UNIVERSIDAD AUTÓNOMA METROPOLITANA
Unidad Iztapalapa

UN JUGADOR vs. UN CASINO:

APUESTAS SECUENCIALES ÓPTIMAS

Reporte de los cursos:
Seminario de Investigación I y Seminario de Investigación II

Licenciatura en Matemáticas

Presenta: Felipe Hernández Cardona
Asesor: Dr. José Raúl Montes de Oca M.
División: Ciencias Básicas e Ingeniería
Departamento de Matemáticas
México D.F., 2013

CONTENIDO

INTRODUCCIÓN

Es sabido (véase [14]) que el hombre primitivo apostaba tanto en ciertas "competencias deportivas" como en juegos de azar. En estas líneas se descubre cómo nace el ser humano como apostador.

Lo siguiente gira en torno a la Nueva Edad de Piedra, el llamado Neolítico, comprendida entre 7,000 y 4,000 años A.C. Es en este tiempo cuando se han encontrado los primeros vestigios acerca del interés del ser humano por el juego.

EL HOMBRE Y EL AZAR, UNA RELACIÓN ANCESTRAL

No es posible señalar el origen del juego en un momento exacto de la historia. La lotería, las apuestas deportivas *on-line* o los modernos casinos de hoy día tienen como base el juego en las antiguas sociedades prehistóricas. El interés por el entretenimiento y la curiosidad por el azar se hayan implícitos en el ser humano.

El hombre primitivo ya se interesaba por el juego miles y miles de años atrás. En la parte occidental del estado de Tennessee en Estados Unidos de Norte América, por ejemplo, se encontró un artefacto que data de hace más de 7,000 años. Se trata de una porción de hueso occipital de venado, tallado y pulimentado totalmente, y atravesado con un asta del mismo animal. Los expertos lo han interpretado como un instrumento lúdico (es decir, de juego) parecido al juego del anillo y la varilla utilizado por los indios del norte de América.

Se han hallado en diversas excavaciones arqueológicas algunos objetos prehistóricos destinados tanto al uso lúdico como a ciertos fines adivinatorios. Se trata de los astrágalos o tabas, huesos del tarso de un mamífero que se usaban como dados. Los populares dados de seis caras modernos tienen su origen en estos huesos de animal.

En ocasiones, cuando en las sociedades prehistóricas había que repartir un bien o distribuir una cierta tarea, como ir de caza, los miembros tribales recurrían a estos dados para encomendar la decisión a los dioses.

Es probable que estos sorteos ante las divinidades se repitieran más tarde simplemente para volver a experimentar la curiosidad ante lo desconocido, la

tentación ante el riesgo y la satisfacción por la ganancia. Cada jugador aportaba entonces un bien y el azar decidía la suerte del mismo.

Competiciones deportivas y apuestas en la Prehistoria

Si se acude a la RAE (Institución española especializada en lexicografía, gramática, ortografía y bases de datos lingüísticos) para develar el significado de la palabra deporte, se encuentra la siguiente acepción: *"actividad física, ejercida como juego o competición, cuya práctica supone entrenamiento y sujeción a normas"*. El hombre primitivo ya realizaba ciertas competiciones deportivas, aunque éstas estuvieran sujetas a unas reglas muy arcaicas y simples.

Algunos expertos aseguran que los juegos de azar tienen su origen en ciertas competiciones deportivas prehistóricas, en las que los rivales, en una carrera o en una lucha, apostaban antes del inicio del evento para obtener así una doble victoria: la moral y la económica. Esto ha sido definido por el antropólogo Geertz como "*apuesta pareja*" (véase [9]).

Los débiles y poco ágiles, imposibilitados en la práctica física, fueron desarrollando otros juegos no físicos que simulaban la competición real, como el backgammon, el juego de mesa más antiguo del que se tiene conocimiento (5,000 A.C.).

Las excavaciones arqueológicas acontecidas recientemente en Ciudad Quemada (Irán) han dado a conocer al mundo el par de dados más antiguos conocidos, dicho par data del 3.000 A.C. Los estudiosos en la materia han dado a conocer que estos dados provienen de un juego de backgammon.

El hombre primitivo, el primero en apostar

Según el antropólogo estadounidense Alfred Kroeber, todos los pueblos prehistóricos del planeta jugaban y tenían sus propios códigos en este aspecto, excluyendo algunos grupos aislados geográficamente. Así, cabe afirmar que el hombre primitivo, además de llevar a cabo ciertos rituales religiosos y manifestaciones culturales, se interesó por *el juego y las apuestas*.

Así se configuró el hombre como apostador, un ser humano que buscaba en la prehistoria un placer por arriesgar alguna propiedad apostándola en un juego o en una práctica física (véase [14]).

CADENAS DE MARKOV

Durante la segunda mitad del siglo XIX, Andrei Markov (matemático ruso) realizó trabajos donde trató ciertos procesos en donde están involucradas componentes aleatorias, los cuales se derivaron en lo que hoy se conoce como Cadenas de Markov (véase [1]).

Grosso modo, las Cadenas de Markov son sucesiones de variables aleatorias en las que el valor de la variable en el futuro depende del valor de la variable en el presente, pero es independiente de la historia de dicha variable, lo cual es conocido como "*la propiedad de Markov"* (véase [16]).

Las Cadenas de Markov, hoy día, se consideran una herramienta esencial en disciplinas como la economía, la ingeniería, la investigación de operaciones y muchas otras (véase [1]). Una de ellas "Juegos de Apuestas", de la cual en las subsecuentes páginas se tratará la forma de analizar un modelo de un juego de apuestas contra un casino, del cual nos interesa saber:

¿Cuál es la estrategia de juego que maximiza la probabilidad de que el jugador gane una riqueza de N pesos antes de que se arruine?

Para ello es necesario abordar una extensión de las cadenas de Markov conocida como procesos de decisión de Markov (véase [16]).

PROCESOS DE DECISIÓN DE MARKOV

Al problema de encontrar una estrategia o política de acción en un problema de decisión secuencial que maximice la recompensa esperada en el tiempo se le conoce como proceso de decisión de Markov (PDM).

El trabajo de Markov está estrechamente ligado a las suposiciones de que el agente siempre conocerá el estado en que se encuentra al momento de iniciar la ejecución de sus acciones (observabilidad total), y que la probabilidad de transición de un estado depende sólo del estado y no de la historia (propiedad de Markov). Los PDM fueron introducidos originalmente por Bellman (véase [3]) y han sido estudiados a profundidad en los campos de análisis de decisiones e investigación de operaciones desde los 60's iniciando con el trabajo seminal de Howard (véase [10]). Entre los textos más importantes en el área de PDM se encuentran los trabajos de Bertsekas (véase [4]) y Puterman (véase [15]).

MAYOR GANANCIA VS. CANTIDAD DE APUESTAS

Claramente, algo no menos relevante que nos interesa saber es cómo optimizar la recompensa así como el tiempo esperado de juegos que debe jugar el apostador para lograr la máxima recompensa y así quedar satisfecho.

Para este fin, es necesario analizar las estrategias óptimas, que son dependiendo del caso la estrategia tímida y la estrategia audaz; estas últimas las abordaremos desde dos puntos de vista: el de Ross (véase [17]) y el de Bak (véase [2]).

OBJETIVO

El objetivo principal de este trabajo es mostrar como los Procesos de Decisión de Markov (PDM) son una herramienta funcional para la modelación en probabilidad de estrategias de Juegos de Apuestas, la optimización de los rendimientos de un apostador y el tiempo de juego; no obstante, no se debe perder de vista el hecho de que con los PDM pueden ser modelados diversos problemas más allá de Juegos de Apuestas, por lo cual, desde la introducción hasta el apéndice, todas las secciones del presente deben tener una carga motivacional para el lector, en un sentido de modelaje de situaciones propias con PDM y tal vez algunas que aún no han sido exploradas.

ORGANIZACIÓN DEL TRABAJO

En este trabajo se abordarán técnicas para que un apostador en un juego obtenga una cierta riqueza deseada antes de perderlo todo, acción tal que además debe maximizar la probabilidad de que esto suceda, para lo cual, en primera instancia, se hará una descripción intuitiva de las ideas principales, proporcionando en un contexto histórico algunas proposiciones y teoremas que fueron introducidos por matemáticos del pasado, uno de estos teoremas es el de Dubins y Savage, ubicado en el capítulo 1, mismo en el que se analizará la cantidad de juegos esperados y duración en el próximo juego, y se bifurcará en dos tipos de juego: el audaz y el tímido.

Posteriormente, en el capítulo 2, será expuesta una herramienta conocida como: Procesos de Decisión de Markov (PDM), ya que, para dar formalismo a lo abarcado en el capítulo anterior, es necesario el uso de ellos. En este segundo capítulo, serán explicados y sustentados resultados como: La Ecuación de

Optimalidad, la cual trata de políticas óptimas, también se encontrará una condición necesaria para que una política sea óptima.

Seguido de lo anterior, se tendrán bases para aplicar esta teoría a Juegos de Apuestas, en el capítulo 3; aquí se desmenuzará y formalizará cómo es que el juego tímido maximiza la probabilidad de que un apostador nunca alcance una fortuna N y el tiempo de juego si $p \geq \frac{1}{2}$, también cómo la estrategia audaz maximiza la probabilidad de alcanzar N en el tiempo n si $p \leq \frac{1}{2}$, entre otros.

En la última parte de este capítulo se tratará un problema de aplicación que definimos como:

Sea un apostador con una riqueza inicial de i-pesos, dónde el casino contra el cual apostará tiene la siguiente política:

- Si se tiene i-pesos, se permite apostar cualquier entero positivo menor o igual a i.
- Si se apuesta a-pesos, entonces:
    a. Se gana a con probabilidad p.
    b. Se pierde a con probabilidad $1 - p$.

Y nos interesa saber: ¿Cuál es la estrategia de juego que maximiza la probabilidad de que el jugador gane una riqueza de N-pesos antes de quedar en la ruina?

Problema tal que se modela como un PDM.

Finalmente se enfatizarán los hechos más importantes de cada capítulo como conclusiones en el capítulo 4.

Capítulo 1. JUEGOS DE APUESTAS: ANTECEDENTES

En este primer capítulo, como su título dice, se tratarán antecedentes de Juegos de Apuestas extraídos de [2], esto de manera intuitiva, no obstante se enunciará una serie de resultados de los cuales no todos serán demostrados. Empezaremos suponiendo la siguiente idea:

Una persona que posee una cantidad igual a A pesos decide arriesgarse con la esperanza de aumentarla a B pesos.

Esto podría representar el último intento de un jugador que ha perdido casi todo su dinero en Las Vegas, y ahora busca ganar lo suficiente para el transporte a casa. Se encuentra con un juego en el que la probabilidad de ganar es p y en una apuesta de A pesos potencialmente paga A pesos, por supuesto, si pierde, pierde ese dinero.

En su último esfuerzo, continúa apostando la mayor cantidad posible hacia el logro, pero sin exceder su meta. Así, si tiene actualmente A pesos y $A < \frac{B}{2}$, arriesga todo, si se trata de más de medio camino de su meta, apuesta la diferencia, $B - A$, cuando gane esa cantidad traerá en su bolso exactamente B pesos. Se detiene sólo cuando, o bien ha perdido todo su dinero, o ha alcanzado su objetivo B.

El jugador podría describirse como desesperado, aunque el adjetivo ansioso parece más apropiado, ya que aparte de la connotación de nerviosismo por el resultado, también sugiere el deseo de llegar a una conclusión lo antes posible. En este capítulo se abordarán ambos aspectos del juego: la posibilidad del jugador de llegar a su objetivo (que llamaremos éxito), y el número esperado de los juegos hasta que, o bien ha logrado un éxito, o ha perdido todo su dinero (que llamaremos fracaso).

Cabe mencionar que, este análisis también se aplica para calmar perfectamente a una persona dispuesta a arriesgar una cantidad fija de los ingresos disponibles por la oportunidad de una determinada suma de dinero.

Además, analizaremos el hecho de que un incremento de las apuestas en el caso clásico incrementa la probabilidad del suceso cuando $p < \frac{1}{2}$, lo cual da pie al teorema

de aproximación de la optimalidad del juego ansioso: "teorema 2 (Dubins y Savage (1976))".

Por último pondremos especial énfasis en $D(A, B, p)$, el número de juegos esperados o, duración en el próximo juego ansioso.

1.1. ANTECEDENTES DE ALGUNAS APLICACIONES.

Jugada tímida: Apostando continuamente 1.

