

Capítulo 1

Espacio muestral, eventos y medidas de probabilidad

El objetivo de la Teoría de Probabilidad es estudiar o modelar, por medio de la herramienta matemática, el comportamiento de fenómenos o experimentos aleatorios. Un experimento es llamado *aleatorio* cuando no se puede asegurar el resultado, es decir, intrínsecamente existe cierta *incertidumbre* sobre el resultado.

Para llevar a cabo tal *modelación probabilista* se requieren de tres conceptos fundamentales. A saber, *espacio muestral*, *evento* y *medida de probabilidad*. Más un cuarto, basado en los anteriores, llamado *variable aleatoria*, el cual juega un papel primordial en el estudio de experimentos aleatorios. El objetivo del presente capítulo es definir dichos conceptos. Los cuales, formarán el marco teórico de nuestro estudio.

1.1. Espacio muestral y eventos

Definición 1.1.1 *El conjunto de todos los resultados posibles de un experimento aleatorio es llamado espacio muestral, y será denotado por Ω .*

Ejemplos.

- (1) El experimento de lanzar una moneda: $\Omega = \{a, s\}$, donde a denota águila y s denota sello.
- (2) El experimento de lanzar un dado: $\Omega = \{1, 2, 3, 4, 5, 6\}$.
- (3) El experimento de observar el tiempo de vida de un aparato eléctrico: $\Omega = [0, \infty)$.
- (4) El experimento de lanzar dos monedas consecutivamente:
 $\Omega = \{(a, a), (a, s), (s, a), (s, s)\}$.
- (5) Experimento de contar el número de autos que cruzan en determinado intervalo de tiempo, digamos por día, por la caseta de peaje:
 $\Omega = \{0, 1, 2, \dots\}$.

Definición 1.1.2 Un evento es una característica de interés en un experimento aleatorio. Normalmente los denotaremos por las letras A, B, C , etc.

- (1.1) El experimento de lanzar una moneda. El evento “la moneda cae águila”: $A = \{a\}$.
- (2.1) El experimento de lanzar un dado. Nos puede interesar $A =$ “la cara muestra un número impar” ó $B =$ “la cara muestra un número divisible por 3”. Por lo tanto, $A = \{1, 3, 5\}$ y $B = \{3, 6\}$.
- (3.1) El experimento de observar el tiempo de vida de un aparato eléctrico: $A =$ “dura más de un año pero menos de 2”, $B =$ “al menos un año”. Entonces, $A = \{(1, 2)\}$ y $B = \{(1, \infty)\}$.
- (4.1) El experimento de lanzar dos monedas consecutivamente. Puede ser de interés, $A =$ “al menos una cara en los dos lanzamientos” ó $B =$ “dos caras”. Luego, $A = \{(a, a), (a, s), (s, a)\}$ y $B = \{(a, a)\}$.
- (5.1) Experimento de contar el número de autos que cruzan por la caseta de peaje en un día. Supongamos que nos interesa, $A =$ “al menos pasaron 105 carros”, $B =$ “pasaron menos de 150 carros”. Entonces, $A = \{105, 106, \dots\}$ y $B = \{0, 1, \dots, 148, 149\}$.

Nos interesa “medir” o calcular la probabilidad de ciertos conjuntos. Los eventos son precisamente a los conjuntos que podemos calcular su probabilidad. Nos gustaría que las operaciones elementales entre eventos tales como complemento, unión, intersección también fuera un evento. Denotaremos por \mathcal{F} a la clase de todos los eventos.

1.2. Medidas de probabilidad

En la Sección 1.1 definimos dos de los tres elementos principales en la Teoría de Probabilidad. El propósito de la presente sección es introducir el concepto de medida de probabilidad, con lo cual estaremos en posición de iniciar nuestro estudio de *modelación de experimentos aleatorios*.

Para lograr lo anterior, primero vamos a definir de manera rigurosa las propiedades que debe cumplir la clase de todos los eventos. La clase $\mathcal{F} \subset \mathcal{P}(\Omega)$, donde $\mathcal{P}(\Omega)$ denota la clase de todos los subconjuntos de Ω también conocido como *conjunto potencia*. En cursos más avanzados, donde se hace uso de la Teoría de la Medida, se puede ver que \mathcal{F} debe ser una σ -álgebra.

Definición 1.2.1 Sea \mathcal{F} una familia no vacía de subconjuntos de Ω . La familia \mathcal{F} es llamada σ -álgebra si satisface las siguientes propiedades:

- (i) Dado $A \in \mathcal{F}$ se cumple $A^c \in \mathcal{F}$, donde A^c denota el complemento de A en Ω .
- (ii) Sea $(A_n)_{n=1}^{\infty}$ cualquier sucesión de conjuntos en \mathcal{F} . Entonces, $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Observación 1.2.2 No es el propósito del presente curso, pero puede verse que en algunos casos hay subconjuntos de Ω que no pertenecen a \mathcal{F} . En consecuencia, hay subconjuntos de que no podremos medir.

Algunas propiedades de σ -álgebras

1. $\Omega \in \mathcal{F}$, y por lo tanto, $\emptyset \in \mathcal{F}$.
2. Sea \mathcal{F} una σ -álgebra y $(A_n)_{n=1}^{\infty}$ una colección de conjuntos en \mathcal{F} . Entonces, $\bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$.
3. Todo conjunto Ω tiene dos σ -álgebras triviales: la más pequeña $\mathcal{F}_0 = \{\Omega, \emptyset\}$, y la más grande $\mathcal{F}_1 = \mathcal{P}(\Omega)$. Además, si \mathcal{F} es cualquier otra σ -álgebra,

$$\mathcal{F}_0 \subset \mathcal{F} \subset \mathcal{F}_1.$$

4. Sean $A, B \in \mathcal{F}$, entonces $A \setminus B \in \mathcal{F}$. En efecto, basta notar que $A \setminus B = A \cap B^c$.

Ahora sí, ya tenemos los elementos necesarios para definir el concepto de medida de probabilidad.

Definición 1.2.3 Una medida de probabilidad sobre el espacio (Ω, \mathcal{F}) , donde Ω es un conjunto y \mathcal{F} una σ -álgebra sobre Ω , es una función de conjuntos $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ tal que:

(i) $\mathbb{P}(\Omega) = 1$.

(ii) Si $(A_n)_{n=0}^{\infty} \subset \mathcal{F}$ es una sucesión de conjuntos disjuntos por pares ($A_i \cap A_j = \emptyset$, $i \neq j$), entonces

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n). \quad (1.1)$$

A la terna $(\Omega, \mathcal{F}, \mathbb{P})$ le llamaremos *espacio de probabilidad*.

Propiedades de la función de probabilidad

Las siguientes propiedades son consecuencia de la definición de función de probabilidad.

1. Si $A \in \mathcal{F}$, entonces $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$. En particular, $\mathbb{P}(\emptyset) = 0$.
2. Considere los eventos A y B . Entonces,

$$\mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(A \cap B).$$

En efecto, notemos que $A = (A \cap B) \cup (A \cap B^c)$. Entonces,

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c),$$

luego,

$$\mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(A \cap B).$$

Por lo tanto, si $A \subset B$ entonces, $\mathbb{P}(A) \leq \mathbb{P}(B)$.

3. Sean A y B dos eventos cualquiera, entonces

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

Proposición 1.2.4 Sea (A_n) una sucesión de eventos tales que $A_n \subset A_{n+1}$, para todo n . Entonces,

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n). \quad (1.2)$$

Demostración: Notemos que, para cada $k \in \mathbb{N}$,

$$A_k = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \cdots \cup (A_k \setminus A_{k-1}),$$

y

$$\cup_{n=1}^{\infty} A_n = A_1 \cup_{n=1}^{\infty} (A_{n+1} \setminus A_n).$$

Por lo tanto, dado que $\{A_1, A_2 \setminus A_1, A_3 \setminus A_2, \dots\}$ es una sucesión de eventos disjuntos se concluye que

$$\begin{aligned} \mathbb{P}(\cup_{n=1}^{\infty} A_n) &= \mathbb{P}(A_1) + \sum_{n=1}^{\infty} \mathbb{P}(A_{n+1} - A_n) \\ &= \mathbb{P}(A_1) + \sum_{n=1}^{\infty} [\mathbb{P}(A_{n+1}) - \mathbb{P}(A_n)] \\ &= \lim_{n \rightarrow \infty} \left[\mathbb{P}(A_1) + \sum_{k=1}^{n-1} (\mathbb{P}(A_{k+1}) - \mathbb{P}(A_k)) \right] \\ &= \lim_{n \rightarrow \infty} \mathbb{P}(A_n). \end{aligned}$$

□

Proposición 1.2.5 *i)(Subaditividad)* Para cada $n \in \mathbb{N}$, sean A_1, \dots, A_n eventos. Entonces,

$$\mathbb{P}\left(\bigcup_{k=1}^n A_k\right) \leq \sum_{k=1}^n \mathbb{P}(A_k). \quad (1.3)$$

Demostración: Vamos a proceder por inducción sobre n . El resultado es válido para $n = 2$ ya que,

$$\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2) \leq \mathbb{P}(A_1) + \mathbb{P}(A_2).$$

Supongamos que se cumple para n . Entonces,

$$\begin{aligned} \mathbb{P}(\cup_{i=1}^{n+1} A_i) &= \mathbb{P}(\cup_{i=1}^n A_i \cup A_{n+1}) \\ &\leq \mathbb{P}(\cup_{i=1}^n A_i) \mathbb{P}(A_{n+1}) \\ &\leq \sum_{i=1}^n \mathbb{P}(A_i), \end{aligned}$$

la última desigualdad es por hipótesis de inducción. □

ii)(σ -Subaditividad) Sean $(A_n)_{n=1}^{\infty}$ una sucesión eventos. Entonces,

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n). \quad (1.4)$$

Demostración:

1.3. Independencia y probabilidad condicional

El concepto de probabilidad condicional es de suma importancia en la teoría de probabilidad dado que, en muchos casos de modelación se conoce información *a priori* sobre el fenómeno y un modo de “aprovechar” dicha información es por medio de la probabilidad condicional. Veamos un ejemplo para clarificar lo anterior.

Ejemplo 1.3.1 *Supongamos que tenemos una baraja de 52 cartas. Sea A el evento de que primera carta sea un as. Entonces,*

$$\mathbb{P}(A) = \frac{4}{52} = \frac{1}{13}.$$

Supongamos ahora que nos damos cuenta que la última carta es el as de espadas, y denotemos por B tal evento. ¿Cuál es la probabilidad de que la primera carta sea un as? Hay 51 cartas para seleccionar, de las cuales 3 son favorables para nuestro evento de interés. Por lo tanto, la probabilidad que buscamos es $3/51$.

El ejemplo anterior nos motiva a introducir la siguiente

Definición 1.3.2 *Sean A y B dos eventos tales que $\mathbb{P}(B) > 0$. La probabilidad condicional del evento A dado B se define como*

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \quad (1.5)$$

Observación 1.3.3 *Cuando $\mathbb{P}(B) = 0$ la probabilidad condicional $\mathbb{P}(A|B)$ no se define como antes.*

Usaremos ahora la definición de probabilidad condicional en el ejemplo anterior. Nos interesa encontrar

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\frac{1}{52} \cdot \frac{3}{51}}{\frac{1}{52}} = \frac{3}{51}.$$

Intuitivamente, detrás de $\mathbb{P}(A|B)$ está que la ocurrencia del evento B nos provee información sobre la ocurrencia del evento A . Por lo tanto, si deseamos decir que A y B son independientes entonces, la ocurrencia del evento B no debe influir en la ocurrencia del evento A . Más formalmente, los eventos A y B son llamados *independientes* si

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \mathbb{P}(A),$$

en otras palabras,

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Vamos a introducir la definición más general de independencia entre eventos.

Definición 1.3.4 *Una colección de eventos $(A_i)_{i \in I}$ es una colección independiente si para todo subconjunto finito $J \subset I$ se cumple*

$$\mathbb{P}(\cap_{i \in J} A_i) = \prod_{i \in J} \mathbb{P}(A_i).$$

Es importante notar que la colección de eventos puede ser finita o infinita. Además, si $(A_i)_{i \in I}$ son independientes entonces son independientes dos a dos. Sin embargo, la recíproca no es cierta como lo muestra el siguiente

Ejemplo 1.3.5 Sea $\Omega = \{1, 2, 3, 4, \}$ y \mathbb{P} la medida uniforme sobre Ω . Considere $A = \{1, 2\}$, $B = \{1, 3\}$ y $C = \{3, 2\}$. Entonces, A , B y C son independientes dos a dos pero no son independientes.

Ejemplo 1.3.6 Una planta obtiene dos genes (los cuales determinan el color de las flores) de manera independiente, cada uno proviene de una planta progenitora. Si los genes son idénticos, entonces las flores adquieren el color correspondiente. Si los genes son distintos, entonces las flores tienen los dos colores. Los genes de los colores rosa (r), violeta (V) y rojo (R) ocurren en la población con proporciones $a : b : c$, de modo que $a + b + c = 1$. Supongamos que seleccionamos una planta al azar; sea A el evento que sus flores sean al menos parcialmente rosas, y sea B el evento de que sus flores tengan dos colores.

a) Encuentre $\mathbb{P}(A)$ y $\mathbb{P}(B)$.

b) Demuestre que A y B son independientes si $a = 2/3$ y $b = c = 1/3$.

Solución:a) Primero notamos que

$$A = rr \cup rV \cup Vr \cup rR \cup Rr.$$

Luego,

$$\mathbb{P}(rr) = \mathbb{P}(r)\mathbb{P}(r) = a^2,$$

dado que el color rosa ocurre con probabilidad a . Por otro lado,

$$\mathbb{P}(rR) = \mathbb{P}(r)\mathbb{P}(R) = ac = \mathbb{P}(Rr),$$

y

$$\mathbb{P}(rV) = ab = \mathbb{P}(Vr).$$

Por lo tanto,

$$\begin{aligned} \mathbb{P}(A) &= a^2 + 2(ac + ab) \\ &= a^2 + 2a(1 - a) \\ &= 1 - (1 - a)^2. \end{aligned}$$

De manera análoga se obtiene que

$$\mathbb{P}(B) = 2(ab + bc + ca).$$

b) Tarea.

Los siguientes resultados son muy útil para calcular probabilidades de eventos cuando conocemos ciertas probabilidades condicionales.

Teorema 1.3.7 (*Ley de probabilidad total*)

(i) Sean A y B dos eventos tales que $0 < \mathbb{P}(B) < 1$. Entonces,

$$\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^c)\mathbb{P}(B^c).$$

ii) Más generalmente, para cualquier partición $(B_n)_{n \in \Lambda}$ ($\Lambda \subset \mathbb{N}$) de Ω , tal que $\mathbb{P}(B_i) > 0$ para todo $i \in \Lambda$, se cumple

$$\mathbb{P}(A) = \sum_{i \in \Lambda} \mathbb{P}(A|B_i)\mathbb{P}(B_i).$$

Teorema 1.3.8 (de Bayes) Sea $A \subset \cup_{i=1}^n B_i$, y $B_i \cap B_j = \emptyset$ para $i \neq j$, entonces

$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(A|B_j)\mathbb{P}(B_j)}{\sum_i \mathbb{P}(A|B_i)\mathbb{P}(B_i)}, \quad \mathbb{P}(A) > 0.$$

1.4. Ejercicios

1. Demuestre que

$$\mathbb{P}(A \cap B) \geq \mathbb{P}(A) + \mathbb{P}(B) - 1,$$

la desigualdad anterior es conocida como desigualdad de Bonferroni. Generalice dicha desigualdad para el caso de n eventos, donde $n \geq 2$.

2. Sea B un evento tal que $\mathbb{P}(B) > 0$. Demuestre que $P : \mathcal{F} \rightarrow [0, 1]$, definida por

$$P(A) = \mathbb{P}(A|B)$$

en una medida de probabilidad sobre (Ω, \mathcal{F}) .

3. Sea A_n una sucesión de eventos tales que $A_{n+1} \subset A_n$, para cada n . Demuestre que

$$\mathbb{P}(\cap_{n=1}^{\infty} A_n) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n).$$

4. Dar un ejemplo para mostrar que independencia dos a dos no implica independencia.
5. Sea $\Omega = \{1, 2, \dots, p\}$ donde p es un número primo, sea \mathcal{F} la σ -álgebra de todos los subconjuntos de Ω , y $\mathbb{P}(A) = |A|/p$, $A \in \mathcal{F}$. Demuestre que si $A, B \in \mathcal{F}$ son independientes entonces al menos un evento es Ω o \emptyset .
6. Sea $(B_k)_{k=1}^n$ una partición de Ω , i.e., los B 's son disjuntos por pares y $\Omega = \cup_{k=1}^n B_k$. Suponga además que para cada k , $\mathbb{P}(B_k) > 0$. Demuestre que, para cada $A \in \mathcal{F}$,

$$\mathbb{P}(A) = \sum_{k=1}^n \mathbb{P}(A|B_k)\mathbb{P}(B_k).$$

Demuestre el análogo para una partición infinita contable.

7. En el ejemplo 1.3.6 demuestre la expresión para $\mathbb{P}(B)$ y la parte b).
8. Un dado se lanza N veces, donde N es un número aleatorio. Sea A_i el evento $N = i$, y suponga que $\mathbb{P}(A_i) = 2^{-i}$, $i \geq 1$. Sea S la suma de los resultados en las caras de los dados. Encuentre las siguientes probabilidades:
- $N = 2$ dado que $S = 4$.
 - $S = 4$ dado que N es par.
 - $N = 2$, dado que $S = 4$ y el primer dado mostró 1.
 - El número mayor en las caras es r , donde S es desconocida.

Capítulo 2

VARIABLES ALEATORIAS Y FUNCIONES DE DISTRIBUCIÓN

2.1. σ -álgebra de Borel en \mathbb{R}

Teorema 2.1.1 (Mínima σ -álgebra generada) Sea Ω un conjunto y \mathcal{T} una familia de subconjuntos de Ω . Entonces, existe una σ -álgebra $\sigma(\mathcal{T})$ tal que (i) $\mathcal{T} \subset \sigma(\mathcal{T})$ y (ii) si \mathcal{F} es otra σ -álgebra sobre Ω tal que $\mathcal{T} \subset \mathcal{F}$, entonces $\sigma(\mathcal{T}) \subset \mathcal{F}$. Se dice que $\sigma(\mathcal{T})$ es la σ -álgebra sobre Ω generada por \mathcal{T} .

Demostración: Sea \mathcal{R} la familia de todas las σ -álgebras sobre Ω que contienen a \mathcal{T} . Entonces,

$$\mathcal{R} = \{\mathcal{F} : \mathcal{F} \text{ es } \sigma\text{-álgebra sobre } \Omega \text{ y } \mathcal{T} \subset \mathcal{F}\}.$$

Es claro que \mathcal{R} es no vacía, ya que $\mathcal{P}(\Omega) \in \mathcal{R}$. Definamos

$$\mathcal{R}^* := \bigcap_{\mathcal{F} \in \mathcal{R}} \mathcal{F}.$$

Demostraremos que \mathcal{R}^* es una σ -álgebra. En efecto,

- (i) $\Omega \in \mathcal{R}^*$, ya que $\Omega \in \mathcal{F}$, para toda $\mathcal{F} \in \mathcal{R}$.
- (ii) Sea $A \in \mathcal{R}^*$. Entonces, $A \in \mathcal{F}$, para toda $\mathcal{F} \in \mathcal{R}$. Por lo tanto, $A^c \in \mathcal{F}$, para toda $\mathcal{F} \in \mathcal{R}$. Luego, $A^c \in \mathcal{R}^*$.
- (iii) Sea (A_n) una sucesión de conjuntos en \mathcal{R}^* . Mostraremos que $\bigcup_{n=1}^{\infty} A_n \in \mathcal{R}^*$. Sabemos que para cada $\mathcal{F} \in \mathcal{R}$, $A_n \in \mathcal{F}$, para todo n . Ahora bien, dado que \mathcal{F} es σ -álgebra, se tiene que $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$ para toda $\mathcal{F} \in \mathcal{R}$. Por lo tanto, $\bigcup_{n=1}^{\infty} A_n \in \mathcal{R}^*$.

En consecuencia, \mathcal{R}^* es σ -álgebra. Para concluir la prueba basta notar que si \mathcal{F} es una σ -álgebra tal que $\mathcal{T} \subset \mathcal{F}$, entonces $\mathcal{F} \in \mathcal{R}$. Lo cual implica $\mathcal{R}^* \in \mathcal{F}$, i.e., $\sigma(\mathcal{T}) = \mathcal{R}^*$ \square

La siguiente definición introduce una σ -álgebra sobre \mathbb{R} , la cual será de mucha utilidad en el resto del curso.

Definición 2.1.2 La σ -álgebra de Borel¹ sobre \mathbb{R} , la cual denotamos por $\mathcal{B}(\mathbb{R})$, es la σ -álgebra generada por la clase de conjuntos $\mathcal{T} := \{(-\infty, x] : x \in \mathbb{R}\}$. Todo conjunto en $\mathcal{B}(\mathbb{R})$ será llamado conjunto de Borel o Boreliano.

2.2. Variables aleatorias

Cuando se realiza algún experimento generalmente se está interesado en "funciones" del resultado del experimento más que el experimento mismo. Tales cantidades de interés son funciones real-valoradas definidas en el espacio muestral. A dichas funciones aleatorias se les llama *variables aleatorias*.

Supongamos que el experimento consiste en lanzar dos dados a la vez, y nos interesa la suma de los números en la cara superior de los dados. En este caso, el espacio muestral está dado por $\Omega = \{(a, b) : a, b = 1, \dots, 6\}$. Por lo tanto, dado que sólo nos interesa la suma de las caras, para nosotros será lo mismo $\{(5, 1), (1, 5), (4, 2), (2, 4), (3, 3)\}$. De manera análoga, $\{(5, 5), (4, 6), (6, 4)\}$ nos dará el mismo resultado.

Definición 2.2.1 Dado un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ una variable aleatoria (v.a.), denotada por X , es una función $X : \Omega \rightarrow \mathbb{R}$ tal que, para todo $x \in \mathbb{R}$

$$\{X \leq x\} := \{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}. \quad (2.1)$$

Decimos que X es un vector aleatorio, si en la definición anterior \mathbb{R} se intercambia por \mathbb{R}^d , para $d \geq 2$.²

Observación 2.2.2 La definición anterior es equivalente a la condición

$$X^{-1}(I) \equiv \{X \in I\} := \{\omega \in \Omega : X(\omega) \in I\} \in \mathcal{F},$$

donde I es cualquier intervalo en \mathbb{R} (ó \mathbb{R}^d). Más generalmente, $X^{-1}(A) \in \mathcal{F}$ para todo $A \in \mathcal{B}(\mathbb{R})$, donde $\mathcal{B}(\mathbb{R})$ denota la σ -álgebra de Borel en \mathbb{R} .

Existen dos clases muy importantes de variables aleatorias, variables aleatorias discretas y variables aleatorias continuas. Las variables aleatorias discretas aparecen en contextos donde el experimento intrínsecamente tiene un conjunto de resultados posibles a lo más contable. Por otro lado, las variables aleatorias continuas aparecen en experimentos donde el conjunto de resultados posibles es no contable. Por ejemplo, la estatura de una persona, el tiempo de falla de un electrodoméstico, la temperatura de cierto compuesto químico, etc.

