Trabajo Monográfico

Procesos de Markov controlados:

la aplicación a un juego de dados

Fabián Crocce

Orientador: Ernesto Mordecki
Centro de Matemática

Licenciatura en Matemática
Facultad de Ciencias
Universidad de la República
Uruguay

17 de mayo de 2007

Resumen

Un juego de dados tradicional de dos jugadores es motivación de un artículo de M. Roters [8] y otro de J.Haigh y M.Roters [5], donde se estudian aplicaciones de teoría de parada óptima y procesos de decisión de Markov a la búsqueda de estrategias óptimas para un jugador. Este trabajo desarrolla algunos resultados preliminares (partiendo de [1], [3] y [6]); además estudia y presenta dichos artículos. También se propone una estrategia heurística para el juego y comparaciones, basadas en simulaciones, con las estrategias óptimas que surgen de los artículos citados.

Palabras Claves: parada óptima; procesos de Markov controlados; teoría de decisión de Markov; juegos de dados; teoría de juegos.

Abstract

A tradicional two player dice game motivates the articles by Roters [8] and Haigh and Roters [5], where an optimal stopping problem and a control Markov problem are used, respectively, to find optimal strategies for one player. In this work we develope some preliminary results (departing from [1], [3] and [6]) and study and expose the contents of this two articles. Finally, we propose a heuristic strategy for the two player game, with corresponding simulations estimating the probability to win for a player that plays against the one player optimal strategies described previously.

Keywords: controlled Markov processess; dice games; game theory; Markov decision theory; optimal stopping.

Índice general

Introducción

Motivación

El problema original que motivó este trabajo fue un juego de dados conocido como el "diezmil", que se juega entre varios jugadores con cinco dados, y la dinámica es de un típico problema de parada óptima: se va acumulando puntaje con sucesivas tiradas de dados y luego de cada paso se puede optar por anotar el puntaje acumulado, y ceder el turno al siguiente, o arriesgarse a perder el puntaje con otra tirada para tratar de incrementarlo. El objetivo es alcanzar determinada cantidad de puntos antes que el resto de los jugadores. El problema que surge es decidir cuando plantarse.

Sobre este juego no conocemos trabajos escritos. En cambio hay dos artículos que estudian un juego, también de dados, con reglas mucho más simples pero con la misma idea de fondo, un problema de parada óptima. Yo lo conocía como "La Codicia", pero considerando que búsquedas en la web en conocidos y prestigiosos buscadores no me dieron resultados alentadores, estoy sospechando que ese nombre lo inventé. Después alguien me dijo que se llamaba "El Uno" y tuve más suerte, encontré foros discutiendo como jugar, una verdadera sociedad de fanáticos. De todos modos en este trabajo me voy a dar el gusto de llamarlo "La Codicia".

La Codicia

Objetivo

Los jugadores tienen por objetivo alcanzar cierto puntaje (T). La forma de alcanzarlo es mediante la acumulación de los puntos ganados en cada turno. Se sortea como será la ronda de manera justa ya que comenzar aumenta la probabilidad de ganar.

Un turno

El jugador tira un dado todas las veces que deseé mientras no obtenga un as como resultado. El puntaje del turno es cero si la secuencia terminó por la aparición de un uno, y es la suma de los resultados de los dados en las sucesivas tirada si el turno terminó por decisión del jugador. O sea que el fin del turno está marcado por la aparición de un uno en el dado o la decisión del jugador de plantarse. Al jugador, luego de cada tirada se le plantea la encrucijada: ¿Anoto lo que gané hasta ahora o arriesgo la ganancia para tratar de aumentarla?

Este trabajo

Consta de tres capítulos además de la introducción. A continuación se hace una breve introducción a cada uno explicitándose las principales fuentes.

Primero: El problema de un turno

Cuando se quiere buscar una buena estrategia para jugar a la codicia surge naturalmente la idea de tratar de hacer muchos puntos por turno para llegar lo antes posible al puntaje objetivo.

Este planteo sirve de motivación para estudiar un problema de parada óptima y su aplicación a este ejemplo concreto, que está desarrollado en el capítulo uno y se basa, principalmente, en un artículo de M. Roters [8]. Además se incluyen las pruebas de los resultados en los que se basa dicho artículo, tomadas esencialmente de un libro de Y. Chow y otros [1]

Segundo: La menor cantidad de turnos

Si bien el problema que se aborda en el capítulo uno ayuda a la hora de buscar buenas estrategias para jugar a la codicia, tiene el problema de concentrarse en un solo turno. El problema que surge ahora es encontrar una estrategia de juego que minimice la esperanza de la cantidad de turnos que requiere alcanzar el puntaje objetivo T.

En un artículo de J. Haigh y M. Roters [5] se estudia este problema como un problema de Procesos de Markov Controlados y se aplican resultados de esta teoría para su resolución concreta. En el capítulo dos se expone dicho artículo y se prueban, en base a los libros de E. Dynkin [3] y O. Hernandez-Lerma y J. Lasserre [6] los resultados en que éste se apoya.

Tercero: Ganarle al contrincante

Todo lo anterior no resuelve el problema concreto de ganarle al contrincante. A priori puede parecer que la estrategia hallada en el capítulo anterior es óptima en el sentido de maximizar la probabilidad de ganar, pero no lo es. Para mostrar esto, en el tercer capítulo se busca una estrategia heurística que mejore la obtenida en el segundo capítulo. Es decir, si un jugador (A) juega con la estrategia obtenida en el capítulo 2 y otro (B) con la obtenida en el capítulo 3, resulta que es más probable que gane el segundo; o sea que la probabilidad de que gane B es mayor a 0,5. Los resultados de este capítulo no se justifican matemáticamente sino que están comprobados empíricamente en base a simulaciones hechas en computadora.

Capítulo 1

Un problema de parada óptima: Maximizar el puntaje de un turno

1.1. Introducción

En este capítulo se aborda un problema de parada óptima que tiene interés en sí mismo, además de aplicarse al problema concreto de maximizar la esperanza de un turno en nuestro juego.

La teoría de la parada óptima busca resolver problemas del tipo: encontrar un momento para tomar determinada acción, basándonos en observaciones a una sucesión de variables aleatorias, con el fin de maximizar una ganancia esperada o minimizar un costo esperado. Una introducción básica con ejemplos sobre esto se puede leer en un libro de T. Ferguson [4] que se encuentra disponible en su página web.

Un turno, en el juego de la codicia, podemos modelarlo mediante variables aleatorias que representen el resultado del dado. Y la ganancia es la suma de dichas variables aleatorias en el caso de que no haya aparecido un uno y cero si es que algún resultado fue uno. La cantidad de variables aleatorias que se necesitan para modelar el turno no está definida a priori.

Para ser más precisos, pensemos que la variable X_i es el resultado del i-ésimo dado. Entonces, si tiramos tres veces, la ganancia que obtenemos es $Y_3 = X_1 + X_2 + X_3$ si los X_i son distintos de uno, e $Y_3 = 0$ en otro caso. Después de n tiradas podemos optar por plantarnos y acumular Y_n puntos o seguir para ver si podemos aumentar la ganancia. El problema que nos planteamos es en-

contrar un tiempo de parada τ^* de modo que la esperanza de la ganancia Y_{τ^*} sea máxima.

1.2. Conceptos previos

Se asume que el lector conoce ciertos aspectos básicos de la teoría de probabilidad: los axiomas, el concepto de variable aleatoria, valor esperado, etc. Sobre esto se puede leer en el libro de V. Petrov y E. Mordecki [7].

1.2.1. Definiciones básicas

Las definiciones que aparecen a continuación fueron tomadas de notas de M. Wschebor [9] para el curso de procesos estocásticos. También se pueden consultar en el libro de E.Cinlar [2].

Definición 1.2.1 (Independencia de un variable aleatoria respecto a una σ-álgebra)**.** Si X es una variable aleatoria en $(\Omega, \mathfrak{F}, \mathbb{P})$ y $\mathfrak{F}_1$ es una sub-σ-álgebra de $\mathfrak{F}$ diremos que X es independiente de la σ-álgebra $\mathfrak{F}_1$ si para cada $A \in \mathfrak{F}_1$ la variable aleatoria $\mathbb{I}_A$ (indicatriz del conjunto A) es independiente de X.

Definición 1.2.2 (Esperanza condicional)**.** Si se tiene un espacio de probabilidad $(\Omega, \mathfrak{F}, \mathbb{P})$, una variable aleatoria, $X : \Omega \to \overline{\mathbb{R}}$, $X \in L^1$ y $\mathfrak{F}'$ una sub-σ-álgebra de $\mathfrak{F}$ podemos definir una medida finita signada (μ) en $\mathfrak{F}'$ de la siguiente manera:

$$\mu(A) := E(X\mathbb{I}(A))$$

Dado que esta medida es absolutamente continua respecto de la medida de probabilidad (restringida en $\mathfrak{F}'$) existe la derivada de Radon-Nikodim $\frac{d\mu}{dP}$ de μ respecto a P.

Definimos la esperanza condicional de X respecto de $\mathfrak{F}'$:

$$E(X|\mathfrak{F}') := \frac{d\mu}{dP}.$$

De este modo $E(X|\mathfrak{F}')$ resulta ser una función de Ω en $\overline{\mathbb{R}}$ que es medible con respecto a $\mathfrak{F}'$ y cumple:

$$\int_A X dP = \int_A E(X|\mathfrak{F}')dP \qquad \forall A \in \mathfrak{F}'.$$

Además si se tuviera otra función $\mathfrak{F}'$-medible que cumpla la condición anterior, sería igual a $E(X|\mathfrak{F}')$ casi seguramente respecto de P.

Algunas propiedades de la esperanza condicional

1. El funcional que asocia a X su esperanza condicional respecto de $\mathfrak{F}'$ es lineal y monótono.
2. Si $\mathfrak{F}_1$ y $\mathfrak{F}_2$ son sub-σ-álgebras de $\mathfrak{F}$ de modo que $\mathfrak{F}_1 \subseteq \mathfrak{F}_2$ vale:
$$E(X|\mathfrak{F}_1) = E(E(X|\mathfrak{F}_2)|\mathfrak{F}_1).$$
3. Si Y es una variable aleatoria medible con respecto a $\mathfrak{F}'$ y $XY \in L^1$, entonces
$$E(XY|\mathfrak{F}') = YE(X|\mathfrak{F}').$$
Es como si la variable medible saliera como constante.
4. Si X es independiente de la σ-álgebra $\mathfrak{F}'$, entonces $E(X|\mathfrak{F}') = E(X)$.

Definición 1.2.3 (Filtración)**.** Una filtración $\{\mathfrak{F}_n\}_{n\in\mathbb{N}}$ en un espacio de probabilidad $(\Omega, \mathfrak{F}, \mathbb{P})$ es una sucesión de sub σ-álgebras de $\mathfrak{F}$ creciente, o sea, $\mathfrak{F}_n \subset \mathfrak{F}_{n+1}$ para todo n en $\mathbb{N}$. Se puede definir de forma más general con otros conjuntos de índices, pero no nos interesa en este trabajo.

Definición 1.2.4 (Proceso estocástico adaptado)**.** Si tenemos una sucesión de variables aleatorias $\{X_n\}_{n\in\mathbb{N}}$ definidas en un mismo espacio de probabilidad $(\Omega, \mathfrak{F}, \mathbb{P})$ y $\{\mathfrak{F}_n\}_{n\in\mathbb{N}}$ es una filtración de $(\Omega, \mathfrak{F}, \mathbb{P})$ diremos que $\{X_n\}_{n\in\mathbb{N}}$ es un proceso estocástico adaptado a la filtración si se cumple que X_n es $\mathfrak{F}_n$-medible para todo n en $\mathbb{N}$.

Definición 1.2.5 (Filtración generada por un proceso)**.** Si tenemos una sucesión de variables aleatorias $\{X_n\}_{n\in\mathbb{N}}$ en $(\Omega, \mathfrak{F}, \mathbb{P})$ y consideramos $\{\mathfrak{F}_n\}_{n\in\mathbb{N}}$ donde $\mathfrak{F}_n$ es la mínima sub-σ-álgebra de $\mathfrak{F}$ que hace medibles a $X_i : i = 1 \dots n$ tenemos, naturalmente, un proceso estocástico adaptado. A la filtración $\{\mathfrak{F}_n\}_{n\in\mathbb{N}}$ se le denomina "Filtración generada por el proceso $\{X_n\}_{n\in\mathbb{N}}$".

Definición 1.2.6 (Martingala, submartingala, supermartingala)**.** Diremos que un proceso estocástico $\{X_n\}_{n\in\mathbb{N}}$ adaptado a la filtración $\{\mathfrak{F}_n\}_{n\in\mathbb{N}}$ de $(\Omega, \mathfrak{F}, \mathbb{P})$ es una martingala (respectivamente submartingala, supermartingala) si cumple:

- $X_n^+ \in L^1 \qquad \forall n \in \mathbb{N}$ (podría haber sido $X_n^- \in L^1$ ó $X_n \in L^1$).
- $E(X_{n+1}|\mathfrak{F}_n) = X_n$ c.s. $\forall n \in \mathbb{N}$ (respectivamente $\geq$, $\leq$).

Definición 1.2.7 (Tiempo de parada)**.** Un tiempo de parada respecto a una filtración $\{\mathfrak{F}_n\}_{n\in\mathbb{N}}$ es una función $T : \Omega \to \mathbb{N} \cup \infty$ tal que, para todo natural n el conjunto $\{\omega : T(\omega) = n\}$ es medible con respecto a $\mathfrak{F}_n$. Y $P(T < \infty) = 1$.

1.2.2. Teoría de parada óptima

Los resultados que aparecen en está sección fueron tomados, esencialmente, de un libro de Y. Chow [1].

Definición 1.2.8. Dado $\{X_n\}_{n\in\mathbb{N}}$ un proceso en $(\Omega, \mathfrak{F}, \mathbb{P})$, $\{\mathfrak{F}_n\}_{n\in\mathbb{N}}$ una filtración de $\mathfrak{F}$, y T un tiempo de parada definimos

$$X_T := \sum_{n\in\mathbb{N}} X_n \mathbb{I}_{\{T=n\}}.$$

Definición 1.2.9 (Clase $\mathcal{C}$)**.** Si $(\Omega, \mathfrak{F}, \mathbb{P})$ es un espacio de probabilidad, $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias en él, y $\{\mathfrak{F}_n\}_{n\in\mathbb{N}}$ una filtración de $\mathfrak{F}$, definimos la clase $\mathcal{C}$ como la clase de tiempos de parada tales que:

$$EX_T^- < \infty.$$

Observación 1.2.10. *Para el caso en que las variables aleatorias son no negativas la clase $\mathcal{C}$ es la de todos los tiempos de parada en $(\Omega, \mathfrak{F}, \mathbb{P})$.*

Definición 1.2.11 (Caso monótono)**.** Sean $(\Omega, \mathfrak{F}, \mathbb{P})$ un espacio de probabilidad, $\{\mathfrak{F}_n\}_{n\in\mathbb{N}}$ una filtración en dicho espacio y $\{X_n\}_{n\in\mathbb{N}}$ un proceso estocástico adaptado a $\{\mathfrak{F}_n\}_{n\in\mathbb{N}}$.

