Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias Físico Matemáticas
Postgrado en Ciencias Matemáticas

Procesos de Decisión de Markov Sensibles al Riesgo: Propiedad de Continuidad y caso Semi Markoviano

Tesis

Que para obtener el título de
Doctor en Ciencias (Matemáticas)

Presenta

María Selene Georgina Chávez Rodríguez

Directores de Tesis

Dr. Rolando Cavazos Cadena
Dr. Hugo Adán Cruz Suárez

Puebla, Pue. Octubre 2016

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo prestado para realizar este trabajo.

Introducción

En este trabajo de tesis se presentan resultados relacionados con problemas de control estocástico [4], [23], [25], [33], [36], [39], los cuales estudian sistemas dinámicos cuyo comportamiento se encuentra determinado de acuerdo con las decisiones que toma un controlador.

En particular, se considera un Proceso de Decisión Semi Markoviano (PDSM), el cual tiene la siguiente dinámica: al tiempo t el sistema comienza en un estado x y el controlador elige una acción (control) a con las implicaciones siguientes:

1. Se incurre en un costo $C(x, a)$ por elegir el control a.
2. El sistema permanece en el estado x por un tiempo aleatorio determinado a partir de una función de distribución $F(\cdot)$.
3. Se genera un costo con tasa ρ, el cual depende del tiempo que el sistema permanezca en x.

Una vez transcurrido el tiempo de permanencia, el sistema hace una transición a un nuevo estado z de acuerdo con una probabilidad dada, y la dinámica anterior es repetida. En particular, cuando la función de distribución $F(\cdot)$ descrita en 2 es constante e igual 1, el PDSM se denomina Proceso de Decisión de Markov (PDM).

En este trabajo de tesis se estudian los procesos antes descritos con la característica de que el controlador tiene un coeficiente de sensibilidad al riesgo $\lambda \in \mathbb{R}$, es decir, en lugar de evaluar la eficiencia de las políticas mediante la esperanza de un costo acumulado, ésta se lleva a cabo a través de la esperanza de una función de utilidad [40] y, por tanto, se considera al criterio de costo promedio λ-sensible al riesgo.

El criterio de rendimiento promedio sensible al riesgo para los PDM fue inicialmente trabajado por Howard & Matheson [26]. En dicho trabajo se estudian modelos finitos y comunicantes a través de la teoría de matrices positivas de Perron-Frobenius. En este contexto, una política estacionaria óptima y su correspondiente costo promedio se obtuvieron de una solución a la ecuación de optimalidad promedio sensible al riesgo; la misma técnica, pero considerando el caso de modelos finitos con una estructura de comunicación general es presentada en [35].

Una manera de analizar a los PDM sensibles al riesgo es mediante el uso de la técnica conocida como «Factor de Descuento Desvaneciente» [2], [24], [33] la cual está basada en operadores contracción cuyos puntos fijos son usados para generar aproximaciones que convergen a la solución de la ecuación de optimalidad. Bajo este contexto, en [21], donde además se usaron resultados de teoría de juegos, se estudió el caso con espacio de estados numerable. Nuevamente, usando la técnica de factor de descuento desvaneciente, en [12] se consideró al criterio de costo total y el caso de espacios de Borel fue estudiado en [18, 19, 20] y [28]. En [5] se considera una función de utilidad general y aplicaciones de PDM en problemas financieros son presentados en [4]. En particular, en [4] se demostró que, cuando se trabaja con costos positivos y se considera a la función de utilidad potencia dada por $U(x) = x^{\gamma}$ con $0 < \gamma < 1$, el criterio promedio óptimo sensible al riesgo coincide con el criterio óptimo neutral al riesgo. Algo similar es probado en [15], pero ahora considerando a la función de utilidad logarítmica, $U(x) = log(x)$. La igualdad entre los criterios óptimos sensibles y neutral al riesgo mostrada por estos autores fue determinante para establecer la motivación de la primera parte del trabajo de tesis: proponer condiciones bajo las cuáles es posible asegurar la continuidad del índice promedio con respecto al índice de sensibilidad al riesgo, λ. La idea detrás de esto es el hecho de que el caso neutral al riesgo coincide con el caso $\lambda = 0$, mientras que el sensible al riesgo ocurre cuando $\lambda \neq 0$.

Por otro lado, a pesar de que el índice promedio para los procesos de decisión semi markovianos ha sido ampliamente estudiado bajo el supuesto de que el controlador es neutral al riesgo y ha sido usado en múltiples aplicaciones, como por ejemplo, en el estudio de sistemas de colas [36], [37] y [39], procesos de producción [32], o problemas de mantenimiento [27], el caso sensible al riesgo apenas está comenzando a ser explorado. Por lo tanto, en la literatura no se contaba con una caracterización del índice promedio sensible al riesgo óptimo a través de una ecuación de optimalidad, un hecho que proporciona la motivación para la segunda parte de este trabajo. Los resultados presentados extienden las conclusiones obtenidas recientemente en [10], donde se estudian las cadenas semi markovianas no controladas.

En virtud de lo analizado anteriormente y en relación con el desarrollo de los procesos de decisión sensibles al riesgo, se plantean los problemas siguientes:

Problema 1: *Para el caso de los procesos de decisión de Markov: Determinar condiciones sobre la ley de transición asegurando que para cada estado, la función de costo promedio óptima es una función continua con respecto al parámetro de sensibilidad al riesgo.*

Problema 2: *Para los procesos de decisión Semi markovianos: Determinar una ecuación de optimalidad cuya solución caracterice a la función de costo promedio óptima y produzca una política óptima.*

El primero de estos problemas es desarrollado a lo largo de los capítulos 1 y 2 de esta tesis y el segundo se presenta en los capítulos 3 y 4. Además, tales resultados se encuentran basados en las referencias [16] y [17], respectivamente.

En relación con el Problema 1, se tienen las hipótesis siguientes: el espacio de estados es numerable y la función de costo es acotada. En este contexto, el objetivo es establecer condiciones para garantizar la continuidad de la función de valor óptimo con respecto al parámetro de sensibilidad al riesgo. Como primer paso para dar solución a este problema se demuestra bajo el supuesto de que la función de costo es acotada, que la función $\lambda \mapsto J^*(\lambda, \cdot)$ es continua en cada valor no nulo de λ, sin embargo, para el caso $\lambda = 0$ no se puede garantizar tal hecho bajo el único supuesto de costos acotados, por ello es que se considera, además de las condiciones estándar de continuidad-compacidad, otro requisito estructural utilizado en el trabajo es una forma general de la condición simultánea de Doeblin, bajo la cual una política estacionaria dada puede tener varias clases de recurrencia, pero en tal caso es posible que la cadena pueda transitar de una clase a otra bajo la acción de una política diferente; ver la Suposición 2.1.2. En este contexto, el resultado principal del trabajo, el Teorema 2.4.2, establece que para cada estado x, el costo promedio óptimo λ-sensible; $J^*(\lambda, x)$, es una función continua del coeficiente de sensibilidad al riesgo en $\lambda = 0$; además, se dan ejemplos para demostrar que (i) el resultado de continuidad no es válido para una ley de transición general, y (ii) si la Suposición 2.1.2 se sustituye por la condición de Lyapunov dada en la Suposición 2.1.3, la cual es un requisito de comunicación más débil, entonces la continuidad de la función $\lambda \mapsto J^*(\lambda, \cdot)$ no puede garantizarse.

En referencia al Problema 2, bajo el supuesto de que el espacio de estados es finito, se logra establecer una ecuación de optimalidad, la cual permite caracterizar al costo promedio óptimo. La solución al Problema 2 se encuentra expuesta en el Capítulo 4, en específico véase Teorema 4.3.1 y el Teorema 4.3.4, los cuales pueden ser descritos como sigue:

(i) Bajo el supuesto de que la ecuación de optimalidad (4.4) tiene una solución, la función de costo promedio λ-sensible es constante y será determinada de manera inmediata, además se garantizará la existencia de una política óptima estacionaria. En adelante, se referirá a este problema como el resultado de verificación.

(ii) Si además de la Suposición 4.1.1, se impone una condición sobre las propiedades de comunicación del proceso de control, entonces es posible asegurar la existencia de una solución a la ecuación de optimalidad (4.4). Este problema será referido como el resultado de existencia.

La metodología usada para establecer los problemas planteados arriba es la siguiente: usando la propiedad de que la distribución del número de transiciones que han ocurrido en un intervalo dado $[0, t]$, $t \geq 0$, decae más rápido que cualquier sucesión geométrica será usada para establecer la prueba del resultado de verificación. Por otro lado, el resultado de existencia será probado combinando (i) los resultados conocidos acerca de la existencia de soluciones a la ecuación de optimalidad sensible al riesgo en el caso de tiempo discreto, y (ii) el Teorema del Valor Intermedio para una función continua definida en un intervalo de la recta real.

La organización de la tesis es la siguiente: en el Capítulo 1 se consideran a los procesos de decisión de Markov sensibles al riesgo y se dan algunos resultados que serán de utilidad para resolver el Problema 1. En el Capítulo 2 se presenta el resultado que asegura la continuidad de la función de valor óptimo sobre el parámetro de sensibilidad al riesgo en el contexto markoviano. En el Capítulo 3 se presenta una breve introducción a los procesos de decisión semi markovianos. En el Capítulo 4, los procesos de decisión semi markovianos sensibles al riesgo serán estudiados, comenzando con el aporte de la ecuación de optimalidad asociada al modelo. Además, se presenta un ejemplo que sirve para mostrar la aplicación de la teoría desarrollada. Finalmente, se presenta una sección de conclusiones.

ÍNDICE GENERAL

Capítulo 1

Procesos de Decisión de Markov y Sensibilidad al Riesgo

En este capítulo se presentan a los Procesos de Decisión de Markov [2], [23], [39], además de definir lo que significa que el controlador sea sensible al riesgo. A través de una función de utilidad se definen las diversas actitudes ante el riesgo que puede presentar un controlador, y considerando estas actitudes se presenta al criterio de rendimiento utilizado a lo largo de los dos primeros capítulos de este trabajo de tesis.

1.1. El Modelo de Decisión de Markov

Definimos un Modelo de Decisión de Markov mediante la quíntupla siguiente:

$$\mathcal{M} = (S, A, \{A(x), x \in S\}, C, P)$$

donde

- El espacio de estados S es un conjunto numerable dotado de la topología discreta.
- El conjunto de acciones o controles A es un espacio métrico.
- Para cada estado $x \in S$, $A(x) \subset A$ es un conjunto no vacío, denominado espacio de acciones admisibles a x; el conjunto

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

  es la clase de parejas estado-acción admisibles.

- Por otro lado, $C \in \mathbb{B}(\mathbb{K})$ es la función de costo por etapa, donde $\mathbb{B}(\mathbb{K})$ denota a la clase de las funciones medibles y acotadas definidas en $\mathbb{K}$.
- $P = [p_{x\,y}(\cdot)]$ es la ley de transición sobre S dado $\mathbb{K}$, es decir, para cada $(x, a) \in \mathbb{K}$ y $z \in S$, $p_{xz}(a) \geq 0$ y $\sum_{y \in S} p_{xy}(a) = 1$.

Este modelo es interpretado como sigue: en cada tiempo $t \in \mathbb{N}$ el controlador observa el estado del sistema, digamos $X_t = x \in S$, y elige una acción (control) $A_t = a \in A(x)$. Entonces, un costo $C(x, a)$ es incurrido y, sin importar la historia del proceso hasta el tiempo t (ver (1.1)), el estado del sistema al tiempo $t + 1$ será $X_{t+1} = y \in S$ con probabilidad $p_{xy}(a)$.

Suposición 1.1.1 *(i) Para cada $x \in S$, $A(x)$ es un subconjunto compacto de A.*

(ii) Para cada $x, y \in S$, las funciones $a \mapsto C(x, a)$ y $a \mapsto p_{x\,y}(a)$ son continuas en $a \in A(x)$.

Políticas. El espacio $\mathbb{H}_t$ de posibles historias hasta el tiempo $t \in \mathbb{N}$ es definido por

$$\begin{aligned} \mathbb{H}_0 &:= S, \\ \mathbb{H}_t &:= \mathbb{K}^t \times S, \ t \geq 1. \end{aligned} \tag{1.1}$$

Un elemento de $\mathbb{H}_t$ es denotado por

$$h_t = (x_0, a_0, \ldots, x_i, a_i, \ldots, x_t),$$

donde $x_i \in S$ y $a_i \in A(x_i)$, $i = 0, 1, 2, \ldots, t$.

Una política $\pi = \{\pi_t\}$ es una sucesión de kérneles estocásticos, es decir, para cada $t \in \mathbb{N}$ y $h_t \in \mathbb{H}_t$, $\pi_t(\cdot|h_t)$ es una medida de probabilidad sobre A concentrada en $A(x_t)$, y para cada subconjunto de Borel $B \subset A$, la función $h_t \mapsto \pi_t(B|h_t)$, $h_t \in \mathbb{H}_t$, es Borel medible. La clase de todas las políticas es denotada por $\mathcal{P}$.

Una política $\pi \in \mathcal{P}$ será estacionaria si existe $f \in \mathbb{F} := \{f : S \to A \mid f(x) \in A(x)\}$. Observe que, $\mathbb{F} \subset \mathcal{P}$.

Dada la política $\pi \in \mathcal{P}$ usada para elegir acciones y el estado inicial $X_0 = x$, la distribución del proceso estado-acción $\{(X_t, A_t)\}$ está determinada de manera única [2], [23], [33]; tal distribución y su correspondiente operador esperanza son denotados por P_x^π y E_x^π, respectivamente. Además, para cada $C \in \mathcal{B}(A)$, $B \in \mathcal{B}(S)$ (donde $\mathcal{B}(\cdot)$ denota a la σ-álgebra de Borel), y $h_t \in \mathbb{H}_t$ y $t = 0, 1, 2, \ldots,$ se tiene que

$$\begin{aligned} P_x^\pi\,(X_0 = x) &= 1, \\ P_x^\pi(\,a_t \in C|\,h_t) &= \pi_t(C\,|h_t\,), \\ P_x^\pi(\,x_{t+1} \in B|\,h_t, a_t) &= p(B\,|x_t, a_t\,). \end{aligned}$$

1.2. Problema de Control Sensible al Riesgo

Se sabe que un individuo (consumidor) debe enfrentar situaciones en las que tendrá que elegir entre diversas alternativas, y dichas alternativas pueden no estar representadas de manera que sea fácil poder decidir cuál de ellas es la mejor. Debido a la necesidad de representar las preferencias de un individuo (consumidor), en Economía se desarrolla el concepto de *función de utilidad*. Formalmente, la teoría de utilidad representa de manera numérica las preferencias de un individuo y, es gracias a los trabajos de J. von Neumann y O. Morgenstern [40] que es posible representar preferencias bajo condiciones de incertidumbre y en tal caso se hablará de funciones de utilidad esperada.

Formalmente, se tiene la siguiente definición

Definición 1.2.1 *Una función $U : \text{dom } U \to \mathbb{R}$ es llamada una* **función de utilidad***, si U es estrictamente creciente y continua sobre dom U.*

Suponemos que un consumidor debe elegir entre varias alternativas y que conoce la probabilidad asignada a cada una de ellas, en este caso las preferencias de un consumidor se hacen bajo condiciones de incertidumbre y por tanto, en este caso se puede hablar de una función de utilidad esperada.

Sea $s \in \hat{X}$, donde $\hat{X}$ denota al conjunto de todas las posibles alternativas que puede elegir el consumidor y suponga que $\hat{X}$ es numerable. Denotamos por $\hat{p}(s)$ a la probabilidad de ocurrencia del estado s. Se tiene que se deben cumplir las siguientes propiedades

1. $\hat{p}(s) \geq 0$.
2. $\sum_{s\in\hat{X}} \hat{p}(s) = 1$.

Formalmente se escribirá el valor esperado de la utilidad de un consumidor de la siguiente forma:

$$E[U(s)] = \sum_{s\in\hat{X}} \hat{p}(s)\, U(s)\,.$$

Sea Y un costo aleatorio, se define el ***equivalente seguro*** como el número real $\mathcal{E} \equiv \mathcal{E}_\lambda(Y)$, el cual satisface que

$$U(\mathcal{E}) = E(U(Y))\,, \tag{1.2}$$

así que el consumidor será indiferente entre pagar la cantidad conocida $\mathcal{E}$ o incurrir en un costo aleatorio Y. El equivalente seguro estará bien definido cuando Y sea acotada y U una función continua y estrictamente creciente. Además, la función de utilidad es única salvo constantes, es decir, la relación entre dos utilidades esperadas no cambia cuando U es reemplazada por $U_1 = aU + b$ donde $a > 0$ y $b \in \mathbb{R}$.

Se sabe que los individuos no presentan actitudes ante el riesgo de la misma manera que otros, por lo tanto las decisiones sobre las preferencias pueden depender completamente de la actitud al riesgo de cada individuo, en tal caso se hablará de la sensibilidad al riesgo del consumidor. Es a través de la función de utilidad U que será posible medir la sensibilidad al riesgo que tiene un consumidor. Sea Y una variable aleatoria no constante, en adelante, se dirá que un consumidor es:

- ***Neutral al riesgo*** si $\mathcal{E}_\lambda(Y) = E(Y)$.
- ***Propenso al riesgo*** si $\mathcal{E}_\lambda(Y) < E(Y)$.
- ***Averso al riesgo*** si $\mathcal{E}_\lambda(Y) > E(Y)$.

Definimos a la prima de riesgo correspondiente a Y como la cantidad dada por:

$$\Delta(Y) = \mathcal{E}_\lambda(Y) - E[Y],$$

dicha cantidad sirve para determinar el grado de aversión al riesgo del consumidor [31].

El resultado dado a continuación permite observar que la prima de riesgo es proporcional a la varianza del costo aleatorio Y, y que la constante de proporcionalidad está dada en función de un cociente de derivadas de la función de utilidad. La prueba de dicho resultado está basada en propiedades de la derivada, la igualdad dada en (1.2) y el uso de la expansión en series de Taylor de la función de utilidad.

Proposición 1.2.2 *Supóngase que la función de utilidad U tiene derivada continua de orden 2 sobre la línea real y que $U' > 0$. Entonces*

$$\frac{\Delta(Y(\sigma))}{\sigma^2} \to \frac{1}{2}\frac{U''(y)}{U'(y)} \quad \textit{cuando } \sigma \to 0.$$

donde $y \in \mathcal{R}$,

En adelante se supondrá que el cociente de proporcionalidad es constante, es decir,

$$\frac{U''(y)}{U'(y)} \equiv \lambda, \quad \lambda, y \in \mathbb{R} \tag{1.3}$$

donde λ será referido como el ***coeficiente de sensibilidad al riesgo*** del consumidor y en tal caso el consumidor evalúa un costo (acotado) aleatorio Y usando la esperanza de $U(Y)$. Observe que, la expresión (1.3) implica inmediatamente que $U(y) = ae^{\lambda y} + b$ para toda $y \in \mathbb{R}$ y $a > 0$. Entonces, para cada $\lambda \in \mathbb{R}$, la ***función de utilidad exponencial*** es especificada como sigue: para cada $x \in \mathbb{R}$,

$$U(x) = \begin{cases} \operatorname{sign}(\lambda)e^{\lambda x} & si\ \lambda \neq 0 \\ x & si\ \lambda = 0 \end{cases}; \tag{1.4}$$

note que para cada $c, x \in \mathbb{R}$,

$$U(x+c) = e^{\lambda c} U(x), \quad \lambda \neq 0. \tag{1.5}$$

El equivalente seguro de Y correspondiente a U está dado como sigue:

$$\mathcal{E}_\lambda[Y] = \begin{cases} \dfrac{1}{\lambda} \log\left(E\left[e^{\lambda Y}\right]\right), & \text{si } \lambda \neq 0, \\ E[Y], & \text{si } \lambda = 0, \end{cases} \tag{1.6}$$

y de esta expresión no es difícil ver que $\mathcal{E}_\lambda[\cdot]$ es monótona y homogénea, es decir, para variables aleatorias (acotadas) Y y W,

$$Y \leq W \text{ c.s.} \implies \mathcal{E}_\lambda[Y] \leq \mathcal{E}_\lambda[W] \tag{1.7}$$

y

$$\mathcal{E}_\lambda[Y + a] = \mathcal{E}_\lambda[Y] + a, \quad a \in \mathbb{R}. \tag{1.8}$$

Note que, al considerar a la función de utilidad exponencial, $\lambda > 0$ corresponde al caso averso al riesgo y cuando $\lambda < 0$ se estará trabajando el caso propenso al riesgo. Finalmente, el caso $\lambda = 0$ corresponde al caso neutral al riesgo.

Ahora, volviendo a considerar un modelo de control de Markov, se supondrá que el controlador tiene un coeficiente de sensibilidad al riesgo constante $\lambda \in \mathbb{R}$ y se usará a la función de utilidad dada en (1.4).

Suponga que el controlador ha usado la política $\pi \in \mathcal{P}$ comenzando en $x \in S$, y sea $J_n(\lambda, \pi, x)$ el equivalente seguro del costo total $\sum_{t=0}^{n-1} C(X_t, A_t)$ incurrido antes del tiempo n, es decir,

$$J_n(\lambda, \pi, x) = \begin{cases} \dfrac{1}{\lambda} \log\left(E_x^\pi\left[e^{\lambda \sum_{t=0}^{n-1} C(X_t, A_t)}\right]\right), & \text{si } \lambda \neq 0, \\ E_x^\pi\left[\sum_{t=0}^{n-1} C(X_t, A_t)\right], & \text{si } \lambda = 0. \end{cases} \tag{1.9}$$

Con esta notación, el ***costo de límite superior promedio λ-sensible*** en $x \in S$ bajo la política π está dado por

$$J(\lambda, \pi, x) = \limsup_{n \to \infty} \frac{1}{n} J_n(\lambda, \pi, x), \tag{1.10}$$

y

$$J^*(\lambda, x) = \inf_{\pi \in \mathcal{P}} J(\lambda, \pi, x), \quad x \in S, \tag{1.11}$$

es la función de costo promedio λ-sensible óptima; una política $\pi^* \in \mathcal{P}$ es λ-óptima si $J(\lambda, \pi^*, x) = J^*(\lambda, x)$ para cada $x \in S$. El ***criterio promedio de límite inferior λ-sensible*** $J_-(\lambda, \pi, x)$ correspondiente a $\pi \in \mathcal{P}$ es definido por

$$J_-(\lambda, \pi, x) = \liminf_{n \to \infty} \frac{1}{n} J_n(\lambda, \pi, x), \tag{1.12}$$

y la correspondiente función de valor promedio de límite inferior λ-sensible es dada por

$$J_*(\lambda, x) = \inf_{\pi \in \mathcal{P}} J_-(\lambda, \pi, x), \quad x \in S, \tag{1.13}$$

así que

$$J_*(\lambda, \cdot) \leq J^*(\lambda, \cdot). \tag{1.14}$$

Por lo que, al buscar resolver el problema de control óptimo planteado en (1.11) y (1.14) el controlador debe elegir una política que optimice al criterio de rendimiento promedio, en otras palabras ,el controlador está modelando preferencias sobre las políticas mediante el criterio de rendimiento, por lo que si asociamos una función de utilidad al criterio de costo promedio se podría decir que una política es preferible a otra en función de cuál tiene mejor utilidad esperada, y en tal caso hablaremos de ***Procesos de Decisión Sensibles al Riesgo***.

