14}$$

Nuevamente por aplicación del Lema de Fatou, se tiene que

$$V_\alpha^*(x) \leq c(x,a) + \Delta_\alpha(x,a)\int_{\mathbb{X}} V_\alpha^*(y)\, Q(dy \mid x,a). \tag{2.15}$$

Por consiguiente, la parte *(b)* se obtiene combinando (2.13) y (2.15).

(c) Recordemos de la Observación 1.3 que $\Delta_\alpha(x,a)$ y $\tau_\alpha(x,a)$ son continuas en $A(x)$ para cada $x \in \mathbb{X}$. Entonces por el Lema 1.2, existe $f^* \in \mathbb{F}$ que satisface (2.11), es decir,

$$V_\alpha^*(x) = c(x,f^*) + \Delta_\alpha(x,f^*)\int_{\mathbb{X}} V_\alpha^*(y)Q(dy \mid x,f^*), \quad x \in \mathbb{X}.$$

Iterando esta ecuación se obtiene

$$V_\alpha^*(x) = E_x^{f^*}\left[\sum_{n=0}^{N-1} \Lambda_\alpha^n c(x_n,a_n) + \Lambda_\alpha^N V_\alpha^*(x_N)\right],$$

de donde

$$V_\alpha^*(x) \geq V_N(f^*,x) \quad \forall x \in \mathbb{X},\ N \geq 1, \tag{2.16}$$

pues $\Lambda_\alpha^N V_\alpha^*(x_N) \geq 0$ para cada $N \geq 1$.

Haciendo $N \to \infty$ en (2.16) se tiene

$$V_\alpha^*(x) \geq V(f^*,x) \quad \forall x \in \mathbb{X}. \tag{2.17}$$

Por otra parte, de (2.1) y (2.2) se deduce que

$$V_\alpha^*(x) \leq V(f^*,x) \quad \forall x \in \mathbb{X}. \tag{2.18}$$

En consecuencia

$$V_\alpha^*(x) = V(f^*,x) \quad \forall x \in \mathbb{X},$$

y por lo tanto, $\pi = \{f^*, f^*, ...\}$ es óptima.

■

2.4. Optimalidad Asintótica

En esta sección presentaremos otro criterio de optimalidad, el cual será central para el desarrollo de este trabajo.

Definición 2.1 *Una política* $\pi \in \Pi$ *es asintóticamente óptima descontada (AOD) si para cada* $x \in \mathbb{X}$*,*

$$E_x^\pi \left[\sum_{t=n}^{\infty} \Lambda_{\alpha,n}^t c(x_t, a_t) \right] - E_x^\pi \left[V_\alpha^*(x_n) \right] \to 0 \quad \textit{cuando} \quad n \to \infty,$$

donde

$$\begin{aligned} \Lambda_{\alpha,n}^n \quad &: \quad = 1 \\ \Lambda_{\alpha,n}^t \quad &: \quad = \prod_{j=n}^{t-1} \Delta_\alpha (x_j, a_j), \quad t > n. \end{aligned} \tag{2.19}$$

El resultado que se introduce a continuación nos facilitará el estudio de la optimalidad asintótica.

Teorema 2.2 *Una política* $\pi \in \Pi$ *es AOD si y sólo si para cada* $x \in \mathbb{X}$*,*

$$E_x^\pi \left[\Phi(x_n, a_n) \right] \to 0 \quad \textit{cuando} \quad n \to \infty,$$

donde

$$\Phi(x,a) := c(x,a) + \Delta_\alpha (x,a) \int_{\mathbb{X}} V_\alpha^*(y) Q(dy \mid x,a) - V_\alpha^*(x), \ (x,a) \in \mathbb{K}.$$

A la función Φ se le conoce como *función de discrepancia.*

La idea intuitiva de optimalidad asintótica es la siguiente. Obsérvese que, por el Teorema 2.1*(b)*, $\Phi(x,a) \geq 0$ para cada $(x,a) \in \mathbb{K}$. Más aún, la *EOCD* es equivalente a la relación

$$\min_{a \in A(x)} \Phi(x,a) = 0,$$

y por el Teorema 2.1*(c)*, la política estacionaria $\pi = \{f, f, ...\}$ es óptima si

$$\Phi(x,f) = 0 \qquad \forall x \in \mathbb{X}.$$

Entonces, una política $\pi \in \Pi$ será *AOD* si en el límite la función de discrepancia correspondiente es cero.

Para demostrar el Teorema 2.2 se hará uso del siguiente resultado, el cual se demostrará al final del presente capítulo.

Lema 2.1 *Para cada $\pi \in \Pi$ y $x \in \mathbb{X}$,*

$$\sum_{t=n}^{\infty} E_x^{\pi}\left[\Lambda_{\alpha,n}^{t}\Phi(x_t,a_t)\right] = E_x^{\pi}\left[\sum_{t=n}^{\infty}\Lambda_{\alpha,n}^{t} c(x_t,a_t) - V_{\alpha}^{*}(x_n)\right].$$

Demostración del Teorema 2.2. Para demostrar que la convergencia $E_x^{\pi}[\Phi(x_n,a_n)] \to 0$ cuando $n \to \infty$ implica que la política π es *AOD*, primero nótese que

$$\Lambda_{\alpha,n}^{t} := \prod_{k=n}^{t-1}\Delta_{\alpha}(x_k,a_k) \leq \rho_{\alpha}^{t-n}. \tag{2.20}$$

Como $E_x^{\pi}[\Phi(x_n,a_n)] \to 0$ cuando $n \to \infty$, entonces dado $\varepsilon > 0$, para n suficientemente grande tenemos que $E_x^{\pi}[\Phi(x_n,a_n)] < \varepsilon$. Esto implica que

$$\sum_{t=n}^{\infty} E_x^{\pi}\left[\Lambda_{\alpha,n}^{t}\Phi(x_t,a_t)\right] \leq \sum_{t=n}^{\infty}\rho_{\alpha}^{t-n}E_x^{\pi}\left[\Phi(x_t,a_t)\right] \leq \varepsilon\sum_{t=n}^{\infty}\rho_{\alpha}^{t-n},$$

de donde

$$\sum_{t=n}^{\infty} E_x^{\pi}\left[\Lambda_{\alpha,n}^{t}\Phi(x_t,a_t)\right] \leq \frac{\varepsilon}{1-\rho_{\alpha}}.$$

Por lo tanto,

$$\sum_{t=n}^{\infty} E_x^{\pi}\left[\Lambda_{\alpha,n}^{t}\Phi(x_t,a_t)\right] \to 0 \quad \textit{cuando} \quad n \to \infty.$$

Luego entonces, por el Lema 2.1,

$$E_x^{\pi}\left[\sum_{t=n}^{\infty}\Lambda_{\alpha,n}^{t} c(x_t,a_t) - V_{\alpha}^{*}(x_n)\right] \to 0 \quad \textit{cuando} \quad n \to \infty,$$

es decir, π es *AOD* (véase la Definición 2.1).

Para la demostración del recíproco, supóngase que π es una política *AOD*. Obsérvese que por (2.19) se tiene que para cada $n \in \mathbb{N}$,

$$\begin{aligned}\sum_{t=n}^{\infty} E_x^{\pi}\left[\Lambda_{\alpha,n}^{t}\Phi(x_t,a_t)\right] \quad &= \quad E_x^{\pi}\left[\Phi(x_n,a_n)\right] + \\ &\quad + \sum_{t=n+1}^{\infty} E_x^{\pi}\left[\Lambda_{\alpha,n}^{t}\Phi(x_t,a_t)\right],\end{aligned}$$

y del Lema 2.1 se sigue que

$$E_x^{\pi}\left[\Phi(x_n,a_n)\right] + \sum_{t=n+1}^{\infty} E_x^{\pi}\left[\Lambda_{\alpha,n}^{t}\Phi(x_t,a_t)\right] \to 0 \quad \textit{cuando} \quad n \to \infty.$$

De lo anterior, y por la no negatividad de $\Lambda^t_{\alpha,n}$ y $\Phi(x_t,a_t)$ para todo t, se concluye que

$$E^\pi_x\left[\Phi(x_n,a_n)\right]\to 0 \quad \textit{cuando} \quad n\to\infty.$$

■

Demostración del Lema 2.1. Supóngase por el momento que para cada $\pi\in\Pi$ y $x\in\mathbb{X}$,

$$\lim_{m\to\infty} E^\pi_x\left[\Lambda^m_{\alpha,n}V^*_\alpha(x_m)\right]=0. \tag{2.21}$$

Primero nótese que

$$\begin{aligned}
\Phi(x_t,a_t) &= c\,(x_t,a_t)+\Delta_\alpha\,(x_t,a_t)\int_{\mathbb{X}} V^*_\alpha\,(y)\,Q(dy\mid x_t,a_t)-V^*_\alpha\,(x_t)\\
&= E^\pi_x\left[\{c\,(x_t,a_t)+\Delta_\alpha\,(x_t,a_t)\,V^*_\alpha\,(x_{t+1})-V^*_\alpha\,(x_t)\}\ \mid h_t,a_t\right].
\end{aligned}$$

De aquí,

$$\begin{aligned}
\sum_{t=n}^{\infty}E^\pi_x\left[\Lambda^t_{\alpha,n}\Phi(x_t,a_t)\right] &= \sum_{t=n}^{\infty}E^\pi_x\left[\Lambda^t_{\alpha,n}E^\pi_x\left[\{c\,(x_t,a_t)\right.\right.\\
&\qquad +\Delta_\alpha\,(x_t,a_t)\,V^*_\alpha\,(x_{t+1})\\
&\qquad\qquad \left.\left.-V^*_\alpha\,(x_t)\ \mid h_t,a_t\}\right]\right]\\
&= \sum_{t=n}^{\infty}E^\pi_x\left[\Lambda^t_{\alpha,n}E^\pi_x\left[c\,(x_t,a_t)\ \mid h_t,a_t\right]\right]\\
&\quad +\sum_{t=n}^{\infty}E^\pi_x\left[\Lambda^t_{\alpha,n}E^\pi_x\left[\{\Delta_\alpha\,(x_t,a_t)\,V^*_\alpha\,(x_{t+1})\right.\right.\\
&\qquad\qquad \left.\left.-V^*_\alpha\,(x_t)\}\ \mid h_t,a_t\right]\right].
\end{aligned}$$

Luego usando el hecho de que $c\,(x_t,a_t)$ y $\Lambda^t_{\alpha,n}$ son medibles con respecto a la $\sigma-$álgebra generada por (h_t,a_t), y aplicando propiedades de esperanza condicional lo anterior implica que

$$\begin{aligned}
\sum_{t=n}^{\infty}E^\pi_x\left[\Lambda^t_{\alpha,n}\Phi(x_t,a_t)\right] &= \sum_{t=n}^{\infty}E^\pi_x\left[\Lambda^t_{\alpha,n}c\,(x_t,a_t)\right]\\
&\quad +\sum_{t=n}^{\infty}E^\pi_x\left[E^\pi_x\left[\left\{\Lambda^t_{\alpha,n}\Delta_\alpha\,(x_t,a_t)\,V^*_\alpha\,(x_{t+1})\right.\right.\right.\\
&\qquad\qquad \left.\left.\left.-\Lambda^t_{\alpha,n}V^*_\alpha\,(x_t)\right\}\ \mid h_t,a_t\right]\right]\\
&= E^\pi_x\left[\sum_{t=n}^{\infty}\Lambda^t_{\alpha,n}c\,(x_t,a_t)\right]\\
&\quad +\sum_{t=n}^{\infty}E^\pi_x\left[\Lambda^{t+1}_{\alpha,n}V^*_\alpha\,(x_{t+1})-\Lambda^t_{\alpha,n}V^*_\alpha\,(x_t)\right],
\end{aligned}$$

donde el intercambio de operadores en el primer término del último renglón es posible dado que, $\Lambda_{\alpha,n}^{t}c\left(x_t,a_t\right)$ es no negativa para cada t.

