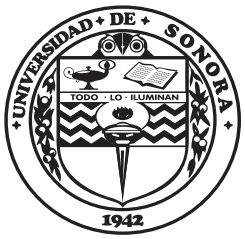

"El saber de mis hijos hará mi grandeza"

UNIVERSIDAD DE SONORA

División de Ciencias Exactas y Naturales

Programa de Posgrado en Matemáticas

Estimación, control y estabilidad en modelos semi-markovianos

T E S I S

Que para obtener el grado académico de:

Doctor en Ciencias
(Matemáticas)

Presenta:

Luz del Carmen Rosas Rosas

Directores de Tesis:
Dr. Jesús Adolfo Minjárez Sosa
Dr. Fernando Luque Vásquez

Hermosillo, Sonora, México, Noviembre 26 de 2010

Universidad de Sonora

Repositorio Institucional UNISON

SINODALES

Dr. Onésimo Hernández Lerma
Centro de Investigación y de Estudios Avanzados del I. P. N., Ciudad de México.

Dr. Héctor Jasso Fuentes
Centro de Investigación y de Estudios Avanzados del I. P. N., Ciudad de México.

Dr. Jesús Adolfo Minjárez Sosa
Universidad de Sonora, Hermosillo, México.

Dr. Oscar Vega Amaya
Universidad de Sonora, Hermosillo, México.

Dr. Fernando Luque Vázquez
Universidad de Sonora, Hermosillo, México.

Índice general

Dedicatoria

A mis Padres

Ramón Rosas Olivas † (1923-2006)
María Consuelo Rosas Borbón

A mi Esposo

Israel Segundo Caballero

A mis Hijos

Jovan Israel
Edmar Daniel
Alan David

A mi Nieto

Kevin Daniel

A mis Hermanos

Ramón, Luis Enrique, Consuelo, José Antonio

A mis Sobrinos

Ramón Omar, Daniela Elizabeth, Corinna Estefanía, Ana Ruth, Juan Daniel, Kimberly Danitza

Y muy especialmente a mi inolvidable Sobrino

Ramón Enrique † (1993-2010)

Así como también a mi Prima-Hermana

Rosario † (1961-2009)

Agradecimientos

Sinodales

Dr. Jesús Adolfo Minjárez-Sosa
Dr. Fernando Luque-Vásquez
Dr. Onésimo Hernández-Lerma
Dr. Óscar Vega-Amaya
Dr. Héctor Jasso-Fuentes

Agradezco a todos y cada uno de ellos por su valioso apoyo respecto a la revisión y la elaboración de esta tesis doctoral.

Asimismo, agradezco a:

Consejo Nacional de Ciencia y Tecnología (CONACYT)
Mis Profesores del Doctorado
Mis Compañeros de estudio y Amigos
Coordinación de Posgrado
Departamento de Matemáticas
División de Ciencias Exactas y Naturales
Universidad de Sonora (UNISON)
Sindicato de Trabajadores Académicos de la UNISON

. . . y por supuesto, ¡Gracias Jehová! por tu constancia en mi camino.

Un verdadero amigo:
Siente tus miedos, pero fortifica tu fe;
Conoce tus debilidades, pero te hace notar tus fortalezas;
Reconoce tu falta de habilidad, pero acentúa tus posibilidades.

William Arthur Ward

Abstract

Semi-Markov control models ($SMCM$s) are a class of continuous time stochastic control models where the distribution of the random times between consecutive decision epochs (holding or sojourn times) is arbitrary. Its evolution in the time can be described as follows. At time of the nth decision epoch T_n, the system is in the state $x_n = x$ and the controller chooses a control $a_n = a \in A(x)$. Then, the system remains in the state x during a nonnegative random time δ_{n+1} with distribution $F(\cdot|x, a)$, and the following happen:

1) an immediate cost $D(x, a)$ is incurred;

2) the system jumps to a new state $x_{n+1} = y$ according to a transition law $Q(\cdot \mid x, a)$; and

3) a cost rate $d(x, a)$ is imposed until the transition occurs.

Once the transition to state y occurs, the process is repeated. In addition, the operation costs are accumulated during the evolution of the system in an infinite horizon according to either the discounted or average cost criterion. Therefore, the optimal control problem is to find control policies directed to minimize the corresponding optimality criterion.

The work deals with $SMCM$s with Borel state and control spaces, possibly unbounded costs, under discounted and average cost criteria. Specifically, our main objective is to study the optimal control problem associated to semi-Markov control systems in which the holding time distribution F is unknown by the controller. This study is developed under three different approaches, which are described as follows.

First, we assume that the holding times are observable and the distribution function F is unknown but independent of the state-action pairs (x, a). In this case, we follow standard schemes on adaptive control, which consist in the combination of statistical estimation methods with control procedures to construct optimal policies under the average criterion. Hence, our approach is a sequel to [29], in which the discounted cost criterion is analyzed. That is, by applying estimation and control procedures we construct asymptotically discounted optimal policies.

The condition that the holding time distribution is independent of the state-action pairs is satisfied in a natural way, for instance, in certain class of controlled queueing systems as well as, in some storage systems (see, e.g., [30, 42]). However, such independence property is not realistic or might be strong for another class of application problems, for example, when the distribution F depends of a parameter which in turns is a function of the state-action pairs.

Considering the last fact, the second approach to study our problem is to assume that F depends on the state and the control selected by the controller in each decision epoch, and additionally it also depends on an unknown and possibly non-observable parameter which may change from stage to stage. In this sense, the distribution of the holding times is partially known by the controller.

In the previous approach, assuming observability of the holding times and independence of F respect to the states and actions, it was possible to apply standard schemes consisting in the combination of estimation methods with control procedures to construct optimal strategies. In general, when it is possible, the unknown component is estimated from historical data using statistical methods. However, in this case, because the dependence on the state-action pairs, as well as the non-observability of the unknown parameter, this standard approach is not possible to apply. Instead, we assume that, at each stage, the only information owing the controller is that the parameter belongs to a suitable set Θ. Hence, the controller is interested in select actions directed to *minimize* the *maximum* cost generated on the corresponding set of parameters.

Specifically, under this setting, we introduce a class of minimax semi-Markov control models to study this semi-Markov optimal control problem. The approach consists in supposing that the controller has an opponent which, at each decision epoch, once the controller choose his/her action, picks a parameter from the set Θ which might depend on the current state of the system and on the action selected. Thus, the goal of the controller is to minimize the maximum cost incurred by the opponent. That is, the controller must select actions guaranteeing the best performance in the worst possible situation. Therefore, our main objective is to show the existence of minimax strategies under both, the discounted and the average criteria.

Finally, our third approach is related to the following stability control problem. Let $V_\alpha(\pi)$ be the total discounted cost under the policy π, and let us assume that it is possible to get an approximate distribution $\tilde{F}(\cdot \mid x, a)$. Under this setting, considering the control problem for $\tilde{F}$ and the corresponding total discounted cost $\tilde{V}_\alpha$, we can obtain a policy $\tilde{\pi}$ minimizing $\tilde{V}_\alpha$. Then, if the controller uses $\tilde{\pi}$ to control the original semi-Markov process, our main objective is to estimate the *optimality*

deviation of the control policy $\tilde{\pi}$ defined as

$$V_\alpha(\tilde{\pi}) - \inf_\pi V_\alpha(\pi).$$

Additionally, we construct asymptotically discounted optimal (ADO) policies applying estimation and control procedures as is shown in [29]. Indeed, if we consider the particular case when F has an unknown density g independent of the state-actions pairs, and the holding times are observable, it is possible to apply some statistical density estimation method to obtain a consistent estimator g_n. Then, using g_n as the approximation of the density of the holding time distribution, the resulting policy will be ADO for the original $SMCM$.

Similar estimation problems of the deviation of optimality has been studied by several authors but for Markov control processes, and under different names, namely, robustness, perturbation, or stability index.

$SMCM$s have been widely studied under several contexts (see, e.g., [4, 5, 20, 21, 22, 26, 27, 28, 32, 33, 47]). In these references it is assumed that all components of the $SMCM$s are known by the controller. In contrast, the main feature of our work is to suppose the distribution of the holding times unknown. To the best of our knowledge the only paper that studies this kind of $SMCM$s is [29] which considers a discounted cost criterion. It is exactly in this point where this thesis takes special importance through the three approaches previously described.

Through the preparation of the work some new interesting problems were arose from the theory developed in this thesis, which could be studied in the future. Indeed, one of them consists to suppose that the holding time distribution function F depends from (x, a) and it is completely unknown by the controller, that is, to consider the non-parametric case. In this situation, instead of to assume a set of parameters, we would consider a set of probability distributions, —from which the opponent would select its actions—; thus, we would have to impose conditions about σ-compactness on this new set.

Of course, other problem to solve consists to estimate the optimality deviation under the average cost criterion. In such situation, the main difficulties would be in the denominator of the expression corresponding to this optimality criterion, —where the holding time is included—, and also, in the transformation from the semi-Markov control model to a Markov control model.

The thesis is structured in five chapters. In Chapter 1, we present the semi-Markov control models and the optimal control problem associated. Next, in Chapter

2, we study the semi-Markov control processes in which we are assuming that the holding time distribution function F is independent of the state-action pairs, and under discounted and average optimality criteria. It is worth noting that, even when the discounted case already was solved in [29], we are including it in this chapter for completeness and an easy reference. In Chapter 3, we analyze the semi-Markov control model as a minimax control problem. In this sense, we introduce all elements to define the discounted and average minimax criteria, to show the existence of minimax policies for both cases. In Chapter 4, we present the result related with the estimation of the optimality deviation of control policies under the discounted criterion. Finally, in Chapter 5, we present our conclusions, and also, we set out some ideas for possible extensions on the theory developed.

Introducción

Los modelos de control semi-markovianos son una clase de modelos de control estocástico en tiempo continuo, donde la distribución de los tiempos entre épocas de decisión consecutivas es arbitraria, y las acciones son elegidas en los tiempos de transición.

Un modelo de control semi-markoviano representa un sistema estocástico controlado, cuya evolución en el tiempo se puede describir de la siguiente manera. Si en el tiempo de la n-ésima época de decisión el sistema se encuentra en el estado $x_n = x$, entonces el controlador elige una acción o control $a_n = a$ y sucede lo siguiente:

1) el sistema permanece en el estado x durante un tiempo aleatorio no negativo δ_{n+1} con distribución F;

2) se genera un costo que depende del estado, el control y el tiempo de permanencia;

3) el sistema se mueve a un nuevo estado $x_{n+1} = y$ de acuerdo a una ley de transición.

Una vez ocurrido lo anterior, el proceso se repite. Durante la evolución del sistema los costos se acumulan de acuerdo a determinado criterio de optimalidad (llamado también índice de funcionamiento); en particular, a lo largo de este trabajo se consideran dos tipos de criterio, el de costo descontado y el de costo promedio. De manera que el problema de control óptimo consiste en encontrar políticas de control que minimicen el criterio de optimalidad en consideración.

Desde un punto de vista más realista suponemos que el controlador carece de información precisa acerca de la distribución del tiempo de permanencia F, ya sea porque desconoce completamente dicha distribución, o bien, la desconoce pero sabe que posee una densidad, o tal vez, la distribución es conocida excepto por un parámetro, entre otras situaciones posibles.

El objetivo del presente trabajo consiste en estudiar el problema de control óptimo asociado a sistemas de control semi-markovianos donde la distribución del tiempo de permanencia F es desconocida por el controlador. El estudio se hará bajo tres puntos de vista diferentes, los cuales describimos a continuación.

Primeramente, supondremos que los tiempos de permanencia son observables y que la distribución F tiene una densidad desconocida que no depende de los pares

estado-acción (x, a). En este escenario seguimos esquemas usuales de control adaptado que combina métodos estadísticos de estimación de la densidad con técnicas de control para la construcción de políticas óptimas. Este problema fue analizado por Luque-Vásquez y Minjárez-Sosa en [29] para el criterio de costo descontado. Nuestro trabajo se centrará en el análisis bajo criterio de costo promedio.

La hipótesis de que la distribución F es independiente del par estado-acción y que el tiempo de permanencia δ es observable, resulta clave para el proceso de estimación de la densidad, ya que un estimador se define en términos de las observaciones históricas de δ. Aunque dicha hipótesis se satisface de manera natural por ciertos sistemas de almacenamiento (véase [42]), ésta podría ser muy fuerte o poco realista para otra clase de problemas.

Considerando este último aspecto, el segundo enfoque para el estudio de nuestro problema supone que F, además de depender de los pares (x, a), depende de un parámetro desconocido, el cual es posiblemente no observable y puede cambiar de una etapa a otra. Tales circunstancias nos impiden aplicar el procedimiento estándar de combinar estimación estadística y control a este caso. En este contexto, supondremos que la única información disponible para el controlador en cada etapa es que el parámetro antes mencionado pertenece a un conjunto Θ. En este caso, el controlador está interesado en seleccionar acciones dirigidas a minimizar el máximo costo que se pueda generar en el conjunto de parámetros.

Para estudiar esta situación introduciremos una clase de modelos de control minimax semi-markovianos conocidos como *juegos contra la naturaleza*. En general, nuestro análisis consiste en suponer que el controlador tiene un oponente, "la naturaleza", quien en cada época de decisión, una vez que el controlador elige su acción, selecciona un parámetro $\theta \in \Theta\,(x, a)$. El objetivo del controlador es construir políticas que minimicen el máximo costo generado por la naturaleza. Este problema será analizado bajo los criterios de optimalidad descontado y promedio.

Finalmente, el tercer enfoque consiste en lo siguiente. Supongamos que es posible obtener una función de distribución $\tilde{F}$ para el tiempo de permanencia, y una política óptima $\tilde{\pi}$ para el problema de control óptimo correspondiente. Entonces, si el controlador usa $\tilde{\pi}$ para controlar el proceso semi-markoviano original, nuestro problema es estimar la desviación de optimalidad de la política $\tilde{\pi}$.

Cada uno de los enfoques previamente descritos dieron lugar a las publicaciones [30, 31, 44], respectivamente.

Los modelos semi-markovianos han sido estudiados ampliamente bajo diferentes contextos: espacio de estado numerable o de Borel, considerando el criterio descontado o promedio, etc., (véase [4, 5, 20, 21, 22, 26, 27, 28, 32, 33, 47]). En cada

una de estas referencias se asume que todas las componentes del modelo de control correspondiente son completamente conocidas. Es justamente en este punto donde este trabajo toma mayor importancia a través de los tres enfoques descritos anteriormente.

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

En el Capítulo 1 se presentan los modelos de control semi-markovianos, así como el problema de control óptimo asociado. Además, se introducen las hipótesis sobre el modelo general.

En el Capítulo 2 se estudian modelos de control semi-markovianos en los cuales se supone que la distribución del tiempo de permanencia F tiene una densidad desconocida independiente del estado y el control bajo los criterios de optimalidad descontado y promedio. Aún cuando el caso descontado fue resuelto previamente en [29], lo incluimos aquí por completez y una fácil referencia.

En el Capítulo 3 se analiza el modelo de control semi-markoviano como un problema de control minimax. En este sentido se introducen todos los elementos para definir los criterios minimax descontado y promedio. En ambos casos se muestra la existencia de políticas minimax.

En el Capítulo 4 se aborda el problema de estimar la desviación de optimalidad de políticas de control. Aquí nuestro estudio se centra únicamente en el caso descontado.

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

Capítulo 1

Procesos de Control Semi-markovianos

1.1. Introducción

En este capítulo definimos los elementos principales del modelo de control semi-markoviano que estaremos estudiando en este trabajo. Asimismo, se introduce el problema de control óptimo correspondiente, así como las hipótesis que garantizan su solución.

Estas hipótesis incluyen condiciones de compacidad y continuidad que garantizan la existencia de minimizadores, y condiciones sobre el crecimiento de las funciones de costo asociadas al problema de control. Estas últimas, en particular, admiten costos no acotados.

Notación.

- A lo largo de este trabajo denotaremos por $\mathbb{N}$, $\mathbb{N}_0$, $\mathbb{R}$ y $\mathbb{R}_+$, al conjunto de los números: enteros positivos, enteros no negativos, reales y reales no negativos, respectivamente. Asimismo, dado un espacio de Borel $\mathbb{Y}$ (es decir, un subconjunto de Borel de un espacio métrico completo y separable), $\mathcal{B}(\mathbb{Y})$ representará su σ-álgebra de Borel. Además, la medibilidad, tanto de conjuntos como de funciones, será respecto a $\mathcal{B}(\mathbb{Y})$.

- Sean $\mathbb{Y}$ y $\mathbb{Z}$ dos espacios medibles. Un *kérnel estocástico* sobre $\mathbb{Y}$ dado $\mathbb{Z}$ es una función $Q(\cdot \mid \cdot)$ tal que:

(a) $Q(\cdot \mid z)$ es una medida de probabilidad sobre $\mathbb{Y}$ para cada $z \in \mathbb{Z}$;

(b) $Q(B \mid \cdot)$ es una función medible sobre $\mathbb{Z}$ para cada $B \in \mathcal{B}(\mathbb{Y})$.

1.2. Modelo de control semi-markoviano

1.2.1. Descripción

Definicin 1.1 *Un modelo de control semi-markoviano (MCSM)*

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

consiste del espacio de estados $\mathbb{X}$ y el conjunto de control (o acciones) $\mathbb{A}$, los cuales se supondrán son espacios de Borel. Para cada $x \in \mathbb{X}$, asociamos un subconjunto medible no vacío $A(x)$ de $\mathbb{A}$, cuyos elementos son los controles admisibles (o acciones) para el estado x. Suponemos que el conjunto

$$\mathbb{K}_A = \{(x, a) : x \in \mathbb{X}, a \in A(x)\} \tag{1.2}$$

de parejas estado-acción admisibles es un subconjunto de Borel de $\mathbb{X} \times \mathbb{A}$. La ley de transición $Q(\cdot \mid \cdot)$ es un kérnel estocástico sobre $\mathbb{X}$ dado $\mathbb{K}_A$, y $F(\cdot \mid x, a)$ es la función de distribución del tiempo de permanencia en el estado $x \in \mathbb{X}$ cuando es elegido el control $a \in A(x)$. Finalmente, las funciones de costo D y d son reales, medibles y posiblemente no-acotadas en $\mathbb{K}_A$.

1.2.2. Interpretación

Un $MCSM$ representa un sistema dinámico que evoluciona en el tiempo de la siguiente manera. Al tiempo de la n-ésima época de decisión (o n-ésima transición) T_n, el sistema está en el estado $x_n = x$ y el controlador elige un control $a_n = a \in A(x)$, con lo cual se genera lo que se describe a continuación:

1. Se incurre en un costo inmediato $D(x, a)$.

2. El sistema permanece en el estado $x_n = x$ durante un tiempo aleatorio no-negativo δ_{n+1} con función de distribución $F(\cdot \mid x, a)$.

3. Luego, en el tiempo $T_{n+1} := T_n + \delta_{n+1}$ ($n \in \mathbb{N}_0$, $T_0 := 0$), el sistema pasa a un nuevo estado $x_{n+1} = y$ de acuerdo a la ley de transición $Q(\cdot \mid x, a)$.

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

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

1.3. Políticas de control

Definicin 1.2 *Dado un MCSM, se define el espacio de historias admisibles hasta la n-ésima época de decisión mediante* $\mathbb{H}_0 := \mathbb{X}$ *y*

$$\mathbb{H}_n := (\mathbb{K}_A \times \mathbb{R}_+)^n \times \mathbb{X}, \quad \text{para } n \in \mathbb{N}.$$

Sea $i_n \in \mathbb{H}_n$, $n \in \mathbb{N}_0$, un vector de la forma

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

donde $(x_k, a_k, \delta_{k+1}) \in \mathbb{K}_A \times \mathbb{R}_+$ para $k = 0, 1, ..., n-1$ y $x_n \in \mathbb{X}$. En lo sucesivo llamaremos a i_n la n-historia.

Definicin 1.3 *Una política de control (o simplemente una política) es una sucesión* $\pi = \{\pi_n\}$ *de kérneles estocásticos* π_n *sobre* $\mathbb{A}$ *dado* $\mathbb{H}_n$*, tal que* $\pi_n(A(x_n) \mid i_n) = 1$ *para cada* $i_n \in \mathbb{H}_n$*,* $n \in \mathbb{N}_0$*.*

Se denotará por Π a la colección de todas las políticas.

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

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

(a) determinística si existe una sucesión $\{\bar{f}_n\}$ *de funciones medibles* $\bar{f}_n : \mathbb{H}_n \to \mathbb{A}$ *tal que* $\pi_n(\cdot \mid i_n)$ *está concentrada en* $\bar{f}_n(i_n)$ *para toda* $i_n \in \mathbb{H}_n$ *y* $n \in \mathbb{N}_0$*; es decir,* $\pi_n(B \mid i_n) = 1_B\left(\bar{f}_n(i_n)\right)$*,* $B \in \mathcal{B}(\mathbb{A})$*.*

(b) markoviana si existe una sucesión $\{f_n\}$ *de funciones medibles* $f_n \in \mathbb{F}$ *tal que* $\pi_n(\cdot \mid i_n)$ *está concentrada en* $f_n(x_n)$ *para toda* $i_n \in \mathbb{H}_n$ *y* $n \in \mathbb{N}_0$*.*

(c) estacionaria si existe $f \in \mathbb{F}$ *tal que* $\pi_n(\cdot \mid i_n)$ *está concentrada en* $f(x_n) \in A(x_n)$ *para toda* $i_n \in \mathbb{H}_n$ *y* $n \in \mathbb{N}_0$*.*

Identificaremos a una política estacionaria con la función f y denotaremos al conjunto de todas las políticas estacionarias como $\mathbb{F}$.

Sea $(\Omega, \mathcal{F})$ el espacio medible canónico consistente del espacio muestral $\Omega := (\mathbb{X} \times \mathbb{A} \times \mathbb{R}_+)^\infty$ y la σ-álgebra producto correspondiente $\mathcal{F}$. Por el Teorema de C.

Ionescu Tulcea (véase, por ejemplo, [1, Teorema 2.7.2]), para cada estado inicial $x \in \mathbb{X}$ y cada política $\pi \in \Pi$, existe una medida de probabilidad P_x^π tal que para todo $B \in \mathcal{B}(\mathbb{A})$, $C \in \mathcal{B}(\mathbb{X})$, $i_n \in \mathbb{H}_n$ con $n \in \mathbb{N}_0$, se tiene

$$P_x^\pi \{x_0 = x\} = 1, \tag{1.3}$$

$$P_x^\pi \{a_n \in B \mid i_n\} = \pi_n (B \mid i_n), \tag{1.4}$$

$$P_x^\pi \{x_{n+1} \in C \mid i_n, a_n, \delta_{n+1}\} = Q (C \mid x_n, a_n), \tag{1.5}$$

y

$$P_x^\pi \{\delta_{n+1} \leq t \mid i_n, a_n\} = F (t \mid x_n, a_n). \tag{1.6}$$

1.4. Criterios de optimalidad

Para formular el problema de control óptimo (PCO) es necesario introducir una función para medir el comportamiento del sistema. Tal función se conoce como *criterio de optimalidad* o *índice de funcionamiento*. A lo largo de esta tesis consideraremos dos tipos de criterios, a saber, costo descontado y costo promedio.

En particular, para el caso de costo descontado supondremos que los costos son descontados en forma continua. Esto es, para un factor de descuento $\alpha > 0$, un costo K_0 generado al tiempo t equivale a un costo $K_0 \exp(-\alpha t)$ al tiempo 0. De modo que, de acuerdo a la interpretación del $MCSM$, la función de *costo por etapa* en este caso toma la forma

$$c_\alpha(x, a) := D (x, a) + d (x, a) \int_0^\infty \int_0^t \exp (-\alpha s)\, ds F (dt \mid x, a), \tag{1.7}$$

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

Las funciones introducidas a continuación se usarán en resultados posteriores.

Definicin 1.5 *Para cada $(x, a) \in \mathbb{K}_A$ se define*

$$\Delta_\alpha (x, a) := \int_0^\infty \exp (-\alpha t)\, F(dt \mid x, a),$$

y

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

Observe que de (1.7) junto con la definición previa se tiene que

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

Esto es,

$$c_\alpha(x,a) = D(x,a) + \tau_\alpha(x,a)\, d(x,a), \qquad (x,a) \in \mathbb{K}_A. \tag{1.8}$$

Definicin 1.6 *Sean $x_0 = x \in \mathbb{X}$, $\pi \in \Pi$, y $\alpha > 0$. Se define el costo total esperado descontado (CD) como*

$$V(x,\pi) := E_x^\pi\left[\sum_{n=0}^\infty \exp(-\alpha T_n)\, c_\alpha(x_n,a_n)\right], \tag{1.9}$$

donde E_x^π denota el operador esperanza con respecto a la medida de probabilidad P_x^π inducida por $\pi \in \Pi$, dado que $x_0 = x$, (véase [3]).

Ahora introduciremos el criterio de costo promedio.

