UNIVERSIDAD DE SONORA
DIVISIÓN DE CIENCIAS EXACTAS Y NATURALES
DEPARTAMENTO DE MATEMÁTICAS

Una Introducción a la Teoría de Martingalas

TESIS

Que para obtener el título de:

Licenciado en Matemáticas

Presenta:

ROSALÍA GUADALUPE HERNÁNDEZ AMADOR

Director: Dr. Jesús Adolfo Minjárez Sosa

Hermosillo, Sonora

Agosto 2002

A mis padres

Este trabajo fue financiado y apoyado por el Consejo Nacional de Ciencia y Tecnología (CONACyT) dentro del proyecto "Control Adaptado de Sistemas Estocásticos", con clave 37239E, bajo la dirección del Dr. Jesús Adolfo Minjárez Sosa.

Índice General

0.1 Introducción

Indudablemente, la probabilidad es una de las ramas de la matemática que ha jugado un papel fundamental en la solución de problemas específicos que se presentan en otras áreas del conocimiento. Esto se debe a que muchos de estos problemas se pueden modelar mediante una familia de variables aleatorias parametrizadas por el tiempo, como por ejemplo, sistemas dinámicos en los que intervienen elementos aleatorios los cuales representan cierta incertidumbre que influye en el comportamiento del sistema, problemas de decisiones secuenciales, dentro de los cuales se encuentran los juegos de apuestas sucesivas, entre otros. A dicha familia de variables aleatorias se les conoce como procesos estocásticos .

Existen muchos tipos de procesos estocásticos cuya clasificación se tiene en base a sus propiedades particulares. Por ejemplo, procesos estacionarios, con incrementos independientes, de Markov, martingalas, etc., aunque esta clasificación no los hace excluyentes. La teoría desarrollada para cada uno de estos procesos es muy amplia y podríamos decir que independientes una de la otra.

De los temas más estudiados en los cursos de procesos estocásticos nivel licenciatura son las cadenas de Markov. Quizá esto se debe a que sus propiedades basadas en la definición de probabilidad condicional, son muy intuitivas y tienen aplicaciones directas y muy sencillas. Por lo contrario, las martingalas, cuyas propiedades se enfocan en la definición de esperanza condicional, es un tema que por lo general se excluye de los cursos, aun siendo igual de importantes que las cadenas de Markov. Esto constituye la motivación principal de desarrollar este trabajo.

Si recurrimos al diccionario de la lengua española para ver el significado de la palabra *martingala* nos podemos topar con lo siguiente:

- calza que los hombres de armas llevaban debajo de los quijotes;
- en el juego del monte, lance que consiste en apostar simultáneamente a cuatro cartas de las del albur contra la quinta;
- combinación que permite ganar el juego;

entre otros, y como sinónimo: artimaña, marrullería, trampa.

Sin embargo, en matemáticas, las martingalas son una clase de procesos estocásticos cuyo surgimiento fue motivado precisamente al modelar juegos de azar justos, aunque en el transcurso de los años, su desarrollo ha alcanzado importantes aplicaciones en otras áreas, así como dentro de la matemática misma [ver, e.g., [4]].

El objetivo de este trabajo es presentar una introducción a la teoría de martingalas. En este sentido nos restringiremos al caso más sencillo: procesos estocásticos discretos a tiempo discreto. Sin embargo, esto no hace que el trabajo carezca de importnancia, ya que existen muchos problemas reales que se modelan en este contexto, como por ejemplo, problemas de apuestas sucesivas, sistemas genéticos, problemas de finanzas, y cierta clase de sistemas de colas. Algunos de estos ejemplos se analizarán a lo largo del trabajo.

La tesis está compuesta de tres capítulos. En el primero de ellos se han resumido algunos resultados fundamentales de la teoría de probabilidad, poniendo énfasis principalmente en el análisis de esperanza condicional y algunas de sus propiedades más importantes. Se expone también la teoría necesaria sobre procesos estocásticos y cadenas de Markov a tiempo discreto, las cuales se retoman en los siguientes capítulos

Durante el segundo capítulo se presenta la definición de martingala así como sus propiedades fundamentales. Primeramente y a manera de motivación se presenta un análisis de estrategias de apuestas, tomando en cuenta que el concepto de martingala en teoría de probabilidad tiene su origen en juegos de apuestas, a saber, proporciona un modelo de un juego de azar justo. Posteriormente se exponen algunos ejemplos que ilustran las definiciones introducidas en el capítulo, y se presenta también un resultado muy importante que nos muestra la relación que guardan las martingalas y las cadenas de Markov.

Finalmente se conluye este trabajo presentando en el tercer capítulo, dos resultados clásicos en teoría de martingalas: el Teorema de Paro Opcional para Martingalas Acotadas y el Teorema de Convergencia de Martingalas. Ilustramos estos resultados mediante algunos ejemplos.

Capítulo 1

Elementos de la Teoría de Probabilidad y Procesos Estocásticos

1.1 Introducción

En este capítulo presentamos los elementos de la teoría de probabilidad y procesos estocásticos que usaremos a lo largo del trabajo. Se dedica especial atención al concepto de esperanza condicional y sus propiedades, lo cual constituye la herramienta principal para el desarrollo del mismo. Muchos detalles estándares de la teoría de probabilidad serán omitidos, pero se procurará incluir los más importantes.

1.2 Probabilidad Condicional

Empezamos esta sección definiendo el espacio de probabilidad fundamental, para después establecer la definición de probabilidad condicional.

Sea Ω el *espacio muestral* de un experimento y $\mathcal{F}$ una σ – *álgebra* de subconjuntos de Ω. Es decir, $\mathcal{F}$ es una familia de subconjuntos de Ω que satisface las siguientes propiedades:

a) El conjunto vacío $\varnothing$ pertenece a $\mathcal{F}$;

b) Si $A_1, A_2, ... \in \mathcal{F}$ entonces $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$;

c) Si $A \in \mathcal{F}$ entonces su complemento A^c pertenece a $\mathcal{F}$.

A los elementos de $\mathcal{F}$ los llamaremos *eventos.*

Definición 1.1 *Una* ***medida de probabilidad*** *P sobre $(\Omega, \mathcal{F})$ es una función*

$$P : \mathcal{F} \to [0, 1]$$

que satisface

a) $P[\varnothing] = 0, \; P[\Omega] = 1;$

b) Si $A_1, A_2, ...$ es una colección de elementos en $\mathcal{F}$ tales que $A_i \cap A_j = \varnothing$ para todo $i \neq j$ entonces

$$P[\bigcup_{n=1}^{\infty} A_n] = \sum_{n=1}^{\infty} P[A_n],$$

y se dice que la medida P es aditivamente numerable.

La terna $(\Omega, \mathcal{F}, P)$ es llamada *espacio de probabilidad.* Es siempre posible asociar un espacio $(\Omega, \mathcal{F}, P)$ a cada experimento y todas las preguntas en las que podamos estar interesados acerca del experimento, podemos formularlas en términos de éste.

Definición 1.2 *Sean $A, B \in \mathcal{F}$ eventos tales que $P[B] \neq 0$. Se define la* ***probabilidad condicional*** *de A dado el evento B como*

$$P[A \mid B] := \frac{P[A \cap B]}{P[B]}.$$

Intuitivamente, esto significa que si sabemos (o bajo el supuesto) que un evento B con probabilidad positiva ha ocurrido, el conjunto de posibles resultados se reduce a B, y entonces solamente cabe la posibilidad de que ocurra la parte de A que pertenece también a B, esto es, el evento $A \cap B$. Así la oportunidad de ocurrencia del evento A dado que ha ocurrido B se convierte en la medida relativa de $A \cap B$ respecto a B.

Definición 1.3 *Los **eventos** A y B se dicen ser **independientes** si*

$$P[A \cap B] = P[A]P[B].$$

En general, una familia de eventos $\{A_i : i \in T\}$ donde T denota un conjunto arbitrario de índices se dice ser independiente si

$$P[\bigcap_{i \in J} A_i] = \prod_{i \in J} P[A_i]$$

para todo subconjunto finito J de T.

Una caracterización para conjuntos independientes en términos de probabilidad condicional es la siguiente: sean A y B eventos tales que $P[B] \neq 0$ y $P[A]$ arbitraria. Entonces A y B son *independientes* si y sólo si

$$P[A \mid B] = P[A],$$

lo cual significa que el saber que ha ocurrido el evento B no cambia la probabilidad de ocurrencia de A.

1.3 Variables y Vectores Aleatorios

Presentamos a continuación la definición de variable y vector aleatorio, así como sus principales características.

Definición 1.4 *Sea $(\Omega\ , \mathcal{F}, P)$ un espacio de probabilidad. Una **variable aleatoria** (v.a.) X es una función*

$$X : \Omega \to \mathbb{R}$$

tal que

$$X^{-1}((-\infty, x]) = \{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F} \quad \text{para cada } x \in \mathbb{R}.$$

Es decir, una variable aleatoria X es una función con valores en los reales que asigna el valor $X(\omega)$ a cada resultado $\omega \in \Omega$ del experimento. Así mismo, se requiere que la imágen inversa bajo X de intervalos de la forma $(-\infty, x]$,

$x \in \mathbb{R}$, pertenezcan a $\mathcal{F}$. Esto nos permite estudiar el comportamiento probabilístico de las variables aleatorias.

Es usual denotar por

$$[X \in (-\infty, x]] \quad \text{o} \quad [X \leq x]$$

al conjunto $X^{-1}((-\infty, x])$.

En muchos casos no es fácil escribir explícitamente el espacio muestral asociado a un experimento aleatorio, lo que hace necesario analizar la característica de interés del experimento por medio de variables aleatorias. En este sentido, es importante estudiar el comportamiento probabilístico de dichas variables. La manera general de hacerlo es mediante la función de distribución.

Definición 1.5 *La* ***función de distribución*** F_X *de una variable aleatoria* X *es la función*

$$F_X : \mathbb{R} \to [0, 1]$$

dada por

$$F_X(x) := P[X \leq x].$$

Obsérvese que F_X satisface las siguientes propiedades:

a) Es no decreciente, es decir, si $x_1 < x_2$ entonces $F_X(x_1) \leq F_X(x_2)$;

b) Es continua por la derecha, esto es,

$$F_X(x^+) := \lim_{h \to 0^+} F_X(x + h) = F_X(x) \quad \text{para cada } x \in \mathbb{R};$$

c)

$$\lim_{x \to \infty} F_X(x) = 1, \qquad \lim_{x \to -\infty} F_X(x) = 0.$$

d) Para cada $x \in \mathbb{R}$,

$$F_X(x^-) = P[X < x]$$

donde $F_X(x^-) := \lim_{h \to 0^+} F_X(x - h)$.

En términos de la función de distribución, podemos clasificar las variables aleatorias como sigue:

Definición 1.6 .

a) *La* ***variable aleatoria*** *X se dice ser* ***absolutamente continua*** *si su función de distribución se puede expresar como*

$$F_X(x) = \int_{-\infty}^{x} f_X(u)du, \qquad x \in \mathbb{R}$$

para alguna función integrable $f_X : \mathbb{R} \longrightarrow [0, \infty)$. f_X es llamada la ***función de densidad*** *de X.*

b) *La* ***variable aleatoria*** *X se dice ser* ***discreta*** *si existe una colección numerable $\{x_1, x_2, ...\} \subset \mathbb{R}$ con $x_i \neq x_j$ para $i \neq j$ tal que*

$$F_X(x) = P[X \leq x] = \sum_{x_n \leq x} P[X = x_n], \qquad x \in \mathbb{R}.$$

La funciòn f_X definida por $f_X(x) := P[X = x]$, $x \in \mathbb{R}$, es llamada la ***función de probabilidad*** *de X.*

Existen situaciones donde es necesario analizar dos o más variables aleatorias simultáneamente. Esto nos lleva a tener que generalizar las definiciones anteriores. En principio tenemos:

Definición 1.7 *Sean $X_1, ..., X_n$ variables aleatorias definidas en el mismo espacio de probabilidad $(\Omega, \mathcal{F}, P)$. Al vector*

$$\mathbf{X} = (X_1, ..., X_n)$$

se le llama ***vector aleatorio de dimensión*** *n. Cuando las componentes de $\mathbf{X}$ son variables aleatorias discretas, decimos que $\mathbf{X}$ es un* ***vector aleatorio discreto****. En el caso de ser continuas $X_1, ..., X_n$, se dice que $\mathbf{X}$ es un* ***vector aleatorio continuo****.*

En base a esta definición usaremos indistintamente los términos "comportamiento conjunto de las variables aleatorias $X_1, ..., X_n$" y "comportamiento del vector $\mathbf{X}$".

