



Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias Físico Matemáticas
Posgrado en Ciencias Matemáticas

Juegos estocásticos y su aplicación en
algunos juegos clásicos iterados

Tesis

presentada para obtener el título de
Maestría en Ciencias Matemáticas

Presenta
Ciria Ruth Briones García (CVU 774000)

Director de Tesis
Dr. Víctor Hugo Vázquez Guevara (CVU
165488)

Puebla, Puebla. Diciembre 2018

A mis abuelitos...

Agradecimientos

Quiero agradecer a todas las personas que creyeron en este proyecto, desde el inicio hasta el final.

Al Consejo Nacional de Ciencia y Tecnología (CONACyT), por su apoyo durante toda la maestría, ya que con este fue posible la realización de esta tesis.

Al Dr. Víctor Hugo Vázquez Guevara por la dedicación al proyecto, la disponibilidad y paciencia que mostró durante todo el desarrollo y sobre todo por la confianza que ha depositado en mi, sin su dirección esto no hubiera sido posible.

A mis sinodales: Dr. Hugo Adán Cruz Suárez, Dr. Francisco Solano Tajonar Sanabria, Dr. Fernando Velasco Luna y el Dr. José Raúl Montes de Oca Machorro, por sus correcciones y contribuciones al trabajo, además de su disponibilidad y deseo de ayudar.

A mis amigos que hicieron amenos los momentos difíciles y sobre todo que se tomaron el tiempo de estudiar conmigo.

Por último, pero no menos importante, a mi familia: papá, mami, Mayis, Julito, Tony, Toño y mi Key, gracias por creer en mi y apoyarme siempre, este trabajo es por ustedes.

Introducción

La teoría de juegos desarrolla modelos matemáticos, estudia dichos modelos y en particular busca encontrar la forma de elegir las estrategias para optimizar el desempeño de los jugadores. Surge como una teoría económica, pero en la actualidad tiene aplicaciones en muchas disciplinas tales como la biología, informática, inteligencia artificial, psicología, etc.

En 1944 se inicia formalmente la teoría de juegos con la publicación del libro *The theory of games and economic behavior* de John Von Neumann y Oskar Morgenstern, aunque antes Cournot, Bertrand y Edgeworth habían planteado ciertas ideas sobre el tema.

John Von Neumann y Oskar Morgenstern estudiaron dos planteamientos: el comportamiento no cooperativo y el cooperativo. En el primero se sabe específicamente lo que cada jugador puede hacer durante el juego y después se busca una estrategia óptima, en el segundo buscaron describir la conducta óptima en juegos con muchos jugadores.

Desde ese momento hubo un desarrollo importante en esta teoría. John Forbes Nash escribió una tesis, bajo la dirección de Albert W. Tucker, en la que expuso por primera vez su solución para juegos no cooperativos, lo que llamó *equilibrio de Nash* y con lo que consiguió un gran reconocimiento entre los especialistas. Esto permitió extender la teoría de juegos no cooperativos más generales que los de suma cero. Durante esa época la mayor parte de las aplicaciones se concentraban en temas de estrategia militar. El equilibrio de Nash es una situación en la que ninguno de los jugadores siente la necesidad de cambiar de estrategia ya que cualquier cambio implicaría una disminución de las ganancias.

Las personas involucradas en una situación de juego tienen intereses que pueden concordar en algunos casos y oponerse

en otros, de manera que pueden perseguir objetivos en común y enfrentarse por otros.

El concepto de juego estocástico fue introducido por Shapley en 1953. A diferencia de los juegos en forma normal, los juegos estocásticos, también llamados procesos de decisión de Markov competitivos, son juegos dinámicos ya que se juegan en varias etapas. Pueden ser vistos como una generalización de los procesos de decisión de Markov, en los que se permite que varios agentes tomen decisiones persiguiendo objetivos diferentes.

En 1997, Filar y Vrieze publicaron el libro *Competitive Markov decision processes* cuyo título refiere a los juegos estocásticos, en el que se tratan ambos temas (véase también [2], [5] y [11]).

La literatura para juegos estáticos es bastante amplia, mientras que para juegos dinámicos no sucede lo mismo. En la tesis estudiaremos algunos artículos sobre los juegos dinámicos, algunos de ellos son los siguientes: para juegos de suma no cero los artículos [1] y [9], y para juegos de suma cero [6], [8] y [10]. Trabajamos con estos artículos debido a que comparten notación, suposiciones, definiciones, entre otros.

En términos muy generales, el modelo consiste en un conjunto de jugadores que en una secuencia de pasos participan en un juego en forma normal, que depende del estado en el que se encuentra el juego y de las decisiones tomadas por los jugadores, además, el pago propio del juego en forma normal, influye en el estado en que estará el juego estocástico en el paso siguiente.

Índice general

Introducción	VIII
Índice de figuras	XIII
Índice de tablas	XVI
Símbolos	XVIII
1. Preliminares	1
1.1. Teoría de juegos	1
1.2. Teoría de juegos estocásticos	3
1.2.1. Modelo del juego estocástico	4
1.3. Estrategias	6
1.4. Criterios de pago	8
2. Juegos de suma cero	11
2.1. Juegos de suma cero	12
2.2. El juego del volado	36
3. Juegos de suma no cero	43
3.1. Juegos de suma no cero	43
3.2. Dilema del Prisionero Iterado	53
Conclusiones	73
Bibliografía	75

Índice de figuras

1.1. Modelo del juego.	5
2.1. Gráfica de S_a , S_s y S_p cuando el jugador 1 elige S_a	38
2.2. Gráfica de S_a , S_s y S_p cuando el jugador 1 elige S_s	38
2.3. Gráfica de S_a , S_s y S_p cuando el jugador 1 elige S_p	39
2.4. Gráfica de S_a , S_s y S_p cuando el jugador 2 elige S_p	39

Índice de tablas

1.1. Forma normal.	3
2.1. Forma normal juego del volado.	36
2.2. Ganancias usando el criterio EAP para el juego del volado.	37
2.3. Mejores respuestas usando el criterio EAP para el juego del volado.	40
3.1. Forma normal del Dilema del Prisionero.	53
3.2. Forma normal del juego.	54
3.3. S_D vs S_D	57
3.4. S_D vs S_C	58
3.5. S_D vs S_G	58
3.6. S_D vs S_T	58
3.7. S_D vs S_A	59
3.8. S_D vs S_Q	59
3.9. S_D vs S_P	59
3.10. S_C vs S_D	59
3.11. S_C vs S_C	60
3.12. S_C vs S_G	60
3.13. S_C vs S_T	60
3.14. S_C vs S_A	60
3.15. S_C vs S_Q	61
3.16. S_C vs S_P	61
3.17. S_G vs S_D	61
3.18. S_G vs S_C	62
3.19. S_G vs S_G	62

3.20. S_G vs S_T	62
3.21. S_G vs S_A	62
3.22. S_G vs S_Q	63
3.23. S_G vs S_P	63
3.24. S_T vs S_D	63
3.25. S_T vs S_C	63
3.26. S_T vs S_G	64
3.27. S_T vs S_T	64
3.28. S_T vs S_A	64
3.29. S_T vs S_Q	64
3.30. S_T vs S_P	65
3.31. S_A vs S_D	65
3.32. S_A vs S_C	65
3.33. S_A vs S_G	66
3.34. S_A vs S_T	66
3.35. S_A vs S_A	66
3.36. S_A vs S_Q	66
3.37. S_A vs S_P	67
3.38. S_Q vs S_D	67
3.39. S_Q vs S_C	67
3.40. S_Q vs S_G	67
3.41. S_Q vs S_T	68
3.42. S_Q vs S_A	68
3.43. S_Q vs S_Q	68
3.44. S_Q vs S_P	68
3.45. S_P vs S_D	69
3.46. S_P vs S_C	69
3.47. S_P vs S_G	69
3.48. S_P vs S_T	69
3.49. S_P vs S_A	70
3.50. S_P vs S_Q	70
3.51. S_P vs S_P	70
3.52. EAP de los jugadores.	71
3.53. La mejor respuesta para ambos jugadores.	71

Símbolos

$\mathcal{B}(X)$	σ -álgebra de Borel de un espacio de Borel X .
$\mathbb{P}(X)$	Familia de medidas de probabilidad en X .
C_b	Espacio de funciones acotadas y continuas.
$\mathbb{P}(X Y)$	Conjunto de todos los kérneos estocásticos sobre X dado Y .
X	Espacio de estados (espacio de Borel).
A	Espacio de acciones para el jugador 1.
B	Espacio de acciones para el jugador 2.
\mathbb{K}_A	Conjunto de restricciones para el jugador 1.
\mathbb{K}_B	Conjunto de restricciones para el jugador 2.
$A(x)$	Conjunto de acciones admisibles para el jugador 1 cuando el sistema se encuentra en el estado x .
$B(x)$	Conjunto de acciones admisibles para el jugador 2 cuando el sistema se encuentra en el estado x .
\mathbb{K}	Conjunto de las 3-tuplas que constan de un estado, una acción para el jugador 1 y una acción para el jugador 2.
$Q(\cdot x, a, b)$	Distribución de probabilidad del siguiente estado del juego.
r_i	Función de recompensa para el jugador i .
H_t	Historia del juego hasta el tiempo t .
π^i	Estrategia aleatoria para el jugador i .
π_t^i	Kernel estocástico para el jugador i en el tiempo t .
Π_i	Familia de todas las estrategias para el jugador i .
Φ_i	Clase de todas las probabilidades de transición para el jugador i .
$\mathbb{P}_A(x)$	Familia de medidas de probabilidad en $A(x)$.
$\mathbb{P}_B(x)$	Familia de medidas de probabilidad en $B(x)$.

Π_{iM}	Conjunto de todas las estrategias de Markov para el jugador i .
Π_{iS}	Conjunto de todas las estrategias de Markov aleatorizadas para el jugador i .
$P_x^{\pi^1 \pi^2}$	Medida de probabilidad para el par de estrategias π^1, π^2 y el estado inicial x .
$E_x^{\pi^1 \pi^2}$	Operador esperanza con respecto a $P_x^{\pi^1 \pi^2}$.
$J_n^{i,0}$	Trayectoria de la muestra de n etapas para el jugador i .
J_n^i	Ganancia esperada en la n -ésima etapa para el jugador i .
$J^{i,0}$	Ganancia media de la trayectoria de la muestra de largo plazo (SPAP).
J^i	Rentabilidad media esperada de largo plazo (EAP).
$L(\cdot)$	Valor inferior del juego.
$U(\cdot)$	Valor superior del juego.
$\mathbb{B}_w(X)$	Espacio lineal de las funciones medibles reales u definidas en X con norma finita w .
$\mathbb{M}_w(X)$	Espacio lineal normado de las medidas finitas con signo $\mu \in X$.
$q(\varphi, \psi)$	Denota una medida de probabilidad invariante única en $\mathbb{M}_w(X)$.
ξ_*	Constante en \mathbb{R} , elemento de una cuádrupla canónica.
u_*	Función medible en $\mathbb{B}_w(X)$, elemento de una cuádrupla canónica.

Capítulo 1

Preliminares

Este capítulo es una breve introducción a la teoría de juegos. La clasificación de los juegos puede hacerse de acuerdo al número de jugadores: biperonal, n -personales; al número de estrategias: finito o infinito; evolución temporal: estático o dinámico; intercambio de información entre los jugadores: cooperativo o no cooperativo; a las ganancias: suma cero o suma no cero; cantidad de información de la que disponen los jugadores: completa o incompleta; cantidad de información que adquieren los jugadores durante el juego: perfecta o imperfecta; etc.

Sin embargo, esta tesis se centra en los juegos estocásticos e iniciamos el capítulo presentando qué es y cómo se define un juego. Para la segunda sección hacemos la transición a un juego estocástico y presentamos el modelo del juego. Las últimas dos secciones hablan de las estrategias que usan los jugadores en este tipo de juegos y los criterios de pago que consideramos en el segundo capítulo de la tesis.

1.1. Teoría de juegos

Para el estudio de la teoría de juegos los conceptos de juego, estrategia y ganancia son básicos y necesarios. Así que

se inicia esta sección con esas definiciones.

Definición 1.1.1. Un *juego* es cualquier situación de decisión caracterizada por una interdependencia estratégica, gobernada por reglas y con un resultado definido.

El resultado que se obtiene es al que llamamos *ganancia*, que depende de las estrategias elegidas por las personas involucradas en el juego.

Definición 1.1.2. Una *estrategia* es una regla para elegir una acción en cada punto donde una decisión puede ser tomada.

Los ejemplos que se presentan durante el desarrollo de la tesis son juegos estáticos sin embargo, en capítulos posteriores se considerarán de forma iterada.

Definición 1.1.3. Un *juego estático* es aquel en el que cada jugador toma solo una decisión y finaliza el juego. Además ningún jugador tiene conocimiento de la decisión tomada por los otros jugadores antes de tomar su propia decisión.

La solución de un juego debería indicar a cada jugador qué resultado esperar y cómo alcanzarlo. Los jugadores intentan obtener el mejor resultado para sus intereses.

Para representar un juego, sus jugadores, estrategias y ganancias usamos la forma normal de un juego que es una descripción tabular de este.

Supongamos que se tiene un juego con dos jugadores, J_1 y J_2 , con el mismo conjunto de estrategias $\{X, Y\}$. Las ganancias de los jugadores están dadas de la siguiente manera: Si J_1 elige X y también J_2 , la ganancia para J_1 es a y para J_2 es b pero si J_2 elige Y las ganancias para los jugadores son c y d , respectivamente. Si J_1 elige Y y J_2 elige X la ganancia de J_1 es e y para J_2 es f , pero si J_2 elige Y las ganancias son g y h , respectivamente. La forma tabular está dada por la Tabla 1.1.

Tabla 1.1: Forma normal.

		J_2	
		X	Y
J_1	X	a, b	c, d
	Y	e, f	g, h

Cuando se habla de la *mejor respuesta* para un jugador, nos referimos a la estrategia con la que el jugador obtiene su mayor ganancia. Un *equilibrio de Nash* es cuando ambos jugadores usan su mejor respuesta, es decir, que los jugadores maximizan sus ganancias.

Con esto es suficiente para dar paso a la teoría de juegos estocásticos.

1.2. Teoría de juegos estocásticos

Iniciamos esta sección con notación y definiciones básicas para poder establecer nuestro modelo del juego estocástico. A partir de este momento trabajaremos con juegos de dos jugadores ya que el caso de n jugadores es análogo.

Los juegos estocásticos son juegos más complejos de los que vimos en la sección anterior, ya que representan un sistema que evoluciona con el tiempo. Para poder definir el modelo del juego necesitamos lo siguiente.

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

Dado un espacio de Borel X , denotamos por $\mathbb{P}(X)$ la familia de medidas de probabilidad en X , dotada de la topología débil $\sigma(\mathbb{P}(X), C_b(X))$, donde C_b representa el espacio de funciones acotadas y continuas en X . Hay que notar que $\mathbb{P}(X)$ es

un espacio de Borel, además, si X es compacto $\mathbb{P}(X)$ también lo es.

Definición 1.2.1. Sean X y Y espacios de Borel. Un kernel estocástico en X dado Y , o también conocida como probabilidad de transición de Y a X , es una función medible $\nu(\cdot|\cdot)$ tal que

- $\nu(\cdot|y)$ es una medida de probabilidad sobre X para cada $y \in Y$, y
- $\nu(D|\cdot)$ es una función medible sobre Y para cada $D \in \mathcal{B}(X)$.

El conjunto de todos los kernels estocásticos sobre X dado Y es denotado por $\mathbb{P}(X|Y)$. Si ν está en $\mathbb{P}(X|Y)$, entonces escribimos sus valores como $\nu(B|y)$ para todo $y \in Y$ y $B \in \mathcal{B}(X)$. Finalmente, si $X = Y$ entonces ν se llama probabilidad de transición de Markov en X .

1.2.1. Modelo del juego estocástico

En la sección anterior se ha definido un kernel estocástico, con ello podemos iniciar nuestro modelo del juego. Pero antes daremos una breve explicación de en qué consisten este tipo de juegos.

Consideraremos el modelo del juego estocástico de suma no cero de dos personas,

$$GM := (X, A, B, \mathbb{K}_A, \mathbb{K}_B, Q, r_1, r_2), \quad (1.1)$$

donde:

- X es el espacio de estados, que se supone de Borel,
- A y B son los espacios de acciones para los jugadores 1 y 2, respectivamente. Se supone que también son espacios de Borel,

- $\mathbb{K}_A \in \mathcal{B}(X \times A)$ y $\mathbb{K}_B \in \mathcal{B}(X \times B)$ son los conjuntos de restricciones y para cada $x \in X$ la x -sección no vacía

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

de A representa el conjunto de acciones disponibles 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 para el jugador 2 en el estado x . Definimos además a

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

- $Q \in \mathbb{P}(X|\mathbb{K})$ es un kernel estocástico, llamado la ley de transición. Si x es el estado en alguna etapa del juego y los jugadores seleccionan las acciones $a \in A(x)$ y $b \in B(x)$, entonces $Q(\cdot|x, a, b)$ es la distribución de probabilidad del siguiente estado del juego,
- $r_i : \mathbb{K} \rightarrow \mathbb{R}$ es una función medible que representa la función de recompensa para el jugador $i = 1, 2$.

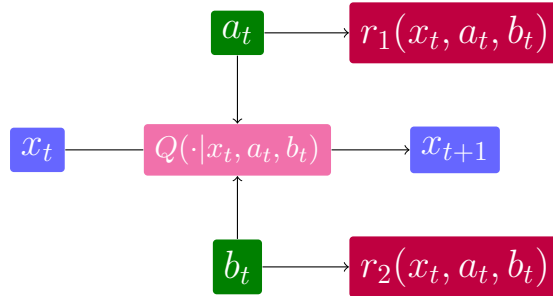


Figura 1.1: Modelo del juego.

El juego transcurre de la manera siguiente. En cada etapa $t = 0, 1, \dots$, los jugadores 1 y 2 observan el estado actual $x \in X$ del sistema, y eligen de manera independiente las acciones $a \in A(x)$ y $b \in B(x)$, respectivamente. Como consecuencia, sucede que:

- (1) el jugador i recibe una recompensa inmediata $r_i(x, a, b)$,
 $i = 1, 2$, y
- (2) el sistema se mueve a un nuevo estado con distribución
 $Q(\cdot|x, a, b)$

El objetivo de cada jugador es maximizar algún criterio de rendimiento.

Cuando el juego es de suma cero, en (1.1) consideramos a $r_i(\cdot) = r(\cdot)$ ($i = 1, 2$) tal que $r_1(\cdot) = -r_2(\cdot)$.

1.3. Estrategias

Como se mencionó anteriormente las estrategias son parte fundamental de un juego, y en los juegos estocásticos no son la excepción. Pero para poder definir las es importante antes mencionar lo siguiente.

