Tesis de Maestría

Juegos estocásticos transitorios y aplicaciones

Fabián Crocce

Orientador: Ernesto Mordecki
Centro de Matemática

Maestría en Matemática (PEDECIBA)
Facultad de Ciencias
Universidad de la República
Uruguay

4 de febrero de 2009

Resumen

Este trabajo desarrolla la teoría de los juegos estocásticos transitorios, basado principalmente en resultados del libro de Filar y Vrieze[10]; éstos son una clase particular de juego estocástico con horizonte infinito, en que se tiene un estado especial en que el juego se considera finalizado, que se alcanza con probabilidad uno, de modo que la suma total (sin descuento) de la ganancia instantánea a lo largo de los infinitos pasos resulta bien definida. En particular se llega a métodos concretos para hallar las estrategias óptimas para estos juegos y se incluye la aplicación a un juego de dados conocido como "la codicia" o "el uno".

Palabras Claves: teoría de juegos; juegos estocásticos; procesos de decisión de Markov competitivos; juegos de dados; juegos estocásticos transitorios.

Abstract

Based mainly in results in the Filar & Vrieze's book [10], theory of transient stochastic games is developed in this work; those are a particular case of infinite horizon stochastic games, in which a special state, considered as a final state, is reached with probability one, ensuring the finiteness of the sum of instant rewards over all steps. Particulary, concrete methods to find optimal strategies for transient stochastic games are reached, and an application to the resolution of a dice game, known as "pig game", is presented.

Keywords: game theory; stochastic games; competitive Markov decision processes; dice games; transient stochastic games.

Índice general

Introducción

La teoría de los *juegos estocásticos* es sin duda parte de la *teoría de juegos*; pero también puede entenderse como una extensión natural de la teoría de los *procesos de Markov controlados*, en que más de un agente toma decisiones persiguiendo objetivos diferentes. De un *juego estocástico* participan una cantidad finita de jugadores. El juego transcurre en una secuencia de instantes o pasos, que puede ser finita o infinita, en que éste puede encontrarse en diferentes estados. En cada uno de los pasos, cada jugador tiene un conjunto finito de posibles acciones, que depende del estado en que se encuentra el juego, del cual debe elegir una. Las acciones tomadas por los jugadores determinan un pago a éstos pero además influye en la dinámica del juego, ya que el estado siguiente es aleatorio y su distribución de probabilidad depende de las acciones tomadas. Si se piensa en este modelo con un solo jugador se tiene un *proceso de Markov controlado* o *proceso de decisión de Markov* clásico.

La clase de *juegos estocásticos* en que se centra este trabajo es la de los *juegos estocásticos transitorios*, en la que los jugadores tienen el objetivo de maximizar la suma de las ganancias esperadas de cada paso en una cantidad infinita de éstos. Es claro que para que la suma de infinitos términos resulte finita hay que pedirle ciertas condiciones al modelo. De hecho, como vamos a ver en el capítulo 3, los *juegos estocásticos transitorios* modelan situaciones en la que se tiene una cantidad no acotada de pasos, pero en cada juego, con probabilidad uno, dicha cantidad es finita.

La motivación para estudiar los *juegos estocásticos transitorios* surge a raíz de un problema concreto, un juego de dados. En mi trabajo monográfico de grado [4] estudié el juego conocido como "la codicia" o "el uno" desde un punto de vista solitario y quedó planteada la idea de hallar una estrategia óptima para el caso competitivo, en el que se enfrentan dos o más jugadores. Enfocado en esta dirección surgió la idea de hacer un seminario para estudiar el libro "Competitive Markov Decision Processes" [10], que trata de *juegos estocásticos* en general, donde encontramos la teoría de los *juegos estocásticos transitorios*, que resultó útil para resolver el problema.

Este trabajo se podría dividir en dos partes: por un lado los tres primeros

capítulos, que son de matemática pura, en los que aparecen resultados teóricos sobre *teoría de juegos*, y *juegos estocásticos*; y por otro lado el cuarto capítulo que es de matemática aplicada y se centra en la resolución del juego de dados. Los capítulos 1 y 2 son preparatorios para el tercer capítulo, se espera que alguien que no conoce de *teoría de juegos*, pero maneje conceptos básicos de *probabilidad* y *procesos estocásticos* pueda entender este trabajo sin gran dificultad. Para consultar aspectos básicos de probabilidad se recomienda [9, 22] y sobre procesos estocásticos se puede consultar en [3, 27].

En el primer capítulo se introducen aspectos básicos de la *teoría de juegos* y algunos resultados clásicos sobre juegos matriciales, que resultan de utilidad para el estudio de los *juegos estocásticos*. Buena parte de los contenidos de esté primer capítulo está basado principalmente en el libro "Uma introdução à teoria econômica dos jogos" [2]; en él se pueden consultar aspectos básicos generales sobre este tema.

El segundo capítulo trata juegos estocásticos en general, el modelo que definimos toma aspectos del modelo de *procesos de decisión de Markov* que aparece en el libro de Hernández-Lerma y Lasserre [13] y tiene aspectos del modelo planteado por Filar y Vrieze en [10]. La primera referencia [13] es excelente para el estudio de los *procesos de decisión de Markov*, así como lo es [10] para los *juegos estocásticos* de dos jugadores.

En el tercer capítulo abordamos la subclase de los *juegos estocásticos* que da nombre a este trabajo, los *juegos estocásticos transitorios*; muchos de los resultados que aparecen en él están basados en el libro de Filar y Vrieze, que es el único libro que encontramos que trata este tema. En particular, en este tercer capítulo, estudiamos a fondo el el problema de hallar estrategias óptimas para esta clase de juegos.

Por último en el cuarto capítulo se muestra el potencial de los resultados obtenidos cuando se tiene el poder de cálculo que brinda un computador; en él se muestra cómo se aplican dichos resultados al problema de "la codicia" y a algunas variantes de este juego; además se pueden ver los resultados obtenidos con los algoritmos desarrollados. También se incluye un pequeño apéndice en que se explican algunos detalles sobre los programas implementados.

Capítulo 1

Nociones sobre Teoría de Juegos

Introducción

La *teoría de juegos* en sus principios surgió como una teoría económica, pero en la actualidad tiene aplicaciones en muchas otras disciplinas tales como la biología, informática, inteligencia artificial, psicología, etc. Al principio del siglo XX se publicaron muchos artículos sobre *teoría de juegos* pero sin gran impacto. Fue a raíz de publicaciones de John von Neumann y del economista Oskar Morgenstern que dicha teoría cobró mayor importancia, sobre todo después de 1940. En particular la publicación en 1944 del libro "Theory of Games and Economic Behavior" [20], junto con la guerra fría y las posibles aplicaciones de la *teoría de juegos* a las estrategias militares, le dieron un fuerte impulso a esta disciplina. En [2, 26] se pueden consultar aspectos básicos sobre *teoría de juegos*.

Los juegos estudiados por la *teoría de juegos*, son modelos matemáticos definidos en cada caso; es decir, no se encuentra una definición de juego con rigurosidad matemática que sea unánime, sino que este concepto engloba una clase de situaciones que comparten un conjunto de características. Tratando de recoger dichas características comunes podríamos decir, en un sentido muy amplio, que un juego se compone de un conjunto de jugadores, un conjunto de estrategias que cada jugador puede elegir, y una especificación de recompensas para las estrategias escogidas. La teoría de juegos desarrolla modelos matemáticos para representar los juegos, estudia dichos modelos y en particular busca encontrar la forma de elegir las estrategias para optimizar el desempeño de los jugadores.

En este primer capítulo se presenta el modelo más básico y clásico de juego, los *juegos en forma normal*. Éstos resultan de utilidad para comprender algunos conceptos, tales como *suma cero*, *estrategia*, *equilibrio de Nash*; y además sirven de base para los *juegos estocásticos*, que son el objetivo de este trabajo.

1.1. Juegos en forma estratégica o normal

Definición 1.1.1 (**Juegos en forma estratégica, o en forma normal**). *Un juego en forma estratégica, o normal, es una 3-úpla* $(J, \boldsymbol{A}, R)$ *donde:*

- J *es un número natural distinto de cero que indica la cantidad de jugadores. A cada jugador lo identificamos con un número* $j = 1, \ldots, J$.
- $\boldsymbol{A} = A_1 \times \ldots \times A_J$ *es el conjunto de J-úplas de acciones posibles por parte de los jugadores, para cada* $j = 1, \ldots, J$, A_j *es un conjunto finito,*

$$A_j = \{1, \ldots, m_j\},$$

  cuyos elementos se denominan ***acciones*** *o* ***estrategias puras*** *del jugador* j.
- $R = (r_1, r_2, \ldots, r_J)$ *es la asignación de recompensas. La recompensa del jugador* j *es un número real que depende de las acciones elegidas por todos los jugadores. En definitiva, se tiene* $r_j : \boldsymbol{A} \to \mathbb{R}$, *de modo que* $r_j(a_1, \ldots, a_J)$ *es la recompensa que recibe el jugador* j *si los jugadores* $k : k = 1 \ldots n$ *toman las acciones* $a_k \in A_k$ *respectivamente.*

Además se tienen las siguientes reglas que rigen el juego:

1. *la cantidad* J *de jugadores, los conjuntos* $A_j : j = 1 \ldots J$, *de acciones posibles de cada jugador, y las funciones de recompensa forman parte de las reglas del juego;*
2. *cada jugador escoge su acción sin conocer, y de manera totalmente independiente, de la elección de los demás jugadores, intentando maximizar su recompensa;*
3. *todos los jugadores conocen las reglas del juego.*

A los juegos que cumplen la regla 1 y 3, que asegura que todos los jugadores son conscientes de la estructura del juego y además saben que los demás jugadores también conocen dicha estructura, se les denomina *juegos de información completa*. A los que cumplen la regla 2 se los denomina *juegos no cooperativos*, o *competitivos*, ya que se impide formar alianzas entre jugadores, para jugar cooperativamente, y se asegura que cada jugador debe jugar para maximizar su propia recompensa. Nuestro interés es en *juegos competitivos de información completa*.

El siguiente ejemplo, que es un clásico en la *teoría de juegos*, muestra la importancia de la hipótesis de no cooperación. Lo presentó Albert Tucker en 1950 en un seminario para psicólogos de la Universidad de Standford.

Ejemplo 1.1.2 (**Dilema del prisionero**). *Dos delincuentes son atrapados por la policía, no hay pruebas para condenarlos salvo por un delito menor, se los aisla a uno del otro y se les propone confesar el delito que cometieron, bajo las siguientes reglas: si ambos confiesan comparten la pena (5 años cada uno); si uno confiesa y el otro no, el que confiesa queda libre y el otro cumple toda la pena (10 años); y si los dos negaran, por falta de pruebas, sólo se los condena por el delito menor (1 año cada uno).*

Para modelar el juego tenemos $J = 2$, cada detenido es un jugador. El conjunto de acciones para ambos jugadores es el mismo, es decir:

$$A_1 = A_2 = \{1 = \text{confesar}, 2 = \text{negar}\}.$$

Es claro que la máxima recompensa es quedar libre; de modo que la recompensa la podemos modelar como el opuesto de los años de cárcel. Tendríamos las funciones r_1 y r_2 definidas así:

$$\begin{array}{rclrcl} r_1(\text{confesar}, \text{confesar}) & = & -5, & r_1(\text{confesar}, \text{negar}) & = & 0, \\ r_1(\text{negar}, \text{confesar}) & = & -10, & r_1(\text{negar}, \text{negar}) & = & -1, \end{array}$$

$$\begin{array}{rclrcl} r_2(\text{confesar}, \text{confesar}) & = & -5, & r_2(\text{confesar}, \text{negar}) & = & -10, \\ r_2(\text{negar}, \text{confesar}) & = & 0, & r_2(\text{negar}, \text{negar}) & = & -1. \end{array}$$

Una forma reducida y más clara de representar el juego, es mediante una matriz en que las filas son las acciones del primer jugador, las columnas son las acciones del segundo jugador, y los valores de la matriz son pares ordenados cuya primera componente representa la recompensa del jugador 1, y la segunda componente la recompensa del jugador 2, para las acciones correspondientes. En este caso tendríamos

j. 1 \ *j.* 2	confesar	negar
confesar	(-5,-5)	(0,-10)
negar	(-10,0)	(-1,-1)

Observar que confesar el delito aumenta la recompensa, independientemente de la opción que tome el otro jugador. De modo que en un juego competitivo ambos jugadores deberían optar por confesar, siendo ambos condenados a 5 años de prisión. Si, en cambio, ambos hubiesen negado, sólo serían condenados a un año cada uno, aumentando la recompensa; pero esto requeriría una decisión en cooperación, confiar en que el otro jugador no nos va a "traicionar".

Definición 1.1.3 (**Estrategia mixta**). *Consideremos un juego en forma estratégica; una estrategia mixta (o simplemente estrategia) μ^j, para el jugador j, es una distribución de probabilidad en el conjunto A_j, de estrategias puras del jugador. La idea es que el jugador no tiene que elegir la acción a tomar*

directamente, sino que puede optar por "sortear" qué acción tomar eligiendo la distribución de probabilidad. En algunas ocasiones es cómodo pensar la estrategia μ^j como un vector $(\mu^j_1, \ldots, \mu^j_{m_j})$, donde $\mu^j_a = \mu^j(a)$ para cada acción $a = 1, \ldots, m_j$; claramente vale que $\sum_{a=1}^{m_j} \mu^j_a = 1$.

Notación 1.1.4.

- *Denominamos Δ_n al subconjunto de $\mathbb{R}^n$ de probabilidades sobre el conjunto $\{1, \ldots, n\}$. Es decir,*

$$\Delta_n = \left\{ (x_1, \ldots, x_n) \in \mathbb{R}^n : \ \forall j \ x_j \geq 0, \ \sum_{j=1}^{n} x_j = 1 \right\}.$$

  Con esta notación una estrategia mixta del jugador j es un elemento de Δ_{m_j}.

- *Llamamos Δ al conjunto de todas las elecciones posibles de estrategias por parte de los jugadores. Es decir*

$$\Delta = \Delta_{m_1} \times \ldots \times \Delta_{m_J}.$$

Si cada jugador $j = 1, \ldots, J$ fija un estrategia mixta μ^j para elegir su acción en el juego queda definido un elemento $\mu = (\mu^1, \ldots, \mu^J) \in \Delta$; además, dado que los jugadores eligen de forma independiente sus acciones, queda definida una distribución de probabilidad asociada a μ, que denotamos con la propia μ, sobre el conjunto $\mathbf{A} = A_1 \times \ldots A_J$ de las posibles acciones. De modo que si $\vec{a} = (a_1, \ldots, a_J) \in \mathbf{A}$ se tiene que

$$\mu(\vec{a}) = \mu^1(a_1) \ldots \mu^J(a_J).$$

Observación 1.1.5. *La estrategia pura $a \in A_j$ del jugador j se obtiene como caso particular de estrategia mixta tomando como distribución de probabilidad en A_j la que acumula toda la probabilidad en a, o sea la función indicatriz de a, que denotamos $\mathbb{I}_a$.*

Definición 1.1.6 (**Recompensa para estrategias mixtas**)**.** *Extendemos la definición de recompensa para contemplar estrategias mixtas, de modo que si μ es una elección de estrategias mixtas $r_j(\mu)$ coincida con el valor esperado de la recompensa, o sea,*

$$r_j(\mu) = \sum_{\vec{a} \in \boldsymbol{A}} r_j(\vec{a}) \mu(\vec{a}).$$

Notar que si μ se compone de estrategias puras, o sea, $\mu = \mathbb{I}_{\vec{a}} = (\mathbb{I}_{a_1}, \ldots, \mathbb{I}_{a_J})$ se cumple

$$r_j(\mu) = r_j(\vec{a})$$

1.1.1. Equilibrio de Nash

Notación 1.1.7. *Si $\mu = (\mu^1, \ldots, \mu^J) \in \Delta$ es una elección de estrategias y $\nu^j \in \Delta_{m_j}$ es una estrategia para el jugador j, denotamos por $\mu[j, \nu^j]$ a la elección de estrategias que surge de cambiar en μ la estrategia tomada por el jugador j y sustituirla por ν^j. Es decir,*

$$\mu[j, \nu^j] = (\mu^1, \ldots, \mu^{j-1}, \nu^j, \mu^{j+1}, \ldots, \mu^J).$$

Definición 1.1.8 (**Equilibrio de Nash**). *Decimos que una elección de estrategias $\mu = (\mu^1, \ldots, \mu^J) \in \Delta$, de los jugadores $1, \ldots, J$ respectivamente, forman un equilibrio de Nash si ningún jugador puede mejorar su recompensa cambiando su estrategia si los demás jugadores no cambian la suya. Es decir, para todo $j \in \{1, \ldots, J\}$ y para toda estrategia $\nu^j \in \Delta_{m_j}$, se tiene*

$$r_j(\mu) \geq r_j(\mu[j, \nu^j]). \tag{1.1}$$

El concepto de *equilibrio de Nash* es de suma importancia en la *teoría de juegos*; lo introdujo John Nash por primera vez en su disertación *"Non-cooperative games" (1950) [16]*, en donde mostró que todas las estrategias óptimas que se habían presentado para distintos juegos no cooperativos, efectivamente cumplían la propiedad de formar un *equilibrio de Nash*. En el artículo *"Non-Cooperative Games" (1951) [17]* también se presenta este concepto que era conocido como *punto de equilibrio*.

En el ejemplo 1.1.2, del "dilema del prisionero", la combinación de estrategias $(\text{confesar}, \text{confesar})$ forma un *equilibrio de Nash*. Este equilibrio tiene la particularidad de lograrse con estrategias puras (no mixtas). Si bien la estrategia $(\text{negar}, \text{negar})$ es mejor, en el sentido de que ambos reducen su pena, resulta ser una estrategia inestable, ya que cada jugador se beneficia cambiando su estrategia si su oponente la conserva; no forma un *equilibrio de Nash*.

No siempre se puede lograr un *equilibrio de Nash* con estrategias puras. En cambio, como lo indica el siguiente teorema, para juegos en forma estratégica, siempre es posible encontrar uno si se trabaja con estrategias mixtas.

Teorema 1.1.9 (**existencia del equilibrio de Nash**). *Todo juego en forma estratégica tiene al menos un equilibrio de Nash.*

Existen muchas demostraciones de este resultado debido al propio Nash, en su artículo "Equilibrium Points in n-Person Games", publicado en 1950, se encuentra una prueba utilizando el teorema de Kakutani. Aquí incluimos una demostración que utiliza el teorema del punto fijo de Brouwer (una demostración de este teorema se puede consultar en el artículo de Milnor [14]).

Teorema 1.1.10 (**del punto fijo, de Brouwer**). *Si C es un subconjunto compacto, convexo y no vacío de un espacio euclideano de dimensión finita y*

$F : C \to C$ es una función continua, entonces F tiene un punto fijo. Es decir, existe $x \in C$ tal que $F(x) = x$.

Antes de demostrar el teorema 1.1.9, vemos algunos lemas previos que son caracterizaciones de un equilibrio de Nash.

Lema 1.1.11. *La elección de estrategias $\mu = (\mu^1, \ldots, \mu^J) \in \Delta$ forma un equilibrio de Nash, si y sólo si ningún jugador puede mejorar su recompensa cambiando su estrategia por una estrategia pura si los demás jugadores mantienen la suya. Es decir, μ forma un equilibrio de Nash, si y sólo si, para cada $j = 1, \ldots, J$ y para cada $a \in A_j$ se tiene que,*

$$r_j(\mu) \geq r_j(\mu[j, \mathbb{I}a])$$

Demostración.

($\Rightarrow$) La demostración del directo es inmediata, ya que si μ forma un equilibrio de Nash, la ecuación (1.1) vale para toda estrategia $\nu^j \in \Delta_{m_j}$, en particular vale tomando $\nu^j = \mathbb{I}a$.

($\Leftarrow$) El recíproco es una consecuencia de que, fijado $j \in \{1, \ldots, J\}$ y ν^j una estrategia del jugador j,

$$\nu^j = \sum_{a=1}^{m_j} \nu^j(a)\mathbb{I}_a,$$

de donde surge que

$$\begin{aligned} r_j(\mu[j, \nu^j]) &= \sum_{a=1}^{m_j} \nu^j(a) r_j(\mu[j, \mathbb{I}_a]) \\ &\leq \sum_{a=1}^{m_j} \nu^j(a) r_j(\mu) \\ &= r_j(\mu), \end{aligned}$$

lo que completa la demostración.

