Apuntes para el curso de

Cálculo de Probabilidades II

Ernesto Barrios Zamudio

6 de enero de 2025

Versión 0.78

Índice

Prefacio **4**

1. Vectores aleatorios **5**

1.1. Espacios de Probabilidad 5
1.2. Vectores aleatorios 6
1.3. Funciones de probabilidad acumulada 11
1.4. Vectores aleatorios continuos 13
1.5. Variables aleatorias independientes 16
1.6. Ejercicios 20

2. Distribuciones condicionales **21**

2.1. Caso discreto 21
2.2. Caso continuo 25
2.3. Esperanza y Varianza Condicional 28
2.4. Ejercicios 33

3. Esperanza y covarianzas de vectores aleatorios **34**

3.1. Recordar 34
3.2. Varianzas y covarianzas 37
3.3. Media y varianza muestrales. 42
3.4. Vector de medias y matriz de covarianzas. 43
3.5. Sumas aleatorias 49
3.6. Mezclas de distribuciones 50
3.6.1. Mezcla de distribuciones normales 50
3.6.2. Mezcla de distribuciones 51
3.7. Ejercicios 53

4. Función generadora de momentos **54**

4.1. Recordar . . . 54
4.2. Suma de variables aleatorias independientes . . . 55
4.3. Caso multivariado . . . 57
4.4. Función característica . . . 58
4.5. Ejercicios . . . 59

5. Distribución multinomial 60

5.1. Distribución binomial . . . 60
5.2. Distribución trinomial . . . 60
5.3. Distribución multinomial . . . 61
5.4. Ejercicios . . . 63

6. Distribución normal multivariada 64

6.1. La distribución normal bivariada . . . 64
6.2. La distribución normal multivariada . . . 68
6.3. Ejercicios . . . 74

7. Transformaciones de variables y vectores aleatorios 75

7.1. Caso univariado . . . 75
7.2. Transformación integral de la probabilidad . . . 79
7.3. Caso multivariado . . . 80
7.4. La distribución t de Student. . . . 86
7.5. Transformación Box-Muller. . . . 87
7.6. Ejercicios . . . 88

8. Suma y cociente de variables aleatorias 89

8.1. Suma de variables aleatorias . . . 90
8.2. Cociente de variables aleatorias . . . 92
8.3. La distribución F . . . 95
8.4. Ejercicios . . . 97

9. Estadísticos de orden 98

9.1. Funciones de distribución y de densidad del r–ésimo estadístico de orden . . . 101
9.2. Función de densidad del rango . . . 102
9.3. Función de densidad conjunta de los estadísticos de orden . . . 103
9.4. Ejercicios . . . 105

10. Desigualdades 106

10.1. Desigualdad de Chebyshev . . . 106
10.2. Desigualdad de Jensen . . . 108
10.3. Ejercicios . . . 108

11. Sucesión de variables aleatorias y teoremas límite **109**

11.1. Modos de convergencia de variables aleatorias 109

11.2. Otros resultados límite 110

11.3. Ley de los grandes números 111

11.4. Teorema central del límite 113

11.5. Ejercicios 116

Referencias **117**

Prefacio

Las condiciones en que nos encontramos este memorable año 2020 ha motivado el trabajo. La imposibilidad de compartir mis notas personales por su desorden me llevó a hacer manuscritos con la mayoría del material del temario de Cálculo de Probabilidades II. Manuscritos que fueron terminados durante el verano, que ofrecí también el curso. En paralelo, comencé a pasar las notas a una presentación más formal usando LaTeX. Este documento es el resultado.

Estos *apuntes* son precisamente eso, unos apuntes o notas para apoyar el curso Cálculo de Probabilidades II que ofrezco regularmente en ITAM.

? (**?**), corresponde a los apuntes del curso de Cálculo de Probabilidades I.

Durante el curso es mi responsabilidad motivar y ligar los distintos temas y en este sentido las notas son de apoyo al desarrollo teórico y técnico de los mismos. No se pretende que los temas sean autocontenidos ni son una versión muy preliminar de algo más elaborado y formal. No es material para ser referenciado.

Cualquier error que identifique, comentario y/o sugerencia serán bienvenido. Diríjalo a Ernesto Barrios `<ebarrios at itam.mx>`.

Ciudad de México, 3 de agosto de 2020

1. Vectores aleatorios

1.1. Espacios de Probabilidad

Considere el experimento $\mathcal{E}$ cuyas salidas o resultados ω no es posible predecir. Decimos que $\mathcal{E}$ es un **experimento aleatorio** (EA). El conjunto de posibles salidas ω lo llaman **espacio muestral** y lo denotan por Ω. Así, $\Omega = \{\omega : \omega$ es salida del experimento aleatorio $\mathcal{E}.\}$ La figura 1 ilustra el experimento aleatorio $\mathcal{E}$ y espacio muestral Ω.

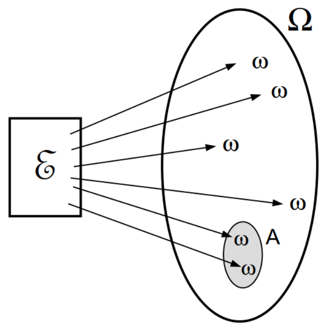


Figura 1: Experimento aleatorio $\mathcal{E}$ arroja salidas impredecibles ω. El espacio muestral Ω es el conjunto de todas las posibles resultados ω. El conjunto A de salidas de interés denota un evento.

Sea $\mathcal{F}$ un álgebra de subconjuntos de Ω. Luego, satisface

i). Si $A, B \in \mathcal{F}$, entonces, $A \cup B \in \mathcal{F}$. Esto es, $\mathcal{F}$ es *cerrado bajo uniones finitas.*

ii). Si $A \in \mathcal{F}$, entonces $A^c \in \mathcal{F}$. Esto es, $\mathcal{F}$ es *cerrado bajo complementos.*

Si además, $\mathcal{F}$ es cerrado bajo uniones numerables de subconjuntos, $\mathcal{F}$ se dice **σ-álgebra** de subconjuntos de Ω. Los subconjuntos de Ω elementos de $\mathcal{F}$ le llamamos **eventos**. Es decir, si $A \subseteq \Omega$, $A \in \mathcal{F}$, entonces, A es un evento.

Sean Ω un espacio muestral y $\mathcal{F}$ una σ-álgebra de subconjuntos de Ω. La pareja $(\Omega, \mathcal{F})$ se le dice un **espacio medible**.

Sea $\mathbb{P}$ una función de la σ-álgebra de eventos a los reales, $\mathbb{P} : \mathcal{F} \to \mathbb{R}$, tal que satisface

K1). Si $A \in \mathcal{F}$, entonces $\mathbb{P}(A) \geq 0$.

K2). $\mathbb{P}(\Omega) = 1$.

K3). Sean $A_1, A_2, \cdots \in \mathcal{F}$, eventos ajenos, $A_i \cap A_j = \emptyset$, para $i \neq j$, entonces $\mathbb{P}\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

entonces $\mathbb{P}$ se dice una **medida de probabilidad** y las propiedades K's anteriores se conocen como los **axiomas de probabilidad** o **axiomas de Kolmogorov**.

Sea Ω un espacio muestral, $\mathcal{F}$ una σ-álgebra de subconjuntos de Ω y $\mathbb{P}$ una medida de probabilidad definida sobre $\mathcal{F}$, entonces, $(\Omega, \mathcal{F}, \mathbb{P})$ se dice un **espacio de probabilidad** (EP).

En el curso de Cálculo de Probabilidades I se derivan varios resultados a partir de los axiomas anteriores. Algunos de ellos se presentan en el siguiente corolario.

Corolario : Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad. Entonces, a). Si $A, B \in \mathcal{F}$ y $A \subseteq B$, entonces $\mathbb{P}(A) \leq \mathbb{P}(B)$.

b). $0 \leq \mathbb{P}(A) \leq 1$, para todo $A \in \mathcal{F}$.

c). Si $A \in \mathcal{F}$, entonces $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

d). $\mathbb{P}\big(\emptyset\big) = 0$.

e). $A, B \in \mathcal{F}$, entonces $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

f). Etcétera ...

Las notas del curso anterior, Cálculo de Probabilidades I, las encuentra en Barrios (2024).

1.2. Vectores aleatorios

Definición : Considere $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad (EP) y (X, Y) una función real bivariada tal que $(X, Y) : \Omega \to \mathbb{R} \times \mathbb{R}$. Si para todo $(x, y) \in \mathbb{R} \times \mathbb{R}$, se tiene que $\big\{\omega \in \Omega : (X, Y)(\omega) = (x, y)\big\} \in \mathcal{F}$, (X, Y) se dice un **vector aleatorio** (*v. a.*) definido sobre el espacio medible $(\Omega, \mathcal{F})$.

Definición : Sea (X, Y) un *v. a.* definido sobre el espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$. Considere la función $f_{XY} : \mathbb{R} \times \mathbb{R} \to [0, 1]$, tal que para todo $(x, y) \in \mathbb{R} \times \mathbb{R}$,

$$f_{XY}(x, y) = \mathbb{P}\big(\{\omega \in \Omega : (X, Y)(\omega) = (x, y)\}\big) = \mathbb{P}(X = x, Y = y)$$

Si $\{(x, y) : f_{XY}(x, y) > 0\}$ es finito o infinito numerable (X, Y) se dice un **vector aleatorio discreto** y f_{XY} su **función masa de probabilidad conjunta** (*f. m. p.*).

Ejemplo : *Dados cargados.* (Vea ejemplo 5.6 de León-García (2008).) Considere el siguiente par de dados representados por el *v. a.* (X, Y)

	X/Y	1	2	3	4	5	6	$\mathbb{P}(X = x)$
	1	2	1	1	$\cdots$		1	7
$\frac{1}{42}$	2	1	2	1	$\cdots$		1	7
	$\vdots$				$\ddots$			
	6	1	1	1	$\cdots$		2	7
	$\mathbb{P}(Y = y)$	7	7				7	42

La suma total dentro de la tabla es 42, de ahí el factor $1/42$ a la izquierda del arreglo, para que éste represente la *f. m. p.* conjunta del *v. a.* (X, Y).

Así pues, en este ejemplo

$$\mathbb{P}(X = x, Y = y) = \begin{cases} \frac{2}{42} & \text{si } x = y \\ \frac{1}{42} & \text{si } x \neq y \end{cases}$$

y la correspondiente *f. m. p.* conjunta del *v. a.* (X, Y) es para $x, y = 1, \ldots, 6$,

$$f_{XY}(x, y) = \frac{2}{42} \mathbb{1}_{\{X=Y\}}(x, y) + \frac{1}{42} \mathbb{1}_{\{X \neq Y\}}(x, y)$$

donde $\mathbb{1}_A$ representa la **función indicadora** del conjunto o evento A. Esto es, $\mathbb{1}_A(\omega) = 1$, si $\omega \in A$ y $\mathbb{1}_A(\omega) = 0$ si $\omega \notin A$.

Note que

I). $f_{XY}(x,y) \geq 0$, para todo $x, y \in \mathbb{R}$.

II). $\sum_{x=1}^{6} \sum_{y=1}^{6} f_{XY}(x,y) = 1$.

Además, por ejemplo,

$$\begin{aligned}
\mathbb{P}(X=2) &= \mathbb{P}(X=2, Y \text{ “}\textit{cualquiera}\text{”}) \\
&= \mathbb{P}\left(X=2 \cap \left(\cup_{y=1}^{6}\{Y=y\}\right)\right) \\
&= \mathbb{P}\left(\cup_{y=1}^{6}\{X=2, Y=y\}\right) \\
&= \sum_{y=1}^{6} \mathbb{P}(X=2, Y=y) \\
&= \frac{1}{42}(1+2+1+\cdots+1) \\
&= \frac{7}{42} = \frac{1}{6}
\end{aligned}$$

Así, para todo $x = 1, \ldots, 6$, se tiene $\mathbb{P}(X=x) = 1/6$, por lo que

$$f_X(x) = \mathbb{P}(X=x) = \frac{1}{6}\mathbb{1}_{\{1,\cdots,6\}}(x)$$

es la **función masa de probabilidad marginal** de la *v. a.* X. Se manera similar, $f_Y(y) = \frac{1}{6}\mathbb{1}_{\{1,\ldots,6\}}(y)$ es la correspondiente *f. m. p.* marginal de Y.

Ejemplo : *Dados honestos.* Considere ahora el caso de los dados “honestos” X y Y. Su *f. m. p.* conjunta es

$$f_{XY}(x,y) = \frac{1}{36}\mathbb{1}_{\{1,\ldots,6\}\times\{1,\ldots,6\}}(x,y)$$

Así, también se tiene

I). $f_{XY}(x,y) \geq 0$, para todo $x, y \in \mathbb{R}$.

II). $\sum_{x=1}^{6} \sum_{y=1}^{6} f_{XY}(x,y) = 1$.

Y nuevamente, la *f. m. p.* marginales de X y de Y son

$$f_X(x) = \frac{1}{6}\mathbb{1}_{\{1,\cdots,6\}}(x) \quad \text{y} \quad f_Y(y) = \frac{1}{6}\mathbb{1}_{\{1,\cdots,6\}}(y)$$

Definición : Sea (X,Y) un *v. a.* con *f. m. p.* conjunta f_{XY} y sea $S_{XY} = \{(x,y) \in \mathbb{R}^2 : f_{XY}(x,y) > 0\}$. S_{XY} se dice el **soporte** del *v. a.* (X,Y), o bien, **soporte de la distribución** de (X,Y).

Definición : Se definen los **conjuntos borelianos** de $\mathbb{R}^2$, $\mathcal{B}(\mathbb{R}^2)$ como los elementos de la σ-álgebra generadas por rectángulos $[a,b) \times [c,d) \subseteq \mathbb{R}^2$, para todo $a < b$ y $c < d$.

Proposición : Sea (X,Y) un *v. a.* con *f. m. p.* conjunta f_{XY} y sea $A \in \mathcal{B}(\mathbb{R}^2)$. Entonces,

$$\mathbb{P}\big((X,Y) \in A\big) = \mathbb{P}_{XY}(A) = \sum_{(x_i,y_j)\in A} f_{XY}(x_i,y_j)$$

donde $\mathbb{P}_{XY}(\cdot)$ se entiende como “la medida de probabilidad *inducida* por la distribución (o *ley de probabilidades*) de (X,Y)”.

Proposición : Sea (X,Y) un *v. a.* con *f. m. p.* conjunta f_{XY} con soporte S_{XY}, entonces

I). $f_{XY}(x, y) \geq 0$, para todo $x, y \in \mathbb{R}$.

II). $\sum_{(x_i,y_j)\in S_{XY}} f_{XY}(x_i, y_j) = 1$,

Demostración:

I). $f_{XY}(x, y) = \mathbb{P}(X = x, Y = y) \geq 0$, por axiomas de probabilidad.

II). Sea $A_{ij} = \{\omega \in \Omega : f_{XY}(x_i, y_j) > 0\}$. Entonces, $\{A_{ij}\}$ forma una *partición* de Ω. Luego, se sigue de los axiomas de probabilidad

$$
\begin{aligned}
1 &= \mathbb{P}(\Omega) \\
&= \mathbb{P}\left(\cup_{i,j} A_{ij}\right) \\
&= \sum_{i.j} \mathbb{P}_{XY}(A_{ij}) \\
&= \sum_{x_i,y_j} \mathbb{P}\left((X,Y) \in A_{ij}\right) \\
&= \sum_{(x_i,y_j)\in S_{XY}} f_{XY}(x_i, y_j)
\end{aligned}
$$

Definición : La función f que satisface I y II de la proposición anterior se dice **función masa de probabilidad propia** o *legítima*.

Proposición : Sea (X, Y) *v. a.* con *f. m. p.* conjunta f_{XY}. Entonces f_X, la *f. m. p.* marginal de X está dada por

$$f_X(x) = \sum_{(x,y_j)\in S_{XY}} f_{XY}(x, y_j), \qquad \text{para todo} \;\; x \in \mathbb{R}$$

De manera similar,

$$f_Y(y) = \sum_{(x_i,y)\in S_{XY}} f_{XY}(x_i, y), \qquad \text{para todo} \;\; y \in \mathbb{R}$$

Demostración: Para todo $x \in \mathbb{R}$,

$$
\begin{aligned}
f_X(x) &= \mathbb{P}(X = x) \\
&= \mathbb{P}(X = x, Y \text{ “}\textit{cualquiera}\text{”}) \\
&= \mathbb{P}\left(\{X = x\} \cap \left(\cup_{(x,y_j)\in S_{XY}} \{Y = y_j\}\right)\right) \\
&= \sum_{(x,y_j)\in S_{XY}} \mathbb{P}\left(X = x, Y = y_j\right) \\
&= \sum_{(x,y_j)\in S_{XY}} f_{XY}(x, y_j)
\end{aligned}
$$

por corolario de probabilidad ya que los eventos $\{Y = y_j\}$ son ajenos para distintos y_j's.

Ejemplo : *Dados cargados*

$$f_X(x) = \sum_{y=1}^{6} f_{XY}(x, y) = \frac{7}{42} = \frac{1}{6}, \quad x = 1, \ldots, 6$$

Ejemplo : *Dados honestos*

$$f_X(x) = \sum_{y=1}^{6} f_{XY}(x,y) = 6\frac{1}{36} = \frac{1}{6}, \quad x = 1, \ldots, 6$$

Note que en ambos ejemplos se tienen las mismas *f. m. p.* marginales pero distinta *f. m. p.* conjunta, lo que da lugar a la siguiente proposición.

Proposición : Sea (X, Y) *v. a.* discreto. A partir de las *f. m. p.* conjunta se puede encontrar las *f. m. p.* marginales, pero no viceversa necesariamente.

Ejemplo : Considere la *f. m. p.* conjunta f dada por

$$f(x,y) = \frac{4xy}{n^2(n+1)^2}$$

para $x, y = 1, \ldots, n$.

a). Muestre que f es una *f. m. p. propia.*

b). Calcule $\mathbb{P}(X = 1)$.

c). Calcule $\mathbb{P}(X = Y)$.

d). Determine $\mathbb{P}(X + Y = n + 1)$.

Solución:

a). I) Claramente $f(x,y) \geq 0$, para todo x, y.

II)

$$\sum_{(x_i,y_j)\in S_{XY}} f(x_i, y_j) = \sum_{x=1}^{n}\sum_{y=1}^{n} \frac{4xy}{n^2(n+1)^2} = \frac{4}{n^2(n+1)^2}\sum_{x=1}^{n} x \sum_{y=1}^{n} y = 1$$

por lo que f es una *f. m. p.* propia.

b).

$$\begin{aligned}
\mathbb{P}(X = 1) &= \sum_{y=1}^{n} f(1, y) \\
&= \frac{4(1)}{n^2(n+1)^2}\sum_{y=1}^{n} y \\
&= \frac{2}{n(n+1)}
\end{aligned}$$

En general,

$$f_X(x) = \mathbb{P}(X = x) = \frac{2x}{n(n+1)}\mathbb{1}_{\{1,\ldots,n\}}(x)$$

c).

$$\begin{aligned}
\mathbb{P}(X = Y) &= \sum_{x=1}^{n} f(x, x) \\
&= \frac{4}{n^2(n+1)^2}\sum_{x=1}^{n} x^2 \\
&= \frac{4}{n^2(n+1)^2}\frac{n(n+1)(2n+1)}{6} \\
&= \frac{2}{3}\frac{2n+1}{n(n+1)}
\end{aligned}$$

d).

$$
\begin{aligned}
\mathbb{P}\left(X+Y=n+1\right) &= \sum_{x=1}^{n} \mathbb{P}\left(X=x, Y=n+1-x\right) \\
&= \sum_{x=1}^{n} f(x, n+1-x) \\
&= \frac{4}{n^2(n+1)^2} \sum_{x=1}^{n} x(n+1-x) \\
&= \frac{4}{n^2(n+1)^2} \left[(n+1) \sum_{1}^{n} x - \sum_{1}^{n} x^2 \right] \\
&= \frac{2n+4}{3n(n+1)}
\end{aligned}
$$

Ejemplo : (Vea ejemplo 5.9 de León-García (2008).) El número de bits N en un mensaje sigue una distribución geométrica con media $(1-\delta)/\delta$ con parámetro $0 < \delta < 1$. Para su envío el mensaje es dividido en bloques de m bits. Sea Q el número de bloques completos y R el número de bits restantes. Entonces se tiene que $N = mQ + R$, donde $N \sim \text{Geom}(\delta)$ con soporte $S_N = \{0, 1, \ldots\}$. Encuentre las distribuciones marginales de Q y R.

Solución: Note que los soportes de N, Q y R son, respectivamente

$$
S_N = \{0, 1, \ldots\}, \quad S_Q = \{0, 1, \ldots\}, \quad S_R = \{0, 1, \ldots, m-1\}
$$

La *f. m. p.* de $N = mQ + R$ es geométrica, luego

$$
f_N(n) = \mathbb{P}\left(N = n\right) = \delta(1-\delta)^n = \delta(1-\delta)^{mq+r}
$$

por lo que la *f. m. p.* conjunta de (Q, R) es

$$
f_{QR}(q, r) = \mathbb{P}\left(Q = q, R = r\right) = \delta(1-\delta)^{mq+r}
$$

a). Sea $q = 0, 1, \ldots,$

$$
\begin{aligned}
f_Q(q) &= \mathbb{P}\left(Q = q\right) \\
&= \sum_{r=0}^{m-1} f_{QR}(q, r) \\
&= \sum_{r=0}^{m-1} \delta(1-\delta)^{mq+r} \\
&= \delta(1-\delta)^{mq} \sum_{0}^{m-1} (1-\delta)^r \\
&= \delta(1-\delta)^{mq} \frac{1-(1-\delta)^m}{1-(1-\delta)} \\
&= \left[1-(1-\delta)^m\right] \left[(1-\delta)^m\right]^q
\end{aligned}
$$

Esto es, Q se distribuye geométricamente con parámetro $1-(1-\delta)^m$

b). Sea $r = 0, \ldots, m-1$.

$$\begin{aligned} f_R(r) &= \mathbb{P}\left(R = r\right) \\ &= \sum_{q=0}^{\infty} f_{QR}(q,r) \\ &= \delta(1-\delta)^r \sum_{q=0}^{\infty} \left[(1-\delta)^m\right]^q \\ &= \frac{1}{1-(1-\delta)^m}\, \delta(1-\delta)^r \end{aligned}$$

Por lo que R se distribuye geométricamente con probabilidad de éxito δ y truncada en m, de ahí la *constante normalizadora* $K = 1/[1-(1-\delta)^m]$.

Ejercicio : Considere el *v. a.* (X,Y) con *f. m. p.* conjunta dada por

$$f(x,y) = \frac{\lambda^y e^{-\lambda} p^x (1-p)^{y-x}}{x!(y-x)!} \mathbb{1}_{\{0,1,\ldots\}}(x) \mathbb{1}_{\{x,x+1,\ldots\}}(y)$$

a). Describa gráficamente el soporte de la distribución.

b). Muestre que marginalmente $X \sim \mathrm{Po}(\lambda p)$ y $Y \sim \mathrm{Po}(\lambda)$.

c). Verifique que $\mathbb{P}\left(X = Y\right) = e^{-\lambda(1-p)}$.

1.3. Funciones de probabilidad acumulada

Definición : Sea (X,Y) *v. a.* bivariado en el espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$. Se define la **función de probabilidad acumulada conjunta** (*f. p. a.*) ó **función de distribución** F_{XY} por

$$F_{XY}(x,y) = \mathbb{P}\left(X \le x, Y \le y\right), \qquad \text{para todo } \ x, y \in \mathbb{R}$$

Propiedades:

1. F_{XY} es no decreciente "en dirección noreste". Esto es, si $x_1 < x_2$, $y_1 < y_2$,

$$F_{XY}(x_1,y_1) \le F_{XY}(x_2,y_2)$$

2. I) $F_{XY}(-\infty, y) = 0$, para todo $y \in \mathbb{R}$.
   II) $F_{XY}(x, -\infty) = 0$, para todo $x \in \mathbb{R}$.
   III) $F_{XY}(\infty, \infty) = 1$.

3. I) $F_X(x) = \lim_{y\to\infty} F_{XY}(x,y)$, para todo $x \in \mathbb{R}$
   II) $F_Y(y) = \lim_{x\to\infty} F_{XY}(x,y)$, para todo $y \in \mathbb{R}$

4. F_{XY} es continua en "dirección suroeste". Esto es,

$$\lim_{\delta,\epsilon\to 0^+} F_{XY}(x+\delta, y+\epsilon) = F_{XY}(x,y)$$

5. Sean $a < b$ y $c < d$, entonces

$$\mathbb{P}\left(a < X \le b,\, c < Y \le d\right) = F_{XY}(b,d) - F_{XY}(b,c) - F_{XY}(a,d) + F_{XY}(a,c)$$

Demostración:

1. Similar al caso univariado.
2. Note por ejemplo que, $F_{XY}(-\infty, y) = \mathbb{P}(X \leq -\infty, Y \leq y) = 0$.
3. $$\lim_{y\to\infty} F_{XY}(x,y) = \mathbb{P}(X \leq x, Y \leq \infty) = \mathbb{P}(X \leq x) = F_X(x)$$
4. Similar al caso univariado.
5. Al restar $F_{XY}(a,d)$ y $F_{XY}(c,b)$ se elimina dos veces el semiplano con vértice (a,c), por lo que se restituye una vez al sumar $F_{XY}(a,c)$.

Nota: Una función $F(x,y)$ que satisface las propiedades anteriores no necesariamente es una función de probabilidad acumulada. Para esto, se debe tener que $F(b,d) - F(b,c) - F(a,d) + F(a,c) \geq 0$ para todo $a < b$ y $c < d$. En efecto, considere

$$F(x,y) := \begin{cases} 0, & x+y < 0 \\ 1, & x+y \geq 0 \end{cases}$$

F satisface las propiedades 1-4 anteriores pero no la propiedad 5, por lo que no define una función de probabilidad acumulada (*f. p. a.*). Si aplica la propiedad 5, se tiene que

$$\mathbb{P}(-1 < X \leq 3, -1 < Y \leq 3) = F(3,3) - F(3,-1) - F(-1,3) + F(-1,-1) = 1-1-1+0 = -1$$

que es un absurdo.

Ejemplo : Considere el *v. a.* (X,Y) con *f. p. a.* dada por

$$F_{XY}(x,y) = \left(1 - e^{-\alpha x}\right)\left(1 - e^{-\beta y}\right), \quad x, y \geq 0$$

Para $\alpha = 1/2$ y $\beta = 1/3$, determine la probabilidad de los siguientes eventos: $A = \{X \leq 1, Y \leq 2\}$, $B = \{X \geq 2, Y > 1\}$, $C = \{2 < X < 3, 1 < Y < 2\}$.

Solución:

I. $$\mathbb{P}(A) = \mathbb{P}(X \leq 1, Y \leq 2) = F_{XY}(1,2) = (1 - e^{-1/2})(1 - e^{-2/3}) = 0.1915$$

II. $$\begin{aligned} \mathbb{P}(B) &= \mathbb{P}(X \geq 2, Y > 1) \\ &= 1 - \mathbb{P}\left(\left[\{X \geq 2\} \cap \{Y > 1\}\right]^C\right) \\ &= 1 - \left[\mathbb{P}\left(\{X < 2\} \cup \{Y \leq 1\}\right)\right] \\ &= 1 - \left[\mathbb{P}(X < 2) + \mathbb{P}(Y \leq 1) - \mathbb{P}(X < 2, Y \leq 1)\right] \\ &= 1 - \left[F_X(2) + F_Y(1) - F(2,1)\right] \\ &= 1 - \left[F(2,\infty) + F(\infty,1) - F(2,1)\right] \\ &= 1 - \left[(1 - e^{-2/2}) + (1 - e^{-1/3}) - (1 - e^{-2/2})(1 - e^{-1/3})\right] \\ &= 0.2636 \end{aligned}$$

III. Se sigue de la continuidad de F_{XY} que $\mathbb{P}(2 < X < 3) = \mathbb{P}(2 < X \leq 3)$ y lo mismo para Y. Luego,

$$\mathbb{P}(2 < X < 3, 1 < Y < 2) = F(3,2) - F(3,1) - F(2,2) + F(2,1) = 0.0294$$

1.4. Vectores aleatorios continuos

Definición : Sea (X, Y) un *v. a.* tal que para todo $A \in \mathcal{B}(\mathbb{R}^2)$,

$$\mathbb{P}\left((X,Y) \in A\right) = \mathbb{P}_{(X,Y)}(A) = \iint_A h(x,y)dxdy$$

para alguna función $h \geq 0$. El vector aleatorio (X, Y) se dice (*absolutamente*) **continuo** con **función de densidad de probabilidad conjunta** (*f. d. p.*) h (*con respecto a la integral*).

Proposición : Sea (X, Y) un *v. a.* con *f. p. a.* conjunta F y *f. d. p.* conjunta f, entonces para toda $x, y \in \mathbb{R}$,

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v)dvdu$$

Demostración: Sea $A = (-\infty, x] \times (-\infty, y]$ y aplique la definición de función de densidad de probabilidad conjunta.

Proposición : Sea (X, Y) *v. a.* continuo con *f. d. p.* conjunta f. Entonces,

I. $f(x, y) \geq 0$, para todo $x, y \in \mathbb{R}$.

II. $\iint_{\mathbb{R}^2} f(x,y)dxdy = 1$.

Demostración:

I. Por definición de *v. a.* continuo, $f \geq 0$.

II. $1 = \mathbb{P}\left((X,Y) \in \mathbb{R}^2\right) = \mathbb{P}_{(X,Y)}(\mathbb{R}^2) = \iint_{\mathbb{R}^2} f(x,y)dxdy$, y nuevamente por definición de *f. d. p.*

Definición : Una función que satisface la proposición anterior se dice *f. d. p.* **propia** o *legítima*.

Proposición : Sea (X, Y) *v. a.* continuo con *f. d. p.* conjunta f y *f. p. a.* conjunta F diferenciable. Entonces,

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$$

Demostración: Se sigue de la definición de *v. a.* absolutamente continua y *f. d. p.*.

Nota: Si la *f. p. a.* F no es diferenciable entonces la *f. d. p.* f no necesariamente existe.

Proposición : Sea (X, Y) un *v. a.* con *f. d. p.* conjunta f. Entonces, para $x_1 < x_2$ y $y_1 < y_2$,

$$\mathbb{P}\left(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2\right) = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f(x,y)dydx$$

Demostración: Se sigue de la definición de *f. d. p.*.

Proposición : Sea (X, Y) *v. a.* con *f. d. p.* y *f. p. a.* conjuntas f y F y marginales f_X, f_Y, F_X, F_Y, respectivamente. Entonces,

I. $f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$, para todo $x \in \mathbb{R}$.

II. $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$, para todo $y \in \mathbb{R}$.

Demostración:

I.

$$\begin{aligned}
f_X(x) = \frac{dF_X(x)}{dx} &= \frac{dF(x,\infty)}{dx} \\
&= \frac{d}{dx}\left[\int_{-\infty}^{x}\int_{-\infty}^{\infty} f(u,v)dvdu\right] \\
&= \frac{d}{dx}\int_{\infty}^{x}\left[\int_{-\infty}^{\infty} f(u,y)dy\right]du \\
&= \int_{-\infty}^{\infty} f(x,y)dy
\end{aligned}$$

por el Teorema Fundamental del Cálculo.

II. De manera similar,

$$\begin{aligned}
f_Y(y) = \frac{dF_Y(y)}{dy} &= \frac{dF(\infty,y)}{dy} \\
&= \frac{d}{dy}\left[\int_{-\infty}^{\infty}\int_{-\infty}^{y} f(u,v)dvdu\right] \\
&= \frac{d}{dy}\int_{-\infty}^{y}\left[\int_{-\infty}^{\infty} f(x,v)dx\right]dv \\
&= \int_{-\infty}^{\infty} f(x,y)dx
\end{aligned}$$

donde se invirtió el orden de integración por el Teorema de Fubini.

Ejemplo : Sean (X,Y) distribuidos *uniformemente* en el cuadrado unitario $C = [0,1] \times [0,1]$, con función de densidad conjunta $f(x,y) = \mathbb{1}_C(x,y)$. Encuentre la correspondiente *f. p. a.* conjunta y las marginales F_X y F_Y.

Solución:

I. Si $x \leq 0$ o $y \leq 0$, $F(x,y) = \mathbb{P}(X \leq x, Y \leq y) = 0$.

II. Si $(x,y) \in C$,

$$\begin{aligned}
F(x,y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)dvdu \\
&= \int_{-\infty}^{x}\int_{-\infty}^{y} \mathbb{1}_C(u,v)dvdu \\
&= \int_{0}^{x}\int_{0}^{y} 1 \cdot dvdu \\
&= xy
\end{aligned}$$

III. Si $0 \leq x \leq 1$ y $y \geq 1$,

$$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} \mathbb{1}_C(u,v)dvdu = \int_{0}^{x}\int_{0}^{1} 1 \cdot dvdu = x$$

IV. Si $x \geq 1$ y $0 \leq y \leq 1$,

$$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} \mathbb{1}_C(u,v)dvdu = \int_0^1\int_0^y 1 \cdot dvdu = y$$

V. Si $x, y \geq 1$, $F(x,y) = 1$.

VI. Ahora bien,

$$F_X(x) = F(x,\infty) = F(x,1) = x$$

Esto es, X, y por simetría Y, se distribuye uniformemente en $[0,1]$.

Ejemplo : Sea (X,Y) *v. a.* con *f. d. p.* conjunta dada por

$$f(x,y) = c\, e^{-(x+y)}\mathbb{1}_{\{0\leq x\leq y<\infty\}}(x,y)$$

I. Bosqueje el soporte de la distribución.

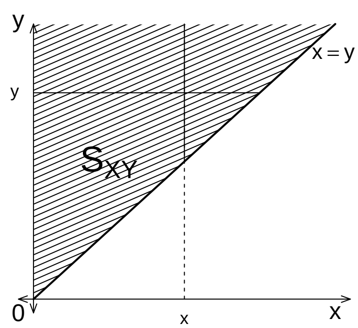
II. Encuentre la constante c de manera que f sea una *f. d. p.* propia.

III. Determine las *f. d. p.* f_X y f_Y.

IV. Verifique que $\mathbb{P}(X+Y\leq 1) = 1 - 2e^{-1} = 0.2642$.

Solución:

I. $S_{XY} = \{(x,y)\in\mathbb{R}^2 | 0\leq x\leq y<\infty\}$.



II.

$$\begin{aligned}
1 &= c\iint_{\mathbb{R}^2} e^{-(x+y)}\mathbb{1}_{\{0\leq x\leq y<\infty\}}(x,y)dxdy\\
&= c\int_0^{\infty}\int_x^{\infty} e^{-(x+y)}dydx\\
&= c\int_0^{\infty} e^{-x}\int_x^{\infty} e^{-y}dy\cdot dx\\
&= c\int_0^{\infty} e^{-2x}dx\\
&= c\frac{1}{K}\int_0^{\infty} Ke^{-2x}dx\\
&= c/K
\end{aligned}$$

donde K es la constante normalizadora de la función de densidad con núcleo e^{-2x}, que corresponde a la distribución exponencial con media $1/2$, por lo que $K = 2$ y por lo tanto $c = 2$.

III. $S_X = (0, \infty)$,

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy = 2e^{-x} \int_{x}^{\infty} e^{-y}dy = 2e^{-2x}\mathbb{1}_{(0,\infty)}(x)$$

Por lo que X se distribuye exponencialmente con media $1/2$.

IV. $S_Y = (0, \infty)$,

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx = 2e^{-y} \int_{0}^{y} e^{-x}dx = 2e^{-y}(1 - e^{-y})\mathbb{1}_{(0,\infty)}(y)$$

V. $A = \big\{(x, y) : x + y \leq 1\big\}$,

$$\begin{aligned}
\mathbb{P}\left(X + Y \leq 1\right) &= \mathbb{P}((X, Y) \in A) \\
&= \mathbb{P}_{XY}(A) \\
&= \iint_A f(x, y)dxdy \\
&= \int_0^{1/2} \int_x^{1-x} 2e^{-(x+y)}dydx \\
&= \int_0^{1/2} 2e^{-2x}dx - 2\int_0^{1/2} e^{-1}dx \\
&= (1 - e^{-1}) - e^{-1} \\
&= 1 - 2e^{-1} \\
&= 0.2642
\end{aligned}$$

1.5. Variables aleatorias independientes

Definición : Sean $X_1, X_2, \ldots, X_n$, variables aleatorias definidas sobre el espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$, Las variables $X_1, \ldots, X_n$ se dicen **variables aleatorias independientes** si para cualesquiera conjuntos borelianos $B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R})$, se tienen que

$$\mathbb{P}(X_1 \in B_1, \ldots, X_n \in B_n) = \mathbb{P}(X_1 \in B_1) \cdots \mathbb{P}(X_n \in B_n)$$

La colección $X_1, X_2, \ldots$ se dicen variables aleatorias independientes si cualquier colección finita de X's es independiente.

Considere ahora $(\Omega, \mathcal{F}, \mathbb{P})$ y (X, Y) el vector aleatorio bivariado con función de distribución conjunta F y marginales F_X, F_Y. Suponga además que para todo $x, y \in \mathbb{R}$, los eventos

$$A_x = \big\{\omega \in \Omega : X(\omega) \leq x\big\} \quad y \quad B_y = \big\{\omega \in \Omega : Y(\omega) \leq y\big\}$$

son independientes. Entonces, para todo $x, y \in \mathbb{R}$,

$$F(x, y) = \mathbb{P}\left(X \leq x, Y \leq y\right) = \mathbb{P}\left(A_x \cap B_y\right) \stackrel{\text{ind}}{=} \mathbb{P}\left(A_x\right)\mathbb{P}\left(B_y\right) = F_X(x)F_Y(y)$$

Además,

$$\begin{aligned}
f(x, y) &= \frac{\partial^2}{\partial x \partial y}F(x, y) \\
&\stackrel{\text{ind}}{=} \frac{\partial}{\partial x}\left[\frac{\partial}{\partial y}F_X(x)F_Y(y)\right] \\
&= \frac{\partial}{\partial x}F_X(x)\frac{\partial}{\partial y}F_Y(y) \\
&= f_X(x)f_Y(y)
\end{aligned}$$

Proposición : Sea (X,Y) un vector aleatorio con función de distribución conjunta F y marginales F_X, F_Y, y con función de densidad de probabilidad conjunta f y marginales f_X y f_Y. X y Y se dicen **variables aleatorias independientes** si y solo si, para todo $x, y \in \mathbb{R}$.

I. $F(x,y) = F_X(x)F_Y(y)$.

II. $f(x,y) = f_X(x)f_Y(y)$.

Si I) se cumple también II) y viceversa.

Nota: Algunos libros de texto utilizan $\perp$ o $\perp\!\!\!\perp$ para denotar independencia. Así, $X \perp Y$ se leería *X y Y son v. a.'s independientes*.

Lema de factorización. Sea (X,Y) un vector aleatorio con función de densidad conjunta f y marginales f_X y f_Y. Entonces, X y Y con variables aleatorias independientes si y solo si existen dos funciones g y h tales que $f(x,y) = g(x)h(y)$, para todo $x, y \in \mathbb{R}$.

Demostración: Suponga que existen funciones g y h tales $f(x,y) = g(x)h(y)$, para todo $x, y \in \mathbb{R}$ y sean $c = \int_{\mathbb{R}} g(x)dx$ y $d = \int_{\mathbb{R}} h(y)dy$. Luego,

$$cd = \int_{\mathbb{R}} g(x)dx \int_{\mathbb{R}} h(y)dy = \iint_{\mathbb{R}^2} g(x)h(y)dxdy = \iint_{\mathbb{R}^2} f(x,y)dxdy = 1$$

por el teorema de Fubini y el hecho de que $f(x,y)$ es la *f. d. p.* conjunta de (X,Y).

Se tiene además que

$$f_X(x) = \int_{\mathbb{R}} f(x,y)dy = \int_{\mathbb{R}} g(x)h(y)dy = g(x)\int_{\mathbb{R}} h(y)dy = dg(x)$$

y similarmente,

$$f_y(y) = \int_{\mathbb{R}} f(x,y)dy = \int_{\mathbb{R}} g(x)h(y)dy = h(y)\int_{\mathbb{R}} g(x)dx = ch(y)$$

Se sigue que, para todo $x, y \in \mathbb{R}$,

$$f_X(x)f_Y(y) = cdg(x)h(y) = cdf(x,y) = f(x,y)$$

por lo que X y Y son variables aleatorias independientes.

Por otro lado, si X y Y son v. a.'s independientes, se tiene que $f(x,y) = f_X(x)f_Y(y)$, para todo $x, y \in \mathbb{R}$. Sean entonces $g(x) = f_X(x)$ y $h(y) = f_Y(y)$ y se sigue el lema.

Proposición : Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad, X y Y variables aleatorias independientes, g y h funciones tales que $g(X)$ y $h(Y)$ son a su vez variables aleatorias. Entonces, $U = g(X)$ y $V = h(Y)$ son también variables aleatorias independientes.

Demostración: Sean $A, B \in \mathcal{B}(\mathbb{R})$ dos conjuntos de Borel cualesquiera. Entonces, puesto que X y Y son independientes se sigue

$$\begin{aligned}
\mathbb{P}(g(X) \in A, h(Y) \in B) &= \mathbb{P}(X \in g^{-1}(A), Y \in h^{-1}(B)) \\
&= \mathbb{P}(X \in g^{-1}(A)) \cdot \mathbb{P}(Y \in h^{-1}(B)) \\
&= \mathbb{P}(g(X) \in A) \cdot \mathbb{P}(h(Y) \in B)
\end{aligned}$$

Definición : Las v. a. X y Y se dicen **dependientes** si para algún $(x,y) \in \mathbb{R}^2$,

I. $F(x,y) \neq F_X(x)F_Y(y)$.

II. $f(x,y) \neq f_X(x)f_Y(y)$.

Ejemplo : *Dados honestos.* X y Y son *v. a.* independientes pues

$$f(x,y) = \frac{1}{36} = \frac{1}{6}\cdot\frac{1}{6} = f_X(x)f_Y(y), \quad \text{para } x,y = 1,\ldots,6$$

Ejemplo : *Dados cargados.* X y Y son *v. a.* dependientes pues

$$f(3,3) = \frac{2}{42} \neq \frac{1}{36} = \frac{1}{6}\cdot\frac{1}{6} = f_X(3)f_Y(3)$$

Ejemplo : *Canal de comunicación.* $N = mQ + R$. Sean $m, n, q, r \in \mathbb{N}_0$, tal que $n = mq + r$,

$$\begin{aligned} f_N(n) &= f_{QR}(q,r) \\ &= \delta(1-\delta)^{mq+r} \\ &= [1-(1-\delta)^m][(1-\delta)^m]^q \cdot \frac{1}{1-(1-\delta^m)}\delta(1-\delta)^r \\ &= f_Q(q)f_R(r) \end{aligned}$$

Por lo que las *v. a.*'s Q y R son independientes.

