

Notas de Probabilidades y Estadística

Capítulos 7 al 12

Víctor J. Yohai

Basadas en apuntes de clase tomados por Alberto Déboli

Diciembre 2003

Capítulo 7

Esperanza Matemática

7.1. Integral de Riemann-Stieljes.

Sea $f : [a, b] \rightarrow \mathbb{R}$ y consideremos una partición del intervalo $[a, b]$ que llamaremos $\pi = \{x_0, x_1, \dots, x_n\}$ tal que $a = x_0 < x_1 < \dots < x_n = b$. Sea un selección de puntos de cada intervalo $\xi = \{\xi_i\}_{1 \leq i \leq n}$, donde $\xi_i \in (x_{i-1}, x_i)$ para $i = 1, 2, \dots, n$.

Definimos la suma de Riemann

$$S_a^b(\pi, \xi, f) = \sum_{i=1}^n f(\xi_i)(x_i - x_{i-1}).$$

Se llama **norma de la partición**

$$\|\pi\| = \max_{1 \leq i \leq n} \{x_i - x_{i-1} : 1 \leq i \leq n\}.$$

Se dice que f es integrable **Riemann** sobre $[a, b]$ con valor $I = \int_a^b f = \int_a^b f(x) dx$ sii para todo $\varepsilon > 0$ existe $\sigma > 0$ tal que si $\|\pi\| < \delta$ entonces

$$|S_a^b(\pi, \xi, f) - I| < \varepsilon.$$

Análogamente se define la integral de **Riemann-Stieljes**. Dadas g, F funciones definidas sobre $[a, b]$ se define la suma de Riemann-Stieljes asociada a la partición π y la selección de puntos ξ

$$S_a^b(\pi, \xi, g, F) = \sum_{i=1}^n f(\xi_i)(F(x_i) - F(x_{i-1})).$$

Se dice que f es integrable **Riemann-Stieljes** sobre $[a, b]$ con valor $I = \int_a^b g dF = \int_a^b g(x) dF(x)$ sii para todo $\varepsilon > 0$ existe $\sigma > 0$ tal que si $\|\pi\| < \delta$ entonces

$$|S_a^b(\pi, \xi, g, F) - I| < \varepsilon.$$

Observación.

1. Si $F(x) = x$, entonces la integral de Riemann-Stieljes es la integral de Riemann.

2. Una condición suficiente aunque no necesaria para que exista la integral de Riemann-Stieljes, es que g se continua y F monótona no decreciente. Si tomamos como F una función de distribución el último requisito se cumplirá.

A continuación enunciamos una serie de propiedades de la integral de Riemann-Stieljes

Algunas propiedades.

P1. Linealidad de la Integral de Riemann-Stieljes respecto de g .

Si $\int_a^b g_1 dF$ y $\int_a^b g_2 dF$ existen y $\alpha_1, \alpha_2 \in \mathbb{R}$ entonces $\int_a^b (\alpha_1 g_1 + \alpha_2 g_2) dF$ existe y además

$$\int_a^b (\alpha_1 g_1 + \alpha_2 g_2) dF = \alpha_1 \int_a^b g_1 dF + \alpha_2 \int_a^b g_2 dF.$$

P2. Linealidad de la Integral de Riemann-Stieljes respecto de F .

Si $\int_a^b g dF_1$ y $\int_a^b g dF_2$ existen y $\alpha_1, \alpha_2 \in \mathbb{R}$ entonces $\int_a^b g d(\alpha_1 F_1 + \alpha_2 F_2)$ existe y además

$$\int_a^b g d(\alpha_1 F_1 + \alpha_2 F_2) = \alpha_1 \int_a^b g dF_1 + \alpha_2 \int_a^b g dF_2.$$

P3. Aditividad respecto del dominio de integración.

Sea $a < b < c$ y supongamos que $\int_a^b g dF$; $\int_b^c g dF$ y $\int_a^c g dF$ existen. Entonces

$$\int_a^c g dF = \int_a^b g dF + \int_b^c g dF.$$

Observación.

La vuelta no vale en general. Una condición suficiente para la vuelta es que F sea no decreciente.

P4. Si F es no decreciente y $g_1 \leq g_2$ sobre $[a, b]$ entonces

$$\int_a^b g_1 dF \leq \int_a^b g_2 dF.$$

En particular teniendo en cuenta que $-|g| \leq g \leq |g|$ se obtiene

$$\left| \int_a^b g dF \right| \leq \int_a^b |g| dF$$

Estamos interesados en extender el dominio de ontegración a toda la recta o a semirectas. Esto lleva a la siguiente definición

Definición. Supongamos que $\int_a^b g dF$ existe para todo $a, b \in \mathbb{R}$. Decimos

que la integral impropia $\int_{-\infty}^{+\infty} g dF = I$ existe sii I es un número real y

$$\lim_{a \rightarrow -\infty; b \rightarrow +\infty} \int_a^b g dF = I \quad (7.1)$$

De manera análoga se define $\int_a^{+\infty} g dF$ o $\int_{-\infty}^b g dF$. Tendremos el siguiente teorema

Teorema. Sea $g \geq 0$ y F no decreciente. Entonces pueden ocurrir dos cosas

(i)

$$M = \sup_{a, b \in \mathbb{R}} \int_a^b g dF < \infty$$

En este caso el límite (7.1) existe y es finito

(ii)

$$M = \sup_{a, b \in \mathbb{R}} \int_a^b g dF = \infty$$

En este caso el límite (7.1) existe y es ∞ . Luego podemos definir $\int_{-\infty}^{+\infty} g dF = \infty$

Sea ahora g de signo arbitrario y F no decreciente. El siguiente teorema vale.

Una condición necesaria y suficiente para que $\int_{-\infty}^{+\infty} g dF$ exista es que

$$\widetilde{M} = \sup_{a, b \in \mathbb{R}} \int_a^b |g| dF < \infty$$

7.2. Definición de Esperanza Matemática.

7.2.1. Heurística

Sea X una variable aleatoria discreta y para fijar ideas supongamos que toma un número finito de valores, x_1, x_2, \dots, x_k , con probabilidades $p_X(x_1); p_X(x_2), \dots, p_X(x_k)$.

Supongamos que se repite un experimento asociado a la variable aleatoria X , n veces en forma independiente y que el resultado x_i se obtiene n_i veces, $1 \leq i \leq k$. Entonces el promedio de todos los valores es

$$\begin{aligned} \bar{x}_n &= \frac{n_1 x_1 + n_2 x_2 + \dots + n_k x_k}{n} \\ &= \frac{n_1}{n} x_1 + \frac{n_2}{n} x_2 + \dots + \frac{n_k}{n} x_k. \end{aligned}$$

Luego pasando al límite y dado que n_j/n se aproxima a $p_X(x_j)$ obtenemos

$$\begin{aligned}\lim_{n \rightarrow +\infty} \bar{x}_n &= \lim_{n \rightarrow +\infty} \left(\frac{n_1}{n} x_1 + \frac{n_2}{n} x_2 + \dots + \frac{n_k}{n} x_k \right) \\ &= x_1 \lim_{n \rightarrow +\infty} \frac{n_1}{n} + x_2 \lim_{n \rightarrow +\infty} \frac{n_2}{n} + \dots + x_k \lim_{n \rightarrow +\infty} \frac{n_k}{n} \\ &= \sum_{j=1}^k x_j p_X(x_j).\end{aligned}$$

Esto motiva la definición de la esperanza matemática de una variable discreta..

7.2.2. Esperanza de una variable aleatoria discreta

Definición. Sea X una variable aleatoria con rango R_X y distribución de probabilidad p_X . Supongamos que

$$\sum_{x \in R_X} |x| p_X(x) < \infty.$$

En tal caso definimos la esperanza matemática de la variable X de la siguiente manera

$$E(X) = \sum_{x \in R_X} x p_X(x).$$

Observación.

1. Se sabe que la convergencia absoluta de la serie garantiza la convergencia de la serie.

2. Supongamos $\sum_{x \in R_X} |x| p_X(x) = \infty$

Denotemos con

$$\begin{aligned}R_X^+ &= \{x \in R_X : x > 0\} \\ R_X^- &= \{x \in R_X : x < 0\}.\end{aligned}$$

Entonces pueden ocurrir tres casos distintos.

(a) $\sum_{x \in R_X^+} x p_X(x) = \infty$ y $\sum_{x \in R_X^-} x p_X(x) = -\infty$.

(b) $\sum_{x \in R_X^+} x p_X(x) = +\infty$ y $\sum_{x \in R_X^-} x p_X(x) > -\infty$.

(c) $\sum_{x \in R_X^+} x p_X(x) < \infty$ y $\sum_{x \in R_X^-} x p_X(x) = -\infty$.

En el caso (a) no se puede definir la esperanza de X . En el caso (b) se puede definir $E(X) = \infty$ en el (c) $E(X) = -\infty$. Es decir para que la esperanza esté definida se requiere que $\sum_{x \in R_X^+} x p_X(x)$ o bien $\sum_{x \in R_X^-} x p_X(x)$ sea finita.

7.2.3. Definición general de esperanza matemática

Ahora queremos definir la esperanza matemática, de manera más general. Supongamos primero que X es una variable aleatoria concentrada en $[a, b]$. Es decir, supongamos que

$$P(a < X < b) = 1.$$

La idea que se utiliza para la definición de la esperanza de esta variable es la siguiente. Se define una sucesión de variables aleatorias discretas X_n que la aproximan y luego conociendo la esperanza de cada una de estas variables, la esperanza de X se define por un paso al límite.

Consideremos para cada n , una partición del intervalo $[a, b]$ en n intervalos de longitud $(b - a)/n$. Para esto consideramos la partición $\pi^n = \{x_0^n, x_1^n, \dots, x_n^n\}$ tal que $a = x_0^n < x_1^n < \dots < x_n^n = b$ y $x_i^n - x_{i-1}^n = \frac{b-a}{n}$.

Elegimos para cada i , $1 \leq i \leq n$, $\xi_i^n \in (x_{i-1}^n, x_i^n]$ y definimos la variable aleatoria

$$X_n = \xi_i^n \text{ si } x \in (x_{i-1}^n, x_i^n].$$

Esta variable toma únicamente un número finito de valores: ξ_i^n , $1 \leq i \leq n$. Además

$$p_{X_n}(\xi_i^n) = F_X(x_i^n) - F_X(x_{i-1}^n).$$

Luego la esperanza de la variable X_n viene dada por

$$\begin{aligned} E(X_n) &= \sum_{i=1}^n \xi_i^n p_{X_n}(\xi_i^n) = \\ &= \sum_{i=1}^n \xi_i^n (F_X(x_i^n) - F_X(x_{i-1}^n)) = S_a^b(\pi^n, \xi^n, x, F), \end{aligned}$$

y se obtiene

$$\lim_{n \rightarrow +\infty} E(X_n) = \lim_{n \rightarrow +\infty} S_a^b(\pi^n, \xi^n, x, F_X) = \int_a^b x dF_X.$$

Por lo tanto definimos la esperanza matemática de X por

$$E(X) = \int_a^b x dF_X.$$

Observación.

Siendo la función x continua y F monótona no decreciente, resulta que $\int_a^b x dF$ existe siempre y por lo tanto también $E(X)$ existe siempre.

Supongamos ahora que X es una variable aleatoria no acotada. El problema que ahora surge es que podría no existir $\int_{-\infty}^{+\infty} x dF$. Sin embargo sabemos

que $M = \int_{-\infty}^{+\infty} |x| dF$ siempre está bien definida, eventualmente con el valor $+\infty$.

Si $M < +\infty$ definimos la esperanza de la variable X similarmente al caso anterior por

$$E(X) = \int_{-\infty}^{+\infty} x dF.$$

Si $M = +\infty$ hay tres casos y el análisis es análogo al que realizamos anteriormente

- (a) $\int_0^{+\infty} x dF = +\infty$ y $\int_{-\infty}^0 x dF = -\infty$.
- (b) $\int_0^{+\infty} x dF = +\infty$ y $\int_{-\infty}^0 x dF > -\infty$.
- (c) $\int_0^{+\infty} x dF < +\infty$ y $\int_{-\infty}^0 x dF = -\infty$.

En el caso (a) la esperanza matemática de X no está definida. En el caso (b) se define $E(X) = +\infty$ y en el (c) $E(X) = -\infty$. Nuevamente la esperanza puede estar no definida y para su definición se requiere que al menos una de las dos integrales $\int_0^{+\infty} x dF$ o $\int_{-\infty}^0 x dF$ converga.

Con esta definición general de esperanza matemática, para el caso de una variable discreta se tienen dos definiciones. Ahora probaremos que efectivamente, la definición general es una extensión de la primera definición dada para el caso discreto, es decir para variables aleatorias discretas ambas definiciones coinciden.

Teorema. Supongamos que X es una variable aleatoria discreta y que $E(X)$ existe y es finita.

Entonces

$$\sum_{x \in R_X} x p_X(x) = \int_{-\infty}^{+\infty} x dF_X$$

Observación

Este resultado vale siempre, pero para fijar ideas vamos a probarlo para el caso en el que a los fines $R_X \cap [a, b]$ es finito para todo a y b . Este es el caso de la mayoría de las variables aleatorias discretas que aparecen en la práctica tomando únicamente valores enteros; por ejemplo, Poisson, Pascal, etc.

Demostración

Teniendo en cuenta que

$$\sum_{x \in R_X} x p_X(x) = \lim_{a \rightarrow -\infty; b \rightarrow +\infty} \sum_{x \in R_X \cap [a, b]} x p_X(x),$$

y que

$$\int_{-\infty}^{+\infty} x dF_X = \lim_{a \rightarrow -\infty; b \rightarrow +\infty} \int_a^b x dF_X,$$

bastará probar que para todo $a < b$

$$\sum_{x \in R_X \cap [a, b]} x p_X(x) = \int_a^b x dF_X. \quad (7.2)$$

Consideremos como antes $\pi^n = \{x_0^n, x_1^n, \dots, x_n^n\}$ una partición del intervalo $[a, b]$, en n intervalos iguales. Luego tenemos $a = x_0^n < x_1^n < \dots < x_n^n = b$ tales $x_i^n - x_{i-1}^n = (b-a)/n$.

Teniendo en cuenta que para todo $n \in \mathbb{N}$, $\|\pi^n\| = (b-a)/n$ es claro que

$$\lim_{n \rightarrow +\infty} \|\pi^n\| = 0.$$

Por la hipótesis supuesta $R_X \cap [a, b]$ es un conjunto finito, digamos

$$R_X \cap [a, b] = \{z_1, z_2, \dots, z_k\}.$$

Llamemos δ

$$\delta = \max_{2 \leq i \leq k} \{z_i - z_{i-1}\}.$$

Va a ser fundamental para esta demostración la elección de los puntos ξ_i^n en cada intervalo $(x_{i-1}^n, x_i^n]$. Procedemos de la siguiente manera.

Si $\|\pi^n\| < \delta$, en cada intervalo hay a lo sumo un elemento de $R_X \cap [a, b]$.

(i) Si

$$(R_X \cap [a, b]) \cap (x_{i-1}^n, x_i^n] \neq \emptyset$$

escogemos algún punto de la intersección como ξ_i^n . Como hemos visto, si $\|\pi^n\| < \delta$ a lo sumo existe uno solo.

(ii) Si

$$(R_X \cap [a, b]) \cap (x_{i-1}^n, x_i^n] = \emptyset$$

escogemos a cualquier punto de $(x_{i-1}^n, x_i^n]$ como ξ_i^n .

Sabemos que

$$\begin{aligned} \lim_{n \rightarrow +\infty} S_a^b(\pi^n, \xi_i^n, x, F) &= \lim_{n \rightarrow +\infty} \sum_{i=1}^n \xi_i^n (F_X(x_i^n) - F_X(x_{i-1}^n)) \\ &= \int_a^{+b} x dF \end{aligned}$$

Veamos cómo se vincula $S_a^b(\pi^n, \xi_i^n, x, F)$ con $E(X)$, definida como una serie. Consideremos dos subconjuntos de índices

$$A = \{i : (R_X \cap [a, b]) \cap (x_{i-1}^n, x_i^n] \neq \emptyset\}$$

y

$$A^c = \{i : (R_X \cap [a, b]) \cap (x_{i-1}^n, x_i^n] = \emptyset\}$$

Entonces podemos realizar la siguiente descomposición:

$$\begin{aligned} S_a^b(\pi^n, \xi_i^n, x, F) &= \sum_{i=1}^n \xi_i^n (F_X(x_i^n) - F_X(x_{i-1}^n)) = \\ &= \sum_{i \in A} \xi_i^n (F_X(x_i^n) - F_X(x_{i-1}^n)) + \sum_{i \in A^c} \xi_i^n (F_X(x_i^n) - F_X(x_{i-1}^n)) \end{aligned}$$

Observemos que $F_X(x_i^n) - F_X(x_{i-1}^n) = 0$ si $i \in A^c$ ya que el intervalo $(x_{i-1}, x_i]$ no contiene elementos de R_X . Luego

$$\sum_{i \in A^c} \xi_i^n (F_X(x_i^n) - F_X(x_{i-1}^n)) = 0$$

y se obtiene

$$S_a^b(\pi^n, \xi_i^n, x, F_X) = \sum_{i \in A} \xi_i^n (F_X(x_i^n) - F_X(x_{i-1}^n))$$

Sea n_0 tal que $(b-a)/n_0 < \delta$. Luego si $n > n_0$, $|\pi^n| < \delta$, y de acuerdo a lo anteriormente observado, en cada intervalo $(x_{i-1}^n, x_i^n]$ hay un solo z_j que coincide con el correspondiente ξ_i^n . En tal caso se tiene que $F_X(x_i^n) - F_X(x_{i-1}^n) = p_X(z_j)$. Por otro lado todo z_j será igual al ξ_i^n para algún i . Luego

$$S_a^b(\pi^n, \xi_i^n, x, F) = \sum_{j=1}^k z_j p_X(z_j) = \sum_{x \in R_X \cap [a, b]} x p_X(x)$$

y por lo tanto

$$\int_a^b x dF = \lim_{n \rightarrow \infty} S_a^b(\pi^n, \xi_i^n, x, F) = \sum_{x \in R_X \cap [a, b]} x p_X(x)$$

y esto prueba (7.2). Esto demuestra el Teorema \square

7.2.4. Esperanza matemática para una variable absolutamente continua.

El siguiente Teorema prueba que en el caso de que X sea una variable aleatoria absolutamente continua $E(X)$ se puede calcular por una integral de Riemann

Teorema. Supongamos que $\int_{-\infty}^{\infty} |x| f_X(x) dx < \infty$. Luego

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx.$$

Demostración. El teorema vale en general. Para facilitar la demostración lo probaremos para el caso que f_X es continua,

Bastará ver que para todo intervalo $[a, b]$, $a < b$ vale que

$$\int_a^b x f_X(x) dx = \int_a^b x dF_X, \quad (7.3)$$

ya que en tal caso el resultado se obtiene pasando al límite.

Consideremos para cada n una partición de puntos equidistantes del intervalo $[a, b]$

$$\pi^n = \{x_0^n, x_1^n, \dots, x_n^n\}$$

tal que $a = x_0^n < x_1^n < \dots < x_n^n = b$ satisfaciendo $x_i^n - x_{i-1}^n = \frac{b-a}{n}$.

Sabemos que $F_X'(x) = f_X(x)$. Por el Teorema del Valor Medio, para todo i , $1 \leq i \leq n$, existe $\xi_i^n \in (x_{i-1}^n, x_i^n]$ tal que

$$F_X(x_i^n) - F_X(x_{i-1}^n) = f_X(\xi_i^n) (x_i^n - x_{i-1}^n). \quad (7.4)$$

Eligiremos estos puntos ξ_i^n , $1 \leq i \leq n$, para formar las sumas de Riemann-Stiljes. Luego

$$S_a^b(\pi^n, \xi^n, x, F_X) = \sum_{i=1}^n \xi_i^n (F_X(x_i^n) - F_X(x_{i-1}^n)), \quad (7.5)$$

y se tendrá que

$$\lim_{n \rightarrow \infty} S_a^b(\pi^n, \xi^n, x, F_X) = \int_a^b x dF_X. \quad (7.6)$$

Usando (7.5) y (7.4) obtenemos que $S_a^b(\pi^n, \xi^n, x, F_X)$ es también una suma de Riemann correspondiente a la función $x f_X$. En efecto

$$\begin{aligned} S_a^b(\pi^n, \xi^n, x, F_X) &= \sum_{i=1}^n \xi_i^n f_X(\xi_i^n) (x_i^n - x_{i-1}^n) \\ &= S_a^b(\pi^n, \xi^n, x f_X(x), x). \end{aligned}$$

Luego

$$\lim_{n \rightarrow \infty} S_a^b(\pi^n, \xi^n, x, F_X) = \int_a^b x f_X(x) dx. \quad (7.7)$$

De (7.6) y (7.7) se obtiene (7.3).

Algunas propiedades de la esperanza matemática

P1. Sea X una variable aleatoria tal que $P_X(\{a\}) = 1$. Entonces

$$E(X) = a.$$

Esto es inmediato teniendo en cuenta X es una variable discreta con $R_X = \{a\}$ y $p_X(a) = 1$. Luego

$$E(X) = \sum_{x \in R_X} x p_X(x) = a.$$

P2. Sea (Ω, \mathcal{A}, P) un espacio de probabilidad y $A \in \mathcal{A}$.

Definimos la variable aleatoria

$$X(\omega) = I_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \notin A. \end{cases}$$

En este caso $R_X = \{0, 1\}$, $p_X(1) = P(A)$, y $p_X(0) = 1 - P(A)$.
Entonces

$$E(X) = 0(1 - P(A)) + 1P(A) = P(A).$$

7.3. Integración por partes

El siguiente teorema permite la integración por partes de una integral de Riemann-Stieljes.

Teorema. Sean g y F funciones definidas sobre $[a, b]$ tales que $\int_a^b g dF$ existe. Supongamos que g sea continua en a . Entonces

$$\int_a^b g dF = g(x) F(x) \Big|_a^b - \int_a^b F dg.$$

Demostración.

Consideremos para cada n , una partición de puntos equidistantes del intervalo $[a, b]$

$$\pi^n = \{x_0^n, x_1^n, \dots, x_n^n\}$$

tal que $a = x_0^n < x_1^n < \dots < x_n^n = b$ donde para todo i , $x_i^n - x_{i-1}^n = \frac{b-a}{n}$.

Elegimos en cada subintervalo $\xi_i^n \in x_{i-1}^n, x_i^n]$ como el extremo derecho

$\xi_i^n = x_i^n$. Entonces

$$\begin{aligned}
S_a^b(\pi^n, \xi^n, g, F) &= \sum_{i=1}^n g(x_i^n) (F(x_i^n) - F(x_{i-1}^n)) = \\
&= \sum_{i=1}^n g(x_i^n) F(x_i^n) - \sum_{i=1}^n g(x_i^n) F(x_{i-1}^n) = \\
&= \sum_{i=1}^{n-1} g(x_i^n) F(x_i^n) + g(b) F(b) - \sum_{i=2}^n g(x_i^n) F(x_{i-1}^n) - g(x_1) F(a) \\
&= \sum_{i=2}^n g(x_{i-1}^n) F(x_{i-1}^n) + g(b) F(b) - \sum_{i=2}^n g(x_i^n) F(x_{i-1}^n) - g(x_1) F(a) \\
&= - \sum_{i=2}^n F(x_{i-1}^n) (g(x_i^n) - g(x_{i-1}^n)) + g(b) F(b) - g(x_1) F(a) \\
&= - \sum_{i=2}^n F(x_{i-1}^n) (g(x_i^n) - g(x_{i-1}^n)) - F(a)(g(x_1) - g(a)) \\
&\quad + g(b)F(b) - g(a)F(a) \\
&= g(x)F(x)|_a^b - \sum_{i=1}^n F(x_{i-1}^n) (g(x_i^n) - g(x_{i-1}^n))
\end{aligned}$$

Como

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n F(x_{i-1}^n) (g(x_i^n) - g(x_{i-1}^n)) = - \int_a^b F dg$$

resulta la tesis del Teorema \square

P3. Dada una función F monótona se tiene

$$\int_a^b dF = F(b) - F(a).$$

Aplicando integración partes s con $g = 1$ y dado que $dg = 0$, obtenemos

$$\int_a^b dF = 1F(x)|_a^b - \int_a^b F dg = F_X(x)|_a^b = F(b) - F(a).$$

Lema. Supongamos que $\int |x|dF_X < \infty$. Entonces vale

(i)
$$\lim_{x \rightarrow +\infty} x(1 - F_X(x)) = 0, \quad (7.8)$$

y

(ii)
$$\lim_{x \rightarrow -\infty} xF_X(x) = 0. \quad (7.9)$$

Demostración.

Teniendo en cuenta que $\int_{-\infty}^{\infty} |x|dF_X$ es finita entonces “las colas” tienden a cero, es decir

$$\lim_{b \rightarrow +\infty} \int_b^{+\infty} x dF_X = 0, \quad (7.10)$$

y

$$\lim_{a \rightarrow -\infty} \int_{-\infty}^a x dF_X = 0. \quad (7.11)$$

Usando P3 obtenemos

$$\int_b^{+\infty} dF_X = \lim_{d \rightarrow \infty} \int_b^d dF_X = \lim_{d \rightarrow \infty} F_X(d) - F_X(b) = 1 - F_X(b),$$

y entonces resulta

$$\int_b^{+\infty} x dF_X \leq b \int_b^{+\infty} dF_X = b(1 - F_X(b)) \geq 0.$$

Luego

$$0 = \lim_{b \rightarrow \infty} \int_b^{+\infty} x dF_X \leq \lim_{b \rightarrow \infty} b(1 - F_X(b)) \geq 0.$$

Luego se deduce (7.10).

(ii) se prueba de manera análoga y se deja como ejercicio. \square

Ahora estamos en condiciones de dar una expresión de la esperanza como sumas de integrales de Riemann.

Teorema. Supongamos que $\int_{-\infty}^{\infty} |x|dF_X < \infty$. Entonces

$$E(X) = \int_0^{+\infty} (1 - F_X(x)) dx - \int_{-\infty}^0 F_X(x) dx. \quad (7.12)$$

Demostración.

Sabemos que

$$E(X) = \int_0^{+\infty} x dF_X + \int_{-\infty}^0 x dF_X.$$

Estudiaremos cada integral por separado. Integrando por partes tenemos que

$$\begin{aligned}
\int_0^b x dF_X &= xF_X(x)|_0^b - \int_0^b F_X(x) dx = \\
&= bF_X(b) - \int_0^b F_X(x) dx = \\
&= bF_X(b) + b - b - \int_0^b F_X(x) dx = \\
&= -b(1 - F_X(b)) + b - \int_0^b F_X(x) dx = \\
&= -b(1 - F_X(b)) + \int_0^b dx - \int_0^b F_X(x) dx = \\
&= -b(1 - F_X(b)) + \int_0^b (1 - F_X(x)) dx.
\end{aligned}$$

Luego pasando al límite y teniendo en cuenta el resultado (7.8) se obtiene

$$\int_0^{+\infty} x dF_X = \int_0^{+\infty} (1 - F_X(x)) dx.$$

Análogamente se prueba

$$\int_{-\infty}^0 x dF_X = - \int_0^{+\infty} F_X(x) dx.$$

De esta dos últimas igualdades se obtiene el Teorema. \square

P4. Sean X e Y dos variables aleatorias tal que $P(X \leq Y) = 1$, y tal que sus esperanzas $E(X)$, $E(Y)$ existen.

Entonces

$$E(X) \leq E(Y).$$

Demostración.

Consideremos el evento $U = \{\omega : X(\omega) \leq Y(\omega)\}$. Claramente $P(U) = 1$ y $P(U^c) = 0$

Podemos escribir

$$\{X \leq x\} = (\{X \leq x\} \cap U) \cup (\{X \leq x\} \cap U^c).$$

Si $\omega \in \{Y \leq x\} \cap U$ entonces $X(\omega) \leq Y(\omega) \leq x$ de manera que

$$\{Y \leq x\} \cap U \subset \{X \leq x\} \cap U.$$

Entonces resulta

$$(\{Y \leq x\} \cap U) \cup (\{X \leq x\} \cap U^c) \subset \{X \leq x\}.$$

Tomando probabilidades y teniendo en cuenta que $P(\{Y \leq x\} \cap U) = P(\{Y \leq x\})$ y $P(\{X \leq x\} \cap U^c) = 0$ se obtiene que

$$P(\{Y \leq x\}) \leq P(\{X \leq x\}),$$

o bien

$$F_Y(x) \leq F_X(x). \quad (7.13)$$

Luego

$$1 - F_X(x) \leq 1 - F_Y(x). \quad (7.14)$$

Teniendo en cuenta (7.12) resulta

$$E(X) = \int_0^{+\infty} (1 - F_X(x)) dx - \int_{-\infty}^0 F_X(x) dx,$$

$$E(Y) = \int_0^{+\infty} (1 - F_Y(x)) dx - \int_{-\infty}^0 F_Y(x) dx.$$

y de (7.13) y (7.14) se obtiene el resultado \square

P5. Supongamos que $P(X = 0) = 1$. Por la propiedad P1 es claro que $E(X) = 0$.

Ahora bien, del hecho de que $E(X) = 0$ no se deduce que $P(X = 0) = 1$. ¿Que condición podemos agregar para que se cumpla? La propiedad P5 responde a esta pregunta.

P6. $E(X) = 0$ y $P(X \geq 0) = 1$ implica que $P(X = 0) = 1$.

Demostración

Supongamos que esta propiedad no es cierta, luego tendríamos una variable aleatoria X tal que $E(X) = 0$, $P(X \geq 0) = 1$ y $P(X = 0) < 1$. Luego teniendo en cuenta que $P(X \geq 0) = 1$ obtenemos que $P(X > 0) = P(X \geq 0) - P(X = 0) = 1 - P(X = 0) = a > 0$.

Ahora consideremos los eventos $A_n = \{X \geq \frac{1}{n}\}$. La sucesión $\{A_n\}$ es monótona creciente, $A_n \subset A_{n+1}$ y

$$\{X > 0\} = \bigcup_{n \in \mathbb{N}} A_n,$$

de manera que

$$\lim_{n \rightarrow \infty} P(A_n) = P(\{X > 0\}) = a > 0.$$

Por lo tanto existe un número natural n_0 tal que $P(A_{n_0}) > a/2$ y entonces

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x dF_X = \int_0^{+\infty} x dF_X = \int_0^{\frac{1}{n_0}} x dF_X + \int_{\frac{1}{n_0}}^{+\infty} x dF_X \geq \\ &\geq \int_{\frac{1}{n_0}}^{+\infty} x dF_X \geq \frac{1}{n_0} \int_{\frac{1}{n_0}}^{+\infty} dF_X = \frac{1}{n_0} \left(1 - F_X\left(\frac{1}{n_0}\right)\right) = \\ &= \frac{1}{n_0} P\left(X > \frac{1}{n_0}\right) = \frac{1}{n_0} \frac{a}{2} > 0. \end{aligned}$$

lo cual es un absurdo ya que contradice la hipótesis. \square

Observación.

La igualdad $\int_{-\infty}^{+\infty} x dF_X = \int_0^{+\infty} x dF_X$ se justifica teniendo en cuenta que $P(X \geq 0) = 1$.

7.4. Esperanza de transformadas de variables aleatorias

7.4.1. Caso discreto

Sea X una variable aleatoria discreta, R_X su rango y p_X su densidad. Sabemos que

$$E(X) = \sum_{x \in R_X} x p_X(x).$$

El siguiente Teorema generaliza esta fórmula

Teorema. consideremos \mathbf{X} un vector aleatorio discreto de dimensión k y sea $g: \mathbb{R}^k \rightarrow \mathbb{R}$ una función medible. Definamos $Y = g(\mathbf{X})$. Entonces

$$E(Y) = \sum_{\mathbf{x} \in R_{\mathbf{X}}} g(\mathbf{x}) p_{\mathbf{X}}(\mathbf{x}).$$

Demostración.

Sea $y \in g(R_{\mathbf{X}}) = R_Y$ y definamos

$$A_y = \{\mathbf{x} \in R_{\mathbf{X}} : g(\mathbf{x}) = y\} = g^{-1}(\{y\}).$$

Es fácil ver que la familia de subconjuntos $\{A_y\}_{y \in R_Y}$ es una partición de $R_{\mathbf{X}}$, es decir $R_{\mathbf{X}} = \bigcup_{y \in R_Y} A_y$ y si $y \neq y'$ entonces $A_y \cap A_{y'} = \emptyset$.

Teniendo en cuenta que

$$p_Y(y) = P_{\mathbf{X}}(A_y) = \sum_{\mathbf{x} \in A_y} p_{\mathbf{X}}(\mathbf{x}),$$

y que para todo $\mathbf{x} \in A_y$ se tiene $g(\mathbf{x}) = y$, obtenemos

$$\begin{aligned} E(Y) &= \sum_{y \in R_Y} y p_Y(y) = \sum_{y \in R_Y} y \sum_{\mathbf{x} \in A_y} p_{\mathbf{X}}(\mathbf{x}) = \\ &= \sum_{y \in R_Y} \sum_{\mathbf{x} \in A_y} g(\mathbf{x}) p_{\mathbf{X}}(\mathbf{x}) = \sum_{\mathbf{x} \in R_{\mathbf{X}}} g(\mathbf{x}) p_{\mathbf{X}}(\mathbf{x}) \end{aligned}$$

ypor lo tanto queda demostrado el Teorema. \square

7.4.2. Caso continuo

Ahora pasamos al caso absolutamente continuo. Sea X una variable aleatoria absolutamente continua y f_X su función de densidad. Sabemos que

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx.$$

El siguiente Teorema es el análogo al Teorema anterior cuando \mathbf{X} es un vector absolutamente continuo

Teorema. Sea \mathbf{X} un vector aleatorio absolutamente continuo de dimensión k , con densidad $f_{\mathbf{X}}$. Sea $g : \mathbb{R}^k \rightarrow \mathbb{R}$ una función medible que toma a lo sumo un numerable de valores y definamos $Y = g(\mathbf{X})$. Luego

$$E(Y) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) dx_1 \dots dx_k. \quad (7.15)$$

Demostración.

Como en el teorema anterior consideramos la partición

$$A_y = \{\mathbf{x} \in R_{\mathbf{X}} : g(\mathbf{x}) = y\} = g^{-1}(\{y\}).$$

En este caso $\mathbb{R}^k = \bigcup_{y \in R_Y} A_y$ y si $y \neq y'$ entonces $A_y \cap A_{y'} = \emptyset$. Además $p_Y(y) = P_{\mathbf{X}}(g^{-1}(\{y\})) = P_{\mathbf{X}}(A_y)$. Entonces usando que para $\mathbf{x} \in A_y$ se tiene $g(\mathbf{x}) = y$ y que

$$\sum_{y \in R_Y} I_{A_y}(\mathbf{x}) = 1$$

obtenemos

$$\begin{aligned}
 E(Y) &= \sum_{y \in R_Y} y p_Y(y) \\
 &= \sum_{y \in R_Y} y P_{\mathbf{X}}(A_y) \\
 &= \sum_{y \in R_Y} y \int \cdots \int_{A_y} f_{\mathbf{X}}(\mathbf{x}) dx_1 \dots dx_k \\
 &= \sum_{y \in R_Y} \int \cdots \int_{A_y} y f_{\mathbf{X}}(\mathbf{x}) dx_1 \dots dx_k \\
 &= \sum_{y \in R_Y} \int \cdots \int_{A_y} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) dx_1 \dots dx_k \\
 &= \sum_{y \in R_Y} \int \cdots \int_{\mathbb{R}^k} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) I_{A_y} dx_1 \dots dx_k \\
 &= \int \cdots \int_{\mathbb{R}^k} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \left(\sum_{y \in R_Y} I_{A_y}(\mathbf{x}) \right) dx_1 \dots dx_k = \\
 &= \int \cdots \int_{\mathbb{R}^k} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) dx_1 \dots dx_k \quad \square
 \end{aligned}$$

P7. Sea X una variable aleatoria. Entonces para todo números reales c y λ

(i) $E(X + c) = E(X) + c.$

(ii) $E(\lambda X) = \lambda E(X).$

Demostración.