□

Lema 1.1.12. *Consideremos, para cada jugador $j = 1, \ldots, J$, y para cada acción $a = 1, \ldots, m_j$ de éste, la función $g_{ja} : \Delta \to \mathbb{R}$, tal que $\forall \mu \in \Delta$,*

$$g_{ja}(\mu) = \text{máx}\{0,\ r_j(\mu[j, \mathbb{I}a]) - r_j(\mu)\}. \tag{1.2}$$

La elección de estrategias $\mu = (\mu^1, \ldots, \mu^J) \in \Delta$ forma un equilibrio de Nash, si y sólo si $g_{ja}(\mu) = 0$ para todo $j = 1, \ldots, J$ y para toda $a = 1, \ldots, m_j$.

Demostración. La demostración es inmediata del lema anterior si se tiene en cuenta que, para toda elección de estrategias $\mu \in \Delta$,

$$g_{ja}(\mu) = 0 \Leftrightarrow r_j(\mu) \geq r_j(\mu[j, \mathbb{I}a]).$$

□

Lema 1.1.13. *Sea la función $F : \Delta \to \Delta$, tal que, si $\mu = (\mu^1, \ldots, \mu^J)$ es una elección de estrategias,*

$$F(\mu) = \nu = (\nu^1, \ldots, \nu^J);$$

donde, si $a \in \{1, \ldots, m_j\}$,

$$\nu_a^j = \nu^j(a) = \frac{\mu_a^j + g_{ja}(\mu^j)}{1 + \sum_{k=1}^{m_j} g_{jk}(\mu^j)}.$$

Se cumple que μ forma un equilibrio de Nash si, y solamente si, $F(\mu) = \mu$, es decir, μ es un punto fijo de F.

Demostración. Primero que nada veamos que la función F está bien definida, es decir, que ν es un elemento de Δ: fijado $j \in \{1 \ldots, J\}$ y $a \in A_j$, se cumple que $\nu_a^j \geq 0$, ya que $g_{jk}(\mu)$ toma valores no negativos para toda $k \in A_j$; además,

$$\begin{aligned}
\sum_{a=1}^{m_j} \nu_a^j &= \sum_{a=1}^{m_j} \frac{\mu_a^j + g_{ja}(\mu)}{1 + \sum_{k=1}^{m_j} g_{jk}(\mu)} \\
&= \frac{\sum_{a=1}^{m_j} [\mu_a^j + g_{ja}(\mu)]}{1 + \sum_{k=1}^{m_j} g_{jk}(\mu)} \\
&= \frac{1 + \sum_{a=1}^{m_j} g_{ja}(\mu)}{1 + \sum_{k=1}^{m_j} g_{jk}(\mu)} = 1,
\end{aligned}$$

lo que demuestra que $\nu^j \in \Delta_{m_j}$, entonces, ν es un elemento de Δ.

($\Rightarrow$) Para demostrar el directo basta observar que, por el lema 1.1.12, el hecho de que μ forme un equilibrio de Nash implica que $g_{ja}(\mu)$ es cero para todo jugador j y toda acción a; lo que a su vez implica que ν_a^j coincide con μ_a^j; es decir, efectivamente μ es un punto fijo de la función F.

($\Leftarrow$) Veamos que vale el recíproco. Sea μ un punto fijo de F, entonces vale la siguiente igualdad, para todo $j = 1, \ldots, J$ y $a = 1, \ldots, m_j$,

$$\mu_a^j = \frac{\mu_a^j + g_{ja}(\mu)}{1 + \sum_{k=1}^{m_j} g_{jk}(\mu)},$$

despejando se obtiene

$$\mu_a^j \sum_{k=1}^{m_j} g_{jk}(\mu) = g_{ja}(\mu). \tag{1.3}$$

Si probamos que $g_{ja}(\mu)$ es nulo para todo $j = 1, \ldots, J$ y $a = 1, \ldots, m_j$, obtenemos que μ forma un equilibrio de Nash como consecuencia del lema 1.1.12. Supongamos, por absurdo, que existe j tal que $\sum_{k=1}^{m_j} g_{jk}(\mu) > 0$; de la igualdad (1.3) resulta que $\mu_a^j > 0$ si y solo si $g_{ja}(\mu) > 0$, lo que es equivalente a que $r_j(\mu[j, \mathbb{I}a]) > r_j(\mu)$. Supongamos, sin pérdida de generalidad, que

$$\mu_1^j > 0, \ldots, \mu_l^j > 0, \mu_{l+1}^j = 0, \ldots, \mu_{m_j}^j = 0,$$

de modo que $\mu_a^j = \sum_{a=1}^{l} \mu_a^j \mathbb{I}_a$, lo que implica que

$$\begin{aligned} r_j(\mu) &= \sum_{a=1}^{l} \mu_a^j r_j(\mu[j, \mathbb{I}a]) \\ &> \sum_{a=1}^{l} \mu_a^j r_j(\mu) \\ &= r_j(\mu), \end{aligned}$$

lo que es absurdo; entonces para todo $j = 1, \ldots, J$ se cumple $\sum_{k=1}^{m_j} g_{jk}(\mu) = 0$, y como los sumandos son no negativos, $g_{jk}(\mu) = 0$ para toda acción k del jugador j, lo que culmina la prueba del lema.

□

Demostración de 1.1.9.

Aplicando el teorema 1.1.10, del punto fijo, con $C = \Delta$, que es compacto y convexo, y con la función F definida en el lema anterior, que es continua, resulta que existe un punto fijo de F que, por el lema anterior, sabemos que forma un equilibrio de Nash.

□

1.2. Juegos matriciales de suma cero

Definición 1.2.1 (**Juego matricial**). *Definimos un juego matricial como un juego en forma estratégica de dos jugadores* ($J = 2$). *Para simplificar la notación denotamos*

$$A = \{1, \ldots, n\} \quad y \quad B = \{1, \ldots, m\}$$

al conjunto de acciones de los jugadores 1 y 2 respectivamente.

La denominación de juego matricial radica en que, como se vio en el ejemplo 1.1.2, estos juegos se pueden representar por una matriz en $M_{n\times m}(\mathbb{R}^2)$, donde las filas corresponden a las acciones que puede tomar el jugador 1, las columnas a las acciones del jugador 2 y las entradas de la matriz son pares ordenados que indican la ganancia del jugador 1 y 2 respectivamente. Por supuesto que

los resultados vistos en general para los juegos en forma estratégica valen para los juegos matriciales.

Definición 1.2.2 (**Juego matricial de suma cero**). *Un juego matricial se dice de suma cero cuando la ganancia de un jugador coincide con la pérdida del contrincante. Es decir,*

$$\forall a \in A,\ \forall b \in B,\ r_1(a,b) = -r_2(a,b).$$

En este caso no utilizamos subíndices, sino que denotamos por r a la función r_1, de modo que r_2 es sencillamente $-r$. Estos juegos se representan mediante una matriz $R \in M_{n\times m}(\mathbb{R})$ de entradas $r_{ab} = r(a,b)$.

Cuando se tiene un juego matricial de suma cero y se eligen las acciones a y b se suele decir que el jugador 2 debe pagar al jugador 1 un monto de $r(a,b)$, de modo que el objetivo del jugador 1 es maximizar r, mientras que el jugador 2 trata de minimizarla.

Como ya fue visto en general, para los juegos en forma normal, la función de recompensa r se extiende a $\Delta = \Delta_n \times \Delta_m$, contemplando así las estrategias mixtas. Recordamos que la forma de extenderla, si $\mu = (\mu_1, \ldots, \mu_n) \in \Delta_n$ y $\nu = (\nu_1, \ldots, \nu_m) \in \Delta_m$, es tomar

$$r(\mu,\nu) = \sum_{a=1}^{n}\sum_{b=1}^{m} r(a,b)\mu_a\nu_b, \tag{1.4}$$

que escrito en forma de producto de vectores por matrices es

$$r(\mu,\nu) = \mu R \nu^T.$$

Definición 1.2.3 (**Valor de un juego matricial**). *Consideremos un juego matricial de suma cero, de matriz R, y llamemos*

$$v_1 = \max_{\mu\in\Delta_n}\min_{\nu\in\Delta_m} r(\mu,\nu)$$

y

$$v_2 = \min_{\nu\in\Delta_m}\max_{\mu\in\Delta_n} r(\mu,\nu).$$

Si se cumple que $v_1 = v_2$ decimos que "el juego tiene un valor", que denotamos $v^*(= v_1 = v_2)$. En caso de haber más de un juego matricial en el contexto denotamos el valor del juego por $\mathrm{val}(R)$.

La continuidad de la función $r : \Delta_n \times \Delta_m \to \mathbb{R}$ y la compacidad de Δ_n y Δ_m aseguran la existencia de v_1 y v_2. En contextos más generales, como el caso en que las estrategias puras posibles son infinitas, se cambia la definición de v_1 y v_2 tomando supremo e ínfimo en lugar de máximo y mínimo.

En general se cumple que v_1 es el máximo valor que el jugador 1 puede asegurarse sin depender de la estrategia de su oponente, o lo que es lo mismo, suponiendo que el oponente juega la mejor estrategia para minimizar r. Análogamente v_2 es el mínimo valor que el jugador 2 puede lograr sin depender del jugador 1.

Definición 1.2.4 (**Estrategia óptima**). *Decimos que $\mu^* \in \Delta_n$ es una estrategia óptima para el jugador 1 si se cumple*

$$\min_{\nu \in \Delta_m} r(\mu^*, \nu) = v_1.$$

Análogamente, una estrategia óptima para el jugador 2, es una estrategia $\nu^* \in \Delta_m$ que verifica

$$\max_{\mu \in \Delta_n} r(\mu, \nu^*) = v_2.$$

Lema 1.2.5. *Se cumple que $v_1 \leq v_2$.*

Demostración. Sabemos que

$$r(\mu, \nu) \geq \min_{\nu' \in \Delta_m} r(\mu, \nu'),$$

tomando máximo en μ obtenemos

$$\max_{\mu \in \Delta_n} r(\mu, \nu) \geq \max_{\mu \in \Delta_n} \min_{\nu' \in \Delta_m} r(\mu, \nu') = v_1,$$

y como la desigualdad anterior vale para todo ν podemos tomar mínimo, obteniendo

$$v_2 = \min_{\nu \in \Delta_m} \max_{\mu \in \Delta_n} r(\mu, \nu) \geq v_1,$$

lo que culmina la prueba.

□

El siguiente teorema caracteriza los equilibrios de Nash, para el caso de juegos matriciales de suma cero, en términos de v_1 y v_2.

Teorema 1.2.6. *En un juego matricial de suma cero, existe un par de estrategias formando un equilibrio de Nash si y sólo si el juego tiene un valor v^*. Además si (μ^*, ν^*) forman el equilibrio de Nash, resulta que $r(\mu^*, \nu^*) = v^*$, y las estrategias son óptimas.*

Demostración.

($\Rightarrow$) Si μ^* y ν^* forman un equilibrio de Nash, para todo μ se cumple

$$r(\mu^*, \nu^*) \geq r(\mu, \nu^*),$$

por lo tanto

$$\begin{aligned} r(\mu^*, \nu^*) &= \max_{\mu \in \Delta_n} r(\mu, \nu^*) \\ &\geq \min_{\nu \in \Delta_m} \max_{\mu \in \Delta_n} r(\mu, \nu) \\ &= v_2; \end{aligned}$$

de manera completamente análoga se tiene que

$$\begin{aligned} r(\mu^*, \nu^*) &= \min_{\nu \in \Delta_m} r(\mu^*, \nu) \\ &\leq \max_{\mu \in \Delta_n} \min_{\nu \in \Delta_m} r(\mu, \nu) \\ &= v_1; \end{aligned}$$

teniendo en cuenta que $v_1 \leq v_2$ por el lema anterior, probamos que

$$v_1 = \min_{\nu \in \Delta_m} r(\mu^*, \nu) = r(\mu^*, \nu^*) = \max_{\mu \in \Delta_n} r(\mu, \nu^*) = v_2$$

lo que demuestra la existencia del valor del juego. Además queda demostrado de la ecuación anterior que $r(\mu^*, \nu^*) = v^*$ y que las estrategias μ^* y ν^* son óptimas para el jugador 1 y 2 respectivamente.

($\Leftarrow$) Supongamos que existe un valor v^*, es decir $v^* = v_1 = v_2$, de la definición de v_1 surge que existe μ^* tal que

$$v_1 = \min_{\nu \in \Delta_m} r(\mu^*, \nu)$$

y análogamente, por la definición de v_2, existe ν^* tal que

$$v_2 = \max_{\mu \in \Delta_n} r(\mu, \nu^*);$$

por lo tanto se tiene que

$$\max_{\mu \in \Delta_n} r(\mu, \nu^*) = \min_{\nu \in \Delta_m} r(\mu^*, \nu).$$

Veamos que (μ^*, ν^*) forman un equilibrio de Nash:

$$\begin{aligned} r(\mu^*, \nu^*) &\geq \min_{\nu \in \Delta_2} r(\mu^*, \nu) \\ &= \max_{\mu \in \Delta_n} r(\mu, \nu^*) \\ &\geq r(\mu', \nu^*) \end{aligned}$$

para todo $\mu' \in \Delta_n$, de modo que el jugador 1 no puede mejorar su desempeño cambiando unilateralmente su estrategia. Análogamente se prueba para el jugador 2, lo que completa la demostración.

□

Corolario 1.2.7 (**Teorema minimax de von Neumann**). *Todo juego matricial de suma cero tiene un valor*

$$v^* = \max_{\mu\in\Delta_n} \min_{\nu\in\Delta_m} r(\mu,\nu) = \min_{\nu\in\Delta_m} \max_{\mu\in\Delta_n} r(\mu,\nu)$$

y estrategias óptimas para ambos jugadores.

Demostración. Por el teorema 1.1.9 sabemos que existe un equilibrio de Nash, por lo tanto, aplicando el teorema anterior obtenemos la igualdad deseada.

□

El teorema anterior vale en contextos más generales, su primera versión fue de von Neumann en 1928 [19] y prueba la existencia de soluciones en estrategias mixtas para juegos de dos jugadores, de suma cero, con una cantidad arbitraria de estrategias puras. Otra demostración del mismo resultado se puede consultar en [21].

Propiedades 1.2.8. *Consideremos dos juegos matriciales, dados por matrices R y R' en $M_{n\times m}(\mathbb{R})$, de entradas r_{ab} y r'_{ab}, respectivamente. Sea $\mathbb{J}$ la matriz en $M_{n\times m}(\mathbb{R})$ cuyas entradas son todas 1.*

1. *Si $R \leq R'$ coordenada a coordenada, entonces* $\operatorname{val}(R) \leq \operatorname{val}(R')$.
2. *Se cumple* $\operatorname{val}(R + k\mathbb{J}) = \operatorname{val}(R) + k$.
3. *Vale la siguiente desigualdad:*

$$|\operatorname{val}(R) - \operatorname{val}(R')| \leq \max_{a,b} |r_{ab} - r'_{ab}| \tag{1.5}$$

Demostración.

1. La afirmación es inmediata si se observa que para cada par de estrategias μ, ν se tiene que

$$r(\mu,\nu) = \mu R\nu^T \leq \mu R'\nu^T = r'(\mu,\nu),$$

y tomando $\max_\mu \min_\nu$ se conserva la desigualdad.

2. Procedemos de forma similar, observando que

$$\mu(R + k\mathbb{J})\nu^T = \mu R\nu^T + k,$$

para todo par de estrategias.

3. Se deduce de las dos partes anteriores, ya que

$$R \leq R' + \max_{a,b} |r_{ab} - r'_{ab}|\mathbb{J},$$

y entonces

$$\text{val}(R) \leq \text{val}(R') + \max_{a,b} |r_{ab} - r'_{ab}|.$$

Análogamente se obtiene

$$\text{val}(R') \leq \text{val}(R) + \max_{a,b} |r_{ab} - r'_{ab}|$$

culminando la prueba.

□

Capítulo 2

Juegos Estocásticos

Introducción

El concepto de *juego estocástico (stochastic game)* fue introducido por Shapley en 1953 [24]. A diferencia de los *juegos en forma normal*, presentados en el capítulo anterior, los *juegos estocásticos*, también llamados *procesos de decisión de Markov competitivos*, son juegos dinámicos ya que se juegan en varias etapas. Pueden ser vistos tanto como una generalización de los *juegos en forma normal*, en que se consideran varias etapas, o como una generalización de los *procesos de decisión de Markov*, en que se permite que varios agentes tomen decisiones persiguiendo objetivos diferentes. A pesar de la fuerte relación que tienen los *juegos estocásticos* con los *procesos de decisión de Markov*, estos dos temas fueron tratados de manera totalmente independientes durante largo tiempo. En 1997 Filar y Vrieze publicaron el libro "Competitive Markov Decision Processes" [10] cuyo título refiere a los *juegos estocásticos*, en el que se tratan ambos temas conjuntamente.

En términos muy generales, el modelo consiste en un conjunto de jugadores que en una secuencia de pasos participan de un *juego en forma normal*, que depende del estado en que se encuentra el juego, y en que las decisiones tomadas por los jugadores, además de determinar el pago propio del juego en forma normal, influye en el estado en que estará el juego estocástico en el paso siguiente. Vamos a ver que los *juegos estocásticos* abarcan las características generales de un juego; es decir, hay un conjunto de jugadores, estrategias para los jugadores y un sistema de retribuciones que sirve para comparar estrategias.

Como ya fue mencionado, el objetivo de este trabajo es estudiar en profundidad una subclase de los *juegos estocásticos* llamados *transitorios*. En este capítulo definimos el concepto de *juego estocástico*, en particular considerando el caso de dos jugadores, para, en el próximo capítulo, dedicarnos a los *juegos estocásticos transitorios*. Vamos a ver que muchos conceptos que aparecen (tales

como *juego de suma cero* y *valor del juego*) son análogos a los definidos en el capítulo 1 para juegos matriciales.

2.1. ¿Qué es un juego estocástico?

Definición 2.1.1 (**Juego estocástico**). *Un juego estocástico queda definido por:*

- *un conjunto finito S, digamos $S = \{1, \ldots, N\}$, de los posibles estados del juego;*
- *para cada estado $i \in S$, se tiene un juego matricial R^i (definido en el capítulo anterior) donde denotamos:*
  - *A^i y B^i a los conjuntos de acciones del jugador 1 y 2 respectivamente,*
  - *$r_1^i,\ r_2^i : A^i \times B^i \to \mathbb{R}$ a las funciones de pago de los jugadores 1 y 2;*
- *una familia de distribuciones de probabilidad sobre S, indexada en el conjunto de configuraciones instantáneas del juego*

$$\mathbb{K} = \{(i, a, b) | i \in S, a \in A^i, b \in B^i\}, \tag{2.1}$$

de modo que si $(i, a, b) \in \mathbb{K}$ asocia $P_{i,a,b}$.

2.1.1. Descripción informal del juego

Denotamos por S_t al estado del juego en el instante $t = 0, 1, 2 \ldots$; el estado inicial se fija de antemano, de modo que $S_0 = i_0$. En cada instante los jugadores participan del juego matricial R^{S_t}, correspondiente al estado en que se encuentra el juego estocástico. El estado siguiente del juego S_{t+1} es aleatorio y depende de S_t y de las acciones tomadas por los jugadores en el juego matricial del paso t, del siguiente modo

$$\mathbb{P}(S_{t+1} = j | S_t = i, A_t = a, B_t = b) = P_{i,a,b}(j).$$

El objetivo de cada jugador en el *juego estocástico* depende de lo que se quiera modelar, pero, para esta descripción informal, supongamos que cada jugador quiere maximizar la ganancia a lo largo de toda la historia, o sea, la suma de las ganancias obtenidas en cada uno de los juego matriciales jugados en cada instante. Entonces, una primera idea que surge, es la de jugar de forma óptima a cada juego matricial; sin embargo esta idea no es buena, ya que mira el problema localmente en el tiempo, sin tener en cuenta la influencia que tienen las decisiones en la dinámica del juego, *puede ser preferible sacrificar ganancia instantánea para ir a estados más favorables que incrementen la ganancia futura.*

Una estrategia para un jugador, intuitivamente, debería ser una forma de elegir en cada paso la acción a tomar en el juego matricial correspondiente a ese paso. La información con la que cuenta el jugador para tomar tal decisión es la historia del juego hasta ese momento. La definición formal de estrategia del *juego estocástico* resulta acorde a esta idea intuitiva.

