Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias Físico Matemáticas

Juegos Markovianos de Suma no Cero con Horizonte Aleatorio

Tesis presentada al

Posgrado en Matemáticas

como requisito para la obtención del grado de

MAESTRO EN CIENCIAS MATEMÁTICAS

por

Adriana González Quiroz

Asesorada por

Dr. Rei Israel Ortega Gutiérrez y Dr. Hugo Adán Cruz Suárez

Puebla Pue.
Julio 2025

Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias Físico Matemáticas

Juegos Markovianos de Suma no Cero con Horizonte Aleatorio

Tesis presentada al

Posgrado en Matemáticas

como requisito para la obtención del grado de

Maestro en Ciencias Matemáticas

por

Adriana González Quiroz

Asesorada por

Dr. Rei Israel Ortega Gutiérrez y Dr. Hugo Adán Cruz Suárez

Puebla Pue.
Julio 2025

Título: Juegos Markovianos de Suma no Cero con Horizonte Aleatorio
Estudiante: Adriana González Quiroz

COMITÉ

Dr. Carlos Camilo Garay
Presidente

Dr. Fernando Velasco Luna
Secretario

Dr. Rubén Blancas Rivera
Vocal

Dr. Bulmaro Juárez Hernández
Vocal

Dr. Rei Israel Ortega Gutiérrez
Asesor

Dr. Hugo Adán Cruz Suárez
Asesor

Agradecimientos

Agradezco profundamente a todas las personas que me acompañaron y apoyaron durante el desarrollo de esta tesis.

A mis padres, Lolis y Carlos. Por su apoyo y compañía, sus palabras de aliento y confianza me llevaron al camino correcto. Este logro también es suyo.

A mi hermano Carlos. Gracias por ser mi compañero de aventuras, por estar conmigo en las buenas y en las malas, el verdadero significado de "Mi vida no sería nada sin ti" lo encontré contigo hermano; si el One Piece fuera una persona, serías tú.

A mis asesores. Dr. Rei, mi tesis estudia la toma de decisiones, pero una teoría no sería capaz de explicar cómo trabajar con usted cambió el juego de la vida para mí, gracias por ayudarme a mejorar, por dedicarme su tiempo y sobre todo por la amistad que desarrollamos. Dr. Hugo, trabajar con usted ha sido un verdadero reto para mí, su vasto conocimiento y experiencia son tan enriquecedores que a veces me siento abrumada por la cantidad de cosas que quisiera aprender de usted, eso mismo me impulsa a esforzarme cada vez más para estar a la altura.

A Monse. Bebé, sigo creyendo que tu existencia le da color a mi vida, tenerte me hace sentir que la suerte existe y que gaste toda la que tenia al conocerte, gracias por ser mi mejor amiga.

A mis grandes amigos. Karlita, no sabes la paz que me haz brindado, tan solo una sonrisa o un abrazo tuyo me hacen muy feliz. Saúl, esta etapa en mi vida no seria lo mismo sin ti, desde platicas serias hasta risas a carcajadas han hecho brillar un día oscuro. Andy, gracias por ser mi amiga, verte y platicar siempre le da un toque dulce a mi vida.

A Cheche. Se que somos personas muy diferentes y por eso chocamos mucho pero quiero que sepas que estoy muy feliz de llamarte mi amigo, no es necesario que lo entiendas pero siempre podrás contar conmigo, te lo digo de corazón.

A Lalo. Gracias por darme tu tiempo, se que no ha sido fácil pero tu presencia le da tranquilidad y alegría a mi vida.

A mis sinodales. Dr. Camilo, Dr. Rubén, Dr. Bulmaro y Dr. Velasco. Gracias por el tiempo dedicado a la lectura y evaluación de este trabajo, así como por sus valiosas observaciones.

A ti. Nunca me arrepentiré de haber sido feliz a tu lado.

A CONACyT. Por el apoyo económico otorgado para la realización de mi maestría.

Índice general

Introducción

El objetivo de este trabajo consiste en demostrar la existencia de un equilibrio de Nash en juegos markovianos de suma no cero con horizonte aleatorio [27, 24], así como en ilustrar dicha existencia mediante ejemplos aplicados a contextos conocidos, tales como el problema de pesquerías y los juegos lineales-cuadráticos. La importancia de considerar un horizonte aleatorio [20], en el estudio de juegos markovianos radica en la necesidad de modelar de manera más realista la duración de estos juegos. Aunque tradicionalmente se asume un horizonte finito o infinito [14], en muchos contextos prácticos la duración del juego es incierta y está sujeta a factores externos impredecibles. Por ejemplo, si dos empresas compiten en un mercado ofreciendo un mismo producto, la eventual quiebra de una de ellas podría interrumpir abruptamente el proceso competitivo, dando lugar a un horizonte de decisión no anticipado.

Los juegos markovianos emergen a partir de la idea principal introducida por el matemático Lloyd Shapley en [38], quien realizó una contribución fundamental a la teoría de juegos. En dicho trabajo, Shapley definió el modelo de juegos estocásticos e inició el análisis de estos bajo horizontes tanto finitos como infinitos. En la actualidad, esta línea de investigación continúa vigente, como se observa en [24], donde se estudian juegos markovianos de suma cero con horizonte infinito. Los fundamentos teóricos que sustentan el desarrollo de esta tesis se apoyan principalmente en los siguientes trabajos; [20] donde Levhari introduce la noción de horizonte aleatorio en el contexto de procesos de decisión de Markov, [34] donde se aborda la teoría de los juegos estocásticos de suma no cero y se estudian equilibrios dinámicos, finalmente [16] incorpora el horizonte aleatorio dentro del marco de los procesos de decisión de Markov. El interés por estudiar juegos markovianos de suma no cero con horizonte aleatorio se justifica por la escasa literatura existente en esta área específica, uno de los pocos trabajos sobre este tema se ubica en [11] donde aborda una clase de juegos de Markov de suma cero en tiempo discreto y con horizonte aleatorio, lo que evidencia una oportunidad relevante para la investigación teórica y aplicada.

La presente investigación se enmarca en el análisis teórico de los juegos markovianos de suma no cero con horizonte aleatorio, utilizando herramientas de la teoría de juegos. Se estudia un juego markoviano de suma no cero con dos jugadores, quienes actúan de manera económicamente racional, en base a ello, se demuestra la existencia de un equilibrio de Nash en este tipo de juegos, bajo la condición de que cada jugador tiene un horizonte aleatorio distinto, los cuales se suponen independientes. Finalmente, con el fin de ilustrar la importancia y aplicabilidad del equilibrio de Nash en este contexto, se recurre a su utilización en la resolución de dos problemas específicos: en el enfoque de recompensas, se presenta la guerra de la gran pesca donde se analiza el comportamiento estratégico de dos países interesados en la explotación de recursos pesqueros ubicados en una misma región marítima y en el enfoque de costos, se desarrolla un juego lineal-cuadrático, fundamentado en el modelo clásico del mismo nombre.

Esta tesis se estructura en tres capítulos. El Capítulo 1 ofrece una introducción a la teoría de juegos, en la que se presenta el concepto de equilibrio de Nash y se abordan los juegos estocásticos. En particular, se introducen los juegos markovianos, junto con el criterio de optimalidad como herramienta fundamental para la evaluación de decisiones. En el Capítulo 2 se abordan los juegos markovianos de suma no cero, haciendo especial énfasis en la incorporación de un horizonte aleatorio. Este capítulo se divide en dos secciones: una orientada al enfoque basado en recompensas y otra centrada en el enfoque basado en costos. Asimismo, se propone un método para resolver este tipo de juegos, fundamentado en la caracterización del equilibrio de Nash mediante el uso de programación dinámica. Finalmente, el Capítulo 3 presenta aplicaciones de los conceptos desarrollados, mediante el estudio de la guerra de la gran pesca y un juego lineal-cuadrático con horizonte aleatorio.

Resumen de Notación y Terminología

Símbolos y abreviaciones

$\mathbb{N}$	Conjunto de enteros positivos
$\mathbb{N}_0$	Conjunto de enteros no negativos
$\mathbb{R}$	Conjunto de números reales
$:=$	Igualdad por definición
$\mathcal{N}$	Número de jugadores (fijado en 2 para esta tesis)
G	Juegos estático
GM	Modelo de juego estocástico
X	Espacio de Estados
$1_D(\cdot)$	Función indicadora del conjunto D
	$1_D(x) := 1 \quad \text{si} \quad x \in D \quad \text{y} \quad 1_D(x) := 0 \quad \text{si} \quad x \notin D$
$\mathbb{K}$	Espacio de estado-acción admisbles
Q	Ley de Transición
$\mathcal{H}$	Horizonte del juego
H_t	Conjunto de historias del juego hasta el tiempo t
$h_t \in H_t$	Historia del juego hasta el tiempo t
$\pi = (\pi_1, \pi_2)$	Multiestrategia
$\Pi = \Pi_1 \times \Pi_2$	Familia de multiestrategias
Ω	Espacio muestral
$\mathcal{F}$	σ-algebra
$P_x^{\pi_1,\pi_2}$	Medida de Probabilidad del proceso de estado del juego
v.a.	Variable aleatoria

Notación

Para el jugador $i = 1, 2$, se utiliza la siguiente notación:

A_i	Conjunto de acciones
$a_i \in A_i$	Acción
$A_i(x)$	Conjunto de acciones admisibles en el estado $x \in X$
r_i/c_i	Función de recompensa o costo
$a_{i,t}$	Acción en el tiempo t
π_i	Estrategia
Π_i	Familia de Estrategias
$\mathbb{F}_i$	Conjunto de Funciones medibles f_i tal que $f_i(x) \in A_i(x)$
$\Pi_{i,M}$	Familia de estrategias de Markov
$\Pi_{i,E}$	Subfamilia de estrategias estacionarias
$\Pi_{i,D}$	Familia de estrategias deterministas
$\Pi_{i,DM}$	Subfamilia de estrategias deterministas de Markov
$\Pi_{i,DE}$	Subfamilia de estrategias estacionarias deterministas

Para el jugador $i = 1, 2$ respecto a recompensas y costos se tiene:

Recompensas	**Costos**	
$\mathcal{J}_{i,N}(\pi, x)$	$\mathcal{V}_{i,N}(\pi, x)$	Recompensa (Costo) total esperada con horizonte finito.
$\mathcal{J}_{i,N_{\beta_i}}(\pi, x)$	$\mathcal{V}_{i,N_{\beta_i}}(\pi, x)$	Recompensa (Costo) total esperada descontada con horizonte finito
$\mathcal{J}_{i,\infty}(\pi, x)$	$\mathcal{V}_{i,\infty}(\pi, x)$	Recompensa (Costo) total esperada descontada con horizonte infinito
$\mathcal{J}_{i,\tau_i}(\pi, x)$	$\mathcal{V}_{i,\tau_i}(\pi, x)$	Recompensa (Costo) total esperada descontada con horizonte aleatorio
$MR_1(\pi_2)$	$\mathcal{MR}_1(\pi_2)$	Conjunto de mejores respuestas para el jugador 1 a la estrategia π_2
$MR_2(\pi_1)$	$\mathcal{MR}_2(\pi_1)$	Conjunto de mejores respuestas para el jugador 2 a la estrategia π_1

Para un espacio de Borel X, se utiliza la siguiente notación (véase [24]):

$\mathscr{B}(X)$	σ-algebra de Borel en X
$\mathbb{P}(X)$	Espacio de medidas de probabilidad en X dotado de la topología débil
$\mathbb{P}(X\|Y)$	Familia de kérneles estocásticos en X dado Y donde Y es un espacio de Borel.

Capítulo 1

Teoría de juegos

La teoría de juegos es un área de la matemática que utiliza modelos para estudiar interacciones en estructuras formalizadas de incentivos (mejor conocidos como juegos) donde interactúan $\mathcal{N}$ individuos para tomar una decisión [2].

Cabe resaltar que uno de los primeros hitos fundamentales en el desarrollo de la teoría de juegos se dio en [30] donde se introduce formalmente la idea de juegos de suma cero. Unos años después, la teoría de juegos alcanzó su apogeo en [31], una obra que pronto fue reconocida como uno de los logros científicos más importantes del siglo XX y extendió su influencia desde la economía a numerosas disciplinas científicas; constituyéndose según sus propios creadores en un instrumento apropiado para desarrollar una teoría del comportamiento económico, y en una aproximación al análisis de situaciones sociales. Surgiendo así la década prodigiosa de esta teoría iniciando en 1950, donde John Forbes Nash presenta [28] en la que se establecen las bases generales de este tipo de juegos y resultados esenciales para la teoría de juegos [27, 29]. En dicho artículo se introducía el concepto de punto de equilibrio (conocido actualmente como equilibrio de Nash) y se prueba su existencia, además debido al trabajo de oligopolios en [7], Nash demostró que el equilibrio de Nash generalizaba el equilibrio clásico en un duopolio de Cournot [22].

De manera introductoria, puede entenderse la teoría de juegos como una situación de toma de decisiones en la que varios individuos, conocidos como jugadores, deberán tomar dichas decisiones o acciones que los llevarán a obtener un resultado que puede ser positivo o negativo dependiendo del escenario. Aunque cada jugador tenga que tomar una decisión para obtener o no un beneficio, el resultado depende no solo de sus propias acciones, sino también de las de los demás.

El objetivo de la teoría de juegos es explicar más adecuadamente la conducta humana frente a la toma de decisiones y contribuir a identificar la mejor respuesta posible para cada uno de los jugadores.

Los juegos pueden clasificarse según diversas características fundamentales. Por ejemplo, en los juegos de suma cero, la ganancia de un jugador representa exactamente la pérdida de otro, mientras que en los de suma no cero, la ganancia de un competidor no depende necesariamente de la pérdida de otro. También pueden dividirse en simultáneos, cuando los jugadores toman decisiones al mismo tiempo, o secuenciales, cuando las decisiones se realizan en un orden determinado. Otra clasificación distingue entre juegos dinámicos, en los que las decisiones se toman por etapas, y juegos estáticos, donde todas las decisiones se realizan en un solo momento. Dentro de estos últimos, se encuentran los juegos markovianos, en los que la evolución depende únicamente del estado actual, y los no markovianos, donde influyen también eventos pasados.

1.1. Juegos estáticos

Definición 1.1.1. Un *juego* es cualquier situación de decisión caracterizada por involucrar:

- $\mathcal{N}$ jugadores,
- un conjunto de estrategias y
- un pago para cada jugador.

Definición 1.1.2. Una *estrategia* (también conocida como *acción*) es una decisión específica que un jugador puede tomar en un momento determinado del juego.

Observación 1.1.3. En el contexto de esta tesis, se empleará el término *acción* para describir las decisiones individuales que los jugadores toman en una etapa específica. Asimismo, el término *estrategia* se utilizará para referirse a la secuencia completa de decisiones a lo largo de todas las etapas, formada por las acciones correspondientes a cada una de ellas.

Definición 1.1.4. Un *pago* describe cuantitativamente el resultado de cada jugador en términos de la cantidad que gana o pierde, este depende no solo de la estrategia que se elige, sino también de las estrategias que eligen los competidores guiados por sus propios intereses. Durante este trabajo se hará la distinción entre si el pago puede ser descrito como una *recompensa* o un *costo*, ya que dependiendo del contexto del juego, si un jugador recibe una recompensa deseará maximizar su pago, en cambio si el jugador está analizando un costo, preferirá minimizarlo.

De ahora en adelante se fija $\mathcal{N} = 2$, es decir, solo se trabajarán juegos de dos jugadores, de esta manera, en todo el trabajo cuando se hable del jugador i, se entenderá que $i = 1, 2$.

Definición 1.1.5. Un *juego estático* es aquél en el que cada jugador toma una única decisión sin conocer previamente las acciones de los demás jugadores. Además, un juego estático con dos jugadores consta de un conjunto de acciones denotado por A_i para el jugador i, de esta forma $a_i \in A_i$ denotará una acción para el jugador i. Sean (a_1, a_2) las acciones para el jugador 1 y 2, respectivamente y sea $r_i/c_i : A_1 \times A_2 \to \mathbb{R}$ la función de recompensa o costo del jugador i, respectivamente. Entonces $r_i/c_i(a_1, a_2)$ es la recompensa o el costo del jugador i si los jugadores eligen (a_1, a_2). Dicho juego será denotado por

$$G = \{A_1, A_2, r_1/c_1, r_2/c_2\}.$$

En un juego estático, una *acción pura* es cada una de las posibles decisiones del jugador.

Cabe resaltar que los juegos estáticos son solo uno de los tipos de juegos que existen, por ejemplo, los juegos dinámicos son aquellos en donde los jugadores actúan en determinado orden y las decisiones se toman de forma consecutiva. Además, los juegos pueden ser representados por medio de una matriz llamada *matriz de juego* en el caso de los juegos estáticos y por medio de un *árbol de decisión* en el caso de juegos dinámicos.