Definicin 1.7 *Para $x_0 = x \in \mathbb{X}$ y $\pi \in \Pi$ definimos el costo esperado promedio (CP)como*

$$J(x,\pi) := \limsup_{n\to\infty} \frac{E_x^\pi\left[\sum_{k=0}^{n-1}\{D(x_k,a_k) + \delta_{k+1} d(x_k,a_k)\}\right]}{E_x^\pi[T_n]}. \tag{1.10}$$

Definicin 1.8 *Para cada $(x,a) \in \mathbb{K}_A$ se define el tiempo medio de permanencia por*

$$\tau(x,a) := \int_0^\infty t F(dt \mid x,a), \tag{1.11}$$

así como el costo por etapa mediante

$$c(x,a) := D(x,a) + \tau(x,a)\, d(x,a). \tag{1.12}$$

Por propiedades de la esperanza condicional y de la definición de T_n (ver Sección 1.2, Capítulo 1) se verifica que

$$J(x,\pi) = \limsup_{n\to\infty} \frac{E_x^\pi\left[\sum_{k=0}^{n-1} c(x_k,a_k)\right]}{E_x^\pi\left[\sum_{k=0}^{n-1} \tau(x_k,a_k)\right]} \quad \forall x \in \mathbb{X}, \pi \in \Pi. \tag{1.13}$$

Dado un $MCSM$, una colección Π de políticas admisibles y un índice de funcionamiento específico, el PCO consiste en encontrar una política $\pi^* \in \Pi$ que minimice el criterio de optimalidad respectivo. Por ejemplo, en el caso del criterio de costo descontado, el PCO consiste en determinar una política $\pi^* \in \Pi$ tal que

$$V^*(x) := \inf_{\pi\in\Pi} V(x,\pi) = V(x,\pi^*) \quad \forall x \in \mathbb{X}, \tag{1.14}$$

en cuyo caso, a π^* la llamaremos *política CD-óptima*. Mientras que, respecto al caso de criterio de costo promedio, el PCO es encontrar una política $\pi^* \in \Pi$ tal que

$$J^*(x) := \inf_{\pi\in\Pi} J(x,\pi) = J(x,\pi^*), \quad x \in \mathbb{X}. \tag{1.15}$$

A la política π^* la llamaremos *política CP-óptima*.

Finalmente, llamaremos a $V^*(\cdot)$ y a $J^*(\cdot)$ la *función de valor óptimo descontado* y la *función de valor óptimo promedio*, respectivamente.

1.5. Hipótesis generales sobre el modelo de control

En esta sección introduciremos condiciones generales sobre $MCSMs$ que estaremos usando a lo largo del trabajo.

Hiptesis 1.1 *Existen constantes positivas ϵ y ζ tal que*

$$F(\zeta \mid x,a) \leq 1-\epsilon \qquad \forall\,(x,a) \in \mathbb{K}_A.$$

Esta hipótesis es una condición de regularidad, la cual garantiza que en intervalos de tiempo finito, no puede ocurrir una infinidad de transiciones del proceso, propiedad que en particular, queda establecida en la parte *(c)* de la proposición siguiente. De hecho, el resultado a continuación es consecuencia de la Hipótesis 1.1, (véase [27, p. 13]).

Proposicin 1.1 *Si se cumple la Hipótesis 1.1, entonces:*

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

(b) $\rho := \inf_{\mathbb{K}_A} \tau(x,a) \geq \epsilon\zeta$,

(c) $P_x^\pi \{\sum_{n=1}^\infty \delta_n = \infty\} = 1 \quad \forall\, x \in \mathbb{X},\, \pi \in \Pi$.

En general, estaremos suponiendo la existencia de una función medible $W : \mathbb{X} \to [1,\infty)$ que domina a las funciones de costo por etapa correspondientes. Denotaremos por $\mathbb{B}_W(\mathbb{X})$ el espacio lineal normado de las funciones medibles $u : \mathbb{X} \to \mathbb{R}$ que satisfacen la condición

$$\|u\|_W := \sup_{x\in\mathbb{X}} \frac{|u(x)|}{W(x)} < \infty. \tag{1.16}$$

Una propiedad de la función W, que estaremos usando repetidamente en el desarrollo de este trabajo, es la siguiente (véase [9, 10, 16]).

Lema 1.1 *Si existen constantes* $b_0 > 0$, $0 < \beta_0 < 1$ *y* $p > 1$, *tal que la función* W *satisface la desigualdad*

$$\int_{\mathbb{X}} W^p(y)\, Q(dy \mid x,a) \leq \beta_0 W^p(x) + b_0 \quad \forall x \in \mathbb{X},\, a \in A(x), \tag{1.17}$$

entonces,

$$\sup_{n\in\mathbb{N}_0} E_x^\pi [W^p(x_n)] < \infty \quad \forall x \in \mathbb{X}, \pi \in \Pi. \tag{1.18}$$

De hecho, la desigualdad (1.17) implica

$$\int_{\mathbb{X}} W(y) Q(dy \mid x,a) \leq \beta W(x) + b \quad \forall\, (x,a) \in \mathbb{K}_A, \tag{1.19}$$

donde $\beta := \beta_0^{1/p}$ y $b := b_0^{1/p}$, y a su vez, (1.19) implica

$$\sup_{n\in\mathbb{N}_0} E_x^\pi [W(x_n)] < \infty, \quad \forall x \in \mathbb{X}, \pi \in \Pi.. \tag{1.20}$$

En cada uno de los siguientes capítulos se impondrán condiciones adicionales sobre el modelo de control, dependiendo del problema específico que se trate.

Capítulo 2

$MCSMs$ con Distribución del Tiempo de Permanencia Desconocida

2.1. Introducción

En este capítulo analizaremos el PCO para $MCSMs$ bajo la suposición de que la función de distribución del tiempo de permanencia F es desconocida por el controlador. Específicamente, enfocaremos nuestro interés en la construcción de políticas de control dirigidas a minimizar el índice de costo descontado y el de costo promedio, para lo cual se implementará una combinación de técnicas de control y métodos adecuados de estimación estadística de F.

Dado que F es desconocida, nótese que de (1.7) y (1.12), los costos por etapa para el costo descontado y el promedio, c_α y c, respectivamente, son también desconocidos. Por tanto, antes de elegir el control a_n en la n-ésima época de decisión, el controlador debe construir un estimador F_n de la distribución F (por consiguiente de c_α o c), y combinar esto con la historia del sistema para seleccionar un control $a = a_n(F_n)$.

Por otra parte, la distribución del tiempo de permanencia F en general depende del estado x y del control a, hecho que en particular dificulta la obtención de una cantidad considerable de observaciones independientes e idénticamente distribuidas (i.i.d.) de los tiempos de permanencia, y en consecuencia, la obtención de estimadores consistentes de la distribución F en cada época de decisión. Por esta razón, a lo largo del presente capítulo supondremos que la distribución F tiene una densidad que no depende de los pares (x, a).

En la primera sección abordaremos el criterio de costo descontado, el cual fue analizado por Luque-Vásquez y Minjárez-Sosa en [29]. Por completez y para fácil referencia, presentaremos únicamente los resultados principales de dicho trabajo, sin incluir sus demostraciones.

2.2. Criterio de costo descontado

2.2.1. Hipótesis y resultados preliminares

Hiptesis 2.1 *Existe una función de densidad g tal que*

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

En la hipótesis siguiente se imponen condiciones sobre la densidad g, con los cuales podemos obtener un método apropiado de estimación.

Hiptesis 2.2 *Existe una constante $q \in (1,2)$ y una función medible $\bar{g} : [0,\infty) \to [0,\infty)$ tal que $g \in L_q([0,\infty))$, $g(s) \leq \bar{g}(s)$ c.d. con respecto a la medida de Lebesgue y*

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

Asimismo, consideraremos también las dos condiciones siguientes, siendo la primera de crecimiento sobre las funciones de costo, y la segunda sobre continuidad y compacidad.

Hiptesis 2.3 *(a) Existe una función medible $W : \mathbb{X} \to [1,\infty)$, asi como constantes positivas $\bar{c}_1$, $\bar{c}_2$, $p > 1$, b_0 y $0 < \beta_0 < 1$ tal que*

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

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

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

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

es continua en $A(x)$.

Hiptesis 2.4 *(a) Para cada $x \in \mathbb{X}$, $A(x)$ es un conjunto compacto. Además, las funciones de costo $D(x,\cdot)$ y $d(x,\cdot)$ son semicontinuas inferiormente (s.c.i.) en $a \in A(x)$ para cada $x \in \mathbb{X}$.*

(b) La ley de transición $Q(\cdot \mid x,a)$ es débilmente continua en $a \in A(x)$ para cada $x \in \mathbb{X}$, es decir, para cada función continua y acotada $v : \mathbb{X} \to \mathbb{R}$,

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

es una función continua y acotada en $A(x)$ para cada $x \in \mathbb{X}$.

Observacin 2.1 *(a) Bajo la Hipótesis 2.1, observe que la Hipótesis 1.1 (Sección 1.5, Capítulo 1) toma la forma*

$$\int_0^{\zeta} g(s)ds \leq 1-\epsilon,$$

lo cual es trivial.

(b) Además, para cada $(x,a) \in \mathbb{K}_A$, las expresiones de la Definición 1.5 toman ahora la forma

$$\Delta_\alpha := \int_0^{\infty} \exp(-\alpha t)\, g(t)dt \qquad y \qquad \tau_\alpha := \frac{1-\Delta_\alpha}{\alpha}, \tag{2.1}$$

respectivamente, por lo cual el costo por etapa (1.8) puede ser reescrito como

$$c_\alpha(x,a) = D(x,a) + \tau_\alpha d(x,a), \qquad (x,a) \in \mathbb{K}_A. \tag{2.2}$$

(c) Asimismo, de (1.6) tenemos que para cada $x \in \mathbb{X}$ y $\pi \in \Pi$,

$$P_x^\pi(\delta_n \leq t) = \int_0^t g(s)\, ds, \tag{2.3}$$

y por la Proposición 1.1(a),

$$\rho_\alpha = \Delta_\alpha < 1. \tag{2.4}$$

(d) Por otra parte, observemos que de (2.1) y la Proposición 1.1(a) se cumple que $\tau_\alpha < 1/\alpha$, por lo cual, de acuerdo con (2.2) y la Hipótesis 2.3(a) vemos que

$$\sup_{a \in A(x)} |c_\alpha(x,a)| \leq \bar{c}_0 W(x), \qquad x \in \mathbb{X}, \tag{2.5}$$

donde $\bar{c}_0 := \bar{c}_1 + \frac{1}{\alpha}\bar{c}_2$.

Usando propiedades de la esperanza condicional se demuestra que una consecuencia de (1.6) y (2.5) junto con la Definición 1.5, es la siguiente, (véase [32]).

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

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

Una primera consecuencia de las Hipótesis previas es la siguiente (véase [16, 32]).

Teorema 2.1 *Supóngase que se satisfacen las Hipótesis 1.1, 2.1-2.4. Entonces:*

(a) La función V^* pertenece a $\mathbb{B}_W(\mathbb{X})$ y satisface la ecuación de optimalidad en costo descontado (EOCD)

$$V^*(x) = \min_{a\in A(x)}\left\{c_{\alpha}(x,a) + \Delta_{\alpha}\int_{\mathbb{X}} V^*(y)\,Q(dy \mid x,a)\right\} \quad \forall x \in \mathbb{X}; \tag{2.6}$$

o equivalentemente,

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

donde

$$\Phi(x,a) := c_{\alpha}(x,a) + \Delta_{\alpha}\int_{\mathbb{X}} V^*(y)\,Q(dy \mid x,a) - V^*(x), \tag{2.7}$$

es una función no negativa debido a (2.6), llamada función de discrepancia. Además

$$|V^*(x)| \leq \bar{M}' W(x), \tag{2.8}$$

con $\bar{M}' := \bar{c}_0\left(1+\frac{b}{1-\beta}\right)\left(\frac{1}{1-\Delta_{\alpha}}\right)$.

(b) Existe $f^* \in \mathbb{F}$ tal que $f^*(x) \in A(x)$ alcanza el mínimo en (2.6), es decir,

$$V^*(x) = c(x,f^*) + \Delta_{\alpha}\int_{\mathbb{X}} V^*(y)\,Q(dy \mid x,f^*) \quad \forall x \in \mathbb{X}, \tag{2.9}$$

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

Optimalidad asintótica. Sea $\{(x_n, a_n)\}$ una sucesión de pares estado-acción correspondiente a la aplicación de una política $\pi \in \Pi$. En [32] se demostró que π es una política óptima si y sólo si

$$E_x^{\pi}\left[\Phi\left(x_n, a_n\right)\right] = 0 \quad \forall n \in \mathbb{N}_0, x \in \mathbb{X};$$

lo cual nos permite analizar la optimalidad de una política mediante la función Φ. En tales circunstancias, la optimalidad de las políticas construidas en [29] fue estudiada en el sentido de la definición siguiente.

Definicin 2.1 *Una política $\pi \in \Pi$ es asintóticamente óptima descontada (AOD) para el MCSM (1.1), si para cada $x \in \mathbb{X}$ se cumple*

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

2.2.2. Estimación y control

Estimación de la densidad. Sean $\delta_1, ..., \delta_n$ realizaciones independientes de variables aleatorias con densidad desconocida g, observadas hasta el momento de la n-ésima época de decisión, y sea $\hat{g}_n\left(s\right) := \hat{g}_n\left(s; \delta_1, ..., \delta_n\right)$, $s \in \mathbb{R}_+$, un estimador arbitrario de g tal que

$$E\left[\left(\left\|g - \hat{g}_n\right\|_q\right)^{q/2}\right] \to 0 \quad \text{cuando} \;\; n \to \infty, \tag{2.10}$$

(véase [2, 13], para ejemplos de estimadores con esta propiedad). Además, sea λ una constante tal que $\Delta_\alpha \leq \lambda < 1$ (véase (2.1) y (2.4)). Definimos el conjunto $\mathbb{D} \subset L_q\left([0, \infty)\right)$ como:

$$\begin{aligned}\mathbb{D} := \Big\{\tilde{\mu} : \tilde{\mu} \text{ es una densidad en } L_q\left([0,\infty)\right), \\ \int_0^{\infty} \exp\left(-\alpha s\right)\tilde{\mu}\left(s\right) ds \leq \lambda, \\ \int_0^{\zeta} \tilde{\mu}\left(s\right) ds < 1 - \epsilon, \\ \tilde{\mu}\left(s\right) \leq \bar{g}\left(s\right) \quad c.d.\Big\}.\end{aligned} \tag{2.11}$$

Nótese que $g \in \mathbb{D}$. Además, como se demostró en [29], el conjunto $\mathbb{D}$ es cerrado y convexo en $L_q\left([0, \infty)\right)$. Con esta propiedad, y aplicando las Proposiciones 2 y 3 en

[23, p. 343], existe una única densidad $g_n \in \mathbb{D}$, la cual es la proyección del estimador $\hat{g}_n$ sobre el conjunto $\mathbb{D}$, y es además un estimador consistente de la densidad g en el sentido de que

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

Construcción de políticas. Para cada $n \in \mathbb{N}$, definimos la *función de costo descontado aproximado* (por etapa) como

$$c_\alpha^n(x,a) := D(x,a) + \tau_\alpha^n d(x,a), \qquad (x,a) \in \mathbb{K}_A, \tag{2.12}$$

(véase (2.1) y (2.2)), donde

$$\tau_\alpha^n := \frac{1-\Delta_\alpha^n}{\alpha} \quad \text{y} \quad \Delta_\alpha^n := \int_0^\infty \exp(-\alpha t)\, g_n(t) dt. \tag{2.13}$$

Nótese que, de la Hipótesis 2.3*(a)*, para cada $n \in \mathbb{N}$ y $x \in \mathbb{X}$ (véase (2.5)),

$$\sup_{a \in A(x)} |c_\alpha^n(x,a)| \le \bar{c}_0 W(x). \tag{2.14}$$

Por otra parte, definimos la sucesión $\{V_n^*\}_{n\in\mathbb{N}}$ de funciones en $\mathbb{B}_W(\mathbb{X})$ como $V_0^* := 0$, y para $n \in \mathbb{N}$,

$$V_n^*(x) := \min_{a \in A(x)} \left\{ c_\alpha^n(x,a) + \Delta_\alpha^n \int_{\mathbb{X}} V_{n-1}^*(y)\, Q(dy \mid x,a) \right\}.$$

De hecho (véase (2.8)), para todo $n \ge 1$, $x \in \mathbb{X}$,

$$|V_n^*(x)| \le \bar{c}_0 \left(1 + \frac{b}{1-\beta}\right)\left(\frac{1}{1-\lambda}\right) W(x),$$

donde λ fue definido en (2.11). Luego, aplicando argumentos estándar sobre la existencia de minimizadores (selectores), por el Teorema 2.1*(b)*, tenemos que para cada $n \in \mathbb{N}$ existe $f_n := f_n^{g_n} \in \mathbb{F}$ tal que

$$V_n^*(x) = c_\alpha^n(x, f_n) + \Delta_\alpha^n \int_{\mathbb{X}} V_{n-1}^*(y)\, Q(dy \mid x, f_n), \qquad x \in \mathbb{X},$$

(el proceso de minimización es para cada $\omega \in \Omega$). Además, cabe señalar que debido a un resultado de Schäl (véase [45]), existe una política estacionaria $\{f_\infty\}$ tal que para cada $x \in \mathbb{X}$, $f_\infty(x) \in A(x)$ es un punto de acumulación de $\{f_n(x)\}$.

Teorema 2.2 *(a) Sea $\hat{\pi} := \{\hat{\pi}_n\}$ la política tal que*

$$\hat{\pi}_n(B \mid i_n) = \hat{\pi}_n(B \mid i_n; g_n) = 1_B(f_n(x_n)) \quad \forall B \in \mathcal{B}(\mathbb{A}), i_n \in \mathbb{H}_n, n \in \mathbb{N},$$

y f_0 un selector fijo. Entonces, bajo las Hipótesis 1.1, 2.1-2.4, se tiene que $\hat{\pi}$ es AOD, es decir,

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

(b) Además, la política $\{f_\infty\} \in \mathbb{F}$ es CD-óptima para el MCSM (1.1).

2.3. Criterio de costo promedio

2.3.1. Hipótesis sobre el modelo de control

Como en el caso previo (criterio de costo descontado), consideramos de nuevo la independencia de F respecto a (x, a) de acuerdo a la siguiente hipótesis.

Hiptesis 2.5 *La función de distribución F es independiente de (x, a), y el tiempo medio de permanencia (véase (1.11)) satisface que*

$$0 < \bar{\tau} =: \int_0^\infty tF(dt \mid x, a) < \infty. \tag{2.15}$$

En este caso, el correspondiente costo por etapa (véase (1.12)) toma la forma

$$c(x, a) = D(x, a) + \bar{\tau} d(x, a), \qquad (x, a) \in \mathbb{K}_A;$$

por lo cual, de (1.13) obtenemos que, para cada $x \in \mathbb{X}$ y $\pi \in \Pi$,

$$\begin{aligned} J(x, \pi) &= \limsup_{n\to\infty} \frac{E_x^\pi\left[\sum_{k=0}^{n-1} c(x_k, a_k)\right]}{n\bar{\tau}} \\ &= \limsup_{n\to\infty} \frac{E_x^\pi\left[\sum_{k=0}^{n-1} \bar{c}(x_k, a_k)\right]}{n}, \end{aligned} \tag{2.16}$$

donde

$$\bar{c}(x, a) := \frac{c(x, a)}{\bar{\tau}}, \qquad (x, a) \in \mathbb{K}_A. \tag{2.17}$$

De (2.15) tenemos que $\bar{\tau}$ es desconocido debido a que la distribución F se desconoce, y por lo tanto, también la función de costo por etapa c (o $\bar{c}$) será desconocida

para el controlador. En tales circunstancias, al igual que en el tratamiento del caso descontado, aquí implementaremos un proceso de estimación del parámetro $\bar{\tau}$ y, consecuentemente del costo $\bar{c}$, que combinado con un procedimiento de control generará una política CP-óptima π^* para el $MCSM$ (1.1). Para tal propósito necesitaremos las siguientes condiciones sobre el modelo de control.

Hiptesis 2.6 *(a) Para cada $x \in \mathbb{X}$, la función de costo $c(x,\cdot)$ es s.c.i. en $A(x)$.*

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

(c) La ley de transición Q es fuertemente continua en $A(x)$, es decir, para cada función medible y acotada $v : \mathbb{X} \to \mathbb{R}$,

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

es una función continua y acotada en $A(x)$ para cada $x \in \mathbb{X}$.

(d) Existe una función $W : \mathbb{X} \to [1,\infty)$ que es medible y estrictamente no acotada tal que para cada $x \in \mathbb{X}$,

$$\sup_{a \in A(x)} |c(x,a)| \leq W(x), \tag{2.18}$$

y

$$a \mapsto \int_{\mathbb{X}} W(y)\, Q(dy \mid x, a) \tag{2.19}$$

es continua en $A(x)$.

Observacin 2.2 *Que la función W sea estrictamente no acotada significa que existe una sucesión de conjuntos compactos $\{K_n\}_{n\in\mathbb{N}}$, $K_n \subset \mathbb{X}$, tal que*

$$\inf_{x \notin K_n} W(x) \to \infty \qquad \textit{cuando} \quad n \to \infty.$$

Hiptesis 2.7 *Existe una medida de probabilidad μ sobre $\mathbb{X}$ y constantes arbitrarias $p > 1$ y $\beta_0 \in (0,1)$ que satisfacen lo siguiente: Para cada $f \in \mathbb{F}$ existe una función medible no negativa ϕ_f sobre $\mathbb{X}$ tal que para cada $x \in \mathbb{X}$ y $C \in \mathcal{B}(\mathbb{X})$:*

(a) $Q(C \mid x, f) \geq \phi_f(x)\, \mu(C)$.

(b) $\int_{\mathbb{X}} W^p(y)\, Q(dy \mid x, f) \leq \beta_0 W^p(x) + \phi_f(x) \int_{\mathbb{X}} W^p(y)\, \mu(dy)$, donde

$$\int_{\mathbb{X}} W^p(y)\, \mu(dy) < \infty.$$

(c) $inf_{f\in\mathbb{F}} \int_{\mathbb{X}} \phi_f(y)\, \mu(dy) > 0$.

Una consecuencia importante de la Hipótesis 2.7 es que para cada $f \in \mathbb{F}$, la cadena de Markov definida por $Q(\cdot \mid \cdot, f)$ es μ-irreducible y *Harris-recurrente positiva*, es decir, es Harris-recurrente (véase [15, p. 189]) y tiene una única medida de probabilidad invariante q_f, es decir, una única medida tal que

$$q_f(C) = \int_{\mathbb{X}} Q(C \mid x, f)\, q_f(dx), \qquad C \in \mathcal{B}(\mathbb{X}). \tag{2.20}$$

Este resultado fue demostrado en [48], donde se asumieron condiciones más débiles que las de la Hipótesis 2.7.

Observacin 2.3 *De la Hipótesis 2.7(a) se tiene que $\phi_f \leq 1$, por lo que la parte (b) implica que para cada $x \in \mathbb{X}$ y $f \in \mathbb{F}$,*

$$\int_{\mathbb{X}} W^p(y)\, Q(dy \mid x, f) \leq \beta_0 W^p(x) + b_0, \tag{2.21}$$

con $b_0 := \int_{\mathbb{X}} W^p(y)\, \mu(dy)$ (véase (1.17), Lema 1.1). Asimismo, como en (1.19), partiendo de (2.21) y por la desigualdad de Jensen obtenemos

$$\int_{\mathbb{X}} W(y)\, Q(dy \mid x, f) \leq \beta W(x) + b, \qquad x \in \mathbb{X},\ f \in \mathbb{F}, \tag{2.22}$$

donde $\beta := (\beta_0)^{1/p}$ y $b := (b_0)^{1/p}$.

2.3.2. La ecuación de optimalidad en costo promedio. Algoritmo de iteración de valores

En el resto de este capítulo supondremos que se satisfacen las Hipótesis 2.5-2.7.

Ecuación de optimalidad en costo promedio. Se dice que el par $(j, h(\cdot))$, consistente de un número real j y una función medible $h : \mathbb{X} \to \mathbb{R}$, es una solución de la ecuación de optimalidad en costo promedio ($EOCP$) para el $MCSM$ (1.1) si para todo $x \in \mathbb{X}$,

$$h(x) = \min_{a \in A(x)} \left\{ c(x,a) - j\bar{\tau} + \int_{\mathbb{X}} h(y)\, Q(dy \mid x, a) \right\}. \tag{2.23}$$

Por (2.17), definiendo $\bar{h}(x) := h(x)/\bar{\tau}$, es evidente que (2.23) es equivalente a

$$\bar{h}(x) + j = \min_{a \in A(x)} \left\{ \bar{c}(x,a) + \int_{\mathbb{X}} \bar{h}(y)\, Q(dy \mid x, a) \right\}, \qquad x \in \mathbb{X}, \tag{2.24}$$

la cual es justamente la *EOCP* para un modelo Markoviano. En tal situación, y considerando el índice de funcionamiento (2.16), podemos establecer el siguiente resultado, el cual es demostrado en [7].