Consideramos los conjuntos A_n definidos mediante:

$$A_n := \{\omega : E(X_{n+1}|\mathfrak{F}_n)(\omega) \leq X_n(\omega)\}.$$

Tiene lugar el caso monótono cuando se cumple:

$$A_i \subseteq A_{i+1} \quad (i = 1, 2, ...), \tag{1.1}$$

y

$$\bigcup_{i=1}^{\infty} A_i = \Omega. \tag{1.2}$$

Una interpretación intuitiva del caso monótono

Si X_n representa la ganancia en el instante n, A_n es el suceso "lo que gané hasta

el momento supera lo que lo que espero ganar si sigo un paso más". En el caso monótono, en virtud de (1.2)*, existe un n tal que* $\omega \in A_n$*. Además si* $\omega \in A_n$*, de la ecuación* (1.1)*, resulta que* $\omega \in A_{n+1}$*. Por lo tanto se pueden distinguir dos etapas: la ganancia esperada crece hasta el primer n en que* $\omega \in A_n$ *y luego decrece.*

Consideremos $\{X_n\}_{n\in\mathbb{N}}$ una sucesión de variables aleatorias en $(\Omega, \mathfrak{F}, \mathbb{P})$ tales que X_n pertenece a L^1 para todo natural n y $\{\mathfrak{F}_n\}_{n\in\mathbb{N}}$ una filtración de $\mathfrak{F}$ de modo que el proceso sea adaptado. El resto de esta sección está dedicado a demostrar el siguiente resultado:

Teorema 1.2.12. *En el caso monótono sea:*

$$S(\omega) := \text{ínf}\{n \in \mathbb{N} : X_n \geq E(X_{n+1}|\mathfrak{F}_n)\}. \tag{1.3}$$

Supongamos que S pertenece a la clase $\mathcal{C}$ *y se cumple:*

$$\liminf_{n\to\infty} \int_{\{S>n\}} X_n^+ dP = 0, \tag{1.4}$$

y que T también pertenece a la clase $\mathcal{C}$ *y se verifica*

$$\liminf_{n\to\infty} \int_{\{T>n\}} X_n^- dP = 0, \tag{1.5}$$

entonces, $EX_S \geq EX_T$*.*

Es decir: S es un tiempo de parada óptimo entre los de la clase $\mathcal{C}$ *que verifican la condición* (1.5)*.*

Observación 1.2.13. *En el caso monótono, S, definida como en* (1.3)*, es un tiempo de parada.*

Demostración. Veamos que el suceso $\{S = n\} \in \mathfrak{F}_n$: para eso observemos que S toma el valor n si y sólo si $X_i < E(X_{i+1}|\mathfrak{F}_i)$ para los i menores que n y $X_n \geq E(X_{n+1}|\mathfrak{F}_n)$. Por lo tanto, se tiene que

$$\{S = n\} = \cap_{i<n}\{X_i < E(X_{i+1}|\mathfrak{F}_i)\} \cap \{X_n \geq E(X_{n+1}|\mathfrak{F}_n)\},$$

donde claramente los conjuntos que se intersectan son medibles con respecto a $\mathfrak{F}_n$, ya que tanto X_i como $E(X_{i+1}|\mathfrak{F}_i)$ es medible con respecto a $\mathfrak{F}_i$, y $\mathfrak{F}_i \subset \mathfrak{F}_n$ para i menor o igual que n.

La otra condición que se debe verificar, de la definición de tiempo de parada, es $P(S < \infty) = 1$, que es inmediata a partir de la condición $\bigcup_{i=1}^{\infty} A_i = \Omega$, que

se cumple por estar en el caso monótono.

□

Corolario 1.2.14. *En el caso monótono, si para todo n natural $X_n \geq 0$ casi seguramente, y se cumple*

$$\liminf_{n\to\infty} \int_{\{S>n\}} X_n dP = 0, \tag{1.6}$$

entonces

$$EX_S \geq EX_T \text{ para todo tiempo de parada } T.$$

Demostración. Como X_n es mayor o igual que cero casi seguramente, la clase $\mathcal{C}$ resulta ser la clase de todos los tiempos de parada. Por lo tanto, la condición de que S pertenezca a la clase $\mathcal{C}$ se satisface automáticamente. Igualmente, por ser la variable no negativa casi seguramente, todos los tiempos de parada cumplen (1.5); además la condición (1.4) coincide con (1.6). Con estas consideraciones, del teorema 1.2.12 se deduce inmediatamente la tesis.

□

Lema 1.2.15. *Si S y T pertenecen a la clase $\mathcal{C}$ y para cada $n \geq 1$ se cumple:*

$$E(X_S|\mathfrak{F}_n) \geq X_n \text{ en el conjunto } \{S > n\} \tag{1.7}$$

y

$$E(X_T|\mathfrak{F}_n) \leq X_n \text{ en el conjunto } \{S = n, T \geq n\}, \tag{1.8}$$

entonces $EX_S \geq EX_T$.

Demostración. Basta observar que

$$\begin{aligned} EX_S &= \textstyle\int_{\{S<T\}} X_S + \int_{\{S\geq t\}} X_S \\ &= \textstyle\sum_{n=1}^{\infty} \int_{\{S=n<T\}} X_n + \sum_{n=1}^{\infty} \int_{\{S\geq T=n\}} E(X_S|\mathfrak{F}_n) \\ &\geq \textstyle\sum_{n=1}^{\infty} \int_{\{S=n<T\}} E(X_T|\mathfrak{F}_n) + \sum_{n=1}^{\infty} \int_{\{S\geq T=n\}} X_n = EX_T. \end{aligned}$$

□

Lema 1.2.16. *Sea S un tiempo de parada tal que para cada natural n*

$$E(X_{n+1}|\mathfrak{F}_n) \geq X_n \text{ en } \{S > n\},$$

entonces $\{X_{min(S,n)}, \mathfrak{F}_n, 1 \leq n < \infty\}$ es submartingala.

Si además existe EX_S *y*

$$\liminf_{n\to\infty} \int_{\{S>n\}} X_n^+ = 0$$

se verifica (1.7).

Demostración. Sea $S(n) = \min(S, n)$. Veamos que existe $EX_{S(n)}$:
Es evidente que:

$$EX_{S(n)}^+ \leq \sum_{i=1}^{n} EX_i^+ < \infty,$$

entonces existe $EX_{S(n)}$.

Para cada $A \in \mathfrak{F}_n$

$$\begin{aligned} \textstyle\int_A X_{S(n)} &= \textstyle\int_{A\{S\leq n\}} X_S + \int_{A\{S>n\}} X_n \\ &\leq \textstyle\int_{A\{S\leq n\}} X_S + \int_{A\{S\geq n+1\}} X_{n+1} \\ &= \textstyle\int_A X_{S(n+1)} = \int_A E(X_{S(n+1)}|\mathfrak{F}_n) \end{aligned}$$

lo que implica que $\{X_{S(n)}\}$ es una $\{\mathfrak{F}_n\}$-submartingala.

Supongamos ahora que EX_S existe. Entonces para cualquier $A \in \mathfrak{F}_n$ $(n \in \mathbb{N})$ en virtud de la propiedad de martingala obtenemos que para cada $m > n$

$$\begin{aligned} \textstyle\int_{A\{S\geq n\}} X_n &= \textstyle\int_{A\{S\geq n\}} X_{S(n)} \leq \int_{A\{S\geq n\}} X_{S(m)} \\ &= \textstyle\int_{A\{n\leq S\leq m\}} X_S + \int_{A\{S>m\}} X_m \\ &\leq \textstyle\int_{A\{n\leq S\leq m\}} X_S + \int_{A\{S>n\}} X_m^+; \end{aligned}$$

ahora podemos considerar $m' \to \infty$ tal que:

$$\int_{\{S>m'\}} X_m^+ \to \liminf_{m\to\infty} \int_{\{S>m\}} X_m^+,$$

que es igual a cero por hipótesis. Obtenemos:

$$\int_{A\{S\geq n\}} X_n \leq \int_{A\{S\geq n\}} X_S = \int_{A\{S\geq n\}} E(X_S|\mathfrak{F}_n),$$

lo que prueba (1.7).

□

Lema 1.2.17. *Sean S y T tiempos de parada en la clase $\mathcal{C}$ tal que para cada $n \in \mathbb{N}$ se cumple*

$$E(X_{n+1}|\mathfrak{F}_n) \leq X_n \text{ en } \{S \leq n\}$$

y

$$\liminf_{n\to\infty} \int_{\{T>n\}} X_n^- = 0.$$

Entonces se verifica (1.8).

Demostración. Sea $r(n) = \text{máx}(S, \text{mín}(T, n))$. Veamos que existe $EX_{r(n)}$: Es claro que

$$EX_{r(n)}^- \leq \sum_{i=1}^{n} EX_i^- + EX_S^- < \infty$$

debido a que X_i son integrables y $S \in \mathcal{C}$. Entonces existe $EX_{r(n)}$. Y para cada $A \in \mathfrak{F}_n$ vale

$$\begin{aligned}
\textstyle\int_A X_{r(n)} &= \textstyle\int_{A\{S>\text{mín}(T,n)\}} X_S + \int_{A\{n\geq T\geq S\}} X_T + \int_{A\{T>n\geq S\}} X_n \\
&\geq \textstyle\int_{A\{S>\text{mín}(T,n)\}} X_S + \int_{A\{n\geq T\geq S\}} X_T + \int_{A\{T\geq n+1>S\}} X_{n+1} \\
&= \textstyle\int_A X_{r(n+1)} = \int_A E(X_{r(n+1)}|\mathfrak{F}_n),
\end{aligned}$$

lo que implica que $\{X_{r(n)}\}$ es una $\{\mathfrak{F}_n\}$-supermartingala.

Como $T \in \mathcal{C}$ tenemos que EX_T existe. Entonces para cualquier $A \in \mathfrak{F}_n$ ($n = 1, 2, \ldots$) podemos escribir, utilizando la propiedad de supermartingala, que para cada $m = n+1, n+2, \ldots$

$$\begin{aligned}
\textstyle\int_{A\{S=n\}\{T\geq n\}} X_n &= \textstyle\int_{A\{S=n\}\{T\geq n\}} X_{r(n)} \geq \int_{A\{S=n\}\{T\geq n\}} X_{r(m)} \\
&= \textstyle\int_{A\{S=n\}\{n\leq T\leq m\}} X_T + \int_{A\{S=n\}\{T>m\}} X_m \\
&\geq \textstyle\int_{A\{S=n\}\{n\leq T\leq m\}} X_T - \int_{A\{S=n\}\{T>m\}} X_m^-;
\end{aligned}$$

consideramos $m' \to \infty$ de forma que:

$$\int_{\{T>m'\}} X_m^- \to \liminf_{m\to\infty} \int_{\{T>m\}} X_m^-,$$

que es igual a cero por hipótesis. Obtenemos:

$$\int_{A\{S=n\}\{T\geq n\}} X_n \geq \int_{A\{S=n\}\{T\geq n\}} X_T = \int_{A\{S=n\}\{T\geq n\}} E(X_T|\mathfrak{F}_n)$$

lo que prueba (1.8).

□

Demostración del teorema 1.2.12. Como $S(\omega)$ está definido como el primer n tal que $\omega \in A_n$ y estamos en el caso monótono (los A_n están encajados), sabemos que $\omega \notin A_n$ en el conjunto $S(\omega) > n$ y $\omega \in A_n$ en el conjunto $S(\omega) \leq n$. Esto se traduce en:

$$E(X_{n+1}|\mathfrak{F}_n) \geq X_n \text{ en } \{S > n\}$$

y

$$E(X_{n+1}|\mathfrak{F}_n) \leq X_n \text{ en } \{S \leq n\}.$$

Además la esperanza de X_S existe dado que S está en la clase C, por lo tanto, en virtud de la hipótesis (1.4), estamos en las hipótesis del lema 1.2.16 de donde surge que vale (1.7).

Si además T está en la clase C y cumple (1.5), estamos en las hipótesis del lema 1.2.17 y vale (1.8).

Ahora estamos en condiciones de aplicar el lema 1.2.15 y se cumple $EX_S \geq EX_T$.

□

1.3. Un problema de parada óptima: Maximizar la ganancia esperada

La solución al problema de los dados planteado en la introducción surge como consecuencia de los resultados que se exponen en esta sección. De todos modos se dará, más adelante en el capítulo, una deducción directa de la solución. Todo lo que aparece desde aquí hasta el final del capítulo, exceptuando algunos resultados simulados, fue tomado del artículo de M. Roters [8].

A lo largo de esta sección consideraremos un espacio de probabilidad $(\Omega, \mathfrak{F}, \mathbb{P})$, $(\mathfrak{F}_n)_{n\in\mathbb{N}}$ una filtración de $\mathfrak{F}$, $\{X_n\}_{n\in\mathbb{N}}$ y $\{\beta_n\}_{n\in\mathbb{N}}$ sucesiones de variables aleatorias en dicho espacio; donde además se se cumplen las siguientes condiciones:

- La sucesión $\{X_n\}_{n\in\mathbb{N}}$ es de variables aleatorias independientes.
- La sucesión $\{\beta_n\}_{n\in\mathbb{N}}$ es de variables aleatorias independientes.
- Para todo $n \in \mathbb{N}$, X_n toma valores en $[0, +\infty)$ casi seguramente y β_n toma valores en [0,1] casi seguramente.
- $\{X_n\}_{n\in\mathbb{N}}$ y $\{\beta_n\}_{n\in\mathbb{N}}$ son procesos estocásticos adaptados.
- Tanto X_{n+1} como β_{n+1} son independientes de $\mathfrak{F}_n \quad \forall n \in \mathbb{N}$.

Definamos una nueva sucesión de variables aleatorias $\{Y_n\}_{n\in\mathbb{N}}$ de la siguiente forma:

$$Y_n := S_n\beta^{(n)},$$

donde

$$S_n := \sum_{i=1}^{n} X_i \text{ y } \beta^{(n)} := \prod_{i=1}^{n} \beta_i,$$

es decir, Y_n es la suma de los n primeros X_i multiplicado por el producto de los n primeros β_i.

Lo que se busca es, bajo ciertas condiciones, encontrar un tiempo de parada τ^* que resuelva el siguiente problema de optimalidad:

$$E(Y_{\tau^*}) = \sup_{\{\tau\in C\}} EY_\tau. \tag{1.9}$$

Para motivar el problema planteado, veamos, sin entrar en mucho detalle, como se aplicaría al el modelo planteado al caso de la codicia: La variable aleatoria X_n corresponde al resultado del dado en la n-ésima tirada. La variable β_n vale cero si X_n es uno y en otro caso vale uno. Observar que de esta forma la variable

Y_n es el puntaje obtenido en el turno si nos plantáramos en la n-ésima tirada, ya que S_n es la suma de los resultados de las n primeras tiradas y $\beta^{(n)}$ vale uno si no salió ningún uno y cero en otro caso. Por lo que el problema planteado en (1.9) coincide con el de encontrar un tiempo de parada óptimo para maximizar la esperanza de la cantidad de puntos del turno.

Teorema 1.3.1. *En las condiciones expuestas arriba, si se cumplen las siguientes condiciones:*

$$E(X_n\beta_n) < \infty, \quad E\beta_n < 1, \quad \forall n \in \mathbb{N}, \tag{1.10}$$

$$\left(\frac{E(X_n\beta_n)}{1 - E\beta_n}\right)_{n\in\mathbb{N}} \text{ es monótona no creciente,} \tag{1.11}$$

$$\sum_{n\in\mathbb{N}} \log P(\beta_n > 0) = -\infty \quad \acute{o} \quad P\left(\sum_{n\in\mathbb{N}} X_n = \infty\right) = 1, \tag{1.12}$$

se tiene que

$$\tau^* := \text{ínf}\,\{n \in \mathbb{N} : E(Y_{n+1}|\mathfrak{F}_n) \leq Y_n\} \tag{1.13}$$

resuelve el problema de optimalidad planteado en la ecuación (1.9)*:*

Demostración. Probaremos que Y_n está en el caso monótono, que Y_n es mayor o igual que cero casi seguramente y que cumple (1.6). Luego aplicaremos el corolario 1.2.14.

Veamos que Y_n es medible con respecto a $\mathfrak{F}_n$: X_n y β_n son $\mathfrak{F}_n$-medibles por hipótesis. Además como $\mathfrak{F}_i$ esta incluido en $\mathfrak{F}_{i+1}$ tenemos que β_i y X_i son medibles con respecto $\mathfrak{F}_i$ para todo i menor o igual que n. Como Y_n se obtiene mediante sumas y productos de funciones $\mathfrak{F}_n$-medibles resulta ser $\mathfrak{F}_n$-medible.

