Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias Físico Matemáticas

Postgrado en Ciencias Matemáticas

Procesos de Decisión Semi-Markovianos con Criterio de Costo Promedio

Tesis

Que para obtener el grado académico de:

Maestro en Ciencias

(Matemáticas)

Presenta:

Carlos Camilo Garay

Director de Tesis:

Dr. Hugo Adán Cruz Suárez

Puebla, Pue. *Febrero 2017*

Introducción

El presente trabajo de tesis se encuentra bajo el contexto de la Teoría de Control, para ser más precisos con la Teoría de los Procesos de Decisión Semi-Markovianos (PDSM, ver [19]), que son procesos estocásticos a tiempo continuo y con horizonte finito o infinito. Un Proceso de Decisión es una sucesión de controles dentro de un tiempo determinado siguiendo una estrategia y pagando un costo por cada decisión realizada.

En un problema de control óptimo se encuentra un sistema dinámico cuyo comportamiento se ve afectado por la elección de alguna de las variables del sistema, que se llaman variables de *control, acción o decisión*. Los controles que se aplican en cualquier instante de tiempo se eligen de acuerdo a "reglas" conocidas como *políticas de control*. Más aún, se da una función conocida como *criterio de rendimiento*, definida sobre el conjunto de políticas, el cual mide en algún sentido la respuesta del sistema cuando se usa la política de control. Así, el **Problema de Control Óptimo** (PCO) consiste en determinar una política de control que optimice un criterio de rendimiento.

El Modelo de Decisión Semi-Markoviano representa un sistema estocástico controlado que se observa en un intervalo de tiempo, el desarrollo se puede describir de la siguiente manera: si en el tiempo de la n-ésima época de decisión el sistema se encuentra en el estado $s_n = s$, entonces el controlador elige una acción $a_n = a$ y sucede lo siguiente: el sistema permanece en el estado s durante un tiempo aleatorio no-negativo δ_{n+1} con distribución H conocida, lo que genera un costo inmediato que depende del estado, la acción y el tiempo de permanencia, el sistema se mueve a un nuevo estado $s_{n+1} = s'$ de acuerdo a una ley de transición, una vez ocurrido lo anterior, el proceso se repite.

A la sucesión de controles que el proceso genera se le conoce como *política*; que es una regla mediante la cual se elige una acción en cada punto de observación del proceso. Para evaluar la calidad de cada política se cuenta con el criterio de rendimiento, que mide la eficiencia de ésta en función de los costos que genera, en la presente tesis se usa el criterio de rendimiento llamado *razón de costo promedio*.

Como se mencionó, el PCO consiste en encontrar una política que optimice un criterio de rendimiento, a ésta se le conoce como *política óptima*, y, al criterio de rendimiento evaluado en tal política óptima se llama *función de valores óptimos*. En los PDSM es de gran interés poder caracterizar las políticas óptimas, es decir, dar una forma explícita de la estrategia que minimice los costos.

Una manera de resolver un PDSM está basado en el principio de Bellman conocido como Programación Dinámica (ver [2]), dicho principio permite resolver problemas en los que es necesario tomar decisiones en etapas sucesivas. Las decisiones elegidas en una etapa condicionan la evolución futura del sistema, afectando a las situaciones en las que el sistema se encontrará y a las decisiones que se plantearán en el futuro.

Los PDSM fueron introducidos en [12], estos modelos han sido estudiados y aplicados, especialmente en líneas de espera controladas. Hoy en día los PDSM son útiles para el estudio de una amplia gama de problemas de optimización, en una variedad de áreas, incluyendo la robótica, optimización de telecomunicaciones, inventarios, economía y en la industria, entre otras (ver [7], [19], [20]). En muchos casos los PDSM proveen modelos más realistas que los Procesos de Decisión de Markov a tiempo discreto ya que los PDSM toman en cuenta el hecho que el tiempo transcurre continuamente.

El trabajo de tesis se centra bajo el criterio de la razón de costo promedio, éste se usa frecuentemente en la teoría de los PDSM. De acuerdo a este criterio, el costo promedio es el límite superior del costo total esperado sobre un número finito de saltos dividido por el tiempo acumulado esperado de estos saltos (ver [6] y [13]).

Este criterio ha sido estudiado por diferentes autores por métodos y conjunto de condiciones diferentes; por ejemplo, para el caso de espacio de estados numerable ver [6], [19] y [23], para el caso de espacio de estados de Borel ver, por ejemplo, [16], en éste último se considera que la función de costos podría no tener cotas superiores ni inferiores, así como el tiempo medio de permanencia se supone acotado y la ley de transición es fuertemente continua (ver Condición 2.2, más adelante). En el trabajo de tesis se consideran espacios de estados y acciones de Borel (ver Apéndice A) al igual que en [16], la diferencia es que en [16] se trabaja con un conjunto de condiciones diferentes a las presentadas en esta tesis, ver Condiciones 2.1 y 2.3, más adelante.

Cabe mencionar que la tesis está fundamentada en [25], en donde se dan condiciones necesarias para resolver el PCO bajo el método de *dos desigualdades de optimalidad promedio* y se garantiza una política de control óptima estacionaria. La aportación de este trabajo de tesis es dar dos ejemplos de líneas de espera que cumplan con las condiciones presentadas en el Capítulo 2, y garantizar la existencia de políticas óptimas además de resolver, los dos ejemplos presentados, numéricamente bajo ciertos valores en los parámetros, esto con la ayuda del software *Mathematica 8.0*, es importante hacer mención que la mayoría de la literatura existente para los PDSM no dan ejemplos en donde se resuelvan numéricamente salvo verificar condiciones y garantizar una política óptima.

La tesis está estructurada de la siguiente manera:

En el Capítulo 1, se presenta un modelo de decisión semi-Markoviano de forma general, así como el problema de decisión con el cual se trabaja.

En el Capítulo 2, se da un conjunto de condiciones necesarias para resolver el PCO bajo el criterio de razón de costo promedio, hacemos énfasis sobre este conjunto de condiciones ya que se usa la técnica de *dos desigualdades de optimalidad promedio* (ver [25]) a diferencia de otros, por ejemplo [16] que trabaja con la Ecuación de Optimalidad de Costo Promedio (EOCP) para probar la existencia de una política estacionaria óptima; se da un conjunto de condiciones aparte de las condiciones estándar (i.e., la condición de crecimiento, la condición en la ley de transición, y la condición de continuidad y compacidad) como las que se dan en [13] y [16], se reemplaza la condición para la ergodicidad geométrica en [16] por una condición donde la *diferencia relativa* de la función de valor óptimo descontado se rige por una función del tipo Lyapunov (ver Condición 2.3, más adelante). Esta condición es una generalización de la condición de ergodicidad geométrica y de las condiciones en [14]. Debido a que las condiciones presentadas en esta tesis son ligeramente "débiles" que aquellas en [13] y [16], y, puesto que la condición de irreducibilidad en [16] así como, la condición de minorización en [13] no se han incluido en este trabajo, no se puede establecer la EOCP para los PDSM. En vez de esto, se reemplaza la EOCP por dos desigualdades de costo óptimas para los PDSM, y asegurar la existencia de una solución a las dos desigualdades usando la transformación de Schweitzer. Además de estas dos desigualdades se prueba la existencia de una política estacionaria óptima, este capítulo está basado en [25].

En el Capítulo 3, se ejemplifica la Teoría de los PDSM en dos modelos de líneas de espera controladas: el primero de ellos controla el parámetro de servicio, el espacio de estados es numerable y el espacio de acciones es un espacio de Borel, bajo ciertas condiciones sobre el modelo se garantiza la existencia de una política estacionaria óptima de costo promedio; en el segundo ejemplo se considera un modelo de líneas de espera $M/M/1$ con periodos de recesos repetidos en el servidor, donde las longitudes de estos periodos son variables aleatorias independientes e idénticamente distribuidas (v.a.i.i.d.) exponencialmente. Se incurren dos costos, uno por cliente en el sistema y otro cada que el servidor se reanuda, se garantiza la existencia de una política que minimiza el criterio de rendimiento. En ambos casos, se le asignan valores específicos a los parámetros para mostrar los costos mínimos por operar en cierto estado del sistema, esto con la ayuda del software *Mathematica 8.0*, en el Apéndice B se muestran los códigos de las funciones de valor iterado para cada línea de espera.

Índice general

Capítulo 1

Preliminares

1.1. Procesos de Decisión Semi-Markovianos

En este capítulo se presenta formalmente el Modelo de Decisión Semi-Markoviano (MDSM) en el cual estamos interesados. Se trabaja de manera general con espacios de estado y acciones de Borel (i.e., un subconjunto de Borel de un espacio métrico separable y completo, ver Apéndice A), funciones de costo no acotados y tiempo medio de espera bajo el criterio de costo promedio (ver, por ejemplo, [9], [10], [16] y [25]).

Definición 1.1 *Un Modelo de Decisión Semi-Markoviano (MDSM), estacionario, a tiempo continuo, consiste de una séxtupla*

$$\Big(\mathbb{S},\ \mathbb{A},\ \{\mathbb{A}(s) : s \in \mathbb{S}\},\ Q,\ H,\ c\Big), \tag{1.1}$$

donde

- $\mathbb{S}$ *es un espacio de Borel, llamado el espacio de estados.*
- $\mathbb{A}$ *es un espacio de Borel, llamado el espacio de acciones.*
- *Para cada* $s \in \mathbb{S}$*,* $\mathbb{A}(s) \subset \mathbb{A}$ *es un conjunto medible y no vacío, cuyos elementos representan las acciones admisibles cuando el sistema se encuentre en el estado* s*.*

- *La ley de transición $Q(\cdot \mid \cdot)$ es un kérnel estocástico (ver Apéndice A) sobre $\mathbb{S}$ dado $\mathbb{K}$, en donde $\mathbb{K} := \{(s,a) \ : \ s \in \mathbb{S},\, a \in \mathbb{A}(s)\}$ es el conjunto de pares estado-acción admisible y es un subconjunto de Borel del espacio $\mathbb{S} \times \mathbb{A}$.*

- *$H(\cdot \mid s,a)$ es la función de distribución del tiempo de permanencia sobre $\mathbb{R}$, para cada $(s,a) \in \mathbb{K}$.*

- *c es una función de costo, la cual es medible sobre $\mathbb{K}$.*

Un MDSM representa un sistema dinámico que evoluciona de la siguiente manera. En el tiempo de la n-ésima época de decisión, t_n, el sistema se encuentra en el estado $s_n = s$ y el controlador elige una acción $a_n = a \in \mathbb{A}(s)$, generándose con ello lo que se describe a continuación:

- Se incurre un costo $c(s,a)$.

- El sistema permanece en dicho estado $s_n = s$ durante un tiempo aleatorio δ_{n+1} con distribución $H(\cdot \mid s,a)$.

- En el tiempo $t_{n+1} := t_n + \delta_{n+1}$ $(n \geq 0,\ t_0 := 0)$, el sistema transita a un nuevo estado $s_{n+1} = s'$ de acuerdo a la distribución $Q(\cdot \mid s,a)$.

- Finalmente, una vez en el estado s' el proceso se repite.

En adelante, se considera un modelo de decisión semi-Markoviano fijo.

Suposición 1.1 $\mathbb{K}$ contiene a la gráfica de una función medible $f : \mathbb{S} \to \mathbb{A}$, de manera que $f(s) \in \mathbb{A}(s)$, para todo $s \in \mathbb{S}$.

1.2. Políticas de Decisión

Consideremos el MDSM como en (1.1), se define el espacio de historias admisibles hasta la n-ésima época de decisión mediante

$$\mathbb{H}_0 := \mathbb{S} \quad \text{y} \quad \mathbb{H}_n := (\mathbb{K} \times \mathbb{R}_+)^n \times \mathbb{S}, \quad \text{para } n \geq 1.$$

De lo anterior se tiene que un elemento $h_n \in \mathbb{H}_n$, llamado n-historia, es un vector de la forma

$$h_n := (s_0, a_0, \delta_1, s_1, a_1, \delta_2, \ldots, s_{n-1}, a_{n-1}, \delta_n, s_n),$$

en donde $(s_k, a_k, \delta_{k+1}) \in \mathbb{K} \times \mathbb{R}_+$, para $k = 0, 1, 2, \ldots, n-1$ y $s_n \in \mathbb{S}$.

Definición 1.2 *Una política es una sucesión $\pi := \{\pi_n\}$, donde cada π_n es un kérnel estocástico sobre $\mathbb{A}$ dado $\mathbb{H}_n$, esto es, cada π_n es una probabilidad condicional $\pi_n(\cdot \mid h_n)$ sobre el conjunto de acciones $A(s_n)$ dada la historia del proceso h_n, tal que satisface $\pi_n(\mathbb{A}(s_n) \mid h_n) = 1$ para todo $h_n \in \mathbb{H}_n$ y $n \geq 0$.*

Se denota por Π al conjunto de todas las políticas.

Sea $\mathbb{F}$ el conjunto de todas las funciones Borel medibles $f : \mathbb{S} \to \mathbb{A}$ tal que $f(s) \in \mathbb{A}(s)$ para todo $s \in \mathbb{S}$. Las funciones en $\mathbb{F}$ se conocen como *selectores* del conjunto de valores que mapean $s \mapsto \mathbb{A}(s)$.

Se denotará a la familia de kérneles estocásticos sobre $\mathbb{A}$ dado $\mathbb{S}$, como $P(\mathbb{A} \mid \mathbb{S})$. Sea Φ el conjunto de todos los kérneles estocásticos φ en $P(\mathbb{A} \mid \mathbb{S})$ tales que para toda $s \in \mathbb{S}$ se tiene que $\varphi(\mathbb{A}(s) \mid s) = 1$. Así, se tiene que $\mathbb{F} \subset \Phi$.

Bajo la Suposición 1.1 se observa que el conjunto $\mathbb{F}$ es no-vacío.

De manera intuitiva, una política $\pi = \{\pi_n\}$ se puede interpretar definiendo una sucesión $\{a_n\}$ de variables aleatorias sobre $\mathbb{A}$, llamadas acciones, tal que, para cada n-historia, h_n, y $n \geq 1$, la distribución de a_n es $\pi_n(\cdot \mid h_n)$, lo cual, por la Definición 1.2, está concentrada en $\mathbb{A}(s_n)$. Esta interpretación de π se formaliza en la ecuación (1.3). A continuación se introducen varias subclases de políticas.

Observación 1.1 *Se dice que $\pi(\cdot \mid h)$ está concentrada en $g(h)$ si, $\pi(C \mid h) = I_C(g(h))$ para cada $C \in \mathfrak{B}(\mathbb{A})$ (ver Apéndice A). Donde I_C es la función indicadora del conjunto C.*

Definición 1.3 *Una política $\pi = \{\pi_n\} \in \Pi$ se dice:*

(a) **Markoviana Aleatorizada** (Π_{RM}). *Si existe una sucesión* $\{\varphi_n\}$ *de kérneles estocásticos* $\varphi_n \in \Phi$, *tales que,*

$$\pi_n(\cdot \mid h_n) = \varphi_n(\cdot \mid s_n), \tag{1.2}$$

para toda $h_n \in \mathbb{H}_n$ *y* $n \geq 0$.

(b) **Markoviana Aleatorizada Estacionaria** (Π_{RS}). *Si (1.2) se cumple para un kérnel estocástico* $\varphi \in \Phi$ *independiente de* n, *i.e.,* $\pi_n(\cdot \mid h_n) = \varphi(\cdot \mid s_n)$ *para toda* $h_n \in \mathbb{H}_n$ *y* $n \geq 0$.

(c) **Determinista** (Π_D). *Si existe una sucesión* $\{g_n\}$ *de funciones medibles con* $g_n : \mathbb{H}_n \to \mathbb{A}$, *tales que, para cada* $h_n \in \mathbb{H}_n$ *y* $n \geq 0$, *se tiene que* $g_n(h_n) \in \mathbb{A}(s_n)$ *y* $\pi_n(\cdot \mid h_n)$ *está concentrada en* $g_n(h_n)$.

(d) **Determinista Markoviana** (Π_{DM}). *Si existe una sucesión* $\{f_n\}$ *de selectores* $f_n \in \mathbb{F}$ *tal que* $\pi_n(\cdot \mid h_n)$ *es la medida de Dirac en* $f_n(s_n) \in \mathbb{A}(s_n)$ *para cada* $h_n \in \mathbb{H}_n$ *y* $n \geq 0$.

(e) **Determinista Markoviana Estacionaria** (Π_{DS}). *Si existe un selector* $f \in \mathbb{F}$ *tal que* $\pi_n(\cdot \mid h_n)$ *es la medida de Dirac en* $f_n(s_n) \in \mathbb{A}(s_n)$ *para cada* $h_n \in \mathbb{H}_n$ *y* $n \geq 0$.

Observación 1.2

- $\Pi_{RS} \subset \Pi_{RM} \subset \Pi$ *y* $\Pi_{DS} \subset \Pi_{DM} \subset \Pi_D \subset \Pi$.
- *Sin pérdida de generalidad, cualquier política estacionaria* $\pi = (f, f, \ldots)$ *se identificará con* $f \in \mathbb{F}$.

1.3. Construcción Canónica

Sea $(\Omega, \mathcal{F})$ el espacio medible que consiste del espacio muestral (canónico), Ω es el espacio producto $(\mathbb{K} \times \mathbb{R}_+)^\infty$ y la correspondiente σ-álgebra producto $\mathcal{F}$. Obsérvese que un elemento $\omega \in \Omega$ tiene la forma $\omega = (s_0, a_0, \delta_1, s_1, a_1, \delta_2, \ldots,)$.

A las variables $s_k \in \mathbb{S}$, $a_k \in \mathbb{A}(s_k)$ y $\delta_k \in \mathbb{R}_+$, se les llama variables de estado, acción y tiempo de transición, respectivamente.

