UNIVERSIDAD DE SONORA

DIVISIÓN DE CIENCIAS EXACTAS Y NATURALES

Programa de Licenciatura en Matemáticas

Control minimax de sistemas estocásticos a tiempo discreto con criterio de costo descontado

T E S I S

Que para obtener el grado académico de:

Licenciado en Matemáticas

Presenta:

Luz Esmeralda Almada Valenzuela

Director de tesis: Dr. Jesús Adolfo Minjárez Sosa.

Hermosillo, Sonora, México 11 de octubre de 2018

Este trabajo fue financiado por el Consejo Nacional de Ciencia y Tecnología (CONACYT) dentro del proyecto “Aproximación, Estimación y Control de Sistemas Estocásticos y Juegos Dinámicos”, con número de referencia CB2015-01/254306, bajo la dirección del Dr. Jesús Adolfo Minjárez Sosa.

Sinodales

Dr. David González Sánchez
Departamento de Matemáticas,
Universidad de Sonora

Dra. Carmen Geraldi Higuera Chan
Departamento de Matemáticas,
Universidad de Sonora

Dr. Jesús Adolfo Minjárez Sosa
Departamento de Matemáticas,
Universidad de Sonora

Dra. Luz del Carmen Rosas Rosas
Departamento de Matemáticas,
Universidad de Sonora

Primeramente a Dios, por
permitirme estar aquí.

A mis padres, por siempre confiar en mi
y darme su apoyo incondicional. Los amo.

A mis padrinos, que siempre
me han alentado a seguir adelante.

Agradecimientos

Agradezco a mis padres Luz y Juan Pedro por siempre apoyarme, darme ánimos, creer en mi y permitirme salir a estudiar fuera a pesar de lo difícil que pudo ser. A mis hermanos Pedro y Joanna ustedes siempre lograban sacarme una sonrisa incluso cuando me encontraba estresada por tener mucha tarea o por presentar algún examen, los quiero lo saben. A los amigos que fui haciendo a lo largo de la carrera: Malena, Ireni, Joselyn, Paola, Félix, Chuy, y muchos otros, a pesar de siempre tener mucho trabajo nunca faltaron los momentos de diversión. A mis amigos que han estado desde siempre Diana, Karina, Brisseida, David y Jassiel, a pesar de no vernos seguido sigo recibiendo su apoyo. A mis profesores, gracias por la paciencia y enseñanzas. Y a todos aquellos que me brindaron su ayuda o me regalaron una sonrisa en un mal día.

Índice general

Introducción ix

1. Modelo de control minimax 1

1.1. Descripción del modelo 1

1.2. Políticas de control 3

1.3. Índices de funcionamiento 5

1.4. Problema de control minimax 6

1.5. Algoritmo de PD con horizonte finito 7

2. Criterios de costo descontado 13

2.1. Introducción 13

2.2. Problema minimax con horizonte finito 14

2.3. Ecuación de Optimalidad Minimax 20

3. Sistemas de control estocástico con distribución desconocida 29

3.1. Introducción 29

3.2. Sistemas de control estocástico 30

3.3. Sistemas de control con ruidos aleatorios no observables 33

3.4. Ejemplo: Proceso de control autorregresivo 37

A. Teorema del Punto fijo 39

B. Espacio medible, espacio de Borel y espacio de probabilidad 46

B.1. Espacio medible y espacio de Borel 46

B.2. Espacio de probabilidad 47

C. Glosario de símbolos y abreviaturas **50**

Bibliografía **52**

Introducción

El objetivo de este trabajo es estudiar sistemas de control minimax a tiempo discreto, los cuales consisten en lo siguiente. En contraste con los problemas de control óptimo estándar, en los que solo hay un único responsable de tomar las decisiones, en los problemas de control minimax existen dos tomadores de decisiones, el controlador mismo y un *"oponente"*. Su evolución en el tiempo se puede describir de la siguiente manera. Al tiempo t, cuando el estado del sistema es $x_t = x$, el controlador elige una acción $a_t = a$ y el oponente elige una acción $b_t = b$. Entonces, el controlador paga un costo $c(x, a, b)$ al oponente, y el sistema se mueve a un nuevo estado $x_{t+1} = y$ de acuerdo a una probabilidad de transición $P_{x,y}(a, b)$. Una vez que el sistema se encuentra en el estado y, de nuevo el controlador y el oponente eligen sus respectivas acciones y el proceso se repite una y otra vez. Los costos se acumulan de acuerdo a un funcional de costo total considerando horizonte, ya sea, finito o infinito. Por lo tanto el objetivo del controlador es minimizar el máximo costo generado por el oponente.

Un caso particular de sistemas de control minimax es cuando la evolución del sistema está determinada por una ecuación en diferencias estocástica de la forma

$$x_{t+1} = F(x_t, a_t, b_t, \xi_t), \quad t = 0, 1, \dots,$$

donde F es una función dada y $\{\xi_t\}$ es una sucesión de variables aleatorias conocido como proceso de perturbaciones aleatorias. En este caso, la probabilidad de transición queda determinada por la función F y la distribución correspondiente de las variables aleatorias.

Las decisiones tanto del controlador como del oponente se eligen por medio de reglas llamadas estrategias o políticas de control las cuales, en términos generales, toman la forma $\pi = \{a_t\}$ y $\gamma = \{b_t\}$, respectivamente. Entonces, si $K(x,\pi,\gamma)$ denota el costo total (índice de funcionamiento) cuando el controlador y el oponente eligen las estrategias π y γ, y el estado inicial es $x_0 = x$, el problema de control minimax es encontrar una estrategia π^* que minimice el máximo de $K(x,\pi,\gamma)$ sobre las estrategias γ del oponente. Es decir, π^* debe satisfacer

$$\inf_{\pi} \sup_{\gamma} K(x,\pi,\gamma) = \sup_{\gamma} K(x,\pi^*,\gamma) \tag{0.0.1}$$

para todo estado inicial x. A la estrategia π^* se le conoce como estrategia minimax. En otras palabras, la estrategia minimax π^* garantiza el mejor rendimiento para el controlador, en la peor situación posible.

De acuerdo a la descripción anterior, podemos decir que un problema de control minimax es un caso especial de un juego suma cero, en el cual solo nos interesa encontrar la estrategia que alcance el valor superior

$$U(x) := \inf_{\pi} \sup_{\gamma} K(x,\pi,\gamma),$$

que es precisamente la estrategia minimax π^*.

Específicamente, en este trabajo estudiaremos sistemas de control minimax considerando el índice de costo descontado con horizonte finito e infinito. Además supondremos que el espacio de acciones del controlador es numerable y el del oponente de Borel. Más aún, consideraremos que el costo es posiblemente no acotado, lo cual se analizará en el contexto de normas ponderadas.

Una de las aplicaciones más importantes de los sistemas de control minimax, la cual abordaremos en el presente trabajo, es en sistemas de control que dependen de parámetros desconocidos. A este tipo de sistemas se le conoce como "juegos contra la naturaleza" ya que al oponente se le puede considerar como la naturaleza, la cual elige en cada etapa el parámetro desconocido. Específicamente, consideraremos un sistema de control que evoluciona de acuerdo a una ecuación de la forma

$$x_{t+1} = F(x_t, a_t, \xi_t), \quad t = 0, 1, ... \tag{0.0.2}$$

donde $\{\xi_t\}$ es una sucesión de variables aleatorias con distribución desconocida. Entonces una estrategia de la naturaleza $\gamma = \{b_t\}$ elegirá una distribución b_t para ξ_t en cada etapa t.

Al problema de encontrar una estrategia óptima en sistemas de control como (0.0.2) con distribución desconocida se le conoce como problema de control adaptado. Este comúnmente se analiza asumiendo que $\{\xi_t\}$ es una sucesión de variables aleatorias observables, independientes e idénticamente distribuidas, lo cual permite implementar métodos de estimación estadística para la distribución común desconocida. En nuestro caso, el análisis de este problema como un juego contra la naturaleza permite que las variables aleatorias no necesariamente sean observables, y más aún, ni con la misma distribución.

La teoría desarrollada en el presente trabajo se basa, principalmente, en los artículos [4, 13] y en los libros [9, 2]. El resto de la bibliografía nos sirvió como referencias para cierto material, las cuales se especificarán en su momento.

Capítulo 1

Modelo de control minimax

En este capítulo introduciremos los elementos necesarios para definir un problema de control minimax. Con este fin, primero se describe el modelo de control minimax y se presenta el concepto de estrategia. A continuación definimos los índices de funcionamiento que miden el comportamiento de las estrategias y planteamos el problema de control minimax. Finalmente se introduce el algoritmo de programación dinámica en su versión minimax para el caso de horizonte finito, el cual proporciona una forma de resolver el problema de control minimax. Este algoritmo nos servirá de motivación para estudiar el caso de problemas con horizonte infinito.

1.1. Descripción del modelo

Como ya mencionamos en la introducción del trabajo, un problema de control minimax es una clase especial de juego suma cero de dos personas. Este hecho, motiva la siguiente definición.

Definición 1.1.1. *Un modelo de control minimax (MCM) en tiempo discreto, es un arreglo*

$$MCM := (\mathbb{X}, A, B, \mathbb{K}_A, \mathbb{K}, P, c) \tag{1.1.1}$$

que consiste de los siguiente elementos:

- $\mathbb{X}$ *es el conjunto de estados y suponemos que es numerable.*

- A *es el espacio de acciones del controlador, el cual suponemos numerable.*

- *B es el espacio de acciones del oponente, que asumimos es un espacio de Borel (Véase Apéndice B, Definición B.1.5).*

Además,

- *denotamos por $\mathbb{K}_A \subset \mathbb{X} \times A$ al conjunto de restricciones para el controlador. Es decir, para cada estado $x \in \mathbb{X}$,*

$$A(x) := \{a \in A : (x,a) \in \mathbb{K}_A\}$$

representa el conjunto de acciones admisibles del controlador en el estado x.

- *Denotamos por $\mathbb{K} \in \mathfrak{B}(\mathbb{X} \times A \times B)$ al conjunto de restricciones para el oponente, de modo que para cada par $(x,a) \in \mathbb{K}_A$*

$$B(x,a) := \{b \in B : (x,a,b) \in \mathbb{K}\}$$

es el conjunto de acciones admisibles para el oponente, cuando el estado es $x \in \mathbb{X}$ y el controlador usa la acción $a \in A(x)$.

- *P representa la ley de transición entre estados. Es decir,*

$$P_{x,y}(a,b) := P[x_{t+1} = y | x_t = x, a_t = a, b_t = b]$$

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

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

Obsérvese que si el conjunto de acciones $B(x,a)$ del oponente no depende de $a \in A(x)$, obtenemos el usual juego de Markov de suma cero, donde los jugadores eligen sus acciones $a \in A(x)$ y $b \in B(x)$ de manera independiente.

El modelo de control minimax (1.1.1) representa un juego dinámico que evoluciona a tiempo discreto. Esto es, denotemos por $x_t \in \mathbb{X}$, $a_t \in A(x_t)$ y $b_t \in B(x_t, a_t)$, el estado del sistema, los controles aplicados por el controlador y por el oponente al tiempo t, respectivamente. Entonces la evolución del sistema la describimos de la siguiente manera. Si el estado al tiempo t es $x_t = x \in \mathbb{X}$, entonces el controlador elige una acción $a_t = a \in A(x)$ y el oponente elige una acción $b_t = b \in B(x, a)$. Como consecuencia de esto, suceden dos cosas:

1. El controlador paga $c(x, a, b)$ al oponente.

2. El sistema se mueve a un nuevo estado $x_{t+1} = y \in \mathbb{X}$ con probabilidad $P_{x,y}(a, b)$.

Una vez que el sistema se encuentra en el estado y, se eligen nuevos controles $a' \in A(y)$ y $b' \in B(y, a')$ y el proceso anterior se repite.

Al tiempo entre dos épocas de decisión consecutivas se le llama *periodo* o *etapa*. Si el número de epocas de decisión es finito, decimos que el modelo tiene horizonte de planeación finito y en otro caso decimos que el horizonte de planeación es infinito.

1.2. Políticas de control

Definición 1.2.1. *Dado un MCM y $t \in \mathbb{N}_0$, definimos los espacios de historias admisibles hasta la etapa t como*

$$H_0 := \mathbb{X}, \quad H_0^{\#} := \mathbb{K}_A \quad y$$

$$H_t := \mathbb{K}^t \times \mathbb{X} \quad y \quad H_t^{\#} := \mathbb{K}^t \times \mathbb{K}_A \quad para \quad t \in \mathbb{N}.$$

Los elementos de H_t y $H_t^{\#}$ son "historias" de la forma

$$h_t = (x_0, a_0, b_0, x_1, a_1, b_1, \cdots, x_t) \quad y \quad h_t^{\#} = (h_t, a_t).$$

Una estrategia (política de control) para el controlador es una sucesión $\pi = \{\pi_t\}$ de funciones $\pi_t : H_t \to A$ tal que $a_t = \pi_t(h_t) \in A(x_t) \quad \forall h_t \in H_t, t \in \mathbb{N}_0$. Denotaremos por Π al conjunto de todas las estrategias del controlador.

