

Capítulo 1

Introducción a la Teoría de Probabilidades

1.1. Introducción

El objetivo de la Teoría de Probabilidades es desarrollar modelos para experimentos que están gobernados por el azar y estudiar sus propiedades y aplicaciones. El modelo fundamental para un experimento de este tipo, como el lanzamiento de un dado, es el Espacio de Probabilidad, que describimos a continuación.

En primer lugar tenemos un conjunto Ω , conocido como el *espacio muestral*, que contiene todos los resultados posibles del experimento. Por ejemplo, si el experimento consiste en lanzar un dado, el espacio muestral es $\Omega = \{1, 2, 3, 4, 5, 6\}$. Si seleccionamos un punto al azar en el intervalo $[0, 1]$, el espacio muestral es $\Omega = [0, 1]$. Si consideramos una sucesión infinita de experimentos con dos resultados posibles: 0 ó 1, el espacio muestral es el conjunto de todas las sucesiones de ceros y unos: $\Omega = \{(a_n)_{n \geq 1}, a_n = 0 \text{ ó } 1\}$.

Los elementos del espacio muestral se denotan por ω y se conocen como los sucesos o eventos elementales.

La segunda componente de nuestro modelo es la clase de los eventos o sucesos \mathcal{F} . Esta clase está compuesta por subconjuntos del espacio muestral y debe satisfacer las siguientes propiedades

A1. $\Omega \in \mathcal{F}$.

A2. Si $A \in \mathcal{F}$ entonces $A^c \in \mathcal{F}$.

A3. Si $A_n \in \mathcal{F}$ para $n \geq 1$ entonces $\bigcup_{n \geq 1} A_n \in \mathcal{F}$.

Una colección de subconjuntos de Ω que satisface estas tres condiciones se conoce como una σ -álgebra. Es sencillo demostrar que si A es cualquier conjunto, el conjunto de partes de A , $\mathcal{P}(A)$ es una σ -álgebra.

En el caso de experimentos sencillos, por ejemplo experimentos con un conjunto finito de resultados, normalmente tomamos como σ -álgebra de eventos el conjunto de partes de Ω . En experimentos más complicados, con una cantidad no-numerable de resultados posibles, no siempre es posible tomar esta opción, y es necesario considerar σ -álgebras más pequeñas.

La tercera y última componente del modelo es una *probabilidad* P definida sobre la clase de conjuntos \mathcal{F} que toma valores sobre el intervalo $[0, 1]$, y que satisface las siguientes propiedades:

P1. Para cualquier evento $A \in \mathcal{F}$,

$$0 = P(\emptyset) \leq P(A) \leq P(\Omega) = 1.$$

P2. Si A_n , $n \geq 1$ es una colección de conjuntos disjuntos 2 a 2, es decir, $A_i \cap A_j = \emptyset$ si $i \neq j$, entonces

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

La terna (Ω, \mathcal{F}, P) se llama un *espacio de probabilidad*. La función P es una (*medida de*) *probabilidad* y como está definida sobre \mathcal{F} , sólo podemos determinar la probabilidad de los conjuntos que están en esta clase; por eso decimos que estos son los conjuntos *medibles*. Las propiedades de \mathcal{F} garantizan que si hacemos las operaciones usuales (unión, intersección, complementos, diferencias, diferencias simétricas) con conjuntos medibles, obtenemos conjuntos medibles. Por ejemplo, si $A \subset B$ son conjuntos medibles entonces $B \setminus A$ también es medible. Como consecuencia de esto y la aditividad de las medidas de probabilidad tenemos que estas son monótonas: si $A \subset B$ son conjuntos medibles, $P(A) \leq P(B)$, ya que $B = A \cup (B \setminus A)$, los eventos en esta unión son disjuntos y por la aditividad

$$P(B) = P(A) + P(B \setminus A) \geq P(A).$$

Ejemplo 1.1

En el caso del lanzamiento de un dado, el espacio muestral es $\Omega = \{1, 2, 3, 4, 5, 6\}$ y los conjuntos medibles \mathcal{F} son todos los subconjuntos de Ω . Para definir la probabilidad P basta decir que todos los eventos elementales tienen la misma probabilidad (que por lo tanto es $1/6$):

$$P(\{i\}) = 1/6, \quad \text{para } i = 1, \dots, 6.$$

Con esta definición, si A es cualquier subconjunto de Ω entonces

$$P(A) = \frac{\text{Card}(A)}{6} = \frac{\text{Card}(A)}{\text{Card}(\Omega)},$$

donde $\text{card}(A)$ denota el cardinal del conjunto A . ▲

Ejemplo 1.2

Si tenemos un experimento que tiene una cantidad numerable de resultados posibles, $\Omega = \{\omega_1, \omega_2, \dots\}$, podemos tomar \mathcal{F} como la colección de todos los subconjuntos de Ω . Si p_1, p_2, \dots son números no-negativos que satisfacen $\sum_n p_n = 1$, podemos definir la probabilidad P por

$$P(A) = \sum_{\omega_i \in A} p_i, \quad \text{para } A \in \mathcal{F}.$$
▲

Ejemplo 1.3

Si el experimento consiste en escoger al azar un número en el intervalo $[0, 1]$, entonces la probabilidad de escoger un número en el intervalo $[c, d] \subset [0, 1]$ debe ser proporcional a la longitud del intervalo, pero como la probabilidad de que el número escogido caiga en el intervalo $[0, 1]$ es 1, vemos que no sólo es proporcional sino que es igual a la longitud del intervalo:

$$P([c, d]) = d - c, \quad \text{para todo } [c, d] \subset [0, 1]. \tag{1.1}$$

Lamentablemente, no es posible definir una medida de probabilidad sobre todos los subconjuntos de $[0, 1]$ que satisfaga la propiedad (1.1). La demostración de este hecho está fuera de los objetivos de este curso, pero esto implica que hay conjuntos que no son 'medibles', es decir, a los cuales no podemos asignarles una probabilidad.

Por lo tanto, es necesario restringirse a una clase más pequeña \mathcal{F} de subconjuntos de $[0, 1]$, que sea una σ -álgebra, es decir, que satisfaga las condiciones A1, A2 y A3. Una posibilidad es usar la clase de los conjuntos borelianos en $[0, 1]$, que es la menor σ -álgebra generada por los subintervalos de $[0, 1]$. Sin embargo, es importante observar que en este caso hay otras σ -álgebras que pueden considerarse. ▲

Dada cualquier colección \mathcal{C} de subconjuntos de Ω , es posible demostrar que existe una σ -álgebra, que denotaremos por $\sigma(\mathcal{C})$, que contiene a \mathcal{C} y que es la menor de todas las σ -álgebras que contienen a \mathcal{C} en el siguiente sentido: Si \mathcal{D} es otra σ -álgebra que contiene a \mathcal{C} , entonces se cumple que $\sigma(\mathcal{C}) \subset \mathcal{D}$. $\sigma(\mathcal{C})$ se conoce como la σ -álgebra *generada* por \mathcal{C} y es posible demostrar que siempre existe y es única.

En el ejemplo 1.3 mencionamos a la σ -álgebra de Borel \mathcal{B} en $[0, 1]$, que tiene gran importancia en el desarrollo de la teoría de la medida, y que introdujimos como la σ -álgebra generada por los subintervalos de $[0, 1]$. De manera equivalente se puede definir como la σ -álgebra generada por la colección de los intervalos abiertos (a, b) , $0 \leq a < b \leq 1$, o los intervalos cerrados $[a, b]$ o los de la forma $(a, b]$, o de la forma $[a, b)$. Es posible demostrar que todas estas definiciones son equivalentes.

También es posible definir la σ -álgebra de Borel como la σ -álgebra generada por los subconjuntos abiertos de $[0, 1]$, y se puede demostrar que esta definición es equivalente a cualquiera de las anteriores. Esta definición tiene la ventaja de que podemos usarla en cualquier espacio que tenga una topología, por ejemplo, en cualquier espacio métrico.

1.2. Probabilidad Condicional

Definición 1.1 La probabilidad condicional $P(A|B)$ del evento A dado el evento B se define por

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{si } P(B) > 0, \quad (1.2)$$

y no está definida, o se le asigna un valor arbitrario, cuando $P(B) = 0$.

A partir de esta definición tenemos la relación

$$P(A \cap B) = P(A|B)P(B). \quad (1.3)$$

Supongamos que saber que el evento B ocurrió no cambia la probabilidad de que A ocurra, es decir $P(A|B) = P(A)$. Entonces la relación (1.3) se convierte en

$$P(A \cap B) = P(A)P(B). \quad (1.4)$$

Definición 1.2 Si (1.4) se satisface decimos que los eventos A y B son *independientes*.

Ley de la Probabilidad Total y el Teorema de Bayes

Sea $\{B_i, i \geq 1\}$ una partición de Ω , es decir, una colección de subconjuntos de Ω que satisface

$$B_i \cap B_j = \emptyset \text{ siempre que } i \neq j \quad \text{y} \quad \Omega = \bigcup_1^{\infty} B_i.$$

Entonces, para cualquier evento A , teniendo en cuenta que los subconjuntos $(A \cap B_i), i \geq 1$ son disjuntos dos a dos,

$$P(A) = P(A \cap \Omega) = P(A \cap (\bigcup_1^{\infty} B_i)) = \sum_1^{\infty} P(A \cap B_i)$$

y ahora, usando la relación (1.3) en cada sumando obtenemos

$$P(A) = \sum_1^{\infty} P(A|B_i)P(B_i) \quad (1.5)$$

que se conoce como la ley de la probabilidad total.

Una consecuencia importante del resultado anterior es el Teorema de Bayes.

Teorema 1.1 (Bayes) Si los eventos $\{B_i, i \geq 1\}$ forman una partición de Ω , para cualquier evento A con $P(A) > 0$ y cualquier índice j ,

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{\infty} P(A|B_i)P(B_i)}. \quad (1.6)$$

Demostración. A partir de la definición de probabilidad condicional obtenemos

$$P(B_j|A) = \frac{P(A \cap B_j)}{P(A)}.$$

Ahora usando (1.3) en el numerador y (1.5) en el denominador obtenemos el resultado. ■

Ejemplo 1.4

Una prueba médica para detectar cierta enfermedad tiene una efectividad de 98% (si una persona padece la enfermedad la prueba es positiva con probabilidad 0.98) y da falsos positivos en 10% de los casos. Por investigaciones epidemiológicas se sabe que un 2% de la población padece la enfermedad. Si un paciente seleccionado al azar resulta positivo en la prueba, ¿cuál es la probabilidad de que tenga la enfermedad?

Usemos las notaciones E y S para un paciente enfermo o sano, respectivamente, y $+$ para una prueba positiva. E y S son eventos complementarios. Nuestra información inicial es:

$$P(+|E) = 0.98; \quad P(+|S) = 0.1; \quad P(E) = 0.02.$$

Queremos calcular $P(E|+)$ y usando el teorema de Bayes tenemos

$$\begin{aligned} P(E|+) &= \frac{P(+|E)P(E)}{P(+|E)P(E) + P(+|S)P(S)} \\ &= \frac{0.98 \times 0.02}{0.98 \times 0.02 + 0.1 \times 0.98} = 0.16 \end{aligned}$$

▲

Otro resultado importante sobre probabilidades condicionales que resulta útil para diversos cálculos es el siguiente. La demostración queda como ejercicio.

Proposición 1.1 Sea A_1, \dots, A_n eventos cualesquiera. Entonces

$$P(A_1 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_2 \cap A_1) \cdots P(A_n|A_1 \cap \dots \cap A_{n-1})$$

1.3. Variables Aleatorias

En este contexto, una variable aleatoria X es una función real definida sobre Ω que satisface ciertas condiciones de medibilidad que describimos a continuación. Está claro que si X toma valores reales, nos va a interesar calcular probabilidades del tipo $P(X \leq a)$, donde $a \in \mathbb{R}$. Por ejemplo, si X representa el ingreso de una familia, o el número de piezas defectuosas en un lote, o el nivel máximo de un río durante cierto año, las probabilidades anteriores son obviamente de interés.

Ahora bien, para que estas probabilidades existan, es necesario que los conjuntos cuyas probabilidades estamos calculando sean 'medibles', es decir, estén en la σ -álgebra \mathcal{F} . Estos conjuntos son de la forma $\{\omega : X(\omega) \leq a\}$, para $a \in \mathbb{R}$. Por lo tanto, la condición que tenemos que pedirle a X para garantizar que podemos asignar una probabilidad a todos estos conjuntos es la siguiente:

M1. Para todo número real a ,

$$\{\omega : X(\omega) \leq a\} \in \mathcal{F}.$$

Una función que satisface esta propiedad se dice que es *medible*. Por lo tanto, definimos una *variable aleatoria* como una función $X : \Omega \rightarrow \mathbb{R}$ que es medible. Usaremos las letras v.a. para abreviar variable aleatoria.

La medibilidad de una función $X : (\Omega, \mathcal{F}) \rightarrow \mathbb{R}$ depende de la σ -álgebra \mathcal{F} . Una función puede ser medible respecto a una σ -álgebra \mathcal{F}_1 y no respecto a otra \mathcal{F}_2 . Sin embargo, está claro a partir de la definición, que si X es medible respecto a \mathcal{F}_1 y $\mathcal{F}_1 \subset \mathcal{F}_2$, entonces X también es medible respecto a \mathcal{F}_2 . La menor σ -álgebra respecto a la cual una variable aleatoria X es medible se conoce como la σ -álgebra *generada por X* , y se denota por $\sigma(X)$. Esta σ -álgebra no es otra cosa que la intersección de todas las σ -álgebras respecto a las cuales X es medible.

Por ejemplo, si X sólo tiene una cantidad numerable de valores posibles x_1, x_2, \dots , los conjuntos

$$A_i = \{\omega : X(\omega) = x_i\}, \quad i = 1, 2, \dots$$

forman una partición numerable de Ω , es decir,

$$\Omega = \bigcup_{i=1}^{\infty} A_i, \quad \text{y} \quad A_i \cap A_j = \emptyset \quad \text{si } i \neq j.$$

En este caso \mathcal{F} está compuesta por los conjuntos \emptyset, Ω y por todos los conjuntos que sean unión de algunos de los A_i .

Ejemplo 1.5

Veamos un ejemplo sencillo. Consideremos el lanzamiento de un dado con el espacio de probabilidad descrito en el ejemplo 1.1 y consideremos la función $X : \Omega \rightarrow \{0, 1\}$ definida de la siguiente manera:

$$X(\omega) = \begin{cases} 1 & \text{si } \omega \text{ es par,} \\ 0 & \text{si } \omega \text{ es impar.} \end{cases}$$

La σ -álgebra generada por X está formada por los conjuntos \emptyset y Ω y por las preimágenes de los valores de la función, que son, respectivamente, los números pares y los impares en Ω :

$$X^{-1}(0) = \{1, 3, 5\}, \quad X^{-1}(1) = \{2, 4, 6\}.$$

Por lo tanto

$$\sigma(X) = \{\emptyset; \{1, 3, 5\}; \{2, 4, 6\}; \Omega\}.$$

▲

Ejemplo 1.6

Consideremos el mismo espacio Ω del ejemplo anterior y sean

$$\mathcal{A} = \mathcal{P}(\Omega), \quad \text{y} \quad \mathcal{F} = \{\emptyset; \{1, 3, 5\}; \{2, 4, 6\}; \Omega\}$$

dos σ -álgebras de subconjuntos de Ω . Sea $X : \Omega \rightarrow \{0, 1\}$ la función definida por $X = \mathbf{1}_{\{1,2,3\}}$ donde $\mathbf{1}_A$ es la función indicadora del conjunto A ($\mathbf{1}_A(\omega) = 1$ si $\omega \in A$ y $\mathbf{1}_A(\omega) = 0$ si $\omega \notin A$). Entonces X es medible respecto a \mathcal{A} pero no respecto a \mathcal{F} .

