Benemérita Universidad
Autónoma de Puebla

Facultad de Ciencias Físico Matemáticas
Posgrado en Ciencias Matemáticas

Aproximaciones Contractivas al Costo Promedio Óptimo Bajo Aversión al Riesgo

Tesis

presentada para obtener el título de
Doctorado en Ciencias Matemáticas

Presenta
Julio Saucedo Zul

Director de Tesis
Dr. Rolando Cavazos Cadena
Dr. Hugo Adán Cruz Suárez

Puebla, Puebla. Noviembre 2021

Índice general

Índice general 1

1. Panorama General 3
- 1.1. Introducción 3
- 1.2. La Organización 4

2. Modelo de Decisión y Sensibilidad al Riesgo 7
- 2.1. Cadenas de Markov Controladas 7
- 2.2. Premio al Riesgo 11
- 2.3. Sensibilidad al Riesgo 13
- 2.4. Utilidad Exponencial y Costo Promedio 15
- 2.5. El Problema 18

3. Ecuación de Optimalidad 21
- 3.1. Introducción 21
- 3.2. Teorema de Verificación 22
- 3.3. Existencia de Soluciones 27
- 3.4. Resultados Auxiliares 27
- 3.5. Soluciones de la Ecuación de Optimalidad 35

4. Enfoque Descontado Bajo Aversión al Riesgo 37
- 4.1. Introducción 37
- 4.2. Modelo de Decisión 39
- 4.3. Aproximaciones Descontadas 44
- 4.4. Resultados Auxiliares 47
- 4.5. Costos con Soporte Compacto 51
- 4.6. Demostración del Teorema de Aproximación 55

5. Resumen y Problemas Abiertos 61
- 5.1. Retrospectiva 61
- 5.2. Dos Problemas Abiertos 62

Bibliografía 63

Capítulo 1

Panorama General

1.1. Introducción

Este trabajo trata sobre cadenas de Markov controladas evolucionando en un espacio numerable. En este modelo un agente, llamado el controlador o tomador de decisiones, observa el estado del sistema en cada tiempo entero no negativo y selecciona una acción entre varias alternativas disponibles, intervención que tiene un efecto doble, a saber, afecta las probabilidades de transición del sistema hacia un nuevo estado, y ocasiona un costo que depende tanto del estado como de la acción aplicada.

En general, el rendimiento de una estrategia de selección de acciones (llamada *política*) se mide usando la sucesión aleatoria de costos incurridos, y en este trabajo se supondrá que el funcionamiento de una política se evalúa usando *la tasa de crecimiento exponencial de los costos acumulados*. Como se verá más adelante, esta selección del criterio de funcionamiento corresponde a un controlador *averso al riesgo* que mide el desempeño de una política usando el costo que, en promedio la acción tomada, está dispuesto a pagar para evitar la incertidumbre ocasionada por el carácter aleatorio de los costos que se observarán durante la evolución del sistema. En el desarrollo subsecuente se introducirá formalmente la idea de aversión al riesgo, pero por el momento es oportuno establecer la siguiente formulación: un agente es averso al riesgo cuando, para evitar incurrir en un costo aleatorio, está dispuesto a pagar una cantidad que excede al valor esperado, así que el criterio del costo promedio estudiado en este trabajo difiere del criterio (neutral al riesgo) que se estudia en la teoría, ahora clásica, de los procesos de decisión Markovianos.

Un tema básico en este contexto es determinar el costo promedio óptimo, esto es, el mínimo de todos los costos promedio que es posible lograr, y el principal objetivo de este trabajo puede formularse como sigue:

• Aproximar el costo promedio óptimo sensible al riesgo mediante los puntos fijos de operadores contractivos.

La solución a este problema es una extensión de un importante resultado clásico en la teoría del criterio del costo promedio bajo neutralidad al riesgo, el cual es conocido como 'el enfoque descontado' al criterio del costo promedio.

1.2. La Organización

La presentación del material subsecuente refleja el proceso de aprendizaje del autor en los últimos tiempos, y ha sido organizada en 4 capítulos adicionales al presente, los cuales son prácticamente autocontenidos de manera que pueden leerse de forma independiente, de tal suerte que la persona interesada en pasar directamente a la principal contribución de este trabajo, puede dirigirse inmediatemente al Capítulo 4, el cual está basado en el artículo '*A Discounted Approach in Communicating Average Markov Decision Chains Under Risk-Aversion*, que fue recientemente publicado en las páginas 585-606 del volúmen 187 de la revista Journal of Optimization Theory and Applications en 2020; vea [59].

La organización del resto del trabajo es la siguiente: En el Capítulo 2 se introducen los conceptos básicos alrededor de los cuales gira este trabajo, de tal manera que la exposición incluye la definición formal de cadena de Markov controlada sobre un espacio de estados numerable, y se formulan las ideas de política de decisión y de índice de funcionamiento de una política. Posteriormente, se discute la noción de sensibilidad al riesgo, se define el criterio del costo promedio para un controlador *sensible al riesgo*, conceptos que permiten formular el principal problema que se estudia en esta tesis, a saber, la aproximación del costo promedio óptimo sensible al riesgo por medio de operadores contractivos. A continuación, en el Capítulo 3 se estudia la caracterización del costo promedio óptimo bajo aversión al riesgo en términos de una ecuación de optimalidad. El propósito principal es analizar una versión ligeramente más general del teorema de existencia de soluciones que fue establecido originalmente por Howard y Matheson [39] y, como en ese artículo, la exposición cubre el caso de modelos con espacio de estados y acciones finitos y utiliza el enfoque algebraico basado en el teorema de Perron-Frobenious para matrices no negativas. Sin embargo, en contraste con la versión original de Howard y Matheson, el argumento presentado no impone condición alguna sobre la *aperiodicidad de las cadenas de Markov inducidas por las* *políticas estacionarias*. La conclusión principal del capítulo puede describirise como sigue: Bajo el supuesto de que el espacio de estados es finito, si cada política estacionaria induce una cadena de Markov *comunicante*, entonces el costo promedio óptimo está caracterizado por una ecuación de optimalidad. Por otro lado, en el Capítulo 4 se presenta la principal contribución de este trabajo, la cual se refiere a procesos de decisión Markovianos con espacio de estados numerable. La hipótesis básica es que el controlador tiene un coeficiente de sensibilidad al riesgo constante y positivo, por lo que el tomador de decisiones es averso al riesgo. En este contexto el funcionamiento de una política de toma de decisiones se mide mediante el correspondiente criterio del costo promedio asociado a una función de costo acotada. Bajo condiciones de comunicación que aseguran que el costo promedio es constante, aunque no necesariamente se caracteriza mediante una ecuación de optimalidad, la principal contribución consiste en formular un criterio descontado mediante el cual es posible aproximar el costo promedio óptimo sensible al riesgo. Finalmente, la exposición concluye con el Capítulo 5 con algunos comentarios breves y la presentación de dos posibles problemas para investigación futura.

Capítulo 2

Modelo de Decisión y Sensibilidad al Riesgo

En este capítulo se introducen los conceptos básicos alrededor de los cuales gira este trabajo. La exposición inicia con la definición de cadena de Markov controlada sobre un espacio de estados numerable, formulando las ideas de política de decisión y de índice de funcionamiento de una política. Posteriormente, se discute la noción de sensibilidad al riesgo, se define el criterio del costo promedio para un controlador *sensible al riesgo*, y se presenta el principal problema que se estudia en esta tesis, a saber, la aproximación del costo promedio óptimo sensible al riesgo por medio de operadores contractivos. Finalmente, el capítulo concluye con un panorama general de la organización del material subsecuente.

2.1. Cadenas de Markov Controladas

Una cadena de Markov controlada $\mathcal{M} = (S, A, \{A(x)\}, C, P)$ es un modelo matemático para un sistema dinámico cuya evolución es influenciada por un *controlador* o *tomador de decisiones*. Las componentes del modelo son como sigue: El conjunto S es el *espacio de estados*, el cual en esta tesis se asume que es un conjunto numerable dotado con la métrica discreta. El espacio métrico A es el *conjunto de acciones*, mientras que, para cada $x \in S$, el subconjunto no vacío $A(x) \subset A$ es el conjunto de *acciones admisibles en el estado* x. Por otro lado $C : \mathbb{K} \to \mathbb{R}$ es la *función de costo por etapa*, donde $\mathbb{K} := \{(x, a) \,|\, a \in A(x), x \in S\}$ es la *clase de parejas admisibles*, y $P = [p_{x,y}(a)]_{(x,a)\in\mathbb{K}, y\in S}$ es la *ley de transición*. La dinámica del sistema, la cual involucra al controlador de manera preponderante, puede describirse como sigue: En cada tiempo $t \in \mathbb{N} := \{0, 1, 2, 3, \ldots\}$ el controlador detecta el estado del sistema $X_t = x \in S$ y, tomado en cuenta el registro de estados y acciones previos a t, selecciona y aplica una acción (control) $A_t = a \in A(X_t)$. Esta intervención tiene dos consecuencias: (i) Se incurre en un costo $C(X_t, A_t) = C(x, a)$, y (ii) La siguiente *propiedad de Markov* es válida: sin importar cual haya sido el registro previo de acciones aplicadas y estados observados, en el tiempo $t+1$ el sistema se moverá al estado $y \in S$ con probabilidad $p_{x,y}(a) \geq 0$, donde las relaciones $\sum_{y \in S} p_{x\,y}(a) = 1$ y $p_{x,y}(a) \geq 0$ son válidas para todo $(x, a) \in \mathbb{K}$ y $y \in S$.

Políticas. La descripción anterior pone de manifiesto que el controlador desempeña un papel esencial en la evolución del sistema a través de la selección y aplicación de acciones (controles). Una *política* es una regla que el tomador de decisiones utiliza para seleccionar acciones. Para describir esta idea formalmente, para cada $n \in \mathbb{N}$ sea $\mathbb{H}_n$ el conjunto de historias posibles hasta el tiempo n, de manera que $\mathbb{H}_0 := S$ y $\mathbb{H}_n = \mathbb{K}^n \times S$ cuando $n \geq 1$, mientras que $\mathbf{h} = (x_0, a_0, \ldots, x_{n-1}, a_{n-1}, x_n)$ denota un elemento arbitrario de $\mathbb{H}_n$, donde $x_n \in S$ y $(x_i, a_i) \in \mathbb{K}$ para $i < n$. Con esta notación, una política se define como una sucesión (especial) $\pi = \{\pi_n\}$ de núcleos estocásticos sobre S dado $\mathbb{H}_n$, i. e.,

(i) Para cada $\mathbf{h}_n \in \mathbb{H}_n$, $\pi_n(\cdot|\mathbf{h}_n)$ es una medida de probabilidad en la familia de Borel del espacio de acciones A, *la cual satisface* $\pi_n(A(x_n)|\mathbf{h}_n) = 1$, y
(ii) Para cada conjunto de Borel $B \subset A$, la función $\mathbf{h}_n \mapsto \pi_n(B|\mathbf{h}_n)$, es medible.

La clase de todas la políticas se denota por $\mathcal{P}$. Dado el estado inicial $x \in S$ y la política π usada para seleccionar acciones, la distribución del proceso $\{(X_t, A_t)\}$ se determina de manera única por medio del teorema de Ionescu-Tulcea [3, 36, 54] y se denota mediante P_x^π. Representando a la historia aleatoria hasta el tiempo n mediante H_n, esto es,

$$H_n := (X_0, A_0, X_1, A_1, \ldots, X_{n-1}, A_{n-1}, X_n), \tag{2.1}$$

la distribución P_x^π se caracteriza por las siguientes relaciones:

$$\begin{aligned} P_x^\pi[X_0 = x] &= 1 \\ P_x^\pi[A_n \in B|H_n = \mathbf{h}_n] &= \pi_n(B|\mathbf{h}_n) \\ P_x^\pi[X_{n+1} = y|H_n, A_n = a_n] &= p_{x_n\,y}(a_n) \end{aligned} \tag{2.2}$$

donde $n \in \mathbb{N}, x, y \in S, \mathbf{h}_n \in \mathbb{H}_n$ y $a_n \in A(x_n)$ son arbitrarios. Por otro lado, sea

$$\mathbb{F} = \prod_{x \in S} A(x) \tag{2.3}$$

la familia de todas las funciones $f : S \to A$ tal que $f(x) \in A(x)$ para toda $x \in S$. Cada $f \in \mathbb{F}$ se puede identificar con una política muy simple para seleccionar acciones: *cada vez que se observe que el sistema ocupa el estado x, aplique la acción $f(x)$*. Esta regla corresponde a la política $\pi^f = \{\pi_n^f\}$ que satisface

$$\pi_n^f(\{f(x_n)\}|\,\mathbf{h}_n) = 1 \tag{2.4}$$

para toda $n \in \mathbb{N}$ y $\mathbf{h}_n \in \mathbb{H}_n$, y en este caso π^f y f se identifican naturalmente, de manera que es legítimo escribir

$$\mathbb{F} \subset \mathcal{P};$$

una política $f \in \mathbb{F}$ se denomina *estacionaria*. Note que (2.2) y (2.4) implican que, bajo una política estacionaria f, la igualdad $A_n = f(X_n)$ ocurre siempre con probabilidad 1, y entonces

$$\begin{aligned} P_x^f[X_{n+1} = y|H_n, A_n] &= P_x^f[X_{n+1} = y|H_n, f(X_n)] \\ &= p_{X_n,y}(f(X_n)) \end{aligned} \tag{2.5}$$

de tal manera que bajo la acción de la política estacionaria f el proceso de estados $\{X_n\}$ es una cadena de Markov con matriz de transición $[p_{x,y}(f(x))]_{x,y\in S}$.

Índice de funcionamiento. Para decidir como seleccionar las acciones que se aplicarán, el controlador necesita una forma para medir el resultado de utilizar una política, y el instrumento que se utiliza con este fin se denomina *índice de desempeño*, o *criterio de funcionamiento*, el cual es una función $V : S \times \mathcal{P} \to \mathbb{R}$, donde para cada $(x, \pi) \in S \times \mathcal{P}$ el valor $V(x, \pi)$ se utiliza para medir el rendimiento de la política π cuando el estado inicial es $X_0 = x$. Dado que los controles se seleccionan por medio de $\pi \in \mathcal{P}$ y que el sistema inicia su evolución en el estado $x \in S$, el criterio $V(x, \pi)$ puede pensarse como una manera de resumir, por medio de un sólo número, el comportamiento probabilístico del proceso de costos $\{C(X_t, A_t)\}$, el cual está determinado por la distribución P_x^π. Comúnmente, el índice $V(x, \pi)$ se forma tomando el valor esperado de alguna función de la sucesión de costos $\{C(X_t, A_t)\}$, como se ilustra en los tres casos discutidos a continuación:

(i) Suponga que el controlador está interesado en el costo incurrido hasta antes del tiempo T, el cual es aleatorio y está dado por $\sum_{t=0}^{T-1} C(X_t, A_t)$; el criterio del costo total (esperado hasta antes del tiempo T) está dado por

$$V_T(x, \pi) = E_x^\pi\left[\sum_{t=0}^{T-1} C(X_t, A_t)\right]. \tag{2.6}$$

(ii) Para $\alpha \in (0,1)$, el criterio del costo (α-)descontado es

$$V_\alpha(x,\pi) = E_x^\pi \left[\sum_{t=0}^{\infty} \alpha^t C(X_t, A_t) \right]. \tag{2.7}$$

Este criterio es ampliamente usado en economía puesto que, definiendo $\alpha = 1/(1+\rho)$ donde ρ es la tasa de interés pagada por un activo sin riesgo, el índice descontado representa el valor en el tiempo $t = 0$ de todos los costos acumulados que se incurrirán en el futuro [55, 66, 67].

(iii) El criterio del costo promedio se define mediante

$$V(x,\pi) = \limsup_{m\to\infty} \frac{1}{m} E_x^\pi \left[\sum_{t=0}^{m-1} C(X_t, A_t) \right]. \tag{2.8}$$

Este criterio se utiliza cuando se tiene la certeza de que el sistema operará un número grande de etapas, pero no se tiene una idea precisa de cuantas.

Después de haber especificado de manera precisa el índice del desempeño $V(x,\pi)$, *el objetivo del controlador* es usar una *política óptima* π^*, esto es, una política π^* que satisfaga

$$V(x,\pi^*) = V^*(x), \quad \forall x \in S,$$

donde $V^*(\cdot)$ es la función de valor óptimo determinada por

$$V^*(x) = \inf_{\pi\in\mathcal{P}} V(x,\pi), \quad \forall x \in S.$$

Aplicaciones de cadenas de Markov controladas con estos y otros criterios han sido presentadas en diversas áreas, como reemplazo de maquinaria, manejo de inventarios, crecimiento económico, control de colas, pesquerías y presas, problemas de ventas y redes de transportes; vea, por ejemplo, [9, 10, 54, 56, 60, 67]. Por otro lado, de los tres criterios anteriores, el índice cuyo estudio requiere más estructura matemática para su estudio es el criterio del costo promedio ya que involucra el comportamiento ergódico del proceso $\{(X_t, A_t)\}$, y su análisis está basado en las propiedades de recurrencia de cadenas de Markov; vea, por ejemplo, [13, 46] para el caso de espacio de estados numerable, o [49, 51] para cadenas de Markov evolucionando sobre un espacio general de Borel. Un estudio detallado de cadenas de decisión de Markov dotadas con varios criterios, incluyendo los tres mencionados antes, puede encontrarse [36–38], así como en [11]. Por otro lado, el criterio fundamental en este trabajo es una variante del criterio del costo promedio definido en (2.8), el cual se formulará considerando la *sensibilidad al riesgo del controlador*, una idea que se discutirá más adelante.

2.2. Premio al Riesgo

La idea que subyace en los criterios definidos en (2.6)–(2.8), es que el controlador valora un costo aleatorio W por medio del valor esperado $E[W]$, de manera que será indiferente entre incurrir el costo aleatorio W o pagar el valor esperado $E[W]$. En cierto sentido, esta visión está justificada por la ley de los grandes números. En efecto, suponga que se realizarán un gran número N de repeticiones (independientes) del experimento que genera el costo aleatorio W, de manera que se observarán costos aleatorios $W_1, W_2, \ldots, W_N$ con la misma distribución ρ_W. Por la ley de los grandes números,

$$\frac{W_1 + W_2 + \cdots, W_N}{N} \approx E[W] = \int_{\mathbb{R}} w\rho_W(dw)$$

y entonces es natural pensar que, cada vez que se vaya a incurrir en el costo aleatorio con distribución ρ_W, el controlador esté dispuesto a pagar la cantidad fija $E[W]$. Sin embargo, es posible imaginar situaciones donde, naturalmente, un costo aleatorio puede valorarse de otra manera. Por ejemplo, considere al dueño de un auto valuado en \$250,000 quien paga un seguro de \$5000 contra accidentes. Suponga que (i) el costo Y asociado con un accidente, el cual es aleatorio, tiene un valor esperado igual a la mitad del precio del auto de manera que $E[Y] = 125,000$ y (ii) según las estadísticas, la probabilidad de que un conductor sufra un accidente el próximo año es de $1/1000$. Así, en lo que a su auto se refiere, el costo aleatorio W que el dueño prevé es $W = Y$ con probablidad $1/1000$ y $W = 0$ con probabilidad $1 - 1/1000$ de manera que $E[W] = E[Y]/1000 = 125$. Sin embargo, el dueño del auto ha pagado \$5000 a cambio de evitar enfrentarse al costo aleatorio W, indicando que W es valorado en una cantidad (mucho) mayor que el valor esperado. Bajo un conjunto de postulados (de racionalidad) introducidos por Von Neumann y Morgetern, es posible demostrar que para un observador particular existe una función U tal que un costo aleatorio W se valorará por medio de

$$E[U(W)] \tag{2.9}$$

(Berger [8]), donde U es (estrictamente) creciente. Luego, al decidir cual de los costos aleatorios W_0 y W_1 incurrir, el observador preferirá W_0 cuando $E[U(W_0)] < E[U(W_1)]$, y será indiferente entre los dos costos aleatorios si $E[U(W_0)] = E[U(W_1)]$. El *equivalente de certeza* (o *equivalente cierto*) del costo aleatorio W es el número real

$$\mathcal{E}_U(W) := U^{-1}\left(E[U(W)]\right). \tag{2.10}$$

En este caso $U(\mathcal{E}_U(W)) = E[U(W)]$, de manera que el controlador será indiferente entre pagar la cantidad fija $\mathcal{E}_U(W)$ o incurrir el costo aleatorio W. Más aún, ante la oferta de pagar una cantidad c en lugar de enfrentar el costo aleatorio W, el controlador aceptará la oferta si $c \leq \mathcal{E}_U(W)$, mientras que la rechazará cuando $c > \mathcal{E}_U(W)$. Así, en el caso del seguro del auto comentado anteriormente, se tiene que $\mathcal{E}_U(W) \geq \$5000$, ya que el dueño aceptó pagar el seguro de \$5000 en lugar de enfrentar el costo aleatorio W. En adelante se supondrá que la función de utilidad U es *continua y estrictamente creciente*, y que cada costo aleatorio W bajo consideración es tal que $U(W)$ tiene esperanza finita, condiciones bajo las cuales el equivalente cierto $\mathcal{E}_U(W)$ está determinado de manera única. Observe ahora que la comparación entre dos utilidades esperadas $E[U(W)]$ y $E[U(W_1)]$ no se altera si U se reemplaza por $\tilde{U} = aU + b$ donde $a > 0$, y por lo tanto *la función de utilidad* de un observador particular esta determinada sólo hasta una transformación lineal (afin) con pendiente positiva.