Una estrategia (política de control) para el oponente es una sucesión $\gamma = \{\gamma_t\}$ de funciones $\gamma_t : H_t^{\#} \to B$ tal que $b_t = \gamma_t(h_t^{\#}) \in B(x_t, a_t) \quad \forall h_t^{\#} \in H_t^{\#}, t \in \mathbb{N}_0$. El conjunto de todas las estrategias del oponente serán denotadas por Γ.

Definición 1.2.2. *Definimos los conjuntos siguientes*

$$\mathbb{F}_A := \{f : \mathbb{X} \to A | \quad f(x) \in A(x), x \in \mathbb{X}\} \qquad y$$

$$\mathbb{F}_B := \{g : \mathbb{X} \times A \to B | \quad g(x,a) \in B(x,a), (x,a) \in \mathbb{K}_A\}$$

La manera de elegir una acción puede depender de la historia hasta la etapa de decisión actual o en su defecto depender solamente del estado del sistema en dicha etapa. Específicamente tenemos la siguiente definición.

Definición 1.2.3.

(a) *Diremos que una estrategia $\pi = \{\pi_t\}$ para el controlador es de Markov si existe una sucesión de funciones $f_t \in \mathbb{F}_A$ tal que $\pi_t(h_t) = f_t(x_t) \forall t \in \mathbb{N}_0$ y en este caso π toma la forma $\pi = \{f_t\}$ y $a_t = f_t(x_t) \in A(x_t)$, es decir, a_t depende solamente del estado actual del sistema.*

(b) *Si $\pi = \{f_t\}$ es una estrategia de Markov y $f_t = f \forall t \in \mathbb{N}_0$, para alguna función $f \in \mathbb{F}_A$, entonces diremos que π es una estrategia estacionaria.*

Las estrategias de Markov y estacionarias para el oponente se definen de manera similar, solo reemplazando $f_t \in \mathbb{F}_A$ por $g_t \in \mathbb{F}_B$.

Sea $(\Omega, \mathscr{F})$ el espacio medible que consiste del espacio muestral $\Omega := (\mathbb{X} \times A \times B)^\infty$ cuyos elementos son las trayectorias $\omega = (x_0, a_0, b_0, ...)$ con $(x_t, a_t, b_t) \in \mathbb{K}$ y de su σ-álgebra producto $\mathscr{F}$. Para cada par de estrategias $\pi \in \Pi$ y $\gamma \in \Gamma$, y cada estado inicial $x \in \mathbb{X}$, existe una probabilidad denotada por $P_x^{\pi,\gamma}$ definida en $(\Omega, \mathscr{F})$ tal que las variables x_t, a_t, b_t satisfacen(Véase [8]):

$$P_x^{\pi,\gamma}[x_0 = x] = 1$$

$$a_t = \pi_t(h_t), \quad b_t = \gamma_t(h_t^\#)$$

$$P_x^{\pi,\gamma}[x_{t+1} = y | h_t, a_t, b_t] = P_{x_t,y}(a_t, b_t)$$

El operador esperanza con respecto a $P_x^{\pi,\gamma}$ es denotado por $E_x^{\pi,\gamma}$. Además, para una función $v : \mathbb{X} \to \mathbb{R}$,

$$E_x^{\pi,\gamma}[v(x_{t+1})|h_t, a_t, b_t] = \sum_{y\in\mathbb{X}} v(y) P_{x,y}(a_t, b_t), \quad \forall t \in \mathbb{N}_0.$$

1.3. Índices de funcionamiento

Un índice de funcionamiento, el cual define el criterio de optimalidad, es una función que "mide" el comportamiento del sistema al utilizar diferentes políticas de control, dado el estado inicial.

A continuación definimos los índices de funcionamiento más comunes.

Definición 1.3.1. *Sea $\alpha \in (0,1]$ un número positivo dado, $x \in \mathbb{X}$, $\pi \in \Pi$ y $\gamma \in \Gamma$. Definimos:*

- *El Costo α- descontado en n etapas con horizonte finito $n \in \mathbb{N}$ por*

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

  Para $\alpha = 1$ denotaremos $V_{n,1}(x,\pi,\gamma) \equiv J_n(x,\pi,\gamma)$.

- *El Costo $\alpha-$descontado con horizonte infinito y $\alpha \in (0,1)$ como*

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

- *El Costo promedio esperado con horizonte infinito*

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

En este trabajo estudiaremos el Modelo de Control Minimax con índice de costo en n-etapas e índice de costo α-descontado.

1.4. Problema de control minimax

Con los elementos descritos anteriormente podemos ahora plantear el problema de control minimax(PCM).

Dado un MCM $(\mathbb{X}, A, B, \mathbb{K}_A, \mathbb{K}, P, c)$, una familia de políticas de control admisible Π para el controlador y Γ para el oponente, junto con uno de los índices de funcionamiento de la Definición 1.3.1 al cual representaremos por

$$K(x, \pi, \gamma),$$

y denotamos

$$K^{\#}(x, \pi) := \sup_{\gamma \in \Gamma} K(x, \pi, \gamma),$$

el problema de control minimax consiste en encontrar una política $\pi^* \in \Pi$ tal que minimice $K^{\#}(x, \cdot)$. Específicamente tenemos la siguiente definición.

Definición 1.4.1. *Sea $K(x, \pi, \gamma)$ cualquier función de costo en la Definición 1.3.1 y sea*

$$K^{\#}(x, \pi) := \sup_{\gamma \in \Gamma} K(x, \pi, \gamma) \tag{1.4.1}$$

Una estrategia $\pi^* \in \Pi$ (para el controlador) se dice ser una estrategia minimax con respecto al índice K si π^* minimiza $K^{\#}(x, \cdot)$ sobre Π para todo $x \in \mathbb{X}$, ésto es,

$$K^{\#}(x, \pi^*) = \inf_{\pi \in \Pi} K^{\#}(x, \pi) = \inf_{\pi \in \Pi} \sup_{\gamma \in \Gamma} K(x, \pi, \gamma) \quad \forall x \in \mathbb{X}. \tag{1.4.2}$$

Observación 1.4.2. *En terminología de teoría de juegos, la función en el lado derecho de (1.4.2), i.e.,*

$$U(x) := \inf_{\pi \in \Pi} \sup_{\gamma \in \Gamma} K(x, \pi, \gamma)$$

se llama valor superior del juego de suma cero (con respecto a K). Por lo tanto, una estrategia minimax es aquella que alcanza el valor superior del juego. Por otro lado, uno podría considerar el problema "dual" en el cual (1.4.1) es reemplazado con

$$K_{\#}(x, \gamma) := \inf_{\pi \in \Pi} K(x, \pi, \gamma)$$

Entonces el valor inferior del juego (con respecto a K) es

$$L(x) := \sup_{\gamma \in \Gamma} K_{\#}(x, \gamma) = \sup_{\gamma \in \Gamma} \inf_{\pi \in \Pi} K(x, \pi, \gamma)$$

y $\gamma^* \in \Gamma$ *es llamado una estrategia maximin si*

$$K_{\#}(x, \gamma^*) = L(x) \quad \forall x \in \mathbb{X}. \tag{1.4.3}$$

En general $L(\cdot) \leq U(\cdot)$*, y si la igualdad se cumple, entonces a* $L(\cdot) = U(\cdot)$ *se le llama función de valor del juego. Si además* π^* *y* γ^* *satisfacen (1.4.2) y (1.4.3), entonces se dice que el par* (π^*, γ^*) *es un equilibrio no cooperativo.*

1.5. Algoritmo de PD con horizonte finito

Para cerrar este capítulo, a continuación veremos el conocido algoritmo de Programación Dinámica con horizonte finito en su versión minimax, el cual motivará el análisis del caso con horizonte infinito de los siguientes capítulos.

Antes de continuar, tengamos en cuenta que analizaremos el problema con índice de costo total esperado $J_N(x, \pi, \gamma)$:

$$J_N(x, \pi, \gamma) = E_x^{\pi,\gamma} \left[\sum_{n=0}^{N-1} c(x_n, a_n, b_n) + C_N(x_N) \right],$$

al cual le estamos añadiendo un costo terminal dado por $C_N(\cdot)$. El objetivo es encontrar una política $\pi^* \in \Pi$ tal que

$$\sup_{\gamma \in \Gamma} J_N(x, \pi^*, \gamma) = \inf_{\pi \in \Pi} \sup_{\gamma \in \Gamma} J_N(x, \pi, \gamma) = J^*(x).$$

La función J^* es la que conocemos como *función de valor minimax*.

Ahora bien, prosigamos enunciando el teorema del algoritmo de Programación Dinámica.

Teorema 1.5.1. *Sean* $J_0, J_1, ..., J_N$ *funciones en* $\mathbb{X}$ *definidas (hacia atrás de* $t = N$ *a* $t = 0$*) por*

$$J_N(x) := C_N(x)$$

$$J_t(x) := \min_{a \in A(x)} \sup_{b \in B(x,a)} \left\{ c(x,a,b) + \sum_{y \in \mathbb{X}} J_{t+1}(y) P_{x,y}(a,b) \right\}$$

Si para cada $t = N-1, N-2, ..., 0$ *existe* $f_t \in \mathbb{F}_A$ *tal que*

$$J_t(x) = \sup_{b \in B(x, f_t(x))} \left\{ c(x, f_t(x), b) + \sum_{y \in \mathbb{X}} J_{t+1}(y) P_{x,y}(f_t(x), b) \right\} \quad \forall x \in \mathbb{X},$$

entonces $\pi^* = \{f_0, f_1, ..., f_{N-1}\}$ *es una estrategia minimax y*

$$J^*(x) = J_0(x) = \sup_{\gamma \in \Gamma} J_N(x, \pi^*, \gamma).$$

La demostración de este teorema requiere algunos resultados preliminares, los cuales veremos a continuación.

Lema 1.5.2. *Sean* $\epsilon > 0$ *y* $\pi \in \Pi$ *fijos. Para* $\gamma \in \Gamma$ *definimos en* $\mathbb{X}$

$$C_t(x, \pi, \gamma) := E^{\pi,\gamma} \left[\sum_{n=t}^{N-1} c(x_n, a_n, b_n) + C_N(x_N) | x_t = x \right] \quad y$$

$$C_N(x, \pi, \gamma) := E^{\pi,\gamma} \left[C_N(x_N) | x_N = x \right] = C_N(x)$$

entonces existe $\widetilde{\gamma} = (\widetilde{\gamma}_0, ..., \widetilde{\gamma}_{N-1})$ *tal que*

$$\forall x \quad C_{N-1}(x, \pi, \widetilde{\gamma}_{N-1}) + \epsilon > C^{\#}_{N-1}(x, \pi) \quad y \tag{1.5.1}$$

$$\forall x \quad C_0(x, \pi, \widetilde{\gamma}_0, ..., \widetilde{\gamma}_{N-1}) + N\epsilon > C^{\#}_0(x, \pi) \tag{1.5.2}$$

donde

$$C^{\#}_t(x, \pi) = \sup_{\gamma \in \Gamma} C_t(x, \pi, \gamma)$$

Demostración.

La demostración sera utilizando inducción hacia atrás.

