EL SABER DE MIS HIJOS
HARÁ MI GRANDEZA

UNIVERSIDAD DE SONORA

DIVISIÓN DE CIENCIAS EXACTAS Y NATURALES

DEPARTAMENTO DE MATEMÁTICAS

Programa de Licenciatura en Matemáticas

Estimación empírica en modelos de control Markovianos descontados

T E S I S

Que para obtener el título de:

Licenciado en Matemáticas

Presenta:

Jessica Liliana Leyva Domínguez

Director de tesis:
Dra. Luz del Carmen Rosas Rosas

Hermosillo, Sonora, México, Septiembre de 2013

Universidad de Sonora

Repositorio Institucional UNISON

"El saber de mis hijos hará mi grandeza"

Excepto si se señala otra cosa, la licencia del ítem se describe como openAccess

QA274.7
.L49

R. T180146

SINODALES

DR. FERNANDO LUQUE VÁSQUEZ
UNIVERSIDAD DE SONORA, HERMOSILLO, MÉXICO

DR. ADOLFO MINJÁREZ SOSA
UNIVERSIDAD DE SONORA, HERMOSILLO, MÉXICO

M.C. CARMEN GERALDI HIGUERA CHAN
UNIVERSIDAD DE SONORA, HERMOSILLO, MÉXICO

DRA. LUZ DEL CARMEN ROSAS ROSAS
UNIVERSIDAD DE SONORA, HERMOSILLO, MÉXICO

UNIVERSIDAD DE SONORA

"El Saber de Mis Hijos Hará Mi Grandeza"

EL SABER DE MIS HIJOS HARÁ MI GRANDEZA

Acta: 45
Foja: 46
Libro: 2
Expediente: 208201573

ACTA DE EXAMEN PROFESIONAL

En la ciudad de Hermosillo, Sonora, siendo las 18:00 horas del día 20 de septiembre de 2013, se reunieron en el Departamento de Matemáticas de la Universidad de Sonora, los integrantes del jurado:

DR. FERNANDO LUQUE VÁSQUEZ
DR. JESÚS ADOLFO MINJÁREZ SOSA
M.C. CARMEN GERALDI HIGUERA CHAN

bajo la presidencia del primero y fungiendo como secretario el último, para realizar el examen profesional a:

JESSICA LILIANA LEYVA DOMINGUEZ

del programa: Licenciatura en Matemáticas

quien de acuerdo a la opción de titulación presentó **tesis profesional** titulada:

"ESTIMACIÓN EMPÍRICA EN MODELOS DE CONTROL MARKOVIANOS"

El jurado, después de debatir entre sí reservada y libremente, emitió el siguiente dictamen:

APROBADO POR UNANIMIDAD

El presidente del jurado dio a conocer al sustentante el resultado de su examen y para constancia se levantó la presente acta.

DR. FERNANDO LUQUE VÁSQUEZ
Presidente

M.C. CARMEN GERALDI HIGUERA CHAN
Secretario

DR. JESÚS ADOLFO MINJÁREZ SOSA
Vocal

DR. JORGE RUPERTO VARGAS CASTRO, jefe del Departamento de Matemáticas de la Universidad de Sonora, hace constar que las firmas que anteceden corresponden al jurado que intervino en el examen profesional.

Hermosillo, Sonora, a 23 de septiembre de 2013

DR. JORGE RUPERTO VARGAS CASTRO
Jefe de Departamento

Nº 3532

Contenido

Dedicatoria

"Toda historia tiene un comienzo y un final, pero en la vida cada final significa un nuevo comienzo."

A mis padres, Amada Alicia Dominguez Gutierrez y Tomas Leyva Chavez.

A mis hermanas, Blanca Irasema y Mercedes Virginia Leyva Dominguez

A toda mi familia.

A mi directora de tesis, Dra. Luz del Carmen Rosas Rosas.

Mil Gracias.

Dedication

Agradecimientos

Me faltan palabras para agradecer a todos aquellos que hicieron posible esto. Primeramente agradezco a Dios por haberme permitido terminar este trabajo y darme fortaleza, pero, sobre todo le doy gracias por darme una familia maravillosa.

A mi madre, Amada Alicia por ser una madre excepcional, por estar conmigo en los momentos más difíciles de mi vida, es mi orgullo, mi razón de existencia y mi fortaleza, también por preocuparse tanto por mí y demostrarme que me quiere mucho, usted siempre me ha apoyado incondicionalmente, especialmente en mis estudios, y ante las adversidades me ha sacado adelante, le agradezco por haberme dado la vida.

A mi padre, Tomas por ser un buen padre, un ejemplo de vida, por estar ahí apoyándome en todo, por darme el valor para salir adelante, por darme esa seguridad que en muchos momentos no la encontraba, por creer en mí, muchas gracias por haberme dado la vida.

A mis dos hermanas.

Blanca Irasema por sus buenos consejos, por ser esa hermana mayor que siempre se ha preocupado por mi desde pequenas, que has estado ahi para no dejarme vencer, y aunque no te lo diga seguido, te quiero mucho hermana.

Mercedes Virginia por motivarme a no parar, aún cuando se presentaron *dificultades me alentabas para continuar en mis estudios, por tus sabios consejos, y en algunas ocaciones por ser mi coniencia, te quiero mucho hermanita.*

A toda mi familia por los buenos deseos brindados durante mi educación.

A mi directora de tesis Dr. Luz del Carmen Rosas Rosas, por su gran ayuda al orientarme y dedicarme de su tiempo para el desarrollo del presente trabajo, así como por confiar y creer en mí, estaré eternamente agradecida. Gracias.

A mi tutor, José Dolores Dávila Galindo, por guiarme durante mi estancia como estudiante de nivel licenciatura en la Unison.

A mis maestros y a todos los profesores del Departamento de Matemáticas de la Universidad de Sonora, gracias por la enorme labor de educación que me brindaron a lo largo de la carrera, todas esas horas de clase donde fui formada como matemática. Especialmente al Dr. Jesús Adolfo Minjárez Sosa, así como al Dr. Fernando Luque Valśquez, por su valiosa colaboración para la realización de este trabajo.

A mis amigos, Guadalupe Morales, Valeria Cienfuegos, Arcelia Moreno, Bogar Murillo, Cecilia Giottonini, Carmen Higuera, Alejandro Duenas, Carmen Romandia y Jesus Ernesto Cruz, gracias por esos momentos de angustia, satisfacción y de fuertes jornadas de estudio, jajajajaja... y sobre todo, por saber que siempre podré contar con su apoyo, fue un honor llevar clases con Ustedes, y aunque no con todos compartí el aula de clases, aún así les agradezco su comprensión, apoyo y entusiasmo en la Unison; no se qué hubiera hecho sin su ayuda.

Al grupo de tesistas del verano de 2013, –gracias a ustedes fue uno de los mejores veranos.

Con carino, Jessica Liliana Leyva

Introducción

La teoría de control óptimo trata con problemas de optimización de sistemas dinámicos cuyo comportamiento puede ser manipulado mediante ciertos controles los cuales se seleccionan por medio de reglas denominadas políticas de control. La calidad de las políticas de control la mide un índice de funcionamiento del sistema el cual representa un costo o una ganancia. De esta forma, el problema de control óptimo consiste en encontrar una política óptima que minimice ó maximice un índice de funcionamiento, según sea el caso.

En el estudio de problemas de control óptimo, los modelos correspondientes se clasifican en: estocásticos ó determinísticos si incluyen ó no componentes aleatorias, respectivamente; asimismo, en tiempo continuo si los controles pueden elegirse en cualquier tiempo, o bien, en tiempo discreto si éstos se seleccionan en un conjunto a lo más numerable (etapas de decisión).

En este trabajo se estudian modelos de control markovianos, los cuales constituyen una clase de modelos de control estocástico en tiempo discreto, y cuya evolución en el tiempo la podemos describir como sigue. Si en la t-ésima etapa de decisión ($t \in \mathbb{N}_0$) el sistema se encuentra en el estado $x_t = x$, entonces el controlador elige una acción o control $a_t = a$ y ocurre lo siguiente: 1) se produce un costo c que depende del estado y la acción elegida; 2) el sistema se mueve a un nuevo estado $x_{t+1} = y$ de acuerdo a una ley de transición. Una vez ocurrido lo anterior, el proceso se repite.

Bajo este escenario los costos de operación se acumulan durante la evolución del sistema y, por lo tanto, el objetivo del controlador consiste en encontrar una política de control que minimice el costo total acumulado, mismo que define el *índice de funcionamiento. En particular, nos enfocaremos en el índice de costo total esperado* α*-descontado.*

Una clase particular de modelos de control markovianos es aquella en la que la dinámica del sistema está modelada por medio de una ecuación en diferencias de la forma

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

donde $\{\xi_t\}$ *es una sucesión de variables aleatorias independientes e idénticamente distribuidas con distribución común* θ*; entonces, la ley de transición de este modelo de control está determinada por la función* F *junto con la distribución* θ*. Bajo este esquema, regularmente se supone que la distribución* θ *es conocida por el controlador, lo cual en algunas situaciones es una hipótesis restrictiva. De modo que, tomando en cuenta este hecho, en este trabajo de tesis consideramos el caso en que* θ *es desconocida, de tal forma que el controlador debe combinar métodos de estimación estadística con técnicas de optimización. En particular, usaremos la distribución empírica para estimar* θ*. A la política que resulta de la combinación estimación y control se le llama* política adaptada*. Entonces, el objetivo del presente trabajo es estudiar la optimalidad de políticas adaptadas bajo el criterio de costo descontado. Sin embargo, como veremos en el Capítulo 3, debido a las características propias del índice descontado, la optimalidad de las políticas adaptadas se estudiará en un sentido asintótico, como se establece en la Definición 3.1, ya que bajo métodos de estimación y control no es posible garantizar la existencia de políticas óptimas.*

El material contenido en esta tesis se encuentra organizado en tres capítulos de la siguiente manera.

En el Capítulo 1 *describimos el modelo de control markoviano en general. Asimismo, planteamos el problema de control óptimo asociado, y además, ilustramos la teoría desarrollada mediante un ejemplo.*

En el Capítulo 2 *analizamos el problema de control óptimo para el caso en que el índice de funcionamiento considerado es el de costo total esperado* α*-descontado. De hecho, imponemos condiciones bajo las cuales demostramos la existencia de una política óptima para el problema de control formulado.*

Mientras que, el Capítulo 3 contiene la parte central de esta tesis, cuyo resultado principal demuestra justamente la existencia de una política asintóticamente óptima descontada para el problema de control óptimo cuando la distribución θ *es desconocida.*

A lo largo de este trabajo asumiremos que $\mathbb{N}$, $\mathbb{N}_0$ *y* $\mathbb{R}$ *denotarán, respectivamente, al conjunto de los números: enteros positivos, enteros no-negativos y reales.*

Capítulo 1

Modelos de Control Markovianos

1.1 Introducción

El propósito de este capítulo se centra en introducir el problema de control óptimo markoviano general, razón por lo cual, a lo largo de sus secciones se describen los tres elementos requeridos para su respectiva formulación, es decir: el modelo de control (markoviano), el conjunto de políticas, así como el índice de funcionamiento (también llamado criterio de optimalidad); asimismo, en su última sección se incluye un ejemplo con el propósito de ilustrar, tanto los elementos antes mencionados como el planteamiento del problema de control óptimo correspondiente.