Un ejemplo de juego estático puede ser el típico piedra, papel o tijeras donde dos jugadores apuestan 10 pesos, en caso de un empate, nadie obtiene una recompensa. Por otro lado, un ejemplo de juego dinámico puede ser una partida de ajedrez donde el ganador recibe 100 pesos del perdedor.

Los juegos pueden ser clasificados según su pago en juegos de suma cero y juegos de suma no cero.

1.1.1. Juegos de suma cero y suma no cero

Supóngase que hay un juego de dos jugadores donde el jugador i tiene su respectivo conjunto de acciones A_i. Cuando el jugador i elige la acción $a_i \in A_i$, el juego se lleva a cabo y el pago de cada jugador es registrado como $r_1/c_1(a_1, a_2)$ para el jugador 1 y $r_2/c_2(a_1, a_2)$ para el jugador 2.

- En un *juego de suma cero*, el pago que un jugador gane, el otro lo pierde. Por ejemplo, si el jugador 1 recibe $r_1/c_1(a_1, a_2)$, entonces el jugador 2 obtiene como pago $r_2/c_2(a_1, a_2) = -r_1/c_1(a_1, a_2)$.
- En un *juego de suma no cero*, el pago de un competidor no depende necesariamente de la pérdida de otro, de esta manera $r_1/c_1(a_1, a_2) + r_2/c_2(a_1, a_2) \neq 0$.

Un caso ilustrativo para juegos de suma cero es la situación donde dos personas apuestan dinero en un lanzamiento de moneda al decidir entre las estrategias águila o sol. En contraste, un ejemplo de suma no cero sería el conocido dilema del prisionero [2].

1.1.2. Equilibrio de Nash

Dado el concepto de juego, se presenta la siguiente pregunta: ¿Es posible elegir una estrategia para cada jugador tal que todos ellos obtengan la mejor respuesta posible? Efectivamente, la teoría de juegos ayuda a encontrar la mejor solución a un juego para todos los jugadores por medio del equilibrio de Nash.

Definición 1.1.6. Sea $G = \{A_1, A_2, r_1, r_2\}$ un juego estático de dos jugadores, el par de acciones puras $(a_1^*, a_2^*) \in A_1 \times A_2$ es un *equilibrio de Nash (para recompensas)* si:

$$r_1(a_1^*, a_2^*) \geq r_1(a_1, a_2^*)$$

$$r_2(a_1^*, a_2^*) \geq r_2(a_1^*, a_2)$$

para cada acción $a_1 \in A_1$ y $a_2 \in A_2$. En otras palabras

$$a_1^* = \arg \max_{a_1 \in A_1} r_1(a_1, a_2^*),$$

y

$$a_2^* = \arg \max_{a_2 \in A_2} r_2(a_1^*, a_2).$$

Definición 1.1.7. Sea $G = \{A_1, A_2, c_1, c_2\}$ un juego estático de dos jugadores, el par de acciones puras $(a_1^*, a_2^*) \in A_1 \times A_2$ es un *equilibrio de Nash (para costos)* si:

$$c_1(a_1^*, a_2^*) \leq c_1(a_1, a_2^*)$$

$$c_2(a_1^*, a_2^*) \leq c_2(a_1^*, a_2)$$

para cada acción $a_1 \in A_1$ y $a_2 \in A_2$. En otras palabras

$$a_1^* = \arg \min_{a_1 \in A_1} c_1(a_1, a_2^*),$$

y

$$a_2^* = \arg \min_{a_2 \in A_2} c_2(a_1^*, a_2).$$

1.2. Juegos Estocásticos

Como se señaló anteriormente, la década de los 50 fue la década en la que se dio forma a la teoría de juegos. Después de la publicación de [31], el matemático Lloyd Shapley contribuyó en la teoría de juegos con su artículo [38], donde definió el modelo de juegos estocásticos, que fue el primer modelo dinámico general de un juego que se estableció.

Un juego estocástico es un juego por etapas, al principio de cada etapa los jugadores están en algún estado del juego y ellos deberán tomar decisiones que lo afecten, además dependiendo del estado en el que se encuentren y de las acciones de todos los jugadores se recibirá un pago para cada jugador. Después el juego se mueve a un nuevo estado aleatoriamente cuya distribución depende del estado previo y las acciones elegidas por los jugadores. El procedimiento se repite en el nuevo estado y el juego continúa por un número finito, aleatorio o infinito de etapas. En este tipo de juegos las estrategias recopilan las acciones utilizadas en cada una de las etapas, en otras palabras, una estrategia para el jugador i es una sucesión de las acciones tomadas por el jugador i en cada una de las etapas del juego. De esta manera, en los juegos estocásticos lo que se desea obtener es el mejor pago posible de todas las etapas y se busca una estrategia para cada jugador que lo otorgue, para ello se utiliza un criterio de optimalidad que mide la efectividad de una estrategia dependiendo del contexto del juego y se empleará la programación dinámica como herramienta fundamental para determinar dicha estrategia.

1.2.1. Tipos de juegos estocásticos

Cabe resaltar que los juegos estocásticos pueden clasificarse en varias categorías según diferentes criterios. Algunas de las más utilizadas son:

(a) Número de Jugadores

  (i) 2 jugadores

  (ii) $\mathcal{N}$ jugadores

(b) Información disponible

  (i) Información perfecta: Los jugadores conocen completamente el estado del juego y las acciones pasadas.

  (ii) Información imperfecta: Algunos elementos del juego, como el estado exacto o las acciones del oponente, pueden no ser completamente observables.

(c) Horizonte temporal

  (i) Horizonte finito: El juego se realiza por un número finito de etapas.

  (ii) Horizonte infinito: No hay un número predefinido de etapas; el juego puede continuar indefinidamente.

  (iii) Horizonte aleatorio: El juego puede finalizar en una etapa aleatoria, determinada por una distribución de probabilidad.

(d) Naturaleza del pago

  (i) Juegos de recompensa: Los jugadores buscan maximizar su pago final.

  (ii) Juegos de costo: Los jugadores buscan minimizar su pago final.

(e) Cumplimiento de la Propiedad de Markov

(i) Markovianos: Descritos en la Definición 1.3.1.

(ii) No Markovianos: Son aquellos donde las funciones de pago y su ley de transición del juego depende del historial completo de estados y acciones.

(f) Suma cero y no cero

(i) Suma cero: La ganancia de un jugador implica la pérdida del otro.

(ii) Suma no cero: La ganancia de un jugador no implica necesariamente la pérdida del otro.

1.3. Juegos Markovianos

Definición 1.3.1. Un **modelo de juego markoviano** es una tupla

$$GM := \{X, (A_i, A_i(x), c_i/r_i)_{i=1,2}, Q, \mathcal{H}\} \tag{1.1}$$

donde

(a) X es un espacio de Borel, el cual es llamado *espacio de estados*.

(b) El jugador i está caracterizado por tres elementos $(A_i, A_i(x), c_i/r_i)$, con

(i) A_i un conjunto de Borel conocido como el *conjunto de acciones del jugador i*.

(ii) $A_i(x)$ una multifunción de X a A_i, la cual define para cada $x \in X$ el *conjunto de acciones admisibles para el jugador i en el estado x*. Sean

$$A(x) = A_1(x) \times A_2(x) \quad \text{y} \quad \mathbb{K} := \{(x, a_1, a_2) | x \in X, (a_1, a_2) \in A(x)\},$$

nótese que los elementos de $A(x)$ son pares ordenados de la forma (a_1, a_2) con $a_1 \in A_1(x)$ y $a_2 \in A_2(x)$, además $\mathbb{K}$ es un subconjunto de Borel de $X \times A_1 \times A_2$ conocido como el *conjunto de estado-acciones admisibles* del juego.

(iii) $r_i/c_i : \mathbb{K} \to \mathbb{R}$ una función medible que representa la *recompensa* $r_i(x, a_1, a_2)$ o bien el *costo* $c_i(x, a_1, a_2)$, para el jugador i.

(c) $Q \in \mathbb{P}(X|\mathbb{K})$ es un kérnel estocástico en X dado $\mathbb{K}$ conocido como *ley de transición* del juego (donde $\mathbb{P}(X|\mathbb{K})$ es el conjunto de todas las medidas de probabilidad de X dado $\mathbb{K}$).

(d) $\mathcal{H}$ es el horizonte del juego, el cual puede ser finito, infinito o aleatorio.

Observación 1.3.2. (a) En un modelo de juego estocástico, los horizontes pueden diferir entre los jugadores sin que esto implique un cambio en la categoría del juego. En otras palabras, dentro de un juego con horizonte finito, cada jugador puede tener un horizonte distinto; es decir, el jugador 1 puede tener un horizonte N_1, mientras que el jugador 2 puede tener un horizonte N_2. De manera análoga, en un juego con horizonte aleatorio, cada jugador puede estar sujeto a un tiempo de finalización aleatorio diferente. En este caso, el horizonte del jugador i puede estar determinado por una variable aleatoria τ_i, que representa el tiempo en el que dicho jugador deja de participar en el juego. En este contexto, la estructura del modelo de juego estocástico presentado en (1.1) puede formalizarse mediante la siguiente tupla:

$$GM = \left\{ X, (A_i, A_i(x), c_i/r_i, \mathcal{H}_i)_{i=1,2}\, , Q \right\}. \tag{1.2}$$

(b) Cabe señalar que, si se considera un modelo de juego estocástico con horizonte finito $N = 1$, dicho juego se reduce en realidad a un juego estático, al tomarse el espacio de estados como el conjunto singular $\{x_0\}$ y la ley de transición Q como la delta de Dirac concentrada en $\{x_0\}$.

Desarrollo del Juego

Paso 1. En cada instante de tiempo $t \in \mathbb{N}_0$, los jugadores observan el estado del juego $x_t = x \in X$.
Paso 2. Los jugadores 1 y 2 seleccionan, de forma independiente y simultánea, sus acciones:

- El jugador 1 elige $a_{1,t} \in A_1(x)$.
- El jugador 2 elige $a_{2,t} \in A_2(x)$.

Así, el par de acciones que toman ambos jugadores en el tiempo t es $(a_{1,t}, a_{2,t}) \in A(x)$.
Paso 3. Ambos jugadores reciben su pago:

- El jugador 1 obtiene $r_1/c_1(x_t, a_{1,t}, a_{2,t})$.
- El jugador 2 obtiene $r_2/c_2(x_t, a_{1,t}, a_{2,t})$.

Paso 4. El juego cambia a un nuevo estado $x_{t+1} = y \in X$ según la ley de transición $Q(\cdot|x_t, a_{1,t}, a_{2,t})$.
Paso 5. Una vez en el nuevo estado, el proceso se repite desde el Paso 2.

Observación 1.3.3. El juego sigue según el horizonte del juego $\mathcal{H}$ y el objetivo de cada jugador es obtener el mejor pago posible (el máximo de sus ganancias o el mínimo de sus pérdidas, dependiendo del contexto del juego).

Observación 1.3.4. Dada la clasificación anterior, se puede establecer que los juegos estocásticos de interés en esta tesis corresponden a los **Juegos Markovianos de suma no cero con horizonte aleatorio** de dos jugadores y se hará la distinción entre si el pago es una recompensa o un costo.

1.3.1. Estrategias

Durante el juego, los jugadores seleccionan las acciones en cada etapa mediante reglas llamadas *estrategias* definidas a continuación.

Definición 1.3.5. Sea $H_0 := X$ y $H_t := \mathbb{K} \times H_{t-1}$ para cada $t \in \mathbb{N}$ un elemento $h_t \in H_t$ está dado por

$$h_t = (x_0, a_{1,0}, a_{2,0}, ..., x_{t-1}, a_{1,t-1}, a_{2,t-1}, x_t)$$

el cual representa la *historia del juego hasta el tiempo* t, donde $(x_j, a_{1,j}, a_{2,j}) \in \mathbb{K}$ para todo $j = 0, 1, ..., t-1$ y $x_t \in X$. Por otra parte, para cada $x \in X$, se define $\mathbb{A}_i(x) := \mathbb{P}(A_i(x))$ para el jugador i, así como los conjuntos de kérneles estocásticos

$$\Phi_1 := \{\varphi_1 \in \mathbb{P}(A_1|X) : \varphi_1(\cdot|x) \in \mathbb{A}_1(x),\ \ \forall x \in X\}$$

$$\Phi_2 := \{\varphi_2 \in \mathbb{P}(A_2|X) : \varphi_2(\cdot|x) \in \mathbb{A}_2(x),\ \ \forall x \in X\}.$$

Definición 1.3.6. Una *estrategia* para el jugador i está definida como una sucesión $\pi_i = \{\pi_{i,t}\}$ de kérneles estocásticos $\pi_{i,t}$ en $\mathbb{P}(A_i|H_t)$ tal que

$$\pi_{i,t}(A_i(x_t)|h_t) = 1, \qquad \forall h_t \in H_t,\ \ t \in \mathbb{N}_0.$$

Se denotará por Π_i a la familia de todas las estrategias para el jugador i.

Sea $\Pi := \Pi_1 \times \Pi_2$, un elemento de Π será denotado por π y llamado *multiestrategia*. Además, sea $\mathbb{F}_i$ el conjunto de todas las funciones medibles $f_i : X \to A_i$ tal que $f_i(x) \in A_i(x)$ para todo $x \in X$.

Clasificación de estrategias

Una estrategia $\pi_i = \{\pi_{i,t}\} \in \Pi_i$ para el jugador i es llamada

(a) **estrategia de Markov** si $\pi_{i,t}$ esta en Φ_i para todo $t \in \mathbb{N}_0$;

(b) **estrategia estacionaria** si

$$\pi_{i,t}(\cdot|h_t) = \varphi_i(\cdot|x_t) \quad \forall h_t \in H_t,\ \ t \in \mathbb{N}_0;$$

para algún kérnel estocástico $\varphi_i \in \Phi_i$, de modo que π_i es de la forma $\pi_i = \{\varphi_i, \varphi_i, ...\}$;

(c) **estrategia determinista** (o **pura**) si existe una sucesión $\{g_{i,t}\}$ de funciones medibles $g_{i,t} : H_t \to A_i$ tal que, para cada $h_t \in H_t$ y $t \in \mathbb{N}_0$, se tiene $g_{i,t}(h_t) \in A_i(x_t)$ y

$$\pi_{i,t}(A_i'|h_t) = 1_{A_i'}[g_{i,t}(h_t)] \text{ para todo } A_i' \in \mathscr{B}(A_i),$$

lo cual significa que $\pi_{i,t}(\cdot|h_t)$ es la medida de Dirac concentrada en $g_{i,t}(h_t)$;

(d) **estrategia determinista de Markov** si existe una sucesión $\{f_{i,t}\}$ de funciones $f_{i,t} \in \mathbb{F}_i$ tal que $\pi_{i,t}(\cdot|h_t)$ es la medida de Dirac concentrada en $f_{i,t}(x_t) \in A_i(x_t)$ para todo $h_t \in H_t$ y $t \in \mathbb{N}_0$;

(e) **estrategia determinista estacionaria** si existe una función $f_i \in \mathbb{F}_i$ tal que $\pi_{i,t}(\cdot|h_t)$ es la medida de Dirac concentrada en $f_i(x_t) \in A_i(x_t)$ para todo $h_t \in H_t$ y $t \in \mathbb{N}_0$.

Notación:

- $\Pi_{i,M}$ es la familia de estrategias de Markov del jugador i.
- $\Pi_{i,E}$ es la subfamilia de estrategias estacionarias del jugador i.
- $\Pi_{i,D}$ es la familia de estrategias deterministas del jugador i.
- $\Pi_{i,DM}$ es la subfamilia de estrategias deterministas de Markov del jugador i.
- $\Pi_{i,DE}$ es la subfamilia de estrategias estacionarias deterministas del jugador i.

Observación 1.3.7.

- $\Pi_{i,E} \subset \Pi_{i,M} \subset \Pi_i$
- $\Pi_{i,DE} \subset \Pi_{i,DM} \subset \Pi_{i,D} \subset \Pi_i$.

En el presente trabajo, las estrategias de interés corresponden a las deterministas de Markov.

1.3.2. Procesos de estado del juego de Markov

Una vez introducido el concepto de juegos markovianos, es posible presentar los procesos de estado en un juego de Markov. Estos son fundamentales para la construcción del espacio de probabilidad asociado al juego, el cual estará representado por un espacio de probabilidad en cada estado. A su vez, en cada estado existirá un proceso de decisión de Markov.

