

Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias Físico Matemáticas
Postgrado en Ciencias Matemáticas

*Procesos de Control Semi-Markovianos en Líneas de
Espera*

Tesis

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

Presenta
Mónica Aguirre Mastranzo

Director de Tesis
Dr. Hugo Adán Cruz Suárez

Puebla, Pue.

Octubre 2012

Introducción

En la presente tesis, se consideran sistemas que se desarrollan en el tiempo, tales que su comportamiento tiene aspectos aleatorios. El desarrollo de estos sistemas se puede observar en tiempo discreto o continuo. En el caso discreto, el eje del tiempo es particionado en segmentos de tamaño fijo, dichos segmentos se llaman etapas o periodos. En un sistema de tiempo continuo los eventos pueden ocurrir en cualquier instante de tiempo y son registrados cuando estos ocurren.

El marco teórico en el que se trabajará se conoce como Procesos de Control Semi-Markovianos (PCSMs) (véase [3], [9] y [14]). Estos procesos son una clase importante de los procesos de control en tiempo continuo. Su 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 $x_n = x$, entonces el controlador elige una acción $a_n = a$ y sucede lo siguiente: el sistema permanece en el estado x durante un tiempo aleatorio no-negativo δ_{n+1} con distribución H conocida, lo que genera una recompensa inmediata que depende del estado, la acción y el tiempo de permanencia, el sistema se mueve a un nuevo estado $x_{n+1} = y$ de acuerdo a una ley de transición, una vez ocurrido lo anterior, el proceso se repite. Con el fin de medir el comportamiento de las recompensas por etapa se define el criterio de rendimiento, para ésto se usará como índice de funcionamiento la recompensa total esperada α -descontada con horizonte infinito. A la sucesión de acciones que se elige en cada época de decisión, se le llama política. De esta manera el Problema de Control Óptimo (PCO) es hallar una política de control que optimice el criterio de rendimiento.

Los Procesos de Control Semi-Markovianos fueron introducidos en Howard, Jewel y DeCani (véase [8]). Estos modelos han sido estudiados y aplicados, especialmente en líneas de espera con control, Ross, Cinlar y Lippman contribuyen a dicha teoría (véase [17], [5] y [13]).

Hoy en día los PCSMs son útiles para el estudio de una amplia gama de

problemas de optimización, en una variedad de áreas, incluyendo la robótica, de control automatizado, economía y en la industria (véase [6], [12] y [14]).

Este trabajo de tesis se encuentra relacionado con la teoría propuesta por Puterman [15], el cual ha servido como motivación para estudiar dichos procesos. El problema que plantea [15], es el estudio de modelos markovianos a tiempo continuo, los cuales involucran el tiempo de permanencia en dichos estados. Además de plantear un ejemplo en líneas de espera controladas. La manera que en [15] se analiza estos modelos es mediante dos formas:

1. Transformando el modelo original a uno equivalente, mediante la técnica de uniformización.
2. Usando directamente los resultados de los Procesos de Control Semi-Markovianos.

La técnica de uniformización consiste en convertir un modelo en tiempo continuo con tasas de transición exponenciales a un modelo con estados independientes y tasas de transición constantes.

El trabajo de tesis se centra en el estudio de Procesos de Control Semi-Markovianos. La referencia básica, para el estudio de los PCSMs se encuentra en [14]. En [14] se trabaja en ambiente de costos no-negativos. En la tesis se presenta el modelo usando recompensas y permitiendo cambios de signo. Bajo condiciones adicionales clásicas en la teoría de los PCSMs se muestra una versión de la Ecuación de Programación Dinámica. La extensión propuesta en la teoría permite el estudio de un problema de línea de espera controlada M/M/n, el cual se encuentra motivado en [15]. En [15] se plantea dicho problema considerando un solo servidor y un solo tipo de cliente, el cual es resuelto únicamente por la técnica de uniformización. Ahora, en este trabajo se propone una extensión a n servidores con dos tipos de clientes, este caso es resuelto por la técnica de uniformización y usando directamente los resultados del Capítulo 1. Cabe mencionar que en cada caso se proveen algoritmos numéricos elaborados en MATLAB.

La tesis se divide en tres capítulos los cuales son distribuidos de la siguiente manera.

En el Capítulo 1, se estudia el Modelo de Control Semi-Markoviano y todas sus componentes bajo el criterio de recompensa descontada. Se presentan una serie de condiciones que garantizan la existencia de políticas óptimas. En Capítulo 2, se presenta el enfoque de uniformización el cual será útil en Capítulo 3.

En el Capítulo 3, se presenta un problema de control en líneas de espera M/M/1 con un solo tipo de cliente, el cual se plantea como un Modelo de Control Semi-Markoviano MCSM, así mismo, se presenta una extensión de dicho problema a un modelo de líneas de espera controlada M/M/n con dos clases de clientes, que también se plantea como un MCSM, el cual se analizará aplicando resultados directos de los PCSMs y el enfoque de uniformización. Además se dan ejemplos numéricos en los que los resultados se pueden visualizar en tablas, realizadas en MATLAB.

ÍNDICE GENERAL

Introducción	III
1. Procesos de Control Semi-Markovianos	1
1.1. Modelo de Control Semi-Markoviano	2
1.2. Políticas de Control	3
1.3. Problema de control óptimo con recompensa descontada . . .	5
1.4. Condiciones sobre el Modelo de Control	6
1.5. Criterio de optimalidad con recompensa descontada	11
1.5.1. Ecuación de Optimalidad	14
2. Uniformización	23
2.1. Uniformización	23
2.1.1. Ejemplo	25
2.2. Modelo Descontado	26
3. Ejemplos	29
3.1. Una línea M/M/1 de espera controlada, con un solo tipo de servicio.	29
3.1.1. Descripción del modelo	30
3.2. Una línea M/M/n de espera controlada, con dos tipos de servicio. 35	
3.2.1. Descripción del modelo	36
3.2.2. Modelo analizado directamente	40
3.2.3. Modelo mediante uniformización	41
3.2.4. Ejemplos numéricos	43
Conclusiones	45

A. Otros resultados	47
A.1. Funciones	47
A.2. Teorema de Ionescu-Tulcea y Lema de Fatou	48
A.3. Operador Contracción	49
B. Códigos	51
B.1. Línea de espera con control M/M/1 con uniformización	51
B.2. Línea de espera con control M/M/n sin uniformización	53
B.3. Línea de espera con control M/M/n con uniformización	58
Bibliografía	61

Capítulo 1

Procesos de Control Semi-Markovianos

El objetivo principal de este capítulo es presentar formalmente el Modelo de Control Semi-Markoviano (MCSM) e introducir el Problema de Control Óptimo Semi-Markoviano (PCO). Para la formulación del PCO se requieren tres elementos: un modelo de control, un conjunto de políticas admisibles y un índice de funcionamiento mediante el cual se evaluará el desempeño del sistema, cuando se usan diferentes políticas. Lo cual se desarrolla en las Secciones 1.1, 1.2 y 1.3 respectivamente.

Para resolver el PCO se requiere que el modelo de control satisfaga un conjunto de condiciones que se introducen en la Sección 1.4. Las condiciones que se consideran en el trabajo son de tres tipos: las primeras condiciones son de continuidad y compacidad; estas permiten asegurar la existencia de maximizadores, las segundas condiciones son de crecimiento, estas nos permiten analizar el PCO con funciones de recompensa y costo no acotados. La tercera condición es de regularidad, la cual es de tipo probabilístico; dicha condición garantiza que los PCSMs son regulares, es decir, que experimenten un número finito de transiciones en intervalos finitos de tiempo (véase [14] y [15]).

La teoría desarrollada en este capítulo, se encuentra motivada en la referencia [14]. En [14] se presenta un estudio de los PCSMs en el contexto de costos no-negativos. En contraste, ahora la teoría se presenta con funciones de recompensa permitiendo cambio de signo.

1.1. Modelo de Control Semi-Markoviano

Definición 1.1.1 *Un Modelo de Control Semi-Markoviano (MCSM)*

$$(\mathbb{X}, \mathbb{A}, \{\mathbb{A}(x) : x \in \mathbb{X}\}, Q, H, D, d) \quad (1.1)$$

consiste en lo siguiente:

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

Observación 1.1.2 *Se supone que el conjunto $\mathbb{K} := \{(x, a) : x \in \mathbb{X}, a \in \mathbb{A}(x)\}$ de pares estado acción admisible, es un subconjunto de Borel del espacio $\mathbb{X} \times \mathbb{A}$. También se supone que \mathbb{K} contiene a la gráfica de una función medible $f : \mathbb{X} \rightarrow \mathbb{A}$, de manera que $f(x) \in \mathbb{A}(x)$, para todo $x \in \mathbb{X}$.*

- La ley de transición $Q(\cdot | \cdot)$ es un kernel estocástico sobre \mathbb{X} dado \mathbb{K} .
- $H(\cdot | x, a)$ es la función de distribución del tiempo de permanencia sobre \mathbb{R} , para cada $(x, a) \in \mathbb{K}$.
- D y d son las funciones de recompensa y costo las cuales son medibles sobre \mathbb{K} .

Un MCSM 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 $x_n = x$ y el controlador elige una acción $a_n = a \in \mathbb{A}(x)$, generándose con ello lo que se describe a continuación:

- Se genera una ganancia inmediata $D(x, a)$.
- El sistema permanece en dicho estado $x_n = x$ durante un tiempo aleatorio no-negativo δ_{n+1} con distribución $H(\cdot | x, a)$.
- En el tiempo $T_{n+1} := T_n + \delta_{n+1}$ ($n \in \mathbb{N}$, $T_0 := 0$), el sistema transita a un nuevo estado $x_{n+1} = y$ de acuerdo a la distribución $Q(\cdot | x, a)$.
- Se produce un costo debido al tiempo de permanencia en el estado x , cuya razón de costo es $d(x, a)$.
- Finalmente, una vez en el estado y el proceso se repite.

1.2. Políticas de Control

Sea un MCSM fijo, se define el espacio de historias admisibles hasta la n -ésima época de decisión mediante

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

y

$$\mathbb{H}_n := (\mathbb{K} \times \mathbb{R}_+)^n \times \mathbb{X} \text{ para } n \in \mathbb{N}.$$

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

$$h_n := (x_0, a_0, \delta_1, x_1, a_1, \delta_2, \dots, x_{n-1}, a_{n-1}, \delta_n, x_n),$$

donde $(x_k, a_k, \delta_{k+1}) \in \mathbb{K} \times \mathbb{R}_+$ para $k = 0, 1, 2, \dots, n-1$ y $x_n \in \mathbb{X}$.

Definición 1.2.1 *Una política de control admisible o simplemente una política es una sucesión $\pi := \{\pi_n\}$, donde cada π_n es un kernel estocástico sobre \mathbb{A} dado \mathbb{H}_n , tales que satisfacen la restricción $\pi(\mathbb{A}(x) \mid h_n) = 1$ para todo $h_n \in \mathbb{H}_n$ y $n \in \mathbb{N}$.*

Se denota por Π al conjunto de todas las políticas.

Sea \mathbb{F} el conjunto de todas las funciones $f : \mathbb{X} \rightarrow \mathbb{A}$, tal que $f(x) \in \mathbb{A}(x)$ para todo $x \in \mathbb{X}$.

Se denota a la familia de kernels estocásticos sobre \mathbb{A} dado \mathbb{X} , como $P(\mathbb{A} \mid \mathbb{X})$.

Sea Φ el conjunto de todos los kernels estocásticos φ en $P(\mathbb{A} \mid X)$ tales que para toda $x \in \mathbb{X}$ se tiene que $\varphi(\mathbb{A}(x) \mid x) = 1$.

Observación 1.2.2 *Se dice que $\pi(\cdot \mid h)$ está concentrada en $g(h)$, si, $\pi(C \mid h) = I_C(g(h))$ para cada $C \in \mathcal{B}(\mathbb{A})$. Donde I_C es la función indicadora del conjunto C .*

Definición 1.2.3 *Una política $\pi \in \Pi$ es:*

- **Markoviana Aleatorizada** (Π_{RM}). *Si existe una sucesión $\{\varphi_n\}$ de kernels estocásticos con $\varphi_n \in \Phi$ (definidas sobre \mathbb{A} dado X), tales que, $\pi_n(\cdot \mid h_n) = \varphi_n(\cdot \mid x_n)$ para toda $h_n \in \mathbb{H}_n$ y $n \in \mathbb{N}$.*

- **Markoviana Aleatorizada Estacionaria** (Π_{RS}). Si existe $\varphi \in \Phi$ kernel estocástico, tal que: $\pi_n(\cdot | h_n) = \varphi(\cdot | x_n)$ para toda $h_n \in \mathbb{H}_n$ y $n \in \mathbb{N}$.
- **Determinista** (Π_D). Si existe una sucesión $\{g_n\}$ de funciones medibles con $g_n : \mathbb{H}_n \rightarrow \mathbb{A}$, tales que, para cada $h_n \in \mathbb{H}_n$ y $n \in \mathbb{N}$, se tiene que $g_n(h_n) \in \mathbb{A}(x_n)$ y $\pi_n(\cdot | h_n)$ está concentrada en $g_n(h_n)$.
- **Determinista Markoviana** (Π_{DM}). Si existe una sucesión $\{f_n\}$ de funciones medibles $f_n : \mathbb{X} \rightarrow \mathbb{A}$ (o $f_n \in \mathbb{F}$), tales que $f_n(x_n) \in \mathbb{A}(x_n)$ y $\pi_n(\cdot | h_n)$ está concentrada en $f_n(x_n)$ para cada $h_n \in \mathbb{H}_n$ y $n \in \mathbb{N}$.
- **Determinista Markoviana Estacionaria** (Π_{DS}). Si existe una función medible $f : \mathbb{X} \rightarrow \mathbb{A}$ (o $f \in \mathbb{F}$), tal que $f(x_n) \in \mathbb{A}(x_n)$ y $\pi_n(\cdot | h_n)$ está concentrada en $f(x_n)$ para cada $h_n \in \mathbb{H}_n$ y $n \in \mathbb{N}$.

Observación 1.2.4 Note que $\Pi_{RS} \subset \Pi_{RM} \subset \Pi$ y $\Pi_{DS} \subset \Pi_{DM} \subset \Pi_D \subset \Pi$.

Debido a la definición anterior, en adelante se identificará al conjunto de todas las políticas estacionarias con \mathbb{F} .

Sea (Ω, \mathcal{F}) el espacio medible canónico en el cual Ω es el espacio producto $(\mathbb{K} \times \mathbb{R}_+)^{\infty}$ y \mathcal{F} la σ -álgebra producto correspondiente. Se observa que un elemento $\omega \in \Omega$ tiene la forma $\omega = (x_0, a_0, \delta_1, x_1, a_1, \delta_2, \dots)$. A las variables $x_k \in \mathbb{X}$, $a_k \in \mathbb{A}(x_k)$ y $\delta_{k+1} \in \mathbb{R}_+$, se les llamará variables de estado, control y tiempo de transición, respectivamente. De acuerdo al Teorema de Ionescu-Tulcea (véase Apéndice A), existe una única medida de probabilidad P_x^π sobre (Ω, \mathcal{F}) . Además, para cada $B \in \mathcal{B}(\mathbb{A})$, $C \in \mathcal{B}(\mathbb{X})$, $h_t \in \mathbb{H}_t$ y $n \in \mathbb{N}$ se tiene que:

$$P_x^\pi(x_0 = x) = 1, \quad (1.2)$$

$$P_x^\pi(a_t \in B | h_n) = \pi_t(B | h_n), \quad (1.3)$$

$$P_x^\pi(x_n \in C | h_n, a_n, \delta_{n+1}) = Q(C | x_n, a_n), \quad (1.4)$$

$$P_x^\pi(\delta_{n+1} \leq t | h_n, a_n) = H(t | x_n, a_n). \quad (1.5)$$

Para una política arbitraria $\pi \in \Pi$, la variable x_n describe el estado del sistema en el tiempo de la n -ésima transición (o época de decisión) T_n y a_n representa el control elegido de acuerdo a la política π .

Se denotará por E_x^π al operador esperanza correspondiente a P_x^π .

1.3. Problema de control óptimo con recompensa descontada

En el transcurso del trabajo para un MCSM se consideran las recompensas descontadas en forma continua, es decir, para un factor de descuento $\alpha \in (0, 1)$, una ganancia k se genera al tiempo t , esto equivale a una ganancia $ke^{-\alpha t}$ en el tiempo $t = 0$, de acuerdo a la interpretación del MCSM, la función de recompensa por etapa, cuando el proceso se encuentra en el estado x y se elige una acción a , toma la forma siguiente:

$$r(x, a) := D(x, a) - d(x, a) \int_0^\infty \int_0^t e^{-\alpha s} ds H(dt | x, a). \quad (1.6)$$

Cada MCSM estará dotado de una función real, llamada función objetivo o criterio de rendimiento, cuyo fin es medir el comportamiento de las recompensas por etapa, para ello se usará como índice de funcionamiento la recompensa total esperada α -descontada.