Definición 2.2.3 Un variable aleatoria X es llamada discreta si esta toma a lo más un número contable de valores. Es decir, si existe una colección de puntos $\{x_1, x_2, \dots\}$ tales que,

$$X(\Omega) := \{X(\omega) : \omega \in \Omega\} = \cup_{n=1}^{\infty} \{x_i : X(\omega) = x_i, \omega \in \Omega\}.$$

¹Émile Borel 1871-1956, matemático y político francés.

²En la terminología de teoría de la medida se dice que, X es una función medible con respecto a la σ -álgebra \mathcal{F} .

En el caso de v.a. discretas definimos la *función de probabilidades* asociada a X de la siguiente manera

$$p(x_n) := \mathbb{P}(X = x_n) \quad n = 1, 2, \dots .$$

Ahora bien, dado que $\mathbb{P}(\Omega) = 1$, y $\Omega = \cup_{n=1}^{\infty} \{\omega \in \Omega : X = x_n\}$, se tiene que

$$\sum_{n=1}^{\infty} p(x_n) = 1.$$

Observación 2.2.4 *Cualquier función no negativa p tal que el conjunto $\{x \in \mathbb{R} : p(x) > 0\}$ es a lo más contable y $\sum_x p(x) = 1$, es función de probabilidades de alguna variable aleatoria discreta.*

Definición 2.2.5 *Diremos que una v.a. X es continua si existe una función f no negativa, definida en \mathbb{R} , tal que para todo conjunto A de números reales*

$$\mathbb{P}(X \in A) = \int_A f(x) dx. \tag{2.2}$$

La función f es llamada *función de densidad de probabilidad* o simplemente *función de densidad* de la v.a. X .

De la relación (2.2) se obtiene que, si f es una función de densidad, entonces

$$\int_{-\infty}^{\infty} f(x) dx = 1.$$

Lo anterior es debido a que, $\{X \in \mathbb{R}\} = \{\omega \in \Omega : X(\omega) \in \mathbb{R}\} = \Omega$ y $\mathbb{P}(\Omega) = 1$. De hecho, si f es cualquier función continua y no negativa tal que

$$\int_{-\infty}^{\infty} f(x) dx = 1,$$

entonces f es la función de densidad de alguna variable aleatoria continua.

Nota 2.2.6 *Es importante hacer notar que existen v.a. que no son continuas ni discretas. A tales variables aleatorias se le conoce como v.a. mixtas. Sin embargo, tales v.a. quedan fuera del alcance del presente curso.*

Veámos algunos ejemplos de v.a.

1. Sea c una constante en \mathbb{R} , defina $X(\omega) = c$ para todo $\omega \in \Omega$, entonces X es v.a., y es llamada v.a. constante. Para ver que X es v.a. basta notar que

$$\{X \leq x\} = \begin{cases} \emptyset, & \text{si } x < c, \\ \Omega, & \text{si } x \geq c. \end{cases}$$

2. Sea A un evento, entonces

$$X(\omega) = \begin{cases} 1, & \omega \in A, \\ 0, & \text{en otro caso,} \end{cases}$$

es v.a. aleatoria. En efecto, sea I cualquier intervalo en \mathbb{R} ,

$$\{X \in I\} = \begin{cases} \Omega, & \text{si } 0 \in I, 1 \in I, \\ A, & \text{si } 0 \notin I, 1 \in I, \\ A^c, & \text{si } 0 \in I, 1 \notin I, \\ \emptyset, & \text{si } 0 \notin I, 1 \notin I. \end{cases}$$

En muchos casos es de interés estudiar funciones de variables aleatorias. Entonces, surge la siguiente pregunta: si X es v.a. ¿para qué funciones g se cumple que $g(X)$ también es v.a.? La siguiente proposición da respuesta a la pregunta.

Proposición 2.2.7 *Sea X una v.a. definida sobre el espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ y $g : \mathbb{R} \mapsto \mathbb{R}$ una función tal que $g^{-1}(I) \in \mathcal{B}(\mathbb{R})$, para todo $I \in \mathcal{B}(\mathbb{R})$, donde $\mathcal{B}(\mathbb{R})$ denota la σ -álgebra de Borel en \mathbb{R}^3 . Entonces,*

$$Y(\omega) := g(X(\omega)), \quad \omega \in \Omega,$$

también es variable aleatoria.

Demostración: Basta probar que $Y^{-1}(I) \in \mathcal{F}$, para todo intervalo I . Se tiene que,

$$\begin{aligned} Y^{-1}(I) &= \{Y \in I\} = \{\omega \in \Omega : g(X(\omega)) \in I\} \\ &= \{\omega \in \Omega : X(\omega) \in g^{-1}(I)\} \\ &= X^{-1}(g^{-1}(I)). \end{aligned}$$

Ahora bien, dado que $g^{-1}(I) \in \mathcal{B}(\mathbb{R})$ y X es v.a., se tiene que $Y^{-1}(I) \in \mathcal{F}$. □

2.3. Funciones de distribución

Para cada $x \in \mathbb{R}$ definimos el conjunto

$$A(x) = \{\omega \in \Omega : X(\omega) \leq x\}.$$

Por la Definición 2.2.1 se tiene que $A(x) \in \mathcal{F}$, en consecuencia $\mathbb{P}(A(x))$ está bien definido.

Definición 2.3.1 *La función de distribución F_X de una v.a. X se define por*

$$F_X(x) = \mathbb{P}(A(x)), \quad x \in \mathbb{R}.$$

Cuando no haya lugar a confusión simplemente escribiremos F en lugar de F_X .

³Se dice que la función g es medible con respecto a la σ -álgebra de Borel $\mathcal{B}(\mathbb{R})$.

Notemos que si F es una función de distribución, entonces F es una función de \mathbb{R} en $[0, 1]$.

Proposición 2.3.2 *Sea F una función de distribución, entonces se cumplen las siguientes propiedades*

a) $\lim_{x \rightarrow -\infty} F(x) = 0$ y $\lim_{x \rightarrow \infty} F(x) = 1$.

b) Si $x < y$, entonces $F(x) \leq F(y)$.

c) F es continua por la derecha, i.e., $F(x+h) \rightarrow F(x)$ cuando $h \downarrow 0$. Además, F tiene límites por la izquierda, i.e., $\lim_{h \downarrow 0} F(x-h)$ existe, y usualmente se denota por $F(x-)$.

Demostración:

a) Sea $(a_n)_{n=1}^{\infty}$ cualquier sucesión tal que $a_n \downarrow -\infty$, y consideremos $A(a_n)$. Entonces, $A(a_1), A(a_2), \dots$ es una sucesión decreciente y tal que $\bigcap_{n=1}^{\infty} A(a_n) = \emptyset$. Por lo tanto, por el ejercicio 3 del Capítulo 1 (por el Problema 3 de la Tarea 1), se tiene que

$$0 = \mathbb{P}(\bigcap_{n=1}^{\infty} A(a_n)) = \lim_{n \rightarrow \infty} \mathbb{P}(A(a_n)) = \lim_{n \rightarrow \infty} F(a_n),$$

es decir, $\lim_{x \rightarrow -\infty} F(x) = 0$. La otra parte es análoga, y se deja como ejercicio. \square

b) Sea $A(x) = \{X \leq x\}$ y $A(x, y) = \{x < X \leq y\}$. Entonces, $A(y) = A(x) \cup A(x, y)$ y la unión es disjunta. Luego,

$$\mathbb{P}(A(y)) = \mathbb{P}(A(x)) + \mathbb{P}(A(x, y)) \geq \mathbb{P}(A(x)),$$

equivalentemente

$$F(y) \geq F(x).$$

\square

c) Vamos a demostrar que F es continua por la derecha. Debemos probar que

$$\lim_{h \downarrow 0} F(x+h) = F(x).$$

En efecto, notemos que $(-\infty, x] = \bigcap_{n=1}^{\infty} (-\infty, x+a_n]$, donde (a_n) es cualquier sucesión de números reales positivos tales que $a_n \downarrow 0$. Luego,

$$\begin{aligned} F(x) &= \mathbb{P}(\bigcap_{n=1}^{\infty} \{X \leq x+a_n\}) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}(X \leq x+a_n) \\ &= \lim_{n \rightarrow \infty} F(x+a_n) \\ &\equiv \lim_{h \downarrow 0} F(x+h), \end{aligned}$$

es decir, F es continua por la derecha.

De manera similar se puede demostrar que F tiene límites por la izquierda. Para ellos es suficiente notar que

$$\mathbb{P}(X < x) = \lim_{n \rightarrow \infty} \mathbb{P}(X \leq x - 1/n) = \lim_{n \rightarrow \infty} F(x - 1/n) \equiv \lim_{h \downarrow 0} F(x - h).$$

\square

Si X es una v.a. continua con función de densidad f , entonces

$$F_X(x) = \int_{-\infty}^x f(y) dy, \quad x \in \mathbb{R}.$$

Como se puede apreciar en la identidad anterior, la función de distribución es útil para encontrar probabilidades asociadas a la variable aleatoria correspondiente. En esa misma dirección se tiene la siguiente

Proposición 2.3.3 *Sea X una variable aleatoria y F su función de distribución. Entonces, para todo $x < y$ en \mathbb{R} se cumple*

$$\mathbb{P}(x < X \leq y) = F(y) - F(x).$$

Demostración: Notemos que, si $A(x) = (-\infty, x]$, entonces

$$\{\omega \in \Omega : x < X(\omega) \leq y\} = A(y) \cap A(x)^c = A(y) \setminus A(x).$$

Por lo tanto,

$$\mathbb{P}(x < X \leq y) = \mathbb{P}(A(y)) - \mathbb{P}(A(x)) = F(y) - F(x).$$

□

2.4. Algunos ejemplos de variables aleatorias conocidas

2.4.1. Discretas

2.4.2. Continuas

1. **Distribución uniforme.** Decimos que X tiene distribución uniforme en el intervalo $[a, b]$, $a, b \in \mathbb{R}$, si X tiene densidad f dada por

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b, \\ 0, & \text{en otro caso.} \end{cases}$$

La función de distribución correspondiente está dada por

$$F(x) = \int_{-\infty}^x f(y) dy = \begin{cases} 0, & x \leq a, \\ \frac{x-a}{b-a} & a < x < b, \\ 1, & x \geq b. \end{cases}$$

Usaremos la notación $X \sim \text{Unif}([a, b])$.

2. **Distribución exponencial.** Sea X una v.a. con función de distribución dada por

$$F(x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-\lambda x}, & x \geq 0. \end{cases}$$

Se dice que X tiene distribución exponencial de parámetro o intensidad $\lambda > 0$, se denota por $X \sim \text{Exp}(\lambda)$. Notemos que

$$F(x) = \int_{-\infty}^x \lambda e^{-\lambda y} dy,$$

es decir, la densidad de X está dada por

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & \text{en otro caso.} \end{cases}$$

La distribución exponencial tiene una propiedad sumamente importante en la teoría de probabilidad y procesos estocásticos, así como también desde el punto de vista de la *modelación estadística*. A saber, para todo $s, t \geq 0$ se cumple

$$\mathbb{P}(X > t + s | X > t) = \mathbb{P}(X > s), \quad (2.3)$$

la propiedad anterior es conocida como *propiedad de pérdida de memoria*⁴. Vamos a ver que se cumple (2.3). Primero notamos que

$$\mathbb{P}(X > x) = \int_x^{\infty} \lambda e^{-\lambda y} dy = e^{-\lambda x}, \quad x \geq 0.$$

Por lo tanto,

$$\begin{aligned} \mathbb{P}(X > t + s | X > t) &= \frac{\mathbb{P}(X > t + s, X > t)}{\mathbb{P}(X > t)} \\ &= \frac{\mathbb{P}(X > t + s)}{\mathbb{P}(X > t)} \\ &= \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} \\ &= e^{-\lambda s} \\ &= \mathbb{P}(X > s). \end{aligned}$$

Nota 2.4.1 Sea $U \sim \text{Unif}(0, 1)$, entonces para $\lambda > 0$ y $x > 0$, se tiene que

$$\begin{aligned} \mathbb{P}\left(-\frac{1}{\lambda} \log(1 - U) \leq x\right) &= \mathbb{P}(1 - U \geq e^{-\lambda x}) \\ &= \mathbb{P}(U \leq 1 - e^{-\lambda x}) \\ &= 1 - e^{-\lambda x}, \end{aligned}$$

es decir, $-\frac{1}{\lambda} \log(1 - U) \sim \text{Exp}(\lambda)$.

⁴La propiedad de pérdida de memoria caracteriza a la distribución exponencial dentro de la clase de distribuciones continuas.

3. **Distribución normal estándar.** Sea ϕ definida por

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \quad x \in \mathbb{R}.$$

Entonces, ϕ es una función de densidad. Sea X la v.a. asociada, se dice que X tiene *distribución normal estándar*. La función de distribución asociada Φ está dada por

$$\Phi(x) = \int_{-\infty}^x \phi(y) dy, \quad x \in \mathbb{R}.$$

La función Φ no se puede calcular de manera explícita. Por lo tanto, métodos numéricos o de simulación de variables aleatorias, son necesarios para conocer aproximaciones de probabilidades de interés.

Nota 2.4.2 Sean $\mu \in \mathbb{R}$ y $\sigma > 0$ constantes dadas. Luego,

$$\begin{aligned} \mathbb{P}(\sigma X + \mu \leq x) &= \mathbb{P}\left(X \leq \frac{x - \mu}{\sigma}\right) \\ &= \int_{-\infty}^{\frac{x - \mu}{\sigma}} \phi(y) dy \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^x e^{-\frac{(z - \mu)^2}{2\sigma^2}} dz. \end{aligned}$$

A la v.a. $Y := \sigma X + \mu$ se le conoce como *variable aleatoria con distribución normal media μ y varianza σ^2* , y se denota por $Y \sim N(\mu, \sigma^2)$. La función de densidad de Y está dada por

$$\phi(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x - \mu)^2}{2\sigma^2}}.$$

Las variables aleatorias continuas tienen la siguiente

Propiedad. Sea X una variable aleatoria continua. Entonces,

$$\mathbb{P}(X = x) = \text{para todo } x \in \mathbb{R}.$$

Demostración: Notemos que $\{x\} = \bigcap_{n=1}^{\infty} (x - \frac{1}{n}, x]$. Luego, como $A_n = (x - \frac{1}{n}, x]$ es una sucesión decreciente se tiene que

$$\begin{aligned} \mathbb{P}(X = x) &= \lim_{n \rightarrow \infty} \mathbb{P}\left(x - \frac{1}{n} < X \leq x\right) \\ &= \lim_{n \rightarrow \infty} \left[F_X(x) - F_X\left(x - \frac{1}{n}\right)\right] \\ &= 0, \text{ dado que } F \text{ es continua.} \end{aligned}$$

□

Capítulo 3

Esperanza condicional

Esperanza condicional es una herramienta fundamental en la Teoría de Procesos Estocásticos. El propósito del presente capítulo es definir dicho concepto y estudiar algunas de sus propiedades más importantes. Trabajaremos en un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$ fijo, i.e., todas las variables aleatorias estarán definidas en dicho espacio de probabilidad sin necesidad de hacer mención explícita de ello.

3.1. Definición de esperanza condicional

Sabemos que si X es una variable aleatoria discreta entonces $\mathbb{E}(X) = \sum_k x_k \mathbb{P}(X = k)$. Sea A un evento tal que $\mathbb{P}(A) \neq 0$, entonces podemos definir

$$\begin{aligned}\mathbb{E}[X|A] &= \sum_k x_k \mathbb{P}(X = x_k|A) \\ &= \sum_k x_k \frac{\mathbb{P}(X = x_k, A)}{\mathbb{P}(A)} \\ &= \sum_k x_k \frac{\mathbb{P}(X \mathbf{1}_A = x_k)}{\mathbb{P}(A)} \\ &= \frac{1}{\mathbb{P}(A)} \mathbb{E}(X \mathbf{1}_A).\end{aligned}$$

Lo anterior no da la esperanza condicional de la variable aleatoria X dado el evento A . Tal concepto se puede extender de la siguiente manera:

Definición 3.1.1 Sean X y Y variables aleatorias discretas. La esperanza condicional de X dado que $Y = y$, donde $f_Y(y) > 0$, se define por

$$\mathbb{E}(X|\{Y = y\}) = \sum_x x f_{X,Y}(x, y) / f_Y(y), \quad (3.1)$$

siempre y cuando la suma sea absolutamente convergente.

Notese que conforme y varia (sobre todos los posibles valores de Y) en la ecuación (3.1), se obtiene una función de Y , la cual denotaremos por $\mathbb{E}(X|Y)$. Entonces, $\mathbb{E}(X|Y)$ es una variable aleatoria tal que

$$\mathbb{E}(X|Y)(\omega) = \mathbb{E}(X|\{Y = y_n\}), \text{ si } Y(\omega) = y_n, \quad (3.2)$$

donde y_1, y_2, \dots son los posibles valores de Y . A la variable aleatoria $\mathbb{E}(X|Y)$ le llamaremos *esperanza condicional de X dado Y* .

Ejemplo 3.1.2 *Considere el lanzamiento de 3 monedas con denominación de 1, 5 y 10 pesos, respectivamente. Sea X la suma de las monedas que caen águila.*

(i) *¿Cual es el valor esperado de X dado que dos monedas caen águila?*

(ii) *Sea Y la suma de las monedas que caen águila, y que además, tienen denominación de 1 ó 5 pesos. ¿Cual es la esperanza condicional de X dado Y ?*

Solución: (i) *El espacio muestral está dado por*

$$\Omega = \{AAA, AAS, ASA, SAA, ASS, SAS, SSA, SSS\}.$$

Sea B el evento que dos monedas caen águila, i.e.,

$$B = \{AAS, ASA, SAA\}$$

Nos interesa determinar $\mathbb{E}(X|B)$. Notemos que, cada punto en B ocurre con probabilidad $1/8$. Luego,

$$\begin{aligned} X(AAS) &= 1 + 5 = 6, \\ X(ASA) &= 1 + 10 = 11, \\ X(SAA) &= 5 + 10 = 15. \end{aligned}$$

Por lo tanto,

$$\mathbb{E}(X|B) = \frac{1}{\frac{3}{8}} \left(6\frac{1}{8} + 11\frac{1}{8} + 15\frac{1}{8} \right) = \frac{32}{3}.$$

(ii) *Ahora observamos que, $Y \in \{0, 1, 5, 6\}$ con probabilidades*

$$\mathbb{P}(Y = 0) = \mathbb{P}(Y = 1) = \mathbb{P}(Y = 5) = \mathbb{P}(Y = 6) = \frac{1}{4}.$$

Finalmente, siguiendo el mismo procedimiento que en (i) obtenemos

$$\begin{aligned} \mathbb{E}(X|\{Y = 0\}) &= 5, & \mathbb{E}(X|\{Y = 1\}) &= 6, \\ \mathbb{E}(X|\{Y = 5\}) &= 10, & \mathbb{E}(X|\{Y = 6\}) &= 11. \end{aligned}$$

Por lo tanto, la esperaza condicional de X dado Y resulta ser

$$\mathbb{E}(X|Y)(\omega) = \begin{cases} 5 & \text{si } Y(\omega) = 0, \\ 6 & \text{si } Y(\omega) = 1, \\ 10 & \text{si } Y(\omega) = 5, \\ 11 & \text{si } Y(\omega) = 6. \end{cases} \quad (3.3)$$

□

Notemos que en el ejemplo anterior $\mathbb{E}(X|Y)$ toma cada valor con la misma probabilidad, es decir, $1/4$. Por lo tanto, $\mathbb{E}(\mathbb{E}(X|Y)) = 8 = \mathbb{E}(X)$. La propiedad anterior no es particular de este ejemplo. Más adelante veremos que tal propiedad se cumple en general.

Ejemplo 3.1.3 Sean (X, Y) un vector aleatorio con función de probabilidad conjunta

$$f(x, y) = \begin{cases} \frac{2}{N(N+1)}, & \text{si } x \leq y, x, y \in \{1, 2, \dots, N\} \\ 0, & \text{en otro caso,} \end{cases}$$

donde N es un entero positivo. Encuentre (i) $\mathbb{E}(X|Y)$ y (ii) $\mathbb{E}(Y|X)$.

Solución: (i) Notemos que

$$f_Y(y) = \sum_{x=1}^y f(x, y) = \frac{2}{N(N+1)}y, \quad y = 1, 2, \dots, N.$$

Luego,

$$\mathbb{E}[X|Y = y] = \sum_{x=1}^y x \frac{f(x, y)}{f_Y(y)} = \frac{1}{y} \sum_{x=1}^y x = \frac{y+1}{2}.$$

Por lo tanto, $\mathbb{E}[X|Y] = \frac{1}{2}(Y+1)$. □

(ii) Porcediendo de manera análogo al inciso anterior se tiene que

$$f_X(x) = \sum_{y=1}^N f(x, y) = \sum_{y=x}^N \frac{2}{N(N+1)} = \frac{2}{N(N+1)}(N+1-x), \quad x = 1, \dots, N.$$

Luego, para $x \in \{1, \dots, N\}$, se tiene que

$$\begin{aligned} E[Y|X = x] &= \sum_{y=1}^N y \frac{f(x, y)}{f_X(x)} \\ &= \frac{1}{N+1-x} \sum_{y=x}^N y \\ &= \frac{x+N}{2}. \end{aligned}$$

Por lo tanto, $\mathbb{E}[Y|X] = \frac{1}{2}(X+N)$. □

Teorema 3.1.4 Sean X una variable aleatoria discreta con esperanza finita y Y cualquier variable aleatoria discreta. Entonces,

(i)

$$\mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}(X), \tag{3.4}$$

siempre que ambos lados existan.