De acuerdo al Teorema de C. Ionescu-Tulcea (ver Apéndice A), para cada $s \in \mathbb{S}$ y cada $\pi \in \Pi$, existe una única medida de probabilidad P_s^π sobre $(\Omega, \mathcal{F})$ y un proceso estocástico $\{s_n, a_n, \delta_{n+1}, n \geq 0\}$ tal que para cada $t \in \mathbb{R}_+$, $B \in \mathfrak{B}(\mathbb{A})$, $C \in \mathfrak{B}(\mathbb{S})$, $h_n \in \mathbb{H}_n$ y $n \geq 0$ se tiene:

$$P_s^\pi(s_0 = s) = 1,$$

$$P_s^\pi(a_n \in B \mid h_n) = \pi_n(B \mid h_n), \tag{1.3}$$

$$P_s^\pi(s_{n+1} \in C \mid h_n, a_n, \delta_{n+1}) = Q(C \mid s_n, a_n), \tag{1.4}$$

$$P_s^\pi(\delta_{n+1} \leq t \mid h_n, a_n) = H(t \mid s_n, a_n). \tag{1.5}$$

Observación 1.3 *Para una política arbitraria $\pi \in \Pi$, la variable s_n describe el estado del sistema en el tiempo de la n-ésima transición (o época de decisión) t_n y a_n representa a la acción elegida de acuerdo a la política π. Nótese que en general, el estado s_n depende de la evolución del sistema en las primeras $n-1$ transiciones; no obstante en el caso de una política estacionaria f, $\{s_n\}$ es una cadena de Markov con probabilidad de transición $Q(\cdot \mid s, f(s))$, lo cual es una consecuencia de las propiedades de esperanza condicional, así como de (1.4) que se identificará como la propiedad de Markov (ver [9]).*

De aquí en adelante se denota por E_s^π al operador esperanza con respecto a la medida de probabilidad P_s^π.

1.4. Problema de Decisión

Cada MDSM está dotado de funciones reales, llamadas función objetivo o criterio de rendimiento, cuyo fin es medir el comportamiento de los costos por etapa, como se mencionó, se usa el criterio de razón de costo promedio.

Consideremos un MDSM fijo y un conjunto de políticas Π, definamos el criterio de costo promedio.

Definición 1.4 El tiempo medio de permanencia *en el estado s cuando se elige la acción $a \in \mathbb{A}(s)$ está dado por*

$$\tau(s,a) := \int_{\mathbb{R}_+} tH(dt \mid s,a). \tag{1.6}$$

Definición 1.5 *Se define la* razón de costo promedio*, para cada $s \in \mathbb{S}$, $\pi \in \Pi$ mediante*

$$J(s,\pi) := \limsup_{n\to\infty} \frac{E_s^\pi\left[\sum_{k=0}^{n-1} c(s_k,a_k)\right]}{E_s^\pi\,(t_n)}. \tag{1.7}$$

Observación 1.4 *Usando propiedades de esperanza condicional tenemos la siguiente equivalencia, para $\pi \in \Pi$ y $s \in \mathbb{S}$,*

$$J(s,\pi) = \limsup_{n\to\infty} \frac{E_s^\pi\left[\sum_{k=0}^{n-1} c(s_k,a_k)\right]}{E_s^\pi\left[\sum_{k=0}^{n-1} \tau(s_k,a_k)\right]}. \tag{1.8}$$

La expresión (1.8) se sigue de la propiedad de Markov (1.4), ya que

$$E_s^\pi[t_n \mid h_n, a_n] = \sum_{k=0}^{n-1} \tau(s_k,a_k), \tag{1.9}$$

para todo $h_n \in \mathbb{H}_n$, $a_n \in \mathbb{A}(s_n)$, $n \geq 0$, tomando esperanza ambos lados de (1.9) tenemos

$$E_s^\pi[t_n] = E_s^\pi[E_s^\pi[t_n \mid h_n,a_n]] = E_s^\pi\left[\sum_{k=0}^{n-1} \tau(s_k,a_k)\right], \text{ para todo } n \geq 0,$$

con $\delta_{k+1} = t_{k+1} - t_k$, para el cual observamos que (1.8) se cumple.

Como se mencionó, el PCO consiste en encontrar una política que optimice el criterio de rendimiento, se da la siguiente definición.

Definición 1.6 *La función $J(s) := \inf_{\pi\in\Pi} J(s,\pi)$ se llama* función de costo promedio óptima *y una política $\pi^* \in \Pi$ se dice de* costo promedio óptima *si $J(s) = J(s,\pi^*)$, para todo $s \in \mathbb{S}$.*

En el siguiente capítulo se dan condiciones necesarias para demostrar el Teorema de Programación Dinámica para el criterio (1.7), éste nos asegura la existencia de una política óptima estacionaria.

Capítulo 2

Razón de Costo Promedio

En este capítulo se analiza el criterio (1.7), se dan condiciones necesarias que garantizan la existencia de una política estacionaria óptima y se resuelve a través de dos desigualdades de optimalidad promedio, cuando el espacio de estados es numerable se tiene la Ecuación de Optimalidad de Costo Promedio, EOCP (ver Observación 2.2 más adelante).

2.1. Condiciones Sobre el Modelo de Control

Supongamos que $\mathbb{S}$ y $\mathbb{A}$ son espacios de Borel con σ-álgebras de Borel $\mathfrak{B}(\mathbb{S})$ y $\mathfrak{B}(\mathbb{A})$, respectivamente. La siguiente condición conocida como *condición de crecimiento*, nos permiten analizar el PCO con costos no acotados y a su vez garantizar la finitud de $J(s,\pi)$.

Condición 2.1 *Condición de crecimiento.*

(a) Existen constantes positivas b, $\beta < 1$ *y una función medible* $w : \mathbb{S} \to [1, \infty)$ *tales que*

$$\int_{\mathbb{S}} w(s')Q(ds' \mid s, a) \leq \beta w(s) + b, \quad \textit{para todo } (s, a) \in \mathbb{K}.$$

(b) Existen constantes positivas L, θ *y* M *tales que*

$$|c(s, a)| \leq Lw(s) \quad y \quad \theta \leq \tau(s, a) \leq M \quad \textit{para toda } (s, a) \in \mathbb{K}.$$

Observación 2.1 *La Condición 2.1(a) se conoce como la desigualdad de Lyapunov (ver [10]).*

Lema 2.1 *Supongamos que se cumple la Condición 2.1. Entonces*

(a) $E_s^\pi[w(s_n)] \leq \beta^n w(s) + \frac{1-\beta^n}{1-\beta} b$, *para todo* $n \geq 0$, $s \in \mathbb{S}$ *y* $\pi \in \Pi$.

(b) $|J(s, \pi)| \leq \frac{bL}{(1-\beta)\theta}$, *para todo* $s \in \mathbb{S}$ *y* $\pi \in \Pi$.

Demostración. *(a)* Por inducción. Para $n = 0$ es trivial. Ahora supóngase que se cumple *(a)* para algún $n \geq 1$. Entonces por (1.4) y por la Condición 2.1*(a)* tenemos

$$\begin{aligned} E_s^\pi[w(s_{n+1})|h_n, a_n, \delta_{n+1}] &= \int_{\mathbb{S}} w(s')P_s^\pi(ds'|h_n, a_n, \delta_{n+1}) \\ &= \int_{\mathbb{S}} w(s')Q(ds'|s_n, a_n) \\ &\leq \beta w(s_n) + b, \end{aligned}$$

que junto con la hipótesis de inducción, implica que

$$\begin{aligned} E_s^\pi[w(s_{n+1})] \leq E_s^\pi[\beta w(s_n) + b] &= \beta E_s^\pi[w(s_n)] + b \\ &= \beta \left[\beta^n w(s) + \frac{1-\beta^n}{1-\beta} b \right] + b \\ &= \beta^{n+1} w(s) + \frac{1-\beta^{n+1}}{1-\beta} b, \end{aligned}$$

y así, la parte *(a)* se cumple.

(b) Se sigue de la Condición 2.1*(b)* y la parte *(a)* que

$$
\begin{aligned}
|J(s,\pi)| &= \limsup_{n\to\infty} \frac{E_s^\pi\left[\sum_{k=0}^{n-1} c(s_k,a_k)\right]}{E_s^\pi\left[\sum_{k=0}^{n-1} \tau(s_k,a_k)\right]} \\
&\leq \limsup_{n\to\infty} \frac{L\sum_{k=0}^{n-1} E_s^\pi\left[w(s_k)\right]}{n\theta} \\
&\leq \limsup_{n\to\infty} \frac{L}{n\theta}\sum_{k=0}^{n-1}\left[\beta^k w(s) + \frac{1-\beta^k}{1-\beta}b\right] \\
&\leq \limsup_{n\to\infty} \frac{L}{n\theta}\left[\frac{1-\beta^n}{1-\beta}w(s) + \frac{bn}{1-\beta}\right] = \frac{bL}{(1-\beta)\theta},
\end{aligned}
$$

para todo $s \in \mathbb{S}$ y $\pi \in \Pi$, por tanto, se concluye la prueba. □

Para asegurar la existencia de una política óptima, se necesita la siguiente condición estándar de continuidad y compacidad, la cual es ampliamente usada (ver [9], [10], [16], [19] y [23]).

Condición 2.2 *Condición de continuidad y compacidad.*

(a) $\mathbb{A}(s)$ *es un conjunto compacto, para cada* $s \in \mathbb{S}$.

(b) *Para cada* $s \in \mathbb{S}$, *la ley de transición* $Q(\cdot \mid s,a)$ *es fuertemente continua en* $a \in \mathbb{A}(s)$, *i.e., para cada función medible y acotada* u *sobre* $\mathbb{S}$, *se tiene que*

$$\int_{\mathbb{S}} u(s')Q(ds' \mid s,a)$$

es una función continua en $a \in \mathbb{A}(s)$.

(c) *Para cada* $s \in \mathbb{S}$, *la función* $c(s,a)$ *es semi-continua inferiormente (l.s.c por sus siglas en inglés, ver Apéndice A) en* $a \in A(s)$, *y las funciones* $\tau(s,a)$ *y* $\int_{\mathbb{S}} w(s')Q(ds' \mid s,a)$ *son continuas en* $a \in \mathbb{A}(s)$, *donde* w *es la función definida en la Condición 2.1.*

2.2. Transformación de Schweitzer

La trasformación de Schweitzer se ha utilizado ampliamente para estudiar la existencia de políticas óptimas en los PDSM bajo el criterio de costo promedio, para el caso de espacios de Borel, ver [16] y para el caso de espacios de estados numerable, ver [5] y [22].

Esta transformación se utilizará para pasar del modelo estudiado a tiempo continuo a un modelo a tiempo discreto, con el cual se podrá resolver el PCO con el criterio de costo descontado y así asegurar la existencia de una política óptima estacionaria. Se introduce dicha transformación.

Sea $\rho \in (0, \theta)$, con θ como en la Condición 2.1, definamos a la función $\hat{c} : \mathbb{K} \to \mathbb{R}$ y el kérnel estocástico $\hat{Q}$ sobre $\mathbb{S}$ dado $\mathbb{K}$ como

$$\hat{c}(s,a) := \frac{c(s,a)}{\tau(s,a)}, \; \hat{Q}(\cdot|s,a) := \frac{\rho}{\tau(s,a)} Q(\cdot|s,a) + \left(1 - \frac{\rho}{\tau(s,a)}\right) \delta_s(\cdot), \tag{2.1}$$

para todo $(s,a) \in \mathbb{K}$, donde $\delta_s(\cdot)$ es la medida de Dirac concentrada en el estado s (ver Apéndice A).

Así, se obtiene un modelo de decisión de Markov a tiempo discreto de la siguiente manera:

$$\Big(\mathbb{S}, \mathbb{A}, \{\mathbb{A}(s) : s \in \mathbb{S}\}, \hat{Q}(\cdot|s,a), \hat{c}(s,a)\Big).$$

Por lo tanto, para cada $\pi \in \Pi$ y $s \in \mathbb{S}$, nuevamente por el Teorema de C. Ionescu-Tulcea, existe una única medida de probabilidad $\hat{P}_s^\pi$ sobre $((\mathbb{S} \times \mathbb{A})^\infty, \mathfrak{B}((\mathbb{S} \times \mathbb{A})^\infty))$ asociado con el PDM a tiempo discreto. Análogamente, denotamos por $\hat{E}_s^\pi$ al operador esperanza con respecto a la medida $\hat{P}_s^\pi$.

Para dar la última condición, necesitamos algunos resultados relacionados con el criterio de costo descontado (ver, por ejemplo, [3] y [19]) para el PDM.

Para cada factor de descuento $\alpha \in (0,1)$, definimos el costo descontado esperado $V_\alpha(\cdot|\cdot)$ y su correspondiente función de valor óptimo para el PDM de la siguiente manera

$$V_\alpha(s,\pi) := \hat{E}_s^\pi \left[\sum_{n=0}^{\infty} \alpha^n \hat{c}(s_n, a_n) \right], \quad V_\alpha^*(s) := \inf_{\pi \in \Pi} V_\alpha(s,\pi),$$

para $s \in \mathbb{S}$ y $\pi \in \Pi$.

Una política $\pi^* \in \Pi$ se dice óptima (α-descontada) si y sólo si $V_\alpha(s,\pi^*) = V_\alpha^*(s)$, para todo $s \in \mathbb{S}$ (ver [3]).

Tenemos el siguiente resultado.

Lema 2.2 *Supongamos que se cumplen las Condiciones 2.1 y 2.2, entonces*

(a) $|V_\alpha(s,\pi)| \leq \frac{L}{(1-\alpha)\theta} w(s) + \frac{bLM}{(1-\alpha)(1-\beta)\theta^2}$*, para todo* $\alpha \in (0,1),\ s \in \mathbb{S}$ *y* $\pi \in \Pi$.

(b) Para cada $\alpha \in (0,1)$*, existe una política óptima* $f_\alpha \in \mathbb{F}$ *(que depende de* α*) tal que* $V_\alpha(s, f_\alpha) = V_\alpha^*(s)$*, para todo* $s \in \mathbb{S}$.

Demostración. *(a)* De (2.1) y de la Condición 2.1*(a)* se tiene que

$$\begin{aligned}
\int_{\mathbb{S}} w(s')\hat{Q}(ds'|s,a) &= \int_{\mathbb{S}} w(s') \left[\frac{\rho}{\tau(s,a)} Q(ds'|s,a) + \left(1 - \frac{\rho}{\tau(s,a)}\right) \delta_s(ds') \right] \\
&= \frac{\rho}{\tau(s,a)} \int_{\mathbb{S}} w(s') Q(ds'|s,a) + \left(1 - \frac{\rho}{\tau(s,a)}\right) w(s) \qquad (2.2) \\
&\leq \frac{\rho}{\tau(s,a)} (\beta w(s) + b) + \left(1 - \frac{\rho}{\tau(s,a)}\right) w(s) \\
&= \left[\frac{(\beta - 1)\rho}{\tau(s,a)} + 1 \right] w(s) + \frac{b\rho}{\tau(s,a)},
\end{aligned}$$

que junto con la Condición 2.1*(b)* se llega

$$\int_{\mathbb{S}} w(s')\hat{Q}(ds'|s,a) \leq \left(1 - \frac{(1-\beta)\rho}{M}\right) w(s) + \frac{\rho b}{\theta},$$

para todo $(s,a) \in \mathbb{K}$. Ahora, por el Lema 2.1*(a)* tenemos

$$\hat{E}_s^\pi[w(s_n)] \leq \left(1 - \frac{(1-\beta)\rho}{M}\right)^n w(s) + \frac{Mb}{1 - \beta\theta},$$

que junto con la Condición 2.1*(b)* se obtiene

$$|V_\alpha(s,\pi)| \leq \frac{L}{\theta}\sum_{n=0}^{\infty}\alpha^n \hat{E}_s^\pi[w(s_n)] \leq \frac{L}{(1-\alpha)\theta}w(s) + \frac{bLM}{(1-\beta)(1-\alpha)\theta^2},$$

luego, se cumple la parte *(a)*.

(b) Para cada $s \in \mathbb{S}$ y cada función medible y acotada u sobre $\mathbb{S}$, por (2.1), (2.2) y la Condición 2.2, $\hat{c}(s,a)$ es inferiormente semi-continua en $a \in \mathbb{A}(s)$, y $\int_{\mathbb{S}} u(s')\hat{Q}(ds'|s,a)$ y $\int_{\mathbb{S}} w(s')\hat{Q}(ds'|s,a)$ son continuas en $a \in \mathbb{A}(s)$. Por tanto, la parte *(b)* se sigue del Teorema 8.3.6 en [10] y de [3] □

Las Condiciones 2.1 y 2.2 garantizan la existencia de una política óptima estacionaria α-descontada, sin embargo, para asegurar la existencia de una política óptima para el criterio de costo promedio, se necesita dar otra condición aparte de las Condiciones 2.1 y 2.2, para esto se introduce el concepto de norma ponderada.

Sea $w : \mathbb{S} \to [1,\infty)$ una función medible, denotamos por $\mathbb{B}_w(\mathbb{S})$ el espacio lineal normado el cual consiste de todas las funciones medibles $u : \mathbb{S} \to \mathbb{R}$ con norma finita, definida por

$$\| u \|_w := \sup_{s\in\mathbb{S}} \frac{| u(s) |}{w(s)},$$

esto es, $\mathbb{B}_w(\mathbb{S}) := \{u | \ \|u\|_w < \infty\}$.