Definición 2.2.1 *Sea U una función de utilidad. El premio al riesgo de una variable aleatoria W con respecto a la función de utilidad U se define como*

$$\Delta_U(W) := \mathcal{E}_U(W) - E[W].$$

Así, dada una función de utilidad U, el premio al riesgo de un costo aleatorio W es la diferencia entre el equivalente de certeza $\mathcal{E}_U(W)$ y el valor esperado $E[W]$.

Un tomador de decisiones con función de utilidad U se llama
(i) *averso al riesgo* si

$$\mathcal{E}_U(W) \geq E[W]$$

para toda variable aleatoria W, mientras que
(ii) el tomador de decisiones es *proclive al riesgo* si

$$\mathcal{E}_U(W) \leq E[W].$$

Por la desigualdad de Jensen, la aversión (proclividad) al riesgo del controlador es equivalente a la convexidad (concavidad) de la correspondiente función de utilidad [28, 57]. Si el controlador es tanto averso como proclive al riesgo se llama *neutral al riesgo*, condición que equivale a que $\mathcal{E}_U(W) = E[W]$ para toda variable aleatoria W, de tal suerte que la función de utilidad es lineal, esto es, $U(x) = ax + b$ donde $a > 0$. En este caso, ya que la función de utilidad está determinada hasta una transformación afín, puede tomarse $U(\cdot)$ como la función identidad en $\mathbb{R}$.

2.3. Sensibilidad al Riesgo

El coeficiente de sensibilidad al riesgo, introducido en Pratt [53], es una medida (local) del grado de aversión o proclividad al riesgo de un controlador que enfrenta un costo aleatorio cercano a un valor específico $w \in \mathbb{R}$. Suponga que Y es una variable aleatoria acotada que satisface

$$E[Y] = 0, \quad \mathrm{Var}\,[Y] = 1, \tag{2.11}$$

y para cada $\sigma \geq 0$ defina

$$W(\sigma) = w + \sigma Y. \tag{2.12}$$

En este caso, $W(0)$ es la constante w y σ^2 es la varianza de $W(\sigma)$, esto es,

$$\Delta_U(W(0)) = 0, \quad \text{y} \quad \mathrm{Var}\,[W(\sigma)] = \sigma^2, \tag{2.13}$$

de manera que $W(\sigma)$ converge en probabilidad a w conforme σ tiende a cero. Intuitivamente, la incertidumbre detrás de $W(\sigma)$ proviene de la positividad de σ, y por lo tanto es interesante comparar el premio al riesgo $\Delta_U(W(\sigma))$ con la varianza σ^2. La siguiente proposición muestra que, para σ 'pequeño', $\Delta_U(W(\sigma))$ es 'prácticamente' proporcional a σ^2.

Proposición 2.3.1 *Sea $w \in \mathbb{R}$ arbitrario pero fijo y, para cada $\sigma > 0$, defina $W(\sigma)$ mediante* (2.12)*, donde Y satisface (2.11). Suponga que la función de utilidad U tiene derivada continua de orden 2 en una vecindad de w y que $U'(w) > 0$. En este caso,*

$$\lim_{\sigma \to 0} \frac{\Delta_U(W(\sigma))}{\sigma^2} = \frac{1}{2}\frac{U''(w)}{U'(w)}.$$

Demostración. Primero se verificará que

$$\frac{d}{d\sigma}\Delta_U(W(\sigma))|_{\sigma=0} = 0.$$

Con este fin, observe que

$$\begin{aligned}
\frac{U(\mathcal{E}_U(W(\sigma))) - U(\mathcal{E}_U(W(0)))}{\sigma} &= \frac{E[U(W(\sigma)) - U(W(0))]}{\sigma} \\
&= \frac{E[U(w + \sigma Y) - U(w)]}{\sigma};
\end{aligned}$$

recordando que U' es continua en una vecindad de w y que la variable aleatoria acotada Y satisface (2.11), el teorema de convergencia acotada implica que

$$\begin{aligned}\lim_{\sigma\to 0}\frac{U(\mathcal{E}_U(W(\sigma)))-U(\mathcal{E}_U(W(0)))}{\sigma}&=E\left[\lim_{\sigma\to 0}\frac{U(w+\sigma Y)-U(w)}{\sigma}\right]\\&=E\left[U'(w)Y\right]\\&=U'(w)E\left[Y\right]=0.\end{aligned}$$

Por lo tanto la función $\sigma\mapsto U(\mathcal{E}_U(W(\sigma))$ tiene derivada nula en $\sigma=0$, y combinando este hecho con la condición $U'>0$, se desprende que $\sigma\mapsto\mathcal{E}_U(W(\sigma))$ es derivable en $\sigma=0$ y que su derivada se anula en ese punto, esto es,

$$\left.\frac{d\mathcal{E}_U(W(\sigma))}{d\sigma}\right|_{\sigma=0}=0.$$

De (2.13) se tiene que

$$\begin{aligned}\lim_{\sigma\to 0}\frac{\Delta_U(W(\sigma))}{\sigma}&=\lim_{\sigma\to 0}\frac{\Delta_U(W(\sigma))-\Delta_U(W(0))}{\sigma}\\&=\left.\frac{d}{d\sigma}\Delta_U(W(\sigma))\right|_{\sigma=0}\\&=\left.\frac{d}{d\sigma}[\mathcal{E}_U(W(\sigma))-E[W(\sigma)]]\right|_{\sigma=0}\\&=\left.\frac{d}{d\sigma}[\mathcal{E}_U(W(\sigma))-w]\right|_{\sigma=0}\\&=0,\end{aligned}$$

donde la cuarta igualdad se debe a que $E[W(\sigma)]=w$ para todo valor de σ. Por lo tanto

$$\Delta_U(W(\sigma))=o(\sigma). \tag{2.14}$$

Combinando esta relación con la igualdad $\mathcal{E}_U(W(\sigma))=w+\Delta_U(W(\sigma)$, un desarrollo de Taylor de $U(\cdot)$ alrededor de w conduce a

$$\begin{aligned}U(\mathcal{E}_U(W(\sigma))&=U(w+\Delta_U(W(\sigma)))\\&=U(w)+U'(w)\Delta_U(W(\sigma))+O([\Delta_U(W(\sigma))]^2)\\&=U(w)+U'(w)\Delta_U(W(\sigma))+o(\sigma^2).\end{aligned}\tag{2.15}$$

Además, teniendo en mente que Y es acotada, una expansión adicional de Taylor alrededor de w implica que

$$U(w+\sigma Y)=U(w)+U'(w)\sigma Y+\frac{1}{2}U''(w)\sigma^2Y^2+R(w,\sigma Y),$$

donde el residuo satisface la desigualdad $|R(w, \sigma Y)| \leq k(w, \sigma)\sigma^2 Y^2$, donde $k(w, \sigma) \to 0$ conforme $\sigma \to 0$: luego,

$$E[|R(w, \sigma Y)|] \leq k(w, \sigma)\sigma^2 E[Y^2] = k(w, \sigma)\sigma^2 = o(\sigma^2).$$

Usando estas dos últimas relaciones desplegadas y recordando las igualdades en (2.11), se sigue que

$$\begin{aligned}
E[U(W(\sigma)] &= E[U(w + \sigma Y)] \\
&= E\left[U(w) + U'(w)\sigma Y + \frac{1}{2}U''(w)\sigma^2 Y^2 + R(w, \sigma Y)\right] \\
&= U(w) + \sigma U'(w)E[Y] + \frac{1}{2}U''(w)\sigma^2 E[Y^2] + E[R(w, \sigma)] \\
&= U(w) + \frac{1}{2}U''(w)\sigma^2 + o(\sigma^2).
\end{aligned}$$

Combinando esta relación con (2.15), se desprende que

$$U'(w)\Delta(W(\sigma)) = \frac{1}{2}U''(w)\sigma^2 + o(\sigma^2),$$

igualdad que conduce de inmediato a la conclusión deseada. ■

Definición 2.3.2 *Dada una función de utilidad U, el coeficiente local de sensibilidad al riesgo en el punto $w \in \mathbb{R}$ se define como*

$$\lambda_U(w) := \frac{U''(w)}{U'(w)}.$$

Con esta notación, la proposición anterior muestra que, para una variable aleatoria W que toma valores en una vecindad 'pequeña' de w, el doble del premio al riesgo $\Delta_U(W)$ es 'practicamente' proporcional a $\mathrm{Var}\,[W]$, y que la constante de proporcionalidad está dada por $\lambda_U(w)$.

2.4. Utilidad Exponencial y Costo Promedio

En el resto del trabajo supondremos que la función de utilidad U tiene primera derivada positiva en $\mathbb{R}$, segunda derivada continua en $\mathbb{R}$ y que el coeficiente de sensibilidad al riesgo $\lambda_U(\cdot)$ es constante y no nulo, digamos $\lambda \in \mathbb{R} \setminus \{0\}$:

$$\lambda_U(w) = \frac{U''(w)}{U'(w)} \equiv \lambda \neq 0, \quad w \in \mathbb{R}.$$

Esta relación equivale a $\log(U'(w)) = \lambda w + \tilde{a}$ donde $\tilde{a}$ es una constante real, de manera que

$$U'(w) = ae^{\lambda w}, \quad w \in \mathbb{R}, \quad \text{donde } a = e^{\tilde{a}} > 0;$$

por lo tanto,

$$U(w) = \frac{a}{\lambda} e^{\lambda w} + b, \quad w \in \mathbb{R}, \quad \text{donde } a > 0 \text{ y } b \in \mathbb{R}.$$

Como una función de utilidad está determinada hasta una transformación lineal con pendiente positiva, se tiene que si $\lambda_U(\cdot) = \lambda$ es constante, entonces la función U puede tomarse como

$$U_\lambda(w) = \operatorname{sign}(\lambda) e^{\lambda w}, \quad w \in \mathbb{R}, \tag{2.16}$$

la cual es la *función de utilidad exponencial* con parámtero $\lambda \neq 0$. El equivalente cierto de un costo aleatorio W con respecto a la función de utilidad U_λ se denota mediante $\mathcal{E}_\lambda(W)$, y está determinado por la igualdad

$$U_\lambda(\mathcal{E}_\lambda(W)) = E[U_\lambda(W)],$$

de tal manera que

$$\mathcal{E}_\lambda(W) = \frac{1}{\lambda} \log\left(E\left[e^{\lambda W}\right]\right). \tag{2.17}$$

A continuación esta noción de equivalente cierto se utilizará para definir un criterio del costo promedio para una cadena de decisión Markoviana, noción que fue introducida en la Sección 2.1. Suponga primero que el sistema va a operar durante n etapas, de tal suerte que el costo aleatorio en que se incurrirá es $\sum_{t=0}^{n-1} C(X_t, A_t)$. Cuando la política $\pi \in \mathcal{P}$ se utiliza para seleccionar las accciones aplicadas y el estado inical es $x \in S$, usando (2.17) se desprende que el equivalente cierto del costo total es

$$J_n(x, \pi) := \frac{1}{\lambda} \log\left(E_x^\pi\left[e^{\lambda \sum_{t=0}^{n-1} C(X_t, A_t)}\right]\right), \tag{2.18}$$

lo que corresponde a un promedio de $J_n(x; \pi)/n$ por etapa, y el mayor punto límite de estos promedios conforme n tiende a ∞ es el *costo promedio superior* $J(x, \pi)$ asociado con la política π en el estado inicial x, esto es,

$$J(x, \pi) := \limsup_{n \to \infty} \frac{1}{n} J_n(x, \pi), \tag{2.19}$$

mientras que el correspondiente costo promedio óptimo está dado por

$$J^*(x) := \inf_{\pi \in \mathcal{P}} J(x, \pi). \tag{2.20}$$

El criterio $J(x, \pi)$ refleja un punto de vista 'pesimista', en el sentido de que mide el desempeño de la política π por medio del mayor punto límite de los costos promedio sensibles al riesgo en un horizonte finito. En contraste, la perspectiva 'optimista' mide el funcionamiento de una política mediante el menor de los puntos límite de la sucesión de costos promedio en horizonte finito, dando origen al *criterio del costo promedio inferior*,

$$J_-(x, \pi) = \liminf_{n\to\infty} \frac{1}{n} J_n(x, \pi), \tag{2.21}$$

cuya función de valor óptimo está dada por

$$J_-^*(x) = \inf_{\pi\in\mathcal{P}} J_-(x, \pi). \tag{2.22}$$

A partir de estas definiciones se desprende de inmediato que

$$J_-(x, \pi) \leq J(x, \pi), \quad \forall x \in S, \quad \pi \in \mathcal{P},$$

y por lo tanto

$$J_-^*(\cdot) \leq J^*(\cdot). \tag{2.23}$$

Bajo las condiciones (de comunicación) que se impondrán en este trabajo, se mostrará que las funciones de valor óptimo superior e inferor $J^*(\cdot)$ y $J_-^*(\cdot)$ coinciden.

Cadenas de decisión dotadas con criterios sensibles al riesgo han sido usadas en diferentes campos, como estudio de inventarios [15, 16], problemas de mantenimiento [19, 32], teoría de aprendizaje [50], y finanzas [5, 6, 12, 65]. Por otro lado, el estudio de cadenas de Markov controladas con criterio del costo promedio puede remontarse, por lo menos, hasta los trabajos fundamentales de Howard y Matheson [39], Jacobson [40], y Jaquette [42, 43]. El caso de modelos con espacio de estados y acciones finitos se estudió primeramente en [39], trabajo donde se utilizó la teoría de Perron-Frobenious sobre matrices positivas [48] para establecer una *ecuación de optimalidad* que caracteriza al costo promedio óptimo, y que a la vez permite determinar una política estacionaria óptima. Ese importante resultado se obtuvo bajo la siguiente condición de comunicación: sin importar la política usada por el controlador, para cada par de estados x y y es posible vistar y cuando el estado inicial es x. La teoría tomó impulso en la década de 1990, con los estudios de [30, 41, 58, 68], los cuales consideraron modelos en tiempo continuo o con espacio de estados de Borel. Modelos discretos, es decir, con espacio de estados numerable y tiempos de observación igual a $\mathbb{N}$, se estudiaron en [29], [33], [47] donde el enfoque estuvo basado en la introducción de un juego estocástico auxiliar. Por otro lado, el criterio del costo promedio (2.19) ha sido estudiado para modelos con con espacio de estados de Borel o costos no acotados en [25–27], donde el índice promedio se estudió por medio de operadores contractivos bajo la condición de que el proceso de estados satisface una condición fuerte de mezclado y que el coeficiente de sensibilidad al riesgo es 'pequeño', mientras que en [35] y [44] se supuso que la función de costo tiene estructura penalizada, en el sentido de que crece sin limite al 'alejarse' de un estado fijo; en estos dos últimos trabajos se usó un enfoque basado en la introducción de un juego auxiliar, mientras que modelos con costos penalizados fueron estudiados via argumentos de programación dinámica en [23]. Por otro lado, en [21] se demostró que los resultados de Howard y Matheson en [39] no pueden trasladarse directamente a modelos no comunicantes; por ejemplo, aún bajo fuertes requerimientos de ergodicidad como la *condición simultánea de Doeblin* (bajo la cual existe un estado z que es accesible desde cualquier otro estado sin importar la política empleada), no es posible asegurar que el costo promedio óptimo esté caracterizado por medio de la ecuación de optimalidad. Adicionalmente, en [17] se mostró que bajo la condición simultánea de Doeblin la ecuación de optimalidad para el criterio (2.19) tiene un a solución si el coeficiente de sensibilidad al riesgo es suficientemente pequeño. En conjunto, estos dos últimos trabajos muestran que obtener una caracterización general del costo promedio óptimo para un valor general del coeficiente de sensibilidad al riesgo es problema interesante que debe estudiarse bajo condiciones fuertes de comunicación.

2.5. El Problema

El problema que se estudia en este trabajo es el siguiente:

• *Aproximar el costo promedio óptimo sensible al riesgo mediante los puntos fijos de operadores contractivos.*

El cual es una extensión al contexto sensible al riesgo de un problema clásico en la teoría de cadenas de Markov controladas. En el caso neutral al riesgo la solución a este problema es un resultado conocido como el *enfoque descontado* al criterio del costo promedio, el cual relaciona los criterios descontado y promedio introducidos en (2.7) y (2.8). Para describir dicho resultado, es necesario introducir la siguiente notación:

Dado $\alpha \in (0, 1)$ sea $V_\alpha^* : S \to \mathbb{R}$ la función de valor óptimo correspondiente al criterio descontado $V_\alpha(x, \pi)$ en (2.7), esto es,

$$V_\alpha^*(x) = \inf_{\pi \in \mathcal{P}} V_\alpha(x, \pi), \quad \forall x \in S.$$

En este caso, es posible verificar que

$$V_\alpha(x) = T_\alpha[V_\alpha](x), \quad \forall x \in S, \tag{2.24}$$

donde para cada función acotada $W : S \to \mathbb{R}$

$$T_\alpha[W](x) = \inf_{a \in A(x)} [C(x,a) + \alpha \sum_{y \in S} p_{x,y}(a) W(y)], \quad \forall x \in S. \tag{2.25}$$

vea, por ejemplo, [36,54,56,67]. Con esta notación, se tiene que T_α es un operador contractivo respecto a la norma del supremo con módulo α, esto es,

$$\|T_\alpha[W] - T_\alpha[\tilde{W}]\| \leq \alpha \|W - \tilde{W}\|$$

y (2.24) establece que V_α^* es punto fijo de T_α, es decir

$$V_\alpha^* = T_\alpha[V_\alpha^*].$$

Ahora, sea

$$V^*(x) = \inf_{\pi \in \mathcal{P}} V(x, \pi), \quad \forall x \in S.$$

la función de valor óptimo correspondiente al criterio del costo promedio $V(x,\pi)$ en (2.8). En este caso, el enfoque descontado establece condiciones bajo las cuales

$$\lim_{\alpha \to 1} (1-\alpha) V_\alpha^*(x) = V^*(x), \quad \forall x \in S. \tag{2.26}$$

Uno de los primeros resultados en esta dirección se encuentra en [56], p. 149, donde la anterior convergencia se estableció bajo la condición de que existe un estado $z \in S$ y $\beta > 0$ tal que

$$p_{x,z}(a) \geq \beta, \quad (x,a) \in \mathbb{K}. \tag{2.27}$$

Posteriormente, la convergencia (2.26) se ha obtenido bajo otras condiciones, como la *condición simultánea de Doeblin* [36], la cual generaliza (2.27) y establece que existe una constante finita B tal que

$$E_x^\pi[T_z] \leq B, \quad \forall x \in S, \quad \pi \in \mathcal{P}; \tag{2.28}$$

vea [2]. El enfoque para obtener la convergencia (2.26) se basa en un hecho fundamental: Bajo la condición (2.28) la función de costo promedio óptimo $V^*(\cdot)$ es constante, digamos g, y dicho valor es el único número real para el cual existe una función acotada $h : S \to \mathbb{R}$ tal que se satisface la siguiente ecuación de optimalidad:

$$g + h(x) = \inf_{a \in A(x)} [C(x,a) + \sum_{y \in S} p_{x\,y}(a) h(y)]. \tag{2.29}$$

Por otro lado, note que (2.24) y (2.25) implican que

$$
\begin{aligned}
&(1-\alpha)V_\alpha^*(x) + \alpha V_\alpha^*(x) \\
&\quad = \inf_{a\in A(x)} [C(x,a) + \alpha \sum_{y\in S} p_{x\,y}(a) V_\alpha^*(y)],
\end{aligned} \tag{2.30}
$$

y comparando estas dos últimas ecuaciones se desprende que, si $(1-\alpha)V_\alpha^*(\cdot)$ es 'aproximadamente constante', digamos g^*, entonces la pareja $(g^*, \alpha V_\alpha^*(\cdot))$ es una solución 'aproximada' de (2.29), de manera que $g^* \approx g$. El argumento que formaliza este enfoque intuitivo consiste en demostrar que, conforme $\alpha \nearrow 1$, $(1-\alpha)V_\alpha^*(x)$ converge y que el límite no depende de x, de manera que si α es cercano a 1, entonces $(1-\alpha)V_\alpha^*(\cdot)$ es aproximadamente igual al costo promedio óptimo.

En el contexto sensible al riesgo en que este trabajo se desarrolla, la solución al problema de aproximación se presentará bajo condiciones que, en general, no implican que el análogo sensible al riesgo de (2.29) se satisface, lo cual es una característica interesante del análisis desarrollado en este trabajo.