1.3. Ecuación de Optimalidad

A continuación se presenta la ***Ecuación de Optimalidad*** (EO) correspondiente al criterio promedio dado en (1.10) utilizada a lo largo del capítulo y la cual será de gran utilidad para el desarrollo de los resultados presentados:

$$U(g_\lambda + h_\lambda(x)) = \inf_{a \in A(x)} \sum_{y \in S} p_{xy}(a) U(C(x, a) + h_\lambda(y)), \quad x \in S, \tag{1.15}$$

donde $g_\lambda \in \mathbb{R}$ y $h_\lambda(\cdot)$ es una función real valuada definida sobre S.

La importancia de esta ecuación en el estudio del índice promedio λ-sensible es mostrada en el siguiente resultado, el cual para el caso averso al riesgo ($\lambda > 0$) ha sido ya establecido, véase por ejemplo [18]. La versión presentada a continuación, a diferencia de otros autores, no impone restricción alguna sobre el signo de λ.

Lema 1.3.1 *Dado $\lambda \in \mathbb{R}$, sean $g_\lambda \in \mathbb{R}$ y $h_\lambda \in \mathcal{B}(S)$ tales que la ecuación de optimalidad (1.15) es válida. En este caso,*

(i) *Para cada $x \in S$ la suma en el lado derecho de (1.15) es una función continua de $a \in A(x)$, y entonces existe $f_\lambda(x) \in A(x)$ tal que minimiza el lado derecho de (1.15).*

(ii) *Para cada estado $x \in S$,*

$$g_\lambda = \lim_{n \to \infty} \frac{1}{n} J_n(\lambda, f_\lambda, x) = J^*(\lambda, x) = J_*(\lambda, x).$$

Demostración. (i) Usando el hecho de que la función de costo C y h_λ son acotadas, por el teorema de convergencia dominada y la Suposición 1.1.1 se sigue la conclusión deseada.

(ii) Por (1.15) se tiene que

$$E_x^\pi \left[U(C(X_0, A_0) + h_\lambda(X_1)) \right] \geq U(g_\lambda + h_\lambda(x)),$$

para cada $x \in S$ y $\pi \in \mathcal{P}$; a partir de esto, usando inducción y la propiedad de Markov, se sigue que para cada entero positivo n

$$\begin{aligned}
& E_x^{\pi}\left[U\left(\sum_{t=0}^{n-1} C(X_t, A_t) + h_\lambda(X_n)\right) \middle| \mathcal{H}_{n-2}, A_{n-2}\right] \\
= \; & E_x^{\pi}\left[U\left(\sum_{t=0}^{n-2} C(X_t, A_t) + C(X_{n-1}, A_{n-1}) + h_\lambda(X_n)\right) \middle| \mathcal{H}_{n-2}, A_{n-2}\right] \\
= \; & U\left(\sum_{t=0}^{n-2} C(X_t, A_t)\right) E_x^{\pi}\left[U\left(C(X_{n-1}, A_{n-1}) + h_\lambda(X_n)\right)\right] \\
\geq \; & U\left(\sum_{t=0}^{n-2} C(X_t, A_t) + g_\lambda + h_\lambda(X_{n-1})\right),
\end{aligned}$$

de donde, aplicando esperanza en la expresión anterior, lleva a

$$E_x^{\pi}\left[U\left(\sum_{t=0}^{n-1} C(X_t, A_t) + h_\lambda(X_n)\right)\right] \geq U(\sum_{t=0}^{n-2} C(X_t, A_t) + g_\lambda + h_\lambda(X_{n-1})),$$

repitiendo el proceso anterior se tiene que

$$\begin{aligned}
E_x^{\pi}\left[U\left(\sum_{t=0}^{n-1} C(X_t, A_t) + \|h_\lambda\|\right)\right] \; & \geq \; E_x^{\pi}\left[U\left(\sum_{t=0}^{n-1} C(X_t, A_t) + h_\lambda(X_n)\right)\right] \\
& \geq \; U(ng_\lambda + h_\lambda(x)),
\end{aligned}$$

donde $\|\cdot\|$ es la norma del supremo y la primera desigualdad es válida debido al hecho de que U es creciente. La última relación, junto con (1.6)–(1.8) implican lo siguiente:

$$\|h_\lambda\| + J_n(\lambda, \pi, x) \geq ng_\lambda + h_\lambda(x),$$

dividiendo ambos lados de esta relación por n y aplicando el límite inferior cuando n tiende a ∞ en ambos lados de la desigualdad resultante, se tiene lo siguiente

$$J_-(\lambda, \pi, x) \geq g_\lambda;$$

ya que la política π y el estado x fueron arbitrarios, de (1.12) y (1.13), se sigue que

$$\liminf_{n\to\infty} \frac{1}{n} J_n(\lambda, f_\lambda, x) \geq J_*(\lambda, x) \geq g_\lambda, \quad x \in S. \tag{1.16}$$

Por otro lado, de la definición de la política f_λ se tiene que

$$E_x^{f_\lambda}\left[U(C(X_0, A_0) + h_\lambda(X_1))\right] = U(g_\lambda + h_\lambda(x)), \quad x \in S,$$

y un argumento de inducción lleva a que, para cada $n = 1, 2, 3, \ldots$,

$$\begin{aligned}
U(ng_\lambda + h_\lambda(x)) \; & = \; E_x^{f_\lambda}\left[U\left(\sum_{t=0}^{n-1} C(X_t, A_t) + h_\lambda(X_n)\right)\right] \\
& \geq \; E_x^{f_\lambda}\left[U\left(\sum_{t=0}^{n-1} C(X_t, A_t) - \|h_\lambda\|\right)\right].
\end{aligned}$$

Usando (1.6)-(1.8) esta relación implica que

$$ng_\lambda + h(x) \geq J_n(\lambda, f_\lambda, x) - \|h_\lambda\|$$

y entonces

$$\begin{aligned} g_\lambda &\geq \limsup_{n\to\infty} \frac{1}{n} J_n(\lambda, f_\lambda, x) \\ &\geq J^*(\lambda, x), \quad x \in S; \end{aligned}$$

donde la segunda desigualdad es consecuencia de (1.11). En consecuencia, la conclusión deseada se sigue combinando esto último con (1.16) y (1.14). ■

Hasta aquí se ha presentado a la ecuación de optimalidad para los Procesos de Decisión de Markov Sensibles al Riesgo y se dio un resultado donde se muestra la importancia de tal ecuación para el estudio del índice promedio óptimo. Sin embargo, estos resultados son ya conocidos y han sido trabajados desde diversas perspectivas y por diversos autores [1], [4],[8]-[14],[18]-[22],[26],[28]. En particular, en [4] y [15] se probó que el criterio promedio óptimo sensible al riesgo coincide con el criterio óptimo neutral al riesgo para las funciones de utilidad potencia y logarítmica, respectivamente. Los resultados obtenidos por estos autores nos motivaron a plantearnos el siguiente problema:

- *¿Bajo qué condiciones se puede garantizar la continuidad del índice promedio óptimo con respecto al parámetro de sensibilidad al riesgo λ?*

El problema arriba planteado será formalmente establecido en el siguiente capítulo.

Capítulo 2

Continuidad del Costo Promedio Óptimo

Cuando se consideran a los Procesos de Decisión de Markov con un criterio de rendimiento promedio y se supone que el controlador cuenta con un coeficiente de sensibilidad al riesgo, puede observarse que varios autores han realizado un trabajo bastante extenso: desde establecer la caracterización del índice promedio óptimo [8], [9], [11], [12] o considerar espacios de estados mas generales [18]-[21], [28], hasta garantizar la validez del algoritmo de iteración de valores [14] o asegurar la existencia de políticas óptimas [22]. Sin embargo, hasta el momento nada se había dicho acerca de la continuidad de la función de valor óptimo con respecto al parámetro de sensibilidad al riesgo.

Por tanto, motivados por los resultados obtenidos recientemente en [4] y [15] con respecto a la relación entre los índices promedio neutral y sensible al riesgo para las funciones de utilidad potencia y logarítmica, es que en este capítulo se garantiza la continuidad para el índice promedio óptimo con respecto a λ. Los resultados presentados a continuación se encuentran motivados en [16].

2.1. Condición Simultánea de Doeblin

En esta sección se presenta la condición necesaria para establecer el resultado principal de la primera parte del trabajo de tesis. Tal condición está relacionada con la forma en la que se comunican los estados del proceso. A continuación se presenta un ejemplo que justifica la elección de dicha condición.

Ejemplo 2.1.1 *Sean el espacio de estados,* $S = \{0, 1, -1\}$ *y el conjunto de controles* $A = \{a\}$*, así que* $A(x) = \{a\}$ *para cada* $x \in S$*. Para este modelo hay una única política estacionaria, la cual se denotará por* f*. Definimos la función de costo como:*

$$C(x, a) \equiv C(x) = x, \quad x \in S,$$

y la ley de transición $[p_{x\,y}(a)] \equiv [p_{x\,y}]$ *dada por*

$$p_{0,x} = \frac{1}{2}, \quad p_{x,x} = 1, \quad x = 1, -1.$$

Es inmediato, de la definición de la ley de transición, que los estados 1 *y* -1 *son absorbentes, entonces, debido a como se encuentra definida la función de costo se tiene que:*

$$\begin{aligned} J(\lambda, x) &= \limsup_{n\to\infty} \frac{1}{n} J_n(\lambda, x) \\ &= x, \quad x = -1, 1, \quad \lambda \in \mathbb{R}. \end{aligned}$$

Suponga que el sistema comienza en $X_0 = 0$*. En este caso al tiempo* 1 *el sistema transitará al estado* 1 *o* -1 *con probabilidad* $1/2$ *y, para cada entero* $n > 1$*, la propiedad de Markov implica que*

$$\begin{aligned} E_0\left[\sum_{t=0}^{n-1} C(X_t)\right] &= C(0) + \frac{1}{2}\left(E_{-1}\left[\sum_{t=0}^{n-2} C(X_t)\right] + E_1\left[\sum_{t=0}^{n-2} C(X_t)\right]\right) \\ &= 0 + \frac{1}{2}\left(-(n-1) + (n-1)\right) \\ &= 0, \end{aligned}$$

y entonces

$$J(0,0) = 0. \tag{2.1}$$

Por otro lado, para $\lambda \neq 0$*,*

$$\begin{aligned} E_0\left[e^{\lambda \sum_{t=0}^{n-1} C(X_t)}\right] &= \frac{1}{2}\left(e^{\lambda(n-2)} + e^{-\lambda(n-2)}\right) \\ &= \frac{1}{2}\left(e^{|\lambda|(n-2)} + e^{-|\lambda|(n-2)}\right), \end{aligned}$$

así que

$$\frac{e^{|\lambda|(n-2)}}{2} \leq E_0\left[e^{\lambda \sum_{t=0}^{n-1} C(X_t)}\right] \leq e^{|\lambda|(n-2)},$$

dicha relación vía (1.9)-(1.11) implican que

$$\begin{aligned} J(\lambda, 0) &= \lim_{n\to\infty} \frac{1}{n\lambda} \log\left(E_0\left[e^{\lambda \sum_{t=0}^{n-1} C(X_t)}\right]\right) \\ &= \frac{|\lambda|}{\lambda} \\ &= \mathit{sign}(\lambda), \quad \lambda \neq 0. \end{aligned}$$

Combinando esto último con (2.1) se sigue que $J(\cdot, 0)$ *no es continua en* $\lambda = 0$*.*

En el ejemplo anterior, la discontinuidad de la función $\lambda \mapsto J^*(\lambda, 0)$ puede adjudicarse al hecho de que los estados -1 y 1 no se comunican.

La siguiente notación será usada en adelante. Para cada conjunto no vacío $F \subset S$, el tiempo de alcance a F, T_F, es definido por

$$T_F := \min\{n \in \mathbb{N} \setminus \{0\} | X_n \in F\} \tag{2.2}$$

donde, de acuerdo con la convención usual, el mínimo del conjunto vacío es $+\infty$; cuando $F = \{z\}$ en lugar de $T_{\{z\}}$ se empleará la notación T_z.

Como se ilustra en el Ejemplo 2.1.1, se requiere imponer una condición sobre la ley de transición para así garantizar la continuidad de la función de valor óptimo con respecto de λ. Dicha condición, llamada la ***Condición Simultánea de Doeblin***, es dada a continuación.

Suposición 2.1.2 *Existen un conjunto finito $F \subset S$ y una constante $M > 0$ satisfaciendo las propiedades siguientes:*

(i) $E_x^f [T_F] \leq M, \quad x \in S \setminus F, \quad f \in \mathbb{F}.$

(ii) Para cada $y \in F$ existe una política $f^y \in \mathbb{F}$ tal que

$$E_x^{f^y} [T_y] \leq M, \quad x \in F \setminus \{y\}. \tag{2.3}$$

Observe que en el Ejemplo 2.1.1 la parte (i) de la Suposición 2.1.2 es válida con $F = \{-1, 1\}$ y $M = 1$, pero la parte (ii) no lo es, debido al hecho de que los estados -1 y 1 son absorbentes.

Se tiene que, bajo la Suposición 1.1.1, la condición simultánea de Doeblin, enunciada en la Suposición 2.1.2, tiene una consecuencia importante en el análisis del criterio de costo promedio neutral al riesgo, tal consecuencia es que la correspondiente ecuación de optimalidad tiene una solución acotada, un hecho que tiene dos implicaciones (véase [2], [23], [33]):

(a) el costo promedio óptimo es constante, además;

(b) existe una política óptima estacionaria.

Mediante la ***Condición de Lyapunov*** dada en [25], la cual impone condiciones bajo las cuales la ecuación de optimalidad tiene una solución posiblemente no acotada, estas dos propiedades también pueden ser aseguradas. Tal condición, bajo la Suposición 1.1.1, puede ser formulada como sigue bajo el contexto de costos acotados [13]:

Suposición 2.1.3 *Existe un estado z tal que*

$$\lim_{n \to \infty} \sup_{f \in \mathbb{F}} E_x^f [T_z I[T_z > n]] = 0, \; x \in S. \tag{2.4}$$

Una pregunta interesante es: ¿el resultado de continuidad que se desea garantizar (Teorema 2.4.2) continúa siendo válido si se reemplaza la Suposición 2.1.2 por la condición dada en (2.4)?. El siguiente ejemplo no controlado garantiza que la respuesta a dicha pregunta es negativa.

Ejemplo 2.1.4 *Suponga que* $S = \mathbb{N}$ *y que* $A = \{a\} = A(x)$ *para cada* $x \in S$. *Considere la función de costo dada por*

$$C(0,a) \equiv C(0) = 0, \text{ y } C(x,a) \equiv C(x) = 1 \text{ para } x \in S \setminus \{0\}, \tag{2.5}$$

y suponga que la ley de transición $[p_{x\,y}(a)] \equiv [p_{x\,y}]$ *está determinada por*

$$p_{0,0} = 1, \quad p_{x,x+1} = \frac{x^2}{(x+1)^2} = 1 - p_{x,0}, \quad x = 1,2,3,\ldots.$$

Se tiene que sólo hay una política, f, *la cual no será denotada explícitamente en el desarrollo del ejemplo.*

Se mostrará que la condición de Lyapunov en (2.4) es válida para este modelo con $z = 0$. *Para ello, del hecho de existir sólo una política, note que (2.4) es equivalente a*

$$E_x[T_0] < \infty, \quad x \in S. \tag{2.6}$$

Para verificar esta propiedad, observe que $E_0[T_0] = 1$, *ya que* 0 *es un estado absorbente. Considere un estado no nulo* x *y observe que la forma de la ley de transición lleva a que para cada entero positivo* n

$$\begin{aligned} P_x[T_0 > n] &= P_x[X_r = x + r,\ 0 < r \le n] \\ &= \frac{x^2}{(x+1)^2}\frac{(x+1)^2}{(x+2)^2}\cdots\frac{(x+n-1)^2}{(x+n)^2} \\ &= \frac{x^2}{(x+n)^2}, \end{aligned} \tag{2.7}$$

y entonces

$$\begin{aligned} E_x[T_0] &= 1 + \sum_{n=1}^{\infty} P_x[T_0 > n] \\ &= 1 + \sum_{n=1}^{\infty} \frac{x^2}{(x+n)^2} \\ &\le 1 + x, \end{aligned}$$

lo cual muestra que (2.6) es válida para $z = 0$.

A continuación, se probará que para cada $x \neq 0$, *la función* $\lambda \mapsto J(\lambda, x)$ *no es continua en cero. Primero, note que*

$$J(\lambda, 0) = 0, \quad \lambda \in \mathbb{R},$$

ya que el estado 0 *es absorbente y* $C(0) = 0$*. En [25] se prueba que la condición de Lyapunov implica que el costo promedio neutral al riesgo* $J(0, x)$ *no depende del estado* x*, por lo tanto*

$$J(0, x) = 0, \quad x \in S. \tag{2.8}$$

Ahora, sean el estado $x \neq 0$ *y* $\lambda > 0$ *arbitrarios, observe que* $X_t \neq 0$ *ocurre* P_x*-casi seguramente sobre el evento* $[t < T_0]$*, así que (2.5) y (2.7) implican lo siguiente:*

$$\begin{aligned} e^{\lambda n} &\geq E_x\left[e^{\lambda \sum_{t=0}^{n-1} C(X_t)}\right] \\ &\geq E_x\left[e^{\lambda \sum_{t=0}^{n-1} C(X_t)} I[T_0 > n]\right] \\ &= \frac{e^{\lambda n} x^2}{(x+n)^2}. \end{aligned}$$

Entonces, usando (1.9) y (1.10), se tiene que

$$\begin{aligned} 1 &\geq \frac{1}{n} J_n(\lambda, x) \\ &> 1 + \frac{1}{\lambda n} \log\left(\frac{x^2}{(x+n)^2}\right), \end{aligned}$$

de donde,

$$1 = \lim_{n \to \infty} \frac{1}{n} J_n(\lambda, x) = J(\lambda, x), \quad x \neq 0, \quad \lambda > 0,$$

y entonces $J(\cdot, x)$ *no es continua en* $\lambda = 0$*, por (2.8).*

En resumen, el modelo presentado en este ejemplo satisface la condición de Lyapunov, pero la función $\lambda \mapsto J^*(\lambda, x)$ *no es continua en cero cuando el estado* x *es no nulo.*

En la siguiente sección se establece el resultado de continuidad para el índice promedio óptimo.

2.2. Continuidad del Índice Promedio Óptimo: $\lambda \neq 0$

En esta sección se establece la continuidad para el caso $\lambda \neq 0$. Considere al modelo $\mathcal{M}$, dado en la Sección 1.1. En adelante, el modelo $\mathcal{M}$ satisface las condiciones dadas en la Suposición 1.1.1.

El problema a resolver consiste en garantizar la continuidad de la función de valor óptimo $J^*(\lambda, \cdot)$ con respecto del coeficiente de sensibilidad al riesgo λ. Tal propiedad, para el caso $\lambda \neq 0$ es inmediata, como puede verse en el resultado siguiente:

Proposición 2.2.1 *Para cada $x \in S$, la función $\lambda \mapsto J^*(\lambda, x)$ es continua sobre $\mathbb{R} \setminus \{0\}$.*

Demostración. Sean $x \in S$, n un entero positivo y λ y ν números reales no nulos, arbitrarios. Ahora, observe que

$$\begin{aligned} \lambda \sum_{t=0}^{n-1} C(X_t, A_t) &= \nu \sum_{t=0}^{n-1} C(X_t, A_t) + (\lambda - \nu) \sum_{t=0}^{n-1} C(X_t, A_t) \\ &\leq \nu \sum_{t=0}^{n-1} C(X_t, A_t) + n|\lambda - \nu| \|C\|, \end{aligned}$$

donde $\|C\| := \sup_{x \in K} |C(x)| < \infty$; la desigualdad anterior junto con (1.9) implican que

$$\begin{aligned} \lambda J_n(\lambda, \pi, x) &= \log \left(E_x^\pi \left[e^{\lambda \sum_{t=0}^{n-1} C(X_t, A_t)} \right] \right) \\ &\leq \log \left(E_x^\pi \left[e^{\nu \sum_{t=0}^{n-1} C(X_t, A_t)} \right] \right) + n|\lambda - \nu| \|C\| \\ &= \nu J_n(\nu, \pi, x) + n|\lambda - \nu| \|C\|, \end{aligned}$$

de donde

$$\lambda \frac{J_n(\lambda, \pi, x)}{n} \leq \nu \frac{J_n(\nu, \pi, x)}{n} + |\lambda - \nu| \|C\|. \tag{2.9}$$

Ahora, suponga que λ y ν son positivos. Tomando el límite superior cuando n tiende a ∞ en ambos lados de (2.9), por (1.10) se sigue que

$$\lambda J(\lambda, \pi, x) \leq \nu J(\nu, \pi, x) + |\lambda - \nu| \|C\|,$$

y entonces, tomando ínfimo sobre $\pi \in \mathcal{P}$ en ambos lados de esta relación, por (1.11) se obtiene que

$$\lambda J^*(\lambda, x) \leq \nu J^*(\nu, x) + |\lambda - \nu| \|C\|;$$

por lo tanto, ya que $\lambda, \nu \in (0, \infty)$ son arbitrarios, se tiene que

$$\lambda J^*(\lambda, x) - \nu J^*(\nu, x)| \leq |\lambda - \nu| \|C\|. \tag{2.10}$$

Si ahora λ y ν son negativos, tomando el límite inferior cuando n tiende a ∞ en ambos lados de (2.9), de (1.10) se sigue que

$$\lambda J(\lambda, \pi, x) \leq \nu J(\nu, \pi, x) + |\lambda - \nu| \|C\|,$$

una relación que, después de tomar el supremo sobre $\pi \in \mathcal{P}$, lleva a

$$\lambda J^*(\lambda, x) \leq \nu J^*(\nu, x) + |\lambda - \nu| \|C\|,$$

e intercambiando los papeles de λ y ν, se sigue que (2.10) es también válida.

En resumen, se ha mostrado que $\lambda J^*(\lambda, x)$ es una función Lipschitz de λ en cada uno de los intervalos $(0, \infty)$ y $(-\infty, 0)$, así que, $\lambda \mapsto J^*(\lambda, x)$ es continua en cada punto en $\mathbb{R} \setminus \{0\}$. ∎

Observación 2.2.2 *Note que la conclusión de la proposición anterior depende únicamente de la cota de la función de costo; sin embargo, como se muestra en el ejemplo 2.1.1, la continuidad de $J^*(\cdot, x)$ en $\lambda = 0$ no puede ser asegurada de manera general suponiendo únicamente costos acotados.*

2.3. La Técnica de Factor Desvaneciente

En esta sección se presentan los resultados auxiliares que serán usados para probar el Teorema 2.4.1, tal resultado tiene un papel fundamental en la prueba del Teorema 2.4.2, el cual garantiza la continuidad del índice promedio óptimo en $\lambda = 0$. La técnica presentada a continuación está basada en el siguiente operador descontado, el cual fue presentado en [18].

