Benemérita Universidad Autónoma de Puebla

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS
LICENCIATURA EN MATEMÁTICAS APLICADAS

CAMINATA ALEATORIA DE LINDLEY EN PROCESOS DE DECISIÓN DE MARKOV: CASO DESCONTADO Y CASO PROMEDIO

TESIS

QUE PARA OBTENER EL TÍITULO DE
LICENCIADO EN MATEMÁTICAS APLICADAS

PRESENTA
RUBEN BLANCAS RIVERA

DIRECTOR DE TESIS
DR. HUGO ADÁN CRUZ SUÁREZ

PUEBLA, PUE. 1 de Septiembre del 2016

Dedicado a mis padres, Liliana y Rubén

Agradecimientos

En primer lugar, agradezco a mis padres Liliana y Rubén, y mi abuelita Socorro que siempre me apoyaron incondicionalmente en la parte moral y económica para poder alcanzar este objetivo.
A mis hermanas Daniela y Pamela, mis mejores amigas que a pesar de la distancia las risas siempre existirán en los pocos momentos que nos vemos.

A mis amigos de la facultad con los cuales pase horas de estudio y momentos agradables, en especial a Isa, por toda la ayuda que me ha brindado, por estar a mi lado inclusive en los momentos y situaciones dificiles siempre motivandome a dar lo mejor de mi. A Tepox y Rocio, porque fueron mis compañeros, amigos y hermanos en los 5 años de Licenciatura.

A mi asesor de tesis, Dr. Hugo Adán Cruz Suárez, el cual tengo un infinito respeto y admiración por su labor en la Facultad, gracias por su apoyo incondicional.
A mi tutor académico, Dra. Lidia A. Hernández Rebollar que desde el primer semestre me ha guiado y aconsejado a lo largo de la licenciatura.

A mis sinodales, Dr. Francisco Tajonar Sanabria, Dra. Hortensia J. Reyes Cervantes y Dr. Fernando Velasco Luna por haber aceptado revisar este trabajo y aportar su conocimiento para la mejora del mismo.

A la Vicerrectoría de Investigación y Estudios de Posgrado (VIEP), por el apoyo en la elaboración de esta tesis.
Gracias.

Introducción

La presente tesis pertenece al área de Procesos Estocásticos y Teoría de Control, específicamente a los Procesos de Decisión de Markov (PDM) a tiempo discreto. Un PDM es utilizado para modelar un sistema que es observado de forma discreta en el tiempo y el cual cuenta con la propiedad de Markov. El trabajo se centra en desarrollar la teoría básica de PDM para el estudio de la caminata aleatoria de Lindley, la cual tiene diversas aplicaciones en las áreas de líneas de espera e inventarios (véase [3] y [15]).
De manera general, un PDM modela un sistema dinámico cuyos estados son observados de manera periódica y es aplicando un control. El desarrollo de un PDM, a través del tiempo, está dado de acuerdo al siguiente procedimiento. En cada tiempo t, $t = 0, 1, ...$, se elige un control que se aplicará dependiendo del estado del sistema. Entonces como consecuencia del estado actual y de haber aplicado un control se paga un costo y el sistema se traslada a un nuevo estado en el instante de tiempo $t+1$, mediante una ley de transición. Al ocurrir un estado en $t+1$ el proceso se repite. De esta manera se obtiene una sucesión de controles a la cual se le denomina política. Con la finalidad de medir la calidad de una política, el PDM está dotado de una función real llamada criterio de rendimiento. Éste trabajo se enfoca en dos tipos de criterios, el criterio descontado y el criterio promedio. El problema de control óptimo consiste en encontrar una política que minimice el criterio de rendimiento, a tal política se le llama óptima, y al criterio de rendimiento evaluado en la política óptima se le llama función de valor.
En el trabajo de tesis se presenta el desarrollo teórico de los criterios descontado y promedio para el caso de Borel. Posteriormente la teoría desarrollada es aplicada a un modelo específico, denominado caminata aleatoria de Lindley. Este modelo se ecuentra motivado en problemas de inventarios, en este contexto los estados representan el nivel de existentes (stock) y las acciones se consideran como la cantidad de productos solicitados. Debido a los costos que se manejan (costos de producción y almacenaje) es necesario para determinar una política (o estrategia) de operación, la cual minimice los costos globales dada por un criterio de rendimiento. En este trabajo se estudia la caminata aleatoria de Lindley bajo los criterios descontado y promedio. En ambos criterios se verifican condiciones para garantizar la existencia de estrategias de operación y se valida el algoritmo

de programación dinámica. La tesis está estructurada de la siguiente manera.
En el Capítulo 1, se definen conceptos básicos de modelos de control, como son políticas y los criterios de rendimientos que son estudiados en esta tesis.
Una vez establecida la construcción de un PDM a tiempo discreto, en el Capítulo 2 se presentan las hipótesis requeridas y el teorema principal del Capítulo, el cual garantiza la existencia de políticas óptimas estacionarias. Los temas presentados en estos capítulos estan basados en [6].
En el Capítulo 3, con los resultados obtenidos en los PDM con criterio de costo descontado se demuestra la existencia de una política óptima estacionaria que minimiza el criterio de costo promedio. Este Capítulo está basado en [16].
Así, en el Capítulo 4 se presenta el modelo propuesto por David Lindley [9]. Primeramente se analizan el caso clásico y posteriormente se observa la necesidad de considerar variables de control, de este modo se propone un modelo controlado de Lindley. Para dicho modelo se verifican las hipótesis dadas en los Capítulos 2 y 3, las cuales garantizan la existencia de políticas óptimas para los dos criterios estudiados. Finalmente, se presenta un ejemplo numérico donde se determinan las funciones de valores óptimos para ambos criterios, dicho algoritmo se elaboro en el software Mathematica 10.1.

Índice general

Capítulo 1

Preliminares

Se presentan conceptos básicos de modelos de control de Markov y los criterios de rendimiento que en este trabajo se estudian.

1.1. Definición del Modelo de Control de Markov

Definición 1.1.1 *Un modelo de control de Markov (MCM), estacionario a tiempo discreto, consiste de una quíntupla:*

$$(X, A, \{A(x)|x \in X\}, Q, c)$$

donde:

(a) X *es un espacio de Borel, llamado espacio de estados;*

(b) A *es un espacio de Borel llamado espacio de acciones o controles;*

(c) $\{A(x)|x \in X\}$ *es una familia de subconjuntos medibles, no vacíos* $A(x)$ *de* A*, donde* $A(x)$ *denota el conjunto de acciones admisible cuando el sistema se encuentra en el estado* $x \in X$*. El conjunto de parejas ordenadas acción y acción admisibles,* $\mathbb{K} := \{(x, a)|x \in X, a \in A(x)\}$*, se supone que es un conjunto medible del espacio producto* $X \times A$*;*

(d) Q *es un kérnel estocástico definido en* X *dado* $\mathbb{K}$ *(ver Definición D.0.5), llamado ley de transición, es decir, para cada* $(x, a) \in \mathbb{K}$*,* $Q(\cdot|x, a)$ *es una medida de probabilidad en* X*, y para cada* $B \subset X$ *medible,* $Q(B|\cdot)$ *es una función medible;*

(e) $c : \mathbb{K} \longrightarrow \mathbb{R}$ *es una función medible y se llama la función costo en un paso.*

La dinámica que describe a este sistema estocástico funciona de la siguiente forma. Si el sistema al tiempo $t \geq 1$, se encuentra en el estado $x_t = x \in X$, y la acción $a_t = a \in A(x)$ es aplicada entonces:
(i) Se paga un costo $c(x, a)$.
(ii) El sistema se traslada a un nuevo estado x_{t+1} mediante la ley de transición $Q(\cdot|x, a)$ sobre X.
Una vez hecha está transición a un nuevo estado, se elige una nueva acción y la dinámica anteriormente descrita se repite.

Hipótesis 1.1.1 *Supóngase que $\mathbb{K}$ contiene la gráfica de una función medible de X a A, es decir, existe una función $f : X \to A$ medible, tal que $f(x) \in A(x)$, para toda $x \in X$. El conjunto de estas funciones es denotado por $\mathbb{F}$ y sus elementos son llamados selectores de la multifunción $x \to A(x)$ (ver Apéndice B).*

1.2. Políticas

Primero definamos el espacio de historias observadas en un proceso de control hasta un tiempo t, denotado por $\mathbb{H}_t$, como:

$$\mathbb{H}_0 = X,$$

$$\mathbb{H}_t = \mathbb{K} \times \mathbb{H}_{t-1} = \mathbb{K}^t \times X,$$

para cada $t \geq 1$. Cada elemento $h_t \in \mathbb{H}_t$ es llamada t-historia, el cual es un vector de la forma $(x_0, a_0, x_1, a_1, \ldots, a_{t-1}, x_t)$, donde $(x_i, a_i) \in \mathbb{K}$ para $i = 0, ..., t-1$ y $x_t \in X$. Obsérvese que para cada $t \geq 1$, $\mathbb{H}_t$ es un subespacio de $\mathbf{H}_t := (X \times A)^t \times X$ y $\mathbf{H}_0 := X$.

Definición 1.2.1 *Una política es una sucesión $\pi = \{\pi_t\}$ de kérneles estocásticos, donde cada π_t está definido sobre A dado $\mathbb{H}_t$ y satisface que: $\pi_t(A(x_t)|h_t) = 1$ para cada $h_t \in \mathbb{H}_t$ con $t \geq 0$. El conjunto de todas las políticas se denota por Π.*

De acuerdo con está definición, una política $\pi = \{\pi_t\}$ puede interpretarse como una sucesión $\{a_t\}$ de variables aleatorias sobre A, tales que para cada t-historia y $t \geq 1$, la distribución de a_t es $\pi_t(\cdot|h_t)$, la cual está concentrada en el conjunto de acciones admisibles $A(x_t)$.
Se denotará a la familia de probabilidades condicionales sobre A dado X, como $\mathcal{P}(A|X)$. Sea Φ el conjunto de todas las probabilidades condicionales φ en $\mathcal{P}(A|X)$ tal que para toda $x \in X$ se tiene $\varphi(A(x)|x) = 1$.

Definición 1.2.2 *Una política $\pi \in \Pi$ es:*

1. ***Markoviana Aleatorizada*** (Π_{RM}). *Si existe una sucesión $\{\varphi_t\} \subseteq \Phi$ (definidas sobre A dado X), tal que $\pi_t(\cdot|h_t) = \varphi_t(\cdot|x_t)$ para toda $h_t \in \mathbb{H}_t$ y $t \geq 0$.*

2. ***Markoviana Aleatorizada Estacionaria*** (Π_{RS}). *Si existe* $\varphi \in \Phi$, *tal que:* $\pi_t(\cdot|h_t) = \varphi(\cdot|x_t)$ *para toda* $h_t \in \mathbb{H}_t$ *y* $t \geq 0$.

3. ***Determinista*** (Π_D). *Si existe una sucesión* $\{g_t\}$ *de funciones medibles* $g_t : \mathbb{H}_t \to A$ *tales que para cada* $h_t \in \mathbb{H}_t$ *y* $t \geq 0$, *se tiene que* $g_t(h_t) \in A(x_t)$ *y* $\pi_t(\cdot|h_t)$ *está concentrada en* $g_t(h_t)$.

4. ***Determinista Markoviana*** (Π_{DM}). *Si existe una sucesión* $\{f_t\} \subseteq \mathbb{F}$ *tal que* $\pi_t(\cdot|h_t)$ *está concentrada en* $f_t(x_t)$ *para cada* $h_t \in \mathbb{H}_t$ *y* $t \geq 0$.

5. ***Determinista Markoviana Estacionaria*** (Π_{DS}). *Si existe una función medible* $f : X \to A$ *(o* $f \in \mathbb{F}$*), tal que,* $f(x_t) \in A(x_t)$ *y* $\pi_t(\cdot|h_t)$ *está concentrada en* $f(x_t)$ *para cada* $h_t \in \mathbb{H}_t$ *y* $t \geq 0$.

Observación 1.2.1 *Por lo anterior tenemos que*

$$\Pi_{RS} \subset \Pi_{RM} \subset \Pi,$$

$$\Pi_{DS} \subset \Pi_{DM} \subset \Pi_D \subset \Pi.$$

1.3. Construcción del Proceso de Markov

Sea $(\Omega, \mathcal{F})$ el espacio medible que consiste del espacio muestral canónico $\Omega := \mathbf{H}_\infty = (X \times A)^\infty$ y $\mathcal{F}$ su correspondiente σ-álgebra producto. Los elementos de Ω son de la forma $w = (x_0, a_0, x_1, a_1, \cdots)$ con $x_t \in X$ y $a_t \in A$ para toda $t \geq 0$, las proyecciones x_t y a_t de Ω sobre X y A son llamados estado y acción respectivamente. Obsérvese que $\mathbb{H}_\infty = \mathbb{K}^\infty \subset \Omega$ es el espacio de historias $(x_0, a_0, x_1, a_1, \cdots)$ con $(x_t, a_t) \in \mathbb{K}$ para toda $t \geq 0$.
Sea $\pi = \{\pi_t\}$ una política de control y ν una medida de probabilidad sobre X. Por el teorema de Ionescu-Tulcea D.0.4 (Ver Apéndice D), existe una única medida de probabilidad P_ν^π en $(\Omega, \mathcal{F})$ tal que, para cada $B \in \mathcal{B}(X)$, $C \in \mathcal{B}(A)$ y $h_t \in \mathbb{H}_t$

$$\begin{aligned} P_\nu^\pi(x_0 \in B) &= \nu(B), \\ P_\nu^\pi(a_t \in C|h_t) &= \pi_t(C|h_t), \\ P_\nu^\pi(x_{t+1} \in B|h_t, a_t) &= Q(B|x_t, a_t). \end{aligned} \tag{1.1}$$

El proceso estocástico $(\Omega, \mathcal{F}, P_\nu^\pi, \{x_t\})$ es llamado **Proceso de Decisión de Markov a tiempo discreto.**

Observación 1.3.1 *Cuando* ν *está concentrada en algún estado* $x = x_0 \in X$, *se utiliza la siguiente notación:*

$$P_x^\pi := P_\nu^\pi \ \ y \ \ E_x^\pi := E_\nu^\pi.$$

1.4. Modelo de Ecuaciones en Diferencias

En algunas aplicaciones la ley de transición Q es inducida por una ecuación en diferencias estocásticas de la forma siguiente:

$$x_{t+1} = F(x_t, a_t, \xi_t),$$

para $t \geq 0$, $x_0 = x$ conocido. Donde $\{\xi_t\}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas (v.a. i.i.d) tomando valores en un espacio S con distribución común μ e independiente del estado x_0 y $F : \mathbb{K} \times S \longrightarrow X$ es una función medible conocida. En este caso la ley de transición Q está dada por:

$$\begin{aligned}
Q(B|x,a) &= P(x_{t+1} \in B | x_t = x, a_t = a) \\
&= P(F(x_t, a_t, \xi_t) \in B | x_t = x, a_t = a) \\
&= P(F(x, a, \xi_t) \in B) \\
&= \int_X I_B(F(x,a,s))\mu(ds) \\
&= E[I_B(F(x,a,\xi))],
\end{aligned} \tag{1.2}$$

con $B \in \mathcal{B}(X)$ y $(x,a) \in \mathbb{K}$, donde I_B es la función indicadora del conjunto B. Obsérvese que cuando la dinámica es determinista, es decir, $x_{t+1} = G(x_t, a_t)$ con $G : \mathbb{K} \to X$, la ley de transición es:

$$Q(B|x,a) = I_B(G(x,a)), \quad B \in \mathcal{B}(X).$$

1.5. Criterio de Rendimiento

Cada PDM estará dotado de una función real, llamada criterio de rendimiento la cual medirá de alguna manera la calidad de cada política, a través de la sucesión de costos que se generan. A continuación se definen los criterios de rendimiento que se utilizarán en este trabajo. Considérese un modelo de control de Markov estacionario y un conjunto de políticas Π.