Se considera un MCSM fijo y un conjunto de políticas Π .

Definición 1.3.1 *Se define la recompensa total esperada α -descontada, para cada $x \in \mathbb{X}$, $\pi \in \Pi$ y $\alpha \in (0, 1)$ como:*

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

donde $r(\cdot, \cdot)$ es la función introducida en (1.6).

Definición 1.3.2 *Una política $\pi^* \in \Pi$ es óptima si*

$$V(x, \pi^*) := \sup_{\pi \in \Pi} V(x, \pi), \quad \forall x \in \mathbb{X}.$$

Definición 1.3.3 *Se define para cada $x \in \mathbb{X}$*

$$V_\alpha^*(x) := \sup_{\pi \in \Pi} V(x, \pi),$$

V_α^* se le llama función de valores óptimos o valor óptimo.

Dado el MCSM (1.1), el problema de control óptimo consiste en encontrar una política óptima, es decir, maximizar la función $\pi \rightarrow V(x, \pi)$ sobre Π , para toda $x \in \mathbb{X}$.

1.4. Condiciones sobre el Modelo de Control

Para resolver el PCO, se requiere que el modelo de control satisfaga un conjunto de condiciones que se exponen en esta sección, las condiciones que se consideran en esta tesis son de tres tipos: condiciones de continuidad y compacidad, condiciones sobre el crecimiento de las funciones de recompensa y una tercera condición de regularidad.

Las primeras condiciones son de continuidad y compacidad, las cuales permiten asegurar la existencia de maximizadores.

Condición 1.4.1 *Condiciones de continuidad y compacidad.*

- a) $\mathbb{A}(x)$ es un conjunto compacto, para cada $x \in \mathbb{X}$.
- b) Las funciones de recompensa y costo $D(x, a)$ y $d(x, a)$ respectivamente, son reales medibles y semi-continuas superiormente (u.s.c.) en $a \in \mathbb{A}(x)$, para cada $x \in \mathbb{X}$,
- c) La ley de transición $Q(\cdot | x, a)$ es fuertemente continua en $a \in \mathbb{A}(x)$, es decir, para cada función medible y acotada v sobre \mathbb{X} , se tiene que,

$$a \mapsto \int_{\mathbb{X}} v(y)Q(dy | x, a),$$

es una función continua y acotada sobre $\mathbb{A}(x)$.

- d) La función $H(t | x, a)$ es continua en $a \in \mathbb{A}(x)$ para cada $x \in \mathbb{X}$ y $t \in \mathbb{R}$.

En seguida se presentan condiciones de crecimiento, las cuales permiten analizar el PCO con recompensa y costos no acotados. Sea $w : \mathbb{X} \rightarrow [1, \infty)$ una función medible. Se denota por $\mathbb{B}_w(\mathbb{X})$ el espacio lineal normado el cual consiste de todas las funciones medibles $u : \mathbb{X} \rightarrow \mathbb{R}$, tales que satisfacen la condición,

$$\| u \|_w := \sup_{x \in \mathbb{X}} \frac{|u(x)|}{w(x)} < \infty. \quad (1.7)$$

Condición 1.4.2 *Condiciones de crecimiento.*

a) Existe una función medible $w : \mathbb{X} \rightarrow [1, \infty)$, así como constantes $c_1, c_2 > 0$ y $\beta \in (0, 1)$, tales que

$$\sup_{a \in \mathbb{A}(x)} |D(x, a)| \leq c_1 w(x), \quad \sup_{a \in \mathbb{A}(x)} |d(x, a)| \leq c_2 w(x), \quad (1.8)$$

y

$$\int_{\mathbb{X}} w(y) Q(dy | x, a) \leq \beta w(x), \quad x \in \mathbb{X}, a \in \mathbb{A}(x). \quad (1.9)$$

b) La función

$$a \mapsto \int_{\mathbb{X}} w(y) Q(dy | x, a),$$

es continua en $\mathbb{A}(x)$, para cada $x \in \mathbb{X}$.

Las demostraciones del Lema 1.4.3 y Proposición 1.4.6, están basadas en la proposición que se encuentra en [14].

Lema 1.4.3 *Mediante la Condición (1.4.2), se satisface la siguiente desigualdad, para cada $\pi \in \Pi$ y $x \in \mathbb{X}$.*

$$\sup_{n \in \mathbb{N}} E_x^\pi [w(x_n)] < \infty.$$

Demostración. Para mostrar que $\sup_{n \in \mathbb{N}} E_x^\pi [w(x_n)] < \infty$ se tiene que para cada $\pi \in \Pi$ y $x \in \mathbb{X}$, se observa que de (1.4) y (1.9) se tiene

$$E_x^\pi [w(x_n) | h_{n-1}, a_{n-1}, \delta_n] = \int_{\mathbb{X}} w(y) Q(dy | x_{n-1}, a_{n-1}) \leq \beta w(x_{n-1}).$$

Al tomar esperanzas en ambos lados de la expresión anterior se obtiene que

$$E_x^\pi [w(x_n)] \leq \beta E_x^\pi [w(x_{n-1})], \quad (1.10)$$

iterando la desigualdad (1.10) para un número n se tiene lo siguiente

$$E_x^\pi [w(x_n)] \leq \beta^n w(x_0), \quad (1.11)$$

es decir, para $n = 1$, se tiene que

$$\begin{aligned} E_x^\pi [w(x_1)] &\leq \beta E_x^\pi [w(x_0)] \\ &\leq \beta w(x_0), \end{aligned}$$

para $n = 2$,

$$\begin{aligned} E_x^\pi [w(x_2)] &\leq \beta E_x^\pi [w(x_1)] \\ &\leq \beta [\beta [w(x_0)]] \\ &\leq \beta^2 w(x_0), \end{aligned}$$

así que para cualquier n se obtiene (1.11).

Como $\beta \in (0, 1)$ y $w(\cdot)$ es finita resulta que

$$E_x^\pi [w(x_n)] \leq w(x) < \infty, \quad n \geq 0.$$

Por lo que,

$$\sup_{n \in \mathbb{N}} E_x^\pi [w(x_n)] < \infty. \quad \blacksquare$$

La siguiente condición es de tipo probabilístico, la cual garantiza que los PCSMs son regulares, es decir, que experimentan un número finito de transiciones en intervalos de tiempo acotados, este hecho se probará en la Proposición (1.4.6). Cabe mencionar que está misma condición se da para el caso no controlado (véase [16])

Condición 1.4.4 *Condición de regularidad.*

Existen reales $\epsilon > 0$ y $\theta > 0$ tales que

$$H(\theta \mid x, a) \leq 1 - \epsilon, \quad (x, a) \in \mathbb{K}.$$

Notación 1.4.5 *Para cada $(x, a) \in \mathbb{K}$, se define*

$$\Delta_\alpha(x, a) := \int_0^\infty e^{-\alpha t} H(dt \mid x, a), \quad (1.12)$$

y

$$\tau_\alpha(x, a) := \frac{1 - \Delta_\alpha(x, a)}{\alpha}.$$

De la relación (1.1) se tiene que:

$$\begin{aligned}
r(x, a) &= D(x, a) - d(x, a) \int_0^\infty \left[\int_0^t e^{-\alpha s} ds \right] H(dt | x, a) \\
&= D(x, a) - d(x, a) \int_0^\infty \left\{ -\frac{1}{\alpha} [e^{-\alpha t} - 1] \right\} H(dt | x, a) \\
&= D(x, a) - d(x, a) \frac{1}{\alpha} \left\{ \int_0^\infty H(dt | x, a) - \int_0^\infty e^{-\alpha t} H(dt | x, a) \right\} \\
&= D(x, a) - d(x, a) \frac{1}{\alpha} \left\{ 1 - \int_0^\infty e^{-\alpha t} H(dt | x, a) \right\}.
\end{aligned}$$

Es decir,

$$r(x, a) = D(x, a) - \tau_\alpha(x, a)d(x, a), \quad (x, a) \in \mathbb{K}. \quad (1.13)$$

Proposición 1.4.6 *Si se cumple la Condición 1.4.2, entonces:*

- a) $\rho_\alpha := \sup_{\mathbb{K}} \Delta_\alpha(x, a) < 1$,
- b) $P_x^\pi [\sum_{n=1}^\infty \delta_n = \infty] = 1, \forall x \in \mathbb{X}, \pi \in \Pi$.

Demostración. a) Mediante la fórmula de integración por partes para integrales de Riemann-Stieltjes

$$\int_\varsigma^v f(t)dg(t) + \int_\varsigma^v g(t)df(t) = f(v)g(v) - f(\varsigma)g(\varsigma),$$

en particular, para $\varsigma = 0, v = \infty, f(t) = e^{-\alpha t}, g(t) = H(t | x, a)$, se tiene

$$\int_0^\infty e^{-\alpha t} H(dt | x, a) + \int_0^\infty H(t | x, a)d(e^{-\alpha t}) = 0. \quad (1.14)$$

La igualdad (1.14) se deduce de las siguientes relaciones:

$$e^{-\alpha t} \rightarrow 0, \quad t \rightarrow \infty \quad \text{y} \quad H(t | x, a) \rightarrow 1, \quad t \rightarrow \infty,$$

para $t = 0$, se tiene que $e^{-\alpha t} = 1$ y $H(t | x, a) = 0$.

Por lo tanto,

$$\begin{aligned}
\Delta_\alpha(x, a) &= \int_0^\infty e^{-\alpha t} H(dt | x, a) = \alpha \int_0^\infty e^{-\alpha t} H(t | x, a) dt \\
&= \alpha \left[\int_0^\theta e^{-\alpha t} H(t | x, a) dt + \int_\theta^\infty e^{-\alpha t} H(t | x, a) dt \right].
\end{aligned}$$

Por la Condición 1.4.4,

$$\begin{aligned} \alpha \int_0^\theta e^{-\alpha t} H(t \mid x, a) dt &\leq \alpha(1 - \epsilon) \int_0^\theta e^{-\alpha t} dt, \quad (x, a) \in \mathbb{K}, \\ &\leq (1 - \epsilon)(1 - e^{-\alpha\theta}); \end{aligned}$$

y además, dado que $H(t \mid x, a) \leq 1$, entonces

$$\begin{aligned} \alpha \int_\theta^\infty e^{-\alpha t} H(t \mid x, a) dt &\leq \alpha \int_\theta^\infty e^{-\alpha t} dt, \quad (x, a) \in \mathbb{K} \\ &\leq e^{-\alpha\theta}. \end{aligned}$$

Esto es,

$$\begin{aligned} \Delta_\alpha(x, a) &\leq (1 - \epsilon)(1 - e^{-\alpha\theta}) + e^{-\alpha\theta} \\ &= 1 - \epsilon(1 - e^{-\alpha\theta}) < 1, \quad (x, a) \in \mathbb{K}. \end{aligned} \tag{1.15}$$

Por consiguiente,

$$\rho_\alpha = \sup_{\mathbb{K}} \Delta_\alpha(x, a) < 1.$$

b) Se supone que para cada $x \in \mathbb{X}$, $\pi \in \Pi$:

$$E_x^\pi [\prod_{n=0}^\infty \Delta_\alpha(x_n, a_n)] = E_x^\pi \left[\exp(-\alpha \sum_{n=1}^\infty \delta_n) \mid x_0, a_0, x_1, a_1, \dots \right], \tag{1.16}$$

y

$$E_x^\pi [\prod_{n=0}^\infty \Delta_\alpha(x_n, a_n)] = 0, \tag{1.17}$$

entonces

$$E_x^\pi \left[\exp(-\alpha \sum_{n=1}^\infty \delta_n) \mid x_0, a_0, x_1, a_1, \dots \right] = 0,$$

luego, por la positividad de la función exponencial se tiene

$$P_x^\pi \left[\sum_{n=1}^\infty \delta_n = \infty \right] = 1, \quad \forall x \in \mathbb{X}, \pi \in \Pi,$$

lo anterior prueba la parte (b), se supone que las igualdades (1.16) y (1.17) son válidas. En seguida se probará la validez de dichas identidades. Para mostrar (1.16), se observa que:

$$\begin{aligned} E_x^\pi [\prod_{n=0}^\infty \Delta_\alpha(x_n, a_n)] &= E_x^\pi \left[\prod_{n=0}^\infty \int_0^\infty e^{-\alpha t} H(dt \mid x_n, a_n) \right] \\ &= E_x^\pi [\prod_{n=0}^\infty E_x^\pi [e^{-\alpha\delta_{n+1}} \mid x_n, a_n]], \end{aligned}$$

se tiene que

$$\begin{aligned} & E_x^\pi \left[\prod_{n=0}^{\infty} E_x^\pi \left[e^{-\alpha\delta_{n+1}} \mid x_n, a_n \right] \right] \\ &= E_x^\pi \left[E_x^\pi \left[\exp\left(-\alpha \sum_{n=1}^{\infty} \delta_n\right) \mid x_0, a_0, x_1, a_1, \dots \right] \right], \end{aligned}$$

por lo que, debido a las propiedades de esperanza condicional, lo anterior implica

$$E_x^\pi \left[\prod_{n=0}^{\infty} \Delta_\alpha(x_n, a_n) \right] = E_x^\pi \left[\exp\left(-\alpha \sum_{n=1}^{\infty} \delta_n\right) \mid x_0, a_0, x_1, a_1, \dots \right],$$

lo cual muestra (1.16).

Para mostrar (1.17), note que de la parte (a) se tiene que $\Delta_\alpha(x, a) < 1$ para todo $(x, a) \in \mathbb{K}$, lo cual implica

$$\prod_{k=0}^n \Delta_\alpha(x_k, a_k) \leq \rho_\alpha^{n+1} \rightarrow 0, \quad n \rightarrow \infty,$$

es decir, $\prod_{n=0}^{\infty} \Delta_\alpha(x_n, a_n)$ converge a cero, de donde

$$E_x^\pi \left[\prod_{n=0}^{\infty} \Delta_\alpha(x_n, a_n) \right] = 0.$$

■

1.5. Criterio de optimalidad con recompensa descontada

En esta sección se muestra la existencia de políticas óptimas estacionarias bajo las condiciones presentadas en la Sección 1.4.

Sean $x \in \mathbb{X}$ y $\pi \in \Pi$,

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

el cual define la recompensa total esperada α -descontada al usar la política π dado el estado inicial $x_0 = x$. Además, la recompensa α -descontada óptima (función de valor óptimo) cuando el estado inicial es $x_0 = x$, se define como

$$V_\alpha^*(x) := \sup_{\pi \in \Pi} V(x, \pi), \quad (1.18)$$

y se dice que $\pi^* \in \Pi$ es una política óptima, si

$$V(x, \pi^*) := \sup_{\pi \in \Pi} V(x, \pi), \quad \forall x \in \mathbb{X}. \quad (1.19)$$

La siguiente proposición muestra una manera diferente de expresar la recompensa total esperada α -descontada.

Proposición 1.5.1 *Para cada $x \in \mathbb{X}$ y $\pi \in \Pi$ se tiene que*

$$V(x, \pi) := E_x^\pi \left[r(x_0, a_0) + \sum_{n=1}^{\infty} \prod_{j=0}^{n-1} \Delta_\alpha(x_j, a_j) r(x_n, a_n) \right], \quad (1.20)$$

donde $\Delta_\alpha(\cdot, \cdot)$ se define en (1.12).