Ejemplo :

$$f(x,y) = 2e^{-(x+y)}\mathbb{1}_{\{0\leq x\leq y<\infty\}}(x,y) \neq 2e^{-2x}2e^{-y}(1-e^{-y}) = f_X(x)f_Y(y)$$

Por lo tanto X y Y no son *v. a.*'s independientes.

Que las *v. a.*'s X y Y sean dependiente pudo preverse notando que en el soporte de (X,Y), $\mathcal{F}_{XY} = \big\{(x,y) : 0 \leq x \leq y < \infty\big\}$, el rango de X depende de Y y viceversa.

Ejemplo : Considere $\boldsymbol{X} = (X_1, X_2)$ *v. a.* con función de distribución conjunta F y *f. d. p.* conjunta

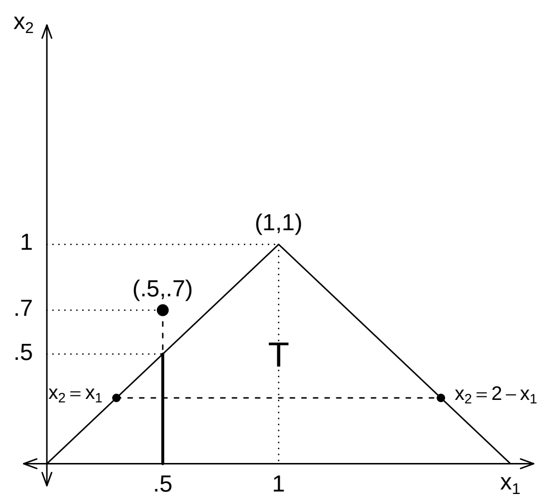
$$f(x_1,x_2) = x_1\mathbb{1}_T(x_1,x_2)$$

donde el soporte es el triángulo T mostrado en la figura.

a). Varifique que $f(x_1,x_2)$ es una *f. d. p.* propia.

b). Calcule $F(.5,.7)$.

c). Determine f_1 y f_2, las *f. d. p.* marginales de X_1 y X_2, respectivamente.



Solución:

a).

$$
\begin{aligned}
\iint_{\mathbb{R}^2} f(x_1,x_2)dx_1dx_2 &= \iint_{\mathbb{R}^2} x_1 \mathbb{1}_T(x_1,x_2)dx_1dx_2 \\
&= \int_0^1 \int_{x_2}^{2-x_2} x_1dx_1 \cdot dx_2 \\
&= \int_0^1 \frac{x_1^2}{2}\Big|_{x_2}^{2-x_2} \cdot dx_2 \\
&= \frac{1}{2}\int_0^1 (4-4x_2)dx_2 \\
&= 1
\end{aligned}
$$

b).

$$
\begin{aligned}
F(.5,.7) &= \mathbb{P}\left(X_1 \le .5, X_2 \le .7\right) \\
&= \mathbb{P}\left(X_1 \le .5, X_2 \le .5\right) \\
&= F(.5,.5) \\
&= \int_0^{.5}\int_0^{x_1} x_1 dx_2 dx_1 \\
&= \frac{1}{24}
\end{aligned}
$$

c). I) ▪ $0 \le x_1 \le 1$,

$$
f_1(x_1) = x_1 \int_0^{x1} dx_2 = x_1^2
$$

▪ $1 \le x_1 \le 2$,

$$
f_1(x_1) = x_1 \int_0^{2-x1} dx_2 = x_1(2-x_1)
$$

Por lo tanto,

$$
f_1(x_1) = x_1^2 \mathbb{1}_{[0,1]}(x_1) + x_1(2-x_1)\mathbb{1}_{(1,2]}(x_1)
$$

II) $0 \le x_2 \le 1$,

$$
f_2(x_2) = \int_{x_2}^{2-x_2} x_1dx_1 = 2(1-x_2)\mathbb{1}_{(0,1])}(x_2)
$$

III) Verifique que las f_1 y f_2 anteriores son funciones de densidad *propias*.

IV) Es claro que las componentes X_1 y X_2 son *dependientes*.

Ejercicio : Calcule $F(x,y)$ para todo $(x,y) \in \mathbb{R}^2$.

Definición : Sean $X_1,\ldots,X_n$, *v.* a.'s con función de distribución marginal $F_{X_1},\ldots,F_{X_n}$, respectivamente, se dicen **variables aleatorias mutuamente independientes**, si para todo $k = 2,\ldots,n$ y subíndices $i_1,\ldots,i_k$, se tiene que su función de distribución conjunta $F_X(x_{i_1},\ldots,x_{i_k})$ es el producto de las correspondientes marginales $F_{X_{i_j}}(x_{i_j})$.. Esto es,

$$
F_X(x_{i_1},\ldots,x_{i_k}) = \prod_{j=1}^{k} F_{X_j}(x_{i_j})
$$

O bien,

$$
f_X(x_{i_1},\ldots,x_{i_k}) = \prod_{j=1}^{k} f_{X_j}(x_{i_j})
$$

donde f corresponde a la función masa de probabilidad o de densidad de probabilidad, según sea el caso.

Definición : Las *v. a.*'s $X_1, \ldots, X_n$ se dicen **variables aleatorias independientes e idénticamente distribuidas** y se denotan por *v.a.i.i.d.*, si son mutuamente independientes y todas tienen la misma distribución en común.

1.6. Ejercicios

Refiérase al Cuaderno de Ejercicios sección 1, Barrios and Chambon (2024).

2. Distribuciones condicionales

Considere el espacio de probabilidad $(\Omega, \mathcal{S}, \mathbb{P})$ y $A \in \mathcal{S}$ con $\mathbb{P}(A) > 0$. Se define la **probabilidad condicional dado el evento** A por

$$\mathbb{P}_A(B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \mathbb{P}\big(B \mid A\big), \qquad \text{para todo } \; B \in \mathcal{S}$$

De la definición anterior se sigue lo que algunos textos llaman la **regla de la multiplicación**: $\mathbb{P}(A \cap B) = \mathbb{P}\big(B \mid A\big)\,\mathbb{P}(A)$.

2.1. Caso discreto

Sea (X, Y) un vector aleatorio discreto con *f. m. p.* conjunta f y marginales f_X y f_Y con soporte S_X y S_Y, respectivamente.

Definición : Se define la **función masa de probabilidad condicional de** X **dado** $Y = y$, por

$$f(x \mid y) = \frac{f(x,y)}{f_Y(y)}, \quad \text{para todo } x \in \mathbb{R}$$

con $y \in \mathcal{S}_Y$, el soporte de Y.

A manera de justificación,

$$f(x \mid y) = \mathbb{P}\big(X = x \mid Y = y\big) = \frac{\mathbb{P}\big(\{X = x\} \cap \{Y = y\}\big)}{\mathbb{P}\big(\{Y = y\}\big)} = \frac{f(x,y)}{f_Y(y)}$$

Así, dado $y_j \in \mathcal{S}_Y$,

$$f(x \mid y_j) = \frac{f(x,y_j)}{f_Y(y_j)}, \qquad \text{para todo } x \in \mathbb{R}$$

De manera similar se tiene,

$$f(y \mid x_i) = \frac{f(x_i,y)}{f_X(x_i)}, \qquad \text{para todo } y \in \mathbb{R}$$

Definición : Se define la **función de distribución condicional** o **función de probabilidad acumulada condicional de** X **dado** $Y = y_j$, por

$$F(x \mid y_j) = \sum_{x_i \le x} f(x_i \mid y_j), \qquad \text{para todo } x \in \mathbb{R}$$

A manera de justificación,

$$\begin{aligned}
F(x \mid y_j) &= \mathbb{P}\big(X \le x \mid Y = y_j\big) \\
&= \sum_{x_i \le x} \frac{\mathbb{P}\big(X = x_i, Y = y_j\big)}{\mathbb{P}(Y = y_J)} \\
&= \sum_{x_i \le x} \frac{f(x_i, y_j)}{f_Y(y_j)} \\
&= \sum_{x_i \le x} f(x_i \mid y_j)
\end{aligned}$$

Ejemplo : *Dados cargados.*

$$f(3 \mid Y = 2) = \mathbb{P}\big(X = 3 \mid Y = 2\big) = \frac{\mathbb{P}(X = 3, Y = 2)}{\mathbb{P}(Y = 2)} = \frac{1/42}{7/42} = \frac{1}{7}$$

$$f(2 \mid Y=2)=\mathbb{P}\big(X=2 \mid Y=2\big)=\frac{\mathbb{P}\left(X=2,Y=2\right)}{\mathbb{P}\left(Y=2\right)}=\frac{2/42}{7/42}=\frac{2}{7}$$

En general, la función masa de probabilidad condicional dado $Y=2$ es

$$f(x \mid Y=2)=\frac{1}{7}\mathbb{1}_{\{1,3,4,5,6\}}(x)+\frac{2}{7}\mathbb{1}_{\{2\}}(x)$$

o bien, en forma de tabla

x	1	2	3	4	5	6	
$f(x \mid 2)$	$\frac{1}{7}$	$\frac{2}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{7}$	1

Proposición : Sea (X,Y) un *v. a.* de componentes independientes. Entonces, si $y_j \in \mathcal{S}_Y$ y para todo $x \in \mathbb{R}$,

$$\begin{aligned} f(x \mid y_j) &= f_X(x) \\ F_X(x \mid y_j) &= F_X(x) \end{aligned}$$

Demostración: X y Y *v. a.*'s independientes, entonces $f(x,y)=f_X(x)f_Y(y)$ para todo $x,y\in\mathbb{R}$. Luego, para $y_j \in \mathcal{S}_Y$,

$$f(x|y_j)=\frac{f(x,y_j)}{f_Y(y_j)} \overset{\text{ind}}{=} \frac{f_X(x)f_Y(y_j)}{f_Y(y_j)}=f_X(x)$$

Se sigue que

$$F(x \mid y_j)=\sum_{x_i\le x} f(x_i \mid y_j) \overset{\text{ind}}{=} \sum_{x_i\le x} f_X(x_i)=F_X(x)$$

Ejemplo : *Canal de comunicación*. Para todo $q=0,1,\ldots,$

$$f(r|q) \overset{\text{ind}}{=} f_R(r)=\frac{1}{1-(1-\delta)^m}\cdot \delta(1-\delta)^r\mathbb{1}_{\{0,1,\ldots,m-1\}}(r)$$

Ejemplo : Sean $X \sim \mathrm{Po}(\lambda_1)$, $Y \sim \mathrm{Po}(\lambda_2)$, *v. a.*'s independientes y sea $Z=X+Y$. Determine la distribución condicional de X dado $X+Y=z$.

Solución:

I). Se determina la *f. m. p.* de $Z=X+Y$ con soporte $\mathcal{S}_Z=\{0,1,\ldots\}$. Sea $z\in\mathcal{S}_Z$,

$$\begin{aligned} \mathbb{P}\left(Z=z\right) &= \mathbb{P}\left(X+Y=z\right) \\ &= \sum_{x=0}^{z}\mathbb{P}\left(X=x,Y=z-x\right) \\ &\overset{\text{ind}}{=} \sum_{x=0}^{z}\mathbb{P}\left(X=x\right)\mathbb{P}\left(Y=z-x\right) \\ &= \sum_{x=0}^{z}\frac{\lambda_1^x e^{-\lambda_1}}{x!}\cdot\frac{\lambda_2^{z-x}e^{-\lambda_2}}{(z-x)!} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}}{z!}\sum_{x=0}^{z}\frac{z!}{x!(z-x)!}\lambda_1^x\lambda_2^{z-x} \\ &= \frac{(\lambda_1+\lambda_2)^z e^{-(\lambda_1+\lambda_2)}}{z!} \end{aligned}$$

Note que los términos de la última suma son $\binom{z}{x}\lambda_1^x\lambda_2^{z-x}$, por lo que se sigue del *teorema del binomio* que la suma total es $(\lambda_1+\lambda_2)^z$ y de ahí la expresión final. Por lo tanto, $Z\sim\mathrm{Po}(\lambda_1+\lambda_2)$.

II). Sea $z = 0, 1, \ldots$ y $x = 0, 1, \ldots, z$. Luego, para $x = 0, 1, \ldots, z$,

$$\begin{aligned}
\mathbb{P}\big(X = x \mid X + Y = z\big) &= \frac{\mathbb{P}(X = x, Y = z - x)}{\mathbb{P}(X + Y = z)} \\
&\overset{\text{ind}}{=} \frac{\mathbb{P}(X = x)\,\mathbb{P}(Y = z - x)}{\mathbb{P}(Z = z)} \\
&= \frac{\lambda_1^x e^{-\lambda_1}/x! \cdot \lambda_2^{z-x} e^{-\lambda_2}/(z-x)!}{(\lambda_1 + \lambda_2)^z e^{-(\lambda_1+\lambda_2)}/z!} \\
&= \frac{z!}{x!(z-x)!}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^x \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{z-x}
\end{aligned}$$

Por lo que para $z = 0, 1, \ldots,$

$$(X \mid X + Y = z) \sim \text{Bin}\left(z, \frac{\lambda_1}{\lambda_1 + \lambda_2}\right)$$

Esto es, condicionado a $X+Y = z$, X sigue una distribución binomial con parámetros, z y probabilidad de éxito $\frac{\lambda_1}{\lambda_1+\lambda_2}$. De manera similar,

$$(Y \mid X + Y = z) \sim \text{Bin}\left(z, \frac{\lambda_2}{\lambda_1 + \lambda_2}\right)$$

Ejemplo : Sean X, Y **variables aleatorias independientes e idénticamente distribuidas (*v.a.i.i.d.*)** geométricamente con probabilidad de éxito p. Entonces, la distribución condicional de X dado $X + Y = z$ es uniforme en $\{0, 1, \ldots, z\}$.

Solución:

I). Para $Z = X + Y$ su soporte es $\mathcal{S}_Z = \{0, 1, \ldots\}$. Sea $q = 1 - p$ y $z = 0, 1, \ldots,$ luego

$$\begin{aligned}
\mathbb{P}(X + Y = z) &= \sum_{x=0}^{z} \mathbb{P}(X = x, Y = z - x) \\
&\overset{\text{ind}}{=} \sum_{x=0}^{z} \mathbb{P}(X = x)\,\mathbb{P}(Y = z - x) \\
&= \sum_{x=0}^{z} pq^x \cdot pq^{z-x} \\
&= p^2 q^z \sum_{x=0}^{z} 1 \\
&= (z+1)p^2 q^z
\end{aligned}$$

Verifique que $f_Z(z) = (z+1)p^2q^z\mathbb{1}_{\{0,1,\ldots\}}(z)$, es una *f. m. p. propia*.

II).

$$\begin{aligned}
f(x \mid z) &= \frac{f(x, z - x)}{f_Z(z)} \\
&\overset{\text{ind}}{=} \frac{f_X(x) f_Y(z - x)}{f_Z(z)} \\
&= \frac{pq^x\; pq^{z-x}}{(z+1)p^2q^z} \\
&= \frac{1}{z+1}\mathbb{1}_{\{0,1,\ldots,z\}}(x)
\end{aligned}$$

que corresponde a la *f. m. p.* de una distribución uniforme (discreta) en los puntos $\{0, 1, \ldots, z\}$.

Ejemplo : *Dados cargados*. Se ha visto ya que

$$f(x \mid 2) = \frac{f(x,2)}{f_Y(2)} = \frac{2}{7}\mathbb{1}_{\{2\}}(x) + \frac{1}{7}\mathbb{1}_{\{1,3,4,5,6\}}(x)$$

Note que

$$\sum_{x=1}^{6} f(x|2) = \frac{1}{7} + \frac{2}{7} + \frac{1}{7} + \cdots + \frac{1}{7} = 1$$

Esto es, $f(x|2)$ es una *f. m. p.* propia. Así, tendría sentido preguntarse por $\mathbb{E}[X \mid Y = 2]$. A saber,

$$\mathbb{E}[X \mid Y = 2] = \mathbb{E}_{Y=2}[X] = \sum_{x=1}^{6} x f(x \mid 2) = 1\frac{1}{7} + 2\frac{2}{7} + \cdots + 6\frac{1}{7} = \frac{23}{7} \approx 3.29$$

De manera similar se puede completar la siguiente tabla

y	$\mathbb{E}[X \mid Y = y]$	$\mathbb{P}(Y = y)$
1	22/7	1/6
2	23/7	1/6
3	24/7	1/6
4	25/7	1/6
5	26/7	1/6
6	27/7	1/6
y	$g(y)$	$f_Y(y)$

Note que $\mathbb{E}[X \mid Y = y] = g(y)$ es una función de $Y = y$. Luego, se puede uno preguntar por $\mathbb{E}[g(Y)]$. Se sigue de la *Ley del Estadístico Inconsciente* (*LEI*),

$$\mathbb{E}[g(Y)] = \sum_{y_j \in \mathcal{S}_Y} g(y_j) f_Y(y_j)$$

Así, en este ejemplo,

$$\begin{aligned}
\mathbb{E}[g(Y)] = \mathbb{E}\left[\mathbb{E}[X \mid Y]\right] &= \sum_{y=1}^{6} \mathbb{E}[X \mid Y = y] f_Y(y_j) \\
&= \sum_{y=1}^{6} \mathbb{E}[X \mid Y = y] \mathbb{P}(Y = y) \\
&= \frac{1}{6}\left[\frac{22}{7} + \cdots + \frac{27}{7}\right] \\
&= \frac{21}{6} \\
&= \mathbb{E}[X]
\end{aligned}$$

El resultado anterior es un caso particular del teorema siguiente:

Proposición : Sean X y Y *v. a.*'s. Entonces,

$$\mathbb{E}[X] = \mathbb{E}\left[\mathbb{E}[X \mid Y]\right]$$

Demostración: Más adelante se demuestra el caso general.

Ejemplo : Sean $X \sim \text{Po}(\lambda_1)$, $Y \sim \text{Po}(\lambda_2)$, *v. a.*'s independientes. Entonces, $Z = X + Y \sim \text{Po}(\lambda_1 + \lambda_2)$. Sea sabe además que $(X \mid X + Y = z) \sim \text{Bin}(z, p)$, con $p = \frac{\lambda_1}{\lambda_1 + \lambda_2}$. Se sigue que $\mathbb{E}[X \mid Z = z] = zp$, por lo que

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}[X \mid Z]\right] &= \mathbb{E}[Zp] \\ &= p\mathbb{E}[Z] \\ &= p(\lambda_1 + \lambda_2) \\ &= \frac{\lambda_1}{\lambda_1 + \lambda_2}(\lambda_1 + \lambda_2) \\ &= \lambda_1 \\ &= \mathbb{E}[X]\end{aligned}$$

Ejemplo : Sean X_1, X_2 variables aleatorias independientes e idénticamente distribuidas (*v.a.i.i.d.*) geométricamente con probabilidad de éxito p. Entonces, $Z = X_1 + X_2$ tiene una *f. m. p.* dada por $f_Z(z) = (z+1)p^2 q^z \mathbb{1}_{\{0,1,\dots\}}(z)$ y $(X_i \mid Z = z)$ se distribuye uniformemente en $\{0, 1, \dots, z\}$.

Luego, $\mathbb{E}[X_1 \mid Z = z] = z/2$. Por lo que

$$\mathbb{E}[X_1] = \mathbb{E}\left[\mathbb{E}[X_1 \mid Z]\right] = \frac{1}{2}\mathbb{E}[Z] = \frac{1}{2}\left(\frac{2q}{p}\right) = \frac{q}{p}$$

Verifique que efectivamente $\mathbb{E}[Z] = 2q/p$.

Teorema de Probabilidad Total (TPT) Sea (X, Y) *v. a.* discreto con *f. m. p.* conjunta f y marginales f_X y f_Y con respectivos soportes $\mathcal{S}_X$ y $\mathcal{S}_Y$. Se tiene entonces,

$$f_X(x) = \sum_{(x,y_j) \in \mathcal{S}_{XY}} f(x \mid y_j) f_Y(y_j)$$

Demostración: Sea $\mathcal{S}_{\mathrm{Y}} = \{y_1, y_2, \dots\}$ el soporte de Y. Los eventos $\{Y = y_j\}$ forman una partición del espacio muestral pues son eventos ajenos cuya unión es todo Ω. Entonces, el teorema se sigue del Teorema de Probabilidad Total para eventos. A saber, para todo $x \in \mathbb{R}$,

$$\begin{aligned}\mathbb{P}\left(\{X = x\}\right) &= \sum_{(x,y)_j \in \mathcal{S}_{\mathrm{Y}}} \mathbb{P}\left(\{X = x\} \cap \{Y = y_j\}\right) \\ &= \sum_{(x,y_j) \in \mathcal{S}_{\mathrm{XY}}} \mathbb{P}\left(X = x \mid Y = y_j\right) \mathbb{P}\left(Y = y_j\right) \\ f_X(x) &= \sum_{(x,y_j) \in \mathcal{S}_{\mathrm{XY}}} f(x \mid y_j) f_Y(y_j)\end{aligned}$$

Regla de Bayes. Sea $y_k \in \mathcal{S}_{\mathrm{Y}}$, entonces

$$f(y_k \mid x) = \frac{f(x, y_k)}{f_X(x)} = \frac{f(x \mid y_k) f_Y(y_k)}{\sum_{(x,y_j) \in \mathcal{S}_{\mathrm{XY}}} f(x \mid y_j) f_Y(y_j)}$$

Demostración: El numerador se sigue de la regla de la multiplicación y el denominador del TPT.

2.2. Caso continuo

Proposición : Sea (X, Y) un *v. a.* continuo con función de densidad de probabilidad conjunta f y marginales f_X y f_Y, respectivamente. Entonces, para $a < b$,

$$\mathbb{P}\left(a \le X \le b \mid Y = y\right) = \int_a^b f(x \mid y) dx$$

donde $f(x \mid y) = f_{X|Y=y}(x) = \dfrac{f(x,y)}{f_Y(y)}$.

Demostración: Para $\delta > 0$ *pequeño*,

$$\begin{aligned}
\mathbb{P}\big(a \le X \le b \mid Y = y\big) &\approx \mathbb{P}\big(a \le X \le b \mid y - \delta \le Y \le y + \delta\big) \\
&= \frac{\mathbb{P}\left(a \le X \le b, y - \delta \le Y \le y + \delta\right)}{\mathbb{P}\left(y - \delta \le Y \le y + \delta\right)} \\
&= \frac{\int_a^b \int_{y-\delta}^{y+\delta} f(u,v)dvdu}{\int_{y-\delta}^{y+\delta} f_Y(v)dv}
\end{aligned}$$

y haciendo $\delta \to 0^+$,

$$\begin{aligned}
\lim_{\delta\to 0^+} \frac{\mathbb{P}\left(a \le X \le b, y - \delta \le Y \le y + \delta\right)}{\mathbb{P}\left(y - \delta \le Y \le y + \delta\right)} &= \lim_{\delta\to 0^+} \frac{\frac{1}{2\delta}\int_a^b \int_{y-\delta}^{y+\delta} f(u,v)dvdu}{\frac{1}{2\delta}\int_{y-\delta}^{y+\delta} f_Y(v)dv} \\
&= \frac{\int_a^b f(x,y)dx}{f_Y(y)} \\
&= \int_a^b \frac{f(x,y)}{f_Y(y)}dx
\end{aligned}$$

suponiendo que $\int_{y-\delta}^{y+\delta} f_Y(v)dv$ es derivable en y, $f_Y(y) > 0$ y que $f(x,y)$ es continua en x.

Resumiendo, para $a < b$,

$$\mathbb{P}\big(a \le X \le b \mid Y = y\big) = \int_a^b f(x \mid y)dx$$

donde $f(x \mid y) = \dfrac{f(x,y)}{f_Y(y)}$ es una *f. d. p.* (con respecto a la integral) para la *f. p. a.* condicional $F(x \mid y) = \mathbb{P}\big(X \le x \mid Y = y\big)$.

Nota: La función de densidad condicionada en un punto es de las aportaciones fundamentales a la Teoría de Probabilidades de A. N. Kolmogorov (1938).

Definición : Sea (X,Y) un *v. a.* Se define la *f. p. a.* de X dado $Y = y$ por

$$F(x \mid y) = \mathbb{P}\big(X \le x \mid Y = y\big) = \int_{-\infty}^x f(u \mid y)du$$

para todo $x \in \mathbb{R}$ y donde $f(x \mid y) = \dfrac{f(x,y)}{f_Y(y)}$ es la *f. d. p.* condicional de X dado $Y = y$.

Nota: la *f. d. p.* condicional de X dado $Y = y$ es una función de densidad propia pues $f(x \mid y) \ge 0$ para todo $x \in \mathbb{R}$ y

$$\int_{\mathbb{R}} f(x \mid y)dx = \int_{-\infty}^{\infty} \frac{f(x,y)}{f_Y(y)}dx = \frac{1}{f_Y(y)} \int_{-\infty}^{\infty} f(x,y)dx = 1$$

En general, si $A \in \mathcal{B}(\mathbb{R})$,

$$\mathbb{P}_{Y=y}\left(X \in A\right) = \mathbb{P}\big(X \in A \mid Y = y\big) = \int_A f(x \mid y)dx$$

Proposición : Sea (X,Y) un *v. a.* con *f. d. p.* conjunta f, marginales f_X, f_Y y condicionales $f(x|y), f(y|x)$. Se tiene entonces la **regla de la multiplicación**

$$f(x,y) = f(x \mid y)f_Y(y) = f(y \mid x)f_X(x)$$

Ejemplo : Sea (X, Y) *v. a.* con *f. d. p.* conjunta f dada por

$$f(x,y) = \frac{1}{y} e^{-\frac{1}{y}(x+y^2)} \mathbb{1}_{(0,\infty)}(x) \mathbb{1}_{(0,\infty)}(y)$$

Encuentre $\mathbb{P}\big(X > 1 \mid Y = y\big)$ y verifique que $\mathbb{P}\big(X > 1 \mid Y = 2\big) = e^{-1/2}$.

Solución:

$$f_Y(y) = \int_{\mathbb{R}} f(x,y) dx = \frac{1}{y} \int_0^\infty e^{-\frac{x}{y} - y} dx = e^{-y} \int_0^\infty \frac{1}{y} e^{-\frac{1}{y} x} dx = e^{-y}$$

Luego, para todo $y > 0$,

$$f(x \mid y) = \frac{\frac{1}{y} e^{-\frac{1}{y}(x+y^2)}}{e^{-y}} = \frac{1}{y} e^{-\frac{1}{y} x} \mathbb{1}_{(0,\infty)}(x),$$

Así, $(X \mid Y = y)$ se distribuye exponencialmente con media y, esto es, $\mathbb{E}[X \mid Y = y] = y$. Entonces,

$$\mathbb{P}\big(X > 1 \mid Y = y\big) = e^{-\frac{1}{y}(1)} = e^{-\frac{1}{y}}, \quad y > 0$$

y

$$\mathbb{P}\big(X > 1 \mid Y = 2\big) = e^{-1/2} = 0.6065$$

Teorema de Probabilidad Total (TPT). Sea (X, Y) un *v. a.* con *f. d. p.* conjunta f y marginales f_X y f_Y. Entonces,

$$f_X(x) = \int_{\mathbb{R}} f(x \mid y) f_Y(y) dy, \qquad \text{para todo } x \in \mathbb{R}$$

Y similarmente,

$$f_Y(y) = \int_{\mathbb{R}} f(y \mid x) f_X(x) dx, \qquad \text{para todo } y \in \mathbb{R}$$

Demostración: La proposición se sigue del hecho $f_X(x) = \int_{\mathbb{R}} f(x,y) dy$ y de la regla de la multiplicación, $f(x,y) = f(x \mid y) f_X(x)$.

Nota: El teorema se cumple también en el caso de un vector aleatorio mixto. Por ejemplo, X un *v. a.* discreta con soporte $\mathcal{S}_{\mathrm{X}}$ y Y una *v. a.* continua. Así,

$$f_X(x) = \mathbb{P}(X = x) = \int_{\mathbb{R}} f(x \mid y) f_Y(y) dy, \qquad \text{para todo } x \in \mathbb{R}$$

y

$$f_Y(y) = \sum_{(x_i, y) \in \mathcal{S}_{\mathrm{XY}}} f(y \mid x_i) f_X(x_i), \qquad \text{para todo } y \in \mathbb{R}$$

Regla de Bayes. Sea (X, Y) un *v. a.* con *f. d. p.* marginales f_X y f_Y y condicionales $f(x \mid y)$ y $f(y \mid x)$. Entonces, si Y es una *v. a.* continua, para todo $y \in \mathbb{R}$,

$$f(y \mid x) = \frac{f(x \mid y) f_Y(y)}{\int_{\mathbb{R}} f(x \mid v) f_Y(v) dv}$$

O bien, si Y es una *v. a.* discreta,

$$f(y \mid x) = \frac{f(x \mid y) f_Y(y)}{\sum_{(x, y_j) \in \mathcal{S}_{\mathrm{XY}}} f(x \mid y_j) f_Y(y_j)}$$

para toda $Y \in \mathbb{R}$.

Ejemplo : Suponga que N el número de accidentes de un automóvil en un año sigue una distribución Poisson con media $\lambda(>0)$ donde λ depende de la persona. Se elije una persona al azar y suponga que a su vez la media de accidentes, $\Lambda=\lambda$, se puede modelar mediante la distribución Gamma con parámetros de forma α y de escala β.

a). Determine la distribución marginal de N.

b). Determine la distribución condicional de Λ dado $N=n$.

Solución: Note que se tiene N una *v. a.* discreta $(N \mid \Lambda=\lambda) \sim \mathrm{Po}(\lambda)$ y Λ una *v. a.* continua con $\lambda \sim \mathrm{Gamma}(\alpha,\beta)$.

a). Sea $n=0,1,\ldots,$

$$\begin{aligned}
f_N(n) &\overset{\text{TPT}}{=} \int_{\mathbb{R}} f(n \mid \lambda) f_\Lambda(\lambda) d\lambda \\
&= \int_0^\infty \frac{\lambda^n e^{-\lambda}}{n!} \frac{\lambda^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} e^{-\lambda/\beta} d\lambda \\
&= \frac{1}{\Gamma(n+1)\beta^\alpha\Gamma(\alpha)} \cdot \frac{1}{K} \int_0^\infty K\, \lambda^{(n+\alpha)-1} e^{-(1+1/\beta)\lambda} d\lambda \\
&= \frac{1}{\Gamma(n+1)\beta^\alpha\Gamma(\alpha)} \cdot \frac{\Gamma(n+\alpha)}{(1+1/\beta)^{n+\alpha}} \cdot 1 \\
&= \frac{(n+\alpha-1)!}{n!(\alpha-1)!} \left(\frac{1}{1+\beta}\right)^\alpha \left(1-\frac{1}{1+\beta}\right)^n \\
&= \binom{n+\alpha-1}{n} p^\alpha q^n
\end{aligned}$$

con $K=\dfrac{(1+1/\beta)^{n+\alpha}}{\Gamma(n+\alpha)}$ es la constante normalizadora de la densidad Gamma con núcleo $\lambda^{(n+\alpha)-1}e^{-(1+1/\beta)\lambda}$ y donde $p=\frac{1}{1+\beta}$ y $q=1-p$. Por lo tanto, marginalmente N sigue una distribución binomial negativa con parámetros α y $1/(1+\beta)$. Esto es, $N \sim \mathrm{BinNeg}\left(\alpha, \dfrac{1}{1+\beta}\right)$.

b). Sea $\lambda>0$,

$$\begin{aligned}
f(\lambda \mid n) &\overset{\text{Bayes}}{=} \frac{f(n \mid \lambda) f_\Lambda(\lambda)}{f_N(n)} \\
&= \frac{\frac{\lambda^n e^{-\lambda}}{\Gamma(n+1)} \cdot \frac{\lambda^{\alpha-1}e^{-\lambda/\beta}}{\beta^\alpha\Gamma(\alpha)}}{\frac{\Gamma(n+\alpha)}{\Gamma(n+1)\Gamma(\alpha)} \left(\frac{1}{1+\beta}\right)^\alpha \left(\frac{\beta}{1+\beta}\right)^n} \\
&= \frac{\left(\frac{1+\beta}{\beta}\right)^{n+\alpha} \lambda^{n+\alpha-1}}{\Gamma(n+\alpha)}\, e^{-\frac{1+\beta}{\beta}\lambda}
\end{aligned}$$

Por lo que condicionalmente dado $N=n$, Λ sigue una distribución Gamma con parámetro de forma $(n+\alpha)$ y parámetro tasa $\frac{1+\beta}{\beta}$, o parámetro de escala $\frac{\beta}{1+\beta}$.

2.3. Esperanza y Varianza Condicional

Definición Se define la **desviación cuadrática media** (DCM) de una variable aleatoria Z respecto a un valor constante θ, por

$$\mathrm{DCM}(Z,\theta) = \mathbb{E}[(Z-\theta)^2]$$

Proposición : Sea Z una variable aleatoria con media y varianza finitas, y sea θ una constante. Entonces,

$$\mathrm{DCM}(Z,\theta) = \mathrm{var}(Z) + (\mathbb{E}[Z]-\theta)^2$$

Demostración

Por facilitar la notación, sea $\mathbb{E}Z = \mathbb{E}[Z]$. Luego,

$$\begin{aligned}
\mathrm{DCM}(Z,\theta) &= \mathbb{E}[(Z-\theta)^2] \\
&= \mathbb{E}\left[\left\{(Z-\mathbb{E}Z)+(\mathbb{E}Z-\theta)\right\}^2\right] \\
&= \mathbb{E}[(Z-\mathbb{E}Z)^2] + \mathbb{E}[(\mathbb{E}Z-\theta)^2] + 2\mathbb{E}\left[(Z-\mathbb{E}Z)(\mathbb{E}Z-\theta)\right] \\
&= \mathbb{E}[(Z-\mathbb{E}Z)^2] + (\mathbb{E}Z-\theta)^2 + 2\left[(\mathbb{E}Z)^2 - \theta\mathbb{E}Z - (\mathbb{E}Z)^2 + \theta\mathbb{E}Z\right] \\
&= \mathbb{E}[(Z-\mathbb{E}Z)^2] + (\mathbb{E}Z-\theta)^2 \\
&= \mathrm{var}(Z) + (\mathbb{E}Z-\theta)^2
\end{aligned}$$

En palabras, lo que la proposición anterior dice que la desviación cuadrática media de una variable aleatoria es igual a la varianza de la *v. a.* más el cuadrado del **sesgo** (desviación media a la constante).

Definición Considere X y Y variables aleatorias (*v. a.*) continuas con $f(x,y)$, $f(y|x)$, y $f_Y(y)$, la función de densidad probabilidad (*f. d. p.*) conjunta, la *f. d. p.* condicional de Y dado $X=x$, y la *f. d. p.* marginal de Y, respectivamente. Se define la **esperanza condicional de Y dado $X=x$**, por

$$\mathbb{E}_x[Y] = \mathbb{E}[Y|X=x] = \int_{\mathbb{R}} y f(y|x)dy$$

O bien, si Y es una variable aleatoria discreta

$$\mathbb{E}_x[Y] = \mathbb{E}[Y|X=x] = \sum_{y_j\in\mathcal{S}_Y} y_j f(y_j|x)$$

Note en las expresiones anteriores $\mathbb{E}_x[Y]$ resulta ser función del valor real x, realización de la *v. a.* X ($X(w)=x$), e.g., $h(x) = \mathbb{E}[Y|X=x]$. Así, $h(X) = \mathbb{E}_X[Y] = \mathbb{E}[Y|X]$ viene a ser ella misma una variable aleatoria y como tal podemos calcular su valor esperado $\mathbb{E}[h(X)]$.

Proposición : Sean X y Y variables aleatorias. Entonces,

$$\mathbb{E}[Y] = \mathbb{E}\big[\mathbb{E}[Y|X]\big]$$

Demostración Sin pérdida de generalidad considere que las *v. a.*'s son continuas. Ahora, $\mathbb{E}[Y|X]$ es una función de la *v. a.* X con *f. d. p.* marginal f_X, luego

$$\begin{aligned}
\mathbb{E}\big[\mathbb{E}[Y|X]\big] &\overset{LEI}{=} \int_{\mathbb{R}} \mathbb{E}_x[Y] f_X(x)dx \\
&= \int_{\mathbb{R}} \mathbb{E}[Y|X=x] f_X(x)dx \\
&= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} y f(y|x)dy\right] f_X(x)dx \\
&= \int_{\mathbb{R}} y \left[\int_{\mathbb{R}} f(y|x) f_X(x)dx\right] dy \\
&= \int_{\mathbb{R}} y \left[\int_{\mathbb{R}} f(x,y)dx\right] dy \\
&= \int_{\mathbb{R}} y f_Y(x)dy \\
&= \mathbb{E}[Y]
\end{aligned}$$

Corolario :
$$\mathbb{E}[XY] = \mathbb{E}\left[X\mathbb{E}[Y|X]\right]$$

Demostración:

$$\begin{aligned}
\mathbb{E}\left[X\mathbb{E}[Y|X]\right] &\overset{LEI}{=} \int_{\mathbb{R}} x\mathbb{E}[Y|X=x]\ f_X(x)dx \\
&= \int_{\mathbb{R}} x \int_{\mathbb{R}} yf(y|x)f_X(x)dy\ dx \\
&= \iint_{\mathbb{R}^2} xyf(x,y)dy\,dx \\
&= \mathbb{E}[XY]
\end{aligned}$$

De igual forma que en el caso del valor esperado condicional se define la varianza condicional.

Definición Sean X y Y *v. a.* Se define la **varianza condicional de Y dado $X = x$**, por

$$\mathrm{var}(Y|X=x) = \mathrm{var}_x(Y) = \mathbb{E}_x\left[(Y - E_x[Y])^2\right] = \mathbb{E}\big[(Y - \mathbb{E}[Y|X=x])^2)|X=x\big]$$

y la variable aleatoria **varianza condicional de Y dado X**, por

$$\mathrm{var}(Y|X) = \mathbb{E}\big[(Y - \mathbb{E}[Y|X])^2)|X\big]$$

Proposición : Sean X y Y variables aleatorias. Entonces,

$$\mathrm{var}(Y) = \mathbb{E}[\mathrm{var}(Y|X)] + \mathrm{var}(\mathbb{E}[Y|X])$$

Demostración: Note que el valor esperado de la *v. a.* Y *condicional* a X es $\mathbb{E}[Y|X]$ y no $\mathbb{E}[Y] = \mathrm{E}Y$. Luego tomando $Z = Y$ y $\theta = \mathrm{E}Y$ en la definición de la desviación cuadrática media (condicional) DCM, se sigue de la proposición 1

$$\begin{aligned}
\mathrm{DCM}_X(Y, \mathrm{E}Y) = \mathbb{E}_X[(Y - \mathrm{E}Y)^2] &= \mathbb{E}[(Y - \mathrm{E}Y)^2|X] \\
&= \mathrm{var}(Y|X) + (\mathbb{E}[Y|X] - \mathrm{E}Y)^2
\end{aligned}$$

y tomando el valor esperado de ambos lados

$$\begin{aligned}
\mathbb{E}\left[\mathbb{E}_X[(Y - \mathrm{E}Y)^2]\right] &= \mathbb{E}\left[\mathrm{var}(Y|X) + (\mathbb{E}[Y|X] - \mathrm{E}Y)^2\right] \\
\mathbb{E}\left[\mathbb{E}[(Y - \mathrm{E}Y)^2|X]\right] &= \mathbb{E}[\mathrm{var}(Y|X)] + \mathbb{E}[(\mathbb{E}[Y|X] - \mathrm{E}Y)^2] \\
\mathbb{E}[(Y - \mathrm{E}Y)^2] &= \mathbb{E}\big[\mathrm{var}(Y|X)\big] + \mathrm{var}\big(\mathbb{E}[Y|X]\big)
\end{aligned}$$

pues se sigue de la proposición 2 que $\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y] = \mathrm{E}Y$. Y el lado izquierdo de la igualdad es precisamente la varianza de Y. Luego, se tiene el resultado.

Las dos proposiciones anteriores se resumen en el siguiente teorema.

Teorema : Sean X y Y variables aleatorias, con valores esperados y varianzas finitas. Entonces, se cumple que:

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

ii) $\mathrm{var}(Y) = \mathbb{E}[\mathrm{var}(Y|X)] + \mathrm{var}(\mathbb{E}[Y|X])$.

Ejemplo : Considere que el número de tareas que requieren de un servicio en un intervalo de tiempo $[0, t]$ es una variable aleatoria N_t que sigue una distribución Poisson de parámetro βt. Suponga también que el tiempo T de procesamiento sigue una distribución exponencial con tiempo medio $1/\alpha$. Encuentre el valor esperado y la varianza del número de requerimientos N que arriban mientras se procesa una tarea.

