Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias Físico Matemáticas

Posgrado en Ciencias Matemáticas

Juegos Markovianos con tiempos de paro, bajo el criterio sensible al riesgo.

Tesis que se presenta como requisito final para obtener el título de

Doctor en Ciencias (Matemáticas)

Presenta: M.C. Jaicer Jonás López Rivero.

Directores de tesis: Dr. Hugo Adán Cruz Suárez
Dr. Rolando Cavazos Cadena.

Puebla, Puebla, Diciembre de 2024.

Dedicatoria

Dedicado a mi familia. En especial a mis padres, esposa e hijos.

Agradecimientos

Quiero expresar mi más sincero agradecimiento a todas las personas que han sido parte fundamental de mi formación, tanto académica como personal. A mis padres, Josefina Rivero y Omar López, cuyo apoyo incondicional ha sido clave para alcanzar este logro, les debo mi gratitud eterna. También quiero agradecer profundamente a mi esposa, Luz Alvarado, y a mis tres hijos, Jaicer Fabián, Adrián Jonás y Camila Valentina, quienes con su amor, paciencia y motivación diaria me han dado el impulso necesario para seguir adelante y cumplir mis sueños.

A quienes me apoyaron para emprender este viaje desde Barquisimeto a Puebla, quiero expresar mi agradecimiento sincero. A mis amigos de la UCLA: Rafael, Uvencio, Karla, Diana, Shaday y Harry, quienes siempre estuvieron presentes con su apoyo, consejos y ánimo en cada paso de este proceso. También quiero agradecer a todas las personas que tuve el honor de conocer en este hermoso país, México. En especial, a Anel, así como a mis compañeros de generación y del Laboratorio de Probabilidad y Estadística.

A los profesores de la FCFM, quienes me brindaron una formación de excelencia, especialmente al Dr. Hugo Cruz, mi asesor, por su orientación, apoyo y dedicación a lo largo de este proceso.

A los integrantes de mi jurado de tesis, por sus valiosas observaciones y contribuciones, las cuales fueron fundamentales para enriquecer mi trabajo.

Al CONAHCYT, por su valioso apoyo económico, y al personal de la FCFM, especialmente a Tere, quien siempre estuvo dispuesta a ayudarme con los trámites académicos, facilitando así mi proceso de integración y desarrollo.

Quiero expresar mi agradecimiento póstumo al Dr. Rolando Cavazos Cadena, cuya invaluable contribución ha sido fundamental para el desarrollo de esta investigación.

A todos ellos, GRACIAS.

Índice general

Introducción

La presente tesis se centra en una clase de juegos de suma cero a tiempo discreto, espacio de estados numerable, transiciones Markovianas y recompensas acotadas. En estos juegos, participan dos jugadores, denominados Jugador I y Jugador II, quienes observan el estado actual del sistema y tienen la capacidad de influir en su evolución mediante la aplicación de acciones en cada época de decisión. El proceso de toma de decisiones es secuencial, comenzando con el Jugador II, quien puede elegir entre detener el juego o permitir que el sistema continúe su evolución. Si decide detenerlo, deberá pagar una recompensa terminal al Jugador I. Si opta por continuar, el Jugador I selecciona una acción, lo que genera dos efectos: primero, la cadena de Markov transita al siguiente estado conforme a la ley de transición; segundo, el Jugador II paga una recompensa inmediata al Jugador I. El proceso anterior se repite en cada nuevo estado al que el juego avanza. A este tipo de juegos se le conoce como *Markov stopping games*, y en español se pueden emplear los términos juegos markovianos con tiempos de paro o juegos de detención de Markov.

Por otro lado, se asume que el jugador I tiene un coeficiente de sensibilidad al riesgo constante $\lambda \neq 0$. En consecuencia, el jugador I evalúa dos recompensas aleatorias diferentes utilizando el valor esperado de una función de utilidad exponencial con este coeficiente de sensibilidad λ. El desempeño de un par de estrategias se evaluará mediante el criterio de recompensa total sensible al riesgo. Así, el objetivo del Jugador I es maximizar su recompensa total sensible al riesgo, mientras que el objetivo del Jugador II es minimizar dicha recompensa para el Jugador I. Esta situación implica que el juego sea de suma cero.

El objetivo general de nuestra investigación es determinar bajo qué condiciones sobre el modelo de control se garantiza la existencia de una solución para el juego. Además, se plantean los siguientes objetivos específicos:

- Caracterizar la función de valor del juego, vía una ecuación de equilibrio.
- Determinar un equilibrio de Nash.

Para alcanzar estos objetivos, se supone que el espacio de acciones admisibles para el Jugador I es un espacio métrico compacto en cada estado, y que tanto la recompensa inmediata como las transiciones del sistema dependen de manera continua de la acción aplicada (véase Supuesto 1.1). Esta suposición es esencial para garantizar la existencia de políticas óptimas, como se explicará en detalle más adelante. Además, se han considerado dos supuestos distintos

para el análisis del modelo: uno basado en un modelo absorbente y otro en un modelo comunicante. En el modelo absorbente, se considera la existencia de un estado absorbente, que se denotará por z, y que presenta dos características principales: (i) tanto la recompensa inmediata como la recompensa terminal son nulas en este estado, y (ii) z es accesible desde cualquier estado inicial bajo cualquier política estacionaria (véase Supuesto 2.1). Un problema interesante es explorar modelos en los que no exista algún estado absorbente. Por ello, también se consideró el modelo de comunicación. En este modelo, se supone que si el Jugador II decide no detener el juego, la cadena de Markov inducida por cualquier política estacionaria adoptada por el Jugador I exhibe propiedades de comunicación y posee una distribución estacionaria (véase Supuesto 3.1).

En general, los procesos de decisión de Markov (PDMs) pueden verse como juegos estocásticos con un solo jugador. Se dispone de una teoría bien establecida de cadenas de Markov controladas [30, 19, 20], y se pueden encontrar aplicaciones, por ejemplo, en el libro de Boucherie y Van Dijk [9], donde se abordan temas relacionados con la detección y tratamiento de enfermedades, transporte, producción, comunicaciones y modelado financiero. En [6] se analizan aplicaciones en el ámbito financiero, mientras que en [7] se estudian criterios sensibles al riesgo.

En el contexto sensible al riesgo, la evaluación de la eficiencia de las políticas se realiza a través de la esperanza de una función de utilidad en lugar de limitarse a la esperanza de una recompensa acumulada. Este enfoque permite tener en cuenta las preferencias individuales del tomador de decisiones respecto al riesgo, ofreciendo así una visión más completa sobre la toma de decisiones en situaciones inciertas. Este concepto se fundamenta en el trabajo de Von Neumann y Morgenstern [35], donde se formalizó la teoría de la utilidad. Este libro es de gran importancia, ya que sentó las bases para la toma de decisiones en situaciones de incertidumbre y riesgo, introduciendo un marco teórico que ha influido profundamente en la economía, la teoría de juegos y otras disciplinas.

Por otro lado, en un contexto neutral al riesgo, el juego descrito anteriormente, con un espacio de estados finito y utilizando el criterio de recompensa total, fue analizado en [27]. Para el caso con espacio de estados numerable, se realizó un análisis en [15], considerando la existencia de un estado absorbente el cual es accesible desde cualquier otro estado. Las conclusiones obtenidas en estos dos artículos son extendidas en [11], donde se asume que bajo cualquier estrategia estacionaria del jugador I, el espacio de estados numerable es una clase recurrente positiva. Además, el caso descontado fue analizado en [12] y en [13].

La teoría de juegos tiene aplicaciones relevantes en diversas áreas, como se explora en [3, 5, 17, 23]. En cuanto a la teoría de los juegos Markovianos, sus fundamentos se encuentran en los artículos de Shapley [33] y Zachrisson [36]. El interés en estos juegos surge de diversas fuentes, siendo especialmente notable en el campo de las matemáticas financieras. En este contexto, muchos problemas se reducen a identificar el momento óptimo para ejecutar un contrato y la mejor estrategia para gestionar el riesgo asociado a la contraparte. Además, la

teoría de tiempos de paro desempeña un papel crucial en el análisis estocástico. Una descripción exhaustiva de esta teoría se puede encontrar en los trabajos de Shiryaev [34] y en Peskir y Shiryaev [29]. Las aplicaciones de esta teoría en las matemáticas financieras están bien documentadas en [8, 28]. En el presente trabajo, se integran las ideas fundamentales de paro óptimo con los PDMs para analizar el juego descrito anteriormente.

El enfoque de este trabajo se fundamenta en el operador T_λ (ver Definición 2.1). Este operador se define sobre un espacio de funciones apropiado, donde el principio de programación dinámica, el problema de paro óptimo y la función de utilidad empleada juegan un papel fundamental en su formulación. Uno de nuestros resultados iniciales es demostrar que este operador tiene puntos fijos. Este punto será crucial para definir las estrategias de los jugadores I y II, las cuales darán lugar a un equilibrio de Nash. Nuestro principal aporte en este trabajo es extender los resultados del caso neutral al contexto sensible al riesgo. Se consideran los dos modelos previamente mencionados, y los resultados más relevantes se presentan en los Teoremas 2.5 y 3.2. Para el modelo comunicante, ofrecemos un ejemplo ilustrativo que cumple con nuestros supuestos y, a partir de este caso particular, presentamos un método numérico para encontrar el punto fijo del operador T_λ y, posteriormente, la estrategia que constituye un equilibrio de Nash. Como resultado de nuestra investigación, se publicó el artículo [25] en 2022 y, más recientemente, el artículo [26] en 2024.

Este trabajo de tesis está organizado en tres capítulos. En el Capítulo 1 se presenta inicialmente la notación básica utilizada, así como una descripción detallada del modelo de decisión y sus componentes. Se analizan las estrategias de decisión admisibles para los jugadores, la sensibilidad al riesgo, la certeza equivalente, el criterio de rendimiento y la definición de un equilibrio de Nash. En el Capítulo 2 se analiza el modelo absorbente. Se presenta el operador T_λ y se destacan las características relevantes del estado absorbente z en relación con W_λ^*, el punto fijo de este operador. Aquí se incluyen también algunos resultados auxiliares que son fundamentales para demostrar la existencia de un equilibrio de Nash. En el Capítulo 3 se aborda el modelo comunicante, donde lo primero que se analizó fueron los resultados que se pierden al no considerar el estado absorbente. La propiedad de comunicación permite establecer directamente la unicidad de W_λ^*. Además, el resultado principal de esta sección es la existencia de un equilibrio de Nash, así como la igualdad entre la función valor del juego y W_λ^*. También se presenta un ejemplo específico de un juego que cumple con todos los supuestos considerados, el cual se analiza numéricamente para complementar la parte teórica. Finalmente, se presentan las conclusiones del trabajo y se plantean problemas futuros. Además, se incluye un apéndice que recopila los teoremas y definiciones utilizadas, seguido de la bibliografía.

Capítulo 1

Juegos Markovianos Sensibles al Riesgo

En este capítulo se ofrece una descripción detallada del modelo de control analizado. Se especifican cada una de las componentes del modelo, así como el espacio de estrategias de los jugadores I y II. También se detalla el criterio de rendimiento utilizado, junto con la definición del equilibrio de Nash.

Antes de avanzar, resulta conveniente introducir la notación básica que se empleará a lo largo del texto. Dado un espacio topológico X, el espacio de Banach $\mathcal{C}(X)$ consta de todas las funciones continuas $C : X \to \mathbb{R}$ cuya norma $\|C\|$ es finita, donde $\|C\| := \sup_{k\in X} |C(k)|$, mientras que $\mathbb{N} := \{0, 1, 2, \ldots\}$. La función indicadora de un evento A se denota por $I[A]$. Además, incluso sin mención explícita, todas las relaciones que involucran esperanzas condicionales son válidas con probabilidad 1 con respecto a la medida de probabilidad subyacente. Por otro lado, $a \wedge b$ y $a \vee b$ se usan como notaciones infijas para mín$\{a, b\}$ y máx$\{a, b\}$, respectivamente, donde $a, b \in \mathbb{R}$. El mínimo del conjunto vacío es $+\infty$ y, finalmente, se utilizará la siguiente convención relativa a las sumatorias:

$$\sum_{t=n}^{m} R(X_t, A_t) := 0, \quad m < n, \quad m, n \in \mathbb{N}. \tag{1.1}$$

1.1. El Modelo de Decisión

A lo largo del texto, $\mathcal{G} = (S, A, \{A(x), x \in S\}, P, R, G)$ representa un juego de suma cero en tiempo discreto de dos jugadores. Las componentes del juego $\mathcal{G}$ son las siguientes:

- S es el espacio de estados, el cual es un conjunto no vacío, numerable y está dotado con la topología discreta.
- A es el espacio de acciones, el cual es un espacio de Borel, es decir, un subconjunto de Borel de un espacio métrico completo y separable.
- $A(x) \subset A$ es el espacio de acciones admisibles para el jugador I en el

estado x, mientras que:

$$\mathbb{K} := \{(x,a)|a \in A(x), x \in S\} \subset S \times A\,,$$

es el correspondiente espacio de parejas estado-acción admisibles.

- $P = [p_{x,y}(\cdot)]_{x,y\in S}$ es la ley de transición en S dado $\mathbb{K}$, de modo que $p_{x,y}(a) \geq 0$ y $\sum_{y\in S} p_{x,y}(a) = 1$ para cada $(x,a) \in \mathbb{K}$.
- $R \in \mathcal{C}(\mathbb{K})$ es la función de recompensa inmediata y $G \in \mathcal{C}(S)$ la función de recompensa terminal.

El juego $\mathcal{G}$ se interpreta de la manera siguiente: en cada época de decisión $t \in \mathbb{N}$, los jugadores I y II observan el estado del sistema, denotado como $X_t = x \in S$. En este contexto, el jugador II debe elegir entre dos acciones: detener el sistema, pagando una recompensa terminal $G(x)$ al jugador I, o permitir que el sistema continúe su evolución. Si opta por esta última opción, el jugador I, utilizando el historial de estados hasta el tiempo t y las acciones anteriores a t, elige una acción $A_t = a \in A(x)$. Esta intervención tiene dos efectos: el jugador I obtiene una recompensa inmediata $R(x,a)$ del jugador II e, independientemente de los estados y acciones anteriores, el sistema transita a $X_{t+1} = y \in S$ con probabilidad $p_{x,y}(a)$; ésta es la propiedad de Markov del proceso de decisión.

Una forma efectiva de obtener ejemplos particulares de *Markov stopping games* es extender el problema de paro óptimo. En un problema de paro óptimo clásico, el sistema evoluciona como una cadena de Markov no controlada, donde un único jugador, en cada época de decisión se enfrenta a dos acciones: detener el sistema o continuar. Para enriquecer este escenario, podemos introducir un segundo jugador que influya en la evolución del sistema a través de sus decisiones. Esta interacción transforma la cadena en una cadena de Markov controlada, permitiendo que las acciones de ambos jugadores afecten el estado del juego. Esta extensión abre nuevas posibilidades para aplicar los *Markov stopping games* en diversos contextos, como la venta de activos, la resolución del problema clásico de la secretaria, y la valoración y ejercicio de opciones financieras, entre otros. A continuación, se presenta un ejemplo que ilustra cómo obtener esta extensión en el contexto de la venta de un activo.

Ejemplo 1.1. *Consideremos un inversor (Jugador I) que posee una propiedad o activo cuyo valor espera que aumente con el tiempo, y a un futuro comprador de dicho activo (Jugador II). En cada época de decisión, el Jugador II debe decidir si acepta la oferta que ha recibido del jugador I y compra la propiedad, o si rechaza esta oferta y solicita nuevas. Las acciones del Jugador I influyen en el sistema, de manera que la nueva oferta puede ser mayor, igual o menor que la oferta anterior. El espacio de estados representa todas las posibles ofertas que pueden surgir durante el horizonte de toma de decisiones.*

En cuanto a las funciones de recompensa, consideramos que la recompensa terminal se define como la función identidad, reflejando así el valor final del activo vendido. La recompensa inmediata se configura como una penalización para el jugador II, ya que se asigna un valor positivo cuando la acción del

Jugador I resulta en una disminución de la oferta, indicando que ha realizado una propuesta mas favorable para el jugador II, y lo incentiva a detener y comprar el activo. Por otro lado, la recompensa es cero si la acción del jugador I lleva a que la nueva oferta sea igual o mayor que la oferta actual.

Este ejemplo ilustra claramente la extensión mencionada anteriormente. Por otro lado, en el contexto del modelo de juego, asumimos el siguiente supuesto a lo largo de este trabajo.

Supuesto 1.1. *(i) Para cada $x \in S$, $A(x)$ es un subconjunto compacto de A.*

(ii) Para cada $x, y \in S$ los mapeos $a \mapsto R(x,a)$ y $a \mapsto p_{x,y}(a)$ son continuos en $a \in A(x)$.

(iii) Para cada $x \in S$, $a \in A(x)$, $G(x) \geq 0$ y $R(x,a) \geq 0$.

Este supuesto, ampliamente utilizado en los PDMs, nos permitirá demostrar la existencia de puntos fijos del operador de equilibrio T_λ y garantizar la existencia de la política del jugador I, la cual se utilizará para probar la existencia del equilibrio de Nash.

Dado que este trabajo busca extender el juego al contexto de la sensibilidad al riesgo, es fundamental introducir conceptos clave en este ámbito, como la sensibilidad al riesgo y la certeza equivalente. Estos conceptos se desarrollarán en la sección siguiente.

1.2. Sensibilidad al Riesgo y la Certeza Equivalente

La sensibilidad al riesgo es un factor crucial en la toma de decisiones, ya que influye en cómo los individuos evalúan y responden a situaciones inciertas. Al considerar el nivel de aversión o propensión al riesgo, los tomadores de decisiones pueden equilibrar posibles beneficios y pérdidas, optimizando así sus elecciones en entornos complejos. Esta comprensión es esencial no solo en contextos financieros, sino también en una variedad de campos, desde la economía hasta la salud pública, donde las decisiones pueden tener consecuencias significativas.

