"El saber de mis hijos hará mi grandeza"

UNIVERSIDAD DE SONORA

Facultad Interdisciplinaria de Ciencias Exactas y Naturales

Licenciatura en Matemáticas

Estimación Empírica en Juegos Estocásticos Finitos No Cooperativos

T E S I S

Que para obtener el título de:

Licenciado en Matemáticas

Presenta:

Luis Pablo Flores Guevara

Directora de tesis: Dra. Alejandra Fonseca Morales

Hermosillo, Sonora, México, 25 de marzo de 2025

SINODALES

Dra. Alejandra Fonseca Morales
Universidad de Sonora, Hermosillo, México

Dra. Carmen Geraldi Higuera Chan
Universidad de Sonora, Hermosillo, México

Dr. Jesus Adolfo Minjarez Sosa
Universidad de Sonora, Hermosillo, México

Dr. Oscar Vega Amaya
Universidad de Sonora, Hermosillo, México

A mi mamá...

Agradecimientos

Gracias a mis sinodales por su tiempo y sus valiosas observaciones.

Índice general

Introducción

Los juegos dinámicos estocásticos son introducidos por primer vez por L. S. Shapley en [10] en 1953. En los años siguientes, importantes contribuciones como [6], [7], [3] y [11] desarrollaron más allá el concepto y dieron pie a grandes resultados en el área. Desde entonces ha surgido una amplia cantidad de literatura que ha expandido y generalizado al concepto, incorporando aspectos como información incompleta, restricciones estratégicas y aprendizaje en el tiempo. Estos avances han permitido que, en la actualidad, los juegos estocásticos dinámicos sean aplicados en un sinnúmero de situaciones, como en el área financiera y en la inteligencia artificial.

La presente tesis se concentra en estudiar juegos dinámicos estocásticos para dos jugadores con un horizonte infinito, es decir, con una cantidad infinita de etapas. Los jugadores serán referidos a lo largo de la tesis como jugador 1 y jugador 2. Un juego dinámico estocástico se juega de la siguiente manera. Primero, los jugadores observan el estado actual x del juego y eligen acciones a y b. Después, el jugador 1 recibe un pago $r_1(x, a, b)$ y el jugador 2 un pago de $r_2(x, a, b)$, para luego pasar a un nuevo estado y estocásticamente siguiendo una ley de transición $Q(\cdot|x, a, b)$, para después repetir el proceso indefinidamente. Los pagos son acumulados a lo largo de todo el juego bajo el criterio de optimalidad del pago esperado α-descontado. Se asume que los jugadores toman una postura no cooperativa, lo cual significa que eligen sus mejores decisiones sin considerar una cooperación "social" entre ellos, siendo así los equilibrios de Nash las soluciones para los juegos dinámicos en este trabajo.

Se prestará especial atención a los juegos cuyos estados $\{x_t\}$ evolucionen según una ecuación en diferencias F de la forma

$$x_{t+1} = F(x_t, a_t, b_t, \xi_t), \quad t = 0, 1, 2, ... \tag{1}$$

Aquí, el par (a_t, b_t) representa a las acciones elegidas por el jugador 1 y el jugador 2, respectivamente, en el tiempo t, mientras que $\{\xi_t\}$ es llamado proceso de perturbación y es una sucesión de variables aleatorias independientes idénticamente distribuidas con distribución θ para ambos jugadores.

Una referencia fundamental para esta tesis es [5], en donde se estudian juegos dinámicos estocásticos de suma cero en los que el conjunto de estados y los conjuntos de acciones de los jugadores 1 y 2 son espacios de Borel. No obstante, en el presente trabajo se abordarán juegos de suma cero, considerando el caso particular donde los conjuntos de estados y acciones son finitos. Esta elección permite simplificar el análisis teórico y facilita la implementación de ejemplos numéricos.

También, se estudian juegos estocásticos de suma no cero con resultados extraídos de [9]. Cabe mencionar que se trata de un trabajo de investigación de años recientes. Los modelos a estudiar de juegos de suma no cero también utilizan conjuntos de estados y acciones finitos. Incluso, en el Capítulo 4, se presenta un ejemplo numérico para esta clase de juegos.

En particular, se tiene el objetivo de estudiar modelos de juegos de suma cero y de suma no cero donde la distribución θ es desconocida. En casos como este, al modelo se le conoce como juego empírico y es llevado a cabo de la siguiente manera. Al tiempo t, luego de que los jugadores observan el estado actual x_t, proceden a estimar θ utilizando la llamada "distribución empírica", obteniendo así un estimador θ_t. Después, los jugadores adaptan sus acciones a dicho estimador, seleccionando las acciones $a_t(\theta_t)$ y $b_t(\theta_t)$. Luego, el juego se mueve a un nuevo estado siguiendo (1) con la distribución desconocida θ. Debido a que, al hacer tender t a infinito, se sabe que la sucesión de estimaciones θ_t se aproxima a la distribución real θ, se tendrá que este método de aproximación permite acercarse de forma asintótica a soluciones que, a pesar de no ser necesariamente equilibrios de Nash, son "casi" equilibrios.

El trabajo está organizado como sigue. En el Capítulo 1 se introduce el modelo de juego y se describen sus elementos. Además, se presenta el concepto de estrategia y se definen las funciones de pago esperado α-descontado para cada jugador y el concepto de equilibrio de Nash. A continuación, en el Capítulo 2, se presenta el contexto de juegos suma cero, se trata la existencia de un equilibrio de Nash y se analiza el concepto de optimalidad asintótica, el cual es útil para identificar "casi" equilibrios. Dimilarmente, estos conceptos se adoptan para juegos de suma no cero en el Capítulo 3. Finalmente, en el Capítulo 4 se termina por mostrar un ejemplo numérico de un juego de suma no cero, llamado "La gran guerra por pescado". En este ejemplo se halla un equilibrio de Nash

y se aproxima a este asintóticamente simulando juegos empíricos. Se incluye un nuevo código que, junto al descrito en [9], permite ordenar y graficar los datos.

Capítulo 1

Juegos Estocásticos

En este capítulo se presentan los juegos estocásticos y se definen los elementos necesarios para estudiarlos. Primero, se introduce el modelo de juego, el cual es una colección de objetos que describen la evolución del juego en el tiempo. Luego, se define al concepto de estrategias para los jugadores y a las funciones de pago esperado para los jugadores. Finalmente, se define el equilibrio de Nash.

1.1. Modelo de juego

El modelo de un juego estocástico de dos jugadores es definido por la colección

$$\mathcal{G} := (X, A, B, Q, r_1, r_2), \tag{1.1}$$

conformada por:

(a) Un conjunto finito X llamado *espacio de estados*.

(b) Conjuntos finitos A y B llamados *espacios de acción* para los jugadores 1 y 2, respectivamente.

(c) La ley de transición entre los estados, $Q(\cdot|\cdot)$. Esto es, si $x \in X$ es el estado del juego en el tiempo t y los jugadores 1 y 2 eligen acciones $a \in A$ y $b \in B$, respectivamente, entonces $Q(\cdot|x, a, b)$ es la distribución del siguiente estado del juego:

$$Q(y|x, a, b) := Pr[x_{t+1} = y|x_t = x, a_t = a, b_t = b], \quad y \in X. \tag{1.2}$$

(d) Finalmente, $r_i : X \times A \times B \to \mathbb{R}$ es la función de ganancia por etapa para el jugador $i = 1, 2$.

Sea $\mathbb{K} := X \times A \times B$. Un elemento $(x, a, b) \in \mathbb{K}$ denota que, en el estado x, el jugador 1 y el jugador 2 eligen las acciones $a \in A$ y $b \in B$, respectivamente.

Un juego estocástico $\mathcal{G}$ se juega de la manera siguiente. En cada tiempo $t \in \mathbb{N}_0 := \{0, 1, 2, ...\}$ los jugadores 1 y 2 observan el estado $x_t = x$ y eligen acciones $a_t = a \in A$ y $b_t = b \in B$, respectivamente. Esto implica dos situaciones: primero, hay una ganancia de $r_1(x, a, b)$ para el jugador 1 y de $r_2(x, a, b)$ para el jugador 2; segundo, el juego pasa a un nuevo estado $x_{t+1} = y \in X$ según la ley de transición $Q(\cdot|x, a, b)$. Una vez que el sistema del juego está en el nuevo estado y, el problema de elección para los jugadores se repite. Como los jugadores están recibiendo ganancias en cada etapa, su objetivo es maximizarlas.

En ocasiones, la evolución del juego está dada por una ecuación estocástica en diferencias dada por el sistema

$$x_{t+1} = F(x_t, a_t, b_t, \xi_t), \qquad t \in \mathbb{N}_0, \tag{1.3}$$

donde $F : \mathbb{K} \times S \to X$ es una función conocida y $\{\xi_t\}$ es una sucesión de variables aleatorias independientes e idénticamente distribuidas (i.i.d.), que están definidas en el espacio de probabilidad $(\Omega, \mathcal{F}, P)$ y que toman valores en un espacio finito S.

Nótese que, si θ es la distribución de probabilidad de la variable aleatoria ξ_t, i.e.,

$$\theta(s) = P[\xi_t = s] \quad \forall s \in S, t \in \mathbb{N}_0, \tag{1.4}$$

entonces, en la etapa t, la ley de transición Q en (1.2) toma la forma

$$\begin{aligned} Q(y|x, a, b) &= Pr[x_{t+1} = y | x_t = x, a_t = a, b_t = b] \\ &= Pr[s \in S : F(x, a, b, s) = y] \\ &= \sum_{s \in S} 1_y[F(x, a, b, s)]\theta(s), \quad y \in X, \end{aligned} \tag{1.5}$$

para $(x, a, b) \in \mathbb{K}$ y donde 1_y denota la función indicadora en y. Esto es, $1_y(x) = 1$ cuando $x = y$ y $1_y(x) = 0$ cuando $x \neq y$.