Condición 2.3 *Existen dos funciones $v_1^*,\ v_2^* \in \mathbb{B}_w(\mathbb{S})$ y un estado $\hat{s} \in \mathbb{S}$ tales que*

$$v_1^*(s) \leq h_\alpha(s) \leq v_2^*(s), \quad \textit{para todo } \ s \in \mathbb{S} \ \ y \ \ \alpha \in (0,1),$$

en donde $h_\alpha(s) := V_\alpha^*(s) - V_\alpha^*(\hat{s})$, a $h_\alpha(s)$ se le conoce como la diferencia relativa *de la función de costo descontada óptima $V_\alpha^*(s)$.*

La Condición 2.3 es similar a aquellas en [9], [19] y [23] para el PDM a tiempo discreto. Cabe mencionar que la función v_1^* en la Condición

2.3 puede no ser acotada inferiormente, y así $h_\alpha(\cdot)$ también puede no ser acotada inferiormente. Sin embargo, la correspondiente $h_\alpha(\cdot)$ en [9], [19] y [23] sí se supone acotada inferiormente.

Para verificar la Condición 2.3 damos algunas condiciones suficientes. Para esto introducimos la siguiente notación.

Para cada $f \in \mathbb{F}$ y el estado $\hat{s}$ (con $\hat{s}$ como en la Condición 2.3), consideremos la variable aleatoria

$$\sigma_{\hat{s}} := \text{ínf}\{n > 0 | s_n = \hat{s}\} \text{ con } \text{ínf}\,\emptyset := \infty,$$

que es el *tiempo de la primer entrada al estado* $\hat{s}$ del proceso $\{s_n,\ n \geq 0\}$ definido por $\hat{Q}(\cdot|\cdot, f(\cdot))$.

Más aún, para cualquier función medible no-negativa $l(s,a)$ sobre $\mathbb{K}$, denotamos por

$$U_f^{\hat{s}}(s,l) := \hat{E}_s^f\left[\sum_{k=0}^{\sigma_{\hat{s}}-1} l(s_k, f(s_k))\right], \text{ para toda } s \neq \hat{s} \text{ y } f \in \mathbb{F},$$

al *costo total esperado del primer paso desde cualquier estado inicial* s al estado $\hat{s}$ bajo la cadena de Markov definida por $\hat{Q}(\cdot|\cdot, f(\cdot))$.

Sea $\mathbb{M}$ el conjunto de las funciones medibles no-negativas $u : \mathbb{S}\backslash\{\hat{s}\} \to \mathbb{R}_+ \cup \{\infty\}$. El siguiente resultado se utiliza con el objetivo de proporcionar una condición suficiente para la comprobación de la Condición 2.3.

Lema 2.3 *Supongamos que* $\hat{P}_s^f(\sigma_{\hat{s}} < \infty) = 1$ *para cada* $f \in \mathbb{F}$ *y* $s \neq \hat{s}$. *Entonces, tenemos*

(a) $U_f^{\hat{s}}(s,l)$ *es la solución no-negativa mínima en* $\mathbb{M}$ *a la siguiente desigualdad:*

$$l(s, f(s)) + \int_{\mathbb{S}\backslash\{\hat{s}\}} u(s')\hat{Q}(ds'|s, f(s)) \leq u(s), \text{ para todo } s \neq \hat{s}. \tag{2.3}$$

(b) Si existen funciones medibles no-negativas $\overline{u}$, $\overline{v} \in \mathbb{B}_w(\mathbb{S})$, *y una constante* $\overline{L} > 0$ *tales que*

$$|\hat{c}(s, f(s))| \leq \overline{L}\overline{v}(s) \ \text{, y}$$

$$\overline{v}(s) + \int_{\mathbb{S}\setminus\{\hat{s}\}} \overline{u}(s')\hat{Q}(ds'|s, f(s)) \leq \overline{u}(s), \quad \textit{para todo} \;\; s \neq \hat{s},$$

entonces, $\hat{E}_s^f \left[\sum_{k=0}^{\sigma_{\hat{s}}-1} |\hat{c}(s_k, f(s_k))|\right] \leq \overline{L}\overline{u}(s)$, *para toda* $s \neq \hat{s}$.

Demostración. *(a)* Para cada $s \neq \hat{s}$ y $f \in \mathbb{F}$, como $l(s, f(s)) \geq 0$ para toda $s \in \mathbb{S}$, tenemos

$$\begin{aligned} U_f^{\hat{s}}(s, l) &= \hat{E}_s^f \left[\sum_{k=0}^{\sigma_{\hat{s}}-1} l(s_k, f(s_k))\right] \\ &= \hat{E}_s^f \left[\sum_{k=0}^{\infty} I_{\{\sigma_{\hat{s}}>k\}} l(s_k, f(s_k))\right] \\ &= \lim_{n\to\infty} \sum_{k=0}^{n} \int_{\mathbb{S}\setminus\{\hat{s}\}} l(s', f(s'))\hat{P}_s^f(s_m \neq \hat{s}, 0 \leq m \leq k-1, s_k \in ds'), \end{aligned}$$

donde I_B es la función indicadora del conjunto B, la última igualdad se obtuvo aplicando la definición de esperanza.

Para cada $n \geq 0$, definamos

$$U_n(s, f) := \sum_{k=0}^{n} \int_{\mathbb{S}\setminus\{\hat{s}\}} l(s', f(s'))\hat{P}_s^f(s_m \neq \hat{s}, 0 \leq m \leq k-1, s_k \in ds'). \tag{2.4}$$

Entonces, tenemos que

$$U_n(s, f) \leq U_{n+1}(s, f) \quad \text{y} \quad \lim_{n\to\infty} U_n(s, f) = U_n^{\hat{s}}(s, l). \tag{2.5}$$

Adicionalmente, definamos un operador T sobre $\mathbb{M}$ de la siguiente manera:

$$Tu(s) := l(s, f(s)) + \int_{\mathbb{S}\setminus\{\hat{s}\}} u(s')\hat{Q}(ds'|s, f(s)),$$

para toda $u \in \mathbb{M}$. Entonces, tenemos que

$$U_{n+1}(s, f) = TU_n(s, f), \;\; \text{para toda} \;\; n \geq -1, \;\; \text{en donde} \;\; U_{-1}(s, f) := 0. \tag{2.6}$$

En efecto, por (2.4) y cálculos directos tenemos que

$$
\begin{aligned}
&U_{n+1}(s,f) \\
&\quad = \sum_{k=0}^{n+1} \int_{\mathbb{S}\setminus\{\hat{s}\}} l(s',f(s'))\hat{P}_s^f(s_m \neq \hat{s}, 0 \leq m \leq k-1, s_k \in ds') \\
&\quad = l(s,f(s)) + \sum_{k=1}^{n+1} \int_{\mathbb{S}\setminus\{\hat{s}\}} l(s',f(s'))\hat{P}_s^f(s_m \neq \hat{s}, 0 \leq m \leq k-1, s_k \in ds') \\
&\quad = l(s,f(s)) + \sum_{k=1}^{n+1} \int_{\mathbb{S}\setminus\{\hat{s}\}} l(s',f(s'))\hat{E}_s^f[I_{\{s_m \neq \hat{s}, 0 \leq m \leq k-1, s_k \in ds'\}}] \\
&\quad = l(s,f(s)) + \sum_{k=1}^{n+1} \int_{\mathbb{S}\setminus\{\hat{s}\}} l(s',f(s'))\hat{E}_s^f[\hat{E}_s^f[I_{\{s_m \neq \hat{s}, 0 \leq m \leq k-1, s_k \in ds'\}}|h_1]] \\
&\quad = l(s,f(s)) + \sum_{k=1}^{n+1} \int_{\mathbb{S}\setminus\{\hat{s}\}} \Big[\int_{\mathbb{S}\setminus\{\hat{s}\}} l(s',f(s')) \\
&\qquad \times \hat{P}_{s''}^f(s_m \neq \hat{s}, 0 \leq m \leq k-2, s_{k-1} \in ds' \Big] \hat{Q}(ds''|s,f(s)) \\
&\quad = l(s,f(s)) + \int_{\mathbb{S}\setminus\{\hat{s}\}} \sum_{k=0}^{n} \Big[\int_{\mathbb{S}\setminus\{\hat{s}\}} l(s',f(s')) \\
&\qquad \times \hat{P}_{s''}^f(s_m \neq \hat{s}, 0 \leq m \leq k-1, s_k \in ds') \Big] \hat{Q}(ds''|s,f(s)) \\
&\quad = l(s,f(s)) + \int_{\mathbb{S}\setminus\{\hat{s}\}} U_n(s'',f)\hat{Q}(ds''|s,f(s)) \\
&\quad = TU_n(s,f).
\end{aligned}
$$

Por tanto, haciendo $n \to \infty$ en (2.6), por (2.5) y el Teorema de la Convergencia Monótona (ver Apéndice A) tenemos lo siguiente

$$
U_f^{\hat{s}}(s,l) = l(s,f(s)) + \int_{\mathbb{S}\setminus\{\hat{s}\}} U_f^{\hat{s}}(s',l)\hat{Q}(ds'|s,f(s)), \quad \text{para toda } \; s \neq \hat{s},
$$

lo cual nos da (2.3). Sea $u \in \mathbb{M}$ una solución arbitraria de (2.3). Entonces

$$
u(s) \geq U_n(s,f), \quad \text{para todo } \; s \neq \hat{s} \; \text{ y } \; n \geq -1. \tag{2.7}
$$

En efecto, (2.7) es trivial para $n = -1$. Supongamos que (2.7) se cumple para $n \geq 0$, luego, por (2.6) obtenemos que

$$
u(s) \geq Tu(s) \geq TU_n(s,f(s)) = U_{n+1}(s,f), \tag{2.8}
$$

así, (2.7) se cumple. Haciendo $n \to \infty$ en (2.8), tenemos $u(s) \geq U_f^{\hat{s}}(s,l)$ para todo $s \neq \hat{s}$, luego la parte *(a)* se cumple.

(b) Se sigue de la condición inicial y la parte *(a)* que

$$\hat{E}_s^f \left[\sum_{k=0}^{\sigma_{\hat{s}}-1} |\hat{c}(s_k, f(s_k))| \right] \leq \overline{L} \hat{E}_s^f \left[\sum_{k=0}^{\sigma_{\hat{s}}-1} \overline{v}(s_k) \right], \quad \text{para todo} \quad s \neq \hat{s},$$

lo cual concluye la prueba. □

En el siguiente lema se dan condiciones equivalentes a la Condición 2.3.

Lema 2.4 *Bajo las Condiciones 2.1 y 2.2, cada una de las siguientes condiciones implican la Condición 2.3.*

(a) *(La condición de ergodicidad w-geométrica uniforme.) Para cada $f \in \mathbb{F}$, existe una medida de probabilidad μ_f sobre $\mathbb{S}$ tal que, para cada $u \in \mathbb{B}_w(\mathbb{S})$, $s \in \mathbb{S}$, $n = 1, 2, \ldots$,*

$$\left| \int_{\mathbb{S}} u(s') \hat{Q}^n(ds'|s, f(s)) - \int_{\mathbb{S}} u(s') \mu_f(ds') \right| \leq \|u\|_w R \eta^n w(s),$$

donde las constantes positivas R y $\eta < 1$ son independientes de f y $\hat{Q}^n(B|s, f(s)) := \hat{P}_s^f(s_n \in B)$, para $B \in \mathfrak{B}(\mathbb{S})$.

(b) *(b1) Existe un conjunto de Borel $C \subset \mathbb{S}$ y una constante $\lambda_1 \in (0,1)$ y $\kappa > 0$, tales que*

$$\int_{\mathbb{S}} w(s') Q(ds'|s, a) \leq \lambda_1 w(s) + \kappa I_C(s),$$

para todo $(s,a) \in \mathbb{K}$ y $\sup_{s \in C} w(s) < \infty$, donde $I_C(\cdot)$ es la función indicadora del conjunto C.

(b2) Existe una constante positiva $\varepsilon \in (0,1)$ y una medida de probabilidad μ concentrada sobre el conjunto de Borel C tal que $Q(D|s,a) \geq \varepsilon \mu(D)$ para cada conjunto de Borel $D \subset C$, $x \in C$ y $a \in \mathbb{A}(s)$.

(c) *(c1) Existe una medida de probabilidad ν sobre $\mathbb{S}$ y una constante $\lambda_2 \in (0,1)$, que cumple: Para cada $f \in \mathbb{F}$ existe una función*

medible no-negativa $0 \leq l_f \leq 1$ *sobre* $\mathbb{S}$*, tal que para cada* $s \in \mathbb{S}$ *y* $B \in \mathfrak{B}(\mathbb{S})$*, se tiene*

(i) $Q(B|s, f(s)) \geq l_f(s)\nu(B)$.

(ii) $\nu(w) := \int_{\mathbb{S}} w(s')\nu(ds') < \infty$ *y* $\inf_F \int_{\mathbb{S}} l_f(s')\nu(ds') > 0$.

(iii) $\int_{\mathbb{S}} w(s')Q(ds'|s, f(s)) \leq \lambda_2 w(s) + l_f(s)\nu(w)$.

(c2) Existe una medida σ*-finita* ψ *sobre* S *para el cual, para cada* $f \in \mathbb{F}$ *la cadena de Markov definida por* $Q(\cdot|\cdot, f(\cdot))$ *es* ψ*-irreducible, esto es, si* $B \in \mathfrak{B}(\mathbb{S})$ *es tal que* $\psi(B) > 0$*, entonces para cada* $s \in \mathbb{S}$ *existe* $n > 0$ *para la cual* $Q^n(B|s, f(s)) > 0$*, donde* $Q^n(B|s, f(s)) := P_s^f(s_n \in B)$.

(d) (d1) Existe una constante $\lambda_3 \in (0, 1)$ *tal que*

$$Q(\{\hat{s}\}|s, f(s)) \geq \lambda_3, \quad \textit{para todo} \;\; s \neq \hat{s} \; \textit{y} \; f \in \mathbb{F}.$$

(d2) Existen funciones medibles $u_1, v_1 \in \mathbb{B}_w(\mathbb{S})$*, y una constante* $L_1 > 0$*, tal que*

$$|c(s, f(s))| \leq L_1 v_1(s), \quad y$$

$$v_1(s) + \int_{\mathbb{S}\setminus\{\hat{s}\}} u_1(s')Q(ds'|s, f(s)) \leq u_1(s), \quad \textit{para todo} \;\; s \neq \hat{s}.$$

Demostración. *(a)* Se sigue del Lema 2.2 que para cada $\alpha \in (0, 1)$, existe una política estacionaria óptima α-descontada $f_\alpha \in \mathbb{F}$ tal que $V_\alpha(s, f_\alpha) = V_\alpha^*(s)$ para toda $s \in \mathbb{S}$. Así, tenemos

$$\begin{aligned}
|h_\alpha(s)| &= |V_\alpha^*(s) - V_\alpha^*(\hat{s})| \\
&= \left| \hat{E}_x^{f_\alpha} \left[\sum_{n=0}^{\infty} \alpha^n \hat{c}(s_n, f_\alpha(s_n)) \right] - \hat{E}_{\hat{x}}^{f_\alpha} \left[\sum_{n=0}^{\infty} \alpha^n \hat{c}(s_n, f_\alpha(s_n)) \right] \right| \\
&\leq \sum_{n=0}^{\infty} \alpha^n \left| \hat{E}_x^{f_\alpha} \left[\hat{c}(s_n, f_\alpha(s_n)) \right] - \hat{E}_{\hat{x}}^{f_\alpha} \left[\hat{c}(s_n, f_\alpha(s_n)) \right] \right|,
\end{aligned}$$

que junto con la condición en la parte *(a)* y la Condición 2.1*(b)*, nos lleva

$$\begin{aligned}
|h_\alpha(s)| &\leq \sum_{n=0}^{\infty} \alpha^n \left| \hat{E}_x^{f_\alpha}\left[\frac{c(s_n, f_\alpha(s_n))}{\tau(s_n, f_\alpha(s_n))}\right] - \hat{E}_{\hat{x}}^{f_\alpha}\left[\frac{c(s_n, f_\alpha(s_n))}{\tau(s_n, f_\alpha(s_n))}\right] \right| \\
&\leq \frac{L}{\theta}\sum_{n=0}^{\infty} \alpha^n \left| \hat{E}_x^{f_\alpha}[w(s_n)] - \hat{E}_{\hat{x}}^{f_\alpha}[w(s_n)] \right| \\
&= \frac{L}{\theta}\sum_{n=0}^{\infty} \alpha^n \left| \int_{\mathbb{S}} w(s')\hat{Q}^n(ds'|s, f_\alpha(s)) - \int_{\mathbb{S}} w(s')\hat{Q}^n(ds'|\hat{s}, f_\alpha(\hat{s})) \right| \\
&= \frac{L}{\theta}\sum_{n=0}^{\infty} \alpha^n \left| \int_{\mathbb{S}} w(s')\hat{Q}^n(ds'|s, f_\alpha(s)) + \int_{\mathbb{S}} w(s')\mu_f(ds') \right. \\
&\qquad\qquad \left. - \int_{\mathbb{S}} w(s')\mu_f(ds') - \int_{\mathbb{S}} w(s')\hat{Q}^n(ds'|\hat{s}, f_\alpha(\hat{s})) \right| \\
&\leq \frac{L}{\theta}\sum_{n=0}^{\infty} \alpha^n (\|w\|_w R\eta^n w(s) + \|w\|_w R\eta^n w(\hat{s})) \\
&\leq \frac{RL}{\theta}\sum_{n=0}^{\infty} \alpha^n \eta^n [w(s) + w(\hat{s})] \\
&\leq \frac{RL}{\theta(1-\eta)}[1 + w(\hat{s})]w(s) =: v_2^*(s),
\end{aligned}$$

y así, la Condición 2.3 se cumple con $v_1^*(s) := -\frac{RL}{\theta(1-\eta)}[1 + w(\hat{s})]w(s)$.