Demostración. Por definición $\Delta_\alpha(x, a)$ se observa que

$$\begin{aligned} & E_x^\pi \left[r(x_0, a_0) + \sum_{n=1}^{\infty} \prod_{j=0}^{n-1} \Delta_\alpha(x_j, a_j) r(x_n, a_n) \right] \\ &= E_x^\pi \left[r(x_0, a_0) + \sum_{n=1}^{\infty} \prod_{j=0}^{n-1} \int_0^\infty e^{-\alpha t} H(dt \mid x_j, a_j) r(x_n, a_n) \right] \\ &= E_x^\pi \left[r(x_0, a_0) + \sum_{n=1}^{\infty} \prod_{j=0}^{n-1} E_x^\pi [e^{-\alpha \delta_{j+1}} \mid x_j, a_j] r(x_n, a_n) \right], \end{aligned}$$

luego por (1.5) se tiene que las variables aleatorias $\delta_1, \delta_2, \dots$, son condicionalmente independientes dado el proceso $x_0, a_0, x_1, a_1, \dots$, así note que,

$$\begin{aligned} \prod_{j=0}^{n-1} E_x^\pi [e^{-\alpha \delta_{j+1}} \mid x_j, a_j] &= E_x^\pi \left[\prod_{j=0}^{n-1} e^{-\alpha \delta_{j+1}} \mid x_j, a_j \right] \\ &= E_x^\pi [e^{-\alpha(\delta_1 + \dots + \delta_n)} \mid x_0, a_0, x_1, a_1, \dots], \end{aligned}$$

entonces

$$\begin{aligned} & E_x^\pi \left[r(x_0, a_0) + \sum_{n=1}^{\infty} \prod_{j=0}^{n-1} \Delta_\alpha(x_j, a_j) r(x_n, a_n) \right] \\ &= E_x^\pi \left[\sum_{n=0}^{\infty} E_x^\pi [e^{-\alpha T_n} \mid x_0, a_0, x_1, a_1, \dots] r(x_n, a_n) \right], \end{aligned}$$

donde T_n ($n \in \mathbb{N}$) es una variable aleatoria, la cual representa la n -ésima época de decisión y que admite la expresión $T_0 = 0$ y $T_n = \sum_{j=0}^n \delta_j$, para cada $n \in \mathbb{N}$. Por las propiedades de esperanza condicional se obtiene que

$$\begin{aligned} & E_x^\pi \left[r(x_0, a_0) + \sum_{n=1}^{\infty} \prod_{j=0}^{n-1} \Delta_\alpha(x_j, a_j) r(x_n, a_n) \right] \\ &= \sum_{n=0}^{\infty} E_x^\pi [e^{-\alpha T_n} r(x_n, a_n)]. \end{aligned}$$

Por lo que,

$$E_x^\pi \left[r(x_0, a_0) + \sum_{n=1}^{\infty} \prod_{j=0}^{n-1} \Delta_\alpha(x_j, a_j) r(x_n, a_n) \right] = V(x, \pi). \quad \blacksquare$$

Para $x \in \mathbb{X}$, $\pi \in \Pi$ y $n \in \mathbb{N}$ se denota por $V_n(x, \pi)$ la recompensa esperada α -descontada hasta la n -ésima transición, es decir,

$$V_n(x, \pi) = E_x^\pi \left[\sum_{j=0}^{n-1} e^{-\alpha T_j} r(x_j, a_j) \right].$$

Se observa que

$$V_n(x, \pi) \rightarrow V(x, \pi).$$

Por otra parte, $V_n(x, \pi)$ puede ser expresado como

$$\begin{aligned} V_1(x, \pi) &= E_x^\pi [r(x_0, a_0)] \\ V_n(x, \pi) &= E_x^\pi \left[r(x_0, a_0) + \sum_{k=1}^{n-1} \prod_{j=0}^{k-1} \Delta_\alpha(x_j, a_j) r(x_k, a_k) \right], \quad n = 2, 3, \dots, \end{aligned}$$

luego, se introduce la notación

$$\begin{aligned} \Lambda_\alpha^0 &: = 1, \\ \Lambda_\alpha^k &: = \prod_{j=0}^{k-1} \Delta_\alpha(x_j, a_j), \quad k \in \mathbb{N}, \end{aligned} \tag{1.21}$$

se tiene

$$V_n(x, \pi) = E_x^\pi \left[\sum_{k=0}^{n-1} \Lambda_\alpha^k r(x_k, a_k) \right], \quad \text{para } n \in \mathbb{N}, \tag{1.22}$$

y

$$V(x, \pi) = \lim_{n \rightarrow \infty} V_n(x, \pi) := E_x^\pi \left[\sum_{k=0}^{\infty} \Lambda_\alpha^k r(x_k, a_k) \right]. \quad (1.23)$$

1.5.1. Ecuación de Optimalidad

El objetivo de esta sección es mostrar que la función de valor óptimo satisface la ecuación de optimalidad con recompensa α -descontada (EO).

Sea $r(x, a)$ como se definió en (1.13), $\Delta_\alpha(x, a)$ definida en (1.12) y $\alpha \in (0, 1)$. Se dice que una función medible $v : \mathbb{X} \rightarrow \mathbb{R}$ es una solución de la EO, si satisface

$$v(x) = \max_{a \in \mathbb{A}(x)} \left\{ r(x, a) + \Delta_\alpha(x, a) \int_{\mathbb{X}} v(y) Q(dy | x, a) \right\}, \quad \forall x \in \mathbb{X}. \quad (1.24)$$

Esto permitirá mostrar la existencia de políticas óptimas. La idea general consiste en lo siguiente.

Se denota por v_n la función de valor óptimo hasta la n -ésima transición, es decir,

$$\begin{aligned} v_0 & : = 0 \\ v_n & : = \sup_{\pi \in \Pi} V_n(x, \pi), \quad x \in \mathbb{X}, \end{aligned}$$

del algoritmo de Programación Dinámica se observa que las funciones v_n pueden ser obtenidas de forma iterada, esto es, para $x \in \mathbb{X}$,

$$v_0 := 0 \text{ y para } n \geq 1,$$

$$v_n(x) = \max_{a \in \mathbb{A}(x)} \left\{ r(x, a) + \Delta_\alpha(x, a) \int_{\mathbb{X}} v_{n-1}(y) Q(dy | x, a) \right\}, \quad (1.25)$$

por lo tanto, si se muestra que $v_n \rightarrow V_\alpha^*$ se puede esperar que V_α^* satisfaga la EO. Esto se formaliza en el Teorema 1.5.5 .

Lema 1.5.2 *Se supone que $\mathbb{A}(x)$, $x \in \mathbb{X}$, es compacto y sea $v : \mathbb{K} \rightarrow \mathbb{R}$ una función medible.*

Si $v(x, a)$ es u.s.c en $a \in \mathbb{A}(x)$ para cada $x \in \mathbb{X}$, entonces

$$v^*(x) := \sup_{a \in \mathbb{A}(x)} v(x, a), \quad (1.26)$$

es medible, y además existe $f^* \in \mathbb{F}$ tal que

$$v^*(x) = v(x, f^*) = \max_{a \in \mathbb{A}(x)} v(x, a), \quad \forall x \in \mathbb{X}. \quad (1.27)$$

Lema 1.5.3 *Se supone que se cumplen las Condiciones 1.4.1(b) y 1.4.4, entonces la función*

$$u(x, a) := \int v(y)Q(dy \mid x, a),$$

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

Demostración. Las pruebas de los Lemas 1.5.2 y 1.5.3 puede consultarse en [10]. ■

A continuación se introduce el operador T_α de Programación Dinámica (PD).

Sea $v : \mathbb{X} \rightarrow \mathbb{R}$ una función medible y $0 \leq \alpha \leq 1$, se denota por $T_\alpha v$ a la función dada por

$$T_\alpha v(x) := \sup_{\mathbb{A}(x)} \left\{ r(x, a) + \Delta_\alpha(x, a) \int_{\mathbb{X}} v(y)Q(dy \mid x, a) \right\}, \quad x \in \mathbb{X}. \quad (1.28)$$

Si el máximo en (1.28) es alcanzado en alguna acción $a \in \mathbb{A}(x)$, para cada $x \in \mathbb{X}$, entonces se escribirá máximo en lugar de supremo.

Mediante el uso de T_α , que se puede reescribir (1.24) y (1.25) como:

$$V_\alpha^* = T_\alpha V_\alpha^*, \quad (1.29)$$

y

$$v_n = T_\alpha v_{n-1} = T_\alpha v_0, \quad (1.30)$$

para todo $n = 0, 1, \dots$, con $v_0 = 0$, respectivamente.

La siguiente proposición fue revisada en [9].

Proposición 1.5.4 *Se supone que se cumplen las Condiciones 1.4.1, 1.4.2 y 1.4.4, para $\alpha \in (0, 1)$ y sea T_α el operador definido en (1.28), entonces T_α es un operador contracción sobre $\mathbb{B}_w(\mathbb{X})$, con módulo $\gamma := \alpha\beta < 1$ y*

$$\|T_\alpha v - T_\alpha v'\|_w \leq \gamma \|v - v'\|_w, \quad \forall v, v' \in \mathbb{B}_w(\mathbb{X}). \quad (1.31)$$

Demostración. Sea v en $\mathbb{B}_w(\mathbb{X})$, entonces por el Lema 1.5.3 y la Condición 1.4.1(b), la función

$$v(x, a) = r(x, a) + \Delta_\alpha(x, a) \int v(y)Q(dy | x, a),$$

es u.s.c en $a \in \mathbb{A}(x)$, para cada $x \in \mathbb{X}$. Por el Lema 1.5.2 se tiene que $T_\alpha v$ es una función medible y existe una $f \in \mathbb{F}$ que satisface (1.27). Por la Condición 1.4.2, se tiene que $T_\alpha v$ tiene w -norma finita, ya que para todo $x \in \mathbb{X}$, y $a \in \mathbb{A}(x)$,

$$\begin{aligned} |v(x, a)| &= \left| r(x, a) + \Delta_\alpha(x, a) \int v(y)Q(dy | x, a) \right| \\ &\leq |r(x, a)| + \Delta_\alpha(x, a) \int |v(y)| Q(dy | x, a) \\ &\leq cw(x) + \Delta_\alpha(x, a) \|v\|_w \int w(y)Q(dy | x, a) \\ &\leq cw(x) + \Delta_\alpha(x, a) \|v\|_w \beta w(x) \\ &\leq (c + \beta \Delta_\alpha(x, a) \|v\|_w)w(x) \\ &< (c + \beta \|v\|_w)w(x). \end{aligned}$$

Para cada $\alpha \in (0, 1)$, se probará que T_α es un operador contracción en el espacio $\mathbb{B}_w(\mathbb{X})$.

Para ello sean v y v' en $\mathbb{B}_w(\mathbb{X})$, tal que

$$v \leq v',$$

entonces

$$T_\alpha v \leq T_\alpha v',$$

esto dice que T_α es un operador monótono. También se observa que para cualquier $r \in \mathbb{R}$,

$$T_\alpha(v + rw)(x) \leq T_\alpha v(x) + (\alpha\beta)rw(x), \quad \forall x \in \mathbb{X}.$$

Por lo que se tiene que T_α es una contracción sobre $\mathbb{B}_w(\mathbb{X})$ con módulo $\gamma = \alpha\beta$. ■

Teorema 1.5.5 *Se supone que se cumplen las Condiciones 1.4.1, 1.4.2 y 1.4.4. Entonces*

a) El valor de la función V_α^* es la única solución de la EO (1.24) en el espacio $\mathbb{B}_w(\mathbb{X})$ y

$$\|v_n - V_\alpha^*\|_w \rightarrow 0 \text{ cuando } n \rightarrow \infty, \quad (1.32)$$

donde,

$$v_n(x) = \max_{a \in \mathbb{A}(x)} \left\{ r(x, a) + \Delta_\alpha(x, a) \int_{\mathbb{X}} v_{n-1}(y) Q(dy | x, a) \right\}.$$

b) Existe $f^* \in \mathbb{F}$ que maximiza el lado derecho en (1.24), es decir,

$$V_\alpha^*(x) = r(x, f^*) + \Delta_\alpha(x, f^*) \int_{\mathbb{X}} V_\alpha^*(y) Q(dy | x, f^*), \quad \forall x \in \mathbb{X}, \quad (1.33)$$

y la política estacionaria $\pi = \{f^*, f^*, \dots\}$ es óptima.

Demostración. Para demostrar a) dado el Lema 1.5.2 y el Teorema del punto fijo de Banach (A.3.2), se tiene que T_α tiene un punto fijo único $u^* \in \mathbb{B}_w(\mathbb{X})$, es decir,

$$T_\alpha u^* = u^*. \quad (1.34)$$

Por lo que para probar la parte (a), se necesita mostrar que

1. $V_\alpha^* \in \mathbb{B}_w(\mathbb{X})$.
2. $V_\alpha^* = u^*$.

Para probar 1, note que de la Condición 1.4.2(a) y por (1.13) se tiene que

$$\begin{aligned} E_x^\pi |r(x_n, a_n)| &= E_x^\pi |D(x_n, a_n) + \tau_\alpha(x_n, a_n)d(x_n, a_n)| \\ &\leq E_x^\pi \left[\bar{c}_1 w(x_n) + \frac{\bar{c}_2}{\alpha} w(x_n) \right] = c E_x^\pi [w(x_n)], \end{aligned} \quad (1.35)$$

de donde $c = \bar{c}_1 + \frac{\bar{c}_2}{\alpha}$.

Por otro lado, para todo $n = 0, 1, \dots$, se obtiene

$$E_x^\pi [w(x_n)] \leq \beta^n w(x), \quad (1.36)$$

y

$$E_x^\pi |r(x_n, a_n)| \leq \beta^n c w(x). \quad (1.37)$$

Como (1.36) se cumple para $n = 0$. Ahora si $n \geq 1$, se tiene que

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

por lo que

$$E_x^\pi [w(x_n)] \leq \beta E_x^\pi [w(x_{n-1})], \quad (1.38)$$

iterando (1.38) como se hizo en Lema 1.4.3, se obtiene (1.36). Para mostrar (1.37), por (1.35) y (1.36) se tiene

$$\begin{aligned} E_x^\pi |r(x_n, a_n)| &\leq c E_x^\pi [w(x_n)] \\ &\leq c \beta^n w(x). \end{aligned}$$

Note que 1 se obtiene por (1.20) y la Proposición 1.4.6(a)

$$\begin{aligned} |V(\pi, x)| &\leq E_x^\pi \left[\sum_{n=0}^{\infty} \prod_{j=0}^n \Delta_\alpha(x_j, a_j) |r(x_n, a_n)| \right] \\ &\leq \sum_{n=0}^{\infty} \rho_\alpha^n E_x^\pi |r(x_n, a_n)| \\ &\leq \sum_{n=0}^{\infty} \rho_\alpha^n c \beta^n w(x) \\ &\leq c w(x) \sum_{n=0}^{\infty} (\rho_\alpha \beta)^n \\ &\leq c \left(\frac{1}{1 - \rho_\alpha \beta} \right) w(x). \end{aligned}$$

Por (1.18) y tomando $M := c \left(\frac{1}{1 - \rho_\alpha \beta} \right)$, de lo anterior se tiene

$$V_\alpha^*(x) \leq M w(x). \quad (1.39)$$

En consecuencia, $V_\alpha^* \in \mathbb{B}_w(\mathbb{X})$.

Para probar 2, note que

$$\lim_{n \rightarrow \infty} \Lambda_\alpha^n E_x^\pi [u(x_n)] = 0, \quad \forall \pi \in \Pi, \quad x \in \mathbb{X}, \quad u \in \mathbb{B}_w(\mathbb{X}). \quad (1.40)$$

Por definición de w -norma se obtiene

$$|u(x_n)| \leq \|u\|_w w(x_n),$$

tomando esperanzas y multiplicando por Λ_α^n

$$\begin{aligned} \Lambda_\alpha^n E_x^\pi |u(x_n)| &\leq \|u\|_w \Lambda_\alpha^n E_x^\pi [w(x_n)] \\ &\leq \|u\|_w \Lambda_\alpha^n \beta^n w(x), \end{aligned}$$

en general para cualquier N se tiene que,

$$u^*(x) = E_x^f \left[\sum_{n=0}^{N-1} \Lambda_\alpha^n r(x_n, f) \right] + \Lambda_\alpha^N E_x^f [u^*(x_N)], \quad \forall N = 1, 2, \dots, \quad (1.42)$$

donde

$$\begin{aligned} \Lambda_\alpha^0 & : = 1, \\ \Lambda_\alpha^k & : = \prod_{j=0}^{k-1} \Delta_\alpha(x_j, a_j), \quad k \in \mathbb{N}, \end{aligned}$$

y tendiendo N a ∞ en (1.42) y por (1.40) se obtiene

$$u^*(x) = E_x^f \left[\sum_{n=0}^{\infty} \Lambda_\alpha^n r(x_n, f) \right] = V(x, f).$$

Así por la Definición 1.18, se observa que

$$u^*(x) \leq V_\alpha^*(x). \quad (1.43)$$

Para ver la otra desigualdad, note que (1.34) implica que

$$u^*(x) \geq r(x, a) + \Delta_\alpha(x, a) \int u^*(y) Q(dy | x, a), \quad \forall (x, a) \in \mathbb{K}. \quad (1.44)$$

Nuevamente iterando la desigualdad (1.44). Para cualquier política $\pi \in \Pi$ y un estado inicial $x \in \mathbb{X}$, se tiene

$$u^*(x) \geq E_x^\pi \left[\sum_{n=0}^{N-1} \Lambda_\alpha^n r(x_n, a_n) \right] + \Lambda_\alpha^N E_x^\pi u^*(x_N), \quad \forall N = 1, 2, \dots,$$

por último, tendiendo N a ∞ , por (1.40), se sigue que

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

como x y π son arbitrarios, se concluye que

$$u^*(x) \geq V_\alpha^*(x), \quad \forall x \in \mathbb{X}. \quad (1.45)$$

Por lo tanto, de las desigualdades (1.45) y (1.43) se obtiene 2.