Por otro lado tenemos,

$$\begin{aligned}
&\sum_{t=n}^{\infty} E_x^{\pi}\left[\Lambda_{\alpha,n}^{t+1}V_\alpha^*\left(x_{t+1}\right)-\Lambda_{\alpha,n}^{t}V_\alpha^*\left(x_t\right)\right] \\
&\qquad\qquad = \lim_{m\to\infty}\left\{\sum_{t=n}^{m} E_x^{\pi}\left[\Lambda_{\alpha,n}^{t+1}V_\alpha^*\left(x_{t+1}\right)-\Lambda_{\alpha,n}^{t}V_\alpha^*\left(x_t\right)\right]\right\} \\
&\qquad\qquad = \lim_{m\to\infty} E_x^{\pi}\left[\Lambda_{\alpha,n}^{m+1}V_\alpha^*\left(x_{m+1}\right)-\Lambda_{\alpha,n}^{n}V_\alpha^*\left(x_n\right)\right] \\
&\qquad\qquad = \lim_{m\to\infty} E_x^{\pi}\left[\Lambda_{\alpha,n}^{m+1}V_\alpha^*\left(x_{m+1}\right)\right]-V_\alpha^*\left(x_n\right),
\end{aligned}$$

puesto que $\Lambda_{\alpha,n}^{n}=1$ (véase (2.19)). De aquí

$$\begin{aligned}
\sum_{t=n}^{\infty} E_x^{\pi}\left[\Lambda_{\alpha,n}^{t}\Phi(x_t,a_t)\right] \quad = \quad & E_x^{\pi}\left[\sum_{t=n}^{\infty}\Lambda_{\alpha,n}^{t}c\left(x_t,a_t\right)\right] \\
& + \lim_{m\to\infty} E_x^{\pi}\left[\Lambda_{\alpha,n}^{m+1}V_\alpha^*\left(x_{m+1}\right)\right]-V_\alpha^*\left(x_n\right),
\end{aligned}$$

de lo cual, por (2.21) se tiene que

$$\sum_{t=n}^{\infty} E_x^{\pi}\left[\Lambda_{\alpha,n}^{t}\Phi(x_t,a_t)\right]=E_x^{\pi}\left[\sum_{t=n}^{\infty}\Lambda_{\alpha,n}^{t}c\left(x_t,a_t\right)-V_\alpha^*\left(x_n\right)\right]$$

para cada $\pi\in\Pi$, $x\in\mathbb{X}$.

Para concluir demostraremos (2.21).

Sea $p\prime$ el exponente conjugado de p, es decir,

$$\frac{1}{p}+\frac{1}{p\prime}=1.$$

La constante p se introdujo en la Hipótesis 1.2*(a)*. Entonces, aplicando la Desigualdad de Hölder se tiene que

$$E_x^{\pi}\left[\Lambda_{\alpha,n}^{m}V_\alpha^*(x_m)\right]\le\left\{E_x^{\pi}\left[\left(\Lambda_{\alpha,n}^{m}\right)^{p\prime}\right]\right\}^{\frac{1}{p\prime}}\left\{E_x^{\pi}\left[\left(V_\alpha^*(x_m)\right)^{p}\right]\right\}^{\frac{1}{p}}.$$

Por otro lado, las relaciones (2.9) y (2.20) implican que

$$\begin{aligned}
&\left\{E_x^{\pi}\left[\left(\Lambda_{\alpha,n}^{m}\right)^{p\prime}\right]\right\}^{\frac{1}{p\prime}}\left\{E_x^{\pi}\left[\left(V_\alpha^*(x_m)\right)^{p}\right]\right\}^{\frac{1}{p}} \\
&\qquad\qquad \le\left\{E_x^{\pi}\left[\left(\rho_\alpha^{m-n}\right)^{p\prime}\right]\right\}^{\frac{1}{p\prime}}\left\{E_x^{\pi}\left[\left(MW(x_n)\right)^{p}\right]\right\}^{\frac{1}{p}} \\
&\qquad\qquad \le\left\{\left(\rho_\alpha^{m-n}\right)^{p\prime}\right\}^{\frac{1}{p\prime}}\left\{M^pE_x^{\pi}\left[W^p(x_n)\right]\right\}^{\frac{1}{p}},
\end{aligned}$$

de lo cual, por la relación (1.12) se obtiene la siguiente expresión

$$E_x^\pi\left[\Lambda_{\alpha,n}^m V_\alpha^*(x_m)\right] \leq \rho_\alpha^{m-n} M\left(1-\frac{b_0}{1-\beta_0}\right)^{\frac{1}{p}} W(x).$$

Haciendo $m \to \infty$ se obtiene (2.21), puesto que $\rho_\alpha < 1$.

■

Capítulo 3

Estimación de la Distribución del Tiempo de Permanencia

3.1. Introducción

En este capítulo estudiaremos el problema de control óptimo para procesos semi-markovianos suponiendo que se desconoce la distribución del tiempo de permanencia H. Ante este caso, el principal objetivo será precisamente el estudio del problema de control óptimo descontado asociado al mismo. Específicamente, enfocaremos nuestro interés sobre la construcción de políticas de control que minimicen el criterio en costo descontado previamente introducido.

La idea general de nuestro tratamiento para la situación anteriormente - descrita, consiste en combinar técnicas de control con métodos adecuados de estimación de H, y el algoritmo resultante de tal combinación es usado para construir una política AOD $\hat{\pi} = \{f_n\}$. Más aún, demostramos la convergencia (en el sentido de Schäl [27]) de $\{f_n\}$ a una política estacionaria óptima $\{f_\infty\}$ del *PCSM*.

3.2. Hipótesis Específicas

En el Capítulo 1 se presentó el *MCSM* estándar, en el cual todos los componentes se suponen conocidos por el controlador. Ahora bajo el contexto actual, siendo H desconocida, obsérvese que por (1.6) el costo c también será desconocido. Por lo tanto, antes de elegir el control a_n en la n-ésima época de decisión, el controlador debe construir un estimador H_n de la distribución H (por consiguiente de c) y combinar ésto con la historia del sistema para seleccionar un control $a = a_n(H_n)$.

Obsérvese que la distribución H depende del estado x_n y el control a_n. Este hecho dificulta la obtención de una cantidad considerable de observaciones independientes e idénticamente distribuidas de los tiempos de permanencia. Y en consecuencia, hace difícil la obtención de estimadores eficientes de la distribución H en cada época de decisión. Dada tal situación, supondremos ahora que la distribución H tiene una densidad que no depende de los pares (x_n, a_n). Específicamente impondremos la siguiente hipótesis.

Hipótesis 3.1 *Existe una función de distribución G (desconocida) con densidad g tal que*

$$H(t \mid x, a) = G(t) = \int_0^t g(s)ds, \quad \forall (x, a) \in \mathbb{K}, \quad t \geq 0.$$

Obsérvese que debido a esta condición y por (1.5), la distribución de δ_n $(n \in \mathbb{N})$ es independiente de la política π, y

$$P_x^\pi \left(\delta_n \leq t\ \right) = \int_0^t g(s)ds. \tag{3.1}$$

Además, bajo esta nueva suposición, señalamos a continuación la forma que toma la Hipótesis 1.3.

Observación 3.1 *Existe $\varepsilon > 0$ y $\theta > 0$ tal que*

$$\int_0^\theta g(s)ds \leq 1 - \varepsilon.$$

Más aún, nótese que la Hipótesis 1.1*(d)* ya no es necesaria.

La condición considerada a continuación impone requerimientos técnicos sobre la densidad g, con los cuales podremos obtener un método apropiado de estimación.

Hipótesis 3.2 *Existe una constante $q \in (1, 2)$ y una función medible*

$$\bar{g} : [0, \infty) \to [0, \infty)$$

tal que $g \in L_q([0, \infty))$, $g(s) \leq \bar{g}(s)$ casi dondequiera con respecto a la medida de Lebesgue y

$$\int_0^\infty \left(\bar{g}(s)\right)^{2-q} ds < \infty.$$

En el siguiente capítulo presentaremos un ejemplo de un sistema de almacenamiento que satisface todas estas hipótesis.

Observación 3.2 *Debido a que H no depende del par (x, a), las expresiones en la Definición 1.6 adquieren la siguiente forma. Para cada $(x, a) \in \mathbb{K}$*

$$\Delta_\alpha = \int_0^\infty e^{-\alpha s} g(s)\, ds,$$

y

$$\tau_\alpha = \frac{1-\Delta_\alpha}{\alpha}; \tag{3.2}$$

*por lo cual (*1.14*) se reescribe ahora como*

$$c(x,a) \quad = \quad D(x,a) + \tau_\alpha d(x,a), \quad (x,a) \in \mathbb{K}\ . \tag{3.3}$$

Observemos además, de la Proposición 1.1(a)*,*

$$\rho_\alpha := \Delta_\alpha < 1. \tag{3.4}$$

Recuerde que para cada $\omega \in \Omega$, las variables $\delta_1(\omega), \delta_2(\omega), ..., \delta_n(\omega)$ representan los tiempos de permanencia hasta el momento de la n-ésima época de decisión; por consiguiente, la condición (3.1) implica que $\delta_1, \delta_2, ..., \delta_n$ son variables aleatorias i.i.d. con densidad común g.

3.3. Estimación de la Densidad

Sea $\hat{g}_n(s) := \hat{g}_n(s; \delta_1, \delta_2, ..., \delta_n)$, $s \in \mathbb{R}_+$, un estimador arbitrario de g tal que para cada $\pi \in \Pi$ y $x \in \mathbb{X}$,

$$E_x^\pi\left[\left(\|g - \hat{g}_n\|_q\right)^{\frac{q}{2}}\right] \to 0 \quad \textit{cuando} \quad n \to \infty. \tag{3.5}$$

(Véase [6].)

Sea λ un número real tal que $\Delta_\alpha \leq \lambda < 1$ (véase (3.4)). Definimos el conjunto $D \subset L_q([0,\infty))$ como:

$$\begin{aligned} D \quad : \quad = \Big\{ \mu : \mu \ &\textit{es una densidad en } L_q([0,\infty)), \\ &\int_0^\infty e^{-\alpha s}\mu(s)ds \leq \lambda, \\ &\qquad \int_0^\theta \mu(s)ds < 1-\varepsilon, \\ &\qquad\qquad \mu(s) \leq \bar{g}(s) \ \ c.d. \Big\} \end{aligned} \tag{3.6}$$

Nótese que $g \in D$. Algunas propiedades del conjunto D se establecen en el siguiente resultado, el cual será demostrado al final del presente capítulo.

Lema 3.1 *El conjunto D es cerrado y convexo en $L_q([0,\infty))$.*

Ahora, tomando en cuenta este resultado y aplicando las Proposiciones 2 y 3 en [11], p. 343, existe una única densidad $g_n \in D$ tal que satisface

$$\|g_n - \hat{g}_n\|_q = \inf_{\mu \in D} \|\mu - \hat{g}_n\|_q, \quad n \in \mathbb{N}. \tag{3.7}$$

Es decir, la densidad g_n es la "mejor aproximación" del estimador $\hat{g}_n$ sobre el conjunto D. Más aún, g_n es un estimador consistente de la densidad g, tal como lo muestra el siguiente resultado.

Proposición 3.1 *Para cada $\pi \in \Pi$ y $x \in \mathbb{X}$ se cumple lo siguiente:*

$$E_x^\pi \left[\int_0^\infty |g(s) - g_n(s)|\, ds \right] \to 0 \quad \text{cuando } n \to \infty. \tag{3.8}$$

Este resultado también será demostrado al final de este capítulo.