Solución: Sea N_t el número de tareas que arriban en el intervalo $[0, t]$, T el tiempo de procesamiento de una tarea, y N el número de tareas que arriban mientras se procesa una de ellas.

a). $T \sim \text{Exp}(\alpha)$. Además, dado $T = t$, $N_t \sim \text{Po}(\beta t)$. Luego, $\mathbb{E}[N_t|T = t] = \beta t$, y por lo tanto
$$\mathbb{E}[N] = \mathbb{E}\left[\mathbb{E}(N_T|T)\right] = \mathbb{E}[\beta T] = \frac{\beta}{\alpha}$$

b). Recuerde que si Z es una *v. a.* entonces $\mathbb{E}[Z^2] = \text{var}(Z) + \mathbb{E}[Z]^2$. Luego, por el inciso anterior se tiene que
$$\mathbb{E}[N^2] = \mathbb{E}[\mathbb{E}[N_T^2|T]] = \mathbb{E}[\beta T + \beta^2 T^2] = \frac{\beta}{\alpha} + 2\frac{\beta^2}{\alpha^2}$$
pues $\mathbb{E}[T^2] = \text{var}(T) + \mathbb{E}[T]^2 = 1/\alpha^2 + 1/\alpha^2 = 2/\alpha^2$. Entonces,
$$\text{var}(N) = \mathbb{E}[N^2] - \mathbb{E}[N]^2 = \left(\frac{\beta}{\alpha} + 2\frac{\beta^2}{\alpha^2}\right) - \frac{\beta^2}{\alpha^2} = \frac{\beta}{\alpha} + \frac{\beta^2}{\alpha^2}$$

c). Por otro lado, puesto que dado $T = t$, N_t sigue una Poisson parámetro βt, se tiene que $\text{var}(N_t|T = t) = \beta t$, y por lo tanto
$$\text{var}(N_T|T) = \beta T$$
y aplicando *ii)* del teorema
$$\begin{aligned}
\text{var}(N) &= \mathbb{E}[\text{var}(N_T|T)] + \text{var}(\mathbb{E}[N_T|T]) \\
&= \mathbb{E}[\beta T] + \text{var}(\beta T) \\
&= \frac{\beta}{\alpha} + \frac{\beta^2}{\alpha^2}
\end{aligned}$$
que coincide con el inciso anterior.

d). Calculemos ahora la *f. m. p.* de N. Por el teorema de probabilidad total, para $k = 0, 1, \ldots,$
$$\begin{aligned}
P(N = k) &= \int_0^\infty P(N_t = k|T = t) f_T(t) dt \\
&= \int_0^\infty e^{-\beta t} \frac{(\beta t)^k}{k!} \alpha e^{-\alpha t} dt \\
&= \frac{\alpha \beta^k}{(\alpha + \beta)^{k+1}} \int_0^\infty \frac{(\alpha + \beta)^{k+1}}{\Gamma(k+1)} t^k e^{-(\alpha+\beta)t} dt \\
&= \frac{\alpha}{\alpha + \beta} \left(\frac{\beta}{\alpha + \beta}\right)^k
\end{aligned}$$
pues el integrando es la *f. d. p.* de una distribución $\text{Gamma}(t; \alpha + \beta, k + 1)$ y por lo tanto integra a 1. Esto es, el número de tareas N que llegan mientras se procesa una de ellas sigue una distribución geométrica de parámetro $p = \alpha/(\alpha + \beta)$. Por lo tanto, su valor esperado y varianza son

$$\begin{aligned}
\mathbb{E}[N] &= \frac{q}{p} & \text{var}(N) &= \frac{q}{p^2} \\
&= \frac{\beta/(\alpha+\beta)}{\alpha/(\alpha+\beta)} & &= \frac{\beta/(\alpha+\beta)}{\alpha^2/(\alpha+\beta)^2} \\
&= \frac{\beta}{\alpha} & &= \frac{\beta(\alpha+\beta)}{\alpha^2} \\
& & &= \frac{\beta}{\alpha} + \frac{\beta^2}{\alpha^2}
\end{aligned}$$

que coinciden con los resultados obtenidos previamente.

Ejemplo : Sea $X \sim \text{unif}(0,1)$ y $Y \sim \text{unif}(0,X)$. Encuentre f_Y, la *f. d. p.* marginal de Y, $\mathbb{E}[Y]$ y $\text{var}(Y)$.

Solución: Note que $Y \sim \text{unif}(0,X)$ define la distribución condicional de Y dado X. Abusando de la notación, $(Y|X=x) \sim \text{unif}(0,x)$.

a). Se sigue del Teorema de Probabilidad Total que para $0 < y < 1$,

$$\begin{aligned}
f_Y(y) &= \int_{\mathbb{R}} f(y|x)\cdot f_X(x)dx \\
&= \int_{\mathbb{R}} \frac{1}{x}\mathbb{1}_{(0,x)}(y)\cdot 1\,\mathbb{1}_{(0,1)}(x)dx \\
&= \int_0^1 \frac{1}{x}\mathbb{1}_{(0,x)}(y)dx \\
&= \int_y^1 \frac{1}{x}dx
\end{aligned}$$

Por lo tanto,

$$f_Y(y) = -\log(y)\mathbb{1}_{(0,1)}(y)$$

Note que f_Y es una *f. d. p.* legítima pues $f_Y(y) \geq 0$, y

$$\int_{\mathbb{R}} f_Y(y)dy = \int_0^1 -\log(y)dy = -\big[y\log(y) - y\big]_0^1 \longrightarrow -\big[0-1+0+0\big] = 1$$

b).

$$\mathbb{E}[Y] = \int_{\mathbb{R}} yf_Y(y)dy = \int_0^1 -y\log(y)dy = -\frac{y^2}{2}\big(\log(y) - \frac{1}{2}\big)\Big|_0^1 \longrightarrow -[0-\frac{1}{4}-0+0) = \frac{1}{4}$$

c).

$$\mathbb{E}[Y^2] = \int_{\mathbb{R}} y^2 f_Y(y)dy = \int_0^1 -y\log(y)dy = -\frac{y^3}{3}\big(\log(y) - \frac{1}{3}\big)\Big|_0^1 \longrightarrow -[0-\frac{1}{9}-0+0) = \frac{1}{9}$$

De donde,

$$\text{var}(Y) = \mathbb{E}[Y^2] - \mathbb{E}^2[Y] = \frac{1}{9} - \frac{1}{16} = \frac{7}{144}$$

d). Por otro lado, utilizado las expresiones del teorema para la media y la varianza condicional

i)

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[X/2] = \frac{1}{2}\mathbb{E}[X] = \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}$$

ii)

$$\begin{aligned}
\mathrm{var}[Y] &= \mathrm{var}(\mathbb{E}[Y|X]) + \mathbb{E}[\mathrm{var}(Y|X)] \\
&= \mathrm{var}(X/2) + \mathbb{E}[X^2/12] \\
&= \frac{1}{4}\cdot\frac{1}{12} + \frac{1}{12}(\frac{1}{12} + \frac{1}{4}) \\
&= \frac{7}{144}
\end{aligned}$$

2.4. Ejercicios

Refiérase al Cuaderno de Ejercicios sección 2, Barrios and Chambon (2024).

3. Esperanza y covarianzas de vectores aleatorios

3.1. Recordar ...

Proposición : Sea X una variable aleatoria definida en el espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$.

a). Si c es una constante tal que $\mathbb{P}(X = c) = 1$. entonces $\mathbb{E}[X] = c$.

b). Si c es un constante y X tiene valor esperado finito, entonces $\mathbb{E}[cX] = c\,\mathbb{E}[X]$.

El siguiente par de teoremas, para los casos univariado y bivariado, se presentan en algunos textos como el **Teorema del Estadístico Inconsciente (**TEI**)**. (Vea por ejemplo, Blitzstein and Hwang (2014))

Proposición : Sea X una *v. a.* y g una función real (*medible*) tal que $g(X)$ es una variable aleatoria. Entonces

$$\mathbb{E}[g(X)] = \begin{cases} \displaystyle\sum_{x_i \in \mathcal{S}_\mathrm{X}} g(x_i)\mathbb{P}(X = x_i) & \text{caso discreto} \\ \displaystyle\int_{\mathbb{R}} g(x) f_X(x) dx & \text{caso continuo} \end{cases}$$

Demostración: Vea Rincón (2014)

Considere ahora (X, Y) un vector aleatorio definido en el *EP* $(\Omega, \mathcal{F}, \mathbb{P})$. Entonces, con los mismos argumentos que en el caso univariado se cumple la TEI.

Proposición : Sea (X, Y) un *v. a.* definido sobre el *EP* $(\Omega, \mathcal{F}, \mathbb{P})$ y sea g una función real medible tal que $Z = g(X, Y)$ es una variable aleatoria, entonces

$$\mathbb{E}[Z] = \mathbb{E}[g(X,Y)] = \sum_{(x_i, y_j) \in \mathcal{S}_\mathrm{XY}} g(x_i, y_j)\mathbb{P}\left(X = x_i, Y = y_j\right)$$

si X y Y *v. a.*'s discretas. Si las *v. a.*'s son continuas con f su *f. d. p.* conjunta, se tiene que

$$\mathbb{E}[Z] = \mathbb{E}[g(X,Y)] = \int_{\mathbb{R}^2} g(x, y) f(x, y) dx dy$$

Proposición : Sean X y Y *v. a.*'s con valor esperado finito.

a). Si $X + Y$ tiene valor esperado finito, entonces $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.

b). Si $\mathbb{P}(X \geq Y) = 1$, entonces $\mathbb{E}[X] \geq \mathbb{E}[Y]$. Aún más, si $\mathbb{E}[X] = \mathbb{E}[Y]$, entonces $\mathbb{P}(X = Y) = 1$.

c). $\mid \mathbb{E}[X] \mid \leq \mathbb{E}[\mid X \mid]$.

Demostración:

a). Sea $Z = g(X,Y) = X + Y$ y sea f la *f. d. p.* conjunta de (X,Y). Entonces,

$$\begin{aligned}
\mathbb{E}[X+Y] &= \mathbb{E}[g(X,Y)] \\
&= \iint_{\mathbb{R}^2} g(x,y)f(x,y)dxdy \\
&= \iint_{\mathbb{R}^2} (x+y)f(x,y)dxdx \\
&= \int_{\mathbb{R}} x\left[\int_{\mathbb{R}} f(x,y)dy\right]dx + \int_{\mathbb{R}} y\left[\int_{\mathbb{R}} f(x,y)dx\right]dy \\
&= \int_{\mathbb{R}} xf_X(x)dx + \int_{\mathbb{R}} yf_Y(y)dy \\
&= \mathbb{E}[X] + \mathbb{E}[Y]
\end{aligned}$$

b). Por facilidad, suponga que X y Y son *v. a.*'s dicretas. Entonces $Z = X - Y$ también es discreta y

$$\mathbb{E}[X] - \mathbb{E}[Y] = \mathbb{E}[X-Y] = \mathbb{E}[Z] = \sum_{z_i} z_i f_Z(z_i)$$

pero como $\mathbb{P}(Z \geq 0) = \mathbb{P}(X \geq Y) = 1$, entonces, $z_i \geq 0$, para todo z_i en $\mathcal{S}_Z$, el soporte de la *v. a.* Z y por lo tanto $\sum_{z_i} z_i f_Z(z_i) \geq 0$.

Por otro lado, si $\mathbb{E}[X] = \mathbb{E}[Y]$, entonces, $\mathbb{E}[Z] = 0 = \sum_{z_i} z_i f_Z(z_i)$. Pero la suma es cero solamente si todos los sumandos (no negativos) son 0. Esto es, si y solo si $z_i f_Z(z_i) = 0$, para todo z_i, por lo que $z_i = 0$. Esto es, el único valor posible de Z es $Z = 0$. Por lo tanto, $\mathbb{P}(Z = 0) = 1 = \mathbb{P}(X = Y)$.

c). Note que $-|X| \leq X \leq |X|$. Luego, $-\mathbb{E}|X| \leq \mathbb{E}[X] \leq \mathbb{E}|X|$. Por lo tanto $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$.

Ejemplo : *Dados cargados*

i).

$$\begin{aligned}
\mathbb{E}[X+Y] &= \sum_{x=1}^{6}\sum_{y=1}^{6}(x+y)\mathbb{P}(X=x, Y=y) \\
&= \Big\{\big[1 + (2 \cdots + 6)\big] + \cdots + \big[6 + (1 + \cdots + 5)\big]\Big\}\frac{1}{42} \\
&\quad + \big\{2(1 + \cdots + 6)\big\}\frac{2}{42} \\
&= \frac{294}{42} = 2\left(\frac{7}{2}\right) = \frac{7}{2} + \frac{7}{2} \\
&= \mathbb{E}[X] + \mathbb{E}[Y]
\end{aligned}$$

ii). Sin embargo, verifique que

$$\mathbb{E}[XY] = \sum_{x=1}^{6}\sum_{y=1}^{6} xy\,\mathbb{P}(X=x, Y=y) = \frac{38}{3} \neq \frac{49}{4} = \mathbb{E}[X]\cdot\mathbb{E}[Y]$$

Proposición : Sea X una *v. a.* con valor esperado finito y M una constante tal que $\mathbb{P}\big(|X| \leq M\big) = 1$. Entonces, $|\mathbb{E}[X]| \leq M$.

Proposición : Sean X y Y *v. a.*'s con media finita. Si X y Y son independientes, entonces $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$.

Demostración: Suponga que (X,Y) es un *v. a.* continuo con *f. d. p.* conjunta f, marginales f_X y f_Y y sea $g(X,Y)=XY$. Se sigue entonces del TEI,

$$\begin{aligned}
\mathbb{E}[XY] &= \mathbb{E}[g(X,Y)] \\
&= \iint_{\mathbb{R}^2} g(x,y) f(x,y) dx dy \\
&= \iint_{\mathbb{R}^2} xy\, f(x,y) dx dy \\
&\overset{\text{ind}}{=} \iint_{\mathbb{R}^2} xy f_X(x) f_Y(y) dx dy \\
&= \int_{\mathbb{R}} x f_X(x) dx \cdot \int_{\mathbb{R}} y f_Y(y) dy \\
&= \mathbb{E}[X]\,\mathbb{E}[Y]
\end{aligned}$$

Proposición : En general, $\mathbb{E}[XY] \neq \mathbb{E}]X]\,\mathbb{E}[Y]$.

Demostración: Revise el ejemplo anterior de los dados cargados.

Nota: $\mathbb{E}[XY] = \mathbb{E}]X]\,\mathbb{E}[Y]$ no implica independencia. En efecto, considere el *v. a.* (X,Y) con *f. m. p.* conjunta dada por la siguiente tabla

$X\backslash Y$	-1	0	$+1$	f_X
-1	$1/3$	0	$1/3$	$2/3$
$+1$	0	$1/3$	0	$1/3$
f_Y	$1/3$	$1/3$	$1/3$	1

Se tiene que

$$\begin{aligned}
\mathbb{E}[X] &= -1\frac{2}{3} + 1\frac{1}{3} = -\frac{1}{3} \\
\mathbb{E}[Y] &= 0 \\
\mathbb{E}[XY] &= (-1)(-1)\frac{1}{3} + \cdots + (1)(1)0 = 0
\end{aligned}$$

Luego, $\mathbb{E}[XY] = \mathbb{E}]X]\,\mathbb{E}[Y]$, pero por ejemplo,

$$f(-1,-1) = \frac{1}{3} \neq \frac{2}{9} = \frac{2}{3} \cdot \frac{1}{3} = f_X(-1) f_Y(-1)$$

por lo que las *v. a.*'s X e Y no son independientes.

Definición : Sea (X,Y) un *v. a.*. Se define el (r,s) **momento conjunto** de (X,Y) por $\mathbb{E}[X^r Y^s]$, siempre que le valor esperado exista.

Definición : Se define el (r,s) **momento central conjunto** de (X,Y) por $\mathbb{E}\left[(X-\mu_X)^r (Y-\mu_Y)^s\right]$.

Definición : Se define el r**-ésimo momento condicional de** X **dado** $Y=y$ por

$$\mathbb{E}[X^r \mid Y=y] = \begin{cases} \displaystyle\sum_{x_i \in \mathcal{S}_X} x_i^r\, \mathbb{P}\left(X = x_i \mid Y = y\right), & \text{caso discreto} \\ \displaystyle\int_{\mathbb{R}} x^r\, f(x \mid y) dx, & \text{caso continuo} \end{cases}$$

siempre que las sumas o integrales existan.

3.2. Varianzas y covarianzas

Proposición : Sean X y Y *v. a.*'s con μ_X y μ_Y sus respectivas medias. Entonces, se cumple

$$\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y) + 2\mathbb{E}\left[(X-\mu_X)(Y-\mu_Y)\right]$$

Demostración:

$$\begin{aligned}
\mathrm{var}(X+Y) &= \mathbb{E}\left[\left\{(X+Y)-\mathbb{E}[X+Y]\right\}^2\right] \\
&= \mathbb{E}\left[\left\{(X-\mu_X)+(Y-\mu_Y)\right\}^2\right] \\
&= \mathbb{E}(X-\mu_X)^2] + \mathbb{E}[(Y-\mu_Y)^2] + 2\mathbb{E}\left[(X-\mu_X)(Y-\mu_Y)\right] \\
&= \mathrm{var}(X) + \mathrm{var}(Y) + 2\mathbb{E}\left[(X-\mu_X)(Y-\mu_Y)\right]
\end{aligned}$$

Definición : Sean X y Y *v. a.*'s. Se define la **covarianza de** X **y** Y por el primer momento central conjunto del *v. a.* (X, Y). Esto es,

$$\mathrm{cov}(X,Y) = \mathbb{E}\left[(X-\mu_X)(Y-\mu_Y)\right] = \sigma_{XY}$$

Corolario : Sea (X, Y) un *v. a.*. Entonces,

$$\begin{aligned}
\mathrm{var}(X+Y) &= \mathrm{var}(X) + \mathrm{var}(Y) + 2\mathrm{cov}(X,Y) \\
\sigma^2_{X+Y} &= \sigma^2_X + \sigma^2_Y + 2\sigma_{XY}
\end{aligned}$$

Nota: La covarianza es una medida de la asociación lineal entre X y Y con unidades resultado del producto de las unidades de X y Y. Por ejemplo, considere el consumo mensual por casa habitación de agua y electricidad. El consumo medio de agua es $\mu_A = 3$ m^3 con una desviación estándar de $\sigma_A = 0.5$ m^3, mientras que el consumo de electricidad es de $\mu_E = 100$ kWh con una desviación estándar $\sigma_E = 18$ kWh. En general se tiene que a mayor consumo de agua mayor el consumo de electricidad y viceversa, luego la covarianza de éstas variables se espera positiva, digamos, $\sigma_{AE} = 5.4$ m$^3 \times$ kWh.

Proposición : Sean X y Y *v. a.*'s. Entonces,

$$\mathrm{cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$$

Demostración:

$$\begin{aligned}
\mathrm{cov}(X,Y) &= \mathbb{E}\left[(X-\mu_X)(Y-\mu_Y)\right] \\
&= \mathbb{E}\left[XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y\right] \\
&= \mathbb{E}\left[XY\right] - \mathbb{E}\left[X\mu_Y\right] - \mathbb{E}\left[Y\mu_X\right] + \mathbb{E}\left[\mu_X\mu_Y\right] \\
&= \mathbb{E}\left[XY\right] - \mu_X\mu_Y - \mu_X\mu_y + \mu_X\mu_Y \\
&= \mathbb{E}[XY] - \mu_X\,\mu_Y
\end{aligned}$$

Corolario : Sean X y Y *v. a.*'s independientes. Entonces $\mathrm{cov}(X,Y) = 0$.

Proposición : Si $\mathrm{cov}(X,Y) = 0$, no necesariamente las *v. a.*'s X y Y son independientes.

Demostración: Considere nuevamente la *f. m. p.* conjunta dada por la tabla

$X\backslash Y$	-1	0	$+1$	f_X
-1	$1/3$	0	$1/3$	$2/3$
$+1$	0	$1/3$	0	$1/3$
f_Y	$1/3$	$1/3$	$1/3$	1

Así, se tiene que $\mathbb{E}[XY] = 0$, $\mathbb{E}[X] = -1/3$ y $\mathbb{E}[Y] = 0$. Luego, $\text{cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0$, pero se vio ya que X e Y no son *v. a.*'s independientes.

Ejemplo : En el caso de los *dados cargados* se tiene que

$$\text{cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \frac{38}{3} - \left(\frac{7}{2}\right)^2 \approx 0.4167$$

mientras que en el caso de los *dados honestos*, $\text{cov}(X,Y) = 0$, por la independencia de X y Y.

Ejemplo : Considere $(\Omega, \mathcal{F}, \mathbb{P})$ un *EP*, $A, B \in \mathcal{F}$ y sean $X = \mathbb{1}_A$ y $Y = \mathbb{1}_B$. Entonces, las *v. a.*'s X y Y siguen una distribución Bernoulli con parámetros de éxito $\mathbb{P}(A)$ y $\mathbb{P}(B)$ respectivamente. Note además que $\mathbb{E}[XY] = \mathbb{E}[\mathbb{1}_A \mathbb{1}_B] = \mathbb{P}(A \cap B)$. Luego, se tiene

$$\begin{aligned}
\text{cov}(X,Y) &= \text{cov}(\mathbb{1}_A, \mathbb{1}_B) \\
&= \mathbb{E}\left[\mathbb{1}_A \mathbb{1}_B\right] - \mathbb{E}[\mathbb{1}_A]\mathbb{E}[\mathbb{1}_B] \\
&= \mathbb{P}(A \cap B) - \mathbb{P}(A)\,\mathbb{P}(B) \\
&= \mathbb{P}\big(A \mid B\big)\,\mathbb{P}(B) - \mathbb{P}(A)\,\mathbb{P}(B) \\
&= \Big[\mathbb{P}\big(A \mid B\big) - \mathbb{P}(A)\Big]\,\mathbb{P}(B)
\end{aligned}$$

I). Si $\mathbb{P}\big(A \mid B\big) > \mathbb{P}(A)$ entonces $\text{cov}(\mathbb{1}_A, \mathbb{1}_B) > 0$. Es decir, la ocurrencia del evento B aumenta la probabilidad de ocurrencia del evento A. Luego, su asociación (covarianza) es positiva.

II). Por el contrario, si la ocurrencia del evento B disminuye la probabilidad de ocurrencia del evento A, la asociación (covarianza) entre ellas es negativa.

$$\mathbb{P}\big(A \mid B\big) < \mathbb{P}(A) \Rightarrow \text{cov}(\mathbb{1}_A, \mathbb{1}_B) < 0$$

Propiedades : Sean X, Y, Z *v. a.*'s con varianza finita, $a, b \in \mathbb{R}$. Entonces se satisface

a). $\text{cov}(a, X) = 0$.

b). $\text{cov}(X, X) = \text{var}(X)$.

c). $\text{cov}(X, Y) = \text{cov}(Y, X)$.

d). $\text{cov}(aX, bY) = ab\text{cov}(X, Y)$.

e). $\text{cov}(X + Y, Z) = \text{cov}(X, Z) + \text{cov}(Y, Z)$.

f). $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y)$.

g). Si X y Y independientes, $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$.

Demostración:

a). $\text{cov}(a, X) = \mathbb{E}[aX] - a\mathbb{E}[X] = 0$.

b). $\text{cov}(X, X) = \mathbb{E}\left[(X - \mu_X)(X - \mu_X)\right] = \mathbb{E}\left[(X - \mu_X)^2\right] = \text{var}(X)$.

c). $\text{cov}(X, Y) = \mathbb{E}\left[(X - \mu_X)(Y - \mu_Y)\right] = \mathbb{E}\left[(Y - \mu_Y)(X - \mu_X)\right] = \text{cov}(Y, X)$.

d).

$$\begin{aligned}
\text{cov}(aX, bY) &= \mathbb{E}\left[(aX - a\mu_X)(bY - b\mu_Y)\right] \\
&= \mathbb{E}\left[a(X - \mu_X)b(Y - \mu_Y)\right] = ab\mathbb{E}\left[(X - \mu_X)(Y - \mu_Y)\right] \\
&= ab\,\text{cov}(X, Y)
\end{aligned}$$

e).

$$\begin{aligned}
\text{cov}(X + Y, Z) &= \mathbb{E}\left[\left((X + Y) - (\mu_X + \mu_Y)\right)(Z - \mu_Z)\right] \\
&= \mathbb{E}\left[\left((X - \mu_X) + (Y - \mu_Y)\right)(Z - \mu_Z)\right] \\
&= \mathbb{E}\left[(X - \mu_X)(Z - \mu_Z) + (Y - \mu_Y)(Z - \mu_Z)\right] \\
&= \mathbb{E}\left[(X - \mu_X)(Z - \mu_Z)\right] + \mathbb{E}\left[(Y - \mu_Y)(Z - \mu_Z)\right] \\
&= \text{cov}(X, Z) + \text{cov}(Y, Z)
\end{aligned}$$

f). Quedó demostrado en proposición anterior.

g). Se sigue pues X y Y *v. a.*'s independientes, luego $\text{cov}(X, Y) = 0$ y del inciso anterior.

Propiedades : Sean $X_1, \ldots, X_n$, $Y_1, \ldots, Y_m$, *v. a.*'s y $a_1, \ldots, a_n, b_1, \ldots, b_m \in \mathbb{R}$. Entonces,

a). $\text{var}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 \text{var}(X_i) + 2\sum_{i>j} a_i a_j \text{cov}(X_i, X_j)$.

b). Si $X_1, \ldots, X_n$ son *v. a.*'s independientes, $\text{var}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 \text{var}(X_i)$

c). $\text{cov}\left(\sum_{i=1}^n a_i X_i, \sum_{j=1}^m b_j Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \text{cov}(X_i, Y_j)$.

Demostración: a)–c) Se siguen de la proposición anterior. Escriba su demostración.

Proposición :

$$\text{cov}\left(X, Y - \mathbb{E}[Y|X]\right) = 0$$

Demostración: $\mathbb{E}[Y - \mathbb{E}[Y|X]] = \mathbb{E}[Y] - \mathbb{E}[\mathbb{E}[Y|X]] = 0$. Luego,

$$\text{cov}\left(X, Y - \mathbb{E}[Y|X]\right) = \mathbb{E}[X(Y - \mathbb{E}[Y|X])] = \mathbb{E}[XY] - \mathbb{E}[X\mathbb{E}[Y|X]] = 0$$

Definición : Sean X y Y con varianza finita. Se define el **coeficiente de correlación lineal** de X y Y por

$$\text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}} = \rho_{XY}$$

Nota: El coeficiente de correlación lineal es al igual que la covarianza, una medida de la asociación lineal entre X y Y, pero adimensional, es decir, sin unidades.

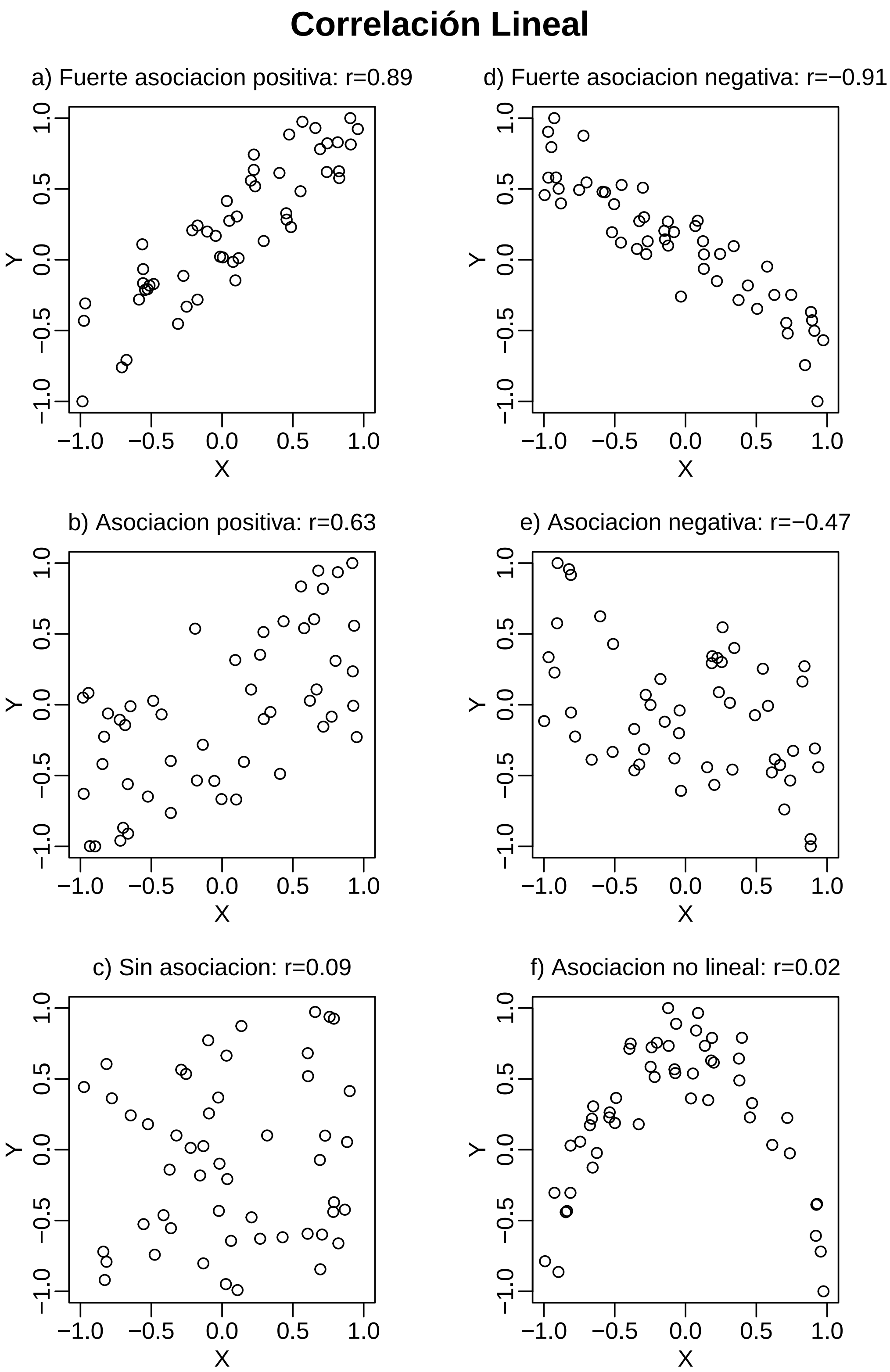


Figura 2: *Realizaciones* del vector aleatorio (X, Y) con distintos niveles de asociación. Por ejemplo, los puntos del panel a) muestran una asociación positiva con un coeficiente de correlación de $r = 0.89$.

Con referencia al ejemplo presentado después de la definición de covarianza, el consumo mensual de agua y electricidad tienen una correlación (lineal) de $\rho_{AE} = 0.6$.

La figura 2 muestra seis páneles con distintos grados de asociación entre las variables X y Y. Por ejemplo, en el panel d) es clara la asociación negativa de las variables, reflejada con un coeficiente de correlación de $r = -0.91$. El panel c) no muestra asociación las variables, con una correlación de $r = -0.11$. Finalmente, note que el panel f) muestra una asociación no lineal entre variables con el correspondiente coeficiente de correlación de $r = -0.15$.

Propiedades : Sea X y Y *v. a.*'s. Entonces,

a). $\operatorname{corr}(X, Y) = \operatorname{corr}(Y, X)$.

b). $-1 \leq \operatorname{corr}(X, Y) \leq +1$.

c). $\operatorname{corr}(X, Y) = \pm 1$, si y solo si $\mathbb{P}(X = aY) = 1$ para algún $a \in \mathbb{R}$.

Demostración:

a). Se sigue de la conmutatividad de la covarianza.

b). Se sigue de la desigualdad de Cauchy–Schwarz que se presenta a continuación.

c). Se sigue de la desigualdad de Cauchy–Schwarz.

Teorema : Desigualdad de Cauchy-Schwarz . Sean X y Y *v. a.*'s con segundo momento finito. Entonces,

$$\left(\mathbb{E}[XY]\right)^2 \leq \mathbb{E}[X^2]\,\mathbb{E}[Y^2]$$

cumpliéndose la igualdad, si y solo si, $\mathbb{P}(Y = 0) = 1$ ó $\mathbb{P}(X = aY) = 1$, para algún $a \in \mathbb{R}$.

Demostración: Se sigue la presentada en Hoel, Port, and Stone (1971)

I). Si $\mathbb{P}(Y = 0) = 1$, entonces $\mathbb{P}(XY = 0) = 1$ y $\mathbb{E}[Y^2] = \mathbb{E}[XY] = 0$ y la igualdad se cumple.

II). Si para algún $a \in \mathbb{R}$, $\mathbb{P}(X = aY) = 1$, entonces $\mathbb{E}[XY] = \mathbb{E}[aY \cdot Y] = a\mathbb{E}[Y^2]$, y

$$\begin{aligned}
\mathbb{E}^2[XY] &= \mathbb{E}^2[aY \cdot Y] \\
&= a^2\mathbb{E}^2[Y^2] \\
&= \left(a^2\mathbb{E}[Y^2]\right)\mathbb{E}[Y^2] \\
&= \mathbb{E}[a^2Y^2]\mathbb{E}[Y^2] \\
&= \mathbb{E}[X^2]\mathbb{E}[Y^2]
\end{aligned}$$

y la igualdad se cumple.

III). Suponga ahora que $\mathbb{P}(Y = 0) < 1$. Luego, $\mathbb{E}[Y^2] > 0$. Ahora, para todo $b \in \mathbb{R}$, se tiene que

$$0 \leq \mathbb{E}[(X - bY)^2] = b^2\mathbb{E}[Y^2] - 2b\mathbb{E}[XY] + \mathbb{E}[X^2]$$

La expresión anterior es un polinomio de grado 2 con coeficiente líder positivo, por lo que tiene un mínimo en $b^* = \dfrac{\mathbb{E}[XY]}{\mathbb{E}[Y^2]}$ y alcanza el valor

$$0 \leq \frac{\mathbb{E}^2[XY]}{\mathbb{E}^2[Y^2]}\mathbb{E}[Y^2] - 2\frac{\mathbb{E}[XY]}{\mathbb{E}[Y^2]}\mathbb{E}[XY] + \mathbb{E}[X^2]$$

Al multiplicar toda la expresión por $\mathbb{E}[Y^2]$ se tiene la desigualdad

$$\mathbb{E}^2[XY] \leq \mathbb{E}[X^2]\mathbb{E}[Y^2]$$

Corolario : X y Y *v. a.*'s con varianza finita, entonces

$$|\operatorname{corr}(X,Y)| \leq 1$$

Demostración: Aplique la desigualdad de Cauchy–Schwarz a las va's $(X-\mu_X)$ y $(Y-\mu_Y)$ y tome raíz cuadrada para mostrar que $|\operatorname{corr}(X,Y)| \leq 1$, o bien,

$$-1 \leq \operatorname{corr}(X,Y) \leq 1$$

3.3. Media y varianza muestrales.

Definición : Sean $X_1, X_2, \ldots,$ *v.a.i.i.d.* con $\mathbb{E}[X_i] = \mu$, $\operatorname{var}(X_i) = \sigma^2$. Se define la **suma, media muestral y varianza muestral**

$$S_n = \sum_{i=1}^{n} X_i, \qquad \bar{X}_n = \frac{1}{n} S_n, \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

respectivamente. Se tiene entonces la siguiente proposición.

Proposición :

a). $\mathbb{E}[\bar{X}_n] = \mu$.

b). $\operatorname{var}(\bar{X}_n) = \sigma^2/n$.

c). $\mathbb{E}[S^2] = \sigma^2$.

Demostración: Note que $\mathbb{E}[S_n] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i] = n\mu$. También,

$$\operatorname{var}(S_n) = \operatorname{var}\big(\sum_{i=1}^{n} X_i\big) \overset{\text{ind}}{=} \sum_{i=1}^{n} \operatorname{var}(X_i) = n\sigma^2$$

a). $\mathbb{E}[\bar{X}] = \mathbb{E}\left[\frac{1}{n}S_n\right] = \frac{1}{n}n\mu = \mu$.

b). $\operatorname{var}(\bar{X}) = \operatorname{var}\left(\frac{1}{n}S_n\right) = \frac{1}{n^2}\operatorname{var}(S_n) \overset{\perp\!\!\!\perp}{=} \frac{1}{n^2}n\sigma^2 = \frac{1}{n}\sigma^2$.

c). Note que $(n-1)S^2 = \sum_{i=1}^n (X_i - \bar{X})^2$. Luego,

$$\begin{aligned}
\mathbb{E}\left[(n-1)S^2\right] &= \mathbb{E}\left[\sum_{i=1}^n (X_i - \bar{X})^2\right] \\
&= \mathbb{E}\left[\sum_{i=1}^n \left\{(X_i - \mu) + (\mu - \bar{X})\right\}^2\right] \\
&= \mathbb{E}\left[\sum_{i=1}^n (X_i - \mu)^2 + \sum_{i=1}^n (\mu - \bar{X})^2 - 2(\bar{X} - \mu)\sum_{i=1}^n (X_i - \mu)\right] \\
&= \mathbb{E}\left[\sum_{i=1}^n (X_i - \mu)^2 + n(\bar{X} - \mu)^2 - 2(\bar{X} - \mu)\cdot n(\bar{X} - \mu)\right] \\
&= \sum_{i=1}^n \mathbb{E}\left[(X_i - \mu)^2\right] - n\mathbb{E}\left[(\bar{X} - \mu)^2\right] \\
&= n\sigma^2 - n\sigma^2/n \\
&= (n-1)\sigma^2
\end{aligned}$$

Por lo tanto, $\mathbb{E}[S^2] = \sigma^2$.

3.4. Vector de medias y matriz de covarianzas.

Definición : Sea $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ un vector aleatorio de dimensión n. Se define su **vector de medias o valor esperado** por $\boldsymbol{\mu} = \mathbb{E}[\boldsymbol{X}]$, donde

$$\mathbb{E}[\boldsymbol{X}] = \begin{bmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_n] \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} = \boldsymbol{\mu}$$

Definición : Sea $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ un vector aleatorio de dimensión n. Se define su **matriz de covarianzas** por $\Sigma = \mathrm{cov}(\boldsymbol{X})$, donde

$$\begin{aligned}
\Sigma &= \mathbb{E}\left[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^T\right] \\
&= \Big(\mathbb{E}\left[(X_i - \mu_i)(X_j - \mu_j)\right]\Big) \\
&= (\sigma_{ij}) \\
\sigma_{ij} &= \mathbb{E}\left[(X_i - \mu_i)(X_j - \mu_j)\right] \\
&= \mathrm{cov}(X_i, X_j)
\end{aligned}$$

O bien,

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_n^2 \end{pmatrix}$$

Nota: El operador esperanza actúa sobre la matriz tomado el valor esperado de cada una de las entradas de la matriz.

Verifique que efectivamente

$$\Sigma = \mathbb{E}\left[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^T\right] = (\sigma_{ij})$$

Proposición : Sea $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ un *v. a.* y $\text{cov}(\boldsymbol{X}) = \Sigma$. Entonces, Σ es simétrica (semi) definida positiva.

Demostración:

I). La simetría de Σ se sigue de la conmutatividad de la covarianza, $\sigma_{ij} = \text{cov}(X_i, X_j) = \text{cov}(X_j, X_i) = \sigma_{ji}$.

II). Se sigue de proposición que se muestra más adelante.

Proposición : Sea $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ un *v. a.* con vector de medias $\boldsymbol{\mu}$ y matriz de covarianzas Σ. Entonces, $\Sigma = \text{cov}(\boldsymbol{X}) = \mathbb{E}\left[\boldsymbol{X}\boldsymbol{X}^T\right] - \boldsymbol{\mu}\boldsymbol{\mu}^T$.

Demostración:

$$\begin{aligned}
\text{cov}(\boldsymbol{X}) &= \mathbb{E}\left[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^T\right] \\
&= \mathbb{E}\left[\boldsymbol{X}\boldsymbol{X}^T - \boldsymbol{\mu}\boldsymbol{X}^T - \boldsymbol{X}\boldsymbol{\mu}^T + \boldsymbol{\mu}\boldsymbol{\mu}^T\right] \\
&= \mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T] - \boldsymbol{\mu}\mathbb{E}[\boldsymbol{X}^T] - \mathbb{E}[\boldsymbol{X}]\boldsymbol{\mu}^T + \boldsymbol{\mu}\boldsymbol{\mu}^T \\
&= \mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T] - \boldsymbol{\mu}\boldsymbol{\mu}^T
\end{aligned}$$

Proposición : Sea $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ un *v. a.* con vector de medias $\boldsymbol{\mu}$ y matriz de covarianzas Σ. Sea $A_{m\times n}$ matriz de constantes. Entonces,

a). $\mathbb{E}[A\boldsymbol{X}] = A\mathbb{E}[\boldsymbol{X}]$.

b). $\text{cov}(A\boldsymbol{X}) = A\Sigma A^T$.

Demostración:

a). Sea $\boldsymbol{Y} = A\boldsymbol{X}$. Para la i-ésima entrada del vector se tiene

$$\begin{aligned}
\big(\mathbb{E}[\boldsymbol{Y}]\big)_i &= \mathbb{E}[Y_i] \\
&= \mathbb{E}\left[\sum_{\ell=1}^{n} a_{i\ell}X_\ell\right] \\
&= \sum_{\ell=1}^{n} a_{i\ell}\mathbb{E}[X_\ell] \\
&= \sum_{\ell=1}^{n} a_{i\ell}\mu_\ell \\
&= (A\boldsymbol{\mu})_i
\end{aligned}$$

b). Sea $B = A\Sigma$, luego $A\Sigma A^T = BA^T$. Entonces, la entrada ij de la matriz,

$$\begin{aligned}
\left(A\Sigma A^T\right)_{ij} &= \left(BA^T\right)_{ij} \\
&= \sum_{u=1}^{n} b_{iu} a_{uj}^T \\
&= \sum_{u=1}^{n} b_{iu} a_{ju} \\
&= \sum_{u=1}^{n} \left(\sum_{v=1}^{n} a_{iv}\sigma_{vu} \right) a_{ju} \\
&= \sum_{u}\sum_{v} a_{iv}a_{ju}\operatorname{cov}(X_v, X_u) \\
&= \sum_{v} a_{iv}\operatorname{cov}(X_v, \sum_{u} a_{ju}X_u) \\
&= \operatorname{cov}\left(\sum_{v} a_{iv}X_v, \sum_{u} a_{ju}X_u \right) \\
&= \operatorname{cov}\left((A\boldsymbol{X})_i, (A\boldsymbol{X})_j \right) \\
&= \left(\operatorname{cov}(A\boldsymbol{X}) \right)_{ij}
\end{aligned}$$

Por lo tanto,

$$\operatorname{cov}(A\boldsymbol{X}) = A\operatorname{cov}(\boldsymbol{X})A^T = A\Sigma A^T$$

Definición : Sean $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ y $\boldsymbol{Y} = (Y_1, \ldots, Y_m)^T$ dos v. a.'s. Se define la **matriz de covarianzas de $\boldsymbol{X}$ y $\boldsymbol{Y}$** por

$$\operatorname{cov}(\boldsymbol{X}, \boldsymbol{Y}) = \left(\operatorname{cov}(X_i, Y_j) \right)_{n\times m}$$

Proposición : Sean $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ y $\boldsymbol{Y} = (Y_1, \ldots, Y_m)^T$ dos v. a.'s, $A_{u\times n}$ y $B_{v\times m}$ dos matrices de constantes. Entonces,

$$\operatorname{cov}(A\boldsymbol{X}, B\boldsymbol{Y}) = A\operatorname{cov}(\boldsymbol{X}, \boldsymbol{Y})B^T$$

Demostración: Se sigue directamente empleando álgebra matricial.

