CONTROL ADAPTADO PARA PROCESOS DE MARKOV CON COSTOS NO ACOTADOS

Tesis que presenta:

Jesús Adolfo Minjárez Sosa

Para la obtención del grado de

Doctor en Ciencias

Febrero de 1998

Director: Dr. Evgueni I. Gordienko

Universidad Autónoma Metropolitana-Iztapalapa
División de Ciencias Básicas e Ingeniería
Departamento de Matemáticas

A mi Fran

Por el buqui.

Agradecimientos

Para la realización de este trabajo recibí el apoyo de muchas personas.

En especial quiero agradecer a mi esposa, a mis padres y hermanos.

Agradezco también al Dr. Evgueni I. Gordienko por haber aceptado ser mi director; a los Doctores Rolando Cavazos, Daniel Hernández, Onésimo Hernández y Raúl Montes de Oca, por las sugerencias y observaciones hechas a este trabajo; a los compañeros Enrique Hugues, Fernando Luque y Oscar Vega por los consejos y sugerencias que me han dado, y por el interés que mostraron en el tema de estudio.

Finalmente quiero agradecer al Consejo Nacional de Ciencia y Tecnología (CONACyT), por el apoyo económico que me brindó; a los Departamentos de Matemáticas de la Universidad de Sonora y la Universidad Autónoma Metropolitana - Iztapalapa; y al Sindicato de Trabajadores Académicos de la Universidad de Sonora.

INDICE

Introducción

El estudio de los sistemas de control estocástico a tiempo discreto ha tomado mucha importancia en los últimos años por sus multiples aplicaciones a problemas que se presentan en otras areas. En muchos de estos problemas, algunas de las componentes del modelo correspondiente no son completamente conocidas. En este sentido, debemos implementar esquemas que nos permitan ir aprendiendo o recolectando información acerca de las componentes desconocidas durante la operación del sistema, y de esta manera elegir una decisión o un control. Si lo anterior es posible realizar, decimos que tenemos un problema de control estocástico adaptado, para el cual debemos de diseñar políticas de control que minimicen un índice de funcionamiento del sistema, por ejemplo, índice de costo descontado o promedio.

En este trabajo nos centraremos en el estudio del problema de control adaptado (PCA) para procesos de Markov, en los cuales la distribución del ruido aleatorio se desconoce.

Aun cuando el estudio de los Procesos de Control de Markov (PCMs) tuvo sus origenes en los años 50's, no fué hasta principios de los 60's cuando aparecieron los primeros trabajos en control adaptado para procesos de Markov, en Bellman (1961), considerando que la distribución del ruido es conocida excepto por un parámetro θ (Control Adaptado Paramétrico). A partir de este trabajo, se desarrollaron diferentes técnicas para resolver el PCA en el caso paramétrico.

Una de estas técnicas consiste en suponer que el parámetro desconocido θ es una variable de estado, asignándole una distribución de probabilidad a priori. Esto nos lleva de manera natural a poder formular el PCA como uno con variables parcialmente observables. A este método se le conoce como formulación bayesiana del PCA [ver, por ejemplo, Rieder (1975), Van Hee (1978), Schäl (1979), Di Massi y Stettner (1995)].

Por otro lado, la formulación no bayesiana de un PCA consiste en suponer que el parámetro desconocido θ pertenece a cierto conjunto conocido Θ. Dentro de esta formulación, existen diferentes técnicas como por ejemplo, procesos de busqueda aleatoria [ver, por ejemplo, Gordienko (1985b)], métodos basados en la aplicación del principio de estimación y control [Mandl (1974)], método del gradiente [El-Fattah (1981)], técnicas basadas en los algoritmos de la teoria

de aprendizaje [Lyubchik y Pozniak (1974), Kurano (1987)], entre otros.

Cada uno de estos métodos tiene sus características propias. Por ejemplo, la convergencia de los algoritmos de los procesos de busqueda aleatoria es muy lenta; las políticas obtenidas mediante los métodos del gradiente y las basadas en algoritmos de la teoría de aprendizaje son aleatorizadas y tienen la restricción de que el espacio de estados y de control deben ser finitos.

Estos métodos son fáciles de implementar debido a que no se necesita estar resolviendo un problema de optimización en cada etapa, como sucede cuando se usa el Principio de Estimación y Control (PEC), lo que algunas veces resulta un problema difícil.

Aun considerando esta desventaja del PEC, tal vez este sea el método más popular y/o universal por su versatilidad ante situaciones generales. Este método fué propuesto, de manera independiente, por Mandl (1974) y Kurano (1972), y es aplicable cuando es posible primero resolver el problema suponiendo parámetros conocidos. Siendo asi, el PEC consiste en obtener el control óptimo en cada etapa para el parámetro desconocido, sustituyendo el valor del parámetro verdadero por el valor del estimador.

En la literatura de los PCMs, al PEC también se le conoce con los nombres de certainty equivalence principle, self-tuning regulator, o naive feedback controller [Bertsekas (1987)].

Una clase de problemas muy conocidos para los cuales el PEC es aplicable son los problemas de control de sistemas lineales con costo cuadrático y ruido gaussiano [ver Caines (1988), Kumar y Varaiya (1986)]. Este tipo de sistemas han sido tan ampliamente estudiado que se han desarrollado técnicas propias y muy efectivas (por ejemplo, ecuaciones de Ricatti), que no son aplicables, o difíciles de extender al caso de sistemas no lineales. Estas técnicas, combinadas con la aplicación del PEC resuelven el PCA para estos sistemas. De los métodos de estimación de parámetros más comunes en los sistemas lineales, están, el método de máxima verosimilitud, mínimos cuadrados, esquemas de mínima varianza [Caines (1988)].

Para sistemas en general, el PEC ha sido usado por varios autores para diseñar políticas adaptadas bajo diferentes tipos de hipótesis en el modelo de control. Por ejemplo, Schäl (1987) y Hernández-Lerma (1987), construyen, respectivamente, PEC- políticas adaptadas bajo el criterio de costo descontado y promedio [ver también hernández-Lerma (1989) y sus referencias].

Una ventaja importante del PEC es que puede aplicarse, de manera natural, cuando la distribución del ruido aleatorio es completamente desconocida. Nos referiremos a este caso como Problema de Control Adaptado No-paramétrico (PCAN). De hecho, este tipo de problemas puede verse como un problema de control adaptado paramétrico observando lo siguiente. Si μ denota la distribución desconocida de las componentes aleatorias del sistema, ésta puede ser considerada como un parámetro (desconocido) tomando valores en un conjunto apropiado de distribuciones de probabilidad sobre el espacio de perturbaciones aleatorias del sistema. De los autores que han aplicado el PEC para resolver un PCAN destacan Gordienko (1985, 1985a), Hernández-Lerma y Marcus (1987), Hernández-Lerma y Cavazos-Cadena (1990), Cavazos-Cadena (1990), Gordienko y Minjárez-Sosa (1996, 1997), Minjárez-Sosa (1998), entre otros.

Aún cuando existe una relación entre el caso paramétrico y el no paramétrico, la diferencia entre las técnicas de solución de ambos problemas es sustancial. Por un lado están las técnicas de estimación de parámetros y por otro las de estimación de la distribución. Cada una de ellas constituye un área de investigación dentro de la estadística, en las cuales se siguen obteniendo nuevos resultados.

El caso de estimación no- paramétrica tal vez sea más complicado. Un método natural para estimar la distribución μ de las componentes aleatorias, es por medio de la distribución empírica. Esta técnica fué usada en problemas de control adaptado por Gordienko (1985), Hernández-Lerma y Marcus (1987), Cavazos-Cadena (1990), entre otros. Otra manera de analizar el problema es suponer que μ es absolutamente continua con respecto a la medida de Lebesgue en $\Re^k$. Esto implica que existe la función de densidad (desconocida) de las componentes aleatorias [Ash (1972)], lo cual nos lleva a tener que implementar métodos de estimación de densidad, estudiados ampliamente por algunos autores [ver Devroye y Györfi (1985), Devroye (1987), Silverman (1986), Bosq (1996), Yakowitz (1985), Haminskii e Ibragimov (1990), Hernández-Lerma (1991)]. De los trabajos donde se aplican técnicas de estimación de densidad en problemas de control adaptado podemos citar, por ejemplo, Hernández-Lerma y Cavazos-Cadena (1990), Gordienko y Minjárez-Sosa (1996, 1997), Minjárez-Sosa (1998).

La ventaja de usar la distribución empírica como método de estimación es que tanto la

distribución μ como el espacio de perturbaciones aleatorias pueden ser arbitrarios; pero la desventaja es que se requiere imponer hipótesis muy restrictivas en el modelo de control [ver, por ejemplo, Hernández-Lerma (1989) y sus referencias]. En el otro caso, a pesar de que se restringe a que μ tenga densidad, existe un gran número de ejemplos donde se cumple este supuesto (sistemas de producción-inventario, ciertos sistemas de colas, etc.), además de que no se necesitan hipótesis restrictivas en el modelo de control.

El problema de control adaptado para procesos de Markov, y en general la teoría de los PCMs, han sido ampliamente estudiados en los casos cuando el espacio de estados y control son un conjunto numerable y/o la función de costo por etapa es acotada, ver por ejemplo, para el problema de control adaptado, Hernández-Lerma (1989) y sus referencias, Gordienko (1985, 1985a,b), Cavazos-Cadena (1990), Hernández-Lerma y Cavazos-Cadena (1990), y para la teoría de los PCMs ver Arapostathis, et. al. (1993) y sus referencias, Bertsekas y Shreve (1978), entre otros. En los últimos años, diferentes autores han extendido los resultados relativos a los PCMs al caso cuando los espacios de estados y control son de Borel y/o el costo por etapa es no-acotado, [ver Hernández-Lerma y Lasserre (1995) y sus referencias, Gordienko y Hernández-Lerma (1995a,b), Gordienko, Montes-de-Oca y Minjárez-Sosa (1997), etc.].

Sin embargo, hasta donde conocemos, el problema de control adaptado para procesos de Markov no se ha extendido a este contexto, lo cual es la motivación principal del presente trabajo.

En este trabajo estamos interesados en estudiar el problema de control adaptado para procesos de Markov a tiempo discreto cuya dinámica es de la forma:

$$x_{t+1} = F(x_t, a_t, \xi_t), \quad t = 0, 1, 2, ...,$$

donde F es una función arbitraria conocida; x_t y a_t representan, respectivamente, el estado y control al tiempo t, tomando valores en un espacio de Borel; y $\{\xi_t\}$ son vectores aleatorios i.i.d. en $\Re^k$ teniendo una densidad desconocida ρ perteneciendo a un conjunto apropiado.

Considerando que los costos por etapa pueden ser no acotados y suponiendo que las realizaciones $\xi_0, \xi_1, \xi_2, ...$ del proceso de perturbaciones aleatorias y el de estados $x_0, x_1, x_2, ...$ son completamente observables, presentamos la construcción de políticas adaptadas asintóticamente

óptimas [Mandl (1974), Schäl (1987)] con respecto al índice de costo descontado y políticas adaptadas óptimas con respecto al índice de costo promedio por etapa.

Específicamente, estudiamos la optimalidad de las políticas adaptadas llamadas PEC, propuesta originalmente por Mandl (1974), Iteración de Valores No-estacionario (IVN), propuesta en Hernández-Lerma y Marcus (1985) y la G- política propuesta en Gordienko (1985). En este sentido, nuestro trabajo se puede considerar como una extensión de los resultados presentados en estos artículos, los cuales consideran costo por etapa acotado.

Un punto clave para lograr los objetivos del trabajo, es garantizar la existencia de soluciones de la ecuación de optimalidad en costo descontado y de la ecuación de optimalidad en costo promedio, combinando estos resultados con métodos apropiados de estimación de la densidad. El hecho de permitir que el costo por etapa sea no-acotado lo hace un problema no trivial, en donde se necesita aplicar nuevas técnicas de demostración y métodos más precisos de estimación de la densidad. Por ejemplo, para el caso descontado, en general no se tienen propiedades contractivas del operador de la ecuación de optimalidad, lo cual nos lleva a tener que imponer hipótesis tipo Lippman [Lippman (1975), Van Nunen (1978)] en las probabilidades de transición del proceso, para después usar los resultados de Hernández-Lerma (1994) y Gordienko y Hernández-Lerma (1995a). Para el caso promedio tampoco se tienen propiedades contractivas y, además de las hipótesis tipo Lippman, necesitamos imponer condiciones de ergodicidad (debido al análisis asintótico que se requiere) y de esta manera usar los resultados de Gordienko y Hernández-Lerma (1995a,b) y Gordienko, Montes-de-Oca y Minjárez-Sosa (1997). Ahora, la construcción de las políticas adaptadas se hace combinando los resultados anteriores con un proceso de estimación de la densidad que consiste en usar como estimador de ρ a la proyección ρ_t de un estimador $\hat{\rho}_t$ sobre un conjunto apropiado de densidades en $L_q(\Re^k)$, donde $E\,\|\hat{\rho}_t - \rho\|_q^r = \mathbf{O}(t^{-\gamma})$, $\gamma > 0$, $r > 1$. Esta propiedad es heredada por el estimador ρ_t. Ejemplos de estimadores $\hat{\rho}_t$ con estas características se pueden encontrar en Hasminskii e Ibragimov (1990).

A lo largo del trabajo, presentamos un ejemplo de un sistema de colas con tasa de servicio controlable, el cual satisface todas las hipótesis que usamos. Otros ejemplos que pueden ser considerados son los procesos autorregresivos.

El trabajo consta de cuatro capítulos y está organizado de la siguiente manera. En el Capítulo I se define el tipo de modelo de control en los que estamos interesados. En los Capítulos II y III se estudia, respectivamente, el problema de control adaptado bajo el índice costo descontado y costo promedio, los cuales están basados en los artículos recientes Gordienko y Minjárez-Sosa (1996, 1997), Gordienko, Montes-de-Oca y Minjárez-Sosa (1997), Minjárez-Sosa (1998). Finalmente, el Capítulo IV consta de las conclusiones y una lista de problemas importantes que consideramos aún no resueltos y que pueden ser un complemeto para este trabajo.

Capítulo 1

Procesos de Control de Markov

1.1 Introducción

En este capítulo daremos los elementos necesarios para definir el problema de control óptimo. Para esto, especificaremos la clase de modelos de control, la clase de políticas admisibles y los criterios de optimalidad con los que trabajaremos. Además, daremos las condiciones sobre el modelo de control para las cuales, la teoría desarrollada a lo largo del trabajo, es aplicable. Concluimos con un ejemplo de un modelo de sistemas de colas el cual retomaremos en cada uno de los siguientes capítulos.

1.2 Modelo de control de Markov

Consideremos un modelo de control de Markov [Dynkin y Yushkevich (1979), Hernández- Lerma (1989)] estacionario a tiempo discreto de la forma particular $(X, A, \Re^k, F, \rho, c)$ donde:

- X es un espacio de Borel [ver A.1.1, Apéndice A] no vacío llamado espacio de estados. A los elementos de X les llamaremos estados.

- A es un espacio de Borel no vacío llamado conjunto de acciones o de control.

A cada $x \in X$, le asociamos un conjunto no vacío $A(x) \in \mathbb{B}(A)$, al cual llamaremos conjunto de acciones admisibles en el estado x, donde $\mathbb{B}(A)$ es la $\sigma-$álgebra de Borel asociada al conjunto A [ver A.1.0, Apéndice A].

Para el resto del trabajo, la medibilidad de conjuntos y funciones será con respecto a la $\sigma-$álgebra de Borel correspondiente.

Supondremos que el conjunto

$$\mathbb{K} = \{(x, a) : x \in X, a \in A(x)\},$$

de las parejas estado-control admisibles, es un subconjunto de Borel del espacio $X \times A$.

- $\Re^k$ es el espacio de perturbaciones aleatorias.

- $F : XA\Re^k \to X$, es una función medible conocida, que representa la dinámica del sistema, esto es,

$$x_{t+1} = F(x_t, a_t, \xi_t), \quad t = 0, 1, 2, ..., \tag{1.1}$$

  donde $x_t \in X$, $a_t \in A(x_t)$ y $\{\xi_t\}$, llamado proceso de perturbación, es una sucesión de variables aleatorias (v.a) independientes e identicamenmte distribuidas (i.i.d.) con valores en $\Re^k$ y con una distribución la cual tiene densisad ρ.

- $c : \mathbb{K} \to \Re$, es una función medible no-negativa, posiblemente no acotada, la cual representa el costo por etapa.

El modelo de control $(X, A, \Re^k, F, \rho, c)$ representa un sistema estocástico controlado que se observa en los tiempos $t = 0, 1, ...$, el cual evoluciona de la siguiente manera. Sea $x_t \in X$ el estado del sistema al tiempo t y a_t la acción elegida. Si $x_t = x$ y $a_t = a$, entonces: 1) se incurre en un costo $c(x, a)$ y 2) el sistema avanza a un nuevo estado x_{t+1} de acuerdo a una distribución de probabilidad $Q_\rho(\cdot/x, a)$ definida como

$$Q_\rho(B/x, a) := Prob\,[F(x_t, a_t, \xi_t) \in B/x_t = x, a_t = a] = \int_{\Re^k} 1_B[F(x, a, s)]\rho(s)ds, \tag{1.2}$$

donde $B \in \mathbb{B}(X)$ y 1_B es la función indicadora del conjunto B. Estando ahora en el nuevo estado, digamos $x_{t+1} = x'$, se aplica un nuevo control $a' \in A(x')$ y se repite el proceso anterior.

En algunos casos, el estado inicial x_0 es escogido de acuerdo a una distribución de probabilidad inicial ν, i.e. $\Pr(x_0 \in B) = \nu(B)$, $B \in \mathbb{B}(X)$. En nuestro caso supondremos que ν está concentrada en un estado dado $x \in X$, es decir, $\Pr[x_0 = x] = 1$.

1.3 Procesos de control de Markov

Definimos el espacio de historias admisibles hasta el tiempo $t = 0, 1, ...,$ como

$$\begin{aligned} \mathbb{H}_0 &: = X, \\ \mathbb{H}_t &: = (\mathbb{K} \times \Re^k)^t \times X. \end{aligned}$$

Un elemento $h_t \in \mathbb{H}_t$ es un vector o historia de la forma:

$$h_t = (x_0, a_0, s_0, ..., x_{t-1}, a_{t-1}, s_{t-1}, x_t)$$

donde $(x_n, a_n) \in \mathbb{K}$, $s_n \in \Re^k$ para $n = 0, 1, ..., t-1$ y $x_t \in X$.

Denotemos por $\mathbb{F}$ al conjunto de funciones medibles $f : X \to A$ tal que $f(x) \in A(x)$, $x \in X$.

Definición 1.1: (a) Una política de control $\pi = \{\pi_t\}$ es una sucesión de kérnels estocásticos [ver A.2.5, Apéndice A] π_t sobre A dado $\mathbb{H}_t$, tal que

$$\pi_t(A(x_t)/h_t) = 1, \ h_t \in \mathbb{H}_t, \ t = 0, 1, ...$$

Al conjunto de todas las políticas lo denotaremos por Π.

(b) Una política de control $\pi = \{\pi_t\}$ es estacionaria si existe una función $f \in \mathbb{F}$ tal que $\pi_t(\cdot/h_t)$ está concentrada en $f(x)$, $t \geq 0$.

Una política estacionaria toma la forma $\mathbf{f} = \{f, f, ...\}$ y nos referiremos a $\mathbb{F}$ como el conjunto de políticas estacionarias.

Sea $(\Omega, \mathcal{F})$ un espacio medible donde $\Omega := (X \times A \times \Re^k)^\infty = X \times A \times \Re^k \times X \times A \times \Re^k \cdots$ y $\mathcal{F}$ la $\sigma-$álgebra producto correspondiente. Un elemento $\omega \in \Omega$ es de la forma

$$\omega = (x_0, a_0, s_0, x_1, a_1, s_1, ...),$$

donde las variables x_t, a_t y s_t, $t = 0, 1, ...$, son las proyecciones de Ω sobre los conjuntos X, A y $\Re^k$ respectivamente.

Es fácil ver que Ω contiene al espacio $\mathbb{H}_\infty := \mathbb{K} \times \Re^k \times \mathbb{K} \times \Re^k \cdots$ de historias admisibles $(x_0, a_0, s_0, x_1, a_1, s_1, ...)$, donde $(x_t, a_t) \in \mathbb{K}$ y $s_t \in \Re^k$, $t \geq 0$.

Sea $\pi \in \Pi$ una política de control arbitraria y $x_0 = x \in X$ el estado inicial. Entonces, existe una única medida de probabilidad P_x^π en $(\Omega, \mathcal{F})$ [ver e.g. Dynkin y Yushkevich (1979), Hinderer (1970)] tal que para toda $B \in \mathbb{B}(X)$, $C \in \mathbb{B}(A)$, $D \in \mathbb{B}(\Re^k)$, $h_t \in \mathbb{H}_t$ y $t = 0, 1, ...$

$$\begin{aligned}
P_x^\pi[x_0 &= x] = 1; \\
P_x^\pi[a_t &\in C \mid h_t] = \pi_t[C \mid h_t]; \\
P_x^\pi[\xi_t &\in D \mid h_t] = \int_D \rho(s)ds; \\
P_x^\pi[x_{t+1} &\in B \mid h_t, a_t] = Q(B \mid x_t, a_t) = \int_{\Re^k} 1_B[F(x_t, a_t, s)]\rho(s)ds. \qquad (1.3)
\end{aligned}$$

donde $a_t = \pi_t(h_t)$.

Al proceso estocástico $(\Omega, \mathcal{F}, P_x^\pi, \{x_t\})$ se le conoce como Proceso de Control de Markov (PCM). En particular, si π es una política estacionaria, el proceso $\{x_t\}$ es de Markov respecto a la medida de probabilidad P_x^π [Hernández-Lerma (1989)].

Denotaremos por E_x^π al operador esperanza respecto a P_x^π y para cada política estacionaria

$\mathbf{f} \in \mathbb{F}$ escribiremos

$$c(x,f) := c(x,f(x)) \quad y \quad F(x,f,s) := F(x,f(x),s)\,, \quad x \in X,\ s \in \Re^k.$$

1.4 Problema de Control Optimo

Para formular el problema de control óptimo (PCO) necesitamos una función que mida el comportamiento del sistema. A esta función se le conoce como índice de funcionamiento y su forma depende del criterio de optimalidad de interés. En este trabajo estudiaremos los criterios de costo descontado y promedio, los cuales están definidos de la siguiente manera.