Definición 1.5.1 *Se define para cada $x \in X$ y $\pi \in \Pi$ el* ***criterio de costo total descontado*** *por*

$$v_{\alpha,N,T}(\pi, x) := E_x^{\pi} \left[\sum_{t=0}^{N-1} \alpha^t c(x_t, a_t) + \alpha^N T(x_N) \right], \tag{1.3}$$

donde $\alpha \in (0,1)$ es llamado factor de descuento, $T : X \to \bar{\mathbb{R}}$ una función medible acotada inferiormente, llamada costo terminal y N es un entero positivo conocido,

llamado horizonte del problema de optimización.
Si $T(x) = 0$ para cada $x \in X$, se denota $v_{\alpha,N}$ en lugar de $v_{\alpha,N,T}$. Cuando $N = \infty$ y $T(x) = 0$ para cada $x \in X$, se denota el criterio de costo total descontado con horizonte infinito como v_α en lugar de $v_{\alpha,\infty}$.

Definición 1.5.2 *Para cada $x \in X$ y $\pi \in \Pi$ se define el **criterio de costo promedio** por*

$$u(\pi, x) := \limsup_{N\to\infty} \frac{1}{N} E_x^\pi \left[\sum_{t=0}^{N-1} c(x_t, a_t) \right]. \tag{1.4}$$

Definición 1.5.3 *La función de valor óptimo α descontada se define como,*

$$V_\alpha^*(x) := \inf_{\pi\in\Pi} V(\pi, x), \ \ x \in X. \tag{1.5}$$

donde $V(\pi, x)$ puede representar $v_{\alpha,N,T}(\pi, x), v_{\alpha,N}(\pi, x)$ o $v_\alpha(\pi, x)$, y la función de valor promedio óptimo como,

$$\bar{V}^*(x) := \inf_{\pi\in\Pi} u(\pi, x), \ \ x \in X. \tag{1.6}$$

El problema de control óptimo descontado consiste en encontrar una política $\pi_1^* \in \Pi$ que cumpla,

$$V_\alpha^*(x) = V(\pi_1^*, x), \quad x \in X, \tag{1.7}$$

y el problema de control óptimo promedio consiste en encontrar una política $\pi_2^* \in \Pi$ tal que,

$$\bar{V}^*(x) = u(\pi_2^*, x), \quad x \in X. \tag{1.8}$$

Las políticas $\pi_1^*, \pi_2^* \in \Pi$ que satisfacen (1.7) y (1.8) respectivamente son llamadas políticas óptimas.

Capítulo 2

Procesos de Decisión de Markov con Criterio de Costo Descontado

En este capítulo se estudia el problema de control óptimo usando el criterio de costo total descontado. Se presenta el caso con horizonte finito, utilizando programación dinámica se resuelve el problema de control óptimo. Posteriormente se aborda el caso con horizonte infinito con ayuda del espacio funciones con normas ponderadas (ver Apéndice A), se estudian las condiciones para la existencia de políticas óptimas.

2.1. Horizonte Finito

Considere $(X, A, \{A(x)|x \in X\}, Q, c)$ un modelo de control de Markov estacionario, definido en el Capítulo 1.

Definición 2.1.1 *Sean $v_1, v_2, ..., v_N$ funciones definidas en X como:*

$$v_N(x) := \alpha^N T(x) \tag{2.1}$$

y para cada $t \in \{N-1, ..., 0\}$,

$$v_t(x) := \min_{a \in A(x)} \left\{ c(x,a) + \alpha \int_X v_{t+1}(y) Q(dy|x,a) \right\}, \tag{2.2}$$

donde cada función v_t es medible para $t \in \{0, ..., N\}$.

El teorema de programación dinámica el cual esta basado en el principio de optimalidad de Bellman [2], tiene como suposición la existencia de selectores $f \in \mathbb{F}$, los cuales minimizan la ecuación de programación dinámica en cada etapa. Con la finalidad de garantizar la existencia de selectores medibles se considera la siguiente hipótesis, cabe señalar que en la literatura de PDM existen condiciones sobre las componentes del modelo de control, las cuáles garantizan dicha hipótesis (véase [7]).

Hipótesis 2.1.1 *Supóngase que para cada $t = 0, ..., N-1$, existe un selector $f_t \in \mathbb{F}$ tal que $f_t(x) \in A(x)$ y alcanza el mínimo en (2.2) para todo $x \in X$, es decir,*

$$v_t(x) = c(x, f_t(x)) + \alpha \int_X v_{t+1}(y) Q(dy|x, f_t(x)). \tag{2.3}$$

Teorema 2.1.1 ***(Teorema de Programación Dinámica).***
Supóngase que la Hipótesis 2.1.1 se satisface entonces la política $\pi^* = \{f_0, ..., f_{N-1}\}$ es óptima y se tiene que,

$$V^*(x) = v_{\alpha,N,T}(\pi^*, x) = v_0(x), \quad x \in X. \tag{2.4}$$

Demostración.
Sean $x \in X$ y $\pi = \{\pi_t\}$ una política arbitraria, se define,

$$\begin{aligned} \mathcal{C}_t(\pi, x) &:= E_x^\pi \left[\sum_{n=t}^{N-1} \alpha^n c(x_n, a_n) + \alpha^N T(x_N) | x_t = x \right] \\ \mathcal{C}_N(\pi, x) &:= E_x^\pi \left[\alpha^N T(x_N) | x_N = x \right] \\ &= \alpha^N T(x) \ , \end{aligned} \tag{2.5}$$

donde $\mathcal{C}_t(\pi, x)$ es el costo total esperado del tiempo t al tiempo terminal N, dado el estado $x_t = x$, con $t = 0, ..., N-1$. Observe que,

$$v_{\alpha,N,T}(\pi, x) = \mathcal{C}_0(\pi, x) \ .$$

Primero se demuestra que,

$$\mathcal{C}_t(\pi, x) \geqslant v_t(x) \ . \tag{2.6}$$

Y en particular si $\pi = \pi^*$ en (2.6) se tiene la igualdad, es decir,

$$\mathcal{C}_t(\pi^*, x) = v_t(x) \ . \tag{2.7}$$

Se procede a demostrar (2.6), primeramente observe que para $t = 0$ es válido, debido a las siguientes relaciones,

$$\mathcal{C}_0(\pi, x) = v_{\alpha,N,T}(\pi, x) \geqslant v_0(x), \qquad x \in X,$$

con $v_{\alpha,N,T}(\pi^*, x) = v_0(x)$ y cuando $\pi = \pi^*$,

$$\mathcal{C}_0(\pi^*, x) = v_{\alpha,N,T}(\pi^*, x) = V^*(x) = v_0(x) \ .$$

Ahora se procede a probar (2.6) y (2.7) por inducción hacia atrás, para $t = N-1, ..., 0$.
Es claro que se cumple (2.6) y (2.7) cuando $t = N$, ya que

$$\mathcal{C}_N(\pi, x) = \alpha^N T(x) = v_N(x) \, .$$

Supóngase que, para alguna $t \in \{0, ..., N-1\}$

$$\mathcal{C}_{t+1}(\pi, x) \geqslant v_{t+1}(x), \qquad x \in X \, ,$$

entonces

$$\begin{aligned}
\mathcal{C}_t(\pi, x) &= E_x^\pi \left[\sum_{n=t}^{N-1} \alpha^n c(x_n, a_n) + \alpha^N T(x_N) \middle| \, x_t = x \right] \\
&= E_x^\pi \left[\alpha^t c(x_t, a_t) \middle| \, x_t = x \right] + E_x^\pi \left[\sum_{n=t+1}^{N-1} \alpha^n c(x_n, a_n) + \alpha^N T(x_N) \middle| \, x_t = x \right] \\
&= \int_A \alpha^t c(x, a) \pi(da|x) + E_x^\pi \left[\sum_{n=t+1}^{N-1} \alpha^n c(x_n, a_n) + \alpha^N T(x_N)) \middle| \, x_t = x \right] \, .
\end{aligned}$$

Note que,

$$\begin{aligned}
&E_x^\pi \left[\sum_{n=t+1}^{N-1} \alpha^n c(x_n, a_n) + \alpha^N T(x_N) \middle| \, x_t = x \right] \\
&\qquad = E_x^\pi \left[E_x^\pi \left[\sum_{n=t+1}^{N-1} \alpha^n c(x_n, a_n) + \alpha^N T(x_N) \middle| \, x_{t+1} = y \right] \middle| \, x_t = x \right] \, .
\end{aligned}$$

Ya que x_{t+1} está en función de x_t y por la Proposición E.0.6 (ver Apéndice E).
Por otro lado,

$$E_x^\pi \left[\sum_{n=t+1}^{N-1} \alpha^n c(x_n, a_n) + \alpha^N T(x_N) \middle| \, x_{t+1} = y \right] = \mathcal{C}_{t+1}(\pi, y) \, .$$

Así,

$$\begin{aligned}
\mathcal{C}_t(\pi, x) &= \int_A c(x,a)\pi(da|x) + E_x^\pi \left[\mathcal{C}_{t+1}(\pi, y)|x_t = x\right] \\
&= \int_A \left\{ c(x,a) + \int_X \mathcal{C}_{t+1}(\pi, y) Q(dy|x,a) \right\} \pi(da|x) \\
&\geqslant \int_A \left\{ c(x,a) + \alpha \int_X v_{t+1}(y) Q(dy|x,a) \right\} \pi(da|x) \\
&\geqslant \min_{a \in A(x)} \left\{ c(x,a) + \alpha \int_X v_{t+1}(y) Q(dy|x,a) \right\} \\
&= v_t(x) \ .
\end{aligned}$$

Por lo tanto,

$$\mathcal{C}_t(\pi, x) \geqslant v_t(x) \qquad \forall t \in \{N-1, ..., 0\}.$$

Si la igualdad (2.7) se cumple para $\pi = \pi^*$, es decir,

$$\mathcal{C}_{t+1}(\pi^*, x) = v_{t+1}(x) \ .$$

se sigue que

$$\begin{aligned}
\mathcal{C}_t(\pi^*, x) &= c(x, f_t(x)) + E_x^\pi[\mathcal{C}_{t+1}(\pi^*, y)|x_t = x] \\
&= c(x, f_t(x)) + \alpha \int_x v_{t+1}(y) Q(dy|x, f_t(x)) \\
&= v_t(x) \ .
\end{aligned}$$

Es así como se ha demostrado el Teorema . □

2.2. Horizonte Infinito

Sea $(X, A, \{A(x)|x \in X\}, Q, c)$ un modelo de control de Markov estacionario, y se considera el criterio de costo total descontado con horizonte infinito v_α con su correspondiente función de valor óptimo V_α^*, dadas en Definición 1.5.1 y Definición 1.5.3, respectivamente.

2.2.1. Condiciones de Optimalidad

En ésta sección se presentan las condiciones necesarias para la existencia de una política óptima estacionaria.

Definición 2.2.1 *Se define la ecuación de valor óptimo α descontada de acuerdo* *a la siguiente expresión funcional,*

$$V_\alpha^*(x) = \min_{A(x)} \left\{ c(x,a) + \alpha \int_X V_\alpha^*(y) Q(dy|x,a) \right\}, \qquad x \in X. \tag{2.8}$$

Con la finalidad de estudiar el criterio descontado (ver Definición 1.3) se consideran tres tipos de Hipótesis. La primera, Hipótesis 2.2.1, se le conoce como condición de continuidad-compacidad para modelos de control de Markov. La segunda, Hipótesis 2.2.2, considera una función de peso w que condiciona el crecimiento de la función de costo, finalmente, la Hipótesis 2.2.3, es una condición de continuidad.

Hipótesis 2.2.1 *Para cada estado $x \in X$.*

(a) $A(x)$ es compacto.

(b) $c(x,a)$ es inferiormente semicontinua (l.s.c) en $A(x)$.

(c) La función $\mu'(x,a) := \int_X \mu(y) Q(dy|x,a)$ es continua en $A(x)$ para cada función $\mu \in \mathbb{B}(X)$, donde $\mathbb{B}(X)$ denota el espacio de Banach de funciones medibles, continuas y acotadas en X, con la norma supremo,

$$\| \mu \|_\infty := \sup_{x \in X} |\mu(x)|.$$

El siguiente teorema da una equivalencia a la Hipótesis 2.2.1 (c).

Teorema 2.2.1 *La Hipótesis 2.2.1 (c) se satisface si y sólo si $\int_X \nu(y) Q(dy|x,a)$ es l.s.c. en $A(x)$ para cada función no negativa $\nu \in \mathbb{B}(X)$.*

Demostración.
Por definición de continuidad es claro que la suficiencia se cumple. Ahora, supóngase que

$$\int_X \nu(y) Q(dy|x,a),$$

es l.s.c en $A(x)$ para cada función no negativa $\nu \in \mathbb{B}(X)$. Sea $\mu \in \mathbb{B}(X)$, se tiene que $\|\mu\|_\infty + \mu$ es no negativa, entonces,

$$\int_X (\|\mu\|_\infty + \mu)(y) Q(dy|y,a) = \mu'(x,a) + \|\mu\|_\infty,$$

es l.s.c en $A(x)$ lo cual implica que $\mu'(x,a)$ es l.s.c en $A(x)$. Análogamente, usando $-\mu$ se obtiene que $\mu'(x,a)$ es u.s.c en $A(x)$. Por lo tanto, como $\mu'(x,\cdot)$ es l.s.c y u.s.c en $A(x)$ entonces es continua. $\square$

Hipótesis 2.2.2 ***(Condición de crecimiento)***
Existen constantes no negativas M y β, con $1 \leq \beta \leq \frac{1}{\alpha}$ y una función de peso $w \geq 1$ sobre X tal que para cada estado $x \in X$,

(a) $\sup_{a\in A(x)} |c(x,a)| \leq Mw(x)$,

(b) $\sup_{a\in A(x)} \int_X w(y)Q(dx|x,a) \leq \beta w(x)$.