2.1.2. Estrategias

Consideramos, para cada instante $t = 0, 1, 2, \ldots$ los conjuntos $\mathbb{H}_t$ de *historias posibles hasta el instante* t definidos por:

$$\begin{aligned} \mathbb{H}_0 &:= S \\ \mathbb{H}_1 &:= \mathbb{K} \times S \\ &\vdots \\ \mathbb{H}_t &:= \mathbb{K}^t \times S \\ &\vdots \end{aligned}$$

donde $\mathbb{K}^t$ es el producto cartesiano de $\mathbb{K}$ por sí mismo t veces. También definimos el conjunto $\mathbb{H} := \cup_{t=0}^{\infty} \mathbb{H}_t$.

Definición 2.1.2 (**Estrategia general**). *Una* ***estrategia general****, o simplemente estrategia,* π *para el jugador* 1 *es una correspondencia que a cada*

$$h = (i_0, a_0, b_0, \ldots, i_t) \in \mathbb{H}$$

asocia $\pi(\ |h)$*, una distribución de probabilidad sobre* A^{i_t}*, es decir una estrategia mixta del jugador 1 en el juego matricial* R^{i_t}*, según la definición 1.1.3.*

Análogamente, una estrategia φ *para el jugador 2, asocia a* h *una estrategia mixta del jugador 2 en el juego matricial* R^{i_t}*, que denotamos* $\varphi(\ |h)$*.*

En la bibliografía en inglés a estas estrategias se las denomina "behavior strategies".

Definición 2.1.3 (**Estrategia pura**). *Se llama* ***estrategia pura*** *para el jugador 1 a una estrategia general* π *para éste cuando la distribución de probabilidad* $\pi(\ |h)$ *acumula toda la probabilidad en un solo elemento, dicho de otra forma,* $\pi(\ |h)$ *es una estrategia pura del jugador 1 en el juego matricial* R^{i_t}*. Para el jugador 2 se define de forma análoga.*

En la sección 2.2.2 se presentan algunas subclases de estrategias que resultan de particular interés.

2.1.3. Definición formal del proceso asociado a un juego estocástico con estrategias fijas

Dado un juego estocástico, un estado inicial i, y estrategias π y φ de los jugadores queremos definir un espacio de probabilidad y un proceso estocástico que modele el transcurso del juego. Entendemos por *transcurso del juego* la secuencia de estados por la que éste pasó y las acciones tomadas por los jugadores.

Consideramos los conjuntos $\mathbb{A}$ y $\mathbb{B}$, de todas las acciones posibles para el jugador 1 y 2 respectivamente, mediante

$$\mathbb{A} = \bigcup_{i=1}^{N} A^i \quad \text{y} \quad \mathbb{B} = \bigcup_{i=1}^{N} B^i.$$

Sea $\Omega_1 = S$, $\Omega_2 = \mathbb{A} \times \mathbb{B}$, $\Omega_3 = S$, $\Omega_4 = \mathbb{A} \times \mathbb{B}$, y así sucesivamente. El espacio muestral en que queremos definir una probabilidad es

$$\Omega = \prod_{k=1}^{\infty} \Omega_k$$

ya que cada elemento $(i_0, (a_0, b_0), i_1, (a_1, b_1), \ldots) \in \Omega$ representa una historia, de estados y acciones tomadas por los jugadores, por la que pasó el juego, de modo que i_t es el estado en que se encontraba el juego en el instante t y (a_t, b_t) las acciones tomadas por los jugadores en dicho instante. Notar que para todo $t = 0, 1, 2, \ldots$

$$\mathbb{H}_t \subseteq \prod_{k=1}^{2t+1} \Omega_k$$

Como los conjuntos Ω_k son de cardinal finito podemos considerar la σ-álgebra de partes sobre ellos, $\mathcal{F}_k = \mathcal{P}(\Omega_k)$. Para cada $k = 1, 2, \ldots$ y para cada $(\omega_1, \ldots, \omega_{k-1})$ la dinámica del juego define la siguiente probabilidad sobre $(\Omega_k, \mathcal{F}_k)$:

- si $k = 1$,

$$\mathbb{P}(\omega_1) = \begin{cases} 1 & \text{si } \omega_1 = i \\ 0 & \text{en otro caso} \end{cases}$$

- si k es par, y $(\omega_1, \ldots, \omega_{k-1}) \in \mathbb{H}_{\frac{k}{2}-1}$

$$\mathbb{P}(\omega_k) = \begin{cases} \pi(a|\omega_1, \ldots, \omega_{k-1})\varphi(b|\omega_1, \ldots, \omega_{k-1}) & \text{si } \omega_k = (a, b) \\ & (a, b) \in A^{\omega_{k-1}} \times B^{\omega_{k-1}} \\ 0 & \text{en otro caso} \end{cases}$$

- si k es par, y $(\omega_1, \ldots, \omega_{k-1}) \notin \mathbb{H}_{\frac{k}{2}-1}$ no importa la distribución de probabilidad ya que estamos sobre un conjunto de probabilidad 0, por lo tanto la podríamos definir de cualquier manera.
- si k es impar diferente de 1, y $\omega_{k-1} = (a, b)$ con $(a, b) \in A^{\omega_{k-2}} \times B^{\omega_{k-2}}$

$$\mathbb{P}(\omega_k) = P_{\omega_{k-2},a,b}$$

- si k es impar diferente de 1, y $\omega_{k-1} = (a, b)$ con $(a, b) \notin A^{\omega_{k-2}} \times B^{\omega_{k-2}}$ no importa la probabilidad que se defina.

Con las probabilidades anteriores estamos en las hipótesis del teorema de Ionescu Tulcea, teorema 2 de la sección 9 del capítulo 2 de [25]. Llamamos $\mathbb{P}_{i,\pi,\varphi}$ a la probabilidad que surge en el teorema, definida en $(\Omega, \mathcal{F})$, donde $\mathcal{F}$ es la σ-álgebra de los cilindros.

Sobre el espacio de probabilidad definido $(\Omega, \mathcal{F}, \mathbb{P}_{i,\pi,\varphi})$ consideramos los procesos estocásticos

$$\{S_t\}_{t=0,1,\ldots},\ \{A_t\}_{t=0,1,\ldots},\ \{B_t\}_{t=0,1,\ldots}$$

que representan el estado, y las acciones de los jugadores 1 y 2 respectivamente en el instante t; definidos como uno espera, si $\omega = (i_0, (a_0, b_0), i_1, (a_1, b_1), \ldots) \in \Omega$ entonces

$$S_t(\omega) = i_t,\ , A_t(\omega) = a_t,\ B_t(\omega) = b_t.$$

También consideramos los elementos aleatorios $H_t : t = 0, 1, \ldots$, de la historia del proceso hasta el instante t, de modo que H_t toma valores en $\prod_{k=1}^{2t+1} \Omega_k$ y

$$H_t(\omega) = (i_0, a_0, b_0, i_1, a_1, b_1, \ldots, i_t).$$

Por como fue definida la probabilidad $\mathbb{P}_{i,\pi,\varphi}$ es muy fácil verificar que:

- el juego comienza en el estado i, es decir

$$\mathbb{P}_{i,\pi,\varphi}(S_0 = i) = 1;$$

- las variables aleatorias H_t toman valores en $\mathbb{H}_t$ con probabilidad 1.
- las acciones elegidas por los jugadores dependen de las estrategias y de la historia del juego y son independientes, de modo que

$$\mathbb{P}_{i,\pi,\varphi}(A_t = a_t | H_t = h_t) = \pi(a_t|h_t)$$

$$\mathbb{P}_{i,\pi,\varphi}(B_t = b_t | H_t = h_t) = \varphi(b_t|h_t)$$

$$\mathbb{P}_{i,\pi,\varphi}(A_t = a_t, B_t = b_t | H_t = h_t) = \pi(a_t|h_t)\varphi(b_t|h_t).$$

- las transiciones entre estados dependen únicamente del último estado y de las acciones

$$\mathbb{P}_{i,\pi,\varphi}(S_{t+1} = i_{t+1}|H_t = h_t, A_t = a_t, B_t = b_t) = P_{i_t,a_t,b_t}(i_{t+1}),$$

donde i_t es el último estado de la historia h_t.

Observación 2.1.4. *En muchas aplicaciones podría quererse que el estado inicial del juego sea aleatorio (un claro ejemplo de esto es cuando en los juegos de naipes se reparten las cartas); el hecho de tener un estado inicial no quita generalidad en este sentido, ya que uno podría definir un estado ficticio i_0 en que los jugadores no tengan acciones y la distribución del estado siguiente sea la distribución inicial que se quiere tener.*

2.2. El problema asociado a un juego estocástico

2.2.1. Criterios de optimalidad

Definición 2.2.1 (**Criterio de optimalidad**). *Dado un juego estocástico como el definido en 2.1.1. Definimos **criterio de optimalidad** como una función que a cada par de estrategias π y φ de los jugadores 1 y 2 respectivamente, y a cada jugador $j = 1, 2$ asocia un vector $v^{\pi,\varphi}(j) \in \mathbb{R}^N$, donde el número $v^{\pi,\varphi}(j)_i$ se denomina "valor del par de estrategias π y φ para el jugador j si el juego comienza en estado i"*

El problema que típicamente se asocia a un juego estocástico es el de, dado un criterio de optimalidad, hallar estrategias que optimicen dicho criterio. En el caso competitivo, que es el que nos interesa en este trabajo, cada jugador $j = 1, 2$ trata de maximizar el valor $v^{\pi,\varphi}(j)_i$; pero dicha ganancia depende tanto de su estrategia como la de su contrincante.

El criterio de optimalidad depende de la naturaleza del problema que se quiera modelar con el *juego estocástico*. A continuación mencionamos algunos de los criterios más comúnmente usados.

Suma descontada

El caso más estudiado, inspirado en modelos económicos, es el de la suma descontada por un factor $\beta \in (0, 1)$. En este caso, el valor del par de estrategias π y φ para el jugador j si el juego comienza en estado i es

$$v^{\pi,\varphi}(j)_i = \sum_{t=0}^{\infty} \beta^t \mathbb{E}_{i,\pi,\varphi} r_j^{S_t}(A_t, B_t),$$

donde $\mathbb{E}_{i,\pi,\varphi}$ es el valor esperado cuando la distribución de probabilidad es $\mathbb{P}_{i,\pi,\varphi}$. Para comprender el factor de descuento puede pensarse que la unidad de tiempo es un mes y que el dinero a medida que se gana se pone en un banco a

una tasa de interés mensual de α. Entonces ganar \$10 en el instante t, genera el mismo beneficio que ganar $10(1+\alpha)$ en el instante $t+1$, de modo que cuanto antes se tenga el dinero mejor es. De la descripción mencionada surge que $\beta = \frac{1}{1+\alpha}$.

El factor β además de cumplir el rol teórico de representar el hecho de que el dinero genera interés, cumple una función práctica muy importante asegurando que cada entrada del vector $v^{\pi,\varphi}(j)$ sea finita independientemente del juego estocástico y de las estrategias:

$$\begin{aligned} |v^{\pi,\varphi}(j)_i| &\leq \sum_{t=0}^{\infty} \beta^t \mathbb{E}_{i,\pi,\varphi} |r_j^{S_t}(A_t, B_t)| \\ &\leq \sum_{t=0}^{\infty} \beta^t \underbrace{\max_{i,a,b} |r_j^i(a,b)|}_{k} \\ &= \frac{k}{1-\beta}. \end{aligned}$$

Promedio

Otro criterio de optimalidad muy usado es la ganancia promedio, es decir,

$$v^{\pi,\varphi}(j)_i = \lim_{n\to\infty} \frac{\sum_{t=0}^{n-1} \mathbb{E}_{i,\pi,\varphi} r_j^{S_t}(A_t, B_t)}{n}.$$

De este modo también se asegura la finitud de V_i sin necesidad de un modelo con descuento. Este criterio de optimalidad nace en 1957, con el artículo [11] de D. Gillette.

Suma con horizonte finito

En algunos casos solo importa el seguimiento del juego en una cantidad finita de pasos, y la función a optimizar es

$$v^{\pi,\varphi}(j)_i = \sum_{t=0}^{N} \mathbb{E}_{i,\pi,\varphi} r_j^{S_t}(A_t, B_t).$$

Suma total con horizonte infinito

Por último aparece el modelo que profundizamos en este trabajo, que es cuando la función a optimizar es sencillamente la suma de las ganancias a lo largo de todos los pasos del juego. Es decir

$$v^{\pi,\varphi}(j)_i = \sum_{t=0}^{\infty} \mathbb{E}_{i,\pi,\varphi} r_j^{S_t}(A_t, B_t).$$

En este caso sí existe el problema de que $v^{\pi,\varphi}(j)_i$ podría ser infinito para algún estado inicial i, en cuyo caso no cumpliría la definición de criterio. Sin embargo hay ciertos juegos estocásticos para los que se cumple que $v^{\pi,\varphi}(j)_i$ es finito independientemente de las estrategias y del estado inicial.

Definición 2.2.2 (**Juego estocástico sumable**). *Un **juego estocástico sumable** es un juego estocástico para el cual el criterio de la suma con horizonte infinito está bien definido.*

2.2.2. Subclases de estrategias

Definición 2.2.3 (**Estrategia semimarkoviana**). *Una estrategia general π es semimarkoviana si cumple que a historias hasta el instante t, que tienen el mismo estado inicial y el mismo estado final, les asocia la misma distribución de probabilidad. Es decir, si $h = (i_0, a_0, b_0, \ldots, i_t)$ y $h' = (i'_0, a'_0, b'_0, \ldots, i'_t)$ se cumple que*

$$i_0 = i'_0,\ i_t = i'_t \Rightarrow \forall i \in S\ \pi(i|h) = \pi(i|h')$$

En caso de tener una estrategia semimarkoviana se suele escribir $\pi(\ |i_t, t, i_0)$ indicando que la distribución depende únicamente de t, i_t, e i_0.

Definición 2.2.4 (**Estrategia markoviana**). *Una estrategia general π es markoviana si cumple que a historias hasta el instante t que tienen el mismo estado final les asocia la misma distribución de probabilidad. Es decir, si $h = (i_0, a_0, b_0, \ldots, i_t)$ y $h' = (i'_0, a'_0, b'_0, \ldots, i'_t)$ se cumple que*

$$i_t = i'_t \Rightarrow \forall i \in S\ \pi(i|h) = \pi(i|h')$$

Siguiendo el mismo criterio que en el caso de las estrategias semimarkovianas, en este caso $\pi(\ |h)$ se abrevia como $\pi(\ |i_t, t)$.

Definición 2.2.5 (**Estrategia estacionaria**). *Una estrategia general π es estacionaria si cumple que a historias con el mismo estado final asocia la misma distribución de probabilidad. Es decir, si $h = (i_0, a_0, b_0, \ldots, i_t)$ y $h' = (i'_0, a'_0, b'_0, \ldots, i'_{t'})$ se cumple que*

$$i_t = i'_{t'} \Rightarrow \forall i \in S\ \pi(i|h) = \pi(i|h')$$

En este caso la distribución con que se elige la acción depende únicamente del estado actual i_t, por lo que en lugar de $\pi(\ |h)$ basta escribir $\pi(\ |i_t)$.

Observación 2.2.6. *Se cumple que toda estrategia estacionaria es markoviana, toda estrategia markoviana es semimarkoviana y toda estrategia semimarkoviana es general.*

Definición 2.2.7 (**Regla de decisión**). *Dado un juego estocástico, una **regla de decisión f para el jugador 1** es una función que a cada estado $i \in S$ asocia una una estrategia f_i del jugador 1 para el juego matricial R^i, es decir f_i*

es una distribución de probabilidad sobre A^i. De forma análoga, una ***regla de decisión*** *g* ***para el jugador 2*** *es una función que a cada estado $i \in S$ asocia una distribución de probabilidad g_i sobre B^i.*

Observación 2.2.8. *Las diferentes clases de estrategias se relacionan con las reglas de decisión de la siguiente manera:*

- *una estrategia estacionaria π se corresponde con una regla de decisión f, de modo que $f_i = \pi(\ |i)$, a veces nos referimos a la estrategia estacionaria como f, la regla de decisión asociada;*
- *una estrategia markoviana π se corresponde con una sucesión de reglas de decisión $(f_0, f_1, \ldots)$, donde $f_{t,i} = \pi(\ |i, t)$;*
- *una estrategia semimarkoviana π se corresponde con una sucesión de reglas de decisión $(f_0^{i_0}, f_1^{i_0}, \ldots)$ para cada estado inicial i_0;*
- *a una estrategia general π se asocia una correspondencia que a cada historia hasta el instante $t-1$ y a cada par de acciones a_{t-1}, b_{t-1} le asocia una regla de decisión $f^{h_{t-1},a_{t-1},b_{t-1}}$.*

Dado un juego estocástico, la forma en que se toman las acciones en el paso t están dadas por reglas de decisión f y g que, en general, dependen de la historia del juego hasta ese momento, con esto queda definida la probabilidad de transición de un paso, de modo que la probabilidad de que $S_{t+1} = j$ dado que $S_t = i$ está dada por la siguiente fórmula:

$$\mathbb{P}(S_{t+1} = j | S_t = i) = \sum_{a \in A^i} \sum_{b \in B^i} f_i(a) g_i(b) P_{i,a,b}(j); \tag{2.2}$$

Estas probabilidades se representan en una matriz $P(f, g)$, tal que $P(f, g)_{i,j} = \mathbb{P}(S_{t+1} = j | S_t = i)$.

2.2.3. Equilibrio de Nash

El concepto de equilibrio de Nash, visto en el capítulo 1 para los juegos matriciales, se puede definir fácilmente en el contexto de los juegos estocásticos, y de los juegos en general.

Definición 2.2.9 (**Equilibrio de Nash para juegos estocásticos**)**.** *Dado un juego estocástico, un criterio de optimalidad, y familias de estrategias Π y Φ para los jugadores, decimos que el par de estrategias π^*, φ^* forma un* ***equilibrio de Nash*** *en las familias de estrategias mencionadas si se cumple para todo estado inicial i*

$$\forall \pi \in \Pi,\ v^{\pi,\varphi^*}(1)_i \leq v^{\pi^*,\varphi^*}(1)_i$$

y

$$\forall \varphi \in \Phi,\ v^{\pi^*,\varphi}(2)_i \leq v^{\pi^*,\varphi^*}(2)_i,$$

es decir, ningún jugador puede mejorar su desempeño cambiando su estrategia unilateralmente.

2.3. Juegos estocásticos de suma cero

En esta sección definimos los juegos estocásticos de suma cero y vemos que algunos conceptos definidos para el caso de juegos matriciales de suma cero, tratados en la sección 1.2, se extienden para este caso.

Definición 2.3.1 (**Juego estocástico de suma cero**). *Un juego estocástico se dice de suma cero cuando la suma de los pagos de cada jugador es cero. O dicho de otra manera, para todo* $(i, a, b) \in \mathbb{K}$

$$r_1^i(a, b) = -r_2^i(a, b).$$

En este caso no utilizamos subíndices para referirnos a la función de pago, sino que denotamos por $r^i(a, b)$ *a la ganancia del jugador 1, y el opuesto es la ganancia del jugador 2.*

Para cualquiera de los criterios de optimalidad mencionados como ejemplo en la sección 2.2.1, se cumple que si el juego es de suma cero entonces $v^{\pi,\varphi}(1) = -v^{\pi,\varphi}(2)$; llamamos $v^{\pi,\varphi}$ a ese vector y lo denominamos "valor del juego para las estrategias (π, φ)".

Definición 2.3.2 (**Valor del juego estocástico**). *Fijado un juego estocástico de suma cero, un criterio de optimalidad* $v^{\pi,\varphi}$*, y familias de estrategias posibles* Π *y* Φ *para ambos jugadores. Si se cumple*

$$\sup_{\pi\in\Pi} \inf_{\varphi\in\Phi} v_i^{\pi,\varphi} = \inf_{\varphi\in\Phi} \sup_{\pi\in\Pi} v_i^{\pi,\varphi},$$

para todo estado inicial i; decimos que el juego tiene un valor con respecto al criterio $v^{\pi,\varphi}$ *en la familia de estrategias mencionadas y llamamos* $v^* \in \mathbb{R}^N$ *a ese valor. Si no se aclara la familia de estrategias posibles se asume que es la de las estrategias generales.*

Para el criterio de optimalidad de la suma descontada, y de la suma con horizonte finito siempre existe el valor del juego si se consideran las estrategias generales. Esto no ocurre necesariamente con el criterio del promedio.

