UNIVERSIDAD DE SONORA

División de Ciencias Exactas y Naturales

Departamento de Matemáticas

"El saber de mis hijos hará mi grandeza"

Algoritmo de Aproximación en Modelos de Control Semi-Markovianos y Markovianos con Costos Descontados

T E S I S

Que para obtener el grado académico de:

Maestro en Ciencias (Matemáticas)

Presenta:

Jazmín Sarahí Flores Gomez

Director de Tesis:

Dra. María Teresa Robles Alcaraz
Dr. Jesús Adolfo Minjárez Sosa

Hermosillo, Sonora, México, Octubre 2022

SINODALES

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

Dr. Saúl Mendoza Palacios
Centro de Investigación y Docencia Económicas, A.C. México

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

Dra. María Teresa Robles Alcaraz
Universidad de Sonora, Hermosillo Sonora, México

Agradecimientos

Quiero comenzar agradeciendo a mi familia por todo el apoyo incondicional que me brindaron durante mi estancia dentro de la maestria en Matemáticas, especialmente a mi abuela Maria Asunción Graciano Rojo y a mi madre Dulce Maria Gómez Graciano, por motivarme a seguir mis sue˜nos.

A mi profesora y Directora de Tesis, Dra.Teresa Robles, asi como a mis profesores sinodales, gracias por su paciencia y por las horas dedicadas a la preparación y revisión de esta tesis.

Por último quiero agradecer a la Universidad de Sonora y al Consejo Nacional de Ciencia y Tecnología (CONACyT) por el apoyo económico brindado que me permitió concluir mis estudios de posgrado.

Jazmin Sarahí Flores Gómez
Hermosillo, Sonora
octubre 2022

Indice General

Introducción

La teoría de control óptimo trata con problemas de optimización de sistemas dinámicos cuyo comportamiento puede verse influenciado por la aplicación de acciones (decisiones o controles), las cuales son seleccionadas mediante reglas conocidas como estrategias o políticas de control. La eficiencia de cada una de tales políticas se mide mediante un índice de funcionamiento del sistema conocido también como criterio de optimalidad, mismo que representa, ya sea un costo o una ganancia. Entonces el problema de control óptimo consiste en encontrar una estrategia óptima tal que, según sea el caso, minimice o maximice un índice de funcionamiento apropiado. Los modelos utilizados en el estudio de los problemas de control óptimo se clasifican en estocásticos o determinísticos si incluyen o no elementos aleatorios, respectivamente; de acuerdo a su índice de funcionamiento, ya sea descontado o promedio; y en tiempo continuo o discreto dependiendo de como se observa el proceso. Por ejemplo, los procesos de control markovianos, son procesos a tiempo discreto donde las acciones se toman en unidades de tiempo fijo. Por otro lado, una clase de procesos a tiempo continuo la constituyen los procesos semi-markovianos donde los controles son elegidos en tiempos aleatorios llamados épocas de decisión.

El objetivo de este trabajo es estudiar métodos de aproximación para resolver el problema de control óptimo descontado en procesos de control markovianos con costos acotados y procesos semi-markovianos con costos no acotados dentro del siguiente contexto. Un esquema general para estudiar el problema de control óptimo con costo descontado consiste en lo siguiente: (1) demostrar que la función de costo óptimo V es una solución de la ecuación de optimalidad, la cual se puede establecer como

$$V = TV$$

donde T es un operador en un espacio de funciones adecuado definido en términos del mínimo sobre el conjunto de controles; (2) a partir de esta ecuación, resolver el problema de minimización correspondiente que nos lleve a obtener políticas de control óptimas. Como lo establece gran parte de la literatura, la solución y análisis de los puntos (1) y (2) en problemas de aplicación concretos no son fáciles. De hecho, muchos trabajos de investigación solo se han enfocado a establecer condiciones que garanticen la existencia de soluciones a la ecuación de óptimalidad respectiva y de políticas óptimas, y no a proponer soluciones especificas. A partir de este hecho, surgen los métodos de aproximación llamados *algoritmo de iteración de valores* y el *algoritmo de iteración de políticas*. Estos proporcionan una manera iterativa de aproximar tanto a la función de valor como a la política óptima mediante una ecuación recursiva de la forma

$$v_{n+1} = Tv_n, \quad n = 0, 1, ... \tag{0.0.1}$$

En este caso la sucesión de funciones $\{v_n\}$ converge a la función de valor óptimo. Estos algoritmos son teóricamente eficientes pero numéricamente intratables en sistemas con un gran número de estados o con espacio de estados infinito debido a la llamada *"maldición de la dimensión"*. Con el fin de resover este problema se proponen los llamados algoritmo de Iteración de Valores Aproximado *(IVA)* y el algoritmo de Iteración de Políticas Aproximado *(IPA)*, los cuales consisten en combinar los algoritmos originales con métodos de aproximación de funciones. Generalmente estos métodos de aproximación de funciones son definidos por medio de operadores de tal manera que el problema es encontrar un operador apropiado T' de tal forma que $T'v$ sea una aproximación de la función v, y aplicarlo en el algoritmo original (0.0.1). Entonces, dependiendo del operador que se considere se obtiene un algoritmo de aproximación (ver , [3], [4], [5], [7], [12], [18], [17], [19], [20]). Muchos de estos trabajos consideran modelos de control con espacios finitos.

Una clase particular de operadores son los llamados "promediadores", los cuales son operadores con las siguientes propiedades:

1. las funciones constantes son puntos fijos del operador;

2. es lineal y positivo;

3. tiene propiedades de continuidad.

Los métodos de interpolación más comunes como lineal, multilineal, splines, kernels, y algunas variantes de los métodos de proyección, son ejemplos de promediadores. Como se establece en los trabajos [18], [17] el uso de promediadores tiene la ventaja de que se pueden aplicar en problemas de control con espacio de estado arbitrario y con costos posiblemente no acotados.

A partir de lo anterior, en este trabajo estudiamos el algoritmo de iteración de políticas aproximado definido por medio de una clase de promediadores. Dicho algoritmo lo establecemos para procesos de control markovianos y semi-markovianos. Mas aún, para el caso markoviano proponemos un algoritmo de estimación-aproximación para sistemas que evolucionan de acuerdo a una ecuación en diferencias estocástica donde la distribución del ruido aleatorio correspondiente es desconocida.

El trabajo está estructurado de la siguiente manera. En el Capítulo 1 se describe un Modelo de Control markoviano y se da solución al problema de control óptimo correspondiente; esto se realiza bajo ciertas condiciones de continuidad y compacidad. Además introducimos un algoritmo de aproximación conocido como algoritmo de iteración de políticas bajo el cual nos aproximamos a la función de valor óptimo. En el Capítulo 2 trabajamos con un modelo de control muy particular, para el cual la dinámica del sistema es descrita por una ecuación en diferencias. Además presentamos un modelo de control estimado para el que se encuentra la solución, la cual nos permitirá aproximarnos a la función de valor óptimo del modelo de control descrito en en el Capítulo 1. En el Capitulo 3 se estudia el modelos de control semi-markoviano estandar bajo un criterio de costo α-descontado y se presentan los principales resultados sobre el problema de control para este caso.

Finalmente en los Capítulos 4 y 5 se estudia un algortimo de iteración de políticas aproximado para los modelos de control markovianos y semi-markovianos. Con este algoritmo se realiza una aproximación a la función de valor óptimo y a las políticas óptimas. Como primer paso a la aproximación se describen los modelos de control perturbados, los cuales estan asociados a los promediadores.

Capítulo 1

Proceso de Control de Markov y Problema de Control Óptimo Descontado

1.1 Introducción

El propósito de este capítulo es plantear y dar solución al problema de control óptimo descontado con costos acotados. Para tal fin, a lo largo de sus secciones describiremos el modelo correspondiente e introduciremos el conjunto de políticas de control, así como el índice de funcionamiento que permitirá medir el desempeño mismo de dichas políticas. También describimos un procedimiento conocido como algoritmo de iteración de políicas que nos permitirá aproximarnos a la función de valor óptimo. Finalmente se presenta el ejemplo de un sistema de inventario que satisface las condiciones de continuidad y compacidad necesarias para reolver el problema de control óptimo correspondiente.

1.2 Modelo de control markoviano

En general, un *modelo de control de Markov* (MCM), denotado por

$$(\mathbf{X}, \mathbf{A}, \{A(x) : x \in \mathbf{X}\}, Q, C) \tag{1.2.1}$$

consta de los siguientes elementos:

i) $\mathbf{X}$ es un espacio de Borel y representa el *espacio de estados* del sistema.

ii) $\mathbf{A}$ representa el *espacio de controles* (o acciones) del sistema y al igual que $\mathbf{X}$ es un espacio de Borel.

iii) $\{A(x) : x \in \mathbf{X}\}$ es la *familia de conjuntos de acciones admisibles*, de forma tal que cada estado $x \in \mathbf{X}$ tiene asociado un conjunto medible y no vacio $A(x) \subset \mathbf{A}$, cuyos elementos son las acciones admisibles cuando el sistema se encuentra en el estado x. Definimos el conjunto

$$\mathbb{K} := \{(x, a) : x \in \mathbf{X}, a \in A(x)\} \tag{1.2.2}$$

de pares estado-acción admisibles, el cual es un subconjunto medible de $\mathbf{X} \times \mathbf{A}$.

iv) $Q(\cdot \mid x, a)$ es un kernel estocástico en $\mathbf{X}$ dado $\mathbb{K}$ llamado *ley de transición.*

v) $C : \mathbb{K} \to \mathbb{R}$ representa una función de *costo por estapa*, la cual es una función medible no negtiva sobre $\mathbb{K}$.

Explícitamente, un modelo de control de Markov representa un sistema estocástico controlado que evoluciona en el tiempo de modo que, para cada $n \in \mathbb{N}$, ocurre que en la n-ésima etapa de decisión el sistema se encuentra en el estado $x_n = x \in \mathbf{X}$ y se elige un control $a_n = a \in A(x)$, entonces sucede lo siguiente:

i) Se genera un costo $C(x, a)$.

ii) El sistema transita a un nuevo estado $x_{n+1} \in \mathbf{X}$ de acuerdo a la medida de probabilidad $Q(\cdot \mid x, a)$, es decir

$$Q(B \mid x, a) = \text{Prob}\left[x_{n+1} \in B \mid x_n = x, a_n = a\right] \quad B \subset \mathbf{X},$$

una vez que el sistema se encuentra en el nuevo estado, se elige un nuevo control y el proceso se repite.

Definición 1.1. *Dado un modelo de control de Markov definimos para cada $n \in \mathbb{N}_0$ el espacio de historias admisibles hasta la etapa n mediante*

$$\mathbf{H}_0 := \mathbf{X}$$

$$\mathbf{H}_n := \mathbb{K}^n \times \mathbf{X} \quad \forall n \in \mathbb{N}.$$

Un elemento $h_n \in \mathbf{H}_n$ consiste de un vector (llamado también n-historia) de la forma

$$h_n = (x_0, a_0, x_1, a_1, \ldots, x_{n-1}, a_{n-1}, x_n)$$

con $(x_k, a_k) \in \mathbb{K}$ para $k = 0, 1, \ldots, n-1$ y $x_n \in \mathbf{X}$.

Definición 1.2. *Una política de control aleatorizada es una sucesión $\pi = \{\pi_n\}$ de kérneles estocásticos en $\mathbf{A}$ dado $\mathbf{H}_n$ tal que*

$$\pi_n\left(A\left(x_n\right) \mid h_n\right) = 1 \quad \forall h_n \in \mathbf{H}_n, n = 0, 1...$$

Denotaremos por Π al conjunto de todas las políticas y por $\mathbb{F}$ al conjunto de todas las funciones medibles $f : \mathbf{X} \to \mathbf{A}$ tal que $f(x) \in A(x)$ para toda $x \in \mathbf{X}$. A estas funciones las llamaremos selectores.

Definición 1.3. *Una política $\pi = \{\pi_n\}$ es*

a) Determinista si existe una sucesión $\{f_n\}$ de funciones medibles $f_n : \mathbf{H}_n \to \mathbf{A}$ tal que, para toda $h_n \in \mathbf{H}_n$ y $n \in \mathbb{N}_0, f_n\left(h_n\right) \in A\left(x_n\right)$ y $\pi_n\left(\cdot \mid h_n\right)$ está concentrada en $f_n\left(h_n\right)$, es decir

$$\pi_n\left(B \mid h_n\right) = I_B\left(f_n\left(h_n\right)\right) \quad \forall B \in \mathcal{B}(\mathbf{A}), \quad h_n \in \mathbf{H}_n.$$

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

$$\pi_n\left(B \mid h_n\right) = I_B\left(f_n\left(x_n\right)\right) \quad \forall B \in \mathcal{B}(\mathbf{A}), \quad h_n \in \mathbf{H}_n.$$

c) Estacionaria si existe $f \in \mathbb{F}$ tal que

$$\pi_n\left(B \mid h_n\right) = I_B\left(f\left(x_n\right)\right) \quad \forall B \in \mathcal{B}(\mathbf{A}), h_n \in \mathbf{H}_n, n \in \mathbb{N}_0.$$

En este caso la política $\pi = \{\pi_n\}$ se identifica con el selector f y a la clase de todas las políticas estacionarias con la familia de los selectores $\mathbb{F}$.

Definición 1.4. *Dada una política* $\pi = \{\pi_t\}$ *definimos la política "t-recorrida"* $\pi^{(t)} = \left\{\pi_n^{(t)}, n = 0, 1, \ldots\right\}$ *como*

$$\pi_n^{(t)}\left(\cdot \mid x_t, a_t, ...x_{t+n}\right) := \pi_{t+n}\left(\cdot \mid h_t, a_t, ...x_{t+n}\right) \quad \textit{para cada} \quad n \geq 0, \quad h_n \in H_n \tag{1.2.3}$$

Consideremos el espacio medible $(\Omega, \mathcal{F})$ para el cual el espacio muestral está definido como $\Omega := (\mathbf{X} \times \mathbf{A})^{\infty}$ y donde $\mathcal{F}$ es la correspondiente sigma-álgebra producto. Los elementos en Ω son sucesiones de la forma

$$\omega = (x_0, a_0, x_1, a_1, \ldots)$$

con $x_k \in \mathbf{X}$ y $a_k \in \mathbf{A}$ para toda $k = 0, 1, \ldots$.

Sea $\pi = \{\pi_n\}$ una política de control arbitraria y ν una medida arbitraria de probabilidad en $\mathbf{X}$ a la cual nos referiremos como "distribución inicial". Entonces, por el Teorema de Ionescu-Tulcea [2, Theorem 2.7.2 p. 109], existe una única medida de probabilidad P_ν^π en $(\Omega, \mathcal{F})$ y procesos estocásticos $\{x_n\}$ y $\{a_n\}$ tal que $P_y^\pi\left(H_\infty\right) = 1$ y para toda $B \in \mathcal{B}(\mathbf{X}), C \in \mathcal{B}(\mathbb{A})$, $h_n \in \mathbf{H}_n, n = 0, 1, \ldots$

$$P_\nu^\pi\left(x_0 \in B\right) = \nu(B); \tag{1.2.4}$$

$$P_\nu^\pi\left(a_n \in C \mid h_n\right) = \pi_n\left(C \mid h_n\right); \tag{1.2.5}$$

$$P_\nu^\pi\left(x_{n+1} \in B \mid h_n, a_t\right) = Q\left(B \mid x_n, a_n\right). \tag{1.2.6}$$

Al operador esperanza con respecto a la medida de probabilidad P_ν^π lo denotaremos por E_ν^π. Si ν es una medida concentrada en el estado inicial $x_0 = x$, entonces escribiremos P_x^π y E_x^π en lugar de P_ν^π y E_ν^π, respectivamente.

1.3 Problema de control óptimo descontado y condiciones de optimalidad

En general, un índice de funcionamiento es una función con la cual se mide el desempeño del sistema bajo diferentes políticas de control, dado el estado inicial. El criterio que usaremos a lo largo del presente trabajo es el costo total descontado, el cual se define a continuación.

Definición 1.5. *Sean* $x_0 = x \in \mathbf{X}$, $\pi \in \Pi$ *arbitrarios, y sea* $\alpha \in (0, 1)$ *una constante dada. Definimos el costo total esperado* α*-descontado con horizonte infinito como:*

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

donde α *representa un factor de descuento.*

La función de valor óptimo correspondiente se define como

$$V^*(x) := \inf_{\pi \in \Pi} V(x, \pi), \quad x \in \mathbf{X}.$$

Entonces, el problema de control óptimo en costo descontado consiste en encontrar una política $\pi^* \in \Pi$ tal que

$$V^*(x) = V(x, \pi^*) \quad \forall x \in \mathbf{X}.$$

En este caso, diremos que π^* es óptima en costo descontado o, simplemente, que π^* es óptima descontada.

Para garantizar que el problema de control óptimo esté bien definido, asi como la existencia de políticas óptimas, se introducen las siguientes condiciones de compacidad y continuidad

Hipótesis 1.1. ***a)*** *$C(x,a)$ es una función continua en $\mathbb{K}$ y acotada por una constante $M>0$;*

b) *$A(x)$ es un subconjunto de $\mathbf{A}$ no-vacío y compacto para cada $x \in \mathbf{X}$ y el mapeo $x \to A(x)$ es continuo;*

c) *$Q(\cdot \mid x,a)$ es débilmente continua en $\mathbb{K}$, es decir, la función*

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

es continua para cada $u \in \mathbf{C}_b(\mathbf{X})$.

Hipótesis 1.2. ***a)*** *$C(x,a)$ es acotada por una constante $M>0$.*

b) *$C(x,\cdot)$ es una función continua en $A(x)$ para cada $x \in \mathbf{X}$.*

c) *La multifunción $x \to A(x)$ toma valores en los subconjuntos compactos de $\mathbf{A}$.*

d) *$Q(\cdot \mid x,a)$ es fuertemente continua en $A(x)$, es decir, la función*

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

es continua para cada $u \in \mathbf{M}_b(\mathbf{X})$ y cada $x \in \mathbf{X}$.

Para cada selector $f \in \mathbb{F}$ y cada función medible $v : \mathbb{K} \to \mathbb{R}$ se define

$$v_f(x) := v(x, f(x)), \quad \forall x \in \mathbf{X}, B \in \mathcal{B}(\mathbf{X}).$$

Con esta notación, la función de costo y probabilidad de transición se reescriben de la siguiente manera

$$C_f(x) = C(x, f(x)) \quad \text{y} \quad Q_f(B \mid x) = Q(B \mid x, f(x)).$$

Observación 1.1. *Notemos que si la función de costo por etapa $C(x,a)$ es continua en $\mathbb{K}$ y acotada por una constante M, entonces*

$$\begin{aligned} V(x,f) &= E_x^f \left[\sum_{n=0}^{\infty} \alpha^n C(x_n, a_n) \right] \\ &\leq E_x^f \left[\sum_{n=0}^{\infty} \alpha^n M \right]. \end{aligned}$$

Entonces

$$\|V(x,f)\|_\infty \leq \alpha \frac{M}{1-\alpha}. \tag{1.3.1}$$

Ahora para cada $f \in \mathbb{F}$, y para cada $u(x) \in \mathbf{M}_b(\mathbf{X})$ se define el operador

$$T_f u(x) := C_f(x) + \alpha \int_{\mathbf{X}} u(y) Q_f(dy \mid x), \quad x \in \mathbf{X}. \tag{1.3.2}$$

Análogamente, para cada $u \in \mathbf{M}_b(\mathbf{X})$, se define el operador

$$Tu(x) := \inf_{a \in A(x)} \left[C(x,a) + \alpha \int_{\mathbf{X}} u(y) Q(dy \mid x,a) \right], \quad x \in \mathbf{X}. \tag{1.3.3}$$

A T se le llama operador de la programación dinámica.

Como consecuencia de la Hipótesis 1.2, se presenta el siguiente resultado el cual es punto de partida para dar solución al problema de control óptimo.

Proposición 1.2. *Suponga que se satisface la Hipótesis 1.2. Entonces*

a) *El operador de programación dinámica T es un operador de contracción módulo α en el espacio de Banach $(\mathbf{M}_b(\mathbf{X}), \|\cdot\|_\infty)$, y por lo tanto, tiene un único punto fijo u^* en $\mathbf{M}_b(\mathbf{X})$. Además para cada función $u \in \mathbf{M}_b(\mathbf{X})$ existe un selector $f_u \in \mathbb{F}$ tal que $Tu(\cdot) = T_{f_u}(\cdot)$, es decir para cada $x \in \mathbf{X}$ se tiene que*

$$Tu(x) = \inf_{a \in A(x)} \left[C(x,a) + \alpha \int_{\mathbf{X}} u(y) Q(dy \mid x,a) \right] = C_{f_u}(x) + \alpha \int_{\mathbf{X}} u(y) Q_{f_u}(dy \mid x).$$

b) *El operador $T_f, f \in \mathbb{F}$, es de contracción módulo α en el espacio de Banach $(\mathbf{M}_b(\mathbf{X}), \|\cdot\|_\infty)$, entonces, por el teorema de punto fijo de Banach, tiene un único punto fijo u_f en $\mathbf{M}_b(\mathbf{X})$.*

Demostración. **a)** Notemos que para cada $u(x) \in \mathbf{M}_b(\mathbf{X})$

$$C(x,a) + \alpha \int_{\mathbf{X}} u(y) Q(dy \mid x,a)$$

es una función medible en $(x,a) \in \mathbb{K}$ y continua en $a \in A(x)$, para cada $x \in \mathbf{X}$, donde $A(x)$ es un conjunto compacto. Entonces por la Proposición B.1 en el Apéndice B, existe un selector $f_u \in \mathbb{F}$ en donde, para cada $x \in \mathbf{X}$ se alcanza el ínfimo sobre $A(x)$, es decir

$$C_{f_u}(x) + \alpha \int_{\mathbf{X}} u(y) Q_{f_u}(dy \mid x) = \inf_{a \in A(x)} \left[C(x,a) + \alpha \int_{\mathbf{X}} u(y) Q(dy \mid x,a) \right],$$

y además

$$\inf_{a \in A(x)} \left[C(x,a) + \alpha \int_{\mathbf{X}} u(y) Q(dy \mid x,a) \right]$$

es una función medible, es decir, $Tu \in \mathbf{M}_b(\mathbf{X})$.

Veamos ahora que el operador de programación dinámica es de contracción.

Sean $u, v \in \mathbf{M}_b(\mathbf{X})$ entonces

$$
\begin{aligned}
|Tu(x) - Tv(x)| =& |\inf_{a\in A(x)} \left[C(x,a) + \alpha \int_{\mathbf{X}} u(y)Q(dy \mid x,a)\right] - \inf_{a\in A(x)} \left[C(x,a) + \alpha \int_{\mathbf{X}} v(y)Q(dy \mid x,a)\right]| \\
&\leq \sup_{a\in A(x)} |\alpha \int_{\mathbf{X}} u(y)Q(dy \mid x,a) - \alpha \int_{\mathbf{X}} v(y)Q(dy \mid x,a)| \\
&< \alpha \sup_{a\in A(x)} |\int_{\mathbf{X}} (u(y) - v(y))Q(dy \mid x,a)| \\
&\leq \alpha \sup_{a\in A(x)} \|u - v\|_\infty \int_{\mathbf{X}} Q(dy \mid x,a)
\end{aligned}
$$

por lo tanto

$$
|Tu(x) - Tv(x)| \leqslant \alpha \|u - v\|_\infty.
$$

Finalmente, tomando supremo sobre $\mathbf{X}$ se tiene que

$$
\|Tu - Tv\|_\infty \leqslant \alpha \|u - v\|_\infty,
$$

lo cual implica que T es un operador de contracción modulo α en el espacio de Banach $(\mathbf{M}_b(\mathbf{X}), \|\cdot\|_\infty)$. Por lo tanto, del teorema del punto fijo de Banach se sigue que existe un único punto fijo u^* del operador T en $\mathbf{M}_b(\mathbf{X})$.

b) De manera análoga se demuestra que el operador T_f es de contracción con módulo α en el espacio de Banach $(\mathbf{M}_b(\mathbf{X}), \|\cdot\|_\infty)$, de manera que, por el teorema de punto fijo de Banach, tiene un único punto fijo u_f en $\mathbf{M}_b(\mathbf{X})$.

□

1.4 Solución al problema de control óptimo descontado

Proposición 1.3. *Si se cumple la Hipótesis 1.2 entonces:*

a) *La función de valor óptimo V^* es el único punto fijo en $\mathbf{M}_b(\mathbf{X})$ del operador T, es decir,*

$$
V^*(x) = TV^*(x) = \min_{a\in A(x)} \left[C(x,a) + \alpha \int_{\mathbf{X}} V^*(y)Q(dy \mid x,a)\right]. \tag{1.4.1}
$$

b) *Una política estacionaria $f \in \mathbb{F}$ es óptima si y sólo si*

$$
V^*(x) = C(x,a) + \alpha \int_{\mathbf{X}} V^*(y)Q(dy \mid x,a) = T_f V^*(x)
$$

c) *Existe una política estacionaria $f^* \in \mathbb{F}$ tal que*

$$
V^* = T_{f^*} V^*.
$$

La demostración de la Proposición 1.3 necesita de un resultado preliminar en el cual se relaciona el punto fijo u_f del operador T_f con el costo $V(f,x)$ bajo la política estacionaria f, y el punto fijo u^* del operador de la programación dinámica con la función de valor óptimo V^*.

Lema 1.4. *Consideremos el punto fijo u_f del operador T_f, entonces, bajo la Hipótesis 1.2:*

a) *$u_f(x) = V(x, f)$ para cada política estacionaria $f \in \mathbb{F}$;*

b) *una política π^* es óptima si y sólo si para cada $x \in \mathbf{X}$*

$$TV(x, \pi^*) = V(x, \pi^*) \tag{1.4.2}$$

Demostración. **a)** Por la unicidad del punto fijo u_f del operador T_f (Proposición 1.2 inciso b)), es suficiente demostrar que $V(x, f)$ satisface la ecuación $T_f V(x, f) = V(x, f)$.

Para esto, expresamos el costo $V(x, f)$ de la siguiente manera

$$V(x, f) = E_x^f \sum_{t=0}^{\infty} \alpha^t C(x_t, a_t) = C(x, f(x)) + \alpha E_x^f \left[\sum_{t=1}^{\infty} \alpha^{t-1} C(x_t, a_t) \right]. \tag{1.4.3}$$

Luego, por propiedades de la esperanza condicional y de la propiedad de Markov (1.2.6), la esperanza del lado derecho de la igualdad (1.4.3) se puede reescribir como

$$\begin{aligned} E_x^f \left[\sum_{t=1}^{\infty} \alpha^{t-1} C(x_t, a_t) \right] &= E_x^f \left[E_x^f \left(\sum_{t=1}^{\infty} \alpha^{t-1} C(x_t, a_t) \mid h_1 \right) \right] \\ &= E_x^f \left[E_{x_1}^f \left(\sum_{t=1}^{\infty} \alpha^{t-1} C(x_t, a_t) \right) \right] \\ &= E_x^f \left[V(x_1, f) \right] \\ &= \int_{\mathbf{X}} V(y, f) Q(dy \mid x, f(x)). \end{aligned}$$

Entonces, para cada $x \in \mathbf{X}$,

$$V(x, f) = C(x, f(x)) + \alpha \int_{\mathbf{X}} V(y, f) Q(dy \mid x, f(x)) = T_f V(x, f), \tag{1.4.4}$$

por lo tanto $u_f(x) = V(x, f)$ para cada política estacionaria $f \in \mathbb{F}$.

b) Supongamos que π^* es una política óptima, demostremos que

i) $V(x, \pi^*) \leq TV(x, \pi^*)$.

ii) $V(x, \pi^*) \geq TV(x, \pi^*)$
Para demostrar i) consideremos $g \in \mathbb{F}$ una política estacionaria arbitraria y sea $\pi' := (g, \pi^*)$ la política que usa g al tiempo $t = 0$ y usa la política óptima π^* desde el tiempo $t = 1$ en adelante.

De manera que

$$\pi'_0(x_0) := g(x_0)$$

y para $t \geq 1$

$$\pi'_t(\cdot \mid x_0, a_0, \ldots, x_{t-1}, a_{t-1}, x_t) := \pi^*_{t-1}(\cdot \mid x_1, a_1, \ldots, x_t).$$

Entonces la óptimalidad de π^* implica que para cada $x \in \mathbf{X}$

$$V(x,\pi^*) \leq V\left(x,\pi'\right) = C(x,g(x)) + \alpha \int_{\mathbf{X}} V(y,\pi^*) Q(dy \mid x, g(x)).$$

Como $g \in \mathbb{F}$ es una política arbitraria, para cada $x \in \mathbf{X}$

$$V(x,\pi^*) \leq \inf_{a\in A(x)} \{C(x,a) + \alpha \int_{\mathbf{X}} V(y,\pi^*) Q(dy \mid x, g(a))\} = TV(x,\pi^*),$$

lo cual demuestra i).

Para la desigualdad en ii) consideramos el costo descontado $V(x,\pi^*)$ de la siguiente manera

$$\begin{aligned}
V(x,\pi^*) &= E_x^{\pi^*} \sum_{t=0}^{\infty} \alpha^t C\left(x_t, a_t\right) \\
&= E_x^{\pi^*} [C\left(x_0, a_0\right) + \sum_{t=1}^{\infty} \alpha^t C\left(x_t, a_t\right)] \\
&= E_x^{\pi^*} [C\left(x_0, a_0\right) + \alpha \sum_{t=1}^{\infty} \alpha^{t-1} C\left(x_t, a_t\right)]
\end{aligned}$$

$$= \int_A \left\{C(x,a) + \alpha \int_X V\left[y, \pi^{*(1)}\right] Q(dy \mid x,a)\right\} \pi_0^*(da \mid x)$$

donde $\pi^{*(1)} = \{\pi_t^{*(1)}\}$ denota la política "1-shifted" en (1.2.3) con $x_0 = x$ y $a_0 = a$

$$\pi_t^{*(1)}\left(\cdot \mid h_t\right) := \pi_{t+1}^*\left(\cdot \mid x_0, a_0, h_t\right) \text{ para } t = 0, 1, \ldots$$

y ya que π^* es una política óptima, entonces

$$\begin{aligned}
V(x,\pi^*) &\geq \int_A \left\{C(x,a) + \alpha \int_{\mathbf{X}} V(y,\pi^*) Q(dy \mid x,a)\right\} \pi_0^*(da \mid x) \\
&\geq \inf_{a\in A(x)} \left\{C(x,a) + \alpha \int V(y,\pi^*) Q(dy \mid x,a)\right\} \\
&= TV(x,\pi^*).
\end{aligned}$$

Ahora sea π^* una política tal que $V\left(x,\pi^*\right) = TV\left(x,\pi^*\right)$, esto es

$$V\left(x,\pi^*\right) = \inf_{a\in A(x)} \left\{C(x,a) + \alpha \int_{\mathbf{X}} V\left(x,\pi^*\right) Q(dy \mid x,a)\right\}, \quad x \in \mathbf{X}.$$

Demostraremos que π^*, es óptima, es decir $V\left(x,\pi^*\right) \leq V(x,\pi)$ para cada pólitica π y para cada estado inicial $x \in \mathbf{X}$.

Para cada t-historia $h_t \in \mathbf{H}_t$, se sigue de la propiedad de Markov que

$$E_x^{\pi}\left[\alpha^{t+1} V\left(x_{t+1},\pi^*\right) \mid h_t, a_t\right] = \alpha^{t+1} \int_{\mathbf{X}} V\left(y,\pi^*\right) Q\left(dy \mid x_t, a_t\right)$$

$$= \alpha^t \left\{ C(x_t, a_t) + \alpha \int_{\mathbf{X}} V(y, \pi^*) Q(dy \mid x_t, a_t) \right\} - \alpha^t C(x_t, a_t)$$
$$\geq \alpha^t V(x_t, \pi^*) - \alpha^t C(x_t, a_t).$$

Equivalentemente

$$\alpha^t V(x_t, \pi^*) - E_x^\pi \left[\alpha^{t+1} V(x_{t+1}, \pi^*) \mid h_t, a_t \right] \leq \alpha^t C(x_t, a_t) \tag{1.4.5}$$

luego, tomando valor esperado y sumando para $t = 0, ...n$ en (1.4.5), obtenemos que

$$V(x, \pi^*) - \alpha^{n+1} E_x^\pi V(x_{n+1}, \pi^*) \leq E_x^\pi \sum_{t=0}^{n} \alpha^t C(x_t, a_t).$$

Finalmente, tomando límite cuando $n \to \infty$ obtenemos que

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

Por lo tanto π^* es óptima.

$\square$

La demostración de la Proposición 1.3, se obtiene de la Proposición 1.2 y del Lema 1.4.

Demostración. **de la Proposición 1.3**

a) Del Lema 1.4 b) una politica π^* es óptima, es decir, $V(x, \pi^*) = V^*(x)$ para cada $x \in \mathbf{X}$, si y sólo si

$$TV^*(x) = V^*(x).$$

La unicidad del punto fijo V^* se tiene de la Proposición 1.2 a).

b) Sea $f \in \mathbb{F}$ una política estacionaria tal que

$$V^*(x) = T_f V^*(x).$$

Entonces, de la Proposición 1.2 y del Lema 1.4 a) se tiene que

$$V^*(x) = u_f(x) = V(x, f).$$

Por lo tanto f es una política óptima. Por otra parte, si $f \in \mathbb{F}$ es óptima, entonces $V^*(x) = u_f(x)$ y por la unicidad del punto fijo u_f, se tiene que

$$V^*(x) = T_f V^*(x).$$

c) Notemos que

$$C(x, a) + \alpha \int_{\mathbf{X}} V^*(y) Q(dy \mid x, a)$$

es una función medible en $(x, a) \in \mathbb{K}$ y continua en $a \in A(x)$, para cada $x \in \mathbf{X}$, donde $A(x)$ es un conjunto compacto. Entonces por la Proposición B.1 en el Apéndice A, existe un selector f^* tal que

$$V^*(x) = C_{f^*}(x) + \alpha \int_{\mathbf{X}} V^*(y) Q_{f^*}(dy \mid x).$$

$\square$

Observación 1.5. *Notemos que bajo la Hipótesis 1.1, se obtienen los mismos resultados de la Proposición 1.2 y 1.3, ahora en el espacio de Banach* $(\mathbf{C}_b(\mathbf{X}), \|\cdot\|_\infty)$, *esto es*

a) *el operador de la programación dinámica* T *es un operador de contracción con modulo* α, *ahora en el espacio de Banach* $(\mathbf{C}_b(\mathbf{X}), \|\cdot\|_\infty)$, *por lo tanto, tiene un único punto fijo en* $\mathbf{C}_b(\mathbf{X})$*;*

b) *la función de valor óptimo* V^* *es el único punto fijo del operador de la programación dinámica* T *en* $\mathbf{C}_b(\mathbf{X})$.

1.5 Algoritmo de iteración de políticas

Bajo la Hipótesis 1.2, es posible aproximar la función de valor óptimo V^* empleando un procedimiento llamado algoritmo de iteración de políticas, el cual se describe a continuación.