Para ver que se cumple (1.32) se tiene que para cada $x \in \mathbb{X}$, $\pi \in \Pi$, de (1.22) y (1.23) se obtiene

$$\begin{aligned}
V(x, \pi) &= E_x^\pi \left[\sum_{k=0}^{\infty} \Lambda_\alpha^k r(x_k, a_k) \right] \\
&= E_x^\pi \left[\sum_{k=0}^{n-1} \Lambda_\alpha^k r(x_k, a_k) \right] + E_x^\pi \left[\sum_{k=n}^{\infty} \Lambda_\alpha^k r(x_k, a_k) \right] \\
&= V_n(x, \pi) + E_x^\pi \left[\Lambda_\alpha^n \sum_{k=n}^{\infty} \Lambda_\alpha^{k-n} r(x_k, a_k) \right]. \tag{1.46}
\end{aligned}$$

Por otro lado para cada $x \in \mathbb{X}$,

$$\begin{aligned}
|v_n(x) - V_\alpha^*(x)| &= \left| \sup_{\pi \in \Pi} V_n(x, \pi) - \sup_{\pi \in \Pi} V(x, \pi) \right| \\
&\leq \sup_{\pi \in \Pi} |V_n(x, \pi) - V(x, \pi)|,
\end{aligned}$$

usando las relaciones (1.46) y (1.37), así por la Proposición 1.4.6(b) se tiene lo siguiente

$$\begin{aligned}
|v_n(x) - V_\alpha^*(x)| &\leq \rho_\alpha^n \sup_{\pi \in \Pi} \sum_{k=n}^{\infty} \rho_\alpha^{k-n} E_x^\pi |r(x_k, a_k)| \\
&\leq \rho_\alpha^n \sup_{\pi \in \Pi} \sum_{k=n}^{\infty} c\beta^n w(x) \rho_\alpha^{k-n} \\
&\leq \rho_\alpha^n \beta^n c w(x) \frac{1}{1 - \rho_\alpha} = M(\rho_\alpha \beta)^n w(x).
\end{aligned}$$

Por lo que de la definición de $\|\cdot\|_w$ lo anterior implica que

$$\|v_n - V_\alpha^*\|_w \leq M(\rho_\alpha \beta)^n,$$

cuando $n \rightarrow \infty$ y usando la Proposición (1.4.6) del hecho que $\rho_\alpha < 1$ y $\beta \in (0, 1)$ se concluye la demostración de parte (a).

b) Para esta prueba, se recuerda que $\Delta_\alpha(x, a)$ y $\tau_\alpha(x, a)$ son continuas en $\mathbb{A}(x)$ para cada $x \in \mathbb{X}$. Entonces por el Lema 1.5.2, existe $f^* \in \mathbb{F}$ que satisface los siguiente

$$V_\alpha^*(x) = r(x, f^*) + \Delta_\alpha(x, f^*) \int_{\mathbb{X}} V_\alpha^*(y) Q(dy | x, f^*), \quad x \in \mathbb{X}.$$

Iterando esta ecuación se tiene que

$$V_\alpha^*(x) = E_x^{f^*} \left[\sum_{n=0}^{N-1} \Lambda_\alpha^n r(x_n, a_n) + \Lambda_\alpha^N V_\alpha^*(x_N) \right],$$

de donde,

$$V_\alpha^*(x) \geq V_N(x, f^*), \quad \forall x \in \mathbb{X}, \quad N \geq 1, \quad (1.47)$$

haciendo tender N a ∞ se tiene que

$$V_\alpha^*(x) \geq V(x, f^*), \quad \forall x \in \mathbb{X}, \quad (1.48)$$

por otra parte, de (1.18) y (1.19) se obtiene que

$$V_\alpha^*(x) \leq V(x, f^*), \quad \forall x \in \mathbb{X}. \quad (1.49)$$

Entonces

$$V_\alpha^*(x) = V(x, f^*), \quad \forall x \in \mathbb{X},$$

y por lo tanto, $\pi = \{f^*, f^*, \dots\}$ es óptima. ■

Capítulo 2

Uniformización

En este capítulo, se presenta la técnica de uniformización, el cual consiste en transformar un Proceso de Control de Markov a tiempo continuo con tasas de transición exponenciales a un modelo equivalente en que los estados son independientes y las tasas de transición son constantes. Dicha técnica descrita en este capítulo será fundamental para el trabajo de tesis, ya que ésta es una manera de analizar el modelo, misma que se describirá en el Capítulo 3.

2.1. Uniformización

Sea S el conjunto de estados numerable, y $X = \{X_t : t \geq 0\}$ un proceso estocástico con espacio de estados S , se dice que X es una cadena de Markov a tiempo continuo, si para cualquier $u \geq 0$ y $j \in S$

$$P\{X_{t+u} = j \mid X_u; t \geq u \geq 0\} = P\{X_{t+u} = j \mid X_t\}.$$

Además, si para todo $u \geq 0$

$$P\{X_{t+u} = j \mid X_u = s\}$$

es independiente de u , entonces la cadena de Markov a tiempo continuo se dice que tiene probabilidades de transición estacionarias u homogéneas.

Dicho proceso se desarrolla como sigue: si en algún instante t , $X_t = s$, el proceso se mantiene en el estado s en un tiempo aleatorio, determinado por una distribución exponencial con parámetro $\beta(s)$, $0 \leq \beta(s) < \infty$, y este

transita al estado j con probabilidad $Q(j | s)$. Se puede resumir el comportamiento probabilístico del proceso en términos de un generador infinitesimal, donde un generador infinitesimal se refiere a una matriz A con componentes

$$A(j | s) = \begin{cases} -[1 - Q(s | s)]\beta(s) & j = s, \\ Q(j | s)\beta(s) & j \neq s. \end{cases}$$

Por consiguiente, los procesos con el mismo generador infinitesimal tienen distribuciones idénticas finita dimensionales, siempre que éstos tengan la misma distribución inicial. Esto sugiere que si se modifica el proceso y al mismo tiempo no se altera su generador infinitesimal, entonces el proceso modificado tendría la misma estructura probabilística.

Cuando $\sup_{s \in S} \beta(s) < \infty$, se deriva un proceso modificado \tilde{X} de X como sigue.

Se elige una constante $C < \infty$ tal que satisfaga

$$\sup_{s \in S} [1 - Q(s | s)] \beta(s) \leq C < \infty, \quad (2.1)$$

y se define un proceso $\{\tilde{X}_t : t \geq 0\}$ con estados independientes, tiempos de permanencia exponenciales $\tilde{\beta}(j) = C$, y probabilidades de transición $\tilde{Q}(j | s)$ dado por

$$\tilde{Q}(j | s) = \begin{cases} 1 - \frac{[1 - Q(s | s)]\beta(s)}{C} & j = s, \\ \frac{Q(j | s)\beta(s)}{C} & j \neq s. \end{cases} \quad (2.2)$$

Donde $\tilde{A}(j | s) = C\tilde{Q}(j | s)$ para $j \neq s$, se dice que \tilde{A} es el generador infinitesimal de \tilde{X} y satisface que

$$\tilde{A} = A,$$

por lo que los dos procesos son equivalentes en distribución. Se refiere a \tilde{X} como la uniformización de X , por que tiene distribución de tiempos de permanencia idénticas en cada estado. La uniformización puede ser vista como un proceso equivalente, en el cual el estado del sistema es observado en tiempos aleatorios los cuales están distribuidos exponencialmente con parámetro C , (véase [15]).

2.1.1. Ejemplo

El siguiente ejemplo muestra la técnica de uniformización.

Considérese una cadena de Markov a tiempo continuo con espacio de estados en $S = \{s_1, s_2\}$, con probabilidades de transición descrita de la siguiente forma:

$$\begin{aligned} Q(s_1 \mid s_1) &= 0, \\ Q(s_2 \mid s_1) &= 1, \\ Q(s_1 \mid s_2) &= 1, \\ Q(s_2 \mid s_2) &= 0. \end{aligned}$$

Supóngase que los tiempos de permanencia en s_1 y s_2 tienen una distribución exponencial con parámetros 2 y 0.8 respectivamente, es decir, $\beta(s_1) = 2$ y $\beta(s_2) = 0.8$. Por lo que el generador infinitesimal para este proceso está dado por

$$A = \begin{bmatrix} -2 & 2 \\ 0.8 & -0.8 \end{bmatrix}.$$

Ahora se deriva la uniformización \tilde{X} para este proceso, donde se elige una constante $C = 4$, la cual satisface la condición dada en (2.1) y usando (2.2), se obtiene que

$$\tilde{Q} = \begin{bmatrix} 0.5 & 0.5 \\ 0.2 & 0.8 \end{bmatrix},$$

usando que

$$\tilde{Q} = I - Q,$$

se tiene lo siguiente

$$\begin{aligned} Q &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0.5 & 0.5 \\ 0.2 & 0.8 \end{bmatrix} \\ &= \begin{bmatrix} 0.5 & -0.5 \\ -0.2 & 0.2 \end{bmatrix}, \end{aligned}$$

así que

$$\begin{aligned} \tilde{A} &= C(I - Q) \\ &= C\tilde{Q}, \end{aligned}$$

haciendo algunas operaciones se tiene que

$$\begin{aligned}\tilde{A} &= 4 \left[\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0.5 & -0.5 \\ -0.2 & 0.2 \end{bmatrix} \right] \\ &= 4 \begin{bmatrix} 0.5 & 0.5 \\ 0.2 & 0.8 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 2 \\ 0.8 & 3.2 \end{bmatrix},\end{aligned}$$

por lo que se cumple que $\tilde{A} = A$ para $j \neq s$, por lo tanto los dos procesos tienen el mismo comportamiento probabilístico.

2.2. Modelo Descontado

Se considera un PCSM, en el cual las épocas de decisión son tomadas en cada transición de estado y los tiempos entre épocas están distribuidos exponencialmente, es decir, la distribución está dada por

$$F(t | s, a) = 1 - e^{-\beta(s,a)t}, \text{ para } t \geq 0.$$

Sea $Q(t | s, a)$ la probabilidad que en la siguiente época de decisión el sistema se encuentre en el estado j , si en la época de decisión actual está en el estado s y se elige una acción $a \in \mathbb{A}(s)$. Se supone que se tiene la misma estructura de recompensa definida en el Capítulo 1. Por lo que $r(s, a)$ es la recompensa esperada descontada entre épocas de decisión la cual satisface,

$$\begin{aligned}r(s, a) &= k(s, a) + c(s, a) E_s^a \left\{ \int_0^{\tau_1} e^{-\alpha t} dt \right\} \\ &= k(s, a) + c(s, a) E_s^a \left\{ [1 - e^{-\alpha \tau_1}] / \alpha \right\} \\ &= k(s, a) + c(s, a) / [\alpha - \beta(s, a)].\end{aligned}$$

Para $f \in \mathbb{F}$, sea $Q_f(j | s) = Q(j | s, f(s))$, $r_f(s) = r(s, f(s))$, $\beta_f(s) = \beta(s, f(s))$, y Q_f es la matriz con componentes $Q_f(j | s)$.

Sea $f^\infty = (f^*, f^*, \dots)$ una política estacionaria determinista y dado un proceso de control Markoviano a tiempo continuo se tiene que

$$\begin{aligned}v_\alpha^{f^\infty}(s) &= r_f(s) + \sum_{j \in S} \left[\int_0^\infty \beta_f(s) e^{-[\alpha - \beta_f(s)]t} \right] Q_f(j | s) v_\alpha^{f^\infty}(j) \\ &= r_f(s) + \frac{\beta_f(s)}{\beta_f(s) + \alpha} \sum_{j \in S} Q_f(j | s) v_\alpha^{f^\infty}(j).\end{aligned}$$

Si $\beta(s, a) = C$ para todo $s \in S$ y $a \in \mathbb{A}(s)$, $\beta_f(s) = C$ para todo $s \in S$, así se tiene que

$$v_\alpha^{f\infty}(s) = r_f(s) + \lambda \sum_{j \in S} Q_f(j | s) v_\alpha^{f\infty}(j),$$

con $\lambda = C/(C + \alpha)$. Por lo que se sigue que

$$v(s) = \max_{a \in A(s)} \left\{ r(s, a) + \lambda \sum_{j \in S} Q_f(j | s) v(j) \right\}.$$

La uniformización puede ser aplicada en los procesos de control Markovianos a tiempo continuo obteniendo un modelo con tasas de transición constante, por lo que los resultados y algoritmos para los modelos descontados a tiempo discreto pueden ser usados directamente (véase [15], [4] y [19]).

Así, $r(s, a)$ depende explícitamente de $\beta(s, a)$, asegurando que el sistema transformado tiene la misma recompensa total descontada con una tasa de recompensa continua $C(s, a)$.

Suposición 2.2.1 *Existe una constante $C < \infty$ para la cual*

$$[1 - Q(s | s, a)] \beta(s, a) \leq C, \quad (2.3)$$

para todo $s \in S$ y $a \in \mathbb{A}(s)$.

Se define una uniformización de un proceso de control de Markov a tiempo continuo con componentes denotados por tilde como sigue. Sea $\tilde{S} = S$, $\tilde{\mathbb{A}}(s) = \mathbb{A}(s)$ para todo $s \in \tilde{S}$,

$$r(s, a) \equiv r(s, a) \frac{\alpha + \beta(s, a)}{\alpha + C}, \quad (2.4)$$

y

$$\tilde{Q}(j | s, a) = \begin{cases} 1 - \frac{[1 - Q(s | s, a)] \beta(s, a)}{C} & j = s, \\ \frac{Q(j | s, a) \beta(s, a)}{C} & j \neq s. \end{cases} \quad (2.5)$$

Sea $\tilde{r}_f(s) \equiv \tilde{r}(s, f(s))$, $Q_f(j | s) \equiv Q(j | s, f(s))$ y \tilde{Q}_f es la matriz con componentes $\tilde{Q}_f(j | s)$. Además, $\tilde{v}_\alpha^{f\infty}$ denota la recompensa esperada descontada para el proceso transformado.

El siguiente resultado se refiere a los modelos originales y uniformizados.

Proposición 2.2.2 *Considérese la Supocisión 2.2.1 y que \tilde{r} y \tilde{Q} satisfacen (2.4) y (2.5). Entonces, para cada $f \in \mathbb{F}$ y $s \in S$,*

$$\tilde{v}_\alpha^{f^\infty}(s) = v_\alpha^{f^\infty}(s).$$

La prueba de la proposición anterior puede consultarse en [15].

Capítulo 3

Ejemplos

El modelo estudiado en el Capítulo 2, tiene algunas aplicaciones en problemas de control, particularmente en líneas de espera. Dicho modelo es analizado permitiendo la elección de acciones en un tiempo aleatorio sobre $[0, \infty)$. En este capítulo se pretende ilustrar un modelo de PCSMs a través de un modelo de línea de espera controlada a tiempo continuo.

Este modelo se puede analizar mediante dos formas.

1. Usando los resultados de los Procesos de Control Semi-Markovianos del Capítulo 1 directamente.
2. Transformando el modelo a un modelo equivalente, analizado mediante uniformización, visto en Capítulo 2.

3.1. Una línea M/M/1 de espera controlada, con un solo tipo de servicio.

El problema se formula como un modelo de control de admisión en forma de un PCSM. En tal sistema, los tiempos de transición están distribuidos exponencialmente y las acciones son elegidas en cada época de decisión. En este caso, se considera el modelo $M/M/1$ discutido en [15], en el cual se supone que todos los tiempos entre llegadas son independientes e idénticamente distribuidos de acuerdo a una distribución exponencial con tasa μ , es decir, el proceso de entrada es Poisson, además todos los tiempos de servicio están distribuidos de acuerdo a otra distribución exponencial con tasa λ .

El objetivo del problema es maximizar la recompensa total esperada descontada, donde $\alpha \in (0, 1)$ es la tasa descontada en tiempo continuo. Se supone que el estado se denota como $\langle s, b \rangle$, es decir, cuando hay s clientes en el sistema y b ocurre, donde $b = 1$, si un cliente entra al sistema y $b = 0$, si no hay llegada de clientes. En esta situación, las épocas de decisiones ocurren cuando llega o parte un cliente.

Para este ejemplo en particular se considera una compañía, que se encarga en proporcionar un servicio a particulares y empresas.