Teorema 2.3 *Si se cumplen las Hipótesis 2.5-2.7, entonces*

(a) Existen $j \in \mathbb{R}$ *y* $\bar{h} \in \mathbb{B}_W(\mathbb{X})$*, tal que* $\left(j, \bar{h}(\cdot)\right)$ *satisface la EOCP (2.24) y, por lo tanto,* $(j, h(\cdot))$ *con* $h(\cdot) = \bar{\tau}\bar{h}(\cdot)$ *satisface (2.23).*

(b) Existe $f^* \in \mathbb{F}$ *tal que*

$$\bar{h}(x) + j = \bar{c}(x, f^*) + \int_{\mathbb{X}} \bar{h}(y)\, Q(dy \mid x, f^*),$$

f^* *es CP-óptima, y* $j = J(x, f^*)$ *para todo* $x \in \mathbb{X}$*.*

Algoritmo de iteración de valores. Para cada $n \in \mathbb{N}$ definimos el *costo esperado en n etapas* bajo la política π y con estado inicial $x \in \mathbb{X}$, como

$$J_n(x, \pi) := \sum_{k=0}^{n-1} E_x^{\pi}\left[\bar{c}(x_k, a_k)\right]. \tag{2.25}$$

Las correspondientes funciones de valor óptimo también conocidas como funciones de iteración de valores son

$$v_0(x) := 0 \quad \text{y} \quad v_n(x) := \inf_{\pi \in \Pi} J_n(x, \pi), \quad n \geq 1.$$

De hecho (véase, por ejemplo, [8]), como consecuencia de las Hipótesis 2.6-2.7 se puede demostrar que $v_n \in \mathbb{B}_W(\mathbb{X})$ para todo $n \in \mathbb{N}_0$. Además, por argumentos de programación dinámica estocástica, tenemos que

$$v_n(x) = \min_{a \in A(x)} \left\{ \bar{c}(x, a) + \int_{\mathbb{X}} v_{n-1}(y)\, Q(dy \mid x, a) \right\}, \quad x \in \mathbb{X}, n \in \mathbb{N}. \tag{2.26}$$

Ahora, sea $z \in \mathbb{X}$ un estado fijo y arbitrario. Definimos una sucesión de funciones h_n y una sucesión de constantes j_n como

$$h_n(x) := v_n(x) - v_n(z), \quad x \in \mathbb{X}, \tag{2.27}$$

$$j_n := v_n(z) - v_{n-1}(z). \tag{2.28}$$

Entonces, de (2.26) tenemos

$$h_n(x) + j_n = \min_{a \in A(x)} \left\{ \bar{c}(x, a) + \int_{\mathbb{X}} h_{n-1}(y)\, Q(dy \mid x, a) \right\}.$$

El algoritmo de iteración de valores consiste en demostrar la convergencia de las sucesiones $\{h_n\}$ y $\{j_n\}$ a una solución $\left(j, \bar{h}(\cdot)\right)$ de la $EOCP$ (2.24). No obstante, para tal propósito es necesario imponer condiciones adicionales.

Denotemos por $\mathcal{G}$ a la clase de las funciones no decrecientes $\varphi : [0, \infty) \to [0, \infty)$ tales que $\varphi(s) \to 0$ cuando $s \to 0^+$.

Hiptesis 2.8 *Sean* $k := (x, a)$, $k' := (x', a')$, d *una métrica en* $\mathbb{K}_A$, *y*

$$\|\mu_0\|_{WT} := \int_{\mathbb{X}} W(y) \, |\mu_0| \, (dy),$$

para cualquier medida finita con signo μ_0 *en* $\mathbb{X}$, *donde* $|\mu_0|$ *representa la variación total de* μ_0. *Para cada* $x \in \mathbb{X}$, *existen funciones* $\varphi_x^{\bar{c}}, \varphi_x^Q \in \mathcal{G}$ *tal que para todo* $a \in A(x)$ *y* $k' \in \mathbb{K}_A$,

$$\left|c(k) - c\left(k'\right)\right| \leq \varphi_x^{\bar{c}}\left(d\left(k, k'\right)\right) \tag{2.29}$$

y

$$\left\|Q(\cdot|k) - Q\left(\cdot|k'\right)\right\|_{WT} \leq \varphi_x^Q\left(d\left(k, k'\right)\right). \tag{2.30}$$

Observacin 2.4 (a) *Como en [8] y [10], es posible demostrar que si los conjuntos* $A(x)$ *no dependen de* $x \in \mathbb{X}$, *es decir,* $A(x) \equiv \mathbb{A}$ *para cada* $x \in \mathbb{X}$, *y* $\mathbb{A}$ *es compacto, entonces la condición (2.30) se puede sustituir por la siguiente: Para todo* $x \in \mathbb{X}$ *existe una función* $\varphi_x^Q \in \mathcal{G}$ *tal que para cada* $x' \in \mathbb{X}$,

$$\sup_{\mathbb{A}} \left\|Q(\cdot|x, a) - Q\left(\cdot|x', a\right)\right\|_{WT} \leq \varphi_x^Q\left(d\left(x, x'\right)\right). \tag{2.31}$$

(b) Además (véase [8]), las Hipótesis 2.6-2.7 implican

$$\sup_{n \in \mathbb{N}} \|h_n\|_W < \infty \qquad y \qquad \sup_{n \in \mathbb{N}} \|v_n\|_W < \infty. \tag{2.32}$$

La convergencia del algoritmo de iteración de valores se establece en el siguiente resultado (véase [8, Teorema 2.6]).

Teorema 2.4 *Si las Hipótesis 2.6-2.8 se cumplen, y* f^* *(véase Teorema 2.3(b)) satisface que*

$$q_{f^*}(U) > 0 \tag{2.33}$$

para cada conjunto abierto no vacío $U \subset \mathbb{X}$, *entonces*

(a) $\lim_{n \to \infty} |j_n - j| = 0$, *y*

(b) $\lim_{n \to \infty} \left|h_n(x) - \bar{h}(x)\right| = 0$ *donde la convergencia es uniforme sobre subconjuntos compactos de* $\mathbb{X}$.

2.4. Construcción de políticas CP-óptimas

Sea $\bar{m}$ una constante positiva tal que $\bar{\tau} \geq \bar{m}$ y $\bar{\tau}_n$ un estimador de $\bar{\tau}$ tal que

$$\bar{\tau}_n \geq \bar{m} \tag{2.34}$$

y

$$E_x^{\pi}\left[|\bar{\tau}_n - \bar{\tau}|\right] = O\left(n^{-\gamma_0}\right) \tag{2.35}$$

para alguna constante positiva γ_0, bajo la política $\pi \in \Pi$ y estado inicial $x \in \mathbb{X}$.

Además, para $n \in \mathbb{N}$ definimos

$$\bar{c}_n(x,a) := \frac{c(x,a)}{\bar{\tau}_n}, \tag{2.36}$$

y para $t \in \mathbb{N}$ fijo, sea

$$J_n^{(\bar{\tau}_t)}(x,\pi) := \sum_{k=0}^{n-1} E_x^{\pi}\left[\bar{c}_t(x_k,a_k)\right],$$

el costo esperado en n etapas (véase (2.25)) bajo la política $\pi \in \Pi$ y estado inicial $x \in \mathbb{X}$, cuando usamos $\bar{\tau}_t$ en lugar de $\bar{\tau}$. Asimismo, sea

$$v_n^{\bar{\tau}_t}(x) := \inf_{\pi \in \Pi} J_n^{(\bar{\tau}_t)}(x,\pi), \quad x \in \mathbb{X},$$

la correspondiente función de valor óptimo. Definimos también $h_n^{\bar{\tau}_t}(\cdot)$ y $j_n^{\bar{\tau}_t}$, de modo similar a (2.27) y (2.28), respectivamente; esto es

$$h_n^{\bar{\tau}_t}(x) := v_n^{\bar{\tau}_t}(x) - v_n^{\bar{\tau}_t}(z), \quad x \in \mathbb{X},$$

y

$$j_n^{\bar{\tau}_t} := v_n^{\bar{\tau}_t}(z) - v_{n-1}^{\bar{\tau}_t}(z).$$

De aquí, y usando (2.26) y (2.36) nótese que

$$v_n^{\bar{\tau}_t}(x) = \min_{a \in A(x)} \left\{ \bar{c}_t(x,a) + \int_{\mathbb{X}} v_{n-1}^{\bar{\tau}_t}(y)\, Q(dy \mid x,a) \right\}, \quad x \in \mathbb{X},\ n \in \mathbb{N}, \tag{2.37}$$

donde $v_0^{\bar{\tau}_t}(x) \equiv 0$. Más aún,

$$h_n^{\bar{\tau}_t}(x) + j_n^{\bar{\tau}_t} = \min_{a \in A(x)} \left\{ \bar{c}_t(x,a) + \int_{\mathbb{X}} h_{n-1}^{\bar{\tau}_t}(y)\, Q(dy \mid x,a) \right\}.$$

Por otra parte, sea ν un número arbitrario tal que $0 < \nu < \gamma_0$, y sea

$$\hat{D} := \beta + b + 1,$$

donde, $\beta := (\beta_0)^{1/p}$ y $b := (b_0)^{1/p}$ (véase Observación 2.3). Definamos ahora la sucesión

$$n_t := \left[\nu \log_{\hat{D}} t\right], \tag{2.38}$$

donde $[x]$ es la parte entera de x. Entonces, de (2.37), para $n_t \geq 1$ obtenemos

$$h^{\bar{\tau}_t}_{n_t}(x) + j^{\bar{\tau}_t}_{n_t} = \min_{a \in A(x)} \left\{ \bar{c}_t(x,a) + \int_{\mathbb{X}} h^{\bar{\tau}_t}_{n_t - 1}(y)\, Q(dy \mid x, a) \right\}.$$

Aplicando argumentos estándar sobre la existencia de selectores (véase, por ejemplo, [43]), tenemos que para cada $t \in \mathbb{N}$, existe $f_t := f_{n_t}$ tal que

$$h^{\bar{\tau}_t}_{n_t}(x) + j^{\bar{\tau}_t}_{n_t} = \bar{c}_t(x, f_t) + \int_{\mathbb{X}} h^{\bar{\tau}_t}_{n_t - 1}(y)\, Q(dy \mid x, f_t), \quad x \in \mathbb{X}. \tag{2.39}$$

Finalmente, establecemos nuestro resultado principal de la siguiente manera.

Teorema 2.5 *Sea $\hat{\pi} := \{\hat{\pi}_t\}$ la política determinada por los selectores f_t en (2.39). Esto es,*

$$\hat{\pi}_t(B \mid i_t) = 1_B(f_t(x_t)) \quad \forall B \in \mathcal{B}(\mathbb{A}), i_t \in \mathbb{H}_t, t \in \mathbb{N},$$

y f_0 un selector fijo. Entonces, bajo las Hipótesis 2.5-2.8, $\hat{\pi}$ es una política CP-óptima para el MCSM (1.1); es decir,

$$j = J(x, \hat{\pi}).$$

2.4.1. Resultados preliminares

La demostración del Teorema 2.5 es consecuencia de los lemas siguientes.

Lema 2.1 *Supóngase que las Hipótesis 2.5-2.6 se cumplen. Entonces, para cada $x \in \mathbb{X}$ y $\pi \in \Pi$:*

(a) $E^{\pi}_x \left[\left| \frac{1}{\bar{\tau}_n} - \frac{1}{\bar{\tau}} \right| \right] = O(n^{-\gamma_0})$; *y*

(b) $E^{\pi}_x \left[\sup_{a \in A(x)} |\bar{c}_n(x,a) - \bar{c}(x,a)| \right] \to 0$, *cuando* $n \to \infty$.

Demostración. *(a)* Dado que $\bar{m} \leq \bar{\tau}_n$ y $\bar{m} \leq \bar{\tau}$ (véase (2.34)), tenemos que

$$0 < \frac{1}{\bar{\tau}\bar{\tau}_n} \leq \frac{1}{\bar{m}^2}.$$

Entonces, de (2.35), para cada $x \in \mathbb{X}$ y $\pi \in \Pi$ se sigue que

$$\begin{aligned} E_x^\pi \left[\left|\frac{1}{\bar{\tau}_n} - \frac{1}{\bar{\tau}}\right|\right] = E_x^\pi \left[\left|\frac{\bar{\tau} - \bar{\tau}_n}{\bar{\tau}\bar{\tau}_n}\right|\right] &\leq \frac{1}{\bar{m}^2} E_x^\pi \left[|\bar{\tau} - \bar{\tau}_n|\right] \\ &= O\left(n^{-\gamma_0}\right) \quad \text{cuando } n \to \infty. \end{aligned}$$

(b) De (2.17), (2.36) y la Hipótesis 2.6, para cada $x \in \mathbb{X}$ y $\pi \in \Pi$ se tiene

$$\begin{aligned} E_x^\pi \left[\sup_{a \in A(x)} |\bar{c}_n(x,a) - \bar{c}(x,a)|\right] &= E_x^\pi \left[\sup_{a \in A(x)} \left|\frac{c(x,a)}{\bar{\tau}_n} - \frac{c(x,a)}{\bar{\tau}}\right|\right] \\ &\leq W(x) E_x^\pi \left[\left|\frac{1}{\bar{\tau}_n} - \frac{1}{\bar{\tau}}\right|\right]; \end{aligned}$$

y entonces, *(b)* se sigue de *(a)*.

■

Definamos la función $\check{\Phi} : \mathbb{K}_A \to \mathbb{R}$ como

$$\check{\Phi}(x,a) := \bar{c}(x,a) + \int_{\mathbb{X}} \bar{h}(y)\, Q(dy \mid x,a) - \bar{h}(x) - j, \tag{2.40}$$

y observe que, de (2.24),

$$\begin{aligned} \min_{a \in A(x)} \check{\Phi}(x,a) = 0, \qquad & \text{y} \\ \check{\Phi}(x,a) \geq 0 \qquad & \forall\, (x,a) \in \mathbb{K}_A. \end{aligned}$$

A continuación introducimos un resultado que caracteriza la optimalidad promedio en términos de la función $\check{\Phi}$, y cuya demostración puede consultarse en [14, 15].

Lema 2.2 *Una política $\pi \in \Pi$ es CP-óptima para el MCSM (1.1) si para cada $x \in \mathbb{X}$*

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

Lema 2.3 *Bajo las Hipótesis 2.6-2.7, para cada* $\pi \in \Pi$ *y* $x \in \mathbb{X}$,

$$E_x^{\pi}\left[\left\|v_{n_t}^{\bar{\tau}_t} - v_{n_t}\right\|_W\right] \to 0 \quad \textit{cuando} \quad t \to \infty.$$

Demostración. De las expresiones (1.16), (2.22), (2.26) y (2.37), para cada $x \in \mathbb{X}$, $t \in \mathbb{N}$,

$$\begin{aligned}\left|v_{n_t}^{\bar{\tau}_t}(x) - v_{n_t}(x)\right| &\leq \sup_{a \in A(x)} \left\{ |c(x,a)| \left|\frac{1}{\bar{\tau}_t} - \frac{1}{\bar{\tau}}\right| \right. \\ &\qquad \left. + \int_{\mathbb{X}} \left|v_{n_t-1}^{\bar{\tau}_t}(y) - v_{n_t}(y)\right| Q(dy \mid x,a) \right\} \\ &\leq W(x) \left|\frac{1}{\bar{\tau}_t} - \frac{1}{\bar{\tau}}\right| + \left\|v_{n_t-1}^{\bar{\tau}_t} - v_{n_t}\right\|_W \{\beta W(x) + b\}.\end{aligned}$$

Por lo tanto, dado que $W(\cdot) \geq 1$,

$$\frac{\left|v_{n_t}^{\bar{\tau}_t}(x) - v_{n_t}(x)\right|}{W(x)} \leq \left|\frac{1}{\bar{\tau}_t} - \frac{1}{\bar{\tau}}\right| + \left\|v_{n_t-1}^{\bar{\tau}_t} - v_{n_t}\right\|_W (\beta + b).$$

De aquí obtenemos

$$\left\|v_{n_t}^{\bar{\tau}_t} - v_{n_t}\right\|_W \leq \left|\frac{1}{\bar{\tau}_t} - \frac{1}{\bar{\tau}}\right| + \hat{D}\left\|v_{n_t-1}^{\bar{\tau}_t} - v_{n_t}\right\|_W.$$

Iterando esta desigualdad, un cálculo directo muestra que

$$\left\|v_{n_t}^{\bar{\tau}_t} - v_{n_t}\right\|_W \leq \left|\frac{1}{\bar{\tau}_t} - \frac{1}{\bar{\tau}}\right| \sum_{k=0}^{n_t-1} \hat{D}^k = \left|\frac{1}{\bar{\tau}_t} - \frac{1}{\bar{\tau}}\right| \frac{1 - \hat{D}^{n_t}}{1 - \hat{D}}.$$

De modo que, de acuerdo a la definición de n_t en (2.38),

$$\begin{aligned}\left\|v_{n_t}^{\bar{\tau}_t} - v_{n_t}\right\|_W &\leq \left|\frac{1}{\bar{\tau}_t} - \frac{1}{\bar{\tau}}\right| \frac{1 - \hat{D}^{\left[\nu \log_{\hat{D}} t\right]}}{1 - \hat{D}} \\ &= \left|\frac{1}{\bar{\tau}_t} - \frac{1}{\bar{\tau}}\right| O(t^{\nu}) \quad \text{cuando} \quad t \to \infty.\end{aligned}$$

Finalmente, del Lema 2.1*(a)* y el hecho que $0 < \nu < \gamma_0$, para cada $\pi \in \Pi$ y $x \in \mathbb{X}$, obtenemos

$$E_x^{\pi}\left[\left\|v_{n_t}^{\bar{\tau}_t} - v_{n_t}\right\|_W\right] = O\left(t^{-\gamma_0}\right) O(t^{\nu}) \to 0 \quad \text{cuando} \quad t \to \infty.$$

■

Observacin 2.5 *Dado que (véase (2.27))*

$$\left\|h_{n_t}^{\bar{\tau}_t} - h_{n_t}\right\|_W \leq 2\left\|v_{n_t}^{\bar{\tau}_t} - v_{n_t}\right\|_W, \tag{2.41}$$

tenemos

$$E_x^\pi\left[\left\|h_{n_t}^{\bar{\tau}_t} - h_{n_t}\right\|_W\right] \to 0 \quad \textit{cuando} \quad t \to \infty. \tag{2.42}$$

Lema 2.4 *Sea $\{\psi_t\}$ una sucesión de variables aleatorias que satisfacen las condiciones siguientes:*

C.1 $\sup_{t\in\mathbb{N}} \psi_t = M_1 < \infty$;

C.2 $E_x^\pi[\psi_t] \to 0$ *cuando* $t \to \infty$, *para cada* $x \in \mathbb{X}$ *y* $\pi \in \Pi$.

Entonces

$$E_x^\pi[W(x_t)\,\psi_t] \to 0 \quad \textit{cuando} \quad t \to \infty. \tag{2.43}$$

Demostración. De *C.2* tenemos la convergencia en probabilidad

$$\psi_t \overset{P_x^\pi}{\to} 0 \quad \text{cuando} \quad t \to \infty. \tag{2.44}$$

Además, de *C.1* y la relación (1.18)

$$\sup_{t\in\mathbb{N}} E_x^\pi[W(x_t)\,\psi_t]^p \leq M_1^p \sup_{t\in\mathbb{N}} E_x^\pi[W^p(x_t)] < \infty.$$

Esto implica (véase, por ejemplo, [1, Lema 7.6.9]) que $\{W(x_t)\,\psi_t\}$ es P_x^π-uniformemente integrable. Por otra parte, para números positivos l_1 y l_2, se tiene para cada $x \in \mathbb{X}$ y $\pi \in \Pi$,

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

De modo que, por la Desigualdad de Chebyshev:

$$P_x^\pi[W(x_t)\,\psi_t > l_1] \leq P_x^\pi\left[\psi_t > \frac{l_1}{l_2}\right] + \frac{E_x^\pi[W(x_t)]}{l_2}, \quad t \in \mathbb{N}, \tag{2.45}$$

para cada $x \in \mathbb{X}$ y $\pi \in \Pi$. Ahora, (2.45) junto con (2.44) y la relación (1.20), implican que $\{W(x_t)\,\psi_t\}$ converge a cero en probabilidad puesto que l_2 es arbitrario, es decir,

$$W(x_t)\,\psi_t \overset{P_x^\pi}{\to} 0 \quad \text{cuando} \quad t \to \infty. \tag{2.46}$$

Finalmente, (2.43) se sigue de (2.46) y el hecho de que $\{W(x_t)\,\psi_t\}$ es P_x^π-uniformemente integrable.

■

Lema 2.5 *Bajo las Hipótesis 2.5-2.8, para cada $x \in \mathbb{X}$ y $\pi \in \Pi$,*

(a) $E_x^\pi \left[\left|j_{n_t}^{\bar{\tau}_t} - j\right|\right] \to 0 \quad cuando \quad t \to \infty.$

(b) $E_x^\pi \left[\left|h_{n_t}^{\bar{\tau}_t}(x_t) - \bar{h}(x_t)\right|\right] \to 0 \quad cuando \quad t \to \infty.$

Demostración. *(a)* Sea $z \in \mathbb{X}$ el estado fijo en (2.27) y (2.28). Luego, sumando y restando el término j_{n_t}, y usando además, las definiciones de j_n y $j_n^{\bar{\tau}_t}$ tenemos que para cada $x \in \mathbb{X}$ y $\pi \in \Pi$,

$$\begin{aligned} E_x^\pi \left[\left|j_{n_t}^{\bar{\tau}_t} - j\right|\right] &\leq E_x^\pi \left[\left|j_{n_t}^{\bar{\tau}_t} - j_{n_t}\right| + \left|j_{n_t} - j\right|\right] \\ &\leq W(z)\, E_x^\pi \left[\left\|v_{n_t}^{\bar{\tau}_t} - v_{n_t}\right\|_W\right] \\ &\quad + W(z)\, E_x^\pi \left[\left\|v_{n_t-1}^{\bar{\tau}_t} - v_{n_t-1}\right\|_W\right] + \left|j_{n_t} - j\right|. \end{aligned}$$

De aquí, el Lema 2.3 y el Teorema 2.4 implican

$$E_x^\pi \left[\left|j_{n_t}^{\bar{\tau}_t} - j\right|\right] \to 0 \quad \text{cuando} \quad t \to \infty,$$

lo cual demuestra la parte *(a)*.

(b) Note que, para cada $t \in \mathbb{N}$,

$$\begin{aligned} \left|h_{n_t}^{\bar{\tau}_t}(x_t) - \bar{h}(x_t)\right| &\leq \left|h_{n_t}^{\bar{\tau}_t}(x_t) - h_{n_t}(x_t)\right| + \left|h_{n_t}(x_t) - \bar{h}(x_t)\right| \\ &\leq \left\|h_{n_t}^{\bar{\tau}_t} - h_{n_t}\right\|_W W(x_t) + \left|h_{n_t}(x_t) - \bar{h}(x_t)\right|. \end{aligned}$$

Por otra parte, (2.32) y la Observación 2.5 implican que las condiciones *C.1* y *C.2* del Lema 2.4 se satisfacen con $\psi_t := \left\|h_{n_t}^{\bar{\tau}_t} - h_{n_t}\right\|_W$. Por lo tanto,

$$E_x^\pi \left[\left\|h_{n_t}^{\bar{\tau}_t} - h_{n_t}\right\|_W W(x_t)\right] \to 0 \quad \text{cuando} \quad t \to \infty. \tag{2.47}$$

Ahora, sea $\{K_n\}$ la sucesión de conjuntos en la Observación 2.2 (véase Hipótesis 2.6) y $\bar{\delta} > 0$ una constante fija pero arbitraria. Para cada $n \in \mathbb{N}$ tenemos que

$$\begin{aligned} E_x^\pi \left[\left|h_{n_t}^{\bar{\tau}_t}(x_t) - \bar{h}(x_t)\right|\right] &= E_x^\pi \left[\left|h_{n_t}^{\bar{\tau}_t}(x_t) - \bar{h}(x_t)\right| 1_{K_n}(x_t)\right] \\ &\quad + E_x^\pi \left[\left|h_{n_t}^{\bar{\tau}_t}(x_t) - \bar{h}(x_t)\right| 1_{K_n^c}(x_t)\right] \\ &=: I_{1,n}(t) + I_{2,n}(t). \end{aligned}$$

Además, de (2.32),

$$\left|h_{n_t}(x) - \bar{h}(x)\right| \leq \tilde{C} W(x), \quad t \in \mathbb{N},\ x \in \mathbb{X},$$

para alguna constante $\tilde{C} < \infty$. De aquí, la relación (1.18) junto con la Observación 2.2 implican que

$$I_{2,n}(t) \leq \tilde{C} E_x^{\pi}\left[W^p(x_t)\, W^{1-p}(x_t)\, 1_{K_n^c}(x_t)\right] \leq \tilde{C} E_x^{\pi}\left[W^p(x_t) \sup_{x \notin K_n} W^{1-p}(x)\right]$$
$$\leq \tilde{C} \sup_{x \notin K_n} W^{1-p}(x) \sup_{t \in \mathbb{N}} E_x^{\pi}\left[W^p(x_t)\right] \leq \bar{\delta}/2, \tag{2.48}$$

para todo $t \in \mathbb{N}$ y algún número n suficientemente grande.