Esta es una variante del problema clásico conocido como "La ruina del apostador" (véase [A.1.2]). En este caso, un jugador con A pesos, continuamente apuesta \$1 frente a un oponente con $(B - A)$ pesos hasta que uno de ellos es "arruinado". Si el jugador con la cantidad original A tiene probabilidad p de ganar cada juego, entonces su probabilidad de no ser arruinado (su oportunidad de acumular B pesos antes de perderlo todo, es decir, antes de llegar a 0) está dada por:

$$P(A, B, p) = \begin{cases} \dfrac{\left(\frac{q}{p}\right)^{A} - 1}{\left(\frac{q}{p}\right)^{B} - 1} & si\ p \neq \frac{1}{2}, q = 1 - p \\ \dfrac{A}{B} & si\ p = \frac{1}{2} \end{cases} . \qquad (1)$$

Obviamente, el problema de nerviosismo de nuestro jugador, puede ser similar a una lucha entre dos individuos, uno con una fortuna inicial de A, y el otro con una fortuna inicial de $(B - A)$, terminando cuando uno de ellos sea arruinado. La única diferencia, por supuesto, es que el jugador nervioso apostará una cantidad diferente en cada juego, con el afán de satisfacerse pero no excederse de su meta.

Para distinguir entre las diferentes pero relacionadas variables en discusión, p y q siempre denotarán la probabilidad de que el apostador gane y pierda respectivamente, en cada apuesta. $\boldsymbol{P}(A, B, p)$ y $\boldsymbol{D}(A, B, p)$ representarán la probabilidad del suceso y el número esperado de juegos (o duración) en el caso clásico, con apuestas uniformes de \$1 en cada uno. $P(A, B, p)$ y $D(A, B, p)$ denotarán la probabilidad y duración correspondientes para el apostador ansioso.

Para este fin, primero presentaremos una extremadamente simple prueba de que $P\left(A, B, \frac{1}{2}\right) = \frac{A}{B}$. Generalizaremos el resultado y mostraremos la idea clave de cómo la prueba fue usada por D'Moivre (véase [5]) para obtener la fórmula (1) para $\boldsymbol{P}(A, B, p)$, incluso cuando $p \neq \frac{1}{2}$.

Entonces, para valores de p en general, encontraremos fórmulas explícitas para $P(A,B,p)$ y $D(A,B,p)$. También citaremos el teorema de Dubins y Savage (véase [6]) en una extrema propiedad de $P(A,B,p)$, derivando un resultado correspondiente para $D(A,B,p)$. En las siguientes dos secciones se concluirán algunas ramificaciones de estas ideas, relacionándolas con los aspectos de certeza de la posibilidad de un evento y loterías, analizaremos más posibilidades teóricas: juego contra un adversario infinitamente rico.

De $P\left(A,B,\frac{1}{2}\right)$ a $\boldsymbol{P}(A,B,p)$ y $\boldsymbol{D}(A,B,p)$.

Insistiendo en que $P\left(A,B,\frac{1}{2}\right)$ es la probabilidad de que un jugador ansioso eventualmente alcance una riqueza B, dado que empieza con A y usa la estrategia ansiosa en un juego con posibilidades de ganar.

Proposición 1. Para todos enteros positivos A,B tales que $A < B$, $P\left(A,B,\frac{1}{2}\right) = \frac{A}{B}$.

Prueba.

Sea $X_0 = A$, y sea X_i la ganancia del jugador en la apuesta i, $i \geq 1$, de modo que $S_n = X_0 + X_1 + X_2 + \cdots + X_n$ representa las posesiones del jugador después de n apuestas. Un simple argumento por inducción muestra que cada X_i tiene un valor positivo y un correspondiente valor negativo con probabilidades iguales. Así, $E[X_i] = 0,\ i \geq 1$, y para todo entero positivo n,

$$E[S_n] = X_0 = A.$$

La probabilidad de que el juego continúe infinitamente es 0 dado que la probabilidad de que el juego continúe más allá de n juegos es a lo más $\frac{1}{2^n}$. Entonces, con probabilidad 1, S_n converge en el límite a una función S, la cual, de acuerdo a la estrategia del jugador, tiene solo dos posibles valores: B con probabilidad $P\left(A,B,\frac{1}{2}\right)$, o 0 con la probabilidad complementaria. Así, por un lado,

$$E[S] = \lim_{n\to\infty} E[S_n] = A,$$

mientras que por el otro lado,

$$E[S] = BP\left(A,B,\frac{1}{2}\right)$$

de la definición de valor esperado. Una comparación de las dos expresiones para $E[S]$ muestra que $P\left(A, B, \frac{1}{2}\right) = \frac{A}{B}$. Una prueba formal de que $E[S] = \lim_{n\to\infty} E[S_n] = A$ puede darse usando el teorema de convergencia en la frontera.

Esta proposición puede ser aplicada con una ligera modificación para obtener el mismo resultado que en el problema clásico cuando $p = \frac{1}{2}$, de hecho la proposición puede ser extendida a una variedad de casos, incluyendo apuestas fijas de algún tamaño, o algún otro tipo de apuesta, justo como el que obtuvimos cuando probabilidades reales son ofrecidas. El único requisito es que el tipo de juego sea:

C1. Existen dos números positivos fijos m y q tales que, en cada juego, la probabilidad de perder por lo menos m es, por lo menos q, o

C1'. Existen dos números positivos fijos m y p tales que, en cada juego, la probabilidad de ganar por lo menos m es por lo menos p (cualquiera de las dos condiciones garantiza que la probabilidad de continuar infinitamente es cero)

C2. El único final posible resulta ser 0 o B.

C3. Cada juego es justo, *i, e.*, $E[X_i] = 0$ para toda $i \geq 1$.

Así, tenemos:

Proposición 2 (Generalización de la proposición 1). La probabilidad de que un apostador con riqueza inicial A alcance una riqueza B es $\frac{A}{B}$. Siempre que la apuesta individual y la estrategia global satisfagan las condiciones C1-C3.

De acuerdo con la condición C3, la sucesión de variables aleatorias $\{S_n\} = \{X_0 + X_1 + \cdots + X_n\}$ es una martingala. (En teoría de probabilidad, la **martingala** - galicismo de *martingale*- es un determinado proceso estocástico. Sin embargo, se conoce comúnmente con este nombre a un método de apuesta en juegos de azar consistente en multiplicar sucesivamente en caso de pérdida una apuesta inicial determinada. En el momento de ganar la apuesta, el proceso se iniciaría de nuevo). La proposición 1 puede ser vista como un ejemplo de teoría de martingala. Es interesante notar, sin embargo, como la propiedad definida por la condición C3, fue usada por D'Moivre para resolver el problema clásico de la ruina de un jugador casi 200 años antes de que se desarrollara la teoría de martingalas.

Las soluciones del problema clásico de la ruina de un jugador datan más o menos de 1654. Aunque la prueba del caso general no fue publicado sino hasta 1711. Edwards (véase [7]) mostró como las ideas expresadas en una serie de cartas entre Pascal y Fermat indican el conocimiento del resultado general. Edwards incluso ofrece una

probable reconstrucción de estas pruebas, basándose en el aprovechamiento de problemas similares y de estas sugerencias se derivó la correspondencia.

Ninguna de las primeras pruebas indican la simplificación de la prueba para $p = \frac{1}{2}$ en la primera prueba publicada de D'Moivre, sin embargo, actualmente se usa el método de generalización de la proposición para resolver el problema clásico en todos los casos, es decir, incluso si $p \neq \frac{1}{2}$. Su ingenio consistía en aprovechar los cambios de los valores de las monedas usadas para las apuestas, asignando dichos valores en cada camino para garantizar que fuera satisfecha la condición C3. Para este fin, imaginó que el jugador con A monedas apiladas, o mejor dicho, que tuvieran la misma unidad de valor, a la moneda en el fondo le es dado el valor $\frac{q}{p}$ y sólo a la que está por arriba le es dado el valor $\left(\frac{q}{p}\right)^2$, etc., hasta la moneda top con un valor de $\left(\frac{q}{p}\right)^A$. Las $(B - A)$ monedas de su oponente son igualmente apiladas y dado el valor $\left(\frac{q}{p}\right)^{A+1}$ para la moneda top, $\left(\frac{q}{p}\right)^{A+2}$ para la que le sigue por debajo, y asi hasta $\left(\frac{q}{p}\right)^B$ para la moneda de hasta abajo. Además, la transferencia de una moneda después de cualquier juego siempre se realiza mediante la colocación de moneda superior del perdedor en la parte superior de la pila del ganador. Así, $E[X_i]$, el valor esperado para el primer jugador en el juego i, esta dada por una combinación de términos de la forma $p\left(\frac{q}{p}\right)^{j+1} - q\left(\frac{q}{p}\right)^j$, todos los cuales son igual a 0. Remplazando A por la nueva fortuna inicial del primer jugador, remplazando B por la suma de la fortuna inicial del segundo jugador, y argumentando como en la proposición de D'Moivre, obtenemos $P(A, B, p) = \left[\left(\frac{q}{p}\right)^1 + \left(\frac{q}{p}\right)^2 + \cdots + \left(\frac{q}{p}\right)^A\right] / \left[\left(\frac{q}{p}\right)^1 + \left(\frac{q}{p}\right)^2 + \cdots + \left(\frac{q}{p}\right)^B\right]$ lo que coincide con la fórmula (1) para valores de p.

Más tarde, D'Moivre atacó el problema de encontrar la duración del problema clásico. Reasignando el valor 1 a cada moneda, notó que la esperanza de ganar para el jugador en cada juego es $(p - q)$. La ganancia total esperada es:

$$[\boldsymbol{P}(A, B, p)](B - A) - [1 - \boldsymbol{P}(A, B, p)]A.$$

Así, usando el hecho de que el producto del número esperado de juegos por la ganancia esperada de juegos sería igual a la ganancia total esperada, D'Moivre concluyo que el número de juegos esperado esta dado por:

$$\boldsymbol{D}(A, B, p) = [A - B\boldsymbol{P}(A, B, p)\,]/(q - p) \quad \text{si } p \neq \tfrac{1}{2}\,. \tag{2}$$

D'Moivre argumentó que no puede ser aplicado si $p = \frac{1}{2}$. En este caso, el resultado puede ser obtenido mediante la resolución de la ecuación diferencial, y se tiene que,

$$\boldsymbol{D}\left(A, B, \frac{1}{2}\right) = A(B - A). \tag{3}$$

Jugada audaz: Incrementando las apuestas en la clásica ruina del jugador.

La relación entre el aprovechamiento del jugador en el problema clásico y el aprovechamiento de nuestro jugador ansioso puede ser vista como sigue. Supongamos que el jugador clásico elevó en un juego su apuesta a \$2, o a alguna cantidad mayor S (técnicamente, por supuesto, esta es la única posibilidad si tanto A como B son divisibles por S). Si este no es el caso, uno podría usar una estrategia mixta, regresando a \$1 por juego, donde la fortuna queda por debajo de S o por encima de $(B - S)$. En nuestra opinión, sin embargo, podríamos simplificar asumiendo que S es un común divisor de A y B. Si $p = \frac{1}{2}$, de acuerdo a nuestra proposición general, el incremento de la apuesta podría no tener efecto en la probabilidad del suceso del jugador. Por otra parte, si $p < \frac{1}{2}$, la probabilidad del suceso crece.

Feller (véase [8]) notó que incrementar las apuestas a S es equivalente a cambiar A y B por $\frac{A}{S}$ y $\frac{B}{S}$ respectivamente, y probó que esto lleva a un incremento de la probabilidad del suceso si $S = 2$. En una más reciente nota, Isaac (véase [11]) dio una buena y natural prueba del resultado general, mostrando que la probabilidad correspondiente de fallar es un decremento en función de S, para todo positivo S. Así,

$$\boldsymbol{P}\left(\frac{A}{S}, \frac{B}{S}, p\right) > \boldsymbol{P}(A, B, p) \text{ para } p < \frac{1}{2},\ S > 1. \tag{4}$$

Dado que la probabilidad de ganar del apostador es la misma que la probabilidad de perder de su oponente, y dado que su oponente está jugando el mismo juego con A remplazado por $(B - A)$ y p remplazado por q, se sigue que $P(A, B, p) = 1 - P((B - A), B, q)$. Junto con la desigualdad anterior, se tiene que:

$$\boldsymbol{P}\left(\frac{A}{S}, \frac{B}{S}, p\right) < \boldsymbol{P}(A, B, p) \text{ para } p > \frac{1}{2},\ S > 1. \tag{5}$$

Dubins y Savage propusieron una prueba no formal de estas desigualdades, pero notaron que para $p < \frac{1}{2}$, la ley fuerte de los grandes números lo simula en buena medida para apuestas grandes. En efecto, como la referencia de la ley fuerte de los grandes números implica que las dos desigualdades anteriores van de la mano con el decremento de la duración del juego. La siguiente proposición amplifica esta idea.

Proposición 3. Para todo $0 \leq p \leq 1$ y toda $S \geq 2$ (la cual divide a A y B), $\boldsymbol{D}\left(\frac{A}{S}, \frac{B}{S}, p\right) < \frac{\boldsymbol{D}(A,B,p)}{S}$.

Prueba.