Proposición : Sea $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ un v. a. con matriz de covarianzas Σ. Entonces, Σ es una matriz (semi) definida positiva.

Demostración: Sea $\boldsymbol{a} = (a_1, \ldots, a_n)^T \in \mathbb{R}^n$, $\boldsymbol{a} \neq 0$ y $Y = \boldsymbol{a}^T\boldsymbol{X}$. Entonces,

$$0 \leq \operatorname{var}(Y) = \operatorname{var}(\boldsymbol{a}^T\boldsymbol{X}) = \boldsymbol{a}^T\operatorname{var}(\boldsymbol{X})(\boldsymbol{a}^T)^T = \boldsymbol{a}^T\Sigma\boldsymbol{a}$$

La desigualdad anterior se cumple para todo $\boldsymbol{a}$, por lo que Σ es (semi) definida positiva.

Nota: Recuerde las siguientes propiedades del operador traza de matrices:

a). Sea $A_{n\times n}$, se define la **traza de la matriz** A por

$$\operatorname{tr}(A) = \sum_{i=1}^{n} a_{ii}$$

b). Si $a \in \mathbb{R}$, $\operatorname{tr}(a) = a$.

c). Sean A y B matrices tales que los productos AB y BA están bien definidos. Entonces,

$$\operatorname{tr}(AB) = \operatorname{tr}(BA)$$

Definición : Sea $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ un *v. a.* con matriz de covarianzas Σ. Algunos textos definen la **variación total** de $\boldsymbol{X}$ por

$$\operatorname{vartot}(\boldsymbol{X}) = \operatorname{tr}(\Sigma) = \sum_{i=1}^{n} \sigma_i^2 = \sum_{i=1}^{n} \operatorname{var}(X_i)$$

Teorema Espectral: *Problema de diagonalización.* Sea $A_{n\times n}$ matriz simétrica (semi) definida positiva. Entonces existen matrices $Q_{n\times n}$ ortonormal y $\Lambda_{n\times n}$ diagonal, $\Lambda = \operatorname{diag}\{\lambda_1, \ldots, \lambda_n\}$, con $\lambda_1 \geq \cdots \geq \lambda_n (\geq) > 0$, los valores propios de A, tal que $A = Q\Lambda Q^T$ y en tal caso, $\Lambda = Q^T A Q$.

Proposición : Sea $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ un *v. a.* con matriz de covarianzas Σ. Entonces, existen matrices Q ortonormal y $\Lambda = \operatorname{diag}\{\lambda_1, \ldots, \lambda_n\}$, con $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$, los valores propios de Σ tales que $\Sigma = Q\Lambda Q^T$.

Demostración: Se sigue del hecho que toda matriz de covarianzas es (semi) definida positiva y del teorema espectral.

Proposición : Sea $\boldsymbol{X} = (X_, \ldots, X_n)^T$ un *v. a.* con matriz de covarianzas Σ y $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$ sus valores propios. Entonces,

$$\operatorname{vartot}(\boldsymbol{X}) = \sum_{i=1}^{n} \sigma^2 = \sum_{I=1}^{n} \lambda_i$$

Demostración:

$$\operatorname{vartot}(\boldsymbol{X}) = \operatorname{tr}(\Sigma) = \operatorname{tr}(Q\Lambda Q^T) = \operatorname{tr}(\Lambda Q^T Q) = \operatorname{tr}(\Lambda) = \sum_{i=1}^{n} \lambda_i$$

Proposición : Sea $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ un *v. a.* con matriz de covarianzas Σ con descomposición espectral $\Sigma = Q\Lambda Q^T$. Entonces $\boldsymbol{Y} = Q^T \boldsymbol{X}$ es un vector aleatorio con componentes no correlacionadas y la misma variación total que $\boldsymbol{X}$.

Demostración: Sea $\Sigma = Q\Lambda Q^T$ la descomposición espectral de la matriz de covarianzas de $\boldsymbol{X}$ y $\boldsymbol{Y} = Q^T \boldsymbol{Y}$. Luego,

$$\operatorname{var}(\boldsymbol{Y}) = \operatorname{var}(Q^T \boldsymbol{X}) = Q^T \operatorname{var}(\boldsymbol{X}) Q = Q^T \Sigma Q = Q^T (Q\Lambda Q^T) Q = \operatorname{diag}\{\lambda_1, \ldots, \lambda_n\}$$

Entonces,

I). $\operatorname{var}(Y_i) = \lambda_i$, $i = 1, \ldots, n$.

II). $\operatorname{cov}(Y_i, Y_j) = 0$, para $i \neq j$.

III). $\operatorname{vartot}(\boldsymbol{Y}) = \sum_{i=1}^{n} \lambda_i = \operatorname{vartot}(X)$, como se vio en proposición anterior.

Definición : Sea $\boldsymbol{X} = (X_1, \ldots, X_n)$ un *v. a.*. Se define la **matriz de correlación de** X por

$$R = \operatorname{corr}(\boldsymbol{X}) = \big(\operatorname{corr}(X_i, X_j)\big) = (\rho_{ij})$$

donde $\rho_{ij} = \dfrac{\operatorname{cov}(X_i, X_j)}{\sqrt{\operatorname{var}(X_i)\operatorname{var}(X_j)}}$.

Nota: Sean $V = \operatorname{diag}\{\sigma_1^2, \ldots, \sigma_n^2\}$, $V^{1/2} = \operatorname{diag}\{\sigma_1, \ldots, \sigma_n\}$, $V^{-1/2} = \operatorname{diag}\{1/\sigma_1, \ldots, 1/\sigma_n\}$. Entonces,

$$R = V^{-1/2}\Sigma V^{-1/2} = V^{-1/2}\left(\Sigma V^{-1/2}\right) = V^{-1/2}\left(\frac{\sigma_{ij}}{\sigma_j}\right) = \left(\frac{\sigma_{ij}}{\sigma_i\sigma_j}\right) = (\rho_{ij})$$

Note que la diagonal principal de la matriz R son $\rho_{ii} = 1$.

Ejemplo : Considere el vector aleatorio $\boldsymbol{X} = (X_1, \ldots, X_5)^T$ con vector de medias $\boldsymbol{\mu}_X$ y matriz de covarianzas Σ_X, dados por

$$\boldsymbol{\mu}_X = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 2 \\ 0 \end{bmatrix} \quad \text{y} \quad \Sigma_X = \begin{bmatrix} 1 & 0 & -1 & 1 & 0 \\ 0 & 2 & -1 & 0 & 1 \\ -1 & -1 & 3 & -1 & 0 \\ 1 & 0 & -1 & 4 & 0 \\ 0 & 1 & 0 & 0 & 5 \end{bmatrix}$$

a). Considere las siguientes definiciones: $Y_1 = X_1 + \cdots + X_5$; $Y_2 = X_5 - X_1$; $Y_3 = (X_2 + X_3 + X_4)/3$; $Y_4 = X_1 + X_2 - X_4 - X_5$; $Y_5 = -3X_3$. Calcule el vector de medias y matriz de covarianzas del *v. a.* $\boldsymbol{Y}$.

b). Encuentre la descomposición espectral de Σ y úsela para determinar una transformación lineal $\boldsymbol{W} = B\boldsymbol{X}$ con sus componentes no correlacionadas.

c). Verifique que $\operatorname{vartot}(\boldsymbol{W}) = \operatorname{vartot}(\boldsymbol{X})$.

d). Para cada una de las entradas del vector $\boldsymbol{W}$ calcule su participación proporcional en la variación total, esto es $\delta_i = \operatorname{var}(W_i)/\operatorname{vartot}(W)$. Determine la proporción acumulada $\Delta_j = \delta_1 + \cdots + \delta_j$ hasta la componente j-ésima para $1 \leq j \leq 5$.

Solución:

a). Para definir el vector $\boldsymbol{Y}$ construya la matriz A por

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ -1 & 0 & 0 & 0 & 1 \\ 0 & 1/3 & 1/3 & 1/3 & 0 \\ 1 & 1 & 0 & -1 & -1 \\ 0 & 0 & -3 & 0 & 0 \end{bmatrix}$$

y se tiene que $\boldsymbol{Y} = A\boldsymbol{X}$. Así $\boldsymbol{\mu}_Y = A\boldsymbol{\mu}_X$ y $\Sigma_Y = A\Sigma_X A^T$ con

$$\boldsymbol{\mu}_Y = \begin{bmatrix} 2.000 \\ 0.000 \\ 0.666 \\ -1.000 \\ 3.000 \end{bmatrix} \quad \text{y} \quad \Sigma_Y = \begin{bmatrix} 13.000 & 5.000 & 1.998 & -7.000 & 30.000 \\ 5.000 & 6.000 & 0.333 & -4.000 & 22.000 \\ 1.998 & 0.333 & 0.554 & -0.999 & 0.666 \\ -7.000 & -4.000 & -0.999 & 8.000 & -17.000 \\ 30.000 & 22.000 & 0.666 & -17.000 & 152.000 \end{bmatrix}$$

b). Sea $\Sigma_X = Q\Lambda Q^T$, la descomposición espectral de Σ_X. Se tiene

$$Q = \begin{bmatrix} 0.161 & -0.271 & 0.027 & 0.390 & 0.865 \\ 0.331 & 0.035 & 0.470 & -0.768 & 0.281 \\ -0.352 & 0.436 & -0.628 & -0.374 & 0.390 \\ 0.360 & -0.664 & -0.579 & -0.277 & -0.132 \\ 0.782 & 0.543 & -0.220 & 0.205 & -0.061 \end{bmatrix}$$

y $\Lambda = \text{diag}\{5.423, 5.064, 2.869, 1.247, 0.397\}$. Luego, el vector $\boldsymbol{W} = Q^T\boldsymbol{X}$

$$\begin{aligned} W_1 &= 0.161X_1 + 0.331X_2 - 0.352X_3 + 0.36X_4 + 0.782X_5 \\ W_2 &= -0.271X_1 + 0.035X_2 + 0.436X_3 - 0.664X_4 + 0.543X_5 \\ W_3 &= 0.027X_1 + 0.47X_2 - 0.628X_3 - 0.579X_4 - 0.22X_5 \\ W_4 &= 0.39X_1 - 0.768X_2 - 0.374X_3 - 0.277X_4 + 0.205X_5 \\ W_5 &= 0.865X_1 + 0.281X_2 + 0.39X_3 - 0.132X_4 - 0.061X_5 \end{aligned}$$

es tal que $\text{cov}(\boldsymbol{W}) = \text{diag}\{5.423, 5.064, 2.869, 1.247, 0.397\}$. Esto es, los elementos fuera de la diagonal son cero por lo que $\boldsymbol{W}$ es de componentes no correlacionadas.

c).

$$\text{vartot}(\boldsymbol{W}) = 5.423 + 5.064 + 2.869 + 1.247 + 0.397 = 15 = 1 + 2 + 3 + 4 + 5 = \text{vartot}(\boldsymbol{X})$$

d). Finalmente, la siguiente tabla muestra las varianzas λ_j y proporciones p_j acumuladas de las componentes del vector aleatorio $\boldsymbol{W}$. Note que las primeras 3 componentes acumulan el 89 %

	W_1	W_2	W_3	W_4	W_5
λ_j	5.423	5.064	2.869	1.247	0.397
$\sum \lambda_j$	5.423	10.488	13.356	14.603	15.000
$\sum p_j$	0.362	0.699	0.890	0.974	1.000

de la variación total.

El panel izquierdo de la figura 3 presenta las varianzas individuales de cada una de las componentes. El panel de la derecha muestra como se va acumulando la variación total.

Los cálculos del ejemplo anterior se realizaron con el lenguaje R mediante el siguiente código:

```
# ----------------------------------------------------------------------
mu.x <- c(0, 1, -1, 2, 3)
Sigma.x <- matrix(c(1, 0, -1, 1, 0,  0, 2, -1, 0, 1,  -1, -1, 3, -1, 0,
 1, 0, -1, 4, 1,  0, 1, 0, 1, 5), nrow=5,  byrow=FALSE)
trSigma.x <- sum(diag(Sigma.x))
A <- matrix(c(1, -1, 0, 1, 0,  1, 0, 0.333, 1, 0,  1, 0, 0.333, 0, -3,
 1, 0, 0.333, -1, 0,  1, 1, 0, -1, 0), nrow=5,  byrow=FALSE)
mu.y <- A%*%mu.x
Sigma.y <- A%*%Sigma.x%*%t(A)
spec.decomp <- eigen(Sigma.x)
lambda <- spec.decomp$values
Q <-   spec.decomp$vectors
mu.w <- t(Q)%*%mu.x
Sigma.w <- t(Q)%*%Sigma.x%*%Q
trSigma.w <- sum(diag(Sigma.w))
# ----------------------------------------------------------------------
```

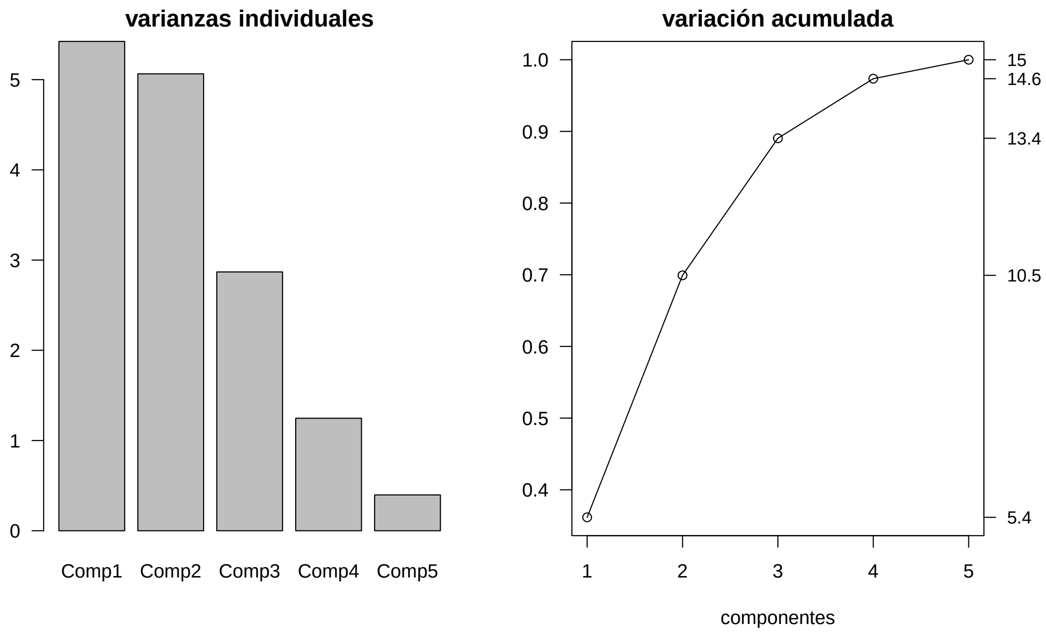


Figura 3: Varianzas individuales $\mathrm{var}(Y_i) = \lambda_i$ y las correspondientes sumas acumuladas.

3.5. Sumas aleatorias

Sea N una variable aleatoria entera no negativa con media μ_N y varianza σ_N^2. Sean $X_1, X_2, \ldots,$ *v. a.*'s independientes con media μ_X y varianza σ_X^2 común e independientes de N. Entonces, si $S_N = \sum_{i=1}^{N} X_i = X_1 + \cdots + X_N$, la suma de un número aleatorio de *v. a.*'s, se tiene

a). $\mathbb{E}[S_N] = \mu_N \mu_X$.

b). $\mathrm{var}(S_N) = \sigma_N^2 \mu_X^2 + \mu_N \sigma_X^2$.

Demostración: Sea $S_N = X_1 + \cdots + X_N$ suma de un número aleatorio de *v. a.*'s. Considerando el *valor esperado iterado*

a). $\mathbb{E}[S_N] = \mathbb{E}\left[\mathbb{E}[S_N \mid N]\right]$,
$\mathbb{E}\left[S_N \mid N = n\right] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = n\mathbb{E}[X_i] = n\mu_X$.
Por lo tanto, $\mathbb{E}\left[S_N\right] = \mathbb{E}\left[N\mu_X\right] = \mu_X \mathbb{E}[N] = \mu_N \mu_X$.

b). Considerando la varianza condicional,

$$\mathrm{var}(S_N) = \mathbb{E}\left[\mathrm{var}(S_N \mid N)\right] + \mathrm{var}(\mathbb{E}\left[S_N \mid N\right])$$

I) $\mathrm{var}(S_N \mid N = n) = \mathrm{var}\left(\sum_{i=1}^{n} X_i\right) \overset{\text{ind}}{=} \sum_{i=1}^{n} \mathrm{var}(X_i) = n\sigma_X^2$.

II) $\mathbb{E}\left[S_N \mid N = n\right] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = n\mu_X$,

$$\mathrm{var}\left(S_N\right) = \mathbb{E}\left[N\sigma_X^2\right] + \mathrm{var}\left(N\mu_X\right) = \mu_N \sigma_X^2 + \mu_X^2 \sigma_N^2$$

Ejemplo : El número de personas que entran a un elevador se distribuye aproximadamente como Poisson de media $\lambda = 2.3$. El peso W de una persona se aproxima mediante una distribución Gamma con $\alpha = 53$ y $\beta = 1.25$, parámetros de forma y escala respectivamente. Calcule el peso medio y la variación que opera el elevador por recorrido.

Solución: Sea N el número de personas por recorrido. Luego, $N \sim \text{Po}(\lambda)$, $\mu_N = 2.3$ y $\sigma_N^2 = 2.3$. W es el peso (kg) por persona con $W \sim \text{Gamma}(\alpha, \beta)$, Luego. $\mu_W = \alpha\beta = 66.25$ kg, $\sigma_W^2 = 82.81$ y $\sigma_W = 9.1$ kg.

$$\mathbb{E}[S_N] = \mu_N \mu_W = 2.3(66.25) = 152.38 \text{ kg}$$

$$\text{var}(S_N) = \sigma_N^2 \mu_W^2 + \mu_N \sigma_W^2 = 2.3 \cdot 66.25^2 + 2.3 \cdot 82.81 = 10,285.31$$

y $\sigma_{S_N} = 101.42$ kg.

3.6. Mezclas de distribuciones

3.6.1. Mezcla de distribuciones normales

Normales contaminadas

Suponga que se observan eventos que se distribuyen normal estándar y que en ocasiones éstos se distribuyen normalmente pero con una mayor varianza $\sigma^2 (> 1)$. Sea $Z \sim \text{N}(0, 1)$ y la variable aleatoria I_p que toma valores de 1 y 0 con probabilidades p y $1-p$, respectivamente. Suponga que Z e I_p son variables aleatorias independientes y defina

$$W = ZI_p + \sigma Z(1 - I_p)$$

Entonces, si $w \in \mathbb{R}$, la función de probabilidad acumulada (*f. p. a.*) está dada por

$$\begin{aligned}
F_W(w) &= P(W \leq w) \\
&= P(W \leq w, I_p = 0) + P(W \leq w, I_p = 1) \\
&= P(W \leq w | I_p = 0) P(I_p = 0) + P(W \leq w | I_p = 1) P(I_p = 1) \\
&= (1 - p) P(W \leq w | I_p = 0) + p P(W \leq w | I_p = 1) \\
&= (1 - p) P(Z \leq w/\sigma) + p P(Z \leq w) \\
&= (1 - p) \Phi(w/\sigma) + p \Phi(w)
\end{aligned}$$

donde Φ es la función de probabilidad acumulada de la distribución normal estándar.

Diferenciando F_W se tiene la correspondiente función de densidad de probabilidad (*f. d. p.*) de W. A saber,

$$f_W(w) = p\phi(w) + \frac{1 - p}{\sigma} \phi\left(\frac{w}{\sigma}\right)$$

donde ϕ es la *f. d. p.* de Z.

La variable aleatoria

$$W = I_p Z + (1 - I_p)(\sigma Z)$$

es una **mezcla de normales**.

Se sigue que por independencia de Z e I_p que

$$\begin{aligned}
\mathbb{E}[W] &= \mathbb{E}\left[ZI_p + \sigma Z(1 - I_p)\right] \\
&= \mathbb{E}[Z]\mathbb{E}[I_p] + \sigma \mathbb{E}[Z]\mathbb{E}[1 - I_p] \\
&= 0
\end{aligned}$$

$$\text{var}(W) = p + \sigma^2(1 - p)$$

Suponga ahora que en realidad se desea la distribución de $X = a + bW$, con $b > 0$. Entonces,

$$\begin{aligned} F_X(x) &= P(X \leq x) = P(a + bW \leq x) = P(W \leq \frac{x-a}{b}) \\ &= F_W(\frac{x-a}{b}) \\ &= (1-p)\Phi(\frac{x-a}{b\sigma}) + p\,\Phi(\frac{x-a}{b}) \end{aligned}$$

que es una mezcla de *f. p. a.* normales. Se tiene además que

$$\begin{aligned} \mathbb{E}[X] &= a \\ \mathrm{var}(X) &= b^2\left(p + \sigma^2(1-p)\right) \end{aligned}$$

3.6.2. Mezcla de distribuciones

Suponga k distribuciones con *f. d. p.* f_i, medias μ_i, varianzas σ_i^2, soportes S_i y con **probabilidades de mezclas** $p_i(> 0)$, $i = 1, \ldots, k$, tal que $1 = p_1 + \cdots + p_k$. Sea $S = \bigcup_{i=1}^k S_i$ y para $x \in S$, considere la *v. a.* X que tiene como función de densidad la función

$$f_X(x) = \sum_{i=1}^{k} p_i f_i(x) \tag{1}$$

Note que f_X es un función de densidad *propia* ya que satisface

a). $f_X(x) \geq 0$, pues $p_i > 0$ y $f_i(x) \geq 0$.

b). $\int_{\mathbb{R}} f_X(x)dx = 1$, pues $\int_{\mathbb{R}} f_X(x)dx = \sum p_i[\int_{\mathbb{R}} f_i(x)dx] = \sum p_i = 1$.

por lo que la función dada por (1) es una función de densidad legítima. Se tiene además que

$$\begin{aligned} \mathbb{E}[X] &= \int_{\mathbb{R}} x f_X(x)dx \\ &= \int_{\mathbb{R}} x \left(\sum_{i=1}^{k} p_i f_i(x)\right) dx \\ &= \sum_{i=1}^{k} p_i \int_{\mathbb{R}} x f_i(x)dx \\ &= \sum_{i=1}^{k} p_i \mathbb{E}[X_i] \\ &= \sum_{i=1}^{k} p_i \mu_i \\ &= \bar{\mu} \end{aligned}$$

que es el promedio de medias ponderado por la probabilidades de mezcla. Similarmente,

$$\begin{aligned}
\mathrm{var}(X) &= \int_{\mathbb{R}} (x-\mu_X)^2 f_X(x)dx \\
&= \int_{\mathbb{R}} (x-\mu_X)^2 \left[\sum_{i=1}^{k} p_i f_i(x)\right] dx \\
&= \sum_{i=1}^{k} p_i \int_{\mathbb{R}} (x-\mu_X)^2 f_i(x)dx \\
&= \sum_{i=1}^{k} p_i \int_{\mathbb{R}} \left[(x-\mu_i)+(\mu_i-\mu_X)\right]^2 f_i(x)dx \\
&= \sum_{i=1}^{k} p_i \int_{\mathbb{R}} (x-\mu_i)^2 f_i(x)dx + \sum_{i=1}^{k} p_i(\mu_i-\mu_X)^2 \int_{\mathbb{R}} f_i(x)dx \\
&= \sum_{i=1}^{k} p_i\sigma_i^2 + \sum_{i=1}^{k} p_i(\mu_i-\bar{\mu})^2
\end{aligned}$$

que corresponde a la suma ponderada de las varianzas más la varianza de las medias.

Note que los resultados anteriores corresponden a *mezclas de distribuciones* y no combinación lineal de variables aleatorias.

Ejemplo (Lista 2, ejercicio 4): El número total de defectos X en un chip sigue una distribución de Poisson parámetro α. Suponga que cada defecto tiene una probabilidad p de caer en una región específica R y que la localización es independiente del número de defectos. Entonces, el número de defectos en R sigue una distribución Poisson con media αp.

En efecto, sea X el número de defectos en el chip y sea N el número de defectos que se localizan en la región R. Luego, dado $X = x$, se sigue que $N \sim \mathrm{Bin}(x, p)$. Por lo tanto, por el **teorema de probabilidad total**, para $n = 0, 1, 2, \ldots$

$$\begin{aligned}
\mathbb{P}(N=n) &= \sum_{x=0}^{\infty} \mathbb{P}(N=n|X=x)\mathbb{P}(X=x), \\
&\qquad \text{que es una mezcla (infinita) de distribuciones, binomial-Poisson} \\
&= \sum_{x=n}^{\infty} \binom{x}{n} p^n(1-p)^{x-n} \cdot \alpha^x e^{-\alpha}/x! \\
&= (\alpha p)^n \frac{e^{-(\alpha p)}}{n!}
\end{aligned}$$

Por lo tanto, el número de defectos en la región R se distribuye Poisson parámetro αp.

Definición : Sea X una variable aleatoria, se dice distribuida **loggamma** con parámetros $\alpha(>0)$ y $\beta(>0)$ si tiene una función de densidad dada por

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{-\frac{1+\beta}{\beta}} (\log x)^{\alpha-1}\ I_{(1,\infty)}(x) \tag{2}$$

y se denota por $X \sim \mathrm{loggamma}(\alpha, \beta)$.

Proposición : Sea $Y \sim \mathrm{Gamma}(\alpha, \beta)$, con $\mathbb{E}[Y] = \alpha\beta$ y $\mathrm{var}(Y) = \alpha\beta^2$, y sea $X = e^Y$. Entonces, $X \sim \mathrm{loggamma}(\alpha, \beta)$, con *f. d. p.* dada por la expresión (2).

Los actuarios han encontrado la mezcla de *gamma* con *loggamma* como un buen modelo para la distribución de reclamaciones.

Suponga $X_1 \sim \text{loggamma}(\alpha_2, \beta_2)$ y $X_2 \sim \text{Gamma}(\alpha_1, \beta_1)$, y la mezcla es p y $(1-p)$. Entonces la distribución de la mezcla X tiene la *f. d. p.* dada por

$$
\begin{aligned}
f(x) \quad = \quad & \left[\frac{1-p}{\beta_2^{\alpha_2}\Gamma(\alpha_2)} x^{\alpha_2 - 1} e^{-x/\beta_2}\right] I_{(0,1]}(x) \qquad (3)\\
& + \left[\frac{p}{\beta_1^{\alpha_1}\Gamma(\alpha_1)} (\log x)^{\alpha_1 - 1} x^{-(\beta_1+1)/\beta_1} + \frac{1-p}{\beta_2^{\alpha_2}\Gamma(\alpha_2)} x^{\alpha_2 - 1} e^{-x/\beta_2}\right] I_{(1,\infty)}(x)
\end{aligned}
$$

Si $\beta_1 < 1/2$, se puede mostrar que la media μ y la varianza σ^2 de la mezcla están dadas respectivamente por

$$
\begin{aligned}
\mathbb{E}[X] &= p(1-\beta_1)^{-\alpha_1} + (1-p)\alpha_2\beta_2 \\
\text{var}(X) &= p\left[(1-2\beta_1)^{-\alpha_1} - (1-\beta_1)^{-2\alpha_1}\right] + (1-p)\alpha_2\beta_2^2 \\
&\quad + p(1-p)\left[(1-\beta_1)^{-\alpha_1} - \alpha_2\beta_2\right]^2
\end{aligned}
$$

Notas:

a). La mezcla de distribuciones son llamadas también **composición de distribuciones**.

b). Las mezclas de distribuciones no tienen porque restringirse al caso finito. Por ejemplo, suponga que N_λ sigue una distribución Poisson de media λ, que a su vez sigue una distribución gamma. Esto es, $N_\lambda \sim \text{Po}(\lambda)$ y $\Lambda \sim \text{Gamma}(\alpha, \beta)$. Entonces, para $k = 0, 1, 2, \ldots$

$$
\begin{aligned}
P(N = k) &= \int_0^\infty P(N_\lambda = k | \Lambda = \lambda) f_\Lambda(\lambda) d\lambda \\
&= \int_0^\infty \frac{\lambda^k e^{-\lambda}}{k!} \frac{1}{\Gamma(\alpha)\beta^\alpha} \lambda^{\alpha-1} e^{-\lambda/\beta} d\lambda \\
&= \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)k!} \frac{\beta^k}{(1+\beta)^{\alpha+k}}
\end{aligned}
$$

que no es mas que la aplicación del *teorema de probabilidad total.*

c). Si en el inciso anterior $\alpha = r \in \mathbb{N}$ y $\beta = \frac{1-p}{p}$, $0 < p < 1$, entonces,

$$
\begin{aligned}
P(N = k) &= \frac{(r+k-1)!}{(r-1)!k!} p^r (1-p)^k \\
&= \binom{r+k-1}{k} p^r (1-p)^k
\end{aligned}
$$

Esto es, N se distribuye marginalmente como una *binomial negativa* con parámetros r y p.

d). La distribución binomial negativa ha sido empleada con éxito en la modelación del número de accidentes.

3.7. Ejercicios

Refiérase al Cuaderno de Ejercicios sección 3, Barrios and Chambon (2024).

4. Función generadora de momentos

4.1. Recordar ...

Definición : Sea X una variable aleatoria (*v. a.*). Se define su **función generadora de momentos** (*f. g. m.*) por

$$m_X(t) = \mathbb{E}[e^{tX}]$$

siempre que el valor esperado exista para $|t| < \delta$, para algún $\delta > 0$.

Proposición : Sea X una *v. a.* con *f. g. m.* m_X. Entonces,

$$m_X(t) = \sum_{r=0}^{\infty} \frac{t^r}{r!}\mathbb{E}[X^r]$$

Corolario : Sea X una *v. a.* con *f. g. m.* m_X diferenciable. Entonces, para $r = 1, 2, \ldots$, se tiene que

$$\mathbb{E}[X^r] = \frac{d^r m_X(t)}{dt^r}\Big|_{t=0}$$

Ejemplo :

- Si $X \sim \mathrm{Ber}(p)$, entonces $m_X(t) = (q + pe^t)$, para todo $t \in \mathbb{R}$, con $q = 1 - p$.
- $N \sim \mathrm{Po}(\lambda)$. entonces, $m_N(t) = \exp\{\lambda(e^t - 1)\}$, para todo $t \in \mathbb{R}$.
- $Z \sim \mathrm{N}(0, 1)$, entonces, $m_Z(t) = e^{t^2/2}$, para todo $t \in \mathbb{R}$.
- $Y \sim \mathrm{Gamma}(\alpha, \beta)$, entonces, $m_Y(t) = (1 - \beta t)^{-\alpha}$, para $t < 1/\beta$ y donde α y β son los parámetros de forma y escala de las distribución, respectivamente.

Proposición : Sea X una *v. a.* con *f. g. m.* m_X y sean a y b constantes. Entonces la *f. g. m.* de $Y = a + bX$ es

$$m_{a+bX}(t) = e^{at} \cdot m_X(bt)$$

Proposición : Sea X una *v. a.*. Si la *f. g. m.* $m_X(t)$ existe para todo $|t| < \delta$, entonces m_X determina de manera única la distribución de X.

Proposición : Una distribución de probabilidad no queda determinada completamente por sus momentos. Esto es, si X *v. a.*, se puede conocer $\mu_r = \mathbb{E}[X^r]$ para todo $r = 1, 2, \ldots$ y aún así no poder determinar completamente su distribución.

Demostración: Considere la distribución **lognormal** con *f. d. p.*

$$f_X(x) = \frac{1}{x\sqrt{\sigma\pi}} \exp\left\{-\frac{1}{2}(\log x)^2\right\} \mathbb{1}_{\mathbb{R}^+}(x)$$

y su *"perturbación"* Y con *f. d. p.*

$$f_Y(y) = f_X(y)\left[1 + \sin(2\pi \log y)\right] \mathbb{1}_{\mathbb{R}^+}(y)$$

En este caso se tiene que $\mathbb{E}[X^r] = \mathbb{E}[Y^r]$, para $r = 1, 2, \ldots$. Por lo que disponer de todos los momentos μ_r no se alcanza a distinguir de qué distribución exactamente se trataría.

4.2. Suma de variables aleatorias independientes

Teorema de Unicidad. Sean X y Y dos *v. a.*'s con F_X y F_Y, m_X y m_Y, sus *f. p. a.* y *f. g. m.* respectivamente,. Si $m_X(t) = m_Y(t)$, para todo $|t| < \delta$, entonces, $F_X(w) = F_Y(w)$ para todo $w \in \mathbb{R}$. Esto es, si dos *v. a.*'s tienen *f. g. m.* y son la misma entonces siguen la misma distribución de probabilidades.

Vea por ejemplo (Mood, Graybill, and Boes 1974) II.4.6 y (Blitzstein and Hwang 2014) 6.4 para leer más sobre funciones generadoras de momentos.

Recuerde, si $Y \sim \mathrm{N}(\mu, \sigma^2)$, su *f. d. p.* es $f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}$ y su *f. g. m.* $m_Y(t) = \exp\{\mu t + \frac{1}{2}\sigma^2 t^2\}$. Por otro lado, si Z sigue una distribución normal estándar, $m_Z(t) = \exp\{\frac{1}{2}t^2\}$. Luego, si a y b constantes y $X = a + bZ$, se sigue de proposición anterior que

$$m_X(t) = e^{at} m_X(bt) = \exp\{at + \frac{1}{2}b^2 t^2\}$$

que coincide con la *f. g. m.* de una distribución normal con media a y varianza b^2. Del teorema de unicidad se concluye que $X \sim \mathrm{N}(a, b^2)$.

Note que de la definición de X se sabía que $\mathbb{E}[X] = a + b\mathbb{E}[Z] = a$ y que $\mathrm{var}(X) = b^2 \mathrm{var}(Z) = b^2$, pero es el teorema de unicidad lo que permite concluir que X sigue una distribución normal.

Proposición : Sea $Z \sim \mathrm{N}(0, 1)$. Entonces $Y = Z^2 \sim \mathrm{Ga}(\alpha = 1/2, \beta = 2)$, por lo que $\mathbb{E}[Y] = 1$.

Demostración: Sea $Y = Z^2$. Entonces, se sigue de TEI que

$$m_Y(t) = \mathbb{E}[e^{tZ^2}] = \int_{\mathbb{R}} e^{tz^2}\phi(z)dz \overset{\text{TEI}}{=} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(1-2t)z^2} dz = \sigma \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}z^2} dz = \sigma$$

donde $\sigma = (1 - 2t)^{-1/2}$. Por lo tanto, $m_Y(t) = (1 - 2t)^{-1/2}$. La conclusión se sigue del teorema de unicidad.

Proposición : Sean X y Y *v. a.*'s independientes con *f. g. m.*'s m_X y m_Y, respectivamente. Entonces,

$$m_{X+Y}(t) = m_X(t) \cdot m_Y(t), \text{ para todo } t$$

Demostración:

$$\begin{aligned} m_{X+Y}(t) &= \mathbb{E}\left[e^{t(X+Y)}\right] \\ &= \mathbb{E}\left[e^{tX} \cdot e^{tY}\right] \\ &= \mathbb{E}\left[e^{tX}\right] \cdot \mathbb{E}\left[e^{tY}\right], \text{ por independencia de } X \text{ y } Y \\ &= m_X(t) \cdot m_Y(t) \end{aligned}$$

Corolario : Sean $X_1, \ldots, X_n$ *v. a.*'s independientes con $m_i(t)$ la correspondiente *f. g. m.* de X_i, $i = 1, \ldots, n$. Entonces, si $S_n = \sum_{i=1}^n X_i$, se tiene que

$$m_{S_n}(t) = \prod_{i=1}^n m_i(t)$$

Corolario : Sean $X_1, \ldots, X_n$ *v.a.i.i.d.*'s con *f. g. m.* $m_X(t)$. Entonces, para $S_n = \sum_{i=1}^n X_i$, se tiene que

$$m_{S_n}(t) = \left[m_X(t)\right]^n$$

Ejemplo : Sean $X_1, \ldots, X_n$ *v.a.i.i.d.* con $X_i \sim \mathrm{Ber}(p)$. Entonces, $S_n = \sum_{i=1}^n X_i \sim \mathrm{Bin}(n, p)$.

Solución: $m_{S_n}(t) = (q + pe^t)^n$, que corresponde a la *f. g. m.* de la distribución binomial con parámetros n y p. Se sigue del teorema de unicidad que $S_n \sim \mathrm{Bin}(n, p)$.

Ejemplo : Sean $Y_1, \ldots, Y_n$, *v. a.*'s independientes con $Y_i \sim \mathrm{Po}(\lambda_i)$. Entonces, $Y_1 + Y_2 \sim \mathrm{Po}(\lambda_1 + \lambda_2)$ y en general $Sn \sim \mathrm{Po}(\sum_{i=1}^n \lambda_i)$.

Solución: En ejemplo anterior se mostró que si X_1 y X_2 eran *v. a.*'s independientes con $X_i \sim \mathrm{Po}(\lambda_1)$, entonces $(X_1 + X_2) \sim \mathrm{Po}(\lambda_1 + \lambda_2)$. Se sigue de corolario anterior que si $S_n = \sum_{i=1}^n Y_i$, entonces,

$$m_{S_n}(t) = \prod_{i=1}^n m_i(t) = \prod_{i=1}^n \exp\{(e^t - 1)\lambda_i\} = \exp\{(e^t - 1)\sum_{i=1}^n \lambda_i\}$$

que corresponde a la *f. g. m.* de la distribución Poisson con media $\sum_{i=1}^n \lambda_i$. La conclusión del ejemplos se sigue del teorema de unicidad.

Ejemplo : Sean $X_1, \ldots, X_n$ *v. a.*'s independientes con $X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$. Entonces, $S_n \sim \mathrm{N}(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$.

Solución: note que

$$m_{S_n}(t) = \prod_{i=1}^n \exp\{\mu_i t + \frac{1}{2}\sigma_i^2 t^2\} = \exp\left\{\left(\sum \mu_i\right) t + \frac{1}{2}\left(\sum \sigma_i^2\right) t^2\right\}$$

El resultado se sigue nuevamente del teorema de unicidad.

Ejemplo : Sean $Y_1, \ldots, Y_n$ *v. a.*'s independientes con $Y_i \sim \mathrm{Gamma}(\alpha_i, \beta)$. Entonces, $S_n \sim \mathrm{Gamma}(\sum \alpha_i, \beta)$.

Definición : Sea Y *v. a.* que sigue una distribución Gamma con $\alpha = n/2$, $\beta = 2$, parámetros de forma y escala respectivamente. Si $\alpha = n/2$, $n \in \mathbb{N}$ y $\beta = 2$, Y se dice que sigue una **distribución Ji-cuadrada con n grados de libertad** y se denota por $Y \sim \chi_n^2$. Esto es, $\mathrm{Gamma}(\frac{n}{2}, 2) \equiv \chi_n^2$, con $n = 1, 2, \ldots$.

Corolario : Sea $Y \sim \chi_n^2$. Entonces $\mathbb{E}[Y] = n$ y $\mathrm{var}(Y) = 2n$.

Corolario : Sean $Y_1, \ldots, Y_n$ *v. a.*'s independientes con $Y_i \sim \chi_{m_i}^2$. Entonces, $S_n \sim \chi_m^2$, donde $m = \sum_{i=1}^n m_i$.

Demostración: se sigue de la definición de χ_n^2 y el ejemplo anterior.

Proposición : Sean $X_1, \ldots, X_n$ *v. a.*'s independientes con $X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$. Entonces,

$$\sum_{i=1}^n \left(\frac{X_i - \mu_i}{\sigma_i}\right)^2 \sim \chi_n^2$$

Demostración: Note que $Z_i = (X_i - \mu_i)/\sigma_i \sim \mathrm{N}(0, 1)$. Luego $Z_i^2 \sim \chi_1^2$ y $\sum_{i=1}^n Z_i^2 \sim \chi_n^2$.

4.3. Caso multivariado

Definición : Sea (X,Y) un vector aleatorio (*v. a.*). Se define la **función generadora de momentos conjunta** de (X,Y) por

$$m_{XY}(u,v) = \mathbb{E}[e^{uX+vY}]$$

siempre que el valor esperado exista para $|u| < \delta$, $|v| < \epsilon$ con $\delta, \epsilon > 0$.

Proposición : Sea (X,Y) *v. a.* con *f. g. m.* conjunta $m_{XY}(u,v)$. Entonces, las correspondientes *f. g. m.* marginales están dadas por

$$m_X(t) = \lim_{v\to 0} m_{XY}(t,v) = m_{XY}(t,0)$$
$$m_Y(t) = \lim_{u\to 0} m_{XY}(u,t) = m_{XY}(0,t)$$

Demostración:

$$m_X(t) = \mathbb{E}[e^{tX}] = \mathbb{E}[e^{tX+0\cdot Y}] = m_{XY}(t,0)$$

Proposición : Sea (X,Y) *v. a.* con *f. g. m.* conjunta $m_{XY}(u,v)$. Entonces, si el momento (r,s) conjunto existe y m_{XY} es diferenciable, se tiene que

$$\mathbb{E}[X^rY^s] = \frac{\partial^{r+s}}{\partial u^r \partial v^s} m_{XY}(u,v)\Big|_{u=0=v}$$

Ejemplo : Sea (X_1,X_2) un *v. a.* con *f. g. m.* conjunta dada por

$$m_{12}(t_1,t_2) = \exp\left\{-t_2 + \frac{1}{2}t_1^2 - t_1t_2 + t_2^2\right\}$$

Encuentre $\mathbb{E}[X_i]$, $\mathrm{var}(X_i)$ y $\mathrm{corr}(X_1,X_2)$.