Luego, fijando algún n de acuerdo con (2.48), del Teorema 2.4*(b)* vemos que

$$I_{1,n}(t) \leq \sup_{x \in K_n} \left|h_{n_t}(x) - \bar{h}(x)\right| \leq \bar{\delta}/2, \tag{2.49}$$

para t suficientemente grande. Finalmente, combinando (2.48) y (2.49) obtenemos el resultado deseado.

■

2.4.2. Demostración del Teorema 2.5

Para cada $t \in \mathbb{N}$ y $(x, a) \in \mathbb{K}_A$ definimos la función $\check{\Phi}_{n_t}^{\bar{\tau}_t}$ como

$$\check{\Phi}_{n_t}^{\bar{\tau}_t}(x, a) := \bar{c}(x, a) + \int_{\mathbb{X}} h_{n_t-1}^{\bar{\tau}_t}(y)\, Q(dy \mid x, a) - h_{n_t}^{\bar{\tau}_t}(x) - j_{n_t}^{\bar{\tau}_t}. \tag{2.50}$$

Luego, de la definición de los minimizadores f_t (véase (2.39)) tenemos

$$\check{\Phi}_{n_t}^{\bar{\tau}_t}(\cdot, f_t(\cdot)) = 0, \quad t \in \mathbb{N}.$$

De aquí,

$$E_x^{\hat{\pi}}\left[\check{\Phi}(x_t, a_t)\right] = E_x^{\hat{\pi}}\left[\left|\check{\Phi}(x_t, a_t) - \check{\Phi}_{n_t}^{\bar{\tau}_t}(x_t, a_t)\right|\right], \tag{2.51}$$

y por lo tanto, de acuerdo al Lema 2.2, es suficiente demostrar la convergencia a cero del lado derecho de (2.51).

Sea $\{(x_t, a_t)\}$ una sucesión de pares estado-acción correspondiente a la aplicación de la política $\hat{\pi}$. Entonces, sumando y restando el término

$$\int_{\mathbb{X}} h_{n_t-1}(y)\, Q(dy \mid x_t, a_t),$$

y por (2.40) y (2.50) obtenemos

$$\begin{aligned}
&\left|\check{\Phi}\left(x_t,a_t\right)-\check{\Phi}_{n_t}^{\bar{\tau}_t}\left(x_t,a_t\right)\right| \\
&\leq\left|\int_{\mathbb{X}}\bar{h}\left(y\right)Q(dy\mid x_t,a_t)-\int_{\mathbb{X}}h_{n_t-1}\left(y\right)Q(dy\mid x_t,a_t)\right| \\
&\quad+\left|\int_{\mathbb{X}}h_{n_t-1}\left(y\right)Q(dy\mid x_t,a_t)-\int_{\mathbb{X}}h_{n_t-1}^{\bar{\tau}_t}\left(y\right)Q(dy\mid x_t,a_t)\right| \\
&\qquad+\left|h_{n_t-1}^{\bar{\tau}_t}\left(x_t\right)-\bar{h}\left(x_t\right)\right|+\left|j_{n_t-1}^{\bar{\tau}_t}-j\right| \qquad (2.52)
\end{aligned}$$

Del Lema 2.5 tenemos que

$$E_x^{\hat{\pi}}\left[\left|h_{n_t-1}^{\bar{\tau}_t}\left(x_t\right)-\bar{h}\left(x_t\right)\right|\right]\to 0, \qquad (2.53)$$

y

$$E_x^{\hat{\pi}}\left[\left|j_{n_t-1}^{\bar{\tau}_t}-j\right|\right]\to 0 \quad \text{cuando} \quad t\to\infty. \qquad (2.54)$$

Por otra parte, nótese que cuando $t\to\infty$,

$$\begin{aligned}
&E_x^{\hat{\pi}}\left[\left|\int_{\mathbb{X}}\bar{h}\left(y\right)Q(dy\mid x_t,a_t)-\int_{\mathbb{X}}h_{n_t-1}\left(y\right)Q(dy\mid x_t,a_t)\right|\right] \\
&\leq E_x^{\hat{\pi}}\left[\int_{\mathbb{X}}\left|\bar{h}\left(y\right)-h_{n_t-1}\left(y\right)\right|Q(dy\mid x_t,a_t)\right] \\
&= E_x^{\hat{\pi}}\left[E_x^{\hat{\pi}}\left[\left|\left\{\bar{h}\left(x_{t+1}\right)-h_{n_t-1}\left(x_{t+1}\right)\right\}\right|\mid x_t,a_t\right]\right] \\
&= E_x^{\hat{\pi}}\left[\left|\bar{h}\left(x_{t+1}\right)-h_{n_t-1}\left(x_{t+1}\right)\right|\right]\to 0. \qquad (2.55)
\end{aligned}$$

Además, de (1.16) y (2.22) tenemos

$$\begin{aligned}
&\left|\int_{\mathbb{X}}h_{n_t-1}\left(y\right)Q(dy\mid x_t,a_t)-\int_{\mathbb{X}}h_{n_t-1}^{\bar{\tau}_t}\left(y\right)Q(dy\mid x_t,a_t)\right| \\
&\leq\int_{\mathbb{X}}\left|h_{n_t-1}\left(y\right)-h_{n_t-1}^{\bar{\tau}_t}\left(y\right)\right|Q(dy\mid x_t,a_t) \\
&\leq\left\|h_{n_t}^{\bar{\tau}_t}-h_{n_t}\right\|_W\left\{\beta W\left(x\right)+b\right\}, \qquad (2.56)
\end{aligned}$$

lo cual, junto con (2.42) y (2.47) implican

$$\lim_{t\to\infty}E_x^{\hat{\pi}}\left[\left\|h_{n_t}^{\bar{\tau}_t}-h_{n_t}\right\|_W\left\{\beta W\left(x_t\right)+b\right\}\right]=0. \qquad (2.57)$$

Finalmente combinando las expresiones (2.51)-(2.57) obtenemos

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

lo cual demuestra que, en efecto, $\hat{\pi}$ es una política CP-óptima para el $MCSM$ (1.1). (véase Lema 2.2).

■

2.5. Ejemplo: un sistema de almacenamiento

Para ilustrar la teoría desarrollada en este capítulo, presentamos un ejemplo de un sistema de almacenamiento. En particular, mostraremos que se satisfacen las Hipótesis 2.5, 2.6, 2.7 y 2.8, que se impusieron en el modelo de control correspondiente al caso promedio. Como se podrá observar, la verificación del cumplimiento de las hipótesis para el caso descontado es similar.

Consideremos un sistema de almacenamiento cuyas entradas están controladas de tal manera que, justo al tiempo en que se acumula cierta cantidad de producto $K^* > 0$ para su admisión al sistema, el controlador elige un valor $a \in [a_*, 1] =: \mathbb{A}$ $(0 < a_* < 1)$, que representa la porción de K^* que habrá de ser admitida. Es decir, aK^* será la cantidad de producto admitido en el sistema. Además, supongamos que existe una tasa de consumo constante $\hat{b} > 0$ (fija) por unidad de tiempo. Entonces, si $x = x_n \in \mathbb{X} := [0, \infty)$ representa la cantidad de producto en el sistema, y $a = a_n \in A\left(x\right)$ las decisiones tomadas por el controlador al tiempo de la n-ésima época de decisión T_n, y además, definiendo el tiempo de permanencia del sistema en el estado x_n como la variable aleatoria $\delta_{n+1} := T_{n+1} - T_n$ $(n \in \mathbb{N}_0)$, el sistema evoluciona de acuerdo a la siguiente ecuación en diferencias

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

De la descripción previa, claramente las épocas de decisión solo dependen del tiempo en que se acumula la cantidad de producto K^*, y no de las variables estado-acción, por lo cual de manera natural tenemos que la distribución de las variables aleatorias δ_{n+1} es independiente de (x_n, a_n).

Supongamos que $\{\delta_n\}$ es una sucesión de variables aleatorias i.i.d. con densidad acotada, continua y positiva g. Consideraremos además la condición siguiente:

Hiptesis 2.9

$$E\left[\delta\right] > \frac{K^*}{\hat{b}}.$$

Lema 2.6 *Bajo la Hipótesis 2.9, existen constantes* λ_0 *y* $p > 1$ *tal que*

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

donde Υ *es la función generadora de momentos de la variable aleatoria*

$$K^* - \hat{b}\delta,$$

es decir

$$\Upsilon(t) = E\left[\exp\left\{t\left(K^* - \hat{b}\delta\right)\right\}\right].$$

Demostración. La Hipótesis 2.9 implica que $\Upsilon'(0) < 0$, y dado que $\Upsilon(0) = 1$, existe $\lambda_0 > 0$ tal que $\Upsilon(\lambda_0) < 1$. Además, debido a la continuidad de Υ, podemos elegir $p > 1$ tal que

$$\Upsilon(p\lambda_0) = E\left[\exp\left\{p\lambda_0\left(K^* - \hat{b}\delta\right)\right\}\right] < 1. \tag{2.58}$$

■

Supongamos ahora que la función de costo c es una función arbitraria s.c.i. en $\mathbb{K}_A$ que satisface (2.29) de la Hipótesis 2.8, y tal que para alguna constante $\bar{K}^* \geq 1$ y para cada $x \in \mathbb{X}$,

$$\sup_{a \in A(x)} |c(x,a)| \leq \bar{K}^* \exp(\lambda_0 x). \tag{2.59}$$

Teorema 2.6 *La Hipótesis 2.9 implica la Hipótesis 2.6.*

Demostración. De la descripción del ejemplo, claramente se cumplen las partes *(a)* y *(b)* de la Hipótesis 2.6, mientras que la desigualdad (2.18) de la parte *(d)* se satisface si definimos

$$W(x) := \bar{K}^* \exp(\lambda_0 x), \quad x \in \mathbb{X}, \tag{2.60}$$

con λ_0 como en el Lema 2.6.

Para verificar la parte *(c)* de la Hipótesis 2.6, sea v una función medible acotada sobre $\mathbb{X}$, y para cada $a \in A(x)$, sea g_a la densidad de $aK^* - \hat{b}\delta$. Obsérvese que

$$g_a(y) = \frac{1}{\hat{b}} g\left(\frac{aK^* - y}{\hat{b}}\right), \quad -\infty < y \leq aK^*. \tag{2.61}$$

Además, para cada $y \in \mathbb{R}$, $a \mapsto g_a(y)$ es continua sobre $A(x)$. Entonces,

$$\begin{aligned}\int_{\mathbb{X}} v(y)\, Q(dy \mid x, a) &= \int_0^{\infty} v\left((x+y)^+\right) g_a(y)\, dy \\ &= v(0) \int_{-\infty}^{-x} g_a(y)\, dy + \int_{-x}^{\infty} v(x+y)\, g_a(y)\, dy \\ &= v(0) \int_{-\infty}^{-x} g_a(y)\, dy + \int_0^{\infty} v(y)\, g_a(y-x)\, dy.\end{aligned}$$

De modo que, por el Teorema de Scheffé,

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

define una función continua, lo cual demuestra la Hipótesis 2.6*(c)*.

Finalmente, reemplazando $v(\cdot)$ por $W(\cdot)$ y usando argumentos similares, obtenemos que (2.19) se satisface.

■

Ahora verificaremos la Hipótesis 2.7. Para tal fin, sea $f \in \mathbb{F}$ una política estacionaria y definamos

$$\phi_f(x) := P\left\{x + f(x) K^* - \hat{b}\delta \leq 0\right\}, \quad x \in \mathbb{X}, \tag{2.62}$$

y $\mu(\cdot) := p_0$, donde p_0 es la medida de Dirac concentrada en $x = 0$. Luego, para $C \in \mathcal{B}([0, \infty))$

$$\begin{aligned}Q(C \mid x, f) &= \int_{\mathbb{R}} 1_C \left(x + f(x) K^* - \hat{b}\delta\right)^+ g(s)\, ds \\ &= P\left\{\left(x + f(x) K^* - \hat{b}\delta\right)^+ \in C\right\} \\ &\geq P\left\{x + f(x) K^* - \hat{b}\delta \leq 0\right\} \mu(C) = \phi_f(x)\, \mu(C),\end{aligned}$$

lo cual implica que la Hipótesis 2.7*(a)* se cumple.

Por otra parte, sea $p > 1$ como en el Lema 2.6. Entonces, para $x \in \mathbb{X}$ y $f \in \mathbb{F}$,

$$\begin{aligned}\int_{\mathbb{X}} W^p(y)\, Q(dy \mid x, f) &= \int_{\mathbb{X}} \left(\bar{K}^* \exp(\lambda_0 y)\right)^p Q(dy \mid x, f) \\ &= \left(\bar{K}^*\right)^p \int_0^{\infty} \exp\left(\lambda_0 p \left(x + f(x) K^* - \hat{b}s\right)^+\right) g(s)\, ds \\ &\leq \left(\bar{K}^*\right)^p P\left\{x + f(x) K^* - \hat{b}\delta \leq 0\right\} \\ &\quad + \left(\bar{K}^* \exp(\lambda_0 x)\right)^p \int_0^{\infty} \exp\left(\lambda_0 p \left(K^* - \hat{b}s\right)\right) g(s)\, ds,\end{aligned}$$

debido a que $f(x) \leq 1$ para todo $x \in \mathbb{X}$. De aquí,

$$\int_{\mathbb{X}} W^p(y)\, Q(dy \mid x, f) \leq \left(\bar{K}^*\right)^p \phi_f(x) + W^p(x)\, \beta_0, \tag{2.63}$$

donde $\beta_0 := \Upsilon(p\lambda_0) < 1$ (véase Lema 2.6). Ahora, observando que

$$\left(\bar{K}^*\right)^p = \left(\bar{K}^* \exp(\lambda_0(0))\right)^p = W^p(0) = \int_{\mathbb{X}} W^p(y)\, Q(dy \mid x, f),$$

se obtiene que la relación (2.63) implica la parte *(b)* de la Hipótesis 2.7.

Y finalmente, la parte *(c)* de la Hipótesis 2.7 es una consecuencia de las siguientes relaciones:

$$\begin{aligned}
\inf_{f\in\mathbb{F}} \int_{\mathbb{X}} \phi_f(y)\, \mu(dy) &= \inf_{f\in\mathbb{F}} \int_{\mathbb{X}} P\left\{y + f(y)\, K^* - \hat{b}\delta \leq 0\right\} \mu(dy) \\
&= \inf_{f\in\mathbb{F}} P\left\{0 + f(0)\, K^* - \hat{b}\delta \leq 0\right\} \\
&\geq P\left\{K^* - \hat{b}\delta \leq 0\right\} > 0,
\end{aligned}$$

donde la última desigualdad se debe a que

$$E\left[K^* - \hat{b}\delta\right] < 0,$$

(véase Hipótesis 2.9). Por consiguiente, la Hipótesis 2.7 se satisface.

De acuerdo a la Observación 2.4*(a)*, para demostrar que la Hipótesis 2.8 se cumple, solo resta comprobar que la relación (2.31) se satisface. Para tal propósito, requeriremos las condiciones que se establecen a continuación.

Hiptesis 2.10 *(a) Para cada $a \in \mathbb{A}$ existe un número $\theta' > 0$ tal que*

$$\int_{-\infty}^{\infty} \exp(\lambda_0 y)\, \rho_{a,\theta'}(y)\, dy < \infty,$$

donde

$$\rho_{a,\theta'}(y) := \sup\left\{g_{a'}(y) : a' \in \mathbb{A} \ y \ \left|a' - a\right| \leq \theta'\right\}, \quad y \in \mathbb{R}.$$

(b) Existen funciones $\varphi \in \mathcal{G}$ y $\phi_a(z)$, $z \in \mathbb{R}$, $a \in \mathbb{A}$, tal que

$$C_* := \sup_{\mathbb{A}} \int_{-\infty}^{\infty} \exp(\lambda_0 z)\, \phi_a(z)\, dz < \infty$$

y

$$|g_a(z+y) - g_a(z)| \leq \varphi(|y|)\, \phi_a(z)$$

para cada z,$y \in \mathbb{R}$ y $a \in \mathbb{A}$.

Teorema 2.7 *La Hipótesis 2.10 implica la relación (2.31).*

Demostración. Primero obsérvese que (2.61) implica

$$0 \leq g_a(y) \leq \xi, \quad y \in \mathbb{R}, a \in \mathbb{A},$$

donde ξ es una cota de g. Además, si denotamos por H_a, $a \in \mathbb{A}$, a la función de distribución de $aK^* - \hat{b}\delta$, se puede ver que para x,$x' \in \mathbb{R}$,

$$\left|H_a(x) - H_a\left(x'\right)\right| \leq \xi \left|x - x'\right|.$$

Ahora, usando el hecho de que

$$Q(C \mid x, a) = \int_{\mathbb{R}} 1_C(y)\, g_a(y - s)\, dy,$$

entonces, bajo la Hipótesis 2.10 tenemos lo siguiente:

$$\begin{aligned}
&\left\|Q(\cdot|x,a) - Q\left(\cdot|x',a\right)\right\|_{WT} \\
&\quad = \int_{\mathbb{X}} W(y) \left|Q(dy \mid x, a) - Q\left(dy \mid x', a\right)\right| \\
&\qquad \leq W(0) \left|P\left\{x + aK^* - \hat{b}\delta \leq 0\right\} - P\left\{x' + aK^* - \hat{b}\delta \leq 0\right\}\right| \\
&\qquad\qquad + \bar{K}^* \int_0^\infty \exp(\lambda_0 y) \left|g_a(y - x) - g_a\left(y - x'\right)\right| dy \\
&\qquad = W(0) \left|H_a(-x) - H_a\left(-x'\right)\right| \\
&\qquad\qquad + \bar{K}^* \int_{-x}^\infty \exp(\lambda_0 (y + x)) \left|g_a(y) - g_a\left(y + x - x'\right)\right| dy \\
&\qquad\qquad \leq W(0)\, \xi \left|x' - x\right| \\
&\qquad\qquad\qquad + \bar{K}^* \exp(\lambda_0 x)\, \varphi\left(\left|x' - x\right|\right) \int_{-\infty}^\infty \exp(\lambda_0 y)\, \phi_a(y)\, dy\ ;
\end{aligned}$$

es decir,

$$\begin{aligned}
&\left\|Q(\cdot|x,a) - Q\left(\cdot|x',a\right)\right\|_{WT} \\
&\qquad\qquad \leq W(0)\, \xi \left|x' - x\right| + C_* \bar{K}^* \exp(\lambda_0 x)\, \varphi\left(\left|x' - x\right|\right) \\
&\qquad\qquad\qquad \leq C_x \left\{\left|x' - x\right| + \varphi\left(\left|x' - x\right|\right)\right\},
\end{aligned}$$

donde $C_x := \max\left\{W(0)\, \xi\ ,\ C_* \bar{K}^* \exp(\lambda_0 x)\right\}$ y C_* es la constante que aparece en la Hipótesis 2.10*(b)*. Luego, definiendo

$$\varphi_x^Q(y) := C_x \left\{|y| + \varphi(|y|)\right\}$$

obtenemos que la desigualdad (2.31) se cumple.

■

Para concluir con el ejemplo, de (2.20) obsérvese que la condición (2.33) se satisface si $Q(U \mid x, a) > 0$ para cada conjunto abierto no vacío $U \subset \mathbb{X}$ y $(x, a) \in \mathbb{K}_A$, lo cual se obtiene como sigue

$$\begin{aligned} Q(U \mid x, a) &= \int_{\mathbb{R}} 1_U \left[(x+y)^+\right] g_a(y)\, dy \\ &= 1_U(0) \int_{-\infty}^{-x} g_a(y)\, dy + \int_0^{\infty} 1_U(y)\, g_a(y-x)\, dy > 0. \end{aligned}$$

Capítulo 3

$MCSMs$ con Distribución del Tiempo de Permanencia Parcialmente Conocida

3.1. Introducción

A diferencia del capítulo anterior, en el presente nuestro enfoque es sobre los procesos de control semi-markovianos cuya distribución del tiempo de permanencia depende, no solo de los pares (x_n, a_n), sino además de cierto parámetro θ desconocido y posiblemente no observable, el cual puede cambiar de etapa en etapa. En este sentido la distribución del tiempo de permanencia toma la forma $F(\cdot \mid x, a, \theta)$.

El hecho de que θ sea no observable nos impide implementar métodos de estimación estadística y control, similares a los del capítulo anterior. Sin embargo, supondremos que en cada etapa la única información que tiene el controlador es que $\theta \in \Theta(x, a)$. En tal situación, el controlador está interesado en seleccionar acciones dirigidas a minimizar el máximo costo generado sobre el conjunto de parámetros.

Específicamente, modelaremos el problema de control asociado al $MCSM$ (1.1), introducido en el Capítulo 1, como una clase de problemas de control minimax conocidos como juegos contra la naturaleza. En términos generales este enfoque consiste en suponer que el controlador tiene un oponente, la naturaleza, el cual elegirá el valor del parámetro en cada etapa. Por consiguiente, el interés del controlador consiste en minimizar el máximo costo que genera la naturaleza.

El objetivo de este capítulo es mostrar la existencia de políticas de control minimax, tanto para el criterio de costo descontado como el de costo promedio. Para lo anterior, este capítulo está estructurado de la siguiente manera. En la Sección 3.2 describimos el modelo de control minimax, mientras que en la Sección 3.3 introducimos los criterios minimax en el caso descontado y promedio. Finalmente, en las Secciones 3.4 y 3.5 mostramos la existencia de políticas minimax descontadas y promedio, respectivamente.

3.2. Modelo de control minimax semi-markoviano

Consideremos un modelo de control minimax semi-markoviano ($MXSM$) de la forma

$$(\mathbb{X}, \mathbb{A}, \Theta, \mathbb{K}_A, \mathbb{K}, Q, F, D, d) \tag{3.1}$$

donde $\mathbb{X}$, $\mathbb{A}$, Q, F, D, d y $\mathbb{K}_A$ son como en (1.1) y (1.2), respectivamente; mientras que, Θ es el espacio de acción (control) del oponente, el cual suponemos es un espacio de Borel. El conjunto $\mathbb{K} \in \mathcal{B}(\mathbb{X} \times \mathbb{A} \times \Theta)$, tal que $(x, a, \theta) \in \mathbb{K}$ implica que $(x, a) \in \mathbb{K}_A$, es el espacio de restricción para el oponente. De aquí, para cada $(x, a) \in \mathbb{K}_A$, asumimos que

$$\Theta(x, a) := \{\theta \in \Theta : \ (x, a, \theta) \in \mathbb{K}\}$$

es un subconjunto medible no vacío de Θ que respresenta al conjunto de acciones admisibles para el oponente cuando el estado es $x \in \mathbb{X}$ y el controlador elige la acción $a \in A(x)$. De acuerdo a este escenario, $F(\cdot \mid x, a, \theta)$ es la función de distribución del tiempo de permanencia en el estado $x \in \mathbb{X}$ cuando es elegido el control $a \in A(x)$ y el oponente selecciona $\theta \in \Theta(x, a)$. Por otra parte, el costo por etapa, es una función real sobre $\mathbb{K}$, medible y posiblemente no acotada, la cual está definida en términos de las funciones de costo D y d acorde al criterio de optimalidad en consideración.

El $MXSM$ (3.1) tiene una interpretación en términos de un juego contra la naturaleza. En efecto, una vez que el controlador selecciona la acción $a_n = a$ cuando el sistema está en el estado $x_n = x$, el oponente selecciona un parámetro $\theta_n = \theta \in \Theta(x, a)$ y el sistema permanece en el estado x durante el tiempo aleatorio δ_{n+1} con distribución $F(\cdot \mid x, a, \theta)$. Luego, el sistema evoluciona de forma similar a un $MCSM$ estándar. Ahora, debido a que el costo por etapa $c_\alpha(x, a, \theta)$ o $c(x, a, \theta)$ depende del parámetro seleccionado en cada etapa, el objetivo del controlador consiste en minimizar el máximo costo generado por la naturaleza.

3.2.1. Políticas de control

La definición de políticas de control para este tipo de problemas es muy similar a la que se presentó en el Capítulo 1. Incluiremos estas definiciones en forma resumida para adaptarlas al contexto minimax.

Sea $\tilde{\mathbb{H}}_0 := \mathbb{X}$, $\tilde{\mathbb{H}}'_0 := \mathbb{K}_A$ y para $n \in \mathbb{N}$ sean $\tilde{\mathbb{H}}_n := \mathbb{K}^n \times \mathbb{X}$ y $\tilde{\mathbb{H}}'_n := \mathbb{K}^n \times \mathbb{K}_A$, de modo que los elementos genéricos de $\tilde{\mathbb{H}}_n$ y $\tilde{\mathbb{H}}'_n$ toman la forma $\tilde{\imath}_n = (x_0, a_0, \theta_0, ..., x_{n-1}, a_{n-1}, \theta_{n-1}, x_n)$ y $\tilde{\imath}'_n = (\tilde{\imath}_n, a_n)$, respectivamente.