1.2 Modelo de control markoviano

1.2.1 Descripción

DEFINICIÓN 1.1 Un *modelo de control markoviano* (*MCM*) en tiempo discreto, denotado por

$$(\mathbb{X}, \mathbb{A}, \{A(x) : x \in \mathbb{X}\}, P, c), \tag{1.1}$$

consta de los siguientes elementos:

- $\mathbb{X}$ *representa el* espacio de estado, *y supondremos que es un conjunto numerable.*

- $\mathbb{A}$ *representa el* espacio de control o acción, *y supondremos que es un conjunto numerable.*

- $\{A(x) : x \in \mathbb{X}\}$ *es la familia de* conjuntos de controles (o acciones) admisibles. *Es decir, cada estado $x \in \mathbb{X}$ tiene asociado un conjunto no vacío $A(x) \subset \mathbb{A}$, cuyos elementos son los controles admisibles cuando el sistema se encuentra en el estado x.*

- *P representa la* ley de transición

$$P_{x,y}(a) := P[x_{t+1} = y \mid x_t = x, a_t = a], \tag{1.2}$$

la cual es una distribución de probabilidad en $\mathbb{X}$ para cada $(x,a) \in \mathbb{K}$, donde

$$\mathbb{K} := \{(x,a) : x \in \mathbb{X}, a \in A(x)\} \tag{1.3}$$

es el conjunto de pares estado-acción admisibles.

- $c : \mathbb{K} \to \mathbb{R}$ *representa la función de* costo por etapa.

1.2.2 Interpretación

Un MCM representa un sistema que evoluciona en el tiempo de modo que, en cada etapa de decisión $t \in \mathbb{N}_0$ *el sistema está en el estado $x_t = x \in \mathbb{X}$ y se elige un control $a_t = a \in A(x)$. Entonces:*

1. *se produce un costo $c(x,a)$;*

2. *luego, el sistema evoluciona al estado $x_{t+1} = x' \in \mathbb{X}$ de acuerdo a la ley de transición (1.2);*

3. *y, una vez que el sistema se encuentra en el estado $x_{t+1} = x'$, el proceso se repite.*

Diremos que el MCM (1.1) tiene horizonte de planeación finito *si el número de etapas N es finito, y en otro caso, diremos que el horizonte de planeación respectivo es infinito.*

Observación 1.1 *(a) En algunas aplicaciones la evolución del sistema está determinada por una ecuación en diferencias de la forma*

$$x_{t+1} = F\left(x_t, a_t, \xi_t\right) \tag{1.4}$$

donde $\{\xi_t\}$ *es una sucesión de variables aleatorias (v.a.'s) independientes e idénticamente distribuidas (i.i.d.) con valores en algún conjunto numerable* $\mathbb{S}$*; mientras que,* $F : \mathbb{X} \times \mathbb{A} \times \mathbb{S} \to \mathbb{X}$ *es una función conocida.*

(b) Si θ *es la función de probabilidad común de las v.a.'s* ξ_t*, es decir,*

$$\theta(k) = P[\xi_t = k] \quad \forall\, k \in \mathbb{S},\ t \in \mathbb{N}_0,$$

entonces para cada $(x,a) \in \mathbb{K}$ *tenemos*

$$P\left[x_{t+1} = x' \mid x_t = x, a_t = a\right] = \sum_{k \in S_F} \theta(k), \tag{1.5}$$

donde

$$S_F := \left\{ s \in \mathbb{S} : F(x,a,s) = x' \right\}.$$

De lo anterior, es posible obtener la representación del sistema correspondiente mediante probabilidades de transición usando (1.5).

1.3 Políticas de control admisibles

DEFINICIÓN 1.2 Dado un *MCM* definimos para cada $t \in \mathbb{N}_0$ el *espacio de historias admisibles hasta la etapa* t mediante $\mathbb{H}_0 := \mathbb{X}$ y

$$\mathbb{H}_t := \mathbb{K}^t \times \mathbb{X} \quad \text{para } t \in \mathbb{N}.$$

De modo que, un elemento de $\mathbb{H}_t$ es un vector (o t-historia) de la forma

$$h_t = (x_0, a_0, \ldots, x_{t-1}, a_{t-1}, x_t)$$

con $(x_k, a_k) \in \mathbb{K}$ para $k = 0, 1, \ldots, t-1$ y $x_t \in \mathbb{X}$.

Una regla de decisión es un procedimiento para elegir un control (acción) en una etapa, el cual puede depender, ya sea, de la historia hasta la etapa t, o bien, únicamente del estado del sistema en dicha etapa.

De hecho, una regla de decisión dependiente de la historia, es una función $f_t : \mathbb{H}_t \to \mathbb{A}$ *tal que* $f_t(h_t) \in A(x_t)$. *Mientras que, si* f_t *depende de* h_t *sólamente a través de* x_t, *diremos que* f_t *es una regla de decisión* markoviana, *y en cuyo caso podemos decir que una regla de este tipo es una función* $f_t : \mathbb{X} \to \mathbb{A}$ *tal que* $f_t(x) \in A(x)$.

DEFINICIÓN 1.3 Una *política de control admisible* (o simplemente una *política*) es una sucesión $\pi = \{f_0, f_1, \ldots\}$ de reglas de decisión. Si las f_t son markovianas diremos que la política π es *markoviana*, y en caso de que $f_t \equiv f$ para alguna $f : \mathbb{X} \to \mathbb{A}$, es decir, $\pi = \{f, f, \ldots\}$, diremos entonces que la política es *estacionaria*.

Denotaremos por Π *al conjunto de todas las políticas. Y, definiendo el conjunto*

$$\mathbb{F} := \{f : \mathbb{X} \to \mathbb{A} \mid f(x) \in A(x)\}, \tag{1.6}$$

debido a la Definición 1.3 y sin pérdida de generalidad, en adelante identificaremos al conjunto de políticas estacionarias con $\mathbb{F}$.

En particular, una política en el caso de un MCM con horizonte de planeación finito N toma la forma $\pi = \{f_0, f_1, \ldots, f_{N-1}\}$.

Por otra parte, en un MCM con horizonte de planeación $N < \infty$, *definimos el espacio muestral como*

$$\Omega_N := \mathbb{K}^N \times \mathbb{X},$$

cuyos elementos son las trayectorias

$$\omega = (x_0, a_0, \ldots, x_{N-1}, a_{N-1}, x_N)$$

con $(x_k, a_k) \in \mathbb{K}$ *para* $k = 0, 1, \ldots, N-1$ *y* $x_N \in \mathbb{X}$; *mientras que, el espacio muestral correspondiente al caso en que* $N = \infty$ *toma la forma*

$$\Omega := \mathbb{K}^\infty,$$

y sus respectivas trayectorias son de la forma

$$\omega = (x_0, a_0, \ldots, x_t, a_t, \ldots).$$

En adelante denotaremos por x_k y a_k ($k \in \mathbb{N}_0$) a las variables de estado y de control en la k-ésima etapa, respectivamente.

Para un estado $x \in \mathbb{X}$ y una política $\pi = \{f_0, f_1, \ldots\} \in \Pi$, existe una probabilidad denotada por P_x^π definida en una familia de subconjuntos de Ω tal que las variables x_k y a_k satisfacen

$$P_x^\pi [x_0 = x] = 1,$$
$$a_k = f_k(h_k) \quad \forall h_k \in \mathbb{H}_k \tag{1.7}$$

y

$$P_x^\pi [x_{t+1} = y \mid h_t, a_t] = P_{x_t,y}(a_t). \tag{1.8}$$

En el caso de horizonte finito ($N < \infty$), la probabilidad P_x^π se define de forma explícita mediante

$$P_x^\pi (x_0, a_0, \ldots, x_{N-1}, a_{N-1}, x_N)$$
$$= \rho_x(x_0) P_{x_0,x_1}(a_0) \cdots P_{x_{N-1},x_N}(a_{N-1}),$$

donde $a_k = f_k(x_0, a_0, \ldots, x_{k-1}, a_{k-1}, x_k)$, $k = 0, 1, \ldots, N-1$, y $\rho_x(\cdot)$ representa la probabilidad concentrada en x.

Observación 1.2 *(a) Denotaremos por E_x^π al operador esperanza con respecto a P_x^π, es decir, si W es una v.a. definida en Ω (o Ω_N), su valor esperado está dado por*

$$E_x^\pi [W] = \sum W(x_0, a_0, \ldots, x_{N-1}, a_{N-1}, x_N) P_x^\pi (x_0, a_0, \ldots, x_{N-1}, a_{N-1}, x_N),$$

donde la suma se toma sobre todas las trayectorias en Ω_N.

(b) Si W es una función de $a_t, \ldots, x_N$, y $h_t \in \mathbb{H}_t$, entonces

$$E_x^\pi [W(x_t, a_t, \ldots, x_N \mid h_t)] = \sum W(x_t, a_t, \ldots, x_N) P_x^\pi (x_t, a_t, \ldots, x_N \mid h_t).$$

(c) Si v es una función de x_{t+1}, entonces

$$E_x^\pi [v(x_{t+1} \mid h_t, a_t)] = \sum_{y \in \mathbb{X}} v(y) P_{x_t,y}(f_t(h_t)). \tag{1.9}$$

1.4 Índice de funcionamiento y problema de control óptimo

En general, un índice de funcionamiento (o criterio de optimalidad) consiste en una función que de alguna manera, "mide" el comportamiento del sistema, digamos el costo total, al utilizar diferentes políticas de control, dado el estado inicial.

Entonces, si $w(\pi,x)$ representa el costo que se genera al utilizar la política π cuando el estado inicial es $x_0 = x$, el problema de control óptimo *(PCO) consiste en determinar una política π^* tal que*

$$w(\pi^*,x) = \inf_{\pi\in\Pi} w(\pi,x) =: \hat{w}(x) \quad \forall x \in \mathbb{X} \tag{1.10}$$

Llamaremos a π^* política óptima, *y a la función obtenida, $\hat{w}(x)$, la* función de valor óptimo.

A continuación incluimos tres índices de funcionamiento usuales, en los cuales, E_x^π denota el operador esperanza cuando se usa específicamente la política π dado que el estado inicial es $x_0 = x$.

DEFINICIÓN 1.4 Sean $x \in \mathbb{X}$ y $\pi \in \Pi$. Se define:

(a) El costo total esperado hasta la N-ésima etapa *por*

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

donde $c_N(x)$ es una función definida para cada $x \in \mathbb{X}$, y puede ser interpretada como un "costo terminal".

(b) El costo total esperado α-descontado *mediante*

$$V_\alpha(\pi,x) := E_x^\pi\left[\sum_{t=0}^{\infty} \alpha^t c(x_t,a_t)\right], \tag{1.11}$$

donde $\alpha \in (0,1)$ representa el factor de descuento.

(c) El costo promedio esperado *por*

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

1.5 Ejemplo: un sistema de inventario

Con el propósito de ilustrar la teoría desarrollada en este capítulo, a continuación consideramos un sistema de inventario con capacidad finita C, en el cual para cada $t \in \mathbb{N}_0$*:*

- x_t *representa el nivel de inventario de determinado artículo al inicio de la etapa* t.
- a_t *representa la cantidad de artículos solicitados a la unidad de producción, a fin de abastecer la unidad de inventario al inicio de la etapa* t*, la cual suponemos es suministrada en forma inmediata.*
- ξ_t *representa la demanda en la etapa* t*, y suponemos que* $\{\xi_t\}$ *es una sucesión de v.a.'s i.i.d. (con valores en* $\mathbb{N}_0$*), y función de probabilidad común* q.