1.2. Estrategias

En cada etapa $t = 0, 1, 2, ...$ del juego, los jugadores observan el estado x_t y seleccionan sus acciones de acuerdo con distribuciones de probabilidad sobre sus respectivos conjuntos de acciones. La secuencia de distribuciones de probabilidad empleadas por cada jugador se denomina estrategia, cuya definición formal se presenta a continuación.

Sean $H_0 := X$ y $H_t := H_{t-1} \times \mathbb{K}$ para $t = 1, 2, 3, ...$ Un elemento $h_t \in H_t$ es de la forma

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

y representa el historial del juego hasta el tiempo t.

Los conjuntos $\mathbb{P}(A)$ y $\mathbb{P}(B)$ representan a las funciones de probabilidad en A y B, respectivamente. Esto es, $\mathbb{P}(A)$ es el conjunto de funciones $\sigma : A \to [0, 1]$ tales que $\sum_{a \in A} \sigma(a) = 1$. $\mathbb{P}(B)$ es definido de forma similar.

Definición 1.2.1. *(a) Una **estrategia** para el jugador 1 es una sucesión $\pi = \{\pi_t\}$ tal que π_t es una función de probabilidad sobre A condicionada por el historial $h_t \in H_t$. Es decir, $\pi_t(\cdot|h_t) \in \mathbb{P}(A),\ \forall h_t \in H_t$.*

Se denota por Π a la familia de estrategias para el jugador 1.

*(b) Una estrategia π para el jugador 1 es **markoviana** si existe una función de probabilidad f_t sobre A tal que $\pi_t(\cdot|h_t) = f_t(\cdot|x_t)$ para cada $t \in \mathbb{N}_0$.*

Se denota por Π_M a la familia de estrategias markovianas para el jugador 1.

*(c) Una estrategia markoviana $\pi = \{f_t\}$ para el jugador 1 es **estacionaria** si $f_t = f$ para toda $t \in \mathbb{N}_0$. En este caso, se denotará a la estrategia estacionaria π por f^∞. Se denota por Π_S al conjunto de estrategias estacionarias del jugador 1.*

Nótese que

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

Observación 1.2.2. *Si $f^\infty(\cdot|\cdot)$ es una estrategia estacionaria para el jugador 1, significa que $f(\cdot|x)$ es una distribución de probabilidad condicional para cada $x \in X$. Para una fácil lectura se denotará $f \in \mathbb{P}(A)$, pues, para cada $x \in X$, se tiene que $f(\cdot|x) \in \mathbb{P}(A)$.*

Los conjuntos de estrategias, estrategias markovianas y de estrategias estacionarias del jugador 2 son definidos de manera similar y denotados por Γ, Γ_M y Γ_S, respectivamente.

Si el juego inicia en el estado $x_0 = x \in X$ y los jugadores siguen el par de estrategias

$(\pi, \gamma) \in \Pi \times \Gamma$, entonces, por el Teorema de Ionescu-Tulcea (véase, [1] o [2]), existe una única medida de probabilidad $P_x^{\pi,\gamma}$. Se denota por $E_x^{\pi,\gamma}$ al operador de esperanza asociado a $P_x^{\pi,\gamma}$.

1.3. Criterio de optimalidad descontada

Definición 1.3.1. *Para cada par de estrategias $(\pi, \gamma) \in \Pi \times \Gamma$ y estado inicial $x_0 = x \in X$, el pago total esperado α-descontado del jugador $i = 1, 2$ se define como*

$$V^i(x, \pi, \gamma) := E_x^{\pi,\gamma}\left[\sum_{t=0}^{\infty} \alpha^t r_i(x_t, a_t, b_t)\right], \quad i = 1, 2 \tag{1.6}$$

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

Note que, si en el juego se obtuvieron los estados $\{x_t\}$ y se eligieron las acciones $\{a_t, b_t\}$, dado que $\mathbb{K}$ es finito, existen M y N tales que $|r_1(x_t, a_t, b_t)| \leq M$ y $|r_2(x_t, a_t, b_t)| \leq N$ para toda $t \in \mathbb{N}_0$. Luego, si $\alpha \in (0, 1)$, se tiene

$$\sum_{t=0}^{\infty} \alpha^t r_1(x_t, a_t, b_t) \leq \sum_{t=0}^{\infty} \alpha^t M = M/(1-\alpha),$$

y

$$\sum_{t=0}^{\infty} \alpha^t r_2(x_t, a_t, b_t) \leq \sum_{t=0}^{\infty} \alpha^t N = N/(1-\alpha),$$

por lo que los pagos esperados α-descontados totales para cada jugador están acotados para cualquier par de estrategias (π, γ).

Se introduce ahora el concepto de equilibrio de Nash, el cual establece una solución para el juego (1.1) bajo un efoque "no cooperativo". Esto significa que los jugadores toman una actitud de no cooperación entre ellos, buscando obtener su mayor ganancia individual.

Definición 1.3.2. *Un par de estrategias $(\pi_*, \gamma_*) \in \Pi \times \Gamma$ es un **equilibrio de Nash** si, para toda $x \in X$,*

$$V^1(x, \pi_*, \gamma_*) \geq V^1(x, \pi, \gamma_*), \quad \forall \pi \in \Pi \tag{1.7}$$

y

$$V^2(x, \pi_*, \gamma_*) \geq V^2(x, \pi_*, \gamma), \quad \forall \gamma \in \Gamma. \tag{1.8}$$

Los pagos en el equilibrio del juego con estado inicial $x \in X$ para los jugadores 1 y 2 son $V^1(x, \pi_*, \gamma_*)$ y $V^2(x, \pi_*, \gamma_*)$, respectivamente.

En términos simples, el concepto de equilibrio de Nash se interpreta como que cada jugador elige su estrategia "óptima" en respuesta a que el otro jugador está jugando con una estrategia que es parte de un equilibrio de Nash. Esto implica que ningún jugador tiene incentivos para desviarse del equilibrio que se está jugando, ya que cualquier cambio individual no le proporcionaría un mayor beneficio, incluso podría recibir un pago menor.

Este capítulo ha permitido comprender el concepto general de los llamados modelos de juegos estocásticos. Bajo el supuesto de que los jugadores deciden no cooperar entre sí, el problema de interés se resume en encontrar equilibrios de Nash para, así, poder dar una solución al juego. Con ello, se sientan las bases para abordar nuevos caminos en la teoría de juegos, como lo son la teoría asintótica y la teoría de los juegos empíricos, mismas en las que se profundiza en el hecho de que, por medio de estimaciones usando distribuciones empíricas, se pueden encontrar soluciones más débiles en cierto sentido respecto a las provenientes de equilibrios de Nash. En los siguientes capítulos se mostrarán estos conceptos para juegos de suma cero y juegos de suma no cero.

Capítulo 2

Estimación y Control en Juegos Suma Cero

Cuando las funciones de pago r_1, r_2 en un juego estocástico como en (1.1) satisfacen que $r_1 + r_2 = 0$, significa que cualquier ganancia obtenida por un jugador representa exactamente la pérdida del otro. Este tipo de juegos se conocen como juegos de suma cero, ya que la suma de los pagos de ambos jugadores es siempre igual a cero en cada resultado posible.

Este capítulo se estudian juegos de suma cero cuyo estado evoluciona siguiendo una ecuación en diferencias como la descrita en (1.3). El objetivo principal es estimar la distribución de probabilidad θ como en (1.4), pues se asume que es desconocida para los jugadores. Usando juegos empíricos se establecen algunos resultados intersantes de aproximación, como el de estimar equilibrios de Nash a través de datos observados.

2.1. Juegos suma cero

En un juego suma-cero donde la evolución del sistema que describe a los estados está dada por una ecuación en diferencias F como en (1.3) el modelo $\mathcal{G}$ en (1.1) toma la forma

$$\mathcal{G}_0 := (X, A, B, F, r), \tag{2.1}$$

donde $r := r_1 = -r_2$.

El problema de juegos se resume a estudiar la función de pago total esperado α-descontado dada por

$$V(x,\pi,\gamma) := E_x^{\pi,\gamma}\left[\sum_{t=0}^{\infty}\alpha^t r(x_t,a_t,b_t)\right], \tag{2.2}$$

con factor de descuento $\alpha \in (0,1)$.

Para cada $x \in X$, los valores inferior y superior del juego se definen respectivamente por:

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

y

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

Definición 2.1.1. *Siguiendo la Definición 1.3.2, en un juego $\mathcal{G}_0$, un par de estrategias $(\pi_*,\gamma_*) \in \Pi\times\Gamma$ es un* ***equilibrio de Nash*** *si*

$$V(x,\pi,\gamma_*) \le V(x,\pi_*,\gamma_*) \le V(x,\pi_*,\gamma) \quad \forall \pi\in\Pi, \gamma\in\Gamma. \tag{2.3}$$

El valor del juego con estado inicial $x \in X$ es $V(x) := V(x,\pi_*,\gamma_*)$.