Para probar que estamos en el caso monótono debemos considerar los conjuntos $A_n = \{\omega : E(Y_{n+1}|\mathfrak{F}_n)(\omega) \leq Y_n(\omega)\}$. Veamos que se cumple (1.1) de la

definición de caso monótono:

$$\begin{aligned} E(Y_{n+1}|\mathfrak{F}_n) &= E(X_{n+1}\beta^{(n)}\beta_{n+1} + S_n\beta^{(n)}\beta_{n+1}|\mathfrak{F}_n) \\ &= E(X_{n+1}\beta^{(n)}\beta_{n+1}|\mathfrak{F}_n) + E(S_n\beta^{(n)}\beta_{n+1}|\mathfrak{F}_n) \\ &= \beta^{(n)}E(X_{n+1}\beta_{n+1}|\mathfrak{F}_n) + \beta^{(n)}S_nE(\beta_{n+1}|\mathfrak{F}_n) \\ &= \beta^{(n)}E(X_{n+1}\beta_{n+1}) + \beta^{(n)}S_nE\beta_{n+1} \\ &= (E(X_{n+1}\beta_{n+1}) + S_nE\beta_{n+1})\beta^{(n)} \end{aligned}$$

La primera igualdad surge de aplicar las definiciones de Y_{n+1}, S_{n+1} y $\beta^{(n+1)}$. La segunda de la monotonía de la esperanza condicional. La tercera de que los factores medibles con respecto a la σ-álgebra condición "salen como constantes". La cuarta de que las variables $X_{n+1}\beta_{n+1}$ y β_{n+1} son independiente de la σ-álgebra $\mathfrak{F}_n$, por lo que la esperanza condicional es igual a la esperanza.

De la igualdad anterior se obtiene que:

$$\begin{aligned} E(Y_{n+1}|\mathfrak{F}_n) \leq Y_n &\Leftrightarrow (E(X_{n+1}\beta_{n+1}) + S_nE\beta_{n+1})\beta^{(n)} \leq Y_n \\ &\Leftrightarrow (S_n - E(X_{n+1}\beta_{n+1}) - S_nE\beta_{n+1})\beta^{(n)} \geq 0 \\ &\Leftrightarrow ((1 - E\beta_{n+1})S_n - E(X_{n+1}\beta_{n+1}))\beta^{(n)} \geq 0 \\ &\Leftrightarrow S_n \geq \tfrac{E(X_{n+1}\beta_{n+1})}{1-E\beta_{n+1}} \quad \acute{o} \quad \beta^{(n)} = 0. \end{aligned} \tag{1.14}$$

Utilizando la hipótesis (1.11) y que S_n es no decreciente se deduce que:

$$S_n \geq \frac{E(X_{n+1}\beta_{n+1})}{1 - E\beta_{n+1}} \Rightarrow S_{n+1} \geq \frac{E(X_{n+2}\beta_{n+2})}{1 - E\beta_{n+2}}.$$

Además sabemos que $\beta^{(n)} = 0$ implica $\beta^{(n+1)} = 0$. Con esto queda probado $A_n \subseteq A_{n+1}$, o sea (1.1).

Ahora probaremos $P(\cup_{n\in\mathbb{N}}A_n) = 1$, o sea, que vale la ecuación (1.2):

Veamos primero

$$\cup_{n\in\mathbb{N}}A_n \supseteq \cup_{n\in\mathbb{N}}\{\beta_n = 0\} \cup \{\lim_{n\to\infty} S_n = \infty\} :$$

Si $\omega \in \{\beta_n = 0\}$ por (1.14) se deduce que $\omega \in A_n$ lo que prueba

$$\cup_{n\in\mathbb{N}}A_n \supseteq \cup_{n\in\mathbb{N}}\{\beta_n = 0\}.$$

Para ver que

$$\cup_{n\in\mathbb{N}} A_n \supseteq \{\lim_{n\to\infty} S_n = \infty\}$$

basta observar que, por la hipótesis (1.11), existe un natural a partir del cual

$$S_n \geq \frac{E(X_{n+1}\beta_{n+1})}{1 - E\beta_{n+1}},$$

ya que S_n tiende a infinito; entonces utilizando de nuevo la ecuación (1.14) obtenemos la inclusión deseada.

Si logramos probar

$$P(\cup_{n\in\mathbb{N}}\{\beta_n = 0\} \cup \{\lim_{n\to\infty} S_n = \infty\}) = 1$$

tenemos lo que queremos. La hipótesis (1.12) nos dice que o bien se cumple

$$P\left(\sum_{n\in\mathbb{N}} X_n = \infty\right) = 1,$$

que implica trivialmente la igualdad anterior, o se cumple

$$\sum_{n\in\mathbb{N}} \log P(\beta_n > 0) = -\infty,$$

que implica

$$\lim_{N\to\infty} \prod_{n=1}^{N} P(\beta_n > 0) = 0,$$

entonces

$$P(\omega \notin \cup_{n\in\mathbb{N}}\{\beta_n = 0\}) = 0$$

y también se cumple lo que queremos.

Hemos demostrado que estamos en el caso monótono. Que $Y_n \geq 0$ casi seguramente es evidente. Veamos que vale (1.6).

De la definición de τ^* y la ecuación (1.14) se deduce que

$$\tau^* = \text{ínf}\left\{n \in \mathbb{N} : S_n \geq \frac{E(X_{n+1}\beta_{n+1})}{1 - E\beta_{n+1}} \quad \acute{o} \quad \beta^{(n)} = 0\right\}. \tag{1.15}$$

Por lo tanto en $\{\tau^* > n\}$ tenemos que $S_n \leq \frac{E(X_{n+1}\beta_{n+1})}{1-E\beta_{n+1}}$ y considerando (1.11) tenemos $S_n \leq \frac{E(X_1\beta_1)}{1-E\beta_1}$. Entonces:

$$0 \leq Y_n \leq S_n \leq \frac{E(X_1\beta_1)}{1 - E\beta_1}. \tag{1.16}$$

Queremos ver que

$$\liminf_{n\to\infty} \int_{\{\tau^*>n\}} Y_n dP = 0,$$

pero

$$\liminf_{n\to\infty} \int_{\{\tau^*>n\}} Y_n dP = \liminf_{n\to\infty} \int_{\Omega} \mathbb{I}_{\{\tau^*>n\}} Y_n dP. \tag{1.17}$$

Para calcular el limite podemos, en virtud de la acotación obtenida en (1.16), utilizar el teorema de convergencia dominada. Tenemos $\tau^*(\omega) > n$ si y sólo si $\omega \notin \cup_{i=1}^n A_i$, pero $P(\cup_{i=1}^n A_i)$ crece a 1 cuando $(n \to \infty)$, por lo que el integrando en el segundo miembro de (1.17) decrece a 0 casi seguramente. Por lo tanto se cumple (1.6).

Ahora estamos en condiciones de aplicar 1.2.14, que asegura que $EY_{\tau^*} \geq EY_\tau$ para todo tiempo de parada τ, concluyendo la demostración.

□

Corolario 1.3.2. *Si a las condiciones del teorema anterior agregamos que las sucesiones* $\{X_n\}_{n\in\mathbb{N}}$ *y* $\{\beta_n\}_{n\in\mathbb{N}}$ *son de variables idénticamente distribuidas, que casi seguramente* β_n *toma valores en* $\{0,1\}$ *y que casi seguramente* X_n *toma valores mayores que cero, se cumple*

$$\tau^* = \tau_{k_0} := \text{ínf}\,\{n \in \mathbb{N} : Y_n \notin (0, k_0)\},$$

donde

$$k_0 := \frac{EY_1}{1 - E\beta_1}.$$

Demostración. Observar que

$$Y_n \notin (0, k_0) \;\; \Leftrightarrow Y_n = 0 \;\; \acute{o} \;\; Y_n \geq \tfrac{EY_1}{1-E\beta_1}.$$

Como $S_n > 0$ casi seguramente, tenemos que $Y_n = 0$ si y sólo si $\beta^{(n)} = 0$. Si $\beta^{(n)} \neq 0$ entonces $\beta^{(n)} = 1$ y $S_n = Y_n$. Además como $\{X_n\}$ y $\{\beta_n\}$ son idénticamente distribuidas $E(X_{n+1}\beta_{n+1}) = EY_1$ y $E\beta_{n+1} = E\beta_1$. Por lo tanto

$$Y_n \notin (0, k_0) \Leftrightarrow S_n \geq \frac{E(X_{n+1}\beta_{n+1})}{1 - E\beta_{n+1}} \;\; \acute{o} \;\; \beta^{(n)} = 0,$$

de (1.15) se deduce la tesis.

□

Definición 1.3.3 (c-reglas)**.** A los tiempos de parada de la forma

$$\text{ínf}\,\{n \in \mathbb{N} : Y_n \notin (0, c)\},$$

como el del teorema anterior, los denominamos *c-reglas* y los denotamos τ_c.

1.4. La ganancia esperada en un caso discreto

En esta sección consideramos, como caso particular de la sección anterior, las sucesiones $\{X_n\}_{n\in\mathbb{N}}$ y $\{\beta_n\}_{n\in\mathbb{N}}$ independientes e idénticamente distribuidas, agregando la hipótesis de que para todo n en $\mathbb{N}$, casi seguramente X_n toma valores en los naturales positivos y β_n toma valores en $\{0,1\}$.

Como aplicación de lo visto en la sección anterior sabemos que hay una *c-regla* que resuelve el problema de maximizar la esperanza de la ganancia; además es claro que ese tiempo de parada coincide, casi seguramente, con la *k'-regla* correspondiente a considerar k' como el primer entero mayor o igual que c. En pos de resolver el problema de calcular EY_{τ_k} para $k \in \mathbb{N}$ tenemos el siguiente teorema.

Teorema 1.4.1. *Si $\{X_n\}_{n\in\mathbb{N}}$ y $\{\beta_n\}_{n\in\mathbb{N}}$ son como se especifica arriba, e $\{Y_n\}_{n\in\mathbb{N}}$ se define como en la sección anterior, se cumple:*

$$EY_{\tau_{k+1}} - EY_{\tau_k} = (EY_1 - k(1 - E\beta_1))P(Y_{\tau_k} = k). \tag{1.18}$$

Demostración. Si $Y_{\tau_k} > k$ entonces $Y_{\tau_k} \geq k+1$ y $\tau_k = \tau_{k+1}$ ya que la primera vez que se alcanza o supera k coincide con la primera vez que se alcanza o supera $k+1$. En cambio si $Y_{\tau_k} = k$ se tiene que $\tau_{k+1} = 1 + \tau_k$ y $Y_{\tau_{k+1}} = k + X_{1+\tau_k}$ en el caso en que $\beta_{1+\tau_k} = 1$ y $Y_{\tau_{k+1}} = 0$ en el otro caso. Por lo tanto:

$$Y_{\tau_{k+1}} = Y_{\tau_k}\mathbb{I}_{\{Y_{\tau_k}>k\}} + (X_{1+\tau_k} + k)\mathbb{I}_{\{Y_{\tau_k}=k\}}\beta_{1+\tau_k}, \tag{1.19}$$

o lo que es lo mismo

$$Y_{\tau_{k+1}} - Y_{\tau_k} = -k\mathbb{I}_{\{Y_{\tau_k}=k\}} + (X_{1+\tau_k} + k)\mathbb{I}_{\{Y_{\tau_k}=k\}}\beta_{1+\tau_k}.$$

Tomando esperanza en ambos miembros se obtiene

$$EY_{\tau_{k+1}} - EY_{\tau_k} = -kP(Y_{\tau_k} = k) + E((X_{1+\tau_k} + k)\mathbb{I}_{\{Y_{\tau_k}=k\}}\beta_{1+\tau_k}). \tag{1.20}$$

Como $\{Y_{\tau_k} = k\} = \bigcup_{j\in\mathbb{N}}\{Y_j = k\} \cap \{\tau_k = j\}$ tenemos

$$\begin{aligned} E((X_{1+\tau_k} + k)\mathbb{I}_{\{Y_{\tau_k}=k\}}\beta_{1+\tau_k}) &= \textstyle\sum_{j\in\mathbb{N}} E((X_{j+1} + k)\mathbb{I}_{\{Y_j=k\}\cap\{\tau_k=j\}}\beta_{j+1}) \\ &= \textstyle\sum_{j\in\mathbb{N}} E(X_{j+1}\beta_{j+1}\mathbb{I}_{\{Y_j=k\}\cap\{\tau_k=j\}}) \\ &\quad + \textstyle\sum_{j\in\mathbb{N}} kE(\beta_{j+1}\mathbb{I}_{\{Y_j=k\}\cap\{\tau_k=j\}}). \end{aligned}$$

Como $\mathbb{I}_{\{Y_j=k\}\cap\{\tau_k=j\}}$ es medible con respecto a la σ-álgebra $\mathfrak{F}_j$ y tanto X_{j+1} como β_{j+1} son independientes de la σ-álgebra $\mathfrak{F}_j$ de la igualdad anterior se obtiene:

$$\begin{aligned} E((X_{1+\tau_k}+k)\mathbb{I}_{\{Y_{\tau_k}=k\}}\beta_{1+\tau_k}) &= \sum_{j\in\mathbb{N}} E(X_{j+1}\beta_{j+1})P(\{Y_j=k\}\cap\{\tau_k=j\}) \\ &\quad + \sum_{j\in\mathbb{N}} kE\beta_{j+1}P(\{Y_j=k\}\cap\{\tau_k=j\}) \\ &= \sum_{j\in\mathbb{N}}(E(X_1\beta_1+kE\beta_1)P(\{Y_j=k\}\cap\{\tau_k=j\}) \\ &= (EY_1+kE\beta_1)P(Y_{\tau_k}=k). \end{aligned}$$

De esta última igualdad y la ecuación (1.20) se deduce la tesis.

□

Observación 1.4.2. *De* (1.19) *se deduce que para* $k\in\mathbb{N}$

$$\{Y_{\tau_{k+1}}\geq k+1\} =_{c.s} \{Y_{\tau_k}>k\}\cup(\{Y_{\tau_k}=k\}\cap\{\beta_{\tau_k+1}=1\}),$$

o lo que es lo mismo

$$\{Y_{\tau_{k+1}}\geq k+1\} =_{c.s} \{Y_{\tau_k}\geq k\}\setminus(\{Y_{\tau_k}=k\}\cap\{\beta_{\tau_k+1}=0\}),$$

de donde se deduce

$$P(Y_{\tau_{k+1}}\geq k+1) = P(Y_{\tau_k}\geq k) - P(Y_{\tau_k}=k)P(\beta_{\tau_k+1}=0). \tag{1.21}$$

Lo que muestra que la sucesión $(P(Y_{\tau_k}\geq k))_{k\in\mathbb{N}}$ *es monótona no creciente, cosa que es bastante intuitiva a priori.*

Observación 1.4.3. *Para poder calcular* EY_{τ_k} *y* $P(Y_{\tau_k}\geq k)$*, a partir de las ecuaciones* (1.18) *y* (1.21)*, se necesita conocer* $P(Y_{\tau_k}=k)$*,* $k\in\mathbb{N}$ *que no se puede calcular en general porque depende de las distribuciones de* X_1 *y* β_1*.*

Corolario 1.4.4. *En las hipótesis del teorema anterior*

$$\tau_{k^*} := \text{ínf}\{n\in\mathbb{N} : Y_n\notin(0,k^*)\},$$

donde

$$k^* = \left\lfloor \frac{EY_1}{1-E\beta_1}+1 \right\rfloor,$$

resuelve el problema de optimalidad planteado en (1.9)*. La notación* $\lfloor x\rfloor$ *representa el máximo natural menor o igual que* x*.*

Demostración. De (1.18) se ve que

$$EY_1 - k(1-E\beta_1)\geq 0 \Rightarrow EY_{\tau_{k+1}}\geq EY_{\tau_k},$$

entonces

$$k \leq \frac{EY_1}{1 - E\beta_1} \Rightarrow EY_{\tau_{k+1}} \geq EY_{\tau_k}.$$

Esto muestra que la sucesión (EY_{τ_i}) es monótona creciente al principio y después es monótona decreciente. El valor más grande que toma es $EY_{\tau_{h+1}}$ si $h = \text{máx}\{i \in \mathbb{N} : i \leq \frac{EY_1}{1-E\beta_1}\}$; o sea $h = \left\lfloor \frac{EY_1}{1-E\beta_1} \right\rfloor$.