(i) Sea $Y = X + c$. Sabemos que $F_Y(x) = F_X(x - c)$. Entonces integrando por partes, haciendo el cambio de variable $x = y - c$ y sumando y restando $F_X(x) x|_{a-c}^{b-c}$ se obtiene

$$\begin{aligned}
 \int_a^b y dF_Y &= y F_Y(y) \Big|_a^b - \int_a^b F_Y(y) dy \\
 &= y F_X(y - c) \Big|_a^b - \int_a^b F_X(y - c) dy = \\
 &= b F_X(b - c) - a F_X(a - c) - \int_{a-c}^{b-c} F_X(x) dx + F_X(x) x \Big|_{a-c}^{b-c} - F_X(x) x \Big|_{a-c}^{b-c}.
 \end{aligned}$$

Luego, integrando por partes resulta

$$\begin{aligned}
 - \int_{a-c}^{b-c} F_X(x) dx &= xF_X(x) \Big|_{a-c}^{b-c} - \int_{a-c}^{b-c} x dF_X = \\
 &= -(b-c)F_X(b-c) + (a-c)F_X(a-c) + \int_{a-c}^{b-c} x dF_X \\
 &= -F_X(x)x \Big|_{a-c}^{b-c} + \int_{a-c}^{b-c} x dF_X.
 \end{aligned}$$

Reemplazando y cancelando algunos términos obtenemos

$$\begin{aligned}
 \int_a^b y dF_Y &= bF_X(b-c) - aF_X(a-c) - F_X(x)x \Big|_{a-c}^{b-c} + \int_{a-c}^{b-c} x dF_X + F_X(x)x \Big|_{a-c}^{b-c} - F_X(x)x \Big|_a^b \\
 &= bF_X(b-c) - aF_X(a-c) + \int_{a-c}^{b-c} x dF_X - (b-c)F_X(b-c) + (a-c)F_X(a-c) \\
 &= \int_{a-c}^{b-c} x dF_X + cF_X(b-c) - cF_X(a-c).
 \end{aligned}$$

Finalmente pasando al límite se obtiene el resultado

$$\begin{aligned}
 E(Y) &= \lim_{a \rightarrow -\infty, b \rightarrow +\infty} \int_a^b y dF_Y = \lim_{a \rightarrow -\infty, b \rightarrow +\infty} \left[\int_{a-c}^{b-c} x dF_X + cF_X(b-c) - cF_X(a-c) \right] = \\
 &= E(X) + c. \quad \square
 \end{aligned}$$

La prueba de (ii) se deja como ejercicio.

Recordemos el concepto de convergencia uniforme

Definición. Sea $(f_n)_{n \geq 1}$ una sucesión de funciones definidas sobre $[a, b]$.

Se dice que la sucesión de funciones $(f_n)_{n \geq 1}$ converge uniformemente a la función f sobre $[a, b]$ si i para cada $\varepsilon > 0$ existe $n_0 \in \mathbb{N}$ tal que si $n \geq n_0$ entonces para todo $x \in [a, b]$

$$|f_n(x) - f(x)| < \varepsilon.$$

Observación.

La diferencia con la convergencia puntual es que el n_0 en este caso sirve para todo x , es decir sólo depende de ε .

La convergencia uniforme implica la puntual pero no al revés. En particular nos interesa la convergencia uniforme de variables aleatorias. Hacemos notar que el límite puntual de funciones medibles y en consecuencia el uniforme es una función medible.

Notación: En el caso de que la convergencia sea uniforme escribiremos $X_n \Rightarrow X$.

Teorema. Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias que convergen uniformemente a una variable aleatoria X . Supongamos que $E(X)$ existe. Entonces

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

Observación

La existencia de $E(X)$ implica la existencia de $E(X_n)$ para todo n a partir de un valor n_0 .

Demostración.

Teniendo en cuenta la convergencia uniforme dado $\varepsilon > 0$ existe $n_0 \in \mathbb{N}$ tal que si $n \geq n_0$ entonces

$$\sup |X_n - X| < \varepsilon.$$

Esto significa si $n \geq n_0$ entonces

$$|X_n - X| < \varepsilon,$$

o bien

$$X - \varepsilon < X_n < X + \varepsilon.$$

Por la monotonia de la esperanza y **P7** se obtiene que si $n \geq n_0$ entonces

$$E(X) - \varepsilon \leq E(X_n) \leq E(X) + \varepsilon.$$

Por lo tanto $\lim E(X_n) = E(X)$. \square

Teorema. Sea $\mathbf{X} = (X_1, X_2, \dots, X_k)$ un vector aleatorio absolutamente continuo con función de densidad $f_{\mathbf{X}}$ y $g : \mathbb{R}^k \rightarrow \mathbb{R}$ una función medible arbitraria. Si definimos la variable aleatoria $Y = g(\mathbf{X})$ entonces

$$E(Y) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}.$$

Observación.

La misma afirmación la probamos anteriormente para una función medible que asume un número a lo sumo numerable de valores. También hay que observar que la variable aleatoria Y no tiene por qué ser absolutamente continua, pues g podría ser constante y en tal caso Y discreta.

La estrategia de la siguiente demostración es la siguiente y se usa a menudo: Se aproxima uniformemente la variable por una sucesión de variables cuya esperanzas satisfacen la propiedad en cuestión. Luego se pasa al límite haciendo uso del resultado que asegura que si X_n converge uniformemente a X entonces $E(X_n)$ converge a $E(X)$

Demostración.

Sea g_n la función definida de la siguiente manera. Dado x , existe algún entero i tal que

$$\frac{i}{n} < g(x) \leq \frac{i+1}{n}.$$

Luego definimos

$$g_n(x) = \frac{i+1}{n}$$

Teniendo en cuenta que $g_n^{-1}(\frac{i+1}{n}) = g^{-1}((\frac{i}{n}, \frac{i+1}{n}])$ se ve que para cada n , g_n es una función medible Borel. Definimos la variable aleatoria $Y_n = g_n(\mathbf{X})$. Teniendo en cuenta que

$$|g_n(x) - g(x)| < \frac{1}{n},$$

se obtiene que $Y_n \Rightarrow Y$.

Como ya hemos demostrado este teorema para funciones g que toman un numerable de valores, se tendrá

$$E(Y_n) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} g_n(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}.$$

Como, por el teorema anterior $\lim_{n \rightarrow \infty} E(Y_n) = E(Y)$, bastará probar

$$\lim_{n \rightarrow +\infty} \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} g_n(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}. \quad (7.16)$$

Luego obtenemos

$$\begin{aligned} & \left| \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} g_n(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} - \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} g(\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \right| \\ &= \left| \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} (g_n(\mathbf{x}) - g(\mathbf{x})) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \right| \\ &\leq \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} |g_n(\mathbf{x}) - g(\mathbf{x})| f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\ &\leq \frac{1}{n} \underbrace{\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}}_{=1} = \frac{1}{n}, \end{aligned}$$

y por lo tanto se cumple (7.16) \square .

Ahora vamos a probar la linealidad de la esperanza.

Teorema. Supongamos que X e Y son variables aleatorias con esperanza finita. Entonces para todo escalar α y β vale que

$$E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y).$$

Observación.

Para su demostración utilizaremos la misma estrategia comentada anteriormente.

Demostración.

Sean X e Y variables aleatorias discretas con esperanza finita. Definimos $Z = \alpha X + \beta Y$ y $g : \mathbb{R}^2 \rightarrow \mathbb{R}$

$$g(x, y) = \alpha x + \beta y.$$

Entonces si $\mathbf{X} = (X, Y)$ se tiene que $Z = g(\mathbf{X})$. Sabemos que para el caso discreto vale

$$\begin{aligned} E(Z) &= \sum_{(x,y) \in R_{(X,Y)}} g(x, y) p_{(X,Y)}(x, y) = \sum_{(x,y) \in R_{(X,Y)}} (\alpha x + \beta y) p_{(X,Y)}(x, y) = \\ &= \alpha \sum_{(x,y) \in R_{(X,Y)}} x p_{(X,Y)}(x, y) + \beta \sum_{(x,y) \in R_{(X,Y)}} y p_{(X,Y)}(x, y). \end{aligned}$$

Si probamos que

$$\sum_{(x,y) \in R_{(X,Y)}} x p_{(X,Y)}(x, y) = E(X),$$

y

$$\sum_{(x,y) \in R_{\mathbf{X}}} y p_{(X,Y)}(x, y) = E(Y).$$

se obtiene la linealidad para el caso discreto.

Bastará probar la primera de estas dos ecuaciones. Consideremos la proyección a la primer eje, $g_1 : \mathbb{R}^2 \rightarrow \mathbb{R}$ donde $g_1(x, y) = x$. Observando que $X = g_1(\mathbf{X})$, tenemos

$$\begin{aligned} E(X) &= E(g_1(\mathbf{X})) = \sum_{(x,y) \in R_{(X,Y)}} g_1(x, y) p_{(X,Y)}(x, y) \\ &= \sum_{(x,y) \in R_{(X,Y)}} x p_{(X,Y)}(x, y). \end{aligned}$$

que es el resultado buscado.

Ahora bien, si X e Y son variables aleatorias arbitrarias, entonces podemos definir dos sucesiones de variables aleatorias $(X_n)_{n \geq 1}$ e $(Y_n)_{n \geq 1}$ tal que $X_n \Rightarrow X$, y $Y_n \Rightarrow Y$

Más precisamente si $X(\omega) \in (\frac{i}{n}; \frac{i+1}{n}]$ definimos $X_n(\omega) = \frac{i+1}{n}$. Es fácil ver que

$$|X_n(\omega) - X(\omega)| < \frac{1}{n}.$$

Análogamente definimos Y_n . Ambas convergencias son uniformes, de manera que $(\alpha X_n + \beta Y_n)_{n \geq 1}$ converge uniformemente a la variable aleatoria $\alpha X + \beta Y = Z$.

Para el caso discreto vimos la linealidad, es decir

$$E(\alpha X_n + \beta Y_n) = \alpha E(X_n) + \beta E(Y_n).$$

Entonces teniendo en cuenta esa propiedad y del resultado que asegura que si X_n converge uniformemente a X entonces $E(X_n)$ converge a $E(X)$ se obtiene

$$\lim_{n \rightarrow \infty} E(\alpha X_n + \beta Y_n) = E(\alpha X + \beta Y).$$

Como además

$$\begin{aligned} \lim_{n \rightarrow \infty} E(\alpha X_n + \beta Y_n) &= \lim_{n \rightarrow \infty} (\alpha E(X_n) + \beta E(Y_n)) \\ &= \alpha \lim_{n \rightarrow \infty} E(X_n) + \beta E(Y_n) \\ &= \alpha E(X) + \beta E(Y), \end{aligned}$$

se prueba la tesis \square .

7.5. Esperanza del producto de variables aleatorias independientes

Otro problema interesante es estudiar la esperanza de un producto de variables aleatorias. Si las variables aleatorias X e Y tienen esperanzas finitas y definimos la variable aleatoria $Z = XY$ entonces no preguntamos ¿cuando vale que $E(Z) = E(XY) = E(X)E(Y)$? Una condición suficiente es la independencia de las variables X e Y .

Teorema. Sean X e Y variables aleatorias independientes con esperanza finita. Si $Z = XY$ entonces

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

Observación

La estrategia para su demostración vuelve a ser la misma que antes.

Demostración

En principio lo probaremos para el caso discreto. Luego aproximaremos uniformemente y finalmente pasaremos al límite.

Sean X e Y variables aleatorias discretas independientes con esperanza finita y definamos $g : \mathbb{R}^2 \rightarrow \mathbb{R}$

$$g(x, y) = xy.$$

Entonces $Z = g(\mathbf{X})$. Sabemos que

$$\begin{aligned}
 E(Z) &= \sum_{(x,y) \in R_{(X,Y)}} g(x,y) p_{(X,Y)}(x,y) \\
 &= \sum_{(x,y) \in R_X \times R_Y} xyp_{(X,Y)}(x,y) \\
 &= \sum_{(x,y) \in R_X \times R_Y} (xp_X(x))(yp_Y(y)) \\
 &= \sum_{x \in R_X} xp_X(x) \sum_{y \in R_Y} yp_Y(y) \\
 &= E(X)E(Y).
 \end{aligned}$$

Observemos que $R_{(X,Y)} \subset R_X \times R_Y$ pero para $(x,y) \in R_X \times R_Y - R_{(X,Y)}$ se tiene $p_{(X,Y)}(x,y) = 0$, lo que justifica la segunda igualdad. La tercera se justifica por el hecho que dado que X e Y son independientes se tiene $p_{(X,Y)}(x,y) = p_X(x)p_Y(y)$.

Ahora vamos a dar la prueba para una variables aleatorias cualquiera. Dada una función medible $g : \mathbb{R}^k \rightarrow \mathbb{R}$ existe una sucesión de funciones medibles $\{g_n\}_{n \geq 1}$ que asumen un número de valores a lo sumo infinito numerable tal que $|g_n - g| < \frac{1}{n}$.

Utilizaremos esto con la función identidad. Es decir, existe una sucesión de funciones medibles $\{g_n\}_{n \geq 1}$ cada una con un sólo valor $\frac{i}{n}$ tal que $|g_n(x) - x| < \frac{1}{n}$.

Sean X e Y independientes y consideremos las sucesiones de variables aleatorias discretas $g_n(X) = X_n$ e $Y_n = g_n(Y)$. Dado que g_n es medible y X e Y son independientes, se tiene que X_n e Y_n son independientes.

Luego, por el caso anterior

$$E(X_n Y_n) = E(X_n)E(Y_n).$$

Ahora teniendo en cuenta que $X_n \Rightarrow X$, $Y_n \Rightarrow Y$

$$\lim_{n \rightarrow \infty} E(X_n Y_n) = \lim_{n \rightarrow \infty} E(X_n) \lim_{n \rightarrow \infty} E(Y_n) = E(X)E(Y).$$

Luego basta probar que $\lim_{n \rightarrow \infty} E(X_n Y_n) = E(XY)$. Para ver esto observemos que

$$\begin{aligned}
 |E(X_n Y_n) - E(XY)| &= |E(X_n Y_n - XY)| \\
 &\leq E|X_n Y_n - XY| \\
 &= E|X_n Y_n - X_n Y + X_n Y - XY| \\
 &= E|X_n (Y_n - Y) + Y (X_n - X)| \\
 &\leq E(|X_n (Y_n - Y)| + |Y (X_n - X)|) \\
 &\leq E(|X_n| |Y_n - Y|) + E(|Y| |X_n - X|).
 \end{aligned}$$

Por la convergencia uniforme existe $n_0 \in \mathbb{N}$ tal que si $n \geq n_0$ entonces $|Y_n - Y| \leq \frac{\varepsilon}{2(E(|X|) + 1)}$, $|X_n - X| \leq \frac{\varepsilon}{2E(|Y|)}$ y además que $E(|X_n|) \leq E(|X|) + 1$.

Luego

$$\begin{aligned} |E(X_n Y_n) - E(XY)| &\leq E(|X_n| |Y_n - Y|) + E(|Y| |X_n - X|) \leq \\ &\leq \text{máx } |Y_n - Y| E(|X_n|) + \text{máx } |X_n - X| E(|Y|) \leq \\ &\leq \frac{\varepsilon}{2(E(|X|) + 1)} (E(|X|) + 1) + \frac{\varepsilon}{2E(|Y|)} (E(|Y|)) = \varepsilon, \end{aligned}$$

y se obtiene el resultado buscado \square .

Damos a continuación un ejemplo que muestra que la recíproca es falsa, es decir de la propiedad $E(XY) = E(X)E(Y)$ no se deduce la independencia de las variables.

Consideremos un vector (X, Y) discreto tal que $R_{(X,Y)} = \{(-1, 0); (1, 0); (0, -1); (0, 1)\}$ y tal que $p(x, y) = 1/4$ para cada $(x, y) \in R_{(X,Y)}$. Sea $g(x, y) = xy$ y $Z = XY = g(X, Y)$

Como para todo $(x, y) \in R_{(X,Y)}$, se tiene $xy = 0$, resulta $P(XY \equiv 0) = 1$. Luego $E(XY) = 0$. También se ve que $R_X = \{-1, 0, 1\}$ y $p_X(-1) = 1/4$, $p_X(0) = 1/2$ y $p_X(1) = 1/4$. por lo tanto que $E(X) = -1(1/4) + 0(1/2) + 1(1/4) = 0$. De manera que se cumple $E(XY) = E(X)E(Y) = 0$. Pero X e Y no son independientes pues $p_X(1) = \frac{1}{4} = p_Y(1)$ y dado que $(1, 1) \notin R_{(X,Y)}$ se tiene $p_{(X,Y)}(1, 1) = 0$.

Sin embargo si X, Y fueran independientes debiera cumplirse

$$p_{(X,Y)}(1, 1) = p_X(1/4)p_Y(1/4) = \frac{1}{4} \frac{1}{4} = \frac{1}{16}.$$

de donde resulta una contradicción. Por lo tanto X e Y no son independientes..

7.6. Esperanza de distribuciones sinétricas

El concepto de esperanza matemática esta ligado con el valor central de la distribución. Ciertas variables llamadas simétricas tienen un centro natural. Por ejemplo aquellas que tienen densidad simétrica respecto a un punto.

Definición. Dada una variable aleatoria X cualquiera, se dice que tiene distribución simétrica respecto de μ si

$$P_X((-\infty, \mu - x]) = P_X([\mu + x, \infty)). \quad (7.17)$$

Proposición. X tiene distribución simétrica respecto de μ si y solo si $Y = X - \mu$ tiene distribución simétrica respecto de 0.

Demostración. Se tiene

$$\begin{aligned} P_X((-\infty, \mu - x]) &= P(X \leq \mu - x) \\ &= P(Y \leq -x) \\ &= P_Y((-\infty, -x]), \end{aligned}$$

y

$$\begin{aligned} P_X([\mu + x, \infty)) &= P(X \geq \mu + x) \\ &= P(Y \geq x) \\ &= P_Y([x, \infty)). \end{aligned}$$

Luego $P_X((-\infty, \mu - x]) = P_X([\mu + x, \infty))$ es equivalente a $P_Y((-\infty, -x]) = P_Y([x, \infty))$ y por lo tanto la proposición es cierta \square .

Proposición (i) Si X es absolutamente continua, entonces X tiene distribución simétrica respecto de μ si y solo si

$$f_X(\mu - x) = f_X(\mu + x). \quad (7.18)$$

(ii) Si X es discreta, entonces X tiene distribución simétrica respecto de μ si y solo si

$$p_X(\mu - x) = p_X(\mu + x).$$

Demostración. Solo probaremos (i) Para facilitar la demostración aunque no es necesario, supondremos que f_X es continua y por lo tanto

$$F'_X(x) = f_X(x). \quad (7.19)$$

Sea X absolutamente continua simétrica respecto de μ satisfaciendo (7.19). En este caso (7.17) es equivalente a

$$F_X(\mu - x) = 1 - F_X(\mu + x),$$

y derivando se obtiene

$$-f_X(\mu - x) = -f_X(\mu + x),$$

y por lo tanto

$$f_X(\mu - x) = f_X(\mu + x).$$

Supongamos ahora que X satisface (7.18), entonces

$$\begin{aligned}
 P_X((-\infty, \mu - x]) &= \int_{-\infty}^{\mu-x} f_X(z) dz \\
 &= \int_{-\infty}^{\mu-x} f_X(\mu - (\mu - z)) dz \\
 &= \int_{-\infty}^{\mu-x} f_X(\mu + (\mu - z)) dz \quad (7.20) \\
 &= \int_{-\infty}^{\mu-x} f_X(2\mu - z) dz.
 \end{aligned}$$

Haciendo la transformación $w = 2\mu - z$ se obtiene

$$\begin{aligned}
 \int_{-\infty}^{\mu-x} f_X(2\mu - z) dz &= - \int_{\infty}^{\mu+x} f_X(w) dw \\
 &= \int_{\mu+x}^{\infty} f_X(w) dw \\
 &= P_X([\mu + x, \infty)). \quad (7.21)
 \end{aligned}$$

De (7.20) y (7.21) resulta (7.17), y por lo tanto X es simétrica respecto de μ \square .

Teorema. Si X es simétrica respecto de μ y $E(X)$ existe y es finita, resulta $E(X) = \mu$ Demostración. Lo demostraremos unicamente en el caso de que X sea absolutamente. Supongamos primero que $\mu = 0$, y por lo tanto por (7.18) $f_X(-x) = f_X(x)$. Luego

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx = \int_{-\infty}^0 x f_X(x) dx + \int_0^{+\infty} x f_X(x) dx, \quad (7.22)$$

y haciendo el cambio de variable $y = -x$, $dy = -dx$, resulta

$$\begin{aligned}
 \int_{-\infty}^0 x f_X(x) dx &= \int_{\infty}^0 y f_X(y) dy \\
 &= - \int_{-0}^{\infty} y f_X(y) dy \\
 &= - \int_{-0}^{\infty} x f_X(x) dx. \quad (7.23)
 \end{aligned}$$

Luego de (7.22) y (7.23) resulta $E(X) = 0$. Tomemos ahora X simétrica respecto de μ . Por el resultado anterior $Y = X - \mu$ es simétrica respecto de 0. Luego $E(Y) = 0 = E(X) - \mu$, Luego $E(X) = \mu$ \square .

La esperanza de una variable aleatoria es una característica de su distribución que describe un valor central. Sin embargo, variables aleatorias con distribuciones muy distintas pueden tener la misma esperanza. Por ejemplo pueden diferir en como se dispersan los valores que toma la variable alrededor de la esperanza: pueden estar más o menos dispersos. Esto nos lleva a definir otras características, en particular características de dispersión.

7.7. Mediana de una variable aleatoria.

Dijimos que la esperanza describe un valor central de una variable aleatoria. En particular, si la variable aleatoria X es simétrica y tiene esperanza finita, entonces esta coincide con su centro de simetría. Una desventaja de la esperanza es que es muy inestable; es decir es muy sensible a las pequeñas perturbaciones; pequeños errores en las observaciones se ven reflejados dramáticamente en los valores de la esperanza.

Otra desventaja se refiere al hecho de su posible no existencia. Puede ocurrir incluso que la distribución sea simétrica y su esperanza no exista. Un ejemplo de ello es la distribución de Cauchy cuya función de densidad es

$$f_X(x) = \frac{1}{\pi} \frac{1}{1+x^2}.$$

El gráfico de esta densidad es parecido a la de la densidad normal aunque las colas tienden a 0 más lentamente. Es una función par y por lo tanto simétrica respecto del eje y . Esta distribución no tiene esperanza puesto que un cálculo sencillo prueba que

$$\frac{1}{\pi} \int_0^{+\infty} x \frac{1}{1+x^2} dx = -\frac{1}{\pi} \int_{-\infty}^0 x \frac{1}{1+x^2} dx = +\infty.$$

Por lo tanto la simetría no garantiza la existencia de la esperanza. En este sentido no es una buena medida de centralidad.

Otra medida de centralidad es la mediana. Si existe un valor que deja la misma probabilidad a su derecha que a la izquierda, ese valor es la mediana. Esto se podrá lograr siempre en el caso de una variable aleatoria continua. Si X es simétrica y existe la esperanza entonces la mediana coincide con la esperanza y ambas con el centro de simetría. Una definición general de mediana es la siguiente

Definición. Se dice que m es una *mediana* de la variable aleatoria X si se cumple que

(i)

$$P(X \geq m) \geq \frac{1}{2}$$

y

(ii)

$$P(X \leq m) \geq \frac{1}{2}.$$

Veremos que siempre existe, y que si no es única, el conjunto de las medianas es conexo, es decir es un intervalo en \mathbb{R} .

Para mostrar esto necesitaremos recurrir a la función

$$F_X^{-1}(y) = \inf A_y,$$

donde $A_y = \{x : F_X(x) \geq y\}$. Hemos visto que el ínfimo es en verdad un mínimo, de manera que $F_X(F_X^{-1}(y)) \geq y$ es decir

$$P(X \leq F_X^{-1}(y)) \geq y. \quad (7.24)$$

Probaremos ahora una propiedad adicional

Lema

$$P(X \geq F_X^{-1}(y)) \geq 1 - y. \quad (7.25)$$

Demostración. Sea $x < F_X^{-1}(y)$, entonces, dado que $F_X^{-1}(y)$ es el mínimo de A_y se tiene que $F_X(x) < y$

Luego si ponemos $x = F_X^{-1}(y) - \frac{1}{n} < F_X^{-1}(y)$ obtenemos

$$F_X\left(F_X^{-1}(y) - \frac{1}{n}\right) < y,$$

es decir

$$P\left(X \leq F_X^{-1}(y) - \frac{1}{n}\right) < y.$$

La sucesión de eventos

$$A_n = \left\{X \leq F_X^{-1}(y) - \frac{1}{n}\right\}$$

es monótona no decreciente y además

$$\bigcup_{n=1}^{\infty} A_n = \{X < F_X^{-1}(y)\}.$$

Luego pasando al límite se tiene

$$\lim_{n \rightarrow \infty} P\left(X \leq F_X^{-1}(y) - \frac{1}{n}\right) \leq y,$$

y además

$$\lim_{n \rightarrow \infty} P\left(X \leq F_X^{-1}(y) - \frac{1}{n}\right) = P(\{X < F_X^{-1}(y)\}).$$

Por lo tanto

$$P(\{X < F_X^{-1}(y)\}) \leq y,$$

o equivalentemente

$$P(\{X \geq F_X^{-1}(y)\}) \geq 1 - y.$$

□.

Teorema. Sea X una variable aleatoria y F_X su distribución. Entonces

(i)

$$F_X^{-1}\left(\frac{1}{2}\right)$$

es una mediana.

(ii) Si m es mediana de X entonces

$$F_X^{-1}\left(\frac{1}{2}\right) \leq m.$$

(iii) Si m_1 y m_2 son medianas de X entonces para todo $m \in (m_1, m_2)$, m es mediana de X .

Demostración

(i) Se deduce de (7.24) y (7.25 tomando $y = 1/2$

(ii) Si m es otra mediana, entonces como $P(X \leq m) \geq 1/2$, resulta que $m \in A_{1/2}$. Como $F_X^{-1}(1/2) = \inf A_{1/2}$ resulta $F_X^{-1}(1/2) > m$

(iii) Se deja como ejercicio. \square

También se propone como ejercicio dar ejemplos de distribuciones en los que el intervalo de las medianas sea cerrado a derecha y ejemplos en los que sea abierto a derecha.

En el caso de que se trate de un intervalo podemos definir la mediana central como el punto medio del intervalo. Es decir si el conjunto de medianas es el intervalo $[a, b)$ o el $[a, b]$; la mediana central es $m_c(X) = \frac{a+b}{2}$.

7.8. Varianza de una variable aleatoria.

No sólo nos interesan definir medidas de centralidad como la media y la mediana, sino también cuan dispersos están los valores que toma la variable alrededor de un valor central.

Tampoco existe una única manera de medir dicha "dispersión". Consideremos una variable aleatoria X . Podríamos considerar la distancia entre los valores que toma X y su esperanza, es decir $|X - E(X)|$ y como es una variable aleatoria calcular su esperanza $E(|X - E(X)|)$. Sin embargo, dado que la función valor absoluto no es derivable en el origen, será conveniente reemplazarla por la función cuadrática.

Definición. Definimos la *varianza* de la variable aleatoria X

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

La *desviación típica* de una variable aleatoria X se define como la raíz cuadrada de la varianza

$$ds(X) = \sqrt{\text{Var}(X)}.$$

Observación.

Es Inmediato observar que $\text{Var}(X) \geq 0$ pues se trata de la esperanza de una variable aleatoria no negativa. También es claro que siempre existe si admitimos como medida el valor $+\infty$.

La varianza tiene las siguientes propiedades

P1 $\text{Var}(X) = E(X^2) - E^2(X)$.

Luego para el caso discreto resulta

$$\text{Var}(X) = \sum_{x \in R_X} x^2 p_X(x) - \sum_{x \in R_X} x p_X(x),$$

y para el continuo

$$\text{Var}(X) = \int_{-\infty}^{\infty} x^2 f_X(x) dx - \left(\int_{-\infty}^{\infty} x f_X(x) dx \right)^2.$$

Demostración.

Teniendo en cuenta las propiedades de la esperanza, se obtiene que:

$$\begin{aligned} \text{Var}(X) &= E((X - E(X))^2) \\ &= E(X^2 - 2E(X)X + E^2(X)) \\ &= E(X^2) - 2E(X)E(X) + E^2(X) \\ &= E(X^2) - E^2(X). \end{aligned}$$

Desde luego consideramos dichas expresiones solo en caso de que estén bien definidas. \square

P2. $\text{Var}(X) = 0$ es equivalente a $P(X = E(X)) = 1$.

Demostración. Supongamos $\text{Var}(X) = E((X - E(X))^2) = 0$. Como $(X - E(X))^2$ es no negativa, resulta que

$$P((X - E(X))^2 = 0) = 1.$$

Esto equivale a que

$$P(X - E(X) = 0) = 1,$$

o

$$P(X = E(X)) = 1.$$

Se deja como ejercicio probar que si

$$P(X = E(X)) = 1,$$

entonces $\text{Var}(X) = 0$. Para eso obsérvese que la variable aleatoria $(X - E(X))^2$ es cero con probabilidad uno. \square

P3. Sea X una variable aleatoria y $Y = \alpha X + \beta$, con α, β escalares. Entonces $\text{Var}(Y) = \alpha^2 \text{Var}(X)$.

Demostración. Como $E(Y) = \alpha E(X) + \beta$ resulta

$$\begin{aligned}\text{Var}(Y) &= E((Y - E(Y))^2) \\ &= E((\alpha X + \beta - (\alpha E(X) + \beta))^2) \\ &= E((\alpha(X - E(X)))^2) \\ &= \alpha^2 E((X - E(X))^2) \\ &= \alpha^2 \text{Var}(X).\end{aligned}$$

□

Se mostrará que en el caso de suma de variables aleatorias independientes, la varianza es aditiva.

P4. Sean X e Y variables aleatorias independientes. Luego si $Z = X + Y$ resulta $\text{Var}(Z) = \text{Var}(X) + \text{Var}(Y)$.

Demostración. Tenemos

$$\begin{aligned}\text{Var}(Z) &= E((Z - E(Z))^2) \\ &= E((X + Y - E(X) - E(Y))^2) \\ &= E([(X - E(X)) + (Y - E(Y))]^2) \\ &= E((X - E(X))^2) + 2E[(X - E(X))(Y - E(Y))] + E((Y - E(Y))^2) \\ &= \text{Var}(X) + 2E[(X - EX)(Y - E(Y))] + \text{Var}(Y).\end{aligned}$$

Luego, bastará probar que

$$E[(X - EX)(Y - E(Y))] = 0.$$

Usando la independencia de X e Y y teniendo en cuenta que

$$E(X - EX) = 0 = E(Y - EY),$$

resulta

$$\begin{aligned}E[(X - EX)(Y - E(Y))] &= E(X - EX)E(Y - E(Y)) \\ &= 0.\end{aligned}\tag{7.26}$$

□

7.9. Covarianza

La ecuación (7.26) motiva la definición del concepto de covarianza:

Definición. Sean X e Y variables aleatorias. Se define la *covarianza* de X e Y como

$$\text{Cov}(X, Y) = E(X - EX)E(Y - E(Y)).$$

Las siguientes propiedades P5 y P6 son inmediatas

$$\mathbf{P5.} \quad \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y).$$

La recíproca es falsa : la covarianza igual a cero no garantiza la independencia de las variables. Para esto revisar el contraejemplo de que $E(XY) = E(X)E(Y)$.

$$\mathbf{P6.} \quad \text{Si } X, Y \text{ son independientes, } \text{Cov}(X, Y) = 0$$

Diremos que dos variables aleatorias X e Y están positivamente correlacionadas si $\text{Cov}(X, Y) > 0$ y negativamente correlacionadas si $\text{Cov}(X, Y) < 0$.

Si $\text{Cov}(X, Y) = E(X - E(X))E(Y - E(Y)) > 0$ $X - EX$ y $Y - E(Y)$ tienden a tener el mismo signo, es decir tienden a situarse del mismo lado de sus respectivas esperanzas. Lo contrario ocurre si $\text{Cov}(X, Y) < 0$.

P7. Si X e Y son variables aleatorias y ponemos $X' = \alpha X + \beta$ e $Y' = \gamma Y + \delta$ entonces

$$\text{Cov}(X', Y') = \alpha\gamma \text{Cov}(X, Y).$$

Demostración. Para probar obsérvese que

$$\begin{aligned} X' - E(X') &= \alpha X + \beta - (\alpha E(X) + \beta) = \alpha(X - E(X)), \\ Y' - E(Y') &= \gamma Y + \delta - (\gamma E(Y) + \delta) = \gamma(Y - E(Y)). \end{aligned}$$

Luego

$$\begin{aligned} E((X' - E(X'))(Y' - E(Y'))) &= E(\alpha\gamma E(X - E(X))(Y - E(Y))) \\ &= \alpha\gamma E(E(X - E(X))(Y - E(Y))). \end{aligned}$$

de donde se obtiene el resultado enunciado. \square

Ahora enunciaremos la desigualdad de Cauchy-Schwarz para variables aleatorias.

Teorema. Desigualdad de Cauchy-Schwarz. Sean X e Y variables aleatorias. Entonces si las varianzas de ambas variables son finitas

$$E^2(XY) \leq E(X^2)E(Y^2), \quad (7.27)$$

y la igualdad ocurre si y solo si existe α tal que $P(Y = \alpha X) = 1$. Además

$$\text{Cov}^2(X, Y) \leq \text{Var}(X)\text{Var}(Y), \quad (7.28)$$

y la igualdad ocurre si y solo si existen escalares α, β tal que

$$P(Y = \alpha X + \beta) = 1. \quad (7.29)$$

Demostración.

Sea $Z = Y - \alpha X$. Entonces

$$Q(\alpha) = E(Z^2) = \alpha^2 E(X^2) + E(Y^2) - 2\alpha E(X, Y) \geq 0.$$

es un polinomio no negativo de segundo grado en α y como tiene a lo sumo una raíz su discriminante es no positivo

$$\Delta = 4E^2(XY) - 4E(X^2)E(Y^2) = 4(E^2(XY) - E(X^2)E(Y^2)) \leq 0.$$

Luego

$$E^2(X, Y) - E^2(X)E^2(Y) \leq 0,$$

y se obtiene el resultado.

La igualdad se cumple si $\Delta = 0$. Esto ocurre si y solo si existe un único α tal que $Q(\alpha) = 0$. Esto es equivalente a que $E((Y - \alpha X)^2) = 0$, y esto a que $P(Y = \alpha X) = 1$.

La desigualdad (7.27) se obtiene aplicando (7.28) a $X^* = X - E(X)$ y $Y^* = Y - E(Y)$. Luego resulta que la correspondiente igualdad se cumple si y solo si existe α tal que

$$P(Y - E(Y) = \alpha(X - E(X))) = 1.$$

Poniendo $\beta = E(Y) + \alpha E(X)$, esto es equivalente a (7.29). \square

Definición. Dadas las variables aleatorias X e Y se define el *cuadrado del coeficiente de correlación* y se denota por $\rho^2(X, Y)$ a

$$\rho^2(X, Y) = \frac{\text{Cov}^2(X, Y)}{\text{Var}(X)\text{Var}(Y)}.$$

También definimos el *coeficiente de correlación* entre X e Y por

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{[\text{Var}(X)]^{\frac{1}{2}} [\text{Var}(Y)]^{\frac{1}{2}}}.$$

De Cauchy-Schwarz se deduce la siguiente propiedad:

P7. $0 \leq \rho(X, Y)^2 \leq 1$ y $-1 \leq \rho(X, Y) \leq 1$. Además $\rho(X, Y)^2 = 1$ es equivalente a que para algún α y β se tenga $P(Y = \alpha X + \beta) = 1$, es decir a que haya una relación lineal perfecta entre las variables X e Y .

7.10. Distribución Normal Bivariada.

Calcularemos ahora $E(Y)$ y $\text{Var}(Y)$ para una variable Y con distribución $N(\mu, \sigma^2)$.

Teorema. Si $Y \sim N(\mu, \sigma^2)$ entonces $E(Y) = \mu$ y $\text{Var}(Y) = \sigma^2$.