Nótese que, si $(\pi_*,\gamma_*) \in \Pi\times\Gamma$ es un equilibrio de Nash, entonces se cumple que

$$\sup_{\pi\in\Pi} V(x,\pi,\gamma_*) \le V(x,\pi_*,\gamma_*) \le \inf_{\gamma\in\Gamma} V(x,\pi_*,\gamma),$$

y

$$U(x) = \inf_{\gamma\in\Gamma}\sup_{\pi\in\Pi} V(x,\pi,\gamma) \le V(x) = V(x,\pi_*,\gamma_*) \le \sup_{\pi\in\Pi}\inf_{\gamma\in\Gamma} V(x,\pi,\gamma) = L(x).$$

Pero también se tiene que

$$L(x) \le V(x) \le U(x),$$

por lo que, uniendo las desigualdades, se obtiene que $(\pi_*,\gamma_*) \in \Pi\times\Gamma$ es un equilibrio de Nash si, y solo si

$$V(x) = L(x) = U(x). \tag{2.4}$$

2.2. Criterio de pago descontado

Si Y es un conjunto finito, se define la norma de una función $u : Y \to \mathbb{R}$ como

$$||u|| := \max_{y \in Y} |u(y)|. \tag{2.5}$$

Sean $x \in X$, $\lambda \in \mathbb{P}(A)$, $\mu \in \mathbb{P}(B)$ y $u : \mathbb{K} \to \mathbb{R}$. Se usará la siguiente notación:

$$u(x, \mu, \lambda) := \sum_{a \in A} \sum_{b \in B} u(x, a, b)\lambda(a)\mu(b). \tag{2.6}$$

Similarmente, para F descrita en $\mathcal{G}_0$,

$$F(x, \lambda, \mu, s) := \sum_{a \in A} \sum_{b \in B} F(x, a, b, s)\lambda(a)\mu(b), \quad s \in S. \tag{2.7}$$

A continuación se mostrará la existencia de una función de valor V y de un par de estrategias estacionarias $(\pi_*, \gamma_*) \in \Pi_S \times \Gamma_S$ tales que son un equilibrio de Nash, es decir, que para todo $x \in X$ y todas $(\pi, \gamma) \in \Pi \times \Gamma$,

$$V(x, \pi, \gamma_*) \leq V(x, \pi_*, \gamma_*) = V(x) \leq V(x, \pi_*, \gamma). \tag{2.8}$$

Con este fin, se define el siguiente operador T en la familia de funciones $u : X \to \mathbb{R}$ por

$$Tu(x) := \inf_{\mu \in \mathbb{P}(B)} \sup_{\lambda \in \mathbb{P}(A)} \left[r(x, \lambda, \mu) + \alpha \sum_{s \in S} u[F(x, \lambda, \mu, s)]\theta(s) \right]. \tag{2.9}$$

El operador T es conocido como operador de Shapley y es un operador de contracción con un punto fijo, propiedades que son útiles para asegurar la existencia de equilibrios de Nash, como se enuncia en los siguientes resultados.

Teorema 2.2.1. *En un juego suma-cero $\mathcal{G}_0$ se tiene que*

(a) el valor del juego V existe y es tal que

$$V = TV;$$

(b) existe un par de estrategias estacionarias $(f_*^\infty, g_*^\infty) \in \Pi_S \times \Gamma_S$ tales que $V(x) = V(x, f_*^\infty, g_*^\infty)$.

Lema 2.2.2. *El operador T es de contracción módulo α.*

Demostración. Para $u, v : X \to \mathbb{R}$, se tiene que

$$\begin{aligned}
||Tu - Tv|| &= \max_{x\in X}\left[\inf_{\mu\in\mathbb{P}(B)}\sup_{\lambda\in\mathbb{P}(A)}\left[r(x,\lambda,\mu)+\alpha\sum_{s\in S}u[F(x,\lambda,\mu,s)]\theta(s)\right]\right.\\
&\quad\left.-\inf_{\mu\in\mathbb{P}(B)}\sup_{\lambda\in\mathbb{P}(A)}\left[r(x,\lambda,\mu)+\alpha\sum_{s\in S}u[F(x,\lambda,\mu,s)]\theta(s)\right]\right]\\
&\leq \max_{x\in X}\inf_{\mu\in\mathbb{P}(B)}\sup_{\lambda\in\mathbb{P}(A)}\left|\alpha\sum_{s\in S}\left(u[F(x,\lambda,\mu,s)]-v[F(x,\lambda,\mu,s)]\right)\theta(s)\right|\\
&\leq \alpha\sum_{s\in S}\max_{y\in X}|u(y)-v(y)|\theta(s)\\
&= \alpha||u-v||.
\end{aligned}$$

■

Por el Lema 2.2.2 y por el Teorema del Punto Fijo de Banach, el operador T tiene un único punto fijo $\hat{V}$, i.e.,

$$T\hat{V} = \hat{V}. \tag{2.10}$$

Además, se cumple que, si $n \to \infty$,

$$||T^n u - \hat{V}|| \to 0, \quad \forall u : X \to \mathbb{R}, \tag{2.11}$$

donde $T^n u = T(T^{n-1}u)$, con $n \geq 1$.

Lema 2.2.3. *Para cada función $u : X \to \mathbb{R}$ existen $f \in \mathbb{P}(A)$ y $g \in \mathbb{P}(B)$ tales que, para toda $x \in X$,*

$$\begin{aligned} Tv(x) &= \sup_{\lambda\in\mathbb{P}(A)} \inf_{\mu\in\mathbb{P}(B)} \left[r(x,\lambda,\mu) + \alpha \sum_{s\in S} u[F(x,\lambda,\mu,s)]\theta(s) \right] \\ &= r(x,f,g) + \alpha \sum_{s\in S} u[F(x,f,g,s)]\theta(s) \\ &= \max_{\lambda\in\mathbb{P}(A)} \left[r(x,\lambda,g) + \alpha \sum_{s\in S} u[F(x,\lambda,g,s)]\theta(s) \right] \\ &= \min_{\mu\in\mathbb{P}(B)} \left[r(x,f,\mu) + \alpha \sum_{s\in S} u[F(x,f,\mu,s)]\theta(s) \right]. \end{aligned}$$

La demostración para este lema puede ser consultada en [5, p. 11]. De el lema se sigue que ínf y sup pueden ser intercambiados en el operador T y que el máximo y mínimo se alcanzan en $\mathbb{P}(A)$ y $\mathbb{P}(B)$, respectivamente.

Teorema 2.2.4. *Dado un juego como en (2.5), se tiene que:*

(a) La función $\hat{V}$ en (2.10) es el valor del juego, i.e., $\hat{V}(\cdot) = V(\cdot)$.

(b) El valor V satisface $TV = V$ y existen $(f_*, g_*) \in \mathbb{P}(A) \times \mathbb{P}(B)$, tales que, para toda $x \in X$,

$$\begin{aligned} V(x) &= r(x,f_*,g_*) + \alpha \sum_{s\in S} V[F(x,f_*,g_*,s)]\theta(s) \\ &= \max_{\lambda\in\mathbb{P}(A)} \left[r(x,\lambda,g_*) + \alpha \sum_{s\in S} V[F(x,\lambda,g_*,s)]\theta(s) \right] \\ &= \min_{\mu\in\mathbb{P}(B)} \left[r(x,f_*,\mu) + \alpha \sum_{s\in S} V[F(x,f_*,\mu,s)]\theta(s) \right] \end{aligned}$$

Por lo que $(f_*^\infty, g_*^\infty) \in \Pi_S \times \Gamma_S$ es un equilibrio de Nash.

2.3. Optimalidad asintótica

Por diversas razones, en ocasiones no será posible obtener el equilibrio de Nash para los jugadores, por lo que se pueden buscar estrategias "casi"óptimas. Esto implica analizar el concepto de optimalidad en un sentido más débil. Para introducir este criterio de "optimalidad", se utiliza la llamada función de discrepancia $D : \mathbb{K} \to \mathbb{R}$:

$$D(x,a,b) := r(x,a,b) + \alpha \sum_{X} V(y)Q(y|x,a,b) - V(x), \tag{2.12}$$

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

Nótese que la ecuación (2.10) es equivalente a

$$\sup_{\lambda\in\mathbb{P}(A)}\inf_{\mu\in\mathbb{P}(B)} D(x,\lambda,\mu) = \inf_{\mu\in\mathbb{P}(B)}\sup_{\lambda\in\mathbb{P}(A)} D(x,\lambda,\mu) = 0. \tag{2.13}$$

Definición 2.3.1. *Una estrategia $\pi_* \in \Pi$ es **descontada asintóticamente óptima** (DAO) para el jugador 1 si*

$$\liminf_{t\to\infty} E_x^{\pi_*,\gamma} D(x_t,a_t,b_t) \geq 0 \quad \forall x\in X, \gamma\in\Gamma. \tag{2.14}$$

De manera semejante, $\gamma_* \in \Gamma$ es DAO para el jugador 2 si

$$\limsup_{t\to\infty} E_x^{\pi,\gamma_*} D(x_t,a_t,b_t) \leq 0 \quad \forall x\in X, \pi\in\Pi. \tag{2.15}$$

A un par de estrategias DAO (π_*,γ_*) se le llama par DAO y cumple que, para cada $x\in X$,

$$\lim_{t\to\infty} E_x^{\pi_*,\gamma_*} D(x_t,a_t,b_t) = 0. \tag{2.16}$$

Estas soluciones para el juego de suma cero "más débiles"son estudiadas en la Sección 2.5 de este capítulo, integrándolas con los conceptos presentados en la siguente Sección 2.4.

2.4. Juego empírico

Se asume que los jugadores desconocen la función de probabilidad θ y que las variables aleatorias $\{\xi_t\}$ pueden ser observadas y registradas. Entonces, utilizando la frecuencia de cada valor $s\in S$, se puede aproximar la probabilidad $\theta(s)$ por medio de una sucesión $\{\theta_t\}$. Tomando $\theta_0\in\mathbb{P}(S)$ arbitrario, se define a la distribución empírica al tiempo $t\in\mathbb{N}$ como

$$\theta_t(s) := \frac{1}{t}\sum_{i=0}^{t-1} 1_s(\xi_i), \quad \forall s\in S. \tag{2.17}$$

De la Ley Fuerte de los Grandes Números se sabe que

$$\theta_t(s)\to\theta(s), \quad P-c.s., \tag{2.18}$$

donde $P-c.s.$ significa casi seguramente respecto a la medida de probabilidad P del espacio $(\Omega,\mathcal{F},P)$.