Definición 2.3.1 *Para $\alpha \in (0, 1)$ y $\lambda \in \mathbb{R} \setminus \{0\}$ el operador $T_{\alpha,\lambda} : \mathbb{B}(S) \to \mathbb{B}(S)$ es definido como sigue: para cada $V \in \mathbb{B}(S)$, la función $T_{\alpha,\lambda}[V]$ está determinada por*

$$U(T_{\alpha,\lambda}[V](x)) = \inf_{a \in A(x)} \left[\sum_{y \in S} p_{xy}(a) U(C(x, a) + \alpha V(y)) \right], \quad x \in S. \tag{2.11}$$

Combinando (1.5) con el hecho de que U es creciente, se puede ver que $T_{\alpha,\lambda}$ es un operador monótono y α-homogéneo, es decir, dados $V, W \in \mathbb{B}(S)$,

$$T_{\alpha,\lambda}[V] \geq T_{\alpha,\lambda}[W]$$

cuando $V \geq W$, y

$$T_{\alpha,\lambda}[V + r] = T_{\alpha,\lambda}[V] + \alpha r,$$

para cada $r \in \mathbb{R}$. Estas propiedades implican que $T_{\alpha,\lambda}$ es un operador contracción en el espacio $\mathbb{B}(S)$ dotado con la norma del supremo, y su módulo es α, es decir,

$$\|T_{\alpha,\lambda}[V] - T_{\alpha,\lambda}[W]\| \leq \alpha \|V - W\|, \quad V, W \in \mathbb{B}(S). \tag{2.12}$$

Por lo tanto, por el Teorema del Punto Fijo de Banach, existe una única función $V_{\alpha,\lambda} \in \mathbb{B}(S)$ satisfaciendo $T_{\alpha,\lambda}[V_{\alpha,\lambda}] = V_{\alpha,\lambda}$; más explícitamente,

$$U(V_{\alpha,\lambda}(x)) = \inf_{a \in A(x)} \left[\sum_{y \in S} p_{xy}(a) U(C(x, a) + \alpha V_{\alpha,\lambda}(y)) \right], \quad x \in S. \tag{2.13}$$

Observe que (2.11) implica que

$$T_{\alpha,\lambda}[0](x) = \inf_{a \in A(x)} C(x, a),$$

así que,

$$\|T_{\alpha,\lambda}[0]\| \leq \|C\|.$$

Usando (2.12) con $V_{\alpha,\lambda}$ y 0 en lugar de V y W, respectivamente, se sigue que

$$(1-\alpha)\|V_{\alpha,\lambda}\| \leq \|C\|. \tag{2.14}$$

A continuación, para cada $x \in S$ se define

$$g_{\alpha,\lambda}(x) = (1-\alpha)V_{\alpha,\lambda}(x) \in [-\|C\|, \|C\|], \quad \alpha \in (0,1), \quad \lambda \in \mathbb{R} \setminus \{0\}, \tag{2.15}$$

y note que el hecho de que $V_{\alpha,\lambda}$ y $C(\cdot,\cdot)$ sean acotados, junto con la Suposición 1.1.1 implican que, para cada $\lambda \neq 0$ y $\alpha \in (0,1)$, existe una política estacionaria $f_{\alpha,\lambda}$ satisfaciendo

$$\begin{aligned} U(V_{\alpha,\lambda}(x)) &= \sum_{y \in S} p_{x\,y}(f_{\alpha,\lambda}(x)) U(C(x, f_{\alpha,\lambda}(x)) + \alpha V_{\alpha,\lambda}(y)) \\ &= E_x^{f_{\alpha,\lambda}}\left[U(C(X_0, A_0,) + \alpha V_{\alpha,\lambda}(X_1))\right], \quad x \in S. \end{aligned} \tag{2.16}$$

Para establecer la prueba del resultado principal (Teorema 2.4.2) se utiliza un resultado dado en la siguiente sección (Teorema 2.4.1), el cual establece que para algún $\delta > 0$, la ecuación de optimalidad correspondiente a un parámetro de sensibilidad al riesgo $\lambda \in (-\delta, \delta) \setminus \{0\}$ tiene solución $(g_\lambda, h_\lambda(\cdot))$, de tal manera que la familia $\{h_\lambda : 0 < |\lambda| < \delta\}$ es acotada en $\mathbb{B}(S)$.

La prueba del Teorema 2.4.1 usa la técnica de factor de descuento desvaneciente, presentada anteriormente, y el siguiente resultado, cuya prueba usa la condición simultánea de Doeblin ([25]).

Lema 2.3.2 *Suponga que la Suposición 2.1.2 es válida. En este caso, existen $b \in (1,\infty)$ y $\rho \in (0,1)$ tales que las siguientes afirmaciones son válidas para cada $n \in \mathbb{N}$:*

(i) $P_x^f[T_F \geq n] \leq b\rho^n$ para cada $x \in S \setminus F$ y $f \in \mathbb{F}$, donde F es el conjunto finito en la Suposición 2.1.2.

(ii) Para cada $y \in F$, la política estacionaria f^y en la segunda parte de la Suposición 2.1.2 satisface que

$$P_x^{f^y}[T_y \geq n] \leq b\rho^n, \quad x \in S \setminus \{y\}.$$

Demostración. Sea N un entero positivo tal que $N \geq 2M$. En este caso, la desigualdad de Markov implica que

$$\begin{aligned} P_x^f[T_F \geq N] &\leq \frac{E_x^f[T_F]}{N} \\ &\leq \frac{M}{N} \leq \frac{1}{2}, \quad x \in S \setminus F, \quad f \in \mathbb{F}. \end{aligned}$$

La expresión anterior, usando un argumento de inducción junto con la propiedad de Markov, implica que para cada estado $x \in S \setminus F$ y $f \in \mathbb{F}$, la desigualdad

$$P_x^f[T_F \geq kN] \leq (1/2)^k$$

es válida para cada entero no negativo k. Dado $n \in \mathbb{N}$ escribimos $n = kN + r$ donde $r \in \{0, 1, 2, \ldots, N-1\}$, y note que

$$P_x^f[T_F \geq n] \leq P_x^f[T_F \geq kN] \leq (1/2)^k,$$

de lo cual

$$P_x^f[T_F \geq n] \leq \tilde{b}\tilde{\rho}^n, \quad x \in S \setminus F, \quad f \in \mathbb{F}, \quad n \in \mathbb{N}, \tag{2.17}$$

donde $\tilde{\rho} = (1/2)^{1/N} \in (0,1)$ y $\tilde{b} = \tilde{\rho}^{-N} \in (1, \infty)$. Ahora, para cada $y \in F$ considere la política $f^y \in \mathbb{F}$ dada en la Suposición 2.1.2. Se mostrará lo siguiente:

$$E_x^{f^y}[T_y] \leq 2M, \quad x \in S \setminus \{y\}. \tag{2.18}$$

Note que la desigualdad es válida para $x \in F \setminus \{y\}$, por (2.3). Para completar la prueba observe que, del hecho que $y \in F$ se tiene que $T_F \leq T_y$, así que

$$\begin{aligned} E_x^{f^y}[T_y] &= E_x^{f^y}[T_F] + E_x^{f^y}[(T_y - T_F)I[T_y > T_F]] \\ &\leq M + E_x^{f^y}[(T_y - T_F)I[T_y > T_F]], \quad x \in S \setminus F, \end{aligned}$$

donde la desigualdad se debe a la Suposición 2.1.2(i). Observando que $X_{T_F} \in F \setminus \{y\}$ sobre el evento $T_y > T_F$ y usando (2.2), la propiedad fuerte de Markov aplicada a la cadena de Markov inducida por f^y lleva, por (2.3), a

$$E_x^{f^y}[(T_y - T_F)I[T_y > T_F] | T_y > T_F, X_{T_F}] = E_{X_{T_F}}[T_y] \leq M,$$

y entonces

$$E_x^{f^y}[(T_y - T_F)I[T_y > T_F]] \leq M,$$

una relación que junto con la desigualdad anterior implica que la desigualdad en (2.18) es también válida cuando $x \in S \setminus F$. Ahora, sea el modelo reducido $\mathcal{M}_{f^y}$ el cual consiste en el elemento $f^y(x)$ como única acción disponible para cada $x \in S$. Para este modelo $\mathcal{M}_{f^y}$, la relación (2.18) establece que la Suposición 2.1.2 es válida con $F = \{y\}$, y entonces la primera parte de la prueba lleva a que existen $b_y \in (1, \infty)$ y $\rho_y \in (0,1)$ tales que

$$P_x^{f^y}[T_y \geq n] \leq b_y \rho_y^n, \quad n \in \mathbb{N}, \quad x \in S \setminus \{y\}.$$

Haciendo $b := \max\{\tilde{b}, b_y, y \in F\} \in (1, \infty)$ y $\rho := \max\{\tilde{\rho}, \rho_y : y \in F\} \in (0,1)$, la expresión en (2.17) lleva a que las conclusiones en (i) y (ii) son válidas. ∎

Lema 2.3.3 *Suponga que las Suposiciones 1.1.1 y 2.1.2 son válidas. Sean $b > 1$ y $\rho \in (0,1)$ como en el Lema 2.3.2 y sean*

$$\delta := \frac{1}{2\|C\|} \log\left(\frac{1+\rho}{2\rho}\right) \quad y \quad B = \log\left(\frac{2b}{1-\rho}\right). \tag{2.19}$$

Con esta notación, para cada $\alpha \in (0,1)$ las afirmaciones siguientes son válidas, donde F es el conjunto finito dado en la Suposición 2.1.2, y el tiempo de alcance T_F es como en (2.2).

(i) Para cada $x \in S \setminus F$ y $f \in \mathbb{F}$,

$$e^{-|\lambda|B/\delta} \leq E_x^f[e^{2\lambda\|C\|T_F}] < e^{|\lambda|B/\delta}, \quad 0 < |\lambda| \leq \delta.$$

(ii) Para cada $z \in F$ la política estacionaria f^z en la Suposición 2.1.2(ii) satisface que

$$e^{-|\lambda|B/\delta} \leq E_x^{f^z}[e^{2\lambda\|C\|T_z}] < e^{|\lambda|B/\delta}, \quad x \in S \setminus \{z\}, \quad 0 < |\lambda| \leq \delta.$$

(iii) Para cada $x \in S \setminus F$, $f \in \mathbb{F}$ y $\lambda \in (-\delta, \delta) \setminus \{0\}$,

$$\begin{aligned} &\lim_{n\to\infty} E_x^f\left[U\left(\sum_{t=0}^{T_F\wedge n-1}[C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{T_F\wedge n})\right)\right] \\ &\quad = E_x^f\left[U\left(\sum_{t=0}^{T_F-1}[C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{T_F})\right)\right]; \end{aligned} \tag{2.20}$$

ver (2.15) para la definición de $g_{\alpha,\lambda}(\cdot)$.

(iv) Para cada $\lambda \in (-\delta, \delta) \setminus \{0\}$, y $z \in F$, la política $f^z \in \mathbb{F}$ en la Suposición 2.1.2(ii) satisface que, para cada $x \in S \setminus \{z\}$,

$$\begin{aligned} &\lim_{n\to\infty} E_x^{f^z}\left[U\left(\sum_{t=0}^{T_z\wedge n-1}[C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{T_z\wedge n})\right)\right] \\ &\quad = E_x^{f^z}\left[U\left(\sum_{t=0}^{T_z-1}[C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{T_z})\right)\right]. \end{aligned} \tag{2.21}$$

Demostración. (*i*) Sean $x \in S \setminus F$ y $f \in \mathbb{F}$ arbitrarios. Por el Lema 2.3.2(i), la desigualdad

$$P_x^f[T_F \geq n] \leq b\rho^n$$

es válida para cada n. Entonces, de (2.19) se sigue que

$$e^{2\delta\|C\|}\rho = \frac{1+\rho}{2} < 1,$$

de lo cual

$$\begin{aligned} E_x^f[e^{2\delta\|C\|T_F}] &= \sum_{k=1}^{\infty} e^{2\delta\|C\|k} P_x^f[T_F = k] \\ &\leq b\sum_{k=1}^{\infty} e^{2\delta\|C\|k}\rho^k \\ &\leq \frac{b}{1-\rho e^{2\|C\|\delta}} \\ &= \frac{b}{1-(1+\rho)/2} \\ &= e^B. \end{aligned}$$

Ahora, sea $\lambda \in (-\delta, \delta) \setminus \{0\}$ arbitrario pero fijo, y note que

$$E_x^f[e^{2\lambda\|C\|T_F}] \leq E_x^f[e^{(|\lambda|/\delta)2\delta\|C\|T_F}];$$

de esta expresión, la desigualdad de Jensen aplicada a la función cóncava $\varphi(x) = x^{|\lambda|/\delta}$ sobre $(0, \infty)$ lleva a

$$E_x^f[e^{2\lambda\|C\|T_F}] \leq E_x^f[e^{2\delta\|C\|T_F}]^{(|\lambda|/\delta)},$$

y junto con la desigualdad anterior, esto último implica que

$$E_x^f[e^{2\lambda\|C\|T_F}] \leq e^{|\lambda|B/\delta}.$$

Observe que si W es una variable aleatoria positiva, entonces $E[W^{-1}] \geq 1/E[W]$, debido a la desigualdad de Jensen. Combinando este hecho con la expresión anterior, se sigue que

$$\begin{array}{rcl} e^{|\lambda|B/\delta} & \geq & E_x^f[e^{2\lambda\|C\|T_F}] \\ & \geq & E_x^f[e^{-2|\lambda|\|C\|T_F}] \\ & \geq & \dfrac{1}{E_x^f[e^{2|\lambda|\|C\|T_F}]} \\ & \geq & e^{-|\lambda|B/\delta}. \end{array}$$

(ii) Usando un argumento análogo al de la parte (i), el Lema 2.3.2(ii) implica la conclusión deseada.

(iii) Sean $\lambda \in (-\delta, \delta) \setminus \{0\}$, $x \in S \setminus F$ y $f \in \mathbb{F}$ arbitrarios, y observe que las siguientes propiedades (a) y (b) se satisfacen:

(a) Para cada entero positivo n, la siguiente identidad es válida sobre el evento $[T_F \leq n]$

$$\sum_{t=0}^{T_F \wedge n-1} [C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{T_F \wedge n})$$
$$= \sum_{t=0}^{T_F-1} [C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{T_F}).$$

Como T_F es finito P_x^f-c. s, se sigue que

$$\lim_{n\to\infty} U\left(\sum_{t=0}^{T_F \wedge n-1} [C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{T_F \wedge n})\right)$$
$$= U\left(\sum_{t=0}^{T_F-1} [C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{T_F})\right), \qquad P_x^f - c.s.$$

(b) Por (1.4) y (2.15),

$$\begin{aligned}
&\left|U\Big(\textstyle\sum_{t=0}^{T_F\wedge n-1}[C(X_t,A_t)-g_{\alpha,\lambda}(X_t)]+\alpha V_{\alpha,\lambda}(X_{T_F\wedge n})\Big)\right| \\
&\leq e^{|\lambda|\sum_{t=0}^{T_F\wedge n-1}|C(X_t,A_t)-g_{\alpha,\lambda}(X_t)|+\alpha|\lambda||V_{\alpha,\lambda}(X_{T_F\wedge n})|} \\
&\leq e^{2|\lambda|(T_F\wedge n)\|C\|+\alpha|\lambda|\|C\|/(1-\alpha)} \\
&\leq e^{2\delta\|C\|T_F}e^{\alpha|\lambda|\|C\|/(1-\alpha)}.
\end{aligned}$$

Como $E_x^f\left[e^{2\delta\|C\|T_F}\right]$ es finito por la parte (i), la propiedad (2.20) se sigue de combinar (a) y (b) con el Teorema de Convergencia Dominada. Un argumento similar usando la parte (ii) puede ser empleado para establecer la convergencia en (2.21). ■

Teorema 2.3.4 *Bajo las Suposiciones 1.1.1 y 2.1.2, existen $\delta > 0$ y $B^* > 0$ tales que, para cada $\lambda \in (-\delta,\delta)\setminus\{0\}$ y $\alpha \in (0,1)$,*

$$\alpha|V_{\alpha,\lambda}(x) - V_{\alpha,\lambda}(z)| \leq B^*, \quad x, z \in F, \tag{2.22}$$

y

$$\alpha|V_{\alpha,\lambda}(x) - V_{\alpha,\lambda}(z)| \leq 2B^*, \quad x \in S, \quad z \in F. \tag{2.23}$$

Demostración. Dado $z \in F$, sea f^z la política estacionaria en la Suposición 2.1.2(ii). De (2.13) se sigue que para cada $x \in S$

$$U(V_{\alpha,\lambda}(x)) \leq \sum_{y\in S} p_{x\,y}(f^z(x))U(C(x,f^z(x)) + \alpha V_{\alpha,\lambda}(y)),$$

dicha desigualdad, al aplicar(1.4) y (2.15), lleva a

$$\begin{aligned}
&U(\alpha V_{\alpha,\lambda}(x)) \\
&\quad\leq \sum_{y\in S} p_{x\,y}(f^z(x))U(C(x,f^z(x)) - g_{\alpha,\lambda}(x) + \alpha V_{\alpha,\lambda}(y)) \\
&\quad= E_x^{f^z}\left[U(C(X_0,A_0) - g_{\alpha,\lambda}(X_0) + \alpha V_{\alpha,\lambda}(X_1))\right], \quad x \in S.
\end{aligned} \tag{2.24}$$

Ahora, se probará por inducción que para cada entero positivo n y $x \in S$, la siguiente expresión es válida

$$\begin{aligned}
&U(\alpha V_{\alpha,\lambda}(x)) \\
&\quad\leq E_x^{f^z}\left[U\left(\sum_{t=0}^{T_z\wedge n-1}[C(X_t,A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{T_z\wedge n})\right)\right].
\end{aligned} \tag{2.25}$$

Para probar (2.25), primeramente note que para $n = 1$ la relación anterior es equivalente a (2.24), ya que $T_z \geq 1$. Suponga ahora que (2.25) es válida para algún entero positivo n. Sea $x \in S$ arbitrario y observe que

$$
\begin{aligned}
&E_x^{f^z}\left[U\left(\sum_{t=0}^{T_z\wedge n-1}[C(X_t,A_t)-g_{\alpha,\lambda}(X_t)]+\alpha V_{\alpha,\lambda}(X_{T_z\wedge n})\right)\right]\\
&=E_x^{f^z}\left[U\left(\sum_{t=0}^{T_z-1}[C(X_t,A_t)-g_{\alpha,\lambda}(X_t)]+\alpha V_{\alpha,\lambda}(X_{T_z})\right)I[T_z\le n]\right]\\
&\quad+E_x^{f^z}\left[U\left(\sum_{t=0}^{n-1}[C(X_t,A_t)-g_{\alpha,\lambda}(X_t)]+\alpha V_{\alpha,\lambda}(X_n)\right)I[T_z> n]\right]. \qquad (2.26)
\end{aligned}
$$

Luego, usando (1.5) note que

$$
\begin{aligned}
&U\left(\sum_{t=0}^{n-1}[C(X_t,A_t)-g_{\alpha,\lambda}(X_t)]+\alpha V_{\alpha,\lambda}(X_n)\right)I[T_z>n]\\
&\qquad=e^{\lambda\sum_{t=0}^{n-1}[C(X_t,A_t)-g_{\alpha,\lambda}(X_t)]}I[T_z>n]U\left(\alpha V_{\alpha,\lambda}(X_n)\right),
\end{aligned}
$$

y observe que (2.24) implica que

$$
\begin{aligned}
&U\left(\alpha V_{\alpha,\lambda}(X_n)\right)\\
&\quad\le\sum_{y\in S}p_{X_n\,y}(f^z(X_n))U(C(X_n,f^z(X_n))-g_{\alpha,\lambda}(X_n)+\alpha V_{\alpha,\lambda}(y))\\
&\quad=E_x^{f^z}\left[U(C(X_n,A_n)-g_{\alpha,\lambda}(X_n)+\alpha V_{\alpha,\lambda}(X_{n+1}))\,|X_k,k\le n\right],
\end{aligned}
$$

donde la igualdad es debida a la propiedad de Markov. Usando (1.5), las dos últimas expresiones implican que

$$
\begin{aligned}
&U\Big(\textstyle\sum_{t=0}^{T_z\wedge n-1}[C(X_t,A_t)-g_{\alpha,\lambda}(X_t)]+\alpha V_{\alpha,\lambda}(X_{T_z\wedge n})\Big)I[T_z>n]\\
&\quad\le E_x^{f^z}[U(\textstyle\sum_{t=0}^{n}[C(X_t,A_t)-g_{\alpha,\lambda}(X_t)]+\alpha V_{\alpha,\lambda}(X_{n+1}))I[T_z>n]|X_k,k\le n],
\end{aligned}
$$

y entonces

$$
\begin{aligned}
&E_x^{f^z}\left[U\left(\sum_{t=0}^{T_z\wedge n-1}[C(X_t,A_t)-g_{\alpha,\lambda}(X_t)]+\alpha V_{\alpha,\lambda}(X_{T_z\wedge n})\right)I[T_z>n]\right]\\
&\quad\le E_x^{f^z}\left[U\left(\sum_{t=0}^{n}[C(X_t,A_t)-g_{\alpha,\lambda}(X_t)]+\alpha V_{\alpha,\lambda}(X_{n+1})\right)I[T_z>n]\right]\\
&\quad=E_x^{f^z}\left[U\left(\sum_{t=0}^{T_z-1}[C(X_t,A_t)-g_{\alpha,\lambda}(X_t)]+\alpha V_{\alpha,\lambda}(X_{T_z})\right)I[T_z=n+1]\right]\\
&\quad+E_x^{f^z}\left[U\left(\sum_{t=0}^{n}[C(X_t,A_t)-g_{\alpha,\lambda}(X_t)]+\alpha V_{\alpha,\lambda}(X_{n+1})\right)I[T_z>n+1]\right];
\end{aligned}
$$

combinando esta relación con (2.26) se sigue que

$$
\begin{aligned}
&E_x^{f^z}\left[U\left(\sum_{t=0}^{T_z\wedge n-1}[C(X_t,A_t)-g_{\alpha,\lambda}(X_t)]+\alpha V_{\alpha,\lambda}(X_{T_z\wedge n})\right)\right]\\
&\quad\le E_x^{f^z}\left[U\left(\sum_{t=0}^{T_z-1}[C(X_t,A_t)-g_{\alpha,\lambda}(X_t)]+\alpha V_{\alpha,\lambda}(X_{T_z})\right)I[T_z\le n+1]\right]\\
&\qquad+E_x^{f^z}\left[U\left(\sum_{t=0}^{n}[C(X_t,A_t)-g_{\alpha,\lambda}(X_t)]+\alpha V_{\alpha,\lambda}(X_{n+1})\right)I[T_z> n+1]\right]\\
&\quad=E_x^{f^z}\left[U\left(\sum_{t=0}^{T_z\wedge(n+1)-1}[C(X_t,A_t)-g_{\alpha,\lambda}(X_t)]+\alpha V_{\alpha,\lambda}(X_{T_z\wedge(n+1)})\right)\right],
\end{aligned}
$$

una relación que junto con la hipótesis de inducción lleva a que (2.25) es válido con $n+1$ en lugar de n, completando el argumento de inducción.