De acuerdo con la formula (2) de D'Moivre:

$$\frac{\boldsymbol{D}(A,B,p)}{\boldsymbol{D}\left(\frac{A}{S},\frac{B}{S},p\right)}=\frac{S[A-B\boldsymbol{P}(A,B,p)]}{A-B\boldsymbol{P}\left(\frac{A}{S},\frac{B}{S},p\right)}.$$

Si $p<\frac{1}{2}$, $\boldsymbol{P}(A,B,p)$ y $\boldsymbol{P}\left(\frac{A}{S},\frac{B}{S},p\right)$ son menores que $\frac{A}{B}$, entonces, ambas expresiones en el lado derecho de la ecuación son positivas. Se sigue que la desigualdad en nuestra proposición es equivalente a la desigualdad (4). Similarmente, si $p>\frac{1}{2}$ la proposición es equivalente a la desigualdad (5). Finalmente, si $p=\frac{1}{2}$, podemos aplicar la formula (3) para obtener un resultado más explicito:

$$\boldsymbol{D}\left(\frac{A}{S},\frac{B}{S},\frac{1}{2}\right)=\frac{\boldsymbol{D}\left(A,B,\frac{1}{2}\right)}{S^2}.$$

■

De acuerdo a la proposición 3, $\boldsymbol{D}(2,6,p)\geq 2\boldsymbol{D}(1,3,p)$ para todo p, y $\boldsymbol{D}\left(2,6,\frac{1}{2}\right)=4\boldsymbol{D}\left(1,3,\frac{1}{2}\right)$. Algunos ejemplos de estos valores y los correspondientes valores de $\boldsymbol{P}$ están dados en la siguiente figura para $0.1\leq p\leq 0.9$.

p	$\boldsymbol{P}(1,3,p)$	$\boldsymbol{P}(2,6,p)$	$\boldsymbol{D}(1,3,p)$	$\boldsymbol{D}(2,6,p)$	$\frac{\boldsymbol{D}(2,6,p)}{\boldsymbol{D}(1,3,p)}$
0.1	0.0110	0.0002	1.2088	2.4989	2.0672
0.2	0.0476	0.0037	1.4286	3.2967	2.3077
0.3	0.1139	0.0277	1.6456	4.5843	2.7859
0.4	0.2105	0.1203	1.8421	6.3910	3.4694
0.5	0.3333	0.3333	2	8	4
0.6	0.4737	0.6090	2.1053	8.2707	3.9286
0.7	0.6203	0.8214	2.1519	7.3213	3.4022
0.8	0.7619	0.9377	2.1429	6.0440	2.8205
0.9	0.8901	0.9877	2.0879	4.9074	2.3504

Figura 1

$P(A,B,p)$ y $D(A,B,p)$ en relación con $\boldsymbol{P}(A,B,p)$ y $\boldsymbol{D}(A,B,p)$

Para obtener la fórmula general para $P(A,B,p)$, una vez más consideraremos las variables aleatorias $S_i=X_0+\cdots+X_i$, las cuales representa la fortuna del jugador

después de i juegos, y sea $R_i = \frac{S_i}{B}$ el correspondiente radio de su fortuna en su última meta para $i \geq 0$ (Así $R_0 = \frac{A}{B}$). Si $R_i < \frac{1}{2}$, R_{i+1} será, ya sea 0 (con probabilidad q) o $2R_i$ (con probabilidad p), dado que S_{i+1} seria 0 o $2S_i$, en los respectivos casos. Similarmente, si $R_i \geq \frac{1}{2}$, R_{i+1} sería, 1 (con probabilidad p) o $2R_i - 1$ (con probabilidad q), dado que S_{i+1} sería B o $S_i - (B - S_i) = 2S_i - B$.

Los cuatro casos nos permiten categorizar la $i - ésima$ apuesta de dos maneras:

a) La $i - ésima$ apuesta será la última (concluyendo con $S_i = 0$, con probabilidad q si $R_{i-1} \leq \frac{1}{2}$; y concluyendo con $S_i = B$, con probabilidad p, si $R_{i-1} \geq \frac{1}{2}$).

b) Habrá una $(i+1) - ésima$ apuesta. En este caso $R_{i-1} \neq \frac{1}{2}$, y R_{i-1} tiene representación binaria $0.b_1b_2b_3$ ... , R_i será igual a cambiar la primera posición $0.b_2b_3b_4$ .... Esto se sigue dado que $R_{i-1} < \frac{1}{2}$ implica que $b_1 = 0$ y $2R_{i-1} = 0.b_2b_3b_4$ ... , mientras que $R_{i-1} > \frac{1}{2}$ implica que $b_1 = 1$ y $0.b_2b_3b_4$ ... $= 2R_{i-1} - 1$. Procediendo inductivamente, entonces, esto se sigue de que el único valor posible para R_i, es 0 o 1, teniendo una representación binaria igual a un cambio de la $i - ésima$ posición de la representación binaria para R_0. Como un ejemplo, en la figura 2 se muestran los posibles valores de R_i, siendo $R_0 = \frac{A}{B} = \frac{11}{32} = 0.01011$ (en base 2). Obviamente, si R_i es igual a 0 o 1, todas las R_j $j > i$ subsecuentes, tienen el mismo valor. Por simplicidad, estos valores heredados de 0 o 1 tienen que ser omitidos.

Para determinar $P(A, B, p)$, sea $W_i$el evento en el que el jugador alcanza una meta de B pesos en el $i - ésimo$ juego. Note que W_i es la intersección de los eventos $R_i = 1$ y $R_j \neq 0$ o 1, $0 < j < i$. Además, R_i puede ser igual a 1, sólo si, $R_{i-1} \geq \frac{1}{2}$. Así, el primer digito binario de R_{i-1}, el cual es igual al primer digito binario de R_0 debe ser 1. Si $R_0 = \frac{A}{B} = \sum \left(\frac{1}{2}\right)^{n_k}$, donde los n_k son enteros positivos, se sigue que $P_r(W_i) > 0$ sí, y sólo si, $i = n_k$ para alguna k. Para asegurar que R_j es distinto de 0 o 1, $0 < j < i$ mientras $R_i = 1$, cada apuesta $1,2, \dots, n_k$ debe terminar en gane con la excepción de los juegos n_j, $j < k$, el cual debe resultar perdedor. Así, $P_r(W_i) = p^{n_k - k + 1} q^{k-1}$. Combinando estos resultados tenemos el teorema 1.

Teorema 1. Para todo $0 < p < 1$ se tiene que,

$$P(A, B, p) = \sum P_r(W_i) = \sum p^{n_k - k + 1} q^{k-1}, \tag{6}$$

donde la sucesión creciente $\{n_k\}$ representa la primera posición de la representación binaria de $\frac{A}{B}$.

El teorema 1 muestra que $P(A,B,p)$ es actualmente una función del radio $r=\frac{A}{B}$, y la probabilidad p. De hecho, para p fijo, $P(A,B,p)=P(r,p)$ es una función continua de r. Si dos radios $r_1\ y\ r_2$ son suficientemente cercanos, sus representaciones binarias coincidirán en los primeros m digitos, y de acuerdo a (6), la diferencia entre los valores asociados de P no pueden exceder $\sum_{n_k>m} p^{n_k-k+1}q^{k-1}$. Así, es fácil ver que es menor que $q^m\ si\ p\leq\frac{1}{2}$, y que $p^m\ si\ p\geq\frac{1}{2}$. Así, la diferencia tiende a 0 conforme m tiende a infinito. Análogamente, es fácil mostrar que P es una función creciente de r para p fijo, y una función creciente de p para r fijo. Note que si $p=q=\frac{1}{2}$, (6) se convierte en:

$$P\left(A,B,\tfrac{1}{2}\right)=\sum\left(\tfrac{1}{2}\right)^{n_k}=\tfrac{A}{B}.$$

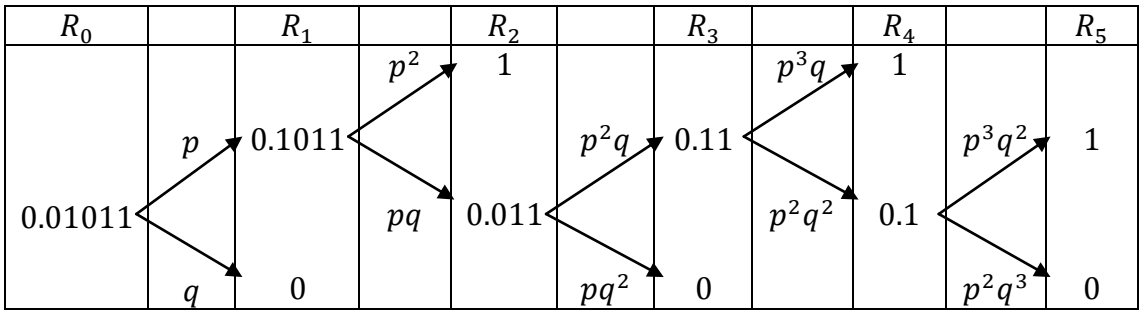

Figura 2

Por ejemplo, en la figura 2, $P(A,B,p)=p^2+p^3q+p^3q^2$.

El hecho de que un incremento en las apuestas en el caso clásico incremente la probabilidad del suceso (con $p<\frac{1}{2}$), sugiere el teorema de aproximación de la optimalidad del jugador ansioso.

Teorema 2 (Dubins y Savage (1976)). Si $p<\frac{1}{2}$, $P(A,B,p)$ es al menos tan grande como la probabilidad ofrecida por alguna estrategia sujeta a una restricción, de que los posibles pagos en cada juego consistan en una perdida igual al monto de la apuesta con probabilidad q, o una ganancia del mismo monto con probabilidad p (véase [6]).

La demostración, involucra algunos resultados generales a cerca de estrategias óptimas y procesos de Markov, y está dada por Dubins y Savage. Así, $P(A,B,p)$ no es únicamente mayor o igual a $\boldsymbol{P}(A,B,p)$, pero es al menos tan grande como alguna estrategia mixta del tipo descrito anteriormente. De hecho, esta última estrategia mixta, involucra la posible más grande apuesta razonable en cada estado.

Ahora, fijaremos nuestra atención en $D(A,B,p)$, el número de juegos esperados, o duración, en el próximo juego ansioso.

Teorema 3. Si $\frac{A}{B}$ es igual a la fracción de términos binarios $\sum_{j=1}^{k}\left(\frac{1}{2}\right)^{n_j}$, entonces:

$$D(A,B,p)=\frac{1}{q}+\left(\frac{1}{p}-\frac{1}{q}\right)\sum_{j=1}^{k-1}\left(\frac{q}{p}\right)^{j-1}p^{n_j}-\left(\frac{q}{p}\right)^{k-2}p^{n_k-1}. \quad (7)$$

Si $\frac{A}{B}$ tiene una representación binaria infinita de la forma anterior, entonces:

$$D(A,B,p)=\frac{1}{q}+\left(\frac{1}{p}-\frac{1}{q}\right)\sum_{j=1}^{\infty}\left(\frac{q}{p}\right)^{j-1}p^{n_j}. \quad (8)$$

Nota. Dado que tanto $D(A,B,p)$ como $P(A,B,p)$ dependen únicamente del radio $\frac{A}{B}$ y de p, podemos tomarlas relativamente primero una u otra. En este caso, la fórmula (7) aplica si $B=2^{n_k}$ para algún entero n_k y (8) es la fórmula indicada para todos los otros casos.

Demostración.

Supóngase que $\frac{A}{B}=\sum\left(\frac{1}{2}\right)^{n_j}$, donde la suma corre desde 1 hasta k si $B=2^{n_k}$, y de 1 a infinito en otro caso. Sea $N(A,B,p)$ el número de juegos hasta que el jugador concluya. Y para todo entero positivo i, sea $d(i)$ la probabilidad de que $N(A,B,p)=i$, con $D(i)$ igual a la probabilidad de que $N(A,B,p)\geq i$.

Por definición, $D(A,B,p)=\sum id(i)$. Lo que encontramos más conveniente, sin embargo, obtenemos $D(A,B,p)$ como suma equivalente de la serie $\sum D(i)$. Para este fin, renombramos la notación R_j que introdujimos en la derivación de la fórmula (6), y notamos que el número de juegos será por lo menos n sí, y sólo si, para toda $j<n-1$,

i. $R_j\neq\frac{1}{2}$,
ii. El $(j+1)-ésimo$ juego resulta ganador si $R_j<\frac{1}{2}$ y,
iii. El $(j+1)-ésimo$ juego perdedor ganador si $R_j>\frac{1}{2}$.

Dada la representación binaria para R_j que es simplemente la representación binaria para $\frac{A}{B}$ empezando con el $(j+1)-ésimo$ dígito, podemos derivar (7) seccionando $\sum D(i)$ en partes, de la cual mostramos la primera, la segunda y la última:

$$\sum_{i=1}^{n_1}D(i)=1+p+p^2+\cdots+p^{n_1-1}$$

$$\sum_{i=1}^{n_2} D(i) = p^{n_1-1}q(1 + p + p^2 + \cdots + p^{n_2-n_1-1})$$

$$\sum_{i=1+n_{k-1}}^{n_k} D(i) = p^{n_{k-1}-k+1}q^{k-1}(1 + p + p^2 + \cdots + p^{n_k-n_{k-1}-1}).$$

Nota. En este caso no necesitamos considerar algún termino adicional en esta serie dada, con $B = 2^{n_k}$, $D(i) = 0$ para toda $i > n_k$.