▲

No es difícil demostrar que si la condición M1 se satisface entonces para cualquier intervalo real I se tiene que

$$\{\omega : X(\omega) \in I\} \in \mathcal{F}, \tag{1.7}$$

y recíprocamente, si (1.7) vale entonces la condición M1 es cierta. Es un poco más difícil demostrar, pero también es cierto, que la condición M1 es equivalente a (1.7) si reemplazamos el intervalo I por un boreliano B . Resumiendo tenemos las siguientes equivalencias:

$$\begin{aligned} \{\omega : X(\omega) \leq a\} \in \mathcal{F}, \text{ para todo } a \in \mathbb{R} &\Leftrightarrow \{\omega : X(\omega) \in I\} \in \mathcal{F}, \text{ para todo intervalo } I \subset \mathbb{R} \\ &\Leftrightarrow \{\omega : X(\omega) \in B\} \in \mathcal{F}, \text{ para todo } B \in \mathcal{B}. \end{aligned}$$

1.4. Distribución de una Variable Aleatoria

Consideremos un espacio de probabilidad (Ω, \mathcal{F}, P) sobre el cual hemos definido una variable aleatoria X con valores reales. Si A es un intervalo de \mathbb{R} y queremos calcular la probabilidad de que la variable X tome valores en A , tenemos que considerar el conjunto $\{\omega : X(\omega) \in A\} = X^{-1}(A)$, que es la pre-imagen de A por la función X . Como la función X es medible, este conjunto está en la colección \mathcal{F} de los conjuntos medibles y en consecuencia podemos calcular su probabilidad. Por lo tanto, si A es un intervalo

$$P(X \in A) = P(\{\omega \in \Omega : X(\omega) \in A\}) = P(X^{-1}(A)). \quad (1.8)$$

Es posible demostrar que esta definición también funciona para conjuntos más complicados. Si llamamos \mathcal{B} a la σ -álgebra generada por los intervalos de \mathbb{R} (como mencionamos anteriormente, \mathcal{B} se conoce como la σ -álgebra de Borel y sus conjuntos son los borelianos de \mathbb{R}) entonces la relación (1.8) vale para todo $A \in \mathcal{B}$.

Esta relación nos permite definir una (medida de) probabilidad P_X sobre $(\mathbb{R}, \mathcal{B})$ inducida por la variable X , de la siguiente manera: Para todo $A \in \mathcal{B}$,

$$P_X(A) = P(X \in A) = P(\{\omega \in \Omega : X(\omega) \in A\}). \quad (1.9)$$

Esta (medida de) probabilidad se conoce como la *distribución* o la *ley* de X y en ocasiones se usa la notación $\mathcal{L}(X)$. Esta probabilidad contiene toda la información probabilística sobre la variable X .

1.5. Funciones de Distribución

Si en la relación (1.9) consideramos subconjuntos A de \mathbb{R} de la forma $(-\infty, x]$ obtenemos la siguiente función $F : \mathbb{R} \rightarrow \mathbb{R}$, que se conoce como la *función de distribución* de X :

$$F(x) = P_X((-\infty, x]) = P(X \leq x).$$

Si tenemos varias variables aleatorias en el mismo contexto, puede resultar útil distinguir sus funciones de distribución usando la notación F_X para la función de distribución de X . Usaremos las letras f.d. para abreviar función de distribución.

Está claro que si conocemos la distribución de una variable aleatoria, entonces podemos determinar la función de distribución correspondiente. El recíproco también es cierto, pero la demostración de este hecho requiere herramientas de teoría de la medida que no están a nuestra disposición en este curso.

Una función de distribución F tiene las siguientes tres propiedades,

FD1. F es continua por la derecha y tiene límites por la izquierda.

FD2. F es creciente (en sentido amplio).

FD3. Si F es una función de distribución,

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

Estas tres propiedades caracterizan a las funciones de distribución: Si una función F satisface estas tres propiedades entonces existe una variable aleatoria X definida sobre un espacio de probabilidad (Ω, \mathcal{F}, P) tal que F es la función de distribución de X . Más aún, es posible tomar $\Omega = \mathbb{R}$.

1.5.1. Variables Discretas

Una variable aleatoria X es *discreta* si toma valores en un conjunto finito $\{x_1, x_2, \dots, x_n\}$ o numerable $\{x_1, x_2, \dots\}$. En el primer caso existen números positivos p_1, \dots, p_n con $p_1 + \dots + p_n = 1$, tales que

$$P(X = x_i) = p_i. \quad (1.10)$$

Llamaremos a esta función $p_X(x_i) = p_i, 1 \leq i \leq n$, la *función de probabilidad o densidad* de X .

De manera similar, en el segundo caso tenemos números positivos $p_i, i \geq 1$ con $\sum_{i=1}^{\infty} p_i = 1$ que satisfacen (1.10) para $i \geq 1$.

En ambos casos las funciones de distribución son funciones de saltos, es decir, funciones que sólo crecen por saltos y que son constantes entre saltos consecutivos. Los saltos están dados por

$$p(x_i) = F(x_i) - F(x_i^-)$$

es decir, que los saltos ocurren en los valores de la función en los puntos x_i y su altura es igual a la probabilidad de este valor. Además

$$F(x) = \sum_{x_i \leq x} p(x_i)$$

Ejemplo 1.7

Una variable con distribución uniforme en el conjunto $\{1, 2, 3, 4\}$ tiene función de probabilidad

$$p(j) = P(X = j) = \frac{1}{4} \quad \text{para } j = 1, 2, 3, 4.$$

En este caso la función de distribución F es una función escalera

$$F(x) = \begin{cases} 0, & \text{para } -\infty \leq x < 1, \\ 1/4, & \text{para } 1 \leq x < 2, \\ 1/2, & \text{para } 2 \leq x < 3, \\ 3/4, & \text{para } 3 \leq x < 4, \\ 1, & \text{para } x \geq 4. \end{cases}$$

▲

1.5.2. Variables Continuas

Una variable X es *continua* si su función de distribución F es continua. Una definición equivalente es que una v.a. X es continua si para cualquier valor x de la variable se tiene que $P(X = x) = 0$. Para la mayoría de estas variables existe una función $f : \mathbb{R} \rightarrow \mathbb{R}$ con $f(x) \geq 0$ para todo x y $\int_{\mathbb{R}} f(x) dx = 1$ que satisface

$$F(x) = \int_{-\infty}^x f(t) dt, \quad (1.11)$$

para todo $x \in \mathbb{R}$. La función f se conoce como la *densidad* de la variable X o de su f.d. F_X .

Hay algunas variables continuas que no tienen esta propiedad, es decir, que su función de distribución no puede obtenerse como la integral de otra función, pero su interés es más bien teórico y no las consideraremos en este curso. Se conocen como variables (o funciones de distribución) continuas singulares. De ahora en adelante todas las f.d. continuas que consideraremos satisfacen (1.11).

Si F es diferenciable en x entonces la densidad de probabilidad está dada por

$$f(x) = \frac{d}{dx} F(x) = F'(x), \quad -\infty < x < \infty.$$

Ejemplo 1.8

Consideremos una variable X con función de distribución

$$F(x) = \begin{cases} 0 & \text{para } x < 0, \\ x^2 & \text{para } 0 \leq x \leq 1, \\ 1 & \text{para } x \geq 1. \end{cases}$$

Esta función tiene derivada (salvo en el punto 1) que está dada por

$$f(x) = \begin{cases} 2x & \text{en } 0 < x < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

que es la densidad en este caso.

Hay variables aleatorias 'mixtas', cuyas f.d. son la combinación de funciones continuas y saltos. ▲

Ejemplo 1.9

La función de distribución

$$F(x) = \begin{cases} 0 & \text{para } x \leq 0, \\ x & \text{para } 0 < x < 1/2, \\ 1 & \text{para } x \geq 1/2. \end{cases}$$

es de este tipo. ▲

Dada cualquier función de distribución F es posible demostrar que existen dos funciones de distribución F_d y F_c , la primera discreta y la segunda continua, y un número $0 \leq \alpha \leq 1$ tales que $F = \alpha F_d + (1 - \alpha) F_c$. Esta descomposición es única.

1.6. Valores Esperados y Momentos

Si X es una v.a. discreta, el *momento de orden n* de X está dado por

$$E[X^n] = \sum_i x_i^n P(X = x_i) \quad (1.12)$$

siempre y cuando la serie en (1.12) converja absolutamente. Si esta serie diverge decimos que el momento no existe.

Si X es una v.a. continua con densidad $f(x)$, el *momento de orden n* de X está dado por

$$E[X^n] = \int_{-\infty}^{\infty} x^n f(x) dx, \quad (1.13)$$

siempre que esta integral converja absolutamente.

El primer momento, que corresponde a $n = 1$, se conoce como la *media*, el *valor esperado* o *esperanza* de X , y lo denotaremos por μ_X . El *momento central* o *centrado* de orden n es el momento de orden n de la variable $X - \mu_X$, siempre que μ_X exista. El primer momento central es cero. El segundo momento central se conoce como la *varianza* de X , denotado por $\text{Var}(X)$. Su raíz cuadrada es la *desviación típica*. Tenemos

$$\text{Var}[X] = E[(X - \mu_X)^2] = E[X^2] - \mu_X^2.$$

La mediana de una v.a. X es cualquier valor ν con la propiedad de que

$$P(X \geq \nu) \geq \frac{1}{2}, \quad \text{y} \quad P(X \leq \nu) \geq \frac{1}{2}.$$

Si X es una variable aleatoria y g es una función (medible) entonces $g(X)$ también es una variable aleatoria. Si X es discreta y toma valores $x_j, j \geq 1$ entonces el valor esperado de $g(X)$ está dado por

$$E[g(X)] = \sum_{j=1}^{\infty} g(x_j) P(X = x_j) \quad (1.14)$$

siempre que la suma converja absolutamente. Si X es continua y tiene densidad f , el valor esperado de $g(X)$ es

$$\mathbb{E}[g(X)] = \int g(x)f_X(x) dx. \quad (1.15)$$

Una fórmula general que abarca ambos casos (y también los casos mixtos) es la siguiente

$$\mathbb{E}[g(X)] = \int g(x) dF_X(x), \quad (1.16)$$

donde F_X es la f.d. de la variable X . La integral que aparece en la fórmula (1.16) es una integral de Lebesgue-Stieltjes, cuya definición está más allá del nivel de este curso. Para nuestros efectos, interpretamos esta integral como la suma que aparece en la fórmula (1.14) si la variable es discreta y como la integral de la ecuación (1.15) si es continua.

A continuación presentamos dos desigualdades sencillas pero fundamentales.

Teorema 1.2 (Desigualdad de Markov) *Si X es una variable aleatoria que satisface $X \geq 0$, entonces, para cualquier $a > 0$,*

$$P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

Demostración. Haremos la demostración en el caso continuo. Si X tiene densidad f ,

$$\begin{aligned} \mathbb{E}[X] &= \int_0^\infty xf(x) dx = \int_0^a xf(x) dx + \int_a^\infty xf(x) dx \\ &\geq \int_a^\infty xf(x) dx \geq \int_a^\infty af(x) dx \\ &= a \int_a^\infty f(x) dx = aP(X \geq a). \end{aligned}$$

■

Corolario 1.1 (Desigualdad de Chebyshev) *Si X es una variable aleatoria con media μ y varianza σ^2 , para cualquier valor de $x > 0$,*

$$P(|X - \mu| \geq x) \leq \frac{\sigma^2}{x^2}.$$

Demostración. Como $(X - \mu)^2$ es una v.a. no negativa, podemos aplicar la desigualdad de Markov con $a = x^2$ y obtenemos

$$P((X - \mu)^2 \geq x^2) \leq \frac{\mathbb{E}[(X - \mu)^2]}{x^2}.$$

Pero como $(X - \mu)^2 \geq x^2$ sí y sólo sí $|X - \mu| > x$, la desigualdad anterior equivale a

$$P(|X - \mu| \geq x) \leq \frac{\mathbb{E}[(X - \mu)^2]}{x^2} = \frac{\sigma^2}{x^2}.$$

■

Para concluir esta sección enunciamos dos resultados fundamentales que son válidos para la integral de Lebesgue-Stieltjes. De nuevo, las demostraciones correspondientes están más allá del nivel de este curso.

Teorema 1.3 (Convergencia Monótona) *Si X_1, X_2, \dots es una sucesión de variables aleatorias acotadas inferiormente que satisface $X_1 \leq X_2 \leq X_3 \dots$ y $X_n \uparrow X$, entonces*

$$\mathbb{E}[X_n] \uparrow \mathbb{E}[X].$$

Teorema 1.4 (Convergencia Dominada) Sea X_i , $i \geq 1$ una sucesión de variables aleatorias que satisfacen $|X_i| \leq Y$, donde Y es una v.a. con $E[Y] < \infty$. Si

$$\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$$

para casi todo ω (es decir, para todo $\omega \in \Omega$ fuera de un subconjunto de probabilidad 0), entonces

$$\lim_{n \rightarrow \infty} E[X_n] = E[X].$$

1.7. Distribuciones Conjuntas e Independencia

Si tenemos un par de variables aleatorias (X, Y) definidas sobre un espacio de probabilidad (Ω, \mathcal{F}, P) , su función de distribución conjunta F_{XY} está definida por

$$F_{XY}(x, y) = F(x, y) = P(X \leq x, Y \leq y).$$

Si ambas variables son discretas y toman valores x_i , $i \geq 1$ e y_j , $j \geq 1$ respectivamente, la función de probabilidad conjunta de X e Y es

$$p_{XY}(x_i, y_j) = P(X = x_i, Y = y_j), \quad i \geq 1, j \geq 1.$$

Una función de distribución conjunta tiene densidad (conjunta) si existe una función f_{XY} de dos variables que satisfice

$$F_{XY}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{XY}(s, t) dt ds, \quad \text{para todo } x, y.$$

La función $F_X(x) = \lim_{y \rightarrow \infty} F(x, y)$ es una f.d. que se conoce como la *función de distribución marginal* de X . Si las variables aleatorias son ambas discretas, las funciones de probabilidad marginales están dadas por

$$p_X(x_i) = \sum_{j=1}^{\infty} p_{XY}(x_i, y_j) \quad y \quad p_Y(y_j) = \sum_{i=1}^{\infty} p_{XY}(x_i, y_j).$$

Si la f.d. F tiene una densidad conjunta f , las densidades marginales de X e Y están dadas, respectivamente, por

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy \quad y \quad f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx.$$

Si X e Y tienen distribución conjunta entonces

$$E[X + Y] = E[X] + E[Y]$$

siempre y cuando todos estos momentos existan.

Independencia

Si para todos los valores de x e y se tiene que $F(x, y) = F_X(x) \times F_Y(y)$ decimos que las variables X e Y son *independientes*. Si las variables son discretas y tienen función de probabilidad conjunta p_{XY} , son independientes si y sólo si

$$p_{XY}(x, y) = p_X(x)p_Y(y). \quad (1.17)$$

De manera similar, si las variables son continuas y tienen densidad conjunta $f_{XY}(x, y)$, son independientes si y sólo si

$$f_{XY}(x, y) = f_X(x)f_Y(y). \quad (1.18)$$

Si X e Y son variables con distribución conjunta, medias μ_X , μ_Y , y varianzas finitas σ_X^2 , σ_Y^2 , la covarianza de X e Y , que escribimos σ_{XY} o $\text{Cov}(X, Y)$, está definida por

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - \mu_X\mu_Y,$$

y decimos que X e Y no están correlacionadas si su covarianza es cero, es decir, $\sigma_{XY} = 0$.

Las variables independientes con varianza finita no están correlacionadas, pero el recíproco no es cierto. Hay variables que no están correlacionadas pero no son independientes.

Dividiendo la covarianza σ_{XY} por las desviaciones típicas σ_X y σ_Y obtenemos el coeficiente de correlación $\rho_{X,Y}$:

$$\rho_{X,Y} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}$$

que satisface $-1 \leq \rho \leq 1$.