□

Observación 1.4.5. *k^* coincide con k' definido en la introducción de esta sección salvo en el caso en que $c \in \mathbb{N}$ que $k^* = k' + 1$ y tanto τ_{k^*} como $\tau_{k'}$ son tiempos de parada óptimos.*

1.5. Un ejemplo: *El juego de la codicia*

En el caso del juego que sirve de motivación a este trabajo la variable X_n corresponde a la tirada n-ésima del dado, que es independiente de las tiradas anteriores, por lo tanto X_n debe tener distribución uniforme en $\{1,2,3,4,5,6\}$. La variable aleatoria β_n debe ser uno si $X_n \in \{2,3,4,5,6\}$ y cero en otro caso, o sea, $\beta_n = \mathbb{I}_{\{2,3,4,5,6\}}(X_n)$.

Por lo tanto las sucesiones $\{X_n\}_{n\in\mathbb{N}}$ y $\{\beta_n\}_{n\in\mathbb{N}}$ son de variables aleatorias independientes, idénticamente distribuidas, de modo que X_n toma valores en $\mathbb{N}$ y β_n toma valores en $\{0,1\}$. Además cumplen

$$E(X_n\beta_n) = \frac{2+3+4+5+6}{6} = \frac{10}{3} \text{ y } \ E\beta_n = \frac{5}{6}.$$

La ecuación (1.11) se verifica trivialmente, ya que la condición de idénticamente distribuidas de las variables X_n y β_n implica que la sucesión numérica que allí aparece sea constante. También es claro que se cumple la condición (1.12), ya que $\sum_{i=1}^{n} X_i \geq n$ casi seguramente.

Estamos en condiciones de aplicar el corolario 1.3.2, que asegura que τ_{20} es un tiempo de parada óptimo.

También estamos en las hipótesis del corolario 1.4.4 por lo que tenemos que τ_{21} también es un tiempo de parada óptimo.

Observación 1.5.1. *De* (1.18) *podemos observar que* $E(Y_{\tau_{21}}) = E(Y_{\tau_{20}})$ *ya que* $EY_1 - 20(1 - E\beta_1) = 0$

Para ver que esas son las únicas *k-reglas* óptimas ($k \in \mathbb{N}$) probaremos que $E(Y_{\tau_{20}}) - E(Y_{\tau_{19}}) > 0$ y que $E(Y_{\tau_{22}}) - E(Y_{\tau_{21}}) < 0$, para eso alcanza con probar que $P(Y_{\tau_{19}} = 19) > 0$ y $P(Y_{\tau_{21}} = 20) > 0$ y de la ecuación (1.18) se deduce lo que queremos. Probaremos, más en general, que $P(Y_{\tau_k} = k) > 0$ para $k \geq 2$

- $P(Y_{\tau_1} = 1) = 0$ ya que el resultado 1 del dado da ganancia 0.
- $P(Y_{\tau_2} = 2) = \frac{1}{6}$ que es la probabilidad de obtener un 2 en la primera tirada.
- $P(Y_{\tau_3} = 3) = \frac{1}{6}$ que es la probabilidad de obtener un 3 en la primera tirada.
- $P(Y_{\tau_4} = 4) = \frac{1}{6} + \frac{1}{36} = \frac{7}{36}$ que es la probabilidad de obtener un 4 en la primera tirada más la de obtener dos veces un 2.

- $P(Y_{\tau_5} = 5) = \frac{1}{6} + \frac{1}{36} + \frac{1}{36} = \frac{2}{9}$ que es la probabilidad de las combinaciones (5), (2,3) y (3,2).
- $P(Y_{\tau_6} = 6) = \frac{1}{6} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{216} = \frac{55}{216}$ que es la probabilidad de las combinaciones (6), (3,3), (2,4), (4,2) y (2,2,2).

Pensemos ahora en $\{Y_{\tau_k} = k\}$ con $k > 6$, podemos partir el conjunto, teniendo en cuenta el resultado del último dado tirado obteniendo:

$$\{Y_{\tau_k} = k\} = \cup_{i=2}^{6}(\{Y_{\tau_{k-i}} = k - i\} \cap \{X_{1+\tau_{k-i}} = i\})$$

(notar que la unión es disjunta). De esto se deduce:

$$\begin{aligned} P(Y_{\tau_k} = k) &= \textstyle\sum_{i=2}^{6} P(Y_{\tau_{k-i}} = k - i)P(X_{1+\tau_{k-i}} = i) \\ &= \frac{\sum_{i=2}^{6} P(Y_{\tau_{k-i}} = k-i)}{6}. \end{aligned} \tag{1.22}$$

La ecuación anterior deja claro que $P(Y_{\tau_k} = k) > 0$ para todo $k \geq 2$; además nos da una forma de calcular dicha probabilidad.

La ecuación (1.18), que para este caso sería

$$EY_{\tau_{k+1}} - EY_{\tau_k} = \left(\frac{10}{3} - \frac{k}{6}\right) P(Y_{\tau_k} = k), \tag{1.23}$$

nos permite, calcular EY_{τ_k} para $k > 1$ sabiendo que $EY_{\tau_1} = EY_1 = \frac{10}{3}$. En la figura 1.1 se muestra una tabla con los valores obtenidos de EY_{τ_k} para k desde 1 hasta 40. Además se agrega una estimación de la varianza hecha mediante simulaciones. Nótese que si bien τ_{20} y τ_{21} son estrategias iguales en cuanto a esperanza del puntaje no sucede lo mismo con la varianza. Se cumple

$$VarY_{\tau_{20}} < VarY_{\tau_{21}}.$$

Cómo se obtuvo la tabla:

Los valores de $P(Y_{\tau_k} = k)$ y EY_{τ_k} se calcularon directamente a partir de las ecuaciones (1.22) y (1.23) respectivamente. Confirmando totalmente los resultados obtenidos por M. Roters [8].

Para estimar $VarY_{\tau_k}$, para cada k entre 1 y 40, se simuló 100.000 veces el puntaje obtenido en un turno en que se juega con la estrategia τ_k. LLamémosle

k	$P(Y\tau_k=k)$	$EY\tau_k$	$VarY\tau_k$	k	$P(Y\tau_k=k)$	$EY\tau_k$	$VarY\tau_k$
1	0,0000	3,3333	3,9	21	0,0950	8,1418	120
2	0,1667	3,3333	3,9	22	0,0908	8,1260	128
3	0,1667	3,8333	4,9	23	0,0866	8,0957	135
4	0,1944	4,3056	6,6	24	0,0828	8,0524	142
5	0,2222	4,8241	9,5	25	0,0792	7,9972	150
6	0,2546	5,3796	13,7	26	0,0758	7,9312	157
7	0,1250	5,9738	19	27	0,0724	7,8554	165
8	0,1674	6,2446	23	28	0,0692	7,7709	172
9	0,1605	6,5795	28	29	0,0661	7,6786	177
10	0,1606	6,8737	34	30	0,0632	7,5794	183
11	0,1550	7,1414	41	31	0,0605	7,4740	189
12	0,1447	7,3739	48	32	0,0578	7,3631	195
13	0,1281	7,5668	55	33	0,0552	7,2475	200
14	0,1314	7,7162	63	34	0,0528	7,1278	203
15	0,1248	7,8476	70	35	0,0505	7,0046	208
16	0,1200	7,9516	78	36	0,0483	6,8784	212
17	0,1140	8,0316	86	37	0,0461	6,7497	216
18	0,1082	8,0886	95	38	0,0441	6,6190	218
19	0,1030	8,1246	103	39	0,0422	6,4867	222
20	0,0997	8,1418	111	40	0,0403	6,3532	225

Figura 1.1: $P(Y_{\tau_k} = k)$, EY_{τ_k}, $VarY_{\tau_k}$, para $k = 1 \dots 40$

$\overline{Y}^i_{\tau_k}$ al resultado obtenido en la i-ésima simulación para la estimación de $VarY_{\tau_k}$. El estimador de la varianza utilizado fue

$$\frac{\sum_{i=1}^{100000} \left(\overline{Y}^i_{\tau_k} - EY_{\tau_k}\right)^2}{100000}$$

Capítulo 2

Procesos de Markov controlados: Minimizar la cantidad de turnos

2.1. Descripción y motivación del problema

En el capítulo anterior resolvimos el problema de maximizar el puntaje de un turno; pero para ganar en el juego lo que se debe hacer es llegar al puntaje objetivo antes que el resto de los jugadores, que a priori requiere alcanzar dicho puntaje en la menor cantidad de turnos posible. Si bien es cierto que maximizar el puntaje de un turno es una buena estrategia cuando se está lejos del objetivo, no ocurre lo mismo cuando se está cerca. Supongamos, por ejemplo que, al empezar un turno, nos faltan 2 puntos para alcanzar el objetivo; es evidente que una vez que tiremos el dado y obtengamos 2 o más puntos debemos plantarnos y no tratar de superar los 20 ó 21 puntos, como sugieren los resultados del capítulo uno.

En este capítulo nos proponemos resolver el problema de minimizar la esperanza de la cantidad de turnos necesarios para alcanzar el puntaje objetivo. Para esto, vamos a necesitar estudiar un poco de "Procesos de Markov controlados", también llamado "Teoría de decisión de Markov".

2.2. Procesos de Markov controlados

Si bien el modelo planteado a continuación y las demostraciones que aparecen las escribí yo, enfocado a probar algunos resultados que son punto de partida del artículo de J. Haigh y M. Roters [5], está basado en los libros de E. Dynkin [3] y O. Hernández-Lerma [6].

2.2.1. La Idea

Los procesos de Markov controlados generalizan las cadenas de Markov. Se tiene una sucesión de variables aleatorias x_n que indican el estado de una partícula en el paso n. La variable x_n toma valores en X que es el conjunto de estados posibles.

La dinámica es la siguiente: Si en el instante i la partícula está en el estado x se debe tomar una acción a, entre el conjunto de acciones posibles para ese estado ($A(x)$). La acción tomada determina la distribución de probabilidad en X para el siguiente estado, además de tener aparejado un costo $c_i(x, a)$.

La diferencia sustancial con una cadena de Markov, es que cuando la partícula se encuentra en un estado dado, no está definida la distribución del siguiente estado hasta que no se toma una acción, de ahí la idea de control. Se obtiene una cadena de Markov en el caso particular en que sólo hay una acción posible por estado, de modo que no hay decisión alguna.

Un problema que se plantea es encontrar un "procedimiento de control", es decir, un criterio para decidir las acciones a tomar, de modo que el valor esperado del costo total (suma de los costos de las acciones tomadas), a lo largo de un intervalo de tiempo, sea mínimo.

2.2.2. Definición del proceso y planteo del problema

En un principio consideraremos el proceso indexado en los naturales del intervalo $[m, n]$. Para definir de forma precisa el proceso necesitamos los siguientes elementos:

1. El conjunto X, de los estados posibles. En nuestro caso consideramos X finito.
2. El conjunto A, de todas las acciones posibles. A también finito.
3. Para cada $x \in X$ un conjunto $A(x) \subseteq A$ que indica las acciones posibles cuando se está en ese estado.

4. Para cada $i \in [m, n-1]$ una función de costos c_i, que a cada par (x, a) con $x \in X$ y $a \in A(x)$ asocia un real no negativo que representa el costo de tomar la acción a cuando se está en el estado x en el instante i. O sea que si el sistema está en el estado $x_i = x$ y se decide tomar la acción a se debe "pagar" $c_i(x, a)$.
5. Para cada par (x, a), tal que $x \in X,\ a \in A(x)$, una probabilidad $p(.|x, a)$ en X que es la distribución del estado siguiente.
6. Una distribución de probabilidad μ en X para el primer estado, x_m, a la que llamaremos distribución inicial.

Definición 2.2.1. A los items 1-5 se les llama modelo del proceso y tiene interés en sí mismo cuando se quiere considerar el proceso en un subintervalo. En general lo anotaremos como Z. Cuando además se tiene la distribución inicial anotaremos Z_μ.

Observación 2.2.2. *Esta definición tiene muchas generalizaciones posibles. Algunas de ellas son:*

- *No exigir que los conjuntos de estados y acciones sean finitos.*
- *Permitir que el conjunto de estados dependa del instante en el que se está.*

El objetivo es encontrar un método de control de manera que el valor esperado del costo total sea mínimo. Para esto hay que definir "método de control" y la función cuya esperanza queremos minimizar.

Definición 2.2.3 (Espacio de caminos en Z)**.** A lo largo del proceso se obtiene una sucesión finita

$$\ell = x_m, a_{m+1}, x_{m+1}, \ldots, x_{n-1}, a_n, x_n,$$

donde $x_i \in X$ es el estado en el instante i y $a_{i+1} \in A(x_i)$ es la acción que se tomó a continuación. Al conjunto de estas posibles sucesiones lo denominamos *Espacio de caminos en Z* y denotaremos $L(Z)$ (o también L en caso en que el modelo esté claro del contexto). A un elemento ℓ de $L(Z)$ le llamamos *camino en Z*.

Definición 2.2.4 (Camino parcial en Z)**.** Un camino parcial es una sucesión finita, como en la definición de *camino en Z*, pero mirada hasta un instante previo i. Es decir, una sucesión

$$h = x_m, a_{m+1}, x_{m+1}, \ldots, x_{i-1}, a_i, x_i,$$

donde i está entre m y $n-1$, y se cumplen las restricciones de pertenencia indicadas antes.

Cada camino ℓ en Z tiene asociado un costo $C(\ell)$ que es la suma de los costos de las acciones tomadas. Es decir:

$$C(\ell) = \sum_{i=m}^{n-1} c_i(x_i, a_{i+1}) \tag{2.1}$$

si

$$\ell = x_m, a_{m+1}, x_{m+1}, \ldots, x_{n-1}, a_n, x_n.$$

Queremos minimizar el valor esperado de la variable aleatoria $C(\ell)$.

Definición 2.2.5 (Estrategia o Método de control)**.** Una estrategia (o método de control) π, es una función que a cada camino parcial

$$h = x_m, a_{m+1}, x_{m+1}, \ldots, x_{i-1}, a_i, x_i$$

en Z, le asigna una distribución de probabilidad $\pi_i(.|h)$ en A concentrada en las acciones posibles para el último estado, $A(x_i)$. *La idea es que dada la secuencia del proceso hasta el instante i, la estrategia asigna la distribución de probabilidad con la que se "sortea" la acción a_{i+1}.*

Denominamos Π a la familia de todas las estrategias posibles para el modelo.

Las siguientes, son subfamilias de estrategias muy importantes:

- Markoviana - Cuando $\pi_i(.|h)$ depende sólo del último estado en vez de depender de todo el camino parcial h. En general en este caso anotamos $\sigma_i(.|x_i)$ en lugar de $\pi_i(.|h)$.
- Estacionaria en el tiempo (o también estacionaria) - Es una estrategia markoviana en que $\pi_i(.|x_i)$ depende sólo del estado x_i y no del instante. Anotamos $\sigma(.|x)$.
- Determinista - Cuando para cualquier camino parcial h la probabilidad $\pi_i(.|h)$ está concentrada en un punto, se dice que la estrategia es *determinista*. Al punto en el cual se concentra la distribución se le anota $\phi_i(h)$.

  Observación 2.2.6. *Si una estrategia es determinista y estacionaria se puede pensar como una función $\sigma : X \to A(X)$ que decide qué acción tomar cuando se está en el estado x.*

Cuando se tiene un modelo con distribución inicial (Z_μ), y se fija una estrategia π, queda definida naturalmente una probabilidad en es espacio $L(Z)$ de los caminos. Dicha probabilidad es la siguiente:

$$\begin{aligned} P(x_m, a_{m+1}, x_{m+1}, \ldots, x_{n-1}, a_n, x_n) = \quad & \mu(x_m)\pi(a_{m+1}|x_m)p(x_{m+1}|x_m, a_{m+1}) \\ & \ldots \pi(a_n|x_m, a_{m+1}, x_{m+1}, \ldots, x_{n-1}) \\ & p(x_n|x_{n-1}, a_n). \end{aligned} \tag{2.2}$$

Definición 2.2.7 (Costo de una estrategia)**.** Si se da una estrategia π en un modelo con distribución inicial (Z_μ), definimos el costo de dicha estrategia como:

$$\begin{aligned} w(\mu, \pi) \quad & := E(C(x_m, a_{m+1}, x_{m+1}, \ldots, x_{n-1}, a_n, x_n)) \\ & = \textstyle\sum_{\ell \in L} C(\ell)P(\ell). \end{aligned} \tag{2.3}$$

Si la distribución inicial asigna todo el peso al estado x se suele denotar $w(x, \pi)$.