Sean $H_0 := X$ y $H_t := \mathbb{K} \times H_{t-1}$ para $t = 1, 2, \dots$. Para cada t , un elemento

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

de H_t representa una historia del juego hasta el tiempo t .

Una estrategia aleatoria π^1 para el jugador 1 es una sucesión $\pi^1 = \{\pi_t^1, t = 0, 1, \dots\}$, de kérneos estocásticos en $\mathbb{P}(A|H_t)$ tal que

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

Denotamos por Π_1 la familia de todas las estrategias para el jugador 1.

Definimos $\mathbb{P}_A(x) := \mathbb{P}(A(x))$ para cada estado $x \in X$, y Φ_1 como la clase de todas las probabilidades de transición

$\phi \in \mathbb{P}(A|X)$ tal que $\phi(\cdot|x)$ esta en $\mathbb{P}_A(x)$ para todo $x \in X$, es decir, $\Phi_1 = \{\phi \in \mathbb{P}(A|X) : \phi(\cdot|x) \in \mathbb{P}_A(x), \forall x \in X\}$.

Los conjuntos Π_2 y Φ_2 para el jugador 2 se definen de forma similar, escribiendo $B(x)$ en lugar de $A(x)$ y $\mathbb{P}_B(x) := \mathbb{P}(B(x))$.

Una estrategia $\pi = \{\pi_t\}$ se llama de Markov aleatorizada si $\pi_t \in \mathbb{P}(A|X)$ para cada $t = 0, 1, \dots$, esto es, cada π_t depende solamente del estado actual del sistema. El conjunto de todas las estrategias de Markov del jugador 1 será denotado por Π_{1M} . Una estrategia de Markov aleatorizada $\pi = \{\pi_t\}$ se dice que es estacionaria si existe $f \in \Phi_1$ tal que $\pi_t = f$ para cada $t = 0, 1, \dots$. En este caso, la estrategia π se identificará con f . Denotamos por Π_{1S} al conjunto de todas las estrategias markovianas aleatorizadas estacionarias del jugador 1. Tenemos, desde luego que

$$\Pi_{1S} \subset \Pi_{1M} \subset \Pi_1.$$

Los conjuntos $\Pi_2, \Pi_{2M}, \Pi_{2S}$ de todas las estrategias, de las estrategias de Markov y las estrategias estacionarias, respectivamente, para el jugador 2 se definen de manera similar.

Sea (Ω, \mathcal{F}) el espacio canónico medible 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 $(\pi^1, \pi^2) \in \Pi_1 \times \Pi_2$ y cada estado inicial $x \in X$, por el Teorema de C. Ionescu-Tulcea, existe una única medida de probabilidad $P_x^{\pi^1, \pi^2}$ 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 de los jugadores 1 y 2, respectivamente, en cada etapa $t = 0, 1, \dots$. El operador de esperanza con respecto a $P_x^{\pi^1, \pi^2}$ se denota por $E_x^{\pi^1, \pi^2}$.

Es importante recordar que los jugadores toman decisiones

de manera independiente, es decir, que para cualquier par de estrategias $\pi^i \in \Pi_i$ ($i = 1, 2$) y cualquier estado inicial $x \in X$, los procesos $\{a_t\}$ y $\{b_t\}$ son condicionalmente independientes, esto quiere decir que

$$P_x^{\pi^1 \pi^2}(a_t \in C, b_t \in D | h_t) = \pi_t^1(C | h_t) \pi_t^2(D | h_t),$$

para todo $C \in \mathcal{B}(A)$, $D \in \mathcal{B}(B)$, $h_t \in H_t$ y $t = 0, 1, \dots$

1.4. Criterios de pago

Ahora podemos establecer el objetivo del juego, acumulando las ganancias de los jugadores durante la evolución del mismo pero consideraremos los siguientes criterios.

Para cada $n = 1, 2, \dots$. Y cada historia $h_\infty := (x_0, a_0, b_0, x_1, a_1, b_1, \dots)$, sea

$$J_n^{i,0}(h_\infty) := \sum_{t=0}^{n-1} r_i(x_t, a_t, b_t), \quad (1.2)$$

la trayectoria de la muestra de n etapas cuando el jugador i ($i = 1, 2$) usa la estrategia $\pi^i \in \Pi_i$, dado el estado inicial $x_0 = x$. Es necesaria toda la historia del juego para definir uno de los criterios (1.2). Cuando el juego es de suma cero la expresión (1.2) quedará de la siguiente forma:

$$J_n^0(h_\infty) := \sum_{t=0}^{n-1} r(x_t, a_t, b_t).$$

La ganancia esperada en la n -ésima etapa es

$$J_n^i(x, \pi^1, \pi^2) := E_x^{\pi^1 \pi^2} \left[\sum_{t=0}^{n-1} r_i(x_t, a_t, b_t) \right], \quad (1.3)$$

con $x \in X$ y $\pi^i \in \Pi_i$ ($i = 1, 2$). Para el caso de suma cero, (1.3) queda expresada como:

$$J_n(x, \pi^1, \pi^2) := E_x^{\pi^1 \pi^2} \left[\sum_{t=0}^{n-1} r(x_t, a_t, b_t) \right].$$

A continuación, definimos la ganancia media de la trayectoria de la muestra de largo plazo (**SPAP**, sample-path average payoff)

$$J^{i,0}(h_\infty) := \liminf_{n \rightarrow \infty} \frac{J_n^{i,0}(h_\infty)}{n}, \quad (1.4)$$

y de manera similar, la rentabilidad media esperada de largo plazo (**EAP**, expected average payoff)

$$J^i(x, \pi^1, \pi^2) := \liminf_{n \rightarrow \infty} \frac{J_n^i(x, \pi^1, \pi^2)}{n}. \quad (1.5)$$

Para juegos de suma cero, (1.4) y (1.4) quedan de la siguiente forma:

$$J^0(h_\infty) := \liminf_{n \rightarrow \infty} \frac{J_n^0(h_\infty)}{n}$$

y

$$J(x, \pi^1, \pi^2) := \liminf_{n \rightarrow \infty} \frac{J_n(x, \pi^1, \pi^2)}{n}.$$

Para introducir los criterios de optimalidad que nos interesan consideremos los siguientes conceptos. Las funciones en X definidas como

$$L(x) := \sup_{\pi^1 \in \Pi_1} \inf_{\pi^2 \in \Pi_2} J(x, \pi^1, \pi^2) \quad \text{y}$$

$$U(x) := \inf_{\pi^2 \in \Pi_2} \sup_{\pi^1 \in \Pi_1} J(x, \pi^1, \pi^2).$$

Se denominan el valor inferior y el valor superior (respectivamente) del juego. Es claro que $L(\cdot) \leq U(\cdot)$ en general, pero si se cumple la igualdad para todo $x \in X$, entonces la función común se llama el valor del juego, que se representa por $V(\cdot)$.

Capítulo 2

Juegos de suma cero

Como mencionamos al inicio del Capítulo 1, hay diferentes formas de clasificar los juegos y que una de ellas son los juegos estocásticos pero dentro de ellos también podemos agruparlos en juegos estocásticos de suma cero y juegos estocásticos de suma no cero. En este capítulo abordaremos los juegos estocásticos de suma cero, el cual esta dividido en dos secciones.

En la primera presentamos los teoremas del artículo [6], que afirman las equivalencias de los conjuntos de estrategias de los diferentes criterios, además de las diferentes suposiciones necesarias para que las equivalencias se cumplan. Esta sección esta dividida en subsecciones con el fin de presentar por separado cada uno de los teoremas y sus demostraciones. El primer teorema nos da la equivalencia entre los conjuntos de estrategias EAP y los pares canónicos, y el segundo, además de esto da una equivalencia de martingala, el último teorema nos asegura la equivalencia entre los conjuntos de estrategias bajo los criterios SPAP y EAP.

La segunda sección es un ejemplo usado para ilustrar los juegos de suma cero usando el criterio EAP. Este ejemplo es un juego clásico en la teoría de juegos, el juego del volado, que es adaptado a un juego estocástico para poder ser estudiado

con la teoría presentada anteriormente.

2.1. Juegos de suma cero

En esta sección tratamos de exponer de forma detallada las pruebas de los teoremas presentados en [6]. Para esto necesitamos algunas definiciones y resultados previos que se presentan a continuación.

Usaremos (1.1) con $r_i(\cdot) = r(\cdot)$ ($i = 1, 2$) ya que $r_1 = -r_2$, así (1.2), (1.3), (1.4) y (1.4) podemos reescribirlas como J_n^0 , J_n , J^0 y J , respectivamente.

Definición 2.1.1. Supongamos que existe valor del juego, $V(\cdot)$. Entonces, se dice que una estrategia π^{1*} en Π_1 es EAP óptima para el jugador 1 si

$$\inf_{\pi^2 \in \Pi_2} J(x, \pi^{1*}, \pi^2) = V(x), \quad \forall x \in X. \quad (2.1)$$

Análogamente, $\pi^{2*} \in \Pi_2$ es EAP óptima para el jugador 2 si

$$\sup_{\pi^1 \in \Pi_1} J(x, \pi^1, \pi^{2*}) = V(x), \quad \forall x \in X. \quad (2.2)$$

Cuando se cumplen ambas igualdades, decimos que (π^{1*}, π^{2*}) es un par EAP óptimo.

Para el caso SPAP introducimos la siguiente definición.

Definición 2.1.2. Supongamos que el juego tiene valor $V(\cdot)$. Entonces, se dice que un par de estrategias $(\pi^{1*}, \pi^{2*}) \in \Pi_1 \times \Pi_2$ es SPAP óptimo si satisface que, para todo $x \in X$ y $\pi^i \in \Pi_i$ ($i = 1, 2$),

$$\begin{aligned} J^0(h_\infty) &= V(x), & P_x^{\pi^{1*}\pi^{2*}} \text{ c.s.}, \\ J^0(h_\infty) &\geq V(x), & P_x^{\pi^{1*}\pi^2} \text{ c.s.}, \\ J^0(h_\infty) &\leq V(x), & P_x^{\pi^1\pi^{2*}} \text{ c.s.} \end{aligned} \quad (2.3)$$

Y, para un último criterio utilizamos la siguiente notación. Para cualquier función $f : \mathbb{K} \rightarrow \mathbb{R}$ y medidas de probabilidad $\varphi \in \mathbb{P}_A(x)$ y $\psi \in \mathbb{P}_B(x)$, escribimos

$$f(x, \varphi, \psi) := \int_{A(x)} \int_{B(x)} f(x, a, b) \psi(db) \varphi(da),$$

siempre que las integrales esten bien definidas. En particular, para r y Q como en (1.1), tenemos que

$$r(x, \varphi, \psi) := \int_{A(x)} \int_{B(x)} r(x, a, b) \psi(db) \varphi(da) \quad (2.4)$$

y

$$Q(\cdot|x, \varphi, \psi) := \int_{A(x)} \int_{B(x)} Q(\cdot|x, a, b) \psi(db) \varphi(da).$$

Definición 2.1.3. Una cuádrupla $(\xi_*, u_*, \varphi_*, \psi_*)$ que consiste en una constante $\xi_* \in \mathbb{R}$, una función medible $u_* : X \rightarrow \mathbb{R}$ y un par $(\varphi_*, \psi_*) \in \Phi_1 \times \Phi_2$ de estrategias estacionarias se dice que es una cuádrupla canónica si se cumple que, para todo $x \in X$

$$\xi_* + u_*(x) = r(x, \varphi_*(x), \psi_*(x)) + \int_X u_*(y) Q(dy|x, \varphi_*(x), \psi_*(x)) \quad (2.5)$$

$$= \max_{\varphi \in \mathbb{P}_A(x)} [r(x, \varphi, \psi_*(x)) + \int_X u_*(y) Q(dy|x, \varphi, \psi_*(x))] \quad (2.6)$$

$$= \min_{\psi \in \mathbb{P}_B(x)} [r(x, \varphi_*(x), \psi) + \int_X u_*(y) Q(dy|x, \varphi_*(x), \psi)] \quad (2.7)$$

En este caso, se dice que (φ_*, ψ_*) es un par canónico de estrategias estacionarias.

La definición anterior esta relacionada con la llamada ecuación de Shapley (o de programación dinámica)

$$\xi_* + u_*(x) = Tu_*(x), \quad \forall x \in X,$$

donde T es el operador minimax definido por

$$\begin{aligned}
 Tu_*(x) &:= \max_{\varphi \in \mathbb{P}_A(x)} \min_{\psi \in \mathbb{P}_B(x)} [r(x, \varphi, \psi) + \\
 &\quad \int_X u_*(y) Q(dy|x, \varphi, \psi)] \\
 &= \min_{\psi \in \mathbb{P}_B(x)} \max_{\varphi \in \mathbb{P}_A(x)} [r(x, \varphi, \psi) + \\
 &\quad \int_X u_*(y) Q(dy|x, \varphi, \psi)].
 \end{aligned} \tag{2.8}$$

Denotamos por $(\Phi_1 \times \Phi_2)_{ca}$ a la familia de pares canónicos, por $(\Phi_1 \times \Phi_2)_{eap}$ a la familia de estrategias EAP óptimas y por $(\Phi_1 \times \Phi_2)_{spap}$ a la familia de pares SPAP óptimos de estrategias estacionarias.

A continuación presentamos algunas suposiciones sobre continuidad-compacidad para los conjuntos de estados y acciones admisibles para los jugadores, condiciones de crecimiento para la función de recompensa, entre otros, que son necesarios para la existencia de estrategias óptimas para los criterios de pago presentados en el Capítulo 1.

Suposición 2.1.1. a) Para cada estado $x \in X$, los conjuntos (no vacíos) $A(x)$ y $B(x)$ de acciones admisibles son compactos.

b) Para cada (x, a, b) en \mathbb{K} , $r(x, \cdot, b)$ es semicontinua superior (u.s.c.) en $A(x)$ y $r(x, a, \cdot)$ es semicontinua inferior (l.s.c.) en $B(x)$.

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

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

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

- d) Existe una constante \bar{r} y una función medible $w(\cdot) \geq 1$ definida en X tal que

$$|r(x, a, b)| \leq \bar{r}w(x), \quad \forall (x, a, b) \in \mathbb{K}, \quad (2.9)$$

y, además, la parte c) se cumple cuando v se reemplaza con w .

Las siguientes suposiciones garantizan que el proceso de estado $\{x_t\}$ es ergódico.

Suposición 2.1.2. Existe una medida de probabilidad $\nu \in \mathbb{P}(X)$, un número positivo $\alpha < 1$, y una función medible $\beta : \mathbb{K} \rightarrow [0, 1]$ tales que para todo $(x, a, b) \in \mathbb{K}$ y $D \in \mathcal{B}(X)$:

- a) $Q(D|x, a, b) \geq \beta(x, a, b)\nu(D)$.
- b) $\int_X w(y)Q(dy|x, a, b) \leq \alpha w(x) + \beta(x, a, b)\| \nu \|_w$, donde $w(\cdot) \geq 1$ es la función de la Suposición 2.1 d) y $\| \nu \|_w := \int w d\nu$.

c) $\inf_{(\varphi, \psi) \in \Phi_1 \times \Phi_2} \int_X \beta(x, \varphi(x)\psi(x))\nu(dx) > 0$.

Suposición 2.1.3. Existe una medida σ -finita λ definida en X con respecto a la cual, para cada par (φ, ψ) en $\Phi_1 \times \Phi_2$, la probabilidad de transición de Markov $Q(\cdot|x, \varphi(x), \psi(x))$ es λ -irreducible.

La siguiente definición y notación son necesarias para las suposiciones restantes.

Definición 2.1.4. $\mathbb{B}_w(X)$ denota el espacio lineal de las funciones medibles reales u definidas en X con norma w finita que se define como

$$\|u\|_w := \sup_{x \in X} \frac{|u(x)|}{w(x)}, \quad (2.10)$$

además, $\mathbb{M}_w(X)$ representa el espacio lineal normado de las medidas finitas con signo μ en X tales que

$$\|\mu\|_w := \int_X w d|\mu| < \infty, \quad (2.11)$$

donde $|\mu| := \mu^+ + \mu^-$ denota la variación total de μ .

Note que la integral $\int u d\mu$ es finita para cada u en $\mathbb{B}_w(X)$ y μ en $\mathbb{M}_w(X)$ porque, por (2.10) y (2.11) tenemos que

$$\left| \int u d\mu \right| \leq \|u\|_w \int w d|\mu| = \|u\|_w \|\mu\|_w < \infty.$$

Si se satisfacen las Suposiciones 2.1.2 y 2.1.3, tenemos lo siguiente:

Para cada par $(\varphi, \psi) \in \Phi_1 \times \Phi_2$ el proceso de estados $\{x_t\}$ es Harris positivo recurrente, en particular, la probabilidad de transición $Q(\cdot|x, \varphi(x), \psi(x))$ admite una medida de probabilidad invariante única en $\mathbb{M}_w(X)$, que será denotada por $q(\varphi, \psi)$; así

$$q(\varphi, \psi)(D) = \int_X Q(D|x, \varphi(x), \psi(x)) q(\varphi, \psi)(dx), \quad \forall D \in \mathcal{B}(X). \quad (2.12)$$

$\{x_t\}$ es w -geométricamente ergódico, es decir, existen constantes positivas $\theta < 1$ y M tales que

$$\left| \int_X u(y) Q^n(dy|x, \varphi(x), \psi(x)) - \int_X u(y) q(\varphi, \psi)(dy) \right| \leq w(x) \|u\|_w M \theta^n \quad (2.13)$$

para todo $u \in \mathbb{B}_w(X)$, $x \in X$ y $n = 0, 1, \dots$, donde Q^n denota la probabilidad de transición de Markov de n pasos.

Otra consecuencia de las Suposiciones 2.1.1, 2.1.2 y 2.1.3 es la siguiente proposición.

Proposición 2.1.1. Si las suposiciones 2.1.1, 2.1.2 y 2.1.3 se satisfacen, entonces el juego de recompensa promedio (esperado) tiene un valor constante, digamos $V(x) = V^*$ para todo

$x \in X$, y existe un par óptimo EAP de estrategias estacionarias.

La Suposición 2.1.3 y la proposición anterior son necesarias para la prueba de nuestro primer teorema.

Suposición 2.1.4. Existe una medida σ -finita γ sobre X y una función de densidad estrictamente positiva $g(x, a, b, \cdot)$ tal que

$$Q(D|x, a, b) = \int_D g(x, a, b, y) \gamma(dy)$$

para todo $D \in \mathcal{B}(X)$ y $(x, a, b) \in \mathbb{K}$.

Note que la Suposición 2.1.4 implica la Suposición 2.1.3 con $\lambda = \gamma$.

El siguiente lema muestra, cómo obtener la ecuación de programación dinámica asociada a un par arbitrario de estrategias en $\Phi_1 \times \Phi_2$.