Solución: Sean m_i la *f. g. m.* marginal de X_i.

I. $m_1(t_1) = m_{12}(t_1,0) = \exp\{t_1^2/2\}$.
Se sigue que $E[X_1] = m_1'(0) = 0$ y consecuentemente $\mathrm{var}(X_1) = \mathbb{E}[X_1^2] = m_1''(0) = 1$

II. $m_2(t_2) = m_{12}(0,t_2) = \exp\{-t_2 + t_2^2\}$.
$E[X_2] = m_2'(0) = -1$, $\mathbb{E}[X_2^2] = m_2''(0) = 3$. Entonces, $\mathrm{var}(X_2) = 2$.

III.
$$\mathbb{E}[X_1X_2] = \frac{\partial^2}{\partial t_1 \partial t_2} m_{12}(t_1,t_2)\Big|_{t_1=0=t_2} = -e^{-t_2+\frac{1}{2}t_1^2-t_1t_2+t_2^2}\Big|_{t_1=0=t_2} = -1$$
Luego, $\mathrm{cov}(X_1,X_2) = \mathbb{E}[X_1X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2] = -1$ y por lo tanto
$$\mathrm{corr}(X_1,X_2) = \frac{\mathrm{cov}(X_1,X_2)}{\sqrt{\mathrm{var}(X_1)\mathrm{var}(X_2)}} = -1/\sqrt{2}$$

Proposición : Sea (X_1,X_2) *v. a.* con *f. g. m.* conjunta m_{12} y marginales m_1 y m_2. X_1 y X_2 son *v. a.*'s independientes si y solo si

$$m_{12}(t_1,t_2) = m_1(t_1)\cdot m_2(t_2), \text{ para todo } \ t_1,t_2$$

Demostración

I. Sean X_1 y X_2 *v. a.*'s independientes. Entonces,

$$m_{12}(t_1,t_2) = \mathbb{E}[e^{t_1X_1+t_2X_2}] \overset{\text{ind}}{=} \mathbb{E}[e^{t_1X_1}]\cdot\mathbb{E}[e^{t_2X_2}] = m_1(t_1)\cdot m_2(t_2)$$

para todo t_1 y t_2 donde está definida la *f. g. m.* conjunta m_{12}.

II. Suponga ahora que $m_{12}(t_1,t_2) = m_1(t_1)\,m_2(t_2)$, para toda t_1 y t_2 y sean Y_1 y Y_2 *v. a.*'s independientes con *f. g. m.* m_i para $i = 1, 2$, respectivamente. Luego la *f. g. m.* conjunta de (Y_1,Y_2) es

$$m_{Y_1Y_2}(t_1,t_2) \overset{\text{ind}}{=} m_1(t_1)\cdot m_2(t_2) = m_{X_1}(t_1)\cdot m_{X_2}(t_2)$$

Se sigue del teorema de unicidad que (X_1,X_2) tiene la misma distribución que (Y_1,Y_2), por lo que para todo z_1, z_2, se tiene

$$F_{X_1X_2}(z_1,z_2) = F_{Y_1Y_2}(z_1,z_2) \overset{\text{ind}}{=} F_{Y_1}(z_1)F_{Y_2}(z_2) = F_{X_1}(z_1)F_{X_2}(z_2)$$

Por lo que $X_1 \sim Y_1$ y $X_2 \sim Y_2$ independientes.

4.4. Función característica

Definición : Se define la **función característica** (*f. c.*) de la *v. a.* X por

$$\varphi_X(t) = \mathbb{E}[e^{itX}]$$

para todo t y con $i = \sqrt{-1}$.

Corolario : Sea X *v. a.* con función generadora de momentos m_X y función característica φ_X. Entonces,

$$\varphi_X(t) = m_X(it)$$

para todo t donde m_X exista.

Proposición : Sea X *v. a.* continua con función de densidad de probabilidad f_X. Entonces, su función característica φ_X siempre existe.

Demostración Recuerde que $e^{ix} = \cos x + i\sin x$, por lo que $|e^{ix}|^2 = \cos^2 x + \sin^2 x = 1$. Luego, se tiene para todo t,

$$|\varphi_X(t)| = \left|\int_{\mathbb{R}} e^{itx} f_X(x)dx\right| \le \int_{\mathbb{R}} |e^{itx}| f_X(x)dx = \int_{\mathbb{R}} 1\cdot f_X(x)dx = 1$$

Así, la función característica *siempre existe*.

La proposición anterior se presentó el caso de X *v. a.* continua, pero es el mismo resultado si X es una *v. a.* discreta. Sustituya la integral por suma.

Recuerde que toda variable aleatoria X tiene su función de probabilidad acumulada o de distribución F_X que tiene asociada una función (densidad, masa o mixta) de probabilidad f_X. La definición de función característica y la proposición anterior muestran que la distribución de X también tiene asociada φ_X.

Fórmula de Inversión [(Harris 1966), Teorema 4-5.4] Sea h cualquier número positivo y sean x y $x+h$ un par de puntos de continuidad de F_X. Entonces,

$$F_X(x+h) - F_X(x) = \lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T} \frac{1-e^{-ith}}{it} e^{-itx}\varphi_X(t)dt$$

Se sigue del resultado anterior que toda función característica tiene asociada una función de distribución por lo que φ_X es una manera de *caracterizar* la distribución de X y de ahí su nombre.

Ejemplo :

- $X \sim \text{Bin}(n, p)$. Entonces, $\varphi_X(t) = m_X(it) = (q + pe^{it})^n$.
- $N \sim \text{Po}(\lambda)$. Entonces, $\varphi_N(t) = \exp\{\lambda(e^{it} - 1)\}$.
- $X \sim \text{N}(\mu, \sigma^2)$. Entonces, $\varphi_X(t) = \exp\{i\mu t - \frac{1}{2}\sigma^2 t^2\}$.
- $Y \sim \text{Gamma}(\alpha, \beta)$. Entonces, $\varphi_Y(t) = (1 - i\beta t)^{-\alpha}$.
- $W \sim \text{Cauchy}(0, 1)$. Entonces, $\varphi_W(t) = e^{-|t|}$.

Se sabe que la distribución Cauchy no tiene valor esperado. Por otro lado, la proposición anterior indica que todas las funciones de probabilidad tiene su correspondiente función característica y el ejemplo muestra que la distribución Cauchy tiene como *f. c.* $\varphi(t) = e^{-|t|}$. En este caso no es posible usar el resultado $\varphi(t) = m_X(it)$ pues la *f. g. m.* no existe y si existiese note que la función φ no es diferenciable en 0.

Proposición : Sea X *v. a.* con *f. p. a.* $F_X(x)$ y *f. c.* $\varphi_X(t)$. Entonces, $\varphi_X(t)$ es real si y solo si F_X es simétrica, en el sentido que $F_X(-x) = 1 - F_X(x)$, para todo x.

Demostración: Vea Apéndice 3 de (Harris 1966).

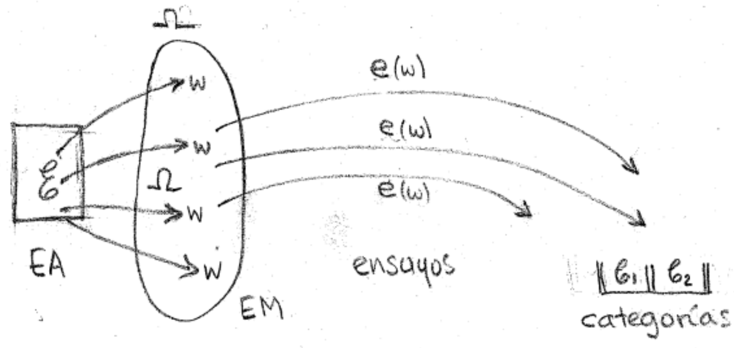
Note que efectivamente la *f. d. p.* de la distribución Cauchy(0,1) es simétrica al rededor de 0 y su *f. c.* es una función real. Lo mismo sucede con la distribución normal centrada en 0, su correspondiente *f. c.* es $\varphi(t) = e^{-\sigma^2 t^2/2}$, función real.

4.5. Ejercicios

Refiérase al Cuaderno de Ejercicios sección 4, Barrios and Chambon (2024).

5. Distribución multinomial

5.1. Distribución binomial



Considere el espacio de probabilidad $(\Omega, \mathcal{S}, \mathbb{P})$ y suponga que un experimento aleatorio arroja salidas $\omega \in \Omega$. Las salidas o **ensayos** $e(\omega)$ son clasificado en una de dos clases o categorías $\mathcal{C}_1$ y $\mathcal{C}_2$ que son *exhaustivas* y *exclusivas*. Exclusivas porque cada ensayo es asignado a una y solo una de las clases y exhaustivo porque todo ensayo es clasificado.

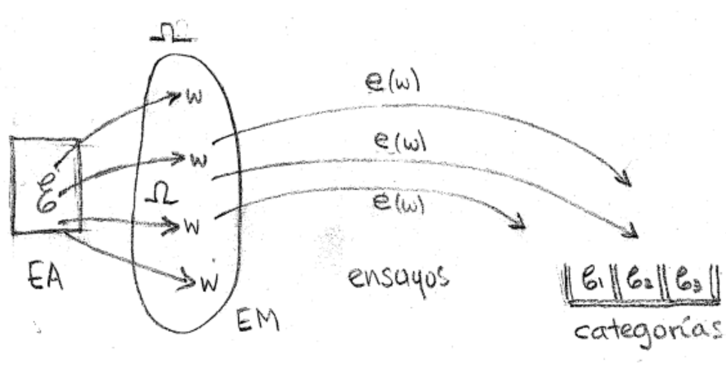
Suponga que la probabilidad de que un ensayo e sea clasificado $\mathcal{C}_i$ es $\mathbb{P}(e \in \mathcal{C}_i) = p_i$. Se sigue de la exhaustividad y exclusividad de las clases que $p_1 + p_2 = 1$.

Considere ahora $\{e_\ell\}_{\ell=1}^n$, n ensayos independientes y sean X_i el número de ensayos que fueron clasificados $\mathcal{C}_i$, $i = 1, 2$. Luego, $X_i = \sum_{\ell=1}^n \mathbb{1}_{\mathcal{C}_i}(e_\ell)$ y $n = X_1 + X_2$. Así, para $x_1 = 0, 1, \ldots, n$,

$$\begin{aligned}
\mathbb{P}(X_1 = x_1) &= \binom{n}{x_1} p_1^{x_1}(1-p_1)^{n-x_1} \\
&= \binom{n}{x_1}\binom{n-x_1}{x_2} p_1^{x_1} p_2^{x_2} \\
&= \frac{n!}{x_1(n-x_1)!}\,\frac{(n-x_1)!}{x_2!(n-x_1-x_2)!}\,p_1^{x_1}p_2^{x_2} \\
\mathbb{P}(X_1 = x_1, X_2 = x_2) &= \frac{n!}{x_1!\,x_2!} pi^{x_1} p_2^{x_2} \mathbb{1}_{\{x_i : x_1 + x_2 = n\}}(x)
\end{aligned}$$

donde $x = (x_1, x_2)$.

5.2. Distribución trinomial



Considere el caso donde cada experimento e puede ser clasificado en una y solo una de tres categorías, exhaustivas y exclusivas, $\mathcal{C}_1, \mathcal{C}_2$ y $\mathcal{C}_3$. Nuevamente, $\mathbb{P}(e \in \mathcal{C}_i) = p_i$, $i = 1, 2, 3$ y $1 = p_1 + p_2 + p_3$.

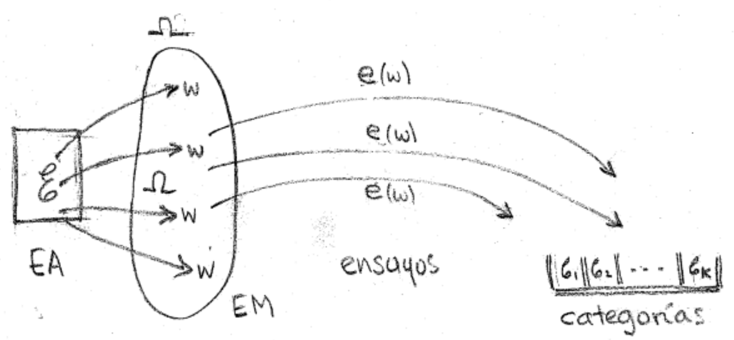
Considere $\{e_\ell\}_{\ell=1}^n$, n ensayos independientes y sean X_i el número de ensayos que fueron clasificados en la categoría $\mathcal{C}_i$. Luego, $X_i = \sum_{\ell=1}^n \mathbb{1}_{\mathcal{C}_i}(e_\ell)$ y $n = X_1 + X_2 + X_3$. Sean $x_i \in \{0, 1, \ldots, n\}$ tal que $n = x_1 + x_2 + x_3$, entonces,

$$\begin{aligned}
\mathbb{P}\left(X_1 = x_1\right) &= \binom{n}{x_1} p_1^{x_1} (1-p_1)^{n-x_1} \\
\mathbb{P}\left(X_1 = x_1, X_2 = x_2\right) &= \binom{n}{x_1} p_1^{x_1} \binom{n-x_1}{x_2} p_2^{x_2} (1-p_1-p_2)^{n-p_1-p_2} \\
\mathbb{P}\left(X_1 = x_1, X_2 = x_2, X_3 = x_3\right) &= \binom{n}{x_1} p_1^{x_1} \binom{n-x_1}{x_2} p_2^{x_2} \binom{n-x_1-x_2}{x_3} p_3^{x_3} \\
&= \frac{n!}{x_1!(n-x_1)!} \frac{(n-x_1)!}{x_2!(n-x_1-x_2)!} \frac{(n-x_1-x_2)!}{x_3!(n-x_1-x_2-x_3)!} \\
&\quad \cdot p_1^{x_1} p_2^{x_2} p_3^{x_3} \\
&= \frac{n!}{x_1!\, x_2!\, x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3}
\end{aligned}$$

Por lo que la función masa de probabilidad conjunta del vector aleatorio $X = (X_1, X_2, X_3)$ está dada por

$$f_X(x) = \frac{n!}{x_1!\, x_2!\, x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3} \mathbb{1}_{\{x_i : x_1 + x_2 + x_3 = n\}}(x)$$

5.3. Distribución multinomial



Suponga ahora el caso donde cada experimento e puede ser clasificado en una y solo una de k categorías, exclusivas y exhaustivas, $\mathcal{C}_1, \ldots, \mathcal{C}_k$. Nuevamente, $\mathbb{P}(e \in \mathcal{C}_i) = p_i$, $i = 1, \ldots, k$ y $1 = p_1 + \cdots + p_k$.

Considere $\{e_\ell\}_{\ell=1}^n$, n ensayos independientes y sean X_i el número de ensayos que fueron clasificados en la categoría $\mathcal{C}_i$. Luego, $X_i = \sum_{\ell=1}^n \mathbb{1}_{\mathcal{C}_i}(e_\ell)$ y $n = X_1 + \cdots + X_k$. De la misma manera al caso de la trinomial, ahora con k categorías, se deriva f_X, la función masa de probabilidad conjunta de $X = (X_1, \ldots, X_k)$.

Sean $x_i \in \{0, 1, \ldots, n\}$ tal que $n = x_1 + \cdots + x_k$, luego

$$\begin{aligned}
f_X(x) &= \mathbb{P}\left(X_1 = x_1, \ldots, X_k = x_k\right) \\
&= \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k} \\
&= \binom{n}{x_1 \cdots x_k} p_1^{x_1} \cdots p_k^{x_k}
\end{aligned}$$

donde $\binom{n}{x_1 \cdots x_k} = \dfrac{n!}{x_1! \cdots x_k!}$ define el llamado **coeficiente multinomial**.

Definición : El vector aleatorio discreto $X = (X_1, \ldots, X_k)$ con *f. m. p.* conjunta dada por

$$f_X(x) = \binom{n}{x_1 \cdots x_k} p_1^{x_1} \cdots p_k^{x_k} \mathbb{1}_{\{x_i \in \mathbb{N}_0 : x_1 + \cdots + x_k = n\}}(x)$$

se dice que sigue una **distribución multinomial** con parámetros $n, p_1, \ldots, p_k$ y se denota $X \sim \text{Multinomial}(n, p_1, \ldots, p_k)$.

Corolario : Sea $X \sim \text{Multinomial}(n, p_1, \ldots, p_k)$. Entonces, marginalmente $X_i \sim \text{Bin}(n, p_i)$. Luego, $\mathbb{E}[X_i] = np_i$ y $\text{var}(X_i) = np_i(1 - p_i)$, para $i = 1, \ldots, k$.

Demostración: Considere la i–ésima clase $\mathcal{C}_i$ y el resto de las otras clases $\mathcal{C}_i^C$. Sea $Y = X_i$, luego $Y \sim \text{Bin}(n, p_i)$, por lo que $\mathbb{E}[X_i] = np_i$ y $\text{var}(X_i) = np_1(1 - p_i)$.

Proposición : Sea $X \sim \text{Multinomial}(n, p_1, \ldots, p_k)$. Entonces, para $i \neq j$, $\text{cov}(X_i, X_j) = -np_ip_j$.

Demostración: Se sigue la idea presentada por Ross (2006).

Sean $\{e_\ell\}_{\ell=1}^n$, n ensayos independientes y $p_i = \mathbb{P}(e_\ell \in \mathcal{C}_i)$, $i = 1, \ldots, k$. Sea X_i el número de ensayos clasificados en la categoría $\mathcal{C}_i$. Luego, $X_i = \sum_{\ell=1}^n \mathbb{1}_{\mathcal{C}_i}(e_\ell)$

I. Suponga $i \neq j$, entonces

$$\begin{aligned}
\text{cov}\left(\mathbb{1}_{\mathcal{C}_i}(e_u), \mathbb{1}_{\mathcal{C}_j}(e_u)\right) &= \mathbb{E}\left[\mathbb{1}_{\mathcal{C}_i}(e_u) \cdot \mathbb{1}_{\mathcal{C}_j}(e_u)\right] - \mathbb{E}\left[\mathbb{1}_{\mathcal{C}_i}(e_u)\right] \cdot \mathbb{E}\left[\mathbb{1}_{\mathcal{C}_j}(e_u)\right] \\
&= -p_ip_j
\end{aligned}$$

pues la esperanza del producto de indicadoras es siempre 0 pues $i \neq j$ y la exclusividad de las clases. Se sigue que

$$\begin{aligned}
\text{cov}\left(X_i, X_j\right) &= \text{cov}\left(\sum_{u=1}^n \mathbb{1}_{\mathcal{C}_i}(e_u), \sum_{v=1}^n \mathbb{1}_{\mathcal{C}_j}(e_v)\right) \\
&= \sum_{u=1}^n \sum_{v=1}^n \text{cov}\left(\mathbb{1}_{\mathcal{C}_i}(e_u), \mathbb{1}_{\mathcal{C}_j}(e_v)\right) \\
&= \sum_{u=1}^n \text{cov}\left(\mathbb{1}_{\mathcal{C}_i}(e_u), \mathbb{1}_{\mathcal{C}_j}(e_u)\right) \\
&= -np_ip_j
\end{aligned}$$

pues si $u \neq v$, se sigue por independencia de los ensayos que $\text{cov}\left(\mathbb{1}_{\mathcal{C}_i}(e_u), \mathbb{1}_{\mathcal{C}_j}(e_v)\right) = 0$.

II. Si $i = j$, $\text{cov}(X_i, X_j) = \text{var}(X_i) = np_i(1 - p_i)$.

Corolario . Sea $X \sim \text{Multinomial}(n, p_1, \ldots, p_k)$. Entonces, para $i \neq j$,

$$\text{corr}(X_i, X_j) = -\sqrt{\frac{p_i}{1 - p_i} \frac{p_j}{1 - p_j}}$$

Demostración: Se sigue directamente de la definición de correlación.

Proposición : Sea el *v. a.* $X \sim \text{Multinomial}(n, p_1, \ldots, p_k)$. Entonces X tine como *f. g. m.* conjunta

$$m_X(t) = \mathbb{E}\left[\mathrm{e}^{t^T X}\right] = \left(p_1 \mathrm{e}^{t_1} + \cdots + p_k \mathrm{e}^{t_k}\right)^n$$

Demostración: La proposición se sigue del **teorema multinomial**. A saber,

$$(a_1 + \cdots + a_k)^n = \sum_{\{i_j \in \mathbb{N}_0 : \sum_1^k i_j = n\}} \binom{n}{i_1 \cdots i_k} a_1^{i_1} \cdots a_k^{i_k}$$

Sea $X = (X_1, \ldots, X_k) \sim \text{Multinomial}(n; p_1, \ldots, p_k)$ con soporte $\mathcal{S}_\text{X}$, esto es, $\mathcal{S}_\text{X} = \left\{x \in \mathbb{N}_0^k : x_1 + \cdots x_k = n\right\}$, luego

$$\begin{aligned}
m_X(t) &\stackrel{\text{def}}{=} \mathbb{E}\left[\mathrm{e}^{\sum t_i X_i}\right] \\
&= \sum_{x \in \mathcal{S}_\text{X}} \mathrm{e}^{t^T x} f_X(x) \\
&= \sum_{\{i_j \in \mathbb{N}_0 : \sum_1^k i_j = n\}} e^{t_1 i_1 + \cdots t_k i_k} \binom{n}{i_1 \cdots i_k} p_1^{i_1} \cdots p_k^{i_k} \\
&= \sum_{\{i_j \in \mathbb{N}_0 : \sum_1^k i_j = n\}} \binom{n}{i_1 \cdots i_k} (\mathrm{e}^{t_1} p_1)^{i_1} \cdots (\mathrm{e}^{t_k} p_k)^{i_k} \\
&= \left(p_1 e^{t_1} + \cdots + p_k \mathrm{e}^{t_k}\right)^n
\end{aligned}$$

Ejemplo : (Wackerly et al. (2008) Ej. 5.120) Considere una muestra con n observaciones de un lote de gran tamaño. Sea p_1 la proporción de artículos con un defecto y p_2 la proporción de artículos con más de un defecto. Luego, $0 \leq p_1 + p_2 \leq 1$. El costo de reparar el lote es $C = Y_1 + 3Y_2$, donde Y_1 y Y_2 denotan el número de artículos con uno o más de un defecto respectivamente. Si $n = 100$, $p_1 = 0.3$ y $p_2 = 0.2$, determine la media y la varianza de C.

Solución: Sea Y_0 el número de artículos sin defecto. Luego $p_0 = 1 - p_1 - p_2$ es la probabilidad de que un artículo no tenga defectos. Entonces, $n = Y_0 + Y_1 + Y_2$ y $Y = (Y_0, Y_1, Y_2) \sim \text{Multinomial}(n, p_0, p_1, p_2)$. Sea $C = Y_1 + 3Y2$ el costo de reparar los artículos. Se sigue que

a). $\mathbb{E}[C] = \mathbb{E}[Y_1] + 3\mathbb{E}[Y_2] = np_1 + 3np_2 = n\,(p_1 + 3p_2) = 100(.3 + 3 \cdot .2) = 90.$

b).

$$\begin{aligned}
\text{var}(C) &= \text{var}(Y_1) + 9\text{var}(Y_2) + 2(3)\text{cov}(Y_1, Y_2) \\
&= n\left[p_1(1-p_1) + 9p_2(1-p_2) - 6p_1p_2\right] \\
&= 100\left[.3(.7) + .2(.8) - 6(.3)(.2)\right] \\
&= 129
\end{aligned}$$

5.4. Ejercicios

Refiérase al Cuaderno de Ejercicios sección 5, Barrios and Chambon (2024).

6. Distribución normal multivariada

6.1. La distribución normal bivariada

Definición : El vector bivariado (X_1, X_2) con función de densidad de probabilidad conjunta dada por

$$f_{12}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times \tag{4}$$
$$\times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right\}$$

se dice que sigue una **distribución normal bivariada** con parámetros μ_1, μ_2, σ_1^2, σ_2^2 y ρ.

Note que el exponente de la función de densidad es la ecuación de una elipse centrada en (μ_1, μ_2), de semiejes σ_1 y σ_2 y rotada ρ. Luego, para valores fijos de $f_{12}(x_1, x_2)$, el lugar geométrico descrito por los puntos $\{(x_1, x_2)\}$ son elipses definidas por la expresión (4). La figura 4 muestra la función de densidad para distintos valores de los parámetros.

Se desea encontrar las funciones densidad de probabilidad (*f. d. p.*) marginales de X_1 y X_2, las *f. d. p.* condicionales de X_i dado $X_j = x_j$ y la correlación entre X_1 y X_2. Para esto se sigue el procedimiento empleado en (Ross 2006).

Proposición : Considere el vector aleatorio (*v. a.*) bivariado (X_1, X_2) con *f. d. p.* conjunta dada por al expresión (4). Se sigue entonces que dado $X_2 = x_2$, X_1 sigue una distribución normal con media $\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2)$ y varianza $\sigma_1^2(1-\rho^2)$.

Demostración: Sea f_{12} la *f. d. p.* conjunta de X_1 y X_2 y sea f_j la marginal de X_j, $j = 1, 2$. En el siguiente desarrollo las constantes k_i resumen constantes y términos que no dependen de x_1. Luego,

$$\begin{aligned}
f(x_1|x_2) &= \frac{f_{12}(x_1, x_2)}{f_2(x_2)} \\
&= k_1 f_{12}(x_1, x_2) \\
&= k_2 \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)\right]\right\} \\
&= k_3 \exp\left\{-\frac{1}{2\sigma_1^2(1-\rho^2)}\left[x_1^2 - 2x_1\left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2-\mu_2)\right)\right]\right\} \\
&= k_4 \exp\left\{-\frac{1}{2\sigma_1^2(1-\rho^2)}\left[x_1 - \left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2-\mu_2)\right)\right]^2\right\}
\end{aligned}$$

y donde la exponencial corresponde al núcleo de la función de densidad de una distribución normal univariada con media $\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2)$ y varianza $\sigma_1^2(1-\rho^2)$. Se concluye entonces que $k_4 = 1/\left[\sqrt{2\pi}\sqrt{\sigma_1^2(1-\rho^2)}\ \right]$ y

$$(X_1|X_2 = x_2) \sim \mathrm{N}\left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2-\mu_2), \sigma_1^2(1-\rho^2)\right)$$

Similarmente se tiene que $(X_2|X_1 = x_1) \sim \mathrm{N}\left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1), \sigma_2^2(1-\rho^2)\right)$.

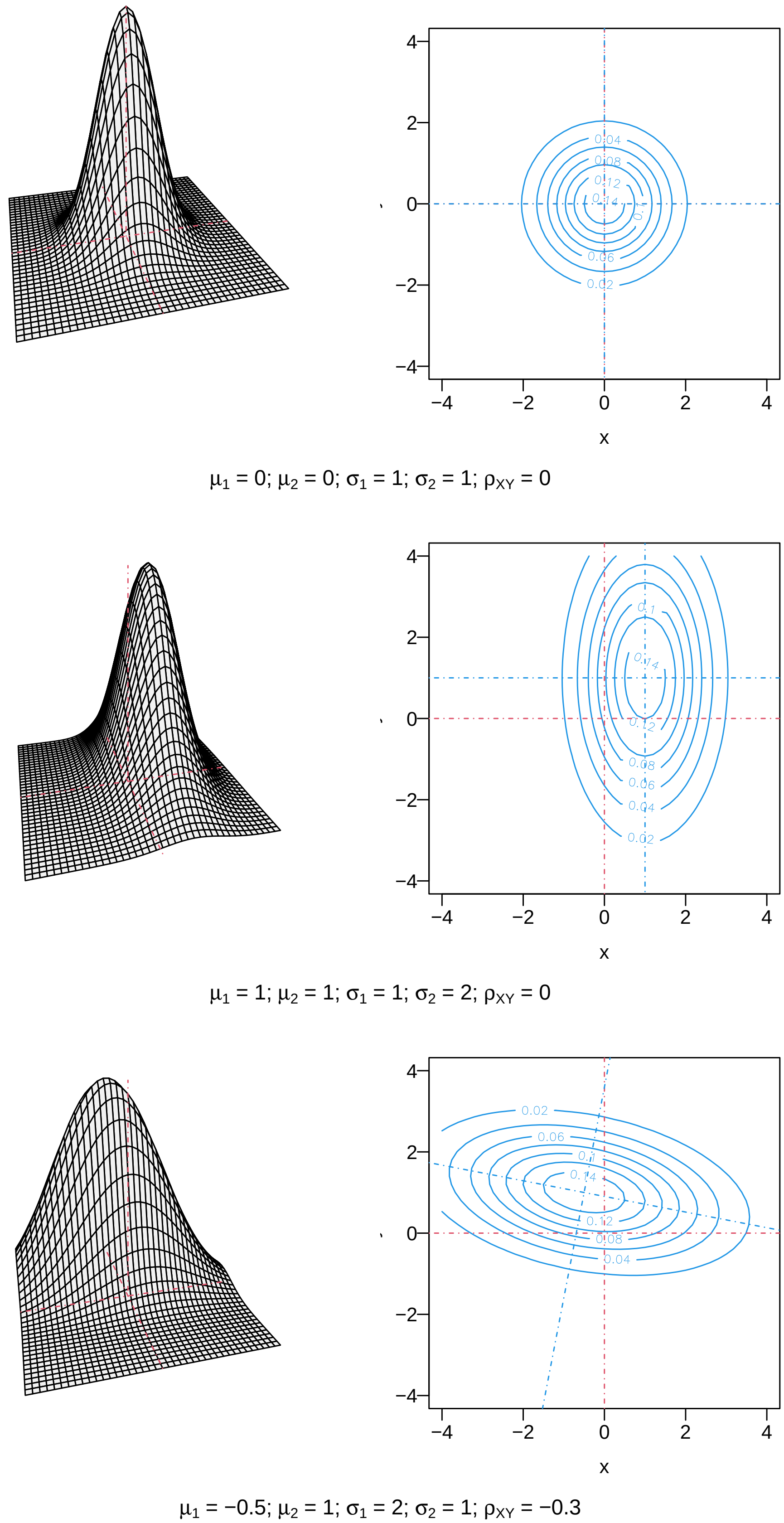


Figura 4: Funciones de densidad y correspondientes curvas de nivel (contornos) de la distribución normal bivariada para distintos valores de medias (μ_i), varianzas (σ_i^2) y coeficiente de correlación (ρ).

Proposición : Sea (X_1, X_2) un vector aleatorio bivariado con función de densidad conjunta dada por la expresión (4), entonces marginalmente X_i se distribuye normal univariada con media μ_i y varianza σ_i^2, $i = 1, 2$.

Demostración: Sea $k = \dfrac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$. Luego, la *f. d. p.* marginal de X_1 será

$$\begin{aligned}
f_1(x_1) &= \int_{\mathbb{R}} f_{12}(x_1, x_2)dx_2 \\
&= k \cdot \int_{\mathbb{R}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right\} dx_2 \\
&= k\sigma_2 \exp\left\{-\frac{z_1^2}{2(1-\rho^2)}\right\} \cdot \int_{\mathbb{R}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[z_2^2 - 2\rho z_1 z_2\right]\right\} dz_2 \\
&\qquad \text{con los cambios de variable } z_i = \frac{x_i - \mu_i}{\sigma_i},\ i = 1, 2 \\
&= k\sigma_2 \exp\left\{-\frac{z_1^2}{2(1-\rho^2)}(1-\rho^2)\right\} \cdot \int_{\mathbb{R}} \exp\left\{-\frac{1}{2(1-\rho^2)}(z_2 - \rho z_1)^2\right\} dz_2 \\
&= k\sigma_2 \exp\left\{-\frac{1}{2}z_1^2\right\} \cdot \sqrt{2\pi(1-\rho^2)} \\
&= \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right\}
\end{aligned}$$

que corresponde a la *f. d. p.* de una distribución normal univariada de media μ_1 y varianza σ_1^2. Entonces, marginalmente $X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$ para $i = 1, 2$.

Proposición : Sea (X_1, X_2) un vector aleatorio bivariado con función de densidad conjunta dada por la expresión (4), entonces el parámetro ρ es el coeficiente de correlación entre las componentes X_1 y X_2. Esto es, $\rho = \mathrm{corr}(X_1, X_2)$.

Demostración: Como se mostró anteriormente, por ejemplo,
$(X_1|X_2 = x_2) \sim \mathrm{N}\left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2), \sigma_1^2(1-\rho^2)\right)$. Luego,

$$\begin{aligned}
\mathbb{E}[X_1X_2|X_2 = x_2] &= \mathbb{E}[x_2X_1|X_2 = x_2] \\
&= x_2\mathbb{E}[X_1|X_2 = x_2] \\
&= x_2\left[\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2)\right]
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[X_1X_2] &= \mathbb{E}\left[\mathbb{E}[X_1X_2|X_2]\right] \\
&= \mathbb{E}\left[\mu_1X_2 + \rho\frac{\sigma_1}{\sigma_2}(X_2 - \mu_2)X_2\right] \\
&= \mu_1\mu_2 + \rho\frac{\sigma_1}{\sigma_2}\mathbb{E}\left[(X_2^2 - \mu_2X_2)\right] \\
&= \mu_1\mu_2 + \rho\frac{\sigma_1}{\sigma_2}\left(\mathbb{E}[X_2^2] - \mu_2^2\right) \\
&= \mu_1\mu_2 + \rho\sigma_1\sigma_2
\end{aligned}$$

Por lo que

$$\mathrm{cov}(X_1, X_2) = \mathbb{E}[X_1X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2] = \rho\sigma_1\sigma_2$$

y se sigue el resultado, $\mathrm{corr}(X_1, X_2) = \rho$.

Definición : El vector bivariado (X_1, X_2) con función de densidad de probabilidad conjunta dada por la expresión (4) se dice que sigue una **distribución normal bivariada** con medias μ_1 y μ_2, varianzas σ_1^2 y σ_2^2 y coeficiente de correlación ρ.

Una vez identificados los parámetros de la distribución, es decir, habiendo interpretado las medias μ's, las desviaciones estándar σ's y el coeficiente de correlación ρ, refiérase nuevamente a la figura 4. El panel superior muestra la densidad y contornos de la función de densidad normal bivariada centrada en cero y varianzas iguales. El panel central muestra el caso de una densidad salida del origen y mayor varianza de la segunda componente. Finalmente, el panel inferior muestra una densidad nuevamente fuera del origen, con distintas varianzas y rotada al tener componentes correlacionadas.

Proposición : Sea (X_1, X_2) un vector aleatorio bivariado con función de densidad conjunta dada por la expresión (4). Las componentes X_1 y X_2 son independientes, si y sólo si $\rho = 0$.

Demostración: Si $\rho = 0$ en la expresión (4), se sigue que

$$f_{12}(x_1, x2) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right\} \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right\} = f_1(x_1)\cdot f_2(x_2)$$

Esto es, la *f. d. p.* conjunta de X_1 y X_2 es el producto de las *f. d. p.* marginales. Luego X_1 y X_2 son componentes independientes.

Y viceversa, si X_1 y X_2 son independientes, entonces su correlación es cero, i.e. $\rho = 0$.

Proposición : Un vector normal (bivariado) siempre induce marginales normales pero no viceversa necesariamente.

Demostración: Se sigue del resultado 2 anterior que las marginales de una distribución normal bivariada son normales. Además, del resultado 4, si las componentes son distribuidas normales independientes, la distribución conjunta es normal bivariada. Sin embargo, considere el siguiente caso:

Sea $X_1 \sim \mathrm{N}(0, 1)$ y defina

$$X_2 = \begin{cases} -X_1 & \text{si } |X_1| \le 1 \\ X_1 & \text{si } |X_1| > 1 \end{cases}$$

La variable aleatoria X_2 sigue también una normal estándar. Considere F_2 la función de probabilidad acumulada de X_2 y verifique que

$$F_2(x) = \mathbb{P}(X_2 \le x) = \mathbb{P}(X_1 \le x) = \Phi(x)$$

para todos los distintos casos: $-\infty < x \le -1$; $-1 < x \le 0$; $0 < x \le 1$ y $x > 1$. Se tiene además que $\mathbb{P}(X_1+X_2 = 0) = \mathbb{P}(|X_1| \le 1) \approx 0.68$, que contradice el hecho de que si (X_1, X_2) sigue una normal bivariada entonces $\mathbb{P}(X_1 + X_2 = 0) = 0$. Así, no siempre componentes normales generan un vector normal multivariado.

Ejercicio : Suponga que las *v. a.*'s X y Y tienen la *f. d. p.* conjunta

$$f_{X,Y}(x, y) = k \cdot \exp\left\{-\frac{1}{2}\left[\frac{4}{3}x^2 + \frac{16}{3}y^2 + \frac{8}{3}xy - 8x - 16y + 16\right]\right\}$$

a). Encuentre la constante k apropiada ($k = \frac{2}{\pi\sqrt{3}}$)
(Sugerencia: De la expresión (4), $k = 1/\left(2\pi\sigma_X\sigma_Y\sqrt{1-\rho}\right)$. Luego, defina tres ecuaciones lineales en σ_X, σ_Y y ρ y resuelva. Mismo caso para obtener las medias μ_X y μ_Y.)

b). Determine las densidades marginales de X y Y.

c). Encuentre los siguientes momentos: $\mathbb{E}[X]$, $\mathrm{var}(Y)$, $\mathrm{cv}(X)$, $\textsc{snr}(Y)$.
(cv $= \sigma/|\mu|$ denota **el coeficiente de variación** y $\textsc{snr}=|\mu|/\sigma$ **la relación señal-ruido**.)

d). Determine $\mathrm{cov}(X,Y)$ y $\mathrm{corr}(Y,X)$.

e). Encuentre la distribución condicional de X dado $Y=2$ y la condicional de Y dado $X=-1$.

f). Determine $\mathbb{E}[X|Y=-1]$ y $\mathrm{var}(Y|X=2)$.

g). Calcule nuevamente $\mathbb{E}[X]$ y $\mathrm{var}(Y)$ utilizando los momentos condicionales $\mathbb{E}[X|Y]$ y $\mathrm{var}(Y|X)$.

6.2. La distribución normal multivariada

Sean Z_1 y Z_2 dos variables aleatorias independiente e idénticamente distribuidas (*v.a.i.i.d.*) normal estándar. Entonces, la *f. d. p.* conjunta se puede escribir como

$$f(z_1,z_2)=\frac{1}{2\pi}e^{-\frac{1}{2}(z_1^2+z_2^2)}=\frac{1}{2\pi}e^{-\frac{1}{2}\boldsymbol{z}^T\boldsymbol{z}}=f(\boldsymbol{z})$$

con $\boldsymbol{z}$ un vector columna bivariado y donde $\boldsymbol{z}^T$ denota el vector fila (z_1,z_2).

Sean ahora $X_i\sim \mathrm{N}(\mu_i,\sigma_i^2)$, $i=1,2$, *v. a.*'s independientes. Entonces, la *f. d. p.* conjunta se puede escribir como

$$\begin{aligned}
f_{12}(x_1,x_2)&=\frac{1}{2\pi\sigma_1\sigma_2}\exp\left\{-\frac{1}{2}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2+\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right\}\\
&=\frac{1}{2\pi\sigma_1\sigma_2}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\begin{pmatrix}\sigma_{11}&0\\0&\sigma_{22}\end{pmatrix}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}\\
&=f_{12}(\boldsymbol{x})
\end{aligned}$$

donde

$$\boldsymbol{x}=\begin{bmatrix}x_1\\x_2\end{bmatrix},\quad \mathbb{E}[\boldsymbol{X}]=\begin{bmatrix}\mu_1\\\mu_2\end{bmatrix},\quad \mathrm{cov}(\boldsymbol{X})=\begin{pmatrix}\sigma_{11}&0\\0&\sigma_{22}\end{pmatrix}\quad \text{y}\quad (\boldsymbol{x}-\boldsymbol{\mu})=\begin{bmatrix}x_1-\mu_1\\x_2-\mu_2\end{bmatrix}$$

Similarmente, si $X_i\sim\mathrm{N}(\mu_i,\sigma_i^2)$, $i=1,2$, *v. a.*'s con $\rho=\mathrm{corr}(X_1,X_2)$, entonces su *f. d. p.* conjunta está dada por la expresión (4) y ésta se puede escribir como

$$f(\boldsymbol{x})=\frac{1}{2\pi|\Sigma|^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}$$

donde $\boldsymbol{x}=(x_1,x_2)^T$ y

$$\boldsymbol{X}=\begin{bmatrix}X_1\\X_2\end{bmatrix},\quad \mu=\mathbb{E}[\boldsymbol{X}]=\begin{bmatrix}\mu_1\\\mu_2\end{bmatrix}\ \text{y}\quad \Sigma=\mathrm{cov}(\boldsymbol{X})=\begin{pmatrix}\sigma_{11}&\sigma_{12}\\\sigma_{21}&\sigma_{22}\end{pmatrix}$$

$\sigma_{ii}=\sigma_i^2$, $\sigma_{12}=\mathrm{cov}(X_1,X_2)=\rho\sigma_1\sigma_2$ y $|\Sigma|$ denota el determinante de la matriz de covarianzas.

Definición : En general se dice que el vector aleatorio p-variado $\boldsymbol{X} = (X_1, \ldots, X_p)^T$ con función de densidad de probabilidad dada por

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = (2\pi)^{-\frac{p}{2}}|\Sigma|^{-\frac{1}{2}} \ \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\} \tag{5}$$

sigue una **distribución normal multivariada** con vector de medias μ y matriz de covarianzas Σ

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}, \quad \boldsymbol{\mu} = \mathbb{E}[\boldsymbol{X}] = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} \quad \text{y} \quad \Sigma = \operatorname{cov}(\boldsymbol{X}) = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix}$$

y se denota como $\boldsymbol{X} \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$.

Note que la distribución normal multivariada queda completamente determinada por sus primeros momentos $\boldsymbol{\mu}$ y Σ.