Considere $f_0 \in \mathbb{F}$ una política estacionaria con costo α-descontado $V(x, f_0) < \infty$. Definamos $w_0(x) := V(x, f_0)$, el cual, como ya se demostró en (1.4.4) satisface

$$w_0(x) = C_{f_0}(x) + \alpha \int w_0(y) Q_{f_0}(dy \mid x) \quad \forall x \in \mathbf{X}.$$

Sea T el operador de programación dinámica definido en (1.3.3). Evaluamos $Tw_0(x)$, esto es

$$Tw_0(x) = \inf_{a \in A(x)} \left[C(x,a) + \alpha \int w_0(y) Q(dy \mid x, a) \right]$$

entonces por la Proposición 1.2 a) existe un selector $f_1 \in \mathbb{F}$ tal que

$$C_{f_1}(x) + \alpha \int w_0(y) Q_{f_1}(dy \mid x) = \inf_{a \in A(x)} \left[C(x,a) + \alpha \int w_0(y) Q(dy \mid x, a) \right] \quad \forall x \in \mathbf{X}.$$

Ahora, consideremos

$$w_1(\cdot) := V(\cdot, f_1).$$

De manera general, el algoritmo consiste en lo siguiente

i) inicio: tomamos $k = 0$ y elija una política $f_0 \in \mathbb{F}$;

ii) evaluación: dada $f_k \in \mathbb{F}$ calculemos el correspondiente costo α-descontado

$$w_k(\cdot) := V(\cdot, f_k);$$

iii) mejoramiento: encuentre $f_{k+1} \in \mathbb{F}$ tal que $Tw_k = T_{f_{k+1}} w_k$ y regrese a la etapa de evaluación.

Observación 1.6. *Bajo la Hipótesis 1.1 el algoritmo de iteración de políticas no está bien definido, puesto que la función* $w_0(x) := V(x, f_0) \in \mathbf{M}_b(\mathbf{X})$, *con lo cuál no está garantizado que la función*

$$\int_{\mathbf{X}} w_0(x) Q(dy \mid x, a)$$

sea continua en $a \in A(x)$, *de manera que, no es posible asegurar la existencia de selectores en la etapa de mejoramiento.*

Proposición 1.7. *Bajo la Hipótesis 1.2, se tiene que la sucesión* $\{w_k\} \in \mathbf{M}_b(\mathbf{X})$ *formada a partir del algoritmo de iteración de políticas es decreciente y converge en la norma del supremo a la función de valor óptimo* V^*. *Además*

$$\|V^* - w_k\|_\infty \leq \frac{2\alpha}{1-\alpha} \|w_k - w_{k-1}\|_\infty \quad \forall k \in \mathbb{N}.$$

Demostración. Para demostrar la convergencia de la sucesión $\{w_k\}$ a V^*, tomemos $f_0 \in \mathbb{F}$ arbitraria y observemos que

$$\begin{aligned} w_0(x) &:= V(x, f_0) \\ &= C_{f_0}(x) + \alpha \int_{\mathbf{X}} w_0(y) Q_{f_0}(dy \mid x) \\ &\geq \inf_{a \in A(x)} \left[C(x,a) + \alpha \int_{\mathbf{X}} w_0(y) Q(dy \mid x, a) \right] \\ &= Tw_0(x). \end{aligned}$$

Por otro lado, existe $f_1 \in \mathbb{F}$ tal que

$$Tw_0(x) = C_{f_1}(x) + \alpha \int_{\mathbf{X}} w_0(y) Q_{f_1}(dy \mid x),$$

entonces

$$w_0(x) \geq C_{f_1}(x) + \alpha \int_{\mathbf{X}} w_0(y) Q_{f_1}(dy \mid x).$$

Iterando la desigualdad anterior obtenemos

$$w_0(x) \geq C_{f_1}(x) + \alpha \int_{\mathbf{X}} \left[C_{f_1}(y) + \alpha \int_{\mathbf{X}} w_0(z) Q_{f_1}(dz \mid y) \right] Q_{f_1}(dy \mid x)$$

$$= C_{f_1}(x) + \alpha \int_{\mathbf{X}} C_{f_1}(y) Q_{f_1}(dy \mid x) + \alpha^2 \int_{\mathbf{X}} \int_{\mathbf{X}} w_0(z) Q_{f_1}(dz \mid y) Q_{f_1}(dy \mid x)$$

$$= E_x^{f_1} \sum_{k=0}^{1} \alpha^t C_{f_1}(x_k) + \alpha^2 E_x^{f_1} w_0(x_2).$$

De esta manera, iterando n veces se tiene que

$$w_0(x) \geq E_x^{f_1} \sum_{k=0}^{n-1} \alpha^k C_{f_1}(x_k) + \alpha^n E_x^{f_1} w_0(x_n). \tag{1.5.1}$$

Como $w_0(x)$ es una función acotada y $\alpha \in (0,1)$, tomando límite cuando $n \to \infty$, obtenemos $\alpha^n E_x^{f_1} w_0(x_n) \to 0$. Por lo tanto

$$w_0(x) \geq Tw_0(x) \geq V_{f_1}(x) =: w_1(x).$$

Repitiendo la argumentación anterior, observemos que de ii) y iii) del algoritmo de iteración de políticas, se tiene que

$$w_k(x) \geq Tw_k(x) \geq w_{k+1}(x), \forall x \in \mathbf{X}, k \in \mathbb{N}_0, \tag{1.5.2}$$

por lo que $\{w_n\}$ es una sucesión decreciente y acotada por abajo ($w_n \geq 0 \quad \forall n$). Entonces existe una función $w \in \mathbf{M}_b(\mathbf{X})$ tal que $w_n \downarrow w$. Luego tomando límite cuando $k \to \infty$ en (1.5.2), por el teorema de convergencia monotona y del intercambio del límite con el ínfimo ([6], Lemma 3.4), tenemos

$$\begin{aligned}
\lim_{k\to\infty} Tw_k(x) &= \lim_{k\to\infty} \inf_{a\in A(x)} \left[C(x,a) + \alpha \int_{\mathbf{X}} w_k(y) Q(dy \mid x,a) \right] \\
&= \inf_{a\in A(x)} \left[C(x,a) + \lim_{k\to\infty} \alpha \int_{\mathbf{X}} w_k(y) Q(dy \mid x,a) \right] \\
&= \inf_{a\in A(x)} \left[C(x,a) + \alpha \int_{\mathbf{X}} w(y) Q(dy \mid x,a) \right] \\
&= Tw(x), \quad x \in \mathbf{X}.
\end{aligned}$$

Por lo tanto

$$\lim_{k\to\infty} Tw_k(x) = Tw(x), \quad x \in \mathbf{X}, \tag{1.5.3}$$

así que de (1.5.2) y la igualdad anterior obtenemos que $w > Tw > w$ lo cual prueba que w es punto fijo del operador de la programación dinámica T, es decir,

$$w(\cdot) = Tw(\cdot).$$

Ahora de la Proposición 1.3 a) el único punto fijo del operador de la programación dinámica T es la función de valor óptimo V^*, por lo tanto $w = V^*$.

Ahora para demostrar la segunda parte de la proposición notemos que

$$\begin{aligned}
\|V^* - w_k\|_\infty &\leq \|V^* - T_{f_k} w_{k-1}\|_\infty + \|T_{f_k} w_{k-1} - w_k\|_\infty \\
&= \|TV^* - Tw_{k-1}\|_\infty + \|T_{f_k} w_{k-1} - T_{f_k} w_k\|_\infty \\
&\leq \alpha \|V^* - w_{k-1}\|_\infty + \alpha \|w_{k-1} - w_k\|_\infty \\
&\leq \alpha \|V^* - w_k\|_\infty + \alpha \|w_k - w_{k-1}\|_\infty + \alpha \|w_{k-1} - w_k\|_\infty .
\end{aligned}$$

De aquí

$$\|V^* - w_k\|_\infty - \alpha \|V^* - w_k\|_\infty \leq \alpha \|w_k - w_{k-1}\|_\infty + \alpha \|w_{k-1} - w_k\|_\infty ,$$

lo cual implica que

$$\|V^* - w_k\|_\infty \leq \frac{2\alpha}{1-\alpha} \|w_k - w_{k-1}\|_\infty .$$

□

1.6 Ejemplo: un sistema de inventario

Considere un sistema de inventario a tiempo discreto, de un solo producto y que cuenta con una capacidad finita $\mathcal{K} > 0$.

Definimos las siguientes variables:

- x_n: nivel de inventario al inicio del periodo $n \in \mathbb{N}_0$;
- a_n: cantidad de producto que se ordena al inicio del periodo $n \in \mathbb{N}_0$;

- ξ_n: demanda del producto durante el periodo $n \in \mathbb{N}_0$.

Se asume que la cantidad de producto que se ordena se abastece de forma inmediata, que la demanada que no se satisface en cada periodo se pierde y que el nivel de inventario inicial es $x_0 = x \in \mathbf{X}$.

Así, se tiene que el espacio de estados y controles son $\mathbf{X} = \mathbf{A} = [0, \mathcal{K}]$, y que el conjunto de controles admisibles cuando el nivel de inventario es $x \in \mathbf{X}$ es $A(x) = [0, \mathcal{K} - x]$.

El sistema de inventario evoluciona de acuerdo a la siguiente ecuación

$$x_{n+1} = (x_n + a_n - \xi_n)^+ = \max(x_n + a_n - \xi_n, 0), \quad n \in \mathbb{N}_0 \tag{1.6.1}$$

donde $\{\xi_n\}$ es una sucesión de variables aleatorias no negativas, independientes e idénticamente distribuidas, y con una función de distribución común F, la cual suponemos admite una función de densidad ρ, es decir

$$F(s) = \int_0^s \rho(t)dt, \forall s \in \mathbb{R}.$$

La expresión (1.6.1) se puede interpretar del siguiente modo: la cantidad de producto en el periodo $n+1$, será la cantidad x_n que existía hasta el periodo anterior, más la cantidad a_n solicitada de producto, menos la demanda que se surte a los clientes. Dado que no hay acumulación de demanda, la cantidad de producto no puede ser negativa, de manera que, si no se logra surtir toda la demanda, el nivel de inventario al siguiente periodo será cero.

Es posible describir la dinámica del sistema de acuerdo a la ley de transición

$$Q(B \mid x, a) := Prob[x_{n+1} \in B \mid x_n = x, a_n = a]$$

$$= Prob[(x_n + a_n - \xi_n)^+ \in B \mid x_n = x, a_n = a]$$

$$= Prob[(x + a - \xi_n)^+ \in B]$$

$$= E_\xi I_B \left[(x + a - \xi_0)^+\right] \quad \forall B \in \mathcal{B}(\mathbf{X}), (x, a) \in \mathbb{K};$$

donde E_ξ representa la esperanza con respecto a la función de distribución $F(\cdot)$.

Observemos que

$$Q(B \mid x, a) = I_B(0)[1 - F(x + a)] + \int_0^{x+a} I_B(x + a - \xi)F(d\xi).$$

La función de costo por etapa para cada $(x, a) \in \mathbb{K}$ está dada por

$$C(x, a) := ca + h(x + a) + pE_\xi(\xi - x - a)^+$$

$$= ca + h(x + a) + p\int (\xi - x - a)^+ \rho(\xi)d\xi; \tag{1.6.2}$$

donde $c, h,$ y p son constantes positivas que representan el costo unitario de producción, el costo unitario por almacenamiento y el costo unitario por demanda insatisfecha, respectivamente.

El resultado que se presenta a continuación, demuestra que bajo condiciones generales el sistema de inventario descrito satisface condiciones de compacidad y continuidad.

Proposición 1.8. *Supongamos que $E_\xi(\xi - x - a)^+$ es finita. Entonces el sistema de inventario cumple con las condiciones de la Hipótesis 1.1*

Demostración. a) La función de costo por etapa

$$C(x,a) := ca + h(x+a) + pE_\xi(\xi - x - a)^+$$

es una función continua en $\mathbb{K}$, ya que $E_\xi(\xi - x - a)^+$ es finita, además también es una función acotada, ya que está definida en el conjunto $\mathbb{K}$ el cual es un conjunto compacto.

b) Es claro que $A(x) = [0, \mathcal{K} - x]$ es un subconjunto no vacío y compacto de $\mathbf{A} = [0, \mathcal{K}]$ para cada $x \in \mathbf{X}$. Veamos también que el mapeo $x \to A(x)$ es continuo. Para esto, considere un subconjunto cerrado $\bar{A}$ de $\mathbf{A} = [0, \mathcal{K}]$. Como $\mathbf{A}$ es compacto, también $\bar{A}$ es compacto; entonces, $a^1 := \min \bar{A}$ y $a^2 := \max \bar{A}$ están bien definidos y pertenecen a $\bar{A}$. Note que el conjunto

$$\{x \in \mathbf{X} : \mathbf{A}(x) \cap \bar{A} \neq \emptyset\} = \left[0, \mathcal{K} - a^1\right]$$

es un subconjunto cerrado en $\mathbf{X}$, por lo que el mapeo es semicontinuo superiormente.

c) Ahora, demostraremos que para cada $B \in \mathcal{B}(\mathbf{X}), (x,a) \in \mathbb{K}$ el kernel de transición

$$Q(B \mid x, a) = E_\xi I_B \left[(x + a - \xi_0)^+\right],$$

es debilmente continuo en $\mathbb{K}$. Notemos que para cada $(x,a) \in \mathbf{K}$ se tiene que el mapeo

$$(x,a) \to \int_{\mathbf{X}} u(y) Q(dy \mid x, a) = E_\xi u\left((x + a - \xi_0)^+\right)$$

es continuo para cada función $u \in \mathbf{C}_b(\mathbf{X})$.

Por lo tanto, el sistema de inventario satisface la Hipótesis 1.1. □

Ahora mostraremos que el sistema de inventario cumple con la Hipotesis 1.2.

Proposición 1.9. *Supongamos que la densidad ρ es continua y acotada en $[0,\infty]$. Entonces el sistema de inventario cumple con la Hipótesis 1.2*

Demostración. Observe que es suficiente demostrar que Q es fuertemente continuo, en $A(x)$, es decir, que la función

$$a \to \int_{\mathbf{X}} u(y) Q(dy \mid x, a) \tag{1.6.3}$$

es continua para cada $u \in \mathbf{M}_b(\mathbf{X})$ y $x \in \mathbf{X}$.

Sea ρ_a la función de densidad de la v.a $y = a - \xi$. Observe que $\rho_a(y) = \rho(a - y)$. Entonces como $\rho(\cdot)$ es continua y acotada, la función $a \to \rho_a(\cdot)$ es continua y acotada. Ahora para $u \in \mathbf{M}_b(\mathbf{X})$. Entonces

$$\int_{\mathbf{X}} u(y) Q(dy \mid x, a) = E[u(x_{t+1}) \mid x_t = x, a_t = a] = \int_{-\infty}^{\infty} u(x+y)^+ \rho_a(y) dy$$

$$= u(0) \int_{-\infty}^{-x} \rho_a(y) dy + \int_0^{\infty} u(y) \rho_a(y - x) dy. \tag{1.6.4}$$

Sea $\{a_n\}$ una sucesión en $A(x), x \in \mathbf{X}$, tal que $a_n \to a$. Entonces $\rho_{a_n}(\cdot) \to \rho_a(\cdot)$. De (1.6.4), el Teorema de Scheffe, y el Teorema de convergencia acotada tenemos, para cada $x \in \mathbf{X}$

$$\int_{\mathbf{X}} u(y)Q(dy \mid x, a_n) = u(0)\int_{-\infty}^{x} \rho_{a_n}(y)dy + \int_0^{\infty} u(y)\rho_{a_n}(y-x)dy$$

$$\to u(0)\int_{-\infty}^{x} \rho_a(y)dy + \int_0^{\infty} u(y)\rho_a(y-x)dy = \int_{\mathbf{X}} u(y)Q(dy \mid x, a).$$

Por lo tanto la función (1.6.3) es continua $\square$

Capítulo 2

Estimación en Modelos Markovianos

2.1 Introducción

Un caso particular del modelo de control de Markov (1.2.1), estudiado en el Capítulo 1, se da cuando la dinámica del sistema está descrita por

$$x_{n+1} = H\left(x_n, a_n, \xi_n\right), \quad n \in \mathbb{N}_0,$$

donde ξ_n, $n \in \mathbb{N}_0$, son variables aleatorias independientes e idénticamente distribuidas, con función de densidad $\rho(\cdot)$ desconocida. En este capítulo se estudiará dicho caso, y se supondrá que las variables aleatorias, ξ_n son observables, lo que nos permitirá obtener una estimación $\rho_t(\cdot)$ de la función de densidad $\rho(\cdot)$ y definir con ello un modelo de control estimado (véase 2.3.1), para el cual se asegura la existencia de la función de valor óptimo correspondiente $V^{*(t)}$. A partir de esta solución $V^{*(t)}$ será posible aproximar la función de valor óptimo V^* del modelo original (1.2.1), donde se obtiene una cota de error por realizar dicho proceso (ver en la Proposición 2.4). Finalmente se retoma el ejemplo de un sistema de inventario presentado en la sección 1.6 del Capítulo 1 para ilustrar el proceso de estimación.

2.2 Modelo de control dado por ecuación en diferencias

Sea el modelo de control de Markov (1.2.1) y considere que este evoluciona de acuerdo a la siguiente ecuación en diferencias

$$x_{n+1} = H\left(x_n, a_n, \xi_n\right), \quad n \in \mathbb{N}_0 \tag{2.2.1}$$

donde $H : \mathbb{K} \times \mathbb{R}^m \to \mathbf{X}$ es una función medible, las variables aleatorias $\xi_n, n \in \mathbb{N}_0$, son independientes e idénticamente distribuidas (i.i.d.) con valores en el espacio euclidiano $\mathbb{R}^m$ y función de densidad $\rho(\cdot)$, la cual suponemos es desconocida para el controlador. Suponemos además que la función de costo por etapa está dada por

$$C(x, a) := \int_{\mathbb{R}^m} \widehat{c}(x, a, \xi)\rho(\xi)d\xi, \quad (x, a) \in \mathbb{K} \tag{2.2.2}$$

donde $\widehat{c} : \mathbb{K} \times \mathbb{R}^m \to \mathbb{R}$ es una función medible.

Para todo $B \in \mathcal{B}(\mathbf{X}), (x, a) \in \mathbb{K}$, la ley de transición para este caso es de la forma

$$Q(B \mid x, a) = \int_{\mathbb{R}^m} I_B(H(x, a, \xi))\rho(\xi)d\xi. \tag{2.2.3}$$

Además, para cada función medible v en $\mathbf{X}$ se satisface la igualdad

$$\int_{\mathbf{X}} v(y)Q(dy \mid x,a) = \int_{\mathbb{R}^m} v((H(x,a,\xi))\rho(\xi)d\xi$$

siempre que las integrales estén bien definidas.

Para este caso particular se plantea el problema de control óptimo descontado del mismo modo que en el Capítulo 1, se consideran los operadores de contracción definidos en (1.3.2), y (1.3.3), y además se usa la misma notación para la función de valor óptimo correspondiente.

A continuación se presentan condiciones bajo las cuales será posible obtener los resultados presentados en el Capítulo 1, dando así solución al problema de control óptimo

Hipótesis 2.1. ***a)*** *$H(\cdot,\cdot,s)$ es continua en $\mathbb{K}$ para cada $s \in \mathbb{R}^m$;*

b) *$\widehat{c}(\cdot,\cdot,s)$ es una función acotada y continua en $\mathbb{K}$ para cada $s \in \mathbb{R}^m$;*

c) *la multifunción $x \to A(x)$ es continua con valores en la familia de los conjuntos compactos de $\mathbf{A}$.*

Observación 2.1. *La Hipótesis 2.1 implica la Hipótesis 1.1. Es decir*

a) *la función de costo $C(x,a) = \int_{\mathbb{R}^m} \widehat{c}(x,a,\xi)\rho(\xi)$*

es una función continua y acotada;

b) *la ley de transición $Q(B \mid x,a) = \int_{\mathbb{R}^m} I_B(H(x,a,\xi))\rho(\xi)d\xi$ es débilmente continua en $\mathbb{K}$, es decir, la función*

$$(x,a) \to \int_{\mathbf{X}} u(y)Q(dy \mid x,a) = \int_{\mathbb{R}^m} u((H(x,a,\xi))\rho(\xi)d\xi$$

es continua para cada $u \in \mathbf{C}_b(\mathbf{X})$

Lo anterior se sigue del Teorema de la convergencia acotada y del hecho de que si u es una función en $\mathbf{C}_b(\mathbf{X})$ entonces también lo es $u((H(x,a,\xi)) \quad \forall x \in \mathbf{X}$

Notemos entonces que el modelo de control (2.2.1) satisface las condiciones de la Hipótesis 1.1, por lo tanto, como ya se dijo en la Observación 1.5 a), los resultados de la Proposición 1.2 son válidos ahora en el espacio de Banach $(\mathbf{C}_b(\mathbf{X}), \|\cdot\|_\infty)$; es decir, se tiene el siguiente resultado.

Proposición 2.2. *Bajo la Hipótesis 2.1 la función de valor óptimo V^* es el único punto fijo del operador de contracción T en el espacio de Banach $(\mathbf{C}_b(\mathbf{X}), \|\cdot\|_\infty)$, esto es*

$$V^*(x) = TV^*(x) = \min_{a \in A(x)} \left[C(x,a) + \alpha \int_{\mathbf{X}} V^*(y)Q(dy \mid x,a) \right],$$

donde la función de costo por etapa $C(x,a)$ y la ley de transición $Q(B \mid x,a)$ están dadas por (2.2.2) y (2.2.3) respectivamente.

2.3 Modelo estimado

Consideremos de nuevo un modelo de control que evoluciona de acuerdo a la ecuación en diferencias (2.2.1), es decir

$$x_{n+1} = H\left(x_n, a_n, \xi_n\right), \quad n \in \mathbb{N}_0$$

donde las variables aleatorias $\xi_n, n \in \mathbb{N}$, tienen densidad $\rho(\cdot)$ desconocida y supongamos que dichas variables son observables. Sea $\boldsymbol{\xi_t} := (\xi_0, \xi_1, \ldots, \xi_{t-1})$ una muestra observada por el controlador. Entonces, basado en $\boldsymbol{\xi_t}$, el controlador obtiene una densidad estimada

$$\rho_t(\cdot) = \rho_t\left(\cdot \mid \boldsymbol{\xi}_t\right)$$

de la densidad desconocida $\rho(\cdot)$, lo que nos permitirá construir un modelo estimado.

Un modelo de control estimado es un arreglo de forma

$$\left(\mathbf{X}, \mathbf{A}, \{A(x) : x \in \mathbf{X}\}, Q^{(t)}, C^{(t)}\right) \tag{2.3.1}$$

donde para cada $(x, a) \in \mathbb{K}$ y $B \in \mathcal{B}(\mathbf{X})$ se definen la función de costo por etapa y el kernel de transición considerando la función de densidad estimada $\rho_t(\cdot)$. Esto es

$$\begin{aligned} Q^{(t)}(B \mid x, a) &:= \int_{\mathbb{R}^m} I_B\left(H(x, a, \xi\right)\rho_t(\xi) d\xi \\ C^{(t)}(x, a) &:= \int_{\mathbb{R}^m} \widehat{c}(x, a, \xi)\rho_t(\xi) d\xi \end{aligned}$$

Definición 2.1. *El costo descontado $V_f^{(t)}$ bajo la política $f \in \mathbb{F}$ para el modelo estimado, se define como*

$$V_f^{(t)}(x) := E_{x,f}^{(t)} \sum_{k=0}^{\infty} \alpha^k C_f^{(t)}\left(x_k^{(t)}\right), x \in \mathbf{X}$$

donde $\left\{x_k^{(t)}\right\}$ denota la cadena de Markov inducida por la política $f \in \mathbb{F}$ en el modelo estimado.

La función de valor óptimo para este caso, se define como

$$V^{*(t)}(x) := \inf_{f \in \mathbb{F}} V_f^{(t)}(x).$$

De manera análoga al operador T definido en (1.3.3), denotamos por T_t al operador de programación dinámica para el modelo estimado, es decir para cada $u \in \mathbf{C}_b(\mathbf{X})$

$$\begin{aligned} T_t u(x) &:= \min_{a \in A(x)} \left\{C^{(t)}(x, a) + \alpha \int_{\mathbb{R}^m} u(y) Q^{(t)}(dy \mid x, a)\right\} \\ &= \min_{a \in A(x)} \left\{C^{(t)}(x, a) + \alpha \int_{\mathbb{R}^m} u(H(x, a, \xi))\rho_t(\xi) d\xi\right\}. \end{aligned}$$

Proposición 2.3. *Bajo la Hipótesis 2.1 se tiene que*

a) *La función de costo en el modelo estimado*

$$C^{(t)}(x,a)=\int_{\mathbb{R}^m}\widehat{c}(x,a,\xi)\rho_t(\xi)d\xi$$

es una función continua y acotada;

b) *el kernel estocástico $Q^{(t)}(\cdot \mid x,a)$ es débilmente continuo en $\mathbb{K}$, es decir, la función*

$$(x,a)\to\int_{\mathbf{X}}u(y)I_B\left(H(x,a,\xi)\right)\rho_t(\xi)d\xi$$

es continua para cada $u\in\mathbf{C}_b(\mathbf{X})$;

c) *el operador T_t es de contracción módulo α en el espacio de Banach $(\mathbf{C}_b(\mathbf{X}),\|\cdot\|_\infty)$ y su único punto fijo es la función de valor óptimo $V^{*(t)}$.*

Demostración. **a)** Notemos que la función de costo por etapa $C^{(t)}(x,a)$ está definida de manera análoga a la función de costo $C(x,a)$, pero considerando ahora la función de densidad estimada $\rho_t(\cdot)$, de manera que análogamente a la Observación 2.1 se tiene que $C^{(t)}(x,a)$ también es continua y acotada.

b) De manera análoga a la Observación 2.1, el kérnel de transición $Q^{(t)}(\cdot \mid x,a)$ del caso estimado es débilmente continuo.

c) Notemos que este caso es análogo al resultado dado en la Proposición 2.2 considerando en este caso la función de costo por etapa y el kernel de transición definidos de acuerdo a la función de densidad estimada $\rho_t(\cdot)$.

□

La siguiente Proposición da una cota de error por estimar la función de valor óptimo V^* en el modelo (2.2.1) mediante la función de valor óptimo $V^{*(t)}$ del modelo estimado. Dicho error será medido en términos de la siguiente cantidad

$$\eta_t:=\int_{\mathbb{R}^m}\left|\rho(\xi)-\rho_t(\xi)\right|d\xi, t\in\mathbb{N}$$

.

Proposición 2.4. *Bajo la Hipótesis 2.1, para cada $t\geq 0$ se cumple*

$$\left\|V^*-V^{*(t)}\right\|_\infty\leq\left[\frac{M}{1-\alpha}+\frac{\alpha M}{(1-\alpha)^2}\right]\eta_t. \tag{2.3.2}$$

Demostración. Obsereve que para cada $t\in\mathbb{N}$, como T_t y T son de contracción, tenemos,

$$\begin{aligned}|V^*(\cdot)-V^{*(t)}(\cdot)|&=\left|TV^*(\cdot)-T_tV^*(\cdot)+T_tV^*(\cdot)-T_tV^{*(t)}(\cdot)\right|\\&\leq|TV^*(\cdot)-T_tV^*(\cdot)|+\alpha\left|V^*(\cdot)-V^{*(t)}(\cdot)\right|,\quad\forall t\in\mathbb{N}.\end{aligned}$$

Entonces

$$|V^*(\cdot)-V^{*(t)}(\cdot)|-\alpha|V^*(\cdot)-V^{*(t)}(\cdot)|\leq|TV^*(\cdot)-T_tV^*(\cdot)|,$$

lo cual implica

$$\left|V^*(\cdot)-V^{*(t)}(\cdot)\right| \leq \frac{1}{1-\alpha}\left|TV^*(\cdot)-T_tV^*(\cdot)\right|. \tag{2.3.3}$$

Por otra parte,

$$\begin{aligned}
|TV^*(x)-T_tV^*(x)| \leq & \sup_{a\in A(x)}\Big[\,|\,C(x,a)-C^{(t)}(x,a)\\
& +\alpha\int_{\mathbb{R}^m} V_*(H(x,a,w))\rho(w)dw-\alpha\int_{\mathbb{R}^m} V_*(H(x,a,w))\rho_t(w)dw\,|\Big]\\
\leq & \sup_{a\in A(x)}\left[\int_{\mathbb{R}^m}\hat{c}(x,a,\xi)\,|\rho(\xi)-\rho_t(\xi)|\,d\xi\right.\\
& \left.+\alpha\int_{\mathbb{R}^m} V^*(H(x,a,\xi))\,|\rho(\xi)-\rho_t(\xi)|\,d\xi\right]\\
\leq & M\int_{\mathbb{R}^m}|\rho(\xi)-\rho_t(\xi)|\,d\xi+\frac{\alpha M}{1-\alpha}\int_{\mathbb{R}^m}|\rho(\xi)-\rho_t(\xi)|\,d\xi,\quad t\in\mathbb{N}\\
\leq & M\eta_t+\frac{\alpha M}{1-\alpha}\eta_t,\quad x\in\mathbf{X}.
\end{aligned}$$

De esta última desigualdad y de (2.3.3) se tiene que

$$\left|V^*(\cdot)-V^{*(t)}(\cdot)\right| \leq \frac{1}{1-\alpha}\left[M\eta_t+\frac{\alpha M}{1-\alpha}\eta_t\right],$$

lo cual demuestra (2.3.2). □

Existen distintos métodos para estimar funciones de densidad, uno de ellos es el método de máxima verosimilitud, el cual se utiliza cuando la densidad desconocida ρ pertenece a una familia paramétrica $\{\rho_\theta : \theta \in, \Theta\}$. Además dicho método permite encontrar estimadores consistentes. Para ilustrar esto se presenta el siguiente ejemplo.

Ejemplo 2.1. *Consideremos el modelo de control (2.2.1), donde*

$$x_{n+1}=H\left(x_n,a_n,\xi_n\right),\quad n\in\mathbb{N}_0$$

y supongamos que la densidad desconocida $\rho(\cdot)$ de las variables aleatorias ξ_n pertenece a la siguiente familia paramétrica

$$\rho_\theta(\xi)=\frac{\varphi(\xi)}{\displaystyle\int_\theta^\infty\varphi(s)ds}\mathbf{I}_{[\theta,\infty)}(\xi),\quad \theta\in\Theta$$

donde $\varphi:\mathbb{R}\to\mathbb{R}_+$ es una función medible tal que $\displaystyle\int_{-\infty}^{\infty}\varphi(s)ds$ es finita y Θ es un subconjunto de $\mathbb{R}^+$.

Para encontrar una estimación ρ_{θ_t} de la densidad ρ, en este trabajo usaremos el método de máxima verosimilitud. Este método consiste en obtener el valor de θ que maximiza la función de verosimilitud, lo cual significa escoger como valor estimado aquel que tiene mayor probabilidad de ocurrir según la muestra observada.

La función de verosimilitud se define como:

$$\mathcal{L}(\theta) := \prod_{i=1}^{n} \rho_\theta\left(\xi_i\right) = \prod_{i=1}^{n} \frac{\varphi(\xi_i)}{\displaystyle\int_\theta^\infty \varphi(s)ds} \mathbf{I}_{[\theta,\infty)}(\xi_i), \quad \theta \in \Theta$$

Observe que, como φ es una función positiva, entonces para $\theta_1 < \theta_2$ se tiene que

$$\int_{\theta_1}^\infty \varphi(s)ds > \int_{\theta_2}^\infty \varphi(s)ds.$$

Entonces

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} \frac{\varphi(\xi_i)}{\displaystyle\int_\theta^\infty \varphi(s)ds}$$

es una función creciente de θ.

Además, el caso de interés se da cuando $\theta \leq \xi_1, \xi_2, ..., \xi_n$, de manera que, el valor de θ que maximiza $\mathcal{L}(\theta)$ es

$$\theta_t := \min\left\{\xi_0, \xi_1, \ldots, \xi_{t-1}\right\},$$

es decir θ_t es el estimador de máxima verosimilitud de θ.

Este estimador es fuertemente consistente, lo cual significa que se cumple

$$\theta_t \to \theta \quad P_\theta - a.s. \text{ cuando } t \to \infty.$$

Para demostrar esta última afirmación, denotemos por F_θ y F_{θ_t} las funciones de distribución de θ y θ_t respectivamente, y sean ρ_θ y ρ_{θ_t} sus densidades correspondientes. Entonces, para cualquier $\varepsilon > 0$, obtenemos

$$\begin{aligned} P_\theta\left[|\theta_t - \theta| > \varepsilon\right] &= P_\theta\left[\theta_t > \theta + \varepsilon\right] \\ &= 1 - F_{\theta_t}(\theta + \varepsilon) \\ &= 1 - (1 - \prod_{i=1}^{t}(1 - F_{\theta_i}(\theta + \varepsilon))) \\ &= \left(1 - F_\theta(\theta + \varepsilon)\right)^t; \end{aligned}$$

entonces,

$$\sum_{t=1}^{\infty} P_\theta\left[|\theta_t - \theta| > \varepsilon\right] \leq \sum_{t=1}^{\infty} \left(1 - F_\theta(\theta + \varepsilon)\right)^t < \infty$$

lo cual combinado con el lema de Borel-Cantelli implica

$$\theta_t \to \theta \quad P_\theta - a.s. \quad \forall \theta \in \Theta.$$

Ahora, se demuestra que

$$\bar{\eta}_t := E\int_{\mathbb{R}^m} |\rho_{\theta_t}(\xi) - \rho_\theta(\xi)|\, d\xi \to 0.$$

Para esto notemos que $\rho_{\theta_t}(\xi) = 0$ para $\xi \leq \theta_t$ lo que implica

$$|\rho_{\theta_t}(\xi) - \rho_\theta(\xi)| = \left\{ \begin{array}{ccc} \rho_\theta(\xi) & si & \theta < \xi \leq \theta_t \\ \rho_{\theta_t}(\xi) - \rho_\theta(\xi) & si & \xi > \theta_t \end{array} \right. .$$

De aquí y notando que $\displaystyle\int_{\theta_t}^{\infty} \rho_{\theta_t}(\xi) d\xi = 1$ se tiene que

$$\begin{aligned}
\eta_t &= \int_\theta^\infty |\rho_{\theta_t}(\xi) - \rho_\theta(\xi)|\, d\xi \\
&= \int_\theta^{\theta_t} |\rho_{\theta_t}(\xi) - \rho_\theta(\xi)|\, d\xi + \int_{\theta_t}^\infty |\rho_{\theta_t}(\xi) - \rho_\theta(\xi)|\, d\xi \\
&= \int_\theta^{\theta_t} \rho_\theta(\xi) d\xi + \int_{\theta_t}^\infty \rho_{\theta_t}(\xi) d\xi - \int_{\theta_t}^\infty \rho_\theta(\xi) d\xi \\
&= \int_\theta^{\theta_t} \rho_\theta(\xi) d\xi + 1 - (1 - \int_\theta^{\theta_t} \rho_\theta(\xi) d\xi) \\
&= 2\int_\theta^{\theta_t} \rho_\theta(\xi) d\xi = 2F_\theta\left(\theta_t\right).
\end{aligned}$$

La consistencia fuerte de los estimadores $\{\theta_t\}$ y la continuidad por la derecha de F_θ implican que

$$\eta_t \to 0 \quad P_\theta - a.s, \quad t \to \infty.$$

Por otra parte tenemos

$$\begin{aligned}
\bar{\eta}_t &= E\int_{\mathbb{R}} |\rho_{\theta_t}(\xi) - \rho_\theta(\xi)|\, d\xi = 2E\left(F_\theta\left(\theta_t\right)\right) \\
&= 2\int_\theta^\infty F_\theta(\xi)\rho_{\theta_t}(\xi) d\xi.
\end{aligned}$$

Además notemos que

$$\begin{aligned}
\rho_{\theta_t}(\xi) &= F'_{\theta_t}(\xi) = \left(1 - (1 - F_\theta(\xi))^t\right)' \\
&= t\left(1 - F_\theta(\xi)\right)^{t-1} \rho_\theta(\xi).
\end{aligned}$$

De las igualdades anteriores se tiene que

$$\bar{\eta}_t = 2\int_\theta^\infty t\rho_\theta(\xi) F_\theta(\xi)\left(1 - F_\theta(\xi)\right)^{t-1} d\xi.$$

Haciendo el cambio de variable $w = F_\theta(\xi)$, obtenemos que

$$\bar{\eta}_t = 2t\int_\theta^\infty w(1-w)^{t-1} dw.$$

Entonces resolviendo con el método de integración por partes y tomando en cuenta que $F_\theta(\theta) = 0$, obtenemos que

$$\bar{\eta}_t = \frac{2}{t+1} \to 0, \text{ cuando } t \to \infty.$$

Ejemplo 2.2. *Este ejemplo se presenta como continuación al ejemplo de un sistema de inventario en la sección 1.6 del Capítulo 1. Para el cual se tiene que*

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

Consideremos que la función de densidad $\rho(\cdot)$ *de la demanda* ξ_n *pertenece a la siguiente familia paramétrica de densidades*

$$\rho_\theta(z) = \lambda \exp(-\lambda(z-\theta)) I_{[\theta,\infty)}(z),$$

donde el parámetro θ *es desconocido y está en algún intervalo* $\Theta = [\theta_1, \theta_2]$ *para* $0 \leq \theta_1 < \theta_2$. *Observemos que* ρ_θ *es un caso particular de la función de densidad*

$$\rho_\theta(\xi) = \frac{\varphi(\xi)}{\displaystyle\int_\theta^\infty \varphi(s)ds} \mathbf{I}_{[\theta,\infty)}(\xi), \quad \theta \in \Theta$$

presentada en el Ejemplo 2.1, tomando $\varphi(z) = \exp(-\lambda z)$.