Ahora, tomando el límite cuando n tiende a ∞ en (2.24), por la convergencia (2.21) en el Lema 2.3.3(ii) se sigue que, para cada $x\in S$

$$
U(\alpha V_{\alpha,\lambda}(x))\le E_x^{f^z}\left[U\left(\sum_{t=0}^{T_z-1}[C(X_t,A_t)-g_{\alpha,\lambda}(X_t)]+\alpha V_{\alpha,\lambda}(X_{T_z})\right)\right];
$$

usando que T_z es finito $P_x^{f^z}$-casi seguramente y que $X_{T_z}=z$ cuando $T_z<\infty$, esta desigualdad y (1.5) llevan a

$$
\begin{aligned}
U(\alpha[V_{\alpha,\lambda}(x)-V_{\alpha,\lambda}(z)]) &\le E_x^{f^z}\left[U\left(\sum_{t=0}^{T_z-1}[C(X_t,A_t)-g_{\alpha,\lambda}(X_t)]\right)\right]\\
&\le E_x^{f^z}\left[U\left(2\|C\|T_z\right)\right],\quad x\in S,
\end{aligned}
\tag{2.27}
$$

donde el hecho de que U es creciente fue usado en el último paso. Considere ahora los siguientes casos:

Caso 1: $\lambda\in(0,\delta)$. Por tanto, $U(w)=e^{\lambda w}$ para cada $w\in\mathbb{R}$, así que (2.27) lleva a

$$
e^{\lambda\alpha[V_{\alpha,\lambda}(x)-V_{\alpha,\lambda}(z)]}\le E_x^{f^z}\left[e^{2\lambda\|C\|T_z}\right],
$$

y entonces

$$
e^{\lambda\alpha[V_{\alpha,\lambda}(x)-V_{\alpha,\lambda}(z)]}\le e^{\lambda B/\delta}
$$

para $x\ne z$, por el Lema 2.3.3(ii); ya que la última desigualdad también es válida para $x=z$ se sigue que

$$
\alpha[V_{\alpha,\lambda}(x)-V_{\alpha,\lambda}(z)]\le\frac{B}{\delta},\quad x\in S.
\tag{2.28}
$$

Caso 2: $\lambda \in (-\delta, 0)$. En este caso, $U(w) = -e^{\lambda w}$ para cada $w \in \mathbb{R}$, y (2.27) lleva a

$$e^{\lambda\alpha[V_{\alpha,\lambda}(x)-V_{\alpha,\lambda}(z)]} \geq E_x^{f^z}\left[e^{2\lambda\|C\|T_z}\right],$$

y por el Lema 2.3.3(ii), esto implica que, para $x \neq z$,

$$e^{\lambda\alpha[V_{\alpha,\lambda}(x)-V_{\alpha,\lambda}(z)]} \geq e^{-|\lambda|B/\delta} = e^{\lambda B/\delta},$$

una desigualdad que también es válida cuando $x = z$. De lo anterior y recordando que λ es negativo, se sigue que (2.28) también es válido en este caso.

Como $z \in F$ fue arbitrario, (2.28) implica que

$$\alpha[V_{\alpha,\lambda}(x) - V_{\alpha,\lambda}(z)] \leq \frac{B}{\delta}$$

para cada $x, z \in F$, así que (2.22) es válido con

$$B^* = \frac{B}{\delta}. \tag{2.29}$$

Ahora, se probará que la constante B^* también satisface (2.23). Para comenzar, note que por (2.22) es suficiente mostrar que la desigualdad en (2.23) es válida cuando $x \in S \setminus F$. Sea $f_{\alpha,\lambda}$ la política estacionaria en (2.16), y multiplicando ambos lados de la igualdad por $e^{-\lambda(1-\alpha)V_{\alpha,\lambda}(x)}$ para obtener, mediante (1.5) y (2.15), que

$$U(\alpha V_{\alpha,\lambda}(x)) = E_x^{f_{\alpha,\lambda}}\left[U\left([C(X_0, A_0) - g_{\alpha,\lambda}(X_0)] + \alpha V_{\alpha,\lambda}(X_1)\right)\right], \quad x \in S.$$

A partir de esta igualdad, de manera similar a lo usado para la prueba de (2.22), usando inducción se sigue que, para cada entero positivo n y $x \in S \setminus F$,

$$U(\alpha V_{\alpha,\lambda}(x)) = E_x^{f_{\alpha,\lambda}}\left[U\left(\sum_{t=0}^{T_F\wedge n-1}[C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{T_F\wedge n})\right)\right],$$

así que

$$U(\alpha V_{\alpha,\lambda}(x)) = E_x^{f_{\alpha,\lambda}}\left[U\left(\sum_{t=0}^{T_F-1}[C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha V_{\alpha,\lambda}(X_{T_F})\right)\right],$$

por (2.20). Multiplicando ambos lados de esta igualdad por $e^{-\lambda\alpha V_{\alpha,\lambda}(z)}$ donde $z \in F$ es arbitrario pero fijo, por (1.5) se sigue que para cada estado x en $S \setminus F$,

$$\begin{aligned} &U\big(\alpha[V_{\alpha,\lambda}(x) - V_{\alpha,\lambda}(z)]\big) \\ &\quad = E_x^{f_{\alpha,\lambda}}\left[U\left(\sum_{t=0}^{T_F-1}[C(X_t, A_t) - g_{\alpha,\lambda}(X_t)] + \alpha[V_{\alpha,\lambda}(X_{T_F}) - V_{\alpha,\lambda}(z)]\right)\right], \end{aligned}$$

es decir,

$$\begin{aligned} & e^{\lambda\alpha[V_{\alpha,\lambda}(x)-V_{\alpha,\lambda}(z)]} \\ & = E_x^{f_{\alpha,\lambda}}\left[e^{\lambda\sum_{t=0}^{T_F-1}[C(X_t,A_t)-g_{\alpha,\lambda}(X_t)]+\lambda\alpha[V_{\alpha,\lambda}(X_{T_F})-V_{\alpha,\lambda}(z)]}\right]. \end{aligned}$$

Combinando esta igualdad con (2.15) y (2.22) se sigue que para cada $x \in S \setminus F$

$$\begin{aligned} E_x^{f^z}\left[e^{-2|\lambda|\|C\|T_F-|\lambda|B^*}\right] & \leq e^{\lambda\alpha[V_{\alpha,\lambda}(x)-V_{\alpha,\lambda}(z)]} \\ & \leq E_x^{f^z}\left[e^{2|\lambda|\|C\|T_F+|\lambda|B^*}\right], \end{aligned}$$

y por el Lema 2.3.3(i), esto lleva a

$$e^{-|\lambda|B/\delta}e^{-|\lambda|B^*} \leq e^{\lambda\alpha[V_{\alpha,\lambda}(x)-V_{\alpha,\lambda}(z)]} \leq e^{|\lambda|B/\delta}e^{|\lambda|B^*}.$$

Finalmente, usando (2.29), esta relación implica que

$$\alpha|V_{\alpha,\lambda}(x) - V_{\alpha,\lambda}(z)| \leq B^* + B/\delta = 2B^*$$

para cada $x \in S \setminus F$, completando la verificación de la desigualdad (2.23). ■

2.4. Continuidad del Índice Promedio Óptimo: $\lambda = 0$

A continuación se presenta el resultado que garantiza la continuidad de la función de valor óptimo sobre el parámetro de sensibilidad al riesgo para el caso $\lambda = 0$. En adelante, se supondrá que $z \in F$ y $\lambda \in (-\delta, \delta) \setminus \{0\}$ son arbitrarios, pero fijos, y para cada $\alpha \in (0, 1)$ la función $h_{\alpha,\lambda} \in \mathbb{B}(S)$ es dada por

$$h_{\alpha,\lambda}(x) = \alpha[V_{\alpha,\lambda}(x) - V_{\alpha,\lambda}(z)], \quad x \in S. \tag{2.30}$$

Note que, por (2.15), $h_{\alpha,\lambda}$ y $g_{\alpha,\lambda}$ satisfacen la siguiente relación:

$$g_{\alpha,\lambda}(x) - g_{\alpha,\lambda}(z) = \frac{1-\alpha}{\alpha}h_{\alpha,\lambda}(x), \tag{2.31}$$

y observe que (2.14), (2.15) y el Teorema 2.3.4 implican que

$$|g_{\alpha,\lambda}(z)| \leq \|C\| \text{ y } h_{\alpha,\lambda}(x)| \leq B, \quad x \in S, \quad \alpha \in (0, 1), \tag{2.32}$$

donde $B := 2B^*$.

De esto último, el método diagonal de Cantor [7] lleva a que existe una sucesión $(\alpha_n) \subset (0, 1)$ tal que

$$\alpha_n \nearrow 1,$$

y los siguientes límites existen:

$$\begin{aligned} \lim_{n\to\infty} g_{\alpha_n,\lambda}(z) & = g^* \in [-\|C\|, \|C\|], \\ \lim_{n\to\infty} h_{\alpha_n,\lambda}(x) & = h^*_\lambda(x) \in [-B, B], \quad x \in S; \end{aligned} \tag{2.33}$$

combinando estas dos últimas expresiones con (2.31) se sigue que

$$\lim_{n\to\infty} g_{\alpha_n,\lambda}(x) = g^*, \quad x \in S. \tag{2.34}$$

Ahora, sea $f_{\alpha,\lambda} \in \mathbb{F}$ la política estacionaria en (2.16). Ya que $\mathbb{F}$ es un espacio métrico compacto, eligiendo una subsucesión de (α_n), se puede suponer que $(f_{\alpha_n,\lambda})$ converge en $\mathbb{F}$:

$$\lim_{n\to\infty} f_{\alpha_n,\lambda}(x) = f^*_\lambda(x) \in A(x), \quad x \in S. \tag{2.35}$$

A continuación se presenta un resultado, mediante el cual se prueba que para $\lambda \in (-\delta, \delta)$ dado, (1.15) tiene una solución acotada. Este resultado puede considerarse como una extensión del presentado en [8], en el cual se supone que

(i) el espacio de estados y el conjunto de acciones son finitos,

(ii) λ es positivo, y

(iii) que la Suposición 2.1.2 es válida para $F \subset S$.

En contraste, en este trabajo se consideran espacios numerables y además no se impone restricción alguna sobre el signo de λ.

Teorema 2.4.1 *Bajo Suposiciones 1.1.1 y 2.1.2, existen números positivos δ y B tales que para cada $\lambda \in (-\delta, \delta) \setminus \{0\}$ la ecuación de optimalidad (1.15) tiene una solución $(g_\lambda, h_\lambda(\cdot))$ satisfaciendo que $\|h_\lambda\| \leq B$.*

Demostración. Se probará que la pareja $(g^*, h^*_\lambda(\cdot))$ en (2.33) es una solución de la ecuación de optimalidad (1.15).

Note que, multiplicando ambos lados de (2.13) por $e^{-\lambda[\alpha V_{\alpha,\lambda}(z)+(1-\alpha)V_{\alpha,\lambda}(x)]}$ se sigue, por (1.5), (2.15) y (2.30), que para cada $x \in S$:

$$U(h_{\alpha,\lambda}(x)) = \inf_{a\in A(x)} \left[\sum_{y\in S} p_{x\,y}(a) U(C(x,a) - g_{\alpha,\lambda}(x) + h_{\alpha,\lambda}(y)) \right], \tag{2.36}$$

y

$$U(h_{\alpha,\lambda}(x)) = \sum_{y\in S} p_{x\,y}(f_{\alpha,\lambda}(x)) U(C(x, f_{\alpha,\lambda}(x)) - g_{\alpha,\lambda}(x) + h_{\alpha,\lambda}(y)), \tag{2.37}$$

donde $f_{\alpha,\lambda}$ es la política estacionaria en (2.16). Además, observe que (1.4), (2.15) y (2.32) implican la siguiente cota:

$$\begin{aligned} |U(C(x,a) - g_{\alpha,\lambda}(x) + h_{\alpha,\lambda}(y))| &\leq e^{|\lambda|(|C(x,a)|+|g_{\alpha,\lambda}(x)|+|h_{\alpha,\lambda}(y)|)} \\ &\leq e^{|\lambda|(2\|C\|+B)}, \quad (x,a) \in \mathbb{K}. \end{aligned} \tag{2.38}$$

Sea $(x,a) \in \mathbb{K}$ fijo. Remplazando α por α_n en (2.36) se obtiene que

$$U(h_{\alpha_n,\lambda}(x)) \leq \sum_{y\in S} p_{x\,y}(a) U(C(x,a) - g_{\alpha_n,\lambda}(x) + h_{\alpha_n,\lambda}(y)),$$

y entonces, tomando el límite cuando n tiende a ∞ en ambos lados de esta desigualdad por el Teorema de Convergencia Dominada, (2.33) y (2.38) se obtiene que

$$U(h_\lambda^*(x)) \leq \sum_{y\in S} p_{x\,y}(a)U(C(x,a) - g_\lambda^* + h_\lambda^*(y)), \quad (x,a)\in\mathbb{K}. \tag{2.39}$$

Ahora, sea $G \subset S$ un conjunto finito y note que (2.37) y (2.38) llevan a

$$\begin{aligned} U(h_{\alpha_n,\lambda}(x)) \quad &\geq \quad \sum_{y\in G} p_{x\,y}(f_{\alpha_n,\lambda}(x))U(C(x,f_{\alpha_n,\lambda}(x)) - g_{\alpha_n,\lambda}(x) + h_{\alpha_n,\lambda}(y)) \\ &\quad - \sum_{y\in S\setminus G} p_{x\,y}(f_{\alpha_n,\lambda}(x))e^{|\lambda|(2\|C\|+B)} \\ &= \quad \sum_{y\in G} p_{x\,y}(f_{\alpha_n,\lambda}(x))U(C(x,f_{\alpha_n,\lambda}(x)) - g_{\alpha_n,\lambda}(x) + h_{\alpha_n,\lambda}(y)) \\ &\quad - \left(1 - \sum_{y\in G} p_{x\,y}(f_{\alpha_n,\lambda}(x))\right) e^{|\lambda|(2\|C\|+B)}, \end{aligned}$$

y entonces, haciendo que n tienda a ∞, por (2.33)-(2.35) y la Suposición 3.2.1 se tiene que

$$\begin{aligned} U(h_\lambda^*(x)) \quad &\geq \quad \sum_{y\in G} p_{x\,y}(f_\lambda^*(x))U(C(x,f_\lambda^*(x)) - g_\lambda^*(x) + h_\lambda^*(y)) \\ &\quad - \left(1 - \sum_{y\in G} p_{x\,y}(f_\lambda^*(x))\right) e^{|\lambda|(2\|C\|+B)}; \end{aligned}$$

haciendo que G se aproxime a S, la relación anterior lleva a

$$U(h_\lambda^*(x)) \geq \sum_{y\in S} p_{x\,y}(f_\lambda^*(x))U(C(x,f_\lambda^*(x)) - g_\lambda^*(x) + h_\lambda^*(y)).$$

Combinando esta desigualdad con (2.39) se sigue que la pareja $(g_\lambda^*, h_\lambda^*(\cdot))$ satisface la ecuación de optimalidad (1.15). Como $\|h_\lambda^*\| \leq B = 2B^*$y $\lambda \in (-\delta,\delta)\setminus\{0\}$ es arbitrario, esto establece la conclusión del Teorema 2.4.1. ■

Después de los preliminares anteriores, el Teorema 2.4.2 será demostrado. En adelante δ y B son los números positivos dados en el Teorema 2.4.1 y para cada $\lambda \in (-\delta,\delta)\setminus\{0\}$, la pareja $(g_\lambda, h_\lambda(\cdot)) \in \mathbb{R}\times\mathbb{B}(S)$ es una solución de la ecuación de optimalidad (1.15) satisfaciendo que

$$\|h_\lambda\| \leq B; \tag{2.40}$$

note que el Lema 1.3.1 lleva a que g_λ es el costo promedio óptimo en cada estado inicial x, así que

$$|g_\lambda| \leq \|C\|. \tag{2.41}$$

Recordando que $U(x) = \operatorname{sign}(\lambda)e^{\lambda x}$ para cada $x \in \mathbb{R}$, dividiendo ambos lados de (1.15) por $\operatorname{sign}(\lambda)\lambda = |\lambda| > 0$ y restando $1/\lambda$ en ambos lados de la igualdad obtenida, se tiene que

$$\frac{e^{\lambda(g_\lambda + h_\lambda(x))} - 1}{\lambda} = \inf_{a \in A(x)} \sum_{y \in S} p_{x\,y}(a) \frac{e^{\lambda(C(x,a) + h_\lambda(y))} - 1}{\lambda}, \tag{2.42}$$

$x \in S$, $0 < |\lambda| < \delta$; y observe que (2.40) y la Suposición 3.2.1 implican que existe $f_\lambda \in \mathbb{F}$ tal que minimiza el lado derecho de (2.42), es decir,

$$\frac{e^{\lambda(g_\lambda + h_\lambda(x))} - 1}{\lambda} = \sum_{y \in S} p_{x\,y}(f_\lambda(x)) \frac{e^{\lambda(C(x,f_\lambda(x)) + h_\lambda(y))} - 1}{\lambda}, \tag{2.43}$$

$x \in S$, $0 < |\lambda| < \delta$. Finalmente, usando que

$$\lim_{\lambda \to 0} \Delta(\lambda) = 0, \text{ donde } \Delta(\lambda) := \sup_{x:\,|x| \le \|C\| + B} \left| \frac{e^{\lambda x} - 1}{\lambda} - x \right|, \tag{2.44}$$

lo cual se deriva de la expansión de Taylor de segundo orden de la función exponencial. Con esta notación, (2.40) implica que

$$\left| \frac{e^{\lambda(C(x,a) + h_\lambda(y))} - 1}{\lambda} - [C(x,a) + h_\lambda(y)] \right| \le \Delta(\lambda).$$

El resultado principal del capítulo es a continuación establecido.

Teorema 2.4.2 *Si las Suposiciones 1.1.1 y 2.1.2 son válidas entonces para cada $x \in S$ la función $\lambda \mapsto J^*(\lambda, x)$ es continua en $\lambda = 0$.*

Demostración. Se probará que

$$\lim_{\lambda \to 0} g_\lambda = J^*(0, x), \quad x \in S. \tag{2.45}$$

Para ello, sea $\{\lambda_n\}_{n \in \mathbb{N}}$ una sucesión arbitraria tal que $\lambda_n \in (-\delta, \delta) \setminus \{0\}$, $n = 0, 1, 2, \ldots$, y

$$\lim_{n \to \infty} \lambda_n = 0. \tag{2.46}$$

Recordando que el conjunto $\mathbb{F}$ de políticas estacionarias es un espacio métrico compacto, (2.41) y (2.40) permiten usar el método diagonal de Cantor para construir una subsucesión $(\lambda_{n_k})_{k \in \mathbb{N}}$ tal que los siguientes límites existen:

$$\begin{aligned} &\lim_{k \to \infty} g_{\lambda_{n_k}} =: g_0^*, \quad \lim_{k \to \infty} h_{\lambda_{n_k}}(x) =: h_0^*(x), \\ &\lim_{k \to \infty} f_{\lambda_{n_k}}(x) =: f_0^*(x), \quad x \in S. \end{aligned} \tag{2.47}$$

Ahora, sea $(x, a) \in \mathbb{K}$ arbitrario y note que (2.42) implica que

$$\frac{e^{\lambda_{n_k}(g_{\lambda_{n_k}} + h_{\lambda_{n_k}}(x))} - 1}{\lambda_{n_k}} \le \sum_{y \in S} p_{x\,y}(a) \frac{e^{\lambda_{n_k}(C(x,a) + h_{\lambda_{n_k}}(y))} - 1}{\lambda_{n_k}},$$

una desigualdad que, usando (2.40) y (2.44), lleva a

$$g_{\lambda_{n_k}} + h_{\lambda_{n_k}}(x) - \Delta(\lambda_{n_k}) \leq \sum_{y\in S} p_{x\,y}(a)(C(x,a) + h_{\lambda_{n_k}}(y)) + \Delta(\lambda_{n_k}).$$

Entonces, como $\Delta(\lambda_{n_k}) \to 0$ cuando $k \to \infty$, tomando el límite cuando k tiende a ∞ en la desigualdad anterior, por (2.47) y el Teorema de Convergencia Dominada se tiene que

$$g_0^* + h_0^*(x) \leq \sum_{y\in S} p_{x\,y}(a)(C(x,a) + h_0^*(y)), \quad (x,a) \in \mathbb{K}. \tag{2.48}$$

Sea $x \in S$ arbitrario y observe que por (2.43) y (2.44) se llega a

$$\begin{aligned} & g_{\lambda_{n_k}} + h_{\lambda_{n_k}}(x) + \Delta(\lambda_{n_k}) \\ & \quad \geq \frac{e^{\lambda_{n_k}(g_{\lambda_{n_k}} + h_{\lambda_{n_k}}(x))} - 1}{\lambda_{n_k}} \\ & \quad \geq \sum_{y\in S} p_{x\,y}(f_{\lambda_{n_k}}(x)) \frac{e^{\lambda_{n_k}(C(x,f_{\lambda_{n_k}}(x)) + h_{\lambda_{n_k}}(y))} - 1}{\lambda_{n_k}} \\ & \quad \geq \sum_{y\in S} p_{x\,y}(f_{\lambda_{n_k}}(x))(C(x,f_{\lambda_{n_k}}(x)) + h_{\lambda_{n_k}}(y)) - \Delta(\lambda_{n_k}). \end{aligned}$$

Entonces, por (2.40), se sigue que para cualquier subconjunto finito y no vacío, $G \subset S$, que

$$\begin{aligned} g_{\lambda_{n_k}} + h_{\lambda_{n_k}}(x) \quad & \geq \quad \sum_{y\in G} p_{x\,y}(f_{\lambda_{n_k}}(x))(C(x,f_{\lambda_{n_k}}(x)) + h_{\lambda_{n_k}}(y)) \\ & - \quad (\|C\| + B)\left[1 - \sum_{y\in G} p_{x\,y}(f_{\lambda_{n_k}}(x))\right] - 2\Delta(\lambda_{n_k}); \end{aligned}$$

haciendo k tender a ∞, la Suposición 3.2.1 y (2.47) implican que

$$\begin{aligned} g_0^* + h_0^*(x) \quad & \geq \quad \sum_{y\in G} p_{x\,y}(f_0^*(x))(C(x,f_0^*(x)) + h_0^*(y)) \\ & - \quad (\|C\| + B)\left[1 - \sum_{y\in G} p_{x\,y}(f_0^*(x))\right], \end{aligned}$$

una desigualdad que incrementando G a S lleva a

$$g_0^* + h_0^*(x) \geq \sum_{y\in S} p_{x\,y}(f_0^*(x))(C(x,f_0^*(x)) + h_0^*(y)), \quad x \in S. \tag{2.49}$$

Combinando esto último con (2.48) se obtiene:

$$g_0^* + h_0^*(x) = \inf_{a\in A(x)}\left[C(x,a) + \sum_{y\in S} p_{x\,y}(a)h_0^*(y)\right], \quad x \in S, \tag{2.50}$$

mostrando que la pareja $(g_0^*, h_0^*(\cdot))$ satisface la ecuación de optimalidad neutral al riesgo correspondiente a la función de utilidad $U_0(x) = x$; ya que h_0^* es una función acotada, implica que

$$g_0^* = J^*(0, x), \quad x \in S.$$

En resumen, se ha demostrado que cada sucesión $(\lambda_n)_{n\in\mathbb{N}}$ satisfaciendo (2.46) tiene una subsucesión $(\lambda_{n_k})_{k\in\mathbb{N}}$, tal que

$$\lim_{k\to\infty} g_{\lambda_{n_k}} = J^*(0, \cdot),$$

una propiedad que es equivalente a (2.45); como $g_\lambda = J^*(\lambda, \cdot)$ cuando $0 < |\lambda| < \delta$, por el Lema 1.3.1 y el Teorema 2.4.1, lleva a

$$\lim_{\lambda\to 0} J^*(\lambda, x) = J^*(0, x)$$

para cada $x \in S$. ■

Observación 2.4.3 *Por (2.49), la política f_0^* en la prueba anterior satisface que, para cada estado x, la acción $f_0^*(x)$ es un minimizador del lado derecho de la ecuación de optimalidad en (2.50), y entonces f^* es una política óptima para el caso neutral al riesgo [23] y [33].*

2.5. Ejemplo

En las secciones anteriores, la continuidad de la función de costo promedio óptima sensible al riesgo $J^*(\lambda, \cdot)$ con respecto a λ ha sido estudiada. En la Proposición 2.2.1 se muestra que para costos acotados tal propiedad es siempre válida en $\lambda \neq 0$, pero la continuidad en cero puede no ser válida, como ocurre para el modelo presentado en el Ejemplo 2.1.1.