Un juego en el que los jugadores utilicen la distribución empírica es llamado juego empírico. Para cada $t \in \mathbb{N}_0$, el modelo del juego empírico es

$$\mathcal{G}^0_{\theta_t} := (X, A, B, Q_{\theta_t}, r), \tag{2.19}$$

con X, A, B y r definidos como en el modelo (2.1) y con $Q_{\theta_t}(\cdot|\cdot)$ tal que para todo $y \in X$ y $(x, a, b) \in \mathbb{K}$,

$$\begin{aligned} Q_{\theta_t}(y|x,a,b) &= \sum_{s\in S} 1_y[F(x,a,b,s)]\theta_t(s) \\ &= \frac{1}{t}\sum_{i=0}^{t} 1_y[F(x,a,b,\xi_i)]. \end{aligned} \tag{2.20}$$

Para cada $t \in \mathbb{N}_0$, la función de pago esperado α-descontado para el juego $\mathcal{G}^0_{\theta_t}$ es

$$V_{\theta_t}(x,\pi,\gamma) := E^{x,\pi,\gamma}_{\theta_t}\left[\sum_{i=0}^{\infty} \alpha^i r(x_i, a_i, b_i)\right]. \tag{2.21}$$

Los jugadores resuelven cada juego utilizando la distribución empírica θ_t en lugar de la distribución original θ, lo que llevará a un equilibrio de estrategias estacionarias $(f^\infty_t, g^\infty_t) \in \Pi_S \times \Gamma_S$ para el juego $\mathcal{G}^0_{\theta_t}$, para cada $t \in \mathbb{N}_0$. Esto se sigue del Teorema 2.2.1 y se presenta en el siguiente resultado.

Teorema 2.4.1. *Existen $(f^\infty_t, g^\infty_t) \in \Pi_S \times \Gamma_S$ tales que para todo $x \in X$*

$$\begin{aligned} V_{\theta_t}(x) &= r(x, f_t, g_t) + \alpha \textstyle\sum_{s\in S} V_{\theta_t}[F(x, f_t, g_t, s)]\theta_t(s) \\ &= \max_{\lambda\in\mathbb{P}(A)}\left[r(x,\lambda,g_t) + \alpha\sum_{s\in S} V_{\theta_t}[F(x,\lambda,g_t,s)]\theta_t(s)\right] \\ &= \min_{\mu\in\mathbb{P}(B)}\left[r(x,f_t,\mu) + \alpha\sum_{s\in S} V_{\theta_t}[F(x,f_t,\mu,s)]\theta_t(s)\right]. \end{aligned}$$

Esto es, de cada juego empírico $\mathcal{G}^0_{\theta_t}, t \in \mathbb{N}_0$ se obtienen los equilibrios $(f^\infty_t, g^\infty_t) \in \Pi_S \times \Gamma_S$. Para el resto de este capítulo, defínase el par de estrategias

$$(\hat{f}, \hat{g}) := (\{f_t\}, \{g_t\}) \in \Pi_M \times \Gamma_M. \tag{2.22}$$

Observación 2.4.2. *Nótese que, si bien el par (f_t, g_t) proviene del par de estrategias estacionarias (f^∞_t, g^∞_t) para el juego $\mathcal{G}^0_{\theta_t}$, el par de estrategias $(\hat{f}, \hat{g})$ para el juego $\mathcal{G}_0$ en general no es de estrategias estacionarias, pero sí markovianas.*

2.5. Estrategias asintóticamente óptimas

A continuación se utilizarán los conceptos introducidos en la Sección 2.3 y 2.4 para analizar el comportamiento asintótico de $(\hat{f}, \hat{g})$ y mostrar que son un par de estrategias DAO.

Se tiene el siguiente resultado.

Lema 2.5.1. *Si $V(\cdot)$ es la función de valor del juego $\mathcal{G}_0$ y $V_{\theta_t}(\cdot)$ es la función de valor del juego empírico $\mathcal{G}^0_{\theta_t}$, para cada $t \in \mathbb{N}_0$, entonces*

$$||V_{\theta_t} - V|| \to 0 \text{ cuando } t \to \infty, \quad P - c.s. \tag{2.23}$$

Demostración. Por argumentos estándares de Programación Dinámica,

$$V(x) = \min_{\mu \in \mathbb{P}(B)} \max_{\lambda \in \mathbb{P}(A)} \left[r(x, \lambda, \mu) + \alpha \sum_{s \in S} V[F(x, \lambda, \mu, s)]\theta(s) \right] \tag{2.24}$$

y

$$V_{\theta_t}(x) = \min_{\mu \in \mathbb{P}(B)} \max_{\lambda \in \mathbb{P}(A)} \left[r(x, \lambda, \mu) + \alpha \sum_{s \in S} V_{\theta_t}[F(x, \lambda, \mu, s)]\theta_t(s) \right]. \tag{2.25}$$

Por otro lado,

$$\begin{aligned}
V(x) - V_{\theta_t}(x) \leq & \left| \min_{\mu \in \mathbb{P}(B)} \max_{\lambda \in \mathbb{P}(A)} \left[r(x, \lambda, \mu) + \alpha \sum_{s \in S} V[F(x, \lambda, \mu, s)]\theta(s) \right] \right. \\
& \left. - \min_{\mu \in \mathbb{P}(B)} \max_{\lambda \in \mathbb{P}(A)} \left[r(x, \lambda, \mu) + \alpha \sum_{s \in S} V_{\theta_t}[F(x, \lambda, \mu, s)]\theta_t(s) \right] \right| \\
\leq & \max_{\mu \in \mathbb{P}(B), \lambda \in \mathbb{P}(A)} \left| \alpha \sum_{s \in S} V[F(x, \lambda, \mu, s)]\theta(s) - \alpha \sum_{s \in S} V_{\theta_t}[F(x, \lambda, \mu, s)]\theta_t(s) \right| \\
\leq & \max_{\mu \in \mathbb{P}(B), \lambda \in \mathbb{P}(A)} \left| \alpha \sum_{s \in S} V[F(x, \lambda, \mu, s)]\theta(s) - \alpha \sum_{s \in S} V[F(x, \lambda, \mu, s)]\theta_t(s) \right| \\
& + \max_{\mu \in \mathbb{P}(B), \lambda \in \mathbb{P}(A)} \left| \alpha \sum_{s \in S} V[F(x, \lambda, \mu, s)]\theta_t(s) - \alpha \sum_{s \in S} V_{\theta_t}[F(x, \lambda, \mu, s)]\theta_t(s) \right| \\
\leq & \max_{\mu \in \mathbb{P}(B), \lambda \in \mathbb{P}(A)} \alpha \sum_{s \in S} |V[F(x, \lambda, \mu, s)]||\theta(s) - \theta_t(s)| \\
& + \max_{\mu \in \mathbb{P}(B), \lambda \in \mathbb{P}(A)} \alpha \sum_{s \in S} |V[F(x, \lambda, \mu, s)] - V_{\theta_t}[F(x, \lambda, \mu, s)]|\theta_t(s) \\
\leq & \alpha ||V|| \sum_{s \in S} |\theta(s) - \theta_t(s)| + \alpha ||V - V_{\theta_t}||
\end{aligned}$$

es decir,

$$\max_x |V(x) - V_{\theta_t}(x)| \leq \alpha ||V|| \sum_{s \in S} |\theta(s) - \theta_t(s)| + \alpha ||V - V_{\theta_t}||.$$

Así,

$$\begin{aligned} ||V - V_{\theta_t}|| \leq & \frac{\alpha ||V||}{1-\alpha} \sum_{s \in S} |\theta(s) - \theta_t(s)| \\ & \to 0, \quad \text{si } t \to \infty, P - c.s. \end{aligned}$$

■

Teorema 2.5.2. $(\hat{f}, \hat{g})$ *es un par de estrategias DAO.*

Demostración. Se mostrará que, para cada $t \in \mathbb{N}_0$,

$$D(x_t, a_t, g_t) \leq 0, \quad \forall a_t \in A, x_t \in X, \tag{2.26}$$

lo que implica que

$$\limsup_{t \to \infty} E_x^{\pi, \hat{g}} D(x_t, a_t, b_t) \leq 0. \tag{2.27}$$

Es decir, se mostrará que $\hat{g}$ es DAO para el jugador 2.

Primero, se define

$$D_t(x, a_t, b_t) = r(x, a_t, b_t) + \alpha \sum_{s \in S} V_{\theta_t}(F[x, a_t, b_t, s])\theta_t(s) - V_{\theta_t}(x). \tag{2.28}$$

Nótese que, para cada $\lambda \in \mathbb{P}(A)$,

$$\begin{aligned} D_t(x, \lambda, g_t) &= r(x, \lambda, g_t) + \alpha \sum_{s \in S} V_{\theta_t}(F[x, \lambda, g_t, s])\theta_t(s) - V_{\theta_t}(x) \\ &\leq \max_{\sigma \in \mathbb{P}(A)} \left[r(x, \sigma, g_t) + \alpha \sum_{s \in S} V_{\theta_t}(F[x, \sigma, g_t, s])\theta_t(s) \right] - V_{\theta_t}(x) \\ &= 0. \end{aligned}$$

Por lo que, para toda $\sigma \in \mathbb{P}(A)$,

$$\begin{aligned} D(x_t, \sigma, g_t) &\leq |D(x_t, \sigma, g_t) - D_t(x_t, \sigma, g_t)| \\ &\leq \sup_{x \in X, \lambda \in \mathbb{P}(B)} |D(x, \lambda, g_t) - D_t(x, \lambda, g_t)|. \end{aligned}$$

Ahora,

$$
\begin{aligned}
&|D(x,\lambda,g_t) - D_t(x,\lambda,g_t)| \\
&\qquad = \Big| r(x,\lambda,g_t) + \alpha \sum_{s\in S} V(F[x,\lambda,g_t,s])\theta(s) - V(x) - r(x,\lambda,g_t) \\
&\qquad\quad - \alpha \sum_{s\in S} V_{\theta_t}(F[x,\lambda,g_t,s])\theta_t(s) + V_{\theta_t}(x) \Big| \\
&\qquad \leq |V_{\theta_t}(x) - V(x)| + \left| \alpha \sum_{s\in S} V(F[x,\lambda,g_t,s])\theta(s) - \alpha \sum_{s\in S} V_{\theta_t}(F[x,\lambda,g_t,s])\theta_t(s) \right| \\
&\qquad \leq |V_{\theta_t}(x) - V(x)| + \left| \alpha \sum_{s\in S} V(F[x,\lambda,g_t,s])\theta(s) - \alpha \sum_{s\in S} V(F[x,\lambda,g_t,s])\theta_t(s) \right| \\
&\qquad\quad + \left| \alpha \sum_{s\in S} V_{\theta_t}(F[x,\lambda,g_t,s])\theta_t(s) + \alpha \sum_{s\in S} V(F[x,\lambda,g_t,s])\theta_t(s) \right| \\
&\qquad \leq (1+\alpha)||V_{\theta_t} - V|| + \alpha||V|| \sum_{s\in S} |\theta(s) - \theta_t(s)|
\end{aligned}
$$

Del Lema 2.5.1 y dado que $\theta_t(s) \to \theta(s), \quad P - c.s.$, se sigue que

$$
|D(x,\lambda,g_t) - D_t(x,\lambda,g_t)| \leq (1-\alpha)||V_{\theta_t} - V|| + \alpha||V|| \sum_{s\in S} |\theta(s) - \theta_t(s)| \to 0,
$$

cuando $t \to \infty$, P-c.s. y entonces

$$
\limsup_{t\to\infty} D(x_t, \sigma, g_t) \leq 0, \quad P - c.s.
$$

como se buscaba.