La evaluación de las estrategias utilizadas por los jugadores en un contexto neutral se fundamenta en la esperanza de la recompensa acumulada hasta que se detiene el juego, sin tener en cuenta el riesgo asociado a la elección de dichas estrategias. Por ejemplo, no se hace distinción entre garantizar una recompensa nula de forma segura y arriesgarse a ganar o perder una cantidad positiva con una probabilidad de $\frac{1}{2}$. Sin embargo, es evidente que, en este caso, el jugador enfrenta un riesgo significativo, que se manifiesta claramente en la varianza del resultado en este ejemplo.

Para evaluar el riesgo de manera adecuada, es fundamental considerar una función de utilidad, que es una representación matemática que refleja las preferencias de un individuo ante diferentes resultados inciertos. Esta función

asigna un valor numérico a cada posible resultado, permitiendo así cuantificar la satisfacción o el bienestar que cada uno de ellos proporciona. Al modelar cómo una persona valora diferentes niveles de riqueza y el riesgo asociado, la función de utilidad ayuda a entender su aversión o propensión al riesgo, lo que resulta crucial para la toma de decisiones informadas en situaciones de incertidumbre. Gracias al aporte de J. von Neumann y O. Morgenstern [35], esta noción se formaliza y se convierte en una herramienta clave para comparar y elegir entre distintas alternativas, guiando al individuo hacia la opción que maximiza su bienestar esperado.

En este contexto, que el jugador I posea un coeficiente de sensibilidad al riesgo constante, denotado por $\lambda \neq 0$, implica que una recompensa aleatoria Y es evaluada a través de la esperanza de su función de utilidad, expresada como $E[U_\lambda(Y)]$. Para cada $\lambda \neq 0$, la función de utilidad asociada, $U_\lambda : \mathbb{R} \to \mathbb{R}$, se define de la siguiente manera:

$$U_\lambda(u) := sgn(\lambda)e^{\lambda u}\,, u \in \mathbb{R}, \tag{1.2}$$

donde $sgn(\lambda)$ es la función signo de λ (Definición A.1).

Es importante destacar que $U_\lambda(\cdot)$ es una función estrictamente creciente, lo que asegura que un aumento en la recompensa conlleva un incremento correspondiente en la utilidad. Esta función de utilidad cumple con la siguiente propiedad:

$$U_\lambda(u+w) = e^{\lambda u}U_\lambda(w), \quad u, w \in \mathbb{R}. \tag{1.3}$$

Esta propiedad implica que la utilidad de una suma de recompensas puede descomponerse de manera exponencial y se utilizará en diversas ocasiones a lo largo del texto.

La elección de esta función de utilidad exponencial se justifica por la suposición de que el jugador I tiene un coeficiente de sensibilidad constante al riesgo. Al ser constante, la sensibilidad riesgo no se ve afectada por el tamaño de las recompensas, lo cual es útil en situaciones donde se quiere modelar de manera uniforme cómo un individuo reacciona ante la incertidumbre. Además, U_λ tiene propiedades como la (1.3) que facilitan considerablemente los cálculos en las demostraciones y permiten reducir las expresiones algebraicas a formas más manejables.

El signo de λ indica la actitud del jugador I frente al riesgo: si $\lambda > 0$, el jugador I es propenso al riesgo, lo que significa que está dispuesto a asumir riesgos a cambio de una posible mayor recompensa; en cambio, si $\lambda < 0$, el jugador I es averso al riesgo, prefiriendo certezas a resultados inciertos. Si el jugador I tiene la opción de elegir entre dos recompensas aleatorias Y_1 y Y_0, prefiere recibir Y_0 cuando se cumple la condición $E[U_\lambda(Y_0)] > E[U_\lambda(Y_1)]$. En caso de que ambas recompensas proporcionen la misma utilidad esperada, es decir, $E[U_\lambda(Y_0)] = E[U_\lambda(Y_1)]$, el jugador I se mostrará indiferente entre Y_0 y Y_1.

La certeza equivalente de una recompensa aleatoria acotada Y con respecto a la función de utilidad U_λ se define como la constante $\mathcal{E}_\lambda(Y) \in \mathbb{R} \cup \{-\infty, \infty\}$,

que satisface la relación:

$$U_\lambda(\mathcal{E}_\lambda(Y)) = E[U_\lambda(Y)],$$

lo que implica que el jugador I es indiferente entre recibir una recompensa aleatoria Y o la correspondiente certeza equivalente $\mathcal{E}_\lambda(Y)$. Obsérvese que la certeza equivalente puede expresarse de la siguiente manera:

$$\mathcal{E}_\lambda(Y) = \frac{1}{\lambda}\log(E[e^{\lambda Y}]). \tag{1.4}$$

La certeza equivalente es crucial en la toma de decisiones porque proporciona una forma de cuantificar la percepción del riesgo. Un individuo que prefiere la certeza a la incertidumbre mostrará una certeza equivalente que es menor o igual al valor esperado de la variable aleatoria Y. Este comportamiento se puede corroborar utilizando la desigualdad de Jensen (Teorema A.1) y la propiedad (1.4), ya que si $\lambda < 0$, se tiene que $\mathcal{E}_\lambda(Y) \leq E[Y]$.

Esta preferencia destaca la importancia de la función de utilidad en la toma de decisiones, ya que su forma determina la naturaleza de la certeza equivalente. Funciones de utilidad cóncavas, como las que a menudo se utilizan para representar la aversión al riesgo, resultarán en certezas equivalentes más bajas en comparación con aquellas funciones que representan una actitud neutral o favorable hacia el riesgo.

Antes de presentar el criterio de rendimiento que se considerará en este trabajo, es fundamental definir el espacio de estrategias de los jugadores I y II.

1.3. Estrategias de Decisión

Para cada $t \in \mathbb{N}$, el espacio de historias admisibles hasta el tiempo t, denotado como $\mathbb{H}_t$, se define de la siguiente manera: para $t = 0$, tenemos $\mathbb{H}_0 := S$, y para $t > 0$, se establece que $\mathbb{H}_t := \mathbb{K} \times \mathbb{H}_{t-1}$. Un elemento genérico de $\mathbb{H}_t$ se representa como $h_t = (x_0, a_0, \ldots, x_i, a_i, \ldots, x_t)$, donde $a_i \in A(x_i)$. Una política $\pi = \{\pi_t\}$ o estrategia de decisión para el jugador I es una sucesión especial de kérneles estocásticos definidos en el espacio de acciones A dado $\mathbb{H}_t$, donde para cada $t \in \mathbb{N}$ y $h_t \in \mathbb{H}_t$, se tiene que $\pi_t(\cdot|h_t)$ es una medida de probabilidad sobre A concentrada en $A(x_t)$, y para cada subconjunto Borel $B \subset A$ el mapeo $h_t \to \pi_t(B|h_t)$, $h_t \in \mathbb{H}_t$, es Borel medible. La clase de todas las políticas constituye la familia de estrategias admisibles para el jugador I y se denota por $\mathcal{P}$.

Por otro lado, cuando el jugador I maneja el sistema mediante π, el control A_t aplicado en el tiempo t pertenece a $B \subset A$ con probabilidad $\pi_t(B|h_t)$, donde $h_t \in \mathbb{H}_t$ es la historia observada del proceso hasta el tiempo t. Dados $\pi \in \mathcal{P}$ y el estado inicial $X_0 = x$, se determina de manera única una medida de probabilidad P_x^π en la σ-álgebra de Borel del espacio $\mathbb{H} := \prod_{t=0}^{\infty} \mathbb{K}$, que incluye todas las realizaciones posibles del proceso estado-acción $\{(X_t, A_t)\}$. El operador esperanza correspondiente se denota por E_x^π. A continuación,

definimos $\mathbb{F} := \prod_{x\in S} A(x)$ y observamos que $\mathbb{F}$ es un espacio métrico compacto, compuesto por todas las funciones $f : S \to A$ tales que $f(x) \in A(x)$ para cada $x \in S$. Una política π es estacionaria si existe $f \in \mathbb{F}$ tal que la medida de probabilidad $\pi_t(\cdot|h_t)$ está siempre concentrada en $f(x_t)$; en este caso, π y f se identifican naturalmente. Con esta convención, tenemos que $\mathbb{F} \subset \mathcal{P}$.

Asimismo, el espacio $\mathcal{T}$ de estrategias del jugador II está formado por todos los tiempos de paro $\tau : \mathbb{H} \to \mathbb{N} \cup \{\infty\}$ con respecto a la filtración $\{\mathcal{F}_t\}$ definida por:

$$\mathcal{F}_t := \sigma(X_0, A_0, \cdots, X_{t-1}, A_{t-1}, X_t), \tag{1.5}$$

lo que implica que el evento $[\tau = t] \in \mathcal{F}_t$ para cada $t \in \mathbb{N}$. Intuitivamente, esta condición significa que la decisión de parar o no al tiempo n debe basarse únicamente en la información disponible en ese momento, sin considerar ninguna información futura.

Una vez definidos los espacios de las estrategias de los jugadores I y II, es momento de presentar el criterio de rendimiento sensible al riesgo utilizado en este trabajo.

1.4. Criterio de Rendimiento

Dado el estado inicial $X_0 = x \in S$, supongamos que los jugadores I y II conducen el sistema utilizando las estrategias $\pi \in \mathcal{P}$ y $\tau \in \mathcal{T}$, respectivamente. La recompensa total (aleatoria) obtenida por el jugador I hasta que el sistema es detenido en el tiempo τ por el jugador II viene dada por:

$$\sum_{t=0}^{\tau-1} R(X_t, A_t) + G(X_\tau) I[\tau < \infty], \tag{1.6}$$

y la correspondiente certeza equivalente es el índice de rendimiento $V_\lambda(x; \pi, \tau)$ asociado con el par $(\pi, \tau) \in \mathcal{P} \times \mathcal{T}$ en el estado $x \in S$, el cual está dado por:

$$V_\lambda(x; \pi, \tau) = \frac{1}{\lambda} \log \left(E_x^\pi \left[e^{\lambda \left(\sum_{t=0}^{\tau-1} R(X_t, A_t) + G(X_\tau) I[\tau<\infty] \right)} \right] \right). \tag{1.7}$$

Este índice de rendimiento se obtiene a través de la expresión de la certeza equivalente (1.4), considerando la recompensa total (1.6). Dado que tanto R como G son no negativas, se tiene que:

$$V_\lambda(x; \pi, \tau) \geq 0. \tag{1.8}$$

Cuando el jugador II emplea la estrategia τ, el mayor valor de la certeza equivalente que puede alcanzar el jugador I es $\sup_{\pi\in\mathcal{P}} V_\lambda(x; \pi, \tau)$, el cual es una función de x y τ, digamos $\varphi(x; \tau)$. Se supone que el objetivo principal del jugador II es minimizar la utilidad esperada del jugador I, por lo que el jugador II se esforzará en emplear un tiempo de paro $\tilde{\tau}$ tal que $\varphi(x; \tilde{\tau})$ sea lo más cercano posible a $\inf_{\tau\in\mathcal{T}} \varphi(x; \tau)$. Esta última cantidad es el valor superior del juego y está determinado explícitamente por:

$$V_\lambda^*(x) := \inf_{\tau\in\mathcal{T}} \left[\sup_{\pi\in\mathcal{P}} V_\lambda(x; \pi, \tau) \right], x \in S. \tag{1.9}$$

Intercambiando el orden en que se toman el supremo y el ínfimo, se obtiene la siguiente función de valor inferior del juego:

$$V_{\lambda,*}(x) = \sup_{\pi\in\mathcal{P}}\left[\inf_{\tau\in\mathcal{T}} V_\lambda(x;\pi,\tau)\right], x \in S. \tag{1.10}$$

Dado que $\inf_{\tau\in\mathcal{T}} V_\lambda(x;\pi,\tau) \leq V_\lambda(x;\pi,\tau) \leq \sup_{\pi\in\mathcal{P}} V_\lambda(x;\pi,\tau)$, estas definiciones conducen inmediatamente a que:

$$V_{\lambda,*}(\cdot) \leq V^*_\lambda(\cdot)\,. \tag{1.11}$$

Por lo que vemos que la desigualdad anterior entre el valor superior e inferior del juego siempre se cumple. Uno de nuestros objetivos es demostrar que bajo ciertas condiciones, también se cumple la desigualdad contraria, lo cual implica la existencia de un único valor del juego. Para concluir este capítulo, hace falta introducir el concepto del equilibrio de Nash, el cual es uno de los elementos más importantes en este trabajo.

1.5. Estrategias de Equilibrio.

El principal objetivo de este trabajo es establecer la existencia de un par de estrategias $(\pi^*,\tau^*) \in \mathcal{P}\times\mathcal{T}$ que sea un equilibrio de Nash para el juego, cuya definición se presenta a continuación.

Definición 1.1. *El par $(\pi^*,\tau^*) \in \mathcal{P}\times\mathcal{T}$ es un equilibrio de Nash si para cada estado $x \in S$*

$$V_\lambda(x;\pi,\tau^*) \leq V_\lambda(x;\pi^*,\tau^*) \leq V_\lambda(x;\pi^*,\tau)\,,\ \pi\in\mathcal{P}, \tau\in\mathcal{T}\,. \tag{1.12}$$

Analicemos el cómo se interpretan las dos desigualdades presentes en la definición de un equilibrio de Nash. Cuando las estrategias π^* y τ^* realmente usadas por los jugadores I y II forman un equilibrio de Nash, de la primera desigualdad en (1.12) se deduce que, si el jugador II continúa usando la estrategia τ^*, entonces el jugador I no tiene ningún incentivo para cambiar a otra política. Esto se debe a que si decide hacerlo se verá perjudicado ya que obtendría una recompensa menor. De manera similar, la segunda desigualdad en (1.12) implica que, si el jugador I continúa usando π^*, entonces el jugador II no tiene ninguna motivación para cambiar la estrategia τ^* en uso. Si decidiera hacerlo, la recompensa que tendría que pagarle al jugador I sería mayor.

Además, nótese que si (π^*,τ^*) es un equilibrio de Nash, entonces (1.12) implica que:

$$V^*_\lambda(\cdot) \leq \sup_{\pi\in\mathcal{P}} V_\lambda(\cdot;\pi,\tau^*) \leq V_\lambda(\cdot;\pi^*,\tau^*) \leq \inf_{\tau\in\mathcal{T}} V_\lambda(\cdot;\pi^*,\tau) \leq V_{\lambda,*}(\cdot),$$

donde las desigualdades de la izquierda y de la derecha se deben a (1.9) y (1.10), respectivamente, por lo que a través de (1.11) se deduce que las funciones de valor superior e inferior son iguales y coinciden con $V_\lambda(\cdot;\pi^*,\tau^*)$.

En el siguiente capítulo se aborda el problema de encontrar un par de estrategias que sea un equilibro de Nash. Para ello en primer lugar se definirá un operador cuyo punto fijo se utiliza para definir las estrategias de los jugadores que conforman un equilibro de Nash. El análisis se lleva a cabo bajo los Supuestos 1.1 y 2.1.

Capítulo 2

Modelo con un Estado Absorbente

En este capítulo se presentan los primeros resultados obtenidos en la investigación. Se estudia el modelo considerando el supuesto de la existencia de un estado absorbente (Supuesto 2.1) y se demuestra la existencia de un equilibrio de Nash. Los resultados presentados han sido publicados en el artículo [25]. La principal condición estructural es la existencia de un estado absorbente que, independientemente de las estrategias de los jugadores, puede ser alcanzado eventualmente desde cualquier estado inicial. Lo cual se establece en el siguiente supuesto.

Supuesto 2.1. *Existe un estado $z \in S$ para el cual se cumplen las siguientes condiciones:*

(i) Para cada $x \in S$ y $f \in \mathbb{F}$,

$$P_x^f[\tau_z < \infty] = 1, \tag{2.1}$$

donde

$$\tau_z := \min\{n \mid X_n = z\}. \tag{2.2}$$

(ii) $G(z) = 0 = R(z, a)$ y $p_{z,z}(a) = 1$, $a \in A(z)$.

Nótese que τ_z es un tiempo de paro con respecto a la filtración $\{\mathcal{F}_t\}$ dada en (1.5), por lo que una consecuencia directa de esto es que $\tau_z \in \mathcal{T}$. Por otro lado, una vez que el sistema alcance el estado z, no podrá salir de allí y la recompensa acumulada a partir de ese momento será cero. Además, es importante señalar que:

$$X_{\tau_z} = z \text{ en el evento } [\tau_z < \infty]. \tag{2.3}$$

A continuación, se define el operador de equilibrio, el cual es crucial para determinar las estrategias óptimas de los jugadores I y II.

2.1. Operador de Equilibrio

Para encontrar el equilibrio de Nash, en primer lugar se introduce un subconjunto de $\mathcal{C}(S)$ y se define un operador sobre este subconjunto.

Definición 2.1. *(i) Sea G la función de recompensa terminal. El espacio $[\![0, G]\!] \subset \mathcal{C}(S)$ se define como:*

$$[\![0, G]\!] := \{h \in \mathcal{C}(S) \mid 0 \leq h(x) \leq G(x),\ x \in S\}. \tag{2.4}$$

(ii) El operador $T_\lambda : [\![0, G]\!] \to [\![0, G]\!]$ es determinado de la siguiente manera: Para cada $W \in [\![0, G]\!]$ y $x \in S$,

$$\begin{aligned} &U_\lambda(T_\lambda[W](x)) \\ &\quad := \text{mín}\left\{U_\lambda(G(x)),\ \sup_{a\in A(x)}\left[\sum_{y\in S} p_{x,y}(a)U_\lambda(R(x,a)+W(y))\right]\right\}. \end{aligned} \tag{2.5}$$

El operador T_λ modela la decisión secuencial del jugador II, quien debe elegir entre parar, pagando la recompensa terminal $G(x)$, o continuar, lo que implica pagar la recompensa inmediata $R(x, a)$ junto con las recompensas futuras que tendrá que pagar en el nuevo estado y, las cuales dependen de las acciones del jugador I.