3.1.1. Descripción del modelo

A partir de la idea general sobre el sistema, el modelo matemático se formula con las valores de costo y distribuciones definidas a continuación:

Se considera un MCSM

$$(S, \mathbb{A}, \{\mathbb{A}(s) : s \in S\}, Q, H, D, d).$$

■ Espacio de estados

Se define el espacio de estados como el conjunto del número posible de clientes en el sistema, denotado por S .

$$S = \{0, 1, 2, 3, \dots\} \times \{0, 1\}. \quad (3.1)$$

Se tiene que el estado del sistema es observado en cada arribo o partida.

■ Espacio de acciones

Para todo $\langle s, b \rangle \in S$, donde $s \in \{0, 1, 2, 3, \dots\}$, $b \in \{0, 1\}$, el espacio de acciones $\mathbb{A}(s) = \mathbb{A}$, es definido como:

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

es decir, en el estado $\langle s, 0 \rangle$ donde hay s clientes en el sistema pero no hay arribos la acción es continuar, por lo que $\mathbb{A}(\langle s, 0 \rangle) = \{0\}$ y en el estado $\langle s, 1 \rangle$ donde hay s clientes en el sistema y un nuevo cliente llega, el controlador puede admitir o rechazar servicio a la llegada, por lo que $\mathbb{A}(\langle s, 1 \rangle) = \{0, 1\}$, donde la acción 0 es rechazar y 1 es admitir.

■ Distribución del tiempo de servicio

Se tiene que

$$\beta(\langle s, b \rangle, a) = \begin{cases} \mu & \text{si } s = 0, a = 0 \text{ y } b = 0, 1, \\ \mu + \lambda & \text{si } s = 0, a = 1 \text{ y } b = 1 \text{ o } s > 0. \end{cases}$$

donde $\beta(\langle s, b \rangle, a)$ denota las tasas de servicio para cada estado dado la acción, para ver ésto, se puede observar que, cuando el sistema está vacío, es decir, cuando se observa el estado $\langle 0, 0 \rangle$ y $a = 0$ o $\langle 0, 1 \rangle$ y $a = 0$, la siguiente época de decisión ocurre cuando el siguiente cliente entra al sistema de acuerdo a una distribución exponencial con parámetro μ y en cualquier otro estado, la siguiente época de decisión ocurre después de la llegada o culminación del servicio. Si T_a el tiempo hasta el último arribo y T_s el tiempo hasta la siguiente terminación del servicio, entonces el tiempo de la siguiente época de decisión T_e satisface $T_e = \min(T_a, T_s)$, es posible mostrar que T_e sigue una distribución exponencial con parámetro $\mu + \lambda$. Por lo que se tiene que,

$$H(t \mid \langle s, b \rangle, a) = (\mu + \lambda)e^{-(\mu + \lambda)t}. \quad (3.2)$$

■ **Probabilidades de transición**

Las probabilidades de transición satisfacen:

$$Q(j \mid s, a) = \begin{cases} 1, & j = \langle 0, 1 \rangle, a = 0 \text{ y } s = \langle 0, 0 \rangle, s = \langle 0, 1 \rangle, \\ \frac{\mu}{\mu + \lambda}, & j = \langle m - 1, 0 \rangle, s = \langle m, k \rangle, a = 0 \text{ para } m \geq 1, \\ & \text{y } k = 0, 1, \text{ o} \\ \frac{\lambda}{\mu + \lambda}, & j = \langle m, 0 \rangle, s = \langle m, 1 \rangle, a = 1 \text{ para } m \geq 0, \\ & j = \langle m + 1, 1 \rangle, s = \langle m, 1 \rangle, a = 1 \text{ para } m \geq 0, \\ & \text{o} \\ j = \langle m, 1 \rangle, s = \langle m, k \rangle, a = 0 \text{ para } m \geq 1 \text{ y } k = 0, 1, \\ 0, & \text{en otro caso,} \end{cases} \quad (3.3)$$

para ver ésto, se observa que, cuando el sistema está vacío, la siguiente época de decisión ocurre cuando el sistema comienza en $\langle 0, 1 \rangle$. Si en algún instante de tiempo, el sistema contiene m clientes y no hay llegadas, el sistema está disponible para la admisión o un cliente ha rechazado el servicio, entonces el estado en la siguiente época de decisión puede ser $\langle m, 1 \rangle$ o $\langle m - 1, 0 \rangle$, dependiendo si hay una llegada o una terminación de servicio. Si hay m clientes y una llegada, y está ha sido admitida, entonces el siguiente estado puede ser $\langle m + 1, 1 \rangle$ o $\langle m, 0 \rangle$. Las probabilidades de transición se obtienen por lo siguiente $P\{T_a \leq T_s\} = \lambda/(\mu + \lambda)$ y $P\{T_s \leq T_a\} = \mu/(\mu + \lambda)$.

Con excepción de los estados $\langle 0, 0 \rangle$ y $\langle 0, 1 \rangle$, todas las transiciones ocurren con tasa $\mu + \lambda$.

Otra forma de observar esto, es mediante el siguiente diagrama.

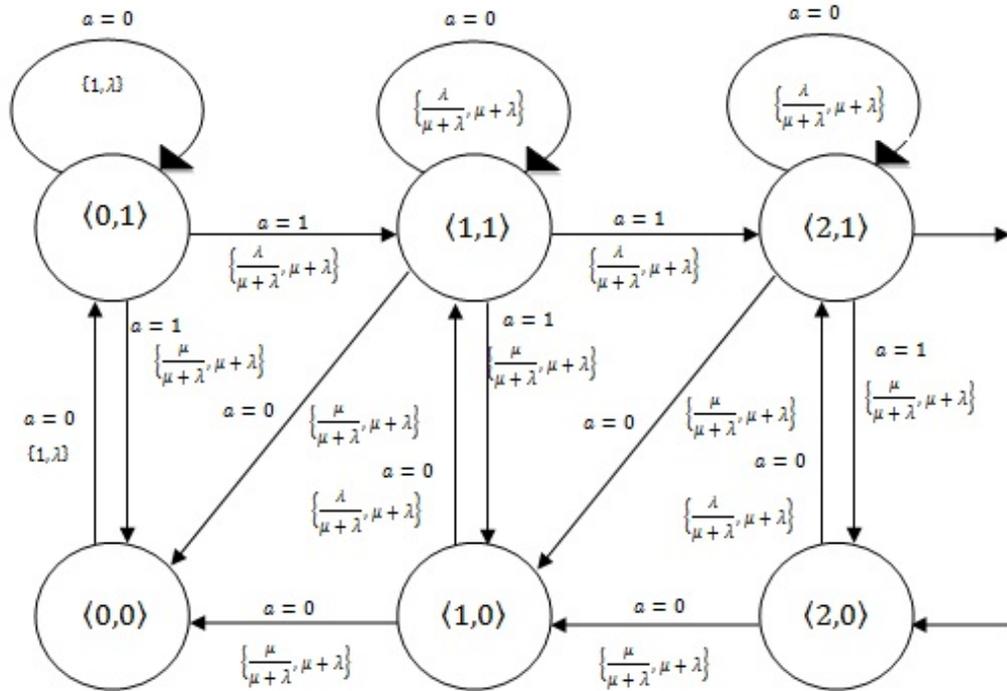


Diagrama 1. Representa la estructura de transición del modelo de admisión. Los números entre llaves representan las probabilidades de transición y las tasas de transición exponencial respectivamente.

■ Funciones de recompensa

La función de recompensa definida en el Capítulo 1 (1.6), está dada por una ganancia inmediata D , más un costo debido al tiempo de permanencia en el estado s , cuya razón de costo es d .

$$r(\langle s, b \rangle, a) = D(\langle s, b \rangle, a) - d(\langle s, b \rangle, a) \int_0^\infty \int_0^t e^{-\alpha s} ds H(dt | \langle s, b \rangle, a). \quad (3.4)$$

Donde $D(\langle s, b \rangle, a) = R$ para $b = 1$ y $D(\langle s, b \rangle, a) = 0$ para $b = 0$. Además se define $d(\langle s, 0 \rangle, a) = c(s)$ y $d(\langle s, 1 \rangle, a) = c(s + 1)$, donde c es una función

continua y creciente para todo $s \in S$, y $c(0) = 0$. Por lo que las funciones de recompensa para cada $s \in S$, son las siguientes.

$$\begin{aligned}
 r(\langle s, 0 \rangle, 0) &= -c(s) \int_0^\infty \int_0^t e^{-\alpha s} ds H(dt | \langle s, 0 \rangle, 0) & (3.5) \\
 &= -\frac{c(s)(\mu + \lambda)}{\alpha} \int_0^\infty [e^{-(\mu+\lambda)t} - e^{-(\alpha+\mu+\lambda)t}] dt \\
 &= -\frac{c(s)}{\alpha + \mu + \lambda},
 \end{aligned}$$

$$\begin{aligned}
 r(\langle s, 1 \rangle, 1) &= R - c(s+1) \int_0^\infty \int_0^t e^{-\alpha s} ds H(dt | \langle s, 1 \rangle, 0) & (3.6) \\
 &= R - \frac{c(s+1)(\mu + \lambda)}{\alpha} \int_0^\infty [e^{-(\mu+\lambda)t} - e^{-(\alpha+\mu+\lambda)t}] dt \\
 &= R - \frac{c(s+1)}{\alpha + \mu + \lambda}.
 \end{aligned}$$

Debido a que los PCSMs permiten tomar decisiones en cualquier instante de tiempo, en este caso, se considera un modelo en el cual las decisiones sólo son tomadas en tiempos de transición. Esto se analiza, transformando el modelo a un modelo equivalente, a través de uniformización (véase Capítulo 2).

Para uniformizar el sistema se elige una constante $C = \mu + \lambda$ y se altera la estructura de transición en aquellos estados. Para esto se aplica (2.5) y se tiene las probabilidades transformadas, en el estado $\langle 0, 0 \rangle$

$$\begin{aligned}
 \tilde{Q}(\langle 0, 1 \rangle | \langle 0, 0 \rangle, 0) &= Q(\langle 0, 1 \rangle | \langle 0, 0 \rangle, 0) \beta(\langle 0, 1 \rangle, 0) / C = \lambda / (\mu + \lambda), \\
 \tilde{Q}(\langle 0, 0 \rangle | \langle 0, 0 \rangle, 0) &= \mu / (\mu + \lambda).
 \end{aligned}$$

En el modelo sin transformar, eligiendo $a = 0$ en el estado $\langle 0, 1 \rangle$, instantáneamente se mueve al estado $\langle 0, 0 \rangle$,

$$\begin{aligned}
 \tilde{Q}(\langle 0, 0 \rangle | \langle 0, 1 \rangle, 0) &= \mu / (\mu + \lambda), \\
 \tilde{Q}(\langle 0, 1 \rangle | \langle 0, 1 \rangle, 0) &= \lambda / (\mu + \lambda).
 \end{aligned}$$

Se tiene que para los demás estados $\tilde{Q}(j | s, a) = Q(j | s, a)$.

Ahora se analiza el modelo descontado. Se asume que $c(0) = 0$, así que la función de recompensa para cada caso en el sistema uniformizado satisface:

$$\begin{aligned}\tilde{r}(\langle 0, 0 \rangle, 0) &= \tilde{r}(\langle 0, 1 \rangle, 0), \\ \tilde{r}(\langle s, 1 \rangle, 1) &= R - \frac{c(s+1)}{\alpha + \mu + \lambda} \quad s \geq 0, \\ \tilde{r}(\langle s, 0 \rangle, 0) &= \tilde{r}(\langle s, 1 \rangle, 0) = -\frac{c(s)}{\alpha + \mu + \lambda}, \quad s \geq 1.\end{aligned}$$

Por lo que las ecuaciones de optimalidad para este sistema son las siguientes:

$$\begin{aligned}v(\langle 0, 0 \rangle) &= \frac{\mu + \lambda}{\alpha + \mu + \lambda} \left[\frac{\mu}{(\mu + \lambda)} v(\langle 0, 0 \rangle) + \frac{\lambda}{(\mu + \lambda)} v(\langle 0, 1 \rangle) \right], \\ v(\langle s, 0 \rangle) &= -\frac{c(s)}{\alpha + \mu + \lambda} + \frac{\mu + \lambda}{\alpha + \mu + \lambda} \left[\frac{\mu}{(\mu + \lambda)} v(\langle s - 1, 0 \rangle) \right. \\ &\quad \left. + \frac{\lambda}{(\mu + \lambda)} v(\langle s, 1 \rangle) \right],\end{aligned}$$

para $s \geq 1$, y

$$v(\langle s, 1 \rangle) = \max \left\{ \begin{array}{l} R - \frac{c(s+1)}{\alpha + \mu + \lambda} \\ + \frac{\mu + \lambda}{\alpha + \mu + \lambda} \left[\frac{\mu}{(\mu + \lambda)} v(\langle s, 0 \rangle) + \frac{\lambda}{(\mu + \lambda)} v(\langle s + 1, 1 \rangle) \right], \\ -\frac{c(s)}{\alpha + \mu + \lambda} \\ + \frac{\mu + \lambda}{\alpha + \mu + \lambda} \left[\frac{\mu}{(\mu + \lambda)} v(\langle s - 1, 0 \rangle) + \frac{\lambda}{(\mu + \lambda)} v(\langle s, 1 \rangle) \right] \end{array} \right\},$$

note que la ecuación $v(\langle s, 0 \rangle)$ simplifica a $v(\langle s, 1 \rangle)$, es decir,

$$v(\langle s, 1 \rangle) = \max \{ R + v(\langle s + 1, 0 \rangle), \quad v(\langle s, 0 \rangle) \}, \quad (3.7)$$

para $s \geq 0$. Se sigue del Teorema 1.5.5(b) que la política estacionaria $\pi^* = (f^*, f^*, \dots)$ derivado de la regla de decisión

$$f^*(s) = \begin{cases} 1 & v_\alpha^*(\langle s + 1, 0 \rangle) - v_\alpha^*(\langle s, 0 \rangle) > -R, \\ 0 & v_\alpha^*(\langle s + 1, 0 \rangle) - v_\alpha^*(\langle s, 0 \rangle) \leq -R, \end{cases} \quad (3.8)$$

es óptima. Por lo tanto $v_\alpha^*(\langle s + 1, 0 \rangle) - v_\alpha^*(\langle s, 0 \rangle)$ es monótona decreciente y es una política de control óptimo.

Un ejemplo numérico se muestra a continuación.

Ejemplo 3.1.1 *Se considera un sistema donde se controla la admisión. Supóngase que la suma total de ganancias está dada por $R = 20$, sea la función de costo por servicio dado por $c(s) = 1 \cdot 3^s, s \in \{0, 1, \dots\}$. Además la tasa de llegada es $\mu = 0 \cdot 05$ y la tasa de servicio es $\lambda = 0 \cdot 2$. Supóngase que la tasa descontada a tiempo discreto es $\gamma = 0 \cdot 9$, entonces la tasa descontada a tiempo continuo es $\alpha = \ln \gamma = 0 \cdot 105$. La constante de uniformización es $C = \mu + \lambda = 0 \cdot 25$. Con esta información se da un conjunto de datos que muestra en la Tabla 1, las diferencias $v_\alpha^*(\langle s+1, 0 \rangle) - v_\alpha^*(\langle s, 0 \rangle)$, para distintos valores de s .*

$s=0$	$s=1$	$s=2$	$s=3$	$s=4$
3.25	1.20	-0.75	-2.76	-4.25
$s=5$	$s=6$	$s=7$	$s=8$	$s=9$
-7.46	-10.42	-13.87	-17.53	-21.32
$s=10$	$s=11$	$s=12$	$s=13$	$s=14$
-26.51	-33.46	-42.67	-54.79	-70.66

Tabla 1

De acuerdo a la regla de decisión $f^*(s)$, en la Tabla 1 se puede ver que de acuerdo a la regla de decisión el total de clientes que serán admitidos es 9, además se observa que las diferencias $v_\alpha^*(\langle s+1, 0 \rangle) - v_\alpha^*(\langle s, 0 \rangle)$, para distintos valores de s decrecen.

El siguiente ejemplo es una extensión del análisis del modelo M/M/1 con un solo tipo de servicio, el problema se puede extender a múltiples servidores con múltiples tipos de servicio.

3.2. Una línea M/M/n de espera controlada, con dos tipos de servicio.

Se considera nuevamente una compañía de Servicios, en el cual tiene segmentado sus mercados en 2 clases, tipo 1 y tipo 2. El problema se puede plantear como un modelo de líneas de espera M/M/n donde las tasas de llegada son Poisson con parámetros λ_1 y λ_2 respectivamente. Además, se supone las tasas de servicio tipo 1 y 2, están dadas por μ_1 y μ_2 respectivamente.