Teorema 2.2.8. *Se verifica la fórmula*

$$w(\mu, \pi) = \sum_{x \in X} \mu(x) w(x, \pi).$$

Demostración. Surge inmediatamente de la definición.

□

Definición 2.2.9 (Costo de la distribución inicial μ)**.** Fijado un modelo, definimos el costo de la distribución (μ) de la siguiente manera:

$$w^*(\mu) := \inf_{\pi \in \Pi} w(\mu, \pi).$$

Si la distribución μ está concentrada en el estado x al costo de la distribución se le suele llamar *costo del estado x para el modelo Z* y se anota $w^*(x)$.

Lema 2.2.10. *Vale la desigualdad*

$$w^*(\mu) \leq \sum_{x \in X} \mu(x) w^*(x).$$

Demostración. Supongamos, por absurdo, que $w^*(\mu) > \sum_{x \in X} \mu(x) w^*(x)$, entonces existiría $\epsilon > 0$ tal que $w^*(\mu) = \sum_{x \in X} \mu(x) w^*(x) + \epsilon$. Para cada $x \in X$

consideremos una estrategia π_x tal que $w(x, \pi_x) < w^*(x) + \epsilon$. Sea π una estrategia que coincide con π_x cuando el primer estado es x. Con esta estrategia combinada se tiene, gracias al teorema 2.2.8:

$$w(\mu, \pi) = \sum_{x \in X} \mu(x) w(x, \pi_x) < \sum_{x \in X} \mu(x)(w^*(x) + \epsilon) = w^*(\mu),$$

lo que es absurdo por definición de $w^*(\mu)$.

□

Lema 2.2.11. *Vale la siguiente desigualdad:*

$$w^*(\mu) \geq \sum_{x \in X} \mu(x) w^*(x).$$

Demostración. Supongamos, por absurdo, que $w^*(\mu) < \sum_{x \in X} \mu(x) w^*(x)$, entonces existe una estrategia π tal que $w(\mu, \pi) < \sum_{x \in X} \mu(x) w^*(x)$. Pero por el teorema 2.2.8 sabemos que $w(\mu, \pi) = \sum_{x \in X} \mu(x) w(x, \pi)$. Entonces existe algún x tal que $w(x, \pi) < w^*(x)$, lo que es absurdo por la definición de $w^*(x)$.

□

Como consecuencia de los dos lemas anteriores se obtiene el siguiente teorema:

Teorema 2.2.12. *El costo de una distribución inicial μ en un modelo Z cumple la siguiente identidad:*

$$w^*(\mu) = \sum_{x \in X} \mu(x) w^*(x).$$

Definición 2.2.13 (Estrategia óptima para Z_μ)**.** Es una estrategia π^*_μ que realiza el ínfimo:

$$w(\mu, \pi^*_\mu) = \inf_{\pi \in \Pi} w(\mu, \pi) = w^*(\mu).$$

Definición 2.2.14 (Estrategia uniformemente óptima)**.** Es una estrategia π^* que es óptima para toda distribución inicial μ, es decir

$$w(\mu, \pi^*) = \inf_{\pi \in \Pi} w(\mu, \pi) = w^*(\mu)$$

para toda distribución de probabilidad μ en X.

Teorema 2.2.15. *Si para cada $x \in X$, tenemos una estrategia π^*_x óptima para la distribución inicial que concentra todo su peso en x, o sea*

$$w(x, \pi^*_x) = w^*(x),$$

y definimos la estrategia π^* de modo que coincida con π_x^* si el primer estado fue x, o sea

$$\pi^*(.|h) := \pi_x^*(.|h) \text{ si el primer estado de } h \text{ es } x,$$

ésta resulta ser uniformemente óptima.

Demostración. Por el teorema 2.2.12, sabemos que

$$\begin{aligned} w(\mu, \pi^*) &= \textstyle\sum_{x \in X} \mu(x) w(x, \pi^*) \\ &= \textstyle\sum_{x \in X} \mu(x) w(x, \pi_x^*) \qquad \text{(por la definición de } \pi^*) \\ &\leq \textstyle\sum_{x \in X} \mu(x) w(x, \pi) = w(\mu, \pi) \end{aligned}$$

para toda estrategia $\pi \in \Pi$. La desigualdad se debe a la optimalidad de π_x^*. Esto concluye la demostración.

□

Observación 2.2.16. *El teorema anterior muestra que a la hora de buscar estrategias óptimas globales, alcanza con encontrar estrategias óptimas para cada estado y combinarlas.*

2.2.3. Solución al problema

Teorema 2.2.17. *Dado un modelo Z, como el definido en 2.2.1, existe una estrategia π^* que es óptima para ese modelo.*

Demostración. Para definir una estrategia hay que dar una familia finita de distribuciones de probabilidad en el conjunto A, que es finito. Cada distribución de probabilidad se puede ver como una cantidad finita de variables en $[0, 1]$ cuya suma es 1. Por lo tanto una estrategia es una cantidad finita de variables en un conjunto compacto. La función que queremos minimizar es continua como función de dichas variables, ya que son sólo sumas y productos, según (2.3). Entonces podemos asegurar que se alcanza el ínfimo, lo que prueba la existencia de la estrategia óptima.

□

Definición 2.2.18 (Modelo derivado)**.** Dado un modelo Z llamamos Z' al modelo que resulta de la observación del modelo Z luego del primer paso. Es decir en el intervalo $[m + 1, n]$ en lugar de $[m, n]$. A este nuevo modelo se le llama modelo derivado. Para Z' utilizaremos la notación w' y $w^{*'}$, para referirnos a los costos, en lugar de w y w^*.

Teorema 2.2.19 (Ecuación fundamental)**.** *Dado un modelo Z y una estrategia π se cumple la siguiente igualdad:*

$$w(x,\pi) = \sum_{a\in A} \pi(a|x)(c_m(a,x) + w'(p_{x,a}, \pi_{x,a})), \tag{2.4}$$

donde

- *$p_{x,a}$ es la distribución $p(.|x,a)$ del modelo Z (que toma el rol de distribución inicial cuando se mira el modelo a partir del segundo paso y la acción tomada previamente fue a)*
- *$\pi_{x,a}$ es la estrategia π mirada a partir del segundo paso, cuando el primer estado fue x y la acción tomada fue a, es decir,*

$$\pi_{x,a}(.|x_{m+1}, a_{m+2}, \ldots, x_i) = \pi(.|x, a, x_{m+1}, a_{m+2}, \ldots, x_i).$$

Demostración. Empecemos por ver la relación entre el costo y la probabilidad de un camino en $L(Z)$ y un camino en $L(Z')$: un camino en $L(Z)$ se puede descomponer en el primer paso concatenado con un camino en $L(Z')$. Si $\ell = x, a_{m+1}, x_{m+1}, a_{m+2}, \ldots, x_n$ se tiene

$$C(\ell) = c_m(x, a_{m+1}) + C(\ell') \tag{2.5}$$

y

$$P(\ell) = \pi(a_{m+1}|x)P(\ell') \tag{2.6}$$

siendo $\ell' = x_{m+1}, a_{m+2}, \ldots, x_n$. Al escribir $P(\ell)$ se omitió la distribución inicial ya que concentra todo el peso en el estado x. De la definición de costo de una estrategia, teniendo en cuenta que un camino en $L(Z)$ que no empieza en el estado x tiene probabilidad cero y las ecuaciones (2.5) y (2.6), surge:

$$\begin{aligned} w(x,\pi) &= \textstyle\sum_{a_{m+1}\in A,\ \ell'\in L'} [c_m(x,a_{m+1}) + C(\ell')]\pi(a_{m+1}|x)P(\ell') \\ &= \textstyle\sum_{a_{m+1}\in A,\ \ell'\in L'} c_m(x,a_{m+1})\pi(a_{m+1}|x)P(\ell') \\ &\quad + \textstyle\sum_{a_{m+1}\in A,\ \ell'\in L'} C(\ell')\pi(a_{m+1}|x)P(\ell') \\ &= \textstyle\sum_{a_{m+1}\in A} \pi(a_{m+1}|x)[c_m(x,a_{m+1}) + \underbrace{\sum_{\ell'\in L'} C(\ell')P(\ell')}_{w'(p_a,\pi_a)}]. \end{aligned} \tag{2.7}$$

□

2.2.4. Horizonte infinito

Ahora consideraremos el modelo visto antes, pero en un intervalo infinito, $[m, \infty)$. El modelo es prácticamente el mismo. Lo que hay que tener en cuenta es que ahora hay una cantidad infinita de funciones de costo (c_i). Y además, el costo de una estrategia es una suma infinita. Dado que estamos considerando costos no negativos, no hay problema con la definición de esa suma infinita, aunque podría dar infinito. En este trabajo nos vamos a limitar a modelos para los que existe alguna estrategia de costo finito, o lo que es lo mismo, modelos cuyo valor es finito. Si consideramos una distribución inicial μ, el costo de una estrategia π es, en este caso,

$$\begin{aligned} w(\mu, \pi) &= E\left(\textstyle\sum_{i=m}^{\infty} c_i(x_i, a_{i+1})\right) \\ &= \textstyle\sum_{i=m}^{\infty} E(c_i(x_i, a_{i+1})) \\ &= \lim_{n\to\infty} \textstyle\sum_{i=m}^{n-1} E(c_i(x_i, a_{i+1})). \end{aligned} \tag{2.8}$$

La última igualdad da la idea de que se pueden generalizar resultados del caso finito, ya que la función a la que se le está tomando límite es es el costo de la estrategia en el caso $[m, n]$.

Teorema 2.2.20 (Ecuación fundamental (horizonte infinito))**.** *Dado un modelo Z, en el caso infinito, y una estrategia π, vale la igualdad:*

$$w(x, \pi) = \sum_{a\in A} \pi(a|x)(c_m(a, x) + w'(p_{x,a}, \pi_{x,a})), \tag{2.9}$$

al igual que en el caso finito.

Demostración. Consideremos para cada $n > m$ el modelo Z^n definido en $[m, \infty)$ a partir de Z, pero con función de costos c^n definida de la siguiente forma:

$$c_i^n(x, a) := \begin{cases} c_i(x, a) & \text{si } i < n \\ 0 & \text{si } i \geq n \end{cases}$$

El modelo Z^n es, formalmente, un modelo de horizonte infinito, pero como consecuencia de la definición de los costos, que valen cero a partir de n es totalmente equivalente a un modelo de horizonte finito $[m, n]$. Entonces para Z^n vale la ecuación fundamental. Además, si llamamos $w^n(\mu, \pi)$ al costo de la estrategia π para el modelo Z^n, se cumple que $w^n(\mu, \pi)$ crece a $w(\mu, \pi)$. Hasta

ahora tenemos:

$$\begin{aligned} w(\mu,\pi) &= \lim_{n\to\infty} \sum_{a\in A} \pi(a|x)(c_m(a,x) + w^{n'}(p_{x,a},\pi_{x,a})) \\ &= \sum_{a\in A} \pi(a|x)(c_m(a,x) + \lim_{n\to\infty} w^{n'}(p_{x,a},\pi_{x,a})). \end{aligned}$$

Veamos que el límite ($n \to \infty$) de $w^{n'}(p_{x,a},\pi_{x,a})$ es $w'(p_{x,a},\pi_{x,a})$: por definición de costo tenemos,

$$w^{n'}(p_{x,a},\pi_{x,a}) = E\left(\sum_{i=1}^{\infty} c_i^n(x_i,a_{i+1})\right).$$

Como c^n crece a c, se puede pasar al límite gracias al teorema de convergencia monótona y se obtiene lo que queremos.

□

Teorema 2.2.21. *Para el modelo Z, en horizonte infinito, se verifica*

$$w^*(x) = \min_{a\in A}(c_m(x,a) + w^{*'}(p_{x,a})).$$

Demostración. A partir de la ecuación fundamental, obtenida en el teorema anterior, se tiene inmediatamente que

$$w(x,\pi) \geq \min_{a\in A}(c_m(x,a) + w'(p_{x,a},\pi_{x,a})) \geq \min_{a\in A}(c_m(x,a) + w^{*'}(p_{x,a})),$$

de donde surge que:

$$\min_{a\in A}(c_m(x,a) + w^{*'}(p_{x,a})) \leq w^*(x)$$

y queda probada una desigualdad. Veamos, por absurdo, que la desigualdad es en realidad una igualdad. Supongamos que la desigualdad es estricta, entonces existe $\epsilon > 0$ tal que

$$\min_{a\in A}(c_m(x,a) + w^{*'}(p_{x,a})) + \epsilon = w^*(x).$$

Dado que $w^{*'}(p_{x,a})) = \inf_{\pi\in\Pi'} w'(p_{x,a},\pi)$, existe una estrategia cuyo valor es tan cercano a $w^{*'}(p_{x,a})$ como queramos. Sea π' una estrategia tal que $w'(p_{x,a},\pi) \leq w^{*'}(p_{x,a}) + \frac{\epsilon}{2}$. Ahora consideremos la estrategia π en Z que en el primer paso toma una acción a que realiza el mínimo en el segundo miembro de la igualdad que queremos probar (tal acción existe ya que el mínimo es entre una cantidad finita de acciones, y la denominamos **una acción óptima**), y a partir del segundo paso coincide con π'. A partir de la ecuación fundamental, surge que:

$$w(x,\pi) = c_m(x,a) + w'(p_{x,a},\pi) \leq c_m(x,a) + w^{*'}(p_{x,a}) + \frac{\epsilon}{2} < w^*(x).$$

Habríamos conseguido una estrategia que mejora el óptimo, lo que resulta absurdo.

□

Corolario 2.2.22. *Para n arbitrario se verifica*

$$\inf_{\pi\in\Pi} w(x,\pi) = \inf_{\pi^1\in\Pi^1} w(x,\pi^1) = \ldots = \inf_{\pi^n\in\Pi^n} w(x,\pi^n),$$

donde Π^i *es el conjunto de las estrategias que en los primeros i pasos toman la acción óptima.*

Demostración. La primera igualdad sale del teorema anterior, y las siguientes de aplicar el mismo en forma sucesiva a los modelos derivados.

□

Corolario 2.2.23. *Una estrategia* π^∞ *que en todos los pasos toma la acción óptima, o sea, que pertenece a* Π^i *para todo i, y en caso de haber más de una decide con algún criterio determinista, es una estrategia óptima. Además es markoviana y determinista.*

Demostración. Si no fuera óptima $w(x,\pi^\infty) > w^*(x)$, entonces existiría un natural n tal que $\sum_{i=m}^{n-1} E(c_i(x_i, a^*_{i+1})) > w^*(x)$. Consideremos $\pi^n \in \Pi^n$, se cumple

$$\begin{aligned} w(x,\pi^n) &= \textstyle\sum_{i=m}^{n-1} E(c_i(x_i, a^*_{i+1})) + \sum_{i=n}^{\infty} E(c_i(x_i, a_{i+1})) \\ &\geq \textstyle\sum_{i=m}^{n-1} E(c_i(x_i, a^*_{i+1})) > w^*(x). \end{aligned}$$

Esto contradice el corolario anterior.
Que es markoviana y determinista es evidente a partir de la definición.

□

2.2.5. Horizonte Infinito, caso homogéneo

La idea en está sección, es definir una subclase de los modelos definidos en las secciones anteriores, y estudiar las particularidades de la estrategia óptima obtenida en la sección anterior.

Definición 2.2.24 (Modelo Homogéneo)**.** Decimos que un modelo es homogéneo cuando la función de costos no depende del instante. Es decir, cuando $c_i(x,a) = c_j(x,a)$ para todo i, j en $[m, +\infty)$.

Recordemos que la estrategia óptima que conocemos (π^*) es la que cuando se encuentra en el estado x y en el instante i toma la acción a_{i+1} que cumple

$$c_i(x, a_{i+1}) + w^{*'}(p_{x,a_{i+1}}) = \min_{a \in A}(c_m(x, a) + w^{*'}(p_{x,a})),$$

donde $w^{*'}(p_{x,a_i})$ representa el costo óptimo de la distribución inicial p_{x,a_i} del modelo que empieza en el instante $i + 1$.