La suma parcial anterior puede ser simplificada como:

$$\frac{1-p^{n_1}}{q} + \frac{1}{p}(p^{n_1} - p^{n_2}) + \frac{1}{p}\frac{q}{p}(p^{n_2} - p^{n_3}) + \cdots + \frac{1}{p}\frac{q^{k-2}}{p}(p^{n_{k-1}} - p^{n_k}).$$

Y combinando los términos iguales de potencias de p, ganancias de la fórmula (7). La fórmula (8) se sigue de que k tiende a infinito, y observando que, dado que $n_k \geq k$, la expresión final de (7) está acotada por pq^{k-2} y tiende a cero cuando k tiende al infinito.

Observaciones.

a) La probabilidad de que el juego continúe infinitamente es 0 dado que la probabilidad de que el juego continúe más allá de n partidas es a lo más $\frac{1}{2^n}$.

b) La probabilidad de que un apostador con una riqueza inicial A alcance una riqueza B $(A < B)$ es $\frac{A}{B}$ siempre que la apuesta y la estrategia en cada una de las partidas satisfagan:

   1. Existen $m, q \epsilon \mathbb{N}$ tales que, en cada juego, la probabilidad de perder por lo menos m es, por lo menos q, y
   2. El único final posible resulta ser 0 o B, y
   3. Cada juego es justo, es decir, $E[X_i] = 0$ para toda $i \geq 1$.

c) La $i-$ésima apuesta será la última (concluyendo con $S_i = 0$, con probabilidad q si $R_{i-1} \leq \frac{1}{2}$; y concluyendo con $S_i = B$, con probabilidad p, si $R_{i-1} \geq \frac{1}{2}$).

d) Habrá una $(i+1)-$ésima apuesta. En este caso $R_{i-1} \neq \frac{1}{2}$, y R_{i-1} tiene representación binaria $0.b_1b_2b_3$ ... , R_i será igual a cambiar la primera posición $0.b_2b_3b_4$ .... Esto se sigue dado que $R_{i-1} < \frac{1}{2}$ implica que $b_1 = 0$ y $2R_{i-1} = 0.b_2b_3b_4$ ... , mientras que $R_{i-1} > \frac{1}{2}$ implica que $b_1 = 1$ y $0.b_2b_3b_4$ ... $= 2R_{i-1} - 1$. Procediendo inductivamente, entonces, esto se sigue de que el único valor posible para R_i, es 0 o 1, teniendo una representación binaria igual a un cambio de la $i-$ésima posicion de la

representación binaria para R_0. Obviamente, si R_i es igual a 0 o 1, todas las $R_j, j > i$ subsecuentes tienen el mismo valor.

e) El hecho de que un incremento en las apuestas en el caso clásico incremente la probabilidad del suceso (cuando $p < \frac{1}{2}$) sugiere el teorema de aproximación del jugador ansioso (teorema 2).

Capítulo 2. PROCESOS DE DECISION DE MARKOV

Los procesos de decisión de Markov (véase [17]), son una clase de procesos estocásticos la cual trata básicamente el problema de encontrar una política óptima que maximice la recompensa esperada en el tiempo.

Estos procesos están estrechamente ligados a las suposiciones de que el sujeto siempre conocerá el estado en que se encuentra al momento de iniciar las acciones, y que la probabilidad de transición de un estado depende sólo del estado presente, también conocido como la propiedad de Markov (véase [A.1]).

En este capítulo consideraremos un proceso que se observa en puntos de tiempo discreto, es decir, considerando un espacio de estados numerable; una vez observado el estado del proceso, en otras palabras, su resultado en el tiempo t, se debe tomar una decisión, aunque técnicamente nos referiremos a esto último como elegir una acción, la cual condicionará el estado del proceso en la siguiente etapa.

Supongamos por un momento que este proceso fuera un modelo macroeconómico y nuestro interés, conocer la tasa de inflación dentro de un año. Entonces, es claro que existirán acciones apropiadas que podríamos llevar a cabo para mantener constante, incrementar o disminuir el valor de la variable en cuestión "la tasa de inflación", lo interesante sería poder elegir la deseada, más aún, saber cómo elegirla. A esta situación la llamaremos "elección de la política óptima". La importancia de tomar la decisión óptima debe ser clara, ya que de no hacerlo el efecto podría ser gravemente negativo, en un sentido que debemos precisar como contrario.

2.1 PROCESOS DE DECISIÓN DE MARKOV

Considérese un proceso que se observa en puntos de tiempo discreto para estar en cualquiera de los estados posibles $1, 2, \dots, M$. Después de observar el estado del proceso, una acción debe ser elegida.

Sea A el conjunto de todas las acciones posibles, finito.

Si el proceso está en el estado i en el tiempo n y la acción a es elegida, entonces el próximo estado del sistema es determinado por las probabilidades de transición $P_{ij}(a)$ (véase [A.1]).

Si tomamos X_n denotando el estado del proceso en el tiempo n y la acción a_n es elegida en el tiempo n, entonces lo anterior es equivalente a afirmar que,

$$P\{X_{n+1} = j|X_0, a_0, X_1, a_1, \dots, X_n = i, a_n = a\} = P_{ij}(a).$$

Así, las probabilidades de transición son funciones únicamente del estado presente y la acción subsecuente.

Definición. Política es una regla para la elección de acciones.

Nos restringiremos a políticas de la forma: la acción que predice para cada tiempo depende únicamente del estado del proceso en dicho tiempo, y permitiremos que la política sea aleatoria. En otras palabras, una política β es un conjunto de números $\beta = \{\beta_i(a) : a \epsilon A,\ i = 1, \dots, M\}$ con la siguiente interpretación: si el proceso está en el estado i, entonces, la acción a debe ser escogida con probabilidad $\beta_i(a)$.

Claramente, es necesario tener:

$0 \leq \beta_i(a) \leq 1$, para todas i, a y,

$\sum_a \beta_i(a) = 1$, para toda i.

Dada una política β, la sucesión de los estados $\{X_n, n = 0, 1, \dots\}$ constituye una cadena de Markov con probabilidades de transición $P_{ij}(\beta)$ dada por:

$$P_{ij}(\beta) = P\{[X_{n+1} = j|X_n = i]|\beta\} = \sum_a P_{ij}(a)\beta_i(a).$$

Donde la última igualdad, se sigue de la elección de una acción condicionada en el estado i. Supongamos que para cada elección de una política β, la cadena de Markov $\{X_n : n = 0, 1, \dots\}$ resultante es ergódica (véase [A.1.1]).

Para cualquier política β, sea π_{ia} el límite de la probabilidad (o estado estacionario (véase [A.1])) de que el proceso se encuentre en el estado i y la acción a sea elegida si la política β es empleada, esto es,

$$\pi_{ia} = \lim_{n\to\infty} P\{[X_n = i, a_n = a\,]|\beta\}.$$

El vector $\pi = (\pi_{ia})$ debe satisfacer:

(i) $\pi_{ia} \geq 0$ para todo i, a,
(ii) $\sum_i \sum_a \pi_{ia} = 1$ y,
(iii) $\sum_a \pi_{ja} = \sum_i \sum_a \pi_{ia} P_{ij}(a)$ para toda j.

Las ecuaciones (i) y (ii) son obvias, y la ecuación (iii) es análoga a la ecuación:

$\pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij}$, $j \geq 0$,

dado que el lado izquierdo de la igualdad es la probabilidad del estado estacionario en el estado j, y el lado derecho es la misma probabilidad calculada por el condicionamiento en el estado y la acción elegida en la etapa anterior. La justificación es el siguiente teorema:

Teorema. Para una cadena de Markov ergódica e irreducible, $\lim_{n\to\infty} P_{ij}^n$ existe y es independiente del estado i. Más aún, $\pi_j = \lim_{n\to\infty} P_{ij}^n$, $j \geq 0$. Entonces, π_j es la única solución no negativa de $\pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij}$, $j \geq 0$, donde $\sum_{j=0}^{\infty} \pi_j = 1$.

Así, para alguna política β existe un vector $\pi = (\pi_{ia})$ que satisface (i), (ii) y (iii) y la interpretación es que π_{ia} es igual a la probabilidad del estado estacionario en el estado i, eligiendo la acción a, para $\beta_i \epsilon \beta$. Más aún, resulta que el reciproco también es cierto. Es decir, para algún vector $\pi = (\pi_{ia})$ que satisfaga (i), (ii) y (iii), existe una política β tal que si β es usada, entonces la probabilidad del estado estacionario en el estado i eligiendo la acción a, es igual a π_{ia}. Para verificar este último argumento, supongamos que $\pi = (\pi_{ia})$ es un vector que satisface (i), (ii) y (iii). Entonces, sea la política $\beta = (\beta_i(a))$,

$$\beta_i(a) = P\{\beta \text{ elija } a | \text{estado } i\} = \frac{\pi_{ia}}{\sum_a \pi_{ia}}.$$

Ahora, sea P_{ia} el límite de la probabilidad de estar en el estado i, eligiendo a con la política β. Necesitamos mostrar que $P_{ia} = \pi_{ia}$. Para lo cual, primero notemos que $\{P_{ia}: i = 1, \dots, M , a \epsilon A\}$ son los límites de las probabilidades de la cadena de Markov bidimensional $\{(X_n , a_n), n \geq 0\}$. Entonces, por el teorema anterior, existe una única solución de:

(i') $P_{ia} \geq 0$,
(ii') $\sum_i \sum_a P_{ia} = 1$ y,
(iii') $P_{ja} = \sum_i \sum_{a'} P_{ia'} P_{ij}(a') \beta_j(a)$,

donde, de (iii'), se sigue que,

$$P\{X_{n+1} = j, a_{n+1} = a | X_n = i, a_n = a'\} = P_{ij}(a') \beta_j(a).$$

Entonces,

$$\beta_j(a) = \frac{\pi_{ja}}{\sum_a \pi_{ja}}.$$

Vemos que (P_{ia}) es la única solución de:

$P_{ia} \geq 0$,
$\sum_i \sum_a P_{ia} = 1$ y,
$P_{ja} = \sum_i \sum_{a'} P_{ia'} P_{ij}(a') \frac{\pi_{ja}}{\sum_a \pi_{ja}}$.

Así, para mostrar que $P_{ia} = \pi_{ia}$, necesitamos mostrar que,

$\pi_{ia} \geq 0$,
$\sum_i \sum_a \pi_{ia} = 1$ y,
$\pi_{ja} = \sum_i \sum_{a'} \pi_{ia'} P_{ij}(a') \frac{\pi_{ja}}{\sum_a \pi_{ja}}$.

De los dos primeros se siguen (i) y (ii) y ya que la tercera es equivalente a:

$\sum_a \pi_{ja} = \sum_i \sum_{a'} \pi_{ia'} P_{ij}(a')$,

se sigue (iii).

De esta forma mostramos que el vector $\pi = (\pi_{ia})$ satisface (i), (ii) y (iii) sí, y sólo si, existe una política β tal que π_{ia} es igual a la probabilidad del estado estacionario, en el estado i eligiendo la acción a. De hecho, la política β está definida por:

$$\beta_i(a) = \frac{\pi_{ia}}{\sum_a \pi_{ia}}.$$

El procedimiento es muy importante en la determinación de la política óptima. Por ejemplo, supongamos que una recompensa $R(i, a)$ es obtenida sin importar la acción escogida a en el estado i. Entonces $R(X_i, a_i)$ representaría la recompensa obtenida en el tiempo i, el promedio de la recompensa esperada por unidad de tiempo bajo la política β puede ser expresada como sigue,

$$\text{Promedio de la recompensa esperada bajo } \beta = \lim_{n\to\infty} E_\beta \left[\frac{\sum_{i=1}^{n} R(X_i, a_i)}{n}\right].$$

Ahora, si π_{ia} denota la probabilidad del estado estacionario en el estado i eligiendo la acción a, se sigue que el límite de la esperanza total en el tiempo n es igual a:

$$\lim_{n\to\infty} E[R(X_i, a_i)] = \sum_i \sum_a \pi_{ia} R(i, a).$$

Lo cual implica que,

$$\text{Promedio de la recompensa esperada bajo } \beta = \sum_i \sum_a \pi_{ia} R(i,a).$$

Entonces, el problema de determinar la política que maximiza el promedio total esperado es:

$$\max_{\pi=(\pi_{ia})} \sum_i \sum_a \pi_{ia} R(i,a)$$

$$\text{sujeto a } \begin{cases} \pi_{ia} \geq 0 \quad \text{para todas } i,a \\ \sum_i \sum_a \pi_{ia} = 1 \\ \sum_a \pi_{ja} = \sum_i \sum_a \pi_{ia} P_{ij}(a) \quad \text{para toda } j. \end{cases}$$

Observaciones.

a) Sin embargo, este problema de maximización es un caso especial de lo que es conocido como programación lineal y puede ser resuelto con un algoritmo de programación lineal estándar conocido como método simplex.

b) Si $\pi^* = (\pi_{ia}^*)$ maximiza al problema, entonces la política óptima estará dada por β^*, donde:

$$\beta_i^*(a) = \frac{\pi_{ia}^*}{\sum_a \pi_{ia}^*}.$$

2.2. MAXIMIZACIÓN DE RECOMPENSAS PROGRAMACIÓN DINÁMICA POSITIVA

En esta sección consideraremos modelos en los cuales nos interesa maximizar los rendimientos esperados. En particular, asumiremos un espacio de estados contable y un espacio de acciones finito, y supondremos que si una acción a se toma estando en el estado i, entonces una recompensa esperada $R(i,a)$, que se supone no negativa, será ganada.