Para una colección X_1, \dots, X_n de variables aleatorias la distribución conjunta se define como

$$F(x_1, \dots, x_n) = F_{X_1 \dots X_n}(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n).$$

Si

$$F_{X_1 \dots X_n}(x_1, \dots, x_n) = F_{X_1}(x_1) \cdots F_{X_n}(x_n)$$

para todos los valores posibles de x_1, \dots, x_n decimos que las variables X_1, \dots, X_n son independientes.

Una función de distribución conjunta $F(x_1, \dots, x_n)$ tiene densidad de probabilidad $f(t_1, \dots, t_n)$ si

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f(t_1, \dots, t_n) dt_1 \cdots dt_n,$$

para todos los valores de x_1, \dots, x_n .

Para variables X_1, \dots, X_n y funciones arbitrarias h_1, \dots, h_m de n variables,

$$E\left[\sum_{j=1}^m h_j(X_1, \dots, X_n)\right] = \sum_{j=1}^m E[h_j(X_1, \dots, X_n)],$$

siempre que todos estos momentos existan.

Proposición 1.2 Si X, Y son v.a.i. con primer momento finito, entonces el producto XY también tiene primer momento finito y

$$E(XY) = E(X)E(Y)$$

Este resultado se extiende a cualquier colección finita de variables independientes.

Demostración. Vamos a hacer la demostración en el caso discreto. Sean $\{x_i, i \geq 1\}$ y $\{y_j, j \geq 1\}$ los conjuntos de valores de X y Y , respectivamente, y sea $p_{XY}(x, y)$ la función de probabilidad conjunta con marginales $p_X(x)$ y $p_Y(y)$. por independencia tenemos que $p_{XY}(x, y) = p_X(x)p_Y(y)$.

El valor esperado del producto XY es

$$\begin{aligned} E(XY) &= \sum_{x,y} xy p_{XY}(x, y) \\ &= \sum_{x,y} xy p_X(x) p_Y(y) \end{aligned}$$

Podemos separar la suma sobre x de la suma sobre y y obtenemos

$$= \sum_x x p_X(x) \sum_y y p_Y(y) = E(X)E(Y)$$

■

Corolario 1.2 Si X e Y son independientes y tienen varianzas respectivas σ_X^2 y σ_Y^2 entonces la varianza de la suma $Z = X + Y$ es la suma de las varianzas:

$$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2.$$

Esta propiedad se extiende al caso de n variables independientes.

Demostración.

$$\begin{aligned} \text{Var}(X + Y) &= \mathbb{E} \left[((X + Y) - \mathbb{E}(X + Y))^2 \right] = \mathbb{E} \left[(X - \mathbb{E}(X) + Y - \mathbb{E}(Y))^2 \right] \\ &= \mathbb{E} \left[(X - \mathbb{E}(X))^2 \right] + \mathbb{E} \left[(Y - \mathbb{E}(Y))^2 \right] + 2 \mathbb{E} \left[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y)) \right] \\ &= \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y). \end{aligned}$$

Este resultado es general, pues no hemos usado hasta ahora la independencia de las variables. Si las variables no están correlacionadas, y en particular si son independientes, la covarianza vale 0 y se tiene el resultado del teorema. ■

Sumas y Convoluciones

Si X e Y son variables aleatorias independientes con f.d. F_X y F_Y , respectivamente, entonces la f.d. de la suma $Z = X + Y$ es la convolución de F_X y F_Y :

$$F_Z(z) = \int_{-\infty}^{\infty} F_X(z-t) dF_Y(t) = \int_{-\infty}^{\infty} F_Y(z-t) dF_X(t).$$

Si X e Y toman valores en los enteros no-negativos con funciones de probabilidad respectivas p_X y p_Y entonces

$$p_Z(n) = P(Z = n) = \sum_{i=0}^n P(X = i)P(Y = n - i) = \sum_{i=0}^n p_X(i)p_Y(n - i) = \sum_{i=0}^n p_X(n - i)p_Y(i).$$

Si consideramos la situación en la cual X tienen Y densidades f_X y f_Y , respectivamente, la densidad f_Z de la suma es la convolución de las densidades f_X y f_Y :

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-t)f_Y(t) dt = \int_{-\infty}^{\infty} f_Y(z-t)f_X(t) dt.$$

1.8. Algunas Distribuciones Importantes

1.8.1. Distribuciones Discretas

Distribución de Bernoulli

Una variable aleatoria de Bernoulli toma valores 1 y 0 con probabilidades respectivas p y $q = 1 - p$, donde $0 < p < 1$. Si el resultado del experimento es 1 decimos que ha ocurrido un *éxito* y p es entonces la probabilidad de éxito. La media y varianza son, respectivamente,

$$\mathbb{E}[X] = p, \quad \text{Var}[X] = p(1 - p).$$

Si A es un evento y $\mathbf{1}_A$ es la función indicadora de A , entonces $\mathbf{1}_A$ es una variable de Bernoulli con parámetro $p = \mathbb{E}[\mathbf{1}_A] = P(A)$.

Distribución Binomial

Consideremos una colección X_1, X_2, \dots, X_n de variables independientes de Bernoulli con probabilidad de éxito p . Sea S el total de éxitos en los n experimentos de Bernoulli, es decir, $S = \sum_1^n X_i$. La distribución de S es binomial con parámetros n y p , es decir, la función de probabilidad es

$$p_S(k) = P(S = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \text{para } k = 0, 1, \dots, n. \quad (1.19)$$

Usaremos la notación $X \sim \text{Bin}(n, p)$ para indicar que la variable X tiene distribución binomial de parámetros n y p .

Podemos determinar el valor esperado usando la definición

$$E[S] = E\left[\sum_1^n X_i\right] = \sum_1^n E[X_i] = np,$$

y usando independencia podemos calcular la varianza,

$$\text{Var}[S] = \text{Var}\left[\sum_1^n X_i\right] = \sum_1^n \text{Var}[X_i] = np(1-p).$$

Distribución Geométrica

En el mismo esquema anterior, sea Z el número de ensayos antes del primer éxito, es decir, $Z = k$ sí y sólo sí los primeros $k-1$ ensayos resultan en fracaso, cada uno con probabilidad $(1-p)$ y el k -ésimo es un éxito. En este caso Z tiene distribución geométrica con función de probabilidad

$$p_Z(k) = p(1-p)^{k-1}, \quad \text{para } k = 1, 2, \dots \quad (1.20)$$

y los primeros dos momentos son

$$E[Z] = \frac{1}{p}, \quad \text{Var}[Z] = \frac{1-p}{p^2}.$$

En ciertas ocasiones se define la distribución geométrica asociada al número de fracasos hasta el primer éxito, es decir, $Z' = Z - 1$. Con esta definición la función de probabilidad es

$$p_{Z'}(k) = p(1-p)^k, \quad \text{para } k = 1, 2, \dots \quad (1.21)$$

En este caso $E[Z'] = E[Z] - 1 = (1-p)/p$ y $\text{Var}[Z'] = \text{Var}[Z] = (1-p)/p^2$.

Distribución Binomial Negativa

Sea ahora W_k el número de ensayos antes del k -ésimo éxito, para un entero fijo $k \geq 1$. La distribución de esta variable se conoce como la binomial negativa de parámetros k y p . Para que W_k tome el valor r es necesario que haya exactamente $k-1$ éxitos en los primeros $r-1$ ensayos y luego un éxito en el ensayo r . La probabilidad de la primera condición es binomial mientras que la probabilidad de la segunda es p , lo cual nos da la función de probabilidad de W_k :

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

Otra manera de escribir W_k es como la suma de k variables independientes con distribución geométrica (1.20): $W_k = Z_1 + \dots + Z_k$. Usando esta relación es inmediato que

$$E[W_k] = \frac{k}{p}, \quad \text{Var}[W_k] = \frac{k(1-p)}{p^2}.$$

Distribución de Poisson

La distribución de Poisson de parámetro $\lambda > 0$ tiene función de probabilidad

$$p(k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad \text{para } k = 0, 1, \dots \quad (1.22)$$

Usaremos la notación $X \sim \mathcal{Pois}(\lambda)$ para esta distribución. El desarrollo en serie de potencias de la función exponencial es

$$e^\lambda = 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \dots$$

y vemos que $\sum_0^\infty p(k) = 1$. Usando de nuevo este desarrollo podemos calcular el valor esperado para una variable X con esta distribución

$$E[X] = \sum_{k=0}^{\infty} k p(k) = \sum_{k=1}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda.$$

La misma idea nos permite calcular

$$E[X(X-1)] = \sum_{k=0}^{\infty} k(k-1)p(k) = \sum_{k=2}^{\infty} k(k-1) \frac{\lambda^k e^{-\lambda}}{k!} = \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} = \lambda^2.$$

A partir de esta relación obtenemos $E[X(X-1)] = E[X^2] - E[X] = \lambda^2$, de donde $E[X^2] = \lambda^2 + \lambda$ y $\text{Var}[X] = E[X^2] - (E[X])^2 = \lambda$.

Entre otras razones, la distribución de Poisson es importante porque aparece como límite de la distribución binomial cuando $n \rightarrow \infty$ y $p \rightarrow 0$ de modo que $np \rightarrow \lambda > 0$. Este resultado se conoce como la ley de 'eventos raros'.

Distribución Multinomial

Las variables X_1, \dots, X_k , con valores en el conjunto $\{0, 1, \dots, n\}$, tienen una distribución multinomial si su función de probabilidad conjunta es

$$P(X_1 = r_1, \dots, X_k = r_k) = \begin{cases} \frac{n!}{r_1! \cdots r_k!} p_1^{r_1} \cdots p_k^{r_k}, & \text{si } r_1 + \cdots + r_k = n, \\ 0 & \text{si no.} \end{cases}$$

donde $p_i > 0$ para $i = 1, \dots, k$ y $p_1 + \cdots + p_k = 1$.

Para esta distribución $E[X_i] = np_i$, $\text{Var}[X_i] = np_i(1 - p_i)$ y $\text{Cov}(X_i X_j) = -np_i p_j$.

1.8.2. Distribuciones Continuas

Distribución Normal

Una variable X tiene distribución normal de parámetros μ y σ^2 si su densidad es

$$\varphi(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty.$$

Usaremos la notación $X \sim \mathcal{N}(\mu, \sigma^2)$. Los parámetros μ y σ^2 representan el valor esperado y la varianza de la variable X . La densidad $\varphi(x; \mu, \sigma^2)$ es simétrica respecto a μ .

El caso $\mu = 0$, $\sigma^2 = 1$ se conoce como la densidad normal típica o estándar. Si $X \sim \mathcal{N}(\mu, \sigma^2)$ entonces $Z = (X - \mu)/\sigma$ tiene una distribución normal típica. De esta manera los cálculos de probabilidad siempre pueden reducirse al caso estándar. La densidad típica es

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < \infty,$$

y la función de distribución correspondiente es

$$\Phi(x) = \int_{-\infty}^x \varphi(t) dt, \quad -\infty < x < \infty.$$

Distribución Lognormal

Si $\log V$ tiene una distribución normal, decimos que V tiene distribución lognormal. Recíprocamente, si $X \sim \mathcal{N}(\mu, \sigma^2)$ entonces $V = e^X$ es una variable con distribución lognormal. Haciendo un cambio de variable para la densidad obtenemos

$$f_V(v) = \frac{1}{\sqrt{2\pi\sigma v}} \exp\left\{-\frac{1}{2}\left(\frac{\log v - \mu}{\sigma}\right)^2\right\}, \quad v \geq 0.$$

La media y varianza son, respectivamente,

$$\begin{aligned} E[V] &= \exp\left\{\mu + \frac{1}{2}\sigma^2\right\} \\ \text{Var}[V] &= \exp\left\{2\left(\mu + \frac{1}{2}\sigma^2\right)\right\}(e^{\sigma^2} - 1) \end{aligned}$$

Distribución Exponencial

Una variable T no-negativa tiene una distribución exponencial con parámetro $\lambda > 0$ si su densidad es

$$f_T(t) = \begin{cases} \lambda e^{-\lambda t} & \text{para } t \geq 0, \\ 0 & \text{para } t < 0. \end{cases}$$

La función de distribución correspondiente es

$$F_T(t) = \begin{cases} 1 - e^{-\lambda t} & \text{para } t \geq 0, \\ 0 & \text{para } t < 0. \end{cases}$$

Usaremos la notación $X \sim \text{Exp}(\lambda)$. La media y varianza están dadas por

$$E[T] = \frac{1}{\lambda}, \quad \text{Var}[T] = \frac{1}{\lambda^2}.$$

Una de las propiedades fundamentales de distribución exponencial es la falta de memoria, que explicamos a continuación. Supongamos que T es el tiempo de vida de cierto componente y, dado que el componente ha durado hasta el instante t queremos obtener la distribución condicional del tiempo de vida remanente $T - t$. Equivalentemente, para $x > 0$ queremos determinar la distribución condicional $P(T > t + x | T > t)$. Aplicando la definición de probabilidad condicional obtenemos

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

Por lo tanto,

$$P(T > t + x | T > t) = e^{-\lambda x} = P(T > x)$$

y un componente que ha durado hasta el instante t tiene una vida remanente que es estadísticamente igual a la de un componente nuevo.

La *función de riesgo* ('hazard') o *de falla* $r(s)$ de una variable no-negativa S con densidad continua $g(s)$ y f.d. $G(s) < 1$ se define como

$$r(s) = \frac{g(s)}{1 - G(s)}, \quad \text{para } s > 0.$$

Calculemos ahora

$$\begin{aligned} P(s < S \leq s + \Delta s | S > s) &= \frac{P(s < S \leq s + \Delta s)}{P(S > s)} \\ &= \frac{g(s)\Delta s}{1 - G(s)} + o(\Delta s) \\ &= r(s)\Delta s + o(\Delta s). \end{aligned}$$

Por lo tanto un componente que ha durado hasta el tiempo s fallará en el intervalo $(s, s + \Delta s]$ con probabilidad condicional $r(s)\Delta s + o(\Delta s)$, lo que motiva el nombre de función de riesgo o fallas.

Podemos invertir la relación de la definición integrando

$$-r(s) = \frac{-g(s)}{1 - G(s)} = \frac{d[1 - G(s)]/ds}{1 - G(s)} = \frac{d(\log(1 - G(s)))}{ds}$$

para obtener

$$-\int_0^t r(s) ds = \log[1 - G(t)],$$

o

$$G(t) = 1 - \exp\left\{-\int_0^t r(s) ds\right\}, \quad t \geq 0,$$

que nos da la f.d. explícitamente en términos de la tasa de fallas.

La distribución exponencial puede ser caracterizada como la única distribución continua que tiene tasa de fallas constante $r(t) = \lambda$. La tasa de fallas no cambia en el tiempo, otra consecuencia de la propiedad de falta de memoria.

Distribución Uniforme

Una variable aleatoria U tiene distribución uniforme en el intervalo $[a, b]$, con $a < b$, si tiene la densidad de probabilidad

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

La función de distribución es

$$F_U(x) = \begin{cases} 0 & \text{para } x \leq a, \\ \frac{x-a}{b-a} & \text{para } a < x \leq b, \\ 1 & \text{para } x > b. \end{cases}$$

Usaremos la notación $U \sim \mathcal{U}[a, b]$. La media y varianza son

$$E[U] = \frac{1}{2}(a + b) \quad \text{y} \quad \text{Var}[U] = \frac{(b - a)^2}{12}.$$

Distribución Gamma

La distribución Gamma con parámetros $\alpha > 0$ y $\lambda > 0$ tiene densidad de probabilidad

$$f(x) = \frac{\lambda}{\Gamma(\alpha)} (\lambda x)^{\alpha-1} e^{-\lambda x}, \quad \text{para } x > 0,$$

donde

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx, \quad \text{para } \alpha > 0.$$

Usaremos la notación $X \sim \Gamma(\alpha, \lambda)$ para esta distribución.