(ii) Para toda función $g: \mathbb{R} \rightarrow \mathbb{R}$ medible y acotada, se tiene

$$\mathbb{E}[g(Y)\mathbb{E}[X|Y]] = \mathbb{E}[g(Y)X].$$

Demostración: (i) Siempre que las sumatorias sean absolutamente convergentes se tiene que,

$$\begin{aligned}
 \mathbb{E}(\mathbb{E}(X|Y)) &= \sum_y \mathbb{E}(X|\{Y = y\})f_Y(y) \\
 &= \sum_y \left(\sum_x \frac{xf_{X,Y}(x,y)}{f_Y(y)} \right) f_Y(y) \\
 &= \sum_x xf_X(x) \\
 &= \mathbb{E}(X).
 \end{aligned}$$

(ii) Sea $g : \mathbb{R} \rightarrow \mathbb{R}$ cualquier función medible y acotada, entonces

$$\begin{aligned}
 \mathbb{E}[g(Y)\mathbb{E}[X|Y]] &= \sum_k g(y_k)\mathbb{E}[X|Y = y_k]\mathbb{P}(Y = y_k) \\
 &= \sum_k g(y_k) \left(\sum_j x_j \frac{\mathbb{P}(X = x_j, Y = y_k)}{\mathbb{P}(Y = y_k)} \right) \mathbb{P}(Y = y_k) \\
 &= \sum_k g(y_k) \sum_j x_j \mathbb{P}(X = x_j, Y = y_k) \\
 &= \sum_{k,j} g(y_k)x_j \mathbb{P}(X = x_j, Y = y_k) \\
 &\equiv \mathbb{E}[g(Y)X].
 \end{aligned}$$

□

Observación 3.1.5 La esperanza condicional $\mathbb{E}[X|Y]$ está bien definida. En efecto, se $h : \mathbb{R} \rightarrow \mathbb{R}$ una función medible tal que $h(Y)$ tiene esperanza finita y $\mathbb{E}[g(Y)h(Y)] = \mathbb{E}[g(Y)\mathbb{E}[X|Y]]$ para cualquier función g medible y acotada. Luego,

$$\sum_k g(y_k)h(y_k)\mathbb{P}(Y = y_k) = \sum_{k,j} g(y_k)x_j \mathbb{P}(X = x_j, Y = y_k).$$

Ahora bien, la identidad anterior se cumple para todo g , en particular para $f = \mathbf{1}_{\{y_k\}}$, se tiene que

$$h(y_k)\mathbb{P}(Y = y_k) = \sum_j x_j \mathbb{P}(X = x_j, Y = y_k),$$

es decir,

$$h(y_k) = \mathbb{E}[X|Y = y_k], \text{ para todo } k.$$

La observación anterior nos permite dar una definición de esperanza condicional de una variable aleatoria dada otra variable aleatoria sin el supuesto de que estas sean discretas.

Definición 3.1.6 (Esperanza condicional de una variable aleatoria dada otra variable aleatoria) Sea X una variable aleatoria con esperanza finita y Y cualquier variable aleatoria. Si existe una función medible $h : \mathbb{R} \rightarrow \overline{\mathbb{R}}$ tal que $h(Y)$ tiene media finita y

$$\mathbb{E}[g(Y)h(Y)] = \mathbb{E}[g(Y)X],$$

para cualquier función medible $g : \mathbb{R} \rightarrow \mathbb{R}$ medible y acotada, entonces se dice que $h(Y)$ es una versión de esperanza condicional $\mathbb{E}[X|Y]$ y se define

$$\mathbb{E}[X|Y] = h(Y) \text{ y } \mathbb{E}[X|Y = y] = h(y), \text{ } y \in \mathbb{R}.$$

Ejemplo 3.1.7 Una gallina pone X huevos, donde X es Poisson con parámetro λ . Cada huevo es fecundado con probabilidad p , independientemente de los otros, produciendo así Y pollos. Demuestre que $\rho(X, Y) = \sqrt{p}$.

Solución: Observemos que, condicional en $X = k$, Y tiene distribución binomial $\text{Bin}(k, p)$. Por lo tanto, $\mathbb{E}(Y|\{X = k\}) = kp$. Más generalmente,

$$\mathbb{E}(Y|X) = Xp.$$

Entonces, por el teorema anterior se tiene que

$$\mathbb{E}(XY) = \mathbb{E}(\mathbb{E}(XY|X)) = \mathbb{E}(X\mathbb{E}(Y|X)) = \mathbb{E}(X^2p) = (\lambda^2 + \lambda)p.$$

De manera similar, obtenemos

$$\begin{aligned} \mathbb{E}(Y^2) &= \mathbb{E}(\mathbb{E}(Y^2|X)) \\ &= \mathbb{E}(Xp(1-p) + X^2p^2) \\ &= \lambda p(1-p) + (\lambda^2 + \lambda)p^2 \\ &= \lambda p + \lambda^2 p^2. \end{aligned}$$

Finalmente, tenemos que

$$\begin{aligned} \rho(X, Y) &= \frac{\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)}{(\text{Var}(X)\text{Var}(Y))^{1/2}} \\ &= \frac{(\lambda^2 + \lambda)p - \lambda \cdot \lambda p}{(\lambda(\lambda p + \lambda^2 p^2 - \lambda^2 p^2))^{1/2}} \\ &= \sqrt{p}. \end{aligned}$$

□

Ejemplo 3.1.8 Sea (X_i) una sucesión de v.a. i.i.d., y sea Y una v.a. con valores en los enteros no negativos independiente de la sucesión (X_i) . Defina $S_Y = \sum_{i=1}^Y X_i$. Demuestre que

$$\text{Var}(S_Y) = \mathbb{E}(X_1^2) \text{Var}(Y) + \mathbb{E}(Y) \text{Var}(X_1).$$

Solución: Por el ejercicio 2 de la tarea 2 tenemos que,

$$\mathbb{E}(S_Y) = \mathbb{E}(X_1)\mathbb{E}(Y).$$

Ahora bien, por el Teorema 3.1.4 se obtiene

$$\begin{aligned} \mathbb{E}(S_Y^2) &= \mathbb{E}(\mathbb{E}(S_Y^2|Y)) \\ &= \sum_{k=0}^{\infty} \mathbb{E}(S_k^2) p_k = \mathbb{P}(Y = k) \\ &= \sum_{k=0}^{\infty} p_k \left(\sum_{i=1}^k \mathbb{E}(X_i^2) + \sum_{i=1}^k \sum_{j \neq i}^k \mathbb{E}(X_i X_j) \right) \\ &= \sum_{k=0}^{\infty} p_k (k\mathbb{E}(X_1^2) + k(k-1)[\mathbb{E}(X_1)]^2) \\ &= \mathbb{E}(Y\mathbb{E}(X_1^2) + Y(Y-1)[\mathbb{E}(X_1)]^2) \\ &= \mathbb{E}(Y) (\mathbb{E}(X_1^2) - [\mathbb{E}(X_1)]^2) + \mathbb{E}(Y^2)[\mathbb{E}(X_1)]^2. \end{aligned}$$

El resultado se sigue usando la identidad $\text{Var}(S_Y) = \mathbb{E}(S_Y^2) - [\mathbb{E}(S_Y)]^2$.

3.2. Propiedades de la esperanza condicional

Teorema 3.2.1 Sean a y b constantes, g una función de valor real, y suponga que X , Y y Z son conjuntamente distribuidas. Entonces,

1. $\mathbb{E}(a|Y) = a$.
2. $\mathbb{E}(aX + bZ|Y) = a\mathbb{E}(X|Y) + b\mathbb{E}(Z|Y)$.
3. $\mathbb{E}(X|Y) \geq 0$ si $X \geq 0$.
4. $\mathbb{E}(X|Y) = \mathbb{E}(X)$ si X e Y son independientes.
5. $\mathbb{E}(Xg(Y)|Y) = g(Y)\mathbb{E}(X|Y)$.
6. $\mathbb{E}(X|Y, g(Y)) = \mathbb{E}(X|Y)$.
7. $\mathbb{E}(\mathbb{E}(X|Y, Z)|Y) = \mathbb{E}(X|Y)$.

Demostración: (1) Sabemos que $f_{a,Y}(a, y) = f_Y(y)$. Entonces,

$$\mathbb{E}(a|Y) = a \frac{f_{a,X}(a, y)}{f_Y(y)} = a.$$

□

(2) Tarea.

(3) Si $X \geq 0$, entonces cada sumando en la definición de esperanza condicional será no-negativo. Por lo tanto, $\mathbb{E}(X|Y) \geq 0$. \square

(4) Para cada y en el rango de Y , tenemos que

$$\mathbb{E}(X|Y)(\omega) = \mathbb{E}(X|Y = y), \quad Y(\omega) = y.$$

Luego, por definición de esperanza condicional se tiene que

$$\begin{aligned} \mathbb{E}(X|Y = y) &= \sum_x x \frac{f_{X,Y}(x, y)}{f_Y(y)} \\ &= \sum_x x \frac{f_X(x) f_Y(y)}{f_Y(y)} \\ &= \mathbb{E}(X). \end{aligned}$$

Entonces, $\mathbb{E}(X|Y)(\omega) = \mathbb{E}(X)$, para todo $\omega \in \Omega$. \square

(5) Notemos que, conjunto y perteneciente al rango de Y se tiene

$$\begin{aligned} \mathbb{E}(Xg(Y)|\{Y = y\}) &= \frac{\sum_x xg(y)f_{X,Y}(x, y)}{f_Y(y)} \\ &= g(y) \frac{\sum_x x f_{X,Y}(x, y)}{f_Y(y)} \\ &= g(y) \mathbb{E}(X|\{Y = y\}). \end{aligned}$$

Entonces, $\mathbb{E}(Xg(Y)|Y) = g(Y)\mathbb{E}(X|Y)$.

El resultado anterior se puede interpretar de la siguiente manera: esperanza condicional es una manera de “medir” la información que aporta Y sobre la variable aleatoria $Xg(Y)$. En consecuencia, al menos intuitivamente, se tiene que conociendo el valor de Y automáticamente conocemos el valor de $g(Y)$, y por lo tanto, puede tratarse como una constante dentro de la esperanza condicional. \square

(6) Tarea.

(7) Supongamos que $Y = y$ y que $Z = z$, entonces

$$\mathbb{E}(X|\{Y = y\}, \{Z = z\}) = \frac{\sum_x x f_{X,Y,Z}(x, y, z)}{f_{Y,Z}(y, z)}.$$

Entonces, por definición tenemos que, para cada ω tal que $Y(\omega) = y$

$$\begin{aligned}
\mathbb{E}(\mathbb{E}(X|Y, Z)|Y) &= \sum_z \mathbb{E}(X|Y = y, Z = z) \frac{f_{Y,Z}(y, z)}{f_Y(y)} \\
&= \sum_z \sum_x x \frac{f_{X,Y,Z}(x, y, z)}{f_{Y,Z}(y, z)} \frac{f_{Y,Z}(y, z)}{f_Y(y)} \\
&= \sum_x x \frac{\sum_z f_{X,Y,Z}(x, y, z)}{f_Y(y)} \\
&= \sum_x x \frac{f_{X,Y}(x, y)}{f_Y(y)} \\
&= \mathbb{E}(X|\{Y = y\}).
\end{aligned}$$

De lo anterior se sigue el resultado, ya que se cumple para cada y en el rango de Y . \square

Concluimos la presente sección con un resultado que nos dice que, $\mathbb{E}(X|Y)$ es la variable aleatoria que se encuentra a una menor distancia de X de todas las variables aleatorias que se pueden determinar a partir de Y .

Teorema 3.2.2 *Sea h cualquier función de valor real tal que $\mathbb{E}(h(Y)^2) < \infty$. Entonces,*

$$\mathbb{E}[(X - h(Y))^2] \geq \mathbb{E}[(X - \mathbb{E}(X|Y))^2]. \quad (3.5)$$

Más aún, si h es tal que

$$\mathbb{E}[(X - h(Y))^2] = \mathbb{E}[(X - \mathbb{E}(X|Y))^2],$$

entonces

$$\mathbb{E}[(h(Y) - \mathbb{E}(X|Y))^2] = 0.$$

Demstración:

$$\begin{aligned}
\mathbb{E}[(X - h(Y))^2] &= \mathbb{E}[(X - \mathbb{E}(X|Y) + \mathbb{E}(X|Y) - h(Y))^2] \\
&= \mathbb{E}[(X - \mathbb{E}(X|Y))^2] + \mathbb{E}[(\mathbb{E}(X|Y) - h(Y))^2] \\
&\quad + 2\mathbb{E}[(X - \mathbb{E}(X|Y))(\mathbb{E}(X|Y) - h(Y))].
\end{aligned}$$

Ahora bien, por Teorema 3.1.4 tenemos que

$$\begin{aligned}
\mathbb{E}[(X - \mathbb{E}(X|Y))(\mathbb{E}(X|Y) - h(Y))] &= \mathbb{E}[\mathbb{E}((X - \mathbb{E}(X|Y))(\mathbb{E}(X|Y) - h(Y))|Y)] \\
&= \mathbb{E}[(\mathbb{E}(X|Y) - h(Y))\mathbb{E}[(X - \mathbb{E}(X|Y))|Y]],
\end{aligned}$$

donde la segunda igualdad es debido al Teorema 3.2.1 (5). Luego, observemos que $\mathbb{E}[(X - \mathbb{E}(X|Y))|Y] = 0$. Entonces,

$$\mathbb{E}[(X - h(Y))^2] = \mathbb{E}[(X - \mathbb{E}(X|Y))^2] + \mathbb{E}[(\mathbb{E}(X|Y) - h(Y))^2]. \quad (3.6)$$

Para terminar la prueba note que $\mathbb{E}[(\mathbb{E}(X|Y) - h(Y))^2] \geq 0$, y por lo tanto, obtenemos (3.5).

Por último, si $\mathbb{E}[(X - h(Y))^2] = \mathbb{E}[(X - \mathbb{E}(X|Y))^2]$ de (3.6) se concluye que $\mathbb{E}[(\mathbb{E}(X|Y) - h(Y))^2] = 0$, es decir, $h(Y) = \mathbb{E}(X|Y)$ salvo en un conjunto de probabilidad cero. \square

Ejemplo 3.2.3 En una fiesta n de los asistentes se quita el sombrero. Se ponen los n sombreros en un contenedor, y cada persona selecciona uno al azar. Decimos que ocurre un “coincidencia” si una persona selecciona su propio sombrero. ¿Cuál es la probabilidad de que no ocurra ninguna coincidencia? ¿Cuál es la probabilidad de que ocurran exactamente k coincidencias?

Solución: Sea E el evento que no ocurra ninguna coincidencia, sea $p_n := \mathbb{P}(E)$. Definamos el evento $M :=$ “la primera persona selecciona su propio sombrero”. Entonces,

$$p_n = \mathbb{P}(E) = \mathbb{P}(E|M)\mathbb{P}(M) + \mathbb{P}(E|M^c)\mathbb{P}(M^c).$$

Notemos que, $\mathbb{P}(E|M) = 0$, (al menos ocurre una coincidencia). Luego,

$$p_n = \mathbb{P}(E|M^c)\frac{n-1}{n}. \quad (3.7)$$

Ahora bien, note que $\mathbb{P}(E|M^c)$ es la probabilidad de que no ocurra ninguna coincidencia cuando $n-1$ personas seleccionan de $n-1$ sombreros, y que además, hay una persona cuyo sombrero no está dentro de los $n-1$. Lo anterior puede pasar de dos formas mutuamente excluyentes: (1) no ocurre ninguna coincidencia y la persona extra no selecciona el sombrero extra (el sombrero perteneciente a la primera persona en seleccionar); ó (2) no ocurre ninguna coincidencia y la persona extra selecciona el sombrero extra. La probabilidad de (1) es exactamente p_{n-1} , lo anterior es considerando que el sombrero extra pertenece a la persona extra. Por otro lado, la probabilidad de (2) está dada por $\frac{1}{n-1}p_{n-2}$. Entonces,

$$\mathbb{P}(E|M^c) = p_{n-1} + \frac{1}{n-1}p_{n-2}.$$

Por lo tanto, combinando la ecuación anterior con (3.7) tenemos,

$$p_n = \frac{n-1}{n}p_{n-1} + \frac{1}{n}p_{n-2},$$

equivalentemente,

$$p_n - p_{n-1} = -\frac{1}{n}(p_{n-1} - p_{n-2}). \quad (3.8)$$

Ahora bien, dado que p_n es la probabilidad de que no ocurra ninguna coincidencia cuando n personas seleccionan un sombrero, tenemos

$$p_1 = 0, \quad p_2 = \frac{1}{2},$$

y por lo tanto, de (3.8) resulta que

$$p_3 - p_2 = -\frac{p_2 - p_1}{3} = -\frac{1}{3!},$$

es decir,

$$p_3 = \frac{1}{2!} - \frac{1}{3!}.$$

De manera análoga, obtenemos

$$p_4 - p_3 = -\frac{p_3 - p_2}{4} = \frac{1}{4!},$$

es decir,

$$p_4 = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!}.$$

Procediendo de manera similar, obtenemos que

$$p_n = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \dots + \frac{(-1)^n}{n!}.$$

Para responder la segunda pregunta considere un grupo fijo de k personas. La probabilidad que ellos, y solamente ellos, seleccionen sus propios sombreros está dada por

$$\frac{1}{n} \frac{1}{n-1} \dots \frac{1}{n-(k-1)} p_{n-k} = \frac{(n-k)!}{n!} p_{n-k},$$

donde p_{n-k} es la probabilidad que las $n-k$ personas restantes, que seleccionan dentro de sus propios sombreros, no haya ninguna coincidencia. Ahora bien, dado que hay exactamente $\binom{n}{k}$ formas diferentes de seleccionar un grupo de k personas, la probabilidad de que haya exactamente k coincidencias está dada por

$$\begin{aligned} \binom{n}{k} \frac{(n-k)!}{n!} p_{n-k} &= \frac{p_{n-k}}{k!} \\ &= \frac{\frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \dots + \frac{(-1)^{n-k}}{(n-k)!}}{k!}. \end{aligned}$$

Por lo tanto, para n suficientemente grande, se tiene que la probabilidad de que haya exactamente k coincidencias es aproximadamente $\frac{e^{-1}}{k!}$. \square

Caso absolutamente continuo.

Hasta ahora, en los ejemplos, hemos puesto mucho énfasis en caso de variables (vectores) aleatorias discretas. Sin embargo, hemos dado la definición general de esperanza condicional de una variable aleatoria dado otra variable aleatoria. En el caso absolutamente continuo se tiene lo siguiente: sea (X, Y) un vector aleatorio con función de densidad conjunta $f_{X,Y}$ tal que $\mathbb{E}(X) < \infty$. Entonces, una función de densidad para la esperanza condicional $\mathbb{E}(X|Y)$ esta dada por

$$\frac{f_{X,Y}(x, y)}{f_Y(y)}, \text{ siempre que } f_Y(y) > 0.$$

Comunmente se usa la notación

$$f_{X|Y}(x|y) \equiv \frac{f_{X,Y}(x, y)}{f_Y(y)}.$$

3.3. Más ejemplos

Ejemplo 3.3.1 Supongamos que el número de accidentes que tiene una persona en un año tiene distribución Poisson con parámetro Y , de modo que, para cada $y > 0$, el porcentaje de personas para las cuales $Y > y$ es igual a $\lambda e^{-\lambda y}$, donde λ es una constante positiva. Si X es el número de accidentes en un año de una persona seleccionada al azar, encuentre i) la distribución de X y $\mathbb{E}(X)$, ii) la distribución condicional de Y dado que $X = x$, para $x \in \{0, 1, \dots\}$ y iii) $\mathbb{E}(Y|X)$.

Solución: i) Sabemos que $X|Y$ tiene distribución Poisson de parámetro Y . Entonces, por Ley de Probabilidad Total, para cada $x \in \{0, 1, \dots\}$, se tiene que

$$\begin{aligned}\mathbb{P}(X = x) &= \int_0^\infty \mathbb{P}(X = x|Y = y) f_Y(y) dy \\ &= \int_0^\infty \frac{e^{-y} y^x}{x!} \lambda e^{-\lambda y} dy \\ &= \frac{\lambda}{x!} \int_0^\infty y^x e^{-(\lambda+1)y} dy,\end{aligned}$$

vamos a completar la integral anterior para que sea una Gama($x, \lambda + 1$), entonces

$$\begin{aligned}\mathbb{P}(X = x) &= \frac{\lambda}{(\lambda + 1)^{x+1}} \int_0^\infty \frac{(\lambda + 1)^{x+1} y^x e^{-(\lambda+1)y}}{x!} dy \\ &= \frac{\lambda}{\lambda + 1} \left(\frac{1}{\lambda + 1} \right)^x.\end{aligned}$$

Por lo tanto, X tiene distribución geométrica de parámetro $p = \lambda/(\lambda + 1)$, puesto que

$$\mathbb{P}(X = x) = p(1 - p)^x, \quad x = 0, 1, \dots$$

Entonces,

$$\mathbb{E}(X) = \frac{1 - p}{p} = \frac{1}{\lambda} \equiv \int_0^\infty \mathbb{E}(X|Y = y) \lambda e^{-\lambda y} dy = \int_0^\infty y \lambda e^{-\lambda y} dy.$$

ii) Para $x \in \{0, 1, \dots\}$ y $y > 0$ se tiene que

$$f_{Y|X}(x|y) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{f_{X|Y}(x|y) f_Y(y)}{f_X(x)},$$

notemos que no conocemos $f_{X,Y}$. Sin embargo, en el último término si conocemos todos los factores. Por lo tanto,

$$f_{Y|X}(x|y) = \frac{\frac{y^x e^{-y}}{x!} \lambda e^{-\lambda y}}{\frac{\lambda}{\lambda+1} \left(\frac{1}{\lambda+1}\right)^x} = \frac{(\lambda + 1)^{x+1} y^x e^{-(\lambda+1)y}}{x!}.$$

iii) De la parte ii) observamos que $Y|X$ tiene distribución gama con parámetros $x + 1$ y $\lambda + 1$. Luego, recordando que si $Z \sim \text{Gamma}(\alpha, \beta)$, entonces $\mathbb{E}(Z) = \alpha/\beta$. Por lo tanto,

$$\mathbb{E}(Y|X) = \frac{X + 1}{\lambda + 1}.$$

□

Ejemplo 3.3.2 Sean X y Y dos variables aleatorias con esperanza finita tales que $\mathbb{E}(X|Y) = \mathbb{E}(Y)$. Supongamos que XY también tiene esperanza finita, demuestre que $\text{Cov}(X, Y) = 0$.

Solución: Sabemos que

$$\begin{aligned}\text{Cov}(X, Y) &\equiv \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \\ &= \mathbb{E}[\mathbb{E}(XY|X)] - \mathbb{E}(X)\mathbb{E}(Y), \text{ (por Teorema 3.1.4 (i))} \\ &= \mathbb{E}[X\mathbb{E}(Y|X)] - \mathbb{E}(X)\mathbb{E}(Y), \text{ (por Teorema 3.2.1 5)} \\ &= \mathbb{E}[X\mathbb{E}(Y)] - \mathbb{E}(X)\mathbb{E}(Y) \\ &= 0.\end{aligned}$$

□

Ejemplo 3.3.3 *Supongamos que el número de personas que suben a un elevador, en la planta baja de un edificio de N pisos, tiene distribución Poisson con parámetro λ . Supongamos que cada persona deja el elevador al azar, en cualquiera de los N pisos, independientemente de donde bajen los demás. Encuentre el número esperado de paradas que hace el elevador hasta que bajan todas las personas.*

Solución: sea Y en número de personas que sube al elevador en la planta baja y X en número de paradas necesarias para que el elevador quede vacío. Definamos las variables aleatorias X_1, X_2, \dots, X_N como sigue:

$$X_i = \begin{cases} 1, & \text{si el elevador para en el piso } i, \\ 0, & \text{en otro caso.} \end{cases}$$

Entonces, $X = \sum_{i=1}^N X_i$, y para cada $k \in \{0, 1, \dots, \}$ se tiene que

$$\mathbb{P}(X_1 = 0|Y = k) = \mathbb{P}(\text{ninguna persona baja en el piso } i) = \left(1 - \frac{1}{N}\right)^k,$$

es decir,

$$\mathbb{E}(X_i|Y = k) = 1 - \left(1 - \frac{1}{N}\right)^k.$$

Luego,

$$\mathbb{E}(X|Y = k) = N \left[1 - \left(1 - 1/N\right)^k\right].$$

Por lo tanto, por la Ley de Probabilidad Total,

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k=0}^{\infty} \mathbb{E}(X|Y = k)\mathbb{P}(Y = k) \\ &= \sum_{k=0}^{\infty} N \left[1 - \left(1 - 1/N\right)^k\right] \frac{\lambda^k e^{-\lambda}}{k!} \\ &= N \left(1 - e^{-\lambda/N}\right).\end{aligned}$$

□

3.4. Ejercicios

Los ejercicios marcados con * son para entregar.

