
Universidad de Sonora
División de Ciencias Exactas y Naturales
Departamento de Matemáticas



**Estrategias Adaptadas para juegos
Markovianos de suma cero**

Tesis

para obtener grado académico de Maestro en Ciencias (Matemáticas)
presenta:

Carmen Geraldí Higuera Chan

Director de Tesis:
Dr. Jesús Adolfo Minjárez Sosa

Septiembre de 2011

SINODALES

Dr. Fernando Luque Vásquez
Departamento de Matemáticas, Universidad de Sonora

Dr. Héctor Jasso Fuentes
Departamento de Matemáticas, CINVESTAV

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

Dr. Oscar Vega Amaya
Departamento de Matemáticas, Universidad de Sonora

Índice general

Índice general	III
Introducción	v
1. Juegos Markovianos de suma cero	1
1.1. Formulación de un juego suma cero	1
1.2. Estrategias	3
1.3. Criterios de Optimalidad	4
2. Criterio de optimalidad descontado	7
2.1. Preliminares	7
2.2. Juegos con horizonte finito	9
2.3. Juegos con horizonte infinito	14
2.4. Estrategias asintóticamente óptimas	22
3. Criterio de optimalidad promedio	33
3.1. Teorema de existencia	34
3.2. Estimación de la densidad	39
3.3. Construcción de estrategias óptimas	40
4. Conclusiones	48
A. Espacios de Borel, Kernels y multifunciones	50
A.1. Espacios de Borel y kernels estocásticos	50
A.2. Multifunciones y selectores	51
B. Semicontinuidad y Lema de Fatou	53
B.1. Funciones semicontinuas	53
B.2. Lema de Fatou generalizado	54

<i>ÍNDICE GENERAL</i>	IV
C. Varios resultados	57
C.1. Teorema de punto fijo de Banach	57
C.2. Teorema minimax de Fan	57
C.3. Teorema del Máximo de Berge y Teorema de Selección medible	58
Bibliografía	61

Introducción

En este trabajo estudiaremos juegos markovianos de suma cero.

En términos generales, un juego suma cero consiste en lo siguiente. Se tienen dos jugadores con objetivos opuestos. En el tiempo t , supongamos que el estado del juego es $x_t = x$, entonces independientemente cada jugador elige una acción o control a_t y b_t , respectivamente, de determinado conjunto, y sucede lo siguiente:

- 1) El jugador 1 recibe un pago $r(x, a, b)$ del jugador 2, es decir, $r(x, a, b)$ representa la ganancia para el jugador 1 y el costo para el jugador 2.
- 2) El juego se mueve a nuevo estado $x_{t+1} = x'$ de acuerdo a una determinada distribución de probabilidad.
- 3) Una vez en el estado x' el proceso se repite.

Los jugadores seleccionan sus acciones mediante reglas conocidas como estrategias de control, y el pago se acumula durante la evolución del juego a tiempo discreto considerando el criterio de optimalidad descontado y el de pago promedio por etapa. Por lo tanto, el objetivo del jugador 1 es determinar estrategias para maximizar su ganancia, mientras que las estrategias del jugador 2 estarían dirigidas a minimizar su costo.

Una clase particular de juegos suma cero se tiene cuando la evolución en el tiempo es modelada por medio de ecuaciones en diferencia de la forma

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

donde F es una función conocida y $\{\xi_t\}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas, llamado proceso de perturbación, con densidad común ρ . En este caso la distribución de probabilidad que representa la ley de transición entre los estados la determina la densidad ρ junto con la función F .

El objetivo principal de este trabajo de tesis es establecer condiciones para la existencia de estrategias óptimas para los jugadores tanto para el criterio de optimalidad descontado como para el criterio de optimalidad promedio cuando los espacios de estados y de acciones de ambos jugadores son de Borel, y el pago es posiblemente no acotado. Además, en el contexto de los juegos del tipo (1) construiremos estrategias (casi) óptimas en el caso descontado y óptimas en el caso promedio cuando la densidad ρ es desconocida por los jugadores.

Nuestra metodología para el estudio de estos problemas es la siguiente. Primeramente estudiamos el criterio descontado con horizonte finito N . Luego, a partir de estos resultados, analizamos el problema con horizonte infinito haciendo $N \rightarrow \infty$. Además, el criterio de optimalidad promedio se estudia aplicando el método de *factor de descuento desvaneciente*, es decir, como límite del caso descontado. Por otro lado, en el problema, tanto para el caso descontado como para el promedio, cuando la densidad ρ es desconocida, construimos para cada jugador un estimador estadístico de la densidad. Entonces la decisión de cada jugador es adaptada a la correspondiente estimación para determinar sus estrategias. Las estrategias que combinan estimación y control se les conocen como *estrategias adaptadas*.

El trabajo está estructurado de la siguiente manera. En el Capítulo 1 se introducen los elementos necesarios para la descripción de un juego estocástico suma cero así como los criterios de optimalidad que estudiaremos. Enseguida, en el Capítulo 2, se estudia el criterio descontado en el cual se introducen las hipótesis que garantizan la existencia de estrategias óptimas tanto para el caso de horizonte finito como el infinito. Además se presenta la construcción de estrategias adaptadas asintóticamente descontadas óptimas para los jugadores. Después, tomando en cuenta los resultados obtenidos, en el Capítulo 3 estudiamos el criterio de optimalidad promedio como límite del caso descontado, y se presentan las condiciones que garantizan la existencia de estrategias óptimas. De igual forma, se construyen estrategias adaptadas, las cuales a diferencia del caso descontado, éstas resultan ser óptimas.

Capítulo 1

Juegos Markovianos de suma cero

En este capítulo se introducen los elementos necesarios para describir un juego estocástico suma cero. Además presentaremos la definición de estrategia así como los índices de funcionamiento que estudiaremos a lo largo del trabajo.

Finalmente, veremos algunos resultados que serán útiles en la construcción de estrategias adaptadas para los criterios descontado y promedio en los Capítulos 2 y 3 respectivamente.

1.1. Formulación de un juego suma cero

Consideraremos el modelo de un juego estocástico de dos personas con suma cero de la siguiente forma

$$(X, A, B, \mathbb{K}_A, \mathbb{K}_B, Q, r) \tag{1.1}$$

donde:

- i) X es el espacio de estados del juego, el cual suponemos que es un espacio de Borel (ver Apéndice A.1).
- ii) A y B son los espacios de acciones para los jugadores 1 y 2, respectivamente, y también suponemos que son espacios de Borel.
- iii) \mathbb{K}_A y \mathbb{K}_B son subconjuntos de Borel no vacíos de $X \times A$ y $X \times B$, respectivamente. Además para cada $x \in X$, la x -sección

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

es un subconjunto no vacío de A que representa el conjunto de acciones admisibles para el jugador 1 en el estado x . Análogamente, la x -sección $B(x) := \{b \in B : (x, b) \in \mathbb{K}_B\}$ denota el conjunto de acciones admisibles para el jugador 2 en el estado x . Luego definimos

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

el cual es un subconjunto de Borel de $X \times A \times B$.

- iv) Q es un kernel estocástico en X dado \mathbb{K} (ver Apéndice A), llamado la ley de transición entre los estados. Es decir $Q(\cdot | x, a, b)$ es la distribución condicional de x_{t+1} dado $x_t = x$, $a_t = a$ y $b_t = b$ para todo $t \geq 0$.
- v) $r : \mathbb{K} \rightarrow \mathbb{R}$ es una función medible que denota la función de pago, y representa la ganancia para el jugador 1 y el costo para el jugador 2.

El modelo (1.1) representa un juego dinámico que evoluciona en el tiempo de la siguiente forma. En cada etapa $t = 0, 1, \dots$, los jugadores 1 y 2 observan el estado actual del juego $x_t = x \in X$, e independientemente los jugadores eligen las acciones $a \in A(x)$ y $b \in B(x)$, respectivamente. Como consecuencia de esto, pasa lo siguiente: (1) el jugador 1 recibe inmediatamente una ganancia $r(x, a, b)$; (2) el jugador 2 tiene un costo $r(x, a, b)$, y (3) el sistema se mueve a un nuevo estado de acuerdo a la distribución $Q(\cdot | x, a, b)$. Así el objetivo del jugador 1 es maximizar su ganancia, mientras que el del jugador 2 es minimizar su costo.

Observación 1.1.1. En algunos casos, la evolución del juego se puede modelar mediante una ecuación recursiva de la forma

$$x_{n+1} = F(x_n, a_n, b_n, \xi_n), \quad n = 0, 1, \dots, \quad (1.2)$$

donde $F : X \times A \times B \times \mathbb{R}^k \rightarrow X$ es una función medible y el proceso de perturbación $\{\xi_n\}$ está formado por vectores aleatorios independientes e idénticamente distribuidos en \mathbb{R}^k , para algún entero fijo k , con densidad común ρ . En este caso, el kernel estocástico Q queda determinado como

$$Q(D|x, a, b) := \int_{\mathbb{R}^k} I_D[F(x, a, b, s)] \rho(s) ds, \quad D \in \mathcal{B}(X), \quad (1.3)$$

donde $I_D(\cdot)$ es la función indicadora del conjunto D .

Con el fin de especificar claramente los elementos de un juego estocástico cuando estamos trabajando en este contexto expresaremos al modelo del juego de la siguiente forma:

$$(X, A, B, \mathbb{K}_A, \mathbb{K}_B, F, \rho, r). \quad (1.4)$$

1.2. Estrategias

Consideremos el modelo de un juego estocástico como en (1.1), y para cada $t = 0, 1, \dots$, definimos el espacio de historias admisibles hasta la t -ésima etapa como

$$H_0 = X \text{ y } H_t := \mathbb{K}^t \times X = \mathbb{K} \times H_{t-1} \text{ para } t = 1, 2, \dots \quad (1.5)$$

Un elemento de H_t es un vector al que llamaremos historia, de la forma

$$h_t = (x_0, a_0, b_0, \dots, x_{t-1}, a_{t-1}, b_{t-1}, x_t). \quad (1.6)$$

Definición 1.2.1. Una estrategia aleatorizada π para el jugador 1, es una sucesión $\pi = \{\pi_t, t = 0, 1, \dots\}$ de kernels estocásticos π_t en $\mathbb{P}(A|H_t)$ (ver Definición A.1.1) tal que

$$\pi(A(x_t)|h_t) = 1 \quad \forall h_t \in H_t, t = 0, 1, \dots$$

Denotamos por Π a la familia de estrategias del jugador 1.

Definición 1.2.2. Una estrategia $\pi = \{\pi_t\}$ se dice ser:

(a) *Markoviana (o de Markov)* si $\pi_t \in \mathbb{P}(A|X)$ para cada $t = 0, 1, \dots$, esto es, cada π_t depende solo del estado actual x_t del sistema.

Al conjunto de las estrategias de Markov del jugador 1 lo denotaremos por Π_M .

(b) *Estacionaria* si existe $f \in \mathbb{P}(A|X)$ tal que tal que $\pi_t = f$ para cada $t = 0, 1, \dots$

En este caso la estrategia estacionaria π será identificada con f .

Denotaremos Π_S al conjunto de las estrategias estacionarias del jugador 1.

Tenemos, por supuesto, las siguientes contenciones

$$\Pi_S \subset \Pi_M \subset \Pi.$$

Similarmente definimos las estrategias para el jugador 2, y denotamos por $\Gamma, \Gamma_M, \Gamma_S$

al conjunto de todas las estrategias, de las estrategias de Markov y de las estrategias estacionarias, respectivamente, para el jugador 2.

Sea (Ω, \mathcal{F}) el espacio medible (canónico) que consiste del espacio muestral $\Omega := (X \times A \times B)^\infty$ y su σ -álgebra producto \mathcal{F} . Entonces para cada par de estrategias $(\pi, \gamma) \in \Pi \times \Gamma$ y cada estado inicial $x \in X$, por el Teorema de C. Ionescu Tulcea (ver [3]), existe una única medida de probabilidad $P_x^{\pi\gamma}$ y un proceso estocástico $\{(x_t, a_t, b_t), t = 0, 1, \dots\}$ definido en (Ω, \mathcal{F}) de manera canónica, donde x_t, a_t y b_t representan el estado y las acciones del jugador 1 y 2, respectivamente, en cada etapa $t = 0, 1, \dots$. El operador esperanza con respecto a $P_x^{\pi\gamma}$ es denotado por $E_x^{\pi\gamma}$.

1.3. Criterios de Optimalidad

Para cada estado inicial $x_0 = x$, estrategias $\pi \in \Pi$, $\gamma \in \Gamma$, y $n \in \mathbb{N}$, definiremos el pago total esperado hasta la n -ésima etapa como:

$$J_n(x, \pi, \gamma) = E_x^{\pi\gamma} \left[\sum_{t=0}^{n-1} r(x_t, a_t, b_t) \right]. \quad (1.7)$$

Así mismo, sea $\alpha \in (0, 1)$. Definimos el pago descontado esperado hasta la n -ésima etapa como:

$$V_n(x, \pi, \gamma) := E_x^{\pi\gamma} \left[\sum_{t=0}^{n-1} \alpha^t r(x_t, a_t, b_t) \right]. \quad (1.8)$$

El pago promedio esperado con horizonte infinito se define como

$$J(x, \pi, \gamma) = \liminf_{n \rightarrow \infty} \frac{J_n(x, \pi, \gamma)}{n} \quad (1.9)$$

y el pago descontado esperado con horizonte infinito como

$$V_\alpha(x, \pi, \gamma) := E_x^{\pi\gamma} \left[\sum_{t=0}^{\infty} \alpha^t r(x_t, a_t, b_t) \right], \quad (1.10)$$

para cada par de estrategias (π, γ) y cada estado inicial x . El número α es llamado el “factor de descuento”.

Definición 1.3.1. Para cada $n = 1, 2, \dots$, definimos las funciones

$$L_n^\alpha(x) := \sup_{\pi \in \Pi} \inf_{\gamma \in \Gamma} V_n(x, \pi, \gamma)$$

y

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

Las funciones $L_n^\alpha, U_n^\alpha : X \rightarrow \mathbb{R}$ son llamadas el valor inferior y el valor superior, del juego en n etapas, respectivamente.

Es claro que $L_n^\alpha(\cdot) \leq U_n^\alpha(\cdot)$, y si además se tiene que $L_n^\alpha(x) = U_n^\alpha(x)$ para todo $x \in X$, entonces la función común es llamada el valor del juego en n etapas, y la denotamos por $V_n^*(\cdot)$.

Definición 1.3.2. Consideremos un juego en n etapas.

i) Una estrategia $\pi^* \in \Pi$ se dice ser óptima para el jugador 1 si

$$U_n^\alpha(x) \leq V_n(x, \pi^*, \gamma) \quad \text{para cada } \gamma \in \Gamma \text{ y } x \in X.$$

ii) Una estrategia $\gamma^* \in \Gamma$ se dice ser óptima para el jugador 2 si

$$L_n^\alpha(x) \geq V_n(x, \pi, \gamma^*) \quad \text{para cada } \pi \in \Pi \text{ y } x \in X.$$

Si i) y ii) se satisfacen, entonces (π^*, γ^*) se dice ser un par óptimo de estrategias. Equivalentemente, (π^*, γ^*) es un par de estrategias óptimo para el juego en n etapas, si para todo $x \in X$,

$$U_n^\alpha(x) = \inf_{\gamma \in \Gamma} V_n(x, \pi^*, \gamma) \text{ y } L_n^\alpha(x) = \sup_{\pi \in \Pi} V_n(x, \pi, \gamma^*).$$

La función de valor y las estrategias óptimas para el caso α -descontado con horizonte infinito, así como para los casos de pago total y promedio, son definidas de manera similar, y denotaremos por V_α^*, J_n^* y J^* el valor del juego correspondiente.

Existe otra manera de definir optimalidad, por ejemplo, en el caso α -descontado con horizonte infinito, si el juego tiene un valor V_α^* , una estrategia $\pi^* \in \Pi$ se dice ser ϵ -óptima para el jugador 1, $\epsilon \geq 0$, si

$$V_\alpha^*(\cdot) - \epsilon \leq \inf_{\gamma \in \Gamma} V_\alpha(\cdot, \pi^*, \gamma). \quad (1.11)$$

Similarmente, una estrategia $\gamma^* \in \Gamma$ se dice ser ϵ -óptima para el jugador 2 si

$$V_\alpha^*(\cdot) + \epsilon \geq \sup_{\pi \in \Pi} V_\alpha(\cdot, \pi, \gamma^*). \quad (1.12)$$

De aquí que una estrategia es óptima cuando es 0-óptima.

Un par de estrategias (π^*, γ^*) es llamado un par óptimo o punto silla y observemos que (π^*, γ^*) es un par óptimo si y sólo si

$$V_\alpha(\cdot, \pi, \gamma^*) \leq V(\cdot, \pi^*, \gamma^*) \leq V(\cdot, \pi^*, \gamma) \quad \forall \pi \in \Pi, \quad \forall \gamma \in \Gamma.$$

Definiciones similares de estrategias ϵ -óptimas y puntos silla se tienen para el criterio de costo promedio.

Para concluir esta sección introduciremos el concepto de optimalidad asintótica para pares de estrategias el cual es mas débil que el criterio de optimalidad visto anteriormente. Este concepto de optimalidad es una adaptación a juegos estocásticos de la noción de optimalidad asintótica para procesos de control de Markov propuesta por Schäll en [26] (ver también [9, 10, 14]), y será usada en el estudio de las estrategias adaptadas para el caso descontado.

Definamos la función de discrepancia $D : \mathbb{K} \rightarrow \mathbb{R}$ como:

$$D(x, a, b) := r(r, a, b) + \alpha \int_{\mathbb{R}^k} V_\alpha^* [(x, a, b, s)] \rho(s) ds - V_\alpha^*(x).$$

Definición 1.3.3. Una estrategia $\pi^* \in \Pi$ se dice ser asintóticamente descontada (AD)-óptima para el jugador 1 si

$$\liminf_{n \rightarrow \infty} E_x^{\pi^* \gamma} D(x_n, a_n, b_n) \geq 0 \quad \forall x \in X, \quad \gamma \in \Gamma.$$

Similarmente, $\gamma^* \in \Gamma$ se dice ser AD -óptima para el jugador 2 si

$$\limsup_{n \rightarrow \infty} E_x^{\pi^* \gamma^*} D(x_n, a_n, b_n) \leq 0 \quad \forall x \in X, \quad \pi \in \Pi.$$

Un par de estrategias AD -óptimas es llamado un par AD -óptimo. En este caso, si (π^*, γ^*) es un par AD -óptimo, entonces, para cada $x \in X$,

$$\lim_{n \rightarrow \infty} E_x^{\pi^* \gamma^*} D(x_n, a_n, b_n) = 0.$$

Capítulo 2

Criterio de optimalidad descontado

Una vez establecidos los elementos necesarios para la descripción de un juego estocástico suma cero, en este capítulo se analiza el criterio de optimalidad descontado con horizonte finito e infinito (ver (1.8) y (1.10)).

Introducimos también un conjunto de hipótesis que garantizan la existencia de pares de estrategias óptimas de acuerdo a este criterio de optimalidad, para lo cual presentamos una serie de resultados preliminares que son consecuencia de dichas hipótesis.

Además en el contexto de ecuaciones en diferencias (ver Observación 1.1.1), consideramos el caso en que los jugadores desconocen la densidad ρ , y se construirán estrategias adaptadas que serán asintóticamente óptimas.