Definición 1.8 *La **función de distribución conjunta** $F_{X_1,...,X_n}$ de las variables aleatorias $X_1, ..., X_n$ definidas sobre el mismo espacio de probabilidad (Ω, F, P), se define como la función*

$$F_{X_1,...,X_n} : \mathbb{R}^n \to [0, 1]$$

dada por

$$F_{X_1,...,X_n}(x_1, ..., x_n) := P[X_1 \leq x_1, ..., X_n \leq x_n].$$

Además, si existe una función integrable

$$f_{X_1,...,X_n} : \mathbb{R}^n \to [0, \infty)$$

tal que

$$F_{X_1,...,X_n}(x_1, ..., x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{X_1,...,X_n}(u_1, ..., u_n) du_1 ... du_n,$$

*decimos que $X_1, ..., X_n$ tienen distribución conjunta continua. En este caso $f_{X_1,...,X_n}$ se llama **función de densidad conjunta** de las variables $X_1, ..., X_n$.*

Similarmente, para el caso discreto se tiene que para todo $(x_1, ..., x_n) \in \mathbb{R}^n$

$$F_{X_1,...,X_n}(x_1, ..., x_n) = \sum_{k_1 \leq x_1} \cdots \sum_{k_n \leq x_n} f_{X_1,...,X_n}(k_1, ..., k_n),$$

*donde $f_{X_1,...,X_n}(k_1, ..., k_n) := P[X_1 = k_1, ..., X_n = k_n]$, $(k_1, ..., k_n) \in \mathbb{R}^n$. A $f_{X_1,...,X_n}$ se le conoce como **función de probabilidad conjunta** de $X_1, ..., X_n$, o función de probabilidad del vector $\mathbf{X} = (X_1, ..., X_n)$.*

Sea $\mathbf{X} = (X_1, ..., X_n)$ un vector aleatorio discreto con función de probabilidad $f_{X_1,...,X_n}$ y sea f_{X_i} la función de probabilidad de la variable aleatoria X_i. Entonces

$$f_{X_i}(x_i) = \sum_{x_j, j \neq i} f_{X_1,...,X_n}(x_1, ..., x_n).$$

Esto es, la función de probabilidad de X_i se puede obtener sumando los valores de $f_{X_1,...,X_n}(x_1, ..., x_n)$ sobre todo x_j con $j \neq i$. En este caso llamamos a f_{X_i} la *función de probabilidad marginal* de X_i.

Definición 1.9 *Sean $X_1, ..., X_n$ variables aleatorias definidas en el mismo espacio de probabilidad. Decimos que $X_1, ..., X_n$ son* ***independientes*** *si para toda colección de intervalos $\{J_i\}_{i=1}^n$ con $J_i \subset \mathbb{R}$, se tiene que*

$$P[X_1 \in J_1, ..., X_n \in J_n] = \prod_{i=1}^{n} P[X_i \in J_i].$$

En general, una familia de variables aleatorias $\{X_t : t \in T\}$ es independiente si para todo subconjunto finito J de T las variables $\{X_t : t \in J\}$ son independientes.

Si $X_1, ..., X_n$ son independientes, entonces $X_1, ..., X_k$ son independientes para $k < n$. Mas aún, $X_{i_1}, ..., X_{i_k}$ son independientes para todos los enteros $1 \leq i_1 < ... < i_k \leq n$.

La independencia de variables aleatorias se puede caracterizar por medio de las funciones de distribución correspondientes, así como por medio de las funciones de probabilidad para el caso discreto como lo muestran los siguientes resultados:

Teorema 1.10 *Las variables aleatorias $X_1, ..., X_n$ son independientes si y sólo si la función de distribución conjunta $F_{X_1,...,X_n}$ de $X_1, ..., X_n$ se puede escribir como*

$$F_{X_1,...,X_n}(x_1, ..., x_n) = \prod_{i=1}^{n} F_{X_i}(x_i), \quad \textit{para cada } (x_1, ..., x_n) \in \mathbb{R}^n,$$

donde F_{X_i} es la función de distribución de la variable aleatoria X_i.

Teorema 1.11 *Sea $\mathbf{X} = (X_1, ..., X_n)$ un vector aleatorio discreto con función de probabilidad $f_{X_1,...,X_n}$, y sea f_{X_i} la función de probabilidad de la variable aleatoria X_i, $i = 1, ..., n$. Entonces $X_1, ..., X_n$ son independientes si y sólo si*

$$f_{X_1,...,X_n}(x_1, ..., x_n) = \prod_{i=1}^{n} f_{X_i}(x_i), \quad (x_1, ..., x_n) \in \mathbb{R}^n.$$

- Para los objetivos del trabajo es suficiente restringirnos al estudio de variables aleatorias discretas. En este sentido, el resto del trabajo se centrará principalmente en definiciones y resultados que involucran solo a este tipo de variables aleatorias.

1.4 Esperanza y Esperanza Condicional

El concepto de esperanza condicional es la herramienta principal en el estudio de martingalas. Es por esto que se dedica la presente sección a su análisis. Primero se verá la noción de valor esperado o esperanza de una variable aleatoria. Posteriormente, se analiza el concepto de esperanza condicional, el cual se realiza primero condicionando sobre un evento y finalmente sobre una variable aleatoria discreta.

Definición 1.12 *La **esperanza** (valor esperado) de una variable aleatoria discreta X con función de probabilidad f_X se define por*

$$E[X] := \sum_x x f_X(x),$$

siempre que la suma sea absolutamente convergente.

Observemos que para un evento $A \in F$ y la variable aleatoria $1_A : \Omega \to \mathbb{R}$ definida como

$$1_A(\omega) := \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \in A^c \end{cases}$$

se tiene que

$$E[1_A] = P(A).$$

Algunas propiedades que se deducen directamente de la definición son las siguientes:

a) Supongamos que la variable aleatoria X es constante casi seguramente (c.s.), es decir, $P[X = c] = 1$ para alguna constante $c \in \mathbb{R}$, entonces $E[X] = c$.

b) Sea X una variable aleatoria esencialmente acotada, es decir, $P[|X| \leq M] = 1$ para alguna constante $M \in \mathbb{R}$, entonces $|E[X]| \leq M$.

Otras propiedades de la esperanza se resumen en los siguientes resultados:

Teorema 1.13 *Sea* $\mathbf{X} = (X_1, ..., X_n)$ *un vector aleatorio discreto,* $h : \mathbb{R}^n \to \mathbb{R}$*, y supongamos que la variable aleatoia* $Y = h(\mathbf{X})$ *es tal que su esperanza está definida. Entonces*

$$E[Y] = E[h(\mathbf{X})] = \sum_{\mathbf{x}\in\mathbb{R}^n} h(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x})$$

donde $f_{\mathbf{X}}$ *es la función de probabilidad conjunta de* $X_1, ..., X_n$.

Teorema 1.14 *Sea* $\mathbf{X} = (X_1, ..., X_n)$ *un vector aleatorio y supongamos que* $Y_i = g_i(\mathbf{X})$, $i = 1, ..., n$ *son variables aleatorias con esperanza finita. Entonces:*

a) Para todas las constantes $c_1, ..., c_n$

$$E[c_1 Y_1 + ... + c_n Y_n] = \sum_{i=1}^{n} c_i E[Y_i].$$

b) Si $P[Y_1 \leq Y_2] = 1$, *entonces*

$$E[Y_1] \leq E[Y_2].$$

Teorema 1.15 *Sean* $X_1, ..., X_n$ *variables aleatorias independientes con esperanza finita. Entonces*

$$E[X_1 \cdot X_2 \cdots X_n] = \prod_{i=1}^{n} E[X_i].$$

1.4.1 Esperanza Condicional

Antes de definir esperanza condicional, primero veamos la siguiente defininción:

Definición 1.16 *Sean* X *y* Y *variables aleatorias discretas definidas en el mismo espacio de probabilidad, y sea* $x \in \mathbb{R}$ *tal que* $f_X(x) > 0$. *La* ***función de probabilidad condicional*** *de* Y *dado que* $X = x$ *se define como*

$$f_{Y|X}(y \mid x) := \frac{P[X = x, Y = y]}{f_X(x)}, \quad y \in \mathbb{R}.$$

Es claro que para cada $x \in \mathbb{R}$, donde $f_X(x) > 0$, se tiene

a) $f_{Y|X}(y \mid x) \geq 0, \quad y \in \mathbb{R}$.

b) $\sum\limits_y f_{Y|X}(y \mid x) = 1$.

En base a esta definición tenemos:

Definición 1.17 *Sean X y Y variables aleatorias discretas definidas en el mismo espacio de probabilidad con función de probabilidad condicional $f_{Y|X}$. La* ***esperanza condicional de** Y **dado** $X = x$* *se define como*

$$E[Y \mid X = x] := \sum_y y f_{Y|X}(y \mid x)$$

donde la suma es sobre el conjunto numerable de valores de y para los cuales $f_{Y|X}(y \mid x) > 0$.

Como para cada $x \in \mathbb{R}$ donde $f_X(x) > 0$, $f_{Y|X}(y \mid x)$ es una función de probabilidad, la esperanza condicional $E[Y \mid X = x]$ satisface las siguientes propiedades, las cuales pueden demostrarse directamente de la definición:

a) $E[c_1 Y_1 + \ldots + c_n Y_n \mid X = x] = \sum\limits_{i=1}^{n} c_i E[Y_i \mid X = x]$ para todo $c_1, \ldots, c_n \in \mathbb{R}$.

b) $E[h(X) \mid X = x] = h(x)$ para cualquier función $h : \mathbb{R} \to \mathbb{R}$.

c) $E[Y \mid X = x] = E[Y]$ si X y Y son independientes.

En vista de la definición anterior, podemos ver a $E[Y \mid X = x]$ como una función de x:

$$g(x) = E[Y \mid X = x]. \tag{1.1}$$

De acuerdo a este hecho tenemos la siguiente definición:

Definición 1.18 *La* ***esperanza condicional de** Y **dado** X* *se define como*

$$E[Y \mid X] := g(X),$$

donde g está dada por (1.1).

La diferencia esencial entre $E[Y \mid X = x]$ y $E[Y \mid X]$ es que $E[Y \mid X = x]$ es un número mientras que $E[Y \mid X]$ es una variable aleatoria. De hecho, $E[Y \mid X]$ se obtiene reemplazando x por X en la expresión para $E[Y \mid X = x]$. En este sentido, estaremos utilizando el término *casi seguramente* (c.s.) para comparar la esperanza condicional con una variable aleatoria. Es decir, escribiremos

$$E[Y \mid X] = Z \quad \text{c.s.},$$

donde c.s. significa que $P[E[Y \mid X] = Z] = 1$.

Dentro de las principales propiedades de la esperanza condicional tenemos:

Teorema 1.19 .

a) La esperanza condicional $g(X) = E[Y \mid X]$ satisface

$$E[g(X)h(X)] = E[Yh(X)]$$

para cualquier función $h : \mathbb{R} \to \mathbb{R}$ para la cual ambas esperanzas existen.

b) Sea $x \in \mathbb{R}$ tal que $f_X(x) > 0$. Entonces

$$E[XY \mid X = x] = xE[Y \mid X = x].$$

c) Si $H : \mathbb{R} \to \mathbb{R}$ es una función inyectiva, entonces

$$E[Y \mid H(X)] = E[Y \mid X] \quad \text{c.s.}$$

Demostración. .

a) Sea $f_X(x)$ la función de probabilidad de X. Por el Teorema 1.13, tenemos

$$\begin{aligned}
E[g(X)h(X)] &= \sum_x g(x)h(x)f_X(x) = \sum_x E[Y \mid X = x]h(x)f_X(x) \\
&= \sum_x \left[\sum_y y f_{Y|X}(y \mid x)\right] h(x)f_X(x) \\
&= \sum_{x,y} yh(x)f_{X,Y}(x,y) \\
&= E[Yh(X)].
\end{aligned}$$

b) Si $x \neq 0$, este resultado es una consecuencia directa de la Definición 1.17 . En efecto,

$$\begin{aligned} E[XY \mid X = x] &= \sum_z z f_{XY|X}(z \mid x) \\ &= \sum_z z \frac{P[X = x, XY = z]}{f_X(x)} \\ &= \sum_z z \frac{P[X = x, Y = \frac{z}{x}]}{f_X(x)} \end{aligned}$$

y haciendo $y = \frac{z}{x}$ se tiene

$$\begin{aligned} E[XY \mid X = x] &= \sum_{yx} yx \frac{P[X = x, Y = y]}{f_X(x)} \\ &= x \left(\sum_y y \frac{P[X = x, Y = y]}{f_X(x)} \right) \\ &= xE[Y \mid X = x]. \end{aligned}$$

El caso $x = 0$ se sigue trivialmente.

c) Aplicando el hecho de que $P[X = x] = P[H(X) = H(x)]$ para cada $x \in \mathbb{R}$, se sigue directamente el resultado.