Para cada política $\pi \in \Pi$ y estado inicial $x \in X$, definimos el *costo total descontado* como

$$V_\alpha(\pi,x) := E_x^\pi \left[\sum_{t=0}^{\infty} \alpha^t c(x_t, a_t) \right] \tag{1.4}$$

donde $\alpha \in (0,1)$ es el factor de descuento dado; y el *costo promedio por etapa* como

$$J(\pi,x) := \limsup_{n\to\infty} \frac{J_n(\pi,x)}{n}, \tag{1.5}$$

donde

$$J_n(\pi,x) := \sum_{t=0}^{n-1} E_x^\pi [c(x_t,a_t)] \tag{1.6}$$

es el costo esperado en n etapas cuando se usa la política $\pi \in \Pi$ y el estado inicial es $x \in X$.

Dado un modelo de control $(X, A, \Re^k, F, \rho, c)$, una familia de políticas admisibles Π y un índice de funcionamiento, el PCO consiste en encontrar una política π^* tal que minimice dicho índice.

Por ejemplo, para el criterio de costo descontado, el PCO es encontrar una política $\pi^* \in \Pi$ tal que

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

En este caso decimos que π^* es $\alpha-$óptima.

Para el criterio de costo promedio, el PCO es encontrar una política $\pi^* \in \Pi$ tal que

$$J^*(x) := \inf_{\pi \in \Pi} J(\pi, x) = J(\pi^*, x), \quad x \in X, \tag{1.8}$$

y diremos que π^* es costo promedio óptima (CP-óptima).

A las funciones $V_\alpha^*(\cdot)$ y $J^*(\cdot)$ las llamaremos costo descontado óptimo y costo promedio óptimo respectivamente, y cuando sea claro el contexto en el que estemos trabajando solo las llamaremos funciones de valor óptimo.

El objetivo de este trabajo es estudiar el problema de control adaptado para procesos de Markov, para lo cual supondremos que la densidad ρ de las variables aleatorias ξ_t en (1.1) es desconocida. Esto nos lleva a tener que implementar métodos de estimación estadística, y en este sentido, la política que resulta al combinar estimación y control se la llama adaptada. Por lo tanto, el problema de control adaptado consiste en encontrar una política adaptada que minimice cierto índice de funcionamiento.

Para el estudio del problema de control adaptado, supondremos que las realizaciones $\xi_0, \xi_1, \xi_2, ...$, del proceso de perturbaciones, y el de estados $x_0, x_1, x_2, ...$, son completamente observables.

1.5 Hipótesis generales sobre el modelo de control

La construcción de las políticas adaptadas en los siguientes capítulos está basada en la existencia de minimizadores medibles $f \in \mathbb{F}$. Para garantizar dicha existencia, es necesario imponer condiciones de continuidad y compacidad en el modelo de control. Antes de establecer estas condiciones veamos la definición formal de minimizadores:

Definición 1.2. Sea v una función real definida en $\mathbb{K}$. Decimos que $f \in \mathbb{F}$ es un

a) minimizador medible de v si

$$v(x, f(x)) = \inf_{a \in A(x)} v(x, a), \quad x \in X.$$

b) $\delta-$ minimizador medible, $\delta > 0$, de v si

$$v(x, f(x)) \leq \inf_{a \in A(x)} v(x, a) + \delta, \quad x \in X.$$

Claramente, si $f \in \mathbb{F}$ es un minimizador medible entonces f es un $\delta-$minimizador medible para cada $\delta > 0$. Esto implica que para garantizar la existencia de $\delta-$minimizadores no es necesario imponer condiciones tan restritictivas en el modelo de control, comparándolas con las que garantizan la existencia de minimizadores.

A continuación estableceremos dos tipos de hipótesis que garantizan la existencia de δ-minimizadores y minimizadores medibles.

Sea $W : X \to [1, \infty)$ una función medible dada. Denotemos por L_W^∞ al espacio lineal normado de todas las funciones medibles $u : X \to \Re$ con

$$\|u\|_W := \sup_{x \in X} \frac{|u(x)|}{W(x)} < \infty. \tag{1.9}$$

Hipótesis H1.

a) Para cada $u \in L_W^\infty$ el conjunto

$$\left\{ (x, a) : \int_{\Re^k} u[F(x, a, s)]\rho(s)ds \leq r \right\}$$

es de Borel en $\mathbb{K}$, para cada $r \in \Re$;

b) para cada $x \in X$, $A(x)$ es un conjunto $\sigma-$ compacto;

c) para cada $x \in X$, la función $a \to c(x, a)$ es semi-continua inferiormente (s.c.i.) en $A(x)$ y $\sup_{A(x)} |c(x, a)| \leq W(x)$.

Lema 1.3: Supongamos que la Hipótesis H1 se cumple y $u \in L_W^\infty$. Entonces la función

$$u^*(x) := \inf_{a \in A(x)} \left\{ c(x, a) + \int_{\Re^k} u[F(x, a, s)]\rho(s)(ds) \right\}, \quad x \in X, \tag{1.10}$$

es medible y para cada $\delta > 0$ existe un $\delta-$minimizador, es decir, existe $f \in \mathbb{F}$ tal que

$$c(x,f)+\int_{\Re^k} u[F(x,f,s)]\rho(s)(ds) \leq u^*(x)+\delta,\ \ x\in X.$$

La demostración de este lema es una consecuencia del Corolario 4.3 en Rieder (1978).

En general tenemos la siguiente definición:

Definición 1.4. Sea $\delta > 0$ un número arbitraio. Decimos que una política $\pi \in \Pi$ es

a) $\delta-\alpha-$óptima si $V_\alpha(\pi,x) \leq V_\alpha^*(x)+\delta,\ x\in X.$

b) $\delta-$CP-óptima si $J(\pi,x) \leq J^*(x)+\delta,\ \ x\in X.$

Hipótesis H1'.

a) H1(a), (c) se cumplen.

b) Para cada $x \in X$, $A(x)$ es un conjunto compacto.

c) Para cada $u \in L_W^\infty$, $x \in X$, la función

$$a \longrightarrow \int_{\Re^k} u[F(x,a,s)]\rho(s)ds$$

es semi-continua inferiormente [ver A.2.0, Apéndice A].

Lema 1.5. Supongamos que la Hipótesis H1' se cumple y $u \in L_W^\infty$. Entonces la función u^* definida en (1.10) es medible y existe $f \in \mathbb{F}$ tal que

$$u^*(x) = c(x,f)+\int_{\Re^k} u[F(x,f,s)]\rho(s)(ds),\ \ x\in X.$$

De nuevo, la demostración de este lema es una consecuencia del Corolario 4.3 en Rieder (1978).

A lo largo del trabajo usaremos $\sup_X$, $\inf_X$, $\sup_{A(x)}$ y $\inf_{A(x)}$ para denotar $\sup_{x\in X}$, $\inf_{x\in X}$, $\sup_{a\in A(x)}$ y $\inf_{a\in A(x)}$ respectivamente.

En cada uno de los siguientes capítulos, estableceremos otro grupo de hipótesis que imponen condiciones a la función de densidad ρ de las variables aleatorias ξ_t en (1.1). Estas condiciones dependen fuertemente del tipo de índice de funcionamiento que se esté analizando.

1.6 Ejemplo

Consideremos un proceso de control de la forma

$$x_{t+1} = (x_t + a_t\eta_t - \chi_t)^+, \quad t = 0, 1, 2, ..., \tag{1.11}$$

con $x_0 = x$ conocido, espacio de estados $X = [0, \infty)$, conjunto de controles admisibles $A(x) = A$, $x \in X$, donde A es un subconjunto compacto del intervalo $(0, \theta]$, para algún $\theta \in \Re$, con $\theta \in A$. Además, $\{\eta_t\}$ y $\{\chi_t\}$ son sucesiones independientes de variables aleatorias i.i.d.

La relación (1.11) describe algunos modelos de control de sistemas de almacenamiento, como por ejemplo, sistemas de producción- inventario ($\eta_t = 1$), sistemas de colas, etc. [ver Dynkin y Yushkevish (1979), Hernández-Lerma (1989), etc.]

Para fijar ideas, centramos nuestra atención en un sistema de colas con un servidor del tipo $GI/GI/1/\infty$ con tasa de servicio controlable. En este caso, x_t y $a_t\eta_t$ denotan, respectivamente, el tiempo de espera y el tiempo de servicio del $t-$ ésimo cliente y χ_t el tiempo entre arribos del $t-$ ésimo y el $(t+1)-$ ésimo cliente. Además, a_t es el recíproco de la tasa de servicio controlable u_t ($a_t = 1/u_t$) para el $t-$ ésimo cliente.

Supondremos la existencia de las funciones de densidad ρ_{η_0} y ρ_{χ_0} de las variables aleatorias η_0 y χ_0 respectivamente, las cuales son acotadas, continuas y estrictamente positivas en $[0, \infty)$. Además supondremos que la densidad $\rho := \rho_{\eta_0}\rho_{\chi_0}$ es desconocida, pero las realizaciones $\eta_0, \eta_1, ..., \eta_{t-1}$, $\chi_0, \chi_1, ..., \chi_{t-1}$ y los estados x_t son observables en el momento de tomar la decisión a_t. Un caso particular donde se cumple este supuesto, es cuando el sistema de espera es un canal de trasmisión de mensajes. En este caso, x_t representa el tiempo que tarda para empezar a ser procesado el $t-$ésimo mensaje, $a_t\eta_t$ tiempo de transmisión del $t-$ésimo mensaje

y χ_t el tiempo entre arribos del $t-$ ésimo y el $(t+1)-$ ésimo mensaje [ver, por ejemplo, Weber y Stidham (1987)].

Este ejemplo será retomado en cada uno de los siguientes capítulos para ejemplificar la teoría desarrollada.

Capítulo 2

Criterio de Costo Descontado

2.1 Introducción

En este capítulo estudiamos el problema de control adaptado bajo el índice de costo descontado (1.4). Proponemos un método de estimación de la densidad ρ de las variables aleatorias ξ_t en (1.1) para construir políticas adaptadas. Dichas políticas eligen la acción a_t como una función de la historia del proceso y del estimador ρ_t de ρ.

El índice de costo descontado tiene la característica que depende fuertemente de los controles o acciones aplicadas en las primeras etapas, que es donde el método de estimación no proporciona mucha información acerca de la densidad ρ. Esto implica que por lo general no existe una política adaptada $\alpha-$óptima, y este hecho nos lleva a tener que considerar otro criterio de optimalidad llamado *optimalidad asintótica descontada.*

En particular, en este capítulo, mostramos que la política adaptada llamada Principio de Estimación y Control [Mandl (1974)] y la que resulta al implementar esquemas de Iteración de Valores No-estacionario [Hernández-Lerma y Marcus (1985)] son asintóticamente óptimas descontadas. Estos resultados se encuentran en el articulo Gordienko y Minjárez-Sosa (1997).

2.2 Ecuación de optimalidad en costo descontado

Para el resto del capítulo, fijamos una función $W(\cdot)$ satisfaciendo la Hipótesis H1(c) y caracterizaremos el conjunto de densidades que definen la clase de PCMs para los cuales la teoría

desarrollada es aplicable.

Definición 2.1. Sea $\bar{\rho}$: $\Re^k \to \Re$ una función medible no negativa y $\varepsilon \in (0, 1/2)$, fijos. Denotando $q := 1 + 2\varepsilon$, definimos el conjunto $D_0 = D_0(\bar{\rho}, L, \beta_0, b_0, p, q)$ como el conjunto de todas las densidades μ en $\Re^k$ que satisfacen lo siguiente:

a) $\mu \in L_q(\Re^k)$;

b) existe una constante L tal que para cada $z \in \Re^k$,

$$\|\Delta_z \mu\|_{L_q} \leq L\,|z|^{1/q}, \tag{2.1}$$

donde $\Delta_z \mu(x) := \mu(x+z) - \mu(x)$, $x \in \Re^k$ y $|\cdot|$ es la norma Euclidiana en $\Re^k$;

c) $\mu(s) \leq \bar{\rho}\,(s)$ casi dondequiera con respecto a la medida de Lebesgue;

d) para cada $x \in X$, $a \in A(x)$

$$\int_{\Re^k} W^p[F(x,a,s)]\mu(s)ds \leq \beta_0 W^p(x) + b_0; \tag{2.2}$$

donde $p > 1$, $\beta_0 < 1$, $b_0 < \infty$ son arbitrarios pero fijos.

Hipótesis H2.

a) La densidad ρ de las v.a. ξ_t en (1.1) pertenece a D_0.

b) La función

$$\varphi(s) := \sup_X [W(x)]^{-1} \sup_{A(x)} W[F(x,a,s)] \tag{2.3}$$

es finita para cada $s \in \Re^k$.

c) $\int_{\Re^k} \varphi^2(s) \left|\bar{\rho}\,(s)\right|^{2-q} ds < \infty$.

Observación 2.2. La función φ en (2.3) puede ser no-medible. En este caso supondremos que existe una función $\bar{\varphi}$, satisfaciendo la Hipótesis H2(c), tal que $\varphi \leq \bar{\varphi}$.

Proposición 2.3. Supongamos que $k = 1$. Una condición suficiente para que se cumpla (2.1) es la siguiente: Existe un conjunto finito $G \subset \Re$ (posiblemente vacío) y una constante $M \geq 0$ tal que,

i) μ tiene una derivada acotada μ' en $\Re \setminus G$ la cual pertenece a L_q;

ii) la función $|\mu'(x)|$ es no creciente para $x \geq M$ y no decreciente para $x \leq -M$.

Observemos que G incluye puntos de discontinuidad de μ si estos existen.

Demostración. Consideremos los siguientes casos:

I.- $|z| \geq 1$.

Usando las definiciones de $\|\cdot\|_q$ y Δ_z, y el hecho de que $\int_{\Re} \mu(x)dx = \int_{\Re} \mu(x+z)dx$ tenemos:

$$\|\Delta_z \mu\|_q^q = \int_{\Re} |\Delta_z \mu(x)|^q \, dx \leq \int_{\Re} (|\mu(x+z)| + |\mu(x)|)^q \, dx \leq$$

$$M_0 \int_{\Re} |\mu(x)|^q \, dx \leq M_0' \leq M_0' |z|, \tag{2.4}$$

para alguna constante $M_0 > 0$ y $M_0' > 0$. La relación (2.4) implica (2.1) con $L := (M_0')^{1/q}$.

II.- $|z| < 1$.

Sea $G = \{x_0, x_1, ..., x_m\}$. Sin pérdida de generalidad, supongamos que $z \geq 0$, $-M - z < x_0 < ... < x_m < M$. Ahora,

$$\|\Delta_z \mu\|_q^q = \int_{-\infty}^{-M-z} |\Delta_z \mu(x)|^q \, dx + \int_{-M-z}^{-M} |\Delta_z \mu(x)|^q \, dx + \int_{-M}^{x_0} |\Delta_z \mu(x)|^q \, dx + \int_{x_0}^{x_1} |\Delta_z \mu(x)|^q \, dx + ... +$$

$$\int_{x_m}^{M} |\Delta_z \mu(x)|^q \, dx + \int_{M}^{\infty} |\Delta_z \mu(x)|^q \, dx. \tag{2.5}$$

Por otro lado, bajo la condición (i), para cada x, $z \notin G$, existe $x^* \in (x, x+z)$ tal que

$$\left| \frac{\Delta_z \mu(x)}{z} \right| = |\mu'(x^*)| \leq L_0$$

o

$$|\Delta_z \mu(x)| \leq |z| \, L_0 \tag{2.6}$$

para alguna constante $0 < L_0 < \infty$. De aquí tenemos que

$$|\Delta_z \mu(x)| \leq |z| \max_{x \leq y \leq x+z} \mu'(y), \;\; x, \, z \notin G.$$

Ahora, por la condición (ii) tenemos

$$|\Delta_z \mu(x)| \leq |z| \, |\mu'(x+z)| \, , \; si \; x \in (-\infty, -M-z); \tag{2.7}$$

$$|\Delta_z \mu(x)| \leq |z| \, |\mu'(x)| \, , \; si \; x \in (M, \infty). \tag{2.8}$$

Usando las relaciones (2.6)-(2.8) en (2.5) y por la condición (i) obtenemos:

$$\|\Delta_z \mu\|_q^q \;\; \leq \;\; \int_{-\infty}^{-M-z} |z|^q \, |\mu'(x+z)|^q \, dx + \int_{-M-z}^{-M} L_0^q \, |z|^q \, dx + \int_{-M}^{x_0} L_0^q \, |z|^q \, dx + \int_{x_0}^{x_1} L_0^q \, |z|^q \, dx + ... +$$

$$\int_{x_m}^{M} L_0^q \, |z|^q \, dx + \int_{M}^{\infty} |z|^q \, |\mu'(x)|^q \, dx \leq L_1 \, |z|^q + L_2 \, |z|^q + ... + L_{m+5} \, |z|^q \leq L^* \, |z|^q \leq L^* \, |z| \, ,$$

para algunas constantes $L_i > 0$, $i = 1, ..., m+5$, y $L^* > 0$. La última desigualdad es por el hecho de que $|z| < 1$. De aquí, $\|\Delta_z \mu\|_q \leq (L^*)^{1/q} \, |z|^{1/q}$. Tomando $L := (L^*)^{1/q}$ llegamos a que (2.1) se cumple.

Una demostración similar se tiene cuando $z < 0$. Basta tomar a los elementos de G satisfaciendo $-M < x_0 < ... < x_m < M - z$. ■

El siguiente lema establece propiedades importantes de la función W y la función de valor óptimo V_α^* que serán usadas a lo largo del trabajo.

Lema 2.4. Supongamos que las Hipótesis H1(c) y H2(a) se cumplen. Entonces,

a) para cada $x \in X$, $a \in A(x)$,

$$\int_{\Re^k} W[F(x,a,s)]\rho(s)ds \leq \beta W(x) + b, \tag{2.9}$$

donde $\beta = \beta_0^{1/p}$, $b = b_0^{1/p}$;

b) $\sup_{t\geq 1} E_x^\pi[W^p(x_t)] < \infty$, $\sup_{t\geq 1} E_x^\pi[W(x_t)] < \infty$, para cada $\pi \in \Pi$, $x \in X$;

c) existe una constante C tal que

$$V_\alpha^*(x) \leq CW(x) \qquad x \in X. \tag{2.10}$$

Demostración. a) Por (2.2) tenemos

$$\int_{\Re^k} W[F(x,a,s)]\rho(s)ds \leq \left[\int_{\Re^k} W^p[F(x,a,s)]\rho(s)ds\right]^{1/p} \leq$$

$$[\beta_0 W^p(x) + b_0]^{1/p} \leq \beta_0^{1/p} W(x) + b_0^{1/p}.$$

Tomando $\beta := \beta_0^{1/p}$ y $b := b_0^{1/p}$ obtenemos la parte a).

b) De (2.2) y (1.3) obtenemos, para cada $t \in \mathbb{N}$,

$$E_x^\pi[W^p(x_t) \mid h_t] = \int_{\Re^k} W^p[F(x_{t-1}, a_{t-1}, s)]\rho(s)ds \leq \beta_0 W^p(x_{t-1}) + b_0, \quad x \in X, \ \pi \in \Pi.$$

Tomando esperanza E_x^π en ambos lados de esta relación,

$$E_x^\pi[W^p(x_t)] \leq \beta_0 E_x^\pi[W^p(x_{t-1})] \ + b_0\,, \quad x \in X, \ \pi \in \Pi, \ t \in \mathbb{N}.$$

Iterando esta desigualdad y usando el hecho de que $\beta_0 < 1$ y $W^p(\cdot) \geq 1$, obtenemos

$$E_x^\pi[W^p(x_t)] \leq \beta_0^t W^p(x) \ + (1 + \beta_0 + ... + \beta_0^{t-1})b_0 \leq W^p(x) + b_0/(1-\beta_0)$$

$$\leq [1 + b_0/(1-\beta_0)]W^p(x) < \infty.$$

De manera similar, usando (2.9) se prueba que

$$E_x^\pi[W(x_t)] \leq [1 + b/(1-\beta)]W(x) < \infty \tag{2.11}$$

c) De la relación (2.11) y la Hipótesis H1(c), tenemos que para cada $t \geq 0$

$$E_x^\pi[c(x_t, a_t)] \leq E_x^\pi[W(x_t)] \leq [1 + b/(1-\beta)]W(x). \tag{2.12}$$

Finalmente de (1.4) y (1.7) concluimos que

$$V_\alpha^*(x) \leq V_\alpha(\pi, x) \leq CW(x), \tag{2.13}$$

donde $C := [1/(1-\alpha)][1 + b/(1-\beta)]$. ■

En el siguiente teorema se muestra que la función de valor óptimo V_α^* satisface la *ecuación de optimalidad* y se garantiza la existencia de $\delta-$ minimizadores bajo el criterio de costo descontado. Este resultado es clave para mostrar la existencia de políticas adaptadas asintóticamente óptimas.

Teorema 2.5. Supogamos que las Hipótesis H1 y H2(a) se cumplen. Entonces

a) la función de valor óptimo $V_\alpha^*(\cdot)$ satisface la ecuación de optimalidad α-descontada, i.e.

$$V_\alpha^*(x) = \inf_{a \in A(x)} \left\{ c(x,a) + \alpha \int_{\Re^k} V_\alpha^*[F(x,a,s)]\rho(s)ds \right\}, \quad x \in X; \tag{2.14}$$

b) para cada $\delta > 0$, existe una política estacionaria $\mathbf{f} \in \mathbb{F}$ tal que

$$c(x,f) + \alpha \int_{\Re^k} V_\alpha^*[F(x,f,s)]\rho(s)ds \leq V_\alpha^*(x) + \delta, \quad x \in X. \tag{2.15}$$

La parte a) es el Teorema 4.1(b) en Hernández-Lerma (1994), mientras que la parte b) se sigue del Lema 1.3.

2.3 Estimación de la densidad

En esta sección desarrollamos el método de estimación de la densidad ρ el cual lo usaremos en la construcción de las políticas adaptadas en la próxima sección. Para esto, denotemos por $\xi_0, \xi_1, ..., \xi_{t-1}$ a las realizaciones independientes, observadas hasta el tiempo $t-1$, de vectores aleatorios con la densidad desconocida $\rho \in D_0$. Sea $\hat{\rho}_t := \hat{\rho}_t(s; \xi_0, \xi_1, ..., \xi_{t-1})$, $s \in \Re^k$ un estimador arbitrario de ρ tal que $\hat{\rho}_t \in L_q$ y, para algún $\gamma > 0$,

$$E \left\| \rho - \hat{\rho}_t \right\|_q^{\frac{q}{2}} = \mathbf{O}(t^{-\gamma}) \quad \text{cuando} \quad t \to \infty. \tag{2.16}$$

Supondremos que los estimadores $\hat{\rho}_t$, $t \in \mathbb{N}$, no necesariamente son densidades.