Si α es entero y sumamos α variables exponenciales independientes con parámetro λ , esta suma X_α tiene distribución Gamma con parámetros α y λ . Los parámetros de esta distribución son

$$E[X_\alpha] = \frac{\alpha}{\lambda} \quad \text{Var}[X_\alpha] = \frac{\alpha}{\lambda^2}.$$

Distribución Beta

La densidad beta con parámetros $\alpha > 0$ y $\beta > 0$ es

$$f(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} & \text{para } 0 < x < 1, \\ 0 & \text{en otro caso.} \end{cases}$$

Los parámetros de esta distribución son

$$E[X] = \frac{\alpha}{\alpha + \beta} \quad \text{Var}[X] = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}.$$

Distribución Normal Bivariada

Sean $\sigma_X > 0$, $\sigma_Y > 0$, μ_X , μ_Y y ρ con $-1 < \rho < 1$ números reales. Para x e y reales definimos la forma cuadrática

$$Q(x, y) = \frac{1}{1 - \rho^2} \left\{ \left(\frac{x - \mu_X}{\sigma_X} \right)^2 - 2\rho \left(\frac{x - \mu_X}{\sigma_X} \right) \left(\frac{y - \mu_Y}{\sigma_Y} \right) + \left(\frac{y - \mu_Y}{\sigma_Y} \right)^2 \right\}$$

Definimos la distribución normal o gaussiana conjunta para las variables X e Y por la función de densidad

$$\phi_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2} Q(x, y) \right\},$$

para $-\infty < x, y < \infty$. Los momentos de la distribución son

$$E[X] = \mu_X, \quad E[Y] = \mu_Y, \quad \text{Var}[X] = \sigma_X^2, \quad \text{Var}[Y] = \sigma_Y^2,$$

y

$$\text{Cov}[X, Y] = E[(X - \mu_X)(Y - \mu_Y)] = \rho\sigma_X\sigma_Y.$$

Si definimos la matriz de varianzas y covarianzas del vector (X, Y) por

$$\Sigma = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix}$$

entonces podemos escribir la densidad anterior como

$$\phi_{\mathbf{X}}(\mathbf{x}) = \frac{1}{2\pi} \frac{1}{\sqrt{\det \Sigma}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}.$$

donde $\mathbf{X} = (X, Y)$, $\mathbf{x} = (x, y)$, $\det \Sigma$ es el determinante de Σ , Σ^{-1} la matriz inversa y \mathbf{x}' es el vector traspuesto de \mathbf{x} . Esta expresión para la densidad se puede generalizar al caso de vectores gaussianos de dimensión n .

1.9. Probabilidad y Esperanza Condicional

1.9.1. El Caso Discreto

Definición 1.3 Sean X, Y variables aleatorias discretas. La función o densidad de probabilidad condicional $p_{X|Y}(x|y)$ de X dado $Y = y$ se define por

$$p_{X|Y}(x|y) = \frac{P(X = x, Y = y)}{P(Y = y)} \quad \text{si } P(Y = y) > 0,$$

y no está definida, o se le asigna un valor arbitrario, si $P(Y = y) = 0$.

En términos de la densidad conjunta y de la densidad marginal de Y , $p_{X,Y}(x, y)$ y $p_Y(y) = \sum_x p_{X,Y}(x, y)$ respectivamente, la definición es

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}, \quad \text{si } p_Y(y) > 0.$$

Observamos que $p_{X|Y}(x|y)$ es una densidad de probabilidad en x para cada y fijo:

$$p_{X|Y}(x|y) \geq 0, \quad \sum_x p_{X|Y}(x|y) = 1, \quad \text{para todo } y.$$

Por lo tanto, podemos definir la función de distribución condicional de X dado que $Y = y$ como la f.d. asociada a la función de probabilidad $p_{X|Y}(x|y)$ (siempre que $p_Y(y) > 0$):

$$F_{X|Y}(x|y) = \sum_{z \leq x} p_{X|Y}(z|y) = \frac{1}{p_Y(y)} \sum_{z \leq x} p_{X,Y}(z, y).$$

La ley de la probabilidad total es

$$P(X = x) = \sum_y P(X = x|Y = y)P(Y = y) = \sum_y p_{X|Y}(x|y)p_Y(y).$$

Ejemplo 1.10

Supongamos que X tiene distribución binomial de parámetros p y N , donde N a su vez tiene distribución de Poisson con media λ . ¿Cuál es la distribución de X ?

Tenemos que

$$p_{X|N}(k|n) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \text{para } k = 0, 1, \dots, n$$

$$p_N(n) = \frac{\lambda^n e^{-\lambda}}{n!}, \quad \text{para } n = 0, 1, \dots$$

Usando la ley de la probabilidad total

$$\begin{aligned} P(X = k) &= \sum_{n=0}^{\infty} p_{X|N}(k|n)p_N(n) = \sum_{n=k}^{\infty} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \frac{\lambda^n e^{-\lambda}}{n!} \\ &= \frac{\lambda^k e^{-\lambda} p^k}{k!} \sum_{n=k}^{\infty} \frac{[\lambda(1-p)]^{n-k}}{(n-k)!} = \frac{(\lambda p)^k e^{-\lambda}}{k!} e^{\lambda(1-p)} \\ &= \frac{(\lambda p)^k e^{-\lambda p}}{k!} \end{aligned}$$

para $k = 0, 1, \dots$, es decir, $X \sim \mathcal{Pois}(\lambda p)$. ▲

Sea g una función tal que $E[g(X)] < \infty$. Definimos la esperanza condicional de $g(X)$ dado $Y = y$ por la fórmula

$$E[g(X)|Y = y] = \sum_x g(x)p_{X|Y}(x|y) \quad \text{si } p_Y(y) > 0,$$

y la esperanza condicional no está definida para valores y tales que $p_Y(y) = 0$. La ley de la probabilidad total para esperanzas condicionales es

$$E[g(X)] = \sum_y E[g(X)|Y = y]p_Y(y).$$

La esperanza condicional $E[g(X)|Y = y]$ es una función de la variable real y , que denotaremos $\varphi(y)$. Si evaluamos esta función φ en la variable aleatoria Y obtenemos una nueva variable aleatoria $\varphi(Y)$, que denotamos $E[g(X)|Y]$:

$$E[g(X)|Y](\omega) = E[g(X)|Y = Y(\omega)].$$

Podemos ahora escribir la ley de la probabilidad total como

$$E[g(X)] = E[E[g(X)|Y]]$$

Ejemplo 1.11

Consideremos un dado tetrahedral con 4 resultados posibles: 1, 2, 3 y 4 y probabilidades respectivas $p(i) = p_i$ para $i = 1, \dots, 4$. Lanzamos el dado dos veces y definimos X como el producto de los resultados y Y como su suma. A continuación presentamos una descripción de los resultados posibles del experimento y de los valores de las variables X e Y .

$\Omega :$	(1, 1)	(1, 2)	(1, 3)	(1, 4)	$X :$	1	2	3	4	$Y :$	2	3	4	5
	(2, 1)	(2, 2)	(2, 3)	(2, 4)		2	4	6	8		3	4	5	6
	(3, 1)	(3, 2)	(3, 3)	(3, 4)		3	6	9	12		4	5	6	7
	(4, 1)	(4, 2)	(4, 3)	(4, 4)		4	8	12	16		5	6	7	8

Calculemos ahora la probabilidad condicional $p_{X|Y}(x|y)$ para algunos valores de y . Por ejemplo, si $Y = 2$, el único resultado posible es (1, 1) y el valor de X en este caso es 1. Por lo tanto

$$p_{X|Y}(x|2) = \begin{cases} 1 & \text{si } x = 1, \\ 0 & \text{en otro caso.} \end{cases}$$

Algo similar ocurre cuando $y = 3, 7$ u 8 : Para cada uno de estos valores de la variable Y hay un sólo valor de la variable X , que por lo tanto ocurre condicionalmente con probabilidad 1.

Para los otros valores de y la situación es distinta, pues hay varios valores posibles de x . Veamos, como ejemplo, el caso $y = 5$; tenemos dos valores posibles de X : 4, que corresponde a los eventos elementales (4, 1) y (1, 4), y 6, que corresponde a los eventos elementales (3, 2) y (2, 3). Por lo tanto,

$$p_{X|Y}(4|5) = \frac{P(X = 4, Y = 5)}{P(Y = 5)} = \frac{p_1p_4}{p_1p_4 + p_2p_3},$$

$$p_{X|Y}(6|5) = \frac{P(X = 6, Y = 5)}{P(Y = 5)} = \frac{p_2p_3}{p_1p_4 + p_2p_3}.$$

De manera similar se calculan los otros valores de la función de probabilidad condicional, que son para $Y = 4$:

$$p_{X|Y}(3|4) = \frac{2p_1p_3}{2p_1p_3 + p_2^2}, \quad p_{X|Y}(4|4) = \frac{p_2^2}{2p_1p_3 + p_2^2},$$

para $Y = 6$:

$$p_{X|Y}(8|6) = \frac{2p_2p_4}{2p_2p_4 + p_3^2}, \quad p_{X|Y}(9|6) = \frac{p_3^2}{2p_2p_4 + p_3^2}.$$

En consecuencia vemos que para cada valor de la variable Y tenemos una función de probabilidad sobre los posibles valores de X . Veamos ahora los distintos valores de la esperanza condicional,

$$\begin{aligned} E[X|Y = 2] &= 1, & E[X|Y = 3] &= 2 \\ E[X|Y = 4] &= 3 \frac{2p_1p_3}{2p_1p_3 + p_2^2} + 4 \frac{p_2^2}{2p_1p_3 + p_2^2} \\ E[X|Y = 5] &= 4 \frac{p_1p_4}{p_1p_4 + p_2p_3} + 6 \frac{p_2p_3}{p_1p_4 + p_2p_3} \\ E[X|Y = 6] &= 8 \frac{2p_2p_4}{2p_2p_4 + p_3^2} + 9 \frac{p_3^2}{2p_2p_4 + p_3^2} \\ E[X|Y = 7] &= 12, & E[X|Y = 8] &= 16. \end{aligned}$$

Para el caso particular en el cual el dado es simétrico y todos los valores tienen la misma probabilidad los valores de las tres esperanzas centrales en la expresión anterior son

$$E[X|Y = 4] = \frac{10}{3}; \quad E[X|Y = 5] = 5; \quad E[X|Y = 6] = \frac{25}{3}.$$

Por lo tanto, $E[X|Y]$ es una función de los valores de Y , y como Y es una variable aleatoria, también lo es $E[X|Y]$. La siguiente tabla muestra los valores de Y , los valores asociados de $E[X|Y]$ y las probabilidades correspondientes, y representa una descripción de la variable aleatoria $E[X|Y]$.

y	$E[X Y = y]$	$P(Y = y)$
2	1	1/16
3	2	1/8
4	10/3	3/16
5	5	1/4
6	25/3	3/16
7	12	1/8
8	16	1/16

▲

Propiedades.

Como la esperanza condicional de $g(X)$ dado $Y = y$ es la esperanza respecto a la densidad de probabilidad condicional $p_{X|Y}(x|y)$, las esperanzas condicionales se comportan en muchos aspectos como esperanzas ordinarias.

Suponemos que X e Y tienen distribución conjunta, $c \in \mathbb{R}$, g es una función tal que $E[|g(X)|] < \infty$, h es una función acotada y ν es una función en \mathbb{R}^2 tal que $E[|\nu(X, Y)|] < \infty$.

- $E[c_1g_1(X_1) + c_2g_2(X_2)|Y = y] = c_1 E[g_1(X_1)|Y = y] + c_2 E[g_2(X_2)|Y = y]$.
- Si $g \geq 0$ entonces $E[g(X)|Y = y] \geq 0$.
- $E[\nu(X, Y)|Y = y] = E[\nu(X, y)|Y = y]$.
- $E[g(X)|Y = y] = E[g(X)]$ si X e Y son v.a.i.
- $E[g(X)h(Y)|Y = y] = h(y) E[g(X)|Y = y]$.
- $E[g(X)h(Y)] = \sum_y h(y) E[g(X)|Y = y]p_Y(y) = E[h(Y) E[g(X)|Y]]$.

Como consecuencia de 1, 5 y 6 obtenemos

- $E[c|Y = y] = c$,

8. $E[h(Y)|Y = y] = h(y)$,
9. $E[g(X)] = \sum_y E[g(X)|Y = y]p_Y(y) = E[E[g(X)|Y]]$.

Ejemplo 1.12

Usando la propiedad 9 podemos obtener la esperanza de X en el ejemplo anterior:

$$E[X] = E[E[X|Y]] = 1 \times \frac{1}{16} + 2 \times \frac{1}{8} + \dots + 16 \times \frac{1}{16} = 6.25$$

y hemos hallado la esperanza de X sin haber usado su función de distribución. ▲

1.9.2. El Caso Continuo

Sean X, Y v.a. con distribución conjunta continua de densidad $f_{X,Y}(x, y)$. Definimos la densidad condicional $f_{X|Y}(x|y)$ para la variable X dado que $Y = y$ por la fórmula

$$f_{X|Y}(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)} \quad \text{si } f_Y(y) > 0,$$

y no está definida si $f_Y(y) = 0$. La f.d. condicional para X dado $Y = y$ se define por

$$F_{X|Y}(x|y) = \int_{-\infty}^x \frac{f_{XY}(s, y)}{f_Y(y)} ds \quad \text{si } f_Y(y) > 0.$$

La densidad condicional tiene las propiedades que uno esperaría. En particular

$$P(a < X < b, c < Y < d) = \int_c^d \left(\int_a^b f_{X|Y}(x|y) dx \right) f_Y(y) dy$$

y haciendo $c = -\infty, d = \infty$ obtenemos la ley de probabilidad total

$$P(a < X < b) = \int_{-\infty}^{\infty} \left(\int_a^b f_{X|Y}(x|y) dx \right) f_Y(y) dy.$$

Ejemplo 1.13

Sea (X, Y) un vector gaussiano de dimensión 2, de densidad

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_X^2} - 2\rho\frac{xy}{\sigma_X\sigma_Y} + \frac{y^2}{\sigma_Y^2}\right)\right\} \quad (1.23)$$

La variable X es gaussiana, centrada, de varianza σ_X^2 y densidad

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left\{-\frac{x^2}{2\sigma_X^2}\right\}. \quad (1.24)$$

La densidad condicional de Y dado que $X = x$ es, por lo tanto,

$$f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2\sigma_Y^2(1-\rho^2)}\left(y - \frac{\sigma_Y}{\sigma_X}\rho x\right)^2\right\}$$

que es una densidad gaussiana de media $\rho x \sigma_Y / \sigma_X$ y varianza $\sigma_Y^2(1 - \rho^2)$. ▲

Si g es una función para la cual $E[|g(X)|] < \infty$, la esperanza condicional de $g(X)$ dado que $Y = y$ se define como

$$E[g(X)|Y = y] = \int g(x)f_{X|Y}(x|y) dx \quad \text{si } f_Y(y) > 0.$$

Estas esperanzas condicionales satisfacen también las propiedades 1-5 que listamos anteriormente. La propiedad 6 es en este caso,

$$E[g(X)h(Y)] = E[h(Y)E[g(X)|Y]] = \int h(y)E[g(X)|Y = y]f_Y(y) dy$$

válida para cualquier h acotada y suponiendo $E[|g(X)|] < \infty$. Cuando $h \equiv 1$ obtenemos

$$E[g(X)] = E[E[g(X)|Y]] = \int E[g(X)|Y = y]f_Y(y) dy.$$

Ejemplo 1.14

Si (X, Y) es un vector gaussiano bidimensional cuya densidad está dada por (1.23), entonces

$$E[Y|X = x] = \int_{-\infty}^{\infty} yf_{Y|X}(y|x) dy$$

es la esperanza condicional de Y dado que $X = x$. A partir de (1.24) vemos que la densidad condicional es gaussiana de media $\rho x \sigma_Y / \sigma_X$, de donde

$$E[Y|X] = \frac{\sigma_Y}{\sigma_X} \rho X.$$

▲

Podemos reunir los casos discreto y continuo en una sola expresión:

$$E[g(X)h(Y)] = E[h(Y)E[g(X)|Y]] = \int h(y)E[g(X)|Y = y]dF_Y(y)$$

y

$$E[g(X)] = E[E[g(X)|Y]] = \int E[g(X)|Y = y]dF_Y(y).$$

1.9.3. El Caso Mixto

Hasta ahora hemos considerado dos casos para el vector (X, Y) : ambas variables discretas o ambas continuas. En esta sección consideraremos los dos casos mixtos posibles. Comenzamos por el caso continuo-discreto. En ambos casos supondremos, sin pérdida de generalidad, que la variable discreta toma valores en los enteros positivos.