■

El Teorema 1.19 tiene las siguientes consecuencias importantes:

- Si $h(x) = 1$ para todo $x \in \mathbb{R}$, en *a)*, entonces $E[g(X)] = E[Y]$. Es decir,

  $$E[E[Y \mid X]] = E[Y],$$

  a lo cual se le conoce como la *propiedad de la doble esperanza.*

- Sea $B \subseteq \mathbb{R}$ y tomemos $h(x) = 1_B(x)$, $x \in \mathbb{R}$ en *a)*. Entonces $E[g(X)1_B(X)] = E[Y1_B(X)]$, es decir,

  $$E[E[Y \mid X]1_B(X)] = E[Y1_B(X)].$$

De aquí,

$$E[Y 1_B(X)] = \sum_x E[Y \mid X = x] 1_B(x) f_X(x). \qquad (1.2)$$

- Tomando $B = \mathbb{R}$ en (1.2), obtenemos

$$E[Y] = \sum_x E[Y \mid X = x] f_X(x).$$

  Esta expresión constituye otra forma de calcular esperanzas, la cual es muy práctica en muchas situaciones.

- De la parte *b)* del teorema concluímos,

$$E[XY \mid X] = XE[Y \mid X] \quad \text{c.s.}$$

  De hecho, si $f_1, f_2 : \mathbb{R} \to \mathbb{R}$ son funciones arbitrarias, es fácil ver que

$$E[f_1(X) f_2(Y) \mid X] = f_1(X) E[f_2(Y) \mid X] \quad \text{c.s.}$$

En los siguientes dos capítulos estaremos usando la esperanza condicional de una variable aleatoria dado un vector aleatorio. La definición en este caso, así como sus propiedades, se pueden extender de las anteriores.

Definición 1.20 *Sea Y una variable discreta y $\mathbf{X} = (X_1, ..., X_n)$ un vector aleatorio discreto. Se define el* ***valor esperado de Y dado $\mathbf{X} = \mathbf{x}$*** *como*

$$E[Y \mid \mathbf{X} = \mathbf{x}] := \sum_y y f_{Y|\mathbf{X}}(y \mid \mathbf{x}),$$

donde $f_{Y|\mathbf{X}}(y \mid \mathbf{x}) := \frac{P[Y=y, \mathbf{X}=\mathbf{x}]}{P[\mathbf{X}=\mathbf{x}]}$.

Del mismo modo, se tiene la siguiente definición:

Definición 1.21 *Sea $g : \mathbb{R}^n \to \mathbb{R}$ una función dada por*

$$g(x_1, ..., x_n) := E[Y \mid X_1 = x_1, ..., X_n = x_n].$$

Se define la ***esperanza condicional de Y dado $\mathbf{X}$*** *como*

$$E[Y \mid \mathbf{X}] := g(\mathbf{X}).$$

Considerando estas definiciones y el Teorema 1.19, llegamos a las siguientes propiedades:

Teorema 1.22 *Sean* $Y, Y_1, ..., Y_k$ *variables aleatorias discretas,* $\mathbf{X} = (X_1, ..., X_n)$ *un vector aleatorio discreto,* $h : \mathbb{R}^n \to \mathbb{R}$ *y* $f : \mathbb{R} \to \mathbb{R}$ *funciones arbitrarias,* $H : \mathbb{R}^n \to \mathbb{R}^n$ *una función inyectiva, y* $c_1, ..., c_k \in \mathbb{R}$. *Entonces*

a) $E[E[Y \mid \mathbf{X}]h(\mathbf{X})] = E[Yh(\mathbf{X})]$;

b) $E[f(Y)h(\mathbf{X}) \mid \mathbf{X}] = h(\mathbf{X})E[f(Y) \mid \mathbf{X}] \quad c.s.$;

c) $E[Y \mid H(\mathbf{X})] = E[Y \mid \mathbf{X}] \quad c.s$;

d) $E[c_1Y_1 + ... + c_kY_k \mid \mathbf{X}] = \sum_{i=1}^{k} c_iE[Y_i \mid \mathbf{X}]$.

Teorema 1.23 *Sea* X *una variable aleatoria y* $\mathbf{Y_1}, \mathbf{Y_2}$ *vectores aleatorios discretos. Entonces*

$$E[E[X \mid \mathbf{Y_1}, \mathbf{Y_2}] \mid \mathbf{Y_1}] = E[X \mid \mathbf{Y_1}] \quad c.s.$$

Demostración. Haremos la prueba para el caso de vectores $\mathbf{Y_1}, \mathbf{Y_2}$ 1-dimensionales. El caso general se sigue de manera análoga. Fijemos y_1 y observemos que $P[(Y_1, Y_2) = (y', y'') \mid Y_1 = y_1] = 0$ para todo $y' \neq y_1$. Ahora, sea

$$g(y', y'') = E[X \mid Y_1 = y', Y_2 = y''].$$

Entonces

$$\begin{aligned}
E[E[X \mid Y_1, Y_2] \mid Y_1 = y_1] &= \sum_{y', y''} g(y', y'')P[Y_1 = y', Y_2 = y'' \mid Y_1 = y_1] \\
&= \sum_{y''} g(y_1, y'')P[Y_1 = y_1, Y_2 = y'' \mid Y_1 = y_1] \\
&= \sum_{y''}\sum_{x} xP[X = x \mid Y_1 = y_1, Y_2 = y'']P[Y_1 = y_1, Y_2 = y'' \mid Y_1 = y_1] \\
&= \sum_{y''}\sum_{x} x\frac{P[X = x, Y_1 = y_1, Y_2 = y'']}{P[Y_1 = y_1]} \\
&= \sum_{x} xP[X = x \mid Y_1 = y_1] \\
&= E[X \mid Y_1 = y_1].
\end{aligned}$$

Puesto que y_1 es arbitrario, se tiene el resultado. ■

1.5 Procesos Estocásticos

Intuitivamente un proceso estocástico es una colección de variables aleatorias que puede ser usada como modelo matemático de los resultados de una serie de fenómenos aleatorios. Esto ha permitido que un buen número de sus aplicaciones se presenten en diversas áreas del conocimiento.

En esta sección introducimos los elementos básicos de la teoría de procesos estocásticos y presentamos una clase particular de ellos, los llamados cadenas de Markov, lo cual retomaremos en los capítulos siguientes.

Definición 1.24 *Un* ***proceso estocástico*** *es una familia* $\mathbf{X} = \{X(t) : t \in T\}$ *de variables aleatorias definidas en* $(\Omega, \mathcal{F}, P)$*, donde* $T \subseteq \mathbb{R}$ *es el conjunto de parámetros.*

El hecho de que $X(t)$ sea una variable aleatoria implica que el proceso estocástico se puede ver como una colección de funciones reales definidas sobre $T \times \Omega$, es decir, $X(\cdot,\cdot) : T \times \Omega \to \mathbb{R}$ las cuales satisfacen:

a) Para cada $t \in T$ se tiene que $X(t,\cdot)$ es una variable aleatoria.

b) Para cada $\omega \in \Omega$, la función $X(\cdot,\omega)$ es una función real.

A la función $X(\cdot,\omega) : T \to \mathbb{R}$ le llamaremos *realización* o *trayectoria del proceso estocástico.*

Es común pensar que el índice t representa el tiempo; así $X(t)$ representará la posición o estado del proceso al tiempo t. Sin embargo, es importante notar que el conjunto de parámetros no necesariamente se relaciona con el tiempo físico de ocurrencia de los eventos modelados por la colección de variables aleatorias.

Cuando el conjunto de parámetros T es numerable, se dice que $\mathbf{X} = \{X(t) : t \in T\}$ es un *proceso con parámetro* (o *tiempo*) *discreto.* En este caso, y sin pérdida de generalidad, tomamos $T = \{0, 1, ...\}$ de tal manera que el proceso estocástico $\mathbf{X}$ toma la forma $\{X_0, X_1, ...\}$.

Definición 1.25 *El **espacio de estados** S del proceso $\mathbf{X} = \{X(t) : t \in T\}$ es el conjunto de todos los posibles valores que pueden tomar las variables aleatorias $X(t)$ para cada $t \in T$.*

A cada elemento en S se le llama *estado del proceso.* Si S es finito o numerable, decimos que $\mathbf{X}$ *es un proceso discreto* y si S es un intervalo decimos que *el proceso es continuo.*

- En este trabajo solo consideraremos procesos estocásticos discretos a tiempo discreto. En este caso el proceso $\mathbf{X}$ será una sucesión de variables aleatorias discretas definidas en el mismo espacio de probabilidad.

A continuación presentamos una clase particular de este tipo de procesos.

1.5.1 Cadenas de Markov

La teoría de cadenas de Markov es muy extensa. Aquí solo presentaremos algunas nociones básicas acerca de esta clase de procesos que posteriormente utilizaremos en el análisis de martingalas, en los capítulos siguientes.

Definición 1.26 *Sea $\mathbf{X} = \{X_n : n = 0, 1, ...\}$ un proceso estocástico discreto. Decimos que $\mathbf{X}$ es una **cadena de Markov** si*

$$P[X_{n+1} = j \mid X_n = i, X_{n-1} = x_{n-1}, ..., X_0 = x_0] = P[X_{n+1} = j \mid X_n = i], \tag{1.3}$$

para todos los estados $x_0, ..., x_{n-1}, i, j \in S$ y todo $n \geq 0$.

A la relación (1.3) se le conoce como *propiedad de Markov* y tiene la siguiente interpretación: la distribución condicional de cualquier estado futuro X_{n+1} dado los estados pasados $X_0, ..., X_{n-1}$ y el estado presente X_n, es independiente de los estados pasados y solo depende del estado presente.

Las probabilidades condicionales

$$p_{ij}(n, n+1) := P[X_{n+1} = j \mid X_n = i]$$

son llamadas *probabilidades de transición* o *probabilidades en un paso*, y representan la probabilidad de pasar al estado j en el tiempo $n+1$, estando en el estado i al tiempo n; o bien, la probabilidad de pasar del estado i al estado j en el paso del tiempo n al $n+1$.

Si las probabilidades de transición no dependen del tiempo n, decimos que la cadena de Markov $\mathbf{X} = \{X_n : n = 0, 1, ...\}$ es una *cadena homogénea* o *estacionaria en el tiempo* y escribimos

$$p_{ij} := P[X_1 = j \mid X_0 = i] = P[X_{n+1} = j \mid X_n = i] \quad n = 1, ...; i, j \in S.$$

- En adelante, nos restringiremos al estudio de cadenas homogéneas.

Definición 1.27 *Sea $\mathbf{X} = \{X_n; n \geq 0\}$ una cadena de Markov. La matriz $P = (p_{ij})$ cuyos elementos son las probabilidades de transición en un paso, se llama* ***matriz de transición*** *de la cadena de Markov.*

Si el espacio de estados de la cadena de Markov es finito, digamos con k elementos, entonces P es una matriz de $k \times k$. En caso contrario, P es una matriz infinita.

Las probabilidades de transición satisfacen las siguientes propiedades:

a) $p_{ij} \geq 0$;

b) $\sum_{j \in S} p_{ij} = 1$ para todo $i \in S$.

Esta última propiedad significa que la suma de cada renglón de la matriz P es 1. Una matriz que satisface esta propiedad se llama *matriz estocástica*. Si además, la suma de las columnas también es 1, la matriz es llamada *matriz doblemente estocástica*.