Demostración. Tomemos primero una variable X con distribución $N(0, 1)$. Mostraremos que $E(X) = 0$ y $\text{Var}(X) = 1$. La densidad de X es

$$f(x) = \frac{1}{(2\pi)^{1/2}} e^{-x^2/2}.$$

Como X es simétrica respecto de 0, para mostrar que $E(X) = 0$, bastara mostrar que $E(|X|) < \infty$. Tenemos que

$$\begin{aligned} E(|X|) &= \int_{-\infty}^{\infty} |x|f(x)dx \\ &= 2 \int_0^{\infty} xf(x)dx \\ &= \frac{2}{(2\pi)^{1/2}} \int_0^{\infty} xe^{-x^2/2}dx. \end{aligned} \tag{7.30}$$

Definamos $u = x^2/2$. Luego $du = x$ y entonces calculemos la integral indefinida

$$\begin{aligned} \int xe^{-x^2/2}dx &= \int e^{-u}du \\ &= -e^{-u} \\ &= -e^{-x^2/2}. \end{aligned} \tag{7.31}$$

Luego por (7.30) y (7.31) se tiene

$$\int_{-\infty}^{\infty} |x|f(x)dx = \frac{2^{1/2}}{\pi^{1/2}} \left[-e^{-x^2/2} \right]_0^{\infty} = \frac{2^{1/2}}{\pi^{1/2}} < \infty.$$

Vamos ahora a calcular la integral indefinida

$$\int x^2 e^{-x^2/2} dx.$$

Haciendo $u = x$ y $dv = xe^{-x^2/2}dx$, se tiene $du = dx$ y por (7.31) $v = -e^{-x^2/2}$. Luego

$$\begin{aligned} \int x^2 e^{-x^2/2} dx &= \int u dv \\ &= uv - \int v du \\ &= -xe^{-x^2/2} + \int e^{-x^2/2} dx. \end{aligned}$$

Luego

$$\int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx = -[xe^{-x^2/2}]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-x^2/2} dx,$$

y como $[xe^{-x^2/2}]_{-\infty}^{\infty} = 0$, resulta

$$\int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx = \int_{-\infty}^{\infty} e^{-x^2/2} dx.$$

Entonces se tiene

$$\begin{aligned} \text{Var}(X) &= \int_{-\infty}^{\infty} x^2 f(x) dx \\ &= \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx \\ &= \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{\infty} e^{-x^2/2} dx \\ &= \int_{-\infty}^{\infty} f(x) dx \\ &= 1. \end{aligned}$$

De acuerdo a su definición, la distribución $N(\mu, \sigma^2)$ es la distribución de $Y = \sigma X + \mu$, con $X \sim N(\mu, \sigma^2)$. Luego $E(Y) = \sigma E(X) + \mu = \mu$ y $\text{Var}(Y) = \sigma^2 \text{Var}(X) = \sigma^2$.

Observación De acuerdo a este resultado, los parámetros de una distribución normal coinciden con la esperanza y la varianza.

Vamos a definir una distribución bivariada normal con medias, varianzas y covarianzas arbitrarias.

Queremos definir la distribución conjunta de un vector $\mathbf{Y} = (Y_1, Y_2)$ de forma tal que $Y_1 \sim N(\mu_1, \sigma_1^2)$ y $Y_2 \sim N(\mu_2, \sigma_2^2)$, y tal que $\text{Cov}(Y_1, Y_2) = \sigma_{12}$. Los valores $\mu_1, \mu_2, \sigma_1^2 > 0, \sigma_2^2 > 0$ y σ_{12} son arbitrarios aunque como veremos se tendrán que cumplir ciertas restricciones.

Teniendo en cuenta la desigualdad de Cauchy-Schwarz se deberá tener que $\sigma_{12}^2 \leq \sigma_1^2 \sigma_2^2$. Ahora bien si queremos una distribución bivariada absolutamente continua, no podrá cumplirse $\sigma_{12}^2 = \sigma_1^2 \sigma_2^2$, ya que en este caso (Y_1, Y_2) estaría sobre una recta que es un conjunto de superficie 0. Luego se deberá cumplir $\sigma_{12}^2 < \sigma_1^2 \sigma_2^2$.

Sea la matriz Σ definida por

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}.$$

Luego $\det(\Sigma) = \sigma_1^2 \sigma_2^2 - \sigma_{12}^2 > 0$.

Definamos la matriz de covarianza del vector \mathbf{Y} por

$$\Sigma_{\mathbf{Y}} = \begin{pmatrix} \text{Var}(Y_1) & \text{Cov}(Y_1, Y_2) \\ \text{Cov}(Y_1, Y_2) & \text{Var}(Y_2) \end{pmatrix}.$$

Como $\det(\Sigma) = \sigma_1^2 \sigma_2^2 - \sigma_{12}^2 > 0$ y $\sigma_1^2 > 0$, Σ resulta simétrica y definida positiva. Luego tiene al menos una "raíz cuadrada" en el sentido de que

existe una matriz (no unica)

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

tal que

$$\Sigma = AA^t,$$

donde A^t designa su transpuesta.. Definamos el vector $\mathbf{Y} = A\mathbf{X} + \mu$ donde

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}; \mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix},$$

siendo X_1 , y X_2 variables aleatorias independientes $N(0, 1)$. Entonces tenemos el siguiente Teorema

Teorema.

- (i) Y_1 tiene distribución $N(\mu_1, \sigma_1^2)$ y Y_2 distribución $N(\mu_2, \sigma_2^2)$
- (ii) $\text{cov}(Y_1, Y_2) = \sigma_{12}$
- (iii) La densidad del vector \mathbf{Y} está dada por

$$f(\mathbf{y}) = \frac{1}{2\pi \det(\Sigma)^{\frac{1}{2}}} \exp\left(\frac{-1}{2} (\mathbf{y} - \mu)^t \Sigma^{-1} (\mathbf{y} - \mu)\right).$$

- (iv) La forma cuadrática $Q(\mathbf{y}) = (\mathbf{y} - \mu)^t \Sigma^{-1} (\mathbf{y} - \mu)$ es igual a

$$\frac{1}{\sigma_1^2 \sigma_2^2 (1 - \rho^2)} \left[(y_1 - \mu_1)^2 \sigma_2^2 + (y_2 - \mu_2)^2 \sigma_1^2 - 2(y_1 - \mu_1)(y_2 - \mu_2) \sigma_{12} \right].$$

Demostración.

Observemos que el vector \mathbf{Y} satisface

$$\begin{aligned} Y_1 &= a_{11}X_1 + a_{12}X_2 + \mu_1, \\ Y_2 &= a_{21}X_1 + a_{22}X_2 + \mu_2. \end{aligned}$$

Como $E(X_1) = E(X_2) = 0$, resulta

$$E(Y_1) = \mu_1, \quad E(Y_2) = \mu_2.$$

Ademas como $\text{cov}(X_1, X_2) = 0$, $\text{Var}(X_1) = \text{Var}(X_2) = 1$, resulta

$$\begin{aligned} \text{Var}(Y_1) &= a_{11}^2 \text{Var}(X_1) + a_{12}^2 \text{Var}(X_2) \\ &= a_{11}^2 + a_{12}^2. \end{aligned}$$

Similarmente

$$\text{Var}(Y_2) = a_{21}^2 + a_{22}^2,$$

y como $E(X_1 X_2) = 0$.

$$\begin{aligned} \text{Cov}(Y_1, Y_2) &= E((a_{11}X_1 + a_{12}X_2)(a_{21}X_1 + a_{22}X_2)) \\ &= a_{11}a_{21}E(X_1^2) + a_{12}a_{22}E(X_2^2) + (a_{12}a_{21} + a_{11}a_{22})E(X_1 X_2) \\ &= a_{11}a_{21} + a_{12}a_{22}. \end{aligned}$$

Luego

$$\begin{aligned} \Sigma_{\mathbf{Y}} &= \begin{pmatrix} a_{11}^2 + a_{12}^2 & a_{11}a_{21} + a_{12}a_{22} \\ a_{11}a_{21} + a_{12}a_{22} & a_{21}^2 + a_{22}^2 \end{pmatrix} \\ &= AA^T \\ &= \Sigma \\ &= \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}. \end{aligned}$$

De acuerdo a lo que ya sabemos como Y_1 e Y_2 son combinaciones lineales de normales independientes serán normales. Luego de acuerdo a los cálculos realizados de esperanzas y varianzas se cumplirá que Y_1 es $N(\mu_1, \sigma_1^2)$, Y_2 es $N(\mu_2, \sigma_2^2)$ y además $\text{Cov}(Y_1, Y_2) = \sigma_{12}$. Esto prueba (i) y (ii).

Vamos a calcular la distribución conjunta del vector \mathbf{Y} .

Comencemos escribiendo la distribución conjunta del vector \mathbf{X} . Como X_1 y X_2 son independientes, la distribución conjunta de \mathbf{X} es el producto de las marginales.

$$\begin{aligned} f_{\mathbf{X}}(\mathbf{x}) &= \frac{1}{2\pi} \exp\left(-\frac{x_1^2}{2}\right) \exp\left(-\frac{x_2^2}{2}\right) \\ &= \frac{1}{2\pi} \exp\left(-\frac{(x_1^2 + x_2^2)}{2}\right) \\ &= \frac{1}{2\pi} \exp\left(-\frac{1}{2}\mathbf{xx}'\right), \end{aligned}$$

donde $\mathbf{xx}' = \|\mathbf{x}\|^2$.

Teniendo en cuenta que $\mathbf{X} = A^{-1}(\mathbf{Y} - \mu)$ se obtiene que el Jacobiano $J = \frac{1}{\det(A)}$. Además, como $\Sigma = AA^t$ se obtiene que $\det(A)^2 = \det(\Sigma)$ o sea $\det(A) = \det(\Sigma)^{\frac{1}{2}}$ y por lo tanto $J = \frac{1}{\det(\Sigma)^{\frac{1}{2}}}$.

Entonces teniendo en cuenta que $(A^t)^{-1}A^{-1} = \Sigma^{-1}$ y usando la fórmula para transformaciones de vectores aleatorios, resulta

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}) &= \frac{1}{2\pi \det(\Sigma)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{y} - \mu)^t (A^t)^{-1} A^{-1} (\mathbf{y} - \mu)\right) \\ &= \frac{1}{2\pi \det(\Sigma)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{y} - \mu)^t \Sigma^{-1} (\mathbf{y} - \mu)\right). \end{aligned}$$

Calculemos el determinante de Σ

$$\det(\Sigma) = \sigma_1^2 \sigma_2^2 - \sigma_{12}^2 = \sigma_1^2 \sigma_2^2 \left(1 - \frac{\sigma_{12}^2}{\sigma_1^2 \sigma_2^2}\right) = \sigma_1^2 \sigma_2^2 (1 - \rho^2).$$

Luego la inversa de Σ viene dada por

$$\Sigma^{-1} = \frac{1}{\sigma_1^2 \sigma_2^2 (1 - \rho^2)} \begin{pmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{12} & \sigma_1^2 \end{pmatrix}.$$

Entonces la forma cuadrática se puede escribir como

$$\begin{aligned} (\mathbf{y} - \boldsymbol{\mu})^t \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu}) &= (\mathbf{y} - \boldsymbol{\mu})^t \frac{1}{\sigma_1^2 \sigma_2^2 (1 - \rho^2)} \begin{pmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{12} & \sigma_1^2 \end{pmatrix} (\mathbf{y} - \boldsymbol{\mu}) \\ &= \frac{1}{\sigma_1^2 \sigma_2^2 (1 - \rho^2)} ((y_1 - \mu_1)^2 \sigma_2^2 + (y_2 - \mu_2)^2 \sigma_1^2 - \\ &\quad 2(y_1 - \mu_1)(y_2 - \mu_2) \sigma_{12}). \end{aligned}$$

Luego se tiene

$$\begin{aligned} &(\mathbf{y} - \boldsymbol{\mu})^t \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu}) \\ &= \frac{1}{1 - \rho^2} \left(\frac{(y_1 - \mu_1)^2}{\sigma_1^2} + \frac{(y_2 - \mu_2)^2}{\sigma_2^2} - 2 \frac{\sigma_{12}}{\sigma_1^2 \sigma_2^2} (y_1 - \mu_1)(y_2 - \mu_2) \right) \\ &= \frac{1}{1 - \rho^2} \left(\frac{(y_1 - \mu_1)^2}{\sigma_1^2} + \frac{(y_2 - \mu_2)^2}{\sigma_2^2} - 2 \frac{\rho}{\sigma_1 \sigma_2} (y_1 - \mu_1)(y_2 - \mu_2) \right). \end{aligned}$$

□.

Observación. El Teorema anterior se demostró para el caso de dos variables. Sin embargo la densidad normal multivariada de cualquier dimensión tiene una expresión similar al punto (iii).

Observación. El máximo valor de $f_{\mathbf{Y}}$ se logra cuando se hace mínimo el exponente de la exponencial, esto es en $\mathbf{y} = \boldsymbol{\mu}$. Por otro lado las curvas de nivel $f_{\mathbf{Y}}(\mathbf{y}) = c$ (c constante) son elipses cuyas direcciones principales vienen dadas por los autovectores de Σ^{-1} . Si la $\text{Cov}(Y_1, Y_2) = 0$ entonces, la matriz Σ es diagonal y las direcciones son paralelas a los ejes coordenados.

Definición. Se dice que el vector \mathbf{Y} tiene una *distribución normal bivariada* con media $\boldsymbol{\mu}$ y matriz de covarianza Σ , que se denotará por $N(\boldsymbol{\mu}, \Sigma)$ si su función densidad es

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{2\pi \det(\Sigma)^{\frac{1}{2}}} \exp\left(\frac{-1}{2} (\mathbf{y} - \boldsymbol{\mu})^t \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu})\right).$$

Capítulo 8

Teoría de la Predicción.

8.1. Error cuadrático medio y predictores óptimos

En esta sección veremos como utilizar ciertas variables conocidas para predecir otras variables que no se pueden observar en un determinado momento. Por ejemplo se quiere predecir la cantidad de lluvia que mañana caerá en determinada región, utilizaré otras variables que se puedan medir hoy. En algún sentido quisiéramos encontrar el predictor que se aproxime más a la variable a predecir, entre todas los predictores pertenecientes a un conjunto dado.

Sea \mathcal{P} un conjunto de predictores para la variable aleatoria Y que forman un espacio vectorial. Cada elemento de \mathcal{P} es una variables aleatoria observable. Supongamos que se quiere predecir a Y a través de $\hat{Y} \in \mathcal{P}$. ¿Cómo se puede medir la bondad de un predictor \hat{Y} cualquiera? Se pueden considerar las siguientes alternativas

Definición. El *error cuadrático medio* del predictor \hat{Y} para predecir Y está dado por

$$\text{ECM}(\hat{Y}, Y) = E\left(\left(Y - \hat{Y}\right)^2\right)$$

y el *error absoluto medio*

$$\text{EAM}(\hat{Y}, Y) = E\left(\left|Y - \hat{Y}\right|\right).$$

Si usamos como criterio de bondad de un predictor el error cuadrático medio, diremos que $\hat{Y}_0 \in \mathcal{P}$ es un *predictor óptimo de Y en \mathcal{P}* , si dado otro $\hat{Y} \in \mathcal{P}$ se tiene

$$\text{ECM}(\hat{Y}_0, Y) \leq \text{ECM}(\hat{Y}, Y).$$

A continuación damos un criterio suficiente para obtener un predictor óptimo usando el criterio del error cuadrático medio.

Teorema. Una condición suficiente para que $\hat{Y}_0 \in \mathcal{P}$ sea un predictor óptimo usando el criterio del error cuadrático medio es que

$$E\left(\left(Y - \hat{Y}_0\right) \hat{Y}\right) = 0 \quad (8.1)$$

para todo $\hat{Y} \in \mathcal{P}$. Además, si \hat{Y}_0 satisface (8.1), es esencialmente el único predictor óptimo. Es decir si $\hat{Y} \in \mathcal{P}$ satisface $\text{ECM}(\hat{Y}_0, Y) = \text{ECM}(\hat{Y}, Y)$ entonces $P(\hat{Y} = \hat{Y}_0) = 1$.

Observación.

La condición (8.1) se puede interpretar como que $(Y - \hat{Y}_0)$ es ortogonal a todo elemento de \mathcal{P} cuando el producto escalar está definido por $\langle Y, X \rangle = E(YX)$ en el espacio de Hilbert de las variables aleatorias.

Demostración

Sea $\hat{Y} \in \mathcal{P}$. Entonces

$$\begin{aligned} \text{ECM}(\hat{Y}, Y) &= E\left(\left(Y - \hat{Y}\right)^2\right) = E\left(\left[\left(Y - \hat{Y}_0\right) + \left(\hat{Y}_0 - \hat{Y}\right)\right]^2\right) = \\ &= E\left(\left(Y - \hat{Y}_0\right)^2\right) + E\left(\left(\hat{Y}_0 - \hat{Y}\right)^2\right) - 2E\left(\left(\hat{Y}_0 - \hat{Y}\right)\left(Y - \hat{Y}_0\right)\right). \end{aligned}$$

Usando la condición de ortogonalidad, como $\hat{Y}_0 - \hat{Y} \in \mathcal{P}$ se tiene

$$E\left(\left(\hat{Y}_0 - \hat{Y}\right)\left(Y - \hat{Y}_0\right)\right) = 0,$$

y luego

$$\begin{aligned} \text{ECM}(\hat{Y}, Y) &= E\left(\left(Y - \hat{Y}_0\right)^2\right) + E\left(\left(\hat{Y}_0 - \hat{Y}\right)^2\right) \\ &\geq E\left(\left(Y - \hat{Y}_0\right)^2\right) \\ &= \text{ECM}(\hat{Y}_0, Y), \end{aligned}$$

y por lo tanto \hat{Y}_0 es óptimo.

Además si \hat{Y} fuera también óptimo se tendría $E\left(\left(\hat{Y}_0 - \hat{Y}\right)^2\right) = 0$ y siendo $\left(\hat{Y}_0 - \hat{Y}\right)^2 \geq 0$ resultaría $P(\hat{Y} = \hat{Y}_0) = 1 \square$.

La condición (8.1) es generalmente difícil de verificar. Si \mathcal{P} tiene dimensión finita, esta condición se puede simplificar. Esto queda establecido en el siguiente Teorema.

Teorema. Sea $\{\hat{Y}_1, \dots, \hat{Y}_k\}$ una base de \mathcal{P} . Una condición suficiente para que $\hat{Y}_0 \in \mathcal{P}$ sea óptimo con el criterio del error cuadrático medio es que

$$E\left(\left(Y - \hat{Y}_0\right) \hat{Y}_i\right) = 0, \quad 1 \leq i \leq k. \quad (8.2)$$

Además \widehat{Y}_0 es esencialmente el único predictor óptimo. Es decir si $\widehat{Y} \in \mathcal{P}$ satisface $\text{ECM}(\widehat{Y}_0, Y) = \text{ECM}(\widehat{Y}, Y)$ entonces $P(\widehat{Y} = \widehat{Y}_0) = 1$

Demostración. Bastará mostrar que se satisface el requerimiento del Teorema anterior. Sea \widehat{Y} cualquier elemento de \mathcal{P} , entonces existen escalares $\alpha_1, \dots, \alpha_k$ tal que $\widehat{Y} = \sum_{i=1}^k \alpha_i \widehat{Y}_i$. Luego si para $i = 1, 2, \dots, k$ se cumple que

$$E\left(\left(Y - \widehat{Y}_0\right) \widehat{Y}_i\right) = 0,$$

resulta también que

$$E\left(\left(Y - \widehat{Y}_0\right) \widehat{Y}\right) = E\left(\left(Y - \widehat{Y}_0\right) \sum_{i=1}^k \alpha_i \widehat{Y}_i\right) = \sum_{i=1}^k \alpha_i E\left(\left(Y - \widehat{Y}_0\right) \widehat{Y}_i\right) = 0. \square$$

8.2. Predictores constantes

Se pueden considerar distintos conjuntos de predictores. Comenzaremos con los predictores constantes.

Sea (Ω, \mathcal{A}, P) un espacio de probabilidad, Y una variable aleatoria a predecir y consideremos

$$\mathcal{P}_1 = \{\widehat{Y} : \widehat{Y} \text{ es una variable aleatoria constante}\}.$$

El siguiente teorema determina el predictor óptimo perteneciente a \mathcal{P}_1 .

Teorema. El predictor $\widehat{Y}_0 = E(Y)$ es el de menor error cuadrático medio en \mathcal{P}_1 . Además $\text{EMC}(\widehat{Y}_0, Y) = \text{Var}(Y)$.

Demostración. Una base de \mathcal{P}_1 es $\{\widehat{Y}_1\}$ donde $\widehat{Y}_1 = 1$. Como

$$E\left(\left(Y - \widehat{Y}_0\right) 1\right) = E(Y - E(Y)) = E(Y) - E(Y) = 0,$$

resulta $\widehat{Y}_0 = E(Y)$ el predictor de menor error cuadrático medio.

Además

$$\begin{aligned} \text{EMC}(\widehat{Y}_0, Y) &= E((Y - \widehat{Y}_0)^2) \\ &= E((Y - E(Y))^2) \\ &= \text{Var}(Y). \end{aligned}$$

□

Designamos el predictor óptimo para Y en \mathcal{P}_1 por $\widehat{Y}_{0,C}$. En la práctica únicamente se usa un predictor constante si no se observan otras variables vinculadas a Y .

8.3. Predictores lineales

Sea ahora (Ω, \mathcal{A}, P) un espacio de probabilidad, Y una variable aleatoria a predecir y X otra variable aleatoria observada. Consideremos el siguiente conjunto de predictores

$$\mathcal{P}_2 = \{ \hat{Y} : \hat{Y} = \alpha X + \beta \}.$$

\mathcal{P}_2 es el conjunto de variables aleatorias que se obtiene por una transformación lineal de la variable X . Claramente $\mathcal{P}_1 \subset \mathcal{P}_2$, y por lo tanto el error cuadrático medio del predictor óptimo en \mathcal{P}_2 será menor o igual que el del predictor óptimo en \mathcal{P}_1 . Por esta razón, si denotamos por $\hat{Y}_{0,L}$ el predictor óptimo en \mathcal{P}_2 , resulta claro que $ECM(Y, \hat{Y}_{0,L}) \leq ECM(Y, \hat{Y}_{0,C})$.

El siguiente Teorema caracteriza el predictor óptimo en \mathcal{P}_2 .

Teorema. (i) El predictor de menor error cuadrático medio en \mathcal{P}_2 está dado por $\hat{Y}_{0,L} = \alpha X + \beta$ con

$$\beta = E(Y) - \alpha E(X) \quad (8.3)$$

y

$$\alpha = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}. \quad (8.4)$$

(ii) El error cuadrático medio de $\hat{Y}_{0,L}$ está dado por

$$ECM(\hat{Y}_{0,L}, Y) = \text{Var}(Y) - \frac{\text{Cov}^2(X, Y)}{\text{Var}(X)}. \quad (8.5)$$

Demostración. Una base de \mathcal{P}_2 es $\{\hat{Y}_1, \hat{Y}_2\}$ donde $\hat{Y}_1 = X$ y $\hat{Y}_2 = 1$. Luego el predictor óptimo $\hat{Y}_{0,L}$ debe satisfacer

$$E((Y - \alpha X - \beta) X) = 0 \quad (8.6)$$

y

$$E((Y - \alpha X - \beta) 1) = 0. \quad (8.7)$$

De la condición (8.6) se obtiene

$$EY) - \alpha E(X) - \beta = 0,$$

de donde resulta (8.3)

Ahora multiplicando (8.6) por $E(X)$ resulta

$$E((Y - \alpha X - \beta) E(X)) = 0,$$

y restándola de (8.7) obtenemos

$$E((Y - \alpha X - \beta)(X - E(X))) = 0$$

Reemplazando β por (8.3) obtenemos

$$E((Y - \alpha X - E(Y) + \alpha E(X))(X - E(X))) = 0,$$

y por lo tanto

$$E((Y - E(Y)) - \alpha(X - E(X))(X - E(X))) = 0.$$

Entonces distribuyendo la esperanza se obtiene

$$\begin{aligned} \text{Cov}(X, Y) &= E[(Y - E(Y))(X - E(X))] \\ &= \alpha E[(X - E(X))^2] = \\ &\alpha \text{Var}(X), \end{aligned}$$

y por lo tanto resulta (8.4).

Ahora calcularemos el error cuadrático medio de $\widehat{Y}_{0,L}$. Usando (8.3) obtenemos

$$\begin{aligned} \text{ECM}(\widehat{Y}_{0,L}, Y) &= E((Y - \alpha X - \beta)^2) \\ &= E[(Y - \alpha X - E(Y) + \alpha E(X))^2] = \\ &= E[((Y - E(Y)) - \alpha(X - E(X)))^2] = \\ &= E((Y - E(Y))^2 + \alpha^2 E((X - E(X))^2) \\ &\quad - 2\alpha E((Y - E(Y))(X - E(X))). \end{aligned}$$

Luego, usando (8.5) se obtiene

$$\begin{aligned} \text{ECM}(\widehat{Y}_{0,L}, Y) &= \text{Var}(Y) + \alpha^2 \text{Var}(X) - 2\alpha \text{Cov}(X, Y) = \\ &= \text{Var}(Y) + \frac{\text{Cov}^2(X, Y)}{\text{Var}(X)} - 2 \frac{\text{Cov}^2(X, Y)}{\text{Var}(X)} \\ &= \text{Var}(Y) - \frac{\text{Cov}^2(X, Y)}{\text{Var}(X)}. \end{aligned}$$

□

Para evaluar cuanto mejora el error cuadrático medio cuando se usa $\widehat{Y}_{0,L}$ en vez de $\widehat{Y}_{0,C}$, calculemos su decrecimiento relativo

$$\begin{aligned} & \frac{\text{ECM}(\widehat{Y}_{0,C}, Y) - \text{ECM}(\widehat{Y}_{0,L}, Y)}{\text{ECM}(\widehat{Y}_{0,L}, Y)} \\ &= \frac{\text{Var}(Y) - \left(\text{Var}(Y) - \frac{\text{Cov}^2(X, Y)}{\text{Var}(X)}\right)}{\text{ECM}(\widehat{Y}_{0,C}, Y)} \\ &= \frac{\frac{\text{Cov}^2(X, Y)}{\text{Var}(X)}}{\text{Var}(Y)} = \frac{\text{Cov}^2(X, Y)}{\text{Var}(X) \text{Var}(Y)} = \rho^2(X, Y). \end{aligned}$$

Esto permite interpretar coeficiente $\rho^2(X, Y)$ como el decrecimiento relativo del error cuadrático medio cuando se usa in predictor lineal basado en X en vez de un predictor constante. Por lo tanto $\rho^2(X, Y)$ mide la utilidad de la variable X para predecir Y por una función lineal. Observemos que nuevamente se obtiene la desigualdad de Cauchy-Schwarz. En efecto, como $0 \leq \text{ECM}(\widehat{Y}_{0,C}, Y) - \text{ECM}(Y_{0,L}, Y) \leq \text{ECM}(Y, \widehat{Y}_{0,L})$, se obtiene $0 \leq \rho^2(X, Y) \leq 1$.

Veremos ahora el significado de los casos extremos $\rho^2(X, Y) = 1$ y $\rho^2(X, Y) = 0$. $\rho^2(X, Y) = 1$ es equivalente a $\text{ECM}_L(Y, \widehat{Y}_{0,L}) = 0$ y esto es equivalente $E\left(\left(Y - \widehat{Y}_{0,L}\right)^2\right) = 0$ y esto a

$$P\left(\left(Y = \widehat{Y}_{0,L}\right)\right) = P\left((Y = \alpha X + \beta)\right) = 1.$$

Es decir $\rho^2(X, Y) = 1$ es equivalente a que hay una relación lineal perfecta entre X e Y con probabilidad 1.

Existen dos posibilidades para $\rho^2(X, Y) = 1$ o bien $\rho(X, Y) = 1$ o $\rho(X, Y) = -1$. El signo de $\rho(X, Y)$ coincide con el de $\text{Cov}(X, Y)$ que es el mismo que el de la pendiente del predictor lineal óptimo. Luego $\rho(X, Y) = 1$ indica que la relación entre la X y la Y es creciente y $\rho(X, Y) = -1$ que la relación es decreciente.

Veremos ahora como se interpreta $\rho^2 = 0$. En este caso $\text{ECM}(\widehat{Y}_{0,L}, Y) = \text{ECM}(\widehat{Y}_{0,C}, Y)$ y $\text{Cov}(X, Y) = 0$. Por lo tanto $\alpha = 0$, y se puede concluir que la variable X no tiene utilidad para predecir Y cuando se utilizan predictores constantes.

Se deja como ejercicio probar que la recta $Y = \alpha X + \beta$ pasa por el punto $(E(X), E(Y))$, . es decir cuando $X = E(X)$ la predicción de Y es $E(Y)$.

Capítulo 9

Esperanza y distribución condicional

9.1. Caso discreto

Sean dos variables aleatorias discretas X, Y definidas sobre un mismo espacio de probabilidad (Ω, \mathcal{A}, P) . Sea $R_X = \{x : p_X(x) > 0\}$ y $R_Y = \{y : p_Y(y) > 0\}$. Luego, para cada $x \in R_X$ definimos la función de probabilidad de Y condicional $X = x$ como

$$p_{Y|X}(y|x) = \frac{p_{XY}(x, y)}{p_X(x)}.$$

Para cada $x \in R_X$ fijo esta función es una función de densidad de probabilidad ya que

$$\sum_{y \in R_Y} p_{Y|X}(y|x) = \sum_{y \in R_Y} \frac{p_{XY}(x, y)}{p_X(x)} = \frac{1}{p_X(x)} \sum_{y \in R_Y} p_{XY}(x, y) = \frac{p_X(x)}{p_X(x)} = 1,$$

y representa la distribución de Y una vez conocido que el valor de $X = x$.

Si se tienen dos vectores discretos $\mathbf{X} = (X_1, \dots, X_k)$, $\mathbf{Y} = (Y_1, \dots, Y_h)$ podemos definir una noción análoga. Sea $R_{\mathbf{X}} = \{\mathbf{x} \in R^k : p_{\mathbf{X}}(\mathbf{x}) > 0\}$, luego para todo $\mathbf{x} \in R_{\mathbf{X}}$ definimos

$$p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \frac{p_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y})}{p_{\mathbf{X}}(\mathbf{x})}, \quad (9.1)$$

y también se tendrá

$$\sum_{\mathbf{y} \in R_{\mathbf{Y}}} p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = 1.$$

Sea ahora Y una variable aleatoria y \mathbf{X} un vector aleatorio de dimensión k . La esperanza condicional de la variable Y condicional $\mathbf{X} = \mathbf{x}$ se define

como la esperanza de Y utilizando como distribución de esta variable la distribución determinada por (9.1). Es decir esta esperanza condicional se define por

$$E(Y|\mathbf{X} = \mathbf{x}) = \sum_{y \in R_y} yp_{Y|\mathbf{X}}(y|\mathbf{x}). \quad (9.2)$$

Este valor representa la esperanza de la variable Y una vez conocido que la variable X ha tomado el valor x .

Llamemos $g(\mathbf{x}) = E(Y|\mathbf{X} = \mathbf{x})$, luego $g(\mathbf{x}) : R_X \rightarrow R$. Vamos a definir ahora una variable aleatoria que llamaremos esperanza de Y condicional X , y que notaremos por $E(Y|\mathbf{X})$. Esta variable se define por

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

Vamos ahora a mostrar el siguiente Teorema

Teorema 1. Si Y tiene esperanza finita, entonces se tiene que $E(E(Y|\mathbf{X})) = E(Y)$.

Demostración. Tenemos que

$$E(E(Y|\mathbf{X})) = E(g(\mathbf{X})) = \sum_{\mathbf{x} \in R_{\mathbf{x}}} g(\mathbf{x})p_{\mathbf{X}}(\mathbf{x}).$$

Utilizando que $g(\mathbf{x})$ viene dado por (9.2), se tiene

$$\begin{aligned} E(E(Y|\mathbf{X})) &= \sum_{\mathbf{x} \in R_{\mathbf{x}}} \left(\sum_{y \in R_Y} yp_{Y|\mathbf{X}}(y|\mathbf{x}) \right) p_{\mathbf{X}}(\mathbf{x}) \\ &= \sum_{\mathbf{x} \in R_{\mathbf{x}}} \left(\sum_{y \in R_Y} y \frac{p_{\mathbf{X}Y}(\mathbf{x}, y)}{p_{\mathbf{X}}(\mathbf{x})} \right) p_{\mathbf{X}}(\mathbf{x}) \\ &= \sum_{\mathbf{x} \in R_{\mathbf{x}}} \left(\sum_{y \in R_Y} yp_{\mathbf{X}Y}(\mathbf{x}, y) \right) \\ &= \sum_{y \in R_Y} y \left(\sum_{\mathbf{x} \in R_{\mathbf{x}}} p_{\mathbf{X}Y}(\mathbf{x}, y) \right) \\ &= \sum_{y \in R_Y} yp_Y(y) \\ &= E(Y). \end{aligned}$$

Luego el Teorema queda demostrado. \square

Ejemplo. Supongamos que se hace una primera serie de n tiradas de una moneda y sea X el número de caras obtenido. En base al resultado de la primera serie de tiradas, se inicia una segunda serie de X tiradas. Sea Y el número de caras obtenidas en esta segunda serie. Calcular la $E(Y)$.

Si $X = x$, la distribución de Y condicional $X = x$ es binomial $\text{Bi}(0,50, x)$. Luego $g(x) = E(Y|X = x) = 0,50x$. Luego $E(Y|X) = g(X) = 0,50X$, y por lo tanto $E(Y) = E(E(Y|X)) = 0,50E(X)$. Como X es $\text{Bi}(0,50, n)$, entonces $E(X) = 0,5n$. Por lo tanto $E(Y) = 0,25n$.

Teorema 2. Si \mathbf{X} , \mathbf{Y} son dos vectores aleatorios independientes, entonces se tiene

(i) $p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = p_{\mathbf{Y}}(\mathbf{y})$

(ii) Si Y es una variable aleatoria y $E(Y)$ existe y es finita entonces $E(Y|\mathbf{X} = \mathbf{x}) = E(Y)$.

(iii) Sean \mathbf{X} e \mathbf{Y} son dos vectores aleatorios tales $p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = p(\mathbf{y})$ para todo $\mathbf{x} \in R_{\mathbf{X}}$. Entonces $p_{\mathbf{Y}}(\mathbf{y}) = p(\mathbf{y})$, y \mathbf{X} e \mathbf{Y} son independientes.

Demostración. Partes (i) y (ii) del Teorema son inmediatas de las definiciones.

Para probar (iii) veamos que $p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = p(\mathbf{y})$ implica que

$$p_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y}) = p_{\mathbf{X}}(\mathbf{x})p(\mathbf{y}), \tag{9.3}$$

y por lo tanto

$$p_{\mathbf{Y}}(\mathbf{y}) = \sum_{\mathbf{x} \in R_{\mathbf{X}}} p_{\mathbf{X}}(\mathbf{x})p(\mathbf{y}) = p(\mathbf{y}) \sum_{\mathbf{x} \in R_{\mathbf{X}}} p_{\mathbf{X}}(\mathbf{x}) = p(\mathbf{y}).$$

Luego reemplazando en (9.3) se obtiene

$$p_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y}) = p_{\mathbf{X}}(\mathbf{x})p_{\mathbf{Y}}(\mathbf{y}), \tag{9.4}$$

y esto implica que \mathbf{X} e \mathbf{Y} son independientes. \square

Teorema 3. Si $P(Y = c) = 1$, entonces, cualquiera sea el vector \mathbf{X} , se tiene

(i) $p_{Y|\mathbf{X}}(c|\mathbf{x}) = 1$.

(ii) $E(Y|\mathbf{X} = \mathbf{x}) = c$.