Al igual que en el caso anterior, nuestro objetivo será maximizar la recompensa total esperada descontada, donde $\alpha \in (0, 1)$ es la tasa de descuento a tiempo continuo.

3.2.1. Descripción del modelo

A partir de la idea general sobre el sistema, el modelo matemático se formula con los valores de costo y distribuciones definidas a continuación:

Nuevamente se considera un MCSM

$$(S, \mathbb{A}, \{\mathbb{A}(s) : s \in S\}, Q, H, D, d).$$

■ Espacio de estados

Se supone que el estado del sistema se denota como $s = \langle i, j, b \rangle$, donde hay i clientes tipo 1 y j clientes tipo 2 en el sistema y b ocurre, donde $b = 1$ denota la llegada de un cliente tipo 1 y $b = -1$ denota la salida de un cliente tipo 1. Similarmente, se denota $b = 2$ como la llegada de un cliente tipo 2 y $b = -2$ como la salida del cliente tipo 2. Supóngase que hay un total de n servidores disponibles, así que $i + j \leq n$, en todos los tiempos. Por lo que el espacio de estados se define como sigue.

$$S = \{1, 2, 3, \dots\} \times \{1, 2, 3, \dots\} \times \{-1, -2, 1, 2\}.$$

■ Espacio de acciones

Las épocas de decisión se toman cuando un cliente llega o parte del sistema. Cuando un cliente llega, las posibles acciones, a , deberán ser admitir al cliente, denotado por $a = 1$ o rechazar denotado por $a = 0$. Cuando no hay llegadas o cuando hay una salida, la única acción posible debe ser continuar, denotado por $a = 0$. Para todo $s \in S$, el espacio de acciones $\mathbb{A}(s) = \mathbb{A}$, es definido como:

$$\mathbb{A} = \{0, 1\}.$$

■ Probabilidades de transición

Sea β_{ijba} la tasa de transición al estado $\langle i, j, b \rangle$ dado la acción a . Sean las probabilidades de transición denotadas por $Q(k|\langle i, j, b \rangle)$, que representa la probabilidad de transición del estado $\langle i, j, b \rangle$ al estado $k = \langle i^*, j^*, b^* \rangle$.

Se puede observar, que el estado $\langle i, j, 1 \rangle$ cuando la decisión es aceptar a un cliente tipo 1 que llega, $a = 1$, el número de clientes tipo 1 en el sistema incrementa a $i + 1$, el siguiente evento podría ser la partida de un cliente tipo 1, es decir, $b = -1$, la partida de un cliente tipo 2, $b = -2$, el arribo de un cliente tipo 1, $b = 1$, el arribo de un cliente tipo 2, $b = 2$, es decir,

3.2 Una línea M/M/n de espera controlada, con dos tipos de servicio.

- * Cuando hay una partida de un cliente tipo 1 el sistema transitará al estado $\langle i, j, -1 \rangle$ con una tasa de $(i + 1)\mu_1$, donde hay $(i + 1)$ clientes en el sistema.
- * Cuando hay una partida de un cliente tipo 2, el sistema transitará al estado $\langle i + 1, j - 1, -2 \rangle$ con una tasa de $j\mu_2$,
- * Cuando hay una llegada de tipo 1, el sistema transitará al estado $\langle i + 1, j, 1 \rangle$ con una tasa λ_1 .
- * Cuando hay una llegada de tipo 2, el sistema transitará al estado $\langle i + 1, j, 2 \rangle$ con tasa λ_2 .

Donde cualquiera de los cuatro eventos podrán ocurrir, por lo que en la siguiente época de decisión ocurre con una tasa $\beta_{ij11} = (i+1)\mu_1 + j\mu_2 + \lambda_1 + \lambda_2$, y la probabilidad de transición dado el estado $\langle i, j, 1 \rangle$, donde $i + j \leq n$, es descrita como sigue:

$$Q(k|\langle i, j, 1 \rangle) = \begin{cases} (i + 1)\mu_1/\beta_{ij11}, & k = \langle i, j, -1 \rangle, \\ j\mu_2/\beta_{ij11}, & k = \langle i + 1, j - 1, -2 \rangle, j \geq 1, \\ \lambda_1/\beta_{ij11}, & k = \langle i + 1, j, 1 \rangle, \\ \lambda_2/\beta_{ij11}, & k = \langle i + 1, j, 2 \rangle. \end{cases}$$

de manera particular se puede ver en el siguiente Diagrama 2.

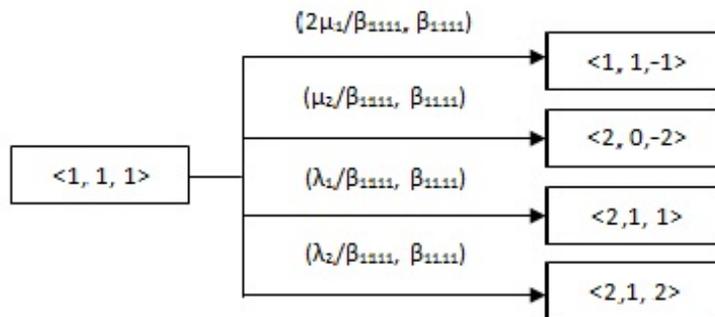


Diagrama 2

De manera análoga se puede describir las probabilidades de transición para los estados $\langle i, j, 2 \rangle$ y $\langle i, j, b \rangle$, donde $i + j \leq n$ y $b \in \{-2, -1, 1, 2\}$ escritas como sigue:

En el estado $\langle i, j, 2 \rangle$, donde $i + j \leq n$, es decir, cuando un cliente tipo 2 llega y aún hay servidores disponibles, y la acción es admitir, es decir, $a = 1$, las probabilidades de transición pueden ser escritas como:

$$Q(k | \langle i, j, 2 \rangle) = \begin{cases} i\mu_1/\beta_{ij21}, & k = \langle i-1, j+1, -1 \rangle, \quad i \geq 1, \\ (j+1)\mu_2/\beta_{ij21}, & k = \langle i, j-1, -2 \rangle, \\ \lambda_1/\beta_{ij21}, & k = \langle i, j+1, 1 \rangle, \\ \lambda_2/\beta_{ij21}, & k = \langle i, j+1, 2 \rangle. \end{cases}$$

donde la tasa de transición del estado es $\beta_{ij21} = i\mu_1 + (j+1)\mu_2 + \lambda_1 + \lambda_2$.

En todos los estados $\langle i, j, b \rangle$, donde $i + j \leq n$ y $b \in \{-2, -1, 1, 2\}$, y la acción es rechazar el servicio o continuar, $a = 0$, las probabilidades de transición pueden ser escritas como:

$$Q(k | \langle i, j, 2 \rangle) = \begin{cases} i\mu_1/\beta_{ijb0}, & k = \langle i-1, j, -1 \rangle, \quad i > 1, \\ j\mu_2/\beta_{ijb0}, & k = \langle i, j-1, -2 \rangle, \quad j > 1, \\ \lambda_1/\beta_{ijb0}, & k = \langle i, j, 1 \rangle, \\ \lambda_2/\beta_{ijb0}, & k = \langle i, j, 2 \rangle. \end{cases}$$

donde la tasa de transición del estado es $\beta_{ijb0} = i\mu_1 + j\mu_2 + \lambda_1 + \lambda_2$.

Note que

$$\begin{aligned} \beta_{ij11} &= \beta_{(i+1)jb0}, \\ \beta_{ij21} &= \beta_{i(j+1)b0}. \end{aligned}$$

■ Distribución de tiempos de servicio

Sea $H(t | \langle i, j, b \rangle, a)$ la función de distribución para cada $\langle i, j, b \rangle \in S$, donde los tiempos de servicio para cada estado siguen una distribución exponencial con parámetro β_{ij11} , β_{ij21} y β_{ijb0} respectivamente, así que podemos escribir a H como sigue:

$$\begin{aligned} H(t | \langle i, j, b \rangle, a) &= \beta_{ij11} e^{-\beta_{ij11}t} I_{\langle i, j, 1 \rangle}(\langle i, j, b \rangle) + \\ &\quad \beta_{ij21} e^{-\beta_{ij21}t} I_{\langle i, j, 2 \rangle}(\langle i, j, b \rangle) + \beta_{ijb0} e^{-\beta_{ijb0}t} I_{\langle i, j, -1 \rangle}(\langle i, j, b \rangle), \end{aligned}$$

donde para cada $b \in \{-1, -2, 1, 2\}$

$$I_{\langle i, j, b \rangle}(k) = \begin{cases} 1 & \langle i, j, b \rangle = k, \\ 0 & \langle i, j, b \rangle \neq k. \end{cases}$$

▪ **Funciones de recompensa.**

Supóngase que cada arribo de un cliente tipo 1 retribuye una ganancia de K_1 , y cada arribo de un cliente tipo 2 retribuye una ganancia de K_2 . Además los costos de permanencia en el sistema debido al uso de los servidores, son $c_1(i)$ cuando hay i clientes tipo 1 en el sistema, y $c_2(j)$ cuando hay j clientes tipo 2 en el sistema. De tal manera se definen las funciones de costo para los diferentes casos:

$$\begin{aligned} d(\langle i, j, 1 \rangle, 1) &= c_1(i+1) + c_2(j), \\ d(\langle i, j, 2 \rangle, 1) &= c_1(i) + c_2(j+1), \\ d(\langle i, j, b \rangle, 0) &= c_1(i) + c_2(j), \end{aligned}$$

para $b \in \{-1, -2, 1, 2\}$.

De acuerdo a (1.6), la recompensa esperada descontada $r(\langle i, j, b \rangle, a)$ entre épocas de decisión, dado el sistema en el estado $\langle i, j, b \rangle$ y la acción a , deberá ser escrita como:

$$\begin{aligned} r(\langle i, j, 1 \rangle, 1) &= K_1 - [c_1(i+1) + c_2(j)] \int_0^\infty \left[\int_0^t e^{-\alpha s} ds \right] H(dt | \langle i, j, 1 \rangle, 1) \\ &= K_1 - \frac{[c_1(i+1) + c_2(j)]}{\alpha} \int_0^\infty (1 - e^{-\alpha t}) \beta_{ij11} e^{-\alpha \beta_{ij11} t} dt \\ &= K_1 - \frac{c_1(i+1) + c_2(j)}{\alpha + \beta_{ij11}}, \end{aligned}$$

de manera similar se obtiene los demás casos,

$$\begin{aligned} r(\langle i, j, 2 \rangle, 1) &= K_2 - \frac{c_1(i) + c_2(j+1)}{\alpha + \beta_{ij21}}, \\ r(\langle i, j, -1 \rangle, 0) &= -\frac{c_1(i) + c_2(j)}{\alpha + \beta_{ijb0}}, \end{aligned}$$

para $b \in \{-2, -1, 1, 2\}$.

En esta parte del trabajo, se analizará el ejemplo planteado, usando directamente los resultados de los PCSMs, y analizando mediante la técnica de uniformización, es decir, transformando el modelo a un modelo equivalente.

3.2.2. Modelo analizado directamente

Se considera el MCSM

$$(S, \mathbb{A}, \{\mathbb{A}(s) : s \in S\}, Q, H, D, d),$$

usando la descripción del modelo anterior, se tiene que la ecuación de optimalidad tiene la siguiente forma:

$$V(\langle i, j, b \rangle) = \max_{a \in \mathbb{A}(x)} \left\{ r(\langle i, j, b \rangle, a) + \Delta_\alpha(\langle i, j, b \rangle, a) \sum_k Q(k | \langle i, j, b \rangle) V(k) \right\},$$

donde $\Delta_\alpha(\langle i, j, b \rangle, a)$ está definida en (1.12), así que para los diferentes casos se tiene que,

$$\Delta_\alpha(\langle i, j, 1 \rangle, 1) = \frac{\beta_{ij11}}{\alpha + \beta_{ij11}},$$

$$\Delta_\alpha(\langle i, j, 2 \rangle, 1) = \frac{\beta_{ij21}}{\alpha + \beta_{ij21}},$$

y

$$\Delta_\alpha(\langle i, j, b \rangle, 0) = \frac{\beta_{ijb0}}{\alpha + \beta_{ijb0}}.$$

Por lo que las funciones de valor para cada caso son las siguientes:

$$V(\langle i, j, -1 \rangle) = -\frac{c_1(i) + c_2(j)}{\alpha + \beta_{ijb0}} + \frac{1}{\alpha + \beta_{ijb0}} \left[\begin{array}{l} i\mu_1 V(\langle i-1, j, -1 \rangle) + j\mu_2 V(\langle i, j-1, -2 \rangle) \\ + \lambda_1 V(\langle i, j, 1 \rangle) + \lambda_2 V(\langle i, j, 2 \rangle) \end{array} \right].$$

y $V(\langle i, j, -1 \rangle) = V(\langle i, j, -2 \rangle)$.

$$V(\langle i, j, 1 \rangle) = \max \left\{ \begin{array}{l} K_1 - \frac{c_1(i+1) + c_2(j)}{\alpha + \beta_{ij11}} \\ + \frac{1}{\alpha + \beta_{ij11}} \left[\begin{array}{l} (i+1)\mu_1 V(\langle i+1, j, -1 \rangle) \\ + j\mu_2 V(\langle i+1, j-1, -2 \rangle) \\ + \lambda_1 V(\langle i+1, j, 1 \rangle) \\ + \lambda_2 V(\langle i+1, j, 2 \rangle) \\ V(\langle i, j, -1 \rangle) \end{array} \right] \end{array} \right\},$$

$$V(\langle i, j, 2 \rangle) = \max \left\{ \begin{array}{c} K_2 - \frac{c_1(i) + c_2(j+1)}{\alpha + \beta_{ij21}} \\ + \frac{1}{\alpha + \beta_{ij21}} \left[\begin{array}{c} i\mu_1 V(\langle i-1, j+1, -1 \rangle) \\ + (j+1)\mu_2 V(\langle i, j, -2 \rangle) \\ + \lambda_1 V(\langle i, j+1, 1 \rangle) \\ + \lambda_2 V(\langle i, j+1, 2 \rangle) \\ V(\langle i, j, -1 \rangle) \end{array} \right] \end{array} \right\}.$$

Por lo que se tiene que

$$\Delta_{ij1} = + \frac{1}{\alpha + \beta_{ij11}} \left[\begin{array}{c} K_1 - \frac{c_1(i+1) + c_2(j)}{\alpha + \beta_{ij11}} \\ (i+1)\mu_1 V(\langle i, j, -1 \rangle) + j\mu_2 V(\langle i+1, j-1, -2 \rangle) \\ + \lambda_1 V(\langle i+1, j, 1 \rangle) + \lambda_2 V(\langle i+1, j, 2 \rangle) \\ - V(\langle i, j, -1 \rangle), \end{array} \right]$$

y

$$\Delta_{ij2} = + \frac{1}{\alpha + \beta_{ij21}} \left[\begin{array}{c} K_2 - \frac{c_1(i) + c_2(j+1)}{\alpha + \beta_{ij21}} \\ i\mu_1 V(\langle i-1, j+1, -1 \rangle) + (j+1)\mu_2 V(\langle i, j, -2 \rangle) \\ + \lambda_1 V(\langle i, j+1, 1 \rangle) + \lambda_2 V(\langle i, j+1, 2 \rangle) \\ - V(\langle i, j, -1 \rangle). \end{array} \right]$$

Así que, bajo el Teorema 1.5.5(b) la política estacionaria

$\pi^* = (f^*, f^*, \dots)$ derivado de las reglas de decisión

$$f_1^*(\langle i, j, 1 \rangle) = \begin{cases} 1 & \Delta_{ij1} > -R, \\ 0 & \Delta_{ij1} \leq -R. \end{cases}$$

y

$$f_2^*(\langle i, j, 2 \rangle) = \begin{cases} 1 & \Delta_{ij2} > -R, \\ 0 & \Delta_{ij2} \leq -R. \end{cases}$$

es óptima. Por lo tanto, Δ_{ij1} y Δ_{ij2} son monótonas decrecientes y políticas óptimas.

3.2.3. Modelo mediante uniformización

Se analizará este modelo usando uniformización (véase Capítulo 2).