Sea $(\Omega', \mathcal{F}')$ el espacio medible que consiste en el espacio muestral $\Omega' = \mathbb{K}^\infty$ y su σ-álgebra producto $\mathcal{F}'$. Siguiendo los argumentos estándar (véase, por ejemplo, [34]), del Teorema de Ionescu-Tulcea se tiene que para cada par de estrategias $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$ y un estado inicial $x_0 = x \in X$, existe una única medida de probabilidad $P_x^{\pi_1,\pi_2}$ y un proceso estocástico $\{(x_t, a_{1,t}, a_{2,t})\}$, donde $x_t \in X$ y $(a_{1,t}, a_{2,t}) \in A_1(x) \times A_2(x)$ representan el estado y las acciones de los jugadores, respectivamente, en la etapa $t \in \mathbb{N}_0$ se cumple, para $D \in \mathscr{B}(X)$, $A_i' \in \mathscr{B}(A_i)$

$$
\begin{aligned}
P_x^{\pi_1,\pi_2}\left[x_0 \in D\right] &= \delta_x(D) \\
P_x^{\pi_1,\pi_2}\left[a_{1,t} \in A_1' \mid h_t\right] &= \pi_{1,t}\left(A_1' \mid h_t\right) \\
P_x^{\pi_1,\pi_2}\left[a_{2,t} \in A_2' \mid h_t\right] &= \pi_{2,t}\left(A_2' \mid h_t\right) \\
P_x^{\pi_1,\pi_2}\left[a_{1,t} \in A_1', a_{2,t} \in A_2' \mid h_t\right] &= \pi_{1,t}\left(A_1' \mid h_t\right)\pi_{2,t}\left(A_2' \mid h_t\right) \\
P_x^{\pi_1,\pi_2}\left[x_{t+1} \in D \mid h_t, a_{1,t}, a_{2,t}\right] &= Q\left(D \mid x_t, a_{1,t}, a_{2,t}\right)
\end{aligned}
$$

donde $\delta_x(\cdot)$ es la medida de Dirac concentrada en x. Se denota por $E_x^{\pi_1,\pi_2}$ al operador esperanza con respecto a $P_x^{\pi_1,\pi_2}$.

El proceso estocástico $\{x_t\}$ definido en $(\Omega', \mathcal{F}', P_x^{\pi_1,\pi_2})$ se llama proceso de estados del juego.

1.4. Criterios de Optimalidad

Una vez introducido el concepto de estrategia, surge la siguiente cuestión: ¿cómo analizar las estrategias del juego para cada jugador? Para abordar esta interrogante, se introduce el siguiente concepto.

Un criterio de optimalidad (también denominado función objetivo o criterio de rendimiento) permite evaluar la efectividad de una estrategia en función del contexto del juego. En la literatura, los criterios que se consideran con mayor frecuencia son:

- Criterio total. Evalúa la suma total de recompensas obtenidas a lo largo del juego.

- Criterio descontado con horizonte finito. Aplica un factor de descuento a las recompensas futuras, considerando un horizonte temporal finito.

- Criterio descontado con horizonte infinito. Aplica un factor de descuento a las recompensas futuras, considerando un horizonte temporal infinito.

- Criterio aleatorio. Introduce un enfoque probabilístico en la evaluación de la estrategia, considerando distribuciones de recompensas en lugar de valores fijos. Además tendrá dos variantes, el descontado o no descontado dependiendo de si se introduce un factor de descuento o no.

Estos criterios pueden distinguirse según si se aplican a un contexto de recompensas [34, 13] o costos [14, 15], dependiendo de la naturaleza del juego analizado.

1.4.1. Recompensas

(a) **Recompensa total esperada con horizonte finito.**
Sean $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$ y $x \in X$. La recompensa total esperada con horizonte finito N para el jugador i se define por

$$\mathcal{J}_{i,N}(x, \pi_1, \pi_2) := \mathbb{E}_x^{\pi_1,\pi_2}\left[\sum_{t=0}^{N} r_i(x_t, a_{1,t}, a_{2,t}) + R_{i,N+1}(x_{N+1})\right],$$

donde $R_{i,N+1}$ es una función medible en X, llamada función de recompensa terminal.

(b) **Recompensa total esperada descontada con horizonte finito.**
Sean $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$ y $x \in X$. La recompensa total esperada descontada con horizonte finito N para al jugador i se define por

$$\mathcal{J}_{i,N_{\beta_i}}(x, \pi_1, \pi_2) := \mathbb{E}_x^{\pi_1,\pi_2}\left[\sum_{t=0}^{N} \beta_i^t r_i(x_t, a_{1,t}, a_{2,t}) + \beta_i^{N+1} R_{i,N+1}(x_{N+1})\right],$$

donde $R_{i,N+1}$ es una función medible en X, llamada función de recompensa terminal y $\beta_i \in (0,1)$ es llamado factor de descuento del jugador i.

(c) **Recompensa total esperada descontada con horizonte infinito.**
Sean $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$ y $x \in X$. La recompensa total esperada descontada con horizonte infinito para el jugador i se define por

$$\mathcal{J}_{i,\infty}(x, \pi_1, \pi_2) := \mathbb{E}_x^{\pi_1,\pi_2}\left[\sum_{t=0}^{\infty} \beta_i^t r_i(x_t, a_{1,t}, a_{2,t})\right],$$

donde $\beta_i \in (0,1)$ es llamado factor de descuento del jugador i.

(d) **Recompensa total esperada descontada con horizonte aleatorio.**
Sean $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$ y $x \in X$. La recompensa total esperada descontada con horizonte aleatorio para el jugador i se define por

$$\mathcal{J}_{i,\tau_i}(x, \pi_1, \pi_2) := \mathbb{E}_i\left[\sum_{t=0}^{\tau_i} \beta_i^t r_i(x_t, a_{1,t}, a_{2,t})\right], \tag{1.3}$$

donde $\beta_i \in (0,1)$ es llamado factor de descuento del jugador i, τ_i es una v.a. discreta en $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$ para el jugador i con distribución conocida dada por $\mathbb{P}_i(\tau_i = t) := \rho_{i,t}, \quad t = 0, 1, 2, \ldots, T_i$, donde T_i es un número entero positivo o $T_i = \infty$ y $\mathbb{E}_i$ representa el valor esperado con respecto a la distribución conjunta del proceso $\{(x_t, \pi_{1,t}, \pi_{2,t})\}$ y τ_i con $i = 1, 2$.

Observación 1.4.1. Tal como se indicó previamente, esta tesis se basa en el estudio de modelos con horizonte aleatorio; en consecuencia, el criterio de optimalidad adoptado será el correspondiente a dicho horizonte. No obstante, los demás criterios también son relevantes, ya que, como se demostrará en el Capítulo 2, el criterio con horizonte aleatorio es, bajo ciertas condiciones, equivalente a los criterios con horizonte finito o infinito.

1.4.2. Costos

(a) **Costo total esperado con horizonte finito.**
Sean $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$ y $x \in X$. El costo total esperado con horizonte finito N para el jugador i se define por

$$\mathcal{V}_{i,N}(x, \pi_1, \pi_2) := \mathbb{E}_x^{\pi_1,\pi_2}\left[\sum_{t=0}^{N} c_i(x_t, a_{1,t}, a_{2,t}) + C_{i,N+1}(x_{N+1})\right],$$

donde $C_{i,N+1}$ es una función medible en X, llamada función de costo terminal.

(b) **Costo total esperado descontado con horizonte finito.**
Sean $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$ y $x \in X$. El costo total esperado descontado con horizonte finito N para al jugador i se define por

$$\mathcal{V}_{i,N_{\beta_i}}(x, \pi_1, \pi_2) := \mathbb{E}_x^{\pi_1,\pi_2}\left[\sum_{t=0}^{N} \beta_i^t c_i(x_t, a_{1,t}, a_{2,t}) + \beta_i^{N+1} C_{i,N+1}(x_{N+1})\right],$$

donde $C_{i,N+1}$ es una función medible en X, llamada función de costo terminal y $\beta_i \in (0, 1)$ es llamado factor de descuento del jugador i.

(c) **Costo total esperado descontado con horizonte infinito.**
Sean $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$ y $x \in X$. El costo total esperado descontado con horizonte infinito para el jugador $i = 1, 2$ se define por

$$\mathcal{V}_{i,\infty}(x, \pi_1, \pi_2) := \mathbb{E}_x^{\pi_1,\pi_2}\left[\sum_{t=0}^{\infty} \beta_i^t c_i(x_t, a_{1,t}, a_{2,t})\right],$$

donde $\beta_i \in (0, 1)$ es llamado factor de descuento del jugador i.

(d) **Costo total esperado con horizonte aleatorio.**
Sean $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$ y $x \in X$. El costo total esperado con horizonte aleatorio para el jugador i se define por

$$\mathcal{V}_{i,\tau_i}(x, \pi_1, \pi_2) := \mathbb{E}_i \left[\sum_{t=0}^{\tau_i} c_i(x_t, a_{1,t}, a_{2,t}) \right], \tag{1.4}$$

donde τ_i es una v.a. discreta en $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$ para el jugador i con distribución conocida dada por $\mathbb{P}_i(\tau_i = t) := \rho_{t_i}, \quad t = 0, 1, 2, \ldots, T_i$, donde T_i es un número entero positivo o $T_i = \infty$.

Observación 1.4.2. Cabe señalar que, en los criterios de optimalidad referentes a horizonte aleatorio, se adopta la convención de utilizar el descontado en recompensas y el no descontado en costos.

1.5. Equilibrio de Nash en juegos estocásticos

Para finalizar, es fundamental reflexionar sobre el objetivo del juego y la utilidad de los criterios de optimalidad. Estos criterios permiten evaluar la calidad de una estrategia; sin embargo, el propósito final es identificar la mejor estrategia para cada jugador, aquella que le proporcione el mayor beneficio posible.

Este objetivo se alcanza a través del equilibrio de Nash, el cual varía dependiendo de si el análisis se realiza en términos de costos o recompensas. En un escenario basado en recompensas, los jugadores buscarán maximizar su pago, mientras que, si el criterio es el costo, su objetivo será minimizarlo.

1.5.1. Recompensas

Definición 1.5.1. Una multiestrategia $\pi^* = (\pi_1^*, \pi_2^*) \in \Pi_1 \times \Pi_2$ es un **equilibrio de Nash** (en recompensas) del juego markoviano con horizonte aleatorio si

$$\mathcal{J}_{1,\tau_1}(x, \pi_1^*, \pi_2^*) \geq \mathcal{J}_{1,\tau_1}(x, \pi_1, \pi_2^*)$$

y

$$\mathcal{J}_{2,\tau_2}(x, \pi_1^*, \pi_2^*) \geq \mathcal{J}_{2,\tau_2}(x, \pi_1^*, \pi_2)$$

para todo $\pi_1 \in \Pi_1$, $\pi_2 \in \Pi_2$ y $x \in X$.

Definición 1.5.2. Sea $\pi_2 \in \Pi_2$ una estrategia (fija) para el jugador 2. Se define el **conjunto de mejores respuestas para el jugador 1 a la estrategia** π_2 (en recompensas) como

$$MR_1(\pi_2) = \left\{ \pi_1^* \in \Pi_1 : \mathcal{J}_{1,\tau_1}(x, \pi_1^*, \pi_2) = \sup_{\pi_1 \in \Pi_1} \mathcal{J}_{1,\tau_1}(x, \pi_1, \pi_2) \right\}.$$

Análogamente, el **conjunto de mejores respuestas para el jugador 2 a una estrategia** $\pi_1 \in \Pi_1$ del jugador 1 (en recompensas) se define como

$$MR_2(\pi_1) = \left\{ \pi_2^* \in \Pi_2 : \mathcal{J}_{2,\tau_2}(x, \pi_1, \pi_2^*) = \sup_{\pi_2 \in \Pi_2} \mathcal{J}_{2,\tau_2}(x, \pi_1, \pi_2) \right\}.$$

Una definición equivalente a la Definición 1.5.1 es la siguiente:

Definición 1.5.3. Se dice que una multiestrategia $\pi^* = (\pi_1^*, \pi_2^*) \in \Pi$ es un **equilibrio de Nash** (en recompensas) si

$$\pi_1^* \in MR_1(\pi_2^*) \text{ y } \pi_2^* \in MR_2(\pi_1^*).$$

Equivalentemente, π^* es un equilibrio de Nash (en recompensas) si

$$\mathcal{J}_{1,\tau_1}^*(x) := \mathcal{J}_{1,\tau_1}(x, \pi_1^*, \pi_2^*) = \sup_{\pi_1 \in \Pi_1} \mathcal{J}_{1,\tau_1}(x, \pi_1, \pi_2^*)$$

y

$$\mathcal{J}_{2,\tau_2}^*(x) := \mathcal{J}_{2,\tau_2}(x, \pi_1^*, \pi_2^*) = \sup_{\pi_2 \in \Pi_2} \mathcal{J}_{2,\tau_2}(x, \pi_1^*, \pi_2).$$

Las dos igualdades anteriores se conocen como **funciones de valor del juego** (en recompensas) para cada jugador.

1.5.2. Costos

Definición 1.5.4. Una multiestrategia $\pi^* = (\pi_1^*, \pi_2^*) \in \Pi$ es un **equilibrio de Nash** (en costos) del juego markoviano con horizonte aleatorio si

$$\mathcal{V}_{1,\tau_1}(x, \pi_1^*, \pi_2^*) \leq \mathcal{V}_{1,\tau_1}(x, \pi_1, \pi_2^*)$$

y

$$\mathcal{V}_{2,\tau_2}(x, \pi_1^*, \pi_2^*) \leq \mathcal{V}_{2,\tau_2}(x, \pi_1^*, \pi_2)$$

para todo $\pi_1 \in \Pi_1$, $\pi_2 \in \Pi_2$ y $x \in X$.

Definición 1.5.5. Sea $\pi_2 \in \Pi_2$ una estrategia (fija) para el jugador 2. Se define el **conjunto de mejores respuestas para el jugador 1 a la estrategia** π_2 (en costos) como

$$\mathcal{MR}_1(\pi_2) = \left\{ \pi_1^* \in \Pi_1 : \mathcal{V}_{1,\tau_1}(x, \pi_1^*, \pi_2) = \inf_{\pi_1 \in \Pi_1} \mathcal{V}_{1,\tau_1}(x, \pi_1, \pi_2) \right\}.$$

Análogamente, **el conjunto de mejores respuestas para el jugador 2 a una estrategia** $\pi_1 \in \Pi_1$ del jugador 1 (en costos) se define como

$$\mathcal{MR}_2(\pi_1) = \left\{ \pi_2^* \in \Pi_2 : \mathcal{V}_{2,\tau_2}(x, \pi_1, \pi_2^*) = \inf_{\pi_2 \in \Pi_2} \mathcal{V}_{2,\tau_2}(x, \pi_1, \pi_2) \right\}.$$

Una definición equivalente a la Definición 1.5.4 es la siguiente:

Definición 1.5.6. Se dice que una multiestrategia $\pi^* = (\pi_1^*, \pi_2^*) \in \Pi$ es un **equilibrio de Nash** (en costos) si

$$\pi_1^* \in \mathcal{MR}_1(\pi_2^*) \text{ y } \pi_2^* \in \mathcal{MR}_2(\pi_1^*).$$

Equivalentemente, π^* es un equilibrio de Nash (en costos) si

$$\mathcal{V}_{1,\tau_1}^*(x) := \mathcal{V}_{1,\tau_1}(x, \pi_1^*, \pi_2^*) = \inf_{\pi_1 \in \Pi_1} \mathcal{V}_{1,\tau_1}(x, \pi_1, \pi_2^*)$$

y

$$\mathcal{V}_{2,\tau_2}^*(x) := \mathcal{V}_{2,\tau_2}(x, \pi_1^*, \pi_2^*) = \inf_{\pi_2 \in \Pi_2} \mathcal{V}_{2,\tau_2}(x, \pi_1^*, \pi_2).$$

Las dos igualdades anteriores se conocen como **funciones de valor del juego** (en costos) para cada jugador.

En este capítulo se introdujo la teoría de juegos desde una perspectiva general, comenzando con un repaso histórico y la presentación de definiciones fundamentales. Se expuso el concepto de Equilibrio de Nash, el cual constituye una de las soluciones más reconocidas dentro de esta teoría. Posteriormente, se abordaron los juegos estocásticos y sus principales variantes, lo que permitió sentar las bases para el desarrollo del enfoque adoptado en esta tesis.

En particular, se presentó el modelo de juego markoviano que será utilizado a lo largo del trabajo, describiendo su estructura y propiedades esenciales. Asimismo, se discutió el análisis de estrategias en este tipo de juegos, haciendo énfasis en el criterio de optimalidad como herramienta central para evaluar decisiones en entornos dinámicos.

Finalmente, se estableció una distinción conceptual entre juegos con recompensas y juegos con costos, lo cual resulta relevante para la organización del capítulo siguiente, donde se abordarán juegos markovianos con horizonte aleatorio, estructurados según dicha diferenciación.