Una política para el controlador es una sucesión $\pi := \{\pi_n\}$ de kérneles estocásticos sobre $\mathbb{A}$ dado $\tilde{\mathbb{H}}_n$, tal que $\pi_n(A(x_n) \mid \tilde{\imath}_n) = 1$ para todo $\tilde{\imath}_n \in \tilde{\mathbb{H}}_n$ y $n \in \mathbb{N}_0$.

En particular, denotaremos por $\Pi_{\mathbb{A}}$ al conjunto de todas las políticas para el controlador, y por $\mathbb{F}_{\mathbb{A}} \subset \Pi_{\mathbb{A}}$ al conjunto de todas las políticas estacionarias. Como es usual (véase Definición 1.4*(c)*), cada política estacionaria $\pi \in \mathbb{F}_{\mathbb{A}}$ será identificada con alguna función medible $f : \mathbb{X} \to \mathbb{A}$ tal que $\pi_n(\cdot \mid \tilde{\imath}_n)$ está concentrado en $f(x_n) \in A(x_n)$ para todo $\tilde{\imath}_n \in \tilde{\mathbb{H}}_n$ y $n \in \mathbb{N}_0$. Mientras que, una política para el oponente es una sucesión $\pi' = \{\pi'_n\}$ de kerneles estocásticos sobre Θ dado $\tilde{\mathbb{H}}'_n$, tal que $\pi'_n(\Theta(x_n, a_n) \mid \tilde{\imath}'_n) = 1$ para todo $\tilde{\imath}'_n \in \tilde{\mathbb{H}}'_n$ y $n \in \mathbb{N}_0$. De este modo, denotaremos por Π_Θ al conjunto de todas las políticas para el oponente, y por $\mathbb{F}_\Theta \subset \Pi_\Theta$ al conjunto de todas las políticas estacionarias. Además, análogamente identificaremos cada política estacionaria $\pi' \in \mathbb{F}_\Theta$ con alguna función medible $g : \mathbb{X} \times \mathbb{A} \to \Theta$ tal que $\pi'_n(\cdot \mid \tilde{\imath}'_n)$ está concentrado en $g(x_n, a_n) \in \Theta(x_n, a_n)$ para todo $\tilde{\imath}'_n \in \tilde{\mathbb{H}}'_n$ y $n \in \mathbb{N}_0$.

3.2.2. Hipótesis sobre el modelo de control minimax

Introduciremos aquí dos clases de condiciones sobre el $MXSM$ (3.1). La primera es la condición de regularidad análoga a la Hipótesis 1.1 (véase Capítulo 1), mientras que la segunda contiene condiciones de continuidad y compacidad con el propósito de garantizar la existencia de selectores minimax.

Hiptesis 3.1 *Existen constantes positivas ϵ_0 y ζ_0 tal que*

$$F(\zeta_0 \mid x, a, \theta) \leq 1 - \epsilon_0 \qquad \forall\, (x, a, \theta) \in \mathbb{K}.$$

Hiptesis 3.2 *(a) Las funciones de costo $D(x, a)$ y $d(x, a)$ son s.c.i. en $\mathbb{K}_A$. Además, existe una función continua $W : \mathbb{X} \to [1, \infty)$ y una constante $M_0 > 0$ tal que*

$$\max\{D(x, a), d(x, a)\} \leq M_0 W(x) \qquad \forall\, (x, a) \in \mathbb{K}_A.$$

(b) La ley de transición Q es débilmente continua en $\mathbb{K}_A$, es decir, para cada función continua y acotada $u : \mathbb{X} \to \mathbb{R}$, la función

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

es continua en $\mathbb{K}_A$.

(c) La función

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

es continua en $\mathbb{K}_A$.

(d) La multifunción $x \mapsto A(x)$ es semicontinua superiormente (s.c.s.) y, el conjunto $A(x)$ es compacto para cada $x \in \mathbb{X}$.

(e) La multifunción $(x,a) \mapsto \Theta(x,a)$ es semicontinua inferiormente (s.c.i.) y, para cada $(x,a) \in \mathbb{K}_A$, $\Theta(x,a)$ es un conjunto σ-compacto.

Denotaremos por $\mathbb{L}_W(\mathbb{X}) \subset \mathbb{B}_W(\mathbb{X})$ al subespacio de funciones s.c.i. en $\mathbb{B}_W(\mathbb{X})$.

Observacin 3.1 *(a) La Hipótesis 3.2(b) (véase [15, p. 176]) se puede sustituir por la condición equivalente: Para cada función $u : \mathbb{X} \to \mathbb{R}$ s.c.i. y acotada inferiormente, la función*

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

es s.c.i. en $\mathbb{K}_A$.

(b) Se puede demostrar que $\mathbb{L}_W(\mathbb{X})$ es un subconjunto cerrado de $\mathbb{B}_W(\mathbb{X})$. Por lo cual, dado que $\mathbb{B}_W(\mathbb{X})$ es un espacio de Banach, se tiene que $\mathbb{L}_W(\mathbb{X})$ es un subespacio completo de $\mathbb{B}_W(\mathbb{X})$.

3.3. Criterios de optimalidad minimax

Para estudiar el criterio minimax de costo descontado, asumiremos de nuevo que los costos son descontados en forma continua. Entonces, para un factor de descuento fijo $\alpha > 0$ (véase (1.7)), definimos el costo por etapa para el criterio descontado como

$$c_\alpha(x,a,\theta) := D(x,a) + d(x,a) \int_0^\infty \int_0^t \exp(-\alpha s)\, ds F(dt \mid x,a,\theta), \qquad (x,a,\theta) \in \mathbb{K}. \tag{3.2}$$

Definicin 3.1 *Para cada $(x,a,\theta) \in \mathbb{K}$ definimos*

$$\Delta_\alpha(x,a,\theta) := \int_0^\infty \exp(-\alpha s)\, F(ds \mid x,a,\theta) \tag{3.3}$$

y

$$\tau_\alpha(x,a,\theta) := \frac{1 - \Delta_\alpha(x,a,\theta)}{\alpha}, \tag{3.4}$$

(véase Definición 1.5).

Observacin 3.2 *Utilizando un procedimiento similar al de la Sección 1.4 se deduce que el costo por etapa (3.2) toma la forma*

$$c_\alpha(x,a,\theta) = D\,(x,a) + \tau_\alpha(x,a,\theta) d\,(x,a)\,, \qquad (x,a,\theta) \in \mathbb{K}. \tag{3.5}$$

Definicin 3.2 *Para cada estado inicial $x \in \mathbb{X}$, cada par de políticas (π,π') en $\Pi_{\mathbb{A}} \times \Pi_\Theta$ y $\alpha > 0$ fijo, definimos el costo total esperado descontado como*

$$V(x,\pi,\pi') := E_x^{\pi\pi'}\left[\sum_{n=0}^{\infty} \exp\left(-\alpha T_n\right) c_\alpha(x_n,a_n,\theta_n)\right], \tag{3.6}$$

donde $E_x^{\pi\pi'}$ denota el operador esperanza respecto a la medida de probabilidad $P_x^{\pi\pi'}$ inducida por (π,π'), dado $x_0 = x$.

Para cada $x \in \mathbb{X}$ y $\pi \in \Pi_{\mathbb{A}}$, sea

$$V'\,(x,\pi) := \sup_{\pi' \in \Pi_\Theta} V(x,\pi,\pi').$$

Entonces, el problema de control minimax semi-markoviano descontado para el controlador consiste en encontrar una política $\pi^* \in \Pi_{\mathbb{A}}$ tal que

$$V'\,(x,\pi^*) = \inf_{\pi \in \Pi_{\mathbb{A}}} V'(x,\pi) = \inf_{\pi \in \Pi_{\mathbb{A}}} \sup_{\pi' \in \Pi_\Theta} V(x,\pi,\pi') =: V^*\,(x)\,, \quad x \in \mathbb{X}. \tag{3.7}$$

En este caso, a π^* la llamaremos *política descontada-minimax*, y a V^* la *función de valor óptimo descontado-minimax*.

Para definir el criterio minimax de costo promedio, definimos primero el *tiempo medio de permanencia* en el estado $x \in \mathbb{X}$, cuando el controlador y el oponente seleccionan $a \in A\,(x)$ y $\theta \in \Theta\,(x,a)$, respectivamente, como

$$\tau\,(x,a,\theta) := \int_0^\infty t F\,(dt \mid x,a,\theta)\,, \qquad (x,a,\theta) \in \mathbb{K}. \tag{3.8}$$

En tal situación, bajo argumentos similares a los empleados en (1.13), para $x \in \mathbb{X}$ y $(\pi,\pi') \in \Pi_{\mathbb{A}} \times \Pi_\Theta$, definimos el *costo promedio esperado* por

$$J(x,\pi,\pi') := \limsup_{n\to\infty} \frac{E_x^{\pi\pi'}\left[\sum_{k=0}^{n-1} c(x_k,a_k,\theta_k)\right]}{E_x^{\pi\pi'}\left[\sum_{k=0}^{n-1} \tau(x_k,a_k,\theta_k)\right]}, \tag{3.9}$$

donde

$$c\left(x,a,\theta\right) := D\left(x,a\right) + \tau\left(x,a,\theta\right) d\left(x,a\right), \qquad (x,a,\theta) \in \mathbb{K}, \tag{3.10}$$

es el costo por etapa para el criterio promedio.

Si definimos

$$J'\left(x,\pi\right) := \sup_{\pi' \in \Pi_\Theta} J(x,\pi,\pi'), \qquad x \in \mathbb{X},\, \pi \in \Pi_{\mathbb{A}}, \tag{3.11}$$

entonces el problema de control minimax semi-markoviano promedio para el controlador consiste en encontrar una política $\pi^* \in \Pi_{\mathbb{A}}$ tal que

$$J'\left(x,\pi^*\right) = \inf_{\pi \in \Pi_{\mathbb{A}}} J'(x,\pi) = \inf_{\pi \in \Pi_{\mathbb{A}}} \sup_{\pi' \in \Pi_\Theta} J(x,\pi,\pi') =: J^*\left(x\right), \quad x \in \mathbb{X}. \tag{3.12}$$

Llamaremos a π^* una *política promedio-minimax* y a J^* la *función de valor óptimo promedio-minimax* .

De acuerdo a lo anterior, nuestro objetivo es obtener condiciones para la existencia de políticas minimax, tanto en el caso descontado como en el promedio para el $MXSM$ (3.1).

3.4. Criterio de costo descontado minimax

En esta sección analizaremos el criterio de costo descontado minimax, el cual fue definido mediante las expresiones (3.2)-(3.7).

Nótese primero que de (3.3), y similarmente al resultado de la Proposición 2.1, podemos reescribir el criterio de optimalidad (3.6) como:

$$\begin{aligned} V(x,\pi,\pi') := E_x^{\pi\pi'} \Big[& c_\alpha(x_0,a_0,\theta_0) \\ & + \sum_{n=1}^{\infty} \prod_{k=0}^{n-1} \Delta_\alpha\left(x_k,a_k,\theta_k\right) c_\alpha(x_n,a_n,\theta_n) \Big], \\ & x \in \mathbb{X}, \left(\pi,\pi'\right) \in \Pi_{\mathbb{A}} \times \Pi_\Theta. \end{aligned} \tag{3.13}$$

Una primera consecuencia de las Hipótesis 3.1-3.2*(a)* es la siguiente.

Lema 3.1 *Bajo las Hipótesis 3.1 y 3.2(a) tenemos:*

(a) $\bar{\rho}_\alpha := \sup_{(x,a,\theta)\in\mathbb{K}} \Delta_\alpha(x,a,\theta) < 1$;

(b) Para alguna constante positiva M_1',

$$|c_\alpha(x,a,\theta)| \leq M_1' W(x) \qquad \forall\,(x,a,\theta) \in \mathbb{K}.$$

Demostración. La demostración de la parte *(a)* es semejante a la de Proposición 1.1*(a)*. Para la parte *(b)* obsérvese que de *(a)* y (3.4),

$$\tau_\alpha(x,a,\theta) \leq \frac{1}{\alpha} \qquad \forall\,(x,a,\theta) \in \mathbb{K}.$$

Luego, por la Hipótesis 3.2*(a)*, la parte *(b)* se sigue de la expresión (3.5) considerando

$$M_1' := \left(1 + \frac{1}{\alpha}\right) M_0.$$

■

Como la función W no necesariamente es acotada (Hipótesis 3.2), el Lema 3.1*(b)* nos permite suponer que el costo por etapa no necesariamente es acotado. Sin embargo, para obtener nuestros resultados, como en el capítulo anterior, es necesario imponer condiciones para regular el crecimiento de W. Además, necesitamos condiciones de continuidad en la función de distribución F.

Hiptesis 3.3 *(a) Existe una constante* $\bar{\beta}$ *tal que* $1 \leq \bar{\beta} < 1/\bar{\rho}_\alpha$, *y para todo* $(x,a) \in \mathbb{K}_A$

$$\int_{\mathbb{X}} W(y)\, Q(dy \mid x,a) \leq \bar{\beta} W(x). \tag{3.14}$$

(b) Para cada $t \geq 0$, $F(t \mid x,a,\theta)$ *es continua en* $\mathbb{K}$.

Observacin 3.3 *(a) Para todo* $x \in \mathbb{X}$, $n \in \mathbb{N}_0$ *y* $(\pi,\pi') \in \Pi_{\mathbb{A}} \times \Pi_\Theta$, *de (3.14) se tiene que*

$$E_x^{\pi\pi'}[W(x_{n+1})] \leq \bar{\beta} E_x^{\pi\pi'}[W(x_n)],$$

lo cual a su vez implica

$$E_x^{\pi\pi'}[W(x_{n+1})] \leq \left(\bar{\beta}\right)^{n+1} E_x^{\pi\pi'}[W(x)]. \tag{3.15}$$

Por lo tanto, de (3.13) y el Lema 3.1, para cada $x \in \mathbb{X}$ *y* $(\pi, \pi') \in \Pi_{\mathbb{A}} \times \Pi_{\Theta}$,

$$\begin{aligned} \left|V(x,\pi,\pi')\right| &\leq E_x^{\pi\pi'}\left[\ \left|c_\alpha(x_0,a_0,\theta_0) + \sum_{n=1}^{\infty}(\bar{\rho}_\alpha)^n\, c_\alpha(x_n,a_n,\theta_n)\right|\ \right] \\ &\leq M_1' W(x) \sum_{n=0}^{\infty}\left(\bar{\beta}\bar{\rho}_\alpha\right)^n = \frac{M_1' W(x)}{1-\bar{\beta}\bar{\rho}_\alpha}. \end{aligned}$$

Entonces,

$$\|V^*\|_W \leq \frac{M_1'}{1-\bar{\beta}\bar{\rho}_\alpha}. \tag{3.16}$$

(b) De la Hipótesis 3.3(b), se tiene que la función $\Delta_\alpha(x,a,\theta)$ *es continua en* $\mathbb{K}$ *(véase (3.3)); mientras que, por la Hipótesis 3.2(a) la función de costo* $c_\alpha(x,a,\theta)$ *en (3.5) es s.c.i. en* $\mathbb{K}_A$ *y continua en* θ.

Para $u \in \mathbb{B}_W(\mathbb{X})$ y $(x,a,\theta) \in \mathbb{K}$, definamos

$$H_\alpha(u,x,a,\theta) := c_\alpha(x,a,\theta) + \Delta_\alpha(x,a,\theta)\int_{\mathbb{X}} u(y)\, Q(dy \mid x,a)$$

y

$$T_\alpha u(x) := \inf_{a\in A(x)} \sup_{\theta\in\Theta(x,a)} H_\alpha(u,x,a,\theta). \tag{3.17}$$

Lema 3.2 *Si las Hipótesis 3.1-3.3 se satisfacen, entonces:*

(a) El operador T_α *es una contracción con módulo* $\bar{\beta}\bar{\rho}_\alpha < 1$.

(b) T_α *transforma* $\mathbb{L}_W(\mathbb{X})$ *en sí mismo.*

(c) Para cada $u \in \mathbb{L}_W(\mathbb{X})$, *existe* $f^* \in \mathbb{F}_{\mathbb{A}}$ *tal que*

$$T_\alpha u(x) = \sup_{\theta\in\Theta(x,f^*)} H_\alpha(u,x,f^*,\theta), \qquad x \in \mathbb{X}.$$

Demostración. *(a)* Sean u, $u' \in \mathbb{B}_W(\mathbb{X})$. Entonces, de la definición en (1.16), junto con el Lema 3.1*(a)* y la Hipótesis 3.3*(a)*, tenemos

$$\begin{aligned} H_\alpha(u,x,a,\theta) &\leq H_\alpha\left(u',x,a,\theta\right) + \Delta_\alpha(x,a,\theta)\int_{\mathbb{X}}\left|u(y)-u'(y)\right| Q(dy \mid x,a) \\ &\leq H_\alpha\left(u',x,a,\theta\right) + \bar{\beta}\bar{\rho}_\alpha\left\|u-u'\right\|_W W(x) \qquad \forall\,(x,a,\theta) \in \mathbb{K}, \end{aligned}$$

lo cual implica (véase (3.17))

$$T_\alpha u\,(x) \leq T_\alpha u'\,(x) + \bar{\beta}\bar{\rho}_\alpha \left\|u - u'\right\|_W W\,(x) \quad \forall x \in \mathbb{X}.$$

Por lo tanto,

$$T_\alpha u\,(x) - T_\alpha u'\,(x) \leq \bar{\beta}\bar{\rho}_\alpha \left\|u - u'\right\|_W W\,(x) \quad \forall x \in \mathbb{X}. \tag{3.18}$$

Un argumento similar demuestra que

$$T_\alpha u'\,(x) - T_\alpha u\,(x) \leq \bar{\beta}\bar{\rho}_\alpha \left\|u - u'\right\|_W W\,(x) \quad \forall x \in \mathbb{X}. \tag{3.19}$$

Luego, combinando (3.18) y (3.19) obtenemos

$$\left\|T_\alpha u - T_\alpha u'\right\|_W \leq \bar{\beta}\bar{\rho}_\alpha \left\|u - u'\right\|_W,$$

con lo cual queda demostrada la parte *(a)*.

(b) Sea $u \in \mathbb{L}_W\,(\mathbb{X})$, y definamos

$$\hat{H}\,(u, x, a) := \sup_{\theta \in \Theta(x,a)} H_\alpha\,(u, x, a, \theta)\,, \quad (x, a) \in \mathbb{K}_A.$$

Para comprobar la parte *(b)* es suficiente con demostrar:

(i) Para alguna constante $\hat{M} < \infty$, $\left|\hat{H}\,(u, x, a)\right| \leq \hat{M} W\,(x)$ para todo $(x, a) \in \mathbb{K}_A$; y

(ii) $\hat{H}$ es s.c.i. en $\mathbb{K}_A$.

Por definición de T_α (véase (3.17)), y la afirmación en *(i)* se tiene

$$|T_\alpha u\,(x)| \leq \hat{M} W\,(x) \quad \forall x \in \mathbb{X},$$

lo cual implica que

$$\|T_\alpha u\|_W < \infty. \tag{3.20}$$

Ahora, si se satisface la afirmación *(ii)*, por *(i)* tenemos que la función

$$(x, a) \mapsto \hat{H}\,(u, x, a) + \hat{M} W\,(x) \tag{3.21}$$

es no negativa y s.c.i. en $\mathbb{K}_A$. De lo anterior, por la Hipótesis 3.2*(d)* y debido a un resultado de Schäl en [45], (véase también [15, Proposición D.5]), deducimos que

$$\begin{aligned} H\,(u, x) &:= \inf_{a \in A(x)} \left\{\hat{H}\,(u, x, a) + \hat{M} W\,(x)\right\} \\ &= T_\alpha u\,(x) + \hat{M} W\,(x) \end{aligned} \tag{3.22}$$

es una función s.c.i. en $\mathbb{X}$. Luego, por la continuidad de $W(\cdot)$, obtenemos que

$$T_\alpha u(x) = H(u,x) - \hat{M}W(x)$$

es entonces una función s.c.i. en $\mathbb{X}$. Este hecho junto con (3.20) implica que T_α transforma $\mathbb{L}_W(\mathbb{X})$ en sí mismo.

Para concluir con la parte *(b)* ahora demostraremos los puntos *(i)* y *(ii)*.

Sea $u \in \mathbb{L}_W(\mathbb{X})$. De (1.16), junto con el Lema 3.1*(b)* y la Hipótesis 3.3*(a)*, el punto *(i)* es inmediato definiendo

$$\hat{M} := M_1' + \bar{\beta}\bar{\rho}_\alpha \|u\|_W .$$

Por otra parte, obsérvese que

$$|u(x)| \leq \|u\|_W W(x), \quad x \in \mathbb{X}.$$

De modo que, debido a la continuidad de $W(\cdot)$, la función

$$v(x) := u(x) + \|u\|_W W(x)$$

es no negativa y s.c.i., lo que implica que (véase Observación 3.1*(a)*)

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

es una función s.c.i. en $\mathbb{K}_A$. Entonces, por la Hipótesis 3.2*(c)*, la función

$$\int_{\mathbb{X}} u(y)\, Q(dy \mid x,a) = \int_{\mathbb{X}} v(y)\, Q(dy \mid x,a) - \|u\|_W \int_{\mathbb{X}} W(y)\, Q(dy \mid x,a)$$

es s.c.i. en $\mathbb{K}_A$. Por consiguiente, (véase Observación 3.3*(b)*), $H_\alpha(u,x,a,\theta)$ es s.c.i. en $\mathbb{K}$.

Ahora, sea $\{(x_j,a_j)\} \subset \mathbb{K}_A$ una sucesión convergente a $(x,a) \in \mathbb{K}_A$ y, sea $\theta \in \Theta(x,a)$ arbitrario. Entonces, dado que $\Theta(x,a)$ es σ-compacto (véase Hipótesis 3.2*(e)*), existe $\theta_j \in \Theta(x_j,a_j)$ tal que $\theta_j \to \theta$. De aquí,

$$\begin{aligned}
\liminf_{j\to\infty} \hat{H}(u,x_j,a_j) &= \liminf_{j\to\infty} \sup_{\theta\in\Theta(x_j,a_j)} H_\alpha(u,x_j,a_j,\theta) \\
&\geq \liminf_{j\to\infty} H_\alpha(u,x_j,a_j,\theta_j) \\
&\geq H_\alpha(u,x,a,\theta),
\end{aligned}$$

donde la última desigualdad se debe a la semicontinuidad inferior de H_α. Dado que θ fue escogida arbitrariamente, tenemos que

$$\liminf_{j\to\infty} \hat{H}\left(u, x_j, a_j\right) \geq \hat{H}\left(u, x, a\right),$$

lo cual implica que $\hat{H}$ es s.c.i. en $\mathbb{K}_A$. Esto completa la demostración de la parte *(b)*.

(c) Debido a la no negatividad y a la semicontinuidad inferior de la función definida en (3.21), así como a la Hipótesis 3.2*(d)*, argumentos estándar sobre la existencia de minimizadores (véase, por ejemplo, [43, 45]) implican que existe $f^* \in \mathbb{F}_{\mathbb{A}}$ tal que

$$\inf_{a\in A(x)} \left\{\hat{H}\left(u, x, a\right) + \hat{M}W\left(x\right)\right\} = \hat{H}\left(u, x, f^*\right) + \hat{M}W\left(x\right).$$

Por lo tanto, combinando esta relación con (3.22) obtenemos la parte *(c)*.

■

Observacin 3.4 *Debido a que T_α es un operador de contracción que transforma $\mathbb{L}_W\left(\mathbb{X}\right)$ en sí mismo (véase Lema 3.2) y dado que $\mathbb{L}_W\left(\mathbb{X}\right) \subset \mathbb{B}_W\left(\mathbb{X}\right)$ es completo (véase Observación 3.1(b)), por el Teorema de Punto Fijo de Banach, existe una función única $\tilde{u} \in \mathbb{L}_W\left(\mathbb{X}\right)$ tal que para todo $x \in \mathbb{X}$,*

$$\tilde{u}\left(x\right) = T_\alpha \tilde{u}\left(x\right), \tag{3.23}$$

y

$$\left\|T_\alpha^n u - \tilde{u}\right\|_W \leq \left(\bar{\beta}\bar{\rho}_\alpha\right)^n \left\|u - \tilde{u}\right\|_W \qquad \forall u \in \mathbb{L}_W\left(\mathbb{X}\right),\ n \in \mathbb{N}_0. \tag{3.24}$$

Definicin 3.3 *Para cada $n \in \mathbb{N}$, $x \in \mathbb{X}$ y $(\pi, \pi') \in \Pi_{\mathbb{A}} \times \Pi_\Theta$, definimos el costo descontado esperado en n etapas (véase (3.13)) como*

$$V_n\left(x, \pi, \pi'\right) := \begin{cases} E_x^{\pi\pi'}\left[c_\alpha\left(x_0, a_0, \theta_0\right)\right], & n = 1 \\ E_x^{\pi\pi'}\left[c_\alpha\left(x_0, a_0, \theta_0\right)\right. & \\ \qquad \left. + \sum_{j=1}^{n-1} \prod_{k=0}^{j-1} \Delta_\alpha\left(x_k, a_k, \theta_k\right) c_\alpha\left(x_j, a_j, \theta_j\right)\right], & n \geq 2. \end{cases}$$

Además, definimos la sucesión $\{v_n\}$ en $\mathbb{L}_W\left(\mathbb{X}\right)$ como $v_0 := 0$, y para $n \in \mathbb{N}$,

$$v_n\left(x\right) := T_\alpha v_{n-1}\left(x\right), \qquad x \in \mathbb{X}.$$

3.5. Existencia de políticas descontadas-minimax

Ahora establecemos el resultado principal correspondiente al criterio de costo descontado de la siguiente manera.