Por último, presentaremos de manera general otros métodos para la construcción de estrategias adaptadas.

2.1. Preliminares

Antes de introducir los resultados principales, estableceremos la siguiente notación.

Para cada $x \in X$, $\mathbb{P}_A(x) := \mathbb{P}(A(x))$ y $\mathbb{P}_B(x) := \mathbb{P}(B(x))$. Además notemos que las multifunciones $x \mapsto \mathbb{P}_A(x)$ y $x \mapsto \mathbb{P}_B(x)$ son medibles, y con valores en los conjuntos compactos si $A(x)$ y $B(x)$ son compactos (ver Proposición A.2.1).

Definimos $\mathbb{F}_{\mathbb{P}_A} := \{f \in \mathbb{P}(A|X) : f(x) \in \mathbb{P}_A(x)\}$ y $\mathbb{F}_{\mathbb{P}_B} := \{f \in \mathbb{P}(B|X) : f(x) \in \mathbb{P}_B(x)\}$. A los elementos de $\mathbb{F}_{\mathbb{P}_A}$ y $\mathbb{F}_{\mathbb{P}_B}$ se les llama selectores medibles para los jugadores 1 y

2 respectivamente.

Para medidas de probabilidad $\mu \in \mathbb{P}_A(x)$ y $\lambda \in \mathbb{P}_B(x)$, y cualquier función medible $u : \mathbb{K} \rightarrow \mathbb{R}$ definimos:

$$u(x, \mu, \lambda) := \int_{A(x)} \int_{B(x)} u(x, a, b) \mu(da) \lambda(db), \quad x \in X.$$

De esta manera para las funciones r y Q establecidas anteriormente, y para cada conjunto de Borel $D \subset \mathcal{B}(X)$ tenemos

$$r(x, \mu, \lambda) := \int_{A(x)} \int_{B(x)} r(x, a, b) \mu(da) \lambda(db), \quad (2.1)$$

$$Q(D|x, \mu, \lambda) := \int_{A(x)} \int_{B(x)} Q(D|x, a, b) \mu(da) \lambda(db), \quad (2.2)$$

y

$$\int_X u(y) Q(dy|x, \mu, \lambda) := \int_{B(x)} \int_{A(x)} \int_X u(y) Q(dy|x, a, b) \mu(da) \lambda(db). \quad (2.3)$$

Definición 2.1.1. Sea $w : X \rightarrow [1, \infty)$ una función medible. Para cada función medible $u : X \rightarrow \mathbb{R}$ definimos su w -norma como

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

Denotaremos por $\mathbb{B}_w(X)$ al espacio de Banach de las funciones con w -norma finita. Además, $\mathbb{L}_w(X)$ ($\mathbb{C}_w(X)$) es el conjunto de funciones $u \in \mathbb{B}_w(X)$ que son semicontinuas inferiormente (continuas).

Para cada $u \in \mathbb{B}_w(X)$ y $(x, a, b) \in \mathbb{K}$, definimos

$$H(u; x, a, b) := r(x, a, b) + \alpha \int_X u(y) Q(dy|x, a, b). \quad (2.4)$$

Con la notación de antes, definimos el operador

$$T_\alpha u(x) := \sup_{\mu \in \mathbb{P}_A(x)} \inf_{\lambda \in \mathbb{P}_B(x)} H(u; x, \mu, \lambda) \quad \forall u \in \mathbb{B}_w(X). \quad (2.5)$$

Más adelante podremos asegurar que el supremo y el ínfimo se alcanzan, así que podemos reemplazar \inf y \sup por $\mbox{mín}$ y $\mbox{máx}$, respectivamente. De esta manera tendremos

$$T_\alpha u(x) = \mbox{máx}_{\mu \in \mathbb{P}_A(x)} \mbox{mín}_{\lambda \in \mathbb{P}_B(x)} H(u; x, \mu, \lambda) \quad \forall u \in \mathbb{B}_w(X). \quad (2.6)$$

2.2. Juegos con horizonte finito

Los resultados de existencia del valor del juego, así como de pares óptimos los obtendremos bajo dos tipos de hipótesis. La primera de ellas es la siguiente.

Hipótesis 1. **(a)** Para cada estado $x \in X$, los conjuntos de acciones admisibles $A(x)$ y $B(x)$ son compactos.

(b) Para cada $(x, a, b) \in \mathbb{K}$, $r(x, \cdot, b)$ es semicontinua superiormente en $A(x)$, y $r(x, a, \cdot)$ es semicontinua inferiormente en $B(x)$.

(c) Para cada $(x, a, b) \in \mathbb{K}$ y cada función acotada y medible v en X , las funciones

$$\int_X v(y)Q(dy|x, \cdot, b) \text{ y } \int_X v(y)Q(dy|x, a, \cdot)$$

son continuas en $A(x)$ y $B(x)$, respectivamente.

(d) Existe una constante $M > 0$ y una función medible $w : X \rightarrow \mathbb{R}$ tal que $w(x) \geq 1$ para cada $x \in X$ y

$$|r(x, a, b)| \leq Mw(x). \quad (2.7)$$

Además (c) se cumple cuando v es remplazada con w .

(e) Existe una constante $1 < \nu < \frac{1}{\alpha}$ tal que

$$\int_X w(y)Q(dy|x, a, b) \leq \nu w(x) \quad \forall (x, a, b) \in \mathbb{K}. \quad (2.8)$$

La condición (e) puede ser remplazada por

(e)' Existe un conjunto de Borel no vacío $C \subset X$ tal que para algún $\beta \in (0, 1)$ y $\eta > 0$, tenemos

$$\int_X w(y)Q(dy|x, a, b) \leq \beta w(x) + \eta I_C(x),$$

para cada $(x, a, b) \in \mathbb{K}$. En este caso $I_C(x)$ es la función indicadora del conjunto C y w es la función definida en (d).

El siguiente teorema muestra como podemos intercambiar el máximo y mínimo de la definición (2.5).

Lema 2.2.1. Supongamos que la Hipótesis 1 se cumple. Entonces para cada u en $\mathbb{B}_w(X)$:

$$(a) \quad T_\alpha u(x) = \max_{\lambda \in \mathbb{P}_B(x)} \min_{\mu \in \mathbb{P}_A(x)} H(u; x, \mu, \lambda),$$

(b) Existe $f_0 \in \mathbb{F}_{\mathbb{P}_A}$ y $g_0 \in \mathbb{F}_{\mathbb{P}_B}$ tal que, para todo $x \in X$,

$$\begin{aligned} T_\alpha u(x) &= \max_{\mu \in \mathbb{P}_A(x)} H(u; x, \mu, g_0(x)) \\ &= \min_{\lambda \in \mathbb{P}_B(x)} H(u; x, f_0(x), \lambda) \\ &= H(u; x, f_0(x), g_0(x)) \end{aligned}$$

(c) $T_\alpha u$ es un elemento de $\mathbb{B}_w(X)$.

Demostración. Elijamos una función u en $\mathbb{B}_w(X)$.

(a) Por las Hipótesis 1(c) y 1(d) la integral en (2.4) es continua en $a \in A(x)$ y $b \in B(x)$ (ver Lema 8.3.7(a) en [13]). Este hecho y la Hipótesis 1(b) implican que para cada $(x, a, b) \in \mathbb{K}$ la función $H(u; x, \cdot, b)$ es semicontinua superiormente en $A(x)$, y $H(u; x, a, \cdot)$ es semicontinua inferiormente en $B(x)$. De aquí la función $H(u; x, \mu, \lambda)$ es semicontinua superiormente en $\mu \in \mathbb{P}_A(x)$, y semicontinua superiormente en $\lambda \in \mathbb{P}_B(x)$. Más aún, $H(u; x, \mu, \lambda)$ es concava en μ y convexa en λ . Así por el Teorema Minimax de Fan (ver Teorema C.2.1), obtenemos la parte (a).

(b) Definamos

$$H_1(x, \mu) := \min_{\lambda \in \mathbb{P}_B(x)} H(u; x, \mu, \lambda),$$

para todo $x \in X$ y $\mu \in \mathbb{P}_A(x)$. Como podemos notar, en la prueba de la parte (a), $H_1(x, \cdot)$ es semicontinua superiormente en $\mathbb{P}_A(x)$. Por tanto, por Proposición A.1.1 y Teorema C.3.4, existe $f_0 \in \mathbb{F}_{\mathbb{P}_A}$ tal que

$$H_1(x, f_0(x)) = \max_{\mu \in \mathbb{P}_A(x)} H_1(x, \mu) \quad \forall x \in X.$$

En consecuencia, obtenemos

$$H_1(x, f_0(x)) = \max_{\mu \in \mathbb{P}_A(x)} \min_{\lambda \in \mathbb{P}_B(x)} H(u; x, \mu, \lambda). \quad (2.9)$$

Definamos el operador

$$T_\alpha u(x) := \max_{\mathbb{P}_A(x)} \min_{\mathbb{P}_B(x)} H(u; x, \mu, \lambda), \quad \forall u \in \mathbb{B}(X).$$

Por (2.9), tenemos

$$T_\alpha u(x) = \min_{\lambda \in \mathbb{P}_B(x)} H(u; x, f_0(x), \lambda).$$

De manera similar, sea

$$H_2(x, \lambda) := \max_{\mu \in \mathbb{P}_A(x)} H(u; x, \mu, \lambda).$$

Entonces existe $g_0 \in \mathbb{F}_{\mathbb{P}_B}$ tal que

$$T_\alpha u(x) = \min_{\lambda \in \mathbb{P}_A(x)} H(u; x, \mu, g_0(x)).$$

(c) Como $|u(\cdot)| \leq \|u\|_w w(\cdot)$, de (2.7) y (2.8) obtenemos, para cualquier (x, a, b) en \mathbb{K} ,

$$\begin{aligned} |H(u; x, a, b)| &\leq Mw(x) + \|u\|_w \alpha \int_X w(y) Q(dy|x, a, b) \\ &\leq (M + \alpha \nu \|u\|_w) w(x). \end{aligned}$$

Así, por el inciso (b) y esto último, $T_\alpha u$ pertenece a $\mathbb{B}_w(X)$.

□

Una consecuencia del lema anterior es el siguiente resultado que garantiza la existencia del valor del juego con horizonte finito.

Teorema 2.2.2. Supongamos que se satisface la Hipótesis 1. Entonces el juego estocástico con horizonte finito tiene un valor y ambos jugadores tienen estrategias de Markov óptimas. Mas aún, si V_n^* es la función de valor para el juego con horizonte n , entonces $V_n^* \in \mathbb{B}_w(X)$ y $V_n^*(x) = T_\alpha V_{n-1}^*(x)$ para cada $n \geq 2$.

Demostración. La prueba se sigue por inducción. Para $n = 1$ se sigue directamente de (1.8) y Lema 2.2.1 con $u(\cdot) \equiv 0$. Supongamos ahora que el resultado es válido para $n - 1$ ($n \geq 2$). Sean $\pi_{n-1} = (f_1, f_2, \dots, f_{n-1})$ y $\gamma_{n-1} = (g_1, g_2, \dots, g_{n-1})$ un par

de estrategias óptimas para los jugadores 1 y 2, respectivamente, en el juego con n etapas. Entonces

$$V_{n-1}^*(\cdot) = V_{n-1}(\cdot, \pi_{n-1}, \gamma_{n-1}).$$

Sean $U_n^\alpha(\cdot)$ y $L_n^\alpha(\cdot)$ las funciones de valor superior y inferior, respectivamente (ver Definición 1.3.1). Elijamos $g \in \mathbb{F}_{\mathbb{P}_B}$ arbitraria y definamos $\gamma^g := (g, \gamma_{n-1})$. Notemos que, por definición de U_n^α ,

$$U_n^\alpha(x) \leq \sup_{\pi \in \Pi} V_n(x, \pi, \gamma^g)$$

De aquí, para cada $x \in X$,

$$U_n^\alpha(x) \leq \sup_{\mu \in \mathbb{P}_A(x)} \left[r(x, \mu, g(x)) + \alpha \int_X \sup_{\pi \in \Pi} V_{n-1}(y, \pi, \gamma_{n-1}) Q(dy|x, \mu, g(x)) \right],$$

esto es,

$$U_n^\alpha(x) \leq \sup_{\mu \in \mathbb{P}_A(x)} \left[r(x, \mu, g(x)) + \alpha \int_X V_{n-1}(y) Q(dy|x, \mu, g(x)) \right].$$

Sea $H_{n-1}(x, \mu, \lambda) := H(V_{n-1}; x, \mu, \lambda)$. Por tanto,

$$U_n^\alpha(x) \leq \sup_{\mu \in \mathbb{P}_A(x)} H_{n-1}(x, \mu, g(x))$$

y, como $g \in \mathbb{F}_{\mathbb{P}_B}$ es arbitraria,

$$U_n^\alpha(x) \leq \inf_{\lambda \in \mathbb{P}_B(x)} \sup_{\mu \in \mathbb{P}_A(x)} H_{n-1}(x, \mu, \lambda)$$

para cada $x \in X$. Por tanto, por el Lema 2.2.1, llegamos a

$$U_n^\alpha(x) \leq T_\alpha V_{n-1}^*(x).$$

Similarmente, tenemos

$$L_n^\alpha(x) \geq T_\alpha V_{n-1}^*(x).$$

De estas dos últimas desigualdades, concluimos que $U_n^\alpha(x) = L_n^\alpha(x) = T_\alpha V_{n-1}^*(x)$, lo cual muestra que el juego con horizonte n tiene un valor $V_n^* = T_\alpha V_{n-1}^*$. Además, por el Lema 2.2.1, concluimos que $V_n^* \in \mathbb{B}_w(X)$ y existen $f_0 \in \mathbb{F}_{\mathbb{P}_A}$ y $g_0 \in \mathbb{F}_{\mathbb{P}_B}$ tales que para cada $f \in \mathbb{F}_{\mathbb{P}_A}$ y $g \in \mathbb{F}_{\mathbb{P}_B}$

$$V_n(x) = H_{n-1}(x, f_0(x), g_0(x)),$$

$$V_n(x) \geq H_{n-1}(x, f(x), g_0(x)),$$

$$V_n(x) \leq H_{n-1}(x, f_0(x), g(x)).$$

Por último, sean $\pi_n = (f_0, f_1, \dots, f_{n-1})$ y $\gamma_n = (g_0, g_1, \dots, g_{n-1})$. Entonces concluimos que π_n y γ_n son óptimas para los jugadores 1 y 2, respectivamente, para el juego estocástico en n etapas. \square

Ahora estableceremos otro resultado de existencia para este tipo de juegos considerando la siguiente hipótesis.

Hipótesis 2. (a) Las correspondencias $x \mapsto A(x)$ y $x \mapsto B(x)$ son continuas con valores compactos.

(b) La función $r : \mathbb{K} \rightarrow \mathbb{R}$ es continua.

(c) El kernel Q es débilmente continuo (ver Definición A.1.2).

(d) Existe una función $w : X \rightarrow [1, \infty)$ tal que $|r(x, a, b)| \leq w(x)$ para cada $(x, a, b) \in \mathbb{K}$.

(e) La función

$$(x, a, b) \mapsto \int_X w(y)Q(dy|x, a, b)$$

es continua en \mathbb{K} .

(f) Existen constantes positivas $\beta < 1$ y c tal que para cada $(x, a, b) \in \mathbb{K}$,

$$\int_X w(y)Q(dy|x, a, b) \leq \beta w(x) + c.$$

Teorema 2.2.3. Supongamos que la Hipótesis 2 se cumple. Entonces las conclusiones del Teorema 2.2.2 son válidas. Más aún, $V_n^* \in \mathbb{C}_w(X)$.

Demostración. La prueba se sigue por inducción y se hace aplicando argumentos de programación dinámica (como en la demostración del Teorema 2.2.2). En este caso, para demostrar la continuidad de las funciones V_n^* y la existencia de estrategias markovianas se usan los Teoremas C.3.3 y C.3.4. \square

2.3. Juegos con horizonte infinito

En esta sección, consideraremos juegos estocásticos con horizonte infinito. Probaremos que el valor del juego V_α^* es un punto fijo del operador T_α in (2.6), esto es, $V_\alpha^* = T_\alpha V_\alpha^*$, y mostraremos que la sucesión $\{V_n^\alpha\}$ definida en la sección anterior, converge geoméricamente a V_α en la w -norma.

De nuevo el resultado se establecerá bajo las Hipótesis 1 y 2 de la sección anterior.

Consideraremos la función de pago α -descontado $V_\alpha(x, \pi, \gamma)$ y sus correspondientes funciones de valor inferior y superior, $L_\alpha(\cdot)$ y $U_\alpha(\cdot)$, respectivamente. El objetivo es mostrar que $L_\alpha(\cdot) = U_\alpha(\cdot)$, de tal forma que el valor para el juego con horizonte infinito existe y es $V_\alpha^*(\cdot)$. Para hacer esto primero introducimos algunos resultados preliminares.

Lema 2.3.1. Bajo la Hipótesis 1, el operador T_α definido en (2.5) es un operador de contracción en $\mathbb{B}_w(X)$ con módulo $\tau := \nu\alpha$ (con ν como en (2.8)).

Demostración. Primero notemos que T_α es un operador monótono, es decir, si u y \tilde{u} son funciones en $\mathbb{B}_w(X)$, y $u \geq \tilde{u}$, entonces $T_\alpha u(x) \geq T_\alpha \tilde{u}(x)$ para todo $x \in X$. Por otro lado, por (2.8), para cualquier número real $k \geq 0$ se cumple:

$$T_\alpha(u + kw)(x) \leq T_\alpha u(x) + \nu\alpha kw(x), \quad (2.10)$$

para todo $x \in X$ y $u \in \mathbb{B}_w(X)$. Ahora, para verificar que T_α es de contracción, elijamos u y \tilde{u} en $\mathbb{B}_w(X)$. Como $u \leq \tilde{u} + w \|u - \tilde{u}\|_w$, la monotonía de T_α y (2.10) con $k = \|u - \tilde{u}\|_w$ implican

$$T_\alpha u(x) \leq T_\alpha(\tilde{u} + kw)(x) \leq T_\alpha \tilde{u}(x) + \nu\alpha kw(x),$$

esto es,

$$T_\alpha u(x) - T_\alpha \tilde{u}(x) \leq \nu\alpha \|u - \tilde{u}\|_w w(x), \quad x \in X.$$

Si intercambiamos u y \tilde{u} obtenemos

$$T_\alpha u(x) - T_\alpha \tilde{u}(x) \geq -\nu\alpha \|u - \tilde{u}\|_w w(x),$$

así que

$$|T_\alpha u(x) - T_\alpha \tilde{u}(x)| \leq \nu\alpha \|u - \tilde{u}\|_w w(x).$$

Por lo tanto, haciendo $\tau := \nu\alpha$, obtenemos $\|T_\alpha u - T_\alpha \tilde{u}\| \leq \tau \|u - \tilde{u}\|_w$, y con esto queda demostrado el lema pues $u, \tilde{u} \in \mathbb{B}_w(x)$ fueron arbitrarias. \square

Lema 2.3.2. Sean M , w y ν como en la Hipótesis 1, y además, sean $\pi \in \Pi$ y $\gamma \in \Gamma$ estrategias arbitrarias para los jugadores 1 y 2, respectivamente, y $x \in X$ el estado inicial. Entonces para cada $t = 0, 1, \dots$

- (a) $E_x^{\pi\gamma} w(x_t) \leq \nu^t w(x)$,
- (b) $|E_x^{\pi\gamma} r(x_t, a_t, b_t)| \leq M\nu^t w(x)$,
- (c) $\lim_{t \rightarrow \infty} \alpha^t E_x^{\pi\gamma} u(x_t) = 0$ para toda $u \in \mathbb{B}_w(X)$.