Ahora, el estimador de ρ que usaremos lo definimos como la proyección ρ_t de $\hat{\rho}_t$ sobre el conjunto de densidaes $D := D_1 \cap D_2$, donde0

$$D_1 := \{\mu : \mu \text{ es una densidad en } \Re^k, \mu \in L_q \text{ y } \mu(s) \leq \bar{\rho}\,(s) \; c.s.\}; \tag{2.17}$$

$$\begin{aligned} D_2 := \Big\{\mu : \mu \text{ es una densidad en } \Re^k, \mu \in L_q, \textstyle\int W[F(x,a,s)]\mu(s)ds \\ \leq \beta W(x) + b, \; (x,a) \in \mathbb{K}\} \end{aligned} \tag{2.18}$$

La existencia del estimador ρ_t la garantiza el Lema 2.6 enunciado mas adelante, debido a que el Lema 2.6 muestra que D es un conjunto cerrado y convexo en L_q, y de esta manera, usando un resultado sobre la existencia de la "mejor aproximación " en L_q [ver, por ejemplo, Proposiciónes 2 y 3, p. 343 en Köthe (1969)], existe una única densidad $\rho_t \in D$ satisfaciendo

$$\left\| \rho_t - \hat{\rho}_t \right\|_q = \inf_D \left\| \mu - \hat{\rho}_t \right\|_q , \quad t \in \mathbb{N}. \tag{2.19}$$

Es decir, la densidad ρ_t es la mejor aproximación del estimador $\hat{\rho}_t$ sobre el conjunto D.

Observemos que bajo la Hipótesis H2 y el Lema 2.4(a) tenemos que $\rho \in D_0 \subset D$.

Lemma 2.6. Bajo la Hipótesis H2(b), (c), el conjunto $D = D_1 \cap D_2$ definido en (2.17) y (2.18) es un subconjunto cerrado y convexo de L_q.

Demostración. Primero mostraremos que D es cerrado. Sea $\mu_n \in D$ una sucesión tal que $\mu_n \xrightarrow{L_q} \mu \in L_q$. Supongamos que $\mu \notin D_1$, es decir, existe $A \subset \Re^k$ con $m(A) > 0$ tal que $\mu(s) > \bar{\rho}(s)$, $s \in A$, donde m es la medida de Lebesgue en $\Re^k$. Entonces, para algún $\delta > 0$ y $A' \subset A$ con $m(A') > 0$ tenemos que

$$\mu(s) > \bar{\rho}(s) + \delta, \quad s \in A'. \tag{2.20}$$

Ahora, como $\mu_n \in D_1$, $n \in \mathbb{N}$, existe $H \subset \Re^k$ con $m(H) = 0$ tal que

$$\mu_n(s) \leq \bar{\rho}(s), \quad s \in \Re^k \setminus H, \quad n \in \mathbb{N}. \tag{2.21}$$

Combinando (2.20) y (2.21) obtenemos

$$|\mu(s) - \mu_n(s)| \geq \delta, \quad s \in A' \cap (\Re^k \setminus H), \quad n \in \mathbb{N}.$$

Como $m(A' \cap (\Re^k \setminus H)) > 0$, tenemos que μ_n no converge a μ en medida, lo cual contradice a la convergencia en L_q. Por lo tanto $\mu \in D_1$.

Ahora, para mostrar que $\mu \in D_2$, usando el hecho

$$\int_{\Re^k} W[F(x,a,s)]\mu_n(s)ds \leq \beta W(x) + b, \quad (x,a) \in \mathbb{K}, \ n \in \mathbb{N},$$

es suficiente mostrar que

$$\int_{\Re^k} W[F(x,a,s)]\mu_n(s)ds \to \int_{\Re^k} W[F(x,a,s)]\mu(s)ds \quad \text{cuando} \quad n \to \infty,$$

para toda $(x,a) \in \mathbb{K}$.

Por (2.3) $W[F(x,a,s)] \leq W(x)\varphi(s)$, $(x,a) \in \mathbb{K}$, $s \in \Re^k$. De aquí, para cualquier $(x,a) \in \mathbb{K}$ fijo, y $\varepsilon = (q-1)/2$

$$I \;:=\; \left| \int_{\Re^k} W[F(x,a,s)][\mu_n(s)-\mu(s)]ds \right| \le W(x) \left| \int_{\Re^k} \varphi(s)[\mu_n(s)-\mu(s)]ds \right|$$

$$\le\; W(x) \int_{\Re^k} \varphi(s) \left|\mu_n(s)-\mu(s)\right|^{(1-2\varepsilon)/2} \left|\mu_n(s)-\mu(s)\right|^{(1+2\varepsilon)/2} ds \qquad (2.22)$$

Aplicando la desigualdad de Hölder en el último término de (2.22) y tomando en cuenta que $\mu, \mu_n \in D_1$ obtenemos

$$I \;\le\; W(x) \left[\int_{\Re^k} \varphi^2(s) \left|\mu_n(s)-\mu(s)\right|^{1-2\varepsilon} ds \right]^{\frac{1}{2}} \left[\int_{\Re^k} \left|\mu_n(s)-\mu(s)\right|^{1+2\varepsilon} ds \right]^{\frac{1}{2}}$$

$$\le\; W(x) \left[\int_{\Re^k} \varphi^2(s) \left|2\,\bar{\rho}\,(s)\right|^{1-2\varepsilon} ds \right]^{\frac{1}{2}} \left[\int_{\Re^k} \left|\mu_n(s)-\mu(s)\right|^{1+2\varepsilon} ds \right]^{\frac{1}{2}}$$

$$\le\; MW(x) \left[\int_{\Re^k} \left|\mu_n(s)-\mu(s)\right|^{1+2\varepsilon} ds \right]^{\frac{1}{2}}. \qquad (2.23)$$

para alguna constante M. La última desigualdad se debe a la Hipótesis H2(c).

Como $q = 1+2\varepsilon$ y $\mu_n \xrightarrow{L_q} \mu$, el lado derecho de (2.23) tiende a cero cuando $n \to \infty$.

Finalmente, para demostrar que D es cerrado, mostraremos que μ es una densidad en $\Re^k$. Para esto, observemos que $\mu \ge 0$ c.s..

Por otro lado, similarmente a (2.23),

$$\left| 1 - \int_{\Re^k} \mu(s)ds \right| \le \int_{\Re^k} \left|\mu_n(s)-\mu(s)\right| ds$$

$$\le \left[\int_{\Re^k} \left[2\,\bar{\rho}\,(s)\right]^{1-2\varepsilon} ds \right]^{\frac{1}{2}} \left[\int_{\Re^k} \left|\mu_n(s)-\mu(s)\right|^{1+2\varepsilon} ds \right]^{\frac{1}{2}}$$

$$\leq M_1 \left[\int_{\Re^k} |\mu_n(s) - \mu(s)|^{1+2\varepsilon} \, ds \right]^{\frac{1}{2}} \to 0, \quad as \quad n \to \infty,$$

para alguna constante $M_1 > 0$.

La convexidad de D_1y D_2 es verificada directamente de (2.17) y (2.18). ■

Tomando en cuenta las desigualdades (2.22) y (2.23) obtenemos el siguiente resultado.

Corolario 2.7. Bajo la Hipótesis H2,

$$\int_{\Re^k} \varphi(s) \, |\rho_t(s) - \rho(s)| \, ds \leq M \|\rho_t - \rho\|_q^{q/2}, \; t \in \mathbb{N}.$$

En efecto, basta comparar el segundo término de (2.22) con el último término de (2.23), sustituir ρ_t y ρ por μ_n y μ respectivamente y recordar que $q = 1 + 2\varepsilon$.

A continuación mostramos un ejemplo de un estimador que satisface la relación (2.16).

Ejemplo 2.8. [Hasminskii e Ibragimov (1990)]. Sea $\{z_t\}$ una sucesión de números reales positivos tal que $\lim_{t\to\infty} z_t^k/t = 0$ y $z_t \to \infty$ cuando $t \to \infty$. Definimos

$$\hat{\rho}_t(s) = \hat{\rho}_t(s; z_t, \xi_0, \xi_1, ..., \xi_{t-1}) := \frac{1}{t} \sum_{i=0}^{t-1} V_{z_t}(s - \xi_i), \quad s \in \Re^k, \tag{2.24}$$

donde $V_z(y)$ es un kernel del tipo de Vallée Poussin [ver, por ejemplo, Devroye (1987)]:

$$V_z(y) = \prod_{n=1}^{k} \frac{\cos zy_n - \cos 2zy_n}{\pi z y_n^2}, \qquad y = (y_1, y_2, ..., y_k) \in \Re^k, \; z > 0.$$

En particular, como se muestra en Hasminskii e Ibragimov (1990), si escogemos $z_t = t^r$ donde $r = sq(q-1)/[q(s+1) - 1] < 1$ y $s = 1/kq$, entonces (2.1) implica que el estimador (2.24) satisface la relación (2.16) con $\gamma := r/2$. ■

Definimos la pseudo norma (posiblemente tomando valores infinitos) en el espacio de todas las densidades μ en $\Re^k$ por

$$\|\mu\| := \sup_X [W(x)]^{-1} \sup_{A(x)} \int_{\Re^k} W[F(x,a,s)]\mu(s)ds. \tag{2.25}$$

Teorema 2.9. Supongamos que las Hipótesis H1 y H2 se cumplen. Entonces

$$E\,\|\rho_t - \rho\| = \mathbf{O}(t^{-\gamma}) \quad \text{cuando} \quad t \to \infty.$$

Demostración: De (2.25) y (2.3) tenemos

$$\|\rho_t - \rho\| \leq \int_{\Re^k} \sup_X [W(x)]^{-1} \sup_{A(x)} W[F(x,a,s)]\,|\rho_t(s) - \rho(s)|\,ds$$

$$= \int_{\Re^k} \varphi(s)\,|\rho_t(s) - \rho(s)|\,ds, \quad t \in \mathbb{N}.$$

Por el Corolario 2.7 obtenemos:

$$\|\rho_t - \rho\| \leq M\,\|\rho_t - \rho\|_q^{q/2}, \quad t \in \mathbb{N}, \tag{2.26}$$

para alguna constante M.

Por otro lado,

$$\|\rho_t - \rho\|_q \leq \|\rho_t - \hat{\rho}_t\|_q + \|\hat{\rho}_t - \rho\|_q \leq 2\,\|\hat{\rho}_t - \rho\|_q,$$

donde la última desigualdad se sigue de (2.19) y el hecho de que $\rho \in D$. De aquí,

$$\|\rho_t - \rho\|_q^{q/2} \leq 2^{q/2}\,\|\hat{\rho}_t - \rho\|_q^{q/2}, \quad t \in \mathbb{N}. \tag{2.27}$$

Combinando (2.26) y (2.27) y usando (2.16) obtenemos el resultado deseado. ■

Observación 2.10. Puede suceder que $\|\rho_t - \rho\|$ no sea una variable aleatoria. En este caso el Teorema 2.9 muestra la existencia de cotas superiores medibles b_t para $\|\rho_t - \rho\|$ tal que $\lim_{t\to\infty} Eb_t = 0$ (ver Observación 2.2).

2.4 Políticas adaptadas

Como se mencionó en la introducción del capítulo, las características propias del criterio de costo descontado hacen que no necesariamente existan políticas adaptadas α- óptimas. Es por esta razón que en este caso usamos el criterio de optimalidad asintótica introducido originalmente por Mandl (1974) y usado despues para el caso descontado por Schäl (1987).

Definición 2.11. a) [Schäl (1987)] Una política π se dice que es asintóticamente óptima descontada si para cada $x \in X$,

$$E_x^\pi [\Phi(x_t, a_t)] \to 0 \text{ cuando } t \to \infty,$$

donde $a_t = \pi_t(h_t)$ y

$$\Phi(x,a) := c(x,a) + \alpha \int_{\Re^k} V_\alpha^*[F(x,a,s)]\rho(s)ds - V_\alpha^*(x), \quad (x,a) \in \mathbb{K}, \tag{2.28}$$

es llamada *función de discrepancia,* la cual es no negativa en vista del Teorema 2.5.

b) Sea $\delta \geq 0$. Una política π es $\delta-$asintóticamente óptima descontada si para cada $x \in X$,

$$\lim_{t\to\infty} E_x^\pi [\Phi(x_t, a_t)] \leq \delta.$$

Φ es llamada función de discrepancia porque puede ser interpretada como una medida de la "desviación de optimalidad". Para mas detalle ver, por ejemplo, Hernández-Lerma (1989).

Para construir las políticas adaptadas que presentamos en este capítulo, nos basamos en las ideas usadas en Hernéndez-Lerma y Cavazos-Cadena (1990), las cuales consisten en remplazar a la densidad desconocida ρ por sus estimadores ρ_t y hacer uso de las correspondientes ecua-

ciones de optimalidad. Por esta razón, necesitamos extender algunos resultados de la Seccion 2 al contexto de las densidades $\rho_t \in D$, definidas en la Sección 3.

Hipótesis H3. Para cada $u \in L_W^\infty$, $t \in \mathbb{N}$ y $r \in \Re$, el conjunto

$$\left\{(x,a) : \int_{\Re^k} u[F(x,a,s)]\rho_t(s)ds \leq r\right\}$$

es de Borel en $\mathbb{K}$.

Bajo la Hipótesis H3, los resultados del Lema 2.4 y del Teorema 2.5, con ρ_t en lugar de ρ, son válidos debido a que para demostrarlos, además de las Hipótesis H1(b), (c), solo se usa la desigualdad (2.9) y la Hipótesis H1(a) [ver Hernández-Lerma (1994)], las cuales pueden ser sustituidas por el hecho de que $\rho_t \in D_2$ y la Hipótesis H3 respectivamente. Con esto tenemos el siguiente resultado.

Teorema 2.12. Bajo las Hipótesis H1(b), (c) y H3 tenemos:

a) Para cada $t \in \mathbb{N}$ existe una función $V_t \in L_W^\infty$ tal que

$$V_t(x) = \inf_{A(x)} \left\{c(x,a) + \alpha \int_{\Re^k} V_t[F(x,a,s)]\rho_t(s)ds\right\}, \quad x \in X. \tag{2.29}$$

b) Para cada $t \in \mathbb{N}$ y $\delta_t > 0$, existe una política estacionaria $\mathbf{f}_t \in \mathbb{F}$ tal que

$$c(x,f_t) + \alpha \int_{\Re^k} V_t[F(x,f_t,s)]\rho_t(s)ds \leq V_t(x) + \delta_t, \quad x \in X. \tag{2.30}$$

c) Existe una constante C^* tal que $\sup_{t\geq 1} \|V_t\|_W \leq C^*$.

d) Si $\bar{V}_0 \equiv 0$ y

$$\bar{V}_t\,(x) = \inf_{A(x)} \left\{c(x,a) + \alpha \int_{\Re^k} \bar{V}_{t-1}\,[F(x,a,s)]\rho_t(s)ds\right\}, \quad x \in X, \quad t \in \mathbb{N}, \tag{2.31}$$

entonces $\left\|\bar{V}_t\right\|_W \leq \bar{C}$ para alguna constante $\bar{C}$; y para cada $\bar{\delta}_t > 0$ existe una política estacionaria $\bar{f}_t \in \mathbb{F}$ tal que

$$c(x, \bar{f}_t) + \alpha \int_{\Re^k} \bar{V}_{t-1}\,[F(x, \bar{f}_t, s)]\rho_t(s)ds \leq \bar{V}_t\,(x) + \bar{\delta}_t\,, \quad x \in X. \tag{2.32}$$

A continuación definimos las políticas adaptadas π^* y $\bar{\pi}$ las cuales se pueden considerar como una extensión de las políticas adaptadas llamadas "Principio de Estimación y Control" (PEC-política), propuesta en Mandl (1974), y la que se obtiene al aplicar esquemas de "Iteración de Valores No estacionarios" (IVN- política), propuesta en Hernández-Lerma y Marcus (1985).

Definición 2.13. Sean $\{\delta_t\}$ y $\left\{\bar{\delta}_t\right\}$ sucesiones arbitrarias de números positivos; $\{f_t\}$ y $\left\{\bar{f}_t\right\}$ sucesiones de funciones satisfaciendo (2.30) y (2.32) respectivamente.

a) La PEC- política daptada $\pi^* = \{\pi_t^*\}$ se define como

$$\pi_t^*(h_t) = \pi_t^*(h_t; \rho_t) := f_t(x_t), \quad h_t \in \mathbb{H}_t, \quad t \in \mathbb{N},$$

donde $\pi_0^*(x)$ es cualquier control fijo.

b) La IVN- política adaptada $\bar{\pi} = \{\bar{\pi}_t\}$ se define como

$$\bar{\pi}_t\,(h_t) = \bar{\pi}_t\,(h_t; \rho_t) := \bar{f}_t\,(x_t), \quad h_t \in \mathbb{H}_t, \quad t \in \mathbb{N},$$

donde $\bar{\pi}_0\,(x)$ es cualquier control fijo.

Supongamos que $\{\delta_t\}$ y $\{\bar{\delta}_t\}$ convergen y sea $\delta := \lim_{t\to\infty} \delta_t$, $\bar{\delta} := \lim_{t\to\infty} \bar{\delta}_t$. El resultado principal de este capítulo es el siguiente:

Teorema 2.14. Bajo las Hipótesis H1, H2 y H3 la política adaptada π^* es $\delta-$asintóticamente óptima descontada y la política adaptada $\bar{\pi}$ es $\bar{\delta}$ − asintóticamente óptima descontada.

En particular, si $\delta = \bar{\delta} = 0$ entonces las políticas π^* y $\bar{\pi}$ son asintóticamnete óptimas descontadas.

Obsevación 2.15. En el resto del trabajo estaremos usando repetidamente las siguientes desigualdades:

$$|u(x)| \leq \|u\|_W W(x) \tag{2.33}$$

y

$$\int_{\Re^k} |u[F(x,a,s)]|\, \mu(s)ds \leq \|u\|_W [\beta W(x) + b] \tag{2.34}$$

para toda $u \in L_W^\infty$, $\mu \in D$, $x \in X$ y $a \in A(x)$. La desigualdad (2.33) es consecuencia de la definición de $\|\cdot\|_W$ y (2.34) se sigue del Lema 2.4(a) y al definición del conjunto D.

La demostración del Teorema 2.14 está basada en el siguiente resultado.

Lema 2.16. Bajo las Hipótesis H1, H2 y H3, para cada $x \in X$ y $\pi \in \Pi$,

$$a) \lim_{t\to\infty} E_x^\pi \|V_t - V^*\|_W = 0 \qquad \text{y} \qquad b) \lim_{t\to\infty} E_x^\pi \left\|\bar{V}_t - V^*\right\|_W = 0\,.$$

Demostración. a) Para cada $\mu \in D$ definimos el operador

$$T_\mu u(x) = \inf_{A(x)} \left\{ c(x,a) + \alpha \int_{\Re^k} u[F(x,a,s)]\mu(s)ds \right\}, \quad x \in X,\ u \in L_W^\infty. \tag{2.35}$$

Por la Hipótesis H1(c), la definición del conjunto D y (2.34) tenemos que T mapea L_W^∞ en si mismo ($T : L_W^\infty \to L_W^\infty$).

Sea $\theta \in (\alpha, 1)$ un número arbitrario fijo, y $\bar{W}\ (x) := W(x) + d$, $x \in X$, donde $d := b\,(\theta/\alpha - 1)^{-1}$. Denotamos por $L_{\bar{W}}^\infty$ al espacio de funciones medibles $u : X \to \Re$ con la norma

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

De aquí tenemos que

$$\|u\|_{\bar{W}} \leq \|u\|_W\,,\ u \in L^\infty_W. \tag{2.36}$$

Por otro lado,

$$\|u\|_W = \sup_X \left\{ \frac{u(x)}{W(x)+d} \cdot \frac{W(x)+d}{W(x)} \right\} \leq \|u\|_{\bar{W}} \left(1 + d/\inf_X W(x)\right). \tag{2.37}$$

De (2.36) y (2.37) tenemos que $L^\infty_W = L^\infty_{\bar{W}}$ y las normas $\|\cdot\|_W$ y $\|\cdot\|_{\bar{W}}$ son equivalentes. De esta manera, para demostrar la parte (a) es suficiente mostrar que

$$\lim_{t\to\infty} E^\pi_x \|V_t - V^*_\alpha\|_{\bar{W}} = 0. \tag{2.38}$$

En el Lema 2 de Van Nunen y Wessels (1978) fué demostrado que la desigualdad

$$\int_{\Re^k} W[F(x,a,s)]\mu(s)ds \leq W(x) + b$$

implica que el operador T_μ en (2.35) es una contracción con respecto a la norma $\|\cdot\|_{\bar{W}}$ con modulo θ, es decir,

$$\|T_\mu v - T_\mu u\|_{\bar{W}} \leq \theta \|v - u\|_{\bar{W}}\,,\quad v,\, u \in L^\infty_W\,. \tag{2.39}$$

Por lo tanto. de (2.14) y (2.39) tenemos que la función V^*_α es el único punto fijo, en L^∞_W, del operador T_ρ, mientras que la función V_t es el único punto fijo, en L^∞_W, de T_{ρ_t}, $t \in \mathbb{N}$; esto es

$$T_\rho V^*_\alpha = V^*_\alpha,\quad T_{\rho_t} V_t = V_t. \tag{2.40}$$

De aquí

$$\|V^*_\alpha - V_t\|_{\bar{W}} = \left\|T_\rho V^*_\alpha - T_{\rho_t} V_t\right\|_{\bar{W}} \leq \left\|T_\rho V^*_\alpha - T_{\rho_t} V^*_\alpha\right\|_{\bar{W}} + \left\|T_{\rho_t} V^*_\alpha - T_{\rho_t} V_t\right\|_{\bar{W}}$$

$$\leq \left\|T_\rho V^*_\alpha - T_{\rho_t} V^*_\alpha\right\|_{\bar{W}} + \theta \|V^*_\alpha - V_t\|_{\bar{W}}\,,$$

o

$$\|V_\alpha^* - V_t\|_{\bar{W}} \leq \frac{1}{1-\theta} \left\| T_\rho V_\alpha^* - T_{\rho_t} V_\alpha^* \right\|_{\bar{W}}, \quad t \in \mathbb{N}. \tag{2.41}$$

Por otro lado, usando (2.25), el Lema 2.4(c) y el hecho de que $[\bar{W}\ (\cdot)]^{-1} < [W(\cdot)]^{-1}$, obtenemos

$$\left\| T_\rho V_\alpha^* - T_{\rho_t} V_\alpha^* \right\|_{\bar{W}} \leq \alpha \sup_X [\bar{W}\ (x)]^{-1} \sup_{A(x)} \int_{\Re^k} V_\alpha^*[F(x,a,s)]\, |\rho(s) - \rho_t(s)|\, ds$$

$$\leq \alpha C \sup_X [W(x)]^{-1} \sup_{A(x)} \int_{\Re^k} W[F(x,a,s)]\, |\rho(s) - \rho_t(s)|\, ds = \alpha C \|\rho - \rho_t\|, \quad t \in \mathbb{N}. \tag{2.42}$$

Como ρ_t no depende de π y x, tenemos que $E_x^\pi \|\rho - \rho_t\| = E \|\rho - \rho_t\|$. Por lo tanto, combinando las desigualdades (2.41) y (2.42) con el Teorema 2.9, concluimos que (2.38) es válido.

b) Usando argumentos similares a la demostración dela parte (a), de las relaciones (2.14), (2.31), (2.39) y (2.42), obtenemos

$$\left\| V_\alpha^* - \bar{V}_{t+1} \right\|_{\bar{W}} \leq \left\| T_\rho V_\alpha^* - T_{\rho_t} V_\alpha^* \right\|_{\bar{W}} + \theta \left\| V_\alpha^* - \bar{V}_t \right\|_{\bar{W}} \leq \alpha C \|\rho - \rho_t\| + \theta \left\| V_\alpha^* - \bar{V}_t \right\|_{\bar{W}}.$$

Por lo tanto, para cada $x \in X$, $\pi \in \Pi$, $t \in \mathbb{N}$,

$$E_x^\pi \left\| V_\alpha^* - \bar{V}_{t+1} \right\|_{\bar{W}} \leq \alpha C E_x^\pi \|\rho - \rho_t\| + \theta E_x^\pi \left\| V_\alpha^* - \bar{V}_t \right\|_{\bar{W}}, \tag{2.43}$$

Ahora, por el Lema 2.4(c), Teorema 2.12(d) y la equivalencia de las normas $\|\cdot\|_W$ y $\|\cdot\|_{\bar{W}}$ tenemos que $\lambda := \limsup_{t\to\infty} E_x^\pi \left\| V_\alpha^* - \bar{V}_t \right\|_{\bar{W}} < \infty$. Tomando lim sup cuando $t \to \infty$ en ambos lados de (2.43) y aplicando el Teorema 2.9, llegamos a que $\lambda \leq \theta\lambda$, lo cual implica que $\lambda = 0$. Esto demuestra la parte (b). ■

Demostración del Teorema 2.14.