Demostración. Tenemos que

$$\{\mathbf{X} = \mathbf{x}\} = (\{\mathbf{X} = \mathbf{x}\} \cap \{Y = c\}) \cup (\{\mathbf{X} = \mathbf{x}\} \cap \{Y \neq c\}).$$

Como $P(\{\mathbf{X} = \mathbf{x}\} \cap \{Y \neq c\}) = 0$, se tiene

$$\begin{aligned} p_{\mathbf{X}}(\mathbf{x}) &= P(\mathbf{X} = \mathbf{x}) = P(\mathbf{X} = \mathbf{x}, Y = c) \\ &= p_{\mathbf{X}Y}(x, c). \end{aligned}$$

Por lo tanto

$$\begin{aligned} p_{Y|\mathbf{X}}(c|x) &= \frac{p_{\mathbf{X}Y}(x, c)}{p_{\mathbf{X}}(\mathbf{x})} \\ &= 1. \end{aligned}$$

Como en este caso $R_Y = \{c\}$, se tiene

$$\begin{aligned} E(Y|\mathbf{X} = \mathbf{x}) &= \sum_{y \in R_Y} y p_{Y|\mathbf{X}}(y|\mathbf{x}) \\ &= c p_{Y|\mathbf{X}}(c|\mathbf{x}) \\ &= c, 1 \\ &= c, \end{aligned}$$

y el Teorema queda demostrado. \square

Sean ahora dos vectores aleatorios discretos, $\mathbf{X} = (X_1, \dots, X_k)$, $\mathbf{Y} = (Y_1, \dots, Y_j)$, y sea $Z = h(\mathbf{X}, \mathbf{Y})$, donde $h : R^{k+j} \rightarrow R$ es una función medible. El siguiente teorema muestra como se calcula $E(Z|\mathbf{X} = \mathbf{x})$.

Teorema 4. Sean \mathbf{X} , \mathbf{Y} dos vectores aleatorios de dimensiones k y j , y sea $h : R^{k+j} \rightarrow R$ una función medible. Definamos la variable aleatoria $Z = h(\mathbf{X}, \mathbf{Y})$, y supongamos que tiene esperanza finita. Entonces para todo $\mathbf{x} \in R_{\mathbf{X}}$ se tiene

$$E(Z|\mathbf{X} = \mathbf{x}) = \sum_{\mathbf{y} \in R_{\mathbf{Y}}} h(\mathbf{x}, \mathbf{y}) p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}).$$

Demostración. Comenzaremos calculando la función de probabilidad conjunta de (\mathbf{X}, Z) . Sea $R_Z^{\mathbf{x}} = \{z : z = h(\mathbf{x}, \mathbf{y}) : \mathbf{y} \in R_{\mathbf{Y}}\}$, y para todo $z \in R_Z^{\mathbf{x}}$ definamos $A_z^{\mathbf{x}} = \{\mathbf{y} : h(\mathbf{x}, \mathbf{y}) = z\}$. Es fácil ver que:

Si $z \neq z'$ entonces $A_z^{\mathbf{x}} \cap A_{z'}^{\mathbf{x}} = \phi$, y que

$$\bigcup_{z \in R_Z^{\mathbf{x}}} A_z^{\mathbf{x}} = R_{\mathbf{Y}}. \quad (9.5)$$

Es inmediato que

$$p_{\mathbf{X}Z}(\mathbf{x}, z) = \begin{cases} P(\mathbf{X} = \mathbf{x}, \mathbf{Y} \in A_z^{\mathbf{x}}) = \sum_{\mathbf{y} \in A_z^{\mathbf{x}}} p_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y}) & \text{si } \mathbf{x} \in R_{\mathbf{X}}, z \in R_Z^{\mathbf{x}} \\ 0 & \text{en otro caso,} \end{cases}$$

y luego, para $\mathbf{x} \in R_{\mathbf{X}}$ se tiene

$$p_{Z|\mathbf{X}}(z|\mathbf{x}) = \frac{p_{\mathbf{X}Z}(\mathbf{x}, z)}{p_{\mathbf{X}}(\mathbf{x})} = \begin{cases} \sum_{\mathbf{y} \in A_z^{\mathbf{x}}} \frac{p_{\mathbf{X}\mathbf{Y}}(\mathbf{x}, \mathbf{y})}{p_{\mathbf{X}}(\mathbf{x})} & \text{si } z \in R_Z^{\mathbf{x}} \\ 0 & \text{en otro caso,} \end{cases}$$

y por lo tanto se tiene

$$p_{Z|\mathbf{X}}(z|\mathbf{x}) = \begin{cases} \sum_{\mathbf{y} \in A_z^{\mathbf{x}}} p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) & \text{si } z \in R_Z^{\mathbf{x}} \\ 0 & \text{en otro caso.} \end{cases} \quad (9.6)$$

Luego utilizando (9.6) se tiene

$$\begin{aligned}
 E(Z|\mathbf{X} = \mathbf{x}) &= \sum_{z \in R_Z^x} z p_{Z|\mathbf{X}}(z|\mathbf{x}) \\
 &= \sum_{z \in R_Z^x} z \sum_{\mathbf{y} \in A_z^x} p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) \\
 &= \sum_{z \in R_Z^x} \sum_{\mathbf{y} \in A_z^x} z p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}),
 \end{aligned}$$

y como para $y \in A_z^x$, se tiene $h(\mathbf{x}, \mathbf{y}) = z$, utilizando (9.5) obtenemos

$$\begin{aligned}
 E(Z|\mathbf{X} = \mathbf{x}) &= \sum_{z \in R_Z^x} \sum_{\mathbf{y} \in A_z^x} h(\mathbf{x}, \mathbf{y}) p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) \\
 &= \sum_{\mathbf{y} \in R_{\mathbf{Y}}} h(\mathbf{x}, \mathbf{y}) p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}),
 \end{aligned}$$

probando por lo tanto el Teorema. \square

El Teorema 4 se puede interpretar como que $E(Z|\mathbf{X} = \mathbf{x})$ se calcula como la esperanza de $h(\mathbf{Y}, \mathbf{x})$ (variable aleatoria que depende unicamente de la variable aleatoria \mathbf{Y} , ya que \mathbf{x} es tratada como si fuera constante) utilizando $p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$ como distribución de \mathbf{Y}

Vamos a ver que del Teorema 4 se deducen las siguientes propiedades de la esperanza condicional.

Propiedad 1. Sean \mathbf{X} un vector aleatorio de dimensión k y \mathbf{Y} un vector aleatorio de dimensión j , y sea $r : R^k \rightarrow R$ y $s : R^j \rightarrow R$. Entonces se tiene

$$E(r(\mathbf{X})s(\mathbf{Y})|\mathbf{X} = \mathbf{x}) = r(\mathbf{x})E(s(\mathbf{Y})|\mathbf{X} = \mathbf{x}).$$

Demostración. Utilizando el Teorema 4 con $h(\mathbf{x}, \mathbf{y}) = r(\mathbf{x})s(\mathbf{y})$ se tiene

$$\begin{aligned}
 E(r(\mathbf{X})s(\mathbf{Y})|\mathbf{X} = \mathbf{x}) &= \sum_{\mathbf{y} \in R_{\mathbf{Y}}} r(\mathbf{x})s(\mathbf{y})p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) \\
 &= r(\mathbf{x}) \sum_{\mathbf{y} \in R_{\mathbf{Y}}} s(\mathbf{y})p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) \\
 &= r(\mathbf{x})E(s(\mathbf{Y})|\mathbf{X} = \mathbf{x}),
 \end{aligned}$$

y luego la propiedad queda demostrada. \square

Propiedad 2 Sea \mathbf{X} un vector aleatorio de dimensión k , y sea $r : R^k \rightarrow R$. Luego $E(r(\mathbf{X})|\mathbf{X} = \mathbf{x}) = r(\mathbf{x})$.

Demostración. La demostración resulta de la Propiedad 1 tomando $s(\mathbf{y}) = 1$, ya que entonces

$E(r(\mathbf{X})|\mathbf{X} = \mathbf{x}) = r(\mathbf{x})E(1|\mathbf{X} = \mathbf{x})$. Luego por el Teorema 3 resulta la Propiedad 2.

Propiedad 3. Si Y_1 e Y_2 son variables aleatorias con esperanza finita, y \mathbf{X} es un vector aleatorio, entonces

$$E(c_1Y_1+c_2Y_2|\mathbf{X} = \mathbf{x}) = c_1E(Y_1|\mathbf{X} = \mathbf{x}) + c_2E(Y_2|\mathbf{X} = \mathbf{x}).$$

Demostración. Sea $\mathbf{Y} = (Y_1, Y_2)$ y definamos $h(\mathbf{x}, \mathbf{y}) = c_1y_1+c_2y_2$, $h_1(\mathbf{x}, \mathbf{y}) = y_1$ y $h_2(\mathbf{x}, \mathbf{y}) = y_2$. Entonces se tiene $h(\mathbf{x}, \mathbf{y}) = c_1h_1(\mathbf{x}, \mathbf{y}) + c_2h_2(\mathbf{x}, \mathbf{y})$. Luego tenemos

$$\begin{aligned} E(c_1Y_1 + c_2Y_2|\mathbf{X} = \mathbf{x}) &= E(h(\mathbf{X}, \mathbf{Y})|\mathbf{X} = \mathbf{x}) \\ &= \sum_{\mathbf{y} \in R_{\mathbf{Y}}} h(\mathbf{x}, \mathbf{y})p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) \\ &= \sum_{\mathbf{y} \in R_{\mathbf{Y}}} (c_1h_1(\mathbf{x}, \mathbf{y}) + c_2h_2(\mathbf{x}, \mathbf{y}))p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) \\ &= c_1 \sum_{\mathbf{y} \in R_{\mathbf{Y}}} h_1(\mathbf{x}, \mathbf{y})p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) + c_2 \sum_{\mathbf{y} \in R_{\mathbf{Y}}} h_2(\mathbf{x}, \mathbf{y})p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) \\ &= c_1E(h_1(\mathbf{X}, \mathbf{Y})|\mathbf{X} = \mathbf{x}) + c_2E(h_2(\mathbf{X}, \mathbf{Y})|\mathbf{X} = \mathbf{x}) \\ &= c_1E(Y_1|\mathbf{X} = \mathbf{x}) + c_2E(Y_2|\mathbf{X} = \mathbf{x}), \end{aligned}$$

y la Propiedad 3 queda demostrada. \square

Propiedad 4. Sea Y una variable aleatoria discreta con esperanza finita y \mathbf{X} un vector aleatorio discreto de dimensión k . Luego si $g(\mathbf{x}) = E(Y|\mathbf{X} = \mathbf{x})$, entonces para toda $t : R^k \rightarrow R$ medible tal que $Yt(\mathbf{X})$ tiene esperanza finita resulta

$$E((Y - g(\mathbf{X}))t(\mathbf{X})) = 0.$$

Demostración. Sea $Z = h(\mathbf{X}, Y) = (Y - g(\mathbf{X}))t(\mathbf{X})$. Luego bastará demostrar que

$$E(Z) = 0.$$

Utilizando el Teorema 1 bastará demostrar que

$$E(Z|\mathbf{X}) = 0. \tag{9.7}$$

De acuerdo a la Propiedad 1, tenemos que

$$E(Z|\mathbf{X} = \mathbf{x}) = t(\mathbf{x})E((Y - g(\mathbf{X}))|\mathbf{X} = \mathbf{x}),$$

y por lo tanto

$$E(Z|\mathbf{X}) = t(\mathbf{X})E((Y - g(\mathbf{X}))|\mathbf{X}).$$

Luego para mostrar (9.7) bastará demostrar que

$$E(Y - g(\mathbf{X})|\mathbf{X}) = 0.$$

Pero esto es cierto ya que por propiedades 2, 3 y 4 se tiene

$$\begin{aligned} E(Y - g(\mathbf{X})|X) &= E(Y|\mathbf{X}) - E(g(\mathbf{X})|\mathbf{X}) \\ &= E(Y|\mathbf{X}) - g(\mathbf{X}) \\ &= g(\mathbf{X}) - g(\mathbf{X}) \\ &= 0, \end{aligned}$$

y por lo tanto queda demostrada esta propiedad. \square

Propiedad 5. Sea Y una variable aleatoria discreta varianza finita y \mathbf{X} un vector aleatorio discreto de dimensión k . Luego $\hat{Y} = g(\mathbf{X}) = E(Y|\mathbf{X})$ es el único predictor con menor error cuadrático medio en la clase de predictores $\mathcal{P} = \{ \hat{Y} = t(\mathbf{X}) : t \text{ medible, } \text{Var}(t(\mathbf{X})) < \infty \}$

Demostración. De acuerdo al Teorema que da una condicion suficiente para un predictor óptimo en una espacio vectorial \mathcal{P} , bastará mostrar que si $\text{Var}(t(\mathbf{X})) < \infty$ entonces $E(|Yt(\mathbf{X})|) < \infty$. Esto resulta de la desigualdad de Cauchy-Schwartz. \square

9.2. Caso general

Vamos ahora dar una definición de $E(Y|\mathbf{X})$ para el caso de una variable Y cualesquiera , y un vector \mathbf{X} cualquiera de dimensión k . Ambos, Y y \mathbf{X} no tienen porque ser discretos ni absolutamente continuos

Definición. La variable aleatoria esperanza de Y condicional \mathbf{X} se define por $E(Y|\mathbf{X}) = g(\mathbf{X})$, donde $g : R^k \rightarrow R$ es una función medible tal que

$$E((Y - g(\mathbf{X}))t(\mathbf{X})) = 0 \tag{9.8}$$

para toda $t : R^k \rightarrow R$ medible tal que $Yt(\mathbf{X})$ tiene esperanza finita . Definiremos $E(Y|\mathbf{X} = \mathbf{x}) = g(\mathbf{x})$.

La Propiedad 4 demostrada anteriormente muestra que en el caso de Y y \mathbf{X} discretos esta definición coincide con la dada anteriormente, y por lo tanto en este caso siempre existe.

El siguiente teorema muestra que siempre existe una única variable aleatoria $g(\mathbf{X}) = E(Y|\mathbf{X})$ satisfaciendo (9.8).

Teorema 6. Sea Y una variable aleatoria con esperanza finita y sea \mathbf{X} un vector aleatorio cualquiera de dimensión k . Luego:

- (i) Siempre existe una función medible $g : R^k \rightarrow R$ satisfaciendo (9.8)
- (ii) Si g_1 y g_2 son dos funciones medibles satisfaciendo (9.8), entonces

$$P(g_1(\mathbf{X}) = g_2(\mathbf{X})) = 1.$$

Demostración. (i) no lo demostraremos en general en este curso. Mas adelante haremos una demostración para el caso absolutamente continuo.

(ii) Sean g_1 y g_2 son dos funciones medibles satisfaciendo (9.8), entonces

$$E((Y - g_1(\mathbf{X}))t(\mathbf{X})) = 0 \quad (9.9)$$

y

$$E((Y - g_2(\mathbf{X}))t(\mathbf{X})) = 0 \quad (9.10)$$

para toda $t(\mathbf{X})$ tal que $Yt(\mathbf{X})$ tenga esperanza finita. Luego restando (9.10) de (9.9) se obtiene

$$E((g_2(\mathbf{X}) - g_1(\mathbf{X}))t(\mathbf{X})) = 0,$$

y tomando $t(\mathbf{X}) = g_2(\mathbf{X}) - g_1(\mathbf{X})$ resulta

$$E((g_2(\mathbf{X}) - g_1(\mathbf{X}))^2) = 0.$$

Esto implica que

$$\begin{aligned} P((g_2(\mathbf{X}) - g_1(\mathbf{X}))^2 = 0) &= P(g_2(\mathbf{X}) = g_1(\mathbf{X})) \\ &= 1. \end{aligned}$$

□

Vamos ahora a demostrar que todas las propiedades de esperanza condicional que valían para el caso discreto también valen para la definición general.

Teorema 1 '. Si Y tiene esperanza finita, entonces $E(E(Y|\mathbf{X})) = E(Y)$.

Demostración. Apliquemos (9.8) con $t(\mathbf{X}) = 1$. Luego se tiene

$$\begin{aligned} 0 &= E(Y - g(\mathbf{X})) \\ &= E(Y) - E(g(\mathbf{X})) \\ &= E(Y) - E(E(Y|\mathbf{X})), \end{aligned}$$

y por lo tanto se cumple el Teorema 1 '. □

Teorema 2 '. Sean Y una variable aleatoria con esperanza finita y \mathbf{X} un vector aleatorio independientes. Entonces se tiene $E(Y|\mathbf{X}) = E(Y)$.

Demostración. Veamos que poniendo $g(\mathbf{X}) = E(Y)$ se cumple (9.8). En efecto dado que $(Y - E(Y))$ y $t(\mathbf{X})$ son independientes se tiene

$$E((Y - E(Y))t(\mathbf{X})) = E(Y - E(Y))E(t(\mathbf{X})).$$

Luego como $E(Y - E(Y)) = E(Y) - E(Y) = 0$, se tiene demostrado el Teorema 2.

Teorema 3 '. Si $P(Y = c) = 1$, entonces, cualquiera sea el vector \mathbf{X} , se tiene $E(Y|\mathbf{X}) = c$.

Demostración. Poniendo $g(\mathbf{X}) = c$, resulta inmediatamente (9.8).

Vamos ahora a probar las propiedades 1-4 para la definición general de $E(Y|\mathbf{X})$. \square

Propiedad 1 '. Sean \mathbf{X} un vector aleatorio de dimensión k y \mathbf{Y} un vector aleatorio de dimensión j , y sea $r : R^k \rightarrow R$ y $s : R^j \rightarrow R$. Entonces se tiene

$$E(r(\mathbf{X})s(\mathbf{Y})|\mathbf{X}) = r(\mathbf{X})E(s(\mathbf{Y})|\mathbf{X}).$$

Demostración. Vamos a probar que si ponemos $g(\mathbf{X}) = r(\mathbf{X})E(s(\mathbf{Y})|\mathbf{X})$, entonces (9.8) se cumple. En efecto

$$\begin{aligned} E((r(\mathbf{X})s(\mathbf{Y}) - g(\mathbf{X}))t(\mathbf{X})) &= E((r(\mathbf{X})s(\mathbf{Y}) - r(\mathbf{X})E(s(\mathbf{Y})|\mathbf{X}))t(\mathbf{X})) \\ &= E((s(\mathbf{Y}) - E(s(\mathbf{Y})|\mathbf{X}))m(\mathbf{X})), \end{aligned}$$

con $m(\mathbf{X}) = r(\mathbf{X})t(\mathbf{X})$. Luego por la definición de $E(s(\mathbf{Y})|\mathbf{X})$ obtenemos $E((s(\mathbf{Y}) - E(s(\mathbf{Y})|\mathbf{X}))m(\mathbf{X})) = 0$. Por lo tanto la propiedad queda demostrada. \square

Propiedad 2 ' Sea X un vector aleatorio de dimensión k y sea $r : R^k \rightarrow R$, una función medible. Luego $E(r(\mathbf{X})|\mathbf{X}) = r(\mathbf{X})$.

Demostración. Igual que en el caso discreto se obtiene de la Propiedad 1 tomando $s(\mathbf{Y}) = 1$. \square

Propiedad 3 '. Si Y_1 e Y_2 son variables aleatorias con esperanza finita, y \mathbf{X} es un vector aleatorio, entonces

$$E(c_1Y_1 + c_2Y_2|\mathbf{X}) = c_1E(Y_1|\mathbf{X}) + c_2E(Y_2|\mathbf{X}).$$

Demostración. Vamos a ver que se cumple (9.8) poniendo

$$g(\mathbf{X}) = c_1E(Y_1|\mathbf{X}) + c_2E(Y_2|\mathbf{X}).$$

En efecto si $Z = c_1Y_1 + c_2Y_2$ usando la linealidad de la esperanza y la definición de esperanza condicional se tiene

$$\begin{aligned} E((Z - g(\mathbf{X}))t(\mathbf{X})) &= E((c_1(Y_1 - E(Y_1|\mathbf{X})) + c_2(Y_2 - E(Y_2|\mathbf{X})))t(\mathbf{X})) \\ &= c_1E((Y_1 - E(Y_1|\mathbf{X}))t(\mathbf{X})) + c_2E((Y_2 - E(Y_2|\mathbf{X}))t(\mathbf{X})) \\ &= c_1 \cdot 0 + c_2 \cdot 0 \\ &= 0, \end{aligned}$$

y la propiedad queda demostrada. \square

La generalización de propiedad 4 usando la definición general de $E(Y|\mathbf{X})$ es obvia a partir de la definición.

Propiedad 5'. Sea Y una variable aleatoria con varianza finita y \mathbf{X} un vector aleatorio de dimensión k . Luego $\widehat{Y} = g(\mathbf{X}) = E(Y|\mathbf{X})$ es el único predictor con menor error cuadrático medio en la clase de predictores $\mathcal{P} = \{\widehat{Y} = t(\mathbf{X}) : t \text{ medible, } \text{Var}(t(\mathbf{X})) < \infty\}$

Demostración. Es totalmente similar a Propiedad 5. \square

De acuerdo a esta propiedad $E(Y|\mathbf{X})$ es el predictor de Y óptimo basado en cualquier función medible (lineal o no lineal) de \mathbf{X} . Por esta razón lo denotaremos con $\widehat{Y}_{O,NL}$.

9.3. Caso continuo

Supongamos ahora que tenemos dos vectores $\mathbf{X} = (X_1, \dots, X_k)$ e $\mathbf{Y} = (Y_1, \dots, Y_j)$ de dimensiones k y j respectivamente con distribución conjunta absolutamente continua y densidad $f_{\mathbf{X},\mathbf{Y}}$, y sea $h : R^{k+j} \rightarrow R$ una función medible. Definamos la densidad de \mathbf{Y} condicional $\mathbf{X} = \mathbf{x}$ por

$$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \frac{f_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y})}{f_{\mathbf{X}}(\mathbf{x})}.$$

Es fácil ver que para cada \mathbf{x} fijo con $f_{\mathbf{X}}(\mathbf{x}) > 0$, la función $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$ es una densidad para el vector \mathbf{Y} . Es decir se tendrá

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) dy_1 \dots dy_j = 1.$$

El siguiente Teorema es una versión para el caso continuo del Teorema 4.

Teorema 6. Sea $Z = h(\mathbf{X}, \mathbf{Y})$ una variable con esperanza finita, luego se tiene que

$$\begin{aligned} E(Z|\mathbf{X} = \mathbf{x}) &= g(\mathbf{x}) \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h(\mathbf{x}, \mathbf{y}) f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) dy_1 \dots dy_j. \end{aligned}$$

Demostración: Para facilitar la notación en la demostración, supondremos que tanto X como Y son variables aleatorias en vez de vectores. Pero excepto por la notación más complicada, la demostración para vectores es similar, ya que solamente se deben reemplazar las integrales simples por integrales múltiples.

De acuerdo a (9.8) será suficiente probar que

$$E((h(X, Y) - g(X))t(X)) = 0,$$

o equivalentemente

$$E((h(X, Y)t(X))) = E(g(X)t(X)). \quad (9.11)$$

Por un lado tenemos que

$$E((h(X, Y)t(X))) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y)t(x)f_{XY}(x, y)dxdy. \quad (9.12)$$

Además se tiene que

$$\begin{aligned} E(g(X)t(X)) &= \int_{-\infty}^{\infty} g(x)t(x)f_X(x)dxdy \\ &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} h(x, y)f_{Y|X}(y|x)dy \right] t(x)f_X(x)dx. \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y)t(x)f_{XY}(x, y)dxdy. \end{aligned} \quad (9.13)$$

Ecuaciones (9.12) y (9.13) prueban (9.11). \square

Definición . Sean dos vectores aleatorios \mathbf{X} e \mathbf{Y} de dimensiones k y j respectivamente. Luego dado $B \in \beta^j$ (conjunto Boreliano de dimensión j), la probabilidad de que $\mathbf{Y} \in B$, condicional $\mathbf{X} = \mathbf{x}$ que se denotará con $P_{\mathbf{Y}|\mathbf{X}}(B|\mathbf{X} = \mathbf{x})$ está dado por

$$P_{\mathbf{Y}|\mathbf{X}}(B|\mathbf{X} = \mathbf{x}) = E(I_B(\mathbf{Y})|\mathbf{X} = \mathbf{x}),$$

donde I_B es la función indicadora del conjunto B . La probabilidad de que $\mathbf{Y} \in B$, condicional \mathbf{X} que se denotará por $P_{\mathbf{Y}|\mathbf{X}}(B|\mathbf{X})$ está dado por

$$P_{\mathbf{Y}|\mathbf{X}}(B|\mathbf{X}) = E(I_B(\mathbf{Y})|\mathbf{X}).$$

La justificación de esta definición está dada por el hecho que

$$P_{\mathbf{Y}}(B) = E(I_B(\mathbf{Y})).$$

En efecto $I_B(\mathbf{Y})$ toma valor 1 con probabilidad $P_{\mathbf{Y}}(B)$ y 0 con probabilidad $1 - P_{\mathbf{Y}}(B)$. Luego $E(I_B(\mathbf{Y})) = 1P_{\mathbf{Y}}(B) + 0(1 - P_{\mathbf{Y}}(B)) = P_{\mathbf{Y}}(B)$.

En el caso discreto, de acuerdo al Teorema 4, se tendrá

$$\begin{aligned} P_{\mathbf{Y}|\mathbf{X}}(B|\mathbf{X} = \mathbf{x}) &= E(I_B(\mathbf{Y})|\mathbf{X} = \mathbf{x}) \\ &= \sum_{\mathbf{y} \in R_{\mathbf{Y}}} I_B(\mathbf{y})p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) \\ &= \sum_{\mathbf{y} \in R_{\mathbf{Y}} \cap B} p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}). \end{aligned}$$

En el caso absolutamente continuo, de acuerdo al Teorema 6 se tiene

$$\begin{aligned} P_{\mathbf{Y}|\mathbf{X}}(B|\mathbf{X} = \mathbf{x}) &= E(I_B(\mathbf{Y})|\mathbf{X} = \mathbf{x}) \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} I_B(\mathbf{y}) f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) d\mathbf{y} \\ &= \int_B \int f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) d\mathbf{y}. \end{aligned}$$

Observamos que $f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$ actúa como una verdadera densidad, en el sentido de que para calcular la probabilidad condicional de un evento B hay que integrar esta función sobre ese conjunto.

De acuerdo al Teorema 1 se tendrá

$$E(P_{\mathbf{Y}|\mathbf{X}}(B|\mathbf{X})) = P_{\mathbf{Y}}(B).$$

Para el caso discreto y continuo podemos definir la función de distribución de \mathbf{Y} condicional $\mathbf{X} = \mathbf{x}$, la cual se denotará por $F_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$ y estarán definidas respectivamente por

$$\begin{aligned} F_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) &= P_{\mathbf{Y}|\mathbf{X}}\left(\prod_{i=1}^j (-\infty, y_i] | \mathbf{X} = \mathbf{x}\right) \\ &= \sum_{\mathbf{z} \in R_{\mathbf{Y}} \cap \{z_1 \leq y_1\} \dots \cap \{z_j \leq y_j\}} p_{\mathbf{Y}|\mathbf{X}}(\mathbf{z}|\mathbf{x}). \end{aligned}$$

$$\begin{aligned} F_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) &= P_{\mathbf{Y}|\mathbf{X}}\left(\prod_{i=1}^j (-\infty, y_i] | \mathbf{X} = \mathbf{x}\right) \\ &= \int_{-\infty}^{y_j} \dots \int_{-\infty}^{y_1} f_{\mathbf{Y}|\mathbf{X}}(\mathbf{z}|\mathbf{x}) d\mathbf{y}. \end{aligned}$$

Es fácil ver que para cada \mathbf{x} fijo $F_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$ es una verdadera función de distribución del vector \mathbf{Y} , en el sentido que cumple con las propiedades que caracterizan a una función de distribución.

9.4. Varianza condicional

Definición: Sea $\mathbf{X} = (X_1, \dots, X_k)$ un vector aleatorio e Y una variable aleatoria con varianza finita. Entonces la varianza de Y condicional $\mathbf{X} = \mathbf{x}$ se define como

$$\text{Var}(Y|\mathbf{X} = \mathbf{x}) = E((Y - E(Y|\mathbf{X} = \mathbf{x}))^2 | \mathbf{X} = \mathbf{x}),$$

y esta varianza puede considerarse como la varianza de variable \mathbf{X} una vez que se conoce que $\mathbf{X} = \mathbf{x}$. Denotemos por $q(\mathbf{x}) = \text{Var}(Y|\mathbf{X} = \mathbf{x})$, luego

$q : R^k \rightarrow R$. Llamaremos varianza condicional de Y condicional \mathbf{X} a la variable aleatoria

$$\text{Var}(Y|\mathbf{X}) = q(\mathbf{X}) = E((Y - E(Y|\mathbf{X}))^2|\mathbf{X}). \quad (9.14)$$

Desarrollando el cuadrado en (9.14) y utilizando la Propiedad 3 ' se obtiene

$$\begin{aligned} \text{Var}(Y|\mathbf{X}) &= E([Y^2 + E^2(Y|\mathbf{X}) - 2YE(Y|\mathbf{X})]|\mathbf{X}) \\ &= E(Y^2|\mathbf{X}) + E^2(Y|\mathbf{X}) - 2E(Y|\mathbf{X})E(Y|\mathbf{X}) \\ &= E(Y^2|\mathbf{X}) - E^2(Y|\mathbf{X}). \end{aligned}$$

El siguiente Teorema vincula la varianza condicional al error cuadrático medio del predictor óptimo no lineal $\hat{Y}_{O,NL} = E(Y|\mathbf{X})$.

Teorema 7. Supongamos que Y es una variable aleatoria con varianza finita, \mathbf{X} un vector aleatorio, y sea $\hat{Y}_{O,NL} = E(Y|\mathbf{X})$, el mejor predictor no lineal de Y basado en \mathbf{X} . Luego se tiene

- (i) $\text{ECM}(\hat{Y}_{O,NL}, Y) = E(\text{Var}(Y|\mathbf{X}))$.
- (ii) $E(\text{Var}(Y|\mathbf{X})) \leq \text{Var}(Y)$.
- (iii) $E(\text{Var}(Y|\mathbf{X})) = \text{Var}(Y)$ si y solo si $P(E(Y|\mathbf{X}) = E(Y)) = 1$.

Demostración. Aplicando el Teorema 1 y utilizando la definición (9.14) se tiene

$$\begin{aligned} \text{ECM}(\hat{Y}_{O,NL}, Y) &= E((Y - E(Y|\mathbf{X}))^2) \\ &= E(E((Y - E(Y|\mathbf{X}))^2|\mathbf{X})) \\ &= E(\text{Var}(Y|\mathbf{X})), \end{aligned}$$

y por lo tanto queda demostrado parte (i) del Teorema.

Como $\hat{Y}_{O,NL}$ es el predictor con menor error cuadrático medio en la clase de predictores $\mathcal{P} = \{\hat{Y} : \hat{Y} = t(\mathbf{X}), \text{Var}(t(\mathbf{X})) < \infty\}$, y como el predictor óptimo constante $\hat{Y}_{O,C} = E(Y) \in \mathcal{P}$, se tiene

$$\begin{aligned} E(\text{Var}(Y|\mathbf{X})) &= \text{ECM}(\hat{Y}_{O,NL}, Y) \\ &\leq \text{ECM}(\hat{Y}_{O,C}, Y) \\ &= E((Y - E(Y))^2) \\ &= \text{Var}(Y) \end{aligned}$$

y por un teorema anterior la igualdad vale si y solo si $P(\hat{Y}_{O,NL} = \hat{Y}_{O,C}) = 1$. \square

Capítulo 10

Convergencia de Variables Aleatorias.

10.1. Convergencia de funciones

En primer lugar recordaremos algunos tipos de convergencia en espacio de funciones.

Convergencia Puntual.

Sea $\{f_n\}_{n \in \mathbb{N}}$ una sucesión de funciones definidas sobre un conjunto Ω y que toma valores reales. Se dice que f_n *converge puntualmente* a otra función f si para todo $x \in \Omega$ y para todo $\varepsilon > 0$, existe $n_0 \in \mathbb{N}$ dependiendo de ε y de x tal que si $n \geq n_0$ entonces $|f_n(x) - f(x)| < \varepsilon$.

Observación.

En general n_0 depende de ε y x , es decir $n_0 = n_0(x, \varepsilon)$. Cuando la elección de n_0 puede hacerse con independencia de x , se tiene la siguiente noción de convergencia.

Convergencia Uniforme.

Sea $\{f_n\}_{n \in \mathbb{N}}$ una sucesión de funciones definidas sobre un conjunto Ω y que toma valores reales. Se dice que f_n *converge uniformemente en Ω* a otra función f si para todo $\varepsilon > 0$, existe $n_0 \in \mathbb{N}$ tal que si $n \geq n_0$ entonces $|f_n(x) - f(x)| < \varepsilon$ para todo $x \in \Omega$.

Observación.

Es inmediato ver que si $\{f_n\}_{n \in \mathbb{N}}$ converge uniformemente en Ω entonces $\{f_n\}_{n \in \mathbb{N}}$ converge puntualmente. La recíproca es falsa. Por ejemplo si definimos $f_n(x) = x^n$ para $x \in [0, 1]$ entonces la sucesión converge puntualmente a la función

$$f(x) = \begin{cases} 0 & \text{si } 0 \leq x < 1 \\ 1 & \text{si } x = 1 \end{cases}$$

para todo $x \in [0, 1]$ pero no converge uniformemente en $[0, 1]$.

Veremos ahora algunos tipos de convergencia para variables aleatorias que hacen uso de la estructura del espacio de probabilidades

Existen varios tipos de convergencia, pero en este curso consideraremos sólo dos: convergencia casi segura y convergencia en probabilidad.

10.2. Convergencia casi segura y en probabilidad.

Consideremos un espacio de probabilidades (Ω, \mathcal{A}, P) . Sea $\{X_n\}_{n \in \mathbb{N}}$ una sucesión de variables aleatorias definidas sobre este espacio y X otra variable aleatoria también definida sobre el mismo espacio.

Definición. Diremos que $\{X_n\}_{n \in \mathbb{N}}$ converge casi seguramente a X ($X_n \rightarrow X$ c.s.) si

$$P(\{\omega : X_n(\omega) \rightarrow X(\omega)\}) = 1. \quad (10.1)$$

Observación.

En teoría de la medida, este tipo de convergencia se denomina convergencia en casi todo punto y se la nota $X_n \rightarrow X$ p.p.

Definición. Diremos que $\{X_n\}_{n \in \mathbb{N}}$ converge en probabilidad a X si para todo $\varepsilon > 0$ se tiene

$$\lim_{n \rightarrow +\infty} P(\{|X_n - X| \geq \varepsilon\}) = 0. \quad (10.2)$$

Notación: Si la sucesión de variables aleatorias $\{X_n\}_{n \in \mathbb{N}}$ converge en probabilidad a la variable aleatoria X escribiremos $X_n \rightarrow_p X$

Observaciones.

1. La convergencia en probabilidad significa que fijado $\varepsilon > 0$ la probabilidad que la distancia entre X_n y X se puede hacer menor que ε con probabilidad tan cercana a 1 como se quiera con tal de tomar n suficientemente grande.

2. En Teoría de la Medida la convergencia en probabilidad se denomina convergencia en medida.

Proposición. Sea $\{X_n\}_{n \in \mathbb{N}}$ una sucesión de variables aleatorias definidas sobre un espacio de probabilidad (Ω, \mathcal{A}, P) y X otra variable aleatoria definida sobre el mismo espacio. Son equivalentes

(i) $X_n \rightarrow_p X$

(ii) Para todo $\varepsilon > 0$ y todo $\delta > 0$ existe $n_0 \in \mathbb{N}$ tal que si $n \geq n_0$ entonces

$$\lim_{n \rightarrow +\infty} P(\{|X_n - X| \geq \varepsilon\}) \leq \delta.$$

(iii) Para todo $\varepsilon > 0$, existe $n_0 \in \mathbb{N}$ tal que si $n \geq n_0$ entonces

$$\lim_{n \rightarrow +\infty} P(\{|X_n - X| \geq \varepsilon\}) \leq \varepsilon.$$

Demostración.

(ii) es equivalente a (i) como consecuencia directa de la definición de convergencia en probabilidad. La equivalencia entre (ii) y (iii) se deja como ejercicio \square .

El siguiente Teorema establece que la convergencia casi segura (10.1) implica la convergencia en probabilidad (10.2).