Capítulo 3

Ecuación de Optimalidad

En este capítulo se estudia la caracterización del costo promedio óptimo bajo aversión al riesgo en términos de una ecuación de optimalidad. El propósito principal es analizar el teorema de existencia de soluciones que fue establecido originalmente por Howard y Matheson [39] y, como en ese artículo, la exposición cubre el caso de modelos con espacio de estados y acciones finitos y utiliza el enfoque algebraico basado en el teorema de Perron-Frobenious para matrices no negativas [31, 48]. Sin embargo, en contraste con la versión original de Howard y Matheson, el argumento presentado a continuación no impone condición alguna sobre la *aperiodicidad de las cadenas de Markov inducidas por las políticas estacionarias*. La conclusión principal del capítulo puede describirise como sigue: bajo la condición de que cada política estacionaria induce una cadena de Markov *comunicante*, se demuestra que el costo promedio óptimo está caracterizado por la ecuación de optimalidad.

3.1. Introducción

En este capítulo se estudian cadenas de Marov controladas con espacio de estados finito, contexto en el que se analiza la caracterización del costo promedio óptimo suponiendo que el controlador es averso al riesgo y que el modelo es *comunicante*, esto es, bajo cada política estacionaria cualquier estado $y \in S$ puede ser alcanzado sin importar el estado inicial x. Las principales conclusiones son que el costo promedio óptimo no depende del estado inicial, y que su valor común se caracteriza mediante una sola ecuación de optimalidad. Suponiendo adicionalmente que el conjunto de acciones es finito, este resultado fue establecido por Howard y Matheson en [39], utilizando un enfoque algebraico basado en el teorema de Perron-Frobenious sobre matrices positivas, y posteriormente este resultado se extendió al caso de un espacio compacto de acciones en [22], donde se utilizaron técnicas de programación dinámica aplicadas a un problema auxiliar de costo total hasta el retorno a un estado especificado de antemano. En las siguientes secciones se estudian estos resultados presentando argumentos alternativos a los usados en los dos trabajos antes mencionados.

La presentación del capítulo ha sido organizada de la siguiente manera: En la Sección 3.2 se introduce una ecuación de optimalidad sensible al riesgo, la cual, cuando tiene solución, determina el costo promedio óptimo y permite obtener políticas estacionarias óptimas respecto al criterio del costo promedio, conclusiones que se establecen en el Teorema 3.2.1, el cual se conoce como el teorema de verificación. A continuación, en la Sección 3.3 el resultado fundamental de Howard y Matheson sobre la existencia de soluciones de la ecuación de optimalidad bajo aversión al riesgo se enuncia como el Teorema 3.3.1; dicho resultado se refiere a modelos finitos (es decir, con espacios de estados y acciones finitos) y requiere imponer condiciones estrictas de comunicación, las cuales se enuncian formalmente en la Hipótesis 3.3.1. La demostración de este resultado se basa en los resultados auxiliares establecidos en la Sección 3.4 sobre matrices no negativas y comunicantes, particularmente en las relaciones de Collatz-Wielandt [48] que caracterizan el valor propio de módulo máximo, propiedades que se utilizan para demostrar el teorema de existencia de soluciones en la Sección 3.5.

3.2. Teorema de Verificación

En esta sección se formula la caracterización del costo promedio óptimo a través de la ecuación de optimalidad, la cual se formula a continuación como la relación (3.1).

Teorema 3.2.1 [Teorema de verificación.] *Sea $\mathcal{M}$ una cadena de Markov controlada como en la Sección 2.1, y suponga que la constante $g \in \mathbb{R}$ y $h : S \to \mathbb{R}$ satisfacen*

$$e^{\lambda g + \lambda h(x)} = \inf_{a \in A(x)} \left[e^{\lambda C(x,a)} \sum_{y \in S} p_{x\,y}(a) e^{\lambda h(y)} \right], \quad \forall x \in S, \quad (3.1)$$

donde $h(\cdot)$ es acotada, esto es,

$$\|h\| = \sup_{x \in S} |h(x)| < \infty.$$

En estas circunstancias, las siguientes afirmaciones son válidas:

(i) Para cada política $\pi \in \mathcal{P}$

$$J_{-}(x;\pi) \geq g;$$

vea (2.21).

(ii) Para cada $\varepsilon > 0$*, existe* $f_\varepsilon \in \mathbb{F}$ *tal que*

$$e^{\lambda(g+\varepsilon)+\lambda h(x)} \geq e^{\lambda C(x,f_\varepsilon(x))} \sum_{y\in S} p_{x\,y}(f_\varepsilon(x)) e^{\lambda h(y)}, \quad \forall x \in S. \quad (3.2)$$

y esta política f_ε *satisface*

$$g+\varepsilon \geq \limsup_{n\to\infty} \frac{1}{n+1} J_n(x;f_\varepsilon) = J(x,f_\varepsilon), \quad \forall x \in S.$$

(iii) Las funciones de costo promedio óptimo superior e inferior coinciden y son iguales a g*, esto es,*

$$J_{-}^{*}(x) = g = J^{*}(x), \quad \forall x \in S.$$

(iv) Si la política estacionaria $f \in \mathbb{F}$ *es tal que*

$$e^{\lambda g+\lambda h(x)} = e^{\lambda C(x,f(x))} \sum_{y\in S} p_{x\,y}(f(x)) e^{\lambda h(y)}, \quad \forall x \in S, \quad (3.3)$$

entonces

$$g = \lim_{n\to\infty} \frac{1}{n+1} J_n(x;f), \quad \forall x \in S.$$

Demostración. Sea $(g,h(\cdot))$ una solución de la ecuación (3.1). Primeramente, observe que para cada $x \in S$ y $\pi \in \mathcal{P}$ las siguientes relaciones ocurren con probabilidad 1 con respecto a la distribución P_x^π:

$$\begin{aligned} e^{\lambda g+\lambda h(X_t)} &\leq e^{\lambda C(X_t,A_t)} \sum_{y\in S} p_{X_t\,y}(A_t) e^{\lambda h(y)} \\ &= E_x^\pi \left[e^{\lambda C(X_t,A_t)+\lambda h(X_{t+1})} \middle| H_t, A_t \right], \end{aligned} \quad (3.4)$$

donde H_t es la historia del proceso hasta el tiempo t, y la propiedad de Markov se utilizó para establecer la igualdad; vea (2.1) y (2.2).

(i) Considere un estado $x \in S$ y una política $\pi \in \mathcal{P}$ arbitrarios, y note que la variable aleatoria $e^{\sum_{t=0}^{n-1} \lambda C(X_t,A_t)}$ es $\sigma(H_n)$-medible, de tal manera que para todo entero positivo n

$$
\begin{aligned}
&E_x^\pi\left[e^{\lambda\sum_{t=0}^{n}C(X_t,A_t)+\lambda h(X_{n+1})}\Big|\, H_n, A_n\right]\\
&\qquad= e^{\lambda\sum_{t=0}^{n-1}C(X_t,A_t)}E_x^\pi\left[e^{\lambda C(X_n,A_n)+\lambda h(X_{n+1})}\Big|\, H_n, A_n\right]\\
&\qquad\geq e^{\lambda\sum_{t=0}^{n-1}C(X_t,A_t)}e^{\lambda g+\lambda h(X_n)}\\
&\qquad= e^{\lambda g}e^{\lambda\sum_{t=0}^{n-1}C(X_t,A_t)+\lambda h(X_n)},
\end{aligned}
$$

donde la relación (3.4) se usó para establecer la desigualdad. Por lo tanto

$$
\begin{aligned}
&E_x^\pi\left[e^{\lambda\sum_{t=0}^{n}C(X_t,A_t)+\lambda h(X_{n+1})}\right]\\
&\qquad\qquad\geq e^{\lambda g}E_x^\pi\left[e^{\lambda\sum_{t=0}^{n-1}C(X_t,A_t)+\lambda h(X_n)}\right].
\end{aligned}\tag{3.5}
$$

Observe que poniendo $t=0$ en (3.4) se obtiene que

$$
\begin{aligned}
&E_x^\pi\left[e^{\lambda C(X_0,A_0)+\lambda h(X_1)}\right]\\
&\qquad\geq e^{\lambda g+\lambda h(x)},\quad \forall x\in S,\quad \pi\in\mathcal{P},
\end{aligned}\tag{3.6}
$$

y combinando estas dos últimas relaciones, por medio de un argumento de inducción se concluye que

$$
\begin{aligned}
&E_x^\pi\left[e^{\lambda\sum_{t=0}^{n}C(X_t,A_t)+\lambda h(X_{n+1})}\right]\\
&\qquad\geq e^{\lambda(n+1)g+\lambda h(x)},\quad \forall x\in S,\quad \pi\in\mathcal{P},\quad n\in\mathbb{N},
\end{aligned}\tag{3.7}
$$

de tal manera que, usando (2.18) y recordando que $h(\cdot)$ es acotada,

$$
\begin{aligned}
e^{\lambda\|h\|+\lambda J_n(x;\pi)} &= e^{\lambda\|h\|}E_x^\pi\left[e^{\lambda\sum_{t=0}^{n-1}C(X_t,A_t)}\right]\\
&\geq E_x^\pi\left[e^{\lambda\sum_{t=0}^{n-1}C(X_t,A_t)+\lambda h(X_n)}\right]\\
&\geq e^{\lambda ng+\lambda h(x)}\geq e^{\lambda ng-\lambda\|h\|},\quad n=1,2,3,\ldots,
\end{aligned}
$$

Por lo tanto

$$
J_n(x;\pi)\geq ng-2\|h\|,
$$

y como $n\in\mathbb{N}\setminus\{0\}$ es arbitrario, por medio de (2.21) se sigue que

$$
J_-(x;\pi)=\liminf_{n\to\infty}\frac{1}{n}J_n(x;\pi)\geq g,
$$

demostrando la parte (i).

(ii) Sea $(g, h(\cdot))$ una solución de (3.1). Por la definición de ínfimo, para cada $\varepsilon > 0$ y $x \in S$ existe una acción $f_\varepsilon(x) \in A(x)$ tal que

$$e^{\lambda(g+\varepsilon)+\lambda h(x)} \geq e^{\lambda C(x, f_\varepsilon(x))} \sum_{y \in S} p_{x\,y}(f_\varepsilon(x)) e^{\lambda h(y)}, \quad \forall x \in S. \tag{3.8}$$

Recordando que la igualdad

$$P_x^{f_\varepsilon}[A_t = f_\varepsilon(X_t)] = 1 \tag{3.9}$$

es siempre válida y usando la propiedad de Markov, la desigualdad (3.8) equivale a

$$\begin{aligned} &e^{\lambda(g+\varepsilon)+\lambda h(X_t)} \\ &\quad \geq e^{\lambda C(X_t, A_t)} \sum_{y \in S} p_{X_t\,y}(A_t) e^{\lambda h(y)} \\ &\quad = E_x^{f_\varepsilon}\left[e^{\lambda C(X_t, A_t)+\lambda h(X_{t+1})} \,\middle|\, H_t \right] \ P_x^{f_\varepsilon}\text{-c. s.}, \ \forall x \in S, \ t \in \mathbb{N}. \end{aligned} \tag{3.10}$$

Sea $x \in S$ arbitrario pero fijo y observe que $e^{\sum_{t=0}^{n-1} \lambda C(X_t, A_t)}$ es medible respecto a $\sigma(H_n)$, de tal manera que

$$\begin{aligned} &E_x^{f_\varepsilon}\left[e^{\lambda \sum_{t=0}^{n} C(X_t, A_t)+\lambda h(X_{n+1})} \,\middle|\, H_n \right] \\ &\qquad = e^{\lambda \sum_{t=0}^{n-1} C(X_t, A_t)} E_x^{f_\varepsilon}\left[e^{\lambda C(X_n, A_n)+\lambda h(X_{n+1})} \,\middle|\, H_n \right] \\ &\qquad \leq e^{\lambda \sum_{t=0}^{n-1} C(X_t, A_t)} e^{\lambda(g+\varepsilon)+\lambda h(X_n)} \\ &\qquad = e^{\lambda(g+\varepsilon)} e^{\lambda \sum_{t=0}^{n-1} C(X_t, A_t)+\lambda h(X_n)}, \end{aligned}$$

donde la relación (3.10) se usó para establecer la desigualdad. Por lo tanto

$$\begin{aligned} &E_x^{f_\varepsilon}\left[e^{\lambda \sum_{t=0}^{n} C(X_t, A_t)+\lambda h(X_{n+1})} \right] \\ &\qquad \leq e^{\lambda(g+\varepsilon)} E_x^{f_\varepsilon}\left[e^{\lambda \sum_{t=0}^{n-1} C(X_t, A_t)+\lambda h(X_n)} \right]. \end{aligned} \tag{3.11}$$

Notando ahora que el caso $t = 0$ en (3.10) establece que

$$E_x^{f_\varepsilon}\left[e^{\lambda C(X_0, A_0)+\lambda h(X_1)} \right] \leq e^{\lambda(g+\varepsilon)+\lambda h(x)},$$

un argumento de inducción combinando este hecho con (3.11) conduce a

$$\begin{aligned} &E_x^{f_\varepsilon}\left[e^{\lambda \sum_{t=0}^{n-1} C(X_t, A_t)+\lambda h(X_{n+1})} \right] \\ &\qquad \leq e^{\lambda n(g+\varepsilon)+\lambda h(x)}, \quad x \in S, \quad n \in \mathbb{N} \setminus \{0\}, \end{aligned} \tag{3.12}$$

y combinando esta relación con (2.18) se obtiene que, para todo entero positivo n y $x \in S$,

$$\begin{aligned} e^{-\lambda\|h\|+\lambda J_n(x;f_\varepsilon)} &= e^{-\lambda\|h\|} E_x^{f_\varepsilon}\left[e^{\lambda\sum_{t=0}^{n-1} C(X_t,A_t)}\right] \\ &\leq E_x^{f_\varepsilon}\left[e^{\lambda\sum_{t=0}^{n-1} C(X_t,A_t)+\lambda h(X_n)}\right] \\ &\leq e^{\lambda n(g+\varepsilon)+\lambda h(x)} \leq e^{\lambda n(g+\varepsilon)+\lambda\|h\|}. \end{aligned}$$

Por lo tanto $J_n(x;f_\varepsilon) \leq n(g+\varepsilon)+2\|h\|$, y como $n \in \mathbb{N}\setminus\{0\}$ es arbitrario, por medio de (2.19) se sigue que

$$J(x;f_\varepsilon) = \limsup_{n\to\infty} \frac{1}{n} J_n(x;f_\varepsilon) \leq g+\varepsilon,$$

estableciendo la parte (ii).
(iii) Sea $\varepsilon > 0$ arbitrario y sea $f_\varepsilon \in \mathbb{F}$ la política estacionaria en (3.2). En este caso la parte (ii) implica que

$$g+\varepsilon \geq J(\cdot;f_\varepsilon) \geq J^*(\cdot)$$

donde la desigualdad se debe a (2.20). Por lo tanto, puesto que $\varepsilon > 0$ es arbitrario,

$$g \geq J^*(\cdot).$$

Por otro lado, via (2.22) y la parte (i) se obtiene

$$J_-^*(\cdot) = \inf_{\pi\in\mathcal{P}} J(\cdot;\pi) \geq g,$$

y combinando estas dos últimas relaciones con (2.23) se desprende que $J^*(\cdot) = g = J_-^*(\cdot)$.

(iv) Suponga que $f \in \mathbb{F}$ satisface (3.2). En este caso (3.3) ocurre con $\varepsilon = 0$, y un argumento paralelo al usado en la parte (ii) conduce a

$$g \geq J(x;f) = \limsup_{n\to\infty} \frac{1}{n} J_n(x;f);$$

puesto que

$$J_-(x;f) = \liminf_{n\to\infty} \frac{1}{n} J_n(x;f) \geq g$$

se deduce que

$$g = \limsup_{n\to\infty} \frac{1}{n} J_n(x;f) = \liminf_{n\to\infty} \frac{1}{n} J_n(x;f),$$

concluyendo el argumento. ■

3.3. Existencia de Soluciones

En esta seccción se formulan condiciones bajo las cuales existen soluciones $(g, h(\cdot))$ de la ecuación de optimalidad (3.1).

Hipótesis 3.3.1 *(i) El espacio de estados S y el espacio de acciones A son finitos;*

(ii) Bajo la acción de cada política estacionaria $f \in \mathbb{F}$, el espacio de estados es comunicante, esto es,

$$\text{Para cada } x, y \in S \text{ existe } n \equiv n(f, x, y) > 0 \text{ tal que}$$
$$P_x^f[X_n = y] > 0. \tag{3.13}$$

Teorema 3.3.1 *Sea $\mathcal{M} = (S, A, \{A(x)\}_{x \in S}, P, C)$ una cadena de decisión Markoviana que satisface la Hipótesis 3.3.1. En este contexto, para cualquier coeficiente de sensibilidad al riesgo $\lambda > 0$, existen $g \in \mathbb{R}$ y $h : S \to \mathbb{R}$ tales que, para todo $x \in S$,*

$$e^{\lambda g + \lambda h(x)} = \min_{a \in A(x)} \left[e^{\lambda C(x,a)} \sum_{y \in S} p_{x\,y}(a) e^{\lambda h(y)} \right]. \tag{3.14}$$

Este teorema es una versión algo más general que el resultado original de [39], donde, además de la Hipótesis 3.3.1, adicionalmente se suposo que para cada $f \in \mathbb{F}$ la cadena de Markov inducida por f es aperiódica, condición que equivale a que exista un entero $N \equiv N(f)$ tal que

$$P_x^f[X_N = y] > 0, \quad x, y \in S. \tag{3.15}$$

El Teorema 3.3.1 se demostrará en la Sección 3.5, después de establecer los resultados auxiliares necesarios en la siguiente sección.

3.4. Resultados Auxiliares

En esta sección se presentan los resultados técnicos preliminares que se usarán para demostrar el Teorema 3.3.1. En [39] ese resultado se estableció usando que el valor propio positivo de una matriz no negativa es mayor que el módulo de cualquier otro valor propio, un hecho que es parte del teorema clásico de Perron-Frobenious [31, 48]. En este trabajo el enfoque que se usará para demostrar el Teorema 3.3.1 pone de relieve el importante papel del concepto de *matriz comunicante* en el estudio del criterio promedio sensible al riesgo.

Definición 3.4.1 *Sea A una matriz no negativa de orden $k \times k$, esto es,*

$$A_{i\,j} \geq 0, \quad i, j = 1, 2, 3, \ldots, k.$$

En estas circunstancias, A es comunicante si para cada par de enteros $i, j \in \{1, 2, \ldots, k\}$ existe una sucesión $i_0, i_1, \ldots, i_r$ con componentes en $\{1, 2, 3, \ldots, k\}$ tal que

$$i_0 = i, \quad i_r = j, \quad y \quad A_{i_{t-1}, i_t} > 0, \quad t = 1, 2, \ldots, r. \tag{3.16}$$

Observación 3.4.1 *(i) La sucesión $i_0 = i, i_1, \ldots, i_r = j$ en (3.16) se denomina un camino desde i hacia j con longitud r. No es difícil ver que si existe un camino de i hacia j, entonces existe un camino de longitud menor que k.*

(ii) Teniendo en mente la definición de multiplicación matricial, el comentario anterior implica que una matriz no negativa de orden $k \times k$ es comunicante si y sólo si

$$\sum_{t=0}^{k-1} A^t > 0,$$

donde para una matriz M, la desigualdad $M > 0$ siginifica que todas las componentes de M son mayores que cero.

Se usará la siguiente notación. Un vector genérico en $\mathbb{R}^k$ se considera un vector columna y se denota por una letra en negritas, mientras que sus componentes se escriben con la letra normal correspondiente, como en $\mathbf{x} = (x_1, x_2, \ldots, x_k)'$. Por otro lado, $\mathbb{1} = (1, 1, \ldots, 1)' \in \mathbb{R}^k$ denota al vector con todas sus componentes iguales a 1, mientras que para una matriz cuadrada A, la n-ésima potencia se escribe como A^n, de manera que $A^0 = I$ es la matriz identidad, y $A^t = A \times A^{t-1}$ para $t \geq 1$. Finalmente, desigualdades y operaciones con vectores se interpretan componente a componente: por ejemplo, si $\mathbf{x} = (x_1, x_2, \ldots, x_k)'$ y $\mathbf{y} = (y_1, y_2, \ldots, y_k)'$ entonces

$$\mathbf{x} \leq \mathbf{y} \iff x_i \leq y_i, \quad i = 1, 2, \ldots, k,$$

y

$$\mathbf{x}^\alpha = (x_1^\alpha, x_2^\alpha, \ldots, x_k^\alpha)'$$

siempre que el lado derecho tenga sentido. Por otro lado, si $\mathbb{K}$ es un espacio topológico, $\mathcal{C}(\mathbb{K})$ denota la clase de todas las funciones reales y acotadas definidas en $\mathbb{K}$, y $\|C\|: = \sup_{x \in \mathbb{K}} |C(x)| < \infty$ representa la norma del supremo de $C \in \mathcal{C}(\mathbb{K})$. Dado un evento N, la correspondiente función indicadora se representa mediante $I[N]$, y todas las relaciones que involucran esperanzas condicionales son válidas con probabilidad 1 con respecto a la medida de probabilidad en contexto.