Observación 2.2.1 *Un caso particular donde la Hipótesis 2.2.2 se cumple es cuando c es acotada sobre $\mathbb{K}$, es decir, existe una constante positiva M_1 tal que $|c(x,a)| \leq M_1$ para cada $(x,a) \in \mathbb{K}$.*

Hipótesis 2.2.3 *Para cada estado $x \in X$ la función,*

$$w'(x,a) := \int_X w(y)Q(dx|x,a),$$

es continua en $A(x)$.

La siguiente proposición satisface algunas propiedades que se implican de la Hipótesis 2.2.2.

Proposición 2.2.1 *Supóngase que la Hipótesis 2.2.2 se satisface con estado inicial $x_0 = x \in X$, entonces,*

(a) $E_x^\pi [w(x_t)] \leq \beta^t w(x)$*, para cada* $t \geq 0$.

(b) $|v_\alpha(\pi,x)| \leq M\frac{w(x)}{1-\gamma}$*, donde* $\gamma := \beta\alpha$.

(c) $\lim_{t\to\infty} \alpha^t E_x^\pi [|\mu(x_t)|] = 0$ *para cada función* $\mu \in \mathbb{B}_w(X)$.

Demostración.
(a) El caso cuando $t = 0$, se sigue ya que, $x_0 = x$ y

$$E_x^\pi[w(x_0)] = \beta^0 w(x_0).$$

Si $t \geq 1$, usando la Hipótesis 2.2.2 y (1.1),

$$\begin{aligned} E_x^\pi [w(x_t)|h_{t-1}, a_{t-1}] &= \int_X w(y)Q(dy|x_{t-1}, a_{t-1}) \\ &\leq \beta w(x_{t-1}), \end{aligned}$$

por consiguiente,

$$\begin{aligned} E_x^\pi[w(x_t)] &\leq \beta E_x^\pi[w(x_{t-1})] \\ &\leq \beta^2 E_x^\pi[w(x_{t-2})] \\ &\vdots \\ &\leq \beta^t w(x). \end{aligned}$$

(b) Por inciso (a) e Hipótesis 2.2.2 (a), se tiene que para cada $t \geq 1$,

$$E_x^\pi[|c(x_t, a_t)|] \leq M\beta^t w(x). \tag{2.9}$$

Finalmente,

$$\begin{aligned} |v_\alpha(\pi, x)| &\leq \sum_{t=0}^{\infty} \alpha^t E_x^\pi[|c(x_t, a_t)|] \\ &\leq \frac{Mw(x)}{1-\gamma}, \end{aligned}$$

(c) Por la definición de $w - norma$ dada en A.1 (ver Apéndice A), y (2.9), se sigue para cada $\mu \in \mathbb{B}_w(X)$,

$$E_x^\pi[|\mu(x_t)|] \leq \|\mu\|_w E_x^\pi[w(x_t)] \leq \|\mu\|_w \beta^t w(x). \tag{2.10}$$

Tomando $t \to \infty$ en (2.10),

$$\lim_{t\to\infty} \alpha^t E_x^\pi[\mu(x_t)] = 0.$$

Con lo cual se ha demostrado la Proposición. □

Proposición 2.2.2 *Supóngase que la Hipótesis 2.2.2 se satisface. Entonces*

$$C(x) := \sum_{t=0}^{\infty} \alpha^t c_t(x) < \infty \tag{2.11}$$

donde $c_0(x) := \sup_{A(x)} \mid c(x, a) \mid$ *y*

$$c_t(x) := \sup_{a\in A(x)} \left\{ \int_X c_{t-1}(y) Q(dy|x, a) \right\}, \quad t \geq 1. \tag{2.12}$$

Recíprocamente, si

(a) $C \geq 1$.

(b) Se satisface la desigualdad (2.11) para algún α_0 *con* $0 < \alpha_0 < \alpha$*, entonces la Hipótesis 2.2.2 se cumple con* $M := 1$*,* $w(x) := C(x)$ *y* $\beta := \alpha_0^{-1}$.

Demostración.
Si la Hipótesis 2.2.2 se cumple, entonces es fácil probar por inducción que,

$$c_t(x) \leq M\beta^t w(x) \quad x \in X, \;\; t = 0, 1, ...$$

Así, usando que $0 < \alpha\beta < 1$ se obtiene,

$$C(x) \leq \frac{Mw(x)}{1-\alpha\beta} < \infty, \qquad x \in X.$$

Recíprocamente, por (b) y (2.12),

$$\begin{aligned}
\int_X C(y)Q(dy|x,a) &= \sum_{t=0}^{\infty} \int c_t(y)Q(dy|x,a) \\
&\leq \sum_{t=0}^{\infty} \alpha_0^t c_{t+1}(x) \\
&= \alpha_0^{-1}\left[C(x) - c_0(x)\right].
\end{aligned}$$

Así,

$$\int_X C(y)Q(dy|x,a) \leq \alpha_0^{-1}C(x).$$

Además,

$$\begin{aligned}
c_0(x)\alpha_0^{-1}(x) + 1 &\leq c_0\alpha_0^{-1} + \int C(y)Q(dy|x,a) \\
&= \alpha_0^{-1}C(x).
\end{aligned}$$

Por lo tanto $\sup_{a\in A(x)} |c(x,a)| \leq C(x)$ para cada $x \in X$. $\square$

2.2.2. Existencia de una Política Óptima Estacionaria

En esta sección usando las hipótesis anteriores se demuestra el principal resultado de este capítulo.
A continuación se define el algoritmo de iteración de valores óptimos.

Definición 2.2.2 *El algoritmo de iteración de valores óptimos α descontada, se define como,*

$$v_n(x) := \min_{a\in A(x)} \left\{ c(x,a) + \alpha \int_X v_{n-1}(y)Q(dy|x,a) \right\}, \tag{2.13}$$

para cada $n \geq 1$ y $x \in X$, con $v_0(\cdot) \equiv 0$.
Si $n \geq 1$, v_n es el costo óptimo en el n-ésimo paso, i.e.,

$$v_n(x) := \inf_{\pi\in\Pi} V_n(\pi, x), \qquad x \in X,$$

donde

$$V_n(\pi, x) := E_x^\pi \left[\sum_{t=0}^{n-1} \alpha^t c(x_t, a_t) \right].$$

La siguiente definición, presenta el operador de programación dinámica T_α.

Definición 2.2.3 *Dada una función medible $\mu : X \longrightarrow \mathbb{R}$ y una constante α tal que $0 \leq \alpha \leq 1$, denotamos por T_α a la siguiente función,*

$$T_\alpha(\mu(x)) := \inf_{a \in A(x)} \left\{ c(x,a) + \alpha \int_X \mu(y) Q(dy|x,a) \right\}, \quad x \in X. \tag{2.14}$$

Si el ínfimo en (2.14) se alcanza en alguna acción $a \in A(x)$ para cada $x \in X$, entonces se escribe mín en lugar de ínf.
Más adelante, se demuestra que para cada $0 < \alpha < 1$, T_α es un operador contracción en el espacio $\mathbb{B}_w(X)$, para esto se requiere el siguiente Lema.

Lema 2.2.1 *Supóngase que la Hipótesis 2.2.1 se satisface, entonces,*

la función

$$\mu(x,a) := \int \mu(y) Q(dy|x,a), \tag{2.15}$$

es continua en $A(x)$ para cada $x \in X$ y cada función μ en $\mathbb{B}_w(X)$.

Demostración.
Sea $\mu \in \mathbb{B}_w(X)$, tal que $|\mu(x)| \leq mw(x)$ para cada $x \in X$, donde $m := \|\mu\|_w$. Entonces $\mu_m := \mu + mw$ es una función no negativa en $\mathbb{B}_w(X)$, la cual es límite de una sucesión de funciones medibles no decrecientes acotadas $\{\mu_k\} \subseteq \mathbb{B}_w(X)$. Sea $x \in X$, fijo y sea $\{a_n\}$ una sucesión en $A(x)$ que converge a $a \in A(x)$. Por lo tanto, como $\mu_k \uparrow \mu_m$, cuando $k \to \infty$, usando Hipótesis 2.2.1 (c),

$$\begin{aligned} \liminf_{n\to\infty} \int_X \mu_m(y) Q(dy|x,a_n) &\geq \liminf_{n\to\infty} \int_X \mu_k(y) Q(dy|x,a_n) \\ &= \int_X \mu_k(y) Q(dy|x,a). \end{aligned}$$

Tomando, $k \to \infty$ en la relación y aplicando el Teorema de Convergencia Monótona, se obtiene que

$$\liminf_{n\to\infty} \int_X \mu_m(y) Q(dy|x,a_n) \geq \int_X \mu_m(y) Q(dy|x,a),$$

por consiguiente, $\int_X \mu_m(y) Q(dy|x,a_n)$ es l.s.c. en $A(x)$ lo cual implica que $\mu'(x,a)$ es l.s.c. en $A(x)$. En otras palabras $\mu'(x,a)$ es l.s.c. en $A(x)$ para cada función μ en $\mathbb{B}_w(X)$.
Aplicando el procedimiento anterior para la función $-\mu$ en lugar de μ, se obtiene que $\mu'(x,a)$ es u.s.c., de ahí que $\mu'(x,\cdot)$ es continua en $A(x)$.
□

Lema 2.2.2 *Sea $\nu : \mathbb{K} \to \mathbb{R}$ función medible. Definamos*

$$\nu^*(x) := \inf_{a \in A(x)} \nu(x,a), \qquad x \in X. \tag{2.16}$$

(a) Si $\nu(x,\cdot)$ es l.s.c. en $A(x)$ para cada $x \in X$, entonces existe un selector $f \in \mathbb{F}$ tal que $f(x) \in A(x)$ alcanza el mínimo en (2.16) para todo $x \in X$, es decir,

$$\nu^*(x) = \nu(x, f(x)), \quad x \in X. \tag{2.17}$$

y ν^* es una función medible.

(b) Si el selector $x \to A(x)$ es u.s.c. y ν es l.s.c y acotada inferiormente en $\mathbb{K}$, entonces existe un selector $f \in \mathbb{F}$ que satisface (2.17) y, además, ν^* es l.s.c. y acotada inferiormente en X.

(c) Suponga que $x \to A(x)$ es u.s.c., ν es l.s.c. y, además,

$$\sup_{a \in A(x)} |\nu(x,a)| \leq kw(x) \quad \forall \in X, \tag{2.18}$$

donde k es una constante positiva y $w(\cdot) \geq 1$ es una función continua en X, entonces existe un selector $f \in \mathbb{F}$ que satisface (2.17), ν^* es l.s.c., y

$$|\nu^*(x)| \leq kw(x) \quad \forall x \in X. \tag{2.19}$$

es decir, ν^* es una función l.s.c. en el espacio $\mathbb{B}_w(X)$ y satisface que $\| \nu^* \| \leq k$.

Demostración
Para la demostración de (a) y (b), puede consultarse en [11] y [12].
En este apartado nos enfocamos a demostrar (c), la cual se sigue al aplicar (b) a la función l.s.c no negativa

$$\mu(x,a) := \nu(x,a) + kw(x).$$

□
El siguiente Teorema demuestra que el operador de programación dinámica dado en Definición 2.2.3 es un operador contracción en el espacio de funciones $\mathbb{B}_w(X)$.

Proposición 2.2.3 *Supóngase que las Hipótesis 2.2.1, 2.2.2 y 2.2.3 se satisfacen y sea T_α el operador de programación dinámica. Entonces*

(a) T_α es un operador contracción en $\mathbb{B}_w(X)$ con constante de contracción $\gamma := \alpha\beta < 1$; es decir, T_α es operador en $B_w(X)$ y

$$\| T_\alpha(\mu_1) - T_\alpha(\mu_0) \|_w \leq \gamma \| \mu_1 - \mu_0 \|_w \quad \forall \mu_1, \mu_0 \in \mathbb{B}_w(X). \tag{2.20}$$

(b) Para cada función $\mu \in \mathbb{B}_w(X)$ existe un selector $f \equiv f_\mu$ en $\mathbb{F}$ tal que

$$T_\alpha(\mu(x)) = c(x,f) + \alpha \int_X \mu(y)Q(dy|x,f), \quad \forall x \in X. \tag{2.21}$$

Demostración
Sea $\mu \in \mathbb{B}_w(X)$ entonces, por el Lema 2.2.1 y la Hipótesis 2.2.1 (b), la función

$$\mu(x,a) := c(x,a) + \alpha \int_X \mu(y)Q(dy|x,a),$$

es l.s.c. en $A(x)$ para cada $x \in X$. Por lo tanto, del Lema 2.2.2 (a), se tiene que $T_\alpha(u)$ es una función medible y existe $f \in \mathbb{F}$ que satisface (2.21).
Por otro lado, $T_\alpha(u)$ tiene $w - norma$ finita, ya que, por la Hipótesis 2.2.2

$$\begin{aligned} |\mu(x,a)| &\leq Mw(x) + \alpha \parallel \mu \parallel_w \int_X w(y)Q(dy|x,a) \\ &\leq (M + (\alpha\beta) \parallel \mu \parallel_w)w(x), \end{aligned}$$

para cada $x \in X$ y $a \in A(x)$. Más aún, T_α es un operador monótono ($u_1 \leq u_2$ implica que $T_\alpha(u_1) \leq T_\alpha(u_2)$). Así, usando Proposición C.0.2 (ver Apéndice C), para complementar la prueba es suficiente demostrar que,

$$T_\alpha(\mu + rw) \leq T(\mu) + \gamma rw \qquad \forall \mu \in \mathbb{B}_w(X), r \in \mathbb{R}.$$

Por lo cual, para cada $r \in \mathbb{R}$,

$$\begin{aligned} T_\alpha((\mu + rw)(x)) &= \inf_{a \in A(x)} \left\{ c(x,a) + \alpha \int_X (\mu(y) + rw(y))Q(dy|x,a) \right\} \\ &\leq T_\alpha(\mu(x)) + \alpha\beta w(x), \end{aligned}$$

así T_α es un operador contracción sobre $\mathbb{B}_w(X)$ con módulo $\gamma := \alpha\beta$. □
Usando T_α, las ecuaciones (2.8) y (2.13) se pueden reescribir como

$$V_\alpha^* = T_\alpha V_\alpha^* \tag{2.22}$$

y

$$v_n = T_\alpha v_{n-1} = T_\alpha^n v_0 \qquad \forall n = 0, 1, ..., \quad con \quad v_0 = 0. \tag{2.23}$$

Teorema 2.2.1 *Supóngase que las Hipótesis 2.2.1, 2.2.2 y 2.2.3 se satisfacen. Sea β dada en la Hipótesis 2.2.2 (b). Se define $\gamma = \alpha\beta$, entonces*

(a) la función de valor óptimo α descontada, V_α^* es la única solución de la

ecuación de valor óptimo (2.8) en el espacio $\mathbb{B}_w(X)$*, y*

$$\| v_n - V_\alpha^* \|_w \leq \frac{M\gamma^n}{(1-\gamma)}, \qquad n = 0, 1, ... \tag{2.24}$$

donde M *es la constante dada en la Hipótesis 2.2.2.*

(b) Existe un selector $f^* \in \mathbb{F}$ *tal que* $f^*(x) \in A(x)$ *alcanza el mínimo en la ecuación (2.8) para cada* $x \in X$*, es decir,*

$$V_\alpha^*(x) = c(x, f^*(x)) + \int_X V_\alpha^*(y) Q(dy|x, f^*(x)), \quad x \in X. \tag{2.25}$$

y una política $f_*^\infty \in \Pi_{DS}$ *la cual es óptima, recíprocamente, si* $f_*^\infty \in \Pi_{DS}$ *es óptima entonces satisface (2.25).*

(c) Una política π^* *es óptima si y sólo si* $v_\alpha(\pi^*, \cdot)$*, satisface la ecuación de valor óptimo.*

(d) Si una política óptima existe, entonces existe una política estacionaria determinista la cual es óptima.