Definición 3.1 *Sea $\alpha > 0$ un factor de descuento fijo. Para cada $n \in \mathbb{N}$, se define*

$$\Delta_n := \int_0^\infty e^{-\alpha s} g_n(s) ds, \tag{3.9}$$

$$\tau_n := \frac{1 - \Delta_n}{\alpha}. \tag{3.10}$$

Por lo anterior y de (1.15) puede observarse que $\Delta_n < 1$, y por consiguiente $\tau_n < 1/\alpha$. Además, cuando $n \to \infty$, (3.8) implica que para cada $\pi \in \Pi$ y $x \in \mathbb{X}$,

$$E_x^\pi |\Delta_\alpha - \Delta_n| \to 0 \tag{3.11}$$

y

$$E_x^\pi |\tau_\alpha - \tau_n| \to 0. \tag{3.12}$$

3.4. Construcción de Políticas

Definición 3.2 *Para cada $n \in \mathbb{N}$, se define la función de costo por etapa aproximado como*

$$c_n(x,a) = D(x,a) + \tau_n d(x,a), \quad (x,a) \in \mathbb{K}. \tag{3.13}$$

Nótese que por la Hipótesis 1.2*(a)*, para cada $n \in \mathbb{N}$ y $x \in \mathbb{X}$

$$\sup_{a \in A(x)} |c_n(x,a)| \leq \bar{c}W(x), \tag{3.14}$$

donde $\bar{c} = \bar{c}_1 + \frac{\bar{c}_2}{\alpha}$ (véase Proposición 2.2, Capítulo 2).

Además, definimos la sucesión $\{V_n\}$ de funciones en $\mathbb{B}_W(\mathbb{X})$ para cada $x \in \mathbb{X}$ como:

$$\begin{aligned} V_0(x) \quad &: \quad = 0 \quad \textit{y para} \quad n \geq 1, \\ V_n(x) \quad &: \quad = \min_{a \in A(x)} \left\{ c_n(x,a) + \Delta_n \int_{\mathbb{X}} V_{n-1}(y) Q(dy \mid x,a) \right\}. \end{aligned} \tag{3.15}$$

De hecho, para todo $n \in \mathbb{N}$ y $x \in \mathbb{X}$ (véase (2.9)),

$$|V_n(x)| \leq \bar{c} \left(1 + \frac{b}{1-\beta}\right) \left(\frac{1}{1-\lambda}\right) W(x) \tag{3.16}$$

donde λ fue introducida en (3.6).

Aplicando el Teorema 2.1*(c)*, sobre la existencia de minimizadores, se tiene que bajo las Hipótesis 1.1, 1.2 y 3.1, para cada $n \in \mathbb{N}$, existe $f_n = f_n^{g_n} \in \mathbb{F}$ tal que

$$V_n(x) = c_n(x, f_n) + \Delta_n \int_{\mathbb{X}} V_{n-1}(y) Q(dy \mid x, f_n), \quad x \in \mathbb{X}, \tag{3.17}$$

donde la minimización es para cada $\omega \in \Omega$.

Teorema 3.1 *Sea $\hat{\pi} = \{\hat{\pi}_n\}$ la política definida por $\hat{\pi}_n(h_n) = \hat{\pi}_n(h_n; g_n) := f_n(x_n)$, $n \in \mathbb{N}$ y $\hat{\pi}_0$ alguna acción fija. Entonces, bajo las Hipótesis 1.1, 1.2, 1.3, 3.1 y 3.2, $\hat{\pi}$ es AOD.*

La demostración de este Teorema requiere de los dos resultados siguientes.

Lema 3.2 *Bajo las hipótesis 1.1, 1.2, 1.3, 3.1 y 3.2,*

$$\lim_{n \to \infty} E_x^{\pi} \|V_n - V_\alpha^*\|_W = 0 \quad \forall \pi \in \Pi, x \in \mathbb{X}. \tag{3.18}$$

Demostración. Primero definamos los operadores

$$Tu\,(x) := \min_{a\in A(x)} \left\{ c(x,a) + \Delta_\alpha \int_{\mathbb{X}} u(y) Q(dy \mid x,a) \right\} \tag{3.19}$$

y

$$T_m u\,(x) := \min_{a\in A(x)} \left\{ c_m(x,a) + \Delta_m \int_{\mathbb{X}} u(y) Q(dy \mid x,a) \right\} \tag{3.20}$$

para $x \in \mathbb{X}$, $u \in \mathbb{B}_W(\mathbb{X})$ y $m \in \mathbb{N}$. Observe que por las Hipótesis 1.1 y 1.2, junto con (3.14), T y T_m mapean $\mathbb{B}_W(\mathbb{X})$ en sí mismo. Además de (2.10) y (3.15)

$$TV_\alpha^* = V_\alpha^* \quad \text{y} \quad T_n V_{n-1} = V_n, \; n \in \mathbb{N}. \tag{3.21}$$

Ahora fijemos $\rho \in (\lambda, 1)$ arbitrario (recuérdese la definición de λ en (3.6)) y definamos $\bar{W}(x) := W(x) + d$, para $x \in \mathbb{X}$, donde $d := b\,(\rho/\lambda - 1)^{-1}$. Además, sea $\mathbb{B}_{\bar{W}}(\mathbb{X})$ el espacio de todas las funciones medibles $u : \mathbb{X} \to \mathbb{R}$ cuya norma se define como

$$\|u\|_{\bar{W}} := \sup_{x\in\mathbb{X}} \frac{|u(x)|}{\bar{W}(x)} < \infty.$$

Nótese que como $\rho > \lambda$ y $b > 0$ tenemos que $d > 0$. De aquí,

$$\|u\|_{\bar{W}} := \sup_{x\in\mathbb{X}} \frac{|u(x)|}{W(x)+d} \leq \sup_{x\in\mathbb{X}} \frac{|u(x)|}{W(x)} =: \|u\|_W \tag{3.22}$$

Pero además, como $W(x) \geq 1$ entonces

$$\begin{aligned} \|u\|_W &= \sup_{x\in\mathbb{X}} \left\{ \frac{|u(x)|}{W(x)} \cdot \frac{\bar{W}(x)}{\bar{W}(x)} \right\} \\ &= \sup_{x\in\mathbb{X}} \left\{ \frac{|u(x)|}{\bar{W}(x)} \cdot \frac{W(x)+d}{W(x)} \right\} \\ &= \sup_{x\in\mathbb{X}} \left\{ \frac{|u(x)|}{\bar{W}(x)} \cdot \left(1 + \frac{d}{W(x)}\right) \right\} \\ &\qquad \leq \|u\|_{\bar{W}}\,(1+d)\,. \end{aligned}$$

De ésto y por (3.22) tenemos que

$$\|u\|_{\bar{W}} \leq \|u\|_W \leq \|u\|_{\bar{W}}\,(1+d)\,, \tag{3.23}$$

lo cual implica que $\|\cdot\|_W$ y $\|\cdot\|_{\bar{W}}$ son equivalentes y, por consiguiente, la demostración de (3.18) se reduce a mostrar que

$$\lim_{n\to\infty} E_x^{\pi} \left\| V_n - V_\alpha^* \right\|_{\bar{W}} = 0 \quad \forall \pi \in \Pi, x \in \mathbb{X}. \tag{3.24}$$

Una consecuencia del Lema 2 en [29] es que la desigualdad (1.11) implica que, para cada $m \in \mathbb{N}$, el operador T_m es una contracción con respecto a la norma $\|\cdot\|_{\bar{W}}$ con módulo ρ, ésto es

$$\left\| T_m u - T_m v \right\|_{\bar{W}} \leq \rho \left\| u - v \right\|_{\bar{W}}, \quad \forall\, u, v \in \mathbb{B}_W(\mathbb{X}),\ m \in \mathbb{N}. \tag{3.25}$$

De modo que por (3.21), para cada $n \in \mathbb{N}$

$$\begin{aligned} \left\| V_\alpha^* - V_{n+1} \right\|_{\bar{W}} &= \left\| TV_\alpha^* - T_{n+1} V_n \right\|_{\bar{W}} \\ &\leq \left\| TV_\alpha^* - T_{n+1} V_\alpha^* \right\|_{\bar{W}} + \left\| T_{n+1} V_\alpha^* - T_{n+1} V_n \right\|_{\bar{W}} \\ &\leq \left\| TV_\alpha^* - T_{n+1} V_\alpha^* \right\|_{\bar{W}} + \rho \left\| V_\alpha^* - V_n \right\|_{\bar{W}} \end{aligned} \tag{3.26}$$

Por otro lado, de (3.13) y (3.3) se deduce que para cada $x \in \mathbb{X}$ y $n \in \mathbb{N}$,

$$\begin{aligned} \sup_{a\in A(x)} |c(x,a) - c_n(x,a)| &\leq \bar{c}_2 \left|\tau_\alpha - \tau_n\right| W(x) && (3.27) \\ &\leq \bar{c}_2 \left|\tau_\alpha - \tau_n\right| \{W(x) + d\} \\ &\leq \bar{c}_2 \left|\tau_\alpha - \tau_n\right| \bar{W}(x) && (3.28) \end{aligned}$$

Ahora obsérvese que de (3.19) y (3.20), así como por las relaciones (2.9) y (1.11),

$$\begin{aligned} \left| TV_\alpha^*(x) - T_{n+1} V_\alpha^*(x) \right| &= \Bigg| \min_{a\in A(x)} \left\{ c(x,a) + \Delta_\alpha \int_{\mathbb{X}} V_\alpha^*(y) Q(dy \mid x,a) \right\} \\ &\quad - \min_{a\in A(x)} \Big\{ c_{n+1}(x,a) \\ &\qquad + \Delta_{n+1} \int_{\mathbb{X}} V_\alpha^*(y) Q(dy \mid x,a) \Big\} \Bigg| \\ &\leq \sup_{a\in A(x)} \Big\{ \left| c(x,a) - c_{n+1}(x,a) \right| \\ &\qquad + \left| \Delta_\alpha - \Delta_{n+1} \right| \int_{\mathbb{X}} V_\alpha^*(y) Q(dy \mid x,a) \Big\} \\ &\leq \bar{c}_2 \left| \tau_\alpha - \tau_{n+1} \right| \bar{W}(x) \\ &\qquad + \left| \Delta_\alpha - \Delta_{n+1} \right| M \int_{\mathbb{X}} W(y) Q(dy \mid x,a) \Big\} \\ &\leq \bar{c}_2 \left| \tau_\alpha - \tau_{n+1} \right| \bar{W}(x) \\ &\qquad + \left| \Delta_\alpha - \Delta_{n+1} \right| M \left\{ \beta W(x) + b \right\} \Big\}. \end{aligned}$$

De aquí, como $\bar{W}(\cdot) > W(\cdot) \geq 1$,

$$\begin{aligned} \|TV_\alpha^* - T_{n+1}V_\alpha^*\|_{\bar{W}} \quad \leq \quad & \bar{c}_2 \left|\tau_\alpha - \tau_{n+1}\right| \\ & +M\left(\beta+b\right)\left|\Delta_\alpha - \Delta_{n+1}\right|, \qquad n \in \mathbb{N}. \qquad (3.29) \end{aligned}$$

Combinando (3.26) con (3.29), y tomando esperanza se obtiene que para cada $n \in \mathbb{N}$, $\pi \in \Pi$ y $x \in \mathbb{X}$,

$$\begin{aligned} E_x^\pi \left\|V_\alpha^* - V_{n+1}\right\|_{\bar{W}} \quad \leq \quad & \bar{c}_2 E_x^\pi \left|\tau_\alpha - \tau_{n+1}\right| \\ & + M\left(\beta+b\right) E_x^\pi \left|\Delta_\alpha - \Delta_{n+1}\right| \\ & + \rho E_x^\pi \left\|V_\alpha^* - V_n\right\|_{\bar{W}}. \qquad (3.30) \end{aligned}$$

Por otro lado, de (2.9) y (3.16)

$$\begin{aligned} E_x^\pi \left\|V_\alpha^* - V_{n+1}\right\|_{\bar{W}} \quad = \quad & E_x^\pi \left[\sup_{x\in\mathbb{X}} \frac{\left|V_\alpha^*(x) - V_{n+1}(x)\right|}{\bar{W}(x)}\right] \\ & \leq E_x^\pi \left[\sup_{x\in\mathbb{X}} \frac{\left|V_\alpha^*(x)\right| + \left|V_{n+1}(x)\right|}{W(x)+d}\right] \\ & \leq E_x^\pi \left[\sup_{x\in\mathbb{X}} \frac{MW(x) + \bar{c}\left(1+\frac{b}{1-\beta}\right)\left(\frac{1}{1-\lambda}\right)W(x)}{W(x)}\right] \\ & \qquad \forall \pi \in \Pi, \quad x \in \mathbb{X}. \end{aligned}$$

Ésto es, para cada $\pi \in \Pi$ y $x \in \mathbb{X}$,

$$E_x^\pi \left\|V_\alpha^* - V_{n+1}\right\|_{\bar{W}} \leq M + \bar{c}\left(1+\frac{b}{1-\beta}\right)\left(\frac{1}{1-\lambda}\right),$$

de donde

$$\eta := \limsup_{n\to\infty} E_x^\pi \left\|V_\alpha^* - V_{n+1}\right\|_{\bar{W}} < \infty \qquad \forall \pi \in \Pi, \quad x \in \mathbb{X}.$$

En consecuencia, de (3.11) y (3.12), tomando límite superior cuando $n \to \infty$ en ambos lados de (3.30) se obtiene que $\eta \leq \rho\eta$, lo cual implica que $\eta = 0$ debido a que $\rho < 1$. Ésto demuestra (3.24), y por lo tanto el Lema 3.2.