Proposición : $\boldsymbol{X} \sim \mathrm{N}_p(\mu, \Sigma)$ tiene componentes independientes si y sólo si $\sigma_{ij} = 0$ para $i \neq j$. Esto es, $\boldsymbol{X}$ distribuida normal multivariada tiene componentes independientes si y solo si la matriz de covarianzas Σ es una matriz diagonal.

Demostración: Si $\boldsymbol{X}$ tiene componentes independientes $\sigma_{ij} = 0$ para $i \neq j$. Si supone ahora que $\sigma_{ij} = 0$ para $i \neq j$, se sigue que $\Sigma = \operatorname{diag}\{\sigma_1^2, \cdots, \sigma_p^2\}$ y

$$\begin{aligned} f_{\boldsymbol{X}}(\boldsymbol{x}) &= (2\pi)^{-\frac{p}{2}}|\Sigma|^{-\frac{1}{2}} \ \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\} \\ &= \frac{1}{(2\pi)^{p/2}\sigma_1\cdots\sigma_p} \exp\left\{-\frac{1}{2}\sum_{i=1}^p\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2\right\} \\ &= \prod_{i=1}^p \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left\{-\frac{1}{2}\left(\frac{x_i-\mu_i}{\sigma_i}\right)^2\right\} \\ &= \prod_{i=1}^p f_i(x_i) \end{aligned}$$

Esto es, la *f. d. p.* conjunta es el producto de las *f. d. p.* marginales. Luego, las componentes son independientes.

Proposición : Sea $\boldsymbol{X} \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$. Entonces su función generadora de momentos (*f. g. m.*) conjunta está dada por

$$m_{\boldsymbol{X}}(\boldsymbol{t}) = \mathbb{E}[e^{\boldsymbol{t}^T\boldsymbol{X}}] = \exp\left\{\boldsymbol{t}^T\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}^T\Sigma\boldsymbol{t}\right\} = \exp\left\{\sum_{i=1}^p \mu_i t_i + \frac{1}{2}\sum_{i,j} t_i t_j \sigma_{ij}\right\}$$

Demostración: Vea Graybill (1961).

Proposición : Sea $\boldsymbol{X} \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$ y $A_{q\times p}$ con rango$(A) = q \leq p$. Entonces, $\boldsymbol{Y} = A\boldsymbol{X} \sim \mathrm{N}_q(A\boldsymbol{\mu}, A\Sigma A^T)$. Esto es, una transformación lineal de normales es normal.

Demostración: Sea $\boldsymbol{Y} = A\boldsymbol{X}$. Entonces, la *f. g. m.* del *v. a.* $\boldsymbol{Y}$ está dada por

$$\begin{aligned} m_{\boldsymbol{Y}}(\boldsymbol{t}) &= \mathbb{E}\left[e^{\boldsymbol{t}^T\boldsymbol{Y}}\right] = \mathbb{E}\left[e^{\boldsymbol{t}^T(A\boldsymbol{X})}\right] = \mathbb{E}\left[e^{(A^T\boldsymbol{t})^T\boldsymbol{X}}\right] \\ &= \exp\left\{(A^T\boldsymbol{t})^T\boldsymbol{\mu} + \frac{1}{2}(A^T\boldsymbol{t})^T\Sigma(A^T\boldsymbol{t})\right\} \\ &= \exp\left\{\boldsymbol{t}^T(A\boldsymbol{\mu}) + \frac{1}{2}\boldsymbol{t}^T(A\Sigma A^T)\boldsymbol{t}\right\} \end{aligned}$$

que corresponde a la *f. g. m.* de una normal multivariada de vector de medias $A\boldsymbol{\mu}$ y matriz de covarianzas $(A\Sigma A^T)$. Se sigue del teorema de unicidad que $A\boldsymbol{X} \sim \mathrm{N}_q(A\boldsymbol{\mu}, A\Sigma A^T)$.

Proposición : La combinación lineal de las componentes de una normal multivariada es normal.

Demostración: Sea $\boldsymbol{X} \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$ y $A = (a_1, \ldots, a_p)$. Se sigue del resultado anterior que

$$Y = A\boldsymbol{X} = \sum_{i=1}^{p} a_i X_i \sim \mathrm{N}\Big(\sum_{i=1}^{n} a_i\mu_i, \sum_{ij} a_i a_j \sigma_{ij}\Big)$$

Proposición : Si $\boldsymbol{X} \sim \mathrm{N}_p(\mu, \Sigma)$, cualquier reordenamiento de las componentes de $\boldsymbol{X}$ también es normal.

Demostración: Sea A una matriz de permutación del vector $\boldsymbol{X}$. Entonces A es una matriz con un 1 por fila y columna y los demás elementos 0. Luego $A\boldsymbol{X}$ es el vector $\boldsymbol{X}$ con sus componentes reordenados. La proposición se sigue aplicando resultado anterior.

Proposición : Marginales de normal es normal.

Esto es, si $\boldsymbol{X} \sim \mathrm{N}_p(\mu, \Sigma)$ y $\boldsymbol{X}$ se particiona como $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix}$, con el subvector $\boldsymbol{X}_i = (X_{i_1}, \ldots, X_{i_{p_i}})^T$, para $i = 1, 2$ y las correspondiente particiones $\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}$ y $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, entonces $\boldsymbol{X}_i \sim \mathrm{N}_{p_i}\big(\boldsymbol{\mu}_i, \Sigma_{ii}\big)$.

Demostración: Considere por ejemplo $\boldsymbol{X}_1$. Entonces la matriz $A = \big[I_{p_1} : 0\big]_{p_1 \times p}$ es tal que $\boldsymbol{X}_1 = A\boldsymbol{X}$ y $\boldsymbol{X}_1 \sim \mathrm{N}_{p_1}\big(\boldsymbol{\mu}_1, \Sigma_{11}\big)$, con $\boldsymbol{\mu}_1 = A\boldsymbol{\mu}$ y $\Sigma_{11} = A\Sigma A^T$.

Proposición : Sea $\boldsymbol{X} \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$ y X con la partición definida en el resultado anterior. Entonces, condicionalmente

$$(\boldsymbol{X}_1 | \boldsymbol{X}_2 = \boldsymbol{x}_2) \sim \mathrm{N}_{p_1}\left(\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2),\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$$

Demostración: Vea Graybill (1961).

Note que si $\boldsymbol{X}$ es un vector normal bivariado con $p_1 = 1 = p_2$, se sigue del resultado anterior que, por ejemplo,

$$(X_1 | X_2 = x_2) \sim \mathrm{N}\left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2), \sigma_1^2(1 - \rho^2)\right)$$

Proposición : Sea $\boldsymbol{X} \sim \mathrm{N}_p(\boldsymbol{\mu}, \Sigma)$ con (Q, Λ) la descomposición espectral de la matriz Σ, entonces $\boldsymbol{Y} = Q^T \boldsymbol{X}$ tiene componentes normales independientes. Aún más, $\mathrm{var}(Y_i) = \lambda_i$, el i-ésimo valor propio de Σ y $\sum \sigma_i^2 = \sum \lambda_i$.

Demostración: Sea (Q, Λ) la descomposición espectral de la matriz de covarianzas Σ. Entonces, la matriz Q es de columnas ortonormales correspondientes a los vectores propios de la matriz Σ y $\Lambda = \mathrm{diag}\{\lambda_1, \ldots, \lambda_p\}$, los correspondientes valores propios, con $\lambda_1 \geq \cdots \geq \lambda_p > 0$ y tales que $\Sigma = Q\Lambda Q^T$.

Defina $\boldsymbol{Y} = Q^T \boldsymbol{X}$. Se sigue del resultado (8) que $\boldsymbol{Y} \sim \mathrm{N}_p(Q^T \boldsymbol{\mu}, \Lambda)$, pues

$$\mathrm{cov}(\boldsymbol{Y}) = \mathrm{cov}(Q^T \boldsymbol{X}) = Q^T \Sigma Q = Q^T (Q\Lambda Q^T) Q = (Q^T Q)\Lambda(Q^T Q) = \Lambda$$

por la ortogonalidad de Q. Entonces, $\mathrm{var}(Y_i) = \lambda_i$, $i = 1, \ldots, p$ y la $\mathrm{cov}(Y_i, Y_j) = 0$ para $i \neq j$. Así $\boldsymbol{Y}$ es distribuida normalmente con componentes no correlacionadas, luego son independientes. Finalmente, se sigue de las propiedades del operador traza que

$$\mathrm{vartot}(\boldsymbol{X}) = \sum \sigma_i^2 = \mathrm{tr}(\Sigma) = \mathrm{tr}(Q\Lambda Q^T) = \mathrm{tr}(Q^T Q\Lambda) = \mathrm{tr}(\Lambda) = \sum \lambda_i = \mathrm{vartot}(\boldsymbol{Y})$$

La interpretación práctica del resultado anterior es que la transformación del vector aleatorio $\boldsymbol{X}$ por $\boldsymbol{Y} = Q^T \boldsymbol{X}$ produce componentes normales independientes que no comparten información entre ellas y todas juntas reconstruyen totalmente la información contenida en $\boldsymbol{X}$.

Aún más, puesto que $\mathrm{var}(Y_i) = \lambda_i$ y éstos vienen ordenados $\lambda_1 \geq \cdots \geq \lambda_p$, usualmente basta con los primeras componentes de Y para colectar la mayoría de la información contenida en $\boldsymbol{X}$ interpretada como sus varianzas. Vea (Schott 1997).

Ejemplo : Considere $\boldsymbol{X} \sim N_5(\boldsymbol{\mu}, \Sigma)$, con

$$\boldsymbol{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 2 \\ 0 \end{bmatrix} \quad \text{y} \quad \Sigma = \begin{bmatrix} 1 & 0 & -1 & 1 & 0 \\ 0 & 2 & -1 & 0 & 1 \\ -1 & -1 & 3 & -1 & 0 \\ 1 & 0 & -1 & 4 & 1 \\ 0 & 1 & 0 & 1 & 5 \end{bmatrix}$$

a). ¿Cómo se distribuye X_5?

b). Encuentre la distribución marginal de X_1 y la de X_4. ¿Cuál es su correlación?

c). Encuentre la distribución marginal del vector aleatorio $(X_3, X_2)^T$.

d). Muestre que X_2 y X_4 son *v. a.*'s independientes.

e). Encuentre la distribución condicional de X_4 dado $X_1 = -2$.

f). Calcule $\mathbb{P}(X_4 - X_2 > 0)$.

g). Calcule $\mathbb{P}(X_4 - X_2 > 0 | X_3 = 1, X_5 = -1)$.

h). Encuentre una matriz $A_{5\times 5}$ tal que el *v. a.* $\boldsymbol{Y} = A\boldsymbol{X}$ tenga componentes independientes. (Use alguna aplicación de computadora, como MATLAB, Python o R.)

i). Verifique que $\mathrm{vartot}(\boldsymbol{X}) = \mathrm{vartot}(\boldsymbol{Y})$.

j). Para cada una de las entradas del vector $\boldsymbol{Y}$ calcule su participación proporcional en la variación total, esto es $\delta_i = \mathrm{var}(Y_i)/\mathrm{vartot}(Y)$. Determine la proporción acumulada $\Delta_j = \delta_1 + \cdots + \delta_j$ hasta la componente j-ésima para $1 \leq j \leq 5$.

Solución:

a). De la quinta componente del vector de medias μ y la entrada $(5,5)$ de la matriz de covarianzas Σ, $X_5 \sim \mathrm{N}_1(0,5)$.

b). De manera similar al inciso anterior, $X_1 \sim \mathrm{N}_1(0,1)$ y $X_4 \sim \mathrm{N}_1(2,4)$. Ahora, la covarianza entre las componentes X_1 y X_4 se obtiene de la entrada $(1,4)$ o $(4,1)$ de la matriz Σ. Así, $\mathrm{cov}(X_1,X_4) = (\sigma_{14}) = (\Sigma)_{14} = 1$.

c). Elija las correspondientes componentes del vector de medias μ y la matriz de covarianzas Σ. Así,

$$\begin{bmatrix} X_3 \\ X_2 \end{bmatrix} \sim \mathrm{N}_2\left(\begin{bmatrix} -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 & -1 \\ -1 & 2 \end{bmatrix}\right)$$

Más formalmente, se sigue de la proposición 8 anterior definiendo $A = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$ y determinando la distribución de $A\boldsymbol{X} = \begin{bmatrix} X_3 \\ X_2 \end{bmatrix}$.

d). $\mathrm{cov}(X_2,X_4) = (\Sigma)_{24} = 0$ y la normalidad implica la independencia de X_2 y X_4.

e). Se sigue de la proposición 1 que

$$(X_i|X_j = x_j) \sim \ \mathrm{N}_1\left(\mu_i + \rho_{ij}\frac{\sigma_i}{\sigma_j}(x_j - \mu_j),\ \sigma_i^2(1-\rho_{ij}^2)\right)$$

Note que $\rho_{ij}\dfrac{\sigma_i}{\sigma_j} = \dfrac{\sigma_{ij}}{\sigma_i\sigma_j}\cdot\dfrac{\sigma_i}{\sigma_j} = \dfrac{\sigma_{ij}}{\sigma_j^2}$. Luego,

$$\rho_{41} = \frac{\sigma_{41}}{\sqrt{\sigma_4^2\cdot\sigma_1^2}} = \frac{1}{\sqrt{4\cdot 1}} = \frac{1}{2} \quad \text{y} \quad \rho_{41}\frac{\sigma_4}{\sigma_1^2} = \frac{\sigma_{41}}{\sigma_1^2} = \frac{1}{1} = 1$$

Entonces,

$$(X_4|X_1 = -2) \sim \ \mathrm{N}_1\left(2 + \frac{1}{1}(-2-0),\ 4\left(1-(1/2)^2\right)\right) \equiv \mathrm{N}_1(0,3)$$

f). $(X_4 - X_2) \sim \mathrm{N}_1\left(\mu_4 - \mu_2, \sigma_4^2 + \sigma_2^2 - 2\sigma_{42}\right) \equiv \mathrm{N}(1,6)$. Entonces,

$$\mathbb{P}(X_4 - X_2 > 0) = 1 - \mathbb{P}(X_4 - X_2 \le 0) = 1 - \Phi\left(\frac{0-1}{\sqrt{6}}\right) = \Phi(1/\sqrt{6}) \approx 0.6584$$

g). Defina

$$\boldsymbol{Y}_1 = \begin{bmatrix} X_4 \\ X_2 \end{bmatrix}, \boldsymbol{Y}_2 = \begin{bmatrix} X_3 \\ X_5 \end{bmatrix}, \quad \begin{bmatrix} \boldsymbol{Y}_1 \\ - \\ \boldsymbol{Y}_2 \end{bmatrix} \sim \mathrm{N}_4\left(\begin{bmatrix} 2 \\ 1 \\ - \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 & 0 & | & -1 & 1 \\ 0 & 2 & | & -1 & 1 \\ - & - & | & - & - \\ -1 & -1 & | & 3 & 0 \\ 1 & 1 & | & 0 & 5 \end{bmatrix}\right)$$

De la distribución condicional se sigue

$$(\boldsymbol{Y}_1|\boldsymbol{Y}_2 = \boldsymbol{y}_2) \sim \mathrm{N}_2\left(\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{y}_2 - \boldsymbol{\mu}_2),\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$$

con $\boldsymbol{y}_2 = (1,-1)'$ se tiene

$$\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{y}_2 - \boldsymbol{\mu}_2) = \begin{bmatrix} 2 \\ 1 \end{bmatrix} + \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 3 & 0 \\ 0 & 5 \end{bmatrix}^{-1}\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix} - \begin{bmatrix} -1 \\ 0 \end{bmatrix}\right) = \frac{1}{15}\begin{bmatrix} 17 \\ 2 \end{bmatrix}$$

$$\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix} - \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 5 \end{bmatrix}^{-1} \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix} = \frac{1}{15} \begin{bmatrix} 52 & -8 \\ -8 & 22 \end{bmatrix}$$

Luego,

$$\left(\begin{bmatrix} X_4 \\ X_2 \end{bmatrix} \Big| X_3 = 1, X_5 = -1 \right) \sim \mathrm{N}_2 \left(\frac{1}{15} \begin{bmatrix} 17 \\ 2 \end{bmatrix}, \frac{1}{15} \begin{bmatrix} 52 & -8 \\ -8 & 22 \end{bmatrix} \right)$$

y

$$\big(X_4 - X_2 | X_3 = 1, X_5 = -1\big) \sim \mathrm{N}_1 \left(\frac{17-2}{15}, \frac{52+22-2(-8)}{15} \right) \equiv \mathrm{N}_1 \left(\frac{15}{15}, \frac{90}{15} \right)$$

Finalmente,

$$\mathbb{P}\left(X_4 - X_2 > 0 | X_3 = 1, X_5 = -1\right) = 1 - \Phi\left(-\frac{1}{\sqrt{6}}\right) \approx 0.6584$$

Note que la diferencia de las probabilidades se debe a que el el subvector condicionante (X_3, X_5) covaría con el subvector (X_4, X_1) afectando su distribución.

h). Sea $\Sigma = Q\Lambda Q^T$ la descomposición espectral de Σ, la matriz de covarianzas del vector aleatorio X. Luego,

$$Q = \begin{bmatrix} -0.1691 & -0.2945 & 0.0000 & -0.3510 & 0.8726 \\ -0.2468 & 0.0534 & 0.5774 & 0.7302 & 0.2639 \\ 0.3106 & 0.5215 & -0.5774 & 0.3823 & 0.3900 \\ -0.5574 & -0.4680 & -0.5774 & 0.3479 & -0.1260 \\ -0.7095 & 0.6476 & -0.0000 & -0.2763 & -0.0300 \end{bmatrix}$$

y $\Lambda = \mathrm{diag}\,\{6.133, 4.360, 3.000, 1.098, 0.409\}$.

Sea ahora $A = Q'$ y se define la transformación $\boldsymbol{Y} = A\boldsymbol{X}$. El vector aleatorio tiene por ejemplo, la componente $Y_1 = -0.1691X_1 - 0.2468X_2 + 0.3106X_3 - 0.5574X_4 - 0.7095X_5$. Se sigue además de la proposición 13 que Y tiene matriz de covarianzas $\Lambda = \mathrm{diag}\,\{6.133, 4.360, 3.000, 1.098, 0.409\}$, diagonal, por lo que Y es de componentes independientes, con varianzas $\mathrm{var}(Y_1) = 6.133, \ldots, \mathrm{var}(Y_5) = 0.409$.

Más adelante se incluye código en R.

i). Se verifica que

$$\mathrm{vartot}(\boldsymbol{X}) = \sum_{i=1}^{5} \sigma_i^2 = 1 + \cdots + 5 = 15 = 6.133 + \cdots + 0.409 = \sum_{j=1}^{5} \lambda_j = \mathrm{vartot}(\boldsymbol{Y})$$

j). Finalmente, la siguiente tabla muestra las varianzas λ_j y proporciones p_j acumuladas de las componentes del vector aleatorio $\boldsymbol{Y}$. Note que las primeras 3 componentes

	Y_1	Y_2	Y_3	Y_4	Y_5
λ_j	6.133	4.360	3.000	1.098	0.409
$\sum \lambda_j$	6.133	10.493	13.493	14.591	15.000
$\sum p_j$	0.409	0.700	0.900	0.973	1.000

acumulan el 90 % de la variación total.

Código R

```
mu <- c(0,1,-1,2,0)          # Define el vector de medias
Sigma <- matrix(c(1,0,-1,1,0,  0,2,-1,0,1,  -1,-1,3,-1,0,
  1,0,-1,4,1, 0,1,0,1,5), nrow=5)    # Define la matriz de covarianzas
tt <- eigen(Sigma)           # Obtiene la descomposici{on espectral de Sigma
Q <- tt$vectors              # Matriz de vectores propios
lambda <- tt$values          # Valores propios
print(Q)
print(lambda)
print(diag(lambda))          # Construye una matriz diagonal a partir de un vector
print(sum(diag(Sigma)))      # Extrae la diagonal principal de una matriz y suma
print(cumsum(lambda))        # Suma y acumuala parcialmente
print(cumsum(lambda)/15)
```

6.3. Ejercicios

Refiérase al Cuaderno de Ejercicios sección 6, Barrios and Chambon (2024).

7. Transformaciones de variables y vectores aleatorios

7.1. Caso univariado

Sea Z la variable aleatoria distribuida normal estándar, $Z \sim \mathrm{N}(0,1)$. Z tiene a ϕ como función de densidad de probabilidad y Φ como función de probabilidad acumulada o función de distribución y donde, para todo $z \in \mathbb{R}$,

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad \text{y} \qquad \Phi(z) = \int_{-\infty}^{z} \phi(u)du$$

Considere ahora a $\mu \in \mathbb{R}$, $\sigma > 0$ y defina $X = \mu + \sigma Z$. Se desea encontrar la *f. d. p.* de la *v. a.* X.

En la sección 4 correspondiente a la función generadora de momentos se mostró que X así definida seguía una distribución normal con media μ y varianza σ^2. Este resultado se volverá a demostrar pero en esta ocasión trabajando directamente con la *f. p. a.* para aplicar después la derivada para obtener la *f. d. p.*. Esto es,

$$f_X(x) = \frac{d}{dx} F_X(x)$$

Así, sea $X = \mu + \sigma Z$. Luego su *f. p. a.* será

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leq x) = \mathbb{P}(\mu + \sigma Z \leq x) \\ &= \mathbb{P}\left(Z \leq \frac{x-\mu}{\sigma} \right) \\ &= \Phi\left(\frac{x-\mu}{\sigma} \right), \qquad \text{para todo } x \in \mathbb{R} \end{aligned}$$

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

$$\begin{aligned} f_X(x) &= \frac{d}{dx} F_X(x) \\ &= \frac{d}{dx}\left[\Phi\left(\frac{x-\mu}{\sigma} \right) \right] \\ &= \Phi'\left(\frac{x-\mu}{\sigma} \right) \frac{1}{\sigma} \\ &= \frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma} \right) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2}\left(\frac{x-\mu}{\sigma} \right)^2 \right\} \end{aligned}$$

que corresponde efectivamente a la *f. d. p.* de la distribución normal con media μ y varianza σ^2.

Ejemplo : Sea $U \sim \mathrm{unif}(0,1)$, $\lambda > 0$ y la *v. a.* Y definida por $Y = -\frac{1}{\lambda}\log(1-U)$. Encuentre la *f. d. p.* de Y.

Solución: Sea F_Y la *f. p. a.* de la *v. a.* Y. Note, que Y toma valores positivos. Sea $y > 0$,

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

que es la *f. p. a.* de una distribución exponencial con media $1/\lambda$.

Ejemplo : Sea X una *v. a.* con F_X y f_X su *f. p. a.* y *f. d. p.* respectivamente. Sean $a, b \in \mathbb{R}$, $b \neq 0$ y defina $Y = a + bX$. Encuentre f_Y, la *f. d. p.* de Y.

Solución: Sea F_Y la *f. p. a.* de Y. Entonces,

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(a+bX \leq y) = \mathbb{P}(bX \leq y-a) = (*)$$

i) Si $b > 0$,

$$(*) = \mathbb{P}\left(X \leq \frac{y-a}{b}\right) = F_X\left(\frac{y-a}{b}\right)$$

y consecuentemente

$$\begin{aligned}
f_Y(y) = \frac{d}{dy}F_Y(y) &= \frac{d}{dy}\left[F_X\left(\frac{y-a}{b}\right)\right] \\
&= F'_X\left(\frac{y-a}{b}\right)\frac{1}{b} \\
&= \frac{1}{b}f_X\left(\frac{y-a}{b}\right)
\end{aligned}$$

ii) Si $b < 0$,

$$(*) = \mathbb{P}\left(X \geq \frac{y-a}{b}\right) = 1 - F_X\left(\frac{y-a}{b}\right)$$

y consecuentemente

$$\begin{aligned}
f_Y(y) = \frac{d}{dy}F_Y(y) &= \frac{d}{dy}\left[1 - F_X\left(\frac{y-a}{b}\right)\right] \\
&= -F'_X\left(\frac{y-a}{b}\right)\frac{1}{b} \\
&= \frac{1}{|b|}f_X\left(\frac{y-a}{b}\right)
\end{aligned}$$

Así, ambos casos quedan considerados: si $Y = a + bX$,

$$f_Y(y) = \frac{1}{|b|}f_X\left(\frac{y-a}{b}\right)$$

Con la misma idea se tiene el siguiente teorema.

Teorema de Transformación (Mood et al. (1974))

Sea X una variable aleatoria continua con F_X y f_X sus funciones de distribución y de densidad de probabilidad, respectivamente y con soporte $\mathcal{S}_X$. Suponga además que

I). Sea $y = g(x)$ una función uno-a-uno de $\mathcal{S}_X$ a $\mathcal{S}_Y$.

II). $x = h(y) = g^{-1}(y)$ es derivable con respecto a y, continua y no cero para todo $y \in \mathcal{S}_Y$. La función h es la inversa de la función g.

Entonces, la *f. d. p.* de $Y = g(X)$ está dada por

$$f_Y(y) = f_X(h(y)) \left| \frac{d}{dy} h(y) \right| \mathbb{1}_{\mathcal{S}_Y}(y)$$

Demostración: Sea $Y = g(X)$ y $X = g^{-1}(Y) = h(Y)$.

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = (*)$$

i) Si g es monótona creciente

$$(*) = \mathbb{P}\left(X \leq h(y)\right) = F_X\left(h(y)\right)$$

y consecuentemente

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) = \frac{d}{dy}\left[F_X\left(h(y)\right)\right] \\ &= F_X'\left(h(y)\right) \frac{dh(y)}{dy} \\ &= f_X\left(h(y)\right) \frac{dh(y)}{dy} \end{aligned}$$

ii) Si g es monóntona decreciente h también lo es y consecuentemente $\frac{dh}{dy} < 0$. Así,

$$(*) = \mathbb{P}\left(X \geq h(y)\right) = 1 - F_X\left(h(y)\right)$$

y entonces

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) = \frac{d}{dy}\left[1 - F_X\left(h(y)\right)\right] \\ &= -F_X'\left(h(y)\right) \frac{dh(y)}{dy} \\ &= f_X\left(h(y)\right) \left|\frac{dh(y)}{dy}\right| \end{aligned}$$

Así, ambos casos quedan considerados si $Y = g(X)$,

$$f_Y(y) = f_X\left(h(y)\right) \left|\frac{dh(y)}{dy}\right|$$

Ejemplo : Sea $\Theta \sim \text{unif}\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ y $W = \tan(\Theta)$ Encuentre f_W, la *f. d. p.* de la *v. a.* W.

Solución: $S_\Theta = (-\frac{\pi}{2}, \frac{\pi}{2})$, entonces el soporte de $W = \tan(\Theta)$ es $\mathcal{S}_W = (-\infty, \infty)$.

$$w = g(\theta) = \tan(\theta), \quad \theta = h(w) = \arctan(w)$$

y $\left|\frac{dh(w)}{dw}\right| = \frac{1}{1+w^2}$.

Se sigue del teorema de transformación que para todo $w \in \mathbb{R}$,

$$\begin{aligned} f_W(w) &= f_\Theta\left(h(w)\right) \cdot \left|\frac{dh(w)}{dw}\right| \\ &= \frac{1}{\pi} \mathbb{1}_{(-\frac{\pi}{2}, \frac{\pi}{2})}(\arctan(w)) \cdot \frac{1}{1+w^2} \\ &= \frac{1}{\pi(1+w^2)} \end{aligned}$$

puesto que si $\Theta \sim \text{unif}\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, $f_\Theta(\theta) = \frac{1}{\pi}\mathbb{1}_{(-\frac{\pi}{2},\frac{\pi}{2})}(\theta)$.

Ejemplo : Sea X *v. a.* continua con *f. d. p.* f_X. Encuentre la *f. d. p.* f_Y de la *v. a.* $Y = X^2$.

Solución:

Note que $Y = g(X) = X^2$ no es una función uno-a-uno en todo los reales por lo que no es posible aplicar el teorema de transformación así como está enunciado pues asume que la función g es precisamente uno-a-uno. Sin embargo es posible resolver el problema como se hizo al inicio de la sección, construyendo la función de distribución y derivando para obtener la función de densidad.

Sea F_Y la *f. p. a.* de Y. Luego,

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \le y) = \mathbb{P}(X^2 \le y) = \mathbb{P}(-\sqrt{y} \le X \le \sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \end{aligned}$$

Por lo que

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}F_Y(y) = F_X'(\sqrt{y})\left(\frac{1}{2}y^{-1/2}\right) - F_X'(-\sqrt{y})\left(-\frac{1}{2}y^{-1/2}\right) \\ &= \frac{1}{2\sqrt{y}}\left[f_X(\sqrt{y}) + f_X(-\sqrt{y})\right] \end{aligned}$$

Por casos como el anterior se modifica el teorema de transformación.

Teorema de Transformación: Sea X una *v. a.* continua con *f. d. p.* f_X y con soporte $\mathcal{S}_X$. Sea $Y = g(X)$ tal que, el soporte de X se puede separar como $\mathcal{S}_X = I_1 \cup \cdots \cup I_m$, donde la función g es uno-a-uno de I_k a $\mathcal{S}_Y$ para $k = 1, \ldots, m$. Con las mismas condiciones del teorema enunciado anteriormente, se tiene

$$f_Y(y) = \sum_{k=1}^{m} f_X(h_k(y)) \left|\frac{dh_k(y)}{dy}\right| \mathbb{1}_{\mathcal{S}_Y}(y)$$

donde las h_k son las funciones inversas de g sobre I_k. Esto es, $h_k(y) = g^{-1}(y) \in I_k$, para $k = 1, \ldots, m$.

Ejemplo : Sea X *v. a.* que sigue la **distribución Laplaciana** de parámetros 0 y 1 con *f. d. p.* dada por

$$f_X(x) = \frac{1}{2}e^{-|x|}\mathbb{1}_{\mathbb{R}}(x)$$

Sea $Y = X^2$. Encontrar la *f. d. p.* de Y.

Solución: Sea $y = g(x) = x^2$; $x = h_1(y) = +\sqrt{y}$, $x = h_2(y) = -\sqrt{y}$; $|\frac{dh}{dy}| = \frac{1}{2}y^{-1/2}$.

Luego, para todo $y > 0$,

$$\begin{aligned} f_Y(y) &= f_X\left(h_1(y)\right)\left|\frac{dh_1(y)}{dy}\right| + f_X\left(h_2(y)\right)\left|\frac{dh_2(y)}{dy}\right| \\ &= \frac{1}{2}e^{-\sqrt{y}} \cdot \frac{1}{2\sqrt{y}} + \frac{1}{2}e^{-\sqrt{y}} \cdot \frac{1}{2\sqrt{y}} \\ &= \frac{1}{2\sqrt{y}}e^{-\sqrt{y}} \end{aligned}$$

7.2. Transformación integral de la probabilidad

El siguiente teorema es de gran importancia teórica y práctica, con aplicación en varios de los algoritmos de simulación fundamental en el cálculo estadístico.

Teorema de la transformación integral de la probabilidad:. Sea X una *v. a.* con *f. p. a.* F_X continua estrictamente creciente. Entonces $U = F_X(X)$ se distribuye uniformemente en el intervalo $(0,1)$. Viceversa, si $U \sim \text{unif}(0,1)$ y $Y = F_X^{-1}(U)$, entonces Y sigue la misma distribución de X.

Demostración:

i) Sea $U = F_X(X)$ con F_U su función de distribución. Entonces, los valores que toma U están en el intervalo $(0,1)$ pues es una probabilidad. Luego, el soporte de U es $\mathcal{S}_U = (0,1)$. Entonces, para $0 < u < 1$, se tiene

$$\begin{aligned} F_U(u) &= \mathbb{P}(U \le u) = \mathbb{P}\left(F_X(X) \le u\right) = \mathbb{P}\Big(X \le F_X^{-1}(u)\Big) \\ &= F_X\Big(F_X^{-1}(u))\Big) \\ &= u \end{aligned}$$

que corresponde a la función de distribución de la distribución uniforme en el intervalo $(0,1)$.

ii) Por otro lado, sean $U \sim \text{unif}(0,1)$ y $Y = F_X^{-1}(U)$. Se sigue,

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \le y) = \mathbb{P}\Big(F_X^{-1}(U) \le y\Big) = \mathbb{P}\left(U \le F_X(y)\right) \\ &= F_U\left(F_X(y))\right) \\ &= F_X(y) \end{aligned}$$

por lo que Y tiene a F_X como *f. p. a.*. Luego, Y y X siguen la misma distribución.

Ejemplo : Sea $\theta > 0$ y $X \sim \text{Exp}(\theta)$, tal que $\mathbb{E}[X] = \theta$. $F_X(x) = 1 - e^{-x/\theta} = u$. Entonces, de acuerdo al teorema anterior $F_X^{-1}(u) = -\theta \log(1-u)$ sería una realización del la distribución exponencial. Esto es, $Y = -\theta \log(1-U)$ sigue la distribución exponencial con valor esperado θ.

La figura 5 muestra los histogramas de 5000 realizaciones de la distribución uniforme y las correspondientes de la distribución exponencial de media $\theta = 3$ de acuerdo al resultado de ejemplo anterior.

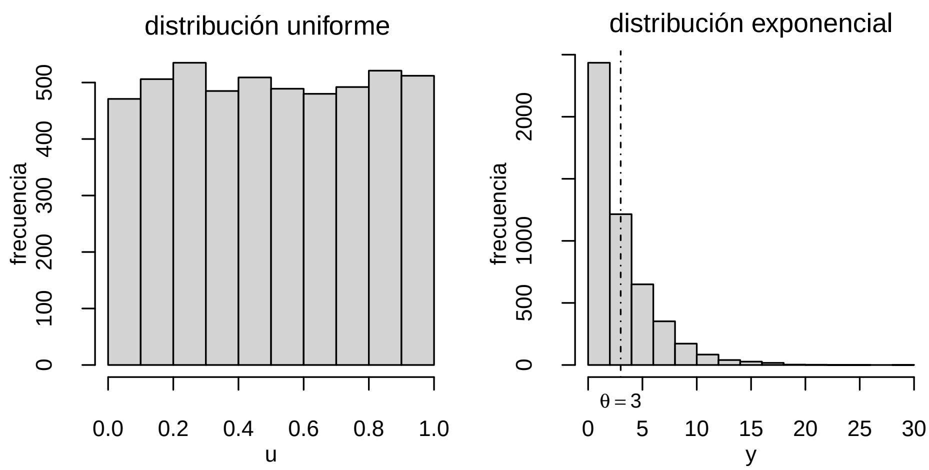


Figura 5: Histogramas de 5000 realizaciones de la distribución uniforme y las correspondientes de la distribución exponencial de media θ de acuerdo al teorema integral de la probabilidad.

7.3. Caso multivariado

Ejemplo : Sean Z_1 y Z_2 *v.a.i.i.d.* distribuidas normal estándar. Sean $X_1 = Z_1 + Z_2$ y $X_2 = Z_1 - Z_2$. Muestre que $\boldsymbol{X} = (X_1, X_2)^T$ tienen componentes independientes.

Solución:

a). Puesto que Z_1 y Z_2 son independientes, entonces $X_1 = Z_1 + Z_2 \sim \mathrm{N}(0,2)$ y $X_2 = Z_1 - Z_2 \sim \mathrm{N}(0,2)$. Además,

$$\begin{aligned}\mathrm{cov}(X_1, X_2) &= \mathrm{cov}(Z_1 + Z_2, Z_1 - Z_2)\\ &= \mathrm{cov}(Z_1, Z_1) - \mathrm{cov}(Z_1, Z_2) + \mathrm{cov}(Z_2, Z_1) - \mathrm{cov}(Z_2, Z_2)\\ &= 0\end{aligned}$$

y por normalidad se sigue la independencia.

b). De manera alternativa

$$\boldsymbol{Z} = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \sim \mathrm{N}_2(\boldsymbol{\mu}_Z, \Sigma_Z), \text{ donde } \boldsymbol{\mu}_Z = Z = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ y } \Sigma_Z = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

y considere ahora la matriz $A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$. Entonces, $\boldsymbol{X} = A\boldsymbol{Z}$ con

$$\boldsymbol{\mu}_X = A\boldsymbol{\mu}_Z = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ y } \Sigma_X = A\Sigma_Z A^T = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

Nuevamente, puesto que $\mathrm{cov}(X_1, X_2) = 0$ y por la normalidad se sigue la independencia.

Note que en el ejemplo anterior la normalidad de $\boldsymbol{X}$ se sigue de que sus componentes son normales no correlacionadas.

Teorema : Sobre transformaciones multivariadas. (Mood et al. (1974))

Sea $X = (X_1, \ldots, X_n)^T$ un vector aleatorio con función de densidad de probabilidad *conjunta* $f_X(x) = f_{X_1 \cdots X_n}(x_1, \ldots, x_n)$. Sea $\mathcal{S} = \{x : f_X(x) > 0\}$ el soporte de la distribución y suponga que $\mathcal{S}$ puede descomponerse en los subconjuntos $\mathcal{S}_1, \ldots, \mathcal{S}_m$, tales que, $y_1 = g_1(x), \ldots, y_n = g_n(x)$, forma una transformación *uno-a-uno* de $\mathcal{S}_k$ en $\mathcal{R} \subset \mathbb{R}^n$, $k = 1, \ldots, m$. Sean entonces $x_1 = h_{1k}(y), \ldots, x_n = h_{nk}(y)$, que denoten la transformación inversa de $\mathcal{R}$ en $\mathcal{S}_k$. Para $k = 1, \ldots, m$, denote el determinante jacobiano de la transformación h,

$$Jh_k = \begin{vmatrix} \frac{\partial h_{1k}(y)}{\partial y_1} & \frac{\partial h_{1k}(y)}{\partial y_2} & \cdots & \frac{\partial h_{1k}(y)}{\partial y_n} \\ \frac{\partial h_{2k}(y)}{\partial y_1} & \frac{\partial h_{2k}(y)}{\partial y_2} & \cdots & \frac{\partial h_{2k}(y)}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_{nk}(y)}{\partial y_1} & \frac{\partial h_{nk}(y)}{\partial y_2} & \cdots & \frac{\partial h_{nk}(y)}{\partial y_n} \end{vmatrix}$$

Suponga además que las derivadas parciales incluidas en Jh_k son continuas sobre $\mathcal{R}$ y que el determinante jacobiano $Jh_k \neq 0$, $k = 1 \ldots, m$. Entonces, la *f. d. p. conjunta* del *v. a.* $Y = (Y_1, \ldots, Y_n)^T$ está dado por

$$f_Y(y) = f_{Y_1 \cdots Y_n}(y_1, \ldots, y_n) = \sum_{k=1}^{m} f_X(h_k(y)) |Jh_k(y)|$$

para $y = (y_1, \ldots, y_n) \in \mathcal{R}$ y $h_k : \mathbb{R}^n \to \mathbb{R}^n$, con $h_k = (h_{1k}, \ldots, h_{nk})$ y $|Jh_k|$ denota el valor absoluto del determinante jacobiano.

Ejemplo : Sean Z_1 y Z_2 *v.a.i.i.d.* $\mathrm{N}(0,1)$. $X_1 = Z_1 + Z_2$ y $X_2 = Z_1 - Z_2$. Encuentre la *f. d. p.* conjunta de X.

Solución:

$$X_1 = g_1(Z) = Z_1 + Z_2, \qquad Z_1 = h_1(X) = \frac{X_1 + X_2}{2}$$

$$X_2 = g_2(Z) = Z_1 - Z_2, \qquad Z_2 = h_2(X) = \frac{X_1 - X_2}{2}$$

$$Jh(x) = \begin{vmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{vmatrix} = -1/4 - 1/4 = -1/2 \;\text{ y }\; |Jh(x)| = 1/2$$

Se sigue del teorema de transformación y por independencia de las Z_i's,

$$\begin{aligned} f_X(x) &= f_Z\left(h(x)\right) |Jh(x))| \\ &= \phi(h_1(x))\, \phi(h_2(x))\, |Jh(x)| \\ &= (2\pi)^{-1} \exp\left\{ -\frac{1}{2} \left[\left(\frac{x_1 + x_2}{2} \right)^2 + \left(\frac{x_1 - x_2}{2} \right)^2 \right] \right\} \frac{1}{2} \\ &= \frac{1}{(2\pi)2} \exp\left\{ -\frac{1}{2} \left[\left(\frac{x_1}{\sqrt{2}} \right)^2 + \left(\frac{x_2}{\sqrt{2}} \right)^2 \right] \right\} \\ &= \frac{1}{\sqrt{2\pi}\sqrt{2}} \exp\left\{ -\frac{1}{2} \left(\frac{x_1}{\sqrt{2}} \right)^2 \right\} \cdot \frac{1}{\sqrt{2\pi}\sqrt{2}} \exp\left\{ -\frac{1}{2} \left(\frac{x_2}{\sqrt{2}} \right)^2 \right\} \\ &= f(x_1) \cdot f(x_2) \end{aligned}$$

con f la *f. d. p.* de una distribución $\mathrm{N}(0,2)$. Finalmente, puesto que $f_X(x) = f(x_1) \cdot f_(x_2)$, las *v. a.*'s X_1 y X_2 son independientes idénticamente distribuidas.

Ejemplo : Sea $X \sim \mathrm{N}_n(\mu, \Sigma)$ y $A_{n\times n}$ matriz no singular. Entonces, $Y = AX \sim \mathrm{N}_n(A\mu, A\Sigma A^T)$

Solución: Sean $Y = AX = (Y_1, \ldots, Y_n)^T$ y $B = A^{-1}$.