1. *Sean X y Y dos variables aleatorias discretas, y g una función de real-valuada. Demuestre lo siguiente

$$\mathbb{E}(X|Y, g(Y)) = \mathbb{E}(X|Y).$$

2. *Sea (X, Y) un vector aleatorio con función de densidad conjunta dada por

$$f_{X,Y}(x, y) = \begin{cases} \frac{6}{N(N^2-1)}(y-x), & \text{si } x < y \text{ y } x, y \in \{1, \dots, N\}, \\ 0, & \text{en otro caso.} \end{cases}$$

Encuentre la función de densidad condicional de: (i) X dado Y ; (ii) Y dado X .

3. *Sea X y Y variables aleatorias independientes con distribución Poisson. Sea $Z := X + Y$. Encuentre la distribución condicional de X dado Z .
4. *Una moneda muestra águila con probabilidad p . Sea X_n el número de lanzamientos necesarios para obtener un corrida de n águilas consecutivas. Demuestre que

$$\mathbb{E}(X_n) = \sum_{k=1}^n p^{-k}.$$

Sugerencia: recuerde que $\mathbb{E}(X_n) = \mathbb{E}[\mathbb{E}(X_n|X_{n-1})]$.

5. Sea X_1, \dots, X_n variables aleatorias independientes e idénticamente distribuidas, con esperanza finita. Demuestre que, para cualquier $k \in \{1, \dots, n\}$, se cumple

$$\mathbb{E} \left(X_k \mid \sum_{i=1}^n X_i \right) = \frac{1}{n} \sum_{i=1}^n X_i.$$

6. *Se eligen, al azar y sin reemplazo, dos tarjetas de una urna que contiene N tarjetas numeradas del 1 al N , con $N \geq 1$. Sean X y Y el menor y mayor, respectivamente, de los números en la tarjetas seleccionadas. Encuentre $\mathbb{E}(X|Y)$ y $\mathbb{E}(Y|X)$.
7. Supongamos que (X_i) es una sucesión de variables aleatorias independientes tales que,

$$X_i|P_i \sim \text{Ber}(P_i), \quad i = 1, 2, \dots, n,$$

donde $P_i \sim \text{Beta}(\alpha, \beta)$. Defina $Y_n = \sum_{i=1}^n X_i$. Encuentre $\mathbb{E}(Y_n)$ y $\text{Var}(Y_n)$.

8. *Dos jugadores A y B tienen n monedas. Se las reparten de la siguiente manera: lanzan cada moneda y A obtiene las que resultan “águila”, digamos X , entonces B obtiene las restantes $n - X$ monedas.

Luego, A y B juegan volados independientes y justos, cada vez que A gana (la moneda cae águila) B le da una moneda al jugador A ; y cada vez que pierde le da una moneda a B . El juego termina cuando uno de ellos se queda sin monedas.

Sea D_X el número de volados jugados. Encuentre $\mathbb{E}(D_X)$, y demuestre que $\rho(X, D_X) = 0$.

9. Sean X y Y dos variables aleatorias independientes, ambas con distribución uniforme en el conjunto $\{1, \dots, N\}$. Encuentre $\mathbb{E}(X|Y - X)$ y $\mathbb{E}(Y|Y - X)$.

3.5. Caminatas aleatorias simples

En esta sección veremos nuestro primer ejemplo de proceso estocástico llamado *caminata aleatoria*. La caminata aleatoria es un proceso simple de describir. Sin embargo, eso no quiere decir que sea sencillo estudiarla. Por otro lado, nos permite presentar algunos problemas que son de interés en procesos mucho más complicados. Comencemos con la definición de proceso estocástico.

Definición 3.5.1 *Un proceso estocástico es una colección de variables aleatorias, $\{Z_t, t \in T\}$, definidas sobre un mismo espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$, donde T es un conjunto de índices.*

Para propósitos del presente curso, $T = \mathbb{Z}_+$ (ó \mathbb{R}_+) y todas las variables aleatorias Z_t , $t \in T$, toman valores en \mathbb{Z} .

Antes de definir la caminata aleatoria conviene recordar la definición de independencia de una colección de variables aleatorias. Se dice que $(X_n, n \geq 1)$ es una sucesión de variables independientes, si para cada n -éada de enteros (k_1, k_2, \dots, k_n) distintos se cumple que las variables aleatorias $X_{k_1}, X_{k_2}, \dots, X_{k_n}$ son independientes.

Ahora ya tenemos todos los elementos necesarios para definir nuestro proceso estocástico de interés:

Definición 3.5.2 *Sea $(X_n, n \geq 1)$ una sucesión de variables aleatorias i.i.d. con distribución común dada por*

$$\mathbb{P}(X_1 = 1) = 1 - \mathbb{P}(X_1 = -1) = 1 - q = p.$$

La sucesión $(S_n, n \geq 0)$, donde

$$S_n = S_0 + \sum_{i=1}^n X_i,$$

es llamada *caminata aleatoria simple*. En general, S_0 puede ser constante o una variable aleatoria, se dice que la caminata inicia en S_0 . Si $p = q = \frac{1}{2}$ es llamada *caminata aleatoria simple simétrica*.

Las caminatas aleatorias son útiles para modelar varios fenómenos: podemos usarlo para modelar la posición de una partícula que se mueve en los enteros, a cada paso la partícula puede avanzar o retroceder por un paso con probabilidad p y $1 - p$, respectivamente. Además, la dirección (subir o bajar) es independiente de los pasos anteriores. Asimismo pueden servir para modelar un juego de apuestas donde en cada jugada se pierde o se gana una unidad.

Las caminatas aleatorias simples se grafican en el plano cartesiano con los puntos $(n, S_n)_{n=0}^{\infty}$ uniendo los puntos vecinos con líneas rectas con pendiente 1 ó -1. A la gráfica resultante se le llama *trayectoria o realización*, y dado que es trata de una sucesión de variables aleatorias, para $\omega \in \Omega$ se tiene una trayectoria o realización.

3.5.1. Propiedades de las caminatas aleatorias simples

Lema 3.5.3 *Toda caminata aleatoria simple $\{S_n, n \geq 0\}$, con $S_0 = a$, posee las siguientes propiedades:*

(i) *Homogeneidad espacial:*

$$\mathbb{P}(S_n = j | S_0 = a) = \mathbb{P}(S_n = j + b | S_0 = a + b).$$

(ii) Homogeneidad temporal, para todo $n, m \in \mathbb{Z}$:

$$\mathbb{P}(S_n = j | S_0 = a) = \mathbb{P}(S_{n+m} = j | S_m = a).$$

(iii) Propiedad de Markov, para todo $n, m \in \mathbb{Z}$,

$$\mathbb{P}(S_{n+m} = j | S_0, S_1, \dots, S_n) = \mathbb{P}(S_{n+m} = j | S_n). \quad (3.9)$$

Demostración: (i) Veamos el lado izquierdo

$$\mathbb{P}(S_n = j | S_0 = a) = \mathbb{P}\left(\sum_{i=1}^n X_i = j - a\right).$$

Análogamente, el lado derecho satisface

$$\mathbb{P}\left(\sum_{i=1}^n X_i = j - a\right) = \mathbb{P}(S_n = j + b, S_0 = a + b) = \mathbb{P}\left(\sum_{i=1}^n X_i = j + b - (a + b)\right).$$

□

(ii) Procederemos como en (i). El lado derecho es igual a

$$\begin{aligned} & \frac{\mathbb{P}(S_0 + \sum_{i=1}^{n+m} X_i = j, S_0 + \sum_{i=1}^m X_i = a)}{\mathbb{P}(S_0 + \sum_{i=1}^m X_i = a)} \\ &= \frac{\mathbb{P}(\sum_{i=m+1}^{n+m} X_i = j - a, S_0 + \sum_{i=1}^m X_i = j)}{\mathbb{P}(S_0 + \sum_{i=1}^m X_i = j)} \\ &= \mathbb{P}\left(\sum_{i=m+1}^{n+m} X_i = j - a\right) \text{ (independencia)} \\ &= \mathbb{P}\left(\sum_{i=1}^n X_i = j - a\right), \end{aligned}$$

la última igualdad es debido al hecho que el vector (X_1, \dots, X_n) tiene la misma distribución que el vector $(X_{m+1}, X_{m+1}, \dots, X_{n+m})$. Un cálculo similar, pero más simple, demuestra la igualdad deseada. □

(iii) Sean s_0, s_1, \dots, s_n enteros tales que

$$\begin{aligned} & \mathbb{P}(S_{n+m} = j | S_0 = s_0, S_1 = s_1, \dots, S_n = s_n) \\ &= \frac{\mathbb{P}(S_0 = s_0, S_1 = s_1, \dots, S_n = s_n, S_{n+m} = s_{n+m})}{\mathbb{P}(S_0 = s_0, S_1 = s_1, \dots, S_n = s_n)} \\ &= \frac{\mathbb{P}(S_0 = s_0, X_1 = s_1 - s_0, X_2 = s_2 - s_1, \dots, X_n = s_n - s_{n-1}, \sum_{i=n+1}^{n+m} X_i = s_{n+m} - s_n)}{\mathbb{P}(S_0 = s_0, X_1 = s_1 - s_0, X_2 = s_2 - s_1, \dots, X_{n-1} = s_{n-1} - s_{n-2}, X_n = s_n - s_{n-1})} \\ &= \mathbb{P}\left(\sum_{i=n+1}^{n+m} X_i = s_{n+m} - s_n\right). \end{aligned}$$

Por otro lado, tenemos que

$$\begin{aligned}\mathbb{P}(S_{n+m} = s_{n+m} | S_n = s_n) &= \frac{\mathbb{P}(S_{n+m} = s_{n+m}, S_n = s_n)}{\mathbb{P}(S_n = s_n)} \\ &= \mathbb{P}\left(\sum_{i=n+1}^{n+m} X_i = s_{n+m} - s_n\right).\end{aligned}$$

□

Observación 3.5.4 (i) *Cualquier proceso estocástico, $\{Z_n, n \geq 0\}$, que cumpla la propiedad (3.9) es llamado cadena de Markov a tiempo discreto, es decir,*

$$\mathbb{P}(Z_{n+m} = j | Z_0, Z_1, \dots, Z_n) = \mathbb{P}(Z_{n+m} = j | Z_n).$$

(ii) *Se dirá que la probabilidad $\mathbb{P}(S_n = j | S_0 = a)$ es la probabilidad de transición del estado a al estado j en n pasos.*

En el siguiente resultado calcularemos las probabilidades de transición para la caminata aleatoria simple.

Lema 3.5.5 *Para todo $a, b \in \mathbb{Z}$ y $n \geq 0$, se tiene que*

$$\mathbb{P}(S_n = b | S_0 = a) = \begin{cases} \binom{n}{(n+b-a)/2} p^{(n+b-a)/2} q^{(n-b+a)/2} & \text{si } (n+b-a)/2 \in \mathbb{Z}, \\ 0 & \text{en otro caso.} \end{cases}$$

Demostración: Se tiene que, una realización que lleva del punto $(0, a)$ al punto (n, b) en n pasos tiene r pasos hacia arriba $(+1)$ y l pasos hacia abajo (-1) , donde r, l son tales que $l + r = n$ y $r - l = b - a$. Lo anterior es debido a que, $S_n = r(+1) + l(-1) = b - a$. Resolviendo las ecuaciones anteriores obtenemos que,

$$r = \frac{n + b - a}{2} \text{ y } l = \frac{n - b + a}{2}.$$

Ahora bien, cada realización que lleva de a a b en n pasos tiene probabilidad $p^r q^l$, y hay $\binom{n}{(n+b-a)/2}$ realizaciones posibles. Por lo que el resultado se sigue. □

Note que, la prueba del resultado anterior se basa en el conteo de trayectorias, i.e., “casos favorables” / “casos posibles”. Esta es una propiedad de muy interesante y que ha llamado la atención no sólo de la comunidad probabilista sino que también es explotada en teoría combinatoria, teoría de juegos, entre otras.

En lo que sigue procederemos a calcular probabilidades asociadas a la caminata aleatoria simple mediante las herramientas estudiadas hasta el momento. A saber, por medio de condicionamientos.

Definición 3.5.6 *Para cada $j \in \mathbb{Z}$, el primer tiempo de llegada al estado j se define por*

$$T_j = \min\{n \geq 0 : S_n = j\}.$$

Proposición 3.5.7 Para cada $j \in \mathbb{Z}$, sea h_j la probabilidad de que una caminata aleatoria que parte del estado j llegue al estado 0 antes de llegar al estado N , i.e., $h_j = \mathbb{P}(T_0 < T_N | S_0 = j)$. Entonces,

$$h_j = \begin{cases} \frac{\left(\frac{q}{p}\right)^j - \left(\frac{q}{p}\right)^N}{1 - \left(\frac{q}{p}\right)^N} & p \neq q, \\ 1 - \frac{j}{N} & p = \frac{1}{2}. \end{cases}$$

Demostración: Condicionando en la primera transición obtenemos la ecuación,

$$h_j = ph_{j+1} + qh_{j-1},$$

para $j \in \{1, 2, \dots, N-1\}$. Además, notemos que $h_0 = 1$ y $h_N = 0$. Reescribiendo la ecuación anterior se obtiene

$$h_n = ph_{n+1} + qh_{n-1} \iff q(h_{n+1} - h_n) = p(h_{n+1} - h_n), \quad n \geq 1. \quad (3.10)$$

El caso simétrico: $p = q = 1/2$. En este caso, se tiene la ecuación

$$h_n - h_{n-1} = h_{n+1} - h_n, \quad n \geq 1.$$

Por lo tanto, la recta h_n tiene una pendiente constante $c := h_{n+1} - h_n$, en consecuencia

$$h_n = 1 + \sum_{j=1}^n (h_j - h_{j-1}) = 1 + nc, \quad 1 \leq n \leq N.$$

Ahora bien, recordando que $h_N = 0$ obtenemos que $c = -1/N$, i.e., $h_n = 1 - n/N$.

El caso general: $p \neq q$. Definamos la sucesión $(x_n, n \geq 0)$ como sigue, $x_0 \in \mathbb{R}$ (se determinará más adelante) y $x_n = h_n - h_{n-1}$, para $1 \leq n \leq N$. De la ecuación en el lado derecho de (3.10) obtenemos que la sucesión $(x_n, n \geq 0)$ satisface la relación $x_{n+1} = \frac{q}{p}x_n$, para $1 \leq n \leq N$. Por lo tanto,

$$x_{n+1} = \left(\frac{q}{p}\right)^n x_0, \quad 0 \leq n \leq N. \quad (3.11)$$

Luego, dado que

$$h_n = h_0 + \sum_{j=1}^n (h_j - h_{j-1}),$$

la ecuación (3.11) implica

$$\begin{aligned} h_n &= h_0 + x_0 \sum_{j=1}^n \left(\frac{q}{p}\right)^j \\ &= h_0 + x_0 \left(\frac{q}{p}\right) \frac{1 - \left(\frac{q}{p}\right)^n}{1 - \left(\frac{q}{p}\right)}. \end{aligned} \quad (3.12)$$

Haciendo uso del hecho que, $h_N = 0$ y $h_0 = 1$, obtenemos que

$$0 = 1 + x_0 \binom{q}{p} \frac{1 - \left(\frac{q}{p}\right)^n}{1 - \left(\frac{q}{p}\right)},$$

de donde se sigue que

$$x_0 = -\frac{p}{q} \frac{1 - \frac{q}{p}}{1 - \left(\frac{q}{p}\right)^N}.$$

Finalmente, de (3.12) se concluye que

$$h_n = \frac{\left(\frac{q}{p}\right)^n - \left(\frac{q}{p}\right)^N}{1 - \left(\frac{q}{p}\right)^N}.$$

□

Corolario 3.5.8 *Para cada $j \in \mathbb{N}$ se tiene que*

$$\mathbb{P}(T_0 < \infty | S_0 = j) = \begin{cases} 1 & \text{si } q \leq p, \\ \left(\frac{q}{p}\right)^j & \text{si } q > p. \end{cases}$$

Demostración: Para cada n , sea $A_n := \{T_0 < T_n\}$. Notemos que $A_n \subset A_{n+1}$, dado que $T_n \leq T_{n+1}$, para cada n . Además, observemos que

$$\{T_0 < \infty\} = \cup_{n=1}^{\infty} \{T_0 < T_n\}.$$

Por lo tanto, dado que (A_n) es una sucesión creciente, la continuidad de la medida de probabilidad implica

$$\lim_{n \rightarrow \infty} \mathbb{P}(A_n | S_0 = j) = \mathbb{P}(T_0 < \infty | S_0 = j).$$

Luego, el resultado se sigue de la proposición anterior. □

Sean $a, b \in \mathbb{Z}$ y $n \in \mathbb{N}$. Sea $N_{n(a,b)}$ el número de trayectorias que van de a a b en n pasos y $N_{n(a,b)}^0$ las trayectorias que unen a y b en n pasos; y que además, pasan por 0 al menos una vez.

Teorema 3.5.9 (*Principio de Reflexión*) *Para cada $a, b \in \mathbb{N}$ se tiene que*

$$N_{n(a,b)}^0 = N_{n(-a,b)}.$$

Demostración: Haciendo bosquejo podemos ver que cada trayectoria que lleva de $(0, a)$ a (b, n) cruza el eje x por lo menos una vez, sea $(k, 0)$ el punto donde esto ocurre por primera vez. Reflejando el segmento de la trayectoria anterior se obtiene una trayectoria de $(0, a)$ a (b, n) y que pasa por el eje x por lo menos una vez. Luego, haciendo lo mismo en el sentido opuesto obtenemos el resultado. □

Lema 3.5.10 Para todo $a, b \in \mathbb{Z}$, se cumple que

$$N_{n(a,b)} = \binom{n}{\frac{1}{2}(n+b-a)}.$$

Veamos el siguiente resultado importante, el cual es una consecuencia del lema anterior.

Teorema 3.5.11 (Teorema de las votaciones (Ballot Theorem)) Sea $b \in \mathbb{N}$, entonces el número de realizaciones que van de $(0, 0)$ a (n, b) y que no visitan al eje x después del primer paso está dado por

$$\frac{b}{n} N_{n(0,b)}.$$

Demostración: Notemos que, las trayectorias que nos interesa contar en el primer paso se encuentran en $(1, 1)$. Por lo tanto, en número de trayectorias de interés está dado por

$$\begin{aligned} N_{n-1(1,b)} - N_{n-1(1,b)}^0 &= N_{n-1(1,b)} - N_{n-1}(-1, b) \\ &= \frac{(n-1)!}{\left(\frac{n-b}{2}\right)! \left(\frac{n+b-2}{2}\right)!} - \frac{(n-1)!}{\left(\frac{n-b-2}{2}\right)! \left(\frac{n+b}{2}\right)!} \\ &= \frac{(n-1)!}{\left(\frac{n-b}{2}\right)! \left(\frac{n+b}{2}\right)!} \left(\frac{n+b}{2} - \frac{n-b}{2} \right) \\ &= \frac{b}{n} N_{n(0,b)}. \end{aligned}$$

□

Veamos ahora porque el resultado anterior se llama Teorema de las votaciones. Supongamos que tenemos dos candidatos A y B , y que A obtiene a votos y B obtiene b votos, donde $a > b$. Cual es la probabilidad de que A tenga la ventaja durante toda la votación?

Supongamos que $X_i = 1$ si el i -ésimo individuo vota por el candidato A y vale -1 si vota el candidato B . Supongamos que cualquier combinación de votos es igualmente probable, i.e., cada una tiene probabilidad $\binom{\alpha+\beta}{\alpha}$. La trayectoria que deben seguir las votaciones para que A tenga las preferencias durante toda la jornada de votaciones va del punto $(0, 0)$ al punto $(\alpha + \beta, \alpha - \beta)$. Por lo tanto, por el Teorema 3.5.11 está dada por

$$\frac{\alpha - \beta}{\alpha + \beta} N_{\alpha+\beta(0, \alpha-\beta)} \frac{1}{\binom{\alpha+\beta}{\alpha}} = \frac{\alpha - \beta}{\alpha + \beta}.$$

El siguiente resultado es una aplicación del Principio de Reflexión (Teorema 3.5.9)

Teorema 3.5.12 Supongamos que $S_0 = 0$, entonces para todo $n \geq 0$ se cumple que

$$\mathbb{P}(S_1 S_2 \cdots S_n \neq 0, S_n = b) = \frac{|b|}{n} \mathbb{P}(S_n = b) \quad (3.13)$$

Demostración: Supongamos que $S_0 = 0$ y $S_n = b > 0$. Notemos que, $S_1 S_2 \cdots S_n \neq 0$ si y sólo si la caminata aleatoria no visita el eje x en el intervalo de tiempo $[1, n]$. Por lo tanto, por el Teorema 3.5.11 se tiene que el número de tales trayectorias es

$$\frac{b}{n} N_n(0, b)$$

y por argumentos similares a los del Lema 3.5.5 se sigue que hay $(n+b)/2$ pasos hacia arriba y $(n-b)/2$ pasos hacia abajo. Por lo tanto,

$$\begin{aligned}\mathbb{P}(S_1 S_2 \cdots S_n \neq 0, S_n = b) &= \frac{b}{n} N_n(0, b) p^{(n+b)/2} q^{(n-b)/2} \\ &= \frac{b}{n} \binom{n}{\frac{1}{2}(n+b)} p^{(n+b)/2} q^{(n-b)/2} \\ &= \frac{b}{n} \mathbb{P}(S_n = b | S_0 = 0).\end{aligned}$$

El caso $b < 0$ es similar, concluyendo así que

$$\mathbb{P}(S_1 S_2 \cdots S_n \neq 0, S_n = b) = \frac{|b|}{n} \mathbb{P}(S_n = b).$$

□

Observación 3.5.13 *Notemos que, la ecuación (3.13) implica*

$$\mathbb{P}(S_1 S_2 \cdots S_n \neq 0) = \frac{1}{n} \mathbb{E}(|S_n|).$$

Ahora vamos a analizar el comportamiento de los máximos de una caminata aleatoria. Sea $M_n := \{S_k : 1 \leq k \leq n\}$, $n \geq 1$, el máximo de (S_n) hasta el tiempo n . Tenemos el siguiente

Teorema 3.5.14 *Supongamos que $S_0 = 0$. Entonces, para cada $r \geq 1$, se cumple*

$$\mathbb{P}(M_n \geq r, S_n = b) = \begin{cases} \mathbb{P}(S_n = b), & \text{si } b \geq r, \\ \left(\frac{q}{p}\right)^{r-b} \mathbb{P}(S_n = 2r - b), & \text{si } b < r. \end{cases}$$

Demostración: Supongamos que $r \geq 1$ y que $b < r$, pues el caso $b \geq r$ es trivial. Sea $N_n^r(0, b)$ el número de realizaciones que van del $(0, 0)$ a (n, b) y que pasan por el estado r al menos una vez. Sea i_r el primer tiempo al cual la caminata visita el estado r , reflejando la trayectoria entre i_r y n en la recta r se obtiene una trayectoria que va de $(0, 0)$ a $(n, 2r - b)$. Ahora bien, a una de estas trayectorias le aplicamos la transformación inversa y obtenemos una que va de $(0, 0)$ a (n, b) y que además pasa por r . Entonces, se obtiene que

$$N_n^r(0, b) = N_n(0, 2r - b),$$

y sabemos que cada una de tales realizaciones tiene probabilidad $p^{(n+b)/2} q^{(n-b)/2}$. Por lo tanto,

$$\begin{aligned}\mathbb{P}(M_n \geq r, S_n = b) &= N_n^r(0, b) p^{(n+b)/2} q^{(n-b)/2} \\ &= \left(\frac{q}{p}\right)^{r-b} N_n(0, 2r - b) p^{(n+2r+b)/2} q^{(n-2r+b)/2} \\ &= \left(\frac{q}{p}\right)^{r-b} \mathbb{P}(S_n = 2r - b).\end{aligned}$$

□

Una pregunta interesante que podemos hacernos es la siguiente, ¿Cuál es la probabilidad de que (S_n) , $S_0 = 0$, alcance el nivel b por primera vez al tiempo n ? Sea $f_b(n)$ tal probabilidad.