La herramienta principal que se usará para establecer la existencia de una solución de la ecuación de optimalidad (3.1) es el siguiente resultado, el cual es parte de las conclusiones del teorema de Perron- Frobenious [31, 48].

Teorema 3.4.2 *Sea A una matriz de orden $k \times k$ con componentes no negativas,y suponga que A es comunicante. En este contexto, las afirmaciones a continuación son válidas.*

(i) A tiene un valor propio positivo que admite un vector propio positivo (esto es un vector con todas sus componentes mayores que cero). Más precisamente, existen $\mu > 0$ y $\mathbf{m} \in (0, \infty)^k$ tales que

$$A\mathbf{m} = \mu\mathbf{m};$$

el par $(\mu, \mathbf{m})$ se denomina un eigenpar positivo *de A.*

(ii) El valor propio positivo μ en la parte (i) es único y es igual a la tasa de crecimiento de los iterados multiplicativos de A, esto es, para cada vector no nulo $\mathbf{x} \in [0, \infty)^k$,

$$\lim_{n\to\infty} [A^n \mathbf{x}]^{1/n} = \mu\mathbb{1}.$$

(iii) Un sub-eigenvector o super-eigenvector no negativo *de A correspondiente a μ es, necesariamente, un eigenvector. Más precisamente,*

$$\begin{aligned} &si\ \mathbf{x} \in [0, \infty)^k \\ &A\mathbf{x} \geq \mu\mathbf{x} \quad \text{ó} \quad A\mathbf{x} \leq \mu\mathbf{x} \implies A\mathbf{x} = \mu\mathbf{x}. \end{aligned}$$

Como se hará patente en el argumento usado para demostrar este resultado, la demostración de este teorema es más sencilla cuando todas las componentes de A son positivas, un caso que será analizado incialmente. Así mismo, el argumento siguiente mostrará que la parte (ii) se desprende de inmediato de la parte (i), mientras que la parte (iii) es consecuencia de la condición de comunicación. La siguiente notación será útil.

Definición 3.4.3 *Sea A una matriz positiva de orden $k \times k$.*

(i) Para cada $\mathbf{x} = (x_1, \ldots, x_k)' \in (0, \infty)^k$ defina

$$\mu(\mathbf{x}) = \min \left\{ \frac{[A\mathbf{x}]_i}{x_i} \,\middle|\, i = 1, 2, \ldots, k \right\}, \tag{3.17}$$

y

$$\mu^* = \sup_{\mathbf{x} \in (0,\infty)^k} \mu(\mathbf{x}). \tag{3.18}$$

(ii) El número a asociado con la matriz A se define mediante

$$a = \max\left\{ \frac{A_{i\,j}}{A_{r\,j}} \,\middle|\, i, r, j = 1, 2, \ldots k \right\},$$

mientras que el conjunto $\mathcal{K}$ está dado por

$$\mathcal{K} = \left\{ \mathbf{x} \in (0, \infty)^k \,\middle|\, \frac{1}{ak} \leq x_i \leq \frac{a}{k},\ i = 1, 2, \ldots, k \right\}.$$

Note que $\mu(\cdot) > 0$, puesto que todas las componentes de A son positivas, una propiedad que también implica que a está bien definido como un número real positivo, de manera que $\mathcal{K}$ es un subconjunto compacto del cono $(0, \infty)^k$. Las siguientes propiedades se desprenden directamente de la Definición 3.4.3:

$$\mu(c\mathbf{x}) = \mu(\mathbf{x}), \quad \text{y} \quad A\mathbf{x} \geq \mu(\mathbf{x})\mathbf{x}, \quad \mathbf{x} \in (0, \infty)^k, \quad c > 0. \tag{3.19}$$

A continuación, se mostrará que μ^* en (3.18) es el valor propio positivo de la matriz A que admite un vector propio positivo; tal caracterización es una relación de Collatz-Wielandt [48]. El punto de partida hacia la demostración del Teorema 3.4.2 es el siguiente.

Lema 3.4.4 *Sea A una matriz de orden $k \times k$ tal que*

$$A_{i\,j} > 0, \quad i, j = 1, 2, 3, \ldots, k.$$

Usando la notación de la Definición 3.4.3, las siguientes afirmaciones son válidas.

(i) Para cada $\mathbf{x} \in (0, \infty)^k$

$$\mu(A\mathbf{x}) \geq \mu(\mathbf{x}).$$

(ii) La inclusión

$$\frac{A\mathbf{x}}{\mathbb{1}' A\mathbf{x}} \in \mathcal{K}$$

es válida para todo $\mathbf{x} \in (0, \infty)^k$.

(iii) En la definición de μ^* el máximo sobre $x \in (0, \infty)^k$ puede reemplazarse por el máximo sobre $\mathbf{x} \in \mathcal{K}$, esto es,

$$\mu^* = \sup_{\mathbf{x} \in \mathcal{K}} \mu(\mathbf{x}).$$

(iv) Existe $\mathbf{y}^* \in \mathcal{K} \subset (0, \infty)^k$ tal que $\mu^* = \mu(\mathbf{y}^*)$.

Demostración. (i) Sea $\mathbf{x} \in (0,\infty)^k$ arbitrario pero fijo, y defina $\mathbf{z} = A\mathbf{x}$. Usando la desigualdad en (3.19) se desprende que $A\mathbf{z} = A(A\mathbf{x}) \geq A(\mu(\mathbf{x})\mathbf{x}) = \mu(\mathbf{x})A\mathbf{x}$, es decir, $A\mathbf{z} \geq \mu(\mathbf{x})\mathbf{z}$, una relación que conduce a $\mu(A\mathbf{x}) = \mu(\mathbf{z}) \geq \mu(\mathbf{x})$, por la Definición 3.4.3(i).

(ii) A partir de la especificación del número a en la Definición 3.4.3(ii), se tiene que

$$A_{i\,j} \leq aA_{r\,j}, \quad r, i, j = 1, 2, 3, \ldots, k. \tag{3.20}$$

Dado un índice r entre 1 y k, esto implica que, para cada $\mathbf{x} \in (0,\infty)^k$,

$$\mathbb{1}'A\mathbf{x} = \sum_{i=1}^{k}\sum_{j=1}^{k} A_{i\,j}x_j \leq \sum_{i=1}^{k}\sum_{j=1}^{k} aA_{r\,j}x_j = a\sum_{i=1}^{k}[A\mathbf{x}]_r = ak[A\mathbf{x}]_r$$

de manera que

$$\frac{1}{ak} \leq \frac{[A\mathbf{x}]_r}{\mathbb{1}'A\mathbf{x}}, \quad r = 1, 2, 3, \ldots, k. \tag{3.21}$$

Análogamente, para un índice fijo r y $\mathbf{x} \in (0,\infty)^k$, usando (3.20) se obtiene que

$$k[A\mathbf{x}]_i = \sum_{r=1}^{k}\sum_{j=1}^{k} A_{i\,j}x_j \leq \sum_{r=1}^{k}\sum_{j=1}^{k} aA_{r\,j}x_j = a\mathbb{1}'A\mathbf{x},$$

y entonces

$$\frac{[A\mathbf{x}]_i}{\mathbb{1}'A\mathbf{x}} \leq \frac{a}{k}, \quad i = 1, 2, 3, \ldots, k,$$

y combinando esta relación con (3.21) se obtiene la conclusión deseada; vea (3.4.3).

(iii) Observe que (3.18) y la parte (i) implican que

$$\mu^* = \sup_{\mathbf{x}\in(0,\infty)^k} \mu(A\mathbf{x}),$$

un hecho que combinado con la igualdad en (3.19) conduce a

$$\mu^* = \sup_{\mathbf{x}\in(0,\infty)^k} \mu\left(\frac{A\mathbf{x}}{\mathbb{1}'A\mathbf{x}}\right) = \sup_{\mathbf{y}\in\mathcal{K}} \mu(\mathbf{y}).$$

donde la segunda igualdad se debe a la parte (ii).

(iv) Puesto que $\mathcal{K}$ es un conjunto compacto, y $\mu(\cdot)$ es una función continua, se tiene que existe $\mathbf{y}^* \in \mathcal{K}$ tal que $\mu(\mathbf{y}^*) = \sup_{\mathbf{y}\in\mathcal{K}} \mu(\mathbf{y})$, y entonces $\mu^* = \mu(\mathbf{y}^*)$, por la parte (iii). ∎

A continuación, el lema anterior se usará para demostrar el principal resultado de esta sección.

Demostración del Teorema 3.4.2. El argumento se ha dividido en dos casos:

Caso 1: Todas las componentes de A son positivas.

En este entorno, sea μ^* el número en (3.18) y, usando el Lema 3.4.4(iv), seleccione un vector $\mathbf{y}^* \in (0,\infty)^k$ tal que $\mu^* = \mu(\mathbf{y}^*)$, de manera que

$$A\mathbf{y}^* \geq \mu^*\mathbf{y}^*,$$

por (3.19).

(i) Se mostrará que

$$A\mathbf{y}^* = \mu^*\mathbf{y}^*. \tag{3.22}$$

Para verificar esta relación, sea

$$\mathbf{z} = A\mathbf{y}^* - \mu^*\mathbf{y}^*$$

y note que $\mathbf{z} \geq 0$. Ahora, *suponga* que $\mathbf{z}$ es un vector no nulo. En esta circunstancia, usando que $A_{i\,j} > 0$ para todo índice i, j, se desprende que $A\mathbf{z} > 0$, es decir,

$$A(A\mathbf{y}^*) - \mu^* A\mathbf{y}^* > 0,$$

de tal manera que existe $\varepsilon > 0$ tal que $A(A\mathbf{y}^*) - \mu^* A\mathbf{y}^* \geq \varepsilon\mu^* A\mathbf{y}^*$, esto es, $A(A\mathbf{y}^*) - (1+\varepsilon)\mu^* A\mathbf{y}^* > 0$, y entonces

$$A(A\mathbf{y}^*) > (1+\varepsilon)\mu^* A\mathbf{y}^*.$$

Por la Definición 3.4.3, esta desigualdad implica que $\mu(A\mathbf{y}^*) > (1+\varepsilon)\mu^*$, contradiciendo el hecho de que $\mu^*(>0)$ es el máximo de la función $\mu(\cdot)$. Esta contradicción proviene del supuesto de que el vector $\mathbf{z}$ es no nulo, y se concluye que $\mathbf{z} = \mathbf{0}$, igualdad que equivale a (3.22). Luego, el par $(\mu, \mathbf{m}) \equiv (\mu^*, \mathbf{y}^*)$ satisface la primera conclusión del Teorema 3.4.2.

(ii) Observe que $A\mathbf{m} = \mu\mathbf{m}$ implica que $A^n\mathbf{m} = \mu^n\mathbf{m}$ para todo entero $n > 0$, de tal forma que $[A^n\mathbf{m}]^{1/n} = \mu[\mathbf{m}]^{1/n}$, y entonces

$$\lim_{n\to\infty} [A^n\mathbf{m}]^{1/n} = \mu \lim_{n\to\infty} [\mathbf{m}]^{1/n} = \mu\mathbb{1}. \tag{3.23}$$

Ahora sea $\mathbf{x} \in [0,\infty)^k$ un vector no nulo arbitrario pero fijo, y note que $\tilde{\mathbf{x}} = A\mathbf{x} > 0$. Puesto que $\mathbf{m}$ tiene componentes positivas, se sigue que existen constantes positivas c_0 y c_1 tales que $c_0\mathbf{m} \leq \tilde{\mathbf{x}} \leq c_1\mathbf{m}$, de tal manera que

$$c_0 A^{n-1}\mathbf{m} \leq A^{n-1}\tilde{\mathbf{x}} = A^n\mathbf{x} \leq c_1 A^{n-1}\mathbf{m};$$

tomando la raíz n-ésima y después el límite conforme n tiende a ∞ en esta relación, por medio de (3.23) se desprende que $[A^n\mathbf{x}]^{1/n} \to \mu\mathbb{1}$, lo cual implica la unicidad del valor propio positivo μ.

(iii) Suponga que $\mathbf{x} \in [0,\infty)^k$ satisface $A\mathbf{x} \geq \mu\mathbf{x}$, de manera que $\mathbf{z} = A\mathbf{x} - \mu\mathbf{x} \geq 0$. *Suponga además que* $\mathbf{z}$ *es no nulo*. En este caso $\mathbf{x} \neq \mathbf{0}$ y, usando que $A_{i\,j} > 0$ para todos los índices i, j, se sigue que $A\mathbf{z} > 0$, es decir,

$$A\mathbf{z} = A(A\mathbf{x} - \mu\mathbf{x}) = A(A\mathbf{x}) - \mu A(\mathbf{x}) > 0.$$

Por lo tanto, existe $\varepsilon > 0$ tal que $A\mathbf{y} - (1+\varepsilon)\mu\mathbf{y} \geq 0$, donde $\mathbf{y} = A\mathbf{x}(> 0)$, y entonces $A^n\mathbf{y} \geq (1+\varepsilon)^n\mu^n\mathbf{y}$ para $n = 1, 2, 3, \ldots$; así, $\lim_{n\to\infty}[A^n\mathbf{y}]^{1/n} \geq (1+\varepsilon)\mu\mathbb{1}$, lo cual contradice la parte (ii). Por lo tanto si $\mathbf{x} \in [0,\infty)^k$ satisface que $A\mathbf{x} \geq \mu\mathbf{x}$, entonces $A\mathbf{x} = \mu\mathbf{x}$. De forma similar, se puede verificar que si $A\mathbf{x} \leq \mu\mathbf{x}$ para algún vector no nulo $\mathbf{x} \in [0,\infty)^k$, entonces $A\mathbf{x} = \mu\mathbf{x}$. Esto completa la demostración del Teorema 3.4.2 cuando todas las componentes de A son mayores que cero.

Caso 2: La matriz no negativa y comunicante A es arbitraria.

Defina las matrices $\hat{A}$ y $\tilde{A}$ por

$$\hat{A} = I + A \quad \text{y} \quad \tilde{A} = \hat{A}^k$$

y observe que

$$\tilde{A} = \sum_{r=0}^{k} \binom{k}{r} A^r > 0,$$

donde la Observación 3.4.1(ii) se uso para establecer la desigualdad. Usando el caso 1 aplicado a la matriz $\tilde{A}$, se desprende que existe un par $(\tilde{\mu}, \tilde{\mathbf{m}})$ que satisface

$$\tilde{\mu} > 0 \quad \text{y} \quad \tilde{\mathbf{m}} \in (0,\infty)^k$$

y

$$\tilde{A}\tilde{\mathbf{m}} = \tilde{\mu}\tilde{\mathbf{m}}. \tag{3.24}$$

En este caso

$$\tilde{\mu}\mathbb{1} = \lim_{n\to\infty}[\tilde{A}^n\mathbf{m}]^{1/n} \geq \lim_{n\to\infty}[\mathbf{m}]^{1/n} = \mathbb{1},$$

donde la desigualdad se debe a que $\tilde{A} \geq I$. Luego,

$$\tilde{\mu} \geq 1.$$

Defina

$$\hat{\mu} = [\tilde{\mu}]^{1/k}, \quad \mathbf{m} = \left[\sum_{r=0}^{k-1} \hat{\mu}^{k-1-r}\hat{A}^r\right]\tilde{\mathbf{m}} \quad \text{y} \quad \mu = \hat{\mu} - 1, \tag{3.25}$$

y observe la siguiente factorización:

$$
\begin{aligned}
\tilde{A} - \tilde{\mu} I &= \hat{A}^k - \hat{\mu}^k I \\
&= (\hat{A} - \hat{\mu} I) \left[\sum_{r=0}^{k-1} \hat{\mu}^{k-1-r} \hat{A}^r \right] \\
&= \left[\sum_{r=0}^{k-1} \hat{\mu}^{k-1-r} \hat{A}^r \right] (\hat{A} - \hat{\mu} I).
\end{aligned} \tag{3.26}
$$

Usando que $\hat{A} \geq A$ y $\tilde{\mu} \geq 1$, la Observación 3.4.1(ii) implica que la matriz entre corchetes en la anterior expresión tiene componentes positivas. Por lo tanto, puesto que $\tilde{\mathbf{m}} > 0$, se desprende que $\mathbf{m} > 0$. Por otro lado, note que (3.24)–(3.26) implican que

$$
0 = (\tilde{A} - \tilde{\mu} I)\tilde{\mathbf{m}} = (\hat{A} - \hat{\mu} I) \left[\sum_{r=0}^{k-1} \hat{\mu}^{k-1-r} \hat{A}^r \right] \tilde{\mathbf{m}} = (\hat{A} - \hat{\mu} I)\mathbf{m}.
$$

Por lo tanto, $\hat{A}\mathbf{m} = \hat{\mu}\hat{\mathbf{m}}$, igualdad que equivale a

$$
A\mathbf{m} = \mu \mathbf{m}; \tag{3.27}
$$

vea (3.23) y (3.25).

(i) Puesto que $\mathbf{m}$ tiene coordenadas positivas, usando (3.27) es suficiente mostrar que $\mu > 0$. Para lograr este objetivo, observe que $\mu \neq 0$. En efecto, si $\mu = 0$ la anterior relación desplegada implica que $A\mathbf{m} = 0$, y entonces, puesto que A tienen componentes no negativas y $\mathbf{m} \in (0, \infty)^k$, se desprende que la matriz A es nula, contradiciendo que A es comunicante. Por otro lado, como ya se notó, la relación $\tilde{\mu} \geq 1$ ocurre, de manera que $\mu \geq 0$ (por (3.25)), y entonces μ tiene todas sus componentes positivas.

(ii) Esta parte puede establecerse usando un argumento similar al empleado en el Caso 1.