(1.5.1) se sigue de la definición de supremo. Ademas, existe $\widetilde{\gamma}_{N-2}$ tal que

$$C_{N-2}(x,\pi,\widetilde{\gamma}_{N-2},\widetilde{\gamma}_{N-1})+\epsilon > \sup_{\gamma_{N-2}} C_{N-2}(x,\pi,\gamma_{N-2},\widetilde{\gamma}_{N-1}).$$

Entonces

$$\begin{aligned}
C_{N-2}(x,\pi,\widetilde{\gamma}_{N-2},\widetilde{\gamma}_{N-1})+2\epsilon &> \sup_{\gamma_{N-2}} C_{N-2}(x,\pi,\gamma_{N-2},\widetilde{\gamma}_{N-1})+\epsilon \\
&= \sup_{\gamma_{N-2}} \left[c(x,\pi_{N-2},\gamma_{N-2}) + \sum_{y\in\mathbb{X}} C_{N-1}(y,\pi,\widetilde{\gamma}_{N-1}) P_{x,y}(\pi,\gamma_{N-2}) \right] + \epsilon \\
&= \sup_{\gamma_{N-2}} c(x,\pi_{N-2},\gamma_{N-2}) + \sum_{y\in\mathbb{X}} (C_{N-1}(y,\pi,\widetilde{\gamma}_{N-1})+\epsilon) P_{x,y}(\pi,\gamma_{N-2}) \\
&\geq \sup_{\gamma_{N-2}} c(x,\pi_{N-2},\gamma_{N-2}) + \sum_{y\in\mathbb{X}} C^{\#}_{N-1}(y,\pi) P_{x,y}(\pi,\gamma_{N-2}).
\end{aligned}$$

Tenemos que

$$C^{\#}_{N-2}(y,\pi) \geq C_{N-1}(y,\pi,\gamma_{N-1}) \quad \forall \gamma_{N-1}$$

entonces

$$\begin{aligned}
C_{N-2}(x,\pi,\widetilde{\gamma}_{N-2},\widetilde{\gamma}_{N-1})+2\epsilon &\geq \sup_{\gamma_{N-2}} c(x,\pi_{N-2},\gamma_{N-2}) + \sup_{\gamma_{N-1}} \sum_{y\in\mathbb{X}} C_{N-1}(y,\pi,\gamma_{N-1}) P_{x,y}(\pi,\gamma_{N-2}) \\
&= \sup_{\gamma_{N-1},\gamma_{N-2}} C_{N-2}(x,\pi,\gamma_{N-2},\gamma_{N-1}) \\
&= C^{\#}_{N-2}(x,\pi).
\end{aligned}$$

Similarmente

$$\begin{aligned}
C_{N-(k-1)}(\cdot)+(k-1)\epsilon &> C^{\#}_{N-(k-1)}(\cdot) \quad \text{implica} \\
C_{N-k}(\cdot)+k\epsilon &> C^{\#}_{N-k}(\cdot).
\end{aligned}$$

■

Corolario 1.5.3. *Sea $\pi \in \Pi$. Entonces existe una sucesión $\gamma^{(j)} \subseteq \Gamma$ tal que*

$$C_t(x,\pi,\gamma^(j)) > C^{\#}_t(x,\pi) - \frac{N}{j} \quad \forall x \in \mathbb{X}$$

para todo $t = 0,1,...,N-1$.

A continuación demostramos el Teorema 1.5.1.

Demostración. (**Teorema 1.5.1**)

Sea π arbitraria. Para $\gamma \in \Gamma$ definimos

$$C_t(x,\pi,\gamma) := E^{\pi,\gamma}\left[\sum_{n=t}^{N-1} c(x_n,a_n,b_n) + C_N(x_N)|x_t = x\right]$$

y

$$C_N(x,\pi,\gamma) := E^{\pi,\gamma}\left[C_N(x_N)|x_N = x\right] = C_N(x).$$

En particular, notemos que

$$J_N(x,\pi,\gamma) = C_0(x,\pi,\gamma),$$

pues,

$$\begin{aligned} C_0(x,\pi,\gamma) &= E^{\pi,\gamma}\left[\sum_{n=0}^{N-1} c(x_n,a_n,b_n) + C_N(x_N)|x_0 = x\right] \\ &= E_x^{\pi,\gamma}\left[\sum_{n=0}^{N-1} c(x_n,a_n,b_n) + C_N(x_N)\right] \\ &= J_N(x,\pi,\gamma). \end{aligned}$$

Ahora bien, definamos

$$C_t^{\#}(x,\pi) = \sup_{\gamma\in\Gamma} C_t(x,\pi,\gamma).$$

Mostraremos que para todo $t = 0,1,...,N$,

$$C_t^{\#}(x,\pi) \geq J_t(x), \tag{1.5.3}$$

y en el caso que $\pi = \pi^*$ se tiene que $C_t^{\#}(x,\pi^*) = J_t(x)$. Esto es suficiente para la demostración debido a que para $t = 0$ tendríamos

$$J_0(x) \leq C_0^{\#}(x,\pi) \quad \text{y} \quad J_0(x) = C_0^{\#}(x,\pi^*).$$

Entonces,

$$\begin{aligned} J_0(x) &= \sup_{\gamma \in \Gamma} C_0(x, \pi^*, \gamma) \\ &= \sup_{\gamma \in \Gamma} J_N(x, \pi^*, \gamma). \end{aligned} \tag{1.5.4}$$

Por otro lado

$$J_0(x) \leq C_0^{\#}(x, \pi) = \sup_{\gamma \in \Gamma} C_0(x, \pi, \gamma),$$

es decir

$$J_0(x) \leq \sup_{\gamma \in \Gamma} C_0(x, \pi, \gamma) = \sup_{\gamma \in \Gamma} J_N(x, \pi, \gamma) \quad \forall \pi \in \Pi.$$

En particular

$$J_0(x) \leq \inf_{\pi \in \Pi} \sup_{\gamma \in \Gamma} J_N(x, \pi, \gamma) = J^*(x) \tag{1.5.5}$$

Por lo que de (1.5.4) y (1.5.5) se sigue que

$$J_0(x) = J^*(x) = \sup_{\gamma \in \Gamma} J_N(x, \pi^*, \gamma).$$

Ahora bien, para la demostración de (1.5.3) utilizaremos inducción matemática.

- El caso $t = N$ se sigue por definición, pues

$$C_N^{\#}(x, \pi) = C_N(x) = J_N(x).$$

- Supongamos que (1.5.3) se cumple para $t = k + 1$, *i.e.*,

$$C_{k+1}^{\#}(x, \pi) \geq J_{k+1}(x) \quad \text{con igualdad si} \quad \pi = \pi^*. \tag{1.5.6}$$

  Notemos que

$$\begin{aligned} C_k^{\#}(x, \pi) &= \sup_{\gamma_k, \dots, \gamma_{N-1}} C_k(x, \pi, \gamma_k, \dots, \gamma_{N-1}) \\ &= \sup_{\gamma_k} \left\{ c(x, \pi_k, \gamma_k) + \sup_{\gamma_{k+1}, \dots, \gamma_{N-1}} \sum_{y \in \mathbb{X}} C_{k+1}(y, \pi, \gamma_{k+1}, \dots, \gamma_{N-1}) P_{x,y}(\pi_k, \gamma_k) \right\} \end{aligned} \tag{1.5.7}$$

Por otro lado, sea $j \in \mathbb{N}$ y $\{\gamma^{(j)}\} \subseteq \Gamma$ como en el Corolario 1.5.3, entonces

$$\begin{aligned}
\sup_{\gamma_{k+1},..,\gamma_{N-1}} \sum_{y\in\mathbb{X}} C_{k+1}(y,\pi,\gamma_{k+1},...,\gamma_{N-1})P_{x,y}(\pi_k,\gamma_k) &\geq \sum_{yin\mathbb{X}} C_{k+1}(y,\pi,\gamma^{(j)})P_{x,y}(\pi_k,\gamma_k) \\
&\geq \sum_{y\in\mathbb{X}} (C^{\#}_{k+1}(y,\pi) - \frac{N}{j})P_{x,y}(\pi_k,\gamma_k) \\
&\geq \sum_{y\in\mathbb{X}} J_{k+1}(y)P_{x,y}(\pi_k,\gamma_k) - \frac{N}{j}.
\end{aligned}$$

Luego

$$\begin{aligned}
C^{\#}_k(x,\pi) &\geq \sup_{\gamma_k}\left\{c(x,\pi_k,\gamma_k) + \sum_{y\in\mathbb{X}} J_{k+1}(y)P_{x,y}(\pi_k,\gamma_k)\right\} - \frac{N}{j} \\
&= \sup_{b\in B(x,a)}\left\{c(x,\pi_k,b) + \sum_{y\in\mathbb{X}} J_{k+1}(y)P_{x,y}(\pi_k,b)\right\} - \frac{N}{j} \\
&\geq \min_{a\in A(x)} \sup_{b\in B(x,a)}\left\{c(x,a,b) + \sum_{y\in\mathbb{X}} J_{k+1}(y)P_{x,y}(a,b)\right\} - \frac{N}{j} \\
&= J_k(x) - \frac{N}{j}.
\end{aligned}$$

Haciendo $j \to \infty$, obtenemos $C^{\#}_k(x,\pi) \geq J_k(x)$. Si $\pi = \pi^*$, entonces por (1.5.7) y (1.5.6)

$$\begin{aligned}
C^{\#}_k(x,\pi^*) &\leq \sup_{\gamma_k}\left\{c(x,\pi^*_k,\gamma_k) + \sum_{y\in\mathbb{X}} C^{\#}_{k+1}(y,\pi^*)P_{x,y}(\pi^*_k,\gamma_k)\right\} \\
&= \sup_{b\in(x,f_k(x))}\left\{c(x,f_k,b) + \sum_{y\in\mathbb{X}} J_{k+1}(y)P_{x,y}(f_k,b)\right\} \\
&= J_k(x).
\end{aligned}$$

Por lo tanto $C^{\#}_k(x,\pi^*) = J_k(x)$.

■

Observación 1.5.4. *Es importante observar que el algoritmo anterior es válido cuando los costos son no estacionarios, es decir, en cada etapa inducimos un costo* $c_t(x,a,b)$. *La demostración es prácticamente la misma.*

Capítulo 2

Criterios de costo descontado

2.1. Introducción

En este capítulo analizaremos problemas de control minimax bajo el criterio de costo descontado, introducidos en la Definición 1.3.1, considerando el problema con horizonte finito y con horizonte de planeación infinito.

Para el análisis del problema con horizonte de planeación finito utilizaremos el algoritmo de Programación Dinámica, mismo que se abordó en el Capítulo 1, para lo cual será necesario realizar algunos cambios. Para el caso de costo descontado, analizaremos el problema por medio de la ecuación de optimalidad minimax, donde demostraremos que la función de valor óptimo es la única solución.

Consideraremos primero el problema con índice de costo en n-etapas $V_{n,\alpha}(x,\pi,\gamma)$, que consiste en encontrar una política $\pi^* \in \Pi$ tal que

$$\sup_{\gamma\in\Gamma} V_{n,\alpha}(x,\pi^*,\gamma) = \inf_{\pi\in\Pi}\sup_{\gamma\in\Gamma} V_{n,\alpha}(x,\pi,\gamma) =: V^*_{n,\alpha}(x).$$

A la función $V^*_{n,\alpha}$ se le llama *función de valor minimax en n-etapas.*

De manera similar se tiene el Problema de Control Minimax para el caso descontado. Se considera el índice de costo α-descontado con horizonte infinito $V_\alpha(x,\pi,\gamma)$, y similarmente, el problema consiste en encontrar una política $\pi^* \in \Pi$ tal que

$$\sup_{\gamma\in\Gamma} V_\alpha(x,\pi^*,\gamma) = \inf_{\pi\in\Pi}\sup_{\gamma\in\Gamma} V_\alpha(x,\pi,\gamma) =: V_\alpha^*(x).$$

Ahora bien, el estudio de estos problemas se realizará asumiendo costo por etapa posiblemente no acotado dentro del contexto de normas ponderadas. Con tal objetivo supondremos que se cumplen las siguientes condiciones:

Hipótesis 2.1

(a) Existe una constante $\bar{c} \geq 0$ y una función $w(\cdot) \geq 1$ en $\mathbb{X}$ tal que

$$|c(x,a,b)| \leq \bar{c}w(x) \quad \forall (x,a,b) \in \mathbb{K}.$$

(b) Existe una constante $\beta > 0$ tal que

$$\widehat{w}(x,a,b) \leq \beta w(x) \quad \forall (x,a,b) \in \mathbb{K}$$

donde

$$\widehat{w}(x,a,b) := \sum_{y\in\mathbb{X}} w(y) P_{x,y}(a,b).$$

(c) Para cada $(x,a) \in \mathbb{K}_A$, $B(x,a)$ es σ-compacto (Vease Apéndice B, Definición B.2.4).