Capítulo 2

Juegos Markovianos de suma no cero con horizonte aleatorio

El estudio de juegos markovianos con horizonte aleatorio es importante porque estos modelos reflejan situaciones del mundo real donde la duración del juego no es fija, sino que depende de eventos probabilísticos, uno de los aportes principales en esta teoría surge en [20] donde se introduce el horizonte aleatorio. Además dicho horizonte tiene aplicaciones en economía, finanzas, inteligencia artificial y otros campos donde la incertidumbre en el tiempo de finalización es clave.

Para resaltar la relevancia de estos juegos, se pueden ilustrar algunas aplicaciones en problemas reales.

- Finanzas: Inversiones, seguros o deudas donde el final del juego puede depender de eventos estocásticos (quiebra de una empresa o persona, vencimiento de un contrato) [4].
- Medicina: Tratamientos médicos donde el fin del juego depende de la recuperación del paciente o durante una epidemia donde depende de la propagación de la enfermedad [10].
- Energía y recursos naturales: Gestión de recursos con eventos de interrupción aleatoria, como la coordinación de la carga de un vehículo en múltiples periodos [6].

En estos juegos, el impacto de las estrategias y equilibrios radica en que los jugadores deben adaptar sus decisiones según la probabilidad de que el juego termine en cualquier momento, lo que implica la necesidad de ajustar estrategias. Los equilibrios de Nash se pueden encontrar utilizando técnicas como la inducción hacia atrás estocástica, que permite determinar las mejores decisiones en función de las posibles finalizaciones del juego.

Una vez que se ha dado la interpretación de lo que ocurre en un juego markoviano con horizonte aleatorio, es necesario revisar la literatura existente sobre el tema. En este caso, se está trabajando con un modelo de dos jugadores, pero si se considera el modelo con un solo jugador, en realidad se está tratando con un proceso de decisión de Markov [13, 14, 15]. Por lo tanto, para abordar este tema, los estudios previos sobre procesos de decisión de Markov son de gran utilidad. Por ejemplo, en [37], se analiza un problema en el que el horizonte tiene una distribución geométrica, y se observa que este caso es equivalente a un problema de control óptimo descontado con horizonte infinito, mostrando una equivalencia entre los criterios de rendimiento. Sin embargo, lo que más podría contribuir al enfoque de este estudio es el tratamiento de los procesos de decisión de Markov con horizonte aleatorio, como se presenta en [9, 16], donde se resuelve este tipo de problemas mediante programación dinámica, estos trabajos constituyen una base clave para el desarrollo de esta tesis.

Es importante destacar que la literatura existente sobre los juegos markovianos con horizonte aleatorio es limitada. En [24] se presenta una introducción a los juegos markovianos de suma cero, abordando además el criterio de optimalidad en los contextos de horizonte finito e infinito. Por otro lado, en [11] se analiza una clase de juegos markovianos de suma cero en tiempo discreto y con horizonte aleatorio. No obstante, los juegos de Markov de suma no cero con horizonte aleatorio continúan siendo un problema abierto, lo cual constituye una de las principales motivaciones de este trabajo.

2.1. Recompensas

En esta sección se considera el modelo de juego markoviano de suma no cero con horizonte aleatorio,

$$GM = \{X, (A_i, A_i(x), r_i, \tau_i)_{i=1,2}, Q\}$$

descrito en la Definición 1.3.1 con la notación de (1.2). Además, en este modelo τ_i representa una variable aleatoria discreta para el jugador i, asociada a un espacio de probabilidad $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$. Se asume que la distribución de τ_i es conocida y está dada por $\mathbb{P}_i(\tau_i = t) := \rho_{i,t}$ con $t = 0, 1, 2, ..., T$ donde T es un número entero positivo o $T = \infty$, además τ_1 y τ_2 son independientes.

Dado que se trata de un juego de suma no cero con recompensas, cada jugador busca maximizar su propio beneficio sin que ello implique necesariamente la pérdida del otro, lo que añade complejidad a su análisis y solución.

Para resolver este juego, es necesario analizar las estrategias de cada jugador utilizando un criterio de optimalidad, en este caso será el de *recompensa total esperada descontada con horizonte aleatorio* para el jugador i presentado en la ecuación (1.3):

$$\mathcal{J}_{i,\tau_i}\left(x,\pi_1,\pi_2\right)=\mathbb{E}_i\left[\sum_{t=0}^{\tau_i}\beta_i^t r_i\left(x_t,a_{1,t},a_{2,t}\right)\right],$$

con $x\in X$, $(\pi_1,\pi_2)\in\Pi_1\times\Pi_2$ y $\beta_i\in(0,1)$.

El objetivo es determinar una multiestrategia $\pi^*=(\pi_1^*,\pi_2^*)\in\Pi_1\times\Pi_2$ que constituya un equilibrio de Nash para el juego. Para ello, se empleará programación dinámica, cuyo desarrollo se presentará en la siguiente subsección.

Antes de abordar la programación dinámica, se analizará una observación clave sobre la relación entre los distintos criterios de optimalidad, considerando los casos de horizonte aleatorio, finito e infinito, lo que permitirá una mejor comprensión del problema y sus soluciones.

Suposición 2.1.1. Para cada $x\in X$ y $\pi\in\Pi$ el proceso inducido $\{(x_t,a_{1,t},a_{2,t})|t=0,1,2,...\}$ es independiente de τ_i.

Lema 2.1.2. (a) Bajo la Suposición 2.1.1 se tiene que

$$\begin{aligned}\mathcal{J}_{i,\tau_i}\left(x,\pi_1,\pi_2\right)&=\mathbb{E}_i\left[\sum_{t=0}^{\tau_i}\beta_i^t r_i\left(x_t,a_{1,t},a_{2,t}\right)\right]\\&=\mathbb{E}_i\left[\mathbb{E}_i\left[\sum_{t=0}^{\tau_i}\beta_i^t r_i\left(x_t,a_{1,t},a_{2,t}\right)\middle|\tau_i\right]\right]\\&=\sum_{n=0}^{T}\mathbb{E}_x^{\pi_1,\pi_2}\left[\sum_{t=0}^{n}\beta_i^t r_i\left(x_t,a_{1,t},a_{2,t}\right)\right]\rho_{i,n}\\&=\sum_{t=0}^{T}\sum_{n=t}^{T}\mathbb{E}_x^{\pi_1,\pi_2}\left[\beta_i^t r_i\left(x_t,a_{1,t},a_{2,t}\right)\right]\rho_{i,n}\\&=\mathbb{E}_x^{\pi_1,\pi_2}\left[\sum_{t=0}^{T}P_{i,t}\beta_i^t r_i\left(x_t,a_{1,t},a_{2,t}\right)\right],\end{aligned}\tag{2.1}$$

$(\pi_1,\pi_2)\in\Pi_1\times\Pi_2, x\in X$, donde $P_{i,t}:=\sum_{n=t}^{T}\rho_{i,n}=\mathbb{P}_i(\tau_i\geq t), t=0,1,2,\ldots,T$. Con lo anterior se muestra una equivalencia entre un criterio descontado con horizonte aleatorio τ_i y un el criterio descontado con horizonte T, con recompensa por etapa no homogénea dado por $P_{i,t}\beta_i^t r_i$ y recompensa terminal igual a cero.

(b) De esta manera, el criterio de optimalidad descontado con horizonte aleatorio se divide en dos casos, cuando

- $T<\infty$ (soporte finito)
- $T=\infty$ (soporte infinito)

ambos casos se resuelven por medio de diferentes técnicas, durante este trabajo se analizará el caso en el cual $T < \infty$.

Observación 2.1.3. Cuando se tiene el criterio no descontado con horizonte aleatorio en recompensas y el horizonte tiene distribución geométrica para el jugador i, con parámetro p_i, $0 < p_i < 1$, se tiene que $\mathbb{P}_i(\tau_i = n) = p_i q_i^n$, $n = 0, 1, 2, \ldots$, donde $q_i = 1 - p_i$. Por lo tanto,

$$\begin{aligned} P_{i,t} &= \sum_{n=t}^{\infty} p_i q_i^n \\ &= p_i q_i^t \sum_{n=0}^{\infty} q_i^n \\ &= q_i^t \end{aligned}$$

con $t \in \mathbb{N}_0$. Entonces

$$\begin{aligned} \mathcal{J}_{i,\tau_i}(x, \pi_1, \pi_2) &= \mathbb{E}_i \left[\sum_{t=0}^{\tau_i} r_i(x_t, a_{1,t}, a_{2,t}) \right] \\ &= \mathbb{E}_x^{\pi_1,\pi_2} \left[\sum_{t=0}^{T} P_{i,t} r_i(x_t, a_{1,t}, a_{2,t}) \right] \\ &= \mathbb{E}_x^{\pi_1,\pi_2} \left[\sum_{t=0}^{T} q_i^t r_i(x_t, a_{1,t}, a_{2,t}) \right]. \end{aligned}$$

De este modo, se observa que se obtiene el criterio descontado con horizonte finito con factor de descuento q_i, similar a como se afirma en [37]. De este modo, se sustenta la idea de que un problema de decisión de Markov con horizonte aleatorio, incluso si tiene una distribución diferente, puede analizarse como un problema descontado no homogéneo con horizonte finito, siendo esta la forma en que se ha abordado el problema.

Similarmente, cuando se analizan costos no descontados y el horizonte aleatorio se distribuye geométricamente entonces el criterio puede verse como un descontado con horizonte finito.

2.1.1. Programación dinámica

La programación dinámica [3] es una técnica matemática utilizada para resolver problemas de optimización secuenciales, en los cuales las decisiones presentes afectan los estados futuros del sistema. Su aplicación en Procesos de Decisión de Markov [13, 14, 15] permite modelar sistemas donde la evolución futura depende únicamente del estado actual y la acción tomada, siguiendo una estructura probabilística.

En particular, dentro del marco de los juegos estocásticos [16], esta herramienta permite identificar las mejores estrategias para los jugadores mediante las ecuaciones de Bellman, las cuales han pasado a conocerse actualmente como ecuaciones de programación dinámica, adaptadas a contextos con múltiples agentes.

Las ecuaciones de programación dinámica son clave en este análisis, ya que permiten descomponer un problema en subproblemas más pequeños, cuya solución conjunta lleva al resultado deseado. Intuitivamente, estas ecuaciones evalúan el estado actual considerando todas las posibles transiciones futuras, las ponderan por su probabilidad y calculan una expectativa de la recompensa futura de cada decisión. En juegos estocásticos, esta expectativa se ajusta según la interacción entre jugadores, sus estrategias y la incertidumbre del entorno.

Este enfoque de programación dinámica, aplicado a juegos estocásticos y en particular a juegos markovianos de suma no cero con horizonte aleatorio, proporciona herramientas clave para el análisis y diseño de mejores estrategias en entornos dinámicos e inciertos.

2.1.2. Solución por programación dinámica en recompensas

Es importante volver a mencionar que el criterio de optimalidad a analizar corresponde al de recompensa total esperada descontada con un horizonte aleatorio descrito en (1.3), el cual se redujo a

$$\mathcal{J}_{i,\tau_i}(x,\pi_1,\pi_2) = \mathbb{E}_x^{\pi_1,\pi_2}\left[\sum_{t=0}^{T} P_{i,t}\beta_i^t r_i(x_t,a_{1,t},a_{2,t})\right]$$

gracias al Lema 2.1.2 con $\pi \in \Pi$, $x \in X$, $0 < \beta_i < 1$, $P_{i,t} = \mathbb{P}_i(\tau_i \geq t), t = 0,1,2,\ldots,T$ y T finito.

Se considera la siguiente suposición.

Suposición 2.1.4.

(a) $A_i(x)$ es compacto para cada $x \in X$ y $i = 1,2$.

(b) La función de recompensa por etapa para el jugador i, $r_i(x,a_1,a_2)$, es semicontinua superiormente en $(a_1,a_2) \in A_1(x) \times A_2(x)$.

(c) La ley de transición Q es débilmente continua en $(a_1,a_2) \in A_1(x) \times A_2(x)$.

El siguiente teorema proporciona la ecuación de programación dinámica, la cual será de gran utilidad para resolver los juegos de suma no cero con horizonte aleatorio.

Teorema 2.1.5. Sean $J_{i,0}, J_{i,1}, ..., J_{i,T+1}$ funciones sobre X para el jugador i definidas por

$$J_{i,T+1}(x) := 0, \quad x \in X$$

y para $t = T, T-1, ..., 0$,

$$J_{i,t}(x) := \max_{a_i \in A_i(x)} \left[P_{i,t}\beta_i^t r_i(x, a_1, a_2) + \int_X J_{i,t+1}(y) Q(dy|x, a_1, a_2) \right], \quad x \in X. \quad (2.2)$$

Bajo la Suposición 2.1.4, estas funciones son medibles en $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$ y para cada $t = 0, 1, ..., T$, existe $f_{i,t}^* \in \mathbb{F}_i$ para cada jugador tal que $f_{i,t}^*(x) \in A_i(x)$ alcanza el máximo en la ecuación (2.2) para todo $x \in X$, esto es

$$J_{i,t}(x) = P_{i,t}\beta_i^t r_i(x, f_{1,t}^*(x), f_{2,t}^*(x)) + \int_X J_{i,t+1}(y) Q(dy|x, f_{1,t}^*(x), f_{2,t}^*(x)),$$

$x \in X$ y $t = 0, 1, ..., T$. Entonces la multiestrategia $\pi^* = (\pi_1^*, \pi_2^*) \in \Pi_1 \times \Pi_2$ con $\pi_i^* = \{f_{i,0}^*, f_{i,1}^*, ..., f_{i,T}^*\}$ cumple:

$$\mathcal{J}_{i,\tau_i}(x, \pi_1^*, \pi_2^*) = \mathcal{J}_{i,\tau_i}^*(x) = J_{i,0}(x), \quad \forall x \in X,$$

además, π^* es un equilibrio de Nash.

Demostración. La demostración se desarrolla del siguiente modo. En primer lugar, se resuelve el problema desde la perspectiva del jugador 1, bajo el supuesto de que ambos jugadores actúan de manera económicamente racional. En consecuencia, se asume que el jugador 2 emplea su estrategia π_2^*, y el jugador 1 toma sus decisiones considerando dicha estrategia. Es importante destacar que las estrategias π_1^* y π_2^* existen, dado que corresponden a sucesiones finitas de funciones cuya existencia está garantizada por la Suposición 2.1.4.

Sean $\pi_1 = \{a_{1,t}\}_{t=0}^T \in \Pi_1$ una estrategia arbitraria para el jugador 1 y $\pi_2^* = \{f_{2,t}^*\}_{t=0}^T \in \mathbb{F}_2$ una estrategia fija para el jugador 2. Además, defina

$$R_{1,t}(x, \pi_1, \pi_2^*) := \mathbb{E}^{\pi_1, \pi_2^*} \left[\sum_{n=t}^{T} P_{1,n}\beta_1^n r_1(x_n, a_{1,n}, f_{2,n}^*) \middle| X_t = x \right], \quad t = 0, 1, \ldots, T,$$

$$R_{1,T+1}(x, \pi_1, \pi_2^*) := 0.$$

Observe que

$$R_{1,0}(x, \pi_1, \pi_2^*) = \mathbb{E}^{\pi_1, \pi_2^*} \left[\sum_{n=0}^{T} P_{1,n}\beta_1^n r_1(x_n, a_{1,n}, f_{2,n}^*) \middle| X_t = x \right] = \mathcal{J}_{1,\tau_1}(x, \pi_1, \pi_2^*) \quad (2.3)$$

Cabe resaltar que $R_{1,t}$ es conocida como la recompensa del tiempo t en adelante.

A continuación se demostrará que se cumplen los siguientes resultados:

1) $R_{1,t}(x,\pi_1,\pi_2^*) \leq J_{1,t}(x), \quad t=0,\dots,T+1.$

2) Si $\pi_1=\pi_1^*$, $R_{1,t}(x,\pi_1^*,\pi_2^*) = J_{1,t}(x)$.