Demostración. (a) El caso $t = 0$ se sigue trivialmente. Ahora, si $t \geq 1$, tenemos

$$\begin{aligned} E_x^{\pi\gamma} [w(x_t) | h_{t-1}, a_{t-1}, b_{t-1}] &= \int_X w(y) Q(dy | x_{t-1}, a_{t-1}, b_{t-1}) \\ &\leq \nu w(x_{t-1}) \quad \text{por (2.8)}. \end{aligned}$$

Por lo tanto tenemos la desigualdad $E_x^{\pi\gamma} w(x_t) \leq \nu E_x^{\pi\gamma} w(x_{t-1})$, la cual iterando obtenemos (a).

(b) Observe que con la Hipótesis 1(d) tenemos

$$|r(x_t, a_t, b_t)| \leq Mw(x_t) \quad \forall t = 0, 1, \dots,$$

así que por el inciso (a),

$$E_x^{\pi\gamma} |r(x_t, a_t, b_t)| \leq M\nu^t w(x).$$

(c) Por definición de w -norma y por la parte (a), obtenemos

$$E_x^{\pi\gamma} |u(x_t)| \leq \|u\|_w E_x^{\pi\gamma} w(x_t) \leq \|u\|_w \nu^t w(x),$$

de donde se sigue (c). □

Definición 2.3.1. Sean $f \in \mathbb{F}_{\mathbb{P}_A}$, $g \in \mathbb{F}_{\mathbb{P}_B}$, y H como en (2.4). Definimos el operador

$$R_{fg} : \mathbb{B}_w(X) \rightarrow \mathbb{B}_w(X), \quad u \mapsto R_{fg}u,$$

por

$$R_{fg}u(x) := H(u; x, f(x), g(x)) \quad \forall x \in X. \quad (2.11)$$

Lema 2.3.3. El operador R_{fg} es un operador de contracción con módulo $\tau := \alpha\nu$ en $\mathbb{B}_w(X)$ y $V_\alpha(x, f, g)$ es su único punto fijo en $\mathbb{B}_w(X)$.

Demostración. Aplicando argumentos similares como en la prueba del Lema 2.3.1, se sigue que R_{fg} es un operador de contracción en $\mathbb{B}_x(X)$ con módulo $\tau := \alpha\nu$. Por lo tanto, R_{fg} tiene un único punto fijo u_{fg} en $\mathbb{B}_w(X)$, es decir,

$$u_{fg} = R_{fg}u_{fg}, \quad (2.12)$$

con lo cual tenemos que u_{fg} es la única solución en $\mathbb{B}_w(X)$ de la ecuación

$$u_{fg}(x) = r(x, f(x), g(x)) + \alpha \int_X u_{fg}(y) Q(dy|x, f(x), g(x)), \quad \forall x \in X. \quad (2.13)$$

Más aún, iterando (2.12) o (2.13) llegamos a

$$u_{fg}(x) = R_{fg}^n u_{fg}(x) = E_x^{fg} \left[\sum_{t=0}^{n-1} \alpha^t r(x_t, f(x_t), g(x_t)) \right] + \alpha^n E_x^{fg} u_{fg}(x_n)$$

para toda $x \in X$ y $n \geq 1$, donde $E_x^{fg} u(x_n) = \int_X u(y) Q^n(dy|x, f, g)$, y $Q^n(\cdot|x, f, g)$ la n -ésima transición del kernel del proceso de Markov $\{x_t\}$ cuando usamos las estrategias f y g . Finalmente, por el Lema 2.3.2(c) y haciendo $n \rightarrow \infty$ tenemos, de la definición de V_α , que $u_{fg}(x) = V_\alpha(x, f, g)$ para todo $x \in X$. □

Ahora estableceremos uno de los resultados principales en esta sección, el cual establece, entre otras cosas, que la sucesión $\{V_n\}$ converge geoméricamente a V_α en la w -norma, así como la existencia de un par de estrategias óptimas. Para la demostración se consideran los resultados establecidos en el Teorema 2.2.2 para el caso de horizonte finito n , para después hacer $n \rightarrow \infty$.

Teorema 2.3.4. Bajo la Hipótesis 1 y dadas las constantes ν como en (2.8) y $\tau := \nu\alpha$, tenemos:

(a) La función de valor α -descontado V_α^* es la única función en el espacio $\mathbb{B}_w(X)$ que satisface la ecuación $T_\alpha V_\alpha^* = V_\alpha^*$, y

$$\|V_n^* - V_\alpha^*\|_w \leq M\tau^n / (1 - \tau) \quad \forall n = 1, 2, \dots \quad (2.14)$$

(b) Existe un par de estrategias óptimo.

Demostración. Por el Lema 2.3.2 y el Teorema de punto fijo de Banach (Proposición C.1.1), T_α tiene un único punto fijo V^* en $\mathbb{B}_w(X)$, es decir,

$$T_\alpha V^* = V^*, \quad (2.15)$$

y

$$\|T_\alpha^n u - V^*\|_w \leq \tau^n \|u - V^*\|_w \quad \forall u \in \mathbb{B}_w(X), n = 0, 1, \dots \quad (2.16)$$

De aquí, para probar la parte (a) necesitamos mostrar que

(i) V_α^* está en $\mathbb{B}_w(X)$, con norma $\|V_\alpha^*\|_w \leq M/(1 - \tau)$, y

(ii) $V_\alpha^* = V^*$.

En efecto, usando

$$V_n^* = T_\alpha V_{n-1}^* = T_\alpha^n V_0 \quad \forall n = 0, 1, \dots, V_0 = 0, \quad (2.17)$$

(2.14) se seguirá de (2.17) y (2.16) con $u \equiv 0$.

Para probar (i), consideremos $\pi \in \Pi$ y $\gamma \in \Gamma$ estrategias arbitrarias para los jugadores 1 y 2, respectivamente, y $x \in X$ un estado inicial arbitrario, entonces (i) se deduce del Lema 2.3.2(b) y observando que

$$|V_\alpha(x, \pi, \gamma)| \leq \sum_{t=0}^{\infty} \alpha^t E_x^{\pi\gamma} |r(x_t, a_t, b_t)| \leq Mw(x)/(1 - \tau),$$

con $\tau := \alpha\nu$. En consecuencia, como $\pi \in \Pi, \gamma \in \Gamma$ y $x \in X$ fueron arbitrarias,

$$|V_\alpha^*(x)| \leq Mw(x)/(1 - \tau).$$

Para probar (ii) notemos que por la igualdad $V^* = T_\alpha V^*$ y el Lema 2.2.1, existen $f_* \in \mathbb{F}_{\mathbb{P}_A}$ y $g_* \in \mathbb{F}_{\mathbb{P}_B}$ tales que, para todo $x \in X$,

$$\begin{aligned} V^*(x) &= \sup_{\mu \in \mathbb{P}_A(x)} H(V^*; x, \mu, g_*(x)) \\ &= \inf_{\lambda \in \mathbb{P}_B(x)} H(V^*; x, f_*(x), \lambda) \\ &= H(x, f_*(x), g_*(x)). \end{aligned} \quad (2.18)$$

Observemos que (2.18) puede ser escrito como

$$V^*(x) = r(x, f_*(x), g_*(x)) + \alpha \int_X V^*(y) Q(dy|x, f_*(x), g_*(x)).$$

Entonces se sigue del Lema 2.3.3, que $V^*(x) = V_\alpha(x, f_*, g_*)$. Por tanto, tenemos

$$V_\alpha(x, f_*, g_*) = \sup_{\mu \in \mathbb{P}_A(x)} \left[r(x, \mu, g_*(x)) + \alpha \int_X V_\alpha(y, f_*, g_*) Q(dy|x, \mu, g_*(x)) \right],$$

para todo $x \in X$. De aquí, aplicando resultados estándar de programación dinámica (ver por ejemplo [12], [13]), se sigue que

$$V_\alpha(x, f_*, g_*) = \sup_{\pi \in \Pi} V_\alpha(x, \pi, g_*). \quad (2.19)$$

De manera similar, considerando ahora el ínfimo en (2.18) obtenemos

$$V_\alpha(x, f_*, g_*) = \inf_{\gamma \in \Gamma} V_\alpha(x, f_*, \gamma). \quad (2.20)$$

De (2.19) y (2.20) concluimos

$$\begin{aligned} V_\alpha(x, f_*, g_*) &= \sup_{\pi \in \Pi} V_\alpha(x, \pi, g_*) \geq \inf_{\gamma \in \Gamma} \sup_{\pi \in \Pi} V_\alpha(x, \pi, \gamma), \\ V_\alpha(x, f_*, g_*) &= \inf_{\gamma \in \Gamma} V_\alpha(x, f_*, \gamma) \leq \sup_{\pi \in \Pi} \inf_{\gamma \in \Gamma} V_\alpha(x, \pi, \gamma), \end{aligned}$$

lo cual a su vez implican

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

Esto prueba que el juego estocástico tiene un valor, que la función de valor es $V_\alpha^*(x) = V_\alpha(x, f_*, g_*) = V^*(x)$ para todo $x \in X$ y que f_* , g_* son estrategias óptimas para los jugadores 1 y 2, respectivamente. □

Ahora estableceremos un resultado análogo en el contexto de la Hipótesis 2.

Teorema 2.3.5. Supongamos que la Hipótesis 2 se cumple. Entonces el juego estocástico con costo descontado tiene un valor V_α^* y ambos jugadores tienen estrategias estacionarias óptimas $f_* \in \mathbb{F}_{\mathbb{P}_A}$ y $g_* \in \mathbb{F}_{\mathbb{P}_B}$. Más aún, $V_\alpha^* \in \mathbb{C}_w(X)$ y

$$\begin{aligned}
V_\alpha^*(x) &= T_\alpha V_\alpha^*(x) = r(s, f_*(x), g_*(x)) + \alpha \int_X V_\alpha^*(y) Q(dy|x, f_*(x), g_*(x)) \quad (2.21) \\
&= \max_{\mu \in \mathbb{P}_A(x)} \left[r(x, \mu, g_*(x)) + \alpha \int_X V_\alpha^*(y) Q(dy|x, \mu, g_*(x)) \right] \\
&= \min_{\lambda \in \mathbb{P}_B(x)} \left[r(x, f_*(x), \lambda) + \alpha \int_X V_\alpha^*(y) Q(dy|x, f_*(x), \lambda) \right]
\end{aligned}$$

para cada $x \in X$.

Demostración. Primero mostremos que V_α^* es un punto fijo en $\mathbb{C}_w(X)$ del operador T_α para cada $\alpha \in (0, 1)$. Por el Teorema 2.2.3, para cada $n \geq 0$, tenemos

$$\begin{aligned}
V_n^*(x) &= \min_{\gamma \in \Gamma} \max_{\pi \in \Pi} E_x^{\pi\gamma} \left(\sum_{t=0}^{n-1} \alpha^t r(x_t, a_t, b_t) \right) \\
&= \max_{\pi \in \Pi} \min_{\gamma \in \Gamma} E_x^{\pi\gamma} \left(\sum_{t=0}^{n-1} \alpha^t r(x_t, a_t, b_t) \right)
\end{aligned}$$

y

$$V_n^*(x) = T_\alpha^n V_0(x),$$

para cada $x \in X$, $n \geq 1$. De la Hipótesis 2(e), uno puede probar fácilmente por inducción que

$$E_x^{\pi\gamma}(w(x_n)) \leq \lambda^n w(x) + \varphi(w) \sum_{k=0}^{n-1} \lambda^k \leq w(x) + \frac{\varphi(w)}{1-\lambda}, \quad x \in X, (\pi, \gamma) \in \Pi \times \Gamma. \quad (2.22)$$

Ahora, supongamos que $n > k$ para $k \in \mathbb{N}$ y notemos que

$$\begin{aligned}
|V_n^*(x) - V_k^*(x)| &\leq \sup_{\pi \in \Pi} \sup_{\gamma \in \Gamma} \left| E_x^{\pi\gamma} \left(\sum_{t=0}^{n-1} \alpha^t r(x_t, a_t, b_t) \right) - E_x^{\pi\gamma} \left(\sum_{t=0}^{k-1} \alpha^t r(x_t, a_t, b_t) \right) \right| \\
&\leq \sup_{\pi \in \Pi} \sup_{\gamma \in \Gamma} E_x^{\pi\gamma} \left(\sum_{t=k}^{n-1} \alpha^t |r(x_t, a_t, b_t)| \right) \\
&\leq \sup_{\pi \in \Pi} \sup_{\gamma \in \Gamma} E_x^{\pi\gamma} \left(\sum_{t=k}^{n-1} \alpha^t w(x_t) \right).
\end{aligned}$$

Así que de (2.22), se sigue que

$$|V_n^*(x) - V_k^*(x)| \leq \sum_{t=k}^{n-1} \alpha^t w(x) d \leq \frac{\alpha^k}{1-\alpha} w(x) d$$

para alguna constante d . Esto implica que la sucesión $\{V_n^*\}$ es de Cauchy en $\mathbb{C}_w(X)$. Sea

$$V_\alpha^*(x) = \lim_{n \rightarrow \infty} V_n^*(x), \quad x \in X.$$

Notemos que $V_\alpha^* \in \mathbb{C}_w(X)$ y

$$\|V_n^* - V_\alpha^*\|_w = \sup_{x \in X} \frac{|V_n^*(x) - V_\alpha^*(x)|}{w(x)} \rightarrow 0 \text{ cuando } n \rightarrow \infty. \quad (2.23)$$

Para $n \geq 2$, es un hecho que

$$V_n^*(x) = T_\alpha^n V_0(x) = T_\alpha V_{n-1}^*(x).$$

Además tenemos que $T_\alpha V_{n-1} \rightarrow T_\alpha V_\alpha^*$ cuando $n \rightarrow \infty$. En efecto, notemos que

$$\begin{aligned} |T_\alpha V_n^*(x) - T_\alpha V_\alpha^*(x)| &\leq \max_{a \in A(x)} \max_{b \in B(x)} \int_X \frac{|V_n^*(y) - V_\alpha^*(y)|}{w(y)} w(y) Q(dy|x, a, b) \\ &\leq \|V_n^* - V_\alpha^*\|_w \max_{a \in A(x)} \max_{b \in B(x)} \int_X w(y) Q(dy|x, a, b). \end{aligned}$$

Ahora de la Hipótesis 2(f) y (2.23), concluimos que $T_\alpha V_n \rightarrow T_\alpha V_\alpha^*$ cuando $n \rightarrow \infty$, con lo cual tenemos que V_α^* es un punto fijo de T_α . La existencia de f_* y g_* que satisfacen la optimalidad y las igualdades deseadas, se siguen aplicando argumentos similares a la demostración del Teorema 2.3.4 usando en este caso los Teoremas C.3.3 y C.3.4. \square

2.3.1. Resultados adicionales

Concluimos la presente sección presentando un nuevo conjunto de hipótesis propuesto por Küenle en [18] para mostrar la existencia del valor del juego. Estas hipótesis son mas débiles que las Hipótesis 1 y 2, pero solo garantizan la existencia de estrategias ϵ -óptimas (ver (1.11) y (1.12)).

Hipótesis 3. (a) La correspondencia $x \mapsto A(x)$ es semicontinua inferiormente y $A(x)$ es completo para cada $x \in X$.

(b) La correspondencia $x \mapsto B(x)$ es semicontinua superiormente y $B(x)$ es compacto para cada $x \in X$.

(c) La función $r(\cdot, \cdot, \cdot) \geq 0$ es semicontinua inferiormente en \mathbb{K} .

(d) El kernel Q es débilmente continuo.

(e) Existe una función medible $w \geq 1$ en X y constantes positivas $\beta < 1$, M y c , tal que las siguientes desigualdades se satisfacen para $(x, a, b) \in \mathbb{K}$:

$$(e.1) \quad 0 \leq r(x, a, b) \leq Mw(x);$$

$$(e.2) \quad \int_X w(y)Q(dy|x, a, b) \leq \beta w(x) + c.$$

Teorema 2.3.6 (Küenle [18]). Bajo la Hipótesis 3 tenemos lo siguiente:

(a) El juego tiene un valor $V_\alpha^* \in \mathbb{L}_w$.

(b) Existe una estrategia estacionaria g_* tal que

$$V^*(x) = \sup_{\mu \in \mathbb{P}A(x)} \left[r(x, \mu, g_*) + \alpha \int_{\mathbb{R}^k} V_\alpha^*(y)Q(dy|x, \mu, g_*) \right] \quad \forall x \in X;$$

más aún, para cada $\epsilon > 0$ existe una estrategia estacionaria f_ϵ tal que

$$V^*(x) - \epsilon \leq \inf_{\lambda \in \mathbb{P}B(x)} \left[r(x, f_\epsilon, \lambda) + \alpha \int_{\mathbb{R}^k} V_\alpha^*(y)Q(dy|x, f_\epsilon, \lambda) \right] \quad \forall x \in X.$$

(c) f_ϵ es ϵ -óptima para el jugador 1 y g_* es óptima para el jugador 2, es decir,

$$V^*(\cdot) - \epsilon \leq \inf_{\gamma \in \Gamma} V_\alpha(\cdot, f_\epsilon, \gamma),$$

$$V^*(\cdot) = \sup_{\pi \in \Pi} V_\alpha(\cdot, \pi, g_*).$$

2.4. Estrategias asintóticamente óptimas

En esta sección estudiaremos juegos estocásticos suma cero en el contexto del modelo (1.4). Es decir, supondremos que la dinámica del juego es modelada por medio de ecuaciones en diferencia de la forma (ver Observación 1.1.1):

$$x_{n+1} = F(x_n, a_n, b_n, \xi_n), n = 0, 1, \dots,$$

y más aún, supondremos que la densidad ρ de los vectores aleatorios ξ_n es desconocida por los jugadores. Bajo estas circunstancias, antes de elegir las acciones a_n y b_n en la etapa n , los jugadores deben implementar un método de estimación estadística para obtener los estimadores ρ_n^1 y ρ_n^2 de ρ , y combinarlos con la historia del juego para construir sus estrategias.

Por otro lado, como podemos observar, el criterio de costo descontado depende fuertemente de las decisiones que toman los jugadores en las primeras etapas del juego, precisamente donde la información de la densidad ρ es escasa. Es decir, en las primeras etapas los métodos de estimación proporcionan una información deficiente de la densidad. Esto implica que en general no existen estrategias adaptadas óptimas para los jugadores, y por lo tanto la optimalidad de las estrategias construidas con procesos de estimación y control será estudiado en un sentido asintótico como en la Definición 1.3.3.

Los resultados los obtendremos bajo la siguiente hipótesis.