De lo anterior se observa que:

- $\mathbb{X}$=$\mathbb{A}$=$\{0,1,...,C\}$
- *Debido a la capacidad del sistema, si* $x_t = x$*, entonces* ***solo tendría sentido*** *solicitar a la unidad de producción una cantidad de artículos dada por* $a_t = a \in A(x) = \{0,1,...,C-x\}$*; esto es, cada* $x \in \mathbb{X}$ *tiene asociado un conjunto no vacío* $A(x) \subset \mathbb{A}$ *(de controles admisibles cuando el sistema está en el estado* x*).*
- *En tales condiciones, la dinámica de las variables de estado puede modelarse mediante el sistema de ecuaciones en diferencias*

$$x_{t+1} = (x_t + a_t - \xi_t)^+, \quad t \in \mathbb{N}_0 \text{ y } x_0 = \tilde{x} \in \mathbb{X}. \tag{1.12}$$

- *Consideremos que la evolución de este sistema se ha observado hasta la etapa* t*, de manera tal que se conoce la historia correspondiente mediante los valores específicos de* $x_0, a_0, x_1, a_1, ..., x_t, a_t$*, y supongamos además que,* *en particular, $x_t = x$ y $a_t = a$. De aquí, usando (1.12) y el hecho de que las v.a.'s ξ_t son i.i.d. con función de probabilidad común $q(\cdot)$, entonces*

$$\begin{aligned}
&P[x_{t+1} = y \mid x_0, a_0, x_1, a_1, \ldots, x_{t-1}, a_{t-1}, x_t = x, a_t = a] \qquad (1.13)\\
&= P\left[(x_t + a_t - \xi_t)^+ = y \mid \right.\\
&\qquad\qquad \left. x_0, a_0, x_1, a_1, \ldots, x_{t-1}, a_{t-1}, x_t = x, a_t = a\right]\\
&= P\left[(x + a - \xi_t)^+ = y\right]\\
&= \sum_{\xi \in W} q(\xi),
\end{aligned}$$

donde $W = \{\xi \in \mathbb{N}_0 \,:\, (x + a - \xi)^+ = y\}$; de lo cual se desprende de manera natural que, la probabilidad en (1.13) ***depende únicamente*** *del último estado observado ($x_t = x$) y del control respectivo ($a_t = a$), sin importar la $(t-1)$-historia del sistema, ni el valor de t. Es decir, para todo $x, y \in \mathbb{X}$, $a \in A(x)$ y $t \in \mathbb{N}_0$ se tiene*

$$\begin{aligned}
P[x_{t+1} = y \mid x_0, a_0, x_1, a_1, \ldots, x_{t-1}, a_{t-1}, x_t = x, a_t = a]&\\
= P\left[(x_t + a_t - \xi_t)^+ = y \mid x_t = x, a_t = a\right]&\\
= P_{x,y}(a)\,,&
\end{aligned}$$

que es la ley de transición del sistema correspondiente (en una etapa).

- *Finalmente, definiendo las constantes λ y h como sigue*

$$\begin{aligned}
\lambda &: \text{ precio (unitario) de producción,}\\
h &: \text{ costo (unitario) de almacenamiento,}
\end{aligned}$$

tenemos que el costo por etapa, para $t \in \mathbb{N}_0$, queda determinado por

$$c(x, a) = \lambda a + hE_{\xi_t}\left[(x + a - \xi_t)^+\right].$$

Cabe señalar que en lo que resta de este trabajo consideraremos el índice de costo total esperado α-descontado introducido en (1.11). Asimismo, como veremos más adelante, nos enfocaremos al caso de horizonte de planeación infinito.

Capítulo 2

Criterio de Costo Descontado

2.1 Introducción

En este capítulo analizaremos el PCO asociado al MCM en (1.1) bajo un criterio de costo total esperado α*-descontado (véase (1.11)). Cabe mencionar que, por lo regular, este tipo de índice de funcionamiento encuentra aplicaciones en problemas en los cuales tiene una interpretación económica (o monetaria). En tal situación, se introduce un factor de descuento al costo, debido al hecho de que, cierta cantidad de dinero en el presente tiene menos valor en el futuro. De hecho, en muchos problemas el factor de descuento* α *se interpreta como* $\alpha = 1/(1+i)$*, donde* i *denota la tasa de interés. De modo que,* α^t *representa el valor presente de la moneda* t *períodos después.*

A lo largo de esta segunda parte del trabajo, bajo condiciones específicas sobre costo por etapa acotado y finitud de los conjuntos de acciones admisibles, estableceremos y demostraremos resultados relevantes que resuelven el PCO correspondiente al mencionado índice de funcionamiento.

2.2 Criterio de costo descontado

Retomando de (1.11) en el Capítulo 1, recordemos que para $x \in \mathbb{X}$ *y* $\pi \in \Pi$*, la expresión*

$$V_\alpha(\pi, x) := E_x^\pi \left[\sum_{t=0}^{\infty} \alpha^t c(x_t, a_t) \right]$$

define el costo total esperado α-descontado al usar la política π cuando el estado inicial es $x_0 = x$, con $\alpha \in (0,1)$ como factor de descuento. En cuyo caso, obsérvese que usando (1.10) el PCO respectivo consiste de manera específica en encontrar una política $\pi^* \in \Pi$ tal que ***minimice*** *la función introducida en (1.11), es decir,*

$$V_\alpha(\pi^*, x) = \inf_{\pi \in \Pi} V_\alpha(\pi, x) \quad \forall x \in \mathbb{X}.$$

De lo anterior (véase (1.10)), se tiene además que la función (que para distinguir aquí llamaremos) de valor α-óptimo *cuando el estado inicial es $x_0 = x$ queda definida como*

$$V^*(x) := \inf_{\pi \in \Pi} V_\alpha(\pi, x), \quad x \in \mathbb{X}, \tag{2.1}$$

mientras que, llamaremos a π^* una política α-óptima *(para el modelo MCM (1.1)) si*

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

2.3 Condiciones

En lo sucesivo asumiremos que se cumplen las siguientes condiciones:

Hipótesis 2.1 *(a) Para cada $x \in \mathbb{X}$, $A(x)$ es un conjunto finito.*

(b) Existe una constante $M > 0$ tal que

$$|c(x,a)| \leq M, \quad \forall (x,a) \in \mathbb{K}. \tag{2.2}$$

Se incluye a continuación una consecuencia importante de la Hipótesis previa.

Proposición 2.2 *La Hipótesis 2.1(b) implica que el índice $V_\alpha(\pi, x)$ en (1.11) está acotado.*

Demostración. *Primero, para cada $t \in \mathbb{N}_0$ definamos las v.a.'s X_t y Y_t como sigue*

$$X_t := \alpha^t c(x_t, a_t) \ \text{ y } \ Y_t := |X_t|.$$

Nótese que, por una propiedad del valor absoluto se tiene

$$P[X_t \leq Y_t] = 1 \quad \forall t \in \mathbb{N}_0.$$

De lo cual, por el Teorema A.2(d), para cada $x \in \mathbb{X}$ y $\pi \in \Pi$ se cumple

$$E_x^\pi [X_t] \leq E_x^\pi [Y_t] \quad \forall t \in \mathbb{N}_0;$$

mientras que, de (2.2) vemos que

$$E_x^\pi [Y_t] \leq M\alpha^t < \infty \quad \forall t \in \mathbb{N}_0.$$

De aquí, por el Teorema A.2(a)-(c) se tiene que, para cada $x \in \mathbb{X}$ y $\pi \in \Pi$:

$$\begin{aligned} |V_\alpha(\pi, x)| &= \left| \sum_{t=0}^{\infty} E_x^\pi \left[\alpha^t c(x_t, a_t)\right] \right| \\ &\leq M \sum_{t=0}^{\infty} \alpha^t = \frac{M}{1-\alpha} < \infty \quad \forall x \in \mathbb{X} \text{ y } \pi \in \Pi, \end{aligned} \tag{2.3}$$

debido a que $\alpha \in (0,1)$. Lo cual demuestra la Proposición.

■

De hecho, dicha propiedad nos facilitará el análisis, ya que nos permitirá apoyarnos en la teoría de ecuaciones sobre espacios lineales normados, a fin de establecer los principales resultados de optimalidad α-descontada.

Para lo anterior, denotaremos por $B(\mathbb{X})$ al espacio lineal normado consistente de todas las funciones acotadas $v : \mathbb{X} \to \mathbb{R}$. Además, definamos la norma de $v \in B(\mathbb{X})$ como

$$\|v\| := \sup_{x \in \mathbb{X}} |v(x)|. \tag{2.4}$$

Observación 2.3 *Como consecuencias directas nótese que:*

(a) $B(\mathbb{X})$ es un espacio de Banach.

(b) De (2.1) y la Proposición 2.2 se tiene que $V^* \in B(\mathbb{X})$, y

$$|V^*(x)| \leq \frac{M}{1-\alpha} \quad \forall x \in \mathbb{X}.$$

2.4 Ecuación de optimalidad

A continuación introduciremos un elemento, el cual es la clave para caracterizar y obtener políticas óptimas.

DEFINICIÓN 2.1 Diremos que una función $u \in B(\mathbb{X})$ es una solución de la *ecuación de optimalidad* α*-descontada* (EO) si

$$u(x) = \min_{a \in A(x)} \left\{ c(x,a) + \alpha \sum_{y \in \mathbb{X}} u(y) P_{x,y}(a) \right\} \quad \forall x \in \mathbb{X}. \tag{2.5}$$

El objetivo de este capítulo es demostrar que, bajo la Hipótesis 2.1, la función de valor α*-óptimo (véase (2.1)) satisface la EO, lo cual nos permitirá mostrar la existencia de políticas óptimas para el modelo MCM (1.1). Para tal fin introduciremos nueva notación, así como algunos resultados preliminares.*

Primeramente, para cada $u \in B(\mathbb{X})$ *definimos el operador*

$$Tu(x) := \min_{a \in A(x)} \left\{ c(x,a) + \alpha \sum_{y \in \mathbb{X}} u(y) P_{x,y}(a) \right\}, \quad x \in \mathbb{X}, \tag{2.6}$$

y para $f \in \mathbb{F}$ *(véase (1.6)),*

$$T_f u(x) := c(x,f) + \alpha \sum_{y \in \mathbb{X}} u(y) P_{x,y}(f), \quad x \in \mathbb{X}. \tag{2.7}$$

Se tiene también que

$$T^t u := T\left[T^{t-1} u\right], \ t \in \mathbb{N}, \ \ y \ \ T^0 u := u;$$

y análogamente, para cada $f \in \mathbb{F}$

$$T_f^t := T_f\left[T_f^{t-1} u\right], \ t \in \mathbb{N}, \ \ y \ \ T_f^0 u := u.$$

Observación 2.4 *Nótese que:*

(a) En términos del operador T, la EO queda expresada como

$$u = Tu, \quad u \in B(\mathbb{X}).$$

(b) La Hipótesis 2.1(a) garantiza que existe $f \in \mathbb{F}$ tal que

$$Tu = T_f u, \quad u \in B(\mathbb{X}).$$

(c) De la Hipótesis 2.1(b), se tiene que para cada $u \in B(\mathbb{X})$ y $t \in \mathbb{N}_0$,

$$T^t u \in B(\mathbb{X})$$

y, además

$$T_f^t u \in B(\mathbb{X}), \quad f \in \mathbb{F}.$$

Los dos resultados que se demuestran a continuación resaltan algunas propiedades importantes de ambos operadores, T y T_f, previamente definidos.

Proposición 2.5 *Bajo la Hipótesis 2.1(b), T y T_f $(f \in \mathbb{F})$ son operadores de contracción (módulo α) sobre $B(\mathbb{X})$ con la norma introducida en (2.4), ésto es, para cada par de funciones $u, v \in B(\mathbb{X})$:*

(a) $\|Tu - Tv\| \leq \alpha \|u - v\|$, y

(b) $\|T_f u - T_f v\| \leq \alpha \|u - v\|$.

Demostración.