Observación 2.1.1. *Si c es acotada, entonces la función $w(\cdot)$ puede ser tomada como una constante, por ejemplo, $w \equiv 1$, entonces la Hipóteis 2.1(b) puede ser omitida.*

2.2. Problema minimax con horizonte finito

Sea $w(\cdot) \geq 1$ como en la Hipótesis (2.1). Denotaremos por $\mathbb{B}_w(\mathbb{X})$ al espacio lineal normado (Véase Apéndice Teorema A.0.9) de las funciones u en $\mathbb{X}$ que tienen w-norma finita $\|u\|_w < \infty$, definida como

$$\|u\|_w := \sup_{x\in\mathbb{X}} \frac{|u(x)|}{w(x)}. \tag{2.2.1}$$

Para cada $u \in \mathbb{B}_w(\mathbb{X})$, $0 < \alpha \leq 1$, y $(x, a, b) \in \mathbb{K}$, sea

$$H_\alpha(u; x, a, b) := c(x, a, b) + \alpha \sum_{y \in \mathbb{X}} u(y) P_{x,y}(a, b) \tag{2.2.2}$$

$$H_\alpha^{\#}(u; x, a) := \sup_{b \in B(x,a)} H_\alpha(u; x, a, b) \tag{2.2.3}$$

$$T_\alpha u(x) := \inf_{a \in A(x)} H_\alpha^{\#}(u; x, a) \tag{2.2.4}$$

Es decir,

$$T_\alpha u(x) = \inf_{a \in A(x)} \sup_{b \in B(x,a)} \left[c(x, a, b) + \alpha \sum_{y \in \mathbb{X}} u(y) P_{x,y}(a, b) \right].$$

El operador T_α se conoce generalmente como *operador de programación dinámica.*

Obtendremos la solución al problema de control minimax mediante el algoritmo de programación dinámica, estudiado en el capítulo anterior, pero adaptado al presente contexto, tal como lo muestra el siguiente resultado.

Teorema 2.2.1. *Sea $\alpha > 0$ fijo y definamos la siguiente sucesión de funciones*

$$v_{0,\alpha} = 0$$

$$v_{n,\alpha} := T_\alpha v_{n-1,\alpha} = T_\alpha^n v_{0,\alpha} \quad \forall n = 1, 2, ... \tag{2.2.5}$$

Entonces para cada $n \in \mathbb{N}$

(a) $v_{n,\alpha} \in \mathbb{B}_w(\mathbb{X})$.

(b) *Existe $f_n \in \mathbb{F}_A$ tal que*

$$v_{n,\alpha}(x) = H_\alpha^{\#}(v_{n-1,\alpha}; x, f_n(x)) \quad \forall x \in \mathbb{X},$$

es decir,

$$v_{n,\alpha}(x) = \sup_{b \in B(x, f_n(x))} \left[c(x, f_n(x), b) + \alpha \sum_{y \in \mathbb{X}} v_{n-1,\alpha}(y) P_{x,y}(f_n(x), b) \right],$$

(c) *$v_{n,\alpha}$ es el costo óptimo en n-etapas, esto es,*

$$v_{n,\alpha}(x) = \inf_{\pi\in\Pi} \sup_{\gamma\in\Gamma} V_{n,\alpha}(x,\pi,\gamma) \quad \forall x \in \mathbb{X}.$$

(d) *La política de Markov $\pi^n := \{f_n, f_{n-1}, ..., f_1\}$ es una estrategia minimax para el problema de n-etapas; esto es,*

$$v_{n,\alpha}(x) = \sup_{\gamma\in\Gamma} V_{n,\alpha}(x,\pi^n,\gamma) \quad \forall x \in \mathbb{X}.$$

Antes de demostrar el Teorema 2.2.1 introduciremos primero algunos resultados preliminares. Observemos en particular que el siguiente lema nos da la existencia de estrategias minimax para una problema minimax de una etapa con función de costo $v(x,a,b)$.

Lema 2.2.2. *Sea $v : \mathbb{K} \to \mathbb{R}$ una función tal que*

$$|v(x,a,b)| \leq \overline{v}w(x) \quad \forall (x,a,b) \in \mathbb{K} \tag{2.2.6}$$

para alguna constante $\overline{v} \geq 0$. Sea

$$v^{\#}(x,a) := \sup_{b\in B(x,a)} v(x,a,b) \quad y \quad v^*(x) := \inf_{a\in A(x)} v^{\#}(x,a).$$

Entonces

(a) $|v^{\#}(x,a)| \leq \overline{v}w(x) \quad \forall (x,a) \in \mathbb{K}_A$;

(b) *$v^* \in \mathbb{B}_w(\mathbb{X})$ y además existe $f \in \mathbb{F}_A$ tal que*

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

y por lo tanto,

$$v^*(x) = \sup_{b\in B(x,f(x))} v(x,f(x),b) \quad \forall x \in \mathbb{X}.$$

Demostración.

(a) La desigualdad se sigue claramente de (2.2.6).

(b) De (a) tenemos que

$$|v^{\#}(x,a)| \leq \overline{v}w(x) \quad \forall (x,a) \in \mathbb{K}_A,$$

y por lo tanto, tenemos que

$$|v^*(\cdot)| \leq \overline{v}w(\cdot).$$

De aquí podemos concluir que $v^* \in \mathbb{B}_w(\mathbb{X})$. Ahora, recordemos que $A(x)$ es finito, por lo que podemos asegurar que existe $f \in \mathbb{F}_A$ tal que minimice el lado derecho

$$v^*(x) = \min_{a \in A(x)} v^{\#}(x,a),$$

i.e.,

$$v^*(x) = v^{\#}(x, f(x)) \quad \forall x \in \mathbb{X}.$$

■

En (a) del siguiente lema reemplazaremos la función v de (2.2.6) por la función $H_\alpha(u;\cdot)$ de (2.2.2), con $u \in \mathbb{B}_w(\mathbb{X})$. De aquí tenemos, que el operador T_α mapea $\mathbb{B}_w(\mathbb{X})$ en sí mismo. Específicamente se obtiene el siguiente resultado.

Lema 2.2.3. *Sea u una función arbitraria en $\mathbb{B}_w(\mathbb{X})$, y sea $T_\alpha u$ como en (2.2.4). Entonces:*

(a) $|H_\alpha(u;x,a,b)| \leq [\overline{c} + \alpha\beta\|u\|_w]w(x), \quad \forall (x,a,b) \in \mathbb{K},$

(b) $T_\alpha u \in \mathbb{B}_w(\mathbb{X})$ *y más aún, existe $f \in \mathbb{F}_A$ tal que, para toda $x \in \mathbb{X}$*

$$T_\alpha u(x) = \sup_{b \in B(x,f(x))} \left[c(x,f(x),b) + \alpha \sum_{y \in \mathbb{X}} u(y) P_{x,y}(f(x),b) \right].$$

Demostración.

(a) De (2.2.1) y la Hipótesis 2.1(b), tenemos que

$$\sum_{y \in \mathbb{X}} |u(y)| P_{x,y}(a,b) \leq \|u\|_w \sum_{y \in \mathbb{X}} w(y) P_{x,y}(a,b) \leq \|u\|_w \beta w(x)$$

para todo $(x,a,b) \in \mathbb{K}$. De la desigualdad anterior junto con la Hipótesis 2.1(a) y (2.2.2) obtenemos

$$\begin{aligned}
|H_\alpha(u;x,a,b)| &= |c(x,a,b) + \alpha \sum_{y \in \mathbb{X}} u(y) P_{x,y}(a,b)| \\
&\leq |c(x,a,b)| + \alpha \sum_{y \in \mathbb{X}} |u(y)| P_{x,y}(a,b) \\
&\leq \bar{c} w(x) + \alpha \|u\|_w \beta w(x) \\
&\leq [\bar{c} + \alpha \|u\|_w \beta] w(x).
\end{aligned}$$

(b) Se sigue de (a) y del Lema 2.2.2(b) con $v(\cdot) = H_\alpha(u;\cdot)$. En efecto, de (a) tenemos que $H_\alpha(u;\cdot)$ cumple la hipótesis del Lema 2.2.2 ,por lo tanto, se tiene que $v^* \in \mathbb{B}_w(\mathbb{X})$, donde en este caso, tenemos que v^* toma la forma:

$$v^*(x) = \inf_{a \in A(x)} \sup_{b \in B(x,a)} H_\alpha(u;x,a,b) = T_\alpha u(x),$$

es decir, $T_\alpha u \in \mathbb{B}_w(\mathbb{X})$. Del Lema 2.2.2(b) se observa también que existe $f \in \mathbb{F}_A$ tal que

$$\begin{aligned}
T_\alpha u(x) &= \sup_{b \in B(x,f(x))} H_\alpha(u;x,a,b) \quad \forall x \in \mathbb{X}. \\
&= \sup_{b \in B(x,f(x))} \left[c(x,a,b) + \alpha \sum_{y \in \mathbb{X}} u(y) P_{x,y}(a,b) \right].
\end{aligned}$$

∎

Tomando en cuenta los lemas anteriores, a continuación demostramos el Teorema 2.2.1.

Demostración. (Teorema 2.2.1)

(a) Procedemos por inducción. Como $v_{0,\alpha}(\cdot) \equiv 0$ entonces (a) se sigue trivialmente para $n = 0$.

Supongamos ahora que $v_{n-1,\alpha} \in \mathbb{B}_w(\mathbb{X})$ para alguna $n \geq 1$. Ahora, debemos mostrar que $v_{n,\alpha} \in \mathbb{B}_w(\mathbb{X})$. Por definición tenemos que

$$v_{n,\alpha} = T_\alpha v_{n-1,\alpha},$$

de la hipótesis de inducción tenemos que $v_{n-1,\alpha} \in \mathbb{B}_w(\mathbb{X})$, entonces del Lema 2.2.3 tenemos que $T_\alpha v_{n-1,\alpha} \in \mathbb{B}_w(\mathbb{X})$.

(b) Se sigue inmediatamente de (a) y Lema 2.2.3(b).

(c) y **(d)** Se siguen de (b) y el algoritmo de programación dinámica descrito en el Teorema 1.5.1. En efecto, cuando el costo esperado es de la forma

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

con función de costo

$$c_t(x, a, b) := \alpha^t c(x, a, b)$$

las ecuaciones de PD serían:

$$J_N(x) := \alpha^N C_N(x),$$

$$J_t(x) := \min_{a \in A(x)} \sup_{b \in B(x,a)} \left\{\alpha^t c(x, a, b) + \sum_{y \in \mathbb{X}} J_{t+1}(y) P_{x,y}(a, b)\right\}.$$

Reescribiendo estas funciones en términos de las funciones $v_{t,\alpha} := \alpha^{-(N-t)} J_{N-t}$, $t = 0, ..N$ obtenemos

$$v_{0,\alpha}(x) := C_N(x)$$

$$v_{t,\alpha}(x) := \min_{a \in A(x)} \sup_{b \in B(x,a)} \left\{c(x, a, b) + \alpha \sum_{y \in \mathbb{X}} v_{t-1}(y) P_{x,y}(a, b)\right\},$$

las cuales cumplen las conclusiones del Teorema 1.5.1, es decir, los puntos (c) y (d).

■

2.3. Ecuación de Optimalidad Minimax

Ahora estudiaremos el problema usando el índice en costo α- descontado con horizonte infinito, para lo cual, recordemos que el índice de funcionamiento en este caso está definido como

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

y la función de valor minimax como

$$V_\alpha^*(x) := \inf_{\pi\in\Pi}\sup_{\gamma\in\Gamma} V_\alpha(x,\pi,\gamma).$$

Antes de establecer el resultado principal, veamos el teorema siguiente.

Teorema 2.3.1. *Supongamos que se cumple la Hipótesis 2.1, y que además la constante β en (b) satisface que*

$$1 \le \beta < \frac{1}{\alpha}. \tag{2.3.1}$$

Entonces existe una función $v^* : \mathbb{X} \to \mathbb{R}$ que es la única en $\mathbb{B}_w(\mathbb{X})$ que satisface $v^* = T_\alpha v^*$. Más aún, se satisface

$$\lim_{n\to\infty} \alpha^n E_x^{\pi,\gamma}[v^*(x_n)] = 0 \quad \forall \pi\in\Pi, \gamma\in\Gamma, x\in\mathbb{X},$$

$$y \quad \|v_{n,\alpha} - v^*\|_w \le \bar{c}\frac{(\alpha\beta)^n}{1-\alpha\beta} \quad \forall n\ge 0,$$

con $\bar{c}$ y $v_{n,\alpha}$ como en la Hipótesis 2.1(a) y (2.2.5), respectivamente y con $v_{0,\alpha}\equiv 0$.

Para demostrar el teorema previo, primero estableceremos algunos resultados generales.

Lema 2.3.2. *Sea T un operador que mapea $\mathbb{B}_w(\mathbb{X})$ en si mismo. Supongamos que*

(i) *T es monótono, i.e., $u \leq u'$ implica $Tu \leq Tu'$.*

(ii) *Existe una constante $0 < \tau < 1$ tal que*

$$T(u + rw) \leq Tu + \tau rw \quad \forall u \in \mathbb{B}_w(\mathbb{X}) \quad y \quad r \geq 0.$$

Entonces T es un mapeo de contracción módulo τ, es decir,

$$\|Tu - Tu'\|_w \leq \tau \|u - u'\|_w \quad \forall u, u' \in \mathbb{B}_w(\mathbb{X}),$$

y por lo tanto, existe una única función $u^* \in \mathbb{B}_w(\mathbb{X})$ que satisface

$$(a) u^* = Tu^* \quad y \quad (b) \|T^n u - u^*\|_w \leq \tau^n \|u - u^*\|_w \quad \forall u \in \mathbb{B}_w(\mathbb{X}), n \geq 0. \tag{2.3.2}$$

Además, si existe una constante $k \geq 0$ tal que

$$\|Tu\|_w \leq k + \tau \|u\|_w \quad \forall u \in \mathbb{B}_w(\mathbb{X}),$$

entonces

$$\|u^*\|_w \leq \frac{k}{1 - \tau}.$$

Demostración.