Hipótesis 4. Existe una función medible $\tilde{\rho}$ en \mathbb{R}^k tal que la densidad ρ satisface las siguientes condiciones:

- (a) $\rho(\cdot) \leq \tilde{\rho}(\cdot)$ casi dondequiera respecto a la medida de Lebesgue.
- (b) Para cada $s \in \mathbb{R}^k$ la función $F(\cdot, \cdot, \cdot, s)$ es continua.
- (c) Mas aún, existen constantes $\beta_0 \in (0, 1)$, $b_0 \geq 0$, $p > 1$, y $M > 0$ tales que para todo $(x, a, b) \in \mathbb{K}$ se tiene que

$$0 \leq r(x, a, b) \leq Mw(x),$$

$$\int_{\mathbb{R}^k} w^p [F(x, a, b, s)] \rho(s) ds \leq \beta_0 w^p(x) + b_0.$$

- (d) Hipótesis 2(a), (b) y (e) se cumplen.

Observación 2.4.1. (a) La Hipótesis 4 (b) implica que la función

$$(x, a, b) \rightarrow \int_{\mathbb{R}^k} v(F(x, a, b, s))\mu(ds) \quad (2.24)$$

es continua para cada función continua y acotada v en X y cada medida de probabilidad $\mu(\cdot)$ en \mathbb{R}^k . Entonces, el kernel estocástico Q dado en (1.3) es débilmente continuo.

(b) Por la desigualdad de Jensen, la Hipótesis 4(c) implica la Hipótesis 2(f) con $\beta := \beta_0^{1/p}$ y $c := b_0^{1/p}$, esto es,

$$\int_{\mathbb{R}^k} w(F(x, a, b, s))\rho(s)ds \leq \beta w(x) + c. \quad (2.25)$$

Entonces por la parte (a) la Hipótesis 4 implica la Hipótesis 2, y por lo tanto, los resultados del Teorema 2.3.5 se cumplen.

(c) Adicionalmente, notemos que con la Hipótesis 4(c) se tiene

$$\sup_{n \geq 0} E_x^{\pi\gamma} [w^n(x_n)] < \infty \text{ y } \sup_{n \geq 0} E_x^{\pi\gamma} [w(x_n)] < \infty, \quad (2.26)$$

para cada par $(\pi, \gamma) \in \Pi \times \Gamma$ y $x \in X$ (ver [9], [13]).

2.4.1. Estimación de la densidad

En esta parte construiremos los estimadores de la densidad ρ . El método de estimación que presentaremos es una adaptación a juegos estocásticos del que se usó en [9, 14] para control adaptado de procesos de Markov, y de esta manera solo expondremos las ideas principales.

Sean $\xi_0, \xi_1, \dots, \xi_n, \dots$ realizaciones independientes de vectores aleatorios con densidad $\rho(\cdot)$, y denotemos por $\widehat{\rho}_n^i(s) := \widehat{\rho}_n^i(s; \xi_0, \xi_1, \dots, \xi_{n-1})$, $s \in \mathbb{R}^k$, $n \in \mathbb{N}$, los estimadores de la densidad para los jugadores $i = 1, 2$ tales que

$$E \|\widehat{\rho}_n^i - \rho\|_1 = E \int_{\mathbb{R}^k} |\widehat{\rho}_n^i(s) - \rho(s)| ds \rightarrow 0, \text{ cuando } n \rightarrow \infty. \quad (2.27)$$

Estimadores que satisfacen la propiedad (2.27) se pueden ver en [6] (Capítulo 9, Teorema 9.2).

Consideremos la clase \mathcal{D} formada por las densidades $\sigma(\cdot)$ que satisfacen las siguientes condiciones:

D.1 $\sigma(\cdot) \leq \tilde{\rho}(\cdot)$ casi dondequiera respecto a la medida de Lebesgue;

D.2 $\int_{\mathbb{R}^k} w[F(x, a, b, s)]\sigma(s)ds \leq \beta w(x) + c$ para todo $(x, a, b) \in \mathbb{K}$, donde las constantes β y c son como en la Observación 2.4.1(b).

La clase de densidades \mathcal{D} es un subconjunto cerrado y convexo de \mathbb{L}_1 bajo la Hipótesis siguiente (ver [9]).

Hipótesis 5. La funcion

$$\psi(s) := \sup_{x \in X} \sup_{a \in A(x)} \sup_{b \in B(x)} \frac{1}{w(x)} w[F(x, a, b, s)], \quad s \in \mathbb{R}^k, \quad (2.28)$$

es finita y satisface la condición

$$\int_{\mathbb{R}^k} \psi^2(s) \tilde{\rho}(s) ds < \infty. \quad (2.29)$$

Entonces de ([6], p. 169), para cada $n \in \mathbb{N}$ existe $\rho_n^i(\cdot) \in \mathcal{D}$ tal que

$$\|\rho_n^i - \tilde{\rho}_n^i\|_1 = \inf_{\sigma \in \mathcal{D}} \|\sigma - \tilde{\rho}_n^i\|_1, \quad i = 1, 2. \quad (2.30)$$

Notemos que $E \|\rho_n^i - \rho\|_1 \rightarrow 0$, dado que

$$\|\rho_n^i - \rho\|_1 \leq \|\rho_n^i - \tilde{\rho}_n^i\|_1 + \|\tilde{\rho}_n^i - \rho\|_1 \leq 2 \|\tilde{\rho}_n^i - \rho\|_1 \quad \forall n \in \mathbb{N}.$$

Las densidades $\{\rho_n^i(\cdot)\}$ serán usadas para obtener las estrategias adaptadas para cada jugador, pero primero expresaremos su propiedad de convergencia usando la siguiente norma, la cual es mas apropiada para nuestros fines. Para una función medible $\sigma : \mathbb{R}^k \rightarrow \mathbb{R}$ definimos

$$\|\sigma\| := \sup_{x \in X} \sup_{a \in A(x), b \in B(x)} \frac{1}{w(x)} \int_{\mathbb{R}^k} w[F(x, a, b, s)] |\sigma(s)| ds. \quad (2.31)$$

Notemos que la condición D.2 garantiza que $\|\sigma\| < \infty$ para cualquier densidad σ en \mathcal{D} .

Lema 2.4.1. Si las Hipótesis 4 y 5 se cumplen, entonces

$$E \|\rho_n^i - \rho\| \rightarrow 0 \text{ cuando } n \rightarrow \infty \text{ para } i = 1, 2.$$

Demostración. De las desigualdad de Cauchy-Schwarz, y las relaciones (2.28) y (2.29),

$$\begin{aligned}
\|\rho_n^i - \rho\| &\leq \int_{\mathbb{R}^k} \psi(s) |\rho_n^i(s) - \rho(s)| ds \\
&\leq \left(\int_{\mathbb{R}^k} \psi^2(s) |\rho_n^i(s) - \rho(s)| ds \right)^{1/2} \left(\int_{\mathbb{R}^k} |\rho_n^i(s) - \rho(s)| ds \right)^{1/2} \\
&\leq 2^{1/2} \left(\int_{\mathbb{R}^k} \psi^2(s) \tilde{\rho}(s) ds \right)^{1/2} \left(\int_{\mathbb{R}^k} |\rho_n^i(s) - \rho(s)| ds \right)^{1/2} \\
&\leq C' \left(\int_{\mathbb{R}^k} |\rho_n^i(s) - \rho(s)| ds \right)^{1/2} = C' \|\rho_n^i - \rho\|_1,
\end{aligned}$$

donde $C' := (2 \int_{\mathbb{R}^k} \psi^2(s) \tilde{\rho}(s) ds)^{1/2}$. El resultado se sigue de la última desigualdad dado que $E \|\rho_n^i - \rho\|_1 \rightarrow 0$ cuando $n \rightarrow \infty$. \square

2.4.2. Construcción de estrategias

Como se comentó al principio de la sección, la construcción de las estrategias se hará bajo la Hipótesis 4, la cual implica la Hipótesis 2 (ver Observación 2.4.1).

Sea $\theta \in (\alpha, 1)$ y $d := c[\theta/\alpha - 1]^{-1}$ constantes fijas. Definimos la función

$$\widehat{w}(\cdot) := w(\cdot) + d, \quad (2.32)$$

así como la correspondiente \widehat{w} -norma

$$\|u\|_{\widehat{w}} = \sup_{x \in X} \frac{|u(x)|}{\widehat{w}(x)} < \infty,$$

para $u \in \mathbb{B}_w(X)$. Observemos que las normas $\|\cdot\|_{\widehat{w}}$ y $\|\cdot\|_w$ son equivalentes dado que

$$\|u\|_{\widehat{w}} \leq \|u\|_w \leq (1 + d) \|u\|_{\widehat{w}}.$$

Por lo tanto, $u \in \mathbb{B}_w(X)$ si y solo si $\|u\|_{\widehat{w}} < \infty$.

Por otro lado, de [27] tenemos que si $\rho \in \mathcal{D}$ entonces la función \widehat{w} satisface la desigualdad

$$\alpha \int_{\mathbb{R}^k} \widehat{w}(F(x, a, b, s)) \rho(s) ds \leq \theta \widehat{w}(x) \quad \forall (x, a, b) \in \mathbb{K}.$$

De aquí, siguiendo argumentos similares a la demostración del Lema 2.3.1, podemos demostrar que bajo la Hipótesis 4 y el hecho de que $\rho \in \mathcal{D}$ esta desigualdad implica que el operador T_α es de contracción de $\mathbb{C}_w(X)$ en sí mismo con módulo θ respecto a la \widehat{w} -norma. Por el Teorema de Punto Fijo de Banach (ver Proposición C.1.1) T_α tiene un único punto fijo, y por el Teorema 2.3.5, este punto fijo es el valor del juego V_α^* , es decir,

$$TV_\alpha^* = V_\alpha^* \quad (2.33)$$

Además,

$$\|T_\alpha^n u - V_\alpha^*\|_{\widehat{w}} \rightarrow 0 \quad \forall u \in \mathbb{C}_w(X),$$

lo cual implica que

$$\|T_\alpha^n u - V_\alpha^*\|_w \rightarrow 0 \quad \forall u \in \mathbb{C}_w(X).$$

Las estrategias adaptadas serán obtenidas usando el método de iteración de valores cuando los jugadores usan los estimadores de densidad $\{\rho_n^i\}$ definidos en (2.30) como la densidad verdadera. Entonces introducimos, para $i = 1, 2$, los operadores

$$T_n^i v(x) := \inf_{\lambda \in \mathbb{P}_A(x)} \sup_{\mu \in \mathbb{P}_B(x)} \left[r(x, \mu, \lambda) + \alpha \int_{\mathbb{R}^k} v(F(x, \mu, \lambda, s)) \rho_n^i(s) ds \right],$$

para $x \in X, n \in \mathbb{N}$. Notemos que bajo la Hipótesis 4 y Observación 2.4.1 estos operadores son de contracción del espacio $(\mathbb{C}_w(X), \|\cdot\|_{\widehat{w}})$ en sí mismo, con módulo θ . Así, en este caso, las funciones de iteración de valores “minimax” para el jugador i ($= 1, 2$)

$$U_n^i := T_n^i U_{n-1}^i, \quad n \in \mathbb{N}, \quad U_0^i \equiv 0, \quad (2.34)$$

pertenecen al espacio $\mathbb{C}_w(X)$. Más aún, existe una sucesión $\{\lambda_n^*\} \subset \mathbb{F}_{\mathbb{P}_B}$ para el jugador 2 tal que

$$U_n^2(x) = \sup_{\mu \in \mathbb{P}_A(x)} \left[r(x, \mu, \lambda_n^*) + \alpha \int_{\mathbb{R}^k} U_{n-1}^2(F(x, \mu, \lambda_n^*, s)) \rho_n^2(s) ds \right], \quad (2.35)$$

para todo $x \in X, n \in \mathbb{N}$. De manera análoga, existe una sucesión $\{\mu_n^*\} \subset \mathbb{F}_{\mathbb{P}_A}$ tal que

$$U_n^1(x) = \inf_{\lambda \in \mathbb{P}_B(x)} \left[r(x, \mu_n^*, \lambda) + \alpha \int_{\mathbb{R}^k} U_{n-1}^1(F(x, \mu_n^*, \lambda, s)) \rho_n^1(s) ds \right], \quad (2.36)$$

para todo $x \in X$, $n \in \mathbb{N}$. Además, se puede mostrar que para algunas constantes positivas C_1 y C_2 ,

$$|U_n^i(x)| \leq C_i w(x) \quad \forall n \in \mathbb{N}_0, \quad x \in X, \quad i = 1, 2. \quad (2.37)$$

Finalmente, tenemos el resultado principal de esta sección.

Teorema 2.4.2. Supongamos que las Hipótesis 4 y 5 se cumplen. Entonces las estrategias $\pi^* = \{\mu_n^*\}$ y $\gamma^* = \{\lambda_n^*\}$ son AD -óptimas para los jugadores 1 y 2, respectivamente. Así, en particular,

$$\lim_{n \rightarrow \infty} E_x^{\pi^*, \gamma^*} D(x_n, a_n, b_n) = 0 \quad \forall x \in X.$$

2.4.3. Demostración del resultado principal

La demostración del Teorema 2.4.2 es consecuencia de los siguientes lemas.

Lema 2.4.3. Para cada $i = 1, 2$,

$$\|T_n^i V_\alpha^* - TV_\alpha^*\|_w \leq \alpha \|V_\alpha^*\|_w \|\rho_n^i - \rho\| \quad \forall n \in \mathbb{N}. \quad (2.38)$$

Así, para cada $(\pi, \gamma) \in \Pi \times \Gamma$ y $x \in X$,

$$\lim_{n \rightarrow \infty} E_x^{\pi, \gamma} \|T_n^i V_\alpha^* - TV_\alpha^*\|_w = 0. \quad (2.39)$$

Demostración. Notemos que (2.39) se sigue de (2.38) y del Lema 2.4.1, ya que $\|\rho_n^i - \rho\|$ no depende de $(\pi, \gamma) \in \Pi \times \Gamma$ y $x \in X$. Para demostrar (2.38), notemos que para todo $x \in X$, $n \in \mathbb{N}$, e $i = 1, 2$, tenemos

$$\begin{aligned}
& |T_n^i V_\alpha^*(x) - TV_\alpha^*(x)| \\
& \leq \sup_{\lambda \in \mathbb{P}_B(x)} \sup_{\mu \in \mathbb{P}_A(x)} \left| \alpha \int_{\mathbb{R}^k} V_\alpha^*(F(x, \mu, \lambda, s)) \rho_n^i(s) ds - \alpha \int_{\mathbb{R}^k} V_\alpha^*(F(x, \mu, \lambda, s)) \rho(s) ds \right| \\
& \leq \sup_{\lambda \in \mathbb{P}_B(x)} \sup_{\mu \in \mathbb{P}_A(x)} \alpha \int_{\mathbb{R}^k} |V_\alpha^*(F(x, \mu, \lambda, s))| |\rho_n^i(s) - \rho(s)| ds \\
& \leq \alpha \|V_\alpha^*\|_w \sup_{\lambda \in \mathbb{P}_B(x)} \sup_{\mu \in \mathbb{P}_A(x)} \int_{\mathbb{R}^k} w(F(x, \mu, \lambda, s)) |\rho_n^i(s) - \rho(s)| ds \\
& \leq \alpha \|V_\alpha^*\|_w \|\rho_n^i - \rho\| w(x).
\end{aligned}$$

Esta desigualdad junto con la definición de $\|\cdot\|_w$ demuestran (2.38). \square

Lema 2.4.4. Para cada $i = 1, 2$, $(\pi, \gamma) \in \Pi \times \Gamma$, y $x \in X$,

$$\lim_{n \rightarrow \infty} E_x^{\pi, \gamma} \|U_n^i - V_\alpha^*\|_w = 0$$

Demostración. Primero notemos que para todo $n \in \mathbb{N}$ e $i = 1, 2$,

$$\begin{aligned}
\|U_n^i - V_\alpha^*\|_{\widehat{w}} &= \|T_n^i U_{n-1}^i - TV_\alpha^*\|_{\widehat{w}} \quad (\text{ver (2.32) para la definición de } \widehat{w}) \\
&\leq \|T_n^i U_{n-1}^i - T_n^i V_\alpha^*\|_{\widehat{w}} + \|T_n^i V_\alpha^* - TV_\alpha^*\|_{\widehat{w}}.
\end{aligned}$$

Entonces, dado que T_n^i es un operador de contracción con módulo θ con respecto a la \widehat{w} -norma, del Lema 2.4.3 se sigue que

$$\|U_n^i - V_\alpha^*\|_{\widehat{w}} \leq \theta \|U_{n-1}^i - V_\alpha^*\|_{\widehat{w}} + \alpha \|V_\alpha^*\|_w \|\rho_n^i - \rho\|. \quad (2.40)$$

Más aún, como $E \|\rho_n^i - \rho\| \rightarrow 0$ existe una constante positiva M' tal que

$$E_x^{\pi, \gamma} \|U_n^i - V_\alpha^*\|_{\widehat{w}} \leq \theta E_x^{\pi, \gamma} \|U_{n-1}^i - V_\alpha^*\|_{\widehat{w}} + M'.$$

Así, iterando esta desigualdad obtenemos

$$\begin{aligned}
 E_x^{\pi, \gamma} \|U_n^i - V_\alpha^*\|_{\hat{w}} &\leq \theta^n \|V_\alpha^*\|_{\hat{w}} + M' \sum_{i=0}^{n-1} \theta^i \\
 &\leq \frac{M' + \|V_\alpha^*\|_{\hat{w}}}{1 - \theta},
 \end{aligned}$$

lo cual implica que

$$L := \limsup_{n \rightarrow \infty} E_x^{\pi, \gamma} \|U_n^i - V_\alpha^*\|_{\hat{w}} < \infty.$$

Ahora tomando esperanza en (2.40) y limsup cuando n tiende a infinito, tenemos que $0 \leq L \leq \theta L$, con lo cual obtenemos $L = 0$. Esto demuestra el lema. \square

De acuerdo al Teorema 2.4.2, el objetivo es demostrar que las estrategias π^* y γ^* son AD -óptimas (ver Definición 1.3.3). Para esto definimos “las funciones de discrepancia aproximantes” para el jugador $i = 1, 2$, como:

$$D_n^i(x, \mu, \lambda) := r(x, \mu, \lambda) + \alpha \int_{\mathbb{R}^k} U_{n-1}^i(F(x, \mu, \lambda, s)) \rho_n^i(s) ds - U_n^i(x)$$

para $x \in X$, $\mu \in \mathbb{P}_A(x)$, $\lambda \in \mathbb{P}_B(x)$.

Lema 2.4.5. Para todo $x \in X$, $n \in \mathbb{N}$, se tiene que

$$\sup_{\mu \in \mathbb{P}_A(x)} \sup_{\lambda \in \mathbb{P}_B(x)} |D(x, \mu, \lambda) - D_n^i(x, \mu, \lambda)| \leq w(x) \eta_n^i,$$

donde

$$\eta_n^i := \|U_n^i - V_\alpha^*\|_w + (\beta + c) \|U_n^i - V_\alpha^*\|_w + \alpha \|V_\alpha^*\|_w \|\rho_n^i - \rho\| \quad (2.41)$$

para todo $n \in \mathbb{N}$.