Teorema Sea $\{X_n\}_{n \in \mathbb{N}}$ una sucesión de variables aleatorias definidas sobre un espacio de probabilidad (Ω, \mathcal{A}, P) y X otra variable aleatoria definida sobre el mismo espacio. Entonces

(i) La sucesión X_n converge casi seguramente a X sii

$$\lim_{m \rightarrow \infty} P(\cup_{n=m}^{\infty} \{|X_n - X| \geq \varepsilon\}) = 0. \quad (10.3)$$

(ii) Si X_n converge casi seguramente a X entonces X_n converge en probabilidad a la variable aleatoria X .

Demostración

Llamemos A al conjunto de los puntos ω de Ω donde $X_n(\omega) \rightarrow X(\omega)$. Luego

$$A = \{\omega \in \Omega : X_n(\omega) \rightarrow X(\omega)\}.$$

Decir que $\omega \in A$ es equivalente a decir que para todo $\varepsilon > 0$ existe $m \in \mathbb{N}$ tal que para todo $n \geq m$ se tiene $|X_n(\omega) - X(\omega)| < \varepsilon$. Entonces, si para cada $\varepsilon > 0$ definimos

$$B_{n,\varepsilon} = \{\omega \in \Omega : |X_n(\omega) - X(\omega)| < \varepsilon\}.$$

el conjunto A resulta

$$A = \bigcap_{\varepsilon > 0} \left(\bigcup_{m=1}^{\infty} \bigcap_{n \geq m} B_{n,\varepsilon} \right).$$

En realidad como basta elegir ε tan chico como se quiera, nos podemos limitar a tomar $\varepsilon = 1/k$. Luego también tenemos

$$A = \bigcap_{k=1}^{\infty} \left(\bigcup_{m=1}^{\infty} \bigcap_{n \geq m} B_{n, \frac{1}{k}} \right).$$

Sabemos que la convergencia casi segura se define por $P(A) = 1$ o equivalentemente por $P(A^c) = 0$. Observemos que

$$A^c = \bigcup_{k=1}^{\infty} \left(\bigcap_{m=1}^{\infty} \bigcup_{n \geq m} B_{n, \frac{1}{k}}^c \right).$$

Luego, como A^c es una unión numerable, $P(A^c) = 0$ si y sólo si para todo $k \in \mathbb{N}$ se tiene

$$\mathbb{P} \left(\bigcap_{m=1}^{\infty} \bigcup_{n \geq m} B_{n, \frac{1}{k}}^c \right) = 0,$$

y como $B_{n,\varepsilon}^c$ es crecientemente con ε , esto es equivalente a que para todo $\varepsilon > 0$

$$P\left(\bigcap_{m=1}^{\infty} \bigcup_{n \geq m} B_{n,\varepsilon}^c\right) = 0. \quad (10.4a)$$

Definamos

$$C_{m,\varepsilon} = \bigcup_{n \geq m} B_{n,\varepsilon}^c.$$

Claramente, para todo $\varepsilon > 0$ la sucesión $\{C_{m,\varepsilon}\}_{m \geq 1}$ es monótona no decreciente, de manera que

$$P\left(\bigcap_{m=1}^{\infty} \bigcup_{n \geq m} B_{n,\varepsilon}^c\right) = P\left(\bigcap_{m=1}^{\infty} C_{m,\varepsilon}\right) = \lim_{m \rightarrow \infty} P(C_{m,\varepsilon}).$$

Luego se tendrá que (10.4a) es equivalente a

$$\lim_{m \rightarrow \infty} P(C_{m,\varepsilon}) = 0,$$

y esto es equivalente a

$$\lim_{m \rightarrow \infty} P\left(\bigcup_{n \geq m} B_{n,\varepsilon}^c\right) = 0$$

Pero como

$$B_{n,\varepsilon}^c = \{|X_n - X| \geq \varepsilon\},$$

(i) queda demostrado.

(ii) Supongamos que $X_n \rightarrow X$ c.s. Luego se cumple (10.3) y como

$$\{|X_m - X| \geq \varepsilon\} \subset \bigcup_{n=m}^{\infty} \{|X_n - X| \geq \varepsilon\},$$

resulta

$$\lim_{m \rightarrow \infty} P(\{|X_m - X| \geq \varepsilon\}) = 0.$$

Por lo tanto $X_n \rightarrow_p 0$. \square

Veremos que la recíproca es falsa, incluso puede ocurrir que exista convergencia en probabilidad, pero que el conjunto de los puntos donde haya convergencia sea vacío.

10.3. Preservación de la convergencia por funciones continuas

Los siguientes dos Teoremas muestran que las funciones continuas preservan los dos tipos de convergencia que hemos definido: convergencia en probabilidad y convergencia casi segura.

Teorema. Sea $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ continua y supongamos que las sucesiones de variables aleatorias $(X_n)_{n \geq 1}$, $(Y_n)_{n \geq 1}$ convergen casi seguramente a las variables aleatorias X e Y . Entonces $(g(X_n, Y_n))_{n \geq 1}$ converge casi seguramente a la variable aleatoria $g(X, Y)$.

Observación. La propiedad vale en general para $g : \mathbb{R}^k \rightarrow \mathbb{R}$ continua. Si $(X_n^j)_{n \geq 1} \rightarrow X^j$ c.s para $j = 1, 2, \dots, k$ entonces

$$g(X_n^1, X_n^2, \dots, X_n^k) \rightarrow g(X^1, X^2, \dots, X^k) \text{ c.s.}$$

Demostración.

Sean $A = \{\omega : X_n(\omega) \rightarrow X(\omega)\}$ y $B = \{\omega : Y_n(\omega) \rightarrow Y(\omega)\}$. Como $P(A) = P(B) = 1$, también se tendrá $P(A \cap B) = 1$. En efecto

$$0 \leq P((A \cap B)^c) = P(A^c \cup B^c) \leq P(A^c) + P(B^c) = 0.$$

Ahora si $\omega \in A \cap B$ entonces $X_n(\omega) \rightarrow X(\omega)$ y $Y_n(\omega) \rightarrow Y(\omega)$. Luego, por la continuidad de g se tiene $g(X_n(\omega), Y_n(\omega)) \rightarrow g(X(\omega), Y(\omega))$. Por lo tanto

$$A \cap B \subset \{\omega : g(X_n(\omega), Y_n(\omega)) \rightarrow g(X(\omega), Y(\omega))\},$$

y en consecuencia como

$$1 = P(A \cap B) \leq P(\{\omega : g(X_n(\omega), Y_n(\omega)) \rightarrow g(X(\omega), Y(\omega))\}) \leq 1,$$

se obtiene la tesis \square .

Corolario.

Como $g(x, y) = x + y$ es continua resulta que si $Y_n \rightarrow Y$ c.s y $X_n \rightarrow X$ c.s entonces $X_n + Y_n \rightarrow X + Y$ c.s

Análogamente si $g(x, y) = xy$, resulta que si $Y_n \rightarrow Y$ c.s y $X_n \rightarrow X$ c.s entonces $X_n Y_n \rightarrow XY$ c.s

Además si $g(x, y) = \frac{x}{y}$ entonces si $Y_n \rightarrow Y$ c.s con $P(Y = 0) = 0$ y $X_n \rightarrow X$ c.s entonces $\frac{X_n}{Y_n} \rightarrow \frac{X}{Y}$ c.s.

Demostración.

Inmediata \square

Para demostrar una propiedad similar para la convergencia en probabilidad necesitamos algunos resultados previos. Comenzamos probando que toda variable aleatoria es “acotada en probabilidad”.

Lema 1. Sea X una variable aleatoria. Dado $\varepsilon > 0$ existe K tal que

$$P(|X| \geq K) < \varepsilon.$$

Observación.

Esto dice que los valores de X están dentro de un compacto, con probabilidad tan cercana a uno como se quiera.

Demostración.

Consideremos la sucesión de conjuntos

$$A_n = \{|X| \geq n\}.$$

Esta sucesión es monótona decreciente, es decir, $A_{n+1} \subset A_n$ y además $\bigcap_{n=1}^{\infty} A_n = \emptyset$. Entonces

$$\lim_{n \rightarrow \infty} P(A_n) = 0.$$

Luego, dado $\varepsilon > 0$ existe $n_0 \in \mathbb{N}$ tal que $P(A_{n_0}) < \varepsilon$, es decir

$$P(A_{n_0}) = P(\{|X| \geq n_0\}) < \varepsilon.$$

Luego el lema es cierto tomando $K = n_0$. \square

Probaremos ahora un resultado más fuerte: Sucesiones de variables que convergen en probabilidad están acotadas en probabilidad uniformemente.

Lema 2. Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias que converge en probabilidad a la variable aleatoria X . Entonces dado $\varepsilon > 0$ existe K tal que $P(|X| \geq K) < \varepsilon$ y tal que para todo n

$$P(|X_n| \geq K) < \varepsilon.$$

Observación

Demostración.

En primer lugar podemos hallar, de acuerdo al lema 1, K_0 de forma tal que

$$P(|X| \geq K_0) < \frac{\varepsilon}{2}.$$

Teniendo en cuenta que

$$|X_n| \leq |X_n - X| + |X|,$$

se prueba fácilmente que

$$\{|X_n| \leq K_0 + 1\} \subset \{|X_n - X| \geq 1\} \cup \{|X| \geq K_0\}. \quad (10.5)$$

Debido a que $X_n \rightarrow_p X$ en probabilidad podemos encontrar n_0 tal que si $n \geq n_0$

$$P(|X_n - X| \geq 1) < \frac{\varepsilon}{2}.$$

Tomando probabilidades an ambos miembros de (10.5) obtenemos

$$\begin{aligned} P(\{|X_n| \leq K_0 + 1\}) &\leq P(\{|X_n - X| \geq 1\}) + P(\{|X| \geq K_0\}) \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\ &= \varepsilon \end{aligned}$$

para todo $n \geq n_0$. Además por el Lema 1, para cada i tal que $1 \leq i \leq n_0$, podemos encontrar K_i tal que $P(\{|X_i| \geq K_i\}) \leq \varepsilon$. Luego tomando

$$K = \max \left\{ \max_{1 \leq i \leq n_0} \{K_i\}, K_0 + 1 \right\},$$

se obtiene la tesis. \square

Ahora estamos en condiciones de probar la propiedad de que las funciones continuas conservan la convergencia en probabilidad.

Teorema. Sea $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ continua y supongamos que las sucesiones $(X_n)_{n \geq 1}$, $(Y_n)_{n \geq 1}$ convergen en probabilidad a las variables aleatorias X e Y . Entonces $(g(\bar{X}_n, Y_n))_{n \geq 1}$ converge en probabilidad a la variable aleatoria $g(X, Y)$.

Observación

Vale la misma observación hecha para el caso de la convergencia casi segura en cuanto a que el Teorema vale para funciones continuas definidas en \mathbb{R}^k .

Demostración.

Queremos probar que dado $\varepsilon > 0$ existe $n_0 \in \mathbb{N}$ tal que si $n \geq n_0$

$$P(|g(X_n, Y_n) - g(X, Y)| \geq \varepsilon) < \varepsilon. \quad (10.6)$$

De acuerdo al lema anterior podemos hallar un K tal que simultáneamente

$$\begin{aligned} P(|X_n| \geq K) &< \frac{\varepsilon}{6} \quad \forall n \\ P(|X| \geq K) &< \frac{\varepsilon}{6} \\ P(|Y_n| \geq K) &< \frac{\varepsilon}{6} \quad \forall n \\ P(|Y| \geq K) &< \frac{\varepsilon}{6}. \end{aligned}$$

Esto puede lograrse considerando primero un K_1 que cumpla con las dos primeras, después un K_2 que cumpla con las siguientes y tomando $K = \max\{K_1, K_2\}$.

Sea

$$C = [-K, K] \times [-K, K].$$

Dado que g es continua y C compacto entonces g es uniformemente continua en C . Luego existe $\delta > 0$ tal que si $|x - x'| < \delta$, $|y - y'| < \delta$ y $|x|, |x'|, |y|, |y'| \leq K$ entonces

$$|g(x, y) - g(x', y')| < \varepsilon. \quad (10.7)$$

Por la convergencia en probabilidad existe $n_0 \in \mathbb{N}$ tal que si $n \geq n_0$ entonces

$$P(|X_n - X| \geq \delta) < \frac{\varepsilon}{6} \quad (10.8)$$

$$P(|Y_n - Y| \geq \delta) < \frac{\varepsilon}{6} \quad (10.9)$$

Esto se logra considerando un valor n_1 para la sucesión $(X_n)_{n \geq 1}$, un valor n_2 para la sucesión $(Y_n)_{n \geq 1}$ y luego tomando $n_0 = \max\{n_1, n_2\}$.

Ahora definimos los conjuntos

$$A_{1n} = \{|X_n - X| \geq \delta\}$$

$$A_{2n} = \{|Y_n - Y| \geq \delta\}$$

$$A_{3n} = \{|X_n| \geq K\}$$

$$A_{4n} = \{|Y_n| \geq K\}$$

$$A_{5n} = \{|X| \geq K\}$$

$$A_{6n} = \{|Y| \geq K\}.$$

Si bien A_{4n} , A_{5n} no dependen de n , usamos la notación por conveniencia. Vamos a mostrar que si llamamos

$$B = \bigcup_{i=1}^6 A_{in},$$

entonces

$$\{|g(X_n, Y_n) - g(X, Y)| \geq \varepsilon\} \subset B.$$

Para esto debemos mostrar que en B^c se tiene

$$|g(X_n, Y_n) - g(X, Y)| < \varepsilon. \quad (10.10)$$

En efecto, como

$$B^c = \left(\bigcup_{i=1}^6 A_{in}\right)^c = \bigcap_{i=1}^6 A_{in}^c,$$

resulta que cuando B^c es cierto X_n, X, Y_n, Y están en el compacto C y además $|X_n - X| \leq \delta$ y $|Y_n - Y| \leq \delta$. Luego por (10.7) resulta (10.10).

Luego para todo $n \geq n_0$

$$P(\{|g(X_n, Y_n) - g(X, Y)| \geq \varepsilon\}) \leq P(B) \leq \sum_{i=1}^6 P(A_{in}) < 6 \frac{\varepsilon}{6} = \varepsilon,$$

y el Teorema queda demostrado. \square

Análogamente a lo observado para la convergencia casi segura se tienen los siguientes corolarios.

Corolario.

Tomando $g(x, y) = x + y$, entonces si $Y_n \rightarrow^P Y$ y $X_n \rightarrow^P X$ resulta $X_n + Y_n \rightarrow^P X + Y$.

Si $g(x, y) = xy$, entonces si $Y_n \rightarrow^P Y$ y $X_n \rightarrow^P X$ resulta $X_n Y_n \rightarrow^P XY$

Además si $g(x, y) = \frac{x}{y}$ entonces si $Y_n \rightarrow^P Y$ con $P(Y = 0) = 0$ y $X_n \rightarrow^P X$ resulta $\frac{X_n}{Y_n} \rightarrow^P \frac{X}{Y}$.

Demostración.

Inmediata \square

10.4. Ley débil de los grandes números

Desigualdad de Markov

Sea X una variable aleatoria y g una función no negativa, par y no decreciente en el módulo, esto es si $|x| > |y|$ entonces $g(x) \geq g(y)$. Supongamos además que $g(X)$ tiene esperanza finita, es decir que $E(g(X)) < \infty$.

Entonces

$$P(\{|X| \geq \varepsilon\}) \leq \frac{E(g(X))}{g(\varepsilon)}.$$

Demostración

Consideremos el conjunto $A = \{|X| \geq \varepsilon\}$. Entoces $\{A, A^c\}$ es una partición del espacio muestral Ω . Luego $I_A(x) + I_{A^c}(x) = 1$, y como todas las variables son no negativas y $g(x)$ no decreciente en $|x|$, tenemos

$$\begin{aligned} g(X) &= g(X) I_A(X) + g(X) I_{A^c}(X) \\ &\geq g(X) I_A(X) \\ &\geq \varepsilon I_A(X). \end{aligned}$$

Luego tomando esperanza

$$E(g(X)) \geq g(\varepsilon) E(I_A) = g(\varepsilon) P(\{|X| \geq \varepsilon\}).$$

De esta desigualdad se obtiene inmediatamente el resultado buscado. \square

En particular tomando $g(x) = x^2$ se obtiene la siguiente versión de la *desigualdad de Tchebichev*

$$P(\{|X| \geq \varepsilon\}) \leq \frac{E(X^2)}{\varepsilon^2}.$$

Por otro lado si consideramos la variable aleatoria $X - E(X)$ obtenemos la versión (clásica) de la desigualdad de Tchebichev

$$P(\{|X - E(X)| \geq \varepsilon\}) \leq \frac{E([X - E(X)]^2)}{\varepsilon^2} = \frac{\text{Var}(X)}{\varepsilon^2}.$$

Tomando complementos esta desigualdad puede escribirse como

$$P(\{|X - E(X)| < \varepsilon\}) \geq 1 - \frac{\text{Var}(X)}{\varepsilon^2}.$$

Luego si $\text{Var}(X)$ es pequeña (o sea hay poca dispersión), la probabilidad de que la variable X tome valores en el intervalo $(E(X) - \varepsilon, E(X) + \varepsilon)$ se hará grande.

Ahora estamos en condiciones de estudiar la ley de los grandes números en sus dos versiones: débil y fuerte. La importancia de estas leyes, es que permite dar fundamento matemático a la heurística que interpreta la esperanza de una variable aleatoria como el valor al cual tiende el promedio de varias realizaciones de la variable correspondientes a la repetición de experimentos independientes. También permite fundamentar la noción heurística de la probabilidad de un evento como el valor límite de las frecuencias relativas con que ocurre el evento cuando se repiten muchos experimentos independientes. La ley débil expresa estos resultados en términos de convergencia en probabilidad y la ley fuerte en término de convergencia casi segura.

Teorema (Ley débil de los grandes números). Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias no correlacionadas, es decir $\text{Cov}(X_i, X_j) = 0$ si $i \neq j$, tal que $E(X_i) = \mu_i$ y $\text{Var}(X_i) = \sigma_i^2$ para cada $i = 1, 2, \dots$. Consideramos la sucesión de variables aleatorias $(\bar{X}_n)_{n \geq 1}$ donde \bar{X}_n es el promedio de las primeras n variables. Luego

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i,$$

y sea $\bar{\mu}_n = E(\bar{X}_n)$ dada por

$$\bar{\mu}_n = \frac{1}{n} \sum_{i=1}^n \mu_i.$$

Entonces si

$$\lim_{n \rightarrow \infty} \left(\frac{1}{n^2} \sum_{i=1}^n \sigma_i^2 \right) = 0, \quad (10.11)$$

se tiene

$$\bar{X}_n - \bar{\mu}_n \xrightarrow{P} 0.$$

Demostración

Se tiene que

$$\text{Var}(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^n \sigma_i^2,$$

y por Tchebichev

$$\begin{aligned} P(|\bar{X}_n - \bar{\mu}_n| \geq \varepsilon) &\leq \frac{\text{Var}(\bar{X}_n)}{\varepsilon^2} \\ &= \frac{1}{\varepsilon^2 n^2} \sum_{i=1}^n \sigma_i^2. \end{aligned}$$

Tomando límite resulta que

$$\begin{aligned} \lim_{n \rightarrow \infty} P(|\bar{X}_n - \bar{\mu}_n| \geq \varepsilon) &\leq \frac{1}{\varepsilon^2} \lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{i=1}^n \sigma_i^2, \\ &= 0 \end{aligned}$$

y luego el Teorema queda demostrado. \square

Observaciones.

1 Si X_n es una sucesión de variables aleatorias independientes, entonces las variables X_n son no correlacionadas y el Teorema puede aplicarse.

2. Una condición suficiente para que se cumpla (10.11) es que $\{\sigma_i^2\}$ sea una sucesión acotada. En efecto, si $\sigma_i^2 \leq K$ para todo i , se obtiene

$$\frac{1}{n^2} \sum_{i=1}^n \sigma_i^2 \leq \frac{Kn}{n^2} = \frac{K}{n} \rightarrow 0.$$

En particular, esta condición se cumple si todas las variables tienen igual varianza.

3. Si todas las variables tienen igual media digamos $\mu_i = \mu$, se tiene que $\bar{\mu}_n = \mu$, y entonces $\bar{X}_n - \mu \xrightarrow{P} 0$ es equivalente

$$\bar{X}_n \xrightarrow{p} \mu.$$

4. En particular si $\{X_n\}_{n \geq 1}$ es una sucesión de variables no correlacionadas igualmente distribuidas con $E(X_n) = \mu$ y $\text{Var}(X_n) = \sigma^2$, se tendrá $\bar{X}_n \xrightarrow{p} \mu$.

5. Veremos ahora como esta ley débil permite fundamentar el concepto de probabilidad de un evento. Sea (Ω, \mathcal{A}, P) un espacio de probabilidad y A un evento. Supongamos que realizamos n experimentos independientes y definimos

$$X_i(\omega) = \begin{cases} 1 & \text{si en el experimento } i, \omega \in A \\ 0 & \text{si en el experimento } i, \omega \notin A. \end{cases}$$

Definamos

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Se tiene

$$E(X_i) = 1 \cdot P(A) + 0 \cdot P(A^c) = P(A),$$

y como $X_i^2 = X_i$

$$\begin{aligned} \text{Var}(X_i) &= E(X_i^2) - E(X_i)^2 \\ &= E(X_i) - E(X_i)^2 \\ &= P(A) - P(A)^2 \\ &= P(A)(1 - P(A)). \end{aligned}$$

Luego, como además las variables X_i son independientes, de acuerdo a la ley débil de los grandes números se tendrá

$$\bar{X}_n \xrightarrow{P} E(X_i) = P(A). \quad (10.12)$$

Obsérvese que \bar{X}_n es la frecuencia relativa de ocurrencia del evento A . Entonces (10.12) puede interpretarse como que la frecuencia relativa de ocurrencia del evento A tiende en probabilidad a su probabilidad.

10.5. Ley fuerte de los grandes números

Para probar la ley fuerte de los grandes números necesitaremos algunos lemas previos

Lema (Desigualdad de Kolmogorov). Sean X_1, \dots, X_n variables independientes con $E(X_i) = 0$. Supongamos que $\sigma_i^2 = \text{Var}(X_i) < \infty$ y consideremos las sumas parciales $S_i = \sum_{j=1}^i X_j$. Entonces

$$P\left(\max_{1 \leq i \leq n} |S_i| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2} \sum_{i=1}^n \sigma_i^2. \quad (10.13)$$

Observación.

Vamos a mostrar que la desigualdad de Kolmogorov es un refinamiento de la desigualdad de Tchebichev. Para ver esto, apliquemos la desigualdad de Tchebichev a las variable aleatoria S_n . Luego

$$P(|S_n| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \text{Var}(S_n) = \frac{1}{\varepsilon^2} \sum_{i=1}^n \sigma_i^2. \quad (10.14)$$

Observemos que $S_n \leq \max_{1 \leq i \leq n} |S_i|$ de manera que

$$\{|S_n| \geq \varepsilon\} \subset \{\max_{1 \leq i \leq n} |S_i| \geq \varepsilon\},$$

y por lo tanto

$$P(\{|S_n| \geq \varepsilon\}) \leq P\left(\{\max_{1 \leq i \leq n} |S_i| \geq \varepsilon\}\right).$$

Luego resulta que (10.13) implica (10.14)

Demostración.

Sea

$$A = \{\max_{1 \leq i \leq n} |S_i| \geq \varepsilon\},$$

y consideremos para cada i los conjuntos

$$A_i = \{|S_1| < \varepsilon, |S_2| < \varepsilon, \dots, |S_{i-1}| < \varepsilon, |S_i| \geq \varepsilon\}.$$

Estos eventos son disjuntos de a dos y forman una partición de A . Luego

$$A = \bigcup_{i=1}^n A_i,$$

y por lo tanto se deduce que

$$I_A = \sum_{i=1}^n I_{A_i}.$$

Luego como $S_n^2 I_{A^c} \geq 0$ se deduce que

$$S_n^2 = S_n^2 I_A + S_n^2 I_{A^c} \geq S_n^2 I_A = S_n^2 \sum_{i=1}^n I_{A_i}.$$

Tomando esperanza en ambos miembros resulta

$$E(S_n^2) \geq \sum_{i=1}^n E(S_n^2 I_{A_i}). \quad (10.15)$$

Para cada término $S_n^2 I_{A_i}$ resulta

$$S_n^2 I_{A_i} = (S_i + T_i)^2 I_{A_i} = S_i^2 I_{A_i} + T_i^2 I_{A_i} + 2S_i T_i I_{A_i}, \quad (10.16)$$

donde

$$T_i = \sum_{j \geq i} X_j.$$

Vamos ahora a probar que $E(S_i T_i I_{A_i}) = 0$. Por un lado observamos que S_i depende solo de X_1, \dots, X_i y lo mismo ocurre con I_{A_i} . Como T_i depende solo de X_{i+1}, \dots, X_n , resulta que $S_i I_{A_i}$ es independiente de T_i . Luego como $E(T_i) = 0$ se obtiene

$$E(S_i T_i I_{A_i}) = E([S_i I_{A_i}] T_i) = E(S_i I_{A_i}) E(T_i) = 0. \quad (10.17)$$

Tomando esperanza en (10.16) y teniendo en cuenta (10.17) y que en A_i se tiene $|S_i| \geq \varepsilon$

$$\begin{aligned} E(S_n^2 I_{A_i}) &= E(S_i^2 I_{A_i}) + E(T_i^2 I_{A_i}) \\ &\geq E(S_i^2 I_{A_i}) \\ &\geq \varepsilon E(I_{A_i}) \\ &= \varepsilon P(A_i). \end{aligned}$$

Luego tomando esperanzas en (10.15) se obtiene

$$E(S_n^2) \geq \sum_{i=1}^n E(S_n^2 I_{A_i}) \geq \varepsilon^2 \sum_{i=1}^n P(A_i) = \varepsilon^2 P(A),$$

o sea

$$P(A) \leq \frac{E(S_n^2)}{\varepsilon^2} = \frac{1}{\varepsilon^2} \sum_{i=1}^n \sigma_i^2. \quad \square$$

Para probar la ley fuerte de los grandes números necesitamos también el siguiente Lema.

Lema. Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias. Una condición suficiente para que

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

es que para todo $\varepsilon > 0$ exista una sucesión creciente de enteros positivos $r_1 < r_2 < \dots < r_n \dots$ que puede depender de ε tal que

$$\sum_{i=1}^{\infty} P\left(\bigcup_{n=r_i}^{r_{i+1}-1} B_{n\varepsilon}^c\right) < \infty, \quad (10.18)$$

donde $B_{n\varepsilon} = \{|X_n - X| < \varepsilon\}$.

Demostración

Recordemos el resultado ya probado que

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

sii

$$\lim_{m \rightarrow \infty} P\left(\bigcup_{n \geq m} B_{n\varepsilon}^c\right) = 0. \quad (10.19)$$

Supongamos que se cumple (10.18). Veremos que entonces se cumple (10.19)

Sea $\varepsilon > 0$, entonces (10.18) implica que existe i_0 tal que

$$\sum_{i=i_0}^{\infty} P \left(\bigcup_{n=r_i}^{r_{i+1}-1} B_{n\varepsilon}^c \right) < \varepsilon$$

Pero entonces

$$P \left(\bigcup_{n=r_{i_0}}^{\infty} B_{n\varepsilon}^c \right) = P \left(\bigcup_{i=i_0}^{\infty} \bigcup_{n=r_i}^{r_{i+1}-1} B_{n\varepsilon}^c \right) \leq \sum_{i=i_0}^{\infty} P \left(\bigcup_{n=r_i}^{r_{i+1}-1} B_{n\varepsilon}^c \right) < \varepsilon.$$

Esto implica que (10.19) se cumple. \square

Teorema (Ley fuerte de los grandes números).

Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias independientes tal que $E(X_i) = \mu_i$ y $\text{Var}(X_i) = \sigma_i^2$ para cada $i \in \mathbb{N}$. Consideremos la sucesión de variables aleatorias $(\bar{X}_n)_{n \geq 1}$ definida por

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

y sus respectivas medias

$$\bar{\mu}_n = E(\bar{X}_n) = \frac{1}{n} \sum_{i=1}^n \mu_i.$$

Entonces si

$$\sum_{i=1}^{\infty} \frac{\sigma_i^2}{i^2} < \infty, \quad (10.20)$$

se tiene

$$\bar{X}_n - \bar{\mu}_n \rightarrow 0 \text{ c.s..}$$

Demostración.

Basta probar el teorema suponiendo que para todo i , $\mu_i = 0$. Para ver esto, supongamos que el teorema vale cuando para todo i , $\mu_i = 0$. Ahora supongamos el caso general, esto es que para cada i , $E(X_i) = \mu_i$. Consideremos nuevas variables $Y_i = X_i - \mu_i$. Entonces $E(Y_i) = 0$ y $\text{Var}(Y_i) = \text{var}(X_i) = \sigma_i^2$. Las variables Y_i son independientes y luego se cumple $\bar{Y}_n \rightarrow 0$ c.s. Pero como $\bar{Y}_n = \bar{X}_n - \bar{\mu}_n$, resulta también $\bar{X}_n - \bar{\mu}_n \rightarrow 0$ c.s. Luego para demostrar el teorema podemos supondremos que $\mu_i = 0$ para todo i .

Usaremos el lema anterior tomando $r_i = 2^{i-1}$. Luego si llamamos

$$\lambda_i = P \left(\bigcup_{n=2^i}^{2^{i+1}-1} B_{n\varepsilon}^c \right),$$

bastará demostrar que

$$\sum_{i=1}^{\infty} \lambda_i < \infty.$$

Observemos que si llamamos $S_n = \sum_{i=1}^n X_i$ tenemos $\bar{X}_n = S_n/n$. Luego

$$\begin{aligned} \lambda_i &= P \left(\bigcup_{n=2^i}^{2^{i+1}-1} B_{n\varepsilon}^c \right) \\ &= P \left(\bigcup_{n=2^i}^{2^{i+1}-1} \{|\bar{X}_n| \geq \varepsilon\} \right) \\ &= P \left(\bigcup_{n=2^i}^{2^{i+1}-1} \{|S_n| \geq n\varepsilon\} \right) \leq \\ &\leq P \left(\bigcup_{n=2^i}^{2^{i+1}-1} \{|S_n| \geq 2^i \varepsilon\} \right) \end{aligned} \quad (10.21)$$

Para aplicar la desigualdad de Kolmogorov, debemos reindexar las variables.

Sean

$$\begin{aligned} X_1^* &= S_{2^i} \\ X_2^* &= X_{2^i+1} \\ X_3^* &= X_{2^i+2} \\ &\dots \\ X_{2^i}^* &= X_{2^{i+1}-1} \end{aligned}$$

y definamos

$$S_i^* = X_1^* + X_2^* + \dots + X_i^*$$

entonces se tiene que

$$\begin{aligned} S_{2^i} &= S_1^* \\ S_{2^i+1} &= S_2^* \\ &\dots \\ S_{2^{i+1}-1} &= S_{2^i}^* \end{aligned}$$

Como

$$\text{Var}(X_1^*) = \sum_{i=1}^{2^i} \sigma_i^2, \text{Var}(X_2^*) = \sigma_{2^i+1}^2, \dots, \text{Var}(X_{2^i}^*) = \sigma_{2^{i+1}-1}^2,$$

se tiene

$$\sum_{j=1}^{2^i} \text{Var}(X_j^*) = \sum_{j=1}^{2^{i+1}-1} \sigma_j^2.$$

Luego, aplicando la desigualdad de Kolmogorov obtenemos

$$\begin{aligned} P\left(\bigcup_{n=2^i}^{2^{i+1}-1} \{|S_n| \geq 2^i \varepsilon\}\right) &= P\left(\max_{1 \leq n \leq 2^i} |S_n^*| > \varepsilon 2^i\right) \\ &\leq \frac{1}{4^i \varepsilon^2} \sum_{j=1}^{2^i} \text{Var}(X_j^*) \\ &\leq \frac{1}{4^i \varepsilon^2} \sum_{j=1}^{2^{i+1}-1} \sigma_j^2. \end{aligned} \quad (10.22)$$

Entonces de (10.21) y (10.22) obtenemos para cada i

$$\lambda_i \leq \frac{1}{4^i \varepsilon^2} \sum_{j=1}^{2^{i+1}-1} \sigma_j^2,$$

y cambiando el orden de sumación se tiene

$$\begin{aligned} \sum_{i=1}^{\infty} \lambda_i &\leq \sum_{i=1}^{\infty} \frac{1}{4^i \varepsilon^2} \sum_{j=1}^{2^{i+1}-1} \sigma_j^2 \\ &= \frac{1}{\varepsilon^2} \sum_{j=1}^{\infty} \sigma_j^2 \sum_{2^{i+1}-1 \geq j} \frac{1}{4^i}. \end{aligned} \quad (10.23)$$

La desigualdad $2^{i+1} - 1 \geq j$ es equivalente a

$$i \geq \frac{\log(j+1)}{\log(2)} - 1 = i_0(j),$$

y entonces podemos escribir

$$\begin{aligned} \sum_{2^{i+1}-1 \geq j} \frac{1}{4^i} &= \sum_{i \geq i_0(j)} \frac{1}{4^i} \\ &= a_0 \left(\frac{1}{1 - \frac{1}{4}} \right) \\ &= \frac{4}{3} a_0, \end{aligned} \quad (10.24)$$

donde a_0 es el primer término de la serie geométrica.

$$\sum_{i \geq i_0(j)} \frac{1}{4^i}. \quad (10.25)$$

Por otro lado $2^{i+1} - 1 \geq j$ implica que $4^i \geq j^2/4$, es decir para todos los términos de la serie geométrica (10.25) obtenemos

$$\frac{1}{4^i} \leq \frac{4}{j^2},$$

y en particular se tendrá

$$a_0 \leq \frac{4}{j^2}. \quad (10.26)$$

Entonces por (10.24 y (10.26) se tiene

$$\sum_{2^{i+1}-1 \geq j} \frac{1}{4^i} = \frac{4}{3} a_0 \leq \frac{4}{3} \frac{4}{j^2} = \frac{16}{3} \frac{1}{j^2},$$

y de acuerdo a (10.23) se tiene

$$\sum_{i=1}^{\infty} \lambda_i \leq \frac{16}{3\varepsilon^2} \sum_{j=1}^{\infty} \frac{\sigma_j^2}{j^2} < \infty.$$

Esto prueba la Ley Fuerte de los Grandes Números. \square

Observaciones.

La condición (10.20) se cumple si todas las varianzas están acotadas. En efecto, si existe una constante K tal que para todo i , $\sigma_i^2 \leq K$ entonces como se tiene

$$\sum_{i=1}^{\infty} \frac{1}{i^2} < \infty,$$

resulta

$$\sum_{i=1}^{\infty} \frac{\sigma_i^2}{i^2} \leq K \sum_{i=1}^{\infty} \frac{1}{i^2} < \infty.$$

Para el caso en que para todo i , $\mu_i = \mu$, $\sigma_i^2 = \sigma^2$ se cumple efectivamente que

$$\sigma^2 \sum_{i=1}^{\infty} \frac{1}{i^2} < \infty,$$

y por lo tanto

$$\overline{X}_n - \mu \rightarrow 0 \text{ c.s.},$$

o equivalentemente

$$\overline{X}_n \rightarrow \mu \text{ c.s..}$$

Todas las consideraciones posteriores a la ley debil que discuten como esta fundamenta las nociones heurísticas de esperanza de un variable aleatoria y de probabilidad de un evento siguen valiendo, reemplazando la convergencia en probabilidad por convergencia casi segura.

10.6. Teorema de la Convergencia Dominada

Ahora daremos una demostración del Teorema de Convergencia Dominada (Lebesgue).

Antes necesitamos el siguiente caso particular de convergencia dominada

Lema.

Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias, no negativas y Z una variable aleatoria no negativa con $E(Z) < \infty$ que domina todos los términos de la sucesión, es decir $0 \leq X_n \leq Z$.

Entonces si $X_n \xrightarrow{P} 0$ se tiene

$$E(X_n) \rightarrow 0.$$

Observación

Recordemos que si $Z \geq 0$ la condición de $E(Z) < \infty$ es equivalente a $\int_0^{+\infty} z dF_Z < \infty$ y esto es equivalente a $\lim_{k \rightarrow \infty} \int_k^{+\infty} z dF_Z = 0$.