(b) Por (2.1) y la parte *(b)* tenemos

$$\begin{aligned}
\int_{\mathbb{S}} w(s')\hat{Q}(ds'|s,a) &= \frac{\rho}{\tau(s,a)}\int_{\mathbb{S}} w(s')Q(ds'|s,a) + \left(1 - \frac{\rho}{\tau(s,a)}\right) w(s) \\
&\leq \frac{\rho}{\tau(s,a)}[\lambda_1 w(s) + \kappa I_C(s)] + \left(1 - \frac{\rho}{\tau(s,a)}\right) w(s) \\
&\leq \frac{\rho\lambda_1 w(s)}{\theta} + \frac{\rho\kappa}{\theta} I_C(s) + \left(1 - \frac{\rho}{M}\right) w(s) \\
&\leq \left(1 - \frac{\rho(1-\lambda_1)}{\theta}\right) w(s) + \frac{\rho\kappa}{\theta} I_C(s),
\end{aligned}$$

para todo $(s,a) \in \mathbb{K}$, y $\hat{Q}(D|s,a) \geq \frac{\rho}{\tau(s,a)} Q(D|s,a) \geq \frac{\rho}{M}\varepsilon\mu(D)$, para cada $D \subset C$, $s \in C$ y $a \in \mathbb{A}(s)$, lo que junto con el Teorema 2.3 en [17] tenemos que para cada $f \in \mathbb{F}$ la cadena de Markov definida por $\hat{Q}(\cdot|\cdot, f(\cdot))$ es ergódica w-geométrica uniforme. Por tanto, *(b)* implica *(a)*, y así, se llega a la Condición 2.3.

(c) Por el Lema 5.5 en [16] y por el Teorema 7.3.10 en [10], se concluye que para cada $f \in \mathbb{F}$ la cadena de Markov definida por $\hat{Q}(\cdot|\cdot, f(\cdot))$ es ergódica w-geométrica uniforme. Por tanto, *(c)* implica *(a)*, y así, se llega a la Condición 2.3.

(d) Para cada $s \neq \hat{s}$ y $f \in \mathbb{F}$, por *(d1)* y (2.1) tenemos

$$\begin{aligned} \hat{Q}(\{\hat{s}\}|s, f(s)) &\geq \frac{\rho}{\tau(s,a)} Q(\{\hat{s}\}|s, f(s)) \\ &\geq \frac{\rho}{M} Q(\{\hat{s}\}|s, f(s)) \geq \frac{\rho\lambda_3}{M} > 0. \end{aligned} \tag{2.9}$$

Así,

$$\hat{P}_s^f(s_k \neq \hat{s}, 1 \leq k \leq n) \leq \left(1 - \frac{\rho\lambda_3}{M}\right)^n, \text{ para toda } n \geq 1, \tag{2.10}$$

en efecto, por (2.9) se llega a que

$$\hat{P}_s^f(s_k \neq \hat{s}) = 1 - \hat{Q}(\{\hat{s}\}|s, f(s)) \leq 1 - \frac{\rho\lambda_3}{M},$$

luego se cumple (2.10) para $n = 1$. Supongamos que (2.10) se cumple para $n > 1$. Entonces, por (2.9) y la hipótesis de inducción se obtiene que

$$\begin{aligned} \hat{P}_s^f(s_k \neq \hat{s}, 1 \leq k \leq n+1) &= \hat{E}_s^f[I_{\{s_k \neq \hat{s}, 1 \leq k \leq n+1\}}] \\ &= \hat{E}_s^f[\hat{E}_s^f[I_{\{s_k \neq \hat{s}, 1 \leq k \leq n+1\}}|h_n, f(s_n)]] \\ &= \hat{E}_s^f[I_{\{s_k \neq \hat{s}, 1 \leq k \leq n\}}[1 - \hat{Q}(\{\hat{s}\}|s_n, f(s_n))]] \\ &\leq \left(1 - \frac{\rho\lambda_3}{M}\right) \hat{P}_s^f(s_k \neq \hat{s}, 1 \leq k \leq n) \\ &\leq \left(1 - \frac{\rho\lambda_3}{M}\right)^{n+1}, \end{aligned}$$

luego, se cumple (2.10). Por lo tanto, $\hat{P}_s^f(\cap_{k=1}^{\infty}\{s_k \neq \hat{s}\}) = 0$, lo cual nos da $\hat{P}_s^f(\sigma_{\hat{s}} < \infty) = 1$, para todo $s_k \neq \hat{s}$ y $f \in \mathbb{F}$. Más aún, por (2.1) y la condición *(d2)* tenemos

$$|\hat{c}(s, f(s))| = \left|\frac{c(s, f(s))}{\tau(s, f(s))}\right| \leq \frac{L_1 v_1(s)}{\tau(s, f(s))}, \text{ para todo } s \neq \hat{s}.$$

Sea $\overline{u}(s) := \frac{u_1(s)}{\rho}$ y $\overline{v}(s) := \frac{v_1(s)}{\tau(s,f(s))}$. Luego, por (2.1)

$$
\begin{aligned}
&\overline{v}(s) + \int_{\mathbb{S}\setminus\{\hat{s}\}} \overline{u}(s')\hat{Q}(ds'|s, f(s)) \\
&= \frac{v_1(s)}{\tau(s, f(s))} + \int_{\mathbb{S}\setminus\{\hat{s}\}} \frac{u_1(s')}{\tau(s, f(s))} Q(ds'|s, f(s)) + \left(1 - \frac{\rho}{\tau(s, f(s))}\right) \frac{u_1(s)}{\rho} \\
&= \frac{v_1(s)}{\tau(s, f(s))} + \frac{1}{\tau(s, f(s))} \int_{\mathbb{S}\setminus\{\hat{s}\}} u_1(s')Q(ds'|s, f(s)) - \frac{u_1(s)}{\tau(s, f(s))} + \frac{u_1(s)}{\rho},
\end{aligned}
$$

lo que, junto con la condición *(d2)* conlleva a

$$
\overline{v}(s) + \int_{\mathbb{S}\setminus\{\hat{s}\}} \overline{u}(s')\hat{Q}(ds'|s, f(s)) \le \frac{u_1(s)}{\rho} = \overline{u}(s), \quad \text{para todo } \ s \neq \hat{s}.
$$

Por lo tanto, por el Lema 2.3 se tiene

$$
\hat{E}_s^f \left[\sum_{k=0}^{\sigma_{\hat{s}}-1} |\hat{c}(s_k, f(s_k))| \right] \le \frac{L_1}{\rho} u_1(s), \quad \text{para todo } \ s \neq \hat{s} \text{ y } f \in \mathbb{F}. \tag{2.11}
$$

Más aún, obsérvese que para cada $f \in \mathbb{F}$ y $s \neq \hat{s}$,

$$
\begin{aligned}
1 + \int_{\mathbb{S}\setminus\{\hat{s}\}} \frac{M}{\rho\lambda_3} \hat{Q}(ds'|s, f(s)) &= 1 + \frac{M}{\rho\lambda_3} [1 - \hat{Q}(\{\hat{s}\}|s, f(s))] \\
&\le 1 + \frac{M}{\rho\lambda_3} \left(1 - \frac{\rho\lambda_3}{M}\right) = \frac{M}{\rho\lambda_3},
\end{aligned}
$$

que, junto con el Lema 2.3 obtenemos lo siguiente

$$
\hat{E}_s^f[\sigma_{\hat{s}}] \le \frac{M}{\rho\lambda_3}, \quad \text{para todo } \ s \neq \hat{s} \text{ y } f \in \mathbb{F}. \tag{2.12}
$$

Por otro lado, se sigue del Lema (2.2) que para cada $\alpha \in (0, 1)$, existe un política estacionaria óptima α-descontada $f_\alpha \in \mathbb{F}$ tal que $V_\alpha(s, f_\alpha) = V_\alpha^*(s)$ para todo $s \in \mathbb{S}$. Así,

$$
\begin{aligned}
h_\alpha(s) &= V_\alpha(s, f_\alpha) - V_\alpha(\hat{s}, f_\alpha) \\
&= \hat{E}_s^{f_\alpha} \left[\sum_{n=0}^{\infty} \alpha^n \hat{c}(s_n, f_\alpha(s_n)) \right] - V_\alpha(\hat{s}, f_\alpha) \\
&= \hat{E}_s^{f_\alpha} \left[\sum_{n=0}^{\sigma_{\hat{s}}-1} \alpha^n \hat{c}(s_n, f(s_n)) \right] + \hat{E}_s^{f_\alpha} \left[\sum_{n=\sigma_{\hat{s}}}^{\infty} \alpha^n \hat{c}(s_n, f_\alpha(s_n)) \right] - V_\alpha(\hat{s}, f_\alpha) \\
&= \hat{E}_s^{f_\alpha} \left[\sum_{n=0}^{\sigma_{\hat{s}}-1} \alpha^n \hat{c}(s_n, f(s_n)) \right] + \hat{E}_s^{f_\alpha} [\alpha^{\sigma_{\hat{s}}} - 1] V_\alpha(\hat{s}, f_\alpha),
\end{aligned}
$$

lo que junto con la desigualdad $1-\alpha^{\sigma_{\hat{s}}} \leq (1-\alpha)\sigma_{\hat{s}}$ implican que

$$|h_\alpha(s)| \leq \hat{E}_s^{f_\alpha}\left[\sum_{n=0}^{\sigma_{\hat{s}}-1} |\hat{c}(s_n, f_\alpha(s_n))|\right] + (1-\alpha)|V_\alpha(\hat{s}, f_\alpha)|\hat{E}_s^{f_\alpha}[\sigma_{\hat{s}}]. \quad (2.13)$$

Más aún, por el Lema 2.2 se tiene que

$$(1-\alpha)|V_\alpha(\hat{s}, f_\alpha)| \leq \frac{L}{\theta}w(\hat{s}) + \frac{bLM}{(1-\beta)\theta^2}, \ \text{ para todo } \ \alpha \in (0,1). \quad (2.14)$$

Por lo tanto, de (2.11)-(2.14) tenemos

$$|h_\alpha(s)| \leq \frac{L_1}{\rho}u_1(s) + \frac{LM}{\theta\rho\lambda_3}w(\hat{s}) + \frac{bLM^2}{(1-\beta)\rho\lambda_3\theta^2} =: v_2^*(s),$$

luego, se cumple la Condición 2.3 con $v_1^*(s) := -[\frac{L_1}{\rho}u_1(s) + \frac{LM}{\theta\rho\lambda_3}w(\hat{s}) + \frac{bLM^2}{(1-\beta)\rho\lambda_3\theta^2}]$. □

La condición *(b)* de Lema 2.4 es la misma que en [13] y [14], mientras que la condición *(c)* es la misma que en [16]. Como se puede ver, las condiciones *(a)* y *(b)* implican la condición de ergodicidad, por tanto, el Lema 2.4 indica que las condiciones usadas en [13], [14] y [16] son suficientes para que se cumplan la Condición 2.3 y la de ergodicidad.

2.3. Existencia de Políticas Óptimas

El objetivo de esta sección es probar la existencia de una política estacionaria óptima bajo el criterio de costo promedio, como se dijo en la parte introductoria no podemos usar la EOCP puesto que las condiciones dadas aquí son ligeramente más débiles que las ya existentes (ver, por ejemplo, [13] y [16]), para probar esto se reemplazará la EOCP por dos desigualdades de optimalidad promedio y asegurar la existencia de una solución a estas dos desigualdades usando la transformación de Schweitzer.

Lema 2.5 *Bajo las Condiciones 2.1, 2.2 y 2.3, existen una única constante $\hat{g}$, dos funciones $\hat{h}_1^*, \hat{h}_2^* \in \mathbb{B}_w(\mathbb{S})$, y un política $f \in \mathbb{F}$, que satisfacen* *las dos desigualdades de optimalidad promedio:*

$$\hat{g} + \hat{h}_1^*(s) \leq \inf_{a \in \mathbb{A}(s)} \left\{ \hat{c}(s,a) + \int_{\mathbb{S}} \hat{h}_1^*(s')\hat{Q}(ds'|s,a) \right\}, \tag{2.15}$$

$$\hat{g} + \hat{h}_2^*(s) \geq \inf_{a \in \mathbb{A}(s)} \left\{ \hat{c}(s,a) + \int_{\mathbb{S}} \hat{h}_2^*(s')\hat{Q}(ds'|s,a) \right\} \tag{2.16}$$

$$= \hat{c}(s, \hat{f}(s)) + \int_{\mathbb{S}} \hat{h}_2^*(s')\hat{Q}(ds'|s, \hat{f}(s)) \tag{2.17}$$

para todo $s \in \mathbb{S}$.

Demostración. Sea $\hat{s}$ como en la Condición 2.3, y sea $\{\alpha_n\}$ cualquier sucesión creciente de factores de descuento tal que $\alpha_n \to 1$ cuando $n \to \infty$. Por el Lema 2.2*(a)*, $(1-\alpha_n)V^*_{\alpha_n}(\hat{s})$ está acotado para $n \geq 1$. Por consiguiente, existe una subsucesión, $\{\alpha_k\}$, de $\{\alpha_n\}$ y una constante $\hat{g}$ que satisfacen

$$\lim_{k \to \infty} (1-\alpha_k)V^*_{\alpha_k}(\hat{s}) = \hat{g}, \quad \hat{h}_1^* := \limsup_{k \to \infty} h_{\alpha_k}(s). \tag{2.18}$$

Como $|h_{\alpha_k}| \leq |v_1^*| + |v_2^*|$, $\hat{h}_1^*$ pertenece a $\mathbb{B}_w(\mathbb{S})$ (por la Condición 2.3), retomando que $h_\alpha(s) = V^*_\alpha(s) - V^*_\alpha(\hat{s})$, del Teorema 8.3.6*(a)* en [10] tenemos que

$$(1-\alpha)V^*_\alpha(\hat{s}) + h_\alpha(s) = \min_{a \in \mathbb{A}(s)} \left\{ \hat{c}(s,a) + \alpha \int_{\mathbb{S}} h_\alpha(s')\hat{Q}(ds'|s,a) \right\},$$

para todo $s \in \mathbb{S}$, lo que conlleva a que

$$(1-\alpha_k)V^*_{\alpha_k}(\hat{s}) + h_{\alpha_k}(s) \leq \hat{c}(s,a) + \int_{\mathbb{S}} \alpha_k h_{\alpha_k}(s')\hat{Q}(ds'|s,a), \tag{2.19}$$

para todo $s \in \mathbb{S}$ y $a \in \mathbb{A}(s)$. Por la extensión del Lema de Fatou (ver Apéndice A), haciendo $k \to \infty$ en (2.19) y por (2.18) se llega a que

$$\hat{g} + \hat{h}_1^*(s) \leq \hat{c}(s,a) + \int_{\mathbb{S}} \hat{h}_1^*(s')\hat{Q}(ds'|s,a),$$

para todo $s \in \mathbb{S}$, $a \in \mathbb{A}(s)$, lo cual se cumple (2.15).

Ahora veamos (2.16). Sea

$$\hat{h}_2^*(s) := \liminf_{k \to \infty} h_{\alpha_k}(s) \in \mathbb{B}_w(\mathbb{S}),$$

para cada $s \in \mathbb{S}$, siempre que

$$\hat{h}_2^*(s) = \lim_{k\to\infty} g_{\alpha_k}(s), \text{ donde}$$
$$g_{\alpha_k}(s) := \inf\{h_{\alpha_m}(s) : m \geq k\} \leq h_{\alpha_k}(s),$$

para toda $k \geq 1$. Similarmente, del Teorema 8.3.6*(b)* en [10] y como $h_{\alpha_k} \geq g_{\alpha_k}$, se obtiene que

$$\begin{aligned}(1-\alpha_k)V_{\alpha_k}^*(\hat{s}) + h_{\alpha_k}(s) &= \min_{a\in\mathbb{A}(s)} \left\{\hat{c}(s,a) + \alpha_k \int_{\mathbb{S}} h_{\alpha_k}(s')\hat{Q}(ds'|s,a)\right\} \\ &\geq \min_{a\in\mathbb{A}(s)} \left\{\hat{c}(s,a) + \alpha_k \int_{\mathbb{S}} g_{\alpha_k}(s')\hat{Q}(ds'|s,a)\right\}.\end{aligned} \tag{2.20}$$

Como $\alpha_{k+1} \geq \alpha_k > 0$ y $g_{\alpha_{k+1}} - g_{\alpha_1} \geq g_{\alpha_k} - g_{\alpha_1} \geq 0$, se observa que los límites

$$\lim_{k\to\infty} \min_{a\in\mathbb{A}(s)} \left\{\hat{c}(s,a) + \alpha_k \int_{\mathbb{S}} (g_{\alpha_k}(s') - g_{\alpha_1}(s'))\hat{Q}(ds'|s,a)\right\}$$

y

$$\lim_{k\to\infty} \min_{a\in\mathbb{A}(s)} \left\{\hat{c}(s,a) + \alpha_k \int_{\mathbb{S}} g_{\alpha_k}(s')\hat{Q}(ds'|s,a)\right\} \tag{2.21}$$

existen. Ahora, por (2.18) y (2.20)

$$\hat{g} + \hat{h}_2^*(s) \geq \lim_{k\to\infty} \min_{a\in\mathbb{A}(s)} \left\{\hat{c}(s,a) + \alpha_k \int_{\mathbb{S}} g_{\alpha_k}(s')\hat{Q}(ds'|s,a)\right\}. \tag{2.22}$$

Por otro lado, para cada $k \geq 1$ fijo, por la Condición 2.2, existe $a_k(s) \in \mathbb{A}(s)$ (que depende de k y de s) tal que

$$\begin{aligned}&\min_{a\in\mathbb{A}(s)} \left\{\hat{c}(s,a) + \alpha_k \int_{\mathbb{S}} g_{\alpha_k}(s')\hat{Q}(ds'|s,a)\right\} \\ &\qquad = \hat{c}(s,a_k(s)) + \alpha_k \int_{\mathbb{S}} g_{\alpha_k}(s')\hat{Q}(ds'|s,a_k(s)).\end{aligned} \tag{2.23}$$

Puesto que $\mathbb{A}(s)$ es un conjunto compacto, existe una sub-sucesión, $\{a_{k_i}(s)\}$, de $\{a_k(s)\}$ tal que el límite $\lim_{i\to\infty} a_{k_i}(s)$ existe; lo denotamos por $a'(s) \in \mathbb{A}(s)$. Se observa que $\|g_{\alpha_k}\|_w \leq \|v_1^*\|_w + \|v_2^*\|_w$ para todo $k \geq 1$, de la Condición 2.2*(b)*, (2.21)-(2.23), y la extensión del Lema de

Fatou se tiene que

$$\begin{aligned}
\hat{g} + \hat{h}_2^*(s) &\geq \lim_{i\to\infty} \min_{a\in\mathbb{A}(s)} \left\{ \hat{c}(s,a) + \alpha_{k_i} \int_{\mathbb{S}} g_{\alpha_{k_i}}(s')\hat{Q}(ds'|s,a) \right\} \\
&= \lim_{i\to\infty} \min_{a\in\mathbb{A}(s)} \left[\hat{c}(s,a_{k_i}(s)) + \alpha_{k_i} \int_{\mathbb{S}} g_{\alpha_{k_i}}(s')\hat{Q}(ds'|s,a_{k_i}(s)) \right] \\
&\geq \hat{c}(s,a'(s)) + \int_{\mathbb{S}} \hat{h}_2^*(s')\hat{Q}(ds'|s,a'(s)) \\
&\geq \min_{a\in\mathbb{A}(s)} \left\{ \hat{c}(s,a) + \int_{\mathbb{S}} \hat{h}_2^*(s')\hat{Q}(ds'|s,a'(s)) \right\}.
\end{aligned}$$

Así, se cumple (2.16). Más aún, (2.16) junto con el *Teorema de Selección Medible* (ver [10], pág. 50) implican la existencia de $\hat{f} \in \mathbb{F}$ satisfaciéndose (2.17), luego, se completa la prueba. □

Ahora se demuestra el teorema sobre la existencia de una política óptima de costo promedio.

