

Capítulo 1

Introducción

1.1. Introducción

Falta

1.2. Probabilidad y Esperanza Condicional

1.2.1. El Caso Discreto

Definición 1.1 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 1.1

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 1.2

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,

$$p_{X|Y}(4|5) = \frac{P(X = 4, Y = 5)}{P(Y = 5)} = \frac{p_1 p_4}{p_1 p_4 + p_2 p_3},$$

$$p_{X|Y}(6|5) = \frac{P(X = 6, Y = 5)}{P(Y = 5)} = \frac{p_2 p_3}{p_1 p_4 + p_2 p_3}.$$

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

$$p_{X|Y}(3|4) = \frac{2p_1 p_3}{2p_1 p_3 + p_2^2}, \quad p_{X|Y}(4|4) = \frac{p_2^2}{2p_1 p_3 + p_2^2},$$

para $Y = 6$:

$$p_{X|Y}(8|6) = \frac{2p_2 p_4}{2p_2 p_4 + p_3^2}, \quad p_{X|Y}(9|6) = \frac{p_3^2}{2p_2 p_4 + p_3^2}.$$

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,

$$\begin{aligned} E[X|Y = 2] &= 1, & E[X|Y = 3] &= 2 \\ E[X|Y = 4] &= 3 \frac{2p_1 p_3}{2p_1 p_3 + p_2^2} + 4 \frac{p_2^2}{2p_1 p_3 + p_2^2} \\ E[X|Y = 5] &= 4 \frac{p_1 p_4}{p_1 p_4 + p_2 p_3} + 6 \frac{p_2 p_3}{p_1 p_4 + p_2 p_3} \\ E[X|Y = 6] &= 8 \frac{2p_2 p_4}{2p_2 p_4 + p_3^2} + 9 \frac{p_3^2}{2p_2 p_4 + p_3^2} \\ E[X|Y = 7] &= 12, & E[X|Y = 8] &= 16. \end{aligned}$$

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 $E[|\nu(X, Y)|] < \infty$.

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

Como consecuencia de 1, 5 y 6 obtenemos

7. $E[c | Y = y] = c$,
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 1.3

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. ▲

1.2.2. 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 1.4

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

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_X^2} - 2\rho\frac{xy}{\sigma_X\sigma_Y} + \frac{y^2}{\sigma_Y^2}\right)\right\} \quad (1.1)$$

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

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left\{-\frac{x^2}{2\sigma_X^2}\right\}. \quad (1.2)$$

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 1.5

Si (X, Y) es un vector gaussiano bidimensional cuya densidad está dada por (1.1), entonces

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

es la esperanza condicional de Y dado que $X = x$. A partir de (1.2) vemos que la densidad condicional es gaussiana de media $\rho x\sigma_Y/\sigma_X$, de donde

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

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

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

Proposición 1.1 Sea X, Y variables aleatorias tales que el vector (X, Y) tiene densidad $f_{X,Y}$. Si existen funciones $\alpha : \mathbb{R} \rightarrow \mathbb{R}^+$, $\beta : \mathbb{R}^2 \rightarrow \mathbb{R}^+$ tales que, para todo $x, y \in \mathbb{R}$,

$$f_{X,Y}(x, y) = \alpha(y)\beta(x, y) \quad y \quad \int_{\mathbb{R}} \beta(x, y) dx = 1.$$

Entonces $\alpha(y) = f_Y(y)$ y $\beta(x, y) = f_{X|Y}(x|y)$.

Demostración. Observamos que

$$f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x, y) dx = \int_{\mathbb{R}} \alpha(y)\beta(x, y) dx = \alpha(y) \int_{\mathbb{R}} \beta(x, y) dx = \alpha(y)$$

y, teniendo en cuenta que

$$f_{XY}(x, y) = f_Y(y)f_{X|Y}(x|y) = \alpha(y)\beta(x, y)$$

obtenemos que $\beta(x, y) = f_{X|Y}(x|y)$. ■

1.2.3. 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 1.6

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 1.7

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

Usando la ley de probabilidad total

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

1.2.4. 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 1.8

- a) Colas: N representa el número de clientes, X_i es el tiempo de atención de cada cliente, T es el tiempo total de atención.
- b) 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.
- c) Población: N representa el número de plantas, X_i es el número de semillas de cada planta, T es el total de semillas.
- d) 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