Teorema 3.1 *Si las Hipótesis 3.1-3.3 se cumplen, entonces:*

(a) La función de valor óptimo descontado-minimax (3.7) es la única solución en $\mathbb{L}_W(\mathbb{X})$ *que satisface*

$$V^*(x) = T_\alpha V^*(x), \qquad x \in \mathbb{X}.$$

(b) Para cada $n \in \mathbb{N}$,

$$\|v_n - V^*\|_W \leq M_1' \frac{\left(\bar{\beta}\bar{\rho}_\alpha\right)^n}{1 - \bar{\beta}\bar{\rho}_\alpha}.$$

(c) Existe $f^* \in \mathbb{F}_\mathbb{A}$ *tal que para cada* $x \in \mathbb{X}$,

$$V^*(x) = \sup_{\theta \in \Theta(x, f^*(x))} \left\{ c\left(x, f^*(x), \theta\right) + \Delta_\alpha\left(x, f^*(x), \theta\right) \int_\mathbb{X} V^*(y)\, Q(dy \mid x, f^*(x)) \right\},$$

y además, f^* *es una política descontada-minimax para el controlador (véase* $MXSM$ *(3.1)), es decir, para cada* $x \in \mathbb{X}$,

$$V^*(x) = \sup_{\pi' \in \Pi_\Theta} V\left(x, f^*, \pi'\right).$$

Demostración. *(a)-(b)* Primero obsérvese que $v_n = T_\alpha^n v_0$ para todo $n \in \mathbb{N}_0$. Entonces, tomando $u = v_0$ en (3.24) tenemos

$$\|v_n - \tilde{u}\|_W \leq \left(\bar{\beta}\bar{\rho}_\alpha\right)^n \|\tilde{u}\|_W \qquad \forall n \in \mathbb{N}_0.$$

Por lo tanto, de (3.16) y (3.23), las partes *(a)* y *(b)* del teorema quedarán demostradas si se comprueba que $\tilde{u} = V^*$.

Sea $f \in \mathbb{F}_\mathbb{A}$ un selector tal que (véase Lema 3.2*(c)*)

$$\tilde{u}(x) = \sup_{\theta \in \Theta(x, f(x))} H_\alpha\left(\tilde{u}, x, f(x), \theta\right), \qquad x \in \mathbb{X}.$$

Entonces,

$$\begin{aligned}\tilde{u}(x) \geq\ & c_\alpha(x, f(x), \theta) \\ & + \Delta_\alpha(x, f(x), \theta) \int_{\mathbb{X}} \tilde{u}(y)\, Q(dy \mid x, f(x)) \\ & \qquad \forall x \in \mathbb{X}, \theta \in \Theta(x, f(x)). \end{aligned} \tag{3.25}$$

Ahora, sea $\pi' \in \Pi_\Theta$ una política arbitraria para el oponente y $\{(x_n, f(x_n), \theta_n)\}$ una sucesión de ternas estado-acción correspondientes a la aplicación de las políticas f y π'. Luego, iterando la desigualdad (3.25) se obtiene que para cada $n \in \mathbb{N}$,

$$\begin{aligned}\tilde{u}(x) \geq\ & E_x^{f\pi'} \Big[c_\alpha(x_0, f(x_0), \theta_0) \\ & + \sum_{j=1}^{n-1} \prod_{k=0}^{j-1} \Delta_\alpha(x_k, f(x_k), \theta_k)\, c_\alpha(x_j, f(x_j), \theta_j) \Big] \\ & + E_x^{f\pi'} \left[\prod_{k=0}^{n-1} \Delta_\alpha(x_k, f(x_k), \theta_k)\, \tilde{u}(x_n) \right] \quad \forall x \in \mathbb{X}. \end{aligned}$$

Es decir, para cada $n \in \mathbb{N}$,

$$\begin{aligned}\tilde{u}(x) \geq V_n\left(x, f, \pi'\right) + E_x^{f\pi'} \left[\prod_{k=0}^{n-1} \Delta_\alpha(x_k, f(x_k), \theta_k)\, \tilde{u}(x_n) \right] \\ \forall x \in \mathbb{X}. \end{aligned} \tag{3.26}$$

Obsérvese que de (3.15) se tiene que para cada $n \in \mathbb{N}$,

$$\begin{aligned} E_x^{f\pi'} \left[\prod_{k=0}^{n-1} \Delta_\alpha(x_k, f(x_k), \theta_k)\, \tilde{u}(x_n) \right] &\leq (\bar{\rho}_\alpha)^n\, E_x^{f\pi'} [\ |\tilde{u}(x_n)|\ ] \\ &\leq \left(\bar{\beta}\bar{\rho}_\alpha\right)^n \|\tilde{u}\|_W\, W(x), \quad x \in \mathbb{X}. \end{aligned}$$

De esta desigualdad y haciendo $n \to \infty$ en (3.26) se obtiene lo siguiente

$$\tilde{u}(x) \geq V\left(x, f, \pi'\right) \qquad \forall x \in \mathbb{X}, \pi' \in \Pi_\Theta, \tag{3.27}$$

lo cual implica que

$$\tilde{u}(x) \geq \sup_{\pi' \in \Pi_\Theta} V\left(x, f, \pi'\right) \qquad \forall x \in \mathbb{X}. \tag{3.28}$$

En consecuencia,

$$\tilde{u}(x) \geq V^*(x) \qquad \forall x \in \mathbb{X}. \tag{3.29}$$

Por otra parte, dado que las funciones c_α y Δ_α son continuas en θ (véase Observación 3.3*(b)*), nótese que de (3.23) junto con la Hipótesis 3.2*(e)*, y aplicando un teorema de selección medible adecuado (véase, por ejemplo, [43]), se tiene que para cada $\varepsilon > 0$, existe $g : \mathbb{K}_A \to \Theta$, con $g(x,a) \in \Theta(x,a)$, tal que

$$\begin{aligned}\tilde{u}(x) \leq \inf_{a\in A(x)} \Big\{ & c_\alpha(x,a,g(x,a)) \\ & + \Delta_\alpha(x,a,g(x,a)) \int_{\mathbb{X}} \tilde{u}(y)\, Q(dy \mid x,a) \Big\} + \varepsilon \\ \leq\ & c_\alpha(x,a,g(x,a)) \\ & + \Delta_\alpha(x,a,g(x,a)) \int_{\mathbb{X}} \tilde{u}(y)\, Q(dy \mid x,a) + \varepsilon \\ & \qquad \forall x \in \mathbb{X}, a \in A(x). \end{aligned} \tag{3.30}$$

Sea $\pi \in \Pi_{\mathbb{A}}$ una política arbitraria. Entonces, de modo similar como en (3.26), iterando (3.30) obtenemos

$$\tilde{u}(x) \leq V_n(x,\pi,g) + E_x^{\pi g}\left[\prod_{k=0}^{n-1} \Delta_\alpha(x_k,a_k,g(x,a))\, \tilde{u}(x_n)\right] + \frac{\varepsilon}{1-\bar{\rho}_\alpha} \qquad \forall x \in \mathbb{X}.$$

Entonces, (véase (3.27)), haciendo $n \to \infty$ tenemos

$$\tilde{u}(x) \leq V(x,\pi,g). \tag{3.31}$$

Luego, como

$$V(x,\pi,g) \leq \sup_{\pi' \in \Pi_\Theta} V\left(x,\pi,\pi'\right) \qquad \forall x \in \mathbb{X},$$

y $\pi \in \Pi_{\mathbb{A}}$ es arbitraria, entonces (3.31) implica

$$\tilde{u}(x) \leq \inf_{\pi\in\Pi_{\mathbb{A}}} \sup_{\pi'\in\Pi_\Theta} V\left(x,\pi,\pi'\right) = V^*(x) \qquad \forall x \in \mathbb{X},$$

lo cual combinado con (3.29) conduce a que

$$\tilde{u} = V^*.$$

(c) La existencia de $f^* \in \mathbb{F}_{\mathbb{A}}$ es consecuencia de lo establecido en la parte *(a)* y del Lema 3.2*(c)*. Además, aplicando argumentos similares como en la demostración

de la parte *(a)* (ver (3.28)) obtenemos

$$V^*(x) := \inf_{\pi\in\Pi_{\mathbb{A}}} \sup_{\pi'\in\Pi_\Theta} V\left(x,\pi,\pi'\right)$$
$$\geq \sup_{\pi'\in\Pi_\Theta} V\left(x,f^*,\pi'\right) \qquad \forall x\in\mathbb{X}.$$

Por lo tanto,

$$V^*(x) = \sup_{\pi'\in\Pi_\Theta} V\left(x,f^*,\pi'\right) \qquad \forall x\in\mathbb{X},$$

ésto es, f^* es una política descontada-minimax para el controlador. (Véase (3.7).)

■

3.6. Criterio de costo promedio minimax

Para analizar el problema de control minimax semi-markoviano promedio, definido anteriormente mediante las expresiones (3.8)-(3.12), necesitamos imponer algunas condiciones de acotamiento sobre el tiempo medio de permanencia. Esto es, se requiere de la existencia de constantes positivas m y M tales que satisfagan

$$m \leq \tau(x,a,\theta) \leq M \qquad \forall\,(x,a,\theta)\in\mathbb{K}. \tag{3.32}$$

Nótese que la primera desigualdad en (3.32) es consecuencia de la Hipótesis 3.1 puesto que para todo $(x,a,\theta)\in\mathbb{K}$,

$$\tau(x,a,\theta) \geq \int_{\zeta_0}^{\infty} tF(dt \mid x,a,\theta) \geq \zeta_0\epsilon_0 =: m,$$

(véase (3.1)). De aquí,

$$\inf_{(x,a,\theta)\in\mathbb{K}} \tau(x,a,\theta) \geq m. \tag{3.33}$$

La segunda desigualdad se introduce en la siguiente hipótesis.

Hiptesis 3.4 *Existe una constante positiva M tal que para todo $(x,a,\theta)\in\mathbb{K}$,*

$$\tau(x,a,\theta) \leq M.$$

A continuación demostraremos dos propiedades importantes, las cuales se obtienen como consecuencia de las Hipótesis 3.2-3.4.

Proposicin 3.1 *Bajo las Hipótesis 3.2(a) y 3.4 se tiene que, para alguna constante positiva* M_2*,*

$$|c(x,a,\theta)| \leq M_2 W(x) \qquad \forall\,(x,a,\theta) \in \mathbb{K}. \tag{3.34}$$

Demostración. De la definición de $c(x,a,\theta)$ en (3.10) y por la Hipótesis 3.2*(a)* vemos que, para todo $(x,a,\theta) \in \mathbb{K}$,

$$|c(x,a,\theta)| \leq |1+\tau(x,a,\theta)|\, M_0 W(x);$$

expresión de la cual, como consecuencia de la Hipótesis 3.4, obtenemos justamente la desigualdad (3.34), donde

$$M_2 := (1+M)\, M_0.$$

■

Proposicin 3.2 *Las Hipótesis 3.3(b) y 3.4 implican que la función* $\tau(x,a,\theta)$ *es continua en* $\mathbb{K}$*.*

Demostración. Primero, sea $\{(x_n,a_n,\theta_n)\} \subset \mathbb{K}$ una sucesión tal que

$$(x_n,a_n,\theta_n) \to (x,a,\theta) \quad \text{cuando} \;\; n \to \infty.$$

Entonces, para cada $t \geq 0$,

$$F(t \mid x_n,a_n,\theta_n) \to F(t \mid x,a,\theta) \quad \text{cuando} \;\; n \to \infty$$

debido a la continuidad de la función F en $\mathbb{K}$. Esto a su vez, implica que para cada $t \geq 0$,

$$\varrho_n(t) := 1 - F(t \mid x_n,a_n,\theta_n) \to 1 - F(t \mid x,a,\theta) \quad \text{cuando} \;\; n \to \infty.$$

De acuerdo con la expresión (3.8), nótese que para cada $(x,a,\theta) \in \mathbb{K}$ podemos también escribir

$$\tau(x,a,\theta) = \int_0^\infty \{1 - F(t \mid x,a,\theta)\}\, dt. \tag{3.35}$$

Sea

$$\varrho^*(t) := \sup_{n\in\mathbb{N}} \varrho_n(t), \quad t \geq 0.$$

Por la Hipótesis 3.4, existe una subsucesión $\{(x_{n_k}, a_{n_k}, \theta_{n_k})\}$ de $\{(x_n, a_n, \theta_n)\}$ tal que para cada $t \geq 0$,

$$\varrho_{n_k}(t) \uparrow \varrho^*(t) \quad \text{cuando} \quad k \to \infty.$$

De aquí, obsérvese que

$$\int_0^\infty \varrho^*(t)\, dt = \int_0^\infty \lim_{k\to\infty} \varrho_{n_k}(t)\, dt = \lim_{k\to\infty} \int_0^\infty \varrho_{n_k}(t)\, dt \leq M;$$

por lo cual, debido a (3.35) se tiene que para cada $(x, a, \theta) \in \mathbb{K}$,

$$\begin{aligned}\tau(x, a, \theta) &= \int_0^\infty tF(dt \mid x, a, \theta) \\ &= \lim_{k\to\infty} \int_0^\infty tF(dt \mid x_{n_k}, a_{n_k}, \theta_{n_k}) = \lim_{k\to\infty} \tau(x_{n_k}, a_{n_k}, \theta_{n_k}).\end{aligned}$$

Esto demuestra lo requerido.

■

Para el análisis asintótico del criterio de costo promedio, requerimos imponer la siguiente condición de ergodicidad.

Hiptesis 3.5 *Existe una medida de probabilidad μ sobre $\mathbb{X}$, una función semicontinua superiormente (s.c.s.) $\phi : \mathbb{K}_A \to [0, 1]$ y una constante $\gamma \in (0, 1)$, tales que*

(a) $Q(C \mid x, a) \geq \phi(x, a)\, \mu(C)$ *para todo* $(x, a) \in \mathbb{K}_A$ *y* $C \in \mathcal{B}(\mathbb{X})$.

(b) $\int_{\mathbb{X}} W(y)\, Q(dy \mid x, a) \leq \gamma W(x) + \phi(x, a)\, \mu(W)$, *donde*

$$\mu(W) := \int_{\mathbb{X}} W(y)\, \mu(dy) < \infty.$$

(c) $\int_{\mathbb{X}} \phi(x, f(x))\, \mu(dx) > 0$ *para cada* $f \in \mathbb{F}_{\mathbb{A}}$.

Observacin 3.5 *Como en el Capítulo 2 (véase Hipótesis 2.7), la Hipótesis 3.5 implica que para cada $f \in \mathbb{F}_{\mathbb{A}}$, la cadena de Markov definida por $Q(\cdot \mid \cdot, f)$ es μ-irreducible y Harris recurrente positiva (véase [48]).*

Sean $f \in \mathbb{F}_{\mathbb{A}}$ y $Q^n(\cdot \mid \cdot, f)$ el kernel de transición en n etapas asociado a f. Entonces, por la Hipótesis 3.5*(b)*,

$$\begin{aligned}\int_{\mathbb{X}} W(y)\, Q^n(dy \mid x, f) &\leq \gamma^n W(x) \\ &\quad + \left(1 + \gamma + \cdots + \gamma^{n-1}\right) \mu(W) \\ &\leq \left\{1 + \frac{\mu(W)}{1-\gamma}\right\} W(x),\end{aligned}$$

lo cual implica que para toda función $v \in \mathbb{B}_W(\mathbb{X})$,

$$\lim_{n\to\infty} \frac{1}{n} \int_{\mathbb{X}} v(y)\, Q^n(dy \mid x, f) = 0. \tag{3.36}$$

3.7. Existencia de políticas promedio-minimax

Ahora introduciremos el resultado principal correspondiente al criterio de costo promedio como sigue.

Teorema 3.2 *Supóngase que se cumplen las Hipótesis 3.1-3.5. Entonces existe una constante j^*, una función $h^* \in \mathbb{L}_W(\mathbb{X})$ y una política $f^* \in \mathbb{F}_{\mathbb{A}}$ tales que para todo $x \in \mathbb{X}$,*

$$\begin{aligned} h^*(x) &= \inf_{a\in A(x)} \sup_{\theta\in\Theta(x,a)} \left\{ c(x,a,\theta) - j^*\tau(x,a,\theta) \right. \\ &\qquad \left. + \int_{\mathbb{X}} h^*(y)\, Q(dy \mid x,a) \right\} \\ &= \sup_{\theta\in\Theta(x,f^*(x))} \left\{ c(x,f^*(x),\theta) - j^*\tau(x,f^*(x),\theta) \right. \\ &\qquad \left. + \int_{\mathbb{X}} h^*(y)\, Q(dy \mid x,f^*(x)) \right\}. \end{aligned}$$

Además, j^* es el costo óptimo promedio-minimax y f^* es una política promedio-minimax, es decir, para todo $x \in \mathbb{X}$,

$$\begin{aligned} j^* &= \sup_{\pi'\in\Pi_\Theta} J\left(x, f^*, \pi'\right) \\ &= \inf_{\pi\in\Pi_{\mathbb{A}}} \sup_{\pi'\in\Pi_\Theta} J\left(x,\pi,\pi'\right). \end{aligned}$$

La demostración del Teorema 3.2 se basa en una transformación del modelo de control minimax semi-markoviano, la cual fue introducida en [46] y aplicada por varios autores a modelos de control semi-markovianos (ver, por ejemplo, [27, 28]). Ahora introduciremos dicho procedimiento en nuestro contexto como sigue.

Sea τ un número real tal que

$$0 < \tau < m, \tag{3.37}$$

(véase (3.33)). Definimos la función $\hat{c} : \mathbb{K} \to \mathbb{R}$, así como el kernel estocástico $\hat{Q}$ sobre $\mathbb{X}$ dado $\mathbb{K}$ mediante

$$\hat{c}\,(x, a, \theta) := \frac{c\,(x, a, \theta)}{\tau\,(x, a, \theta)}, \tag{3.38}$$

y

$$\hat{Q}(C \mid x, a, \theta) := \frac{\tau}{\tau\,(x, a, \theta)} Q\,(C \mid x, a) + \left\{1 - \frac{\tau}{\tau\,(x, a, \theta)}\right\} \delta_x\,(C)\,, \tag{3.39}$$

donde $\delta_x\,(\cdot)$ es la medida de Dirac concentrada en x.

De lo anterior, la función de costo $\hat{c}$ y el kernel estocástico $\hat{Q}$ definen el modelo de control de Markov minimax

$$\left(\mathbb{X}, \mathbb{A}, \Theta, \mathbb{K}_A, \mathbb{K}, \hat{Q}, \hat{c}\right). \tag{3.40}$$

En tales circunstancias, nuestro trabajo consiste entonces en analizar el problema de control minimax correspondiente al modelo (3.40), cuya solución demuestra el Teorema 3.2. Para tal fin, requeriremos los siguientes resultados.

Lema 3.3 *Si las Hipótesis 3.1-3.2, 3.3(b) y 3.4 se cumplen, entonces la Hipótesis 3.2(b)-(c), así como la relación (3.34) se satisfacen cuando c, Q y $\mathbb{K}_A$ son reemplazadas por $\hat{c}$, $\hat{Q}$ y $\mathbb{K}$, respectivamente.*

Demostración. Primero demostraremos que la ley de transición $\hat{Q}$ es débilmente continua. Sea $u : \mathbb{X} \to \mathbb{R}$ una función continua y acotada, entonces por (3.39) tenemos

$$\int_{\mathbb{X}} u\,(y)\,\hat{Q}(dy \mid x, a, \theta) = \frac{\tau}{\tau\,(x, a, \theta)} \int_{\mathbb{X}} u\,(y)\,Q(dy \mid x, a) + \left\{1 - \frac{\tau}{\tau\,(x, a, \theta)}\right\} u\,(x)\,.$$

De modo que, debido a (3.33), así como a la Hipótesis 3.2*(b)* y a la Proposición 3.2, de lo anterior concluimos que la función

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

es continua en $\mathbb{K}$. Luego, de forma similar se demuestra que la función

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

es continua en $\mathbb{K}$.

Finalmente, obsérvese que de (3.33) y (3.34),

$$|\hat{c}\,(x,a,\theta)| = \frac{|c\,(x,a,\theta)|}{\tau\,(x,a,\theta)} \leq M_* W\,(x) \qquad \forall\,(x,a,\theta) \in \mathbb{K},$$

donde

$$M_* := \frac{M_2}{m}.$$

■

El modelo (3.40) satisface la Hipótesis 3.5, `mutatis mutandis`, de acuerdo al resultado a continuación.

Lema 3.4 *Bajo las Hipótesis 3.1-3.5 existe una medida de probabilidad μ sobre $\mathbb{X}$, una función s.c.s. $\hat{\phi} : \mathbb{K} \to [0,1]$, y una constante $\hat{\gamma} \in (0,1)$, tales que*

(a) $\hat{Q}(C \mid x,a,\theta) \geq \hat{\phi}\,(x,a,\theta)\,\mu\,(C)$ *para todo* $(x,a,\theta) \in \mathbb{K}$ *y* $C \in \mathcal{B}\,(\mathbb{X})$.

(b) $\int_{\mathbb{X}} W\,(y)\,\hat{Q}(dy \mid x,a,\theta) \leq \hat{\gamma} W\,(x) + \hat{\phi}\,(x,a,\theta)\,\mu\,(W)$, *donde*

$$\mu\,(W) := \int_{\mathbb{X}} W\,(y)\,\mu(dy) < \infty.$$

(c) $\int_{\mathbb{X}} \hat{\phi}\,(x, f\,(x)\,,\theta)\,\mu(dx) > 0$ *para cada* $f \in \mathbb{F}_{\mathbb{A}}$.

Demostración. Sean

$$\hat{\phi}\,(x,a,\theta) := \frac{\tau}{\tau\,(x,a,\theta)}\phi\,(x,a)\,, \qquad (x,a,\theta) \in \mathbb{K},$$
$$\hat{\gamma} := 1 - \frac{\tau}{M}\,(1-\gamma)\,.$$

Nótese que, de la expresión (3.37) junto con la Proposición 3.2 y por la Hipótesis 3.5, se cumple que $\hat{\phi} : \mathbb{K} \to [0,1]$, $\hat{\phi}$ es s.c.s. y $0 < \hat{\gamma} < 1$. Por lo tanto:

(a) la Hipótesis 3.5*(a)* implica que

$$\hat{Q}(C \mid x,a,\theta) \geq \hat{\phi}\,(x,a,\theta)\,\mu\,(C) \quad \forall\,(x,a,\theta) \in \mathbb{K}, C \in \mathcal{B}\,(\mathbb{X})\,;$$

(b) de la Hipótesis 3.5*(b)* se obtiene

$$\int_{\mathbb{X}} W(y)\,\hat{Q}(dy \mid x,a,\theta) \leq \hat{\gamma} W(x) + \hat{\phi}(x,a,\theta) \int_{\mathbb{X}} W(y)\,\mu(dy);$$

y finalmente,

(c) por la Hipótesis 3.5*(c)* concluimos que para cada $f \in \mathbb{F}_{\mathbb{A}}$,

$$\int_{\mathbb{X}} \hat{\phi}(x, f(x), \theta)\,\mu(dx) > 0.$$

■

Ahora, como consecuencia del Lema 3.4, estamos en condiciones de establecer el siguiente resultado para el modelo minimax markoviano (3.40), el cual tomamos del Teorema 5.2 en [6] y reproducimos a continuación (véase además [24, 25]).