Al considerar un modelo homogéneo de horizonte infinito, el costo de una distribución inicial no depende del instante de comienzo. Para convencerse, basta reescribir la ecuación (2.1) para el caso infinito homogéneo. Una consecuencia de esto es que $w' = w$ y la ecuación de optimalidad toma la forma:

$$w^*(x) = \min_{a \in A}(c(x, a) + w^*(p_{x,a})),$$

donde, en virtud del teorema 2.2.12 (si bien fue demostrado en el caso finito se puede observar que en la demostración no se utilizó para nada la finitud, por lo tanto vale también para este caso), $w^*(p_{x,a}) = \sum_{y \in X} p_{x,a}(y) w^*(y)$, entonces:

$$w^*(x) = \min_{a \in A}(c(x, a) + \sum_{x \in X} p_{x,a}(x) w^*(x)). \tag{2.10}$$

Observación 2.2.25. *La estrategia hallada es markoviana, determinista, y estacionaria (ya que la acción que toma depende sólo del estado actual y no del instante).*

Corolario 2.2.26. *Un corolario muy importante, de la observación anterior, es que a la hora de buscar estrategias óptimas en modelos como estos nos podemos restringir a la subclase de estrategias markovianas, deterministas, estacionarias. Es decir:*

$$\inf_{\pi \in \Pi} w(\mu, \pi) = \inf_{\pi \in \Pi_{(m,d,e)}} w(\mu, \pi),$$

donde $\Pi_{(m,d,e)}$ es la clase de estrategias markovianas, deterministas, estacionarias.

Como ya fue dicho, este tipo de estrategias se puede ver como una función que a cada estado asigna una acción, por lo que se suelen representar como una función $\sigma : X \to A$.

2.3. Minimizar el número de turnos en la codicia como un problema de control

En esta sección nos planteamos el problema de encontrar una estrategia de juego que minimice la esperanza de la cantidad de turnos que lleva alcanzar un puntaje objetivo, que es un natural T.

Vamos a plantearlo como un problema de control, aplicando los resultados de la sección anterior. Para esto debemos decidir el conjunto de estados, el conjunto de acciones, las probabilidades de transición y las funciones de costo, para que el problema se ajuste a lo que queremos.

El resto de este capítulo se basa en el artículo de J. Haigh y M. Roters [5].

2.3.1. El modelo

El conjunto X de los estados posibles va a depender del objetivo T por lo que lo llamaremos $X(T)$ y se define así:

$$X(T) := \{(t, r) : 1 \leq t \leq T,\ r = 0, 2, 3, \ldots, t + 5\}, \tag{2.11}$$

donde

- t: Es el objetivo teniendo en cuenta el puntaje que ya se obtuvo, es decir, si en los turnos anteriores acumulamos n puntos ahora $t = T - n$.
- r: Es el puntaje acumulado en el turno corriente. Es claro que no puede ser uno ya que un as en el dado indica que se termina el turno sin obtener puntaje.

Consideremos un ejemplo para entender la definición. Supongamos que el juego es a superar 200 puntos, entonces $T = 200$.

Al principio se está en el estado $(200, 0)$, se tira el dado una vez y se obtiene el valor 4, entonces pasamos al estado $(200, 4)$; se decide seguir el turno y se obtiene un 6, el nuevo estado es $(200, 10)$; ahora supongamos que decidimos parar, el nuevo estado es $(190, 0)$, ya que se nos contabiliza el puntaje alcanzado en el turno, nuestro nuevo objetivo es superar 190 puntos. Empieza el segundo turno, se tira el dado y se obtiene un 3, se pasa al estado $(190, 3)$; se vuelve a tirar el dado y se obtiene un 1, entonces se pierde el puntaje acumulado y se vuelve al estado $(190, 0)$.

Cuando se está en un estado (t, r) con $r \geq t$ consideramos el juego terminado. Esa es la razón por la cual consideramos estados a los pares (t, r) que cumplen $r \leq t + 5$. Supongamos que se está en el estado $(20, 18)$ eso quiere decir que tenemos que llegar a 20 y ya tenemos 18. Decidimos seguir y volvemos a tirar y obtenemos un 6, pasamos al estado $(20, 24)$ en que el juego está terminado, no tiene sentido seguir tirando el dado.

El conjunto de acciones posibles es muy sencillo, ya que la decisión que uno toma cuando juega a la codicia es seguir o parar. Por lo tanto definimos:

$$A := \{0\text{-}parar, 1\text{-}seguir\}.$$

Los conjuntos $A((t, r))$ con $(t, r) \in X$ serán

$$A((t,r)) := \begin{cases} \{1\} & \text{si } r = 0 \\ \{0,1\} & \text{si } 0 < r < t \\ \{0\} & \text{si } r > t \end{cases}$$

La idea de esta definición de las acciones posibles, es que cuando un turno recién empieza, $(r = 0)$, no es razonable plantarse, implica perder un turno sin ganar nada. Durante un turno las dos acciones son razonables. Cuando el juego terminó la acción "seguir" no tiene sentido.

Antes de definir la función de costos veamos como son las probabilidades de transición. Debemos dar una distribución de probabilidad $p(.|x, a)$ para el estado siguiente, que respete la dinámica del juego, para cada par (x, a) con $x \in X$ y $a \in A(x)$. Si se parte de un estado $(t, 0)$ la única acción posible es *seguir* y la probabilidad de transición es:

$$P((t',r')|(t,0),1) := \begin{cases} \frac{1}{6} & \text{si } t' = t \text{ y } r = 0,2,3,4,5,6 \\ 0 & \text{en otro caso} \end{cases}$$

En caso de partir de un estado (t, r) con $0 < r < t$ hay que dar las probabilidades de transición para las dos acciones posibles.

$$P((t',r')|(t,r),1) := \begin{cases} \frac{1}{6} & \text{si } t' = t \text{ y } r' = r + i \text{ con } i = 0,2,3,4,5,6 \\ 0 & \text{en otro caso} \end{cases}$$

$$P((t',r')|(t,r),0) := \begin{cases} 1 & \text{si } t' = t \text{ y } r' = 0 \\ 0 & \text{en otro caso} \end{cases}$$

Si se está en un estado (t, r) con $r \geq t$, la una acción posible es parar, y no cambia el estado. En este caso la probabilidad de transición es:

$$P((t', r')|(t, r), 0) := \begin{cases} 1 & \text{si } t' = t \text{ y } r' = r \\ \\ 0 & \text{en otro caso} \end{cases}$$

La función de costos, debe contar los turnos que insumió un juego, ya que esa es la variable que cuyo valor esperado queremos minimizar. En el modelo que definimos, el comienzo de un turno está marcado por el pasaje por un estado (t, r) con $r = 0$. Pensemos en esto con más cuidado: El juego arranca en el estado $(T, 0)$ y empezar a jugar ya asegura que por lo menos vamos a necesitar un turno. A medida que se continúa un turno, sin que salga un uno, se va pasando por estados (T, r) con r positivo y no se pierden turnos. Si sale un as se vuelve al estado $(T, 0)$ y se pierde un turno. Si se decide parar se va al estado $(T - r, 0)$ y se pierde un turno. Entonces una buena función de costo es la que a los estados $(t, 0)$ les asocia costo uno; a los estados (t, r) con $0 < r < t$ les asigna costo cero independientemente de la acción tomada; y les asocia costo nulo a los estados (t, r) con r mayor o igual que t. La función de costo resultante es homogénea, es decir, no varía en el tiempo:

$$c((t, r), a) := \begin{cases} 1 & \text{si } r = 0 \\ \\ 0 & \text{en otro caso} \end{cases} \tag{2.12}$$

Observación 2.3.1. *El modelo definido es homogéneo y el horizonte es infinito, ya que a priori no se sabe cuantas tiradas puede llevar alcanzar el objetivo. Además, cumple que fijada una estrategia π resulta que $w((T, 0), \pi)$ es la esperanza de la cantidad de turnos, que lleva alcanzar el objetivo, jugando con la estrategia π.*

2.3.2. Solución al problema

Como aplicación de los resultados de la sección 2.2.5 tenemos una solución óptima para el problema, es la que resulta de tomar la acción que realiza el mínimo en las ecuaciones de optimalidad (2.10), en este caso serían:

- Para los estados de la forma $(t, 0)$

$$w^*((t, 0)) = 1 + \frac{1}{6}\left(w^*((t, 0)) + \sum_{k=2}^{6} w^*((t, k))\right), \tag{2.13}$$

  notar que el mínimo no aparece ya que hay una sola acción posible. El 1 que aparece sumado corresponde al costo de la acción.

- si (t,r) es tal que $2 \leq r \leq t-1$

$$w^*((t,r)) = \text{mín}\left\{ w^*((t-r,0)),\ \frac{1}{6}\left(w^*((t,0)) + \sum_{i=2}^{6} w^*((t,r+i)) \right) \right\} \tag{2.14}$$

donde, la primera parte del mínimo anterior corresponde al mínimo que se alcanza tomando la acción de parar en (t,r). La otra parte corresponde al mínimo valor esperado si se decide seguir. Si resulta que las dos acciones posibles llevan al mismo valor de $w((t,r))$ acordamos seguir el turno.

El objetivo que tenemos ahora es calcular efectivamente la estrategia óptima. Para eso necesitaremos un poco de notación.

Definición 2.3.2 (Estrategias de juego)**.** Les llamamos estrategias de juego a las funciones $\sigma : X(T) \to A$ que respetan $\sigma(x) \in A(x)$.

Observación 2.3.3. *Las estrategias de juego no son otra cosa que las estrategias markovianas, deterministas y estacionarias. Sabemos, por 2.2.26, que si nos restringimos a estas estrategias para buscar optimizar la cantidad de turnos "no perdemos nada".*

A partir de ahora a la estrategia óptima la denotamos σ_0 (también $\sigma_{T,0}$ en caso de confusión). Entonces $\sigma_0 : X \to \{0,1\}$, asigna la acción que realiza el mínimo en las ecuaciones de optimalidad y en caso de igualdad decide seguir.

A la esperanza de la cantidad de turnos que requiere terminar el juego, si nos encontramos en (t,r) y jugamos con la estrategia σ, le llamamos $B_\sigma(t,r)$. Si la estrategia es la óptima ($\sigma = \sigma_0$) llamamos $B_0(t,r)$ en lugar de $B_{\sigma_0}(t,r)$.

El valor de la estrategia óptima para alcanzar el objetivo T se denota $M(T)$. Notar que:

$$M(T) = B_0(T,0).$$

Veamos como se relacionan B_0 con w^*. Es claro que $B_0(T,0) = w^*((T,0))$. En el caso de los estados (t,r) con $0 < r < t$ sucede que $w^*((t,r))$ cuenta un turno de menos porque no considera el turno actual. Por lo tanto tenemos que

$$B_0(t,r) = w^*((t,r)) + 1.$$

En el caso de los estados (t,r) con $r > t$ no nos queda más que definir $B_0(t,r) = 1$ ya que el juego está terminado y sólo se debe contar el turno actual.

Considerando que para los estados (t, r) con $r > t$ se tiene $w^*((t, r)) = 0$, ya que el sistema queda en ese estado absorbente y no hay costos, la relación entre B_0 y w^* se resume:

$$B_0(t,r) = \begin{cases} w^*(t,r) \text{ si } r = 0 \\ \\ w^*(t,r) + 1 \qquad \text{en otro caso} \end{cases} \tag{2.15}$$

A partir de la ecuación (2.15) podemos reescribir las ecuaciones de optimalidad (2.13) y (2.14) en función de B_0. De (2.13) obtenemos

$$B_0(T,0) = 1 + \frac{1}{6}\left(B_0(T,0) + \sum_{k=2}^{6}(B_0(T,k) - 1)\right),$$

o lo que es lo mismo,

$$M(T) = B_0(T,0) = \frac{1}{6}\left(1 + M(T) + \sum_{k=2}^{6} B_0(T,k)\right). \tag{2.16}$$

Análogamente de (2.14) se tiene que para $r = 2, 3, \ldots T-1$

$$B_0(T,r) - 1 = \text{mín}\left\{B_0(T-r,0),\ \frac{1}{6}\left(B_0(T,0) + \sum_{k=2}^{6}(B_0(T,r+k) - 1)\right)\right\},$$

de donde se deduce

$$B_0(T,r) = \text{mín}\left\{1 + M(T-r),\ \frac{1}{6}\left(1 + M(T) + \sum_{k=2}^{6} B_0(T,r+k)\right)\right\}. \tag{2.17}$$

Observación 2.3.4. *La función M cumple $M(1) = M(2) = \frac{6}{5}$*

Demostración. La ecuación (2.16) nos permite despejar $M(T)$ para obtener:

$$M(T) = \frac{1 + \sum_{k=2}^{6} B_0(T,k)}{5}.$$

En los casos $T = 1$, $T = 2$, $B_0(T,k)$ resulta ser uno. □

Sólo con las ecuaciones (2.16) y (2.17) no podemos calcular la estrategia σ_0 y $M(T)$ en general, ya que se da una dependencia cruzada que no permite llegar a un paso base. Un ejemplo que muestra la dependencia es el siguiente: si quisiéramos calcular $M(T)$ para T mayor que dos, intentaríamos hacerlo mediante la recursión planteada en la ecuación (2.16), pero necesitamos conocer

$B_0(T, j)$ con $j = 2, 3, 4, 5, 6$, para lo que trataríamos de utilizar la ecuación (2.17), que requiere conocer $M(T)$.

Para resolver este problema construimos una sucesión que converge a $M(T)$, para luego, con la ayuda de un computador, poder calcularlo. Los siguientes lemas apuntan a eso.

Lema 2.3.5. *Si $T \geq 2$, la cantidad mínima esperada de turnos para alcanzar el objetivo T decrece con T, es decir, $M(T) \geq M(T-1)$.*

Demostración. Podemos partir de la estrategia $\sigma_{T,0}$ para definir otra que alcance $T-1$ de la siguiente manera:

$$\sigma_1(t, r) := \sigma_{T,0}(t+1, r), \qquad 2 \leq r < t \leq T-1.$$

Esta nueva estrategia está definida en $X(T-1)$. Con ella, la cantidad de turnos que lleva alcanzar $T-1$ es la cantidad de turnos que requiere alcanzar T jugando con la estrategia óptima, o sea, $B_{\sigma_1}(T-1, 0) = M(T)$. Obviamente, por definición, $M(T-1) \leq B_{\sigma_1}(T-1, 0)$.

□

Lema 2.3.6. *Supongamos que $M_1(T) \leq M(T)$ y definamos:*

$$M_2(T) := \frac{1}{6}\left(1 + M_1(T) + \sum_{k=2}^{6} B^1(T, k)\right), \tag{2.18}$$

donde $B^1(T, r)$ vale uno para $r \geq T$ y para $r = 2, 3, \ldots, T-1$ está definido de la siguiente forma:

$$B^1(T, r) = \text{mín}\left\{1 + M(T-r),\ \frac{1}{6}\left(1 + M_1(T) + \sum_{k=2}^{6} B^1(T, r+k)\right)\right\}. \tag{2.19}$$

Entonces, se cumple la siguiente desigualdad:

$$M(T) \geq M_2(T) \geq M_1(T).$$

Demostración. Como $M(T) \geq M_1(T)$ podemos escribir $M_1(T) = M(T) - \epsilon$, donde, $\epsilon \geq 0$. De la desigualdad entre $M_1(T)$ y $M(T)$, en virtud de (2.17) y (2.19), se ve claramente que $B_1(T, r) \leq B_0(T, r)$ para $r = 2, 3, \ldots T-1$. Con esto, de (2.16) y (2.18) se deduce que $M_2(T) \leq M(T)$ con lo que probamos una de las desigualdades.

Para probar que $M_2(T)$ es menor o igual que $M_1(T)$ necesitamos primero analizar la diferencia entre B_0 y B^1.