Lema 2.1.1. Supongamos que se cumplen las suposiciones 2.1.1, 2.1.2 y 2.1.3. Entonces, para cada par de estrategias $(\varphi, \psi) \in \Phi_1 \times \Phi_2$ tenemos lo siguiente:

(a) La constante (finita)

$$j(\varphi, \psi) := \int_X r(x, \varphi(x), \psi(x)) q(\varphi, \psi)(dx),$$

con $q(\varphi, \psi)$ como en (2.12), es tal que (1.4) y (1.4) pueden escribirse, para todo $x \in X$, como

$$J^0(h_\infty) = \lim_{n \rightarrow \infty} \frac{J_n^0(h_\infty)}{n} = j(\varphi, \psi) \quad P_x^{\varphi, \psi} \text{ c.s.} \quad (2.14)$$

y

$$J(x, \varphi, \psi) = \lim_{n \rightarrow \infty} \frac{J_n(x, \varphi, \psi)}{n} = j(\varphi, \psi) \quad (2.15)$$

respectivamente.

(b) La función $h_{\varphi,\psi}$ definida en X como

$$\begin{aligned} h_{\varphi,\psi}(x) &:= \lim_{n \rightarrow \infty} [J_n(x, \varphi, \psi) - nj(\varphi, \psi)] \\ &= \sum_{t=0}^{\infty} E_x^{\varphi,\psi} [r(x_t, \varphi(x_t), \psi(x_t)) - j(\varphi, \psi)] \end{aligned}$$

pertenece a $\mathbb{B}_w(X)$ (el conjunto de la Definición 2.1.4), y, además, su w -norma es independiente de (φ, ψ) ; de hecho,

$$\|h_{\varphi,\psi}\|_w \leq \frac{\bar{r}M}{1-\theta}, \quad \forall (\varphi, \psi) \in \Phi_1 \times \Phi_2,$$

con \bar{r} como en (2.9), y M y θ como en (2.13).

(c) El par $(j(\varphi, \psi), h_{\varphi,\psi})$ en $\mathbb{R} \times \mathbb{B}_w(X)$ es la única solución de la ecuación de programación.

$$\begin{aligned} j(\varphi, \psi) + h_{\varphi,\psi}(x) &= r(x, \varphi(x), \psi(x)) + \\ &\int_X h_{\varphi,\psi}(y) Q(dy|x, \varphi(x), \psi(x)) \end{aligned} \quad (2.16)$$

que satisface la condición

$$\int_X h_{\varphi,\psi}(x) q(\varphi, \psi)(dx) = 0.$$

También se puede obtener la ecuación de programación (2.16) usando un “argumento de contradicción”.

Lema 2.1.2. Supongamos que se satisfacen las suposiciones 2.2 y 2.3. Sea (φ, ψ) un par arbitrario en $\Phi_1 \times \Phi_2$, y definimos

$$P(\cdot|x) := Q(\cdot|x, \varphi(x), \psi(x)) \quad \text{y} \quad \mu(\cdot) := q(\varphi, \psi)(\cdot), \quad (2.17)$$

si una función u en $\mathbb{B}_w(X)$ es P -subarmónica (o subinvariante), es decir,

$$u(x) \leq \int_X u(y) P(dy|x), \quad \forall x \in X,$$

entonces $u(\cdot)$ es una constante $\mu - c.d.$; de hecho,

$$u(\cdot) = \inf_{x \in X} u(x) = \int_X u d\mu, \quad \mu - c.d.$$

(con P y μ como en (2.17)). De manera similar, si u es P -subarmónica (o superinvariante), es decir,

$$u(x) \geq \int u(y)P(dy|x),$$

para todo $x \in X$, entonces

$$u(\cdot) = \sup_{x \in X} u(x) = \int_X u d\mu \quad \mu - c.d.$$

Lema 2.1.3. Supongamos que se cumplen las suposiciones 2.1.2 y 2.1.4, y sea γ como en la última suposición. Entonces, para cualquier par (φ, ψ) en $(\Phi_1 \times \Phi_2)$,

- (a) γ es equivalente a la medida de probabilidad invariante $q(\varphi, \psi)$; por lo tanto,
- (b) si $q(\varphi, \psi)(D) = 0$, entonces $Q(D|x, a, b) = 0$ para todo $(x, a, b) \in \mathbb{K}$.

El siguiente lema es resultado de cálculos sencillos usando (2.10) y la desigualdad en la Suposición 2.1.2 (b).

Lema 2.1.4. Suponga que (2.9) y las Suposiciones 2.1.1 (a) y 2.1.2 (b) se satisfacen, y sea $k := 1 + \frac{\|v\|_w}{1-\alpha}$. Entonces para cualquier $\pi^i \in \Pi_i$ ($i = 1, 2$), $x \in X$ y $n = 0, 1, \dots$. Se tiene que

- (a) $E_x^{\pi^1 \pi^2} w(x_n) \leq kw(x)$,
- (b) $E_x^{\pi^1 \pi^2} |r(x_n, a_n, b_n)| \leq \bar{r}kw(x)$.
En particular, por (a) y (2.10), para cada función $u \in \mathbb{B}_w(X)$ tenemos que

- (c) $E_x^{\pi^1 \pi^2} |u(x_n)| \leq \|u\|_w k w(x)$, y entonces
- (d) $\lim_{n \rightarrow \infty} \frac{E_x^{\pi^1 \pi^2} |u(x_n)|}{n} = 0$, y la convergencia es uniforme sobre $\Pi_1 \times \Pi_2$.

Consideremos el operador minimax T en (2.8). Sea $u \in \mathbb{B}_w(X)$ una función arbitraria y definimos:

$$H(u; x, a, b) := r(x, a, b) + \int_X u(u) Q(dy|x, a, b)$$

para $(x, a, b) \in \mathbb{K}$.

Lema 2.1.5. Suponga que se cumplen las suposiciones 2.1.1 y 2.1.2 b). Entonces Tu esta en $\mathbb{B}_w(X)$ para cada u en $\mathbb{B}_w(X)$, y existen estrategias estacionarias $\varphi^* \in \Phi_1$ y $\psi^* \in \Phi_2$ tales que, para todo $x \in X$,

$$\begin{aligned} Tu(x) &= H(u; x, \varphi^*(x), \psi^*(x)) \\ &= \max_{\varphi \in \mathbb{P}_A(x)} H(u; x, \varphi, \psi^*(x)) \\ &= \min_{\psi \in \mathbb{P}_B(x)} H(u; x, \varphi^*(x), \psi). \end{aligned}$$

Definimos $(\Phi_1 \times \Phi_2)_{ca}$ como el conjunto de pares canónicos y $(\Phi_1 \times \Phi_2)_{eap}$ como el conjunto de pares de estrategias bajo el criterio EAP.

Equivalencia de los conjuntos $(\Phi_1 \times \Phi_2)_{eap}$ y $(\Phi_1 \times \Phi_2)_{ca}$.

Esta sección contiene el primer teorema que afirma la igualdad entre pares de estrategias canónicas y pares de estrategias bajo el criterio EAP, además de su demostración.

Teorema 2.1.1. Si las suposiciones 2.1.1, 2.1.2 y 2.1.4 se satisfacen, entonces se tiene

$$(\Phi_1 \times \Phi_2)_{ca} = (\Phi_1 \times \Phi_2)_{eap}.$$

De hecho, existe una cuádrupla canónica $(\xi_*, u_*, \varphi_*, \psi_*)$ con u_* en $\mathbb{B}_w(X)$ y (por la Proposición 2.1.1) $\xi_* = V^*$.

Demostración. Primero mostraremos que $(\Phi_1 \times \Phi_2)_{ca} \subset (\Phi_1 \times \Phi_2)_{cap}$. Para esto, sea $(\xi_*, u_*, \varphi_*, \psi_*)$ una cuádrupla canónica, con $(\varphi_*, \psi_*) \in \Phi_1 \times \Phi_2$, $\xi_* \in \mathbb{R}$ y $u_* \in \mathbb{B}_w(X)$. Por la Definición 2.1.3 tenemos que (2.5) se cumple, sea $x_0 = x$ y considere esta ecuación para $n = 0, 1$

$$\begin{aligned} \xi_* + u_*(x) = & r(x, \varphi_*(x), \psi_*(x)) + \\ & \int_X u_*(x_1) Q(dx_1 | x, \varphi_*(x), \psi_*(x)), \end{aligned} \quad (2.18)$$

$$\begin{aligned} \xi_* + u_*(x_1) = & r(x_1, \varphi_*(x_1), \psi_*(x_1)) + \\ & \int_X u_*(x_2) Q(dx_2 | x_1, \varphi_*(x_1), \psi_*(x_1)). \end{aligned} \quad (2.19)$$

Despejando $u_*(x_1)$ de (2.19) y sustituyendo en (2.18) tenemos que

$$\begin{aligned} \xi_* + u_*(x) = & r(x, \varphi_*(x), \psi_*(x)) + \int_X [-\xi_* + r(x_1, \varphi_*(x_1), \psi_*(x_1))] \\ & + \int_X u_*(x_2) Q(dx_2 | x_1, \varphi_*(x_1), \psi_*(x_1)) Q(dx_1 | x, \varphi_*(x), \psi_*(x)). \end{aligned}$$

De aquí que,

$$\begin{aligned} 2\xi_* + u_*(x) = & r(x, \varphi_*(x), \psi_*(x)) \\ & + \int_X r(x_1, \varphi_*(x_1), \psi_*(x_1)) Q(dx_1 | x, \varphi_*(x), \psi_*(x)) \\ & + \int_X \int_X u_*(x_2) Q(dx_2 | x_1, \varphi_*(x_1), \psi_*(x_1)) Q(dx_1 | x, \varphi_*(x), \psi_*(x)) \\ & = E_x^{\varphi_*, \psi_*} \left[\sum_{t=0}^1 r(x_t, \varphi_*(x_t), \psi_*(x_t)) \right] \\ & + \int_X u_*(x_2) \int_X Q(dx_1 | x, \varphi_*(x), \psi_*(x)) Q(dx_2 | x_1, \varphi_*(x_1), \psi_*(x_1)) \\ & = E_x^{\varphi_*, \psi_*} \left[\sum_{t=0}^1 r(x_t, \varphi_*(x_t), \psi_*(x_t)) \right] + E_x^{\varphi_*, \psi_*} [u_*(x_2)]. \end{aligned}$$

Haciendo este procedimiento n veces obtenemos la siguiente igualdad

$$\begin{aligned} n\xi_* + u_*(x) &= E_x^{\varphi_*\psi_*} \left[\sum_{t=0}^{n-1} r(x_t, \varphi_*(x_t), \psi_*(x_t)) \right] + E_x^{\varphi_*\psi_*} [u_*(x_n)] \\ &= J_n(x, \varphi_*, \psi_*) + E_x^{\varphi_*\psi_*} [u_*(x_n)]. \end{aligned} \quad (2.20)$$

Multiplicando por $\frac{1}{n}$ ambos lados de la segunda igualdad de (2.20), calculando el límite cuando $n \rightarrow \infty$, tenemos que $\frac{u_*(x)}{n} \rightarrow 0$ y por el Lema 2.1.4 (d), se sigue que $\frac{E_x^{\varphi_*\psi_*} [u_*(x_n)]}{n} \rightarrow 0$, así se tiene que

$$\xi_* = \lim_{n \rightarrow \infty} \frac{J_n(x, \varphi_*, \psi_*)}{n},$$

y por (2.15) concluimos que

$$\xi_* = J(x, \varphi_*, \psi_*) = j(\varphi_*, \psi_*)$$

para todo $x \in X$. Solo nos resta mostrar que $\xi_* = V^*$.

De la definición 2.1.3 tenemos que (2.6) se cumple, así que

$$\xi_* + u_*(x) \geq r(x, \varphi, \psi_*(x)) + \int_X u_*(y) Q(dy|x, \varphi, \psi_*(x)) \quad (2.21)$$

para cualesquiera $\varphi \in \mathbb{P}_A(x)$ y $x \in X$.

Realizando un procedimiento similar al que se hizo de (2.18)-(2.20) tomando (2.21) en lugar de (2.5), obtenemos que

$$n\xi_* + u_*(x) \geq J_n(x, \varphi, \psi_*) + E_x^{\varphi\psi_*} [u_*(x_n)]. \quad (2.22)$$

Multiplicando ambos lados por $\frac{1}{n}$, calculando del límite inferior cuando $n \rightarrow \infty$, $\frac{u_*(x)}{n} \rightarrow 0$ y por el Lema 2.1.4 (d), tenemos que $\frac{E_x^{\varphi\psi_*} [u_*(x_n)]}{n} \rightarrow 0$, así

$$\begin{aligned} \xi_* &\geq \liminf_{n \rightarrow \infty} \frac{J_n(x, \varphi, \psi_*)}{n} = J(x, \varphi, \psi_*) \\ &\geq \liminf_{n \rightarrow \infty} \frac{J_n(x, \pi^1, \psi_*)}{n} = J(x, \pi^1, \psi_*) \\ &\geq \inf_{\pi^2 \in \Pi_2} J(x, \pi^1, \pi^2), \end{aligned}$$

para todo $\pi^1 \in \Pi_1$ y $x \in X$. Entonces por definición de $L(x)$, $\xi_* \geq L(x) = V^*$ para todo $x \in X$.

De manera similar, usando (2.7) en lugar de (2.6) tenemos que

$$\xi_* \leq J(x, \varphi_*, \pi^2) \leq \inf_{\pi^1 \in \Pi_1} J(x, \pi^1, \pi^2)$$

para cualesquiera $\pi^2 \in \Pi_2$ y $x \in X$, y por definición de $U(x)$, $\xi_* \leq U(x) = V^*$ para todo $x \in X$. Así $\xi_* = V^*$ y con esto concluimos que $(\Phi_1 \times \Pi_2)_{ca} \subset (\Phi_1 \times \Phi_2)_{eap}$.

Ahora, mostremos la otra contención, sea $(\varphi_*, \psi_*) \in \Pi_1 \times \Pi_2$ un par EAP óptimo, por Definición (2.1.1) se tiene que (2.1) y (2.2) se cumplen, y por la Proposición 2.1.1, tenemos que $V(x) = V^*$ para todo $x \in X$ y por (2.15) obtenemos que

$$V^* = J(x, \varphi_*, \psi_*) = j(\varphi_*, \psi_*), \quad (2.23)$$

para todo $x \in X$. Por el Lema 2.1.1 (b) y (c), existe una función $u_0 \in \mathbb{B}_w(X)$ definida por

$$u_0(x) := \sum_{t=0}^{\infty} E_x^{\varphi_* \psi_*} [r(x_t, \varphi_*(x_t), \psi_*(x_t)) - j(\varphi_*, \psi_*)],$$

además, $(j(\varphi_*, \psi_*), u_0)$ es la única solución de

$$\begin{aligned} j(\varphi_*, \psi_*) = V^* + u_0(x) &= r(x, \varphi_*(x), \psi_*(x)) \\ &+ \int_X u_0(y) Q(dy|x, \varphi_*(x), \psi_*(x)) \end{aligned} \quad (2.24)$$

para todo $x \in X$. De esto que

$$V^* + u_0(x) \geq \min_{\psi \in \mathbb{P}_B(x)} [r(x, \varphi_*(x), \psi) + \int_X u_0(y) Q(dy|x, \varphi_*(x), \psi)], \quad (2.25)$$

por el Lema 2.1.5 existe $\psi_0 \in \Phi_2$ tal que

$$\begin{aligned} \min_{\psi \in \mathbb{P}_B(x)} [r(x, \varphi_*(x), \psi) + \int_X u_0(y) Q(dy|x, \varphi_*(x), \psi)] &= \\ r(x, \varphi_*(x), \psi_0(x) + \int_X u_0(y) Q(dy|x, \varphi_*(x), \psi_0(x)) & \end{aligned} \quad (2.26)$$

para todo $x \in X$, entonces

$$V^* + u_0(x) \geq r(x, \varphi_*(x), \psi_0(x)) + \int_X u_0(y)Q(dy|x, \varphi_*(x), \psi_0(x)). \quad (2.27)$$

Repitiendo el procedimiento que se usó para obtener (2.20) pero ahora usando (2.27) en lugar de (2.5) obtenemos que

$$nV^* + u_0(x) \geq E_x^{\varphi_*\psi_0} \left[\sum_{t=0}^{n-1} r(x_t, \varphi_*(x_t), \psi_*(x_t)) \right] + E_x^{\varphi_*\psi_*} [u_0(x_n)].$$

Multiplicando ambos lados por $\frac{1}{n}$, calculando el límite inferior cuando $n \rightarrow \infty$ tenemos lo siguiente

$$\begin{aligned} V^* &\geq \liminf_{n \rightarrow \infty} \frac{E_x^{\varphi_*\psi_0} [\sum_{t=0}^{n-1} r(x_t, \varphi_*(x_t), \psi_0(x_t))]}{n} \\ &= \liminf_{n \rightarrow \infty} \frac{E_x^{\varphi_*\psi_0} [\sum_{t=0}^{n-1} r(x_t, a_t, bx_t)]}{n} \\ &\geq \liminf_{n \rightarrow \infty} \frac{E_x^{\varphi_*\pi^2} [\sum_{t=0}^{n-1} r(x_t, a_t, bx_t)]}{n} = \inf_{\pi^2 \in \Pi_2} J(x, \varphi_*, \pi^2). \end{aligned}$$

Esto porque se tomó a $\psi_* \in \Pi_2$ y después encontramos $\psi_0 \in \Phi_2$ tal que se cumple (2.26), así por (2.1)

$$V^* \geq J(x, \varphi_*, \psi_0) \geq \inf_{\pi^2 \in \Pi_2} J(x, \varphi_*, \pi^2) = V^*.$$

Por lo tanto,

$$J(x, \varphi_*, \psi_0) = j(\varphi_*, \psi_0) = V^* \quad (2.28)$$

para todo $x \in X$, y así la mejor respuesta para el jugador 2 a la estrategia óptima φ_* del jugador 1 es ψ_0 .

Ahora usando (2.28) en lugar de (2.23) y el mismo argumento que se utilizó para obtener (2.22), tenemos que existe $u_1 \in \mathbb{B}_w(X)$ tal que

$$V^* + u_1(x) = r(x, \varphi_*(x), \psi_0(x)) + \int_X u_1(y)Q(dy|x, \varphi_*(x), \psi_0(x)) \quad (2.29)$$

para todo $x \in X$. Restando (2.29) de (2.27) tenemos que

$$u_0(x) - u_1(x) \geq \int_X [u_0(y) - u_1(y)] Q(dy|x, \varphi_*(x), \psi_0(x)),$$

para todo $x \in X$, casi siempre, de modo que $u_0 - u_1$ es superarmónico con respecto a la probabilidad de transición de Markov $Q(\cdot|x, \varphi_*(x), \psi_0(x))$.