Para cada $t \in \mathbb{N}$ definimos las funciones $\Phi_t^* : \mathbb{K} \to \Re$ y $\bar{\Phi}_t : \mathbb{K} \to \Re$ como:

$$\Phi_t^*(x,a) \;:=\; c(x,a) + \alpha \int_{\Re^k} V_t[F(x,a,s)]\rho_t(s)ds - V_t(x);$$

$$\bar{\Phi}_t\,(x,a) \;:=\; c(x,a) + \alpha \int_{\Re^k} \bar{V}_{t-1}\,[F(x,a,s)]\rho_t(s)ds - \bar{V}_t\,(x).$$

Por el Teorema 2.12(a), (d) ambas funciones son no negativas.

Usando las definiciones de Φ_t^* y Φ; y sumando y restando el término $\alpha \int_{\Re^k} V_t[F(x,a,s)]\rho(s)ds$, tenemos

$$\begin{aligned}
|\Phi_t^*(x,a) - \Phi(x,a)| \;\leq\; & |V_\alpha^*(x) - V_t(x)| + \alpha \int_{\Re^k} V_t[F(x,a,s)]\,|\rho_t(s) - \rho(s)|\,ds \\
& + \alpha \int_{\Re^k} |V_t[F(x,a,s)] - V_\alpha^*[F(x,a,s)]|\,\rho(s)ds \\
\leq\; & \|V_\alpha^* - V_t\|_W\, W(x) + \alpha C^* \int_{\Re^k} W[F(x,a,s)]\,|\rho_t(s) - \rho(s)|\,ds \\
& + \alpha[\beta W(x) + b]\,\|V_t - V_\alpha^*\|_W\,,
\end{aligned}$$

para cada $(x,a) \in \mathbb{K}$, $t \in \mathbb{N}$ [ver Teorema 2.12(c) y Lema 2.4(a)]. Por lo tanto, usando la definición de $\|\cdot\|$ en (2.25), la equivalencia de las normas $\|\cdot\|_W$ y $\|\cdot\|_{\bar{W}}$ y las desigualdades (2.41) y (2.42),

$$\begin{aligned}
\sup_X [W(x)]^{-1} \sup_{A(x)} |\Phi_t^*(x,a) - \Phi(x,a)| \;&\leq\; \frac{\alpha C}{1-\theta}\,\|\rho_t - \rho\| + \alpha C^*\,\|\rho_t - \rho\| \\
+\alpha\left[\beta + \frac{b}{\inf_X W(x)}\right]\frac{\alpha C}{1-\theta}\,\|\rho_t - \rho\| \;&=\; C'\,\|\rho_t - \rho\|\,, \qquad (2.44)
\end{aligned}$$

donde $C' = \alpha C^* + \{1 + \alpha\,[\beta + b/\inf_X W(x)]\}\,(1-\theta)^{-1}\alpha C$.

Por otro lado, de la definición de la política π^* [ver Definición 2.13(a)] y de las funciones f_t en (2.30) tenemos que $\Phi_t^*(\cdot, \pi_t^*(\cdot)) \leq \delta_t$, $t \in \mathbb{N}$. De esta manera

$$\begin{aligned}\Phi(x_t,\pi_t^*(h_t)) &\leq |\Phi(x_t,\pi_t^*(h_t)) - \Phi_t^*(x_t,\pi_t^*(h_t)) + \delta_t| \leq \sup_{A(x_t)} |\Phi(x_t,a) - \Phi_t^*(x_t,a)| + \delta_t \\ &\leq W(x_t) \sup_X [W(x)]^{-1} \sup_{A(x)} |\Phi(x,a) - \Phi_t^*(x,a)| + \delta_t \\ &\leq W(x_t)\zeta_t + \delta_t \ , \ \ t \in \mathbb{N}, \end{aligned} \tag{2.45}$$

donde $\zeta_t := C' \|\rho_t - \rho\|$. De (2.45) tenemos que

$$E_x^{\pi^*}[\Phi(x_t,a_t)] \leq E_x^{\pi^*}[W(x_t)\zeta_t] + \delta_t,$$

y por lo tanto, para mostrar que la política π^* es $\delta-$ asintóticamente óptima descontada (ver Definición 2.11) es suficiente mostrar que

$$\limsup_{t\to\infty} E_x^{\pi^*}[W(x_t)\zeta_t] = 0. \tag{2.46}$$

Para esto, observemos primero que si $\mu \in D_2$ entonces por (2.18) y (2.25)

$$\|\mu\| \leq \sup_X [W(x)]^{-1}[\beta W(x) + b] \leq \beta + b/\inf_X W(x). \tag{2.47}$$

Usando esta desigualdad, es fácil ver que $\sup_{t\geq 1} \|\rho_t - \rho\| \leq C_1 < \infty$ para alguna constante C_1. De aquí y por el Teorema 2.9 tenemos la convergencia en probabilidad:

$$\zeta_t \xrightarrow{P_x^{\pi^*}} 0 \text{ cuando } t \to \infty. \tag{2.48}$$

Además, por el Lema 2.4(b) obtenemos

$$\sup_{t\geq 1} E_x^{\pi^*}[W(x_t)\zeta_t]^p < (C')^p C_1^p \sup_{t\geq 1} E_x^{\pi^*}[W^p(x_t)] < \infty.$$

Esto implica que la sucesión $\{W(x_t)\zeta_t\}$ es $P_x^{\pi^*}-$uniformemente integrable [ver Lema 7.6.9, p. 301 en Ash (1972)]. De esta manera, usando un criterio de convergencia de integrales [ver, por ejemplo, Teorema 7.5.2 en Ash (1972)], la relación (2.46) queda demostrada si

$$W(x_t)\zeta_t \xrightarrow{P_x^{\pi^*}} 0 \text{ cuando } t \to \infty. \tag{2.49}$$

Para esto tenemos:

$$\begin{aligned} P_x^{\pi^*}[W(x_t)\zeta_t > \tau] &= P_x^{\pi^*}[W(x_t)\zeta_t > \tau, W(x_t) < l] + P_x^{\pi^*}[W(x_t)\zeta_t > \tau, W(x_t) > l] \\ &\leq P_x^{\pi^*}[\zeta_t > \frac{\tau}{l}] + P_x^{\pi^*}[W(x_t) > l] \leq P_x^{\pi^*}\left[\zeta_t > \frac{\tau}{l}\right] + \frac{E_x^{\pi^*}[W(x_t)]}{l}, \end{aligned}$$

donde τ y l son números positivos arbitrarios. Por lo tanto, la relación (2.49) se obtiene haciendo $t \to \infty$ en esta desigualdad y aplicando (2.48) y el Lema 2.4(b).

Las ideas para la demostración de la segunda parte del teorema son similares. Primero podemos mostrar que

$$\sup_X [W(x)]^{-1} \sup_{A(x)} \left|\bar{\Phi}_{t+1}(x,a) - \Phi(x,a)\right| \leq$$

$$\left\|V_\alpha^* - \bar{V}_{t+1}\right\|_W + \alpha \bar{C} \|\rho_{t+1} - \rho\| + \alpha[\beta + b/\inf_X W(x)] \left\|\bar{V}_t - V_\alpha^*\right\|_W := \bar{\zeta}_t, \ t \in \mathbb{N}.$$

De nuevo, $\lim_{t\to\infty} E_x^{\bar{\pi}}[\bar{\zeta}_t] = 0$ (ver Lema 2.16(b) y Teorema 2.9), y las variables aleatorias $\bar{\zeta}_t$ son uniformemente acotadas debido al Lema 2.4(c), Teorema 2.12(d) y (2.47). La demostración se concluye repitiendo los argumentos de la primera parte. ■

2.5 Ejemplo

En esta sección mostramos que el ejemplo presentado en el Capítulo I, Sección 1.6, satisface las hipótesis usadas a lo largo del capítulo (H1, H2, H3). Para esto supondremos que se cumple la siguiente hipótesis:

Hipótesis E1.

a) La densidad $\rho := \rho_{\eta_0}\rho_{\chi_0}$ pertenece a $L_q(\Re^2)$ y satisface la desigualdad

$$\|\Delta_z \rho\|_{L_q} \leq L\,|z|^{1/q},\ z \in \Re^2, \tag{2.50}$$

para algunas constantes $L < \infty$, $q > 1$;

b) $E(\varsigma) < 0$, donde $\varsigma := \theta\eta - \chi$;

c) existen constantes $M_1 < \infty$, $M_2 < \infty$, $\alpha > 0$, $r > (q-1)/(2-q)$, tal que

$$\rho_\eta(s_1) \leq M_1 e^{-\alpha s_1},\ s_1 \geq 0 \tag{2.51}$$

y

$$\rho_\chi(s_2) \leq M_2 \min\left\{1, s_2^{-(1+r)}\right\},\ s_2 \geq 0.$$

Una consecuencia de (2.51) es que si $\bar{\lambda} \in (0, \alpha/\theta)$

$$Ee^{\bar{\lambda}\varsigma} = Ee^{-\bar{\lambda}\chi} \int e^{\bar{\lambda}\theta s_1} \rho_\eta(s_1) ds_1 \leq Ee^{-\bar{\lambda}\chi} M_1 \int e^{(\bar{\lambda}\theta - \alpha)s_1} ds_1 < \infty. \tag{2.52}$$

Consideremos la función $H(z) := Ee^{z\varsigma}$. Entonces la Hipótesis E1(b) implica que $H(0) = 1$ y $H'(0) = E(\varsigma) < 0$. De aquí y tomando en cuanta (2.52), existe un número positivo λ^* tal que $H(z) < 1$ para $z \in (0, \lambda^*)$. Fijamos un $\lambda > 0$ tal que

$$\lambda < \min\{\lambda^*, \alpha(2-q)/(2\theta)\}.$$

Observemos que por la continuidad de la función H podemos elegir $p > 1$ tal que

$$H(p\lambda) := \beta_0 < 1. \tag{2.53}$$

Tomamos $W(x) := \bar{b}e^{\lambda x},\ x \in [0, \infty)$, donde $\bar{b}$ es una constante arbitraria.

Ahora, para verificar que la Hipótesis H2(a) se cumple, es suficiente mostrar que la densidad ρ satisface las condiciones (c) y (d) de la Definición 2.1.

Definiendo

$$\bar{\rho}(s) := M_1 M_2 e^{-\alpha s_1} \min\left\{1, s_2^{-(1+r)}\right\},\ s = (s_1, s_2) \in \Re^2, \tag{2.54}$$

claramente tenemos que $\rho \leq \bar{\rho}$ lo cual muestra que la condición (c) se cumple.

Por otro lado, observemos que

$$\begin{aligned}\int_{\Re^2} 1_{(-\infty,y]}[(x+as_1-s_2)^+]\rho(s)ds &= Prob[(x+a\eta-\chi)^+ \leq y] \\ &= Prob[x+a\eta-\chi \leq y], \quad s=(s_1,s_2).\end{aligned}$$

De aquí,

$$\int_{\Re^2} 1_{(-\infty,0]}[(x+as_1-s_2)^+]\rho(s)ds = Prob[x+a\eta-\chi \leq 0], \quad s=(s_1,s_2). \tag{2.55}$$

Además, si $B \in \mathbb{B}(0,\infty)$,

$$\int_{\Re^2} 1_B[(x+as_1-s_2)^+]\rho(s)ds = Prob[x+a\eta-\chi \in B], \quad s=(s_1,s_2). \tag{2.56}$$

De (2.55) y (2.56), tomando p y β_0 como en (2.53), $b_0 := W^p(0)$ y $R = (0,\infty)\times(0,\infty)$ tenemos

$$\begin{aligned}\int_{\Re^2} W^p[F(x,a,s)]\rho(s)ds &= b_0 Prob[x+a\eta-\chi \leq 0] + \left(\bar{b}\right)^p \int_R e^{\lambda p(x+as_1-s_2)}\rho(s)ds \\ &\leq b_0 + \left(\bar{b}\right)^p e^{\lambda px}\int_R e^{\lambda p(as_1-s_2)}\rho(s)ds \leq b_0 + W^p(x)Ee^{\lambda p\varsigma} \\ &= b_0 + H(p\lambda)W^p(x) = b_0 + \beta_0 W^p(x), \quad x\in X,\ a\in A. \end{aligned} \tag{2.57}$$

Esto muetra que la Hipótesis H2(a) se cumple.

Para verificar las Hipótesis H2 (b), (c), sea $K_1 := \{(x,a)\in \mathbb{K} : x+as_1 \leq s_2\}$ y $K_2 := \{(x,a)\in \mathbb{K} : x+as_1 > s_2\}$. De (2.3), para cada $s=(s_1,s_2)\in \Re_2^+$ tenemos

$$\varphi(s) = \sup_X \left(\bar{b}\right)^{-1} e^{-\lambda x}\,\bar{b}\sup_{A(x)} e^{\lambda(x+as_1-s_2)^+} = \sup_X e^{-\lambda x}\sup_{A(x)} e^{\lambda(x+as_1-s_2)^+}$$

$$= \max\left\{\sup_{K_1} e^{-\lambda x}, \sup_{K_2} e^{-\lambda x} e^{\lambda(x+as_1-s_2)}\right\} = \max\left\{1, e^{\lambda(\theta s_1 - s_2)}\right\} < \infty. \qquad (2.58)$$

Por otro lado, observemos que como $\lambda > 0$, $\theta > 0$, $s_1 \geq 0$, $s_2 \geq 0$,

$$\varphi(s) = \max\left\{1, e^{\lambda\theta s_1} e^{-\lambda s_2}\right\} \leq \max\left\{1, e^{\lambda\theta s_1}\right\} = e^{\lambda\theta s_1}, \quad s = (s_1, s_2) \in \Re_2^+.$$

De esta manera, considerando (2.54),

$$\begin{aligned}\int_{\Re^2} \varphi^2(s)\,(\bar{\rho}(s))^{1-2\varepsilon}\,ds &\leq (M_1 M_2)^{1-2\varepsilon} \int_{\Re^2} e^{2\lambda\theta s_1} e^{-\alpha s_1(1-2\varepsilon)} \min\left\{1, s_2^{-(1+r)(1-2\varepsilon)}\right\} ds \\ &= (M_1 M_2)^{1-2\varepsilon} \int_{\Re^2} e^{[2\lambda\theta - \alpha(1-2\varepsilon)]s_1} \min\left\{1, s_2^{-(1+r)(1-2\varepsilon)}\right\} ds. \quad (2.59)\end{aligned}$$

De la Hipótesis E1(c) es fácil ver que $2\lambda\theta - \alpha(1-2\varepsilon) < 0$, lo cual implica que la integral (2.59) es finita. De aquí y (2.58) tenemos que las Hipótesis H2(b), (c) se satisfacen.

Las hipótesis de medibilidad H1(a) y H3 se demuestran directamente. Finalmente, la Hipótesis H1(c) se satisface si tomamos una función de costo por etapa no negativa y s.i.c., arbitraria $c : [0,\infty) \times A \to [0,\infty)$ tal que

$$\sup_A c(x,a) \leq \bar{b}\, e^{\lambda x}, \quad x \in [0,\infty).$$

Capítulo 3

Criterio de Costo Promedio

3.1 Introducción

En este capítulo estudiamos el problema de control adaptado bajo el índice de costo promedio por etapa. Apoyándonos en el método de estimación de la densidad ρ de las v.a. ξ_t, dado en el capítulo anterior, presentamos la construcción de dos políticas adaptadas. La primera de ellas se contruye aplicando una variante del ya conocido algoritmo de iteración de valores no estacionario, y la segunda analizando el costo promedio como límite del caso descontado. A esta última técnica se le conoce como "enfoque del factor de descuento desvaneciente" [vanishing discount factor approach, Blackwell (1962)].

A diferencia del criterio de costo descontado, en el caso promedio sí podemos mostrar que las políticas adaptadas son CP- óptimas, pero su estudio requiere de una herramienta matemática mas sofisticada debido a las dificultades ocasionadas por el análisis asintótico que este índice requiere. Por ejemplo, bajo este criterio, en general no tenemos propiedades contractivas de operadores, y más aun, necesitamos imponer condiciones de ergodicidad restrictivas sobre la clase de los PCMs considerados. Esto último nos permite usar los resultados recientes de Gordienko y Hernández-Lerma (1995a,b) sobre la existencia de soluciones de la desigualdad y ecuación de optimalidad en costo promedio, y sobre la convergencia del algoritmo de iteración de valores, para mostrar la optimalidad de las políticas adaptadas construidas.

Los resultados expuestos en este capítulo sobre control adaptado se encuentran en los

artículos Gordienko y Minjárez-Sosa(1997) y Minjárez-Sosa (1998).

3.2 Hipótesis de optimalidad en costo promedio

Para el resto del capítulo, fijamos una función $W(\cdot)$ satisfaciendo la Hipótesis H1(c). Al igual que en el capítulo anterior, caracterizaremos al conjunto de densidades que definen la clase de PCMs que estaremos trabajando.

Sean d_1 y d_2 métricas en X y A respectivamente, y sea d la métrica en $\mathbb{K}$ definida como:

$$d[(x,a),(x',a')] := \max\{d_1(x,x'), d_2(a,a')\}, \tag{3.1}$$

para todo (x,a) y (x',a') en $\mathbb{K}$. Además, sea $\mathcal{G}$ la clase de todas las funciones no decrecientes $g : [0,\infty) \to [0,\infty)$ tal que $g(s) \to 0$ cuando $s \downarrow 0$.

Definición 3.1. Sea $\bar{\rho}$: $\Re^k \to \Re$ una función medible no negativa y $\varepsilon \in (0,1/2)$, fijos. Denotando $q := 1 + 2\varepsilon$, definimos el conjunto $D_0' = D_0'(\bar{\rho}, L, \beta_0, b_0, p, q, m, \psi, g^Q)$, como el conjunto de todas las densidades μ en $\Re^k$ que satisfacen lo siguiente [ver Def. 2.1]:

a) $\mu \in L_q(\Re^k)$.

b) Existe una constante L tal que para cada $z \in \Re^k$,

$$\|\Delta_z \mu\|_{L_q} \leq L\,|z|^{1/q}.$$

c) $\mu(s) \leq \bar{\rho}\,(s)$ casi dondequiera respecto a la medida de Lebesgue.

d) Para cada política $\mathbf{f} \in \mathbb{F}$ el proceso de Markov x_t^f con probabilidad de transición [ver (1.2)]

$$Q_\mu(B \mid x, f) := \int_{\Re^k} 1_B[F(x,f,s)]\mu(s)ds, \;\; B \in \mathbb{B}(X)$$

es Harris-recurrente positivo [ver Definición A.2.7, Apéndice A].

c) Existe una medida de probabilidad m en $(X,\mathbb{B}(X))$ y un número no negativo $\beta_0 < 1$, y para cada $f \in \mathbb{F}$ una función no negativa $\psi_f : X \to \Re$ tal que para cualquier $x \in X$ y $B \in \mathbb{B}(X)$,

i) $Q_\mu(B \mid x, f) \geq \psi_f(x) m(B)$;

ii) $\int_{\Re^k} W^p[F(x,f,s)]\mu(s)ds \leq \beta_0 W^p(x) + \psi_f(x) \int_X W^p(y)m(dy)$ para algún $p > 1$,
$\int_X W^p(y)m(dy) < \infty$;

iii) $\inf_{f\in\mathbb{F}} \int_X \psi_f(x)m(dx) := \bar{\psi} > 0$.

f) Para cada $x \in X$, existe una función $g_x^Q \in \mathcal{G}$ tal que para toda $a \in A(x)$ y $(x', a') \in \mathbb{K}$,

$$\|Q_\mu(\cdot \mid k) - Q_\mu(\cdot \mid k')\|_W \leq g_x^Q[d(k,k')],$$

donde $k := (x,a)$, $k' := (x',a')$, Q es como en (1.2) y

$$\|\lambda\|_W := \int_X W(y)\,|\lambda|\,(dy),$$

para cualquier medida finita signada λ en X, con $|\lambda| :=$ variación total de λ.