Un ejemplo clásico de cadena de Markov es el siguiente:

Ejemplo 1.28 *Sean $Z_1, Z_2, ...$ variables aleatorias independientes tales que $P[Z_i = 1] = p$ y $P[Z_i = -1] = q = 1 - p$ para $i = 1, 2, ...$, y sea*

$$S_n = S_0 + Z_1 + ... + Z_n$$

donde S_0 es constante. Observemos que

$$\begin{aligned}
&P[S_{n+1} = y \mid S_n = x, S_{n-1} = s_{n-1}, ..., S_0 = s_0] = \\
&= \frac{P[S_{n+1} = y, S_n = x, S_{n-1} = s_{n-1}, ..., S_0 = s_0]}{P[S_n = x, S_{n-1} = s_{n-1}, ..., S_0 = s_0]} \\
&= \frac{P[Z_{n+1} = y - x, Z_n = x - s_{n-1}, ..., S_0 = s_0]}{P[Z_n = x - s_{n-1}, ..., S_0 = s_0]} \\
&= \frac{P[Z_{n+1} = y - x]P[Z_n = x - s_{n-1}, ..., S_0 = s_0]}{P[Z_n = x - s_{n-1}, ..., S_0 = s_0]} \\
&= P[Z_{n+1} = y - x].
\end{aligned}$$

Por otro lado,

$$\begin{aligned}
P[S_{n+1} = y \mid S_n = x] &= \frac{P[S_{n+1} = y, S_n = x]}{P[S_n = x]} \\
&= \frac{P[Z_{n+1} = y - x, S_n = x]}{P[S_n = x]} \\
&= P[Z_{n+1} = y - x],
\end{aligned}$$

de aquí que $\{S_n : n = 0, 1, ...\}$ es una cadena de Markov con probabilidad de transición

$$p_{xy} = P[Z_1 = y - x].$$

Ejemplo 1.29 *Consideremos una población de N bolas las cuales pueden ser de uno de dos colores: negra ó blanca, y sea X_n la variable aleatoria que mide el número de bolas negras en esta población. Supongamos que de esta población, podemos obtener otra del mismo tamaño, de la siguiente manera: tomamos N bolas nuevas, y el color de cada una de ellas se determina de acuerdo a los resultados que se obtienen al extraer con reemplazamiento N bolas de la poblacion anterior, y denotemos por X_{n+1} a la variable que mide el número de bolas negras presentes en esta nueva población. Entonces claramente $\{X_n : n = 0, 1, ...\}$ es una cadena de Markov con espacio de estados*

$\{0, ..., N\}$. *Observemos que si al tiempo n se tienen x bolas negras, entonces*

$$P[X_{n+1} = y \mid X_n = x] = \binom{N}{y} \left(\frac{x}{N}\right)^y \left(1 - \frac{x}{N}\right)^{N-y}$$

pues se tiene probabilidad $\frac{x}{N}$ de obtener una bola negra.

Definición 1.30 *La* ***probabilidad de transición en*** n ***pasos***, $n \geq 0$, *para una cadena de Markov homogénea* $\mathbf{X}$ *se define por*

$$p_{ij}^n := P[X_{m+n} = j \mid X_m = i] \quad \textit{para todo } m \geq 0;\ i, j \in S.$$

Observemos que

$$p_{ij}^n = P[X_{m+n} = j \mid X_m = i] = P[X_n = j \mid X_0 = i], \quad \forall m \geq 0;\ i, j \in S.$$

Para $n = 1$, se tiene que

$$p_{ij}^1 = p_{ij}$$

y para $n = 0$,

$$p_{ij}^0 = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases} .$$

Definición 1.31 *Sea* $\mathbf{X} = \{X_n : n = 0, 1, ...\}$ *una cadena de Markov. Se define la* ***matriz de transición en*** n ***pasos***, *como la matriz* $P^{(n)} := (p_{ij}^n)$, *donde* p_{ij}^n *son las probabilidades de transición en* n *pasos.*

La probabilidad de transición en n pasos, $n \geq 0$, satisface:

a) $p_{ij}^n \geq 0$ para todo $i, j \in S$;

b) $\sum_{j \in S} p_{ij}^n = 1$ para todo $i \in S$.

De *b)* se sigue que $P^{(n)}$ es una matriz estocástica.

Teorema 1.32 *Sea* $\mathbf{X} = \{X_n; n \geq 0\}$ *una cadena de Markov. Entonces*

$$p_{ij}^{n+m} = \sum_{k \in S} p_{ik}^{n} p_{kj}^{m}, \quad \forall n, m \geq 0;\ i, j \in S. \tag{1.4}$$

La ecuación anterior es llamada *ecuación de Chapman-Kolmogorov* y nos proporciona una manera práctica de calcular las probabilidades de transición en n pasos. Además podemos observar que es equivalente a la ecuación matricial

$$P^{(n+m)} = P^{(n)} P^{(m)}$$

en el caso finito.

De (1.4), se sigue que podemos expresar la probabilidad de transición en n pasos como

$$p_{ij}^{n} = \sum_{k \in S} p_{ik}^{m} p_{kj}^{n-m} \qquad \text{para todo } 0 \leq m \leq n,$$

o equivalentemente

$$P^{(n)} = P^{(m)} P^{(n-m)} = P^{(n-m)} P^{(m)}.$$

Por otro lado, observemos que $P^{(n)} = P^n$. Esto significa que la matriz de transición en n pasos se puede calcular como la $n-$ésima potencia de la matriz P, siempre y cuando la cadena de Markov tenga un número finito de estados.

Capítulo 2

Martingalas

2.1 Introducción

En matemáticas, las martingalas son una clase de procesos estocásticos cuyo surgimiento fue motivado al modelar juegos de azar justos. Sin embargo, el desarrollo de la teoría de martingalas ha ido más allá del análisis de juegos justos y durante el transcurso de los años ha alcanzado grandes aplicaciones en otras áreas, así como dentro de la matemática misma (ver, e.g., [4]).

En este capítulo presentamos la teoría de martingalas para procesos discretos. A manera de motivación, en la Sección 2.2 se hace un análisis de estrategias de apuesta para juegos, y después en la Sección 2.3 se introduce formalmente la definición de martingala así como sus propiedades básicas. Finalmente, en la Sección 2.4 presentamos la relación entre cadenas de Markov y martingalas.

2.2 Estrategias de Apuestas

Existen muchos tipos de juegos de apuestas y por lo tanto un gran número de estrategias que utilizan los jugadores para apostar. Nos centraremos en una clase particular de juegos a los cuales se les conoce en la literatura como *juegos de apuestas sucesivas.* Estos consisten en apostar sucesivamente cierta cantidad, la cual se fija mediante cierta regla, y de tal manera que en cada juego se puede perder o ganar la cantidad apostada. Ejemplos de este tipo de juegos pueden ser apostar a cierto color en la ruleta sucesivamente durante varios juegos; lanzar sucesivamente una moneda al aire y estar apostando a una de las caras u otro juego de azar semejante.

Consideremos un juego de apuestas sucesivas y supongamos que p es la probabilidad de ganar en cualesquiera de las jugadas, donde $p \in (0,1)$.

Sea Z_n la variable aleatoria que toma el valor 1 si se gana en la $n-$ésima jugada y -1 si se pierde, es decir, estas variables tienen distribución

$$p = P[Z_n = 1], \quad P[Z_n = -1] = 1 - p \quad \text{para cada } n \in \mathbb{N};$$

y supongamos que $Z_0, Z_1, ...$ son independientes (como es el caso de lanzamiento de monedas o el juego de ruleta), donde Z_0 denotará la fortuna inicial del jugador, la cual será una cantidad conocida.

Podemos definir una estrategia de apuestas como sigue:

Definición 2.1 *Una* ***estrategia de apuestas*** *es una sucesión de funciones* $\{\xi_n : n = 1, 2, ...\}$, *donde*

$$\xi_{n+1} : \{Z_0\} \times \{-1, 1\}^n \to [0, \infty),$$

tales que $\xi_{n+1}(Z_0, ..., Z_n)$ *es la cantidad a apostar en la* $(n+1)-$*ésima jugada para* $n = 0, 1, ...$

Observemos que

$$Z_{n+1} \cdot \xi_{n+1}(Z_0, ..., Z_n)$$

representa la ganancia (ó pérdida) en el $(n+1)-$ésimo juego. Además, si S_{n+1} es la fortuna acumulada hasta el tiempo $n+1$, entonces

$$S_{n+1} = S_0 + \sum_{i=1}^{n+1} Z_i \cdot \xi_i(Z_0, ..., Z_{i-1}) \quad \text{para } n = 0, 1, ..., \tag{2.1}$$

o bien,

$$S_{n+1} = S_n + Z_{n+1} \cdot \xi_{n+1}(Z_0, ..., Z_n) \quad \text{para } n = 0, 1, ..., \tag{2.2}$$

donde $S_0 = Z_0$, es decir S_0 es la fortuna inicial del jugador.

A continuación presentamos algunos ejemplos de estrategias de apuestas.

Ejemplo 2.2 *Sea c un número real positivo y definamos*

$$\xi_{n+1}(Z_0,...,Z_n) = \begin{cases} c & si\ S_n \geq c \\ 0 & si\ S_n < c \end{cases}$$

para cada $n \geq 0$. *Entonces* $\{\xi_n : n = 1,2,...\}$ *es una estrategia de apuestas. Además, en este caso*

$$S_n = S_0 + c\sum_{i=1}^{n} Z_i 1_{[S_{i-1}\geq c]} \quad para\ n = 1,2,.... \tag{2.3}$$

Observemos que el aplicar esta estrategia implica que en cada jugada siempre se apostará una cantidad constante c, cuando la fortuna sea suficiente, independientemente de los resultados obtenidos en las jugadas anteriores.

Ejemplo 2.3 *Sea* $\alpha \in (0,1]$ *y definamos* ξ_n *como*

$$\xi_{n+1}(Z_0,...,Z_n) = \alpha \cdot S_n \quad para\ n = 0,1,....$$

Se tiene que $\{\xi_n : n = 1,2,...\}$ *define una estrategia de apuestas. Notemos que cuando* $\alpha = 1$, *el jugador apuesta toda su fortuna acumulada en cada jugada mientras que para* $\alpha \in (0,1)$ *solo apuesta un porcentaje de ella.*

En muchos de los juegos de apuestas, el criterio para decidir la cantidad a apostar se basa solamente en el resultado del último juego. En este caso, la estrategia se representa por funciones

$$\xi_{n+1} : \{-1,1\} \to [0,\infty) \quad \text{para } n = 0,1,....$$

Un ejemplo de este tipo de estrategia de apuestas es el siguiente:

Ejemplo 2.4 *En primer lugar se determina la cantidad que se desea ganar, la cual denotamos por C. En el primer juego se procede a apostar precisamente esta cantidad. Si se gana, el ciclo ha terminado; si se pierde, se apuesta el doble nuevamente al mismo resultado en la siguiente jugada. Si se gana se habrá ganado en esa jugada* $2C$, *y como en la anterior se perdió* C *la ganancia neta del jugador será de* C. *Si se ha perdido en la segunda jugada se procede a apostar* $4C$ *nuevamente al mismo resultado (el doble de lo perdido en la jugada anterior). Si se gana en esa jugada (la tercera), se* *habrá ganado entonces $4C$, y como en las jugadas anteriores se ha perdido C y $2C$, la ganacia neta es de C. Así sucesivamente se continúa jugando hasta ganar una sola vez, con lo que la ganancia neta será precisamente C. En resumen, la estrategia puede describirse como sigue: se apuesta en un principio C y cada vez que se pierda se apuesta el doble de la cantidad perdida en la jugada anterior; se procede de esta manera hasta ganar una sola vez. De aquí, podemos representar esta estrategia como*

$$\xi_{n+1}(Z_n) = \begin{cases} 2^n C & \text{si } Z_n = -1 \\ 0 & \text{si } Z_n = 1 \end{cases}$$

para $n = 1, 2, ...$ y $\xi_1 = C$.
Notemos que si T es la variable aleatoria que representa el número de juegos que se realizan hasta ganar por primera vez, y suponiendo que $P[Z_n = 1] = \frac{1}{2}$ para cada n, entonces $P[T < \infty] = 1$, pues bajo estas condiciones T sigue una distribución geométrica con parámetro $\frac{1}{2}$. Lo anterior significa que aplicando esta estrategia, se ganará con probabilidad 1 en alguna ocasión.

Es claro que el desarrollo del juego depende fuertemente del valor de $p = P[Z_i = 1]$, $i = 1, 2, ...$ Este hecho lo establece el siguiente resultado.