Por (2.2.1), se tiene que para cada par $u, u' \in \mathbb{B}_w(\mathbb{X})$

$$u \leq u' + \|u - u'\|_w w.$$

Por lo tanto, de (i) y (ii) con $r := \|u - u'\|_w$ se deduce que

$$\begin{aligned} Tu &\leq T(u' + \|u - u'\|_w w) \\ &\leq Tu' + \tau \|u - u'\|_w w. \end{aligned}$$

Entonces

$$Tu - Tu' \leq \tau \|u - u'\|_w w.$$

Similarmente

$$Tu - Tu' \geq -\tau \|u - u'\|_w w$$

. Por consiguiente

$$\|Tu - Tu'\|_w \leq \tau \|u - u'\|_w \quad \forall u, u' \in \mathbb{B}_w(\mathbb{X}).$$

Para finalizar, nótese que (2.3.2) se sigue del Teorema del Punto fijo de Banach (Véase apéndice A, Teorema A.0.11), y entonces

$$\|u^*\|_w = \|Tu^*\|_w \leq k + \tau \|u^*\|_w$$

de aquí,

$$\|u^*\|_w - \tau \|u^*\|_w \leq k.$$

Factorizando $\|u^*\|_w$ obtenemos que

$$\|u^*\|_w [1 - \tau] \leq k,$$

es decir

$$\|u^*\|_w \leq \frac{k}{1 - \tau}.$$

■

Antes de pasar a la demostración del Teorema 2.3.1, veremos que para todo $u \in \mathbb{B}_w(\mathbb{X}), \pi \in \Pi, \gamma \in \Gamma$ y $x \in \mathbb{X}$,

$$\lim_{n \to \infty} \alpha^n E_x^{\pi,\gamma} [u(x_n)] = 0. \tag{2.3.3}$$

En efecto, usando la Hipótesis 2.1 (b) tendremos que

$$\begin{aligned} E_x^{\pi,\gamma} [w(x_{n+1}) | h_t, a_t] &= \sum_{y \in \mathbb{X}} w(y) P_{x_n, y}(a_n, b_n) \\ &\leq \beta [w(x_n)], \end{aligned}$$

de modo que, calculando esperanza $E_x^{\pi,\gamma}$ en ambos lados

$$E_x^{\pi,\gamma}[w(x_{n+1})] \leq \beta E_x^{\pi,\gamma}[w(x_n)] \quad \forall n \in \mathbb{N}_0.$$

Luego, iterando esta desigualdad obtenemos

$$\begin{aligned} E_x^{\pi,\gamma}[w(x_{n+1})] &\leq \beta E_x^{\pi,\gamma}[w(x_n)] \\ &\leq \beta[\beta E_x^{\pi,\gamma}[w(x_{n-1})]] \\ &= \beta^2 E_x^{\pi,\gamma}[w(x_{n-1})] \\ &\leq \beta^3 E_x^{\pi,\gamma}[w(x_{n-2})] \\ &\vdots \\ &\leq \beta^{n+1} w(x), \end{aligned}$$

esto es, para toda $n \in \mathbb{N}_0$ por lo que se satisface

$$E_x^{\pi,\gamma}[w(x_n)] \leq \beta^n w(x).$$

Por lo tanto, de (2.3.1)

$$\lim_{n\to\infty} \alpha^n E_x^{\pi,\gamma}[w(x_n)] = 0.$$

Esto implica (2.3.3), ya que de (2.2.1) se tiene que

$$E_x^{\pi,\gamma}[u(x_n)] \leq \|u\|_w E_x^{\pi,\gamma}[(w(x_n)] \quad \forall u \in \mathbb{B}_w(\mathbb{X}).$$

Teniendo esto en consideración prosigamos con la demostración del Teorema 2.3.1.

Demostración. (Teorema 2.3.1)

Primero usaremos el lema anterior para demostrar que T_α es un mapeo de contracción en $\mathbb{B}_w(\mathbb{X})$ módulo $\tau := \alpha\beta < 1$, esto es

$$\|T_\alpha u - T_\alpha u'\|_w \leq \tau\|u - u'\|_w \quad \forall u, u' \in \mathbb{B}_w(\mathbb{X}). \tag{2.3.4}$$

Por el Lema 2.2.3, T_α mapea $\mathbb{B}_w(\mathbb{X})$ en si mismo. Por otro lado, es evidente que T_α

es monótono. Entonces, para obtener (2.3.4) es suficiente verificar que T_α satisface (ii) del lema anterior. Para esto, nótese que de la Hipótesis 2.1(b),

$$\sup_{b \in B(x,a)} \sum_{y \in \mathbb{X}} w(y) P_{x,y}(a,b) \leq \beta w(x) \quad \forall (x,a) \in \mathbb{K}_A.$$

Esta desigualdad y (2.2.4) implican (ii) para $T = T_\alpha$ y $\tau := \alpha\beta$. En efecto,

$$\begin{aligned} T_\alpha(u + rw) &= \inf_{a \in A(x)} \sup_{b \in B(x,a)} \left[c(x,a,b) + \alpha \sum_{y \in \mathbb{X}} (u + rw)(y) P_{x,y}(a,b) \right] \\ &= \inf_{a \in A(x)} \sup_{b \in B(x,a)} \left[c(x,a,b) + \alpha \sum_{y \in \mathbb{X}} u(y) P_{x,y}(a,b) + \alpha \sum_{y \in \mathbb{X}} rw(y) P_{x,y}(a,b) \right] \\ &\leq \inf_{a \in A(x)} \left[\sup_{b \in B(x,a)} [c(x,a,b) + \alpha \sum_{y \in \mathbb{X}} u(y) P_{x,y}(a,b)] + \alpha\beta rw(x) \right] \\ &= T_\alpha u + \tau rw. \end{aligned}$$

Entonces se cumple (2.3.4). Por lo tanto, existe una única función $v^* \in \mathbb{B}_w(\mathbb{X})$ tal que satisface $v^* = T_\alpha v^*$ y más aún, (tomando $u = v_{0,\alpha}$ en (2.3.2)(b))

$$\|v_{n,\alpha} - v^*\|_w \leq \tau^n \|v^*\|_w \quad \forall n \geq 0, \tau = \alpha\beta. \tag{2.3.5}$$

Finalmente, observe que el Lema 2.2.3(a) nos da

$$\|T_\alpha u\|_w \leq \overline{c} + \tau \|u\|_w \quad \forall u \in \mathbb{B}_w(\mathbb{X}).$$

En particular, si $u = v^*$, entonces

$$\|v^*\|_w \leq \frac{\overline{c}}{1 - \tau}.$$

Ahora, por la desigualdad anterior y (2.3.5):

$$\begin{aligned} \|v_{n,\alpha} - v^*\|_w &\leq \tau^n \|v^*\|_w \\ &\leq \frac{\tau^n \overline{c}}{1 - \tau}. \end{aligned}$$

Por lo tanto, como v^* satisface que

$$\lim_{n\to\infty} \alpha^n E_\alpha^{\pi,\gamma}[v^*(x_n)] = 0 \quad \forall \pi \in \Pi, \gamma \in \Gamma, x \in \mathbb{X},$$

la prueba está completa. ∎

Para finalizar con este capítulo, veamos el siguiente teorema, ya que nos proporciona la solución a nuestro problema de control minimax.

Teorema 2.3.3.

(a) *La función de valor minimax V_α^* es la única función en $\mathbb{B}_w(\mathbb{X})$ que satisface*

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

(b) *Para cada $n \in \mathbb{N}$,*

$$\|v_{n,\alpha} - V_\alpha^*\|_w \le \frac{\bar{c}(\alpha\beta)^n}{1-\alpha\beta}.$$

(c) *Existe $f^* \in \mathbb{F}_A$ tal que*

$$V_\alpha^*(x) = \sup_{b \in B(x,f^*(x))} \left[c(x,f^*,b) + \alpha \sum_{y\in\mathbb{X}} V_\alpha^*(y) P_{x,y}(f^*,b) \right].$$

Además $\pi^* = \{f^*\}$ es minimax, i.e.

$$V_\alpha^*(x) = \sup_{\gamma\in\Gamma} V_\alpha(x,\pi^*,\gamma).$$

Demostración.

(a) y **(b)** Primeramente nótese que,

$$v_{n,\alpha}(x) = T_\alpha v_{n-1,\alpha}(x) = T_\alpha^n v_{0,\alpha}, \quad x \in \mathbb{X}, n \in \mathbb{N}.$$

Por otro lado del Teorema 2.3.1, sabemos que existe $v^* \in \mathbb{B}_w(\mathbb{X})$ tal que

$$v^*(x) = T_\alpha v^*(x) \quad \text{y} \quad \|v_{n,\alpha} - v^*\|_w \le \frac{\bar{c}(\alpha\beta)^n}{1-\alpha\beta}.$$

Entonces lo que debemos mostrar es que $v^* = V_\alpha^*$. En tal situación, sea $f \in \mathbb{F}_A$ tal que

$$\begin{aligned} v^*(x) &= \inf_{a\in A(x)} \sup_{b\in B(x,a)} \left[c(x,a,b) + \alpha \sum_{y\in\mathbb{X}} v^*(y) P_{x,y}(a,b) \right] \\ &= \sup_{b\in B(x,f(x))} \left[c(x,f,b) + \alpha \sum_{y\in\mathbb{X}} v^*(y) P_{x,y}(f,b) \right], \end{aligned}$$

la cual existe debido a que $A(x)$ es finito. Entonces, para todo $b \in B(x, f(x))$,

$$v^*(x) \geq c(x,f,b) + \alpha \sum_{y\in\mathbb{X}} v^*(y) P_{x,y}(f,b). \tag{2.3.6}$$

Sea $\gamma \in \Gamma$ una estrategia arbitraria y consideremos la estrategia estacionaria $\pi = \{f\}$. Entonces iterando (2.3.6) se obtiene

$$v^*(x) \geq E_x^{f,\gamma} \left[\sum_{k=0}^{n-1} \alpha^k c(x_k, f, b_k) \right] + \alpha^n E_x^{\pi,\gamma} v^*(x_n). \tag{2.3.7}$$

Ahora, tomando límite cuando $n \to \infty$ en (2.3.7) se observa que

$$\begin{aligned} v^*(x) &\geq E_x^{f,\gamma} \left[\sum_{k=0}^{\infty} \alpha^k c(x_k, f, b_k) \right] \\ &= V_\alpha(x,\pi,\gamma) \quad \forall x \in \mathbb{X}, \gamma \in \Gamma. \end{aligned}$$

Puesto que esto es valido para toda $\gamma \in \Gamma$, en particular se tiene que

$$\begin{aligned} v^*(x) &\geq \sup_{\gamma\in\Gamma} V_\alpha(x,\pi,\gamma) \\ &\geq \inf_{\pi\in\Pi} \sup_{\gamma\in\Gamma} V_\alpha(x,\pi,\gamma) \\ &= V_\alpha^*(x), x \in \mathbb{X}. \end{aligned} \tag{2.3.8}$$

Entonces $v^*(x) \geq V_\alpha^*(x)$.

Con el fin de demostrar la desigualdad opuesta, nótese que como $B(x,a)$ es σ-compacto entonces para cada $\epsilon > 0$ existe $g : \mathbb{K}_A \to B$ con $g(x,a) \in B(x,a)$ y tal que

$$\begin{aligned} v^*(x) &= \inf_{a\in A(x)} \left[c(x,a,g) + \alpha \sum_{y\in\mathbb{X}} v^*(y) P_{x,y}(a,g) \right] + \epsilon \\ &\leq c(x,a,g) + \alpha \sum_{y\in\mathbb{X}} v^*(y) P_{x,y}(a,g) + \epsilon \quad \forall x \in \mathbb{X}, a \in A(x). \end{aligned} \tag{2.3.9}$$

Iterando (2.3.9)

$$v^*(x) \leq E_x^{\pi,g} \left[\sum_{k=0}^{n-1} \alpha^k c(x_k, a_k, g) \right] + \alpha^n E_x^{\pi,g} v^*(x_n) + \frac{\epsilon}{1-\alpha}.$$

Haciendo $n \to \infty$,

$$v^*(x) \leq V_\alpha(x,\pi,g) + \frac{\epsilon}{1-\alpha}.$$

Y puesto que ϵ es arbitrario,

$$v^*(x) \leq V_\alpha(x,\pi,g), \quad x \in \mathbb{X}, \pi \in \Pi.$$

Por otro lado, como $\mathbb{F}_B \subset \Gamma$,

$$V_\alpha(x,\pi,g) \leq \sup_{g\in\mathbb{F}_B} V_\alpha(x,\pi,g) \leq \sup_{\gamma\in\Gamma} V_\alpha(x,\pi,\gamma),$$

es decir

$$v^* \leq \sup_{\gamma\in\Gamma} V_\alpha(x,\pi,\gamma), x \in \mathbb{X}, \pi \in \Pi.$$

Finalmente, dado que $\pi \in \Pi$ es arbitrario, entonces

$$v^*(x) \leq \inf_{\pi\in\Pi} \sup_{\gamma\in\Gamma} V_\alpha(x,\pi,\gamma) = V_\alpha^*(x).$$

Por lo tanto $v^* = V_\alpha^*$.