Teorema 2.1 *Bajo las Condiciones 2.1, 2.2 y 2.3, se cumple lo siguiente.*

(a) Existen una única constante g^*, dos funciones $h_1^*, h_2^* \in \mathbb{B}_w(\mathbb{S})$, y una política estacionaria $f^* \in \mathbb{F}$, que satisfacen las siguientes dos desigualdades de optimalidad promedio:

$$h_1^*(s) \leq \inf_{a\in\mathbb{A}(s)} \left\{ c(s,a) - g^*\tau(s,a) + \int_{\mathbb{S}} h_1^*(s')Q(ds'|s,a) \right\}, \tag{2.24}$$

$$h_2^*(s) \geq \inf_{a\in\mathbb{A}(s)} \left\{ c(s,a) - g^*\tau(s,a) + \int_{\mathbb{S}} h_2^*(s')Q(ds'|s,a) \right\} \tag{2.25}$$

$$= c(s,f^*(s)) - g^*\tau(s,f^*(s)) + \int_{\mathbb{S}} h_2^*(s')Q(ds'|s,f^*(s)), \tag{2.26}$$

para todo $s \in \mathbb{S}$.

(b) $g^* = \inf_{\pi\in\Pi} J(s,\pi)$, para todo $s \in \mathbb{S}$.

(c) Cualquier política estacionaria $f \in \mathbb{F}$ que cumpla con el mínimo en (2.25) es óptima, y así f^* en (2.26) es una política estacionaria óptima de costo promedio.

Demostración. *(a)* Sea $g^* := \hat{g}$, $h_1^*(s) := \rho\hat{h}_1^*(s)$, y $h_2^*(s) := \rho\hat{h}_2^*(s)$ para todo $s \in \mathbb{S}$. Claramente $h_1^*, h_2^* \in \mathbb{B}_w(\mathbb{S})$. Entonces por (2.1) y (2.15) tenemos

$$\begin{aligned}
&\hat{c}(s,a) + \int_{\mathbb{S}} \hat{h}_1^*(s')\hat{Q}(ds'|s,a) \\
&= \frac{c(s,a)}{\tau(s,a)} + \frac{1}{\tau(s,a)}\int_{\mathbb{S}} h_1^*(s')Q(ds'|s,a) + \frac{1}{\rho}\left(1 - \frac{\rho}{\tau(s,a)}\right)h_1^*(s) \\
&\geq g^* + \frac{1}{\rho}h_1^*(s), \quad \text{para todo } (s,a) \in \mathbb{K},
\end{aligned}$$

lo cual implica (2.24). Análogamente, por (2.1) y (2.17) se tiene

$$\begin{aligned}
&\hat{c}(s,\hat{f}(s)) + \int_{\mathbb{S}} \hat{h}_2^*(s')\hat{Q}(ds'|s,\hat{f}(s)) \\
&= \frac{c(s,\hat{f}(s))}{\tau(s,\hat{f}(s))} + \frac{1}{\tau(s,\hat{f}(s))}\int_{\mathbb{S}} h_2^*(s')Q(ds'|s,\hat{f}(s)) \\
&\quad + \frac{1}{\rho}\left(1 - \frac{\rho}{\tau(s,\hat{f}(s))}\right)h_2^*(s) \\
&\leq g^* + \frac{1}{\rho}h_2^*(s), \quad \text{para todo } s \in \mathbb{S},
\end{aligned}$$

lo cual implica (2.25). Más aún, (2.25) junto con el *Teorema de Selección Medible* (ver [10]) aseguran que existe $f^* \in \mathbb{F}$ que satisface (2.26), luego, se completa la prueba de la parte *(a)*.

(b) Para cada $\pi \in \Pi$ y $s \in \mathbb{S}$ de (2.24) se tiene

$$h_1^*(s_k) \leq c(s_k,a_k) - g^*\tau(s_k,a_k) + \int_{\mathbb{S}} h_1^*(s')Q(ds'|s_k,a_k),$$

para todo $s_k \in \mathbb{S}$, $a_k \in \mathbb{A}(s_k)$ y $k \geq 0$, lo que junto con (1.4) tenemos que

$$E_s^\pi[h_1^*(s_k)] \leq E_s^\pi[c(s_k,a_k) - g^*\tau(s_k,a_k) + h_1^*(s_{k+1})], \tag{2.27}$$

para todo $k \geq 0$. Consecuentemente, sumando (2.27) sobre $k = 0,1,\ldots,n-1$, obtenemos que

$$h_1^*(s) \leq E_s^\pi\left[\sum_{k=0}^{n-1} c(s_k,a_k)\right] - g^*E_s^\pi\left[\sum_{k=0}^{n-1}\tau(s_k,a_k)\right] + E_s^\pi h_1^*(s_n),$$

para todo $n \geq 1$, así

$$g^* \leq \frac{E_s^\pi[\sum_{k=0}^{n-1} c(s_k, a_k)]}{E_s^\pi[\sum_{k=0}^{n-1} \tau(s_k, a_k)]} + \frac{E_s^\pi h_1^*(s_n) - h_1^*(s)}{E_s^\pi[\sum_{k=0}^{n-1} \tau(s_k, a_k)]}, \quad \text{para todo } n \geq 1. \tag{2.28}$$

Por el Lema 2.1, tenemos

$$\begin{aligned} 0 &\leq \lim_{n\to\infty} \frac{|E_s^\pi h_1^*(s_n) - h_1^*(s)|}{E_s^\pi[\sum_{k=0}^{n-1} \tau(s_k, a_k)]} \\ &\leq \lim_{n\to\infty} \left[\frac{(1+\beta^n)\|h_1^*\|_w w(s)}{n\theta} + \frac{b\|h_1^*\|_w}{n\theta(1-\beta)}\right] = 0, \end{aligned}$$

lo que junto con (2.28) se llega a que

$$g^* \leq J(s, \pi), \quad \text{para todo } s \in \mathbb{S} \text{ y } \pi \in \Pi.$$

Por tanto,

$$g^* \leq J^*(s) = \inf_{\pi\in\Pi} J(s, \pi), \quad \text{para todo } s \in \mathbb{S}. \tag{2.29}$$

Similarmente, por (2.26)

$$h_2^*(s_k) \geq c(s_k, f^*(s_k)) - g^*\tau(s_k, f^*(s_k)) + \int_{\mathbb{S}} h_2^*(s')Q(ds'|s_k, f^*(s_k)), \tag{2.30}$$

para todo $s \in \mathbb{S}$ y $k \geq 0$. Similarmente como en la prueba de (2.28), por (2.30) se obtiene

$$g^* \geq \frac{E_s^{f^*}[\sum_{k=0}^{n-1} c(s_k, f^*(s_k))]}{E_s^{f^*}[\sum_{k=0}^{n-1} \tau(s_k, f^*(s_k))]} + \frac{E_s^{f^*} h_2^*(s_n) - h_2^*(s)}{E_s^{f^*}[\sum_{k=0}^{n-1} \tau(s_k, f^*(s_k))]}, \tag{2.31}$$

para todo $n \geq 1$. Notemos que

$$\begin{aligned} 0 &\leq \lim_{n\to\infty} \frac{|E_s^{f^*} h_2^*(s_n) - h_2^*(s)|}{E_s^{f^*}[\sum_{k=0}^{n-1} \tau(s_k, f^*(s_k))]} \\ &\leq \lim_{n\to\infty} \left[\frac{(1+\beta^n)\|h_2^*\|_w w(s)}{n\theta} + \frac{b\|h_2^*\|_w}{n\theta(1-\beta)}\right] = 0, \end{aligned}$$

y haciendo tender $n \to \infty$ en (2.31), obtenemos

$$g^* \geq J(s, f^*), \quad \text{para todo } s \in \mathbb{S}. \tag{2.32}$$

De (2.29) y (2.32), $g^* = J(s, f^*) = \inf_{\pi\in\Pi} J(s,\pi)$, para todo $s \in \mathbb{S}$, luego se cumple la parte *(b)*.

(c) La parte *(c)* se sigue inmediato de los incisos *(a)* y *(b)*. □

Observación 2.2 *(a) Las dos desigualdades de optimalidad promedio (2.24) y (2.25) son una generalización de las dos desigualdades de optimalidad promedio en [8] para PDM a tiempo discreto y de la EOCP en [13], [14] y [16] para PDSM (ver [25]).*

(b) Cuando $\mathbb{S}$ *es* ***numerable****, el* argumento de diagonalización estándar *(ver [8]), nos ayuda para mostrar la existencia de una subsucesión* $\{h_{\alpha_k}(s)\}$*, tal que el límite* $h^* := \lim_{k\to\infty} h_{\alpha_k}(s)$ *existe para todo* $s \in \mathbb{S}$*. Por tanto, se tiene que* $h_1^*(s) = h_2^*(s) = h^*(s)$*, para todo* $s \in \mathbb{S}$*, consecuentemente las desigualdades (2.24) y (2.25) coinciden, y así la EOCP se obtiene (ver [8] y [25]). Para los ejemplos del Capítulo 3, se considera el espacio de estados numerable por lo cual la EOCP se cumple.*

2.4. Metodología

A continuación se explica la metodología a seguir para resolver el PCO en el caso semi-Markoviano. Los pasos se describen a continuación:

- Se considera un modelo de decisión semi-Markoviano fijo, estacionario, a tiempo continuo.
- La transformación de Schweitzer; (ver Subsección 2.2) se usa para asegurar la existencia de políticas óptimas estacionarias para los PDSM (ver [5] y [19]), bajo esta transformación se pasa de un modelo en tiempo continuo a un modelo a tiempo discreto.
- Con el modelo discreto se resuelve el PCO con el criterio de costo descontado, (ver, por ejemplo, [3] y [19]), también conocido como el factor de descuento desvaneciente.

- Una vez resuelto el PCO bajo el criterio descontado, se regresa al modelo original, esto haciendo tender al factor de descuento desvaneciente a uno.

Capítulo 3

Ejemplos

En este capítulo se ejemplifica la teoría de los PDSM junto con el criterio de rendimiento que se estudió en el Capítulo 2, conocido como la razón de costo promedio.

Se estudian dos modelos de línea de espera; el primero de ellos es un modelo de colas en donde el parámetro del servicio está controlado, la variable del estado representa el número de clientes en el sistema. En el segundo ejemplo se considera a una línea de espera con recesos controlados. Es importante mencionar que, en el primero ejemplo, el espacio de acciones se considera de Borel y en el segundo se toma un conjunto compacto. El espacio de estados es numerable para ambos casos.

Se verifica que existe una política estacionaria óptima de costo promedio en ambos casos para valores específicos en los parámetros y se muestran los costos mínimos por operar en cierto estado del sistema, así como la política que minimiza el criterio de rendimiento, esto con la ayuda de *Mathematica 8.0*, en el Apéndice B se muestran los códigos de las funciones de valor iterado para cada ejemplo.

3.1. Línea de Espera con Servicios Controlados

3.1.1. Descripción del Modelo

Consideremos una línea de espera con control sobre el parámetro de servicio, la variable del estado denota el número de clientes en el sistema. Así, el espacio de estados está representado por $\mathbb{S} := \{0, 1, 2, \ldots\}$. Los tiempos de interarribos son independientes e idénticamente distribuidos de manera exponencial con parámetro $\lambda^* > 0$, mientras que los tiempos de servicio son independientes de los tiempos de arribo y siguen una distribución uniforme sobre $[0, \mu^*]$, donde el parámetro μ^* está controlado. El agente controlador puede cambiar únicamente el servicio cuando se complete uno de éstos o bien cuando haya un arribo y el sistema se encuentre vacío. Cuando el sistema está en el estado $s = 0$, naturalmente no se necesita ningún servicio. Cuando el sistema se encuentre en el estado $s \in \{1, 2, \ldots\}$, suponemos que el agente toma una acción de un conjunto dado $\mathbb{A}(s) := [\mu_1^*, \mu_2^*]$ con $\mu_2^* > \mu_1^* > 0$. Así, se obtiene el espacio de acciones de la siguiente manera,

$$\mathbb{A}(s) = \begin{cases} \{0\}, & s = 0, \\ [\mu_1^*, \mu_2^*], & s \in \{1, 2, \ldots\}, \quad \mu_2^* > \mu_1^* > 0. \end{cases}$$

Más aún, el sistema se incurre en un costo de almacenamiento de tasa $c_1(s)$ por unidad de tiempo cuando hay s clientes en el sistema, y un costo de tasa $c_2(a)$ por unidad de tiempo por haber elegido el parámetro de servicio a.

Se formula este sistema como un PDSM. La ley de transición $Q(\cdot|s, a)$, la distribución del tiempo de permanencia $H(\cdot|s, a)$ y la función de costos están dadas de la siguiente manera:

$$Q(1|0, 0) = 1, \tag{3.1}$$

y para cada $s \geq 1$ y $a \in \mathbb{A}(s)$,

$$Q(s'|s,a) = \begin{cases} \frac{1}{a}\int_0^a \frac{e^{-\lambda^* t}(\lambda^* t)^{s'-s+1}}{(s'-s+1)!}dt, & s' \geq s-1, \\ 0, & e.o.c. \end{cases} \tag{3.2}$$

$$H(t|s,a) = \begin{cases} 1-e^{-\lambda^* t}, & s=0,\ a=0,\ t\geq 0, \\ \frac{t}{a}, & s\geq 1,\ a\in\mathbb{A}(s),\ t\in[0,a], \\ 0, & e.o.c. \end{cases} \tag{3.3}$$

$$c(s,a) = \begin{cases} \frac{1}{\lambda^*}[c_1(0)+c_2(0)], & s=0,\ a=0, \\ \frac{a}{2}[c_1(s)+c_2(a)], & e.o.c. \end{cases} \tag{3.4}$$

Observación 3.1

- *La ley de transición $Q(s'|s,a)$ dada en la ecuación (3.2) nos representa la probabilidad de pasar del estado s al estado s' y haber elegido la acción a. Ésta sigue una distribución Poisson con parámetro $\lambda^* t$.*

- *La distribución del tiempo de permanencia $H(t|s,a)$, ecuación (3.3), para $s=0$, $a=0$ y $t\geq 0$, es igual a la distribución acumulada de la exponencial con parámetro $\lambda^* t$. Para $s\geq 1$, $H(t|s,a)$ sigue una distribución uniforme en $[0,a]$. En cualquier otro caso $H(t|s,a)$ es nula.*

- *La función de costos depende de la suma de las tasas por unidad de tiempo cuando hay s clientes en el sistema y por haber elegido el parámetro de servicio a, $c_1(s)$ y $c_2(a)$, respectivamente.*

El objetivo es encontrar condiciones que garanticen la existencia de una política estacionaria óptima de costo promedio. Para esto, considérese el siguiente conjunto de condiciones.

(C1) Las contantes λ^*, μ_1^* y μ_2^* satisfacen $0<\mu_1^*<\mu_2^*\leq \frac{3}{2(e-1)\lambda^*}$.

(C2) La función $c_2(a)$ es continua en $a\in[\mu_1^*,\mu_2^*]$.

(C3) Existe una constante positiva L^* tal que $|c_1(s)|\leq L^*(s+1)$ para todo $s\in\mathbb{S}$.

Bajo las condiciones anteriores, tenemos el siguiente resultado.