Para una política π, sea

$$V_\pi(i) = E_\pi[\sum_{t=0}^{\infty} R(X_t, a_t) | X_0 = i], \quad i \geq 0,$$

por lo tanto, $V_\pi(i)$ es la ganancia total esperada bajo π cuando $X_0 = i$. Debido a que $R(i,a) \geq 0$, $V_\pi(i)$ está bien definida, aunque podría ser infinita.

También sea:

$$V(i) = \sup_{\pi} V_{\pi}(i).$$

Una política π^* se dice que es óptima si,

$$V_{\pi^*}(i) = V(i) \quad \text{para todo } i \geq 0.$$

Teorema 1. La Ecuación de Optimalidad:

$$V(i) = \max_a \left[R(i,a) + \sum_{j=0}^{\infty} P_{ij}(a)V(j)\right], \quad i \geq 0.$$

Por desgracia, no resulta que la política determinada por la ecuación funcional de este teorema sea una política óptima. De hecho, resulta que una política óptima no tiene por qué existir. Considere lo siguiente:

Ejemplo para el cual no existe una política óptima. Supongamos que existen dos acciones y las probabilidades de transición están dadas por:

$$P_{00}(1) = P_{00}(2) = 1$$

$$P_{ii+1}(1) = 1, \quad i > 0$$

$$P_{i0}(2) = 1, \quad i > 0.$$

Las recompensas están dadas por:

$$R(0,1) = R(0,2) = 0$$

$$R(i,1) = 0, \quad i > 0$$

$$R(i,2) = 1 - \frac{1}{i}, \quad i > 0.$$

En otras palabras, en el estado i, no se puede obtener una recompensa inmediata e ir al estado $(i+1)$, o si obtenemos una recompensa inmediata de $\left(1 - \frac{1}{i}\right)$, no tendremos recompensa en el futuro.

Es fácil de ver que, mientras $V(i) = 1$, $V_{\pi}(i) < 1$ para toda política π y toda i. Entonces, no existe una política óptima. De hecho, la ecuación de optimalidad está dada por:

$$V(i) = \max\left[V(i+1), 1 - \frac{1}{i}\right],$$

y ya que $V(i) \equiv 1$, se sigue que la política determinada por la ecuación de optimalidad es la que siempre elige la acción 1, y por tanto el rendimiento esperado es 0.

Entonces, no es necesario que exista una política óptima. Sin embargo, una cosa que se puede y se debe demostrar es que, si la función de rendimiento para una política dada satisface la ecuación de optimalidad, entonces esta política es óptima.

Teorema 2. Sea V_f la función de rendimiento esperado para la política estacionaria f. Si V_f satisface la ecuación de optimalidad (teorema 1), entonces f es óptima. Esto es, si

$$V_f(i) = \max_a \left[R(i,a) + \sum_j P_{ij}(a) V_f(j)\right], \quad i \geq 0, \qquad (*)$$

entonces:

$$V_f(i) = V(i) \quad \text{para toda } i.$$

Demostración.

De la hipótesis se tiene que,

$$V_f(i) \geq R(i,a) + \sum_j P_{ij}(a) V_f(j) \quad \text{para toda } a.$$

Nota: Por supuesto, la igualdad se alcanza cuando $a = f(i)$.

Ya que el lado derecho de $(*)$ es el rendimiento esperado si inicialmente tomamos la acción a y luego seguimos con f, podemos interpretar $(*)$ como: usar f es mejor que hacer cualquier cosa en la primera etapa y luego cambiar a f. Pero si hacemos cualquier cosa en la primera etapa, entonces dependiendo en qué punto cambiemos a f podría ser mejor que hacer cualquier cosa en otro periodo y luego cambiar a f. Entonces, vemos que usar f es mejor que hacer cualquier cosa en dos etapas y luego cambiar a f. Repitiendo este argumento, mostramos que usando f, es mejor que hacer cualquier cosa en n-etapas y luego cambiar a f.

Esto es, para una política π,

$$V_f(i) \geq E_\pi\left[\sum_{t=0}^{n-1} R(X_t, a_t) | X_0 = i\right] + E_\pi\left[V_f(X_n)\right].$$

Sin embargo, ya que toda recompensa $R(i,a)$ es no negativa, se sigue que $V_f(i) \geq 0$, así, obtenemos que,

$$V_f(i) \geq E_\pi\left[\sum_{t=0}^{n-1} R(X_t, a_t) | X_0 = i\right].$$

Haciendo $n \to \infty$, tenemos que:

$$V_f(i) \geq V_\pi(i),$$

lo cual prueba el resultado.

Observaciones.

a) La prueba del teorema 2 muestra que no necesariamente se requiere que $R(i,a) \geq 0$ para que el teorema sea válido. Más débilmente, una condición suficiente podría ser que, $V_f(i) \geq 0$ para toda i. De hecho, una condición suficiente aún más débil podría ser que $\lim_n \inf E_n\left[V_f(X_n)|X_0 = i\right] \geq 0$ para toda i y toda π.

b) Se sigue de los teoremas 1 y 2 que una política estacionaria f es óptima sí y sólo si, la función de rendimiento esperado satisface la ecuación de optimalidad, esto es, si para cada estado inicial, usar f es mejor que hacer cualquier cosa para esta etapa y luego cambiar a f. Esto nos proporciona un método para comprobar si una determinada política puede ser óptima o no, y es particularmente útil en los casos en los cuales existe una política óptima obvia.

c) Ya que asumimos que todas las recompensas son no negativas, se dice que, problemas que se ajustan a este marco son del tipo de programación dinámica positiva.

Capítulo 3. APLICACIONES DE PROCESOS DE DECISIÓN DE MARKOV EN JUEGOS DE APUESTAS

El hombre apostador, se configuró en la prehistoria, como un ser humano que buscaba un placer por arriesgar alguna propiedad apostándola en un juego o en una práctica física. Una de las tantas áreas donde pueden aplicarse los procesos de decisión de Markov es en Juegos de Apuestas.

Con el fin de plantar una buena idea, imaginemos un individuo que llega a un casino con la firme intención de salir victorioso de él, es decir, haber ganado luego de apostar ciertas cantidades de dinero durante algunas partidas de un juego.

Hay algo sumamente importante que este jugador debe saber, esto es, la probabilidad con la que gozaría ganado al apostar su dinero en una partida, o en su defecto, con la que padecería perdiendo. Dependiendo el valor de esta probabilidad y su complementaria, el apostador debe elegir cuanto de su dinero tirara por apuesta con el fin de llevarse a casa una cantidad mayor que el fijó previamente. Existen dos posibilidades:

- Una es, apostar 1 en cada partida hasta alcanzar su meta.
- La otra, apostar todo o la parte de ese todo que le permita alcanzar su meta.

En ambos casos existe el riesgo de perderlo todo si no sabe donde detenerse, por lo cual en este capítulo, trataremos la forma de elegir la estrategia óptima de juego, y no menos importante el momento en el que debe parar para no arruinarse.

3.1. JUEGOS DE APUESTAS CON PROCESOS DE DECISIÓN DE MARKOV

Considere la siguiente situación. Un individuo que posee i pesos apuesta en un casino. El apostador del casino permite algunas apuestas de la siguiente manera: si se poseen i pesos entonces se tiene permitido apostar una cantidad menor o igual a i pesos. Además, si se apuesta j, entonces puede suceder que:

a) Se gana j con probabilidad p, o
b) Se pierde j con probabilidad $1-p$.

¿Qué estrategia de juego maximiza la probabilidad de que el individuo alcance una fortuna N antes de arruinarse?

Tenemos que mostrar que si $p \geq \frac{1}{2}$ (es decir, si se está jugando un juego favorable), entonces la estrategia óptima es la estrategia tímida que apuesta siempre 1 hasta alcanzar o N o 0. Sin embargo, si $p \leq \frac{1}{2}$, entonces la estrategia audaz, la cual apuesta siempre la fortuna actual o esa parte de ella que le permita llegar a N si ha ganado, es la óptima.

¿Este modelo se ajusta al marco del capítulo anterior? Para mostrar esto, sea el espacio de estados el conjunto $\{0,1,\dots,N\}$ y decimos que el estado i es donde la fortuna es i. Ahora definimos la estructura de la recompensa como:

$$R(i,a) = 0, \quad i \neq N, \quad \text{para toda } a$$

$$R(N,a) = 1$$

$$P_{N0}(a) = P_{00}(a) = 1.$$

En otras palabras, una recompensa de \$1 es ganada sí, y sólo si, la fortuna actual nunca es N y así, la recompensa total esperada es justamente la probabilidad de que la fortuna actual nunca sea N. Por lo tanto, este problema no necesariamente tiene que ajustarse al marco del capítulo anterior.

Para determinar una política óptima, primero notemos que, si la fortuna actual es i, entonces nunca se tendría que pagar para apostar más que $N-i$. Esto es, en el estado i, se puede limitar la elección de apuestas a $1,2,\dots,\min\{i, N-i\}$.

Sabemos por el teorema 2 de la sección 1.2 que, si $U(i)$ la recompensa de alguna política estacionaria, satisface:

$$U(i) \geq pU(i+k) + (1-p)U(i-k) \quad \text{para todo } 0 < i < N \text{ y } k \leq \min\{i, N-1\}$$

entonces, esta política es óptima.

La estrategia tímida

Definición. La estrategia tímida es la estrategia que siempre apuesta1.

En el fondo, esta estrategia transforma al juego en la ruina del apostador clásico o modelo de caminata aleatoria, y $U(i)$, la probabilidad de alcanzar N antes que arruinarse cuando se empieza con i, $i < N$, satisface:

$$U(i) = \begin{cases} \dfrac{1-\left(\frac{q}{p}\right)^{i}}{1-\left(\frac{q}{p}\right)^{N}} & \text{si } \ p \neq \dfrac{1}{2} \\ \dfrac{i}{N} & \text{si } \ p = \dfrac{1}{2}\,. \end{cases}$$

Proposición 1. Si $p \geq \frac{1}{2}$, la estrategia tímida maximiza la probabilidad de nunca alcanzar una fortuna de N.

Prueba.

Si $p = \frac{1}{2}$, entonces $U(i) = \frac{i}{N}$ trivialmente satisface la desigualdad anterior. Cuando $p > \frac{1}{2}$ se puede mostrar que,

$$\frac{1-\left(\frac{q}{p}\right)^{i}}{1-\left(\frac{q}{p}\right)^{N}} \geq p\left[\frac{1-\left(\frac{q}{p}\right)^{i+k}}{1-\left(\frac{q}{p}\right)^{N}}\right] + q\left[\frac{1-\left(\frac{q}{p}\right)^{i-k}}{1-\left(\frac{q}{p}\right)^{N}}\right]$$

o

$$\left(\frac{q}{p}\right)^{i} \leq p\left(\frac{q}{p}\right)^{i+k} + q\left(\frac{q}{p}\right)^{i-k}$$

o

$$1 \leq p\left(\frac{q}{p}\right)^{k} + q\left(\frac{p}{q}\right)^{k}$$

o equivalentemente

$$1 \leq p\left[\left(\frac{q}{p}\right)^{k} + \left(\frac{p}{q}\right)^{k-1}\right].$$

Note que esto se mantiene si $k = 1$, así este resultado quedará probado si podemos mostrar que $\left(\frac{q}{p}\right)^{k} + \left(\frac{p}{q}\right)^{k-1}$ es una función creciente de k, lo cual se probará en el siguiente lema.

Lema. $f(x) \equiv \left(\frac{1-p}{p}\right)^x + \left(\frac{p}{1-p}\right)^{x-1}$ es creciente para $x \geq 1$ cuando $p > \frac{1}{2}$.

Prueba.

$$f'(x) = \left(\frac{1-p}{p}\right)^x log\left(\frac{1-p}{p}\right) + \left(\frac{p}{1-p}\right)^{x-1} log\left(\frac{p}{1-p}\right)$$

$$= \left[\left(\frac{p}{1-p}\right)^{x-1} - \left(\frac{1-p}{p}\right)^x\right] log\left(\frac{p}{1-p}\right) \geq 0.$$

Esto prueba el lema y también completa la prueba de la proposición anterior.