Bajo la condición simultánea de Doeblin, en la Suposición 2.1.2, se establece en el Teorema 2.4.2 que $J^*(\lambda, \cdot)$ es también continua en $\lambda = 0$. Note que la primera parte de la Suposición 2.1.2 asegura que, bajo cualquier política estacionaria, la clase de estados recurrentes es no vacía, pero puede haber varias clases recurrentes; si tal es el caso, la segunda parte de la Suposición 2.1.2 garantiza que, empleando otra política estacionaria, una clase recurrente dada es accesible desde cualquier otra, una condición que no se cumple en el Ejemplo 2.1.1. Por otro lado, la condición simultánea de Doeblin es un requerimiento fuerte que, mediante la existencia de una solución acotada de la ecuación de optimalidad neutral al riesgo, asegura lo siguiente:

(a) el costo promedio óptimo (neutral al riesgo) es constante, y

(b) existe una política óptima estacionaria.

Tales propiedades pueden ser garantizadas por la condición de Lyapunov (2.4), bajo la cual, la ecuación de optimalidad neutral al riesgo tiene una solución (en general) no acotada. Sin embargo, se demostró en el Ejemplo 2.1.4 que, bajo la Suposición 2.1.3, la función $\lambda \mapsto J^*(\lambda, x)$ puede tener una discontinuidad en $\lambda = 0$. En este punto, es interesante observar una característica común en los Ejemplos 2.1.1 y 2.1.4, a saber, en estos ejemplos la discontinuidad en cero de la función $\lambda \mapsto J^*(\lambda, x)$ ocurre cuando *x es un estado transitorio*. Este hecho resalta el importante papel de los estados transitorios al determinar el costo promedio sensible al riesgo [11] y, naturalmente, lleva a considerar la siguiente pregunta:

Suponga que la condición de Lyapunov es válida, y el espacio de estados es una clase comunicante bajo cada política estacionaria, de modo que ningún estado es transitorio. En este contexto, ¿es verdad que la función $\lambda \mapsto J^*(\lambda, x)$ es continua en cero para cada estado x?

Como se muestra en el siguiente ejemplo, la respuesta a esta pregunta es negativa.

Ejemplo 2.5.1 *Considere un conjunto numerable S^* cuyos elementos son denotados por x^*, donde $x = 1, 2, 3, \ldots$, y se define*

$$S = \mathbb{N} \cup S^* = \mathbb{N} \cup \{x^* \mid x = 1, 2, 3, \ldots\}.$$

La ley de transición $[p_{s,s_1}]$ con $s, s1 \in S$ es determinada como sigue: Para $x = 1, 2, 3, \ldots$,

$$\begin{aligned} 1 - p_{x,0} &= p_{x,x+1} = \frac{x^2}{(x+1)^2} = p_{x^*,(x+1)^*} = 1 - p_{x^*,0}, \\ p_{0,x^*} &= \frac{\gamma}{x^3} = p_{0,x} \end{aligned} \tag{2.51}$$

donde $\gamma^{-1} = \sum_{x=1}^{\infty} \frac{2}{x^3}$. Finalmente, se define la función de costo $C : S \to \mathbb{R}$ por

$$C(x) = 1, \quad C(x^*) = -1, \quad x = 1, 2, 3, \ldots, \text{ y } C(0) = 0; \tag{2.52}$$

haciendo al conjunto de controles, A, igual a un conjunto unitario, las cantidades anteriores determinan un PDM con una única política estacionaria, la cual no será indicada explícitamente en el análisis siguiente.

En la Proposición 2.5.2 se muestra que la Suposición 2.1.3 se satisface en este ejemplo, y además, se determina el costo promedio neutral al riesgo.

Proposición 2.5.2 *En el Ejemplo 2.5.1 las siguientes afirmaciones son válidas:*

(i) La condición de Lyapunov (2.4) se satisface con $z = 0$.

(ii) El proceso $(X_t)_{t\in\mathbb{N}}$ es irreducible, es decir, $P_w[T_y < \infty] > 0$ para cada $w, y \in S$.

(iii) La función de costo promedio neutral al riesgo es nula: $J(0, y) = 0$ para cada $y \in S$.

Demostración. (i) Dado que sólo hay una política estacionaria, es suficiente verificar que

$$E_y[T_0] < \infty, \quad y \in S. \tag{2.53}$$

Para establecer esta afirmación note que, de manera análoga al argumento utilizado en el Ejemplo 2.1.4, la definición de la ley de transición en (2.51) implica que para cada entero positivo n

$$\begin{aligned} P_x[T_0 > n] &= P_x[X_r = x + r,\, 0 \le r \le n] \\ &= \frac{x^2}{(x+n)^2} \\ &= P_{x^*}[X_r = (x+r)^*,\, 0 \le r \le n] \\ &= P_{x^*}[T_0 > n], \qquad x = 1, 2, 3, \ldots, \end{aligned} \tag{2.54}$$

y entonces

$$\begin{aligned} E_{x^*}[T_0] &= E_x[T_0] \\ &\le \sum_{n=0}^{\infty} \frac{x^2}{(x+n)^2} \\ &\le 1 + x, \qquad x = 1, 2, 3, \ldots, \end{aligned}$$

mientras que, por la propiedad de Markov,

$$\begin{aligned} E_0[T_0] &= 1 + \sum_{x=1}^{\infty} (p_{0,x} E_x[T_0] + p_{0,x^*} E_{x^*}[T_0]) \\ &\le 1 + \gamma \sum_{x=1}^{\infty} [2(1+x)/x^3] < \infty. \end{aligned}$$

(ii) Observando que $P_0[X_1 = y] > 0$ para cada $y \in S \setminus \{0\}$, la irreducibilidad del proceso de estados se obtiene de (2.53).

(iii) Combinando la parte anterior y (2.53) se implica que la ley de transición en (2.51) tiene una única distribución invariante $(\nu_y)_{y \in S}$ y entonces el teorema ergódico implica que, para cada $y \in S$,

$$J(0, y) = \lim_{n \to \infty} \frac{1}{n} E_y \left[\sum_{t=0}^{n-1} C(X_t) \right] = \sum_{x \in S} \nu_x C(x).$$

Por otro lado, usando (2.51) no es difícil ver que $\nu_x = \nu_{x^*}$ para cada $x = 1, 2, 3, \ldots$, una propiedad que junto con (2.52) lleva a que $J(0, \cdot) = 0$. ■

A continuación, la función de costo promedio correspondiente a un coeficiente de sensibilidad al riesgo no nulo será determinada.

Proposición 2.5.3 *Para el modelo en el Ejemplo 2.5.1 las siguientes propiedades son válidas.*

(i) *Para cada $\lambda \neq 0$, el costo promedio λ-sensible $J(\lambda, \cdot)$ es constante.*

(ii) *Si $\lambda > 0$ entonces $J(\lambda, \cdot) = 1$, y*

(iii) *$J(\lambda, \cdot) = -1$ para cada $\lambda < 0$.*

Demostración. (i) Sean $\lambda \in \mathbb{R} \setminus \{0\}$ y $w, y \in S$ arbitrarios. Usando la Proposición 2.5.2(ii), elegimos $k \in \mathbb{N}$ tal que $P_w[X_k = y] > 0$ y note que, para $n > k$ la propiedad de Markov y (2.52) implican que

$$\begin{aligned} e^{\lambda J_n(\lambda,w)} &= E_w\left[e^{\lambda \sum_{t=0}^{n-1} C(X_t)}\right] \\ &\geq E_w\left[e^{\lambda \sum_{t=0}^{k-1} C(X_t)} I[X_k = y] e^{\lambda \sum_{t=k}^{n-1} C(X_t)}\right] \\ &\geq e^{-k|\lambda|} P_w[X_k = y] E_y\left[e^{\lambda \sum_{t=0}^{n-k-1} C(X_t)}\right] \\ &= e^{-k|\lambda|} P_w[X_k = y] e^{\lambda J_{n-k}(\lambda,y)} \end{aligned}$$

y entonces

$$\lambda \frac{J_n(\lambda, w)}{n} \geq \frac{\log(e^{-k|\lambda|} P_w[X_k = y])}{n} + \lambda \frac{J_{n-k}(\lambda, y)}{n}. \tag{2.55}$$

Si $\lambda > 0$, tomando el límite superior cuando n tiende a ∞ en ambos lados de esta desigualdad, se sigue que

$$\lambda J(\lambda, w) \geq \lambda J(\lambda, y); \tag{2.56}$$

cuando $\lambda < 0$, si tomamos el límite inferior cuando n tiende a ∞ en (2.55), la relación en (2.56) también es válida. Por lo tanto, ya que los estados w y y son arbitrarios, (2.56) implica que $J(\lambda, \cdot)$ es constante para cada coeficiente de sensibilidad al riesgo λ no nulo.

(ii) Suponga que $\lambda > 0$. Usando que $C(x) = 1$ para cada $x = 1, 2, 3, \ldots$, se sigue que para cada entero positivo n

$$\begin{aligned} e^{\lambda J_n(\lambda,1)} &= E_1\left[e^{\lambda \sum_{t=0}^{n-1} C(X_t)}\right] \\ &\geq E_1\left[e^{\lambda \sum_{t=0}^{n-1} C(X_t)} I[X_r = 1 + r, 0 \leq r < n]\right] \\ &= e^{\lambda n} P_1\left[X_r = 1 + r,\ 0 \leq r < n\right] \\ &= e^{\lambda n} \frac{1}{n^2}, \end{aligned}$$

donde (2.54) se utilizó para establecer la última igualdad. Así,

$$\frac{J_n(\lambda, 1)}{n} \geq 1 - \frac{2\log(n)}{\lambda n},$$

y entonces

$$J(\lambda, 1) = \limsup_{n \to \infty} \frac{J_n(\lambda, 1)}{n} \geq 1;$$

recordando que $C(\cdot) \leq 1$, se sigue que $J(\lambda, 1) = 1$ de donde $J(\lambda, \cdot) = 1$, por la parte (i).

(*iii*) Suponga que $\lambda < 0$. Usando que $C(x^*) = -1$ para cada $x = 1, 2, 3, \ldots$ y observe que para cada entero positivo n

$$\begin{aligned} e^{\lambda J_n(\lambda, 1^*)} &= E_{1^*}\left[e^{\lambda \sum_{t=0}^{n-1} C(X_t)}\right] \\ &\geq E_{1^*}\left[e^{\lambda \sum_{t=0}^{n-1} C(X_t)} I[X_r = (x+r)^*, 0 \leq r < n]\right] \\ &= e^{-\lambda n} P_{1^*}\left[X_r = (x+r)^*, 0 \leq r < n\right] \\ &= e^{-\lambda n} \frac{1}{n^2}; \end{aligned}$$

como λ es negativa, de esta relación se sigue

$$\frac{J_n(\lambda, 1^*)}{n} \leq -1 + \frac{2\log(n)}{n\lambda},$$

una relación que combinada con la desigualdad $C(\cdot) \geq -1$ lleva a

$$J(\lambda, x) = \limsup_{n\to\infty} \frac{J_n(\lambda, x)}{n} = -1;$$

de esto, la parte (i) implica que

$$J(\lambda, \cdot) = -1.$$

■

Para el modelo dado en el Ejemplo 2.5.1, la condición de Lyapunov dada en (2.4) es válida y, cada estado es positivo recurrente, por las partes (i) y (ii) de la Proposición 2.5.2. Sin embargo, las afirmaciones (ii) y (iii) de la Proposición 2.5.3 y la tercera parte de la Proposición 2.5.2 implican que, para cada estado y, la función $\lambda \mapsto J(\lambda, y)$ no es continua en $\lambda = 0$, ni por la izquierda ni por la derecha.

Capítulo 3

Procesos Semi Markovianos

En este capítulo se presentan algunos resultados relacionados con los Procesos de Decisión Semi Markovianos [23], [29], [33], [38] [39], éstos serán únicamente enunciados y servirán para establecer los resultados obtenidos en el Capítulo 4.

3.1. Procesos de Decisión Semi Markovianos

Un modelo de control semi markoviano (MCSM) es un sistema dinámico cuyos estados cambian en tiempos aleatorios bajo la influencia de un controlador.

Un MCSM consiste de una séxtupla

$$\mathcal{S} = \left(S, A, \{A(x)\}_{x\in S}, C, \{\rho_{x,a}(\cdot)\}_{x\in S, a\in A(x)}, \{F_{x,a}\}_{x\in S, a\in A(x)}, P\right),$$

donde

- S es un conjunto numerable dotado de la topología discreta y es llamado espacio de *estados*;
- A es un espacio métrico, llamado espacio de *acciones o controles*;
- $\{A(x)\}_{x\in S}$ es una familia de subconjuntos medibles, no vacíos $A(x)$ de A, donde $A(x)$ denota el conjunto de *acciones admisibles* cuando el sistema se encuentra en el estado $x \in S$. El conjunto $\mathbb{K}$ de parejas de estados acciones admisibles, está definido por

  $$\mathbb{K} = \{(x, a)\, | x \in S, a \in A(x)\},$$

  y se supondrá que es un conjunto medible del espacio producto $S \times A$;
- $C : \mathbb{K} \to \mathbb{R}$ es una función medible y se llama la *función de costo en un paso*.

- Para cada $(x,a) \in \mathbb{K}$ la función $\rho_{x,a}\colon [0,\infty) \to \mathbb{R}$ es la tasa de costo de permanencia correspondiente a la aplicación del control $a \in A(x)$ en el estado x.
- $F_{x,a}(\cdot)$ denota a la función de distribución del tiempo de espera en el estado x cuando el control a es aplicado; se supondrá que $F_{x,a}$ tiene soporte sobre el rayo positivo, es decir

$$F_{x,a}(0) = 0, \quad (x,a) \in \mathbb{K}. \tag{3.1}$$

- Finalmente, $P = [p_{x,y}(a)]$ es la ley de transición y satisface que $\sum_{y \in S} p_{x,y}(a) = 1$ para cada $(x,a) \in \mathbb{K}$.

La dinámica que describe a este sistema estocástico funciona de la siguiente forma: si en la n-ésima transición el sistema se encuentra en el estado $X_n = x \in S$, y la acción $A_n = a \in A(x)$ es aplicada; entonces ocurre lo siguiente:

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

(ii) El sistema permanece en x durante un tiempo (aleatorio) de espera S_n cuya función de distribución está dada por $F_{x,a}$.

(iii) Un costo de permanencia con tasa $\rho_{x,a}$ es incurrido, mientras el sistema permanece en x.

(iv) Sin importar la historia del proceso hasta el tiempo T_n (ver (3.4)), una vez que S_n ha transcurrido el sistema transita a otro estado $X_{n+1} = y \in S$ con probabilidad $p_{x,y}(a)$.

Una vez hecha esta transición a un nuevo estado, se elige una nueva acción y la dinámica anteriormente descrita se repite.

Observación 3.1.1 *Se puede ver en (iv) que el proceso de decisión satisface una propiedad de tipo Markov.*

Note que la $n-$ésima transición es completada al tiempo T_n, donde

$$T_0 = 0 \quad \text{y} \quad T_n = \sum_{i=0}^{n-1} S_i, \quad n = 1, 2, 3, \ldots, \tag{3.2}$$

de este modo el número de transiciones N_t en el intervalo $[0,t]$ es dado por

$$N_t = \sup\{n \in \mathbb{N} : T_n \leq t\}, \quad t \geq 0. \tag{3.3}$$

Por otro lado, la historia (información) registrada por el controlador hasta el tiempo T_n está dada por $\mathcal{H}_n$, donde

$$\mathcal{H}_0 = X_0, \text{ y } \mathcal{H}_n = (X_0, A_0, S_0, \ldots, X_{n-1}, A_{n-1}, S_{n-1}, X_n), \quad n \geq 1, \tag{3.4}$$

así que para cada $t \geq 0$ y $n \in \mathbb{N}$, se tiene que T_n es $\sigma(\mathcal{H}_n)$–medible, y

$$[N_t \geq n] = [T_n \leq t] \in \sigma(\mathcal{H}_n). \tag{3.5}$$

Políticas. Considere un modelo de control semi markoviano y defina para cada $n = 0, 1, 2, \ldots$, al espacio $\mathbb{H}_n$ de historias admisibles hasta el término de la $n-$ésima transición por

$$\begin{aligned}
&\mathbb{H}_0 := S,\\
&\mathbb{H}_n := \mathbb{K} \times (0, \infty) \times \mathbb{H}_{n-1}, \ \textit{para } n = 1, 2, 3, \ldots.
\end{aligned}$$

Un elemento de $\mathbb{H}_n$ es denotado por

$$h_n = (x_0, a_0, s_0, x_1 \ldots, x_{n-1}, a_{n-1}, s_{n-1}, x_n),$$

donde $x_i \in S$, $a_i \in A(x_i)$ y $s_i > 0$.

Definición 3.1.2 *Una política $\pi = \{\pi_n\}$ es una sucesión de kérneles estocásticos, donde cada π_n está definido sobre A dado $\mathbb{H}_n$ y satisface que $\pi_n(A(x_n)|\, h_n) = 1$ para toda $h_n \in \mathbb{H}_n$ y $n = 0, 1, 2, \ldots$. El conjunto de todas las políticas será denotado por $\mathcal{P}$.*

De acuerdo con esta definición, una política $\pi = \{\pi_n\}$ puede interpretarse como una regla para la elección de acciones, la cual en cada época de decisión T_n puede depender del estado actual así como de los estados, controles y tiempos de espera previos.

Dado un estado inicial $X_0 = x$ y la política $\pi \in \mathcal{P}$ usada para elegir acciones, la distribución del proceso $\{(X_n, A_n, S_n)\}_{n\in\mathbb{N}}$ está determinada de manera única vía el Teorema de Ionescu-Tulcea; ver por ejemplo [2]. Tal distribución es denotada por P_x^π, mientras que E_x^π denota al correspondiente operador esperanza. Las siguientes relaciones de Markov se satisfacen casi seguramente, bajo P_x^π: para cada $x, y \in S$, $\tilde{A} \in \mathcal{B}(A)$ (donde $\mathcal{B}(A)$ denota a la σ-álgebra de Borel de A) y $n \in \mathbb{N}$,

$$\begin{aligned}
P_x^\pi[X_0 = x] &= 1,\\
P_x^\pi[A_n \in \tilde{A} \,|\, \mathcal{H}_n] &= \pi_n(\tilde{A} \,|\, \mathcal{H}_n),\\
P_x^\pi[S_n \leq t \,|\, \mathcal{H}_n, A_n] &= F_{X_n, A_n}(t), \quad t \geq 0,\\
P_x^\pi[X_{n+1} = y \,|\, \mathcal{H}_n, A_n, S_n] &= p_{X_n, y}(A_n).
\end{aligned} \tag{3.6}$$

El tiempo promedio de espera en el estado x cuando la acción a es elegida está dado por

$$\tau(x, a) := \int_0^\infty t F_{x,a}(dt), \ \ (x, a) \in \mathbb{K}.$$

3.2. Criterio de Rendimiento y Problema de Control Óptimo

Dado un modelo de control semi markoviano, para cada $x \in S$ y $\pi \in \mathcal{P}$ definimos al criterio de costo promedio esperado como

$$J(\pi, x) = \limsup_{n\to\infty} \frac{E_x^\pi\left[\sum_{t=0}^{n-1} C(x_t, a_t)\right]}{E_x^\pi[T_n]}, \tag{3.7}$$

y,

$$J^*(x) = \inf_{\pi\in\mathcal{P}} J(\pi, x) \tag{3.8}$$

es el costo promedio óptimo en x; una política $\pi^* \in \mathcal{P}$ es óptima si $J(\pi^*, x) = J^*(x)$ para cada $x \in S$.

Usando propiedades de la esperanza condicional, se puede escribir (3.7) como

$$J(\pi, x) = \limsup_{n\to\infty} \frac{E_x^\pi\left[\sum_{t=0}^{n-1} C(x_t, a_t)\right]}{E_x^\pi\left[\sum_{t=0}^{n-1} \tau(x_t, a_t)\right]}.$$

En efecto,

$$\begin{aligned} E_x^\pi[T_n] &= E_x^\pi\left[\sum_{t=0}^{n-1} S_t\right] \\ &= \sum_{t=0}^{n-1} E_x^\pi\left[E_x^\pi(S_t \mid \mathcal{H}_t, a_t)\right] \\ &= \sum_{t=0}^{n-1} E_x^\pi\left[\int_0^\infty t F_{x_t, a_t}(dt)\right] \\ &= E_x^\pi\left[\sum_{t=0}^{n-1} \tau(x_t, a_t)\right]. \end{aligned}$$

Por lo que el problema de control óptimo planteado en (3.8) consiste en determinar una política que minimice al criterio de rendimiento.

Las siguientes suposiciones permiten asegurar la existencia de una política que minimiza al criterio de rendimiento.

Suposición 3.2.1 (*i*) *Para cada $x \in S$, $A(x)$ es un subconjunto compacto de A.*

(*ii*) *Para cada $x, y \in S$, las funciones $a \mapsto C(x,a)$ y $a \mapsto p_{xy}(a)$ son continuas para cada $a \in A(x)$.*

iii) *Para cada $x \in S$, $\tau(x, \cdot)$ es continua sobre $A(x)$, y existen constantes positivas m y M tales que*

$$m \leq \tau(x, a) \leq M$$

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

A continuación se presentan a la ecuación de optimalidad y el resultado que garantiza la existencia de una política óptima, además de caracterizar al índice promedio óptimo.

Definición 3.2.2 *Una pareja* $(g, h(\cdot))$ *donde* g *es un número real y* h *es una función medible y acotada, será solución de la ecuación de optimalidad si para cada* $x \in S$,

$$h(x) = \min_{a \in A(x)} \left\{ C(x, a) - g\tau(x, a) + \int h(y)\, p(dy\,|x, a) \right\}.$$

La prueba del siguiente resultado puede ser consultada en: [23], [29], [33] ó [39].

Teorema 3.2.3 *Considere que la Suposición 3.2.1 es válida. Si la pareja* $(g, h(\cdot))$ *es solución de la ecuación de optimalidad 3.2.2. Entonces existe una política* $f \in \mathbb{F}$ *tal que minimiza el lado derecho de 3.2.2, es decir,*

$$h(x) = C(x, f) - g\tau(x, f) + \int h(y)\, p(dy\,|x, f).$$

Entonces, f *es una política óptima y se tiene que* $J^*(x) = g$.

De manera análoga a lo presentado en el Sección 1.2, a continuación se relacionan a los procesos de decisión semi markovianos con el concepto de sensibilidad al riesgo.