■

Para la demostración del siguiente resultado haremos uso de las relaciones

$$|u(x)| \leq \|u\|_W W(x) \qquad (3.31)$$

y

$$\int_{\mathbb{X}} |u(y)|\, Q(dy \mid x, a) \leq \|u\|_W \left[\beta W(x) + b\right], \qquad \forall x \in \mathbb{X}, \quad a \in A(x), \qquad (3.32)$$

las cuales son consecuencia de (1.8) y (1.11).

Ahora definamos la *función de discrepancia aproximada* Φ_n (véase el Teorema 2.2) para cada $n \in \mathbb{N}$, $(x, a) \in \mathbb{K}$, como

$$\Phi_n(x,a) \quad : \quad = c_n(x,a) + \Delta_n \int_{\mathbb{X}} V_{n-1}(y)\, Q(dy \mid x,a) - V_n(x), \tag{3.33}$$

y denotemos para $n \in \mathbb{N}$,

$$\Psi_n \quad : \quad = \sup_{x \in \mathbb{X}} \left\{ [W(x)]^{-1} \sup_{a \in A(x)} |\Phi(x,a) - \Phi_n(x,a)| \right\}. \tag{3.34}$$

Lema 3.3 *Bajo las hipótesis 1.1, 1.2, 1.3, 3.1 y 3.2, para cada $x \in \mathbb{X}$ y $\pi \in \Pi$,*

$$\lim_{n\to\infty} E_x^\pi [\Psi_n] = 0. \tag{3.35}$$

Además,

$$\lim_{n\to\infty} E_x^\pi [W(x_n)\, \Psi_n] = 0. \tag{3.36}$$

Demostración. De las definiciones de Φ y Φ_n, sumando y restando el término $\Delta_n \int_{\mathbb{X}} V_\alpha^*(y)\, Q(dy \mid x,a)$, así como reacomodando términos y aplicando la Desigualdad del Triángulo, se observa que

$$\begin{aligned} |\Phi(x,a) - \Phi_n(x,a)| \leq{} & |c(x,a) - c_n(x,a)| + |V_n(x) - V_\alpha^*(x)| \\ & + \Delta_n \int_{\mathbb{X}} |V_\alpha^*(y) - V_{n-1}(y)|\, Q(dy \mid x,a) \\ & + |\Delta_\alpha - \Delta_n| \int_{\mathbb{X}} V_\alpha^*(y)\, Q(dy \mid x,a). \end{aligned} \tag{3.37}$$

De este modo, usando conjuntamente las relaciones (3.27), (3.31), (3.32), (2.9), (1.11) y dado que $\Delta_n < 1$, $n \in \mathbb{N}$, de lo anterior se tiene

$$\begin{aligned} |\Phi(x,a) - \Phi_n(x,a)| \quad \leq \quad & \bar{c}_2 |\tau_\alpha - \tau_n| W(x) + \|V_n - V_\alpha^*\|_W W(x) \\ & + \|V_\alpha^* - V_{n-1}\|_W \{\beta W(x) + b\} \\ & + |\Delta_\alpha - \Delta_n| M \{\beta W(x) + b\}. \end{aligned} \tag{3.38}$$

Luego

$$\begin{aligned} \Psi_n \quad \leq \quad & \bar{c}_2 |\tau_\alpha - \tau_n| + \|V_n - V_\alpha^*\|_W \\ & + (\beta + b) \|V_\alpha^* - V_{n-1}\|_W + M(\beta + b) |\Delta_\alpha - \Delta_n|, \end{aligned}$$

de donde, para cada $x \in \mathbb{X}$ y $\pi \in \Pi$,

$$\begin{aligned} E_x^\pi [\Psi_n] \quad &\leq \quad \bar{c}_2 E_x^\pi |\tau_\alpha - \tau_n| + E_x^\pi \|V_n - V_\alpha^*\|_W \\ &\qquad + (\beta + b) E_x^\pi \|V_\alpha^* - V_{n-1}\|_W + M(\beta + b) E_x^\pi |\Delta_\alpha - \Delta_n| . \end{aligned}$$

Por lo tanto, aplicando (3.11), (3.12) y (3.18) demostramos la relación (3.35).

Con respecto a la demostración de (3.36), obsérvese en primer lugar que de la Desigualdad del Triángulo, así como por la Definición 1.6 junto con la relación (1.15) y la Observación 3.2, para cada $n \in \mathbb{N}$

$$|\tau_\alpha - \tau_n| \leq |\tau_\alpha| + |\tau_n| < \frac{2}{\alpha}$$

y

$$|\Delta_\alpha - \Delta_n| \leq |\Delta_\alpha| + |\Delta_n| < 2.$$

De acuerdo a (3.38) se tiene

$$\begin{aligned} |\Phi(x,a) - \Phi_n(x,a)| \quad &\leq \quad \frac{2\bar{c}_2}{\alpha} W(x) + \|V_n - V_\alpha^*\|_W W(x) \\ &\qquad + \|V_\alpha^* - V_{n-1}\|_W \{\beta W(x) + b\} \\ &\qquad\qquad + 2M\{\beta W(x) + b\}, \end{aligned}$$

lo cual implica (véase Lema 1.1, Capítulo 1)

$$\sup_{n\in\mathbb{N}} \Psi_n \leq M_1 < \infty \tag{3.39}$$

para alguna constante M_1. Además, de (3.35) obtenemos la convergencia en probabilidad

$$\Psi_n \overset{P_x^\pi}{\to} 0 \quad \textit{cuando} \quad n \to \infty. \tag{3.40}$$

Ahora, nótese que por (3.39)

$$\sup_{n\in\mathbb{N}} E_x^\pi [W(x_n)\Psi_n]^p \leq (M_1)^p \sup_{n\in\mathbb{N}} E_x^\pi [W^p(x_n)],$$

y por consiguiente, aplicando el Lema 1.1 se deduce que

$$\sup_{n\in\mathbb{N}} E_x^\pi [W(x_n)\Psi_n]^p < \infty. \tag{3.41}$$

Entonces por la Proposición D.3 (véase Apéndice D) se tiene que $\{W(x_n)\Psi_n\}$ es $P_x^\pi-$uniformemente integrable.

Por otra parte, obsérvese que para números arbitrarios positivos l_1 y l_2,

$$\begin{aligned} \left[W(x_n)\Psi_n > l_1\right] \quad &= \quad \left[W(x_n)\Psi_n > l_1, W(x_n) \leq l_2\right] \\ &\qquad \cup \left[W(x_n)\Psi_n > l_1, W(x_n) > l_2\right] \\ &\subset \quad \left[\Psi_n > \frac{l_1}{W(x_n)}, W(x_n) \leq l_2\right] \\ &\qquad\qquad \cup \left[W(x_n) > l_2\right], \end{aligned}$$

de donde

$$\left[W(x_n)\Psi_n > l_1\right] \subset \left[\Psi_n > \frac{l_1}{l_2}\right] \cup \left[W(x_n) > l_2\right],$$

así que

$$P_x^\pi\left[W(x_n)\Psi_n > l_1\right] \leq P_x^\pi\left[\Psi_n > \frac{l_1}{l_2}\right] + P_x^\pi\left[W(x_n) > l_2\right], \quad n \in \mathbb{N}.$$

Aplicando la Desigualdad de Chebyshev sobre el segundo término en la parte derecha de esta desigualdad se obtiene

$$P_x^\pi\left[W(x_n)\Psi_n > l_1\right] \quad \leq \quad P_x^\pi\left[\Psi_n > \frac{l_1}{l_2}\right] + \frac{E_x^\pi[W(x_n)]}{l_2}, \quad n \in \mathbb{N}, \tag{3.42}$$

ya que por definición $W(\cdot) > 0$. Por (3.40) y el Lema 1.1, obtenemos que $\{W(x_n)\Psi_n\}$ converge a cero en probabilidad, es decir,

$$W(x_n)\Psi_n \stackrel{P_x^\pi}{\to} 0 \quad \textit{cuando} \quad n \to \infty. \tag{3.43}$$

Finalmente (3.36) se sigue de (3.43) y del hecho de que $\{W(x_n)\Psi_n\}$ es P_x^π-uniformemente integrable.

■

3.4.1. Demostración del Teorema 3.1.

Obsérvese que de la definición de la política $\hat{\pi} = \{\hat{\pi}_n\}$ así como la correspondiente a la función Φ_n, se tiene que $\Phi_n(\cdot, \hat{\pi}_n(\cdot)) = 0$, $n \in \mathbb{N}$, (véase (3.17)). Por consiguiente,

$$\begin{aligned}
\Phi(x_n, \hat{\pi}_n(h_n)) &= |\Phi(x_n, \hat{\pi}_n(h_n)) - \Phi_n(x_n, \hat{\pi}_n(h_n))| \\
&\leq \sup_{A(x_n)} |\Phi(x_n, a) - \Phi_n(x_n, a)| \\
&\leq W(x_n) \sup_{\mathbb{X}} [W(x)]^{-1} \sup_{A(x)} |\Phi(x, a) - \Phi_n(x, a)| \\
&\leq W(x_n)\Psi_n, \quad n \in \mathbb{N},
\end{aligned}$$

de donde,

$$E_x^{\hat{\pi}}\left[\Phi(x_n, a_n)\right] \leq E_x^{\hat{\pi}}\left[W(x_n)\Psi_n\right], \quad n \in \mathbb{N}.$$

Y puesto que Φ es no negativa, la relación (3.36) implica que

$$E_x^{\hat{\pi}}\left[\Phi(x_n, a_n)\right] \to 0 \quad \textit{cuando} \quad n \to \infty.$$

En consecuencia, por el Teorema 2.2 se deduce la optimalidad asintótica de $\hat{\pi}$.

■

3.5. Comportamiento Límite de Minimizadores

Concluimos el capítulo presentando un análisis sobre el comportamiento límite de la sucesión de minimizadores $\{f_n\}$ definida en (3.17). Para ésto, primero introduciremos la siguiente definición.

Definición 3.3 *Diremos que una sucesión $\{f_n\} \subset \mathbb{F}$ converge en el sentido de Schäl [27] si existe $f_\infty \in \mathbb{F}$ tal que $f_\infty(x) \in A(x)$ es un punto de acumulación de $\{f_n(x)\} \subset A(x)$ para cada $x \in \mathbb{X}$. Es decir, para cada $x \in \mathbb{X}$ existe una subsucesión $\{n_i(x)\}$ de $\{n\}$ tal que*

$$f_{n_i(x)}(x) \to f_\infty(x) \quad \text{cuando} \quad i \to \infty.$$

Por el resultado de Schäl (en [27]), el cual es reproducido en [8] Proposición *D.7*, tenemos que si $A(x)$ es compacto para cada $x \in \mathbb{X}$, entonces cualquier sucesión en $\mathbb{F}$ converge en el sentido de Schäl. De aquí, bajo la Hipótesis 1.1 $A(x)$ es compacto, entonces existe una política estacionaria $\{f_\infty\}$ tal que para cada $x \in \mathbb{X}$, $f_\infty(x) \in A(x)$ es un punto de acumulación de la sucesión $\{f_n(x)\}$ definida en (3.17).