Teorema 3.5.15 *Para cada $n \geq 1$ se cumple que*

$$f_b(n) = \frac{|b|}{n} \mathbb{P}(S_n = b).$$

Demostración: Supongamos que $b > 0$. Notemos que,

$$\begin{aligned} f_b(n) &\equiv \mathbb{P}(M_{n-1} = S_{n-1} = b-1, S_n = b) \\ &= (\mathbb{P}(M_{n-1} = S_{n-1} = b-1, S_n = b) \mathbb{P}(M_{n-1} = S_{n-1} = b-1)) \\ &= p \mathbb{P}(M_{n-1} = S_{n-1} = b-1) \\ &= p [\mathbb{P}(M_{n-1} \geq b-1, S_{n-1} = b-1) - \mathbb{P}(M_{n-1} \geq b, S_{n-1} = b-1)] \\ &= p \left[\mathbb{P}(S_{n-1} = b-1) - \frac{q}{p} \mathbb{P}(S_{n-1} = b-1) \right] \\ &= \frac{b}{n} \mathbb{P}(S_n = b), \end{aligned}$$

donde en la penúltima igualdad usamos el Teorema 3.5.14.

El caso $b < 0$ se obtiene de manera similar. □

Ejemplo 3.5.16 *Sea $(S_n)_{n \geq 0}$ una caminata aleatoria simple con $S_0 = 0$. Para cada $r \neq 0$ definamos V_r como el número de visitas al estado r antes de que la cadena regrese a su estado inicial.*

(i) *Demuestre que $\mathbb{E}(V_r) = 1$.*

(ii) *Dar un criterio para determinar si el número de visitas a 0 es finito o infinito.*

Solución: (i) Sea A_n el evento que “al tiempo n la caminata visita el estado r y no ha visitado el estado 0 hasta ese instante”. Entonces, $V_r \equiv \sum_{n \geq 1} \mathbf{1}_{A_n}$. Por otro lado, se tiene que

$$\mathbb{E}(\mathbf{1}_{A_n}) = \mathbb{P}(A_n) = \mathbb{P}(S_n = r, S_1 S_2 \cdots S_n \neq 0) = \frac{|r|}{n} \mathbb{P}(S_n = r) \equiv f_r(n).$$

Por lo tanto,

$$\mathbb{E}(V_r) = \mathbb{E}\left(\sum_{n \geq 1} \mathbf{1}_{A_n}\right) = \sum_{n \geq 1} f_r(n) = 1.$$

Vamos a ver demostrar la última igualdad. Note que,

$$\sum_{n \geq 1} f_r(n) \equiv \mathbb{P}(S_n = r, \text{ para algún } n) := f_r.$$

Condicionando en el primer salto, i.e., en S_1 obtenemos la ecuación

$$f_r = \frac{1}{2}(f_{b+1} - f_{b-1}), \quad b > 0,$$

con condición inicial $f_0 = 1$. Resolviendo la ecuación obtenemos que $f_b = 1$. Lo mismo se puede hacer para el caso $b < 0$.

(ii) Sea R el número total de visitas al estado 0. Notemos que, $R = \sum_{n \geq 0} \mathbf{1}_{\{S_n = 0\}}$, entonces

$$\mathbb{E}(R) = \sum_{n \geq 0} \mathbb{P}(S_n = 0) = \sum_{k \geq 0} \binom{2k}{k} p^k q^k. \quad (3.14)$$

Notemos que, $pq \leq 1/4$ y $pq = 1/4$ si y sólo si $p = 1/2$. Luego, usando la identidad de Stirling $n! \sim n^{n+\frac{1}{2}} e^{-n} \sqrt{2\pi}$, se tiene que para k suficientemente grande

$$\binom{2k}{k} \sim \frac{(2k)^{2k+1/2} e^{-2k}}{\sqrt{2\pi} (k^{k+1/2} e^{-2})^2} = (\sqrt{2\pi})^{-1} 2^{2k+1/2} k^{-1/2},$$

es decir, el término general en la serie está dado por la aproximación

$$\binom{2k}{k} p^k q^k \sim (\sqrt{\pi})^{-1} k^{-1/2} (4pq)^k.$$

Por lo tanto, la serie que aparece en (3.14) no es convergente para $p = 1/2$ ya que el término general es de orden de $k^{-1/2}$. Por otro lado, si $p \neq 1/2$, se tiene que $4pq < 1$, en consecuencia (3.14) es convergente. □

3.6. Ejercicios

1. Sea $\{S_n, n \geq 0\}$ la caminata aleatoria simple simétrica con $S_0 = 0$, y defina $T =: \{n \geq 1 : S_n = 0\}$ el primer tiempo de regreso al punto de inicio. Demuestre que

$$\mathbb{P}(T = 2n) = \frac{1}{2n-1} \binom{2n}{n} 2^{-2n}.$$

Deduzca de lo anterior que $\mathbb{E}(T^\alpha) < \infty$ si, y sólo si, $\alpha < \frac{1}{2}$. *Sugerencia:* recuerde la fórmula de Stirling, $n! \sim n^{n+\frac{1}{2}} e^{-n} \sqrt{2\pi}$.

2. Sea $\{S_n, n \geq 0\}$ la caminata aleatoria simple simétrica con $S_0 = 0$ y sea $M_n = \max_{n \geq 0} S_n$. Demuestre que

$$\mathbb{P}(M_n = r) = \mathbb{P}(S_n = r) + \mathbb{P}(S_n = r+1), \quad r \geq 0.$$

3. Sea $\{S_n, n \geq 0\}$ la caminata aleatoria simple simétrica con $S_0 = 0$.
 - a) Demuestre que

$$\mathbb{P}(S_1 S_2 \cdots S_{2m} \neq 0) = \mathbb{P}(S_{2m} = 0), \quad m \geq 1.$$

b) Sea $\alpha_{2n}(2k)$ la probabilidad de que la última visita a 0 antes del tiempo $2n$ ocurrió en el tiempo $2k$. Justifique que

$$\alpha_{2n}(2k) = \mathbb{P}(S_{2k} = 0) \mathbb{P}(S_1 S_2 \cdots S_{2n-2k} \neq 0).$$

c) Pruebe que

$$\alpha_{2n}(2k) = \mathbb{P}(S_{2k} = 0)\mathbb{P}(S_{2n-2k} = 0).$$

Capítulo 4

Funciones generadoras

4.1. Funciones generadoras de probabilidades

Definición 4.1.1 Sea $\mathbf{X} = (X_1, X_2, \dots, X_n)$ un vector aleatorio discreto con función de probabilidades conjuntada $f_{\mathbf{X}}(x_1, x_2, \dots, x_n)$, definimos la función generadora de probabilidades del vector \mathbf{X} por

$$\begin{aligned} G_{\mathbf{X}}(s_1, s_2, \dots, s_n) &= \sum_{x_1, x_2, \dots, x_n} s_1^{x_1} s_2^{x_2} \cdots s_n^{x_n} f_{\mathbf{X}}(x_1, x_2, \dots, x_n), \quad |s_i| \leq 1, \quad i = 1, \dots, n, \\ &\equiv \mathbb{E} [s_1^{X_1} s_2^{X_2} \cdots s_n^{X_n}]. \end{aligned}$$

De la definición anterior obtenemos que, la función generadora de probabilidades (f.g.p.) de X_i está dada por

$$G_{X_i}(s) = G_{\mathbf{X}}(1, \dots, 1, s, 1, \dots, 1) = \mathbb{E} [s^{X_i}], \quad |s| \leq 1,$$

donde s aparece en la i -ésima entrada.

Notemos que, en general $G_X(s)$ está bien definida para todo $|s| \leq 1$. En efecto,

$$|G_X(s)| \leq \left| \sum_x s^x \mathbb{P}(X = x) \right| \leq \sum_x |s|^x \mathbb{P}(X = x) \leq \sum_x \mathbb{P}(X = x) = 1.$$

Sin embargo, puede extenderse el rango de definición de G_X . Al número $R > 0$ tal que $|G_X(s)| < \infty$, $|s| < R$, se le llama *radio de convergencia*.

Ejemplo 4.1.2 (i) Supongamos que $X \sim \text{Bin}(n, p)$. Entonces,

$$\begin{aligned} G_X(s) &= \sum_{k=0}^n s^k \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=0}^n \binom{n}{k} (ps)^k (1-p)^{n-k} \\ &= (q + ps)^n, \quad q := 1 - p. \end{aligned}$$

(ii) Si $X \sim \text{Poisson}(\lambda)$, se tiene

$$\begin{aligned} G_X(s) &= \sum_{k=0}^{\infty} s^k \frac{\lambda^k}{k!} e^{-\lambda} \\ &= e^{-\lambda} e^{-\lambda s} = e^{-\lambda(1+s)} \end{aligned}$$

Note que, en ambos ejemplos $R = \infty$.

Vamos a ver ahora la utilidad de la f.g.p.

Teorema 4.1.3 Sea X una variable aleatoria no negativa tal que $\mathbb{P}(X = n) \equiv p_n$, $n = 0, 1, \dots$ y con función generadora de probabilidades G . Entonces,

(i) $G(s)$ es diferenciable en todo $|s| < 1$, y su derivada está dada por

$$G'(s) = \sum_{n=1}^{\infty} np_n s^{n-1}.$$

Para $s = 1$,

$$G'(1) := \lim_{s \uparrow 1} \sum_{n=1}^{\infty} np_n s^{n-1} \text{ finito o infinito.}$$

(ii) Para cada $k \geq 1$, se tiene la derivada k -ésima está dada por

$$G^{(k)} = \sum_{n=1}^{\infty} \frac{n!}{(n-k)!} p_n s^{n-k}.$$

(iii) G determina la distribución de X , es decir, $(p_n)_{n \geq 0}$.

Demostración: Sólo vamos a demostrar parte (iii). Por definición se tiene que

$$G(0) = \mathbb{P}(X = 0) = p_0.$$

Ahora bien, para cada $k \geq 1$,

$$G^{(k)}(0) = \lim_{s \downarrow 0} \sum_{n=1}^{\infty} \frac{n!}{(n-k)!} p_n s^{n-k} = kp_k,$$

entonces

$$p_k = \frac{1}{k!} G^{(k)}(0).$$

Por lo tanto, G determina $(p_k)_{k \geq 0}$. □

La parte (iii) del teorema anterior nos dice que, para conocer (su distribución) a una variable aleatoria es suficiente con conocer su f.g.p. Por otro lado, conociendo la distribución de una variable aleatoria se determina su f.g.p.

Como un corolario del teorema anterior tenemos el siguiente resultado.

Corolario 4.1.4 Sea X una variable aleatoria con función generadora de probabilidades G . Entonces,

$$\mathbb{E}(X) = G'(1).$$

Más generalmente, el k -ésimo momento factorial, $\mu^{(k)}$, de X está dado por

$$\begin{aligned}\mu^{(k)} &:= \mathbb{E}[X(X-1)(X-2)\cdots(X-k+1)] \\ &= G^{(k)}(1).\end{aligned}$$

En particular, del corolario anterior se sigue que

$$\text{Var}(X) = G^{(2)}(1) + G'(1) - (G'(1))^2 \quad (4.1)$$

El siguiente resultado nos habla de la función generadora de probabilidades conjunta cuando hay independencia.

Teorema 4.1.5 Suponga que X y Y tienen función generadora de probabilidades conjunta $G(s, t)$. Entonces, X y Y son independientes si y sólo si $G(s, t) = G(s, 1)G(1, t)$.

Demostración: \Rightarrow) Por definición tenemos que

$$\begin{aligned}G(s, t) &= \mathbb{E}(s^X t^Y) \\ &= \mathbb{E}(s^X) \mathbb{E}(t^Y), \text{ (independencia)} \\ &= G(s, 1)G(1, t).\end{aligned}$$

\Leftarrow) Notemos que,

$$\begin{aligned}G(s, 1)G(1, t) &= \left(\sum_x s^x \mathbb{P}(X = x) \right) \left(\sum_y t^y \mathbb{P}(Y = y) \right) \\ &= \sum_x \sum_y s^x t^y \mathbb{P}(X = x) \mathbb{P}(Y = y).\end{aligned}$$

Por otro lado,

$$G(s, t) = \sum_{x,y} s^x t^y \mathbb{P}(X = x, Y = y).$$

Luego, para que se cumpla $G(s, t) = G(s, 1)G(1, t)$ se debe tener que

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y), \text{ para todo } x, y,$$

la última relación es justamente la definición de independencia. \square

Ejemplo 4.1.6 (Continuando con el Ejemplo 3.1.7) Una gallina pone X huevos, donde X es Poisson con parámetro λ . Cada huevo es fecundado con probabilidad p , independientemente de los otros. Sea Y el número de huevos fecundados y Z los restantes, produciendo así Y pollos. Demuestre que Y y Z son independientes.

Solución: Condicionalmente en $X = x$, $Y \sim \text{Bin}(x, p)$. Entonces,

$$\mathbb{E}[s^Y | X = x] = (ps + q)^x.$$

Luego,

$$\begin{aligned} \mathbb{E}(s^Y t^Z) &= \mathbb{E}(s^Y t^{X-Y}) \\ &= \mathbb{E} \{ \mathbb{E}[(s/t)^Y t^X | X] \} \\ &= \mathbb{E} \{ t^X \mathbb{E}[(s/t)^Y | X] \} \\ &= \mathbb{E}[t^X (ps/t + q)^X] \\ &= \mathbb{E}[(ps + qt)^X]. \end{aligned}$$

Recordando que X es tiene distribución Poisson tenemos que

$$\mathbb{E}(s^Y t^Z) = \exp\{\lambda(ps + qt - 1)\} = \exp\{\lambda p(s - 1)\} \exp\{\lambda q(t - 1)\},$$

usando el teorema anterior obtenemos que Y y Z son independientes. Además, se observa que $Y \sim \text{Poisson}(\lambda p)$ y $Z \sim \text{Poisson}(\lambda q)$.

Vamos a concluir la sección con un resultado que será muy útil mas adelante.

Proposición 4.1.7 Sean N y $(X_i)_{i \geq 1}$ variables aleatorias, supongamos que N es no negativa y que, para cada $i \geq 1$

$$\mathbb{E}(s^{X_i}) = G(s),$$

es decir, las X_i 's tienen la misma distribución. Entonces, $Z := \sum_{i=1}^N X_i$ tiene función generadora de probabilidades

$$G_Z(s) = G_n(G(s)).$$

Demostración: Condicionando tenemos que,

$$\begin{aligned} \mathbb{E}(s^Z) &= \mathbb{E}[\mathbb{E}(s^Z | N)] \\ &= \mathbb{E}[\mathbb{E}(s^{X_1}) \cdots \mathbb{E}(s^{X_N})] \\ &= \mathbb{E}[G(s)^N] \\ &= G_N(G(s)). \end{aligned}$$

□

4.2. Una breve introducción a procesos de Galtoa-Watson

Supongamos que una población de partículas (moléculas, virus, etc.) evoluciona de la siguiente manera. La población inicial al tiempo $n = 0$ con una partícula, al tiempo $n = 1$ dicha partícula muere y da origen a un número aleatorio (X) de partículas idénticas entre si y su progenitora y, a tiempos subsecuentes $n = 2, 3, \dots$, cada individuo evoluciona de la misma manera (muriendo y

ramificandose) produciendo así X partículas. Supondremos que X tiene función de probabilidades $f_X(k)$, $k = 0, 1, 2, \dots$. Vamos a suponer que el número de partículas que produce cada individuo es independiente de los demás y que tiene la misma distribución que X .

Vamos a denotar por Z_n el tamaño de la población al tiempo n , entonces $Z = \{Z_n : n \geq 0\}$ es un proceso estocástico en cual, a cada tiempo, no da el total de la población. Sea $(X_i^n, n \geq 0, i \geq 1)$ un colección de variables aleatorias independientes todas con función de probabilidades f_X . Entonces, el proceso Z se puede describir de la siguiente manera, $Z_0 = 1$ y

$$Z_{n+1} = \sum_{i=1}^{Z_n} X_i^n, \quad n = 0, 1, \dots, \quad (4.2)$$

donde X_i^n representa el número de descendientes que produce el i -ésimo individuo presente en la generación n . Una consecuencia de la independencia de la colección (X_i^n) es que el proceso Z es una cadena de Markov a tiempo discreto, ver la Observación 3.5.4 (i). El proceso Z es llamado *proceso de Galton-Watson*¹. El proceso de Galton-Watson ha sido fuente de inspiración para procesos de ramificación mucho más generales, los cuales conforman un área de investigación dentro de la teoría de probabilidad y procesos estocásticos por su riqueza en la variedad de modelos y su interacción con otras áreas de las matemáticas.

A principios del presente capítulo vimos que la función generadora de probabilidades es muy útil cuando se trabaja con variables aleatorias que toman valores en los enteros no negativos, que es caso del proceso de Galton-Watson. Supongamos que X tiene función generadora de probabilidades $G(s)$, $|s| < 1$ Entonces,

$$G_n(s) := \mathbb{E}(s^{Z_n}).$$

Notemos que $G_1(s) = G(s)$. Luego, por identidad (4.2) y la Proposición 4.1.7 se tiene que para $|s| \leq 1$

$$G_{n+1}(s) = G_n(G(s)), \quad \text{para todo } n \geq 1, \quad (4.3)$$

es decir, G_n es la convolución de G consigo misma n veces.

Proposición 4.2.1 *Supongamos que $\mu = \mathbb{E}(X)$ y $\sigma^2 = \text{Var}(X)$. Entonces, para cada $n \geq 1$,*

$$\mathbb{E}(Z_n) = \mu^n,$$

y

$$\text{Var}(Z_n) = \begin{cases} n\sigma^2, & \text{si } \mu = 1, \\ \sigma^2 \mu^{n-1} \frac{\mu^n - 1}{\mu - 1}, & \text{si } \mu \neq 1. \end{cases}$$

Demostración: Sabemos que $\mathbb{E}(Z_n) = G'_n(s)|_{s=1}$. Entonces, de (4.2) se sigue que

$$\mathbb{E}(Z_n) = \mu \mathbb{E}(Z_{n-1}) = \mu G'_{n-1}(1),$$

el resultado se sigue por iteración.

¹Francis Galton propuso la pregunta sobre la probabilidad de extinción de apellidos aristocráticos en Inglaterra en 1873 y Henry William Watson lo resolvió; y el 1874 escribieron el paper “On the probability of extinction of families”.

Ahora vamos a demostrar la segunda afirmación. Diferenciando dos veces en (4.3) y recordando que $G(1) = 1$ y se obtiene

$$G_n''(1) = G^{(2)}(1) \left[G_{n-1}^{(2)}(1) \right]^2 + G'(1)G_{n-1}^{(2)}(1).$$

Luego, el resultado se concluye usando la fórmula (4.1). En efecto, para $\mu = 1$ se tiene

$$G^{(2)}(1) = \sigma^2,$$

y

$$G_n^{(2)}(1) = \sigma^2 + G_{n-1}^{(2)}(1),$$

lo cual junto con el hecho $G_1'(1) = G'(1)$ implica

$$G_n^{(2)}(1) = \sigma^2 n, \quad n \geq 0.$$

...

□

En general hay muy pocos casos en los que se puede encontrar una expresión explícita para G_n . Uno de ellos es el caso en que la ramificación sigue una ley geométrica como lo muestra el siguiente

Ejemplo 4.2.2 (*Ramificación geométrica*) Supongamos que $G(s) = q(1 - ps)^{-1}$ ($p + q = 1$), $|s| < \frac{1}{p}$, es decir, $\mathbb{P}(X = k) = qp^k$, $k \geq 0$. En tal caso se tiene (Ver ejercicio 3)

$$G_n(s) = \begin{cases} \frac{n-(n-1)s}{n+1-ns}, & p = \frac{1}{2}, \\ \frac{q[p^n - q^n - ps(p^{n-1} - q^{n-1})]}{p^{n+1} - q^{n+1} - ps(p^n - q^n)}, & p \neq \frac{1}{2}. \end{cases} \quad (4.4)$$

Una de las preguntas importantes acerca del proceso Z es conocer la probabilidad de extinción al tiempo n , es decir, conocer $\mathbb{P}(Z_n = 0) = G_n(0)$ así como también $\lim \mathbb{P}(Z_n = 0)$. En el presente ejemplo se pueden encontrar de manera explícita. En efecto, de (4.4) tenemos que

$$\mathbb{P}(Z_n = 0) = \begin{cases} \frac{n}{n+1}, & p = q, \\ \frac{q(p^n - q^n)}{p^{n+1} - q^{n+1}}, & p \neq q. \end{cases}$$

Por lo tanto,

$$\lim_{n \rightarrow \infty} \mathbb{P}(Z_n = 0) = \begin{cases} 1, & p \leq q, \\ \frac{q}{p}, & p > q. \end{cases}$$

Observación 4.2.3 En el ejemplo anterior sabemos que $\mathbb{E}(Z_1) = p/q \leq 1$ si y sólo si $p \leq q$. Luego, en tal caso, como $\lim_{n \rightarrow \infty} \mathbb{E}(Z_n) = 0$ ya que $\mathbb{E}(Z_n) = [\mathbb{E}(Z_1)]^n$. Lo cual nos indica que, de alguna manera, Z_n tiende a cero cuando n tiende a infinito.

La propiedad anterior no es propia de caso geométrico como lo muestra en siguiente

Teorema 4.2.4 Se cumple que $\lim_{n \rightarrow \infty} \mathbb{P}(Z_n = 0) = \eta$ existe. Más aún, η es la menor raíz no negativa de la ecuación $G(s) = s$, $0 \leq s \leq 1$.

Demostración: Sea $\eta_n := \mathbb{P}(Z_n = 0)$, y sea η la menor raíz no negativa de de la ecuación $G(s) = s$, $0 \leq s \leq 1$. Vamos a demostrar que $\eta_n \rightarrow \eta$.