Por el Lema 2.1.2, existe una constante k_1 y un conjunto de Borel $D \subset X$, con $q(\varphi_*, \psi_0)(D) = 1$ tal que

$$\begin{aligned} k_1 &= u_0(x) - u_1(x), \quad \forall x \in D_1, \\ k_1 + u_1(x) &= u_0(x), \quad \forall x \in D_1. \end{aligned}$$

Definimos

$$v_0(\cdot) := u_0(\cdot) = u_1(\cdot) + k_1 \quad \text{en } D_1 \text{ y}$$

$$v_0(x) := \min_{\psi \in \mathbb{P}_B(x)} [r(x, \varphi_*(x), \psi) + \int_X u_0(y) Q(dy|x, \varphi_*(x), \psi)] - V^* \quad (2.30)$$

para $x \in D_1^C$. Entonces, como $q(\varphi_*, \psi_0)(D_1^C) = 0$, por el Lema 2.1.3 (b) tenemos que $Q(D_1^C|x, a, b) = 0$ para todo $(x, a, b) \in \mathbb{K}$, lo cual implica que en la integral de (2.30) podemos reemplazar u_0 por v_0 y así obtener

$$V^* + v_0(x) = \min_{\psi \in \mathbb{P}_B(x)} [r(x, \varphi_*(x), \psi) + \int_X v_0(y) Q(dy|x, \varphi_*(x), \psi)] \quad (2.31)$$

para todo $x \in X$. Retomando (2.22) y calculando el máximo, es decir,

$$\begin{aligned} V^* + u_0(x) &\leq \max_{\varphi \in \mathbb{P}_A(x)} [r(x, \varphi, \psi_*(x)) + \int_X u_0(y) Q(dy|x, \varphi, \psi_*(x))] \\ &= r(x, \varphi_0(x), \psi_*(x)) + \int_X u_0(y) Q(dy|x, \varphi_0(x), \psi_*(x)) \end{aligned} \quad (2.32)$$

para algún $\varphi_0 \in \Phi_1$. Repitiendo el procedimiento que se ha usado durante toda la prueba (iteración) ahora para esta última desigualdad, multiplicando por $\frac{1}{n}$ y calculando el límite

inferior cuando $n \rightarrow \infty$, tenemos que $V^* \leq J(x, \varphi_0, \psi_*) \leq \max_{\pi^1 \in \Pi_1} J(x, \pi_1, \psi_*) = V^*$, de aquí que

$$j(\varphi_0, \psi_*) = J(x, \varphi_0, \psi_*) = V^*, \quad \text{para todo } x \in X. \quad (2.33)$$

De nuevo, usamos (2.28) en lugar de (2.23) y el mismo argumento que se utilizó para hallar (2.22), tenemos que existe $u_2 \in \mathbb{B}_w(X)$ tal que

$$V^* + u_2(x) = r(x, \varphi_0(x), \psi_*(x)) + \int_X u_2(y) Q(dy|x, \varphi_0(x), \psi_*(x)) \quad (2.34)$$

para todo $x \in X$. Restando (2.34) de (2.32) obtenemos que

$$u_0(x) - u_2(x) \leq \int_X [u_0(y) - u_2(y)] Q(dy|x, \varphi_0(x), \psi_*(x)) \quad (2.35)$$

para todo $x \in X$. Por argumentos utilizados en la obtención de (2.29)-(2.31) existe un conjunto de Borel D_2 con $q(\varphi_0, \psi_*)(D_2) = 1$ y una función $v_1(\cdot) \in \mathbb{B}_w(X)$ tales que

$$v_1(\cdot) = u_0(\cdot) \quad \text{en } D_2 \quad (2.36)$$

y

$$V^* + v_1(x) = \max_{\varphi \in \mathbb{P}_A(x)} [r(x, \varphi, \psi_*(x)) + \int_X v_1(y) Q(dy|x, \varphi, \psi_*(x))] \quad (2.37)$$

para todo $x \in X$. Sea $D = D_1 \cap D_2$ y definamos $u_*(\cdot) := u_0(\cdot)$ en D y

$$u_*(\cdot) := Tu_0(\cdot) - V^* \quad \text{en } D^C. \quad (2.38)$$

Por el Lema 2.1.3 (a), $\gamma(D_1^C) = \gamma(D_2^C) = 0$, y así $\gamma(D) = 0$, que por la Suposición 2.1.4 se tiene que $Q(D^C|x, a, b) = 0$ para todo $(x, a, b) \in \mathbb{K}$.

Por lo tanto, en lugar de (2.38) podemos escribir

$$V^* + u_*(x) = Tu_*(x), \quad \forall x \in X,$$

que junto con (2.22), (2.29)-(2.31) y (2.36)-(2.37) nos lleva a que $(V^*, u_*, \varphi_*, \psi_*)$ es una cuádrupla canónica. ■

Equivalencia entre los conjuntos $(\Phi_1 \times \Phi_2)_{eap}$, $(\Phi_1 \times \Phi_2)_{ca}$ y con una condición de martingala.

Este segundo teorema afirma la equivalencia entre los criterios EAP y pares canónicos, que se refieren a pares de estrategias que cumplan con las ecuaciones de Shapley. Además, equivalentes a una condición de martingala.

Pero antes habrá que definir algunos conceptos necesarios para la condición de martingala.

Sea \mathcal{F}_n la σ -álgebra generada por (x_t, a_t, b_t) para $t = 0, \dots, n$, esto es

$$\mathcal{F}_n := \sigma\{x_0, a_0, b_0, \dots, x_n, a_n, b_n\}. \quad (2.39)$$

Además, sean J_n^0 como en (1.2) y ξ_* , u_* como en el Teorema 2.1, y luego definamos el proceso estocástico

$$M_n(h_\infty) := J_n^0(h_\infty) + u_*(x_n) - n\xi_* \quad \text{para } n = 1, 2, \dots, \quad (2.40)$$

con $M_0(h_\infty) = u_*(x_0)$. Finalmente, sea $\Delta : \mathbb{K} \rightarrow \mathbb{R}$ la llamada función de discrepancia dada por

$$\Delta(x, a, b) := r(x, a, b) + \int_X u_*(y)Q(dy|x, a, b) - u_*(x) - \xi_*. \quad (2.41)$$

Teorema 2.1.2. Bajo las hipótesis del Teorema 2.1.1, las siguientes expresiones son equivalentes:

- a) El par $(\varphi_*, \psi_*) \in \Phi_1 \times \Phi_2$ es EAP óptimo.
- b) Para cada $x \in X$

$$\begin{aligned} \Delta(x, \varphi_*(x), \psi_*(x)) &= \max_{\varphi \in \mathbb{P}_A(x)} \Delta(x, \varphi, \psi_*(x)) \\ &= \min_{\psi \in \mathbb{P}_B(x)} \Delta(x, \varphi_*(x), \psi) = 0. \end{aligned} \quad (2.42)$$

- c) Para cada $x \in X$, $\pi^1 \in \Pi_1$, y $\pi^2 \in \Pi_2$

- c1) $\{M_n(h_\infty), \mathcal{F}_n\}$ es una $P_x^{\varphi_*, \psi_*}$ - martingala,
- c2) $\{M_n(h_\infty), \mathcal{F}_n\}$ es una $P_x^{\varphi_*, \pi^2}$ - submartingala, y
- c3) $\{M_n(h_\infty), \mathcal{F}_n\}$ es una $P_x^{\pi^1, \psi_*}$ - supermartingala.

Demostración. Notemos que (2.42) es otra forma de escribir (2.5)-(2.7). Así, por el Teorema 2.1.1 tenemos la equivalencia entre a) y b). Para probar que b) implica c), sea $x \in X$ un estado arbitrario y $\pi^i \in \Pi_i$ también arbitraria. Sea $\mathcal{F}_n := \sigma\{x_0, a_0, b_0, \dots, x_n, a_n, b_n\}$. Notemos ahora que

$$E_x^{\pi^1 \pi^2}[u_*(x_{n+1})|\mathcal{F}_n] = \int_X u_*(y)Q(dy|x_n, a_n, b_n), \quad (2.43)$$

de (2.41) se tiene que

$$\begin{aligned} E_x^{\pi^1 \pi^2}[\Delta(x_n, a_n, b_n)|\mathcal{F}_n] &= E_x^{\pi^1 \pi^2}[r(x_n, a_n, b_n)] \\ &\quad + E_x^{\pi^1 \pi^2}[E_x^{\pi^1 \pi^2}[u_*(x_{n+1})|\mathcal{F}_n]|\mathcal{F}_n] \\ &\quad - E_x^{\pi^1 \pi^2}[u_*(x_n)|\mathcal{F}_n] - E_x^{\pi^1 \pi^2}[\xi_*|\mathcal{F}_n] \\ &= r(x_n, a_n, b_n) \\ &\quad + E_x^{\pi^1 \pi^2}[u_*(x_{n+1})|\mathcal{F}_n] - u_*(x_n) - \xi_*. \end{aligned} \quad (2.44)$$

Consideremos ahora el proceso estocástico dado en (2.40). Del Lema 2.1.4 (b), (c), se deduce que este proceso es integrable con respecto a $P_x^{\pi^1 \pi^2}$, es decir,

$$E_x^{\pi^1 \pi^2}|M_n(h_\infty)| < \infty, \quad \text{para cada } n = 0, 1, \dots$$

Además,

$$\begin{aligned} M_{n+1}(h_\infty) &= \sum_{t=0}^n r(x_t, a_t, b_t) + u_*(x_{n+1}) - (n+1)\xi_* \\ &= \sum_{t=0}^{n-1} r(x_t, a_t, b_t) + r(x_n, a_n, b_n) + u_*(x_{n+1}) + u_*(x_n) \\ &\quad - u_*(x_n) - n\xi_* - \xi_* \\ &= M_n(h_\infty) + r(x_n, a_n, b_n) + u_*(x_{n+1}) - u_*(x_n) + \xi_*, \end{aligned}$$

así, por (2.44) se tiene que

$$\begin{aligned} E_x^{\pi^1 \pi^2}[M_{n+1}(h_\infty)|\mathcal{F}_n] &= E_x^{\pi^1 \pi^2}[M_n(h_\infty) + r(x_n, a_n, b_n) \\ &\quad + u_*(x_{n+1}) - u_*(x_n) + \xi_*] \\ &= M_n(h_\infty) + E_x^{\pi^1 \pi^2}[\Delta(x_n, a_n, b_n)|\mathcal{F}_n]. \end{aligned} \quad (2.45)$$

De aquí que, (2.42) implica (c), pues si

$$\Delta(x, \varphi_*(x), \psi_*(x)) = 0, \quad \forall x \in X, \quad (2.46)$$

tendríamos por (2.45) que

$$E_x^{\varphi_* \psi_*}[M_{n+1}(h_\infty)|\mathcal{F}_n] = M_n(h_\infty), \quad \forall n = 0, 1, \dots,$$

y por lo tanto, $\{M_n(h_\infty)|\mathcal{F}_n\}$ es una $P_x^{\varphi_* \psi_*}$ -martingala. De manera análoga, si

$$\min_{\psi \in \mathbb{P}_B(x)} \Delta(x, \varphi_*(x), \psi) = 0, \quad \forall x \in X, \quad (2.47)$$

entonces,

$$\Delta(x, \varphi_*(x), \psi) \geq 0, \quad \forall x \in X, \psi \in \mathbb{P}_B(x), \quad (2.48)$$

de (2.45) tenemos que

$$E_x^{\varphi_* \pi^2}[M_{n+1}(h_\infty)|\mathcal{F}_n] \geq M_n(h_\infty), \quad \forall n = 0, 1, \dots,$$

por lo tanto, $\{M_n(h_\infty)|\mathcal{F}_n\}$ es una $P_x^{\varphi_* \pi^2}$ -submartingala. Si

$$\max_{\varphi \in \mathbb{P}_A(x)} \Delta(x, \varphi, \psi_*(x)) = 0, \quad (2.49)$$

entonces,

$$\Delta(x, \varphi, \psi_*(x)) \leq 0 \quad \forall x \in X, \varphi \in \mathbb{P}_A(x). \quad (2.50)$$

Así, de (2.45) tenemos que

$$E_x^{\pi^1, \psi_*}[M_{n+1}(h_\infty)|\mathcal{F}_n] \leq M_n(h_\infty) \quad \forall n = 0, 1, \dots,$$

por lo tanto $\{M_n(h_\infty)|\mathcal{F}_n\}$ es una $P_x^{\pi^1\psi^*}$ -supermartingala. Y así, se tiene que b) implica c).

Para la implicación opuesta, (c) implica (b), notemos que si calculamos esperanzas en ambos lados de (2.45) tenemos que

$$E_x^{\pi^1\pi^2}[M_{n+1}(h_\infty)] = E_x^{\pi^1\pi^2}[M_n(h_\infty)] + E_x^{\pi^1\pi^2}[\Delta(x_n, a_n, b_n)]. \quad (2.51)$$

Por lo tanto, si se cumple (c1), $E_x^{\pi^1\pi^2}[\Delta(x_n, a_n, b_n)] = 0$ entonces $E_x^{\varphi_*, \psi^*}[\Delta(x_n, a_n, b_n)] = 0$, y para $n = 0$ tenemos que (2.46) se cumple. De manera análoga, si se cumple (c2), de (2.51) obtenemos que

$$E_x^{\varphi_*\pi^2} \Delta(x_n, \varphi_*(x_n), \pi_n^2(h_n)) \geq 0$$

para todo $x \in X$, $h_n \in H_n$, y $n = 0, 1, \dots$. En particular, para $n = 0$

$$\Delta(x, \varphi_*(x), \pi_0^2(x)) \geq 0 \quad \forall x \in X,$$

y así, como $\pi^2 \in \Pi_2$ es arbitraria, tenemos (2.47) y (2.48). Finalmente, de (2.51) y (c3) obtenemos (2.49) y (2.50). ■

Definimos $(\Phi_1 \times \Phi_2)_{spap}$ como el conjunto de los pares de estrategias bajo el criterio SPAP.

Equivalencia de los conjuntos $(\Phi_1 \times \Phi_2)_{eap}$, $(\Phi_1 \times \Phi_2)_{spap}$ y $(\Phi_1 \times \Phi_2)_{ca}$.

Este último teorema muestra la equivalencia entre los conjuntos de estrategias de pares canónicos, pares bajo el criterio EAP y con el criterio SPAP.

Teorema 2.1.3. Supongamos que las hipótesis del Teorema 2.1.1 se satisfacen y, además, existe una constante $\hat{r} \geq 0$ tal que

$$r^2(x, a, b) \leq \hat{r}w(x), \quad \forall (x, a, b) \in \mathbb{K}. \quad (2.52)$$

Entonces un par de estrategias en $\Phi_1 \times \Phi_2$ es EAP óptimo si y solo si es SPAP óptimo; entonces, por el Teorema 2.1.1

$$(\Phi_1 \times \Phi_2)_{eap} = (\Phi_1 \times \Phi_2)_{ca} = (\Phi_1 \times \Phi_2)_{spap}.$$

Demostración. Usaremos la siguiente notación para el criterio SPAP,

$$J_n^0(\Delta; h_\infty) := \sum_{t=0}^{n-1} \Delta(x_t, a_t, b_t), \quad (2.53)$$

$$J^0(\Delta; h_\infty) := \liminf_{n \rightarrow \infty} \frac{J_n^0(\Delta; h_\infty)}{n},$$

donde $x_0 = x \in X$ y $\pi^i \in \Pi_i$ ($i = 1, 2$) son arbitrarios. Además, con \mathcal{F}_n como en (2.39), consideremos el siguiente proceso estocástico

$$\begin{aligned} Y_t(\pi^1, \pi^2) &:= u_*(x_t) - E_x^{\pi^1 \pi^2} [u_*(x_t) | \mathcal{F}_{t-1}] \\ &= u_*(x_t) - \int_X u_*(y) Q(dy | x_{t-1}, a_{t-1}, b_{t-1}) \end{aligned} \quad (2.54)$$

para $t = 1, 2, \dots$, y

$$S_n(\pi^1, \pi^2) := \sum_{t=1}^n Y_t(\pi^1, \pi^2). \quad (2.55)$$

De (2.43), (2.44) y (2.53), también podemos escribir este último proceso como

$$\begin{aligned} S_n(\pi^1, \pi^2) &= \sum_{t=1}^n u_* \left[(x_t) - \int_X u_*(y) Q(dy | x_{t-1}, a_{t-1}, b_{t-1}) \right] \\ &= \sum_{t=0}^{n-1} [-u_*(x_0) + u_*(x_t) + u_*(x_n)] \\ &\quad - \sum_{t=1}^n \left[\int_X u_*(y) Q(dy | x_{t-1}, a_{t-1}, b_{t-1}) \right] \\ &= u_*(x_n) - u_*(x_0) \end{aligned}$$

$$\begin{aligned}
 & - \sum_{t=0}^{n-1} \left[\int_X u_*(y) Q(dy | x_t, a_t, b_t) - u_*(x_t) \right] \\
 & = u_*(x_n) - u_*(x_0) \\
 & - \sum_{t=0}^{n-1} \left[r(x_t, a_t, b_t) + \int_X u_*(y) Q(dy | x_t, a_t, b_t) \right. \\
 & \left. - u_*(x_t) - \xi_* \right] + \sum_{t=0}^{n-1} r(x_t, a_t, b_t) - n\xi_* \\
 & = u_*(x_n) - u_*(x_0) - \sum_{t=0}^{n-1} \Delta(x_t, a_t, b_t) \\
 & + \sum_{t=0}^{n-1} r(x_t, a_t, b_t) - n\xi_*.
 \end{aligned}$$

Y así,

$$S_n(\pi^1, \pi^2) = u_*(x_t) - u_*(x_0) - J_n^0(\Delta; h_\infty) + J_n^0(h_\infty) - n\xi_*. \quad (2.56)$$

Por otro lado, de (2.54) y (2.55) tenemos que

$$\begin{aligned}
 E_x^{\pi^1 \pi^2} [S_{n+1} | \mathcal{F}_n] & = E_x^{\pi^1 \pi^2} \left[\sum_{t=1}^{n+1} Y_t | \mathcal{F}_n \right] \\
 & = \sum_{t=1}^n Y_t + E_x^{\pi^1 \pi^2} [Y_{n+1} | \mathcal{F}_n] \\
 & = S_n + E_x^{\pi^1 \pi^2} [u_*(x_{n+1} \\
 & - \int_X u_*(y) Q(dy | x_n, a_n, b_n)) | \mathcal{F}_n] \\
 & = S_n + E_x^{\pi^1 \pi^2} [u_*(x_{n+1})] - E_x^{\pi^1 \pi^2} [u_*(x_{n+1})] \\
 & = S_n.
 \end{aligned}$$

Tenemos entonces que $\{S_n(\pi^1, \pi^2)\}$ es una $P_x^{\pi^1, \pi^2}$ -martingala con respecto a (\mathcal{F}_n) .