Definición 2.3.3 (**Estrategia óptima**). *Dado un juego estocástico de suma cero, un criterio de optimalidad* $v^{\pi,\varphi}$*, y familias de estrategias posibles* Π *y* Φ *para ambos jugadores, decimos que una estrategia* $\pi^* \in \Pi$ *es óptima para el jugador 1 si para todo estado inicial i se cumple*

$$\inf_{\varphi\in\Phi} v_i^{\pi^*,\varphi} = \sup_{\pi\in\Pi} \inf_{\varphi\in\Phi} v_i^{\pi,\varphi}$$

y análogamente, la estrategia φ^* *es óptima para el jugador 2 si para todo i se cumple*

$$\sup_{\pi\in\Pi} v_i^{\pi,\varphi^*} = \inf_{\varphi\in\Phi} \sup_{\pi\in\Pi} v_i^{\pi,\varphi}$$

Lema 2.3.4. *Se cumple que*

$$\sup_{\pi\in\Pi}\inf_{\varphi\in\Phi} v_i^{\pi,\varphi} \leq \inf_{\varphi\in\Phi}\sup_{\pi\in\Pi} v_i^{\pi,\varphi}$$

Demostración. La demostración de este lema es análoga a la del lema 1.2.5 sobre juegos matriciales, cambiando máximo por supremo y mínimo por ínfimo:

Sabemos que

$$v_i^{\pi,\varphi} \geq \inf_{\varphi'\in\Phi} v_i^{\pi,\varphi'},$$

tomando supremo en π obtenemos

$$\sup_{\pi\in\Pi} v_i^{\pi,\varphi} \geq \sup_{\pi\in\Pi}\inf_{\varphi'\in\Phi} v_i^{\pi,\varphi'}$$

y como la desigualdad anterior vale para todo φ, tomando ínfimo en φ se obtiene el resultado buscado.

□

Teorema 2.3.5. *Dado un juego estocástico de suma cero, un criterio de optimalidad $v^{\pi,\varphi}$, y familias de estrategias posibles Π y Φ para ambos jugadores, se cumple que si el par de estrategias (π^*,φ^*) forma un equilibrio de Nash entonces el juego tiene un valor $v^* = v^{\pi^*,\varphi^*}$ y las estrategias son óptimas.*

Demostración. Cualquiera sea el estado inicial i, si π^* y φ^* forman un equilibrio de Nash, para todo $\pi \in \Pi$ se cumple

$$v_i^{\pi^*,\varphi^*} \geq v_i^{\pi,\varphi^*},$$

por lo tanto

$$\begin{aligned} v_i^{\pi^*,\varphi^*} &= \sup_{\pi\in\Pi} v_i^{\pi,\varphi^*} \\ &\geq \inf_{\varphi\in\Phi}\sup_{\pi\in\Pi} v_i^{\pi,\varphi}; \end{aligned}$$

igualmente

$$\begin{aligned} v_i^{\pi^*,\varphi^*} &= \inf_{\varphi\in\Phi} v_i^{\pi^*,\varphi} \\ &\leq \sup_{\pi\in\Pi}\inf_{\varphi\in\Phi} v_i^{\pi,\varphi}; \end{aligned}$$

teniendo en cuenta que la desigualdad dada por el lema anterior, probamos tanto la existencia del valor como la optimalidad de las estrategias.

□

Capítulo 3

Juegos estocásticos transitorios

Introducción

A priori podríamos decir que un *juego estocástico transitorio* es un *juego estocástico* que a partir de un momento pasa a un estado especial en que se considera el juego terminado. Es común en la bibliografía no considerar el estado especial y directamente pedir que las probabilidades de transición a los diferentes estados no sumen 1; de modo que si no sale "sorteado" ningún estado se considera que el juego terminó; de hecho, en el artículo de Shapley [24] que crea los *juegos estocásticos* se considera de este modo.

Si bien esta clase de juegos no es más que una subclase de los juegos estocásticos, tratados en el capítulo anterior, le dedicamos un capítulo por tratarse del modelo en el que nos interesa profundizar para resolver las aplicaciones que motivan este trabajo.

En este capítulo abordamos el problema asociado a un *juego estocástico transitorio de suma cero*. Concretamente se demuestra que los juegos de esta clase siempre tienen un valor (ver definición 2.3.2) y que siempre hay estrategias óptimas que resultan ser estacionarias; dichas estrategias se componen de estrategias mixtas óptimas para ciertos juegos matriciales, que para hallarlos hay que resolver ecuaciones análogas a las de Bellman para *procesos de control de Markov*; en este caso competitivo en lugar del máximo o mínimo de la ecuación de Bellman aparece el valor de un juego matricial.

3.1. Condición de transitoriedad

Para hablar de *juegos estocásticos transitorios* consideramos un juego estocástico, en que al estado N le asignamos una interpretación especial, significa que el juego terminó; es decir:

1. el estado N es absorbente, o sea, para todo par de acciones (a, b) de los jugadores se cumple
$$P_{N,a,b}(N) = 1; \text{ y}$$
2. la ganancia es cero si el juego se encuentra en ese estado, sin importar las acciones tomadas por los jugadores,
$$r_1^N(a, b) = r_2^N(a, b) = 0.$$

Una vez que el juego entra en el estado N las acciones que tomen los jugadores dejan de afectar la dinámica; ya que con probabilidad uno el juego se mantendrá en ese estado y la ganancia de los jugadores de ahí en más es cero. Obviamente el hecho de haberle asignado al estado N, y no a otro, el carácter de estado final no es ninguna restricción, sino una convención para simplificar la notación.

Dados un juego estocástico (definido en 2.1.1), en el que el estado N cumple con las condiciones mencionadas, un par de estrategias π y φ de los jugadores, denominamos *condición de transitoriedad* a la siguiente condición:

$$\forall i \in S, \quad \sum_{t=0}^{\infty} \mathbb{P}_{i,\pi,\varphi}(S_t \neq N) < \infty; \tag{3.1}$$

que se puede reescribir así:

$$\forall i \in S, \quad \sum_{t=0}^{\infty} \sum_{j=1}^{N-1} \mathbb{P}_{i,\pi,\varphi}(S_t = j) < \infty$$

Teorema 3.1.1. *La condición de transitoriedad asegura que la suma total con horizonte infinito resulte finita, es decir, para todo estado inicial $i \in S$ y $j = 1, 2$ se cumple:*

$$v^{\pi,\varphi}(j)_i = \sum_{t=0}^{\infty} \mathbb{E}_{i,\pi,\varphi} r_j^{S_t}(A_t, B_t) < \infty,$$

Demostración. Basta observar que, cualquiera sea el estado inicial i,

$$\begin{aligned}
|\mathbb{E}_{i,\pi,\varphi} r_j^{S_t}(A_t, B_t)| &= \left| \sum_{k \in S} \sum_{(a,b) \in A^k \times B^k} r_j^k(a, b) \mathbb{P}_{i,\pi,\varphi}(S_t = k, A_t = a, B_t = b) \right| \\
&= \left| \sum_{k=1}^{N-1} \sum_{(a,b) \in A^k \times B^k} r_j^k(a, b) \mathbb{P}_{i,\pi,\varphi}(S_t = k, A_t = a, B_t = b) \right| \\
&\leq K \mathbb{P}_{i,\pi,\varphi}(S_t \neq N)
\end{aligned}$$

donde $K = \text{máx}_{k\in S, a\in A^k, b\in B^k} |r_j^k(a,b)|$. Entonces

$$\sum_{t=0}^{\infty} |\mathbb{E}_{i,\pi,\varphi} r_j^{S_t}(A_t, B_t)| \leq K \sum_{t=0}^{\infty} \mathbb{P}_{i,\pi,\varphi}(S_t \neq N) < \infty$$

por la condición de transitoriedad. Por lo tanto la serie converge, ya que converge absolutamente. □

3.2. Un juego estocástico sumable

Definición 3.2.1 (**Juego estocástico transitorio**). *Decimos que un* ***juego estocástico es transitorio*** *cuando se cumple la condición* (3.1) *de transitoriedad para todo par de estrategias generales.*

Observación 3.2.2. *De la definición anterior y la observación 3.1.1 se deduce inmediatamente que un juego estocástico transitorio es sumable.*

El criterio de optimalidad considerado para los juegos transitorios es el de la suma total con horizonte infinito, tratado el la sección 2.2.1; de modo que de aquí en más cuando hagamos referencia a $v^{\pi,\varphi}(j)$ nos referimos a

$$v^{\pi,\varphi}(j)_i = \sum_{t=0}^{\infty} \mathbb{E}_{i,\pi,\varphi} r_j^{S_t}(A_t, B_t).$$

El hecho de tener un juego estocástico sumable nos asegura que tal criterio está bien definido.

Teorema 3.2.3. *Dado un juego estocástico transitorio, si denominamos τ al tiempo de espera hasta que S_t valga N, o sea*

$$\tau = \text{ínf}\{t|\ S_t = N\}$$

o $\tau = \infty$ si $\{t|\ S_t = N\}$ es vacío, se cumple que $\mathbb{P}_{i,\pi,\varphi}(\tau < \infty) = 1$ para cualquier par de estrategias π, φ de los jugadores 1 y 2 respectivamente y cualquiera sea el estado inicial i.

Demostración. Consideremos el juego dado, cambiando la función de pago de modo que $r_j^K(a,b) = 1$ si $k \neq N$ y $r^N(a,b) = 0$. El nuevo juego estocástico sigue siendo transitorio, ya que la transitoriedad del juego no depende en absoluto de la función de pago. Ahora bien, con este nuevo juego si $\omega \in \{\tau = \infty\}$ se cumple

$$\sum_{t=0}^{\infty} r_j^{S_t(\omega)}(A_t(\omega), B_t(\omega)) = \infty,$$

de modo que si para un par de estrategias (π, φ) se cumpliera que τ toma el valor infinito con probabilidad positiva $\mathbb{P}_{i,\pi,\varphi}(\{\tau = \infty\}) > 0$ entonces se tendría que

$$\sum_{t=0}^{\infty} \mathbb{E}_{i,\pi,\varphi} r_j^{S_t}(A_t, B_t) = \infty$$

para algún estado inicial i. Lo que es absurdo por la observación 3.1.1.

□

El resultado anterior da una idea intuitiva de qué casos se pueden modelar con un juego estocástico transitorio: problemas en que si bien la cantidad de pasos no se puede acotar a priori se cumple que es finita con probabilidad 1.

Como la definición de *juego estocástico transitorio* requiere que se cumpla la condición de transitoriedad para una cantidad infinita de estrategias, en la práctica, resulta difícil verificar si se está en ese caso; más adelante vemos que para que un juego estocástico sea transitorio basta pedir que se cumpla la condición de transitoriedad solo para las estrategias estacionarias puras, que son una cantidad finita.

Teorema 3.2.4. *Dado un juego estocástico transitorio, se cumple que para cualquier par de estrategias generales π, φ y cualquier estado inicial $i \in S$,*

$$\mathbb{P}_{i,\pi,\varphi}(S_N = N) > 0$$

Demostración. Definimos, para cada instante t, el conjuntos R_t de los estados alcanzables en t pasos. Es decir

$$R_t := \{k \in S,\ \mathbb{P}_{i,\pi,\varphi}(S_t = k) > 0)\}.$$

Queremos demostrar que $N \in R_N$, aunque es claro que alcanza con probar $N \in R_t$, para algún $t < N$, ya que el estado N es absorbente. **Afirmación:** para todo n se cumple que, o bien $N \in R_n$, o $R_{n+1} \nsubseteq \cup_{t=0}^{n} R_t$; de modo que si $N \notin R_n$ se cumple

$$\# \cup_{t=0}^{n+1} R_t > \# \cup_{t=0}^{n} R_t$$

y como la cantidad total de estados es N no se puede cumplir $R_{n+1} \nsubseteq \cup_{t=0}^{n} R_t$ más de N veces, de modo que $N \in R_N$. Probemos ahora la **afirmación**: supongamos, por absurdo, que existen un estado inicial i, estrategias π y φ, y un instante n de modo que

$$N \notin R_n \text{ y } R_{n+1} \subseteq \cup_{t=0}^{n} R_t.$$

Entonces si para cada estado $k \in \cup_{t=0}^{n} R_t$ se usan las reglas de decisión que determinarían el par de estrategias π y φ tendríamos que, con probabilidad uno, el estado siguiente pertenece al conjunto $\cup_{t=0}^{n} R_t$; consideremos un par de

estrategias f y g estacionarias que a cada uno de los estados k asocie dicha regla. Es claro que con ese par de estrategias, con probabilidad uno, el juego se mantiene siempre en el conjunto $\cup_{t=0}^{n} R_t$, de modo que

$$\mathbb{P}_{i,f,g}(S_t \neq N) = 1,$$

lo que determina que no se cumpla la condición de transitoriedad.

□

3.3. Caso de suma cero

Si bien definimos *juegos estocásticos transitorios* en general, el caso en que nos interesa profundizar es el de *suma cero*.

3.3.1. Optimización en estrategias estacionarias

El teorema siguiente demuestra la existencia del valor de un *juego estocástico transitorio de suma cero* entre las estrategias estacionarias e indica cómo se calculan las estrategias óptimas. Más adelante, en el teorema 3.3.4 y en el corolario 3.3.8, vemos que las estrategias que se definen en el teorema que sigue son óptimas entre todas las estrategias, que es el principal objetivo del capítulo.

Teorema 3.3.1 (**Existencia del valor entre estrategias estacionarias**)**.** *Todo juego estocástico transitorio y de suma cero tiene un valor v^* con respecto a las estrategias estacionarias. O sea que si consideramos un juego estocástico transitorio de suma cero y Π, Φ los conjuntos de estrategias estacionarias de los jugadores 1 y 2 respectivamente, se cumple que para todo estado inicial $i \in S$*

$$\sup_{\pi\in\Pi} \operatorname{ínf}_{\varphi\in\Phi} v_i^{\pi,\varphi} = \operatorname{ínf}_{\varphi\in\Phi} \sup_{\pi\in\Pi} v_i^{\pi,\varphi} = v_i^*;$$

además v^* es el único vector en $\mathbb{R}^N$ que cumple: $v_N^* = 0$ y

$$v_i^* = \operatorname{val}\left[r^i(a,b) + \sum_{k\in S} P_{i,a,b}(k) v_k^* \right]_{a\in A^i,\ b\in B^i}, \tag{3.2}$$

donde $\operatorname{val}[f(a,b)]_{a\in A^i,\ b\in B^i}$ denota el valor del juego matricial de matriz con filas $a \in A^i$, columnas $b \in B^i$ y entradas $f(a,b)$, que sabemos que existe por el teorema minimax de von Neumann para juegos matriciales 1.2.7.

Además ambos jugadores tienen estrategias (estacionarias) óptimas. La estrategia óptima del jugador 1 es f^*, tal que f_i^* es una estrategia óptima para el jugador 1 en el juego matricial

$$\left[r^i(a,b) + \sum_{k\in S} P_{i,a,b}(k) v_k^* \right]_{a\in A^i,\ b\in B^i};$$

análogamente la estrategia óptima g^* del jugador 2 se compone de las estrategias óptimas para dicho jugador en el mismo juego matricial.

Antes de pasar a la demostración del teorema necesitamos algunas notaciones:

- Denotamos por $\mathbb{R}_0^N$ al conjunto de vectores de $\mathbb{R}^N$ con última coordenada 0.
- Para cada par (f, g) de reglas de decisión (ver definición 2.2.7) de los jugadores, consideramos el vector $r(f, g) \in \mathbb{R}_0^N$ de modo que la entrada i-ésima sea $r^i(f_i, g_i)$, el valor esperado del pago (definido en la ecuación (1.4)) en el juego matricial R^i cuando se utilizan las estrategias f_i, g_i, de modo que

$$r(f, g)_i = \sum_{a \in A^i} \sum_{b \in B^i} r^i(a, b) f_i(a) g_i(b).$$

- Como fue visto en el capítulo 2, fijadas las reglas de decisión (f, g), que usarán los jugadores en el instante de tiempo t, podemos escribir las probabilidades de transición en una matriz $P(f, g)$.

Lema 3.3.2. *Si fijamos un juego estocástico transitorio de suma cero con estrategias estacionarias f y g para el jugador 1 y 2 respectivamente, se cumple la siguiente igualdad:*

$$v^{f,g} = \sum_{t=0}^{\infty} P(f, g)^t r(f, g),$$

donde $v^{f,g}$ es el vector de valor del juego para las estrategias f y g, definido en el capítulo 2. Por la característica especial del estado N, tanto $v^{f,g}$ como $r(f, g)$ son vectores $\mathbb{R}_0^N$, y los estamos utilizando como vectores columna.

Demostración. Si reescribimos la definición de la "suma total con horizonte infinito", utilizando las estrategias estacionarias f y g, tenemos que

$$v_i^{f,g} = \sum_{t=0}^{\infty} \mathbb{E}_{i,f,g} r^{S_t}(A_t, B_t). \tag{3.3}$$

Por la estacionareidad de las estrategias, las probabilidades de transición entre los estados del juego están dadas por la matriz $P(f, g)$, independientemente de t. De modo que las potencias de la matriz dan la probabilidad de transición en varios pasos, obteniendo que

$$\mathbb{P}_{i,f,g}(S_t = j) = (P(f, g)^t)_{i,j}.$$

Utilizando la notación introducida para representar el valor esperado del pago instantáneo cuando se está en el estado j y se juega con las estrategias estacionarias f y g, podemos escribir un sumando de (3.3) como sigue:

$$\mathbb{E}_{i,f,g} r^{S_t}(A_t, B_t) = \sum_{j=1}^{N} (P(f,g)^t)_{i,j} r(f,g)_j,$$

que no es otra cosa que la i-ésima entrada del vector $P(f,g)^t r(f,g)$. Al considerar la suma en t se obtiene lo que queremos.

□

Lema 3.3.3. *Si v es un vector de $\mathbb{R}_0^N$ y f, g son estrategias estacionarias para un juego estocástico transitorio de suma cero tal que*

$$v \leq r(f,g) + P(f,g)v, \tag{3.4}$$

considerando la desigualdad coordenada a coordenada, entonces se cumple

$$v \leq v^{f,g},$$

siendo $v^{f,g}$ el valor del juego.

Demostración. Si en la parte derecha de la inecuación (3.4) sustituimos v utilizando la propia inecuación obtenemos

$$v \leq r(f,g) + P(f,g)r(f,g) + P(f,g)^2 v,$$

si volvemos a hacer lo mismo, al cabo de k veces obtenemos

$$v \leq \sum_{t=0}^{k} P(f,g)^t r(f,g) + P(f,g)^{k+1} v.$$

Veamos que al hacer tender k a infinito, la parte derecha de la inecuación tiende a $v^{f,g}$, lo que culminaría la prueba. Sabemos, por el lema anterior, que $\sum_{t=0}^{k} P(f,g)^t r(f,g)$ tiende a $v^{f,g}$, por lo que bastaría ver que $P(f,g)^{k+1}v$ tiende a cero. Veamos qué pasa con el valor absoluto del elemento i-ésimo, con $i = 1, \ldots, N-1$, de $P(f,g)^{k+1}v$:

$$\begin{aligned}
|(P(f,g)^{k+1}v)_i| &\leq \sum_{j=1}^{N} \mathbb{P}_{i,f,g}(S_{k+1} = j)|v_j| \\
&= \sum_{j=1}^{N-1} \mathbb{P}_{i,f,g}(S_{k+1} = j)|v_j| \\
&\leq \sum_{j=1}^{N-1} \mathbb{P}_{i,f,g}(S_{k+1} = j)||v||_\infty \\
&= \mathbb{P}_{i,f,g}(S_{k+1} \neq N)||v||_\infty
\end{aligned}$$

donde la primera igualdad se debe a que $v_N = 0$. Por definición de juego estocástico transitorio sabemos que $\mathbb{P}_{i,f,g}(S_{k+1} \neq N)$ tiende a cero, ya que es el término general de una serie convergente, lo que culmina la demostración.

□

Demostración del teorema 3.3.1.

La prueba se divide en dos partes: primero probamos que existe un único vector $v^* \in \mathbb{R}_0^N$ que cumple

$$v_i^* = \text{val}\left[r^i(a,b) + \sum_{j\in S} P_{i,a,b}(j)v_j^*\right]_{a\in A^i,\ b\in B^i} \tag{3.5}$$

para todo $i = 1, \ldots, N$; y luego vemos que las estrategias f^* y g^* definidas en el enunciado son estrategias óptimas y el valor del juego con esas estrategias coincide con v^*.