3.3. Problemas de Control y Sensibilidad al Riesgo

Considerando lo expuesto en la Sección 1.2 con relación a la sensibilidad al riesgo del controlador y la función de utilidad exponencial, procederemos a establecer al criterio de rendimiento utilizado a lo largo de los restantes capítulos.

Note que la teoría desarrollada en la Sección 3.1 corresponde con el caso neutral al riesgo, y en adelante será referido así.

Por tanto, considerando lo expuesto anteriormente, el problema de control en el caso sensible al riesgo queda de la siguiente manera:

Supóngase que el sistema ha sido conducido por el controlador hasta un tiempo $t > 0$, así que en los tiempos de arribo

$$T_0, T_1, \ldots, T_{N_t},$$

los estados siguientes serán visitados:

$$X_0, X_1, \ldots, X_{N_t},$$

respectivamente. Luego, para cada entero no negativo $k \leq N_t$, el control A_k será aplicado en el estado X_k, incurriendo en un costo $C(X_k, A_k)$ inmediato. Para los costos de permanencia, note que para $k < N_t$, el sistema permanecerá en X_k durante el intervalo $[T_k, T_{k+1}) \subset [0, t]$. Como $T_{k+1} = T_k + S_k$, se sigue que el sistema permanecerá en X_k durante S_k unidades de tiempo, incurriendo en el costo de permanencia dado por

$$\int_0^{S_k} \rho_{X_k, A_k}(r)\, dr.$$

Por otro lado, al tiempo T_{N_t} el sistema transita al estado X_{N_t}, y permanece ahí durante el intervalo $[T_{N_t}, t]$, ya que la siguiente transición ocurre en el tiempo $T_{N_t+1} > t$; por tanto, dentro del intervalo de observación $[0, t]$, el sistema permanece en X_{N_t} un intervalo de longitud $t - T_{N_t}$, con el correspondiente costo de permanencia

$$\int_0^{t-T_{N_t}} \rho_{X_{N_t}, A_{N_t}}(r)\, dr.$$

Por lo tanto, el costo total incurrido hasta el tiempo $t > 0$ está dado por

$$\begin{aligned} \mathcal{C}_t \quad := \quad & \sum_{k=0}^{N_t-1} \left[C(X_k, A_k) + \int_0^{S_k} \rho_{X_k, A_k}(r)\, dr \right] \\ & + C(X_{N_t}, A_{N_t}) + \int_0^{t-T_{N_t}} \rho_{X_{N_t}, A_{N_t}}(r)\, dr. \end{aligned} \tag{3.9}$$

Sin importar el estado inicial y la política empleada, se mostrará en el Lema 4.2.1 dado más adelante, que N_t es finito P_x^π-casi seguramente para cada $t > 0$, $x \in S$ y $\pi \in \mathcal{P}$, entonces $\mathcal{C}_t$ también lo es, por la Suposición 4.1.1, dada en el Capítulo 2.

En adelante se supondrá que el controlador tiene un coeficiente de sensibilidad al riesgo constante $\lambda > 0$, es decir, el controlador es averso al riesgo; y que evaluará el costo aleatorio Y mediante $E[U(Y)]$ donde $U(\cdot)$ es la función de utilidad exponencial dada en (1.4).

Ahora, suponga que el estado inicial es $X_0 = x \in S$ y que el sistema ha sido llevado por el controlador hasta un tiempo $t > 0$ usando la política $\pi \in \mathcal{P}$. El costo total incurrido en el intervalo de tiempo $[0, t]$ es dado en (3.9), y en lugar de enfrentar la cantidad aleatoria $\mathcal{C}_t$ el controlador acepta pagar el equivalente seguro

$$J_{t,\lambda}(x, \pi) = \frac{1}{\lambda} \log\left(E_x^\pi \left[e^{\lambda \mathcal{C}_t} \right] \right).$$

El criterio de costo promedio (long-run) λ-sensible en el estado x bajo la política π estará dado por:

$$J_\lambda(x, \pi) := \limsup_{t \to \infty} \frac{1}{t} J_{t,\lambda}(x, \pi).$$

El costo promedio óptimo λ-sensible en el estado x es

$$J_\lambda^*(x) := \inf_{\pi \in \mathcal{P}} J_\lambda(x, \pi),$$

y una política $\pi^* \in \mathcal{P}$ es λ-promedio óptima si $J_\lambda(x, \pi) = J_\lambda^*(x)$ para cada estado x.

En el capítulo siguiente se desarrolla la teoría referente a los procesos de decisión semi markovianos sensibles al riesgo y se presenta a la ecuación de optimalidad mediante la cual se caracteriza al índice promedio óptimo.

Capítulo 4

Ecuación de Optimalidad

En este capítulo se procederá a establecer una ecuación de optimalidad correspondiente al criterio promedio sensible al riesgo para los procesos de decisión semi markovianos. Bajo el supuesto de que existe una solución para tal ecuación se caracterizará a la función de costo promedio óptima y, se determinará una política óptima. Por tanto, también se presenta un resultado que garantiza la existencia de una solución a la ecuación de optimalidad. Los resultados presentados en este capítulo se encuentran basados en [17]. Dicho trabajo es una extensión del artículo [10], el cual estudia el caso no controlado.

4.1. Ecuación de Optimalidad

Considere al modelo $\mathcal{S} = (S, A, \{A(x)\}, C, \{\rho_{x,a}(\cdot)\}, \{F_{x,a}\}, [p_{x,y}(\cdot)])$ dado en el Capítulo 3, se supondrá que el espacio de estados S es un conjunto finito dotado de la topología discreta, y que el conjunto de acciones A es un espacio métrico.

En adelante se supondrá que el modelo $\mathcal{S}$ satisface las siguientes condiciones,

Suposición 4.1.1 *(i) y (ii) de la Suposición 3.2.1, además*

(iii) La familia $\{F_{x,a}\}_{(x,a)\in\mathbb{K}}$ tiene soporte sobre un intervalo compacto y es débilmente continua, es decir, existe $B > 0$ tal que

$$F_{x,a}(B) = 1, \quad (x, a) \in \mathbb{K}, \tag{4.1}$$

y para cada $x \in S$ y $u \in \mathbb{B}([0, B])$, la función $a \mapsto \int_0^B u(s)\, dF_{x,a}(S)$ es continua en $a \in A(x)$.

(iv) Para cada $x \in S$, la función $(a, s) \mapsto \rho_{x,a}(s)$ es continua para cada $(a, s) \in A(x) \times [0, B]$.

Observación 4.1.2 *Observe que para cada $x \in S$, el conjunto $A(x) \times [0, B]$ es un conjunto compacto ya que $A(x)$ lo es, y entonces la parte (iv) de la suposición* *anterior implica que*

$$\sup_{(a,s)\in A(x)\times[0,B]} |\rho_{x,a}(s)| < \infty.$$

Usando que el espacio de estados es finito, se sigue que

$$B_\rho := \sup_{(x,a)\in\mathbb{K},\ s\in[0,B]} |\rho_{x,a}(s)| < \infty. \tag{4.2}$$

Recordando lo planteado en la Sección 3.3,

$$J_{t,\lambda}(x,\pi) = \frac{1}{\lambda}\log\left(E_x^\pi\left[e^{\lambda\mathcal{C}_t}\right]\right).$$

El criterio de costo promedio (long-run) λ-sensible en el estado x bajo la política π estará dado por:

$$J_\lambda(x,\pi) = \operatorname*{limsup}_{t\to\infty}\frac{1}{t}J_{t,\lambda}(x,\pi).$$

El costo promedio óptimo λ-sensible en el estado x es

$$J_\lambda^*(x) = \inf_{\pi\in\mathcal{P}} J_\lambda(x,\pi),$$

y una política $\pi^*\in\mathcal{P}$ es λ-promedio óptima si $J_\lambda(x,\pi) = J_\lambda^*(x)$ para cada estado x.

El principal objetivo de este capítulo es determinar una ecuación de optimalidad cuya solución caracterice a la función de costo promedio óptima $J_\lambda^*(\cdot)$ y por tanto produzca una política óptima. Sin embargo, tal objetivo conlleva dos problemas por sí mismo:

i) Primero, se debe probar que J_λ^* puede ser obtenido como una solución de la ecuación de optimalidad, y

ii) Segundo, deben darse condiciones que aseguren la existencia de una solución para la ecuación de optimalidad

El primero de estos problemas será establecido bajo la Suposición 4.1.1, mientras que el segundo será obtenido mediante una propiedad de comunicación dada en la Suposición 4.1.3, la cual como será discutido más adelante, es necesaria para obtener una caracterización general del costo promedio óptimo λ-sensible en términos de una sola ecuación.

A continuación se presenta a la ***Ecuación de Optimalidad*** usada a lo largo del capítulo, dicha ecuación representa el primer aporte de este capítulo. Para cada $x\in S$,

$$e^{\lambda h(x)} = \inf_{a\in A(x)} E_x^\pi\left[e^{\lambda\left[C(X_0,A_0)+\int_0^{S_0}\rho_{X_0,A_0}(t)dt - gS_0 + h(X_1)\right]}\middle| X_0 = x, A_0 = a\right], \tag{4.3}$$

donde g es un número real y $h(\cdot)$ es una función real definida sobre el espacio de estados S. Usando las relaciones de Markov dadas en (3.6), la igualdad (4.3) puede escribirse para cada $x \in S$, como

$$e^{\lambda h(x)} = \inf_{a \in A(x)} \left[e^{\lambda C(x,a)} \int_0^B e^{\lambda \left[\int_0^s \rho_{x,a}(t)dt - gs \right]} dF_{x,a}(s) \sum_{y \in S} p_{x,y}(a) e^{\lambda h(y)} \right]. \tag{4.4}$$

Se tiene que mientras en la ecuación de optimalidad para el caso neutral al riesgo únicamente se considera a la esperanza de $\int_0^{S_0} \rho_{x,a}(t)dt$ y S_0 [29], [39], en el contexto de sensibilidad al riesgo se involucra a la función de distribución $F_{x,a}$. Una razón para esta diferencia es, como ya fue señalado en [10], la no linealidad de la función $Y \mapsto \mathcal{E}_\lambda[Y]$.

Además de la Suposición 4.1.1, se supondrá la siguiente condición:

Suposición 4.1.3 *La cadena de Markov $\{X_n\}$ es comunicante bajo cada política estacionaria, es decir, dados $f \in \mathbb{F}$ y $x, y \in S$, existen un entero positivo $n_{x,y,f} \equiv n$ y estados $x_k, 0 \leq k \leq n$, tales que*

$$x_0 = x, \; x_n = y, \quad y \; p_{x_{k-1},x_k}(f(x_{k-1})) > 0 \quad para \; 1 \leq k \leq n.$$

La importancia de las Suposiciones 4.1.1 y 4.1.3 en el desarrollo del capítulo es discutida a continuación:

Observación 4.1.4 *(i) Considere el caso en el que los tiempos de permanencia S_k son idénticamente 1, es decir, las funciones de distribución $F_{x,a}$ satisfacen que $F_{x,a}(1) = 1$ y $F_{x,a}(t) = 0$ para $t < 1$. En este contexto, en [1] y [9] se probó que si bajo la acción de una política estacionaria la cadena de Markov $\{X_n\}$ tiene estados transitorios, o más de una clase recurrente, entonces la función costo promedio óptimo $J^*_\lambda(\cdot)$ no es necesariamente constante, y entonces (4.4) no tiene solución [1], [9]. Por lo tanto, la Suposición 4.1.3 es necesaria para establecer un resultado general que asegure la existencia de soluciones a la ecuación de optimalidad (4.4).*

(ii) Dada una solución $(g, h(\cdot))$ de la ecuación de optimalidad (4.4), la idea detrás de la prueba del Teorema 4.3.1 es comparar el equivalente seguro del costo total relativo $\mathcal{C}_t - tg$, incurrido hasta un tiempo $t > 0$, con el correspondiente a $\tilde{\mathcal{C}}_t - gT_{N_t+1}$, donde $\tilde{\mathcal{C}}_t$ es el costo total incurrido hasta completar la primera transición posterior a t, lo cual ocurre en el tiempo T_{N_t+1}. Tal comparación es posible gracias a la condición (4.1), otra ventaja es que permite presentar los resultados obtenidos de una manera simplificada. En el caso opuesto, es decir, la condición (4.1) no es válida, es necesario introducir un concepto más: el de la distribución condicional sobre los tiempos de permanencia $S_n - r$ dado que $S_n > r$ para $r > 0$. Tal distribución fue usada en [10], sin embargo, tal supuesto propicia la necesidad de incluir otros resultados relacionados con la distribución condicional.

4.2. Distribución del Número de Transiciones

En esta sección se presentan las herramientas necesarias para establecer la prueba de los resultados principales: Teoremas 4.3.1 y Teorema 4.3.4. Estas herramientas están relacionadas con el número de transiciones, N_t, que han ocurrido en el intervalo $[0,t]$ para $t>0$ y una desigualdad, la cual involucra a la ecuación de optimalidad.

Lema 4.2.1 *Bajo la Suposición 4.1.1 las siguientes afirmaciones son válidas:*

(i) *Sea $x \in S$ arbitrario y suponga que la función $R : A(x) \times [0,B] \to \mathbb{R}$ es continua. En este caso, la función*

$$a \mapsto \int_0^B R(a,s)dF_{x,a}(s)$$

es continua para cada $a \in A(x)$.

(ii) *Dado $\alpha \in (0,1)$, existe un entero $m_\alpha > 0$ tal que*

$$\int_0^B e^{-\mu s}\, dF_{x,a}(s) \le \alpha, \quad (x,a) \in \mathbb{K}, \quad \mu \ge m_\alpha.$$

(iii) *Para cada $\alpha \in (0,1)$, $t \ge 0$ y $n \in \mathbb{N}$,*

$$P_x^\pi\left[N_t \ge n\right] \le \alpha^n e^{m_\alpha t}, \quad x \in S, \quad \pi \in \mathcal{P},$$

donde m_α es como en la parte (ii), y entonces

$$P_x^\pi[N_t < \infty] = 1. \tag{4.5}$$

Demostración. (i) Sean un estado arbitrario x y $\{a_n\} \subset A(x)$ una sucesión convergente, digamos

$$a^* = \lim_{n\to\infty} a_n,$$

y note que $R(a^*,\cdot)$ es continua en $[0,B]$, así que

$$\lim_{n\to\infty}\int_0^B R(a^*,s)dF_{x,a_n}(s) = \int_0^B R(a^*,s)dF_{x,a^*}(s),$$

por la Suposición 4.1.1(iii). Observe que $\|R(a_n,\cdot)-R(a^*,\cdot)\| \to 0$, por el Lema del Tubo (véase, por ejemplo Teorema 26.8 en [30]), y entonces

$$\left|\int_0^B R(a_n,s)dF_{x,a_n}(s) - \int_0^B R(a^*,s)dF_{x,a_n}(s)\right| \le \|R(a_n,\cdot)-R(a^*,\cdot)\| \to 0.$$

Usando la desigualdad del triángulo, las dos últimas expresiones implican que

$$\lim_{n\to\infty}\int_0^B R(a_n,s)dF_{x,a_n}(s) = \int_0^B R(a^*,s)dF_{x,a^*}(s).$$

(ii) Sean $x \in S$ y $\alpha \in (0,1)$ arbitrarios, pero fijos. Para cada $n \in \mathbb{N}$ se define $g_n : A(x) \to (0,\infty)$ por

$$g_n(a) = \int_0^B e^{-ns} dF_{x,a}(s), \quad a \in A(x),$$

y note que $g_n(\cdot)$ es continua, por la Suposición 4.1.1(iii), mientras que (3.1) y el Teorema de Convergencia Dominada implican lo siguiente:

$$\lim_{n\to\infty} g_n(a) = 0$$

para cada $a \in A(x)$.

Como $g_n(\cdot) \geq g_{n+1}(\cdot)$, la compacidad del conjunto de acciones $A(x)$ lleva, por el Teorema de Dini (véase, por ejemplo Teorema 8.2.6 en [3]), a

$$\lim_{n\to\infty} \|g_n(\cdot)\| = 0,$$

así que existe un entero $m_{x,\alpha}$ satisfaciendo

$$g_{m_{x,\alpha}}(a) = \int_0^B e^{-m_{x,\alpha}s} dF_{x,a}(s) \leq \alpha, \quad a \in A(x).$$

Recordando que el espacio de estados es finito, se obtiene que

$$m_\alpha := \max\{m_{x,\alpha} \,|\, x \in S\}$$

es un número finito y la expresión anterior lleva a

$$\int_0^B e^{-\mu s} dF_{x,a}(s) \leq \alpha$$

para cada $(x,a) \in \mathbb{K}$ cuando $\mu \geq m_\alpha$.

(iii) Sean $x \in S$, $\pi \in \mathcal{P}$ y $\alpha \in (0,1)$ arbitrarios. Dado un entero positivo n, note que la tercera igualdad en las relaciones de Markov (3.6) implican que, condicionando sobre $X_k, A_k, k = 0,1,2,\ldots,n-1$, los tiempos de permanencia $S_0, S_1, \ldots, S_{n-1}$ son independientes con función de distribución $F_{X_0,A_0}, F_{X_1,A_1}, \ldots, F_{X_{n-1},A_{n-1}}$, respectivamente. Entonces, eligiendo m_α como en la parte (ii), se sigue que

$$\begin{aligned} E_x^\pi \left[e^{-m_\alpha[S_0+S_1+\cdots+S_{n-1}]} \middle| X_k, A_k, 0 \leq k < n \right] &= \prod_{k=0}^{n-1} \int_0^B e^{-m_\alpha s} dF_{X_k,A_k}(s) \\ &\leq \alpha^n, \end{aligned}$$

y entonces $E_x^\pi\left[e^{-m_\alpha T_n}\right] \leq \alpha^n$; ver (3.2). Esta relación lleva a que, para cualquier $t \geq 0$,

$$e^{-m_\alpha t} P_x^\pi[T_n \leq t] \leq E_x^\pi\left[e^{-m_\alpha T_n}\right] \leq \alpha^n.$$

Usando la igualdad $[N_t \geq n] = [T_t \leq t]$ en (3.5), se sigue

$$P_x^\pi[N_t \geq n] \leq \alpha^n e^{m_\alpha t}$$

para cada $n \in \mathbb{N}$. Como $\alpha \in (0,1)$ se tiene que

$$P_x^\pi[N_t = \infty] = \lim_{n\to\infty} P_x^\pi[N_t \geq n] = 0,$$

estableciendo (4.5). ■

Teorema 4.2.2 *Sea $(g, h(\cdot))$ una solución de la ecuación (4.4). Bajo la Suposición 4.1.1, las afirmaciones siguientes son válidas:*

(i) La siguiente desigualdad es válida para cualesquiera $x \in S$, $\pi \in \mathcal{P}$ y $t > 0$:

$$e^{\lambda h(x)} \leq E_x^\pi \left[e^{\lambda \left[\sum_{k=0}^{N_t} \left(C(X_k, A_k) + \int_0^{S_k} \rho_{X_k, A_k}(s)\, ds \right) - g T_{N_t+1} + h(X_{N_t+1}) \right]} \right]. \qquad (4.6)$$

(ii) Para cada $x \in S$, el término entre corchetes en (4.4) tiene un minimizador $f^*(x) \in A(x)$, y la política $f^* \in \mathbb{F}$ satisface que, para cada $x \in S$ y $t > 0$,

$$e^{\lambda h(x)} = E_x^{f^*} \left[e^{\lambda \left[\sum_{k=0}^{N_t} \left(C(X_k, A_k) + \int_0^{S_k} \rho_{X_k, A_k}(s)\, ds \right) - g T_{N_t+1} + h(X_{N_t+1}) \right]} \right]. \qquad (4.7)$$

Demostración. (i) Sean $x \in S$ y $\pi \in \mathcal{P}$ arbitrarios, pero fijos. Note que (4.4) lleva a que, para cada $a \in A(x)$,

$$\begin{aligned} e^{\lambda h(x)} &\leq e^{\lambda C(x,a)} \int_0^B e^{\lambda \left[\int_0^s \rho_{x,a}(t)\, dt - gs \right]} dF_{x,a}(s) \sum_{y \in S} p_{x,y}(a) e^{\lambda h(y)} \qquad (4.8) \\ &= E_x^\pi \left[e^{\lambda \left[C(X_0, A_0) + \int_0^{S_0} \rho_{X_0, A_0}(t)\, dt - g S_0 + h(X_1) \right]} \,\middle|\, X_0 = x, A_0 = a \right]. \end{aligned}$$

Por (3.6), se sigue que para cada $n \in \mathbb{N}$:

$$e^{\lambda h(X_n)} \leq E_x^\pi \left[e^{\lambda \left[C(X_n, A_n) + \int_0^{S_n} \rho_{X_n, A_n}(t)\, dt - g S_n + h(X_{n+1}) \right]} \,\middle|\, \mathcal{H}_n, A_n \right]. \qquad (4.9)$$

Se probará por inducción que para cada entero no negativo n, la siguiente relación es válida:

$$\begin{aligned} e^{\lambda h(x)} &\leq E_x^\pi \left[e^{\lambda \left[\sum_{k=0}^{N_t} \left(C(X_k, A_k) + \int_0^{S_k} \rho_{X_k, A_k}(t)\, dt \right) - g T_{N_t+1} + h(X_{N_t+1}) \right]} I[N_t \leq n] \right] \\ &+ E_x^\pi \left[e^{\lambda \left[\sum_{k=0}^{n} \left(C(X_k, A_k) + \int_0^{S_k} \rho_{X_k, A_k}(t)\, dt \right) - g T_{n+1} + h(X_{n+1}) \right]} I[N_t > n] \right]. \qquad (4.10) \end{aligned}$$