Demostración

Vamos a demostrar que dado $\varepsilon > 0$ existe n_0 tal que si $n \geq n_0$ entonces $E(X_n) < \varepsilon$.

Dado $K > 0$ (arbitrario) particionamos al espacio de la siguiente manera

$$\Omega = \left\{ X_n \leq \frac{\varepsilon}{3} \right\} \cup \left\{ \frac{\varepsilon}{3} < X_n \leq K \right\} \cup \{ X_n > K \}.$$

Entonces

$$\begin{aligned} X_n &= X_n I_{\{X_n \leq \varepsilon/3\}} + X_n I_{\{\varepsilon/3 < X_n \leq K\}} + X_n I_{\{X_n > K\}} \\ &\leq \frac{\varepsilon}{3} + K I_{\{X_n > \varepsilon/3\}} + Z I_{\{Z > K\}}. \end{aligned} \quad (10.27)$$

Tomando esperanza en ambos miembros se tiene

$$E(X_n) \leq \frac{\varepsilon}{3} + KP \left(X_n > \frac{\varepsilon}{3} \right) + E(Z I_{\{Z > K\}}). \quad (10.28)$$

Se deja como ejercicio probar que

$$E(Z I_{\{Z > K\}}) = \int_K^{+\infty} z dF_Z. \quad (10.29)$$

Dado que $E(Z) < \infty$ existe K_0 tal que

$$E(Z I_{\{Z > K_0\}}) < \frac{\varepsilon}{3}. \quad (10.30)$$

Una vez elegido K_0 , usando que $X_n \xrightarrow{P} 0$, podemos encontrar n_0 tal que para todo $n \geq n_0$ se tiene

$$P \left(X_n > \frac{\varepsilon}{3} \right) < \frac{\varepsilon}{3K_0}.$$

Luego de (10.27), (10.28), (10.29) y (10.30) resulta que para todo $n \geq n_0$

$$E(X_n) \leq \frac{\varepsilon}{3} + K_0 \frac{\varepsilon}{3K_0} + \frac{\varepsilon}{3} = \varepsilon,$$

y el Lema queda demostrado. \square

Ahora probaremos el Teorema de la Convergencia Dominada en el caso general

Teorema. Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias tal que existe un variable $Z \geq 0$ con $E(Z) < \infty$ y $|X_n| \leq Z$ para todo n . Entonces si $X_n \xrightarrow{P} X$ se tendrá

$$E(X_n) \rightarrow E(X).$$

Demostración

Debemos probar que

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

Ahora bien, por una propiedad de la esperanza

$$|E(X_n) - E(X)| \leq |E(X_n - X)| \leq E(|X_n - X|),$$

de manera que bastará con probar que

$$\lim_{n \rightarrow \infty} E(|X_n - X|) = 0. \quad (10.31)$$

Sea

$$Y_n = |X_n - X| \geq 0,$$

luego como $X_n \xrightarrow{P} X$ resulta $Y_n \xrightarrow{P} 0$. Se deja como ejercicio probar que $P(|X| \leq Z) = 1$. Luego $Y_n \leq |X_n| + |X| \leq 2Z$, y estamos en la situación del lema anterior. Por lo tanto podemos concluir que $E(Y_n) \rightarrow 0$. Luego (10.31) se cumple y el Teorema queda demostrado \square

Capítulo 11

Convergencia en Distribución

11.1. Definición de convergencia en distribución

Tanto la convergencia casi segura como la convergencia en probabilidad se basan en el concepto de "proximidad entre variables aleatorias". Veremos ahora un tipo de convergencia que se basa en la proximidad entre funciones de distribución.

Definición. Sea $\{F_n\}_{n \geq 1}$ una sucesión de funciones de distribución definidas sobre \mathbb{R} y F otra función de distribución. Diremos que la sucesión F_n converge débilmente a F si para todo punto x de continuidad de F existe convergencia puntual. Es decir, si F es continua en x entonces

$$\lim_{n \rightarrow \infty} F_n(x) = F(x).$$

Notación. Si $\{F_n\}_{n \geq 1}$ converge débilmente en distribución a F escribiremos

$$F_n \rightarrow^d F.$$

Observación

Recordemos que una función de distribución definida sobre \mathbb{R} se caracteriza por las propiedades P1, P2, P3 y P4 y que el conjunto de puntos donde es discontinua es a lo sumo numerable.

Definición. Sea $\{X_n\}_{n \geq 1}$ una sucesión de variables aleatorias y F una función de distribución. Diremos que la sucesión X_n converge en distribución a F si $\{F_{X_n}\}_{n \geq 1}$ converge débilmente a F .

Notación. Si $\{X_n\}_{n \geq 1}$ converge en distribución a F escribiremos

$$X_n \rightarrow^D F.$$

Observación

Por extensión también diremos que $\{X_n\}_{n \geq 1}$ converge en distribución a X si $F_{X_n} \rightarrow^d F_X$.

Al decir que $\{X_n\}_{n \geq 1}$ convergen en distribución a X hay un abuso de lenguaje puesto que las variables X_n no se aproximan X , sino la función de distribución de X_n se aproxima a la función de distribución de X .

Consideremos el caso donde X e Y son dos variables independientes con distribución $N(0, 1)$. Definamos para todo n , $X_n = X$ entonces $X_n \xrightarrow{D} Y$ y sin embargo como las variables X e Y son independientes, X no se “aproxima” a Y .

Veamos ahora la relación que existe entre la convergencia en probabilidad y la convergencia en distribución.

Proposición.

Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias y X otra variable aleatoria.

Entonces

$$X_n \xrightarrow{P} X$$

implica que

$$X_n \xrightarrow{D} X$$

Demostración

Sea F_X la función de distribución de X y x un punto de continuidad.

La inclusión

$$\{X_n \leq x\} \subset \{X \leq x + \varepsilon\} \cup \{|X_n - X| \geq \varepsilon\}$$

es fácil de verificar y se deja como ejercicio. Tomado probabilidades en ambos miembros se obtiene

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

Tomando límite superior en ambos miembros y teniendo en cuenta que

$$\lim_{n \rightarrow \infty} P(|X_n - X| \geq \varepsilon) = 0 \tag{11.1}$$

se obtiene

$$\overline{\lim}_{n \rightarrow \infty} F_{X_n}(x) \leq F_X(x + \varepsilon),$$

y haciendo que $\varepsilon \rightarrow 0$, en virtud de la continuidad de F en x se tiene

$$\overline{\lim}_{n \rightarrow \infty} F_{X_n}(x) \leq F_X(x). \tag{11.2}$$

Ahora hacemos algo similar a izquierda de x . Consideramos la inclusión

$$\{X \leq x - \varepsilon\} \subset \{X_n \leq x\} \cup \{|X_n - X| \geq \varepsilon\}.$$

Tomado probabilidades en ambos miembros se obtiene

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

Tomando límite inferior en ambos miembros y usando (11.1) se obtiene

$$F(x - \varepsilon) \leq \underline{\lim}_{n \rightarrow \infty} F_{X_n}(x),$$

y haciendo que $\varepsilon \rightarrow 0$, en virtud de la continuidad de F_X en x

$$F(x) \leq \underline{\lim}_{n \rightarrow \infty} F_{X_n}(x). \quad (11.3)$$

De (11.2) y (11.3) resulta

$$\overline{\lim}_{n \rightarrow \infty} F_{X_n}(x) \leq F_X(x) \leq \underline{\lim}_{n \rightarrow \infty} F_{X_n}(x),$$

y como

$$\underline{\lim}_{n \rightarrow \infty} F_{X_n}(x) \leq \overline{\lim}_{n \rightarrow \infty} F_{X_n}(x),$$

debe ser

$$\overline{\lim}_{n \rightarrow \infty} F_{X_n}(x) = \underline{\lim}_{n \rightarrow \infty} F_{X_n}(x) = F_X(x).$$

Luego existe el límite de $\{F_{X_n}\}$ en el punto x y además

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F(x). \quad \square$$

Observación.

La recíproca no vale en general. Pero si en el caso que $P(X = C) = 1$, donde C es una constante. Es decir, si $X_n \xrightarrow{D} C$, entonces $X_n \xrightarrow{P} C$. Dejamos la prueba como ejercicio.

11.2. Funciones característica

Una herramienta muy importante para la demostración del Teorema Central del Límite es, la función característica asociada a una distribución. Para esto necesitaremos el concepto de variable aleatoria compleja.

11.2.1. Variables aleatorias complejas

Definición: Sea (Ω, \mathcal{A}, P) un espacio de probabilidad. Se dice que X es una *variable aleatoria compleja* si $X : \Omega \rightarrow \mathbb{C}$ (\mathbb{C} indica el conjunto de números complejos) es de la forma $X = X_1 + iX_2$ con X_1 y X_2 variables aleatorias reales.

Definición. Sea la variable aleatoria compleja $X = X_1 + iX_2$, donde X_1 y X_2 tienen esperanza finita. Definimos la *esperanza de X* como $E(X) = E(X_1) + iE(X_2)$.

Observación

$E(X) \in \mathbb{C}$. La parte real e imaginaria de la esperanza son respectivamente $\text{Re}(E(X)) = E(X_1)$ y $\text{Im} E(X) = E(X_2)$.

Definición. Diremos que dos variables aleatorias complejas $X = X_1 + iX_2$ e $Y = Y_1 + iY_2$ son independientes si el vector aleatorio $\mathbf{X} = (X_1, X_2)$ es independiente del vector aleatorio $\mathbf{Y} = (Y_1, Y_2)$.

Algunas propiedades

Veamos ahora que para variables aleatorias complejas independientes se tiene el siguiente resultado esperado

P1. Sean $X = X_1 + iX_2$ e $Y = Y_1 + iY_2$ dos variables aleatorias complejas independientes. Entonces

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

Demostración

La demostración se basa en el cálculo directo usando la definición y la propiedad análoga para variables aleatorias reales independientes

$$\begin{aligned} E(XY) &= E[(X_1 + iX_2)(Y_1 + iY_2)] \\ &= E[(X_1Y_1 - X_2Y_2) + i(X_2Y_1 + Y_2X_1)] \\ &= E(X_1Y_1 - X_2Y_2) + iE(X_2Y_1 + Y_2X_1) = \\ &= E(X_1Y_1) - E(X_2Y_2) + iE(X_2Y_1) + iE(Y_2X_1) = \\ &= E(X_1)E(Y_1) - E(X_2)E(Y_2) + iE(X_2)E(Y_1) + iE(Y_2)E(X_1) \\ &= (E(X_1) + iE(X_2))(E(Y_1) + iE(Y_2)) \\ &= E(X)E(Y). \end{aligned}$$

P2. Sea una variable compleja $X = X_1 + iX_2$. Entonces

$$|E(X)| \leq E(|X|).$$

Demostración

Podemos suponer que $E(X) \neq 0$ pues en tal caso la desigualdad se cumple.

Siendo $E(X) = E(X_1) + iE(X_2) \in \mathbb{C}$ podemos escribir

$$E(X) = re^{i\theta}$$

para cierto $r > 0$.

Consideremos la variable aleatoria compleja $Y = e^{-i\theta}X$ y verifiquemos que su esperanza es real

$$\begin{aligned} E(Y) &= E(e^{-i\theta}X) \\ &= e^{-i\theta}E(X) \\ &= r > 0. \end{aligned}$$

Hemos probado con anterioridad que la propiedad se cumple para esperanzas de variables aleatorias reales. Luego

$$|E(Y)| \leq E(|Y|).$$

De acá se deduce la tesis, pues

$$\begin{aligned}
 |E(X)| &= r \\
 &= E(Y) \\
 &= |E(Y)| \\
 &\leq E(|Y|) \\
 &= E(|X|). \quad \square
 \end{aligned}$$

11.2.2. Definición de función característica y propiedades

Definición Sea X una variable aleatoria y F_X su función de distribución. Definimos la *función característica de X* por la función $\phi_X : \mathbb{R} \rightarrow \mathbb{C}$ asociada a F_X de la siguiente manera

$$\begin{aligned}
 \phi_X(t) &= E(\exp(itX)) = \\
 &= E(\cos(tX) + iE(\sin(tX))).
 \end{aligned}$$

Observación

Como las variables $\cos(tX)$, $\sin(tX)$ son acotadas, las esperanzas de estas variables existen y son finitas.

El motivo de la introducción de la función característica es poder estudiar más fácilmente la distribución de suma de variables aleatorias independientes. Mientras que la función de distribución de esta suma (que se obtiene por convoluciones) puede ser muy complicada, su función característica, como se desprende de la propiedad P3 que veremos a continuación, es muy simple. Por otro lado como veremos mas adelante hay una correspondencia biunívoca entre función de distribución y función característica. Luego conociendo la función característica, de una variable aleatoria, se puede obtener su función de distribución.

P3. Sean X e Y dos variables aleatorias independientes. Entonces para todo $t \in \mathbb{R}$

$$\phi_{X+Y}(t) = \phi_X(t) \phi_Y(t).$$

Demostración.

Observando que $\exp(itX)$, $\exp(itY)$ son variables aleatorias independientes se tiene

$$\begin{aligned}
 \phi_{X+Y}(t) &= E(\exp(it(X+Y))) \\
 &= E(\exp(itX) \exp(itY)) = \\
 &= E(\exp(itX)) E(\exp(itY)) \\
 &= \phi_X(t) \phi_Y(t). \quad \square
 \end{aligned}$$

P4. Sea X una variable aleatoria. Entonces

$$|\phi_X| \leq 1$$

Demostración.

$$\begin{aligned} |\phi_X| &= |E(\exp(itX))| \\ &\leq E(|\exp(itX)|) \\ &= E(1) \\ &= 1. \quad \square \end{aligned}$$

P4. $\phi_X(0) = E(1) = 1$. La demostración es inmediata

Ahora enunciamos dos teoremas básicos que usaremos más adelante. El lector puede encontrar la prueba de estos teoremas en el libro de Barry R. James, "Probabilidad: un curso em nivel intermediário". En particular la interpretación de ϕ_X como la transformada de Fourier de F_X y la fórmula clásica de inversión.

Teorema 1. Sean X e Y dos variables aleatorias.

Entonces si

$$\phi_X = \phi_Y,$$

también se tiene

$$F_X = F_Y.$$

Teorema 2. (Continuidad de Paul Levy). Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias, $(F_{X_n})_{n \geq 1}$ la correspondiente sucesión de funciones de distribución y $(\phi_{X_n})_{n \geq 1}$ la correspondiente sucesión de funciones características asociadas. Entonces

$$F_{X_n} \rightarrow^D F_X$$

si y sólo si para todo $t \in \mathbb{R}$

$$\phi_{X_n}(t) \rightarrow \phi(t).$$

Teorema 3. Sea X una variable aleatoria. Entonces ϕ_X es continua en todo punto.

Demostración.

Sea $t \in \mathbb{R}$ y consideremos una sucesión $(h_n)_{n \geq 1} \subset \mathbb{R}$ tal que $h_n \rightarrow 0$. Queremos probar que

$$\lim_{n \rightarrow \infty} \phi_X(t + h_n) = \phi_X(t).$$

Teniendo en cuenta que

$$\phi_X(t + h_n) = E(\cos((t + h_n)X)) + iE(\sin((t + h_n)X)),$$

bastará con probar que si $n \rightarrow +\infty$ entonces

$$E(\cos((t+h_n)X)) \rightarrow E(\cos(tX)),$$

y

$$E(\sin((t+h_n)X)) \rightarrow E(\sin(tX)).$$

Probaremos que $E(\cos((t+h_n)X)) \rightarrow E(\cos(tX))$ cuando $n \rightarrow +\infty$, la otra propiedad es análoga.

Consideremos la sucesión de variables aleatorias

$$Y_n = \cos((t+h_n)X).$$

Se comprueba fácilmente que Y_n está dominada por la variable aleatoria $Z = 1$, es decir para todo n

$$|Y_n| = |\cos((t+h_n)X)| \leq 1.$$

Además si $Y = \cos(tX)$, por la continuidad de la función coseno, se tiene convergencia puntual de Y_n a Y , es decir para todo $\omega \in \Omega$

$$Y_n(\omega) \rightarrow Y(\omega).$$

Luego, por el Teorema de Convergencia Dominada se obtiene

$$E(Y_n) \rightarrow E(Y). \quad \square$$

Observación.

Se puede probar algo más fuerte: ϕ_X es uniformemente continua (ver el libro de Barry R. James).

Veamos como opera una función característica sobre una transformación afín de la variable aleatoria.

P4. Sea X una variable aleatoria e $Y = aX + b$. Entonces para todo $t \in \mathbb{R}$

$$\phi_{aX+b}(t) = \exp(ibt) \phi_X(at).$$

Demostración.

Para todo $t \in \mathbb{R}$ se tiene

$$\begin{aligned} \phi_Y(t) &= \phi_{aX+b}(t) \\ &= E(\exp(it(aX+b))) \\ &= E(\exp(it(aX)) \exp(itb)) \\ &= \exp(ibt) E(\exp(it(aX))) \\ &= \exp(ibt) \phi_X(at). \quad \square \end{aligned}$$

Ahora queremos caracterizar a las funciones características a valores reales. Para esto recordemos el concepto de variable aleatoria simétrica respecto del origen

Una variable aleatoria X se dice que es simétrica respecto del origen sii para todo $x \geq 0$ se tiene que

$$P(X \leq -x) = P(X \geq x). \quad (11.4)$$

El siguiente Teorema permite dar una definición equivalente

Teorema 4. X es simétrica respecto del origen, si y solo si $F_{-X} = F_X$

Demostración.

Supongamos primero que X satisface $F_{-X} = F_X$. De (11.4) obtenemos

$$P(X \leq -x) = P(-X \leq -x), \quad \forall x \geq 0,$$

o equivalentemente

$$F_X(x) = F_{-X}(x), \quad \forall x \leq 0. \quad (11.5)$$

Consideremos ahora $x > 0$. Como se tiene

$$\{X < -x\} = \cup_{n=1}^{\infty} \{X \leq -x - 1/n\},$$

y

$$\{X > x\} = \cap_{n=1}^{\infty} \{X \geq x + 1/n\},$$

donde $\{X \leq -x - 1/n\}$ y $\{X \geq x + 1/n\}$ son sucesiones monótonas crecientes y decrecientes respectivamente, tendremos

$$P(X < -x) = \lim_{n \rightarrow \infty} P(X \leq -x - 1/n),$$

y

$$P(X > x) = \lim_{n \rightarrow \infty} P(X \geq x + 1/n).$$

Luego de (11.4) resulta que para todo $x > 0$

$$P(X < -x) = P(X > x)$$

y luego

$$1 - P(X < -x) = 1 - P(X > x),$$

o equivalentemente

$$P(X \geq -x) = P(X \leq x),$$

o equivalentemente

$$P(-X \leq x) = P(X \leq x).$$

Por lo tanto

$$F_{-X}(x) = F_X(x), \quad \forall x > 0. \quad (11.6)$$

De (11.5) y (11.6) resulta $F_{-X} = F_X$ para todo x .

Supongamos ahora que $F_{-X} = F_X$. Vamos a probar que X es simétrica respecto de 0, es decir que se cumple (11.4).

Sea $x \geq 0$. Luego como $F_{-X}(-x) = F_X(-x)$, se tendrá

$$P(-X \leq -x) = P(X \leq -x),$$

o equivalentemente

$$P(X \geq x) = P(X \leq -x),$$

y por lo tanto se cumple (11.4) \square .

Teorema 5. ϕ_X es real si X es simétrica respecto del origen. En este caso ϕ_X es par

Demostración.

Supongamos primero que X sea simétrica respecto del origen. Como para todo $t \in \mathbb{R}$

$$\phi_X(t) = E(\cos(tX)) + iE(\sin(tX)),$$

para mostrar que ϕ_X es simétrica bastará ver que $E(\sin(tX)) = 0$.

Teniendo en cuenta que X es simétrica se tiene que $F_X = F_{-X}$ de manera que $E(g(X)) = E(g(-X))$ para cualquier g medible, entonces si para cada $t \in \mathbb{R}$ se toma $g(x) = \sin(tx)$ se obtiene

$$E(\sin(tX)) = E(\sin(-tX)) = -E(\sin(tX)),$$

y por lo tanto $E(\sin(tX)) = 0$.

Además,

$$\begin{aligned} \phi_X(-t) &= E(\cos(X(-t))) \\ &= E(\cos(Xt)) \\ &= \phi_X(t). \end{aligned}$$

Luego ϕ_X es par.

Supongamos ahora que ϕ_X es real, esto es $E(\sin(tX)) = 0$. Entonces teniendo en cuenta que la función coseno es par y la función seno impar tendremos para todo $t \in \mathbb{R}$

$$\begin{aligned} \phi_X(t) &= E(\cos(tX)) + iE(\sin(tX)) \\ &= E(\cos(tX)), \end{aligned}$$

y

$$\begin{aligned} \phi_{-X}(t) &= E(\cos(t(-X))) + iE(\sin(t(-X))) \\ &= E(\cos(tX)) - iE(\sin(tX)) \\ &= E(\cos(tX)) \end{aligned}$$

Luego $\phi_X(t) = \phi_{-X}(t)$ y entonces por el Teorema 1, se obtiene que $F_X = F_{-X}$ y por el Teorema 4 que X es simétrica respecto del origen \square

Definición. Momentos de orden k . Sea X una variable aleatoria.

Definimos el *momento de orden $k > 0$ de X* como el número

$$\mu_k = E\left(X^k\right),$$

cuando este valor existe y el *momento absoluto de orden $k > 0$ de X* como el número

$$\mu_k^* = E\left(|X|^k\right).$$

Observación.

Si k es par entonces $\mu_k = \mu_k^*$. Además siempre se tiene que $\mu_k < \infty$ sii $\mu_k^* < \infty$, es decir la integrabilidad absoluta de $|X|^k$ equivale a la de X^k .

En particular $E(X) = \mu$ y $\text{Var}(X) = \mu_2 - \mu_1^2$.

Teorema. Si $\mu_k^* < \infty$ entonces para todo $i < k$ se tiene $\mu_i^* < \infty$.

Demostración.

Sea $i < k$. Se tiene

$$|X|^i = I_{\{|X| \leq 1\}} |X|^i + I_{\{|X| > 1\}} |X|^i.$$

Como

$$I_{\{|X| \leq 1\}} |X|^i \leq I_{\{|X| \leq 1\}}$$

y

$$I_{\{|X| > 1\}} |X|^i \leq |X|^i \leq |X|^k,$$

obtenemos

$$|X|^i \leq I_{\{|X| \leq 1\}} + |X|^k.$$

Tomando esperanza en ambos miembros resulta

$$\mu_i^* \leq P(\{|X| \leq 1\}) + \mu_k^* < \infty,$$

y esto demuestra el Teorema. \square

11.3. Momentos y función característica

11.3.1. Derivación dentro del signo esperanza

Para hacer un desarrollo de Taylor de la función característica, necesitaremos sus derivadas. Como la función característica está definida como una esperanza, será conveniente encontrar condiciones bajo las cuales se pueda intercambiar el orden en el que se deriva y se toma esperanza.

Sea $g(x, t)$ una función de dos variables a valores reales, medible respecto de la primera variable y derivable respecto de la segunda variable. Sea g_2 definida por

$$g_2(x, t) = \frac{\partial g(x, t)}{\partial t}.$$

Sea X una variable aleatoria, entonces para cada t , $Y_t = g(X, t)$ es también una variable aleatoria. Supongamos que $E(|Y_t|) < \infty$ y consideremos la función $h(t) = E(Y_t) = E(g(X, t))$. El siguiente Teorema nos da condiciones suficientes para que $h'(t) = E(g_2(X, t))$.

Teorema. Supongamos que en $t = t_0$ se cumplan las siguientes condiciones (i) existe $\varepsilon > 0$ tal que

$$\sup_{|t-t_0| \leq \varepsilon} \{ |g_2(X, t)| \} \leq Z,$$

donde $E(Z) < \infty$, (ii) Para todo x la función $g_2(x, t)$ es continua respecto a la segunda variable en $t = t_0$. Luego $h'(t_0) = E(g_2(X, t_0))$.

Demostración.

Sea $(r_n)_{n \geq 1}$ una sucesión de números reales no creciente que converge a 0 y tal que $|r_n| \leq \varepsilon$. Bastará demostrar que

$$\lim_{n \rightarrow +\infty} \frac{h(t_0 + r_n) - h(t_0)}{r_n} = E(g_2(X, t_0)).$$

Utilizando el teorema del valor medio existe $r_n^* = r_n^*(X)$ tal que $|r_n^*(X)| \leq r_n \leq \varepsilon$ y tal que

$$\frac{g(X, t_0 + r_n) - g(X, t_0)}{r_n} = g_2(X, t_0 + r_n^*(X)).$$

Luego

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{h(t_0 + r_n) - h(t_0)}{r_n} &= \lim_{n \rightarrow \infty} E \left(\frac{g(X, t_0 + r_n) - g(X, t_0)}{r_n} \right) \\ &= \lim_{n \rightarrow \infty} E(g_2(X, t_0 + r_n^*(X))). \end{aligned}$$

Por lo tanto bastará con mostrar que

$$\lim_{n \rightarrow \infty} E(g_2(X, t_0 + r_n^*(X))) = E(g_2(X, t_0)). \quad (11.7)$$

Ahora bien $r_n^*(X) \rightarrow 0$ y por la continuidad de g_2 en $t = t_0$, $\{g_2(X, t_0 + r_n^*(X))\}_{n \geq 1}$ converge puntualmente a la función $g_2(X, t_0)$. Además se cumple que $\sup_{n \in \mathbb{N}} |g_2(X, t_0 + r_n^*(X))| \leq Z$, con $E(Z) < \infty$. Luego aplicando el teorema de la convergencia dominada se obtiene (11.7). \square

11.3.2. Derivadas de la función característica y momentos

Dada una variable aleatoria X , sabemos que $\phi_X(t) = E(\exp(itX))$. Procedamos de manera ingenua, sin preocuparnos por la justificación, y derivemos sucesivamente dentro del signo esperanza

$$\begin{aligned}\phi_X^{(1)}(t) &= E(iX \exp(itX)) = iE(X \exp(itX)) \\ \phi_X^{(2)}(t) &= E(i^2 X^2 \exp(itX)) = i^2 E(X^2 \exp(itX)) \\ &\dots \\ \phi_X^{(n)}(t) &= E(i^n X^n \exp(itX)) = i^n E(X^n \exp(itX)).\end{aligned}$$

El siguiente Teorema permite justificar estas expresiones.

Teorema. Supongamos que $\mu_n^* < \infty$. Luego se cumple que

$$\phi_X^{(n)}(t) = i^n E(X^n \exp(itX)). \quad (11.8)$$

Demostración. Demostraremos el Teorema por inducción en n . Para $n = 0$ es cierto ya que $\phi_X(t) = E \exp(itX)$ por definición. Supongamos que el Teorema es cierto para n . Vamos a demostrar que es cierto para $n + 1$. Supongamos que $\mu_{n+1}^* < \infty$, luego por el Teorema anterior $\mu_n^* < \infty$ y luego la fórmula (11.8) es cierta para n . Entonces, tenemos que

$$\begin{aligned}\phi_X^{(n)}(t) &= i^n E(X^n \exp(itX)) \\ &= i^n (E(X^n \cos(tX)) + iE(X^n \sin(tX))).\end{aligned} \quad (11.9)$$

Sea $g(x, t) = x^n \cos(tx)$. Luego $g_2(x, t) = -x^{n+1} \sin(tx)$ es continua y $|g_2(X, t)| \leq |X|^{n+1}$. Como $E(|X|^{n+1}) < \infty$, por el Teorema anterior se tendrá que si $h(t) = E(X^n \cos(tx))$, entonces

$$\begin{aligned}h'(t) &= E(g_2(X, t)) \\ &= -E(X^{n+1} \sin(tX)).\end{aligned} \quad (11.10)$$

Similarmente si $h^*(t) = E(X^n \sin(tx))$, luego

$$h^{*'}(t) = E(X^{n+1} \cos(tX)). \quad (11.11)$$

Luego por (11.10), (11.11), derivando (11.9) se tendrá

$$\phi_X^{(n+1)}(t) = i^n (h'(t) + h^{*'}(t)) \quad (11.12)$$

$$= i^n (E(-X^{n+1} \sin(tX)) + iE(X^{n+1} \cos(tX))). \quad (11.13)$$

Multiplicando por i y dividiendo por i se obtiene

$$\phi_X^{(n+1)}(t) = i^{n+1} ((1/i)E(-X^{n+1} \sin(tX)) + E(X^{n+1} \cos(tX))),$$

y usando que $1/i = -i$

$$\begin{aligned}\phi_X^{(n+1)}(t) &= i^{n+1}(iE(X^{n+1} \sin(tX)) + E(X^{n+1} \cos(tX))) \\ &= i^{n+1}E(X^{n+1} \exp(itX))\end{aligned}$$

y por lo tanto el Teorema queda demostrado. \square .

Observemos entonces que de acuerdo al Teorema anterior que si $\mu_n^* < \infty$ resulta

$$\begin{aligned}\phi_X^{(n)}(0) &= i^n E(X^n) \\ &= i^n \mu_n.\end{aligned}$$

En particular

$$\phi'_X(0) = i\mu_1 \tag{11.14}$$

y

$$\phi''_X(0) = -\mu_2. \tag{11.15}$$

Ahora estamos en condiciones de probar que la función característica de la distribución $X \sim N(0, 1)$ es su densidad salvo una constante.

11.4. Función característica de una distribución normal.

Para la prueba del Teorema Central de Límite, necesitamos calcular la función característica de una distribución normal. Dado que si $X \sim N(\mu, \sigma^2)$ se puede escribir como $X = \sigma Y + \mu$, donde $Y \sim N(0, 1)$ de acuerdo P4, solo se necesitará calcular ϕ_X para el caso $\mu = 0$ y $\sigma^2 = 1$.

Teorema 6. Sea $X \sim N(0, 1)$. Luego la función característica de X es

$$\phi^*(t) = \exp\left(-\frac{1}{2}t^2\right).$$

Demostración. Como X es simétrica respecto del origen, ϕ^* es real y par.

Consideremos dos variables aleatorias independientes $X_1 \sim N(0, 1)$, $X_2 \sim N(0, 1)$ y definamos $Y = u_1 X_1 + u_2 X_2$ con $u_1 \geq 0$, $u_2 \geq 0$. Entonces $Y \sim N(0, u_1^2 + u_2^2)$.

Podemos expresar a Y como un múltiplo de una variable $N(0, 1)$. En efecto

$$\begin{aligned}Y &= \sqrt{u_1^2 + u_2^2} \frac{Y}{\sqrt{u_1^2 + u_2^2}} \\ &= \sqrt{u_1^2 + u_2^2} Z,\end{aligned}$$

donde

$$Z = \frac{Y}{\sqrt{u_1^2 + u_2^2}}$$

tiene distribución $N(0, 1)$.

Calcularemos ϕ_Y de dos manera distintas. Por un lado, usando P4

$$\phi_Y(t) = \phi_{\sqrt{u_1^2 + u_2^2}Z}(t) \quad (11.16)$$

$$= \phi^* \left(\sqrt{u_1^2 + u_2^2} t \right). \quad (11.17)$$

Por otro lado siendo Y suma de variables aleatorias independientes, usando P1 y recordando que $u_1 \geq 0$ y $u_2 \geq 0$, se tiene que

$$\begin{aligned} \phi_Y(t) &= \phi_{u_1 X_1 + u_2 X_2}(t) \\ &= \phi_{u_1 X_1}(t) \phi_{u_2 X_2}(t) \end{aligned} \quad (11.18)$$

$$\begin{aligned} &= \phi^*(u_1 t) \phi^*(u_2 t) \\ &= \phi^* \left(\sqrt{u_1^2} t \right) \phi^* \left(\sqrt{u_2^2} t \right). \end{aligned} \quad (11.19)$$

De (11.16) y (11.19) se obtiene

$$\phi^* \left(\sqrt{u_1^2 + u_2^2} t \right) = \phi^* \left(\sqrt{u_1^2} t \right) \phi^* \left(\sqrt{u_2^2} t \right), \quad (11.20)$$

y haciendo $t = 1$

$$\phi^* \left(\sqrt{u_1^2 + u_2^2} \right) = \phi^* \left(\sqrt{u_1^2} \right) \phi^* \left(\sqrt{u_2^2} \right) \quad (11.21)$$

Definamos g^* como la composición de ϕ^* con la raíz cuadrada, es decir

$$g^*(u) = \phi^*(\sqrt{u}).$$

Luego por (11.21) se tiene

$$g^*(u_1^2 + u_2^2) = g^*(u_1^2) g^*(u_2^2).$$

Luego, si ponemos $v_1 = u_1^2$ y $v_2 = u_2^2$ entonces para todo $v_1, v_2 \geq 0$ obtenemos

$$g^*(v_1 + v_2) = g^*(v_1) g^*(v_2). \quad (11.22)$$

Entonces para todo $v \geq 0$ se tiene

$$\begin{aligned} g^*(v) &= g^* \left(\frac{v}{2} + \frac{v}{2} \right) = \\ &= \left(g^* \left(\frac{v}{2} \right) \right)^2 \geq 0. \end{aligned}$$

Observación. Ecuación (11.22) recuerda la caracterización de la distribución exponencial como una distribución con "falta de memoria". Luego para caracterizar a g^* procederemos de igual manera.

Por inducción se puede probar que dados $v_1 \geq 0, v_2 \geq 0, \dots, v_n \geq 0$ entonces

$$g^* \left(\sum_{i=1}^n v_i \right) = \prod_{i=1}^n g^*(v_i). \quad (11.23)$$

Luego usando (11.23) se obtiene que para todo n natural

$$g^*(n) = g^* \left(\underbrace{1 + 1 + \dots + 1}_{n \text{ veces}} \right) \quad (11.24)$$

$$= [g^*(1)]^n \quad (11.25)$$

Usando (11.23) y (11.24) se obtiene que para todo m y n naturales

$$\begin{aligned} [g^*(1)]^n &= g^*(n) \\ &= g^* \left(m \frac{n}{m} \right) \\ &= g^* \left(\underbrace{\frac{n}{m} + \frac{n}{m} + \dots + \frac{n}{m}}_{m \text{ veces}} \right) \\ &= \left[g^* \left(\frac{n}{m} \right) \right]^m, \end{aligned}$$

y entonces

$$g^* \left(\frac{n}{m} \right) = [g^*(1)]^{\frac{n}{m}}.$$

Luego para todo $r \in \mathbb{Q}$ positivo se tiene

$$g^*(r) = [g^*(1)]^r.$$

Por la continuidad de g^* y la densidad de \mathbb{Q} en \mathbb{R} , se concluye que para todo $x \in \mathbb{R}_{\geq 0}$

$$g^*(x) = [g^*(1)]^x.$$

Ahora veamos que

$$0 < g^*(1) < 1. \quad (11.26)$$

Como $g^*(1)$ es real con $|g^*(1)| \leq 1$ y $g^*(1) \geq 0$, para demostrar (11.26) se deberá mostrar que $g^*(1) \neq 0$ y que $g^*(1) \neq 1$.

Supongamos que $g^*(1) = 0$. Entonces para todo $t \in \mathbb{R}_{\geq 0}$

$$\phi^*(\sqrt{t}) = g^*(t) = [g^*(1)]^t = 0.$$

Esto es absurdo, pues si $t = 0$ se tendría $\phi^*(0) = 0$ y según P3 $\phi^*(0) = 1$. Supongamos que $g^*(1) = 1$ entonces

$$\begin{aligned}\phi^*(1) &= \phi^*(\sqrt{1}) \\ &= g^*(1) \\ &= 1.\end{aligned}$$

Ahora como ϕ^* es real, $\phi^*(1) = E(\cos(X))$. Entonces $g^*(1) = 1$ se puede escribir como

$$E(1) = E(\cos(X))$$

luego

$$E(1 - \cos(X)) = 0$$

Pero siendo la variable aleatoria $1 - \cos(X)$ no negativa se concluye que

$$P(\cos(X) = 1) = 1.$$

Esto no puede ser cierto puesto que $\{x \in R : \cos(x) = 1\}$ es un conjunto de puntos numerable, de manera que como la distribución normal es absolutamente continua, su probabilidad es cero.