Caso 1: Continuo-Discreto

Sean X y N v.a. con distribución conjunta donde N toma valores $0, 1, 2, \dots$. La función de distribución condicional $F_{X|N}(x|n)$ de X dado que $N = n$ es

$$F_{X|N}(x|n) = \frac{P(X \leq x, N = n)}{P(N = n)} \quad \text{si } P(N = n) > 0,$$

y la función de distribución condicional no está definida para otros valores de n . Es sencillo verificar que $F_{X|N}(x|n)$ es una f.d. en x para cada valor fijo de n para el cual esté definida.

Supongamos que X es continua y $F_{X|N}(x|n)$ es diferenciable en x para todo n con $P(N = n) > 0$. Definimos la densidad condicional de X dado $N = n$ por

$$f_{X|N}(x|n) = \frac{d}{dx} F_{X|N}(x|n).$$

De nuevo, $f_{X|N}(x|n)$ es una densidad en x para los n para los cuales está definida y tiene las propiedades que uno esperaría, por ejemplo

$$P(a \leq X \leq b, N = n) = \int_a^b f_{X|N}(x|n) p_N(n) dx, \quad \text{para } a < b.$$

Usando la ley de la probabilidad total obtenemos la densidad marginal de X ,

$$f_X(x) = \sum_n f_{X|N}(x|n) p_N(n).$$

Supongamos que g es una función para la cual $E[|g(X)|] < \infty$. La esperanza condicional de $g(X)$ dado que $N = n$ se define por

$$E[g(X)|N = n] = \int g(x) f_{X|N}(x|n) dx.$$

Esta esperanza condicional satisface las propiedades anteriores y en este caso la ley de la probabilidad total es

$$E[g(X)] = \sum_n E[g(X)|N = n] p_N(n) = E[E[g(X)|N]].$$

Caso 2: Discreto-Continuo

Consideremos ahora un vector (N, X) . Supongamos que X tiene una distribución continua de densidad $f_X(x)$ y, dado el valor x de X , N es discreta con función de probabilidad $p_{N|X}(n|x)$ para $n \geq 0$. Podemos pensar que X es un parámetro (aleatorio) de la distribución de N , y una vez conocido el valor de este parámetro la distribución de N está completamente determinada.

La función de probabilidad condicional de N dado X es

$$p_{N|X}(n|x) = P(N = n|X = x) = \frac{f_{N,X}(n, x)}{f_X(x)}$$

siempre que $f_X(x) > 0$, donde $f_{N,X}(n, x)$ es la densidad de probabilidad conjunta del vector (N, X) . La función de distribución condicional correspondiente es

$$F_{N|X}(n, x) = \frac{1}{f_X(x)} \sum_{k=0}^n P(N = k|X = x) = \frac{1}{f_X(x)} \sum_{k=0}^n p_{N|X}(k|x)$$

Ejemplo 1.15

Suponemos que $X \sim \text{Bin}(p, N)$ con $p \sim \mathcal{U}[0, 1]$. ¿Cuál es la distribución de X ?

$$\begin{aligned} P(X = k) &= \int_{\mathbb{R}} P(X = k|p = \xi) f_p(\xi) d\xi \\ &= \int_0^1 \frac{N!}{k!(N-k)!} \xi^k (1-\xi)^{N-k} d\xi \\ &= \frac{N!}{k!(N-k)!} \frac{k!(N-k)!}{(N+1)!} = \frac{1}{N+1}, \quad k = 0, \dots, N. \end{aligned}$$

es decir, X tiene distribución uniforme en los enteros $0, 1, \dots, N$. ▲

Ejemplo 1.16

Sea $Y \sim \mathcal{Exp}(\theta)$ y dado $Y = y$, X tiene distribución de Poisson de media y . Queremos hallar la ley de X .

Usando la ley de probabilidad total

$$\begin{aligned} P(X = k) &= \int_0^\infty P(X = k|Y = y) f_Y(y) dy \\ &= \int_0^\infty \frac{y^k e^{-y}}{k!} \theta e^{-\theta y} dy \\ &= \frac{\theta}{k!} \int_0^\infty y^k e^{-(1+\theta)y} dy \\ &= \frac{\theta}{k!(1+\theta)^{k+1}} \int_0^\infty u^k e^{-u} du \\ &= \frac{\theta}{(1+\theta)^{k+1}}, \quad k = 0, 1, \dots \end{aligned}$$

▲

1.9.4. Sumas Aleatorias

Con frecuencia encontramos sumas de la forma $T = X_1 + \dots + X_N$, donde el número de sumandos es una variable aleatoria. Consideremos una sucesión X_1, X_2, \dots de v.a.i.i.d. y sea N una v.a. discreta, independiente de X_1, X_2, \dots con densidad $p_N(n) = P(N = n)$, $n = 0, 1, \dots$. Definimos la suma aleatoria T como

$$T = \begin{cases} 0 & \text{si } N = 0, \\ X_1 + \dots + X_N & \text{si } N > 0. \end{cases}$$

Ejemplos 1.17

- Colas: N representa el número de clientes, X_i es el tiempo de atención de cada cliente, T es el tiempo total de atención.
- Seguros: N representa el número de reclamos en un período de tiempo dado, X_i es el monto de cada reclamo y T es el monto total de los reclamos en el período.
- Población: N representa el número de plantas, X_i es el número de semillas de cada planta, T es el total de semillas.
- Biometría: N es el tamaño de la población, X_i es el peso de cada ejemplar y T representa el peso total de la muestra.

Momentos de una Suma Aleatoria

Supongamos que X_k y N tienen momentos finitos

$$\begin{aligned} E[X_k] &= \mu, & \text{Var}[X_k] &= \sigma^2, \\ E[N] &= \nu, & \text{Var}[N] &= \tau^2. \end{aligned}$$

y queremos determinar media y varianza de $T = X_1 + \dots + X_N$. Veamos que

$$E[T] = \mu\nu, \quad \text{Var}[T] = \nu\sigma^2 + \mu^2\tau^2.$$

Tenemos

$$\begin{aligned} E[T] &= \sum_{n=0}^{\infty} E[T|N = n]p_N(n) = \sum_{n=0}^{\infty} E[X_1 + \cdots + X_N|N = n]p_N(n) \\ &= \sum_{n=0}^{\infty} E[X_1 + \cdots + X_n|N = n]p_N(n) = \sum_{n=0}^{\infty} E[X_1 + \cdots + X_n]p_N(n) \\ &= \sum_{n=0}^{\infty} n\mu p_N(n) = \mu\nu. \end{aligned}$$

Para determinar la varianza comenzamos por

$$\begin{aligned} \text{Var}[T] &= E[(T - \mu\nu)^2] = E[(T - N\mu + N\mu - \nu\mu)^2] \\ &= E[(T - N\mu)^2] + E[\mu^2(N - \nu)^2] + 2E[\mu(T - N\mu)(N - \nu)]. \end{aligned}$$

Calculemos cada uno de estos sumandos por separado, el primero es

$$\begin{aligned} E[(T - N\mu)^2] &= \sum_{n=0}^{\infty} E[(T - N\mu)^2|N = n]p_N(n) \\ &= \sum_{n=1}^{\infty} E[(X_1 + \cdots + X_n - n\mu)^2|N = n]p_N(n) \\ &= \sum_{n=1}^{\infty} E[(X_1 + \cdots + X_n - n\mu)^2]p_N(n) \\ &= \sigma^2 \sum_{n=1}^{\infty} np_N(n) = \nu\sigma^2. \end{aligned}$$

Para el segundo tenemos

$$E[\mu^2(N - \nu)^2] = \mu^2 E[(N - \nu)^2] = \mu^2\tau^2$$

y finalmente el tercero es

$$\begin{aligned} E[\mu(T - N\mu)(N - \mu)] &= \mu \sum_{n=0}^{\infty} E[(T - n\mu)(n - \mu)|N = n]p_N(n) \\ &= \mu \sum_{n=0}^{\infty} (n - \mu) E[(T - n\mu)|N = n]p_N(n) \\ &= 0. \end{aligned}$$

La suma de estos tres términos demuestra el resultado.

Distribución de una Suma Aleatoria

Supongamos que los sumandos X_1, X_2, \dots son v.a.i. continuas con densidad de probabilidad $f(x)$. Para $n \geq 1$ fijo, la densidad de la suma $X_1 + \cdots + X_n$ es la n -ésima convolución de la densidad $f(x)$, que denotaremos por $f^{(n)}(x)$ y definiremos recursivamente por

$$\begin{aligned} f^{(1)}(x) &= f(x), \\ f^{(n)}(x) &= \int f^{(n-1)}(x-u)f(u) du \quad \text{para } n > 1. \end{aligned}$$

Como N y X_1, X_2, \dots son independientes, $f^{(n)}(x)$ es también la densidad condicional de $T = X_1 + \cdots + X_N$ dado que $N = n \geq 1$.

Supongamos que $P(N = 0) = 0$, es decir, que la suma aleatoria siempre tiene al menos un sumando. Por la ley de la probabilidad total, T es continua y tiene densidad marginal

$$f_T(x) = \sum_{n=1}^{\infty} f^{(n)}(x)p_N(n).$$

Observación 1.1 Si N puede valer 0 con probabilidad positiva entonces $T = X_1 + \dots + X_N$ es una v.a. mixta, es decir, tiene componentes discreta y continua. Si suponemos que X_1, X_2, \dots son continuas con densidad $f(x)$, entonces

$$P(T = 0) = P(N = 0) = p_N(0)$$

mientras que para $0 < a < b$ ó $a < b < 0$,

$$P(a < T < b) = \int_a^b \left(\sum_{n=1}^{\infty} f^{(n)}(x)p_N(n) \right) dx$$

▲

Ejemplo 1.18 (Suma Geométrica de Variables Exponenciales)

Supongamos que

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

$$p_N(n) = \beta(1 - \beta)^{n-1} \quad n = 1, 2, \dots$$

Comenzamos por hallar la convolución de las densidades exponenciales

$$\begin{aligned} f^{(2)}(x) &= \int f(x-u)f(u) du = \int \mathbf{1}_{\{x-u \geq 0\}}(u) \lambda e^{-\lambda(x-u)} \mathbf{1}_{\{u \geq 0\}}(u) \lambda e^{-u} \\ &= \lambda^2 e^{-\lambda x} \int_0^x du = x \lambda^2 e^{-\lambda x} \end{aligned}$$

para $x \geq 0$. La siguiente convolución es

$$\begin{aligned} f^{(3)}(x) &= \int f^{(2)}(x-u)f(u) du = \int \mathbf{1}_{\{x-u \geq 0\}}(u) \lambda^2 (x-u) e^{-\lambda(x-u)} \mathbf{1}_{\{u \geq 0\}}(u) \lambda e^{-u} du \\ &= \lambda^3 e^{-\lambda x} \int_0^x (x-u) du = \frac{x^2}{2} \lambda^3 e^{-\lambda x} \end{aligned}$$

para $x \geq 0$. Procediendo inductivamente obtenemos que

$$f^{(n)}(x) = \frac{x^{n-1}}{(n-1)!} \lambda^n e^{-\lambda x}$$

La densidad de $T = X_1 + \dots + X_N$ es

$$\begin{aligned} f_T(t) &= \sum_{n=1}^{\infty} f^{(n)}(t)p_N(n) = \sum_{n=1}^{\infty} \frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t} \beta(1 - \beta)^{n-1} \\ &= \lambda \beta e^{-\lambda t} \sum_{n=1}^{\infty} \frac{(\lambda(1 - \beta)t)^{n-1}}{(n-1)!} = \lambda \beta e^{-\lambda t} e^{\lambda(1-\beta)t} \\ &= \lambda \beta e^{-\lambda \beta t} \end{aligned}$$

para $t \geq 0$, y por lo tanto $T \sim \mathcal{Exp}(\lambda\beta)$.

▲

1.10. Funciones Generadoras de Probabilidad

Consideremos una v.a. ξ con valores enteros positivos y distribución de probabilidad

$$P(\xi = k) = p_k, \quad k = 0, 1, \dots$$

La función generadora de probabilidad (f.g.p.) $\phi(s)$ asociada a la v.a. ξ (o equivalentemente a su distribución (p_k)) se define por

$$\phi(s) = E[s^\xi] = \sum_{k=0}^{\infty} s^k p_k, \quad 0 \leq s \leq 1. \quad (1.25)$$

A partir de la definición es inmediato que si ϕ es una f.g.p. entonces

$$\phi(1) = \sum_{k=0}^{\infty} p_k = 1.$$

Resultados Fundamentales:

1. La relación entre funciones de probabilidad y funciones generadoras es 1-1. Es posible obtener las probabilidades (p_k) a partir de ϕ usando la siguiente fórmula

$$p_k = \frac{1}{k!} \left. \frac{d^k \phi(s)}{ds^k} \right|_{s=0}. \quad (1.26)$$

Por ejemplo,

$$\phi(s) = p_0 + p_1 s + p_2 s^2 + \dots \Rightarrow p_0 = \phi(0)$$

$$\frac{d\phi(s)}{ds} = p_1 + 2p_2 s + 3p_3 s^2 + \dots \Rightarrow p_1 = \left. \frac{d\phi(s)}{ds} \right|_{s=0}$$

2. Si ξ_1, \dots, ξ_n son v.a.i. con funciones generadoras $\phi_1(s), \phi_2(s), \dots, \phi_n(s)$ respectivamente, la f. g. p. de su suma $X = \xi_1 + \xi_2 + \dots + \xi_n$ es el producto de las funciones generadoras respectivas

$$\phi_X(s) = \phi_1(s)\phi_2(s)\dots\phi_n(s). \quad (1.27)$$

3. Los momentos de una variable que toma valores en los enteros no-negativos se pueden obtener derivando la función generadora:

$$\frac{d\phi(s)}{ds} = p_1 + 2p_2 s + 3p_3 s^2 + \dots,$$

por lo tanto

$$\left. \frac{d\phi(s)}{ds} \right|_{s=1} = p_1 + 2p_2 + 3p_3 + \dots = E[\xi]. \quad (1.28)$$

Para la segunda derivada tenemos

$$\frac{d^2\phi(s)}{ds^2} = 2p_2 + 3 \cdot 2p_3 s + 4 \cdot 3p_4 s^2 + \dots,$$

evaluando en $s = 1$,

$$\begin{aligned} \left. \frac{d^2\phi(s)}{ds^2} \right|_{s=1} &= 2p_2 + 3 \cdot 2p_3 + 4 \cdot 3p_4 \dots \\ &= \sum_{k=2}^{\infty} k(k-1)p_k \\ &= E[\xi(\xi-1)] = E[\xi^2] - E[\xi] \end{aligned} \quad (1.29)$$

de modo que

$$\mathbb{E}[\xi^2] = \left. \frac{d^2 \phi(s)}{ds^2} \right|_{s=1} + \mathbb{E}[\xi] = \left. \frac{d^2 \phi(s)}{ds^2} \right|_{s=1} + \left. \frac{d\phi(s)}{ds} \right|_{s=1},$$

y en consecuencia

$$\text{Var}[\xi] = \mathbb{E}[\xi^2] - (\mathbb{E}[\xi])^2 = \left. \frac{d^2 \phi(s)}{ds^2} \right|_{s=1} + \left. \frac{d\phi(s)}{ds} \right|_{s=1} - \left(\left. \frac{d^2 \phi(s)}{ds^2} \right|_{s=1} \right)^2.$$

Ejemplo 1.19

Supongamos que $\xi \sim \text{Pois}(\lambda)$:

$$p_k = P(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \dots$$

Su función generadora de probabilidad es

$$\begin{aligned} \phi(s) &= \mathbb{E}[s^\xi] = \sum_{k=0}^{\infty} s^k \frac{\lambda^k}{k!} e^{-\lambda} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(s\lambda)^k}{k!} = e^{-\lambda} e^{\lambda s} \\ &= e^{-\lambda(1-s)} \end{aligned}$$

Entonces,

$$\left. \frac{d\phi(s)}{ds} \right|_{s=1} = \lambda e^{-\lambda(1-s)}, \quad \left. \frac{d\phi(s)}{ds} \right|_{s=1} = \lambda \quad (1.30)$$

$$\left. \frac{d^2 \phi(s)}{ds^2} \right|_{s=1} = \lambda^2 e^{-\lambda(1-s)}, \quad \left. \frac{d^2 \phi(s)}{ds^2} \right|_{s=1} = \lambda^2 \quad (1.31)$$

y obtenemos

$$\mathbb{E}[\xi] = \lambda, \quad \text{Var}(\xi) = \lambda^2 + \lambda - (\lambda)^2 = \lambda.$$

▲

1.10.1. Funciones Generadoras de Probabilidad y Sumas de V. A. I.

Sean ξ, η v.a.i. con valores $0, 1, 2, \dots$ y con funciones generadoras de probabilidad

$$\phi_\xi(s) = \mathbb{E}[s^\xi], \quad \phi_\eta(s) = \mathbb{E}[s^\eta], \quad |s| < 1,$$

entonces la f.g.p. de la suma $\xi + \eta$ es

$$\phi_{\xi+\eta}(s) = \mathbb{E}[s^{\xi+\eta}] = \mathbb{E}[s^\xi s^\eta] = \mathbb{E}[s^\xi] \mathbb{E}[s^\eta] = \phi_\xi(s) \phi_\eta(s) \quad (1.32)$$

El recíproco también es cierto, si $\phi_{\xi+\eta}(s) = \phi_\xi(s) \phi_\eta(s)$ entonces las variables ξ y η son independientes.