Proposicin 3.3 *Bajo las Hipótesis 3.1-3.5 existe una constante $\hat{\jmath}$, una función $\hat{h} \in \mathbb{L}_W(\mathbb{X})$ y una política estacionaria $\hat{f} \in \mathbb{F}_{\mathbb{A}}$, tal que para toda $x \in \mathbb{X}$,*

$$\begin{aligned} \hat{\jmath} + \hat{h}(x) &= \operatorname*{ínf}_{a \in A(x)} \sup_{\theta \in \Theta(x,a)} \left\{ \hat{c}(x,a,\theta) \right. \\ &\qquad \left. + \int_{\mathbb{X}} \hat{h}(y)\,\hat{Q}(dy \mid x,a,\theta) \right\} \qquad (3.41) \\ &= \sup_{\theta \in \Theta\left(x,\hat{f}(x)\right)} \left\{ \hat{c}\left(x, \hat{f}(x), \theta\right) \right. \\ &\qquad \left. + \int_{\mathbb{X}} \hat{h}(y)\,\hat{Q}(dy \mid x, \hat{f}(x), \theta) \right\}. \qquad (3.42) \end{aligned}$$

3.7.1. Demostración del Teorema 3.2

Primero, de (3.41) tenemos que para cada $(x,a) \in \mathbb{K}_A$,

$$\begin{aligned} \hat{\jmath} + \hat{h}(x) \leq \sup_{\theta \in \Theta(x,a)} &\left\{ \hat{c}(x,a,\theta) \right. \\ &\left. + \int_{\mathbb{X}} \hat{h}(y)\,\hat{Q}(dy \mid x,a,\theta) \right\}. \end{aligned}$$

Observe además que, de la expresión (3.39) y por la Hipótesis 3.5*(b)*, la función

$$\theta \longmapsto \int_{\mathbb{X}} \hat{h}\,(y)\,\hat{Q}(dy \mid x, a, \theta)$$

es continua para todo $(x, a) \in \mathbb{K}_A$. En tal situación, a consecuencia de la Hipótesis 3.2*(e)* y aplicando un teorema de selección medible apropiado (véase, por ejemplo, [43]), para cada $\varepsilon > 0$ existe $g \in \mathbb{F}_\Theta$ tal que

$$\hat{\jmath} + \hat{h}\,(x) \leq \hat{c}\,(x, a, g\,(x, a)) + \int_{\mathbb{X}} \hat{h}\,(y)\,\hat{Q}(dy \mid x, a, g\,(x, a)) + \frac{\varepsilon}{M} \qquad \forall\,(x, a) \in \mathbb{K}_A. \tag{3.43}$$

Luego, de las expresiones (3.38) y (3.39), junto con la Hipótesis 3.4 obtenemos

$$\hat{\jmath}\tau\,(x, a, g\,(x, a)) \leq c\,(x, a, g\,(x, a)) + \int_{\mathbb{X}} \tau\hat{h}\,(y)\,Q(dy \mid x, a) - \tau\hat{h}\,(x) + \varepsilon,$$

lo cual implica

$$h^*\,(x) \leq c\,(x, a, g\,(x, a)) - j^*\tau\,(x, a, g\,(x, a)) + \int_{\mathbb{X}} h^*\,(y)\,Q(dy \mid x, a) + \varepsilon \tag{3.44}$$

con

$$j^* := \hat{\jmath} \quad \text{y} \quad h^*\,(\cdot) := \tau\hat{h}\,(\cdot)\,. \tag{3.45}$$

Dado que ε es arbitrario, tenemos

$$h^*\,(x) \leq \sup_{\theta \in \Theta(x,a)} \left\{ c\,(x, a, \theta) - j^*\tau\,(x, a, \theta) + \int_{\mathbb{X}} h^*\,(y)\,Q(dy \mid x, a) \right\} \qquad \forall\,(x, a) \in \mathbb{K}_A,$$

y, por lo tanto,

$$h^*\,(x) \leq \inf_{a \in A(x)} \sup_{\theta \in \Theta(x,a)} \left\{ c\,(x, a, \theta) - j^*\tau\,(x, a, \theta) + \int_{\mathbb{X}} h^*\,(y)\,Q(dy \mid x, a) \right\} \qquad \forall x \in \mathbb{X}. \tag{3.46}$$

Ahora, por la expresión (3.42), existe $\hat{f} \in \mathbb{F}_{\mathbb{A}}$ tal que para todo $x \in \mathbb{X}$,

$$\hat{\jmath} + \hat{h}\,(x) = \sup_{\theta \in \Theta\left(x, \hat{f}(x)\right)} \left\{ \hat{c}\left(x, \hat{f}\,(x)\,, \theta\right) + \int_{\mathbb{X}} \hat{h}\,(y)\,\hat{Q}(dy \mid x, \hat{f}\,(x)\,, \theta) \right\}.$$

De aquí, para todo $x \in \mathbb{X}$ y $\theta \in \Theta\left(x, \hat{f}(x)\right)$,

$$\hat{\jmath} + \hat{h}(x) \geq \hat{c}\left(x, \hat{f}(x), \theta\right) + \int_{\mathbb{X}} \hat{h}(y)\, \hat{Q}(dy \mid x, \hat{f}(x), \theta),$$

de donde, como en (3.44),

$$h^*(x) \geq c\left(x, \hat{f}(x), \theta\right) - j^* \tau\left(x, \hat{f}(x), \theta\right) + \int_{\mathbb{X}} h^*(y)\, Q(dy \mid x, \hat{f}(x)).$$

Dado que θ es arbitrario en $\Theta\left(x, \hat{f}(x)\right)$, tenemos que

$$h^*(x) \geq \sup_{\theta \in \Theta\left(x, \hat{f}(x)\right)} \left\{ c\left(x, \hat{f}(x), \theta\right) - j^* \tau\left(x, \hat{f}(x), \theta\right) + \int_{\mathbb{X}} h^*(y)\, Q(dy \mid x, \hat{f}(x)) \right\},$$

y esto implica que

$$h^*(x) \geq \operatorname*{ínf}_{a \in A(x)} \sup_{\theta \in \Theta(x,a)} \left\{ c(x, a, \theta) - j^* \tau(x, a, \theta) + \int_{\mathbb{X}} h^*(y)\, Q(dy \mid x, a) \right\}. \tag{3.47}$$

Luego, combinando las expresisones (3.46) y (3.47) obtenemos

$$h^*(x) = \operatorname*{ínf}_{a \in A(x)} \sup_{\theta \in \Theta(x,a)} \left\{ c(x, a, \theta) - j^* \tau(x, a, \theta) + \int_{\mathbb{X}} h^*(y)\, Q(dy \mid x, a \right\}, \quad x \in \mathbb{X}. \tag{3.48}$$

Además, observe que las expresiones en (3.45), correspondientes a la función $h^* \in \mathbb{L}_W(\mathbb{X})$ y a la constante j^*, junto con las Proposiciones 3.2 y 3.3 demuestran la existencia de $f^* \in \mathbb{F}_{\mathbb{A}}$ tal que para todo $x \in \mathbb{X}$,

$$h^*(x) = \sup_{\theta \in \Theta(x, f^*(x))} \left\{ c(x, f^*(x), \theta) - j^* \tau(x, f^*(x), \theta) + \int_{\mathbb{X}} h^*(y)\, Q(dy \mid x, f^*(x)) \right\}. \tag{3.49}$$

Para concluir, demostraremos que j^* es el costo óptimo promedio-minimax y f^* es una política promedio-minimax. Para este fin, nótese que partiendo de (3.49), para todo $\theta \in \Theta\left(x, f^*\left(x\right)\right)$, $x \in \mathbb{X}$,

$$h^*\left(x\right) \geq c\left(x, f^*\left(x\right), \theta\right) - j^* \tau\left(x, f^*\left(x\right), \theta\right) + \int_{\mathbb{X}} h^*\left(y\right) Q(dy \mid x, f^*\left(x\right)). \tag{3.50}$$

Sea $\pi' \in \Pi_\Theta$ una política arbitraria. Entonces iterando la expresión (3.50) obtenemos

$$\begin{aligned} h^*\left(x\right) \geq\ & E_x^{f^*\pi'}\left[\sum_{k=0}^{n-1} c\left(x_k, a_k, \theta_k\right)\right] \\ & - j^* E_x^{f^*\pi'}\left[\sum_{k=0}^{n-1} \tau\left(x_k, a_k, \theta_k\right)\right] \\ & + E_x^{f^*\pi'}\left[h^*\left(x_n\right)\right], \quad n \in \mathbb{N}, \end{aligned} \tag{3.51}$$

lo cual, a su vez, implica

$$j^* \geq \frac{E_x^{f^*\pi'}\left[\sum_{k=0}^{n-1} c\left(x_k, a_k, \theta_k\right)\right] + E_x^{f^*\pi'}\left[h^*\left(x_n\right)\right] - h^*\left(x\right)}{E_x^{f^*\pi'}\left[\sum_{k=0}^{n-1} \tau\left(x_k, a_k, \theta_k\right)\right]}, \tag{3.52}$$

donde

$$E_x^{f^*\pi'}\left[h^*\left(x_n\right)\right] = \int_{\mathbb{X}} h^*\left(y\right) Q^n(dy \mid x, f^*\left(x\right)).$$

Por otro lado (véase (3.33) e Hipótesis 3.4), observe que

$$nm \leq E_x^{f^*\pi'}\left[\sum_{k=0}^{n-1} \tau\left(x_k, a_k, \theta_k\right)\right] \leq nM, \tag{3.53}$$

y además, por (3.36) y el hecho que $h^* \in \mathbb{L}_W\left(\mathbb{X}\right)$, tenemos

$$\frac{E_x^{f^*\pi'}\left[h^*\left(x_n\right)\right]}{n} \to 0 \quad \text{cuando } n \to \infty. \tag{3.54}$$

Entonces, tomando lím sup cuando $n \to \infty$ en (3.52) y aplicando (3.53) y (3.54), se sigue que (véase (3.9))

$$j^* \geq J\left(x, f^*\left(x\right), \pi'\right). \tag{3.55}$$

Como π' es arbitraria, obtenemos

$$j^* \geq \sup_{\pi' \in \Pi_\Theta} J\left(x, f^*(x), \pi'\right) \geq \inf_{\pi \in \Pi_\mathbb{A}} \sup_{\pi' \in \Pi_\Theta} J\left(x, \pi, \pi'\right). \tag{3.56}$$

Por otra parte, de (3.48), para todo $(x, a) \in \mathbb{K}_A$,

$$h^*(x) \leq \sup_{\theta \in \Theta(x,a)} \left\{ c(x, a, \theta) - j^* \tau(x, a, \theta) + \int_\mathbb{X} h^*(y)\, Q(dy \mid x, a) \right\}.$$

Luego, por la expresión (3.43), para $\varepsilon_0 > 0$ existe $g \in \mathbb{F}_\Theta$ tal que

$$h^*(x) \leq c(x, a, g(x, a)) - j^* \tau(x, a, g(x, a)) + \int_\mathbb{X} h^*(y)\, Q(dy \mid x, a) + \varepsilon_0.$$

Si $\pi \in \Pi_\mathbb{A}$ es una política arbitraria, entonces aplicando argumentos similares como en (3.51)-(3.55) y usando el hecho de que (véase (3.33) e Hipótesis 3.4)

$$nM \geq E_x^{\pi g}\left[\sum_{k=0}^{n-1} \tau(x_k, a_k, \theta_k)\right] \geq nm, \quad n \in \mathbb{N},$$

tenemos que

$$h^*(x) \leq E_x^{\pi g}\left[\sum_{k=0}^{n-1} c(x_k, a_k, \theta_k)\right] - j^* E_x^{\pi g}\left[\sum_{k=0}^{n-1} \tau(x_k, a_k, \theta_k)\right] + E_x^{\pi g}\left[h^*(x_n)\right] + n\varepsilon_0,$$

de donde

$$j^* \leq J(x, \pi, g) \leq \sup_{\pi' \in \Pi_\Theta} J\left(x, \pi, \pi'\right).$$

Puesto que π es arbitraria, de lo anterior se deduce que

$$j^* \leq J(x, \pi, g) \leq \inf_{\pi \in \Pi_\mathbb{A}} \sup_{\pi' \in \Pi_\Theta} J\left(x, \pi, \pi'\right). \tag{3.57}$$

Y finalmente, combinando (3.56) y (3.57) obtenemos

$$j^* = \sup_{\pi' \in \Pi_\Theta} J\left(x, f^*, \pi'\right) = \inf_{\pi \in \Pi_\mathbb{A}} \sup_{\pi' \in \Pi_\Theta} J\left(x, \pi, \pi'\right),$$

lo cual completa la demostración.

■

Capítulo 4

Estimación de la Desviación de Optimalidad en $MCSMs$ Descontados

4.1. Introducción

Siguiendo en el escenario de que la distribución del tiempo de permanencia F en el $MCSM$ (1.1) es desconocida, en este capítulo nos centraremos en resolver el siguiente problema. Supongamos que es posible obtener una distribución "aproximante" $\tilde{F}(\cdot \mid x, a)$ de $F(\cdot \mid x, a)$. Entonces, considerando el PCO asociado a $\tilde{F}$ y el correspondiente costo total descontado $\tilde{V}$, podemos obtener una política $\tilde{\pi}$ que minimice a $\tilde{V}$. Luego, si el controlador usa $\tilde{\pi}$ para controlar el proceso semi-markoviano original, nuestro objetivo principal consiste en estimar la *desviación de optimalidad* de la política de control $\tilde{\pi}$ definida como

$$\mathcal{D}(x, \tilde{\pi}) := V(x, \tilde{\pi}) - \inf_{\pi \in \Pi} V(x, \pi). \tag{4.1}$$

Específicamente, nuestro resultado principal en términos generales establece que

$$\mathcal{D}(x, \tilde{\pi}) \leq \Lambda\left(\partial\left(F, \tilde{F}\right)\right),$$

donde $\Lambda : [0, \infty) \to [0, \infty)$ es una función tal que $\Lambda(s) \to 0$ cuando $s \to 0^+$, y ∂ es alguna "distancia" entre F y $\tilde{F}$.

4.2. Condiciones sobre el modelo de control

Consideremos el $MCSM$ (1.1)

$$\mathcal{M} := (\mathbb{X}, \mathbb{A}, \{A(x) : x \in \mathbb{X}\}, Q, F, D, d),$$

como fue descrito en el Capítulo 1, y el $MCSM$ correspondiente a la distribución "aproximante" $\tilde{F}$, denotado por

$$\tilde{\mathcal{M}} := (\mathbb{X}, \mathbb{A}, \{A(x) : x \in \mathbb{X}\}, Q, \tilde{F}, D, d);$$

asociado a una función de distribución $\tilde{F}(\cdot \mid x, a)$, donde las componentes $\mathbb{X}$, $\mathbb{A}$, $A(x)$, Q, D y d son como en el modelo $\mathcal{M}$. Asimismo, asociado con $\tilde{\mathcal{M}}$ usaremos además la notación siguiente (véase Definición 1.5, Proposición 1.1*(a)* y las expresiones (1.8), (1.9) y (1.14)). Para cada $(x, a) \in \mathbb{K}_A$ (véase (1.2)),

$$\tilde{c}_\alpha(x, a) := D(x, a) + \tilde{\tau}_\alpha(x, a)\, d(x, a),$$

lo cual representa el costo por etapa, donde

$$\tilde{\tau}_\alpha(x, a) := \frac{1 - \tilde{\Delta}_\alpha(x, a)}{\alpha} \tag{4.2}$$

y

$$\tilde{\Delta}_\alpha(x, a) := \int_0^\infty \exp(-\alpha t)\, \tilde{F}(dt \mid x, a). \tag{4.3}$$

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

$$\tilde{V}(x, \pi) := E_x^\pi \left[\sum_{n=0}^{\infty} \exp\left(-\alpha \tilde{T}_n\right) \tilde{c}_\alpha(x_n, a_n) \right]$$

es el costo total esperado descontado cuando se usa la política $\pi \in \Pi$ y el estado inicial es $x \in \mathbb{X}$, donde $\tilde{T}_n := \tilde{T}_{n-1} + \tilde{\delta}_n$, $n \in \mathbb{N}$, $\tilde{T}_0 := 0$, y $\tilde{\delta}_n$ es la variable aleatoria que representa el tiempo de permanencia del sistema en el estado x_{n-1}, y cuya distribución es $\tilde{F}$. Finalmente,

$$\tilde{V}(x) := \inf_{\pi \in \Pi} \tilde{V}(x, \pi)$$

es la correspondiente función de valor óptimo descontado.

De manera similar a los capítulos anteriores, requerimos dos clases de hipótesis. La primera de ellas es la condición de regularidad semejante a la Hipótesis 1.1 en el Capítulo 1; mientras que, la segunda contiene las condiciones necesarias para la existencia de minimizadores.

Hiptesis 4.1 *Existen constantes positivas $\zeta, \tilde{\zeta}, \epsilon$ y $\tilde{\epsilon}$ tal que, para cada $(x, a) \in \mathbb{K}_A$,*

$$F(\zeta \mid x, a) \leq 1 - \epsilon \quad \text{y} \quad \tilde{F}(\tilde{\zeta} \mid x, a) \leq 1 - \tilde{\epsilon}.$$

En un esquema similar al de la parte *(a)* en la Proposición 1.1, se tiene el resultado siguiente.

Proposicin 4.1 *Bajo la Hipótesis 4.1 se tiene que*

$$\rho_\alpha := \sup_{(x,a)\in\mathbb{K}_A} \Delta_\alpha(x,a) < 1 \quad y \quad \tilde{\rho}_\alpha := \sup_{(x,a)\in\mathbb{K}_A} \tilde{\Delta}_\alpha(x,a) < 1.$$

Hiptesis 4.2 (a) *Las funciones $F(t \mid x,a)$ y $\tilde{F}(t \mid x,a)$ son continuas en $A(x)$ para cada $x \in \mathbb{X}$ y $t \geq 0$.*

(b) Para cada $x \in \mathbb{X}$, $A(x)$ es un conjunto compacto, y las funciones de costo $D(x,a)$ y $d(x,a)$ son s.c.i. en $A(x)$. Además, existe una función medible $W : \mathbb{X} \to [1,\infty)$ y constantes positivas β_1, c_1 y c_2 tal que

$$1 \leq \beta_1 < \min\{1/\rho_\alpha, 1/\tilde{\rho}_\alpha\}, \tag{4.4}$$

$$\sup_{a\in A(x)} |D(x,a)| \leq c_1 W(x), \quad \sup_{a\in A(x)} |d(x,a)| \leq c_2 W(x) \quad \forall x \in \mathbb{X}, \tag{4.5}$$

$$\int_{\mathbb{X}} W(y)\, Q(dy \mid x,a) \leq \beta_1 W(x), \quad (x,a) \in \mathbb{K}_A. \tag{4.6}$$

(c) Para cada $x \in \mathbb{X}$,

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

es una función continua en $A(x)$.

(d) La ley de transición $Q(\cdot \mid x,a)$ es débilmente continua en $A(x)$.

Como una consecuencia de las Hipótesis 4.1-4.2 (véase, por ejemplo, [32]) tenemos el resultado siguiente.

Proposicin 4.2 *Supongamos que se cumplen las Hipótesis 4.1-4.2. Entonces,*

(a) Existen políticas óptimas estacionarias f^* y $\tilde{f}$ para los modelos $\mathcal{M}$ y $\tilde{\mathcal{M}}$, respectivamente.

(b) Las funciones de valor óptimo V^* y $\tilde{V}$ satisfacen la ecuación de optimalidad correspondiente, es decir, para cada $x \in \mathbb{X}$:

$$V^*(x) = \min_{a\in A(x)} \left\{ c_\alpha(x,a) + \Delta_\alpha(x,a) \int_{\mathbb{X}} V^*(y)\, Q(dy \mid x,a) \right\},$$

y

$$\tilde{V}(x) = \min_{a\in A(x)} \left\{ \tilde{c}_\alpha(x,a) + \tilde{\Delta}_\alpha(x,a) \int_{\mathbb{X}} \tilde{V}(y)\, Q(dy \mid x,a) \right\}.$$

Obsérvese ahora que, de acuerdo a la Proposición 4.2, para cada $x \in \mathbb{X}$ tenemos

$$V^*(x) = \inf_{\pi \in \Pi} V(x, \pi) = V(x, f^*) \tag{4.7}$$

y

$$\tilde{V}(x) = \inf_{\pi \in \Pi} \tilde{V}(x, \pi) = \tilde{V}\left(x, \tilde{f}\right). \tag{4.8}$$

4.3. Resultados principales

4.3.1. Estimación de la desviación de optimalidad

En el contexto de la Proposición 4.2, nuestro objetivo consiste en estimar la desviación de optimalidad de la política estacionaria $\tilde{f}$. Así pues, podemos establecer nuestro resultado principal como sigue.

Teorema 4.1 *Supongamos que las Hipótesis 4.1-4.2 se satisfacen. Entonces existe una constante positiva M^* tal que para toda $x \in \mathbb{X}$,*

$$\mathcal{D}\left(x, \tilde{f}\right) := V\left(x, \tilde{f}\right) - V^*(x) \leq M^* \left\|\tilde{F} - F\right\| W(x); \tag{4.9}$$

donde

$$\left\|\tilde{F} - F\right\| := \sup_{(x,a) \in \mathbb{K}_A} \int_0^\infty \exp(-\alpha t) \left|\mu_F(\cdot \mid x, a) - \mu_{\tilde{F}}(\cdot \mid x, a)\right|(dt), \tag{4.10}$$

y $\left|\mu_F(\cdot \mid x, a) - \mu_{\tilde{F}}(\cdot \mid x, a)\right|$ representa la variación total de la medida con signo $\mu_F(\cdot \mid x, a) - \mu_{\tilde{F}}(\cdot \mid x, a)$, siendo μ_F y $\mu_{\tilde{F}}$ las medidas inducidas por las funciones de distribución F y $\tilde{F}$, respectivamente, para cada par fijo $(x, a) \in \mathbb{K}_A$.

4.3.2. Optimalidad asintótica

Como en el trabajo [29] de Luque-Vásquez y Minjárez-Sosa, (véase Sección 2.2, Capítulo 2), consideraremos ahora el caso especial que establece la siguiente condición.

Hiptesis 4.3 *Las distribuciones F y $\tilde{F}$ tienen densidades g y $\tilde{g}$, respectivamente, las cuales son independientes de los pares estado-acción (x, a); es decir,*

$$F(t \mid x, a) = \int_0^t g(s)\, ds, \qquad y \qquad \tilde{F}(t \mid x, a) = \int_0^t \tilde{g}(s)\, ds.$$

Observe que, bajo esta hipótesis tenemos que Δ_α y τ_α, asi como $\tilde{\Delta}_\alpha$ y $\tilde{\tau}_\alpha$ son a su vez independientes de (x, a); esto es,

$$\Delta_\alpha(x,a) := \Delta_\alpha = \int_0^\infty \exp(-\alpha s)\, g(s)\, ds,$$
$$\tau_\alpha := \frac{1-\Delta_\alpha}{\alpha},$$

y similarmente, para $\tilde{\Delta}_\alpha$ y $\tilde{\tau}_\alpha$, (véase (4.2) y (4.3)).

Además, note que

$$\tilde{\Delta}_\alpha - \Delta_\alpha = \int_0^\infty \exp(-\alpha s)\, \{\tilde{g}(s) - g(s)\}\, ds,$$

y en tal caso, la desigualdad (4.9) toma la forma

$$\mathcal{D}\left(x, \tilde{f}\right) \leq M^* \|\tilde{g} - g\|\, W(x) \qquad \forall x \in \mathbb{X}, \tag{4.11}$$

donde

$$\|\tilde{g} - g\| := \int_0^\infty \exp(-\alpha s)\, |\tilde{g}(s) - g(s)|\, ds.$$

En particular (ver Sección 2.2), suponiendo que los tiempos de permanencia son observables, sean $\delta_1, ..., \delta_n$ realizaciones independientes de variables aleatorias, observadas hasta el momento de la n-ésima época de decisión, con la densidad desconocida g, y sea $g_n(s) := g_n(s; \delta_1, ..., \delta_n)$, $s \in \mathbb{R}_+$, un estimador arbitrario de g tal que

$$E\left[\int_0^\infty |g_n(s) - g(s)|\, ds\right] \to 0 \quad \text{cuando } n \to \infty, \tag{4.12}$$

donde g_n es una densidad.

Denotemos ahora por

$$f_n := f_n^{g_n} \tag{4.13}$$

a la política estacionaria óptima correspondiente a la densidad g_n, en cuyo caso el proceso de minimización se lleva a cabo casi seguramente ($c.s.$) respecto a la medida de probabilidad P inducida por la distribución común de las variables aleatorias $\delta_1, ..., \delta_n$. Luego, aplicando (4.11) con g_n en lugar de $\tilde{g}$, obtenemos la desviación de optimalidad esperada

$$E[\mathcal{D}(x, f_n)] \leq M^* W(x)\, E[\|g - g_n\|] \to 0 \quad \text{cuando } n \to \infty,\ x \in \mathbb{X}, \tag{4.14}$$

donde E es el operador esperanza respecto a la medida de probabilidad P. Además, como una consecuencia de (4.14) tenemos el siguiente resultado.

Teorema 4.2 *Bajo la Hipótesis 4.3, la política $\hat{\pi} = \{f_n\}$ es asintóticamente óptima descontada puntual, lo cual significa que para cada $x \in \mathbb{X}$,*

$$E\left[\Phi\left(x, f_n\right)\right] \to 0 \quad \textit{cuando } n \to \infty,$$

donde Φ es la función de discrepancia definida en (2.7).

4.3.3. Resultados preliminares

Para la demostración de los Teoremas 4.1 y 4.2 usaremos los siguientes resultados. Algunos de ellos ya fueron usados en los capítulos anteriores, pero para una fácil referencia los incluimos de nuevo.