(iii) Sea $\mathbf{x} \in [0, \infty)^k$ un vector tal que $A\mathbf{x} \geq \mu \mathbf{x}$. En este caso (3.4) y (3.25) impican que

$$
\hat{A}\mathbf{x} - \hat{\mu}\mathbf{x} \geq 0,
$$

así como $\tilde{A}\mathbf{x} \geq \tilde{\mu}\mathbf{x}$. A partir de esta desigualdad, una aplicación del Caso 1 a la matriz $\tilde{A}$ implica que $(\tilde{A} - \tilde{\mu} I)\mathbf{x} = 0$, y combinando este hecho con la factorización en (3.26) se desprende que

$$
\left[\sum_{r=0}^{k-1} \hat{\mu}^{k-1-r} \hat{A}^r \right] (\hat{A} - \hat{\mu} I)\mathbf{x} = 0;
$$

Como ya se observó, todas las componentes de la matriz encerrada en paréntesis rectangulares son positivas, mientras que el vector

$\hat{A}\mathbf{x} - \hat{\mu}\mathbf{x}$ tiene coordenadas no negativas, así que la relación desplegada arriba implica que $(\hat{A} - \hat{\mu}I)\mathbf{x} = 0$, lo cual es equivalente a $A\mathbf{x} = \mu\mathbf{x}$, por (3.4) y (3.25). Usando un argumento similar es posible mostrar que si $A\mathbf{x} \leq \mu\mathbf{x}$ para algún $\mathbf{x} \in [0, \infty)^k$, entonces $A\mathbf{x} = \mu\mathbf{x}$. ■

3.5. Soluciones de la Ecuación de Optimalidad

En esta sección se usará el Teorema 3.4.2 para demostrar la existencia de soluciones de la ecuación de optimalidad (3.1). Sea $\mathcal{M} = (S, A, \{A(x)\}_{x\in S}, P, C)$ un proceso de decisión Markoviano con espacios de estados y acciones finitos, de manera que el espacio $\mathbb{F} = \prod_{x\in S} A(x)$ de políticas estacionarias también es finito. Para cada $f \in \mathbb{F}$ defina la matriz $A(f)$ cuyas filas y columnas se etiquetan por los elementos de S como sigue;

$$A(f)_{x\,y} = e^{\lambda C(x,f(x))} p_{x\,y}(f(x)), \quad x, y \in S. \tag{3.28}$$

A continuación, suponga que $x, y \in S$ y el entero positivo n son tales que $P_x^f[X_n = y] > 0$. En este caso , existen estados

$$x_1, x_2, \ldots x_{n-1} \in S$$

tales que

$$P_x^f[X_1 = x_1, X_2 = x_2, \ldots, X_{n-1} = x_{n-1}, X_n = y] > 0,$$

relación que explícitamente se escribe como

$$p_{x\,x_1}(f(x))p_{x_1\,x_2}(f(x))\cdots p_{x_{n-2}\,x_{n-1}}(f(x))p_{x_{n-1}\,y}(f(x)) > 0,$$

desigualdad que *via* (3.28) equivale a

$$A(f)_{x\,x_1}A(f)_{x_1\,x_2}A(f)_{x_2\,x_3}\cdots A(f)_{x_{n-2}\,x_{n-1}}A(f)_{x_{n-1}\,y} > 0,$$

de tal manera que, bajo la condición (3.13) en el Teorema 3.3.1, cada matriz $A(f)$ es comunicante (vea Definición 3.4.1). Por lo tanto, una aplicación del Teorema 3.4.2 implica que para cada $f \in \mathbb{F}$, existen $\mu(f) > 0$ y $\mathbf{m}(f) \in (0, \infty)^k$ tales que

$$\mu(f)\mathbf{m}(f) = A(f)\mathbf{m}(f). \tag{3.29}$$

Defina

$$\mu^* = \min_{f\in\mathbb{F}} \mu(f) \tag{3.30}$$

y, recordando que $\mathbb{F}$ es finito, seleccione $f^* \in \mathbb{F}$ tal que

$$\mu(f^*) = \mu^*. \tag{3.31}$$

Ahora defina

$$\mathbf{m}^* = \mathbf{m}(f^*).$$

Con esta notación se tiene que $\mu^* \mathbf{m}^* = A(f^*)\mathbf{m}^*$, igualdad que puede escribirse como

$$\mu^* \mathbf{m}^*_x = e^{\lambda C(x, f^*(x))} \sum_{y \in S} p_{x\,y}(f^*(x)) \mathbf{m}^*_y, \quad \forall x \in S; \tag{3.32}$$

Demostración del Teorema 3.3.1. Se mostrará que

$$\mu^* \mathbf{m}^*_x = \min_{a \in A(x)} \left[e^{\lambda C(x,a)} \sum_{y \in S} p_{x\,y}(a) \mathbf{m}^*_y \right], \quad \forall x \in S. \tag{3.33}$$

Con este fin, para cada $x \in S$ seleccione un minimizador $\tilde{f}(x) \in A(x)$ del lado derecho de esta ecuación, de tal manera que

$$\begin{aligned} & e^{\lambda C(x, \tilde{f}(x))} \sum_{y \in S} p_{x\,y}(\tilde{f}(x)) \mathbf{m}^*_y \\ & \quad = \min_{a \in A(x)} \left[e^{\lambda C(x,a)} \sum_{y \in S} p_{x\,y}(a) \mathbf{m}^*_y \right] \\ & \quad \le e^{\lambda C(x, f^*(x))} \sum_{y \in S} p_{x\,y}(f^*(x)) \mathbf{m}^*_y \; = \mu^* \mathbf{m}^*_x, \end{aligned} \tag{3.34}$$

donde se utilizó la igualdad (3.32) en el último paso. Como $x \in S$ es arbitrario, usando (3.28) esta relación equivale a $A(\tilde{f})\mathbf{m}^* \le \mu^* \mathbf{m}^*$, y entonces, de $\mu^* \le \mu(\tilde{f})$ (por (3.30)), y por lo tanto

$$A(\tilde{f})\mathbf{m}^* \le \mu^* \mathbf{m}^* \le \mu(\tilde{f})\mathbf{m}^*.$$

Luego, iniciando con la relación $A(\tilde{f})\mathbf{m}^* \le \mu(\tilde{f})\mathbf{m}^*$, el Teorema 3.4.2(iii) aplicado a la matriz $A(\tilde{f})$ permite concluir que $A(\tilde{f})\mathbf{m}^* = \mu(\tilde{f})\mathbf{m}^*$, igualdad que combinada con la anterior relación desplegada implica que $A(\tilde{f})\mathbf{m}^* = \mu^* \mathbf{m}^*$, esto es, para cada $x \in S$,

$$e^{\lambda C(x, \tilde{f}(x))} \sum_{y \in S} p_{x\,y}(\tilde{f}(x)) \mathbf{m}^*_y = \mu^* \mathbf{m}^*_x;$$

vea (3.28). Combinando esta igualdad y (3.34) se tiene que (3.33) ocurre. Para finalizar, defina

$$g = \frac{1}{\lambda} \log(\mu^*) \quad \text{y} \quad h(x) = \frac{1}{\lambda} \log(\mathbf{m}^*_x), \quad \forall x \in S.$$

Con esta notación (3.33) equivale a la ecuación de optimalidad (3.1). ■

Capítulo 4

Enfoque Descontado Bajo Aversión al Riesgo

Este capítulo trata sobre procesos de decisión Markovianos con espacio de estados numerable. La hipótesis básica es que el controlador tiene un coeficiente de sensibilidad al riesgo constante y positivo, por lo que el tomador de decisiones es averso al riesgo. En este contexto el funcionamiento de una política de toma de decisiones se mide mediante el correspondiente criterio del costo promedio asociado a una función de costo acotada. Bajo condiciones de comunicación que aseguran que el costo promedio es constante, aunque no necesariamente se caracteriza mediante una ecuación de optimalidad, la principal contribución consiste en formular un criterio descontado mediante el cual es posible aproximar el costo promedio óptimo sensible al riesgo.

4.1. Introducción

En este capítulo se estudian *procesos de decisión Markovianos* (PDMs) con espacio de estados numerable bajo el supuesto básico de que el controlador tiene un coeficiente de sensibilidad al riesgo constante y positivo, de manera que es averso al riesgo. Para valorar el funcionamiento de una estrategia de control se utiliza el criterio (superior) del costo promedio asociado a una función de costo acotada y no negativa, mientras que en lo que refiere al entorno de trabajo, además de suponer condiciones de continuidad y compacidad que son estándar en la teoría de PDMs, las siguientes dos condiciones sobre la ley de transición determinan el marco de referencia para el análisis subsecuente:

(i) El espacio de estados es comunicante bajo la acción de cualquier política estacionaria, y

(ii) La condición simultánea de Doeblin se satisface.

Dentro de este contexto, recientemente se ha demostrado que es posible que el costo promedio óptimo no se determine mediante la ecuación de optimalidad usual [20]; sin embargo, para un *índice descontado* asociado a un juego auxiliar de suma cero, *la función de valor óptimo* es el punto fijo de un operador contractivo, de tal manera que *se caracteriza a través de una ecuación*, la cual es similar a la ecuación de optimalidad para el costo promedio. Luego, es interesante estudiar el siguiente 'problema del enfoque descontado':

• *Aproximar* el costo promedio óptimo con sensibilidad al riego mediante los puntos fijos asociados con una familia de operadores contractivos.

Este problema es clásico en la teoría de PDMs con criterio promedio neutral al riesgo [2, 54], y los resultados presentados más adelante en este trabajo extienden las conclusiones obtenidas en [26], donde se supuso que el coeficiente de sensibilidad al riesgo es 'suficientemente pequeño', y en [1], donde se estudiaron modelos con espacio de estados finito. El problema anterior conduce naturalmente a estudiar la siguiente cuestión, la cual es interesante por sí misma.

• Estudiar la igualdad de las funciones de valor óptimo para los criterios de costo promedio superior e inferior.

Bajo el supuesto de que el tomador de decisiones es averso al riesgo, el criterio del costo promedio fue inicialmente estudiado para PDMs con espacio de estados finito en un importante trabajo de Howard y Matheson [39], y el interés en el tema ha sido estimulado fuertemente por las aplicaciones, por ejemplo, en finanzas [5]. PDMs conducidos por un controlador sensible al riesgo sobre un espacio de estados finito o numerable se estudian en [14, 24, 33, 34], o [62, 63], mientras que modelos con un espacio de estados de Borel se analizan en [25–27], y [44]. Aplicaciones a problemas de finanzas se encuentran en [5, 52, 64], y otros criterios con sensibilidad al riesgo se estudian en [5, 6, 45, 61], mientras que *juegos estocásticos* con índice sensible al riesgo se analizan en [4, 7].

La principal contribución de este trabajo consiste en presentar una solución a los dos problemas planteados líneas arriba, y se establece formalmente en el Teorema 4.3.1 de la Sección 4.3. En términos generales, la estrategia para establecer ese resultado puede reumirse como sigue: Primero se establecerá el teorema para el caso de una función de costo con soporte compacto, y entonces el caso general se obtendrá a través de un proceso de aproximación.

La presentación del capítulo ha sido organizada de la siguiente manera: En la Sección 4.2 se describe formalmente el modelo de decisión y se define el criterio del costo promedio con sensibilidad al riesgo. Posteriormente, en la Sección 4.3 se introduce la familia de funciones de valor óptimo descontado, y el resultado principal donde se presenta la solución a los problemas planteados anteriormente se enuncia en el Teorema 4.3.1. A continuación, en la Sección 4.4 se estudian propiedades básicas de acotamiento de las funciones de valor relativo descontado, las cuales se utilizan en la Sección 4.5 para establecer el teorema principal para funciones de costo con soporte compacto, resultado que se aplica en la Sección 4.6 para demostrar el Teorema 4.3.1 para el caso de una función de costo general en el espacio de las funciones acotadas.

4.2. Modelo de Decisión

Como punto de partida, es conveniente introducir la siguiente notación, la cual se usará consistentemente en el resto del capítulo. El conjunto de enteros no negativos se denota mediante $\mathbb{N}$, mientras que, dado un espacio topológico $\mathbb{K}$, el espacio $\mathcal{C}(\mathbb{K})$ consiste de todas las funciones continuas $C : \mathbb{K} \to \mathbb{R}$ que son acotadas, esto es, $\|C\| < \infty$, donde $\|C\| := \sup_{x\in\mathbb{K}} |C(x)| < \infty$ es la norma del supremo de C. Por otro lado, dado un evento N, la correspondiente función indicadora se denota mediante $I[N]$ y, por convención, se entenderá que cualquier relación entre variables aleatorias ocurre casi seguramente con respecto a la medida de probabilidad bajo consideración.

Sea $\mathcal{M} = (S, A, \{A(x)\}_{x\in S}, C, P)$ un proceso de decisión Markoviano, donde el *espacio de estados* S es un conjunto numerable y no vacío dotado con la topología discreta, el espacio métrico A es el *conjunto de acciones* y, para cada $x \in S$, $A(x)$ es la clase de acciones (controles) admisibles en el estado x. Por otro lado, $C : \mathbb{K} \to \mathbb{R}$ es la función de costo, donde $\mathbb{K} := \{(x, a) : a \in A(x), x \in S\}$ es la familia de pares admisibles, mientras que la ley de transición $P = [p_{x,y}(a)]$ es un núcleo estocástico sobre S dado $\mathbb{K}$. La interpretación del modelo $\mathcal{M}$ es la siguiente: En cada tiempo $t \in \mathbb{N}$ el tomador de decisiones observa el estado X_t de un sistema dinámico, digamos $X_t = x \in S$, y aplica una acción $A_t = a$, la cual se selecciona dentro del conjunto $A(x)$ de acciones permitidas en el estado x. Como consecuencia de esa intervención, se incurre en un costo $C(x, a)$ e, independientemente del registro de acciones y estados previos, en el tiempo $t + 1$ el sistema se moverá al estado $X_{t+1} = y$ con probabilidad $p_{x,y}(a) \geq 0$, donde $\sum_{y \in S} p_{x,y}(a) = 1$; ésta última es la propiedad de Markov del proceso de decisiones. Las siguientes hipótesis se supondrán en el resto del capítulo.

Hipótesis 4.2.1 *(i) Para cada $x \in S$, $A(x)$ es un subconjunto compacto de A;*
(ii) Para cada estado x, la función $a \mapsto p_{x,y}(a)$ es continua en $a \in A(x)$.

Hipótesis 4.2.2 *$C(\cdot) \geq 0$ y $C \in \mathcal{C}(\mathbb{K})$.*

Políticas. Para cada $n \in \mathbb{N}$, sea $\mathbb{H}_n$ el conjunto de historias posibles del proceso de decisión hasta el tiempo $n \in \mathbb{N}$, esto es, $\mathbb{H}_0 := S$ y $\mathbb{H}_n := \mathbb{K}^n \times S$, $n \geq 1$; un elemento genérico de $\mathbb{H}_n$ se denota por $\mathbf{h}_n = (x_0, a_0, x_1, a_1, \ldots, x_{n-1}, a_{n-1}, x_n)$, donde $x_n \in S$ y $(x_i, a_i) \in \mathbb{K}$ para $0 \leq i < n$. Una *política* $\pi = \{\pi_t\}_{t \in \mathbb{N}}$ es una sucesión de núcleos estocásticos π_n sobre A dado $\mathbb{H}_n$, los cuales satisfacen la relación $\pi_n(A(x_n)|\mathbf{h}_n) = 1$ para cada $n \in \mathbb{N}$. Si la historia observada hasta el tiempo n es $\mathbf{h}_n \in \mathbb{H}_n$, entonces bajo la política π la acción A_n será selecccionada por el controlador dentro de un conjunto de Borel $B \subset A$ con probabilidad $\pi_n(B|\mathbf{h}_n)$; la clase de todas las políticas se denota por $\mathcal{P}$. Dado el estado inicial $x \in S$, y la política $\pi \in \mathcal{P}$ utilizada para seleccionar acciones, la distribución del proceso $\{(X_t, A_t)\}$ se determina de manera única por medio del teorema de Tulcea [36, pp. 4–6], [54, pp. 22–25]. Dicha distribución se denota mediante P_x^π, mientras que E_x^π indica el corespondiente operador de valor esperado. Ahora, sea $\mathbb{F} := \prod_{x \in S} A(x)$ la clase de todas las funciones $f : S \to A$ con la propiedad de que $f(x) \in A(x)$ para cada $x \in S$, y observe que $\mathbb{F}$ es un espacio métrico compacto con respecto a la topología producto, de manera que una sucesión $\{f_n\} \subset \mathbb{F}$ converge a $f \in \mathbb{F}$ si, y sólo si , $\lim_{n\to\infty} f_n(x) = f(x)$ para cada $x \in S$. Una política $\pi = \{\pi_n\}$ es estacionaria si existe $f \in \mathbb{F}$ tal que, para cada $n \in \mathbb{N}$ y $\mathbf{h}_n \in \mathbb{H}_n$, la medida $\pi_n(\cdot|\mathbf{h}_n)$ está concentrada en el punto $f(x_n)$, esto es, $\pi_n(\{f(x_n)\}|\mathbf{h}_n) = 1$, y en este caso la igualdad $1 = P_x^f[A_n = f(X_n)]$ es siempre válida. La clase de políticas estacionarias se identifica de manera natural con $\mathbb{F}$ y, con esta convención, $\mathbb{F} \subset \mathcal{P}$. Observe que bajo la acción de cada política $f \in \mathbb{F}$, el proceso de estados $\{X_t\}$ es una cadena de Markov con mecanismo de transición invariante determinado por la matriz estocástica $[p_{x,y}(f(x))]_{x,y \in S}$. En el resto del capítulo $\{\mathcal{F}_n\}$ denota la filtración inducida por el proceso de estados $\{X_n\}$, esto es,

$$\mathcal{F}_n = \sigma(X_k, A_k, 0 \leq k < n, X_n), \quad n \in \mathbb{N}. \tag{4.1}$$

Criterio del Costo Promedio. En el desarrollo subsecuente una condición fundamental es que el tomador de decisiones tiene un coeficiente constante de sensibilidad al riesgo, el cual se denota por λ y se supone que $\lambda > 0$, de manera que un costo aleatorio Y se valora por medio de la esperanza de $U_\lambda(Y)$, donde la función de utilidad $U_\lambda : \mathbb{R} \to \mathbb{R}$ está dada por

$$U_\lambda(x) := e^{\lambda x}, \quad \forall x \in S.$$

Así, cuando el controlador se enfrenta al problema de elegir entre dos costos aleatorios W y Y, preferirá incurrir en el costo W si $E[e^{\lambda W}] < E[e^{\lambda Y}]$, mientras que será indiferente entre cualquiera de los costos cuando $E[e^{\lambda W}] = E[e^{\lambda Y}]$. El ($\lambda$-)equivalente cierto de un costo aleatorio Y se define como el número

$$\mathcal{E}_\lambda(Y) := \log(E[e^{\lambda Y}])/\lambda,$$

de manera que $U_\lambda(\mathcal{E}_\lambda(Y)) = E[U_\lambda(Y)]$, y entonces el controlador es indiferente entre incurrir el costo aleatorio Y o pagar la cantidad constante $\mathcal{E}_\lambda(Y)$. Dada la política $\pi \in \mathcal{P}$ utilizada para seleccionar acciones y el estado inical $X_0 = x \in S$, sea $J_n(\pi, x)$ el equivalente cierto del costo total $\sum_{t=0}^{n-1} C(X_t, A_t)$ en los primeros n tiempos de decisión, esto es,

$$J_n(\pi, x) := \frac{1}{\lambda} \log\left(E_x^\pi \left[e^{\lambda \sum_{t=0}^{n-1} C(X_t, A_t)} \right] \right), \quad n \in \mathbb{N} \setminus \{0\}, \quad (4.2)$$

lo cual corresponde a una media de $J_n(\pi, x)/n$ por decisión. El límite superior de estos promedios conforme n tiende a ∞, es el costo promedio (superior λ-)sensible al riesgo bajo la política π en el estado x:

$$J(\pi, x) := \limsup_{n \to \infty} \frac{1}{n} J_n(\pi, x), \quad (4.3)$$

y la función de costo promedio (λ-)óptimo está dada por

$$J^*(x) := \inf_{\pi \in \mathcal{P}} J(\pi, x), \quad x \in S, \quad (4.4)$$

mientras que una política $\pi^* \in \mathcal{P}$ es (λ-)óptima, si $J(\pi^*, \cdot) = J^*(\cdot)$. Similarmente, el costo promedio inferior λ-sensible al riesgo bajo la política π en el estado x está dado por

$$\tilde{J}(\pi, x) := \liminf_{n \to \infty} \frac{1}{n} J_n(\pi, x), \quad (4.5)$$

mientras que la correspondiente función de valor óptimo está definida por

$$\tilde{J}^*(x) := \inf_{\pi \in \mathcal{P}} \tilde{J}(\pi, x); \quad (4.6)$$

note que

$$\tilde{J}^*(\cdot) \leq J^*(\cdot). \tag{4.7}$$

Bajo las condiciones de comunicación introducidas más adelante, se mostrará que la igualdad ocurre en esta última relación, de manera que las funciones de valor óptimo para los costos promedio superior e inferior coinciden.

Observación 4.2.1 *En la Sección 4.6 se considerarán varias funciones de costo de manera simultánea. En ese contexto, la función de costo C se incorporará a la notación (4.2)–(4.7). Por ejemplo, $J_n(C, \pi, x), J(C, \pi, x)$ y $J^*(C, \pi, x)$ se utilizarán en lugar de $J_n(\pi, x), J(\pi, x)$ y $J^*(x)$, respectivamente. Observe que (4.2) implica que la siguiente propiedad de monotonicidad es siempre válida:*

$$C \geq D \implies J_n(C, \pi, x) \geq J_n(D, \pi, x),$$

y por medio de (4.3)–(4.6) se desprende que

$$C \geq D \implies \begin{cases} J(C, \pi, \cdot) \geq J(D, \pi, \cdot), & \pi \in \mathcal{P}, \\ J^*(C, \cdot) \geq J^*(D, \cdot), \\ \tilde{J}^*(C, \cdot) \geq \tilde{J}^*(D, \cdot). \end{cases} \tag{4.8}$$

El siguiente resultado, conocido como *lema de verificación*, es un instrumento útil para analizar el criterio de costo promedio; una demostración puede encontrarse, por ejemplo, en [33].

Lema 4.2.1 *Suponga que $g \in \mathbb{R}$ y la función acotada $h : S \to \mathbb{R}$ son tales que la siguiente ecuación de optimalidad es válida:*

$$e^{\lambda g + \lambda h(x)} = \inf_{a \in A(x)} \left[e^{\lambda C(x,a)} \sum_{y \in S} p_{x\,y}(a) e^{\lambda h(y)} \right], \quad \forall x \in S. \tag{4.9}$$

En este caso, bajo las Hipótesis 4.2.1 y 4.2.2, las siguietes afirmaciones (i)–(iii) son válidas:
(i) La función de costo promedio óptimo $J^*(\cdot)$ es constante e igual a g;
(ii) $J^*(x) = \tilde{J}^*(x)$ para cada $x \in S$;
(iii) Existe una política $f \in \mathbb{F}$ tal que, para cada estado x, el término entre corchetes en (4.9) alcanza su valor mínimo en la acción $a = f(x) \in A(x)$, y tal política f es óptima respecto a los costos promedio superior e inferior; más aún,

$$J^*(x) = \lim_{n \to \infty} \frac{1}{n} J_n(f, x) = g = \tilde{J}^*(x), \quad \forall x \in S.$$

Los primeros resultados acerca de la *existencia* de soluciones de la ecuación de optimalidad pueden encontrarse en [39]. Más recientemente, suponiendo que el espacio de estados es finito, en [18] se establecieron condiciones necesarias y suficientes para que (4.9) admita una solución $(g, h(\cdot))$ para cada función de costo $C \in \mathcal{C}(\mathbb{K})$. En términos generales, las conclusiones de ese artículo establecen que la ecuación de optimalidad para el costo promedio posee una solución si, y sólo si, el espacio de estados es una clase comunicante bajo cualquier política estacionaria. Una versión de esta condición, la cual se enuncia líneas más adelante, se usará en el presente contexto de modelos con espacio numerable. Primero, es conveniente introducir la siguiente notación: Dado un conjunto $F \subset S$, el tiempo T_F de arribo a F se define como

$$T_F := \inf\{n \geq 1 : X_n \in F\} \tag{4.10}$$

donde, por convención, el ínfimo del conjunto vacío es ∞. Combinando esta igualdad con (4.1) se desprende que $[T_F = n] \in \mathcal{F}_n$ para cada $n \in \mathbb{N}$, y entonces T_F es un tiempo de paro respecto a la filtración $\{\mathcal{F}_n\}$.