1) Se resolverá por inducción hacia atrás sobre t.
Base inductiva. Para $t=T+1$ se tiene que

$$R_{1,T+1}(x,\pi_1,\pi_2^*) = 0 = J_{1,T+1}(x)$$

por definición de $R_{1,t}$ y $J_{1,t}$.
Paso inductivo. Supóngase que para algún $t=T,T-1,\dots,0$ se cumple, es decir,

$$R_{1,t+1}(x,\pi_1,\pi_2^*) \leq J_{1,t+1}(x), \quad \forall x \in X.$$

Ahora,

$$\begin{aligned}
&R_{1,t}(x,\pi_1,\pi_2^*)\\
&= \mathbb{E}^{\pi_1,\pi_2^*}\left[\sum_{n=t}^{T} P_{1,n}\beta_1^n r_1(x_n,a_{1,n},f_{2,n}^*)\Big| X_t=x\right]\\
&= \mathbb{E}^{\pi_1,\pi_2^*}\left[P_{1,t}\beta_1^t r_1(x_t,a_{1,t},f_{2,t}^*) + \sum_{n=t+1}^{T} P_{1,t}\beta_1^n r_1(x_n,a_{1,n},f_{2,n}^*)\Big| X_t=x\right]\\
&= \int_{A_1}\left[P_{1,t}\beta_1^t r_1(x,a_1,f_2^*) + \int_X R_{1,t+1}(y,\pi_1,\pi_2^*)Q(dy|x,a_1,f_2^*)\right]\pi_{1,t}(da_1|x,f_2^*)
\end{aligned}$$

Así, por hipótesis de inducción,

$$\begin{aligned}
&R_{1,t}(x,\pi_1,\pi_2^*)\\
&\leq \int_{A_1}\left[P_{1,t}\beta_1^t r_1(x,a_1,f_2^*) + \int_X J_{1,t+1}(y)Q(dy|x,a_1,f_2^*)\right]\pi_{1,t}(da_1|x,f_2^*)\\
&\leq \left(\int_{A_1}\pi_{1,t}(da_1|x,df_2^*)\right)\max_{a_1\in A_1(x)}\left[P_{1,t}\beta_1^t r_1(x,a_1,f_2^*) + \int_X J_{1,t+1}(y)Q(dy|x,a_1,f_2^*)\right]\\
&= \max_{a_1\in A_1(x)}\left[P_{1,t}\beta_1^t r_1(x,a_1,f_2^*) + \int_X J_{1,t+1}(y)Q(dy|x,a_1,f_2^*)\right]\\
&= J_{1,t}(x).
\end{aligned}$$

Por lo tanto,

$$R_{1,t}(x,\pi_1,\pi_2^*) \leq J_{1,t}(x).$$

2) Por otro lado, si $R_{1,t+1}(x,\pi_1,\pi_2^*) = J_{1,t+1}(x), \quad x\in X$ con $\pi_1=\pi_1^*$, $\pi_{1,t}(\cdot|h_t)$ es la medida de Dirac concentrada en $f_{1,t}^*(x_t)$, entonces se mantiene la igualdad en los cálculos anteriores, obteniendo

$$R_{1,t}(x, \pi_1^*, \pi_2^*) = J_{i,t}(x)$$

Luego, como $R_{1,t}(x, \pi_1, \pi_2^*) \leq J_{1,t}(x)$, en particular para $t = 0$ y de (2.3) se tiene que

$$\mathcal{J}_{1,\tau_1}(x, \pi_1, \pi_2^*) \leq J_{1,0}(x)$$

y para $\pi_1 = \pi_1^*$,

$$\mathcal{J}_{1,\tau_1}(x, \pi_1^*, \pi_2^*) = J_{1,0}(x). \tag{2.4}$$

De manera análoga, el procedimiento se aplica para el jugador 2, considerando que, al actuar de forma económicamente racional, el jugador 1 empleará su estrategia π_1^*. Bajo esta premisa, se obtiene:

$$\mathcal{J}_{2,\tau_2}(x, \pi_1^*, \pi_2^*) = J_{2,0}(x). \tag{2.5}$$

Así, en virtud de las ecuaciones (2.4) y (2.5), y dado que el máximo existe, este coincide con el supremo, lo que permite concluir que:

$$\mathcal{J}^*_{1,\tau_1}(x) = \sup_{\pi_1 \in \Pi_1} \mathcal{J}_{1,\tau_1}(x, \pi_1, \pi_2^*)$$

y

$$\mathcal{J}^*_{2,\tau_2}(x) = \sup_{\pi_2 \in \Pi_2} \mathcal{J}_{2,\tau_2}(x, \pi_1^*, \pi_2),$$

así $\pi^* = (\pi_1^*, \pi_2^*)$ corresponde a un equilibrio de Nash.

■

Observación 2.1.6. Las funciones $J_{i,0}, J_{i,1}, ..., J_{i,T+1}$ del Teorema 2.1.5, son fundamentales para la determinación de la solución del juego, pues permiten construir una estrategia que constituye un equilibrio de Nash en el caso de recompensas.

Observación 2.1.7. Para facilitar el desarrollo subsecuente, se hace el siguiente cambio de variable.
Sea $U_{i,T+1}(x) := 0$ para todo $x \in X$ y

$$U_{i,t} := \frac{J_{i,t}}{P_{i,t}\beta_i^t}$$

$t \in \{0, 1, 2, ..., T-1\}$. Entonces sustituyendo en (2.2) se obtiene la siguiente expresión

$$P_{i,t}\beta_i^t U_{i,t}(x) = \max_{a_i \in A_i(x)} \left[P_{i,t}\beta_i^t r_i(x, a_1, a_2) + \int_X P_{i,t+1}\beta_i^{t+1} U_{i,t+1}(y) Q(dy|x, a_1, a_2) \right]$$

de donde, considerando que $P_{i,t}\beta_i^t > 0$, se obtiene que

$$U_{i,t}(x) = \max_{a_i \in A_i(x)} \left[r_i(x, a_1, a_2) + \frac{P_{i,t+1}}{P_{i,t}} \beta_i \int_X U_{i,t+1}(y) Q(dy|x, a_1, a_2) \right].$$

Entonces la ecuación de programación dinámica (2.2) es equivalente a

$$U_{i,t}(x) = \max_{a_i \in A_i(x)} \left[r_i(x, a_1, a_2) + \alpha_{i,t}\, \beta_i \int_X U_{i,t+1}(y) Q(dy|x, a_1, a_2) \right] \tag{2.6}$$

donde

$$\alpha_{i,t} := \frac{P_{i,t+1}}{P_{i,t}} \quad t \in \{0, 1, 2, ..., T-1\}.$$

Note que puede obtenerse la siguiente interpretación probabilística:

$$\begin{aligned} \alpha_{i,t} &= \frac{\mathbb{P}_i(\tau_i \geq t+1, \tau_i \geq t)}{\mathbb{P}_i(\tau_i \geq t)} \\ &= \mathbb{P}_i(\tau_i \geq t+1 | \tau_i \geq t). \end{aligned}$$

Observación 2.1.8. Existen diversas variantes de la ecuación de programación dinámica. Por ejemplo, considerando la dinámica de un juego estocástico de dos jugadores, descrita por la siguiente ecuación,

$$x_{t+1} = F(x_t, a_{1,t}, a_{2,t}, \xi_t)$$

equivalentemente (2.6) se puede expresar como

$$U_{i,t}(x) = \max_{a_i \in A_i(x)} \left[r_i(x, a_1, a_2) + \alpha_{i,t}\, \beta_i \mathbb{E} \left[U_{i,t+1}(F(x, a_1, a_2, \xi_t)) \right] \right], \tag{2.7}$$

para $t = 0, 1, ..., T$ teniendo en cuenta que $U_{i,T+1}(x) = 0$ [14].

2.2. Costos

En esta sección se lleva a cabo un procedimiento completamente análogo al de la sección de recompensas, pero aplicado al criterio de costo total esperado con horizonte aleatorio. La principal diferencia es que, en vez de maximizar recompensas, los jugadores buscan minimizar costos. Además, en este caso, se trabaja con costos no descontados.

Por consiguiente, se considera el modelo de juego markoviano de suma no cero con horizonte aleatorio,

$$GM = \{X, (A_i, A_i(x), c_i, \tau_i)_{i=1,2}, Q\}$$

descrito en la Definición 1.3.1 con la notación de (1.2). Además, en este modelo τ_i representa una variable aleatoria discreta para el jugador i, asociada a un espacio de probabilidad $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$. Se asume que la distribución de τ_i es conocida y está dada por $\mathbb{P}_i(\tau_i = t) := \rho_{i,t}$ con $t = 0, 1, 2, ..., T$ donde T es un número entero positivo o $T = \infty$, además τ_1 y τ_2 son independientes.

Para resolver este juego, es necesario analizar las estrategias de cada jugador utilizando un criterio de optimalidad, en este caso será el de *costo total esperado con horizonte aleatorio* para el jugador i presentado en la ecuanción (1.4):

$$\mathcal{V}_{i,\tau_i}(x, \pi_1, \pi_2) = \mathbb{E}_i \left[\sum_{t=0}^{\tau_i} c_i(x_t, a_{1,t}, a_{2,t}) \right],$$

con $x \in X$, $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$.

El objetivo es determinar una multiestrategia $\pi^* = (\pi_1^*, \pi_2^*) \in \Pi_1 \times \Pi_2$ que constituya un equilibrio de Nash para el juego. Para lograrlo, se recurrirá nuevamente a la programación dinámica.

De manera similar a lo realizado en la sección de recompensas, se analizará una observación sobre la relación entre los criterios de optimalidad: aleatorio, finito e infinito.

Lema 2.2.1. (a) Bajo la Suposición 2.1.1 se tiene que

$$\begin{aligned}
\mathcal{V}_{i,\tau_i}(x, \pi_1, \pi_2) &= \mathbb{E}_i \left[\sum_{t=0}^{\tau_i} c_i(x_t, a_{1,t}, a_{2,t}) \right] \\
&= \mathbb{E}_i \left[\mathbb{E}_i \left[\sum_{t=0}^{\tau_i} c_i(x_t, a_{1,t}, a_{2,t}) \middle| \tau_i \right] \right] \\
&= \sum_{n=0}^{T} \mathbb{E}_x^{\pi_1,\pi_2} \left[\sum_{t=0}^{n} c_i(x_t, a_{1,t}, a_{2,t}) \right] \rho_{i,n} \\
&= \sum_{t=0}^{T} \sum_{n=t}^{T} \mathbb{E}_x^{\pi_1,\pi_2} \left[c_i(x_t, a_{1,t}, a_{2,t}) \right] \rho_{i,n} \\
&= \mathbb{E}_x^{\pi_1,\pi_2} \left[\sum_{t=0}^{T} P_{i,t} c_i(x_t, a_{1,t}, a_{2,t}) \right], \qquad (2.8)
\end{aligned}$$

$(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2, x \in X$, donde $P_{i,t} := \sum_{n=t}^{T} \rho_{i,n} = \mathbb{P}_i(\tau_i \geq t), t = 0, 1, 2, \ldots, T$. Con lo anterior se muestra una equivalencia entre un criterio con horizonte aleatorio τ_i y un criterio descontado con horizonte T, con costo por etapa no homogéneo dado por $P_{i,t} c_i$ y costo terminal igual a cero.

(b) De esta manera, el criterio de optimalidad con horizonte aleatorio se divide en dos casos, cuando

- $T < \infty$ (soporte finito)

- $T = \infty$ (soporte infinito)

Al igual que en el caso de recompensas, ambos enfoques requieren métodos de resolución distintos; en lo que sigue, se continuará el análisis restringiéndolo al caso en el cual $T < \infty$.

2.2.1. Solución por programación dinámica en costos

Conviene reiterar que el criterio de optimalidad a examinar corresponde al costo total esperado con un horizonte aleatorio, tal como se describe en (1.4), y que ha sido reducido a

$$\mathcal{V}_{i,\tau_i}(x, \pi_1, \pi_2) = \mathbb{E}_x^{\pi_1,\pi_2}\left[\sum_{t=0}^{T} P_{i,t} c_i(x_t, a_{1,t}, a_{2,t})\right]$$

en Lema 2.2.1 con $(\pi_1, \pi_2) \in \Pi_1 \times \Pi_2$, $x \in X$, $P_{i,t} = \mathbb{P}_i(\tau_i \geq t), t = 0, 1, 2, \ldots, T$ y T finito.

Se considera la siguiente suposición.

Suposición 2.2.2. Considere la Suposición 2.1.4 incisos (a) y (c), sustituyendo el inciso (b) por el siguiente:

(b′) La función de costo por etapa para el jugador $i = 1, 2$, $c_i(x, a_1, a_2)$, es semicontinua inferiormente en $(a_1, a_2) \in A_1(x) \times A_2(x)$.

El siguiente teorema proporciona la ecuación de programación dinámica en costos.

Teorema 2.2.3. Sean $V_{i,0}, V_{i,1}, ..., V_{i,T+1}$ funciones sobre X para el jugador i definidas por

$$V_{i,T+1}(x) := 0, \quad x \in X$$

y para $t = T, T-1, ..., 0$,

$$V_{i,t}(x) := \min_{a_i \in A_i(x)}\left[P_{i,t}c_i(x, a_1, a_2) + \int_X V_{i,t+1}(y)Q(dy|x, a_1, a_2)\right], \quad x \in X. \tag{2.9}$$

Bajo la Suposición 2.2.2, estas funciones son medibles en $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$ y para cada $t = 0, 1, ..., T$, existe $f_{i,t}^* \in \mathbb{F}_i$ para cada jugador tal que $f_{i,t}^*(x) \in A_i(x)$ alcanza el mínimo en la ecuación (2.9) para todo $x \in X$, esto es

$$V_{i,t}(x) = P_{i,t}c_i(x, f_{1,t}^*(x), f_{2,t}^*(x)) + \int_X V_{i,t+1}(y)Q(dy|x, f_{1,t}^*(x), f_{2,t}^*(x)),$$

$x \in X$ y $t = 0, 1, ..., T$. Entonces la multiestrategia $\pi^* = (\pi_1^*, \pi_2^*) \in \Pi_1 \times \Pi_2$ con $\pi_i^* = \{f_{i,0}*, f_{i,1}^*, ..., f_{i,T}^*\}$ cumple:

$$\mathcal{V}_{i,\tau_i}(x, \pi_1^*, \pi_2^*) = \mathcal{V}_{i,\tau_i}^*(x) = V_{i,0}(x), \quad \forall x \in X.$$

además, π^* es un equilibrio de Nash.

Demostración. La demostración es análoga a la del Teorema 2.1.5. ■

Observación 2.2.4. Las funciones $V_{i,0}, V_{i,1}, ..., V_{i,T+1}$, presentadas en el Teorema 2.2.3 para el caso de costos, cumplen un rol equivalente en la caracterización del equilibrio al que desempeñan las funciones $J_{i,0}, J_{i,1}, ..., J_{i,T+1}$ en el caso de recompensas.

Observación 2.2.5. Para facilitar el desarrollo subsecuente, se hace el siguiente cambio de variable.
Sea $W_{i,0}(x) := 0$ para todo $x \in X$ y

$$W_{i,t} := \frac{V_{i,t}}{P_{i,t}}$$

$t \in \{1, 2, ..., T\}$. Entonces sustituyendo en (2.9) se obtiene la siguiente expresión

$$P_{i,t}W_{i,t}(x) = \min_{a_i \in A_i(x)} \left[P_{i,t}c_i(x, a_1, a_2) + \int_X P_{i,t+1}W_{i,t+1}(y)Q(dy|x, a_1, a_2) \right]$$

de donde, considerando que $P_{i,t} > 0$, se obtiene que

$$W_{i,t}(x) = \min_{a_i \in A_i(x)} \left[c_i(x, a_1, a_2) + \frac{P_{i,t+1}}{P_{i,t}} \int_X W_{i,t+1}(y)Q(dy|x, a_1, a_2) \right].$$

Entonces la ecuación de programación dinámica (2.9) es equivalente a

$$W_{i,t}(x) = \min_{a_i \in A_i(x)} \left[c_i(x, a_1, a_2) + \gamma_{i,t} \int_X W_{i,t+1}(y)Q(dy|x, a_1, a_2) \right] \tag{2.10}$$

donde

$$\gamma_{i,t} := \frac{P_{i,t+1}}{P_{i,t}} \qquad t \in \{0, 1, 2, ..., T-1\}.$$

Note que puede obtenerse la siguiente interpretación probabilística:

$$\begin{aligned} \gamma_{i,t} &= \frac{\mathbb{P}_i(\tau_i \geq t, \tau_i \geq t+1)}{\mathbb{P}_i(\tau_i \geq t+1)} \\ &= \mathbb{P}_i(\tau_i \geq t | \tau_i \geq t+1). \end{aligned}$$

Observación 2.2.6. Dada la dinámica de un juego estocástico de dos jugadores

$$x_{t+1} = F(x_t, a_{1,t}, a_{2,t}, \xi_t)$$

equivalentemente (2.10) se puede expresar como

$$W_{i,t}(x) = \min_{a_i \in A_i(x)} \left[c_i(x, a_1, a_2) + \gamma_{i,t}\, \mathbb{E}\left[W_{i,t+1}(F(x, a_1, a_2, \xi_t)) \right] \right], \tag{2.11}$$

para $t = 1, ..., T$ teniendo en cuenta que $W_{i,T+1}(x) = 0$ [14].