Consideremos los siguientes casos

1. Si $f_X(0) = 0$, entonces

$$\eta_n = G_n(0) = 0 = \eta.$$

2. Si $f_X(0) = 1$, entonces

$$\eta_n = G_n(1) = 1 = \eta.$$

3. Supongamos que $f_X(0) + f_X(1) = 1$ y $f_X(0)f_X(1) \neq 0$. Entonces,

$$\begin{aligned} \eta_n = G_n(0) &= 1 - \mathbb{P}(Z_n > 0) \\ &= 1 - [f(1)]^n \\ &\rightarrow 1, \text{ cuando } n \rightarrow \infty, \end{aligned}$$

y en este caso $\eta = 1$.

4. Finalmente, supongamos que $0 < f_X(0) < f_X(0) + f_X(1) < 1$. Notemos que

$$\{Z_n = 0\} \subset \{Z_{n+1} = 0\},$$

entonces

$$\eta_n = \mathbb{P}(Z_n = 0) \leq \mathbb{P}(Z_{n+1} = 0) = \eta_{n+1}.$$

Luego, (η_n) es una sucesión creciente y acotada, y por lo tanto es convergente. Sea $\lambda := \lim_{n \rightarrow \infty} \eta_n$.

Por otro lado, sabemos que $G_{n+1}(0) = G(G_n(0))$, entonces

$$\eta_n = G(\eta_n),$$

en consecuencia si hacemos n tender a infinito, por continuidad se sigue que

$$\lambda = G(\lambda).$$

Para concluir la prueba debemos demostrar que $\lambda = \eta$. Notemos que,

$$\eta_1 = G_1(0) = G(0) \leq G(\eta),$$

y

$$\eta_2 = G_2(0) = G(G(0)) = G(\eta_1) \leq G(\eta) = \eta.$$

Procediendo inductivamente obtenemos que $\eta_n \leq \eta$, entonces $\lambda \leq \eta$. Ahora bien, por hipótesis η es la menor raíz no negativa de la ecuación $G(s) = s$, $0 \leq s \leq 1$. Entonces, $\lambda \geq \eta$. Concluyendo así que

$$\lambda = \eta.$$

□

Puede demostrarse que $\eta = 1$ si $\mathbb{E}(X) < 1$ y $\eta < 1$ si $\mathbb{E}(X) > 1$. Si $\mathbb{E}(X) = 1$, entonces $\eta = 1$ siempre que $\text{Var}(X) > 0$, es decir, siempre que X no sea constante con probabilidad 1.

4.3. Ejercicios

1. Sea X una variable aleatoria con distribución de Poisson de parámetro Y , donde Y tiene distribución de Poisson con parámetro μ . Demuestre que

$$G_{X+Y}(s) = \exp\{\mu[se^{s-1} - 1]\}.$$

2. Sean X_0, X_1, X_2, \dots una variables aleatorias independientes e idénticamente distribuidas todas con distribución logarítmica, es decir,

$$\mathbb{P}(X = k) = \frac{(1-p)^k}{k \log(1/p)}, \quad k \geq 1,$$

donde $0 < p < 1$. Supong que N es independiente de las X_i 's y tiene distribución de Poisson con parámetro λ . Demuestre que $Y := \sum_{i=1}^N X_i$ tiene distribución binomial negativa.

Sedice que Z tiende distribución binomial negativa con parámetros $r \in \mathbb{N}$ y $p \in (0, 1)$ si

$$\mathbb{P}(Z = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}, \quad k = r+1, r+2, \dots$$

Sugerencia: recuerde que $\log(x) = (x-1) - \frac{(x-1)^2}{2} + \frac{(x-1)^3}{3} + \dots$.

3. Verifique la identidad (4.4) de Ejemplo 4.2.2.
4. Sea X una v.a. no-negativa con función generadora de probabilidades $G_X(s)$ tal que $G'_X(1) < \infty$. Demuestre que

$$G(s) = \frac{1}{\mathbb{E}(X)} \frac{1 - G_X(s)}{1 - s}$$

es la función generadora de probabilidades de alguna variable aleatoria Y . ¿Cuándo se tiene que $G(s) = G_X(s)$?

5. Sea (Z_n) un proceso de Galton-Watson con $Z_0 = 1$, $\mathbb{E}(Z_1) = \mu > 0$ y $\text{Var}(Z_1) > 0$. Demuestre que $\mathbb{E}(Z_n Z_m) = \mu^{n-m} \mathbb{E}(Z_m^2)$, $m \leq n$. Luego, encuentre $\rho(Z_n, Z_m)$ en términos de μ .
6. Sea (Z_n) con en ejercicio anterior. Sea G_n la función generadora de probabilidades de Z_n .
 - (a) Encuentre una expresión para G_n cuando la función generadora de Z_1 está dada por $G_1(s) \equiv G(s) = 1 - \alpha(1-s)^\beta$, $0 < \alpha, \beta < 1$.
 - (b) Encuentre $\mathbb{P}(Z_1 = k)$, $k = 0, 1, \dots$
7. Sea Z un proceso de Galton-Watson donde X (el número de descendientes) es tal que $\mathbb{P}(X = 0) = 2/5 = 1 - \mathbb{P}(X = 2)$. Encuentre la probabilidad de extinción de Z .

Capítulo 5

Procesos de Poisson

Hasta ahora hemos visto variables aleatorias, sucesiones de variables aleatorias. Veremos ahora una clase de variables aleatorias parametrizadas por un conjunto de índices, en éste caso el conjunto $[0, \infty)$ (proceso a tiempo continuo), el cuál será pensado con tiempo. En otras palabras, vamos estudiar como una colección de variables aleatorias evolucionan con el tiempo, tales clases de variables aleatorias son conocidas como *procesos estocásticos*. Según las propiedades que se le pidan a las variables aleatorias definirán distintas clases de procesos estocásticos.

El objetivo del presente capítulo es introducir al alumno un primer ejemplo de proceso estocástico, a saber el proceso de Poisson. Los procesos de Poisson sirven como modelos para contar ocurrencia de eventos en el tiempo.

5.1. Definición y propiedades básicas

Definición 5.1.1 *Un proceso de Poisson homogéneo de parámetro (o intensidad) $\lambda > 0$, es un proceso estocástico $(X_t, t \geq 0)$ tal que X_t toma valores en los enteros no-negativos, y además,*

1. $X_0 = 0$.

2. *Para toda sucesión finita de tiempos $0 \leq t_0 < t_1 \cdots, t_n$ se tiene que los incrementos*

$$X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \cdots, X_{t_n} - X_{t_{n-1}},$$

son variables aleatorias independientes.

3. *Para cada $s, t \geq 0$, la variable aleatoria $X_{t+s} - X_s$ tiene una distribución de Poisson con parámetro λt , es decir,*

$$\mathbb{P}(X_{t+s} - X_s = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad k = 0, 1, \dots$$

Una consecuencia inmediata de la definición es que, para todo $0 \leq s < t$ se tiene que

$$\mathbb{E}(X_t - X_s) = \text{Var}(X_t - X_s) = \lambda(t - s).$$

Observación 5.1.2 El proceso de Poisson se puede construir de la siguiente manera: sea $(\xi_i, i \geq 1)$ una sucesión de variables aleatorias independientes todas ellas con distribución exponencial de parámetro $\lambda > 0$. Definamos

$$T_0 = 0 \text{ y } T_n = \sum_{k=1}^n \xi_k, \quad n \geq 1.$$

Luego, sea

$$X_t = \#\{n > 0 : T_n \leq t\} = \sum_{i=1}^{\infty} \mathbf{1}_{\{T_i \leq t\}}, \quad t \geq 0.$$

Entonces, el proceso estocástico $\{X_t, t \geq 0\}$ es un proceso de Poisson de parámetro $\lambda > 0$.

Notemos que el tiempo T_n es el tiempo al cual ocurre el n -ésimo evento, dicho de otra forma, es tiempo que hay que esperar para observar el n -ésimo salto en el proceso de Poisson. Por lo tanto, T_1, T_2, \dots son llamados tiempos de espera.

Las siguientes equivalencias pueden ser muy útiles

$$\{X_t \leq n\} = \{T_n \geq t\}$$

y

$$\{X_t = n\} = \{T_n \leq t < T_{n+1}\}.$$

Ley de eventos raros. La distribución de Poisson aparece de manera natural en fenómenos de conteo. Más precisamente, aparece con límite de la distribución binomial de la siguiente manera: sea $X_{N,p} \sim \text{Bin}(N, p)$ y consideremos $N \rightarrow \infty$ y $p \rightarrow 0$ de modo que $Np \rightarrow \lambda > 0$. Entonces,

$$\lim_{Np \rightarrow \mu} \mathbb{P}(X_{N,p} = k) = \frac{\mu^k e^{-\lambda}}{k!}, \quad k = 0, 1, \dots$$

La propiedad anterior es conocida como *ley de eventos raros*.

El proceso de Poisson se puede pensar con un proceso que “cuenta” la ocurrencia de eventos. Tal interpretación es más clara cuando consideramos Procesos Puntuales de Poisson.

Definición 5.1.3 (*Procesos Puntuales de Poisson*) Sean $s < t$ y $N((s, t])$ una variable aleatoria que cuenta el número de “eventos” que ocurren en el intervalo $(s, t]$. Entonces, N es un proceso puntual de Poisson con intensidad $\lambda > 0$ si:

- Para cada $m = 1, 2, \dots$ y puntos distintos $t_0 = 0 < t_1 < t_2 < \dots < t_m$, las variables aleatorias

$$N((t_0, t_1]), N((t_1, t_2]), \dots, N((t_{m-1}, t_m]),$$

son independientes.

- Para todo $s < t$ la variable aleatoria $N((s, t])$ tiene distribución de Poisson de parámetro $\lambda(t - s)$, es decir,

$$\mathbb{P}(N((s, t]) = k) = \frac{[\lambda(t - s)]^k}{k!} e^{-\lambda(t - s)}, \quad k = 0, 1, \dots$$

Ejemplo 5.1.4 Suponga que la llegada de clientes a solicitar algún servicio sigue un proceso de Poisson con intensidad $\lambda = 2$ por unidad de tiempo. Sea $X(t)$ es número de clientes que han llegado hasta el tiempo t . Determine las siguientes probabilidades:

(a) $\mathbb{P}(X(1) = 2) = \frac{2^2 e^{-2}}{2!}$, dado que $X(2) \sim \text{Poisson}(2)$.

(b) $\mathbb{P}(X(1) = 2, X(3) = 6)$. Usando la propiedad de incrementos independientes obtenemos que la probabilidad deseada es igual a

$$\begin{aligned} \mathbb{P}(X(1) = 2, X(3) - X(1) = 4) &= \mathbb{P}(X(1) = 2)\mathbb{P}(X(3) - X(1) = 4) \\ &= \frac{2^2 e^{-2}}{2!} \frac{4^4 e^{-4}}{4!}, \end{aligned}$$

donde usamos el hecho que $X(3) - X(2) \sim \text{Poisson}(4)$.

(c) $\mathbb{P}(X(1) = 2 | X(3) = 6)$. Por definición de esperanza condicional tenemos que

$$\begin{aligned} \mathbb{P}(X(1) = 2 | X(3) = 6) &= \frac{\mathbb{P}(X(1) = 2, X(3) = 6)}{\mathbb{P}(X(3) = 6)} \\ &= \frac{\frac{2^2 e^{-2}}{2!} \frac{4^4 e^{-4}}{4!}}{\frac{6^6 e^{-6}}{6!}}. \end{aligned}$$

(d) $\mathbb{P}(X(3) = 6 | X(1) = 2)$. De manera similar a (c) obtenemos que

$$\mathbb{P}(X(3) = 6 | X(1) = 2) = \frac{\mathbb{P}(X(1) = 2)\mathbb{P}(X(3) - X(1) = 4)}{\mathbb{P}(X(1) = 2)} = \mathbb{P}(X(3) - X(1) = 4).$$

△

Ejemplo 5.1.5 Un sistema eléctrico recibe descargas de acuerdo a un proceso de Poisson con intensidad $\lambda > 0$. Suponga que el sistema sobrevive a cada descarga con probabilidad α , independientemente de las demás, de modo que la probabilidad de que sobreviva a la k -ésima descarga es α^k . ¿Cuál es la probabilidad de que el sistema esté operando al tiempo t ?

Solución: Sea $X(t)$ el número total de descargas recibidas hasta el tiempo t , entonces $\{X(t), t \geq 0\}$ es un proceso de Poisson de intensidad λ . Luego, sea $A := \{\text{“el sistema está operando al tiempo } t\}$, entonces por la ley de probabilidad total tenemos

$$\begin{aligned} \mathbb{P}(A) &= \sum_{n=0}^{\infty} \mathbb{P}(A | X(t) = n)\mathbb{P}(X(t) = n) \\ &= \sum_{n=0}^{\infty} \alpha^n \frac{(\lambda t)^n e^{-\lambda t}}{n!} \\ &= e^{-\lambda(1-\alpha)}. \end{aligned}$$

5.2. Distribuciones asociadas con el proceso de Poisson

Teorema 5.2.1 Sea T_n el n -ésimo tiempo de espera, i.e., el tiempo en el que transcurre el n -ésimo evento. Entonces, T_n tiene distribución gamma con función de densidad

$$f_{T_n}(t) = \frac{\lambda^n t^{n-1}}{(n-1)!} e^{-\lambda t}, \quad t \geq 0.$$

Demostración: Nótese que $T_n \leq t$ si, y sólo si, hasta el tiempo t han ocurrido al menos n eventos. Por lo tanto,

$$\begin{aligned} F_{T_n}(t) &= \mathbb{P}(T_n \leq t) \\ &= \mathbb{P}(X_t \geq n) \\ &= 1 - \mathbb{P}(X_t \leq n-1) \\ &= 1 - \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{K!} e^{-\lambda t}. \end{aligned}$$

Luego, derivando la última ecuación obtenemos que

$$f_{T_n}(t) = \frac{\lambda^n t^{n-1}}{(n-1)!} e^{-\lambda t}.$$

□

En el siguiente resultado usaremos la notación, $h(x) = o(g(x))$, cuando $x \rightarrow 0$, si

$$\lim_{x \rightarrow 0} \frac{h(x)}{g(x)} = 0.$$

Definamos $S_n = T_{n+1} - T_n$, para $n = 0, 1, \dots$, con $T_0 = 0$. La variable aleatoria S_n es llamada tiempo de permanencia en estado n .

Teorema 5.2.2 *Los tiempos de permanencia S_0, S_1, \dots, S_{n-1} son variables aleatorias independientes e idénticamente distribuidas con distribución exponencial de parámetro $\lambda > 0$, i.e.,*

$$f_{S_k}(s) = \lambda e^{-\lambda s}, \quad s \geq 0.$$

Demostración: Debemos demostrar que

$$f_{S_0, \dots, S_n}(s_0, \dots, s_n) = (\lambda e^{-\lambda s_0}) \dots (\lambda e^{-\lambda s_n}).$$

Vamos a demostrar el resultado para el caso $n = 1$. Nótese que, si S_0 y S_1 son tales que

$$s_0 < S_0 < s_0 + \Delta s_0 \text{ y } s_1 < S_1 < s_1 + \Delta s_1.$$

Entonces no debe ocurrir ningún evento en los intervalos

$$(0, s_0], \quad (s_0 + \Delta s_0, s_0 + \Delta s_0 + s_1]$$

y exactamente un evento debe ocurrir en los intervalos

$$(s_0, s_0 + \Delta s_0] \quad (s_0 + \Delta s_0 + s_1, s_0 + \Delta s_0 + s_1 + \Delta s_1].$$

Por lo tanto,

$$\begin{aligned} &f_{S_0, S_1}(s_0, s_1) \Delta s_0 \Delta s_1 \\ &= \mathbb{P}(s_0 < S_0 \leq s_0 + \Delta s_0, s_1 < S_1 \leq s_1 + \Delta s_1) + o(\Delta s_0 \Delta s_1) \\ &= \mathbb{P}(X(0, s_0] = 0) \mathbb{P}(X(s_0 + \Delta s_0, s_0 + \Delta s_0 + s_1] = 0) \\ &\quad \times \mathbb{P}(X(s_0, s_0 + \Delta s_0] = 1) \mathbb{P}(X(s_0 + \Delta s_0 + s_1, s_0 + \Delta s_0 + s_1 + \Delta s_1] = 1) \\ &\quad + o(\Delta s_0 \Delta s_1). \end{aligned}$$

Luego, dividiendo ambos lados de la identidad anterior por $\Delta s_0 \Delta s_1$ y, haciendo Δs_1 y Δs_1 tender a cero, obtenemos que

$$f_{s_0, s_1}(s_0, s_1) = (\lambda e^{-\lambda s_0}) (\lambda e^{-\lambda s_1}).$$

□

Teorema 5.2.3 *Sea $\{X_t, t \geq 0\}$ un proceso de Poisson de parámetro $\lambda > 0$, y $0 \leq s < t$. Entonces, para cada $0 \leq k \leq n$*

$$\mathbb{P}(X_s = k | X_t = n) = \binom{n}{k} \left(\frac{u}{t}\right)^k \left(1 - \frac{u}{t}\right)^{n-k}.$$

Demostración: Por definición de probabilidad condicional se tiene que

$$\begin{aligned} \mathbb{P}(X_u = k | X_t = n) &= \frac{\mathbb{P}(X_u = k, X_t = n)}{\mathbb{P}(X_t = n)} \\ &= \frac{\mathbb{P}(X_u = k, X_t - X_u = n - k)}{\mathbb{P}(X_t = n)} \\ &= \frac{\mathbb{P}(X_u = k) \mathbb{P}(X_t - X_u = n - k)}{\mathbb{P}(X_t = n)} \quad (\text{prop. de incrementos ind.}) \end{aligned}$$

Por lo tanto,

$$\begin{aligned} \mathbb{P}(X_u = k | X_t = n) &= \frac{\frac{(\lambda u)^k e^{-\lambda u}}{k!} \frac{(\lambda(t-u))^{n-k} e^{-\lambda(t-u)}}{(n-k)!}}{\frac{(\lambda t)^n e^{-\lambda t}}{n!}} \\ &= \binom{n}{k} \left(\frac{u}{t}\right)^k \left(1 - \frac{u}{t}\right)^{n-k}. \end{aligned}$$

□

El teorema anterior no dice que, dado que sabemos que hasta el tiempo t han ocurrido n eventos, entonces éstos se distribuyen de manera uniforme en el intervalo $(0, t)$. El siguiente teorema reafirma el resultado anterior. Pues nos dice que, dado que $X_t = n$, entonces los tiempos de ocurrencia de eventos se distribuyen con los estadísticos de orden de una muestra uniforme.

Teorema 5.2.4 *Dado que $X_t = n$, los tiempos de llegada T_1, T_2, \dots, T_n tienen la misma distribución que los estadísticos de orden de una muestra de tamaño n de una distribución uniforme en $(0, t)$.*

Demostración: Nos interesa conocer la distribución condicional de T_1, T_2, \dots, T_n dado $X_t = n$. Es claro que $T_1 < T_2 < \dots < T_n$. Nótese que, el eventos

$$T_1 = t_1, T_2 = t_2, \dots, T_n = t_n, X_t = n$$

es equivalente al evento

$$S_1 = t_1, S_2 = t_2 - t_1, \dots, S_n = t_n - t_{n-1}, S_{n+1} > t - t_n.$$

Por lo tanto,

$$\begin{aligned}
 f(t_1, \dots, t_n | n) &:= \frac{f(t_1, \dots, t_n, n)}{\mathbb{P}(X_t = n)} \\
 &= \frac{\lambda e^{-\lambda t_1} \lambda e^{-\lambda(t_2 - t_1)} \dots \lambda e^{-\lambda(t_n - t_{n-1})} \lambda e^{-\lambda(t - t_n)}}{\frac{(\lambda t)^n}{n!} e^{-\lambda t}} \\
 &= \frac{n!}{t^n}, \quad 0 < t_1 < t_2 < \dots < t_n < \infty.
 \end{aligned}$$

Para concluir la prueba basta notar que la expresión anterior es precisamente la función de densidad conjunta de los estadísticos de orden de una muestra aleatoria de tamaño n de una distribución uniforme en $(0, t)$. \square

Ejemplo 5.2.5 *Un servicio recibe clientes de acuerdo a un proceso de Poisson con intensidad $\lambda > 0$. Cada cliente paga \$1 a su llegada. Es de interés evaluar la suma de dinero acumulada en el intervalo de tiempo $(0, t]$ descontada al tiempo 0, donde la tasa de descuento es β .*

Solución: *denotemos por M al valor esperado que nos interesa conocer. Entonces,*

$$M = \mathbb{E} \left[\sum_{k=1}^{X(t)} e^{-\beta T_k} \right],$$

donde T_n representa el tiempo de llegada del n -ésimo cliente y $X(t)$ es el total de clientes hasta el tiempo t . Luego, de nuevo por ley de probabilidad total tenemos que

$$M = \sum_{n=1}^{\infty} \mathbb{E} \left[\sum_{k=1}^n e^{-\beta T_k} \middle| X(t) = n \right] \mathbb{P}(X(t) = n). \quad (5.1)$$

Ahora bien, por el Teorema 5.2.4 se tiene que

$$\mathbb{E} \left[\sum_{k=1}^n e^{-\beta T_k} \middle| X(t) = n \right] = \mathbb{E} \left[\sum_{k=1}^n e^{-\beta U_k} \right],$$

donde U_1, U_2, \dots, U_n son variables aleatorias independientes e idénticamente distribuidas, con $U_k \sim \text{Unif}(0, t)$. Por lo tanto,

$$\mathbb{E} \left[\sum_{k=1}^n e^{-\beta T_k} \middle| X(t) = n \right] = n \mathbb{E} [e^{-\beta U_1}] = \frac{n(1 - e^{-\beta t})}{t}.$$

Entonces, usando el hecho que $\{X(t), t \geq 0\}$ es un proceso de Poisson con intensidad λ y la relación (5.1) se sigue que

$$M = \frac{\lambda}{\beta} (1 - e^{-\beta t}).$$

5.3. Proceso de Poisson compuesto

Sea $N := \{N(t), t \geq 0\}$ un proceso de Poisson con intensidad λ y $(Z_i)_{i=1}^{\infty}$ un colección de variables aleatorias independientes e independientes del proceso N . Supongamos que cada Z_i tiene función de distribución F . El proceso de Poisson compuesto, $X := \{X(t), t \geq 0\}$, se define como sigue

$$X(t) = \sum_{i=1}^{N(t)} Z_i, \text{ para cada } t \geq 0. \quad (5.2)$$

Un par de ejemplos donde el proceso de Poisson compuesto puede servir como modelo:

1. **Teoría de riesgo.** Una compañía aseguradora recibe reclamaciones de acuerdo a un proceso de Poisson N y Z_k denota el monto de la k -ésima reclamación. Luego, $X(t) = \sum_{k=1}^{N(t)} Z_k$ es el monto acumulado de reclamaciones hasta el tiempo t .