La prueba para la estrategia $\hat{f}$ del jugador 1 es análoga y de ahí se sigue que el par $(\hat{f}, \hat{g})$ es DAO, lo que concluye la prueba. ■

Durante este capítulo se ha analizado brevemente lo siguiente: primero, se estudiaron algunos resultados relacionados con la existencia de equilibrios de Nash. Posteriormente, se introdujo el concepto de soluciones DAO, seguido de la definición de juego empírico. Esto ha permitido establecer dos resultados muy relevantes: el Lema 2.5.1 y el Teorema 2.5.2, los cuales afirman que las soluciones obtenidas en los juegos empíricos implican el refinamiento de un par de estrategias que son DAO para el juego de suma cero original. De esta manera, se establece un punto de partida para el siguiente capítulo, donde se

desea responder si es posible realizar nuevamente este análisis, pero ahora para juegos estocásticos de suma no cero.

Capítulo 3

Estimación y control en juegos suma no cero

3.1. Criterio de pago descontado

En este capítulo, considérese el modelo $\mathcal{G}$ como en (1.1), bajo el criterio que los estados evolucionan a través de un sistema como en (1.3). Esto es,

$$\tilde{\mathcal{G}}_\theta := (X, A, B, S, F, \theta, r_1, r_2). \tag{3.1}$$

1. X es el espacio de estados, se supone finito.

2. A es el conjunto de acciones del jugador 1, se supone finito. B es el conjunto de acciones del jugador 2, se supone finito.

3. Sea $\{\xi_t\}$ una sucesión de variables aleatorias i.i.d. que están definidas en el espacio de probabilidad $(\Omega, \mathcal{F}, P)$ y toman valores en el conjunto finito S. La función $F : \mathbb{K} \times S \to X$ describe la evolución de los estados del juego mediante la ecuación en diferencias

$$x_{t+1} = F(x_t, a_t, b_t, \xi_t), \quad t \in \mathbb{N}_0. \tag{3.2}$$

4. Se denota por θ la distribución de probabilidad de ξ_t. Es decir,

$$\theta(s) = P[\xi_t = s], \quad \forall s \in S, t \in \mathbb{N}_0. \tag{3.3}$$

Se define la ley de transición entre los estados

$$Q(y|x,a,b) = P[x_{t+1} = y | x_t = x, a_t = a, b_t = b] = \sum_{s \in S} 1_y[F(x,a,b,s)]\theta(s), \quad (3.4)$$

para cada $y \in X$.

5. $r_1 : \mathbb{K} \to \mathbb{R}$, $r_2 : \mathbb{K} \to \mathbb{R}$ son las funciones de pago del jugador 1 y jugador 2, respectivamente.

El pago α-descontado para el jugador $i = 1, 2$ es:

$$V^i(x, \pi, \gamma) = E_x^{\pi,\gamma}\left[\sum_{t=1}^{\infty} \alpha^t r_i(x_t, a_t, b_t)\right]. \quad (3.5)$$

Teorema 3.1.1. *El modelo $\tilde{\mathcal{G}}_\theta$ con pagos descontados V^1 y V^2 tiene un equilibrio de Nash en el espacio de las estrategias estacionarias. Es decir, existe un par $(f^\infty, g^\infty) \in \Pi_S \times \Gamma_S$ tal que, para cada $x \in X$,*

$$V^1(x, f^\infty, g^\infty) \geq V^1(x, \pi, g^\infty), \quad \forall \pi \in \Pi,$$

y

$$V^2(x, f^\infty, g^\infty) \geq V^2(x, f^\infty, \gamma), \quad \forall \gamma \in \Gamma.$$

La demostración puede consultarse en [8].

Las funciones de valor en el equilibrio (f^∞, g^∞) se definen como:

$$V^1(x) := \max_{\pi \in \Pi} V^1(x, \pi, g^\infty) \quad (3.6)$$

y

$$V^2(x) := \max_{\gamma \in \Gamma} V^2(x, f^\infty, \gamma) \quad (3.7)$$

y, por Programación Dinámica satisfacen para todo $x \in X$ las siguientes ecuaciones:

$$\begin{aligned} V^1(x) &= \max_{\lambda \in \mathbb{P}(A)} \left[r_1(x, \lambda, g) + \alpha \sum_{s \in S} V^1[F(x, \lambda, g, s)]\theta(s) \right] \\ &= r_1(x, f, g) + \alpha \sum_{s \in S} V^1[F(x, \lambda, g, s)]\theta(s) \end{aligned} \quad (3.8)$$

$$\begin{aligned} V^2(x) &= \max_{\mu\in\mathbb{P}(B)}\left[r_2(x,f,\mu)+\alpha\sum_{s\in S}V^2[F(x,f,\mu,s]\theta(s)\right] \\ &= r_2(x,f,g)+\alpha\sum_{s\in S}V^2[F(x,f,g,s]\theta(s). \end{aligned} \tag{3.9}$$

Al igual que en el Capítulo 2, el interés radica en el concepto de juego empírico, el cual es introducido en la siguiente sección.

3.2. Juego empírico

De forma semejante a la Sección 2.4, se asume que la función de probabilidad θ es desconocida para los jugadores y que $\{\xi_t\}$ pueden observarse y registrarse. Así, la frecuencia de cada valor $s \in S$ sirve como aproximación para la probabilidad $\theta(s)$.

Para cada $t \in \mathbb{N}$ las **probabilidades empíricas** son

$$\theta_t(s) := \frac{1}{t}\sum_{i=0}^{t-1} 1_s(\xi_i), \quad s \in S, \tag{3.10}$$

con θ_0 arbitrario.

En cada tiempo t, se observa ξ_t y los jugadores utilizan la estimación θ_t para seleccionar sus acciones en función de θ_t. Nuevamente, la ley fuerte de los grandes números, implica que

$$\theta_t(s) \to \theta(s), \quad P-c.s.$$

Por otro lado, para cualquier función $v : S \to \mathbb{R}$, también se tiene que

$$\sum_{s\in S} v(s)\theta_t(s) = \frac{1}{t}\sum_{t=0}^{k-1} v(\xi_t) \to \sum_{s\in S} v(s)\theta(s), \quad P-c.s.$$

Aún más, para cualquier función $u : X \to \mathbb{R}$ y cada $(x,\lambda,\mu) \in X \times \mathbb{P}(A) \times \mathbb{P}(B)$,

$$\left|\sum_{s\in S} u[F(x,\lambda,\mu,s)]\theta_t(s) - \sum_{s\in S} u[F(x,\lambda,\mu,s)]\theta(s)\right| \to 0, \quad P-a.s. \tag{3.11}$$

Se introduce ahora, para cada tiempo $t \in \mathbb{N}_0$ el **modelo del juego empírico**,

$$\tilde{\mathcal{G}}_{\theta_t} := (X, A, B, S, F, \theta_t, r_1, r_2), \tag{3.12}$$

con θ_t definida como en (3.10).

Para cada $t \in \mathbb{N}_0$, la función de pago esperado α-descontado para el jugador 1 en el juego $\tilde{\mathcal{G}}_{\theta_t}$ es

$$V^1_{\theta_t}(x, \pi, \gamma) := E^{x,\pi,\gamma}_{\theta_t}\left[\sum_{i=0}^{\infty} \alpha^i r_1(x_i, a_i, b_i)\right].$$

La función $V^2_{\theta_t}$ de pago esperado α-descontado para el jugador 2 es definida de manera similar.

Del Teorema 3.1.1, se sigue que existe un equilibrio de Nash estacionario $(f_t^\infty, g_t^\infty) \in \Pi_S \times \Gamma_S$ para cada modelo $\tilde{\mathcal{G}}_{\theta_t}$, $t \in \mathbb{N}_0$. Es decir, para cada $x \in X$ y t en $\mathbb{N}_0$

$$V^1_{\theta_t}(x, f_t^\infty, g_t^\infty) \geq V^1_{\theta_t}(x, \pi, g_t^\infty), \quad \forall \pi \in \Pi$$

y

$$V^2_{\theta_t}(x, f_t^\infty, g_t^\infty) \geq V^2_{\theta_t}(x, f_t^\infty, \gamma), \quad \forall \gamma \in \Gamma.$$

Las funciones de valor correspondientes satisfacen

$$\begin{aligned} V^1_{\theta_t}(x) &= \max_{\lambda \in \mathbb{P}(A)}\left[r_1(x, \lambda, g_t) + \alpha \sum_{s \in S} V^1_{\theta_t}[F(x, \lambda, g_t, s)]\theta_t(s)\right] \\ &= r_1(x, f_t, g_t) + \alpha \sum_{s \in S} V^1_{\theta_t}[F(x, f_t, g_t, s)]\theta_t(s), \end{aligned} \tag{3.13}$$

$$\begin{aligned} V^2_{\theta_t}(x) &= \max_{\mu \in \mathbb{P}(B)}\left[r_2(x, f_t, \mu) + \alpha \sum_{s \in S} V^2_{\theta_t}[F(x, f_t, \mu, s)]\theta_t(s)\right] \\ &= r_2(x, f_t, g_t) + \alpha \sum_{s \in S} V^2_{\theta_t}[F(x, f_t, g_t, s)]\theta_t(s). \end{aligned} \tag{3.14}$$

A continuación, se analizará la optimalidad del par $(\hat{f}, \hat{g}) := (\{f_t\}, \{g_t\})$, en el juego original $\tilde{\mathcal{G}}_\theta$ cuando $t \to \infty$ observando su comportamiento asintótico.