Además, del Lema 11.3.11 en [5], obtenemos que

$$\lim_{n \rightarrow \infty} \frac{S_n(\pi^1, \pi^2)}{n} = 0 \quad \text{y} \quad \lim_{n \rightarrow \infty} \frac{u_*(x_n)}{n} = 0.$$

Por lo tanto, multiplicando ambos lados de (2.56) por $\frac{1}{n}$ y considerando el límite cuando $n \rightarrow \infty$, obtenemos que

$$0 = - \lim_{n \rightarrow \infty} \frac{J_n^0(\Delta; h_\infty)}{n} + \lim_{n \rightarrow \infty} \frac{J_n(h_\infty)}{n} - \xi_*$$

$$J^0(h_\infty) = \xi_* + J^0(n, h_\infty), \quad P_x^{\pi^1 \pi^2} - c.s. \quad (2.57)$$

Enseguida, usamos (2.57) para probar la parte $(\Phi_1 \times \Phi_2)_{eap} \subset (\Phi_1 \times \Phi_2)_{spap}$ del Teorema 2.1.3.

Supongamos que $(\varphi_*, \psi_*) \in \Phi_1 \times \Phi_2$ es EAP óptimo. Luego escribimos (2.42) como en (2.46), (2.47), (2.48), (2.49) y (2.50).

De (2.46), (2.57) y (2.14) tenemos (2.3) con $V(\cdot) = \xi_*$, es decir,

$$J^0(h_\infty) = \xi_*, \quad P_x^{\varphi_* \psi_*} - c.s. \quad (2.58)$$

Análogamente, de (2.57) y (2.48)

$$J^0(h_\infty) \geq \xi_*, \quad P_x^{\varphi_* \pi^2} - c.s.,$$

y con (2.50) en lugar de (2.48) tenemos que

$$J^0(h_\infty) \leq \xi_*, \quad P_x^{\pi^1 \psi_*} - c.s. \quad (2.59)$$

Por lo tanto, como $x \in X$, $\pi^1 \in \Pi_1$ y $\pi^2 \in \Pi_2$ fueron arbitrarios, concluimos (de la Definición 2.1.2) que (φ_*, ψ_*) es SPAP óptimo con la ganancia óptima SPAP $V(\cdot) = \xi_*$.

Ahora, supongamos que (φ_*, ψ_*) es SPAP óptimo. Entonces por (2.58), (2.14) y (2.15) vemos que (φ_*, ψ_*) satisface que

$$J(x, \varphi_*, \psi_*) = \xi_*, \quad \forall x \in X. \quad (2.60)$$

Como (2.59) se cumple para todo $\pi^1 \in \Pi_1 \supset \Phi_1$, utilizando nuevamente (2.14) y (2.15) hallamos que

$$J(x, \varphi, \psi_*) = j(\varphi, \psi_*) \leq \xi_*, \quad \forall \varphi \in \Phi_1, x \in X; \quad (2.61)$$

por lo tanto, si fijamos $\psi_* \in \Phi_2$, entonces (2.60), (2.61) y el Corolario 2.1.1 (a) implica que

$$\sup_{\pi^1 \in \Pi_1} J(x, \pi^1, \psi_*) = \xi_*, \quad \forall x \in X. \quad (2.62)$$

De manera análoga, pero ahora con (2.59) en lugar de (2.61) tenemos que

$$\inf_{\pi^2 \in \Pi_2} J(x, \varphi_*, \pi^2) = \xi_*, \quad \forall x \in X. \quad (2.63)$$

Así, por (2.62) y (2.63), el par (φ_*, ψ_*) es EAP óptimo. ■

El siguiente corolario es resultado de la Proposición 2.1.1 y los teoremas anteriores.

Corolario 2.1.1. Considere el MCP, proceso de control de Markov, asociado a $MCM = (X, A, \mathbb{K}_A, Q, r)$ y suponga que se satisfacen las suposiciones 2.1.1, 2.1.2 y 2.1.3. Entonces tenemos lo siguiente:

- a) Existe una estrategia óptima de recompensa media esperada (EAR) φ^* en Φ_1 , el conjunto de estrategias de control estacionarias, y el valor $\hat{V}(\cdot) = \sup_{\pi \in \Pi} \hat{J}(\cdot, \pi)$ para todo $x \in X$ satisface que

$$\hat{J}(x, \varphi^*) = \hat{V}(x) = \sup_{\varphi \in \Phi_1} \hat{J}(x, \varphi) =: \xi^*, \quad \forall x \in X.$$

- b) Y existe una tripla canónica (ξ^*, h^*, φ^*) con $\xi^* \in \mathbb{R}$, $h^* \in \mathbb{B}_w$, y $\varphi^* \in \Phi_1$, es decir, para todo $x \in X$,

$$\begin{aligned} \xi^* + h^*(x) &= \hat{r}(x, \varphi^*(x)) + \int_X h^*(y) \hat{Q}(dy|x, \varphi^*(x)) \\ &= \max_{\varphi \in \mathbb{P}_A(x)} [\hat{r}(x, \varphi) + \int_X h^*(y) \hat{Q}(dy|x, \varphi)]. \end{aligned}$$

En este caso se dice que φ^* es una estrategia canónica. Además, si reemplazamos la Suposición 2.1.3 por la Suposición 2.1.4, entonces las siguientes afirmaciones c) a f) son equivalentes:

- c) $\varphi^* \in \Phi_1$ es EAR óptimo.
 d) $\varphi^* \in \Phi_1$ es una estrategia canónica.
 e) Para cada $x \in X$

$$\hat{\Delta}(x, \varphi^*(x)) = \max_{\varphi \in \mathbb{P}_A(x)} \hat{\Delta}(x, \varphi) = 0,$$

donde $\hat{\Delta} : \mathbb{K}_A \rightarrow \mathbb{R}$ es la (recompensa promedio) función de discrepancia:

$$\hat{\Delta}(x, a) := \hat{r}(x, a) + \int_X h^*(y) \hat{Q}(dy|x, a) - h^*(x) - \xi^*.$$

- f) Para cada $\pi \in \Pi_C$ y $x \in X$, el proceso estocástico (2.40)

$$\hat{M}_n(\hat{h})_\infty := \sum_{t=0}^{n-1} \hat{r}(x_t, a_t) + h^*(x_n) - n\xi^*,$$

para $n = 1, 2, \dots$, con $\hat{M}_0(\hat{h}_\infty) := h^*(x_0)$ y $\hat{h}_\infty = (x_0, a_0, x_1, a_1, \dots)$, es una P_x^π -supermartingala (con respecto a la σ -álgebra generada por $\{x_0, a_0, \dots, x_n, a_n\}$), mientras que $\{\hat{M}_n(\hat{h}_\infty), n = 0, 1, \dots\}$, es una $P_x^{\varphi^*}$ -martingala. Finalmente, si además $\hat{r}(x, a)$ satisface (2.52), entonces cada una de las expresiones c) a f) son equivalentes a lo siguiente:

- g) $\varphi^* \in \Phi$ es la recompensa promedio del camino de la muestra (SPAP) óptima, es decir, (con ξ^* como en a) y los cambios obvios en (1.4)),

$$J^0(\hat{h}_\infty) = \xi^*, \quad P_x^{\varphi^*} \text{ c.s.}, \quad \forall x \in X$$

y

$$J^0(\hat{h}_\infty) \leq \xi^*, \quad P_x^\pi \text{ c.s.}, \quad \forall \pi \in \Pi_C, x \in X.$$

2.2. El juego del volado

El siguiente ejemplo es para ilustrar el caso de los juegos de suma cero y se trabaja con el criterio EAP.

Este juego consiste en lo siguiente: dos amigos quieren decidir a quien le toca ir a comprar el refresco. El primero saca una moneda y la coloca sobre la mesa cubriéndola con la mano, le pide a su compañero que adivine cuál es la posición en que la puso, ¿el águila hacia arriba o el sol? Si es descubierto, tendrá que ir él mismo a la compra, pero si no, le tocará al segundo amigo. La forma normal del juego se muestra en la Tabla 2.1.

Tabla 2.1: Forma normal juego del volado.

		J_2	
		A	S
J_1	A	-1,1	1,-1
	S	1,-1	-1,1

Donde A representa que la moneda se encuentra con águila hacia arriba y S con sol hacia arriba.

La solución de este juego, estático, es el par de estrategias mixtas (π^{1*}, π^{2*}) tal que π^{i*} ($i = 1, 2$) esta definida como: elegir A con probabilidad $\frac{1}{2}$ y elegir S con probabilidad $\frac{1}{2}$.

Luego iteramos el juego, esto representa que los jugadores irán por un refresco en cada etapa del juego, es decir, si el juego consta de n etapas, entonces entre ambos jugadores tendrán n refrescos, y calculamos las ganancias bajo el criterio EAP.

Definimos a $X = \{(-1, 1), (1, -1)\}$ como el conjunto de estados, $A = B = \{0, 1\}$ como el conjunto de acciones para el jugador 1 y el jugador 2, respectivamente, donde 0 representa

que se obtuvo águila y 1 que se obtuvo sol.

El kernel de transición, $Q \in \mathbb{P}(X|\mathbb{K})$ con $\mathbb{K} = \{(x, a, b) | x \in X, a \in A(x), b \in B(x)\}$, se define como:

$$Q(\{(-1, 1)\} | x_t, a_t, b_t) = \begin{cases} 1, & \text{si } a_t = b_t \\ 0, & \text{c.o.c.} \end{cases}$$

$$Q(\{(1, -1)\} | x_t, a_t, b_t) = \begin{cases} 1, & \text{si } a_t \neq b_t \\ 0, & \text{c.o.c.} \end{cases}$$

Las funciones de recompensa quedan definidas como en el caso estático.

Las estrategias que usamos en este ejemplo son las siguientes:

- S_a , elegir siempre águila;
- S_s , elegir siempre sol;
- S_p , elegir águila con probabilidad p y sol con probabilidad $1 - p$ donde $p \in (0, 1)$.

Usando el criterio EAP obtenemos las “ganancias” que se muestran en la Tabla 2.3.

Tabla 2.2: Ganancias usando el criterio EAP para el juego del volado.

		J_2		
		S_a	S_s	S_p
J_1	S_a	-1,1	1,-1	$1 - 2p, -1 + 2p$
	S_s	1,-1	-1,1	$-1 + 2p, 1 - 2p$
	S_p	$1 - 2p, -1 + 2p$	$-1 + 2p, 1 - 2p$	$-4p^2 + 4p - 1, 4p^2 - 4p + 1$

Ahora hay que hallar las estrategias óptimas. Si el jugador 1 elige S_a , el jugador 2 puede elegir entre todas sus opciones, pero ¿qué es lo que más le conviene? Enseguida la Figura 2.1 muestra de las ganancias para el jugador 2.

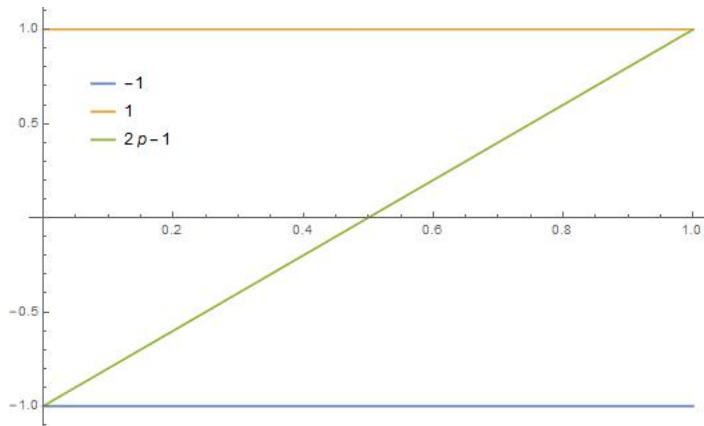


Figura 2.1: Gráfica de S_a , S_s y S_p cuando el jugador 1 elige S_a .

Con esto podemos ver que la mejor respuesta es S_a .

Si el jugador 1 elige S_s , las opciones para el jugador 2 se muestran en la Figura 2.2.

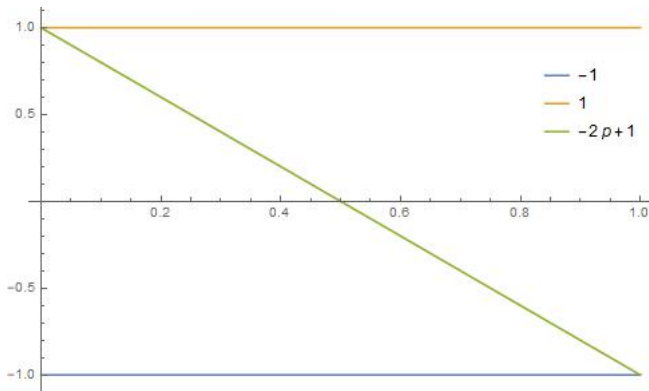


Figura 2.2: Gráfica de S_a , S_s y S_p cuando el jugador 1 elige S_s .

De aquí que la mejor respuesta para el jugador 2 es S_s .

Si el jugador 1 elige S_p , las ganancias del jugador 2 se muestran en la Figura 2.3.

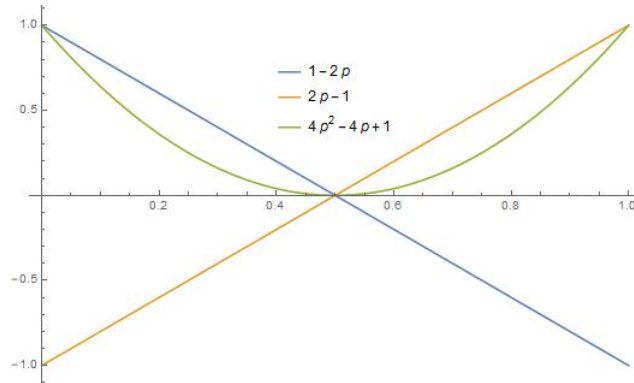


Figura 2.3: Gráfica de S_a , S_s y S_p cuando el jugador 1 elige S_p .

La mejor respuesta para el jugador 2 es S_s si $p < \frac{1}{2}$, S_a si $p > \frac{1}{2}$ y cualquier estrategia cuando $p = \frac{1}{2}$.

Si el jugador 2 elige S_p , las ganancias del jugador 1 se muestran en la Figura 2.4.

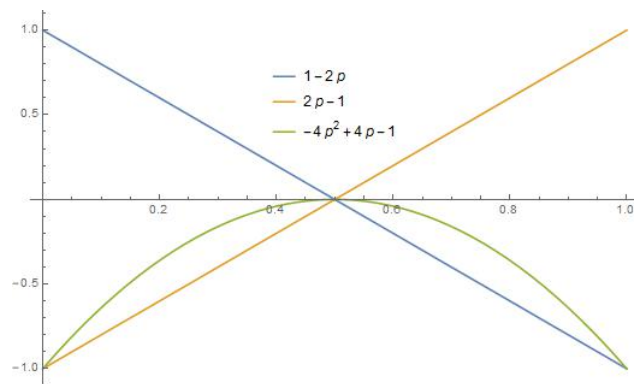


Figura 2.4: Gráfica de S_a , S_s y S_p cuando el jugador 2 elige S_p .

Y la mejor respuesta para el jugador 2 es S_a cuando $p < \frac{1}{2}$, S_s cuando $p > \frac{1}{2}$, mientras que para $p = \frac{1}{2}$ puede elegir cualquiera de las estrategias.

Tabla 2.3: Mejores respuestas usando el criterio EAP para el juego del volado.

		J_2		
		S_a	S_s	S_p
J_1	S_a	$-1, \underline{1}$	$\underline{1}, -1$	$\frac{1-2p}{-1+2p}, -1+2p$
	S_s	$\underline{1}, -1$	$-1, \underline{1}$	$\frac{-1+2p}{-4p^2+4p-1}, 1-2p$
	S_p	$1-2p, \underline{-1+2p}$	$-1+2p, \underline{1-2p}$	$\underline{-4p^2+4p-1}, \underline{4p^2-4p+1}$

Con el análisis anterior podemos observar que tenemos solo un par de estrategias óptimas EAP, (S_p, S_p) para $p = \frac{1}{2}$.

Un sistema de colas controlado a tiempo discreto.

Considere un juego estocástico en el cual el estado variable denota el número de clientes en el sistema. El jugador 1 controla la tasa de llegada $a \in [\lambda_1, \lambda_2]$ con $\lambda_2 > \lambda_1 > 0$, e independientemente el jugador 2 controla la tasa de servicios $b \in [\mu_1, \mu_2]$ con $\mu_2 > \mu_1 > 0$. Cuando el estado es $i \in S = \{0, 1, \dots\}$, y la acción elegida es (a, b) , el jugador 1 recibe una recompensa inmediata $r(i, a, b)$ (o bien el jugador 2 paga un costo $r(i, a, b)$). En otras palabras, se considera el control a tiempo continuo del proceso de salto de Markov con espacio de estado S , donde la tasa de transición del estado i al $i + 1$ es controlada por el jugador 1 y la tasa de transición del estado i al $i - 1$ (para $i > 0$) es controlada por el jugador 2.

En lugar de analizar directamente el modelo a tiempo continuo, se estudia el modelo a tiempo discreto correspondiente. Ya que las tasas de transición en este modelo son acotadas, se puede mostrar que la estrategia promedio estacionaria para el modelo a tiempo discreto es además una estrategia estacionaria óptima para el problema original a tiempo continuo usando la técnica estándar de uniformización y más aún, la probabilidad de transición para el modelo a tiempo discreto queda

como sigue

$$q(0|0, a, b) = \frac{\mu_2}{\mu_2 + \lambda_2}, \quad q(1|0, a, b) = \frac{\lambda_2}{\mu_2 + \lambda_2}, \quad (2.64)$$

y para cada $i \geq 1$, $a \in [\lambda_1, \lambda_2]$, $b \in [\mu_1, \mu_2]$,

$$q(j|i, a, b) = \begin{cases} \frac{b}{\lambda_2 + \mu_2} & \text{si } j = i - 1, \\ 1 - \frac{a+b}{\lambda_2 + \mu_2} & \text{si } j = i, \\ \frac{a}{\lambda_2 + \mu_2} & \text{si } j = i + 1, \\ 0 & \text{c.o.c.} \end{cases} \quad (2.65)$$

usando la renormalización adecuada.

El objetivo es encontrar condiciones bajo las cuales exista un par EAP óptimo de estrategias estacionarias. Para esto, se usan las siguientes suposiciones:

- i)* $e\lambda_2 < \mu_1$, donde e es la constante conocida.
- ii)* Para cada $i \in S$, las funciones $r(i, a, b)$ son continuas en $A(i) \times B(i)$, y $|r(i, a, b)| < Le^i$ para todo $i \in S$ y alguna constante $L > 0$.