Primera parte

Definimos el mapa $U : \mathbb{R}_0^N \to \mathbb{R}_0^N$ de modo que

$$(Uv)_i := \text{val}\left[r^i(a,b) + \sum_{j=1}^{N} P_{i,a,b}(j)v_j\right]_{a\in A^i,\ b\in B^i},$$

notar que U está definido de modo que v^* sea un punto fijo. Por la propiedad 3 de 1.2.8, sabemos que la diferencia entre los valores de dos juegos matriciales de la misma dimensión se acota por la máxima diferencia entre las entradas de las matrices; de modo que

$$\begin{aligned} |Uv - Uw|_i &\leq \max_{a,b} |\sum_{j=1}^{N} P_{i,a,b}(j)(v-w)_j| \\ &\leq \max_{a,b} \sum_{j=1}^{N} P_{i,a,b}(j)|v-w|_j, \end{aligned}$$

donde usamos la notación $|v|$ para hacer referencia al vector $(|v_1|, \ldots, |v_n|)$. Aplicando la desigualdad anterior a $U^{n-1}v$ y $U^{n-1}w$ tenemos la siguiente desigualdad

$$|U^n v - U^n w|_i \leq \max_{a,b} \sum_{j=1}^{N} P_{i,a,b}(j)|U^{n-1}v - U^{n-1}w|_j.$$

Llamando f_n y g_n a las estrategias estacionarias puras que a cada estado i le asocian las acciones a_i y b_i que realizan el máximo en la ecuación anterior,

podemos reescribir la inecuación anterior de forma matricial obteniendo

$$\begin{aligned} |U^n v - U^n w| &\leq P(f_n, g_n)|U^{n-1}v - U^{n-1}w| \\ &\leq P(f_n, g_n)\dots P(f_1, g_1)|v - w|. \end{aligned}$$

donde la segunda desigualdad surge de aplicar la primera repetidas veces. Consideremos $||v||$, la norma del máximo y llamemos i a la coordenada del vector $|U^N v - U^N w|$ que realiza el máximo, o sea $|U^N v - U^N w|_i = ||U^N v - U^N w||$; sea π una estrategia que en los primeros N pasos siga las reglas de decisión $f_1, \dots, f_N$ y sea φ una estrategia para el jugador 2 que en los primeros pasos siga las reglas $g_1, \dots, g_N$. Entonces, basándonos en la desigualdad anterior tenemos que

$$\begin{aligned} ||U^N v - U^N w|| &= |U^N v - U^N w|_i \\ &\leq \sum_{j=1}^{N} \mathbb{P}_{i,\pi,\varphi}(S_N = j)|v - w|_j \\ &= \sum_{j=1}^{N-1} \mathbb{P}_{i,\pi,\varphi}(S_N = j)|v - w|_j \\ &\leq \sum_{j=1}^{N-1} \mathbb{P}_{i,\pi,\varphi}(S_N = j)||v - w|| \\ &= \mathbb{P}_{i,\pi,\varphi}(S_N \neq N)||v - w|| \end{aligned}$$

Sabemos, por el teorema 3.2.4, que $\mathbb{P}_{i,\pi,\varphi}(S_N \neq N) < 1$; además $\mathbb{P}_{i,\pi,\varphi}(S_N \neq N)$ depende de i y de las estrategias estacionarias puras $f_1, \dots, f_N$ y $g_1, \dots, g_N$ que son una cantidad finita. Por lo tanto existe $k < 1$ tal que

$$\forall v, w \in \mathbb{R}_0^N, \quad ||U^N v - U^N w|| \leq k||v - w||,$$

y U resulta ser una contracción en N pasos. Por lo tanto tiene un único punto fijo $v^* \in \mathbb{R}_0^N$, que es el único vector en $\mathbb{R}_0^N$ que cumple las ecuaciones (3.5).

Segunda parte

Consideramos las estrategias estacionarias f^* y g^* definidas en el enunciado. Vamos a probar que para toda estrategia estacionaria f del jugador 1 y g del jugador 2 se cumple:

$$v^{f,g^*} \leq v^* \leq v^{f^*,g}$$

lo que prueba que $v^{f^*,g^*} = v^*$ y que el par de estrategias (f^*, g^*) forman un *equilibrio de Nash*; lo que en virtud del teorema 2.3.5, implica que v^* es el valor del juego y que las estrategias mencionadas son óptimas.

Por la definición de f^* sabemos que f_i^* es una estrategia óptima para el jugador 1 en el juego matricial

$$\left[r^i(a,b) + \sum_{j \in S} P_{i,a,b}(j) v_j^* \right]_{a \in A^i,\ b \in B^i},$$

que tiene valor v_i^*, por lo tanto, para toda estrategia g del jugador 2 se cumple

$$r(f^*, g)_i + \sum_{j=1}^{N} P(f^*, g)_{i,j} v_j^* \geq v_i^*,$$

ya que la parte izquierda de la desigualdad es el valor del juego matricial con las estrategias f^* y g. En notación matricial la desigualdad sería,

$$r(f^*, g) + P(f^*, g) v^* \geq v^* \tag{3.6}$$

Aplicando el lema 3.3.3 se obtiene $v^* \leq v^{f^*,g}$. Análogamente se prueba $v^* \geq v^{f,g^*}$, culminando la demostración.

□

3.3.2. Optimización en estrategias semimarkovianas

El teorema que sigue muestra que, en caso de *juegos estocásticos transitorios*, la restricción a estrategias estacionarias, a la hora de buscar estrategias óptimas, no genera una pérdida con respecto a las estrategias semimarkovianas, que a priori es una familia mucho más rica.

Teorema 3.3.4 (**Valor del juego entre las estrategias semimarkovianas**). *En las hipótesis del teorema 3.3.1, el valor v^* hallado en dicho teorema, coincide con el valor del juego si se consideran estrategias semimarkovianas. Además las estrategias estacionarias f^* y g^* resultan óptimas entre las estrategias semimarkovianas.*

Demostración. Vamos a probar que para toda estrategia semimarkoviana π del jugador 1 y φ del jugador 2 se cumple

$$v^{\pi,g^*} \leq v^{f^*,g^*} \leq v^{f^*,\varphi}$$

coordenada a coordenada, de modo que ni el jugador 1 ni el 2 pueden mejorar su desempeño cambiando la estrategia estacionaria óptima por otra semimarkoviana, lo que prueba que el par de estrategias (f^*, g^*) forma un equilibrio de Nash entre las estrategias semimarkovianas, y por el teorema 2.3.5, se obtiene el resultado deseado.

Fijemos $i \neq N$ una coordenada (el caso $i = N$ es trivial ya que se da la igualdad a cero). Como surge de la observación 2.2.8, para cada estado inicial

i, una estrategia semimarkoviana φ, del jugador 2, se corresponde con una estrategia markoviana $\varphi^i = (g_0, g_1, \ldots)$; de modo que $v_i^{f^*,\varphi} = v_i^{f^*,\varphi^i}$. Por otra parte, de forma análoga a la desarrollada en el lema 3.3.2, se prueba que

$$v^{f^*,\varphi^i} = \sum_{t=0}^{\infty} \left[\prod_{k=0}^{t-1} P(f^*, g_k)\right] r(f^*, g_t),$$

donde $\prod_{k=0}^{-1} P(f^*, g_k) = I$.

De la demostración del teorema 3.3.1, en la ecuación (3.6), surge que para toda regla de decisión g del jugador 2 se cumple

$$r(f^*, g) + P(f^*, g)v^* \geq v^*,$$

que aplicada a $g = g_t$ y multiplicada por $\prod_{k=0}^{t-1} P(f^*, g_k)$ resulta

$$\left[\prod_{k=0}^{t-1} P(f^*, g_k)\right] r(f^*, g_t) + \left[\prod_{k=0}^{t} P(f^*, g_k)\right] v^* \geq \left[\prod_{k=0}^{t-1} P(f^*, g_k)\right] v^*;$$

despejando y sumando en $t = 0, \ldots, T$ se obtiene

$$\sum_{t=0}^{T} \left[\prod_{k=0}^{t-1} P(f^*, g_k)\right] r(f^*, g_t) \geq v^* - \left[\prod_{k=0}^{T} P(f^*, g_k)\right] v^*;$$

tomando límite cuando T tiende a infinito, la parte izquierda tiende a v^{f^*,φ^i}; veamos que

$$\left(\left[\prod_{k=0}^{T} P(f^*, g_k)\right] v^*\right)_i \to 0$$

y se obtiene $v_i^{f^*,\varphi^i} \geq v_i^*$:

$$\begin{aligned}
\left|\left(\left[\prod_{k=0}^{T} P(f^*, g_k)\right] v^*\right)_i\right| &\leq \sum_{j=1}^{N} \mathbb{P}_{i,f^*,\varphi^i}(S_T = j)|v_j^*| \\
&= \sum_{j=1}^{N-1} \mathbb{P}_{i,f^*,\varphi^i}(S_T = j)|v_j^*| \\
&\leq \sum_{j=1}^{N-1} \mathbb{P}_{i,f^*,\varphi^i}(S_T = j)||v^*||_\infty \\
&= \mathbb{P}_{i,f^*,\varphi^i}(S_T \neq N)||v^*||_\infty
\end{aligned}$$

por la condición de transitoriedad tenemos que $\mathbb{P}_{i,f^*,\varphi^i}(S_T \neq N)$ tiende a cero por ser el término general de una serie convergente. Entonces $v_i^{f^*,\varphi} = v_i^{f^*,\varphi^i} \geq v_i^*$ para todo estado inicial i. Análogamente se prueba $v^{\pi,g^*} \leq v^*$ con lo que queda demostrado el teorema.

□

3.3.3. Optimización en estrategias generales

Si bien esta sección aparece dentro de los *juegos estocásticos transitorios de suma cero* tiene resultados que valen en contextos más generales. El objetivo es probar que se obtiene el mismo valor óptimo entre todas las estrategias que restringiéndose a las semimarkovianas. Esto pasa para cualquier juego estocástico si el criterio de optimalidad está basado en los valores esperados del pago instantáneo $\mathbb{E}r_j^{S_t}(A_t, B_t)$.

Además, como vimos en la sección anterior, en el caso de los juegos estocásticos transitorios de suma cero hay estrategias estacionarias que son óptimas entre las semimarkovianas, de modo que dichas estrategias son óptimas entre todas las estrategias.

Definición 3.3.5 (**Equivalencia entre pares de estrategias**). *Decimos que un par de estrategias (π, φ) es equivalente a otro (π', φ') si para todo instante t, para todo estado inicial $i \in S$ y para toda configuración instantánea $(j, a, b) \in \mathbb{K}$ se cumple que la probabilidad de estar en dicha configuración en el instante t habiendo partido del estado i coincide para ambas estrategias. Es decir:*

$$\mathbb{P}_{i,\pi,\varphi}(S_t = j, A_t = a, B_t = b) = \mathbb{P}_{i,\pi',\varphi'}(S_t = j, A_t = a, B_t = b)$$

Observación 3.3.6. *Si un criterio de optimalidad $v^{\pi,\varphi} : \pi \in \Pi, \varphi \in \Phi$, es basado exclusivamente en en los valores esperados del pago instantáneo $\mathbb{E}r_j^{S_t}(A_t, B_t)$, como es el caso de los criterios vistos, resulta que el valor de dos pares de estrategias equivalentes coincide. O sea*

$$\forall i \in S,\ v_i^{\pi,\varphi} = v_i^{\pi',\varphi'}$$

si el par de estrategias (π, φ) equivale al par (π', φ')

Teorema 3.3.7. *Dado un par de estrategias (π, φ) para un juego estocástico*

- *Si φ es semimarkoviana existe una estrategia π' semimarkoviana tal que (π, φ) y (π', φ) son equivalentes.*
- *Análogamente, si π es semimarkoviana existe φ' semimarkoviana tal que (π, φ) y (π, φ') son equivalentes.*

Demostración. Probamos la primera parte, ya que la segunda es completamente análoga. Sea i un estado inicial, observemos que si φ es semimarkoviana, sin importar cuál es la estrategia π del jugador 1, se cumple

$$\begin{aligned}
&\mathbb{P}_{i,\pi,\varphi}(S_t = j, A_t = a, B_t = b) \\
&= \mathbb{P}_{i,\pi,\varphi}(A_t = a | S_t = j, B_t = b)\mathbb{P}_{i,\pi,\varphi}(S_t = j, B_t = b) \\
&= \mathbb{P}_{i,\pi,\varphi}(A_t = a | S_t = j)\mathbb{P}_{i,\pi,\varphi}(S_t = j, B_t = b) \\
&= \mathbb{P}_{i,\pi,\varphi}(A_t = a | S_t = j)\mathbb{P}_{i,\pi,\varphi}(B_t = b | S_t = j)\mathbb{P}_{i,\pi,\varphi}(S_t = j)
\end{aligned} \tag{3.7}$$

donde la segunda igualdad se debe a que la distribución de B_t depende únicamente de S_t y del estado inicial i (por ser φ semimarkoviana), de modo que el suceso $\{B_t = b\}$ no agrega información a la probabilidad del suceso $\{A_t = a\}$ cuando se conocen el estado inicial y S_t.

Definimos la estrategia semimarkoviana π', de modo que cuando el estado inicial es i se asocien a π' las reglas de decisión $(f_0, f_1, \ldots)$ tal que

$$f_{t,j}(a) = \mathbb{P}_{i,\pi,\varphi}(A_t = a | S_t = j).$$

Tenemos que probar que para todo t se cumple que cualquiera sea a, b, y j.

$$\mathbb{P}_{i,\pi,\varphi}(S_t = j, A_t = a, B_t = b) = \mathbb{P}_{i,\pi',\varphi}(S_t = j, A_t = a, B_t = b);$$

según la ecuación (3.7) alcanzaría con probar que para todo a, b, y j valen:

$$\mathbb{P}_{i,\pi',\varphi}(A_t = a | S_t = j) = \mathbb{P}_{i,\pi,\varphi}(A_t = a | S_t = j) \tag{3.8}$$

$$\mathbb{P}_{i,\pi',\varphi}(B_t = b | S_t = j) = \mathbb{P}_{i,\pi,\varphi}(B_t = b | S_t = j) \tag{3.9}$$

$$\mathbb{P}_{i,\pi',\varphi}(S_t = j) = \mathbb{P}_{i,\pi,\varphi}(S_t = j) \tag{3.10}$$

La ecuación (3.8) se cumple por cómo fue definida la estrategia π'. La ecuación (3.9) es cierta porque φ es semimarkoviana, entonces la distribución de B_t está completamente determinada por S_t y el estado inicial, de modo que $\mathbb{P}_{i,\pi,\varphi}(B_t = b | S_t = j)$ es independiente de las estrategia π. Veamos que se cumple (3.10) por inducción: para $t = 0$ se cumple trivialmente, supongamos que es cierto para valores de t menores que n, entonces se cumple

$$\begin{aligned}
&\mathbb{P}_{i,\pi',\varphi}(S_n = j) \\
&= \textstyle\sum_{k,a,b} \mathbb{P}_{i,\pi',\varphi}(S_{n-1} = k, A_{n-1} = a, B_{n-1} = b)P_{k,a,b}(j) \\
&= \textstyle\sum_{k,a,b} \mathbb{P}_{i,\pi,\varphi}(S_{n-1} = k, A_{n-1} = a, B_{n-1} = b)P_{k,a,b}(j) \\
&= \mathbb{P}_{i,\pi,\varphi}(S_n = j),
\end{aligned}$$

donde en la segunda igualdad se utilizó la hipótesis inductiva, sumada a las ecuaciones (3.7), (3.8) y (3.9).

□

Corolario 3.3.8. *En las hipótesis del teorema 3.3.1 el valor v^*, definido en dicho teorema, es el valor del juego considerando la clase de estrategias generales. Y las estrategias estacionarias f^* y g^* que surgen en en el mismo teorema son óptimas.*

Demostración. Por el teorema anterior, si existiera una estrategia general π que mejore el desempeño, es decir, tal que $v^{\pi,g^*} > v^*$, existiría una estrategia semimarkoviana π' que cumpliría la condición; lo que contradice el teorema 3.3.4.

□

3.4. Una condición más sencilla para verificar si un juego estocástico es transitorio

El objetivo de esta sección es probar que para que un juego estocástico sea transitorio basta verificar la condición de transitoriedad para las estrategias estacionarias puras. Esto hace mucho más viable la verificación de la transitoriedad de un juego, ya que las estrategias estacionarias puras son una cantidad finita.

Teorema 3.4.1. *Un juego estocástico es transitorio si y sólo si se cumple la condición* (3.1) *de transitoriedad para todo par de estrategias estacionarias puras.*

Para demostrar el teorema necesitamos definir un *proceso de decisión de Markov* (no competitivo) auxiliar y algunos resultados adicionales sobre éste.

El caso no competitivo de un *proceso de decisión de Markov* (PDM) es análogo al caso competitivo pero con un único jugador. Como no es el objetivo de este trabajo estudiar estos procesos vamos a definirlo para este caso particular; si el lector quisiera profundizar en estos temas puede consultar [8, 13]. Suponemos dado un juego estocástico. Para definir el nuevo proceso consideramos los siguientes elementos:

- El mismo conjunto de estados S.
- Conjuntos de acciones posibles para cada estado i, que denotamos D^i, y en este caso $D^i = A^i \times B^i$. O sea que una acción en el PDM se corresponde con un par de acciones en el juego estocástico.
- Como función de pago vamos a considerar la función ℓ, tal que $\ell(i) = 1$ para todo estado $i \neq N$, y $\ell(N) = 0$. Si bien normalmente el pago depende de la acción tomada, para esta aplicación no es necesario.

- Las probabilidades de transición se heredan del juego estocástico. Es decir, si en el estado i se elige la acción $d = (a, b)$ la probabilidad de transición esta dada por $P_{i,a,b} = P_{i,d}$.

Consideramos un proceso $\{X_t\}_{t=0,1,\ldots}$, a valores en S, que representa el estado del proceso (análogo a S_t del caso competitivo), y un proceso $\{D_t\}_{t=0,1,\ldots}$ que representa las acciones que va tomando el jugador. Las estrategias para el PDM se definen de la misma forma que en el caso competitivo, salvo que las distribuciones de probabilidad son sobre $\mathbb{D} = \mathbb{A} \times \mathbb{B}$. En particular se tienen las mismas clases de estrategias (generales, semimarkovianas, markovianas, estacionarias, puras). Las estrategias también se asocian con reglas de decisión de la misma forma que en el caso competitivo.

La definición formal del PDM, es completamente análoga a la del juego estocástico, dada en la sección 2.1.1. O sea, dado un estado inicial i y una estrategia σ se tiene que $X_0 = i$; el estado del proceso en el instante t se sortea según la probabilidad de transición del PDM, que depende del estado y la acción en el instante anterior. La acción en el instante t se sortea según σ en base a la historia del proceso hasta ese instante.

Dada una estrategia σ para el PDM definimos el valor de la estrategia, de forma análoga al caso competitivo, como un vector w^σ en que la coordenada i-ésima representa la suma sobre todos los instantes de los valores esperados de la ganancia instantánea, en este caso particular tenemos

$$\begin{aligned} w_i^\sigma &= \sum_{t=0}^{\infty} \mathbb{E}_{i,\sigma} \ell(X_t) \\ &= \sum_{t=0}^{\infty} \sum_{j=1}^{N} \mathbb{P}_{i,\sigma}(X_t = j)\ell(j) \\ &= \sum_{t=0}^{\infty} \sum_{j=1}^{N-1} \mathbb{P}_{i,\sigma}(X_t = j) \\ &= \sum_{t=0}^{\infty} \mathbb{P}_{i,\sigma}(X_t \neq N), \end{aligned} \tag{3.11}$$

que podría ser infinito.