Demostración. Sea $x \in X$, $\mu \in \mathbb{P}_A(x)$, $\lambda \in \mathbb{P}_B(x)$ y $n \in \mathbb{N}$ fijo y arbitrario, y denotemos $R^i(x, \mu, \lambda) := |D(x, \mu, \lambda) - D_n^i(x, \mu, \lambda)|$. Entonces, observemos que

$$\begin{aligned}
 R^i(x, \mu, \lambda) &\leq |U_n^i(x) - V_\alpha^*(x)| \\
 &\quad + \alpha \left| \int_{\mathbb{R}^k} U_{n-1}^i(F(x, \mu, \lambda, s)) \rho_n^i(s) ds - \int_{\mathbb{R}^k} V_\alpha^*(F(x, \mu, \lambda, s)) \rho(s) ds \right| \\
 &\leq |U_n^i(x) - V_\alpha^*(x)| + \alpha \int_{\mathbb{R}^k} V_\alpha^*(F(x, \mu, \lambda, s)) |\rho(s) - \rho_n^i(s)| ds \\
 &\quad + \alpha \int_{\mathbb{R}^k} |U_n^i(F(x, \mu, \lambda, s)) - V_\alpha^*(F(x, \mu, \lambda, s))| \rho_n^i(s) ds \\
 &\leq |U_n^i(x) - V_\alpha^*(x)| + \alpha \|V_\alpha^*\|_w \|\rho_n^i - \rho\| w(x) \\
 &\quad + \|U_{n-1}^i - V_\alpha^*\|_w \int_{\mathbb{R}^k} w(F(x, \mu, \lambda, s)) \rho_n^i(s) ds.
 \end{aligned}$$

Ahora, como $\rho_n^i(\cdot) \in \mathcal{D}$, de la Condición D.2, tenemos que

$$\begin{aligned}
 \int_{\mathbb{R}^k} w[F(x, a, b, s)] \rho_n^i(s) ds &\leq \beta w(x) + c \\
 &\leq [\beta + c]w(x),
 \end{aligned}$$

lo cual implica

$$\begin{aligned}
 R^i(x, \mu, \lambda) &\leq |U_n^i(x) - V_\alpha^*(x)| + \alpha \|V_\alpha^*\|_w \|\rho_n^i - \rho\|_w w(x) \\
 &\quad + \|U_{n-1}^i - V_\alpha^*\|_w (\beta + c)w(x).
 \end{aligned}$$

Así,

$$\sup_{\mu \in \mathbb{P}_A(x)} \sup_{\lambda \in \mathbb{P}_B(x)} |D(x, \mu, \lambda) - D_n^i(x, \mu, \lambda)| \leq w(x) \eta_n^i \quad \forall x \in X, \quad n \in \mathbb{N}.$$

□

Lema 2.4.6. Dadas las estrategias $\pi^* = \{\mu_n^*\}$ y $\gamma^* = \{\lambda_n^*\}$ como en (2.35) y (2.36), tenemos

$$\begin{aligned}
 -w(x_n) \eta_n^1 &\leq D(x, \mu_n^*, \lambda) \quad \forall \lambda \in \mathbb{P}_B(x), \quad n \in \mathbb{N}, \\
 D(x, \mu, \lambda_n^*) &\leq w(x_n) \eta_n^2 \quad \forall \mu \in \mathbb{P}_A(x), \quad n \in \mathbb{N}.
 \end{aligned}$$

Demostración. Las desigualdades se siguen directamente del Lema 2.4.5, notando que

$$\inf_{\lambda \in \mathbb{P}_B(x)} D_n^1(x, \mu_n^*, \lambda) = 0,$$

$$\sup_{\mu \in \mathbb{P}_A(x)} D_n^2(x, \mu, \lambda_n^*) = 0$$

para todo $x \in X$, $n \in \mathbb{N}$.

□

Observación 2.4.2. Observemos que de (2.41), y los Lemas 2.4.1 y 2.4.4,

$$\lim_{n \rightarrow \infty} E_x^{\pi, \gamma} \eta_n^i = 0 \quad \text{para } i = 1, 2, (\pi, \gamma) \in \Pi \times \Gamma \text{ y } x \in X.$$

Así,

$$\eta_n^i \xrightarrow{P_x^{\pi, \gamma}} 0 \quad \text{para } i = 1, 2.$$

Más aún, dado que $\|\sigma\| < \infty$ para $\sigma \in \mathcal{D}$, de (2.37) tenemos

$$k_i := \sup_n \eta_n^i < \infty, \quad i = 1, 2.$$

Demostración del Teorema 2.4.2. Del Lema 2.4.6, es suficiente demostrar, para $i = 1, 2$, que

$$E_x^{\pi, \gamma} w(x_n) \eta_n^i \rightarrow 0 \quad \forall x \in X, \pi \in \Pi, \gamma \in \Gamma.$$

Para esto, primero mostraremos que el proceso $\{w(x_n) \eta_n^i\}$ converge a cero en probabilidad respecto a $P_x^{\pi, \gamma}$ para todo $x \in X$, $\pi \in \Pi$, $\gamma \in \Gamma$. Sean l_1 y l_2 constantes positivas arbitrarias, entonces observemos que

$$P_x^{\pi, \gamma} [w(x_n) \eta_n^i > l_1] \leq P_x^{\pi, \gamma} \left[\eta_n^i > \frac{l_1}{l_2} \right] + P_x^{\pi, \gamma} [w(x_n) > l_2] \quad \forall x \in X, n \in \mathbb{N}.$$

Así, por la desigualdad de Chebyshev y (2.26) obtenemos

$$\begin{aligned} P_x^{\pi, \gamma} [w(x_n) \eta_n^i > l_1] &\leq P_x^{\pi, \gamma} \left[\eta_n^i > \frac{l_1}{l_2} \right] + \frac{1}{l_2} E_x^{\pi, \gamma} w(x_n) \\ &\leq P_x^{\pi, \gamma} \left[\eta_n^i > \frac{l_1}{l_2} \right] + \frac{\bar{M}}{l_2}, \end{aligned}$$

para alguna constante $\bar{M} < \infty$. Así, por la Observación 2.4.3,

$$\limsup_{n \rightarrow \infty} P_x^{\pi, \gamma} [w(x_n) \eta_n^i > l_1] \leq \frac{\bar{M}}{l_2}.$$

Como l_2 es arbitrario, tenemos que

$$\lim_{n \rightarrow \infty} P_x^{\pi, \gamma} [w(x_n) \eta_n^i > l_1] = 0, \quad (2.42)$$

es decir, $\{w(x_n)\eta_n^i\}$ converge a cero en probabilidad.

Por otro lado, de la Observación 2.4.1(c) y (2.37), tenemos que la desigualdad

$$\sup_{n \in \mathbb{N}} E_x^{\pi, \gamma} [w(x_n)\eta_n^i]^p \leq k_i^p \sup_{n \in \mathbb{N}} E_x^{\pi, \gamma} w^p(x_n) < \infty$$

es válida para $x \in X$, $\pi \in \Pi$, $\gamma \in \Gamma$. Así de [3] la desigualdad anterior implica que el proceso $\{w(x_n)\eta_n^i\}$ es $P_x^{\pi, \gamma}$ -uniformemente integrable.

Finalmente, usando la integrabilidad uniforme del proceso $\{w(x_n)\eta_n^i\}$ y la convergencia a cero, obtenemos

$$E_x^{\pi, \gamma} w(x_n)\eta_n^i \rightarrow 0.$$

□

Concluimos el capítulo con la siguiente observación.

Observación 2.4.3. Bajo la Hipótesis 3, propuesta por Küenle [18] podemos establecer un resultado similar al Teorema 2.4.2. Basándonos en el Teorema 2.3.5 tenemos que:

Existe una sucesión $\{\lambda_n^*\} \subset \mathbb{F}_{\mathbb{P}_B}$ para el jugador 2 tal que se satisface (2.35) para todo $x \in X$, $n \in \mathbb{N}$. De manera análoga, para cualquier sucesión de números positivos $\{\epsilon_n\}$ que converge a cero existe una sucesión $\{\mu_n^*\} \subset \mathbb{F}_{\mathbb{P}_A}$ tal que

$$U_n^1(x) - \epsilon_n \leq \inf_{\lambda \in \mathbb{P}_B(x)} \left[r(x, \mu_n^*, \lambda) + \alpha \int_{\mathbb{R}^k} U_{n-1}^1(F(x, \mu_n^*, \lambda, s)) \rho_n^1(s) ds \right] \quad (2.43)$$

para todo $x \in X$, $n \in \mathbb{N}$. Además, se puede mostrar (2.37).

De esta forma se tiene el siguiente resultado, cuya demostración es similar al Teorema 2.4.2.

Teorema 2.4.7. Supongamos que las Hipótesis 3, 4(a)-(c) y 5 se cumplen. Entonces las estrategias $\pi^* = \{\mu_n^*\}$ y $\gamma^* = \{\lambda_n^*\}$ son AD -óptimas para los jugadores 1 y 2, respectivamente. Así, en particular,

$$\lim_{n \rightarrow \infty} E_x^{\pi^*, \gamma^*} D(x_n, a_n, b_n) = 0 \quad \forall x \in X.$$

Capítulo 3

Criterio de optimalidad promedio

En este capítulo se analiza el criterio de optimalidad promedio.

Para cada estado inicial $x_0 = x$, estrategias $\pi \in \Pi$, $\gamma \in \Gamma$, y $n \in \mathbb{N}$, recordemos que el pago total esperado hasta la n -ésima etapa es:

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

Así mismo, el pago promedio esperado con horizonte infinito se define como:

$$J(x, \pi, \gamma) = \liminf_{n \rightarrow \infty} \frac{J_n(x, \pi, \gamma)}{n}. \quad (3.1)$$

Como es usual en el estudio del criterio de optimalidad promedio, necesitamos imponer hipótesis de ergodicidad para el análisis asintótico. Bajo estas hipótesis demostraremos la existencia del valor del juego, así como de estrategias óptimas para los jugadores. Para esto usaremos el enfoque del “factor de descuento desvaneciente”, el cual consiste, en términos generales, en estudiar el caso promedio como límite del caso descontado. En este sentido, estaremos haciendo referencia a los resultados del capítulo anterior.

Finalmente, consideraremos juegos cuya dinámica está modelada por medio de ecuaciones en diferencia con distribución del ruido aleatorio desconocida. En este caso, a diferencia del criterio descontado, demostraremos que aplicando procesos de estimación y control obtenemos estrategias adaptadas óptimas.

3.1. Teorema de existencia

Los resultados de este capítulo se obtendrán en el contexto de la siguiente hipótesis.

Hipótesis 6. (a) Las correspondencias $x \mapsto A(x)$ y $x \mapsto B(x)$ son continuas y con valores compactos.

(b) La función $r : \mathbb{K} \rightarrow \mathbb{R}$ es continua.

(c) El kernel Q es débilmente continuo.

(d) Existe una función $w : X \rightarrow [1, \infty)$ tal que $|r(x, a, b)| \leq w(x)$ para cada $(x, a, b) \in \mathbb{K}$

(e) La función

$$(x, a, b) \mapsto \int_X w(y)Q(dy|x, a, b)$$

es continua en \mathbb{K} .

(f) Existe una función de Borel $\delta : \mathbb{K} \rightarrow [0, 1]$ y una medida de probabilidad φ en X tal que

1. $Q(B|x, a, b) \geq \delta(x, a, b)\varphi(B)$ para cada conjunto de Borel $B \subset X$ y $x \in X$.
2. $\int_X \inf_{a \in A(x)} \inf_{b \in B(x)} \delta(x, a, b)\varphi(dx) > 0$.
3. $\varphi(w) := \int_X w(x)\varphi(dx) < \infty$.
4. Para algún $\beta \in (0, 1)$ y cada $(x, a, b) \in \mathbb{K}$, se tiene que

$$\int_X w(y)Q(dy|x, a, b) \leq \beta w(x) + \delta(x, a, b)\varphi(w).$$

Observemos que bajo la Hipótesis 6 los resultados del Teorema 2.3.5 son válidos.

Teorema 3.1.1. Supongamos que la Hipótesis 6 se cumple. Entonces el juego estocástico con pago promedio tiene un valor j^* , que es independiente del estado inicial, y ambos jugadores tienen estrategias óptimas estacionarias. Más aún, $j^* = \lim_{\alpha \rightarrow 1} (1 - \alpha)V_\alpha^*(x)$ para cualquier $x \in X$.

Demostración. Por el Teorema 2.3.5 para cada $\alpha \in (0, 1)$, existe el valor del juego α -descontado V_α^* y estrategias $f_\alpha \in \Pi_S$, $g_\alpha \in \Gamma_S$ que satisfacen (2.21). Fijemos

$x^* \in X$ y consideremos una sucesión $\{\alpha_n\}$ de factores de descuento que converge a 1. Definimos

$$h_n(x) := V_{\alpha_n}^*(x) - V_{\alpha_n}(x^*), \quad J_n := (1 - \alpha_n)V_{\alpha_n}^*(x^*).$$

Entonces por (2.21), se sigue que, para cada $n \in \mathbb{N}$,

$$\begin{aligned} J_n + h_n(x) &= \min_{\lambda \in \mathbb{P}_B(x)} \max_{\mu \in \mathbb{P}_A(x)} \left[r(x, \mu, \lambda) + \alpha_n \int_X h_n(y) Q(dy|x, \mu, \lambda) \right] \\ &= \max_{\mu \in \mathbb{P}_A(x)} \left[r(x, \mu, g_{\alpha_n}(x)) + \alpha_n \int_X h_n(y) Q(dy|x, \mu, g_{\alpha_n}) \right], \quad x \in X. \end{aligned} \quad (3.2)$$

Por el Lema 7 en [24], la sucesión $\{J_n\}$ es acotada y, sin pérdida de generalidad podemos suponer que J_n converge a algún número J^* . Más aún, por el Lema 8 en [24], se sigue que $\|h_n\|_w \leq c$ para alguna constante c . Ahora, fijemos un estado $x_0 \in X$ y consideremos una sucesión arbitraria de estados $\{x_n\}$ tan que $x_n \rightarrow x_0$ cuando $n \rightarrow \infty$. Entonces de (3.2), haciendo $g_n := g_{\alpha_n}$, obtenemos

$$J_n + h_n(x_n) = \max_{\mu \in \mathbb{P}_A(x_n)} \left[r(x_n, \mu, g_n(x_n)) + \alpha_n \int_X h_n(y) Q(dy|x_n, \mu, g_n(x_n)) \right]. \quad (3.3)$$

Sea $\{n_k\}$ una subsucesión de enteros positivos para los cuales

$$\liminf_{n \rightarrow \infty} h_n(x_n) = \lim_{k \rightarrow \infty} h_{n_k}(x_{n_k}).$$

Es claro que $\lim_{k \rightarrow \infty} J_{n_k} = J^*$ y

$$J^* + \liminf_{n \rightarrow \infty} h_n(x_n) = \liminf_{n \rightarrow \infty} [J_n + h_n(x_n)] = \lim_{k \rightarrow \infty} [J_{n_k} + h_{n_k}(x_{n_k})].$$

Entonces de (3.3), obtenemos

$$\begin{aligned} J^* + \liminf_{n \rightarrow \infty} h_n(x_n) \\ = \lim_{k \rightarrow \infty} \max_{\mu \in \mathbb{P}_A(x_{n_k})} \left[r(x_{n_k}, \mu, g_{n_k}(x_{n_k})) + \alpha_{n_k} \int_X h_{n_k}(y) Q(dy|x_{n_k}, \mu, g_{n_k}(x_{n_k})) \right]. \end{aligned}$$

Sea Π_{CS} el conjunto de las estrategias estacionarias continuas para el jugador 1. Por las Proposiciones A.2.1 y A.2.2, Π_{CS} es no-vacío. Elijamos $f \in \Pi_{CS}$. Entonces, tenemos

$$\begin{aligned}
& j^* + \liminf_{n \rightarrow \infty} h_n(x_n) \\
& \geq \liminf_{k \rightarrow \infty} \left[r(x_{n_k}, f(x_{n_k}), g_{n_k}(x_{n_k})) + \alpha_{n_k} \int_X h_{n_k}(y) Q(dy | x_{n_k}, f(x_{n_k}), g_{n_k}(x_{n_k})) \right].
\end{aligned} \tag{3.4}$$

Notemos que $Z := \{x_0\} \cup \{x_n\}$ es compacto en X . Además sabemos que la correspondencia $x \mapsto \mathbb{P}_B(x)$ es continua y toma valores compactos. Esto junto con el Teorema de Berge's (Teorema C.3.3) implican que $\bigcup_{z \in Z} \mathbb{P}_B(z)$ es compacto en \mathbb{P}_B . Por lo tanto, $\{g_{n_k}(x_{n_k})\}$ tiene una subsucesión convergente a algún $\lambda_0 \in \mathbb{P}_B$. Sin pérdida de generalidad, sea $g_{n_k}(x_{n_k}) \rightarrow \lambda_0$, cuando $k \rightarrow \infty$. Por la continuidad de $x \mapsto \mathbb{P}_B(x)$, $\lambda_0 \in \mathbb{P}_B(x_0)$.

Por otro lado, $f(x_{n_k}) \rightarrow f(x_0) \in \mathbb{P}_A(x_0)$, porque $f \in \Pi_{CS}$ y además

$$Q(\cdot | x_{n_k}, f(x_{n_k}), g_{n_k}(x_{n_k})) \rightarrow Q(\cdot | x_0, f(x_0), \lambda_0)$$

débilmente cuando $k \rightarrow \infty$. Por (3.4) y Lema B.2.3, concluimos que

$$\begin{aligned}
& j^* + \liminf_{n \rightarrow \infty} h_n(x_n) \\
& \geq \liminf_{k \rightarrow \infty} r(x_{n_k}, f(x_{n_k}), g_{n_k}(x_{n_k})) + \liminf_{k \rightarrow \infty} \int_X h_{n_k}(y) Q(dy | x_{n_k}, f(x_{n_k}), g_{n_k}(x_{n_k})) \\
& \geq r(x_0, f(x_0), \lambda_0) + \liminf_{k \rightarrow \infty} \int_X h_{n_k}(y) Q(dy | x_{n_k}, f(x_{n_k}), g_{n_k}(x_{n_k})) \\
& \geq r(x_0, f(x_0), \lambda_0) + \int_X \bar{h}_*(y) Q(dy | x_0, f(x_0), \lambda),
\end{aligned}$$

donde \bar{h}_* es el \liminf generalizado (ver apéndice B.2) de la sucesión $\bar{h}_k = h_{n_k}$. Sea $h_*(x_0)$ el \liminf generalizado de $\{h_n\}$. Entonces $h_* \leq \bar{h}_*$, y aplicando esto a lo obtenido anteriormente tenemos

$$j^* + \liminf_{n \rightarrow \infty} h_n(x_n) \geq r(x_0, f(x_0), \lambda_0) + \int_X h_*(y) Q(dy | x_0, f(x_0), \lambda).$$

Además, por el hecho de que un selector continuo $f \in \Pi_{CS}$ puede ser elegido de tal manera que un valor arbitrario de $\mathbb{P}_A(x)$ es asignado al punto x_0 (ver Lema A.2.2), concluimos que

$$\begin{aligned}
 j^* + \liminf_{n \rightarrow \infty} w_n(x_n) &\geq \sup_{f \in \Pi_{CS}} \left[r(x_0, f(x_0), \lambda_0) + \int_X h_*(y) Q(dy|x_0, f(x_0), \lambda_0) \right] \\
 &= \sup_{\mu \in \mathbb{P}_A(x_0)} \left[r(x_0, \mu, \lambda) + \int_X h_*(y) Q(dy|x_0, \mu, \lambda) \right].
 \end{aligned}$$

Como h_* es semicontinua inferiormente (ver Lema B.2.1), podemos escribir

$$j^* + \liminf_{n \rightarrow \infty} h_n(x_n) \geq \min_{\lambda \in \mathbb{P}_B(x_0)} \sup_{\mu \in \mathbb{P}_A(x_0)} \left[r(x_0, \mu, \lambda) + \int_X h_*(y) Q(dy|x_0, \mu, \lambda) \right],$$

y por la definición (B.2), tenemos

$$j^* + h_*(x_0) \geq \min_{\lambda \in \mathbb{P}_B(x_0)} \sup_{\mu \in \mathbb{P}_A(x_0)} \left[r(x_0, \mu, \lambda) + \int_X h_*(y) Q(dy|x_0, \mu, \lambda) \right]. \quad (3.5)$$

Dado que x_0 fue elegida arbitrariamente, entonces en (3.5) podemos reemplazar x_0 por cualquier $x \in X$.