Finalmente si ponemos $c = -\log(g^*(1))$ entonces, $c > 0$ y $g^*(1) = \exp(-c)$. Luego

$$\begin{aligned}g^*(t) &= [g^*(1)]^t \\ &= \exp(-ct), \forall t \geq 0.\end{aligned}$$

Ademas

$$\phi^*(t) = g^*(t^2) = \exp(-ct^2), \forall t \geq 0$$

Como la función $\phi^*(t)$ es par se tendrá

$$\phi^*(t) = \exp(-ct^2), \forall t.$$

Derivando dos veces ϕ^*

$$(\phi^*)^{(1)}(t) = -2ct \exp(-ct^2),$$

$$\begin{aligned}(\phi^*)^{(2)}(t) &= -2c \exp(-ct^2) + 4c^2 t^2 \exp(-ct^2) \\ &= 2c \exp(-ct^2) (2ct^2 - 1),\end{aligned}$$

y evaluando en 0, de acuerdo a (11.15) se tendrá

$$\begin{aligned}-2c &= (\phi^*)^{(2)}(0) \\ &= -\mu_2 \\ &= -1.\end{aligned}$$

Por lo tanto obtenemos que $c = \frac{1}{2}$ y el teorema queda demostrado. \square

11.5. Teorema Central del Límite

El siguiente Lema da el desarrollo de Taylor de la función característica de una variable aleatoria X con $E(X) = 0$ y $\text{var}(X) = 1$.

Lema. Sea X una aleatoria con $E(X) = 0$ y $\text{var}(X) = 1$. Entonces

$$\phi_X(t) = 1 - \frac{t^2}{2} + o_2(t^2),$$

donde $o(t^2)$ es una función tal que

$$\lim_{t \rightarrow 0} \frac{o_2(t^2)}{t^2} = 0.$$

Demostración. Sabemos que $\phi(0) = 1$ y por (11.14) y (11.15) se tiene $\phi'_X(0) = 0$ y $\phi''_X(0) = -1$. Luego usando un desarrollo de Taylor de grado 2 en $t = 0$ se tiene

$$\begin{aligned} \phi_X(t) &= \phi_X(0) + \phi'_X(0)t + \phi''_X(0)\frac{t^2}{2} + o_2(t^2) \\ &= 1 - \frac{t^2}{2} + o(t^2). \end{aligned}$$

Esto demuestra el Lema. \square

11.5.1. Caso de variables independientes igualmente distribuidas

Teorema Central del Límite. Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias independientes idénticamente distribuidas (i.i.d) con varianza finita. Llamemos $\mu = E(X_i)$ y $\sigma^2 = \text{Var}(X_i) > 0$. Sean las sumas parciales

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

y

$$Z_n = \frac{S_n - E(S_n)}{\sqrt{\text{Var}(S_n)}}. \quad (11.27)$$

Entonces

$$Z_n \rightarrow^D N(0, 1). \quad (11.28)$$

Observación

La expresión (11.27) puede reformularse escribiendo

$$Z_n = \frac{\bar{X}_n - E(\bar{X}_n)}{\sqrt{\text{Var}(\bar{X}_n)}},$$

donde

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

es la variable aleatoria "promedio aritmético".

Demostración

En primer lugar veamos que basta con probar el teorema suponiendo que $\mu = 0$ y $\sigma^2 = 1$.

Teniendo en cuenta la independencia de las X_i y la definición de S_n se tiene que

$$\begin{aligned} E(S_n) &= n\mu, \\ \text{Var}(S_n) &= n\sigma^2. \end{aligned}$$

Luego (11.27) se puede escribir como

$$\begin{aligned} Z_n &= \frac{\sum_{i=1}^n (X_i - n\mu)}{\sqrt{n}\sigma} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right) \\ &= \frac{\sum_{i=1}^n X_i^*}{\sqrt{n}}, \end{aligned}$$

donde

$$X_i^* = \frac{X_i - \mu}{\sigma}$$

Claramente las variables X_i^* son i.i.d. con $E(X_i^*) = 0$ y $\text{Var}(X_i^*) = 1$. Luego si el teorema vale para $\mu = 0$ y $\sigma^2 = 1$ vale para μ y σ^2 arbitrarios.

Supondremos entonces que $\mu = 0$ y $\sigma^2 = 1$. De acuerdo al teorema de Levy, bastará probar que para todo $t \in \mathbb{R}$

$$\lim_{n \rightarrow +\infty} \phi_{Z_n}(t) = \exp\left(-\frac{t^2}{2}\right) \quad (11.29)$$

Sabemos que como $\mu = 0$ y $\sigma^2 = 1$, por el Lema anterior para todo $i \in \mathbb{N}$ se tiene

$$\phi_{X_i}(t) = \phi_X(t) = 1 - \frac{t^2}{2} + o_2(t^2),$$

donde $o_2(t^2)$ es una función tal que

$$\lim_{t \rightarrow 0} \frac{o_2(t^2)}{t^2} = 0. \quad (11.30)$$

Como las variables X_i son independientes, podemos aplicar la propiedad P3 de las funciones características y se tiene que para todo n

$$\phi_{S_n}(t) = \prod_{i=1}^n \phi_{X_i}(t) = \left(1 - \frac{t^2}{2} + o_2(t^2)\right)^n.$$

Finalmente teniendo en cuenta que $\mu = 0$ y $\sigma^2 = 1$, resulta $Z_n = S_n/\sqrt{n}$. Luego por la propiedad P4 de las funciones características se obtiene

$$\begin{aligned}\phi_{Z_n}(t) &= \phi_{S_n}\left(\frac{t}{\sqrt{n}}\right) \\ &= \left(1 - \frac{t^2}{2n} + o_2\left(\frac{t^2}{n}\right)\right)^n.\end{aligned}$$

De acuerdo a (11.29), bastará ver que la sucesión de funciones

$$\lim_{n \rightarrow \infty} \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n = \exp\left(-\frac{t^2}{2}\right). \quad (11.31)$$

Para ello escribamos la sucesión de características del siguiente modo

$$\phi_{Z_n}(t) = \left(1 - \frac{1}{n} \left[\frac{t^2}{2} + o_2\left(\frac{t^2}{n}\right)n\right]\right)^n,$$

y luego si hacemos

$$a_n = \left[\frac{t^2}{2} + o_2\left(\frac{t^2}{n}\right)n\right],$$

entonces

$$\phi_{Z_n}(t) = \left(1 - \frac{a_n}{n}\right)^n.$$

Se conoce del cálculo elemental que si $a_n \rightarrow L$ entonces

$$\left(1 - \frac{a_n}{n}\right)^n \rightarrow \exp(-L).$$

Por lo tanto bastará mostrar que en nuestro caso $L = t^2/2$. Equivalentemente bastará con mostrar que

$$\lim_{n \rightarrow \infty} o_2\left(\frac{t^2}{n}\right)n \rightarrow 0.$$

Pero esto resulta de escribir

$$o_2\left(\frac{t^2}{n}\right)n = \frac{o_2\left(\frac{t^2}{n}\right)}{\frac{t^2}{n}} t^2$$

y de observar que como $t^2/n \rightarrow 0$ cuando $n \rightarrow \infty$, de acuerdo a (11.30) se tiene

$$\lim_{n \rightarrow +\infty} \frac{o\left(\frac{t^2}{n}\right)}{\frac{t^2}{n}} = 0.$$

Esto prueba el Teorema. \square

Observaciones.

1. Teniendo en cuenta que $E(\bar{X}_n) = n\mu/n = \mu$ y $\text{Var}(\bar{X}_n) = n\sigma^2/n = \sigma^2/n$, podemos escribir las variables Z_n de la siguiente manera

$$\begin{aligned} Z_n &= \frac{\bar{X}_n - E(\bar{X}_n)}{\sqrt{\text{Var}(\bar{X}_n)}} \\ &= n^{\frac{1}{2}} \frac{(\bar{X}_n - \mu)}{\sigma}. \end{aligned}$$

Luego, de acuerdo a (11.28) tenemos

$$n^{\frac{1}{2}} \frac{(\bar{X}_n - \mu)}{\sigma} \rightarrow^D N(0, 1). \tag{11.32}$$

De acuerdo a la Ley Fuerte de los Grandes Números $\bar{X}_n - \mu \rightarrow 0$ c.s., y por lo tanto también

$$W_n = (\bar{X}_n - \mu)/\sigma \rightarrow 0 \text{ c.s..}$$

Además, recordemos que convergencia casi segura implica convergencia en distribución. Al multiplicar W_n por el factor $n^{\frac{1}{2}}$, de acuerdo a (11.32) deja de tender a 0 y tampoco tiende infinito. Por eso se dice que la velocidad de convergencia de \bar{X}_n a μ es $n^{1/2}$. Se deja como ejercicio probar que si multiplicamos W_n por $n^{1/2+\delta}$ la sucesión converge a $\pm\infty$ en probabilidad. Es decir dado cualquier $K > 0$, tendremos

$$\lim_{n \rightarrow \infty} P(n^{1/2+\delta}|W_n| > K) = 1$$

También se deja como ejercicio que si multiplicamos W_n por $n^{1/2-\delta}$ converge en probabilidad a 0. El exponente $1/2$ es el número exacto por el para que la sucesión no converga ni a 0 ni a $\pm\infty$.

11.5.2. Teorema Central del Límite para variables no igualmente distribuidas

El Teorema Central del Límite sigue valiendo bajo condiciones menos restrictivas. Se puede suprimir la hipótesis de que las distribuciones sean idénticas y aún debilitar la hipótesis de la independencia.

El Teorema de Lindeberg o Teorema Central del Límite Fuerte da una condición suficiente para que una sucesión de variables aleatorias independientes no necesariamente idénticamente distribuidas converga en distribución a la normal estandarizada. Enunciamos este importante teorema sin demostración.

Teorema Central de Lindeberg. Sean $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias independientes con $E(X_i) = \mu_i$ y $\text{Var}(X_i) = \sigma_i^2$ para todo $i \in \mathbb{N}$. Sea

$$s_n^2 = \sum_{i=1}^n \sigma_i^2 = \text{Var}(S_n),$$

donde como antes $S_n = \sum_{i=1}^n X_i$. Definamos las variables aleatorias centradas

$$Y_i = X_i - \mu_i.$$

Una condición suficiente para que

$$Z_n = \frac{S_n - E(S_n)}{\sqrt{\text{Var}(S_n)}} \rightarrow^D \text{N}(0, 1)$$

es que para todo $\varepsilon > 0$

$$\lim_{n \rightarrow +\infty} \frac{\sum_{i=1}^n \int_{\{|y| \geq s_n \varepsilon\}} y^2 dF_{Y_i}}{s_n^2} = 0. \quad (11.33)$$

Demostración

Ver el libro citado de Barry R. James.

Observación.

La condición (11.33) se llama *condición de Lindeberg*. Notemos que como $E(Y_i) = 0$ y $\text{Var}(Y_i) = \text{Var}(X_i) = \sigma_i^2$, se tiene

$$s_n^2 = \sum_{i=1}^n \sigma_i^2 \quad (11.34)$$

$$= \sum_{i=1}^n \text{Var}(Y_i)$$

$$= \sum_{i=1}^n \int_{-\infty}^{+\infty} y^2 dF_{Y_i}$$

$$\sum_{i=1}^n \int_{\{|y| < s_n \varepsilon\}} y^2 dF_{Y_i} + \sum_{i=1}^n \int_{\{|y| \geq s_n \varepsilon\}} y^2 dF_{Y_i}. \quad (11.35)$$

Luego, la condición (11.33) es equivalente a que para todo $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n \int_{\{|y| < s_n \varepsilon\}} y^2 dF_{Y_i}}{\sum_{i=1}^n \int_{-\infty}^{+\infty} y^2 dF_{Y_i}} = 1, \quad (11.36)$$

lo cual se puede interpretar que la contribución de y_i a la varianza de S_n proviene de los valores donde $|Y_i|^2 \leq \varepsilon^2 s_n^2$. Como $s_n^2 = \text{Var}(S_n) = \text{Var}(S_n^*)$, donde $S_n^* = \sum_{i=1}^n Y_i$, resulta que la contribución de Y_i^2 a $\text{Var}(S_n^*)$ corresponde básicamente de los puntos donde $Y_i^2 < \varepsilon s_n^2$, es decir donde Y_i^2 es

pequeña respecto a $E(S_n^{*2})$. Esto está diciendo que con alta probabilidad Y_i^2 es pequeño con respecto a S_n^{*2} . En particular de (11.33) se deduce que para todo $\varepsilon > 0$, existe $n_0(\varepsilon)$ tal que para todo $n \geq n_0$

$$\int_{\{|y| \geq s_n \varepsilon\}} y^2 dF_{Y_i} < s_n^2 \varepsilon$$

para todo $1 \leq i \leq n$. Por otro lado para todo $1 \leq i \leq n$

$$\int_{\{|y| < s_n \varepsilon\}} y^2 dF_{Y_i} \leq s_n^2 \varepsilon.$$

Luego para todo $1 \leq i \leq n$ y $n \geq n_0$ se tiene

$$\sigma_i^2 = \int_{\{|y| < s_n \varepsilon\}} y^2 dF_{Y_i} + \int_{\{|y| \geq s_n \varepsilon\}} y^2 dF_{Y_i} < 2s_n^2 \varepsilon,$$

y por lo tanto que para todo $n \geq n_0$

$$\frac{\max_{1 \leq i \leq n} \sigma_i^2}{\sum_{i=1}^n \sigma_i^2} < 2\varepsilon.$$

Luego

$$\lim_{n \rightarrow \infty} \frac{\max_{1 \leq i \leq n} \sigma_i^2}{\sum_{i=1}^n \sigma_i^2} = 0.$$

Es decir que la varianza de cada variable, sobre la suma de las varianzas tiende a 0.

Del teorema central del límite de Lindeberg se deduce la siguiente versión del Teorema Central del Límite.

Teorema Central del Límite de Liapunov. Sean $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias independientes con $E(X_i) = \mu_i$ y varianza $\text{Var}(X_i) = \sigma_i^2 < \infty$ tal que para algún i_0 , $\sigma_{i_0}^2 > 0$. Llamemos $Y_i = X_i - \mu_i$ a las variable aleatoria centradas. Una condición suficiente para que

$$Z_n = \frac{S_n - E(S_n)}{\sqrt{\text{Var}(S_n)}} \rightarrow^D \text{N}(0, 1)$$

es que exista $\delta > 0$ tal que

$$\lim_{n \rightarrow +\infty} \frac{\sum_{i=1}^n E(|Y_i|^{2+\delta})}{s_n^{2+\delta}} = 0.$$

Esta condición es útil cuando las variables tienen momentos finitos de orden mayor que dos. Para la demostración del corolario ver también el libro de Barry R. James.

Consideremos ahora una sucesión de variables aleatorias $(Y_n)_{n \geq 1}$, donde Y_n tiene distribución $\text{Bi}(n, p)$. Podemos pensar Y_n como el número de éxitos

en n experimentos idénticos independientes, donde la probabilidad de éxito es p . Luego podemos escribir

$$Y_n = \sum_{i=1}^n X_i,$$

donde

$$X_i = \begin{cases} 1 & \text{si resulta éxito en el } i\text{-ésimo experimento} \\ 0 & \text{si no.} \end{cases}$$

Claramente las variables X_i son independientes e igualmente distribuidas. Sabemos que $P(X_i = 1) = p$ y $P(X_i = 0) = 1 - p$. $E(X_i) = p$ y $\text{Var}(Y_i) = p(1 - p)$. Luego, estamos en condiciones de aplicar el Teorema Central del Límite para variables i.i.d.. Entonces

$$\frac{Y_n - E(Y_n)}{\sqrt{\text{Var}(Y_n)}} = \frac{Y_n - np}{\sqrt{np(1 - p)}} \rightarrow^D N(0, 1).$$

Se puede probar que para $n = 20$ la distribución normal es una buena aproximación de la binomial, de manera que a fines prácticos se pueden usar tablas normales para calcular probabilidades binomiales.

11.5.3. Una Aplicación a la binomial.

Se realiza una encuesta para determinar el porcentaje p de votantes que va a votar a un partido C . Se toma una muestra al azar de n votantes y se observan los resultados. Designemos mediante X_i , la variable que toma el valor 1, si la intención declarada del encuestado i es votar al partido C y $X_i = 0$ en caso contrario. Claramente $P(X_i = 1) = p$

La variable

$$Y_n = \sum_{i=1}^n X_i$$

da la cantidad de encuestados que dicen votar al partido C . La variable Y_n es $\text{Bi}(n, p)$

Como desconocemos el parámetro p , podemos estimarlo a partir del promedio

$$\hat{p}_n = \bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$$

Como $E(X_i) = p$, por la ley de los grandes números tendremos $\bar{X}_n \rightarrow p$ c.s.. Lo que queremos saber es cuan grande tiene que ser n para lograr una precisión determinada con cierta probabilidad. Más precisamente supongamos que fijemos una cota e para el error $E_n = \bar{X}_n - p$ (por ejemplo $e = ,05$) y supongamos que queremos conocer aproximadamente la probabilidad de que $|E_n| \leq e$, es decir $P(|E_n| \leq e)$.

Sabemos que

$$\begin{aligned} Z_n &= \frac{Y_n - np}{\sqrt{np(1-p)}} \\ &= \frac{\sum_{i=1}^n X_i - np}{\sqrt{np(1-p)}} \\ &= \sqrt{n} \frac{\bar{X}_n - p}{\sqrt{p(1-p)}} \xrightarrow{D} N(0, 1). \end{aligned}$$

Llamemos

$$a_n = \frac{\sqrt{ne}}{\sqrt{p(1-p)}}, \quad (11.37)$$

y Φ a la función de distribución de una variable $N(0, 1)$. Luego, como Z_n se comporta aproximadamente como una $N(0, 1)$ para n grande, tenemos

$$\begin{aligned} P(|E_n| \leq e) &= P(|\bar{X}_n - p| \leq e) \\ &= P\left(\sqrt{n} \frac{|\bar{X}_n - p|}{\sqrt{p(1-p)}} \leq \frac{\sqrt{ne}}{\sqrt{p(1-p)}}\right) \\ &= P(|Z_n| \leq a_n) \\ &\cong \Phi(a_n) - \Phi(-a_n) \\ &= \Phi(a_n) - (1 - \Phi(a_n)) \\ &= 2\Phi(a_n) - 1, \end{aligned}$$

donde el signo \cong indica aproximadamente. Supongamos ahora que queremos saber que tamaño de muestra se requiere para que $P(|E_n| \leq e)$ sea aproximadamente $1 - \alpha$, donde α es un número pequeño, por ejemplo .05. Entonces se requerirá un valor n tal que

$$2\Phi(a_n) - 1 = 1 - \alpha,$$

o equivalentemente

$$a_n = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right).$$

Reemplazando a_n de acuerdo a (11.37) tendremos

$$\frac{\sqrt{ne}}{\sqrt{p(1-p)}} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right),$$

o equivalentemente

$$n = \frac{p(1-p) \left(\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right)^2}{e^2}.$$

Como p es desconocido podemos elegir el valor más desfavorable. Como n depende en forma creciente de $g(p) = p(1-p)$ deberíamos elegir el

máximo de esta función para $0 \leq p \leq 1$. Observemos que $g'(p) = 1 - 2p = 0$, razón por la cual el único punto crítico es $p = 1/2$, y como $g''(p) = -2 < 0$ corresponde a un máximo relativo. Como en los extremos $g(0) = g(1) = 0$ y $g(1/2) = 1/4$, resulta que el máximo absoluto de g está en $p = 1/2$ y el máximo de g es $1/4$. Luego basta tomar n igual a

$$n = \frac{(\Phi^{-1}(1 - \frac{\alpha}{2}))^2}{4e^2}.$$

Por ejemplo si $e = ,05$ y $\alpha = ,05$, se tendrá buscando en la tabla normal que $\Phi^{-1}(1 - \alpha/2) = \Phi^{-1}(0,975) = 1,96$, y luego

$$n = \frac{(\Phi^{-1}(1 - \frac{\alpha}{2}))^2}{4e^2} = 384,16.$$

Luego, como n tiene que ser entero, bastará tomar $n = 385$.

El valor n calculado nos asegura la probabilidad deseada, pero dado que se reemplaza $p(1-p)$ por una cota superior este valor puede ser más grande que el estrictamente necesario. En la Sección siguiente veremos un Teorema que nos permitirá reemplazar $p(1-p)$ por la estimación $\bar{X}_n(1 - \bar{X}_n)$.

11.6. Teorema de Slutsky

El siguiente Teorema, llamado de Slutsky, tiene numerosas aplicaciones en Estadística.

Teorema . Sean $(X_n)_{n \geq 1}$ e $(Y_n)_{n \geq 1}$ dos sucesiones de variables aleatorias tal que $X_n \xrightarrow{D} X$ e $Y_n \xrightarrow{P} c$, donde X es una variable aleatoria y c una constante. Entonces se tiene

(i)

$$X_n + Y_n \rightarrow X + c,$$

(ii)

$$X_n Y_n \xrightarrow{D} cX,$$

(iii) Si $c \neq 0$ entonces

$$\frac{X_n}{Y_n} \xrightarrow{D} \frac{X}{c}.$$

Probaremos el teorema de Slutsky recurriendo a una serie de lemas previos.

Lema 1. Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias y tal que $X_n \xrightarrow{D} X$ donde X es otra variable aleatoria. Entonces para todo constante $a \in \mathbb{R}$, se tiene

$$aX_n \xrightarrow{D} aX.$$

Demostración

La demostración la haremos distinguiendo tres casos: (i) $a = 0$, (ii) $a > 0$ y (iii) $a < 0$.

(i) Si $a = 0$, entonces es claro que $aX = aX_n = 0$ y por lo tanto el lema se cumple.

(ii) Sea $a > 0$. Queremos probar que para todo punto x de continuidad de F_{aX} vale que

$$\lim_{n \rightarrow +\infty} F_{aX_n}(x) = F_{aX}(x).$$

Calculamos la función de distribución de aX_n

$$\begin{aligned} F_{aX_n}(x) &= P(aX_n \leq x) \\ &= P\left(X_n \leq \frac{x}{a}\right) \\ &= F_{X_n}\left(\frac{x}{a}\right), \end{aligned}$$

y de manera análoga, la función de distribución de aX

$$F_{aX}(x) = F_X\left(\frac{x}{a}\right).$$

Entonces x es un punto de continuidad de F_{aX} si y solo si $\frac{x}{a}$ lo es de F_X .

Ahora bien como $X_n \rightarrow^D X$ vale que para todo x punto de continuidad de F_X

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

En particular eso vale para $\frac{x}{a}$. Esto demuestra el caso (i) $a > 0$.

(ii) Sea $a < 0$. Este caso resulta más complicado de probar. Probaremos en primer lugar que vale para $a = -1$ y después pasaremos al caso general.

Queremos probar que si $X_n \rightarrow^D X$ entonces $-X_n \rightarrow -X$

En primer lugar es fácil ver que en general si X es una variable aleatoria $P(X < a) = F_X(a-)$, donde $F_X(a-)$ es el límite de $F_X(x)$, cuando x tiende a a por la izquierda. Para eso basta con observar que

$$\{X < a\} = \bigcup_{i=1}^{\infty} \left\{X \leq a - \frac{1}{i}\right\}.$$

La sucesión de conjuntos $C_n = \{X \leq a - \frac{1}{n}\}$ es monótona creciente y por lo tanto

$$\begin{aligned} P(\{X < a\}) &= \lim_{n \rightarrow \infty} P\left(X \leq a - \frac{1}{n}\right) \\ &= \lim_{n \rightarrow +\infty} F_X\left(a - \frac{1}{n}\right) \\ &= F_X(a-). \end{aligned}$$

Calcularemos ahora F_{-X} y F_{-X_n} Por un lado

$$\begin{aligned} F_{-X}(x) &= P(-X \leq x) \\ &= P(X \geq -x) \\ &= 1 - P(X < -x) \\ &= 1 - F_X(-x-). \end{aligned}$$

Por otro lado y de manera análoga

$$F_{-X_n}(x) = 1 - F_{X_n}(-x^-).$$

Entonces tenemos que probar que si x es un punto de continuidad de F_{-X} entonces

$$\lim_{n \rightarrow \infty} [1 - F_{X_n}(-x-)] = 1 - F_X(-x-),$$

o equivalentemente tenemos que probar que si x es un punto de continuidad de F_{-X} entonces

$$\lim_{n \rightarrow \infty} F_{X_n}(-x-) = F_X(-x-). \quad (11.38)$$

Como F_{-X} esta definida como

$$F_{-X}(x) = 1 - F_X(-x-),$$

resulta que x es un punto de de continuidad de F_{-X} si y solo si $-x$ lo es de F_X . Por lo tanto en los puntos donde F_{-X} es continua vale que $F_X(-x-) = F_X(-x)$. Por lo tanto (11.38) es equivalente a que

$$\lim_{n \rightarrow \infty} F_{X_n}(-x-) = F_X(-x), \quad (11.39)$$

en los puntos x para los cuales $-x$ es un punto de continuidad de F_X . Como x puede ser cualquiera, esto es equivalente a que

$$\lim_{n \rightarrow \infty} F_{X_n}(x-) = F_X(x), \quad (11.40)$$

para todo punto x que sea de continuidad de F_X .

Por la monotonía de F_{X_n} se tiene que $F_{X_n}(x-) \leq F_{X_n}(x)$. Entonces tomando límite superior en ambos miembros y recordando que la hipótesis de convergencia en distribución implica que $\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$ se obtiene

$$\begin{aligned} \overline{\lim} F_{X_n}(x-) &\leq \overline{\lim} F_{X_n}(x) \\ &= \lim_{n \rightarrow \infty} F_{X_n}(x) \\ &= F_X(x). \end{aligned} \quad (11.41)$$

Para probar la desigualdad contraria, observemos que como F_X es continua en x entonces dado $\varepsilon > 0$ existe $\delta > 0$ tal que $F_X(x) - \varepsilon < F_X(x - \delta)$. Como el conjunto de puntos de discontinuidad de F_X es a lo sumo numerable, podemos elegir $x - \delta$ de forma tal que F_X sea continua en $x - \delta$. Por la monotonía de F_{X_n} resulta

$$F_{X_n}(x^-) \geq F_{X_n}(x - \delta).$$

Tomando límite inferior y recordando que $x - \delta$ es un punto de continuidad se obtiene

$$\begin{aligned} \underline{\lim} F_{X_n}(x^-) &\geq \underline{\lim} F_{X_n}(x - \delta) \\ &= \lim_{n \rightarrow \infty} F_{X_n}(x - \delta) \\ &= F_X(x - \delta) \\ &> F_X(x) - \varepsilon. \end{aligned}$$

Ahora haciendo que $\varepsilon \rightarrow 0$ se tiene

$$\underline{\lim} F_{X_n}(x^-) \geq F_X(x). \quad (11.42)$$

Por lo tanto de (11.41) y (11.42) resulta

$$\overline{\lim} F_{X_n}(x^-) \leq F_X(x) \leq \underline{\lim} F_{X_n}(x^-).$$

Pero como siempre ocurre que $\underline{\lim} F_{X_n}(x) \leq \overline{\lim} F_{X_n}(x^-)$, resulta que

$$\overline{\lim} F_{X_n}(x^-) = F_X(x) = \underline{\lim} F_{X_n}(x^-),$$

y entonces necesariamente existe $\lim F_{X_n}(x^-)$ y además

$$\lim F_{X_n}(x^-) = F_X(x).$$

Esto demuestra (11.40)

Ahora probaremos el Lema para $a < 0$. Para eso escribimos

$$aX_n = (-a)(-X_n).$$

Entonces por un lado como $X_n \xrightarrow{D} X$ se tiene que $-X_n \xrightarrow{D} -X$. Por otro lado si $a < 0$ entonces $-a > 0$ y por el caso (i) $aX_n = (-a)(-X_n) \xrightarrow{D} (-a)(-X) = aX$. \square

Definición. Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias. Decimos que la sucesión está *acotada uniformemente en probabilidad* si dado $\varepsilon > 0$ existe $K > 0$ tal que

$$P(|X_n| \leq K) \geq 1 - \varepsilon.$$

Observación. Recordemos que hemos probado que si $X_n \xrightarrow{P} X$ entonces dado $\varepsilon > 0$ existe $K > 0$ tal que para todo $n \in \mathbb{N}$

$$P(|X_n| \leq K) \geq 1 - \varepsilon$$

y

$$P(|X| \leq K) \geq 1 - \varepsilon.$$

Esto significa que si una sucesión $(X_n)_{n \geq 1}$ converge en probabilidad está acotada uniformemente en probabilidad.

Para la convergencia en distribución, se tiene un resultado análogo.

Lema 2 Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias y X otra variable aleatoria tal que $X_n \xrightarrow{D} X$. Entonces dado $\varepsilon > 0$ existe $K_0 > 0$ tal que para todo $n \in \mathbb{N}$

$$P(|X_n| \leq K_0) \geq 1 - \varepsilon$$

y

$$P(|X| \leq K_0) \geq 1 - \varepsilon.$$

Demostración.

Dado $\varepsilon > 0$ sabemos que existe $K > 0$ tal que

$$P(|X| \leq K) \geq 1 - \frac{\varepsilon}{2}$$

Observemos que si para cierto $K > 0$ vale la desigualdad, entonces también vale para cualquier $K_1 > K$. En efecto, como

$$\{|X| \leq K\} \subset \{|X| \leq K_1\},$$

tomando probabilidades se tiene

$$1 - \varepsilon \leq P(\{|X| \leq K\}) \leq P(\{|X| \leq K_1\}).$$

Luego, como el conjunto de puntos de discontinuidad de F_X es a lo sumo numerable, podemos elegir K de forma tal que F_X sea continua en K y en $-K$. Entonces

$$\begin{aligned} P(\{|X| \leq K\}) &= P(-K \leq X \leq K) \\ &= P(-K < X \leq K) \\ &= F_X(K) - F_X(-K) \end{aligned} \tag{11.43}$$

$$\geq 1 - \frac{\varepsilon}{2}. \tag{11.44}$$

Teniendo en cuenta la convergencia en distribución de X_n a X , resulta

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

y

$$\lim_{n \rightarrow \infty} F_{X_n}(-K) = F_X(-K).$$

Por definición de límite existe $n_1 \in \mathbb{N}$ tal que si $n \geq n_1$ entonces

$$F_{X_n}(K) > F_X(K) - \frac{\varepsilon}{4} \quad (11.45)$$

y también $n_2 \in \mathbb{N}$ tal que si $n \geq n_2$ entonces

$$F_{X_n}(-K) < F_X(-K) + \frac{\varepsilon}{4} \quad (11.46)$$

Luego tenemos

$$\begin{aligned} P(|X_n| \leq K) &= P(-K \leq X_n \leq K) \\ &\geq P(-K < X_n \leq K) \\ &= F_{X_n}(K) - F_{X_n}(-K). \end{aligned}$$

Sea $n_0 = \max\{n_1, n_2\}$. Luego de (11.43), (11.45) y (11.46) resulta que si $n \geq n_0$ se tiene

$$\begin{aligned} P(|X_n| \leq K) &\geq F_{X_n}(K) - F_{X_n}(-K) \\ &> F_X(K) - \frac{\varepsilon}{4} - \left(F_X(-K) + \frac{\varepsilon}{4} \right) \\ &\geq F_X(K) - F_X(-K) - \frac{\varepsilon}{2} \\ &\geq 1 - \frac{\varepsilon}{2} - \frac{\varepsilon}{2} = 1 - \varepsilon/ \end{aligned}$$

Luego hemos conseguido la acotación requerida para X y X_n con $n \geq n_0$. Finalmente para cada $1 \leq j \leq n_0 - 1$, podemos encontrar un número $K_j > 0$ tal que $P(|X_j| \leq K_j) \geq 1 - \varepsilon$. Entonces si ponemos

$$K_0 = \max\{K, K_1, K_2, \dots, K_{n_0-1}\}$$

se cumple

$$P(|X_n| \leq K_0) \geq 1 - \varepsilon, \forall n$$

y

$$P(|X| \leq K_0) \geq 1 - \varepsilon. \square$$

Lema 3. Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias uniformemente acotada en probabilidad y supongamos que $Y_n \xrightarrow{P} 0$ entonces

$$X_n Y_n \xrightarrow{P} 0.$$

Demostración

Utilizado las dos hipótesis dado $\varepsilon > 0$ existe $K > 0$ tal que

$$P(|X_n| \leq K) \geq 1 - \frac{\varepsilon}{2}$$

y $n_0 \in \mathbb{N}$ tal que para todo $n \geq n_0$ se tiene

$$P\left(|Y_n| \geq \frac{\varepsilon}{2K}\right) < \frac{\varepsilon}{2}.$$

Ahora observemos que

$$\{|X_n Y_n| > \varepsilon\} \subset \{|X_n| > K\} \cup \{|Y_n| \geq \frac{\varepsilon}{K}\},$$

ya que si $|X_n| \leq K$ y $|Y_n| < \varepsilon/K$ entonces $|X_n Y_n| \leq \varepsilon$.

Tomando probabilidades tenemos que para todo $n \geq n_0$ resulta

$$\begin{aligned} P(\{|X_n Y_n| > \varepsilon\}) &\leq P(\{|X_n| > K\}) + P\left(\{|Y_n| \geq \frac{\varepsilon}{K}\}\right) \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\ &= \varepsilon. \end{aligned}$$

Esto prueba el Lema. \square

Lema 4. Sean $(X_n)_{n \geq 1}$ e $(Y_n)_{n \geq 1}$ dos sucesiones de variables aleatorias y X otra variable aleatoria tal que $X_n \rightarrow^D X$ y $Y_n \rightarrow^P 0$. Entonces

$$X_n + Y_n \rightarrow^D X.$$

Demostración.

Queremos probar que si x es un punto de continuidad de F_X entonces

$$\lim_{n \rightarrow +\infty} F_{X_n + Y_n}(x) = F_X(x).$$

Sea $\varepsilon > 0$. Dado que el número de puntos de discontinuidad de F_X es a lo sumo numerable, siempre podemos elegir $0 < \varepsilon_1 < \varepsilon$ tal que $x + \varepsilon_1$ sea sea punto de continuidad de F_X . Luego tenemos

$$\{X_n + Y_n \leq x\} \subset \{X_n \leq x + \varepsilon_1\} \cup \{|Y_n| > \varepsilon_1\}$$

pues si $X_n > x + \varepsilon_1$ y $|Y_n| \leq \varepsilon_1$ entonces $X_n + Y_n > x$.

Tomando probabilidades en ambos miembros

$$F_{X_n + Y_n}(x) \leq F_{X_n}(x + \varepsilon_1) + P(|Y_n| > \varepsilon_1). \quad (11.47)$$

Como

$$\lim_{n \rightarrow \infty} F_{X_n}(x + \varepsilon_1) = F_X(x + \varepsilon_1),$$

y

$$\lim_{n \rightarrow \infty} P(Y_n > \varepsilon_1) = 0,$$

se obtiene

$$\begin{aligned} \underline{\lim}(F_{X_n}(x + \varepsilon_1) + P(Y_n > \varepsilon_1)) &= \lim_{n \rightarrow \infty} F_{X_n}(x + \varepsilon_1) + \lim_{n \rightarrow \infty} P(Y_n > \varepsilon_1) \\ &= F_X(x + \varepsilon_1) \\ &\leq F_X(x + \varepsilon). \end{aligned}$$

Luego tomando límite inferior en ambos miembros de (11.47) resulta

$$\overline{\lim} F_{X_n+Y_n}(x) \leq F_X(x + \varepsilon),$$

y haciendo $\varepsilon \rightarrow 0$ resulta

$$\overline{\lim} F_{X_n+Y_n}(x) \leq F_X(x). \quad (11.48)$$

Tomemos ahora $\varepsilon_1 \leq \varepsilon$ y tal que $x - \varepsilon_1$ es un punto de continuidad de F_X . Observamos que también vale

$$\{X_n \leq x - \varepsilon_1\} \subset \{X_n + Y_n \leq x\} \cup \{|Y_n| > \varepsilon_1\},$$

ya que $X_n + Y_n > x$ y $|Y_n| \leq \varepsilon$ implica $X_n + Y_n > x$ y $-\varepsilon \leq -Y_n \leq \varepsilon$ de manera que sumando $X_n > x - \varepsilon$.