Para estimar el parametro θ *se asume que las variables aleatorias* $\xi_n, n \in \mathbb{N}$, *son observables y que una muestra* $\boldsymbol{\xi}_t := (\xi_0, \xi_1, \ldots, \xi_{t-1})$ *está disponible para el controlador. Entonces como ya se desarrollo en el Ejemplo 2.1, el estimador de máxima verosimilitud de* θ *es*

$$\theta_t = \min\{\xi_0, \xi_1, \ldots, \xi_{t-1}\},$$

de manera que, un estimador de la densidad está dado por

$$\rho_t(z) = \lambda \exp\left(-\lambda\left(z - \theta_t\right)\right) I_{[\theta_t,\infty)}(z).$$

Por lo tanto $\theta_t \to \theta \quad P_\theta - c.s.$, *y además se cumple que*

$$\eta_t = \int_\theta^\infty \left|\rho_{\theta_t}(s) - \rho_\theta(s)\right| ds \to 0, \quad t \to \infty.$$

Capítulo 3

Proceso de Control Semi-Markoviano

3.1 Introducción

En este capítulo introducimos los modelos de control semi-markovianos, en los cuales se considera que la transición del sistema de un estado a otro se lleva a cabo en un tiempo aleatorio, es decir, que el tiempo de permanencia en cada estado es una variable aleatoria $\delta_n, n \in \mathbb{N}$. Un primer objetivo de este capítulo es introducir la teoría básica de estos procesos, asi como definir y establecer condiciones bajo las cuales se da solución al problema de control óptimo, bajo un índice de funcionamiento descontado. La teoría se desarrollará asumiendo costos posiblemente no acotados. Entonces, considerando que los procesos semi-markovianos son una generalización de los procesos markovianos, los resultados establecidos en este capítulo extienden los del capítulo anterior.

3.2 Modelo de control semi-markoviano

Un Modelo de control semi-markoviano es un arreglo de la forma

$$(\mathbf{X}, \mathbf{A}, \{A(x) : x \in \mathbf{X}\}, Q, C) \tag{3.2.1}$$

que consta de los siguienes elementos:

i) $\mathbf{X}$ y $\mathbf{A}$ son espacios de Borel y representan los espacios de estados y controles, respectivamente.

ii) $\{A(x) : x \in \mathbf{X}\}$ es la familia de conjuntos de acciones admisibles. Definimos el conjunto

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

de pares estado-acción admisibles.

iii) $Q(\cdot, \cdot \mid x, a)$ es un kernel estocástico en $\mathbf{X} \times \mathbb{R}^+$ dado $\mathbb{K}$.

iv) $C(x, a)$ es una función medible en $\mathbb{K}$ que representa el costo por época de decisión.

Un modelo de control semi-markoviano representa un sistema dinámico que evoluciona a tiempo continuo de modo que, para cada $n \in \mathbb{N}$, en la n-ésima etapa de decisión, el sistema se encuentra en el estado $x_n = x \in \mathbf{X}$ y se elige un control $a_n = a \in A(x)$. Entonces sucede lo siguiente:

i) se genera un costo $C(x, a)$;

ii) el sistema permanece en el estado x durante un tiempo aleatorio δ_{n+1}, y se mueve a un nuevo estado $x_{n+1} = y \in \mathbf{X}$ de acuerdo a la probabilidad conjunta del evento $x_{n+1} \in B$ y $\delta_{n+1} \leq t$ dada por el kernel $Q(B, [0, t] \mid x, a)$, donde $B \in \mathcal{B}(\mathbf{X})$ y $t \in \mathbb{R}^+$, es decir

$$Q(B, [0, t] \mid x_n, a_n) = Prob[x_{n+1} \in B, \delta_{n+1} \leq t \mid x_n = x, a_n = a].$$

Una vez que el sistema se encuentra en el nuevo estado, se elige un nuevo control y el proceso se repite.

En adelante, para cada $n \in \mathbb{N}_0$, las variables coordenadas $x_n \in \mathbf{X}$, $a_n \in A(x_n)$ y $\delta_{n+1} \in \mathbb{R}^+$ denotan el estado del sistema inmediatamente después de la n-ésima transición, el control seleccionado y el tiempo que el sistema permanece en el estado x_n, respectivamente.

A las variables aleatorias $\delta_{n+1}, n \in \mathbb{N}_0$, se les llama tiempos de permanencia y las variables aleatorias

$$T_0 := 0, \quad \text{y } T_n := T_{n-1} + \delta_n \text{ para } n \in \mathbb{N} \tag{3.2.2}$$

se les llama los tiempos de transición o épocas de decisión.

Ahora el espacio de historias admisibles se define considerando los tiempos de permanencia. Es decir, para cada $n \in \mathbb{N}_0$

$$\mathbf{H}_0 := \mathbf{X},$$

$$\mathbf{H}_n := (\mathbb{K} \times \mathbb{R}^+)^n \times \mathbf{X},$$

de manera que un elemento genérico $h_n \in \mathbf{H}_n$ es un vector de la forma

$$h_n = (x_0, a_0, \delta_1, \ldots, \delta_{n-1}, x_{n-1}, a_{n-1}, \delta_n, x_n).$$

En este sentido, las políticas de control son definidas de manera similar que en el caso markoviano (Definición 1.3) pero considerando ahora que las historias admisibles incluyen los tiempos de permanencia.

Denotamos por Π al conjunto de todas las políticas y por $\mathbb{F}$ al conjunto de políticas estacionarias.

Sea $(\Omega, \mathcal{F})$ el espacio medible canónico que consiste del espacio muestral $\Omega := (\mathbf{X} \times \mathbf{A} \times \mathbb{R}^+)^\infty$ y su σ álgebra producto $\mathcal{F}$. Entonces, para cada estado inicial $x \in \mathbf{X}$ y política $\pi \in \Pi$, existe una medida de probabilidad $\mathbb{P}_x^\pi$ en el espacio $(\Omega, \mathcal{F})$ tal que para cada $D \in \mathcal{B}(\mathbf{A}), B \in \mathcal{B}(\mathbf{X}), h_n \in \mathbb{H}_n$ $n \in \mathbb{N}$, [[2] Theorem 2.7.2 p 109] se cumple lo siguiente:

$$\mathbb{P}_x^\pi [x_0 = x] = 1 \tag{3.2.3}$$

$$\mathbb{P}_x^\pi [a_n \in D \mid h_n] = \pi_n (D \mid h_n) \tag{3.2.4}$$

$$\mathbb{P}_x^\pi [x_{n+1} \in B, \delta_{n+1} \leq t \mid h_n, a_n] = Q(B, (0, t] \mid x_n, a_n). \tag{3.2.5}$$

De la misma manera E_x^π denota al operador esperanza correspondiente a $\mathbb{P}_x^\pi$.

Para $t \in \mathbb{R}^+$ y para cada par estado-acción admisible $(x, a) \in \mathbb{K}$ se definen las distribuciones marginales por

$$G(t \mid x, a) := Q(\mathbf{X}, [0, t] \mid x, a), \tag{3.2.6}$$

$$P(B \mid x, a) := Q\left(B, \mathbb{R}^+ \mid x, a\right). \tag{3.2.7}$$

Se dice que un proceso es regular cuando presenta un número finito de transiciones en intervalos de tiempo acotados. Para estudiar dicha propiedad se definen las variables aleatorias

$$N(t) := \sup \{n \geq 1 : T_n \leq t\}, t \in \mathbb{R}^+$$

las cuales cuentan el número de transiciones ocurridas en el intervalo de tiempo $[0, t]$.

Definición 3.1. *Para $\pi \in \Pi, x \in \mathbf{X}$, diremos que un proceso es regular en x si*

$$\mathbb{P}_x^\pi \left[T_n \to \infty \right] = 1$$

Si la propiedad anterior se cumple para todo $x \in \mathbf{X}$ diremos que el proceso es regular.

En la siguiente sección se darán condiciones para asegurar esta condición de regularidad.

3.3 Indice descontado y condiciones de optimalidad

A continuación definimos el criterio de optimalidad, asumiendo que para un modelo de control semi-markoviano los costos son descontados en forma continua. Es decir, dado un factor de descuento $\alpha > 0$ un costo unitario generado al tiempo t es equivalente a un costo $\exp(-\alpha t)$ en el tiempo presente.

Definición 3.2. *Para $\alpha > 0$, se define el costo descontado semi-markoviano por*

$$V(x, \pi) := E_x^\pi \sum_{n=0}^{\infty} e^{-\alpha T_n} C\left(x_n, a_n\right), x \in \mathbf{X}, \pi \in \Pi$$

La función de valor óptimo se define como

$$V^*(x) := \inf_{\pi \in \Pi} V(x, \pi), \quad x \in \mathbf{X}$$

Entonces el problema de control óptimo consiste en encontrar una política π^* tal que

$$V^*(x) = V(x, \pi^*), \quad x \in \mathbf{X}$$

a la cual llamaremos óptima.

Del mismo modo que en el modelo de control markoviano, es necesario establecer condiciones bajo las cuales se garantiza que el criterio de optimalidad para el modelo semi-markoviano está bien definido para cada política, asi como la existencia de políticas óptimas.

Hipótesis 3.1. ***a)*** *Para cada $x \in \mathbf{X}$, el conjunto $A(x)$ es compacto.*

b) *$C(x, a)$ es una función continua en $A(x)$ para cada $x \in \mathbf{X}$.*

c) *La función*

$$a \to \int_{\mathbf{X} \times \mathbb{R}^+} e^{-\alpha t} u(y) Q(dy, dt \mid x, a)$$

es continua para cada $u \in \mathbf{M}_b(\mathbf{X}), x \in \mathbf{X}$.

Hipótesis 3.2. *Existe una función medible $W : \mathbf{X} \to [1, \infty)$, y constantes $\beta \in (0, 1)$, y $\bar{c} \in \mathbb{R}$ tales que para cada $x \in \mathbf{X}$ se tiene que*

a) $|C(x, a)| \leq \bar{c} W(x), \quad \forall a \in A(x),$

b) $\displaystyle\int_{\mathbf{X} \times \mathbb{R}^+} e^{-\alpha t} W(y) Q(dy, dt \mid x, a) \leq \beta W(x), \quad \forall a \in A(x),$

c) *la función*

$$a \to \int_{\mathbf{X}\times\mathbb{R}^{+}} e^{-\alpha t} W(y) Q(dy, dt \mid x, a)$$

es continua en $A(x)$.

Sea $W : \mathbf{X} \to [1, \infty)$ la función medible de la Hipótesis 3.2. Para cada función medible $u : \mathbf{X} \to \mathbb{R}$, se define la semi norma-W como

$$\|u\|_W := \sup_{x\in\mathbf{X}} \frac{|u(x)|}{W(x)}.$$

Consideremos el espacio lineal normado, de funciones medibles $u : \mathbf{X} \to \mathbb{R}$ tales que

$$\|u\|_W < \infty.$$

Este espacio será denotado por $B_W(\mathbf{X})$. Como consecuencia de la Hipótesis 3.2 se obtienen los siguientes resultados.

Proposición 3.1. *Bajo la Hipótesis 3.2, se tiene que para cada* $x \in \mathbf{X}, \pi \in \Pi$, *y* $n \in \mathbb{N}$, *se cumple*

$$E_x^\pi \left[e^{-\alpha T_n} W(x_n)\right] \leq \beta^n W(x) \tag{3.3.1}$$

Demostración. Por propiedades de la esperanza condicional y de la Hipótesis 3.2 b), se tiene que

$$\begin{aligned}
E_x^\pi \left[e^{-\alpha T_n} W(x_n)\right] &= E_x^\pi \left[E_x^\pi \left\{e^{-\alpha T_{n-1}} e^{-\alpha \delta_n} W(x_n) \mid h_{n-1}, a_{n-1}\right\}\right] \\
&= E_x^\pi \left[E_x^\pi \left\{e^{-\alpha T_{n-1}} e^{-\alpha \delta_n} W(x_n) \mid h_{n-1}, a_{n-1}\right\}\right] \\
&= E_x^\pi \left[e^{-\alpha T_{n-1}} E_x^\pi \left\{e^{-\alpha \delta_n} W(x_n) \mid h_{n-1}, a_{n-1}\right\}\right] \\
&= E_x^\pi \left[e^{-\alpha T_{n-1}} \int_{\mathbf{A}} \int_{\mathbf{X}\times\mathbb{R}_+} e^{-\alpha t} W(y) Q(dy, dt \mid x_{n-1}, a)\, \pi(da \mid h_{n-1})\right] \\
&\leq E_x^\pi \left[e^{-\alpha T_{n-1}} \int_{\mathbf{A}} \beta W(x_{n-1})\, \pi(da \mid h_{n-1})\right] \\
&= \beta E_x^\pi \left[e^{-\alpha T_{n-1}} W(x_{n-1})\right].
\end{aligned}$$

Entonces

$$E_x^\pi \left[e^{-\alpha T_n} W(x_n)\right] \leq \beta E_x^\pi \left[e^{-\alpha T_{n-1}} W(x_{n-1})\right].$$

Iterando esta última desigualdad obtenemos que

$$E_x^\pi \left\{e^{-\alpha T_n} W(x_n)\right\} \leq \beta^n W(x).$$

□

Corolario 3.2. *Si se cumple la Hipótesis 3.2, entonces*

a)

$$E_x^\pi \sum_{n=0}^{\infty} \left[e^{-\alpha T_n} W(x_n)\right] \leq \frac{W(x)}{1-\beta}, \quad \forall x \in \mathbf{X}, \pi \in \Pi, n \in \mathbb{N}, \tag{3.3.2}$$

b) *para cada política $\pi \in \Pi$, se tiene*

$$\|V(x,\pi)\|_W \leq \frac{\bar{c}}{1-\beta}.$$

Demostración. **a)** De la desigualdad (3.3.1) y dado que $\beta \in (0,1)$, se tiene que

$$\sum_{n=0}^{\infty} E_x^\pi \left\{e^{-\alpha T_n} W(x_n)\right\} \leq \sum_{n=0}^{\infty} \beta^n W(x) = \frac{W(x)}{1-\beta},$$

de donde se obtiene la desigualdad (3.3.2).

b) De la definición de costo descontado y de la Hipótesis 3.2 a), se tiene que

$$\begin{aligned} |V(x,\pi)| &\leq E_x^\pi \sum_{n=0}^{\infty} e^{-\alpha T_n} |C(x_n, a_n)| \\ &\leq \bar{c} E_x^\pi \sum_{n=0}^{\infty} e^{-\alpha T_n} W(x_n) \\ &\leq \frac{\bar{c}}{1-\beta} W(x). \end{aligned}$$

Por lo tanto,

$$\|V(x,\pi)\|_W \leq \frac{\bar{c}}{1-\beta}.$$

□

Corolario 3.3. *Bajo la Hipótesis 3.2, para cada $u \in B_W(\mathbf{X})$, se tiene que*

$$E_x^\pi e^{-\alpha T_n} u(x_n) \leq W(x)\|u\|_W \beta^n, \quad \forall x \in \mathbf{X}, \pi \in \Pi$$

Demostración. Por propiedades de la esperanza condicional tenemos que

$$\begin{aligned} & E_x^\pi \left\{e^{-\alpha T_n} u(x_n)\right\} \\ &= E_x^\pi \left[E_x^\pi \left\{e^{-\alpha T_{n-1}} e^{-\alpha \delta_n} u(x_n) \mid h_{n-1}, a_{n-1}\right\}\right] \\ &= E_x^\pi \left[e^{-\alpha T_{n-1}} E_x^\pi \left\{e^{-\alpha \delta_n} u(x_n) \mid h_{n-1}, a_{n-1}\right\}\right] \\ &= E_x^\pi \left[e^{-\alpha T_{n-1}} \int_{\mathbf{A}} \int_{\mathbf{X}\times\mathbb{R}_+} e^{-\alpha t} u(y) Q(dy, dt \mid x_{n-1}, a)\, \pi(da \mid h_{n-1})\right] \end{aligned}$$

$$\leq \|u\|_W E_x^{\pi}\left[e^{-\alpha T_{n-1}}\int_{\mathbf{A}}\beta W\left(x_{n-1}\right)\pi\left(da \mid h_{n-1}\right)\right]$$

$$= \|u\|_W \beta E_x^{\pi}\left[e^{-\alpha T_{n-1}} W\left(x_{n-1}\right)\right]$$

$$\leq W(x)\|u\|_W \beta^n, \quad \forall x \in \mathbf{X}, \pi \in \Pi.$$

Esta última desigualdad se sigue de la Proposición 3.1, la cual implica

$$E_x^{\pi} e^{-\alpha T_n} u\left(x_n\right) \leq W(x)\|u\|_W \beta^n, \quad \forall x \in \mathbf{X}, \pi \in \Pi.$$

□

En el siguiente resultado se demuestra que el proceso es regular.

Corolario 3.4. *Para cada politica $\pi \in \Pi$ y estado inicial $x_0 = x \in \mathbf{X}$, la sucesión $T_n, n \in \mathbb{N}$, diverge a infinito $P_x^{\pi}-$ casi seguramente.*

Demostración. Se tiene de (3.3.1) que

$$E_x^{\pi}\left\{e^{-\alpha T_n}\right\} \leq E_x^{\pi}\left\{e^{-\alpha T_n} W\left(x_n\right)\right\} \leq \beta^n W(x) \rightarrow 0,$$

lo cual implica que $T_n \rightarrow \infty \quad P_x^{\pi}$ -casi seguramente, para cada estado inicial $x \in \mathbf{X}$. □

3.4 Solución al problema de control óptimo.

Para cada función $u \in B_W(\mathbf{X})$ se define el operador de programación dinámica en el modelo semi-markoviano como

$$Tu(x) := \inf_{a \in A(x)}\left[C(x, a) + \int_{\mathbf{X} \times \mathbb{R}^+} e^{-\alpha t} u(y) Q(dy, dt \mid x, a)\right], \quad x \in \mathbf{X}. \tag{3.4.1}$$

Proposición 3.5. *Supongamos que se cumplen las Hipótesis 3.1 y 3.2. Entonces:*

a) *Para cada $u \in B_W(\mathbf{X}), Tu \in B_W(\mathbf{X})$ y existe $f \in \mathbb{F}$ tal que*

$$Tu(x) = C_f(x) + \int_{\mathbf{X} \times \mathbb{R}^+} e^{-\alpha t} u(y) Q_f(dy, dt \mid x)$$

b) *T es un operador de contracción con módulo β en el espacio de Banach $\left(B_W(\mathbf{X}), \|\cdot\|_W\right)$.*

Demostración. **a)** De la Hipótesis 3.1 c) para cada $u \in \mathbf{M}_b(\mathbf{X}), x \in \mathbf{X}$ la función

$$a \rightarrow \int_{\mathbf{X} \times \mathbb{R}^+} e^{-\alpha t} u(y) Q(dy, dt \mid x, a)$$

es continua. Por otro lado de la Hipótesis 3.2 la función

$$a \rightarrow \int_{\mathbf{X} \times \mathbb{R}^+} e^{-\alpha t} W(y) Q(dy, dt \mid x, a)$$

es continua en $A(x)$. Por lo que de [10, Lema 8.3.7] se sigue que para cada $u \in B_W(\mathbf{X})$, la función

$$a \to C(x,a) + \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} u(y) Q(dy, dt \mid x, a)$$

es continua en $A(x)$ para cada $x \in \mathbf{X}$. Entonces por un teorema de selección medible ([10] Lema 8.3.8) Tu es una función medible y existe $f \in \mathbb{F}$ tal que

$$Tu(x) = C_f(x) + \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} u(y) Q_f(dy, dt \mid x).$$

Mostraremos ahora que $Tu \in B_W(\mathbf{X})$. Para esto observemos que de la Hipótesis 3.2 a) y b)

$$\begin{aligned}
|Tu(x)| &\le |C(x,a)| + \left| \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} u(y) Q(dy, dt \mid x, a) \right| \\
&\le \bar{c} W(x) + \|u\|_W \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} W(y) Q(dy, dt \mid x, a) \\
&\le \bar{c} W(x) + \|u\|_W \beta W(x).
\end{aligned}$$

Por lo tanto,

$$\|Tu\|_W = \sup_{x\in\mathbf{X}} \frac{|Tu(x)|}{W(x)} \le \bar{c} + \|u\|_W \beta < \infty,$$

lo cual implica que $Tu \in B_W(\mathbf{X})$.

b) Para $u, v \in B_W(\mathbf{X})$ se cumple

$$\begin{aligned}
|Tu(x) - Tv(x)| &\le \left| \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} (u(y) - v(y)) Q(dy, dt \mid x, a) \right| \\
&\le \|u - v\|_W \left| \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} W(y) Q(dy, dt \mid x, a) \right| \\
&\le \beta \|u - v\|_W W(x), \quad x \in \mathbf{X}.
\end{aligned}$$

Por lo tanto T es un operador de contracción módulo β.

$\square$

Para $u \in B_W(\mathbf{X})$ y $f \in \mathbb{F}$ se define el operador

$$T_f u(x) := C_f(x) + \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} u(y) Q_f(dy, dt \mid x)$$

Proposición 3.6. *Bajo la Hipótesis 3.2 a) el operador $T_f, f \in \mathbb{F}$, es un operador de contracción módulo β en el espacio de Banach $(B_W(\mathbf{X}), \|\cdot\|_W)$ y su único punto fijo es $V(x,f)$, es decir ,*

$$V(x,f) = C_f(x) + \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} V(f,y) Q_f(dy, dt \mid x) = T_f V(x,f). \tag{3.4.2}$$

Demostración. Análogamente como con el operador T, se puede demostrar que T_f, es un operador de contracción del espacio Banach $(B_W(\mathbf{X}), \|\cdot\|_W)$ en sí mismo con módulo β. Veamos que $V(x, f)$ es su único punto fijo

$$
\begin{aligned}
V(x,f) &= E_x^f \left[C_f(x_0) + \sum_{n=1}^{\infty} e^{-\alpha T_n} C_f(x_n) \right] \\
&= C_f(x) + E_x^f \left[E_x^f \left[\sum_{n=1}^{\infty} e^{-\alpha\delta_1} e^{-\alpha T_{n-1}} C_f(x_n) \mid h_1 \right] \right] \\
&= C_f(x) + E_x^f \left[e^{-\alpha\delta_1} E_{x_1}^f \left[\sum_{n=0}^{\infty} e^{-\alpha T_n} C_f(x_n) \right] \right] \\
&= C_f(x) + E_x^f \left[e^{-\alpha\delta_1} V(x_1, f) \right] \\
&= C_f(x) + \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} V(y,f) Q_f(dy, dt \mid x) \\
&= T_f V(x,f).
\end{aligned}
$$

□

Una consecuencia de los resultados previos es la siguiente proposición, en la cual se asegura la existencia de políticas óptimas

Proposición 3.7. *Supongamos que se cumplen las Hipótesis 3.1 y 3.2. Entonces:*

a) *La función de valor óptimo V^* es el único punto fijo en $B_W(\mathbf{X})$ del operador T, esto es*

$$V^*(x) = TV^*(x) := \min_{a\in A(x)} \left\{ C(x,a) + \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} V^*(y) Q(dy, dt \mid x, a) \right\}. \tag{3.4.3}$$

b) *Una politica estacionaria $f \in \mathbb{F}$ es óptima si y sólo si alcanza el mínimo en (3.4.3), esto es*

$$V^*(\cdot) = T_f V^*(\cdot).$$

Demostración. **a)** De la Proposición 3.5 b) T es un operador de contracción, por lo que del Teorema del punto fijo de Banach existe un único punto fijo $u^* \in B_W(\mathbf{X})$. Además existe $f \in \mathbb{F}$ tal que

$$u^*(x) = C_f(x) + \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} u^*(y) Q_f(dy, dt \mid x) = T_f u^*(x).$$

Entonces, u^* es el punto fijo de T_f por lo que, de la Proposición 3.6 se tiene que

$$u^*(x) = V(x, f) \geq V^*(x). \tag{3.4.4}$$

Para obtener la desigualdad contraria, notemos que

$$u^*(x) = Tu^*(x) \leq C(x,a) + \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} u^*(y) Q(dy, dt \mid x, a).$$

Entonces

$$0 \leq C(x,a) + \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} u^*(y) Q(dy, dt \mid x, a) - u^*(x)$$

$$= E_x^\pi \left\{ C(x,a) + e^{-\alpha\delta_1} u^*(x_1) - u^*(x_0) \right\}.$$

Ahora, sea $\pi \in \Pi$ una política arbitraria. Para cada estado inicial $x \in \mathbf{X}$ y $m = 0, 1, \ldots$, se cumple que

$$0 \leq E_x^\pi \left\{ C(x_m, a_m) + e^{-\alpha\delta_{m+1}} u^*(x_{m+1}) - u^*(x_m) \mid h_m, a_m \right\}$$

de donde se sigue la desigualdad

$$0 \leq E_x^\pi \left\{ e^{-\alpha T_m} C(x_m, a_m) + e^{-\alpha T_{m+1}} u^*(x_{m+1}) - e^{-\alpha T_m} u^*(x_m) \mid h_m, a_m \right\}.$$

Tomando esperanza E_x^π y sumando para $m = 0, 1, \ldots, n$, obtenemos

$$0 \leq E_x^\pi \left\{ \sum_{m=0}^{n} e^{-\alpha T_m} C(x_m, a_m) + \sum_{m=0}^{n} e^{-\alpha T_{m+1}} u^*(x_{m+1}) - \sum_{m=0}^{n} e^{-\alpha T_m} u^*(x_m) \right\}.$$

Notemos que

$$\sum_{m=0}^{n} e^{-\alpha T_{m+1}} u^*(x_{m+1}) - \sum_{m=0}^{n} e^{-\alpha T_m} u^*(x_m) = e^{-\alpha T_{n+1}} u^*(x_{n+1}) - u^*(x_0),$$

lo cual implica

$$0 \leq E_x^\pi \left\{ \sum_{m=1}^{n} e^{-\alpha T_m} C(x_m, a_m) + e^{-\alpha T_{n+1}} u^*(x_{n+1}) - u^*(x_0) \right\}.$$

Luego por Corolario 3.3 tenemos que

$$u^*(x) \leq E_x^\pi \sum_{m=1}^{n} e^{-\alpha T_m} C(x_m, a_m) + W(x)\|u\|_W \beta^{n+1}.$$

Haciendo n tender a infinito

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

y como $\pi \in \Pi$ es arbitraria, se concluye

$$u^*(x) \leq V^*(x). \tag{3.4.5}$$

De las desigualdades (3.4.4) y (3.4.5) se sigue que

$$u^*(x) = V^*(x).$$

Por lo tanto, la función de valor óptimo V^* es el único punto fijo del operador T.

b) Ahora sea $f^* \in \mathbb{F}$ tal que
$$V^* = TV^* = T_{f^*}V^*.$$
Entonces por la Proposición (3.6)
$$V^*(x) = V(x, f^*)$$
por lo tanto f^* es una política óptima.
Supongamos ahora que f^* una política óptima, es decir
$$V^*(x) = V(x, f^*),$$
nuevamente de la Proposición 3.6 y la parte (a) concluimos que
$$V^*(x) = V\left(x, f^*\right) = T_{f^*}V^*(x) = TV^*(x) \quad \forall x \in \mathbf{X}.$$

□

3.5 Ejemplo: Un sistema de remplazo

Consideremos un sistema que está sujeto a una secuencia de choques, los cuales ocurren de manera aleatoria en el tiempo. Cada choque causa una cantidad aleatoria de daño que se acumula conforme pasa el tiempo y el sistema puede fallar por completo únicamente en el momento de un choque.

Definimos las siguientes variables

i) t_n, $n \in \mathbb{N}$: tiempo en el que ocurre el n-ésimo choque;

ii) $l_n := t_n - t_{n-1}$: tiempos entre choques consecutivos;

iii) z_n: magnitud aleatoria de daño en el tiempo t_n, para cada $n \in \mathbb{N}_0$

iv) x_n: daño acumulado hasta el tiempo t_n.

Se asume que la sucesión t_n, $n \in \mathbb{N}$ de tiempos en los que ocurren los choques es estrictamente creciente y que la probabilidad de falla del sistema es $1 - r(\cdot)$, donde $r : \mathbb{R}_+ \to [0, 1]$, es una función del daño no-creciente, conocida como función de supervivencia. Además la función de distribución condicional del daño z_n dado $x_n = x$ y la función de distribución de los tiempos entre choques consecutivos se denotarán por $J(\cdot \mid x)$ y $H(\cdot)$ respectivamente. Supondremos que estas distribuciones son continuas.

Proceso de remplazo del sistema
En el tiempo $t_n = t$ en el que ocurre un choque, la información disponible para el controlador es el daño $x_n = x$ acumulado hasta ese momento. Dicha información le permite elegir un tiempo de remplazo programado $a \in \mathbf{A} = A(x) = [\theta_1, \theta_2]$, donde θ_1 y θ_2 son constantes tales que $0 < \theta_1 < \theta_2$.

De lo anterior se presentan los siguientes escenarios:

1) El tiempo de remplazo programado es mayor al tiempo para que ocurra el siguiente choque, es decir $l_{n+1} = l < a$, y el sistema falla. Entonces el sistema es remplazado por falla con probabilidad $1 - r(x + z)$, donde $z = z_{n+1}$ es la magnitud del daño al tiempo t_{n+1}, por lo que el siguiente estado del sistema es

$$y = 0.$$

De este caso se genera un costo por falla K_1

2) El tiempo de remplazo programado es mayor al tiempo $l_{n+1} = l$, es decir $l_{n+1} = l < a$, y el sistema no falla con probabilidad $r(x+z)$. Entonces el siguiente estado del sistema es

$$y = x + z.$$

3) Si $l_{n+1} = l \geq a$, entonces el sistema se remplaza en el tiempo programado a, i.e., antes de la falla. En este caso se incurre en un costo por remplazo programado K_2 donde $K_2 < K_1$.

Se considera también un tercer costo K_3 por operación del sistema el cual es proporcional al daño acumulado x.
Observemos que el sistema se remplaza al tiempo que se presenta la falla o al tiempo programado, dependiendo cuál sea menor. De esta manera, sí definimos la variable aleatoria

$$\delta_n := \min\left(l_n, a_n\right).$$

Entonces, los tiempos de transición del sistema están dados por

$$T_0 = 0$$

,

$$T_n = T_{n-1} + \delta_n, \quad n \in \mathbb{N}.$$

El proceso de daño acumulado $\{x_n\}$ es un proceso semi-markoviano con espacios de estados $\mathbf{X} = [0, \infty)$ y espacio de controles $\mathbf{A} = A(x) = [\theta_1, \theta_2]$, con tiempos de permanencia $\{\delta_n\}$.