Observación 3.2.1. *Similarmente al caso en suma-cero, si bien el par (f_t, g_t) proviene del par de estrategias estacionarias (f_t^∞, g_t^∞) para el juego $\tilde{\mathcal{G}}_{\theta_t}$, el par $(\hat{f}, \hat{g})$ de estrategias para el juego $\tilde{\mathcal{G}}^0$ en general no es de estrategias estacionarias, pero sí markovianas.*

3.3. Equilibrio de Nash asintótico

Teorema 3.3.1. *Para cada $t \in \mathbb{N}$, sea (f_t^∞, g_t^∞) un equilibrio de Nash estacionario para el juego empírico $\tilde{\mathcal{G}}_{\theta_t}$. Se asume que existen $(f_*, g_*) : X \times \Omega \to \mathbb{P}(A) \times \mathbb{P}(B)$ tales que*

$$\lim_{t\to\infty}(f_t, g_t) = (f_*, g_*). \tag{3.15}$$

Entonces (f_*^∞, g_*^∞) es un equilibrio de Nash estacionario P-c.s.para el juego $\tilde{\mathcal{G}}_\theta$.

Demostración. Si (f_t^∞, g_t^∞) es un equilibrio de Nash estacionario para $\tilde{\mathcal{G}}_{\theta_t}$, $t \in \mathbb{N}_0$, entonces existen funciones $V^1_{\theta_t}$ y $V^2_{\theta_t}$ que cumplen (3.13), (3.14). Primero, supóngase que

$$\lim_{t\to\infty}(V^1_{\theta_t}, V^2_{\theta_t}) = (V^1, V^2), \tag{3.16}$$

para algunas $V^1, V^2 : X \times \Omega \to \mathbb{R}$. Para cada $\omega \in \Omega$ fijo y $\lambda \in \mathbb{P}(A)$, se afirma que

$$\lim_{t\to\infty}\sum_{s\in S}|V^1_{\theta_t}[F(x,\lambda,g_t,s)] - V^1[F(x,\lambda,g_t,s)]|\theta_t(s) = 0, \quad x \in X, \tag{3.17}$$

y

$$\lim_{t\to\infty}\sum_{s\in S}|V^1[F(x,\lambda,g_t,s)] - V^1[F(x,\lambda,g_*,s)]|\theta_t(s) = 0, \quad x \in X. \tag{3.18}$$

En efecto, nótese que

$$\begin{aligned}\sum_{s\in S}|V^1_{\theta_t}[F(x,\lambda,g_t,s)] - V^1[F(x,\lambda,g_t,s)]|\theta_t(s)& \\ \leq \sum_{s\in S}\max_{x\in X}|V^1_{\theta_t}(x) - V^1(x)|\theta_t(s)& \\ = ||V^1_{\theta_t} - V^1||.&\end{aligned}$$

Como $V^1_{\theta_t} \to V^1$, entonces (3.17) es cierta. Similarmente,

$$\begin{aligned}\sum_{s\in S}|V^1[F(x,\lambda,g_t,s)] - V^1[F(x,\lambda,g_*,s)]|\theta_t(s)& \\ \leq \sum_{s\in S}\sum_{b\in B}|V^1[F(x,\lambda,b,s)]||g_t(b|x) - g_*(b|x)|\theta_t(s)& \\ \leq ||V^1||\sum_{b\in B}|g_t(b|x) - g_*(b|x)|&\end{aligned}$$

Por lo que (3.18) también se cumple. Aún más, por la desigualdad del triángulo,

$$\begin{aligned}\sum_{s\in S} |V^1_{\theta_t}[F(x,\lambda,g_t,s)]\theta_t(s) - V^1[F(x,\lambda,g_*,s)]\theta(s)| \\ \leq \sum_{s\in S} \big[|V^1_{\theta_t}[F(x,\lambda,g_t,s)]\theta_t(s) - V^1[F(x,\lambda,g_t,s)]\theta_t(s)| \\ + |V^1[F(x,\lambda,g_t,s)]\theta_t(s) - V^1[F(x,\lambda,g_*,s)]\theta_t(s)| \\ + |V^1[F(x,\lambda,g_*,s)]\theta_t(s) - V^1[F(x,\lambda,g_*,s)]\theta(s)|\big]\end{aligned}$$

Entonces, por (3.11)-(3.17),

$$\lim_{t\to\infty} \sum_{s\in S} V^1_{\theta_t}[F(x,\lambda,g_t,s)]\theta_t(s) = \sum_{s\in S} V^1[F(x,\lambda,g_*,s)]\theta(s) \quad P-c.s. \tag{3.19}$$

Por otra parte, por (3.13),

$$V^1_{\theta_t}(x) \geq r_1(x,\lambda,g_t) + \alpha \sum_{s\in S} V^1_{\theta_t}[F(x,\lambda,g_t,s)]\theta_t(s) \quad \forall \lambda \in \mathbb{P}(A),$$

para cada $t \in \mathbb{N}_0$. Entonces, si $t \to \infty$ y usando (3.19),

$$V^1(x) \geq r_1(x,\lambda,g_*) + \alpha \sum_{s\in S} V^1[F(x,\lambda,g_*,s)]\theta(s), \quad \lambda \in \mathbb{P}(A) \quad P-c.s. \tag{3.20}$$

Con los cambios apropiados en la demostración de (3.19), se sigue que

$$\lim_{t\to\infty} \sum_{s\in S} V^1_{\theta_t}[F(x,f_t,g_t,s)]\theta_t(s) = \sum_{s\in S} V^1[F(x,f_*,g_*,s)]\theta(s) \quad P-c.s.$$

Luego, si $t \to \infty$ en (3.13) y usando (3.20), se tiene

$$V^1(x) = \max_{\lambda\in\mathbb{P}(A)} \left[r_1(x,\lambda,g_*) + \alpha \sum_{s\in S} V^1[F(x,\lambda,g_*,s)]\theta(s) \right] \tag{3.21}$$

$$= r_1(x,f_*,g_*) + \alpha \sum V^1[F(x,f_*,g_*,s)]\theta(s) \quad P-c.s. \tag{3.22}$$

Se puede mostrar de forma similar que

$$V^2(x) = \max_{\mu\in\mathbb{P}(B)}\left[r_2(x,f_*,\mu) + \alpha\sum_{s\in S} V^2[F(x,f_*,\mu,s)]\theta(s)\right] \tag{3.23}$$

$$= r_2(x,f_*,g_*) + \alpha\sum V^2[F(x,f_*,g_*,s)]\theta(s) \quad P-c.s. \tag{3.24}$$

Entonces (f_*^∞, g_*^∞) es un equilibrio de Nash estacionario P-c.s. cuando (3.16) se cumple.

Se considera ahora a una sucesión $\{(V^1_{\theta_t}, V^2_{\theta_t})\}$ que no sea convergente, pero que satisfaga (3.13)-(3.14). Con $x \in X$ y $\omega \in \Omega$ fijos, es fácil mostrar que, para alguna constante M,

$$|V^1_{\theta_t}(x,\omega)| \le M, \quad |V^2_{\theta_t}(x,\omega)| \le M \quad \forall t \in \mathbb{N}.$$

Por lo que hay una subsucesión convergente $\{(V^1_{t_l}(x,\omega), V^2_{t_l}(x,\omega))\}$ de $\{(V^1_{\theta_t}(x,\omega), V^2_{\theta_t}(x,\omega))\}$. Además, con los mismos índices, $(f_{t_l}, g_{t_l}) \to (f_*, g_*)$. Repitiendo la prueba anterior con $\{(f_{t_l}, g_{t_l})\}$ y $\{(V^1_{t_l}, V^2_{t_l})\}$ se muestra que (f_*^∞, g_*^∞) es un equilibrio de Nash estacionario P-c.s., lo que completa la prueba del teorema. ■

Considere el par de estrategias

$$(\hat{f}, \hat{g}) = (\{f_t\}, \{g_t\}) \in \Pi_M \times \Gamma_M, \tag{3.25}$$

tomado de (f_t^∞, g_t^∞), que es un equilibrio de Nash para $\tilde{\mathcal{G}}_{\theta_t}$.

Sea $(f_*, g_*) \in \Pi_S \times \Gamma_S$ un equilibrio de Nash estacionario de $\tilde{\mathcal{G}}_\theta$ y $V^1(x) := V^1_\theta(x, f_*, g_*)$, $V^2(x) := V^2_\theta(x, f_*, g_*)$ los pagos de equilibrio correspondientes.

Se definen las **funciones de discrepancia** $D^1, D^2 : \mathbb{K} \to \mathbb{R}$ para los jugadores 1 y 2, respectivamente:

$$D^1(x,a,b) := V^1(x) - \left[r_1(x,a,b) + \alpha\sum_{s\in S} V^1[F(x,a,b,s)]\theta(s)\right] \tag{3.26}$$

y

$$D^2(x,a,b) := V^2(x) - \left[r_2(x,a,b) + \alpha\sum_{s\in S} V^2[F(x,a,b,s)]\theta(s)\right] \tag{3.27}$$

Entonces las ecuaciones (3.21) y (3.23) son equivalentes a las relaciones

$$0 = D^1(x,f_*,g_*) \le D^1(x,\lambda,g_*), \quad \forall\lambda \in \mathbb{P}(A), x \in X,$$

y

$$0 = D^2(x, f_*, g_*) \leq D^2(x, f_*, \mu), \quad \forall \mu \in \mathbb{P}(B), x \in X.$$

Aún más, si las estrategias en (3.25) satisfacen (3.15), se tiene

$$\lim_{t\to\infty} D^1(x, f_t, g_t) = 0 \tag{3.28}$$

y

$$\lim_{t\to\infty} D^2(x, f_t, g_t) = 0 \tag{3.29}$$

para cada $x \in X$. Estas igualdades pueden no cumplirse si no se satisface (3.15), pero se espera que puedan ser ciertas para una subsucesión de $(\{f_t\}, \{g_t\})$. Esto motiva la siguiente definición.