Al igual que como fue visto en el caso competitivo, dado un estado inicial i, y una estrategia σ, si en un cierto paso t del PDM se aplica la regla de decisión h quedan definidas las probabilidades de transición en ese paso para el proceso $\{X_t\}$; estas probabilidades las representamos en una matriz $P(h)$, de modo que

$P(h)_{jk} = \mathbb{P}_{i,\sigma}(X_{t+1} = k | X_t = j)$. En el caso de una estrategia estacionaria h las potencias de la matriz $P(h)$ coinciden con las probabilidades de transición en varios pasos; de modo que se cumple

$$\mathbb{P}_{i,h}(X_t = j) = P(h)^t_{ij}$$

Lema 3.4.2. *Si h es una estrategia estacionaria para el PDM se cumple:*

$$w^h = \sum_{t=0}^{\infty} P(h)^t \ell,$$

donde ℓ es el vector de $\mathbb{R}^N$ $(1, \ldots, 1, 0)$. Además de la ecuación anterior surge inmediatamente que

$$w^h = \ell + P(h) w^h.$$

Demostración. Según la ecuación (3.11) basta ver que $(P(h)^t \ell)_i$ coincide con $\sum_{j=1}^{N-1} \mathbb{P}_{i,h}(X_t = j)$; pero

$$\begin{aligned} (P(h)^t \ell)_i &= \sum_{j=1}^{N} P(h)^t_{ij} \ell_j = \sum_{j=1}^{N-1} P(h)^t_{ij} \\ &= \sum_{j=1}^{N-1} \mathbb{P}_{i,h}(X_t = j) \end{aligned} \tag{3.12}$$

lo que culmina la prueba.

□

Lema 3.4.3. *Si h es una estrategia estacionaria del PDM, para la que cualquiera sea el estado inicial i vale*

$$\sum_{t=0}^{\infty} \mathbb{P}_{i,h}(X_t \neq N) < \infty;$$

y w es un vector de $\mathbb{R}_0^N$ tal que

$$w \leq \ell + P(h) w, \tag{3.13}$$

considerando la desigualdad coordenada a coordenada, con ℓ definida como en el lema anterior; entonces se cumple

$$w \leq w^h.$$

Demostración. Si en la parte derecha de la inecuación (3.13), al igual que en la demostración de 3.3.3, sustituimos w utilizando la propia inecuación, al cabo de k veces obtenemos

$$w \leq \sum_{t=0}^{k} P(h)^t \ell + P(h)^{k+1} w.$$

Veamos que al hacer tender k a infinito, la parte derecha de la inecuación tiende a w^h, lo que culminaría la prueba. Del lema anterior surge que $\sum_{t=0}^{k} P(h)^t \ell$ tiende a w^h, por lo que bastaría ver que $P(h)^{k+1}\ell$ tiende a cero. De la ecuación (3.12) se deduce que $(P(h)^{k+1}\ell)_i = \mathbb{P}_{i,h}(X_{k+1} \neq N)$, que es el término general de una serie que, por hipótesis, es convergente.

□

Observación 3.4.4. *Dado un juego estocástico, a cada par de estrategias puras (π, φ) del juego se le asocia una estrategia pura π en el PDM; de modo que si para cierta historia h, $\pi(a|h) = 1$ y $\varphi(b|h) = 1$ entonces $\pi(d|h) = 1$, con $d = (a, b)$. La correspondencia es biunívoca.*

Lema 3.4.5. *Dado un juego estocástico que cumple la condición* (3.1) *de transitoriedad para todo par de estrategias estacionarias puras existe un vector $w = (w_1, \ldots, w_{N-1}, 0)$ tal que*

$$w_i = \max_{(a,b)\in A^i \times B^i} \left\{ 1 + \sum_{j=1}^{N} P_{i,a,b}(j) w_j \right\}$$

para todo $i = 1, \ldots, N-1$.

Demostración. Consideramos el PDM asociado al juego estocástico. Dado que para todo par de estrategias estacionarias puras se cumple la condición de transitoriedad, tenemos que si f y g son estrategias estacionarias puras

$$\sum_{t=0}^{\infty} \mathbb{P}_{i,f,g}(S_t \neq N) < \infty;$$

por la observación anterior y la dinámica del PDM, se cumple que para toda estrategia estacionaria pura h del PDM

$$\sum_{t=0}^{\infty} \mathbb{P}_{i,h}(X_t \neq N) < \infty.$$

Según la ecuación (3.11), la parte izquierda de la inecuación anterior coincide con w_i^h. Como la cantidad de estrategias estacionarias puras es finita podemos definir el vector $w = (w_1, \ldots, w_N)$, tal que

$$w_i = \max_h w_i^h,$$

tomando el máximo sobre las estrategias estacionarias puras. Si llamamos h_i a la estrategia estacionaria pura que realiza el máximo anterior, para cada $i =$

$1, \ldots, N-1$, se tiene

$$\begin{aligned} w_i = w_i^{h_i} &= 1 + \sum_{j=1}^{N} P(h_i)_{ij} w_j^{h_i} \\ &\leq 1 + \sum_{j=1}^{N} P(h_i)_{ij} w_j \\ &\leq \max_{d \in D^i} \left\{ 1 + \sum_{j=1}^{N} P_{i,d}(j) w_j \right\}, \end{aligned}$$

donde la primera igualdad surge del "además" en el lema 3.4.2, la primera desigualdad se debe a que $w_j^{h_i} \leq w_j$ por ser w_j el máximo entre una clase de estrategias que contiene a h_i. Para entender la última desigualdad basta observar que h_i es una estrategia pura, por lo que existe $d_i \in D^i$ tal que $h_i(d_i) = 1$ y $\forall_j \; P(h_i)_{ij} = P_{i,d_i}(j)$, entonces al tomar máximo en d nos aseguramos la desigualdad.

Consideremos la estrategia estacionaria pura h tal que para cada estado $i = 1, \ldots, N-1$, $h(i)$ acumula toda la probabilidad en la acción d_i que realiza el máximo en

$$\max_{d \in D^i} \left\{ 1 + \sum_{j=1}^{N} P_{i,d}(j) w_j \right\}.$$

Entonces se cumple para todo $i = 1, \ldots, N-1$

$$w_i \leq 1 + P(h)_{ij} w_j,$$

o en forma matricial

$$w \leq \ell + P(h) w,$$

lo que implica, según el lema 3.4.3, que $w \leq w^h$. Por como fue definida w, la desigualdad estricta no se puede cumplir para ninguna coordenada, entonces $w = w^h$ y cumple las ecuaciones deseadas.

□

Demostración del teorema 3.4.1.
Dado un juego estocástico en las hipótesis del teorema, sabemos, por el lema anterior, que existe un vector $w = (w_1, \ldots, w_{N-1}, 0)$ tal que

$$w_i = \max_{(a,b) \in A^i \times B^i} \left\{ 1 + \sum_{j=1}^{N} P_{i,a,b}(j) w_j \right\}$$

para todo $i = 1, \ldots, N-1$. Por lo tanto, para cualquier estrategia estacionaria

h del PDM, se cumple

$$\forall i \in [1 \dots N], \quad w_i \geq 1 + \sum_{j=1}^{N} P(h)_{ij} w_j,$$

que podemos escribirlo en forma matricial usando el vector $\ell = (1, \dots, 1, 0) \in \mathbb{R}^N$ como

$$w \geq \ell + P(h)w$$

Ahora consideremos una estrategia semimarkoviana σ del PDM; fijado el estado inicial i, la estrategia se corresponde con una secuencia de reglas de decisión $h_0, h_1, h_2 \dots$; multiplicando la desigualdad anterior aplicada a h_t por $\prod_{k=0}^{t-1} P(h_k)$ se obtiene

$$\prod_{k=0}^{t-1} P(h_k)w \geq \prod_{k=0}^{t-1} P(h_k)\ell + \prod_{k=0}^{t} P(h_k)w,$$

que despejando y sumando en $t = 0$ hasta $t = T$ resulta

$$\sum_{t=0}^{T} \prod_{k=0}^{t-1} P(h_k)\ell \leq w - \prod_{k=0}^{T} P(h_k)w$$

como la parte derecha de la desigualdad se puede acotar independientemente de T tenemos que todas las entradas de

$$\sum_{t=0}^{\infty} \prod_{k=0}^{t-1} P(h_k)\ell$$

son finitas. En particular la entrada i-ésima es

$$\sum_{t=0}^{\infty} \sum_{j=1}^{N-1} \mathbb{P}_{i,\sigma}(X_t = j) < \infty.$$

Dadas un par de estrategias semimarkovianas π y φ se pueden combinar para obtener una estrategia semimarkoviana del PDM, y la dinámica es la misma, por lo tanto, usando la ecuación anterior, se obtiene

$$\sum_{t=0}^{\infty} \sum_{j=1}^{N-1} \mathbb{P}_{i,\pi,\varphi}(S_t = j) < \infty,$$

por lo que vale la condición de transitoriedad para las estrategias semimarkovianas. Resta ver que vale la condición para todo par de estrategias generales. Sean π y φ estrategias generales de los jugadores 1 y 2. Para cada estado inicial $i \in S$,

$$\mathbb{P}_{i\pi\varphi}(S_t = j, A_t = a, B_t = b)$$

$$= \mathbb{P}_{i\pi\varphi}(A_t = a, B_t = b | S_t = j)\mathbb{P}_{i\pi\varphi}(S_t = j).$$

Definimos la estrategia semimarkoviana $\sigma = (h_0, h_1, \ldots)$ del PDM de modo que

$$h_{t,j}(a,b) = \mathbb{P}_{i\pi\varphi}(A_t = a, B_t = b | S_t = j).$$

Si probamos que, para todo instante t se cumple que para todo $j \in S$

$$\mathbb{P}_{i,\sigma}(X_t = j) = \mathbb{P}_{i\pi\varphi}(S_t = j),$$

aplicando luego el resultado para estrategias semimarkovianas tendríamos lo que queremos. Veámoslo por inducción: para t=0 se cumple la igualdad trivialmente; supongamos que es cierta para valores de t menores que n, entonces

$$\begin{aligned}
\mathbb{P}_{i,\pi,\varphi}(S_n = j) &= \sum_{k,a,b} \mathbb{P}_{i,\pi,\varphi}(S_{n-1} = k, A_{n-1} = a, B_{n-1} = b) P_{k,a,b}(j) \\
&= \sum_{k,a,b} \mathbb{P}_{i,\pi,\varphi}(A_{n-1} = a, B_{n-1} = b | S_{n-1} = k) \mathbb{P}_{i,\pi,\varphi}(S_{n-1} = k) P_{k,a,b}(j) \\
&= \sum_{k,a,b} h_{n-1,k}(a,b) \mathbb{P}_{i,\sigma}(X_{n-1} = k) P_{k,a,b}(j) \\
&= \mathbb{P}_{i,\sigma}(X_n = j),
\end{aligned}$$

y la igualdad se cumple para todo t, completando la prueba.

□

Capítulo 4

Aplicación a "la codicia"

4.1. Introducción

Si bien el estudio de los *juegos estocásticos*, y en particular el de los *juegos estocásticos transitorios*, se volvió un fin en sí mismo en el transcurso de este trabajo, la motivación para estudiarlos surgió de la idea de jugar de forma óptima al juego de dados conocido como "la codicia" o "el uno" o en inglés "pig game". Por lo tanto, nuestro ejemplo de aplicación de esta teoría es la resolución del juego de dados.

En mi monografía de licenciatura [4] ya había estudiado aspectos aplicables a este juego, pero desde un punto de vista solitario; y había quedado planteada la necesidad de resolverlo en el caso competitivo, ya que algunas simulaciones hechas sugerían que la estrategia "óptima" hallada no lo era en realidad, cuando se trataba de maximizar la probabilidad de ganar.

En este capítulo se muestra cómo se aplican los resultados obtenidos en el capítulo anterior al juego de dados para obtener un algoritmo que arriba a la estrategia óptima. Además se presentan los resultados obtenidos por dicho algoritmo y se comparan con las estrategias conocidas para el caso solitario.

4.2. Las reglas del juego

La codicia se juega entre dos jugadores y el objetivo final es ser el primero en alcanzar o superar 200 puntos (o algún puntaje establecido a priori). Para lograr el objetivo, los jugadores van acumulando puntaje por turnos alternados. El primer turno debe ser sorteado, ya que ser el que empieza supone ventaja.

Un turno: En un turno el jugador tira un dado todas las veces que quiera o hasta que obtenga un as como resultado. Si el turno termina por decisión del jugador (sin que aparezca un as), éste anota la suma de los resultados obtenidos

en el turno, sino anota 0 puntos y cede el dado al contrincante. En definitiva, luego de cada tirada, si el resultado del dado no fue un as, el jugador que tiene el dado se enfrenta a las siguientes opciones:

- tirar de nuevo, arriesgando a perder todo el puntaje acumulado en el turno actual, para tratar de incrementar el rendimiento del turno;
- parar, anotando la suma de los resultados obtenidos, y ceder el turno al contrincante.

Claramente, el problema de cómo maximizar el desempeño en un turno, es un problema de parada óptima.

4.3. Algunos resultados previos

4.3.1. Optimizar por turno

En el artículo de M.Roters [23] se estudia el problema de parada óptima de un turno. Es decir, cómo maximizar el valor esperado del puntaje de un turno; es claro que jugar maximizando el puntaje por turno parece una buena estrategia. Para este problema se encuentra que parar cuando uno acumuló 20 o más puntos resulta óptimo; al igual que parar cuando se llega a 21 o más, que es equivalente. Este resultado se obtiene de forma muy sencilla comparando el valor esperado si se decide *tirar* y el valor esperado si se decide *parar*. Sobre este problema se puede leer en [4].

4.3.2. Minimizar la esperanza de la cantidad de turnos

En un segundo artículo J.Haigh y M.Roters [12] estudian una estrategia mejor, que consiste en tratar de minimizar el valor esperado de la cantidad de turnos que requiere alcanzar los 200 puntos. Esta estrategia, al igual que la que optimiza por turno, no tiene en cuenta el puntaje del contrincante, por eso les llamamos estrategias solitarias. Los métodos usados para hallar esta estrategia son de *procesos de decisión de Markov* no competitivos. El resultado que se obtiene en este segundo caso se muestra en la figura 4.1, en que para cada par (α, τ), donde α es el puntaje ya anotado por el jugador y τ es el puntaje del turno actual, se indica si se debe parar (zona en negro) o tirar(zona en gris).

Por ejemplo, si el jugador tiene ya ganados 100 puntos y en el turno actual ha acumulado 10 puntos se debe ir a la coordenada (100,10) y se observa que es la zona gris, por lo tanto se debe seguir tirando. En cambio si en el turno actual ya se juntaron 25 puntos se debería parar. Notar que para valores bajos de α el límite entre tirar o parar se encuentra alrededor de los 20 puntos, que es la estrategia que maximiza el puntaje esperado de un turno. Sin embargo cuando

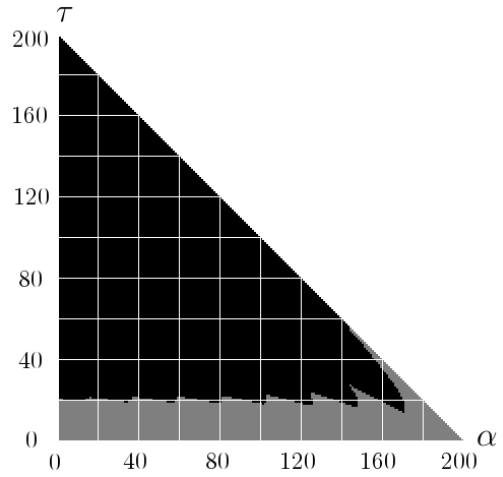


Figura 4.1: Estrategia solitaria que minimiza el valor esperado de la cantidad de turnos. En la zona gris se debe *tirar*, mientras que en la zona negra se debe *parar*

el juego se acerca a su fin la estrategia cambia un poco. Sobre esta estrategia también se puede profundizar en mi trabajo monográfico de grado [4].

4.4. “La codicia” como juego estocástico transitorio

4.4.1. Estrategias en el juego de dados

Pensemos en estrategias para el jugador 1; ya que la simetría del juego lleva a una correspondencia trivial entre las estrategias de ambos jugadores, de modo que una estrategia para el jugador 1 también puede ser entendida como una estrategia para el jugador 2.

Informalmente, una estrategia para el juego, es una forma de elegir qué acción tomar. Las acciones en este caso son *tirar* y *parar*, y la información con la que cuenta el jugador para tomar la decisión es el transcurso del juego hasta el instante actual. Entonces, inspirados en la definición de estrategia general, digamos que una estrategia para el jugador 1, es una regla que a cada historia del juego, que termina en un momento en que el jugador 1 debe decidir, asocia una distribución de probabilidad en el conjunto $\{$*tirar*,*parar*$\}$.

Las estrategias solitarias, mencionadas en la sección anterior, si bien resultan formas muy razonables de jugar al juego de dados, utilizan muy poca información para tomar la decisión. Es claro que son un caso particular de las estrategias consideradas. ¿Pero son óptimas?

4.4.2. El problema planteado

Dadas dos estrategias π y φ, llamemos $p_{\pi,\varphi}$ a la probabilidad de que gane el jugador 1 cuando éste juega con la estrategia π y su contrincante lo hace con la estrategia φ, por lo tanto la probabilidad de que gane el jugador 2 es $1 - p_{\pi,\varphi}$ y es claro, por la simetría del juego, que se cumple

$$p_{\varphi,\pi} = 1 - p_{\pi,\varphi}.$$

Decimos que la estrategia π es *mejor o igual* que φ, si se cumple

$$p_{\pi,\varphi} \geq p_{\varphi,\pi}$$

o lo que es lo mismo, si $p_{\pi,\varphi} \geq \frac{1}{2}$. Y decimos que la estrategia π^* es *óptima* si resulta ser *mejor o igual* que cualquier otra estrategia. Obviamente si el contrincante juega con la propia estrategia π^* las probabilidades de ganar serán iguales para cada jugador. Por lo tanto jugar con una estrategia *óptima* asegura ganar con probabilidad de al menos $\frac{1}{2}$

Notar que si uno conociera la estrategia de su contrincante, podría encontrar una estrategia que aumente la probabilidad de ganar con respecto a la óptima. El mérito de la estrategia *óptima* es que es buena independientemente del contrincante.

El problema que nos planteamos es el de encontrar, si es que existe, una estrategia óptima.

4.4.3. Modelado del juego

Para aplicar los aspectos estudiados para juegos estocásticos a la resolución de nuestro problema debemos plantearlo como un problema de *juegos estocásticos*; es decir, definir un conjunto de estados, acciones para los jugadores, función de pago, probabilidades de transición y un criterio de optimalidad.

Los estados

Si uno piensa en el transcurso del juego, paso a paso, hay 4 aspectos que van variando: quien tiene el dado, puntaje del jugador 1, puntaje del jugador 2 y puntaje del turno actual del que está jugando. Por lo tanto consideramos como estados a las 4-úplas (j, α, β, τ), donde

- $j = 1$ o 2 según a quién le toca jugar,
- $\alpha \in [0 \dots 199]$ es el puntaje que el jugador 1 ya tiene ganado y anotado,
- $\beta \in [0 \dots 199]$ es el puntaje del jugador 2 ya ganado y anotado,

- τ es el puntaje del turno actual del jugador j y toma valores en $[0 \dots 205-\alpha]$ si $j = 1$ y en $[0 \dots 205 - \beta]$ si $j = 2$.

además se tiene un estado inicial s_i en el que se sortea quién empieza y un estado final s_f absorbente en el que el juego terminó, el que en la teoría le llamamos N.

Las acciones

Debemos indicar, para cada estado, cuáles son las acciones posibles de cada jugador; veamos las del jugador 1:

- en los estados s_i y s_f el jugador 1 no tiene opciones, para ser coherentes con la teoría deberíamos decir que el jugador 1 tiene una sola acción posible, llamémosle *esperar*;
- en los estados de la forma $(2, \alpha, \beta, \tau)$ es el jugador 2 quien debe decidir, entonces el jugador 1 solo puede *esperar*;
- en los estados de la forma $(1, \alpha, \beta, 0)$ el turno del jugador 1 recién empieza, por lo que solo tiene sentido la acción *tirar*;
- a los estados $(1, \alpha, \beta, \tau)$, en que $\alpha + \tau \geq 200$, los llamamos estados ganadores, y la única acción posible en ellos es *parar*;
- y en el resto de los estados, los de la forma $(1, \alpha, \beta, \tau)$ con $0 < \tau < 200 - \alpha$ el jugador 1 puede optar por las dos acciones, *tirar* o *parar*.

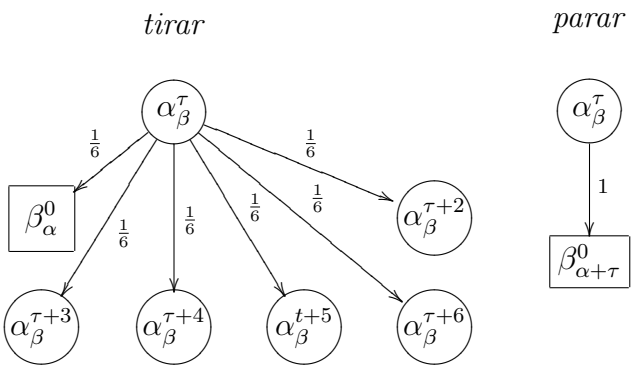
Las acciones posibles para el jugador 2 son totalmente simétricas.