Teorema 3.2 *Bajo las hipótesis 1.1, 1.2, 1.3, 3.1 y 3.2, la política estacionaria $\{f_\infty\}$ es óptima para el MCSM.*

En la demostración de este resultado se utilizará el lema que se enuncia y demuestra a continuación.

Lema 3.4 *Bajo las hipótesis 1.1, 1.2, 1.3, 3.1 y 3.2, para cada $x \in \mathbb{X}$ y $\pi \in \Pi$ se satisface*

$$\lim_{n\to\infty} E_x^\pi \left[\sup_{A(x)} \left| \Delta_n \int_{\mathbb{X}} V_{n-1}(y) Q(dy \mid x,a) - \Delta_\alpha \int_{\mathbb{X}} V_\alpha^*(y) Q(dy \mid x,a) \right| \right] = 0 \ . \tag{3.44}$$

Demostración. Obsérvese primero que para cada $n \in \mathbb{N}$, $x \in \mathbb{X}$ y $a \in A(x)$, sumando y restando el término $\Delta_n \int_{\mathbb{X}} V_\alpha^*(y) Q(dy \mid x,a)$, y usando el hecho de que $\Delta_n < 1$ para cada $n \in \mathbb{N}$, tenemos

$$\begin{aligned} \left| \Delta_n \int_{\mathbb{X}} V_{n-1}(y) Q(dy \mid x,a) - \Delta_\alpha \int_{\mathbb{X}} V_\alpha^*(y) Q(dy \mid x,a) \right| & \\ \leq \int_{\mathbb{X}} \left| V_\alpha^*(y) - V_{n-1}(y) \right| Q(dy \mid x,a) & \\ + \left| \Delta_n - \Delta_\alpha \right| \int_{\mathbb{X}} V_\alpha^*(y) Q(dy \mid x,a). & \end{aligned}$$

Siguiendo argumentos similares a los usados en (3.37) y (3.38) se observa

$$\sup_{a\in A(x)}\left|\Delta_n\int_{\mathbb{X}}V_{n-1}(y)Q(dy\mid x,a)-\Delta_\alpha\int_{\mathbb{X}}V_\alpha^*(y)Q(dy\mid x,a)\right|$$
$$\leq \|V_\alpha^*-V_{n-1}\|_W\{\beta W(x)+b\}$$
$$+|\Delta_\alpha-\Delta_n|\,M\{\beta W(x)+b\}.$$

Tomando E_x^π en ambos lados de esta desigualdad, y aplicando (3.11) y (3.18) se obtiene

$$\lim_{n\to\infty}E_x^\pi\left[\sup_{A(x)}\left|\Delta_n\int_{\mathbb{X}}V_{n-1}(y)Q(dy\mid x,a)-\Delta_\alpha\int_{\mathbb{X}}V_\alpha^*(y)Q(dy\mid x,a)\right|\right]=0$$
$$\forall\pi\in\Pi, x\in\mathbb{X}.$$

■

Demostración del Teorema 3.2.

Para demostrar que la política estacionaria $\{f_\infty\}$ es óptima procedamos como sigue.

Fijemos un $x\in\mathbb{X}$ arbitrario, y recordemos que

$$f_{n_i(x)}(x)\to f_\infty(x)\quad \textit{cuando}\quad i\to\infty.$$

Además, de (3.17) y haciendo $n_i=i$ (por conveniencia notacional) se tiene

$$V_i(x)=c_i(x,f_i)+\Delta_i\int_{\mathbb{X}}V_{i-1}(y)Q(dy\mid x,f_i)\,. \tag{3.45}$$

Ahora bien,

$$\begin{aligned}\Delta_i\int_{\mathbb{X}}V_{i-1}(y)Q(dy\,|x,f_i) &= \Delta_i\int_{\mathbb{X}}V_{i-1}(y)Q(dy\,|x,f_i)\\ &\quad-\Delta_\alpha\int_{\mathbb{X}}V_\alpha^*(y)Q(dy\,|x,f_i)\\ &\quad+\Delta_\alpha\int_{\mathbb{X}}V_\alpha^*(y)Q(dy\,|x,f_i)\,,\quad i\in\mathbb{N}.\end{aligned}$$

Tomando E_x^π,

$$\begin{aligned}E_x^\pi\left[\Delta_i\int_{\mathbb{X}}V_{i-1}(y)Q(dy\,|x,f_i)\right] &= E_x^\pi\left[\Delta_i\int_{\mathbb{X}}V_{i-1}(y)Q(dy\,|x,f_i)\right.\\ &\quad\left.-\Delta_\alpha\int_{\mathbb{X}}V_\alpha^*(y)Q(dy\,|x,f_i)\right]\\ &\quad+E_x^\pi\left[\Delta_\alpha\int_{\mathbb{X}}V_\alpha^*(y)Q(dy\,|x,f_i)\right],\\ &\qquad i\in\mathbb{N},\ \forall\pi\in\Pi, x\in\mathbb{X};\end{aligned}$$

de lo cual, por (3.44) se obtiene

$$\liminf_{i\to\infty} E_x^\pi \left[\Delta_i \int_{\mathbb{X}} V_{i-1}(y)Q(dy \mid x, f_i)\right]$$
$$= \liminf_{i\to\infty} E_x^\pi \left[\Delta_\alpha \int_{\mathbb{X}} V_\alpha^*(y)Q(dy \mid x, f_i)\right]$$
$$\forall \pi \in \Pi,\ x \in \mathbb{X}.$$

Pero por el Lema de Fatou "Generalizado" y la Hipótesis 1.1*(a)*, de la parte derecha en la expresión anterior se tiene que para cada $x \in \mathbb{X}$ y $\pi \in \Pi$,

$$\liminf_{i\to\infty} E_x^\pi \left[\Delta_\alpha \int_{\mathbb{X}} V_\alpha^*(y)Q(dy \mid x, f_i)\right] \geq \Delta_\alpha \int_{\mathbb{X}} V_\alpha^*(y)Q(dy \mid x, f_\infty),$$

es decir,

$$\liminf_{i\to\infty} E_x^\pi \left[\Delta_i \int_{\mathbb{X}} V_{i-1}(y)Q(dy \mid x, f_i)\right] \quad \geq \quad \Delta_\alpha \int_{\mathbb{X}} V_\alpha^*(y)Q(dy \mid x, f_\infty)$$
$$\forall \pi \in \Pi, x \in \mathbb{X}.$$

Luego, tomando esperanza E_x^π y límite inferior cuando $i \to \infty$ en (3.45) obtenemos

$$V_\alpha^*(x) \geq c(x, f_\infty) + \Delta_\alpha \int_{\mathbb{X}} V_\alpha^*(y)Q(dy \mid x, f_\infty). \tag{3.46}$$

Finalmente, dado que x es arbitrario, por (2.10) se cumple la igualdad en (3.46) para cada $x \in \mathbb{X}$, y en consecuencia, por el Teorema 2.1*(c)*, $\{f_\infty\}$ es óptima (para el *MCSM*).

■

3.6. Demostraciones

3.6.1. Demostración del Lema 3.1.

Para demostrar que D es cerrado, considérese una sucesión $\{\mu_n\} \subset D$ tal que

$$\mu_n \overset{L_q}{\to} \mu \quad \textit{cuando} \quad n \to \infty. \tag{3.47}$$

Por demostrar entonces que $\mu \in D$.

Primero se mostrará que $\mu(s) \leq \bar{g}(s)$ c.d., lo cual se hará por contradicción. Supóngase que existe $A \subset [0, \infty)$ con $m\,(A) > 0$ tal que $\mu(s) > \bar{g}(s)$, $s \in A$, donde m es la medida de Lebesgue en $\mathbb{R}$. Entonces, para algún $\varepsilon > 0$ y $A' \subset A$ con $m\,(A') > 0$ se tiene

$$\mu(s) > \bar{g}(s) + \varepsilon,\ s \in A'. \tag{3.48}$$

Por otro lado, dado que $\mu_n \in D$ para $n \in \mathbb{N}_0$, existe $B_n \subset [0,\infty)$ con $m(B_n) = 0$, tal que

$$\mu_n(s) \leq \bar{g}(s), \quad s \in [0,\infty)\backslash B_n, \ n \in \mathbb{N}_0. \tag{3.49}$$

Luego restando (3.48) de (3.49) se obtiene

$$\mu_n(s) - \mu(s) \leq -\varepsilon, \quad \forall s \in A' \cap ([0,\infty)\backslash B_n), \quad n \in \mathbb{N}_0, \tag{3.50}$$

de donde,

$$-\mu_n(s) + \mu(s) \geq \varepsilon, \quad \forall s \in A' \cap ([0,\infty)\backslash B_n), \quad n \in \mathbb{N}_0. \tag{3.51}$$

Así que de (3.50) y (3.51) se genera la siguiente relación

$$|\mu_n(s) - \mu(s)| \geq \varepsilon, \quad s \in A' \cap ([0,\infty)\backslash B_n), \quad n \in \mathbb{N}_0.$$

No obstante, del hecho de que $m(A' \cap ([0,\infty)\backslash B_n) = m(A') > 0$, se deduce entonces que μ_n no converge a μ en medida, lo cual es una contradicción a la suposición (3.47). Por lo tanto, $\mu(s) \leq \bar{g}(s)$ c.d.

Para demostrar que μ es una densidad, nótese que

$$\int_0^\infty \mu_n(s)ds = 1 \quad \forall n \in \mathbb{N},$$

debido a que $\mu_n \in D$ es una densidad para cada $n \in \mathbb{N}$. Esto implica que

$$\begin{aligned}
\left|1 - \int_0^\infty \mu(s)ds\right| &= \left|\int_0^\infty \mu_n(s)ds - \int_0^\infty \mu(s)ds\right| \\
&= \left|\int_0^\infty \{\mu_n(s) - \mu(s)\}\, ds\right| \\
&\leq \int_0^\infty |\mu_n(s) - \mu(s)|\, ds.
\end{aligned}$$

Pero como

$$|\mu_n(s) - \mu(s)| = |\mu_n(s) - \mu(s)|^{\frac{2-q}{2}} \, |\mu_n(s) - \mu(s)|^{\frac{q}{2}},$$

entonces aplicando la Desigualdad de Hölder se obtiene

$$\begin{aligned}
\left|1 - \int_0^\infty \mu(s)ds\right| &\leq \left\{\int_0^\infty \left(|\mu_n(s) - \mu(s)|^{\frac{2-q}{2}}\right)^2 ds\right\}^{\frac{1}{2}} \\
&\qquad \times \left\{\int_0^\infty \left(|\mu_n(s) - \mu(s)|^{\frac{q}{2}}\right)^2 ds\right\}^{\frac{1}{2}} \\
&\leq \left\{\int_0^\infty |\mu_n(s) - \mu(s)|^{2-q}\, ds\right\}^{\frac{1}{2}} \\
&\qquad \times \left\{\int_0^\infty |\mu_n(s) - \mu(s)|^{q}\, ds\right\}^{\frac{1}{2}}.
\end{aligned} \tag{3.52}$$

Por otra parte, de (3.47) se sabe que

$$\left\{\int_0^\infty |\mu_n(s)-\mu(s)|^q \, ds\right\}^{\frac{1}{q}} \to 0 \quad \textit{cuando} \quad n\to\infty, \tag{3.53}$$

mientras que, por la Desigualdad del Triángulo y usando el hecho de que $\mu(s) \leq \bar{g}(s)$ c.d. (previamente demostrado) se tiene

$$\begin{aligned} |\mu_n(s)-\mu(s)|^{2-q} &\leq (|\mu_n(s)|+|\mu(s)|)^{2-q} \\ &\leq (2\bar{g}(s))^{2-q}. \end{aligned} \tag{3.54}$$

Entonces, considerando (3.53) y (3.54) en la relación (3.52) se deduce que

$$\begin{aligned} \left|1-\int_0^\infty \mu(s)ds\right| &\leq \left\{\int_0^\infty (2\bar{g}(s))^{2-q}\, ds\right\}^{\frac{1}{2}} \\ &\quad \times \left\{\left(\int_0^\infty |\mu_n(s)-\mu(s)|^q\, ds\right)^{\frac{1}{q}}\right\}^{\frac{q}{2}} \\ &\quad \to 0 \;\; \textit{cuando} \;\; n\to\infty. \end{aligned}$$

lo cual implica que $\int_0^\infty \mu(s)ds = 1$. Además, $\mu \geq 0$ c.d. debido a (3.47). Por lo tanto, μ es una densidad.