(c) De (a) y Lema 2.2.3 (b) tenemos que existe $f^* \in \mathbb{F}_A$ tal que

$$V_\alpha^*(x) = \sup_{b\in B(x,f^*(x))} \left[c(x,f^*,b) + \alpha \sum_{y\in\mathbb{X}} V_\alpha^*(y) P_{x,y}(f^*,b) \right].$$

Aplicando argumentos similares a los utilizados para la demostración de (a) y

(b) (Véase (2.3.8)) se tiene que

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

En consecuencia,

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

Es decir, la política $\pi^* = \{f^*\}$ es minimax.

■

Capítulo 3

Sistemas de control estocástico con distribución desconocida

3.1. Introducción

En este capítulo analizaremos una aplicación de la teoría de control minimax al estudio de sistemas de control estocástico que dependen de parámetros desconocidos. En este caso, el oponente es considerado como la “naturaleza”que de cierta manera elige el parámetro desconocido en cada tiempo t. Específicamente, trataremos sistemas de control estocástico que evolucionan de acuerdo a una ecuación en diferencias de la forma

$$x_{t+1} = F(x_t, a_t, \xi_t), \quad t \in \mathbb{N}_0 \tag{3.1.1}$$

donde x_t y a_t representan el estado del sistema y el control al tiempo t respectivamente, y $\{\xi_t\}$ es una sucesión de variables aleatorias independientes con distribución desconocida. Entonces, las estrategias de la naturaleza toman la forma $\gamma = \{b_t\}$ con la cual elige una distribución b_t de la variable aleatoria ξ_t en cada tiempo $t \in \mathbb{N}_0$.De modo que, el problema del controlador es minimizar el máximo costo que genera la naturaleza, es decir, encontrar una política π^* que garantice el mejor rendimiento en la peor situación posible. A este tipo de problemas de control minimax se les conoce como *juegos contra la naturaleza*.

Comúnmente sistemas como (3.1.1), en los que la distribución de las variables aleatorias ξ_t es desconocida, se estudian con técnicas de control adaptado que involucran procesos de estimación y control. Esto último es posible siempre y cuando $\{\xi_t\}$ sea una sucesión de v.a. *i.i.d.* observables. Ahora bien, modelarlo como un juego contra la naturaleza nos permite considerar que $\{\xi_t\}$ sean v.a. no necesariamente observables y no necesariamente con la misma distribución. Con el fin de establecer claramente ambos casos, daremos una exposición general de los sistemas de control adaptado, para después introducir los juegos contra la naturaleza.

3.2. Sistemas de control estocástico

Sea

$$\mu_c = (\mathbb{X}, A, S, F, \theta, \hat{c}) \tag{3.2.1}$$

un modelo de control estocástico con espacio de estados y de control $\mathbb{X}$ y A, respectivamente, y S el espacio de perturbaciones aleatorias (todos los espacios se suponen numerables). Además $F : \mathbb{X} \times A \times S \to \mathbb{X}$ es la función que define la dinámica del sistema, esto es

$$x_{t+1} = F(x_t, a_t, \xi_t), \quad t \in \mathbb{N}_0, \tag{3.2.2}$$

donde $\{\xi_t\}$ es una sucesión de variables aleatorias *i.i.d.* que toman valores en S con función de probabilidad común $\theta \in \mathbb{P}(S)$, es decir

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

Finalmente, $\hat{c} : \mathbb{X} \times A \times S \to \mathbb{R}$ representa la función de costo.

El modelo (3.2.1) tiene la siguiente interpretación:
Al tiempo t, cuando el sistema se encuentra en el estado $x_t = x \in \mathbb{X}$ el controlador elige un control $a_t = a \in A$ y se presenta un ruido aleatorio ξ_t. Entonces, se genera un

costo $\hat{c}(x_t, a_t, \xi_t)$ y el sistema avanza a un nuevo estado de acuerdo a la probabilidad

$$P[x_{t+1} = y | x_t = x, a_t = a] = \sum_{k \in S_F} \theta(k),$$

donde

$$S_F = \{s \in S : F(x, a, s) = y\}.$$

Luego el proceso se repite una y otra vez.

Las acciones se eligen mediante las políticas de control definidas en el Capítulo 1. Además, los costos se acumulan de acuerdo al criterio de costo descontado. Es decir, para cada estado inicial $x \in \mathbb{X}$, cada política $\pi \in \Pi$ y $\alpha \in (0,1)$,

$$V(\pi, x) = E_x^\pi \left[\sum_{t=0}^{\infty} \alpha^t \hat{c}(x_t, a_t, \xi_t) \right] \tag{3.2.3}$$

El problema de control óptimo es encontrar una política $\pi^* \in \Pi$ tal que

$$V(\pi^*, x) = \inf_{\pi \in \Pi} V(\pi, x) =: V^*(x), \quad x \in \mathbb{X}. \tag{3.2.4}$$

A la política π^* se le llama política óptima y a la función V^* función de valor óptimo.

El esquema general para resolver el PCO es el siguiente:
Bajo condiciones apropiadas se demuestra que V^* satisface la ecuación de optimalidad. Esto es, para cada $x \in \mathbb{X}$

$$V^*(x) = \min_{a \in A(x)} \left[\sum_{s \in S} \{\hat{c}(x, a, s) + \alpha V^*[F(x, a, s)]\} \theta(s) \right]. \tag{3.2.5}$$

Y como $A(x)$ es finito, existe $f^* \in \mathbb{F}_A$ tal que

$$V^*(x) = \sum_{s \in S} \{\hat{c}(x, f^*(x), s) + \alpha V^*[F(x, f^*(x), s)]\} \theta(s). \tag{3.2.6}$$

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

3.2.1. Control adaptado

Consideremos un sistema de control estocástico como (3.2.2) donde $\{\xi_t\}$ es una sucesión de v.a. observables *i.i.d.*, con distribución desconocida θ. Observe que en este caso no es posible seguir el procedimiento estándar dado en (3.2.3)-(3.2.6) para resolver el Problema de Control Óptimo, ya que la ecuación de optimalidad depende de θ. Por lo tanto, en este caso se aplican técnicas de estimación estadística de θ combinados con el proceso de control para construir políticas, y al problema de control correspondiente se le llama *Problema de Control Adaptado*. Este nombre viene del proceso de estimación y control el cual consiste en lo siguiente.

En cada etapa, antes de elegir el control, el controlador determina un estimador θ_t de θ, y entonces el controlador adapta su decisión a tal estimador para obtener el control $a_t = f_t^{\theta_t}(x_t)$.

Uno de los procedimientos para resolver el problema de control adaptado es como sigue. Sea θ_t un estimador de θ, por ejemplo la distribución empírica:

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

Es conocido, por la *Ley Fuerte de los Grandes Números*, que

$$\theta_t(k) \to \theta(k) \quad \text{c.s., cuando} \quad t \to \infty.$$

Consideremos el modelo de control estimado

$$\mu_t = (\mathbb{X}, A, S, F, \theta_t, \hat{c}),$$

y denotemos por V_t la función de valor correspondiente. Entonces, bajo ciertas condiciones se tiene que para cada $t \in \mathbb{N}$ (Véase (3.2.5)),

$$V_t(x) = \min_{a\in A(x)}\left[\sum_{s\in S}\{\hat{c}(x,a,s) + \alpha V_t[F(x,a,s)]\}\theta_t(s)\right], \tag{3.2.7}$$

y más aún (ver (3.2.6)), existe $f_t = f_t^{\theta_t} \in \mathbb{F}_A$ tal que

$$V_t(x) = \sum_{s \in S} \{\hat{c}(x, f_t, s) + \alpha V_t[F(x, f_t, s)]\}\theta_t(s).$$

El objetivo es estudiar la optimalidad de la política adaptada $\overline{\pi} = \{f_t\}$.

Debido a las características del índice de costo descontando, en general, una política adaptada no es óptima. En tal situación, su optimalidad se estudia en sentido asintótico, lo cual significa que

$$\lim_{t \to \infty} E_x^{\overline{\pi}} \Phi(x_t, f_t) = 0,$$

donde

$$\Phi(x, a) = \sum_{s \in S} \{\hat{c}(x, a, s) + \alpha V^*[F(x, a, s)]\}\theta(s) - V^*(x), \quad x \in \mathbb{X}.$$

La teoría y desarrollo de este tipo de problemas se puede consultar en [14, 5, 8, 13] .

3.3. Sistemas de control con ruidos aleatorios no observables

Considérese nuevamente un modelo de control estocástico como en (3.2.1). En este caso supondremos que $\{\xi_t\}$ es una sucesión de variables aleatorias independientes, <u>no necesariamente idénticamente distribuidas ni observables</u> que toman valores en S, es decir, la distribución puede cambiar de periodo a periodo y es desconocida, pero está restringida a pertenecer a cierta clase $B \subset \mathbb{P}(S)$. Suponemos que el estado inicial x_0 es independiente de la sucesión $\{\xi_t\}$.

Es importante notar que en este caso no es posible aplicar el proceso de estimación y control como en los problemas de control adaptado contemplados en la sección anterior, ya que la hipótesis de que el proceso $\{\xi_t\}$ es observable es necesaria para construir estimadores, como en el caso de la distribución empírica.

La condición de observabilidad se presenta en muchas situaciones. Por ejemplo, en control de inventarios, colas y control de presas, donde las perturbaciones son, respectivamente, la demanda del producto, el tiempo entre arribos y la entrada de agua, las cuales podemos suponer observables. Sin embargo, hay otros casos donde las perturbaciones son ruidos realmente aleatorios y es imposible observarlos. Por ejemplo, en economía, finanzas, control de poblaciones (pesca, epidemias, etc), existen tantos factores externos que influyen en el sistema dinámico que es prácticamente necesario modelar las perturbaciones como un ruido aleatorio que no es observable. Esta es precisamente la clase de situaciones que estamos interesados en estudiar en esta sección, para lo cual tendremos que el oponente será la "naturaleza", quien en cada periodo t elegirá (de un conjunto dado B) una distribución para ξ_t.

Específicamente, nuestro objetivo es proponer un modelo de control minimax para modelar la situación anteriormente mencionada. Para ello, introduciremos algunas definiciones y resultados, que nos serán de utilidad.

Definición 3.3.1. *Sea $\boldsymbol{Z}$ un espacio de Borel. Denotamos por $\mathbb{P}(\boldsymbol{Z})$ a la familia de medidas de probabilidad de $\boldsymbol{Z}$, dotada con la convergencia de la topología débil. Es decir, una sucesión $\{\theta_n\} \subset \mathbb{P}(\boldsymbol{Z})$ converge débilmente a θ si*

$$\int_{\boldsymbol{Z}} u d\theta_n \to \int_{\boldsymbol{Z}} u d\theta$$

para cada función u continua y acotada.

Lema 3.3.2. *Sea $\boldsymbol{Z}$ un espacio de Borel. Entonces*

(a) $\mathbb{P}(\boldsymbol{Z})$ *es un espacio de Borel (Véase [10], página 91).*

(b) *Si ademas $\boldsymbol{Z}$ es compacto, entonces $\mathbb{P}(\boldsymbol{Z})$ es compacto (Véase [15], Teorema II 6.4).*

(c) *Si $\boldsymbol{Z}$ es σ-compacto entonces*

$$\mathbb{P}_0(\boldsymbol{Z}) := \{\theta \in \mathbb{P}(\boldsymbol{Z}) : \int_Z \|s\| d\theta(s) < \infty\}$$

es σ-compacto (Véase [11]).