Como consecuencia, si $\xi_1, \xi_2, \dots, \xi_m$ son v.a.i.i.d. con valores en $\{0, 1, 2, \dots\}$ y f.g.p. $\phi(s) = \mathbb{E}[s^\xi]$ entonces

$$\mathbb{E}[s^{\xi_1 + \dots + \xi_m}] = \phi^m(s) \quad (1.33)$$

¿Qué ocurre si el número de sumandos es aleatorio?

Proposición 1.3 Sea N una v.a. con valores enteros no-negativos e independiente de ξ_1, ξ_2, \dots con f.g.p. $g_N(s) = E[s^N]$ y consideremos la suma

$$X = \xi_1 + \dots + \xi_N.$$

Sea $h_X(s) = E[s^X]$ la f.g.p. de X . Entonces

$$h_X(s) = g_N(\phi(s)). \quad (1.34)$$

Demostración.

$$\begin{aligned} h_X(s) &= \sum_{k=0}^{\infty} P(X = k) s^k \\ &= \sum_{k=0}^{\infty} \left(\sum_{n=0}^{\infty} P(X = k | N = n) P(N = n) \right) s^k \\ &= \sum_{k=0}^{\infty} \left(\sum_{n=0}^{\infty} P(\xi_1 + \dots + \xi_n = k | N = n) P(N = n) \right) s^k \\ &= \sum_{k=0}^{\infty} \left(\sum_{n=0}^{\infty} P(\xi_1 + \dots + \xi_n = k) P(N = n) \right) s^k \\ &= \sum_{n=0}^{\infty} \left(\sum_{k=0}^{\infty} P(\xi_1 + \dots + \xi_n = k) s^k \right) P(N = n) \\ &= \sum_{n=0}^{\infty} \phi^n(s) P(N = n) = g_N(\phi(s)) \end{aligned}$$

■

Ejemplo 1.20

Sea N una variable aleatoria con distribución de Poisson de parámetro λ . Dado el valor de N , realizamos N experimentos de Bernoulli con probabilidad de éxito p y llamamos X al número de éxitos. En este caso ξ_i tiene distribución de Bernoulli y su f.g.p. es

$$\phi_\xi(s) = E[s^{\xi}] = sp + q$$

mientras que $N \sim \mathcal{Pois}(\lambda)$ con f.g.p.

$$g_N(s) = E[s^N] = e^{-\lambda(1-s)}$$

según vimos en el ejemplo 1.19. Por la proposición anterior obtenemos que la f.g.p. de X es

$$h_X(s) = g_N(\phi_\xi(s)) = g_N(q + sp) = \exp \left\{ -\lambda(1 - q - sp) \right\} = \exp \left\{ -\lambda p(1 - s) \right\}$$

que es la f.g.p. de una distribución de Poisson de parámetro λp .

▲

1.11. Funciones Generadoras de Momentos.

Dada una variable aleatoria X , o su función de distribución F , vamos a definir otra función generadora, como

$$M_X(t) = E(e^{tX}).$$

siempre que este valor esperado exista.

Notemos que cuando el recorrido de X son los enteros no-negativos, $M_X(t) = \phi_X(e^t)$. Si X está acotada, M_X está bien definida para todo t real; en cambio, si X no está acotada, es posible que el dominio de M no sea el conjunto de todos los reales. En todo caso, p siempre está definida en cero, y $M(0) = 1$.

Si la función M está definida en un entorno de $t = 0$, entonces las series

$$M_X(t) = E(e^{tX}) = E\left(1 + \sum_{n=1}^{\infty} \frac{t^n X^n}{n!}\right) = 1 + \sum_{n=1}^{\infty} \frac{t^n}{n!} E(X^n)$$

son convergentes y en consecuencia se puede derivar término a término. Obtenemos

$$M'_X(0) = E(X); \quad M''_X(0) = E(X^2) \quad \text{y en general } M_X^{(n)}(0) = E(X^n).$$

Es por esta última propiedad que esta función se conoce como *función generadora de momentos* (f.g.m.).

Ejemplos 1.21

1. Si $X \sim Bin(n, p)$ veamos que $M(t) = (pe^t + 1 - p)^n$: Un cálculo directo muestra que

$$M(t) = \sum_{j=0}^n e^{jt} \binom{n}{j} p^j (1-p)^{n-j} = (pe^t + 1 - p)^n,$$

2. Si $X \sim Exp(\lambda)$, es decir, si $P(X \leq x) = 1 - e^{-\lambda x}$, para $x \geq 0$, entonces $M(t) = \lambda/(\lambda - t)$ para $t < \lambda$.

El resultado se obtiene a partir del cálculo

$$M(t) = \int_0^{\infty} \lambda e^{-\lambda x} e^{tx} dx = \lambda \left. \frac{e^{(t-\lambda)x}}{t-\lambda} \right|_0^{\infty} = \frac{\lambda}{\lambda - t}.$$

Observamos que en este caso, $M(t)$ no está definida si $t \geq \lambda$.

3. Si $X \sim \mathcal{N}(0, 1)$, es decir, si $P(X \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-x^2/2} dx$, entonces $M(t) = e^{t^2/2}$.

Calculemos

$$M(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-t)^2} e^{t^2/2} dx = e^{t^2/2}$$

ya que $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-t)^2} dx = 1$ puesto que el integrando es la densidad de una variable aleatoria con distribución $\mathcal{N}(t, 1)$

Observación 1.2 Por la forma en la cual hemos definido la función generadora de momentos, cuando las f.g.m. de dos variables aleatorias X_1, X_2 coinciden para todos los valores de t en un entorno de $t = 0$, entonces las distribuciones de probabilidad de X_1 y X_2 deben ser idénticas. Este resultado lo enunciamos en el próximo teorema, sin demostración

Teorema 1.5 Si X tiene función generadora de momentos $M(t)$ que está definida en un entorno $(-a, a)$ de 0, entonces $M(t)$ caracteriza a la distribución de X , es decir, si otra variable Y tiene la misma función generadora de momentos, las distribuciones de X e Y coinciden.

La función generadora de momentos resulta particularmente útil cuando consideramos sucesiones de variables aleatorias, como lo muestra el siguiente teorema que enunciamos sin demostración.

Teorema 1.6 (de Continuidad) Sea $F_n(x)$, $n \geq 1$ una sucesión de f.d. con funciones generadores de momento respectivas $M_n(t)$, $n \geq 1$, que están definidas para $|t| < b$. Supongamos que cuando $n \rightarrow \infty$, $M_n(t) \rightarrow M(t)$ para $|t| \leq a < b$, donde $M(t)$ es la función generadora de momentos de la distribución $F(x)$. Entonces $F_n(x) \rightarrow F(x)$ cuando $n \rightarrow \infty$ para todo punto x en el cual F es continua.

Veamos una aplicación del teorema anterior para demostrar el Teorema de de Moivre y Laplace.

Teorema 1.7 (de Moivre-Laplace) Sea $S_n \sim \text{Bin}(n, p)$ para $n \geq 1$ y $q = 1 - p$. Definimos

$$T_n = \frac{S_n - np}{(npq)^{1/2}}$$

Entonces para todo $x \in \mathbb{R}$,

$$P(T_n \leq x) \rightarrow \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

Demostración. Recordemos que S_n es la suma de n v.a.i. con distribución de Bernoulli de parámetro p : $S_n = \sum_{i=1}^n X_i$. Usamos esto para calcular la función generadora de momentos de T_n .

$$\begin{aligned} \mathbb{E}(e^{tT_n}) &= \mathbb{E} \left[\exp \left(\frac{t(S_n - np)}{(npq)^{1/2}} \right) \right] = \mathbb{E} \left[\exp \left(\frac{t(\sum_{i=1}^n (X_i - p))}{(npq)^{1/2}} \right) \right] \\ &= \mathbb{E} \left[\prod_{i=1}^n \exp \left(\frac{t(X_i - p)}{(npq)^{1/2}} \right) \right] = \prod_{i=1}^n \mathbb{E} \left[\exp \left(\frac{t(X_i - p)}{(npq)^{1/2}} \right) \right] \\ &= \left(\mathbb{E} \left[\exp \left(\frac{t(X_1 - p)}{(npq)^{1/2}} \right) \right] \right)^n \\ &= \left(p \exp \left(\frac{t(1-p)}{(npq)^{1/2}} \right) + q \exp \left(\frac{-pt}{(npq)^{1/2}} \right) \right)^n. \end{aligned} \quad (1.35)$$

Ahora hacemos un desarrollo de Taylor para las dos exponenciales que aparecen en esta última expresión para obtener

$$p \exp \left(\frac{t(1-p)}{(npq)^{1/2}} \right) = p \left(1 + \frac{qt}{(npq)^{1/2}} + \frac{q^2 t^2}{2npq} + \frac{C_1 q^3 t^3}{3!(npq)^{3/2}} \right) \quad (1.36)$$

$$q \exp \left(\frac{-pt}{(npq)^{1/2}} \right) = q \left(1 - \frac{pt}{(npq)^{1/2}} + \frac{p^2 t^2}{2npq} + \frac{C_2 p^3 t^3}{3!(npq)^{3/2}} \right). \quad (1.37)$$

La suma de estas dos expresiones nos da $1 + \frac{t^2}{2n} + O(n^{-3/2})$ y sustituyendo en (1.35) obtenemos

$$\mathbb{E}(e^{tT_n}) = \left(1 + \frac{t^2}{2n} + O(n^{-3/2}) \right)^n \rightarrow e^{t^2/2}$$

que es la f.g.m. de la distribución normal típica. ■

1.12. Simulación de Variables Aleatorias

Los generadores de números aleatorios simulan valores de la distribución $\mathcal{U}[0, 1]$, pero con frecuencia nos interesa simular valores de otras distribuciones. Vamos a estudiar en esta sección dos métodos para generar valores a partir de una función de distribución F .

1.12.1. Método de la Distribución Inversa

Este método se basa en el siguiente resultado:

Proposición 1.4 Sea X una variable aleatoria con función de distribución F_X y sea g una función estrictamente creciente. Definimos $Y = g(X)$ y sea F_Y la función de distribución de esta variable. Entonces

$$F_Y(y) = F_X(g^{-1}(y)). \quad (1.38)$$

Demostración. Como g es estrictamente creciente los eventos $\{X \leq g^{-1}(y)\}$ y $\{g(X) \leq y\}$ son iguales. Por lo tanto,

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$$

■

Si g es estrictamente decreciente entonces $F_Y(y) = 1 - F_X(g^{-1}(y))$.

Corolario 1.3 Sea F una función de distribución estrictamente creciente para los x tales que $0 < F(x) < 1$ y sea $U \sim \mathcal{U}[0, 1]$. Entonces la variable $Z = F^{-1}(U)$ tiene distribución F .

Demostración. La función de distribución de U es $F_U(u) = u$ para $u \in [0, 1]$. Entonces

$$F_Z(z) = F_U(F(z)) = F(z) \quad (1.39)$$

de modo que Z tiene función de distribución F . ■

Observación 1.3 El resultado anterior es cierto en general si utilizamos la *inversa generalizada* F^{\leftarrow} de la función F cuando esta no sea estrictamente creciente, que se define por la siguiente expresión:

$$F^{\leftarrow}(y) = \inf\{x : F(x) \geq y\}$$

Por lo tanto, para cualquier función de distribución F , la variable aleatoria $Z = F^{\leftarrow}(U)$ tiene función de distribución F . Para ver que esto es cierto observamos que, a partir de la definición, es fácil demostrar que

$$F^{\leftarrow}(y) \leq t \Leftrightarrow y \leq F(t); \quad F^{\leftarrow}(y) > t \Leftrightarrow y > F(t).$$

Usando esto obtenemos

$$F_Z(z) = P(Z \leq z) = P(F^{\leftarrow}(U) \leq z) = P(U \leq F(z)) = F(z).$$

■

El Corolario 1.3 y la Observación 1.3 nos dan un método para simular una variable aleatoria con función de distribución F : Generamos el valor u de una variable uniforme en $[0, 1]$ y evaluamos la inversa generalizada en u : $F^{\leftarrow}(u)$. Sin embargo, dependiendo de la naturaleza de la función de distribución F , es posible que la inversa generalizada tenga una expresión complicada o incluso no sea posible escribirla en términos de funciones elementales, como ocurre en el caso de las variables Gaussianas. Por esta razón hay métodos particulares que resultan más eficientes en muchos casos.

Ejemplos 1.22

1. **Variables Discretas.** Si queremos simular una variable aleatoria finita X con valores x_1, \dots, x_n y probabilidades respectivas p_1, \dots, p_n , podemos dividir el intervalo $[0, 1]$ en subintervalos usando las probabilidades p_i :

$$[0, p_1); \quad [p_1, p_1 + p_2); \quad [p_1 + p_2, p_1 + p_2 + p_3); \quad \dots \quad \left[\sum_{j < n} p_j, 1 \right].$$

Ahora generamos una variable U con distribución uniforme en $[0, 1]$ y si el valor cae en el i -ésimo intervalo le asignamos a X el valor x_i . Como la probabilidad de que U caiga en el intervalo i es igual a la longitud del intervalo, que es p_i , vemos que

$$P(X = x_i) = p_i, \quad \text{para } 1 \leq i \leq n.$$

Esta es una implementación del método de la distribución inversa. Desde el punto de vista computacional es conveniente ordenar los valores según el tamaño de las p_i , colocando estas probabilidades de mayor a menor, porque para identificar el intervalo en cual cae U tenemos que comparar con p_1 , luego con $p_1 + p_2$, y así sucesivamente hasta obtener el primer valor mayor que U . Ordenar las probabilidades hace que se maximice la probabilidad de que U esté en los primeros intervalos, y esto reduce el número de comparaciones que hay que hacer en promedio para obtener el valor de X .