Siguiendo argumentos similares al procedimiento previo es posible demostrar que, cuando $n\to\infty$, obtenemos los resultados dados continuación:

$$\int_0^\theta \mu_n(s)ds \to \int_0^\theta \mu(s)ds, \tag{3.55}$$

$$\int_0^\infty e^{-\alpha s}\mu_n(s)ds \to \int_0^\infty e^{-\alpha s}\mu(s)ds, \tag{3.56}$$

relaciones que, a su vez implican respectivamente lo siguiente

$$\int_0^\theta \mu(s)ds \leq 1-\varepsilon \qquad \text{y} \qquad \int_0^\infty e^{-\alpha s}\mu(s)ds \leq \lambda.$$

Por consiguiente $\mu \in D$, y por lo tanto D es cerrado.

Para demostrar la convexidad de D, sean $\gamma \in [0,1]$ y $\mu_1, \mu_2 \in D$. Denotemos por $\mu(s) = \gamma\mu_1(s) + (1-\gamma)\mu_2(s)$.

Claramente,

$$\mu(s) \geq 0$$

y

$$\int_0^\infty \mu(s)ds = 1.$$

Además, ya que $\mu_1(s) \leq \bar{g}(s)$ c.d. y $\mu_2(s) \leq \bar{g}(s)$ c.d., tenemos que $\mu(s) \leq \bar{g}(s)$ c.d. Por último, aplicando las propiedades de linealidad de la integral, se demuestra fácilmente que

$$\int_0^\infty e^{-\alpha s}\mu(s)ds \leq \lambda \quad \text{y} \quad \int_0^\theta \mu(s)ds \leq 1-\varepsilon.$$

Por lo tanto, $\mu \in D$, lo cual implica que D es convexo.

■

3.6.2. Demostración de la Proposición 3.1.

Primero obsérvese que por (3.7),

$$\|g_n - \hat{g}_n\|_q \leq \|g - \hat{g}_n\|_q .$$

De aquí,

$$\begin{aligned} \|g - g_n\|_q &\leq \|g - \hat{g}_n\|_q + \|\hat{g}_n - g_n\|_q \\ &\leq 2\|g - \hat{g}_n\|_q \end{aligned} \tag{3.57}$$

Por otro lado, sea $M' < \infty$ tal que $\int_0^\infty (\bar{g}(s))^{2-q}\, ds \leq M'$ (véase Hipótesis 3.2, Sección 3.2), y como $g_n, g \in D$, tenemos que

$$g_n(s) \leq \bar{g}(s)\ c.s. \quad y \quad g(s) \leq \bar{g}(s)\ c.s.$$

Por lo tanto, debido a (3.57),

$$\begin{aligned} \int_0^\infty |g(s) - g_n(s)|\, ds &= \int_0^\infty |g(s) - g_n(s)|^{\frac{2-q}{2}} |g(s) - g_n(s)|^{\frac{q}{2}}\, ds \\ &\leq \left\{\int_0^\infty (2\bar{g}(s))^{2-q}\, ds\right\}^{\frac{1}{2}} \\ &\qquad \times \left\{\int_0^\infty |g(s) - g_n(s)|^q\, ds\right\}^{\frac{1}{2}} \\ &\qquad \leq 2^{\frac{2-q}{2}} (M')^{\frac{1}{2}} \|g - g_n\|_q^{\frac{q}{2}} \\ &\qquad\quad \leq 2^{\frac{2-q}{2}} (M')^{\frac{1}{2}}\, 2^{\frac{q}{2}} \|g - \hat{g}_n\|_q^{\frac{q}{2}} . \end{aligned}$$

Finalmente, tomando esperanza en ambos lados de la desigualdad anterior, la relación (3.5) implica (3.8).

■

Capítulo 4

Ejemplo: Un Sistema de Almacenamiento

En este capítulo presentamos un ejemplo de un sistema de almacenamiento para ilustrar la teoría desarrollada. En particular, mostraremos que se satisfacen las hipótesis que se impusieron en el modelo de control.

Consideremos un sistema de almacenamiento cuyas entradas están controladas de tal manera que, justo al tiempo en que se acumula cierta cantidad de producto $K > 0$ para su admisión al sistema, el controlador elige un valor $a \in [a_*, 1] =: A(x) \equiv \mathbb{A}$ $(0 < a_* < 1)$, que representa la porción de K que habrá de ser admitida; es decir, aK será la cantidad de producto admitido en el sistema.

Dada la situación anterior, $x_n \in \mathbb{X} := [0, \infty)$ representa la cantidad de producto en el sistema, mientras que $a_n \in A(x)$ son las decisiones tomadas por el controlador al tiempo de la n-ésima época de decisión T_n. Ahora bien, definiendo el tiempo de permanencia del sistema en el estado x_n como la variable aleatoria $\delta_{n+1} := T_{n+1} - T_n$ $(n \in \mathbb{N}_0)$, y asumiendo el número $b > 0$ (fijo) como la cantidad consumida por unidad de tiempo de dicho producto (tasa de consumo), nótese que $b\delta_{n+1}$ será entonces interpretado como la cantidad de producto que sale del sistema justo al tiempo de la $(n+1)$-ésima época de decisión.

Obsérvese entonces que conforme a lo descrito, este sistema evoluciona de acuerdo a la siguiente ecuación en diferencias

$$x_{n+1} = (x_n + a_n K - b\delta_{n+1})^+, \quad n \in \mathbb{N}_0.$$

De la descripción anterior, es claro que las épocas de decisión sólo dependen del tiempo en que se acumula la cantidad de producto K, y no de las variables estado-control. De aquí, naturalmente tenemos que la distribución de las variables aleatorias δ_{n+1} es independiente de (x_n, a_n).

Asumiremos que $\{\delta_n\}$ es una sucesión de variables aleatorias i.i.d. con densidad $g \in L_q([0, \infty))$, para alguna $q \in (1, 2)$, que satisface las Hipótesis 1.3 y

3.2. Respecto a este punto, nótese que la Hipótesis 1.3 (i.e. Observación 3.1) se cumple si $G(0) < 1$, donde (véase la Hipótesis 3.1)

$$G(t) = \int_0^t g(s)ds,$$

mientras que si tomamos

$$\bar{g}(s) := M' \min\left\{1, s^{-(1+r)}\right\} \qquad s > 0,$$

para alguna constante $r > 0$ apropiada, entonces la Hipótesis 3.2 se cumple para un gran número de densidades.

Para verificar las Hipótesis 1.1 y 1.2 necesitamos asumir lo siguiente.

Hipótesis 4.1

$$E[\delta_1] > \frac{K}{b}.$$

Lema 4.1 *Bajo la Hipótesis 4.1, existen constantes $\lambda_0 > 0$ y $p > 1$ tales que*

$$\Theta(p\lambda_0) < 1,$$

donde Θ es la función generadora de momentos de la variable aleatoria $K - b\delta$.

Demostración. Por definición se tiene que

$$\Theta(t) = E\left[e^{t(K-b\delta)}\right].$$

Ahora, como

$$\Theta'(0) = E[K - b\delta],$$

la Hipótesis 4.1 implica que $\Theta'(0) < 0$, es decir, Θ es una función decreciente en $t = 0$. Además, como $\Theta(0) = 1$, concluimos que existe $\lambda_0 > 0$ tal que $\Theta(\lambda_0) < 1$.

Finalmente, por la continuidad de Θ existe $p > 1$ tal que

$$\Theta(p\lambda_0) = E\left[e^{p\lambda_0(K-b\delta)}\right] < 1. \tag{4.1}$$

■

Para demostrar que se cumplen las Hipótesis 1.1 y 1.2, supondremos que las funciones de costo D y d son funciones arbitrarias s.c.i. en $A(x)$, tal que para alguna constante $\bar{K} \geq 1$ y para cada $x \in \mathbb{X}$,

$$\sup_{a \in A(x)} D(x,a) \leq \bar{K}e^{\lambda_0 x} \quad \text{y} \quad \sup_{a \in A(x)} d(x,a) \leq \bar{K}e^{\lambda_0 x}. \tag{4.2}$$

Teorema 4.1 *La Hipótesis 4.1 implica las Hipótesis 1.1 y 1.2.*

Demostración. De la descripción del ejemplo, es claro que las Hipótesis 1.1*(a)*, 1.1*(d)* y la relación (1.9) de la Hipótesis 1.2*(a)* se cumplen si definimos

$$W(x) := \bar{K}e^{\lambda_0 x}, \quad x \in \mathbb{X},$$

con λ_0 como en el Lema 4.1.

Para demostrar la relación (1.10), sea $p > 1$ (como en Lema 4.1). Entonces, para $(x,a) \in \mathbb{K}$,

$$\begin{aligned} \int_{\mathbb{X}} \bar{K}^p e^{\lambda_0 p y} Q(dy \mid x,a) &= \int_0^\infty \bar{K}^p e^{\lambda_0 p (x+aK-bs)^+} g(s)\, ds \\ &\leq \bar{K}^p P\left[x + aK - bs \leq 0\right] \\ &\quad + \left(\bar{K}e^{\lambda_0 x}\right)^p \int_0^\infty e^{\lambda_0 p (K-bs)} g(s)\, ds, \end{aligned} \tag{4.3}$$

ya que $a \leq 1$. Pero como

$$\int_0^\infty e^{\lambda_0 p (K-bs)} g(s)\, ds =: E\left[e^{\lambda_0 p (K-b\delta)}\right],$$

entonces por el Lema 4.1 (véase (4.1)) y la relación (4.3)

$$\int_{\mathbb{X}} \bar{K}^p e^{\lambda_0 p y} Q(dy \mid x,a) \leq W^p(x)\, \beta_0 + \bar{K}^p,$$

donde $\beta_0 := \Theta(p\lambda_0)$. Ésto demuestra la relación (1.10).