Para describir la dinámica del sistema consideramos los casos 1), 2) y 3) ya mencionados. Esto es

1) Si $l_{n+1} = l < a$, y ocurre una falla. Entonces el sistema es remplazado por falla con probabilidad $1 - r(x+z)$, donde $z = z_{n+1}$ es la magnitud del daño al tiempo t_{n+1}, donde el siguiente estado del sistema es $y = 0$. Supongamos que $t < a$, entonces para cada $B \in \mathcal{B}(\mathbf{X})$

$$Q(B, [0,t] \mid x, a) = Prob(x_{n+1} \in B, \delta_n \leq t \mid x, a)$$

$$= I_B(0) \int_0^t H(ds) \int_0^\infty (1 - r(x+z)) J(dz \mid x).$$

Si $a < t$. Entonces

$$Q(B, [0,t] \mid x, a) = I_B(0) \int_0^a H(ds) \int_0^\infty (1 - r(x+z)) J(dz \mid x).$$

Por lo tanto, para cada $B \in \mathcal{B}(\mathbf{X}), t \in \mathbb{R}^+$

$$Q(B, [0,t] \mid x, a) = I_B(0) \int_0^{\min(a,t)} H(ds) \int_0^\infty (1 - r(x+z)) J(dz \mid x).$$

2) Si $l_{n+1} = l < a$, y el sistema no falla con probabilidad $r(x+z)$, donde el siguiente estado del sistema es $y = x + z$. Entonces para cada $B \in \mathcal{B}(\mathbf{X}), t \in \mathbb{R}^+$

$$Q(B, [0,t] \mid x, a) = \int_0^{\min(a,t)} H(ds) \int_{(x+z)\in B} r(x+z) J(dz \mid x).$$

3) El sistema es reemplazado en el tiempo programado. Entonces

$$Q(B, [0,t] \mid x, a) = I_B(0) I_{(t>a)} \{1 - H(a)\}, \quad B \in \mathcal{B}(\mathbf{X}), t \in \mathbb{R}^+.$$

Por lo tanto la dinámica del sistema está descrita de acuerdo al kernel conjunto Q dado por

$$\begin{aligned} Q(B, [0,t] \mid x, a) = & \int_0^{\min(a,t)} H(ds) \int_{(x+z)\in B} r(x+z) J(dz \mid x) \\ & + I_B(0) \int_0^{\min(a,t)} H(ds) \int_0^{\infty} (1 - r(x+z)) J(dz \mid x) \\ & + I_B(0) I_{(t>a)} \{1 - H(a)\}, \quad B \in \mathcal{B}(\mathbf{X}), t \in \mathbb{R}^+. \end{aligned}$$

Calculemos la distribución marginal del tiempo. Para esto debemos considerar 2 casos:

Caso 1) Supongamos que $t < a$, entonces tenemos que

$$\begin{aligned} G(t \mid x, a) &= Q(\mathbf{X}, (0,t] \mid x, a) = Prob(x_{n+1} \in \mathbf{X}, \delta_{n+1} \le t \mid x_n = x, a_n = a) \\ &= \int_0^t H(ds) \int_0^{\infty} r(x+z) J(dz \mid x) + \int_0^t H(ds) \int_0^{\infty} (1 - r(x+z)) J(dz \mid x) \\ &= \int_0^t H(ds) \left[\int_0^{\infty} J(dz \mid x) \right] = \int_0^t H(ds) = H(t). \end{aligned}$$

Caso 2) Supongamos que $t > a$, entonces tenemos que

$$\begin{aligned} G(t \mid x, a) &= \int_0^a H(ds) \int_0^{\infty} r(x+z) J(dz \mid x) + \int_0^a H(ds) \int_0^{\infty} [1 - r(x+z)] J(dz \mid x) + 1 - H(a) \\ &= \int_0^a H(ds) \int_0^{\infty} J(dz \mid x) + 1 - H(a) = 1 \end{aligned}$$

De lo anterior se sigue que la distribución marginal en el tiempo es

$$G(t \mid x, a) = Q(\mathbf{X}, (0,t] \mid x, a) = \begin{cases} H(t), & t < a \\ 1 & a \le t \end{cases}$$

Calculemos ahora la distribución marginal de la variable de estados:

$$
\begin{aligned}
P(B \mid x,a) &= Q\left(B, \mathbb{R}^{+} \mid x,a\right) \\
&= H(a)\int_{(x+z)\in B} r(x+z)J(dz \mid x) \\
&+ I_B(0)H(a)\int_0^{\infty}(1-r(x+z))J(dz \mid x) \\
&+ I_B(0)\{1-H(a)\}.
\end{aligned}
$$

Definamos la siguiente función

$$
\widehat{c}(x,a,\delta,y) := \begin{cases} K_1 + K_3x & \text{si} \quad l<a, y\equiv 0 \\ K_2 + K_3x & \text{si} \quad a<l, y\equiv 0 \\ K_3x & \text{si} \quad l<a, y>0. \end{cases}
$$

Por lo tanto, la función de costo en cada época de decisión está dada por

$$
C(x,a) = \int \widehat{c}(x,a,\delta,y)Q(dy,d\delta \mid x,a).
$$

Ahora consideremos de nuevo los casos 1), 2) y 3):

1) Si $l<a$, $y>0$

$$
C(x,a) = K_3x\int_0^a H(ds)\int_0^{\infty} r(x+z)J(dz \mid x).
$$

2) Si $l<a$, $y=0$

$$
C(x,a) = (K_1+K_3x)H(a)\left[\int_0^{\infty}(1-r(x+z))J(dz \mid x)\right].
$$

3) Si $l>a$, $y=0$

$$
C(x,a) = (K_2+K_3x)[1-H(a)].
$$

A partir de lo anterior, la función de costo en cada época de decisión toma la forma:

$$
C(x,a) = K_1H(a)\left\{\int_0^{\infty}(1-r(x+z))J(dz \mid x)\right\} + K_2(1-H(a)) + K_3x.
$$

Una vez establecidos todos los elementos del modelo de control semi-markoviano, verificaremos que bajo ciertas condiciones se satisfacen las Hipótesis 3.1 y 3.2.

Notemos que para cada $u \in \mathbf{M}_b(\mathbf{X})$ se tiene que

$$
\int_{\mathbf{X}\times\mathbb{R}^{+}} e^{-\alpha t}u(y)Q(dy,dt \mid x,a) = \int_0^a e^{-\alpha t}H(dt)\int_0^{\infty} u(x+z)r(x+z)J(dz \mid x)
$$

$$+ u(0) \int_0^a e^{-\alpha t} H(dt) \int_0^\infty (1 - r(x+z)) J(dz \mid x)$$

$$+ u(0) \int_a^\infty e^{-\alpha t} H(dt). \tag{3.5.1}$$

Entonces, por la continuidad de H, las funciones

$$a \to C(x,a) \quad \text{y} \quad a \to \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} u(y) Q(dy, dt \mid x, a)$$

son continuas. Claramente el conjunto $A(x) = [\theta_1, \theta_2]$ es compacto para cada $x \in \mathbf{X}$. Por lo tanto se cumple la Hipótesis 3.1.
Para asegurarnos que se cumplan las condiciones de la Hipótesis 3.2, introducimos la siguiente Hipótesis.

Hipótesis 3.3. *Existen constantes $\alpha_0 > 0$ y $q > 0$ tales que*

$$\sup_{a\in[\theta_1,\theta_2]} \left[\int_0^a e^{-\alpha_0 t} H(dt) \int_0^\infty (e^{qz} r(z) + (1 - r(z)) J(dz) + e^{-\alpha_0 a}(1 - H(a)) \right] < 1$$

Proposición 3.8. *Supongamos que se cumple la Hipótesis 3.3. Entonces el sistema de remplazo satisface las condiciones de las Hipótesis 3.2*

Demostración. Para demostrar que se tiene la Hipótesis 3.2 a), consideremos la función $W(x) = e^{qx}$. Notemos que

$$|C(x,a)| \leq |K_1 H(a) \left\{ \int_0^\infty (1 - r(x+z)) J(dz \mid x) \right\}| + |K_2(1 - H(a))| + |K_3 x|$$

$$\leq K_1 + K_2 + K_3 x.$$

Entonces, eligiendo una constante adecuada $\bar{c} \in \mathbb{R}$, tal que

$$K_1 + K_2 + K_3 x \leq \bar{c} e^{qx}$$

se cumple la Hipótesis 3.2 a).

Ahora observemos que de (3.5.1) y considerando $W(x) = e^{qx}$,

$$\begin{aligned} \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} W(y) Q(dy, dt \mid x, a) &= \int_0^a e^{-\alpha t} H(dt) \int_0^\infty e^{q(x+z)} r(x+z) J(dz) \\ &+ \int_0^a e^{-\alpha t} H(dt) \int_0^\infty (1 - r(x+z)) J(dz) + \int_a^\infty e^{-\alpha t} H(dt) \\ &\leq e^{qx} \beta(\alpha, x, a) \end{aligned}$$

donde

$$\beta(\alpha, x, a) := \int_0^a e^{-\alpha t} H(dt) \left[\int_0^\infty \left(e^{qz} r(x+z) + e^{-qx}(1 - r(x+z))\right) J(dz)\right]. \\ + \int_a^\infty e^{-\alpha t} H(dt).$$

Por lo tanto

$$\int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} W(y) Q(dy, dt \mid x, a) \le e^{qx} \beta(\alpha, x, a). \tag{3.5.2}$$

Recordemos que la función de supervivencia $r : \mathbb{R}^+ \to [0, 1]$ es una función no creciente, y supongamos que esta es una función derivable. Consideremos la función

$$I(x) := e^{qz} r(x+z) + e^{-qx}(1 - r(x+z)).$$

Entonces

$$I'(x) = r'(x+z)\left(e^{qz} - e^{-qx}\right) - qe^{-qx}(1 - r(x+z)).$$

Observemos que $r'(x+z) < 0$ y $e^{qz} - e^{-qz} > 0$, entonces $I'(x) < 0$, lo cual implica que $I(x)$ es no creciente. De aqui, el máximo valor de $I(x)$ se alcanza en $x = 0$, esto es

$$I(0) = 1 + r(z)\left(e^{qz} - 1\right) \ge 1,$$

y por lo tanto

$$\begin{aligned}
\beta(\alpha, x, a) &= \int_0^a e^{-\alpha t} H(dt) \left[\int_0^\infty (e^{qz} r(x+z) + e^{-qx}(1 - r(x+z))) J(dz)\right] + \int_a^\infty e^{-\alpha t} \\
&\le \int_0^a e^{-\alpha t} H(dt) \left[\int_0^\infty \left(e^{qz} r(z) + (1 - r(z))\right) J(dz)\right] + \int_a^\infty e^{-\alpha t} H(dt) \\
&= \beta(\alpha, 0, a).
\end{aligned}$$

Ahora, de la Hipótesis 3.3, existe $\alpha_0 > 0$ tal que

$$\sup_{a\in[\theta_1,\theta_2]} \beta\left(\alpha_0, x, a\right) \le \sup_{a\in[\theta_1,\theta_2]} \beta\left(\alpha_0, 0, a\right) < 1.$$

Notemos que si $\alpha_0 < \alpha$ entonces $\beta\left(\alpha_0, x, a\right) \ge \beta(\alpha, x, a)$. De esta manera, para $\alpha > \alpha_0$, y de (3.5.2) se cumple

$$\begin{aligned}
\int_{\mathbf{X}\times\mathbb{R}_+} e^{-\alpha t} W(y) Q(dy, dt \mid x, a) &\le e^{qx} \beta(\alpha, x, a) \\
&\le e^{qx} \beta(\alpha, 0, a) \le e^{qx} \beta\left(\alpha_0, 0, a\right) \\
&\le e^{qx} \sup_{a\in[\theta_1,\theta_2]} \beta\left(\alpha_0, 0, a\right) \le \beta W(x)
\end{aligned}$$

donde $\beta := \sup_{a\in[\theta_1,\theta_2]} \beta\left(\alpha_0, 0, a\right) < 1$. Por lo tanto se cumple la Hipótesis 3.2 b).

Ahora, considerando (3.5.1), la función

$$a \to \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} W(y) Q(dy, dt \mid x, a) = \int_0^a e^{-\alpha t} H(dt) \int_0^\infty u(x+z) r(x+z) J(dz \mid x)$$

$$+ W(0) \int_0^a e^{-\alpha t} H(dt) \int_0^\infty (1 - r(x+z)) J(dz \mid x)$$

$$+ W(0) \int_a^\infty e^{-\alpha t} H(dt)$$

es continua en $A(x)$.

□

3.6 Modelo Semi-Markoviano truncado

En esta sección realizaremos un procedimiento de truncamiento del espacio de estados y a partir de esto definiremos un modelo de control restringido a un conjunto compacto. El objetivo es plantear y dar solución al problema de control óptimo dentro de este conjunto y hacer una aproximación a la función de valor óptimo V^*, esto mediante la solución encontrada para el espacio truncado. Esto se puede ver en la Proposición 3.16. Para tal fin introducimos primero algunos conceptos necesarios.

Definición 3.3. *Sea (S, B_s) un espacio medible arbitrario .*

a) *Una familia Γ de medidas finitas definidas en S es tensa si para cada $\epsilon > 0$ existe un conjunto compacto O tal que $\gamma(O^c) < \epsilon$ para toda $\gamma \in \Gamma$*

b) *Una función $\varphi : S \to \mathbb{R}$ es un momento en S, si existe una sucesión creciente de conjuntos compactos $S_k \uparrow S$ tal que*

$$\lim_{k\to\infty} \inf_{x \notin S_k} \varphi(x) = \infty$$

El siguiente resultado establece una relación entre estos dos conceptos.

Proposición 3.9. *Una familia Γ de medidas finitas definidas en S es tensa si existe un momento $\varphi \geq 1$ tal que*

$$\sup_{\gamma\in\Gamma} \int_S \varphi(y)\gamma(dy) < \infty. \tag{3.6.1}$$

Demostración. Ver [[10], Teorema 12.2.15 pág. 216]. □

Consideremos el modelo de control semi-markoviano (3.2.1)

$$(\mathbf{X}, \mathbf{A}, \{A(x) : x \in \mathbf{X}\}, Q, C)$$

que satisface las condiciones de las Hipótesis 3.1 y 3.2. Supongamos que existe una sucesión creciente de conjuntos compactos $\mathbf{X}_k \uparrow \mathbf{X}$. Por la Hipótesis 3.2 b) tenemos que la sucesión

$$\frac{1}{W(x)} \int_{\mathbf{X}\times\mathbb{R}^+} e^{-at} W(y) I_{\mathbf{X}_k}(y) Q(dy, dt \mid x, a)$$

converge crecientemente a

$$\frac{1}{W(x)}\int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t}W(y)Q(dy,dt\mid x,a).$$

Observemos que

$$\frac{1}{W(x)}\int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t}W(y)Q(dy,dt\mid x,a) = \frac{1}{W(x)}\int_{\mathbf{X}\times\mathbb{R}^+} e^{-at}W(y)I_{\mathbf{X}_k}(y)Q(dy,dt\mid x,a)$$
$$+\frac{1}{W(x)}\int_{\mathbf{X}\times\mathbb{R}^+} e^{-at}W(y)I_{\mathbf{X}_k^c}(y)Q(dy,dt\mid x,a).$$

En consecuencia tenemos que

$$\frac{1}{W(x)}\int_{\mathbf{X}\times\mathbb{R}^+} e^{-at}W(y)I_{\mathbf{X}_k^c}(y)Q(dy,dt\mid x,a)\downarrow 0.$$

De aquí, para cada $\epsilon>0$ existe una colección de subconjuntos compactos $\mathbf{X}_\epsilon(x,a)$ $(x,a)\in\mathbb{K}$ de $\mathbf{X}$ tales que

$$\frac{1}{W(x)}\int_{\mathbf{X}_\epsilon^c(x,a)\times\mathbb{R}^+} e^{-\alpha t}W(y)Q(dy,dt\mid x,a)\leq\epsilon \tag{3.6.2}$$

donde $\mathbf{X}_\epsilon^c(x,a)=\mathbf{X}\backslash\mathbf{X}_\epsilon(x,a)$.

Para obtener un truncamiento uniforme del espacio de estados es necesario que el conjunto compacto de la desigualdad 3.6.2 sea el mismo para toda $(x,a)\in\mathbb{K}$.

Para garantizar la existencia de dicho conjunto compacto se establece la siguiente condición.

Hipótesis 3.4. ***a)*** *Existe $\eta>0$ tal que W^η es un momento en $\mathbf{X}$ donde W es la función medible de la Hipótesis 3.2;*

b) *además*

$$\sup_{(x,a)\in\mathbb{K}}\frac{1}{W(x)}\int_{\mathbf{X}\times\mathbb{R}^+} e^{-at}W^{(1+\eta)}(y)Q(dy,dt\mid x,a)<\infty. \tag{3.6.3}$$

El siguiente resultado establece que el truncamiento del espacio de estados es uniforme.

Lema 3.10. *Supongamos que se cumplen las Hipótesis 3.2 b) y 3.4. Entonces para cada $\epsilon>0$ existe un conjunto compacto $\mathbf{X}_\epsilon$ tal que*

$$\int_{\mathbf{X}_\epsilon^c\times\mathbb{R}^+} e^{-\alpha t}W(y)Q(dy,dt\mid x,a)\leq\epsilon W(x)\quad\forall(x,a)\in\mathbb{K}$$

Demostración. Para $B\in\mathbf{X},(x,a)\in\mathbb{K}$ se define la medida

$$\gamma_{(x,a)}(B):=\frac{1}{W(x)}\int_{\mathbf{X}\times\mathbb{R}^+} e^{-at}W(y)I_B(y)Q(dy,dt\mid x,a), B\in\mathcal{B}(\mathbf{X})$$

Demostraremos que $\{\gamma_{(x,a)}(\cdot)\}$ es una familia de medidas tensa.

De de la Hipótesis 3.2 b), para cada $B \in \mathbf{X}$, $(x,a) \in \mathbb{K}$ tenemos que

$$\gamma_{(x,a)}(B) = \frac{1}{W(x)} \int_{\mathbf{X}\times\mathbb{R}^+} e^{-at} W(y) I_B(y) Q(dy, dt \mid x, a)$$

$$\leq \frac{1}{W(x)} \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} W(y) Q(dy, dt \mid x, a) \leq \beta;$$

por lo tanto, estas medidas $\gamma_{(x,a)}(\cdot)$ son finitas y además están uniformemente acotadas.

Por otro lado, consideremos $\varphi(\cdot) = W^\eta(\cdot)$. Entonces de la Hipótesis 3.4 tenemos que

$$\sup_{\gamma\in\Gamma} \int_S \varphi(y) \gamma_{(x,a)}(dy) = \sup_{(x,a)\in\mathbb{K}} \int_{\mathbf{X}\times\mathbb{R}^+} W^\eta(y) \gamma_{(x,a)}(dy)$$

$$= \sup_{(x,a)\in\mathbb{R}} \frac{1}{W(x)} \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} W^{(1+\eta)}(y) Q(dy, dt \mid x, a) < \infty.$$

De la Observación 3.9 se sigue que la familia de medidas $\{\gamma_{(x,a)}(\cdot)\}$, $(x,a) \in \mathbb{K}$ es tensa. Entonces, por la definición 3.3, para cada $\epsilon > 0$ existe un subconjunto compacto $\mathbf{X}_\epsilon \subset \mathbf{X}$ tal que

$$\gamma_{(x,a)}(\mathbf{X}_\epsilon^c) = \frac{1}{W(x)} \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} W(y) I_{\mathbf{X}_\epsilon^c} Q(dy, dt \mid x, a) \leq \epsilon \quad \forall (x,a) \in \mathbb{K}$$

lo cual es equivalente a

$$\int_{\mathbf{X}_\epsilon^c\times\mathbb{R}^+} e^{-\alpha t} W(y) Q(dy, dt \mid x, a) \leq \epsilon W(x) \quad \forall (x,a) \in \mathbb{K} \tag{3.6.4}$$

donde $\mathbf{X}_\epsilon^c = \mathbf{X} \backslash \mathbf{X}_\epsilon$. □

Como ya se mencionó, el propósito de esta sección es estudiar el problema de control óptimo hasta que el proceso salga del espacio "truncado" $\mathbf{X}_\epsilon$, por lo que es necesario definir el criterio de óptimalidad descontado restringido al espacio $\mathbf{X}_\epsilon$. Para tal fin definimos primero la siguiente variable

$$\tau = \tau(\epsilon) := \inf \{n \geq 1 : x_n \in \mathbf{X}_\epsilon^c\}$$

donde para el conjunto vacío $\emptyset$ se tiene que $\inf \emptyset := +\infty$.

Definición 3.4. *Para $x \in \mathbf{X}_\epsilon, \pi \in \Pi$, se define el costo total esperado α-descontado*

$$V_\epsilon(x, \pi) := E_x^\pi \left[\sum_{n=0}^\infty e^{-\alpha T_n} C(x_n, a_n) I_{[n<\tau]} \right]$$

y la función de valor óptimo

$$V_\epsilon^*(x) = \inf_{\pi\in\Pi} V_\epsilon(x, \pi).$$

De manera que el problema de control óptimo restringido al espacio $\mathbf{X}_\epsilon$ consiste en encontrar una política $\pi^* \in \Pi$ tal que

$$V_\epsilon^*(x) := V_\epsilon(x, \pi^*), x \in \mathbf{X}_\epsilon$$

Proposición 3.11. *Supongamos que se cumplen las Hipótesis 3.1, 3.2 y 3.4. Entonces para cada política $\pi \in \Pi$ y estado inicial $x \in \mathbf{X}_\epsilon$*

$$\left| E_x^\pi \sum_{n=0}^{\infty} e^{-\alpha T_n} C\left(x_n, a_n\right) I_{[n \geq \tau]} \right| \leq \frac{\bar{c}\epsilon}{(1-\beta)^2} W(x).$$

Demostración. Para cada $n \in \mathbb{N}$ y $k < n$, usando propiedades de la esperanza condicional, de la Hipótesis 3.2 b) se tiene que

$$E_x^\pi e^{-\alpha T_n} W\left(x_n\right) I_{[\tau=k]} = E_x^\pi \left[E_x^\pi \left(e^{-\alpha T_n} W\left(x_n\right) I_{[\tau=k]} \mid h_{n-1}, a_{n-1}\right)\right]$$

$$= E_x^\pi \left[e^{-\alpha T_{n-1}} I_{[\tau=k]} E_x^\pi \left(e^{-\alpha \delta_n} W\left(x_n\right) \mid h_{n-1}, a_{n-1}\right)\right]$$

$$= E_x^\pi \left[e^{-\alpha T_{n-1}} I_{[\tau=k]} \int_{\mathbf{A}} \int_{\mathbf{X} \times \mathbb{R}_+} e^{-\alpha t} W(y) Q\left(dy, dt \mid x_{n-1}, a\right) \pi\left(da \mid h_{n-1}\right)\right]$$

$$\leq E_x^\pi \left[e^{-\alpha T_{n-1}} I_{[\tau=k]} \int_{\mathbf{A}} \beta W\left(x_{n-1}\right) \pi\left(da \mid h_{n-1}\right)\right]$$

$$\leq \beta E_x^\pi \left[e^{-\alpha T_{n-1}} I_{[\tau=k]} W\left(x_{n-1}\right)\right].$$

Por lo tanto

$$E_x^\pi e^{-\alpha T_n} W\left(x_n\right) I_{[\tau=k]} \leq \beta E_x^\pi \left[e^{-\alpha T_{n-1}} I_{[\tau=k]} W\left(x_{n-1}\right)\right].$$

Iterando esta desigualdad obtenemos

$$E_x^\pi e^{-\alpha T_n} W\left(x_n\right) I_{[\tau=k]} \leq \beta^{n-k} E_x^\pi \left[e^{-\alpha T_k} I_{[\tau=k]} W\left(x_k\right)\right] \tag{3.6.5}$$

Notemos que

$$E_x^\pi \left[e^{-\alpha T_k} I_{[\tau=k]} W\left(x_k\right)\right] = E_x^\pi \left[E_x^\pi \left(e^{-\alpha T_k} W\left(x_k\right) I_{[\tau=k]} \mid h_{k-1}, a_{k-1}\right)\right]$$

$$= E_x^\pi \left[e^{-\alpha T_{k-1}} I_{[x_1 \in \mathbf{X}_\epsilon, \ldots, x_{k-1} \in \mathbf{X}_\epsilon]} E_x^\pi \left(e^{-\alpha \delta_k} W\left(x_k\right) I_{[x_k \in \mathbf{X}_\epsilon^c]} \mid h_{k-1}, a_{k-1}\right)\right]$$

$$= E_x^\pi \left[e^{-\alpha T_{k-1}} I_{[x_1 \in \mathbf{X}_\epsilon, ., x_{k-1} \in \mathbf{X}_\epsilon]} \int_{\mathbf{X}_\epsilon^c \times \mathbb{R}^+} e^{-\alpha t} W(y) Q(dy, dt \mid x, a)\right]$$

$$\leq \epsilon E_x^\pi \left[e^{-\alpha T_{k-1}} I_{[x_1 \in \mathbf{X}_\epsilon, ., x_{k-1} \in \mathbf{X}_\epsilon]} W\left(x_{k-1}\right)\right],$$

donde esta última desigualdad se tiene del Lema 3.10.

Observemos ahora que de la Proposición 3.1

$$E_x^\pi \left[e^{-\alpha T_{k-1}} I_{[x_1 \in \mathbf{X}_\epsilon, \ldots, x_{k-1} \in \mathbf{X}_\epsilon]} W\left(x_{k-1}\right)\right] \leq E_x^\pi \left[e^{-\alpha T_{k-1}} W\left(x_{k-1}\right)\right] \leq \beta^{k-1} W(x). \tag{3.6.6}$$

Por lo tanto, de las desigualdades (3.6.5) y (3.6.6), se sigue

$$E_x^\pi e^{-\alpha T_n} W\left(x_n\right) I_{[\tau=k]} \leq \epsilon \beta^{n-1} W(x),$$

lo cuál a su vez implica

$$\begin{aligned} E_x^\pi e^{-\alpha T_n} W\left(x_n\right) I_{[\tau \leq n]} &= E_x^\pi \sum_{k=1}^{n} e^{-\alpha T_n} W\left(x_n\right) I_{[\tau=k]} \\ &\leq n \epsilon \beta^{n-1} W(x). \end{aligned}$$

Por último, de esta desigualdad y por la Hipótesis 3.2 a) concluimos

$$\begin{aligned} E_x^\pi \sum_{n=0}^{\infty} e^{-\alpha T_n} C\left(x_n, a_n\right) I_{[\tau \leq n]} &\leq \bar{c} E_x^\pi \sum_{n=0}^{\infty} e^{-\alpha T_n} W\left(x_n\right) I_{[\tau \leq n]} \\ &\leq \bar{c} \epsilon W(x) \sum_{n=0}^{\infty} n \beta^{n-1} \leq \frac{\bar{c} W(x)}{(1-\beta)^2} \epsilon \end{aligned}$$

□

Consideremos el espacio $B_W(\mathbf{X}_\epsilon)$ de las funciones medibles $u : \mathbf{X}_\epsilon \to \mathbb{R}$ tales que

$$\|u\|_W^\epsilon := \sup_{x \in \mathbf{X}_\epsilon} \frac{|u(x)|}{|W(x)|} < \infty$$

Para el estudio del problema de control óptimo correspondiente, definimos para cada $f \in \mathbb{F}$, $x \in \mathbf{X}_\epsilon$ $u \in B_W(\mathbf{X}_\epsilon)$ los operadores de la programación dinámica "truncados" como

$$T_{\epsilon,f} u(x) := C(x, f) + \int_{\mathbf{X}_\epsilon \times \mathbb{R}_+} e^{-\alpha t} u(y) Q(dy, dt \mid x, f) \tag{3.6.7}$$

$$T_\epsilon u(x) := \min_{a \in A(x)} \left[C(x, a) + \int_{\mathbf{X}_\epsilon \times \mathbb{R}_+} e^{-\alpha t} u(y) Q(dy, dt \mid x, a) \right]. \tag{3.6.8}$$

Observación 3.12. *Notemos que de la Hipótesis 3.1 b) y c), para cada $u \in \mathbf{M}_b(\mathbf{X}), x \in \mathbf{X}_\epsilon$, la función*

$$\begin{aligned} a \to\ & C(x, a) + \int_{\mathbf{X}_\epsilon \times \mathbb{R}^+} e^{-\alpha t} u(y) Q(dy, dt \mid x, a) \\ &= C(x, a) + \int_{\mathbf{X} \times \mathbb{R}^+} e^{-\alpha t} u(y) I_{\mathbf{X}_\epsilon}(y) Q(dy, dt \mid x, a) \end{aligned}$$

es continua.

Proposición 3.13. *Supomgamos que se cumplen las Hipótesis 3.1 y 3.2. Entonces:*

a) *El operador de la programación dinámica T_ϵ es un operador de contracción modulo β en el espacio de Banach $(B_W(\mathbf{X}_\epsilon), \|\cdot\|_W^\epsilon)$ y por lo tanto existe una función u^ϵ tal que*

$$T_\epsilon u^\epsilon = u^\epsilon.$$

Además, para cada función $u \in B_W(\mathbf{X}_\epsilon)$ existe $f \in \mathbb{F}$ tal que

$$T_\epsilon u(x) = T_{\epsilon,f} u(x) \quad \forall x \in \mathbf{X}_\epsilon.$$

b) *$T_{\epsilon,f}$ es un operador de contracción módulo β en el espacio de Banach $(B_W(\mathbf{X}_\epsilon), \|\cdot\|_W^\epsilon)$. Entonces, existe una función $u_f^\epsilon \in B_W(\mathbf{X}_\epsilon)$ tal que*

$$T_{\epsilon,f} u_f^\epsilon = u_f^\epsilon.$$

Demostración. **a)** De las Hipótesis 3.1 c) para cada $u \in \mathbf{M}_b(\mathbf{X}), x \in \mathbf{X}_\epsilon$ la función

$$a \to \int_{\mathbf{X}_\epsilon \times \mathbb{R}^+} e^{-\alpha t} u(y) Q(dy, dt \mid x, a)$$

es continua. Por otro lado de la Hipótesis 3.2 la función

$$a \to \int_{\mathbf{X}_\epsilon \times \mathbb{R}^+} e^{-\alpha t} W(y) Q(dy, dt \mid x, a)$$

es continua en $A(x)$. Por lo que de [10, Lema 8.3.7] se sigue que para cada $u \in B_W(\mathbf{X}_\epsilon)$, la función

$$a \to C(x,a) + \int_{\mathbf{X}_\epsilon \times \mathbb{R}^+} e^{-\alpha t} u(y) Q(dy, dt \mid x, a)$$

es continua en $A(x)$ para cada $x \in \mathbf{X}_\epsilon$. Entonces por un teorema de selección medible ([10] Lema 8.3.8) $Tu(x)$ es una función medible y existe $f \in \mathbb{F}$ tal que

$$Tu(x) = C_f(x) + \int_{\mathbf{X}_\epsilon \times \mathbb{R}^+} e^{-\alpha t} u(y) Q_f(dy, dt \mid x) = T_{\epsilon,f} u(x) \quad \forall x \in \mathbf{X}_\epsilon.$$

La demostración de que T_ϵ es un operador de contracción modulo β se hace de manera análoga a la demostración de la Proposición 3.5 b). Por lo tanto del Teorema del punto fijo de Banach existe una función u^ϵ tal que

$$T_\epsilon u^\epsilon = u^\epsilon.$$

b) De manera análoga se demuestra que el operador $T_{\epsilon,f}$ es un operador de contracción módulo β en el espacio de Banach $(B_W(\mathbf{X}_\epsilon), \|\cdot\|_W^\epsilon)$. Entonces, existe una función $u_f^\epsilon \in B_W(\mathbf{X}_\epsilon)$ tal que

$$T_{\epsilon,f} u_f^\epsilon = u_f^\epsilon.$$

□

Proposición 3.14. *Supongamos que se cumplen las Hipótesis 3.1, 3.2 y 3.4. Entonces*

a) $T_{\epsilon,f}V_\epsilon(\cdot,f) = V_\epsilon(\cdot,f)$ *para toda* $f \in \mathbb{F}$.

b) $V_\epsilon(\cdot,f) = V_\epsilon^*(\cdot)$ *si y solo si* $V_\epsilon^*(\cdot) = T_{\epsilon,f}V_\epsilon^*(\cdot)$.

Demostración. **a)** Sea $u \in B_W(\mathbf{X}_\epsilon)$. Entonces para cada $x \in \mathbf{X}_\epsilon$, $f \in \mathbb{F}$, tenemos que

$$\begin{aligned}
T^2_{\epsilon,f}u(x) &:= T_{\epsilon,f}(T_{\epsilon,f}u(x)) \\
&= C(x,f) + \int_{\mathbf{X}_\epsilon\times\mathbb{R}^+} e^{-\alpha s}\left(C(z,f) + \int_{\mathbf{X}_\epsilon\times\mathbb{R}^+} e^{-\alpha t}u(y)Q(dy,dt\mid z,f)\right)Q(dz,ds\mid x,f) \\
&= C(x,f) + \int_{\mathbf{X}_\epsilon\times\mathbb{R}^+} e^{-\alpha s}C(z,f)Q(dz,ds\mid x,f) \\
&+ \int_{\mathbf{X}_\epsilon\times\mathbb{R}^+} e^{-\alpha s}\int_{\mathbf{X}_\epsilon\times\mathbb{R}^+} e^{-\alpha t}u(y)Q(dy,dt\mid z,a)Q(dz,ds\mid x,f) \\
&= C(x,f) + \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha s}C(z,f)I_{\mathbf{X}_\epsilon}(dz)Q(dz,ds\mid x,f) \\
&+ \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha s}I_{\mathbf{X}_\epsilon}(dz)\left(\int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t}u(y)I_{\mathbf{X}_\epsilon}(dy)Q(dy,dt\mid z,a)\right)Q(dz,ds\mid x,f) \\
&= C(x,f) + \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha s}C(z,f)I_{[x_1\in\mathbf{X}_\epsilon]}Q(dz,ds\mid x,f) \\
&+ \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha s}I_{[x_1\in\mathbf{X}_\epsilon]}\left(\int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t}u(y)I_{[x_1\in\mathbf{X}_\epsilon,x_2\in\mathbf{X}_\epsilon]}Q(dy,dt\mid z,a)\right)Q(dz,ds\mid x,f) \\
&= C(x,f) + E_x^f\left[e^{-\alpha T_1}C_f(x_1)I_{[1<\tau]}\right] + E_x^f\left[e^{-\alpha T_2}u(x_2)I_{[2<\tau]}\right] \\
&= E_x^f\sum_{k=0}^{1} e^{-\alpha T_k}C_f(x_k)I_{[k<\tau]} + E_x^f\left[e^{-\alpha T_2}u(x_2)I_{[2<\tau]}\right].
\end{aligned}$$

Notemos que iterando esta igualdad obtenemos que para todo $x \in \mathbf{X}_\epsilon$, $f \in \mathbb{F}$, $u \in B_W(\mathbf{X}_\epsilon)$, y $n \in \mathbb{N}$

$$T^n_{\epsilon,f}u(x) := T_{\epsilon,f}(T^{n-1}_{\epsilon,f}u(x)) = E_x^f\sum_{k=0}^{n-1} e^{-\alpha T_k}C_f(x_k)I_{[k<\tau]} + E_x^f\left[e^{-\alpha T_n}u(x_n)I_{[n<\tau]}\right]. \quad (3.6.9)$$

Tomando límite en (3.6.9) tenemos que

$$\lim_{n\to\infty} T^n_{\epsilon,f}u(x) = E_x^f\sum_{k=0}^{\infty} e^{-\alpha T_k}C(x_k,a_k)I_{[k<\tau]} + \lim_{n\to\infty}\left|E_x^f\left[e^{-\alpha T_n}u(x_n)I_{[n<\tau]}\right]\right|.$$

Notemos que de la Proposición 3.3, si $n \to \infty$

$$\left|E_x^f\left[e^{-\alpha T_n}u(x_n)I_{[n<\tau]}\right]\right| \le \left|E_x^f\left[e^{-\alpha T_n}u(x_n)\right]\right| \le W(x)\|u\|_W\beta^n \to 0. \quad (3.6.10)$$

Entonces de (3.6.9) y (3.6.10), para cada $x \in \mathbf{X}_\epsilon$, tenemos que

$$\lim_{n\to\infty} T^n_{\epsilon,f}u(x) = E_x^f\sum_{k=0}^{\infty} e^{-\alpha T_k}C(x_k,a_k)I_{[k<\tau]} = V_\epsilon(x,f).$$

Por otra parte, de la Proposición 3.13 sabemos que $u_f^\epsilon(x)$ es punto fijo del operador $T_{\epsilon,f}$, esto es

$$T_{\epsilon,f}u_f^\epsilon(x) = u_f^\epsilon(x).$$

Entonces del Teorema del punto fijo de Banach tenemos que para cada $x \in \mathbf{X}_\epsilon$

$$u_f^\epsilon(x) = \lim_{n\to\infty} T_{\epsilon,f}^n u(x).$$

Por lo tanto, de la unicidad del punto fijo del operador $T_{\epsilon,f}$ tenemos que para toda $f \in \mathbb{F}$

$$T_{\epsilon,f}V_\epsilon(\cdot, f) = V_\epsilon(\cdot, f).$$

b) Supongamos que $V_\epsilon(\cdot, f) = V_\epsilon^*(\cdot)$ entonces del inciso a) y por la unicidad del punto fijo del operador $T_{\epsilon,f}$ se sigue que para toda $f \in \mathbb{F}$

$$T_{\epsilon,f}V_\epsilon(\cdot, f) = V_\epsilon(\cdot, f) = V_\epsilon^*(\cdot) = T_{\epsilon,f}V_\epsilon^*(\cdot)$$

por otra parte, si $V_\epsilon^*(\cdot) = T_{\epsilon,f}V_\epsilon^*(\cdot)$, de nuevo por el inciso a) se sigue que

$$V_\epsilon(\cdot, f) = V_\epsilon^*(\cdot).$$

$\square$

El resultado que se enuncia a continuación da solución al problema de control óptimo restringido al espacio $\mathbf{X}_\epsilon$.