Este operador tiene algunas propiedades importantes, las cuales se presentan a continuación.

- $T_\lambda[W] \in [\![0, G]\!]$, para toda $W \in [\![0, G]\!]$. Esto lo podemos verificar usando que $U_\lambda(\cdot)$ es creciente y que R y G son no negativas.
- La relación entre el estado absorbente z y T_λ es la siguiente:

$$T_\lambda[W](z) = W(z) = 0, \quad W \in [\![0, G]\!]. \tag{2.6}$$

- Definimos el orden $\leq$ en el espacio de funciones $[\![0, G]\!]$ de la siguiente manera: $V \leq W$ si y solo si $V(x) \leq W(x)$ para todo $x \in S$. Con esta definición, T_λ es un operador monótono creciente, es decir, para $V, W \in [\![0, G]\!]$ se tiene que:

$$V \leq W \Rightarrow T_\lambda[V] \leq T_\lambda[W]. \tag{2.7}$$

Para demostrar esta última propiedad, sean $y \in S$,$(x, a) \in \mathbb{K}$ y $V, W \in [\![0, G]\!]$ tales que $V \leq W$. Así, obtenemos que:

$$\begin{aligned} &\sup_{a\in A(x)}\left[\sum_{y\in S} p_{xy}(a)U_\lambda(R(x,a)+V(y))\right] \\ &\qquad \leq \sup_{a\in A(x)}\left[\sum_{y\in S} p_{xy}(a)U_\lambda(R(x,a)+W(y))\right]. \end{aligned}$$

Por otro lado,

$$\begin{aligned} U_\lambda(T_\lambda[V](x)) &= \min\left\{U_\lambda(G(x)),\ \sup_{a\in A(x)}\left[\sum_{y\in S} p_{x,y}(a)U_\lambda(R(x,a)+V(y))\right]\right\} \\ &\leq \min\left\{U_\lambda(G(x)),\ \sup_{a\in A(x)}\left[\sum_{y\in S} p_{x,y}(a)U_\lambda(R(x,a)+W(y))\right]\right\} \\ &= U_\lambda(T_\lambda[W](x)), \end{aligned}$$

de donde se sigue que $T_\lambda[V] \leq T_\lambda[W]$.

Otra característica importante de T_λ es que este operador tiene puntos fijos. Para demostrarlo, se necesitan un par de resultados previos: el Lema 2.1 y el Teorema 2.1, los cuales se presentan a continuación.

Lema 2.1. *(i) Consideremos una familia $\{S_k\}$ de subconjuntos finitos de S tal que:*

$$S = \bigcup_{k=1}^{\infty} S_k\,,\ S_k \subset S_{k+1}\,,\ k \in \mathbb{N}\,, \tag{2.8}$$

y para cada $x \in S$, $k \in \mathbb{N}$ definimos:

$$\delta_k(x) := \sup_{a\in A(x)}\left[1 - \sum_{y\in S_k} p_{xy}(a)\right] = \sup_{a\in A(x)} \sum_{y\in S\setminus S_k} p_{xy}(a), \tag{2.9}$$

entonces,

$$\lim_{k\to\infty} \delta_k(x) = 0\,,\ x \in S\,.$$

(ii) Si $\{W_n\} \subset \mathcal{C}(S)$ es tal que:

$$c := \sup_{n\in\mathbb{N}} \|W_n\| < \infty \quad y \quad \lim_{n\to\infty} W_n(y) = 0\,,\ y \in S. \tag{2.10}$$

En este caso, para cada $x \in S$

$$\sup_{a\in A(x)}\left[e^{\lambda R(x,a)} \sum_{y\in S} p_{xy}(a)|W_n(y)|\right] \to 0 \ \text{ cuando } \ n \to \infty.$$

Demostración. (i) Dado que los conjuntos S_k son finitos, del Supuesto 1.1 se obtiene que para cada $k \in \mathbb{N}$ y $x \in S$ el mapeo $a \mapsto \sum_{y\in S_k} p_{x,y}(a)$ es continuo en el compacto $A(x)$, mientras que utilizando las condiciones en (2.8) se deduce que:

$$\sum_{y\in S_k} p_{xy}(a) \nearrow \sum_{y\in S} p_{xy}(a) = 1 \ \text{ cuando } \ k \to \infty,$$

de modo que el Teorema de Dini (Teorema A.2) implica que la convergencia es uniforme en el espacio $A(x)$, es decir:

$$\sup_{a\in A(x)}\left[1 - \sum_{y\in S_k} p_{xy}(a)\right] \to 0 \text{ cuando } \ k \to \infty.$$

(ii) Fijemos $x \in S$ y para cada $k \in \mathbb{N}$ se tiene que:

$$\begin{aligned}
&\sup_{a\in A(x)} e^{\lambda R(x,a)} \sum_{y\in S} p_{x,y}(a)|W_n(y)| \\
&\quad \leq \sup_{a\in A(x)} e^{\lambda R(x,a)} \sum_{y\in S_k} p_{x,y}(a)|W_n(y)| + \sup_{a\in A(x)} e^{\lambda R(x,a)} \sum_{y\in S\setminus S_k} p_{x,y}(a)|W_n(y)| \\
&\quad \leq e^{|\lambda|\,\|R\|} \left(\max_{y\in S_k} |W_n(y)| + c \sup_{a\in A(x)} \sum_{y\in S\setminus S_k} p_{x,y}(a) \right) \\
&\quad = e^{|\lambda|\,\|R\|} \left(\max_{y\in S_k} |W_n(y)| + c\delta_k(x) \right),
\end{aligned}$$

donde (2.10) se utilizó para establecer la segunda desigualdad, y la igualdad se debe a (2.9). Recordando que los conjuntos S_k son finitos, la convergencia en (2.10) produce que:

$$\limsup_{n\to\infty} \left| \sup_{a\in A(x)} e^{\lambda R(x,a)} \sum_{y\in S} p_{x,y}(a)|W_n(y)| \right| \leq e^{|\lambda|\,\|R\|} c\,\delta_k(x), \quad x \in S,$$

y entonces, como $k \in \mathbb{N}$ es arbitrario, la conclusión se desprende de la parte (i). □

El siguiente resultado establece que T_λ es un operador continuo con respecto a la topología de la convergencia puntual en el espacio $[\![0, G]\!]$.

Teorema 2.1. *Supongamos que la sucesión $\{W_n\} \subset [\![0, G]\!]$ converge puntualmente a una función $V : S \to \mathbb{R}$, esto es,*

$$\lim_{n\to\infty} W_n(x) = V(x)\,,\ x \in\ S\,. \tag{2.11}$$

Entonces se tiene que:

$$V \in [\![0, G]\!]\ y\ \lim_{n\to\infty} T_\lambda[W_n](x) = T_\lambda[V](x)\,,\ x \in\ S\,.$$

Demostración. Nótese que (2.4) y (2.11) implican que $V \in [\![0, G]\!]$. Sea

$$\triangle_n(x) := \sup_{a\in A(x)} \left[e^{\lambda R(x,a)} \sum_{y\in S} p_{xy}(a)|U_\lambda(W_n(y)) - U_\lambda(V(y))| \right]. \tag{2.12}$$

Luego, usando (1.3) observe que:

$$\begin{aligned}
&\sup_{a\in A(x)} \sum_{y\in S} p_{x,y}(a) U_\lambda(R(x,a)+W_n(y)) \\
&\quad = \sup_{a\in A(x)} \left[e^{\lambda R(x,a)} \sum_{y\in S} p_{x,y}(a) U_\lambda(W_n(y)) \right] \\
&\quad = \sup_{a\in A(x)} \left[e^{\lambda R(x,a)} \sum_{y\in S} p_{x,y}(a) U_\lambda(V(y)) \right. \\
&\qquad\qquad \left. + e^{\lambda R(x,a)} \sum_{y\in S} p_{x,y}(a)[U_\lambda(W_n(y)) - U_\lambda(V(y))] \right] \\
&\quad \le \sup_{a\in A(x)} \left[e^{\lambda R(x,a)} \sum_{y\in S} p_{x,y}(a) U_\lambda(V(y)) \right] \\
&\qquad\qquad + \sup_{a\in A(x)} \left[e^{\lambda R(x,a)} \sum_{y\in S} p_{x,y}(a) |U_\lambda(W_n(y)) - U_\lambda(V(y))| \right],
\end{aligned}$$

y una aplicación adicional de (1.3) conduce a

$$\begin{aligned}
&\sup_{a\in A(x)} \sum_{y\in S} p_{x,y}(a) U_\lambda(R(x,a)+W_n(y)) \\
&\qquad \le \sup_{a\in A(x)} \sum_{y\in S} p_{x,y}(a) U_\lambda(R(x,a)+V(y)) + \Delta_n(x),
\end{aligned} \tag{2.13}$$

mientras que la desigualdad

$$\begin{aligned}
&\sup_{a\in A(x)} \sum_{y\in S} p_{x,y}(a) U_\lambda(R(x,a)+V(y)) \\
&\qquad \le \sup_{a\in A(x)} \sum_{y\in S} p_{x,y}(a) U_\lambda(R(x,a)+W_n(y)) + \Delta_n(x),
\end{aligned}$$

puede establecerse de forma similar. Combinando la definición de T_λ en (2.5) con (2.13) y la desigualdad anterior, resulta que:

$$U_\lambda(T_\lambda[W_n](x)) \le U_\lambda(T_\lambda[V](x)) + \Delta_n(x)$$

y

$$U_\lambda(T_\lambda[V](x)) \le U_\lambda(T_\lambda[W_n](x)) + \Delta_n(x),$$

de modo que:

$$|U_\lambda(T_\lambda[W_n](x)) - U_\lambda(T_\lambda[V](x))| \le \Delta_n(x). \tag{2.14}$$

Observe ahora que (1.2) y (2.11) implican que:

$$\lim_{n\to\infty} [U_\lambda(W_n(y)) - U_\lambda(V(y))] = 0, \quad y \in S.$$

Además, utilizando que $\|W\| \le \|G\| < \infty$, si $W \in [\![0, G]\!]$, las inclusiones $W_n, V \in [\![0, G]\!]$ y (1.2) dan como resultado que $\|U_\lambda(W_n(\cdot))\|, \|U_\lambda(V(\cdot))\| \le e^{|\lambda|\,\|G\|}$. Por lo tanto, se tiene que:

$$\|U_\lambda(W_n(\cdot)) - U_\lambda(V(\cdot))\| \le 2e^{|\lambda|\|G\|}.$$

Utilizando el Lema 2.1(ii) con $U_\lambda(W_n) - U_\lambda(V)$ en lugar de W_n, los dos hechos anteriores y (2.12) implican que $\lim_{n\to\infty} \Delta_n(\cdot) = 0$, una convergencia que a través de (2.14) conduce a que:

$$\lim_{n\to\infty} U_\lambda(T_\lambda[W_n](x)) = U_\lambda(T_\lambda[V](x)),$$

para cada $x \in S$. Por otro lado, como $U_\lambda(\cdot)$ es estrictamente creciente y continua, se deduce que $T_\lambda[W_n](x) \to T_\lambda[V](x)$ cuando $n \to \infty$ para todo estado x. □

2.2. Existencia de puntos fijos

El resultado que demuestra la existencia de puntos fijos del operador T_λ se presenta a continuación, y la prueba está apoyada en la propiedad presentada en el Teorema 2.1.

Teorema 2.2. *Bajo el Supuesto 1.1, se tiene que existe un punto fijo del operador T_λ, esto es, existe una función $W_\lambda^* \in [\![0, G]\!]$ que satisface que:*

$$W_\lambda^* = T_\lambda[W_\lambda^*]\,. \tag{2.15}$$

Demostración. Definimos $W_{n,\lambda} := 0$ para $n = 0$ y $W_{n,\lambda} := T_\lambda^n[0]$ para $n \in \mathbb{N} \setminus \{0\}$. Observemos que, para cada $n \in \mathbb{N}$, se tiene que:

$$\begin{aligned} W_{n+1,\lambda} &= T_\lambda^{n+1}[0] \\ &= T_\lambda[T_\lambda^n[0]] \\ &= T_\lambda[W_{n,\lambda}]. \end{aligned} \tag{2.16}$$

Por otro lado, $W_{0,\lambda} = 0 \in [\![0, G]\!]$ y $W_{1,\lambda} = T_\lambda[0] \in [\![0, G]\!]$, de donde se deduce que $W_{0,\lambda} \leq W_{1,\lambda}$. Ahora supongamos que esta propiedad se cumple para $n \in \mathbb{N}$, es decir $W_{n,\lambda} \leq W_{n+1,\lambda}$ y probemos que se cumple para $n + 1$.

$$\begin{aligned} W_{n,\lambda} \leq W_{n+1,\lambda} &\Rightarrow T_\lambda[W_{n,\lambda}] \leq T_\lambda[W_{n+1,\lambda}] \\ &\Rightarrow W_{n+1,\lambda} \leq W_{n+2,\lambda}, \end{aligned}$$

donde se utilizó las propiedades en (2.7) y (2.16). Además, como las funciones $W_{k,\lambda}$ pertenecen a $[\![0, G]\!]$ se sigue que:

$$0 \leq W_{n,\lambda} \leq W_{n+1,\lambda} \leq G\,.$$

Así, para cada $y \in S$ la sucesión $\{W_{n,\lambda}(y)\}$ es creciente, acotada y por la tanto convergente. Luego por el Lema 2.1, existe $\hat{W} \in [\![0, G]\!]$ tal que:

$$\lim_{n\to\infty} W_{n,\lambda}(y) = \hat{W}(y),\ y \in S\,,$$

y además

$$\lim_{n\to\infty} T_\lambda[W_{n,\lambda}](x) = T_\lambda[\hat{W}](x)\,,\ x \in\ S\,.$$

Luego, tomando límite cuando n tiende a ∞, en ambos lados de (2.16), junto con lo mostrado previamente, nos conduce a que:

$$\hat{W} = T_\lambda[\hat{W}]\,.$$

Esto muestra que $\hat{W}$ es un punto fijo de T_λ. □

Mediante la Definición 2.1, la expresión presentada en (2.5) puede escribirse de forma equivalente como sigue: Para todo $x \in S$,

$$\begin{aligned} &U_\lambda(W_\lambda^*(x)) \\ &\quad = \text{mín}\left\{U_\lambda(G(x)),\ \sup_{a\in A(x)} \sum_{y\in S} p_{x,y}(a) U_\lambda(R(x,a) + W_\lambda^*(y))\right\}. \end{aligned} \tag{2.17}$$

Además, utilizando que G está acotada, la inclusión $W_\lambda^* \in [\![0, G]\!]$ y el Supuesto 1.1 implican que existe una política $f^* \in \mathbb{F}$ tal que, para todo $x \in S$,

$$\begin{aligned} &\sum_{y\in S} p_{x,y}(f^*(x)) U_\lambda(R(x, f^*(x)) + W_\lambda^*(y)) \\ &\qquad = \sup_{a\in A(x)} \left[\sum_{y\in S} p_{x,y}(a) U_\lambda(R(x,a) + W_\lambda^*(y))\right]. \end{aligned} \tag{2.18}$$

Asimismo, observando que $W_\lambda^* \geq 0$, el Supuesto 2.1(ii) y (2.17) implican que $U_\lambda(W_\lambda^*(z)) = U_\lambda(G(z)) = U_\lambda(0)$, y entonces

$$W_\lambda^*(z) = 0 = G(z). \tag{2.19}$$

Las estrategias para los jugadores I y II que conforman un equilibrio de Nash se definen utilizando el punto fijo W_λ^*, para ello definimos el subconjunto S^* del espacio de estados como sigue

$$S^* := \{x \in S \mid W_\lambda^*(x) = G(x)\}, \tag{2.20}$$

y sea τ^* el tiempo de alcance al conjunto S^*, esto es:

$$\tau^* := \text{mín}\{n \in \mathbb{N} \mid X_n \in S^*\}, \tag{2.21}$$

de modo que τ^* es un tiempo de paro con respecto a la filtración $\{\mathcal{F}_t\}$ en (1.5), es decir, τ^* pertenece al espacio $\mathcal{T}$ de estrategias admisibles para el jugador II. A partir de este punto, cuando se aparezcan f^* y τ^*, se entenderá que nos referimos a las definidas en (2.18) y (2.21) respectivamente.

Por otro lado, se tiene que bajo el Supuesto 2.1, $S^* \neq \emptyset$, ya que $z \in S^*$, por (2.19) y (2.20), y entonces

$$\tau^* \leq \tau_z. \tag{2.22}$$

2.3. Equilibrio de Nash

En el marco determinado por los Supuestos 1.1 y 2.1, se demostrará que existe un equilibrio de Nash con respecto al índice de recompensa total sensible al riesgo (1.7). Antes de presentar el resultado principal, nuestra atención estará centrada en algunos resultados auxiliares que se utilizarán en la demostración de este.

El inciso (ii) del Lema 2.2 extenderá la propiedad del inciso (i) en el Supuesto 2.1 a la clase de todas las políticas del jugador I.

Lema 2.2. *Para cada $x \in S$, y $n \in \mathbb{N}$, definimos*

$$M_n(x) := \sup_{\pi \in \mathcal{P}} P_x^\pi[\tau_z > n] \in [0,1]\,. \tag{2.23}$$

Con esta notación, las siguientes afirmaciones son válidas:

(i) $\lim_{n\to\infty} M_n(x) = 0,\ x \in S\,.$

(ii) $P_x^\pi[\tau_z < \infty] = 1$ *para cada* $x \in S$ *y* $\pi \in \mathcal{P}$.