Demostración
(a) Por la Proposición 2.2.3 y el Teorema del Punto Fijo de Banach, T_α tiene un único punto fijo μ^* en $\mathbb{B}_w(X)$, i.e.,

$$T_\alpha \mu^* = \mu^* \tag{2.26}$$

y

$$\| T_\alpha^n \mu - \mu^* \|_w \leq \gamma^n \| \mu - \mu^* \|_w, \quad \forall \mu \in \mathbb{B}_w(X), \; n = 0, 1, ... \tag{2.27}$$

Para concluir la parte (a) del teorema, es necesario demostrar lo siguiente:

(I) $V_\alpha^* \in \mathbb{B}_w(X)$, con $\| V_\alpha^* \|_w \leq \frac{M}{(1-\gamma)}$.

(II) $V_\alpha^* = \mu^*$.

Si (I) y (II) se satisfacen entonces (2.24) del Teorema se sigue de (2.27) y (2.23) con $\mu \equiv 0$.
Para probar (I), sea $\pi \in \Pi$ una política arbitraria y sea $x \in X$ un estado inicial cualquiera. Por la Proposición 2.2.1, se tiene que

$$|V_\alpha^*(x)| \leq \frac{Mw(x)}{(1-\gamma)},$$

con lo cual se concluye (I). Para demostrar (II), por la Proposición 2.2.1,

$$\lim_{t\to\infty} \alpha^t E_x^t[\mu(x_t)] = 0 \qquad \forall \pi \in \Pi, \; x \in X, \; \mu \in \mathbb{B}_w(X).$$

Ahora se considera la igualdad $\mu^* = T_\alpha \mu^*$ dada en (2.26). Por la Proposición 2.2.2 (b), existe un selector $f \in \mathbb{F}$ tal que

$$\mu^*(x) = c(x,f) + \alpha \int_X \mu^*(y) Q(dy|x,f), \qquad x \in X. \tag{2.28}$$

Iterando (2.28), de $t=0$ a $t=n-1$, se obtiene,

$$\mu^*(x) = E_x^f[\sum_{t=0}^{n-1} \alpha^t c(x,f)] + \alpha^n E_x^f[\mu^*(x_n)],$$

tomando $n \to \infty$

$$\mu^*(x) = E_x^f[\sum_{t=0}^{\infty} \alpha^t c(x,f)] = v_\alpha(f,x).$$

Por lo tanto, por definición de V_α^*,

$$\mu^*(x) \geq V_\alpha^*(x).$$

Para demostrar la desigualdad inversa, notar que ecuación (2.26) implica que

$$\mu^*(x) \leq c(x,a) + \alpha \int_X \mu^*(y) Q(dy|x,a), \quad (x,a) \in \mathbb{K}. \tag{2.29}$$

Así, para cualquíer política $\pi \in \Pi$ y un estado inicial $x \in X$, usando (2.29)

$$\alpha^t E_x^\pi[c(x_t,a_t) + \alpha\mu^*(x_{t+1}) - \mu^*(x_t)|h_t,a_t] \geq 0 \qquad \forall t = 0,1,\ldots\,.$$

Por lo tanto, tomando esperanza y sumando desde $t=0$ a $t=n-1$, se obtiene

$$\mu^*(x) \leq E_x^\pi\left[\sum_{t=0}^{n-1} \alpha^t c(x_t,a_t)\right] + \alpha^n E_x^\pi\left[\mu^*(x_n)\right], \qquad \forall n = 1,2,\ldots\,.$$

Finalmente, tomando $n \to \infty$ en la anterior desigualdad y usando la Proposición 2.2.1,

$$\mu^*(x) \leq v_\alpha(\pi,x)\,,$$

Como π y x son arbitrarios, entonces

$$\mu^*(x) \leq V_\alpha^*(x), \quad x \in X.$$

Por lo tanto, (II) queda demostrado.
(b) La existencia de un selector $f^* \in \mathbb{F}$ que satisface (2.25), se sigue de la parte (a) y Proposición 2.2.2 (b).
Recíprocamente, para cualquier política determinista estacionaria $f_*^\infty \in \Pi_{DS}$

satisface la igualdad en la ecuación de valor óptimo α descontada (2.8),

$$v_\alpha(f_*^\infty, x) = c(x, f_*^\infty) + \alpha \int_X v_\alpha(f_*^\infty, y) Q(dy|x, f_*^\infty), \quad x \in X. \tag{2.30}$$

Debido a que si se expande el lado derecho de la ecuación (1.3) con $\pi = f_*^\infty \in \Pi_{DS}$, se obtiene

$$v_\alpha(f_*^\infty, x) = c(x, f_*^\infty) + \alpha E_x^\pi \left[\sum_{t=1}^{\infty} \alpha^{t-1} c(x, f_*^\infty) \right], \quad x \in X,$$

usando la propiedad de Markov (1.1) y la Definición 1.5.1 con $\pi = f_*^\infty$, se obtiene que f_*^∞ satisface (2.30).
Por lo tanto, si f_*^∞ es óptima se tiene que $v_\alpha(f_*^\infty, \cdot) = V_\alpha^*(\cdot)$, así (2.25) se sigue de (2.30).
Finalmente, parte (c) se sigue de (b) y (d) es consecuencia de (b) y (c). $\square$

Capítulo 3

Procesos de Decisión de Markov con Criterio de Costo Promedio

En esta sección se aborda el problema de control óptimo usando el criterio de costo promedio $u(\pi, x)$, dado en la Definición 1.5.2. Para demostrar la existencia de una politíca óptima con este criterio, se usan los resultados obtenidos en el capítulo anterior.

3.1. Condiciones de Optimalidad

En esta sección se presentan las condiciones necesarias para la existencia de una politica óptima estacionaria con criterio de costo promedio, además de algunos lemas preliminares que se requieren para nuestro principal resultado. Se retoma la función de valor promedio óptimo, la cual esta dada por,

$$\bar{V}^*(x) = \inf_{\pi \in \Pi} u(\pi, x), \qquad x \in X. \tag{3.1}$$

Se busca una política óptima estacionaria $\pi^* \in \Pi_{DS}$ tal que,

$$\bar{V}^*(x) = u(\pi^*, x). \tag{3.2}$$

Para esto, supóngase que la Hipótesis 2.2.1 se satisface. Las Hipótesis 2.2.2 y 2.2.3 son remplazadas por la siguiente condición.

Hipótesis 3.1.1 *Supóngase que existe una función de peso, $w_1 \geq 1$ sobre X y constantes positivas, K, τ con $\tau < 1$ y b tal que,*

(a) $\int_X w_1(y) Q(dy|x, a) \leq \tau w(x) + b$, *para cada* $(x, a) \in \mathbb{K}$,

(b) $|c(x, a)| \leq K w_1(x)$ *para cada* $(x, a) \in \mathbb{K}$.

Hipótesis 3.1.2 *Supóngase que la función de peso $w_1 : X \to [1, \infty)$, satisface que,*

$$w_1'(x,a) := \int_X w_1(y) Q(dy|x,a), \tag{3.3}$$

es continua en $A(x)$.

Con el siguiente Lema, se garantiza que los resultados obtenidos en el capítulo anterior con respecto a la existencia de una política óptima estacionaria que optimice el criterio de costo descontado se satisfacen si suponemos las Hipótesis 2.2.1, 3.1.1 y 3.1.2.

Lema 3.1.1 *Supóngase que la Hipótesis 3.1.1 se satisface entonces la Hipótesis 2.2.2 se cumple.*

Demostración.
De la Proposición 2.2.2, tomando a $C(x)$ y $c_t(x)$ como en (2.11) y (2.12), respectivamente, a excepción de c_0 la cual se redefine como,

$$c_0(x) := 1 + \sup_{a \in A(x)} |c(x,a)|.$$

Observe que, por (a) de la Proposición 2.2.2, se tiene que $c_0 \leq M w_1$ con $M_1 := 1 + K$. Entonces

$$c_t(x) \leq M_1 w_1(x) \tau^t + M_1 b \sum_{j=0}^{t-1} \tau^j, \quad t \in \mathbb{N},$$

y así,

$$C(x) \leq M_1 w_1(x)/(1-\alpha\tau) + M_1 b\alpha/(1-\alpha)(1-\alpha\tau).$$

Por tanto, por el recíproco de la Proposición 2.2.2, existen una constante M y β que satisfacen la Hipótesis 2.2.2. $\square$

Observación 3.1.1 *Para efectos prácticos se redefine la función de peso w_1 dada en la Hipótesis 3.1.2, como $w := w_1$.*

Los dos siguientes lemas son herramientas para nuestro resultado principal.

Lema 3.1.2 *Supóngase que la Hipótesis 3.1.1 se satisface. Entonces*

(a) $E_x^\pi[w(x_t)] \leq \tau^t w(x) + \frac{b(1-\tau^t)}{1-\tau}$, *para cada* $t \geq 0$, $x \in X$ *y* $\pi \in \Pi$.

(b) $|u(x,\pi)| \leq \frac{bK}{1-\tau}, \quad x \in X,\ \pi \in \Pi.$

Demostración
(a) La desigualdad es válida para $t = 0$. Supóngase que $t \geq 1$, usando la Hipótesis 3.1.1 y (1.1), se tiene que

$$\begin{aligned} E_x^\pi[w(x_t)|h_{t-1}, a_{t-1}] &= \int_X w(y)Q(dy|x_{t-1}, a_{t-1}) \\ &\leq \tau w(x_{t-1}) + b. \end{aligned}$$

Por consiguiente

$$\begin{aligned} E_x^\pi[w(x_t)] &\leq \tau E_x^\pi[w(x_{t-1})] + b \\ &\leq \tau^2 E_x^\pi[w(x_{t-2})] + b + b\tau \\ &\vdots \\ &\leq \tau^t w(x) + b + b\tau + \cdots + b\tau^{t-1} \\ &= \tau^t w(x) + \frac{b(1-\tau^t)}{1-\tau}. \end{aligned}$$

(b) Usando la Hipótesis 3.1.1 e inciso anterior

$$\begin{aligned} |u(x,\pi)| &= \left| \limsup_{N\to\infty} \frac{1}{N} E_x^\pi \left[\sum_{t=0}^{N-1} c(x_t, a_t) \right] \right| \\ &\leq \limsup_{N\to\infty} \frac{1}{N} E_x^\pi \left[\sum_{t=0}^{N-1} |c(x_t, a_t)| \right] \end{aligned}$$

$$\begin{aligned} &\leq \limsup_{N\to\infty} \frac{K}{N} \sum_{t=0}^{N-1} \left(\tau^t w(x) + \frac{b(1-\tau^t)}{1-\tau} \right) \\ &= \frac{bK}{1-\tau}. \quad \square \end{aligned}$$

En el siguiente lema, se retoma el criterio de costo α descontada definido en 1.5.1.

Lema 3.1.3 *Supóngase que las Hipótesis 2.2.1 y 3.1.1 se satisfacen entonces,*

$$|v_\alpha(x,\pi)| \leq K\frac{w(x)}{(1-\alpha)} + K\frac{b}{(1-\tau)(1-\alpha)},$$

para cada $x \in X$ y $\pi \in \Pi$.

Demostración. Por el Lema 3.1.2 y la Hipótesis 3.1.1 (a), se tiene

$$\begin{aligned}
|v_\alpha(x,\pi)| &\leq K\sum_{t=0}^{\infty}\alpha^t E_x^\infty[w(x_t)] \\
&\leq K\sum_{t=0}^{\infty}\alpha^t\left(\tau^t w(x)+\frac{1-\tau^t}{1-\tau}b\right) \\
&\leq K\sum_{t=0}^{\infty}\alpha^t\left(w(x)+\frac{b}{1-\tau}\right) \\
&\leq \frac{Kw(x)}{1-\alpha}+\frac{Kb}{(1-\tau)(1-\alpha)}. \quad \square
\end{aligned}$$

Hipótesis 3.1.3 *Existen dos funciones, $\nu_1, \nu_2 \in \mathbb{B}_w(X)$, y algún estado, $x_0 \in X$, tal que*

$$\nu_1(x) \leq h_\alpha(x) \leq \nu_2(x) \tag{3.4}$$

para cada $x \in X$ y $\alpha \in (0,1)$, donde $h_\alpha(x) := V_\alpha^*(x) - V_\alpha^*(x_0)$, es llamada diferencia relativa de la función $V_\alpha^*(x)$.

La siguiente proposición es una consecuencia de la Hipótesis 3.1.3, la cual demuestra que $h_\alpha \in \mathbb{B}_w(X)$.