Con estas suposiciones se obtiene

Proposición 2.2.1. Bajo las suposiciones *i)* e *ii)*, el sistema de colas controlado a tiempo discreto satisface las condiciones del Teorema 2.1.1, por lo que existe un par EAP óptimo de estrategias estacionarias.

La prueba de este resultado puede revisarse en [13].

Capítulo 3

Juegos de suma no cero

En este capítulo trabajamos el caso de los juegos de suma no cero, para ello usaremos el criterio EAP. En la primera sección se presenta la definición de equilibrio de Nash y un teorema que nos habla de la existencia de equilibrios en los juegos de suma no cero.

La segunda sección muestra un ejemplo conocido en la Teoría de juegos, el Dilema del Prisionero Iterado, que hemos adaptado a un juego estocástico y hallamos los equilibrios de Nash con un conjunto de estrategias que hemos proporcionado, dentro de ellas la famosa *tit-for-tat* o conocida también como *toma y daca*.

3.1. Juegos de suma no cero

En esta sección usaremos el criterio EAP y se darán condiciones para poder hallar equilibrios de Nash de un juego estocástico.

Definición 3.1.1. Un par de estrategias (π^{1*}, π^{2*}) se llama un equilibrio de Nash (para el criterio EAP) si

$$J^1(x, \pi^{1*}, \pi^{2*}) \geq J^1(x, \pi^1, \pi^{2*}), \quad \forall \pi^1 \in \Pi_1, x \in X,$$

y

$$J^2(x, \pi^{1*}, \pi^{2*}) \geq J^2(x, \pi^{1*}, \pi^2), \quad \forall \pi^2 \in \Pi_2, x \in X.$$

En este caso, (2.4) queda expresada de la siguiente forma

$$r_i(x, \varphi, \psi) := \int_{A(x)} \int_{B(x)} r_i(x, a, b) \psi(db) \varphi(da),$$

con $x \in X$, $\varphi \in \mathbb{P}_A(x)$ y $\psi \in \mathbb{P}_B(x)$.

Además de las suposiciones de la sección anterior necesitaremos otras más y haremos un cambio en la Suposición 2.1.1 (b). Para cada $(x, a, b) \in \mathbb{K}$, $r_1(x, a, \cdot)$ es u.s.c. sobre $A(x)$ y $r_2(x, \cdot, b)$ es u.s.c. sobre $B(x)$.

Suposición 3.1.1. La densidad de transición $g(x, a, b, y)$, como en (2.1.4) y λ también, es tal que

$$\lim_{n \rightarrow \infty} \int_X |g(x, a^n, b^n, y) - g(x, a, b, y)| w(y) \gamma(dy) = 0, \quad \forall x \in X$$

si $a^n \rightarrow a$ en $A(x)$ y $b^n \rightarrow b$ en $B(x)$, donde $w(\cdot)$ es la función de la Suposición 2.1.1 (d).

Las suposiciones siguientes necesitan que (2.4) tenga una estructura con recompensa aditiva y ley de transición aditiva.

Suposición 3.1.2. Existen kérneos subestocásticos $Q_1 \in \mathbb{P}(X|\mathbb{K}_A)$ y $Q_2 \in \mathbb{P}(X|\mathbb{K}_B)$ tales que

$$Q(\cdot|x, a, b) = Q_1(\cdot|x, a) + Q_2(\cdot|x, b)$$

cualesquiera $x \in X$, $a \in A(x)$, $b \in B(x)$. Más aún, $Q_1(D|x, \cdot)$ y $Q_2(D|x, \cdot)$ son continuas sobre $A(x)$ y $B(x)$ respectivamente, para cada $D \in \mathcal{B}(X)$.

Suposición 3.1.3. Para $i = 1, 2$ existen funciones medibles $r_{i1} : \mathbb{K}_A \rightarrow \mathbb{R}$ y $r_{i2} : \mathbb{K}_B \rightarrow \mathbb{R}$, tales que

- (a) $r_i(x, a, b) = r_{i1}(x, a) + r_{i2}(x, b)$ para todo $x \in X$, $a \in A$, $b \in B$.
Más aún, para cada $x \in X$,
- (b) las funciones $r_{i1}(x, \cdot)$ y $r_{i2}(x, \cdot)$ son continuas sobre $A(x)$ y $B(x)$, respectivamente, y
- (c) $\max_{a \in A(x)} |r_{i1}(x, a)| \leq w(x)$, y $\max_{b \in B(x)} |r_{i2}(x, b)| \leq w(x)$.

Note que (c) y la condición $\gamma \in (\mathbb{M})_w(X)$ en la Suposición 2.1.4 producen que

$$\int_X \max_{a \in A(x)} |r_{i1}(x, a)| \gamma(dx) < \infty, \int_X \max_{b \in B(x)} |r_{i2}(x, b)| \gamma(dx) < \infty.$$

La importancia del siguiente teorema esta en que existen juegos que no tienen equilibrios de Nash, así, el teorema nos brinda las condiciones necesarias para asegurar que nuestro juego tiene al menos un equilibrio de Nash.

La prueba aquí expuesta para este teorema es la misma que en [1]. Y para esta necesitamos los siguientes lemas y teorema.

Lema 3.1.1. Supongamos que se cumplen las Suposiciones 2.1.1, 2.1.2 y 2.1.3. Entonces para cada $\psi \in \Phi_2$ fijo, existe una estrategia estacionaria $\varphi_* \in \Phi_2$, que es la recompensa media esperada óptima, para el modelo de control de Makov en $MCM_1(\psi) = (X, A, \mathbb{K}_A, Q_\psi, r_{1,\psi})$, es decir,

$$J^1(x, \varphi_*, \psi) = \max_{\pi^1 \in \Pi_1} J^1(\pi^1, \psi) := \rho_1^*(\psi) \quad (3.1)$$

para todo $x \in X$. Más aún, existe una función $h_{\varphi^*, \psi}^1 \in \mathbb{B}_w(X)$ tal que $(\rho_1^*(\psi), h_{\varphi^*, \psi}^1)$ es la única solución en $\mathbb{R} \times \mathbb{B}_w(X)$ de la ecuación

$$\rho_1^*(\psi) + h_{\varphi^*, \psi}^1(x) = r_1(x, \varphi^*(x), \psi(x)) + \int_X h_{\varphi^*, \psi}^1(y) Q(dy|x, \varphi^*(x), \psi(x)) \quad (3.2)$$

$$\begin{aligned}
 &= \max_{\varphi \in \mathbb{P}_A(x)} [r_1(x, \varphi, \psi(x)) + \\
 &\quad \int_X h_{\varphi^*, \psi}^1(y) Q(dy | x, \varphi, \psi(x))] \tag{3.3}
 \end{aligned}$$

para todo $x \in X$, y tal que $\int_X h_{\varphi^*, \psi}^1(y) q(\varphi^*, \psi)(dy) = 0$, con $q(\varphi^*, \psi)$ como en (2.12).

Teorema 3.1.1 (Punto fijo de Fan [4]). Sea L un espacio lineal topológico convexo local y K un conjunto convexo y compacto en L . Sea $\mathcal{R}(X)$ la familia de todos los subconjuntos convexos (no vacíos) de K . Entonces para cualquier transformación f u.s.c. que va de un punto en K a un conjunto en $\mathcal{R}(X)$, existe un punto $x_0 \in K$ tal que $x_0 \in f(x_0)$.

Lema 3.1.2. Supongamos que se cumplen las Suposiciones 2.1.1, 2.1.2 y 2.1.3, y sea $(\varphi, \psi) \in \Phi_1 \times \Phi_2$ un par arbitrario de estrategias estacionarias. Entonces para $(i = 1, 2)$ tenemos: El EAP en 1.4 satisface que:

$$j^i(x, \varphi, \psi) = \lim_{n \rightarrow \infty} \frac{J_n^i(x, \varphi, \psi)}{n} = \rho_i(\varphi, \psi), \tag{3.4}$$

donde

$$\rho_i(\varphi, \psi) := \int_X r_i(x, \varphi(x), \psi(x)) q(\varphi, \psi)(dx), \tag{3.5}$$

con $q(\varphi, \psi)$ como en (2.12).

Ahora podemos enunciar el teorema y pasar enseguida a su prueba.

Teorema 3.1.2. Bajo las Suposiciones 2.1.1, 2.1.2, 2.1.4 y 3.1.1 - 3.1.3, existe un par $(\varphi_*, \psi_*) \in \Phi_1 \times \Phi_2$ que es un equilibrio de Nash.

Demostración. De (3.4) y el Corolario 2.1.1 (a), podemos escribir $\rho_1^*(\psi)$ en (3.1) como

$$\rho_1^*(\psi) = \rho_1(\varphi_*, \psi) = \max_{\varphi \in \Phi_1} \rho_1(\varphi, \psi). \tag{3.6}$$

De manera análoga, para cada $\varphi \in \Phi_1$, existe $\psi_* \in \Phi_2$ tal que

$$\rho_2^*(\varphi) = \rho_2(\varphi, \psi_*) = \max_{\psi \in \Phi_2} \rho_2(\varphi, \psi). \quad (3.7)$$

Utilizamos a continuación (3.6) y (3.7) para introducir una multifunción τ de $\Phi_1 \times \Phi_2$ a $2^{\Phi_1 \times \Phi_2}$ como sigue: para cada par $(\varphi, \psi) \in \Phi_1 \times \Phi_2$ definimos

$$\tau(\varphi, \psi) := \{(\varphi_*, \psi_*) \mid \rho_1(\varphi_*, \psi) = \rho_1^*(\psi), \rho_2(\varphi, \psi_*) = \rho_2^*(\varphi)\}. \quad (3.8)$$

Para completar la prueba del Teorema 3.1.2 procederemos en dos pasos.

Paso 1. Inducir una topología sobre Φ_i ($i = 1, 2$) con respecto a la cual S_i es compacto y metrizable.

Paso 2. Mostrar que la función es semicontinua superior (u.s.c.), es decir, si

- (i) $(\varphi_n, \psi_n) \rightarrow (\varphi_\infty, \psi_\infty)$ en $\Phi_1 \times \Phi_2$, y
- (ii) $(\varphi_{*n}, \psi_{*n}) \in \tau(\varphi_n, \psi_n)$ es tal que $(\varphi_{*n}, \psi_{*n}) \rightarrow (\varphi_{*\infty}, \psi_{*\infty})$, entonces $(\varphi_{*\infty}, \psi_{*\infty})$ esta en $\tau(\varphi_\infty, \psi_\infty)$.

De estos dos pasos y del Teorema 3.1.1, se seguirá que la multifunción tiene un punto fijo $(\varphi_*, \psi_*) \in \Phi_1 \times \Phi_2$, tal que

$$(\varphi_*, \psi_*) \in \tau(\varphi_*, \psi_*). \quad (3.9)$$

Finalmente, de (3.6)-(3.9) concluiremos que (φ_*, ψ_*) es un equilibrio de Nash.

En el Paso 1 usaremos la topología inducida por Warga [12]. Sea F_1 el espacio de las funciones medibles de Banach $f : \mathbb{K}_A \rightarrow \mathbb{R}$ es tal que $f(x, a)$ es continua en $a \in A(x)$ para cada $x \in X$ y

$$\|f\| := \int_X \max_{a \in A(x)} |f(x, a)| \gamma(dx) < \infty,$$

con γ como en la Suposición 2.1.4. Identificaremos dos estrategias estacionarias φ y φ' en Φ_1 si $\varphi = \varphi'$ γ -c.s., y, por otro lado, $\varphi \in \Phi_1$ puede identificarse con el funcional $\Delta_\varphi \in F_1^*$ dado por

$$\Delta_\varphi(f) := \int_X \int_A f(x, a) \varphi(da|x) \gamma(dx).$$

Por otro lado, Φ_1 puede identificarse como un subconjunto de F_1^* , y crear Φ_1 con la topología débil se puede mostrar que Φ_1 es compacto y metrizable. Con el conjunto Φ_2 se trabaja de forma análoga.

Para continuar con el paso 2, suponga que

$$(\varphi_n, \psi_n) \rightarrow (\varphi_\infty, \psi_\infty) \in \Phi_1 \times \Phi_2, \quad (3.10)$$

y que

$$(\varphi_{*n}, \psi_{*n}) \in \tau(\varphi_n, \psi_n), \quad \forall n \quad (3.11)$$

es tal que

$$(\varphi_{*n}, \psi_{*n}) \rightarrow (\varphi_{*\infty}, \psi_{*\infty}). \quad (3.12)$$

Por (3.11) y la definición (3.8) de τ , junto con (3.2), para todo $x \in X$ tenemos

$$\begin{aligned} \rho_1^*(\psi_n) + h_{\varphi_{*n}, \psi_n}^1(x) &= r_1(x, \varphi_{*n}(x), \psi_n(x)) \\ &\quad + \int_X h_{\varphi_{*n}, \psi_n}^1(y) Q(dy|x, \varphi_{*n}(x), \psi_n(x)) \end{aligned} \quad (3.13)$$

y, análogamente,

$$\begin{aligned} \rho_2^*(\varphi_n) + h_{\varphi_n, \psi_{*n}}^2(x) &= r_2(x, \varphi_n(x), \psi_{*n}(x)) \\ &\quad + \int_X h_{\varphi_n, \psi_{*n}}^2(y) Q(dy|x, \varphi_n(x), \psi_{*n}(x)). \end{aligned} \quad (3.14)$$

Obsérvese que, por las Suposiciones 3.1.2 y 3.1.3, para cada $D \in \mathcal{B}(X)$, las funciones $Q_1(D|x, a)$ y $r_{i1}(x, a)$ están en

F_1 , mientras que $Q_2(D|x, b)$ y $r_{i2}(x, b)$ están en F_2 . Por lo tanto, por (3.10) y (3.12) tenemos que

$$\int_X r_1(x, \varphi_{*n}(x), \psi_n(x)) \gamma(dx) \rightarrow \int_X r_1(x, \varphi_{*\infty}(x), \psi_\infty(x)) \gamma(dx), \quad (3.15)$$

y análogamente para $i = 2$. Además, para cualquier $D \in \mathcal{B}(X)$, se tiene que

$$\int_X Q(D|x, \varphi_{*n}(x), \psi_n(x)) \gamma(dx) \rightarrow \int_X Q(D|x, \varphi_{*\infty}(x), \psi_\infty(x)) \gamma(dx), \quad (3.16)$$

y similarmente para $(\varphi_n, \psi_{*n}) \rightarrow (\varphi_\infty, \psi_{*\infty})$.

El siguiente lema será de ayuda para mostrar que nuestras estrategias son óptimas.

Lema 3.1.3. Existe una subsucesión $\{m\}$ de $\{n\}$ y números $\hat{\rho}_1$ y $\hat{\rho}_2$ tal que

$$\rho_1^*(\psi_m) = \rho_1(\varphi_{*m}, \psi_m) \rightarrow \hat{\rho}_1 \quad (3.17)$$

y

$$\rho_2^*(\varphi_m) = \rho_2(\varphi_m, \psi_{*m}) \rightarrow \hat{\rho}_2. \quad (3.18)$$

Demostración. Sea $\rho_i(\varphi, \psi)$ como en (3.5). A continuación, mostramos que, para $i = 1, 2$,

$$|\rho_i(\varphi, \psi)| \leq \frac{\bar{r} \|v\|_w}{1 - \alpha} \quad \forall (\varphi, \psi) \in \Phi_1 \times \Phi_2, \quad (3.19)$$

con \bar{r} como en (2.9), y v y α como en la Suposición 2.1.2. Claramente, (3.19) implica (3.17) y (3.18).

Para probar (3.19), note que la Suposición 2.1.2 (b) produce que

$$\int_X w(y) Q(dy|x, a, b) \leq \alpha w(x) + \|v\|_w \quad (3.20)$$

ya que $\beta(x, a, b) \leq 1$. Ahora sea (φ, ψ) un par arbitrario en $\Phi_1 \times \Phi_2$. Integrando ambos lados de (3.20) con respecto a $\varphi(da|x)$ y $\psi(db|x)$, y luego integrando con respecto a la medida de probabilidad invariante $q(\varphi, \psi)$ tenemos que

$$\int_X w(y)q(\varphi, \psi)(dy) \leq \alpha \int_X q(\varphi, \psi)(dy) + \|v\|_w,$$

y, por lo tanto,

$$\int_X w(y)q(\varphi, \psi)(dy) \leq \frac{\|v\|_w}{1 - \alpha}.$$

Esta última desigualdad, junto con (2.9) y (3.5), nos da

$$\begin{aligned} |\rho_i(\varphi, \psi)| &\leq \int_X |r_i(x, \varphi(x), \psi(x))|q(\varphi, \psi)(dx) \\ &\leq \bar{r} \int_X w(y)q(\varphi, \psi)(dy) \\ &\leq \frac{\bar{r}\|v\|_w}{1 - \alpha}, \end{aligned}$$

es decir, (3.19). Esto nos lleva a su vez a que las sucesiones $\{\rho_1(\varphi_{*n}, \psi_n)\}$ y $\{\rho_2(\varphi_n, \psi_{*n})\}$ son uniformemente acotadas, y así se sigue el Lema \blacksquare

Por conveniencia de notación, vamos a escribir a la subsucesión $\{m\} \subset \{n\}$ en (3.17) y (3.18) como la sucesión original, $\{n\}$. Además, sean

$$u_n(\cdot) := h_{\varphi_{*n}, \psi_n}^1(\cdot), \quad \text{y} \quad \tilde{u}_n(\cdot) := \frac{u_n(\cdot)}{w(\cdot)}. \quad (3.21)$$

La constante $m_0 := \frac{\bar{r}M}{1-\theta}$ satisface que

$$|\tilde{u}_n(x)| \leq m_0$$

Sea U el espacio de todas las clases de γ -equivalencia de funciones medibles real-valuadas u sobre X tales que $|u(x)| \leq m_0$ γ -c.d. Por el Teorema de Alaoglu (página 424 en [3]),

U es un subconjunto compacto y metrizable de $L^\infty(\gamma) \equiv L^\infty(X, \mathcal{B}(X))$ equipado con la topología relativa débil $\sigma(L^\infty(\gamma), L^1(\gamma))$. Por lo tanto, podemos suponer que $\{\tilde{u}_n\}$ converge en la topología débil a alguna función \tilde{u}_* en $L^\infty(\gamma)$. Sea $u_*(x) := \tilde{u}_*(x)w(x)$ para todo $x \in X$. Entonces, como en la prueba del Teorema 4 en [19], usando la Suposición 3.1.1, se puede mostrar cuando $n \rightarrow \infty$ que

$$\max_{a \in A(x)} \max_{b \in B(x)} \left| \int_X (u_n(y) - u_*(y)) Q(dy|x, a, b) \right| \rightarrow 0 \quad \forall x \in X, \quad (3.22)$$

con $u_n(\cdot)$ como en (3.21). A su vez, de (3.22) y de la Suposición 3.1.2 tenemos que

$$\max_{\varphi \in \mathbb{P}_A(x)} \max_{\psi \in \mathbb{P}_B(x)} \left| \int_X (u_n(y) - u_*(y)) Q(dy|x, \varphi, \psi) \right| \rightarrow 0, \quad (3.23)$$

para todo $x \in X$.