Para esto, note que integrando con respecto a $\pi_0(\cdot|x)$ en (4.8) se sigue que

$$\begin{aligned}
e^{\lambda h(x)} &\leq E_x^\pi\left[e^{\lambda\left[C(X_0,A_0)+\int_0^{S_0}\rho_{X_0,A_0}(t)\,dt-gS_0+h(X_1)\right]}\right]\\
&= E_x^\pi\left[e^{\lambda\left[C(X_0,A_0)+\int_0^{S_0}\rho_{X_0,A_0}(t)\,dt-gS_0+h(X_1)\right]}I[N_t=0]\right]\\
&+ E_x^\pi\left[e^{\lambda\left[C(X_0,A_0)+\int_0^{S_0}\rho_{X_0,A_0}(t)\,dt-gS_0+h(X_1)\right]}I[N_t>0]\right];
\end{aligned}$$

como $T_1 = S_0$, esta última expresión inmediatamente lleva a que (4.10) es válida para $n = 0$. Ahora, suponga que (4.10) es válida para cierto entero no negativo n, y observe que

$$\begin{aligned}
&e^{\lambda\left[\sum_{k=0}^{n}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,dt\right)-gT_{n+1}+h(X_{n+1})\right]}I[N_t>n]\\
&\quad= e^{\lambda\left[\sum_{k=0}^{n}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,dt\right)-gT_{n+1}\right]}I[N_t\geq n+1]e^{\lambda h(X_{n+1})}\\
&\quad\leq e^{\lambda\left[\sum_{k=0}^{n}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,dt\right)-gT_{n+1}\right]}I[N_t\geq n+1]\\
&\quad\times E_x^\pi\left[e^{\lambda\left[C(X_{n+1},A_{n+1})+\int_0^{S_{n+1}}\rho_{X_{n+1},A_{n+1}}(t)\,dt-gS_{n+1}+h(X_{n+2})\right]}\middle|\mathcal{H}_{n+1},A_{n+1}\right]\\
&\quad\leq E_x^\pi\left[e^{\lambda\left[\sum_{k=0}^{n+1}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,dt\right)-g[T_{n+1}+S_{n+1}]+h(X_{n+2})\right]}\right.\\
&\qquad\left. I[N_t\geq n+1]\middle|\mathcal{H}_{n+1},A_{n+1}\right],
\end{aligned}$$

donde (4.9) fue usado para establecer la primera desigualdad, mientras que el hecho de que las variables $I[N_t \geq n+1]$ y

$$\sum_{k=0}^{n}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,dt\right)-gT_{n+1}+h\left(X_{n+1}\right)$$

son $\sigma(\mathcal{H}_{n+1})$-medibles fue usado en el último paso. Como $T_{n+2} = T_{n+1} + S_{n+1}$, por (3.2), se sigue de la expresión anterior que

$$\begin{aligned}
&E_x^\pi\left[e^{\lambda\left[\sum_{k=0}^{n}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,dt\right)-gT_{n+1}+h(X_{n+1})\right]}I[N_t>n]\right]\\
&\leq E_x^\pi\left[e^{\lambda\left[\sum_{k=0}^{n+1}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,dt\right)-gT_{n+2}+h(X_{n+2})\right]}I\left[N_t\geq n+1\right]\right]\\
&= E_x^\pi\left[e^{\lambda\left[\sum_{k=0}^{n+1}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,dt\right)-gT_{n+2}+h(X_{n+2})\right]}I[N_t=n+1]\right]\\
&+E_x^\pi\left[e^{\lambda\left[\sum_{k=0}^{n+1}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,dt\right)-gT_{n+2}+h(X_{n+2})\right]}I[N_t>n+1]\right]\\
&= E_x^\pi\left[e^{\lambda\left[\sum_{k=0}^{N_t}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,dt\right)-gT_{N_t+1}+h(X_{N_t+1})\right]}I[N_t=n+1]\right]\\
&+E_x^\pi\left[e^{\lambda\left[\sum_{k=0}^{n+1}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,dt\right)-gT_{n+2}+h(X_{n+2})\right]}I[N_t>n+1]\right],
\end{aligned}$$

y combinando esta relación con la hipótesis de inducción se sigue que (4.10) es también válido con $n+1$ en lugar de n, esto es

$$\begin{aligned}
e^{\lambda h(x)} &\leq E_x^\pi \left[e^{\lambda\left[\sum_{k=0}^{N_t}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,dt\right)-gT_{N_t+1}+h(X_{N_t+1})\right]} I[N_t \leq n]\right] \\
&+E_x^\pi \left[e^{\lambda\left[\sum_{k=0}^{n}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,dt\right)-gT_{n+1}+h(X_{n+1})\right]} I[N_t > n]\right] \\
&\leq E_x^\pi \left[e^{\lambda\left[\sum_{k=0}^{N_t}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,dt\right)-gT_{N_t+1}+h(X_{N_t+1})\right]} I[N_t \leq n]\right] \\
&+E_x^\pi \left[e^{\lambda\left[\sum_{k=0}^{N_t}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,dt\right)-gT_{N_t+1}+h(X_{N_t+1})\right]} I[N_t = n+1]\right] \\
&+E_x^\pi \left[e^{\lambda\left[\sum_{k=0}^{n+1}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,dt\right)-gT_{n+2}+h(X_{n+2})\right]} I[N_t > n+1]\right] \\
&= E_x^\pi \left[e^{\lambda\left[\sum_{k=0}^{N_t}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,dt\right)-gT_{N_t+1}+h(X_{N_t+1})\right]} I[N_t \leq n+1]\right] \\
&+E_x^\pi \left[e^{\lambda\left[\sum_{k=0}^{n+1}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,dt\right)-gT_{n+2}+h(X_{n+2})\right]} I[N_t > n+1]\right],
\end{aligned}$$

completando el argumento de inducción. Ahora, se verá el comportamiento de los dos términos del lado derecho de (4.10) cuando n tiende a ∞. Primero, note que el Teorema de Convergencia Monótona y (4.5) implican lo siguiente:

$$\begin{aligned}
&\lim_{n\to\infty} E_x^\pi \left[e^{\lambda\left[\sum_{k=0}^{N_t}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,dt\right)-gT_{N_t+1}+h(X_{N_t+1})\right]} I[N_t \leq n]\right] \\
&\quad = E_x^\pi \left[e^{\lambda\left[\sum_{k=0}^{N_t}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,dt\right)-gT_{N_t+1}+h(X_{N_t+1})\right]}\right]. \qquad (4.11)
\end{aligned}$$

Con respecto al segundo término en el lado derecho de (4.10) recordando (3.2) y (4.1), y notando que las desigualdades $S_k \leq B$ y $T_{n+1} \leq (n+1)B$ son válidas P_x^π-casi seguramente, se obtiene que

$$\begin{aligned}
&e^{\lambda\left[\sum_{k=0}^{n}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,dt\right)-gT_{n+1}+h(X_{n+1})\right]} I[N_t > n] \\
&\qquad \leq e^{\lambda(n+1)[\|C\|+B(B_\rho+|g|)]+\lambda\|h\|} I[N_t \geq n+1] \qquad (4.12)
\end{aligned}$$

entonces

$$\begin{aligned}
&E_x^\pi \left[e^{\lambda\left[\sum_{k=0}^{n}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,dt\right)-gT_{n+1}+h(X_{n+1})\right]} I[N_t > n]\right] \\
&\qquad \leq e^{\lambda(n+1)[\|C\|+B(B_\rho+|g|)]+\lambda\|h\|} P_x^\pi[N_t \geq n+1].
\end{aligned}$$

Haciendo $\alpha = e^{-\lambda[\|C\|+B(B_\rho+|g|)]}/2$, combinando lo anterior con el Lema 4.2.1(ii) se sigue que

$$\begin{aligned}
&E_x^\pi \left[e^{\lambda\left[\sum_{k=0}^{n}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(t)\,dt\right)-gT_{n+1}+h(X_{n+1})\right]} I[N_t > n]\right] \\
&\qquad \leq e^{\lambda\|h\|}(1/2)^{n+1} e^{m_\alpha t} \to 0 \text{ cuando } n \to \infty.
\end{aligned}$$

Después de tomar el límite cuando n tiende a ∞ en ambos lados de (4.10), la última expresión y (4.11) llevan a (4.6), es decir

$$e^{\lambda h(x)} \leq E_x^\pi \left[e^{\lambda \left[\sum_{k=0}^{N_t} \left(C(X_k, A_k) + \int_0^{S_k} \rho_{X_k, A_k}(t)\, dt \right) - g T_{N_t+1} + h(X_{N_t+1}) \right]} I[N_t \leq n] \right],$$

obteniendo así lo deseado.

(ii) Sea $x \in S$ arbitrario, y note que la Suposición 4.1.1(iv) y el Teorema de Convergencia Dominada implican que

$$R(a,s) := e^{\lambda \left[\int_0^s \rho_{x,a}(t)\, dt - gs \right]}$$

es una función continua de $(a,s) \in A(x) \times [0,B]$, y entonces la función

$$a \mapsto \int_0^B e^{\lambda \left[\int_0^s \rho_{x,a}(t)\, dt - gs \right]} dF_{x,a}(s)$$

es continua en su dominio $A(x)$, por el Lema 4.2.1(i). Combinando esta propiedad con Suposición 4.1.1(i), se sigue que el término entre llaves en (4.4) es una función continua de $a \in A(x)$, y entonces la compacidad de $A(x)$ asegura la existencia de un minimizador $f^*(x) \in A(x)$. Se sigue que la política estacionaria f^* satisface

$$\begin{aligned} e^{\lambda h(x)} &= e^{\lambda C(x, f^*(x))} \int_0^B e^{\lambda \left[\int_0^s \rho_{x,f^*(x)}(t) dt - gs \right]} dF_{x,f^*(x)}(s) \sum_{y \in S} p_{x,y}(f^*(x)) e^{\lambda h(y)} \\ &= E_x^{f^*} \left[e^{\lambda \left[C(X_0, A_0) + \int_0^{S_0} \rho_{X_0, A_0}(t)\, dt - g S_0 + h(X_1) \right]} \right], \quad x \in S. \end{aligned}$$

A partir de esta igualdad, (4.7) puede ser obtenida usando un argumento de inducción similar al usado en la parte (i) para obtener (4.6) de (4.8). ■

4.3. Caracterización del Índice Promedio Óptimo

En esta sección se presentan al Teorema 4.3.1 y Teorema 4.3.4.

Para comenzar, note que (3.2) y (3.3) implican que

$$T_{N_t} \leq t < T_{N_t+1} = T_{N_t} + S_{N_t},$$

así que, para cada $t > 0$, las siguientes relaciones siempre son válidas P_x^π-casi seguramente:

$$0 \leq t - T_{N_t} \leq S_{N_t} \leq B \text{ y } T_{N_t+1} - t \leq S_{N_t} \leq B; \tag{4.13}$$

ver (4.1) para las desigualdades relacionadas a B.

Recordando la expresión (3.9) para el costo total $\mathcal{C}_t$ incurrido hasta un tiempo $t > 0$, note que

$$\sum_{k=0}^{N_t}\left(C(X_k, A_k) + \int_0^{S_k} \rho_{X_k,A_k}(s)\,ds\right) - gT_{N_t+1}$$
$$= (\mathcal{C}_t - tg) + \int_{t-T_{N_t}}^{S_{N_t}} \rho_{X_{N_t},A_{N_t}}(s)\,ds - (T_{N_t+1} - t)g.$$

Combinando esta igualdad con (4.2) y (4.13) se sigue que

$$\left|\sum_{k=0}^{N_t}\left(C(X_k, A_k) + \int_0^{S_k} \rho_{X_k,A_k}(s)\,ds\right) - gT_{N_t+1} - (\mathcal{C}_t - tg)\right| \leq B(B_\rho + |g|). \tag{4.14}$$

Teorema 4.3.1 *[Verificación.] Supóngase que la igualdad (4.4) es satisfecha por la pareja $(g, h(\cdot))$. Bajo la Suposición 4.1.1 las siguientes afirmaciones son válidas.*

(i) $J^*_\lambda(x) = g$ para cada $x \in S$.

(ii) Para cada $x \in S$, el término entre corchetes del lado derecho de (4.4) tiene un minimizador $f^*(x) \in A(x)$, y la política estacionaria f^* es óptima, es decir, $J_\lambda(\cdot, f^*) = g$. Más aún,

$$\lim_{t\to\infty} \frac{1}{t} J_{t,\lambda}(x, f^*) = g, \quad x \in S.$$

Demostración. Sean $x \in S$ y $\pi \in \mathcal{P}$ arbitrarios. Note que (4.6) implica que

$$e^{-2\lambda\|h\|} \leq E_x^\pi\left[e^{\lambda\left[\sum_{k=0}^{N_t}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(s)\,ds\right)-gT_{N_t+1}\right]}\right],$$

una relación que junto con (4.14) lleva a

$$e^{-2\lambda\|h\|} \leq E_x^\pi\left[e^{\lambda[\mathcal{C}_t - tg + B(B_\rho+|g|)]}\right],$$

de donde

$$e^{\lambda[-2\|h\| - B(B_\rho+|g|)+tg]} \leq E_x^\pi\left[e^{\lambda\mathcal{C}_t}\right],$$

en consecuencia

$$g \leq \liminf_{t\to\infty} \frac{1}{\lambda t}\log\left(E_x^\pi\left[e^{\lambda\mathcal{C}_t}\right]\right) \leq J_\lambda(x, \pi); \tag{4.15}$$

ya que la política π es arbitraria, se sigue que

$$g \leq J^*_\lambda(x). \tag{4.16}$$

Ahora, sea $f^* \in \mathbb{F}$ tal que, para cada $x \in S$, la acción $f^*(x)$ minimiza el lado derecho de (4.4). En este caso, la igualdad (4.7) dada en el Teorema 4.2.2(ii) lleva a

$$e^{2\lambda\|h\|} \geq E_x^{f^*}\left[e^{\lambda\left[\sum_{k=0}^{N_t}\left(C(X_k,A_k)+\int_0^{S_k}\rho_{X_k,A_k}(s)\,ds\right)-gT_{N_t+1}\right]}\right],$$

y usando (4.14) la ecuación anterior implica que

$$e^{2\lambda\|h\|} \geq E_x^{f^*}\left[e^{\lambda[\mathcal{C}_t - tg - B(B_\rho + |g|)]}\right],$$

de donde,

$$e^{2\lambda\|h\| + \lambda B(B_\rho + |g|) + \lambda tg} \geq E_x^{f^*}\left[e^{\lambda \mathcal{C}_t}\right],$$

entonces

$$g \geq \limsup_{t\to\infty} \frac{1}{\lambda t} \log\left(E_x^{f^*}\left[e^{\lambda\mathcal{C}_t}\right]\right) = J_\lambda(x, f^*).$$

Combinando esta relación con (4.15) y (4.16), se sigue que

$$J_\lambda^*(x) = g \text{ y } J_\lambda(x, f^*) = \lim_{t\to\infty} \frac{1}{\lambda t} E_x^{f^*}\left[e^{\lambda\mathcal{C}_t}\right];$$

esto completa la prueba ya que el estado x es arbitrario. ■

La prueba del resultado de existencia está basada en resultados ya establecidos del criterio promedio sensible al riesgo para el proceso a tiempo discreto $\{X_n\}$, los cuales son dados a continuación. Sea $\tilde{\mathcal{P}} \subset \mathcal{P}$ la clase de políticas satisfaciendo que para entero positivo n, el kérnel π_n depende únicamente de los estados

$$X_0, A_0, X_1, A_1, \ldots, X_{n-1}, A_{n-1}, X_n,$$

es decir, para cada $h_n \in \mathbb{H}_n$,

$$\pi(\cdot|h_n) = \pi(\cdot|x_0, a_0, \ldots, x_{n-1}, a_{n-1}, x_n).$$

Dada una función $D \in \mathbb{B}(\mathbb{K})$, se define al índice promedio a tiempo discreto en $x \in S$ cuando la política $\pi \in \tilde{\mathcal{P}}$ es usada, por

$$V_{\lambda,D}(x,\pi) := \limsup_{n\to\infty} \frac{1}{\lambda n} \log\left(E_x^\pi\left[e^{\lambda\sum_{k=0}^{n-1} D(X_k, A_k)}\right]\right),$$

y sea la función de valor λ-óptima promedio a tiempo discreto dada por

$$V_{\lambda,D}^*(x) := \inf_{\pi\in\tilde{\mathcal{P}}} V_{\lambda,D}(x,\pi), \quad x \in S.$$

Se tiene que $D \mapsto V_{\lambda,D}^*(x)$ es monótona y aditivamente homogénea, es decir, para $D, D_1 \in \mathbb{B}(\mathbb{K})$ y $c \in \mathbb{R}$,

$$V_{\lambda,D}^*(\cdot) \leq V_{\lambda,D_1}^*(\cdot), \text{ si } D \leq D_1, \text{ y } V_{\lambda,c+D}^*(\cdot) = c + V_{\lambda,D}^*(\cdot). \tag{4.17}$$

Como $D \leq D_1 + \|D - D_1\|$ se sigue que

$$V_{\lambda,D}^*(\cdot) \leq V_{\lambda,D_1}^*(\cdot) + \|D - D_1\|,$$

e intercambiando los papeles de D y D_1 se tiene

$$\|V_{\lambda,D}^* - V_{\lambda,D_1}^*\| \leq \|D - D_1\|.$$

Por otro lado, usando que $V^*_{\lambda,0} = 0$, la propiedad de monotonía en (4.17) implica que para $D, D_1 \in \mathbb{B}(S)$,

$$V^*_{\lambda,D} \leq 0 \leq V^*_{\lambda,D_1} \text{ cuando } D \leq 0 \leq D_1. \tag{4.18}$$

El siguiente resultado sobre $V_{\lambda,D}(\cdot,\cdot)$ sirve como herramienta clave en la prueba del resultado de existencia y fue tomado de [12].

Teorema 4.3.2 *Supóngase que para cualesquiera $x, y \in S$, la función $a \mapsto p_{x,y}(a)$ es continua en $a \in A(x)$. Bajo las Suposiciones 4.1.1(i) y 4.1.3, las siguientes afirmaciones son válidas:*

(i) Para cada $D \in \mathbb{B}(S)$ existen $\mu_D \in \mathbb{R}$ y $h_D : S \to \mathbb{R}$ tales que

$$e^{\lambda[\mu_D + h_D(x)]} = \inf_{a \in A(x)} \left[e^{\lambda D(x,a)} \sum_{y \in S} p_{x,y}(a) e^{\lambda h(y)} \right], \quad x \in S,$$

y

$$\mu_D = V_{\lambda,D}(\cdot). \tag{4.19}$$

(ii) Para cualesquiera $D, D_1 \in \mathbb{B}(S)$,

$$|\mu_D - \mu_{D_1}| \leq \|D - D_1\|. \tag{4.20}$$

Lema 4.3.3 *Suponga que la Suposición 4.1.1 es válida, y para cada $g \in \mathbb{R}$ se define la función $D_g : \mathbb{K} \to \mathbb{R}$ por*

$$D_g(x,a) = C(x,a) + \frac{1}{\lambda} \log \left(\int_0^B e^{\lambda\left[\int_0^s \rho_{x,a}(t)\,dt - gs\right]} \, dF_{x,a}(s) \right), \quad (x,a) \in \mathbb{K}. \tag{4.21}$$

Con esta notación, las siguientes afirmaciones son válidas:

(i) $D_g \in \mathbb{B}(\mathbb{K})$ para cada $g \in \mathbb{R}$.

(ii) $\|D_g - D_{g_1}\| \leq B|g - g_1|, \quad g, g_1 \in \mathbb{R}$.

(iii) Existe $g^- \geq 0$ tal que $D_{g^-} \leq 0$.

(iv) Para cada $\beta > 0$, existe $M_\beta > 0$ tal que

$$\int_0^B e^{\mu s} \, dF_{x,a}(s) > \beta, \quad \mu \geq M_\beta, \quad (x,a) \in \mathbb{K}.$$

(v) $D_{g^+} \geq 0$ para algún $g^+ \leq 0$.

Demostración. (i) Sean $x \in S$ y $g \in \mathbb{R}$ arbitrarios. Como en la prueba del Teorema 4.2.2(ii), la función

$$a \mapsto \int_0^B e^{\lambda\left(\int_0^s \rho_{x,a}(t)\,dt - gs\right)} \, dF_{x,a}(s)$$

es continua en su dominio $A(x)$, y entonces la continuidad de $C(x,\cdot)$ lleva a que $D_g(x,\cdot)$ también es continua. Dado que el espacio de estados está dotado con la topología discreta, se sigue que $D_g \in \mathbb{B}(\mathbb{K})$.

(ii) Observando que la desigualdad

$$e^{\lambda\left[\int_0^s \rho_{x,a}(t)\,dt - gs\right]} \le e^{\lambda\left[\int_0^s \rho_{x,a}(t)\,dt - g_1 s\right]} e^{\lambda B|g-g_1|}$$

es válida para cada $s \in [0, B]$, la conclusión deseada se sigue de (4.21).

(iii) Sea $\alpha = e^{-\lambda\|C\|}/2$ y elegir m_α como en el Lema 4.2.1(ii). Usando (4.2), note que para cada $(x, a) \in \mathbb{K}$,

$$\int_0^B e^{\lambda\left[\int_0^s \rho_{x,a}(t)\,dt - (B_\rho+\gamma)s\right]}\,dF_{x,a}(s) \le \int_0^B e^{-\lambda\gamma s}\,dF_{x,a}(s) \le \alpha, \quad \gamma \ge m_\alpha/\lambda.$$

Entonces, haciendo $g^- = B_\rho + m_\alpha/\lambda$, esta última desigualdad y (4.21) llevan a que

$$D_{g^-}(x, a) \le \|C\| + \log(\alpha)/\lambda \le \|C\| - \|C\| - \log(2)/\lambda < 0$$

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

(iv) Sea $x \in S$ arbitrario. Por la Suposición 4.1.1, la función

$$r_n(a) \mapsto \frac{1}{\int_0^B e^{ns}\,dF_{x,a}(s)}$$

es continua en $a \in A(x)$ para cada $n \in \mathbb{N}$. Más aún, (3.1) y el Teorema de Convergencia Monótona implican que $r_n \searrow 0$, en consecuencia, recordando que $A(x)$ es compacto, el Teorema de Dini (véase, por ejemplo Teorema 8.2.6 en [3]) implica que para cada $\beta > 0$ existe $M_{x,\beta} \in \mathbb{N}$ tal que

$$\frac{1}{\int_0^B e^{ns}\,dF_{x,a}(s)} \le \frac{1}{\beta}, \quad a \in A(x), \quad n \ge M_{x,\beta},$$

y la conclusión se obtiene, haciendo $M_\beta = \max\{M_{x,\beta} \,|\, x \in S\}$.

(v) Sean $\beta = e^{\lambda\|C\|} > 0$ y M_β como en la parte (iv). Usando (4.2), note que

$$\begin{aligned} \int_0^B e^{\lambda\left[\int_0^s \rho_{x,a}(t)\,dt + (B_\rho + M_\beta/\lambda)s\right]}\,dF_{x,a}(s) &\ge \int_0^B e^{M_\beta s}\,dF_{x,a}(s) \\ &\ge \beta = e^{\lambda\|C\|}, \end{aligned}$$

$(x, a) \in \mathbb{K}$ una relación que vía (4.21), lleva a $D_{g^+} \ge 0$ para $g^+ := -(B_\rho + M_\beta/\lambda)$. ■

Usando los preliminares anteriores, la existencia de soluciones para la ecuación de optimalidad (4.4) puede ser establecida como sigue.

Teorema 4.3.4 *[Existencia de Solución.] Bajo las Suposiciones 4.1.1 y 4.1.3 existen $g \in \mathbb{R}$ y $h : S \to \mathbb{R}$ tales que la ecuación de optimalidad (4.4) se satisface.*

Demostración. Para cada $g \in \mathbb{R}$, considere la función $D_g \in \mathbb{B}(\mathbb{K})$ definida en (4.21). Por el Teorema 4.3.2, la función de costo promedio sensible al riesgo a tiempo discreto $V_{\lambda,D_g}(\cdot)$ es la constante μ_{D_g}, mientras que el Lema 4.3.3(ii) y (4.20) llevan a que la función $g \mapsto \mu_{D_g}$ es continua en $g \in \mathbb{R}$. Ahora, sean g^+ y g^- como en partes (iii) y (v) del Lema 4.3.3, así que $D_{g^+} \geq 0$ y $D_{g^-} \leq 0$. En este caso,

$$\mu_{D_{g^+}} \geq 0 \text{ y } \mu_{D_{g^-}} \leq 0,$$

por (4.18) y (4.19). De este punto, el Teorema del Valor Intermedio implica la existencia de un número real g^* entre g^+ y g^- tal que

$$\mu_{D_{g^*}} = 0,$$

y entonces el Teorema 4.3.2(i) asegura que para cierta función $h_{D_{g^*}} : S \to \mathbb{R}$ la igualdad siguiente es válida para cada $x \in S$:

$$e^{\lambda h_{D_{g^*}}(x)} = \inf_{a \in A(x)} \left[e^{\lambda D_{g^*}(x,a)} \sum_{y \in S} p_{x,y}(a) e^{\lambda h_{D_{g^*}}(y)} \right]$$

$$= \inf_{a \in A(x)} \left[e^{\lambda C(x,a)} \int_0^B e^{\lambda \left[\int_0^s \rho_{x,a}(t)\, dt - g^* s \right]} dF_{x,a}(s) \sum_{y \in S} p_{x,y}(a) e^{\lambda h_{D_{g^*}}(y)} \right],$$

donde la segunda igualdad es debida a (4.21). Entonces, la pareja $(g^*, h_{D_{g^*}}(\cdot))$ satisface la ecuación de optimalidad (4.4). ■

4.4. Ejemplo: Problema de Mantenimiento

En esta sección se presenta un ejemplo, el cual fue introducido en un contexto neutral al riesgo en [39] y a continuación es dada la versión considerando un controlador sensible al riesgo.