Demostración. Obsérvese que la inclusión $[\tau_z > n+1] \subset [\tau_z > n]$ y (2.23) conducen a

$$M_{n+1} \leq M_n, \quad n \in \mathbb{N}, \tag{2.24}$$

y entonces

$$M(x) := \lim_{n\to\infty} M_n(x) \in [0,1] \tag{2.25}$$

existe para todo $x \in S$; como $P_z^\pi[\tau_z = 0] = 1$ para todo $\pi \in \mathcal{P}$, por (2.2), se sigue que $M_n(z) = 0$ para todo n positivo, así que

$$M(z) = 0. \tag{2.26}$$

Dado $(x,\tilde{a}) \in \mathbb{K}$ y una política $\pi \in \mathcal{P}$, definimos la nueva política $\pi_{x,\tilde{a}} = \{\pi_{x,\tilde{a},n}\}$ como sigue: para cada $t \in \mathbb{N}$ y $h_t \in \mathbb{H}_t$, $\pi_{x,\tilde{a},t}(\cdot|h_t) = \pi_{t+1}(\cdot|x,\tilde{a},h_t)$. Luego, usando (2.2), observemos que $[\tau_z > n+1] = [X_k \neq z, 0 \leq k \leq n+1]$ y que una aplicación de la propiedad de Markov nos da que para cada $\pi \in \mathcal{P}$, $n \in \mathbb{N}$ y $(x,\tilde{a}) \in \mathbb{K}$ con $x \neq z$

$$\begin{aligned} P_x^\pi[\tau_z > n+1 | A_0 = \tilde{a}] &= \sum_{y \in S\setminus\{z\}} p_{x,y}(\tilde{a}) P_y^{\pi_{x,\tilde{a}}}[\tau_z > n] \\ &\leq \sum_{y \in S\setminus\{z\}} p_{x,y}(\tilde{a}) M_n(y) \\ &\leq \sup_{a \in A(x)} \sum_{y \in S\setminus\{z\}} p_{x,y}(a) M_n(y), \end{aligned}$$

donde la primera desigualdad se debe a (2.23). Por lo tanto,

$$P_x^\pi[\tau_z > n+1] \leq \sup_{a \in A(x)} \sum_{y \in S\setminus\{z\}} p_{x,y}(a) M_n(y), \quad x \neq z.$$

Utilizando el teorema de convergencia dominada (Teorema A.3), junto con (2.23) y (2.25), se deduce que:

$$M(x) \leq \sup_{a\in A(x)} \sum_{y\in S\setminus\{z\}} p_{x,y}(a)M(y), \quad x \in S.$$

Ahora, usando que $M(\cdot)$ está acotado, observemos que el Supuesto 1.1 implica que existe una política $\hat{f} \in \mathbb{F}$ tal que $\sup_{a\in A(x)} \sum_{y\in S\setminus\{z\}} p_{x,y}(a)M(y) = \sum_{y\in S\setminus\{z\}} p_{x,y}(\hat{f}(x))M(y)$ para todo estado x, y entonces

$$M(x) \leq \sum_{y\in S\setminus\{z\}} p_{x,y}(\hat{f}(x))M(y) = \sum_{y\in S} p_{x,y}(\hat{f}(x))M(y), \quad x \in S;$$

ver (2.26) para la igualdad. Combinando esta relación con la propiedad de Markov, se deduce que para cada estado inicial $x \in S$ y $n \in \mathbb{N}$,

$$M(X_n) \leq E_x^{\hat{f}}[M(X_{n+1})|X_n] = E_x^{\hat{f}}[M(X_{n+1})|\mathcal{F}_n], \quad P_x^{\hat{f}}\text{-c. s.},$$

por lo que $\{(M(X_n), \mathcal{F}_n)\}$ es una submartingala con respecto a $P_x^{\hat{f}}$. Dado que $M(\cdot)$ está acotado, el teorema de paro opcional da como resultado que, para cada $x \in S$ y $n \in \mathbb{N}$,

$$M(x) \leq E_x^{\hat{f}}[M(X_{\tau_z\wedge n})] = E_x^{\hat{f}}[M(X_n)\ I[\tau_z > n]] \leq P_x^{\hat{f}}[\tau_z > n],$$

donde, recordando que $M(z) = 0$, la igualdad se obtuvo de (2.3), y la inclusión en (2.25) se utilizó en el último paso. Dado que:

$$\lim_{n\to\infty} P_x^{\hat{f}}[\tau_z > n] = P_x^{\hat{f}}[\tau_z = \infty] = 0,$$

por el Supuesto 2.1(i), lo anterior da como resultado que $M(\cdot) = 0$, estableciendo la parte (i). Para establecer la afirmación (ii), combinamos (2.23) con la parte (i) para obtener:

$$\begin{aligned} P_x^{\pi}[\tau_z = \infty] &= \lim_{n\to\infty} P_x^{\pi}[\tau_z > n] \\ &\leq \lim_{n\to\infty} M_n(x) \\ &= M(x) \\ &= 0, \forall x \in S \text{ y } \forall \pi \in \mathcal{P}. \end{aligned}$$

□

El Lema 2.3 a continuación muestra que el espacio de estrategias del jugador II puede reducirse a la clase de tiempos de paro finitos.

Lema 2.3. *Para todo* $(\pi, \tau) \in \mathcal{P} \times \mathcal{T}$,

$$V_\lambda(\cdot, \pi, \tau) = V_\lambda(\cdot, \pi, \tau \wedge \tau_z). \tag{2.27}$$

Demostración. Sean $x \in S$ y $(\pi, \tau) \in \mathcal{P} \times \mathcal{T}$. Usando que $P_x^{\pi}[\tau_z < \infty] = 1$, por el Lema 2.2, los Supuestos 1.1(ii) y 2.1 junto con (2.2) dan como resultado que:

$$\text{en } [\tau_z < \infty], \ X_{\tau_z} = z \ \text{ y } \ R(X_n, A_n) = G(X_n) = 0 \ \text{ para } \ n \geq \tau_z. \tag{2.28}$$

Ahora, consideremos los siguientes escenarios:

- $[\tau = \infty] \cap [\tau_z < \infty]$

En este caso tenemos que $\tau \wedge \tau_z = \tau_z$, y además se tiene que $R(X_t, A_t) = 0$ para $t \geq \tau \wedge \tau_z$ y $G(X_{\tau\wedge\tau_z}) = 0$, así que

$$\sum_{t=0}^{\tau-1} R(X_t, A_t) = \sum_{t=0}^{\tau\wedge\tau_z-1} R(X_t, A_t)$$

y

$$G(X_\tau)I[\tau < \infty] = 0 = G(X_{\tau\wedge\tau_z})I[\tau \wedge \tau_z < \infty].$$

Por lo tanto,

$$\begin{aligned}&\sum_{t=0}^{\tau-1} R(X_t, A_t) + G(X_\tau)I[\tau < \infty] \\ &\quad = \sum_{t=0}^{\tau\wedge\tau_z-1} R(X_t, A_t) + G(X_{\tau\wedge\tau_z})I[\tau \wedge \tau_z < \infty] \quad \text{en} \quad [\tau = \infty, \tau_z < \infty];\end{aligned}$$

como $P_x^\pi[\tau_z < \infty] = 1$, por el Lema 2.2$(ii)$, se sigue que

$$\begin{aligned}&E_x^\pi\left[I[\tau = \infty]U_\lambda\left(\sum_{t=0}^{\tau-1} R(X_t, A_t) + G(X_\tau)I[\tau < \infty]\right)\right] \\ &\quad = E_x^\pi\left[I[\tau = \infty]U_\lambda\left(\sum_{t=0}^{\tau\wedge\tau_z-1} R(X_t, A_t) + G(X_{\tau\wedge\tau_z})I[\tau \wedge \tau_z < \infty]\right)\right].\end{aligned}$$

- $[\tau_z \leq \tau < \infty]$

En este caso $\tau_z = \tau \wedge \tau_z$ y mediante (2.28) se deduce que

$$G(X_\tau)I[\tau < \infty] = G(X_\tau) = 0 = G(X_{\tau_z}) = G(X_{\tau\wedge\tau_z})I[\tau \wedge \tau_z < \infty]$$

así como

$$\sum_{t=0}^{\tau-1} R(X_t, A_t) = \sum_{t=0}^{\tau_z-1} R(X_t, A_t) = \sum_{t=0}^{\tau\wedge\tau_z-1} R(X_t, A_t),$$

de modo que

$$\begin{aligned}&E_x^\pi\left[I[\tau_z \leq \tau < \infty]\, U_\lambda\left(\sum_{t=0}^{\tau-1} R(X_t, A_t) + G(X_\tau)I[\tau < \infty]\right)\right] \\ &\quad = E_x^\pi\left[I[\tau_z \leq \tau < \infty]\, U_\lambda\left(\sum_{t=0}^{\tau\wedge\tau_z-1} R(X_t, A_t) + G(X_{\tau\wedge\tau_z})I[\tau \wedge \tau_z < \infty]\right)\right].\end{aligned}$$

- $[\tau < \infty, \tau < \tau_z]$

En este último caso $\tau = \tau \wedge \tau_z$, por lo que

$$\sum_{t=0}^{\tau-1} R(X_t, A_t) + G(X_\tau)I[\tau < \infty] = \sum_{t=0}^{\tau\wedge\tau_z-1} R(X_t, A_t) + G(X_{\tau\wedge\tau_z})I[\tau \wedge \tau_z < \infty],$$

y entonces

$$E_x^{\pi}\left[I[\tau<\infty,\tau<\tau_z]\,U_\lambda\left(\sum_{t=0}^{\tau-1}R(X_t,A_t)+G(X_\tau)I[\tau<\infty]\right)\right]$$
$$=E_x^{\pi}\left[I[\tau<\infty,\tau<\tau_z]\,U_\lambda\left(\sum_{t=0}^{\tau\wedge\tau_z-1}R(X_t,A_t)+G(X_{\tau\wedge\tau_z})I[\tau\wedge\tau_z<\infty]\right)\right].$$

Ya que $1=I[\tau=\infty]+I[\tau_z\le\tau<\infty]+I[\tau<\infty,\tau<\tau_z]$, las tres igualdades de los casos anteriores implican que:

$$E_x^{\pi}\left[U_\lambda\left(\sum_{t=0}^{\tau-1}R(X_t,A_t)+G(X_\tau)I[\tau<\infty]\right)\right]$$
$$=E_x^{\pi}\left[U_\lambda\left(\sum_{t=0}^{\tau\wedge\tau_z-1}R(X_t,A_t)+G(X_{\tau\wedge\tau_z})I[\tau\wedge\tau_z<\infty]\right)\right].$$

A través de (1.2) y (1.7), esta relación conduce a que

$$U_\lambda(V_\lambda(x;\pi,\tau))=U_\lambda(V_\lambda(x;\pi,\tau\wedge\tau_z)),$$

y (2.27) se deduce utilizando que $U_\lambda(\cdot)$ es estrictamente creciente. □

Lema 2.4. *Para todo $n\in\mathbb{N}$, $x\in S$ y $\tau\in\mathcal{T}$,*

$$U_\lambda(W_\lambda^*(x))$$
$$\le\sum_{k=0}^{n}E_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{k-1}R(X_t,A_t)+W_\lambda^*(X_k)\right)I[\tau=k]\right]$$
$$+E_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{n}R(X_t,A_t)+W_\lambda^*(X_{n+1})\right)I[\tau\ge n+1]\right]. \quad (2.29)$$

Demostración. Para empezar, obsérvese que (2.17) y (2.18) implican que, para cada estado x,

$$U_\lambda(W_\lambda^*(x))\le\sum_{y\in S}p_{x,y}(f^*(x))U_\lambda(R(x,f^*(x))+W_\lambda^*(y)), \quad (2.30)$$

una relación que mediante la propiedad de Markov implica que, para cada $x\in S$ y $n\in\mathbb{N}$,

$$U_\lambda(W_\lambda^*(X_n))\le E_x^{f^*}\left[U_\lambda(R(X_n,A_n)+W_\lambda^*(X_{n+1}))|\,\mathcal{F}_n\right]. \quad (2.31)$$

A continuación, la desigualdad en (2.29) se verificará mediante inducción. Sean $x\in S$ y $\tau\in\mathcal{T}$ arbitrarios. Combinando la convención (1.1) con las

relaciones $\tau \geq 0$ y $P_x^{f^*}[X_0 = x] = 1$, se deduce que:

$$\begin{aligned}
&U_\lambda(W_\lambda^*(x)) \\
&\quad = U_\lambda(W_\lambda^*(X_0))I[\tau = 0] + U_\lambda(W_\lambda^*(X_0))I[\tau \geq 1] \\
&\quad = U_\lambda\left(\sum_{t=0}^{0-1} R(X_t, A_t) + W_\lambda^*(X_0)\right) I[\tau = 0] + U_\lambda(W_\lambda^*(X_0))I[\tau \geq 1] \\
&\quad \leq U_\lambda\left(\sum_{t=0}^{0-1} R(X_t, A_t) + W_\lambda^*(X_0)\right) I[\tau = 0] \\
&\qquad\qquad + I[\tau \geq 1]E_x^{f^*}\left[U_\lambda(R(X_0, A_0) + W_\lambda^*(X_1))\middle|\, \mathcal{F}_0\right] \\
&\quad = U_\lambda\left(\sum_{t=0}^{0-1} R(X_t, A_t) + W_\lambda^*(X_0)\right) I[\tau = 0] \\
&\qquad\qquad + E_x^{f^*}\left[U_\lambda(R(X_0, A_0) + W_\lambda^*(X_1))I[\tau \geq 1]\middle|\, \mathcal{F}_0\right], \quad P_x^{f^*}\text{-c. s.}
\end{aligned}$$

donde la desigualdad se debe a (2.31) con $n = 0$, y la inclusión $[\tau \geq 1] \in \mathcal{F}_0$ se utilizó para establecer la última igualdad. Después de tomar esperanza con respecto a $P_x^{f^*}$, la desigualdad anterior muestra que se cumple (2.29) para el caso $n = 0$. Ahora supongamos que (2.29) es válida para $n \in \mathbb{N}$ y observemos que:

$$\begin{aligned}
&U_\lambda\left(\sum_{t=0}^{n} R(X_t, A_t) + W_\lambda^*(X_{n+1})\right) I[\tau \geq n+1] \\
&\qquad = U_\lambda\left(\sum_{t=0}^{n} R(X_t, A_t) + W_\lambda^*(X_{n+1})\right) I[\tau = n+1] \\
&\qquad\qquad + U_\lambda\left(\sum_{t=0}^{n} R(X_t, A_t) + W_\lambda^*(X_{n+1})\right) I[\tau \geq n+2]
\end{aligned}$$

mientras que, utilizando (1.3),

$$\begin{aligned}
&U_\lambda\left(\sum_{t=0}^{n} R(X_t, A_t) + W_\lambda^*(X_{n+1})\right) I[\tau \geq n+2] \\
&\quad = e^{\lambda\sum_{t=0}^{n} R(X_t, A_t)} I[\tau \geq n+2] U_\lambda\left(W_\lambda^*(X_{n+1})\right) \\
&\quad \leq e^{\lambda\sum_{t=0}^{n} R(X_t, A_t)} I[\tau \geq n+2] E_x^{f^*}\left[U_\lambda(R(X_{n+1}, A_{n+1}) + W_\lambda^*(X_{n+2}))\middle|\, \mathcal{F}_{n+1}\right] \\
&\quad = E_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{n+1} R(X_t, A_t) + W_\lambda^*(X_{n+2})\right) I[\tau \geq n+2]\middle|\, \mathcal{F}_{n+1}\right]
\end{aligned}$$

donde (2.31) con $n+1$ en lugar de n se utilizó para establecer la desigualdad, y la segunda igualdad se obtuvo combinando (1.3) con el hecho de que la variable aleatoria $e^{\lambda\sum_{t=0}^{n} R(X_t, A_t)} I[\tau \geq n+2]$ es $\mathcal{F}_{n+1}$-medible. Estos dos últimos hechos implican que:

$$\begin{aligned}
&E_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{n} R(X_t, A_t) + W_\lambda^*(X_{n+1})\right) I[\tau \geq n+1]\right] \\
&\qquad \leq E_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{n} R(X_t, A_t) + W_\lambda^*(X_{n+1})\right) I[\tau = n+1]\right] \\
&\qquad\qquad + E_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{n+1} R(X_t, A_t) + W_\lambda^*(X_{n+2})\right) I[\tau \geq n+2]\right].
\end{aligned}$$

Combinando esta relación con la hipótesis de inducción, se deduce que (2.29) se cumple con $n+1$ en lugar de n, completando el argumento de inducción. □

Lema 2.5. *Dado $x \in S$, sean $f \in \mathbb{F}$ y $\tau \in \mathcal{T}$ tales que*

$$P_x^f[\tau < \infty] = 1 \quad y \quad V_\lambda(x; f, \tau) < \infty.$$

En este caso

$$\lim_{n\to\infty} E_x^f \left[\left| U_\lambda \left(\sum_{k=0}^{n} R(X_t, A_t) \right) \right| I[\tau > n+1] \right] = 0. \tag{2.32}$$