Este método también funciona para variables discretas con una cantidad infinita de valores. La misma observación sobre el ordenamiento de los valores de las probabilidades es válida.

2. **Distribución de Bernoulli.** Un caso particular sencillo es el de la distribución de Bernoulli con probabilidad de éxito p . Para generar un valor de la variable X con esta distribución, generamos U y si $U < p$, $X = 1$ y si no, $X = 0$.
3. **Distribución Uniforme Discreta.** Sea X una variable aleatoria que toma valores $\{x_1, x_2, \dots, x_n\}$ con igual probabilidad. Para simular esta distribución generamos un número aleatorio $U \in (0, 1]$, dividimos el intervalo $[0, 1]$ en n intervalos iguales y le asignamos a la variables el valor x_k si

$$\frac{k-1}{n} < U \leq \frac{k}{n},$$

es decir, el valor de la variable es x_k con $k = \lceil Un \rceil$, donde $\lceil a \rceil$ es la función *techo* y representa el menor entero que es mayor o igual a a .

4. **Variables Continuas.** Si X es una variable continua con función de distribución F invertible, para simular X basta generar una variable uniforme U y poner $X = F^{-1}(U)$. Esto es consecuencia del corolario 1.3. Por ejemplo, si queremos simular una v.a. X con función de distribución $F(x) = x^n$ para $0 < x < 1$, observamos que F es invertible y su inversa es $F^{-1}(u) = u^{1/n}$. Por lo tanto basta generar una variables uniforme U y poner $X = U^{1/n}$.
5. **Distribución Uniforme Continua.** Si queremos simular la distribución $\mathcal{U}[a, b]$ generamos U uniforme en $[0, 1]$ y usamos la transformación $u \mapsto a + u(b - a)$.
6. **Distribución Exponencial.** Si $X \sim \text{Exp}(\lambda)$ su f.d. está dada por $F(x) = 1 - e^{-\lambda x}$. La inversa de esta función es

$$F^{-1}(u) = -\frac{1}{\lambda} \log(1 - u).$$

Por lo tanto para generar X podemos generar una uniforme U y ponemos $X = -\ln(1 - U)/\lambda$. Observamos ahora que si U tiene distribución uniforme en $(0, 1)$, $1 - U$ también. Por lo tanto, para simular esta distribución a partir de una variable $U \sim \mathcal{U}(0, 1)$ basta hacer la transformación $-\ln(U)/\lambda$.

1.12.2. Método de Rechazo

Variables Discretas

Supongamos que tenemos un método eficiente para simular una variable Y que tiene función de probabilidad $\{q_j, j \geq 1\}$. Podemos usar este método como base para simular otra variable X con función de probabilidad diferente $\{p_j, j \geq 1\}$, siempre que las dos variables tengan el mismo conjunto de valores posibles o al menos cuando los valores de X sean un subconjunto de los valores de Y . La idea es simular

primero la variable Y y luego aceptar este valor para la variable X con probabilidad proporcional a p_Y/q_Y .

Sea c una constante tal que

$$\frac{p_j}{q_j} \leq c \quad \text{para todo } j \text{ tal que } p_j > 0, \quad (1.40)$$

entonces el algoritmo para el método de rechazo es el siguiente,

Algoritmo.

- Paso 1. Simulamos una variable Y con función de probabilidad q_j .
- Paso 2. Generamos una variable uniforme U .
- Paso 3. Si $U < p_Y/cq_Y$, ponemos $X = Y$ y paramos. Si no, regresamos al paso 1.

Veamos que este método efectivamente produce una variable con distribución p_j . Calculemos primero la probabilidad de obtener el valor j en una sola iteración:

$$\begin{aligned} P(Y = j \text{ y este valor sea aceptado}) &= P(Y = j)P(\text{Aceptar}|Y = j) \\ &= q_j P\left(U < \frac{p_j}{cq_j}\right) \\ &= q_j \frac{p_j}{cq_j} = \frac{p_j}{c}. \end{aligned}$$

Si sumamos ahora sobre los valores posibles j obtenemos la probabilidad de que el valor de la variable generada sea aceptado:

$$P(\text{Aceptar el valor de } Y) = \sum_j \frac{p_j}{c} = \frac{1}{c},$$

Es decir, cada interacción resulta en un valor que es aceptado con probabilidad $1/c$ y esto ocurre de manera independiente, de modo que la distribución del número de iteraciones necesarias para aceptar un valor es geométrica con parámetro $1/c$. En consecuencia

$$\begin{aligned} P(X = j) &= \sum_n P(j \text{ es aceptado en la iteración } n) \\ &= \sum_n \left(1 - \frac{1}{c}\right)^{n-1} \frac{p_j}{c} = p_j. \end{aligned}$$

Como el número de iteraciones es geométrico con parámetro $1/c$, en promedio es necesario realizar c iteraciones para aceptar un valor. Por lo tanto conviene escoger c lo más pequeño posible, siempre que satisfaga (1.40).

Ejemplo 1.23

Supongamos que queremos generar una variable aleatoria con la siguiente distribución: $P(X = j) = p_j$ para $j = 1, 2, 3, 4$ y $p_1 = 0.20, p_2 = 0.15, p_3 = 0.25, p_4 = 0.4$ usando el método de rechazo. Vamos a usar una variable Y con distribución uniforme sobre los valores $1, 2, 3, 4$ y por lo tanto podemos tomar

$$c = \max\left\{\frac{p_j}{q_j} : 1 \leq j \leq 4\right\} = \frac{0.4}{0.25} = 1.6$$

y utilizar el algoritmo descrito anteriormente. En este caso en promedio hacemos 1.6 iteraciones por cada valor aceptado para la variable que queremos generar.

Variables Continuas

Este método funciona exactamente igual que en el caso discreto. Supongamos que tenemos una manera eficiente de generar una variable aleatoria con densidad $g(x)$ y queremos generar otra variable que tiene densidad $f(x)$ con el mismo conjunto de valores posibles. Podemos hacer esto generando Y con distribución g y luego aceptando este valor con probabilidad proporcional a $f(Y)/g(Y)$.

Sea c una constante tal que

$$\frac{f(y)}{g(y)} \leq c \quad \text{para todo } y,$$

entonces tenemos el siguiente algoritmo para generar una variable con densidad f .

Algoritmo.

- Paso 1. Generamos Y con densidad g .
- Paso 2. Generamos un número aleatorio U .
- Paso 3. Si $U \leq \frac{f(Y)}{cg(Y)}$ ponemos $X = Y$ y paramos. Si no, volvemos al paso 1.

Al igual que en caso discreto tenemos el siguiente resultado que justifica el método y que presentamos sin demostración.

Teorema

- (i) La variable generada con el método del rechazo tiene densidad f .
- (ii) El número de iteraciones necesarias en el algoritmo es una variable geométrica con media c .

Ejemplo 1.24

Vamos a usar el método de rechazo para generar una variable aleatoria con densidad

$$f(x) = 20x(1-x)^3, \quad 0 < x < 1.$$

Como esta variable aleatoria está concentrada en el intervalo $(0, 1)$, usaremos el método de rechazo con la distribución uniforme

$$g(x) = 1, \quad 0 < x < 1.$$

Para determinar la menor constante c que satisface $f(x)/g(x) < c$ para todo $x \in (0, 1)$ calculamos el máximo de

$$\frac{f(x)}{g(x)} = 20x(1-x)^3.$$

Derivando esta expresión e igualando a cero obtenemos la ecuación

$$20[(1-x)^3 - 3x(1-x)^2] = 0$$

con soluciones 1 y $1/4$. Esta última solución corresponde al máximo y por lo tanto

$$\frac{f(1/4)}{g(1/4)} = 20 \frac{1}{4} \left(\frac{3}{4}\right)^3 = \frac{135}{64} \equiv c.$$

En consecuencia

$$\frac{f(x)}{cg(x)} = \frac{256}{27} x(1-x)^3$$

y el algoritmo es

Algoritmo.

- Paso 1. Generamos dos números aleatorios U_1 y U_2 .
- Paso 2. Si $U_2 \leq 256U_1(1-U_1)^3/27$ ponemos $X = U_1$ y paramos. Si no, volvemos al paso 1.

En promedio, el paso 1 se realiza $c = \frac{256}{27} \approx 2.11$ veces por cada número generado.

1.12.3. Métodos Particulares

Distribución Binomial

Una manera sencilla de simular una variable con distribución binomial de parámetros n y p es generar n variables de Bernoulli con probabilidad de éxito p y sumarlas. Esto resulta un poco pesado si n es grande, pero en este caso podemos usar el Teorema Central del Límite, (teorema 1.7).

Otra posibilidad es usar el método de la transformada inversa junto con la siguiente relación iterativa para la distribución binomial:

$$\frac{P(S_n = i + 1)}{P(S_n = i)} = \frac{n!(n-i)!}{(i+1)!(n-i-1)!n!} \frac{p^{i+1}(1-p)^{n-i-1}}{p^i(1-p)^{n-i}} = \frac{n-i}{i+1} \frac{p}{1-p},$$

es decir,

$$P(S_n = i + 1) = \frac{n-i}{i+1} \frac{p}{1-p} P(S_n = i).$$

En consecuencia, generamos una variable uniforme U y comparamos con $P(X = 0) = (1-p)^n$. Si U es menor que este valor ponemos $X = 0$, en caso contrario multiplicamos $P(X = 0)$ por $pn/(1-p)$ para obtener $P(X = 1)$ y comparamos. Si U es menor que este valor ponemos $X = 1$, en caso contrario repetimos el procedimiento hasta conseguir el valor de X . El algoritmo se puede describir como sigue:

- Paso 1: Generamos una variable uniforme U .
- Paso 2: Ponemos $a = p/(1-p)$; $b = (1-p)^n$; $c = b$; $i = 0$.
- Paso 3: Si $U < c$ ponemos $X = i$ y paramos.
- Paso 4: $b = ab(n-i)/(i+1)$; $c = c + b$; $i = i + 1$.
- Paso 5: Vamos al paso 3.

Distribución de Poisson

Al igual que para la distribución binomial, tenemos una relación recursiva para la función de probabilidad que permite aplicar el método de la transformada inversa para generar la distribución de Poisson:

$$P(X = i + 1) = \frac{\lambda}{i + 1} P(X = i),$$

que es sencilla de demostrar. El algoritmo es el siguiente:

- Paso 1: Generamos una variable uniforme U .
- Paso 2: Ponemos $a = e^{-\lambda}$; $b = a$; $i = 0$.
- Paso 3: Si $U < b$ ponemos $X = i$ y paramos.
- Paso 4: $a = \lambda a/(i + 1)$; $b = b + a$; $i = i + 1$.
- Paso 5: Vamos al paso 3.

Distribución Geométrica

Una manera de generar variables con distribución geométrica es generar una sucesión de variables de Bernoulli hasta obtener el primer éxito, es decir, generamos una sucesión de números aleatorios en $[0, 1]$ hasta obtener el primero que sea menor que p . Sin embargo, si p es pequeño esto puede ser lento (toma en promedio $1/p$ pasos). Para evitar esto podemos seguir el método alternativo que describimos a continuación. Sea X una v.a. con distribución geométrica de parámetro p , $0 < p < 1$ y sea U un número aleatorio en $[0, 1]$. Definimos Y como el menor entero que satisface la desigualdad $1 - q^Y \geq U$. Entonces

$$\begin{aligned} P(Y = j) &= P(1 - q^j \geq U > 1 - q^{j-1}) \\ &= q^{j-1} - q^j = q^{j-1}(1 - q) = q^{j-1}p, \end{aligned}$$

de modo que Y también tiene una distribución geométrica de parámetro p . Por lo tanto, para generar Y basta resolver la ecuación que la define, es decir,

$$Y = \left\lceil \frac{\log(1-u)}{\log q} \right\rceil$$

pero como $1-u$ y u tienen la misma distribución, podemos usar

$$Y = \left\lceil \frac{\log(u)}{\log q} \right\rceil.$$

Distribución Binomial Negativa

Observamos que una variable con distribución binomial negativa de parámetros k y p es la suma de k variables geométricas con parámetro p : una por cada éxito en la sucesión de ensayos. Esta observación es útil para generar variables con esta distribución: si u_j , $j = 1, \dots, k$ son números aleatorios en $[0, 1]$, la siguiente expresión produce el valor de una variable con distribución binomial negativa:

$$\sum_{j=1}^k \left\lceil \frac{\log(u_j)}{\log q} \right\rceil.$$

Distribución Normal

La función de distribución normal Φ no se puede escribir en términos de funciones simples, y lo mismo ocurre con su inversa, lo que dificulta la aplicación del método de la transformada inversa. Sin embargo existen otros métodos y uno de los más populares es el de Box-Muller, también conocido como el método polar.

Aún cuando la justificación del método no es complicada, requiere algunos conceptos que no hemos introducido, así que vamos a describir el método sin demostrar que efectivamente lo que obtenemos es el valor de una variable normal. El algoritmo es el siguiente:

- Paso 1: Generamos variables uniformes U_1 y U_2 .
- Paso 2: Ponemos $V_1 = 2U_1 - 1$; $V_2 = 2U_2 - 1$; $S = V_1^2 + V_2^2$.
- Paso 3: Si $S > 1$ regresamos al paso 1.
- Paso 4: X y Y son variables normales típicas independientes:

$$X = \sqrt{\frac{-2 \log S}{S}} V_1, \quad Y = \sqrt{\frac{-2 \log S}{S}} V_2.$$

1.12.4. Generación de Variables Aleatorias en R

El lenguaje R tiene incorporadas una serie de rutinas para generar variables aleatorias. La sintaxis precisa de la instrucción correspondiente depende de la distribución, pero todas tienen el formato común `rdist`, donde *dist* designa la distribución; por ejemplo, para generar valores a partir de la distribución normal usamos `rnorm`. Según la distribución, puede ser necesario especificar uno o varios parámetros. La tabla que presentamos a continuación presenta las distribuciones más comunes, los parámetros requeridos y sus valores por defecto. `n` representa siempre el tamaño de la muestra.

Distribución	Función en R
Binomial	<code>rbinom(n, size, prob)</code>
Poisson	<code>rpois(n, lambda)</code>
Geométrica	<code>rgeom(n, prob)</code>
Hipergeométrica	<code>rhyper(nn, m, n, k)</code>
Binomial Negativa	<code>rnbinom(n, size, prob)</code>
Multinomial	<code>rmultinom(n, size, prob)</code>
Uniforme	<code>runif(n, min=0, max=1)</code>
Exponencial	<code>rexp(n, rate=1)</code>
Gaussiana	<code>rnorm(n, mean=0, sd=1)</code>
Gamma	<code>rgamma(n, shape, scale=1)</code>
Weibull	<code>rweibull(n, shape, scale=1)</code>
Cauchy	<code>rcauchy(n, location=0, scale=1)</code>
Beta	<code>rbeta(n, shape1, shape2)</code>
t	<code>rt(n, df)</code>
Fisher	<code>rf(n, df1, df2)</code>
χ^2	<code>rchisq(n, df)</code>
Logística	<code>rlogis(n, location=0, scale=1)</code>
Lognormal	<code>rlnorm(n, meanlog=0, sdlog=1)</code>

Además, R tiene la función `sample` que permite obtener muestras con o sin reposición de conjuntos finitos de valores. La sintaxis es

```
sample(x, size, replace = FALSE, prob = NULL)
```

donde

- `x` es el conjunto a partir del cual queremos obtener la muestra, escrito como un vector,
- `size` es el tamaño de la muestra,
- `replace` permite indicar si se permiten repeticiones (`replace = TRUE`) o no y finalmente
- `prob` es un vector de probabilidades si se desea hacer un muestreo pesado y no uniforme.

1.13. Convergencia de Variables Aleatorias

Hay varios modos de convergencia en la Teoría de Probabilidades. Vamos a considerar algunos de ellos a continuación. Sea $X_n, n \geq 1$ una sucesión de variables aleatorias definidas en un espacio de probabilidad común (Ω, \mathcal{F}, P) y sea X otra variable definida sobre este mismo espacio.