(a) Primero tenemos que para cada $u,v \in B(\mathbb{X})$, $x \in \mathbb{X}$ *y* $a \in A(x)$ *se cumple*

$$\begin{aligned}
c(x,a)+\alpha\sum_{y\in\mathbb{X}}u(y)P_{x,y}(a) &= c(x,a)+\left[\alpha\sum_{y\in\mathbb{X}}v(y)P_{x,y}(a)\right.\\
&\quad\left.-\alpha\sum_{y\in\mathbb{X}}v(y)P_{x,y}(a)\right]+\alpha\sum_{y\in\mathbb{X}}u(y)P_{x,y}(a)\\
&\leq c(x,a)+\alpha\sum_{y\in\mathbb{X}}v(y)P_{x,y}(a)\\
&\quad+\alpha\sum_{y\in\mathbb{X}}|u(y)-v(y)|P_{x,y}(a)\\
&\leq c(x,a)+\alpha\sum_{y\in\mathbb{X}}v(y)P_{x,y}(a)\\
&\quad+\alpha\sup_{y\in\mathbb{X}}|u(y)-v(y)|.
\end{aligned}$$

Ahora, tomando el mínimo sobre $A(x)$ *en ambos lados de esta desigualdad, y de acuerdo con (2.4) y (2.6), vemos que para cada* $x \in \mathbb{X}$*:*

$$Tu(x) \leq Tv(x)+\alpha\|u-v\|,$$

lo cual es equivalente con la expresión

$$Tu(x)-Tv(x) \leq \alpha\|u-v\| \quad \forall x\in\mathbb{X}. \tag{2.8}$$

Luego, siguiendo un procedimiento completamente análogo es posible observar que además se tiene

$$Tv(x)-Tu(x) \leq \alpha\|u-v\| \quad \forall x\in\mathbb{X}. \tag{2.9}$$

De manera que, (2.8) y (2.9) implican que

$$|Tu(x)-Tv(x)| \leq \alpha\|u-v\| \quad \forall x\in\mathbb{X}. \tag{2.10}$$

Finalmente, tomando supremo sobre $\mathbb{X}$ *en (2.10) se obtiene la afirmación de la parte (a).*

(b) La demostración de esta parte sigue un esquema similar al previo.

■

Proposición 2.6 *(a) El operador T tiene un único punto fijo en $B(\mathbb{X})$.*

(b) Para cada $f \in \mathbb{F}$, el operador T_f tiene un único punto fijo en $B(\mathbb{X})$.

Demostración. *Ambas afirmaciones son consecuencia directa de la Proposición 2.5 y el Teorema de Punto Fijo (véase Apéndice C).*

■

2.5 Resultados

Los resultados que se presentan y demuestran en esta sección están orientados a resolver la EO bajo las condiciones impuestas en la Hipótesis 2.1.

Proposición 2.7 *(a) El punto fijo de T_f es $V_\alpha(f,\cdot)$, es decir,*

$$V_\alpha(f,x) = T_f V_\alpha(f,x) \quad \forall x \in \mathbb{X}. \tag{2.11}$$

(b) Una política $\pi = \{f_t\}$ es α-óptima si, y solo si, $V_\alpha(\pi,x)$ es punto fijo de T.

Demostración. *(a) Nótese que de (1.11), junto con los Teoremas A.1, A.2(c) y A.3 (véase Apéndice A) se obtiene lo siguiente*

$$\begin{aligned}
V_\alpha(f,x) \quad : &= E_x^f\left[\sum_{t=0}^{\infty} \alpha^t c(x_t,a_t)\right] \\
&= c(x,f) + \alpha E_x^f\left[\sum_{t=1}^{\infty} \alpha^{t-1} c(x_t,a_t)\right] \\
&= c(x,f) + \alpha E_x^f\left[E_x^f\left[\sum_{t=1}^{\infty} \alpha^{t-1} c(x_t,a_t)\,\middle|\, x_1,a_1\right]\right] \\
&= c(x,f) + \alpha E_x^f\left[V_\alpha(f,x_1)\right] \\
&= c(x,f) + \alpha \sum_{y\in\mathbb{X}} V_\alpha(f,y) P_{x,y}(f) \quad \forall x \in \mathbb{X},
\end{aligned}$$

es decir, se cumple (2.11), y en consecuencia, $V_\alpha(f\,,\cdot)$ *es el punto fijo de* T_f.

(b) Primero, supongamos que

$$u(x) = V_\alpha(\pi, x)$$

es punto fijo de T. *Entonces,*

$$u(x) = \min_{a \in A(x)} \left\{ c(x,a) + \alpha \sum_{y \in \mathbb{X}} u(y) P_{x,y}(a) \right\}. \tag{2.12}$$

Sea $\pi' = \left\{ f_t' \right\}$ *una política arbitraria. Obsérvese que, de acuerdo con (1.7), (1.8) y (1.9), se cumple lo siguiente*

$$\begin{aligned} E_x^{\pi'}\left[\alpha^{t+1} u(x_{t+1}) \middle| h_t, a_t\right] &= \sum_{y \in \mathbb{X}} \alpha^{t+1} u(y) P_x^{\pi'}\left[x_{t+1} = y \middle| h_t, a_t\right] \\ &= \alpha^{t+1} \sum_{y \in \mathbb{X}} u(y) P_{x_t,y}\left(f_t'(h_t)\right). \end{aligned}$$

De aquí y por (2.12), nótese que

$$\begin{aligned} E_x^{\pi'}\left[\alpha^{t+1} u(x_{t+1}) \middle| h_t, a_t\right] &= \alpha^{t+1} \sum_{y \in \mathbb{X}} u(y) P_{x_t,y}\left(f_t'(h_t)\right) \pm \alpha^t c\left(x_t, f_t'(h_t)\right) \\ &= \alpha^t \left[c\left(x_t, f_t'(h_t)\right) + \alpha \sum_{y \in \mathbb{X}} u(y) P_{x_t,y}\left(f_t'(h_t)\right) \right] \\ &\qquad - \alpha^t c\left(x_t, f_t'(h_t)\right) \\ &\geq \alpha^t u(x_t) - \alpha^t c\left(x_t, f_t'(h_t)\right). \end{aligned}$$

Es decir,

$$\alpha^t c\left(x_t, f_t'(h_t)\right) \geq \alpha^t u(x_t) - E_x^{\pi'}\left[\alpha^{t+1} u(x_{t+1}) \middle| h_t, a_t\right],$$

de lo cual, por los Teoremas A.2(c) y A.3 se tiene

$$E_x^{\pi'}\left[\alpha^t c(x_t, a_t)\right] \geq \alpha^t E_x^{\pi'}\left[u(x_t)\right] - \alpha^{t+1} E_x^{\pi'}\left[u(x_{t+1})\right],$$

expresión en la que, sumando de ambos lados desde $t = 0$ *hasta* n, *vemos que*

$$E_x^{\pi'}\left[u(x_0)\right] - \alpha^{n+1} E_x^{\pi'}\left[u(x_{n+1})\right] \leq E_x^{\pi'}\left[\sum_{t=0}^{n} \alpha^t c(x_t, a_t)\right],$$

donde, tomando límite cuando $n \to \infty$, debido a que $u(x)$ es acotada y $\alpha \in (0,1)$, se obtiene que

$$u(x) \leq V_\alpha(\pi', x),$$

ésto es,

$$V_\alpha(\pi, x) \leq V_\alpha(\pi', x)$$

y, como π' es arbitraria,

$$V_\alpha(\pi, x) = V^*(x).$$

Por consiguiente, π es una política α-óptima.

Ahora, supóngase que π es una política α-óptima, es decir,

$$u(x) = V_\alpha(\pi, x) = V^*(x).$$

Demostraremos que

$$u \geq Tu \quad y \quad u \leq Tu, \tag{2.13}$$

simultáneamente.

Para demostrar la primera desigualdad en (2.13) considérese la expresión

$$u(x) = E_x^\pi \left[\sum_{t=0}^{\infty} \alpha^t c(x_t, a_t) \right],$$

de la cual, por el Teorema A.3 vemos que

$$u(x) = c(x, f_0) + \alpha E_x^\pi \left[V_\alpha(\pi', x_1) \right]$$

donde

$$\pi' = \{f_t\}_{t \in \mathbb{N}}.$$

De aquí,

$$u(x) \geq c(x, f_0) + \alpha E_x^\pi [u(x_1)]$$

por lo que se obtiene

$$u(x) \geq \min_{a \in A(x)} \left\{ c(x,a) + \alpha \sum_{y \in \mathbb{X}} u(y) P_{x,y}(a) \right\},$$

lo cual demuestra que $u \geq Tu$.

Con el fin de demostrar la segunda desigualdad en (2.13), sea $g \in \mathbb{F}$ *arbitraria y definiendo la política*

$$\pi' = \{g, \pi\}$$

se tiene que

$$u(x) \leq V_\alpha(\pi', x),$$

de donde

$$u(x) \leq c(x, g) + \alpha E_x^{\pi'}[V_\alpha(\pi, x_1)].$$

Como $u(x_1) = V_\alpha(\pi, x_1)$, *entonces de la desigualdad previa se tiene*

$$u(x) \leq c(x, g) + \alpha \sum_{y \in \mathbb{X}} u(y) P_{x,y}(g),$$

y, dado que $g \in \mathbb{F}$ *es arbitraria, entonces*

$$u(x) \leq \min_{a \in A(x)} \left\{ c(x, a) + \alpha \sum_{y \in \mathbb{X}} u(y) P_{x,y}(a) \right\},$$

lo cual conduce a que $u \leq Tu$.

■

Teorema 2.8 *(a)* V^* *es la única solución acotada de la EO.*

(b) $\pi = \{f\}$ *es una política* α*-óptima si, y solo si,* f *minimiza el lado derecho de la EO, es decir,*

$$V^*(x) = c(x, f) + \alpha \sum_{y \in \mathbb{X}} V^*(y) P_{x,y}(f).$$

Demostración. *(a) Debido a que* T *es un operador de contracción y* $B(\mathbb{X})$ *es un espacio de Banach, entonces por el Teorema de Punto Fijo (véase Apéndice C) existe* $u \in B(\mathbb{X})$ *tal que*

$$\begin{aligned} u(x) &= Tu(x) \\ &= \min_{a \in A(x)} \left\{ c(x, a) + \alpha \sum_{y \in \mathbb{X}} u(y) P_{x,y}(a) \right\}. \end{aligned}$$

Sea $g \in \mathbb{F}$ tal que

$$u(x) = T_g u(x).$$

Luego, por la Proposición 2.7(a)

$$u(x) = V_\alpha(g,x),$$

lo cual implica que $\pi = \{g\}$ es una política α-óptima, y entonces

$$u(x) = V^*(x).$$

(b) Primero supongamos que $\pi = \{f\}$ es una política α-óptima. Entonces, de la Proposición 2.7(b)

$$TV_\alpha(f,x) = V_\alpha(f,x) = V^*(x), \tag{2.14}$$

es decir,

$$\min_{a \in A(x)} \left\{ c(x,a) + \alpha \sum_{y \in \mathbb{X}} V_\alpha(f,y) P_{x,y}(a) \right\} = \min_{a \in A(x)} \left\{ c(x,a) + \alpha \sum_{y \in \mathbb{X}} V^*(y) P_{x,y}(a) \right\},$$

y, dado que por la Proposición 2.7(a)

$$V_\alpha(f,x) = T_f V_\alpha(f,x),$$

por (2.14) tenemos

$$c(x,f) + \alpha \sum_{y \in \mathbb{X}} V^*(y) P_{x,y}(f) = \min_{a \in A(x)} \left\{ c(x,a) + \alpha \sum_{y \in \mathbb{X}} V^*(y) P_{x,y}(a) \right\}.$$

Por consiguiente, f minimiza el lado derecho de la EO.

Supongamos ahora que f minimiza el lado derecho de la EO, en tal situación

$$V^*(x) = T_f V^*(x),$$

y, como por la Proposición 2.7(a) se tiene

$$V^*(x) = V_\alpha(f,x),$$

entonces $\pi = \{f\}$ es una política α-óptima.