Las transiciones

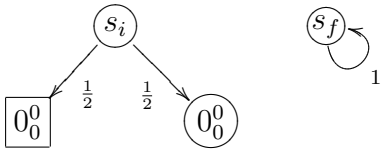
Para representar gráficamente las probabilidades de transición representamos los estados del jugador 1 con un círculo y los del jugador 2 con un cuadrado así:

$$(1, \alpha, \beta, \tau) = \textcircled{α^{τ}_{β}} \qquad (2, \alpha, \beta, \tau) = \boxed{\beta^{\tau}_{\alpha}}.$$

La dinámica del juego, y la semántica de los estados y las acciones, determinan las probabilidades de transición. Veamos las probabilidades de transición del jugador 1 según la acción que toma en un estado de la forma $(1, \alpha, \beta, \tau)$ con $\alpha + \tau < 200$:



En el diagrama se ve que cuando el jugador decide tirar tiene probabilidad $\frac{1}{6}$, por la eventualidad de que salga un as, de perder el puntaje del turno τ y tener que ceder el dado al contrincante. En los estados del jugador 1 en que este ya ganó, es decir, los de la forma $(1, \alpha, \beta, \tau)$ con $\alpha + \tau \geq 200$, la transición es con probabilidad 1 al estado final, ya que el juego terminó. Para el jugador 2 las transiciones son simétricas. En el caso de los estados especiales s_i y s_f, las transiciones no dependen de las acciones de los jugadores, ya que estos tienen una sola opción, y son así:



La función de pago

Vamos a definir un juego de suma cero, entonces basta con dar la función de pago para el jugador 1. Como nuestro objetivo es maximizar la probabilidad de ganar, queremos definir los pagos de modo que si gana el jugador 1, la ganancia total sea 1 para éste y -1 para el jugador 2 y en el caso contrario, si gana el jugador 2, la ganancia sea 0 para ambos. De ese modo el valor esperado de la ganancia total coincide con la probabilidad de que gane el jugador 1. El modelo definido en los capítulos anteriores permite, para cada estado i, definir una función de pago r^i que a cada par de acciones asocie un número real, pero en nuestro caso el pago no va a depender de las acciones tomadas, en definitiva r^i es una constante que depende de i. Definimos, para cada $i \in S$, la constante r^i, mediante,

$$r^i = \begin{cases} 1 & \text{si } i = (1, \alpha, \beta, \tau) \text{ con } \alpha + \tau \geq 200 \\ 0 & \text{en cualquier otro caso} \end{cases}$$

Estudiando con un poco de cuidado el modelo planteado, uno observa que si gana el jugador 1 el juego pasa una y solo una vez por un estado ganador del jugador 1 y si éste pierde nunca se pasa por uno de estos estados.

Con esto hemos terminado de definir el modelo de juego estocástico asociado al juego de dados. Ahora veamos que tal modelo es un modelo de juego estocástico transitorio. Para eso, en virtud del teorema 3.4.1, basta probar que se cumple la condición (3.1) de transitoriedad para todo par de estrategias estacionarias puras. Consideremos fijas dos estrategias estacionarias puras, queremos ver que

$$\sum_{t=0}^{\infty} \mathbb{P}(S_t \neq N) < \infty$$

donde N es el estado s_f. La forma en que fueron definidas las acciones posibles de los jugadores no permiten perpetuar el juego indefinidamente, ya que al inicio del turno es obligatorio tirar, y cuando el juego está ganado es obligatorio parar. Gracias a esto, sabemos que si, por ejemplo, en el dado sale el número 6 repetido 70 veces seguidas no hay forma de evitar que el juego llegue al estado final. Sea γ la probabilidad de que ocurra tal cosa, obviamente $\gamma > 0$. Por lo mencionado antes se cumple

$$\mathbb{P}(S_t \neq N) < 1 - \gamma \text{ si } 70 \leq t < 140$$

y en general

$$\mathbb{P}(S_t \neq N) < (1 - \alpha)^n \text{ si } 70n \leq t < 70(n + 1),$$

de modo que dominamos la serie por una serie convergente, lo que nos asegura que se cumple la condición de transitoriedad.

4.5. La estrategia óptima

Ahora que tenemos un modelo de juego estocástico transitorio podemos aplicar el resultado obtenido en el teorema 3.3.1, para hallar la estrategia estacionaria óptima, que vimos que es óptima entre todas las estrategias generales. En la demostración de dicho teorema se define la aplicación U, que resulta ser una contracción, y su único punto fijo es el vector v^* del enunciado. Por lo tanto sabemos que

$$v^* = \lim_{n \to \infty} U^n(v),$$

donde v es un vector inicial cualquiera. También surge de la demostración de 3.3.1 que v_i^* es el valor del juego partiendo del estado i cuando se juega con las estrategias óptimas. Una vez que se conoce el vector v^* las estrategias estacionarias óptimas consisten en, para cada estado i, jugar de forma óptima al siguiente juego matricial:

$$\left[r^i(a, b) + \sum_{j \in S} P_{i,a,b}(j) v_j^* \right]_{a \in A^i,\ b \in B^i}.$$

La notación usada en el capítulo 3 es un poco distinta a la utilizada en esta aplicación. En particular no tenemos numerados los estados del 1 al N, entonces no tiene muchos sentido hablar de vectores; pero eso es un problema menor, en lugar de hablar de vectores, pensemos en funciones que a cada estado asocian un número real. Si v^* es la función asociada a la estrategia óptima es claro que $v^*(s_f) = 0$, porque el estado s_f es absorbente y no genera ganancia.

Si traducimos la definición de la aplicación U dada en el teorema 3.3.1 a nuestro caso particular, resulta:

$$Uv(s) = \begin{cases} \frac{1}{2}v(1,0,0,0) + \frac{1}{2}v(2,0,0,0) & \text{si } s = s_i \\ v(1,\alpha,\beta,0)_{tirar} & \text{si } s = (1,\alpha,\beta,0) \\ \max\{v(1,\alpha,\beta,\tau)_{parar}, v(1,\alpha,\beta,\tau)_{tirar}\} & \text{si } s = (1,\alpha,\beta,\tau) :\ \alpha+\tau < 200 \\ 1 & \text{si } s = (1,\alpha,\beta,\tau) :\ \alpha+\tau \geq 200 \\ v(2,\alpha,\beta,0)_{tirar} & \text{si } s = (2,\alpha,\beta,0) \\ \min\{v(2,\alpha,\beta,\tau)_{parar}, v(2,\alpha,\beta,\tau)_{tirar}\} & \text{si } s = (2,\alpha,\beta,\tau) :\ \beta+\tau < 200 \\ 0 & \text{en cualquier otro caso} \end{cases}$$

donde

$$\begin{aligned} v(1,\alpha,\beta,\tau)_{tirar} &= \tfrac{1}{6}v(2,\alpha,\beta,0) + \tfrac{1}{6}\textstyle\sum_{k=2}^{6} v(1,\alpha,\beta,\tau+k) \\ v(1,\alpha,\beta,\tau)_{parar} &= v(2,\alpha+\tau,\beta,0) \\ v(2,\alpha,\beta,\tau)_{tirar} &= \tfrac{1}{6}v(1,\alpha,\beta,0) + \tfrac{1}{6}\textstyle\sum_{k=2}^{6} v(2,\alpha,\beta,\tau+k) \\ v(2,\alpha,\beta,\tau)_{parar} &= v(1,\alpha,\beta+\tau,0) \end{aligned}$$

Notar que en lugar del valor de un juego matricial lo sustituimos por el máximo o el mínimo entre las dos acciones posibles, según de quién sea el turno; es muy fácil verificar que en un juego matricial en que sólo el jugador 1 tiene acciones para elegir el valor del juego es el máximo entre las entradas de la matriz y la estrategia óptima es la estrategia pura correspondiente a elegir la acción que realiza el máximo. Análogamente sucede con el jugador 2 y el mínimo.

Una vez conocido v^*, para hallar la estrategia óptima se debe elegir, para cada estado $(1,\alpha,\beta,\tau) :\ 0 < \tau < 200-\alpha$, la acción que realiza el máximo entre

$$v^*(1,\alpha,\beta,\tau)_{parar} = v^*(2,\alpha+\tau,\beta,0)$$

y

$$v^*(1,\alpha,\beta,\tau)_{tirar} = \frac{1}{6}v^*(2,\alpha,\beta,0) + \frac{1}{6}\sum_{k=2}^{6} v^*(1,\alpha,\beta,\tau+k);$$

y ese valor máximo, que coincide con $v^*(1,\alpha,\beta,\tau)$, es la probabilidad de que gane el jugador 1 si ambos juegan con la estrategia óptima y se parte del estado $(1,\alpha,\beta,\tau)$.

Para decir cuál es la estrategia óptima deberíamos decir qué acción tomar en 4.000.000 de estados. Por lo que optamos por poner algunas gráficas, ver figura 4.2, con puntaje del oponente fijo, que son representativas de cómo es la estrategia. En [5] se encuentra un documento con la totalidad de las gráficas.

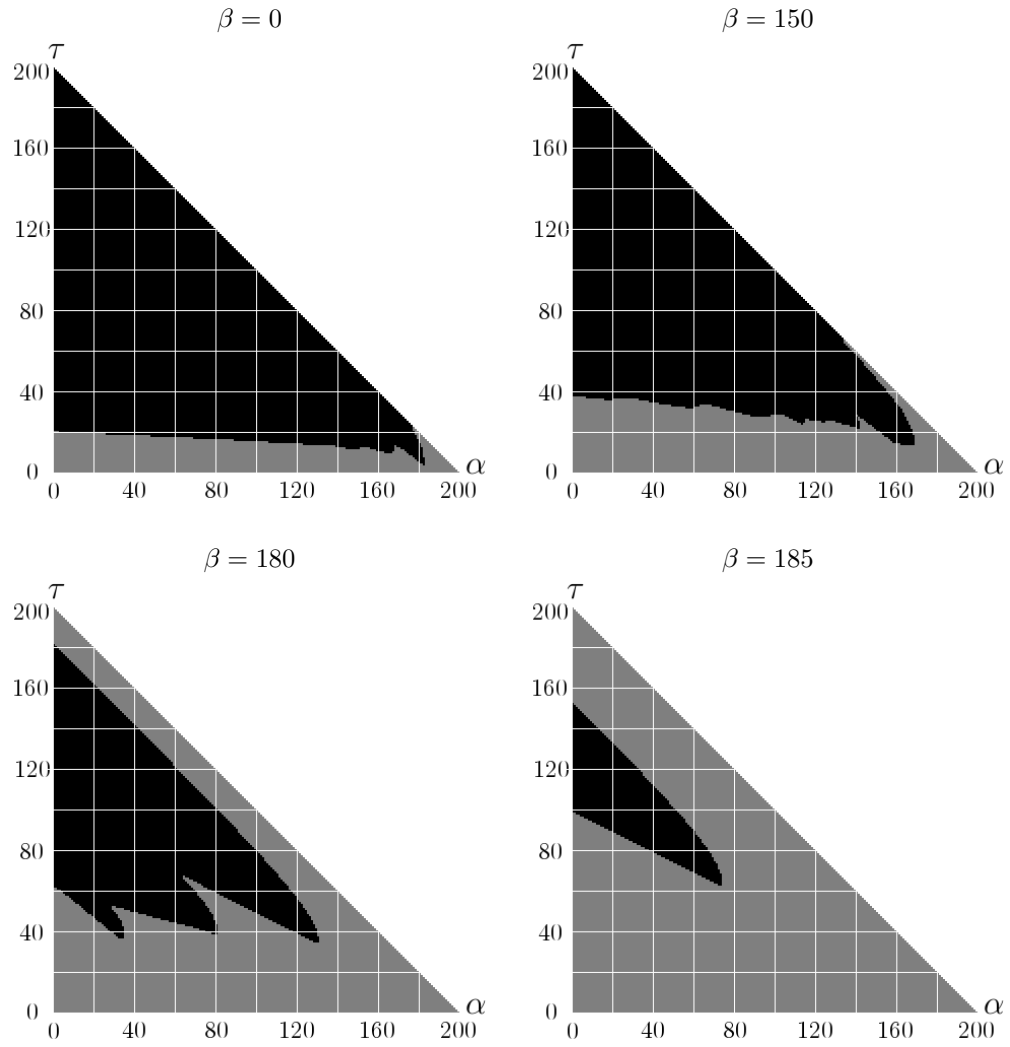


Figura 4.2: Parte de la estrategia óptima, para puntaje del contrincante = 0,150,180 y 185. En la zona gris se debe *tirar* y en la zona negra *parar*

Algunos comentarios sobre las solución

- Si miramos el gráfico cuando el oponente tiene 0 puntos, vemos que al principio (cuando α es cercano a cero) la estrategia es muy parecida a *parar en 20*, que como fue mencionado, maximiza la esperanza del puntaje de cada turno. Esto tiene la interpretación natural de que cuando se está lejos de terminar el juego conviene acercarse al final con "pasos" lo más grande posibles.
- En general, cuando el oponente tiene el mismo puntaje que uno, se observa que la estrategia óptima sugiere parar en 20. Habría que agregar más gráficas para visualizar bien este hecho; pero su interpretación es clara, si se está en igualdad de condiciones se debe intentar maximizar el puntaje del turno.
- Por ejemplo, la gráfica correspondiente a 150 puntos del oponente ($\beta = 150$) es similar a la de $\beta = 0$ pero más arriesgada, es decir, en cada turno pretende ganar más puntaje, aunque en valor esperado significa ganar menos puntaje. Esto se explica porque al juego le quedan pocos turnos (o al menos eso sucede en valor esperado) entonces hay que tratar de hacer más puntaje por turno para poder alcanzar al contrincante antes de que gane.
- Para puntajes del oponente de 187 en adelante ($\beta \geq 187$), la gráfica es toda gris. Es decir, sin importar cuánto puntaje se tenga anotado, si el oponente está tan próximo a ganar, hay que tratar de alcanzar los 200 puntos es un solo turno. Es claro que darle el dado supone un riesgo demasiado grande.

En el artículo [18] se estudia la estrategia óptima, entre las estacionarias puras, para este mismo juego con horizonte 100 en lugar de 200. En dicho artículo los autores plantean la ecuación de tipo Bellman que debería cumplir una estrategia estacionaria para sea la mejor entre dicha clase de estrategia, y observan que por no haber el factor de descuento en la suma podría ocurrir que el método iterativo que plantean no converja, pero observan que en este caso particular sí converge. La diferencia fundamental de este trabajo es que se justifica matemáticamente el método que arriba a la solución y se prueba que si bien la solución hallada es estacionaria y pura, ésta es óptima entre todas las estrategias.

Comparación de la solución óptima con la estrategia solitaria Para comparar la estrategia óptima con la estrategia que minimiza la esperanza de la cantidad de turnos (solitaria), dada en la figura 4.1, lo que se hizo fue hacer un algoritmo que simula un millón de veces el juego en que el jugador 1 juega con la estrategia óptima y el jugador 2 con la otra. Al correr 10 veces dicho algoritmo se obtuvieron los siguientes resultados:

Simulación	Veces que ganó 1	Veces que ganó 2
1	520153	479847
2	520079	479921
3	520017	479983
4	519008	480992
5	520830	479170
6	519256	480744
7	519863	480137
8	518897	481103
9	519955	480045
10	520903	479097

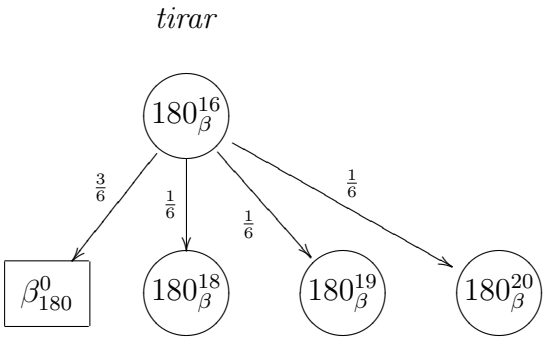
que sugieren que la probabilidad de ganar cuando se juega con la estrategia óptima es de aproximadamente 0,52

4.6. Variantes al juego

4.6.1. Modelo con rebote

Una variante a "la codicia" que resulta interesante estudiar, para ver como cambian las estrategias, es cuando el jugador gana sólo si llega al puntaje exacto. Por ejemplo, si el jugador 1 tiene 180 puntos ya anotados, y en el turno actual acumuló 23 puntos, puede plantarse y ganar en el modelo clásico; sin embargo en este nuevo caso perdería el turno (como si hubiese sacado un as en el dado).

El modelo para resolver este caso es muy similar. Naturalmente, las diferencias surgen en los estados del jugador 1 $(1, \alpha, \beta, \tau)$ con $\alpha + \tau > 195$ cuando éste decide *tirar*, y en los simétricos del jugador 2, que es cuando aparece el riesgo de no ganar por "pasarse del objetivo". Vemos con un ejemplo concreto cómo serían las transiciones en este caso, usando la misma representación gráfica de los estados que en el modelo clásico.



en el ejemplo se ve que hay 3 resultados en que se le da el dado al contrincante (si sale un 1, 5, 6). Los estados de la forma $(1, \alpha, \beta, \tau)$ con $\alpha + \tau > 200$ ya no son necesarios, ya que nunca hay transiciones hacia ellos. La función de pago es exactamente la misma que en el caso clásico. Los gráficos de la figura 4.3 muestran la estrategia óptima para algunos puntajes del contrincante ($\beta = 0, 150, 180, 198$). Un documento con la totalidad de las gráficas se puede consultar en [6].

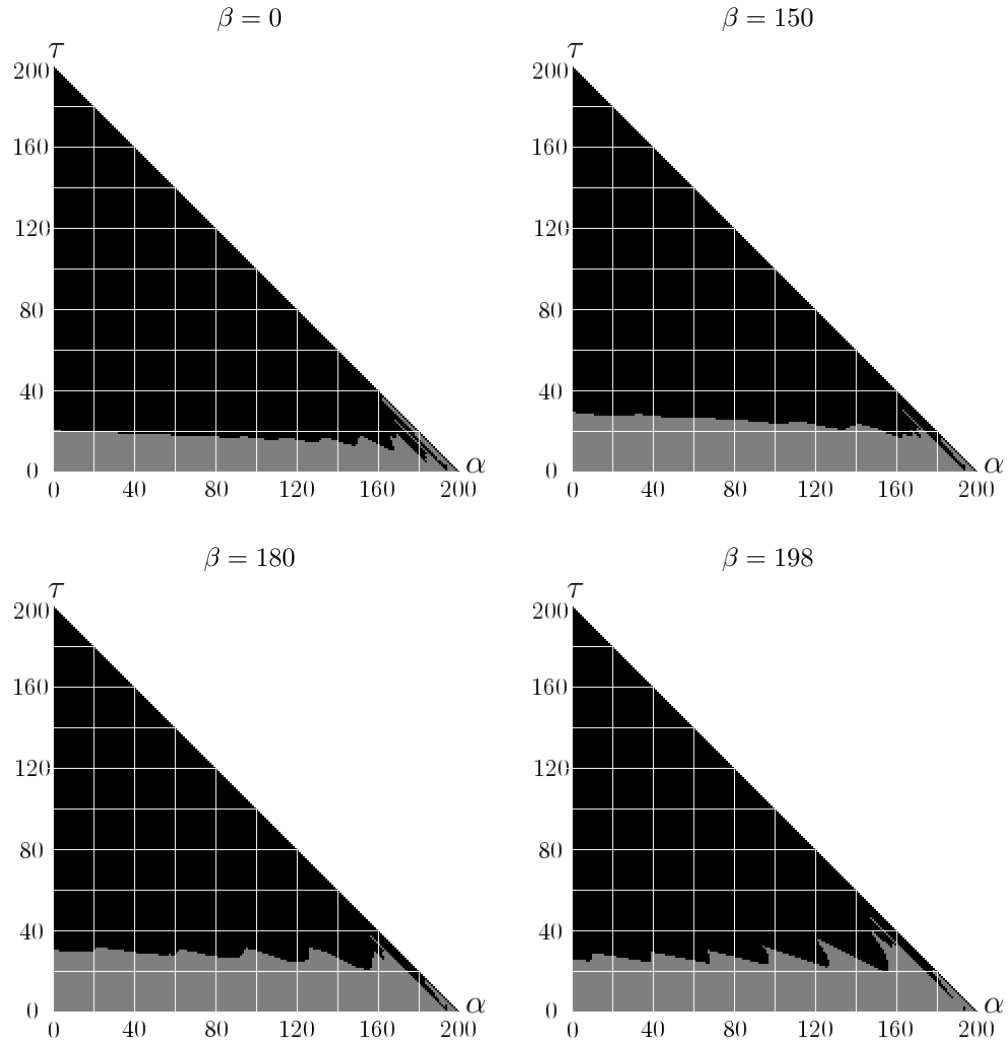


Figura 4.3: Parte de la estrategia óptima para el modelo con rebote, para puntaje del contrincante = 0,150,180 y 198. En la zona gris se debe *tirar* y en la zona negra *parar*

Algunos comentarios sobre las solución

- Al igual que en la estrategia óptima para el juego clásico se da que para valores bajos del puntaje del contrincante y del puntaje propio la estrategia

óptima es similar a *parar en 20*.