Ahora, como la función h_* es semicontinua inferiormente, por (3.5) y un teorema de selección medible minimax (ver Teorema C.2.1), existe alguna $g^* \in \Gamma_S$ tal que

$$j^* + h_*(x) \geq r(x, \mu, g^*(x)) + \int_X h_*(y) Q(dy|x, \mu, g^*(x)), \quad (3.6)$$

para cada $x \in X$ y $\mu \in \mathbb{P}_A(x)$. Iterando (3.6), podemos mostrar de manera estándar que

$$j^* \geq \sup_{\pi \in \Pi} J(x, \pi, g^*) \geq U(x), \quad (3.7)$$

para cada $x \in X$ (donde U el valor superior del juego).

Por otro lado, sea $f_{\alpha_n} = f_n$ una estrategia óptima estacionaria para el jugador 1 en el juego estocástico α_n -descontado. De acuerdo a (3.6) necesitamos mostrar que $j^* \leq L(x)$. De (3.2) y el teorema de minimax de Fan (ver Teorema C.2.1), se puede ver fácilmente que

$$j_n + h_n(x_n) = \min_{\lambda \in \mathbb{P}_B(x_n)} \left[r(x_n, f_n(x_n), \lambda) + \alpha_n \int_X h_n(y) Q(dy|x_n, f_n(x_n), \lambda) \right].$$

Supongamos que $x_n \rightarrow x_0$ y consideremos el \limsup generalizado $w^*(x_0)$. Proce-
diendo de manera similar a lo anterior, tomamos de nuevo una subsucesión $\{n_k\}$ de
enteros positivos tales que

$$\limsup_{n \rightarrow \infty} h_n(x_n) = \lim_{k \rightarrow \infty} h_{n_k}(x_{n_k}).$$

Entonces usando el Lema B.2.3 y (2.21) existe $f^* \in \Pi_{CS}$ tal que

$$j^* + h^*(x) \leq r(x, f^*(s), \lambda) + \int_X h^*(y) Q(dy|x, f^*(x), \lambda) \quad (3.8)$$

para cada $x \in X$ y $\lambda \in \mathbb{P}_B(x)$. De nuevo por argumentos de programación dinámica,
basandonos en (3.8), se muestra que

$$j^* \leq \inf_{\gamma \in \Gamma} J(x, f^*, \gamma) \leq L(x), \quad (3.9)$$

para cada $x \in X$. Por (3.7) y (3.9), concluimos

$$j^* = \sup_{\pi \in \Pi} J(x, \pi, g^*) = \inf_{\gamma \in \Gamma} J(x, f^*, \gamma) = J(x, f^*, g^*), \quad (3.10)$$

esto es, el juego tiene un valor $J^*(x) = j^*$, y f^*, g^* son estrategias óptimas estacio-
narias para los jugadores 1 y 2, respectivamente. □

Observación 3.1.1. Para la demostración del Teorema 3.1.1 se aplicó el enfoque
del “factor de descuento desvaneciente”, es decir, los resultados del caso promedio los
obtuvimos haciendo $\alpha \nearrow 1$ del caso α -descuento. En particular observemos que si
definimos, para $\alpha \in (0, 1)$ y $z \in X$ un estado fijo y arbitrario,

$$j_\alpha := (1 - \alpha)V_\alpha^*(z) \text{ y } h_\alpha(x) := V_\alpha^*(x) - V_\alpha^*(z), \quad (3.11)$$

entonces, por el Teorema 3.1.1,

$$\lim_{t \rightarrow \infty} j_{\alpha_t} = j^*, \quad (3.12)$$

para cualquier sucesión $\{\alpha_t\}$ de factores de descuento tal que $\alpha_t \nearrow 1$. Mas aún,

$$\sup_{\alpha \in (0, 1)} \|h_\alpha\|_w < \infty. \quad (3.13)$$

3.2. Estimación de la densidad

En esta sección presentaremos el proceso de estimación de la densidad ρ de los vectores aleatorios ξ_n , cuando la dinámica del juego es modelada por ecuaciones en diferencia introducidas en la Observación 1.1.1.

Escencialmente, el método de estimación es muy similar al que presentamos para el caso descontado, para lo cual expondremos las ideas principales para una fácil referencia a lo largo del capítulo.

Hipótesis 7. Existe una función medible $\tilde{\rho} \in \mathbb{R}^k$ tal que la densidad ρ satisface las siguientes condiciones:

- (a) $\rho(\cdot) \leq \tilde{\rho}(\cdot)$ casi dondequiera respecto a la medida de Lebesgue;
- (b) Para cada $s \in \mathbb{R}^k$, $F(\cdot, \cdot, \cdot, s)$ es continua.
- (c) Sea w la función en la Hipótesis 6. Entonces existen constantes $\beta_0 \in (0, 1)$, $b_0 \geq 0$, $p > 1$ y $M > 0$ tales que para todo $(x, a, b) \in \mathbb{K}$ se tiene que

$$0 \leq r(x, a, b) \leq Mw(x)$$

$$\int_{\mathbb{R}^k} w^p [F(x, a, b, s)] \rho(s) ds \leq \beta_0 w^p(x) + b_0.$$

La razón para imponer la Hipótesis 7(c) es para poder aplicar los argumentos usados para el caso descontado. En particular, notemos que los comentarios de la Observación 2.4.1 son válidos.

Sean $\xi_0, \xi_1, \dots, \xi_n, \dots$ realizaciones independientes de vectores aleatorios con densidad $\rho(\cdot)$, y $\hat{\rho}_n^i(s) := \hat{\rho}_n^i(s; \xi_0, \xi_1, \dots, \xi_{n-1})$, $s \in \mathbb{R}^k$, $n \in \mathbb{N}$ estimadores de la densidad para los jugadores $i = 1, 2$ tales que

$$E \|\hat{\rho}_t^i - \rho\|_1^{p'} = E \left(\int_{\mathbb{R}^k} |\hat{\rho}_t^i(s) - \rho(s)| ds \right)^{p'} = \mathcal{O}(t^{-\delta}) \text{ cuando } t \rightarrow \infty, \quad (3.14)$$

para alguna constante $\delta > 0$, donde $1/p + 1/p' = 1$.

Ahora, consideremos la clase \mathcal{D} formada por las densidades $\sigma(\cdot)$ que satisfacen las siguientes condiciones:

D.1 $\sigma(\cdot) \leq \tilde{\rho}(\cdot)$ casi dondequiera respecto a la medida de Lebesgue;

D.2 $\int_{\mathbb{R}^k} w[F(x, a, b, s)]\sigma(s)ds \leq \beta w(x) + c$ para todo $(x, a, b) \in \mathbb{K}$, donde las constantes β y c son como en la Observación 2.4.1.

Como en el caso descontado, bajo la Hipótesis 5 la clase de densidades \mathcal{D} es un subconjunto cerrado y convexo de \mathbb{L}_1 . Entonces de [6], para cada $t \in \mathbb{N}$, sea $\rho_t(\cdot) \in \mathcal{D}$ la mejor aproximación de $\hat{\rho}_t(\cdot)$ en el conjunto \mathcal{D} (ver (2.30)), esto es,

$$\|\rho_t^i - \hat{\rho}_t^i\|_1 = \inf_{\sigma \in \mathcal{D}} \|\sigma - \hat{\rho}_t^i\|_1.$$

La densidad $\rho_t^i(\cdot)$ es la que usaremos para obtener las estrategias adaptadas para el jugador $i = 1, 2$, y su convergencia se establece en el siguiente resultado cuya demostración es similar a la del Lema 2.4.1.

Lema 3.2.1. Bajo las Hipótesis 5 y 7, tenemos

$$E \|\rho_t - \rho\|^{p'} = \mathcal{O}(t^{-\delta}) \text{ cuando } t \rightarrow \infty,$$

donde la norma es definida como en (2.31).

3.3. Construcción de estrategias óptimas

Fijemos una sucesión de factores de descuento $\{\hat{\alpha}_t\}$ tal que $1 - \hat{\alpha}_t = \mathcal{O}(t^{-\nu})$ cuando $t \rightarrow \infty$, y

$$\lim_{n \rightarrow \infty} \frac{\kappa(n)}{n} = 0, \quad (3.15)$$

donde $0 < \nu < \delta/\kappa(3p')$, δ y p' como en el Lema 3.2.1, y $\kappa(n)$ es el número de cambios de valor de $\{\hat{\alpha}_t\}$ para $t = 0, 1, \dots, n$.

Tomando en cuenta la Condición D.2, la Observación 2.4.1(b) y el Teorema 2.3.5, tenemos que para cada $t \in \mathbb{N}_0$ existe una función $V_{\hat{\alpha}_t}^{\rho_t^i} \in \mathcal{C}_w$ tal que

$$\|V_{\hat{\alpha}_t}^{\rho_t^i}\|_w \leq \frac{M}{1 - \hat{\alpha}_t} \quad (3.16)$$

y

$$V_{\hat{\alpha}_t}^{\rho_t^i}(x) = \max_{\mu \in \mathbb{P}_A(x)} \min_{\lambda \in \mathbb{P}_B(x)} \left[r(x, \mu, \lambda) + \hat{\alpha}_t \int_{\mathbb{R}^k} V_{\hat{\alpha}_t}^{\rho_t^i}(F(x, \mu, \lambda, s)) \rho_t(s) ds \right], \quad (3.17)$$

o equivalentemente

$$\begin{aligned}
 j_{\hat{\alpha}_t}^{\rho_t^i} + h_{\hat{\alpha}_t}^{\rho_t^i}(x) &= \max_{\mu \in \mathbb{P}_A(x)} \min_{\lambda \in \mathbb{P}_B(x)} \left[r(x, \mu, \lambda) + \hat{\alpha}_t \int_{\mathbb{R}^k} h_{\hat{\alpha}_t}^{\rho_t^i}(F(x, \mu, \lambda, s)) \rho_t(s) ds \right] \\
 &= \max_{\mu \in \mathbb{P}_A(x)} \min_{\lambda \in \mathbb{P}_B(x)} \left[r(x, \mu, \lambda) + \hat{\alpha}_t \int_{\mathbb{R}^k} h_{\hat{\alpha}_t}^{\rho_t^i}(F(x, \mu, \lambda, s)) \rho_t(s) ds \right] \\
 &= \min_{\lambda \in \mathbb{P}_B(x)} \max_{\mu \in \mathbb{P}_A(x)} \left[r(x, \mu, \lambda) + \hat{\alpha}_t \int_{\mathbb{R}^k} h_{\hat{\alpha}_t}^{\rho_t^i}(F(x, \mu, \lambda, s)) \rho_t(s) ds \right]
 \end{aligned}$$

donde, para cada estado fijo $z \in X$, $j_{\hat{\alpha}_t}^{\rho_t^i} := (1 - \hat{\alpha}_t) V_{\hat{\alpha}_t}^{\rho_t^i}(z)$ y $h_{\hat{\alpha}_t}^{\rho_t^i}(x) := V_{\hat{\alpha}_t}^{\rho_t^i}(x) - V_{\hat{\alpha}_t}^{\rho_t^i}(z)$ (ver (3.11)). Más aún, existen sucesiones $\{\mu_t^*\} \subset \Pi_S$ y $\{\lambda_t^*\} \subset \Gamma_S$ tales que

$$j_{\hat{\alpha}_t}^{\rho_t^i} + h_{\hat{\alpha}_t}^{\rho_t^i}(x) = \max_{\mu \in \mathbb{P}_A(x)} \left[r(x, \mu, \lambda_t^*) + \hat{\alpha}_t \int_{\mathbb{R}^k} h_{\hat{\alpha}_t}^{\rho_t^i}(F(x, \mu, \lambda_t^*)) \rho_t(s) ds \right], \quad (3.18)$$

y

$$j_{\hat{\alpha}_t}^{\rho_t^i} + h_{\hat{\alpha}_t}^{\rho_t^i}(x) = \min_{\lambda \in \mathbb{P}_B(x)} \left[r(x, \mu_t^*, \lambda) + \hat{\alpha}_t \int_{\mathbb{R}^k} h_{\hat{\alpha}_t}^{\rho_t^i}(F(x, \mu_t^*, \lambda)) \rho_t(s) ds \right]. \quad (3.19)$$

El resultado principal de esta sección lo estableceremos de la siguiente manera.

Teorema 3.3.1. Supongamos que se cumplen las Hipótesis 5, 6 y 7. Entonces las estrategias $\pi^* = \{\mu_t^*\}$ y $\gamma^* = \{\lambda_t^*\}$ son óptimas en el caso promedio para el jugador 1 y 2, respectivamente, esto es

$$j^* \geq J(x, \pi^*, \gamma), \quad \forall \gamma \in \Gamma, x \in X, \quad (3.20)$$

$$j^* \leq J(x, \pi, \gamma^*), \quad \forall \pi \in \Pi, x \in X. \quad (3.21)$$

Por lo tanto (π^*, γ^*) es un par óptimo.

3.3.1. Demostración del resultado principal

Para la demostración del Teorema 3.3.1 seguiremos algunas ideas introducidas en la Subsección 2.4.2 las cuales estableceremos en la siguiente observación.

Observación 3.3.1. Supongamos que la Hipótesis 7 se cumple. Para cada $t \in \mathbb{N}_0$, definimos $\theta_t := (1 - \widehat{\alpha}_t)/2 \in (\widehat{\alpha}_t, 1)$, $e_t := c(\theta_t/\widehat{\alpha}_t - 1)^{-1}$, y la función $w_t(x) := w(x) + e_t$ para $x \in X$. Ahora, consideremos el espacio $\mathbb{B}_{w_t}(X)$ de funciones $u : X \rightarrow \mathbb{R}$ con w_t -norma finita, esto es

$$\|u\|_{w_t} := \sup_{x \in X} \frac{|u(x)|}{w_t(x)} < \infty, \quad t \in \mathbb{N}_0.$$

Observemos que, para cada $t \in \mathbb{N}_0$, esta norma es equivalente a la w -norma debido a que

$$\|u\|_{w_t} \leq \|u\|_w \leq l_t \|u\|_{w_t}, \quad (3.22)$$

donde

$$l_t := 1 + \frac{2c}{1 - \widehat{\alpha}_t}. \quad (3.23)$$

Entonces para cualquier densidad σ que satisface la condición D.2, el operador $T_\sigma^{\widehat{\alpha}_t}$ es de contracción respecto a la w_t -norma con módulo θ_t , es decir

$$\|T_\sigma^{\widehat{\alpha}_t} v - T_\sigma^{\widehat{\alpha}_t} u\|_{w_t} \leq \theta_t \|v - u\|_{w_t} \quad \forall v, u \in \mathbb{B}_w(X), t \in \mathbb{N}_0. \quad (3.24)$$

Lema 3.3.2. Dadas las Hipótesis 5, 6 y 7, tenemos, para cada $x \in X$ y $(\pi, \gamma) \in \Pi \times \Gamma$, cuando $t \rightarrow \infty$

$$(a) E_x^{\pi, \gamma} \left\| h_{\widehat{\alpha}_t} - h_{\widehat{\alpha}_t}^{\rho_t^i} \right\|_w^{p'} \rightarrow 0, \quad (b) E_x^{\pi, \gamma} \left[\left\| h_{\widehat{\alpha}_t} - h_{\widehat{\alpha}_t}^{\rho_t^i} \right\|_w w(x_t) \right] \rightarrow 0.$$

Demostración. (a) Observemos que para cada $t \in \mathbb{N}_0$, $\left\| h_{\widehat{\alpha}_t} - h_{\widehat{\alpha}_t}^{\rho_t^i} \right\|_w \leq 2 \|V_{\widehat{\alpha}_t} - V_{\widehat{\alpha}_t}^{\rho_t^i}\|_w$. Por tanto, de (3.22), esta parte será demostrada si

$$l_t^{p'} E_x^{\pi, \gamma} \|V_{\widehat{\alpha}_t} - V_{\widehat{\alpha}_t}^{\rho_t^i}\|_w^{p'} \rightarrow 0, \quad \text{cuando } t \rightarrow \infty. \quad (3.25)$$

Del Teorema 2.3.5 y (3.17), para cada $t \in \mathbb{N}_0$, $T_\rho^{\widehat{\alpha}_t} V_{\widehat{\alpha}_t} = V_{\widehat{\alpha}_t}$ y $T_{\rho_t}^{\widehat{\alpha}_t} V_{\widehat{\alpha}_t}^{\rho_t^i} = V_{\widehat{\alpha}_t}^{\rho_t^i}$. Así de (3.24) tenemos

$$\|V_{\widehat{\alpha}_t} - V_{\widehat{\alpha}_t}^{\rho_t^i}\|_{w_t} \leq \|T_\rho^{\widehat{\alpha}_t} V_{\widehat{\alpha}_t} - T_{\rho_t}^{\widehat{\alpha}_t} V_{\widehat{\alpha}_t}^{\rho_t^i}\|_{w_t} + \theta_t \|V_{\widehat{\alpha}_t} - V_{\widehat{\alpha}_t}^{\rho_t^i}\|_{w_t}$$

lo cual implica

$$l_t \left\| V_{\widehat{\alpha}_t} - V_{\widehat{\alpha}_t}^{\rho_t^i} \right\|_{w_t} \leq \frac{l_t}{1 - \theta_t} \left\| T_\rho^{\widehat{\alpha}_t} V_{\widehat{\alpha}_t} - T_{\rho_t}^{\widehat{\alpha}_t} V_{\widehat{\alpha}_t}^{\rho_t^i} \right\|_{w_t}, \quad t \in \mathbb{N}_0. \quad (3.26)$$

Ahora de (3.16), (2.31) y usando el hecho $[w_t(\cdot)]^{-1} < [w(\cdot)]^{-1}$,

$$\begin{aligned}
 & \left\| T_{\rho}^{\widehat{\alpha}_t} V_{\widehat{\alpha}_t} - T_{\rho_t^i}^{\widehat{\alpha}_t} V_{\widehat{\alpha}_t} \right\|_{w_t} \\
 & \leq \widehat{\alpha}_t \sup_{x \in X} [w_t(x)]^{-1} \sup_{\mu \in \mathbb{P}_A(x), \lambda \in \mathbb{P}_B(x)} \int_{\mathbb{R}^k} V_{\alpha}(F(x, \mu, \lambda, s)) |\rho(s) - \rho_t^i(s)| ds \\
 & \leq \frac{M \widehat{\alpha}_t}{1 - \widehat{\alpha}_t} \sup_{x \in X} [w_t(x)]^{-1} \sup_{\mu \in \mathbb{P}_A(x), \lambda \in \mathbb{P}_B(x)} \int_{\mathbb{R}^k} V_{\alpha}(F(x, \mu, \lambda, s)) |\rho(s) - \rho_t^i(s)| ds \\
 & \leq \frac{M}{1 - \widehat{\alpha}_t} \|\rho - \rho_t^i\|. \tag{3.27}
 \end{aligned}$$

Por otro lado, notemos que $\widehat{\alpha}_t$ y θ_t satisfacen la relación

$$\frac{1}{(1 - \theta_t)(1 - \widehat{\alpha}_t)^2} = \mathbf{O}(t^{3\nu}), \quad \text{cuando } t \rightarrow \infty. \tag{3.28}$$

Combinando (3.26)-(3.28), y usando (3.23) obtenemos

$$\begin{aligned}
 l_t^{p'} \left\| V_{\widehat{\alpha}_t} - V_{\widehat{\alpha}_t}^{\rho_t^i} \right\|_{w_t}^{p'} & \leq M^{p'} \left[\frac{1}{(1 - \theta_t)(1 - \widehat{\alpha}_t)} + \frac{2d}{(1 - \theta_t)(1 - \widehat{\alpha}_t)^2} \right]^{p'} \|\rho - \rho_t^i\| \\
 & = M^{p'} \mathbf{O}(t^{3p'\nu}) \|\rho - \rho_t^i\|^{p'} \quad \text{cuando } t \rightarrow \infty. \tag{3.29}
 \end{aligned}$$

Finalmente, tomando esperanza $E_x^{\pi\gamma}$ en ambos lados de (3.29), por el Lema 3.2.1 y el hecho de que $3\nu p' < \delta$, la relación (3.25) se cumple, lo cual demuestra (a).