Sea

$$C = n \max(\mu_1, \mu_2) + \lambda_1 + \lambda_2.$$

Puesto que la constante C es elegida para ser al menos tan grande como la mayor tasa vista en el sistema original. Se necesita ajustar las probabilidades de transición y recompensa bajo el sistema de uniformización como sigue:

$$\tilde{Q}(k|\langle i, j, b \rangle) = \begin{cases} 1 - (1 - q(k|\langle i, j, b \rangle))\beta_{ijba}/C & k = \langle i, j, b \rangle \\ q(k|\langle i, j, b \rangle)\beta_{ijba}/C & k \neq \langle i, j, b \rangle, \end{cases}$$

$$\tilde{r}(\langle i, j, b \rangle, a) = r(\langle i, j, b \rangle, a) \frac{(\alpha + \beta_{ijba})}{\alpha + C}.$$

La ecuación de optimalidad tiene la siguiente forma:

$$V(\langle i, j, b \rangle) = \max_{a \in \mathbb{A}(x)} \left\{ \tilde{r}(\langle i, j, b \rangle, a) + \frac{C}{\alpha + C} \sum_k \tilde{q}(k|\langle i, j, b \rangle) V(k) \right\}.$$

También se puede escribir estas funciones de valor en forma expandida como:

$$V(\langle i, j, 1 \rangle) = \max \left\{ \frac{1}{\alpha + C} \left[\begin{aligned} & \left(K_1 - \frac{c_1(i+1) + c_2(j)}{\alpha + \mu_1} \right) \left(\frac{\alpha + \beta_{ij11}}{\alpha + C} \right) + \\ & (i+1)\mu_1 V(\langle i, j, -1 \rangle) + \\ & j\mu_2 V(\langle i+1, j-1, -2 \rangle) + \\ & \lambda_1 V(\langle i+1, j, 1 \rangle) + \\ & \lambda_2 V(\langle i+1, j, 2 \rangle) + \\ & (C - \beta_{ij11}) V(\langle i, j, 1 \rangle) \\ & V(\langle i, j, -1 \rangle), \end{aligned} \right] \right\}, \quad (3.9)$$

$$V(\langle i, j, 2 \rangle) = \max \left\{ \frac{1}{\alpha + C} \left[\begin{aligned} & \left(K_2 - \frac{c_1(i) + c_2(j+1)}{\alpha + \mu_2} \right) \left(\frac{\alpha + \beta_{ij21}}{\alpha + C} \right) + \\ & (i\mu_1 V(\langle i-1, j+1, -1 \rangle) + \\ & (j+1)\mu_2 V(\langle i, j, -2 \rangle) + \\ & \lambda_1 V(\langle i, j+1, 1 \rangle) + \\ & \lambda_2 V(\langle i, j+1, 2 \rangle) \\ & + (C - \beta_{ij21}) V(\langle i, j, 2 \rangle) \\ & V(\langle i, j, -1 \rangle), \end{aligned} \right] \right\}, \quad (3.10)$$

$$\begin{aligned} V(\langle i, j, -1 \rangle) &= \left(-\frac{(\alpha - 1)(c_1(i) + c_2(j))}{\alpha^2} \right) \left(\frac{\alpha + \beta_{ijb0}}{\alpha + C} \right) \\ &+ \frac{1}{\alpha + C} i\mu_1 V(\langle i-1, j, -1 \rangle) \\ &+ j\mu_2 V(\langle i, j-1, -2 \rangle) + \lambda_1 V(\langle i, j, 1 \rangle) + \lambda_2 V(\langle i, j, 2 \rangle), \end{aligned} \quad (3.11)$$

y $V(\langle i, j, -1 \rangle) = V(\langle i, j, -2 \rangle)$. Simplificando las ecuaciones (3.9) y (3.10) se tiene que

$$V(\langle i, j, 1 \rangle) = \max \left\{ \begin{array}{l} K_1 + V(\langle i + 1, j, -1 \rangle), \\ V(\langle i, j, -1 \rangle). \end{array} \right.$$

$$V(\langle i, j, 2 \rangle) = \max \left\{ \begin{array}{l} K_2 + V(\langle i, j + 1, -1 \rangle), \\ V(\langle i, j, -1 \rangle). \end{array} \right.$$

Se sigue del Teorema 1.5.5(b) que la política estacionaria $\pi^* = (f^*, f^*, \dots)$ derivado de las reglas de decisión

$$f_1^*(\langle i, j, 1 \rangle) = \begin{cases} 1 & v_\alpha^*(\langle i + 1, j, 0 \rangle) - v_\alpha^*(\langle i, j, 0 \rangle) > -R, \\ 0 & v_\alpha^*(\langle i + 1, j, 0 \rangle) - v_\alpha^*(\langle i, j, 0 \rangle) \leq -R, \end{cases}$$

$$f_2^*(\langle i, j, 2 \rangle) = \begin{cases} 1 & v_\alpha^*(\langle i, j + 1, 0 \rangle) - v_\alpha^*(\langle i, j, 0 \rangle) > -R, \\ 0 & v_\alpha^*(\langle i, j + 1, 0 \rangle) - v_\alpha^*(\langle i, j, 0 \rangle) \leq -R, \end{cases}$$

es óptima. Por lo tanto, $\Delta_{ij1} = v_\alpha^*(\langle i + 1, j, 0 \rangle) - v_\alpha^*(\langle i, j, 0 \rangle)$,

$\Delta_{ij2} = v_\alpha^*(\langle i, j + 1, 0 \rangle) - v_\alpha^*(\langle i, j, 0 \rangle)$ son monótonas decrecientes, y son políticas óptimas.

3.2.4. Ejemplos numéricos

A continuación se presenta un ejemplo numérico.

Ejemplo 3.2.1 *Se supone que las sumas globales de ganancias son $K_1 = 20$ y $K_2 = 10$ para los clientes tipo 1 y tipo 2 respectivamente. Sean las tasas de los costos $c_1(i) = 1 \cdot 3^i$ y $c_2(j) = 1 \cdot 8^j$. Sea $n = 9$ el número máximo de servidores. Supóngase también que la tasa de arribos para los clientes tipo 1 y tipo 2 son $\lambda_1 = 0.05$ y $\lambda_2 = 0.2$ respectivamente y sus tasas de servicio son $\mu_1 = 0.5$ y $\mu_2 = 0.25$ respectivamente. Además, que la tasa descontada en tiempo discreto es $\lambda = 0.9$, entonces la tasa descontada en tiempo continuo es $\alpha = \ln \lambda = 0.105$.*

En la Tabla 2 se da el caso sin uniformización y con los datos dados, se muestra las diferencias Δ_{ij1} y Δ_{ij2} , para diferentes valores de i y j .

j	$i=0, b=1$	$i=0, b=2$
0	19.4308	6.3035
1	19.4275	5.745
2	19.4261	3.9105
3	19.4252	1.4045
4	19.4247	-2.5455
5	19.4243	-9.1927
6	19.4240	-20.1881
7	19.4236	-38.3642
8	19.4234	-52.5188
9	19.4233	-68.0543

Tabla 2. Directamente

En Tabla 3 se da el caso uniformizado, donde la constante de uniformización es $C = n \max(\mu_1, \mu_2) + \lambda_1 + \lambda_2 = 4.75$ y con los mismos datos, se obtiene

j	$i=0, b=1$	$i=0, b=2$
0	19.4973	7.0372
1	19.4973	5.7781
2	19.4973	3.9105
3	19.4973	1.3926
4	19.4973	-2.5502
5	19.4973	-9.1927
6	19.4973	-20.1881
7	19.4973	-38.3642
8	19.4973	-51.5183
9	19.4303	-62.1543

Tabla 3. Uniformizado

De acuerdo a la regla de decisión f_1^* y f_2^* , para ambos casos, se puede observar en la Tabla 1 y 2, cuando $i = 0$ y $b = 1$ siempre serán admitidos un número n de clientes en el sistema, para el caso $i = 0, b = 2$ no se admiten más clientes en el sistema cuando hay alrededor de 5 clientes tipo 2.

Conclusiones

En la presente tesis se centró en el estudio de la teoría de los (PCSMs).

En el Capítulo 1, se presentan los resultados principales sobre los PCSMs, en éste se define el modelo de control y se da una interpretación de él. Se define una política, un costo asociado al problema el cual pueden tomar valores positivos y negativos, y se menciona el criterio de rendimiento α -descontado con horizonte infinito, el cual es necesario para evaluar la calidad de las políticas de control.

En el Capítulo 2, se presenta la técnica de uniformización, el cual consiste en transformar un modelo de tiempo continuo con tasas de transición exponenciales a un modelo con estados independientes y tasas de transición constantes. Dicha técnica fue necesaria ya que, parte del trabajo de tesis consistió en hacer un análisis en un problema de control, mediante dos formas: primero utilizando los resultados directos de los PCSMs y segundo usando la técnica de uniformización, esto se analizó en el Capítulo 3 en el que se plantean ejemplos en líneas de espera.

En el Capítulo 3, se planteó un problema de control en líneas de espera M/M/1, en la forma de un MCSM, modelo planteado por Puterman [15], en el cual se analiza usando los resultados del Capítulo 2. En este mismo capítulo se presenta una extensión del problema de control en líneas de espera M/M/1 a un problema de múltiples servidores con 2 clases de clientes. Este ejemplo es planteado como un MCSM, y que es analizado mediante el uso directo de los PCSMs y usando la técnica de uniformización. Estos resultados, también fueron analizados en algunos ejemplos numéricos, los cuales se elaboraron algoritmos computacionales realizados en MATLAB, estos resultados arrojaron que el uso directo como el uso de dicha técnica, son equivalentes.

En resumen las aportaciones de la tesis son las siguientes:

1. Se presentó un estudio del MCSM mediante programación dinámica. A

diferencia de la literatura existente se presenta en términos de recompensas y permite el cambio de signo.

2. Se considera un ejemplo de líneas de espera M/M/n controlado, para el cual se caracteriza una estrategia óptima de operación.
3. Para los modelos dinámicos estudiados se proveen algoritmos numéricos elaborados en MATLAB.

Para trabajos futuros se pretende continuar el estudio de los PCSMs, ahora considerando el caso en el que la distribución del tiempo de permanencia es desconocida, esto nos llevaría a la implementación de Procesos de Control Adaptado.

Apéndice A

Otros resultados

A.1. Funciones

Definición A.1.1 Sea \mathbb{X} un espacio topológico y v una función de \mathbb{X} en $\mathbb{R} \cup \{\infty\}$. Se dice que la función v es:

a) *Semi-continua inferiormente (l.s.c.)* si el conjunto

$$\{x \in \mathbb{X} : v(x) \leq r\},$$

es cerrado en \mathbb{X} para cada $r \in \mathbb{R}$.

b) *Semi-continua superiormente (u.s.c.)* si el conjunto

$$\{x \in \mathbb{X} : v(x) \geq r\},$$

es cerrado en \mathbb{X} para cada $r \in \mathbb{R}$.

Sea \mathbb{X} y \mathbb{Y} espacios de Borel.

Definición A.1.2 Un kernel estocástico o probabilidad de transición sobre \mathbb{X} dado \mathbb{Y} se define como una función $Q(dx | y)$ tal que:

a) $Q(\cdot | y)$ es una medida de probabilidad sobre \mathbb{X} para cada $y \in \mathbb{Y}$.

b) $Q(B | y)$ es una función medible sobre \mathbb{Y} para cada $B \in \mathcal{B}(\mathbb{X})$.

Definición A.1.3 Sea $Q(dx | y)$ un kernel estocástico sobre \mathbb{X} dado \mathbb{Y} . Se dice que:

Q es fuertemente continua si la función:

$$y \mapsto \int_{\mathbb{X}} v(x)Q(dx | y) \quad (\text{A.1})$$

es continua y acotada en $y \in \mathbb{Y}$, para cada función medible y acotada v sobre \mathbb{X} .

A.2. Teorema de Ionescu-Tulcea y Lema de Fatou

Teorema A.2.1 Sea $\mathbb{X}_0, \mathbb{X}_1, \dots$ 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 ν una medida de probabilidad arbitraria en \mathbb{X}_0 y, para cada $n \in \mathbb{N}$, sea $Q_n(d_{x_{n+1}} | y_n)$ un kernel estocástico sobre \mathbb{X}_{n+1} dado \mathbb{Y}_n . Entonces existe una única medida de probabilidad Q_ν sobre \mathbb{Y} tal que, para cada rectángulo medible $B_0 \times \cdots \times B_n$ en \mathbb{Y}_n ,

$$\begin{aligned} Q_\nu(B_0 \times \cdots \times B_n) &= \int_{B_0} \nu(dx_0) \int_{B_1} Q_0(d_{x_1} | x_0) \int_{B_2} Q_1(d_{x_2} | x_0, x_1) \\ &\quad \cdots \int_{B_n} Q_{n-1}(d_{x_n} | x_0, \dots, x_{n-1}). \end{aligned}$$

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

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

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

A.3. Operador Contracción

Definición A.3.1 Sea (S, d) un espacio métrico completo. Un operador $T : S \rightarrow S$ es llamado una contracción si existe un número $0 \leq \gamma < 1$ tal que

$$d(Ts_1, Ts_2) \leq \gamma d(s_1, s_2), \quad \forall s_1, s_2 \in S$$

donde γ es llamado el módulo de T .

Proposición A.3.2 (Teorema del Punto Fijo de Banach.) Un operador contracción T sobre un espacio métrico completo (S, d) tiene un único punto fijo s^* y además

$$d(T^n s, s^*) \leq \gamma^n d(s, s^*) \quad \forall s \in S, \quad n \in \mathbb{N}.$$

Teorema A.3.3 Sea T un operador monótono de $\mathbb{B}_w(\mathbb{X})$ a si mismo. Si existe un número positivo $\gamma < 1$ tal que

$$T(u + \gamma v) \leq Tu + \gamma v, \quad \forall u \in \mathbb{B}_w(\mathbb{X}), \quad v \in \mathbb{R}, \quad (\text{A.2})$$

entonces T es una contracción con módulo γ .

Apéndice B

Códigos

En esta sección se presenta el desarrollo de los códigos usados en este trabajo. Los siguientes programas fueron compilados en MATLAB 7.8 (R2009a)

B.1. Línea de espera con control M/M/1 con uniformización

MM1.m

```
%%Par\U{e1}metros conocidos
m=0.5; %valor de mu
l=0.2; %valor de lamda
a=0.105; %valor de alfa
R=20; %valor del beneficio fijo
j=40; %n\U{fa}mero de iteraciones
k=20; %n\U{fa}mero de clientes

%Inicializaci\U{f3}n de matrices
c=m+l+a;
V=zeros(j,k);
U=zeros(j,k);
D=zeros(j,k);
V;
U;
```

```

for n=1:j %iteraciones
for s=1:k-1 %n\U{fa}mero de clientes en el sistema
if s==1
U(n+1,s)=(-1)*(((1.3)^s)/c)+(1/c)*(l*V(n,s));
else
U(n+1,s)=(-1)*(((1.3)^s)/c)+(1/c)*(m*U(n,s-1)+l*V(n,s));
end
if s==1
V(n+1,s)=max(R-(((1.3)^(s+1))/c)+(1/c)*(m*U(n,s)+l*V(n,s+1)),
(-1)*(((1.3)^s)/c)+(1/c)*(l*V(n,s)));
else
V(n+1,s)=max(R-(((1.3)^(s+1))/c)+(1/c)*(m*U(n,s)+l*V(n,s+1)),
(-1)*(((1.3)^s)/c)+(1/c)*(m*U(n,s-1)+l*V(n,s)));
end
end
end
U;
V;
for n=1:j
for s=1:k-1
D(n,s)=U(n,s+1)-U(n,s);
end
end
D
D1=zeros(1,k-2);
for s=1:k-2
D1(s)=D(j,s);
end
D1
plot(D1,'b');
title('');
xlabel('clientes en el sistema, s');
ylabel('D(s)');

```

B.2. Línea de espera con control M/M/n sin uniformización

```

MMS(SINUNIF).m

%%Par\U{e1}metros conocidos
K1=20; K2=10; M=9; l1=0.05; l2=0.2; m1=0.5; m2=0.25;
a=0.105; C=4.75; n_iter=100; n_cp=11; n_co=20;

V=zeros(n_cp,n_co,n_iter); %V=v(<s,j,1>)
U=zeros(n_cp,n_co,n_iter); %U=v(<s,j,2>)
W=zeros(n_cp,n_co,n_iter); %W=v(<s,j,b>) b={-1,-2}

for n=2:n_iter
    for s=1:n_cp-1
        for j=1:n_co-1
            if s==1 && j==1 %(s=0, j=0)
                W(s,j,n)=(-1)*((1.3)^(s-1)+(1.8)^(j-1))/(a+((s-1)*m1
                    +(j-1)*m2+l1+l2))+1/(a+((s-1)*m1
                    +(j-1)*m2+l1+l2))*((l1*V(s,j,n-1)
                    +l2*U(s,j,n-1)));
            end
            if s==1 && j>1 %(s=0, j>0)
                W(s,j,n)=(-1)*((1.3)^(s-1)+(1.8)^(j-1))/(a+((s-1)*m1
                    +(j-1)*m2+l1+l2))+1/(a+((s-1)*m1
                    +(j-1)*m2+l1+l2))*((j-1)*m2*W(s,j-1,n-1)
                    +l1*V(s,j,n-1)+l2*U(s,j,n-1));
            end
            if s>1 && j==1 %(s>0, j=0)
                W(s,j,n)=(-1)*((1.3)^(s-1)+(1.8)^(j-1))/(a+((s-1)*m1
                    +(j-1)*m2+l1+l2))+1/(a+((s-1)*m1
                    +(j-1)*m2+l1+l2))*((s-1)*m1*W(s-1,j,n-1)
                    +l1*V(s,j,n-1)+l2*U(s,j,n-1));
            end
            if s>1 && j>1 %(s>0, j>0)
                W(s,j,n)=(-1)*((1.3)^(s-1)+(1.8)^(j-1))/(a+((s-1)*m1
                    +(j-1)*m2+l1+l2))+1/(a+((s-1)*m1
                    +(j-1)*m2+l1+l2))*((s-1)*m1*W(s-1,j,n-1)