Proposición 2.5 *Sea $\{\xi_n : n = 1, 2, ...\}$ una estrategia arbitraria, S_n definido por (2.2) y $p = P[Z_n = 1]$ para todo $n \in \mathbb{N}$,*

a) Si $p = \frac{1}{2}$, entonces $E[S_{n+1} \mid S_n, ..., S_0] = S_n$ c.s.;

b) Si $p \geq \frac{1}{2}$, entonces $E[S_{n+1} \mid S_n, ..., S_0] \geq S_n$ c.s.;

c) Si $p \leq \frac{1}{2}$, entonces $E[S_{n+1} \mid S_n, ..., S_0] \leq S_n$ c.s..

Demostración. Observemos que la relación $H(Z_0, ..., Z_n) := (S_0, ..., S_n)$ es uno a uno, donde S_n es la fortuna del jugador acumulada hasta la $n-$ésima jugada, $n = 0, 1, ...$ (ver (2.1)). Por lo tanto, aplicando el Teorema 1.22 *c)*,

$$E[S_{n+1} \mid Z_n, ..., Z_0] = E[S_{n+1} \mid S_n, ..., S_0] \quad \text{c.s..} \tag{2.4}$$

Por otro lado, de (2.2), el Teorema 1.22 *b)*, y el hecho de que las variables Z_i's son independientes, se tiene

$$\begin{aligned} E[S_{n+1} \mid Z_n, ..., Z_0] &= E[S_n + Z_{n+1} \cdot \xi_{n+1}(Z_0, ..., Z_n) \mid Z_n, ..., Z_0] \qquad (2.5) \\ &= E[S_n \mid Z_n, ..., Z_0] + E[Z_{n+1} \cdot \xi_{n+1}(Z_0, ..., Z_n) \mid Z_n, ..., Z_0] \\ &= S_n + \xi_{n+1}(Z_0, ..., Z_n) E[Z_{n+1} \mid Z_n, ..., Z_0] \\ &= S_n + \xi_{n+1}(Z_0, ..., Z_n) E[Z_{n+1}] \end{aligned}$$

Si $p = \frac{1}{2}$ entonces $E[Z_i] = 0$ para todo $i \in \mathbb{N}$. Por lo tanto, la parte *a)* se sigue de (2.4) y (2.5). Similarmente, si $p \leq \frac{1}{2}$, entonces $E[Z_i] = 2p - 1 \leq 0$ para cada i. Ahora, como las variables aleatorias ξ_i's son no negativas, (2.4) y (2.5) implican la parte *b)*. La parte *c)* se sigue análogamente a la parte *b)*. ■

Observemos que de la Proposición 2.5 y del Teorema 1.22 *a)*, obtenemos:

a) Si $p = \frac{1}{2}$, entonces $E[S_{n+1} - S_n] = 0$ para $n = 0, 1, ...$, lo cual significa que el jugador no espera incrementar ni disminuir su fortuna durante el juego. En tal caso decimos que el juego es *un juego justo.*

b) $E[S_{n+1} - S_n] \geq 0$ para $n = 0, 1, ...$ cuando $p \geq \frac{1}{2}$, es decir, el juego es favorable al jugador pues se espera que obtenga un incremento de su fortuna en cada jugada.

c) En cambio, si $p \leq \frac{1}{2}$, el juego será desfavorable al jugador ya que $E[S_{n+1} - S_n] \leq 0$ para cada $n = 0, 1, ...$

Además, observemos que los resultados anteriores se obtienen independientemente de la estrategia que se utiliza al jugar.

2.3 Martingalas

De acuerdo a la Proposición 2.5, introducimos la siguiente definición:

Definición 2.6 *Sean* $\mathbf{X} = \{X_n : n = 0, 1, ...\}$ *y* $\mathbf{Y} = \{Y_n : n = 0, 1, ...\}$ *dos procesos estocásticos discretos. Decimos que*

a) $\mathbf{Y}$ *es una* ***martingala*** *con respecto a la sucesión* $\mathbf{X}$ *si para todo* $n \geq 0$,

$$E[|Y_n|] < \infty \quad y \quad E[Y_{n+1} \mid X_n, ..., X_0] = Y_n \quad c.s;$$

b) $\mathbf{Y}$ *es una* ***submartingala*** *con respecto a la sucesión* $\mathbf{X}$ *si para todo* $n \geq 0$,

$$E[|Y_n|] < \infty \quad y \quad E[Y_{n+1} \mid X_n, ..., X_0] \geq Y_n \quad c.s;$$

c) $\mathbf{Y}$ *es una* ***supermartingala*** *con respecto a la sucesión* $\mathbf{X}$ *si para todo* $n \geq 0$,

$$E[|Y_n|] < \infty \quad y \quad E[Y_{n+1} \mid X_n, ..., X_0] \leq Y_n \quad c.s.$$

Observemos que $\mathbf{Y}$ es una martingala respecto a $\mathbf{X}$ si y sólo si $\mathbf{Y}$ es submartingala y supermartingala respecto a $\mathbf{X}$.

De la Proposición 2.5 y la Definición 2.6 tenemos que si $p = \frac{1}{2}$, entonces $\{S_n : n = 0, 1, ...\}$ (ver (2.2)) es una martingala respecto a si misma. Similarmente, si $p \geq \frac{1}{2}$ entonces $\{S_n : n = 0, 1, ...\}$ es una submartingala y en el caso de ser $p \leq \frac{1}{2}$ se tiene que $\{S_n : n = 0, 1, ...\}$ es una supermartingala respecto a si misma. Con esto, podemos decir que *una martingala es un modelo de un juego justo.*

A continuación presentamos otros ejemplos que ilustran la Definición 2.6.

Ejemplo 2.7 *Sean* $\eta_0, \eta_1, ...$ *una sucesión de variables aleatorias discretas e independientes con* $E[|\eta_n|] < \infty$ *para* $n = 0, 1, ...,$ *y sea*

$$Y_n = \eta_0 + ... + \eta_n.$$

Claramente, $E[|Y_n|] < \infty$.

Por otro lado, observemos que $Y_{n+1} = Y_n + \eta_{n+1}$. *Ahora, aplicando el Teorema 1.22 b) y por la independencia de las variables aleatorias* η_i*'s se tiene*

$$\begin{aligned} E[Y_{n+1} \mid \eta_n, ..., \eta_0] &= E[Y_n \mid \eta_n, ..., \eta_0] + E[\eta_{n+1} \mid \eta_n, ..., \eta_0] \\ &= Y_n + E[\eta_{n+1}] \end{aligned}$$

Si $E[\eta_n] = 0$ *para todo* n, *entonces* $\{Y_n : n = 0, 1, ...\}$ *es una martingala respecto a* $\{\eta_n : n = 0, 1, ...\}$. *Si ahora tomamos* $E[\eta_n] \geq 0$ *para cada* $n = 0, 1, ...,$ *entonces es claro que* $\{Y_n : n = 0, 1, ...\}$ *es una submartingala respecto a* $\{\eta_n : n = 0, 1, ...\}$, *y si* $E[\eta_n] \leq 0$ *para todo* n, *entonces* $\{Y_n : n = 0, 1, ...\}$ *es una supermartingala respecto a* $\{\eta_n : n = 0, 1, ...\}$.

Ejemplo 2.8 .

a) Sean $\eta_0, \eta_1, \ldots$ variables aleatorias discretas, independientes, con $E[|\eta_n|] < \infty$, tales que $E[\eta_n] = 0$ y $var[\eta_n] = E[\eta_n^2] = \sigma_n^2 < \infty$ para $n = 0, 1, \ldots$. Definamos

$$S_n = S_0 + \eta_1, \ldots, +\eta_n,$$

donde S_0 es una constante, y sea

$$v_n = \sum_{i=1}^{n} \sigma_i^2$$

la varianza de S_n. Entonces, el proceso $\{Y_n : n = 0, 1, \ldots\}$ donde $Y_n = S_n^2 - v_n$, es una martingala respecto a $\{S_n : n = 0, 1, \ldots\}$. En efecto, primero observemos que $E[S_n^2] < \infty$, $n \geq 0$. Luego,

$$\begin{aligned} E[|Y_n|] &= E[|S_n^2 - v_n|] \\ &\leq E[S_n^2] + v_n < \infty. \end{aligned}$$

Por otra parte tenemos

$$\begin{aligned} Y_{n+1} - Y_n &= (S_n + \eta_{n+1})^2 - S_n^2 - (v_{n+1} - v_n) \\ &= (S_n + \eta_{n+1})^2 - S_n^2 - \sigma_{n+1}^2 \\ &= 2\eta_{n+1} S_n + \eta_{n+1}^2 - \sigma_{n+1}^2 \end{aligned}$$

Además, por el Teorema 1.22 b)

$$E[2\eta_{n+1} S_n \mid S_n, \ldots, S_0] = 2S_n E[\eta_{n+1} \mid S_n, \ldots, S_0] = 2S_n E[\eta_{n+1}] = 0 \quad c.s.$$

pues la independencia de las variables aleatorias η_i's implica que η_{n+1} es independiente de $S_0, \ldots, S_n$. De aquí y aplicando el hecho de que $E[\eta_{n+1}^2] = \sigma_{n+1}^2$, se tiene que

$$E[\eta_{n+1}^2 - \sigma_{n+1}^2 \mid S_n, \ldots, S_0] = 0 \quad c.s.$$

Por tanto

$$\begin{aligned} E[Y_{n+1} - Y_n \mid S_n, \ldots, S_0] &= E[2\eta_{n+1} S_n + \eta_{n+1}^2 - \sigma_{n+1}^2 \mid S_n, \ldots, S_0] \\ &= E[2\eta_{n+1} S_n \mid S_n, \ldots, S_0] + E[\eta_{n+1}^2 - \sigma_{n+1}^2 \mid S_n, \ldots, S_0] \\ &= 0 \quad c.s., \end{aligned}$$

lo cual significa que $\{Y_n : n = 0, 1, \ldots\}$ es una martingala respecto a $\{S_n : n = 0, 1, \ldots\}$.

b) Continuando con el ejemplo anterior, definamos ahora

$$T_n = S_n^2,$$

entonces, por la independencia de η_i's y b) del Teorema 1.22

$$\begin{aligned}E[T_{n+1} \mid \eta_n, ..., \eta_0] &= E[(S_n + \eta_{n+1})^2 \mid \eta_n, ..., \eta_0] \\ &= E[S_n^2 + 2S_n\eta_{n+1} + \eta_{n+1}^2 \mid \eta_n, ..., \eta_1] \\ &= S_n^2 + 2S_nE[\eta_{n+1} \mid \eta_n, ..., \eta_1] + E[\eta_{n+1}^2] \\ &= T_n + E[\eta_{n+1}^2] \geq T_n.\end{aligned}$$

Por lo tanto, el proceso $\{T_n : n = 0, 1, ...\}$ es una submartingala respecto a $\{\eta_n : n = 0, 1, ...\}$.

Ejemplo 2.9 *Sean $\xi_0, \xi_1, ...$ variables aleatorias independientes e idénticamente distribuidas tales que*

$$P[\xi_n = 1] = P[\xi_n = -1] = \frac{1}{2}$$

y definamos

$$S_n = \xi_0 + ... + \xi_n.$$

Entonces $\{Y_n : n = 0, 1, ...\}$, donde $Y_n = S_n^2 - n$, es una martingala respecto a $\{\xi_n : n = 0, 1, ...\}$. En efecto, como

$$|S_n| = |\xi_0 + ... + \xi_n| \leq |\xi_0| + ... + |\xi_n| = n + 1, \text{ c.s.}$$

se sigue que

$$E[|Y_n|] = E[|S_n^2 - n|] \leq E[S_n^2] + n \leq (n+1)^2 + n < \infty.$$

Ahora, puesto que $S_{n+1}^2 = \xi_{n+1}^2 + 2\xi_{n+1}S_n + S_n^2$ y ξ_{n+1} es independiente de $\xi_0, ..., \xi_n$, tenemos

$$\begin{aligned}E[S_{n+1}^2 \mid \xi_n, ..., \xi_0] &= E[\xi_{n+1}^2 \mid \xi_n, ..., \xi_0] + 2E[\xi_{n+1}S_n \mid \xi_n, ..., \xi_0] + E[S_n^2 \mid \xi_n, ..., \xi_0] \\ &= E[\xi_{n+1}^2] + 2S_nE[\xi_{n+1}] + S_n^2 \\ &= 1 + S_n^2,\end{aligned}$$

esto implica que

$$\begin{aligned}E[Y_{n+1} \mid \xi_n, ..., \xi_0] &= E[S_n^2 - n - 1 \mid \xi_n, ..., \xi_0] \\ &= S_n^2 - n = Y_n,\end{aligned}$$

por lo que $\{Y_n : n = 0, 1, ...\}$ es una martingala respecto a $\{\xi_n : n = 0, 1, ...\}$.