Así, cuando $p > \frac{1}{2}$, esto es, cuando el juego es favorable para el jugador, entonces continuaría con la mínima apuesta hasta que, o alcanzara la meta o quebrara. Antes vimos el caso en el que $p < \frac{1}{2}$, ahora supongamos que el objetivo no es alcanzar alguna meta preasignada, sino maximizar el tiempo de juego. Ahora mostraremos que si $p \geq \frac{1}{2}$, entonces el juego tímido estocásticamente maximiza el tiempo de juego. Esto es, para cada n, la probabilidad de ser capaz de jugar n o más veces antes de quebrar es maximizada por la estrategia tímida.

Proposición 2. Si $p \geq \frac{1}{2}$, entonces el juego tímido estocásticamente maximiza el tiempo de juego.

Prueba.

Asumiendo que una recompensa de 1 es obtenida si se es capaz de jugar por lo menos n veces, vemos que este problema también encaja en el marco del caso positivo. Entonces, debemos mostrar que, es mejor jugar tímidamente que hacer una apuesta inicial de k, $k \leq i$, y entonces jugar tímidamente. Sin embargo se sigue por la proposición anterior que, la estrategia tímida maximiza la probabilidad de alcanzar $(i + k)$ antes que $(i - k)$, y tomar por lo menos una unidad de tiempo. Más formalmente, siendo $U_n(i)$ la probabilidad de ser capaz de jugar por lo menos n veces, dado que la fortuna inicial es i y se juegue tímidamente, se obtiene, condicionando que en el tiempo T la fortuna alcanzada sea o $(i - k)$ o $(i + k)$ y el valor X alcanzado, que:

$$U_n(i) = \mathrm{E}[U_{n-T}(X)]$$

$$\geq \mathrm{E}[U_{n-1}(X)]$$

$$= U_{n-1}(i+k)P(X=i+k) + U_{n-1}(i-k)P(X=i-k)$$

$$\geq pU_{n-1}(i+k) + qU_{n-1}(i-k).$$

La primera desigualdad se sigue del hecho de que $U_n(i)$ es una función decreciente de n y $T \geq 1$, mientras que la segunda desigualdad se sigue de que $P\{X = i+k\} \geq p$ por la primera proposición, y:

$$U_{n-1}(i+k) \geq U_{n-1}(i-k).$$

■

Resulta que si $p < \frac{1}{2}$, entonces la estrategia tímida estocásticamente no maximiza el tiempo de juego. Supongamos $p = 0.1$ y que se empieza con una fortuna inicial de 2. La probabilidad de ser capaz de jugar por lo menos 5 juegos si se juega tímidamente es $1 - (0.9)^2 - 2(0.9)^3(0.1) = 0.0442$. Por otro lado, si se apuesta inicialmente 2 y se juega tímidamente, entonces la probabilidad de jugar por lo menos 5 juegos es 0.1. Sin embargo, es cierto que el juego tímido maximiza el tiempo de juego esperado.

Proposición 3. Si $p < \frac{1}{2}$, entonces el juego tímido maximiza el tiempo esperado de juego.

Prueba.

Sea $U(i)$ el número esperado de apuestas hechas antes de arruinarse, dado que se empieza con i y siempre se apuesta 1. Para calcular $U(i)$, sea X_j la ganancia de la $j-ésima$ apuesta y T el número de apuestas antes de quebrar. Entonces, ya que:

$$\sum_{j=1}^{T} X_j \equiv -i,$$

tenemos por la ecuación de Wald que:

$$-i = E[X]E[T]$$

o

$$U(i) = E[T] = \frac{i}{1-2p}.$$

Ya que maximizando el tiempo esperado de juego fallamos bajo el caso positivo (recibimos una recompensa de 1 cada vez que se es capaz de continuar jugando), el resultado se sigue ya que la desigualdad:

$$U(i) \geq 1 + pU(i+k) + qU(i-k), \;\; 1 \leq k \leq i$$

es equivalente a:

$$\frac{i}{1-2p} \geq 1 + p\frac{i+k}{1-2p} + (1-p)\frac{i-k}{1-2p}$$

o

$$0 \geq 1 - 2p + pk - (1-p)k$$

o

$$k(1-2p) \geq 1 - 2p.$$

Lo cual se sigue de que $k \geq 1$.

■

Así, cuando $p < \frac{1}{2}$, la estrategia tímida es óptima cuando el objetivo es maximizar la cantidad esperada de veces que se es capaz de jugar antes de quebrar.

No obstante, ahora debemos mostrar que si el objetivo es alcanzar una fortuna N, entonces la estrategia óptima es la audaz. De hecho, mostraremos que este es el caso incluso donde uno es el número de apuestas permitidas.

La estrategia audaz

Definición. La estrategia audaz es la estrategia en la que, si la fortuna actual es i,

a) Apuesta i, si $i \leq \frac{N}{2}$.
b) Apuesta $N - i$, si $i \geq \frac{N}{2}$.

Sea $U_n(i)$ la probabilidad de nunca alcanzar N habiendo empezado con i, permitiendo a lo más n apuestas, usando la estrategia audaz y condicionando la salida de la primera apuesta como sigue:

$$U_n(i) = \begin{cases} pU_{n-1}(2i) & \text{si } i \leq \frac{N}{2} \\ p + qU_{n-1}(2i - N) & \text{si } i \geq \frac{N}{2}. \end{cases} \qquad (*)$$

Con condiciones de frontera $U_n(0) = 0$, $U_n(N) = 1$, $n \geq 0$, $U_0(i) = 0$, $i < N$.

Ahora estamos listos para la próxima proposición.

Proposición 4. Cuando $p \leq \frac{1}{2}$, para cada $n > 0$, la estrategia audaz maximiza la probabilidad de alcanzar una fortuna N en el tiempo n.

Prueba.

Por el teorema 2 de la sección 2.1 es suficiente probar que:

$$U_{n+1}(r) \geq pU_n(r + s) - qU_n(r - s), \;\; s \leq \min\{r, N - r\},$$

o equivalentemente que:

$$U_{n+1}(r) - pU_n(r + s) - \; qU_n(r - s) \geq 0, \;\; s \leq \min\{r, N - r\}.$$

Debemos verificar la desigualdad por inducción sobre n, y ya que esto es inmediato para $n = 0$, asumiremos que:

$$U_n(i) - \; pU_{n-1}(i + k) - \; qU_{n-1}(i - k) \geq 0, \;\; k \leq \min\{i, N - i\} \quad (H.I.).$$

Para establecer la segunda desigualdad, existen cuatro casos que debemos considerar.

Caso 1. $r + s \leq \frac{N}{2}$. En este caso, tenemos por $(*)$ que:

$$\begin{aligned} U_{n+1}(r) - pU_n(r + s) - \; qU_n(r - s) \\ = p[U_n(2r) - pU_{n-1}(2r + 2s) - \; qU_{n-1}(2r - 2s)]. \end{aligned}$$

Caso 2. $r - s \geq \frac{N}{2}$. La prueba es justamente como en el caso 1, excepto que se usa la segunda ecuación de $(*)$ en lugar de la primera.

Caso 3. $r \leq \frac{N}{2} \leq r + s, \;\; s \leq r$. En este caso, por $(*)$ tenemos que:

$$\begin{aligned} U_{n+1}(r) - pU_n(r + s) - \; qU_n(r - s) \\ = p[U_n(2r) - p - \; qU_{n-1}(2r + 2s - N) - qU_{n-1}(2r - 2s)]. \end{aligned}$$

Ahora, $2r \geq r + s \geq \frac{N}{2}$, por lo cual, la desigualdad anterior:

$$= p[p + qU_{n-1}(4r - N) - p - qU_{n-1}(2r + 2s - N) - qU_{n-1}(2r - 2s)]$$

$$= q[pU_{n-1}(4r - N) - pU_{n-1}(2r + 2s - N) - pU_{n-1}(2r - 2s)]$$

$$= q\left[U_n\left(2r - \frac{N}{2}\right) - pU_{n-1}(2r + 2s - N) - pU_{n-1}(2r - 2s)\right].$$

Donde la última igualdad se sigue de $(*)$ ya que $0 \leq 2r - \frac{N}{2} \leq \frac{N}{2}$.

Ahora, si $s \geq \frac{N}{4}$, entonces, ya que $p \leq q$, tenemos que la última igualdad es al menos

$$q\left[U_n\left(2r - \frac{N}{2}\right) - pU_{n-1}(2r + 2s - N) - qU_{n-1}(2r - 2s)\right].$$

Sin embargo, esta última expresión es no negativa por la hipótesis de inducción para $i = 2r - \frac{N}{2}$ y $k = 2s - \frac{N}{2}$.

Por otro lado, si $s < \frac{N}{4}$, entonces, ya que $p \leq q$, la última igualdad es al menos

$$q\left[U_n\left(2r - \frac{N}{2}\right) - qU_{n-1}(2r + 2s - N) - pU_{n-1}(2r - 2s)\right].$$

La cual es no negativa por la hipótesis de inducción para $i = 2r - \frac{N}{2}$ y $k = \frac{N}{2} - 2s$.

Caso 4. $r - s \leq \frac{N}{2} \leq r$. La prueba de este caso es análoga a la del anterior.

Corolario. Con tiempo ilimitado, el juego audaz maximiza la probabilidad de nunca alcanzar N cuando $p \leq \frac{1}{2}$.

Prueba.

Si denotamos a $U(r)$ la probabilidad de alcanzar N antes que 0, empezando con r y usando la estrategia audaz, entonces:

$$U(r) = \lim_{n\to\infty} U_n(r).$$

Y por la desigualdad siguiente a $(*)$ se tiene que,

$$U(r) \geq pU(r + s) + qU(r - s), \;\; s \leq \min\{r, N - r\}.$$

Entonces el retorno del juego audaz satisface la ecuación de optimalidad, así, por el teorema 2 de la sección 2.1, es óptimo.

Observaciones.

a) La estrategia tímida maximiza tanto el tiempo de juego como la probabilidad de que un jugador nunca alcance una fortuna N si $p \geq \frac{1}{2}$.

b) La función $f(x) = \left(\frac{1-p}{p}\right)^x + \left(\frac{p}{1-p}\right)^{x-1}$ es creciente cuando $p > \frac{1}{2}$ para toda $x \geq 1$.

c) La estrategia tímida maximiza el tiempo esperado de juego cuando $p < \frac{1}{2}$.

d) La estrategia audaz maximiza la probabilidad de que un jugador alcance una fortuna N en el tiempo n, pero también maximiza la probabilidad de que nunca la alcance si el tiempo es limitado, para toda n cuando $p \leq \frac{1}{2}$.

3.2. UN MODELO DE UN JUEGO DE APUESTAS CONTRA UN CASINO

- Tenemos una persona con cierta riqueza inicial de $\boldsymbol{i}$ pesos;
- Política del Casino:

Si se tienen $\boldsymbol{i}$ pesos se permite apostar cualquier entero positivo menor ó igual a i; además si se apuestan $\boldsymbol{a}$ pesos entonces:

a) se gana $\boldsymbol{a}$ con probabilidad p.

b) se pierde $\boldsymbol{a}$ con probabilidad $1 - p$.

Pregunta: ¿Cuál es la estrategia de juego que maximiza la probabilidad de que la persona gane una riqueza de N pesos (N fijo) antes de que se arruine?

Este juego se modela como un PDM de la siguiente manera:

- $S = \{0,1, \ldots, N\}$.
- Si la riqueza presente del jugador es i pesos, él nunca pagará más de $(N - i)$ pesos, (observe que $i + a \leq N \Rightarrow a \leq N$) y se limitará su selección de apuestas a: $1, 2, \ldots, \min\{i, N - i\}$.

Entonces:

$$A = \left\{0, 1, \ldots, \left[\frac{N}{2}\right]\right\}$$

$$A(0) = \{0\}$$

$$A(i) = \{1, 2, \dots, \min\{i, N-i\}\}, i \in S$$

$$p_{ii+a}(a) = p, \qquad p_{ii-a}(a) = q = 1-p$$

$$p_{N0}(a) = 1, \qquad p_{00}(0) = 1$$

$$R(i, a) = 0, \qquad i \neq N, \qquad \text{para toda } a$$

$$R(N, 0) = 1.$$

Nota: Cuando $a = 1$, se obtiene la caminata aleatoria con un estado absorbente.

Entonces, una recompensa de 1 peso es alcanzada sí, y sólo si, su riqueza presente alcanza N, y por tanto la recompensa total esperada es, justamente la probabilidad de que la riqueza alcance N en algún tiempo, sin pasar por 0.