Proposición 3.1.1 *Supóngase que la Hipótesis 3.1.3 se cumple entonces,*

$$|h_\alpha| \leq |v_1| + |v_2|.$$

Demostración. Usando que

$$-v_2(x) \leq -h_\alpha(x) \leq -v_1(x) \leq |v_1(x)|,$$

entonces

$$-|v_1(x)| \leq h_\alpha(x) \leq v_2(x) \leq |v_2(x)|,$$

luego

$$h_\alpha(x) \leq |v_1(x)| + |v_2(x)|, \quad x \in X.$$

Análogamente se prueba que

$$-(|v_1(x)| + |v_2(x)|) \leq h_\alpha(x), \quad x \in X.$$

Por lo tanto

$$|h_\alpha| \leq |v_1| + |v_2|. \ \square$$

Observación 3.1.2 *Se observa que en la Hipótesis 3.1.3,* ν_1 *puede no ser inferiormente acotada. Esta hipótesis es dada por primera vez en [16].*

Definición 3.1.1 *Una medida* σ*-finita* m *en* $\mathcal{B}(X)$ *se dice* ***invariante****, para* $\{x_t\}$ *(o para el Kérnel* Q*) si*

$$m(B) = \int_X Q(B|x)m(dx), \quad B \in \mathcal{B}(X). \tag{3.5}$$

Usando la notación dada en [6] para medidas $\sigma - finitas$*, se tiene que la ecuación (3.5) es equivalente a*

$$m = Qm.$$

Por lo tanto, m *es medida invariante si está es punto fijo de* Q*.*

Para verificar la Hipótesis 3.3.1, en [16], se demuestra el siguiente lema.

Lema 3.1.4 *Sea* w *la función de peso dada en la Hipótesis 3.1.1. Entonces, bajo las Hipótesis 2.2.1, 3.1.1, 3.1.2 y supóngase que para cada* $f \in \mathbb{F}$*, el kérnel* $Q(\cdot|x, f(x))$ *tiene una única medida de probabilidad invariante,* m_f*, y, más aún, existe una función medible,* l_f*,* $0 < l_f < 1$ *(dependiente de* f*), en* X*, una medida de probabilidad* ν *en* X *y constantes,* $\delta_2 > 0$ *y* β_2 *con* $0 < \beta_2 < 1$*, independiente de* f*, tal que*

(a) $Q(B|x, f(x)) \geq l_f(x)$*, para cada* $B \in \mathcal{B}(S)$*,* $x \in X$*.*

(b) $\int_X l_f(y)\nu(dy) \geq \delta_2$*.*

(c) $\nu(w) := \int_X w(y)\nu(dy) < \infty$*.*

(d) $\int_X w(y)Q(dy|x, f(x)) \leq \beta_2 w(x) + l_f(x)\nu(w)$ *para cada* $x \in X$*.*

Entonces la Hipótesis 3.1.3 se satisface.

3.2. Existencia de una Política Óptima Estacionaria

En esta sección se demuestra nuestro principal resultado.

Teorema 3.2.1 *Bajo las Hipótesis 2.2.1, 3.1.1, 3.1.3 y 3.1.2, las siguientes afirmaciones son válidas.*

(a) *Existe una única constante* $g^* \in \mathbb{R}$, *dos funciones* $h_1^*, h_2^* \in \mathbb{B}_w(X)$ *y una política estacionaria* $f \in \mathbb{F}$, *las cuales satisfacen las dos desigualdades de valor óptimo promedio*

$$g^* + h_1^*(x) \leq \min_{a \in A(x)} \left\{ c(x,a) + \int_X h_1^*(y) Q(dy|x,a) \right\}, \qquad \forall x \in X. \tag{3.6}$$

$$\begin{aligned} g^* + h_2^* &\geq \min_{a \in A(x)} \left\{ c(x,a) + \int_X h_2^*(y) Q(dy|x,a) \right\} \\ &= c(x, f^*) + \int_X h_2^*(y) Q(dy|x, f^*(x)), \qquad x \in X. \end{aligned} \tag{3.7}$$

(b) $g^* = \inf_{\pi \in \Pi} u(x,\pi)$ *para cada* $x \in X$.

(c) *Cualquier política estacionaria,* $f^* \in \mathbb{F}$, *que alcanza el mínimo en (3.7) es óptima para el criterio de costo promedio.*

Demostración.
(a) Sea x_0 dada en la Hipótesis 3.1.3, y sea $\{\alpha_n\}$ una sucesión no creciente arbitraria de factores de descuento tal que $\alpha_n \to 1$ cuando $n \to \infty$. Por el Lema 3.1.3, $(1-\alpha_n)V_{\alpha_n}^*(x_0)$ es acotada para cada $n \geq 1$. Por lo tanto, existe una subsucesión $\{\alpha_{n_k}\}$ de α_n y una constante g^*, que satisface

$$\lim_{k \to \infty} (1-\alpha_k) V_{\alpha_k}^*(x_0) = g^*, \qquad h_1^*(x) := \limsup_{k \to \infty} h_{\alpha_k}(x). \tag{3.8}$$

Por la Proposición 3.1.1, $|h_{\alpha_k}| \leq |\nu_1| + |\nu_2|$, entonces $h_1^* \in \mathbb{B}_w(X)$. Así, por (2.8),

$$(1-\alpha)V_\alpha^*(x_0) + h_\alpha(x) = \min_{a \in A(x)} \left\{ c(x,a) + \alpha \int_X h_\alpha(y) Q(dy|x,a) \right\}, \quad x \in X, \tag{3.9}$$

luego,

$$(1-\alpha_k)V_{\alpha_k}^*(x_0) + h_{\alpha_k}(x) \leq c(x,a) + \int_X \alpha_k h_{\alpha_k}(y) Q(dy|x,a), \quad x \in X, a \in A(x). \tag{3.10}$$

Aplicando el Lema A.0.1 (ver Apéndice A), tomando $t \to \infty$ en (3.10) y usando (3.9), se obtiene,

$$g^* + h_1^*(x) \leq \min_{a \in A(x)} \{ c(x,a) + \int_X h_1^*(y) Q(dy|x,a) \}, \quad x \in X.$$

Por lo tanto, la desigualdad (3.6) se cumple. Ahora, para probar (3.7), sea $x \in X$ y,

$$h_2^*(x) := \liminf_{k \to \infty} h_{\alpha_k}(x) \in \mathbb{B}_w(X),$$

con

$$h_2^*(x) = \lim_{k\to\infty} g_{\alpha_k}(x),$$

donde

$$g_{\alpha_k}(x) := \inf\{h_{\alpha_m}(x) : m \geq k\} \leq h_{\alpha_k}, \qquad \forall k \geq 1.$$

Similarmente, por (2.8) y usando que $h_{\alpha_k} \geq g_{\alpha_k}$, se tiene,

$$\begin{aligned}(1-\alpha_k)V^*_{\alpha_k}(x_0) + h_{\alpha_k}(x) &= \min_{a\in A(x)}[c(x,a) + \alpha_k \int_X h_{\alpha_k}(y)Q(dy|x,a)] \\ &\geq \min_{a\in A(x)}[c(x,a) + \alpha_k \int_X g_{\alpha_k}(y)Q(dy|x,a)].\end{aligned} \tag{3.11}$$

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

$$\lim_{k\to\infty} \min_{a\in A(x)}\{c(x,a) + \alpha_k \int_X (g_{\alpha_k}(y) - g_{\alpha_1}(y))Q(dy|x,a)\},$$

y

$$\lim_{k\to\infty} \min_{a\in A(x)}\{c(x,a) + \alpha_k \int_X g_{\alpha_k}(y)Q(dy|x,a)\},$$

existen. Por (3.11), entonces

$$g^* + h_2^*(x) \geq \lim_{k\to\infty} \min_{a\in A(x)} \left\{c(x,a) + \alpha_k \int_X g_{\alpha_k}(y)Q(dy|x,a)\right\}.$$

Por otro lado, para cada $k \geq 1$ fijo, por el Lema 2.16 (a), existe $a_k \in A(x)$ (que depende de x y k) tal que

$$\min_{a\in A(x)}\{c(x,a) + \alpha_k \int_X g_{\alpha_k}(y)Q(dy|x,a\}$$

$$= c(x,a_k) + \alpha_k \int_X g_{\alpha_k}(y)Q(dy|x,a_k(x)).$$

Como $A(x)$ es compacto, existe una subsucesión $\{a_{k_i}(x)\}$, de $\{a_k(x)\}$ tal que $\lim_{i\to\infty} a_{k_i}(x)$ existe; el cual denotamos por $a_1(x) \in A(x)$. Note que $\|g_{\alpha_k}\|_w \geq \|\nu_1\|_w + \|\nu_2\|_w$ para cada $k \geq 1$, de la Hipótesis 2.2.1 y el Lema 2.2.1, se obtiene que,

$$g^* + h_2^*(x) \geq \lim_{i\to\infty} \left\{c(x,a) + \alpha_{k_i} \int_X g_{\alpha_i}(y)Q(dy|x,a)\right\}$$

$$
\begin{aligned}
&= \lim_{i\to\infty} \left[c(x, a_k(x) + \alpha_{k_i} \int_X g_{\alpha_{k_i}}(y) Q(dy|x, a_{k_i}(x)) \right] \\
&\geq c(x, a_1(x)) + \alpha_{k_i} \int_X h_2^*(y) Q(dy|x, a(x)) \\
&\geq \min_{a\in A(x)} \left\{ c(x, a) + \int_X h_2^*(y) Q(dy|x, a) \right\}.
\end{aligned}
$$

Por lo tanto la desigualdad (3.7) se satisface. Además de (3.7) junto con el Lema 2.16 (a) implica que existe $f^* \in \mathbb{F}$ tal que, se satisface la igualdad en (3.7). Por lo tanto, la prueba de la parte (a) está completa.
(b) Para cada $\pi \in \Pi$ y $x \in X$, de (3.6) se sigue que

$$
g^* + h_1^*(x_t) \leq c(x_t, a_t) + \int_X h_1^*(y) Q(dy|x_t, a_t), \tag{3.12}
$$

con lo cual, usando (1.1), y tomando esperanza en (3.12) se llega a

$$
g^* + E_x^\pi[h_1^*(x_t)] \leq E_x^\pi[c(x_t, a_t)] + E_x^\pi[h_1^*(x_{t+1})] \tag{3.13}
$$

para cada $t \geq 0$, de lo cual se sigue que,

$$
g^* + \frac{h_1^*(x)}{N} \leq \frac{E_x^\pi[\sum_{t=0}^{N-1} c(x_t, a_t)]}{N} + \frac{E_x^\pi[h_1^*(x_N)]}{N},
$$

para $N \geq 1$. Por otro lado, por el Lema 3.1.2 (a) se obtiene que,

$$
E_\pi^*[|h_1^*(x_N)|] \leq \|h_1^*\|_w \left[\tau^N w(x) + \frac{1-\tau^N}{1-\tau} b \right].
$$

Por lo tanto, $\lim_{N\to\infty} E_x^\pi[h_1^*(x_N)]/N = 0$, de lo cual, junto con (3.13),

$$
g^* \leq u(x, \pi), \quad x \in X, \pi \in \Pi.
$$

Así,

$$
g^* \leq \inf_{\pi\in\Pi} u(x, \pi), \quad x \in X. \tag{3.14}
$$

Similarmente, por (3.7) se tiene,

$$
g^* + h_2^*(x_t) \geq c(x_t, f^*(x_t)) + \int_X h_2^*(y) Q(dy|x_t, f^*(x_t)), \tag{3.15}
$$

para cada $t \geq 0$ y $x_t \in X$.
Análogamente, se demuestra usando (3.15) que,

$$
g^* \geq u(x, f^*), \quad x \in X.
$$

Por (3.14) y (3.15), se sigue que,

$$g^* = u(x, f^*) = \inf_{\pi \in \Pi} u(x, \pi),$$

completando la demostración de parte (b).
(c) La demostración de la parte (c) se sigue de las demostraciones de (a) y (b).
$\square$

Capítulo 4

Caminata Aleatoria de Lindley

En este capítulo se presenta el modelo que David Lindley, el cual fue propuesto por primera vez en [9]. El modelo es útil tanto en sistemas de inventarios como en líneas de espera (ver [16] y [13]). Primero será descrito en el contexto de sistema de inventarios y enseguida se darán las especificaciones para líneas de espera. Al final se realiza una aplicación numérica del modelo.

4.1. Descripción del Modelo en Inventarios

Un inventario es un conjunto de mércancias o artículos acumulados en un almacén en espera de ser vendidos o utilizados en un proceso de producción. En este caso estamos interesados en la modelación del flujo de mercancía en el inventario, observándolo como un sistema dinámico estocástico. Los conceptos básicos en los sistemas de inventarios son:

- Demanda. Cantidad de bienes o servicios que se ofrecen.
- Tiempo de espera. El tiempo que transcurre desde que se hace el pedido hasta que la empresa recibe el producto.
- Tamaño del pedido. Número de artículos que conforman el orden del pedido.
- Nivel de inventario. Número de artículos que se encuentran en el inventario.
- Punto de reorden. Nivel de inventario en el que la empresa define en que momento hacer un nuevo pedido.

4.2. Caminata de Lindley

Considere un modelo de inventarios con un único producto cuya dinámica está dada por:

$$x_{t+1} = (x_t - \xi_t)^+, \tag{4.1}$$

con $t \in \mathbb{N} \cup \{0\}$ y la notación $r^+ := \text{máx}(r, 0)$, donde:

- x_t representa la cantidad de cierto producto al tiempo t,
- ξ_t representa la demanda del producto durante el periodo t.

En este contexto se tiene que el espacio de estados está dado por $X := [0, \infty)$ y supóngase que $\{\xi_t\}$ es una sucesión v.a. i.i.d. definidas en $[0, \infty)$ y con esperanza finita. Se demuestra que el estado $x = 0$ es absorbente:

$$\begin{aligned} P[x_1 = 0|x_0 = 0] &= P[(-\xi_0)^+ = 0] \\ &= P[-\xi_0 < 0] \\ &= P[\xi_0 \geq 0] \\ &= 1, \end{aligned} \tag{4.2}$$

la última ecuación se debe a que la v.a. ξ_0 tiene rango en $[0, \infty)$.
La relación (4.2) demuestra que el estado 0 es absorbente, por lo cual en la práctica no es un modelo conveniente de implementar. Se busca tener un nivel mínimo en inventario con la finalidad de suplir demandas en cada periodo y no caer en pérdidas. Por esta razón, ante una demanda, se solicita cierta cantidad de producto que ingrese al sistema. Lo anterior resuelve el problema de no caer en el estado absorbente cero, sin embargo, queda por responder cual es la cantidad óptima de producto solicitado en cada periodo de observación, esta problemática puede ser abordada desde el punto de vista de la teoría de control, considerando a la variable de control como la cantidad de productos solicitados o producidos.