Demostración. Como G es acotada, de (1.7) y (1.8) se deduce que la condición $V_\lambda(x; f, \tau) < \infty$ es equivalente a que:

$$E_x^f \left[e^{\lambda \sum_{k=0}^{\tau-1} R(X_t, A_t)} \right] \in (0, \infty), \tag{2.33}$$

por lo que $P_x^f[e^{\lambda \sum_{k=0}^{\tau-1} R(X_t,A_t)} < \infty] = 1$. Combinando este hecho con la condición $P_x^f[\tau < \infty] = 1$ resulta que:

$$(1 \vee e^{\lambda \sum_{k=0}^{\tau-1} R(X_t,A_t)}) I[\tau > n+1] \to 0 \quad \text{cuando } n \to \infty \quad P_x^f\text{-c. s.},$$

y entonces (2.33) y el teorema de convergencia dominada implican que:

$$E_x^f \left[(1 \vee e^{\lambda \sum_{k=0}^{\tau-1} R(X_t,A_t)}) I[\tau > n+1] \right] \to 0 \quad \text{cuando } n \to \infty;$$

Esta convergencia y la desigualdad $1 \vee e^{\lambda \sum_{k=0}^{\tau-1} R(X_t,A_t)} \geq e^{\lambda \sum_{k=0}^{n} R(X_t,A_t)}$ conducen a $\lim_{n\to\infty} E_x^f \left[e^{\lambda \sum_{k=0}^{n} R(X_t,A_t)} I[\tau > n+1] \right] = 0$, y la conclusión deseada (2.32) se sigue vía (1.2). □

Lema 2.6. *(i) Para todo $x \in S$, $\pi \in \mathcal{P}$ y $n = 1, 2, \ldots$*

$$\begin{aligned} E_x^\pi \left[U_\lambda \left(\sum_{t=0}^{n-1} R(X_t, A_t) + W_\lambda^*(X_n) \right) I[\tau^* > n] \right] \\ \geq E_x^\pi \left[U_\lambda \left(\sum_{t=0}^{n} R(X_t, A_t) + W_\lambda^*(X_{n+1}) \right) I[\tau^* = n+1] \right] \\ + E_x^\pi \left[U_\lambda \left(\sum_{t=0}^{n} R(X_t, A_t) + W_\lambda^*(X_{n+1}) \right) I[\tau^* > n+1] \right] \end{aligned}$$

(ii) Para todo $n \in \mathbb{N}$, $x \in S \setminus S^*$ y $\pi \in \mathcal{P}$,

$$\begin{aligned} U_\lambda(W_\lambda^*(x)) \geq \sum_{k=1}^{n} E_x^\pi \left[U_\lambda \left(\sum_{t=0}^{k-1} R(X_t, A_t) + W_\lambda^*(X_k) \right) I[\tau^* = k] \right] \\ + E_x^\pi \left[U_\lambda \left(\sum_{t=0}^{n-1} R(X_t, A_t) + W_\lambda^*(X_n) \right) I[\tau^* > n] \right]. \end{aligned} \tag{2.34}$$

Demostración. En primer lugar, observemos que $U_\lambda(W_\lambda^*(x)) < U_\lambda(G(x))$ cuando $x \notin S^*$, como se deduce de (2.15) y (2.20). Por lo tanto,

$$\begin{aligned} & U_\lambda(W_\lambda^*(x)) \\ & \quad = \sup_{a\in A(x)} \sum_{y\in S} p_{x,y}(a) U_\lambda\left(R(x,a) + W_\lambda^*(y)\right) \\ & \quad \geq \sum_{y\in S} p_{x,y}(a) U_\lambda\left(R(x,a) + W_\lambda^*(y)\right), \quad x \in S \setminus S^*, \quad a \in A(x). \end{aligned} \tag{2.35}$$

(i) Sea $\pi \in \mathcal{P}$ arbitrario y, usando que $X_t \notin S^*$ para $0 \leq t < \tau^*$, por (2.21), lo anterior y la propiedad de Markov dan como resultado que para cada $n \in \mathbb{N}$ la siguiente relación se mantiene casi seguramente con respecto a P_x^π:

$$\begin{aligned} U_\lambda(W_\lambda^*(X_n)) &\geq \sum_{y\in S} p_{X_n,y}(A_n) U_\lambda\left(R(X_n,A_n) + W_\lambda^*(y)\right) \\ &= E_x^\pi\left[U_\lambda\left(R(X_n,A_n) + W_\lambda^*(X_{n+1})\right) \middle| \mathcal{F}_n, A_n\right] \text{ en } [\tau^* > n]. \end{aligned}$$

Multiplicando ambos lados de esta desigualdad por $e^{\lambda\sum_{t=0}^{n-1} R(X_t,A_t)} I[\tau^* > n]$, la cual es una variable aleatoria $\mathcal{F}_n$-medible, una aplicación de (1.3) conduce a:

$$\begin{aligned} & U_\lambda\left(\sum_{t=0}^{n-1} R(X_t,A_t) + W_\lambda^*(X_n)\right) I[\tau^* > n] \\ & \qquad \geq E_x^\pi\left[U_\lambda\left(\sum_{t=0}^{n} R(X_t,A_t) + W_\lambda^*(X_{n+1})\right) I[\tau^* > n] \middle| \mathcal{F}_n, A_n\right]. \end{aligned}$$

A partir de este punto, la conclusión sigue tomando esperanza con respecto a P_x^π y utilizando la igualdad $I[\tau^* > n] = I[\tau^* = n+1] + I[\tau^* > n+1]$.

(ii) El argumento es por inducción sobre n. Sean $x \in S \setminus S^*$ y $\pi \in \mathcal{P}$ arbitrarios, y observemos que (2.35) conduce a $U_\lambda(W_\lambda^*(x)) \geq E_x^\pi\left[U_\lambda\left(R(X_0,A_0) + W_\lambda^*(X_1)\right)\right]$; ya que $P_x^\pi[\tau^* > 0] = 1$, por (2.21) se deduce que:

$$\begin{aligned} U_\lambda\left(W_\lambda^*(x)\right) \geq\ & E_x^\pi\left[U_\lambda\left(R(X_0,A_0) + W_\lambda^*(X_1)\right) I[\tau^* = 1]\right] \\ & + E_x^\pi\left[U_\lambda\left(R(X_0,A_0) + W_\lambda^*(X_1)\right) I[\tau^* > 1]\right], \end{aligned}$$

una expresión equivalente a (2.34) con $n = 1$. Supongamos ahora que (2.34) es válida para algún $n \in \mathbb{N}$. En este caso, los cálculos directos que combinan la parte (i) con la hipótesis de inducción muestran que (2.34) también se cumple con $n+1$ en lugar de n, completando el argumento de inducción. □

La prueba de la verificación de la existencia de un equilibrio de Nash es bastante técnica y para facilitar la presentación los pasos esenciales se han establecido por separado en los Teoremas 2.3 y 2.4 a continuación.

Teorema 2.3. *Para cada $\tau \in \mathcal{T}$,*

$$W_\lambda^*(\cdot) \leq V_\lambda(\cdot; f^*, \tau). \tag{2.36}$$

Demostración. Por el Lema 2.3, sin pérdida de generalidad τ puede ser sustituido por $\tau \wedge \tau_z$, y entonces el Supuesto 2.1 arroja que es suficiente establecer la conclusión bajo la condición de que τ es un tiempo de paro finito:

$$P_x^{f^*}[\tau < \infty] = 1, \quad x \in S. \tag{2.37}$$

Como (2.36) ciertamente se cumple si $V_\lambda(\cdot; f^*, \tau) = \infty$, en el siguiente argumento se supondrá que:

$$V_\lambda(\cdot; f^*, \tau) < \infty. \tag{2.38}$$

Obsérvese que (1.3) y la inclusión $W_\lambda^* \in [\![0, G]\!]$ dan como resultado que:

$$\begin{aligned}\left| U_\lambda\left(\sum_{t=0}^{n} R(X_t, A_t) + W_\lambda^*(X_{n+1})\right)\right| &= \left| e^{\lambda W_\lambda^*(X_{n+1})} U_\lambda\left(\sum_{t=0}^{n} R(X_t, A_t))\right)\right| \\ &\leq e^{|\lambda| \|G\|} \left| U_\lambda\left(\sum_{t=0}^{n} R(X_t, A_t))\right)\right|\end{aligned}$$

Nótese que, a través del Lema 2.5, (2.37) y (2.38) implican que:

$$\lim_{n\to\infty} E_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{n} R(X_t, A_t)\right) I[\tau > n+1]\right] = 0,$$

y combinando esta convergencia con la desigualdad anterior se deduce que:

$$E_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{n} R(X_t, A_t) + W_\lambda^*(X_{n+1})\right) I[\tau > n+1]\right] \to 0 \ \text{ cuando } n \to \infty.$$

Por otro lado, como $U_\lambda(\cdot)$ tiene signo constante, el teorema de convergencia monótona (Teorema A.4) arroja inmediatamente que:

$$\begin{aligned}&\lim_{n\to\infty} \sum_{k=0}^{n} E_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{k-1} R(X_t, A_t) + W_\lambda^*(X_k)\right) I[\tau = k]\right] \\ &\quad = \sum_{k=0}^{\infty} E_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{k-1} R(X_t, A_t) + W_\lambda^*(X_k)\right) I[\tau = k]\right] \\ &\quad = E_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{\tau-1} R(X_t, A_t) + W_\lambda^*(X_\tau)\right) I[\tau < \infty]\right] \\ &\quad \leq E_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{\tau-1} R(X_t, A_t) + G(X_\tau)\right) I[\tau < \infty]\right] \\ &\quad = U_\lambda(V_\lambda(x, f^*, \tau)),\end{aligned}$$

donde la desigualdad se debe a la inclusión $W_\lambda^* \in [\![0, G]\!]$ y a la monotonía de $U_\lambda(\cdot)$ y, utilizando (2.37), la última igualdad se debe a (1.2) y (1.7). Tomando límite cuando n tiende a ∞ en el lado derecho de (2.29), las dos expresiones anteriores dan como resultado que $U_\lambda(W_\lambda^*(x)) \leq U_\lambda(V_\lambda(x, f^*, \tau))$ y entonces (2.36) sigue usando que $U_\lambda(\cdot)$ es estrictamente creciente. □

Teorema 2.4. *Para todo $x \in S$*

$$V_\lambda(x;\pi,\tau^*) \leq W_\lambda^*(x), \quad \pi \in \mathcal{P}. \tag{2.39}$$

Demostración. En primer lugar, nótese que (2.22) y Lema 2.2(ii) implican que:

$$P_x^\pi[\tau^* < \infty] = 1, \quad x \in S. \tag{2.40}$$

Ahora, sea $\pi \in \mathcal{P}$ arbitrario y supongamos que $x \in S^*$, de modo que (2.20) y (2.21) dan como resultado que:

$$W_\lambda^*(x) = G(x) \quad \text{y} \quad P_x^\pi[\tau^* = 0] = 1,$$

mientras que (1.1) y (1.7) conducen a que $V_\lambda(x;\pi,\tau^*) = W_\lambda^*(x)$, y entonces (2.39) se cumple con igualdad. A continuación, se verificará la conclusión deseada cuando el estado inicial x no pertenece a S^*. Consideremos la siguiente afirmación:

$$\begin{aligned} &\text{Para todo } x \in S \setminus S^*, \text{ y } \pi \in \mathcal{P}, \\ &\liminf_{n\to\infty} E_x^\pi \left[U_\lambda \left(\sum_{t=0}^{n-1} R(X_t, A_t) + W_\lambda^*(X_n) \right) I[\tau^* > n] \right] \geq 0. \end{aligned} \tag{2.41}$$

Observando que $U_\lambda(\cdot) > 0$ cuando λ es positivo, queda claro que la afirmación anterior se cumple si $\lambda > 0$. Para completar la prueba de (2.41), supongamos que $\lambda < 0$ y observemos que (1.2) y la no negatividad de R y W_λ^* dan como resultado que $\left| U_\lambda \left(\sum_{t=0}^{n-1} R(X_t, A_t) + W_\lambda^*(X_n) \right) I[\tau^* > n] \right| \leq I[\tau^* > n]$, por (1.2), y mediante (2.40) se deduce que como $n \to \infty$,

$$E_x^\pi \left[\left| U_\lambda \left(\sum_{t=0}^{n-1} R(X_t, A_t) + W_\lambda^*(X_n) \right) I[\tau^* > n] \right| \right] \leq P_x^\pi[\tau^* > n] \to 0,$$

una convergencia que produce inmediatamente que (2.41) se cumple con igualdad cuando λ es negativo. A continuación, utilizando que la función $U_\lambda(\cdot)$ no tiene cambios de signo para λ fijo, el teorema de convergencia monótona da como resultado que:

$$\begin{aligned} &\lim_{n\to\infty} \sum_{k=1}^{n} E_x^\pi \left[U_\lambda \left(\sum_{t=0}^{k-1} R(X_t, A_t) + W_\lambda^*(X_k) \right) I[\tau^* = k] \right] \\ &= \sum_{k=1}^{\infty} E_x^\pi \left[U_\lambda \left(\sum_{t=0}^{k-1} R(X_t, A_t) + W_\lambda^*(X_k) \right) I[\tau^* = k] \right] \\ &= E_x^\pi \left[U_\lambda \left(\sum_{t=0}^{\tau^*-1} R(X_t, A_t) + W_\lambda^*(X_{\tau^*}) \right) I[\tau^* < \infty] \right] \\ &= U_\lambda(V_\lambda(x;\pi,\tau^*)), \end{aligned}$$

donde la última igualdad se deriva de la combinación de (1.7) y (2.40). Para concluir, tomamos límite inferior cuando n tiende a ∞ en el lado derecho de (2.34) para obtener, a través de la igualdad anterior y (2.41), que

$$U_\lambda(W_\lambda^*(x)) \geq U_\lambda(V_\lambda(x;\pi,\tau^*)),$$

una desigualdad que usando que U_λ estrictamente creciente conduce a que $W_\lambda^*(x) \geq V_\lambda(x; \pi, \tau^*)$, mostrando que (2.39) también es válida para $x \in S \setminus S^*$. □

Por último, se utilizarán los dos teoremas anteriores para establecer la existencia de un equilibrio de Nash.

Teorema 2.5. *Bajo los Supuestos 1.1 y 2.1, se cumplen las siguientes afirmaciones:*

(i) Para todo $x \in S$,

$$V_\lambda(x; f^*, \tau^*) = W_\lambda^*(x).$$

(ii) El par $(f^*, \tau^*) \in \mathbb{F} \times \mathcal{T}$ constituye un equilibrio de Nash.

Demostración. Por los Teoremas 2.3 y 2.4 se tiene que:

$$V_\lambda(\cdot; \pi, \tau^*) \leq W_\lambda^*(\cdot) \leq V_\lambda(\cdot; f^*, \tau), \quad (\pi, \tau) \in \mathcal{P} \times \mathcal{T}.$$

Fijando $(\pi, \tau) = (f^*, \tau^*)$ se deduce que $W_\lambda^*(\cdot) = V_\lambda(\cdot; f^*, \tau^*)$, estableciendo la parte (i), y combinando este hecho con las desigualdades anteriores se deduce de la Definición 1.1 que (f^*, τ^*) es un equilibrio de Nash, completando la prueba. □

Un hecho importante en el Teorema anterior es que dado que la función de valor del juego $V_\lambda(\cdot; f^*, \tau^*)$ es igual a $W_\lambda^*(x)$, se tiene inmediatamente que el operador T_λ tiene un punto fijo único.

En esta sección se presentó el resultado principal considerando el supuesto de la existencia un estado absorbente. Para una mejor presentación de la prueba, se utilizaron resultados auxiliares relacionados con los espacios de estrategias de los jugadores y algunas desigualdades importantes. Además, los teoremas de convergencia como el teorema de convergencia dominada y el teorema de convergencia monótona, junto con el teorema del paro opcional, fueron herramientas claves que nos permitieron concluir nuestra investigación de manera exitosa. La función valor del juego fue caracterizada mediante una ecuación de equilibrio y se determinó un equilibrio de Nash para el juego. En el próximo capítulo, consideramos el modelo comunicante con el objetivo de analizar un supuesto más general, ya que en muchos casos no siempre es posible garantizar la existencia de un estado absorbente.

Capítulo 3

Modelo Comunicante

En este capítulo, consideramos el modelo del juego reemplazando el Supuesto 2.1 por el supuesto de comunicación (Supuesto 3.1). Se demuestra que el juego tiene solución, es decir, existe un equilibrio de Nash para el juego. Además, se proporciona un ejemplo especifico de un juego en donde se cumplen nuestros supuestos, y se trabaja con este ejemplo de manera numérica. Los resultados presentados han sido publicados en el artículo [26].

Estudiar la existencia de equilibrios de Nash para juegos Markovianos con tiempos de paro en modelos más generales que no satisfacen el Supuesto 2.1, es un problema interesante, ya que, al considerar un supuesto más general, se puede determinar si los resultados obtenidos son específicos del caso absorbente o si son aplicables a una variedad más amplia de situaciones. Además, en la práctica, puede que en las aplicaciones no se tengan las condiciones que garanticen la existencia de un estado absorbente. Es por ello que el análisis posterior viene determinado por el siguiente requisito: Si el jugador II no detiene el juego, la cadena de Markov inducida por cualquier política estacionaria del jugador I es comunicante, lo cual es formalizado en el siguiente supuesto.

Supuesto 3.1. (i) *Para cada $f \in \mathbb{F}$, la cadena de Markov inducida por f es comunicante, esto es, dados cualesquiera $x, y \in S$, existe un entero positivo n $(n = n(x, y, f))$ y estados $x_1, x_2 \ldots, x_{n-1} \in S$ tales que:*

$$x_0 = x, x_n = y \;\; y \;\; p_{x_{i-1},x_i}(f(x_{i-1})) > 0, \;\; i = 1, 2, \ldots, n.$$

(ii) *Para cada $f \in \mathbb{F}$, existe una distribución de probabilidad $\rho_f(\cdot)$ en S tal que:*

$$\rho_f(y) = \sum_{x \in S} \rho_f(x) p_{x,y}(f(x)), \;\; y \in S.$$

(iii) *Existe un estado z_0 tal que $R(z_0, a) > 0$ para toda $a \in A(z_0)$.*

El Supuesto 3.1(i) es bien conocido en la literatura sobre PDMs sensibles al riesgo (ver, por ejemplo, [11] y [32]). En particular, en [32] se emplea para garantizar la unicidad de las soluciones de la ecuación de optimalidad asociada con el criterio de costo promedio, siempre que la ecuación admita una solución

acotada. En este trabajo, se aplica en las demostraciones de los Teoremas 3.1 y 3.2 para garantizar la finitud de los tiempos de paro.