Considere que se observa el funcionamiento de una máquina de manera periódica para determinar su nivel de funcionamiento actual y decidir si se lleva a cabo reparación alguna. La máquina puede encontrarse en alguno de los siguientes niveles de funcionamiento: $\{1, 2, \ldots, d\}$. Suponga que encontrarse en un nivel de funcionamiento x es mejor que estar en el nivel $x+1$, y el nivel 1 es el de condiciones perfectas. Si al momento de la revisión, el equipo se encuentra en un nivel de funcionamiento x y no se realiza reparación alguna entonces al siguiente día la máquina se encontrará en un estado y con probabilidad p_{xy}, y se supondrá que $p_{xy} = 0$ para $y > x$, es decir, la máquina no puede auto repararse, y además que $\sum_{y \geq x} p_{xy} = 1$. El nivel de funcionamiento $x = d$ representa una falla del equipo que requiere una reparación mayor, la cual tomará un día de mantenimiento extra. Se tiene que para los estados $1 < x < d$ se puede elegir entre llevar a cabo una reparación preventiva o permitir que el equipo labore por un periodo más. Si el equipo es reparado entonces su nivel de funcionamiento en el siguiente periodo será $x = 1$. Una reparación preventiva en el estado x, tendrá un costo $C(x)$; en caso de una reparación mayor se pagará un costo fijo C. Así el objetivo es determinar una regla de mantenimiento la cual optimizará los costos de funcionamiento. Suponga que el cambio en el nivel de funcionamiento es mediante una distribución uniforme $(0,1)$.

El problema anterior puede ser interpretado como un PDSM, donde las componentes del modelo están dadas a continuación:

1. Los estados representan los niveles de funcionamiento, es decir,

$$S = \{1, 2, \ldots, d, d+1\}\,;$$

donde el estado $d+1$ representa que se realizó una reparación mayor al equipo y por lo tanto, se requiere de un día extra para terminarla.

2. Las acciones están dadas por

$$A = \{0, 1, 2\}\,,$$

donde 0 representa que no se realizó reparación, 1 que una reparación de prevención fue hecha y 2 si una reparación mayor es realizada.

3. Las acciones admisibles están dadas por

$$\begin{array}{rcl} A(1) &=& \{0\}, \\ A(x) &=& \{0,1\}, \ \ \text{para } 1 < x < d, \\ A(d) &=& A(d+1) = \{2\}. \end{array}$$

4. La ley de transición está dada por,

$$\begin{array}{rcl} p_{xy}(0) &=& p_{xy}, \ \text{para } 1 \leq x < n \\ p_{x1}(1) &=& 1, \ \text{para } 1 < x < n \\ p_{d,d+1}(2) &=& p_{d+1,1}(2) = 1, \end{array}$$

y para los restantes casos, $p_{xy}(a) = 0$.

5. Los costos de funcionamiento y la tasa de costo por espera son:

$$C(x,0) = 0, \ C(x,1) = C\,(x)\,, \ C(d,2) = C, \ C(d+1,2) = 0.$$
$$\rho(x,0) = a, \ \rho(x,1) = b, \ \rho(n,2) = c, \ \rho(n+1,2) = 0.$$

Por lo tanto, de acuerdo con (4.4), la ecuación de optimalidad puede escribirse como:

$$e^{\lambda h(x)} = \min_{a \in A(x)} \left[e^{\lambda C(x,a)} \int_0^1 e^{\lambda (\int_0^s \rho_{x,a}(t)dt - gs)} ds \sum_{y \in S} p_{xy}(a) e^{\lambda h(y)} \right].$$

Ahora, se presenta un caso numérico del problema planteado para el cual se determinará la regla de funcionamiento óptimo así como el costo promedio óptimo por periodo mediante el algoritmo de iteración de políticas, presentado a continuación:

- Paso 1: Elegimos una política estacionaria R.

- Paso 2: Para la política R, se determinan las soluciones de $\{g(R), v_x(R)\}$, resolviendo el siguiente sistema de ecuaciones lineales:

$$\begin{aligned} e^{\lambda V_x} &= e^{\lambda C(x,R)} \int_0^B e^{\lambda(\int_0^s \rho_{x,R}(t)dt - gs)} dF_{x,R}(s) \sum_{y \in S} p_{xy}(R) e^{\lambda V_y},\ x \in S, \\ e^{\lambda V_s} &= 1 \end{aligned} \tag{4.22}$$

donde $s \in S$ es un estado arbitrario.

- Paso 3: Para cada estado $x \in S$, determinamos una acción a_x alcanzando el mínimo en:

$$\min_{a \in A(x)} \left[e^{\lambda C(x,a)} \int_0^B e^{\lambda(\int_0^s \rho_{x,a}(t)dt - g(R)s)} dF_{x,a}(s) \sum_{y \in S} p_{xy}(a) e^{\lambda V_y(R)} \right] \tag{4.23}$$

donde $g(R)$ y $V_y(R)$ fueron determinados en el paso anterior; la nueva política $\hat{R}$ es obtenida eligiendo $\hat{R}_x = a_x$.

- Paso 4: Si la nueva política $\hat{R} = R$, entonces el algoritmo se detiene con la política R. De otra manera, regresamos al paso 2 con $\hat{R}$ en lugar de R.

Se supondrá que $d = 3$, entonces $S = \{1, 2, 3, 4\}$, $A(1) = \{0\}$, $A(2) = \{0, 1\}$ y $A(3) = A(4) = \{2\}$, y que el coeficiente de sensibilidad al riesgo es $\lambda = 1$. Por cuestiones de simplicidad se supondrá que la tasa del costo de permanencia, $\rho(x, a) = 0$ para cada $(x, a) \in \mathbb{K}$ y que los costos de transición están dados por $C(x, 0) = C(4, 2) = 0$, para $x \in S$ y $C(2, 1) = 2,\ C(3, 2) = 4$. La ley de transición está dada por

$$p_{xy}(0) = \begin{pmatrix} 0{,}6 & 0{,}4 & 0 \\ 0 & 0{,}5 & 0{,}5 \end{pmatrix}.$$

Usando lo anterior, se tiene que la ecuación (4.22) se puede escribir como

$$\begin{aligned} e^{V_x} &= e^{C(x,R)} \frac{1 - e^{-g}}{g} \sum_{y \in S} p_{xy}(R) e^{V_y},\ x \in S, \\ e^{V_s} &= 1. \end{aligned}$$

Entonces, aplicando el algoritmo de iteración de políticas con la política inicial dada por $R = (0, 0, 2, 2)$, al resolver el sistema (4.22) para cada $x \in S$ se tiene que

$$\begin{aligned} e^{V_1} &= \frac{1-e^{-g}}{g}\left[0{,}6e^{V_1}+0{,}4e^{V_2}\right], \\ e^{V_2} &= \frac{1-e^{-g}}{g}\left[0{,}5e^{V_2}+0{,}5e^{V_3}\right], \\ e^{V_3} &= \frac{1-e^{-g}}{g}e^{4+V_4}, \\ e^{V_4} &= \frac{1-e^{-g}}{g}e^{V_1} \end{aligned} \tag{4.24}$$

y además

$$e^{\lambda V_4} = 1. \tag{4.25}$$

De la última igualdad en (4.24) y (4.25), se sigue que

$$1 = \frac{1-e^{-g}}{g}e^{V_1},$$

entonces

$$e^{v_1} = \frac{g}{1-e^{-g}}. \tag{4.26}$$

Ahora, sustituyendo (4.25) en (4.24) se tiene que

$$e^{V_3} = \frac{1-e^{-g}}{g}e^{4},$$

$$\begin{aligned} e^{V_2} &= \frac{1-e^{-g}}{g}0{,}5\left[e^{V_2}+e^{V_3}\right] \\ &= \frac{1-e^{-g}}{g}0{,}5\left[e^{V_2}+\frac{1-e^{-g}}{g}e^{4}\right], \end{aligned}$$

de donde

$$e^{V_2} = \frac{\left(1-e^{-g}\right)^2 e^4}{g\left(2g-1+e^{g}\right)}.$$

Luego, usando la expresión anterior junto con (4.26), se sigue que

$$e^{V_1} = 0{,}6 + 0{,}4\frac{\left(1-e^{-g}\right)^2 e^4}{g\left(2g-1+e^{g}\right)},$$

es decir,

$$0{,}6 + 0{,}4\frac{\left(1-e^{-g}\right)^2 e^4}{g\left(2g-1+e^{g}\right)} - \frac{g}{1-e^{-g}} = 0,$$

resolviendo la ecuación anterior mediante el uso del programa Mathematica, se tiene que $g = g\left(R\right) = 2{,}35268$ y entonces

$$\begin{aligned} e^{V_1} &= 2{,}59998 \\ e^{V_2} &= 4{,}9995 \\ e^{V_3} &= 20{,}9994 \\ e^{V_4} &= 1. \end{aligned}$$

Ahora, se determinará $a \in A\left(x\right)$ para cada x, tal que minimice a la expresión (4.23). Observe que, debido a las especificaciones del modelo únicamente debe determinarse los minimizadores en el estado 2, sea

$$M_2\left(a, R\right) := e^{C(2,a)} \frac{1 - e^{-g(R)}}{g\left(R\right)} \sum_{y \in S} p_{2y}(a) e^{V_y(R)},\ a \in A\left(2\right).$$

Entonces

$$\begin{aligned} M_2\left(0, R\right) &= e^{\lambda C(2,0)} \frac{1 - e^{-2{,}35268}}{2{,}35268} \sum_{y \in S} p_{2y}(0) e^{V_y(R)} \\ &= 0{,}384619 \left[0{,}5 e^{V_2(R)} + 0{,}5 e^{V_3(R)}\right] \\ &= 4{,}99993, \end{aligned}$$

y

$$\begin{aligned} M_2\left(1, R\right) &= e^{\lambda C(2,1)} \frac{1 - e^{-2{,}35268}}{2{,}35268} \sum_{y \in S} p_{2y}(1) e^{V_y(R)} \\ &= e^2 (0{,}384619) \left[e^{V_1(R)}\right] \\ &= 7{,}38907. \end{aligned}$$

Por lo que, la acción óptima para el estado 2 es $a = 0$ y por lo tanto, $\hat{R} = (0, 0, 2, 2)$.

Finalmente como las políticas R y $\hat{R}$ coinciden, hemos terminado y entonces $R = (0, 0, 2, 2)$ es la política óptima y $g = 2{,}35268$ es el costo mínimo por periodo.

Conclusiones

En esta tesis se trabajó con la teoría referente a los procesos de decisión en los casos markoviano y semi markoviano. Además, en ambos casos se supuso que el controlador contaba con un coeficiente de sensibilidad al riesgo λ. Para medir la eficiencia de una política se utilizó el criterio de rendimiento promedio λ-sensible al riesgo.

Bajo este contexto, en el trabajo se resolvieron dos problemas:

• El primero de ellos está relacionado con la continuidad sobre el parámetro de sensibilidad al riesgo en el caso markoviano y,

• En el segundo se presentó una caracterización del índice promedio óptimo para el caso de procesos de decisión semi markovianos.

En el primer problema se consideraron los Procesos de Decisión de Markov con espacio de estados numerable y función de costo acotada. Bajo estas condiciones, la continuidad de la función de valor óptimo $J^*(\lambda, \cdot)$ sobre el parámetro de sensibilidad al riesgo fue analizada. En la Proposición 2.2.1 se probó que la continuidad de la función de valor óptimo es, bajo condiciones sobre la función de costo, válida para cada $\lambda \neq 0$, sin embargo, para el caso $\lambda = 0$ tal propiedad no siempre está garantizada, por lo que fue necesario dar un supuesto sobre la ley de transición (Suposición 2.1.2), bajo el cual es posible asegurar que para cada estado x, la función de costo promedio óptima λ-sensible $J^*(\lambda, x)$ es una función continua de λ en cero. También, se presentan ejemplos para demostrar que (i) el resultado de continuidad no es válido para una ley de transición general, y (ii) si la Suposición 2.1.2 se sustituye por la condición de Lyapunov, la cual es un requisito de comunicación de recurrencia menos fuerte, entonces la continuidad de la función $\lambda \mapsto J^*(\lambda, \cdot)$ no puede garantizarse.

En el segundo problema se consideraron los Procesos de Decisión Semi markovianos (PDSM) sensibles al riesgo, donde a diferencia de los Procesos de Decisión de Markov sensibles al riesgo, la teoría no ha sido ampliamente desarrollada. En esta parte se consideraron PDSM con espacio de estados finito. En este caso, se presentó la ecuación de optimalidad, y considerando algunas condiciones sobre las componentes del modelo, se probó que la solución de tal ecuación caracteriza a la función de costo promedio óptimo. Posteriormente, bajo un supuesto sobre la ley de transición fue posible asegurar la existencia de una solución a la ecuación de optimalidad. Los resultados dados en esta parte pueden considerarse como una extensión de los obtenidos en [10], donde se trabajó el caso semi markoviano no controlado. Además, se presenta un problema de mantenimiento óptimo el cual ejemplifica la teoría desarrollada en este capítulo, tal problema es resuelto mediante el algoritmo de iteración de políticas.

En resumen, en el presente trabajo de tesis se da solución a dos problemas: la continuidad del índice promedio óptimo para los PDMs y, la caracterización del índice promedio óptimo para los PDSM considerando en ambos casos que el controlador poseé un coeficiente de sensibilidad al riesgo λ. Algunas consecuencias de estos resultados son: aproximar un PDM sensible al riesgo con un PDM neutral al riesgo, esto considerando que el caso sensible al riesgo coincide con $\lambda \neq 0$, mientras que el caso neutral ocurre cuando $\lambda = 0$ y usando el resultado de continuidad sobre el parámetro de sensibilidad al riesgo obtenido en este trabajo junto con los métodos de aproximación ya establecidos en el caso neutral. En el caso semi markoviano, el aporte resulta una pieza clave para poder establecer más resultado sobre el índice promedio, como podrían ser establecer métodos de aproximación para su solución o incluso buscar un resultado análogo al de continuidad dado en el caso markoviano.

Un problema interesante que podría derivarse de los resultados aquí presentados sería el suponer que los tiempos de permanencia no son acotados casi seguramente ya que, como fue planteado en la Observación 4.1.4 la prueba del Teorema 4.3.1 está basada en la comparación de los costos $\mathcal{C}_t - tg$ y la cota B dada en la condición (4.1). De acuerdo con lo anterior, sería necesario considerar condiciones del tipo siguiente:

- Existen constantes $B > 0$ y $b > 0$ tales que

$$F_{x,a,t}(B) \geq b, \quad (x, a) \in \mathbb{K}, \tag{4.27}$$

  donde $F_{x,a,t}(\cdot)$ es la distribución condicional sobre los tiempos de permanencia $S_n - r$ dado que $S_n > r$ para $r > 0$.

La necesidad de garantizar una cota con respecto a los tiempos de permanencia tiene vital importancia para establecer la prueba del Teorema 4.2.2, el cual es fundamental para la prueba del Teorema 4.3.1, donde el hecho de que los tiempos S_k sean acotados casi seguramente por B jugó un papel fundamental, como puede verse en la ecuación (4.13). En la prueba del Teorema 4.2.2, la condición (4.1) permite establecer la cota dada en la ecuación (4.12), es por ello que ahora se requerirá algún resultado similar, pero considerando lo dado en la ecuación (4.27). Por lo que, sería necesario estudiar las distribuciones $F_{x,a,t}(\cdot)$ con la finalidad de determinar cotas para los tiempos de permanencia.

Bibliografía

[1] Alanís-Durán, A., & Cavazos-Cadena, R. (2012). An optimality system for finite average Markov decision chains under risk-aversion. Kybernetika, 48(1), 83-104.

[2] Arapostathis, A., Borkar, V. S., Fernández-Gaucherand, E., Ghosh, M. K., & Marcus, S. I. (1993). Discrete-time controlled Markov processes with average cost criterion: a survey. SIAM Journal on Control and Optimization, 31(2), 282-344.

[3] Bartle, R. G., & Sherbert, D. R. (1992). Introduction to real analysis. New York: Wiley.

[4] Bäuerle, N., & Rieder, U. (2011). Markov decision processes with applications to finance. Springer Science & Business Media.

[5] Bäuerle, N., & Rieder, U. (2013). More risk-sensitive Markov decision processes. Mathematics of Operations Research, 39(1), 105-120.

[6] Baykal-Gürsoy, M., & Gürsoy, K. (2007). Semi-Markov decision processes. Probability in the Engineering and Informational Sciences, 21(04), 635-657.

[7] Binder, P. M. (2008). Philosophy of science: Theories of almost everything. Nature, 455(7215), 884-885.

[8] Cavazos-Cadena, R. (2003). Solution to the risk-sensitive average cost optimality equation in a class of Markov decision processes with finite state space. Mathematical methods of operations research, 57(2), 263-285.

[9] Cavazos-Cadena, R. (2009). Solutions of the average cost optimality equation for finite Markov decision chains: risk-sensitive and risk-neutral criteria. Mathematical Methods of Operations Research, 70(3), 541-566.

[10] Cavazos-Cadena, R. (2016). A Poisson equation for the risk-sensitive average cost in semi-Markov chains. Discrete Event Dynamic Systems, 1-24.

[11] Cavazos-Cadena, R., & Fernández-Gaucherand, E. (1999). Controlled Markov chains with risk-sensitive criteria: Average cost, optimality equations, and optimal solutions. Mathematical Methods of Operations Research, 49(2), 299-324.

[12] Cavazos-Cadena R., & Fernández-Gaucherand E. (2002). Risk-sensitive control in communicating average Markov decision chains, In: M. Dror, P. LÉcuyer and F. Szidarovsky (Eds.): Modelling Uncertainty: An examination of Stochastic Theory, Methods and Applications, Kluwer, 525-544.

[13] Cavazos-Cadena, R., & Hernández-Lerma, O. (1992). Equivalence of Lyapunov stability criteria in a class of Markov decision processes. Applied Mathematics and Optimization, 26(2), 113-137.

[14] Cavazos-Cadena, R., & Montes-de-Oca, R. (2003). The value iteration algorithm in risk-sensitive average Markov decision chains with finite state space. Mathematics of Operations Research, 28(4), 752-776.

[15] Cavazos-Cadena, R., & Hernández-Hernández, D. (2015). A Characterization of the Optimal Certainty Equivalent of the Average Cost via the Arrow-Pratt Sensitivity Function. Mathematics of Operations Research, 41(1), 224-235.

[16] Chávez-Rodríguez, S., Cavazos-Cadena, R., & Cruz-Suárez, H. (2015). Continuity of the optimal average cost in Markov decision chains with small risk-sensitivity. Mathematical Methods of Operations Research, 81(3), 269-298.

[17] Chávez-Rodríguez, S., Cavazos-Cadena, R., & Cruz-Suárez, H. (2016). Controlled Semi-Markov Chains with Risk-Sensitive Average Cost Criterion. Journal of Optimization Theory and Applications, 170(2), 670-686.

[18] Di Masi, G. B., & Stettner, L. (1999). Risk-sensitive control of discrete-time Markov processes with infinite horizon. SIAM Journal on Control and Optimization, 38(1), 61-78.

[19] Di Masi, G. B., & Stettner, L. (2000). Infinite horizon risk sensitive control of discrete time Markov processes with small risk. Systems & control letters, 40(1), 15-20.

[20] Di Masi, G. B., & Stettner, L. (2007). Infinite horizon risk sensitive control of discrete time Markov processes under minorization property. SIAM Journal on Control and Optimization, 46(1), 231-252.

[21] Hernández-Hernández, D., & Marcus, S. I. (1996). Risk sensitive control of Markov processes in countable state space. Systems & control letters, 29(3), 147-155.

[22] Hernández-Hernández, D., & Marcus, S. I. (1999). Existence of risk-sensitive optimal stationary policies for controlled Markov processes. Applied Mathematics and Optimization, 40(3), 273-285.

[23] Hernández-Lerma, O. (1989). Adaptive markov control processes. Springer Science & Business Media.

[24] Hernández-Lerma, O., & Lasserre, J. B. (1991). Discrete-time Markov control processes: basic optimality criteria. Springer Science & Business Media.

[25] Hordijk, A. (1974). Dynamic programming and Markov potential theory. MC Tracts, 51, 1-134.

[26] Howard, R. A., & Matheson, J. E. (1972). Risk-sensitive Markov decision processes. Management science, 18(7), 356-369.

[27] Hu, Q., & Yue, W. (2003). Optimal replacement of a system according toa semi-Markov decision process in a semi-Markov environment. Optimization Methods and Software, 18(2), 181-196.

[28] Jaśkiewicz A. (2007). Average optimality for risk-sensitive control with general state space. The annals of applied probability, 654-675.

[29] Luque-Vásquez, F., & Hernández-Lerma, O. (1999). Semi-Markov control models with average costs. Applicationes mathematicae, 26(3), 315-331.

[30] Munkres, J. R. (1984). Elements of algebraic topology. Addison-Wesley.

[31] Pratt, J. W. (1992). Risk aversion in the small and in the large. In Foundations of Insurance Economics, 83-98. Springer Netherlands.

[32] Pinedo, M. (2008) Scheduling: theory, algorithms and systems. Springer, New York.

[33] Puterman, M. L. (2014). Markov decision processes: discrete stochastic dynamic programming. John Wiley & Sons.

[34] Schweitzer, P. J. (1971). Iterative solution of the functional equations of undiscounted Markov renewal programming. Journal of Mathematical Analysis and Applications, 34(3), 495-501.

[35] Sladký, K. (2008). Growth rates and average optimality in risk-sensitive Markov decision chains. Kybernetika, 44(2), 205-226.

[36] Sennott, L. I. (1999). Stochastic dynamic programming and the control of queueing systems. John Wiley & Sons.

[37] Stidham Jr, S.,& Weber, R. (1993). A survey of Markov decision models for control of networks of queues. Queueing systems, 13(1-3), 291-314.

[38] Stokey, N. L. & Lucas, R. E. (1989). Recursive methods in economic dynamics. Harvard University Press.

[39] Tijms, H. C. (2003). A first course in stochastic models. John Wiley & Sons.

[40] Von Neumann, J., & Morgenstern, O. (1944). Theory of games and economic behavior. Princeton: Princeton university press.

[41] Wei, Q., & Guo, X. (2012). New average optimality conditions for semi-Markov decision processes in Borel spaces. Journal of Optimization Theory and Applications, 153(3), 709-732.