4.3. Caminata de Lindley Controlada

Se presenta el modelo de inventarios con control (o caminata aleatoria de Lindley controlada), el cual está dado de la siguiente forma:

$$x_{t+1} = (x_t + a_t - \xi_t)^+, \tag{4.3}$$

con $t \in \mathbb{N} \cup \{0\}$ y $x_0 = x$.
Tenemos que x_t y ξ_t representan lo mismo que en el caso sin control, y a_t denota la cantidad de producto ordenada que se proporciona al principio del periodo t.
El espacio de estados es $X = [0, \infty)$ y supóngase que la cantidad de producto

ordenada a_t se encuentra en el intervalo $A := [0, \theta]$, para alguna constante $\theta > 0$, independientemente del nivel de stock, es decir, que el conjunto de acciones admisibles es $A(x) = A$ para cada $x \in X$.
Además la sucesión de v.a. i.i.d. $\{\xi_t\}$ cumple que:

- cada variable aleatoria ξ_t tiene rango en $[0, \infty)$,
- si ξ es un valor genérico de la sucesión de v.a. i.i.d. $\{\xi_t\}$, entonces tienen una función de densidad continua y acotada Δ con función de distribución F,
- el valor esperado de la demanda es finito y se supone mayor que la cantidad de producto solicitado, es decir, $\theta < \mu := E[\xi] < \infty$.

Finalmente, para completar la descripción del modelo de control dado en la Definición (1.1.1), introduciremos una función de costo en un paso, la cual considera costos del tipo siguiente:

costo de producción + costo de almacenaje - ingresos de ventas,

de esta forma consideremos la siguiente función de costo:

$$c(x, a) = pa + m(x + a) - kE[\min(x + a, \xi)], \tag{4.4}$$

donde

- p es el costo de producción por unidad,
- m es el costo de almacenaje por unidad,
- k es el precio de venta por unidad,

con k, p, m constantes positivas y además,

$$m + p \leq k. \tag{4.5}$$

La ecuación (4.5) nos dice que no podemos tener costos por unidad mayores al precio de venta por unidad.

Observación 4.3.1 *Note que la función de costo dada en (4.4) no necesariamente es positiva. Si la función toma valores negativos en la práctica se tendrían ganancias.*

Observación 4.3.2 *En líneas de espera, las variables de la dinámica dada en la ecuación (4.3), denotan*

- x_t *número de clientes en el sistema al tiempo* t*,*
- a_t *número de clientes a los cuales se les permite acceso al sistema,*
- ξ_t*, número de servicios concluidos al tiempo* t*.*

Y, para la función de costo dada en (4.4),

- p *es el costo por arribo de clientes,*
- m *es el costo de servicio,*
- k *es la ganancia por clientes atendidos.*

Es importante estudiar los casos de cola finita e infinita de sistemas de espera. En el caso de capacidad finita es importante controlar la entrada de clientes ya que un sobrecupo puede provocar inestabilidad en el sistema. Por otro lado, en el caso de capacidad infinita, al admitir a todos los clientes que buscan servicio, si se considera la función de costos por clientes en el sistema resultaría en la práctica no rentable aceptar a todos los clientes.

Se procede a demostrar que las Hipótesis 2.2.1, 3.1.1 y 3.1.2 se satisfacen.
La Hipótesis 2.2.1 (a) se cumple ya que el conjunto de acciónes admisibles $A(x) = [0, \theta]$ es compacto para cada $x \in X$.
La función de costo es continua, en efecto, recordando que

$$\text{mín}(x, y) = \frac{x + y - |x - y|}{2}, \qquad x, y \in \mathbb{R}.$$

Así,

$$\begin{aligned} kE[\text{mín}(x + a, \xi)] &= kE[\frac{x + a + \xi - |x + a - \xi|}{2}] \\ &= \frac{k}{2}(x + a) + \frac{k}{2}\mu - \frac{k}{2}E[|x + a + \xi|]. \end{aligned} \tag{4.6}$$

con $(x, a) \in \mathbb{K}$. De este modo, solo es necesario demostrar que la función $g(x, a) := E[|x+a+\xi|]$ es continua, lo cual se garantiza en la siguiente proposición.

Proposición 4.3.1 *La función* g *es continua en* $\mathbb{K}$*.*

Demostración.
Considere $\{x_n\}$ y $\{a_n\}$ sucesiones convergentes en X y A, con límites x y a, respectivamente. Ahora, definimos h_n y h como:

$$\begin{aligned} h_n(s) &= |x_n + a_n - s|\Delta(s), \\ h(s) &= |x + a - s|\Delta(s), \end{aligned}$$

note que $h_n(s) \to h(s)$, cuando $n \to \infty$ para $s \in [0, \infty)$, pues Δ es no negativa. Por otra parte,

$$\begin{aligned} h_n(s) &\leq (|x_n| + |a_n| + s)\Delta(s) \\ &\leq (M + a)\Delta(s)\ . \end{aligned}$$

La última desigualdad se cumple ya que $\{x_n\}$ y $\{a_n\}$ son convergentes, por tanto, son acotadas ambas por alguna constante positiva M, así

$$\begin{aligned} \int_X h_n(s)\Delta(s)ds \leq \int_X (M + s)\Delta(s)ds \\ = M + \mu < \infty. \end{aligned}$$

Ahora por el Teorema de Convergencia Dominada, tenemos

$$\begin{aligned} \lim_{n\to\infty} g(x_n, a_n) &= \lim_{n\to\infty} \int_X |x_n + a_n - s|\Delta(s)ds \\ &= \int_X \lim_{n\to\infty} |x_n + a_n|\Delta(s)ds \\ &= \int_X |x + a - s|\Delta(s)ds \\ &= g(x, a). \end{aligned}$$

Así, g es continua en $\mathbb{K}$. □
De la Proposición 4.3.1, se garantiza que la función de costo, (4.4), es continua en $\mathbb{K}$, con lo cual la Hipótesis 2.2.1 (b) se cumple.
Procederemos a demostrar que la Hipótesis 2.2.1 (c) se satisface.

Proposición 4.3.2 *La función* $\mu'(x, a) = \int_X \mu(y)Q(dy, x, a)$ *es continua en* $A(x)$*, para cada* $x \in X$.

Demostración.
Sean $x \in X$ fijo y $\mu \in \mathbb{B}(X)$, por (1.2) para $a \in A(x)$ se tiene,

$$\begin{aligned} \mu(x, a) &= \int_0^\infty \mu[(x + a - s)^+]ds \\ &= \mu(0)[1 - F(x + a)] + \int_0^{x+a} \mu(x + a - s)\Delta(s)ds\ . \end{aligned} \tag{4.7}$$

Haciendo un cambio de variable en la última integral obtenemos que

$$\mu'(x, a) = \mu(0)[1 - F(x + a)] + \int_0^{x+a} \mu(s)\Delta(x + a - s)ds\ ,$$

para cada $a \in A(x)$. Sea $\{a_n\}$ una sucesión convergente en $A(x)$, donde su límite es $a \in A(x)$, luego

$$\lim_{n\to\infty} \mu'(x,a_n) = \lim_{n\to\infty} \left[\mu(0)[1 - F(x+a_n)] + \int_0^{x+a_n} \mu(s)\Delta(x+a_n-s)ds \right]$$

$$= \lim_{n\to\infty} \left[\mu(0)[1 - F(x+a_n)] + \int_0^{\infty} I_{[0,x+a_n]}\mu(s)\Delta(x+a_n-s)ds \right] .$$

Como Δ es continua y acotada entonces F también lo es, en consecuencia,

$$\lim_{n\to\infty} \mu'(x,a_n) = \mu(0)[1 - F(x+a)] + \lim_{n\to\infty} \int_0^{\infty} I_{[0,x_n,a_n]}u(s)\Delta(x+a_n-s)ds.$$

Debido a que se satisface la siguiente propiedad,

$$\liminf[0, x + a_n] \subset \limsup[0, x + a_n] \subset [0, x + a],$$

se tiene que $I_{[0,x+a_n]}$ converge a $I_{[0,x+a]}$ casi seguramente, por tanto,

$$\lim_{n\to\infty} \mu'(x,a_n) = \mu(x,a),$$

es decir, μ' es continua en $A(x)$. $\square$
Por lo tanto por la Proposición 4.3.2 y la Hipótesis 2.2.1 (c) se satisface.

Proposición 4.3.3 *Para el modelo de control de Lindley, la Hipótesis 3.1.1 se satisface.*

Demostración.
Para demostrar que la Hipótesis 3.1.1 se cumple, se tiene que encontrar una función de peso $w : X \to [1,\infty)$ que satisfaga las condiciones dadas en la Hipótesis 3.1.1.
Para esto considere la función generadora de momentos ψ_a de la variable aleatoria $z_a := a - \xi$, con $a \in A(x)$,

$$\psi_a(r) = E[exp(r(a-\xi))], \quad r \geq 0.$$

Como $\psi_a(0) = 1$ y $\psi'_a(0) < 1$ entonces existe un número positivo ρ tal que

$$\psi_a(\rho) < 1.$$

Se define

$$w(x) := exp(\rho x), \tag{4.8}$$

para cada $x \in X$. Note que $w \geq 1$, ya que $\rho \geq 0$ y $x \geq 0$. Entonces, de (4.7) con $\mu := w$, se tiene

$$w'(x,a) = w(0)[1 - F(x+a)] + w(x)\int_0^{x+a} exp(\rho(a-s))\Delta(s)ds,$$

como $w(0) = 1$, $[1 - F(x+a)] \leq 1$ y $\rho(a-s) \leq \rho(\theta - s)$, para cada $a \in A$, se obtiene

$$\begin{aligned} w'(x,a) &\leq [1 - F(x+a)] + \psi_\theta(\rho)w(x) \\ &\leq \tau w(x) + b, \forall x \in X, \end{aligned} \tag{4.9}$$

con

$$\tau := \psi_\theta(\rho), \quad b := 1.$$

Por lo tanto, la Hipótesis 3.1.1 (a), se cumple. Por otro lado, usando (4.4) se tiene,

$$\begin{aligned} |c(x,a)| &\leq |pa + m(x+a) - E[\text{mín}(x+a,\xi)]| \\ &\leq (p+m)a + mx + \mu \\ &\leq (p+m)\theta + mx + \mu. \end{aligned}$$

Por (4.5) y como $\theta < \mu$,

$$\sup_{a \in A(x)} |c(x,a)| \leq k(x + 2\mu),$$

para cada $x \in X$. Por lo tanto, para una constante positiva M_1 suficientemente grande,

$$\sup_{a \in A(x)} |c(x,a)| \leq M_1 \exp(\rho(x + 2\mu)),$$

para cada $x \in X$. Sea $M := M_1 \exp(2\rho\mu)$, por tanto,

$$|c(x,a)| \leq Mw(x), \quad x \in X.$$

Así, la Hipótesis 3.1.1 (b) se satisface. □
Demostrar la Hipótesis 3.1.2 (c), es análogo a la demostración dada en la Proposición 4.3.2, tomando $\mu = w$. Con lo cual se ha demostrado el siguiente teorema.

Teorema 4.3.1 *Con las condiciones dadas en el modelo de Lindley controlado (4.3), se tiene que se satisfacen las Hipótesis 2.2.1, 3.1.1 y 3.1.2 entonces existe una política óptima estacionaria que minimiza el criterio de costo descontado con horizonte infinito.*

Ahora se demostrará que el modelo de control de Lindley dado en (4.3), satisface la Hipótesis 3.1.3. Para esto, usaremos el Lema 3.1.4.
El siguiente lema demuestra que para éste modelo de control existe una medida de probabilidad invariante m_f.

Lema 4.3.1 *Para cada $f \in \mathbb{F}$, sea $\{x_t^f, t = 0, 1, ...\}$ la cadena de Markov definida en (4.3), cuando $a_t = f(x_t)$, para cada t, es decir,*

$$x_{t+1}^f = [x_t^f + f(x_t) - \xi_t]^+, \quad t \in \mathbb{N}_0, \tag{4.10}$$

con $x_0^f = x \in X$, entonces para cada $f \in \mathbb{F}$, $\{x_t^f\}$ es recurrente positiva, entonces por la Proposición E.0.10 (ver Apéndice E), tiene una única medida de probabilidad invariante m_f.

Demostración.
Sea $\{x_t\}$ la cadena de Markov dada en ecuación (4.10) cuando $f(x) := \theta$ para cada estado x, es decir,

$$x_{t+1}^\theta = (x_t^\theta + \theta - \xi_t)^+, \quad t \in \mathbb{N}_0. \tag{4.11}$$

Definiendo $y_t := \theta - \xi_t$ la ecuación (4.11) se transforma a una caminata aleatoria de la forma:

$$x_{t+1}^\theta = (x_t^\theta + y_t)^+. \tag{4.12}$$

Por lo tanto, como $E[|y_0|] \leq \theta + \mu < \infty$ y $E[y_0] = \theta - \mu < 0$, la cadena de Markov $\{x_t^\theta\}$ es recurrente positiva. En particular, $E_0[\tau^\theta] < \infty$, donde τ^θ denota el tiempo de alcance al estado $x = 0$, dado el estado inicial $x_0^\theta = 0$. Sea $f \in \mathbb{F}$ y sea τ^f el tiempo de alcance al estado $x = 0$ dado $x_0^f = 0$. Como $A(x) = [0, \theta]$ se tiene que $f(x) \leq \theta$ para cada $x \in X$, y en consecuencia, $x_t^f \leq x_t^\theta$ para cada $t \in \mathbb{N}_0$. Entonces

$$E_0[\tau^f] \leq E_0[\tau^\theta] < \infty, \tag{4.13}$$

por lo tanto, $\{x_t^f\}$ es recurrente positiva. Dado $f \in \mathbb{F}$ arbitrario, usando la Proposición E.0.10 (ver Apéndice E), se tiene que existe una medida de probabilidad invariante m_f para el modelo de control de Lindley.
Más aún, sea $f \in \mathbb{F}$ y δ la medida de Dirac en $x = 0$, se define

$$l_a(x) := \bar{F}(x + a), \quad x \in X, a \in A(x).$$
$$l_f(x) := \bar{F}(x + f(x)) \quad x \in X.$$

Como $0 \leq f(x) \leq \theta$ para cada $x \in X$, entonces

$$\begin{aligned} l_f(x) &:= \bar{F}(x + f(x)) \\ &= P[x + f(x) \leq \xi] \\ &\geq P[x + \theta \leq \xi] \\ &= \bar{F}(\theta + x). \end{aligned}$$

Por lo tanto,

$$l_f(x) \leq \bar{F}(\theta + x),$$

para cada $x \in X$. Usando (1.2) se tiene para cada $B \in \mathcal{B}(X)$ y $x \in X$,

$$\begin{aligned} Q(B|x, f(x)) &= E[I_B((x + f(x) - \xi)^+)] \\ &= \int_X I_B((x + f(x) - s)^+)\Delta(s)ds \\ &= \int_0^{x+f(x)} I_B(x + f(x) - s)\Delta(s)ds + \int_{x+f(x)}^{\infty} I_B(0)\Delta(s)ds. \end{aligned}$$

Como, $\int_{x+f(x)}^{\infty} I_B(0)\Delta(s)ds \geq 0$, se sigue que,

$$\begin{aligned} Q(B|x, f(x)) &\geq I_B(0) \int_{x+f(x)}^{\infty} \Delta(s)ds \\ &= \delta(B)\bar{F}(x + f(x)) \\ &= \delta(B)l_f(x). \end{aligned}$$

Así, $Q(B|x, f(x)) \geq \delta(B)l_f(x)$. Por otro lado, por (4.9),

$$\int_X w(y)Q(dy|x, f(x)) \leq \psi_\theta(\rho)w(x) + l_f(x),$$

para cada $x \in X$. Luego, por propiedades de la medida de Dirac cuando $x = 0$, véase el Lema E.0.9 (ver Apéndice E),

$$\begin{aligned} \int_X l_f(y)\delta(dy) &= l_f(0) \\ &\geq \bar{F}(\theta) > 0, \end{aligned}$$

para cada $f \in \mathbb{F}$. Además,

$$\delta(w) := \int_X w(y)\delta(dy) = 1,$$

de lo cual, las Hipótesis del Lema 3.1.4 se satisfacen.
En resumen tenemos el siguiente teorema.