En este capítulo se abordó la relevancia del horizonte aleatorio en el estudio de juegos markovianos, estructurando el contenido en dos secciones, una centrada en juegos con recompensas y otra en juegos con costos. En ambos casos, se consideró el modelo markoviano introducido en el capítulo anterior, adaptándolo al contexto de horizonte aleatorio y a juegos de suma no cero.

Particular atención se prestó al análisis del criterio de optimalidad bajo este nuevo enfoque, lo cual requirió el establecimiento de una serie de suposiciones clave. Estas suposiciones permitieron aplicar un teorema fundamental del capítulo, el cual introduce funciones que constituyen una herramienta central para identificar equilibrios de Nash en este tipo de juegos.

Las funciones obtenidas mediante este enfoque serán fundamentales en el capítulo siguiente, donde se aplicarán para resolver casos específicos: uno relacionado con juegos en el contexto de pesquerías, y otro en un entorno lineal-cuadrático (LQ).

Capítulo 3

Ejemplos

En este capítulo, se aplicarán los conceptos abordados en los Capítulos 1 y 2 para resolver dos ejemplos de juegos markovianos de suma no cero con horizonte aleatorio. El primer ejemplo se desarrollará bajo el enfoque de recompensas descontadas y abordará el problema del juego de la gran pesca, mientras que el segundo se basará en costos no descontados y analizará el modelo lineal-cuadrático (LQ). En ambos casos, se determinan equilibrios de Nash de forma explícita.

3.1. Juego de la gran pesca

En las últimas décadas, se han documentado diversos conflictos internacionales vinculados con los derechos de pesca y la explotación de cuerpos de agua compartidos. Estos enfrentamientos surgen cuando múltiples actores buscan acceder a recursos acuáticos limitados, generando tensiones sobre su uso y conservación. Ejemplos notables incluyen la Guerra del Fletán entre Canadá y España en 1995 [25], la Guerra del Salmón del Pacífico en la década de 1990 [23], y disputas actuales en regiones como Yucatán [8], donde pescadores locales han confrontado a foráneos por prácticas de pesca no reguladas de pulpo.

Uno de los conflictos más representativos en este ámbito es la denominada Guerra del Bacalao entre Islandia y el Reino Unido [12], ocurrida entre los años 1950 y 1970. Este enfrentamiento, centrado en la ampliación unilateral de las zonas de pesca por parte de Islandia, no solo tuvo implicaciones geopolíticas, sino que también inspiró estudios en teoría de juegos para analizar situaciones donde múltiples entidades buscan maximizar beneficios al extraer recursos de un cuerpo de agua limitado, evitando su agotamiento. En este contexto, Levhari y Mirman en [21] introdujeron un modelo para representar estos conflictos pesqueros, conocido desde ese entonces como “The Great Fish War” o “ El juego de la gran pesca”.

El estudio del juego de la gran pesca es relevante debido a la persistencia de conflictos similares en la actualidad, como la ya mencionada en Yucatán. Asimismo, la estructura de nuestro modelo podría adaptarse al análisis de otros recursos renovables, como los bosques, donde también se presentan problemáticas asociadas a la explotación compartida y las cuales ya se han analizado con un enfoque de la teoría de juegos [19].

En el estudio del Juego de la Gran Pesca, la noción de horizonte temporal es fundamental para definir la mejor estrategia de extracción de recursos. El horizonte finito representa situaciones en las que los jugadores conocen de antemano la duración del juego, lo cual introduce un incentivo a agotar el recurso conforme se acerca el final del periodo de decisión, fenómeno conocido como efecto de último periodo. En contraste, el horizonte infinito modela un entorno donde el juego se prolonga indefinidamente, permitiendo estrategias más cooperativas y sostenibles, dado que los jugadores internalizan el impacto de sus acciones sobre el futuro del recurso. Finalmente, el horizonte aleatorio introduce una incertidumbre sobre la duración del juego, ya que este puede finalizar abruptamente debido a factores exógenos como eventos ambientales, colapso del recurso o decisiones políticas. Esta formulación más realista fuerza a los jugadores a balancear los beneficios inmediatos con el riesgo de que no haya oportunidades futuras para explotar el recurso, promoviendo estrategias que integran tanto la explotación eficiente como la preservación del sistema ecológico.

Es relevante destacar parte de la literatura existente en la que se fundamenta este trabajo. Levhari y Mirman [21] además de introducir el modelo del juego de la gran pesca, analizaron el problema mediante juegos dinámicos con horizonte finito así como su relación con juegos descontados de horizonte infinito. Entre los desarrollos posteriores, se encuentran los trabajos de Nowak como [17, 32], donde se estudian las condiciones necesarias para la existencia de equilibrios de Nash en versiones no cooperativas del juego de la gran pesca con suma no cero. A su vez, investigaciones más recientes como las de [5, 33] han extendido el modelo a contextos de acuicultura, formulando problemas de maximización de utilidad sujetos a funciones de producción pesquera, los cuales se resuelven mediante técnicas basadas en la ecuación de Euler, estas contribuciones también han explorado formulaciones en tiempo discreto con horizonte infinito y han incorporado enfoques difusos para modelar la incertidumbre inherente a los sistemas pesqueros.

Planteamiento del juego

Supóngase una situación en la que dos países están interesados en la extracción de recursos pesqueros provenientes de una misma región marítima. Cada uno de estos países posee su propia función de utilidad, la cual depende de la cantidad de peces capturados en cada periodo. Ambos actores buscan maximizar la esperanza de la suma total de sus utilidades descontadas a lo largo del tiempo considerando un horizonte aleatorio. Dado que se asume que ambos países actúan de manera económicamente racional, cada uno considera las posibles acciones del otro al momento de tomar decisiones, con el fin de asegurar la mejor estrategia de extracción posible.

Este problema se describe mediante el modelo de juego estocástico que se muestra a continuación.

El espacio de estados está dado por $X = [0, \infty)$, representando el volumen de peces presentes en el ecosistema compartido por ambos jugadores. Por su parte, los conjuntos de acciones de cada jugador están definidos como $A_1 = A_2 = [0, \infty)$, lo cual representa las posibles cantidades de peces que pueden ser extraídas en un periodo determinado.

En cada estado del sistema, el crecimiento natural del recurso biológico se modela mediante la función de crecimiento poblacional de los peces, denotada por $h : X \to [0, \infty)$, definida como $h(x) = x^{\delta}$, con $\delta > 0$. Esta función describe el crecimiento natural del recurso biológico en función de la población existente $x \in X$. A partir de esta función de crecimiento, se asignan las acciones admisibles de cada jugador mediante los conjuntos $A_1(x) = A_2(x) = \left[0, \frac{x^{\delta}}{M}\right]$ donde $M \geq 1$ y $\frac{1}{M}$ representa la proporción de peces de la biomasa que puede ser extraída para una unidad de biomasa. Esto significa que la cantidad máxima de peces que cada país puede extraer en un determinado periodo depende directamente del tamaño actual de la población, determinada por el estado $x \in X$. En otras palabras, la función de crecimiento no solo modela la regeneración del recurso, sino que también impone una restricción biológica sobre las decisiones de explotación, asegurando que la extracción no exceda la capacidad natural del ecosistema en cada momento.

De este modo, la dinámica que rige la evolución del sistema de pesca compartido entre los dos países se describe mediante la siguiente ecuación en diferencias:

$$x_{t+1} = F(x_t, a_{1,t}, a_{2,t}, \xi_t) = h(x_t - a_{1,t} - a_{2,t})\xi_t = (x_t - a_{1,t} - a_{2,t})^{\delta}\xi_t,$$

donde $x_0 = x$ representa el estado inicial conocido del sistema y $a_{i,t} \in A_i(x_t)$ es la cantidad de peces extraída por el jugador i en el periodo $t \in \mathbb{N}_0$. El término x_t representa el stock de peces disponible en dicho periodo. Se impone la condición $x_t > a_{1,t} + a_{2,t}$ casi seguramente (c.s.) con respecto a $\mathbb{P}_x^{\pi_1,\pi_2}$, lo cual garantiza que la suma de las capturas de ambos países no exceda la población disponible en el ecosistema en ese momento. Esta ecuación captura la evolución de la población de peces en función de las decisiones de extracción tomadas por los dos jugadores. El término $h(x_t - a_{1,t} - a_{2,t}) = (x_t - a_{1,t} - a_{2,t})^{\delta}$ representa la cantidad de peces que sobrevive o permanece en el ecosistema después de la extracción conjunta, mientras que el factor multiplicativo aleatorio ξ_t introduce una componente estocástica que refleja perturbaciones ambientales o biológicas, tales como fluctuaciones en la tasa natural de mortalidad, enfermedades, migraciones o condiciones adversas. Se considera que la secuencia $\{\xi_t\}$ está compuesta por variables aleatorias independientes e idénticamente distribuidas, con valores en $S = [0, \infty)$, cuya esperanza cumple $\mathbb{E}[\ln \xi_t] = \mu < \infty$ y supóngase que ξ_0 tiene densidad continua Δ. Esta fuente de incertidumbre es crucial para modelar la variabilidad inherente a los ecosistemas naturales.

En cuanto a los objetivos de cada jugador, se definen funciones de recompensa que representan los beneficios obtenidos por cada país en cada periodo. Para el jugador i, la función de recompensa $r_i : \mathbb{K} \to \mathbb{R} \cup \{-\infty, +\infty\}$ está dada por:

$$r_i(x, a_1, a_2) = \ln a_i, \quad \text{con } x \in X,\ a_i \in A_i(x),$$

esta expresión refleja que la utilidad recibida por cada jugador en un periodo está asociada logarítmicamente a la cantidad de peces que ha capturado en dicho periodo. La elección de una función logarítmica como recompensa implica una preferencia por rendimientos decrecientes: es decir, mientras más se captura, el beneficio marginal disminuye, lo que promueve comportamientos de pesca más conservadores y sostenibles en el largo plazo.

Además, se considera un horizonte aleatorio τ_i definido en $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$ para el jugador i que representa el riesgo del agotamiento del recurso.

De este modo, el criterio de optimalidad a considerar corresponde a la *recompensa total esperada descontada en un horizonte aleatorio*. Este horizonte puede ser interpretado como la duración incierta del juego, la cual puede finalizar abruptamente debido a factores exógenos como eventos ambientales, colapso del recurso o decisiones políticas como se mencionó anteriormente.

$$\mathcal{J}_{i,\tau_i}(x, \pi_1, \pi_2) = \mathbb{E}_i \left[\sum_{t=0}^{\tau_i} \beta_i^t r_i(x_t, a_{1,t}, a_{2,t}) \right].$$

Solución del juego

Lema 3.1.1. El problema de pesquerías planteado en esta sección cumple la Suposición 2.1.4 (a), (b) y (c).

Demostración. (a) Nótese que para cada $x \in X$ el intervalo $\left[0, \frac{x^\delta}{M}\right]$ es cerrado y acotado en los reales, entonces por el Teorema de Heine-Borel [1], $A_1(x)$ y $A_2(x)$ son compactos.

(b) Se demostrará que la función $r_i(x, a_1, a_2) = \ln a_i$, con $i = 1, 2$ es superiormente semicontinua con ayuda de la Definición A.0.1 del Apéndice A. Ahora, para cada i fijo considere la sucesión $(a_{i,n})$ tal que $a_{i,n} \to a_i$. Entonces

$$\limsup_{n\to\infty} \ln a_{i,n} = \ln(\limsup_{n\to\infty} a_{i,n}).$$

Dado que la función logaritmo es monótona creciente, se cumple que:

$$\ln(\limsup_{n\to\infty} a_{i,n}) \leq \ln a_i,$$

siempre que $a_{i,n} \to a_i$. Esto prueba que $\ln a_i$ es semicontinua superiormente.

(c) Para demostrar que la ley de transición es débilmente continua se utilizará la Definición A.0.2 del Apéndice A. Es decir, se demostrará que

$$v'(x, a_1, a_2) = \int v(y) Q(dy|x, a_1, a_2)$$

es continua y acotada en $\mathbb{K}$, para toda función $v : X \to \mathbb{R}$ continua y acotada.

Sea $v(y)$ una función continua y acotada. Utilizando el teorema de cambio de variable con $(x, a_1, a_2) \in \mathbb{K}$, $s \in S$ y considerando

$$w = (x - a_1 - a_2)^\delta s$$

se tiene

$$v'(x, a_1, a_2) = \int v(w) \Delta \left(\frac{w}{(x - a_1 - a_2)^\delta} \right) \frac{1}{(x - a_1 - a_2)^\delta} dw.$$

Para demostrar que v' es continua en (a_1, a_2), se hará uso del Teorema de la Convergencia Dominada, notando que

- La función dentro de la integral es continua para $x - a_1 - a_2 > 0$ por la continuidad de v y de Δ.
- Por otro lado como v es acotada y Δ es una densidad de probabilidad (en particular, acotada casi seguramente), la función dentro de la integral está acotada por una función integrable.

Por el Teorema de la Convergencia Dominada, v' es continua en (a_1, a_2) y de esta manera Q es débilmente continua.

■

Lema 3.1.2. Las funciones $U_{i,0}, U_{i,1}, ..., U_{i,T+1}$ para el jugador i y el equilibrio de Nash del juego con horizonte aleatorio están dados por

$$U_{1,t}(x) = K_t \ln x + C_t$$

$$U_{2,t}(x) = \widehat{K}_t \ln x + \widehat{C}_t$$

y

$$\pi^* = (f^*_{1,t}(x), f^*_{2,t}(x))$$

para cada $x \in X$, donde

$$f^*_{1,t}(x) = \frac{\widehat{K}_t - 1}{\widehat{K}_t K_t} x,$$
$$f^*_{2,t}(x) = \frac{K_t - 1}{\widehat{K}_t K_t} x,$$

$$K_t = 1 + \alpha_{1,t}\beta_1 \delta K_{t+1},$$
$$\widehat{K}_t = 1 + \alpha_{2,t}\beta_2 \delta \widehat{K}_{t+1},$$
$$C_t = \ln\left(\frac{\widehat{K}_t - 1}{\widehat{K}_t K_t - 1}\right) + (K_t - 1)\ln\left(\frac{(\widehat{K}_t - 1)(K_t - 1)}{\widehat{K}_t K_t - 1}\right) + \alpha_{1,t}\beta_1[K_{t+1}\mu + C_{t+1}],$$
$$\widehat{C}_t = \ln\left(\frac{K_t - 1}{\widehat{K}_t K_t - 1}\right) + (\widehat{K}_t - 1)\ln\left(\frac{(\widehat{K}_t - 1)(K_t - 1)}{\widehat{K}_t K_t - 1}\right) + \alpha_{2,t}\beta_2[\widehat{K}_{t+1}\mu + \widehat{C}_{t+1}].$$

Demostración. Con base en el Lema 3.1.1, se concluye que el juego de la gran pesca satisface la Suposición 2.1.4. En consecuencia, el Teorema 2.1.5 garantiza la existencia de un equilibrio de Nash y permite emplear las funciones establecidas en (2.2).

Para simplificar los cálculos, se utilizarán las funciones definidas en la ecuación (2.7), de esta manera se tiene que $U_{i,T+1}(x) = 0$ y

$$\begin{aligned}
U_{i,t}(x) &= \max_{a_i \in A_i(x)} \left[r_i(x, a_1, a_2) + \alpha_{i,t}\, \beta_i \mathbb{E}\left[U_{i,t+1}(F(x, a_1, a_2, \xi_t))\right]\right] \\
&= \max_{a_i \in \left[0, \frac{x^\delta}{M}\right]} \left[\ln a_i + \alpha_{i,t}\beta_i \mathbb{E}\left[U_{i,t+1}\left((x - a_1 - a_2)^\delta \xi\right)\right]\right].
\end{aligned}$$

La demostración procederá por inducción hacia atrás.