Definición 3.3.2. *Un par de estrategias markovianas $(\pi, \gamma) = (\{f_t\}, \{g_t\}) \in \Pi_M \times \Gamma_M$ se dice ser un equilibrio de Nash asintótico (ENA) para el juego $\tilde{\mathcal{G}}_\theta$, con pagos como en (3.5), si existe un equilibrio de Nash en estrategias estacionarias $(f_*^\infty, g_*^\infty) \in \Pi_S \times \Gamma_S$ tal que*

$$\liminf_{t\to\infty} |D^1(x, f_t, g_t)| = 0$$

y

$$\liminf_{t\to\infty} |D^2(x, f_t, g_t)| = 0,$$

donde D^1 y D^2 son las funciones de discrepancia relacionadas a $V^1(x) = V^1(x, f_*^\infty, g_*^\infty)$ y $V^2(x) = V^2(x, f_*^\infty, g_*^\infty)$, respectivamente.

Teorema 3.3.3. *Para cada $t \in \mathbb{N}_0$, sea (f_t^∞, g_t^∞) un equilibrio de Nash estacionario del juego empírico $\tilde{\mathcal{G}}_{\theta_t}$. Entonces $(\tilde{\pi}, \tilde{\gamma}) = (\{f_t\}, \{g_t\})$ es un equilibrio de Nash asintótico P-c.s. para el juego $\tilde{\mathcal{G}}_\theta$*

Demostración. Se fija $\omega \in \Omega$. Dado que $\mathbb{P}(A) \times \mathbb{P}(B)$ es compacto, existe el par

$$(f_*(\cdot|x), g_*(\cdot|x)) \in \mathbb{P}(A) \times \mathbb{P}(B)$$

que es punto de acumulación de la sucesión $\{(f_t(\cdot|x), g_t(\cdot|x))\}$ para cada $x \in X$. Luego, para cada $\omega \in \Omega$, hay una subsucesión $\{f_{t_m}, g_{t_m}\}$ tal que

$$\lim_{m\to\infty} (f_{t_m}, g_{t_m}) = (f_*, g_*). \tag{3.30}$$

Del Teorema 3.3.1, (f_*^∞, g_*^∞) es un equilibrio de Nash estacionario para $\tilde{\mathcal{G}}_\theta$ P-c.s. Si

$V^1(x) := V^1(x, f_*^\infty, g_*^\infty)$ y $V^2(x) := V^2(x, f_*^\infty, g_*^\infty)$ para cada $x \in X$, entonces se tiene

$$V^1(x) = r_1(x, f_*, g_*) + \alpha \sum_{s \in S} V^1[F(x, f_*, g_*, s)]\theta(s), \quad P - c.s.$$

y

$$V^2(x) = r_2(x, f_*, g_*) + \alpha \sum_{s \in S} V^2[F(x, f_*, g_*, s)]\theta(s), \quad P - c.s.$$

Nótese que

$$\lim_{m \to \infty} \sum_{s \in S} V^1[F(x, f_{t_m}, g_{t_m}, s)]\theta(s) = \sum_{s \in S} V^1[F(x, f_*, g_*, s)]\theta(s)$$

y

$$\lim_{m \to \infty} \sum_{s \in S} V^2[F(x, f_{t_m}, g_{t_m}, s)]\theta(s) = \sum_{s \in S} V^2[F(x, f_*, g_*, s)]\theta(s).$$

Entonces,

$$\lim_{m \to \infty} D^1(x, f_{t_m}, g_{t_m}) = \lim_{m \to \infty} D^2(x, f_{t_m}, g_{t_m}) = 0, \quad P - c.s.$$

lo que demuestra lo que se buscaba.

■

En el presente capítulo se ha llevado a cabo el análisis de la estimación asintótica para juegos estocásticos finitos de suma no cero. Se ha profundizado en cómo, usando los conceptos de juego empírico, pueden obtenerse, bajo ciertas hipótesis, equilibrios de Nash asintóticos (ENA) para el juego original. Esto ha permitido abordar la problemática de encontrar equilibrios de Nash para el juego, aunque en un sentido más débil, usando observaciones de las variables aleatorias involucradas. Este recorrido conduce de manera natural a la búsqueda de modelos que ejemplifiquen los resultados teóricos obtenidos. Con dicho objetivo en mente, en el siguiente capítulo se muestran un par de ejemplos, que incluyen algunas simulaciones.

Capítulo 4

Ejemplo de juego suma no cero

4.1. Gran guerra por los pescados (Great fish war)

A continuación se presenta un ejemplo de un juego de suma no cero.

Considérese una versión de "the Great fish war", introducido en [4], en la que dos países compiten por pesca en una misma zona. La población de peces crece según una ecuación en diferencia inferida de leyes naturales, pero afectada también por las acciones de los participantes, cada uno de los cuales hace sus estrategias tomando en cuenta las acciones del otro, obteniendo después de cada etapa cierta utilidad en función a la cantidad de peces pescados. Se asume que los dos participantes actúan en un duopolio y buscan maximizar sus utilidades descontadas.

Sean $X = \{0.0, 0.1, ..., 0.9, 1.0\}$, $A = B = \{0.0, 0.1, 0.2, 0.3\}$ y $S = \{0.5, 0.6, ..., 1.4, 1.5\}$. La variable aleatoria ξ toma valores en S con probabilidad $\theta(s), \quad s \in S$. El sistema evoluciona según la función

$$F(x,a,b,\xi) = \begin{cases} y & \text{si} \quad y \in X \quad \text{y} \quad y - 0.05 \leq \xi[\text{máx}(x-a-b,0)]^{\alpha} < y + 0.05, \\ 1 & \text{si} \quad 1.05 \leq \xi[\text{máx}(x-a-b,0)]^{\alpha} \end{cases} . \quad (4.1)$$

Con $0 < \alpha < 1$. Las funciones de pago por etapa son

$$r_1(x,a,b) = \begin{cases} \sqrt{a} & \text{si} \quad a+b \leq x, \\ \sqrt{\frac{a}{a+b}x} & \text{si} \quad a+b > x \end{cases} \quad (4.2)$$

y

$$r_2(x,a,b) = \begin{cases} \sqrt{b} & \text{si} \quad a+b \leq x, \\ \sqrt{\frac{b}{a+b}x} & \text{si} \quad a+b > x \end{cases} \tag{4.3}$$

para el jugador 1 y 2, respectivamente.

Para fines ilustrativos se supone que ξ toma valores en S con probabilidades de una distribución binomial de parámetros $(n,p) = (10, 0.3)$, $\alpha = 0.7$ y $\beta = 0.85$. Se hacen los cálculos de un equilibrio de Nash simétrico con estrategias estacionarios y luego se simulan los juegos empíricos correspondientes con el fin de encontrar un ENA.

Se utiliza una versión modificada de un programa en Python en [9] en el que se estima la distribución y se calculan estrategias de equilibrio para cada juego $\mathcal{G}_{\theta_m}$ para un solo jugador y son enviadas a un archivo de texto. También, se utiliza otro código de Python creado por el autor y mostrado a continuación, para leer, limpiar y ordenar los datos obtenidos del archivo de texto y para graficar a la evolución de las estrategias obtenidas en los juegos empíricos.

```
import numpy as np
import matplotlib.pyplot as plt

#para leer los datos de un archivo .txt
def leer_txt(file):
    with open(file, 'r') as file:
        lst = [float(num) for num in file.read().split(',') if num.strip
    ()]
    return lst

#para limpiar y organizar los datos
def organize(lst):
    final = [np.zeros((4, 20)) for _ in range(11)]
    lst = [elem for i, elem in enumerate(lst) if i % 56 != 0]
    lst = [elem for i, elem in enumerate(lst, start=1) if i % 5 != 0]
    lst = [lst[i:i + 44][::-1] for i in range(0, len(lst), 44)]
    lst = [item for sublist in lst for item in sublist]
    lst_ = lst
    for a in range(11):
        for b in range(4):
            for c in range(20):
                final[a][b][c] = lst_[44*c + 4*a + b]
    return final

```

```
file = "Fishery_Game.txt"
lista_ = leer_txt(file)
lista = organize(lista_)

def graficar_matriz(M, estado):
    M = np.array(M)
    num_filas, num_columnas = M.shape
    x = np.arange(num_columnas)

    plt.figure(figsize=(10, 6))

    markers = ['o', 's', '^', 'd']

    # Graficar cada fila de la matriz como una línea con diferente marcador
    for i in range(num_filas):
        plt.plot(x, M[i], marker=markers[i % len(markers)], label=f"acción 0.{i}")


    plt.xlabel("m")
    plt.ylabel("Estrategias")
    plt.title(f"estado {estado}")

    # Modificar etiquetas del eje x
    etiquetas_x = []
    exponent = 0
    for i in range(num_columnas):
        if i % 2 == 0:
            etiquetas_x.append(f"1e+{exponent}")
        else:
            etiquetas_x.append(f"5e+{exponent}")
            exponent += 1

    etiquetas_x[-1] = "Equilibrio"  # Reemplazar la última etiqueta
    plt.xticks(x, etiquetas_x, rotation=-45)

    plt.legend()
    plt.grid(True)
    plt.show()

estados = [0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0]
for i in range(11):
    graficar_matriz(lista[i], estados[i])
```

Observación 4.1.1. *A pesar de que el programa obtenido de [9] calcula solamente las estrategias de equilibrio de un jugador, debido a que el juego es simétrico (i.e. los conjuntos de acción A y B son iguales y r_1, r_2 son simétricas), las estrategias de equilibrio correspondientes al otro jugador deben ser las mismas.*

A continuación se presentan las gráficas con los resultados de las estrategias obtenidas con estimaciones para juegos empíricos $\mathcal{G}_{\theta_m}$. Cada una de ellas muestra para cada estado, en el eje y, las probabilidades con las que se elige cada acción mientras m aumenta. Se agrega, también, en la última columna, llamada "Equilibrio", la estrategia de equilibrio del juego original $\mathcal{G}_\theta$. Nótese que, como lo indica la teoría, las estrategias estimadas convergen a las reales.