El supuesto 3.1(ii) requiere una distribución invariante. Esta condición se ha utilizado en este tipo de juegos para el caso neutral (ver [11]) para demostrar la existencia de un equilibrio de Nash. De manera similar, en este documento se adopta esta misma condición con el mismo objetivo. La existencia de una distribución invariante de este tipo se puede garantizar mediante el teorema de Perron-Frobenius en el caso finito o aplicando una condición de Harris en el caso numerable (véase [21]). Además, observe que la propiedad de comunicación del inciso (i) implica que la distribución invariante ρ_f de la cadena de Markov inducida por cualquier $f \in \mathbb{F}$ satisface que:

$$\rho_f(x) > 0, \quad x \in S. \tag{3.1}$$

El Supuesto 3.1(iii) desempeña un papel fundamental en la obtención de los dos resultados principales de este trabajo, como lo son la unicidad del punto fijo del operador de equilibrio y la existencia de un equilibrio de Nash para el juego. En el ejemplo siguiente, verificamos que los Supuestos 1.1 y 3.1 se cumplen para un juego $\mathcal{G}$ en especifico.

Ejemplo 3.1. *Sea N un entero positivo fijo y consideremos un juego $\mathcal{G}$ con las componentes siguientes:*

- *Espacio de estados $S = \mathbb{N}$.*
- *Espacio de acciones $A = \{b_1, b_2, \ldots, b_N\}$, donde $0 < b_1 < b_2 < \cdots < b_N < 1$ y $b_N + b_1 < 1$.*
- *$A(x) = A$, para todo $x \in S$.*
- *Las funciones de recompensa inmediata y recompensa terminal vienen dadas por*
  $$R(x,a) = \begin{cases} 0 & si\ x \geq N \\ \dfrac{1}{ax+1} & si\ x < N \end{cases}, \text{ para todo } (x,a) \in \mathbb{K} \text{ y } G(x) = \frac{N}{x+1},$$
  para todo $x \in S$.
- *La ley de transición controlada es descrita como sigue:*
  $$p_{0,1}(a) = 1,$$
  $$p_{x,x+1}(a) = a,$$
  $$p_{x,x-1}(a) = 1 - a,$$
  para todo $x \neq 0$ y $a \in A$.

Para este juego, se observa que el Supuesto 1.1 se cumple. Por otro lado, la función de transición de una cadena de nacimiento y muerte en los enteros no negativos es de la forma:

$$P(x,y) = \begin{cases} q_x & si\ y = x-1 \\ r_x & si\ y = x \\ p_x & si\ y = x+1, \end{cases}$$

donde $q_0 = 0$, $p_x + q_x + r_x = 1$, *para* $x \in \mathbb{N}$ *y la cadena es recurrente positiva si*

$$\sum_{x=1}^{\infty} \frac{p_0 \cdots p_{x-1}}{q_1 \cdots q_x} < \infty. \tag{3.2}$$

Basado en la ley de transición controlada del juego, se tiene que cada $f \in \mathbb{F}$ *induce una cadena de nacimiento y muerte irreducible. Esto es debido al hecho de que* $b_i > 0$ *para todo* $i \in \{1, 2, \ldots, N\}$, *así se cumple el inciso* (i) *del Supuesto 3.1. Además, la cadena inducida también será recurrente positiva ya que la expresión equivalente a (3.2) para este juego es la siguiente*

$$\begin{aligned} \sum_{x=1}^{\infty} \frac{f(1) \cdots f(x-1)}{(1-f(1)) \cdots (1-f(x))} &\leq \frac{1}{b_N} \sum_{x=1}^{\infty} \left(\frac{b_N}{1-b_1} \right)^x \\ &< \infty, \end{aligned}$$

donde la desigualdad se debe a que $b_1 \leq f(x) \leq b_N$, $\forall x \in S$, *y la convergencia es debida a que* $b_N + b_1 < 1$. *Por lo tanto, el inciso* (ii) *del Supuesto 3.1 también se cumple porque una cadena de Markov irreducible y recurrente positiva tiene una única distribución estacionaria.*

3.1. Unicidad del punto fijo

En la Sección 2.2 se demostró que bajo el Supuesto 1.1 se tiene que el operador T_λ tiene puntos fijos. En esta sección se prueba la unicidad del punto fijo del operador T_λ, cuya prueba no depende de la existencia del equilibrio de Nash. Esta es una de las primeras diferencias que se tienen es esta sección con respecto a la anterior.

El argumento de la prueba de unicidad se basa en los dos lemas auxiliares presentados a continuación, los cuales que se apoyan en gran medida en el Supuesto 3.1.

Lema 3.1. *(i) Los siguientes límites existen:*

$$W_{\lambda_0} := \lim_{n \to \infty} T_\lambda^n[0], \quad W_{\lambda_1} := \lim_{n \to \infty} T_\lambda^n[G]. \tag{3.3}$$

Además,

(ii) $W_{\lambda_0} = T_\lambda [W_{\lambda_0}]$ *y* $W_{\lambda_1} = T_\lambda [W_{\lambda_1}]$.

(iii) Si $W_\lambda^* \in [\![0, G]\!]$ *es un punto fijo del operador* T_λ, *es decir,* $W_\lambda^* = T_\lambda[W_\lambda^*]$, *entonces*

$$W_{\lambda_0} \leq W_\lambda^* \leq W_{\lambda_1}. \tag{3.4}$$

Demostración. (i) Definimos $W_{0,\lambda_0} := 0, W_{0,\lambda_1} := G$ y $W_{n,\lambda_0} := T_\lambda^n[0], W_{n,\lambda_1} := T_\lambda^n[G]$ para $n \in \mathbb{N} \setminus \{0\}$ y observemos que:

$$W_{n+1,\lambda_0} := T_\lambda[W_{n,\lambda_0}], \quad W_{n+1,\lambda_1} := T_\lambda[W_{n,\lambda_1}], \quad n \in \mathbb{N}. \tag{3.5}$$

Como $W_{0,\lambda_0} = 0, W_{1,\lambda_0} = T_\lambda[0] \in [\![0, G]\!]$ y $W_{0,\lambda_1} = G, W_{1,\lambda_1} = T_\lambda[G] \in [\![0, G]\!]$ se deduce que $W_{0,\lambda_0} \leq W_{1,\lambda_0}$ y $W_{1,\lambda_1} \leq W_{0,\lambda_1}$, entonces combinando esto con un argumento de inducción y la propiedad (2.7) produce inmediatamente que:

$$0 \leq W_{n,\lambda_0} \leq W_{n+1,\lambda_0} \leq G \quad \text{y} \quad 0 \leq W_{n+1,\lambda_1} \leq W_{n,\lambda_1} \leq G, \quad n \in \mathbb{N},$$

donde las desigualdades extremas se deben a que las funciones W_{n,λ_0} y W_{n,λ_1} pertenecen a $[\![0, G]\!]$ para todo $n \in \mathbb{N}$. De ello se deduce que las sucesiones $\{W_{n,\lambda_0}(y)\}_{n\in\mathbb{N}}$ y $\{W_{n,\lambda_1}(y)\}_{n\in\mathbb{N}}$ son monótonas y acotadas, de modo que:

$$\lim_{n\to\infty} T_\lambda^n[0](y) := W_{\lambda_0}(y) \quad y \quad \lim_{n\to\infty} T_\lambda^n[G](y) := W_{\lambda_1}(y)$$

existen para todo $y \in S$.

(ii) El Teorema 2.1 permite afirmar que:

$$W_{\lambda_0}, W_{\lambda_1} \in [\![0, G]\!] \tag{3.6}$$

y además

$$\lim_{n\to\infty} T_\lambda[W_{n,\lambda_0}](x) = T_\lambda[W_{\lambda_0}](x) \quad y \quad \lim_{n\to\infty} T_\lambda[W_{n,\lambda_1}](x) = T_\lambda[W_{\lambda_1}](x),$$

para todo $x \in S$. Así, tomando límite cuando n tiende a ∞ en ambos lados de las igualdades en (3.5), junto con lo anterior se tiene que $W_{\lambda_0} = T_\lambda[W_{\lambda_0}]$ y $W_{\lambda_1} = T_\lambda[W_{\lambda_1}]$, mostrando que W_{λ_0} y W_{λ_1} son puntos fijos del operador T_λ.

(iii) Sea $W_\lambda^* \in [\![0, G]\!]$ tal que $W_\lambda^* = T_\lambda[W_\lambda^*]$ y observe que:

$$W_\lambda^* = T_\lambda^n[W_\lambda^*], \quad n \in \mathbb{N}.$$

Combinando las desigualdades $0 \leq W_\lambda^* \leq G$ con la propiedad (2.7) del operador T_λ, resulta que $T_\lambda^n[0] \leq T_\lambda^n[W_\lambda^*] \leq T_\lambda^n[G]$ para todo $n \in \mathbb{N}$, una relación que a través de lo observado anteriormente y (3.3) conduce a $W_{\lambda_0} \leq W_\lambda^* \leq W_{\lambda_1}$. □

Observación 3.1. *Notemos que si existe $\hat{x} \in S$ tal que $W_{\lambda_0}(\hat{x}) = G(\hat{x})$, entonces por (3.6) y (3.4) se tiene que $W_{\lambda_0}(\hat{x}) = W_{\lambda_1}(\hat{x}) = G(\hat{x})$.*

En el Lema 3.1 hemos caracterizado a los puntos fijos del operador T_λ a través de (3.4). Con lo que para probar la unicidad bastaría con mostrar que $W_{\lambda_0} \geq W_{\lambda_1}$. Por otro lado, bajo el supuesto absorbente teníamos que $S^* \neq \emptyset$, ya que $z \in S^*$. Al eliminar este supuesto no podemos garantizar que esta condición se cumpla. Sin embargo, en el siguiente resultado se demuestra que con el supuesto de comunicación también se tiene S^* es distinto de vacío.

Lema 3.2. *Bajo los Supuestos 1.1 y 3.1, se tiene que $S^* \neq \emptyset$.*

Demostración. El argumento de la prueba es por contradicción. Así, supongamos que $S^* = \emptyset$, de modo que $G(x) \neq W_\lambda^*(x) = T_\lambda[W_\lambda^*](x)$ para todo $x \in S$. En este caso la igualdad

$$\begin{aligned} U_\lambda(W_\lambda^*(x)) &= \sup_{a\in A(x)} \left[\sum_{y\in S} p_{x,y}(a) U_\lambda(R(x,a) + W_\lambda^*(y)) \right] \\ &= \sum_{y\in S} p_{x,y}(f^*(x)) U_\lambda(R(x, f^*(x)) + W_\lambda^*(y)), \end{aligned}$$

es siempre válida por (2.5) y (2.18). Esta última expresión la podemos reescribir como:

$$U_\lambda(W^*_\lambda(x) - R(x, f^*(x))) = \sum_{y\in S} p_{x,y}(f^*(x))U_\lambda(W^*_\lambda(y)), \tag{3.7}$$

utilizando (1.3). Luego, como la función $U_\lambda(\cdot)$ es estrictamente creciente y R no negativa, se tiene que:

$$U_\lambda(W^*_\lambda(x)) \geq U_\lambda(W^*_\lambda(x) - R(x, f^*(x))),$$

lo que nos conduce a la siguiente desigualdad:

$$U_\lambda(W^*_\lambda(x)) \geq \sum_{y\in S} p_{x,y}(f^*(x))U_\lambda(W^*_\lambda(y)).$$

De la desigualdad anterior obtenemos la siguiente expresión:

$$U_\lambda(W^*_\lambda(x)) + \delta(x) = \sum_{y\in S} p_{x,y}(f^*(x))U_\lambda(W^*_\lambda(y)), \tag{3.8}$$

donde

$$\delta(x) := \sum_{y\in S} p_{x,y}(f^*(x))U_\lambda(W^*_\lambda(y)) - U_\lambda(W^*_\lambda(x)) \leq 0,\ x \in S.$$

El Supuesto 3.1 (ii) nos garantiza la existencia de $\rho_{f^*}(\cdot)$, la distribución invariante de la cadena de Markov inducida por f^* y se sigue que:

$$\begin{aligned}
\sum_{x\in S} \rho_{f^*}(x)\left[U_\lambda(W^*_\lambda(x)) + \delta(x)\right] &= \sum_{x\in S} \rho_{f^*}(x)\left[\sum_{y\in S} p_{x,y}(f^*(x))U_\lambda(W^*_\lambda(y))\right] \\
&= \sum_{y\in S}\left[\sum_{x\in S} \rho_{f^*}(x)p_{x,y}(f^*(x))\right] U_\lambda(W^*_\lambda(y)) \\
&= \sum_{y\in S} \rho_{f^*}(y)U_\lambda(W^*_\lambda(y)),
\end{aligned}$$

de donde se obtiene que:

$$\sum_{x\in S} \rho_{f^*}(x)\delta(x) = 0\,.$$

Como $\delta(\cdot) \leq 0$, esta última igualdad y (3.1) dan como resultado que $\delta(\cdot) = 0$, por lo que (3.7) y (3.8) implican que:

$$U_\lambda(W^*_\lambda(x) - R(x, f^*(x))) = U_\lambda(W^*_\lambda(x)).$$

Usando que $U_\lambda(\cdot)$ es estrictamente creciente se tiene que $R(x, f^*(x)) = 0$, para todo $x \in S$, en contradicción con el Supuesto 3.1(iii). □

Una vez demostrados los dos lemas anteriores, ahora presentamos el resultado de unicidad de W^*_λ.

Teorema 3.1. *Bajo los Supuestos 1.1 y 3.1, se tiene que existe un único punto fijo del operador T_λ, esto es, existe una única función $W_\lambda^* \in [\![0, G]\!]$ que satisface que:*

$$W_\lambda^* = T_\lambda[W_\lambda^*]\,. \tag{3.9}$$

Demostración. Sean W_{λ_0} y W_{λ_1} los puntos fijos del operador T_λ definidos en (3.3). Por demostrar tenemos que

$$W_{\lambda_0} \geq W_{\lambda_1}. \tag{3.10}$$

Sea $x \in S$, entonces se tiene que:

$$\begin{aligned}
&U_\lambda(W_{\lambda_0}(x)) \\
&= U_\lambda(T_\lambda\left[W_{\lambda_0}\right](x)) \\
&= \text{mín}\left\{U_\lambda(G(x)),\ \sup_{a\in A(x)}\left[\sum_{y\in S} p_{x,y}(a)U_\lambda(R(x,a)+W_{\lambda_0}(y))\right]\right\} \\
&\leq \text{mín}\left\{U_\lambda(G(x)),\ \sup_{a\in A(x)}\left[\sum_{y\in S} p_{x,y}(a)U_\lambda(R(x,a)+W_{\lambda_1}(y))\right]\right\} \\
&+ \sup_{a\in A(x)}\left[e^{\lambda R(x,a)}\sum_{y\in S} p_{x,y}(a)|U_\lambda(W_{\lambda_0}(y)) - U_\lambda(W_{\lambda_1}(y))|\right] \\
&= U_\lambda(T_\lambda\left[W_{\lambda_1}\right](x)) + \sup_{a\in A(x)}\left[e^{\lambda R(x,a)}\sum_{y\in S} p_{x,y}(a)|U_\lambda(W_{\lambda_0}(y)) - U_\lambda(W_{\lambda_1}(y))|\right] \\
&\leq U_\lambda(W_{\lambda_1}(x)) + e^{|\lambda|\|R\|}\sup_{a\in A(x)}\left[\sum_{y\in S} p_{x,y}(a)|U_\lambda(W_{\lambda_0}(y)) - U_\lambda(W_{\lambda_1}(y))|\right].
\end{aligned}$$

Como $U_\lambda(W_{\lambda_0}) - U_\lambda(W_{\lambda_1})$ está acotada, del Supuesto 1,1 se deduce que existe $\tilde{f} \in \mathbb{F}$ tal que:

$$\begin{aligned}
&\sum_{y\in S} p_{x,y}(\tilde{f}(x))|U_\lambda(W_{\lambda_0}(y)) - U_\lambda(W_{\lambda_1}(y))| \\
&= \sup_{a\in A(x)}\left[\sum_{y\in S} p_{x,y}(a)|U_\lambda(W_{\lambda_0}(y)) - U_\lambda(W_{\lambda_1}(y))|\right], \quad x \in S,
\end{aligned}$$

de modo que:

$$U_\lambda(W_{\lambda_0}(x)) - U_\lambda(W_{\lambda_1}(x)) \leq e^{|\lambda|\|R\|}\sum_{y\in S} p_{x,y}(\tilde{f}(x))|U_\lambda(W_{\lambda_0}(y)) - U_\lambda(W_{\lambda_1}(y))|.$$

Mientras que la desigualdad

$$U_\lambda(W_{\lambda_1}(x)) - U_\lambda(W_{\lambda_0}(x)) \leq e^{|\lambda|\|R\|}\sum_{y\in S} p_{x,y}(\tilde{f}(x))|U_\lambda(W_{\lambda_0}(y)) - U_\lambda(W_{\lambda_1}(y))|,$$

se obtiene intercambiando los roles de W_{λ_0} y W_{λ_1}, por lo tanto, concluimos que:

$$|U_\lambda(W_{\lambda_0}(x)) - U_\lambda(W_{\lambda_1}(x))| \leq e^{|\lambda|\|R\|}\sum_{y\in S} p_{x,y}(\tilde{f}(x))|U_\lambda(W_{\lambda_0}(y)) - U_\lambda(W_{\lambda_1}(y))|.$$

Luego, como $W_{\lambda_0} \leq W_{\lambda_1}$, se tiene que:

$$\begin{aligned} U_\lambda(W_{\lambda_0}(x)) - U_\lambda(W_{\lambda_1}(x)) &\geq e^{|\lambda|\|R\|} \sum_{y\in S} p_{x,y}(\tilde{f}(x)) U_\lambda(W_{\lambda_0}(y)) - U_\lambda(W_{\lambda_1}(y)) \\ &\geq \sum_{y\in S} p_{x,y}(\tilde{f}(x)) U_\lambda(W_{\lambda_0}(y)) - U_\lambda(W_{\lambda_1}(y)). \end{aligned}$$

Esta relación y la propiedad de Markov implican que para todo $x \in S$ y $n \in \mathbb{N}$,

$$\begin{aligned} U_\lambda(W_{\lambda_0}(X_n)) - U_\lambda(W_{\lambda_1}(X_n)) &\geq \sum_{y\in S} p_{X_n,y}(\tilde{f}(X_n)) U_\lambda(W_{\lambda_0}(y)) - U_\lambda(W_{\lambda_1}(y)) \\ &= E_x^{\tilde{f}} \left[U_\lambda(W_{\lambda_0}(X_{n+1})) - U_\lambda(W_{\lambda_1}(X_{n+1})) | \mathcal{F}_n \right], \end{aligned}$$

así se tiene que $\{U_\lambda(W_{\lambda_0}(X_n)) - U_\lambda(W_{\lambda_1}(X_n)), \mathcal{F}_n\}$ es una supermartingala con respecto a $P_x^{\tilde{f}}$.