Primero, se analiza lo que ocurre cuando $t = T$.

$$\begin{aligned}
U_{1,T}(x) &= \max_{a_1 \in \left[0, \frac{x^\delta}{M}\right]} \left\{\ln a_1 + \alpha_{1,T}\beta_1 \mathbb{E}\left[U_{1,T+1}\left((x - a_1 - a_2)^\delta \xi\right)\right]\right\} \\
&= \max_{a_1 \in \left[0, \frac{x^\delta}{M}\right]} \left\{\ln a_1 + \alpha_{1,T}\beta_1 \mathbb{E}[0]\right\} \\
&= \max_{a_1 \in \left[0, \frac{x^\delta}{M}\right]} \left\{\ln a_1\right\} = \ln \frac{x^\delta}{M} = \delta \ln x - \ln M
\end{aligned}$$

donde $f^*_{1,T}(x) = \frac{x^\delta}{M}$, para $x \in X$.

$$\begin{aligned}
U_{2,T}(x) &= \max_{a_2\in\left[0,\frac{x^\delta}{M}\right]} \left\{\ln a_2 + \alpha_{2,T}\beta_2\mathbb{E}\left[U_{2,T+1}\left(\left(x-a_1-a_2\right)^\delta \xi\right)\right]\right\}\\
&= \max_{a_2\in\left[0,\frac{x^\delta}{M}\right]} \left\{\ln a_2 + \alpha_{2,T}\beta_2\mathbb{E}[0]\right\}\\
&= \max_{a_2\in\left[0,\frac{x^\delta}{M}\right]} \left\{\ln a_2\right\} = \ln\frac{x^\delta}{M} = \delta\ln x - \ln M
\end{aligned}$$

donde $f^*_{2,T}(x) = \frac{x^\delta}{M}$, para $x \in X$.
A continuación, se supone $U_{1,t}(x) = K_t \ln x + C_t$ y $U_{2,t}(x) = \widehat{K}_t \ln x + \widehat{C}_t$ considerando que $K_T = \widehat{K}_T = \delta$ y $C_T = \widehat{C}_T = -\ln M$.

$$\begin{aligned}
U_{1,t-1}(x) &= \max_{a_1\in\left[0,\frac{x^\delta}{M}\right]} \left\{\ln a_1 + \alpha_{1,t-1}\beta_1\mathbb{E}\left[U_{1,t}\left((x-a_1-a_2)^\delta\xi\right)\right]\right\}\\
&= \max_{a_1\in\left[0,\frac{x^\delta}{M}\right]} \left\{\ln a_1 + \alpha_{1,t-1}\beta_1\mathbb{E}\left[K_t\ln\left((x-a_1-a_2)^\delta\xi\right)+C_t\right]\right\}\\
&= \max_{a_1\in\left[0,\frac{x^\delta}{M}\right]} \left\{\ln a_1 + \alpha_{1,t-1}\beta_1\mathbb{E}\left[\delta K_t\ln\left(x-a_1-a_2\right)+K_t\ln\xi+C_t\right]\right\}\\
&= \max_{a_1\in\left[0,\frac{x^\delta}{M}\right]} \left\{\ln a_1 + \alpha_{1,t-1}\beta_1\left[\delta K_t\ln\left(x-a_1-a_2\right)\right)+K_t\mu+C_t]\right\}\\
&= \max_{a_1\in\left[0,\frac{x^\delta}{M}\right]} \left\{\ln a_1 + \alpha_{1,t-1}\beta_1\delta K_t\ln\left(x-a_1-a_2\right)+\alpha_{1,t-1}\beta_1\left[K_t\mu+C_t\right]\right\}.
\end{aligned}$$

Derivando con respecto a a_1 e igualando a cero se consigue

$$a_1 = \frac{x-a_2}{1+\alpha_{1,t-1}\beta_1\delta K_t}.$$

Ahora, se obtiene la función $U_{2,t-1}$ del jugador 2 en el tiempo $t-1$.

$$\begin{aligned}
U_{2,t-1}(x) &= \max_{a_2\in\left[0,\frac{x^\delta}{M}\right]} \left\{\ln a_2 + \alpha_{2,t-1}\beta_2\mathbb{E}\left[U_{2,t}\left((x-a_1-a_2)^\delta\xi\right)\right]\right\}\\
&= \max_{a_2\in\left[0,\frac{x^\delta}{M}\right]} \left\{\ln a_2 + \alpha_{2,t-1}\beta_2\mathbb{E}\left[\widehat{K}_t\ln\left((x-a_1-a_2)^\delta\xi\right)+\widehat{C}_t\right]\right\}\\
&= \max_{a_2\in\left[0,\frac{x^\delta}{M}\right]} \left\{\ln a_2 + \alpha_{2,t-1}\beta_2\mathbb{E}\left[\delta\widehat{K}_t\ln\left(x-a_1-a_2\right)+\widehat{K}_t\ln\xi+\widehat{C}_t\right]\right\}\\
&= \max_{a_2\in\left[0,\frac{x^\delta}{M}\right]} \left\{\ln a_2 + \alpha_{2,t-1}\beta_2\left[\delta\widehat{K}_t\ln\left(x-a_1-a_2\right)\right)+\widehat{K}_t\mu+\widehat{C}_t]\right\}\\
&= \max_{a_2\in\left[0,\frac{x^\delta}{M}\right]} \left\{\ln a_2 + \alpha_{2,t-1}\beta_2\delta\widehat{K}_t\ln\left(x-a_1-a_2\right)+\alpha_{2,t-1}\beta_2\left[\widehat{K}_t\mu+\widehat{C}_t\right]\right\}.
\end{aligned}$$

Derivando con respecto a a_2 e igualando a cero

$$a_2 = \frac{x-a_1}{1+\alpha_{2,t-1}\beta_2\delta\widehat{K}_t}.$$

Resolviendo simultáneamente $a_1 = \frac{x-a_2}{1+\alpha_{1,t-1}\beta_1\delta K_t}$ y $a_2 = \frac{x-a_1}{1+\alpha_{2,t-1}\beta_2\delta \widehat{K}_t}$ se consigue

$$a_1 = \frac{\alpha_{2,t-1}\beta_2\delta\widehat{K}_t x}{(1+\alpha_{1,t-1}\beta_1\delta K_t)\left(1+\alpha_{2,t-1}\beta_2\delta\widehat{K}_t\right)-1}$$

y

$$a_2 = \frac{\alpha_{1,t-1}\beta_1\delta K_t x}{(1+\alpha_{1,t-1}\beta_1\delta K_t)\left(1+\alpha_{2,t-1}\beta_2\delta\widehat{K}_t\right)-1}.$$

Por lo tanto

$$f^*_{1,t-1}(x) = \frac{\alpha_{2,t-1}\beta_2\delta\widehat{K}_t x}{(1+\alpha_{1,t-1}\beta_1\delta K_t)\left(1+\alpha_{2,t-1}\beta_2\delta\widehat{K}_t\right)-1}$$

y

$$f^*_{2,t-1}(x) = \frac{\alpha_{1,t-1}\beta_1\delta K_t x}{(1+\alpha_{1,t-1}\beta_1\delta K_t)\left(1+\alpha_{2,t-1}\beta_2\delta\widehat{K}_t\right)-1}.$$

De esta manera, sustituyendo $f^*_{1,t-1}(x)$ y $f^*_{2,t-1}(x)$ se obtiene

$$\begin{aligned}U_{1,t-1}(x) &= (1+\alpha_{1,t-1}\beta_1\delta K_t)\ln x + \ln\left(\frac{\alpha_{2,t-1}\beta_2\delta\widehat{K}_t}{(1+\alpha_{1,t-1}\beta_1\delta K_t)\left(1+\alpha_{2,t-1}\beta_2\delta\widehat{K}_t\right)-1}\right)\\ &+ \alpha_{1,t-1}\beta_1\delta K_t \ln\left(\frac{\alpha_{1,t-1}\alpha_{2,t-1}\beta_1\beta_2\delta^2 K_t\widehat{K}_t}{(1+\alpha_{1,t-1}\beta_1\delta K_t)\left(1+\alpha_{2,t-1}\beta_2\delta\widehat{K}_t\right)-1}\right) + \alpha_{1,t-1}\beta_1\left[K_t\mu + C_t\right].\end{aligned}$$

Similarmente

$$\begin{aligned}U_{2,t-1}(x) &= (1+\alpha_{2,t-1}\beta_2\delta \widehat{K}_t)\ln x + \ln\left(\frac{\alpha_{1,t-1}\beta_1\delta K_t}{(1+\alpha_{1,t-1}\beta_1\delta K_t)\left(1+\alpha_{2,t-1}\beta_2\delta\widehat{K}_t\right)-1}\right)\\ &+ \alpha_{2,t-1}\beta_2\delta \widehat{K}_t \ln\left(\frac{\alpha_{1,t-1}\alpha_{2,t-1}\beta_1\beta_2\delta^2 K_t\widehat{K}_t}{(1+\alpha_{1,t-1}\beta_1\delta K_t)\left(1+\alpha_{2,t-1}\beta_2\delta\widehat{K}_t\right)-1}\right) + \alpha_{2,t-1}\beta_2\left[\widehat{K}_t\mu + \widehat{C}_t\right].\end{aligned}$$

De esta manera

$$K_{t-1} = 1+\alpha_{1,t-1}\beta_1\delta K_t,$$

$$\widehat{K}_{t-1} = 1+\alpha_{2,t-1}\beta_2\delta \widehat{K}_t,$$

$$C_{t-1} = \ln\left(\frac{\widehat{K}_{t-1}-1}{\widehat{K}_{t-1}K_{t-1}-1}\right) + (K_{t-1}-1)\ln\left(\frac{(\widehat{K}_{t-1}-1)(K_{t-1}-1)}{\widehat{K}_{t-1}K_{t-1}-1}\right) + \alpha_{1,t-1}\beta_1[K_t\mu + C_t],$$

$$\widehat{C}_{t-1} = \ln\left(\frac{K_{t-1}-1}{\widehat{K}_{t-1}K_{t-1}-1}\right) + (\widehat{K}_{t-1}-1)\ln\left(\frac{(\widehat{K}_{t-1}-1)(K_{t-1}-1)}{\widehat{K}_{t-1}K_{t-1}-1}\right) + \alpha_{2,t-1}\beta_2[\widehat{K}_t\mu + \widehat{C}_t],$$

$$f^*_{1,t-1}(x) = \frac{\widehat{K}_{t-1}-1}{\widehat{K}_{t-1}K_{t-1}}x,$$

$$f^*_{2,t-1}(x) = \frac{K_{t-1}-1}{\widehat{K}_{t-1}K_{t-1}}x.$$

Por lo tanto, el enunciado ha sido demostrado. ■

3.2. Juego lineal-cuadrático

Los problemas lineales-cuadráticos (LQ), también conocidos como linear cuadratic regulators (LQR), se caracterizan por involucrar sistemas dinámicos lineales sujetos a una función de costo cuadrática de una etapa. Estos problemas se encuentran entre los más ampliamente utilizados en disciplinas como la ingeniería, la economía y diversas áreas del conocimiento [14]. En el ámbito de la ingeniería, por ejemplo, el costo total puede interpretarse como una abstracción de la energía [26], o bien aplicarse al control de vehículos aéreos no tripulados [36, 40].

El problema LQ se extiende de manera natural al campo de la teoría de juegos cuando múltiples agentes o jugadores interactúan dentro de un sistema dinámico, cada uno con objetivos particulares. En este contexto, se abordan los denominados juegos lineales-cuadráticos (LQ), que combinan la dinámica lineal del sistema con funciones de costo cuadráticas para cada jugador. Ejemplos de investigaciones sobre el LQ en juegos se encuentran en estudios que consideran un horizonte finito o uno infinito [35, 39], y sus aplicaciones abarcan distintos ámbitos, como en [18], donde el LQ en juegos permite determinar la capacidad de carga de un manipulador en un sistema de bucle cerrado.

De manera análoga a lo observado en el juego de la gran pesca, en el contexto de juegos lineales-cuadráticos, la introducción de un horizonte aleatorio resulta de particular importancia debido a la incertidumbre asociada a la duración del juego. Esta incertidumbre puede reflejar situaciones en las que la interacción estratégica entre los jugadores se ve interrumpida de forma abrupta por factores externos. Por ejemplo, en un juego LQ que modela la explotación conjunta de un recurso renovable, dos empresas pueden enfrentar tensiones políticas que provoquen la finalización imprevista del juego.

Planteamiento del juego

Se considera un juego lineal-cuadrático (LQ) con dos jugadores descrito mediante el modelo de juego estocástico que se muestra a continuación.

El espacio de estados y el conjunto de acciones para el jugador i con $i = 1, 2$, son

$$X = A_1 = A_2 = \mathbb{R}.$$

Cada jugador enfrenta una función de costo cuadrática, que depende tanto del estado $x \in X$ como de sus respectivas acciones $a_1 \in A_1(x)$ y $a_2 \in A_2(x)$:

$$c_i(x, a_1, a_2) = x^2 + a_i^2.$$

La dinámica del sistema está descrita por la siguiente ecuación en diferencias:

$$x_{t+1} = F\left(x_t, a_{1,t}, a_{2,t}, \xi_t\right) = x_t + a_{1,t} + a_{2,t} + \xi_t.$$

Además, se considera un horizonte aleatorio τ_i definido en $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$ para el jugador i que representa el riesgo del agotamiento del recurso.

Condición 3.2.1. (a) Las variables aleatorias ξ_t, para cada $t = 0, 1, \ldots$ son independientes e idénticamente distribuidas con valores en $S = \mathbb{R}$. Supóngase que ξ_0 tiene densidad continua Δ, con media 0 y varianza σ^2, es decir,

$$E(\xi_t) = \int s\Delta(s)ds = 0,$$

y

$$E(\xi_t^2) = \int s^2\Delta(s)ds = \sigma^2 < \infty.$$

(b) Para cada $x \in X$, las acciones admisibles para el jugador i son

$$A_i(x) = [-x, x].$$

De este modo, el criterio de optimalidad a considerar corresponde al *costo total esperado con horizonte aleatorio*:

$$\mathcal{V}_{i,\tau_i}\left(x, \pi_1, \pi_2\right) = \mathbb{E}_i\left[\sum_{t=0}^{\tau_i} c_i\left(x_t, a_{1,t}, a_{2,t}\right)\right].$$

Solución del juego

Lema 3.2.2. El problema de LQ planteado en esta sección cumple la Suposición 2.2.2 (a), (b) y (c).

Demostración. (a) Nótese que para cada $x \in X$ el intervalo $[-x, x]$ es cerrado y acotado en los reales, entonces por el Teorema de Heine-Borel, $A_1(x)$ y $A_2(x)$ son compactos.

(b′) Para el jugador i, su función de costo $c_i(x, a_1, a_2) = x^2 + a_i^2$ es continua en $(a_1, a_2) \in A_1(x) \times A_2(x)$, por lo tanto es semicontinua inferiormente $(a_1, a_2) \in A_1(x) \times A_2(x)$.

(c) Para demostrar que la ley de transición es débilmente continua se utilizará la Definición A.0.2 del Apéndice A. Es decir, se demostrará que

$$v'(x, a_1, a_2) = \int v(y) Q(dy|x, a_1, a_2)$$

es continua y acotada en $\mathbb{K}$, para toda función $v : X \to \mathbb{R}$ continua y acotada.

Sea $v(y)$ una función continua y acotada. Utilizando el teorema de cambio de variable con $(x, a_1, a_2) \in \mathbb{K}$, $s \in S$ y considerando

$$w = x + a_1 + a_2 + s$$

se tiene

$$v'(x, a_1, a_2) = \int v(w) \Delta \left(w - x - a_1 - a_2\right) dw.$$

Para demostrar que v' es continua en (a_1, a_2), se hará uso del Teorema de la Convergencia Dominada, notando que

- La función dentro de la integral es continua por la continuidad de v y de Δ.
- Por otro lado como v es acotada y Δ es una densidad de probabilidad (en particular, acotada casi seguramente), la función dentro de la integral está acotada por una función integrable.

Por el Teorema de la Convergencia Dominada, v' es continua en (a_1, a_2) y de esa manera Q es débilmente continua.

■

Lema 3.2.3. Las funciones $W_{i,0}, W_{i,1}, ..., W_{i,T+1}$ para el jugador i y el equilibrio de Nash del juego LQ están dados por

$$W_{1,t}(x) = D_t x^2 + G_t$$
$$W_{2,t}(x) = \widehat{D}_t x^2 + \widehat{G}_t$$

y

$$\pi^* = (f^*_{1,t}(x), f^*_{1,t}(x))$$

para cada $x \in X$, donde

$$f^*_{1,t}(x) = -\frac{\gamma_{1,t} D_{t+1} x}{1 + \gamma_{1,t} D_{t+1} + \gamma_{2,t} \widehat{D}_{t+1}},$$

$$f^*_{2,t}(x) = -\frac{\gamma_{2,t}\widehat{D}_{t+1}x}{1+\gamma_{1,t}D_{t+1}+\gamma_{2,t}\widehat{D}_{t+1}},$$

$$D_t = 1 + \frac{(1+\gamma_{1,t}D_{t+1})(\gamma_{1,t}D_{t+1})}{(1+\gamma_{1,t}D_{t+1}+\gamma_{2,t}\widehat{D}_{t+1})^2},$$

$$\widehat{D}_t = 1 + \frac{(1+\gamma_{2,t}\widehat{D}_{t+1})(\gamma_{2,t}\widehat{D}_{t+1})}{(1+\gamma_{1,t}D_{t+1}+\gamma_{2,t}\widehat{D}_{t+1})^2},$$

$$G_t = \gamma_{1,t}(D_{t+1}\sigma^2 + G_{t+1}),$$

$$\widehat{G}_t = \gamma_{2,t}(\widehat{D}_{t+1}\sigma^2 + \widehat{G}_{t+1}).$$

Demostración. Con base en el Lema 3.2.2, se concluye que el juego lineal-cuadrático satisface la Suposición 2.2.2. En consecuencia, el Teorema 2.2.3 garantiza la existencia de un equilibrio de Nash y permite emplear las funciones establecidas en (2.9).