Observación 3.2. a) De la condición e(i), $\psi_f \leq 1$. Por lo tanto, la condición e(ii) implica que para cada $x \in X$, $a \in A(x)$,

$$\int_{\Re^k} W^p[F(x,f,s)]\mu(s)ds \leq \beta_0 W^p(x) + b_0 \tag{3.2}$$

donde $b_0 := \int_X W^p(y)m(dy) < \infty$.

b) El conjunto D_0' es mas restrictivo que el conjunto de densidades D_0 usado en el capítulo anterior para el caso descontado, debido a las características del índice de costo promedio mencionadas en la introducción del presente capítulo. Observemos que las condiciones 3.1(a)-(c) en la definición de D_0', junto con (3.2), son escencialmente las mismas que se usaron para definir el conjunto D_0.

c) Considerando la observación anterior, bajo las Hipótesis H1(c) y suponiendo que $\rho \in D_0'$, tenemos que los resultados del Lema 2.4 son válidos en este contexto, es decir,

i) para cada $x \in X$, $a \in A(x)$,

$$\int_{\Re^k} W[F(x,a,s)]\rho(s)(ds) \leq \beta W(x) + b, \tag{3.3}$$

donde $\beta = \beta_0^{1/p}$, $b = b_0^{1/p}$ con β_0 y b_0 como en (3.2);

ii) para cada $\pi \in \Pi$, $x \in X$,

$$\sup_{t\geq 1} E_x^\pi[W^p(x_t)] < \infty, \qquad \sup_{t\geq 1} E_x^\pi[W(x_t)] < \infty. \tag{3.4}$$

d) La condición 3.1(d) implica que [ver Definición A.2.7 Apéndice A] para cada $\mathbf{f} \in \mathbb{F}$ y $\mu \in D_0'$, $Q_\mu(\cdot \mid x, f)$ tiene una medida de probabilidad invariante q_f, es decir,

$$q_f(B) = \int_X Q_\mu(B \mid x, f) q_f(dx), \quad B \in \mathbb{B}(X). \tag{3.5}$$

e) Es fácil mostrar que si los conjuntos $A(x)$ no dependen de $x \in X$, es decir, $A(x) \equiv A$ para todo $x \in X$, entonces la condición (f) puede sustituirse por la siguiente: Para todo $x \in X$ existe una función $g_x^Q \in \mathcal{G}$ tal que para todo $x' \in X$

$$\sup_A \|Q(\cdot \mid x, a) - Q(\cdot \mid x', a)\|_W \leq g_x^Q[d_1(x, x')].$$

La siguiente hipótesis es equivalente a la Hipótesis H2 la cual la reescribiremos en el contexto de este capítulo, y nuevamente consideraremos la Observación 2.2.

Hipótesis H4.

a) La densidad ρ pertenece a D_0'.

b) La función

$$\varphi(s) := \sup_X [W(x)]^{-1} \sup_{A(x)} W[F(x,a,s)], \tag{3.6}$$

es finita para cada $s \in \Re^k$.

c) $\int_{\Re^k} \varphi^2(s) \left|\bar{\rho}\,(s)\right|^{1-2\varepsilon} ds < \infty$.

El estudio de los PCMs bajo el criterio de costo promedio puede hacerse por medio de la desigualdad o ecuación de optimalidad. Claramente, las hipótesis que garantizan la existencia de una solución a la desigualdad de optimalidad no son tan restrictivas comparándolas con las que garantizan una solución a la ecuación de optimalidad.

Teorema 3.3 Supongamos que las Hipótesis H1 y H4(a) se cumplen. Entonces existe una constante j^* y una función ϕ en L_W^∞ tal que

$$j^* + \phi(x) \geq \inf_{A(x)} \left[c(x,a) + \int_{\Re^k} \phi[F(x,a,s)]\rho(s)ds \right], \tag{3.7}$$

y además, $j^* = \inf_{\pi \in \Pi} J(\pi, x)$ para toda $x \in X$, es decir, j^* es el costo óptimo.

Este resultado es el Teorema 2.6 en Gordienko y Hernández-Lerma (1995a). De hecho,para que se cumpla (3.7), no es necesario que ρ satisfaga la condición (f) de la Definición 3.1.

Observación 3.4. En Gordienko y Hernández-Lerma (1995a) se mostro que $j^* = \limsup_{\alpha \nearrow 1} j_\alpha$ donde j^* es el costo promedio óptimo, $j_\alpha := (1-\alpha)V_\alpha^*(z)$, $\alpha \in (0,1)$, $z \in X$ y V_α^* es la función de valor $\alpha-$ óptimo. Siguiendo los mismos argumentos usados para mostrar este hecho, obtenemos también que $j^* = \liminf_{\alpha \nearrow 1} j_\alpha$. Por lo tanto,

$$\lim_{t \to \infty} j_{\alpha_t} = j^*, \tag{3.8}$$

para cualquier sucesión $\{\alpha_t\}$ de factores de descuento tal que $\alpha_t \nearrow 1$ [ver también Gordienko (1985)]. Inclusive, (j^*, ϕ), con $\phi(x) := \liminf_{t \to \infty} \phi_{\alpha_t}(x)$, $x \in X$, es una solución de (3.7), donde $\phi_\alpha(x) := V_\alpha(x) - V_\alpha(z)$. También se demuestra que

$$\sup_{\alpha \in (0,1)} \|\phi_\alpha\|_W < \infty. \tag{3.9}$$

En la Sección 3.5 del presente capítulo, mostramos que las condiciones del Teorema 3.3

garantizan la optimalidad promedio de una política adaptada construida como límite del caso descontado. De lo contrario, para mostrar la optimalidad promedio de la política adaptada de la sección 3.4, necesitamos garantizar la existencia de una solución a la ecuación de optimalidad. Este problema no es difícil bajo el supuesto de que el costo por etapa sea acotado y/o el espacio de estados sea numerable. Pero en el contexto de nuestro trabajo (costo por etapa posiblemente no acotado y espacios generales), este es un problema no trivial al cual se le deben imponer condiciones restrictivas a las componentes del modelo de control como la Hipótesis H1' y la siguiente:

Hipótesis H5.

a) La multifunción $x \to A(x)$ es continua respecto a la métrica de Hausdorff [ver Definición A.2.4, Apéndice A].

b) Para cada $x \in X$, existe una función $g_x^c \in \mathcal{G}$ tal que para toda $a \in A(x)$ y $(x', a') \in \mathbb{K}$,

$$|c(k) - c(k')| \leq g_x^c[d(k, k')],$$

donde $k := (x, a)$, $k' := (x', a')$ y d es como en (3.1).

Si $A(x) \equiv A$ para todo $x \in X$, la Hipótesis H5 puede ser sustituida por la siguiente [ver Observación 3.2(e)]:

Hipótesis H5'. Para cada $x \in X$, existe una función $g_x^c \in \mathcal{G}$ tal que para toda $x' \in X$,

$$\sup_A |c(x, a) - c(x', a)| \leq g_x^c[d_1(x, x')].$$

Teorema 3.5 a) Bajo las Hipótesis H1', H4(a) y H5, existe una constante j^*, una función ϕ en L_W^∞ y una política estacionaria $\mathbf{f}^* \in \mathbb{F}$ tal que

$$j^* + \phi(x) = \inf_{A(x)} \left[c(x, a) + \int_{\Re^k} \phi[F(x, a, s)]\rho(s)ds \right]$$

$$= \quad c(x, f^*) + \int_{\Re^k} \phi[F(x, f^*, s)]\rho(s)ds, \quad x \in X, \tag{3.10}$$

y $j^* = \inf_{\pi \in \Pi} J(\pi, x) = J(\mathbf{f}^*, x)$ para toda $x \in X$.

b) Bajo la Hipótesis H1, para cada $\delta > 0$, existe una política $\mathbf{f} \in \mathbb{F}$ tal que

$$c(x, f) + \int_{\Re^k} \phi[F(x, f, s)]\rho(s)ds \leq j^* + \phi(x) + \delta, \quad x \in X. \tag{3.11}$$

La parte (a) es el Teorema 2.8 en Gordienko y Hernández-Lerma (1995a), mientras que la parte (b) se sigue del Lema 1.3.

Decimos que un par (j^*, ϕ) es una solución de la Ecuación de Optimalidad en Costo Promedio (EOCP) si satisface la relación (3.10).

Observación 3.6. Iterando la relación (3.11), obtenemos

$$J_n(f^*, x) + E_x^{f^*}[\phi(x_n)] \leq n(j^* + \delta) + \phi(x), \; x \in X, \; n \in \mathbb{N}, \tag{3.12}$$

donde J_n fué definido en (1.6). Dividiendo por n en ambos lados de (3.12), por el hecho de que $\phi \in L_W^\infty$, por (3.4) y haciendo $n \to \infty$, tenemos que

$$J(\mathbf{f}^*, x) \leq j^* + \delta, \; x \in X,$$

donde j^* es el costo promedio óptimo. De la Definición 1.4(b) concluimos que $\mathbf{f}^* \in \mathbb{F}$ es una política $\delta-$CP- óptima.

Para cada $n = 1, 2, ...,$ y estado inicial $x \in X$, definimos la función de valor óptimo de un problema de control de n etapas como

$$v_n(x) := \inf_{\pi \in \Pi} J_n(\pi, x) \tag{3.13}$$

donde $J_n(\pi, x)$ fué definida en (1.6). Más adelante, en el Lema 3.7, mostraremos que la función v_n, $n \in \mathbb{N}$, está bien definida y además que pertenece al espacio L_W^∞.

Ahora, sea $z \in X$ un estado fijo y arbitrario. Definimos una sucesión de funciones ϕ_n y de constantes j_n como:

$$\phi_n(x) := v_n(x) - v_n(z), \quad x \in X; \tag{3.14}$$

$$j_n := v_n(z) - v_{n-1}(z). \tag{3.15}$$

Si (j^*, ϕ) es una solución a la EOCP y si $\lim_{n\to\infty} \phi_n(x) = \phi(x)$, $x \in X$, $\lim_{n\to\infty} j_n = j^*$, decimos que el Algoritmo de Iteración de Valores converge. El poder garantizar esta convergencia será un punto clave para mostrar la optimalidad de la IVN-política adaptada.

Antes de establecer el resultado referente a la convergencia del algoritmo de iteración de valores, presentaremos algunas propiedades de las funciones v_n.

Lema 3.7. Supongamos que las Hipótesis H1 y H4(a) se cumplen. Entonces, $v_n \in L_W^\infty$, $n \in \mathbb{N}$, y si $v_0 := 0$,

$$v_n(x) = \inf_{A(x)} \left\{ c(x,a) + \int_{\Re^k} v_{n-1}[F(x,a,s)]\rho(s)ds \right\}, \quad x \in X. \tag{3.16}$$

Además, para cada $\delta > 0$, existe $f_n \in \mathbb{F}$, $n \in \mathbb{N}$, tal que

$$c(x, f_n) + \int_{\Re^k} v_{n-1}[F(x, f_n, s)]\rho(s)ds \leq v_n(x) + \delta, \quad x \in X. \tag{3.17}$$

Demostración. La relación (3.16) es la Ecuación de Programación Dinámica la cual se sigue de la Hipótesis H1 [ver Bertsekas and Shreve(1978), Cap. 8]. El hecho de que $v_n \in L_W^\infty$ es consecuencia de (3.16), Hipótesis H1(c) y H4(a) [tomando en cuenta la Observación 3.2(c)]. Finalmente, la medibilidad de las funciones v_n y la existencia de las funciones f_n satisfaciendo (3.17) se sigue del Lema 1.3 usando recurrentemente la ecuación (3.16). ■

Observación 3.8. a) Si en el Lema 3.7 sustituimos la Hipótesis H1 por H1', podemos garantizar

(ver Lema 1.5) la existencia de $f_n \in \mathbb{F}$, $n \in \mathbb{N}$, tal que

$$v_n(x) = c(x, f_n) + \int_{\Re^k} v_{n-1}[F(x, f_n, s)]\rho(s)ds, \quad x \in X.$$

b) De (3.16), es fácil mostrar que

$$j_n + \phi_n(x) = \inf_{A(x)} \left\{ c(x,a) + \int_{\Re^k} \phi_{n-1}[F(x,a,s)]\rho(s)ds \right\}, \quad x \in X. \tag{3.18}$$

Además, como $v_n \in L_W^\infty$, de (3.14) tenemos que $\phi_n \in L_W^\infty$, $n \in \mathbb{N}$. De hecho, en Gordienko y Hernández-Lerma (1995b) se mostro que bajo las Hipótesis H1' y H4(a),

$$\sup_{n \in \mathbb{N}} \|\phi_n\|_W < \infty. \tag{3.19}$$

Teorema 3.9. Si las Hipótesis H1', H4(a) y H5 se cumplen, y además f^* [ver Teorema 3.5] es tal que

$$q_{f^*}(U) > 0, \tag{3.20}$$

para cada conjunto abierto no vacío $U \subset X$, entonces el algoritmo de iteración de valores converge, es decir,

a) $\lim_{n \to \infty} |j_n - j^*| = 0$;

b) $\phi_n \to \phi$ cuando $n \to \infty$ uniformemente sobre subconjuntos compactos de X.

Este Teorema fué demostrado en Gordienko y Hernández-Lerma(1995b), Teorema 2.6.

3.3 Estimación de la Densidad

Las ideas que seguiremos para resolver el problema de la estimación de $\rho \in D_0'$ son completamente similares a las que se usaron en el caso del criterio de costo descontado.

Sea $\hat{\rho}_t := \hat{\rho}_t(s; \xi_0, \xi_1, ..., \xi_{t-1})$, $s \in \Re^k$, un estimador arbitrario de ρ tal que $\hat{\rho}_t \in L_q$, y para algún $\gamma > 0$,

$$E\,\|\rho - \hat{\rho}_t\|_q^{\frac{qp'}{2}} = \mathbf{O}(t^{-\gamma}) \quad \text{cuando} \quad t \to \infty, \tag{3.21}$$

donde $1/p + 1/p' = 1$. Nuevamente supondremos que los estimadores $\hat{\rho}_t$, $t \in \mathbb{N}$, no necesariamente son densidaes.

Definimos el conjunto de densidades $D' := D'_1 \cap D'_2$, donde

$$\begin{aligned} D'_1 &:= \{\mu : \mu \text{ es una densidad en } \Re^k, \mu \in L_q \text{ y } \mu(s) \leq \bar{\rho}\,(s) \text{ c.d.}\}; \\ D'_2 &:= \Big\{\mu : \mu \text{ es una densidad en } \Re^k, \mu \in L_q, \int W[F(x,a,s)]\mu(s)ds \\ &\qquad \leq \beta W(x) + b, \ (x,a) \in \mathbb{K}\Big\} \end{aligned} \tag{3.22}$$

Ver Observación 3.2 para la definición de las constantes β y b.

Usando argumentos completamente similares a los de la Sección 3, Cap. II, mostramos que D' es un conjunto cerrado y convexo en L_q, de tal manera que podemos garantizar la existencia de una única densidad $\rho_t \in D'$ [ver Lema 2.6] satisfaciendo

$$\|\rho_t - \hat{\rho}_t\|_q = \inf_D \|\mu - \hat{\rho}_t\|_q\,, \quad t \in \mathbb{N}. \tag{3.23}$$

Usaremos a ρ_t como estimador de la densidad ρ. Observemos que bajo las Hipótesis H4 y la relación (3.3), $\rho \in D'_0 \subset D'$.

Observación 3.10. En [Hasminskii e Ibragimov (1990)] se muestra que el estimador (2.24), en el Ejemplo 2.8, también satisface la relación (3.21) con $\gamma := rp'/2 > 0$.

Una consecuencia inmediata del Teorema 2.9 es la siguiente.

Teorema 3.11. Bajo las Hipótesis H1 o H1'y H2, tenemos que

$$E\,\|\rho_t - \rho\|^{p'} = \mathbf{O}(t^{-\gamma}) \quad \text{cuando } t \to \infty,$$

donde $\|\cdot\|$ es la pseudo norma definida en (2.25) y γ es como en la Observación 3.10.

3.4 Política adaptada IVN

Para construir la política adaptada IVN (IVN-Política), seguiremos algunas ideas aplicadas para la construcción de las políticas adaptadas en el caso descontado. Remplazaremos a la densidad desconocida ρ por sus estimadores ρ_t y analizaremos las ecuaciones de optimalidad con horizonte finito correspondientes.

Para el resto de la sección, supondremos que la función $W(\cdot)$, (además de satisfacer la Hipótesis H1(c)), es estrictamente no acotada, es decir, existe una sucesión de conjuntos compactos $K_n \subset X$, $n = 1, 2, ...$, tal que

$$\inf_{x \notin K_n} W(x) \to \infty, \quad \text{cuando} \quad n \to \infty. \tag{3.24}$$

Para cada $t \in \mathbb{N}$ fijo, sea $J_n^{(\rho_t)}(\pi, x) := E_x^{\pi,\rho_t}\left[\sum_{t=0}^{n-1} c(x_t, a_t)\right]$ el costo esperado en n etapas para el proceso (1.1) en el cual las v.a. $\xi_1, \xi_2, ...$ tienen la misma densidad ρ_t, y sea $v_n^{(\rho_t)}(x) := \inf_\pi J_n^{(\rho_t)}(\pi, x)$, $x \in X$, la correspondiente función de valor óptimo. También definimos las sucesiones $\phi_n^{(\rho_t)}(\cdot)$ y $j_n^{(\rho_t)}$ como (3.14) y (3.15) respectivamente. De hecho tenemos que bajo las Hipótesis H1(b),(c) [ver Lema 3.7],

$$v_n^{(\rho_t)}(x) = \inf_{A(x)} \left\{ c(x,a) + \int_{\Re^k} v_{n-1}^{(\rho_t)}[F(x,a,s)]\rho_t(s)ds \right\}, \quad x \in X, \ t \in \mathbb{N}, \tag{3.25}$$

donde $v_n^{(\rho_t)} \equiv 0$. De aquí [ver Observación 3.8(b)],

$$j_n^{(\rho_t)} + \phi_n^{(\rho_t)}(x) = \inf_{A(x)} \left\{ c(x,a) + \int_{\Re^k} \phi_{n-1}^{(\rho_t)}[F(x,a,s)]\rho_t(s)ds \right\}, \ x \in X, \ t \in \mathbb{N}.$$

Además, como $\rho_t \in D'$, tenemos que [ver Lema 3.7] $v_n^{(\rho_t)} \in L_W^\infty$ y por lo tanto $\phi_n^{(\rho_t)} \in L_W^\infty$.

Hipótesis H6. Para cada $u \in L_W^\infty$, $t \in \mathbb{N}$, $r \in \Re$, el conjunto

$$\left\{ (x,a) : \int_{\Re^k} u[F(x,a,s)]\rho_t(s)ds \leq r \right\}$$

es de Borel en $\mathbb{K}$.

Definición 3.12. Sea ν un número real arbitrario tal que $0 < \nu < \gamma/p'$. Denotando $B := \beta + \frac{b}{\inf_X W(x)}$, definimos la sucesión n_t de números enteros como: $n_t := [\nu \log_B t]$ si $B > 1$; $n_t := [t^\nu]$ si $B = 1$, donde $[x]$ es la parte entera de x; y $n_t := t$ si $B < 1$.

Para el resto del trabajo, la sucesión n_t permanecerá fija. Ver (3.21) y Observación 3.10 para la definición de p' y γ.

Considerando el Teorema 3.5(b) tenemos el siguiente resultado.

Teorema 3.13. Bajo las Hipótesis H1(b), (c) y H6, para cada $t \in \mathbb{N}$, $\delta_{n_t} > 0$, existe una política estacionaria $\mathbf{f}_{n_t} \in \mathbb{F}$ tal que

$$c(x, f_{n_t}) + \int_{\Re^k} \phi_{n_t-1}^{(\rho_t)}[F(x, f_{n_t}, s)]\rho_t(s)ds \leq j_{n_t}^{(\rho_t)} + \phi_{n_t}^{(\rho_t)}(x) + \delta_{n_t},\ x \in X, \tag{3.26}$$

o,

$$c(x, f_{n_t}) + \int_{\Re^k} v_{n_t-1}^{(\rho_t)}[F(x, f_{n_t}, s)]\rho_t(s)ds \leq v_{n_t}^{(\rho_t)}(x) + \delta_{n_t},\ x \in X. \tag{3.27}$$

Considerando este resultado, definimos la IVN- política adaptada de la siguiente manera:

Definition 3.14. Sea $\{\delta_n\}$ una sucesión arbitraria de números positivos y $\{f_{n_t}\}$ una sucesión de funciones satisfaciendo (3.26) o (3.27) para cada cada $t \in \mathbb{N}$. Definimos la IVN- política adaptada $\bar{\pi} = \{\bar{\pi}_t\}$ como:

$$\bar{\pi}_t(h_t) = \bar{\pi}_t(h_t; \rho_t) := f_{n_t}(x_t),\ \ h_t \in \mathbb{H}_t,\ t = 1, 2, ...,$$

donde $\bar{\pi}_0(x)$ es cualquier control fijo.

Observación 3.15. Otra forma de obtener una política del tipo IVN, como en la Definición 3.14, es garantizando la existencia de minimizadores f_{n_t} satisfaciendo la igualdad en (3.26) o

(3.27). Esto significa que debemos imponer condiciones mas restrictivas como H1'(b), además de extender las hipótesis referentes a la continuidad de ρ [ver H1'(c) y condición (f) en la Definición 3.1] al estimador ρ_t, $t \in \mathbb{N}$. Esto último podria resultar dificil o poco práctico de veificar en problemas específicos.

Suponiendo que $\{\delta_n\}$ converge, sea $\delta := \lim_{n\to\infty} \delta_n$. El resultado que garantiza la optimalidad de la IVN- política adaptada es el siguiente:

Teorema 3.16. Supongamos que las Hipótesis H1', H4, H5, H6 y la condición (3.20) se cumplen. Entonces la IVN- política adaptada $\bar{\pi}$ is $\delta-$CP- óptima [ver Def. 1.4(b)].

En particular, si $\delta = 0$ entonces la política $\bar{\pi}$ es CP- óptima.

Observación 3.17. Podemos tener un resultado análogo al Teorema 3.16 referente a la política contruida en el contexto que se mencionó en la Observación 3.15. Este resultado podría mostrar directamente la CP- optimalidad de dicha política pero bajo las condiciones restrictivas sobre el estimador ρ_t que se mencionaron.

La demostración del Teorema 3.16 está basada en los siguientes lemas.

Lemma 3.18. [Mandl (1974)]. Una política $\pi \in \Pi$ es CP- óptima si para todo $x \in X$,

$$E_x^{\pi}\left[\Phi(x_t, a_t)\right] \to 0 \quad \text{cuando } t \to \infty,$$

donde $a_t = \pi_t(h_t)$ y

$$\Phi(x,a) := c(x,a) + \int_{\Re^k} \phi[F(x,a,s)]\rho(s)ds - \phi(x) - j^*, \quad (x,a) \in \mathbb{K}$$

es llamada función de discrepancia en costo promedio, la cual es no negativa debido al Teorema 3.5.

De la demostración del Teorema 3.18 [Mandl (1974), Hernández-Lerma (1989)], es fácil ver que una política $\pi \in \Pi$ es δ-CP- óptima si para toda $x \in X$,

$$\lim_{t\to\infty} E_x^{\pi}\left[\Phi(x_t, a_t)\right] \leq \delta. \tag{3.28}$$

Observación 3.19. En la demostración de los siguientes resultados usaremos repetidamente la siguiente desigualdad:

$$(c_1 + c_2 + \cdots + c_n)^{p'} \leq c_{p'}\left(c_1^{p'} + c_2^{p'} + \cdots + c_n^{p'}\right), \tag{3.29}$$

donde $c_i > 0$, $i = 1, 2, ..., n$ y $c_{p'} > 0$ es una constante dependiendo de p'. Además estaremos usando las desigualdades (2.33) y (2.34), con las constantes β y b definidas en la Observación 3.2, las cuales las expresaremos de nuevo para una fácil referencia:

$$u(x) \leq \|u\|_W W(x) \tag{3.30}$$

y

$$\int_{\Re^k} u[F(x,a,s)]\mu(s)ds \leq \|u\|_W \left[\beta W(x) + b\right] \tag{3.31}$$

para toda $u \in L_W^{\infty}$, $\mu \in D$, $x \in X$ y $a \in A(x)$.