Proposición 3.1 *Bajo las condiciones* (C1), (C2) y (C3), *el sistema de colas descrito anteriormente satisface las Condiciones 2.1, 2.2 y 2.3. Por tanto, por el Teorema 2.1, existe una política estacionaria óptima de costo promedio.*

Demostración. Sea $\beta := \frac{e^{\lambda^* \mu_2^*(e-1)}-1}{\lambda^* \mu_2^* e(e-1)}$, $b := e$ y $w(s) := e^s$ para todo $s \in \mathbb{S}$. Por la condición *(C1)*, tenemos $0 < \beta < 1$. Verifiquemos primero la Condición 2.1. De (3.1) tenemos que

$$\sum_{s' \in \mathbb{S}} w(s')Q(s'|0,0) = e \leq \beta w(0) + b.$$

Además, por (3.2) tenemos que, para todo $s \geq 1$ y $a \in \mathbb{A}(s)$,

$$Q(s'|s,a) = \frac{1}{\lambda^* a}(1 - e^{-\lambda^* a}) - \sum_{k=1}^{s'-s+1} \frac{(\lambda^* a)^{k-1}}{k!} e^{-\lambda^* a}, \quad s' \geq s-1. \tag{3.5}$$

Por tanto, para todo $s \geq 1$ y $a \in \mathbb{A}(s)$, se sigue de (3.5) que

$$\begin{aligned}
\sum_{s' \in \mathbb{S}} w(s')Q(s'|s,a) &= \sum_{s'=s-1}^{\infty} e^{s'} \left[\frac{1}{\lambda^* a}(1 - e^{-\lambda^* a}) - \sum_{k=1}^{s'-s+1} \frac{(\lambda^* a)^{k-1}}{k!} e^{-\lambda^* a} \right] \\
&= \sum_{s'=s-1}^{\infty} e^{s'} \frac{1}{\lambda^* a} \left[e^{\lambda^* a} - \sum_{k=0}^{s'-s+1} \frac{(\lambda^* a)^k}{k!} \right] e^{-\lambda^* a} \\
&= \frac{1}{\lambda^* a} e^{-\lambda^* a} \sum_{s'=s-1}^{\infty} e^{s'} \sum_{k=s'-s+2}^{\infty} \frac{(\lambda^* a)^k}{k!} \\
&= \frac{1}{\lambda^* a} e^{-\lambda^* a} e^{s-1} \sum_{k=1}^{\infty} \sum_{n=0}^{k-1} \frac{(\lambda^* a)^k}{k!} e^n \\
&= \frac{1}{\lambda^* a e} e^{-\lambda^* a} \sum_{k=1}^{\infty} \frac{(e^k - 1)(\lambda^* a)^k}{(e-1)k!} w(s) \\
&= \frac{1}{\lambda^* a e(e-1)} [e^{\lambda^* a(e-1)} - 1] w(s).
\end{aligned} \tag{3.6}$$

Para cada $s \geq 1$, sea

$$z(a) := \frac{1}{\lambda^* e(e-1)a} [e^{\lambda^* a(e-1)} - 1], \quad \text{para todo } a \in \mathbb{A}(s).$$

Calculando la derivada de la función $z(a)$, obtenemos

$$\frac{dz(a)}{da} = \frac{e^{\lambda^* a(e-1)}}{\lambda^* e(e-1)a^2}[\lambda^*(e-1)a + e^{-\lambda^*(e-1)a} - 1], \quad \text{para todo } a \in \mathbb{A}(s),$$

lo que junto con la desigualdad $1 - e^{-m} \leq m$, para todo $m > 0$ tenemos que $\frac{dz(a)}{da} \geq 0$, para todo $a \in \mathbb{A}(s)$. Así, por (3.6) se llega a que

$$\sum_{s' \in \mathbb{S}} w(s')Q(s'|s,a) \leq \frac{1}{\lambda^* \mu_2^* e(e-1)}[e^{\lambda^* \mu_2^*(e-1)} - 1]w(s) \leq \beta w(s) + b, \tag{3.7}$$

así, se sigue la Condición 2.1*(a)*.

Por otro lado, de la condición *(C2)*, existe una constante $c^* > |c_2(0)|$ tal que $|c_2(a)| \leq c^*$, para todo $a \in [\mu_1^*, \mu_2^*]$. Sea $L := \frac{L^* + q^*}{\lambda^*}$. Entonces, por (3.4), las condiciones *(C1)* y *(C3)* se obtiene que

$$|c(s,a)| \leq L(s+1) \leq Lw(s), \quad \text{para todo } (s,a) \in \mathbb{K}.$$

Luego por (3.3)

$$\tau(s,a) = \begin{cases} \frac{1}{\lambda^*}, & s = 0 \text{ y } a = 0, \\ \frac{a}{2}, & e.o.c, \end{cases} \tag{3.8}$$

lo que junto con la condición *(C1)* nos lleva a que

$$\frac{\mu_1^*}{2} \leq \tau(s,a) \leq \frac{1}{\lambda^*}, \quad \text{para todo } (s,a) \in \mathbb{K},$$

y así se cumple la Condición 2.1*(b)*.

Por la descripción del modelo, (3.5), (3.6), (3.8) y la condición *(C2)*, se observa que la Condición 2.2 se cumple.

Finalmente, para verificar la Condición 2.3, sea $C := \{0,1\}$, $\lambda_1 := \beta$, $\kappa = b$, $\mu(\cdot) := \delta_1(\cdot)$ y $\varepsilon := \min_{a \in [\mu_1^*, \mu_2^*]}\{\frac{1}{\lambda^* a}(1 - e^{-\lambda^* a}) - e^{-\lambda^* a}\}$.

Entonces, por (3.1), (3.6) y (3.7) se obtienen

$$\sum_{s' \in \mathbb{S}} w(s')Q(s'|s,a) \leq \beta w(s) + \kappa I_C(s), \quad \text{para todo } (s,a) \in \mathbb{K},$$

y

$$Q(D|s,a) \geq \varepsilon \mu(D), \quad \text{para todo } D \subset C, \ s \in C \text{ y } a \in \mathbb{A}(s),$$

lo que junto con el Lema 2.4*(b)* hacen que la Condición 2.3 se cumpla, lo que concluye la demostración. □

3.1.2. Transformación del Modelo

Como se dijo con anterioridad, la transformación de Schweitzer nos permite trabajar con el modelo descontado para encontrar políticas óptimas.

De (2.1) y (3.8) se obtienen fácilmente el costo y la ley de transición transformados.

$$\hat{c}(s,a) = \begin{cases} c_1(0) + c_2(0), & s = 0,\ a = 0; \\ c_1(s) + c_2(a), & s \geq 1,\ a \in \mathbb{A}(s), \end{cases}$$

y

$$\hat{Q}(1|0,0) = \lambda\rho,$$

para $s' \geq s - 1$

$$\hat{Q}(s'|s,a) = \lambda\rho Q(s'|s,a) + (1 - \lambda\rho) I_{s'}(s).$$

Ahora veamos el criterio de costo descontado, del Lema 2.2 (consecuentemente de [10]) tenemos que las funciones de valor iterado se definen de la siguiente manera, para $\alpha \in (0,1)$,

$$v_n(s) = \min_{a \in \mathbb{A}(s)} \left[\hat{c}(s,a) + \alpha \sum_{s'=1}^{\infty} v_{n-1}(s') \hat{Q}(ds'|s,a) \right], \tag{3.9}$$

para todo $n \geq 1$ y $s \in \mathbb{S}$, con $v_0(\cdot) := 0$.

El Teorema de Programación Dinámica establece, entre otras cosas, que la sucesión $\{v_n\}$ converge geométricamente en la w-norma a la función de valor óptimo, V^*, (ver [10]).

De (2.1), (3.5) y (3.9) se tiene que, para $s \geq 1$,

$$\begin{aligned} v_n(s) = \min_{a \in \mathbb{A}(s)} \Big\{ c_1(s) + c_2(a) + \alpha \sum_{s'=1}^{\infty} v_{n-1}(s') \Big[\lambda\rho \Big(\frac{1}{\lambda^* a} (1 - e^{-\lambda^* a}) \\ - \sum_{k=1}^{s'-s+1} \frac{(\lambda^* a)^{k-1}}{k!} e^{-\lambda^* a} \Big) + (1 - \lambda\rho) I_{s'}(s) \Big] \Big\}. \end{aligned} \tag{3.10}$$

3.1.3. Aproximación Numérica al Valor Óptimo Promedio del Modelo

A continuación se presenta un ejemplo numérico del modelo descrito en la Sección 3.1.

Para esto, sean

$$c_1(s) := s \quad \text{y} \quad c_2(a) := a^2,$$

esto para fines de dar el ejemplo con funciones concretas.

Ejemplo 3.1 *La Tabla 3.1 muestra los costos de valor iterado para* $n \geq 0$*, junto con la acción* a *que depende de la función de costos iterada* $v_n(s)$*, la cual minimiza los costos por estar en el estado* s*.*

$v_0(1),\, a_{v_0(1)}$	$v_0(2),\, a_{v_0(2)}$	$v_0(3),\, a_{v_0(3)}$ $\cdots$	$v_0(s),\, a_{v_0(s)}$
$v_1(1),\, a_{v_1(1)}$	$v_1(2),\, a_{v_1(2)}$	$v_1(3),\, a_{v_1(3)}$ $\cdots$	$v_1(s),\, a_{v_1(s)}$
$v_2(1),\, a_{v_2(1)}$	$v_2(2),\, a_{v_2(2)}$	$v_2(3),\, a_{v_2(3)}$ $\cdots$	$v_2(s),\, a_{v_2(s)}$
$v_3(1),\, a_{v_3(1)}$	$v_3(2),\, a_{v_3(2)}$	$v_3(3),\, a_{v_3(3)}$ $\cdots$	$v_3(s),\, a_{v_3(s)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$v_n(1),\, a_{v_n(1)}$	$v_n(2),\, a_{v_n(2)}$	$v_n(3),\, a_{v_n(3)}$ $\cdots$	$v_n(s),\, a_{v_n(s)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$

Tabla 3.1: Función de costos de valor iterado.

Usando el software *Mathematica 8.0*, se aproximaron los valores v_n dados en (3.10). En el Apéndice B se muestra el código para el Ejemplo 3.1, se dan valores específicos a los parámetros para encontrar la política óptima.

En la Tabla 3.2 se muestran los costos mínimos por operar en el estado $s \in \{1, 2, 3\}$ hasta la décima iteración junto con la acción a que minimiza dichos costos. Análogamente, en la Tabla 3.3 se muestran los costos mínimos por operar en los estados $s = 4$ y $s = 5$ hasta la décima iteración junto con la acción que los minimiza.

$v_n(1)$	a	$v_n(2)$	a	$v_n(3)$	a
0	0	0	0	0	0
1.00002	0.0574786	2.00002	0.0574786	3.00002	0.0574786
1.08614	0.0574786	2.17541	0.0574786	3.27111	0.147992
1.09355	0.0574786	2.19079	0.0574786	3.29552	0.152817
1.09419	0.0574786	2.19214	0.0574786	3.29772	0.153292
1.09425	0.0574786	2.19226	0.0574786	3.29791	0.153341
1.09425	0.0574786	2.19227	0.0574786	3.29793	0.153345
1.09425	0.0574786	2.19227	0.0574786	3.29793	0.153346
1.09425	0.0574786	2.19227	0.0574786	3.29793	0.153346
1.09425	0.0574786	2.19227	0.0574786	3.29793	0.153346
1.09425	0.0574786	2.19227	0.0574786	3.29793	0.153346

Tabla 3.2: Costos de los primeros tres estados hasta la décima iteración.

Recordemos que el Teorema de PD establece que la sucesión de funciones de valor iterado converge a la función de valor óptimo (ver [3]), esto es, que

$$\lim_{n\to\infty} v_n(s) = V^*(s), \tag{3.11}$$

es decir, para todo $\varepsilon > 0$, existe $N \in \mathbb{N}$tal que si $n > N$ entonces $\mid v_n(s) - V^*(s) \mid < \frac{\varepsilon}{2}$.
Luego

$$\mid v_n(s) - v_{n-1}(s) \mid \leq \mid v_n(s) - V^*(s) \mid + \mid v_{n-1}(s) - V^*(s) \mid < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \tag{3.12}$$

Observemos que conforme n aumenta los costos $v_n(s)$ se acercan a su valor óptimo $V^*(s)$ a partir de la iteración $n \geq 4$. Ahora, si $\varepsilon = 0.01$ podemos observar de las Tablas 3.1 y 3.2 que se cumple (3.12) para este ejemplo. Del Lema 2.5, tenemos que $\hat{h}^* = \lim_{k\to\infty}(1-\alpha_k)V^*_{\alpha_k}(\hat{s})$, cuando α es cercano a 1, los minimizadores a se estabilizan a partir de $n \geq 4$.

A manera de conclusión del Ejemplo 3.1 se calculan y muestran los costos hasta la décima iteración y a partir de ésta, dichos costos se acer-

$v_n(4)$	a	$v_n(5)$	a
0	0	0	0
4.00002	0.057478	5.00002	0.057478
4.37003	0.186861	5.47376	0.186861
4.40411	0.186861	5.5184	0.186861
4.40724	0.186861	5.52259	0.186861
4.40753	0.186861	5.52298	0.186861
4.40755	0.186861	5.52302	0.186861
4.40755	0.186861	5.52302	0.186861
4.40755	0.186861	5.52302	0.186861
4.40755	0.186861	5.52302	0.186861
4.40755	0.186861	5.52302	0.186861

Tabla 3.3: $v_n(4)$ **y** $v_n(5)$ **para** $n \in \{0, \ldots 10\}$**.**

can a su valor óptimo $V^*(s)$ mostrando la acción que los minimiza en cada etapa. Para este caso, de las Tablas 3.4 y 3.5 se deduce la siguiente regla de decisión, $f^*(s) = 0.0574786$ para $s \in \{1, 2\}$, $f^*(s) = 0.153346$ para $s = 3$ y $f^*(s) = 0.18681$ para $s \in \{4, 5\}$.

3.2. El Modelo $M/M/1$ con Recesos

Los sistemas de colas con recesos han sido estudiados en [4] y [24], estos modelos se pueden clasificar en dos categorías: los modelos en donde al servidor se le permiten recesos múltiples y modelos donde el servidor puede ser extraído. En el primer caso, la duración de un receso es conducido por un proceso externo (esto es, los periodos de receso son v.a.i.i.d., ver por ejemplo, [7] y [15]), mientras que en el segundo caso, el tiempo de receso es impulsado por el proceso de llegada (ver [11]). En ambos casos, el servidor puede ser apagado en cualquier etapa de culminación de servicio de un cliente y puede ser activado sólo en una etapa de llegada de un cliente en el caso del servidor extraíble, y sólo al final de un período de receso en el caso de recesos repetidos.

3.2.1. Descripción del Modelo

Consideramos un sistema de colas $M/M/1$ en donde el servidor atiende a los clientes hasta que la cola se vacía, y después toma un receso de tiempo arbitrario. Al volver de este período, el servidor puede o bien seguir apagado o puede ser activado, siempre que la cola no se encuentre vacía. Las longitudes de estos recesos son variables aleatorias independientes e idénticamente distribuidas. Suponemos que los procesos de tiempos de interarribos, los de servicios y de tiempo de receso son independientes entre sí. Sean $1/\lambda$, $1/\mu$ y $1/\nu$ el tiempo medio de interarribos, el tiempo medio de servicio y el tiempo medio de recesos, respectivamente. Asumimos que un costo $\gamma > 0$ se incurre cada vez que el servidor se reanuda.

Para formular el problema de control como un Proceso de Decisión semi-Markoviano, tenemos que introducir más puntos de observación. Concretamente, vamos a suponer que el sistema se observa en cada arribo, en cada finalización de servicio y en cada tiempo de salto de un proceso Poisson $\mathbf{V}$, independiente de los procesos de entrada y de servicio, con parámetro ν. El proceso $\mathbf{V}$ desempeña el papel de un proceso de receso *virtual* en el sentido de que, si se produce un salto en $\mathbf{V}$ cada vez que el servidor está en reposo, entonces este salto puede tomarse como el tiempo de finalización del próximo receso. Esto se sigue de la propiedad de pérdida de memoria de la distribución exponencial, la independencia de los tiempos de llegadas, de servicios y el proceso $\mathbf{V}$, y el hecho de que el tiempo medio de interarribos para el proceso $\mathbf{V}$ es el mismo que el tiempo medio de recesos. En cada punto de observación que corresponde a un salto en $\mathbf{V}$, una decisión será tomada ya sea en reanudar o no el servidor (siempre que la cola no esté vacía). Esto se deduce de la descripción del problema de control y de la definición del proceso virtual $\mathbf{V}$. En todos los demás puntos de observación, se tomarán acciones ficticias que no afectarán el comportamiento del sistema.

El Proceso de Decisión semi-Markoviano se construye tal que el estado del proceso en la n-ésima época de decisión t_n, $n \in \mathbb{N}$, esté representado por una terna ordenada $(X_n,\ Y_n,\ Z_n) \in \mathbb{N} \times \{0,1\}^2$, donde X_n es la longitud de la cola al tiempo t_n, $Y_n \in \{0,1\}$ describe la actividad del servidor al tiempo t_n, esto es, $Y_n = 1$ si el servidor está activo y $Y_n = 0$ si está en reposo y $Z_n = I_{(t_n \in \mathbf{V})}$, es decir, $Z_n = 1$ si t_n es un tiempo de salto del proceso $\mathbf{V}$ y $Z_n = 0$, en caso contrario.