Sea $\theta_t \in \mathbb{P}(S)$ la función de probabilidad de la variable aleatoria ξ_t, es decir,

$$\theta_t(k) = P[\xi_t = k], \quad t \in \mathbb{N}, \quad k \in S.$$

El objetivo es estudiar el problema de control planteado, como un problema de control minimax. Entonces, el procedimiento será definir un modelo de control minimax como en (1.1.1) tal que satisfaga la Hipótesis 2.1. Por lo tanto, los resultados del capítulo anterior serán válidos.

Consideremos el modelo de control estocástico μ_c definido en (3.2.1)

$$\mu_c = (\mathbb{X}, A, S, F, \theta, \hat{c}),$$

y supongamos que existe $w : \mathbb{X} \to [1, \infty)$ tal que la función de costo $\hat{c} : \mathbb{X} \times A \times S \to \mathbb{R}$ satisface que

$$|\hat{c}(x, a, s)| \leq \bar{c}w(x) \quad \forall (x, a, s) \in \mathbb{K}_A \times S.$$

Definimos la distribución condicional de x_{t+1} dado que $(x_t, a_t, \theta_t) = (x, a, \theta)$, donde $(x, a) \in \mathbb{K}_A$ y $\theta_t \in \mathbb{P}(S)$ representa la distribución de las v.a's ξ_t, como

$$P_{x,y}(a, \theta) := P[x_{t+1} = y | x_t = x, a_t = a, \theta_t = \theta] = \sum_{s \in S_F} \theta(s), \tag{3.3.1}$$

con

$$S_F := \{s \in S : F(x, a, s) = y\}.$$

Supondremos que para alguna constante $\beta \in [1, \frac{1}{\alpha})$ se cumple que

$$\sum_{y \in \mathbb{X}} w(y) P_{x,y}(a, \theta) \leq \beta w(x). \tag{3.3.2}$$

Con estos elementos es posible definir el modelo de control minimax asociado al modelo de control (3.2.1) como

$$\mu := (\mathbb{X}, A, B, \mathbb{K}_A, \mathbb{K}, P, c)$$

donde

$$B := \mathbb{P}(S),$$

$$\mathbb{K} := \{(x, a, \theta) | (x, a) \in \mathbb{K}_A, \theta \in B\}$$

y $c : \mathbb{K} \to \mathbb{R}$ es una función definida como

$$c(x, a, \theta) := \sum_{s \in S} \hat{c}(x, a, s)\theta(s). \tag{3.3.3}$$

Observemos que por el Lema 3.3.2(a) el conjunto B es un espacio de Borel. Más aún, si S es finito entonces B será compacto y que S sea numerable implica que el conjunto de medidas con esperanza finita sobre S será σ- compacto. Entonces, independientemente de la forma del conjunto S, el conjunto $B = \mathbb{P}(S)$ entra en el contexto de nuestro trabajo.

El modelo μ se puede considerar como un juego contra la naturaleza en el siguiente sentido: En cada tiempo $t \in \mathbb{N}$ el controlador observa el estado $x = x_t \in \mathbb{X}$, elige una acción $a = a_t \in A(x)$ y la naturaleza (oponente) elige una distribución $\theta_t \in B$ de la variable aleatoria ξ_t. Entonces se genera un costo dado por (3.3.3) y el sistema avanza a un nuevo estado de acuerdo a la ley de transición (3.3.1), luego se repite el proceso. Ahora bien, el objetivo del controlador es minimizar el máximo costo que genera la naturaleza.

Obsérvese que la relación (3.3.2) toma la forma

$$\sum_{s \in S} w[F(x, a, s)]\theta(s) \leq \beta w(x), \quad x \in \mathbb{X}, a \in A(x) \quad \text{con} \quad \beta \in [1, \frac{1}{\alpha}). \tag{3.3.4}$$

De donde, con el fin de aplicar los resultados del capítulo anterior, es suficiente demostrar que el costo c definido en (3.3.3) satisface la Hipótesis 2.1(a), lo cual se deduce de las relaciones siguientes:

$$\begin{aligned}
|c(x,a,\theta)| &= |\sum_{s\in S} \hat{c}(x,a,s)\theta(s)| \\
&\leq \sum_{s\in S} |\hat{c}(x,a,s)\theta(s)| \\
&\leq \sum_{s\in S} \overline{c}w(x)|\theta(s)| \\
&= \overline{c}w(x) \sum_{s\in S} \theta(s) \\
&= \overline{c}w(x).
\end{aligned}$$

En consecuencia, esta clase de juegos contra la naturaleza es posible resolverlos utilizando los métodos tratados en el Capítulo 2.

3.4. Ejemplo: Proceso de control autorregresivo

Consideremos un proceso de control que evoluciona de acuerdo a la ecuación en diferencias

$$x_{t+1} = a_t x_t + \xi_t, \quad t \in \mathbb{N}_0,$$

con x_0 conocido, y donde el espacio de estados es $\mathbb{X} = \mathbb{N}_0$,para cada $x \in \mathbb{X}$ el conjunto de acciones $A(x) = A$ finito, con $\overline{a} = \text{máx}\, A$, y $\{\xi_t\}$ es una sucesión de v.a. no observables ni idénticamente distribuidas y con distribución desconocida, definidas en un conjunto S, todos con esperanza acotada por $\overline{\xi}$.

Sea $x^* \in \mathbb{X}$ un estado fijo. Suponga que el objetivo del controlador es seleccionar acciones dirigidas a que el proceso $\{x_t\}$ permanezca lo más cerca posible del estado x^*. Bajo este objetivo, definimos la función de costo por etapa como sigue

$$\hat{c}(x,a,s) = |x - x^*|, \quad (x,a,s) \in \mathbb{K}_A \times S.$$

Definiendo para cada $x \in \mathbb{X}$ la función $w(x) = x + x^* + 1$ se tiene que

$$\begin{aligned}
\sum_{s\in S} w[F(x,a,s)]\theta(s) &= \sum_{s\in S} w[ax+s]\theta(s) \\
&= \sum_{s\in S} [ax+s+x^*+1]\theta(s) \\
&\leq \overline{a}x + x^* + 1 + \sum_{s\in S} s\theta(s) \\
&= \overline{a}x + x^* + 1 + \overline{\xi} \\
&\leq \overline{a}(x + x^* + 1) + x^* + 1 + \overline{\xi} \\
&\leq \overline{a}w(x) + x^* + 1 + \overline{\xi} \\
&\leq (\overline{a} + x^* + 1 + \overline{\xi})w(x), \quad x \in \mathbb{X}.
\end{aligned}$$

Por lo tanto, la relación (3.3.4) se cumple para

$$\alpha \leq \frac{1}{\overline{a} + x^* + 1 + \overline{\xi}} \qquad \text{y} \qquad \beta = \overline{a} + x^* + 1 + \overline{\xi}.$$

Este ejemplo satisface las condiciones necesarias para modelarse como un problema de control minimax. Por lo que para su resolución basta aplicar los métodos vistos en el capítulo 2. Es decir es una clase de juego contra la naturaleza que es posible resolver gracias a la teoría desarrollada en este trabajo.

Apéndice A

Teorema del Punto fijo

Definición A.0.1. *Una norma en un espacio vectorial es una función* $\|\cdot\| : X \to \mathbb{R}$*, que satisface las siguientes propiedades:*

(i) $\|x\| \geq 0$,

(ii) $\|x\| = 0$ *sí y sólo si* $x = 0$,

(iii) $\|\alpha x\| = |\alpha|\|x\|$,

(iv) $\|x + y\| \leq \|x\| + \|y\|$.

donde $x, y \in X$ *son vectores arbitrarios y* α *es cualquier escalar.*

Definición A.0.2. *Un espacio normado* X *es un espacio vectorial con una norma definida en él.*

Definición A.0.3. *Un espacio métrico es un par* (S, d)*, donde* S *es un conjunto no vacío, y* $d : S \times S \to \mathbb{R}$ *es una función tal que, para* $x, y, z \in S$ *arbitrarios satisface las siguientes propiedades:*

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

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

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

(iv) $d(x, y) \leq d(x,\ z) + d(z, y)$ *(desigualdad del triángulo).*

A la función d se le conoce como métrica.

Teorema A.0.4. *Una norma en X define una métrica d en X dada por*

$$d(x,y) = \|x-y\| \quad \forall x,y \in \mathbb{X},$$

que es llamada la métrica inducida por la norma.

Demostración. Debemos ver que satisface las 4 propiedades de una métrica. Sean $x,y,z \in X$ arbitrarios.

(i) Por la propiedad (ii) en la Definición A.0.1 se tiene

$$d(x,x) = \|x-x\| = \|0\| = 0.$$

(ii) Si $x \neq y$ entonces $x-y \neq 0$, y por las propiedades (i) y (ii) en la Definición A.0.1 se observa que

$$d(x,y) = \|x-y\| > 0.$$

(iii) Nótese que de la propiedad (iii) en la Definición A.0.1

$$\begin{aligned} d(x,y) &= \|x-y\| = \|(-1)(y-x)\|, \\ &= = |-1|\|y-x\| = \|y-x\|, \\ &= d(y,x). \end{aligned}$$

(iv) Finalmente, de la propiedad (iv) en Definición A.0.1 se sigue que

$$\begin{aligned} d(x,y) &= \|x-y\| = \|(x-z)+(z-y)\|, \\ &\leq \|x-z\| + \|z-y\|, \\ &= d(x,z) + d(z,y). \end{aligned}$$

∎

Definición A.0.5. *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 A.0.6. *Un espacio de Banach es un espacio normado completo con la métrica inducida por la norma.*

Definición A.0.7. *Sea $B(X)$ el espacio de las funciones acotadas definidas en X con la norma del supremo*

$$\|u\| := \sup_{x\in X} |u(x)|.$$

Es decir, $u \in B(X)$ si $\|u\| < \infty$.

Teorema A.0.8. *El espacio de las funciones acotadas $B(X)$ es un espacio de Banach.*

Demostración. Véase [16], Teorema 7.15, página 151. ■

Teorema A.0.9. *El espacio $\mathbb{B}_w(X)$ de las funciones u en X que tienen w-norma finita $\|u\|_w < \infty$, donde $w : X \to [1, \infty)$ y*

$$\|u\|_w = \sup_{x\in X} \frac{|u(x)|}{w(x)}$$

es de Banach.

Demostración. Sea $\{u_n\}$ una sucesión de Cauchy en B_w, entonces $\{\frac{u_n}{w}\}$ es de Cauchy en $B(X)$ con la norma del supremo. Puesto que $B(X)$ es un espacio de Banach, entonces $\{\frac{u_n}{w}\}$ converge a una función $f \in B(X)$, es decir,

$$\|\frac{u_n}{w} - f\| < \epsilon \quad \forall\epsilon > 0$$

De aquí se obtiene que

$$\begin{aligned}
\|\frac{u_n}{w} - f\| &= \sup_{x\in X}\left[\frac{u_n(x)}{w(x)} - f(x)\right] \\
&= \sup_{x\in X}\left[\frac{u_n(x - w(x)f(x))}{w(x)}\right] \\
&= \|u_n - wf\|_w \\
&< \epsilon.
\end{aligned}$$

Es decir,

$$\|u_n - fw\|_w < \epsilon,$$

lo cual implica que $\{u_n\}$ converge a fw. Finalmente, como fw pertenece a B_w, demostramos lo que se quería. ∎

Definición A.0.10. *Sea* $(S,\ d)$ *un espacio métrico. Se dice que un operador*

$$T : S \to S$$

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

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

Teorema A.0.11. ***(Teorema del Punto Fijo de Banach)*** *Si* (S, d) *es un espacio métrico completo y* $T : S \to S$ *es un operador de contracción en* S*, entonces:*

(a) *Existe un único* $x \in S$ *tal que*

$$Tx = x.$$

(b) *Para cada* $x_0 \in S$*,*

$$\lim_{n \to \infty} T^n x_0 = x.$$

Demostración. **(a)** Primero construiremos una sucesión $\{x_n\}$ en S y demostraremos que es Cauchy, es decir tendremos que converge en S; posteriormente demostraremos que el límite x de la sucesión es un punto fijo de T y finalmente demostramos que dicho punto es único.

Sea $x_0 \in S$ arbitrario, definimos la "sucesión iterativa" $\{x_n\}$ por

$$\begin{aligned} x_0 & \quad , \\ x_1 &= Tx_0, \\ x_2 &= Tx_1 = T^2 x_0, \\ &\vdots \\ x_n &= T^n x_0, \\ &\vdots \end{aligned} \tag{A.0.1}$$

La cual es la sucesión de imágenes de x_0 al aplicar el operador T repetidas veces. Veamos que es Cauchy.