Empecemos por comparar $B_0(T, T-1)$ con $B_1(T, T-1)$. De las ecuaciones (2.17) y (2.19) se ve que dicha diferencia es menor o igual a la diferencia que existe entre

$$\frac{1}{6}\left(1 + M(T) + \sum_{k=2}^{6} B_0(T, T-1+k)\right)$$

y

$$\frac{1}{6}\left(1 + M_1(T) + \sum_{k=2}^{6} B^1(T, T-1+k)\right),$$

que es $\frac{1}{6}(M(T) - M_1(T))$, ya que los sumandos de ambas sumatorias son todos iguales a uno por ser $T - 1 + k \geq T$. Además sabemos que dicha diferencia está acotada por $\frac{\epsilon}{6}$. Entonces obtenemos:

$$B^1(T, T-1) \geq B_0(T, T-1) - \frac{\epsilon}{6}.$$

Para $B_0(T, T-2)$ y $B_1(T, T-2)$, con exactamente el mismo argumento llegamos a:

$$B^1(T, T-2) \geq B_0(T, T-2) - \frac{\epsilon}{6}.$$

A la hora de acotar la diferencia de $B_0(T, T-3)$ con $B^1(T, T-3)$ hay uno de los sumandos de la sumatoria que deja de ser uno y hay que considerarlo. Por lo que la cota para esta diferencia es la diferencia entre $\frac{1}{6}(M(T) + B_0(T, T-1))$ y $\frac{1}{6}(M_1(T) + B^1(T, T-1))$, que está acotada por $\frac{1}{6}(\epsilon + \frac{\epsilon}{6})$, entonces:

$$B^1(T, T-3) \geq B_0(T, T-3) - \frac{7\epsilon}{36}.$$

Siguiendo con este razonamiento una cantidad finita de veces llegamos a que existe una constante $K(T)$ positiva, que no depende de ϵ tal que:

$$B^1(T, r) \geq B_0(T, r) - \epsilon K(T) \text{ para todo } r \geq 2.$$

Además se puede observar mediante un argumento recursivo que la constante K se puede tomar menor que 1.

Con esto, de (2.18), obtenemos:

$$M_2(T) \geq \frac{1}{6}\left(1 + M(T) + \sum_{k=2}^{6} B_0(T, k)\right) - L\epsilon,$$

donde $L = \frac{1+5k}{6}$. O lo que es lo mismo por (2.17):

$$M_2(T) \geq M(T) - L\epsilon.$$

Como L es menor que uno y $M_1(T) = M(T) - \epsilon$ tenemos la desigualdad que faltaba.

□

Observación 2.3.7. *En las hipótesis del lema anterior pedimos $M_1(T) < M(T)$. Si definimos $M_1(T) := M(T-1)$, en virtud del lema 2.3.5, podemos aplicarlo.*

Corolario 2.3.8. *Aplicando el resultado del lema anterior sucesivas veces se obtiene una sucesión $M_1(T), M_2(T), M_3(T), \ldots$ que crece a $M(T)$*

Demostración. En la demostración del lema anterior llegamos a

$$M(T) \geq M_2(T) \geq M(T) - L\epsilon,$$

donde ϵ es la diferencia entre $M(T)$ y $M_1(T)$. Ahora, como $M(T) \geq M_2(T)$ podemos, gracias a la observación 2.3.7, aplicar el lema nuevamente para obtener un $M_3(T)$ tal que

$$M(T) \geq M_3(T) \geq M(T) - L\epsilon',$$

donde $\epsilon' = M(T) - M_2(T) \leq L\epsilon$, por lo tanto se cumplirá:

$$M(T) \geq M_3(T) \geq M(T) - L^2\epsilon.$$

Siguiendo este razonamiento sucesivamente obtenemos que

$$M(T) \geq M_n(T) \geq M(T) - L^{n-1}\epsilon,$$

y como $0 < L < 1$ vale la tesis.

□

Observación 2.3.9. *Si uno conoce $M(S)$ para S menor que T puede estimar $M(T)$ utilizando la sucesión anterior. Con los resultados presentes hasta este punto se puede, con una computadora, hallar σ_0. En la sección siguiente se analiza este hecho en detalle para hallar efectivamente la solución óptima.*

2.4. Observaciones sobre la solución

Lema 2.4.1. *Si x es un entero menor o igual que $T-2$ se cumple:*

$$B_0(T+1, T+1-x) \geq B_0(T, T-x).$$

Demostración. Si x no es positivo obviamente se da la igualdad, ya que ambos miembros valen uno. Veamos el caso en que x es mayor que cero. Consideremos $\sigma_0 = \sigma_{0,T+1}$ la estrategia optima en $X(T+1)$. Definamos σ_2 y σ_3 en $X(T)$ como sigue:

$$\sigma_2(t,r) := \begin{cases} \sigma_0(T+1, r+1) & \text{si } t = T \text{ y } r = 2, 3, \ldots, T-1 \\ \sigma_0(t,r) & \text{si } t \leq T-1 \text{ y } r = 2, 3, \ldots, t-1 \end{cases}$$

$$\sigma_3(t,r) := \sigma_0(t+1, r) \qquad 2 \leq r < t \leq T.$$

Consideremos la cadena de Markov A que comienza en el estado $(T+1, T+1-x)$ y se rige por la estrategia σ_0; al mismo tiempo consideremos la cadena B que empieza en el estado $(T, T-x)$ y usa la estrategia σ_2 hasta la primera vez que visita el estado $(T, 0)$, si es que esto sucede, y luego sigue con la estrategia σ_3.

Si A llega a un estado absorbente sin visitar el estado $(T+1, 0)$ resultará que A y B alcanzarán sus puntajes objetivo a la misma vez. Si, en cambio, A visita el estado $(T+1, 0)$, que indica que en el primer turno se obtuvo un uno, la cadena B al mismo tiempo estará en el estado $(T, 0)$ y a partir de ese momento jugará con la estrategia σ_3. Es fácil ver que B alcanzará el objetivo al mismo tiempo o antes que A.

Encontramos una estrategia, no estacionaria, para B que alcanza el objetivo en una cantidad de pasos menor o igual que la estrategia óptima para A. A su vez sabemos que hay una estrategia estacionaria que al menos iguala el valor esperado de la cantidad de turnos de la estrategia usada para B. Esto prueba el resultado.

□

Teorema 2.4.2. *Existe un natural T', tal que si $T \geq T'$ existe un natural $R(T) < T$ que es el mínimo que cumple que siempre que r sea mayor o igual que $R(T)$ conviene plantarse. O sea, si $\sigma_{0,T}$ es la estrategia óptima en $X(T)$ se cumple*

$$\sigma_{0,T}(T, r) = 0 \text{ si } r \geq R(T)$$

y

$$\sigma_{0,T}(T, R(T)-1) = 1.$$

Además $R(T+1) \leq R(T) + 1$.

Demostración. Para que se cumpla el teorema lo único que debe pasar es que $\sigma_{0,T}(T, T-1)$ sea igual a cero, a los efectos de que el número $R(T)$ sea menor estricto que T. De la ecuación (2.17) surge que $\sigma_{0,T}(T, T-1)$ vale cero si y sólo si:

$$1 + M(1) < \frac{1}{6}\left(1 + M(T) + \sum_{k=2}^{6} B_0(T, T-1+k)\right) = \frac{6 + M(T)}{6}.$$

Despejando de la ecuación (2.16) se obtiene que $M(1)$ vale $\frac{6}{5}$ por lo que la condición anterior queda:

$$1 + \frac{6}{5} < \frac{6 + M(T)}{6}.$$

Por lo tanto $\sigma_{0,T}(T, T-1) = 0$ si y sólo si $M(T) > \frac{36}{5}$. Sabemos que M(T) es creciente y claramente se cumple que $lim_{T\to\infty} M(T) = \infty$. Entonces T' será el menor T tal que $M(T) > \frac{36}{5}$. De los cálculos hechos en la próxima sección, cuyos resultados se presentan en la tabla de la figura 2.1 se obtiene que T' es 57.

Veamos ahora, $R(T+1) \leq R(T)+1$. Para eso veremos que fijando un T mayor que T' si $\sigma_0(T, T-j)$ es cero para $j = 1, 2, \ldots, k$ entonces $\sigma_0(T+1, T+1-j)$ también será cero. De acá se deduce inmediatamente la tesis.
Para que $\sigma_0(T, T-j)$ sea cero es necesario y suficiente, en virtud de la ecuación (2.17), que se cumpla la siguiente condición:

$$1 + M(j) < \frac{1}{6}\left(1 + M(T) + \sum_{k=2}^{6} B_0(T, T-j+k)\right).$$

Sabemos, por los lemas 2.3.5 y 2.4.1, que el segundo término es menor o igual que $\frac{1}{6}\left(1 + M(T+1) + \sum_{k=2}^{6} B_0(T+1, T+1-j+k)\right)$. Por lo tanto se cumple

$$1 + M(j) < \frac{1}{6}\left(1 + M(T+1) + \sum_{k=2}^{6} B_0(T+1, T+1-j+k)\right),$$

que es equivalente a que $\sigma_0(T+1, T+1-j)$ sea cero. Concluyendo la demostración.

□

2.5. Los cálculos

La idea en esta sección es aplicar los resultados anteriores, en una situación concreta, para calcular $\sigma_0(t,r)$, $M(t)$ y $B_0(t,r)$ para $0 < r < t \leq T$. Obviamente fijado el objetivo T, la cantidad de datos a calcular es finita. Una vez hechos estos cálculos tendremos la forma de jugar de modo que la esperanza de la cantidad de turnos que requiere alcanzar dicho objetivos sea mínima.

Los resultados en los que se basan las simulaciones son: la observación 2.3.8 que permite estimar $M(T)$ a partir de $M(T-1)$; y la ecuación (2.17) que da una forma recursiva de calcular $B_0(T,r)$ a partir de $M(T)$, $M(T-r)$ y $B_0(T,r+k)$.

Notar que calcular $B_0(t,r)$ para $1 < r < t \leq T$ tiene como consecuencia conocer la estrategia $\sigma_0(t,r)$, ya que la manera de calcular $B_0(t,r)$ es mediante la fórmula (2.17), que establece:

$$B_0(T,r) = \text{mín}\left\{1 + M(T-r),\ \frac{1}{6}\left(1 + M(T) + \sum_{k=2}^{6} B_0(T,r+k)\right)\right\}.$$

Y sabemos que si el mínimo es $\frac{1}{6}\left(1 + M(T) + \sum_{k=2}^{6} B_0(T,r+k)\right)$ debemos seguir $(\sigma_0(t,r) = 1)$ y en otro caso parar $(\sigma_0(t,r) = 0)$.

Veamos, primero, un pseudocódigo que muestra como se aplica el resultado obtenido en 2.3.8 para calcular $M(T)$ conociendo $M(t)$ para $t < T$:

Cálculo de $M(T)$:

1. $M(T) := 0$
2. $M_1(T) := M(T-1)$
3. Mientras $M(T)$ sea distinto de $M_1(T)$
   - *a)* $M(T) := M_1(T)$
   - *b)* para r empezando de $t-1$ bajando hasta 2
     1) $B(T,r) := \text{mín}\{1 + M(T-r), \frac{1}{6}(1 + M(T) + \sum_{k=2}^{6} B(T,r+k))\}$
   - *c)* $M_1(T) := \frac{1}{6}(1 + M(T) + \sum_{k=2}^{6} B(T,k))$

Observaciones sobre el algoritmo

- Notar que para calcular $B(T,r)$ se necesita conocer algunos $B(T,r')$ con $r' > r$. Es por eso que la iteración se hace disminuyendo la variable r.

- La condición del *mientras* tiene sentido en una computadora, trabajando con números exactos nunca se va a dar la igualdad de un paso a otro, esto es equivalente a parar cuando los términos de la sucesión están ϵ-próximos al número buscado, donde ϵ es la precisión de la máquina.

Ahora que tenemos calculado, o mejor dicho sabemos calcular, $M(t)$ para $t \leq T$, estamos en condiciones de calcular $B_0(T, r)$ para $1 < r < T$ y la estrategia $\sigma_0(T, r)$ para $1 < r < T$. El siguiente pseudocódigo realiza esos cálculos basándose en la ecuación (2.17):

Cálculo de $B_0(T, r)$ y $\sigma_0(T, r)$:

1. para r empezando de $t - 1$ bajando hasta 2
   a) $aux\_parar := 1 + M(T - r)$
   b) $aux\_seguir := \frac{1}{6}(1 + M(T) + \sum_{k=2}^{6} B(T, r + k))$
   c) si aux_parar es menor que aux_seguir entonces
      1) $\sigma(T, r) := 0$
      2) $B_0(T, r) := aux\_parar$

      sino
      1) $\sigma(T, r) := 1$
      2) $B_0(T, r) := aux\_seguir$

El siguiente pseudocódigo combina los anteriores para lograr el objetivo de la sección:

Hallar solución óptima y M(t) para todo $t \leq T$

1. para t empezando en 2 hasta T
   a) Cálculo de M(t)
   b) Cálculo de $B_0(t, r)$ y $\sigma_0(t, r)$:

2.5.1. Resultados para $T = 200$

Aquí se presentan los resultados obtenidos para $T = 200$ de la ejecución de los algoritmos presentados anteriormente. Los resultados que aparecen confirman y amplían los que aparecen en el artículo de J. Haigh y M. Roters [5].

Vale aclarar que con los cálculos para $T = 200$ se tiene la forma óptima (en el sentido de minimizar la esperanza de la cantidad de turnos) de jugar cuando el juego es a alcanzar t para cualquier t menor que 200.

M (T) para T = 1..209

	0	1	2	3	4	5	6	7	8	9
0		1,20	1,20	1,24	1,29	1,34	1,41	1,50	1,55	1,62
10	1,69	1,77	1,86	1,95	2,03	2,13	2,22	2,33	2,44	2,55
20	2,66	2,79	2,92	3,05	3,19	3,34	3,49	3,65	3,82	4,00
30	4,10	4,19	4,29	4,39	4,49	4,59	4,69	4,80	4,91	5,02
40	5,13	5,25	5,37	5,49	5,62	5,75	5,88	6,01	6,15	6,29
50	6,43	6,58	6,73	6,85	6,95	7,05	7,16	7,27	7,38	7,49
60	7,60	7,72	7,84	7,96	8,08	8,20	8,32	8,45	8,58	8,71
70	8,84	8,97	9,11	9,25	9,39	9,52	9,63	9,74	9,85	9,96
80	10,08	10,19	10,30	10,42	10,54	10,66	10,78	10,90	11,03	11,15
90	11,28	11,41	11,53	11,66	11,80	11,93	12,07	12,20	12,32	12,43
100	12,55	12,66	12,77	12,89	13,01	13,12	13,24	13,36	13,48	13,61
110	13,73	13,85	13,98	14,11	14,23	14,36	14,49	14,62	14,76	14,89
120	15,01	15,13	15,24	15,36	15,47	15,59	15,71	15,83	15,94	16,06
130	16,19	16,31	16,43	16,56	16,68	16,81	16,93	17,06	17,19	17,32
140	17,45	17,58	17,71	17,82	17,94	18,06	18,17	18,29	18,41	18,53
150	18,65	18,77	18,89	19,01	19,13	19,26	19,38	19,51	19,63	19,76
160	19,89	20,02	20,15	20,28	20,40	20,52	20,64	20,76	20,87	20,99
170	21,11	21,23	21,35	21,47	21,59	21,71	21,84	21,96	22,08	22,21
180	22,33	22,46	22,59	22,72	22,84	22,97	23,10	23,22	23,34	23,46
190	23,58	23,69	23,81	23,93	24,05	24,17	24,29	24,42	24,54	24,66
200	24,79	24,91	25,03	25,16	25,29	25,41	25,54	25,67	25,80	25,92

Figura 2.1: Función M

La tabla de la figura 2.1 muestra el valor esperado $M(T)$ de la cantidad de turnos que requiere el juego si se juega con la estrategia óptima.

El valor de $\sigma_0(t, r)$, o sea, la estrategia óptima para este caso, se muestra a continuación:

- Si $t \leq 29$ se tiene $\sigma_0(t, r) = 1$. Esto quiere decir que conviene tratar de alcanzar todos los puntos en un turno solo.
- Si $t > 13$ y $r \leq 13$, entonces $\sigma_0(t, r) = 1$. Es decir, salvo en los casos triviales, nunca conviene plantarse con menos de 14 puntos.
- Si $t \geq 57$ y $r \geq 24$, entonces $\sigma_0(t, r) = 0$. Notar que el teorema 2.4.2 nos aseguraba que para los $t \geq 57$ había un número $R(t) < t$ tal que si $r \geq R(t)$ $\sigma_0(t, r) = 0$. Ahora podemos agregar $R(t) \leq 24$

Los resultados intermedios, que no aparecen en los puntos anteriores, se muestran en las gráficas siguientes, figuras 2.2, 2.3 y 2.4. Los cuadraditos en blanco indican 1 (que se debe seguir) y los negros 0 (que se debe parar).