El proceso de Poisson compuesto es la piedra angular del llamado *modelo clásico de riesgo*, el cual se define por

$$Z(t) = z + ct - X(t), \quad t \geq 0,$$

donde z representa el capital inicial de la compañía y c es la prima que cobra la aseguradora (i.e., cantidad de dinero que recibe por unidad de tiempo). En tal modelo $Z(t)$ representa el capital de la compañía al tiempo t .

El modelo clásico de riesgo puede ser objeto de estudio de un curso completo de *teoría de riesgo*.

2. **Nivel de inventarios.** Se surte la demanda de acuerdo a un proceso de Poisson y Z_k denota la magnitud de la k -ésima venta. Por lo tanto, $X(t)$ denota el total de mercancía sustraído de inventario.

De manera similar al caso cuando estudiamos la suma Poisson compuesta de hecho, $X(t)$ es una suma Poisson compuesta, se puede dar una expresión para la distribución de $X(t)$ en términos de sus componentes. Más precisamente, sea $F(y) = \mathbb{P}(Z_i \leq y)$ y denotemos por

$$F^{*n}(y) := \mathbb{P}(Z_1 + \cdots + Z_n \leq y) = \int_{-\infty}^{\infty} F^{*(n-1)}(z - y) dF(y),$$

la convolución de Y_1, \dots, Y_n , y

$$F^{*0}(z) = \begin{cases} 1, & z \geq 0, \\ 0, & z < 0. \end{cases}$$

Entonces,

$$\begin{aligned}
 \mathbb{P}(X(t) \leq x) &= \mathbb{P}\left(\sum_{k=1}^{N(t)} Z_k \leq x\right) \\
 &= \sum_{n=0}^{\infty} \mathbb{P}\left(\sum_{k=1}^{N(t)} Z_k \leq x \mid N(t) = n\right) \mathbb{P}(N(t) = n) \\
 &= \sum_{n=0}^{\infty} \mathbb{P}\left(\sum_{k=1}^n Z_k \leq x\right) \mathbb{P}(N(t) = n) \\
 &= \sum_{n=0}^{\infty} \frac{(\lambda t)^n e^{-\lambda t}}{n!} F^{*n}(x),
 \end{aligned}$$

note que, en la tercera igualdad usamos la independencia entre la sucesión (Z_k) y el proceso de Poisson N .

Ejemplo 5.3.1 (Modelo de descargas) Sea $N(t)$ el número de descargas que recibe un sistema eléctrico hasta el tiempo t y Z_k el daño causado por la k -ésima descarga. Supondremos que el daño es positivo ($\mathbb{P}(Z_i \geq 0) = 1$), que los daños son independientes y que además son aditivos. Por lo tanto, el daño acumulado hasta el tiempo t es un proceso de Poisson compuesto

$$X(t) = \sum_{k=1}^{N(t)} Z_k, \quad t \geq 0.$$

Supongamos que el sistema sigue en operación siempre y cuando el daño sea menor o igual que algún valor crítico a , y falla en caso contrario. Encuentre el valor esperado del tiempo a la falla.

Solución: sea T el tiempo de falla. Entonces,

$$\{T > t\} \iff \{X(t) \leq a\}.$$

Luego,

$$\mathbb{P}(T > t) = \mathbb{P}(X(t) \leq a) = \sum_{n=0}^{\infty} \frac{(\lambda t)^n e^{-\lambda t}}{n!} F^{*n}(a),$$

donde en la última igualdad usamos ley de probabilidad total y la independencia entre las variables aleatorias.

Por otro lado, sabemos que $\mathbb{E}(T) = \int_0^{\infty} \mathbb{P}(T > t) dt$ ya que T es una variable aleatoria no

negativa. Entonces,

$$\begin{aligned}
 \mathbb{E}(T) &= \int_0^\infty \left(\sum_{n=0}^\infty \frac{(\lambda t)^n e^{-\lambda t}}{n!} F^{*n}(a) \right) dt \\
 &= \sum_{n=0}^\infty \left(\int_0^\infty \frac{(\lambda t)^n e^{-\lambda t}}{n!} dt \right) F^{*n}(a) \\
 &= \sum_{n=0}^\infty \frac{1}{\lambda} \left(\int_0^\infty \frac{\lambda^{n+1} t^n e^{-\lambda t}}{\Gamma(n+1)} dt \right) F^{*n}(a) \\
 &= \lambda^{-1} \sum_{n=0}^\infty F^{*n}(a),
 \end{aligned}$$

la última relación es gracias a que el integrando es la densidad de una variable aleatoria con distribución gama, y por lo tanto, el valor de la integral es 1. La segunda igualdad es porque podemos intercambiar suma con integral ya que los sumandos son no negativos.

Cuando $Z_k \sim \text{Exp}(\mu)$ podemos decir un poco más. En efecto, en tal caso $Z_1 + Z_2 + \dots + Z_n \sim \text{Gama}(n, \mu)$. Luego,

$$F^{*n}(z) = 1 - \sum_{k=0}^{n-1} \frac{(\mu z)^k}{k!} e^{-\mu z} = \sum_{k=n}^\infty \frac{(\mu z)^k}{k!} e^{-\mu z}.$$

Entonces,

$$\begin{aligned}
 \sum_{n=0}^\infty F^{*n}(a) &= \sum_{n=0}^\infty \sum_{k=n}^\infty \frac{(\mu z)^k}{k!} e^{-\mu z} \\
 &= \sum_{k=0}^\infty \sum_{n=0}^k \frac{(\mu z)^k}{k!} e^{-\mu z} \\
 &= \sum_{k=0}^\infty (1+k) \frac{(\mu z)^k}{k!} e^{-\mu z} \\
 &= 1 + \mu a.
 \end{aligned}$$

Por lo tanto,

$$\mathbb{E}(T) = \frac{1 + \mu a}{\lambda}.$$

□

5.4. Ejercicios

1. (Superposición de procesos de Poisson) Suponga que moscas y avispas caen en la sopa de acuerdo dos procesos de Poisson independientes con intensidad λ y μ , respectivamente. Argumente que el número de insectos que caen en la sopa forma un proceso de Poisson

con intensidad $\lambda + \mu$. El ejercicio consiste en verificar que la suma de procesos de Poisson independientes es un proceso de Poisson cuya intensidad es la suma de las intensidades de los sumandos.

2. Una masa radioactiva emite partículas de acuerdo a un proceso de Poisson con intensidad $\lambda = 2$ por minuto.
 - (i) ¿Cuál es la probabilidad de que la prime partícula aparezca después de los 3 minutos pero antes de los 5 minutos.
 - (ii) ¿Determine la probabilidad de que exactamente una partícula sea emitida en el intervalo de 3 a 5 minutos?
3. Sea $(N_t, t \geq 0)$ un proceso de Poisson con intensidad $\lambda > 0$, el cual es independiente de una variable aleatoria no-negativa T con media μ y varianza σ^2 . Encuentre
 - (i) $\text{Cov}(T, N(T))$.
 - (ii) $\text{Var}(N(T))$.
4. Suponga que la gente llega a la parada del autobús de acuerdo a un proceso de Poisson de parámetro $\lambda > 0$. El autobús parte a al tiempo t . Sea X el tiempo de espera total de los pasajeros que alcanzan el autobús. Nos interesa determinar $\text{Var}(X)$. Sea $N(t)$ el número de pasajeros que llegan al tiempo t .
 - (i) Encuentre $\mathbb{E}(X|N(t))$.
 - (ii) Verique que $\text{Var}(X|N(t)) = \frac{N(t)t^2}{12}$.
 - (iii) Determine $\text{Var}(X)$.
5. Sea $\{X(t), t \geq 0\}$ un proceso de Poisson con intensidad λ . Para cada $n = 1, 2, \dots$, encuentre la esperanza de T_1 dado que $X(1) = n$.
6. Un servicio recibe clientes de acuerdo a un proceso de Poisson de intensidad $\lambda > 0$. Supongamos que sabemos que durante la primera hora llegan 5 clientes. El tiempo que tarda cada cliente en servicio es una variable aleatoria con distribución exponencial de parámetro α independientemente de los demás, y luego se retira. ¿Determine la probabilidad de que la tienda quede vacía al terminar la primera hora?
7. Un sistema recibe descargas de acuerdo a un proceso de Poisson de parámetro λ . Cada descarga daña el sistema independientemente de las otras (también del número de descargas), y los daños se acumulan de manera aditiva. Determine la media y la varianza del daño total cuando los daños tienen distribución exponencial de parámetro θ .

Capítulo 6

Convergencia de variables aleatorias

El objetivo del presente capítulo es estudiar algunos tipos de convergencia de variables aleatorias. Iniciaremos con la definición de los distintos modos de convergencia.

6.1. Definiciones básicas

Definición 6.1.1 Sean X_1, X_2, \dots y X variables aleatorias definidas sobre el mismo espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$; con funciones de distribución F_1, F_2, \dots y F , respectivamente.

1. $\{X_n\}$ converge a X casi seguramente (c.s.), si

$$\mathbb{P}(\{\omega \in \Omega : X_n(\omega) \rightarrow_n X(\omega)\}) = 1.$$

Usaremos la siguiente notación: $X_n \xrightarrow[n]{c.s.} X$.

2. $\{X_n\}$ converge a X en media p -ésima, $p \geq 1$, si $\mathbb{E}(|X_n|^p) < \infty$, para toda n y

$$\mathbb{E}(|X_n - X|^p) \rightarrow_n 0.$$

Se usará la notación común: $X_n \xrightarrow[n]{L^p} X$.

3. $\{X_n\}$ converge en probabilidad a X , si

$$\mathbb{P}(|X_n - X| > \epsilon) \rightarrow_n 0, \text{ para todo } \epsilon > 0.$$

La convergencia en probabilidad se denota por: $X_n \xrightarrow[n]{\mathbb{P}} X$.

4. $\{X_n\}$ converge en distribución a X , si

$$F_n(x) \rightarrow_n F(x),$$

para todo $x \in \mathbb{R}$ tal que F es continua en x . Usaremos la notación: $X_n \xrightarrow[n]{d} X$.

¹Es importante hacer notar que, para este tipo de convergencia, no es necesario que las variables aleatorias estén definidas sobre el mismo espacio de probabilidad.

Notemos que, la convergencia casi segura es prácticamente la convergencia sucesiones en análisis ya que, dado $\omega \in \Omega$, se tiene que $\{X_n(\omega)\}$ es una sucesión de números reales. Por lo tanto, el conjunto donde $\{X_n(\omega)\}$ no converge tiene probabilidad cero. Por otro lado, convergencia en media p -ésima en análisis es conocida como convergencia en L^p .

Vale la pena mencionar que los tipos de convergencia definidos anteriormente no son equivalentes, de hecho, en general se cumple el siguiente

Teorema 6.1.2 *En general se cumple que*

$$X_n \xrightarrow{\text{c.s.}} X \Rightarrow X_n \xrightarrow{\mathbb{P}} X, \quad (6.1)$$

$$X_n \xrightarrow{L^p} X \Rightarrow X_n \xrightarrow{\mathbb{P}} X, \quad (6.2)$$

y

$$X_n \xrightarrow{\mathbb{P}} X \Rightarrow X_n \xrightarrow{d} X. \quad (6.3)$$

En este curso nos concentraremos en estudiar algunas de las propiedades de convergencia en probabilidad y convergencia en distribución. Cabe mencionar que, los otros modos de convergencia también son importantes en la teoría de probabilidad.

6.2. Convergencia en probabilidad

La primera cosa que nos gustaría saber es si el límite está bien definido. El objetivo de la siguiente proposición es precisamente dar respuesta a la pregunta anterior.

Proposición 6.2.1 (*Unicidad del límite*) Sea $\{X_n\}$ una sucesión variables aleatorias tal que $X_n \xrightarrow{\mathbb{P}} X$ y $X_n \xrightarrow{\mathbb{P}} Y$, entonces $\mathbb{P}(X = Y) = 1$.

Demostración: Por la desigualdad de triángulo tenemos que

$$|X - Y| \leq |X_n - X| + |X_n - Y|.$$

Entonces, para cada $\epsilon > 0$ se tiene que

$$\begin{aligned} \{|X - Y| > \epsilon\} &\subset \{|X_n - X| + |X_n - Y| > \epsilon\} \\ &\subset \{|X_n - X| > \epsilon/2\} \cup \{|X_n - Y| > \epsilon/2\}. \end{aligned}$$

Por lo tanto, por la definición de convergencia en probabilidad tenemos que

$$\mathbb{P}(|X - Y| > \epsilon) = 0, \text{ para todo } \epsilon > 0.$$

Finalmente, notemos que

$$\{|X - Y| > 0\} = \cup_{n=1}^{\infty} \{|X - Y| > 1/n\},$$

es decir,

$$\mathbb{P}(|X - Y| > 0) \leq \sum_{n=1}^{\infty} \mathbb{P}(|X - Y| > 1/n) = 0.$$

□

Proposición 6.2.2 Sean $\{X_n\}$ y $\{Y_n\}$ dos sucesiones de variables aleatorias tales que $X_n \xrightarrow{\mathbb{P}} X$ y $Y_n \xrightarrow{\mathbb{P}} Y$. Entonces, para todo $c \in \mathbb{R}$ se tiene que

$$cX_n + Y_n \xrightarrow{\mathbb{P}} cX + Y.$$

Demostración: Es claro que si $c = 0$ no hay nada que probar. Por lo tanto, supongamos que $c \neq 0$. Usando argumentos similares a los usados en la Proposición 6.2.1 se obtiene que, para todo $\epsilon > 0$

$$\{|cX_n - cX + Y_n - Y| > \epsilon\} \subset \{|cX_n - cX| > \epsilon\} \cup \{|Y_n - Y| > \epsilon\}.$$

Entonces,

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(|cX_n - cX + Y_n - Y| > \epsilon) \\ \leq \lim_{n \rightarrow \infty} [\mathbb{P}(|X_n - X| > \epsilon/|c|) + \mathbb{P}(|Y_n - Y| > \epsilon)] \\ = 0. \end{aligned}$$

□

Ejemplo 6.2.3 Sea $\Omega = (0, 1]$ y \mathbb{P} la medida que asigna a cada intervalo su longitud. Dicha medida es conocida con medida de Lebesgue sobre Ω . Para cada $n \in \mathbb{N}$, definimos $X_n = \mathbf{1}_{(0, 1/n)}$, es decir, para cada $\omega \in \Omega$

$$X_n(\omega) = \begin{cases} 1, & \text{si } \omega < 1/n, \\ 0, & \text{en otro caso.} \end{cases}$$

Demuestre que $X_n \xrightarrow{\mathbb{P}} 0$.

Solución: sea $\epsilon > 0$, entonces

$$\begin{aligned} \mathbb{P}(|X_n| > \epsilon) &= \mathbb{P}(X_n > \epsilon) \\ &= \begin{cases} 1/n, & \text{si } \epsilon < 1, \\ 0, & \text{en otro caso.} \end{cases} \end{aligned}$$

Por lo tanto,

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n| > \epsilon) = 0,$$

lo cual demuestra que $X_n \xrightarrow{\mathbb{P}} 0$.

◁

Ejemplo 6.2.4 Sean Ω y \mathbb{P} como en el ejemplo anterior. Considere la siguiente sucesión

$$X_n = \begin{cases} \mathbf{1}_{(0,1/n]}, & \text{si } n \text{ es impar,} \\ \mathbf{1}_{(1/n,1]}, & \text{si } n \text{ es par.} \end{cases}$$

En este caso, se tiene que $\{X_n\}$ no converge en probabilidad.

En efecto, sea $\epsilon > 0$, entonces para cada n impar se tiene que

$$\mathbb{P}(|X_n| > \epsilon) \begin{cases} 1/n, & \text{si } \epsilon < 1 \\ 0, & \text{en otro caso.} \end{cases}$$

Por lo tanto, $X_{2n+1} \xrightarrow[n]{\mathbb{P}} 0$.

Por otro lado, para n par se tiene que

$$\mathbb{P}(|X_n - 1| > \epsilon) \begin{cases} 1/n, & \text{si } \epsilon < 1 \\ 0, & \text{en otro caso.} \end{cases}$$

Lo cual implica que $X_{2n} \xrightarrow[n]{\mathbb{P}} 1$. ◁

Ejemplo 6.2.5 (Variables aleatorias i.i.d.) Sea $\{X_n\}$ un sucesión de variables aleatorias independientes e idénticamente distribuidas, con media μ y variación finita σ^2 . Definamos

$$S_n = \sum_{i=1}^n X_i.$$

Entonces, $\frac{S_n}{n} \xrightarrow[n]{L^2} \mu$.

Solución: notemos que

$$\begin{aligned} \mathbb{E} \left(\left| \frac{S_n}{n} - \mu \right|^2 \right) &= \mathbb{E} \left| \frac{1}{n} \sum_{i=1}^n (X_i - \mu) \right|^2 \\ &= \frac{1}{n^2} \sum_{i=1}^n \mathbb{E} (X_i - \mu)^2, \quad (\text{independencia}) \\ &= \frac{1}{n^2} \sum_{i=1}^n \sigma^2 \\ &= \frac{\sigma^2}{n}. \end{aligned}$$

Por lo tanto, haciendo $n \rightarrow \infty$, obtenemos que $\frac{S_n}{n} \xrightarrow[n]{L^2} \mu$. ◁

El ejemplo anterior contiene la demostración de la Ley débil de los grandes números, la cual enunciamos en el siguiente

Teorema 6.2.6 (*Leyes de los grandes números*) Sea $\{X_n\}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas, con media $\mu < \infty$. Entonces,

(i) (*Ley débil de los grandes números*)

$$\frac{1}{n} \sum_{i=1}^n X_n \xrightarrow{\mathbb{P}} \mu.$$

(ii) (*Ley fuerte de los grandes números*)

$$\frac{1}{n} \sum_{i=1}^n X_n \xrightarrow{c.s.} \mu.$$

Demostraremos la parte (i) bajo el supuesto adicional que las variables aleatorias tienen segundo momento finito. La demostración de la parte (ii) requiere de algunos conceptos más avanzados, y por lo tanto, por el momento postpondremos su demostración.

Demostración: Sea $\epsilon > 0$, luego por la desigualdad de Chebyshev tenemos que

$$\mathbb{P} \left(\left| \frac{1}{n} \sum_{i=1}^n X_n - \mu \right| > \epsilon \right) \leq \frac{1}{\epsilon^2} \mathbb{E} \left| \frac{1}{n} \sum_{i=1}^n X_n - \mu \right|^2.$$

El resultado se concluye aplicando el Ejemplo 6.2.5. □

Observemos que, la parte (i) del teorema anterior es un caso particular de (6.2).

Lema 6.2.7 Sean $\{X_n, X\}$ variables aleatorias tales que $X_n \xrightarrow{L^p} X$, para algún $p \geq 1$. Entonces, $X_n \xrightarrow{\mathbb{P}} X$.

Demostración: La demostración usa la misma idea de la prueba del Teorema 6.2.6 (i). Sea $\epsilon > 0$, entonces

$$\mathbb{P}(|X_n - X| > \epsilon) \leq \frac{\mathbb{E}|X_n - X|^p}{\epsilon^p} \rightarrow 0, \text{ cuando } n \rightarrow \infty.$$

□

El concepto de convergencia en probabilidad es importante en la teoría estadística ya que, si $\hat{\theta}_n := \hat{\theta}(X_1, X_2, \dots, X_n)$ es un estimador para un parámetro θ , se dice que $\hat{\theta}_n$ es un estimador consistente si $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta$.

6.3. Convergencia en distribución

Vamos a iniciar la presente sección analizando algunos ejemplos de convergencia en distribución.

Ejemplo 6.3.1 Para cada n sea X_n una variable aleatoria con distribución exponencial de parámetro $\lambda_n > 0$. Entonces,

$$\mathbb{P}(X_n \leq x) = 1 - e^{-\lambda_n x} \xrightarrow[n]{} \begin{cases} 1 - e^{-\lambda x}, & \text{si } \lambda_n \rightarrow \lambda \in (0, \infty), \\ 0, & \text{si } \lambda_n \rightarrow 0, \\ 1, & \text{si } \lambda_n \rightarrow \infty. \end{cases}$$

Por lo tanto,

1. Si $\lambda_n \rightarrow \lambda \in (0, \infty)$, entonces $X_n \xrightarrow[n]{D} X$, donde X es una variable aleatoria exponencial de parámetro λ .
2. Si $\lambda_n \rightarrow 0$, entonces $X_n \xrightarrow[n]{D} X$, donde X es una variable aleatoria que toma el valor $+\infty$ con probabilidad 1.
3. Si $\lambda_n \rightarrow \infty$, entonces $X_n \xrightarrow[n]{D} X$, donde X es una variable aleatoria tal que $\mathbb{P}(X = 0) = 1$. Veamos que la afirmación anterior es cierta. Sea F_X la función de distribución de X , entonces

$$F_X(x) = \begin{cases} 0, & x < 0, \\ 1, & x \geq 0. \end{cases}$$

Luego,

$$F_n(x) \xrightarrow[n]{} F_X(x),$$

para todo punto de continuidad x de F_X , lo cual prueba la afirmación.

◁

Ejemplo 6.3.2 (Poisson-convergencia en distribución) Anteriormente probamos que, si X_n es una variable aleatoria binomial de parámetros n y $p_n \in (0, 1)$ de tal forma que $np_n \rightarrow \lambda$, entonces, para cada $k = 0, 1, 2, \dots$, se tiene que

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_n = k) = \frac{\lambda^k e^{-\lambda}}{k!}.$$

Por lo tanto,

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = \sum_{k=0}^{[x]} \frac{\lambda^k e^{-\lambda}}{k!},$$

donde $[x]$ denota la parte entera de x . Así que, $X_n \xrightarrow[n]{D} X$, donde X es una variable aleatoria con distribución de Poisson de parámetro λ .

◁

Ejemplo 6.3.3 (Convergencia a la exponencial) Sea $\{U_i\}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas, con distribución uniforme en $(0, 1)$. Definamos $M_n = \max\{U_1, \dots, U_n\}$ y $Z_n = n(1 - M_n)$. Demuestre que $Z_n \xrightarrow{d} \mathcal{E}$, donde \mathcal{E} es una variable aleatoria con distribución exponencial de parámetro 1.

Solución:

Notemos que,

$$\begin{aligned} \mathbb{P}(Z_n \leq z) &= \mathbb{P}(M_n \geq 1 - z/n) \\ &= 1 - \mathbb{P}(M_n < 1 - z/n) \\ &= 1 - \prod_{i=1}^n \mathbb{P}(U_i < 1 - z/n) \\ &= 1 - (1 - z/n)^n. \end{aligned}$$

Por lo tanto,

$$\lim_{n \rightarrow \infty} \mathbb{P}(Z_n \leq z) = 1 - e^{-z}.$$

◁

Ejemplo 6.3.4 Sea $\{X_n\}$ una sucesión de variables aleatorias independientes todas con distribución exponencial de parámetro 1. Defina $M_n = \max\{X_1, \dots, X_n\}$. Demuestre que

$$M_n - \ln n \xrightarrow{d} X,$$

donde X es una variable aleatoria con función de distribución

$$F(x) = e^{-e^{-x}}, \quad x \in \mathbb{R}.$$

Solución:

$$\begin{aligned} \mathbb{P}(M_n - \ln n \leq x) &= \mathbb{P}(X_1 \leq x + \ln n, \dots, X_n \leq x + \ln n) \\ &= (1 - e^{-x - \ln n})^n \\ &= (1 - e^{-x}/n)^n. \end{aligned}$$

Por lo tanto,

$$\lim_{n \rightarrow \infty} \mathbb{P}(M_n - \ln n \leq x) = e^{-e^{-x}}.$$

◁

Es importante hacer notar que, si $\{X_n\}$ converge en distribución y cada X_n tiene función de densidad, no necesariamente las funciones de densidad convergen. Vamos a ver un ejemplo donde se ilustra tal situación.