Proposición 3.15. *Supongamos que se cumplen las Hipótesis 3.1, 3.2 y 3.4. Entonces*

a) *La función de valor óptimo V_ϵ^* es el único punto fijo en $B_W(\mathbf{X}_\epsilon)$ del operador T_ϵ, es decir,*

$$V_\epsilon^*(x) = T_\epsilon V_\epsilon^*(x) = \min_{a\in A(x)} \left[C(x,a) + \int_{\mathbf{X}_\epsilon\times\mathbb{R}^+} e^{-\alpha t} V_\epsilon^*(y) Q(dy, dt \mid x, a) \right]. \qquad (3.6.11)$$

b) *Existe $f_\epsilon \in \mathbb{F}$ tal que*

$$V_\epsilon^*(\cdot) = T_{\epsilon,f_\epsilon}V_\epsilon^*(\cdot),$$

y f_ϵ es una política óptima.

Demostración. **a)** De la Proposición 3.13 a) sabemos que existe u^ϵ punto fijo de T_ϵ y que existe $f_\epsilon \in \mathbb{F}$ tal que

$$u^\epsilon = T_{\epsilon,f_\epsilon}u^\epsilon.$$

Entonces, u^ϵ es el punto fijo de T_{ϵ,f_ϵ}, por lo que

$$u^\epsilon = V_\epsilon(x, f_\epsilon) \geq V_\epsilon^*. \qquad (3.6.12)$$

Para la desigualdad contraria, observemos que para cada $x \in \mathbf{X}_\epsilon$, $a \in A(x)$

$$u^\epsilon(x) \leq C(x,a) + \int_{\mathbf{X}_\epsilon\times\mathbb{R}^+} e^{-\alpha t} u^\epsilon(y) Q(dy, dt \mid x, a).$$

Entonces
$$0 \leq C(x,a) + \int_{\mathbf{X}_\epsilon \times \mathbb{R}^+} e^{-\alpha t} u^\epsilon(y) Q(dy, dt \mid x, a) - u^\epsilon(x).$$

$$= C\left(x,a\right) I_{[0<\tau]} + \int_{\mathbf{X} \times \mathbb{R}^+} e^{-\alpha t} u^\epsilon(y) I_{\mathbf{X}_\epsilon}(dy) Q\left(dy, dt \mid x_0, a_0\right) - u^\epsilon\left(x_0\right) I_{[0<\tau]}.$$

Por lo tanto para cada $m = 0, 1, 2, \ldots$ se cumple

$$0 \leq C\left(x_m, a_m\right) I_{[m<\tau]} + \int_{\mathbf{X}_e \times \mathbb{R}_+} e^{-\alpha t} u^\epsilon(y) Q\left(dy, dt \mid x_m, a_m\right) - u^\epsilon\left(x_m\right) I_{[m<\tau]},$$

de donde se sigue la desigualdad

$$0 \leq E_x^\pi \left[e^{-\alpha T_m} C\left(x_m, a_m\right) I_{[m<\tau]} + e^{-\alpha T_{m+1}} u^\epsilon\left(x_{m+1}\right) I_{[m<\tau]} - e^{-\alpha T_m} u^\epsilon\left(x_m\right) I_{[m<\tau]} \mid h_m, a_m \right].$$

Sumando desde $m = 0, 1, 2. \cdots, n$ y tomando esperanza tenemos

$$0 \leq E_x^\pi \left[\sum_{m=0}^{n} e^{-\alpha T_m} C\left(x_m, a_m\right) I_{[m<\tau]} + e^{-\alpha T_{n+1}} u^\epsilon\left(x_{n+1}\right) I_{[m<\tau]} - u^\epsilon\left(x_0\right) \right],$$

lo cual implica

$$E_x^\pi \left[\sum_{m=0}^{n} e^{-\alpha T_m} C\left(x_m, a_m\right) I_{[m<\tau]} \right] + E_x^\pi \left[e^{-\alpha T_{n+1}} u^\epsilon\left(x_{n+1}\right) I_{[m<\tau]} \right] \geq u^\epsilon(x). \tag{3.6.13}$$

Luego si $n \to \infty$ del Corolario 3.3 tenemos que

$$\begin{aligned} E_x^\pi \left[e^{-\alpha T_{n+1}} u^\epsilon\left(x_{n+1}\right) I_{[m<\tau]} \right] &\leq E_x^\pi \left[e^{-\alpha T_{n+1}} u^\epsilon\left(x_{n+1}\right) \right] \\ &\leq W(x) \|u^\epsilon\| \beta^n \to 0, \end{aligned}$$

de manera que tomando límite cuando $n \to \infty$ en (3.6.13)

$$E_x^\pi \left[\sum_{m=0}^{\infty} e^{-\alpha T_m} C\left(x_m, a_m\right) I_{[m<\tau]} \right] \geq u^\epsilon(x).$$

De aquí se sigue

$$V_\epsilon^*(x) \geq u^\epsilon(x). \tag{3.6.14}$$

Combinando (3.6.12) y (3.6.14) llegamos a que

$$V_\epsilon^*(x) = u^\epsilon(x).$$

Por lo tanto la función de valor óptimo $V_\epsilon^*(x)$ es el único punto fijo en $B_W\left(\mathbf{X}_\epsilon\right)$ del operador T_ϵ, es decir, se tiene la igualdad (3.6.11).

b) Del inciso a) se tiene que

$$V_\epsilon^*(x) = T_\epsilon V_\epsilon^*(x),$$

y de la Proposición 3.14 b) se sigue que existe $f_\epsilon \in \mathbb{F}$ tal que

$$V_\epsilon^*(\cdot) = T_{\epsilon,f_\epsilon} V_\epsilon^*(\cdot)$$

y f_ϵ es una política óptima.

□

Finalmente presentamos un resultado que nos provee una cota para el error producido en el proceso de truncamiento.

Proposición 3.16. *Supongamos que se cumplen las Hipótesis 3.1, 3.2 y 3.4. Entonces*

a) *Para cada política $\pi \in \Pi$ y estado inicial $x \in \mathbf{X}_\epsilon$,*

$$\|V(\cdot,\pi) - V_\epsilon(\pi,\cdot)\|_W^\epsilon \leq \frac{\bar{c}\epsilon}{(1-\beta)^2}$$

b)

$$\|V^* - V_\epsilon^*\|_W^\epsilon \leq \frac{\bar{c}\epsilon}{(1-\beta)^2}$$

Demostración. **a)** Dado $\epsilon > 0, x \in \mathbf{X}_\epsilon, \pi \in \Pi$, observemos que

$$\begin{aligned} V(x,\pi) &:= E_x^\pi \sum_{n=0}^{\infty} e^{-\alpha T_n} C\left(x_n, a_n\right) I_{[n<\tau]} + E_x^\pi \sum_{n=0}^{\infty} e^{-\alpha T_n} C\left(x_n, a_n\right) I_{[\tau \leq n]} \\ &= V_\epsilon(x,\pi) + E_x^\pi \sum_{n=0}^{\infty} e^{-\alpha T_n} C\left(x_n, a_n\right) I_{[\tau \leq n]}. \end{aligned}$$

Entonces

$$V(x,\pi) - V_\epsilon(x,\pi) = E_x^\pi \sum_{n=0}^{\infty} e^{-\alpha T_n} C\left(x_n, a_n\right) I_{[\tau \leq n]}.$$

De la Proposición 3.11, para cada $x \in \mathbf{X}_\epsilon$

$$\begin{aligned} |V(x,\pi) - V_\epsilon(x,\pi)| &= \left| E_x^\pi \sum_{n=0}^{\infty} e^{-\alpha T_n} C\left(x_n, a_n\right) I_{[\tau \leq n]} \right| \\ &\leq \frac{\bar{c} W(x)}{(1-\beta)^2} \epsilon, \end{aligned}$$

lo cual es equivalente a

$$\frac{|V(x,\pi) - V_\epsilon(x,\pi)|}{W(x)} \leq \frac{\bar{c}\epsilon}{(1-\beta)^2}.$$

Por lo tanto, tomando supremo sobre $x \in \mathbf{X}_\epsilon$ concluimos que

$$\|V(\cdot,\pi) - V_\epsilon(\cdot,\pi)\|_W^\epsilon \leq \frac{\bar{c}\epsilon}{(1-\beta)^2}.$$

b) Notemos que

$$|V^*(x) - V^*_\epsilon(x)| = \left| \inf_{\pi\in\Pi} V(\cdot,\pi) - \inf_{\pi\in\Pi} V_\epsilon(x,\pi) \right|$$

$$\leq \sup_{\pi\in\Pi} |V(x,\pi) - V_\epsilon(x,\pi)|$$

$$\leq \frac{\bar{c}W(x)}{(1-\beta)^2}\epsilon, \quad x \in \mathbf{X}.$$

De aquí se sigue que

$$\|V^* - V^*_\epsilon\|^\epsilon_W \leq \frac{\bar{c}\epsilon}{(1-\beta)^2}.$$

.

$\square$

Capítulo 4

Aproximación de Modelos de Control Markovianos

4.1 Introducción

En este capítulo retomamos el modelo de control markoviano estudiado en el Capítulo 1, donde nuestro objetivo es introducir una clase de métodos de aproximación a la solución del problema de control óptimo. Para este fin definimos una clase de operadores de aproximación, e introducimos un Modelo de Control de Markov perturbado (4.3.1), el cual está asociado a estos operadores, y es una aproximación del modelo original. Entonces en términos generales, nuestro método consiste en aplicar el algoritmo de iteración de políticas sobre este modelo perturbado, con el cual obtenemos una aproximación a la política óptima. Más aún, con el fin de precisar la aproximación, obtenemos cotas de error específicas. Dichos resultados los ilustramos con el ejemplo del sistema de inventario descrito en la Sección 1.6 del Capítulo 1.

4.2 Operadores de aproximación

Definición 4.1. *Un operador $L : \mathbf{M}(\mathbf{X}) \to \mathbf{M}(\mathbf{X})$ es un promediador si, satisface las siguientes condiciones:*

a) $L(I_X) = I_X$

b) *L es un operador lineal;*

c) *L es un operador positivo, esto es, $Lu(\cdot) \geq 0$ para cada $u(\cdot) \geq 0$ en $\mathbf{M}(\mathbf{X})$;*

d) *L satisface la siguiente propiedad de continuidad*

$$\textit{si} \quad \nu_n(\cdot) \downarrow 0, \nu_n \in M(X) \quad \textit{entonces} \quad L\nu_n(\cdot) \downarrow 0.$$

Para un operador $L : \mathbf{M}(\mathbf{X}) \to \mathbf{M}(\mathbf{X})$ la imagen Lu denotará la aproximación de u en $\mathbf{M}(\mathbf{X})$.
A continuación se presentan algunos ejemplos de estos operadores.

Para el análisis de algunos ejemplos unidimensionales, considérese $\mathbf{X} = [0, \theta]$ y una sucesión de puntos en $\mathbf{X}$, tales que $x_1 = 0 < x_2 < \cdots < x_k = \theta$, que generan la partición

$$\{D_i : D_1 = [x_1, x_2] \text{ y } D_i = (x_i, x_{i+1}], \quad i = 2, 3, \cdots, k-1\}.$$

Ejemplo 4.1. ***Aproximador constante por pedazos.***
Para cada $v \in \mathbf{M}_b(\mathbf{X})$, $y_i \in D_i$, se define el operador L como

$$Lv(x) = v(y_i), \quad x \in D_i.$$

Veamos que L es un promediador

Demostración. (a) Sea $x \in \mathbf{X}$ y supóngase que $x \in D_i$. Entonces,

$$LI_{\mathbf{X}}(x) = I_{\mathbf{X}}(y_i) = 1 = I_{\mathbf{X}}(x)$$

(b) Sean $u, v \in \mathbf{M}_b(\mathbf{X})$ y a, b, escalares reales. Considere $x \in \mathbf{X}$ y, sin pérdida de generalidad supóngase que $x \in D_i$. Entonces

$$\begin{aligned} L(au + bv)(x) &= (au + bv)(y_i) \\ &= au(y_i) + bv(y_i) \\ &= aLu(x) + bLv(x). \end{aligned}$$

(c) Si $v \geq 0$ y sea $x \in \mathbf{X}$, tal que $x \in D_i$. Entonces,

$$Lv(x) = v(y_i) \geq 0.$$

Como esto sucede para cualquier x, entonces L es un operador positivo.

(d) Sea $\{v_n\}_{n=1}^{\infty}$, tal que $v_n \in \mathbf{M}_b(\mathbf{X})$ y $v_n \downarrow 0$, considere un $x \in \mathbf{X}$ cualquiera y sin pérdida de generalidad, supóngase que $x \in D_i$. Entonces,

$$Lv_n(x) = v_n(y_i) \downarrow 0.$$

Por lo tanto el aproximador constante por pedazos es un promediador. □

Ejemplo 4.2. *Aproximador de interpolación lineal*
Para $v \in \mathbf{M}_b(\mathbf{X})$, se define el aproximador L como

$$Lv(x) = b_i(x)v(x_i) + \overline{b_i}(x)v(x_{i+1}), \quad x \in D_i, \quad i = 1, 2, \cdots, k-1,$$

donde

$$b_i(x) = (x_{i+1} - x) / (x_{i+1} - x_i), \quad x \in D_i, \quad i = 1, 2, \cdots, k-1$$

y

$$\overline{b_i}(x) = 1 - b_i(x).$$

Veamos que L es un promediador

Demostración. (a) Supóngase que $x \in D_i$. Puesto que

$$\begin{aligned} LI_{\mathbf{X}}(x) &= b_i(x)I_{\mathbf{X}}(x_i) + \overline{b_i}(x)I_{\mathbf{X}}(x_{i+1}) \\ &= b_i(x)I_{\mathbf{X}}(x) + \overline{b_i}(x)I_{\mathbf{X}}(x) \\ &= I_{\mathbf{X}}(x) \end{aligned}$$

(b) Sean $u, v \in \mathbf{M}_a(\mathbf{X}), a, b$, escalares reales. Considere $x \in \mathbf{X}$ y, sin pérdida de generalidad, supóngase que $x \in D_i$. Entonces

$$\begin{aligned} L(au+bv)(x) &= b_i(x)(au+bv)(x_i) + \bar{b}_i(x)(au+bv)(x_{i+1}) \\ &= b_i(x)\left[au(x_i) + bv(x_i)\right] + \overline{b_i}(x)\left[au(x_{i+1}) + bv(x_{i+1})\right] \\ &= a\left[b_i(x)u(x_i) + \overline{b_i}(x)u(x_{i+1})\right] + b\left[b_i(x)v(x_i) + \overline{b_i}(x)v(x_{i+1})\right] \\ &= aLu(x) + bLv(x). \end{aligned}$$

(c) Sean $v \geq 0$ y $x \in \mathbf{X}$, tal que $x \in D_i$, entonces

$$Lv(x) = b_i(x)v(x_i) + \bar{b}_i(x)v(x_{i+1}) \geq 0,$$

ya que $0 \leq b_i(x) \leq 1, 0 \leq \bar{b}_i(x) \leq 1$ y $v(x_i) \geq 0 \forall i = 1, 2, \cdots, k$. Entonces L es un operador positivo.

(d) Sea $\{v_n\}_{n=1}^{\infty}$, tal que $v_n \in \mathbf{M}_a(\mathbf{X})$ y $v_n \downarrow 0$, considere un $x \in \mathbf{X}$ cualquiera y sin pérdida de generalidad supongase que $x \in D_i$, entonces

$$Lv_n(x) = b_i(x)v_n(x_i) + \overline{b_i}(x)v_n(x_{i+1}) \downarrow 0$$

ya que $v(x_i) \downarrow 0$ para todo $i = 1, 2, \cdots, k$.

Por lo anterior, el interpolador lineal es un promediador. □

Definición 4.2. *Un promediador L es*

a) débilmente continuo si $Lv(\cdot)$ está en $\mathbf{C}(\mathbf{X})$ para cada $v(\cdot)$ en $\mathbf{C}(\mathbf{X})$;

b) fuertemente continuo si $Lv(\cdot)$ está en $\mathbf{C}(\mathbf{X})$ para cada $v(\cdot)$ en $\mathbf{M}(\mathbf{X})$

A continuación se presentan algunas propiedades importantes de los operadores de aproximación.

Proposición 4.1. *Sea L un promediador. Definamos el mapeo $L(\cdot \mid \cdot) : \mathcal{B}(\mathbf{X}) \times \mathbf{X} \to [0, 1]$, como*

$$L(B \mid x) := LI_B(x), \quad x \in X, \quad B \in \mathcal{B}(\mathbf{X}). \tag{4.2.1}$$

Entonces $L(\cdot \mid \cdot)$ es una probabilidad de transición en $\mathbf{X}$, esto es, $L(\cdot \mid x)$ es una medida de probabilidad en $\mathbf{X}$ para cada $x \in \mathbf{X}$ y $L(D \mid \cdot)$ es una función medible para cada $B \in \mathcal{B}(\mathbf{X})$.

Demostración. Se demostrará primero que $L(\cdot \mid x)$ es una medida de probabilidad en $\mathbf{X}$ para cada $x \in \mathbf{X}$.

i) $L(\mathbf{X} \mid x) := LI_{\mathbf{X}}(x) = I_{\mathbf{X}}(x) \equiv 1$ para todo $x \in \mathbf{X}$

ii) $L(\cdot \mid x) \geqslant 0$, por la definición 4.1 c)

iii) Sean $\{B_i\}_{i=1}^n$, subconjuntos ajenos de $\mathcal{B}(\mathbf{X})$. Entonces,

$$\begin{aligned} L\left(\cup_{i=1}^n B_i \mid x\right) &:= LI_{\cup_{i=1}^n B_i}(x) \\ &= L\left[\sum_{i=1}^n I_{B_i}(x)\right] \\ &= \sum_{i=1}^n \left[LI_{B_i}(x)\right], \text{ (por la Definición 4.1)} \\ &= \sum_{i=1}^n L\left(B_i \mid x\right), \quad \text{(por (4.2.1))} \end{aligned}$$

por lo que $L(\cdot \mid x)$ es finitamente aditiva.

Para ver que es numerablemente aditiva, considérese una sucesión $\{B_i\}_{i=1}^{\infty}$, de subconjuntos de $\mathcal{B}(\mathbf{X})$ tales que $B_i \downarrow \emptyset$. Entonces,

$$L\left(B_i \mid x\right) = LI_{B_i}(x) \downarrow LI_{\emptyset}(x) = 0,$$

por las Definiciones 4.1(d) y 4.1 (a). Por lo tanto, $L(\cdot \mid x)$ es numerablemente aditiva.

Se concluye que $L(\cdot \mid x)$ es una medida de probabilidad en $\mathbf{X}$ para cada $x \in \mathbf{X}$.

Por otra parte, $I_B : \mathbf{X} \to [0,1], \forall B \in \mathcal{B}(\mathbf{X})$, es una función medible y además acotada, es decir, $I_B \in \mathbf{M}_b(\mathbf{X})$, y como por definición $L : \mathbf{M}_b(\mathbf{X}) \to \mathbf{M}_b(\mathbf{X})$, entonces $LI_B \in \mathbf{M}_b(\mathbf{X})$. Por lo tanto $L(B \mid \cdot)$ es una función medible para cada $B \in \mathcal{B}(\mathbf{X})$.

$\square$

Proposición 4.2. *Sea L un promediador y $L(\cdot \mid \cdot)$ la probabilidad de transición definida en (4.2.1). Entonces,*

$$\int_{\mathbf{X}} v(y) L(dy \mid x) = Lv(x), \quad \forall x \in \mathbf{X}, v \in \mathbf{M}_b(\mathbf{X}).$$

Demostración. La demostración se hará primero para funciones indicadoras, luego para funciones simples y finalmente se generaliza para una función arbitraria.
Sea $v(x) = I_B(x)$, para algún $B \in \mathcal{B}(\mathbf{X})$. Entonces por la Proposición 4.1

$$\int_{\mathbf{X}} I_B(y) L(dy \mid x) = L(B \mid x) = LI_B(x)$$

Por lo tanto, la proposición se cumple para funciones indicadoras.
Ahora supóngase que v es una función simple, es decir, que tiene la forma

$$v(x) = \sum_{i=1}^n I_{B_i}(x), \text{ donde } B_i \in \mathcal{B}(\mathbf{X}), i = 1, 2, \cdots, n.$$

Entonces

$$\begin{aligned}
\int_{\mathbf{X}}\left[\sum_{i=1}^{n} I_{B_i}(x)\right] L(dy \mid x) &= \sum_{i=1}^{n}\int_{\mathbf{X}} I_{B_i}(x)L(dy \mid x) \\
&= \sum_{i=1}^{n} LI_{Bi}(x), \\
&= L\left(\sum_{i=1}^{n} I_{B_i}\right)(x),
\end{aligned}$$

por lo que la proposición se cumple para suma de indicadoras.
Consideremos ahora $v \in \mathbf{M}_b(\mathbf{X})$ una función no-negativa arbitraria. Existe una sucesión de funciones simples v_n que convergen crecientemente a v. Entonces, el teorema de convergencia monótona implica que

$$Lv_n(x) = \int_{\mathbf{X}} v_n(y)L(dy \mid x) \uparrow \int_{\mathbf{X}} v(y)L(dy \mid x)$$

para cada $x \in \mathbf{X}$.

Por otra parte, como $v_n \uparrow v$, entonces

$$\begin{aligned}
v - v_n &\downarrow 0 \\
L(v - v_n) &\downarrow 0, \quad \text{(por la Definición 4.1(d))} \\
Lv - Lv_n &\downarrow 0, \text{ (por la Definición 4.1 (b))} \\
Lv_n &\uparrow Lv.
\end{aligned}$$

Por la unicidad del límite, concluimos que

$$Lv(x) = \int_{\mathbf{X}} v(y)L(dy \mid x), \forall v \geq 0, x \in \mathbf{X}.$$

Entonces, la proposición se cumple para funciones no-negativas.
Finalmente se considera el caso general. Sea $v \in \mathbf{M}_b(\mathbf{X})$ arbitrario y v^+, v^- la parte positiva y negativa de v, respectivamente. Entonces,

$$\begin{aligned}
\int_{\mathbf{X}} v(y)L(dy \mid x) &= \int_{\mathbf{X}} \left[v^+(y) - v^-(y)\right] L(dy \mid x) \\
&= \int_{\mathbf{X}} v^+(y)L(dy \mid x) - \int_{\mathbf{X}} v^-(y)L(dy \mid x) \\
&= Lv^+(x) - Lv^-(x) \\
&= L\left[v^+(x) - v^-(x)\right] \\
&= Lv(x),
\end{aligned}$$

con lo cual, queda demostrada la proposición. □

Proposición 4.3. *Observemos que un promediador L, satiaface las siguientes condiciones*

a) *L es monótono;*

b) *L es un operador no expansivo con respecto a la norma del supremo, es decir, para todo $u, v \in \mathbf{M}_b(\mathbf{X})$*

$$\|Lu - Lv\|_\infty \leq \|u - v\|_\infty \tag{4.2.2}$$

Demostración. **a)** Sean $u, v \in \mathbf{M}_b(\mathbf{X})$, tales que $u \geq v$. Entonces, $u - v \geq 0$, lo cual implica que

$$\begin{aligned} L(u - v) &\geq 0, \\ Lu - Lv &\geq 0, \\ Lu &\geq Lv. \end{aligned}$$

Por lo tanto, L es monótono.

b) Sean $u, v \in \mathbf{M}_b(\mathbf{X})$

$$\begin{aligned} Lu(x) - Lv(x) &= \int_{\mathbf{X}} u(y)L(dy \mid x) - \int_{\mathbf{X}} v(y)L(dy \mid x) \\ &= \int_{\mathbf{X}} [u(y) - v(y)]L(dy \mid x) \end{aligned}$$

Entonces,

$$|Lu(x) - Lv(x)| \leq \sup_{x \in \mathbf{X}} |u(x) - v(x)| = \|u - v\|_\infty$$

lo cual prueba que L es no-expansivo con respecto a $\|\cdot\|_\infty$.

$\square$

4.3 Modelo markoviano perturbado

Sea L un promediador y consideremos un modelo de control markoviano fijo

$$(\mathbf{X}, \mathbf{A}, \{A(x) : x \in \mathbf{X}\}, Q, C)$$

Un modelo de Markov pertubado, denotado por

$$(\mathbf{X}, \mathbf{A}, \{A(x); x \in \mathbf{X}\}, \widetilde{C}, \widetilde{Q}), \tag{4.3.1}$$

es un modelo en el cual, el espacio de estados, el espacio de controles y el conjunto de controles admisibles, son iguales a los del modelo de control original, mientras que la función de costo por etapa $\widetilde{C}$ y el kérnel estocástico $\widetilde{Q}$, se definen para cada $x \in \mathbf{X}$, $B \in \mathcal{B}(\mathbf{X})$, $f \in \mathbb{F}$ como:

$$\widetilde{C}_f(x) := LC_f(x) = \int_{\mathbf{X}} C_f(y)L(dy \mid x), \tag{4.3.2}$$

$$\widetilde{Q}_f(B \mid x) := LQ_f(B \mid x) = \int_{\mathbf{X}} Q_f(B \mid y)L(dy \mid x) \tag{4.3.3}$$

Notemos que las igualdades (4.3.2) y (4.3.3) se deben a las propiedades de los promediadores establecidas en la Proposición 4.2

Observación 4.4. *Notemos que de (4.3.3), $\widetilde{Q}_f(\cdot \mid \cdot)$ es la composición de las probabilidades de transición Q y L y por lo tanto $\widetilde{Q}_f(\cdot \mid \cdot)$ es una probabilidad de transición en $\mathbf{X}$. Entonces del teorema de Ionescu-Tulcea ([2], Theorem 2.72. p109) se tiene que para cada estado inicial $x_0 = x \in \mathbf{X}$, y cada política $f \in \mathbb{F}$, existe una medida de probabilidad $\widetilde{P}_x^f$ y una cadena de Markov $\{\widetilde{x}_n\}$ definidas en el espacio $(\Omega, \mathcal{F})$ tal que $\widetilde{Q}_f(\cdot \mid \cdot)$ es la probabilidad de transición en un paso de $\{\widetilde{x}_n\}$. Denotamos por $\widetilde{E}_x^f$ al operador esperanza con respecto a la medida de probabilidad $\widetilde{P}_x^f$.*

El criterio de optimalidad y la función de valor óptimo para el modelo de control perturbado se definen a continuación.

Definición 4.3. *Sea $\alpha \in (0,1)$ un factor de descuento fijo; se define el costo descontado como*

$$\widetilde{V}(x,f) := \widetilde{E}_x^f\left[\sum_{k=0}^{\infty} \alpha^k \widetilde{C}_f\left(\tilde{x}_k\right)\right], x \in \mathbf{X}, f \in \mathbb{F}.$$

Diremos que una política estacionaria f^* es óptima si satisface la igualdad

$$\widetilde{V}(\cdot, f^*) = \inf_{f \in \mathbb{F}} \widetilde{V}(x,f) =: \widetilde{V}^*(x),$$

y en cuyo caso llamaremos a $\widetilde{V}^*(x)$ la función de valor α-óptimo.

Para dar solución al problema de control óptimo correspondiente, de manera análoga a capítulos anteriores se definen los siguientes operadores.

Para cada $f \in \mathbb{F}$, se define el operador $\widetilde{T}_f : \mathbf{M}_b(\mathbf{X}) \to \mathbf{M}_b(\mathbf{X})$ como

$$\widetilde{T}_f u(x) := \widetilde{C}_f(x) + \alpha \int_{\mathbf{X}} u(y) \widetilde{Q}_f(dy \mid x).$$

Definimos también para cada $u \in \mathbf{C}_b(\mathbf{X})$ el operador de programación dinámica para el modelo perturbado como

$$\widetilde{T}u(x) := \inf_{f \in \mathbb{F}} \left[\widetilde{C}_f(x) + \alpha \int_{\mathbf{X}} u(y) \widetilde{Q}_f(dy \mid x)\right] \quad x \in \mathbf{X}, \tag{4.3.4}$$

Lema 4.5. ***a)*** *Para toda función $u \in \mathbf{M}_b(\mathbf{X})$*

$$\widetilde{T}_f u(x) = L T_f u(x), \quad \forall x \in \mathbf{X}$$

b) *$\widetilde{T}_f$ es un operador de contracción en el espacio $(\mathbf{M}_b(\mathbf{X}), \|\cdot\|_\infty)$ con módulo α.*

c) *$\widetilde{V}_f$ es el punto fijo del operador $\widetilde{T}_f$, esto es,*

$$\widetilde{V}_f(x) = \widetilde{C}_f(x) + \alpha \int_{\mathbf{X}} \widetilde{V}_f(y) \widetilde{Q}_f(dy \mid x) = \widetilde{T}_f \widetilde{V}_f(x) \tag{4.3.5}$$

Demostración. **a)** Por la Proposición 4.1 tenemos

$$\begin{aligned}\widetilde{C}_f(x) + \alpha \int_{\mathbf{X}} u(y) \widetilde{Q}_f(dy \mid x) &= LC_f(x) + \alpha \int_{\mathbf{X}} \int_{\mathbf{X}} u(y) Q_f(dy \mid z) L(dz \mid x) \\ &= LT_f u(x)\end{aligned}$$

b) Este resultado se sigue notando del inciso anterior que $\widetilde{T}_f$ es la composición de un operador de contracción (T_f) con módulo α y un operador no expansivo (L).

c) Por el inciso (b) $\widetilde{T}_f$ tiene un único punto fijo $u_f \in \mathbf{M}_b(\mathbf{X})$, es decir,

$$u_f(x) = \widetilde{C}_f(x) + \alpha \int_{\mathbf{X}} u_f(y) \widetilde{Q}_f(dy \mid x).$$

Por otra parte, notemos que por propiedades de la esperanza condicional y la propiedad de Markov se cumple que

$$\begin{aligned}
\widetilde{V}_f(x) &= \widetilde{E}_x^f \left[\widetilde{C}_f\left(x_0\right) + \sum_{n=1}^{\infty} \alpha^n \widetilde{C}_f\left(\tilde{x}_n\right) \right] \\
&= \widetilde{C}_f(x) + \widetilde{E}_x^f \left\{ \widetilde{E}_x^f \left[\alpha \sum_{n=1}^{\infty} \alpha^{n-1} \widetilde{C}_f\left(\tilde{x}_n\right) \mid h_1 \right] \right\} \\
&= \widetilde{C}_f(x) + \widetilde{E}_x^f \left\{ \widetilde{E}_{x_1}^f \left[\alpha \sum_{n=0}^{\infty} \alpha^n \tilde{C}_f\left(\tilde{x}_n\right) \right] \right\} \\
&= \widetilde{C}_f(x) + \widetilde{E}_x^f \left\{ \alpha \widetilde{V}_f\left(\tilde{x}_1\right) \right\} \\
&= \widetilde{C}_f(x) + \alpha \int_{\mathbf{X}} \widetilde{V}_f(y) \widetilde{Q}_f(dy \mid x).
\end{aligned}$$

Entonces, $u_f = \widetilde{V}_f$ es el punto fijo del operador $\widetilde{T}_f$, esto es,

$$\widetilde{V}_f(x) = \widetilde{C}_f(x) + \alpha \int_{\mathbf{X}} \widetilde{V}_f(y) \widetilde{Q}_f(dy \mid x) = \widetilde{T}_f \widetilde{V}_f(x)$$

□

Observemos que se cumple

$$\begin{aligned}
\left\| \widetilde{T}_f^n u - \widetilde{V}_f \right\|_{\infty} &= \left\| \widetilde{T}_f \widetilde{T}_f^{n-1} u - \widetilde{T}_f \tilde{V}_f \right\|_{\infty} \\
&\leq \alpha \left\| \widetilde{T}_f^{n-1} u - \widetilde{V}_f \right\|_{\infty} \leq \cdots \leq \alpha^n \left\| u - \tilde{V}_f \right\|_{\infty}
\end{aligned}$$

Los siguientes resultados se establecerán en el contexto de la Hipótesis 1.1 del Capítulo 1. Para una fácil referencia la estableceremos de nuevo.

Hipótesis 1.1

a) $C(x,a)$ es una función continua en $\mathbb{K}$ y acotada por una constante $M > 0$;

b) $\mathbf{A}(x)$ es un suconjunto de $\mathbf{A}$ no-vacío y compacto para cada $x \in \mathbf{X}$ y el mapeo $x \to \mathbf{A}(x)$ es continuo;

c) $Q(\cdot \mid x, a)$ es débilmente continua en $\mathbb{K}$, es decir, la función

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

es continua para cada $u \in \mathbf{C}_b(\mathbf{X})$.

Lema 4.6. *Supongamos que se cumple la Hipótesis 1.1 que L es un promediador débilmente continuo. Entonces:*

a) *Para toda función $u \in \mathbf{C}_b(\mathbf{X})$*

$$\widetilde{T}u(x) = LTu(x), \quad \forall x \in \mathbf{X}. \tag{4.3.6}$$

b) *$\widetilde{T}$ es es un operador de contracción en el espacio $(\mathbf{C}_b(\mathbf{X}), \|\cdot\|_\infty)$ con módulo α.*

c) *Para cada $u \in \mathbf{C}_b(\mathbf{X})$ existe un selector medible $f \in \mathbb{F}$ tal que*

$$\widetilde{T}u(x) = \widetilde{T}_f u(x) = \widetilde{C}_f(x) + \alpha \int_{\mathbf{X}} u(y)\widetilde{Q}_f(dy \mid x) \tag{4.3.7}$$

Demostración. **a)** Observe que para $u \in \mathbf{C}_b(\mathbf{X})$ y $f \in \mathbb{F}$,

$$Tu(x) \leq C_f(x) + \alpha \int_{\mathbf{X}} u(y)Q_f(dy \mid x), x \in \mathbf{X}.$$

Ahora, como L es un operador monótono tenemos que para cada $x \in \mathbf{X}$

$$\begin{aligned} LTu(x) &\leq L\left\{C_f(x) + \alpha \int_{\mathbf{X}} u(y)Q_f(dy \mid x)\right\} \\ &= \widetilde{C}_f(x) + \alpha \int_{\mathbf{X}} u(y)\widetilde{Q}_f(dy \mid x). \end{aligned}$$

Tomando ínfimo sobre $f \in \mathbb{F}$ en la desigualdad anterior

$$\begin{aligned} LTu(x) &\leq \inf_{f \in \mathbb{F}} \left\{\widetilde{C}_f(x) + \alpha \int_{\mathbf{X}} u(y)\widetilde{Q}_f(dy \mid x)\right\} \\ &= \widetilde{T}u(x). \end{aligned}$$

Para obtener la desigualdad contraria, note que para cada u en $\mathbf{C}_b(\mathbf{X})$ existe f_u tal que

$$Tu(x) = C_{f_u}(x) + \alpha \int_{\mathbf{X}} u(y)Q_{f_u}(dy \mid x) \forall x \in \mathbf{X}. \tag{4.3.8}$$

Por otra parte, notemos que para $x \in \mathbf{X}$,

$$\begin{aligned} \widetilde{T}u(x) &\leq \widetilde{C}_{f_u}(x) + \alpha \int_{\mathbf{X}} u(y)\widetilde{Q}_{f_u}(dy \mid x) \\ &= L\left\{C_{f_u}(x) + \alpha \int_{\mathbf{X}} u(y)Q_{f_u}(dy \mid x)\right\} \\ &= LTu(x), \end{aligned}$$

donde la última igualdad se sigue de (4.3.8). Por lo tanto,

$$\widetilde{T}u(x) = LTu(x).$$

b) Para cada $u \in \mathbf{C}_b(\mathbf{X})$, por la Observación 1.5 se cumple $Tu \in \mathbf{C}_b(\mathbf{X})$. Además, como L es un promediador débilmente continuo, $LTu \in \mathbf{C}_b(\mathbf{X})$. Entonces, por la parte (a), $\widetilde{T}u \in \mathbf{C}_b(\mathbf{X})$, es decir, $\widetilde{T}$ mapea $\mathbf{C}_b(\mathbf{X})$ en sí mismo. Ahora, usando que el operador L es no expansivo, se sigue que

$$\|\widetilde{T}u - \tilde{T}v\|_\infty = \|LTu - LTv\|_\infty \leq \|Tu - Tv\|_\infty \leq \alpha\|u - v\|_\infty,$$

lo cual demuestra la parte (b).

c) Sea $u \in \mathbf{C}_b(\mathbf{X})$. Por (4.3.8) existe f_u tal que

$$Tu(x) = C_{f_u}(x) + \alpha \int_{\mathbf{X}} u(y) Q_{f_u}(dy \mid x).$$

Luego por la propiedad de linealidad de L se tiene

$$\widetilde{T}u(x) = \widetilde{C}_{f_u}(x) + \alpha \int_{\mathbf{X}} u(y)\widetilde{Q}_{f_u}(dy \mid x) \forall x \in \mathbf{X},$$

lo cual demuestra el resultado deseado.