(b) Sea $\overline{C} := (E_x^{\pi\gamma} [w^p(x_t)])^{1/p} < \infty$ (ver Observación 2.4.1 (c)). Entonces, aplicando la desigualdad de Holder en la parte (a), obtenemos

$$E_x^{\pi\gamma} \left\| h_{\widehat{\alpha}_t} - h_{\widehat{\alpha}_t}^{\rho_t^i} \right\|_w w(x_t) \leq \overline{C} \left(E_x^{\pi\gamma} \left[\left\| h_{\widehat{\alpha}_t} - h_{\widehat{\alpha}_t}^{\rho_t^i} \right\|_w^{p'} \right] \right)^{1/p'} \rightarrow \infty, \quad \text{cuando } t \rightarrow \infty.$$

Esto demuestra el lema. □

Demostración del Teorema 3.3.1. Primero demostraremos la relación (3.21). Para esto sea $\pi = \{\pi_t\} \in \Pi$ una estrategia arbitraria para el jugador 1. Definimos

$$\begin{aligned}\eta_t &= r(x_t, \pi_t, \lambda_t^*) + \widehat{\alpha}_t \int_{\mathbb{R}^k} h_{\widehat{\alpha}_t} [F(x_t, \pi_t, \lambda_t^*, s)] \rho(s) ds - J_{\alpha_t} - h_{\widehat{\alpha}_t}(x_t) \\ &= r(x_t, \pi_t, \lambda_t^*) + \widehat{\alpha}_t E_x^{\pi\gamma^*} [h_{\widehat{\alpha}_t}(x_{t+1}) | h_t] - J_{\alpha_t} - h_{\widehat{\alpha}_t}(x_t).\end{aligned}$$

De aquí

$$E_x^{\pi\gamma^*} r(x_t, a_t, b_t) = J_{\widehat{\alpha}_t} + E_x^{\pi\gamma^*} [h_{\widehat{\alpha}_t}(x_t) - \widehat{\alpha}_t h_{\widehat{\alpha}_t}(x_{t+1})] + E_x^{\pi\gamma^*}(\eta_t),$$

lo cual implica, para $n \geq k \geq 1$

$$\begin{aligned}n^{-1} E_x^{\pi\gamma^*} \left[\sum_{t=1}^n r(x_t, a_t, b_t) \right] &= n^{-1} \sum_{t=1}^n J_{\widehat{\alpha}_t} + n^{-1} E_x^{\pi\gamma^*} \left[\sum_{t=1}^{k-1} (h_{\widehat{\alpha}_t}(x_t) - \widehat{\alpha}_t h_{\widehat{\alpha}_t}(x_{t+1})) \right] \\ &\quad + n^{-1} E_x^{\pi\gamma^*} \left[\sum_{t=k}^n (h_{\widehat{\alpha}_t}(x_t) - \widehat{\alpha}_t h_{\widehat{\alpha}_t}(x_{t+1})) \right] \\ &\quad + n^{-1} E_x^{\pi\gamma^*} \left[\sum_{t=1}^n \eta_t \right] =: I_1(n) + I_2(n) + I_3(n) + I_4(n).\end{aligned}\tag{3.30}$$

Observemos que por (3.12) cuando $n \rightarrow \infty$,

$$I_1(n) \rightarrow j^*.\tag{3.31}$$

También tenemos

$$I_2(n) \rightarrow 0, \text{ cuando } n \rightarrow \infty.\tag{3.32}$$

Por otro lado, sea $\overline{C} < \infty$ una constante tal que $E_x^{\pi\gamma^*} [h_{\alpha}(x_t)] < C'$, para todo $\alpha \in (0, 1)$ y $t \in \mathbb{N}_0$ (ver Obsevación 2.4.1 (c)). Denotemos por $\alpha_1^*, \alpha_2^*, \dots, \alpha_{\kappa(n)}^*$, $n \leq 1$, los diferentes valores de $\widehat{\alpha}_t$ para $t \leq n$ (ver (3.15)), y usando el hecho que $\{\widehat{\alpha}_t\}$ es una sucesión no decreciente, tenemos

$$\begin{aligned}
 0 &\leq I_3(n) \\
 &= n^{-1} E_x^{\pi\gamma^*} \left[\sum_{t=1}^n (h_{\hat{\alpha}_t}(x_t) - \hat{\alpha}_t h_{\hat{\alpha}_t}(x_t)) \right] + n^{-1} E_x^{\pi\gamma^*} \left[\sum_{t=1}^n \hat{\alpha}_t (h_{\hat{\alpha}_t}(x_t) - h_{\hat{\alpha}_t}(x_{t+1})) \right] \\
 &\leq (1 - \alpha_k) C' + n^{-1} C' \sum_{i=1}^{\kappa(n)} \alpha_i^* \\
 &\leq (1 - \alpha_k) C' + C' \kappa(n) n^{-1}
 \end{aligned}$$

Así, como k es arbitraria y $\hat{\alpha}_t \nearrow 1$ cuando $t \rightarrow \infty$, la codición (3.15) implica que

$$\lim_{n \rightarrow \infty} I_3(n) = 0. \quad (3.33)$$

Ahora mostraremos que

$$\lim_{n \rightarrow \infty} I_4(n) \leq 0. \quad (3.34)$$

Sumando y restando los términos $\hat{\alpha}_t \int_{\mathbb{R}^k} h_{\hat{\alpha}_t}^{\rho_t^2} [F(x_t, \pi_t, \lambda_t^*, s)] \rho(s) ds$ y $\hat{\alpha}_t \int_{\mathbb{R}^k} h_{\hat{\alpha}_t}^{\rho_t^2} [F(x_t, \pi_t, \lambda_t^*, s)] \rho_t^2(s) ds$ obtenemos

$$\begin{aligned}
 \eta_t &= \hat{\alpha}_t \int_{\mathbb{R}^k} h_{\hat{\alpha}_t} [F(x_t, \pi_t, \lambda_t^*, s)] \rho(s) ds - \hat{\alpha}_t \int_{\mathbb{R}^k} h_{\hat{\alpha}_t}^{\rho_t^2} [F(x_t, \pi_t, \lambda_t^*, s)] \rho(s) ds \\
 &+ \hat{\alpha}_t \int_{\mathbb{R}^k} h_{\hat{\alpha}_t}^{\rho_t^2} [F(x_t, \pi_t, \lambda_t^*, s)] \rho(s) ds - \hat{\alpha}_t \int_{\mathbb{R}^k} h_{\hat{\alpha}_t}^{\rho_t^2} [F(x_t, \pi_t, \lambda_t^*, s)] \rho_t^2(s) ds \\
 &+ r(x_t, \pi_t, \lambda_t^*) + \hat{\alpha}_t \int_{\mathbb{R}^k} h_{\hat{\alpha}_t}^{\rho_t^2} [F(x_t, \pi_t, \lambda_t^*, s)] \rho_t^2(s) ds - j_{\hat{\alpha}_t} - h_{\hat{\alpha}_t}(x_t).
 \end{aligned}$$

Entonces de la definición de w -norma, (2.31), usando el hecho que $|u(x)| \leq \|u\|_w w(x)$, $u \in \mathbb{L}_w$, $x \in X$, y (3.19)

$$\begin{aligned}
 \eta_t &\leq \widehat{\alpha}_t \left\| h_{\widehat{\alpha}_t} - h_{\widehat{\alpha}_t}^{\rho_t^2} \right\|_w [(x_t) + d] + \widehat{\alpha}_t w(x_t) \left\| h_{\widehat{\alpha}_t}^{\rho_t^2} \right\|_w \|\rho - \rho_t^2\| \\
 &+ \max_{\mu \in \mathbb{P}_A(x)} \left\{ r(x_t, \pi_t, \lambda_t^*) + \widehat{\alpha}_t \int_{\mathbb{R}^k} h_{\widehat{\alpha}_t}^{\rho_t^2} [F(x_t, \pi_t, \lambda_t^*, s)] \rho_t^2(s) ds \right\} - J_{\widehat{\alpha}_t} - h_{\widehat{\alpha}_t}(x_t) \\
 &\leq \widehat{\alpha}_t \left\| h_{\widehat{\alpha}_t} - h_{\widehat{\alpha}_t}^{\rho_t^2} \right\|_w [(x_t) + d] + \widehat{\alpha}_t w(x_t) \left\| h_{\widehat{\alpha}_t}^{\rho_t^2} \right\|_w \|\rho - \rho_t^2\| \\
 &+ J_{\widehat{\alpha}_t}^{\rho_t^2} + h_{\widehat{\alpha}_t}^{\rho_t^2}(x_t) - J_{\widehat{\alpha}_t} - h_{\widehat{\alpha}_t}(x_t). \tag{3.35}
 \end{aligned}$$

Ahora del Lema 3.2.3

$$\lim_{t \rightarrow \infty} E_x^{\pi\gamma^*} \left[\widehat{\alpha}_t \left\| h_{\widehat{\alpha}_t} - h_{\widehat{\alpha}_t}^{\rho_t^2} \right\|_w [\beta w(x_t) + c] \right] = 0. \tag{3.36}$$

Mas aún, de la definición de $\widehat{\alpha}_t$ y (3.16)

$$\left\| h_{\widehat{\alpha}_t}^{\rho_t^2} \right\|_w \leq 2 \left\| V_{\widehat{\alpha}_t}^{\rho_t^2} \right\|_w \leq \frac{2}{1 - \widehat{\alpha}_t} = \mathbf{O}(t^\nu) \text{ cuando } t \rightarrow \infty.$$

Así, tomando esperanza y aplicando la desigualdad de Holder, obtenemos

$$\begin{aligned}
 0 &\leq E_x^{\pi\gamma^*} \left[w(x_t) \left\| h_{\widehat{\alpha}_t}^{\rho_t^2} \right\|_w \|\rho - \rho_t^2\| \right] \leq \left([\mathbf{O}(t^\nu)]^{p'} E_x^{\pi\gamma^*} \|\rho - \rho_t^2\|^{p'} \right)^{1/p'} \\
 &= \left[\mathbf{O}(t^{\nu p' - \delta}) \right]^{1/p'} \rightarrow 0 \text{ cuando } t \rightarrow \infty, \tag{3.37}
 \end{aligned}$$

dado que $\nu < \delta/p'$.

Por otro lado, observemos que

$$J_{\widehat{\alpha}_t}^{\rho_t^2} - J_{\widehat{\alpha}_t} \leq (1 - \widehat{\alpha}_t) \left\| V_{\widehat{\alpha}_t}^{\rho_t^2} - V_{\widehat{\alpha}_t} \right\|$$

y

$$h_{\widehat{\alpha}_t}^{\rho_t^2}(x_t) - h_{\widehat{\alpha}_t}(x_t) \leq \left\| h_{\widehat{\alpha}_t}^{\rho_t^2} - h_{\widehat{\alpha}_t} \right\|_w w(x_t).$$

Así del Lema 3.2.3 obtenemos

$$\lim_{t \rightarrow \infty} E_x^{\pi\gamma^*} (J_{\widehat{\alpha}_t}^{\rho_t^2} - J_{\widehat{\alpha}_t}) \leq 0 \tag{3.38}$$

y

$$\lim_{t \rightarrow \infty} E_x^{\pi \gamma^*} (h_{\hat{\alpha}_t}^{\rho_t^2}(x_t) - h_{\hat{\alpha}_t}(x_t)) \leq 0. \quad (3.39)$$

Por lo tanto, combinando (3.35)-(3.39) obtenemos (3.34). Finalmente, haciendo $n \rightarrow \infty$ en (3.30), de (3.1) y (3.31)-(3.34), obtenemos (3.20), es decir,

$$J(x, \pi, \gamma^*) \leq j^*, \quad \forall \pi \in \Pi, x \in X.$$

La demostración de (3.20) es similar a la anterior, solamente considerando $\gamma \in \Gamma$ arbitraria y la densidad respectiva ρ_t^1 . \square

Capítulo 4

Conclusiones

En este trabajo se estudiaron juegos estocásticos de suma cero bajo los criterios de optimalidad descontado y promedio. Específicamente este estudio se hizo bajo dos tipos hipótesis generales con las cuales se puede garantizar la existencia del valor del juego así como de pares de estrategias óptimos para los jugadores.

Otro problema que se estudio fue la construcción de estrategias adaptadas en juegos suma cero cuando la dinámica es modelada por medio de ecuaciones en diferencia, y la distribución del ruido aleatorio es desconocida por ambos jugadores. Es decir, se implementaron técnicas de estimación estadística de la distribución y proceso de optimización para la construcción de estrategias, las cuales resultan asintóticamente óptimas para el caso descontado, y óptimas para el caso promedio.

Básicamente estos resultados son una extensión a juegos de lo métodos de control adaptado para procesos de Markov establecidos en [9, 10, 14, 20, 21, 26].

En estos sistemas, los estimadores de la densidad se definen en términos de observaciones históricas del ruido aleatorio, lo cual implica que la hipótesis de observabilidad del proceso de perturbación es muy importante para desarrollar la teoría.

A partir de esto último surgen problemas interesantes que se pueden abordar con futuros trabajos. En efecto, podemos estudiar juegos estocásticos donde el proceso de perturbación es no observable, los cuales, siguiendo una metodología similar, se pueden abordar extendiendo los resultados de procesos de control de Markov para

sistemas parcialmente observables.

Para concluir haremos una breve descripción sobre la bibliografía que se utilizó para la elaboración del presente trabajo.

La teoría general sobre procesos de control de Markov se estudió en [5, 10, 12, 13], y la teoría de juegos estocásticos en [17, 24, 25]. Los métodos de estimación y control en sistemas estocásticos se obtuvieron de [9, 10, 11, 14, 20, 21]. Además, los resultados del Capítulo 2 sobre la existencia de estrategias óptimas así como la construcción de estrategias adaptadas fueron estudiadas en [22, 24, 25], mientras que los del Capítulo 3 sobre el criterio promedio se estudiaron en [17, 23]. El resto de la bibliografía sirvió de apoyo.

Apéndice A

Espacios de Borel, Kernels y multifunciones

A.1. Espacios de Borel y kernels estocásticos

Un espacio topológico siempre estará dotado con la σ -álgebra de Borel $\mathcal{B}(X)$, esto es, la más pequeña σ -álgebra de subconjuntos de X que contiene a todos los conjuntos abiertos en X . Así, cuando nos referimos tanto a conjuntos como a funciones, *medible* significa *Borel medible*.

Un subconjunto de Borel X de un espacio métrico completo y separable es llamado un *espacio de Borel*, y su σ -álgebra es denotada por $\mathcal{B}(X)$. Un subconjunto de Borel de un espacio de Borel es un espacio de Borel.

Definición A.1.1. Un kernel estocástico de X en Y es una función $P(\cdot|\cdot)$ tal que

- (i) $P(\cdot|y)$ es una medida de probabilidad en X para cada $y \in Y$ fijo, y
- (ii) $P(D|\cdot)$ es una función medible en Y para cada $D \in \mathcal{B}(X)$ fijo.

El conjunto de los kernels estocásticos de X en Y los denotaremos por $\mathbb{P}(X|Y)$.

Dado un espacio X denotamos por $\mathbb{P}(X)$ a la familia de medidas de probabilidad en X .

Decimos que $\mathbb{P}(X)$ esta dotado de la topología débil, esto es, $\mu_n \rightarrow \mu$ débilmente

si

$$\int u d\mu_n \rightarrow \int u d\mu \quad \text{para cada función continua y acotada } u. \quad (\text{A.1})$$

Proposición A.1.1. Si X es un espacio de Borel, entonces $\mathbb{P}(X)$ es de Borel. Si, además X es compacto, entonces $\mathbb{P}(X)$ también lo es.

Demostración. Ver el Corolario 7.25.1 y la Proposición 7.22 en [5]. □

Definición A.1.2. Un kernel estocástico $Q \in \mathbb{P}(X|Y)$ se dice ser

(a) *débilmente continua* (o satisface la propiedad de Feller) si la función

$$y \mapsto \int v(x)Q(dx|y) \quad (\text{A.2})$$

es continua y acotada en Y para cada función v continua y acotada en X ;

(b) *fuertemente continua* (o satisface la propiedad fuerte de Feller) si la función en (A.2) es continua y acotada en Y para cada función v acotada en X .

A.2. Multifunciones y selectores

Sean X y A espacios de Borel (no vacíos).

Una *multifunción* (también llamada *correspondencia* o *función conjunto-valorada*) φ de X en A es una función tal que $\varphi(x)$ es un subconjunto no vacío de A para cada $x \in X$. (Una función $F : X \rightarrow A$ es un ejemplo de una multifunción). La *gráfica* de una multifunción es el subconjunto de $X \times A$ definido como

$$Gr(\varphi) := \{(x, a) | x \in X, a \in \varphi(x)\}. \quad (\text{A.3})$$

Para cada subconjunto B de A definamos el siguiente conjunto

$$\varphi^{-1}(B) := \{x \in X | \varphi(x) \cap B \neq \emptyset\}.$$

Definición A.2.1. Una multifunción φ de X en A se dice ser

(a) *Borel medible* si $\varphi^{-1}(G)$ es un subconjunto de Borel en X para cada cada conjunto abierto $G \subset A$;

- (b) *semicontinua superiormente* (s.c.s.) si $\varphi^{-1}(B)$ es cerrado en X para cada conjunto cerrado $B \subset A$;
- (c) *semicontinua inferiormente* (s.c.i.) si $\varphi^{-1}(G)$ es abierto en X para cada conjunto abierto $G \subset A$;
- (d) *continua* si es s.c.s. y s.c.i.

Sea φ una multifunción Borel medible de X en A , denotaremos por \mathbb{F} al conjunto de funciones medibles $f : X \rightarrow A$ con $f(x) \in \varphi(x)$ para todo $x \in X$. Una función $f \in \mathbb{F}$ es llamada un selector de la multifunción φ .