Para simplificar los cálculos, se utilizarán las funciones definidas en la ecuación (2.11), de esta manera se tiene que $W_{i,T+1}(x) = 0$ y

$$\begin{aligned} W_{i,t}(x) &= \min_{a_i \in A_i(x)} \left[c_i(x,a_1,a_2) + \gamma_{i,t}\,\mathbb{E}\left[W_{i,t+1}(F(x,a_1,a_2,\xi_t))\right]\right]. \\ &= \min_{a_1 \in [-x,x]} \left[x^2 + a_1^2 + \gamma_{1,t}\mathbb{E}\left[W_{1,t+1}\left(x+a_1+a_2+\xi\right)\right]\right] \end{aligned}$$

La demostración procederá por inducción hacia atrás.

Primero, se analiza lo que ocurre cuando $t = T$.

$$\begin{aligned} W_{1,T}(x) &= \min_{a_1 \in [-x,x]} \left[x^2 + a_1^2 + \gamma_{1,T}\mathbb{E}\left[W_{1,T+1}\left(x+a_1+a_2+\xi\right)\right]\right] \\ &= \min_{a_1 \in [-x,x]} \left[x^2 + a_1^2 + \gamma_{1,T}\mathbb{E}[0]\right] \\ &= \min_{a_1 \in [-x,x]} \left[x^2 + a_1^2\right] \\ &= x^2 \end{aligned}$$

donde $f^*_{1,T}(x) = 0$, para $x \in X$.

$$\begin{aligned} W_{2,T}(x) &= \min_{a_2 \in [-x,x]} \left[x^2 + a_2^2 + \gamma_{2,T}\mathbb{E}\left[W_{2,T+1}\left(x+a_1+a_2+\xi\right)\right]\right] \\ &= \min_{a_2 \in [-x,x]} \left[x^2 + a_2^2 + \gamma_{2,T}\mathbb{E}[0]\right] \\ &= \min_{a_2 \in [-x,x]} \left[x^2 + a_2^2\right] \\ &= x^2 \end{aligned}$$

donde $f_{2,T}^*(x) = 0$, para $x \in X$.
A continuación, se supone $W_{1,t}(x) = D_t x^2 + G_t$ y $W_{2,t}(x) = \widehat{D}_t x^2 + \widehat{G}_t$. considerando que $D_T = \widehat{D}_T = 1$ y $G_T = \widehat{G}_T = 0$.

$$\begin{aligned} W_{1,t}(x) &= \min_{a_1 \in [-x,x]} \left[x^2 + a_1^2 + \gamma_{1,t} \mathbb{E}\left[W_{1,t+1}\left(x + a_1 + a_2 + \xi\right)\right]\right] \\ &= \min_{a_1 \in [-x,x]} \left[x^2 + a_1^2 + \gamma_{1,t} D_{t+1} \left[(x+a_2)^2 + 2a_1(x+a_2) + a_1^2 + \sigma^2 \right] + \gamma_{1,t} G_{t+1} \right]. \end{aligned}$$

Derivando con respecto a a_1 e igualando a cero se consigue

$$a_1 = -\frac{\gamma_{1,t} D_{t+1}\left(x + a_2\right)}{1 + \gamma_{1,t} D_{t+1}}.$$

Similarmente para el jugador 2, se tiene

$$\begin{aligned} W_{2,t}(x) &= \min_{a_2 \in [-x,x]} \left[x^2 + a_2^2 + \gamma_{2,t} \mathbb{E}\left[W_{2,t+1}\left(x + a_1 + a_2 + \xi\right)\right]\right] \\ &= \min_{a_2 \in [-x,x]} \left[x^2 + a_2^2 + \gamma_{2,t} \widehat{D}_{t+1} \left[(x+a_1)^2 + 2a_2(x+a_1) + a_2^2 + \sigma^2 \right] + \gamma_{2,t} \widehat{G}_{t+1} \right] \end{aligned}$$

y

$$a_2 = -\frac{\gamma_{2,t} \widehat{D}_{t+1}\left(x + a_1\right)}{1 + \gamma_{2,t} \widehat{D}_{t+1}}.$$

Se resuelve simultáneamente para obtener

$$a_1 = -\frac{\gamma_{1,t} D_{t+1} x}{1 + \gamma_{1,t} D_{t+1} + \gamma_{2,t} \widehat{D}_{t+1}}$$

y

$$a_2 = -\frac{\gamma_{2,t} \widehat{D}_{t+1} x}{1 + \gamma_{1,t} D_{t+1} + \gamma_{2,t} \widehat{D}_{t+1}}.$$

Por lo tanto

$$f_{1,t}^*(x) = -\frac{\gamma_{1,t} D_{t+1} x}{1 + \gamma_{1,t} D_{t+1} + \gamma_{2,t} \widehat{D}_{t+1}}$$

y

$$f_{2,t}^*(x) = -\frac{\gamma_{2,t} \widehat{D}_{t+1} x}{1 + \gamma_{1,t} D_{t+1} + \gamma_{2,t} \widehat{D}_{t+1}}.$$

De esta manera, sustituyendo $f_{1,t-1}^*(x)$ y $f_{2,t-1}^*(x)$ se obtiene

$$W_{1,t}(x) = \left(1 + \frac{(1+\gamma_{1,t}D_{t+1})(\gamma_{1,t}D_{t+1})}{(1+\gamma_{1,t}D_{t+1}+\gamma_{2,t}\widehat{D}_{t+1})^2}\right)x^2 + \gamma_{1,t}(D_{t+1}\sigma^2 + G_{t+1});$$

$$W_{2,t}(x) = \left(1 + \frac{(1+\gamma_{2,t}\widehat{D}_{t+1})(\gamma_{2,t}\widehat{D}_{t+1})}{(1+\gamma_{1,t}D_{t+1}+\gamma_{2,t}\widehat{D}_{t+1})^2}\right)x^2 + \gamma_{2,t}(\widehat{D}_{t+1}\sigma^2 + \widehat{G}_{t+1}).$$

Entonces

$$D_t = 1 + \frac{(1+\gamma_{1,t}D_{t+1})(\gamma_{1,t}D_{t+1})}{(1+\gamma_{1,t}D_{t+1}+\gamma_{2,t}\widehat{D}_{t+1})^2},$$

$$\widehat{D}_t = 1 + \frac{(1+\gamma_{2,t}\widehat{D}_{t+1})(\gamma_{2,t}\widehat{D}_{t+1})}{(1+\gamma_{1,t}D_{t+1}+\gamma_{2,t}\widehat{D}_{t+1})^2},$$

$$G_t = \gamma_{1,t}(D_{t+1}\sigma^2 + G_{t+1}),$$

$$\widehat{G}_t = \gamma_{2,t}(\widehat{D}_{t+1}\sigma^2 + \widehat{G}_{t+1}).$$

Por lo tanto, el enunciado ha sido demostrado.

■

En este último capítulo se aplicó la teoría desarrollada en los capítulos anteriores para analizar y proponer soluciones a dos juegos de especial relevancia: un juego en el contexto de pesquerías y un juego de tipo lineal-cuadrático (LQ). Ambos representan problemas actuales con múltiples aplicaciones en economía, gestión de recursos naturales y teoría del control.

La metodología se basó principalmente en el uso de programación dinámica, herramienta clave para establecer y caracterizar los equilibrios de Nash de cada juego. Se demostró que dichos equilibrios adoptan formas estructuradas y particulares, lo que facilita tanto su interpretación como su posible implementación en contextos reales.

Conclusiones

La presente tesis tuvo como objetivo principal demostrar la existencia de un equilibrio de Nash en juegos markovianos de suma no cero con horizonte aleatorio, así como ilustrar su aplicabilidad en contextos relevantes. Este propósito se abordó desde un enfoque teórico riguroso, incorporando herramientas fundamentales de la teoría de juegos y extendiendo el análisis a un tipo de modelo poco explorado en la literatura actual.

El trabajo se sustentó en el desarrollo formal del modelo Markoviano, considerando jugadores racionales, donde las decisiones se toman bajo incertidumbre respecto a la duración del proceso. La inclusión de un horizonte aleatorio permitió representar de forma más realista situaciones donde el tiempo de interacción entre los jugadores no está previamente determinado, como ocurre en numerosos contextos económicos y ecológicos.

Una contribución destacada fue la formulación de un marco general para juegos Markovianos de suma no cero con horizonte aleatorio, diferenciando entre enfoques basados en recompensas y en costos. A partir de ciertas hipótesis estructurales, se presentó un procedimiento basado en programación dinámica para determinar los equilibrios de Nash. Este método no solo facilitó la demostración de existencia de equilibrios, sino que también permitió describir su estructura de forma clara.

Desde el punto de vista aplicado, se mostraron dos ejemplos representativos: un juego de la gran pesca y un juego de tipo lineal-cuadrático. Ambos confirmaron la pertinencia del enfoque propuesto, evidenciando que los equilibrios obtenidos son consistentes.

Como trabajo a futuro, se plantea abordar la problemática en escenarios donde se cuenta con un soporte infinito para la variable aleatoria que representa el horizonte del problema, además de explorar el caso donde los horizontes son dependientes entre sí, lo cual introduce desafíos adicionales tanto en la formulación matemática del modelo como en la caracterización de los equilibrios asociados. Esta extensión permitiría ampliar el alcance y aplicabilidad de los resultados obtenidos, especialmente en contextos donde la duración del juego es incierta.

Apéndice A

Conceptos

Definición A.0.1. Una función $f : X \to \mathbb{R}$ es **semicontinua superiormente** en un punto x_0 si

$$\limsup_{x \to x_0} f(x) \leq f(x_0).$$

Es decir, para toda sucesión $x_n \to x_0$, se cumple que

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

Definición A.0.2. La ley de transición Q es débilmente continua si

$$v'(x, a_1, a_2) := \int v(y) Q(dy|x, a_1, a_2)$$

es continua y acotada en $\mathbb{K}$ para cada función continua y acotada v en X.

Bibliografía

[1] T. M. Apostol. Análisis matemático, 2a. *Edición, Editorial Reverté, España*, 518, 1976.

[2] E. Barron. *Game Theory An Introduction.* Wiley, second edition, 2013.

[3] R. Bellman. *Dynamic Programming.* Princeton University Press, 1957.

[4] A. Bressan, A. Marigonda, K. T. Nguyen, and M. Palladino. A stochastic model of optimal debt management and bankruptcy. *SIAM Journal on Financial Mathematics*, 8(1):841–873, 2017.

[5] C. Camilo-Garay, R. I. Ortega-Gutiérrez, and H. Cruz-Suárez. Optimal strategies for a fishery model applied to utility functions. *Mathematical Biosciences and Engineering*, 18(1):518–529, 2021.

[6] Y. Chen, S. Zou, and J. Lygeros. Game theoretic stochastic energy coordination under a distributed zeroth-order algorithm. *IFAC-PapersOnLine*, 53(2):4070–4075, 2020.

[7] A. Cournot. *Recherches sur les principes mathématiques de la théorie des richesses.* Librairie de L. Hachette, 1838.

[8] J. M. Crespo-Guerrero, J. M. Casado Izquierdo, et al. Pesca y comercialización del pulpo en yucatán:¿ un proceso extractivista impulsado por la unión europea? *Geográfica Venezolana*, 2024.

[9] H. Cruz-Suárez, R. Ilhuicatzi-Roldán, and R. Montes-de Oca. Markov decision processes on borel spaces with total cost and random horizon. *Journal of Optimization Theory and Applications*, 162:329–346, 2014.

[10] C. Eksin, J. S. Shamma, and J. S. Weitz. Disease dynamics in a stochastic network game: a little empathy goes a long way in averting outbreaks. *Scientific reports*, 7(1):44122, 2017.

[11] D. González-Sánchez, F. Luque-Vásquez, and J. A. Minjárez-Sosa. Zero-sum markov games with random state-actions-dependent discount factors: existence of optimal strategies. *Dynamic Games and Applications*, 9:103–121, 2019.

[12] G. J. Guðmundsson. The cod and the cold war. *Scandinavian Journal of History*, 31(2):97–118, 2006.

[13] O. Hernández-Lerma. *Adaptive Markov control processes*, volume 79. Springer Science & Business Media, 2012.

[14] O. Hernández-Lerma and J. B. Lasserre. *Discrete-time Markov control processes: basic optimality criteria*, volume 30. Springer Science & Business Media, 2012.

[15] O. Hernández-Lerma and J. B. Lasserre. *Further topics on discrete-time Markov control processes*, volume 42. Springer Science & Business Media, 2012.

[16] R. Ilhuicatzi-Roldan. *Procesos de decision de Markov con horizonte aleatorio*. PhD thesis, Benemérita Universidad Autónoma de Puebla, 2013.

[17] A. Jaśkiewicz and A. S. Nowak. Non-zero-sum stochastic games. *Handbook of dynamic game theory*, pages 1–64, 2018.

[18] M.H. Korayem, H. Esfandiar, and R. Dargahi. Determining load carrying capacity of a manipulator by game theory: Closed-loop nonzero-sum differential game approach. *Journal of Mechanical Science and Technology*, 30:5197–5205, 2016.

[19] J. Kotlarz. Application of nash equilibrium for developing an optimal forest harvesting strategy in torun forest district. *arXiv preprint arXiv:2403.03555*, 2024.

[20] D. Levhari and L. J. Mirman. Savings and consumption with an uncertain horizon. *Journal of Political Economy*, 85(2):265–281, 1977.

[21] D. Levhari and L. J. Mirman. The great fish war: an example using a dynamic cournot-nash solution. *The Bell Journal of Economics*, pages 322–334, 1980.

[22] J. Mayberry, J. Nash and M. Shubik. *A Comparison of Treatments of a Duopoly Situation*. Econometrica, Vol. 21, No. 1, 1953.

[23] K. A. Miller. North american pacific salmon: A case of fragile cooperation. *FAO Fisheries Report*, pages 105–122, 2003.

[24] J. A. Minjárez-Sosa. *Zero-Sum discrete-time Markov games with unknown disturbance distribution: discounted and average criteria*. Springer Nature, 2020.

[25] P. C. Missios and C. Plourde. The canada-european union turbot war: A brief game theoretic analysis. *Canadian Public Policy/Analyse de Politiques*, pages 144–150, 1996.

[26] K. Mori. Approximate linear quadratic regulator problem and its application to optimal control in discrete-time lti systems. In *Intelligent Control and Innovative Computing*, pages 53–64. Springer, 2011.

[27] J. Nash. *Equilibrium points in n-person games.* National Academy of Science, Vol. 36, No. 1, 1950.

[28] J. Nash. *Non-cooperative games.* Department of Mathematics, Princeton University, 1950.

[29] J. Nash. *Two-person cooperative games.* Santa Mónica: RAND Corporation, 1950.

[30] J. Neumann. Zur theorie der gesellschaftsspiele. *Mathematische annalen*, 100(1):295–320, 1928.

[31] J. Neumann and O. Morgenstern. *Theory of Games and Economic Behavior.* Princeton University Press, Princeton, 1944.

[32] A. Nowak et al. A note on an equilibrium in the great fish war game. *Economics Bulletin*, 17(2):1–10, 2006.

[33] R. I. Ortega-Gutiérrez, R. Montes-de Oca, and H. Cruz-Suárez. Characterization of a cournot–nash equilibrium for a fishery model with fuzzy utilities. *Journal of Mathematics*, 2024(1):6885051, 2024.

[34] R.I. Ortega Gutierrez. La ecuación de euler para la solución de juegos estocásticos descontados. Tesis de Maestria, Benemérita Universida Autonoma de Puebla, 2011.

[35] G.P. Papavassilopoulos and G.J. Olsder. On the linear-quadratic, closed-loop, no-memory nash game. *Journal of optimization theory and applications*, 42(4):551–560, 1984.

[36] H. Purnawan, E. B. Purwanto, et al. Design of linear quadratic regulator (lqr) control system for flight stability of lsu-05. In *Journal of Physics: Conference Series*, volume 890, page 012056. IOP Publishing, 2017.

[37] M. L. Puterman. *Decision Process: Discrete Stochastic Dynamic Programming.* John Wiley & Sons, 1994.

[38] L. S. Shapley. *Stochastic Games.* PNAS 39 (10): 1095-1100, 1953.

[39] Z. J. Williams. *Distributed game-theoretic trajectory planning for multi-agent interactions.* PhD thesis, University of Illinois at Urbana-Champaign, 2023.

[40] J. Willis, J. Johnson, and R. W. Beard. State-dependent lqr control for a tilt-rotor uav. In *2020 American Control Conference (ACC)*, pages 4175–4181. IEEE, 2020.