A continuación veremos algunos resultados que resumen las principales propiedades de las martingalas.

Proposición 2.10 *Sea* $\mathbf{Y} = \{Y_n : n = 0, 1, ...\}$ *una martingala con respecto a* $\mathbf{X} = \{X_n : n = 0, 1, ...\}$, *entonces:*

a) $E[Y_{n+m} \mid X_n, ..., X_0] = Y_n$ *c.s. para todo* $n, m \geq 0$.

b) $E[Y_n] = E[Y_0]$ *para todo* n.

Demostración.

a) Sean $Z_1 = (X_0, ..., X_n)$ y $Z_2 = (X_{n+1}, ..., X_{n+m-1})$. Por el Teorema 1.23 y el hecho de que $\mathbf{Y}$ es martingala, tenemos

$$\begin{aligned} E[Y_{n+m} \mid X_n, ..., X_0] &= E[E[Y_{n+m} \mid X_{n+m-1}, ...X_n, ..., X_0] \mid X_n, ..., X_0] \\ &= E[Y_{n+m-1} \mid X_n, ..., X_0], \end{aligned}$$

e iterando, se obtiene el resultado.

b) Este resultado se sigue directamente de *a)* y la propiedad de la doble esperanza, pues

$$E[Y_n] = E[E[Y_n \mid X_1]] = E[Y_1].$$

■

Observemos que si $\mathbf{Y}$ es una submartingala con respecto a $\mathbf{X}$, entonces de la demostración de la Proposición 2.10 podemos concluir que $E[Y_n] \geq E[Y_0]$, y similarmente si $\mathbf{Y}$ es una supermartingala con respecto a $\mathbf{X}$, $E[Y_n] \leq E[Y_0]$ para todo n.

Antes de continuar con las propiedades de las martingalas veamos la siguiente definición:

Definición 2.11 *Sea* τ *una variable aleatoria positiva que toma valores en el conjunto* $\{0, 1, ...\} \cup \{\infty\}$. *Decimos que* τ *es un* ***tiempo de paro*** *respecto al proceso* $\mathbf{Y} = \{Y_n : n = 0, 1, ...\}$ *si el evento* $[\tau = n]$ *es determinado por las variables aleatorias* $Y_0, ..., Y_n$.

Intuitivamente podemos decir que una variable aleatoria τ con rango $\{0, 1, ...\} \cup \{\infty\}$ es un tiempo de paro respecto al proceso $\{Y_n : n = 0, 1, ...\}$, si la ocurrencia (ó no ocurrencia) del evento *"paramos al tiempo n"* se puede determinar observando los valores del proceso hasta el tiempo n, es decir, observando los valores de $Y_0, ..., Y_n$

Ejemplo 2.12 *Sea $\{Y_n : n = 0, 1, ...\}$ un proceso discreto y $[a, b] \subset \mathbb{R}$. Entonces*

$$\tau_1 = \min\{n : Y_n \in [a, b]\}$$

es un tiempo de paro. En efecto, es suficiente observar que para cada n, podemos escribir el evento "paramos al tiempo n", $[\tau_1 = n]$, como

$$[\tau_1 = n] = [Y_0 \notin [a, b]] \cap ... \cap [Y_{n-1} \notin [a, b]] \cap [Y_n \in [a, b]].$$

Por lo tanto, la ocurrencia de los eventos del lado derecho de la igualdad la podemos determinar observando el proceso hasta el tiempo n.

En particular, para el proceso $\{S_n : n = 0, 1, ...\}$ donde S_n está dada por (2.2), se tiene que

$$V_y = \min\{n : S_n = y\}, \tag{2.6}$$

es un tiempo de paro para cada $y \in \mathbb{R}$. Además, en este caso V_y representa el primer tiempo en que la fortuna S_n ha alcanzado la cantidad y.

Es por esto que en el problema de apuestas sucesivas, un tiempo de paro es la variable aleatoria que podría representar el momento en que termina el juego, digamos el primer tiempo en que la fortuna acumulada llega a cierta cantidad específica como puede ser la ruina, o más en general, el primer tiempo en que la fortuna tome valores fuera de un rango.

Sea τ un tiempo de paro respecto a un proceso $\{Y_n : n = 0, 1, ...\}$, y definamos la variable aleatoria $\tau \wedge n$ como

$$\tau \wedge n := \min\{\tau, n\}.$$

Al proceso $\{Y_{\tau \wedge n} : n = 0, 1, ...\}$ definido por

$$Y_{\tau \wedge n} := \begin{cases} Y_n & \text{si } n \leq \tau \\ Y_\tau & \text{si } n > \tau \end{cases}$$

se le conoce como *proceso de paro*. Una propiedad importante de este proceso la establece el siguiente resultado.

Proposición 2.13 *Sea τ un tiempo de paro respecto a $\{Y_n : n = 0, 1, ...\}$.*

a) Si $\{Y_n : n = 0, 1, ...\}$ es una martingala, entonces $\{Y_{\tau\wedge n} : n = 0, 1, ...\}$ también es una martingala;

b) Si $\{Y_n : n = 0, 1, ...\}$ es una submartingala, entonces $\{Y_{\tau\wedge n} : n = 0, 1, ...\}$ es una submartingala;

c) Si $\{Y_n : n = 0, 1, ...\}$ es una supermartingala, entonces $\{Y_{\tau\wedge n} : n = 0, 1, ...\}$ es una supermartingala.

Demostración. *Probaremos primero que $\{Y_{\tau\wedge n} : n = 0, 1, ...\}$ es una martingala, es decir, $E[|Y_{\tau\wedge n}|] < \infty$ y $E[Y_{\tau\wedge n} \mid Y_{\tau\wedge(n-1)}, ..., Y_{\tau\wedge 1}] = Y_{\tau\wedge(n-1)}$ c.s.*

Primero observemos que $E[|Y_{\tau\wedge n}|] < \infty$ pues $\{Y_n : n = 0, 1, ...\}$ es martingala. Sea

$$1_n = \begin{cases} 1 & si\, n \leq \tau \\ 0 & si\, n > \tau \end{cases},$$

es decir, 1_n es 1 si no hemos parado después de haber observado $Y_1, ..., Y_{n-1}$. Entonces

$$Y_{\tau\wedge n} = Y_{\tau\wedge(n-1)} + 1_n(Y_n - Y_{n-1}), \tag{2.7}$$

pues si $n \leq \tau$ entonces $Y_{\tau\wedge n} = Y_n$, $Y_{\tau\wedge(n-1)} = Y_{n-1}$ y $1_n = 1$; ahora, si $n > \tau$, entonces $Y_{\tau\wedge(n-1)} = Y_{\tau\wedge n} = Y_\tau$.

Por otro lado, de (2.7), usando el hecho de que $Y_{\tau\wedge(n-1)}$ y 1_n están determinadas por $Y_1, ...Y_{n-1}$ y de que $\{Y_n\}$ es una martingala,

$$\begin{aligned} E[Y_{\tau\wedge n} \mid Y_{n-1}, ..., Y_1] &= E[Y_{\tau\wedge(n-1)} + 1_n(Y_n - Y_{n-1}) \mid Y_{n-1}, ..., Y_1] \\ &= Y_{\tau\wedge(n-1)} + 1_n E[Y_n - Y_{n-1} \mid Y_{n-1}, ..., Y_1] \\ &= Y_{\tau\wedge(n-1)} \end{aligned}$$

Ahora, si conocemos los valores de $Y_1, ...Y_{n-1}$, entonces conocemos también los valores de $Y_{\tau\wedge 1}, ...Y_{\tau\wedge(n-1)}$. Por lo tanto, usando el Teorema 1.23 y la relación (2.7)

$$\begin{aligned} E[Y_{\tau\wedge n} \mid Y_{\tau\wedge(n-1)}, ..., Y_{\tau\wedge 1}] &= E[E[Y_{\tau\wedge n} \mid Y_{\tau\wedge(n-1)}, ..., Y_{\tau\wedge 1}, Y_{n-1}, ..., Y_1] \mid Y_{\tau\wedge(n-1)}, ..., Y_{\tau\wedge 1}] \\ &= E[E[Y_{\tau\wedge n} \mid Y_{n-1}, ..., Y_1] \mid Y_{\tau\wedge(n-1)}, ..., Y_{\tau\wedge 1}] \\ &= E[Y_{\tau\wedge(n-1)} \mid Y_{\tau\wedge(n-1)}, ..., Y_{\tau\wedge 1}] = Y_{\tau\wedge(n-1)}, \end{aligned}$$

de aquí que $\{Y_{\tau\wedge n} : n = 0, 1, ...\}$ es una martingala. Los resultados b) y c) se prueban análogamente a la parte a). ■

La Proposición 2.13 nos dice que es imposible convertir un juego justo en un juego no justo; un juego favorable en uno desfavorable; y uno desfavorable en uno favorable, sin importar el criterio que se haya decidido para parar de jugar.

2.4 Martingalas y Cadenas de Markov

El hecho de que un proceso estocástico sea una martingala nos lleva a obtener conclusiones y resultados interesantes como veremos en el siguiente capítulo. Sin embargo, muchos de los procesos importantes carecen de esta propiedad, como por ejemplo $\{S_n : n = 0, 1, ...\}$ si $p \neq \frac{1}{2}$ (ver Proposición 2.5).

Bajo estas circunstancias, lo que se buscaría es una transformación del proceso cuyo resultado sea una martingala, de tal forma que los resultados que se tienen de este nuevo proceso nos permita obtener conclusiones importanes del proceso original. Es decir, si $\{X_n : n = 0, 1, ...\}$ es un proceso estocástico, nuestro problema es encontrar una función h tal que el proceso $\{h(X_n) : n = 0, 1, ...\}$ resulte ser una martingala.

Desafortunadamente la existencia de dicha transformación solo se puede garantizar para las cadenas de Markov, como lo muestra el siguiente resultado.

Proposición 2.14 *Sea $\mathbf{X} = \{X_n : n = 0, 1, ...\}$ una cadena de Markov homogénea en el tiempo con probabilidades de transición p_{xy}, $x, y \in \{0, 1, ...\}$, y sea h una función del estado x tal que*

$$h(x) = \sum_{y=0}^{\infty} p_{xy} h(y). \tag{2.8}$$

Entonces el proceso $\{M_n : n = 0, 1, ...\}$ donde $M_n := h(X_n)$, es una martingala respecto a $\mathbf{X}$.

Demostración. Este resultado se sigue de la propiedad de Markov, pues

$$\begin{aligned}
E[M_{n+1} \mid X_n = x, X_{n-1} = x_{n-1}, ..., X_0 = x_0] \\
&= E[h(X_{n+1}) \mid X_n = x, X_{n-1}, ..., X_0 = x_0] \\
&= E[h(X_{n+1}) \mid X_n = x] = \sum_{y=0}^{\infty} p_{xy} h(y) = h(x).
\end{aligned}$$

De aquí que

$$E[M_{n+1} \mid X_n, X_{n-1}, ..., X_0] = h(X_n) = M_n.$$

■

Como observamos en la sección anterior, el proceso $\{S_n : n = 0, 1, ...\}$ donde S_n representa la fortuna acumulada durante el juego hasta el tiempo n, no es, en general, una martingala. Sin embargo, como veremos a continuación, podemos transformar el proceso $\{S_n : n = 0, 1, ...\}$ a una martingala en el caso donde $p \neq \frac{1}{2}$, usando la Proposición 2.14. Para esto y por simplificación de cálculos, supondremos que el jugador aplica la estrategia del Ejemplo 2.2, donde el jugador apuesta una unidad en cada jugada..