El PDSM se construye de la siguiente manera. Sea $\mathbb{S} := \mathbb{N} \times \{0,1\}^2$ el espacio de estados, y sea $\mathbb{A}(x,y,z) \subset \mathbb{A} = \{0,1\}$ el conjunto de todas las acciones disponibles cuando el sistema se encuentra en el estado $s = (x,y,z) \in \mathbb{S}$. Asumimos que

$$\mathbb{A}(x,1,1) = \{1\}, \quad x \geq 1;$$

$$\mathbb{A}(x,1,0) = \begin{cases} \{1\}, & x \geq 1; \\ \{0\}, & x = 0; \end{cases}$$

$$\mathbb{A}(x,0,1) = \begin{cases} \{0,1\}, & x \geq 1; \\ \{0\}, & x = 0; \end{cases}$$

$$\mathbb{A}(x,0,0) = \{0\}, \quad x \geq 1,$$

donde la acción 1 (respectivamente 0) se toma si la decisión es reanudar el servidor (apagar respectivamente). Notemos que los únicos estados cuando más de una acción está disponible son los estados $(x,0,1)$ con $x \geq 1$, esto es, cuando ocurre un salto en $\mathbf{V}$ $(z=1)$, el servidor está en reposo $(y=0)$ y que la cola no esté vacía $(x \geq 1)$. En los demás casos, la definición de una sola acción disponible es arbitraria.

Sea $Q(\cdot \mid s; a)$ la distribución de probabilidad del siguiente estado visitado por el sistema dado que se encuentra en el estado $s=(x,y,z) \in \mathbb{S}$ y la acción $a \in \mathbb{A}(x,y,z)$ se elige. Las probabilidades de transición están dadas por

$$\begin{aligned}
Q(x-1,1,0 \mid x,1,1;1) &= \mu/\beta, && x \geq 1;\\
Q(x+1,1,0 \mid x,1,1;1) &= \lambda/\beta, && x \geq 1;\\
Q(x,1,1 \mid x,1,1;1) &= \nu/\beta, && x \geq 1;
\end{aligned}$$

$$\begin{aligned}
Q(x-1,1,0 \mid x,1,0;1) &= \mu/\beta, && x \geq 1;\\
Q(x+1,1,0 \mid x,1,0;1) &= \lambda/\beta, && x \geq 1;\\
Q(x,1,1 \mid x,1,0;1) &= \nu/\beta, && x \geq 1;\\
Q(0,0,1 \mid 0,1,0;0) &= \nu/(\lambda+\nu);\\
Q(1,0,0 \mid 0,1,0;0) &= \lambda/(\lambda+\nu);
\end{aligned}$$

$$\begin{aligned}
Q(x-1,1,0 \mid x,0,1;1) &= \mu/\beta, && x \geq 1;\\
Q(x+1,1,0 \mid x,0,1;1) &= \lambda/\beta, && x \geq 1;\\
Q(x,1,1 \mid x,0,1;1) &= \nu/\beta, && x \geq 1;\\
Q(x+1,0,0 \mid x,0,1;0) &= \lambda/(\lambda+\nu), && x \in \mathbb{N};\\
Q(x,0,1 \mid x,0,1;0) &= \nu/(\lambda+\nu), && x \in \mathbb{N};
\end{aligned}$$

$$\begin{aligned}
Q(x+1,0,0 \mid x,0,0;0) &= \lambda/(\lambda+\nu), && x \geq 1;\\
Q(x,0,1 \mid x,0,0;0) &= \nu/(\lambda+\nu), && x \geq 1,
\end{aligned}$$

donde $\beta := \lambda + \mu + \nu$.

Definición 3.1 *Sean $s=(x,y,z) \in \mathbb{S}$ y $a \in \mathbb{A}(x,y,z)$, definamos*

$$c(s,a) := xz + \gamma(1-y)z \cdot I_{\{1\}}(a), \tag{3.13}$$

donde γ es el costo que se incurre cada vez que el servidor se reanuda y $I_{\{1\}}(a)$ es la función indicadora de $\{a = 1\}$, es decir, toma valor uno si $a = 1$ y cero en otro caso.

Distribución del tiempo de servicio.

Denotemos por $\hat{\beta}$ la tasa de servicio para cada estado s dado la acción a. Sean $s = (x, y, z) \in \mathbb{S}$ y $a \in \mathbb{A}(s)$, entonces

$$\hat{\beta}(s,a) = \begin{cases} \beta, & x \geq 1,\ y,\ z \in \{0,1\},\ a = 1; \\ \lambda + \nu, & (x,y,z) \in \mathbb{S},\ a = 0; \end{cases}$$

así,

$$H(t \mid s,a) = \begin{cases} \beta e^{-\beta t}, & x \geq 1,\ \ y, z \in \{0,1\},\ a = 1; \\ (\lambda+\nu) e^{-(\lambda+\nu)t}, & (x,y,z) \in \mathbb{S},\ a = 0, \end{cases} \tag{3.14}$$

sigue una distribución exponencial con parámetro β para $x \geq 1$ y $a = 1$, y con parámetro $\lambda + \nu$ en otro caso. De (1.6) y de (3.14) se tiene que el tiempo medio de permanencia está dado de la siguiente manera, para $s \in \mathbb{S}$,

$$\tau(s,a) = \begin{cases} \frac{1}{\beta} & a = 1; \\ \frac{1}{(\lambda+\nu)} & a = 0. \end{cases} \tag{3.15}$$

Ahora se procede a la transformación de Schweitzer, esto para analizar el caso descontado y poder asegurar la existencia de una política estacionaria óptima para el modelo.

3.2.2. Transformación del Modelo $M/M/1$ con Recesos

De (2.1) y (3.13) se obtienen la función de costos y la ley de transición para el modelo $M/M/1$ del Ejemplo 3.2,

$$\hat{c}(s,a) = \begin{cases} \beta x, & (x,1,1),\ x \geq 1; \\ \beta(x+\gamma), & (x,0,1),\ x \geq 1,\ a=1; \\ x(\lambda+\nu), & (x,0,1),\ x \geq 1,\ a=0; \\ 0 & e.o.c., \end{cases} \tag{3.16}$$

recordando de (2.1) que, $\rho \in (0,\theta)$, con θ como en la Condición 2.1. Se elige $\rho = \frac{1}{\lambda+\mu+\nu+1}$, así,

$$\begin{aligned} \hat{Q}(x-1,1,0 \mid x,1,1;1) &= \rho\mu, && x \geq 1; \\ \hat{Q}(x+1,1,0 \mid x,1,1;1) &= \rho\lambda, && x \geq 1; \\ \hat{Q}(x,1,1 \mid x,1,1;1) &= \rho\nu + (1-\rho\beta), && x \geq 1; \end{aligned}$$

$$\begin{aligned} \hat{Q}(x-1,1,0 \mid x,1,0;1) &= \rho\mu, && x \geq 1; \\ \hat{Q}(x+1,1,0 \mid x,1,0;1) &= \rho\lambda, && x \geq 1; \\ \hat{Q}(x,1,1 \mid x,1,0;1) &= \rho\nu, && x \geq 1; \\ \hat{Q}(0,0,1 \mid 0,1,0;0) &= \rho\nu; \\ \hat{Q}(1,0,0 \mid 0,1,0;0) &= \rho\lambda; \end{aligned}$$

$$\begin{aligned} \hat{Q}(x-1,1,0 \mid x,0,1;1) &= \rho\mu, && x \geq 1; \\ \hat{Q}(x+1,1,0 \mid x,0,1;1) &= \rho\lambda, && x \geq 1; \\ \hat{Q}(x,1,1 \mid x,0,1;1) &= \rho\nu, && x \geq 1; \\ \hat{Q}(x+1,0,0 \mid x,0,1;0) &= \rho\lambda, && x \in \mathbb{N}; \\ \hat{Q}(x,0,1 \mid x,0,1;0) &= \rho\nu + (1-\rho(\lambda+\nu)), && x \in \mathbb{N}; \end{aligned}$$

$$\begin{aligned} \hat{Q}(x+1,0,0 \mid x,0,0;0) &= \rho\lambda, && x \geq 1; \\ \hat{Q}(x,0,1 \mid x,0,0;0) &= \rho\nu, && x \geq 1. \end{aligned}$$

Sea $(x,y,z) \in \mathbb{S},\ x \geq 1$ y $a \in \mathbb{A}(x,y,z)$, a partir del Lema 2.2 y del Teorema 4.1 en [3] tenemos que, para $\alpha \in (0,1)$

$$V(x,1,1) = \min_{a\in\mathbb{A}(x,1,1)} \Big\{\hat{c}(x,1,1;a) + \alpha \sum_{(x',y',z')\in\mathbb{S}} V(x',y',z')\hat{Q}(x',y',z' \mid x,1,1;a)\Big\}$$

como $a \in \mathbb{A}(x,1,1) = \{1\}$, para $x \geq 1$, entonces de (3.16) tenemos que

$$\hat{c}(x,1,1;1) = \beta x,$$

así

$$V(x,1,1) = \beta x + \alpha\Big[V(x-1,1,0)\rho\mu + V(x+1,1,0)\rho\lambda + V(x,1,1)(\rho\nu + 1 - \rho\beta)\Big], \tag{3.17}$$

con $\beta := \lambda + \mu + \nu$. Similarmente se deducen las siguientes ecuaciones de optimalidad.

$$V(x,1,0) = \alpha\mu\rho V(x-1,1,0) + \alpha\lambda\rho V(x+1,1,0) + \alpha\nu\rho V(x,1,1), \quad \text{para} \ \ x \geq 1; \tag{3.18}$$

$$V(0,1,0) = \alpha\lambda\rho V(1,0,0) + \alpha\nu\rho V(0,0,1); \tag{3.19}$$

$$V(0,0,1) = \alpha\lambda\rho V(1,0,0) + \alpha\nu\rho V(0,0,1) + \alpha[1-\rho(\lambda+\nu)]V(0,0,1); \tag{3.20}$$

para $x \geq 1$,

$$V(x,0,1) = \min\Big\{(\lambda+\nu)x + \alpha\lambda\rho V(x+1,0,0) + \alpha\nu\rho V(x,0,1) + \alpha[1-\rho(\lambda+\nu)]V(x,0,1); \beta(x+\gamma) + \alpha\mu\rho V(x-1,1,0) + \alpha\lambda\rho V(x+1,1,0) + \alpha\nu\rho V(x,1,1)\Big\}; \tag{3.21}$$

$$V(x,0,0) = \alpha\lambda\rho V(x+1,0,0) + \alpha\nu\rho V(x,0,1) \ \text{ para } x \geq 1. \tag{3.22}$$

A continuación se reducirá el número de incógnitas involucradas en el conjunto de ecuaciones (3.17)-(3.22).

Primero, combinando (3.19) y (3.20)

$$V(0,0,1) = V(0,1,0) + \alpha[1-\rho(\lambda+\nu)]V(0,0,1), \tag{3.23}$$

despejando $V(0,1,0)$ de (3.23)

$$\Big(1-\alpha[1-\rho(\lambda+\nu)]\Big)V(0,0,1) = V(0,1,0). \tag{3.24}$$

Sustituyendo (3.24) en (3.19)

$$V(1,0,0) = \frac{1-\alpha+\alpha\lambda\rho}{\alpha\lambda\rho}V(0,0,1). \tag{3.25}$$

Ahora, sustituyendo (3.18) y (3.22) en (3.21) se obtiene

$$\begin{aligned} V(x,0,1) = \min\Big\{&(\lambda+\nu)x + V(x,0,0) + \alpha(1-\rho(\lambda+\nu))V(x,0,1);\\ &\beta(x+\gamma) + V(x,1,0)\Big\} \ \text{ para } \ x \geq 1. \end{aligned} \tag{3.26}$$

Introduciendo (3.26) en (3.22) para $x \geq 1$

$$\begin{aligned} V(x,0,0) =& \alpha\lambda\rho V(x+1,0,0) + (\alpha\rho\nu)\min\Big\{(\lambda+\nu)x + V(x,0,0)\\ &+ \alpha(1-\rho(\lambda+\nu))V(x,0,1); \beta(x+\gamma) + V(x,1,0)\Big\}. \end{aligned} \tag{3.27}$$

Observemos de (3.17) y (3.18) que

$$V(x,1,1) = \beta x + V(x,1,0) + (1-\rho\beta)V(x,1,1) \ \text{ para } \ x \geq 1,$$

de aquí que

$$\rho\beta V(x,1,1) = \beta x + V(x,1,0). \tag{3.28}$$

Finalmente, introduciendo (3.28) en (3.18), para $x \geq 1$

$$V(x,1,0) = \alpha\mu\rho V(x-1,1,0) + \alpha\lambda\rho V(x+1,1,0) + \frac{\alpha\nu}{\beta}\Big[\beta x + V(x,1,0)\Big]. \tag{3.29}$$

La expresión (3.29) define una ecuación en diferencias de orden dos, quedando como sigue

$$\left(1-\frac{\alpha\nu}{\beta}\right)V(x,1,0)=\alpha\nu x+\alpha\lambda\rho V(x+1,1,0)+\alpha\mu\rho V(x-1,1,0). \tag{3.30}$$

Resolviendo la ecuación en diferencias (3.30). Primeramente veamos que la solución general a la ecuación (3.30) está dada de la siguiente forma

$$V(x,1,0)=a\beta_1^x+b\beta_2^x+cx+d, \tag{3.31}$$

con β_1 y β_2 raíces del polinomio característico (en z)

$$-(\alpha\lambda\rho)z^2+\left(1-\frac{\alpha\nu}{\beta}\right)z-\alpha\mu\rho,$$

con $0<\beta_1<1<\beta_2$. La ecuación (3.30) se puede ver como

$$-\alpha\lambda\rho V(x+1,1,0)+\left(1-\frac{\alpha\nu}{\beta}\right)V(x,1,0)-\alpha\mu\rho V(x-1,1,0)=\alpha\nu x. \tag{3.32}$$

Encontremos la solución a la ecuación homogénea

$$-\alpha\lambda\rho V(x+1,1,0)+\left(1-\frac{\alpha\nu}{\beta}\right)V(x,1,0)-\alpha\mu\rho V(x-1,1,0)=0. \tag{3.33}$$

La ecuación característica de (3.33) está dada de la siguiente forma

$$-\alpha\lambda\rho z^2+\left(1-\frac{\alpha\nu}{\beta}\right)z-\alpha\mu\rho=0. \tag{3.34}$$

Supóngase que β_1 y β_2 son soluciones a la ecuación (3.34), luego la solución general de (3.33) es

$$V(x,1,0)=a\beta_1^x+b\beta_2^x. \tag{3.35}$$

Para el caso no-homogéneo, se propone una solución particular, digamos

$$V(x,1,0)=cx+d, \tag{3.36}$$

sustituyendo (3.36) en (3.32) se tiene que

$$-\alpha\lambda\rho\big(c(x+1)+d\big)+\left(1-\frac{\alpha\nu}{\beta}\right)(cx+d)-\alpha\mu\rho\big(c\big(x-1)+d\big)=\alpha\nu x, \tag{3.37}$$

de (3.37) se obtienen

$$cx\left(\left(1-\frac{\alpha\nu}{\beta}\right)-\alpha\lambda\rho-\alpha\mu\rho\right)=\alpha\nu x,$$

y

$$-\alpha\lambda\rho(c+d)+\left(1-\frac{\alpha\nu}{\beta}\right)d+\alpha\mu\rho(c-d)=0.$$

Despejando se llega a que

$$c=\frac{\alpha\nu}{1-\alpha\left(\frac{\nu}{\beta}+\lambda\rho+\mu\rho\right)}, \tag{3.38}$$

$$d=\frac{\alpha^2\nu\rho(\lambda-\mu)}{1-\alpha\left(\frac{\nu}{\beta}+\lambda\rho+\mu\rho\right)}. \tag{3.39}$$

Así, la solución general de (3.32) es la suma de (3.35) y (3.36), lo cual se llega a (3.31).

Observemos que $\frac{V(x,1,0)}{x}$ es uniformemente acotada en $\mathbb{N}$, podemos ver de (3.31) que necesariamente $b=0$ pues $\beta_2>1$. Para encontrar el coeficiente a usamos como condición inicial $x=0$, de (3.31) se tiene

$$V(0,1,0)=a+\frac{\alpha^2\nu\rho(\lambda-\mu)}{1-\alpha\left(\frac{\nu}{\beta}+\lambda\rho+\mu\rho\right)}, \tag{3.40}$$

luego sustituyendo (3.24) en (3.40) se llega a que

$$a=\Big(1-\alpha[1-\rho(\lambda+\nu)]\Big)V(0,0,1)+\frac{\alpha^2\nu\rho(\mu-\lambda)}{1-\alpha\left(\frac{\nu}{\beta}+\lambda\rho+\mu\rho\right)}.$$

Finalmente

$$\begin{aligned}V(x,1,0)=&\Big(1-\alpha[1-\rho(\lambda+\nu)]\Big)V(0,0,1)+\frac{\alpha^2\nu\rho(\mu-\lambda)}{1-\alpha\left(\frac{\nu}{\beta}+\lambda\rho+\mu\rho\right)}\beta_1^x\\&+\frac{\alpha\nu}{1-\alpha\left(\frac{\nu}{\beta}+\lambda\rho+\mu\rho\right)}x+\frac{\alpha^2\nu\rho(\lambda-\mu)}{1-\alpha\left(\frac{\nu}{\beta}+\lambda\rho+\mu\rho\right)}.\end{aligned} \tag{3.41}$$

En la siguiente sección se presentan los resultados obtenidos al hacer la programación de los costos $V_n(x,y,z)$ obtenidos en esta sección, esto con la ayuda del software *Mathematica 8.0*.