Tomando probabilidades resulta

$$F_{X_n}(x - \varepsilon_1) \leq F_{X_n+Y_n}(x) + P(|Y_n| > \varepsilon_1),$$

y pasando al límite inferior, como $x - \varepsilon_1$ es un punto de continuidad de F_X se obtiene

$$F_X(x - \varepsilon_1) \leq \underline{\lim} F_{X_n+Y_n}(x).$$

Además, como

$$F_X(x - \varepsilon) \leq F_X(x - \varepsilon_1),$$

resulta

$$F_X(x - \varepsilon) \leq \underline{\lim} F_{X_n+Y_n}(x).$$

Luego tomando límite cuando $\varepsilon \rightarrow 0$, y dado que F_X es continua en x , tenemos

$$F_X(x) \leq \underline{\lim} F_{X_n+Y_n}(x). \quad (11.49)$$

De (11.48) y (11.49) se obtiene

$$\overline{\lim} F_{X_n+Y_n}(x) \leq F_X(x) \leq \underline{\lim} F_{X_n+Y_n}(x),$$

y esto implica que

$$\lim_{n \rightarrow \infty} F_{X_n+Y_n}(x) = F_X(x). \quad \square$$

Lema 5. Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias y X otra variable aleatoria tal que $X_n \rightarrow^D X$. Si a es constante, entonces

$$X_n + a \rightarrow^D X + a.$$

Demostración.

Tenemos

$$\begin{aligned} F_{X_n+a}(x) &= P(X_n + a \leq x) \\ &= P(X_n \leq x - a) \\ &= F_{X_n}(x - a), \end{aligned}$$

y

$$\begin{aligned}F_{X+a}(x) &= P(X + a \leq x) \\ &= P(X \leq x - a) \\ &= F_X(x - a).\end{aligned}$$

Poe lo tanto si x es un punto de continuidad de F_{X+a} entonces $x - a$ es un punto de continuidad de F_X de manera que aplicando la hipótesis y lo anterior

$$\begin{aligned}\lim_{n \rightarrow +\infty} F_{X_n+a}(x) &= \lim_{n \rightarrow +\infty} F_{X_n}(x - a) \\ &= F_X(x - a) \\ &= F_{X+a}(x). \quad \square\end{aligned}$$

Ahora estamos en condiciones de probar el teorema de Slutsky

Lema 6. Sea $(X_n)_{n \geq 1}$ una sucesión de variables aleatorias tal que $X_n \rightarrow^p c$, donde c es una constante. Luego si g es una función medible continua en c , se tiene

$$Y_n = g(X_n) \rightarrow^p g(c).$$

Demostración. Dado $\varepsilon > 0$ existe $\delta > 0$ tal que $|x - c| \leq \delta$ implica $|g(x) - g(c)| \leq \varepsilon$. Luego

$$\{|g(x) - g(c)| > \varepsilon\} \subset \{|x - c| > \delta\},$$

y tomando probabilidades y límites

$$\lim_{n \rightarrow \infty} P(|g(x) - g(c)| > \varepsilon) \leq \lim_{n \rightarrow \infty} P(|x - c| > \delta) = 0.$$

Luego

$$\lim_{n \rightarrow \infty} P(|g(x) - g(c)| > \varepsilon) = 0,$$

y el Lema queda probado. \square

Demostración del Teorema de Slutsky.

(i) Podemos escribir

$$X_n + Y_n = (X_n + c) + (Y_n - c).$$

Sabemos por el Lema 5 que

$$X_n + c \rightarrow^D X + c,$$

y

$$Y_n - c \rightarrow^P 0.$$

y aplicando el Lema 4

$$X_n + Y_n \rightarrow^D X + c.$$

(ii) Escibimos el producto de la siguiente manera

$$X_n Y_n = cX_n + (Y_n - c) X_n.$$

Sea

$$Z_n = (Y_n - c) X_n,$$

y

$$U_n = cX_n.$$

Por un lado sabemos que $(Y_n - c) \rightarrow^P 0$ y que $(X_n)_{n \geq 1}$ esta uniformemente acotada en probabilidad entonces aplicando el Lema 3 se tiene que

$$Z_n \rightarrow^P 0,$$

y aplicando el Lema 1

$$U_n \rightarrow^D cX.$$

Finalmente, aplicando el Lema 4

$$X_n Y_n = U_n + Z_n \rightarrow^D cX.$$

(iii) Como $c \neq 0$ y la función $g(y) = 1/y$ es continua en $y = c$, resulta por Lema 6 que

$$\frac{1}{Y_n} \rightarrow^P \frac{1}{c}.$$

Luego como

$$\frac{X_n}{Y_n} = \left(\frac{1}{Y_n} \right) X_n.$$

(iii) resulta aplicando (ii).

Ahora daremos una aplicación de estos conceptos resolviendo el siguiente problema de Estadística

11.7. Aplicación a intervalos de confianza

Problema Sea X una variable aleatoria respecto de la cual desconocemos su función de distribución F . Por ejemplo, puede tratarse del “peso de una lata de arvejas” que es una variable aleatoria que varia de lata en lata. La distribución de X no tiene porque ser normal. Supongamos que $\mu = E(X)$ y $\sigma^2 = \text{Var}(X)$ son parámetros desconocidos de la variable que dependen de F . (los parámetros de población). Se toma una muestra aleatoria de tamaño n , y se obtienen las variables aleatorias X_1, X_2, \dots, X_n . Estas variables sean independientes e idénticamente distribuidas con distribución F .

Como por la ley fuerte de los grandes números que $\overline{X}_n \rightarrow \mu$ c.s., podemos tomar como estimación del parámetro μ el promedio aritmético de las muestras \overline{X}_n

$$\hat{\mu}_n = \overline{X}_n.$$

Este valor para n grande estará próximo a la media verdadera μ , y el error será

$$E_n = \overline{X}_n - \mu.$$

Una pregunta natural es para un valor de n determinado, tratar de acotar el error E_n con probabilidad alta.

Teniendo en cuenta que la varianza se define $\sigma^2 = E(X^2) - [E(X)]^2$ podemos estimar la varianza de la siguiente manera

$$\hat{\sigma}_n^2 = \frac{\sum_{i=1}^n X_i^2}{n} - \frac{(\sum_{i=1}^n X_i)^2}{n}.$$

Teniendo en cuenta la ley de los grandes números se tendrá que

$$\frac{\sum_{i=1}^n X_i^2}{n} \rightarrow E(X^2) \text{ c.s.},$$

y

$$\frac{\sum_{i=1}^n X_i}{n} \rightarrow E(X) \text{ c.s.}$$

Luego como el cuadrado es una función continua se tendrá

$$\hat{\sigma}_n^2 \rightarrow E(X^2) - E^2(X) = \sigma^2 \text{ c.s.}$$

Por el Teorema Central del Límite

$$\sqrt{n} \frac{\overline{X}_n - \mu}{\sigma} \rightarrow^D N(0, 1). \quad (11.50)$$

Como sabemos que $\hat{\sigma}_n \rightarrow^P \sigma$, se tendrá

$$\frac{\sigma}{\hat{\sigma}_n} \rightarrow^P 1. \quad (11.51)$$

Luego teniendo en cuenta (11.50) y (11.51), y aplicando el teorema de Slutsky resulta

$$Z_n = \sqrt{n} \frac{\overline{X}_n - \mu}{\hat{\sigma}_n} = \sqrt{n} \frac{\overline{X}_n - \mu}{\sigma} \frac{\sigma}{\hat{\sigma}_n} \rightarrow^D N(0, 1).$$

Es decir, si se reemplaza σ por su estimador $\hat{\sigma}_n$ en (11.50), la convergencia en distribución no cambia.

Ahora consideremos un valor α , $0 < \alpha < 1$ llamado nivel de significación; generalmente se toma $\alpha = 0,01$; $\alpha = 0,05$. Buscamos en la tabla de la

distribución normal un valor $z_{\alpha/2}$ tal que $P(Z > \alpha/2) = \alpha/2$ donde Z es una variable $N(0, 1)$. Luego por simetría también se tendrá $P(Z < -z_{\alpha/2}) = \frac{\alpha}{2}$.

Ahora bien si $Z_n \xrightarrow{D} Z$ con $Z \sim N(0, 1)$ entonces también $-Z_n \xrightarrow{D} -Z$. Como $-Z$ también es $N(0, 1)$ tenemos que para n grande

$$P(-z_{\alpha/2} \leq -Z_n \leq z_{\alpha/2}) \approx 1 - \alpha,$$

donde \approx indica aproximadamente es decir

$$P\left(-z_{\alpha/2} \leq \sqrt{n} \frac{\mu - \bar{X}_n}{\hat{\sigma}_n} \leq z_{\alpha/2}\right) \approx 1 - \alpha,$$

y despejando

$$P\left(\bar{X}_n - \frac{z_{\alpha/2} \hat{\sigma}_n}{\sqrt{n}} \leq \mu \leq \bar{X}_n + \frac{z_{\alpha/2} \hat{\sigma}_n}{\sqrt{n}}\right) \approx 1 - \alpha. \quad (11.52)$$

Luego fijando α se se puede garantizar que μ se encuentra en el intervalo

$$\left[\bar{X}_n - \frac{z_{\alpha/2} \hat{\sigma}_n}{\sqrt{n}}; \bar{X}_n + \frac{z_{\alpha/2} \hat{\sigma}_n}{\sqrt{n}}\right].$$

sea aproximadamente $1 - \alpha$. Este intervalo se llama *intervalo de confianza*. Obsérvese que hay dos parámetros que pueden variar, el nivel de significación α y el tamaño de la muestra n . Cuando α decrece $z_{\alpha/2}$ aumenta y consecuentemente aumenta la longitud intervalo de confianza. Como contrapartida también aumenta la probabilidad que contenga a μ . En cambio cuando n crece y se mantiene el α constante, la longitud del intervalo decrece, tendiendo a 0 cuando n tiende a infinito.

Obsérvese que otra manera de escribir (11.52) es como

$$P\left(|E_n| \leq \frac{z_{\alpha/2} \hat{\sigma}_n}{\sqrt{n}}\right) \approx 1 - \alpha.$$

Es decir, la tenemos acotado el error $|E_n|$ por $z_{\alpha/2} \hat{\sigma}_n / \sqrt{n}$ con probabilidad aproximada $1 - \alpha$.

Veremos ahora un teorema de convergencia en distribución muy útil.

En primer lugar recordemos que si $(X_n)_{n \geq 1}$ es una sucesión de variables aleatorias i.i.d entonces

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \xrightarrow{D} N(0, 1)$$

o equivalentemente por el Lema 1

$$\sqrt{n} (\bar{X}_n - \mu) \xrightarrow{D} N(0, \sigma^2).$$

Supongamos que g sea una función continua en μ . Parece natural preguntarse si $\sqrt{n}(g(\bar{X}_n) - g(\mu))$ converge en distribución y en caso de respuesta positiva a que distribución. El siguiente Teorema responde esta pregunta.

Teorema. Sea $(Y_n)_{n \geq 1}$ una sucesión de variables aleatorias y $(a_n)_{n \geq 1}$ una sucesión de números reales tal que $a_n \rightarrow \infty$. Consideremos la sucesión de variables aleatorias $Z_n = a_n(Y_n - \mu)$ y supongamos que $Z_n \xrightarrow{D} X$. Sea $g : \mathbb{R} \rightarrow \mathbb{R}$ una función g con derivada continua en un entorno de μ . Entonces

(a)

$$W_n = a_n(g(Y_n) - g(\mu)) \xrightarrow{D} g'(\mu)X.$$

(b) Si $X \sim N(0, \sigma^2)$ entonces $g'(\mu)X \sim N(0, [g'(\mu)]^2 \sigma^2)$.

Demostración.

(a) Por el lema 2, la sucesión $a_n(Y_n - \mu)$ está uniformemente acotada en probabilidad. Si consideramos a la sucesión $\{a_n\}_{n \geq 1}$ de números reales como una sucesión de variables aleatorias constantes; es claro que

$$\frac{1}{a_n} \rightarrow_p 0.$$

Luego de acuerdo al lema 3

$$(Y_n - \mu) = \frac{1}{a_n}(a_n(Y_n - \mu)) \rightarrow_p 0,$$

o equivalentemente

$$Y_n \rightarrow_p \mu.$$

Como g es continua y derivable en un entorno de μ podemos aplicar el Teorema del Valor Medio y encontrar un punto intermedio ξ_n entre Y_n y μ tal que

$$W_n = a_n g'(\xi_n)(Y_n - \mu).$$

Además como $Y_n \rightarrow_p \mu$ resulta que también $\xi_n \rightarrow_p \mu$. Por la continuidad de g' y el Lema 6 se tiene

$$g'(\xi_n) \rightarrow_p g'(\mu).$$

Aplicando la parte (ii) del Teorema de Slutsky se obtiene

$$W_n = g'(\xi_n)Z_n \rightarrow g'(\mu)X.$$

(b) Se deduce de (a) pues si $X \sim N(0, \sigma^2)$ entonces $g'(\mu)X \sim N(0, [g'(\mu)]^2 \sigma^2)$. \square

Capítulo 12

Procesos de Poisson.

12.1. Procesos de punto.

Supongamos que observan sucesos que ocurren en el tiempo en forma aleatoria. Por ejemplo, los sucesos pueden ser la llegada de clientes a un negocio, llamadas telefónicas que llegan a una central, la emisión de partículas por cierto material radioactivo, etc.

Más formalmente, para cada valor $t \geq 0$, denominemos $X(t)$ la cantidad de sucesos que ocurrieron desde un instante inicial 0 hasta t . Luego supondremos que para cada t , $X(t)$ es una variable aleatoria que toma valores enteros no negativos. Además tendremos naturalmente que $X(0) = 0$, y si $t_2 > t_1$, entonces $X(t_2) \geq X(t_1)$. Todas las variables aleatorias $X(t)$, $t \geq 0$ estarán definidas sobre un mismo espacio de probabilidad (Ω, \mathcal{A}, P) , pero como la construcción de este espacio es sumamente complicada no daremos detalles sobre el mismo. Digamos solamente que un posible espacio muestral puede estar dado por

$$\Omega = \{\omega : \mathbb{R}_{\geq 0} \rightarrow \mathbb{N}_{\geq 0}; \text{ donde } \omega \text{ es no decreciente y continua a derecha}\}.$$

Luego X puede pensarse entonces dependiendo de $t \in \mathbb{R}_{\geq 0}$ y $\omega \in \Omega$, $X(t) = X(t, \omega) = \omega(t)$

Los procesos $X(t)$ que miden la cantidad de veces que ocurre un suceso hasta el tiempo t , se denominan *procesos de punto*.

12.2. Axiomática de los Procesos de Poisson

Los procesos de Poisson, son procesos de punto particulares que satisfacen los siguientes cuatro axiomas.

A1. Homogeneidad.

Supongamos que $t_2 > t_1 \geq 0$, $t_4 > t_3 \geq 0$ y además $t_4 - t_3 = t_2 - t_1$. Entonces las variables aleatorias

$$X(t_2) - X(t_1) \text{ y } X(t_4) - X(t_3)$$

tienen la misma distribución. Observando que $X(t_2) - X(t_1)$ es el número de sucesos que ocurrieron entre t_1 y t_2 , esto significa que la distribución del número de sucesos ocurridos en un período de tiempo, depende solo de la longitud de ese período.

A2. Independencia.

Consideremos dos periodos de tiempo esencialmente disjuntos (a lo sumo pueden tener en común un punto) $[t_1, t_2]$, $[t_3, t_4]$, $t_1 < t_2 \leq t_3 < t_4$. Entonces las variables aleatorias

$$X(t_2) - X(t_1) \text{ y } X(t_4) - X(t_3)$$

son independientes. Esto significa que el número de sucesos que ocurre en un período de tiempo de tiempo $[t_1, t_2]$ es independiente del número de sucesos que ocurre en el período $[t_3, t_4]$, donde $t_3 \geq t_2$. Luego el hecho de tener información sobre el número de sucesos del período $[t_1, t_2]$ no puede ser utilizada para predecir el número de sucesos del período $[t_3, t_4]$. Los periodos considerados no tienen porque ser de igual longitud.

Los axiomas A3 y A4 son de carácter más técnico.

A3. Sea

$$g_1(t) = P(X(t) = 1),$$

entonces

$$g_1'(0) = \lambda > 0,$$

es decir

$$\lim_{t \rightarrow 0} \frac{P(X(t) = 1)}{t} = \lambda > 0.$$

Esto es equivalente a que

$$P(X(t) = 1) = \lambda t + o_1(t), \tag{12.1}$$

donde

$$\lim_{t \rightarrow 0} \frac{o_1(t)}{t} = 0. \tag{12.2}$$

A4.

$$\lim_{t \rightarrow 0} \frac{P(X(t) > 1)}{t} = 0,$$

o equivalentemente existe $o_2(t)$ tal que

$$P(X(t) > 1) = o_2(t), \tag{12.3}$$

donde o_2 satisface

$$\lim_{t \rightarrow 0} \frac{o_2(t)}{t} = 0. \quad (12.4)$$

Para modelar un proceso real como un proceso de Poisson se requiere de la verificación de este conjunto de axiomas. Existen muchos procesos reales que no responden a este modelo.

12.3. Distribución de un proceso de Poisson

El siguiente teorema caracteriza la distribución de los procesos de Poisson

Teorema. Si $X(t)$ es un proceso de punto que satisface A1, A2, A3 y A4 entonces $X(t)$ tiene distribución de Poisson con parámetro λt , es decir $X(t) \sim P(\lambda t)$.

Demostración.

Para cada n dividimos el intervalo $[0, t]$ en n subintervalos de igual longitud que denominaremos I_i^n , $1 \leq i \leq n$. Más precisamente consideramos la partición regular del intervalo $[0, t]$ con $n + 1$ puntos

$$\pi^n = \left\{ 0, \frac{t}{n}, \frac{2t}{n}, \dots, \frac{(n-1)t}{n}, t \right\}.$$

Esta partición determina n subintervalos

$$I_i^n = \left[\frac{(i-1)t}{n}, \frac{it}{n} \right], 1 \leq i \leq n.$$

El número de sucesos que ocurre en I_i^n es

$$V_i^n = X\left(\frac{it}{n}\right) - X\left(\frac{(i-1)t}{n}\right).$$

Por A1, las variables V_i^n , $1 \leq i \leq n$, tienen la misma distribución que $X(t/n)$ y por el axioma A2 son independientes.

Para cada i definimos el vector aleatorio

$$\mathbf{Z}_i^n = (Z_{i1}^n, Z_{i2}^n, Z_{i3}^n)$$

donde

$$Z_{i1}^n = \begin{cases} 1 & \text{si } V_i^n = 0 \\ 0 & \text{si } V_i^n \neq 0, \end{cases}$$

$$Z_{i2}^n = \begin{cases} 1 & \text{si } V_i^n = 1 \\ 0 & \text{si } V_i^n \neq 1, \end{cases}$$

$$Z_{i3}^n = \begin{cases} 1 & \text{si } V_i^n > 1 \\ 0 & \text{si } V_i^n \leq 1. \end{cases}$$

El suceso $Z_{i1}^n = 1$ indica que en el intervalo I_i^n no ocurrió ningún suceso, $Z_{i2}^n = 1$ que ocurrió solo 1, y $Z_{i3}^n = 1$ que ocurrió más de uno. Es claro que siempre ocurre una y sólo una de esas tres posibilidades y por lo tanto

$$Z_{i1}^n + Z_{i2}^n + Z_{i3}^n = 1.$$

Por otro la distribución del vector \mathbf{Z}_i^n es multinomial, digamos con parámetros de probabilidad p_{1n}, p_{2n}, p_{3n} y para una única repetición. Luego

$$\mathbf{Z}_i^n \sim M(p_{1n}, p_{2n}, p_{3n}, 1),$$

donde

$$p_{1n} = P\left(X\left(\frac{t}{n}\right) = 0\right),$$

$$p_{2n} = P\left(X\left(\frac{t}{n}\right) = 1\right),$$

$$p_{3n} = P\left(X\left(\frac{t}{n}\right) > 1\right).$$

Usando (12.2) y (12.3) resulta

$$p_{2n} = \lambda \frac{t}{n} + o_1\left(\frac{t}{n}\right), \quad (12.5)$$

y

$$p_{3n} = o_2\left(\frac{t}{n}\right). \quad (12.6)$$

Finalmente

$$p_{1n} = 1 - p_{2n} - p_{3n} \quad (12.7)$$

$$= 1 - \lambda \frac{t}{n} - o_1\left(\frac{t}{n}\right) - o_2\left(\frac{t}{n}\right)$$

$$= 1 - \lambda \frac{t}{n} - o_3\left(\frac{t}{n}\right), \quad (12.8)$$

donde

$$o_3(t) = o_1(t) + o_2(t).$$

Claramente, de (12.2) y (12.3) resulta

$$\lim_{t \rightarrow 0} \frac{o_3(t)}{t} = 0. \quad (12.9)$$

Como las variables V_i^n , $1 \leq i \leq n$ son independientes, y como el vector \mathbf{Z}_i^n depende solo de V_i , los vectores \mathbf{Z}_i^n , $1 \leq i \leq n$ también son independientes.

Ahora definimos las variables

$$Y_1^n = \sum_{i=1}^n Z_{i1}^n,$$

$$Y_2^n = \sum_{i=1}^n Z_{i2}^n,$$

$$Y_3^n = \sum_{i=1}^n Z_{i3}^n.$$

Claramente Y_1^n es el número de intervalos en los que no ocurre ningún suceso, Y_2^n el número en los que ocurre exactamente uno e Y_3^n el número en los que ocurren más de un suceso. Luego, la distribución del vector $\mathbf{Y}^n = (Y_1^n, Y_2^n, Y_3^n)$ es multinomial con parámetros de probabilidad p_{1n}, p_{2n}, p_{3n} y n repeticiones. Por lo tanto podemos escribir

$$\mathbf{Y}^n = (Y_1^n, Y_2^n, Y_3^n) \sim M(p_{1n}, p_{2n}, p_{3n}, n).$$

Sea A_n el evento “en ningún intervalo ocurre más de un suceso”. Es decir

$$A_n = \{Y_3^n = 0\}.$$

Veremos que

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

o equivalentemente

$$\lim_{n \rightarrow \infty} P(A_n^c) = 0.$$

Observemos que

$$A_n^c = \bigcup_{i=1}^n \{Z_{i3}^n = 1\},$$

pues si en algún intervalo ocurre el suceso más de una vez entonces existe algún i tal que la variable $Z_{i3}^n = 1$ y recíprocamente.

Luego, como $P(Z_{i3}^n = 1) = p_{3n}$, usando (12.6) resulta

$$\begin{aligned} P(A_n^c) &= P\left(\bigcup_{i=1}^n \{Z_{i3}^n = 1\}\right) \\ &\leq \sum_{i=1}^n P(Z_{i3}^n = 1) = np_{3n} = no_2 \left(\frac{t}{n}\right). \end{aligned}$$

Como $t/n \rightarrow 0$, por (12.4) resulta

$$\begin{aligned} \lim_{n \rightarrow \infty} P(A_n^c) &= \lim_{n \rightarrow \infty} \left(\frac{o_2\left(\frac{t}{n}\right)}{\frac{t}{n}} t \right) \\ &= t \lim_{n \rightarrow \infty} \left(\frac{o_2\left(\frac{t}{n}\right)}{\frac{t}{n}} \right) \\ &= 0. \end{aligned} \tag{12.10}$$

Calculemos ahora la probabilidad de que hasta el momento t hayan ocurrido k sucesos. Tenemos

$$P(X(t) = k) = P(\{X(t) = k\} \cap A_n) + P(\{X(t) = k\} \cap A_n^c).$$

Pasando al límite y teniendo en cuenta (12.10) resulta

$$\lim_{n \rightarrow +\infty} P(\{X(t) = k\} \cap A_n) = 0,$$

y entonces

$$P(X(t) = k) = \lim_{n \rightarrow +\infty} P(\{X(t) = k\} \cap A_n).$$

Pero es claro que el evento $\{X(t) = k\} \cap A_n$ se caracteriza por

$$\{X(t) = k\} \cap A_n = \{Y_1^n = n - k, Y_2^n = k, Y_3^n = 0\},$$

y luego

$$P(X(t) = k) = \lim_{n \rightarrow +\infty} P\{Y_1^n = n - k, Y_2^n = k, Y_3^n = 0\}.$$

Teniendo en cuenta que la distribución del vector \mathbf{Y}^n es $M(p_{1n}, p_{2n}, p_{3n}, n)$, obtenemos

$$\begin{aligned} P(X(t) = k) &= \lim_{n \rightarrow +\infty} \frac{n!}{(n-k)!k!} p_{1n}^{n-k} p_{2n}^k p_{3n}^0 \\ &= \frac{1}{k!} \lim_{n \rightarrow +\infty} \left(\prod_{i=1}^k (n-i+1) \right) \\ &\quad \cdot \left(1 - \lambda \frac{t}{n} + o_1\left(\frac{t}{n}\right) \right)^{n-k} \left(\lambda \frac{t}{n} + o_2\left(\frac{t}{n}\right) \right)^k \end{aligned}$$

Como

$$\left(\lambda \frac{t}{n} + o_2\left(\frac{t}{n}\right) \right)^k = \frac{1}{n^k} \left(\lambda t + n o_2\left(\frac{t}{n}\right) \right)^k,$$

tenemos

$$\begin{aligned} P(\{X(t) = k\}) &= \frac{1}{k!} \lim_{n \rightarrow +\infty} \left(\prod_{i=1}^k \frac{(n-i+1)}{n} \right) \\ &\quad \cdot \left(1 - \lambda \frac{t}{n} + n o_1\left(\frac{t}{n}\right) \right)^{n-k} \left(\lambda t + n o_2\left(\frac{t}{n}\right) \right)^k, \end{aligned}$$

o bien

$$P(\{X(t) = k\}) = \frac{1}{k!} \lim_{n \rightarrow \infty} B_n C_n D_n E_n, \quad (12.11)$$

donde

$$\begin{aligned} B_n &= \prod_{i=1}^k \frac{n-i+1}{n} \\ C_n &= \left(1 - \lambda \frac{t}{n} + o_1\left(\frac{t}{n}\right)\right)^n \\ D_n &= \left(1 - \lambda \frac{t}{n} + o_1\left(\frac{t}{n}\right)\right)^{-k} \\ E_n &= \left(\lambda t + n o_2\left(\frac{t}{n}\right)\right)^k. \end{aligned}$$

Comencemos calculando el límite de B_n

$$\begin{aligned} \lim_{n \rightarrow +\infty} B_n &= \lim_{n \rightarrow +\infty} \prod_{i=1}^k \frac{n-i+1}{n} \\ &= \prod_{i=1}^k \left(\lim_{n \rightarrow +\infty} \frac{n-i+1}{n} \right) \\ &= \prod_{i=1}^k \left(1 - \lim_{n \rightarrow +\infty} \left(\frac{i-1}{n} \right) \right) \\ &= 1^k \quad (12.12) \\ &= 1. \quad (12.13) \end{aligned}$$

El límite de C_n se puede calcular de la siguiente manera

$$\begin{aligned} \lim_{n \rightarrow +\infty} C_n &= \lim_{n \rightarrow +\infty} \left(1 - \lambda \frac{t}{n} + o_1\left(\frac{t}{n}\right) \right)^n \\ &= \lim_{n \rightarrow +\infty} \left(1 - \frac{1}{n} \left(\lambda t - n o_1\left(\frac{t}{n}\right) \right) \right)^n \\ &= \lim_{n \rightarrow +\infty} \left(1 - \frac{a_n}{n} \right)^n. \end{aligned}$$

donde

$$a_n = \lambda t - n o_1\left(\frac{t}{n}\right).$$

Como en (12.10) se puede demostrar que

$$\lim_{n \rightarrow \infty} n o_1\left(\frac{t}{n}\right) = 0,$$

y entonces resulta

$$\lim_{n \rightarrow +\infty} a_n = \lambda t.$$

Por lo tanto

$$\begin{aligned} \lim_{n \rightarrow +\infty} C_n &= \lim_{n \rightarrow +\infty} \left(1 - \frac{a_n}{n}\right)^n \\ &= \exp\left(-\lim_{n \rightarrow \infty} a_n\right) \\ &= \exp(-\lambda t). \end{aligned} \tag{12.14}$$

Por otro lado, como $t/n \rightarrow 0$ y $o_1(t/n) \rightarrow 0$, resulta

$$\begin{aligned} \lim_{n \rightarrow +\infty} D_n &= \lim_{n \rightarrow +\infty} \left(1 - \lambda \frac{t}{n} + o_1\left(\frac{t}{n}\right)\right)^{-k} \\ &= 1^{-k} \\ &= 1. \end{aligned} \tag{12.15}$$

Finalmente, como $\lim_{n \rightarrow +\infty} n o_2(t/n) = 0$, resulta

$$\begin{aligned} \lim_{n \rightarrow +\infty} E_n &= \lim_{n \rightarrow +\infty} \left(\lambda t + n o_2\left(\frac{t}{n}\right)\right)^k \\ &= (\lambda t)^k. \end{aligned} \tag{12.16}$$

Usando (12.11), (12.12), (12.14), (12.15) y (12.16) obtenemos

$$P(\{X(t) = k\}) = \exp(-\lambda t) \frac{(\lambda t)^k}{k!}.$$

Esto prueba que $X(t) \sim P(\lambda t)$. \square

12.4. Tiempos de espera

Sea T_1 la variable aleatoria definida como “el tiempo necesario hasta que ocurra el primer suceso”. Calcularemos ahora su distribución.

Teorema. T_1 tiene distribución exponencial con parámetro λt ($E(\lambda t)$).
Demostración.

$$\begin{aligned} F_{T_1}(t) &= P(T_1 \leq t) \\ &= P(X(t) > 0) \\ &= 1 - P(X(t) = 0) \\ &= 1 - \exp(-\lambda t). \end{aligned}$$

Luego $T_1 \sim E(\lambda)$. \square

Otro problema de interés es la distribución de los tiempos sucesivos de ocurrencia de los sucesos. Definamos T_2 como “el tiempo de espera hasta que ocurra el segundo suceso ” entonces $T_2 - T_1$ tiene la misma distribución que T_1 . No daremos una demostración formal de este hecho. Heurísticamente, este resultado puede justificarse de la siguiente manera. $T_2 - T_1$ es el tiempo de espera para el primero suceso luego del instante T_1 . Como por A1 el proceso es homogéneo, este tiempo de espera debería tener la misma distribución que T_1 . Además por A2, T_1 está determinado por $X(t)$ con $t \leq t_1$ y $T_2 - T_1$ por $X(t)$, $t > T_1$, resulta que T_1 es independiente de $T_2 - T_1$.

Definamos ahora T_i como el tiempo de espera para que ocurran i sucesos. Luego, un argumento similar puede aplicarse, y tendremos el siguiente Teorema que enunciaremos sin demostración.

Teorema. Las variables aleatorias $T_1, T_2 - T_1, T_3 - T_2, \dots, T_i - T_{i-1}, \dots$ son i. i. d. con distribución $E(\lambda)$.

Corolario. El tiempo de espera T_i tiene distribución $\Gamma(i, \lambda)$.

Demostración.

Podemos escribir a la variable T_i como una suma telescópica

$$T_i = T_1 + (T_2 - T_1) + (T_3 - T_2) + \dots + (T_i - T_{i-1}).$$

Recordando que $E(\lambda) = \Gamma(1, \lambda)$ y teniendo en cuenta que T_i una suma de variables independientes todas $\Gamma(1, \lambda)$ resulta que $T_i \sim \Gamma(i, \lambda)$. \square

12.5. Procesos de Poisson en el plano

Los procesos de Poisson se puede generalizar al plano. No vamos a describir estos procesos con detalle, pero daremos una breve presentación. Un ejemplo de este tipo de procesos podría ser los que representan la ubicación de los árboles en un bosque.

Consideramos ahora el plano en vez de la recta. Supongamos que en ciertos puntos del plano ocurren sucesos en forma aleatoria, como por ejemplo la presencia de un árbol. Luego para cada boreliano B del plano tendremos la variable aleatoria $X(B)$ que representa la cantidad de sucesos que han ocurrido en B (por ejemplo, la cantidad de árboles). Los axiomas de un proceso de Poisson en el plano son los siguientes

AP1. Homogeneidad.

Dado un boreliano, notemos con A su área. Supongamos que $B_1, B_2 \in \mathcal{B}^2$ son borelianos del plano tal que $A(B_1) = A(B_2)$ entonces las variables aleatorias

$$X(B_1) \text{ y } X(B_2)$$

tienen la misma distribución. Esto dice que la distribución del número de sucesos que ocurre en una región del plano solo depende de su área.

AP2. Independencia.

Consideremos dos borelianos del plano esencialmente disjuntos $B_1, B_2 \in \mathcal{B}^2$, es decir tal que $A(B_1 \cap B_2) = 0$. Entonces las variables aleatorias $X(B_1)$ y $X(B_2)$ son independientes. Esto significa que cuando las regiones B_1 y B_2 tienen área en común igual a 0, entonces la información de lo que ocurre en una región B_1 no contiene ninguna información respecto de lo que ocurre en la región B_2 .

AP3.

$$\lim_{A(B) \rightarrow 0} \frac{P(\{X(B) = 1\})}{A(B)} = \lambda > 0,$$

o bien

$$P(\{X(B) = 1\}) = \lambda A(B) + o_1(A(B)).$$

AP4.

$$\lim_{A(B) \rightarrow 0} \frac{P(\{X(B) > 1\})}{A(B)} = 0,$$

o equivalentemente existe $o_2(t)$ tal que

$$P(\{X(B) > 1\}) = o_2(A(B)).$$

El siguiente Teorema se demuestra de manera totalmente análoga al correspondiente para procesos de Poisson en la recta.

Teorema. Si $X(B)$ es un proceso que satisface AP1, AP2, AP3 y AP4 entonces la distribución de $X(B)$ es Poisson con parámetro $\lambda A(B)$, $X(B) \sim P(\lambda A(B))$.

Supongamos que se elija un punto cualquiera del plano (x_0, y_0) , y sea D_1 la distancia al punto más cercano donde ocurre un suceso (en el ejemplo, sería la distancia al árbol más próximo), D_2 al punto donde ocurre segundo suceso más próximo, ..., D_i al punto donde ocurre el i -ésimo suceso más próximo

El siguiente Teorema nos da la distribución de D_1^2

Teorema. La distribución de D_1^2 es $E(\pi\lambda)$.

Demostración.

Sea $d > 0$ y sea C el círculo con centro en (x_0, y_0) y radio $d^{1/2}$. Decir que $D_1 \leq d^{1/2}$ es lo mismo que decir que en C ocurrió algún suceso. Luego

$$\begin{aligned} \{D_1^2 \leq d\} &= \{D_1 \leq d^{1/2}\} \\ &= \{X(C) > 0\} \\ &= \{X(C) = 0\}^c \end{aligned}$$

Luego tomando probabilidades y teniendo en cuenta que $A(C) = \pi d$

$$\begin{aligned} P(D_1^2 \leq d) &= 1 - P(X(C) = 0) \\ &= 1 - \exp(-\lambda A(C)) \\ &= 1 - \exp(-\lambda \pi d) \end{aligned}$$

y por lo tanto D_1^2 tiene distribución $E(\pi\lambda)$.

El siguiente Teorema del cual no se dará demostración es análogo al correspondiente Teorema para procesos de Poisson en la recta.

Teorema. Las variables aleatorias $D_1^2, D_2^2 - D_1^2, D_3^2 - D_2^2, \dots, D_i^2 - D_{i-1}^2, \dots$ son i. i. d. con distribución $E(\pi\lambda)$.

Como corolario tendremos

Corolario. La variable aleatoria D_i^2 tiene distribución $\Gamma(i, \pi\lambda)$.

Bibliografía.

1. Barry R. James. 1981. "Probabilidade: um curso em nível intermediário". IMPA. Instituto de Matemática Pura e Aplicada -CNPq. Brasil.
2. William Feller. 1978. "Introducción a la teoría de probabilidades y sus aplicaciones". Volumen I y II. Ed. Limusa.