- En este caso con rebote, si bien se observa que la estrategia es un poco más arriesgada a medida que el juego se acerca al final, no ocurre el comportamiento extremo de tratar de ganar en un turno cuando el oponente está muy cerca de ganar. Esto es muy razonable, ya que en el modelo clásico estar muy cerca del final da una probabilidad altísima de ganar en un turno, mientras que en el caso con rebote, por cerca que se esté de llegar al objetivo, la probabilidad de ganar en cierto turno nunca es mucho mayor que $\frac{1}{6}$, porque al final hay que obtener un resultado exacto para poder ganar.
- Tener como puntaje acumulado 194 puntos da una ventaja muy importante, porque es un puntaje en que se puede ganar en un tiro (obteniendo un 6), pero aún no se corre riesgo de pasarse. Esa característica especial se nota en la estrategia óptima cuando el oponente tiene dicho puntaje, es decir la estrategia óptima para $\beta = 194$, es sensiblemente diferente que la de $\beta = 193$ e incluso que la de $\beta = 195$.

4.6.2. Maximizar valor esperado de la diferencia de puntaje

En "La codicia" el objetivo es ganar. Genera el mismo beneficio ganar cuando el contrincante tiene 5 puntos que cuando el contrincante tiene 198. El juego que queremos estudiar ahora es igual a la codicia, salvo que el perdedor tiene que pagar al ganador lo que le faltó para llegar a 200 puntos. Por ejemplo, si el ganador fue el jugador 2, y el jugador 1 quedó con 170 puntos, el jugador 1 debe pagar al jugador 2 un monto de 30.

El modelo para este caso es igual al del juego clásico, salvo en la asignación de recompensas. Las recompensas en esta variante tampoco dependen de las acciones tomadas por los jugadores y son:

$$r(s) = \begin{cases} 200 - \beta & \text{si } s = (1, \alpha, \beta, \tau) \text{ con } \alpha + \tau \geq 200 \\ \alpha - 200 & \text{si } s = (2, \alpha, \beta, \tau) \text{ con } \beta + \tau \geq 200 \\ 0 & \text{en cualquier otro caso} \end{cases}$$

La figura 4.4 muestra la estrategia óptima para valores del oponente ($\beta = 0, 150, 170, 180$). La totalidad de las gráficas se puede consultar en [7].

Algunos comentarios sobre las solución

- En este caso es mucho más difícil entender el porqué de la estrategia óptima, ya que la función de pago es mucho más compleja que en las aplicaciones anteriores.

- Si bien las estrategias en este caso son bastante distintas a las del modelo clásico se repite el hecho de que para valores bajos del puntaje del contrincante y del puntaje propio la estrategia óptima es similar a *parar en 20*.
- Para puntajes del contrincante mayores a 182 la estrategia óptima sugiere tratar de ganar en un turno. Este comportamiento es para puntajes más bajos que en el caso clásico, que recién a los 187 puntos del contrincante ocurría esto.

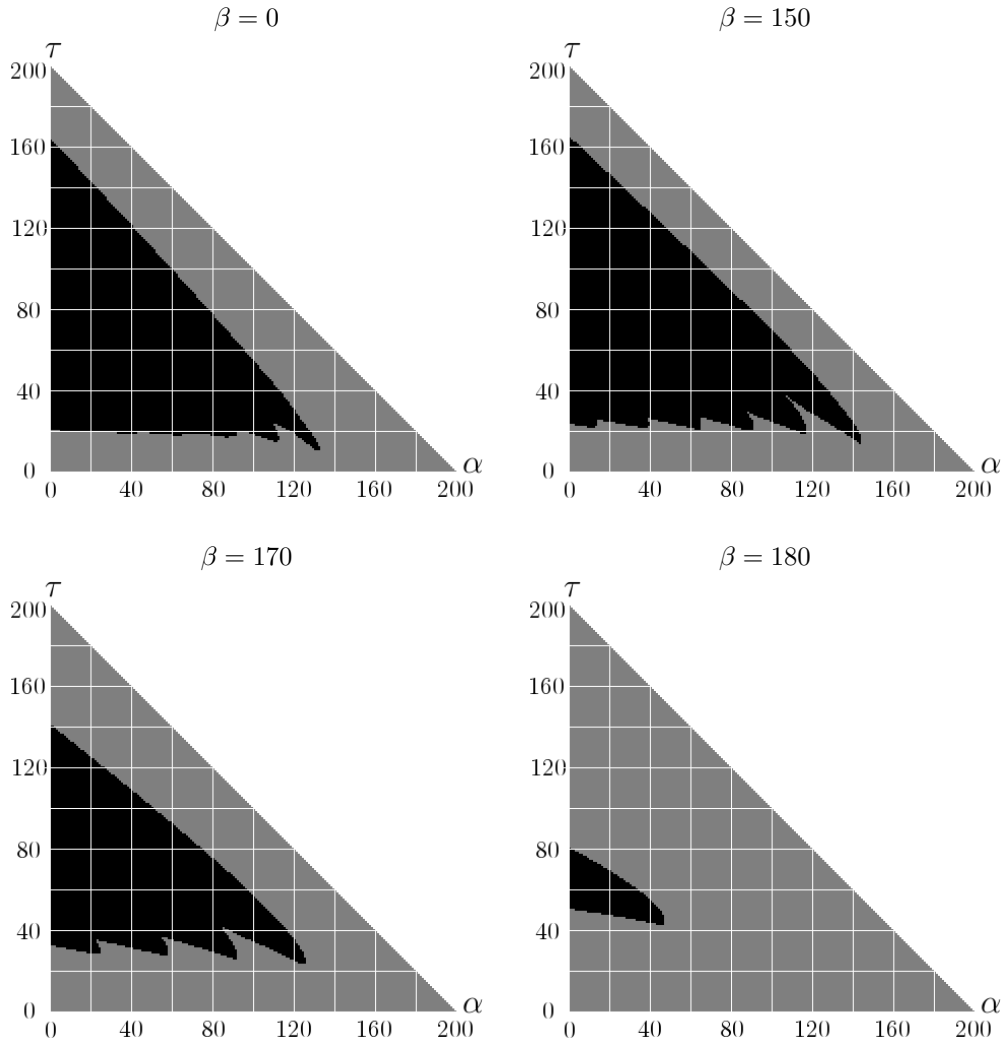


Figura 4.4: Parte de la estrategia óptima cuando se quiere maximizar la diferencia de puntaje, para puntaje del contrincante = 0,150,170 y 180. En la zona gris se debe *tirar* y en la zona negra *parar*

Apéndice A

Datos sobre las implementaciones realizadas

Para resolver los problemas que se presentan en el el capítulo 4 se realizaron cuatro programas. Uno para cada una de las variantes de "la codicia" y otro para simular las jugadas entre la estrategia óptima con la óptima solitaria. Todos ellos fueron implementados en el lenguaje C, un lenguaje muy potente y eficiente para la implementación de algoritmos.

Los programas para hallar las estrategias óptimas en las diferentes variantes de la codicia tienen una estructura muy similar. Las principales variables del programa son:

- vector v: el vector v se utiliza para iterar aplicando el operador U y hallar el valor del juego; es un arreglo de números reales indexado en α, β y τ; si bien para modelar los estados del juego se precisaría una variable más, para indicar el jugador, por la simetría del juego alcanza con tener los valores del juego para los estados del jugador 1.
- estrategia: estrategia es un arreglo de caracteres, también indexado en α, β y τ; en cada posición del arreglo se indica si la acción óptima es *tirar* o *parar*

Un pseudocódigo de muy alto nivel de los mismos es el siguiente:

```
inicializar el vector v
repetir
   actualizar v aplicando U
   calcular estrategia
hasta (condición de parada)
generar gráficos
```

A continuación vemos con un grado mayor de detalle qué es cada línea:

- `inicializar el vector v`: en el caso del modelo clásico de la codicia y del modelo con rebote inicializamos todas las entradas del vector v en 0,5, para el caso del modelo con rebote las inicializamos en 100. Cabe aclarar que si se inicializa en otros valores el resultado final no cambia, aunque puede tardar más en estabilizarse el vector.
- `actualizar v aplicando U`: Consiste en aplicar el operador U, definido en cada caso, al vector v y guardar el resultado en el mismo vector v, para esto se requiere iterar en los estados e ir guardando los resultados de Uv en cada estado en una estructura auxiliar.
- `calcular estrategia`: En cada paso de la repetición se obtiene un nuevo vector v, lo que permite calcular la estrategia que correspondería si ese fuera el vector de valor; con ese procedimiento se va obteniendo una secuencia de estrategias que va convergiendo a la óptima; la razón por la que calculamos estrategias intermedias es para utilizarlas en la condición de parada.
- `condición de parada`: Este es un aspecto fundamental, la condición de parada establece cuándo consideramos suficiente la cantidad de iteraciones y detenemos el algoritmo; más abajo vemos los criterios que se utilizaron.
- `generar gráficas`: Simplemente es una pequeña subrutina que genera gráficas mostrando la estrategia óptima; se utilizó la biblioteca "pngwriter" para hacer los dibujos.

Condición de parada

Si uno tuviera un poder de cálculo que le permita hacer operaciones con números reales sin ningún margen de error, resultaría ideal iterar, aplicándole el operador U al vector v, tantas veces como le de el tiempo; y de esa forma obtener la mejor aproximación, a la que esté dispuesto a esperar, del punto fijo de U, que hemos llamado v^*. Como la cantidad de números reales que representa la computadora es finita y las operaciones tienen error sucede que a partir de un momento seguir aplicando el operador U no genera una mejor aproximación del vector v^*, porque la diferencia que se generaría entre Uv y v es similar al error introducido por la máquina.

En nuestra aplicación el problema mencionado en el párrafo anterior no genera un perjuicio importante, ya que la cantidad de *estrategias estacionarias puras* es finita, lo que sugiere que a partir de un momento podemos alcanzar la estrategia óptima, aún sin tener el vector v^* fielmente calculado.

Al principio, la condición de parada elegida había sido $||Uv-v|| < \varepsilon$, es decir, parar cuando aplicar U al vector v genere una diferencia pequeña; ajustando el valor de ϵ con valores pequeños se obtuvieron buenos resultados. Después decidimos calcular en cada paso de la iteración la estrategia que generaría parar en ese

paso y compararla con la estrategia obtenida en el paso anterior, mostrando en pantalla la cantidad de cambios (es decir, en cuántos de los 4 000 000 de estados se modifica la acción óptima); si en una secuencia de pasos la estrategia permanece incambiada es porque convergió a la estrategia óptima. Observando la cantidad de iteraciones que requería que se estabilice la estrategia se estimó que parar cuando en 100 iteraciones consecutivas no se obtienen cambios es más que razonable.

La tabla a continuación muestra cuántas iteraciones requirió cada caso (contando las últimas 100 en que no se modifica la estrategia), y cuál fue la variación que se obtuvo al aplicar U en la última de las iteraciones.

	codicia clásica	con rebote	máxima diferencia
cantidad de iteraciones	444	457	743
$\|\|v - Uv\|\|$ en la última	1,8 10^{-7}	2,9 10^{-7}	1,5 10^{-5}

Si el lector quisiera el código de los programas implementados puede solicitármelos por correo electrónico a `fabian@cmat.edu.uy`.

Conclusiones

Generales

El estudio de los *juegos estocásticos transitorios* supuso un importante esfuerzo para adquirir los conceptos previos sobre *teoría de juegos*, y *juegos estocásticos* en general, ya que éstos son conceptos nuevos para mi. Otro aspecto que resultó ser una dificultad es la falta de material sobre el caso transitorio de los *juegos estocásticos*. Como fue mencionado en la introducción, el único libro que encontramos que incluye este tema es *Competitive Markov Decision Processes* [10], en él el modelo transitorio se trata entre muchos otros que son similares, y algunos resultados que allí aparecen no están demostrados al detalle sino que se hace referencia a ideas de resultados anteriores. De modo que poder "aislar" este caso en un trabajo autocontenido requirió reestructurar el tema, cambiar demostraciones, etc. En particular, el modelo planteado no es exactamente el que aparece en [10], sino que es una generalización del modelo para el caso no competitivo de [13].

El principal resultado de este trabajo es que los *juegos estocásticos transitorios* siempre tienen un valor y hay estrategias óptimas para ambos jugadores, además el teorema 3.3.1 sugiere una forma de hallar tales estrategias. Es interesante y a la vez muy razonable, que para hallar las estrategias óptimas para muchos modelos de juegos estocásticos, en particular los transitorios, haya que resolver ecuaciones del estilo de las propuestas por Bellman en [1] para resolver problemas de control óptimo en procesos de Markov. La ecuación (3.2) tiene la misma idea de fondo que las "ecuaciones de Bellman", ya que se optimiza teniendo en cuenta a la vez la ganancia instantánea y el valor esperado de la ganancia futura; claro que optimizar en el caso competitivo es resolver un juego matricial, mientras que en el caso de un jugador es simplemente tomar máximo o mínimo. También es interesante observar que la estrategia óptima se pueda encontrar entre las estrategias estacionarias, que, al menos a priori, es una clase de estrategias muy restrictiva.

Sobre la aplicación

El método para hallar la solución óptima en un *juego estocástico transitorio de suma cero*, dado en el teorema 3.3.1, es sin duda un resultado teórico de mucho valor; pero en la práctica sería inútil si hay que valerse de papel y lápiz para resolver un juego. Sin embargo, en alrededor de un minuto, y con un algoritmo relativamente sencillo, la computadora llega a hallar la estrategia óptima para las diferentes variantes del juego estudiado. Se recomienda consultar las estrategias completas, que como fue mencionado se encuentran en [5, 6, 7], y comparar los resultados con lo que dice la intuición.

Si bien la aplicación presentada deja ver el potencial de los resultados teóricos, estos últimos permiten una clase de juegos más general, en que los pagos instantáneos dependan de las acciones de los jugadores y ambos jugadores tomen decisiones en un mismo momento en lugar de ir alternando por turnos.

Una posible variante de "la codicia", que surgió en el seminario sobre juegos estocásticos mencionado en la introducción, y "aprovecha más" el modelo, es la siguiente: En cada turno ambos jugadores deben elegir la cantidad de dados que quieren tirar (de 1 a 10 por ejemplo) y si en alguno de sus dados sale un as anotan cero, de lo contrario anotan la suma de sus dados. En este modelo deciden a la misma vez los dos jugadores y no se tiene el problema de parada óptima de un turno. Su planteo matemático es una consecuencia directa de los resultados del capítulo 3, y su implementación quedó planteada como un posible trabajo futuro.

Bibliografía

[1] R. Bellman: *Dynamic Programming*, Princeton University Press, New Jersey, 1957.

[2] H. Bortolossi, G. Garbagio y B. Sartini: *Uma introdução à teoria econômica dos jogos*, Publicações matemáticas, 26º Colóquio Brasileiro de Matemática, IMPA, 2007.

[3] E. Cinlar, *Introduction to Stochastic Processes*, Prentice-Hall, Inc., Englewood Cliffs, New Yersey, USA, 1975.

[4] F. Crocce: *Trabajo monográfico. Procesos de Markov controlados; la aplicación a un juego de dados*, 2007, disponible en

`http://www.cmat.edu.uy/cmat/monografias/2007/fabian.pdf`.

[5] F. Crocce: *Estrategia óptima en "La Codicia"*, disponible en

`http://www.cmat.edu.uy/cmat/docentes/g1/fabian/archivos/codicia_optima.pdf`.

[6] F. Crocce: *Estrategia óptima para la variante de "La Codicia" en que hay rebote*, disponible en

`http://www.cmat.edu.uy/cmat/docentes/g1/fabian/archivos/rebote_optima.pdf`.

[7] F. Crocce: *Estrategia óptima en la variante de "La Codicia" en que se intenta maximizar la diferencia de puntaje*, disponible en

`http://www.cmat.edu.uy/cmat/docentes/g1/fabian/archivos/maxdif_optima.pdf`.

[8] E.B. Dynkin y A.A.Yushkevich, *Controlled Markov Processes*, Springer-Verlag New York Inc., U.S.A., 1979.

[9] W. Feller: *An Introduction to Probability Theory and Its Applications*, volumen 1, 3th Ed. John Wiley & Sons, Inc. 1968.

[10] J. Filar y K. Vrieze: *Competitive Markov Decision Processes*, Springer-Verlang, New York, 1997.

[11] D. Gillette: Stochastic Games with Zero Stop Probabilities. En A. Tucker, M. Dresher y P. Wolfe, editores, *Contributions to the Theory of Games,* Princeton University Press, Princeton, New Yersey 1957. Annals of Mathematics Studies 39.

[12] J. Haigh y M. Roters: Optimal Strategy in a Dice Game. *Journal of Applied Probability*, volumen 37, pp. 1110-1116, 2000.

[13] O. Hernández-Lerma y J.B. Lasserre: *Discrete-Time Markov Control Processes*, Springer, New York, 1996.

[14] J. Milnor: Analytics Proofs of the "Hairy Ball Theorem" and the Brouwer Fixed Point Theorem. *The American Mathematical Monthly*, volumen 85, nº 7, pp. 521-524, 1978.

[15] J. Nash: Equilibrium Points in n-Person Games, *Proc. Natl. Acad. Sci. USA.* 36(1), pp. 48–49, 1950.

[16] J. Nash: Non-Cooperative Games, dissertation, 1950, disponible en

`http://www.princeton.edu/mudd/news/faq/topics/Non-Cooperative_Games_Nash.pdf.`

[17] J. Nash: Non-Cooperative Games, *Annals of mathematics*, volumen 54, nº 2, 1951.

[18] T. Neller y C.Presser: Optimal Play of the Dice Game Pig. *The UMAP Journal*, **25.1**, pp. 25-47, 2004.

[19] J. von Neumann: Zur Theorie der Gesellschaftsspiele. *Mathematische Annalen, volumen 100, pp. 295-320.* Traducido por S. Bargmann: On the Theory of Games of Strategy, en A. Tucker y R. Luce, editores: *Contributions to the Theory of Games*, volumen 4, pp. 13-42, Princeton University Press, 1959.

[20] J. von Neumann y O. Morgenstern: *Theory of Games and Economic Behavior*, Princeton University Press, 1944.

[21] H. Nikaidô: On von Neumann's minimax theorem. *Pacific J. Math*, volumen 4, nº 1, pp. 65-72, 1954, disponible en

`http://projecteuclid.org/euclid.pjm/1103044955.`

[22] V. Petrov y E. Mordecki, *Teoría de la Probabilidad*, 2ª edición, Montevideo: DIRAC, 2008.

[23] M. Roters: Optimal Stopping in a Dice Game. *Journal of Applied Probability*, volumen 35, pp. 229-235, 1998.

[24] L.S. Shapley: Stochastic games *Proc. Nat. Acad. Science*, volumen 39, pp. 1095-1100, 1953.

[25] A.N. Shiryaev: *Probability*, 2nd. ed, Springer, 1996.

[26] T. Turocy y B. von Stengel, *Game Theory*, CDAM Research Report, LSE-CDAM-2001-09, 2001, disponible en

`http://www.cdam.lse.ac.uk/Reports/Files/cdam-2001-09.pdf.`

[27] M. Wschebor, *Notas para el curso de Introducción a los Procesos Estocásticos*, Centro de Matemática, UDELAR, 2001, disponible en

`http://www.cmat.edu.uy/~wschebor/Archivos/notas2.pdf.`

Índice alfabético