Ejemplo 2.15 *Consideremos el juego de apuestas sucesivas del Ejemplo 2.2 con $c = 1$. En este caso tenemos*

$$S_{n+1} = S_0 + \sum_{i=1}^{n+1} Z_i = S_n + Z_{n+1} \quad \textit{para } n = 0, 1, ...$$

Se vió en el Ejemplo 1.28 que el proceso $\{S_n : n = 0, 1, ...\}$ es una cadena de Markov con probabilidad de transición

$$p_{xy} = P[Z_1 = y - x]. \tag{2.9}$$

Ahora, tomando

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

donde $p = P[Z_n = 1]$, $n \in \mathbb{N}$ *y aplicando (2.9), se tiene*

$$\begin{aligned}\sum_y p_{xy}\left(\frac{1-p}{p}\right)^y &= \sum_y P[Z_1 = y - x]\left(\frac{1-p}{p}\right)^y \\ &= p\left(\frac{1-p}{p}\right)^{x+1} + (1-p)\left(\frac{1-p}{p}\right)^{x-1} \\ &= \left(\frac{1-p}{p}\right)^x.\end{aligned}$$

Es decir, la función h satisface (2.8), y por lo tanto, el proceso $\{\left(\frac{1-p}{p}\right)^{S_n} : n = 0, 1, ...\}$ *es una martingala respecto a* $\{S_n : n = 0, 1, ...\}$.

Capítulo 3

Teorema de Paro Opcional y Convergencia de Martingalas

3.1 Introducción

Concluimos el trabajo presentando dos resultados clásicos en teoría de martingalas: el Teorema de Paro Opcional para Martingalas Acotadas y el Teorema de Convergencia para Martingalas.

Presentaremos algunos ejemplos en los cuales se pondrá énfasis en la importancia de ambos resultados en las aplicaciones.

3.2 Teorema de Paro Opcional

Teorema 3.1 *Teorema de paro opcional para martingalas acotadas. Sean $\{Y_n : n = 0, 1, ...\}$ una martingala respecto a $\{X_n : n = 0, 1, ...\}$ y τ un tiempo de paro respecto a $\{X_n : n = 0, 1, ...\}$ con $P[\tau < \infty] = 1$. Si existe una constante K tal que $|Y_{\tau \wedge n}| \leq K$ c.s. para todo n, entonces*

$$E[Y_\tau] = E[Y_0].$$

Demostración. De la Proposición 2.13 tenemos que $\{Y_{\tau \wedge n} : n = 0, 1, ...\}$ es una martingala, y por lo tanto, de la Proposición 2.10,

$$E[Y_{\tau \wedge n}] = E[Y_0] \text{ para todo } n.$$

Además, de la condición $|Y_{\tau\wedge n}| \leq K$ c.s. para todo n, tenemos que $|Y_\tau| \leq K$ c.s. por lo tanto

$$\begin{aligned}|E[Y_0] - E[Y_\tau]| &= |E[Y_{\tau\wedge n}] - E[Y_\tau]| \\ &\leq \left|E[(Y_{\tau\wedge n} - Y_\tau)1_{[\tau>n]}]\right| + \left|E[(Y_{\tau\wedge n} - Y_\tau)1_{[\tau\leq n]}]\right| \\ &\leq 2KP[\tau > n].\end{aligned}$$

Tomando límite cuando $n \to \infty$ en esta expresión, y usando la hipótesis $P[\tau < \infty] = 1$, llegamos a que $|E[Y_0] - E[Y_\tau]| = 0$, lo cual implica que $E[Y_0] = E[Y_\tau]$. ■

El hecho de que una martingala modele un juego justo nos hace pensar que nuestra fortuna no se verá incrementada sustancialmente durante el juego, independientemente del criterio que se utilize para dejar de jugar. Observemos que esta propiedad de las martingalas la establece el Teorema 3.1.

Ejemplo 3.2 *Supongamos que en un juego de apuestas sucesivas un jugador sigue la estrategia del Ejemplo 2.2 con $c = 1$, es decir, que en cada turno apuesta siempre una unidad. Sea $p = P[Z_n = 1]$, $n \in \mathbb{N}$ y S_n la fortuna acumulada hasta la $n-$ésima jugada definida en (2.3). Supondremos además que la fortuna inicial del jugador es $S_0 = x$. El objetivo es calcular la probabilidad de que el jugador incremente su fortuna a cierta cantidad "b" antes de que disminuya a cierta cantidad "a". El caso $a = 0$ se le conoce como problema de la ruina del jugador.*

En términos del tiempo de paro V_y definido en (2.6), lo que queremos calcular es $P_x[V_b < V_a]$, donde $x \in (a, b)$, a, b, x números enteros no negativos y $P_x(\cdot) := P[\cdot \mid S_0 = x]$. Este cálculo se deriva de una simple aplicación del Teorema 3.1, considerando el tiempo de paro $V = min\{n > 0 : S_n \notin (a, b)\}$ y usando el hecho de que los eventos $[S_V = b]$ y $[V_b < V_a]$ son equivalentes, implica que

$$P_x[S_V = b] = P_x[V_b < V_a]. \tag{3.1}$$

Para esto, primero veamos que $P_x[V < \infty] = 1$. En efecto, independientemente del valor de la fortuna inicial $S_0 = x$, existe una trayectoria con probabilidad positiva $(s_0, ..., s_{b-a}, ...)$ del proceso $\{S_n : n = 0, 1, ...\}$ tal que $S_{b-a} \notin (a, b)$, lo cual implica que $P[S_{b-a} \notin (a, b)] > 0$. Por lo tanto, existe $k > 0$ tal que $P_x[V \leq k] \geq \gamma > 0$, lo cual implica que $P_x[V > k] \leq 1 - \gamma$.

Ahora aplicando el hecho de que $\{S_n : n = 0, 1, ...\}$ *es una cadena de Markov (ver Ejemplo(2.15)) se puede ver que* $P_x[V > 2k] \leq (1-\gamma)^2$. *En general tendremos (ver, e.g.[5])*

$$P_x[V > nk] \leq (1-\gamma)^n \quad \text{para todo } n \geq 1.$$

Haciendo $n \to \infty$ *en esta expresión llegamos a que* $P_x[V < \infty] = 1$.

Como se mostró en la Proposición 2.5, solo cuando $p = \frac{1}{2}$ *el proceso* $\{S_n : n = 0, 1, ...\}$ *es una martingala. En este sentido, el cálculo de* $P_x[V_b < V_a]$ *lo haremos en dos partes:*

a) Consideremos primero el caso $p = \frac{1}{2}$. *Es claro que* $S_{V\wedge n} \in [a, b]$ *para todo* $n \geq 0$, *y por lo tanto* $\{S_{V\wedge n} : n = 0, 1, ...\}$ *es una martingala acotada. Además, por el hecho de que* $\xi_{n+1}(Z_0, ..., Z_n) = 1$ *(es decir, la fortuna solo puede incrementar o disminuir en una unidad),* $S_V \in \{a, b\}$. *Entonces, del Teorema 3.1 se sigue que*

$$x = E_x[S_0] = E_x[S_V] = aP_x[S_V = a] + bP_x[S_V = b],$$

donde $E_x[\cdot] := E_x[\cdot \mid S_0 = x]$, *es decir,* $E_x[\cdot]$ *es el operador esperanza correspondiente a la probabilidad* $P_x(\cdot)$. *Ahora, como* $P_x[S_V = a] = 1 - P_x[S_V = b]$, *tenemos*

$$x = a + (b-a)P_x[S_V = b].$$

Finalmente, por (3.1) concluimos

$$P_x[V_b < V_a] = \frac{x-a}{b-a}, \quad \forall x \in (a, b).$$

b) Si $p \neq \frac{1}{2}$, $\{S_n : n = 0, 1, ...\}$ *no es una martingala. Sin embargo, por el Ejemplo(2.15),* $\{h(S_n) : n = 0, 1, ...\}$ *es una martingala donde* $h(y) = \left(\frac{1-p}{p}\right)^y$. *Aún mas, como* $S_{V\wedge n} \in [a, b]$ *para todo* $n \geq 0$, *entonces* $\{h(S_{V\wedge n}) : n = 0, 1, ...\}$ *es una martingala acotada, y por lo tanto, del Teorema 3.1*

$$E_x[h(S_V)] = E_x[h(S_0)] = E_x\left[\left(\frac{1-p}{p}\right)^{S_0}\right] = \left(\frac{1-p}{p}\right)^x.$$

Por otro lado, siguiendo argumentos similares al caso $p = \frac{1}{2}$, tenemos

$$\begin{aligned} E_x[h(S_V)] &= h(a)P_x[S_V = a] + h(b)P_x[S_V = b] \\ &= \left(\frac{1-p}{p}\right)^a P_x[S_V = a] + \left(\frac{1-p}{p}\right)^b P_x[S_V = b] \\ &= \left(\frac{1-p}{p}\right)^a + P_x[S_V = b]\left[\left(\frac{1-p}{p}\right)^b - \left(\frac{1-p}{p}\right)^a\right], \end{aligned}$$

por tanto

$$P_x[V_b < V_a] = \frac{\left(\frac{1-p}{p}\right)^x - \left(\frac{1-p}{p}\right)^a}{\left(\frac{1-p}{p}\right)^b - \left(\frac{1-p}{p}\right)^a}, \quad \forall x \in (a,b).$$

3.3 Teorema de Convergencia de Martingalas

Antes de establecer el teorema, primero veamos algunas definiciones y resultados.

Definición 3.3 *Sean $X_1, X_2, ..., X$ variables aleatorias definidas en el mismo espacio de probabilidad. Decimos que X_n* ***converge a*** *X* ***en*** *L^2, si $E[|X_n - X|^2] \to 0$ cuando $n \to \infty$.*

Lema de Fatou *Si $\{Y_n : n = 0, 1, ...\}$ es una sucesión de variables aleatorias tales que $Y_n \geq Y$ c.s. para todo n y alguna variable aleatoria Y tal que $E[|Y|] < \infty$, entonces*

$$E[\liminf Y_n] \leq \liminf E[Y_n].$$

Desigualdad de Doob-Kolmogorov *Si $\{Y_n : n = 1, 2, ...\}$ es una martingala respecto a $\{X_n : n = 1, 2, ...\}$, entonces*

$$P[\max_{1 \leq i \leq n} |Y_i| \geq \varepsilon] \leq \frac{1}{\varepsilon^2} E[Y_n^2] \quad \text{para } \varepsilon > 0.$$

Demostración. Sean

$$\begin{aligned} &A_0 = \Omega, \\ &A_k = [|Y_1| < \varepsilon, ..., |Y_k| < \varepsilon] \quad \text{para } k = 1, 2, ... \\ &B_k = A_{k-1} \cap [|Y_k| \geq \varepsilon], \end{aligned}$$

observemos que $\bigcup_{i=1}^{n} B_i = [\max_{1 \leq i \leq n} |Y_i| \geq \varepsilon]$, y además

$$\begin{aligned} A_k \cup B_k &= [|Y_1| < \varepsilon, ..., |Y_{k-1}| < \varepsilon, |Y_k| < \varepsilon] \cup [|Y_1| < \varepsilon, ..., |Y_{k-1}| < \varepsilon, |Y_k| \geq \varepsilon] \\ &= (A_{k-1} \cap [|Y_k| < \varepsilon]) \cup (A_{k-1} \cap [|Y_k| \geq \varepsilon]) \\ &= A_{k-1}, \quad k = 1, 2, ... \end{aligned}$$

Por lo tanto

$$\begin{aligned} A_k \cup (\bigcup_{i=1}^{k} B_i) &= A_k \cup B_1 \cup ... \cup B_k \\ &= A_{k-1} \cup B_1 \cup ... \cup B_{k-1} \\ &= A_{k-2} \cup B_1 \cup ... \cup B_{k-2} \\ &= ... = A_1 \cup B_1 \\ &= [|Y_1| < \varepsilon] \cup (A_0 \cap [|Y_1| \geq \varepsilon]) \\ &= [|Y_1| < \varepsilon] \cup (\Omega \cap [|Y_1| \geq \varepsilon]) \\ &= [|Y_1| < \varepsilon] \cup [|Y_1| \geq \varepsilon] = \Omega, \end{aligned}$$

además, ésta es una unión ajena. En efecto, es claro que $B_i \cap B_j = \varnothing$ si $i \neq j$, y como $A_k \cap B_i = \varnothing$ para cada $i \leq k$, se sigue que en efecto, para cada $k = 1, 2, ...$ se tiene

$$A_k \cap (\bigcup_{i=1}^{k} B_i) = (A_k \cap B_1) \cup ... \cup (A_k \cap B_k) = \varnothing,$$

Así, de lo anterior tenemos,

$$\begin{aligned} E[Y_n^2] &= E[Y_n^2 \cdot 1_{A_n}] + E[Y_n^2 \cdot 1_{\bigcup_{i=1}^{n} B_i}] \\ &= E[Y_n^2 \cdot 1_{A_n}] + \sum_{i=1}^{n} E[Y_n^2 \cdot 1_{B_i}] \geq \sum_{i=1}^{n} E[Y_n^2 \cdot 1_{B_i}], \end{aligned}$$

por ser $Y_n^2 \geq 0$.