Definición 1.4 La sucesión X_n converge *puntualmente* a X si para todo $\omega \in \Omega$ se cumple que

$$\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega).$$

Notación: $X_n \rightarrow X$.

Definición 1.5 La sucesión X_n converge *casi seguramente* o *con probabilidad 1* a X si existe un conjunto nulo $N \in \mathcal{F}$ tal que para todo $\omega \notin N$ se cumple que

$$\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega).$$

Notación: $X_n \rightarrow X$ *c.s.* o *c.p.1*, o también $X_n \xrightarrow{c.s.} X$.

Definición 1.6 La sucesión X_n converge *en probabilidad* a X si dado cualquier $\varepsilon > 0$ se tiene que

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0.$$

Notación: $X_n \xrightarrow{P} X$.

Definición 1.7 La sucesión X_n converge *en L^p* , $1 \leq p < \infty$, a X si $E[|X_n|^p] < \infty$ y

$$\lim_{n \rightarrow \infty} E[|X_n - X|^p] = 0.$$

Notación: $X_n \xrightarrow{L^p} X$ o también $X_n \rightarrow X$ en L^p .

Definición 1.8 La sucesión X_n converge *en distribución* a X

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x), \quad \text{para todo } x \in \mathcal{C}(F_X),$$

donde $\mathcal{C}(F_X)$ es el conjunto de puntos de continuidad de F_X .

Notación: $X_n \xrightarrow{D} X$. También usaremos la notación $X_n \xrightarrow{D} F_X$

Observación 1.4

1. Cuando consideramos la convergencia c.s. consideramos para cada $\omega \in \Omega$, si la sucesión de números reales $X_n(\omega)$ converge al número real $X(\omega)$. Si esto ocurre fuera de un conjunto de ω de medida 0, decimos que hay convergencia c.s.
2. La convergencia en L^2 se conoce usualmente como convergencia en media cuadrática.
3. En la definición de convergencia en distribución las variables sólo aparecen a través de sus funciones de distribución. Por lo tanto las variables no tienen que estar definidas en un mismo espacio de probabilidad.
4. Es posible demostrar que una función de distribución tiene a lo sumo una cantidad numerable de discontinuidades. Como consecuencia $\mathcal{C}(F_X)$ es la recta real, excepto, posiblemente, por un conjunto numerable de puntos.
5. Es posible demostrar que en cualquiera de estos modos de convergencia el límite es (esencialmente) único.

▲

Ejemplo 1.25

Sea $X_n \sim \Gamma(n, n)$. Veamos que $X_n \xrightarrow{P} 1$ cuando $n \rightarrow \infty$.

Observamos que $E[X_n] = 1$ mientras que $\text{Var}[X] = 1/n$. Usando la desigualdad de Chebyshev obtenemos que para todo $\varepsilon > 0$,

$$P(|X_n - X| > \varepsilon) \leq \frac{1}{n\varepsilon^2} \rightarrow 0 \quad \text{cuando } n \rightarrow \infty.$$

▲

Ejemplo 1.26

Sean X_1, X_2, \dots v.a.i. con densidad común

$$f(x) = \begin{cases} \alpha x^{-\alpha-1}, & \text{para } x > 1, \alpha > 0, \\ 0, & \text{en otro caso.} \end{cases}$$

y sea $Y_n = n^{-1/\alpha} \max_{1 \leq k \leq n} X_k$, $n \geq 1$. Demuestre que Y_n converge en distribución y determine la distribución límite.

Para resolver este problema vamos a calcular la f.d. común:

$$F(x) = \int_1^x \alpha x^{-\alpha-1} dy = 1 - x^{-\alpha}$$

siempre que $x > 1$ y vale 0 si no. Por lo tanto, para cualquier $x > 0$,

$$\begin{aligned} F_{Y_n}(x) &= P(\max_{1 \leq k \leq n} X_k \leq xn^{1/\alpha}) = (F(xn^{1/\alpha}))^n \\ &= \left(1 - \frac{1}{nx^\alpha}\right)^n \rightarrow e^{-x^{-\alpha}} \quad \text{cuando } n \rightarrow \infty. \end{aligned}$$

▲

Ejemplo 1.27 (La Ley de los Grandes Números)

Esta es una versión débil de la LGN. Sean X_1, X_2, \dots v.a.i.i.d. con media μ y varianza finita σ^2 y pongamos $S_n = X_1 + \dots + X_n$, $n \geq 1$. La Ley (Débil) de los Grandes Números dice que para todo $\varepsilon > 0$,

$$P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) \rightarrow 0 \quad \text{cuando } n \rightarrow \infty,$$

es decir

$$\frac{S_n}{n} \xrightarrow{P} \mu \quad \text{cuando } n \rightarrow \infty.$$

La prueba de esta proposición sigue de la desigualdad de Chebyshev:

$$P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0 \quad \text{cuando } n \rightarrow \infty.$$

▲

Ejemplo 1.28 (Aproximación de Poisson)

Sea $X_n \sim \text{Bin}(n, \frac{\lambda}{n})$, entonces

$$X_n \xrightarrow{D} \text{Pois}(\lambda)$$

Vemos esto

$$\begin{aligned} P(X_n = k) &= \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{n(n-1)\cdots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{n(n-1)\cdots(n-k+1)}{n^k} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right)}{\left(1 - \frac{\lambda}{n}\right)^k} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \end{aligned}$$

Si ahora hacemos $n \rightarrow \infty$ la primera fracción tiende a 1 porque k y λ están fijos, mientras que

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$$

Por lo tanto

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

▲

1.13.1. Relación entre los Distintos Tipos de Convergencia

En esta sección nos planteamos investigar la relación entre los distintos tipos de convergencia que hemos definido, y en particular exploramos la posibilidad de poder ordenar estos conceptos.

I. Convergencia c.s. implica Convergencia en Probabilidad.

Es posible demostrar que $X_n \xrightarrow{c.s.} X$ cuando $n \rightarrow \infty$ sí y sólo sí para todo $\varepsilon > 0$ y δ , $0 < \delta < 1$, existe n_0 tal que, para todo $n > n_0$

$$P\left(\bigcap_{m>n} \{|X_m - X| < \varepsilon\}\right) > 1 - \delta \quad (1.41)$$

o equivalentemente

$$P\left(\bigcup_{m>n} \{|X_m - X| > \varepsilon\}\right) < \delta.$$

Como, para $m > n$,

$$\{|X_m - X| > \varepsilon\} \subset \bigcup_{m>n} \{|X_m - X| > \varepsilon\},$$

la sucesión también converge en probabilidad. El siguiente ejemplo muestra que el recíproco es falso.

Ejemplo 1.29

Sean X_1, X_2, \dots v.a.i. tales que

$$P(X_n = 1) = 1 - \frac{1}{n} \quad \text{y} \quad P(X_n = n) = \frac{1}{n}, \quad n \geq 1.$$

Claramente,

$$P(|X_n - 1| > \varepsilon) = P(X_n = n) = \frac{1}{n} \rightarrow 0, \quad \text{cuando } n \rightarrow \infty,$$

para todo $\varepsilon > 0$, es decir,

$$X_n \xrightarrow{P} 1 \quad \text{cuando } n \rightarrow \infty.$$

Veamos ahora que X_n no converge c.s. a 1 cuando $n \rightarrow \infty$. Para todo $\varepsilon > 0$, $\delta \in (0, 1)$ y $N > n$ tenemos

$$\begin{aligned} P\left(\bigcap_{m>n} \{|X_m - X| < \varepsilon\}\right) &= P\left(\lim_N \bigcap_{m=n+1}^N \{|X_m - X| < \varepsilon\}\right) \\ &= \lim_N P\left(\bigcap_{m=n+1}^N \{|X_m - X| < \varepsilon\}\right) \\ &= \lim_N \prod_{m=n+1}^N P(|X_m - 1| < \varepsilon) \\ &= \lim_N \prod_{m=n+1}^N P(X_m = 1) = \lim_N \prod_{m=n+1}^N \left(1 - \frac{1}{m}\right) \\ &= \lim_N \prod_{m=n+1}^N \frac{m-1}{m} = \lim_N \frac{n}{N} = 0, \end{aligned}$$

para cualquier n . Esto muestra que no existe n_0 para el cual (1.41) valga, y por lo tanto X_n no converge c.s. a 1 cuando $n \rightarrow \infty$.

▲

II. Convergencia en L^p implica Convergencia en Probabilidad

Usando la desigualdad de Markov, para $\varepsilon > 0$ fijo

$$P(|X_n - X| > \varepsilon) \leq \frac{1}{\varepsilon^p} E[|X_n - X|^p] \rightarrow 0$$

cuando $n \rightarrow \infty$, lo que muestra la conclusión.

En este caso el recíproco tampoco es cierto. Para empezar, $E[|X_n - X|]$ no tiene por qué existir, pero aun si existe puede ocurrir que haya convergencia en probabilidad sin que haya convergencia en L^p .

Ejemplo 1.30

Sea $\alpha > 0$ y sea X_1, X_2, \dots v.a. tales que

$$P(X_n = 1) = 1 - \frac{1}{n^\alpha} \quad \text{y} \quad P(X_n = n) = \frac{1}{n^\alpha}, \quad n \geq 1.$$

Como

$$P(|X_n - 1| > \varepsilon) = P(X_n = n) = \frac{1}{n^\alpha} \rightarrow 0, \quad \text{cuando } n \rightarrow \infty,$$

para todo $\varepsilon > 0$, tenemos que

$$X_n \xrightarrow{P} 1 \quad \text{cuando } n \rightarrow \infty.$$

Por otro lado

$$E[|X_n - 1|^p] = 0^p \cdot \left(1 - \frac{1}{n^\alpha}\right) + |n - 1|^p \frac{1}{n^\alpha} = \frac{(n-1)^p}{n^\alpha},$$

de donde obtenemos que

$$E[|X_n - 1|^p] \rightarrow \begin{cases} 0, & \text{para } p < \alpha, \\ 1, & \text{para } p = \alpha, \\ +\infty, & \text{para } p > \alpha, \end{cases} \quad (1.42)$$

Esto muestra que $X_n \xrightarrow{P} 1$ cuando $n \rightarrow \infty$ si $p < \alpha$ pero X_n no converge en L^p si $p \geq \alpha$. Por lo tanto, convergencia en L^p es más fuerte que convergencia en probabilidad. \blacktriangle

Observación 1.5 Si $\alpha = 1$ y las variables son independientes, cuando $n \rightarrow \infty$

$$\begin{aligned} X_n &\xrightarrow{P} 1, \\ X_n &\xrightarrow{c.s.}, \\ E[X_n] &\rightarrow 2, \\ X_n &\xrightarrow{L^p} 1 \quad \text{para } 0 < p < 1, \\ X_n &\xrightarrow{L^p} \quad \text{para } p \geq 1. \end{aligned}$$

\blacktriangle

Observación 1.6 Si $\alpha = 2$ y las variables son independientes, cuando $n \rightarrow \infty$

$$\begin{aligned} X_n &\xrightarrow{P} 1, \\ X_n &\xrightarrow{c.s.} 1, \\ E[X_n] &\rightarrow 1, \quad \text{y} \quad \text{Var}[X_n] \rightarrow 1 \\ X_n &\xrightarrow{L^p} 1 \quad \text{para } 0 < p < 2, \\ X_n &\xrightarrow{L^p} \quad \text{para } p \geq 2. \end{aligned}$$

\blacktriangle

III. Convergencia en L^p y Convergencia c.s. son Independientes

Ninguna de las dos implica la otra, y esto lo podemos ver de las observaciones anteriores. En el primer caso, para $0 < p < 1$ hay convergencia en L^p mientras que no hay convergencia c.s. En el segundo hay convergencia c.s. pero no hay convergencia en L^p para $p \geq 2$.

IV. Convergencia en Probabilidad implica Convergencia en Distribución

Sea $\varepsilon > 0$, entonces

$$\begin{aligned} F_{X_n}(x) &= P(X_n \leq x) \\ &= P(\{X_n \leq x\} \cap \{|X_n - X| \leq \varepsilon\}) + P(\{X_n \leq x\} \cap \{|X_n - X| > \varepsilon\}) \\ &\leq P(\{X \leq x + \varepsilon\} \cap \{|X_n - X| \leq \varepsilon\}) + P(|X_n - X| > \varepsilon) \\ &\leq P(X \leq x + \varepsilon) + P(|X_n - X| > \varepsilon) \end{aligned}$$

es decir,

$$F_{X_n}(x) \leq F_X(x + \varepsilon) + P(|X_n - X| > \varepsilon). \quad (1.43)$$

De manera similar se demuestra que

$$F_X(x - \varepsilon) \leq F_{X_n}(x) + P(|X_n - X| > \varepsilon). \quad (1.44)$$

Como $X_n \xrightarrow{P} X$ cuando $n \rightarrow \infty$ obtenemos, haciendo $n \rightarrow \infty$ en (1.43) y (1.44),

$$F_X(x - \varepsilon) \leq \liminf_{n \rightarrow \infty} F_{X_n}(x) \leq \limsup_{n \rightarrow \infty} F_{X_n}(x) \leq F_X(x + \varepsilon)$$

Esta relación es válida para todo x y todo $\varepsilon > 0$. Para demostrar la convergencia en distribución suponemos que $x \in \mathcal{C}(F_X)$ y hacemos $\varepsilon \rightarrow 0$. Obtenemos

$$F_X(x) \leq \liminf_{n \rightarrow \infty} F_{X_n}(x) \leq \limsup_{n \rightarrow \infty} F_{X_n}(x) \leq F_X(x),$$

por lo tanto

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x),$$

y como esto vale para cualquier $x \in \mathcal{C}(F_X)$ obtenemos la convergencia en distribución. \blacktriangle

Observación 1.7 Observamos que si F_X tiene un salto en x , sólo podemos concluir que

$$F_X(x-) \leq \liminf_{n \rightarrow \infty} F_{X_n}(x) \leq \limsup_{n \rightarrow \infty} F_{X_n}(x) \leq F_X(x),$$

y $F_X(x) - F_X(x-)$ es el tamaño del salto. Esto explica por qué sólo se toman en cuenta los puntos de continuidad en la definición de convergencia en distribución.

Como mencionamos anteriormente, la convergencia en distribución no requiere que las variables estén definidas en un mismo espacio de probabilidad, y por lo tanto es más débil que los otros modos de convergencia. El siguiente ejemplo muestra que aun cuando las distribuciones conjuntas existan, existen variables que convergen sólo en distribución.

Ejemplo 1.31

Sea X una variable con distribución simétrica, continua y no-degenerada y definimos X_1, X_2, \dots por $X_{2n} = X$ y $X_{2n-1} = -X$, $n = 1, 2, \dots$. Como $X_n \xrightarrow{D} X$ para todo n , tenemos, en particular, $X_n \xrightarrow{D} X$ cuando $n \rightarrow \infty$. Por otro lado, como X tiene distribución no-degenerada existe $a > 0$ tal que $P(|X| > a) > 0$ (¿por qué?). En consecuencia, para todo $\varepsilon > 0$, $0 < \varepsilon < 2a$,

$$P(|X_n - X| > \varepsilon) = \begin{cases} 0, & \text{para } n \text{ par,} \\ P(|X| > \frac{\varepsilon}{2}) > 0, & \text{para } n \text{ impar.} \end{cases}$$

Esto muestra que X_n no puede converger en probabilidad a X cuando $n \rightarrow \infty$, y en consecuencia tampoco c.s. o en L^p . ▲

Podemos resumir todo lo anterior en el siguiente teorema.

Teorema 1.8 Sean X y X_1, X_2, \dots variables aleatorias, entonces, cuando $n \rightarrow \infty$,

$$X_n \xrightarrow{c.s.} X \implies X_n \xrightarrow{P} X \implies X_n \xrightarrow{D} X$$
$$\uparrow$$
$$X_n \xrightarrow{L^p} X$$

Ninguna de las implicaciones se puede invertir.