Dada una estrategia f, se puede verificar que, para cada $k = 0, 1, \dots,$

$$R(x_0, a_0) + R(x_1, a_1) + \cdots + R(x_2, a_2) + \cdots + R(x_N, a_N)$$
$$= Ind\,([x_0 = N] \cup [x_0 \neq N, x_1 = N] \cup \dots \cup$$
$$\cup [x_0 \neq N, x_1 \neq N, \dots, x_{k-1} \neq N, x_k = N])\ \text{c.s.}$$

Entonces, si $k \to \infty$:

$$\sum_t R(x_t, a_t) = Ind([x_0 = N] \cup [x_0 \neq N, x_1 = N] \cup \cdots)\ \text{c.s.}$$

Por tanto, para cada $i = 0, 1, \dots, N$:

$$V(f, i) = E^f[\sum_t R(x_t, a_t) | x_0 = i] = P^f[([x_0 = N] \cup$$
$$\cup [x_0 \neq N, x_1 = N] \cup \cdots) | x_0 = i].$$

Solución para $p > \frac{1}{2}$.

Considérese la estrategia $\tau(i) = 1,\ i \neq 0,$ y $\tau(0) = 0$

(τ es la estrategia **tímida**).

Entonces,

$$V(\tau, i) = \frac{1 - \left(\frac{q}{p}\right)^i}{1 - \left(\frac{q}{p}\right)^N}, i \in S.$$

- Se puede demostrar, usando el hecho de que la función h: $[1, \infty) \to \mathbb{R}$ dada por:

$$h(x) = \left(\frac{q}{p}\right)^x + \left(\frac{p}{q}\right)^{x-1}$$

es creciente cuando $p > \frac{1}{2}$ que,

$$V(\tau, i) \geq pV(\tau, i+a) + qV(\tau, i-a), \qquad \text{para todas } a, i$$

$\Leftrightarrow \quad V(\tau, i) \geq R(i, a) + \sum_j V(\tau, j) p_{ij}(a)$, para todas a, i

$\Downarrow$ (Programación Dinámica)

τ es óptima.

Observaciones.

a) Para $p = \frac{1}{2}$, también la estrategia óptima es τ, y

$$V(\tau, i) = \frac{i}{N}, i \epsilon S.$$

b) Para $p < \frac{1}{2}$, la estrategia óptima está dada por:

$$\begin{cases} \alpha(i) = i, & \text{si } i \leq \frac{N}{2} \\ \alpha(i) = N - i, & \text{si } i \geq \frac{N}{2} \end{cases}$$

(α es la estrategia **audaz**).

Nota: en este caso no existe una fórmula explícita para $V(\alpha, i),\ i \epsilon S$ (véase [17]).

Capítulo 4. CONCLUSIONES

Las Cadenas de Markov pueden ser consideradas como una de las grandes aportaciones de las matemáticas. No necesariamente hay que ser experto en la materia para entender en qué consisten y el uso que se les puede dar en diferentes situaciones de la vida diaria.

Por ejemplo, en las sociedades primitivas ya se realizaban apuestas y podría "*apostar*" que aquellas personas si bien no conocían las Cadenas de Markov y por supuesto los Procesos de Decisión de Markov, seguramente se dieron cuenta cuál era la mejor forma de apostar, evidentemente dicha forma de aprenderlo fue empírica.

Una derivación de las cadenas de Markov son los Procesos de Decisión de Markov, los cuales tratan en principio el problema de encontrar una política óptima que maximice la recompensa esperada a lo largo del tiempo.

Se ha demostrado que tanto las Cadenas como los Procesos de Decisión de Markov tienen un gran valor en diversas áreas, particularmente en este trabajo tratamos como intervienen en los *juegos de apuestas en un casino* ya que nos ayudan a predecir cuál es la estrategia óptima con la que un jugador debe actuar para satisfacer sus ganancias así como la duración del juego, es decir, cuántas partidas jugará y cuánto apostará en cada una de ellas.

Una vez que se ha hecho un tanto teórico el conocimiento sobre los Procesos de Decisión de Markov, no debe ser tan difícil aplicar los modelos a otras circunstancias tal vez más comunes como por ejemplo:

Cuántas horas de estudio debería invertir un estudiante diariamente para obtener una calificación. Podríamos suponer que un estudiante x posee una cantidad de i horas disponibles diariamente para estudiar, n la cantidad de días que lo hará y la calificación estaría en un rango de 0 a 1.

Así, se podría decir que todos deberíamos tener un mínimo conocimiento acerca de nuestros anfitriones en este trabajo, las cadenas y los procesos de decisión de Markov para tener un mejor desempeño en la vida diaria.

A continuación se presentarán algunos resultados medulares que fueron tratados a lo largo de este trabajo, pretendiendo seguir la cronología del mismo.

Procesos de Decisión de Markov

Sean un espacio de estados S y un conjunto de acciones A:

$\pi = \pi_{ia}$ satisface:

i. $\pi_{ia} \geq 0$ para todo i, a.

ii. $\sum_i \sum_a \pi_{ia} = 1$.

iii. $\sum_a \pi_{ja} = \sum_i \sum_a \pi_{ia} P_{ij}(a)$ para toda j.

Sí, y sólo si, existe una política β tal que π_{ia} sea igual a la probabilidad de que el proceso se encuentre en el estado estacionario en $i \epsilon S$ y $a \epsilon A$ sea elegida dado que β es usada.

Además, si $R(X_i, a_i)$ es la recompensa obtenida en el tiempo i, en el límite la esperanza total en el tiempo n es el promedio de la recompensa esperada bajo β:

$\beta = \sum_i \sum_a \pi_{ia} R(i, a)$.

De donde, el problema de maximización es:

$\max_{\pi=(\pi_{ia})} \sum_i \sum_a \pi_{ia} R(i, a)$

$$\text{sujeto a } \begin{cases} \pi_{ia} \geq 0 \quad \text{para todas } i, a \\ \sum_i \sum_a \pi_{ia} = 1 \\ \sum_a \pi_{ja} = \sum_i \sum_a \pi_{ia} P_{ij}(a) \quad \text{para toda } j. \end{cases}$$

El cual se resuelve como un problema de programación lineal.

La prueba del teorema 2 del capítulo 1 muestra que no necesariamente se requiere que $R(i, a) \geq 0$ para que el teorema sea válido.

Si V_f es la función de rendimiento esperado para la política estacionaria f, entonces más débilmente, una condición suficiente podría ser que, $V_f(i) \geq 0$ para toda i.

De hecho, una condición suficiente aún más débil podría ser que $\lim_n \inf E_\pi\left[V_f(X_n) | X_0 = i\right] \geq 0$ para toda i y toda π.

De los teoremas 1 y 2 del capítulo 1, una política estacionaria f es óptima sí y sólo si, la función de rendimiento esperado satisface la ecuación de optimalidad, esto es, si para cada estado inicial, usar f es mejor que hacer cualquier cosa para esta etapa y luego cambiar a f.

Por lo tanto, esto nos proporciona un método para comprobar si una determinada política puede ser óptima o no, y es particularmente útil en los casos en los cuales existe una política óptima obvia.

Ejemplo de aplicación:

Sea $S = \{0,1, \dots N\}$ el espacio de estados, llamemos a $i \epsilon S$ donde la fortuna es i. Una recompensa de \$1 es ganada sí, y sólo si, la fortuna actual nunca es N.

Objetivo: Determinar una política óptima.

Procedimiento: Primero notemos que, si la fortuna actual es i entonces, nunca se tendría que pagar por apostar más que $N - i$, es decir, en el estado i se puede limitar la elección de apuesta a: $1,2, \dots, \min\{i, N - i\}$.

Conclusión: Esta política es óptima.

Justificación: La justificación es que $U(i)$ la recompensa de alguna política estacionaria satisface:

$$U(i) \geq pU(i + k) + (1 - p)U(i - k) \quad \text{para todo } 0 < i < N \text{ y } k \leq \min\{i, N - 1\}.$$

Estrategia tímida es aquella que siempre apuesta 1.

En el ejemplo anterior, la estrategia tímida transforma el juego en la ruina del apostador clásico o caminata aleatoria.

Nótese que $U(i)$, la probabilidad de alcanzar N antes que 0 cuando se empieza con una fortuna i, $i < N$, satisface:

$$U(i) = \begin{cases} \dfrac{1 - \left(\frac{q}{p}\right)^{i}}{1 - \left(\frac{q}{p}\right)^{N}} & \text{si } \ p \neq \dfrac{1}{2} \\ \dfrac{i}{N} & \text{si } \ p = \dfrac{1}{2}. \end{cases}$$

La estrategia tímida maximiza tanto el tiempo de juego como la probabilidad de que un jugador nunca alcance una fortuna N si $p \geq \frac{1}{2}$.

La estrategia tímida maximiza el tiempo esperado de juego antes de quebrarnos cuando $p < \frac{1}{2}$.

Para todos enteros positivos A, B tales que $A < B$, $P\left(A, B, \frac{1}{2}\right) = \frac{A}{B}$.

La probabilidad de que el juego continúe infinitamente es 0 ya que la probabilidad de que el juego continúe más allá de n partidas es a lo más $\frac{1}{2^n}$.

Estrategia audaz es tal que, si la fortuna es i entonces,

a) Apuesta i, si $i \leq \frac{N}{2}$.
b) Apuesta $N - i$, si $i \geq \frac{N}{2}$.

La estrategia audaz maximiza la probabilidad de que un jugador alcance una fortuna N en el tiempo n cuando $p \leq \frac{1}{2}$.

También maximiza la probabilidad de que nunca la alcance si el tiempo es limitado, para toda n cuando $p \leq \frac{1}{2}$.

Sean $S_i = X_0 + \cdots + X_i$ y $R_i = \frac{S_i}{B}$ $(i \geq 0)$ la fortuna de un jugador después de n juegos y el radio de su fortuna en la última meta respectivamente.

Note que $R_0 = \frac{A}{B}$.

Si $R_i < \frac{1}{2}$ entonces, R_{i+1} será:

i. 0 con probabilidad q
ii. $2R_i$ con probabilidad p,

dado que S_{i+1} sería: 0 o $2S_i$ en los respectivos casos.

Similarmente, si $R_i \geq \frac{1}{2}$, R_{i+1} sería:

iii. 1 con probabilidad p
iv. $2R_i - 1$ con probabilidad q,

dado que S_{i+1} sería: B o $S_i - (B - S_i) = 2S_i - B$ respectivamente.

Los cuatro casos nos permiten categorizar la $i - ésima$ apuesta de dos maneras:

a) La $i - ésima$ apuesta será la última (concluyendo con $S_i = 0$, con probabilidad q si $R_{i-1} \leq \frac{1}{2}$; y concluyendo con $S_i = B$, con probabilidad p, si $R_{i-1} \geq \frac{1}{2}$).

b) Habrá una $(i+1)-ésima$ apuesta. En este caso $R_{i-1} \neq \frac{1}{2}$, y R_{i-1} tiene representación binaria $0.b_1b_2b_3$ ... , R_i será igual a cambiar la primera posición $0.b_2b_3b_4$ .... Esto se sigue dado que $R_{i-1} < \frac{1}{2}$ implica que $b_1 = 0$ y $2R_{i-1} = 0.b_2b_3b_4$ ... , mientras que $R_{i-1} > \frac{1}{2}$ implica que $b_1 = 1$ y $0.b_2b_3b_4$ ... $= 2R_{i-1} - 1$. Procediendo inductivamente, entonces, esto se sigue de que el único valor posible para R_i, es 0 o 1, teniendo una representación binaria igual a un cambio de la $i-ésima$ posicion de la representación binaria para R_0. Obviamente, si R_i es igual a 0 o 1, todas las R_j $j > i$ subsecuentes, tienen el mismo valor.

APÉNDICE

Este Apéndice está basado en las siguientes referencias:

[13]Maitra, A. and Sudderth, W., ***Discrete Gambling and Stochastic Games***, Springer, (2008).

[16] Ross, S., ***Introduction to Probability Models***, 9th ed., Academic Press, (2007).

[17] Ross, S., ***Intoduction to Stochastic Dynamic Programming***, Academic Press, (1983).

A.1. CADENAS DE MARKOV

Proceso estocástico

Un *proceso estocástico* es un concepto matemático que sirve para caracterizar una sucesión de variables aleatorias que evolucionan en función de otra variable, generalmente el tiempo. Cada una de las variables aleatorias del proceso tiene su propia función de distribución de probabilidad y, entre ellas, pueden estar correlacionadas o no.

Cada variable o conjunto de variables sometidas a influencias o impactos aleatorios constituye un proceso estocástico.

Un proceso estocástico se puede definir equivalentemente de dos formas diferentes:

c) Como un conjunto de realizaciones temporales y un índice aleatorio que selecciona una de ellas.
d) Como un conjunto de variables aleatorias X_t indexadas por un índice t, dado $t\epsilon T$, con $T \subseteq \mathbb{R}$.

T puede ser continuo si es un intervalo o discreto si es numerable. Las variables aleatorias X_t toman valores en el espacio probabilístico.

Cadenas De Markov

En la teoría de la probabilidad, se conoce como *Cadena de Markov* a un tipo especial de proceso estocástico discreto en el que la probabilidad de que ocurra un evento depende del evento inmediatamente anterior. En efecto, estas cadenas tienen memoria. "Recuerdan" el último evento y esto condiciona las posibilidades de los eventos futuros. Esta dependencia del evento anterior distingue a las cadenas de Markov de las series de eventos independientes, como tirar una moneda al aire o un dado.