Para cada $u \in \mathbb{B}_W\left(\mathbb{X}\right)$ y $(x,a) \in \mathbb{K}_A$, de (4.6) y de la definición (1.16) de $\|\cdot\|_W$ tenemos

$$|u(x)| \leq \|u\|_W W\left(x\right) \tag{4.15}$$

y

$$\int_{\mathbb{X}} u\left(y\right) Q(dy \mid x,a) \leq \beta_1 \|u\|_W W\left(x\right). \tag{4.16}$$

Además (ver (1.20)),

$$\sup_{n\in\mathbb{N}_0} E_x^{\pi}\left[W\left(x_n\right)\right] < \infty \tag{4.17}$$

y, además, existen constantes positivas M' y $\tilde{M}'$ tales que, para todo $\pi \in \Pi$ y $x \in \mathbb{X}$, (véase (2.8)):

$$V^*\left(x\right) \leq V\left(x,\pi\right) \leq M'W\left(x\right) \qquad \text{y} \qquad \tilde{V}\left(x\right) \leq \tilde{V}\left(x,\pi\right) \leq \tilde{M}'W\left(x\right).$$

Por otra parte, obsérvese que para todo $(x,a) \in \mathbb{K}_A$,

$$\left|\tilde{\Delta}_\alpha(x,a) - \Delta_\alpha(x,a)\right| \leq \left\|\tilde{F} - F\right\|. \tag{4.18}$$

De aquí, y por definición de los costos c_α y $\tilde{c}_\alpha$, junto con (4.5) obtenemos,

$$|c_\alpha(x,a) - \tilde{c}_\alpha(x,a)| \leq \frac{c_2}{\alpha}\left\|\tilde{F} - F\right\| W\left(x\right) \quad \forall\,(x,a) \in \mathbb{K}_A. \tag{4.19}$$

Además, para cada función $u \in \mathbb{B}_W\left(\mathbb{X}\right)$, de la expresiones (4.16) y (4.18), y por definición de ρ_α (véase Proposición 4.1), tenemos que para todo $(x,a) \in \mathbb{K}_A$

$$\left|\tilde{\Delta}_\alpha(x,a) - \Delta_\alpha(x,a)\right| \int_{\mathbb{X}} u\left(y\right) Q(dy \mid x,a) \leq \beta_1 \left\|\tilde{F} - F\right\| \|u\|_W W\left(x\right) \tag{4.20}$$

y

$$\Delta_\alpha(x,a)\int_{\mathbb{X}} u\,(y)\,Q(dy\mid x,a)\leq \beta_1\rho_\alpha\,\|u\|_W\,W\,(x)\,. \tag{4.21}$$

De forma similar, reemplazando Δ_α y ρ_α por $\tilde{\Delta}_\alpha$ y $\tilde{\rho}_\alpha$, respectivamente, en las desigualdades (4.20) y (4.21), obtenemos expresiones semejantes, en particular,

$$\tilde{\Delta}_\alpha(x,a)\int_{\mathbb{X}} u\,(y)\,Q(dy\mid x,a)\leq \beta_1\tilde{\rho}_\alpha\,\|u\|_W\,W\,(x)\,.$$

Por lo anterior, usando la notación

$$V_{\tilde{f}}\,(x):=V\left(x,\tilde{f}\right),$$

donde $\tilde{f}$ es la política óptima estacionaria para el modelo $\tilde{\mathcal{M}}$, de la desigualdad (4.15) obtenemos

$$\begin{aligned}\tilde{\Delta}_\alpha(x,\tilde{f})\int_{\mathbb{X}}\left|V\left(y,\tilde{f}\right)-\tilde{V}\left(y,\tilde{f}\right)\right|Q(dy\mid x,\tilde{f})&\\ \leq \beta_1\tilde{\rho}_\alpha\left\|V_{\tilde{f}}-\tilde{V}\right\|_W W\,(x)\,,\quad x\in\mathbb{X}.&\end{aligned} \tag{4.22}$$

Lema 4.1 *Existen constantes positivas M_1^* y M_2^*, tales que*

$$\left\|V_{\tilde{f}}-\tilde{V}\right\|_W\leq M_1^*\left\|\tilde{F}-F\right\| \tag{4.23}$$

y

$$\left\|\tilde{V}-V^*\right\|_W\leq M_2^*\left\|\tilde{F}-F\right\|. \tag{4.24}$$

Demostración. Primero obsérvese que de la propiedad de Markov, para una política estacionaria $f\in\mathbb{F}$ y cada $x\in\mathbb{X}$,

$$V\,(x,f)=c_\alpha\,(x,f)+\Delta_\alpha\,(x,f)\int_{\mathbb{X}} V\,(y,f)\,Q(dy\mid x,f).$$

Similarmente para $\tilde{V}$. Entonces,

$$\begin{aligned}\left|V_{\tilde{f}}\,(x)-\tilde{V}\,(x)\right|&=\left|V\left(x,\tilde{f}\right)-\tilde{V}\left(x,\tilde{f}\right)\right|\\ &\leq\left|\left\{c_\alpha\left(x,\tilde{f}\right)+\Delta_\alpha\left(x,\tilde{f}\right)\int_{\mathbb{X}} V\left(y,\tilde{f}\right)Q(dy\mid x,\tilde{f})\right\}\right.\\ &\quad\left.-\left\{\tilde{c}_\alpha\left(x,\tilde{f}\right)+\tilde{\Delta}_\alpha\left(x,\tilde{f}\right)\int_{\mathbb{X}}\tilde{V}\left(y,\tilde{f}\right)Q(dy\mid x,\tilde{f})\right\}\right|.\end{aligned}$$

Luego, sumando y restando el término

$$\tilde{\Delta}_\alpha\left(x,\tilde{f}\right)\int_{\mathbb{X}} V\left(y,\tilde{f}\right)Q(dy\mid x,\tilde{f})$$

obtenemos que para cada $x\in\mathbb{X}$,

$$\begin{aligned}\left|V_{\tilde{f}}(x)-\tilde{V}(x)\right| &\leq \left|c_\alpha\left(x,\tilde{f}\right)-\tilde{c}_\alpha\left(x,\tilde{f}\right)\right| \\ &\quad+\left|\Delta_\alpha\left(x,\tilde{f}\right)-\tilde{\Delta}_\alpha\left(x,\tilde{f}\right)\right|\int_{\mathbb{X}}\left|V\left(y,\tilde{f}\right)\right|Q(dy\mid x,\tilde{f}) \\ &\quad+\tilde{\Delta}_\alpha\left(x,\tilde{f}\right)\int_{\mathbb{X}}\left|V\left(y,\tilde{f}\right)-\tilde{V}\left(x,\tilde{f}\right)\right|Q(dy\mid x,\tilde{f})\,.\end{aligned}$$

Por lo cual, dado que $W(\cdot)\geq 1$, de las expresiones (4.19), (4.20) y (4.21), tenemos

$$\begin{aligned}\left\|V_{\tilde{f}}-\tilde{V}\right\|_W &\leq \frac{c_2}{\alpha}\left\|\tilde{F}-F\right\|+\beta_1\left\|V_{\tilde{f}}\right\|_W\left\|\tilde{F}-F\right\| \\ &\quad+\beta_1\tilde{\rho}_\alpha\left\|V_{\tilde{f}}-\tilde{V}\right\|_W. \qquad (4.25)\end{aligned}$$

Ahora, como $\beta_1\tilde{\rho}_\alpha<1$ (véase (4.4)), de la desigualdad (4.25) obtenemos

$$\left\|V_{\tilde{f}}-\tilde{V}\right\|_W\leq M_1^*\left\|\tilde{F}-F\right\|,$$

donde

$$M_1^*:=\frac{\frac{c_2}{\alpha}+\beta_1\left\|V_{\tilde{f}}\right\|_W}{1-\beta_1\tilde{\rho}_\alpha},$$

con lo cual queda demostrada la desigualdad (4.23).

Por otra parte, a efecto de demostrar la desigualdad (4.24), obsérvese que de la Proposición 4.2*(b)*,

$$\begin{aligned}\left|\tilde{V}(x)-V^*(x)\right| &= \left|\min_{a\in A(x)}\left\{\tilde{c}_\alpha(x,a)+\tilde{\Delta}_\alpha(x,a)\int_{\mathbb{X}}\tilde{V}(y)\,Q(dy\mid x,a)\right\}\right. \\ &\quad\left.-\min_{a\in A(x)}\left\{c_\alpha(x,a)+\Delta_\alpha(x,a)\int_{\mathbb{X}}V^*(y)\,Q(dy\mid x,a)\right\}\right| \\ &\leq \sup_{a\in A(x)}\left\{\left|\tilde{c}(x,a)-c(x,a)\right|\right. \\ &\quad+\left|\tilde{\Delta}_\alpha(x,a)\int_{\mathbb{X}}\tilde{V}(y)\,Q(dy\mid x,a)\right. \\ &\quad\left.\left.-\Delta_\alpha(x,a)\int_{\mathbb{X}}V^*(y)\,Q(dy\mid x,a)\right|\right\}. \qquad (4.26)\end{aligned}$$

Ahora, sumando y restando el término

$$\Delta_\alpha(x,a)\int_{\mathbb{X}}\tilde{V}(y)\,Q(dy\mid x,a),$$

y aplicando argumentos similares a los de la demostración de (4.26) obtenemos

$$\begin{aligned}\left|\tilde{V}(x)-V^*(x)\right| \leq \frac{c_2}{\alpha}\left\|\tilde{F}-F\right\|W(x)+\beta_1\left\|\tilde{V}\right\|_W\left\|\tilde{F}-F\right\|W(x)\\ +\beta_1\rho_\alpha W(x)\left\|\tilde{V}-V^*\right\|_W.\end{aligned}$$

Luego, la expresión anterior implica que

$$\left\|\tilde{V}-V^*\right\|_W \leq \frac{c_2}{\alpha}\left\|\tilde{F}-F\right\|+\beta_1\left\|\tilde{V}\right\|_W\left\|\tilde{F}-F\right\|+\beta_1\rho_\alpha\left\|\tilde{V}-V^*\right\|_W,$$

de lo cual

$$\left\|\tilde{V}-V^*\right\|_W \leq M_2^*\left\|\tilde{F}-F\right\|,$$

donde

$$M_2^* := \frac{\frac{c_2}{\alpha}+\beta_1\left\|\tilde{V}\right\|_W}{1-\beta_1\rho_\alpha}. \quad \blacksquare$$

4.3.4. Demostración del Teorema 4.1

Nótese primero que $\tilde{V}\left(x,\tilde{f}\right)=\tilde{V}(x)$, $x\in\mathbb{X}$. Entonces,

$$\begin{aligned}\mathcal{D}\left(x,\tilde{f}\right) &= V\left(x,\tilde{f}\right)-V^*(x)\\ &= \left\{V\left(x,\tilde{f}\right)-\tilde{V}\left(x,\tilde{f}\right)\right\}+\left\{\tilde{V}(x)-V^*(x)\right\}.\end{aligned}$$

Este hecho, junto con la expresión (4.15) y el Lema 4.1 implica que

$$\begin{aligned}\mathcal{D}\left(x,\tilde{f}\right) &\leq \left|V\left(x,\tilde{f}\right)-\tilde{V}\left(x,\tilde{f}\right)\right|+\left|\tilde{V}(x)-V^*(x)\right|\\ &\leq M^*\left\|\tilde{F}-F\right\|W(x) \quad \forall x\in\mathbb{X},\end{aligned}$$

donde

$$M^* := M_1^*+M_2^*. \quad \blacksquare$$

4.3.5. Demostración del Teorema 4.2

Primero (véase (2.7)), para cada $n \in \mathbb{N}$, definamos la función de discrepancia aproximada como

$$\Phi_n(x,a) := c_\alpha^n(x,a) + \Delta_\alpha^n \int_{\mathbb{X}} V_n^*(y)\, Q(dy \mid x,a) - V_n^*(x), \quad (x,a) \in \mathbb{K}_A, \tag{4.27}$$

donde (véase (2.12) y (2.13))

$$c_\alpha^n(x,a) = D(x,a) + \tau_\alpha^n d(x,a),$$

$$\tau_\alpha^n = \frac{1-\Delta_\alpha^n}{\alpha}$$

y

$$\Delta_\alpha^n = \int_0^\infty \exp(\alpha t)\, g_n(t)\, dt;$$

mientras que,

$$V_n^*(x) := \inf_{\pi\in\Pi} V_n(x,\pi), \quad x \in \mathbb{X},$$

y

$$V_n(x,\pi) := E_x^\pi \left[\sum_{k=0}^\infty \exp(-\alpha T_k)\, c_\alpha^n(x_k,a_k)\right].$$

Nótese que, para cada $n \in \mathbb{N}$, de las expresiones (4.13) y (4.27), junto con la Proposición 4.2 se tiene que, en particular,

$$\Phi_n(x,f_n) = 0 \quad c.s.,\ x \in \mathbb{X}.$$

Por lo tanto, puesto que Φ_n es una función no negativa, para cada $x \in \mathbb{X}$ podemos escribir

$$\begin{aligned}\Phi(x,f_n) &= |\Phi(x,f_n) - \Phi_n(x,f_n)| \\ &\le \sup_{a\in A(x)} |\Phi(x,a) - \Phi_n(x,a)| \quad c.s.\ \forall x \in \mathbb{X}.\end{aligned} \tag{4.28}$$

Por otra parte, de la definición de Φ y Φ_n en (2.7) y (4.27), respectivamente, tenemos

$$\begin{aligned}|\Phi(x,a) - \Phi_n(x,a)| \le |c_\alpha(x,a) - c_\alpha^n(x,a)| + |V^*(x) - V_n^*(x)| \\ + \Big|\Delta_\alpha \int_{\mathbb{X}} V^*(y)\, Q(dy \mid x,a) \\ -\Delta_\alpha^n \int_{\mathbb{X}} V_n^*(y)\, Q(dy \mid x,a)\Big| \quad c.s.\end{aligned}$$

Luego, sumando y restando el término

$$\Delta_\alpha \int_{\mathbb{X}} V_n^*(y)\, Q(dy \mid x,a)$$

obtenemos, *c.s.*,

$$\begin{aligned}|\Phi(x,a) - \Phi_n(x,a)| \leq |c_\alpha(x,a) - c_\alpha^n(x,a)| + |V^*(x) - V_n^*(x)| \\ + \Delta_\alpha \int_{\mathbb{X}} |V^*(y) - V_n^*(y)|\, Q(dy \mid x,a) \\ + |\Delta_\alpha - \Delta_\alpha^n| \int_{\mathbb{X}} |V^*(y)|\, Q(dy \mid x,a).\end{aligned}$$

Por lo tanto, usando las expresiones (4.15), (4.19), (4.20), (4.22) y (4.24), se sigue que para cada $n \in \mathbb{N}$, *c.s.* tenemos:

$$\begin{aligned}|\Phi(x,a) - \Phi_n(x,a)| \leq \frac{c_2}{\alpha} \|g_n - g\| W(x) + M_2^* \|g_n - g\| W(x) \\ + \beta_1 \rho_n M_2^* \|g_n - g\| W(x) \\ + \beta_1 \|V^*\|_W \|g_n - g\| W(x).\end{aligned}$$

De lo anterior, considerando que $\beta_1 \rho_n < 1$ *c.s.* y por la desigualdad (4.28), para cada $x \in \mathbb{X}$ tenemos lo siguiente

$$E[\Phi(x, f_n)] \leq M_0^* E[\|g_n - g\|] W(x) \qquad \forall n \in \mathbb{N},$$

donde

$$M_0^* := \frac{c_2}{\alpha} + 2M_2^* + \beta_1 \|V^*\|_W.$$

En consecuencia, dado que $\Phi(\cdot,\cdot) \geq 0$, de (4.12) obtenemos que para cada $x \in \mathbb{X}$,

$$E[\Phi(x, f_n)] \to 0 \qquad \text{cuando } n \to \infty. \quad \blacksquare$$

Capítulo 5

Conclusiones y Problemas Abiertos

En este trabajo se estudiaron $MCSMs$ considerando que el controlador desconoce la distribución del tiempo de permanencia F. El estudio se realizó analizando el problema de control óptimo correspondiente desde tres escenarios distintos. En el primero de ellos se supuso que la distribución del tiempo de permanencia F tiene una densidad (desconocida), la cual es independiente del estado y control. Después se consideró que F, además de depender del estado y control, también depende de un parámetro desconocido, el cual es no observable y que puede cambiar de etapa en etapa. Finalmente, se estudió el problema de estimación de la optimalidad de políticas que son óptimas para el modelo correspondiente a una función de distribución aproximada $\tilde{F}$.

Es importante señalar que el estudio de estos problemas inició con el artículo [29] de Luque-Vásquez y Minjárez-Sosa, quienes consideraron el caso de criterio de costo descontado, y en este sentido, nuestro trabajo toma como punto de partida sus resultados.

En [29] se consideró que F tiene una densidad g que es independiente de los pares estado-acción (x, a), suposición bajo la cual fue posible implementar métodos de estimación estadística y control, similares al caso de control adaptado para procesos Markovianos (véase, por ejemplo, [9, 10, 14, 17, 18, 19, 34, 35, 36, 37, 38, 39]). Luego, bajo este mismo escenario se analizó el problema para el criterio de costo promedio.

Aún cuando la condición sobre la independencia de F respecto a (x, a) puede ser una hipótesis fuerte, existen algunos problemas que la satisfacen, tal como es el caso del ejemplo incluido al final del Capítulo 2, (véase también [42]). No obstante, tomando en consideración dicho aspecto, en el Capítulo 3 proponemos una alternativa que conlleva a evitar esta suposición, pero considerando un problema paramétrico. Este enfoque consiste en modelar el problema correspondiente al $MCSM$ como un problema de control minimax semi-markoviano, —conocido también como *juego contra la naturaleza*—, donde precisamente, la naturaleza elige el parámetro desconocido en cada etapa. Un problema importante que se puede abordar en un futuro consiste

en suponer que la distribución F depende de (x, a) y que, además, es completamente desconocida, es decir, considerar el caso no paramétrico. Para este problema, en lugar de tener un conjunto de parámetros $\Theta(x, a)$ como es nuestro caso, habría que considerar un conjunto de distribuciones de probabilidad, —a partir del cual la naturaleza seleccionaría sus acciones—, y en consecuencia se tendrían que imponer condiciones de σ-compacidad sobre este nuevo conjunto, entre otros aspectos, lo cual hace esta situación más interesante desde el punto de vista matemático.

Otra forma de evitar la hipótesis de independencia de F respecto a (x, a), consiste en plantear el problema como fue hecho en el Capítulo 4. Es decir, tener manera de medir la calidad de la aproximación de una función de distribución conocida $\tilde{F}$, en términos de la optimalidad de la política correspondiente a $\tilde{F}$.

Vale la pena mencionar que en este último escenario consideramos el caso particular en que las funciones de distribución F y $\tilde{F}$ tienen, respectivamente, densidad independiente del estado y control, y nuestros resultados también demuestran optimalidad asintótica como fue obtenida por Luque-Vásquez y Minjárez-Sosa en [29].

Un problema que queda pendiente en este enfoque consiste en estimar la desviación de optimalidad considerando el criterio de costo promedio. Cabe señalar que, para este caso, las principales dificultades se presentan en la expresión del índice de funcionamiento, ya que el denominador contempla el tiempo de permanencia, así como en el proceso de transformación del $MCSM$ a un modelo de control markoviano.

Por otra parte, en [40] Minjárez-Sosa J.A. y Luque-Vásquez F. analizaron el problema de la construcción de políticas óptimas en juegos semi-markovianos de suma cero para el caso descontado, donde se considera que la distribución del tiempo de permanencia es desconocida por uno de los jugadores. Un problema que aún está abierto consiste en analizar el criterio de optimalidad promedio.

Los problemas analizados en el presente trabajo dieron lugar a las publicaciones [30, 31, 44], tomando como base la bibliografía siguiente. La teoría de control estocástico en general fue consultada en [14, 15, 16]. Los resultados sobre procesos de control semi-markovianos fueron revisados en [26, 27, 28, 32]; mientras que, para el análisis de las técnicas implementadas nos basamos en [6, 9, 11, 12, 19, 34, 38, 41], y para los procesos de control semi-markovianos con distribución del tiempo de permanencia desconocida se consultó [29, 40]. La bibliografía restante sirvió como apoyo.

Bibliografía

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

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

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

[4] Federgruen A., Tijms H.C. (1978) *The optimality equation in average cost denumerable state semi-Markov decision problems, recurrency conditions and algorithms.* J. Appl. Probab. 15: 356-373.

[5] Federgruen A., Schweitzer P.J., Tijms H.C. (1983) *Denumerable undiscounted semi-Markov decision processes with unbounded rewards.* Math. Oper. Res. 8,2: 298-313.

[6] González-Trejo T.J., Hernández-Lerma O., Hoyos-Reyes L.F. (2003) *Mi-nimax control of discrete-time stochastic systems.* SIAM J. Control Optim. 41: 1626-1659.

[7] Gordienko E.I., Hernández-Lerma O. (1995) *Average cost Markov control processes with weighted norms: existence of canonical policies.* Appl. Math. 23: 199–218.

[8] Gordienko E.I., Hernández-Lerma O. (1995) *Average cost Markov control processes with weighted norms: value iteration.* Appl. Math. 23: 219–237.

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

[10] Gordienko E.I., Minjárez-Sosa J.A. (1998) *Adaptive control for discrete-time Markov processes with unbounded costs: average criterion.* ZOR - Math. Methods of Oper. Res. 48, pp. 37–55.

[11] Gordienko E.I., Salem-Silva F. (1998) *Robustness inequality for Markov control processes with unbounded costs.* Systems Control Lett. 33: 125–130.

[12] Gordienko E.I., Salem-Silva F. (2000) *Estimates of Markov control processes with unbounded costs.* Kybernetika 36,2: 195–210.

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

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

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

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

[17] Hilgert N., Minjárez-Sosa J.A. (2001) *Adaptive policies for time-varying stochastic systems under discounted criterion.* Math. Methods of Oper. Res. 54, 491-505.

[18] Hilgert N., Minjárez-Sosa J.A. (2003) *Limiting average cost adaptive control problem for time-varying stochastic systems.* Bol. Soc. Mat. Mexicana. 3, 9 197-212.

[19] Hilgert N., Minjárez-Sosa J.A. (2006) *Adaptive control of stochastic systems with unknown disturbance distribution: discounted criteria.* Math. Methods of Oper. Res. 63, 3, pp. 443-460.

[20] Jaskiewicz A. (2001) *An approximation approach to ergodic semi-Markov control processes.* Math Methods Oper. Res. 54: 1-19.

[21] Jaskiewicz A. (2004) *On the equivalence of two expected average cost criteria for semi-Markov control processes.* Math. Oper. Res. 29: 326-338.

[22] Jaskiewicz A. (2007) *A fixed point approach to solve the average cost optimality equation for semi-Markov decision with Feller transition probabilities.* Comm. Stat.-Theory and Methods 36: 2559-2575.

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

[24] Küenle H.-U. (1991) *On the optimality of* (s, S)*-strategies in a minimax inventory model with average cost criterion.* Optimization, 22, pp. 123-138.

[25] Kurano M. (1987) *Minimax strategies for average cost stochastic games with an application to inventory models.* J. Oper. Res. Soc. Japan. 30, 232-247.

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

[27] Luque-Vásquez F. (1997) *Modelos de Control Semi-Markovianos en Espacios de Borel.* Tesis Doctoral. UNAM, México.

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

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

[30] Luque-Vásquez F., Minjárez-Sosa J.A., Rosas-Rosas L.C. (2010) *Semi-Markov control processes with unknown holding times distribution under an average cost criterion.* Appl. Math. Optim. 61: 317-336.

[31] Luque-Vásquez F., Minjárez-Sosa J.A., Rosas-Rosas L.C. (2010) *Semi-Markov control models with partially known holding times distribution: discounted and average criteria.* (Aceptado para su publicación en Acta Appl. Math.)

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

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

[34] Minjárez-Sosa J.A. (1998) *Control Adaptado para Procesos de Markov con Costos no Acotados.* Tesis Doctoral. UAM-Iztapalapa, México.

[35] Minjárez-Sosa J.A. (1999) *Adaptive policies for discrete-time Markov control processes with unbounded costs: average and discounted criteria.* Morfismos, 3: 41-67.

[36] Minjárez-Sosa J.A. (1999) *Nonparametric adaptive control for discrete-time Markov processes with unbounded costs under average criteria.* Appl. Math. (Warsaw) 26: 267–280.

[37] Minjárez-Sosa J.A. (2002) *Average optimality for adaptive Markov control processes with unbounded costs and unknown disturbance distribution.* In Markov Processes and Controlled Markov Chains, Ch.7. Hou Zhenting, Jerzy A. Filar and Anyue Chen, Editors, Kluwer.

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

[39] Minjárez-Sosa J.A. (2008) *Empirical estimation in average Markov control processes.* Applied Mathematics Letters 21, 459–464.

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

[41] Montes-de-Oca R., Sakhanenko A.I., Salem-Silva F. (2003) *Estimates for pertubations of general discounted Markov control chains.* Appl. Math. 30: 287-304.

[42] Puterman M.L. (1994) *Markov Decision Processes. Discrete Stochastic Dynamic Programming.* Wiley & Sons, New York.

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

[44] Rosas-Rosas L.C. (2010) *Estimation of the deviation of optimality in discounted semi-Markov control models.* (Sometido para publicación).

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

[46] Schweitzer P.J. (1971) *Iterative solution of the functional equations of undiscounted Markov renewal programming.* J. Math. Anal. Appl. 34: 495-501.

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

[48] Vega-Amaya O. (2003) *The average cost optimality equation: a fixed point approach.* Bol. Soc. Mat. Mexicana 9: 185-195.