Lemma 3.20. Bajo las Hipótesis H1 o H1' y H4, para cada $x \in X$ y $\pi \in \Pi$, cuando $t \to \infty$,

$$\text{a) } E_x^{\pi}\left\|v_{n_t}^{(\rho_t)} - v_{n_t}\right\|_W^{p'} \to 0; \quad \text{b) } E_x^{\pi}\left\|\phi_{n_t}^{(\rho_t)} - \phi_{n_t}\right\|_W^{p'} \to 0.$$

Demostración. a) De (3.16) y (3.25), sumando y restando el término $\int_{\Re^k} v_{n_t-1}[F(x,a,s)]\rho_t(s)ds$ tenemos,

$$\left|v_{n_t}^{(\rho_t)}(x) - v_{n_t}(x)\right| \leq \sup_{a\in A(x)}\left\{\int_{\Re^k}\left|v_{n_t-1}^{(\rho_t)}[F(x,a,s)] - v_{n_t-1}[F(x,a,s)]\right|\rho_t(s)ds\right\}$$

$$+ \sup_{a \in A(x)} \left\{ \int_{\Re^k} v_{n_t-1}[F(x,a,s)] \, |\rho_t(s) - \rho(s)| \, ds \right\}, \quad x \in X, \; t \in \mathbb{N}.$$

Usando el hecho $\rho_t \in D$, $t \in \mathbb{N}$, de (2.25) y (3.31) obtenemos

$$\left| v_{n_t}^{(\rho_t)}(x) - v_{n_t}(x) \right| \leq \left\| v_{n_t-1}^{(\rho_t)} - v_{n_t-1} \right\|_W [\beta W(x) + b] + \|v_{n_t-1}\|_W \, \|\rho_t - \rho\| \, W(x), \quad x \in X.$$

Como $W(\cdot) \geq 1$, dividiendo por $W(\cdot)$ en ambos lados de la desigualdad y usando la definición de $\|\cdot\|_W$ y de la constante B [ver Def. 3.12],

$$\left\| v_{n_t}^{(\rho_t)} - v_{n_t} \right\|_W \leq B \left\| v_{n_t-1}^{(\rho_t)} - v_{n_t-1} \right\|_W + \|v_{n_t-1}\|_W \, \|\rho_t - \rho\|, \; t \in \mathbb{N}.$$

Iterando esta desigualdad y usando el hecho de que $v_0 \equiv v_0^{(\rho_t)} \equiv 0$, $t \in \mathbb{N}$, obtenemos, para $B \neq 1$

$$\left\| v_{n_t}^{(\rho_t)} - v_{n_t} \right\|_W \leq \|\rho_t - \rho\| \sum_{k=0}^{n_t-1} B^k \left\| v_{n_t-(k+1)} \right\|_W \leq C' \, \|\rho_t - \rho\| \, \frac{1 - B^{n_t}}{1 - B}, \tag{3.32}$$

y para $B = 1$

$$\left\| v_{n_t}^{(\rho_t)} - v_{n_t} \right\|_W \leq C' \, \|\rho_t - \rho\| \, n_t, \tag{3.33}$$

donde la constante C' es de (3.19).

De esta manera, para $B > 1$, de la definición de n_t y (3.32) se sigue que

$$\left\| v_{n_t}^{(\rho_t)} - v_{n_t} \right\|_W^{p'} = \|\rho_t - \rho\|^{p'} \, \mathbf{O}(t^{\nu p'}) \;\; \text{cuando} \;\; t \to \infty. \tag{3.34}$$

Tomando esperanza en ambos lados de (3.34), por el Teorema 3.11 y por el hecho de que $E_x^\pi \|\rho_t - \rho\| = E \|\rho_t - \rho\|$ (debido a que ρ_t no depende de π y x), obtenemos

$$E_x^\pi \left\| v_{n_t}^{(\rho_t)} - v_{n_t} \right\|_W^{p'} = \mathbf{O}\left(t^{\nu p' - \gamma}\right) \;\; \text{cuando} \;\; t \to \infty.$$

El resultado deseado se sigue del hecho de que $\nu p' < \gamma$ [ver Def. 3.10].

Finalmente, cuando $B \leq 1$, de nuevo usando la definición de n_t, de (3.32) y (3.33), tenemos

que $E_x^\pi \left\| v_{n_t}^{(\rho_t)} - v_{n_t} \right\|_W^{p'}$ es de orden $\mathbf{O}(t^{-s})$ para algún $s > 0$. Esto muestra la parte (a).

b) Esta parte se sigue facilmente de (a) y del hecho de que

$$\left\| \phi_{n_t}^{(\rho_t)} - \phi_{n_t} \right\|_W \leq 2 \left\| v_{n_t}^{(\rho_t)} - v_{n_t} \right\|_W .$$

■

Lema 3.21. Bajo las Hipótesis H1', H4, H5 y la condición (3.20), para cada $x \in X$ y $\pi \in \Pi$, cuando $t \to \infty$,

$$\text{a) } E_x^\pi \left| j_{n_t}^{(\rho_t)} - j^* \right|^{p'} \to 0; \qquad \text{b) } E_x^\pi \left| \phi_{n_t}(x_t) - \phi(x_t) \right| \to 0.$$

Demostración. a) Sea $z \in X$ el estado fijo en (3.14) y (3.15). Para cada $t \in \mathbb{N}$,

$$\begin{aligned} \left| j_{n_t}^{(\rho_t)} - j^* \right| &\leq \left| j_{n_t}^{(\rho_t)} - j_{n_t} \right| + \left| j_{n_t} - j^* \right| \\ &\leq \left| v_{n_t}^{(\rho_t)}(z) - v_{n_t}(z) \right| + \left| v_{n_t-1}^{(\rho_t)}(z) - v_{n_t-1}(z) \right| + \left| j_{n_t} - j^* \right| . \end{aligned}$$

De esta manera, la relación (3.29), Teorema 3.9 y Lema 3.20 demuestran la parte (a).

b) Sea $\{K_n\}$ la sucesión de conjuntos compactos en (3.24) y $\delta > 0$ fijo y arbitrario. Para cualquier $n \in \mathbb{N}$, tenemos que

$$\begin{aligned} E_x^\pi \left| \phi(x_t) - \phi_{n_t}(x_t) \right| &= E_x^\pi \left\{ \left| \phi(x_t) - \phi_{n_t}(x_t) \right| 1_{K_n}(x_t) \right\} \\ + E_x^\pi \left\{ \left| \phi(x_t) - \phi_{n_t}(x_t) \right| 1_{\bar{K}_n}(x_t) \right\} &= : I_{1,n}(t) + I_{2,n}(t). \end{aligned}$$

Además, por (3.19)

$$\left| \phi(x) - \phi_{n_t}(x) \right| \leq C'' W(x), \; t \in \mathbb{N}, \; x \in X,$$

para alguna constante $C'' < \infty$.

De aquí, por (3.4) y (3.24)

$$
\begin{aligned}
I_{2,n}(t) &\leq C'' E_x^{\pi}\left\{(W(x_t))^p\,(W(x_t))^{1-p}\,1_{\bar{K}_n}(x_t)\right\} \\
&\leq C'' \sup_{x\notin K_n}(W(x))^{1-p}\sup_{t\in\mathbb{N}} E_x^{\pi}\left\{(W(x_t))^p\right\} \leq \delta/2, \qquad (3.35)
\end{aligned}
$$

para toda $t \in \mathbb{N}$ y n suficientemente grande.

Ahora, fijando algún n tal que (3.35) se cumple, por el Teorema 3.9(b) obtenemos,

$$
I_{1,n}(t) \leq \sup_{x\in K_n} |\phi(x) - \phi_{n_t}(x)| \leq \delta/2,
$$

para toda t suficientemente grande. Esto prueba la parte (b). ■

Demostración del Teorema 3.16.

Para cada $t, n \in \mathbb{N}$ definimos la función $\Phi_n^{(\rho_t)}$ como:

$$
\Phi_n^{(\rho_t)}(x,a) := c(x,a) + \int_{\Re^k} \phi_{n-1}^{(\rho_t)}[F(x,a,s)]\rho_t(s)ds - \phi_n^{(\rho_t)}(x) - j_n^{(\rho_t)}, \quad (x,a) \in \mathbb{K}.
$$

Por la definición de la política $\bar{\pi}$ y de las funciones f_{n_t} (3.26), (3.27), tenemos que $\Phi_{n_t}^{(\rho_t)}(\cdot, \bar{\pi}_t(\cdot)) \leq \delta_{n_t}$, $t \in \mathbb{N}$. De aquí,

$$
E_x^{\bar{\pi}}\left(\Phi(x_t,a_t)\right) \leq E_x^{\bar{\pi}}\left|\Phi(x_t,a_t) - \Phi_{n_t}^{(\rho_t)}(x_t,a_t)\right| + \delta_{n_t}, \qquad (3.36)
$$

donde la función Φ fué definida en el Lema 3.18.

En vista del Lema 3.18, es suficiente mostrar la convergencia a cero de la esperanza en el lado derecho de (3.36).

Sea $\{(x_t,a_t)\}$ una sucesión de pares estado-control correspondientes a la aplicación de la política adaptada $\bar{\pi}$. Usando la definición de Φ y $\Phi_{n_t}^{(\rho_t)}$ obtenemos

$$
\left|\Phi(x_t,a_t) - \Phi_{n_t}^{(\rho_t)}(x_t,a_t)\right| \leq |\phi(x_t) - \phi_{n_t}(x_t)| + \left|\phi_{n_t}(x_t) - \phi_{n_t}^{(\rho_t)}(x_t)\right| + \left|j^* - j_{n_t}^{(\rho_t)}\right|
$$

$$
\begin{aligned}
&+\left|\int_{\Re^k}\phi[F(x_t,a_t,s)]\rho(s)ds-\int_{\Re^k}\phi_{n_t-1}[F(x_t,a_t,s)]\rho(s)ds\right| \\
&+\left|\int_{\Re^k}\phi_{n_t-1}[F(x_t,a_t,s)]\rho(s)ds-\int_{\Re^k}\phi_{n_t-1}^{(\rho_t)}[F(x_t,a_t,s)]\rho_t(s)ds\right| \\
= &: I_1(t)+I_2(t)+I_3(t)+I_4(t)+I_5(t). \qquad (3.37)
\end{aligned}
$$

Por el Lema 3.21 $E_x^{\bar{\pi}}(I_1(t)) \to 0$ y $E_x^{\bar{\pi}}(I_3(t)) \to 0$ cuando $t \to \infty$. Por otro lado, por (3.30) y la desigualdad de Hölder

$$
\begin{aligned}
E_x^{\bar{\pi}}(I_2(t)) &\leq E_x^{\bar{\pi}}\left(\left\|\phi_{n_t}-\phi_{n_t}^{(\rho_t)}\right\|_W W(x_t)\right) \\
&\leq \left(E_x^{\bar{\pi}}\left[\left\|\phi_{n_t}-\phi_{n_t}^{(\rho_t)}\right\|_W^{p'}\right]\right)^{1/p'}\left(E_x^{\bar{\pi}}\left[W^p(x_t)\right]\right)^{1/p} \to 0 \text{ as } t\to\infty, \qquad (3.38)
\end{aligned}
$$

debido al Lema 3.20 y (3.4). Ahora, por Lema 3.21(b),

$$
\begin{aligned}
E_x^{\bar{\pi}}(I_4(t)) &\leq E_x^{\bar{\pi}}\left\{\int_{\Re^k}\left|\phi[F(x_t,a_t,s)]-\phi_{n_t-1}[F(x_t,a_t,s)]\right|\rho(s)ds\right\} \\
&= E_x^{\bar{\pi}}\left\{E_x^{\bar{\pi}}\left\{\left|\phi[F(x_t,a_t,\xi_t)]-\phi_{n_t-1}[F(x_t,a_t,\xi_t)]\right|\right\} \mid x_t,a_t\right\} \\
&= E_x^{\bar{\pi}}\left\{\left|\phi(x_{t+1})-\phi_{n_t-1}(x_{t+1})\right|\right\} \to 0 \text{ as } t\to\infty.
\end{aligned}
$$

Finalmente, para el término $I_5(t)$, sumando y restando $\int_{\Re^k}\phi_{n_t-1}^{(\rho_t)}[F(x,a,s)]\rho_t(s)ds$, por (3.31) y (2.25)

$$
\begin{aligned}
I_5(t) &\leq \int_{\Re^k}\left|\phi_{n_t-1}[F(x_t,a_t,s)]-\phi_{n_t-1}^{(\rho_t)}[F(x_t,a_t,s)]\right|\rho(s)ds \\
+\int_{\Re^k}\phi_{n_t-1}^{(\rho_t)}[F(x_t,a_t,s)]\left|\rho(s)-\rho_t(s)\right|ds &\leq \left\|\phi_{n_t-1}-\phi_{n_t-1}^{(\rho_t)}\right\|_W\left[\beta W(x_t)+b\right] \\
&+\left\|\phi_{n_t-1}^{(\rho_t)}\right\|_W W(x_t)\left\|\rho-\rho_t\right\| = : J_1(t)+J_2(t). \qquad (3.39)
\end{aligned}
$$

De (3.38) y el Lema 3.20, es fácil ver que $\lim_{t\to\infty} E_x^{\bar{\pi}}(J_1(t)) = 0$.

Ahora, de (3.25) y (3.31), para cada $x \in X$, $n \in \mathbb{N}$ tenemos

$$v_n^{(\rho_t)}(x) \leq W(x) + \sup_{a \in A(x)} \int_{\Re^k} v_{n-1}^{(\rho_t)}[F(x,a,s)]\rho_t(s)ds \leq W(x) + \left\|v_{n-1}^{(\rho_t)}\right\|_W [\beta W(x) + b].$$

De aquí,

$$\left\|v_n^{(\rho_t)}\right\|_W \leq 1 + B\left\|v_{n-1}^{(\rho_t)}\right\|_W, \quad n \in \mathbb{N}.$$

Iterando esta desigualdad obtenemos

$$\left\|v_n^{(\rho_t)}\right\|_W \leq \sum_{k=0}^{n-1} B^k = \frac{1-B^n}{1-B}, \quad \text{if} \quad B \neq 1,$$

and

$$\left\|v_n^{(\rho_t)}\right\|_W \leq n, \quad \text{if} \quad B = 1.$$

De la definición de n_t, es fácil ver que existe una constante C_1' tal que

$$\left\|v_{n_t}^{(\rho_t)}\right\|_W \leq C_1' t^\nu, \quad t \in \mathbb{N},$$

y usando el hecho de que $\left\|\phi_{n_t}^{(\rho_t)}\right\|_W \leq 2\left\|v_{n_t}^{(\rho_t)}\right\|_W$, existe una constante C_1 tal que

$$\left\|\phi_{n_t}^{(\rho_t)}\right\|_W \leq C_1' t^\nu, \quad t \in \mathbb{N}. \tag{3.40}$$

De esta manera, aplicando la desigualdad de Hölder a $E_x^{\bar{\pi}}(J_2(t))$ y usando (3.40) obtenemos

$$E_x^{\bar{\pi}}(J_2(t)) \leq C_1 t^\nu \left(E_x^{\bar{\pi}}[W^p(x_t)]\right)^{1/p} \left(E_x^{\bar{\pi}}\left[\|\rho - \rho_t\|^{p'}\right]\right)^{1/p'} \to 0 \text{ cuando } t \to \infty,$$

debido a (3.4), al Teorema 3.11 y al hecho de que $\nu < \gamma/p'$.

Con esto concluimos la demostración del Teorema 3.16. ■

3.5 Política adaptada como límite del caso descontado

En esta sección construimos una política adaptada apoyándonos fuertemente en el análisis de los resultados del caso descontado. Originalmente esta política fué propuesta por Gordienko (1985)

y estudiada después por Hernández-Lerma y Cavazos-Cadena (1990), los cuales le llamaron política adaptada de Gordienko (G-política).

La teoría desarrollada en esta sección se hace en el contexto de las Hipótesis H1, H4 y H6, las cuales son menos restrictivas que las usadas en la sección anterior. Por lo tanto [ver Observación 3.2], los resultados del Capítulo II, referentes al criterio de costo descontado, del Teorema 3.3, junto con la Observación 3.4, son válidos.

La relación entre el criterio de costo promedio y el descontado se tiene observando que, de la definición de j_α y ϕ_α [ver Observación 3.4], la ecuación (2.14) y la desigualdad (2.15) son equivalentes, respectivamente, a

$$j_\alpha + \phi_\alpha(x) = \inf_{A(x)} \left[c(x,a) + \alpha \int_{\Re^k} \phi_\alpha[F(x,a,s)]\rho(s)ds \right], \quad x \in X, \ \alpha \in (0,1), \tag{3.41}$$

y

$$c(x,f) + \alpha \int_{\Re^k} \phi_\alpha[F(x,f,s)]\rho(s)ds \leq j_\alpha + \phi_\alpha(x) + \delta, \quad x \in X, \ \alpha \in (0,1). \tag{3.42}$$

Sea ν un número real arbitrario tal que $0 < \nu < \gamma/(3p')$ [ver (3.21) y Observación 3.10 para la definición de γ]. Fijamos una sucesión arbitraria no decreciente de factores de descuento $\{\alpha_t\}$, tal que $1 - \alpha_t = \mathbf{O}(t^{-\nu})$ y

$$\lim_{n\to\infty} \frac{\kappa(n)}{n} = 0, \tag{3.43}$$

donde $\kappa(n)$ es el número de cambios de valor de $\{\alpha_t\}$ sobre $[0,n]$.

Para cada $t \in \mathbb{N}$ y $\mu \in D'$, definimos el operador $T_{\mu,\alpha_t} \equiv T_\mu : L_W^\infty \to L_W^\infty$ como

$$T_\mu u(x) := \inf_{A(x)} \left\{ c(x,a) + \alpha_t \int_{\Re^k} u[F(x,a,s)]\mu(s)ds \right\}, \quad x \in X, \ u \in L_W^\infty. \tag{3.44}$$

Sea $\rho_t \in D'$ el estimador definido en la Sección 3. Bajo las Hipótesis H1(b), (c), H4 y H6, los resultados del Lema 2.4(c) y de los Teoremas 2.5 y 2.12 son válidos si usamos α_t en lugar

de α. De aquí, tenemos el siguiente resultado.

Teorema 3.22.

a) Supongamos que la Hipótesis H1(b) se cumple y ρ satisface la condición (3.3). Entonces, para cada $t \in \mathbb{N}$, $T_\rho V_{\alpha_t} = V_{\alpha_t}$, $T_{\rho_t} V_{\alpha_t}^{(\rho_t)} = V_{\alpha_t}^{(\rho_t)}$ y

$$V_{\alpha_t}(x) \leq \frac{C}{1-\alpha_t} W(x), \qquad V_{\alpha_t}^{(\rho_t)}(x) \leq \frac{C}{1-\alpha_t} W(x),\ x \in X. \tag{3.45}$$

b) Bajo las Hipótesis H1(b), (c) y H6, para cada $t \in \mathbb{N}$, $\delta_t > 0$, existe una política $\hat{\mathbf{f}}_t \in \mathbb{F}$ tal que

$$c(x, \hat{f}_t) + \alpha_t \int_{\Re^k} V_{\alpha_t}^{(\rho_t)}[F(x, \hat{f}_t, s)]\rho_t(s)ds \leq V_{\alpha_t}^{(\rho_t)}(x) + \delta_t, \quad x \in X, \tag{3.46}$$

o [ver (3.42)]

$$c(x, \hat{f}_t) + \alpha_t \int_{\Re^k} \phi_{\alpha_t}^{(\rho_t)}[F(x, \hat{f}_t, s)]\rho(s)ds \leq j_{\alpha_t}^{(\rho_t)} + \phi_{\alpha_t}^{(\rho_t)}(x) + \delta_t, \quad x \in X, \tag{3.47}$$

donde $\phi_{\alpha_t}^{(\rho_t)}(x) := V_{\alpha_t}^{(\rho_t)}(x) - V_{\alpha_t}^{(\rho_t)}(z)$, $z \in X$.

De (3.41) y definiendo $j_{\alpha_t}^{(\rho_t)} := (1-\alpha_t)V_{\alpha_t}^{(\rho_t)}(z)$, $z \in X$, tenemos que

$$j_{\alpha_t}^{(\rho_t)} + \phi_{\alpha_t}^{(\rho_t)}(x) = \inf_{A(x)} \left[c(x,a) + \alpha_t \int_{\Re^k} \phi_{\alpha_t}^{(\rho_t)}[F(x,a,s)]\rho_t(s)ds \right],\ x \in X,\ t \in \mathbb{N}. \tag{3.48}$$

Definición 3.23. Sea $\{\delta_t\}$ una sucesión arbitraria de números positivos y $\left\{\hat{f}_t\right\}$ una sucesión de funciones satisfaciendo (3.46) o (3.47), para cada $t \in \mathbb{N}$. Definimos la G- política adaptada $\hat{\pi} = \{\hat{\pi}_t\}$ como $\hat{\pi}_t(h_t) = \hat{\pi}_t(h_t; \rho_t) := \hat{f}_t(x_t)$, $t \in \mathbb{N}$, donde $\hat{\pi}_0(x)$ es cualquier control fijo.

Suponiendo que $\delta := \lim_{t\to\infty} \delta_t < \infty$, el objetivo es mostrar que la G- política es $\delta-$ CP-óptima, como lo establece el siguiente teorema:

Teorema 3.24. Bajo las Hipótesis H1, H4 y H6, la G- política es $\delta-$ CP- óptima, i.e., para cada $x \in X$, $J(\hat{\pi}, x) \leq j^* + \delta$, donde j^* es el costo promedio óptimo como en el Teorema 3.3.