```

```

        +(j-1)*m2*W(s,j-1,n-1)+l1*V(s,j,n-1)
        +l2*U(s,j,n-1));
end

if s==1 && j==1 %(s=0, j=0)
    V(s,j,n)=max(K1-((1.3)^(s-1+1)+(1.8)^(j-1))/(a
        +((s-1+1)*m1+(j-1)*m2+l1+l2))
        +(1/(a+((s-1+1)*m1+(j-1)*m2+l1+l2)))*((s-1+1)
        *m1*W(s,j,n-1)+l1*V(s+1,j,n-1)+l2*U(s+1,j,n-1)),
        W(s,j,n-1));
end

if s==1 && j>1 %(s=0, j>0)
    V(s,j,n)=max(K1-((1.3)^(s-1+1)+(1.8)^(j-1))/(a
        +((s-1+1)*m1+(j-1)*m2+l1+l2))+(1/(a+((s-1+1)
        *m1+(j-1)*m2+l1+l2)))*((s-1+1)*m1*W(s,j,n-1)
        +(j-1)*m2*W(s,j-1,n-1)+l1*V(s,j,n-1)+l2
        *U(s,j,n-1)),W(s,j,n-1));
end

if s>1 && j==1 %(s>0, j=0)
    V(s,j,n)=max(K1-((1.3)^(s-1+1)+(1.8)^(j-1))/(a
        +((s-1+1)*m1+(j-1)*m2+l1+l2))+(1/(a+((s-1+1)
        *m1+(j-1)*m2+l1+l2)))*((s-1+1)*m1*W(s-1,j,n-1)
        +l1*V(s+1,j,n-1)+l2*U(s+1,j,n-1)),W(s,j,n-1));
end

if s>1 && j>1 %(s>0, j>0)
    V(s,j,n)=max(K1-((1.3)^(s-1+1)+(1.8)^(j-1))/(a
        +((s-1+1)*m1+(j-1)*m2+l1+l2))+(1/(a+((s-1+1)
        *m1+(j-1)*m2+l1+l2)))*((s-1+1)*m1*W(s-1,j,n-1)
        +(j-1)*m2*W(s,j-1,n-1)+l1*V(s+1,j,n-1)
        +l2*U(s+1,j,n-1)),W(s,j,n-1));
end

if s==1 && j==1 %(s=0, j=0)
    U(s,j,n)=max(K2-((1.3)^(s-1)+(1.8)^(j-1+1))/(a
        +((s-1)*m1+(j-1+1)*m2+l1+l2))+(1/(a+((s-1)
        *m1+(j-1+1)*m2+l1+l2)))*((j-1+1)*m2*W(s,j,n-1)
        +l1*V(s,j+1,n-1)+l2*U(s,j+1,n-1)),W(s,j,n-1));
end

```

```

if s==1 && j>1 %(s=0, j>0)
    U(s,j,n)=max(K2-((1.3)^(s-1)+(1.8)^(j-1+1))/(a+((s-1)
        *m1+(j-1+1)*m2+l1+l2))+1/(a+((s-1)*m1+(j-1+1)
        *m2+l1+l2)))*((j-1+1)*m2*W(s,j,n-1)
        +l1*V(s,j+1,n-1)+l2*U(s,j+1,n-1)),W(s,j,n-1));
end
if s>1 && j==1 %(s>0, j=0)
    U(s,j,n)=max(K2-((1.3)^(s-1)+(1.8)^(j-1+1))/(a
        +((s-1)*m1+(j-1+1)*m2+l1+l2))+1/(a+((s-1)*m1
        +(j-1+1)*m2+l1+l2)))*((s-1)*m1*W(s-1,j+1,n-1)
        +(j-1+1)*m2*W(s,j,n-1)+l1*V(s,j+1,n-1)
        +l2*U(s,j+1,n-1)),W(s,j,n-1));
end
if s>1 && j>1 %(s>0, j>0)
    U(s,j,n)=max(K2-((1.3)^(s-1)+(1.8)^(j-1+1))/(a+((s-1)*m1
        +(j-1+1)*m2+l1+l2))+1/(a+((s-1)*m1
        +(j-1+1)*m2+l1+l2)))*((s-1)*m1*W(s-1,j+1,n-1)
        +(j-1+1)*m2*W(s,j,n-1)+l1*V(s,j+1,n-1)
        +l2*U(s,j+1,n-1)),W(s,j,n-1));
end
end
end
end
U;
V;
W;
D1=zeros(n_cp,n_co,n_iter);
D2=zeros(n_cp,n_co,n_iter);
for n=2:n_iter
    for s=1:n_cp-1
        for j=1:n_co-1
            %D1(s,j,n)=K1+W(s+1,j,n)-W(s,j,n); %arribo tipo1
            if s==1 && j==1 %(s=0, j=0)
                D1(s,j,n)=K1-((1.3)^(s-1+1)+(1.8)^(j-1))/(a+((s-1+1)*m1
                    +(j-1)*m2+l1+l2))+1/(a+((s-1+1)*m1
                    +(j-1)*m2+l1+l2)))*(m1*W(s,j,n-1)
                    +l1*V(s+1,j,n-1)+l2*U(s+1,j,n-1))-W(s,j,n-1);
            end
        end
    end
end

```

```

if s==1 && j>1 %(s=0, j>0)
    D1(s,j,n)=K1-((1.3)^(s-1+1)+(1.8)^(j-1))/(a+((s-1+1)*m1
        +(j-1)*m2+l1+l2))+1/(a+((s-1+1)*m1
        +(j-1)*m2+l1+l2))*((j-1)*m2*W(s+1,j-1,n-1)
        +l1*V(s+1,j,n-1)+l2*U(s+1,j,n-1))-W(s,j,n-1);
end
if s>1 && j==1 %(s>0, j=0)
    D1(s,j,n)=K1-((1.3)^(s-1+1)+(1.8)^(j-1))/(a+((s-1+1)*m1
        +(j-1)*m2+l1+l2))+1/(a+((s-1+1)*m1
        +(j-1)*m2+l1+l2))*((s-1+1)*m1*W(s,j,n-1)
        +l1*V(s+1,j,n-1)+l2*U(s+1,j,n-1))-W(s,j,n-1);
end
if s>1 && j>1 %(s>0, j>0)
    D1(s,j,n)=K1-((1.3)^(s-1+1)+(1.8)^(j-1))/(a+((s-1+1)*m1
        +(j-1)*m2+l1+l2))+1/(a+((s-1+1)*m1
        +(j-1)*m2+l1+l2))*((s-1+1)*m1*W(s,j,n-1)
        +(j-1)*m2*W(s+1,j-1,n-1)+l1*V(s+1,j,n-1)
        +l2*U(s+1,j,n-1))-W(s,j,n-1);
end

%D2(s,j,n)=K2+W(s,j+1,n)-W(s,j,n); %arriba tipo2
if s==1 && j==1 %(s=0, j=0)
    D2(s,j,n)=K2-((1.3)^(s-1)+(1.8)^(j-1+1))/(a+((s-1)*m1
        +(j-1+1)*m2+l1+l2))+1/(a+((s-1)*m1
        +(j-1+1)*m2+l1+l2))*((l1*V(s,j+1,n-1)
        +l2*U(s,j+1,n-1))-W(s,j,n-1);
end
if s==1 && j>1 %(s=0, j>0)
    D2(s,j,n)=K2-((1.3)^(s-1)+(1.8)^(j-1+1))/(a+((s-1)*m1
        +(j-1+1)*m2+l1+l2))+1/(a+((s-1)*m1
        +(j-1+1)*m2+l1+l2))*((j-1+1)*m2*W(s,j,n-1)
        +l1*V(s,j+1,n-1)+l2*U(s,j+1,n-1))-W(s,j,n-1);
end
if s>1 && j==1 %(s>0, j=0)
    D2(s,j,n)=K2-((1.3)^(s-1)+(1.8)^(j-1+1))/(a+((s-1)*m1
        +(j-1+1)*m2+l1+l2))+1/(a+((s-1)*m1
        +(j-1+1)*m2+l1+l2))*((s-1)*m1*W(s-1,j+1,n-1)
        +l1*V(s,j+1,n-1)+l2*U(s,j+1,n-1))-W(s,j,n-1);
end

```

```

end
if s>1 && j>1 %(s>0, j>0)
    D2(s,j,n)=K2-(((1.3)^(s-1)+(1.8)^(j-1+1)))/(a+((s-1)*m1
        +(j-1+1)*m2+l1+l2))+1/(a+((s-1)*m1
        +(j-1+1)*m2+l1+l2))*((s-1)*m1*W(s-1,j+1,n-1)
        +(j-1+1)*m2*W(s,j,n-1)+l1*V(s,j+1,n-1)
        +l2*U(s,j+1,n-1))-W(s,j,n-1);
end
end
end
end
D1;
D2;
d1=zeros(1,n_co,n_iter);
d2=zeros(1,n_co,n_iter);
for n=1:n_iter
    for j=1:n_co
        d1(1,j,n)=D1(2,j,n);
        d2(1,j,n)=D2(2,j,n);
    end
end
d1;
d2;
d=zeros(1,n_co-3);
D=zeros(1,n_co-3);
for j=1:n_co-3
    d(j)=d1(1,j,n_iter);
    D(j)=d2(1,j,n_iter);
end
d
D
V(:,:,n_iter)
U(:,:,n_iter)
W(:,:,n_iter)
plot(d)
hold on;
plot(D,'r')
hold off;

```

B.3. Línea de espera con control M/M/n con uniformización

MMS(UNIF).m

```

%%Par\U{e1}metros conocidos
K1=20; K2=10; M=9; l1=0.05; l2=0.2; m1=0.5; m2=0.25;
a=0.105; C=4.75; n_iter=100; n_cp=11; n_co=20;

V=zeros(n_cp,n_co,n_iter); %V=v(<s,j,1>)
U=zeros(n_cp,n_co,n_iter); %V=v(<s,j,2>)
W=zeros(n_cp,n_co,n_iter); %V=v(<s,j,b>) b={-1,-2}

for n=2:n_iter
  for s=1:n_cp-1
    for j=1:n_co-1
      if s==1 && j==1
        W(s,j,n)=(-1)*((1.3)^(s-1)+(1.8)^(j-1))/(a+((s-1)*m1
          +(j-1)*m2+l1+l2))+1/(a+((s-1)*m1
          +(j-1)*m2+l1+l2))*((l1*V(s,j,n-1)
          +l2*U(s,j,n-1)));
      end
      if s==1 && j>1
        W(s,j,n)=(-1)*((1.3)^(s-1)+(1.8)^(j-1))/(a+((s-1)*m1
          +(j-1)*m2+l1+l2))+1/(a+((s-1)*m1
          +(j-1)*m2+l1+l2))*((j-1)*m2*W(s,j-1,n-1)
          +l1*V(s,j,n-1)+l2*U(s,j,n-1));
      end
      if s>1 && j==1
        W(s,j,n)=(-1)*((1.3)^(s-1)+(1.8)^(j-1))/(a+((s-1)*m1
          +(j-1)*m2+l1+l2))+1/(a+((s-1)*m1
          +(j-1)*m2+l1+l2))*((s-1)*m1*W(s-1,j,n-1)
          +l1*V(s,j,n-1)+l2*U(s,j,n-1));
      end
      if s>1 && j>1

```

```

W(s,j,n)=(-1)*((1.3)^(s-1)+(1.8)^(j-1))/(a+((s-1)*m1
      +(j-1)*m2+l1+l2))+1/(a+((s-1)*m1
      +(j-1)*m2+l1+l2))*((s-1)*m1*W(s-1,j,n-1)
      +(j-1)*m2*W(s,j-1,n-1)+l1*V(s,j,n-1)
      +l2*U(s,j,n-1));
end
V(s,j,n)=max(K1+W(s+1,j,n-1),W(s,j,n-1));
U(s,j,n)=max(K2+W(s,j+1,n-1),W(s,j,n-1));
end
end
end
U;
V;
W;
D1=zeros(n_cp,n_co,n_iter);
D2=zeros(n_cp,n_co,n_iter);
for n=1:n_iter
  for s=1:n_cp-1
    for j=1:n_co-1
      D1(s,j,n)=K1+W(s+1,j,n)-W(s,j,n); %arribo tipo1
      D2(s,j,n)=K2+W(s,j+1,n)-W(s,j,n); %arribo tipo2
    end
  end
end
D1; % diferencia DELTA(i,j,1)
D2; % diferencia DELTA(i,j,2)
d1=zeros(1,n_co,n_iter);
d2=zeros(1,n_co,n_iter);
for n=1:n_iter
  for j=1:n_co
    d1(1,j,n)=D1(2,j,n);
    d2(1,j,n)=D2(2,j,n);
  end
end
d1; % valores de DELTA(1,j,n iteraciones)
d2; % valores de DELTA(1,j,n iteraciones)
d=zeros(1,n_co-3);
D=zeros(1,n_co-3);

```

```
for j=1:n_co-3
    d(j)=d1(1,j,n_iter);
    D(j)=d2(1,j,n_iter);
end
d
D
V(:,:,n_iter)
U(:,:,n_iter)
W(:,:,n_iter)
plot(d)
hold on;
plot(D,'r')
hold off;
```

Bibliografía

- [1] Alagoz, O. Ayvaci, M. Uniformization in Markov Decision Processes. Wiley Encyclopedia of Operations Research and Management Science, 2011.
- [2] Baykal-G, M. Semi-Markov Decision Processes. Wiley Encyclopedia of Operations Research and Management Science, 2010.
- [3] Bertsekas, D. P. & Shreve, S. E, Stochastic Optimal Control: The Discrete Time Case, Academic Press, New York, 197
- [4] Beutler , F. J. Ross, K. W. Uniformization for Semi-Markov Decision Processes Under Stationary Policies. 1987.
- [5] Cinlar, C. W. Introduction to Stochastic Processes. Prentice-Hall, 1975.
- [6] Díaz, R. Pazos Arias J. J. Suárez González, A. Teoría de Colas y Simulación de Eventos Discretos. Person Publications Company, 2004.
- [7] Dshalalow, J. H. Advances in Queueing: Theory, Methods, and Open Problems. CRC-Press, 1995.
- [8] Howard, R. A. Semi-Markovian decision Processes. Proc. Intern, Stat. Inst,1963.
- [9] Hernández-Lerma, O. & Lasserre J. B., Discrete Time Markov Control Processes. Springer Verlag, 1996.
- [10] Hernández-Lerma, O. & Lasserre J. B., Further Topics on Discrete-Time Markov Control Processes. Springer Verlag, 1999.
- [11] Kalashnikov, V. Mathematical Methods in Queueing Theory. Springer, 2010.

-
- [12] Linn I. S. Stochastic Dynamic Programming and the Control of Queuing Systems, John Wiley and Sons, 1999.
 - [13] Steven A. L. Applying a New Device in the Optimization of Exponential Queuing Systems. Operations Research, 1973.
 - [14] Luque, V. F. Semi-Markov control process with unknown holding times distribution under a discounted criterion. Springer-Verlag, 2005.
 - [15] Puterman, M. L. Markov Decision Processes: Discrete Stochastic Dynamic Programming, Wiley, New York, 1994.
 - [16] Rincon, L. Introducción a los Procesos Estocásticos. Departamente de Matemáticas, UNAM 2012.
 - [17] Ross, S. M. Average Cost Semi-Markov Decision Processes. J. Appl, 1970.
 - [18] Shaler S. J. Optimal Control of Queuing Systems. Chapman & Hall/CR, 2011.
 - [19] Serfozo, R. F. An Equivalence Between Continuous and Discrete Time Markov Desicion Processes. INFORMS, 1979.