Teorema 4.3.2 *Con las condiciones dadas en el modelo de Lindley controlado (4.3), se tiene que se satisfacen las Hipótesis 2.2.1, 3.1.1, 3.1.2 y 3.1.3 entonces existe una política óptima estacionaria que optimiza el criterio de costo promedio.*

4.4. Aplicación Numérica

En esta sección se darán aproximaciones numéricas a la función de valor óptimo (ver Definición 1.5), para los criterios en estudio de este trabajo.

4.4.1. Políticas Óptimas con Criterio de Costo Descontado

Primeramente se buscan políticas en Π_{DM}, para esto, se toma el algoritmo de iteración de valores óptimos con la función de costo dada en (4.4), para cada $n \geq 1$ y $x \in X$,

$$v_n(x) = \min_{a\in[0,\theta]} \left\{pa + m(x+a) - kL(x+a) + \alpha E[v_{n-1}(x+a-\xi_t)^+]\right\}, \quad (4.14)$$

con $v_0(\cdot) \equiv 0$, donde

$$L(y) = kE[\min\{y, \xi\}] = ky[1 - F(y)] + k\int_0^y s\Delta(s)ds \quad (4.15)$$

y

$$E[v_{n-1}(x+a-\xi_t)^+] = \int_X v_{n-1}(y)Q(dy|x,a).$$

La ecuación (4.14), es equivalente a

$$v_n(x) = \min_{y\in[0,\theta+x]} \left[(p+m)y - L(y) + \alpha E[v_{n-1}(x+a-\xi)^+]\right] - px, \quad (4.16)$$

donde $y := x + a$, para cada $x \in X$, $a \in A$.
Se define $G_n(y) := py + my - L(y) + \alpha E[v_{n-1}(y-\xi)^+]$, para cada $n \in \mathbb{N}$. El siguiente teorema demuestra que cada función G_n es convexa.

Proposición 4.4.1 *Para cada $n \in \mathbb{N}$, G_n es una función convexa en X.*

Demostración.
La prueba de hará por inducción. Para $n = 1$,

$$G_1(y) = (p+m)y - L(y), \quad (4.17)$$

para cada $y \in [x, \theta+x]$ es una función convexa, ya que la función $L(y)$ es cóncava y por tanto, $-L(y)$ es convexa en $[0, \theta+x]$ y la recta $(p+m)y$ tambien lo es, así,

G_1 es convexa en $[x, x+\theta]$.
Supongamos que v_{n-1} es convexa, entonces para n tenemos,

$$\begin{aligned} v_n(x) &= \min_{a\in[0,\theta]} [pa + m(x+a) - L(x+a) + \alpha E[v_{t-1}(x+a-\xi_t)^+]] \\ &= \min_{y\in[x,\theta+x]} \big[(p+m)y - L(y) + \alpha E[v_{t-1}(y-\xi)^+]\big] - px, \end{aligned}$$

donde $y = x+a$. Se debe mostrar que

$$G_n(y) = (p+m)y - kL(y) + \alpha E[v_{n-1}(y-\xi)^+],$$

$y \in [x, x+\theta]$ y $x \in X$ es convexa. Sabemos que $(p+m) - L(y)$ es una función convexa, por el caso $n=1$, solo falta ver que $E[v_{t-1}(y-\xi)^+]$ también lo es. Para ello sea,

$$\begin{aligned} W(y) &= E[v_{n-1}(y-\xi)^+], \\ &= \int_X [v_{n-1}(y-s)^+]\Delta(s)ds. \end{aligned}$$

Sean y_1, $y_2 \in [0,\infty)$ y $0 < \lambda < 1$, entonces,

$$\begin{aligned} W(\lambda y_1 + (1-\lambda)y_2) &= \int_X [v_{n-1}(\lambda y_1 + (1-\lambda)y_2 - s)^+]\Delta(s)ds \\ &= v_{n-1}(0)[1 - F(\lambda y_1 + (1-\lambda)y_2)] \\ &+ \int_0^{\lambda y_1 + (1-\lambda)y_2} v_{n-1}(\lambda y_1 + (1-\lambda)y_2)s\Delta(s)ds, \end{aligned}$$

como $s = \lambda s + (1-\lambda)s$ $v_{n-1}(0) = \lambda v_{n-1}(0) + (1-\lambda)v_{n-1}(0)$ y usando la hipótesis inductiva que v_{n-1} es una función convexa tenemos,

$$\begin{aligned} W(\lambda y_1 + (1-\lambda)y_2) &= \lambda \int_X v_{n-1}(y_1-s)^+\Delta(s)ds + (1-\lambda)\int_X v_{n-1}(y_2-s)^+\Delta(s)ds \\ &= \lambda W(y_1) + (1-\lambda)W(y_2). \quad \square \end{aligned}$$

Con la ayuda del resultado anterior, se busca una política óptima en Π_{DM}. Para esto, se retoma la ecuación (4.16), como cada función G_n es convexa, entonces se busca el punto crítico en cada función $G_n, n \in \mathbb{N}$ y éste será mínimo. Así, para $n=1$, tenemos de (4.14),

$$v_1(y) = \min_{y\in[x,\theta+x]} \{G_1(y)\} - px.$$

Derivando $G_1(y)$ con respecto a y, se obtiene,

$$G_1'(y) = p + m - L'(y).$$

Por consiguiente, usando (4.15) se tiene que

$$\begin{aligned} L'(y) &= k[1-F(y)] - ky\Delta(y) + ky\Delta(y). \\ &= k[1-F(y)]. \end{aligned}$$

Por lo tanto,

$$G_1'(y) = p + m - k[1-F(y)]. \tag{4.18}$$

Igualando a cero, la ecuación (4.18),

$$F(s_1) = (\frac{k-(p+m)}{k}).$$

Dado que F es creciente, existe F^{-1}, así el punto,

$$s_1 = F^{-1}(\frac{k-(p+m)}{k}),$$

minimiza a G_1, y por la Proposición E.0.5 (ver Apéndice E) se tiene que el valor mínimo de v_1 es y^*,

$$y^* = \begin{cases} x, & si\ s_1 < x, \\ s_1, & si\ s_1 \in [x, x+\theta], \\ \theta + x, & si\ s_1 > \theta + x. \end{cases}$$

Así, haciendo el cambio de variable,

$$f_1(x) = \begin{cases} 0, & si\ s_1 < x, \\ s_1 - x, & si\ s_1 \in [x, x+\theta], \\ \theta, & si\ s_1 > \theta + x. \end{cases}$$

Sustituyendo en $v_1(x) = \text{mín}_{a\in[0,\theta]} \{pa + m(x+a) - L(x+a)\}$, se tiene,

$$v_1(x) = \begin{cases} mx + L(x), & si\ s_1 < x, \\ p(s_1 - x) + ms_1 - L(s_1), & si\ s_1 \in [x, x+\theta], \\ p\theta + m(\theta + x) - L(x+\theta), & si\ s_1 > \theta + x. \end{cases}$$

Por otra parte, para cada $n \in \mathbb{N}$, la función $G_n(y)$, $n \in \mathbb{N}$ es convexa para cada $y \in [x, x+\theta]$ y se tiene un mínimo en un punto s_n debido a la Proposición 4.4.1. Por lo tanto por el Lema E.0.5 (ver Apéndice E), al minimizar cada función G_n, se obtiene que los minimizadores son,

$$f_n(x) = \begin{cases} 0, & si\ s_n < x, \\ s_n - x, & si\ s_n \in [x, x+\theta], \\ \theta, & si\ s_n > \theta + x. \end{cases} \tag{4.19}$$

y

$$v_n(x) = \begin{cases} mx - L(x) + \alpha E[v_n(x-\xi_n)^+], & si\ s_n < x, \\ p(s_n - x) + ms_n - L(s_n) + \alpha E[v_{n-1}(s_n - \xi_n)^+], & si\ s_n \in [x, x+\theta], \\ p\theta + m(\theta + x) - L(\theta + x) + \alpha E[v_{n-1}(\theta + x - \xi_n)^+], & si\ s_n > \theta + x. \end{cases}$$

4.4.2. Aproximación Numérica al Valor Óptimo Descontado y Promedio

Considere un sistema de inventario con demandas distribuidas exponencialmente con parámetro λ. En [5] se justifica que en la práctica tiene sentido tomar este tipo de demanda.
El inventario tiene una capacidad $\theta = 10$, entonces el espacio de acciones y acciones admisibles estan dados por

$$A = A(x) = [0, 10], \quad x \in X.$$

Supóngase que $\lambda = 2$, es decir, en promedio se venden 2 productos por cada tiempo en el que se observa el sistema. Los parámetros de la función de costo dada en (4.4), son $k = 60$, $p = 30$ y $m = 20$, por lo tanto, la función de costo es de la forma,

$$c(x, a) = 30p + 20(x + a) - 60E[\min(x + a, \xi)].$$

Usando el software *Mathematica 10.1*, se aproximaron los valores s_n dados en (4.19), de tal forma que buscamos una $N \in \mathbb{N}$ tal que,

$$|v_N(x_0) - v_{N-1}(x_0)| < \epsilon,$$

donde $x_0 = 0$ es el estado inicial, o el número de productos que tenemos al inicio en el inventario y ϵ es el error de aproximación. Por tanto, para los siguientes valores de ϵ, se obtuvo lo siguiente.

ϵ	N	$v_N(x_0)$	s_N
.0001	3	-0.488312	0.0881961
.00001	8	-0.488319	0.0881961

En este ejemplo, el valor s_N, se estabiliza a partir de $N = 3$. Se procede a realizar una simulación del modelo de Lindley controlado. En la siguiente tabla se muestran los valores de la función de costo de algunos estados del sistema.

Tiempo	x_t	Costo
1	12.45	719.13
2	22.31	916.39
3	31.59	1101.98
5	50.52	1480.46
10	99.58	2460.49
20	187.30	4216.08

Donde la primera columna representa el tiempo t, la segunda x_t, el valor del estado en el tiempo t y la tercera columna el costo correspondiente en cada tiempo presentado. Para terminar esta sección, se busca la función de valor óptimo promedio (1.8). Por el Teorema 3.2.1, se tiene que,

$$h^* = \inf_{\pi \in \Pi} u(x, \pi), \qquad x \in X. \tag{4.20}$$

donde,

$$h^* = \lim_{k \to \infty} (1 - \alpha_k) V^*_{\alpha_k}(x_0). \tag{4.21}$$

Para este ejemplo, numéricamente se encontró que los factores de descuento α, cuando son cercanos a 1, se tiene que los minimizadores s_n se estabilizan a algún s a partir del tiempo $N = 9$, como se describe en la siguiente tabla.

α	s
0.9	0.180507
0.99	0.200263
0.999	0.202483
0.9999	0.202708

Tomando en consideración la anterior tabla, podemos establecer para valores α cercanos a 1, que el minimizador se estabiliza aproximadamente en $s = 0.\,202$, con error $\epsilon = 0.\,01$ en el tiempo $N = 10$, con lo cual se obtiene,

$$h^* = -7.\,05707, \tag{4.22}$$

con el estado inicial $x_0 = 0$. Por lo tanto,

$$\bar{V}(x) = -7.\,05707. \tag{4.23}$$

Por lo tanto, al ser un valor negativo en la ecuación (4.23), el valor promedio óptimo arroja ganancias para este ejemplo.

Capítulo 5

Conclusiones

El trabajo de tesis desarrolla la teoría básica de PDM para el estudio de la caminata aleatoria de Lindley. Se trabajarón dos criterios de rendimientos: costo descontado y costo promedio, con ellos se válidaron las hipótesis para la existencia de políticas óptimas estacionarias que minimicen los criterios estudiados. Para resolver el problema de control óptimo promedio se requiere una hipótesis adicional a las ya dadas en el problema de control óptimo descontado, el cual es la estabilidad del proceso de Markov, o en otras palabras la existencia de una medida invariante para el proceso dado. Se validan las hipótesis para el modelo de control de Lindley, por tanto se verifica que para este modelo existen políticas óptimas descontadas y promedio. Se dan aproximaciones numéricas para la función de valor óptimo descontado y promedio. Por último se presentó un ejemplo numérico aplicado a teoría de inventarios.

Los problemas que se consideran como consecuencia del trabajo, y se espera continuar su análisis, son los siguientes.

- Caracterizar las políticas óptimas para el criterio de costo promedio, usando las políticas encontradas en el caso de criterio descontado, ya que en este trabajo se presenta la relación que existen entre las funciones de valores óptimos de estos dos criterios.

- Estudiar la estabilidad del modelo de control de Lindley con criterio de costo descontado.

- Analizar el modelo de control de Lindley considerando el conjunto de acciones A no compacto.

- Generalizando el modelo de control de Lindley, agregando una nueva sucesión de v.a. i.i.d., $\{\mu_t\}$, la cual en el modelo de inventarios significa la cantidad base de producto ordenado.

$$x_{t+1} = (x_t + a_t\mu_t - \xi_t)^+, \qquad t \in \mathbb{N},$$

  se pretende estudiar las condiciones con las cuales podamos garantizar los resultados dados en este trabajo y estudiar la estabilidad del modelo.