Observación 3.25. Para cada $t \in \mathbb{N}$, sea $\theta_t := (1+\alpha_t)/2 \in (\alpha_t, 1)$, $W_t(x) := W(x) + d_t$, $x \in X$, donde $d_t := b\,(\theta_t/\alpha_t - 1)^{-1}$, y $L^\infty_{W_t}$ el espacio de funciones medibles $u : X \to \Re$ con la norma

$$\|u\|_{W_t} := \sup_{x \in X} \frac{|u(x)|}{W_t(x)} < \infty, \quad t \in \mathbb{N}.$$

De (2.36) y (2.37) tenemos que para cada $t \in \mathbb{N}$ [ver demostración del Lema 2.16],

$$\|u\|_{W_t} \leq \|u\|_W \leq l_t \|u\|_{W_t}, \quad t \in \mathbb{N},\ u \in L^\infty_W, \tag{3.49}$$

donde $l_t := 1 + 2b/\left[(1-\alpha_t)\inf_X W(x)\right]$. Además, para cada $t \in \mathbb{N}$, el operador T_μ definido en (3.44) es una contracción con respecto a la norma $\|\cdot\|_{W_t}$ con módulo θ_t, i.e.,

$$\|T_\mu v - T_\mu u\|_{W_t} \leq \theta_t \|v - u\|_{W_t}, \quad v, u \in L^\infty_W,\ t \in \mathbb{N}. \tag{3.50}$$

Lema 3.26. Bajo las Hipótesis H1 y H4, para cada $x \in X$ y $\pi \in \Pi$, cuando $t \to \infty$

$$\text{a) } E^\pi_x \left\| \phi_{\alpha_t} - \phi^{(\rho_t)}_{\alpha_t} \right\|^{p'}_W \to 0, \qquad \text{b) } E^\pi_x \left[\left\| \phi_{\alpha_t} - \phi^{(\rho_t)}_{\alpha_t} \right\|_W W(x_t) \right] \to 0.$$

Demostración. a) De (3.49) y observando que $\left\| \phi_{\alpha_t} - \phi^{(\rho_t)}_{\alpha_t} \right\|_W \leq 2 \left\| V_{\alpha_t} - V^{(\rho_t)}_{\alpha_t} \right\|_W$, es suficiente mostrar

$$\lim_{t \to \infty} E^\pi_x \left\| V_{\alpha_t} - V^{(\rho_t)}_{\alpha_t} \right\|^{p'}_W = 0, \quad x \in X,\ \pi \in \Pi. \tag{3.51}$$

De (3.50) y el Teorema 3.22(a), es fácil mostrar que [ver (2.41)],

$$l_t \left\| V_{\alpha_t} - V^{(\rho_t)}_{\alpha_t} \right\|_{W_t} \leq \frac{l_t}{1-\theta_t} \left\| T_\rho V_{\alpha_t} - T_{\rho_t} V_{\alpha_t} \right\|_{W_t}, \quad t \in \mathbb{N}. \tag{3.52}$$

Por otro lado, por (3.45) y usando el hecho de que $[W_t(\cdot)]^{-1} < [W(\cdot)]^{-1}$, $t \in \mathbb{N}$, obtenemos

$$\begin{aligned}\left\|T_\rho V_{\alpha_t} - T_{\rho_t} V_{\alpha_t}\right\|_{W_t} &\leq \alpha_t \sup_X [W_t(x)]^{-1} \sup_{A(x)} \int_{\Re^k} V_{\alpha_t}[F(x,a,s)]\,|\rho(s) - \rho_t(s)|\,ds \\ &\leq \frac{C\alpha_t}{1-\alpha_t} \sup_X [W(x)]^{-1} \sup_{A(x)} \int_{\Re^k} W[F(x,a,s)]\,|\rho(s) - \rho_t(s)|\,ds \\ &\leq \frac{C}{1-\alpha_t} \|\rho - \rho_t\|\,. \end{aligned} \tag{3.53}$$

Ahora, observemos que [ver definición de α_t y θ_t],

$$\frac{1}{(1-\theta_t)(1-\alpha_t)} = \mathbf{O}(t^{2\nu}) \text{ and } \frac{1}{(1-\theta_t)(1-\alpha_t)^2} = \mathbf{O}(t^{3\nu}). \tag{3.54}$$

Combinando (3.52), (3.53), (3.54), y usando la definición de l_t obtenemos

$$\begin{aligned} l_t^{p'} \left\|V_{\alpha_t} - V_{\alpha_t}^{(\rho_t)}\right\|_{W_t}^{p'} &\leq C^{p'} \left[\frac{1}{(1-\theta_t)(1-\alpha_t)} + \frac{2b}{(1-\theta_t)(1-\alpha_t)^2 \inf_X W(x)}\right]^{p'} \|\rho - \rho_t\|^{p'} \\ &= C^{p'} \mathbf{O}(t^{3p'\nu}) \|\rho - \rho_t\|^{p'}\,. \end{aligned} \tag{3.55}$$

Finalmente, tomando esperanza E_x^π en ambos lados de (3.55), por el Teorema 3.12 y usando el hecho de que $3\nu p' < \gamma$ [ver definición de α_t], mostramos la parte a).

b) Denotando $\bar{C}$:= $(E_x^\pi [W^p(x_t)])^{1/p} < \infty$ [ver (3.4)], aplicando la desigualdad de Holder y por la parte a) tenemos

$$E_x^\pi \left\|\phi_{\alpha_t} - \phi_{\alpha_t}^{(\rho_t)}\right\|_W W(x_t) \leq \bar{C} \left(E_x^\pi \left[\left\|\phi_{\alpha_t} - \phi_{\alpha_t}^{(\rho_t)}\right\|_W^{p'}\right]\right)^{1/p'} \to 0 \text{ cuando } t \to \infty. \tag{3.56}$$

esto demuestra el Lema 3.26. ■

Demostración del Teorema 3.24.

Sea $\{k_t\} := \{(x_t, a_t)\}$ una sucesión de pares estado-control correspondiente a la aplicación de la política $\hat{\pi}$. Definimos

$$
\begin{aligned}
L_t \quad &: \quad = c(k_t) + \alpha_t \int_{\Re^k} \phi_{\alpha_t}[F(k_t, s)]\rho(s)ds - j_{\alpha_t} - \phi_{\alpha_t}(x_t) \qquad (3.57)\\
&= \quad c(k_t) + \alpha_t E_x^{\hat{\pi}}\left[\phi_{\alpha_t}(x_{t+1}) \mid k_t\right] - j_{\alpha_t} - \phi_{\alpha_t}(x_t).
\end{aligned}
$$

Ahora, ordenando términos y tomando esperanza $E_x^{\hat{\pi}}$ obtenemos

$$
\begin{aligned}
E_x^{\hat{\pi}}\left[c(k_t) - j_{\alpha_t}\right] &= E_x^{\hat{\pi}}\left[\phi_{\alpha_t}(x_t)\right] - \alpha_t E_x^{\hat{\pi}}\left[\phi_{\alpha_t}(x_{t+1})\right] + E_x^{\hat{\pi}}\left[L_t\right]\\
&= E_x^{\hat{\pi}}\left[\phi_{\alpha_t}(x_t) - \alpha_t\phi_{\alpha_t}(x_{t+1})\right] + E_x^{\hat{\pi}}\left[L_t\right].
\end{aligned}
$$

De aquí, para $n \geq k \geq 1$

$$
n^{-1}E_x^{\hat{\pi}}\left[\sum_{t=k}^{n} c(k_t) - j_{\alpha_t}\right] = n^{-1}E_x^{\hat{\pi}}\left[\sum_{t=k}^{n}\left(\phi_{\alpha_t}(x_t) - \alpha_t\phi_{\alpha_t}(x_{t+1})\right)\right] + n^{-1}E_x^{\hat{\pi}}\left[\sum_{t=k}^{n} L_t\right]. \qquad (3.58)
$$

Considerando en primer sumando del lado derecho de (3.58), tenemos para $n \geq k \geq 1$,

$$
\begin{aligned}
n^{-1}E_x^{\hat{\pi}}\left[\sum_{t=k}^{n}\left(\phi_{\alpha_t}(x_t) - \alpha_t\phi_{\alpha_t}(x_{t+1})\right)\right] &= n^{-1}E_x^{\hat{\pi}}\left[\sum_{t=k}^{n}\left(\phi_{\alpha_t}(x_t) - \alpha_t\phi_{\alpha_t}(x_t)\right)\right]\\
&\quad + n^{-1}E_x^{\hat{\pi}}\left[\sum_{t=k}^{n}\alpha_t\left(\phi_{\alpha_t}(x_t) - \phi_{\alpha_t}(x_{t+1})\right)\right]. \qquad (3.59)
\end{aligned}
$$

De (3.9), (3.4) y (3.30) tenemos que $E_x^{\hat{\pi}}\left[\phi_\alpha(x_t)\right] < C'$, $\alpha \in (0,1)$, para alguna constante $C' < \infty$. De esta manera, como $\{\alpha_t\}$ es no decreciente obtenemos para $n \geq k \geq 1$,

$$
n^{-1}E_x^{\hat{\pi}}\left[\sum_{t=k}^{n}\left(\phi_{\alpha_t}(x_t) - \alpha_t\phi_{\alpha_t}(x_t)\right)\right] \leq (1 - \alpha_k)C'. \qquad (3.60)
$$

Por otro lado, para cada n, sean α_1^*, α_2^*, ..., α_s^* los distintos valores de α_t para $t \leq n$. Observe que $s = \kappa(n)$ [ver la condición (3.43)]. Entonces tenemos que

$$E_x^{\hat{\pi}}\left[\sum_{t=k}^{n}\alpha_t\left(\phi_{\alpha_t}(x_t)-\phi_{\alpha_t}(x_{t+1})\right)\right]\leq 2C'\alpha_1^*+\ldots+2C'\alpha_s^*\leq 2C'\kappa(n). \tag{3.61}$$

Por lo tanto, de (3.59), (3.60) y (3.61), obtenemos

$$n^{-1}E_x^{\hat{\pi}}\left[\sum_{t=k}^{n}\left(\phi_{\alpha_t}(x_t)-\alpha_t\phi_{\alpha_t}(x_{t+1})\right)\right]\leq(1-\alpha_k)C'+2C'\kappa(n)n^{-1},\ x\in X. \tag{3.62}$$

Ahora, de (3.57) y (3.41) tenemos

$$L_t=c(k_t)+\alpha_t\int_{\Re^k}\phi_{\alpha_t}[F(k_t,s)]\rho(s)ds-\inf_{A(x_t)}\left[c(x_t,a)+\alpha_t\int_{\Re^k}\phi_{\alpha_t}[F(x_t,a,s)]\rho(s)ds\right]$$

$$\leq\left|\alpha_t\int_{\Re^k}\phi_{\alpha_t}[F(k_t,s)]\rho(s)ds-\alpha_t\int_{\Re^k}\phi_{\alpha_t}^{(\rho_t)}[F(k_t,s)]\rho(s)ds\right|$$

$$+\left|\alpha_t\int_{\Re^k}\phi_{\alpha_t}^{(\rho_t)}[F(k_t,s)]\rho(s)ds-\alpha_t\int_{\Re^k}\phi_{\alpha_t}^{(\rho_t)}[F(k_t,s)]\rho_t(s)ds\right|$$

$$+\left|c(k_t)+\alpha_t\int_{\Re^k}\phi_{\alpha_t}^{(\rho_t)}[F(k_t,s)]\rho_t(s)ds-\inf_{A(x_t)}\left[c(x_t,a)+\alpha_t\int_{\Re^k}\phi_{\alpha_t}[F(x_t,a,s)]\rho(s)ds\right]\right|$$

$$=:|I_1(t)|+|I_2(t)|+|I_3(t)|.$$

Usando el hecho de que (3.30) and (3.3),

$$|I_1(t)|\leq\alpha_t\int_{\Re^k}\left|\phi_{\alpha_t}[F(k_t,s)]-\phi_{\alpha_t}^{(\rho_t)}[F(k_t,s)]\right|\rho(s)ds\leq\alpha_t\left\|\phi_{\alpha_t}-\phi_{\alpha_t}^{(\rho_t)}\right\|_W[\beta W(x_t)+b]. \tag{3.63}$$

Tomando esperanza $E_x^{\hat{\pi}}$ en ambos lados de (3.63) y usando el Lema 3.26 obtenemos

$$E_x^{\hat{\pi}} |I_1(t)| \to 0, \text{ as } t \to \infty. \tag{3.64}$$

Para mostrar que $E_x^{\hat{\pi}} |I_2(t)| \to 0$, primero tenemos que de la definición de α_t y (3.45),

$$\left\| \phi_{\alpha_t}^{(\rho_t)} \right\|_W \leq 2 \left\| V_{\alpha_t}^{(\rho_t)} \right\|_W \leq \frac{2C}{1-\alpha_t} = \mathrm{O}(t^{\nu}).$$

Ahora, de la definición (2.25),

$$|I_2(t)| \leq \alpha_t \int_{\Re^k} \phi_{\alpha_t}^{(\rho_t)}[F(k_t,s)] \, |\rho(s) - \rho_t(s)| \, ds \leq \alpha_t W(x_t) \left\| \phi_{\alpha_t}^{(\rho_t)} \right\|_W \|\rho - \rho_t\|. \tag{3.65}$$

Tomando esperanza y aplicando la desigualdad de Holder en (3.65) obtenemos

$$E_x^{\hat{\pi}} |I_2(t)| \leq \bar{C} \left([\mathrm{O}(t^{\nu})]^{p'} \, E_x^{\hat{\pi}} \|\rho - \rho_t\|^{p'} \right)^{1/p'} = \left[\mathrm{O}(t^{\nu p' - \gamma}) \right]^{1/p'} \to 0 \text{ as } t \to \infty, \tag{3.66}$$

debido al hecho de que $\nu < \gamma/p'$ [ver definición de α_t], donde $\bar{C}$ es como (3.56) con π en lugar de $\hat{\pi}$.

Para el término $|I_3(t)|$, de la definición de la política $\hat{\pi}$ y combinando (3.47) y (3.48),

$$\begin{aligned}
|I_3(t)| \;\leq\; & \left| c(k_t) + \alpha_t \int_{\Re^k} \phi_{\alpha_t}^{(\rho_t)}[F(k_t,s)] \rho_t(s) ds - \inf_{A(x_t)} \left\{ c(x_t,a) + \alpha_t \int_{\Re^k} \phi_{\alpha_t}^{(\rho_t)}[F(x_t,a,s)] \rho_t(s) ds \right\} \right| \\
& + \left| \inf_{A(x_t)} \left\{ c(x_t,a) + \alpha_t \int_{\Re^k} \phi_{\alpha_t}^{(\rho_t)}[F(x_t,a,s)] \rho_t(s) ds \right\} \right. \\
& \left. - \inf_{A(x_t)} \left\{ c(x_t,a) + \alpha_t \int_{\Re^k} \phi_{\alpha_t}[F(x_t,a,s)] \rho(s) ds \right\} \right| \\
\leq \; & \delta_t + \alpha_t \sup_{A(x_t)} \left| \int_{\Re^k} \phi_{\alpha_t}^{(\rho_t)}[F(x_t,a,s)] \rho_t(s) ds - \int_{\Re^k} \phi_{\alpha_t}[F(x_t,a,s)] \rho(s) ds \right|.
\end{aligned}$$

De aquí, usando la definición (2.25),

$$\begin{aligned} |I_3(t)| &\leq \delta_t + \alpha_t \sup_{A(x_t)} \int_{\Re^k} \phi_{\alpha_t}^{(\rho_t)}[F(x_t,a,s)]\,|\rho(s)-\rho_t(s)|\,ds \\ &\quad +\alpha_t \sup_{A(x_t)} \int_{\Re^k} \left|\phi_{\alpha_t}^{(\rho_t)}[F(x_t,a,s)] - \phi_{\alpha_t}[F(x_t,a,s)]\right| \rho(s)ds \\ &\leq \delta_t + \alpha_t W(x_t) \left\|\phi_{\alpha_t}^{(\rho_t)}\right\|_W \|\rho-\rho_t\| + \alpha_t \left\|\phi_{\alpha_t} - \phi_{\alpha_t}^{(\rho_t)}\right\|_W [\beta W(x)+b]. \end{aligned}$$

De aquí, de (3.63), (3.64), (3.65) y (3.66), llegamos a que $E_x^{\hat{\pi}}\,|I_3(t)| \to \delta$, cuando $t \to \infty$. por lo tanto

$$E_x^{\hat{\pi}}\,[L_t] \to \delta, \text{ as } t \to \infty. \tag{3.67}$$

Finalmente, de (3.58), (3.62) y (3.67), tenemos que para cualquier $k \geq 1$ y $n \to \infty$,

$$n^{-1} E_x^{\hat{\pi}} \left[\sum_{t=k}^{n} c(k_t) - j_{\alpha_t}\right] = (1-\alpha_k)C' + \mathbf{o}(1) + \delta,\ x \in X.$$

De esta manera, de (3.8), el hecho de que $\lim_{t\to\infty} \alpha_t = 1$, y (1.5),

$$J(\hat{\pi}, x) \leq j^* + \delta,\ x \in X.$$

Con esto concluimos la demostración del teorema. ■

3.6 Ejemplo

Nuevamente retomaremos el ejemplo del Capítulo I y verificaremos que las hipótesis usadas en este capítulo (H1', H4, H5, H6) junto con la condición (3.20) se cumplen.

Supongamos que la Hipótesis E1 de la Sección 2.5 del capítulo anterior se satisface, y además que el costo por etapa es una función medible no negativa satisfaciendo la Hipótesis H5', tal

que para cada $x \in X$, $c(x,\cdot)$ es s.c.i. en A y

$$\sup_A c(x,a) \leq W(x),$$

donde $W(x) := \bar{b}e^{\lambda x}$, $x \in [0,\infty)$, $\lambda > 0$ es como (2.53) y $\bar{b}$ es una constante arbitraria. Claramente, W es una función estrictamente no acotada y las Hipótesis H1 y H1'(a), (b) se cumplen.

Por otro lado, de manera completamente similar al desarrollo que se hizo para verificar las Hipótesis H2(b), (c) y H3, mostramos que las Hipótesis H4(b), (c) y H6 se cumplen, tomando $\bar{\rho}$ como en (2.54).

Para verificar la Hipótesis H1'(c), sea ρ_a, $a \in A$, la densidad de $a\eta - \chi$. Es fácil mostrar que para cada $y \in \Re$,

$$\rho_a(y) = \frac{1}{a}\int_0^\infty \rho_\eta\left(\frac{y+s}{a}\right)\rho_\chi(s)ds = \int_{y/a}^\infty \rho_\eta(s)\,\rho_\chi(as-y)ds. \tag{3.68}$$

Debido a que $\{\eta_t\}$ y $\{\chi_t\}$ son i.i.d., por la continuidad de ρ_η y ρ_χ, el hecho de que ambas estén acotadas y por el Teorema de la Convergencia Acotada, tenemos que el mapeo $a \to \rho_a$ es continuo en A.

Ahora, sea $u \in L_W^\infty$, entonces

$$\begin{aligned}\int_{\Re^2} u[F(x,a,s)]\rho(s)ds &= \int_{\Re^2} u[(x+as_1-s_2)^+]\rho(s)ds = \int_{\Re^2} u[(x+y)^+]\rho_a(y)dy \\ &= u(0)\int_{-\infty}^{-x}\rho_a(y)dy + \int_0^\infty u(y)\rho_a(y-x)dy. \end{aligned}\tag{3.69}$$

Por el Teorema de Scheffé (ver, por ejemplo, Hernández-Lerma (1989)), (3.69) define una función continua en $a \in A$ para cada $x \in X$, lo cual muestra que la Hipótesis H1'(c) se cumple.

Ahora verificaremos la Hipótesis H4(a). Para esto es suficiente mostrar que la densidad $\rho := \rho_\eta\,\rho_\chi$ satisface las condiciones (c)-(f), en la definición del conjunto D_0'.

La condición (c) claramente se cumple con $\bar{\rho}$ como en (2.54). Para verificar la condición (d), sea $a_t = \theta$, $t \in \mathbb{N}$ en (1.11) y denotemos por $\left\{x_t^\theta\right\}$ el correspondiente proceso de Markov, es decir, $x_{t+1}^\theta = (x_t^\theta + \theta\eta_t - \chi_t)^+ := (x_t^\theta + \varsigma)^+$, $t = 0, 1, ...$, donde ς es como en la Hipótesis E1(b). De aquí, bajo la Hipótesis E1(b) tenemos que $\left\{x_t^\theta\right\}$ es Harris recurrente positivo [ver, por ejemplo, Nummelin (1984), ejemplo 5.2]. Esto implica que $E[\tau^\theta] < \infty$ [ver, p. 75 en Nummelin (1984)], donde τ^θ denota el tiempo del primer retorno a $x = 0$ dado el estado inicial $x_0 = 0$.

Ahora, sea $f \in \mathbb{F}$ una política estacionaria arbitraria y $\left\{x_t^f\right\}$ el proceso de Markov correspondiente definido por (1.11), donde $a_t = f(x_t)$, $t = 0, 1, ...$ Sea τ^f el tiempo del primer retorno de $\left\{x_t^f\right\}$ a $x = 0$ dado x_0^f. Como $A = (0, \theta]$, $\theta \in A$, tenemos que $f(x) \leq \theta$, para toda $x \in X$, lo cual implica que $x_t^f \leq x_t^\theta$, $t = 0, 1, ...$ De aquí, $E[\tau^f] \leq E[\tau^\theta] < \infty$, y por el Corolario 5.3 en Nummelin (1984) [ver también Ejemplo 2.3(d) en Nummelin (1984)], el proceso $\left\{x_t^f\right\}$ es Harris recurrente positivo. Como $f \in \mathbb{F}$ se tomó arbitraria, la condición (d) se cumple.

Para verificar la condición (e), nos basaremos en las relaciones (2.55) y (2.56). Definiendo

$$\psi_f(x) := Prob[x + f(x) - \xi_0 \leq 0], \quad f \in \mathbb{F}, \, x \in X,$$

y $m(\cdot) := p_0$, donde p_0 es la medida de Dirac concentrada en $x = 0$, tenemos que de (2.55) y (2.56), para $B \in \mathbb{B}[0, \infty)$

$$\begin{aligned} Q_\rho(B \quad | \quad x, f) &= \int_{\Re^2} 1_B[F(x, f, s)]\rho(s)ds = Prob[x + f(x)\eta - \chi \in B] \\ &\geq Prob[x + f(x)\eta - \chi \leq 0]m(B) = \psi_f(x)m(B), \end{aligned}$$

lo cual muestra que la condición e(i) se cumple.

Por otro lado, sea $b_0 := \int_X W^p(y)m(dy) = W^p(0)$ y $R = (0, \infty) \times (0, \infty)$. Similarmente a (2.57), la condición e(ii) se sigue del siguiente desarrollo:

$$\begin{aligned} \int_{\Re^2} W^p[F(x, a, s)]\rho(s)ds &= b_0\psi_f(x) + \left(\bar{b}\right)^p \int_R e^{\lambda p(x + as_1 - s_2)}\rho(s)ds \\ &\leq b_0\psi_f(x) + H(p\lambda)W^p(x) = b_0\psi_f(x) + \beta_0 W^p(x), \quad x \in X, \, a \in A, \end{aligned}$$

donde β_0 es como (2.53).