■

2.6 Algoritmo de iteración de valores

En esta sección presentamos el resultado que garantiza la convergencia del algoritmo de Iteración de Valores (IterVal) a la función de valor α-óptimo. Para esto, nótese que de la Proposición 2.2, el Teorema 2.1 y la Observación C.1 (ii), para cada $u \in B(\mathbb{X})$ y $t \in \mathbb{N}_0$,

$$\left\|T^t u - V^*\right\| \leq \alpha^t \left\|u - V^*\right\| \tag{2.15}$$

Definamos ahora la sucesión $\{v_t\}$ de funciones de IterVal como sigue,

$$v_0 := 0, \tag{2.16}$$

y para $t \in \mathbb{N}$,

$$v_t(x) := Tv_{t-1}(x) = T^t v_0(x), \tag{2.17}$$

de donde, y por la expresión (2.6), se deduce que

$$v_t(x) = \min_{a \in A(x)} \left\{ c(x,a) + \alpha \sum_{y \in \mathbb{X}} v_{t-1}(y) P_{x,y}(a) \right\}, \quad t \in \mathbb{N}, x \in \mathbb{X}. \tag{2.18}$$

A continuación se enuncia y se demuestre el teorema clave para la mencionada convergencia.

Teorema 2.9 *Bajo la Hipótesis 2.1*

$$\|v_t - V^*\| \to 0 \quad \text{cuando} \quad t \to \infty. \tag{2.19}$$

Además, si

$$0 \leq c(x,a) \leq M \quad \forall (x,a) \in \mathbb{K}, \tag{2.20}$$

entonces

$$v_t \nearrow V^* \quad \text{cuando} \quad t \to \infty. \tag{2.21}$$

Demostración. *Si mayor dificultad puede observarse que la convergencia (2.19) de v_t a V^*, resulta como consecuencia de la expresión (2.15) tomando $u = v_0 = 0$, así como de la Observación 2.1 (b). Por consiguiente, debido a (2.17) y al hecho de que $\alpha \in (0,1)$, se tiene*

$$\|v_t - V^*\| \leq \alpha^t \|V^*\| \leq \frac{\alpha^t M}{1-\alpha} \to 0 \text{ cuando } t \to \infty.$$

Por otra parte, nótese que a consecuencia de (2.20) el operador T es monótono, esto es, si $u, v \in B(\mathbb{X})$ tal que $u \leq v$ no es difícil verificar que

$$Tu \leq Tv. \tag{2.22}$$

Además, (2.16) implica

$$v_0 = 0 \leq \min_{a \in A(x)} \left\{ c(x,a) + \alpha \sum_{y \in \mathbb{X}} v_0(y) P_{x,y}(a) \right\} = \min_{a \in A(x)} \{c(x,a)\} = v_1.$$

De aquí, y por (2.22)

$$Tv_0 \leq Tv_1,$$

es decir,

$$v_0 \leq v_1.$$

Luego, siguiendo un procedimiento inductivo se obtiene que

$$v_t \leq v_{t+1} \quad \forall t \in \mathbb{N}_0;$$

lo cual demuestra precisamente la afirmación (2.21).

■

Capítulo 3

Estimación y Control

3.1 Introducción

En este capítulo estudiaremos un caso particular del MCM definido por medio de la ecuación en diferencias como se estableció en (1.4), donde las perturbaciones aleatorias $\{\xi_t\}$ son v.a's i.i.d. con función de probabilidad θ desconocida por el controlador.

Ante la situación previamente descrita, la idea general de nuestro tratamiento consiste en utilizar métodos adecuados de estimación de θ y técnicas de control, a fin de construir una política asintóticamente óptima descontada (véase Definición 3.1) para el PCO asociado a este esquema específico.

3.2 Modelo de control markoviano: caso específico

Consideremos, como en la Observación 1.1, la ecuación en diferencias

$$x_{t+1} = F(x_t, a_t, \xi_t), \quad t \in \mathbb{N}_0, \tag{3.1}$$

donde $\{\xi_t\}$ es una sucesión de v.a.'s i.i.d. con valores en algún conjunto numerable $\mathbb{S}$ y $F : \mathbb{X} \times \mathbb{A} \times \mathbb{S} \to \mathbb{X}$ es una función dada. Denotemos por θ a la función de probabilidad común de las v.a.'s ξ_t, es decir,

$$\theta(s) := P[\xi_t = s] \quad \forall t \in \mathbb{N}_0,\ s \in \mathbb{S}.$$

Entonces, la dinámica (3.1) define un caso particular de MCM's en tiempo discreto cuya ley de probabillidad de transición está dada como sigue

$$\begin{aligned} P_{x,x'}(a) &= P\left[x_{t+1} = x' \mid x_t = x, a_t = a\right] \\ &= \sum_{k \in S_F} \theta(k), \end{aligned}$$

con

$$S_F := \left\{s \in \mathbb{S} : F(x,a,s) = x'\right\}.$$

Para ser más específicos, denotaremos este modelo de control particular mediante

$$\mathcal{K} := (\mathbb{X}, \mathbb{A}, \{A(x) : x \in \mathbb{X}\}, \mathbb{S}, F, \theta, c), \tag{3.2}$$

donde los elementos $\mathbb{X}$, $\mathbb{A}$, $A(x)$ y c son como en (1.1). Cabe mencionar que aquí asumiremos que θ es ***desconocida*** *por el controlador. En tal situación, usando la distribución empírica (véase Definición B.4, Apéndice B) para estimar θ, el modelo de control $\mathcal{K}$ tiene la siguiente interpretación. En la etapa t el sistema se encuentra en el estado $x_t = x \in \mathbb{X}$ y el controlador usa la distribución empírica con el propósito de obtener un estimador θ_t de la mencionada distribución desconocida θ, es decir, $\{\theta_t\}$ se obtiene procediendo de acuerdo a la expresión*

$$\theta_t(k) = \frac{1}{t} \sum_{j=0}^{t-1} \delta_k(\xi_j), \quad t \in \mathbb{N}, \tag{3.3}$$

donde

$$\delta_k(\xi_j) := \begin{cases} 1 & si\ \xi_j = k \\ 0 & si\ \xi_j \neq k \end{cases}$$

Luego, el controlador combina este proceso con la historia del sistema para seleccionar un control (o acción) $a_t = a \in \mathbb{A}$ adaptado al estimador, de modo que,

$$a = a_t(\theta_t) \in A(x). \tag{3.4}$$

Entonces, se genera un costo $c(x,a)$ y el sistema avanza a un nuevo estado $x_{t+1} = x' \in \mathbb{X}$ de acuerdo a la ley de probabilidad introducida anteriormente dada por

$$\begin{aligned} P_{x,x'}(a) &= P\left[x_{t+1} = x' \mid x_t = x, a_t = a\right] \\ &= \sum_{k \in S_F} \theta(k). \end{aligned}$$

Y una vez que la transición se presenta, el proceso se repite.

Por otra parte, nótese que, si $\{\xi_0, \xi_1, ...\}$ *es una muestra aleatoria, entonces dada una función* u *se tiene que para cada* $t \in \mathbb{N}$,

$$\sum_{k \in \mathbb{S}} u(F(x,a,k))\, \theta_t(k) = \frac{1}{t} \sum_{j=0}^{t-1} u\left(F\left(x,a,\xi_j\right)\right), \quad (x,a) \in \mathbb{K}. \tag{3.5}$$

En efecto, para obtener lo anterior nótese que de (3.3), para cada $(x,a) \in \mathbb{K}$ *podemos escribir*

$$\begin{aligned} \sum_{k \in \mathbb{S}} u(F(x,a,k))\, \theta_t(k) &= \sum_{k \in \mathbb{S}} u(F(x,a,k)) \frac{1}{t} \sum_{j=0}^{t-1} \delta_k(\xi_j) \\ &= \frac{1}{t} \sum_{k \in \mathbb{S}} \left[\sum_{j=0}^{t-1} u(F(x,a,k))\, \delta_k(\xi_j) \right] \\ &= \frac{1}{t} \left[\sum_{j=0}^{t-1} u(F(x,a,k_1))\, \delta_{k_1}(\xi_j) \right. \\ &\qquad \left. + \sum_{j=0}^{t-1} u(F(x,a,k_2))\, \delta_{k_2}(\xi_j) + ... \right]. \end{aligned}$$

De modo que, definiendo para cada $k \in \mathbb{S}$ *los conjuntos* $J_k := \{j : \xi_j = k\}$, *con* $j = 0, 1, ..., t-1$, *vemos que de lo anterior se obtiene (3.5), lo cual se muestra a continuación*

$$\begin{aligned} \sum_{k \in \mathbb{S}} u(F(x,a,k))\, \theta_t(k) &= \frac{1}{t} \left[\sum_{j \in J_{k_1}} u\left(F\left(x,a,\xi_j\right)\right) \right. \\ &\qquad \left. + \sum_{j \in J_{k_2}} u\left(F\left(x,a,\xi_j\right)\right) + ... \right] \\ &= \frac{1}{t} \sum_{j=0}^{t-1} u\left(F\left(x,a,\xi_j\right)\right), \quad (x,a) \in \mathbb{K}. \end{aligned}$$

3.3 Optimalidad asintótica

Retomando lo expuesto en los Capítulos 1 y 2, aquí estamos interesados en usar un criterio de costo total esperado α-descontado (véase (1.11)) a fin de resolver el PCO asociado al modelo $\mathcal{K}$ previamente introducido, el cual consiste en determinar una política α-óptima tal que minimice el mencionado criterio.

Primeramente nótese que, en términos del Teorema 2.8(a), vemos que la EO correspondiente a este escenario específico toma la forma siguiente

$$V^*(x) = \min_{a\in A(x)} \left\{ c(x,a) + \alpha \sum_{k\in\mathbb{S}} V^*(F(x,a,k))\,\theta(k) \right\} \quad \forall x \in \mathbb{X}. \tag{3.6}$$

De lo cual, existe $f^* \in \mathbb{F}$ tal que

$$V^*(x) = c(x,f^*) + \alpha \sum_{k\in\mathbb{S}} V^*(F(x,f^*,k))\,\theta(k),$$

de modo que, la política $\pi = \{f^*\}$ es α-óptima si, y solo si minimiza la parte derecha de (3.6).

Más aún, cabe señalar que si definimos la función $\Phi : \mathbb{K} \to \mathbb{R}$ como

$$\Phi(x,a) := c(x,a) + \alpha \sum_{k\in\mathbb{S}} V^*(F(x,a,k))\,\theta(k) - V^*(x), \tag{3.7}$$

entonces, por el Teorema 2.8(a) y la expresión (2.5), se tiene que la EO en (3.6) es equivalente a la relación

$$\min_{a\in A(x)} \Phi(x,a) = 0,$$

de donde además, se obtiene que la política $\pi = \{f^*\}$ es α-óptima si, y solo si

$$\Phi(x,f^*) = 0 \quad \forall x \in \mathbb{X}. \tag{3.8}$$

En particular, es importante señalar que de acuerdo al procedimiento implementado por el controlador, (basado en estimación combinado con técnicas de control), como puede observarse de (1.11) en la Definición 1.4, el costo total *esperado α-descontado depende fuertemente de las acciones (controles) seleccionadas durante las primeras etapas, que es precisamente cuando la información respecto a la distribución θ resulta deficiente para el estimador, razón por la cual, en estas circunstancias* ***no*** *es posible garantizar en general, la existencia de una política óptima, (véase [4]); de modo que en tal situación, estudiaremos entonces el concepto de optimalidad de una política dada en el sentido asintótico, cuya idea intuitiva, de acuerdo a (3.7) y (3.8) se establece a continuación.*

DEFINICIÓN 3.1 Diremos que una política $\pi \in \Pi$ es *asintóticamente óptima descontada (AOD)* para el modelo $\mathcal{K}$ en (3.2) si para cada $x \in \mathbb{X}$,

$$E_x^{\pi}\left[\Phi\left(x_t, a_t\right)\right] \to 0 \quad \text{cuando} \quad t \to \infty.$$

3.4 Construcción de políticas adaptadas

Primero, definamos la sucesión de funciones $\{V_t\}_{t=0}^{\infty}$ en $B(\mathbb{X})$ como $V_0 \equiv 0$, y para $t \in \mathbb{N}$ mediante la siguiente ecuación recursiva

$$V_t(x) = \min_{a \in A(x)} \left\{ c(x,a) + \alpha \sum_{k \in \mathbb{S}} V_{t-1}\left(F(x,a,k)\right) \theta_t(k) \right\}, \quad x \in \mathbb{X}. \tag{3.9}$$

A continuación presentamos dos propiedades fundamentales de la sucesión previa.