Para verificar la Hipótesis 1.1*(c)*, sea v una función medible y acotada sobre $\mathbb{X}$, y para cada $a \in A(x)$, sea g_a la densidad de $aK - b\delta$. Obsérvese que, por el Teorema de Cambio de Variable,

$$g_a(y) = \frac{1}{b} g\left(\frac{aK - y}{b}\right), \qquad -\infty < y \leq aK.$$

Además, para cada $y \in \mathbb{R}$, $a \mapsto g_a(y)$ es continua en $A(x)$. Ahora considérese lo siguiente,

$$\begin{aligned} \int_0^\infty v\left[(x+y)^+\right] g_a(y)\, dy &= v(0) \int_{-\infty}^{-x} g_a(y)\, dy + \int_{-x}^\infty v(x+y)\, g_a(y)\, dy \\ &= v(0) \int_{-\infty}^{-x} g_a(y)\, dy + \int_0^\infty v(y)\, g_a(y-x)\, dy. \end{aligned}$$

Esto implica que, para $a_1, a_2 \in A(x)$ arbitrarios,

$$\begin{aligned}
&\left| \int_0^\infty v\left[(x+y)^+\right] g_{a_1}(y)\,dy - \int_0^\infty v\left[(x+y)^+\right] g_{a_2}(y)\,dy \right| \\
&\qquad = \left| v(0) \int_{-\infty}^{-x} \left\{ g_{a_1}(y) - g_{a_2}(y) \right\} dy \right. \\
&\qquad\qquad \left. + \int_0^\infty v(y) \left\{ g_{a_1}(y-x) - g_{a_2}(y-x) \right\} dy \right| \\
&\qquad \leq |v(0)| \int_{-\infty}^{-x} \left| g_{a_1}(y) - g_{a_2}(y) \right| dy \\
&\qquad\qquad + \int_0^\infty |v(y)| \left| g_{a_1}(y-x) - g_{a_2}(y-x) \right| dy \\
&\qquad\qquad\qquad \to 0 \quad \textit{cuando} \quad a_1 \to a_2 \,.
\end{aligned}$$

De aquí, aplicando el Teorema de Scheffé (véase Apéndice D, Proposición D.2) tenemos que

$$a \longmapsto \int_{\mathbb{X}} v(y)\, Q(dy \mid x, a) = \int_0^\infty v\left[(x+y)^+\right] g_a(y)\,dy$$

define una función continua, lo cual demuestra que se cumple la Hipótesis 1.1*(c)*.

Finalmente, obsérvese que sustituyendo v por W en el procedimiento previo, y usando argumentos similares se verifica la Hipótesis 1.2*(e)*, con lo cual queda demostrado el teorema.

■

Observación 4.1 *Es importante observar que el proceso de construcción de las políticas asintóticamente óptimas, presentado en el Capítulo* 3*, dependen fuertemente del método de estimación de la densidad* g*. Además, para la construcción del estimador correspondiente, es necesario definir un estimador auxiliar que satisfaga la relación (3.5), a saber*

$$E\left[\left(\|g - \hat{g}_n\|_q\right)^{\frac{q}{2}}\right] \to 0 \quad \textit{cuando} \quad n \to \infty,$$

cuya proyección sobre el conjunto D *nos define el estimador apropiado. De hecho, un ejemplo de estimadores que satisfacen la relación (3.5) es el siguiente (véase [6]). Sea* $\{z_n\}$ *una sucesión de números reales positivos tal que*

$$\lim_{n\to\infty} \frac{z_n^k}{n} = 0$$

y

$$z_n \to \infty \quad \textit{cuando} \quad n \to \infty.$$

Definimos

$$\hat{g}_n(s) = \hat{g}_n(s; z_n, \delta_1, \delta_2, ..., \delta_n) := \frac{1}{n} \sum_{i=1}^n V_{z_n}(s - \delta_i)\,, \qquad s \in \mathbb{R}^k,$$

donde $V_z(y)$ *es un kernel del tipo de Vallée Poussin (véase, por ejemplo, [3]):*

$$V_z(y) = \prod_{n=1}^{k} \frac{\cos zy_n - \cos 2zy_n}{\pi z y_n^2}, \; y = (y_1, ..., y_k) \in \mathbb{R}^k, \; z > 0.$$

Capítulo 5

Conclusiones

En este trabajo se estudiaron los *PCSMs* bajo el criterio de costo descontado y con distribución del tiempo de permanencia desconocida. Aplicando técnicas que combinan métodos de estimación con los de control, se construyeron políticas de control asintóticamente óptimas.

La teoría aquí desarrollada puede ser considerada como una extensión a los trabajos de [2, 12, 13, 18, 23, 25].

La hipótesis clave para obtener los resultados consistió en que la distribución H es independiente de la pareja estado-control (x, a) (Hipótesis 3.1). Esta hipótesis fue necesaria para obtener observaciones independientes de los tiempos de permanencia δ_n durante la evolución del sistema, y de esta forma construir el estimador H_n. Aunque esta suposición podría ser fuerte, existen algunas situaciones en las cuales se satisface, tal como fue mostrado en el Capítulo 4 con el ejemplo del sistema de almacenamiento.

Otro punto que vale la pena observar es la existencia de una densidad de la distribución H. Esto implica que H no puede ser arbitraria, y por lo tanto la hipótesis de nuestro trabajo tiene esta limitación.

La teoría desarrollada puede ser aplicada a problemas de control markoviano, donde la función de costo c sea desconocida. En efecto, el hecho de que en un *PCSM* el costo dependa del tiempo de permanencia conlleva a que este sea parcialmente conocido, a través de su distribución.

En este sentido, en un proceso de control de Markov podemos suponer que el costo por etapa depende, además de la pareja estado-control, de una variable aleatoria cuya distribución es desconocida.

De esta manera, las técnicas de estimación que se implementaron en el presente trabajo pueden ser aplicables a estos procesos, y así construir políticas AOD.

Para la elaboración del trabajo se hizo uso de la siguiente bibliografía.

La teoría de control estocástico en general se estudió en [7, 8]. Los resultados sobre *PCSM* se revisaron en [12, 14, 18]. Para el análisis de las técnicas y métodos generales sobre estimación y control nos basamos en [5, 7, 10, 20, 21], y en particular, para los *PCSMs* con distribución del tiempo de permanencia desconocida se estudiaron [17, 22]. El resto de la bibliografía sirvió de apoyo.

Finalmente, plantearemos algunas ideas sobre posibles extensiones del trabajo.

- El proceso de construcción de las políticas se basa en la implementación de un método estadístico de estimación de densidades. Es decir, dicho proceso es válido si las variables aleatorias δ_n que representan los tiempos de permanencia tienen densidad. Una posible extensión del presente trabajo es suponer que la variable aleatoria δ tiene una distribución arbitraria. Esta situación nos llevaría a tener que implementar métodos de estimación más generales, como podría ser la distribución empírica. Trabajos relacionados con este tipo de técnicas, pero para procesos de control de Markov, se encuentran en [10].

- Cabe señalar que una limitación fuerte del trabajo, es precisamente la hipótesis de que la distribución de los tiempos de permanencia es independiente de las parejas estado-control. De aquí surge un nuevo problema, el cual consistiría en debilitar esta hipótesis.

- Este trabajo se ha enfocado al criterio de costo descontado. Podría ser de interés el problema de control semi-markoviano con distribución de los tiempos de permanencia desconocida, considerando el criterio de costo promedio. La base para abordar este problema podrían ser los trabajos de [15, 16, 19, 31].

Apéndice A

Espacios, Funciones y Multifunciones

Denotaremos por $\mathfrak{B}(\mathbb{X})$ a la σ-álgebra de Borel de un espacio topológico $\mathbb{X}$, es decir, la mínima σ-álgebra de subconjuntos de $\mathbb{X}$ que contiene a todos los abiertos.

Definición A.1 *Un espacio de Borel es un subconjunto de Borel de un espacio métrico, separable y completo.*

El listado a continuación menciona algunos ejemplos de espacios de Borel.
- $\mathbb{R}^n$ con la topología usual.
- Un conjunto numerable con la topología discreta.
- Un espacio métrico compacto.
- El producto finito o numerable de una sucesión de espacios de Borel.

Definición A.2 *Sea $\mathbb{X}$ un espacio topológico y v una función de $\mathbb{X}$ en $\mathbb{R} \cup \{\infty\}$. Se dice que la función v es:*

(a) semi-continua inferiormente (s.c.i.) *si el conjunto*

$$\{x \in \mathbb{X} \ : \ v(x) \leq r\}$$

es cerrado en $\mathbb{X}$ para cada $r \in \mathbb{R}$.

(b) semi-continua superiormente *si el conjunto*

$$\{x \in \mathbb{X} \ : \ v(x) \geq r\}$$

es cerrado en $\mathbb{X}$ para cada $r \in \mathbb{R}$.

Observación A.1 *Sea $\mathbb{X}$ un espacio topológico y v una función de $\mathbb{X}$ en $\mathbb{R}\cup\{\infty\}$. Entonces, respecto a la definición previa, se tiene que*

i) v es s.c.i. si y sólo si $-v$ es s.c.s., y

ii) v es continua si y sólo si v es simultáneamente s.c.i. y s.c.s.

Proposición A.1 *Sea $\mathbb{X}$ un espacio métrico y v una función de $\mathbb{X}$ en $\mathbb{R}\cup\{\infty\}$, entonces v es s.c.i. si y sólo si para cada sucesión $\{x_n\}$ en $\mathbb{X}$, tal que $x_n \to x \in \mathbb{X}$ se tiene que*

$$\liminf_n v(x_n) \geq v(x).$$

Demostración: Véase, por ejemplo, [1] Apéndice A6.

Definición A.3 *Sean $\mathbb{X}$ y $\mathbb{A}$ espacios de Borel. Una multifunción o correspondencia φ de $\mathbb{X}$ en $\mathbb{A}$ es una función definida en $\mathbb{X}$ y cuyo valor $\varphi(x)$, para cada $x \in \mathbb{X}$, es un subconjunto no vacío de $\mathbb{A}$.*

Definición A.4 *Se dice que la multifunción φ es medible de Borel, si para cada conjunto cerrado $F \subset \mathbb{A}$, el conjunto $\{x \in \mathbb{X} : \varphi(x) \subset F\}$ es un subconjunto de Borel en $\mathbb{X}$.*

Definición A.5 *Se dice que la multifunción φ es s.c.i., si para cada conjunto cerrado $F \subset \mathbb{A}$, el conjunto $\{x \in \mathbb{X} : \varphi(x) \subset F\}$ es un cerrado en $\mathbb{X}$.*

Definición A.6 *Se dice que la multifunción φ es s.c.s., si para cada conjunto abierto $G \subset \mathbb{A}$, el conjunto $\{x \in \mathbb{X} : \varphi(x) \subset G\}$ es un abierto en $\mathbb{X}$.*

Apéndice B

Kernel Estocástico y Esperanza Condicional

B.1. Kernel Estocástico

Sean $\mathbb{X}$ e $\mathbb{Y}$ espacios de Borel.

Definición B.1 *Un kernel estocástico o probabilidad de transición sobre* $\mathbb{X}$ *dado* $\mathbb{Y}$ *se define como una función* $Q(dx \mid y)$, *tal que:*

(a) $Q(\cdot \mid y)$ *es una medida de probabilidad sobre* $\mathbb{X}$ *para cada* $y \in \mathbb{Y}$;
(b) $Q(B \mid \cdot)$ *es una función medible sobre* $\mathbb{Y}$ *para cada* $B \in \mathfrak{B}(\mathbb{X})$.

Definición B.2 *Sea* $Q(dx \mid y)$ *un kernel estocástico sobre* $\mathbb{X}$ *dado* $\mathbb{Y}$. *Se dice que:*

(a) Q *es* fuertemente continuo *si la función:*

$$y \mapsto \int_{\mathbb{X}} v(x)\, Q(dx \mid y) \tag{B.1}$$

es continua y acotada en $y \in \mathbb{Y}$, *para cada función medible y acotada* v *sobre* $\mathbb{X}$.

(b) Q *es* débilmente continuo *si la función en* (B.1) *es continua y acotada en* $y \in \mathbb{Y}$, *para cada función continua y acotada* v *sobre* $\mathbb{X}$.