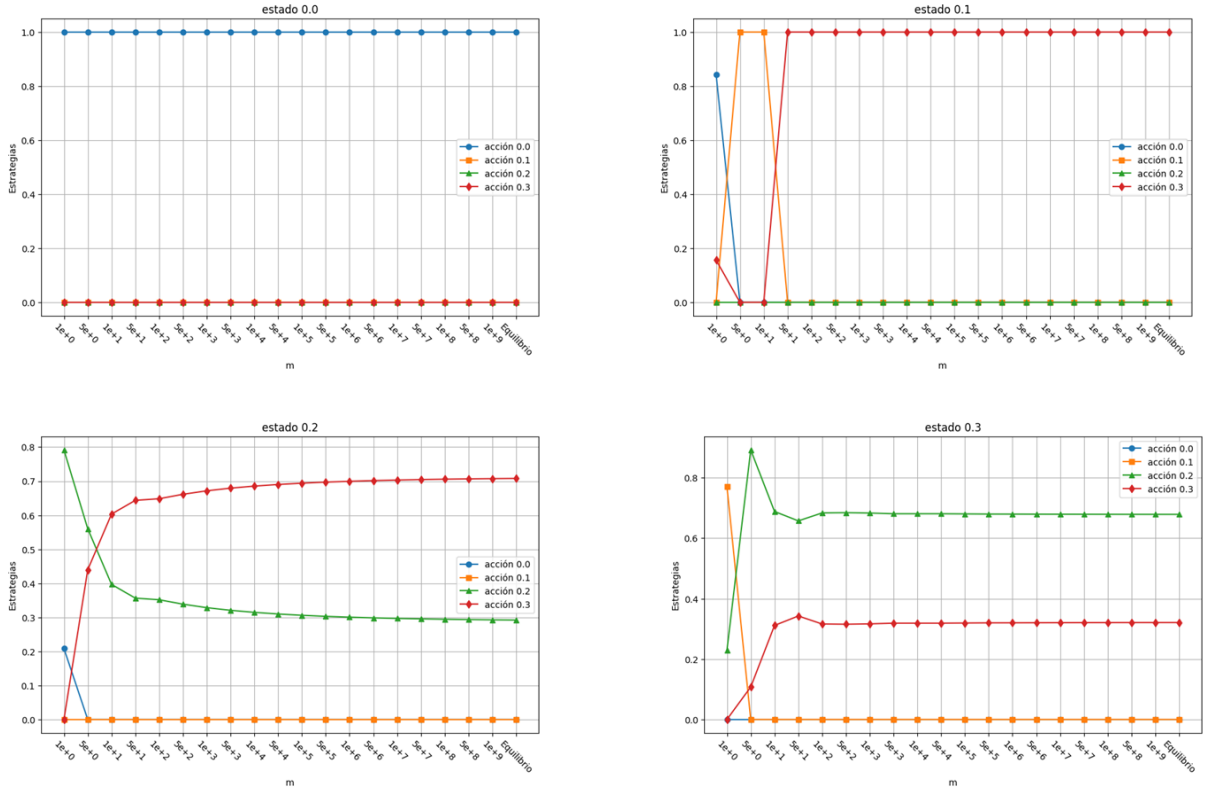

FIGURA 4.1: Estrategias de equilibrio para juegos $\mathcal{G}_{\theta_m}$ simulados. El eje x representa distintos valores de m, mientras que el y representa probabilidades para las acciones. Cada gráfica muestra a estados 0.0, 0.1, 0.2 y 0.3 del juego, respectivamente.

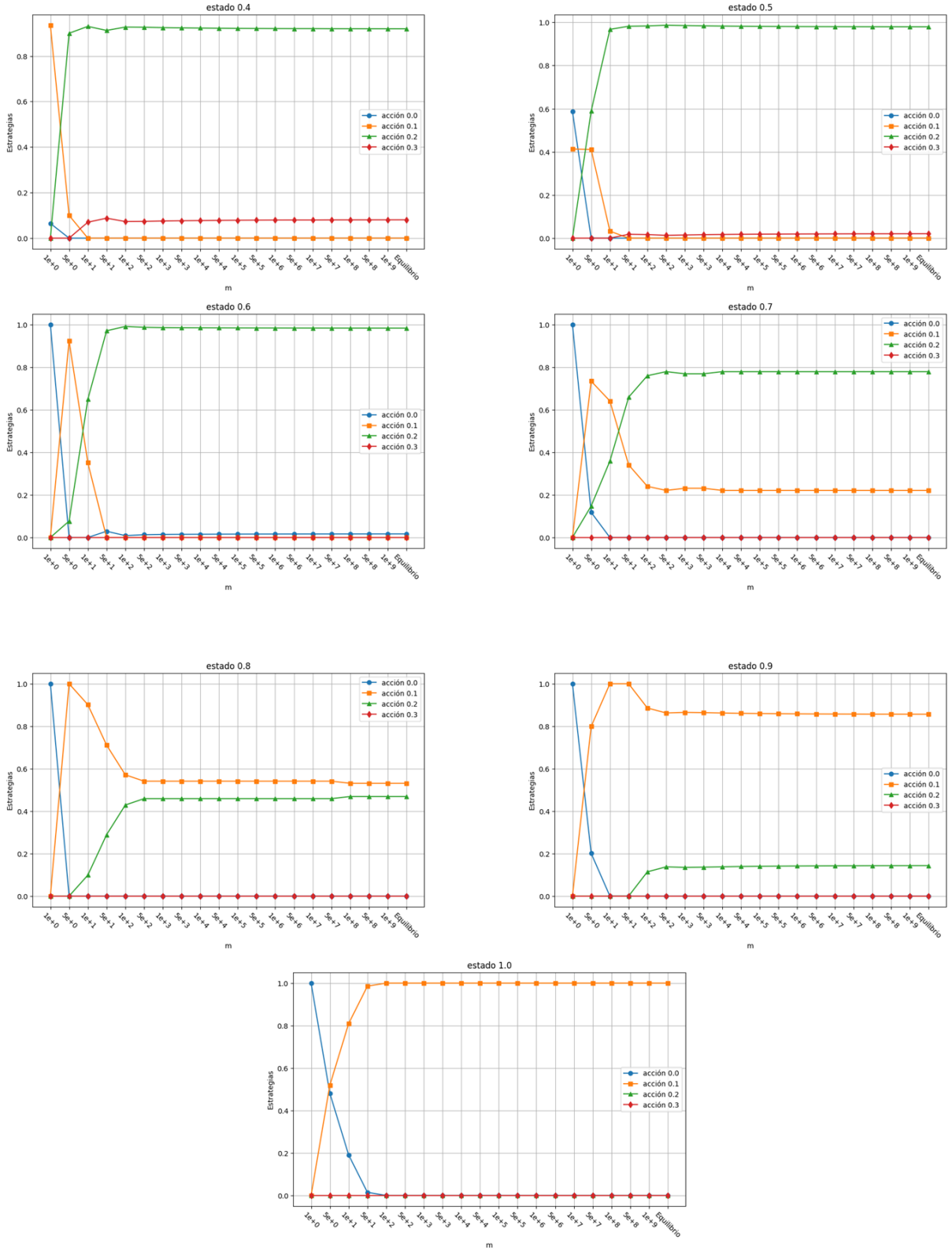

FIGURA 4.2: Estrategias de equilibrio para juegos $\mathcal{G}_{\theta_m}$ simulados. El eje x representa distintos valores de m, mientras que el y representa probabilidades para las acciones. Cada gráfica muestra a estados 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 y 1.0 del juego, respectivamente.

Bibliografía

[1] Ash, R. B. (1972). *Real Analysis and Probability*, volume 11 of *Probability and Mathematical Statistics*. Academic Press. 7

[2] Bertsekas, D. P. and Shreve, S. E. (1978). *Stochastic Optimal Control: The Discrete Time Case*. Academic Press. 7

[3] Fink, A. M. (1964). Equilibrium in a stochastic n-person game. *J. Sci. Hiroshima Univ.*, 28. 1

[4] Levhary, D. and Mirman, L. J. (1980). The great fish war: an example using a dynamic cournot-nash solution. *Bell J. Econom.*, 11:322–334. 29

[5] Minjárez-Sosa, J. A. (2020). *Zero-Sum Discrete-Time Markov Games with Unknown Disturbance Distribution: Discounted and Average Criteria*. SpringerBriefs in Probability and Mathematical Statistics. Springer. 2, 13

[6] Nash, J. F. (1950). Equilibrium points in n-person games. *Proc. Nat. Acad. Sci.*, 36:48–49. 1

[7] Nash, J. F. (1951). Non-cooperative games. *Ann. of Math.*, pages 286–295. 1

[8] Parthasarathy, T. (1973). Discounted, positive, and noncooperative stochastic games. *Int. J. Game Theory*, 2:25–37. 21

[9] Robles-Aguilar, A. D., González-Sánchez, D., and Minjárez-Sosa, J. A. (2022). Empirical approximation of nash equilibria in finite markov games with discounted payoffs. *Asian Journal of Control*, 25:1–13. 2, 3, 30, 32

[10] Shapley, L. S. (1953). Stochastic games. *Proc. Nat. Acad. Sci. U. S. A.*, 39:1095–1100. 1

[11] Takahashi, M. (1964). Equilibrium points of stochastic non-cooperative n-person games. *J. Sci. Hiroshima Univ.*, 28:95–99. 1