3.2.3. Aproximación Numérica al Valor Óptimo Promedio del Modelo $M/M/1$ con Recesos

La matriz que se muestra a continuación contiene los costos $V_n(x,y,z)$ ordenados de la siguiente manera

$$\begin{pmatrix} V_n(1,0,0) & V_n(1,1,0) & V_n(1,0,1) & V_n(1,1,1) \\ V_n(2,0,0) & V_n(2,1,0) & V_n(2,0,1) & V_n(2,1,1) \\ \vdots & \vdots & \vdots & \vdots \\ V_n(m,0,0) & V_n(m,1,0) & V_n(m,0,1) & V_n(m,1,1) \end{pmatrix}.$$

Las Tablas 3.4 y 3.5 muestran los costos mínimos $V_n(x,y,z)$ en forma de matriz como se dió anteriormente, al programarlos en el software *Mathematica 8.0*, esto para $n=\{1,2,5,10\}$, y se le dan valores específicos a los parámetros $\alpha = 1/2,\ \lambda = 1/5,\ \mu = 1/2,\ \nu = 1/2,\ \gamma = 1$ y $m = 10$.

$n=1$				$n=2$			
0.702	0.3965	1.000	1.418	1.362	0.413	1.416	1.413
1.200	1.902	2.000	3.402	2.361	1.952	2.833	3.402
2.650	2.900	3.000	5.400	3.353	2.950	4.250	5.400
3.203	3.900	4.000	7.400	4.345	3.950	5.667	7.400
4.527	4.900	5.000	9.400	5.337	4.950	7.083	9.400
5.965	5.900	6.000	11.400	6.329	5.950	8.500	11.400
6.470	6.900	7.000	13.400	7.320	6.950	9.967	13.400
7.860	7.900	8.000	15.400	8.314	7.950	11.333	15.400
8.400	8.900	9.000	17.400	9.306	8.950	12.750	17.400
9.430	9.900	10.000	19.400	10.298	9.950	14.167	19.400

Tabla 3.4: Costos $V_n(x,y,z)$ de las primeras dos iteraciones.

Del Teorema de PD se tiene que la sucesión de funciones de valor iterado converge a la función de valor óptimo, ver (3.11), esto es, para todo $\varepsilon > 0$, existe $N \in \mathbb{N}$ tal que si $n > N$ entonces $\mid V_n(s) - V^*(s) \mid < \frac{\varepsilon}{2}$.

Luego

$$| V_n(s) - V_{n-1}(s) | \leq | V_n(s) - V^*(s) | + | V_{n-1}(s) - V^*(s) | < \varepsilon. \quad (3.42)$$

Note que cuando n aumenta, los costos $V_n(s)$ se acercan a $V^*(s)$. Ahora, si $\varepsilon = 0.01$ podemos observar de las Tablas 3.4 y 3.5 que se cumple (3.42) para este ejemplo.

$n=5$				$n=10$			
1.392	0.416	2.105	1.416	1.392	0.419	2.105	1.416
2.391	1.956	3.988	3.400	2.391	1.960	3.988	3.400
3.389	2.953	5.840	5.000	3.389	2.953	5.840	5.400
4.387	3.956	7.758	7.400	4.387	3.960	7.758	7.000
5.386	4.958	9.644	9.400	5.386	4.960	9.644	9.400
6.384	5.954	11.566	11.400	6.384	5.954	11.566	11.400
7.380	6.951	13.108	13.000	7.380	6.953	13.108	13.000
8.381	7.952	15.290	15.400	8.381	7.954	15.290	15.000
9.379	8.953	17.193	17.400	9.379	8.958	17.193	17.400
10.377	9.958	19.635	19.400	10.377	9.960	19.635	19.000

Tabla 3.5: Costos $V_n(x, y, z)$ de la quinta y décima iteración.

Elegir, por ejemplo, $s_1 = (3, 0, 0)$, de las tablas podemos observar que $V_1(s_1) = 2.650$, $V_2(s_1) = 3.353$, $V_5(s_1) = 3.389$, $V_{10}(s_1) = 3.389$, así que, conforme n aumenta, los costos $V_n(s)$ van acercándose a $V(s)$, si $\varepsilon = 0.01$ podemos ver que se cumple (3.42). Análogamente se analizan los demás estados. De las Tablas 3.4 y 3.5 se observa que a partir de la décima iteración los costos se estabilizan a un cierto valor para cada estado $(x, y, z) \in \mathbb{S}$, esto es, que se acercan a su valor óptimo.

Capítulo 4

Conclusiones

En esta tesis se trabajó con la teoría de los Procesos de Decisión Semi-Markovianos. El criterio que se utilizó para evaluar la calidad de las políticas fue el criterio conocido como razón de costo promedio.

En el trabajo se consideran dos ejemplos en el cual el espacio de estados es numerable para ambos casos y el espacio de acciones es de Borel para el primero de ellos y en el segundo ejemplo se considera el espacio de acciones compacto.
En el Capítulo 1 se presentó la teoría de los PDSM junto con el criterio de optimalidad con el cual se trabajó. En el Capítulo 2 se dieron condiciones necesarias para resolver el problema de control óptimo y se presenta el Teorema de Programación Dinámica que nos garantiza la existencia de una política óptima estacionaria de costo promedio. En el Capítulo 3, se ejemplificó la Teoría de los PDSM en dos modelos de líneas de espera controladas: en el primero de ellos se controla el parámetro de servicio, el espacio de estados es numerable y el espacio de acciones es un espacio de Borel, se garantizó la existencia de una política estacionaria óptima de costo promedio; para el segundo ejemplo se consideró un modelo de líneas de espera $M/M/1$ con periodos de recesos repetidos en el servidor, se garantiza la existencia de una política que minimiza el criterio de rendimiento. Se le asignaron valores concretos a los parámetros para

mostrar los costos mínimos por operar en cierto estado del sistema, esto se hizo con la ayuda del software *Mathematica 8.0*.

Es importante hacer mención que en la mayoría de la literatura relacionada con los PDSM sólo se estudia la parte teórica, es decir, se dan condiciones necesarias para resolver el problema de control y se garantiza una política óptima estacionaria gracias al Teorema de Programación Dinámica. En este trabajo se estudia la teoría que está basada en [25] para garantizar políticas óptimas estacionarias, además de dar dos ejemplos de líneas de espera en el contexto de la teoría de los PDSM que cumplan con las condiciones presentadas en el Capítulo 2 y garantizar la existencia de políticas óptimas estacionarias, estos ejemplos se resuelven numéricamente para encontrar los costos mínimos por operar en cierto estado y acción, esto con la ayuda del software *Mathematica 8.0*, para valores específicos en los parámetros.

Como trabajo a futuro se hacen las siguientes propuestas:

(a) Considerar el criterio de rendimiento conocido como *time average cost* y hacer un estudio analítico sobre este criterio.

(b) Hacer un análisis comparativo entre ambos criterios, *razón de costo promedio* y *time average cost*, a su vez, observar cómo se comportan los costos para ambos principios.

(c) Proporcionar las condiciones necesarias para probar que ambos criterios son equivalentes en el sentido de que ambos conducen a la misma ecuación de optimalidad.

Apéndice A

Resultados Auxiliares

A.1. Definiciones

Definición A.1 *Sea* (X, τ) *un espacio topológico, la mínima* σ*-álgebra de subconjuntos de* X *que contiene a todos los conjuntos abiertos en* X *es la* σ*-álgebra de Borel, es decir, la* σ*-álgebra generada por* τ*. La denotaremos por* $\mathfrak{B}(X)$*.*

Así, cuando se hable de conjuntos o funciones medibles, se entenderán como *Borel-medibles*.

Definición A.2 X *es un espacio de Borel, si* X *es un subconjunto de Borel de un espacio métrico separable y completo. Un subconjunto de Borel de un espacio de Borel es por sí mismo un espacio de Borel.*

Definición A.3 *Un kérnel estocástico definido sobre* $\mathbb{X}$ *dado* $\mathbb{Y}$ *es una función* $Q(\cdot \mid \cdot)$ *tal que:*

1. $Q(\cdot \mid y)$ *es una medida de probabilidad en* $\mathbb{X}$ *para cada* $y \in \mathbb{Y}$*.*
2. $Q(B \mid \cdot)$ *es una variable aleatoria (función medible) en* $\mathbb{Y}$ *para cada* $B \in \mathfrak{B}(\mathbb{X})$*.*

Definición A.4 *Sea* X *un espacio métrico y* v *una función de* X *a* $\mathbb{R} \cup \{+\infty\}$ *tal que* $v(x) < \infty$ *para al menos un punto* $x \in X$*.*

v se dice semicontinua inferiormente (l.s.c. por sus siglas en inglés), en el punto $x \in X$ si

$$\liminf_{n\to\infty} v(x_n) \geq v(x),$$

para cualquier sucesión $\{x_n\}$ en X que converja a x. La función v se dice l.s.c si lo es en cada punto de X.

Definición A.5 *Una **medida de Dirac** es una medida δ_x sobre un conjunto X definida para un $x \in X$ y cualquier conjunto (medible) $A \subset X$ por*

$$\delta_x(A) = I_A(x) = \begin{cases} 0 & x \notin A, \\ 1 & x \in A, \end{cases}$$

donde I_A es la función indicadora de A.

A.2. Teoremas Auxiliares

Teorema A.1 ***(de C. Ionescu-Tulcea)***

Sea $\mathbb{X}_0, \mathbb{X}_1, \ldots$ una sucesión de espacios de Borel y, para cada $n \in \mathbb{N}$, defínase

$$\mathbb{Y}_n := \mathbb{X}_0 \times \mathbb{X}_1 \times \cdots \times \mathbb{X}_n$$

y

$$\mathbb{Y} := \mathbb{X}_0 \times \mathbb{X}_1 \times \cdots .$$

Sea v una medida de probabilidad arbitraria en $\mathbb{X}_{n+1}$ y, para cada $n \in \mathbb{N}$, sea $Q_n(dx_{n+1} \mid y_n)$ un kérnel estocástico sobre $\mathbb{X}_{n+1}$ dado $\mathbb{Y}_n$. Entonces existe una única medida de probabilidad Q_v sobre $\mathbb{Y}$ tal que, para cada rectángulo medible $B_0 \times \cdots \times B_n$ en $\mathbb{Y}_n$,

$$Q_v(B_0 \times \cdots \times B_n) = \int_{B_0} v(dx_0) \int_{B_1} Q_0(dx_1 \mid x_0) \int_{B_2} Q_1(dx_2 \mid x_0, x_1) \cdots \int_{B_n} Q_{n-1}(dx_n \mid x_0, \ldots, x_{n-1}).$$

Más aún, para cada función medible no-negativa u en $\mathbb{Y}$, la función

$$x \longmapsto \int u(y) Q_x(dy)$$

es medible sobre $\mathbb{X}_0$*, donde* Q_x *denota a* Q_v *cuando* v *es la probabilidad concentrada en* $x \in \mathbb{X}_0$*.*

Demostración. Ver [1]. □

Teorema A.2 ***(de la Convergencia Monótona)***

Sea $(X, \mathfrak{B}, \mu)$ *un espacio medible y* $g_1, g_2, \ldots, g, h$ *funciones medibles*

(a) Si $g_n \geq h$ *para toda* n*, donde* $\int_X h d\mu > -\infty$*, y* $g_n \uparrow g$*, entonces*

$$\int_X g_n d\mu \uparrow \int_X g d\mu.$$

(b) Si $g_n \leq h$ *para toda* n*, donde* $\int_X h d\mu < \infty$*, y* $g_n \downarrow g$*, entonces*

$$\int_X g_n d\mu \downarrow \int_X g d\mu.$$

Demostración. Para una demostración ver [1]. □

Lema A.1 ***(una extensión del Lema de Fatou)***

Sea $\{u_n\}$ *un sucesión acotada en* $\mathbb{B}_w(\mathbb{S})$*, esto es, existe una constante* K *tal que* $\|u_n\|_w \leq K$ *para toda* n*, y defina*

$$u^I(s) := \liminf_{n\to\infty} u_n(s), \quad y \quad u^S(s) := \limsup_{n\to\infty} u_n(s).$$

Entonces para cada $s \in \mathbb{S}$ *y cualquier sucesión* $\{a^n\}$ *en* $\mathbb{A}(s)$ *tal que* $a^n \to a$ *en* $\mathbb{A}(s)$*, tenemos*

$$\liminf_{n\to\infty} \int u_n(s')\hat{Q}(ds'|s, a^n) \geq \int u^I(s')\hat{Q}(ds'|s, a),$$

y

$$\limsup_{n\to\infty} \int u_n(s')\hat{Q}(ds'|s, a^n) \leq \int u^S(s')\hat{Q}(ds'|s, a).$$

Por tanto, si $u_n \to u$ *(esto es,* $u^I = u^S$*), entonces*

$$\lim_{n\to\infty} \int u_n(s')\hat{Q}(ds'|s, a^n) = \int u(s')\hat{Q}(ds'|s, a).$$

Demostración. Ver [10], (pág. 48) □

Apéndice B

Códigos

Los siguientes códigos se realizaron en el programa *Wolfram Mathematica 8.0*

B.1. Código Para el Ejemplo 3.1

s=50;
$\alpha = \mathrm{RandomReal}[1];$
$\lambda = \mathrm{RandomReal}[s];$
$\alpha = \mathrm{RandomReal}[1];$
$\mu_2 = \mathrm{RandomReal}\left[\frac{3}{2(e-1)\lambda}\right];$
$\mu_1 = \mathrm{RandomReal}\left[\mu_2\right];$
$\rho = \mathrm{RandomReal}\left[\frac{\mu_1}{2}\right];$
$v_0 = \mathrm{Table}[0, \{i, 1, s\}];$
$m = 100;$
$Do[$
$\{$
$v_n(\mathrm{x\_}, \mathrm{a\_}) =$
$x + a^2 +$
$\alpha \sum_{y=1}^{s} \Big(v_{n-1}[[y]]\Big((\lambda * \rho)\left(\frac{1-\exp(a(-\lambda))}{a*\lambda} - \sum_{k=1}^{y-x+1} \frac{\exp(a(-\lambda))(a*\lambda)^{k-1}}{k!}\right) +$
$v_n = \mathrm{Table}\left[\mathrm{Minimize}\left[\{v_n[i, a], \mu_1 \leq a \leq \mu_2\}, a\right][[1]], \{i, 1, s\}\right]$

$\}$,

$\{n, 1, m\}]$

$\mathrm{Do}\left[\mathrm{Print}\left[v_i\right], \{i, 0, m\}\right]$

Bibliografía

[1] Ash, R.B. y Doléans-Dade C. A., *"Probability and measure theory"*. Academic Press Elsevier. 2005.

[2] Bellman, R. E., *"Dynamic programming"*. Dover Publications. 2003.

[3] Camilo, C., *"Análisis de una línea de espera usando procesos de decisión semi-markovianos"*. Tesis licenciatura. Facultad de Ciencias Físico Matemáticas, BUAP. 2014.

[4] Doshi, B., *"Queueing systems with vacations - a survey"*. Queueing Syst. **1**, 29-66. 1986.

[5] Federgruen, A., Schweitzer, P. J. and Tijms, H. C., *"Denumerable undiscounted semi-Markov decision processes with unbounded rewards"*. Math. Oper. Res. **8**, 298-313. 1983.

[6] Feinberg, E. A., *"Constrained semi-Markov decision processes with average rewards"*. Math. Methods Oper. Res. **39**, 257-288. 1994.

[7] Gelenbe, E. y Mitrani, I., *"Analysis and synthesis of computer systems"*. Academic Press. 1980.

[8] Guo, X. P. y Zhu, Q. X., *"Average optimality for Markov decision processes in Borel spaces: a new condition and approach"*. J. Appl. Probab. **43**, 318-334. 2006.

[9] Hernández, O. y Lasserre, J. B., *"Discrete-time Markov control processes: Basic optimality criteria"*. Springer-Verlag. 1996.

[10] Hernández, O. y Lasserre, J. B., *"Further topics on discrete-time Markov control processes"*. Springer-Verlag. 1999.

[11] Heyman, D.P. y Sobel, M.J., *"Stochastic models in operations research, stochastic optimization"*. McGraw-Hill Book Company. 1984.

[12] Howard, R. A., *"Semi-Markovian decision processes"*. Proc. Intern. Stat. Inst. 1963.

[13] Jaśkiewicz, A., *"On the equivalence of two expected average cost criteria for semi-Markov control processes"*. Math. Oper. Res. **29**, 326-338. 2004.

[14] Jaśkiewicz, A., *"A fixed point approach to solve the average cost optimality equation for semi-Markov decision processes with Feller trasition probabilities"*. Commun. Stat. Theory Methods **36**, 2559-2575. 2007.

[15] Kella, O., *"Optimal control of the vacation scheme in an M/G/1 queue"*. Oper. Res. **38**, 724-728. 1988.

[16] Luque, F. y Hernández, O., *"Semi-Markov control models with average costs"*. Appl. Math. **26**, 315-331. 1999.

[17] Meyn, S. P. y Tweedie, R. L., *"Computable bounds for geometric convergence rates of Markov chains"*. Ann. Appl. Probab. **4**, 981-1011. 1994.

[18] Mine, H. y Osaki, S., *"Markovian Decision Processes"*. Elsevier. 1970.

[19] Puterman, M. L., *"Markov decision processes: Discrete stochastic dynamic programming"*. Wiley. 1994.

[20] Ross, S. M., *"Applied probability models with optimization Applications"*. Holden-Day. 1970.

[21] Schäl, M., *"On the second optimality equation for semi-Markov decision models"*. Math. Oper. Res. **17**, 470-486. 1992.

[22] Schweitzer, P. J., *"Iterative solution of the functional equations of undiscounted Markov renewal programming"*. J. Math. Anal. Appl. **34**, 495-501. 1971.

[23] Sennot, L. I., *"Stochastic dynamic programming and the control of queueing systems"*. Wiley. 1999.

[24] Teghem, J., *"Control of the service process in a queueing system"*. European J. Op. Res. **23**, 141-158. 1986.

[25] Wei, Q. y Guo, X., *"New average optimality conditions for semi-Markov decision processes in borel spaces"*. J. Optim. Theory Appl. **153**, 709-732. 2012.