A partir de las gráficas se observa algo muy interesante, y es que la estrategia cada vez se parece más a τ_{20} y a τ_{21} cuando t crece. Es intuitivo pensar que si el objetivo a alcanzar está muy lejos conviene tratar de maximizar la esperanza de un turno. Las simulaciones que hice para $t > 250$, cuya tabla no se agregó, muestran que la solución cumple $\sigma(t, r) = 1$ si $r < 20$ y $\sigma(t, r) = 0$ si $r > 21$.

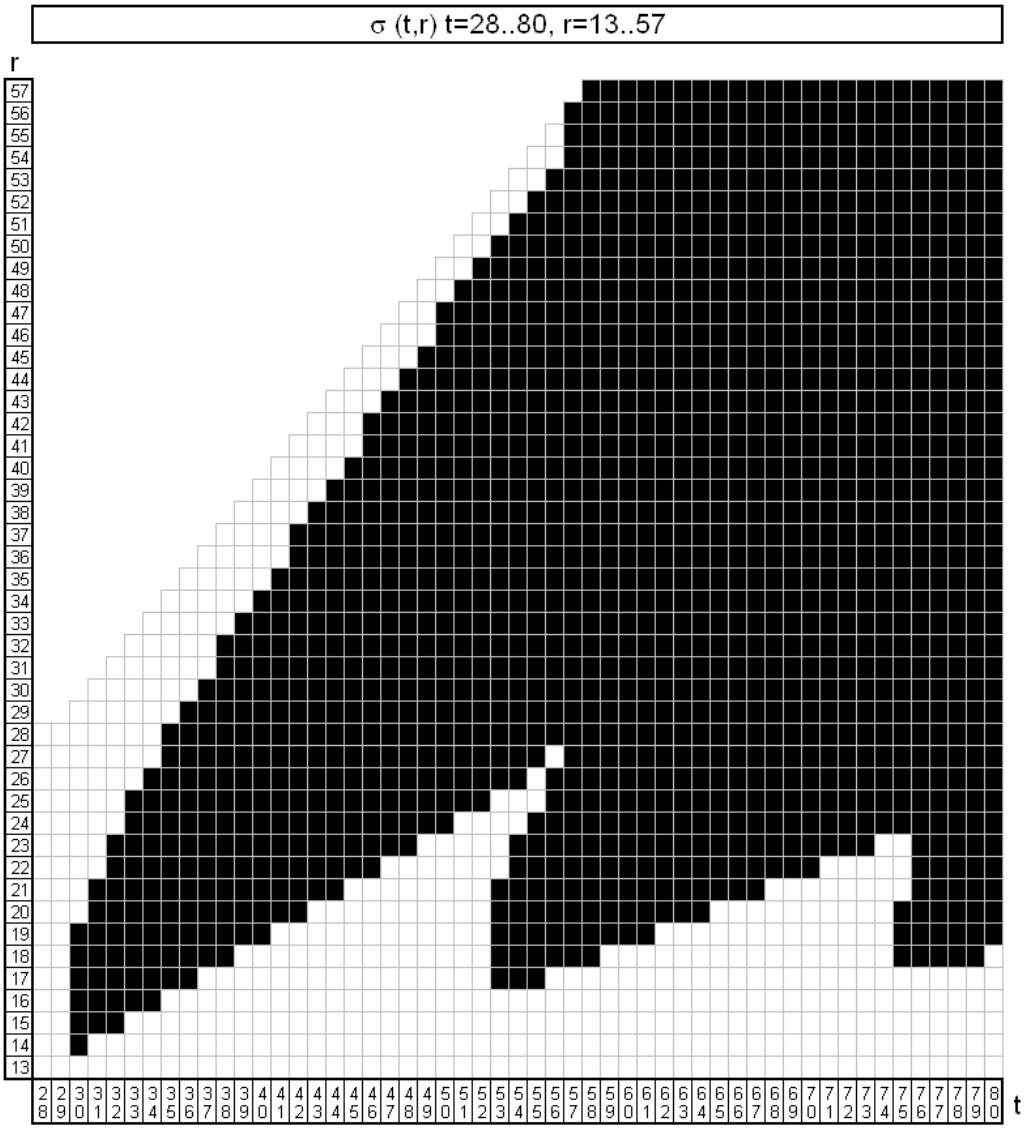

Figura 2.2: Función sigma (primera parte), blanco-seguir, negro-parar

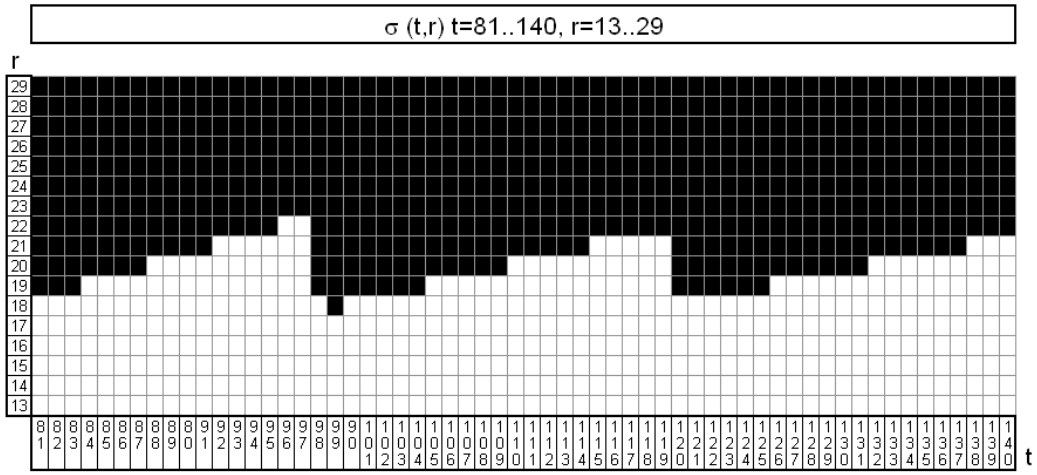

Figura 2.3: función sigma (segunda parte), blanco-seguir, negro-parar

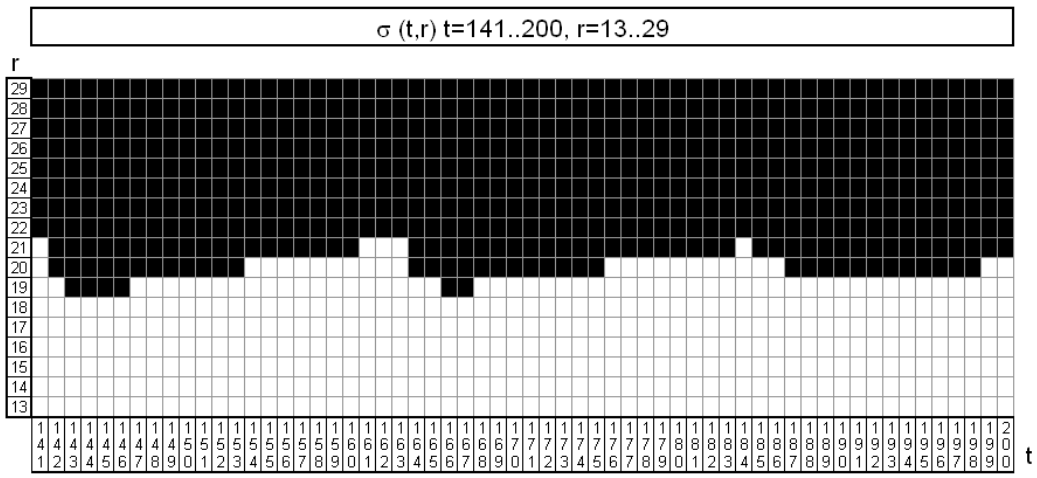

Figura 2.4: función sigma (tercera parte), blanco-seguir, negro-parar

Capítulo 3

Maximizar la probabilidad de ganar

3.1. Motivación

Cuando uno juega a los dados quiere ganar, y una estrategia para jugar se va a considerar buena si tiene una probabilidad alta de ganar. Sin lugar a duda los resultados obtenidos en el capítulo 1 y sobre todo en el capítulo 2 proponen una estrategia que resulta buena, pero no necesariamente óptima cuando se juega de a varios jugadores. La principal carencia que tiene dicha estrategia es no tener en cuenta el puntaje del resto de los jugadores.

Está claro que crear una estrategia que sea óptima para el juego real no es tarea fácil, ya que debería depender de la cantidad de contrincantes y de sus estrategias. En este capítulo tenemos un objetivo mucho más simple: Proponer una estrategia heurística que al jugar contra un otro que juega con la estrategia σ_0, obtenida en el capítulo anterior, tenga más chance de ganar.

3.2. La estrategia heurística

Llamémosle jugador A al que juega con la estrategia σ_0 y p_A a la probabilidad que éste gane. El jugador B es el que juega con la nueva estrategia, que le llamaremos σ_B, y la probabilidad de que gane es p_B. Supongamos que el juego es a 200 puntos, o sea $T = 200$. Y que se sortea quién tira el dado primero. Es claro que $p_A + p_B = 1$. Si el jugador B jugara con la estrategia σ_0, o sea $\sigma_B = \sigma_0$, tendríamos que

$$p_A = p_B = 0{,}5.$$

La idea es proponer una estrategia heurística σ_B de modo que $p_B > 0{,}5$.

Un buen punto de partida para diseñar la estrategia σ_B es la estrategia σ_0. ¿Qué le deberíamos cambiar para aumentar la probabilidad de que gane B ($p_B > 0{,}5$)?

Si en los primeros turnos la suerte nos acompaña, parece razonable seguir con la estrategia σ_0 para ganar lo antes posible. Pero si en los primeros turnos el jugador A sale favorecido, y nos saca una buena ventaja, quizás nos convenga tomar estrategias más arriesgadas que nos den alguna esperanza de ganar. Para ejemplificar este hecho consideremos el siguiente escenario:

Supongamos que empieza nuestro turno y nos faltan 30 puntos ($t_b = 30$) para ganar. Al jugador A le restan sólo 6 puntos ($t_a = 6$) para ganar. Si nos guiamos por la estrategia σ_0, dividiríamos los 30 puntos en dos turnos aproximadamente iguales, lo que implica darle al jugador A la chance de ganarnos. La pregunta que se plantea es:

¿En este caso conviene alejarse de la estrategia que minimiza la esperanza de la cantidad de turnos y tratar de ganar en un turno solo?

Para responder esta pregunta hicimos simulaciones que calculan la proporción de veces que ganamos si jugamos con la estrategia σ_0 y la proporción de veces que ganamos si tratamos de alcanzar los 30 puntos en un solo turno. Se simularon 100.000 jugadas, empezando con los puntajes dichos, para el caso en que el jugador B juega con la estrategia σ_0 y otras 100.000 en que el jugador B intenta alcanzar los 30 puntos en un solo turno. Los resultados obtenidos son los siguientes:

- 0,31 es la proporción de triunfos de B si trata de ganar de una.
- 0,10 para el caso en que B juega con la estrategia σ_0

Evidentemente la nueva estrategia es mejor. Si hacemos este mismo experimento, pero cuando nos faltan 36 puntos y al jugador A le faltan 6 también obtenemos resultados alentadores. Las proporciones para este caso son 0,24 contra 0,08.

En la tabla mostrada en la figura 3.1 se compara las proporciones para $1 \leq t_a \leq 40$ y $30 \leq t_b \leq 100$. No tendría sentido hacerlo para $t_b < 30$ ya que la estrategia σ_0 trataría de alcanzar el puntaje en un turno.

En el eje vertical está t_a y en el horizontal t_b. El cuadrado blanco quiere decir que la estrategia nueva mejora σ_0 (es el caso de los ejemplos presentados), y el cuadrado es negro cuando la proporción de triunfos cuando el jugador B juega con σ_0 es mayor que cuando trata de ganar en un turno. La forma en que se hizo la simulación fue comparando la proporción de triunfos del jugador B si jugaba con una estrategia o con la otra jugando 10.000 veces para cada par (t_A, t_B). Vale aclarar que en los casos críticos, es decir, cuando en la gráfica aparecen dos celdas pegadas de distinto color, que se podría pensar que hay errores de simulación debidos al azar, se volvió a intentar con 100.000 juegos para confirmar los resultados. De modo que es muy poco probable que aparezcan valores cambiados debido a dicho error.

3.2.1. Los resultados obtenidos

Considerando la estrategia σ_B, que en los casos presentados en blanco, en la tabla de la figura 3.1, intenta alcanzar el puntaje en un turno y en cualquier otro caso juega como σ_0, simulamos 100.000 jugadas de nuestro jugador B contra el jugador A y resultó que la proporción de triunfos de nuestro jugador es de aproximadamente 0,52. No es todo lo que hubiésemos querido pero mejoró levemente la estrategia del capítulo 2.

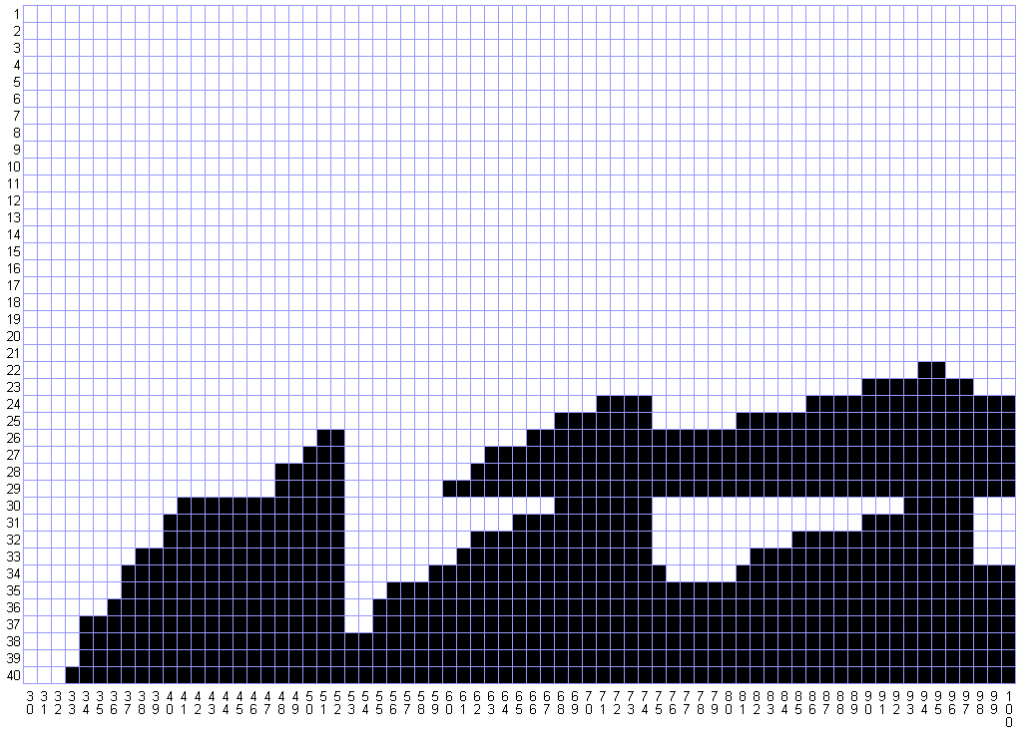

Figura 3.1: Comparación entre proporción de triunfos de B, tratando de alcanzar todo el puntaje en un turno (blanco-si dicha proporción es mayor) y jugando con la estrategia σ_0 (negro-en el caso en que σ_0 es mejor). En el eje vertical se gráfica t_a y en el horizontal t_b.

Bibliografía

[1] Y. Chow, H. Robbins, D. Siegmund, *Great Expectations: The Theory of Optimal Stopping*, Houghton Mifflin Company Boston, USA, 1977

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

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

[4] T. Ferguson, *Optimal Stopping and Applications*, Mathematics Department, UCLA, disponible en

`http://www.math.ucla.edu/~tom/Stopping/Contents.html`

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

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

[7] V. Petrov y E. Mordecki, *Teoría de Probabilidades*, Editorial URSS, agosto 2002

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

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

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