Se tiene que $Y_i = g_i(X) = \sum_{k=1}^n a_{ik} X_i$ y $X_j = h_j(Y) = \sum_{k=1}^n b_{jk} Y_k$, por lo que

$$Jh(y) = \left| \left(\frac{\partial h_i(y)}{\partial y_j} \right) \right| = |\,(b_{ij})\,| = |B| = |A^{-1}| = \frac{1}{|A|}$$

Se sigue del teorema de transformación

$$\begin{aligned}
f_Y(y) &= f_X(h(y))\,|Jh(y)| \\
&= (2\pi)^{-n/2}|\Sigma|^{-1/2} \exp\left\{ -\frac{1}{2} \left(h(y)-\mu\right)^T \Sigma^{-1} \left(h(y)-\mu\right) \right\} ||B|| \\
&= (2\pi)^{-n/2}|\Sigma|^{-1/2} \exp\left\{ -\frac{1}{2} \left(A^{-1}y-\mu\right)^T \Sigma^{-1} \left(A^{-1}y-\mu\right) \right\} ||A^{-1}|| \\
&= (2\pi)^{-n/2}|A\Sigma A^T|^{-1/2} \exp\left\{ -\frac{1}{2} \left[A^{-1}(y-A\mu)\right]^T \Sigma^{-1} \left[A^{-1}(y-A\mu)\right] \right\} \\
&= (2\pi)^{-n/2}|A\Sigma A^T|^{-1/2} \exp\left\{ -\frac{1}{2} (y-A\mu)^T \left(A\Sigma A^T\right)^{-1} (y-A\mu) \right\}
\end{aligned}$$

Y por lo tanto $Y \sim \mathrm{N}(A\mu, A\Sigma A^T)$.

Ejemplo : Sea U_i *v.a.i.i.d.* uniformemente en el intervalo $(0,1)$. Sean $X_1 = U_1 + U_2$ y $X_2 = U_1 - U_2$. Encuentre la *f. d. p.* conjunta de $X = (X_1, X_2)^T$.

Solución: Se tiene del ejercicio anterior que

$$\begin{aligned}
X_1 &= g_1(U) = U_1 + U_2, \qquad & U_1 &= h_1(X) = \frac{X_1 + X_2}{2} \\
X_2 &= g_2(U) = U_1 - U_2, \qquad & U_2 &= h_2(X) = \frac{X_1 - X_2}{2}
\end{aligned}$$

y $|Jh(x)| = 1/2$.

Note que el soporte de $U = (U_1, U_2)^T$ es $\mathcal{S}_U = \big\{(u_1, u_2) : 0 \le u_i \le 1\big\}$, mientras que marginalmente se tiene $\mathcal{S}_1 = \{x : 0 \le x \le 2\}$ y $\mathcal{S}_2 = \{x : -1 \le x \le 1\}$, para X_1 y X_2, respectivamente. Luego, se sigue del teorema de transformación y por independencia de las U_i's,

$$\begin{aligned}
f_X(x) &= f_U\left(h(x)\right) |Jh(x))| \\
&= \mathbb{1}_{(0,1)}(h_1(x))\ \mathbb{1}_{(0,1)}(h_2(x))\, \frac{1}{2} \\
&= \frac{1}{2} \mathbb{1}_{\mathcal{S}_X}(x)
\end{aligned}$$

donde $\mathcal{S}_X$ es el soporte del vector aleatorio X dado por

$$\mathcal{S}_X = \left\{ (x_1, x_2) : 0 \le \frac{x_1 + x_2}{2} \le 1 \ \text{ y } \ 0 \le \frac{x_1 - x_2}{2} \le 1 \right\}$$

que da lugar a las sigueintes desigualdades

$$\begin{aligned}
&i) \quad 0 \le x_1 + x_2 \le 2 \quad \Longrightarrow \quad x_2 \le 2 - x_1 \ \text{ y } \ x_2 \ge -x_1 \\
&ii) \quad 0 \le x_1 - x_2 \le 2 \quad \Longrightarrow \quad x_2 \ge x_1 - 2 \ \text{ y } \ x_2 \le x_1
\end{aligned}$$

y que se muestra en la figura 6

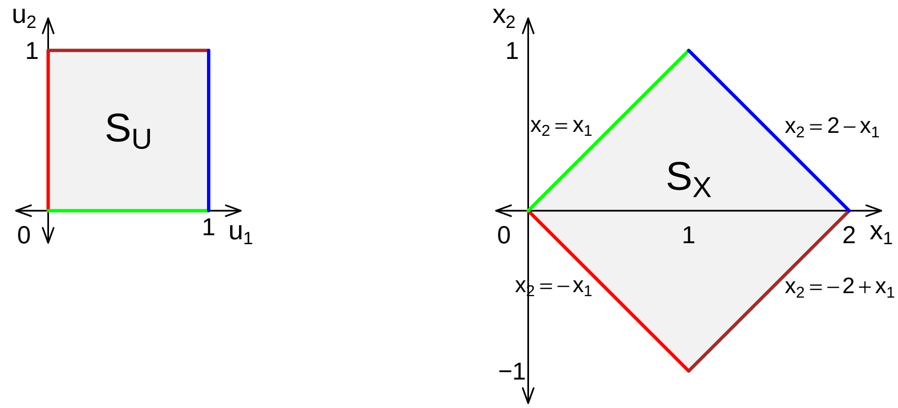


Figura 6: Soportes de las funciones de densidad conjunta de los vectores aleatorios U ($\mathcal{S}_U$) y $X = g(U)$ ($\mathcal{S}_X$) distribuidos uniformente. Se muestran las igualdades que definen los lados del rombo $\mathcal{S}_X$

Ejemplo : Sea $X = (X_1, X_2)$ un vector aleatorio con función de densidad de probabilidad (*f. d. p.*) conjunta dada por

$$f_{\boldsymbol{X}}(x_1, x_2) = \frac{1}{x_1^2 x_2^2} \mathbb{1}_{(1,\infty)}(x_1)\mathbb{1}_{(1,\infty)}(x_2)$$

Encuentre la *f. d. p.* de X_1/X_2.

a). Verifique que $f_{\boldsymbol{X}}$ es una *f. d. p.* propia.

b). Encuentre la *f. d. p.* de $Y = X_1/X_2$.

Solución:

a).

$$\begin{aligned}\int_{\mathbb{R}^2} f_{\boldsymbol{X}}(\boldsymbol{x}))d\boldsymbol{x} &= \iint_{\mathbb{R}^2} \frac{1}{x_1^2 x_2^2} \mathbb{1}_{(1,\infty)}(x_1)\mathbb{1}_{(1,\infty)}(x_2) dx_1\, dx_2 \\ &= \int_1^\infty \frac{1}{x_1^2} dx_1 \cdot \int_1^\infty \frac{1}{x_2^2} dx_2 \\ &= \left(\left[-x^{-1} \right]_1^\infty \right)^2 \\ &= 1\end{aligned}$$

Por lo que efectivamente, $f_{\boldsymbol{X}}$ es una *f. d. p.* propia.

b). Encuentre la *f. d. p.* de $Y = X_1/X_2$. Para poder aplicar el teorema de transformación es necesario tener un vector, por ejemplo $\boldsymbol{Y}$, de la misma dimensión que el vector $\boldsymbol{X}$ que mapee de manera biyectiva el soporte $S_{\boldsymbol{X}}$ al soporte $S_{\boldsymbol{Y}}$. Sean entonces $Y_1 = X_1/X_2$ y $Y_2 = X_2$. Luego,

$$\begin{array}{ccccc} y_1 & = & g_1(\boldsymbol{x}) & = & x_1/x_2 \\ y_2 & = & g_2(\boldsymbol{x}) & = & x_2 \end{array} \qquad \begin{array}{ccccc} x_1 & = & h_1(\boldsymbol{y}) & = & y_1 y_2 \\ x_2 & = & h_2(\boldsymbol{y}) & = & y_2 \end{array}$$

de donde el jacobiano de la transformación $\boldsymbol{h} = (h_1, h_2)$ está dado por

$$|J\boldsymbol{h}(\boldsymbol{y})| = \begin{vmatrix} \frac{\partial h_1(\boldsymbol{y})}{\partial y_1} & \frac{\partial h_1(\boldsymbol{y})}{\partial y_2} \\ \frac{\partial h_2(\boldsymbol{y})}{\partial y_1} & \frac{\partial h_2(\boldsymbol{y})}{\partial y_2} \end{vmatrix} = \begin{vmatrix} y_2 & y_1 \\ 0 & 1 \end{vmatrix} = y_2$$

Observe que las componentes X_i son independientes, por lo que el soporte del vector aleatorios $\boldsymbol{X}$ es $S_{\boldsymbol{X}} = (1, \infty) \times (1, \infty)$. Por otro lado, *marginalmente* $S_{Y_1} = (0, \infty)$, pues con fijo x_1, conforme $x_2 \nearrow \infty$, $y_1 = x_1/x_2 \searrow 0$, mientras que $S_{Y_2} = (1, \infty)$. Sin embargo, se debe cumplir que $x_1 = y_1 y_2 > 1$ lo que implica que $y_1 > 1/y_2$,. Por lo tanto, el soporte *conjunto* es $S_{\boldsymbol{Y}} = \{(y_1, y_2) : y_2 > 1, y_1 > 1/y_2\}$. La figura 7 muestra ambos soportes. En el panel de la derecha se han invertido los ejes para hacer más clara la marginalización de Y_1.

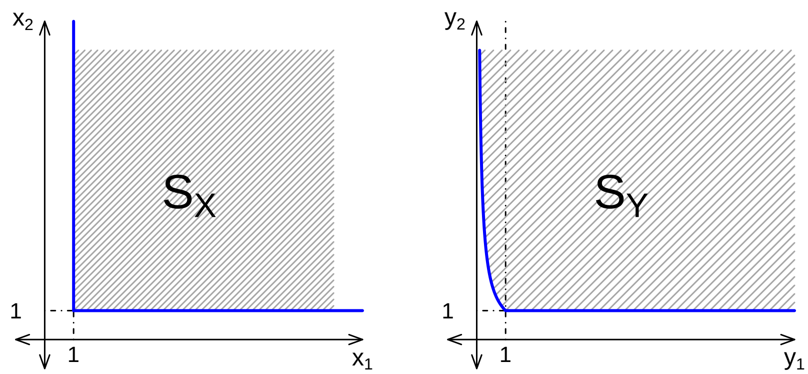


Figura 7: Soportes conjuntos de los vectores aleatorios $\boldsymbol{X}$ y $\boldsymbol{Y}$. Note que se han invertido los ejes del vector $\boldsymbol{Y}$.

Se sigue entonces, del teorema de transformación que

$$\begin{aligned} f_{\boldsymbol{Y}}(\boldsymbol{y}) &= f_{\boldsymbol{X}}\left(\boldsymbol{h}(\boldsymbol{y})\right) \, ||J\boldsymbol{h}(\boldsymbol{y})|| \\ &= \frac{1}{h_1^2(\boldsymbol{y})\, h_2^2(\boldsymbol{y})} \, ||J\boldsymbol{h}(\boldsymbol{y})|| \\ &= \frac{1}{y_1^2 y_2^2 \, y_2^2} \cdot y_2 \\ &= \frac{1}{y_1^2 y_2^3} \mathbb{1}_{\{y_2>1,\ y_1>1/y_2\}}(y_1, y_2) \end{aligned}$$

Del soporte se desprende que las componentes no son independientes.

Como ejercicio verifique que la función de densidad anterior es propia, integrando con respecto a y_1 y y_2 y el orden contrario. Esto es, muestre que

$$1 = \int_{\mathbb{R}^2} f_{\boldsymbol{Y}}(u, v) du dv = \int_{\mathbb{R}^2} f_{\boldsymbol{Y}}(u, v) dv du$$

Finalmente, la densidad de $Y = X_1/X_2$, se sigue de $f_1(y) = \int_{\mathbb{R}} f_{\boldsymbol{Y}}(y_1, y_2) dy_2$. Entonces, para $y_1 > 0$, fijo

$$f_1(y_1) = \int_{\mathbb{R}} \frac{1}{y_1^2 y_2^3} \mathbb{1}_{\{y_2>1,\ y_1>1/y_2\}}(y_1, y_2) dy_2 = \frac{1}{y_1^2} \int_{\max\{1/y_1, 1\}} \frac{1}{y_2^3} dy_2$$

Alternativamente, de la figura 7 se distinguen los siguientes casos:

a) $0 < y_1 \leq 1$,

$$f_1(y_1) = \frac{1}{y_1^2} \int_{1/y_1}^{\infty} \frac{1}{v^3} dv = \frac{1}{2}$$

b) $y_1 > 1$,

$$f_1(y_1) = \frac{1}{y_1^2} \int_{1}^{\infty} \frac{1}{v^3} dv = \frac{1}{2y_1^2}$$

Por lo que la marginal de $Y = X_1/X_2$ está dada por

$$f_Y(y) = \frac{1}{2} \mathbb{1}_{(0,1]}(y) + \frac{1}{2y^2} \mathbb{1}_{(1,\infty)}(y)$$

Proposición : Sean $Z \sim \mathrm{N}(0,1)$ y $Y \sim \chi_n^2$ *v. a.*'s independientes. Entonces la *v. a.* $T = Z/\sqrt{Y/n}$ tiene una función de densidad de probabilidad dada por

$$f_T(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} (1 + t^2/n)^{-(n+1)/2} \, \mathbb{1}_{\mathbb{R}}(t) \tag{6}$$

Demostración: Sea $T = Z/\sqrt{Y/n}$ y defina $U = Y$. Entonces, se tienen las siguientes igualdades

$$\begin{array}{rclcl} T &=& g_1(Z,Y) &=& Z/\sqrt{Y/n} \\ U &=& g_2(Z,Y) &=& Y \end{array} \qquad \text{y} \qquad \begin{array}{rclcl} Z &=& h_1(T,U) &=& T\sqrt{U/n} \\ Y &=& h_2(T,U) &=& U \end{array}$$

y de donde el *jacobiano* de la transformación $h = (h_1, h_2)$ está dado por

$$|Jh(t,u)| = \begin{vmatrix} \frac{\partial h_1(t,u)}{\partial t} & \frac{\partial h_1(t,u)}{\partial u} \\ \frac{\partial h_2(t,u)}{\partial t} & \frac{\partial h_2(t,u)}{\partial u} \end{vmatrix} = \begin{vmatrix} \sqrt{u/n} & \frac{1}{2} t u^{-1/2}/\sqrt{n} \\ 0 & 1 \end{vmatrix} = \sqrt{u/n}$$

Por la independencia de Z y Y se tiene que $f_{ZY}(z,y) = f_Z(z) \cdot f_Y(y)$. Recuerde además que $\chi_n^2 \sim \mathrm{Gamma}(n/2, 2)$. Luego, del teorema cambio de variable se sigue que para todo $t \in \mathbb{R},\ u \in \mathbb{R}^+$,

$$\begin{aligned} f_{TU}(t,u) &= f_{ZY}(h_1(t,u), h_2(t,u)) \cdot |Jh(t,u)| \\ &= f_Z(t\sqrt{u/n}) \cdot f_Y(u) \cdot \sqrt{u/n} \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2 u/n} \cdot \frac{u^{n/2-1} e^{-u/2}}{\Gamma(n/2) 2^{n/2}} \cdot \sqrt{u/n} \\ &= \frac{u^{\frac{n-1}{2}} e^{-\frac{1}{2}(t^2/n+1)u}}{2^{\frac{n+1}{2}} \sqrt{n\pi}\, \Gamma(n/2)} \end{aligned}$$

Por lo tanto, para todo $t \in \mathbb{R}$, la densidad marginal de T está dada por

$$
\begin{aligned}
f_T(t) &= \int_0^\infty f_{TU}(t,u)du \\
&= \int_0^\infty \frac{\left(\frac{u}{2}\right)^{\frac{n-1}{2}} e^{-\frac{1}{2}(t^2/n+1)u}}{2\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} du \\
&= \int_0^\infty \frac{v^{\frac{n-1}{2}} e^{-\lambda v}}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} dv & \text{con } \lambda = \frac{t^2}{n}+1 \text{ y } v=\frac{u}{2} \\
&= \frac{1}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \frac{1}{K} \int_0^\infty K v^{\alpha-1} e^{-\lambda v} dv & \alpha = \frac{n+1}{2} \text{ y } K \text{ la constante} \\
&= \frac{1}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\lambda^{\left(\frac{n+1}{2}\right)}} \int_0^\infty f_W(w;\alpha,\beta)dw & \text{donde } W \sim \text{Gamma}\left(\frac{n+1}{2}, \frac{1}{\lambda}\right) \\
&= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \left(t^2/n+1\right)^{-\frac{n+1}{2}}
\end{aligned}
$$

Ejercicio : Sean Z_1 y Z_2 *v.a.i.i.d.* normal estándar. Sean $Y_1 = Z_1 + Z_2$, y $Y_2 = Z_1/Z_2$. Muestre entonces que marginalmente $Y_1 \sim \text{N}(0,2)$ y $Y_2 \sim \text{Cauchy}(0,1)$. [*Sugerencia*: use el cambio de variable $u = \frac{1}{2}\frac{(1+y_2^2)}{(1+y_2)^2} y_1^2$.]

7.4. La distribución t de Student.

Definición : Sea T variable aleatoria con función de densidad dada por la expresión (6). Entonces T se dice que sigue una **distribución t de *Student* con n grados de libertad** y se denota $T \sim t_n$.

Propiedades.

1. Sean $Z \sim \text{N}(0,1)$ y $Y \sim \chi^2_n$ *v. a.*'s independientes. Entonces la variable aleatoria
$$T = \frac{Z}{\sqrt{Y/n}} \sim t_n$$
2. La distribución de t_n es simétrica alrededor de 0 como se muestra en la Figura 8 para varios valores de n.
3. Para $n > 2$,
$$E(T) = 0, \qquad\qquad V(T) = \frac{n}{n-2}$$
4. La distribución de Cauchy(0,1) es el caso particular de la distribución t de *Student* con 1 grado de libertad.
5. Si $T_n \sim t_n$, para n *grande* (≥ 40) la distribución T_n se aproxima razonablemente a la distribución normal estándar. De hecho, T_n *converge en distribución* a la distribución normal estándar cuando $n \to \infty$.
$$T_n \xrightarrow{D} Z$$
Esto es, para todo $t \in \mathbb{R}$,
$$\lim_{n\longrightarrow\infty} \mathbb{P}(T_n \leq t) = \Phi(t)$$
donde Φ denota la función de probabilidad acumulada de una normal estándar.

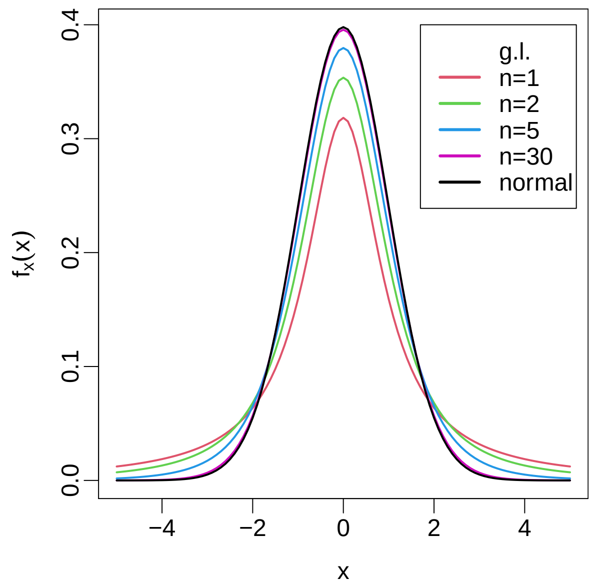


Figura 8: Función de densidad de la distribución t-Student para varios grados de libertad (gl).

7.5. Transformación Box-Muller.

Proposición: Muestre que si U_1 y U_2 son *v.a.i.i.d.* uniformemente en (0, 1) y si se definen

$$\begin{aligned} X_1 &= (-2\ln U_1)^{1/2}\cos 2\pi\, U_2 \\ X_2 &= (-2\ln U_1)^{1/2}\sin 2\pi\, U_2 \end{aligned} \tag{7}$$

entonces X_1 y X_2 son *v.a.i.i.d.* normal estándar.

Demostración Sean U_1 y U_2 *v.a.i.i.d.* uniformemente en el intervalo $(0,1)$. Se mostrará que

$$X_1 = (-2\ln U_1)^{1/2}\cos 2\pi U_2 \quad y \quad X_2 = (-2\ln U_1)^{1/2}\sin 2\pi U_2$$

son a su vez *v.a.i.i.d.* normal estándar. Con este fin se aplicará dos veces el teorema sobre transformaciones antes presentado. Para esto considere las *v. a.*'s Y_i definidas por:

$$\begin{aligned} Y_1 &= g_1(U) = (-2\ln U_1)^{1/2} \\ Y_2 &= g_2(U) = 2\pi U_2 \end{aligned} \qquad \text{y sus funciones inversas} \qquad \begin{aligned} U_1 &= h_1(Y) = e^{-Y_1^2/2} \\ U_2 &= h_2(Y) = Y_2/2\pi \end{aligned}$$

Luego el jacobiano de la transformación $h = (h_1, h_2)$ es

$$Jh(y) = \left|\frac{\partial h_i}{\partial y_j}\right| = \begin{vmatrix} -y_1 e^{-y_1^2/2} & 0 \\ 0 & 1/2\pi \end{vmatrix} = -\frac{y_1}{2\pi}e^{-y_1^2/2}$$

Por el teorema sobre transformaciones, se sigue de la independencia de U_1 y U_2 que $f_U(h) = f_{U_1}(h_1)f_{U_2}(h_2)$

$$f_Y(y) = f_{U_1}\left(h_1(y)\right) f_{U_2}(h_2(y))\,|Jh(y)| = \left[\frac{y_1}{2\pi}e^{-y_1^2/2}I_{[0,1]}(h_1(y))\right]\cdot\left[I_{[0,1]}(h_2(y))\right] \tag{8}$$

pues si $U_i \sim \text{unif}(0,1)$, entonces $f_{U_i}(u) = \mathbb{1}_{[0,1]}(u)$, y donde $\mathbb{1}_{[0,1]}$ es la función indicadora del intervalo $[0,1]$. Note en la expresión anterior que la función de densidad conjunta de Y se puede descomponer como el producto de funciones que dependen exclusivamente de y_1 y

y_2 respectivamente por lo que se sigue que las *v. a.*'s Y_1 y Y_2 son independientes con *f. d. p.* conjunta dada por (8).

Ahora bien, consideremos las variables X's definidas por (7). Entonces,

$X_1 = g_1(Y) = Y_1 \cos Y_2$ $X_2 = g_2(Y) = Y_1 \sin Y_2$	y sus funciones inversas	$Y_1 = h_1(X) = (X_1^2 + X_2^2)^{1/2}$ $Y_2 = h_2(X) = \arctan(X_2/X_1)$

En este caso el jacobiano de la transformación es

$$Jh(x) = \left|\frac{\partial h_i}{\partial x_j}\right| = \begin{vmatrix} x_1(x_1^2+x_2^2)^{-1/2} & x_2(x_1^2+x_2^2)^{-1/2} \\ -x_2(x_1^2+x_2^2)^{-1} & x_1(x_1^2+x_2^2)^{-1} \end{vmatrix} = (x_1^2+x_2^2)^{-1/2}$$

Luego, aplicando nuevamente el teorema de transformación, la *f. d. p.* conjunta de las *v. a.*'s X queda

$$\begin{aligned} f_X(x) &= f_Y(h_1, h_2)|Jh(x)| \\ &= \frac{(x_1^2+x_2^2)^{1/2}}{2\pi}\, e^{-\frac{1}{2}(x_1^2+x_2^2)} \cdot (x_1^2+x_2^2)^{-1/2} \\ &= \frac{1}{\sqrt{2\pi}}\, e^{-x_1^2/2} \cdot \frac{1}{\sqrt{2\pi}}\, e^{-x_2^2/2} \\ &= \phi(x_1) \cdot \phi(x_2) \end{aligned}$$

donde ϕ denota la *f. d. p.* de la distribución normal estándar. Por lo tanto, X_1 y X_2 son *v.a.i.i.d.* con $X_i \sim \mathrm{N}(0,1)$.

La transformación anterior se debe a Box and Muller (1958). Por un tiempo ésta era la manera de generar números seudoaleatorios normalmente distribuidos. Ahora se emplean algoritmos más eficientes.

7.6. Ejercicios

Refiérase al Cuaderno de Ejercicios sección 7, Barrios and Chambon (2024).

8. Suma y cociente de variables aleatorias

Variables aleatorias discretas

Sean X y Y variables aleatorias discretas con función masa de probabilidad conjunta f y marginales f_X y f_Y y soportes $\mathcal{S}_{\mathrm{X}}$ y $\mathcal{S}_{\mathrm{Y}}$, respectivamente. Entonces, se tiene para todo $z \in \mathbb{R}$,

$$\{X + Y = z\} = \bigcup_{x_i \in \mathcal{S}_{\mathrm{X}}} \{X = x_i, Y = z - x_i\}$$

que son eventos disjuntos. Por lo que se sigue (axioma de probabilidad) que

$$\begin{aligned}
\mathbb{P}(X + Y = z) &= \sum_{x_i \in \mathcal{S}_{\mathrm{X}}} \mathbb{P}(X = x_i, Y = z - x_i) \\
&= \sum_{x_i \in \mathcal{S}_{\mathrm{X}}} f(x_i, z - x_i)
\end{aligned}$$

Corolario :

1. Si X y Y son *v. a.*'s independientes, entonces

$$f_{X+Y}(z) = \sum_{x_i \in \mathcal{S}_{\mathrm{X}}} f_X(x_i) f_Y(z - x_i) = \sum_{y_j \in \mathcal{S}_{\mathrm{Y}}} f_X(z - y_j) f_Y(y_j)$$

2. Si X y Y son *v. a.*'s enteras independientes, entonces

$$f_{X+Y}(z) = \sum_{x \in \mathbb{Z}} f_X(x) f_Y(z - x)$$

3. Si X y Y son *v. a.*'s independientes enteras no negativas, entonces

$$f_{X+Y}(z) = \sum_{x=0}^{z} f_X(x) f_Y(z - x)$$

Variables aleatorias continuas

Suponga ahora X y Y *v. a.*'s continuas con *f. d. p.* conjunta f. Sea $Z = \varphi(X, Y)$, donde φ es una función medible sobre $(\Omega, \mathcal{S})$ de manera que Z es una variable aleatoria.

La siguiente idea de como construir funciones de densidad fue tomada del texto de Hoel, Port, and Stone (1971).

Para $z \in \mathbb{R}$, $\{Z \le z\} = \big\{(X, Y) \in A_z\big\}$, donde $A_z = \big\{(x, y) : \varphi(x, y) \le z\big\}$. Luego, si F_Z es la función de probabilidad acumulada de X,

$$F_Z(z) = \mathbb{P}(Z \le z) = \mathbb{P}((X, Y) \in A_z) = \iint_{A_z} f(x, y) dx dy$$

Si existe una función h no negativa tal que para todo $z \in \mathbb{R}$,

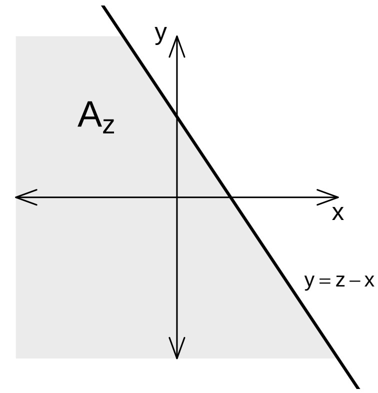
$$\iint_{A_z} f(x, y) dx dy = \int_{-\infty}^{z} h(u) du$$

entones h es una función de densidad para la distribución de probabilidad de $Z = \varphi(X, Y)$.

8.1. Suma de variables aleatorias

Sea $Z = \varphi(X,Y) = X + Y$. Entonces, $A_z = \{(x,y) : x + y \leq z\}$ representa el plano inferior determinado por la recta $x + y = z$. Así,

$$
\begin{aligned}
F_Z(z) &= \iint_{A_z} f(x,y)dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f(x,y)dydx \\
&= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{z} f(x,v-x)dv\right]dx \\
&= \int_{-\infty}^{z}\left[\int_{-\infty}^{\infty} f(x,v-x)dx\right]dv \\
&= \int_{-\infty}^{z} h(v)dv
\end{aligned}
$$



con $h(v) \geq 0$ para todo $v \in \mathbb{R}$, puesto que es la integral de una función no negativa ($f \geq 0$). El cambio del orden de integración se permite por el teorema de Fubini. Luego, h es necesariamente una función de densidad de probabilidad (con respecto a la integral) de la variable aleatoria $Z = \varphi(X,Y)$.

Por lo tanto, la función de densidad de la suma $X + Y$ está dada por

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f(x, z-x)dx = \int_{-\infty}^{\infty} f(z-y, y)dy$$

Corolario :

1. Sean X y Y *v. a.*'s independientes con f_X y f_Y las correspondientes *f. d. p.*. Entonces, la *f. d. p.* de la suma $X + Y$ es

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx = \int_{-\infty}^{\infty} f_X(z-y)f_Y(y)dy$$

2. Sean X y Y *v. a.*'s independientes no negativas con f_X y f_Y las correspondientes *f. d. p.*. Entonces, la *f. d. p.* de la suma $X + Y$ es

$$f_{X+Y}(z) = \int_{0}^{z} f_X(x)f_Y(z-x)dx = \int_{0}^{z} f_X(z-y)f_Y(y)dy$$

Ejemplo : Sean X y Y *v.a.i.i.d.* uniformemente en el intervalo $(0,1)$. Sea $Z = X + Y$. Determine f_{X+Y}, su *f. d. p.*.

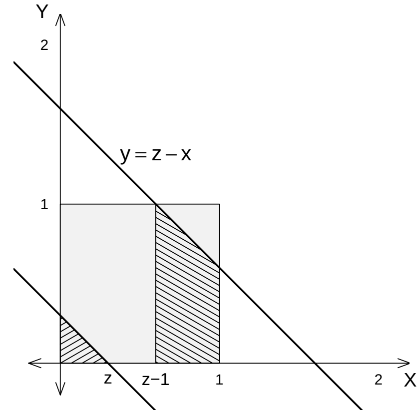
Solución: Sean X y Y *v. a.*'s independientes no negativas y denote por $f(w) = \mathbb{1}_{(0,1)}(w)$ la *f. d. p.* común de X y Y. Entonces,

$$f_Z(z) = \int_0^z f(x)f(z-x)dx = \int_0^z \mathbb{1}_{(0,1)}(x)\mathbb{1}_{(0,1)}(z-x)dx$$

I. Si $z < 0$, entonces $f_Z(z) = 0$.

II. Sea $0 \leq z < 1$.

$$\begin{aligned} f_Z(z) &= \int_0^z \mathbb{1}_{(0,1)}(x)\mathbb{1}_{(0,1)}(z-x)dx \\ &= \int_0^z dx \\ &= z \end{aligned}$$



III. Sea $1 \leq z \leq 2$.

$$\begin{aligned} f_Z(z) &= \int_0^z \mathbb{1}_{(0,1)}(x)\mathbb{1}_{(0,1)}(z-x)dx \\ &= \int_{z-1}^1 dx \\ &= 2-z \end{aligned}$$

Por lo tanto,

$$f_{X+Y}(z) = z\mathbb{1}_{[0,1]}(z) + (2-z)\mathbb{1}_{(1,2]}(z)$$

Ejemplo : Sean X_1 y X_2 *v. a.*'s independientes con $X_i \sim \text{Gamma}(\alpha_i, \lambda)$, $i = 1, 2$. Utilice la *f. d. p.* de las suma de *v. a.*'s para mostrar que $X_1 + X_2 \sim \text{Gamma}(\alpha_1 + \alpha_2, \lambda)$.

Solución: En sección anterior se mostró mediante *f. g. m.*'s el resultado. Ahora, se encontrará la *f. d. p.* de la suma. Para esto, sea $Z = X + Y$.

$f_Z(z) = 0$ para $z < 0$. Sea pues, $z \geq 0$,

$$\begin{aligned} f_Z(z) &= \int_0^z f_X(x)f_Y(z-x)dx \\ &= \int_0^z \frac{\lambda^{\alpha_1}}{\Gamma(\alpha_1)}x^{\alpha_1-1}e^{-\lambda x}\frac{\lambda^{\alpha_2}}{\Gamma(\alpha_2)}(z-x)^{\alpha_2-1}e^{-\lambda(z-x)}dx \\ &= \frac{\lambda^{\alpha_1+\alpha_2}e^{-\lambda z}}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\int_0^z x^{\alpha_1-1}(z-x)^{\alpha_2-1}dx \\ &= \frac{\lambda^{\alpha_1+\alpha_2}e^{-\lambda z}}{\Gamma(\alpha_1)\Gamma(\alpha_2)}\int_0^1 (zu)^{\alpha_1-1}(z-zu)^{\alpha_2-1}zdu \qquad ; u = x/z \\ &= \frac{\lambda^{\alpha_1+\alpha_2}e^{-\lambda z}}{\Gamma(\alpha_1)\Gamma(\alpha_2)}z^{\alpha_1+\alpha_2-1}\int_0^1 u^{\alpha_1-1}(1-u)^{\alpha_2-1}du \\ &= \frac{\lambda^{\alpha_1+\alpha_2}e^{-\lambda z}}{\Gamma(\alpha_1)\Gamma(\alpha_2)}z^{\alpha_1+\alpha_2-1}\cdot B(\alpha_1,\alpha_2) \\ &= \frac{\lambda^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1+\alpha_2)}z^{\alpha_1+\alpha_2-1}e^{-\lambda z} \end{aligned}$$

con el usó el cambio de variable y donde $u = x/z$ y $B(\alpha_1, \alpha_2)$ denota la **función Beta**[1] La función f_Z corresponde a la *f. d. p.* de una distribución Gamma con parámetro de forma $(\alpha_1 + \alpha_2)$ y parámetro tasa λ.

Convolución de funciones

Definición : En el caso de la suma de *v. a.*'s independientes la función f_Z se dice que es la **convolución** de las funciones (de densidad) f_X y f_Y y se denota por $f_Z = f_X * f_Y$. Esto

[1] $B(\alpha_1, \alpha_2) = \int_0^1 u^{\alpha_1-1}(1-u)^{\alpha_2-1}du = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)}$.

es,

$$f_Z(z) = (f_X * f_Y)(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)dx$$

La convolución es una operación entre funciones con propiedades como:

I. *Conmutatividad*: $g * h = h * g$.

II. *Asociatividad*: $f * (g * h) = (f * g) * h$.

Corolario : Sean $X_1, \ldots, X_n$ *v. a.*'s independientes con f_{X_i} la correspondiente *f. d. p.* de X_i. Entonces, si $S_n = X_1 + \cdots + X_n$,

$$f_{S_n}(w) = (f_{X_1} * \cdots * f_{X_n})(w)$$

8.2. Cociente de variables aleatorias

Considere ahora $\varphi(X, Y) = Y/X$. En este caso, para $z \in \mathbb{R}$

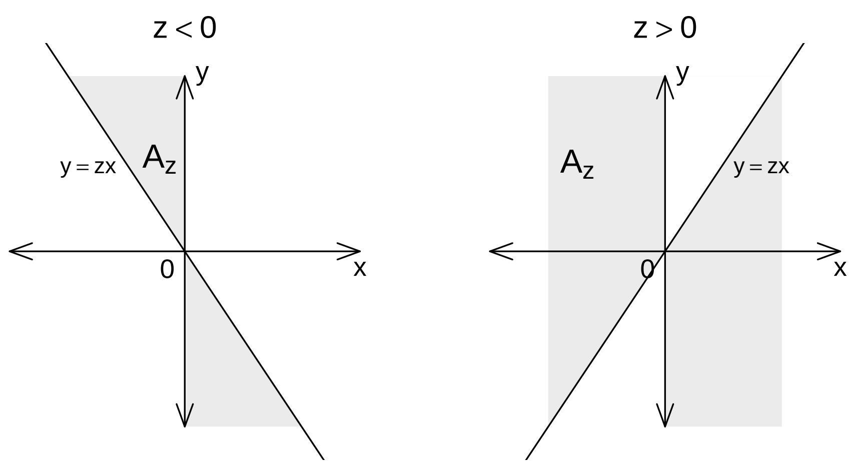
$$A_z = \big\{(x, y) : \varphi(x, y) \leq z\big\} = \left\{(x, y) : \frac{y}{x} \leq z\right\}$$

Note que

$$i) \text{ si } \ x < 0; \quad \frac{y}{x} \leq z \Rightarrow y \geq zx; \quad \text{ y } \quad ii) \text{ si } \ x > 0; \quad \frac{y}{x} \leq z \Rightarrow y \leq zx$$

Entonces,

$$A_z = \big\{(x, y) : x < 0, y \geq zx\big\} \cup \big\{(x, y) : x > 0, y \leq zx\big\}$$



$$F_Z(z) = \mathbb{P}(Z \leq z) = \mathbb{P}\big(\varphi(X, Y) \leq z\big) = \iint_{A_z} f(x, y)dxdy$$

$$
\begin{aligned}
F_Z(z) &= \int_{-\infty}^{0}\int_{zx}^{\infty} f(x,y)dydx + \int_{0}^{\infty}\int_{-\infty}^{zx} f(x,y)dydx \\
&= \int_{-\infty}^{0}\int_{z}^{-\infty} f(x,xv)xdvdx + \int_{0}^{\infty}\int_{-\infty}^{z} f(x,xv)xdvdx; \quad y = xv \\
&= \int_{-\infty}^{0}\int_{-\infty}^{z} -xf(x,xv)dvdx + \int_{0}^{\infty}\int_{-\infty}^{z} xf(x,xv)dvdx \\
&= \int_{-\infty}^{z}\left[\int_{-\infty}^{\infty} |x|f(x,xv)dx\right]dv \\
&= \int_{-\infty}^{z} h(v)dv
\end{aligned}
$$

donde se aplicó el teorema de Fubini para invertir el orden de integración con $h \geq 0$ pues es la integral de una función no negativa. Así, h es una *f. d. p.* para Z. Así, la función de densidad de probabilidad para $Z = Y/X$ está dada por

$$f_{Y/X}(z) = \int_{-\infty}^{\infty} |x|f(x,xz)dx$$

Corolario :

1. Sean X y Y *v. a.*'s independientes con f_X y f_Y sus *f. d. p.* correspondientes. Entonces,
$$f_{Y/X}(z) = \int_{-\infty}^{\infty} |x|f_X(x)f_Y(xz)dx$$
2. Sean X y Y *v. a.*'s independientes no negativas con f_X y f_Y sus *f. d. p.* correspondientes. Entonces,
$$f_{Y/X}(z) = \int_{0}^{\infty} xf_X(x)f_Y(xz)dx$$

Ejemplo : Sean X y Y *v.a.i.i.d.* normal estándar. Determine la distribución de $Z = Y/X$.

Solución: Sea $z \in \mathbb{R}$,

$$
\begin{aligned}
f_{Y/X}(z) &= \int_{-\infty}^{\infty} |x|\phi(x)\phi(zx)dx \\
&= \int_{-\infty}^{\infty} \frac{|x|}{2\pi}e^{-\frac{1}{2}(x^2+z^2x^2)}dx \\
&= \frac{1}{2\pi}\int_{0}^{\infty} e^{-\frac{1}{2}(1+z^2)x^2}(2xdx) \\
&= \frac{1}{2\pi\lambda}\int_{0}^{\infty} \lambda e^{-\lambda u}du; \qquad u = x^2, \lambda = \frac{1}{2}(1+z^2) \\
&= \frac{1}{2\pi\frac{1}{2}(1+z^2)} \\
&= \frac{1}{\pi(1+z^2)}
\end{aligned}
$$

que corresponde a la función de densidad de la distribución Cauchy(0,1).

Ejemplo : Sean X y Y *v.a.i.i.d.*'s $\mathrm{N}(0,\sigma^2)$. Encuentre la *f. d. p.* de $W = Y^2/X^2$.

Solución: Si $X \sim \mathrm{N}(0,\sigma^2)$, $X/\sigma \sim \mathrm{N}(0,1)$ y $X^2/\sigma^2 \sim \chi^2_1 \equiv \mathrm{Gamma}(\frac{1}{2}, 2)$. Luego,

$$W = \frac{Y^2}{X^2} = \frac{U}{V}, \quad \text{con } U = Y^2/\sigma^2,\ V = X^2/\sigma^2 \sim \mathrm{Gamma}(1/2, 2)$$

Se sigue,

$$\begin{aligned} f_W(w) &= \int_0^\infty u \frac{u^{-1/2}e^{-u/2}}{2^{1/2}\Gamma(1/2)} \cdot \frac{(wu)^{-1/2}e^{-wu/2}}{2^{1/2}\Gamma(1/2)} du \\ &= \frac{2w^{-1/2}}{2\pi(1+w)} \int_0^\infty \frac{1}{2}(1+w)e^{-(1+w)u/2} du \\ &= \frac{1}{\pi\sqrt{w}(1+w)} \mathbb{1}_{(0,\infty)}(w) \end{aligned}$$

Ejemplo : Sean X_1, X_2 *v. a.*'s independientes con $X_i \sim \text{Gamma}(\alpha_i, \beta)$. Determine la función de densidad de $Z = X_1/X_2$.

Solución: Sea $Z = X_1/X_2$, entonces el soporte de Z es $\mathbb{R}^+$. Sea pues $z > 0$.

$$\begin{aligned} f_Z(z) &= \int_0^\infty x f_1(zx) f_2(x) dx \\ &= \int_0^\infty x \frac{(zx)^{\alpha_1-1}e^{-zx/\beta}}{\beta^{\alpha_1}\Gamma(\alpha_1)} \cdot \frac{x^{\alpha_2-1}e^{-x/\beta}}{\beta^{\alpha_2}\Gamma(\alpha_2)} dx \\ &= \frac{z^{\alpha_1-1}}{\beta^{\alpha_1+\alpha_2}\Gamma(\alpha_1)\Gamma(\alpha_2)} \frac{1}{K} \int_0^\infty K x^{\alpha_1+\alpha_2-1} e^{-\frac{1+z}{\beta}x} dx \\ &= \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \frac{z^{\alpha_1-1}}{(1+z)^{\alpha_1+\alpha_2}} \end{aligned}$$

donde $K = \dfrac{\left[(1+z)/\beta\right]^{\alpha_1+\alpha_2}}{\Gamma(\alpha_1+\alpha_2)}$ es la constante normalizadora de la función de densidad de una distribución Gamma con parámetro de forma $\alpha_1+\alpha_2$ y $(1+z)/\beta$ como parámetro tasa. Por lo tanto,

$$f_{X_1/X_2}(z) = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \frac{z^{\alpha_1-1}}{(1+z)^{\alpha_1+\alpha_2}} \mathbb{1}_{(0,\infty)}(z)$$

Corolario : Sean $X_i \sim \chi^2_{n_i}$, *v. a.*'s independientes para $i = 1, 2$. Entonces, $Z = X_1/X_2$ tiene *f. d. p.* dada por

$$f_{X_1/X_2}(z) = \frac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})} \frac{z^{\frac{n_1}{2}-1}}{(1+z)^{\frac{n_1+n_2}{2}}} \mathbb{1}_{(0,\infty)}(z)$$

Demostración: Recuerde que si $X_i \sim \chi^2_{n_i}$, entonces $X_i \sim \text{Gamma}(n_i/2, 2)$ y aplique la proposición anterior.