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

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

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

Tenemos

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

Para determinar la varianza comenzamos por

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

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

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

Para el segundo tenemos

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

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 $X_1 + \dots + X_n$ es la n -ésima convolución de la densidad $f(x)$, que denotaremos por $f^{(n)}(x)$ y definimos 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 $T = X_1 + \dots + X_N$ 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 1.9 (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

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

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 $T \sim \text{Exp}(\lambda\beta)$. ▲

1.3. 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

$$\phi(s) = E[s^\xi] = \sum_{k=0}^{\infty} s^k p_k, \quad 0 \leq s \leq 1. \quad (1.3)$$

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

$$p_k = \frac{1}{k!} \left. \frac{d^k \phi(s)}{ds^k} \right|_{s=0}. \quad (1.4)$$

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

$$\phi_X(s) = \phi_1(s)\phi_2(s)\cdots\phi_n(s). \quad (1.5)$$

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

$$\frac{d\phi(s)}{ds} = p_1 + 2p_2s + 3p_3s^2 + \dots,$$

por lo tanto

$$\left. \frac{d\phi(s)}{ds} \right|_{s=1} = p_1 + 2p_2 + 3p_3 + \dots = E[\xi]. \quad (1.6)$$

Para la segunda derivada tenemos

$$\frac{d^2\phi(s)}{ds^2} = 2p_2 + 3 \cdot 2p_3s + 4 \cdot 3p_4s^2 + \dots,$$

evaluando en $s = 1$,

$$\begin{aligned} \left. \frac{d^2\phi(s)}{ds^2} \right|_{s=1} &= 2p_2 + 3 \cdot 2p_3 + 4 \cdot 3p_4 \cdots \\ &= \sum_{k=2}^{\infty} k(k-1)p_k \\ &= E[\xi(\xi-1)] = E[\xi^2] - E[\xi] \end{aligned} \quad (1.7)$$

de modo que

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

y en consecuencia

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

Ejemplo 1.10

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

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

Su función generadora de probabilidad es

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

Entonces,

$$\frac{d\phi(s)}{ds} = \lambda e^{-\lambda(1-s)}, \quad \left. \frac{d\phi(s)}{ds} \right|_{s=1} = \lambda \quad (1.8)$$

$$\frac{d^2\phi(s)}{ds^2} = \lambda^2 e^{-\lambda(1-s)}, \quad \left. \frac{d^2\phi(s)}{ds^2} \right|_{s=1} = \lambda^2 \quad (1.9)$$

y obtenemos

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

▲

1.3.1. 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

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

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

$$\phi_{\xi+\eta}(s) = E[s^{\xi+\eta}] = E[s^\xi s^\eta] = E[s^\xi] E[s^\eta] = \phi_\xi(s) \phi_\eta(s) \quad (1.10)$$

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. $\phi(s) = E[s^\xi]$ entonces

$$E[s^{\xi_1 + \dots + \xi_m}] = \phi^m(s) \quad (1.11)$$

¿Qué ocurre si el número de sumandos es aleatorio?

Proposición 1.2 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

$$h_X(s) = g_N(\phi(s)). \quad (1.12)$$

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 1.11

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.

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

según vimos en el ejemplo 1.10. Por la proposición anterior obtenemos que la f.g.p. de X es

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

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

1.4. Funciones Generadoras de Momentos.

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

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

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

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

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

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

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

Ejemplos 1.12

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

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 1.1 *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 1.2 (de Continuidad) *Sea $F_n(x)$, $n \geq 1$ una sucesión de f.d. con funciones generadoras 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 1.3 (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 .

$$\begin{aligned} E(e^{tT_n}) &= E \left[\exp \left(\frac{t(S_n - np)}{(npq)^{1/2}} \right) \right] = E \left[\exp \left(\frac{t(\sum_{i=1}^n (X_i - p))}{(npq)^{1/2}} \right) \right] \\ &= E \left[\prod_{i=1}^n \exp \left(\frac{t(X_i - p)}{(npq)^{1/2}} \right) \right] = \prod_{i=1}^n E \left[\exp \left(\frac{t(X_i - p)}{(npq)^{1/2}} \right) \right] \\ &= \left(E \left[\exp \left(\frac{t(X_1 - p)}{(npq)^{1/2}} \right) \right] \right)^n \\ &= \left(p \exp \left(\frac{t(1-p)}{(npq)^{1/2}} \right) + q \exp \left(\frac{-pt}{(npq)^{1/2}} \right) \right)^n. \end{aligned} \quad (1.13)$$