Proposición 3.1 [1] *Si se satisface la Hipótesis 2.1, entonces:*

(a) $\|V_t - V^*\| \to 0$ P_x^{π}- a.s.[2] *cuando $t \to \infty$.*

(b) Además, para cada $t \in \mathbb{N}$ existe $f_t = f_t^{\theta_t} \in \mathbb{F}$ tal que minimiza el lado derecho de (3.9), esto es,

$$V_t(x) = c(x, f_t) + \alpha \sum_{k \in \mathbb{S}} V_{t-1}\left(F(x, f_t, k)\right) \theta_t(k), \quad x \in \mathbb{X}.$$

[1] Para su demostración véase la sub-Sección 3.4.2 (p.40).

[2] Convergencia casi segura respecto a la medida de probabilidad P_x^{π}. (Véase Apéndice B).

Ahora definamos la política $\hat{\pi} = \{\hat{\pi}_t\}$ *como*

$$\hat{\pi}_t(h_t) = \hat{\pi}_t(h_t;\theta_t) := f_t(x_t), t \in \mathbb{N}, \tag{3.10}$$

y $\hat{\pi}_0$ *alguna acción fija. El objetivo consiste en demostrar que* $\hat{\pi}$ *es una política AOD, lo cual será consecuencia, además de la Proposición 3.1, de los resultados siguientes.*

Definamos la familia $\mathcal{V}$ *de funciones* $V^* : \mathbb{S} \to \mathbb{R}$ *como sigue*

$$\mathcal{V} := \{V^*(F(x,a,\cdot)) : (x,a) \in \mathbb{K}\}.$$

Nótese que de (2.1) y (2.3) en el Capítulo 2, se tiene que la familia $\mathcal{V}$ *es uniformemente acotada; mientras que, dado que* $\mathbb{S}$ *es numerable, entonces, por la Proposición B.2 observamos que*

$$\eta_t \to 0 \quad P_x^{\pi}\text{- a.s. cuando } t \to \infty, \tag{3.11}$$

donde para cada $t \in \mathbb{N}$,

$$\begin{aligned}\eta_t \ := \ & \sup_{(x,a)\in\mathbb{K}} \left| \sum_{k\in\mathbb{S}} V^*(F(x,a,k))\,\theta_{t-1}(k) \right. \\ & \left. - \sum_{k\in\mathbb{S}} V^*(F(x,a,k))\,\theta(k) \right|. \end{aligned} \tag{3.12}$$

Por otro lado, tenemos la siguiente

Proposición 3.2 [3] *Bajo la Hipótesis 2.1(a), para cada* $x \in \mathbb{X}$ *y* $\pi \in \Pi$:

$$\beta_t := \sup_{(x,a)\in\mathbb{K}} |\Phi(x,a) - \Phi_t(x,a)| \to 0 \quad P_x^{\pi}\text{- a.s.}$$

cuando $t \to \infty$, *donde para cada* $t \in \mathbb{N}$, $\Phi_t : \mathbb{K} \to \mathbb{R}$ *es una función definida como*

$$\Phi_t(x,a) := c(x,a) + \alpha \sum_{k\in\mathbb{S}} V_{t-1}(F(x,a,k))\,\theta_t(k) - V_t(x). \tag{3.13}$$

[3] Para su demostración véase la sub-Sección 3.4.3 (p.42).

3.4.1 Resultado principal

Teorema 3.3 *Bajo la Hipótesis 2.1, la política $\hat{\pi}$ introducida en (3.10) es AOD para el modelo $\mathcal{K}$.*

Demostración. *Obsérvese que de (3.13) y por definición de $\hat{\pi} = \{\hat{\pi}_t\}$ en (3.10) se tiene*

$$\Phi_t(\cdot, \hat{\pi}_t(\cdot)) = 0, \quad t \in \mathbb{N}_0.$$

Por lo que, debido a la no negatividad de Φ (véase (3.7)), para cada $t \in \mathbb{N}_0$ se obtiene lo siguiente

$$\begin{aligned} \Phi(x_t, \hat{\pi}_t(h_t)) &= |\Phi(x_t, \hat{\pi}_t(h_t)) - \Phi_t(x_t, \hat{\pi}_t(h_t))| \\ &\leq \sup_{(x,a)\in\mathbb{K}} |\Phi(x,a) - \Phi_t(x,a)| = \beta_t. \end{aligned}$$

De donde, haciendo $t \to \infty$, de acuerdo a la Proposición 3.2 se obtiene

$$\Phi(x_t, \hat{\pi}_t(h_t)) \to 0 \quad P_x^{\hat{\pi}}\text{- a.s.},$$

lo cual, por la Proposición B.1(a), implica la convergencia en probabilidad, ésto es,

$$\Phi(x_t, \hat{\pi}_t(h_t)) \xrightarrow{P_x^{\hat{\pi}}} 0 \quad \textit{cuando } t \to \infty. \tag{3.14}$$

De aquí, y como además, β_t es uniformemente acotada para cada t, entonces por la Proposición B.1(c), β_t converge en la media de orden r para cada $r \geq 1$, es decir,

$$\beta_t \xrightarrow{r} 0 \quad \textit{cuando } t \to \infty.$$

Por consiguiente,

$$E_x^{\hat{\pi}}[\Phi(x_t, \hat{\pi}_t(h_t))] \leq E_x^{\hat{\pi}}[\beta_t] \to 0 \quad \textit{cuando } t \to \infty,$$

ésto es, la política $\hat{\pi}$ es AOD.

■

*Finalmente, concluimos este capítulo presentando las demostraciones correspondientes a las Proposiciones **3.1** y **3.2**.*

3.4.2 Demostración de la Proposición 3.1

(a) Nótese que, para cada $x \in \mathbb{X}$ y $t \in \mathbb{N}$ se tiene

$$
\begin{aligned}
|V_t(x) - V^*(x)| &= \left| \min_{a\in A(x)} \left\{ c(x,a) + \alpha \sum_{k\in\mathbb{S}} V_{t-1}(F(x,a,k))\, \theta_{t-1}(k) \right\} \right. \\
&\quad \left. - \min_{a\in A(x)} \left\{ c(x,a) + \alpha \sum_{k\in\mathbb{S}} V^*(F(x,a,k))\, \theta(k) \right\} \right| \\
&\leq \alpha \max_{a\in A(x)} \left| \sum_{k\in\mathbb{S}} V_{t-1}(F(x,a,k))\, \theta_{t-1}(k) \right. \\
&\quad \left. - \sum_{k\in\mathbb{S}} V^*(F(x,a,k))\, \theta(k) \right| \qquad (3.15)
\end{aligned}
$$

Luego, sumando y restando el término

$$\alpha \sum_{k\in\mathbb{S}} V^*(F(x,a,k))\, \theta_{t-1}(k)$$

en el argumento del lado derecho en la desigualdad (3.15), y reacomodando términos se tiene que para cada $x \in \mathbb{X}$ y $t \in \mathbb{N}$,

$$
\begin{aligned}
|V_t(x) - V^*(x)| \leq \alpha \Bigg\{ & \max_{a\in A(x)} \sum_{k\in\mathbb{S}} \Big| V_{t-1}(F(x,a,k)) \\
& - V^*(F(x,a,k)) \Big| \, \theta_{t-1}(k) \\
& + \max_{a\in A(x)} \Big| \sum_{k\in\mathbb{S}} V^*(F(x,a,k))\, \theta_{t-1}(k) \\
& - \sum_{k\in\mathbb{S}} V^*(F(x,a,k))\, \theta(k) \Big| \Bigg\}. \qquad (3.16)
\end{aligned}
$$

Por lo cual, para cada $t \in \mathbb{N}$

$$\begin{aligned}\|V_t - V^*\| \leq \; & \alpha \sum_{k\in\mathbb{S}} \sup_{(x,a)\in\mathbb{K}} \Big| V_{t-1}(F(x,a,k)) \\ & \qquad - V^*(F(x,a,k)) \Big| \theta_{t-1}(k) \\ & + \alpha \sup_{(x,a)\in\mathbb{K}} \Big| \sum_{k\in\mathbb{S}} V^*(F(x,a,k))\, \theta_{t-1}(k) \\ & \qquad - \sum_{k\in\mathbb{S}} V^*(F(x,a,k))\, \theta(k) \Big| \Big\}.\end{aligned}$$

Esto es,

$$\|V_t - V^*\| \leq \alpha \|V_{t-1} - V^*\| + \alpha\eta_t, \; t \in \mathbb{N}, \tag{3.17}$$

donde η_t fue definido previamente en (3.12).

Ahora, sea

$$\gamma := \limsup \|V_t - V^*\| < \infty. \tag{3.18}$$

Entonces, por (3.11) y la Proposición B.2 se tiene de (3.17) que

$$\gamma \leq \alpha\gamma \quad c.s.$$

De lo anterior, necesariamente $\gamma \equiv 0$, ya que $\alpha \in (0,1)$. Por otra parte, obsérvese que de (2.4), para cada $t \in \mathbb{N}$

$$\|V_t - V^*\| \geq 0,$$

así que, de la propiedad

$$\liminf \|V_t - V^*\| \leq \limsup \|V_t - V^*\|,$$

se obtiene que

$$\liminf \|V_t - V^*\| = 0 \equiv \gamma,$$

es decir,

$$\|V_t - V^*\| \to 0 \quad P_x^{\pi}\text{- } a.s. \text{ cuando } t \to \infty,$$

lo cual demuestra la afirmación de la parte (a).

(b) Nótese que esta parte es consecuencia directa de la Hipótesis 2.1(a), la cual garantiza la existencia de tales minimizadores.

■

3.4.3 Demostración de la Proposición 3.2

Nótese que de (3.7) y (3.13), sumando y restando el término

$$\alpha \sum_{k\in\mathbb{S}} V^*(F(x,a,k))\,\theta_t(k)$$

y por la Desigualdad del Triángulo se tiene que para cada $(x,a)\in\mathbb{K}$,

$$\begin{aligned}
|\Phi(x,a)-\Phi_t(x,a)| \leq\ & \alpha \left| \sum_{k\in\mathbb{S}} V^*(F(x,a,k))\,\theta(k) \right. \\
& \left. - \sum_{k\in\mathbb{S}} V^*(F(x,a,k))\,\theta_t(k) \right| \\
& + \alpha \sum_{k\in\mathbb{S}} \left| V^*(F(x,a,k)) \right. \\
& \left. - V_{t-1}(F(x,a,k)) \right| \theta_t(k) \\
& + |V_t(x)-V^*(x)|\,.
\end{aligned}$$

De donde,

$$
\begin{aligned}
\sup_{(x,a)\in\mathbb{K}} &\left|\Phi(x,a)-\Phi_t(x,a)\right| \\
&\leq \alpha \sup_{(x,a)\in\mathbb{K}} \left|\sum_{k\in\mathbb{S}} V^*\left(F(x,a,k)\right)\theta(k)\right. \\
&\qquad \left. - \sum_{k\in\mathbb{S}} V^*\left(F(x,a,k)\right)\theta_t(k)\right| \\
&\quad + \alpha \sum_{k\in\mathbb{S}} \sup_{(x,a)\in\mathbb{K}} \left|V^*\left(F(x,a,k)\right)\right. \\
&\qquad \left. - V_{t-1}\left(F(x,a,k)\right)\right| \theta_t(k) \\
&\qquad + \sup_{x\in\mathbb{X}} \left|V_t(x)-V^*(x)\right|.
\end{aligned}
$$

Esto es,

$$
\beta_t \leq \alpha\eta_{t+1} + \alpha\left\|V_{t-1}-V^*\right\| + \left\|V_t-V^*\right\|, \tag{3.19}
$$

(para η_{t+1} véase (3.12)).