Lema 3.1.4. Si $(\varphi_n, \psi_n) \rightarrow (\varphi, \psi)$ en $\Phi_1 \times \Phi_2$, entonces, cuando $n \rightarrow \infty$,

$$\int_X \int_X u(y) Q(dy|x, \varphi_n(x), \psi_n(x)) \gamma(dx) \rightarrow \int_X \int_X u(y) Q(dy|x, \varphi(x), \psi(x)) \gamma(dx)$$

para cualquier función $u \in \mathbb{B}_w(X)$.

Demostración. Sea $u \in \mathbb{B}_w(X)$. Por definición de la convergencia débil de $\varphi_n \rightarrow \varphi$ y $\psi_n \rightarrow \psi$ en Φ_1 y Φ_2 , respectivamente, y por la Suposición 3.1.2, para demostrar el lema basta con demostrar que las funciones

$$(x, a) \rightarrow \int_X u(y) Q_1(dy|x, a) \quad (3.24)$$

y

$$(x, b) \rightarrow \int_X u(y) Q_2(dy|x, b) \quad (3.25)$$

están en F_1 y F_2 , respectivamente. Teniendo esto en cuenta, observemos que $\int_X u(y) Q_i(dy|x, \cdot)$ es continua en $a \in A(x)$ y

$x \in B(x)$, para $i = 1, 2$, respectivamente. Además, por (2.10) y la Suposición 2.1.2(b) (usando que $\beta(x, a, b) \leq 1$), tenemos que

$$\max_{a \in A(x)} \left| \int_X u(y) Q_1(dy|x, a) \right| \leq \|u\|_w (\alpha w(x) + \|v\|_w) \quad \forall x \in X.$$

Por lo tanto, como $\int w d\gamma < \infty$ (por la Suposición 2.1.4), la función en (3.24) está en F_1 . Análogamente, la función en (3.25) está en F_2 ■

Por los Lemas 3.1.3 y 3.1.4, junto con (3.15), (3.16) y (3.23), cuando $n \rightarrow \infty$ en (3.13) obtenemos que (γ -c.s.)

$$\begin{aligned} \hat{\rho}_1 + u_*(x) &= r_1(x, \varphi_{*\infty}(x), \psi_\infty(x)) + \\ &\quad \int_X u_*(y) Q(dy|x, \varphi_{*\infty}(x), \psi_\infty(x)) \\ &= \max_{\varphi \in \mathbb{P}_A(x)} [r_1(x, \varphi, \psi_\infty(x)) + \\ &\quad \int_X u_*(y) Q(dy|x, \varphi, \psi_\infty(x))], \end{aligned} \tag{3.26}$$

en donde la segunda igualdad viene de (3.2)-(3.3) reemplazando (φ_*, ψ) por (φ_{*n}, ψ_n) .

Por último, argumentando como en la última parte de la demostración del Teorema 5.8 en [10], $D \in \mathcal{B}(X)$ el conjunto con $\gamma(D) = 1$ en el que (3.26) se cumple, y sea $h_* : X \rightarrow \mathbb{R}$ tal que $h_*(x) := u_*(x)$ para $x \in D$, y

$$\begin{aligned} h_*(x) &:= \max_{\varphi \in \mathbb{P}_A(x)} [r_1(x, \varphi, \psi_\infty(x)) \\ &\quad + \int_X u_*(y) Q(dy|x, \varphi, \psi_\infty(x))] - \hat{\rho}_1 \end{aligned}$$

para todo x en D^C . Como $\gamma(D^C) = 0$, por el Lema 6.3 en [10], tenemos que $Q(D^C|x, a, b) = 0$ para todo $(x, a, b) \in \mathbb{K}$. Por lo tanto, (3.26) se cumple para todo $x \in X$ cuando $u_*(\cdot)$ se sustituye por $h_*(\cdot)$. Esto implica (por el Lema 3.1.1) que

$$\hat{\rho}_1 = \rho_1^*(\psi_\infty) = \rho_1(\varphi_{*\infty}, \psi_\infty). \tag{3.27}$$

Un argumento análogo, usando (3.11), (3.12) y (3.14) con cambios evidentes, muestra que

$$\hat{\rho}_2 = \rho_2^*(\varphi_\infty) = \rho_2(\varphi_\infty, \psi_{*\infty}). \quad (3.28)$$

En otras palabras, (3.27) y (3.28) indican que, bajo (3.10)-(3.12), el par $(\varphi_{*\infty}, \psi_{*\infty})$ está en $\tau(\varphi_\infty, \psi_\infty)$, y por lo tanto la función de valor fijo definida por (3.8) es u.s.c. Por lo tanto, como ya se señaló, se deduce del Teorema del punto fijo de Fan que τ tiene un punto fijo (como en (3.9), por ejemplo), que completa la demostración del Teorema 3.1.2 ■

3.2. Dilema del Prisionero Iterado

A modo de ilustración, consideremos al Dilema del Prisionero Iterado. Pero antes, veamos en qué consiste el Dilema del Prisionero, es decir, el juego en su forma estática.

Dos ladrones están siendo interrogados por la policía por un delito grave. Están reclusos en celdas separadas y no pueden hablar entre sí. Sin una confesión, la policía solo tiene evidencia suficiente para condenar a los dos ladrones por un cargo menor. La policía hace la siguiente oferta a ambos presos: si solo uno confiesa que ambos cometieron el delito grave, entonces el confesor será puesto en libertad y el otro pasará 5 años en prisión; si ambos confiesan, entonces van a tener una sentencia de 4 años cada uno; si ninguno confiesa, entonces su sentencia será de 2 años por el delito menor.

Tabla 3.1: Forma normal del Dilema del Prisionero.

		J_2	
		C	D
J_1	C	-2,-2	-5,0
	D	0,-5	-4,-4

La forma normal de este juego se presenta en la Tabla 3.1, donde J_i representa al jugador (o prisionero) i , C que el jugador no confiesa, es decir, que coopera con su compañero y D que el jugador confiesa.

El equilibrio de Nash para este juego estático es (S, S) . Sin embargo, nuestro interés es estudiar el caso iterado de este juego, para ello consideremos la Tabla 3.2 para iniciar con este análisis, donde hemos sumado 5 unidades a cada ganancia.

Tabla 3.2: Forma normal del juego.

		J_2	
		C	D
J_1	C	3,3	0,5
	D	5,0	1,1

Ahora consideremos algunas estrategias, $S_C, S_D, S_T, S_A, S_G, S_Q, S_P$, que nos dicen cómo actuar en cada etapa del juego. Las estrategias estan definidas como:

- S_C , jugar siempre C ;
- S_D , jugar siempre D ;
- S_G , inicia cooperando y sigue cooperando hasta que el otro no lo haga, y luego no vuelve a cooperar;
- S_T , inicia cooperando y a continuación hace lo que el otro jugador hizo en la etapa anterior;
- S_A , inicia no cooperando, a continuación hace lo que el otro jugador hizo en la etapa anterior;

- S_Q , coopera con probabilidad q y no coopera con probabilidad $1 - q$, con $q \in (0, 1)$;
- S_P , inicia cooperando y coopera si el otro también lo hace, en otro caso coopera con probabilidad p y no coopera con probabilidad $1 - p$, donde $p \in (0, 1)$.

El modelo del juego estocástico es como el dado en (1.1), en donde:

- $X = \{(0, 5), (1, 1), (3, 3), (5, 5)\}$,
- $A = B = \{0, 1\} = A(x) = B(x)$ para todo $x \in X$, con $C = 1$ y $D = 0$,
- $Q \in \mathbb{P}(X|\mathbb{K})$, con $\mathbb{K} = \{(x, a, b) | x \in X, a \in A(x), b \in B(x)\}$, esta dado por:

$$Q(\{(0, 5)\} | x_t, a_t, b_t) = \begin{cases} 1, & \text{si } a_t = 1, b_t = 0, x_t \in X \\ 0, & \text{c.o.c.} \end{cases}$$

$$Q(\{(1, 1)\} | x_t, a_t, b_t) = \begin{cases} 1, & \text{si } a_t = b_t = 0, x_t \in X \\ 0, & \text{c.o.c.} \end{cases}$$

$$Q(\{(3, 3)\} | x_t, a_t, b_t) = \begin{cases} 1, & \text{si } a_t = b_t = 1, x_t \in X \\ 0, & \text{c.o.c.} \end{cases}$$

$$Q(\{(5, 0)\} | x_t, a_t, b_t) = \begin{cases} 1, & \text{si } a_t = 0, b_t = 1, x_t \in X \\ 0, & \text{c.o.c.} \end{cases}$$

Las estrategias para el jugador 1 quedan definidas de la siguiente forma:

- $S_D = \{\pi_t^1 | \pi_t^1(\{0\} | h_t) = 1, t = 0, 1, \dots\}$.

- $S_C = \{\pi_t^1 | \pi_t^1(\{1\} | h_t) = 1, t = 0, 1, \dots\}$.

En estas dos estrategias la decisión no depende de ningún estado anterior, así que podemos definir ambas estrategias como $\{\pi_t^1 = 1, t = 0, 1, \dots\}$.

- S_G es tal que $\pi_0^1(\{1\} | h_t) = 1$; para $t = 1, 2, \dots$,

$$\pi_t^1(\{1\} | h_t) = \begin{cases} 1, & \text{si } a_{t-1} = b_{t-1} \\ 0, & \text{c.o.c.} \end{cases}$$

y

$$\pi_t^1(\{0\} | h_t) = \begin{cases} 1, & \text{si } a_{t-1} = 0 \text{ o } b_{t-1} = 0 \\ 0, & \text{c.o.c.} \end{cases}$$

Para esta estrategia, como en las dos anteriores, solo dependemos de la decisión del jugador 2 y se coopera siempre que él lo haga, una vez que no sea así se deja de cooperar.

- S_T es tal que $\pi_0^1(\{1\} | h_t) = 1$ y

$$\pi_t^1(\{a_t\} | h_t) = \begin{cases} 1, & \text{si } a_t = b_{t-1}, t = 1, 2, \dots, \\ 0, & \text{c.o.c.} \end{cases}$$

- S_A es tal que $\pi_0^1(\{0\} | h_t) = 1$ y para $t = 1, 2, \dots$,

$$\pi_t^1(\{a_t\} | h_t) = \begin{cases} 1, & \text{si } a_t = b_{t-1} \\ 0, & \text{c.o.c.} \end{cases}$$

Para este par de estrategias dependemos solo de la decisión de la etapa anterior, es decir $t - 1$, del jugador 2 y copiamos su decisión.

- $S_Q =$

$$\{\pi_t^1 | \pi_t^1(\{a_t\} | h_t) = \begin{cases} q, & \text{si } a_t = 1 \\ 1 - q, & \text{si } a_t = 0 \end{cases} t = 0, 1, \dots, \}$$

En esta estrategia tampoco dependemos de algún estado anterior.

- S_P es tal que $\pi_0^1(\{1\}|h_t) = 1$; para $t = 1, 2, \dots$,

$$\pi_t^1(\{1\}|h_t) = \begin{cases} 1, & \text{si } b_{t-1} = 1 \\ p, & \text{si } b_{t-1} = 0 \end{cases}$$

y

$$\pi_t^1(\{0\}|h_t) = \begin{cases} 1 - p, & \text{si } b_{t-1} = 0 \\ 0, & \text{si } b_{t-1} = 1 \end{cases}$$

Para esta última estrategia tenemos que de la decisión del jugador 1 depende de la decisión anterior del jugador 2.

Por lo ya mencionado anteriormente, es claro que trabajamos con estrategias estacionarias y estacionarias markovianas. Luego, por el Teorema 3.1.2 existe un par de estrategias que es un equilibrio de Nash.

Enseguida se calculan las ganancias de los jugadores usando el criterio EAP que esta definido en (2.39), con los cambios en la notación que se mencionaron al inicio del capítulo. Las Tablas 3.3-3.51 ilustran la acción que elige cada jugador en las diferentes etapas del juego.

Supongamos que el jugador 1 elige la estrategia S_D , entonces si el jugador 2 elige:

- S_D , los jugadores actuaran de la siguiente forma:

Tabla 3.3: S_D vs S_D

Etapa	0	1	2	3	4	5	...
J_1	D	D	D	D	D	D	...
J_2	D	D	D	D	D	D	...

La Tabla 3.3 dice que ambos jugadores van a elegir D en cada etapa, que es como se definieron las estrategias.

Por este comportamiento, $J^1(x, S_D, S_D) = 1 = J^2(x, S_D, S_D)$. De aquí en adelante mostraremos el comportamiento de los jugadores en tablas como se hizo en la Tabla 3.3.

- S_C ,

Tabla 3.4: S_D vs S_C

Etapa	0	1	2	3	4	5	...
J_1	D	D	D	D	D	D	...
J_2	C	C	C	C	C	C	...

Así, $J^1(x, S_D, S_C) = 5 = J^2(x, S_C, S_D)$.

- S_G ,

Tabla 3.5: S_D vs S_G

Etapa	0	1	2	3	4	5	...
J_1	D	D	D	D	D	D	...
J_2	C	D	D	D	D	D	...

$J^1(x, S_D, S_G) = 1 = J^2(x, S_G, S_D)$.

- S_T ,

Tabla 3.6: S_D vs S_T

Etapa	0	1	2	3	4	5	...
J_1	D	D	D	D	D	D	...
J_2	C	D	D	D	D	D	...

$J^1(x, S_D, S_T) = 1 = J^2(x, S_T, S_D)$.

- S_A ,

Tabla 3.7: S_D vs S_A

Etapa	0	1	2	3	4	5	...
J_1	D	D	D	D	D	D	...
J_2	D	D	D	D	D	D	...

$$J^1(x, S_A, S_D) = 1 = J^2(x, S_A, S_D).$$

- S_Q ,

Tabla 3.8: S_D vs S_Q

Etapa	0	1	2	3	...
J_1	D	D	D	D	...
J_2	Cq y $D(1-q)$	Cq y $D(1-q)$	Cq y $D(1-q)$	Cq y $D(1-q)$...

En este caso, J_2 coopera con probabilidad C y no coopera con probabilidad $1 - q$, por lo tanto $J^1(x, S_D, S_Q) = 4q + 1 = J^2(x, S_Q, S_D)$.

- S_P ,

Tabla 3.9: S_D vs S_P

Etapa	0	1	2	3	...
J_1	D	D	D	D	...
J_2	C	Cp y $D(1-p)$	Cp y $D(1-p)$	Cp y $D(1-p)$...

$$J^1(x, S_D, S_P) = 4p + 1 = J^2(x, S_P, S_D).$$

Si el jugador 1 ahora elige S_C y el jugador 2:

- S_D ,

Tabla 3.10: S_C vs S_D

Etapa	0	1	2	3	4	5	...
J_1	C	C	C	C	C	C	...
J_2	D	D	D	D	D	D	...

$$J^1(x, S_C, S_D) = 0 = J^2(x, S_D, S_C).$$

- S_C ,

Tabla 3.11: S_C vs S_C

Etapa	0	1	2	3	4	5	...
J_1	C	C	C	C	C	C	...
J_2	C	C	C	C	C	C	...

$$J^1(x, S_C, S_C) = 3 = J^2(x, S_C, S_C).$$

- S_G ,

Tabla 3.12: S_C vs S_G

Etapa	0	1	2	3	4	5	...
J_1	C	C	C	C	C	C	...
J_2	C	C	C	C	C	C	...

$$J^1(x, S_C, S_G) = 3 = J^2(x, S_G, S_C).$$

- S_T ,

Tabla 3.13: S_C vs S_T

Etapa	0	1	2	3	4	5	...
J_1	C	C	C	C	C	C	...
J_2	C	C	C	C	C	C	...

$$J^1(x, S_C, S_T) = 3 = J^2(x, S_T, S_C).$$

- S_A ,

Tabla 3.14: S_C vs S_A

Etapa	0	1	2	3	4	5	...
J_1	C	C	C	C	C	C	...
J_2	D	C	C	C	C	C	...

$$J^1(x, S_C, S_A) = 3 = J^2(x, S_A, S_C).$$

- S_Q ,

Tabla 3.15: S_C vs S_Q

Etapa	0	1	2	3	...
J_1	C	C	C	C	...
J_2	Cq y D1 - q	Cq y D1 - q	Cq y D1 - q	Cq y D1 - q	...

$$J^1(x, S_C, S_Q) = 3q = J^2(x, S_Q, S_C).$$

- S_P ,

Tabla 3.16: S_C vs S_P

Etapa	0	1	2	3	4	5	...
J_1	C	C	C	C	C	C	...
J_2	C	C	C	C	C	C	...

$$J^1(x, S_C, S_P) = 3 = J^2(x, S_P, S_C).$$

Si el jugador 1 elige S_G y el jugador 2 elige:

- S_D ,

Tabla 3.17: S_G vs S_D

Etapa	0	1	2	3	4	5	...
J_1	C	D	D	D	D	D	...
J_2	D	D	D	D	D	D	...

$$J^1(x, S_G, S_D) = 1 = J^2(x, S_D, S_G).$$

- S_C ,

Tabla 3.18: S_G vs S_C

Etapa	0	1	2	3	4	5	...
J_1	C	C	C	C	C	C	...
J_2	C	C	C	C	C	C	...

$$J^1(x, S_G, S_C) = 3 = J^2(x, S_C, S_G).$$

- S_G ,

Tabla 3.19: S_G vs S_G

Etapa	0	1	2	3	4	5	...
J_1	C	C	C	C	C	C	...
J_2	C	C	C	C	C	C	...

$$J^1(x, S_G, S_G) = 3 = J^2(x, S_G, S_G).$$

- S_T ,

Tabla 3.20: S_G vs S_T

Etapa	0	1	2	3	4	5	...
J_1	C	C	C	C	C	C	...
J_2	C	C	C	C	C	C	...

$$J^1(x, S_G, S_T) = 3 = J^2(x, S_T, S_G).$$

- S_A ,

Tabla 3.21: S_G vs S_A

Etapa	0	1	2	3	4	5	...
J_1	C	D	D	D	D	D	...
J_2	D	C	C	C	C	C	...

$$J^1(x, S_G, S_A) = 1 = J^2(x, S_A, S_G).$$

- S_Q ,

Tabla 3.22: S_G vs S_Q

Etapa	0	1	2	3	...
J_1	C	qC y $(1-q)D$	q^2C y $(1-q)^2D$	q^3C y $(1-q)^3D$...
J_2	qC y $(1-q)D$	qC y $(1-q)D$	qC y $(1-q)D$	qC y $(1-q)D$...