²La función de distribución F es llamada distribución Gumbel y juega un papel muy importante en la modelación estadística.

Ejemplo 6.3.5 Sea X_n una variable aleatoria con función de distribución

$$F_n(x) = x - \frac{\sin(2n\pi x)}{2n\pi}, \quad 0 \leq x \leq 1.$$

Entonces, X_n tiene función de densidad, la cual está dada por

$$f_n(x) = F'_n(x) = 1 - \cos(2n\pi x).$$

Luego, la sucesión $\{f_n\}$ no converge. Sin embargo, $X_n \xrightarrow[n]{d} U$, donde U tiene distribución uniforme en $(0,1)$. ◁

Ahora vamos a demostrar la parte (6.3) del Teorema 6.1.2.

Proposición 6.3.6 Sea $\{X_n\}$ un sucesión de variables aleatorias tal que $X_n \xrightarrow[n]{p} X$, entonces $X_n \xrightarrow[n]{D} X$.

Demostración: Dado $\epsilon > 0$, para cada n y $x \in \mathbb{R}$, se tiene que

$$\begin{aligned} F_X(x - \epsilon) &= \mathbb{P}(X \leq x - \epsilon) \\ &= \mathbb{P}(X \leq x - \epsilon, |X_n - X| > \epsilon) + \mathbb{P}(X \leq x - \epsilon, |X_n - X| \leq \epsilon) \\ &= \mathbb{P}(X \leq x - \epsilon, |X_n - X| > \epsilon) + \mathbb{P}(X \leq x - \epsilon, X - \epsilon \leq X_n \leq X + \epsilon) \\ &\leq \mathbb{P}(|X_n - X| > \epsilon) + \mathbb{P}(X_n \leq x) \\ &\leq \mathbb{P}(|X_n - X| > \epsilon) + F_{X_n}(x). \end{aligned}$$

Procediendo de manera análoga podemos obtener que

$$F_{X_n}(x) \leq \mathbb{P}(|X_n - X| > \epsilon) + F_X(x + \epsilon).$$

Entonces, tenemos que

$$F_X(x - \epsilon) - \mathbb{P}(|X_n - X| > \epsilon) \leq F_{X_n}(x) \leq F_X(x + \epsilon) + \mathbb{P}(|X_n - X| > \epsilon),$$

haciendo $n \rightarrow \infty$ y usando que $X_n \xrightarrow[n]{p} X$ obtenemos que

$$F_X(x - \epsilon) \leq \liminf_{n \rightarrow \infty} F_{X_n}(x) \leq \limsup_{n \rightarrow \infty} F_{X_n}(x) \leq F_X(x + \epsilon).$$

Por último, si x es un punto de continuidad de F_X , entonces, haciendo $\epsilon \rightarrow 0$ se obtiene

$$F_X(x) \leq \liminf_{n \rightarrow \infty} F_{X_n}(x) \leq \limsup_{n \rightarrow \infty} F_{X_n}(x) \leq F_X(x),$$

es decir, $X_n \xrightarrow[n]{D} X$. ◻

El recíproco de la proposición anterior en general es falso como lo muestra el siguiente

Ejemplo 6.3.7 Sea X un variable aleatoria normal estándar y, para cada $n \in \mathbb{N}$, definamos

$$X_n = \begin{cases} X & \text{si } n \text{ es impar} \\ -X & \text{si } n \text{ es par.} \end{cases}$$

Entonces, $F_{X_n}(x) = F_X(x)$, para cualquier $x \in \mathbb{R}$. Por lo tanto, $X_n \xrightarrow{D} X$. Sin embargo, notemos que $|X_{2n} - X| = 2|X|$, para todo $n \in \mathbb{N}$. Luego,

$$\mathbb{P}(|X_{2n} - X| > \epsilon) = \mathbb{P}(|X| > \epsilon/2),$$

para cualquier $\epsilon > 0$. En consecuencia,

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_{2n} - X| > \epsilon) = \mathbb{P}(|X| > \epsilon/2) > 0,$$

así que la sucesión $\{X_n\}$ no converge en probabilidad. ◁

Bajo un supuesto adicional el recíproco de la Proposición 6.3 es válido.

Proposición 6.3.8 Sea $\{X_n\}$ una sucesión de variables aleatorias tales que $X_n \xrightarrow{D} c$, donde c es una constante. Entonces,

$$X_n \xrightarrow{P} c.$$

Demostración: Por hipótesis se tiene que

$$\lim_{n \rightarrow \infty} F_n(x) = \begin{cases} 0, & \text{si } x < c, \\ 1, & \text{si } x > c. \end{cases} \quad (6.4)$$

Para cada $\epsilon > 0$ obtenemos

$$\begin{aligned} \mathbb{P}(|X_n - c| > \epsilon) &= \mathbb{P}(X_n - c > \epsilon) + \mathbb{P}(X_n - c < -\epsilon) \\ &= \mathbb{P}(X_n - c > \epsilon) + \mathbb{P}(X_n - c \leq -\epsilon) \\ &= 1 - F_n(c + \epsilon) + F_n(c - \epsilon). \end{aligned}$$

Luego, de (6.4) se concluye que

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - c| > \epsilon) = 0.$$

◻

Es importante hacer notar que, la noción de convergencia en distribución, es de suma importancia es la teoría de probabilidad debido al amplia gama de usos tiene no sólo del punto de vista teórico sino también en las aplicaciones.

En la Proposición 6.2.2 vimos que si dos sucesiones convergen en probabilidad entonces la suma converge en probabilidad a la suma de los límites. En el caso de convergencia no se tiene tal cosa. En efecto, consideremos $X_1, X_2, \dots \sim N(0, 1)$ y $Y_1, Y_2, \dots \sim N(0, 1)$, todas independientes. Sea $X \sim N(0, 1)$ y definamos $Y = -X$, de modo que $Y \sim N(0, 1)$. Es claro que,

$$X_n \xrightarrow[n]{d} X \text{ y } Y_n \xrightarrow[n]{d} Y.$$

Sin embargo, $X_n + Y_n$ no converge en distribución a $X + Y$ ya que $X_n + Y_n \sim N(0, 2)$ y $X + Y$ es la variable aleatoria que toma el valor cero con probabilidad 1.

El siguiente teorema, el cual enunciamos sin demostración, es muy útil en aplicaciones:

Teorema 6.3.9 (*Slutsky*) Sean $\{X_n\}$ y $\{Y_n\}$ sucesiones de variables aleatorias tales que

$$X_n \xrightarrow[n]{d} X \text{ y } Y_n \xrightarrow[n]{\mathbb{P}} c,$$

donde c es una constante. Entonces,

$$(i) \quad X_n + Y_n \xrightarrow[n]{d} X + c.$$

$$(ii) \quad X_n Y_n \xrightarrow[n]{d} cX.$$

$$(iii) \quad \text{Si } c \neq 0, \quad X_n/Y_n \xrightarrow[n]{d} X/c.$$

Una forma equivalente de definir la convergencia en distribución es a través de la convergencia de esperanzas de funciones continuas de las variables aleatorias en cuestión. Lo anterior es el contenido del siguiente teorema, el cual enunciamos sin demostración, ya que para ello se requieren de herramientas más avanzadas de análisis.

Teorema 6.3.10 Una sucesión de variables aleatorias $\{X_n\}$ converge en distribución a una variable aleatoria X sí, y solamente si, para toda función $f : \mathbb{R} \rightarrow \mathbb{R}$ continua y acotada se tiene que

$$\lim_{n \rightarrow \infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X)).$$

El teorema anterior no dice que, para saber cuando una sucesión de variables aleatorias converge en distribución es “suficiente” conocer como se comportan las esperanzas de funciones continuas y acotadas. Este problema puede ser sumamente complicado, ya que la clase de funciones de continuas y acotadas es muy amplia y no siempre es posible conocer dichas esperanzas. Sin embargo, hay una clase de funciones más pequeña las cual determina la convergencia en distribución.

Recordemos que la función generadora de momentos de una variable aleatoria X está definida por

$$M_X(t) = \mathbb{E}(e^{tX}), \tag{6.5}$$

para todo t tal que el lado derecho de (6.5) sea finito.

Ejemplo 6.3.11 1. $X \sim \text{Poisson}(\lambda)$. Entonces, $M_X(t) = e^{-\lambda(e^t-1)}$ para todo $t \in \mathbb{R}$.

2. $X \sim \text{Bin}(n, p)$. Entonces, $M_X(t) = (1 - p + pe^t)^n$.

3. $X \sim N(\mu, \sigma^2)$. Entonces, $M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$. Vamos a encontrar la f.g.m. para $Y \sim N(0, 1)$.

$$\begin{aligned}\mathbb{E}(e^{tY}) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-\frac{x^2}{2}} dx \\ &= e^{\frac{t^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-t)^2}{2}} dx \\ &= e^{\frac{t^2}{2}}.\end{aligned}$$

Finalmente, notemos que $X := \sigma Y + \mu \sim N(\mu, \sigma^2)$, de donde se sigue el resultado.

4. $X \sim \text{Gama}(\alpha, \lambda)$. Entonces, $M_X(t) = \left(\frac{\lambda}{\lambda-t}\right)^\alpha$, $t < \lambda$.

Recordemos que,

$$\Gamma(\alpha) := \int_0^\infty x^{\alpha-1} \lambda^\alpha e^{-\lambda x} dx \quad (\text{No depende de } \lambda!).$$

Entonces,

$$\begin{aligned}\mathbb{E}(e^{tX}) &= \int_0^\infty e^{tx} \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} dx \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(\lambda-t)x} dx, \text{ finita para } t < \lambda \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty \left(\frac{y}{\lambda-t}\right)^{\alpha-1} e^{-y} \frac{dy}{\lambda-t}, \text{ haciendo } y = (\lambda-t)x \\ &= \left(\frac{\lambda}{\lambda-t}\right)^\alpha \frac{1}{\Gamma(\alpha)} \int_0^\infty y^{\alpha-1} e^{-y} dy \\ &= \left(\frac{\lambda}{\lambda-t}\right)^\alpha.\end{aligned}$$

◁

La función generadora de momentos contiene mucha información sobre la variable aleatoria como lo muestra el siguiente

Teorema 6.3.12 *Suponga que X es una variable aleatoria con función generadora de momentos $M(t)$ tal que $M(t) < \infty$, para $|t| < a$, $a > 0$. Entonces, la distribución de X esta determinada de manera única. Además,*

$$M(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} \mathbb{E}(X^k).$$

Notemos que, del teorema anterior se desprende la siguiente relación:

$$\mathbb{E}(X^k) = \frac{d}{dt} M(t)|_{t=0}.$$

Teorema 6.3.13 (Teorema de Continuidad) Sea $\{X_n\}$ una sucesión de variables aleatorias con funciones de distribución $\{F_n\}$, y funciones generadoras $\{M_n\}$, tales que para todo n

$$M_n(t) < \infty, \quad |t| < t_1,$$

para algún $t_1 > 0$. Supongamos que existe una función real M tal que

$$\lim_{n \rightarrow \infty} M_n(t) = M(t), \quad |t| < t_2, \quad t_2 > 0,$$

y $M(t) = \mathbb{E}(e^{tX})$, para alguna variable aleatoria X . Entonces,

$$X_n \xrightarrow[n]{D} X.$$

Es decir, $F_n(x) \xrightarrow[n]{} F_X(x)$, para todo punto de continuidad x de F .

La idea detrás del Teorema de Continuidad es que, cualquier función continua puede aproximarse por combinaciones lineales de funciones exponenciales, dicho resultado puede verse en cursos avanzados de análisis.

Algunos comentarios sobre el teorema del límite central

El teorema del límite central tiene sus orígenes en el teorema de D' Moivre (1733):

Teorema 6.3.14 (de D' Moivre-Laplace) Sea $\{X_n\}$ una sucesión de variables aleatorias tales que $X_n \sim \text{Bin}(n, p)$, $p \in (0, 1)$, entonces

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{X_n - np}{\sqrt{np(1-p)}} \leq x \right) = \Phi(x),$$

donde Φ denota la función de distribución de una variable aleatoria normal estándar, es decir,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy.$$

Demostración:

$$\begin{aligned} \mathbb{E} \left(\exp \left\{ t \frac{X_n - np}{\sqrt{npq}} \right\} \right) &= \left[\mathbb{E} \left(e^{t \frac{X_1 - p}{\sqrt{npq}}} \right) \right]^n \\ &= \left[\mathbb{E} \left(e^{t \frac{X_1}{\sqrt{npq}}} \right) e^{-\frac{tp}{\sqrt{npq}}} \right]^n \\ &= \left[\left(q + pe^{\frac{t}{\sqrt{npq}}} \right) e^{-\frac{tp}{\sqrt{npq}}} \right]^n \\ &= \left(pe^{\frac{tq}{\sqrt{npq}}} + qe^{-\frac{tp}{\sqrt{npq}}} \right)^n \\ &= \left(1 + \frac{t^2}{2n} + o(n^{-\frac{3}{2}}) \right)^n, \end{aligned}$$

Para obtener la última igualdad expandimos ambas exponenciales. Por lo tanto,

$$\lim_{n \rightarrow \infty} \mathbb{E} \left(\exp \left\{ t \frac{X_n - np}{\sqrt{npq}} \right\} \right) = e^{\frac{t^2}{2}},$$

donde el lado derecho es la función generadora de momentos de una variable aleatoria con distribución normal estándar.

Finalmente, por el Teorema de Continuidad se tiene que

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{X_n - np}{\sqrt{np(1-p)}} \leq x \right) = \Phi(x).$$

□

El resultado anterior fue el punto de partida para muchos estudios posteriores, de los cuáles, la escuela rusa tuvo una influencia importante. En particular, Lyapunov (1900) demostró la siguiente generalización: sea $\{X_n\}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas, con tercer momento finito, entonces

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(a < \frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \leq b \right) = \Phi(b) - \Phi(a),$$

donde $\mu = \mathbb{E}(X_1)$ y $\sigma^2 = \text{Var}(X_1)$.

Teorema 6.3.15 (Teorema del límite central) *Sea $\{X_n\}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas con media μ y varianza finita σ^2 . Entonces,*

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \leq x \right) = \Phi(x).$$

Demostración: Vamos a suponer que la función generadora de momentos de X_i existe en una vecindad de 0. Sea $Z_n := \frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}}$ y ψ la función generadora de momentos correspondiente. Entonces, por el supuesto de independencia tenemos que

$$M_{Z_n}(t) = \left[\psi \left(\frac{t}{\sigma\sqrt{n}} \right) \right]^n \exp \left\{ -\frac{n\mu t}{\sigma\sqrt{n}} \right\}.$$

Luego,

$$\ln M_{Z_n}(t) = n \ln \psi \left(\frac{t}{\sigma\sqrt{n}} \right) - \frac{n\mu t}{\sigma\sqrt{n}} = n \left[\ln \psi \left(\frac{t}{\sigma\sqrt{n}} \right) - \frac{\mu t}{\sigma\sqrt{n}} \right].$$

Recordemos que $\psi(0) = 1$, entonces aplicando la regla de L'Hôpital obtenemos que

$$\begin{aligned} \lim_{n \rightarrow \infty} \ln M_{Z_n}(t) &= \frac{t}{2\sigma} \lim_{n \rightarrow \infty} \sqrt{n} \left[\frac{\psi'(t/\sigma\sqrt{n})}{\psi(t/\sigma\sqrt{n})} - \mu \right] \\ &= 2 \left(\frac{t}{2\sigma} \right)^2 \lim_{n \rightarrow \infty} \left\{ \frac{\psi''(t/\sigma\sqrt{n})}{\psi(t/\sigma\sqrt{n})} - \frac{[\psi'(t/\sigma\sqrt{n})]^2}{\psi^2(t/\sigma\sqrt{n})} \right\} \\ &= \frac{t^2}{2\sigma^2} [\psi''(0) - \psi^2(0)] \\ &= \frac{t^2}{2}. \end{aligned}$$

La prueba se concluye usando el Teorema 6.3.13. □

Más tarde, Lindeberg (1922) demostró un teorema del límite central mucho más fuerte: sea sea $\{X_n\}$ una sucesión de variables aleatorias independientes con varianza finita, y tales que, para cualquier $\epsilon > 0$

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{k=1}^n \mathbb{E} [(X_k - \mu_k)^2 \mathbf{1}_{\{|X_k - \mu_k| > \epsilon s_n^2\}}] = 0, \quad (6.6)$$

donde $s_n^2 = \sigma_1^2 + \dots + \sigma_n^2$ y, μ_k y σ_k^2 son la media y la varianza de X_k , respectivamente. Entonces,

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(a < \frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \leq b \right) = \Phi(b) - \Phi(a).$$

Observación 6.3.16 *La condición de Lindeberg (6.6) se cumple cuando las variables aleatorias son idénticamente distribuidas.*

Ejemplo 6.3.17 *Use el TLC para demostrar que*

$$\lim_{n \rightarrow \infty} e^{-n} \sum_{i=0}^n \frac{n^i}{i!} = 1/2.$$

Solución: sea $\{X_n\}$ una sucesión de variables aleatorias independientes con distribución Poisson de parámetro 1. Entonces, $S_n = X_1 + \dots + X_n$ tiene distribución de Poisson de parámetro n . Luego,

$$\mathbb{P}(S_n \leq n) = e^{-n} \sum_{i=0}^n \frac{n^i}{i!}.$$

Por otro lado,

$$\begin{aligned} \mathbb{P}(S_n \leq n) &= \mathbb{P} \left(\frac{S_n - n}{\sqrt{n}} \leq 0 \right) \\ &\xrightarrow{n} \Phi(0) \\ &= \frac{1}{2}. \end{aligned}$$

◁

Ejemplo 6.3.18 *50 números seleccionados al azar se redondean al entero más cercano y después se suman. Suponga que los errores de redondeo están uniformemente distribuidos en $(-\cdot 5, \cdot 5)$. Encuentre la probabilidad de que la suma difiera del valor exacto en más de 3 unidades.*

Solución: sean a_1, \dots, a_{50} los siguientes números que se redondean y U_1, \dots, U_{50} los respectivos errores de redondeo. Entonces, $\mathbb{E}(U_i) = 0$ y $\text{Var}(U_i) = \frac{1}{12}$.

Nos interesa encontrar,

$$\begin{aligned}\mathbb{P}\left(\left|\sum_{i=1}^{50}(a_i + U_i) - \sum_{i=1}^{50} a_i\right| > 3\right) &= \mathbb{P}\left(\left|\sum_{i=1}^{50} U_i\right| > 3\right) \\ &= \mathbb{P}\left(\left|\frac{12}{50} \sum_{i=1}^{50} U_i\right| > 3\frac{12}{50}\right) \\ &\approx 2\Phi\left(-\frac{18}{25}\right) \text{ (por el Teorema 6.3.15)}\end{aligned}$$

◁

6.4. Ejercicios

1. Sean $\{X_n\}$ y $\{Y_n\}$ dos sucesiones de variables aleatorias tales que $X_n \xrightarrow{p} X$ y $\lim_{n \rightarrow \infty} \mathbb{E}(X_n - Y_n)^2 = 0$. Demuestre que

$$Y_n \xrightarrow{p} X.$$

2. Sea $\{Y_i\}$ una sucesión de variables aleatorias independientes, tales que Y_i tiene distribución Bernoulli de parámetro p_i . Demuestre que para cualquier $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\left| \frac{Y_1 + Y_2 + \cdots + Y_n}{n} - \frac{p_1 + p_2 + \cdots + p_n}{n} \right| > \epsilon \right) = 0.$$

3. Sea $\{X_n\}$ una sucesión de variables aleatorias independientes e idénticamente distribuidas, con $\mathbb{E}(X_n) = \mu$ y $\text{Var}X_n = \sigma^2$. Sea $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$. Demuestre que

$$\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \xrightarrow{p} \sigma^2.$$

4. Sea $\{N(t), t \geq 0\}$ un proceso de Poisson con intensidad $\lambda > 0$. Demuestre que

$$\frac{N(t)}{t} \xrightarrow[t \rightarrow \infty]{\mathbb{P}} \lambda.$$

5. Sea X una variable aleatoria con distribución uniforme en el intervalo $(0, 1)$ y, para cada $n \in \mathbb{N}$, sea X_n una variable aleatoria con distribución uniforme en el conjunto $\{\frac{1}{n}, \frac{2}{n}, \dots, \frac{n}{n}\}$.

a) Demuestre directamente que $X_n \xrightarrow[n]{D} X$.

b) Demuestre que la sucesión de funciones generadoras $\{M_{X_n}\}$ converge a M_X .

6. Sea $\{X_n\}$ una sucesión de variables aleatorias independientes, todas con distribución exponencial de parámetro λ y sea Φ la función de distribución de una variable aleatoria con distribución normal estándar. Utilice el teorema del límite central para expresar la función de distribución de $Z = X_1^2 + \cdots + X_n^2$ en términos de Φ .

7. Demuestre que

$$\lim_{n \rightarrow \infty} e^{-n} \sum_{k=0}^{2n} \frac{n^k}{k!} = 1.$$

8. Sean X_n y Y_n variables aleatorias independientes con parámetro n . Demuestre que $(X_n - Y_n)/\sqrt{2n} \xrightarrow[n]{d} X$, donde X es una variable aleatoria normal estándar.

9. Sean Y_1, Y_2, \dots variables aleatorias independientes todas con distribución uniforme en el conjunto $\{0, 1, \dots, 9\}$. Defina $X_n = \sum_{k=1}^n Y_k 10^{-k}$. Demuestre que X_n converge en distribución a la distribución uniforme en $[0, 1]$.

10. Se tienen 100 componentes, los cuales se utilizan en serie, es decir: primero se utiliza el componente 1, al fallar éste se utiliza el componente 2, al fallar este se utiliza el componente 3; y así sucesivamente. Estime la probabilidad de que el tiempo total de vida de los 100 componentes exceda 1200 suponiendo que, para cada $i \in \{1, \dots, 100\}$ el tiempo de vida del componente i tiene distribución:
- a) uniforme en el intervalo $(0, 20 + i/5)$;
 - b) exponencial de parámetro $\lambda_i = 1/(10 + i/10)$.
11. Sea $\{N_t, t \geq 0\}$ un proceso de Poisson de intensidad $\lambda > 0$. Demuestre que

$$\frac{N_t - \lambda t}{\sqrt{\lambda t}} \xrightarrow[n]{d} X,$$

donde X es una variable aleatoria con distribución normal estándar.

12. Use el Teorema del Límite central para evaluar

$$\lim_{n \rightarrow \infty} \frac{1}{(n-1)!} \int_0^n x^{n-1} e^{-x} dx.$$

Sugerencia: considere una suma de exponenciales de parámetro 1.