De manera que, haciendo $t\to\infty$, por (3.11) y la Proposición 3.1(a)*, de (3.19) se obtiene que*

$$
\beta_t \to 0 \quad P_x^\pi\text{- a.s.}
$$

■

Apéndice A

Variables Aleatorias Discretas

Sean $(\Omega, \mathcal{F})$ un espacio medible y P una medida de probabilidad en $\mathcal{F}$.

DEFINICIÓN A.1 Una *variable aleatoria* (v.a.) *real discreta* ξ, definida en un espacio de probabilidad $(\Omega, \mathcal{F}, P)$, es una función con dominio Ω y cuyo rango es un subconjunto de $\mathbb{R}$ a lo más numerable $\{x_1, x_2, \ldots\}$, tal que para cada $j \in \mathbb{N}$:

$$\left\{\omega \in \Omega : \xi(\omega) = x_j\right\} \in \mathcal{F}.$$

DEFINICIÓN A.2 La *función de probabilidad* de una v.a. discreta ξ es la función $f_\xi : \mathbb{R} \to [0,1]$ dada por

$$f_\xi(t) := \begin{cases} P[\xi = t] & \text{si } t \in R_\xi \\ 0 & \text{o.c.} \end{cases}$$

(donde R_ξ denota el rango de ξ), la cual cumple las propiedades a continuación:

(i) $f_\xi(t) \geq 0 \ \ \forall\, t \in \mathbb{R}$, y

(ii) $\sum_{t \in \mathbb{R}} f_\xi(t) = 1$.

Esperanza de v.a.'s discretas: sus propiedades

DEFINICIÓN A.3 Sea ξ una v.a. discreta. Si se satisface al menos una de las condiciones siguientes:

$$\sum_{t>0} t f_\xi(t) < \infty \quad \text{o} \quad \sum_{t<0} t f_\xi(t) > -\infty, \tag{A.1}$$

entonces se define la *esperanza (o valor esperado)* de ξ como

$$E[\xi] := \sum_t t f_\xi(t) \tag{A.2}$$

DEFINICIÓN A.4 Diremos que la v.a. ξ tiene esperanza finita si ambas condiciones en (A.1) se cumplen simultáneamente.

Teorema A.1 *Sean: ξ un n-vector aleatorio con función de probabilidad f_ξ, y h una función tal que $h : \mathbb{R}^n \to \mathbb{R}$. Si la esperanza de la v.a. $Z = h(\xi)$ está bien definida, entonces*

$$E[Z] = \sum_{\mathbf{t}} h(\mathbf{t}) f_\xi(\mathbf{t}).$$

Demostración. *Denotemos por $\{\mathbf{t}_i\}$ y $\{z_j\}$ los distintos "valores" posibles de las v.a.'s ξ y Z, respectivamente. Nótese que, para cada z_j existe al menos un $\mathbf{t}_i$ tal que $z_j = h(\mathbf{t}_i)$. Sea*

$$A_j := \{\mathbf{t}_i : h(\mathbf{t}_i) = z_j\}. \tag{A.3}$$

En tal situación, $\{\mathbf{t} \in A_j\}$ y $\{Z = z_j\}$ denotan exactamente el mismo evento. De modo que,

$$\begin{aligned} P[Z = z_j] &= P[\mathbf{t} \in A_j] \\ &= \sum_{\mathbf{t} \in A_j} h(\mathbf{t}) f_\xi(\mathbf{t}). \end{aligned}$$

De lo anterior,

$$\begin{aligned} \sum_i z_j f_Z(z_j) &= \sum_i z_j P[Z = z_j] \\ &= \sum_j z_j \sum_{\mathbf{t} \in A_j} f_\xi(\mathbf{t}) \\ &= \sum_j \left[\sum_{\mathbf{t} \in A_j} z_j f_\xi(\mathbf{t}) \right]. \end{aligned}$$

Luego, como $h(\mathbf{t}) = z_j$ para $\mathbf{x} \in A_j$, entonces

$$\sum_i z_j f_Z(z_j) = \sum_j \left[\sum_{\mathbf{t} \in A_j} h(\mathbf{t}) \ f_\xi(\mathbf{t}) \right].$$

Finalmente, debido a que de (A.3) los conjuntos A_j son disjuntos para distintos valores de j, y además, su unión es el conjunto de todos los valores posibles de ξ, entonces

$$\sum_i z_j\, f_Z\left(z_j\right) = \sum_{\mathbf{t}} h\left(\mathbf{t}\right)\, f_\xi\left(\mathbf{t}\right).$$

■

Teorema A.2 *Sean ξ_1 y ξ_1 dos v.a.'s con esperanza finita, y sea k una constante.*

(a) Si $P\left[\xi_1 = k\right] = 1$, entonces $E\left[\xi_1\right] = k$.

(b) $E\left[k\xi_1\right] = kE\left[\xi_1\right] < \infty$.

(c) $E\left[\xi_1 + \xi_2\right] < \infty$ y además $E\left[\xi_1 + \xi_2\right] = E\left[\xi_1\right] + E\left[\xi_2\right]$.

(d) Si $P\left[\xi_1 \geq \xi_2\right] = 1$, entonces $E\left[\xi_1\right] \geq E\left[\xi_2\right]$.

(e) $\left|E\left[\xi_1\right]\right| \leq E\left[\left|\xi_2\right|\right]$.

Demostración.

(a) Como $P\left[\xi_1 = k\right] = 1$, entonces

$$f_{\xi_1}\left(t\right) = \begin{cases} 1 & si\ \ t = k \\ 0 & si\ \ t \neq k \end{cases}$$

Por tanto, de (A.2)

$$E\left[\xi_1\right] = kf_{\xi_1}\left(k\right) = k.$$

(b) Sea $h\left(t\right) := kt$. Nótese que

$$\sum_x \left|kt\right| f_{\xi_1}\left(t\right) = \left|k\right| \sum_t \left|t\right| f_{\xi_1}\left(t\right) < \infty,$$

de lo cual, $k\xi_1$ tiene esperanza finita. Así que, por el Teorema A.1

$$E\left[k\xi_1\right] = \sum_t kt f_{\xi_1}\left(t\right) = k \sum_t t f_{\xi_1}\left(t\right) = kE\left[\xi_1\right].$$

(c) Análogamente, sea $h\left(t,s\right) := t + s$. Nótese que

$$\begin{aligned} \sum_{t,s} \left|t+s\right| f_{\xi_1,\xi_2}\left(t,s\right) &\leq \sum_{t,s} \left|t\right| f_{\xi_1,\xi_2}\left(t,s\right) + \sum_{t,s} \left|s\right| f_{\xi_1,\xi_2}\left(t,s\right) \\ &= \sum_t \left|t\right| \sum_s f_{\xi_1,\xi_2}\left(t,s\right) + \sum_s \left|s\right| \sum_t f_{\xi_1,\xi_2}\left(t,s\right) \\ &= \sum_t \left|t\right| \sum_s f_{\xi_1}\left(t\right) + \sum_s \left|s\right| \sum_t f_{\xi_2}\left(s\right) < \infty, \end{aligned}$$

de donde, $\xi_1 + \xi_2$ tiene esperanza finita, y entonces, de nuevo por el Teorema A.1

$$\begin{aligned} E[\xi_1 + \xi_2] &= \sum_{t,s} (t+s) f_{\xi_1,\xi_2}(t,s) \\ &= \sum_{t,s} t f_{\xi_1,Y}(x,s) + \sum_{t,s} y f_{\xi_1,\xi_2}(t,s) \\ &= E[\xi_1] + E[\xi_2]. \end{aligned}$$

(d) Obsérvese que, definiendo la v.a.

$$Z := \xi_1 - \xi_2 = \xi_1 + (-\xi_2), \tag{A.4}$$

entonces, por (b) y (c) tenemos que

$$\begin{aligned} E[\xi_1] - E[\xi_2] &= E[\xi_1 - \xi_2] \\ &= E[Z] = \sum_z z f_Z(z). \end{aligned} \tag{A.5}$$

Dado que por hipótesis

$$P[Z \geq 0] = P[\xi_1 \geq \xi_2] = 1,$$

entonces todos los valores z_j que toma Z (véase (A.4)) tienen que ser no negativos; por lo tanto, de (A.5)

$$\sum_z z f_Z(z) = E[Z] \geq 0,$$

de donde, en efecto

$$E[\xi_1] \geq E[\xi_2].$$

(e) No es difícil observar que para este caso, la demostración se consigue aplicando (b) y (d), ya que

$$-|\xi_1| \leq \xi_1 \leq |\xi_1|$$

implica que

$$-E[|\xi_1|] \leq E[\xi_1] \leq E[|\xi_1|],$$

que es equivalente con lo que se quería demostrar, es decir,

$$|E[\xi_1]| \leq E[|\xi_1|].$$

■

Esperanza condicional de v.a.'s discretas: sus propiedades

DEFINICIÓN A.5 Sean ξ_1 y ξ_2 dos v.a.'s discretas con dominios Ω_1 y Ω_2, respectivamente.

(a) Se define la *función de probabilidad conjunta* de ξ_1 y ξ_2, denotada por f_{ξ_1,ξ_2} ($f_{\xi_1,\xi_2} : (\Omega_1 \times \Omega_2) \to [0,1]$), como

$$f_{\xi_1,\xi_2}(t,s) := P[\xi_1 = t, \xi_2 = s].$$

(b) Se define la *función de probabilidad condicional* de ξ_2 dado ξ_1, denotada por $f_{\xi_2|\xi_1}$ ($f_{\xi_2|\xi_1} : (\Omega_1 \cap \Omega_2) \to [0,1]$), como

$$f_{\xi_2|\xi_1}(s \mid t) := P[\xi_2 = s \mid \xi_1 = t] = \frac{f_{\xi_1,\xi_2}(t,s)}{f_{\xi_1}(t)}, \quad \text{siempre que } f_{\xi_1}(t) > 0.$$

DEFINICIÓN A.6 Sean ξ_1 y ξ_2 dos v.a.'s discretas. Para $t \in \mathbb{R}$ (fijo) tal que $f_{\xi_1}(t) > 0$, se define la *esperanza condicional* de ξ_2 dado $\xi_1 = t$ por

$$E[\xi_2 \mid \xi_1 = t] := \sum_s s f_{\xi_2|\xi_1}(s \mid t).$$

DEFINICIÓN A.7 La esperanza condicional de ξ_2 dado ξ_1 se define como

$$E[\xi_2 \mid \xi_1] := g(\xi_1),$$

donde

$$g(\xi_1) = E[\xi_2 \mid \xi_1 = t].$$

Teorema A.3 $E[\xi_2 \mid \xi_1]$ *tiene la propiedad de la doble esperanza, es decir,*

$$E[E[\xi_2 \mid \xi_1]] = E[\xi_2].$$

Demostración. *Sea*

$$\psi(\xi_1) := E[\xi_2 \mid \xi_1]. \tag{A.6}$$

Nótese que, por (A.6) y el Teorema A.1

$$\begin{aligned}
E[\psi(\xi_1)] &= \sum_t \psi(t) f_{\xi_1}(t) \\
&= \sum_t \left[\sum_s s f_{\xi_2|\xi_1}(s|t)\right] f_{\xi_1}(t) \\
&= \sum_t \left[\sum_s s f_{\xi_1,\xi_2}(t,s)\right] = \sum_s s \left[\sum_t f_{\xi_1,\xi_2}(t,s)\right] \\
&= \sum_s s f_{\xi_2}(s),
\end{aligned}$$

de donde $E[E[\xi_2 \mid \xi_1]] = E[\xi_2]$.