$$J^1(x, S_G, S_Q) = 4q + 1 = J^2(x, S_Q, S_G).$$

- S_P ,

Tabla 3.23: S_G vs S_P

Etapa	0	1	2	3	4	5	...
J_1	C	C	C	C	C	C	...
J_2	C	C	C	C	C	C	...

$$J^1(x, S_G, S_P) = 3 = J^2(x, S_P, S_G).$$

Si el jugador 1 elige S_T y el jugador 2 elige:

- S_D ,

Tabla 3.24: S_T vs S_D

Etapa	0	1	2	3	4	5	...
J_1	C	D	D	D	D	D	...
J_2	D	D	D	D	D	D	...

$$J^1(x, S_T, S_D) = 1 = J^2(x, S_D, S_T).$$

- S_C ,

Tabla 3.25: S_T vs S_C

Etapa	0	1	2	3	4	5	...
J_1	C	C	C	C	C	C	...
J_2	C	C	C	C	C	C	...

$$J^1(x, S_T, S_C) = 3 = J^2(x, S_C, S_T).$$

- S_G ,

Tabla 3.26: S_T vs S_G

Etapa	0	1	2	3	4	5	...
J_1	C	C	C	C	C	C	...
J_2	C	C	C	C	C	C	...

$$J^1(x, S_T, S_G) = 3 = J^2(x, S_G, S_T).$$

- S_T ,

Tabla 3.27: S_T vs S_T

Etapa	0	1	2	3	4	5	...
J_1	C	C	C	C	C	C	...
J_2	C	C	C	C	C	C	...

$$J^1(x, S_T, S_T) = 3 = J^2(x, S_T, S_T).$$

- S_A ,

Tabla 3.28: S_T vs S_A

Etapa	0	1	2	3	4	5	...
J_1	C	D	C	D	C	D	...
J_2	D	C	D	C	D	C	...

$$J^1(x, S_T, S_A) = \frac{5}{2} = J^2(x, S_T, S_G).$$

- S_Q ,

Tabla 3.29: S_T vs S_Q

Etapa	0	1	2	3	...
J_1	C	qC y $(1-q)D$	qC y $(1-q)D$	qC y $(1-q)D$...
J_2	qC y $(1-q)D$	qC y $(1-q)D$	qC y $(1-q)D$	qC y $(1-q)D$...

$$J^1(x, S_T, S_Q) = -q^2 + 3q + 1 = J^2(x, S_Q, S_T).$$

- S_P ,

Tabla 3.30: S_T vs S_P

Etapa	0	1	2	3	4	5	...
J_1	C	C	C	C	C	C	...
J_2	C	C	C	C	C	C	...

$$J^1(x, S_T, S_P) = 3 = J^2(x, S_P, S_T).$$

Si el jugador 1 elige S_A y el jugador 2:

- S_D ,

Tabla 3.31: S_A vs S_D

Etapa	0	1	2	3	4	5	...
J_1	D	D	D	D	D	D	...
J_2	D	D	D	D	D	D	...

$$J^1(x, S_A, S_D) = 1 = J^2(x, S_D, S_A).$$

- S_C ,

Tabla 3.32: S_A vs S_C

Etapa	0	1	2	3	4	5	...
J_1	D	C	C	C	C	C	...
J_2	C	C	C	C	C	C	...

$$J^1(x, S_A, S_C) = 3 = J^2(x, S_C, S_A).$$

- S_G ,

Tabla 3.33: S_A vs S_G

Etapa	0	1	2	3	4	5	...
J_1	D	C	D	D	D	D	...
J_2	C	D	D	D	D	D	...

$$J^1(x, S_A, S_G) = 1 = J^2(x, S_G, S_A).$$

- S_T ,

Tabla 3.34: S_A vs S_T

Etapa	0	1	2	3	4	5	...
J_1	D	C	D	C	D	C	...
J_2	C	D	C	D	C	D	...

$$J^1(x, S_A, S_T) = \frac{5}{2} = J^2(x, S_T, S_A).$$

- S_A ,

Tabla 3.35: S_A vs S_A

Etapa	0	1	2	3	4	5	...
J_1	D	D	D	D	D	D	...
J_2	D	D	D	D	D	D	...

$$J^1(x, S_A, S_A) = 1 = J^2(x, S_A, S_A).$$

- S_Q ,

Tabla 3.36: S_A vs S_Q

Etapa	0	1	2	3	...
J_1	D	$qy + (1-q)D$	$qC y (1-q)D$	$qC y (1-q)D$...
J_2	$qC y (1-q)D$	$qC y (1-q)D$	$qC y (1-q)D$	$qC y (1-q)D$...

$$J^1(x, S_A, S_Q) = -q^2 + 3q + 1 = J^2(x, S_Q, S_A).$$

- S_P ,

Tabla 3.37: S_A vs S_P

Etapa	0	1	2	3	4	...
J_1	D	C	pC y $(1-p)D$	C	pC y $(1-p)D$...
J_2	C	pC y $(1-p)D$	C	pC y $(1-p)D$	C	...

$$J^1(x, S_A, S_P) = \frac{5}{2} + \frac{p}{2} = J^2(x, S_P, S_A).$$

Si el jugador 1 elige S_Q y el jugador 2:

- S_D ,

Tabla 3.38: S_Q vs S_D

Etapa	0	1	2	3	...
J_1	qC y $(1-q)D$	qC y $(1-q)D$	qC y $(1-q)D$	qC y $(1-q)D$...
J_2	D	D	D	D	...

$$J^1(x, S_Q, S_D) = 1 - q = J^2(x, S_D, S_Q).$$

- S_C ,

Tabla 3.39: S_Q vs S_C

Etapa	0	1	2	3	...
J_1	qC y $(1-q)D$	qC y $(1-q)D$	qC y $(1-q)D$	qC y $(1-q)D$...
J_2	C	C	C	C	...

$$J^1(x, S_Q, S_C) = 5 - 2q = J^2(x, S_C, S_Q).$$

- S_G ,

Tabla 3.40: S_Q vs S_G

Etapa	0	1	2	3	...
J_1	qC y $(1-q)D$	qC y $(1-q)D$	qC y $(1-q)D$	qC y $(1-q)D$...
J_2	C	qC y $(1-q)D$	q^2C y $(1-q^2)D$	q^3C y $(1-q^3)D$...

$$J^1(x, S_Q, S_G) = 1 - q = J^2(x, S_G, S_Q).$$

- S_T ,

Tabla 3.41: S_Q vs S_T

Etapa	0	1	2	3	...
J_1	qC y $(1-q)D$	qC y $(1-q)D$	qC y $(1-q)D$	qC y $(1-q)D$...
J_2	C	qC y $(1-q)D$	qC y $(1-q)D$	qC y $(1-q)D$...

$$J^1(x, S_Q, S_T) = -q^2 + 3q + 1 = J^2(x, S_T, S_Q).$$

- S_A ,

Tabla 3.42: S_Q vs S_A

Etapa	0	1	2	3	...
J_1	qC y $(1-q)D$	qC y $(1-q)D$	qC y $(1-q)D$	qC y $(1-q)D$...
J_2	D	qC y $(1-q)D$	qC y $(1-q)D$	qC y $(1-q)D$...

$$J^1(x, S_Q, S_A) = -q^2 + 3q + 1 = J^2(x, S_A, S_Q).$$

- S_Q ,

Tabla 3.43: S_Q vs S_Q

Etapa	0	1	2	3	...
J_1	D	qC y $(1-q)D$	qC y $(1-q)D$	qC y $(1-q)D$...
J_2	qC y $(1-q)D$	qC y $(1-q)D$	qC y $(1-q)D$	qC y $(1-q)D$...

$$J^1(x, S_Q, S_Q) = -q^2 + 3q + 1 = J^2(x, S_Q, S_Q).$$

- S_P ,

Tabla 3.44: S_Q vs S_P

Etapa	0	1	2	3	...
J_1	qC y $(1-q)D$	qC y $(1-q)D$	qC y $(1-q)D$	qC y $(1-q)D$...
J_2	C	qC y $(1-q)(pC$ y $(1-p)D)$	C	qC y $(1-q)(pC$ y $(1-p)D)$...

$$J^1(x, S_Q, S_P) = 1 + q(3 - q) + (1 - q)(4p - pq) = J^2(x, S_P, S_Q).$$

Si el jugador 1 elige S_P y el jugador 2 elige:

- S_D ,

Tabla 3.45: S_P vs S_D

Etapa	0	1	2	3	...
J_1	C	pC y $(1-p)D$	pC y $(1-p)D$	pC y $(1-p)D$...
J_2	D	D	D	D	...

$$J^1(x, S_P, S_D) = 4p + 1 = J^2(x, S_D, S_P).$$

- S_C ,

Tabla 3.46: S_P vs S_C

Etapa	0	1	2	3	4	5	...
J_1	D	C	C	C	C	C	...
J_2	C	C	C	C	C	C	...

$$J^1(x, S_P, S_C) = 3 = J^2(x, S_C, S_P).$$

- S_G ,

Tabla 3.47: S_P vs S_G

Etapa	0	1	2	3	4	5	...
J_1	D	C	D	D	D	D	...
J_2	C	D	D	D	D	D	...

$$J^1(x, S_P, S_G) = 1 = J^2(x, S_G, S_P).$$

- S_T ,

Tabla 3.48: S_P vs S_T

Etapa	0	1	2	3	4	5	...
J_1	D	C	D	C	D	C	...
J_2	C	D	C	D	C	D	...

$$J^1(x, S_P, S_T) = \frac{5}{2} = J^2(x, S_T, S_P).$$

- S_A ,

Tabla 3.49: S_P vs S_A

Etapa	0	1	2
J_1	D	C	$pC y (1-p)D$
J_2	C	$pC y (1-p)D$	C
Etapa	3	4	5
J_1	C	$(1-(1-p)^2)C y (1-p)^2D$	C
J_2	$(1-(1-p)^2)C y (1-p)^2D$	C	$(1-(1-p)^3)C y (1-p)^3D$
Etapa	6	7	...
J_1	$(1-(1-p)^3)C y (1-p)^3D$	C	...
J_2	C	$(1-(1-p)^4)C y (1-p)^4D$...

$$J^1(x, S_P, S_A) = 3 = J^2(x, S_A, S_P).$$

- S_Q ,

Tabla 3.50: S_P vs S_Q

Etapa	0	1	2	...
J_1	C	$(q+p-pq)C y (1-q-p+pq)D$	$(q+p-pq)C y (1-q-p+pq)D$...
J_2	$qC y (1-q)D$	$qC y (1-q)D$	$qC y (1-q)D$...

$$J^1(x, S_P, S_Q) = q^2(p-3) + 3q + 1 - p = J^2(x, S_Q, S_P).$$

- S_P ,

Tabla 3.51: S_P vs S_P

Etapa	0	1	2	3	4	5	...
J_1	C	C	C	C	C	C	...
J_2	C	C	C	C	C	C	...

$$J^1(x, S_P, S_P) = 3 = J^2(x, S_P, S_P).$$

Sean $f(x) = 4x + 1$, $g(x) = 1 - x$, $h(x) = 5 - 2x$,
 $j(x) = -x^2 + 3x + 1$, $k(x) = -x^2(1-p) + x(3-5p) + 1 + 4p$,

$$l(x) = x^2(p - 3) + 3x + 1.$$

La Tabla 3.52 muestra las “ganancias” de los jugadores usando las diferentes estrategias, donde hemos calculado cada una como la suma de las ganancias obtenidas en cada etapa usando el criterio EAP.

Tabla 3.52: EAP de los jugadores.

		J_2						
		S_D	S_C	S_G	S_T	S_A	S_Q	S_P
J_1	S_D	1,1	5,0	1,1	1,1	1,1	$f(q), g(q)$	$f(p), f(p)$
	S_C	0,5	3,3	3,3	3,3	3,3	$3q, h(q)$	3,3
	S_G	1,1	3,3	3,3	3,3	1,1	$f(q), g(q)$	3,3
	S_T	1,1	3,3	3,3	3,3	$\frac{5}{2}, \frac{5}{2}$	$j(q), j(q)$	3,3
	S_A	1,1	3,3	1,1	$\frac{5}{2}, \frac{5}{2}$	1,1	$j(q), j(q)$	$\frac{5+p}{2}, 3$
	S_Q	$g(q), f(q)$	$h(q), 3q$	$g(q), f(q)$	$j(q), j(q)$	$j(q), j(q)$	$j(q), j(q)$	$k(q), l(q)$
	S_P	$f(p), f(p)$	3,3	3,3	3,3	$3, \frac{5+p}{2}$	$l(q), k(q)$	3,3

Ahora hagamos un análisis para comparar las ganancias de los jugadores y ver qué par de estrategias cumple con la definición 3.1.1. Es decir, hallaremos la mejor respuesta para los jugadores.

Tabla 3.53: La mejor respuesta para ambos jugadores.

		J_2						
		S_D	S_C	S_G	S_T	S_A	S_Q	S_P
J_1	S_D	1,1	<u>5,0</u>	1,1	1,1	1,1	<u>$f(q), g(q)$</u>	<u>$f(p)^{***}, f(p)$</u>
	S_C	0, <u>5</u>	3,3	<u>3,3</u>	<u>3,3</u>	<u>3,3</u>	$3q, h(q)$	<u>3*,3</u>
	S_G	1,1	3, <u>3</u>	<u>3,3</u>	<u>3,3</u>	1,1	<u>$f(q), g(q)$</u>	<u>3*,3</u>
	S_T	1,1	3, <u>3</u>	<u>3,3</u>	<u>3,3</u>	$\frac{5}{2}, \frac{5}{2}$	$j(q), j(q)$	<u>3*,3</u>
	S_A	1,1	3, <u>3</u>	1,1	$\frac{5}{2}, \frac{5}{2}$	1,1	$j(q), j(q)$	<u>$\frac{5+p}{2}, 3$</u>
	S_Q	$g(q), f(q)$	$h(q), 3q$	$g(q), f(q)$	$j(q), j(q)$	$j(q), j(q)$	$j(q), j(q)$	<u>$k(q)^{**}, l(q)$</u>
	S_P	<u>$f(p), f(p)^{***}$</u>	3, <u>3*</u>	<u>3,3*</u>	<u>3,3*</u>	<u>3, $\frac{5+p}{2}$</u>	<u>$l(q), k(q)^{**}$</u>	<u>3*,3*</u>

En la Tabla 3.53 se muestra la mejor respuesta de los jugadores, de color azul para el jugador 1 y rojo para el jugador 2. Además, * para $p \in (0, 1)$, ** para $p \in [\frac{1}{2}, \frac{3}{5})$ y *** para $p \in [\frac{3}{5}, 1)$.

Y de esto podemos concluir que los equilibrios de Nash son:

- para $p, q \in (0, 1)$ tenemos a $(S_G, S_G), (S_G, S_T), (S_T, S_G)$ y (S_T, S_T) ;
- para $p \in (0, \frac{1}{2})$ y $q \in (0, 1)$ tenemos a $(S_G, S_G), (S_G, S_T), (S_T, S_G), (S_T, S_T), (S_G, S_P), (S_T, S_P), (S_P, S_G)$ y (S_P, S_T) ;
- para $p \in [\frac{3}{5}, 1)$ y $q \in (0, 1)$ tenemos a $(S_G, S_G), (S_G, S_T), (S_T, S_G), (S_T, S_T), (S_D, S_P)$ y (S_P, S_D) .

Conclusiones y trabajo a futuro

En teoría de juegos el objetivo es encontrar estrategias que optimicen el rendimiento para cada uno de los jugadores, pero esto no siempre es sencillo. En esta tesis se presentaron juegos estocásticos de suma cero y suma no cero.

Los resultados del Capítulo 2 nos aseguran que los 3 criterios, EAP, SPAP y pares canónicos, que se presentaron son equivalentes. Además de ser equivalentes a una condición de martingala. Esto resulta de mucha utilidad pues nos permite elegir el criterio que sea más sencillo al buscar dichos equilibrios.

El juego del volado lo usamos para ilustra este capítulo, que es un juego estático clásico en la teoría de juegos, lo adaptamos a un juego estocástico. Trabajamos con el criterio EAP y algo importante fue que la solución para el juego estático, que es una estrategia mixta, coincide con la solución para el juego estocástico.

Para el Capítulo 3, el resultado que se presentó fue sobre la existencia de equilibrios de Nash y el ejemplo que usamos para este capítulo fue el Dilema del Prisionero Iterado, que fue un juego estudiado en la tesis de licenciatura lo que nos permitió comparar los resultados, usamos el criterio EAP.

El trabajo a futuro es usar los teoremas del Capítulo 2 para juegos estocásticos aplicados en algún ejemplo de la vida real; si fuera posible, pedir menos condiciones a los juegos para que se sigan cumpliendo los teoremas presentados y generalizar los teoremas del Capítulo 2 a juegos de suma no cero.

Bibliografía

- [1] Benítez, Medina, *Existence of Nash equilibria in nonzero-sum ergodic stochastic games in Borel spaces*, MORFISMOS, 5(2), 19-35, 2001.
- [2] Boros, Elbassioni, Gurvich, Makino, *Markov decision processes and stochastic games with total effective payoff*, RUCTOR RESEARCH REPORT, 2014.
- [3] Dunford, Schwartz, *Linear Operators*, INTERSCIENCE PUBLISHERS, Vol 1, 1958.
- [4] Fan, *Fixed point and minimax theorems in locally convex topological spaces*, PROC. NAT. ACAD. SCI., 539-560, 1952.
- [5] Hernández, Lasserre, *Further topics on discrete-time Markov control processes*, SPRINGER SCIENCE+BUSINESS MEDIA, 1999.
- [6] Hernández, Lasserre, *Zero-sum stochastic games in Borel spaces: average payoff criteria*, SIAM J. CONTROL OPTIM, 39(5), 1520-1539, 2001.
- [7] Jassen, *Stochastic games. (Tesis de maestría)*, UNIVERSITEIT LEIDEN, 2006.
- [8] Luque, Vázquez, *Zero-sum semi-Markov games in Borel spaces with discounted payoff*, MORFISMOS, 6(1), 15-29, 2002.

-
- [9] NOWAK, SZAJOWSKI, *Nonzero-sum stochastic games*. In *Stochastic and Differential Games*, M. BARDI, et. al., 297–342, 1999.
- [10] Ramírez, *Existence of optimal strategies for zero-sum stochastic games with discounted payoff*, MORFISMOS, 5(1), 63-83, 2001.
- [11] Soo Chang, Marcus, *Two-person zero-sum Markov games: receding horizon approach*, IEEE TRANSACTIONS ON AUTOMATIC CONTROL, vol. 48, 2003.
- [12] Warga, *Optimal of Differential and Functional Equations*, ACADEMIC PRESS, 1972.
- [13] Yang, Guo, *Zero-Sum Stochastic Games with Average Payoff: New Optimality Conditions*, ACTA MATHEMATICA SINICA, 1201-1216, 2009.