Sea τ_0 el tiempo de alcance al conjunto $S^*_{\lambda_0} = \{x \in S \mid W_{\lambda_0} = G(x)\}$, es decir

$$\tau_0 = \text{mín}\{n \in \mathbb{N} \mid X_n \in S^*_{\lambda_0}\},$$

de modo que τ_0 es un tiempo de paro con respecto a la filtración $\{\mathcal{F}_t\}$ definida en (1.5), esto es, $[\tau_0 = k] \in \mathcal{F}_k$ para todo $k \in \mathbb{N}$. Por otro lado, se tiene que

$$P_x^{\tilde{f}}[\tau_0 < \infty] = 1,$$

por el Supuesto 3.1. Luego utilizando que la función $U_\lambda(W_{\lambda_0}(\cdot)) - U_\lambda(W_{\lambda_1}(\cdot))$ está acotada, el teorema del muestreo opcional conduce a que:

$$U_\lambda(W_{\lambda_0}(x)) - U_\lambda(W_{\lambda_1}(x)) \geq E_x^{\tilde{f}} \left[U_\lambda(W_{\lambda_0}(X_{\tau_0})) - U_\lambda(W_{\lambda_1}(X_{\tau_0})) \right], \, x \in S.$$

Dado que $X_{\tau_0} \in S^*_{\lambda_0}$ en el evento $[\tau_0 < \infty]$, por la Observación 3.1 se tiene que:

$$U_\lambda(W_{\lambda_0}(x)) - U_\lambda(W_{\lambda_1}(x)) \geq 0 \, , \, x \in S.$$

Por lo que (3.10) se obtiene usando que $U_\lambda(\cdot)$ es estrictamente creciente. □

3.2. Equilibrio de Nash

Ahora bajo los Supuestos 1.1 y 3.1, se demostrará que existe un equilibrio de Nash con respecto al índice de recompensa total sensible al riesgo (1.7). En primer lugar, tenemos que el Lema 3.2 garantiza que:

$$S^* \neq \emptyset,$$

donde S^* es el conjunto definido en (2.20).

Por otro lado, recordando que la cadena de Markov asociada a cualquier $f \in \mathbb{F}$ es comunicante y tiene una distribución invariante, por el Supuesto 3.1,

se tiene que el conjunto S^* es accesible desde cualquier estado inicial bajo cada política estacionaria, es decir:

$$P_x^f[\tau^* < \infty] = 1, \quad x \in S, \quad f \in \mathbb{F}, \tag{3.11}$$

además de que:

$$V_\lambda(x, f, \tau^*) < \infty, \quad x \in S, \quad f \in \mathbb{F}. \tag{3.12}$$

La propiedad (3.11) al igual como se hizo en el Lema 2.2 se puede extender a la clase de todas las políticas del jugador I. Esto implica que se cumple la siguiente propiedad:

$$P_x^\pi[\tau^* < \infty] = 1, \quad x \in S, \quad \pi \in \mathcal{P}. \tag{3.13}$$

Para la prueba de la existencia de un equilibrio de Nash necesitamos probar bajo los Supuestos 1.1 y 3.1 las desigualdades presentadas en los Teoremas 2.3 y 2.4. La desigualdad (2.39) del Teorema 2.4 es válida en este caso debido a (3.13). Para probar la desigualdad (2.36), tomamos $x \in S$ arbitrario y consideremos las dos siguientes posibilidades para el par (x, τ):

(i) $P_x^{f^*}[\tau < \infty] = 1$, $x \in S$.

Para este caso la conclusión ya fue probada previamente.

(ii) $P_x^{f^*}[\tau = \infty] > 0$, $x \in S$.

Sea el estado z_0 como en el Supuesto 3.1(iii) y observemos que la propiedad de comunicación da como resultado que:

$$P_x^{f^*}\left[X_n = z_0 \ \ i.o.\right] = 1,$$

donde $i.o.$ significa infinitamente a menudo. Ahora, dado que R es no negativa y que $R\left(z_0, f^*(z_0)\right) > 0$, se deduce que:

$$P_x^{f^*}\left[\sum_{n=0}^{\infty} R\left(X_n, A_n\right) = \infty\right] = 1,$$

y entonces, como el evento $[\tau = \infty]$ tiene probabilidad positiva, se tiene que:

$$\begin{aligned} V_\lambda\left(x; f^*, \tau\right) &= \frac{1}{\lambda} \log\left(E_x^{f^*}\left[e^{\lambda\left(\sum_{t=0}^{\tau-1} R(X_t, A_t) + G(X_\tau) I[\tau < \infty]\right)}\right]\right) \\ &= \frac{1}{\lambda} \log\left(E_x^{f^*}\left[e^{\lambda\left(\left(\sum_{t=0}^{\tau-1} R(X_t, A_t) + G(X_\tau)\right) I[\tau < \infty] + \sum_{t=0}^{\infty} R(X_t, A_t) I[\tau = \infty]\right)}\right]\right) \\ &\geqslant \frac{1}{\lambda} \log\left(E_x^{f^*}\left[e^{\lambda \sum_{t=0}^{\infty} R(X_t, A_t) I[\tau = \infty]}\right]\right) \\ &= \infty, \end{aligned}$$

por lo que la desigualdad en (2.36) también se cumple en este caso.

Se enuncia a continuación el resultado correspondiente al caso comunicante y la prueba de la igualdad de la función valor del juego con el punto fijo es independiente de la existencia de un equilibrio de Nash. Otra característica que distingue los resultados de esta sección con respecto a la anterior.

Teorema 3.2. *Bajo los Supuestos 1.1 y 3.1, se cumplen las siguientes afirmaciones:*

(i) Para todo $x \in S$,

$$V_\lambda(x; f^*, \tau^*) = W_\lambda^*(x).$$

(ii) El par $(f^*, \tau^*) \in \mathbb{F} \times \mathcal{T}$ *constituye un equilibrio de Nash.*

Demostración. (i) Sea $x \in S^*$, entonces (3.11), (2.20) y (2.21) dan como resultado que:

$$W_\lambda^*(x) = G(x) \quad \text{y} \quad P_x^{f^*}[\tau^* = 0] = 1,$$

mientras que (1.1) y (1.7) conducen a que $V_\lambda(x; f^*, \tau^*) = W_\lambda^*(x)$. Ahora se probará que la siguiente igualdad se cumple para todo $n \in \mathbb{N}\backslash\{0\}$ y $x \in S\backslash\{S^*\}$:

$$\begin{aligned} U_\lambda(W_\lambda^*(x)) = \sum_{k=1}^{n} E_x^{f^*} \left[U_\lambda \left(\sum_{t=0}^{k-1} R(X_t, A_t) + W_\lambda^*(X_k) \right) I[\tau^* = k] \right] \\ + E_x^{f^*} \left[U_\lambda \left(\sum_{t=0}^{n-1} R(X_t, A_t) + W_\lambda^*(X_n) \right) I[\tau^* > n] \right]. \end{aligned} \tag{3.14}$$

El argumento es por inducción sobre n. En primer lugar, observemos que $U_\lambda(W_\lambda^*(x)) < U_\lambda(G(x))$ cuando $x \notin S^*$, por (3.9) y (2.20), y entonces se cumple que:

$$\begin{aligned} U_\lambda(W_\lambda^*(x)) &= \sup_{a \in A(x)} \sum_{y \in S} p_{x,y}(a) U_\lambda \left(R(x,a) + W_\lambda^*(y) \right) \\ &= \sum_{y \in S} p_{x,y}(f^*(x)) U_\lambda \left(R(x, f^*(x)) + W_\lambda^*(y) \right) \\ &= E_x^{f^*}[U_\lambda \left(R(X_0, A_0) + W_\lambda^*(X_1) \right)], \quad x \in S \setminus S^*. \end{aligned} \tag{3.15}$$

Como $P_x^{f^*}[\tau^* > 0] = 1$, por (2.21), se sigue que:

$$\begin{aligned} U_\lambda(W_\lambda^*(x)) = E_x^{f^*} \left[U_\lambda \left(R(X_0, A_0) + W_\lambda^*(X_1) \right) I[\tau^* = 1] \right] \\ + E_x^{f^*} \left[U_\lambda \left(R(X_0, A_0) + W_\lambda^*(X_1) \right) I[\tau^* > 1] \right]. \end{aligned}$$

una expresión que es equivalente a (3.14) con $n = 1$. Por otro lado, usando que $X_t \notin S^*$ para $0 \leq t < \tau^*$, por (2.21), la igualdad en (3.15) y la propiedad de Markov dan como resultado que para cada $n \in \mathbb{N}$ la siguiente relación se cumple casi seguramente con respecto a P_x^π:

$$U_\lambda(W_\lambda^*(X_n)) = E_x^{f^*} \left[U_\lambda \left(R(X_n, A_n) + W_\lambda^*(X_{n+1}) \right) \middle| \mathcal{F}_n, A_n \right] \text{ en } [\tau^* > n].$$

Multiplicando ambos lados de esta desigualdad por $e^{\lambda \sum_{t=0}^{n-1} R(X_t, A_t)} I[\tau^* > n]$, la cual es una variable aleatoria $\mathcal{F}_n$-medible, una aplicación de (1.3) nos lleva a que:

$$\begin{aligned} U_\lambda \left(\sum_{t=0}^{n-1} R(X_t, A_t) + W_\lambda^*(X_n) \right) I[\tau^* > n] \\ = E_x^{f^*} \left[U_\lambda \left(\sum_{t=0}^{n} R(X_t, A_t) + W_\lambda^*(X_{n+1}) \right) I[\tau^* > n] \middle| \mathcal{F}_n, A_n \right]. \end{aligned}$$

Ahora tomando esperanza con respecto a $P_x^{f^*}$ y utilizando la igualdad $I[\tau^* > n] = I[\tau^* = n+1] + I[\tau^* > n+1]$ se tiene que:

$$\begin{aligned} &E_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{n-1} R(X_t, A_t) + W_\lambda^*(X_n)\right) I[\tau^* > n]\right] \\ &\qquad = E_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{n} R(X_t, A_t) + W_\lambda^*(X_{n+1})\right) I[\tau^* = n+1]\right] \\ &\qquad\quad + E_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{n} R(X_t, A_t) + W_\lambda^*(X_{n+1})\right) I[\tau^* > n+1]\right]. \end{aligned}$$

Luego, combinando esta igualdad con la hipótesis de inducción, se sigue que (3.14) es válida con $n+1$ en lugar de n. Además, usando que $U_\lambda(\cdot)$ tiene signo constante, el teorema de convergencia monótona da como resultado que:

$$\begin{aligned} &\lim_{n\to\infty} \sum_{k=1}^{n} E_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{k-1} R(X_t, A_t) + W_\lambda^*(X_k)\right) I[\tau^* = k]\right] \\ &= \sum_{k=1}^{\infty} E_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{k-1} R(X_t, A_t) + W_\lambda^*(X_k)\right) I[\tau^* = k]\right] \\ &= E_x^{f^*}\left[U_\lambda\left(\sum_{t=0}^{\tau^*-1} R(X_t, A_t) + W_\lambda^*(X_{\tau^*})\right) I[\tau^* < \infty]\right] \\ &= U_\lambda(V_\lambda(x; f^*, \tau^*)), \end{aligned}$$

donde la última igualdad se deriva de la combinación de (1.7) y (3.11). También se tiene que a través del Lema 2.5, (3.11) y (3.12) implican que:

$$\lim_{n\to\infty} E_x^{f^*}\left[\left|U_\lambda\left(\sum_{k=0}^{n} R(X_t, A_t)\right)\right| I[\tau^* > n+1]\right] = 0.$$

Al tomar el límite cuando n tiende a $+\infty$ en el lado derecho de (3.14), las dos últimas convergencias implican que $U_\lambda(W_\lambda^*(x)) = U_\lambda(V_\lambda(x; f^*, \tau^*))$, una igualdad que usando que U_λ estrictamente creciente conduce a que $W_\lambda^*(x) = V_\lambda(x; f^*, \tau^*)$, con $x \in S \setminus \{S^*\}$.

(ii) Por las desigualdades (2.36) y (2.39), las cuales se cumplen bajo el modelo comunicante que estamos considerando, se tiene que:

$$V_\lambda(\cdot; \pi, \tau^*) \leq W_\lambda^*(\cdot) \leq V_\lambda(\cdot; f^*, \tau), \quad (\pi, \tau) \in \mathcal{P} \times \mathcal{T},$$

y combinando esto con la parte (i) se deduce de la Definición 1.1 que (f^*, τ^*) es un equilibrio de Nash, completando la prueba. □

Para complementar la parte teórica, en la siguiente sección presentamos un ejemplo numérico que ilustra un método para identificar el punto fijo del operador T_λ y, posteriormente, la estrategia que constituye un equilibrio de Nash.

3.3. Un ejemplo numérico

Consideramos el Ejemplo 3.1 e introducimos el Algoritmo 1, que describe los pasos para calcular el punto fijo W_λ^*. El algoritmo especifica cuáles son los datos de entrada necesarios para su ejecución y detalla los elementos que genera como salida.

Algoritmo 1 Método para encontrar el equilibrio de Nash en el Ejemplo 3.1.

Requiere: $\lambda \neq 0$, $\{b_1, b_2, \ldots, b_N\}$, $S = \{1, 2, \ldots, \hat{S}\}$, with $\hat{S} \in \mathbb{N}$, $G(x)$, $R(x, a)$, ϵ.
Asegura: Iter, W_λ^*, f^*, S^*.
$W \leftarrow \mathbb{0}, \hat{W} \leftarrow \mathbb{1}, s \leftarrow \mathbb{0}$ (donde $\mathbb{0}$ y $\mathbb{1}$ representan matrices de ceros y de unos, respectivamente)
Iter $\leftarrow 0$, norm $\leftarrow \|\hat{W} - W\|$, $m \leftarrow 0$.
while norm $> \epsilon$ **do**
  for $l = 1 : N$ **do**
    $s(l) = U_\lambda(R(0, l) + W(1))$.
  end for
  $m = \text{mín}\{U_\lambda(G(0)), \text{máx}(s)\}$.
  $\hat{W}(0) = \log(m/sign(\lambda))/\lambda$.
  for $k = 1 : \hat{S} - 1$ **do**
    for $l = 1 : N$ **do**
      $s(l) = b(l) \cdot U_\lambda(R(k, l) + W(k+1)) + (1 - b(l)) \cdot U_\lambda(R(k, l) + W(k-1))$
    end for
    $m = \text{mín}\{U_\lambda(G(k)), \text{máx}(s)\}$.
    $\hat{W}(k) = \log(m/sign(\lambda))/\lambda$.
  end for
  for $l = 1 : N$ **do**
    $s(l) = U_\lambda(R(\hat{S}, l) + W(\hat{S} - 1))$.
  end for
  $m = \text{mín}\{U_\lambda(G(\hat{S})), \text{máx}(s)\}$.
  $\hat{W}(\hat{S}) = \log(m/sign(\lambda))/\lambda$.
  norm $= \|\hat{W} - W\|$.
  $W \leftarrow \hat{W}$.
  Iter $\leftarrow$ Iter+1.
end while
$W_\lambda^* = \hat{W}$.
Calcula f^* y S^* de acuerdo con (2.18) y (2.20), respectivamente.

Implementamos el Algoritmo 1 en MATLAB, y los resultados numéricos del experimento se presentan en las Tablas 3.1 y 3.2. Es evidente que tanto el número de iteraciones como el tamaño del conjunto S^* aumentan a medida que N crece. Además, se observa una discrepancia notable entre el número de iteraciones para los valores positivos y negativos de λ, y que el tamaño del conjunto S^* varía significativamente al cambiar los valores de λ (ver Figura 3.1). Los resultados permanecen sin cambios a medida que los valores de $\hat{S}$

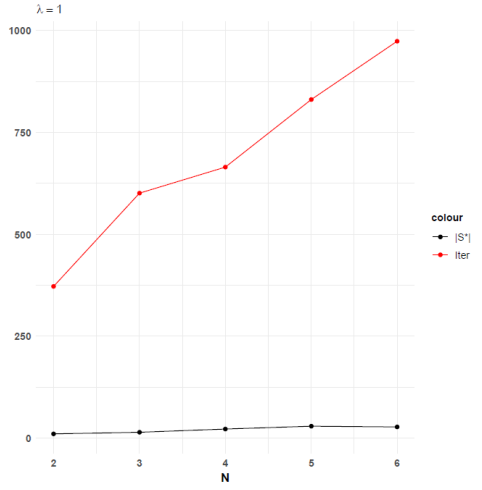
aumentan, manteniendo fijos N y λ. En cuanto a la estrategia f^*, se observó que, para valores positivos de λ, f^* generalmente adopta dos valores: el mínimo y el máximo del espacio de acciones. En contraste, para valores negativos de λ, f^* se mantiene prácticamente constante, asumiendo el valor mínimo del espacio de acciones.