Ahora, de la definición de m,

$$\begin{aligned}\inf_{f\in\mathbb{F}}\int_X \psi_f(x)m(dx) &= \inf_{f\in\mathbb{F}}\int_X Prob[x+f(x)\eta-\chi\leq 0]m(dx)\\ &= \inf_{f\in\mathbb{F}} Prob[f(0)\eta-\chi\leq 0]\geq Prob[\theta\eta-\chi\leq 0]=Prob[\varsigma\leq 0]>0,\end{aligned}$$

donde la última desigualdad es una consecuencia del hecho de que $E[\varsigma] < 0$ [ver Hipótesis E1(b)]. Por lo tanto, la condición e) se cumple.

Por último, para mostrar la condición (f) nos basaremos en la Observación 3.2(e), para lo cual necesitamos imponer una hipótesis adicional:

Hipótesis E2.

a) Para cada $a \in A$ existe un número $\delta > 0$ tal que

$$\int_{-\infty}^{\infty} e^{\lambda y}\rho_{a,\delta}(y)dy < \infty,$$

donde

$$\rho_{a,\delta}(y) := \sup\{\rho_{a'}(y) : a' \in A \text{ y } |a'-a| \leq \delta\},\ y \in \Re;$$

b) existen funciones $g \in \mathcal{G}$ y $\psi_a(z)$, $z \in \Re$, $a \in A$ tal que

$$\sup_A \int_{-\infty}^{\infty} e^{\lambda z}\psi_a(z)dz < \infty,$$

y

$$|\rho_a(z+y)-\rho_a(z)| \leq g(|y|)\psi_a(z)$$

para toda z, $y \in \Re$ y $a \in A$.

Observemos que de (3.68) y el hecho de que ρ_χ es acotada, tenemos que

$$0 \leq \rho_a(y) \leq M,\ y \in \Re,\ a \in A.$$

donde M es una cota de ρ_χ. Además,

$$|F_a(x) - F_a(x')| \leq M\,|x - x'|\,,$$

donde F_a, $a \in A$, es la función de distribución de $a\eta - \chi$. Observemos también que

$$Q_\rho(B \mid x, a) = \int_{\Re} 1_B(y)\rho_a(y - x)dy.$$

De aquí, bajo la Hipótesis E2,

$$\begin{aligned}
\|Q_\rho(\cdot \mid x, a) - Q_\rho(\cdot \mid x', a)\|_W &= \int_X W(y)\,|Q_\rho(dy \mid x, a) - Q_\rho(dy \mid x', a)| \\
&\leq W(0)\,|Prob[x + a\eta - \chi \leq 0] - Prob[x' + a\eta - \chi \leq 0]| \\
&\quad + \bar{b}\int_0^\infty e^{\lambda y}\,|\rho_a(y - x) - \rho_a(y - x')|\,dy \\
&= W(0)\,|F_a(x) - F_a(x')| + \bar{b}\int_{-x}^\infty e^{\lambda(y+x)}\,|\rho_a(y) - \rho_a(y + x - x')|\,dy \\
&\leq W(0)\,|x - x'| + \bar{b}\,e^{\lambda x} g\left(|x - x'|\right)\int_{-\infty}^\infty e^{\lambda y}\psi_a(y)dy \\
&\leq W(0)\,|x - x'| + C'\,\bar{b}\,e^{\lambda x} g\left(|x - x'|\right) \\
&\leq C_x\left[|x - x'| + g\left(|x - x'|\right)\right],
\end{aligned}$$

donde $C_x := \max\left\{W(0), C'\,\bar{b}\,e^{\lambda x}\right\}$ y la constante C' es de la Hipótesis E2(b). Tomando $g_x^Q(y) = C_x\left[|y| + g\left(|y|\right)\right]$, tenemos que la condición (f) se cumple y por lo tanto también la Hipótesis H4(a).

Finalmente, observemos que de (3.5), la condición (3.20) se satisface si $Q_\rho(U/x, a) > 0$ para

cada conjunto abierto $U \subset X$, $x \in X$, $a \in A$. Para mostrar esto, sea U un conjunto abierto arbitrario. Entonces

$$\begin{aligned} Q_\rho(U/x,a) &= \int_{\Re^2} 1_U[F(x,a,s)]\rho(s)ds = \int_{\Re^2} 1_U[(x+y)^+]\rho_a(y)dy \\ &= 1_U(0)\int_{-\infty}^{-x} \rho_a(y)dy + \int_0^{\infty} 1_U(y)\rho_a(y-x)dy > 0,\ x \in X,\ a \in A. \end{aligned}$$

Esto último se sigue de la definición de ρ_a y el hecho de que ρ_η y ρ_χ son estrictamente positivas.

Capítulo 4

Conclusiones y Problemas Abiertos

4.1 Conclusiones

En este trabajo se estudió el problema de control adaptado para procesos de Markov a tiempo discreto con espacios de estado y control de Borel, considerando costo por etapa posiblemente no acotado. El proceso se desarrolla de acuerdo a la ecuación en diferencia

$$x_{t+1} = F(x_t, a_t, \xi_t), \quad t = 0, 1, 2, ...,$$

donde F es una función conocida; x_t, a_t y ξ_t representan el estado, el control y el ruido al tiempo t, respectivamente.

Para especificar la clase de modelos de control, se utilizaron dos tipos de hipótesis básicas, H1 y H1', las cuales garantizan la existencia de δ- minimizadores y minimizadores medibles repectivamente. Respecto a estas hipótesis tenemos que H1'$\Longrightarrow$H1. Además, se usó otra hipótesis que impone condiciones a la función de densidad de las v.a. ξ_t. Dichas condiciones dependen del criterio de optimalidad que se esté analizando.

El motivo por el que se estableció la Hipótesis H1, es el de no tener que imponer condiciones restrictivas, como compacidad y continuidad, en el modelo de control. Esto nos permite ampliar la clase de ejemplos en donde es aplicable la teoría desarrollada en el trabajo.

Suponiendo que las v.a. ξ_t son completamente observables, propusimos una técnica de

estimación estadística de la densidad ρ, la cual nos permitió hacer la construcción de políticas adaptadas bajo los índices de costo descontado y costo promedio.

En el caso descontado se utilizó el criterio de optimalidad asintótica ya que las características de este índice no permiten mostrar, por lo general, que una política adaptada sea óptima. Además, se trabajó en el contexto de la Hipótesis H1 lo cual motivó introducir el concepto de política δ- asintóticamente óptima descontada. Tomando en cuenta lo anterior, se demostró que la PEC- poítica y la IVN- política, propuestas originalmente por Mandl (1974) y Hernández-Lerma y Marcus (1985), respectivamente (considerando costos acotados), son δ- asintóticamente óptimas descontadas, donde δ es el límite de una sucesión de números positivos $\{\delta_t\}$. En el caso que $\delta = 0$, ambas políticas resultan se asintóticamnete óptimas descontadas.

En el Caso promedio sí es posible mostrar que las políticas adaptadas son óptimas. Su estudio se puede hacer por medio de la Ecuación de Optimalidad y Desigualdad de Optimalidad, las cuales se obtienen en el contexto de las Hipótesis H1' y H1, respectivamente (Teoremas 3.3 y 3.5). Imponiendo hipótesis de recurrencia y continuidad en ρ, construimos dos tipos de políticas adaptadas $\delta-$ óptimas.

La primera es una política del tipo IVN (IVN-política) definida recursivamnete por medio de las funciones de valor óptimo en n_t etapas donde n_t es una sucesión de enteros positivos de orden $\mathbf{O}(\nu \log t)$, para algún $\nu > 0$. Para mostrar la optimalidad de esta política nos basamos en los resultados de Gordienko y Hernández-Lerma (1995a,b) referentes a la existencia de soluciones a la ecuación de optimalidad y a la convergencia del algoritmo de iteración de valores en el caso promedio.

La segunda, la cual llamamos G-política, se obtiene analizando el criterio de costo promedio como límite del caso descontado (enfoque del factor de descuento desvaneciente). Esta política fué propuesta originalmente por Gordienko (1985) y retomada después por Hernández-Lerma y Cavazos-Cadena (1990), y en ambos artículos se considero costo por etapa acotado. Para mostrar la optimalidad de la G-política fué suficiente contar con una solución de la desigualdad de optimalidad lo cual implica que no necesitamos imponer condiciones restrictivas en el modelo de control.

Aunque lo anterior puede ser una ventaja de la G-política, esta tiene la desventaja de que

para construirla necesitamos resolver primero el problema del costo descontado en el sentido de que debemos encontrar una solución a la ecuación de optimalidad $\alpha-$ descontada.

4.2 Problemas abiertos

Enunciaremos algunos problemas que consideramos importantes y no resueltos.

1.- La parte difícil para aplicar las políticas adaptadas construidas en el trabajo, es hacer la proyección del estimador $\hat{\rho}_t$ [ver (2.24)] sobre el conjunto D y D' definidos en la Sección 2.3 y 3.3 para el caso descontado y promedio, respectivamente. Un problema interesante es desarrollar algoritmos para obtener dicha proyección.

2.- Sea $J_n(\pi, x)$ el costo total en n etapas cuando se usa la política π y el estado inicial es x, y sea $v_n(x)$ su correspondiente función de valor óptimo. Definimos la desviación promedio en n etapas de la política π como:

$$\Delta_n(\pi, x) := \frac{1}{n} |v_n(x) - J_n(\pi, x)|$$

El problema que proponemos es encontrar cotas superiores de Δ_n para las políticas adaptadas construidas en el trabajo bajo el índice de costo promedio. Por ejemplo, demostrar que

$$\Delta_n(\pi, x) = \mathbf{O}(n^{-r}),$$

para algún $r > 0$, donde π es la NVI- política o la G- política adaptada.

3.- Un punto clave para mostrar la optimalidad de la NVI- política en el caso promedio fué imponer hipótesis de recurrencia y continuidad en la densidad ρ, y de esta manera garantizar la existencia de una solución a la ecuación de optimalidad y la convergencia del algoritmo de iteración de valores. Existe otro tipo de hipótesis que garantizan que se cumplen estos dos puntos. Dichas hipótesis se basan en imponer condiciones a la función de valor óptimo del caso descontado, y aplicar el enfoque del factor de descuento desvaneciente, ver por ejemplo, Montes-de-Oca y Hernández-Lerma (1995, 1996), Montes-de-Oca, Minjárez-Sosa y Hernández-Lerma (1994), Montes-de-Oca (1994), entre otros. El problema de control adaptado en el contexto de

este tipo de hipótesis no se ha estudiado, por lo tanto este sería otro problema abierto.

4.- Un último problema que proponemos es debilitar las hipótesis usadas en el trabajo, especialmente las que se usan para mostrar la optimalidad promedio de las políticas adaptadas construidas en el Capítulo 3.

Apéndice A

Notación y Definiciones

A.1 Notación y terminología

A.1.0.- $\mathbb{B}(X)$: $\sigma-$ álgebra de Borel de un espacio topológico X, es decir, la $\sigma-$ álgebra más pequeña de subconjuntos de X que contiene los conjuntos abiertos.

A.1.1.- Un espacio de Borel es un subconjunto de un espacio métrico, completo y separable.

A.1.2.- Ejemplos de espacios de Borel:

- $\Re^n$ con la topología usual;
- un subconjunto numerable con la topología discreta;
- un espacio métrico compacto;
- el producto, finito o numerable, de una sucesión de espacios de Borel.

A.2 Definiciones importantes

Sean X y Y espacios de Borel.

A.2.0. Sea v una función en X tomando valores en los reales extendidos. Decimos que v es semi-continua inferiormente (s.c.i.) si el conjunto $\{x \in X : v(x) \leq r\}$ es cerrado en X para cada $r \in \Re$ y es semi-continua superiormente (s.c.s.) si $\{x \in X : v(x) \geq r\}$ es cerrado para cada $r \in \Re$.

A.2.1.- Una multifunción o correspondencia Γ de X a Y es una función definida en X cuyo valor $\Gamma(x)$, $x \in X$, es un subconjunto no vacío de Y.

A.2.2.- Sean A y B subconjuntos de X. Definimos las pseudo métricas

$$\begin{aligned} \delta_l(A,B) &= \inf\{\lambda \geq 0 : A \subset \lambda + B\}; \\ \delta_u(A,B) &= \inf\{\lambda \geq 0 : B \subset \lambda + A\}. \end{aligned}$$

Como es usual, $\inf \phi = +\infty$.

A.2.3.- Dados A y B subconjuntos de X distintos del vacío, la pseudo métrica de Hausdorff $\delta(A,B)$ es definida como:

$$\delta(A,B) = \max\{\delta_l(A,B), \delta_u(A,B)\}.$$

A.2.4.- Sea Γ una multifunción de X a Y. Decimos que es continua respecto a la métrica de Hausdorff si, cuando la vemos como función de X a $\mathcal{P}_0(Y)$, esta es continua de X a $(\mathcal{P}_0(Y), \delta)$, donde $\mathcal{P}_0(Y)$ es la familia de subconjuntos de Y distintos del vacío.

A.2.5.- Un kernel estocástico sobre X dado Y es una función $Q(\cdot \mid \cdot)$ tal que, $Q(\cdot \mid y)$ es una medida de probabilidad en X para cada $y \in Y$ y $Q(B \mid \cdot)$ es una función medible sobre Y para cada $B \in \mathbb{B}(X)$.

Sea $\{x_n\}$ una cadena de Markov con espacio de estados X y probabilidad de transición $Q(\cdot \mid \cdot)$.

A.2.6.- Se dice que la cadena $\{x_n\}$ es Harris- recurrente, si existe una medida $\sigma-$ finita λ en $(X,\mathbb{B}(X))$ tal que

$$P(x_n \in B \text{ para algún } n \mid x_0 = x) = 1,$$

para cada $x \in X$ y $B \in \mathbb{B}(X)$ con $\lambda(B) > 0$.

A.2.7.- Se dice que la cadena $\{x_n\}$ es Harris- recurrente positiva, si es Harris- recurrente y Q tiene una medida de probabilidad invariante, es decir, una medida de probabilidad q tal que

$$q(B) = \int_X Q(B \mid x) q(dx), \quad B \in \mathbb{B}(X).$$

Referencias

Agrawal, R. (1991) Minimizing the learning loss in adaptive of Markov chains under the weak accessibility condition. J. Appl. Prob., 28, 779-790.

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

Arapostathis, A., Borkar, V., Fernández-Gaucherand, E., Ghosh, M.K. and Marcus, S.I. (1993) Controlled Markov processes with an average cost criterion: a survey. SIAM J. Control Optim., 31, 282-344.

Bellman, R. (1961) Adaptive Control Processes: A guided Tour. Princeton University Press, Princeton, N.J.

Bertsekas, D.P. and Shreve, S.E. (1978) Stochastic Optimal Control: The discrete Time Case. Academic Press, New York.

Bertsekas, D.P. (1987) Dynamic Programming: Deterministic and Stochastic Models. Prentice-Hall, Englewood Cliffs, N.J.

Blackwell, D. (1962) Discrete dynamic programming. Ann. Math. Statist., 33, 714-726.

Bosq, D. (1996) Nonparametric Statistics for Stochastic Processes. Lecture Notes in Statistics No. 110, springer-Verlag, New York.

Caines, P.E. (1988) Linear Stochastic Systems. John Wiley & Sons, Inc., USA.

Cavazos-Cadena, R. (1990) Nonparametric adaptive control of discounted stochastic system with compact state space. J. Optim. Theory Appl., 65, 191-207.

Devroye, L. (1987) A Course in density Estimation, Birkhäuser, Boston.

Devroye, L. and Györfi, L. (1985) Nonparametric Density Estimation: The L_1 View. Wiley, New York.

Di Massi, G.B. and Stettner, L. (1995) Bayesian ergodic adaptive control of discrete time Markov processes, Stichastic and Stochastics Reports, Vol. 54, pp. 301-316.

Dynkin, E.B. and Yushkevich, A.A. (1979) Controlled Markov Processes. Springer-Verlag, New York.

El-Fattah, Y.M. (1981) Gradient approach for recursive estimation and control in finite Markov chains. Adv. Appl. Probab., 13, 778-803.

Fernández-Gaucherand, E., Arapostathis, A. and Marcus, S.I. (1992) A methodology for the adaptive control of Markov chains under partial state information. Proc. of the 1992 Conf. on Information Sci. and Systems, Princeton, New Jersey, 773-775.

Fernández-Gaucherand, E., Arapostathis, A. and Marcus, S.I. (1993) Analysis of an adaptive control scheme for a partially observed controlled Markov chain. IEEE Trans. Autom. Control, 38, 987-993.

Gordienko, E.I. (1985) Adaptive strategies for certain classes of controlled Markov processes. Theory Probab. Appl., 29, 504-518.

Gordienko, E.I. (1985a) Controlled Markov sequences with slowly varying characteristics II. Adaptive optimal strategies. Soviet J. Comput. Systems Sci., 23, 87-93.

Gordienko, E.I. (1985b) Random search in the problem of adaptive control of Markov processes with discrete time. Soviet J. Comput. Systems Sci., 22, No.4, 56-64.

Gordienko, E. and Hernández-Lerma, O. (1995a) Average cost Markov control processes with weighted norms: existence of canonical policies. Applicationes Math., 23, 199-218.

Gordienko, E. and Hernández-Lerma, O. (1995b) Average cost Markov control processes with weighted norms: value iteration. Applicationes Math., 23, 219-237.

Gordienko, E, Montes-de-Oca, R. and Minjárez-Sosa, J.A. (1997) Approximation of average cost optimal policies for general Markov decision processes with unbounded costs. Math. Methods of Oper. Res., 45 Iss. 245-263.

Gordienko, E. and Minjárez-Sosa, J.A. (1996) Adaptive control for discrete-time Markov processes with unbounded costs: discounted criterion. (Aceptado para su publicación en Kybernetyka)

Gordienko, E. and Minjárez-Sosa, J.A. (1997) Adaptive control for discrete-time Markov processes with unbounded costs: average criterion. (Aceptado para su publicación en ZOR, Math. Meth. of Oper. Res.).

Hasminskii, R. and Ibragimov, I. (1990) On density estimation in the view of Kolmogorov's ideas in approximation theory. Ann. of Statist., 18, 999-1010.

Hernández-Lerma, O. (1987) Approximation and adaptive control of Markov peocesses: average rewards criterion. Kybernetika (Prague) 23, 265-288.

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

Hernández-Lerma, O. (1991) On integrated square errors of recursive nonparametric estimates of nonstationary Markov processes. Probab. and Math. Statist., 12, 22-33.

Hernández-Lerma, O. (1994) Infinite-horizon Markov control processes with undiscounted cost criteria: from average to overtaking optimality. Reporte Interno 165. Departamento de Matemáticas, CINVESTAV-IPN, A.P. 14-740.07000, México, D.F., México.

Hernández-Lerma, O. and Cavazos-Cadena, R. (1990) Density estimation and adaptive control of Markov processes: average and discounted criteria. Acta Appl. Math., 20, 285-307.

Hernández-Lerma, O. and Lasserre, J.B. (1995) Discrete-Time Markov Control Processes. Springer-Verlag, New York.

Hernández- Lerma O. and Marcus, S.I. (1985) Adaptive control of discounted Markov decision chains. J. Optim. Theory Appl., 46, 227-235.

Hernández- Lerma O. and Marcus, S.I. (1987) Adaptive policies for discrete-time stochastic control system with unknown disturbance distribution. Syst. Control Lett., 9, 307-315.

Hinderer, K. (1970) Foundations of Non-Stationary Dynamic Programming with Discrete Time Parameter. Lecture Notes Oper. Res. 33, Springer-Verlag, New York.

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

Kumar, P.R. (1985) A survey of some results in stochastic adaptive control. SIAM J. Control Optim., 23, 329-280.

Kumar, P.R. and Varaiya, P. (1986) Stochastic Systems: Estimation, Identification and Adaptive Control. Prentice-Hall, Englewood Cliffs.

Kurano, M. (1972) Discrete-time markovian decision processes with an unknown parameter - average return criterion. J. Oper. Res. Soc. Japan, 15, 67-76

Kurano, M. (1987) Learning algorithms for Markov decision processes. J. Appl. Probab., 24, 270-276.

Lippman, S.A. (1975) On dynamic programming with unbounded rewards. Manag. Sci., 21, 1225-1233.

Lyubchik, L.M. and Poznyak, A.S. (1974) Learning aotomata in stochastic plant control problems. Autom. Remote Control, 35, 778-789.

Mandl, P. (1974) Estimation and control in Markov chains. Adv. Appl. Probab., 6, 40-60.

Minjárez-Sosa, J.A. (1998) Nonparametric adaptive control for discrete-time Markov processes with unbounded costs under average criterion. (Sometido para su publicación en Systems and Control Letters).

Montes-de-Oca, R. (1994) The average optimality equation for Markov control processes on Borel spaces. Systems and Control Letters, 22, 351-357.

Montes-de-Oca, R., Minjárez-Sosa, J. and Hernández-Lerma, O. (1994) Conditions for average optimality in Markov control processes on Borel spaces. Boletín de la Sociedad Matemática Mexicana 39, 39-50.

Montes-de-Oca, R. and Hernández-Lerma,O. (1995) Conditions for average optimality in Markov control processes with unbounded costs and controls. J. Math. Systems, Estimation and Control, Vol. 5, 4, 459-477.

Montes-de-Oca, R. and Hernández-Lerma, O (1996) Value iteration in average cost Markov control processes on Borel spaces. Acta Applicandae Mathematicae, 42, 203-221.

Rieder, U. (1975) Bayesian dynamic programming. Adv. Appl. Probab., 7, 330-348.

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

Schäl, M. (1979) On dynamic programming and statistical decision theory. Ann. Statist., 7, 432-445.

Schäl, M. (1987) Estimation and control in discounted stochastic dynamic programming. Stochastics, 20, 51-71.

Silverman, B.W. (1986) Density Estimation for Statistics and Data Analysis. Monographs on Statistics and Appl. Probab., No. 26, Chapman &Hall, London.

Stettner, L. (1993) On nearly self-optimizing strategies for a discrete-time uniformly ergodic adaptive model. J. Appl. Math. Optim., 27, 161-177.

Stettner, L. (1995) Ergodic control of Markov process with mixed observation structure. Dissertationes Math., 341, 1-36.

Van **Hee, K.M. (1978)** Bayesian Control of Markov Chains. Mathematical Centre Tract 95, Matematishch Centrum, Amsterdan.

Van Nunen, J.A.E.E. and Wessels, J. (1978) A note on dynamic programming with unbounded rewards. Manag. Sci., 24, 576-580.

Weber, R.R. y Stidham, S. Jr. (1987) Optimal control of service rates in network of queues. Adv. Appl. Proba., 19, 202-218.

Yakowitz, S. (1985) Nonparametric density estimation, prediction and regression for Markov sequences. J. Amer. Statist. Assoc., 80, 215-221.