Observación B.1 *Sea* $Q(dx \mid y)$ *un kernel estocástico sobre* $\mathbb{X}$ *dado* $\mathbb{Y}$. *Entonces, respecto a la definición previa, se tiene la siguiente relación:*

$$(a) \quad \Rightarrow \quad (b)$$

B.2. Esperanza Condicional

Definición B.3 *Sea* $(\Omega, \mathfrak{D}, P)$ *un espacio de probabilidad, y sea* $\mathfrak{S} \subseteq \mathfrak{D}$ *una sub-*σ*-álgebra de* $\mathfrak{D}$*. La esperanza condicional de la variable aleatoria* Z, $E|Z| < \infty$, *dado* $\mathfrak{S}$*, se define como la única c.s. variable aleatoria denotada por* $E[Z \mid \mathfrak{S}]$ *tal que sea* $\mathfrak{S}$*-medible y además:*

$$\int_B E[Z \mid \mathfrak{S}]\, dP = \int_B Z\ dP \qquad \textit{para cada} \ \ B \in \mathfrak{S}.$$

Proposición B.1 *Sean* Z *y* Z' *dos variables aleatorias sobre* $(\Omega, \mathfrak{D}, P)$ *tales que* $E|Z| < \infty$ *y* $E|Z'| < \infty$*, y sean* b_1, b_2 *y* b_3 *constantes reales. Entonces*

(a) $E[b_1 Z + b_2 Z' + b_3 \mid \mathfrak{S}] = b_1 E[Z|\, \mathfrak{S}] + b_2 E[Z' \mid \mathfrak{S}] + b_3$ c.s.

(b) $Z \leq Z'$ c.s. *implica que* $E[Z \mid \mathfrak{S}] \leq E[Z' \mid \mathfrak{S}]$ c.s.

(c) $Z_n \geq 0$ *para cada* n *es tal que* $Z_n \nearrow Z$ c.s. *implica que*

$$E[Z_n \mid \mathfrak{S}] \nearrow E[Z \mid \mathfrak{S}].$$

(d) $E[I_B \mid \mathfrak{S}] = P[B \mid \mathfrak{S}]$, *donde* I_B *es la función indicadora del conjunto* B.

(e) *Sean* $\mathfrak{S}_1, \mathfrak{S}_2$ *dos sub-*σ*-álgebras de* $\mathfrak{D}$ *tales que* $\mathfrak{S}_1 \subset \mathfrak{S}_2 \subset \mathfrak{D}$*. Entonces:*

$$E[E[Z \mid \mathfrak{S}_1] \mid \mathfrak{S}_2] = E[Z \mid \mathfrak{S}_1]$$

y

$$E[E[Z \mid \mathfrak{S}_2] \mid \mathfrak{S}_1] = E[Z \mid \mathfrak{S}_1].$$

(f) Z *es* $\mathfrak{S}$*-medible implica que* $E[Z \mid \mathfrak{S}] = Z$.

(g) $E[E[Z \mid \mathfrak{S}]] = E[Z]$.

Demostración: Véase, por ejemplo, [1] Capítulo 6.

Apéndice C

Teorema de C. Ionescu Tulcea

Proposición C.1 *Sea* $\mathbb{X}_0, \mathbb{X}_1, ...$ *una sucesión de espacios de Borel y, para* $n \in \mathbb{N}_0$, *defínanse*

$$\mathbb{Y}_n := \mathbb{X}_0 \times \mathbb{X}_1 \times \cdots \times \mathbb{X}_n$$

y

$$\mathbb{Y} := \mathbb{X}_0 \times \mathbb{X}_1 \times \cdots$$

Sea ν *una medida de probabilidad arbitraria en* $\mathbb{X}_0$ *y, para cada* $n \in \mathbb{N}_0$, *sea* $Q_n(dx_{n+1} \mid y_n)$ *un kernel estocástico sobre* $\mathbb{X}_{n+1}$ *dado* $\mathbb{Y}_n$. *Entonces existe una única medida de probabilidad* Q_ν *sobre* $\mathbb{Y}$ *tal que, para cada rectángulo medible* $B_0 \times \cdots \times B_n$ *en* $\mathbb{Y}_n$,

$$\begin{aligned} Q_\nu(B_0 \times \cdots \times B_n) &= \int_{B_0} \nu(dx_0) \int_{B_1} Q_0(dx_1 \mid x_0) \int_{B_2} Q_1(dx_2 \mid x_0, x_1) \\ &\qquad \cdots \int_{B_n} Q_{n-1}(dx_n \mid x_0, ..., x_{n-1}). \end{aligned}$$

Más aún, para cada función medible no negativa u *en* $\mathbb{Y}$, *la función*

$$x \longmapsto \int u(y)\, Q_x(dy)$$

es medible sobre $\mathbb{X}_0$, *donde* Q_x *denota a* Q_ν *cuando* ν *es la probabilidad concentrada en* $x \in \mathbb{X}_0$.

Demostración: Véase, por ejemplo, [1] p.109.

Apéndice D

Otros resultados

Proposición D.1 *(Lema de Fatou "Generalizado") Supóngase que para cada* $x \in \mathbb{X}$ *se satisfacen las Hipótesis 1.3(c) y 1.3(e), y sea* $\{u_n\}$ *una sucesión uniformemente acotada en* $\mathbb{B}_W(\mathbb{X})$, *es decir, existe una constante* C^* *tal que* $\|u_n\|_W \leq C^*$ *para cada* n, *y defínanse*

$$u^I(x) := \liminf_{n\to\infty} u_n(x), \qquad y \qquad u^S(x) := \limsup_{n\to\infty} u_n(x).$$

Entonces, para cada $x \in \mathbb{X}$ *en cada sucesión* $\{a^n\} \subset A(x)$ *tal que* $a^n \to a \in A(x)$ *se tiene*

$$\liminf_{n\to\infty} \int_{\mathbb{X}} u_n(y)Q(dy \mid x, a^n) \geq \int_{\mathbb{X}} u^I(y)Q(dy \mid x, a),$$

y

$$\limsup_{n\to\infty} \int_{\mathbb{X}} u_n(y)Q(dy \mid x, a^n) \leq \int_{\mathbb{X}} u^S(y)Q(dy \mid x, a).$$

De aquí, si $u_n \to u$ *(es decir* $u^I = u^S$*), entonces*

$$\lim_{n\to\infty} \int_{\mathbb{X}} u_n(y)Q(dy \mid x, a^n) = \int_{\mathbb{X}} u(y)Q(dy \mid x, a).$$

Demostración: Véase, por ejemplo, [9] p.49.

Proposición D.2 *Sean g_n y g funciones integrables tales que $g_n \to g$ c.d. cuando $n \to \infty$. Entonces*

$$\int |g_n - g| \underset{n\to\infty}{\to} 0 \quad \textit{si y sólo si} \quad \int |g_n| \underset{n\to\infty}{\to} \int |g| \, .$$

En particular, si g_n y g son funciones de densidad de probabilidad, entonces el resultado establecido se conoce como Teorema de Scheffé.

Demostración: Véase, por ejemplo, [32] p.55.

Proposición D.3 *Sea $\{X_i\}_{i\in I}$ una sucesión de variables aleatorias integrables sobre $(\Omega, \mathfrak{D}, P)$. Si $h : [0,\infty) \to [0,\infty)$ es medible,*

$$\frac{h(t)}{t} \underset{n\to\infty}{\to} \infty$$

y

$$\sup_{i\in I} E\left[h\left(|X_i|\right)\right] \underset{n\to\infty}{\to} \infty,$$

entonces X_i son uniformemente integrables.

Demostración: Véase, por ejemplo, [1] p.301.

Bibliografía

[1] Ash R.B. (1972) *Real Analysis and Probability.* Academic Press, New York.

[2] Bhattacharya R.N., Majumdar M. (1989) *Controlled semi-Markov models-the discounted case.* J. Statist. Plann. Inference 21: 365-381.

[3] Devroye L. (1987) *A Course in Density Estimation.* Birkhäuser Verlag, Boston.

[4] Dynkin E.B., Yushkevich A.A. (1979) *Controlled Markov Processes.* Springer-Verlag, New York.

[5] Gordienko E.I., Minjárez-Sosa J.A. (1998) *Adaptive control for discrete-time Markov processes with unbounded costs: discounted criterion.* Kybernetika 34: 217–234.

[6] Hasminskii R., Ibragimov I. (1990) *On density estimation in the view of Kolmogorov´s ideas in approximation theory.* Ann. of Statist. 18: 999-1010.

[7] Hernández-Lerma O. (1989) *Adaptive Markov Control Processes.* Springer-Verlag, New York.

[8] Hernández-Lerma O., Lasserre J.B. (1996) *Discrete-Time Markov Control Processes: Basic Optimality Criteria.* Springer-Verlag, New York.

[9] Hernández-Lerma O., Lasserre J.B. (1999) *Further Topics on Discrete-Time Markov Control Processes.* Springer-Verlag, New York.

[10] Hilgert N., Minjárez-Sosa J.A. (2006) *Adaptive control of stochastic systems with unknown disturbance distribution: discounted criteria.* (Por aparecer en Math. Meth. of Oper. Res.)

[11] Köthe G. (1969) *Topological Vector Spaces I.* Springer-Verlag, New York.

[12] Lippman S.A. (1973) *Semi-Markov decision processes with unbounded rewards.* Management Sci. 19: 717-731.

[13] Lippman S.A. (1975) *On dynamic programming with unbounded rewards.* Management Sci. 21: 1225-1233.

[14] Luque-Vásquez F. (1997) *Modelos de Control Semi-Markoviano en Espacios de Borel.* Tesis Doctoral. Universidad Nacional Autónoma de México.

[15] Luque-Vásquez F. (2002) *Zero-sum semi-Markov games in Borel spaces with discounted payoff: discounted and average payoff.* Bol. Soc. Mat. Mexicana 8: 227-241.

[16] Luque-Vásquez F., Hernández-Lerma O. (1999) *Semi-Markov control models with average costs.* Appl. Math. (Warsaw), 26,3: 315-331.

[17] Luque-Vásquez F., Minjárez-Sosa J.A. (2005) *Semi-Markov control processes with unknown holding times distribution under a discounted criterion.* Math. Meth. Oper. Res. 61: 455-468.

[18] Luque-Vásquez F., Robles-Alcaraz M.T. (1994) *Controlled semi-Markov models with discounted unbounded costs.* Bol. Soc. Mat. Mexicana 39: 51-68.

[19] Luque-Vásquez F., Vega-Amaya O. (2004) *Time and ratio expected average cost optimality for semi-Markov control processes on Borel spaces.* Communications in Statistics: Theory and Methods 33-3: 715-734.

[20] Minjárez-Sosa J.A. (2004) *Approximation and estimation in Markov control processes under a discounted criterion.* Kybernetika 40: 681–690.

[21] Minjárez-Sosa J.A. (1998) *Control Adaptado para Procesos de Markov con Costos no Acotados.* Tesis Doctoral. Universidad Autónoma Metropolitana-Iztapalapa.

[22] Minjárez-Sosa J.A., Luque-Vásquez F. (2006) *Two person zero-sum semi-Markov games with unknown holding times distribution on one side: a discounted payoff criterion.* Reporte Interno No.31. Departamento de Matemáticas, Universidad de Sonora.

[23] Puterman M.L. (1994) *Markov Decision Processes. Discrete Stochastic Dynamic Programming.* John Wiley & Sons.

[24] Rieder U. (1978) *Measurable selection theorems for optimization problems.* Manuscripta Math., 24: 115-131.

[25] Ross S.M. (1970) *Applied Probability Models with Optimization Applications.* Holden-Day San Francisco.

[26] Royden H.L. (1968) *Real Analysis.* Macmillan Publishing Co., Inc., New York.

[27] Schäl M. (1975) *Conditions for optimality and for the limit of n-stage optimal policies to be optimal.* Z. Wahrs. Verw. Gerb. 32: 179–196.

[28] Schäl M. (1987) *Estimation and control in discounted stochastic dynamic programming.* Stochastics 20: 51–71.

[29] Van Nunen J.A.E.E, Wessels J. (1978) *A note on dynamic programming with unbounded rewards.* Manag. Sci. 24: 576–580.

[30] Vega-Amaya O. (1993) *Average optimality in semi-Markov control models on Borel spaces: unbounded costs and controls.* Bol. Soc. Mat. Mexicana 38: 47-60.

[31] Vega-Amaya O., Luque-Vásquez F. (2000) *Sample-path average cost optimality for semi-Markov constrol processes on Borel spaces: unbounded costs and mean holding times.* Applicationes Mathematicae 27: 343-367.

[32] Williams D. (1991) *Probability with Martingales.* Cambridge University Press, New York.