□

Proposición 4.7. *Supongamos que se cumple la Hipótesis 1.1 y que L es un promediador débilmente continuo. Entonces:*

a) *La función de valor óptimo $\widetilde{V}^*$ es el único punto fijo del operador $\widetilde{T}$ en $\mathbf{C}_b(\mathbf{X})$, y existe $f^* \in \mathbb{F}$ tal que*

$$\widetilde{V}^*(x) = \widetilde{T}\widetilde{V}^*(x) = \widetilde{T}_{f^*}\widetilde{V}^*(x) \quad \forall x \in \mathbf{X}.$$

b) *Una politica estacionaria $f \in \mathbb{F}$ es óptima si y sólo $\widetilde{T}\widetilde{V}^*(x) = \widetilde{T}_f\widetilde{V}^*(x)$. Por lo tanto, la política f^* es óptima.*

Demostración. **a)** Por el Lema 4.6 existe una única función $u^* \in \mathbf{C}_b(\mathbf{X})$ y una política estacionaria $f^* \in \mathbb{F}$ tal que

$$u^*(x) = \widetilde{T}u^*(x) = \widetilde{C}_{f^*}(x) + \alpha \int_{\mathbf{X}} u^*(y)\widetilde{Q}_{f^*}(dy \mid x).$$

Entonces, u^* es un punto fijo de $\widetilde{T}_{f^*}$, y por (4.3.7) , $u^*(x) = \widetilde{V}_{f^*}(x)$. Por lo tanto, de la definición de $\widetilde{V}^*$, tenemos

$$u^*(x) = \widetilde{V}_{f^*}(x) \geq \widetilde{V}^*(x).$$

Para obtener la desigualdad contraria, notemos que

$$u^*(x) = \widetilde{T}u^*(x) \leq \widetilde{C}_f(x) + \alpha \int_{\mathbf{X}} u^*(y)\widetilde{Q}_f(dy \mid x), \quad f \in \mathbb{F}.$$

Entonces

$$0 \leq \widetilde{C}_f(x) + \alpha \int_{\mathbf{X}} u^*(y)\widetilde{Q}_f(dy \mid x) - u^*(x)$$

$$= \widetilde{E}_x^f \left\{ \widetilde{C}_f(\widetilde{x}_0) + \alpha u^*(\widetilde{x}_1) - u^*(\widetilde{x}_o) \right\}.$$

Notemos que en general, para cada $x \in \mathbf{X}$ y $h_m \in \mathbb{H}_m, m \in \mathbb{N}_0$, se cumple que

$$0 \leq \widetilde{E}_x^f \left\{ \widetilde{C}_f\left(\widetilde{x}_m\right) + \alpha u^*\left(\widetilde{x}_{m+1}\right) - u^*\left(\widetilde{x}_m\right) \mid h_m \right\},$$

de donde se sigue que

$$0 \leq \widetilde{E}_x^f \left\{ \alpha^m \widetilde{C}_f\left(\widetilde{x}_m\right) + \alpha^{m+1} u^*\left(\widetilde{x}_{m+1}\right) - \alpha^m u^*\left(\widetilde{x}_m\right) \mid h_m \right\}.$$

Tomando valor esperado y sumando se obtiene

$$\begin{aligned} 0 &\leq \widetilde{E}_x^f \left\{ \sum_{m=0}^{n} \alpha^m \widetilde{C}_f\left(\widetilde{x}_m\right) + \sum_{m=0}^{n} \alpha^{m+1} u^*\left(\widetilde{x}_{m+1}\right) - \sum_{m=0}^{n} \alpha^m u^*\left(\widetilde{x}_m\right) \right\} \\ &\leq \widetilde{E}_x^f \left\{ \sum_{m=0}^{n} \alpha^m \widetilde{C}_f\left(\widetilde{x}_m\right) + \alpha^{n+1} u^*\left(\widetilde{x}_{n+1}\right) - u^*\left(x_0\right) \right\}. \end{aligned}$$

Luego tomando límite cuando $n \to \infty$ obtenemos

$$u^*(x) \leq \widetilde{V}_f(x),$$

lo cual implica

$$u^*(x) \leq \widetilde{V}^*(x).$$

Por lo tanto,

$$\widetilde{V}^*(x) = \widetilde{T}\widetilde{V}^*(x).$$

Finalmente, sea $f^* \in \mathbb{F}$ una política estacionaria óptima, es decir, $\widetilde{V}^*(x) = \widetilde{V}_{f^*}(x)$. Por Lema (4.5), $\widetilde{V}_{f^*}$ es punto fijo de $\widetilde{T}_f*$ y $\widetilde{V}^*$ es un punto fijo de T, de donde se sigue

$$\widetilde{T}\widetilde{V}^*(x) = \widetilde{T}_{f^*}\widetilde{V}^*(x).$$

Ahora suponga que para una política $f \in \mathbb{F}$ se cumple $\widetilde{T}\widetilde{V}^*(\cdot) = \widetilde{T}_f\widetilde{V}^*(\cdot)$. Entonces, $\widetilde{V}^*$ es un punto fijo de $\widetilde{T}_f$, lo cual implica que

$$\widetilde{V}^*(\cdot) = \widetilde{V}_f(\cdot).$$

Por lo tanto, f es óptima.

□

4.4 Algoritmo de iteración de políticas aproximado

En esta sección se introduce un algoritmo de iteración de políticas aproximado (IPA). De manera general el algoritmo consiste en lo siguiente

i) Inicio: sean k y $f_0 \in \mathbb{F}$ arbitraria;

ii) etapa de evaluación: dado $f_k \in \mathbb{F}$ calcule $u_k(\cdot) := \widetilde{V}_{f_k}(\cdot)$;

iii) etapa de mejoramiento: calcule la política $f_{k+1} \in \mathbb{F}$ tal que $\widetilde{T}u_k(\cdot) = \widetilde{T}_{f_{k+1}}u_k(\cdot)$y regrese a la etapa (ii).

Proposición 4.8. *Supongamos que se cumple la Hipótesis 1.1 y que L es un promediador fuertemente continuo. Entonces la sucesión $\{u_k\} \in \mathbf{C}_b(\mathbf{X})$ converge decrecientemente y en la norma del supremo a la función de valor óptimo $\widetilde{V}^*$. Ademas,*

$$\left\| \widetilde{V}^* - u_k \right\|_\infty \leq \frac{2\alpha}{1-\alpha} \left\| u_k - u_{k-1} \right\|_\infty \quad \forall k \in \mathbb{N}.$$

Demostración. Veamos que la etapa de mejoramiento en el algoritmo IPA está bien definida. Sabemos que para cada función medible y acotada se tiene que

$$\widetilde{T}_f u(\cdot) = LT_f u(\cdot).$$

En particular para $\widetilde{V}_f$ se tiene que

$$\widetilde{T}_f \widetilde{V}_f = LT_f \widetilde{V}_f.$$

Luego, como $\widetilde{V}_f$ es punto fijo del operador $\widetilde{T}_f$, se tiene que

$$\widetilde{V}_f = L\widetilde{T}_f \widetilde{V}_f.$$

Como L es un promediador fuertemente continuo, $\widetilde{V}_f$ es una función continua y acotada. Entonces, para cada $k \in \mathbb{N}_0$, existe una política $f_{k+1} \in \mathbb{F}$ tal que $\tilde{T}u_k(\cdot) = \widetilde{T}_{f_{k+1}} u_k(\cdot)$. Por lo tanto la etapa de mejoramiento del algoritmo de iteracion de politicas está bien definida.
Para demostrar la convergencia de la sucesión $\{u_k\}$ a $\widetilde{V}^*$, tomemos $f_0 \in \mathbb{F}$ arbitraria y observemos que de (4.3.5) se tiene que

$$\begin{aligned}
u_0(x) &:= \widetilde{V}_{f_0}(x) \\
&= \widetilde{C}_{f_0}(x) + \alpha \int_{\mathbf{X}} u_0(y) \widetilde{Q}_{f_0}(dy \mid x) \\
&\geq \inf_{f \in \mathbb{F}} \left\{ \widetilde{C}_f(x) + \alpha \int_{\mathbf{X}} u_0(y) \widetilde{Q}_f(dy \mid x) \right\} \\
&= \widetilde{T} u_0(x).
\end{aligned}$$

Por otra parte, como sabemos que para cada función continua y acotada existe el selector en el cual el operador de programación dinámica definido para el modelo perturnado alcanza el ínfimo, entonces existe $f_1 \in \mathbb{F}$ tal que

$$\widetilde{T} u_0(x) = \widetilde{C}_{f_1}(x) + \alpha \int_{\mathbf{X}} u_0(y) \widetilde{Q}_{f_1}(dy \mid x),$$

lo cual implica

$$u_0(x) \geq \widetilde{C}_{f_1}(x) + \alpha \int_{\mathbf{X}} u_0(y) \widetilde{Q}_{f_1}(dy \mid x).$$

Iterando la desigualdad anterior de manera análoga a como se hizo en (1.5.1) obtenemos

$$u_0(x) \geq \widetilde{E}_x^{f_1} \sum_{k=0}^{n-1} \alpha^k \widetilde{C}_{f_1}(\widetilde{x}_k) + \alpha^n \widetilde{E}_x^{f_1} u_0(\widetilde{x}_n).$$

Tomando límite cuando $n \to \infty$, resulta

$$u_0(x) \geq \widetilde{V}_{f_1}(x) = u_1(x).$$

Repitiendo la argumentación anterior se obtiene

$$u_k(x) \geq \widetilde{T}u_k(x) \geq u_{k+1}(x), \quad \forall x \in \mathbf{X}, k \in \mathbb{N}_0,$$

por lo que la función $u = \lim u_k$ está bien definida. Además, tomando límite cuando $k \to \infty$, de (1.5.3) tenemos

$$\lim_{k\to\infty} Tu_k(x) = Tu(x).$$

Luego como $\widetilde{T} = LT$, de la igualdad anterior se cumplen las siguientes igualdades

$$\lim_{k\to\infty} \widetilde{T}u_k(x) = \lim_{k\to\infty} LTu_k(x) = L \lim_{k\to\infty} Tu_k(x) = LTu(x) = \widetilde{T}u(x).$$

Entonces,

$$u(x) \geq \widetilde{T}u(x) \geq u(x), \quad \forall x \in \mathbf{X}$$

lo cual prueba que u es punto fijo del operador $\widetilde{T}$, es decir,

$$u(\cdot) = \widetilde{T}u(\cdot).$$

Además, puesto que L es un promediador fuertemente continuo $u \in \mathbf{C}_b(\mathbf{X})$ y luego como la función de valor óptimo es el único punto fijo del operador de programación dinámica definido para el modelo perturbado concluimos que

$$u(\cdot) = \widetilde{V}^*(\cdot).$$

Ahora para demostrar la segunda parte de la proposición notemos que

$$\begin{aligned}
\left\|\widetilde{V}^* - u_k\right\|_\infty &\leq \left\|\widetilde{V}^* - \widetilde{T}_{f_k}u_{k-1}\right\|_\infty + \left\|\widetilde{T}_{f_k}u_{k-1} - u_k\right\|_\infty \\
&= \left\|\widetilde{T}\widetilde{V}^* - \widetilde{T}u_{k-1}\right\|_\infty + \left\|\widetilde{T}_{f_k}u_{k-1} - \widetilde{T}_{f_k}u_k\right\|_\infty \\
&\leq \alpha \left\|\widetilde{V}^* - u_{k-1}\right\|_\infty + \alpha \left\|u_{k-1} - u_k\right\|_\infty \\
&\leq \alpha \left\|\widetilde{V}^* - u_k\right\|_\infty + \alpha \left\|u_k - u_{k-1}\right\|_\infty + \alpha \left\|u_{k-1} - u_k\right\|_\infty,
\end{aligned}$$

lo cual implica que

$$\left\|\widetilde{V}^* - u_k\right\|_\infty - \alpha \left\|\widetilde{V}^* - u_k\right\|_\infty \leq \alpha \left\|u_k - u_{k-1}\right\|_\infty + \alpha \left\|u_{k-1} - u_k\right\|_\infty.$$

Por lo tanto

$$\left\|\widetilde{V}^* - u_k\right\|_\infty \leq \frac{2\alpha}{1-\alpha} \left\|u_k - u_{k-1}\right\|_\infty.$$

$\square$

Introducimos las siguientes constantes para obtener cotas para los errores de aproximación del algoritmo IPA

Definición 4.4. *Sea $\mathbb{F}_0$ la subclase de las políticas estacionarias que contiene las políticas óptimas estacionarias tanto para el modelo original como para el modelo aproximado y las políticas generadas por el algoritmo IPA. Entonces definimos*

$$\delta_C := \sup_{f\in\mathbb{F}_0} \left\| \widetilde{C}_f - C_f \right\|_\infty , \tag{4.4.1}$$

$$\delta_Q := \sup_{x\in\mathbf{X}, f\in\mathbb{F}_0} \left\| \widetilde{Q}_f(\cdot \mid x) - Q_f(\cdot \mid x) \right\|_{TV} , \tag{4.4.2}$$

donde $\|\cdot\|_{TV}$ representa la norma de la variación total para medidas con signo finitas, es decir

$$\|\lambda\|_{TV} := \sup\{|\int v(y)\lambda(dy)| : v \in M_b(X), \|v\|_\infty \le 1\},$$

donde λ es una medida con signo en $\mathbf{X}$.

Notemos que $\forall v \in M_b(\mathbf{X})$

$$\begin{aligned} \left|\int v(y)\lambda(dy)\right| &= \left|\int v(y)\frac{\|v\|_\infty}{\|v\|_\infty}\lambda(dy)\right| \\ &\le \|v\|_\infty \int \left|\frac{v(y)}{\|v\|_\infty}\lambda(dy)\right| \\ &\le \|\lambda\|_{TV}\|v\|_\infty. \end{aligned} \tag{4.4.3}$$

Teorema 4.9. *Supongamos que se cumple la Hipótesis 1.1 y que L es un promediador fuertemente continuo. Sea $u_k \in C_b(\mathbf{X})$ y $f_k \in \mathbb{F}, k \in \mathbb{N}$, las funciones y políticas definidas por el algoritmo IPA. Entonces:*

a)

$$\left\| \widetilde{V}^* - V^* \right\|_\infty \le \frac{1}{1-\alpha}\delta_C + \frac{\alpha M}{(1-\alpha)^2}\delta_Q. \tag{4.4.4}$$

b) *Para todo $k \in \mathbb{N}$*

$$\|V^* - V_{f_k}\|_\infty \le \frac{2}{1-\alpha}\delta_C + \frac{2\alpha M}{(1-\alpha)^2}\delta_Q + \frac{2\alpha}{1-\alpha}\|u_k - u_{k-1}\|_\infty$$

Demostración. **a)** De (1.4.4) y (4.3.5) se tiene que para cada política $f \in \mathbb{F}$,

$$\widetilde{V}_f(x) = \widetilde{C}_f(x) + \alpha\int_{\mathbf{X}} \widetilde{V}_f(y)\widetilde{Q}_f(dy \mid x), \quad V_f(x) = C_f(x) + \alpha\int_{\mathbf{X}} V_f(y)Q_f(dy \mid x) \quad \forall x \in \mathbf{X}.$$

Entonces.

$$\left|\widetilde{V}_f(x) - V_f(x)\right| \le \left|\widetilde{C}_f(x) - C_f(x)\right| + \alpha\left|\int_{\mathbf{X}} \widetilde{V}_f(y)\widetilde{Q}_f(dy \mid x) - \int_{\mathbf{X}} V_f(y)Q_f(dy \mid x)\right|$$

$$\le \left\|\widetilde{C}_f - C_f\right\|_\infty + \alpha\int \left|\widetilde{V}_f(y) - V_f(y)\right| \widetilde{Q}_f(dy \mid x) + \alpha\left|\int V_f(y)\left[\widetilde{Q}_f(dy \mid x) - Q_f(dy \mid x)\right]\right|.$$

$$\leq \left\|\widetilde{C}_f - C_f\right\|_\infty + \alpha \left\|\widetilde{V}_f - V_f\right\|_\infty + \alpha \left\|V_f\right\|_\infty \left\|\widetilde{Q}_f(\cdot \mid x) - Q_f(\cdot \mid x)\right\|_{TV},$$

donde esta última desigualdad se sigue de (4.4.3). Por lo tanto

$$\left|\widetilde{V}_f(x) - V_f(x)\right| \leq \left\|\widetilde{C}_f - C_f\right\|_\infty + \alpha \left\|\widetilde{V}_f - V_f\right\|_\infty + \alpha \left\|V_f\right\|_\infty \left\|\widetilde{Q}_f(\cdot \mid x) - Q_f(\cdot \mid x)\right\|_{TV}.$$

Tomando supremo sobre $x \in \mathbf{X}$, tenemos que

$$\left\|\widetilde{V}_f - V_f\right\|_\infty \leq \left\|\widetilde{C}_f - C_f\right\|_\infty + \alpha \left\|\widetilde{V}_f - V_f\right\|_\infty + \alpha \left\|V_f\right\|_\infty \sup_{x \in \mathbf{X}} \left\|\widetilde{Q}_f(\cdot \mid x) - Q_f(\cdot \mid x)\right\|_{TV}.$$

Ahora como $C(x, a)$ es una función continua en $\mathbb{K}$ y acotada por una constante $M > 0$, aplicando (1.3.1)

$$\left\|\widetilde{V}_f - V_f\right\| (1 - \alpha) \leq \left\|\widetilde{C}_f - C_f\right\|_\infty + \frac{\alpha M}{(1 - \alpha)} \sup_{x \in \mathbf{X}} \left\|\widetilde{Q}_f(\cdot \mid x) - Q_f(\cdot \mid x)\right\|_{TV}.$$

Entonces

$$\left\|\widetilde{V}_f - V_f\right\|_\infty \leq \frac{1}{1 - \alpha} \sup_{f \in \mathbb{F}_0} \left\|\widetilde{C}_f - C_f\right\|_\infty + \frac{\alpha}{(1 - \alpha)^2} M \sup_{x \in \mathbf{X}, f \in \mathbb{F}_0} \left\|\widetilde{Q}_f(\cdot \mid x) - Q_f(\cdot \mid x)\right\|_{TV}$$

$$= \frac{1}{1 - \alpha} \delta_C + \frac{\alpha}{(1 - \alpha)^2} M \delta_Q, \tag{4.4.5}$$

lo cual es equivalente a

$$-\kappa + V_f \leq \widetilde{V}_f \leq V_f + \kappa$$

donde

$$\kappa := \frac{1}{1 - \alpha} \delta_C + \frac{\alpha}{(1 - \alpha)^2} M \delta_Q.$$

Tomando ínfimo sobre la clase de las políticas estacionarias en esta desigualdad llegamos a que

$$-\kappa + V^* \leq \widetilde{V}^* \leq V^* + \kappa.$$

Por lo tanto

$$\left\|\widetilde{V}^* - V^*\right\|_\infty \leq \frac{1}{1 - \alpha} \delta_C + \frac{\alpha M}{(1 - \alpha)^2} \delta_Q.$$

b) Recordemos que $u_k(\cdot) = \widetilde{V}_{f_k}(\cdot)$. Entonces, aplicando la desigualdad (4.4.5) con f_k en lugar de f, tenemos que para toda $k \in \mathbf{N}_0$,

$$\left\|\widetilde{V}_{f_k} - V_{f_k}\right\|_\infty \leq \frac{1}{1 - \alpha} \delta_C + \frac{\alpha}{(1 - \alpha)^2} M \delta_Q. \tag{4.4.6}$$

Por otra parte,

$$\left\|V^* - V_{f_k}\right\|_\infty \leq \left\|V^* - \widetilde{V}^*\right\|_\infty + \left\|\widetilde{V}^* - u_k\right\|_\infty + \left\|\widetilde{V}_{f_k} - V_{f_k}\right\|_\infty.$$

Además recordemos que de la Proposición 4.8 se tiene que

$$\left\|\widetilde{V}^{*}-u_{k}\right\|_{\infty} \leq \frac{2 \alpha}{1-\alpha}\left\|u_{k}-u_{k-1}\right\|_{\infty} \quad \forall \quad k \in \mathbb{N}. \tag{4.4.7}$$

Por lo tanto de las desigualdades (4.4.4), (4.4.6) y (4.4.7) se tiene que

$$\left\|V^{*}-V_{f_{k}}\right\|_{\infty} \leq \frac{2}{1-\alpha} \delta_{C}+\frac{2 \alpha}{(1-\alpha)^{2}} M \delta_{Q}+\frac{2 \alpha}{1-\alpha}\left\|u_{k}-u_{k-1}\right\|_{\infty}.$$

□

4.5 Ejemplo: aproximación en un sistema de inventario

Retomemos el ejemplo de un sistema de inventario introducido en la Sección 1.6 del Capítulo 1. El propósito de esta seccion es hacer uso de la teoria de los promediadores introducida en la Sección 4.2 y realizar una aproximación a la función de valor óptimo, para la cual se calculan las cotas de error δ_c y δ_Q, las cuales son generadas por trabajar con aproximaciones de las funciones de costo y kernel estocástico.
Recordemos que se tiene que el espacio de estados y controles son $\mathbf{X}=\mathbf{A}=[0, \mathcal{K}]$, y que el conjunto de controles admisibles cuando el nivel de inventario es $x \in \mathbf{X}$ es $A(x)=[0, \mathcal{K}-x]$.
Para desarrollar esta aproximación usaremos el promediador de interpolación lineal, el cual definimos a continuación.
Consideremos una partición para el espacio de estados $\mathbf{X}=[0, \mathcal{K}]$, con

$$s_{1}=0<s_{2}<\cdots<s_{k-1}<s_{N}=\mathcal{K} \text{ y sea } \Delta:=\frac{\mathcal{K}}{N}.$$

Definimos para cada $x \in\left[s_{i}, s_{i+1}\right]$, $i=1,2, \ldots, N-1$, el operador L como

$$L v(x):=b_{i}(x) v\left(s_{i}\right)+\left(1-b_{i}(x)\right) v\left(s_{i+1}\right), \tag{4.5.1}$$

donde

$$b_{i}(x)=\frac{s_{i+1}-x}{s_{i+1}-s_{i}}, \quad \bar{b}_{i}(x)=1-b_{i}(x)=\frac{x-s_{i}}{s_{i+1}-s_{i}}.$$

Para calcular las cantidades δ_C y δ_Q definidas en (4.4.1) y (4.4.2) , suponemos que el modelo de inventario satisface las siguientes condiciones adicionales

Hipótesis 4.1. ***a)*** *La demanda esperada $\bar{\xi}:=E_{\xi}(\xi)$ es finita;*

b) *La densidad $\rho(\cdot)$ es acotada y Lipschitz continua con constante l, es decir,*

$$|\rho(y)-\rho(x)| \leq l|y-x|$$

Una política de umbral $S \geqslant 0$ es una política de la forma

$$f_{S}(x)= \begin{cases}S-x & \text{si } 0 \leq x \leq S \\ 0 & \text{si } x>S\end{cases}$$

Bajo la Hipótesis 4.1 existen políticas de umbral óptimo para el sistema de inventario considerando el índice en costo descontado. Este resultado puede verse en [8], [16] y se enuncia a continuación

Observación 4.10. *Si la demanda esperada $\bar{\xi} := E_\xi(\xi)$ es finita y la función de distribución F de las demandas tiene una densidad continua $\rho(\cdot)$ acotada, entonces existe una política de umbral $f_{S^*_\alpha}$ óptima en costo $\alpha-$ descontado. Además, S^*_α satisface la ecuación*

$$F\left(S^*_\alpha\right) = \frac{p-h-c}{p-\alpha c}$$

si $p > c+h$. En caso contrario $S^*_\alpha = 0$.

Para calcular las cotas por realizar la aproximación del sistema de inventario trabajaremos bajo la Hipótesis 4.1 y asumiremos que $p > c+h$. Además definimos

$$U(y) := \int_0^y F(z)dz, \quad y \geq 0.$$

La función de costo por etapa (1.6.2) se puede reescribir como

$$C(x,a) = p\bar{\xi} + ca + (h-p)(x+a) + p\int_0^{x+a} F(\xi)d\xi. \tag{4.5.2}$$

En efecto

$$\begin{aligned}
C(x,a) &= ca + h(x+a) + p\int_{x+a}^{\infty} (\xi - x - a)\rho(\xi)d\xi \\
&= ca + h(x+a) + p[\int_{x+a}^{\infty} \xi\rho(\xi)d\xi - \int_{x+a}^{\infty} (x+a)\rho(\xi)d\xi] \\
&= ca + h(x+a) + p[E_\xi(\xi) - \int_0^{x+a} \xi\rho(\xi)d\xi - (x+a) + (x+a)F(x+a)] \\
&= p\bar{\xi} + ca + (h-p)(x+a) + p[(x+a)F(x+a) - \int_0^{x+a} \xi\rho(\xi)d\xi] \\
&= p\bar{\xi} + ca + (h-p)(x+a) + p\int_0^{x+a} F(\xi)d\xi
\end{aligned}$$

Considerando la expresión (4.5.2) si $0 \leq x \leq S$ entonces

$$C_{f_S}(x) = p\bar{\xi} + c(S-x) + (h-p)(S) + p\int_0^S F(\xi)d\xi$$

$$= -cx + p\bar{\xi} + (c+h-p)S + pU(S).$$

Por otra parte, si $S < x \leq \mathcal{K}$, entonces

$$C_{f_S}(x) = p\bar{\xi} + (h-p)x + pU(x).$$

Por lo tanto, para cada política de umbral f_S se tiene que la función de costo queda de la siguiente manera

$$C_{f_S}(x) = \begin{cases} -cx + p\bar{\xi} + (c+h-p)S + pU(S) & \text{si } 0 \leq x \leq S \\ p\bar{\xi} + (h-p)x + pU(x) & \text{si } S < x \leq \mathcal{K}. \end{cases} \tag{4.5.3}$$

Además, si el umbral $S \in [s_i, s_{i+1}]$, entonces la interpolación lineal de $C_{f_S}(\cdot)$ toma la forma

$$\widetilde{C}_{f_S}(x) = \begin{cases} -cx + p\bar{\xi} + (c+h-p)S + pU(S) & \text{si } 0 \leq x \leq s_i \\ b_i(x)C_{f_S}(s_i) + \bar{b}_i(x)C_{f_S}(s_{i+1}) & \text{si } s_i < x \leq s_{i+1} \\ p\bar{\xi} + (h-p)x + p\widetilde{U}(x) & \text{si } s_{i+1} < x \leq \mathcal{K}, \end{cases} \tag{4.5.4}$$

donde $\widetilde{U}(\cdot)$ es la función de interpolación lineal de $U(\cdot)$.

Consideraemos una política f_S y supongamos que $S \in [s_i, s_{i+1}]$. Entonces de (4.5.3) y (4.5.4) se obtienen los siguientes resultados después de algunos cálculos, los cuales realizamos a detalle en el Apéndice C.

$$\begin{aligned} \left|\widetilde{C}_{f_S}(x) - C_{f_S}(x)\right| &= 0 \quad \forall x \in [0, s_i] \\ \left|\widetilde{C}_{f_S}(x) - C_{f_S}(x)\right| &\leq (2c+h)\Delta \quad \forall x \in [s_i, S] \\ \left|\widetilde{C}_{f_S}(x) - C_{f_S}(x)\right| &\leq (c+2h)\Delta \quad \forall x \in [S, s_{i+1}] \\ \left|\widetilde{C}_{f_S}(x) - C_{f_S}(x)\right| &\leq 2p\Delta \quad \forall x \in [s_{i+1}, \mathcal{K}]\,. \end{aligned}$$

De estas desigulades y como $p > c + h$ se sigue que

$$\left\|C_f - \widetilde{C}_{f_S}\right\|_\infty \leq \max(2c+h, c+2h, 2p)\Delta = 2p\Delta$$

Ahora recordemos que la dinámica del sistema de inventario está descrita por la siguiente ley de transición

$$Q(B \mid x, a) = I_B(0)[1 - F(x+a)] + \int_0^{x+a} I_B(x+a-\xi)F(d\xi).$$

Con el fin de dar expresiones concretas para $Q_{f_S}(\cdot \mid \cdot)$ y su perturbación con el interpolador lineal $\widetilde{Q}_{f_S}(\cdot \mid \cdot)$, se introducen, para cada $y \geq 0$, las siguientes funciones:

$$\begin{aligned} pv(y) &:= \int_0^y v(y)\rho(y-r)dr \\ Pv(y) &:= v(0)[1 - F(y)] + pv(y). \end{aligned}$$

Entonces

$$Q_{f_S}v(x) = \begin{cases} Pv(S) & \text{si } 0 \leq x \leq S \\ Pv(x) & \text{si } S < x \leq \mathcal{K}. \end{cases}$$

Además, si el umbral $S \in [s_i, s_{i+1}]$, entonces

$$\widetilde{Q}_{f_S}v(x) = \begin{cases} Pv(S) & \text{si } 0 \leq x \leq x_i \\ b_i(x)Q_{f_S}(s_i) + \bar{b}_i(x)Q_{f_S}(s_{i+1}) & \text{si } s_i < x \leq s_{i+1} \\ \widetilde{P}v(x) & \text{si } s_{i+1} < x \leq \mathcal{K}. \end{cases}$$

El siguiente resultado nos permitirá encontrar la cota de error generada por trabajar con la aproximación del kernel estocástico.

Lema 4.11. *Bajo la Hipótesis 4.1* $|Pv(y) - Pv(x)| \leq (2M' + l\mathcal{K})\,|y - x|$ *para todo* $x, y \in \mathbb{R}^+$*, donde* M' *es una cota de* $\rho(\cdot)$*.*

Demostración. Observe que para cada $v : \mathbb{R}^+ \to \mathbb{R}$ acotada y para todo $0 \leq x < y$ se tiene

$$\begin{aligned} pv(y) - pv(x) &= \int_0^y v(r)\rho(y-r)dr - \int_0^x v(r)\rho(x-r)dr \\ &= \int_0^x v(r)\rho(y-r)dr + \int_x^y v(r)\rho(y-r)dr - \int_0^x v(r)\rho(x-r)dr \\ &= \int_0^x v(r)[\rho(y-r) - \rho(x-r)]dr + \int_x^y v(r)\rho(y-r)dr. \end{aligned}$$

En particular, si $\|v\|_\infty \leq 1$, resulta que

$$\begin{aligned} |pv(y) - pv(x)| &\leq \int_0^x |\rho(y-r) - \rho(x-r)|dr + \int_x^y \rho(y-r)dr \\ &\leq xl|y-x| + M'|y-x| \\ &\leq \left(\mathcal{K}l + M'\right)|y-x|. \end{aligned} \tag{4.5.5}$$

Entonces,

$$\begin{aligned} |Pv(y) - Pv(x)| &= |v(0)[1 - F(y)] + pv(y) - v(0)[1 - F(x)] - pv(x)| \\ &\leq |F(y) - F(x)| + |pv(y) - pv(x)| \\ &\leq \left|\int_x^y \rho(r)dr\right| + \left(\mathcal{K}l + M'\right)|y-x| \\ &\leq M'|y-x| + \left(\mathcal{K}l + M'\right)|y-x|, \end{aligned}$$

lo cual implica que

$$|Pv(y) - Pv(x)| \leq \left(2M' + \mathcal{K}l\right)|y-x|. \tag{4.5.6}$$

□

Calculemos δ_Q. Para esto observe que si $x \in [0, s_i]$, entonces

$$\left|\widetilde{Q}_{f_S}v(x) - Q_{f_S}v(x)\right| = |Pv(S) - Pv(S)| = 0.$$

Por otra parte para $x \in [s_i, S]$, se tiene

$$\begin{aligned} \widetilde{Q}_{f_S}v(x) - Q_{f_S}v(x) &= b_i(x)Pv(S) + \bar{b}_i(x)Pv\left(s_{i+1}\right) - Pv(S) \\ &= -\left(1 - b_i(x)\right)Pv(S) + \bar{b}_i(x)Pv\left(s_{i+1}\right) \\ &= \bar{b}_i(x)\left[Pv\left(s_{i+1}\right) - Pv(S)\right]. \end{aligned}$$

Entonces,

$$\left|\widetilde{Q}_{f_S}v(x) - Q_{f_S}v(x)\right| \leq \left|Pv\left(s_{i+1}\right) - Pv(S)\right|$$

Utilizando (4.5.6) se tiene

$$\begin{aligned} \left|\widetilde{Q}_{f_S}v(x) - Q_{f_S}v(x)\right| &\leq \left(2M' + \mathcal{K}l\right)|s_{i+1} - S| \\ &\leq \left(2M' + \mathcal{K}l\right)\Delta. \end{aligned}$$

Si $x \in [S, s_{i+1}]$, se cumple que

$$\begin{aligned}\widetilde{Q}_{f_S} v(x) - Q_{f_S} v(x) =& b_i(x) Pv(S) + \bar{b_i}(x) Pv\left(s_{i+1}\right) - Pv(x) \\ =& - b_i(x)[Pv(x) - Pv(S)] + \bar{b_i}(x)\left[Pv\left(s_{i+1}\right) - Pv(x)\right].\end{aligned}$$

Entonces,

$$\left|\widetilde{Q}_{f_S} v(x) - Q_{f_S} v(x)\right| \leq |Pv(x) - Pv(S)| + \left|Pv\left(s_{i+1}\right) - Pv(x)\right|.$$

Utilizando (4.5.6) se tiene

$$\begin{aligned}\left|\widetilde{Q}_{f_s} v(x) - Q_{f_s} v(x)\right| &\leq \left(2M' + \mathcal{K}l\right)|x - S| + \left(2M' + \mathcal{K}l\right)|s_{i+1} - x| \\ &\leq \left(2M' + \mathcal{K}l\right)\Delta + \left(2M' + \mathcal{K}l\right)\Delta \\ &= 2\left(2M' + \mathcal{K}l\right)\Delta.\end{aligned}$$

Si $x \in [s_{i+1}, \mathcal{K}]$ entonces

$$\left|\widetilde{Q}_{f_S} v(x) - Q_{f_S} v(x)\right| = |\widetilde{P}v(x) - Pv(x)|$$

$$\begin{aligned}&= |v(0)[1 - \widetilde{F}(x)] + \widetilde{p}v(x) - v(0)[1 - F(x)] - pv(x)| \\ &\leq |\widetilde{F}(x) - F(x)| + |\widetilde{p}v(x) - pv(x)|.\end{aligned}$$

Considerando $x \in [s_j, s_{j+1}]$ con $j \geq i+1$, se tiene que

$$\begin{aligned}|\widetilde{F}(x) - F(x)| &= \left|b_j(x) F\left(s_j\right) + \bar{b_j}(x) F\left(s_{j+1}\right) - F(x)\right| \\ &= \left|-b_j(x)\left[F(x) - F\left(s_j\right)\right] + \bar{b_j}(x)\left[F\left(s_{j+1}\right) - F(x)\right]\right| \\ &\leq \int_{s_j}^{x} \rho(r) dr + \int_{x}^{s_{j+1}} \rho(r) dr \\ &\leq M'\left(x - s_j\right) + M'\left(s_{j+1} - x\right) \\ &\leq 2M'\Delta.\end{aligned}$$

Además,

$$\begin{aligned}|\widetilde{p}v(x) - pv(x)| &= \left|b_j(x) pv\left(s_j\right) + \bar{b_j}(x) pv\left(s_{j+1}\right) - pv(x)\right| \\ &= \left|b_j(x)\left[pv\left(s_j\right) - pv(x)\right] + \bar{b_j}(x)\left[pv\left(s_{j+1}\right) - pv(x)\right]\right| \\ &\leq \left|pv\left(s_j\right) - pv(x)\right| + \left|pv\left(s_{j+1}\right) - pv(x)\right| \\ &\leq \left(\mathcal{K}l + M'\right)|s_j - x| + \left(\mathcal{K}l + M'\right)|s_{j+1} - x|\end{aligned}$$

por (4.5.5)

$$\leq 2\left(\mathcal{K}l + M'\right)\Delta.$$

Entonces,

$$\begin{aligned}\left|\widetilde{Q}_{f_S} v(x) - Q_{f_S} v(x)\right| &\leq 2M'\Delta + 2\left(\mathcal{K}l + M'\right)\Delta \\ &\leq 2\left(\mathcal{K}l + 2M'\right)\Delta\end{aligned}$$

Las cotas anteriores combinadas con el Teorema 4.9 dejan las siguientes desigualdades:

$$\left\| \widetilde{V}^* - V^* \right\|_\infty \leq \left\{ \frac{2p}{1-\alpha} + \frac{\alpha \left(2Kl + 4M' \right) M}{(1-\alpha)^2} \right\} \Delta$$

y

$$\left\| V^* - V_{f_k} \right\|_\infty \leq \left\{ \frac{2p}{1-\alpha} + \frac{\alpha \left(2Kl + 4M' \right) M}{(1-\alpha)^2} \right\} 2\Delta + \frac{2\alpha}{1-\alpha} \left\| u_k - u_{k-1} \right\|_\infty ,$$

donde f_k es la política tal que

$$\widetilde{T} u_{k-1}(\cdot) = \widetilde{T}_{f_k} u_{k-1}(\cdot),$$

y M es la cota para el costo por etapa.