Sea $M : X \rightarrow Y$ una correspondencia, definimos la correspondencia $\Phi : X \rightarrow \mathbb{P}(Y)$ como

$$\Phi(x) := \mathbb{P}(M(x)) \quad x \in X.$$

Proposición A.2.1. Si M es continua que toma valores compactos (no vacíos), entonces Φ es también continua y con valores compactos.

Demostración. (ver Teorema 3 en [16]). □

El siguiente resultado trata de selectores continuos para multifunciones semicontinuas (ver [19]).

Proposición A.2.2. Supongamos que M es continua con valores compactos (no vacíos). Sea $x_0 \in X$ y $\nu_0 \in \Phi(x_0)$. Entonces Φ admite un selector continuo cuya gráfica contiene el punto (x_0, ν_0) , esto es, existe una función continua $\phi : X \rightarrow Y$ tal que $\phi(x) \in \Phi(x)$ para cada $x \in X$ y $\phi(x_0) = \nu_0$.

Apéndice B

Semicontinuidad y Lema de Fatou

B.1. Funciones semicontinuas

Definición B.1.1. Sea X un espacio métrico, f una función de X en $\mathbb{R} \cup \{\infty\}$ y $f(x_0) < \infty$ para al menos un punto $x_0 \in X$. La función f es semicontinua inferiormente en x_0 si y sólo si, para cualquier sucesión (x_n) tal que $x_n \rightarrow x_0$ tenemos que

$$f(x_0) \leq \liminf_{n \rightarrow \infty} f(x_n). \quad (\text{B.1})$$

Consecuentemente f es semicontinua superiormente en x_0 si y sólo si, para cualquier sucesión (x_n) tal que $x_n \rightarrow x$ tenemos que

$$\limsup_{n \rightarrow \infty} f(x_n) \leq f(x_0).$$

El siguiente resultado es inmediato.

Proposición B.1.1. Sea X un espacio métrico. Una función $f : X \rightarrow \mathbb{R}$ es semicontinua superiormente en $x_0 \in X$, si y sólo si, la función $-f$ es semicontinua inferiormente en x_0 .

El siguiente resultado lo puede ver en [4] (1993, Capítulo 1).

Teorema B.1.2. Sea X un espacio compacto y $f : X \rightarrow \mathbb{R} \cup \{\infty\}$ una función semicontinua inferiormente. Entonces, la función f alcanza el valor mínimo en $x_0 \in X$, es decir $f(x_0) \leq f(x)$ para todo $x \in X$ y el conjunto de puntos donde se alcanza

el valor mínimo es compacto.

Similarmente, una función f semicontinua superiormente con valores reales en un espacio compacto X , alcanza el valor máximo y el conjunto de puntos donde se alcanza el valor máximo es compacto.

B.2. Lema de Fatou generalizado

Dentro de los teoremas básicos de integración están el Teorema de Convergencia Monótona, el Teorema de Convergencia Dominada y el Lema de Fatou, los cuales se pueden ver en [3, 2]. En este caso los resultados que se presentan a continuación, son correspondientes al Lema de Fatou generalizado.

Definición B.2.1. Sea $w : X \rightarrow [1, \infty)$ una función medible. Para cada función medible $u : X \rightarrow \mathbb{R}$ definimos su w -norma como

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

Denotaremos por $\mathbb{B}_w(X)$ al espacio de Banach de las funciones con w -norma finita.

Dada $\{w_n\}$ una sucesión de funciones en $\mathbb{B}_w(X)$, definimos el límite inferior (superior) generalizado como:

$$w_*(x) := \inf \left\{ \liminf_{n \rightarrow \infty} (x_n) : x_n \rightarrow x \right\} \quad (\text{B.2})$$

y

$$w^*(x) := \sup \left\{ \limsup_{n \rightarrow \infty} (x_n) : x_n \rightarrow x \right\}, \quad (\text{B.3})$$

respectivamente.

Lema B.2.1. La función w_* (respectivamente w^*) es semicontinua superiormente (inferiormente).

Demostración. Sea ρ la métrica en el espacio X que define la topología. Para cada $x \in X$ y $r > 0$, hacemos $B(x, r) := \{y : \rho(x, y) < r\}$. Notemos que w_* puede ser escrito en la siguiente forma:

$$w_*(x) = \sup_n \inf_{k \geq n} \left(\inf_{y \in B(x, 1/n)} w_k(y) \right).$$

Fijemos $x \in X$ y supongamos que $w_*(x) > d$ para algún número real d . Para probar que $w_*(x)$ es semicontinua inferiormente en x , necesitamos mostrar que existe $\delta > 0$ tal que $w_*(y) > d$ para cada $y \in B(x, \delta)$. Ya que

$$\inf_{k \geq n} \left(\inf_{y \in B(x, 1/n)} w_k(y) \right) \uparrow w_*(x),$$

existe una N tal que

$$\inf_{k \geq l} \left(\inf_{y \in B(x, 1/l)} w_k(y) \right) > d \quad \forall l > N.$$

Dada $m > 2l > l > N$ y $\delta := 12l$, entonces tenemos

$$\begin{aligned} w_*(y) &= \sup_n \inf_{k \geq n} \left(\inf_{z \in B(x, 1/n)} w_k(z) \right) \geq \inf_{k \geq m} \left(\inf_{z \in B(x, 1/m)} w_k(z) \right) \\ &\geq \inf_{k \geq m} \left(\inf_{z \in B(x, 1/l)} w_k(z) \right) \geq \inf_{k \geq l} \left(\inf_{z \in B(x, 1/l)} w_k(z) \right) > d. \end{aligned}$$

La prueba de la semicontinuidad superior de la función $w^*(x)$ en x es similar a la anterior comenzando con

$$w^*(x) = \inf_n \sup_{k \geq n} \left(\sup_{y \in B(x, 1/n)} w_k(y) \right).$$

□

Lema B.2.2. Sea $\{\mu_n\} \subset \mathbb{P}(X)$ una sucesión que converge débilmente a algún $\mu_0 \in \mathbb{P}(X)$. Si $\{v_n\}$ es una sucesión de funciones Borel medibles no-negativas en X y w_* definida como en (B.2), entonces

$$\int_X w_*(x) \mu_0(dx) \leq \liminf_{n \rightarrow \infty} \int_X v_n(x) \mu_n(dx); \quad (\text{B.4})$$

y si las funciones $\{v_n\}$ son no-positivas, entonces

$$\int_X w^*(x) \mu_0(dx) \geq \limsup_{n \rightarrow \infty} \int_X v_n(x) \mu_n(dx), \quad (\text{B.5})$$

con w^* definida como en (B.3).

Demostración. (ver en Lema 3.2 de [17]). \square

Lema B.2.3. Supongamos que $\{\mu_n\}$ converge débilmente a algún $\mu_0 \in \mathbb{P}(X)$ y $\{w_n\}$ es una sucesión de funciones en $\mathbb{C}_w(X)$ tal que $\|w_n\|_w \leq b$ para toda n y para alguna constante $b > 0$. Si V es una función continua, $\int_X V(x)\mu_m(dx) < \infty$ para cada $m \geq 0$ y

$$\int_X V(x)\mu_m(dx) \rightarrow \int_X V(x)\mu_0(dx) \quad (\text{B.6})$$

cuando $m \rightarrow \infty$, entonces

$$\int_X w_*(x)\mu_0(dx) \leq \liminf_{n \rightarrow \infty} \int_X w_n(x)\mu_n(dx) \quad (\text{B.7})$$

y

$$\int_X w^*(x)\mu_0(dx) \geq \limsup_{n \rightarrow \infty} \int_X w_n(x)\mu_n(dx) \quad (\text{B.8})$$

Demostración. Definamos $v_n(x) := w_n(x) + bV(x)$ y notemos que $v_n \geq 0$. Para alguna $x \in X$ y cualquier sucesión $x_n \rightarrow x$ cuando $n \rightarrow \infty$, tenemos

$$\liminf_{n \rightarrow \infty} v_n(x_n) = bV(x) + \liminf_{n \rightarrow \infty} w_n(x_n).$$

Por tanto $v_*(x) = bV(x) + w_*(x)$, $x \in X$, y consecuentemente

$$\int_X v_*(x)\mu_0(dx) = b \int_X V(x)\mu_0(dx) + \int_X w_*(x)\mu_0(dx)$$

Aplicando (B.4) a la sucesión $\{v_n\}$ y (B.6), fácilmente obtenemos

$$\begin{aligned} \liminf_{n \rightarrow \infty} \int_X w_n(x)\mu_n(dx) + b \int_X V(x)\mu_0(dx) &= \liminf_{n \rightarrow \infty} \int_X v_n(x)\mu_n(dx) \\ &\geq \int_X v_*(x)\mu_0(dx) = \int_X w_*(x)\mu_0(dx) + b \int_X V(x)\mu_0(dx), \end{aligned}$$

lo cual implica (B.7).

Similarmente (B.8) se sigue de (B.5) tomando $v_n(x) := w_n(x) - bV(x) \leq 0$. \square

Apéndice C

Varios resultados

C.1. Teorema de punto fijo de Banach

Definición C.1.1. Dado (X, d) un espacio métrico. Un operador $T : X \rightarrow X$ es llamado una contracción si existe un número $0 \leq \tau < 1$ tal que

$$d(Tx_1, Tx_2) \leq \tau d(x_1, x_2)$$

para todo $x_1, x_2 \in S$. En este caso τ es llamado el módulo de T .

Proposición C.1.1. (Teorema del punto fijo de Banach) Un operador de contracción T en un espacio métrico completo (X, d) tiene un único punto fijo x^* . Mas aún, $d(T^n x, x^*) \leq \tau^n d(x, x^*)$ para todo $x \in X$, $n = 1, 2, \dots$, donde τ es el módulo de T , y $T^n := T(T^{n-1})$ para $n = 1, 2, \dots$, con $T^0 :=$ identidad.

C.2. Teorema minimax de Fan

Antes de enunciar el teorema minimax estableceremos la siguiente definición.

Definición C.2.1. Sea f una función que toma valores reales en el espacio producto $X \times Y$ de dos conjuntos arbitrarios X, Y (no necesariamente topológicos). f se dice ser:

(a) *convexa* en X si para cualquiera dos elementos $x_1, x_2 \in X$ y número $\alpha \in [0, 1]$,

existe un elemento $x_0 \in X$ tal que

$$f(x_0, y) \leq \alpha f(x_1, y) + (1 - \alpha)f(x_2, y) \text{ para todo } y \in Y.$$

(b) *cóncava* en Y si para cualquiera dos elementos $y_1, y_2 \in Y$ y $\beta \in [0, 1]$, existe $y_0 \in Y$ tal que

$$f(x_{y_0}) \geq \beta f(x, y_1) + (1 - \beta)f(x, y_2) \text{ para todo } x \in X.$$

Teorema C.2.1. Sean X, Y dos espacios métricos compactos y $f : X \times Y \rightarrow \mathbb{R}$. Supongamos que, para cada $y \in Y$, $f(x, y)$ es s.c.i. en X ; y para cada $x \in X$, $f(x, y)$ es s.c.s. en Y . Entonces:

(i) La igualdad

$$\min_{x \in X} \max_{y \in Y} f(x, y) = \max_{y \in Y} \min_{x \in X} f(x, y) \quad (\text{C.1})$$

se tiene, si y solo si la siguiente condición se satisface: Para cualquiera dos conjuntos finitos $\{x_1, x_2, \dots, x_n\} \subset X$ y $\{y_1, y_2, \dots, y_m\} \subset Y$, existe $x_0 \in X$ y $y_0 \in Y$ tales que

$$f(x_0, y_k) \leq f(x_i, y_0) \quad (1 \leq i \leq n, 1 \leq k \leq m).$$

(ii) En particular, si f es convexa en X y cóncava en Y , se tiene (C.1).

Demostración. (ver [7]). □

Para ver más resultados minimax vea [8].

C.3. Teorema del Máximo de Berge y Teorema de Selección medible

Los elementos que intervienen en los siguientes resultados están definidos en el Apéndice A.2 y B.1. Cuyas pruebas se pueden consultar en el Capítulo 17 de [1].

Lema C.3.1. Sea $\varphi : X \rightrightarrows Y$ una correspondencia s.c.i. entre espacios métricos y la función $f : Gr\varphi \rightarrow \mathbb{R}$ semicontinua inferiormente. Se define la función real extendida $m : X \rightarrow \overline{\mathbb{R}}$ como

$$m(x) = \sup_{y \in \varphi(x)} f(x, y).$$

Entonces la función m es semicontinua inferiormente.

Lema C.3.2. Sea $\varphi : X \rightrightarrows Y$ una correspondencia s.c.s. entre espacios métricos con valores compactos no vacíos y la función $f : Gr\varphi \rightarrow \mathbb{R}$ semicontinua superiormente. Se define la función $m : X \rightarrow \mathbb{R}$ como

$$m(x) = \max_{y \in \varphi(x)} f(x, y).$$

Entonces la función m es semicontinua superiormente en X .

Teorema C.3.3. (Teorema del Máximo de Berge) Sea $\varphi : X \rightrightarrows Y$ una correspondencia continua entre espacios métricos con valores compactos no vacíos y supongamos que la función $f : Gr\varphi \rightarrow \mathbb{R}$ es continua. Se define la función $m : X \rightarrow \mathbb{R}$ como

$$m(x) = \max_{y \in \varphi(x)} f(x, y)$$

y la correspondencia $\mu : X \rightrightarrows Y$ de maximizadores por

$$\mu(x) = \{y \in \varphi(x) : f(x, y) = m(x)\}.$$

Entonces

- (1) m es continua,
- (2) la correspondencia μ toma valores compactos no vacíos,
- (3) la correspondencia μ es s.c.s.

Teorema C.3.4. (Teorema de Selección medible) Sean X y Y espacios de Borel, $\varphi : X \rightrightarrows Y$ una correspondencia con valores compactos no vacíos y supongamos que la función $u : Gr\varphi \rightarrow \mathbb{R}$ es Borel medible tal que $u(x, \cdot)$ es s.c.s. en $\varphi(x)$ para cada $x \in X$. Entonces existe un selector Borel medible $f : X \rightarrow Y$ para cada φ tal que

$$u(x, f(x)) = \max_{y \in \varphi(x)} f(x, y) \text{ para cada } x \in X.$$

Más aún, la función m definida por $m(x) = \max_{y \in \varphi(x)} u(x, y)$ es Borel medible.

Similarmente, si u es s.c.s. en $\varphi(x)$ para cada $x \in X$, entonces existe un selector medible $g : X \rightarrow Y$ para φ tal que

$$u(x, g(x)) = \min_{y \in \varphi(x)} f(x, y) \text{ para cada } x \in X,$$

y, la función v definida por $v(x) = \min_{y \in \varphi(x)} u(x, y)$ es Borel medible.

Demostración. (ver Himmelberg en [15]).

□

Bibliografía

- [1] C.D. ALIPRANTIS, K.C. BORDER, *Infinite Dimensional Analysis*, Springer-Berlag, Berlin, 2006.
- [2] R. B. ASH, C.A. DOLÉANS-DADE, *Probability and Measure Theory*, Academic Res, 2000.
- [3] R. B ASH, *Real Analysis and Probability*, Academic Press, New York, 1972.
- [4] J.-P. AUBIN, *Optima and equilibria*, Springer-Verlag, Berlin, 1993.
- [5] D. P. BERTSEKAS, S. E. SHREVE, *Stochastic Optimal Control: The Discrete Time Case*, Academic Press, New York, 1978.
- [6] L. DEVROYE, G. LUGOSI, *Combinatorial Methods in Density Estimation*, Springer, New York, 2001.
- [7] K. FAN, *Minimax theorems*, Proc. Nat. Acad. Sci. USA, 39 (1953), pp. 42-47.
- [8] J. B G. FRENK, G. KASSAY, J. KOLUMBÁN, *On equivalent results in minimax theory*, European Journal of Operational Research, 157 (2004), 46-58.
- [9] E. I. GORDIENKO, J. A. MINJÁREZ-SOSA, *Adaptive control for discrete-time Markov processes with unbounded costs: Discounted criterion*, Kybernetika, 34 (1998), pp. 217–234.
- [10] O. HERNÁNDEZ-LERMA, *Adaptive Markov Control Processes*, Springer-Verlag, New York, 1989.
- [11] O. HERNÁNDEZ-LERMA, R. CAVAZOS-CADENA, *Density estimation and adaptive control of Markov processes: Average and discounted criteria*, Acta Appl. Math., 20 (1990), pp. 285–307.

- [12] O. HERNÁNDEZ-LERMA, J. B. LASSERRE, *Discrete-Time Markov Control Processes: Basic Optimality Criteria*, Springer-Verlag, New York, 1996.
- [13] O. HERNÁNDEZ-LERMA, J. B. LASSERRE, *Further Topics on Discrete-Time Markov Control Processes*, Springer-Verlag, New York, 1999.
- [14] N. HILGERT, J. A. MINJÁREZ-SOSA, *Adaptive control of stochastic systems with unknown disturbance distribution: Discounted criteria*, Math. Methods Oper. Res., 63 (2006), pp. 443–460.
- [15] HILMMEBERG, C. J. PARTHASARATHY, VAN VLECK, *Optimal plans for dynamic programming problems*, Math. Oper. Res. 1 (1976), pp. 390-394.
- [16] C. J. HIMMELBERG, F. S. VAN VLECK, *Multifunctions with values in a space of probability measures*, J. Math. Anal. Appl., 50 (1975), pp. 108-112.
- [17] A. JAŚKIEWICZ, A. S. NOWAK, *Zero-sum ergodic stochastic games with Feller transition probabilities*, SIAM J. Control Optim., 45 (2006), pp. 773–789.
- [18] H.-U. KÜENLE, *On Markov games with average reward criterion and weakly continuous transition probabilities*, SIAM J. Control Optim., 45 (2007), pp. 2156–2168.
- [19] E. MICHAEL, *Continuos selections I*, Ann. Math., 63 (1956), pp. 361-382.
- [20] J. A. MINJÁREZ-SOSA, *Nonparametric adaptive control for discrete-time Markov processes with unbounded costs under average criterion*, Appl. Math. (Warsaw), 26 (1999), pp. 267–280.
- [21] J. A. MINJÁREZ-SOSA, *Approximation and estimation in Markov control processes under a discounted criterion*, Kybernetika, 40 (2004), pp. 681–690.
- [22] J. A. MINJÁREZ-SOSA, O. VEGA-AMAYA, *Asymptotically optimal strategies for adaptive zero-sum discounted Markov games*, SIAM J. Control Optim. 48 (2009), pp. 1405-1421.
- [23] J. A. MINJÁREZ-SOSA, O. VEGA-AMAYA, *Average optimal strategies for zero-sum Markov games with unknown disturbance distribution on one side*, (en preparación).

- [24] A. S. NOWAK, *Optimal Strategies in a class of zero-sum ergodic stochastic games*, Math. Methods Oper. Res., 50 (1999), pp. 466-476.
- [25] F. RAMIREZ-REYES, *Existence of optimal strategies for zero-sum stochastic games with discounted payoff*, Dpto. de Matemáticas, CINVESTAV (2000).
- [26] M. SCHÄL, *Estimation and control in discounted stochastic dynamic programming*, Stochastics, 20 (1987), pp. 51-71.
- [27] J. A. E. E. VAN NUNEN, J. WESSELS, *A note on dynamic programming with unbounded rewards*, Management Sci., 24 (1977/78), pp. 576-580.