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

$$p \exp \left(\frac{t(1-p)}{(npq)^{1/2}} \right) = p \left(1 + \frac{qt}{(npq)^{1/2}} + \frac{q^2 t^2}{2npq} + \frac{C_1 q^3 t^3}{3!(npq)^{3/2}} \right) \quad (1.14)$$

$$q \exp \left(\frac{-pt}{(npq)^{1/2}} \right) = q \left(1 - \frac{pt}{(npq)^{1/2}} + \frac{p^2 t^2}{2npq} + \frac{C_2 p^3 t^3}{3!(npq)^{3/2}} \right). \quad (1.15)$$

La suma de estas dos expresiones nos da $1 + \frac{t^2}{2n} + O(n^{-3/2})$ y sustituyendo en (1.13) obtenemos

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

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

1.5. 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 .

1.5.1. Método de la Distribución Inversa

Este método se basa en el siguiente resultado:

Proposición 1.3 *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*

$$F_Y(y) = F_X(g^{-1}(y)). \quad (1.16)$$

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 1.1 *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

$$F_Z(z) = F_U(F(z)) = F(z) \quad (1.17)$$

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 1.1 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 1.13

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 1.1. 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 $X \sim \mathcal{Exp}(\lambda)$ 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 $-\ln(U)/\lambda$.

1.5.2. 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

$$\frac{p_j}{q_j} \leq c \quad \text{para todo } j \text{ tal que } p_j > 0, \quad (1.18)$$

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:

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

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 (1.18).

Ejemplo 1.14

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 1.15

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.

1.5.3. 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 1.3).

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

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

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 $p/(1-p)$ 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:

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

1.5.4. 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.

1.6. 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 1.2 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 1.3 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 1.4 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 1.5 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 1.6 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 1.16

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 1.17

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 1.18 (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 1.19 (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

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

▲

1.6.1. 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 $n > n_0$

$$P\left(\bigcap_{m>n} \{|X_m - X| < \varepsilon\}\right) > 1 - \delta \quad (1.19)$$

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 1.20

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 (1.19) 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 1.21

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

$$E[|X_n - 1|^p] \rightarrow \begin{cases} 0, & \text{para } p < \alpha, \\ 1, & \text{para } p = \alpha, \\ +\infty, & \text{para } p > \alpha, \end{cases} \quad (1.20)$$

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 $n \rightarrow \infty$

$$\begin{aligned} X_n &\xrightarrow{P} 1, \\ X_n &\not\xrightarrow{c.s.}, \\ E[X_n] &\rightarrow 2, \\ X_n &\xrightarrow{L^p} 1 \quad \text{para } 0 < p < 1, \\ X_n &\not\xrightarrow{L^p} \quad \text{para } p \geq 1. \end{aligned}$$

\blacktriangle

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

$$\begin{aligned} X_n &\xrightarrow{P} 1, \\ X_n &\xrightarrow{c.s.} 1, \\ E[X_n] &\rightarrow 1, \quad \text{y} \quad \text{Var}[X_n] \rightarrow 1 \\ X_n &\xrightarrow{L^p} 1 \quad \text{para } 0 < p < 2, \\ X_n &\not\xrightarrow{L^p} \quad \text{para } p \geq 2. \end{aligned}$$

\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,

$$F_{X_n}(x) \leq F_X(x + \varepsilon) + P(|X_n - X| > \varepsilon). \quad (1.21)$$

De manera similar se demuestra que

$$F_X(x - \varepsilon) \leq F_{X_n}(x) + P(|X_n - X| > \varepsilon). \quad (1.22)$$

Como $X_n \xrightarrow{P} X$ cuando $n \rightarrow \infty$ obtenemos, haciendo $n \rightarrow \infty$ en (1.21) y (1.22),

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

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 1.22

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

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

Ninguna de las implicaciones se puede invertir.