Hipótesis 4.2.3 *(i) Bajo cualquier política estacionaria $f \in \mathbb{F}$, el espacio de estados es* comunicante*, esto es, dados $x, y \in S$, existen $n \equiv n(x, y) \in \mathbb{N}$ y $x_i \in S$, $i = 1, 2, \ldots, n$ tales que*

$$x_0 = x, \quad x_n = y, \quad y \quad p_{x_{i-1}, x_i}(f(x_{i-1})) > 0, \quad i = 1, 2, 3, \ldots, n.$$

(ii) [Condición Simultánea de Doeblin.] *Existe $z \in S$ tal que*

$$\sup_{x \in S, f \in \mathbb{F}} E_x^f[T_z] =: B < \infty, \tag{4.11}$$

donde

$$T_z \equiv T_{\{z\}}. \tag{4.12}$$

Observación 4.2.2 *(i) Bajo las Hipótesis 4.2.1–4.2.3, si la ecuación de optimalidad tiene una solución acotada entonces tal solución es esencialmente única. En efecto, si la ecuación (4.9) es satisfecha por los pares $(g, h(\cdot)), (\tilde{g}, \tilde{h}(\cdot)) \in \mathbb{R} \times \mathcal{C}(S)$, entonces $g = \tilde{g}$ y $h(\cdot) - \tilde{h}(\cdot)$ es constante; vea, por ejemplo, [22].*
(ii) Cuando el espacio de estados es finito, entonces no es díficil ver que la Hipótesis 4.2.1 y la propiedad de comunicación en la Hipótesis 4.2.3(i) implican la condición de Doeblin (4.11).

Bajo las Hipótesis 4.2.1–4.2.3, el costo promedio óptimo $J^*(\cdot)$ no siempre se caracteriza por medio de la ecuación de optimalidad (4.9); vea [20]. Por otro lado, en la siguiente sección se introduce una familia $\{T_\alpha\}$ de operadores contractivos en el espacio $\mathcal{C}(S)$, y los correspondientes puntos fijos $\{V_\alpha\}$ se determinan mediante ecuaciones que son similares a (4.9), de manera que es interesante ver si el costo promedio óptimo puede aproximarse por medio de la famila de puntos fijos. El resultado en esta dirección se formula en la siguiente sección.

4.3. Aproximaciones Descontadas

En esta sección se enuncia la principal contribución de este trabajo. Como punto de partida, para cada $\alpha \in]0,1[$ defina $T_\alpha : \mathcal{C}(S) \to \mathcal{C}(S)$ como sigue: Para cada $W \in \mathcal{C}(S)$ y $x \in S$,

$$T_\alpha[W](x) := \frac{1}{\lambda} \log\left(\inf_{a\in A(x)} \left[e^{\lambda C(x,a)} \sum_{y\in S} p_{x,y}(a) e^{\lambda\alpha W(y)} \right]\right). \quad (4.13)$$

A partir de esta definición es posible verificar que T_α es un operador monótono y α-homogéneo, es decir, para cada $W, V \in \mathcal{C}(S)$

(i) $T_\alpha[W] \geq T_\alpha[V]$ si $W \geq V$, y
(ii) $T_\alpha[V + c] = T_\alpha[V] + \alpha c$ para cada $c \in \mathbb{R}$.

Iniciando con $V \leq W + \|V - W\|$, estas propiedades implican que $T_\alpha[V] \leq T_\alpha[W + \|V - W\|] = T_\alpha[W] + \alpha\|V - W\|$, e intercambiando los roles de V y W se sigue que

$$\|T_\alpha[W] - T_\alpha[V]\| \leq \alpha\|W - V\|,$$

de tal suerte que T_α es un *operador contractivo* en el espacio de Banach $\mathcal{C}(S)$ con la norma del supremo. Así, existe una única función $V_\alpha \in \mathcal{C}(S)$ que satisface

$$V_\alpha = T_\alpha[V_\alpha], \quad (4.14)$$

donde el punto fijo V_α se determina mediante el método de aproximaciones sucesivas:

$$\lim_{n\to\infty} \|V_\alpha - T_\alpha^n[W]\| = 0, \quad W \in \mathcal{C}(S). \quad (4.15)$$

Usando (4.13) y (4.14) se desprende que

$$e^{\lambda V_\alpha(x)} = \inf_{a\in A(x)} \left[e^{\lambda C(x,a)} \sum_{y\in S} p_{x,y}(a) e^{\lambda\alpha V_\alpha(y)} \right], \quad x \in S \quad (4.16)$$

y, combinando las Hipótesis 4.2.1 y 4.2.2, es posible verificar que existe $f_\alpha \in \mathbb{F}$ tal que, para cada $x \in S$, $f_\alpha(x)$ minimiza el término dentro de corchetes en la ecuación anterior, de manera que

$$e^{\lambda V_\alpha(x)} = e^{\lambda C(x,f_\alpha(x))} \sum_{y \in S} p_{x,y}(f_\alpha(x)) e^{\lambda\alpha V_\alpha(y)}, \quad x \in S. \qquad (4.17)$$

A continuación, defina el costo (α-)normalizado $g_\alpha(\cdot)$ y la función de costo (α-)relativo $h_\alpha(\cdot)$ como sigue: Para cada $x \in S$,

$$g_\alpha(x) = (1-\alpha)V_\alpha(x), \quad h_\alpha(x) := \alpha[V_\alpha(x) - V_\alpha(z)], \qquad (4.18)$$

donde z es el punto fijo que se introdujo en la Hipótesis 4.2.3(ii). Con esta notación, es posible verificar que (4.16) y (4.17) son equivalentes a

$$e^{\lambda g_\alpha(x) + \lambda h_\alpha(x)} = \inf_{a \in A(x)} \left[e^{\lambda C(x,a)} \sum_{y \in S} p_{x,y}(a) e^{\lambda h_\alpha(y)} \right], \quad x \in S, \qquad (4.19)$$

y

$$e^{\lambda g_\alpha(x) + \lambda h_\alpha(x)} = e^{\lambda C(x,f_\alpha(x))} \sum_{y \in S} p_{x,y}(f_\alpha(x)) e^{\lambda h_\alpha(y)}, \quad x \in S. \qquad (4.20)$$

Observación 4.3.1 *(i) $V_\alpha(\cdot)$ es la función de valor α-descontado de un juego de suma cero con dos jugadores, el cual puede asociarse al modelo $\mathcal{M}$, y (4.16) es equivalente a la ecuación de equilibrio de ese juego; vea, por ejemplo, [33], [34], o [44].*
(ii) Como en el caso del criterio promedio, el costo descontado V_α y el correspondiente operador T_α dependen de la función de costo C. Si varias funciones de costo aparecen en la discusión simultáneamente, se escribirá $T_{\alpha,C}$ and $V_{\alpha,C}$ en lugar de T_α and V_α, respectivamente. Con esta notación, a partir de (4.13) no es difícil ver que la siguiente propiedad de monotonicidad es válida: Para cada $W \in \mathcal{C}(S)$

$$C, D \in \mathcal{C}(\mathbb{K}) \ y \ C \geq D \implies T_{\alpha,C}[W] \geq T_{\alpha,D}[W], \qquad (4.21)$$

mientras que (4.14) y (4.15) pueden escribirse como

$$V_{\alpha,C} = T_{\alpha,C}[V_{\alpha,C}] \quad y \quad \lim_{n\to\infty} \|V_{\alpha,C} - T^n_{\alpha,C}[W]\| = 0. \qquad (4.22)$$

A continuación, suponga que $D \leq C$ donde $D, C \in \mathcal{C}(\mathbb{K})$. En este caso, (4.21) y la anterior igualdad conducen a $T_{\alpha,D}[V_{\alpha,C}] \leq$

$T_{\alpha,C}[V_{\alpha,C}] = V_{\alpha,C}$. A partir de este punto, por medio de un procedimiento de inducción es posible verificar que $T^n_{\alpha,D}[V_{\alpha,C}] \leq V_{\alpha,C}$ para cada $n \in \mathbb{N}$, y entonces la convergencia en (4.22) implica que $V_{\alpha,D} \leq V_{\alpha,C}$. En resumen,

$$D, C \in \mathcal{C}(\mathbb{K}) \text{ y } D \leq C \implies V_{\alpha,D} \leq V_{\alpha,C}. \tag{4.23}$$

La igualdad (4.19) se asemeja a la ecuación de optimalidad para el criterio del costo promedio (4.9), similitud que sugiere que, si $g_\alpha(\cdot)$ es 'aproximadamente' constante, entonces $g_\alpha(y)$ está 'cerca' del costo promedio óptimo, donde $y \in S$ es un estado fijo. El siguiente teorema establece que este razonamiento intuitivo es correcto conforme α crece hacia 1.

Teorema 4.3.1 *Bajo las Hipótesis 4.2.1–4.2.3, existe una constante $g \in \mathbb{R}$ tal que las siguientes afirmaciones son válidas:*
(i) $g = \lim_{\alpha \nearrow 1} g_\alpha(x)$, $x \in S$.
(ii) $J^*(\cdot) = g = \tilde{J}^*(\cdot)$.
Más aún,
(iii) Sea $f \in \mathbb{F}$ un punto de acumulación de la familia $\{f_\alpha\} \subset \mathbb{F}$ conforme $\alpha \nearrow 1$, de tal modo que existe una sucesión $\{\alpha_n\} \subset]0,1[$ tal que

$$\alpha_n \nearrow 1 \quad \text{cuando} \quad n \to \infty, \quad \text{y} \quad \lim_{n\to\infty} f_{\alpha_n}(x) = f(x), \quad x \in S.$$

En este caso,

$$\lim_{n\to\infty} \frac{1}{n} J_n(f, x) = g, \quad \forall x \in S. \tag{4.24}$$

Note que combinando (4.18) con las partes (i) y (ii) de este resultado se desprende que

$$\lim_{\alpha \nearrow 1} (1-\alpha) V_\alpha(x) = J^*(x), \quad \forall x \in S,$$

convergencia que extiende al presente contexto sensible al riesgo, un resultdo clásico que relaciona los criterios del costo promedio y descontado neutrales al riesgo (Araphostatis et al. 1994, Puterman 1994). La demostración del Teorema 4.3.1 depende de los resultados auxiliares que se presentan en las siguientes dos secciones, cuyo principal objetivo es establecer la siguiente conclusión intermedia:

Bajo la condición de que la función de costo C tiene *soporte compacto*, si $(g, h(\cdot))$ es un punto límite conforme α crece hacia 1 de los pares $(g_\alpha(z), h_\alpha(\cdot))$, entonces $h(\cdot)$ es *acotada* y $(g, h(\cdot))$ satisface la ecuación de optimalidad (4.9).

A partir de este resultado, las conclusiones del Teorema 4.3.1 se desprenden de inmediato para el caso en que la función de costo tiene soporte compacto, y el resultado general se deducirá aproximando por abajo a una función de costo acotada por medio de funciones con soporte compacto.

4.4. Resultados Auxiliares

En esta sección se establecen resultados sobre propiedades de acotamiento de la familia $\{h_\alpha\}$ de funciones de valor relativo. Como punto de partida, sea $\alpha \in]0,1[$ y observe que combinando la ecuación (4.13) y la Hipótesis 4.2.2 se tiene que

$$T_\alpha[0](x) = \inf_{a \in A(x)} C(x,a) \geq 0 \tag{4.25}$$

y, por medio de la propiedad de monotonicidad de T_α, se desprende que $T_\alpha^n[0] \geq 0$ para cada $n \in \mathbb{N}$, relación que por medio de (4.15) implica que

$$V_\alpha \geq 0. \tag{4.26}$$

Por otro lado, la desigualdad del triángulo y (4.13)–(4.14) permiten establecer que

$$\begin{aligned}\|V_\alpha\| - \|T_\alpha[0]\| &\leq \|V_\alpha - T_\alpha[0]\| \\ &= \|T_\alpha[V_\alpha] - T_\alpha[0]\| \leq \alpha\|V_\alpha - 0\| = \alpha\|V_\alpha\|;\end{aligned}$$

puesto que (4.25) implica que $\|T_\alpha[0]\| \leq \|C\|$, se desprende que $\|V_\alpha\| - \|C\| \leq \alpha\|V_\alpha\|$, y entonces

$$(1-\alpha)\|V_\alpha\| \leq \|C\|.$$

Por medio de (4.18), esta última relación y (4.26) conducen a

$$0 \leq g_\alpha(\cdot) \leq \|C\| \quad \text{y} \quad \|h_\alpha(\cdot)\| \leq \frac{2\|C\|\alpha}{1-\alpha}, \quad \alpha \in]0,1[. \tag{4.27}$$

El principal objetivo de esta sección es establecer el siguiente resultado.

Teorema 4.4.1 *Bajo las Hipótesis 4.2.1–4.2.3, las siguientes afirmaciones son válidas para cada $x \in S$.*

(i) $\sup_{\alpha \in]0,1[} h_\alpha(x) =: B_1(x) < \infty$.

(ii) $\inf_{\alpha \in]0,1[,\, x \in S} h_\alpha(x) \geq -2\|C\|B$, *donde B es la constante en* (4.11).

La demostración de este teorema es bastante técnica y depende del resultado auxiliar establecido en el Lema 4.4.2 líneas más adelante. En el desarrollo subsecuente, $\{\alpha_n\}$ es una sucesión que satisface

$$0 < \alpha_n < 1, \quad n \in \mathbb{N}, \quad \text{y} \quad \alpha_n \nearrow 1 \quad \text{cuando} \quad n \to \infty. \tag{4.28}$$

Propiedades adicionales de la sucesión $\{\alpha_n\}$ se impondrán cuando sea necesario. Por ejemplo, para cada $\alpha \in]0,1[$ sea $f_\alpha \in \mathbb{F}$ una política estacionaria que satisface (4.20) y, recordando que el conjunto $\mathbb{F}$ de políticas estacionarias es un espacio métrico compacto, note que la sucesión $\{f_{\alpha_n}\}$ tiene un punto de acumulación en $\mathbb{F}$. Por lo tanto, después de tomar una subsucesión (en caso necesario), sin pérdida de generalidad puede suponerse que existe $f \in \mathbb{F}$ tal que

$$\lim_{n\to\infty} f_{\alpha_n}(x) = f(x). \tag{4.29}$$

Lema 4.4.2 *(i) Para cada $n \in \mathbb{N}$ y $x, y \in S$, la función*

$$\varphi \mapsto P_x^\varphi[X_n = y], \quad \varphi \in \mathbb{F}, \text{ es semi-continua inferior.} \tag{4.30}$$

(ii) Existe una función $\tilde{B}_1 : [0,\infty[\times S \to]0,\infty[$ tal que, para cada $w \in S$, $\tilde{B}_1(\cdot, w)$ es creciente, y

$$\limsup_{\alpha\nearrow 1} h_\alpha(w) \le \tilde{B}_1(\|C\|, w) < \infty.$$

Demostración. (i) El argumento es por inducción. Usando que $P_x^\varphi[X_0 = y] = 1$, si $y = x$ y $P_x^\varphi[X_0 = y] = 0$ cuando $y \neq x$, se desprende que (4.30) ocurre para $n = 0$. A continuación, suponga que la afirmación (4.30) es válida para cierto entero k, y observe que para cada $\varphi \in \mathbb{F}$ la propiedad de Markov implica que

$$P_x^\varphi[X_{k+1} = y] = \sum_{v\in S} p_{x,v}(\varphi(x)) P_v^\varphi[X_k = y].$$

Dada $f \in \mathbb{F}$, tomando el límite inferior conforme $\varphi \to f$ en ambos lados de la relación anterior, el lema de Fatou implica que

$$\begin{aligned}
\liminf_{\varphi\to f} P_x^\varphi[X_{k+1} = y] &\ge \sum_{v\in S} \liminf_{\varphi\to f} p_{x,v}(\varphi(x)) P_v^\varphi[X_k = y] \\
&\ge \sum_{v\in S} p_{x,v}(f(x)) P_v^f[X_k = y] \\
&= P_x^f[X_{k+1} = y],
\end{aligned}$$

donde la segunda desigualdad se obtuvo combinando la hipótesis de inducción y la Hipótesis 4.2.1, mientras que la igualdad se debe

a la propiedad de Markov. Esto muestra que (4.30) es válida para $n = k + 1$, completando el argumento de inducción.
(ii) Sea $w \in S$ un estado arbitrario pero fijo, y observe que existe una sucesión $\{\alpha_n\}$ que satisface (4.28) así como

$$\lim_{n\to\infty} h_{\alpha_n}(w) = \limsup_{\alpha\nearrow 1} h_\alpha(w). \tag{4.31}$$

Dado $n \in \mathbb{N}$, sea $f_{\alpha_n} \in \mathbb{F}$ la política en (4.20), de manera que

$$e^{\lambda h_{\alpha_n}(x)} = E_x^{f_{\alpha_n}}\left[e^{\lambda[C(X_0,A_0)-g_{\alpha_n}(X_0)]}e^{\lambda h_{\alpha_n}(X_1)}\right]$$

para cada $x \in S$, igualdad que, por medio de un argumento de inducción, conduce a

$$\begin{aligned} e^{\lambda h_{\alpha_n}(x)} &= E_x^{f_{\alpha_n}}\left[e^{\lambda \sum_{t=0}^{k-1}[C(X_t,A_t)-g_{\alpha_n}(X_t)]}e^{\lambda h_{\alpha_n}(X_k)}\right] \\ &\geq e^{-2\lambda k\|C\|}E_x^{f_{\alpha_n}}\left[e^{\lambda h_{\alpha_n}(X_k)}\right], \quad x \in S, \quad k = 1, 2, \ldots \end{aligned} \tag{4.32}$$

A continuación, tomando una subsucesión (si acaso es necesario), sin pérdida de generalidad se supondrá que (4.29) ocurre para alguna $f \in \mathbb{F}$. Recordando la propiedad de comunicación en la Hipótesis 4.2.3(i), seleccione un entero positivo k^* tal que $P_z^f[X_{k^*} = w] > 0$, donde el estado z es como en la Hipótesis 4.2.3(ii). Además, usando la parte (i), seleccione $N \in \mathbb{N}$ tal que

$$P_z^{f_{\alpha_n}}[X_{k^*} = w] \geq \frac{1}{2}P_z^f[X_{k^*} = w] > 0, \quad n \geq N.$$

Sustituyendo x y k en (4.32) por z y k^*, respectivamente, esta última relación implica que, para cada $n \geq N$

$$\begin{aligned} e^{\lambda h_{\alpha_n}(z)} &\geq e^{-2\lambda k^*\|C\|}E_z^{f_{\alpha_n}}\left[e^{\lambda h_{\alpha_n}(X_{k^*})}\right] \\ &\geq e^{-2\lambda k^*\|C\|}P_z^{f_{\alpha_n}}\left[X_{k^*} = w\right]e^{\lambda h_{\alpha_n}(w)} \\ &\geq \frac{1}{2}e^{-2\lambda k^*\|C\|}P_z^f\left[X_{k^*} = w\right]e^{\lambda h_{\alpha_n}(w)}; \end{aligned}$$

puesto que $h_{\alpha_n}(z) = 0$, por (4.18), se tiene que

$$\infty > 2k^*\|C\| + \frac{1}{\lambda}\log\left(2/P_z^f\left[X_{k^*} = w\right]\right) \geq h_{\alpha_n}(w), \quad n \geq N,$$

una relación que, usando (4.31), implica que $\limsup_{\alpha\nearrow 1} h_\alpha(w) \leq \tilde{B}_1(\|C\|, w) < \infty$, donde $\tilde{B}_1(r, w) := 2k^*r + \frac{1}{\lambda}\log\left(2/P_z^f\left[X_{k^*} = w\right]\right)$. Esto completa la demostración, puesto que $\tilde{B}_1(\cdot, w)$ es creciente. ∎