■

DEFINICIÓN A.8 Sea (ξ_1, ξ_2, ξ_3) un vector aleatorio discreto. Si para $t, s \in \mathbb{R}$ (fijos) $P[\xi_1 = t, \xi_2 = s] > 0$ y además $E[\xi_3]$ está bien definida, entonces la esperanza condicional de ξ_3 dado $\xi_1 = t$ y $\xi_2 = s$ se define por

$$E[\xi_3 \mid \xi_1 = t, \xi_2 = s] := \sum_r r P[\xi_3 = r \mid \xi_1 = t, \xi_2 = s].$$

DEFINICIÓN A.9 La esperanza condicional de ξ_3 dado ξ_1 y ξ_2 se define como

$$E[\xi_3 \mid \xi_1, \xi_2] := g(\xi_1, \xi_2)$$

donde $g(\xi_1, \xi_2) = E[\xi_3 \mid \xi_1 = t, \xi_2 = s]$.

Teorema A.4 $E[E[\xi_3 \mid \xi_1, \xi_2] \mid \xi_1] = E[\xi_3 \mid \xi_1]$.

Demostración. *Del Teorema A.1 y por definición de g se tiene que*

$$
\begin{aligned}
& E\left[g\left(\xi_1,\xi_2\right) \mid \xi_1=x\right] \\
& =\sum_s\left[\sum_r rP\left[\xi_3=r \mid \xi_1=t,\xi_2=s\right]\right] P\left[\xi_2=s \mid \xi_1=t\right] \\
& =\sum_r r\left[\sum_s rP\left[\xi_3=r \mid \xi_1=t,\xi_2=s\right]\right] P\left[\xi_2=s \mid \xi_1=t\right] \\
& =\sum_r r\left[\sum_s \frac{P\left[\xi_3=r,\xi_2=s,\xi_1=t\right]}{P\left[\xi_1=t\right]}\right] \\
& =\sum_r rP\left[\xi_3=r \mid \xi_1=t\right] \\
& =E\left[\xi_3 \mid \xi_1=t\right] \quad \forall t\in R_{\xi_1}.
\end{aligned}
$$

En consecuencia

$$E\left[E\left[\xi_3 \mid \xi_1,\xi_2\right] \mid \xi_1\right]=E\left[\xi_3 \mid \xi_1\right].$$

■

Apéndice B

Convergencia de Variables Aleatorias y Distribución Empírica

Convergencia de v.a.'s

Sean ξ y $\{\xi_t\}$ v.a.'s definidas en un espacio de probabilidad común.

DEFINICIÓN B.1 Diremos que $\{\xi_t\}$ *converge casi seguramente* a ξ, denotado por

$$\xi_t \overset{a.s.}{\rightarrow} \xi$$

o

$$\xi_t \rightarrow \xi \;\; P\text{-} \; a.s. \text{ cuando } t \rightarrow \infty,$$

si

$$P\left[\{\omega \in \Omega : \xi_t(\omega) \rightarrow \xi(\omega) \;\; \text{cuando } t \rightarrow \infty\}\right] = 1.$$

DEFINICIÓN B.2 Diremos que $\{\xi_t\}$ *converge en probabilidad* a ξ, denotado por

$$\xi_t \overset{P}{\rightarrow} \xi$$

si para todo $\varepsilon > 0$,

$$P\left[\{\omega \in \Omega : |\xi_t(\omega) - \xi(\omega)| \geq \varepsilon\}\right] \rightarrow 0 \text{ cuando } t \rightarrow \infty.$$

DEFINICIÓN B.3 Para $r \geq 1$ diremos que $\{\xi_t\}$ *converge en la media de orden r* a ξ, denotado por

$$\xi_t \xrightarrow{r} \xi,$$

si $E[|\xi_t^r|] < \infty$ para todo t y

$$E[|\xi_t - \xi|^r] \to 0 \ \text{ cuando } t \to \infty.$$

Proposición B.1 *Sean ξ y $\{\xi_t\}$ v.a.' s definidas todas en un espacio de probabilidad común.*

(a) Si $\xi_t \xrightarrow{a.s.} \xi$ entonces $\xi_t \xrightarrow{P} \xi$.

(b) Si $\xi_t \xrightarrow{r} \xi$ para todo $r \geq 1$ entonces $\xi_t \xrightarrow{P} \xi$.

(c) Si $\xi_t \xrightarrow{P} \xi$, y además, $P[|\xi_t| \leq K] = 1$ para todo t y alguna constante K, entonces $\xi_t \xrightarrow{r} \xi$ para todo $r \geq 1$.

Demostración. *Véase por ejemplo .[3] p.277.*

Función de distribución empírica

DEFINICIÓN B.4 La *función de distribución empírica* para las v.a.'s $\{\xi_t\}_{t=1}^n$, es la función de distribución denotada por $\theta_t(k) := \theta_t(k;\omega)$ con salto de tamaño $1/t$ en $\xi_i(\omega)$ para cada $i = 1, \ldots, n$, es decir:

$$\theta_t(k) = \frac{1}{t} \sum_{j=0}^{t-1} \delta_k(\xi_j), \ \ t \in \mathbb{N}, \tag{B.1}$$

donde

$$\delta_k(\xi_j) := \begin{cases} 1 & \text{si } \xi_j = k \\ 0 & \text{si } \xi_j \neq k \end{cases}$$

Clase Glivenko-Cantelli

DEFINICIÓN B.5 Sea $\mathcal{H}$ una familia de funciones $h : \mathbb{S} \to \mathbb{R}$. Diremos que $\mathcal{H}$ es una *clase Glivenko-Cantelli* si

$$\sup_{h\in\mathcal{H}} \left| \sum_{k\in\mathbb{S}} h(k)\,\theta_t(k) - \sum_{k\in\mathbb{S}} h(k)\,\theta(k) \right| \to 0 \quad \text{cuando } t \to \infty.$$

Proposición B.2 *Si $\mathcal{H}$ es una familia uniformemente acotada y $\mathbb{S}$ es un conjunto numerable, entonces $\mathcal{H}$ es una* clase Glivenko-Cantelli .

Demostración. *Véase por ejemplo [1] p.17.*

Apéndice C

Teorema de Punto Fijo

Definición C.1 Un espacio métrico es una pareja (S,d), donde S es un conjunto no vacío, y d es una función de $S \times S$ en $\mathbb{R}$ tal que para $x,y,z \in S$ arbitrarios satisface las propiedades siguientes:

(i) $d(x,x) = 0$

(ii) $d(x,y) > 0$ si $x \neq y$

(iii) $d(x,y) = d(y,x)$

(iv) $d(x,y) \leq d(x,z) + d(z,y)$

Definición C.2 Sea (S,d) un espacio métrico. Se dice que (S,d) es un espacio métrico completo si cualquier sucesión de Cauchy en S converge en S.

Definición C.3 Sea (S,d) un espacio métrico. Se dice que un operador

$$T : S \to S$$

es de contracción módulo $\alpha \in (0,1)$, si

$$d(Tx,Ty) \leq \alpha d(x,y) \quad \forall x,y \in S.$$

Teorema C.1 ***(Teorema de Punto Fijo para operadores de contracción)*** *Si (S,d) es un espacio métrico completo y $T : S \to S$ es un operador de contracción, entonces:*

(a) Existe un único $x \in S$ tal que

$$Tx = x.$$

(b) Para cada $y \in S$,

$$\lim_{n\to\infty} T^n y = x.$$

Demostración. (a) *La demostración de la unicidad se hará por contradicción.*

Supongamos la existencia de dos puntos fijos para T. Sean $x, y \in S$ con $x \neq y$ tal que

$$Tx = x \quad y \quad Ty = y, \tag{C.1}$$

de aquí vemos que

$$d(Tx, Ty) = d(x, y). \tag{C.2}$$

Por otra parte, dado que T es operador de contracción se tiene que

$$d(Tx, Ty) \leq \alpha d(x, y). \tag{C.3}$$

Luego, (C.2) y (C.3) implican que

$$d(x, y) \leq \alpha d(x, y),$$

de donde $\alpha \geq 1$, lo cual contradice a la hipótesis de que T es de contracción. Por consiguiente $x = y$.

(b) Sea $y \in S$. Debido a que S es un espacio métrico completo, entonces $\{T^n y\}$ converge si, y solo si

$$d(T^m y, T^n y) \to 0.$$

Por lo cual, supongamos que $m = n + k$ $(m \geq n)$. Ahora nótese que, debido a que T es operador de contracción y por (C.1):

$$\begin{aligned} d\left(T^{n+k}y, T^n y\right) &\leq \alpha d\left(T^{n+k-1}y, T^{n-1}y\right) \leq \alpha^2 d\left(T^{n+k-2}y, T^{n-2}y\right) \\ &\leq \dots \leq \alpha^{n-1} d\left(T^{k+1}y, Ty\right) \\ &= \alpha^{n-1} d\left(T^{k+1}y,\, y\right). \end{aligned} \tag{C.4}$$

Además, de la desigualdad del triángulo (véase Definición C.1) se tiene

$$d\left(T^{k+1}y,\, y\right) \leq d\left(T^{k+1}y, T^k y\right) + d\left(T^k y, T^{k-1}y\right) + \dots + d(Ty,\, y)$$

De aquí y por (C.4) se obtiene

$$d(T^m y, T^n y) \leq \alpha^n \left(\alpha^k + \alpha^{k-1} + ... + \alpha^0\right) d(Ty,\, y),$$

así que, tomando límite de ambos lados cuando $n \to \infty$ se observa que

$$d(T^m y, T^n y) \to 0,$$

en consecuencia, $\{T^n y\}$ es convergente.

Ahora, sea

$$x = \lim_{n\to\infty} T^n y.$$

Como

$$T^{n+1} y \to x \quad \text{cuando } n \to \infty;$$

y, además,

$$T^{n+1} y = T(T^n y) \to Tx \quad \text{cuando } n \to \infty;$$

entonces $Tx = x$.

■

Observación C.2 *(i) Sea $x_0 \in S$ arbitrario. Nótese que podemos definir la sucesión iterativa $\{x_n\}$ como:*

$$\begin{aligned}
x_1 &:= Tx_0, \\
x_2 &:= Tx_1 = T(Tx_0) = T^2 x_0, \\
x_3 &:= Tx_2 = T(Tx_1) = T(T^2 x_0) = T^3 x_0, \\
x_4 &:= Tx_3 = T(Tx_2) = T(T^3 x_0) = T^4 x_0, \\
&\vdots \\
x_n &:= Tx_{n-1} = T(Tx_{n-2}) = \cdots = T^n x_0. \\
&\vdots
\end{aligned}$$

De hecho, $\{x_n\}$ es la sucesión de imágenes de x_0 al aplicar el operador T repetidamente.

[1] *Billingsley P. (1968)* Convergence of Probability Measures. *Editorial John Wiley & Sons, Inc.; New York.*

[2] *Gordienko E.I., Minjárez-Sosa J.A. (1998)* Adaptive control for discrete-time Markov processes with unbounded costs: discounted criterion. *Kyber-*

[10] *Pérez Pérez Aroldo (1996)* Introducción a la Teoría de Control Estocástico. *Tesis de Licenciatura. Universidad de Sonora.*

[11] *Vaart A.W. van der (c1998)* Asymptotic statistics. *Editorial Cambridge University Press; New York, NY. ISBN 978-0-521-49603-9.*

[12] *Vaart A.W. van der, Wellner, J.A. (c1996)* Week Convergence and Empirical Processes: with applications to statistics. *Editorial Springer; New York, NY. ISBN 0-387-94640-3.*