Apéndice A

Espacio de Funciones con Norma Ponderada

Sea X un espacio de Borel, y sea $\mathbb{B}(X)$ el espacio de Banach de funciones medibles acotadas μ en X, con la norma supremo,

$$\|\mu\|_\infty := \sup_{x\in X}\{|\mu(x)|\}.$$

Sea una función medible $w : X \longrightarrow [1,\infty)$ a la cual llamaremos función de peso. Si μ es una función real sobre X se define su $w-norma$ como,

$$\| \mu \|_w := \| \frac{\mu}{w} \|_\infty = \sup_{x\in X}\left\{\frac{|\mu(x)|}{w(x)}\right\}. \tag{A.1}$$

Es claro que si $w \equiv 1$ entonces $w-norma$ y la norma supremo coinciden. Se dice que una función μ es acotada si $\| \mu \|_\infty < \infty$ y $w-acotada$ si $\| \mu \|_w < \infty$. Además w es $w-acotada$ ya que $\| w \|_w = 1$, por lo que w generalmente puede ser una función no acotada pero si $w-acotada$. Por otro lado, si μ es una función acotada y como $w \geq 1$,

$$\| \mu \|_w \leq \| \mu \|_\infty < \infty,$$

entonces μ es $w-acotada$.
Sea $\mathbb{B}_w(X)$ el espacio normado de funciones medibles μ en X que sea $w-acotada$ o llamadas funciones con norma ponderada. Este espacio es completo ya que si $\{\mu_n\}$ es una sucesión de Cauchy con la $w-norma$, entonces $\{\frac{\mu_n}{w}\}$ es sucesión de Cauchy con la norma uniforme, como $\mathbb{B}(X)$ es un espacio de Banach, entonces se puede encontrar una función $\mu \in \mathbb{B}(X)$ la cual sea $w-$ límite de $\{\mu_n\}$. Por lo tanto hemos demostrado la siguiente proposición.

Proposición A.0.1 *$\mathbb{B}_w(X)$ es un espacio de Banach que contiene a $\mathbb{B}(X)$.*

Lema A.0.1 ***Una extensión del Lema de Fatou***

Sea $\{u_n\}$ un sucesión $w-$acotada en $B_w(X)$ y se define

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

Entonces para cualquier estado $x \in X$ y toda sucesión $\{a_n\}$ en $A(x)$ tal que $a_n \to a$ en $A(x)$, tenemos

$$\liminf_{n\to\infty} \int u_n(y)Q(dy|x,a_n) \geq \int u^I(y)Q(dy|x,a),$$

y

$$\limsup_{n\to\infty} \int u_n(y)Q(dy|x,a_n) \leq \int u^S(y)Q(dy|x,a).$$

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

$$\lim_{n\to\infty} \int u_n(y)Q(dy|x,a_n) = \int u(y)Q(dy|x,a).$$

Apéndice B

Multifunciones y Selectores

Sean X y A espacios de Borel no vacíos.

Definición B.0.1 *Una multifunción ψ de X a A es una función tal que para cada $x \in X$ su imagen $\psi(x)$ es un subconjunto no vacío de A, es decir $\psi : X \longrightarrow \boldsymbol{Pot}(A)$, donde **Pot** denota al conjunto potencia de A. La gráfica de ψ es el subconjunto de $X \times A$ definido como*

$$graf(\psi) = \{(x, a) | x \in X, a \in \psi(x)\}.$$

Definición B.0.2 *Sea ψ una multifunción de X a A, se dice que ψ es:*

(a) Borel Medible, si $\psi^{-1}(B)$ es Borel medible en X para cada conjunto abierto B en A.

(b) Semicontinua superiormente (u.s.c.), si $\psi^{-1}(C)$ es cerrado en X para cada conjunto cerrado C en A.

(c) Semicontinua inferiormente (l.s.c.), si $\psi^{-1}(B)$ es abierto en X para cada conjunto abierto B en A.

(d) Cerrada, si $\psi(x)$ es cerrado para cada $x \in X$.

(e) Compacta, si $\psi(x)$ es compacta para cada $x \in X$.

Suponemos que la multifunción es Borel medible, $\nu : graf(\psi) \longrightarrow \mathbb{R}$ es una función medible, y para cada $x \in X$,

$$\nu^*(x) := \inf_{a \in \psi(x)} \nu(x, a).$$

Además, si $\nu(x, \cdot)$ alcanza su mínimo en algún punto de $\psi(x)$, escribimos mín en lugar de ínf.

Apéndice C

Teorema de Punto Fijo de Banach

Definición C.0.3 *Sea (X, d) un espacio métrico completo y f una aplicación. Se dice que f es contractiva si existe una constante K, a la cual se le llama módulo, con $0 < K < 1$, tal que*

$$d(f(x), f(y)) \leq Kd(x, y)$$

para cualesquiera $x, y \in X$.

Definición C.0.4 *Sea (X, d) un espacio métrico completo y f una aplicación. Un punto fijo $x_0 \in X$ de f es tal que, $f(x_0) = x_0$.*

Teorema C.0.1 ***Teorema de punto fijo de Banach*** *Sea (X, d) un espacio métrico completo y f una aplicación sobre X. Entonces existe un único punto fijo de f. Más aún,*

$$d(f^n(x), x_0) \leq K^n d(x, x_0).$$

Proposición C.0.2 *Sea T un operador monótono sobre $\mathbb{B}_w(X)$. Si existe un $\gamma < 1$ tal que*

$$T(u + rw) \leq T(u) + \gamma rw, \qquad \forall u \in \mathbb{B}_w(X).$$

entonces T es un operador contracción con módulo γ.

Demostración Sean $u, v \in \mathbb{B}_w(X)$ entonces $u \leq u + w\cdot \parallel u - v \parallel_w$. Luego, por la monótocidad de T y tomando $r = \parallel u - v \parallel_w$, se tiene que

$$T(u) \leq T(v + rw) \leq T(v) + \gamma rw,$$

lo cual implica que

$$T(u) - T(v) \leq \gamma w \parallel u - v \parallel_w,$$

análogamente se prueba que

$$T(v) - T(u) \geq \gamma w \parallel u - v \parallel_w .$$

Así,

$$|T(u) - T(v)| \leq \gamma w \parallel u - v \parallel_w .$$

Por lo tanto,

$$\parallel T(u) - T(v) \parallel_w \leq \gamma w \parallel u - v \parallel_w .$$

Apéndice D

Kérneles Estocásticos

Sean X y A espacios de Borel.

Definición D.0.5 *Un kérnel estocástico sobre X dado A, es una función $P(\cdot|\cdot)$ tal que:*

(a) $P(\cdot|y)$ es una medida de probabilidad sobre X para cada $y \in Y$,

(b) $P(B|\cdot)$ es una función medible sobre A para cada $B \in \mathcal{B}(X)$.

$\mathbb{B}(X)$ denota el conjunto de funciones medibles y acotadas sobre X.

Teorema D.0.2 *Si $\nu \in \mathbb{X} \times \mathbb{Y}$, entonces la función,*

$$g(y) := \int_X \nu(y)P(dx|y) \in \mathbb{B}(Y).$$

Definición D.0.6 *El kérnel estocástico $P \in P(X|A)$ es,*

1. *Débilmente continuo si la función $y \to \int_X \nu(x)P(dx|y)$ es continua y acotada para cualquier función continua y acotada en X.*

2. *Fuertemente continuo si la función $y \to \int_X \nu(x)P(dx|y)$ es continua y acotada para cualquier función ν acotada en X.*

Es claro que fuertemente continuo implica débilmente continuo.

Proposición D.0.3 *Las siguientes proposiciones son equivalentes:*

1. *P es fuertemente continuo.*

2. *La función $y \to \int_X \nu(x)P(dx|y)$ es l.s.c., para cada ν acotada.*

Proposición D.0.4 ***(Teorema de Ionescu-Tulcea)***
Sea $X_0, X_1, ...,$ una sucesión de espacios de Borel y, para $n \in \mathbb{N}_0$, definimos $Y_n := X_0 \times X_1 \times \cdots \times X_n$ y $Y := \prod_{n=0}^{\infty} X_n$. Sea ν una medida de probabilidad arbitraria

sobre X_0 *y , para cada* $n \in \mathbb{N}_0$, $P_n(dx_{n+1}|y_n)$ *es un kérnel estocástico sobre* X_{n+1} *dado* Y_n*. Entonces existe una única medida de probabilidad* P_ν *sobre* Y *tal que, para cada rectángulo medible* $B_0, \times \cdots \times B_n$ *en* Y_n*,*

$$P_\nu(B_0 \times \cdots \times B_n) = \int_{B_0} \nu(dx_0) \int_{B_1} P_0(dx_1|x_0) \cdots \int_{B_n} P_{n-1}(dx_n|x_0, ..., x_{n-1}).$$

Además, para cualquier función u *medible y no negativa sobre* Y*, la función*

$$x \to \int_X u(y) P_x(dy),$$

es medible en X_0*, donde* P_x *representa a* P_ν *cuando* ν *es la probabilidad concentrada en* $x \in X_0$*.*

Demostración. Véase [1].

Apéndice E

Miscelánea

Definición E.0.7 *Sean (X, d) un espacio métrico y $\nu : X \longrightarrow \mathbb{R} \cup \{+\infty\}$ una función tal que $\nu(x) < \infty$ para algun $x \in X$, la función ν es semicontinua inferior (l.s.c) en $x \in X$ si*

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

para cualquier sucesión $\{x_n\}$ en X convergente a $x \in X$. Si ν es l.s.c para toda $x \in X$, se llama inferiormente continua (l.s.c.). La función ν se dice ser superiormente semicontinua (u.s.c.) si $-\nu$ es l.s.c..

Proposición E.0.5 *Sea $I \subset \mathbb{R}$ un intervalo y $g : I \to \mathbb{R}$ una función convexa. Entonces g es monótona en I, o existe $p \in I$ tal que g es decreciente en $\{x \in I | x \leq p\}$ y g es creciente en $\{x \in I | p \leq x\}$*

Proposición E.0.6 *Sea X una variable aleatoria $P-$ integrable sobre $(\Omega, \mathcal{F}, P)$ y sean $\mathcal{A}$ y $\mathcal{B}$ σ-álgebras contenidas en $\mathcal{F}$. Si $\mathcal{A} \subset \mathcal{B}$ entonces,*

$$E[X|\mathcal{A}] = E\left[E[X|\mathcal{A}]|\mathcal{B}\right] = E\left[E[X|\mathcal{B}]|\mathcal{A}\right].$$

Proposición E.0.7 *Sea ν una variable aleatoria $P-$integrable en $(\Omega, \mathcal{F}, P)$ y $\mathcal{G}$ una sub σ-álgebra de $\mathcal{F}$ entonces $E[E[\nu|\mathcal{G}]] = E[\nu]$.*

Proposición E.0.8 *Sean ν_1 y ν_2 variables aleatorias $P-$ integrables en $(\Omega, \mathcal{F}, P)$ y $\mathcal{G}$, $\mathcal{G}_\infty$ sub σ-álgebras de $\mathcal{F}$. Si ν es $\mathcal{G}$-medible entonces*

$$E[\nu\nu_1|\mathcal{G}] = \nu E[\nu_1|\mathcal{G}]$$

particularmente

$$E[\nu|\mathcal{G}] = \nu.$$

Definición E.0.8 *La medida de Dirac es una medida δ_x sobre un conjunto X, definida para $x \in X$ y cualquier subconjunto medible $A \subset X$, por*

$$\delta_x(A) = 1_A(x) = \begin{cases} 1, & x \in A, \\ 0, u & x \notin A. \end{cases} \tag{E.1}$$

donde 1_A *es la función indicadora de* A.

Proposición E.0.9 *Sea* f *una función medible sobre* X, *entonces*

$$\int_X f(y)\delta_x(dy) = f(x).$$

Definición E.0.9 *Para cada* $B \in \mathcal{B}(X)$, *se definen los tiempos de alcance,*

$$\tau_B := \text{ínf}\{t \geq 1 | x_t \in B\}.$$

Dado un estado inicial $x_0 = x \in X$, se define

$$L(x, B) := P_x[\tau_B < \infty] = P_x[x_t \in B \;\; p.a. \;\; t \geq 1].$$

Definición E.0.10 *La cadena de Markov* $\{x_t\}$ *(o el correspondiente kérnel estocástico* Q*), es llamado Harris recurrente, si existe una medida* $\sigma - finita$ m *en* $\mathcal{B}(X)$ *tal que* $L(x, B) = 1$ *para cada* $x \in X$ *cuando* $m(B) > 0$.

Proposición E.0.10 *Si* Q *es Harris recurrente, entonces existe una medida invariante no trivial* m, *la cual es única salvo multíplos por escalares.*

Observación E.0.1 *En la Proposición E.0.10, si la cadena de Markov es Harris recurrente y la medida invariante es finita, entonces se llamará recurrente positiva.*

Bibliografía

[1] Ash, R. B. y Doléans-Dade, C.A. *Probability and Measure Theory*, Academic Press Elsevier, San Diego, 2005.

[2] Bellman R. *Dynamic Programming*, Dover, 2003.

[3] Bhat, U. N. *An introduction to queueing theory, modeling and analysis in applications*, Birkhauser, 2015.

[4] Bertsekas D.P. *Dynamic Programming: Deterministic and Stochastic Models*, Prentice-Hall, Inc., New Jersey, 1987.

[5] Choy Murphy y Cheong L.F. *Identification of Demand through Statistical Distribution Modeling for Improved Demand*, School of Information Systems, Singapore Management University, 2011.

[6] Hernández-Lerma, O y Laserre, J.B. *Further Topics on Discrete-Time Markov Control Processes*, Springer, New York, 1999.

[7] Hernández-Lerma, O. y Laserre, J.B. *Discrete-Time Markov Control Processes, Basic Optimality Criteria.* Springer, New York, 1989.

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

[9] Lindley D.V. *The theory of queues with single server*, Proc. Cambridge Philos. Soc. 48, p.p. 277-289, 1952.

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

[11] Rieder, U. *Selection theorems for optimization problems*, Manuscripta Math 24, p.p. 115-131, 1978.

[12] Schal, M. *For optimality and for the limit of n-stage optimal policies to be optimal*, Z. Whars. Verw. Geb. 32, p.p. 179-196, 1975.

[13] Sevilla B. *Caracterización de Estrategias Umbral, en Procesos de Decisión de Markov*, Facultad de Ciencias Físico Matemáticas, Benemérita Universidad Autónoma de Puebla, México, 2014.

[14] STROMBERG K.R. *Introduction to Classical Real Analysis*, Wadsworth International Group, Belmont, California, 1981.

[15] TANG Q. y GURAMI T. *Precise estimates fot the ruin probability in finite horizon in a discrete-time model with heavy-tailes insurance and financial risks*, volumen 108, Issue 2, p.p. 299-325, 2003.

[16] XIAPING GUO y QUANXIN ZHU. *Average Optimality for Markov Decision Processes in Borel Spaces: A New Condition and Approach*, Journal of Applied Probability, p.p. 101-111, 2006.

[17] ZACARÍAS ESPINOZA, GABRIEL. *Procesos de Decisión de Markov Descontados*, Facultad de Ciencias Físico Matemáticas, Benemérita Universidad Autónoma de Puebla, México, 2007.