Demostración del Teorema 4.4.1. (i) Sea $w \in S$ un estado arbitrario pero fijo. El Lema 4.4.2(ii) garantiza que existe $\alpha(w) \in]0,1[$ tal que

$$h_\alpha(w) < 1 + \limsup_{\alpha \nearrow 1} h_\alpha(w) =: 1 + \tilde{B}_1(\|C\|, w) < \infty, \quad \alpha \in]\alpha(w), 1[,$$

y la conclusión se desprende obsevando que $h_\alpha(w) \leq \dfrac{2\|C\|\alpha(w)}{1-\alpha(w)}$ para $\alpha \in]0, \alpha(w)[$, una relación que es una consecuencia de (4.27).
(ii) Sea $\alpha \in]0,1[$ arbitrario pero fijo, y note que (4.20) es equivalente a

$$e^{\lambda h_\alpha(x)} = e^{\lambda[C(x, f_\alpha(x)) - g_\alpha(x)]} \sum_{y \in S} p_{x,y}(f_\alpha(x)) e^{\lambda h_\alpha(y)}, \quad \forall x \in S$$

así que, por la propiedad de Markov, la relación

$$\begin{aligned} e^{\lambda h_\alpha(X_n)} &= e^{\lambda[C(X_n, f_\alpha(X_n)) - g_\alpha(X_n)]} \sum_{y \in S} p_{X_n,y}(f_\alpha(X_n)) e^{\lambda h_\alpha(y)} \\ &= E_x^{f_\alpha}\left[e^{\lambda[C(X_n, A_n) - g_\alpha(X_n)] + h_\alpha(X_{n+1})} \middle| \mathcal{F}_n \right] \end{aligned} \tag{4.33}$$

siempre ocurre; vea (4.1). A continuación, para cada $n \in \mathbb{N}$ defina

$$\begin{aligned} Y_0 &= e^{\lambda h_\alpha(X_0)}, \\ Y_n &= e^{\lambda \sum_{t=0}^{n-1}[C(X_t, A_t) - g_\alpha(X_t)] + \lambda h_\alpha(X_n)}, \quad n \geq 1. \end{aligned} \tag{4.34}$$

Con esta notación, (4.33) implica que $Y_0 = E_x^{f_\alpha}[Y_1 | \mathcal{F}_0]$ cuando $n = 0$, mientras que para $n \geq 1$

$$\begin{aligned} Y_n &= E_x^{f_\alpha}[Y_{n+1} | \mathcal{F}_n] \\ &= E_x^{f_\alpha}\left[e^{\lambda \sum_{t=0}^{n}[C(X_t, A_t) - g_\alpha(X_t)] + h_\alpha(X_{n+1})} \middle| \mathcal{F}_n \right] \\ &= e^{\lambda \sum_{t=0}^{n-1}[C(X_t, A_t) - g_\alpha(X_t)]} \\ &\qquad \times E_x^{f_\alpha}\left[e^{\lambda[C(X_n, A_n) - g_\alpha(X_n)] + h_\alpha(X_{n+1})} \middle| \mathcal{F}_n \right] \\ &= e^{\lambda \sum_{t=0}^{n-1}[C(X_t, A_t) - g_\alpha(X_t)]} e^{\lambda h_\alpha(X_n)}, \end{aligned}$$

de donde se desprende que, para cada estado $x \in S$, $\{(Y_n, \mathcal{F}_n)\}$ es una martingala con respecto a $P_x^{f_\alpha}$. Recordando que T_z es un tiempo de paro con respecto a la filtración $\{\mathcal{F}_n\}$, por medio del teorema de paro opcional se desprende que, para cada $x \in S$ y $n \in \mathbb{N}$,

$$Y_0 = E_x^{f_\alpha}\left[Y_{T_z \wedge n} \right].$$

Puesto que $1 = P_x^{f_\alpha}[X_0 = x]$, combinando esta relación con (4.34) se sigue que

$$e^{\lambda h_\alpha(x)} = E_x^{f_\alpha}\left[e^{\lambda \sum_{t=0}^{T_z \wedge n-1}[C(X_t,A_t)-g_\alpha(X_t)]+\lambda h_\alpha(X_{T_z\wedge n})}\right] \quad (4.35)$$

para cada $x \in S$ y $n \in \mathbb{N}$. Para concluir, usando que $h(z) = 0$ y $X_{T_z} = z$ en el evento $[T_z < \infty]$, por (4.10) y (4.12), la igualdad $P_x^{f_\alpha}[T_z < \infty] = 1$ que es consecuencia de (4.11), implica que

$$\begin{aligned}
\lim_{n\to\infty} & \sum_{t=0}^{T_z\wedge n-1} [C(X_t, A_t) - g_\alpha(X_t)] + \lambda h_\alpha(X_{T_z\wedge n}) \\
&= \sum_{t=0}^{T_z-1} [C(X_t, A_t) - g_\alpha(X_t)] + \lambda h_\alpha(X_{T_z}) \\
&= \sum_{t=0}^{T_z-1} [C(X_t, A_t) - g_\alpha(X_t)] \geq -2T_z\|C\|, \quad P_x^{f_\alpha}\text{-c. s.} \quad \forall x \in S,
\end{aligned}$$

donde la desigualdad se debe a (4.27). Así, tomando el límite inferior conforme n tiende a ∞ en (4.35), combinando el lema de Fatou con la relación anterior se obtiene

$$\begin{aligned}
e^{\lambda h_\alpha(x)} &\geq E_x^{f_\alpha}\left[\liminf_{n\to\infty} e^{\lambda \sum_{t=0}^{T_z\wedge n-1}[C(X_t,A_t)-g_\alpha(X_t)]+\lambda h_\alpha(X_{T_z\wedge n})}\right] \\
&\geq E_x^{f_\alpha}\left[e^{-2\lambda\|C\|T_z}\right] \geq e^{-2\lambda\|C\|E_x^{f_\alpha}[T_z]}, \quad \forall x \in S,
\end{aligned}$$

donde la desigualdad de Jensen se usó en el último paso. A partir de este punto, (4.11) implica que $h_\alpha(x) \geq -2\|C\|B$ para cada $x \in S$, completando el argumento. ■

4.5. Costos con Soporte Compacto

En esta sección las conclusiones del Teorema 4.3.1 se demostrarán bajo el siguiente supuesto adicional.

Hipótesis 4.5.1 *Existe un conjunto finito y no vacío $F \subset S$ tal que*

$$x \notin F \implies C(x,a) = 0 \text{ para todo } a \in A(x). \quad (4.36)$$

Teorema 4.5.1 *Suponga que las Hipótesis 4.2.1–4.2.3 y 4.5.1 son válidas. En este contexto, existen $g \in \mathbb{R}$ y $h(\cdot) \in \mathcal{C}(S)$ tales que las siguientes afirmaciones son válidas:*

(i) Para cada $x \in S$,

$$\lim_{\alpha \nearrow 1} g_\alpha(x) = g, \quad y \quad \lim_{\alpha \nearrow 1} h_\alpha(x) = h(x). \tag{4.37}$$

(ii) El par $(g, h(\cdot))$ *satisface la ecuación de optimalidad* (4.9).
(iii) $J^*(\cdot) = g = \tilde{J}^*(\cdot)$.
(iv) Suponga que $f \in \mathbb{F}$ *es tal que*

$$\lim_{n\to\infty} f_{\alpha_n}(x) =: f(x), \quad x \in S, \tag{4.38}$$

donde $\{\alpha_n\} \subset]0,1[$ *satisface que* $\alpha_n \nearrow 1$ *conforme* $n \to \infty$. *En este contexto, para cada* $x \in S$,

$$e^{\lambda g + \lambda h(x)} = e^{\lambda C(x, f(x))} \sum_{y\in S} p_{x,y}(f(x)) e^{\lambda h(y)},$$

y entonces

$$J^*(x) = \lim_{n\to\infty} \frac{1}{n} J_n(f, x) = g = \tilde{J}^*(x), \quad \forall x \in S.$$

La demostración de este teorema depende del siguiente resultado.

Lema 4.5.2 *Bajo las Hipótesis 4.2.1–4.2.3 y 4.5.1, las siguientes afirmaciones (i) y (ii) ocurren:*
(i) Para cada $\alpha \in]0,1[$, *existe* $x^* \equiv x^*(\alpha)$ *tal que*

$$x^* \in F \cup \{z\} \quad y \quad h_\alpha(x) \le h_\alpha(x^*), \quad \forall x \in S, \tag{4.39}$$

donde $z \in S$ *y* $F \subset S$ *son como en las Hipótesis 4.2.3(ii) y 4.5.1, respectivamente.*
(ii) Existe $B_1^* \in [0, \infty)$ *tal que* $\sup_{\alpha\in]0,1[} \|h_\alpha(\cdot)\| \le B_1^*$.

Demostración. (i) Sea $G = F \cup \{z\}$ y note que (4.10) implica que $T_G \le T_z$, desigualdad que por medio de (4.11) conduce a

$$T_G < \infty \quad P_x^{f_\alpha}\text{-a. s.} \tag{4.40}$$

ahora, sea $\alpha \in]0,1[$ arbitrario pero fijo y, observando que el conjunto G es finito, sea $x^* \equiv x^*(\alpha) \in G$ un maximizador de $h_\alpha(\cdot)$ dentro del conjunto G, esto es,

$$x^* \in G \quad \text{y} \quad h_\alpha(y) \le h_\alpha(x^*), \quad y \in G.$$

Luego, para demostrar (4.39) es suficiente mostrar que

$$x \in S \setminus G \implies h_\alpha(x) \le h_\alpha(x^*). \tag{4.41}$$

Para verificar esta implicación, considere las variables Y_n en (4.34) y, usando que T_G es un tiempo de paro con respecto a la filtración $\{\mathcal{F}_n\}$, observe que un argumento similar al que se usó para deducir (4.35), permite establecer que $Y_0 = E_x^{f_\alpha}[Y_{T_G \wedge n}]$, es decir,

$$e^{\lambda h_\alpha(x)} = E_x^{f_\alpha}\left[e^{\lambda \sum_{t=0}^{T_G\wedge n-1}[C(X_t,A_t)-g_\alpha(X_t)]+\lambda h_\alpha(X_{T_G\wedge n})}\right], \quad n \in \mathbb{N}.$$

Suponga ahora que $x \in S \setminus G \subset S \setminus F$ y note los siguientes hechos (a) y (b):

(a) Usando que $P_x^{f_\alpha}[X_0 = x] = 1$, (4.36) implica que

$$1 = P_x^{f_\alpha}[C(X_0, A_0) = 0];$$

(b) Por la definición (4.10), si $1 \leq t < T_G$ entonces $X_t \in S \setminus G \subset S \setminus F$, y la Hipótesis 4.5.1 implican que $C(X_t, A_t) = 0$.

Los hechos (a) y (b) implican que $1 = P_x^{f_\alpha}\left[\sum_{t=0}^{T_G\wedge n-1} C(X_t, A_t) = 0\right]$ y, teniendo en mente que $g_\alpha(\cdot) \geq 0$, por (4.27), la anterior relación desplegada implica que

$$e^{\lambda h_\alpha(x)} \leq E_x^{f_\alpha}\left[e^{\lambda h_\alpha(X_{T_G\wedge n})}\right], \quad n \in \mathbb{N}.$$

Observe ahora que $h_\alpha(X_{T_G\wedge n}) \to h_\alpha(X_{T_G})$ cuando $n \to \infty$ en el evento $[T_G < \infty]$, y entonces, ya que $h_\alpha(\cdot)$ es acotada, por (4.27), combinando el teorema de convergencia dominada con la anterior desigualdad desplegada se obtiene que

$$\begin{aligned} e^{\lambda h_\alpha(x)} &\leq E_x^{f_\alpha}\left[e^{\lambda h_\alpha(X_{T_G})} I[T_G < \infty]\right] \\ &\leq e^{\lambda h_\alpha(x^*)} P_x^{f_\alpha}[T_G < \infty] = e^{\lambda h_\alpha(x^*)}, \end{aligned}$$

donde la segunda desigualdad se obtuvo conjuntando (4.40) con el hecho de que $X_{T_G} \in G$ en el evento $[T_G < \infty]$, y se usó (4.40) en el último paso. La anterior relación desplegada implica que $h_\alpha(x) \leq h_\alpha(x^*)$, lo cual, puesto que $x \in S \setminus G$ es arbitrario, implica (4.41).
(ii) Sea $\alpha \in]0, 1[$ arbitrario pero fijo. Ya que el conjunto F es finito, combinando la parte (i) y el Teorema 4.4.1(i) se desprende que

$$\sup_{x\in S} h_\alpha(x) \leq \max_{x\in F\cup\{z\}} h_\alpha(x) \leq \max_{x\in F\cup\{z\}} B_1(x) =: \hat{B}_1 < \infty,$$

relación que junto con la desigualdad en el Teorema 4.4.1(ii), conduce a $\|h_\alpha\| \leq B_1^* := \max\{|\hat{B}_1|, 2\|C\|B\} < \infty$, donde B es la constante en (4.11). ■

Demostración del Teorema 4.5.1. Observe que (4.27) y el Lema 4.5.2(ii) permiten concluir que, para cada $\alpha \in]0,1[$

$$(g_\alpha(z); h_\alpha(x), x \in S) \in [0, \|C\|\,] \times \prod_{x \in S} [-B_1^*, B_1^*] =: \mathcal{K}. \qquad (4.42)$$

(i) Debido a que el conjunto $\mathcal{K}$ en el lado derecho de esta inclusión es un espacio métrico compacto, la función $\alpha \mapsto (g_\alpha(z); h_\alpha(x), x \in S)$ tiene (al menos) un punto límite en $\mathcal{K}$ conforme $\alpha \nearrow 1$, y entonces, para establecer (4.37) basta con mostrar que la clase $\mathcal{L}$ de tales puntos límite consiste de un sólo punto. Con esto en mente, sea $(g, h(\cdot)) \in \mathcal{L}$ arbitrario, y seleccione una sucesión $\{\alpha_n\}$ que satisfaga (4.28) así como

$$\begin{aligned} &\lim_{n\to\infty} g_{\alpha_n}(z) = g \in [0, \|C\|], \\ &\lim_{n\to\infty} h_{\alpha_n}(x) = h(x) \in [-B_1^*, B_1^*], \quad \forall x \in S. \end{aligned} \qquad (4.43)$$

En este caso, observando que $g_{\alpha_n}(x) - g_{\alpha_n}(z) = \dfrac{1-\alpha_n}{\alpha_n} h_{\alpha_n}(x)$ y $h_{\alpha_n}(z) = 0$ (lo cual se debe a (4.18)), la anterior relación implica que

$$h(z) = 0 \quad \text{y} \quad \lim_{n\to\infty} g_{\alpha_n}(x) = g, \qquad \forall x \in S. \qquad (4.44)$$

Por otro lado, tomando una subsucesión (si es necesario), puede suponerse que (4.29) también ocurre. Luego, sustituyendo α por α_n en (4.20) y tomando el límite inferior conforme n tiende a ∞ en ambos lados de la igualdad resultante, usando el lema de Fatou conjuntamente con las Hipótesis 4.2.1 y 4.2.2, se deprende que

$$e^{\lambda g + \lambda h(x)} \ge e^{\lambda C(x, f(x))} \sum_{y \in S} p_{x,y}(f(x)) e^{\lambda h(y)}, \quad \forall x \in S.$$

Observe ahora que la desigualdad

$$e^{\lambda g_{\alpha_n}(x) + \lambda h_{\alpha_n}(x)} \le e^{\lambda C(x,a)} \sum_{y \in S} p_{x,y}(a) e^{\lambda h_{\alpha_n}(y)}$$

siempre ocurre, por (4.19), y combinando el teorema de convergencia dominada con (4.42)–(4.44), se sigue que

$$e^{\lambda g + \lambda h(x)} \le e^{\lambda C(x,a)} \sum_{y \in S} p_{x,y}(a) e^{\lambda h(y)}, \quad \forall (x,a) \in \mathbb{K}.$$

Combinando las últimas tres relaciones desplegadas se obtiene que

$$\begin{aligned} e^{\lambda g + \lambda h(x)} &= \min_{a \in A(x)} \left[e^{\lambda C(x,a)} \sum_{y \in S} p_{x,y}(a) e^{\lambda h(y)} \right] \\ &= e^{\lambda C(x, f(x))} \sum_{y \in S} p_{x,y}(f(x)) e^{\lambda h(y)}, \quad \forall x \in S. \end{aligned} \qquad (4.45)$$

Así, el par $(g, h(\cdot)) \in \mathbb{R} \times \mathcal{C}(S)$ satisface la ecuación de optimalidad para el criterio del costo promedio sensible al riesgo y, además, $h(z) = 0$; puesto que una pareja $(g, h(\cdot))$ con esas propiedades es única, por la Observación 4.2.2, se concluye que el conjunto $\mathcal{L}$ de puntos límite de $(g_\alpha(z); h_\alpha(x), x \in S)$ conforme $\alpha \nearrow 1$ consiste de un solo punto y, como ya se mencionó, esto completa la demostración de la parte (i). Ahora, observe que (4.45) implica la parte (ii), mientras que la parte (iii) se deprende del Lemma 4.2.1. Para concluir, note que el anterior argumento muestra que si (4.38) ocurre para una política estacionaria $f \in \mathbb{F}$, entonces la segunda igualdad en (4.45) ocurre, y la parte (iv) es consecuencia del Lema 4.2.1(iii). ■

4.6. Demostración del Teorema de Aproximación

La demostración del teorema principal enunciado en la Sección 4.3 se obtendrá combinando el Teorema 4.5.1 con dos lemas auxiliares que se enunciarán más adelante pero, antes de continuar, es conveniente introducir la notación que se utilizará en el argumento. Sea $\{S_n\}$ una sucesión de subconjuntos finitos de S tales que

$$\emptyset \neq S_n \subset S_{n+1}, \quad n \in \mathbb{N}, \qquad \bigcup_{n=0}^{\infty} S_n = S, \tag{4.46}$$

y defina la función de costo truncada $C_n \in \mathcal{C}(S)$ como sigue: Para cada $n \in \mathbb{N}$ y $(x, a) \in \mathbb{K}$,

$$C_n(x, a) := \begin{cases} C(x, a), & x \in S_n, \\ 0, & x \in S \setminus S_n. \end{cases} \tag{4.47}$$

A continuación, usando las Observaciones 4.2.1 y 4.3.1(ii) note que, debido al Teorema 4.5.1, el costo promedio óptimo asociado con C_n es constante, digamos $g_n \in [0, \|C_n\|\,] \subset [0, \|C\|\,]$, y satisface

$$\begin{aligned} &\lim_{\alpha \nearrow 1} (1-\alpha) V_\alpha(C_n, x) \\ &\qquad = g_n = J^*(C_n, x) = \tilde{J}^*(C_n, x), \quad \forall x \in S, \quad n \in \mathbb{N}. \end{aligned} \tag{4.48}$$

Note que las condiciones (4.46) y (4.47), conjuntamente con la Hipótesis 4.2.2, implican que $0 \leq C_n \leq C_{n+1} \leq C$, y entonces

$$\begin{aligned} 0 &\leq (1-\alpha) V_\alpha(C_n, \cdot) \\ &\leq (1-\alpha) V_\alpha(C_{n+1}, \cdot) \leq (1-\alpha) V_\alpha(C, \cdot), \quad \alpha \in]0, 1[, \end{aligned}$$

por (4.23). Tomando el límite inferior conforme α crece hacia 1 en esta última relación, (4.48) implica que

$$g_n \leq g_{n+1} \leq \liminf_{\alpha \nearrow 1} (1-\alpha) V_\alpha(C, x), \quad \forall x \in S. \tag{4.49}$$

Defina

$$g := \lim_{n\to\infty} g_n. \tag{4.50}$$

La primera fase de la demostración del Teorema 4.3.1 es la siguiente.

Lema 4.6.1 *Bajo las Hipótesis 4.2.1–4.2.3, las siguientes afirmaciones ocurren.*
(i) $J^*(C, \cdot) \leq g$.
(ii) $J^*(C, \cdot) = g = \tilde{J}^*(C, \cdot)$.