Tabla 3.1: Desempeño numérico del Algoritmo 1 para diferentes valores de N, manteniendo fijos $\hat{S}$ y λ.

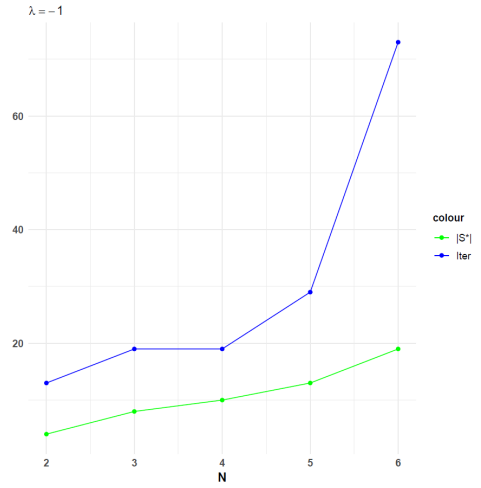
	N	2	3	4	5	6
$\hat{S}$=100000	Iter	372	601	665	831	974
λ=1	$\lvert S^*\rvert$	10	14	22	29	27
$\hat{S}$=100000	Iter	13	19	19	29	73
λ=-1	$\lvert S^*\rvert$	4	8	10	13	19

Tabla 3.2: Desempeño numérico del Algoritmo 1 para diferentes valores de λ, manteniendo fijos $\hat{S}$ y N.

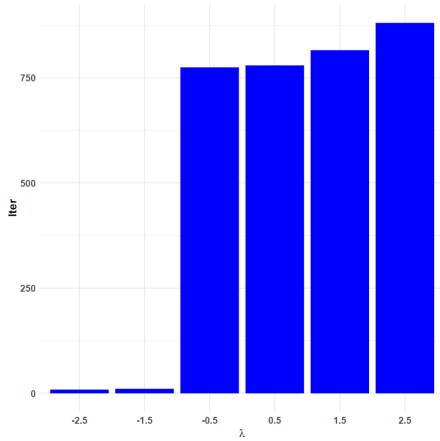
	λ	1/2	-1/2	3/2	-3/2	5/2	-5/2
$\hat{S}$=100000	Iter	780	775	816	11	881	9
N=4	$\lvert S^*\rvert$	10	10	28	5	54	7



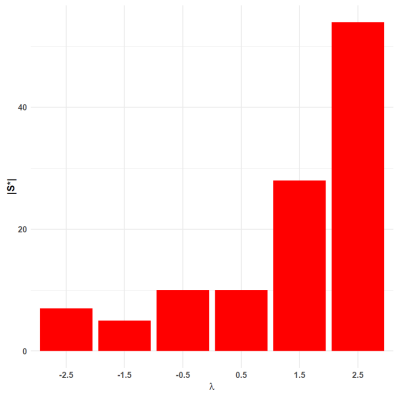
(a) Relación entre el número de iteraciones y el tamaño del conjunto S^* en función de N, con $\lambda = 1$.



(b) Relación entre el número de iteraciones y el tamaño del conjunto S^* en función de N, con $\lambda = -1$.



(c) Número de iteraciones para cada valor de λ.



(d) Tamaño del conjunto S^* para cada valor de λ.

Figura 3.1: Resultados numéricos de la implementación del Algoritmo 1 del Ejemplo 3.1.

Resumen, Conclusiones y Trabajo Futuro

En este trabajo de tesis, se estudió una clase de juegos de suma cero en tiempo discreto, espacio de estados numerable, transiciones Markovianas y recompensas acotadas, mediante el criterio de recompensa total sensible al riesgo. La sensibilidad al riesgo es una característica muy importante que debe tomarse en cuenta en la toma de decisiones. Por lo que es de vital importancia extender los trabajos que se tienen en el caso neutral al caso sensible.

Estudiamos el juego $\mathcal{G}$ considerando dos modelos diferentes como lo fueron el modelo absorbente y el modelo comunicante. Una vez explicada la dinámica del juego, lo siguiente fue enfocarse en el operador de equilibrio. Vimos cuáles eran sus características importantes y nuestro primer resultado fue demostrar que este operador tiene puntos fijos. Esta característica es primordial ya que a partir de este punto fijo definimos las estrategias de los jugadores que constituyen un equilibrio de Nash. Lo siguiente fue utilizar las desigualdades de los Teoremas 2.3 y 2.4, para en base a éllas, enfocar la prueba de la existencia del equilibrio de Nash.

En el modelo absorbente, la prueba de la existencia del equilibrio de Nash permitió demostrar de inmediato la igualdad de la función valor con el punto fijo, así como la unicidad de dicho punto fijo. En el caso del modelo comunicante, se caracterizaron a los puntos fijos del operador T_λ y se demostró que S^* es distinto del conjunto vacío, una propiedad que se había perdido al no considerar el supuesto de la existencia de un estado absorbente. Además, se demostró la unicidad sin depender de la igualdad entre la función valor y W_λ^*, cuya demostración también es independiente de la existencia de un equilibrio de Nash. Se presentó un ejemplo específico de un juego que cumple con los supuestos y que, además, se analizó numéricamente. Por lo tanto, en ambos modelos se obtuvieron los resultados esperados y en el modelo comunicante las demostraciones difieren ligeramente en su enfoque con respecto al modelo absorbente.

Cuando un modelo cuenta con más de un estado absorbente, el espacio de estados se divide en estados transitorios y recurrentes, lo que resulta en múltiples clases de comunicación. Este caso fue analizado en el contexto de los PDMs [1], donde el tomador de decisiones es averso al riesgo y el desempeño de una política de control se mide mediante el criterio del costo promedio a

largo plazo. En [1] no se imponen condiciones de comunicación a la ley de transición, por lo que la función de valor óptimo puede no ser constante. Para abordar esta situación, se introduce el concepto de sistema de optimalidad, que extiende la noción de ecuación de optimalidad y permite caracterizar la función de valor óptimo a través de un sistema de ecuaciones. Considerar este caso en el contexto de los *Markov stopping games* resulta un tema interesante, y podría explorarse en trabajos futuros.

Es importante señalar que una extensión a casos más generales con respecto al espacio de estados es una tarea complicada, ya que como se sabe a partir de la literatura sobre PDMs sensibles al riesgo, no siempre es posible. Un ejemplo de esto se puede ilustrar en la siguiente situación. En 1972 [22], Howard y Matheson demostraron que el costo promedio sensible al riesgo óptimo se determina mediante una ecuación de optimalidad en modelos finitos y comunicantes. Cuarenta años después, se demostró en [14] que el resultado pionero de Howard y Matheson no se puede extender al caso de un espacio de estados numerable. Por lo tanto, un problema futuro interesante es investigar la extensión factible de los resultados presentados en este trabajo a espacios más generales, como los espacios de Borel, y considerar la opción de incorporar posibles recompensas no acotadas.

En el futuro, se podrían considerar otros criterios de rendimiento. Uno de ellos sería la recompensa total descontada sensible al riesgo, que no solo incluiría un coeficiente de sensibilidad, sino también un factor de descuento para tener en cuenta el valor temporal de las recompensas. Este criterio ha sido estudiado en el contexto de los PDMs [18], y en el caso de juegos, se analizó en [37], donde los dos jugadores tienen la posibilidad de detener el juego. Por otro lado, una situación común que se presenta en las matemáticas aplicadas es que los datos necesarios para proponer un modelo matemático presentan ambigüedad, vaguedad o características aproximadas del problema en estudio. Una posibilidad para abordar esta situación es utilizar la teoría difusa. Las recompensas difusas han sido analizadas en el contexto de los PDMs, considerando tanto espacio de estados finito [10] como numerable [16]. Este enfoque resulta particularmente útil en entornos reales, donde las recompensas pueden depender de factores impredecibles o subjetivos. En este sentido, como extensión de este trabajo, se espera ampliar los resultados al ámbito difuso.

Apéndice A

Definiciones y Teoremas Auxiliares

A.1. Definiciones

Definición A.1. *(Función signo [2]) La función signo, $sgn : \mathbb{R} \to \{-1, 0, 1\}$ de un número real es una función por partes que se define de la siguiente manera:*

$$sgn(x) = \begin{cases} -1, & x < 0, \\ 0, & x = 0, \\ 1, & x > 0. \end{cases}$$

A.2. Teoremas

Teorema A.1. *(Desigualdad de Jensen [4]) Sea g una función convexa definida en un intervalo abierto I de números reales, el cual puede ser acotado o no. Sea X una variable aleatoria definida en el espacio de probabilidad $(\Omega, \mathcal{F}, P)$, tal que $X(\omega) \in I$, para todo ω. Supongamos que $E[X]$ es finita. Si $\mathcal{H}$ es una sub σ-álgebra de $\mathcal{F}$, entonces $E[g(X)|\mathcal{H}] \geq g(E[X|\mathcal{H}])$ c.s.. En particular, $E[g(X)] \geq g(E[X])$.*

Teorema A.2. *(Teorema de Dini [24]) Si la sucesión de funciones continuas $f_n : X \to \mathbb{R}$ converge monótonamente a la función continua $f : X \to \mathbb{R}$ en el conjunto compacto X, entonces la convergencia es uniforme.*

Teorema A.3. *(Teorema de convergencia dominada [31]) Sea $X_1, X_2, \ldots$ una sucesión de variables aleatorias tales que $\lim_{n\to\infty} X_n = X$ casi seguramente, y para cada valor de n, $|X_n| \leq Y$, para alguna variable Y con $E[|Y|] < \infty$. Entonces,*

$$\lim_{n\to\infty} E[X_n] = E[\lim_{n\to\infty} X_n].$$

Teorema A.4. *(Teorema de convergencia monótona [31]) Sea $X_1, X_2, \ldots$ una sucesión de variables aleatorias tales que $0 \leq X_1 \leq X_2 \leq \cdots$ y $\lim_{n\to\infty} X_n = X$ casi seguramente. Entonces,*

$$\lim_{n\to\infty} E[X_n] = E[\lim_{n\to\infty} X_n].$$

Bibliografía

[1] Alanís-Durán, A., & Cavazos-Cadena, R., "An optimality system for finite average Markov decision chains under risk-aversion". *Kybernetika, 48(1).* pp. 83-104. 2012.

[2] Alencar, M., & Alencar, R., "*Set, Measure and Probability Theory*". CRC Press. ISBN 9788770228473. 2024.

[3] Altman, E., & Shwartz, A., "Constrained Markov games: Nash equilibria". En J. A. Filar, V. Gaitsgory, & K. Mizukami (Eds.), *Advances in dynamic games and applications* (pp. 213-221). Annals of the International Society of Dynamic Games, vol. 5. Birkhäuser. Online ISBN 978-1-4612-1336-9. 2000.

[4] Ash, R., "*Real analysis and probability*". Academic Press. ISBN-10 0120652013. 1972.

[5] Atar, R., & Budhiraja, A., "A stochastic differential game for the inhomogeneous Laplace equation". *Annals of Probability. 38(2).* pp. 498-531. 2010.

[6] Bäuerle, N., & Rieder, U., "*Markov decision processes with applications to finance*". Springer Science & Business Media. eBook ISBN 978-3-642-18324-9. 2011.

[7] Bäuerle, N., & Rieder, U., "More risk-sensitive Markov decision processes". *Mathematics of Operations Research. 39(1).* pp. 105-120. 2014.

[8] Bielecki, T., Hernández-Hernández, D., & Pliska, S. R., "Risk sensitive control of finite state Markov chains in discrete time, with applications to portfolio management". *Mathematical Methods of Operations Research. 50.* pp. 167-188. 1999.

[9] Boucherie, R., & Van Dijk, N., "*Markov decision processes in practice*". Springer. eBook ISBN 978-3-319-47766-4. 2017.

[10] Carrero-Vera, K., Cruz-Suárez, H., & Montes-de-Oca, R., "Markov decision processes on finite spaces with fuzzy total rewards". *Kybernetika. 58(2).* pp. 180-199. 2022.

[11] Cavazos-Cadena, R., Cantú-Sifuentes, M., & Cerda-Delgado, I.,"Nash equilibria in a class of Markov stopping games with total reward criterion". *Mathematical Methods of Operations Research. 94.* pp. 319-340. 2021.

[12] Cavazos-Cadena, R., & Hernández-Hernández, D., "Nash equilibria in a class of Markov stopping games". *Kybernetika. 48(5).* pp. 1027-1044. 2012.

[13] Cavazos-Cadena, R., & Hernández-Hernández, D., "Exact and approximate Nash equilibria in discounted Markov stopping games with terminal redemption". *Journal of Mathematical Analysis and Applications. 433(2).* pp. 1110-1141. 2016.

[14] Cavazos-Cadena, R., "Characterization of the optimal risk-sensitive average cost in denumerable Markov decision chains". *Mathematics of Operations Research. 43(3).* pp. 1025-1050. 2018.

[15] Cavazos-Cadena, R., Rodríguez-Gutiérrez, L., & Sánchez-Guillermo, D., "Markov stopping games with an absorbing state and total reward criterion". *Kybernetika. 57(3)*,. pp. 474-492. 2021.

[16] Cruz-Suárez, H., Montes-de-Oca, R., & Ortega-Gutiérrez, R., "An extended version of average Markov decision processes on discrete spaces under fuzzy environment". *Kybernetika. 59(1).* pp. 160-178. 2023.

[17] Filar, J., & Vrieze, K., "*Competitive Markov decision processes*". Springer Science & Business Media. eBook ISBN 978-1-4612-4054-9. 2012.

[18] Guo, X., "Risk-sensitive discounted Markov decision processes with unbounded reward functions and Borel spaces". *Stochastics. 96(1).* pp. 649-666. 2024.

[19] Hernández-Lerma, O., "*Adaptive Markov control processes*". Vol. 79. Springer Science & Business Media. eBook ISBN 978-1-4419-8714-3. 2012.

[20] Hernández-Lerma, O., & Lasserre, J., "*Discrete-time Markov control processes: basic optimality criteria*". Vol. 30. Springer Science & Business Media. eBook ISBN 978-1-4612-0729-0. 2012.

[21] Hoel, P., Port, S., & Stone, C., "*Introduction to stochastic processes*". Waveland Press. ISBN-10 0881332674. 1986.

[22] Howard, R., & Matheson, J., "Risk-sensitive Markov decision processes". *Management Science. 18(7).* pp. 356-369. 1972.

[23] Kolokoltsov, V., & Malafeyev, O., "*Understanding game theory*". World Scientific. ISBN 978-9814291712. 2010.

[24] Lima, E., "*Análise real Vol. 1: Funções de uma variável*". IMPA. ISBN 978-65-990528-5-9. 2010.

[25] López-Rivero, J., Cavazos-Cadena, R., & Cruz-Suárez, H., "Risk-sensitive Markov stopping games with an absorbing state". *Kybernetika. 58(1).* pp. 101-122. 2022.

[26] López-Rivero, J., Cruz-Suárez, H., & Camilo-Garay, C., "Nash equilibria in risk-sensitive Markov stopping games under communication conditions". *AIMS Mathematics. 9(9).* pp. 23997-24017. 2024.

[27] Martínez-Cortés, V., "Bipersonal stochastic transient Markov games with stopping times and total reward criteria". *Kybernetika. 57(1).* pp. 1-14. 2021.

[28] Peskir, G., "On the American option problem". *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics. 15(1).* pp. 169-181. 2005.

[29] Peskir, G., & Shiryaev, A., "*Optimal stopping and free-boundary problems*". Birkhäuser Basel. eBook ISBN 978-3-7643-7390-0. 2006.

[30] Puterman, M., "*Markov decision processes: Discrete stochastic dynamic programming*". John Wiley & Sons. ISBN: 978-1-118-62587-3. 2014

[31] Rincón, L., "*Introducción a los procesos estocásticos*". Vol. 63. UNAM, Facultad de Ciencias. ISBN 9786070230448. 2012.

[32] Saucedo-Zul, J., Cavazos-Cadena, R., & Cruz-Suárez, H., "A discounted approach in communicating average Markov decision chains under risk aversion". *Journal of Optimization Theory and Applications. 187.* pp. 585-606. 2020.

[33] Shapley, L., "Stochastic games". *Proceedings of the National Academy of Sciences. 39.* pp. 1095-1100. 1953.

[34] Shiryaev, A., "*Optimal stopping rules*". Vol. 8. Springer Science & Business Media. eBook ISBN 978-3-540-74011-7. 2007.

[35] Von Neumann, J., & Morgenstern, O., "*Theory of games and economic behavior*". 2nd rev. ed.. Princeton University Press. ISBN 978-0691130613. 1947.

[36] Zachrisson, L., "Markov games". En M. Dresher, L. S. Shapley, & A. W. Tucker (Eds.), *Advances in game theory* (Annals of Mathematical Studies, Vol. 52, pp. 211-253). Princeton University Press. eBook ISBN 978-1-4008-8201-4. 1964.

[37] Zhang, W., & Liu, C., "Discrete-time stopping games with risk-sensitive discounted cost criterion". *Mathematical Methods of Operations Research.* pp. 1-30. 2024.