Claramente, $\left\| \widetilde{V}^* - V^* \right\|_\infty$ y $\left\| V^* - V_{f_k} \right\|_\infty$ se pueden hacer arbitrariamente pequeñas tomando una partición suficientemente fina y realizando un número suficientemente grande de iteraciones del algoritmo de iteración de políticas aproximado.

Capítulo 5

Aproximación de Modelos de Control Semi-Markovianos

5.1 Introducción

El objetivo de este capítulo es desarrollar un método de aproximación para procesos de control semi-markovianos, siguiendo el esquema que se aplicó en el capítulo anterior, pero asumiendo costos posiblemente no acotados. Como primer paso, introducimos un modelo de control semi-markoviano perturbado en 5.2.1, el cual está asociado a un operador de aproximación definido. Posteriormente en la Sección 5.3 aplicamos un proceso de truncamiento al modelo perturbado, lo cual nos permite definir un modelo perturbado-truncado de tal modo que dicho modelo hereda las propiedades importantes del modelo original. Posteriormente para aproximar la política óptima del modelo original, implementamos un algoritmo de iteración de políticas en el modelo perturbado-truncado y presentamos cotas de error por aproximar la solución del modelo original por medio de este proceso.

5.2 Modelo semi markoviano perturbado

Consideremos el modelo de control semi-markoviano original (3.2.1),

$$(\mathbf{X}, \mathbf{A}, \{A(x) : x \in \mathbf{X}\}, Q, C),$$

y sea L un promediador (ver Definición 4.1). Entonces un modelo de control semi markoviano perturbado, es un arreglo de la forma

$$(\mathbf{X}, \mathbf{A}, \{A(x) : x \in \mathbf{X}\}, \widetilde{Q}, \widetilde{C}) \tag{5.2.1}$$

donde, para cada $x \in \mathbf{X}, a \in \mathbf{A}$ y $B \in \mathcal{B}(\mathbf{X})$ la función de costo por etapa y el kérnel de transición correspondientes están definidos como

$$\widetilde{C}(x,a) := LC(x,a) = \int_{\mathbf{X}} C(y,a) L(dy \mid x) \tag{5.2.2}$$

$$\widetilde{Q}(B,(0,t] \mid x,a) := LQ(B,(0,t] \mid x,a) \tag{5.2.3}$$

Observación 5.1. *Observemos que de manera análoga a como se mencionó en la Observación 4.4, el kernel conjunto $\widetilde{Q}(\cdot,\cdot \mid \cdot,\cdot)$ es la composición de las probabilidades de transición Q y L. Por lo tanto, $\widetilde{Q}$ es una probabilidad de transición en $\mathbf{X} \times \mathbb{R}^+$ dado $\mathbb{K}$.*

Para cada política $\pi \in \Pi$ y estado inicial $x \in \mathbf{X}$, la ley de transición $\widetilde{Q}$ define una medida de probabilidad $\widetilde{P}_{x,\pi}$ y un proceso estocástico $\left\{\widetilde{x}_n, \widetilde{a}_n, \widetilde{\delta}_n\right\}$ en el espacio $(\Omega, \mathcal{F}) = \left((\mathbf{X} \times \mathbf{A} \times \mathbb{R}^+)^\infty, \mathcal{F}\right)$. Denotemos por $\widetilde{E}_x^\pi$ el operador esperanza con respecto a la medida de probabilidad $P_{x,\pi}$.

A continuación definimos el índice de funcionamiento correspondiente para este modelo.

Definición 5.1. *Para $x \in \mathbf{X}, \pi \in \Pi$ y $\alpha > 0$, se define el costo descontado del modelo semi-markoviano perturbado por*

$$\widetilde{V}(x,\pi) := \widetilde{E}_x^{\pi} \sum_{k=0}^{\infty} e^{-\alpha T_k} \widetilde{C}\left(\widetilde{x}_k, \widetilde{a}_k\right), \quad x \in \mathbf{X}.$$

La función de valor óptimo se define como

$$\widetilde{V}^*(x) := \inf_{\pi} \widetilde{V}(x,\pi).$$

Una política π^* es óptima si

$$\widetilde{V}^*(x) = \widetilde{V}(x,\pi^*)$$

En esta sección asumimos que se cumplen las Hipótesis 3.1 y 3.2, mismas que se trabajaron en el Modelo de control Semi-Markoviano original (3.2.1). De manera que con intención de facilitar la exposición las enunciamos de nuevo.

Hipótesis 3.1

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

b) $C(x,a)$ es una función continua en $A(x)$ para cada $x \in \mathbf{X}$.

c) La función

$$a \to \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} u(y) Q(dy, dt \mid x, a)$$

es continua para cada $u \in \mathbf{M}_b(\mathbf{X}), x \in \mathbf{X}$.

Hipótesis 3.2

Existe una función medible $W : \mathbf{X} \to [1,\infty)$, y constantes $\beta \in (0,1)$, y $\bar{c} \in \mathbb{R}$ tales que para cada $x \in \mathbf{X}$ se tiene que

a) $|C(x,a)| \leq \bar{c}W(x)$

b) $\displaystyle\int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} W(y) Q(dy, dt \mid x, a) \leq \beta W(x)$

c) la función

$$a \to \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} W(y) Q(dy, dt \mid x, a)$$

es continua en $A(x)$.

La Hipótesis que enunciamos a continuación asegura que el modelo semi-markoviano perturbado (3.2.5) hereda propiedades importantes del modelo semi- markoviano original (3.2.1). Es decir, se satisfacen condiciones análogas a las condiciones impuestas por la Hipótesis 3.1 y 3.2 en el modelo semi-markoviano original (3.2.1). Por otra parte, con la finalidad de simplificar la exposición supondremos que $A(x) = \mathbf{A} \quad \forall x \in \mathbf{X}$.

Consideremos la función medible $W : \mathbf{X} \to [1;\infty)$, y la constante $\beta \in (0,1)$ de la Hipótesis 3.2.

Hipótesis 5.1. *El promediador L satisface la siguiente condición.*

a) $\widetilde{W} := L(W) \in B_W(\mathbf{X})$,

b) $\widetilde{\beta} := \beta \|\widetilde{W}\|_W < 1$.

Proposición 5.2. *Bajo las Hipótesis 3.1, 3.2 y 5.1 se cumple lo siguiente*

a) $|\widetilde{C}(x,a)| \leq \widetilde{c}W(x)$ *para toda* $(x,a) \in \mathbb{K}, \text{con}\, \widetilde{c} := \bar{c}\|\widetilde{W}\|_W$;

b) $\widetilde{C}(x,\cdot)$ *es una función continua en* $\mathbf{A}$*, para cada* $x \in \mathbf{X}$*;*

c) *para cada* $u \in B_W(\mathbf{X})$ *la función*

$$a \to \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} u(y) \widetilde{Q}(dy, dt \mid x, a)$$

es continua en $\mathbf{A}$ *para cada* $x \in \mathbf{X}$*;*

d) $\displaystyle\int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} W(y) \widetilde{Q}(dy, dt \mid x, a) \leq \tilde{\beta} W(x)$ *para cada* $(x,a) \in \mathbb{K}$;

e) *la función* $a \to \displaystyle\int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} W(y) \widetilde{Q}(dy, dt \mid x, a)$ *es continua en* $\mathbf{A}$*, para cada* $x \in \mathbf{X}$.

Demostración. **a)** Observe que de la Hipótesis 3.2 a) y la linealidad y monotonía del operador L se obtiene la desigualdad

$$\begin{aligned} |\widetilde{C}(x,a)| = |LC(x,a)| &\leq L\bar{c}W(x) \\ &\leq \bar{c}LW(x) \\ &\leq \bar{c}\widetilde{W}(x) \\ &\leq \bar{c}W(x)\|\widetilde{W}\|_W, \end{aligned}$$

de donde se sigue el resultado.

b) De la Hipótesis 3.1 (b) y la Hipótesis 3.2 a), $C(x,a)$ es continua en $a \in A$ y acotada por $\bar{c}W(x)$. Sea $a_n, n \in \mathbb{N}$, una sucesión que converge a $a \in \mathbf{A}$, y observe que

$$\widetilde{C}(x, a_n) = LC\left(x, a_n\right) = \int_{\mathbf{X}} C\left(y, a_n\right) L(dy \mid x).$$

Entonces por el teorema de convergencia dominada tenemos que

$$\begin{aligned} \lim_{n\to\infty} LC\left(x, a_n\right) &= \lim_{n\to\infty} \int_{\mathbf{X}} C\left(y, a_n\right) L(dy \mid x) \\ &= \int_{\mathbf{X}} \lim_{n\to\infty} C\left(y, a_n\right) L(dy \mid x) \\ &= \int_{\mathbf{X}} C(y,a) L(dy \mid x) = LC(x,a). \end{aligned}$$

Por la tanto $\widetilde{C}(x,a)$ es continua en $a \in \mathbf{A}$.

c) De la Hipótesis 3.1 c), para cada $u \in \mathbf{M}_b(\mathbf{X})$, $x \in \mathbf{X}$ la función

$$a \to \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} u(y) Q(dy, dt \mid x, a)$$

es continua. Por otro lado de la Hipótesis 3.2 la función

$$a \to \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} W(y) Q(dy, dt \mid x, a)$$

es continua en $A(x)$. Por lo que de [10, Lema 8.3.7] se sigue que para cada $u \in B_W(\mathbf{X})$, la función

$$a \to \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} u(y) Q(dy, dt \mid x, a)$$

es continua en $\mathbf{A}$ para cada $x \in \mathbf{X}$. Además de la Hipótesis 3.2 b) se tiene que

$$\begin{aligned}
\int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} u(y) Q(dy, dt \mid x, a) &= \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} \frac{W(x)}{W(x)} u(y) Q(dy, dt \mid x, a) \\
&\leq \int_{\mathbf{X}\times\mathbb{R}^+} \|u\|_W e^{-\alpha t} W(y) Q(dy, dt \mid x, a) \\
&\leq \beta \|u\|_W W(x).
\end{aligned}$$

Por otra parte, observemos que

$$\int_{\mathbf{X}} \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} u(y) Q(dy, dt \mid z, a) L(dz \mid x) = \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} u(y) \widetilde{Q}(dy, dt \mid x, a)$$

de manera que si tomamos una sucesión $\{a_n\}$ que converge a $a \in \mathbf{A}$, entonces por el teorema de convergencia dominada se tiene que

$$\begin{aligned}
\lim_{n\to\infty} \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} u(y) \widetilde{Q}(dy, dt \mid x, a_n) &= \lim_{n\to\infty} \int_{\mathbf{X}} \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} u(y) Q(dy, dt \mid z, a_n) L(dz \mid x) \\
&= \int_{\mathbf{X}} \lim_{n\to\infty} \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} u(y) Q(dy, dt \mid z, a_n) L(dz \mid x) \\
&= \int_{\mathbf{X}} \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} u(y) Q(dy, dt \mid z, a) L(dz \mid x) \\
&= \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} u(y) \widetilde{Q}(dy, dt \mid x, a),
\end{aligned}$$

de donde se sigue el resultado.

d) De la Hipótesis 3.2 b)

$$\int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} W(y) Q(dy, dt \mid x, a) \leq \beta W(x).$$

Entonces de la monotonía del operador L se deduce la desigualdad

$$\int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} W(y) \widetilde{Q}(dy, dt \mid x, a) \leq \beta \widetilde{W}(x) \leq \beta \|\widetilde{W}\|_W W(x) = \widetilde{\beta} W(x) \quad (x, a) \in \mathbb{K}.$$

e) Observemos que

$$\int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} W(y) \widetilde{Q}(dy, dt \mid x, a) = \int_{\mathbf{X}} \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} W(y) Q(dy, dt \mid z, a) L(dz \mid a).$$

Ahora, de manera análoga al inciso b), tomando una sucesión $a_n \to a$, y considerando la Hipótesis 3.2 b) se tiene que

$$
\begin{aligned}
\lim_{n\to\infty}\int e^{-\alpha t}W(y)\widetilde{Q}(dy,dt\mid x,a_n) &= \lim_{n\to\infty}\int_{\mathbf{X}}\int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t}W(y)Q(dy,dt\mid z,a_n)L(dz\mid a_n)\\
&= \int_{\mathbf{X}}\lim_{n\to\infty}\int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t}W(y)Q(dy,dt\mid z,a_n)L(dz\mid a_n)\\
&= \int_{\mathbf{X}}\int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t}W(y)Q(dy,dt\mid z,a)L(dz\mid x)\\
&= \int_{\mathbf{X}\times\mathbb{R}^+} e^{\alpha t}W(y)\widetilde{Q}(dy,dt\mid x,a).
\end{aligned}
$$

□

Por otra parte, de la Proposición 5.2 tenemos que el modelo semi-markoviano perturbado satisface resultados análogos a los presentados en el Capítulo 3 para el modelo semi markoviano original (3.2.1). A cotinuación presentamos dichos resultados.

Proposición 5.3. *Supongamos que se cumplen las Hipótesis 3.1, 3.2 y 5.1. Entonces para cada $x\in\mathbf{X},\pi\in\Pi$, y $n\in\mathbb{N}$, se cumple*

a) $\widetilde{E}_x^{\pi}\left[e^{-\alpha T_n}W\left(\widetilde{x}_n\right)\right]\leq\widetilde{\beta}^n W(\widetilde{x})$;

b) *para cada $u\in B_W(\mathbf{X})$,*

$$
\widetilde{E}_x^{\pi}\left\{e^{-\alpha T_n}u\left(\widetilde{x}_n\right)\right\}\leq\widetilde{\beta}^n W(\widetilde{x})\|u\|_W;
$$

c) $\widetilde{E}_x^{\pi}\sum_{n=0}^{\infty}\left[e^{-\alpha T_n}W\left(\widetilde{x}_n\right)\right]\leq\dfrac{W(\widetilde{x})}{1-\widetilde{\beta}}$;

d) $\|\widetilde{V}(x,\pi)\|_W\leq\dfrac{\bar{c}}{1-\widetilde{\beta}}$.

Demostración.

a) Por propiedades de la esperanza condicional y de Proposición 5.2 d), se tiene que

$$
\begin{aligned}
\widetilde{E}_x^{\pi}\left[e^{-\alpha T_n}W\left(\widetilde{x}_n\right)\right] &= \widetilde{E}_x^{\pi}\left[\widetilde{E}_x^{\pi}\left\{e^{-\alpha T_{n-1}}e^{-\alpha\delta_n}W\left(\widetilde{x}_n\right)\mid h_{n-1},\widetilde{a}_{n-1}\right\}\right]\\
&= \widetilde{E}_x^{\pi}\left[\widetilde{E}_x^{\pi}\left\{e^{-\alpha T_{n-1}}e^{-\alpha\delta_n}W\left(\widetilde{x}_n\right)\mid h_{n-1},\widetilde{a}_{n-1}\right\}\right]\\
&= \widetilde{E}_x^{\pi}\left[e^{-\alpha T_{n-1}}\widetilde{E}_x^{\pi}\left\{e^{-\alpha\delta_n}W\left(\widetilde{x}_n\right)\mid h_{n-1},\widetilde{a}_{n-1}\right\}\right]\\
&= \widetilde{E}_x^{\pi}\left[e^{-\alpha T_{n-1}}\int_{\mathbf{A}}\int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t}W(y)\widetilde{Q}\left(dy,dt\mid\widetilde{x}_{n-1},\widetilde{a}\right)\pi\left(da\mid h_{n-1}\right)\right]\\
&\leq \widetilde{E}_x^{\pi}\left[e^{-\alpha T_{n-1}}\int_{\mathbf{A}}\widetilde{\beta}W\left(\widetilde{x}_{n-1}\right)\pi\left(da\mid h_{n-1}\right)\right]\\
&= \widetilde{\beta}\widetilde{E}_x^{\pi}\left[e^{-\alpha T_{n-1}}W\left(\widetilde{x}_{n-1}\right)\right].
\end{aligned}
$$

Entonces

$$
\widetilde{E}_x^{\pi}\left[e^{-\alpha T_n}W\left(\widetilde{x}_n\right)\right]\leq\widetilde{\beta}\widetilde{E}_x^{\pi}\left[e^{-\alpha T_{n-1}}W\left(\widetilde{x}_{n-1}\right)\right].
$$

Iterando esta última desigualdad obtenemos

$$\widetilde{E}_x^\pi \left\{e^{-\alpha T_n} W\left(\widetilde{x}_n\right)\right\} \leq \widetilde{\beta}^n W(x).$$

b) Por propiedades de la esperanza condicional tenemos que

$$\begin{aligned}
&\widetilde{E}_x^\pi \left\{e^{-\alpha T_n} u\left(\widetilde{x}_n\right)\right\} \\
&= \widetilde{E}_x^\pi \left[\widetilde{E}_x^\pi \left\{e^{-\alpha T_{n-1}} e^{-\alpha \delta_n} u\left(\widetilde{x}_n\right) \mid h_{n-1}, a_{n-1}\right\}\right] \\
&= \widetilde{E}_x^\pi \left[e^{-\alpha T_{n-1}} \widetilde{E}_x^\pi \left\{e^{-\alpha \delta_n} u\left(\widetilde{x}_n\right) \mid h_{n-1}, a_{n-1}\right\}\right] \\
&= \widetilde{E}_x^\pi \left[e^{-\alpha T_{n-1}} \int_{\mathbf{A}} \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} u(y) \widetilde{Q}\left(dy, dt \mid \widetilde{x}_{n-1}, a\right) \pi\left(da \mid h_{n-1}\right)\right] \\
&\leq \|u\|_W \widetilde{E}_x^\pi \left[e^{-\alpha T_{n-1}} \int_{\mathbf{A}} \widetilde{\beta} W\left(\widetilde{x}_{n-1}\right) \pi\left(da \mid h_{n-1}\right)\right] \\
&= \|u\|_W \widetilde{\beta} \widetilde{E}_x^\pi \left[e^{-\alpha T_{n-1}} W\left(\widetilde{x}_{n-1}\right)\right]; \\
&\leq W(x) \|u\|_W \widetilde{\beta}^n, \quad \forall x \in \mathbf{X}, \pi \in \Pi.
\end{aligned}$$

Esta última desigualdad se sigue del inciso anterior, la cual implica

$$\widetilde{E}_x^\pi e^{-\alpha T_n} u\left(\widetilde{x}_n\right) \leq W(x) \|u\|_W \widetilde{\beta}^n, \quad \forall \quad \widetilde{x} \in \mathbf{X}, \pi \in \Pi.$$

Los siguientes incisos se demuestran de manera análoga. □

Otra consecuencia de la Proposición 5.2 es que el proceso perturbado es regular, como lo establece el siguiente resultado.

Corolario 5.4. *Para cada política $\pi \in \Pi$ y estado inicial $\widetilde{x}_0 = \widetilde{x} \in \mathbf{X}$, la sucesión $T_n, n \in \mathbb{N}$, diverge a infinito $\widetilde{P}_x^\pi-$ casi seguramente.*

Demostración. Observemos que de la Proposición 5.3 a) se tiene que

$$\widetilde{E}_x^\pi \left\{e^{-\alpha T_n}\right\} \leq \widetilde{E}_x^\pi \left\{e^{-\alpha T_n} W\left(x_n\right)\right\} \leq \widetilde{\beta}^n W(x) \to 0,$$

lo cual implica que $T_n \to \infty \quad \widetilde{P}_x^\pi$-casi seguramente, para cada estado inicial $\widetilde{x} \in \mathbf{X}$. □

5.3 Modelo Semi-Markoviano perturbado truncado

En esta sección realizaremos un proceso de truncamiento para el modelo semi markoviano perturbado, el cual como ya se mencionó anteriormente, satisface condiciones análogas a las establecidas en las Hipótesis 3.1 y 3.2. De esta manera será posible presentar resultados análogos a los que obtuvimos al truncar el modelo semi markoviano original. Este proceso se llevará a cabo bajo la

Hipótesis 3.4 como se realizó en la Sección 3.6 del Capítulo 3. A continuación establecemos dicha hipótesis para una fácil referencia.

Hipótesis 3.4

a) Existe $\eta > 0$ tal que W^{η} es un momento en $\mathbf{X}$ donde W es la función medible de la Hipótesis 3.2;

b) además

$$\sup_{(x,a)\in\mathbb{K}} \frac{1}{W(x)} \int_{\mathbf{X}\times\mathbb{R}^{+}} e^{-at} W^{(1+\eta)}(y) Q(dy, dt \mid x, a) < \infty$$

Definición 5.2. *Sea $\mathbf{X}_{\epsilon}^{c}$ un conjunto análogo al descrito en (3.6.4) y definamos*

$$\widetilde{\tau} = \widetilde{\tau}(\epsilon) := \inf\left\{n : \widetilde{x}_n \in \mathbf{X}_{\epsilon}^{c}\right\},$$

donde $\inf \phi := +\infty$.

Definición 5.3. *Para $x \in \mathbf{X}_{\epsilon}, \pi \in \Pi$, se define*

$$\widetilde{V}_{\epsilon}(x,\pi) := \widetilde{E}_x^{\pi} \sum_{k=0}^{\infty} e^{-\alpha T_k} \widetilde{C}\left(\widetilde{x}_k, \widetilde{a}_k\right) I_{[k<\widetilde{\tau}]}$$

y la función de valor óptimo

$$\widetilde{V}_{\epsilon}^{*}(x) := \inf_{\pi\in\Pi} \widetilde{V}_{\epsilon}^{*}(x,\pi), \quad x \in \mathbf{X}_{\epsilon}.$$

Para cada $x \in \mathbf{X}_{\epsilon}, f \in \mathbb{F}$ y $u \in B_W\left(\mathbf{X}_{\epsilon}\right)$, se definen los operadores de programación dinámica truncados, de manera análoga a los operadores de programación (3.6.7) y (3.6.8),

$$\widetilde{T}_{\epsilon,f} u(x) := \widetilde{C}(x,f) + \int_{\mathbf{X}_{\epsilon}\times\mathbb{R}^{+}} e^{-\alpha t} u(y) \widetilde{Q}(dy, dt \mid x, f), \tag{5.3.1}$$

$$\widetilde{T}_{\epsilon} u(x) := \min_{a\in A} \left\{\widetilde{C}(x,a) + \int_{\mathbf{X}_{\epsilon}\times\mathbb{R}_{+}} e^{-\alpha t} u(y) \widetilde{Q}(dy, dt \mid x, a)\right\}. \tag{5.3.2}$$

Proposición 5.5. *Bajo las Hipótesis 3.1, 3.2, 3.4 y 5.1 se tiene que:*

a) *Los operadores $\widetilde{T}_{\epsilon,f}$ y $\widetilde{T}_{\epsilon}$ son operadores de contracción en el espacio de Banach $(B_W(\mathbf{X}_\epsilon), \|\cdot\|_W^\epsilon)$ con módulo $\widetilde{\beta}$. Además para cada función $u(\cdot)$ en $B_W(\mathbf{X}_\epsilon)$ existe un selector medible $f \in \mathbb{F}$ tal que*

$$\widetilde{T}_\epsilon u(x) = \widetilde{T}_{\epsilon,f} u(x).$$

b) *$\widetilde{V}_\epsilon(x,f)$ es el único punto fijo del operador $\widetilde{T}_{\epsilon,f}$, esto es,*

$$\widetilde{V}_\epsilon(x,f) = \widetilde{T}_{\epsilon,f}\widetilde{V}_f^\epsilon.$$

Demostración. **a)** La demostración de que $\widetilde{T}_\epsilon$ y $\widetilde{T}_{\epsilon,f}$ son operadores de contracción modulo $\widetilde{\beta}$ se hace de manera análoga a la demostración de la Proposición 3.5 b).

Notemos que de la Proposición 5.2 b) y c) tenemos que para cada $u \in B_W(\mathbf{X}_\epsilon)$ la función

$$a \to \quad \widetilde{C}(x,a) + \int_{\mathbf{X}\times\mathbb{R}^+} e^{-\alpha t} u(y)\widetilde{Q}(dy, dt \mid x, a)$$

es continua en $\mathbf{A}$ para cada $x \in \mathbf{X}_\epsilon$.

Entonces, de manera analoga a como se dijo en la Proposición 3.13 tenemos que por un teorema de selección medible ([10] Lema 8.3.8) Tu es una función medible y existe $f \in \mathbb{F}$ tal que

$$\widetilde{T}_\epsilon u(x) = \widetilde{C}_f(x) + \int_{\mathbf{X}_\epsilon\times\mathbb{R}^+} e^{-\alpha t} u(y)\widetilde{Q}_f(dy, dt \mid x) = \widetilde{T}_{\epsilon,f} u(x) \quad \forall x \in \mathbf{X}_\epsilon.$$

b) Notemos que por la Proposición 5.2 es posible mostrar que para todo $x \in \mathbf{X}_\epsilon, \quad f \in \mathbb{F}, u \in B_W(\mathbf{X}_\epsilon)$, y $n \in \mathbb{N}$,

$$\widetilde{T}_{\epsilon,f}^n u(x) = \widetilde{E}_x^f \sum_{k=0}^n e^{-\alpha T_k} \widetilde{C}_f(\widetilde{x}_k) I_{[k<\widetilde{\tau}]} + \widetilde{E}_x^f \left[e^{-\alpha T_n} u(\widetilde{x}_n) I_{[n<\widetilde{\tau}]}\right],$$

esto de manera análoga a como se realizó en la Proposición 3.14. Por lo que la demostración se realiza dando argumentos análogos a los dados en en la Proposición 3.14.

□

El resultado que presentamos a continuación da solución al problema de control óptimo del modelo Semi-Markoviano, perturbado-truncado.

Proposición 5.6. *Supongamos que se cumplen las Hipótesis 3.1, 3.2 y que L es un promediador. Entonces:*

a) *La función de valor óptimo $\widetilde{V}_\epsilon^*(\cdot)$ es el único punto fijo del operador $\widetilde{T}_\epsilon$ en $B_W(\mathbf{X}_\epsilon)$, es decir*

$$\widetilde{V}_\epsilon^*(x) = \widetilde{T}_\epsilon \widetilde{V}_\epsilon^*(x).$$

Además existe $f^* \in \mathbb{F}$ tal que

$$\widetilde{T}_\epsilon \widetilde{V}_\epsilon^*(x) = \widetilde{T}_{\epsilon,f^*}\widetilde{V}_\epsilon^*(x) = \widetilde{C}(x, f^*) + \int_{\mathbf{X}_\epsilon\times\mathbb{R}^+} e^{-\alpha t}\widetilde{V}_\epsilon^*(y)\widetilde{Q}(dy, dt \mid x, f^*).$$

b) *Una politica estacionaria $f^* \in \mathbb{F}$ es óptima si y sólo si*

$$\widetilde{T}_\epsilon \widetilde{V}_\epsilon^*(\cdot) = \widetilde{T}_{\epsilon,f^*} \widetilde{V}_\epsilon^*(\cdot).$$

Demostración. La demostración de este resultado es consecuencia de la Proposición 5.5 y se hace siguiendo argumentos análogos dados en la Proposición 3.15. □

Ahora con el fin de proporcionar cotas de error por aproximar la función de valor óptimo del modelo truncado V_ϵ^* por medio de la función de valor óptimo $\widetilde{V}_\epsilon^*$ del modelo perturbado-truncado, definimos las siguientes constantes de aproximación, las cuales nos permitirán medir la precisión de aproximación de operador L;

$$\gamma_C := \sup_{(x,a)\in\mathbb{K}_c} \frac{1}{W(x)} |\widetilde{C}(x,a) - C(x,a)|, \tag{5.3.3}$$

$$\gamma_Q := \sup_{(x,a)\in\mathbb{K}_\varepsilon} \frac{1}{W(x)} \|Q(\cdot,\cdot \mid x,a) - \widetilde{Q}(\cdot,\cdot \mid x,a)\|_W, \tag{5.3.4}$$

donde

$$\|Q(\cdot,\cdot \mid x,a) - \widetilde{Q}(\cdot,\cdot \mid x,a)\|_W = \sup_{\|u\|_W \le 1} \left| \int_{\mathbf{X}_\iota \times \mathbb{R}^+} e^{-\alpha t} u(y)[Q(dy,dt \mid x,a) - \tilde{Q}(dy,dt \mid x,a)] \right|.$$

Observemos que se cumplen resultados de aproximación análoagos a los obtenidos en las Proposiciones 3.11 y 3.16. El objetivo de esta sección es encontrar cotas de error por aproximar V_ϵ^* mediante la función $\widetilde{V}_\epsilon^*$.

Teorema 5.7. *Supongamos que se cumplen las Hipótesis 3.1, 3.2, 3.4 y 5.1. Entonces*

$$\left\| \widetilde{V}_\epsilon^* - V_\epsilon^* \right\|_W \le \frac{1}{1-\widetilde{\beta}} \gamma_C + \frac{\bar{c}}{(1-\beta)(1-\widetilde{\beta})} \gamma_Q.$$

Demostración. Observe que $\tilde{V}_\epsilon^*(x)$ y $V_\epsilon^*(x)$ son puntos fijos de los operadores $\widetilde{T}_\epsilon, T_\epsilon$ definidos en (5.3.2) y (3.6.8), respectivamente. Entonces para cada $x \in \mathbf{X}_\epsilon$ se verifica que

$$
\begin{aligned}
\left|\widetilde{V}_\epsilon^*(x) - V_\epsilon^*(x)\right| &= \left|\widetilde{T}_\epsilon \widetilde{V}_\epsilon^*(x) - T_\epsilon V_\epsilon^*(x)\right| \\
&\leq \sup_{a\in\mathbf{A}}\{|\widetilde{C}(x,a) - C(x,a)| \\
&+ \left|\int_{\mathbf{X}_c\times\mathbb{R}^+} e^{-\alpha t}\left(\widetilde{V}_\epsilon^* - V_\epsilon^*\right)(y)\widetilde{Q}(dy,dt \mid x,a)\right| \\
&+ \left|\int_{\mathbf{x}_\epsilon\times\mathbb{R}^+} e^{-\alpha t} V_\epsilon^*(y)(\widetilde{Q}(dy,dt\mid x,a) - Q(dy,dt\mid x,a)\right|\} \\
&\leq \sup_{a\in\mathbf{A}}\{|\widetilde{C}(x,a) - C(x,a)| \\
&+ \left\|\widetilde{V}_\epsilon^* - V_\epsilon^*\right\|_W^\epsilon \int_{\mathbf{x}_\epsilon\times\mathbb{R}^+} e^{-\alpha t} W(y)\widetilde{Q}(dy,dt\mid x,a) \\
&+ \|V_\epsilon^*\|_W^\epsilon \int_{\mathbf{x}_c\times\mathbb{R}^+} e^{-\alpha t}\frac{V_\epsilon^*(y)}{\|V_\epsilon^*\|_W^\epsilon}(\widetilde{Q}(dy,dt\mid x,a) - Q(dy,dt\mid x,a))\Big\} \\
&\leq \sup_{a\in\mathbf{A}}\Big\{|\widetilde{C}(x,a) - C(x,a)| + \left\|\widetilde{V}_\epsilon^* - V_\epsilon^*\right\|_W^\epsilon \widetilde{\beta} W(x) \\
&+ \|V_\epsilon^*\|_W^\epsilon \, \|Q(\cdot,\cdot\mid x,a) - \widetilde{Q}(\cdot,\cdot\mid x,a)\|_W\Big\}.
\end{aligned}
$$

Entonces,

$$
\begin{aligned}
\frac{\left|\widetilde{V}_\epsilon^*(x) - V_\epsilon^*(x)\right|}{W(x)} \leq \sup_{a\in\mathbf{A}}&\left\{\frac{|\widetilde{C}(x,a) - C(x,a)|}{W(x)} + \left\|\widetilde{V}_\epsilon^* - V_\epsilon^*\right\|_W^\epsilon \widetilde{\beta}\right. \\
&\left. + \|V_\epsilon^*\|_W^\epsilon \frac{1}{W(x)}\|Q(\cdot,\cdot\mid x,a) - \widetilde{Q}(\cdot,\cdot\mid x,a)\|_W\right\},
\end{aligned}
$$

de donde se sigue

$$
\left\|\widetilde{V}_\epsilon^* - V_\epsilon^*\right\|_W^\epsilon \leq \gamma_C + \left\|\widetilde{V}_\epsilon^* - V_\epsilon^*\right\|_W^\epsilon \widetilde{\beta} + \|V_\epsilon^*\|_W^\epsilon \gamma_Q.
$$

Por lo tanto

$$
\left\|\widetilde{V}_\epsilon^* - V_\epsilon^*\right\|_W^\epsilon \leq \frac{1}{1-\widetilde{\beta}}\gamma_C + \frac{\|V_\epsilon^*\|_W^\epsilon}{1-\widetilde{\beta}}\gamma_Q. \tag{5.3.5}
$$

Notemos que del Corolario 3.2 (b), la función V_ϵ^* satisface

$$
\|V_\epsilon^*\|_W^\epsilon \leq \frac{\bar{c}}{(1-\beta)}. \tag{5.3.6}
$$

por lo tanto, combinando las desigualdades (5.3.5) y (5.3.6) se obtiene

$$
\left\|\widetilde{V}_\epsilon^* - V_\epsilon^*\right\|_W^\epsilon \leq \frac{1}{1-\widetilde{\beta}}\gamma_C + \frac{\bar{c}}{(1-\beta)(1-\widetilde{\beta})}\gamma_Q
$$

□

Observación 5.8. *Notemos que del Teorema 5.7 y la Proposición 3.16 tenemos lo siguiente:*

$$
\begin{aligned}
\left\|V^* - \widetilde{V}_\epsilon^*\right\|_W^\epsilon &\leq \|V^* - V_\epsilon^*\|_W^\epsilon + \left\|V_\epsilon^* - \widetilde{V}_\epsilon^*\right\|_W^\epsilon \\
&\leq \frac{\bar{c}}{(1-\beta)^2}\epsilon + \frac{1}{1-\widetilde{\beta}}\gamma_C + \frac{\bar{c}}{(1-\beta)(1-\widetilde{\beta})}\gamma_Q.
\end{aligned}
$$

5.4 Aproximación de políticas óptimas

En esta sección aproximamos la función de valor óptimo $\widetilde{V}_\epsilon^*$ del modelo perturbado truncado. El objetivo principal es aproximar la función de valor óptimo V^* del modelo de control semi-markoviano original, para llevar esto a cabo introducimos un algoritmo de iteración de políticas, el cual se aplicará al modelo perturbado-truncado.