Demostración. (i) Sea n un entero no negativo arbitrario. Por el Teorema 4.5.1(iv) aplicado a la función de costo C_n, existe una política estacionaria $f_n \in \mathbb{F}$ tal que

$$e^{\lambda g_n + \lambda h_n(x)} = e^{\lambda C_n(x, f_n(x))} \sum_{y\in S} p_{x\,y}(f_n(x)) e^{\lambda h_n(y)}, \quad \forall x \in S, \tag{4.51}$$

donde $h_n(\cdot) = \lim_{\alpha\nearrow 1} h_\alpha(C_n, \cdot) = \lim_{\alpha\nearrow 1} \alpha(V_\alpha(C_n, \cdot) - V_\alpha(C_n, z))$. Por el Lema 4.4.2(ii), la función h_n satisface $-2\|C_n\|B \leq h_n(\cdot) \leq \tilde{B}_1(\|C_n\|, \cdot)$, donde $\tilde{B}_1(\cdot, w)$ es una función creciente y finita para cada $w \in S$; por lo tanto, observando que $\|C_n\| \leq \|C\|$ se tiene que

$$h_n \in \prod_{x\in S} [-2\|C\|B,\, \tilde{B}_1(\|C\|, x)] =: \mathcal{K}_1, \tag{4.52}$$

y entonces $(f_n, h_n) \in \mathbb{F} \times \mathcal{K}_1$ para cada $n \in \mathbb{N}$. Puesto que $\mathbb{F} \times \mathcal{K}_1$ es un espacio métrico compacto, existe una subsucesión $\{n_k\}$ que crece a ∞ y tal que los siguientes límites existen:

$$\lim_{k\to\infty} f_{n_k}(x) =: \tilde{f}(x), \quad \lim_{k\to\infty} h_{n_k}(x) =: \tilde{h}(x), \quad \forall x \in S. \tag{4.53}$$

Note que, debido a la monotonicidad de $\{g_n\}$ establecida en (4.49), a partir de (4.50) se desprende que

$$\lim_{k\to\infty} g_{n_k} = g.$$

Estas convergencias implican que para cada $x \in S$

$$\lim_{k\to\infty} e^{\lambda g_{n_k} + \lambda h_{n_k}(x)} = e^{\lambda g + \lambda \tilde{h}(x)}$$

mientras que, combinando el lema de Fatou, (4.46), (4.47) y las Hipótesis 4.2.1 y 4.2.2, se obtiene que

$$\begin{aligned}
\liminf_{k\to\infty} & e^{\lambda C_{n_k}(x,f_{n_k}(x))} \sum_{y\in S} p_{x\,y}(f_{n_k}(x)) e^{\lambda h_{n_k}(y)} \\
&\geq e^{\lambda C(x,\tilde f(x))} \sum_{y\in S} \liminf_{k\to\infty} p_{x\,y}(f_{n_k}(x)) e^{\lambda h_{n_k}(y)} \\
&= e^{\lambda C(x,\tilde f(x))} \sum_{y\in S} p_{x\,y}(\tilde f(x)) e^{\lambda \tilde h(y)}, \quad \forall x\in S.
\end{aligned}$$

Así, sustituyendo n por n_k en (4.51) y tomando el límite inferior conforme k tiende a ∞ en ambos lados de la ecuación resultante, las dos últimos hechos desplegados implican que la relación

$$\begin{aligned}
e^{\lambda g+\lambda\tilde h(x)} &\geq e^{\lambda C(x,\tilde f(x))} \sum_{y\in S} p_{x\,y}(\tilde f(x)) e^{\lambda\tilde h(y)} \\
&= E_x^{\tilde f}\left[e^{\lambda C(X_0,A_0)+\lambda\tilde h(X_1)}\right]
\end{aligned}$$

siempre ocurre; usando un argumento de inducción, se sigue que para cada entero positivo n y $x\in S$,

$$\begin{aligned}
e^{\lambda n g+\lambda\tilde h(x)} &\geq E_x^{\tilde f}\left[e^{\lambda\sum_{t=0}^{n-1} C(X_t,A_t)+\lambda\tilde h(X_n)}\right] \\
&\geq E_x^{\tilde f}\left[e^{\lambda\sum_{t=0}^{n-1} C(X_t,A_t)}\right] e^{-2\lambda\|C\|B} \\
&= e^{\lambda J_n(C,\tilde f,x)} e^{-2\lambda\|C\|B},
\end{aligned}$$

donde (4.52) y (4.53) se combinaron para establecer la segunda desigualdad, y la igualdad se obtuvo conjuntando (4.2) con la Observación 4.2.1. Se sigue que $J_n(C,\tilde f,x) \leq ng + \tilde h(x) + 2\|C\|B$, desigualdad que por medio de (4.3) y (4.4) implica que $J^*(C,x) \leq J(C,\tilde f,x) \leq g$.
(ii) Como se observó anteriormente, $C \geq C_n$ para cada entero positivo n y, combinando la propiedad de monotonicidad en (4.8) con (4.48), se tiene que $\tilde J^*(C,\cdot) \geq \tilde J^*(C_n,\cdot) = g_n$. Por lo tanto, (4.50) implica que $\tilde J^*(C,\cdot) \geq g$, y la conclusión se desprende combinando esta desigualdad con (4.7) y la parte (i). ■

Lema 4.6.2 *Bajo las Hipótesis 4.2.1–4.2.3, las siguientes afirmaciones son válidas para cada* $x\in S$.
(i) $\limsup_{\alpha\nearrow 1}(1-\alpha)V_\alpha(C,x) \leq J^*(C,x)$;
(ii) $\lim_{\alpha\nearrow 1}(1-\alpha)V_\alpha(C,x) = g$.

Demostración. (i) Una demostración de esta parte puede encotrarse en [23] o [34].

(ii) Observe que $g \leq \liminf_{\alpha \nearrow 1}(1-\alpha)V_\alpha(C,\cdot)$, por (4.48)–(4.50). A partir de este punto, combinando la parte (i) con el Lema 4.6.1(ii) se obtiene que $\limsup_{\alpha \nearrow 1}(1-\alpha)V_\alpha(C,\cdot) \leq g$. ■

Demostración del Teorema 4.3.1. Sea g la constante en (4.50) y note que la parte (i) se desprende de Lema 4.6.2(ii), mientras que la parte (ii) se demostró en el Lema 4.6.1(ii). Para verificar la tercera parte, para cada $\alpha \in]0,1[$ sea $f_\alpha \in \mathbb{F}$ la política estacionaria que aparece en (4.20), y suponga que

$$\lim_{n\to\infty} f_{\alpha_n}(x) = f(x), \quad \forall x \in S,$$

donde $\{\alpha_n\} \subset]0,1[$ es una sucesión que crece hacia 1. Usando que $\|C_n\| \leq \|C\|$, el Teorema 4.4.1 implica que $\{h_{\alpha_n}(x)\}$ está contenida en $[-2B\|C\|, B_1(x)]$ para cada $x \in S$, y entonces, por medio del método diagonal de Cantor, sin pérdida de generalidad puede suponerse que

$$\lim_{n\to\infty} h_{\alpha_n}(x) =: h(x) \in [-2B\|C\|,\, B_1(x)], \quad \forall x \in S. \tag{4.54}$$

Reemplazando α por α_n en (4.20) y tomando el límite inferior conforme n tiende a ∞ en ambos lados de la igualdad resultante, la parte (i) conjuntamente con las dos últimas relaciones desplegadas conducen a

$$\begin{aligned} e^{\lambda g+\lambda h(x)} &= \liminf_{n\to\infty} e^{\lambda C(x,f_{\alpha_n}(x))} \sum_{y\in S} p_{x,y}(f_{\alpha_n}(x)) e^{\lambda h_{\alpha_n}(y)} \\ &\geq e^{\lambda C(x,f(x))} \sum_{y\in S} \liminf_{n\to\infty} p_{x,y}(f_{\alpha_n}(x)) e^{\lambda h_{\alpha_n}(y)} \\ &\geq e^{\lambda C(x,f(x))} \sum_{y\in S} p_{x,y}(f(x)) e^{\lambda h(y)} \\ &= E_x^f\left[e^{\lambda C(X_0,A_0)+\lambda h(X_1)}\right], \quad \forall x \in S. \end{aligned}$$

donde el lemma de Fatou y la Hipótesis 4.2.2 se aplicaron para establecer la primera desigualdad, y la Hipótesis 4.2.1 se usó en el último paso. A partir de este punto, un argumento de inducción permite concluir que para cada $x \in S$ y $n \in \mathbb{N} \setminus \{0\}$,

$$\begin{aligned} e^{n\lambda g+\lambda h(x)} &\geq E_x^f\left[e^{\lambda \sum_{t=0}^{n-1} C(X_t,A_t)+\lambda h(X_n)}\right] \\ &\geq E_x^f\left[e^{\lambda \sum_{t=0}^{n-1} C(X_t,A_t)}\right] e^{-2\lambda\|C\|B} \\ &= e^{\lambda J_n(f,x)-2\lambda\|C\|B}, \end{aligned}$$

donde se utilizó (4.54) para establecer la segunda desigualdad, y la igualdad se debe a (4.2). Se sigue que $g + [h(x) + 2\|C\|B]/n \geq$

$J_n(f,x)/n$, así que combinando (4.5), (4.6) y la parte (ii), se tiene que para cada estado x,

$$g \geq \limsup_{n\to\infty} \frac{1}{n} J_n(f,x) \geq \liminf_{n\to\infty} \frac{1}{n} J_n(f,x) = \tilde{J}(f,x) \geq \tilde{J}^*(x) = g,$$

completando la demostración. ■

Capítulo 5

Resumen y Problemas Abiertos

5.1. Retrospectiva

En este trabajo se estudiaron procesos de decisión Markovianos con espacio de estados numerable. Las dos premisas fundamentales bajo las cuales se desarrolló el análisis fueron que

(i) El controlador es averso al riesgo y valora un costo aleatorio por medio de una función de utilidad exponencial, de modo que su coeficiente de sensibilidad al riesgo es constante y positivo, y

(ii) El espacio de estados es una clase recurrente positiva bajo el accionar de cualquier política estacionaria.

En este contexto, se estableció una versión sensible al riesgo del 'enfoque descontado', el cual consiste en aproximar el costo promedio óptimo por medio de los puntos fijos de operadores contractivos. El principal resultado en esta dirección se formuló en el Teorema 4.3.1, el cual generaliza un resultado clásico referente a la aproximación del costo promedio óptimo mediante el costo descontado mínimo en el contexto neutral al riesgo. Como resultado adicional que es interesante por sí mismo, se demostró que los criterios promedio superior e inferior tienen la misma función de valor óptimo.

Con respecto al enfoque técnico, es conveniente notar que se demostró que las aproximaciones (4.18) convergen a una solución de la ecuación de optimalidad (4.9) sólo en el caso en que la función de costo tiene soporte compacto, mientras que el caso de una función de costo general se obtuvo mediante un proceso de límite. El enfoque de este trabajo se basó en los resultados de [20], donde bajo las condiciones de comunicación/ergodicidad en las Hipótesis 4.2.3, se estableció una caracterización del costo promedio óptimo suponiendo que el coeficiente de sensibilidad al riesgo del controlador es constante y positivo.

5.2. Dos Problemas Abiertos

El resultado principal de este trabajo, enunciado en el Teorema 4.3.1, se estableció bajo la premisa de que el controlador es averso al riesgo con coeficiente de sensibilidad al riesgo constante y positiva, y las condiciones de comunicación/ergodicidad en la Hipótesis 4.2.3. Luego, es natural considerar los siguientes problemas:

• Extender las conclusiones del Teorema 4.3.1 al caso de proclividad al riesgo, en el cual el controlador tiene coeficiente de sensibilidad al riesgo constante y *negativo*.

• Establecer conclusiones similares a las establecidas en el Teorema 4.3.1 bajo versiones más débiles que las establecidas en la Hipótesis 4.2.3, por ejemplo, suponiendo solamente la condición generalizada de Doeblin

$$\sup_{f\in\mathbb{F}, x\in S} E_x^f[T_F] < \infty,$$

donde F es un subconjunto finito de S.

Bibliografía

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

[2] Arapostathis, A., Borkar, V. S., Fernández-Gaucherand, E., Ghosh, M. K., and Marcus, S. I. Discrete-time controlled Markov processes with average cost criterion: A survey. *SIAM J. Control Optim. 31*, 2 (1993), 282–334.

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

[4] Basu, A., and Ghosh, M. K. Zero-sum risk-sensitive stochastic games on a countable state space. *Stoch. Proc. Appl. 124*, 1 (2014), 961–983.

[5] Bäuerle, N., and Rieder, U. *Markov Decision Processes with Applications to Finance.* Springer, New York, 2011.

[6] Bäuerle, N., and Rieder, U. More risk-sensitive Markov decision processes. *Math. Oper. Res. 39*, 1 (2014), 105–120.

[7] Bäuerle, N., and Rieder, U. Zero-sum risk-sensitive stochastic games. *Stochastic Processes and their Applications. 127*, 2 (2017), 622–642.

[8] Berger, J. O. *Statistical decision theory and Bayesian analysis.* Springer-Verlag, New York, 1985.

[9] Bertsekas, D. P. *Dynamic Programming and Optimal Control, Vol. I.* Athena Scientific Belmont, Massachusetts, 2007.

[10] Bertsekas, D. P. *Dynamic Programming and Optimal Control, Vol. II: Approximate Dynamic Programming.* Athena Scientific Belmont, Massachusetts, 2007.

[11] Bertsekas, D. P., and Shreve, S. E. *Stochastic optimal control: the discrete time case.* Athena Scientific, 2007.

[12] Bielecki, T. R., Hernández-Hernández, D., and Pliska, S. R. Risk sensitive control of finite state markov chains in discrete time, with applications to portfolio management. 167–188.

[13] Billingsley, P. *Probability and Measure 3rd Ed.* Wiley, New York, 1995.

[14] Borkar, V. S., and Meyn, S. P. Risk-sensitive optimal control for Markov decision process with monotone cost. *Math. Oper. Res. 27*, 1 (2002), 192–209.

[15] Bouakiz, M., and Sobel, M. J. Inventory control with an exponential utility criterion. *Operations Research 40* (1992), 603–608.

[16] Caravani, P. On extending linear-quadratic control to non-symmetric risky objective. *Journal of Economic Dynamics and Control 10* (1986), 83–88.

[17] Cavazos-Cadena, R. 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* (2003), 263–285.

[18] Cavazos-Cadena, R. Solutions of the average cost optimality equation for finite Markov decision chains: risk-sensitive and risk-neutral criteria. *Math. Method Oper. Res. 70* (2009), 541–566.

[19] Cavazos-Cadena, R. A Poisson equation for the risk-sensitive average cost in semi-Markov chains. *Discrete Event Dyn. Syst. 26* (2016), 633–656.

[20] Cavazos-Cadena, R. Characterization of the optimal risk-sensitive average cost in denumerable Markov decision chains. *Math. Oper. Res. 43*, 3 (2018), 1025–1050.

[21] Cavazos-Cadena, R., and Fernández-Gaucherand, E. Controlled Markov chains with risk-sensitive criteria: Average cost, optimality equations, and optimal solutions. *Math. Method Oper. Res. 49* (1999), 299–324.

[22] Cavazos-Cadena, R., and Fernández-Gaucherand, E. Risk-sensitive control in communicating average Markov decision chains. In *Modelling Uncertainty: An examination of Stochastic Theory, Methods and Applications (Eds.: M. Dror, P.*

LÉcuyer and F. Szidarovsky). Kluwer, Boston, 2002, pp. 525–544.

[23] Cavazos-Cadena, R., and Salem-Silva, F. The discounted method and equivalence of average criteria for risk-sensitive markov decision processes on borel spaces. *Appl. Math. Optim. 61*, 2 (2010), 167–190.

[24] Denardo, E. V., and Rothblum, U. G. A turnpike theorem for a risk-sensitive Markov decision process with stopping. *SIAM J. Control Optim. 45*, 2 (2006), 414–431.

[25] Di Masi, G. B., and Stettner, L. Risk-sensitive control of discrete time Markov processes with infinite horizon. *SIAM J. Control Optim. 38*, 1 (1999), 61–78.

[26] Di Masi, G. B., and Stettner, L. Infinite horizon risk sensitive control of discrete time Markov processes with small risk. *Syst. Control Lett. 40* (2000), 15–20.

[27] Di Masi, G. B., and Stettner, L. Infinite horizon risk sensitive control of discrete time Markov processes under minorization property. *SIAM J. Control Optim. 46*, 1 (2007), 231–252.

[28] Fitzpatick, P. M. *Advanced Calculus, 2nd Ed.* Thompson, Belmont CA, 2006.

[29] Flemming, W. H., and Hernández-Hernández, D. Risk-sensitive control of finite state machines on an infinite horizon i. *SIAM Journal on Control and Optimization 33* (1997), 1881–1915.

[30] Flemming, W. H., and McEneany, W. M. Risk-sensitive control on an infinite horizon. *SIAM Journal on Control and Optimization* **33** (1995), 1881–1915.

[31] Gantmakher, F. R. *The Theory of Matrices.* Chelsea, London, 1959.

[32] Gosavi, A. A risk-sensitive approach to total productive maintenance. *Automatica 42* (2007), 1321–1330.

[33] Hernández-Hernández, D., and Marcus, S. I. Risk sensitive control of Markov processes in countable state space. *Syst. Control Lett. 29* (1996), 147–155.

[34] Hernández-Hernández, D., and Marcus, S. I. Existence of risk sensitive optimal stationary polices for controlled Markov processes. *Appl. Math. Opt. 40* (1999), 273–285.

[35] Hernández-Hernández, D., and Marcus, S. I. Existence of risk sensitive optimal stationary policies for controlled markov processes. *Applied Mathematics and Optimization 40* (1999), 273–285.

[36] Hernández-Lerma, O. *Adaptive Markov Control Processes.* Springer, New York, 1989.

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

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

[39] Howard, R. A., and Matheson, J. E. Risk-sensitive Markov decision processes. *Manage. Sci. 18*, 7 (1972), 356–369.

[40] Jacobson, D. H. Optimal stochastic linear systems with exponential performance criteria and their relation to stochastic differential games. *IEEE Transactions on Automatic Control 18* (1973), 124–131.

[41] James, M. R., Baras, J. S., and Elliot, R. J. Risk-sensitive optimal control and dynamic games for partially observed discrete-time nonlinear systems. *IEEE Transactons on Automatic Control 39* (1994), 780–792.

[42] Jaquette, S. C. Markov decison processes with a new optimality criterion: discrete time. *The Annals of Statistics 1* (1973), 496–505.

[43] Jaquette, S. C. A utility criterion for Markov decision processes. *Management Sciences 23* (1976), 43–49.

[44] Jaśkiewicz, A. Average optimality for risk sensitive control with general state space. *Ann. Appl. Probab. 17*, 2 (2007), 654–675.

[45] Jaśkiewicz, A., and Nowak, A. S. Stationary Markov perfect equilibria in risk sensitive stochastic overlapping generations models. *Journal of Economic Theory 151* (2014), 411–447.

[46] Loève, M. *Probability Thoery I, 4th Ed.* Springer-Verlag, New York, 1977.

[47] Marcus, S. I. ., Fernández-Gaucherand, E., Hernández-Hernández, D., Coralupi, S., and Frad, P. Risk-sensitive Markov decision processes, in : it Systems & Control in the Twenty-Fisrt Century, progress in systems and control (eds.: C. i. byrnes, b. n. datta, d. s. gilliam, c. f. martin). Birkhäuser. New, York, 1996, pp. 263–279.

[48] Meyer, C. *Matrix Analysis and Applied Linear Algebra.* SIAM, Philadelphia, 2000.

[49] Meyn, S., and Tweedie, R. L. *Markov Chains and Stochastic Stability.* Cambridge University Press, New York, 2009.

[50] Mihatsch, O., and Neuneier, R. Risk-sensitive reinforcement learning. *Machine Learning 49* (2002), 267–290.

[51] Nummelin, E. *General Irreducible Markov Chains and Nonnegative Operators.* Cambridge University Press, London, 1984.

[52] Pitera, M., and Stettner, L. Long run risk sensitive portfolio with general factors. *Math. Meth. Oper. Res 82*, 2 (2016), 265–293.

[53] Pratt, J. W. Risk aversion in the small and in the large. *Econometrica 32* (1964), 122–136.

[54] Puterman, M. L. *Markov Decision Processes: Discrete Stochastic Dynamic Programming.* Wiley, New York, 1994.

[55] Ross, S. M. Average cost semi-Markov processes. *J. App. Prob. 7* (1970), 649–656.

[56] Ross, S. M. *Applied Probability Models with Optimization Applications.* Dover, New York, 1992.

[57] Rudin, W. *Real and Complex Analysisi.* McGraw-Hill, New York, 1987.

[58] Runolfsson, T. The equivalence between infinite horizon control of stochastic systems with exponential-of-integral performance index and stochastic differential games. *IEEE Transactions on Automatic Control 39* (1994), 1551–1563.

[59] Saucedo-Zul, J., Cavazos-Cadena, R., and Cruz-Suárez, H. A. A discounted approach in communicating average markov decision chains under risk-aversion. *Journal of Optimization Theory and Applications 187*, 2 (2020), 585–606.

[60] Sennott, L. I. *Stochastic Dynamic Programming and the Control of Queueing Systems.* Wiley, New York, 1999.

[61] Shen, Y., Stannat, W., and Obermayer, K. Risk-sensitive Markov control processes. *SIAM J. Control Optim. 51*, 5 (2013), 3652–3672.

[62] Sladký, K. Growth rates and average optimality in risk-sensitive Markov decision chains. *Kybernetika 44*, 2 (2008), 205–226.

[63] Sladký, K. Risk-sensitive average optimality in Markov decision processes. *Kybernetika 54*, 6 (2018), 1218–1230.

[64] Stettner, L. Risk sensitive portfolio optimization. *Math. Meth. Oper. Res. 50*, 3 (1999), 463–474.

[65] Stettner, L. Risk-sensitive portfolio optimization with completely and partially observed factors. *IEEE Transactions on Automatic Control 49* (2004), 457–464.

[66] Stokey, N. L., and Lucas, R. E. *Recursive Methods in Economic Dynamics.* Harvard University Press, Cambridge, Massachusetts, 1989.

[67] Tijms, H. C. *A first course in stochastic models.* Wiley, New York, 2003.

[68] Whittle, P. *Risk-sensitive optimal control.* Wiley, New York, 1990.