Como T es un operador de contracción, existe $\alpha \in (0,1)$ tal que para todas $a, b \in S$

$$d(Ta, Tb) \leq \alpha d(a, b). \tag{A.0.2}$$

Luego, por (A.0.1) y (A.0.2)

$$\begin{aligned} d(x_{m+1}, x_m) &= d(Tx_m, T_{m-1}), \\ &\leq \alpha d(x_m, x_{m-1}), \\ &= d(Tx_{m-1}, Tx_{m-2}), \\ &\leq \alpha^2 d(x_{m-1}, x_{m-2}), \\ \cdots &\leq \alpha^m d(x_1, x_0). \end{aligned} \tag{A.0.3}$$

Entonces, por la desigualdad del triángulo y la fórmula para la suma de una sucesión geométrica, para $n > m$ obtenemos

$$\begin{aligned} d(x_m, x_n) &\leq d(x_m, x_{m+1}) + d(x_{m+1}, x_{m+2}) + \ldots + d(x_{n-1}, x_n), \\ &\leq (\alpha^m + \alpha^{m+1} + \cdots + \alpha^{n-1}) d(x_0, x_1), \\ &= \alpha^m \frac{1 - \alpha^{n-m}}{1 - \alpha} d(x_0, x_1). \end{aligned}$$

Como $\alpha \in (0,1)$, en el numerador tenemos $1 - \alpha^{n-m} < 1$. Así,

$$d(x_m, x_n) \leq \frac{\alpha^m}{1-\alpha} d(x_0, x_1) \quad (n > m). \tag{A.0.4}$$

Obsérvese que, en el lado derecho de la desigualdad (A.0.4) tenemos que $\alpha \in (0,1)$ y $d(x_0, x_1)$ son fijos, entonces es posible hacer el lado derecho de (A.0.4) tan pequeño como se quiera tomando m lo suficientemente grande (dado que $n > m$). Esto demuestra que $\{x_m\}$ es de Cauchy, y como S es completo, se tiene que $\{x_m\}$ converge, digamos a x, es decir, $x_m \to x$.

Veamos que ese límite x es un punto fijo de T.

Por la desigualdad del triángulo y (A.0.2) se tiene que

$$\begin{aligned} d(x, Tx) &\leq d(x, x_m) + d(x_m, Tx), \\ &\leq d(x, x_m) + \alpha d(x_{m-1}, x), \end{aligned}$$

y la última suma se puede hacer tan pequeña como cualquier $\epsilon > 0$ dado, pues $x_m \to x$. Entonces, concluimos que $d(x,\ Tx) = 0$, y de aquí que $Tx = x$. Esto demuestra que x es un punto fijo de T.

Además, x es el único punto fijo de T ya que si existieran dos puntos fijos, $Tx = x$ y $T\bar{x} = \bar{x}$, por (A.0.2) se tendria que

$$d(x, \bar{x}) = d(Tx, T\bar{x}) \leq \alpha d(x, \bar{x}),$$

lo cual implica que $d(x, \bar{x}) = 0$ pues $\alpha \in (0,1)$, y así $x = \bar{x}$.

(b) En la demostración del inciso anterior demostramos que para $x_0 \in S$ arbitrario, la sucesión formada por $\{x_n := T^n x_0\}$ converge al único punto fijo x. Por lo tanto, $\lim_{n\to\infty} T^n y = x$. ■

Observación A.0.12. *Una consecuencia del teorema anterior es la siguiente:*

$$d(T^n x_0, x) \leq \alpha^n d(x_0, x), \quad \forall n \in \mathbb{N}. \tag{A.0.5}$$

Demostración. *En efecto, primero observemos que*

$$d(Tx_0, x) = d(Tx_0, Tx) \leq \alpha d(x_0, x).$$

Ahora, supongamos que (A.0.5) se cumple para $n = k$, es decir,

$$d(T^k x_0, x) \leq \alpha^k d(x_0, x). \tag{A.0.6}$$

Entonces,

$$\begin{aligned} d(T^{k+1} x_0, x) &= d(T(T^k x_0), Tx), \\ &\leq \alpha d(T^k x_0, Tx), \\ &\leq \alpha^{k+1} d(x_0, x), \end{aligned}$$

donde la última desigualdad es por (A.0.6). Esto demuestra (A.0.5). ■

Apéndice B

Espacio medible, espacio de Borel y espacio de probabilidad

B.1. Espacio medible y espacio de Borel

El material de este apéndice puede ser consultado en distintos libros ([Véase [1, 6, 7]]). Sin embargo, para fácil referencia presentamos algunas definiciones y resultados de utilidad.

Definición B.1.1. *Una familia $\mathscr{A}$ de subconjuntos de $\mathbb{X}$ se dice ser una σ-álgebra si:*

(i) $\mathbb{X} \in \mathscr{A}$.

(ii) *Si $A \in \mathscr{A}$, entonces $A^c \in \mathscr{A}$.*

(iii) *Si $A_1, A_2, A_3, ... \in \mathscr{A}$, entonces $\bigcup_{n=1}^{\infty} A_n \in \mathscr{A}$.*

A un par ordenado $(\mathbb{X}, \mathscr{A})$ que consiste de un conjunto $\mathbb{X}$ y una σ- álgebra de subconjuntos de $\mathbb{X}$ se le llama **espacio medible**. Cualquier conjunto en $\mathscr{A}$ es llamado **conjunto $\mathscr{A}$-medible**, pero cuando la σ- álgebra es fija y no exista confusión se dice que el conjunto es **medible**.

Definición B.1.2. *Sea $\mathbb{X}$ un conjunto arbitrario no vacío y sea $\mathscr{C}$ una colección de subconjuntos de $\mathbb{X}$. Definimos σ-álgebra generada por $\mathscr{C}$, denotada por $\sigma(\mathscr{C})$,* *mediante*

$$\sigma(\mathscr{C}) := \bigcap\{\mathscr{A}_\alpha : \mathscr{C} \subset \mathscr{A}_\alpha \quad y \quad \mathscr{A}_\alpha \quad es\ \sigma - \text{álgebra en } \mathbb{X}\}.$$

Observación B.1.3. *$\sigma(\mathscr{C})$ es la mínima σ- álgebra que contiene a $\mathscr{C}$.*

Definición B.1.4. *Sea $\mathbb{X}$ un espacio métrico y $\mathscr{C}$ la colección de todos sus conjuntos abiertos. A $\sigma(\mathscr{C})$ se le llama la σ- álgebra de Borel en $\mathbb{X}$ y se denota por $\mathfrak{B}(\mathbb{X})$.*

Definición B.1.5. *Un conjunto de Borel es un elemento de la llamada σ-álgebra de Borel. Además un subconjunto de Borel de un espacio métrico completo y separable es llamado un espacio de Borel.*

Un subconjunto de Borel de un espacio de Borel es un espacio de Borel.

B.2. Espacio de probabilidad

Definición B.2.1. *El espacio muestral de un experimento aleatorio es un conjunto Ω tal que:*

a) *cada posible resultado del experimento se identifica con un elemento de Ω;*

b) *resultados distintos del experimento corresponden a elementos distintos de Ω.*

Definición B.2.2. *Sea $\mathcal{F}$ una σ-álgebra de Ω. Una medida de probabilidad definida en $\mathcal{F}$ es una función que a cada evento $A \in \mathcal{F}$ le asocia un número $P(A)$ que satisface las siguientes condiciones:*

(i) $0 \leq P(A) \leq 1$.

(ii) $P(\Omega) = 1$.

(iii) *Si $A_1, A_2, A_3, ... \in \mathcal{F}$ colección numerable de eventos disjuntos, es decir $A_i \cap A_j = \emptyset$ para $i \neq j$, entonces*

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

Definición B.2.3. *Un espacio de probabilidad es una arreglo de la forma* $(\Omega, \mathcal{F}, P)$*, compuesto por un conjunto* Ω*, una* σ*-álgebra* $\mathcal{F}$ *de subconjuntos de* Ω *y una medida de probabilidad* P *en* $(\Omega, \mathcal{F})$*.*

Dado un espacio $\mathbb{X}$ denotamos por $\mathbb{P}(\mathbb{X})$ a la familia de medidas de probabilidad en $\mathbb{X}$, dotada con la convergencia de la topología débil. Es decir, una sucesión $\{\theta_n\} \subset \mathbb{P}(\mathbb{X})$ converge débilmente a θ si

$$\int_{\mathbb{X}} u d\theta_n \to \int_{\mathbb{X}} u d\theta$$

para cada función continua y acotada u.

Definición B.2.4. *Un conjunto se dice ser* σ*-compacto si es la unión numerable de conjuntos compactos.*

Lema B.2.5. *Sea* $\mathbb{X}$ *un espacio de Borel. Entonces,*

(a) $\mathbb{P}(\mathbb{X})$ *es un espacio de Borel (Véase [10], página 91).*

(b) *Si ademas* $\mathbb{X}$ *es compacto, entonces* $\mathbb{P}(\mathbb{X})$ *es compacto (Véase [15], Teorema II 6.4).*

(c) *Si* $\mathbb{X}$ *es* σ*-compacto entonces,*

$$\mathbb{P}_0(\mathbb{X}) := \{\theta \in \mathbb{P}(\mathbb{X}) : \int_{\mathbb{X}} \|s\| d\theta(s) < \infty\}$$

es σ*-compacto (Véase [11]).*

B.2.1. Ley Fuerte de los grandes números

Teorema B.2.6. *Sean* $X_1, X_2, ..., X_n$ *variables aleatorias (i.i.d) con media finita* μ*. Entonces*

$$\overline{X_n} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

converge a μ *casi seguramente, es decir,*

$$P[\lim_{n \to \infty} \overline{X_n} = \mu] = 1.$$

Demostración. Véase [3], Teorema 3.30. ■

Apéndice C

Glosario de símbolos y abreviaturas

Símbolos

- $\mathbb{N}$: Números naturales.
- $\mathbb{N}_0$: $\mathbb{N} \cup \{0\}$.
- $\mathbb{R}$: Números reales.
- $\mathfrak{B}(\cdot)$: σ- álgebra de borel.
- $\mathbb{B}_w(\mathbb{X})$: Espacio lineal normado de las funciones con w-norma finita.

Abreviaturas

- Modelo de Control Minimax (MCM).
- Programación Dinámica (PD).
- Problema de Control Minimax (PCM).
- Problema de Control Óptimo (PCO).
- Independientes e idénticamente distribuidas (*i.i.d.*).
- Es decir (*i.e.*).

- Variable aleatoria (v.a).
- Casi seguramente (c.s).

Bibliografía

[1] Bartle, R. G. *The Elements of Integration and Lebesgue Measure*. 1995.

[2] Bertsekas, D. P. *Dynamic Programming: Deterministic and Stochastic Models*. Prentice-Hall, 1987.

[3] Breiman, L. *Probability*. Classics in Applied Mathematics. SIAM: Society for Industrial and Applied Mathematics, 1992.

[4] González-Trejo, J., Hernández-Lerma, O., and Hoyos-Reyes, L. F. Minimax control of discrete-time stochastic systems. *SIAM Journal on Control and Optimization 41*, 5 (2002), 1626–1659.

[5] Gordienko, E. I., and Minjárez-Sosa, J. A. Adaptive control for discrete-time markov processes with unbounded costs: discounted criterion. *Kybernetika 34* (1998), 217–234.

[6] Grabinsky, G. *Teoría de la medida*. No. 515.42 G7. 2011.

[7] Grimmett, G. R., and Stirzaker, D. R. *Probability and Random Processes*, 3 ed. Oxford University Press, 2001.

[8] Hernández-Lerma, O. *Adaptive Markov Control Processes*, 1 ed. Applied Mathematical Sciences 79. Springer-Verlag New York, 1989.

[9] Hernández-Lerma, O., and Lasserre, J. B. *Further topics on discrete-time Markov control processes*. Springer Science & Business Media, 2012.

[10] Hinderer, K. *Foundations of Non-stationary Dynamic Programming with Discrete Time Parameter.* Lecture Notes in Operations Research and Mathematical Systems 33. Springer-Verlag Berlin Heidelberg, 1970.

[11] Luque-Vásquez, F., and Minjárez-Sosa, J. A. A note on the σ-compactness of sets of probability measures on metric spaces. *Statistics & Probability Letters 84* (2014), 212–214.

[12] Luque-Vásquez, F., Minjárez-Sosa, J. A., and Vega-Amaya, O. *Introducción a la Teoría de Control Estocástico.* Universidad de Sonora, 1996.

[13] Minjárez-Sosa, J. Estimación empírica en sistemas de control de markov. Por aparecer.

[14] Minjárez-Sosa, J. A. Approximation and estimation in markov control processes under a discounted criterion. *Kybernetika 40* (2004), 681–690.

[15] Parthasarathy, K. R. *Probability measures on metric spaces.* Probability and mathematical statistics. Academic Press, 1967.

[16] Rudin, W. *Principles of Mathematical Analysis.* New York : McGraw-Hill, 1976.