Considere $g_0 \in \mathbb{F}$ una política estacionaria con costo α-descontado $\widetilde{V}_\epsilon(x, g_0) < \infty$. Definamos $u_0(x) := \widetilde{V}_\epsilon(x, g_0)$, el cual, de manera análoga a como se demostró en (1.4.4) satisface

$$u_0(x) = \widetilde{C}_{g_0}(x) + \int e^{-\alpha t} u_0(y) \widetilde{Q}_{g_0}(dy, dt \mid x) \quad \forall x \in \mathbf{X}_\epsilon.$$

Sea $\widetilde{T}_\epsilon$ el operador de la programación dinámica definido en (5.3.2). Evaluamos $\widetilde{T}_\epsilon u_0(x)$, esto es

$$\widetilde{T}_\epsilon u_0(x) = \min_{a \in A(x)} \left[\widetilde{C}(x,a) + \int_{\mathbf{X}_\epsilon \times \mathbb{R}^+} e^{-\alpha t} u_0(y) \widetilde{Q}(dy, dt \mid x, a) \right],$$

luego por la Proposición 5.5 b) existe un selector medible $g_1 \in \mathbb{F}$ tal que

$$\widetilde{C}_{g_1}(x) + \int_{\mathbf{X}_\epsilon \times \mathbb{R}^+} e^{-\alpha t} u_0(y) \widetilde{Q}_{g_1}(dy, dt \mid x) = \min_{a \in A(x)} \left[\widetilde{C}(x,a) + \int_{\mathbf{X}_\epsilon \times \mathbb{R}^+} e^{-\alpha t} u_0(y) \widetilde{Q}(dy, dt \mid x, a) \right].$$

Ahora consideremos

$$u_1(x) = \widetilde{V}_\epsilon(x, g_1).$$

De manera general el algoritmo de iteración políticas aproximado-truncado consiste en lo siguiente

i) Inicio: sea $g_0 \in \mathbb{F}$ arbitraria y sea $n = 0$

ii) Evaluación: dado $g_n \in \mathbb{F}$ calcule $u_n(\cdot) := \widetilde{V}_\epsilon(\cdot, g_n)$

iii) Mejoramiento: encuentre $g_{n+1} \in \mathbb{F}$ tal que $\widetilde{T}_\epsilon u_n(\cdot) = \widetilde{T}_{\epsilon, g_{n+1}} u_n(\cdot)$ y regrese al paso (ii).

Proposición 5.9. *Supongamos que se satisfacen las Hipótesis 3.1, 3.2 y 5.1 y L es un promediador. Entonces, la sucesión $\{u_n(\cdot)\} \in B_W(\mathbf{X}_\epsilon)$ converge decrecientemente y uniformemente en la norma $\|\cdot\|_W^\epsilon$ a la función de valor óptimo $\widetilde{V}_\epsilon^*(\cdot)$; además, para todo $n \in \mathbb{N}$, se cumple que*

$$\left\| \widetilde{V}_\epsilon^* - u_n \right\|_W^\epsilon \leq \frac{2\widetilde{\beta}}{1 - \widetilde{\beta}} \|u_n - u_{n-1}\|_W^\epsilon . \tag{5.4.1}$$

Demostración. Veamos que la etapa de mejoramiento está bien definida. Para esto notemos que bajo las Hipótesis 3.1, 3.2 y 5.1 se tiene de la Proposición 5.5 b) que para cada función $u(\cdot)$ en $B_W(\mathbf{X}_\epsilon)$ existe un selector medible $f \in \mathbb{F}$ tal que

$$\widetilde{T}_\epsilon u(x) = \widetilde{T}_{\epsilon, f} u(x),$$

lo cual asegura la existencia de políticas en la etapa de mejoramieto.

Notemos ahora que para cada $u \in B_W(\mathbf{X}_\epsilon)$

$$\begin{aligned} LT_\epsilon u(x) =& L\left\{\min_{a\in A(x)}\left[C(x,a)+\int_{\mathbf{X}_\epsilon\times\mathbb{R}^+} e^{-\alpha t}u(y)Q(dy,dt\mid x,a)\right]\right\} \\ =& \min_{a\in A}\left\{\widetilde{C}(x,a)+\int_{\mathbf{X}_e\times\mathbb{R}^+} e^{-\alpha t}u(y)\widetilde{Q}(dy,dt\mid x,a)\right\} = \widetilde{T}_\epsilon u(x). \end{aligned}$$

Además por la Proposición 5.3 b), para cada $u \in B_W(\mathbf{X})$, se tiene que

$$\widetilde{E}_x^\pi\left\{e^{-\alpha T_n}u(\widetilde{x}_n)\right\} \le \widetilde{\beta}^n W(\widetilde{x})\|u\|_W.$$

De manera que la desigualdad (5.4.1) se puede demostrar de manera análoga a la demostración de la Proposición 4.8. □

A continuación enunciamos el resultado principal de esta sección el cual nos da una cota de error por aproximar la función de valor óptimo V^* del modelo origal por medio de la solución del modelo perturbado truncado. Esta cota tiene tres fuentes de error, un primer error es producido por truncar el espacio de estados, mientras que el segundo y tercer sumando son errores por aproximar con el promediador L; y el último sumando es el error por detener el algoritmo de iteración de políticas aproximado.

Teorema 5.10. *Supongamos que se cumplen las Hipótesis 3.1, 3.2 y 5.1, y sea* $u_n(\cdot) \in B_W(\mathbf{X}_\epsilon), g_n \in \mathbb{F}, n \in \mathbb{N}$, *las funciones y políticas definidas en el algoritmo de iteración de políticas aproximado truncado. Entonces*

$$\begin{aligned} \|V^*-V_{g_n}\|_W^\epsilon \le & \frac{2\bar{c}}{(1-\beta)^2}\epsilon \\ & + \frac{1}{1-\widetilde{\beta}}\gamma_C + \frac{\bar{c}}{(1-\beta)(1-\widetilde{\beta})}\gamma_Q \\ & + \frac{2\widetilde{\beta}}{1-\widetilde{\beta}}\|u_n-u_{n-1}\|_W^\epsilon. \end{aligned}$$

Demostración. Para cada $f \in \mathbb{F}$, tenemos que

$$\|V^*-V_f\|_W^\epsilon \le \left\|V^*-\widetilde{V}_\epsilon^*\right\|_W^\epsilon + \|V_{\epsilon,f}-V_f\|_W^\epsilon + \left\|\widetilde{V}_\epsilon^*-V_{\epsilon,f}\right\|_W^\epsilon.$$

Note que de la Observación 5.8 el primer término del lado derecho de la desigualdad está acotado por

$$\left\|V^*-\widetilde{V}_\epsilon^*\right\|_W^\epsilon \le \frac{\bar{c}}{(1-\beta)^2}\epsilon + \frac{1}{1-\widetilde{\beta}}\gamma_C + \frac{\bar{c}}{(1-\beta)(1-\widetilde{\beta})}\gamma_Q, \tag{5.4.2}$$

mientras que de la Proposición 3.16 tenemos

$$\|V_{\epsilon,f}-V_f\|_W^\epsilon \le \frac{\bar{c}\epsilon}{(1-\beta)^2}. \tag{5.4.3}$$

Ahora para el tercer sumando, observe que

$$\left\|\widetilde{V}_\epsilon^*-V_{\epsilon,f}\right\|_W^\epsilon \le \left\|\widetilde{V}_\epsilon^*-\widetilde{V}_{\epsilon,f}\right\|_W^\epsilon + \left\|\widetilde{V}_{\epsilon,f}-V_{\epsilon,f}\right\|_W^\epsilon. \tag{5.4.4}$$

Como las funciónes $\widetilde{V}_{\epsilon,f}$ y $V_{\epsilon,f}$ son puntos fijos de los operadores $\widetilde{T}_{\epsilon,f}$, y $T_{\epsilon,f}$, respectivamente, por argumentos análogos a la demostración del Teorema 5.7, el segundo sumando del lado derecho de (5.4.4) es acotado por

$$\left\|\widetilde{V}_{\epsilon,f} - V_{\epsilon,f}\right\|_W^\epsilon \leq \frac{1}{1-\widetilde{\beta}}\gamma_C + \frac{\bar{c}}{(1-\beta)(1-\widetilde{\beta})}.\gamma_Q \tag{5.4.5}$$

Ahora tomemos $f := g_n$ y $u_n = \widetilde{V}_{\epsilon,f}$. Por la Proposición 5.9, para el primer sumando del lado derecho de (5.4.4) tenemos

$$\left\|\widetilde{V}_\epsilon^* - \widetilde{V}_{\epsilon,g_n}\right\|_W^\epsilon = \left\|\widetilde{V}_\epsilon^* - u_n\right\|_W^\epsilon \leq \frac{2\widetilde{\beta}}{1-\widetilde{\beta}}\left\|u_n - u_{n-1}\right\|_W^\epsilon .$$

Así, de (5.4.5), (5.4.4) y la última desigualdad

$$\left\|\widetilde{V}_\epsilon^* - V_{\epsilon,f}\right\|_W^\epsilon \leq \frac{1}{1-\widetilde{\beta}}\gamma_C + \frac{\bar{c}}{(1-\beta)(1-\widetilde{\beta})}\gamma_Q + \frac{2\widetilde{\beta}}{1-\widetilde{\beta}}\left\|u_k - u_{k-1}\right\|_W^\epsilon . \tag{5.4.6}$$

Por lo tanto, combinando (5.4.2), (5.4.3) y (5.4.6), concluimos

$$\begin{aligned}\left\|V^* - V_f\right\|_W^\epsilon \leq &\frac{2\bar{c}}{(1-\beta)^2}\epsilon \\ &+ \frac{2}{1-\widetilde{\beta}}\gamma_C + \frac{2\bar{c}}{(1-\beta)(1-\widetilde{\beta})}\gamma_Q \\ &+ \frac{2\widetilde{\beta}}{1-\widetilde{\beta}}\left\|u_k - u_{k-1}\right\|_W^\epsilon .\end{aligned}$$

□

5.5 Ejemplo: aproximación del problema de reemplazo

En esta sección retomamos el ejemplo del sistema de remplazo descrito en la Sección 3.5 del Capítulo 3. Un primer objetivo es realizar un truncamiento uniforme del espacio de estados considerando las funciones W y r propuestas con anterioridad para el sistema de reemplazo. Además se establecerán condiciones bajo las cuales se obtienen estimaciones de las cotas de error γ_C y γ_Q.

Recordemos las siguientes variables para el sistema de reemplazo

i) t_n, $n \in \mathbb{N}$: tiempo en el que ocurre el n-ésimo choque;

ii) $l_n := t_n - t_{n-1}$: tiempos entre choques consecutivos;

iii) z_n: magnitud aleatoria de daño en el tiempo t_n, para cada $n \in \mathbb{N}_0$;

iv) x_n: daño acumulado hasta el tiempo t_n.

La dinámica del sistema de remplazo está descrita de acuerdo al kernel conjunto Q dado por

$$\begin{aligned}Q(B,[0,t]\mid x,a)&=\int_0^{\min(a,t)}H(ds)\int_{(x+z)\in B}r(x+z)J(dz\mid x)\\&+I_B(0)\int_0^{\min(a,t)}H(ds)\int_0^{\infty}(1-r(x+z))J(dz\mid x)\\&+I_B(0)I_{(t>a)}\{1-H(a)\},\quad B\in\mathcal{B}(\mathbf{X}),t\in\mathbb{R}^+,\end{aligned}$$

donde $r(\cdot)$ es una función de daño acumulado, $J(\cdot)$ la función de distribución del daño w_n, y $H(\cdot)$ la función de distribución de los tiempos entre choques consecutivos. Supondremos que ambas funciones tienen densidades j y h espectivamente, es decir

$$H(t)=\int_0^t h(s)ds,\quad J(y)=\int_0^y j(s)ds.$$

Sea $\epsilon>0$ fijo. Definimos $\mathbf{X}_\epsilon=[0,M]$ para alguna $M>0$ tal que

$$1-J(M)=1-\int_0^M j(s)ds<\epsilon.$$

Veamos que es posible obtener un truncamiento uniforme del espacio de estados $\mathbf{X}$, como se realizó en la Sección 3.6 del Capítulo 3 en el Lema 3.10, el cual establece la siguiente desigualdad

$$\int_{\mathbf{X}_\epsilon^c\times\mathbb{R}^+}e^{-\alpha t}W(y)Q(dy,dt\mid x,a)\le\epsilon W(x)\quad\forall(x,a)\in\mathbb{K}.\tag{5.5.1}$$

Para demostrar (5.5.1) recordemos que para cada $u\in\mathbf{M}_b(\mathbf{X})$, de (3.5.1) y considerando ahora el conjunto $\mathbf{X}_\epsilon^c$ en lugar de $\mathbf{X}$, se tiene que

$$\int_{\mathbf{X_\epsilon^c}\times\mathbb{R}^+}e^{-\alpha t}u(y)Q(dy,dt\mid x,a)=\int_0^a e^{-\alpha t}H(dt)\int_M^\infty u(x+z)r(x+z)J(dz\mid x).\tag{5.5.2}$$

Considerando $W(x)=e^{qx}$ y $r(x)=e^{-bx}$ con $b>q>0$ en (5.5.2) tenemos que para cada $x\in\mathbf{X}_\epsilon=[0,M]$

$$\begin{aligned}\int_{\mathbf{X}_c^c\times\mathbb{R}^+}e^{-\alpha t}W(y)Q(dy,dt\mid x,a)&=\int_0^a e^{-\alpha t}H(dt)\int_M^\infty e^{q(z+x)}e^{-b(z+x)}J(dz)\\&=\int_0^a e^{-\alpha t}H(dt)e^{qx}\int_M^\infty e^{-w(b-q)}e^{-bx}J(dz)\\&\le e^{qx}\int_M^\infty J(dz)\le e^{qx}(1-J(M))<W(x)\epsilon.\end{aligned}$$

Por lo tanto se tiene un truncamiento uniforme para el espacio de estados $\mathbf{X}$.

Para obtener estimaciones de γ_C y γ_Q es necesario establecer condiciones a la densidad $j(\cdot)$ y $r(\cdot)$.

Hipótesis 5.2. *Supongamos que $j(\cdot)$ y $r(\cdot)$ son Lipschitz continuas con módulo b_j and b_r, respectivamente, es decir,*

$$|j(x)-j(y)| \leq b_j|y-x|, |r(x)-r(y)| \leq b_r|y-x| \forall x, y \in [0, M].$$

Observemos que se cumple

$$\int_{\mathbf{X}_c\times\mathbb{R}^+} e^{-\alpha t}u(y)Q(dy, dt \mid x, a) = \int_0^a e^{-\alpha t}h(t)dt \int_0^{M-x} u(x+z)r(x+z)j(z)dz$$

$$+u(0)\int_0^a e^{-\alpha t}h(t)dt \int_0^{\infty}(1-r(x+z))j(z)dz$$

$$+ u(0)\int_a^{\infty} e^{-\alpha t}h(t)dt. \tag{5.5.3}$$

Para obtener las cotas de error, consideremos el operador de interpolación lineal.

Consideremos una partición para $\mathbf{X}_\epsilon = [0, M]$, con

$$s_1 = 0 < s_2 < \cdots < s_{k-1} < s_N = M \text{ y sea } \Delta_s := \frac{M}{N}.$$

Entonces para cada función medible u en $\mathbf{X}_\epsilon$, el operador L está definido para $x \in [s_i, s_{i+1}]$ como

$$Lu(x) = \frac{s_{i+1}-x}{s_{i+1}-s_i}u(s_i) + \frac{x-s_i}{s_{i+1}-s_i}u(s_{i+1}),$$

con $x \in [s_i, s_{i+1}], i = 1, 2, \cdots, N-1$.

Luego

$$\widetilde{Q}(\cdot, \cdot \mid x, a) = b(x)Q(\cdot, \cdot \mid s_i, a) + \bar{b}(x)Q(\cdot, \cdot \mid s_{i+1}, a),$$

donde

$$b(x) = \frac{s_{i+1}-x}{s_{i+1}-s_i} \text{ y } \bar{b}(x) = \frac{x-s_i}{s_{i+1}-s_i}$$

con $x \in [s_i, s_{i+1}], i = 1, 2, \cdots, N-1$.

Ahora notemos que

$$\begin{aligned}
&\int_{\mathbf{X}_c\times\mathbb{R}^+} e^{-\alpha t}u(y)[\widetilde{Q}(dy, dt \mid x, a) - Q(dy, dt \mid x, a)] \\
&= \int_{\mathbf{X}\epsilon\times\mathbb{R}^+} e^{-\alpha t}u(y)[b(x)Q(dy, dt \mid s_i, a) + \bar{b}(x)Q(dy, dt \mid s_{i+1}, a) - Q(dy, dt \mid x, a)] \\
&= b(x)\left\{\int_{\mathbf{X}_\epsilon\times\mathbb{R}^+} e^{-\alpha t}u(y)Q(dy, dt \mid s_i, a) - \int_{\mathbf{X}_\epsilon\times\mathbb{R}^+} e^{-\alpha t}u(y)Q(dy, dt \mid x, a)\right\} \\
&+ \bar{b}(x)\left\{\int_{\mathbf{X}_\epsilon\times\mathbb{R}^+} e^{-\alpha t}u(y)Q(dy, dt \mid s_{i+1}, a) - \int_{\mathbf{X}_\epsilon\times\mathbb{R}^+} e^{-\alpha t}u(y)Q(dy, dt \mid x, a)\right\}.
\end{aligned}$$

Considerando la igualdad (5.5.3), realizando el cambio de variable $y = x+z$ $y = s_i+z$, y observando que $W(y)r(y) = e^{y(q-b)} < 1$, $\|u\|_W \leq 1, u(0) \leq W(0) = 1$ es posible mostrar que se cumple

$$\gamma_Q \leq 2\left(|j| + Mb_j + b_r\right)\Delta s$$

y

$$\gamma_C \leq 2\left(K_1 b_r + K_3\right)\Delta s,$$

donde $\Delta s := \frac{M}{N}$.

Finalmente combinando estos resultados con el Teorema 5.10 obtenemos las siguientes cotas

$$\|V^* - V_{g_n}\|_W^\epsilon \leq e_1 + e_2 + e_3,$$

donde

$$\begin{aligned}
e_1 &:= \frac{2\bar{c}}{(1-\beta)^2}\epsilon,\\
e_2 &:= \frac{1}{1-\widetilde{\beta}}\gamma_C + \frac{\bar{c}}{(1-\beta)(1-\widetilde{\beta})}\gamma_Q,\\
e_3 &:= \frac{2\widetilde{\beta}}{1-\widetilde{\beta}}\|u_n - u_{n-1}\|_W^\epsilon.
\end{aligned}$$

Observemos que e_1 es el error por truncamiento, e_2 es el error por aproximación por usar el promediador L y e_3 es el error por detener el algoritmo de iteración de políticas, los cuales pueden hacerse arbitrariamente pequeños.

Apéndice A

Notación y terminologia

A lo largo del presente trabajo usaremos la siguiente notación:

- $\mathbb{R}$: conjunto de números reales
- $\mathbb{R}^+$: conjunto de los números reales positivos.
- $\mathbb{N}$: conjunto de los números naturales
- $\mathbb{N}_0$: conjunto de los números enteros no negativos

Un espacio de Borel, es un subconjunto de Borel de un espacio métrico separable y completo. Sea $\mathbf{X}$ un espacio de Borel.

- $\mathcal{B}(\mathbf{X})$: sigma-álgebra de Borel, esto es, la mínima sigma-álgebra que contiene la familia de subconjuntos abiertos de $\mathbf{X}$
- $\mathbf{M}(\mathbf{X})$: espacio de funciones $u : \mathbf{X} \to \mathbb{R}$ medibles
- $\mathbf{M}_b(\mathbf{X})$: subespacio de funciones medibles y acotadas
- $\mathbf{C}_b(\mathbf{X})$: espacio de funciones $u : \mathbf{X} \to \mathbb{R}$ continuas y acotadas.
- Para cada $v \in \mathbf{M}_b(\mathbf{X})$ la norma del supremo se define como

$$\|v\|_\infty := \sup_{x \in \mathbf{X}} |v(x)|$$

- Para cada $B \in \mathcal{B}(\mathbf{X})$, denotamos por I_B su función indicadora.

Dados dos espacios de Borel $\mathbf{X}$ y $\mathbf{Y}$, un kérnel estocástico $Q(\cdot \mid \cdot)$ en $\mathbf{X}$ dado $\mathbf{Y}$ es un mapeo tal que:

i) $Q(\cdot \mid y)$ es una medida de probabilidad en $\mathbf{X}$ para cada $y \in \mathbf{Y}$

ii) $Q(B \mid \cdot)$ es una función medible en $\mathbf{Y}$ para cada $B \in \mathcal{B}(\mathbf{X})$.

Apéndice B

Condición de selección medible

Definición B.1. *una multifunción* ψ *de* $\mathbf{X}$ *a* $\mathbf{A}$ *es una función tal que* $\psi(x)$ *es un subconjunto no vacío de* $\mathbf{A}$ *para cada* $x \in \mathbf{X}$. *La gráfica de una multifunción* ψ *es un subconjunto de* $\mathbf{X} \times \mathbf{A}$ *definido como*

$$\mathrm{Gr}(\psi) := \{(x, a) \mid x \in X, a \in \psi(x)\}$$

Para cada subconjunto B de $\mathbf{A}$, se define $\psi^{-1}[B] := \{x \in X \mid \psi(x) \cap B \neq \phi\}$.

Definición B.2. *Una multifunción* ψ *de* $\mathbf{X}$ *a* $\mathbf{A}$ *se dice que es*

a) Borel-medible si $\psi^{-1}[G]$ *es un subconjunto de Borel de* $\mathbf{X}$ *para cada conjunto abierto* $G \subset \mathbf{A}$

b) Semi continua superiormente si $\psi^{-1}[F]$ *es cerrado en* $\mathbf{X}$ *para cada conjunto cerrado* $F \subset \mathbf{A}$

c) semi continua inferiormente si $\psi^{-1}[G]$ *es abierto en* $\mathbf{X}$ *para cada conjunto abierto* $G \subset \mathbf{A}$

d) continua si es a la vez semi continua superiormente y semi continua inferiormente.

Proposición B.1. *Sea* $D : X \to A(X)$ *una multifunción Borel-medible, y sea* $v(x, a)$ *una función medible en* $\mathbb{K}$ *tal que* $v(x, a)$ *es semi continua superiormente en* $a \in D(x)$ *para cada* $x \in X$. *Entonces*

a) Existe un selector $f : X \to A$ *tal que*

$$v(x, f(x)) = \max_{a \in D(x)} v(x, a) \text{ for every } x \in X$$

y la función $v^*(x) := \max_{a \in D(x)} v(x, a)$ *es medible*

b) Si D *es semi continua superiormente y* v *es semi continua superiormente y acotada entonces* v^* *es semi continua superiormente y acotada*

c) Si D *es continua y* v *es continua y acotada, entonces* v^* *es continua y acotada*

Demostración. Himmelberg et al. (1976); Schäl (1975); Bertsekas and Shreve (1978). □

Apéndice C

Cotas de aproximación para el sistema de inventario

Para calcular a detalle las cotas de aproximación para el sistema de inventario en el Capítulo 4 Sección 4.5, trabajaremos bajo la Hipótesis 4.1 y asumiremos que $p > c + h$. Además definimos

$$U(y) := \int_0^y F(z)dz, \quad y \geq 0.$$

Recordemos la función de costo (4.5.2)

$$C(x, a) = p\bar{\xi} + ca + (h - p)(x + a) + p\int_0^{x+a} F(\xi)d\xi.$$

Entonces, para cada política de umbral f_S se tiene que la función de costo toma la siguiente forma

$$C_{f_S}(x) = \begin{cases} -cx + p\bar{\xi} + (c + h - p)S + pU(S) & \text{si } 0 \leq x \leq S \\ p\bar{\xi} + (h - p)x + pU(x) & \text{si } S < x \leq \mathcal{K}. \end{cases}$$

Además, si el umbral $S \in [s_i, s_{i+1}]$, entonces la interpolación lineal de $C_{f_S}(\cdot)$ toma la forma

$$\widetilde{C}_{f_S}(x) = \begin{cases} -cx + p\bar{\xi} + (c + h - p)S + pU(S) & \text{si } 0 \leq x \leq s_i \\ b_i(x)C_{f_S}(s_i) + \bar{b}_i(x)C_{f_S}(s_{i+1}) & \text{si } s_i < x \leq s_{i+1} \\ p\bar{\xi} + (h - p)x + p\widetilde{U}(x) & \text{si } s_{i+1} < x \leq \mathcal{K}, \end{cases}$$

donde $\widetilde{U}(x)$es la función de interpolación lineal de $U(\cdot)$.

Consideraemos una política f_S y supongamos que $S \in [s_i, s_{i+1}]$. Entonces se obtienen los siguientes resultados

$$\begin{aligned} \left|\widetilde{C}_{f_S}(x) - C_{f_S}(x)\right| &= 0 \quad \forall x \in [0, s_i] \\ \left|\widetilde{C}_{f_S}(x) - C_{f_S}(x)\right| &\leq (2c + h)\Delta \quad \forall x \in [s_i, S] \\ \left|\widetilde{C}_{f_S}(x) - C_{f_S}(x)\right| &\leq (c + 2h)\Delta \quad \forall x \in [S, s_{i+1}] \\ \left|\widetilde{C}_{f_S}(x) - C_{f_S}(x)\right| &\leq 2p\Delta \quad \forall x \in [s_{i+1}, \mathcal{K}]. \end{aligned}$$

- Si $x \in [0, s_i]$ es claro que

$$|\widetilde{C}_{fS}(x) - C_{f_S}(x)| = 0$$

- Si $x \in [s_i, S]$ entonces

$$|\widetilde{C}_{fS}(x) - C_{f_S}(x)| = |b_i(x)[C_{f_S}(s_i) - C_{f_S}(x)] + \bar{b}i(x)[C_{f_S}(s_{i+1}) - C_{f_S}(x)]|.$$

Notemos que

$$\begin{aligned} C_{f_S}(s_i) - C_{f_S}(x) &= -cs_i + p\bar{\xi} + (c+h-p)S + pU(S) + cx - p\bar{\xi} - (c+h-p)S - pU(S) \\ &= c(x - s_i) < c\Delta. \end{aligned}$$

Por otro lado

$$\begin{aligned} C_{f_S}(s_{i+1}) - C_{f_S}(x) &= p\bar{\xi} + (h-p)s_{i+1} + pU(s_{i+1}) + cx - p\bar{\xi} - (c+h-p)S - pU(S) \\ &= c(x - S) + (h-p)(s_{i+1} - S) + +p(U(s_{+1}) - U(S)). \end{aligned}$$

Notemos que

$$U(s_{+1}) - U(S) = \int_S^{s_{i+1}} F(z)dz < \Delta \tag{C.0.1}$$

Entonces

$$C_{f_S}(s_{i+1}) - C_{f_S}(x) < c\Delta + (h-p)\Delta + p\Delta.$$

Por lo tanto

$$|\widetilde{C}_{fS}(x) - C_{f_S}(x)| < 2c\Delta + (h-p)\Delta + p\Delta = (2c+h)\Delta.$$

- Si $x \in [S, s_{i+1}]$ entonces

$$|\widetilde{C}_{fS}(x) - C_{f_S}(x)| = |b_i(x)[C_{f_S}(s_i) - C_{f_S}(x)] + \bar{b}_i(x)[C_{f_S}(s_{i+1}) - C_{f_S}(x)]|.$$

Notemos que

$$\begin{aligned} C_{f_S}(s_i) - C_{f_S}(x) &= -cs_i + p\bar{\xi} + (c+h-p)S + pU(S) - p\bar{\xi} - (h-p)x - pU(x) \\ &= c(S - s_i) + (h-p)(S - x) + P(U(S) - U(x)) \\ &< c\Delta + (h-p)\Delta + p[\int_o^S F(z)dz - \int_o^x F(z)dz]. \end{aligned}$$

Por otro lado

$$\begin{aligned} C_{f_S}(s_{i+1}) - C_{f_S}(x) &= p\bar{\xi} + (h-p)s_{i+1} + pU(s_{i+1}) - p\bar{\xi} - (h-p)x - pU(x) \\ &= (h-p)(s_{i+1} - x) + p(U(s_{i+1}) - U(x)) \\ &< (h-p)\Delta + p[\int_0^{s_{i+1}} F(z)dz - \int_0^x F(z)dz]. \end{aligned}$$

Entonces

$$\begin{aligned} \widetilde{C}_{fS}(x) - C_{f_S}(x) &= b_i(x)[C_{f_S}(s_i) - C_{f_S}(x)] + \bar{b}_i(x)[C_{f_S}(s_{i+1}) - C_{f_S}(x)] \\ &\leq b_i(x)[c\Delta + (h-p)\Delta + p[\int_o^S F(z)dz - \int_o^x F(z)dz]] \\ &+ b_i\bar{(}x)[(h-p)\Delta + p[\int_0^{s_{i+1}} F(z)dz - \int_0^x F(z)dz]] \\ &= b_i(x)[c\Delta + (h-p)\Delta + p(-\int_s^x F(z)dz)] \\ &+ b_i\bar{(}x)[(h-p)\Delta + p\int_x^{s_{i+1}} F(z)dz]. \end{aligned}$$

Entonces

$$\begin{aligned}
|\widetilde{C}_{fS}(x) - C_{f_S}(x)| &= |b_i(x)[c\Delta + (h-p)\Delta + p(-\int_s^x F(z)dz)] + \bar{b_i(x)}[(h-p)\Delta + p\int_x^{s_{i+1}} F(z)dz]| \\
&\leq |c\Delta + (h-p)\Delta + p(-\int_s^x F(z)dz)]| + |(h-p)\Delta + p\int_x^{s_{i+1}} F(z)dz| \\
&\leq |c\Delta + (h-p)\Delta| + |p(-\int_s^x F(z)dz)| + (h-p)\Delta + p\Delta \\
&< c\Delta + 2(h-p)\Delta + 2p\Delta = (c+2h)\Delta.
\end{aligned}$$

- Finalmente, si $x \in [s_{i+1}, \mathcal{K}]$ entonces

$$\begin{aligned}
|\widetilde{C}_{fS}(x) - C_{f_S}(x)| &= |p\bar{\xi} + (h-p)x + p\widetilde{U}(x) - p\bar{\xi} - (h-p)x - pU(x)| \\
&= |p(\widetilde{U}(x) - U(x))| = |p[b_i(x)U(s_i) + \bar{b}_i(x)U(s_{i+1}) - U(x)]| \\
&= |p[b_i(x)(U(s_i) - U(x)) + \bar{b_i(x)}(U(s_{i+1}) - U(x))]| \\
&< |pb_i(x)(-\int_{s_i}^x F(z)dz)| + |p\bar{b_i(x)}(-\int_{s_{i+1}}^x F(z)dz)| \\
&< 2p\Delta.
\end{aligned}$$

Bibliografia

[1] Maria Teresa Robles Alcaraz. *Algoritmo de Iteración de Políticas Aproximado en Modelos Markovianos y Semi-Markovianos bajo Criterio de Costo Descontado.* Universidad de Sonora, Departamento de Matemáticas, 2016.

[2] Ash Robert B. *Real Analysis and Probability.* Academic Press, 1972.

[3] Tsitsiklis JN Bersetkas DP. *Neuro-Dynamic Programming. Solving the curse of Dimensionality.* Athena Scientific, Belmont, MA, 1995.

[4] Lewis ME Cooper WL, Henderson SG. *Convergence of simulation-based policy iteration, Probability in the Engineering and Information Sciences.* 17, 2003.

[5] Bersetkas DP. *Approximate policy iteration: a survey and some new methods, Journal of Control Theory and Applications.* 9, 2011.

[6] K. Hinderer. *Foundations of Non-Stationary Dynamic Programming with Discrete Time Parameter.* SpringerVerlag, 1970.

[7] Rust J. *Numerical dynamic programming in economics, in: Handbook of Computational Economics, Vol 1.* H.M. Amman, D.A. Kendrick, J. Rust, J. eds., Elsevier, 1996.

[8] López Borbón Joaquín. *Aproximación Numérica para Iteración de Valores en Procesos de Control Markoviano con Espacios Generales y Costos Acotados.* Universidad de Sonora, Departamento de Matemáticas, 2015.

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

[10] O.Hernández-Lerma and J.B. Lasserre. *Farther Topics on DiscreteTime Markov Control Processes.* Springer-Verlag ,New York, 1999.

[11] Yudi Pawitan. *In All Likelihhood.* Oxford Science Publications, Great Britain, first edition, 2001.

[12] Ma J Powell WB. *A review of stochastic algorithms with continuous value funtion approximation and some new approximate policy iteration algorithms for multidimensional continuous applications, Journal of Control Theory and Applications.* 9, 2011.

[13] U. Rieder. *Measurable selection theorems for optimization problems.* nuscripta Math, 1978.

[14] S.M. Ross. *Applied Probability Models with Optimization Applications,.* Holden-Day, 1970.

[15] Robert J. Serfling. *Approximation Theorems of Mathematical Statistics.* John Wiley and Sons, 1980.

[16] Vega-Amaya and R. Montes de Oca. *Application of average dynamic programming to inventory systems, Mathematical Methods of Operations Research.* Research 47, 1998.

[17] Lopez-Borbón J Vega-Amaya O. *A Perturbation approach for a class of discounted approximate value iteration algorithms,*. JournaL of Dynamics and Games, 2016.

[18] Robles-Alcaráz MT Vega-Amaya O, Minjarez-Sosa JA. *Estimate and approximate policy iteration algorithm for discounted Markov decision models with bounded cost and Borel spaces.* Risk Decision Analysis, 6, 2, 2017.

[19] Powell WB. *Approximate Dynamic Programming. Solving the Curse of Dimensionality.* John Wiley Sons Inc, 2007.

[20] Powell WB. *Perspectives of approximate dynamic programming.* Ann. Oper. Res., 2012.