En matemáticas, se define como un proceso estocástico discreto que cumple con la *propiedad de Markov*, es decir, si se conoce la historia del sistema hasta su instante actual, su estado presente resume toda la información relevante para describir en probabilidad su estado futuro.

Una cadena de Markov es una sucesión X_1, X_2, X_3, ... de variables aleatorias. El rango de estas variables, es llamado espacio de estados, el valor de X_n es el estado del proceso en el tiempo n. Si la distribución de probabilidad condicional de X_{n+1} en estados pasados es una función de X_n por sí sola, entonces:

$$P(X_{n+1} = x_{n+1} | X_n = x_n\,,\, X_{n-1} = x_{n-1}\,, \ldots,\, X_2 = x_2\,,\, X_1 = x_1,$$

$$X_0 = x_0) = P(X_{n+1} = x_{n+1} | X_n = x_n).$$

Donde x_i es el estado del proceso en el instante i. La identidad mostrada es la *propiedad de Markov*.

Cadenas homogéneas y no homogéneas

- Una cadena de Markov se dice *homogénea* si la probabilidad de ir del estado i al estado j en un paso no depende del tiempo en el que se encuentra la cadena, esto es:

$$P(X_n = j | X_{n-1} = i) = P(X_1 = j | X_0 = i), \text{para todo } n \text{ y para cualesquiera } i, j.$$

Si para alguna pareja de estados y para algún tiempo n la propiedad antes mencionada no se cumple diremos que la cadena de Markov es *no homogénea*.

Probabilidades de transición y matriz de transición

- La probabilidad de ir del estado *i* al estado *j* en *n* unidades de tiempo es,

$$P_{ij}^{(n)} = P(X_n = j | X_0 = i).$$

- En la probabilidad de transición en un paso se omite el superíndice de modo que queda,

$$P_{ij} = P(X_1 = j | X_0 = i).$$

- Un hecho importante es que las probabilidades de transición en n pasos satisfacen la ecuación de *Chapman - Kolmogorov*, esto es, para cualquier k tal que $0 < k < n$ se cumple que,

$$P_{ij}^{(n)} = \sum_{r \epsilon E} P_{ir}^{(k)} P_{rj}^{(n-k)},$$

donde E denota el espacio de estados.

- Cuando la cadena de Markov es homogénea, muchas de sus propiedades útiles se pueden obtener a través de su matriz de transición, definida entrada a entrada como $A_{ij} = P_{ij}$, esto es, la entrada ij corresponde a la probabilidad de ir del estado i al j en un paso.

$\boldsymbol{P} = \begin{Vmatrix} P_{00} & P_{01} & P_{02} & \cdots \\ P_{10} & P_{11} & P_{12} & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ P_{i0} & P_{i1} & P_{i2} & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{Vmatrix}$ denota la matriz de transición en un paso del estado i al j.

Del mismo modo se puede obtener la matriz de transición en n pasos como:

$$A_{ij}^{(n)} = P_{ij}^{(n)}, \text{ donde } P_{ij}^{(n)} = P(X_n = j | X_0 = i).$$

Vector de probabilidad invariante

- Se define la distribución inicial $\pi(x) = P(X_0 = x)$.
- Diremos que un vector de probabilidad v (finito o infinito numerable) es invariante para una cadena de Markov si $v\boldsymbol{P} = v$, donde $\boldsymbol{P}$ denota la matriz de transición de la cadena de Markov. Al vector de probabilidad invariante también se le llama *distribución estacionaria* o *distribución de equilibrio*.

Recurrencia

En una cadena de Markov con espacio de estados E, si $x \epsilon E$ se define:

$L_x = P(X_n = x \text{ para algún } n \epsilon \mathbb{N} | X_0 = x)$ y diremos que,

- x es estado *recurrente* si $L_x = 1$.
- x es *transitorio* si $L_x < 1$.
- x es *absorbente* si $P_{x,x} = 1$.

A.1.1. TIPOS DE CADENAS DE MARKOV

Cadenas positivo-recurrentes

Una cadena de Markov se dice *positivo-recurrente* si todos sus estados son positivo-recurrentes. Si la cadena es además irreducible es posible demostrar que existe un único vector de probabilidad invariante y está dado por:

$$\pi_x = \frac{1}{\mu_x}.$$

Cadenas regulares

Una cadena de Markov se dice *regular* (también *primitiva* o *ergódica*) si existe alguna potencia positiva de la matriz de transición cuyas entradas sean todas estrictamente mayores que cero.

Cuando el espacio de estados Ees finito, si $\boldsymbol{P}$ denota la matriz de transición de la cadena se tiene que:

$$\lim_{n\to\infty} \boldsymbol{P}^n = W,$$

donde W es una matriz con todos sus renglones iguales a un mismo vector de probabilidad $\boldsymbol{w}$, que resulta ser el vector de probabilidad invariante de la cadena. En el caso de cadenas regulares, éste vector invariante es único.

Cadenas absorbentes

Una cadena de Markov con espacio de estados finito se dice absorbente si se cumplen las dos condiciones siguientes:

1. La cadena tiene al menos un estado absorbente.
2. De cualquier estado no absorbente se accede a algún estado absorbente.

Si denotamos como A al conjunto de todos los estados absorbentes y a su complemento como D, tenemos los siguientes resultados:

- Su matriz de transición siempre se puede llevar a una de la forma:

$$\boldsymbol{P} = \begin{pmatrix} \mathrm{Q} & \mathrm{R} \\ 0 & \mathrm{I} \end{pmatrix},$$

donde la submatriz Q corresponde a los estados del conjunto D, I es la matriz identidad, 0 es la matriz nula y R alguna submatriz.

- $P_x(T_A < \infty) = 1$, esto es, no importa en donde se encuentre la cadena, eventualmente terminará en un estado absorbente.

A.1.2. LA RUINA DE UN APOSTADOR

Considere un apostador quien para cada partida del juego tiene probabilidad p de ganar una unidad y probabilidad $q = 1 - p$ de perder una unidad. Suponiendo que las partidas son independientes, cuál es la probabilidad de que, empezando con i unidades, el apostador alcance una fortuna N antes que 0 (la ruina).

Si X_n denota la fortuna del jugador en el tiempo n, entonces el proceso $\{X_n : n = 0, 1, 2, \dots\}$ es una cadena de Markov con probabilidades de transición:

$$P_{00} = P_{NN} = 1$$

$$P_{ii+1} = p = 1 - P_{ii-1}, \quad i = 1, \dots, N-1.$$

Esta cadena de Markov tiene tres clases, $\{0\}, \{1, \dots, N-1\}$ y $\{N\}$; la primera y la tercera clases son recurrentes y la segunda transitoria. Dado que cada estado transitorio solo fue tocado una cantidad finita de veces, se sigue que después de un tiempo finito, el jugador tendrá, o riqueza N o riqueza 0 (la ruina).

Sea P_i, $i = 0, 1, \dots, N$, la probabilidad de que, empezando con i, la fortuna del apostador eventualmente será N. Como consecuencia del resultado de la partida inicial del juego tenemos que,

$$P_i = pP_{i+1} + qP_{i-1}, \ i = 1, \dots, N-1.$$

O equivalentemente, dado que $p + q = 1$,

$$pP_i + qP_i = pP_{i+1} + qP_{i-1}$$

o,

$$P_{i+1} - P_i = \frac{q}{p}\left(P_i - P_{i-1}\right), \quad i = 1, \ldots, N-1.$$

Cuando $P_0 = 0$, se tiene que,

$$\begin{aligned}
P_2 - P_1 &= \frac{q}{p}\,(P_1 - P_0) = \frac{q}{p}\,P_1 \\
P_3 - P_2 &= \frac{q}{p}\,(P_2 - P_1) = \left(\frac{q}{p}\right)^2 P_1 \\
&\vdots \\
P_i - P_{i-1} &= \frac{q}{p}\,(P_{i-1} - P_{i-2}) = \left(\frac{q}{p}\right)^{i-1} P_1 \\
&\vdots \\
P_N - P_{N-1} &= \frac{q}{p}\,(P_{N-1} - P_{N-2}) = \left(\frac{q}{p}\right)^{N-1} P_1.
\end{aligned}$$

Sumando las primeras $(i-1)$ ecuaciones, tenemos que:

$$P_i - P_1 = P_1\left(\left(\frac{q}{p}\right) + \left(\frac{q}{p}\right)^2 + \cdots + \left(\frac{q}{p}\right)^{i-1}\right)$$

o,

$$P_i = \begin{cases} \dfrac{1-\left(\frac{q}{p}\right)^i}{1-\left(\frac{q}{p}\right)} P_1 \;, & \text{si } \dfrac{q}{p} \neq 1 \\ iP_1 \;, & \text{si } \dfrac{q}{p} \neq 1. \end{cases}$$

Ahora, usando el hecho que $P_N = 1$, obtenemos que:

$$P_1 = \begin{cases} \dfrac{1-\left(\frac{q}{p}\right)}{1-\left(\frac{q}{p}\right)^N} \;, & \text{si } p \neq \dfrac{1}{2} \\ \dfrac{1}{N} \;, & \text{si } p = \dfrac{1}{2}\,. \end{cases}$$

Entonces,

$$P_i = \begin{cases} \dfrac{1-\left(\frac{q}{p}\right)^i}{1-\left(\frac{q}{p}\right)^N} & , \quad \text{si } p \neq \dfrac{1}{2} \\ \dfrac{i}{N} & , \quad \text{si } p = \dfrac{1}{2} . \end{cases}$$

Notando que si $N \to \infty$,

$$P_i = \begin{cases} 1-\left(\frac{q}{p}\right)^i & , \quad \text{si } p > \dfrac{1}{2} \\ 0 & , \quad \text{si } p \leq \dfrac{1}{2} . \end{cases}$$

Como consecuencia, si $p > \frac{1}{2}$, existe una probabilidad generosa de que la fortuna del jugador se incremente indefinidamente; mientras que si $p \leq \frac{1}{2}$, el jugador podría, con probabilidad 1, quebrar contra un adversario infinitamente rico.

BIBLIOGRAFÍA

[1] Basharin, G., Langville, A., Naumov, V., "***The life and the work of A. A. Markov***", Linear Algebra and its Applications Vol. **386**, p.p. 3–26, (2004), disponible en: https://netfiles.uiuc.edu/meyn/www/spm_files/Markov-Work-and-life.pdf, consultado: 25/07/2012.

[2] Bak, J., "***The anxious gambler's ruin***", Mathematics Magazine Vol. **74** No.3, pp. 182-193, (Jun., 2001), Mathematical Association of America.

[3] Bellman, R.E., ***Dynamic Programming***, Princeton U. Press, Princeton, N.J., (1957).

[4] Bertsekas, D. P., ***Dynamic Programming,*** Prentice Hall, Eaglewood Cliffs, NJ. MA, (1987).

[5] De Moivre, A., "***De mensura sortis***", Phil. R. Soc. **27**, p.p. 213-264, (1711). Translated in Internat. Statist. Rev. **52**, No. 3, 237-262 (1984).

[6] Dubbins, L. E. and Savage, L. J., ***Inequalities for Stochastic Processess; How to Gamble If You Must***, Dover Publications (1976).

[7] Edwards, A. W. F., "***Pascal's problem: The Gambler´s Ruin***", Internat. Statist. Rev. **51**, p.p. 73-79, (1983).

[8] Feller, W., ***An Introduction to Probability Theory and Its Applications***, 3rd ed., Wiley, New York (1968).

[9] Geertz, C., ***La Interpretación de las Culturas***, Gedisa 12ª ed., Barcelona (2003).

[10] Howard, R. A., ***Dynamic Programming and Markov Processes***, MIT Press, Cambridge, Massachusetts, (1960).

[11] Isaac, R. , "***Bold play is best: a simple proof***", Mathematics Magazine Vol. **72**, p.p. 405-407, (1999).

[12] Siegrist, K., "***How to gamble if you must***", Vol. **8**, Department of Mathematical Sciences, University of Alabama in Huntsville, Mathematical Association of America disponible en: www.maa.org/joma/.../siegrist/redblack.pdf, consultado: 23/07/2012.

[13]Maitra, A. and Sudderth, W., ***Discrete Gambling and Stochastic Game***, Springer, (2008).

[14]Martínez, M. A., "***Historia de las apuestas: la prehistoria***", (2011), disponible en: http://suite101.net/article/historia-de-las-apuestas-la-prehistoria-a49559, consultado: 01/12/2012.

[15] Puterman, M.L., ***Markov Decision Processes,*** Wiley, New York, (1994).

[16]Ross, S., ***Introduction to Probability Models*** 9th ed., Academic Press, (1983).

[17]Ross, S., ***Intoduction to Stochastic Dynamic Programming***, Academic Press, (1983).