Por otro lado, para cada $i = 1, ..., n$,

$$\begin{aligned} E[Y_n^2 \cdot 1_{B_i}] &= E[(Y_n - Y_i + Y_i)^2 \cdot 1_{B_i}] \\ &= E[(Y_n - Y_i)^2 \cdot 1_{B_i}] \pm 2E[(Y_n - Y_i)Y_i \cdot 1_{B_i}] + E[Y_i^2 \cdot 1_{B_i}]. \end{aligned}$$

Notemos que

$$E[(Y_n - Y_i)^2 \cdot 1_{B_i}] \geq 0, \quad i = 1, ..., n.$$

Además,

$$E[Y_i^2 \cdot 1_{B_i}] \geq E[\varepsilon^2 \cdot 1_{B_i}] = \varepsilon^2 E[1_{B_i}] = \varepsilon^2 P[B_i], \quad i = 1, ..., n$$

Ahora, aplicando *a)* y *b)* del Teorema 1.22

$$\begin{aligned} E[(Y_n - Y_i)Y_i \cdot 1_{B_i}] &= E[E[(Y_n - Y_i)Y_i \cdot 1_{B_i} \mid X_i, ...X_1]] \\ &= E[Y_i \cdot 1_{B_i} E[(Y_n - Y_i) \mid X_i, ...X_1]] \\ &= 0. \end{aligned}$$

Esto último se debe a que por *a)* en el Teorema 2.10, $E[Y_n \mid X_i, ...X_1] = Y_i$ y $E[Y_i \mid X_i, ...X_1] = Y_i$, para $i = 1, ..., n$.

Por lo tanto,

$$E[Y_i^2] \geq \sum_{i=1}^{n} \varepsilon^2 P[B_i].$$

Finalmente como $\bigcup_{i=1}^{n} B_i = [\max_{1 \leq i \leq n} |Y_i| \geq \varepsilon]$, se sigue que

$$E[Y_i^2] \geq \sum_{i=1}^{n} \varepsilon^2 P[B_i] = \varepsilon^2 P[\max_{1 \leq i \leq n} |Y_i| \geq \varepsilon].$$

■

Teorema 3.4 *Teorema de convergencia de martingalas. Si $\{Y_n : n = 1, 2, ...\}$ es una martingala respecto a $\{X_n : n = 1, 2, ...\}$ tal que $E[Y_n^2] < M < \infty$ para algún M y todo n, entonces existe una variable aleatoria Y tal que Y_n converge a Y casi seguramente y en L^2.*

Demostración. Mostremos primero que $\{E[Y_n^2] : n = 1, 2, ...\}$ es una sucesión convergente. Para esto, observemos que

$$E[Y_m(Y_{m+n} - Y_m)] = E[Y_m E[Y_{m+n} - Y_m \mid X_m, ...X_1]] = 0, \quad m, n \geq 1$$

ya que por *a)* de la Proposición 2.10

$$E[Y_{m+n} \mid X_m, ...X_1] = Y_m, \quad \text{c.s..}$$

Además,

$$E[Y_{m+n}^2] = E[Y_m^2] + E[(Y_{m+n} - Y_m)^2], \tag{3.2}$$

pues

$$\begin{aligned} E[Y_m^2] + E[(Y_{m+n} - Y_m)^2] &= E[Y_m^2] + E[Y_{m+n}^2 - 2Y_{m+n}Y_m + Y_m^2] \\ &= E[Y_{m+n}^2] + E[Y_m^2 - 2Y_{m+n}Y_m + Y_m^2] \\ &= E[Y_{m+n}^2] + E[2Y_m^2 - 2Y_{m+n}Y_m] \\ &= E[Y_{m+n}^2] - 2E[Y_{m+n}Y_m - Y_m^2] \\ &= E[Y_{m+n}^2] + 2E[Y_m(Y_{m+n} - Y_m)] \\ &= E[Y_{m+n}^2]. \end{aligned}$$

Por lo tanto,

$$E[Y_{m+n}^2] \geq E[Y_m^2] \quad \text{para cada } n.$$

Se sigue que $\{E[Y_n^2] : n = 1, 2, ...\}$ es una sucesión no decreciente y acotada por M (por hipótesis), por lo tanto $\{E[Y_n^2] : n = 1, 2, ...\}$ converge. Supongamos, sin pérdida de generalidad, que

$$E[Y_n^2] \to M \quad \text{cuando } n \to \infty. \tag{3.3}$$

Enseguida mostraremos que la sucesión $\{Y_n(\omega) : n = 1, 2, ...\}$ converge casi seguramente. Para ésto probaremos que el conjunto de puntos donde la sucesión es de Cauchy tiene probabilidad 1.

Sea

$$C = \{\omega \in \Omega : \{Y_n(\omega)\} \text{ es de Cauchy }\}.$$

Entonces

$$C = \{\omega \in \Omega : \text{para cada } \varepsilon > 0, \exists m(\varepsilon, \omega) \text{ tal que } |Y_{m+i} - Y_{m+j}|(\omega) < \varepsilon \ \forall i, j \geq 1\}.$$

Por la desigualdad del triángulo

$$|Y_{m+i} - Y_{m+j}| \leq |Y_{m+i} - Y_m| + |Y_{m+j} - Y_m|, \quad \text{c.s.}$$

así que

$$\begin{aligned} C &= \{\omega \in \Omega : \text{para cada } \varepsilon > 0, \exists m \text{ tal que } |Y_{m+i} - Y_m|(\omega) < \varepsilon\ \forall i \geq 1\} \\ &= \bigcap_{\varepsilon>0} \bigcup_m \{|Y_{m+i} - Y_m| < \varepsilon \text{ para todo } i \geq 1\}. \end{aligned}$$

Por lo tanto el complemento de C se debe expresar como

$$C^c = \bigcup_{\varepsilon>0} \bigcap_m \{|Y_{m+i} - Y_m| \geq \varepsilon \text{ para algún } i \geq 1\},$$

o bien,

$$C^c = \bigcup_{\varepsilon>0} \bigcap_m A_m(\varepsilon),$$

donde $A_m(\varepsilon) = [|Y_{m+i} - Y_m| \geq \varepsilon \text{ para algún } i \geq 1]$.

Ahora, observemos que

$$A_m(\varepsilon) \subseteq A_m(\varepsilon') \quad \text{si} \quad \varepsilon \geq \varepsilon',$$

por lo que

$$P[C^c] = \lim_{\varepsilon \to 0^+} P[\bigcap_m A_m(\varepsilon)] \leq \lim_{\varepsilon \to 0^+} \lim_{m \to \infty} P[A_m(\varepsilon)].$$

Probaremos que $P[C^c] = 0$. Para esto es suficiente mostrar que $P[A_m(\varepsilon)] \to 0$ cuando $m \to \infty$ para todo $\varepsilon > 0$. Es aquí donde usaremos la Desigualdad de Doob-Kolmogorov.

Para m fijo, definamos la sucesión $\{S_n : n = 1, 2, ...\}$ como $S_n = Y_{m+n} - Y_m$. Observemos que $\{S_n : n = 1, 2, ...\}$ es una martingala respecto a sí misma, ya que por a) en Proposición 2.10

$$\begin{aligned} E[S_{n+1} \mid X_{m+n}, ..., X_1] &= E[Y_{m+n+1} - Y_m \mid X_{m+n}, ..., X_1] \\ &= E[Y_{m+n+1} \mid X_{m+n}, ..., X_1] - E[Y_m \mid X_{m+n}, ..., X_1] \\ &= Y_{m+n} - Y_m = S_n. \end{aligned}$$

De aquí y (1.23) tenemos

$$
\begin{aligned}
E[S_{n+1} \mid S_n, ..., S_1] &= E[E[S_{n+1} \mid X_{m+n}, ..., X_1] \mid S_n, ..., S_1] \\
&= E[S_n \mid S_n, ..., S_1] = S_n.
\end{aligned}
$$

Aplicando entoces la Desigualdad de Doob-Kolmogorov a esta martingala tenemos

$$P[|Y_{m+i} - Y_m| \geq \varepsilon \text{ para algún } 1 \leq i \leq n] \leq \frac{1}{\varepsilon^2} E[(Y_{m+n} - Y_m)^2].$$

Haciendo $n \to \infty$ y usando las ecuaciones (3.2) y (3.3)se tiene

$$P[A_m(\varepsilon)] \leq \frac{1}{\varepsilon^2}(M - E[Y_m^2])$$

y por tanto $P[A_m(\varepsilon)] \to 0$ cuando $m \to \infty$.

Con esto hemos probado que $\{Y_n(\omega) : n = 1, 2, ...\}$ es Cauchy convergente en C, por lo tanto existe una variable aleatoria Y tal que $Y_n \xrightarrow{c.s.} Y$.

Por otra parte, probar que Y_n converge a Y en L^2 se sigue directamente del Lema de Fatou y las ecuaciones (3.2) y (3.3):

$$
\begin{aligned}
E[(Y_n - Y)^2] &= E[\liminf_{m \to \infty} (Y_n - Y_m)^2] \leq \liminf_{m \to \infty} E[(Y_n - Y_m)^2] \\
&= M - E[Y_n^2] \to 0 \text{ cuando } n \to \infty.
\end{aligned}
$$

■

Ejemplo 3.5 *Recordemos el Ejemplo 1.29. Encontramos que existe una cadena de Markov* $\{X_n : n = 0, 1, ...\}$ *que toma valores en* $\{0, 1, ..., N\}$ *con probabilidades de transición dadas por*

$$p_{xy} = \binom{N}{y} \left(\frac{x}{N}\right)^y \left(1 - \frac{x}{N}\right)^{N-y}. \tag{3.4}$$

Observemos que por la propiedad de Markov

$$E[X_{n+1} \mid X_n = x_n, ..., X_0 = x_0] = E[X_{n+1} \mid X_n = x_n],$$

y ademásEntonces por la propiedad de Markov y (3.4)

$$
\begin{aligned}
E[X_{n+1} \mid X_n = x_n] &= \sum_{j=0}^{N} jP[X_{n+1} = j \mid X_n = x_n] = \sum_{j=0}^{N} jp_{x_n j} \\
&= \sum_{j=0}^{N} \binom{N}{j} \left(\frac{x_n}{N}\right)^j \left(1 - \frac{x_n}{N}\right)^{N-j} \\
&= N\left(\frac{x_n}{N}\right) = x_n,
\end{aligned}
$$

pues $\sum_{j=0}^{N} \binom{N}{j} \left(\frac{x_n}{N}\right)^j \left(1 - \frac{x_n}{N}\right)^{N-j}$ *es la expresión para la esperanza de una variable aleatoria con distribución binomial con parámetros* N *y* $\frac{x_n}{N}$. *De aquí que* $\{X_n : n = 0, 1, \ldots\}$ *es una martingala.*

Por otro lado, como $E[X_n^2] \leq N^2 < \infty$, *del Teorema 3.4 se sigue que existe* X *tal que*

$$X_n \xrightarrow{\text{c.s.}} X.$$

Bibliografía

[1] L. Breiman (1986), *Probability*, Addison-Wesley.

[2] Tomasz Bojdecki (1985), *Ciencia, ¿Qué es y para qué sirve una martingala?*, págs. 59-65.

[3] J. L. Doob (1971), *The American Mathematical monthly, What is a Martingale?*, págs. 451-463.

[4] L. Dubins and L.J. Savage (1976), *Inequalities for Stochastic Processes*, Dover, New York.

[5] R. Durrett (1999), *Essentials of Stochastic Processes*, Springer, New York.

[6] W. Feller (1985), Introducción a la Teoría de Probabilidades y sus Aplicaciones, Vol. II, Segunda Edición, LIMUSA, México, D.F..

[7] G.R. Grimmett and D.R. Stirzaker (1992), *Probability and Random Processes*, Oxford Sc. Pub.

[8] S.M.Ross (1983), *Stochastic Processes*, John Wiley & Sons.