Proposición : Sean $X \sim \chi^2_m$ y $Y \sim \chi^2_n$ independientes, entonces $W = \dfrac{X/m}{Y/n}$ tiene una *f. d. p.* dada por

$$f_W(w) = \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \left(\frac{m}{n}\right)^{m/2} \frac{w^{\frac{m}{2}-1}}{(1+\frac{m}{n}w)^{\frac{m+n}{2}}} \mathbb{1}_{(0,\infty)}(w) \tag{9}$$

Demostración. Sea $W = kZ$ y aplique la transformación del cociente Z de la proposición anterior por el factor $k = n/m$. Esto es, por el teorema de transformación, $f_W(w) = \frac{1}{k} f_Z\left(\frac{1}{k}w\right)$ y se sigue la proposición.

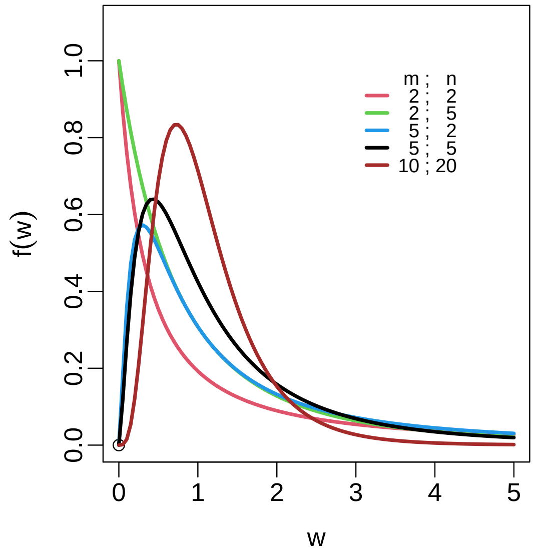


Figura 9: Función de densidad de la distribución F para distintos valores de los grados de libertad m y n.

8.3. La distribución F

Definición : Sea W la *v. a.* con *f. d. p.* dada por la expresión (9). Entonces, W se dice que sigue la **distribución F con m y n grados de libertad** y se denota por $W \sim F_{m,n}$.

Corolario : Si $X \sim \chi^2_m$ y $Y \sim \chi^2_n$ independientemente, entonces

$$W = \frac{X/m}{Y/n} \sim F_{m,n}$$

La figura 9 muestra la función de densidad de la distribución F para distintos valores de los grados de libertad. La flexibilidad de la función de densidad es resultado de los *dos parámetros* de la distribución.

El siguiente par de proposiciones muestra algunas propiedades de la distribución F. Su demostración se sigue, considerando que si $W \sim F_{m,n}$, W podría escribirse como $W = \frac{n}{m}XY^{-1}$, producto de *v. a.*'s independientes distribuidas ji-cuadrada con m y n grados de libertad respectivamente.

Proposición : Sea $W \sim F_{m,n}$, entonces

1. $\mathbb{E}[W] = \dfrac{n}{n-2}$, si $n > 2$.

2. $\text{var}(W) = \dfrac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$, si $n > 4$.

Demostración: $W \sim F_{m,n}$, entonces se puede pesar que $\dfrac{X/m}{Y/n}$, con X y Y independientes distribuidos χ^2 con m y n grados de libertad. Luego $W^r = kX^rY^{-r}$. Entonces, por la independencia de X y Y

$$\mathbb{E}[W^r] = k\mathbb{E}[X^r]\,\mathbb{E}[Y^{-r}]$$

Para el cálculo $\mathbb{E}[Y^{-r}]$ se utiliza el siguiente resultado de la distribución Gamma:
Si $W \sim \text{Gamma}(\alpha, \beta)$, entonces $\mathbb{E}[Y^r] = \dfrac{\Gamma(\alpha + r)}{\Gamma(\alpha)}\beta^r$, para todo $r > -\alpha$.

Aplique las ideas anteriores para demostrar la proposición.

Proposición : Sea $W \sim F_{m,n}$, entonces

1. Si $T \sim t_n$ (t-Student con n grados de libertad), entonces $T^2 \sim F_{1,n}$.
2. $W^{-1} \sim F_{n,m}$.
3. $\mathbb{P}(W \leq w) = 1 - \mathbb{P}(W^{-1} \leq w^{-1})$.
4. Sea $0 < p < 1$ y $F(p; m, n)$ el p-ésimo percentil de la distribución de W, esto es, $p = \mathbb{P}(W \leq F(p; m, n))$, entonces se tiene que
$$F(p; m, n) = \frac{1}{F(1 - p; n, m)}$$

Demostración: Defina la v. a. W como el cociente de v. a.'s independientes distribuidas χ^2. Por ejemplo, $W = \dfrac{X/m}{Y/n}$, donde $X \sim \chi^2_m$ y $Y \sim \chi^2_n$. Vea la lista de ejercicios.

Ejemplo : Sean $W \sim F_{m,n}$ y $p = 0.25$, $m = 3$ y $n = 6$. Con la notación en la proposición anterior y utilizando tablas de probabilidad o alguna aplicación

$$0.25 = \mathbb{P}\left(F_{3,6} \leq F(0.25; 3, 6)\right) = \mathbb{P}\left(W \leq 0.423\right) = \mathbb{P}\left(W \leq \frac{1}{2.422}\right)$$

donde $2.422 = F(0.75; 6, 3)$.

Corolario : Sean $X_1, \ldots, X_{n+1}$, *v.a.i.i.d.* $\mathrm{N}(0, \sigma^2)$. Para $k = 1, \ldots, n+1$,

$$\frac{X_1^2 + \cdots + X_k^2}{\sigma^2} \sim \chi_k^2,$$

Entonces,

I.
$$\frac{n-k}{k} \cdot \frac{X_1^2 + \cdots + X_k^2}{X_{k+1}^2 + \cdots + X_n^2} \sim F_{k,n-k}$$

II.
$$\frac{nX_{n+1}^2}{X_1^2 + \cdots + X_n^2} \sim F_{1,n}$$

8.4. Ejercicios

Refiérase al Cuaderno de Ejercicios sección 8, Barrios and Chambon (2024).

9. Estadísticos de orden

En Estadística uno se refiere a una **muestra aleatoria (*m. a.*) de tamaño** n de una población X con función de densidad de probabilidad (*f. d. p.*) f_X, a una colección $X_1, \ldots, X_n$ de variables aleatorias independientes e idénticamente distribuidas con *f. d. p.* común f_X. Así, se puede denotar como

$$\boldsymbol{X} = (X_1, \ldots, X_n), \; m. \; a. \;\text{ de }\; X \sim f_X$$

Un **estadístico** es una función de la muestra aleatoria que no depende de parámetros desconocidos. Por ejemplo, $\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$, $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, $X_m = \text{mín}\{X_1, \ldots, X_n\}$, mediana$(\boldsymbol{X})$, $X_M = \text{máx}\{X_1, \ldots, X_n\}$, etc. Por otro lado, $\sum_{i=1}^{n}(X_i - \mu)^2$, no es un estadístico pues depende del parámetro desconocido μ.

Sea $\boldsymbol{X} = (X_1, \ldots, X_n)$ una nuestra aleatoria de $X \sim f$. Las X_i's son *v.a.i.i.d.*. Se definen los **estadísticos de orden** de la *m. a.* $\boldsymbol{X}$ a las *v. a.*'s Y_i's que representan la *muestra ordenada* y son tales que, para todo $\omega \in \Omega$,

$$Y_1(\omega) \leq Y_2(\omega) \leq \cdots \leq Y_n(\omega)$$

Así, para todo $\omega \in \Omega$,

$$\begin{aligned}
Y_1(\omega) &= \text{mín}\{X_1(\omega), \ldots, X_n(\omega)\} \\
Y_2(\omega) &= \text{la segunda realización más pequeña de las } X_i\text{'s} \\
&\;\vdots \\
Y_r(\omega) &= \text{ es tal que } \; Y_1(\omega) \leq \cdots \leq Y_{r-1}(\omega) \leq Y_r(\omega) \leq Y_{r+1}(\omega) \leq \cdots \leq Y_n(\omega) \\
&\;\vdots \\
Y_n(\omega) &= \text{máx}\{X_1(\omega), \ldots, X_n(\omega)\}
\end{aligned}$$

Definición : Sea $\boldsymbol{X} = (X_1, \ldots, X_n)$ una *m. a.* de X, se define a Y_r como el r**-ésimo estadístico de orden** y se denota por $X_{(r)} = Y_r$, $r = 1, \ldots, n$.

Por ejemplo, el primer estadístico de orden es $X_{(1)} = \text{mín}\{X_1, \ldots, X_n\}$.

Definición : Sea $\boldsymbol{X} = (X_1, \ldots, X_n)$ una *m. a.* de X, se define el **rango** de la muestra como $R(\boldsymbol{X}) = X_{(n)} - X_{(1)}$, esto es

$$\text{Rango}(\boldsymbol{X}) = \text{máx}(\boldsymbol{X}) - \text{mín}(\boldsymbol{X})$$

Ejemplo : Considere un mecanismo compuesto por n componentes idénticas e independientes con tiempos de vida $T_1 \ldots, T_n$.

a). Si las componentes están conectadas en *serie*, $T_{(1)} = \text{mín}\{T_1, \ldots, T_n\}$ es el tiempo de operación del mecanismo.

b). Si las componentes están conectadas en *paralelo*, $T_{(n)} = \text{máx}\{T_1, \ldots, T_n\}$ es el tiempo de operación del mecanismo.

Ejemplo : Suponga una línea de producción de partes, por ejemplo tornillos, supuestamente idénticos. Sean $X_1 \ldots, X_n$ las longitudes de los tornillos, si $X_{(1)}$ y $X_{(n)}$ están dentro de especificaciones, entonces *todos* los tornillos lo estarán. Note además que el rango de la muestra, $R(X) = X_{(n)} - X_{(1)}$, es una medida de la variación de la producción.

Tabla 1: Muestra observada de tamaño 8 y estadísticos de orden.

i	x_i	orden	$x_{(i)}$
1	0.56	5	0.01
2	0.36	3	0.25
3	0.25	2	0.36
4	0.78	6	0.50
5	0.01	1	0.56
6	0.50	4	0.78
7	0.84	7	0.84
8	0.97	8	0.97

$$\text{Rango de la muestra} = x_{\text{máx}} - x_{\text{mín}} = 0.97 - 0.01 = 0.96$$

Ejemplo : La tabla 1 muestra la realización x_i de una muestra de tamaño 8 de una distribución uniforme en el intervalo $(0,1)$ y los correspondientes estadísticos de orden $x_{(i)}$.

Ejemplo : La tabla 2 muestra la realización de 12 muestras aleatorias de tamaño 8 de una distribución uniforme $(0,1)$, los máximos y mínimos y el correspondiente rango de la muestra. Las muestras se presentan en la figura 10. En la figura se se identifican los máximos ($\triangle$) y mínimos ($\bigtriangledown$). En el margen derecho se ven las correspondientes realizaciones (12) de los estadísticos de orden *máximo* y *mínimo*. Note que la distribución de éstos es diferente a la distribución original de las X's. En la figura se resalta (zona sombreada) la quinta muestra, cuyos estadísticos de orden se mostraron antes en la tabla 1.

Tabla 2: Realización 12 muestras de tamaño 8 de la distribución uniforme (0,1).

	muestras											
	1	2	3	4	5	6	7	8	9	10	11	12
x_1	0.62	0.82	0.29	0.09	0.56	0.72	0.22	0.91	0.85	0.12	0.05	0.61
x_2	0.02	0.59	0.28	0.88	0.36	0.67	0.20	1.00	0.17	0.97	0.95	0.43
x_3	1.00	0.25	0.14	0.32	0.25	0.73	0.01	0.84	0.46	0.20	0.44	0.23
x_4	0.85	0.54	0.62	0.36	0.78	0.99	0.25	0.32	0.70	0.86	0.96	0.63
x_5	0.69	0.41	0.18	0.06	0.01	0.19	0.23	0.41	0.40	0.74	0.05	0.17
x_6	0.82	0.32	0.51	0.52	0.50	0.41	0.34	0.36	0.83	0.10	0.29	0.58
x_7	0.43	0.67	0.10	0.04	0.84	0.58	0.23	0.01	0.27	0.27	0.43	0.59
x_8	0.91	0.50	0.87	0.10	0.97	0.27	0.46	0.04	0.97	0.06	0.82	0.59
$x_{\text{mín}}$	0.02	0.25	0.10	0.04	0.01	0.19	0.01	0.01	0.17	0.06	0.05	0.17
$x_{\text{máx}}$	1.00	0.82	0.87	0.88	0.97	0.99	0.46	1.00	0.97	0.97	0.96	0.63
rango	0.97	0.57	0.77	0.84	0.96	0.81	0.45	0.99	0.80	0.91	0.91	0.46

Ejemplo : El panel de la izquierda de la figura 11 muestra el histograma de $N = 2000$ muestras de tamaño $n = 25$ de una distribución Gamma de parámetros $\alpha = 2.7$ y $\beta = 1.5$. Los puntos en la base del histograma muestran una de las muestras en particular. El décimo estadístico de orden $x_{(10)}$ se resalta en color rojo. El panel la derecha exhibe el histograma y la curva suavizada de la misma correspondiente a los 10-ésimos estadísticos de orden de las N=2000 muestras. Note que la distribución de los X_{10}'s no es la distribución original de $X \sim \text{Ga}(2.7, 1.5)$.

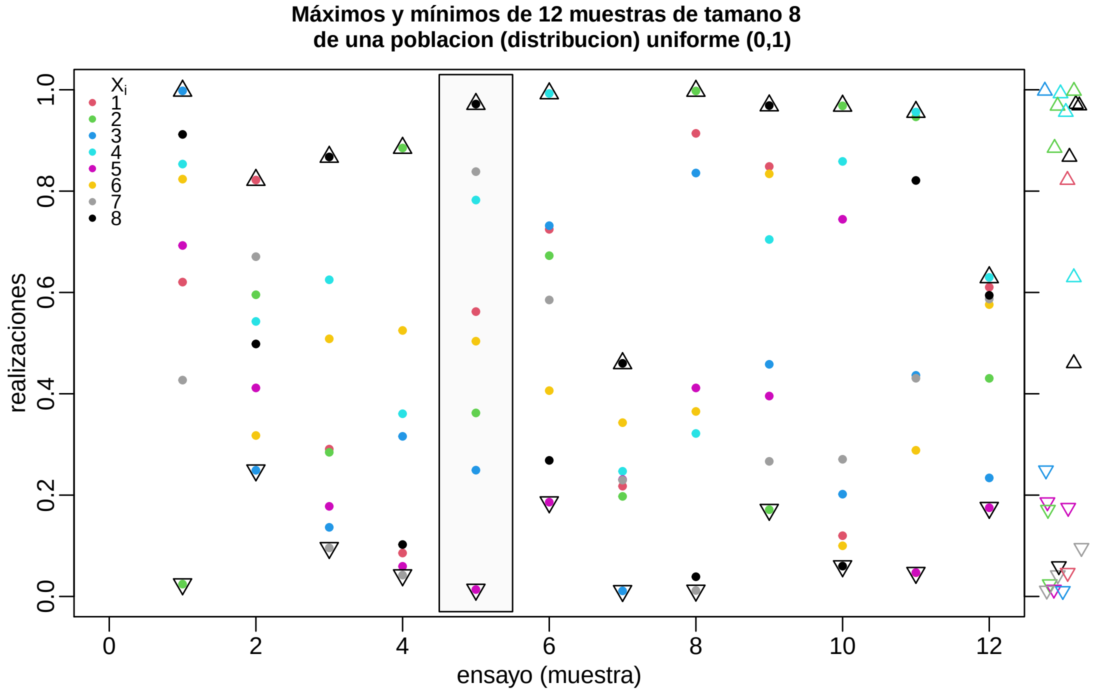


Figura 10: Muestras de tamaño 8 de la distribución uniforme (0,1). Código de colores en la esquina superior izquierda. △ encierra el máximo de la muestra, ▽ el mínimo.

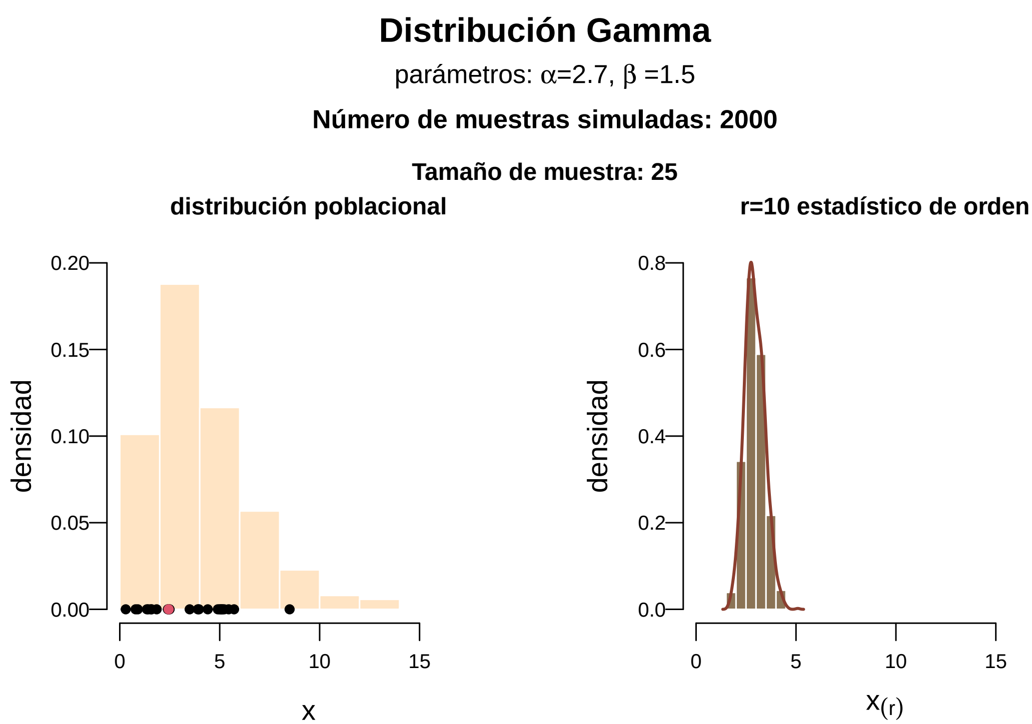


Figura 11: Histograma de $N = 2000$ realizaciones de muestras de tamaño $n = 25$ de una distribución Gamma(2.7,1.5) y el histograma de los correspondiente estadísticos de orden $r = 10$.

9.1. Funciones de distribución y de densidad del r–ésimo estadístico de orden

Sea $\boldsymbol{X} = (X_, \ldots, X_n)$ una *m. a.* de X con *f. p. a.* F y *f. d. p.* f. Sea $x \in \mathbb{R}$, la probabilidad de que exactamente r de los X_i's hayan caído en $(-\infty, x]$ y $(n-r)$ en (x, ∞) es

$$\binom{n}{r} F(x)^r [1 - F(x)]^{n-r}$$

El evento $\left\{X_{(r)} \leq x\right\}$ ocurre si y solo si r o más de las X_i's caen en $(\infty, x]$. Entonces, si $F_r \equiv F_{X_{(r)}}$ denota la *f. p. a.* r-ésimo estadístico de orden, $r = 1, \ldots, n$, se tiene que para todo $x \in \mathbb{R}$,

$$F_r(x) = \mathbb{P}\left(X_{(r)} \leq x\right) = \sum_{k=r}^{n} \binom{n}{k} F(x)^k [1 - F(x)]^{n-k}$$

Así por ejemplo,

$$\begin{aligned}
F_n(x) &= F(x)^n \\
F_1(x) &= 1 - \big[1 - F(x)\big]^n
\end{aligned}$$

O bien, alternativamente

$$\begin{aligned}
F_n(x) &= \mathbb{P}\left(X_n \leq x\right) = \mathbb{P}\big(\text{máx}\{X_i\} \leq x\big) \\
&= \mathbb{P}\left(X_1 \leq x, \ldots, X_n \leq x\right) \\
&\overset{\text{ind}}{=} \mathbb{P}\left(X_1 \leq x\right) \cdots \mathbb{P}\left(X_n \leq x\right) \\
&= F(x)^n
\end{aligned}$$

$$\begin{aligned}
F_1(x) &= \mathbb{P}\left(X_1 \leq x\right) = \mathbb{P}\big(\text{mín}\{X_i\} \leq x\big) \\
&= 1 - \mathbb{P}\big(\text{mín}\{X_i\} \geq x\}\big) \\
&= 1 - \mathbb{P}\left(X_1 \geq x, \ldots, X_n \geq x\right) \\
&\overset{\text{ind}}{=} 1 - \mathbb{P}\left(X_1 \geq x\right) \cdots \mathbb{P}\left(X_n \geq x\right) \\
&= 1 - \big[1 - F(x)\big]^n
\end{aligned}$$

Las correspondientes *f. d. p.*'s, para $x \in \mathbb{R}$

$$\begin{aligned}
f_n(x) &= \frac{d}{dx} F_n(x) = n f(x) F(x)^{n-1} \\
f_1(x) &= \frac{d}{dx} F_1(x) = n f(x) \big[1 - F(x)\big]^{n-1}
\end{aligned}$$

Y, para el r-ésimo estadístico de orden, $r = 1, \ldots, n$, se tiene

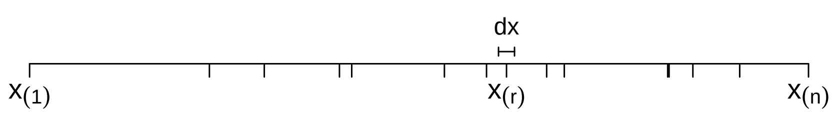
$$f_r(x) = n \binom{n-1}{r-1} f(x) F(x)^{r-1} \big[1 - F(x)\big]^{n-r} \tag{10}$$

Para la justificación del resultado anterior se presentan dos argumentos a). De acuerdo a Hoel et al. (1971)

$$
\begin{aligned}
f_r(x) &= \frac{d}{dx}F_r(x)\\
&= \sum_{k=r}^{n}\binom{n}{k}\Big\{kf(x)F(x)^{k-1}\left[1-F(x)\right]^{n-k}\\
&\qquad -(n-k)f(x)F(x)^{k}\left[1-F(x)\right]^{n-k-1}\Big\}\\
&= \sum_{k=r}^{n}\frac{n!}{(k-1)(n-k)!}f(x)F^{k-1}\left[1-F(x)\right]^{n-k}\\
&\qquad -\sum_{k=r}^{n-1}\frac{n!}{k!(n-k-1)!}f(x)F^{k}(x)\left[1-F(x)\right]^{n-k-1}\\
&= \sum_{k=r}^{n}\frac{n!}{(k-1)(n-k)!}f(x)F^{k-1}\left[1-F(x)\right]^{n-k}\\
&\qquad -\sum_{k=r+1}^{n}\frac{n!}{(k-1)!(n-k)!}f(x)F^{k-1}(x)\left[1-F(x)\right]^{n-k}
\end{aligned}
$$

Después de algunas cancelaciones se obtiene f_r dado en la expresión (10).

b). De acuerdo a Blitzstein and Hwang (2014)



Considere dx la diferencial alrededor de x y f_r la *f. d. p.* de $X_{(r)}$. Luego, $f_r(x)dx$ es la probabilidad de que el r-ésimo estadístico de orden $X_{(r)}$ caiga en le intervalo de longitud dx alrededor del punto x. Entonces,

$$
f_r(x)dx = \underbrace{\binom{n-1}{r-1}F(x)^{r-1}}_{(b)}\cdot\underbrace{nf(x)dx}_{(a)}\cdot\underbrace{\binom{n-r}{n-r}[1-F(x)]^{n-r}}_{(c)}
$$

donde,

(a) Se elije aquel X_i de los n posibles y $f(x)dx$ es la probabilidad de que caiga en el intervalo dx alrededor de x.

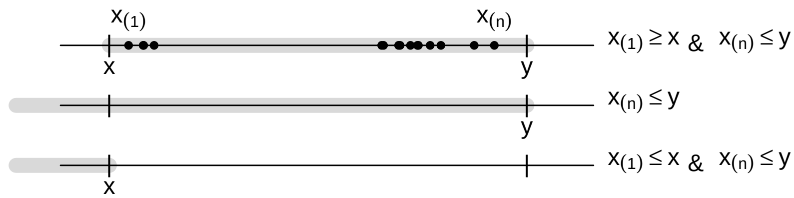
(b) Se elijen los $(r-1)$ X_i's de $(n-1)$ posibles que caerán a la izquierda de x.

(c) Es la probabilidad de que los $(n-r)$ restantes X_i's caigan a la derecha de x.

Cancelando dx de ambos lados se tiene la expresión (10), la *f. d. p.* de $X_{(r)}$.

9.2. Función de densidad del rango

Considere ahora el rango de la muestra $R = X_{(n)} - X_{(1)}$. Para determinar la *f. d. p.* de R se necesita la *f. d. p.* conjunta de $X_{(1)}$ y $X_{(n)}$. Para esto, sean $x < y$,



$$
\begin{aligned}
\mathbb{P}\left(X_{(1)} > x, X_{(n)} \le y\right) &= \mathbb{P}\left(x < X_1 \le y, \ldots, x < X_n \le y\right) \\
&\overset{\text{ind}}{=} \left[\mathbb{P}\left(x < X \le y\right)\right]^n \\
&= \left[F(y) - F(x)\right]^n
\end{aligned}
$$

Luego, la función de distribución conjunta de (Y_1, Y_2), el primero y último estadístico de orden, es

$$
\begin{aligned}
F_{1n}(x,y) &= \mathbb{P}\left(X_{(1)} \le x, X_{(n)} \le y\right) \\
&= \mathbb{P}\left(X_{(n)} \le y\right) - \mathbb{P}\left(X_{(1)} > x, X_{(n)} \le y\right) \\
&= \left[F(y)\right]^n - \left[F(y) - F(x)\right]^n
\end{aligned}
$$

Se sigue que para $x < y$,

$$
f_{1n}(x,y) = \frac{\partial^2 F_{1n}(x,y)}{\partial x \partial y} = n(n-1) f(x) f(y) \left[F(y) - F(x)\right]^{n-2}
$$

Por lo tanto, para $R = X_{(n)} - X_{(1)},\ r \ge 0$,

$$
f_R(r) = \int_{\mathbb{R}} f_{1n}(u, r+u) du = n(n-1) \int_{-\infty}^{\infty} f(u) f(r+u) \left[F(r+u) - F(u)\right]^{n-2} du
$$

9.3. Función de densidad conjunta de los estadísticos de orden

Sea $\boldsymbol{X} = (X_1, \ldots, X_n)$ una *m. a.* de $X \sim f$. Se sigue de la independencia y lo idénticamente distribuidos de los elementos de la muestra que la *f. d. p.* conjunta de $\boldsymbol{X}$ es

$$
f_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{i=1}^{n} f(x_i)
$$

Sean $Y_i = X_{(i)}$, el i-ésimo estadístico de orden, $i = 1, \ldots, n$. Entonces, $X \longrightarrow Y$ es una mera permutación de las componentes de X. Por ejemplo, si $\boldsymbol{X} = (X_1, X_2, X_3)$, se tienen $3! = 6$ posibles permutaciones

$$
\boldsymbol{X} = (X_1, X_2, X_3) \quad \longrightarrow \quad
\begin{array}{l}
\boldsymbol{Y} = (X_1 < X_2 < X_3) \\
\boldsymbol{Y} = (X_1 < X_3 < X_2) \\
\boldsymbol{Y} = (X_2 < X_1 < X_3) \\
\boldsymbol{Y} = (X_2 < X_3 < X_1) \\
\boldsymbol{Y} = (X_3 < X_1 < X_2) \\
\boldsymbol{Y} = (X_3 < X_2 < X_1)
\end{array}
$$

Para la permutación particular $\boldsymbol{X} \to \boldsymbol{Y} = (X_2 < X_1 < X_3)$, se tendría

$$
\begin{array}{ll}
Y_1 = g_1(\boldsymbol{X}) = X_2 & \quad X_1 = h_1(\boldsymbol{Y}) = Y_2 \\
Y_2 = g_2(\boldsymbol{X}) = X_1 & \quad X_2 = h_2(\boldsymbol{Y}) = Y_1 \\
Y_3 = g_1(\boldsymbol{X}) = X_3 & \quad X_3 = h_3(\boldsymbol{Y}) = Y_3
\end{array}
$$

$h(\boldsymbol{Y}) = (Y_2, Y_1, Y_3)$. Luego,

$$|Jh(\boldsymbol{y})| = \begin{vmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1$$

que es una matriz de permutación. Luego, para esta permutación particular, el teorema de transformación

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = f_{\boldsymbol{X}}(h(\boldsymbol{y}))|Jh(\boldsymbol{y})| = \prod_{i=1}^{3} f(y_i) \cdot 1$$

pero hay 3! distintas permutaciones, todas con la misma *f. d. p.* conjunta. Por lo tanto

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \sum_{\{j:\text{permutaciones}\}} f_{\boldsymbol{X}}(h_j(\boldsymbol{y}))|Jh_J(\boldsymbol{y})| = 3! \prod_{i=1}^{3} f(y_i) \mathbb{1}_{\{y_1<y_2<y_3\}}(\boldsymbol{y})$$

En general, para $i = 1, \ldots, n$, se tiene $y_i = g_i(x) = x_j$, para algún j y éstos no se repiten. Luego, $x_i = h_i(\boldsymbol{y})$ que sería la permutación inversa y $|Jh| = 1$. Entonces, por el teorema de cambio de variable

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = \sum_{\{j:\text{permutaciones}\}} f_X(h_j(\boldsymbol{y}))|Jh_j(\boldsymbol{y})| = n! \prod_{i=1}^{n} f(y_i) \mathbb{1}_{\{y_1<\cdots<y_n\}}(\boldsymbol{y})$$

Ejemplo : (Ross (2006)) A lo largo de una milla se distribuyen "al azar" tres personas. Determine la probabilidad de que no haya dos personas a menos de d millas.

Solución: Note que si $d \geq 1/2$ necesariamente al menos dos personas estarán a menos de d millas. Así pues, sea $d < 1/2$. Sea X_i la posición de la i-ésima persona. Sean $Y_i = X_{(i)}$, los correspondientes estadísticos de orden. Luego, la probabilidad solicitada es $p = \mathbb{P}(A)$ donde $A = \{Y_i > Y_{i-1} + d; i = 2, 3\}$. Entonces,

$$\begin{aligned}
\mathbb{P}(A) &= \int_A f_{\boldsymbol{Y}}(\boldsymbol{y})d\boldsymbol{y} \\
&= \int_0^{1-2d} \int_{y_1+d}^{1-d} \int_{y_2+d}^{1} 3! dy_3 dy_2 dy_1 \\
&= 6 \int_0^{1-2y} \int_{y_1+d}^{1-d} (1 - y_2 - d) dy_2 dy_1 \\
&\vdots \\
&= (1 - 2d)^3
\end{aligned}$$

El resultado puede entenderse a n personas. En este caso $d < 1/(n-1)$ y $p = \big[1 - (n-1)d\big]^n$.

Ejemplo : Sean $X_1, \ldots, X_n$ variables aleatorias independientes con $X_i \sim \text{Exp}(\lambda_i)$, $i = 1, \ldots, n$. Muestre que

$$\mathbb{P}\left(X_k = \text{mín}\{X_i\}\right) = \frac{\lambda_k}{\sum_{i=1}^{n} \lambda_i}$$

Solución:

$$\begin{aligned}
\mathbb{P}\left(X_k = \text{mín}\{X_i\}\right) &= \mathbb{P}\left(\cap_{i\neq k}\{X_k \leq X_i\}\right) \\
&\overset{\text{TPT}}{=} \int_0^\infty \mathbb{P}\left(\cap_{i\neq k}\{X_i \geq x | X_k = x\}\right) f_X(x)dx \\
&\quad \text{por independencia de las } X_i\text{'s} \\
&= \int_0^\infty \prod_{i\neq k} \mathbb{P}\left(X_i \geq x\right) \cdot f_X(x)dx \\
&= \int_0^\infty \prod_{i\neq k} e^{-\lambda_i x} \cdot \lambda_k e^{-\lambda_k x} dx \\
&= \lambda_k \int_0^\infty e^{-\sum_{i=1}^n \lambda_i x} dx \\
&= \frac{\lambda_k}{\sum_{i=1}^n \lambda_i}
\end{aligned}$$

9.4. Ejercicios

Refiérase al Cuaderno de Ejercicios sección 9, Barrios and Chambon (2024).

10. Desigualdades

Desigualdad de Markov: Sea X una variable aleatoria positiva con media finita, $\mu_X < \infty$. Entonces, para todo $c > 0$, se tiene

$$\mathbb{P}\left(X \geq c\right) \leq \frac{\mathbb{E}[X]}{c} = \frac{\mu_X}{c}$$

Demostración: Sin pérdida de generalidad suponga que X es una *v. a.* continua positiva con *f. d. p.* f. Sea $\mathbb{E}[X] = \mu_X$, entonces,

$$\begin{aligned}
\mu_X = \mathbb{E}[X] &= \int_0^\infty x f(x) dx \\
&\geq \int_c^\infty x f(x) dx \\
&\geq c \int_c^\infty f(x) dx \\
&= c\mathbb{P}\left(X \geq c\right)
\end{aligned}$$

y se sigue la desigualdad.

10.1. Desigualdad de Chebyshev

Desigualdad de Chebyshev: Sea X una variable aleatoria con media μ_X y varianza σ_X^2 finita. Entonces

$$\mathbb{P}\left(|X - \mu_X| \geq \epsilon\right) \leq \frac{\sigma_X^2}{\epsilon^2}$$

Demostración: Sean $\mu_X = \mathbb{E}[X]$ y $\sigma_X^2 = \mathrm{var}(X)$. Defina $Y = (X - \mu_X)^2$. Entonces, Y es una *v. a.* positiva y por la desigualdad anterior

$$\begin{aligned}
\mathbb{P}\left(Y \geq \epsilon^2\right) &\leq \frac{\mathbb{E}[Y]}{\epsilon^2} \\
\mathbb{P}\left((X - \mu_X)^2 \geq \epsilon^2\right) &\leq \frac{\mathbb{E}[(X-\mu)^2]}{\epsilon^2} \\
\mathbb{P}\left(|X - \mu_X| \geq \epsilon\right) &\leq \frac{\sigma_X^2}{\epsilon^2}
\end{aligned}$$

Algunos textos presentan alternativamente la desigualdad de Chebyshev en términos de la desviación estándar. A saber, para todo $k > 0$,

$$\mathbb{P}\left(|X - \mu_X| < k\sigma_X\right) \geq 1 - \frac{1}{k^2}$$

Ejercicio : Verifique el resultado anterior.

En ocasiones la desigualdad de Chebyshev puede ser muy conservadora pero en otros en que *no puede ser mejorada*, en el sentido que la cota se alcanza,

Ejemplo : Considere la *v. a.* X con la siguiente *f. m. p.*

x	-1	0	$+1$
p	$1/8$	$6/8$	$1/8$

En este caso se tiene $\mathbb{E}[X] = 0$, $\mathrm{var}(X) = \mathbb{E}[X^2] = 2/8$ y

$$\begin{aligned}
\mathbb{P}\left(|X - \mu_X| \geq 2 \cdot \frac{1}{2}\right) &= \mathbb{P}\left(|X| \geq 1\right) = \frac{2}{8} \\
&\leq \frac{\sigma_X^2}{\epsilon^2} = \frac{1/4}{1^2} = \frac{1}{4}
\end{aligned}$$

En el caso anterior se alcanza la igualdad por lo que la Desigualdad de Chebyshev no puede ser mejorada. Sin embargo, en ocasiones la desigualdad pues resultar demasiado conservadora al indicar, por ejemplo que $\mathbb{P}\left(|x-\mu_X| \geq \epsilon\right) \leq 1$.

Deigualdad de Chebyshev de un solo lado: Sea X una *v. a.* con media cero y varianza finita. Entonces, para todo $a > 0$,

$$\mathbb{P}\left(X \geq a\right) \leq \frac{\sigma_X^2}{a^2 + \sigma_X^2}$$

Demostración: Vea Ross (2006).

Ejemplo : El número de clientes por día es una caja tiene una media de 20 clientes y una desviación estándar de 2 clientes. ¿Qué diría de la probabilidad de tener mañana entre 17 y 23 clientes?

Solución: $\mu_X = \mathbb{E}[X] = 20$ y $\sigma_X^2 = \text{var}(X) = 2^2 = 4$. Se sigue de la desigualdad de Chebyshev que

$$\begin{aligned}
\mathbb{P}\left(17 \leq X \leq 23\right) &= \mathbb{P}\left(16 < X < 24\right) \\
&= \mathbb{P}\left(-4 < X - \mu < 4\right) \\
&= \mathbb{P}\left(|X-\mu| < 4\right) \\
&= \mathbb{P}\left(|X-\mu| < 2(2)\right) \\
&\geq 1 - \frac{1}{2^2} = \frac{3}{4}
\end{aligned}$$

Note que si $\sigma = 1$, entonces $k = 4$ para cubrir el intervalo $[17, 23]$ y en tal caso

$$\mathbb{P}\left(17 \leq X \leq 23\right) = 1 - \frac{1}{16} = \frac{15}{16}$$

Esto es, si se reduce la dispersión (varianza) aumenta la probabilidad del intervalo $[17, 23]$.

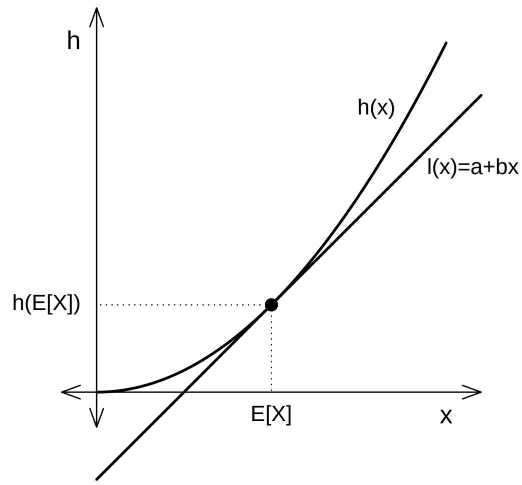
Ejemplo : Se sabe por experiencia que el tiempo medio para reparar una máquina es de 6.2 h y una desviación estándar de 3.52 h. Suponga que un empleado nuevo se lleva 22.5 h en reparar una máquina. ¿Considera usted que necesita de capacitación o entrenamiento?

Solución: Sea Y el tiempo (aleatorio) de reparación de una máquina. Si se aplica la Desigualdad de Chebyshev de un lado se tiene

$$\mathbb{P}\left(Y \geq 225\right) = \mathbb{P}\left((Y-6.2) \geq 16.3\right) \leq \frac{3.52^2}{16.3^2 + 3.52^2} = 0.044$$

La probabilidad es relativamente baja, por lo que se sugiere entrenamiento.

10.2. Desigualdad de Jensen



Definición : Una función h continua en $\mathbb{R}$ se dice **convexa** si para todo $x_0 \in \mathbb{R}$, existe una línea recta $\ell(x) = a + bx$ que pasa por $(x_0, h(x_0))$ y queda por debajo de $h(x)$. Esto es, $h(x) \geq \ell(x)$, para todo $x \in \mathbb{R}$.

Desigualdad de Jensen. Sea X una *v. a.* continua con esperanza finita y h una función convexa tal que $h(\mathbb{E}[X])$ existe. Entonces, se tiene que

$$\mathbb{E}\left[h(X)\right] \geq h\left(\mathbb{E}[X]\right)$$

Demostración: Puesto que h es una función convexa, existe una recta $\ell(x) = a + bx$ que pasa por $\big(\mathbb{E}[X], h(\mathbb{E}[X])\big)$ y satisface que $h(x) \geq \ell(x) = a + bx$. (La recta ℓ es la de la definición de arriba de función convexa.) Se tiene que $h(X) \geq \ell(X)$ y tomando valor esperado de ambos lados

$$\mathbb{E}[h(X)] \geq \mathbb{E}[\ell(X)] = \mathbb{E}[a + bX] = a + b\mathbb{E}[X] = \ell(\mathbb{E}[X]) = h(\mathbb{E}[X])$$

y se tiene la desigualdad.

Ejemplo : Sea $X \sim \text{unif}(0, 1)$ y $h(x) = x^2$. Entonces, $E[X] = 1/2$ y $h(\mathbb{E}[X]) = 1/4$.

Por otro lado, $\mathbb{E}[h(X)] = \mathbb{E}[X^2] = 1/3$. Así se cumple que

$$\mathbb{E}[h(X)] = \frac{1}{3} \geq \frac{1}{4} = h(\mathbb{E}[X])$$

10.3. Ejercicios

Refiérase al Cuaderno de Ejercicios sección 10, Barrios and Chambon (2024).

11. Sucesión de variables aleatorias y teoremas límite

Las definiciones sobre modos de convergencia de variables aleatorias que aquí se presentan se han tomado del libro de Prof. Roussas (1997).

11.1. Modos de convergencia de variables aleatorias

Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad (*EP*) y sean X y $X_1, X_2, \ldots$ variables aleatorias definidas sobre ese *EP*. A continuación se presentan distintos modos de convergencia de sucesiones de *v. a.*'s.

Definición : Se dice que $\{X_n\}$ **converge casi seguramente** o que **converge con probabilidad 1** a X y se denota por $X_n \xrightarrow{\text{cs}} X$, o bien, $X_n \xrightarrow{\text{cp1}} X$, si para "casi todo" $\omega \in \Omega$ [2], $X_n(\omega) \longrightarrow X(\omega)$.

Esto es, $X_n \xrightarrow{\text{cs}} X$, si $\forall\ \epsilon > 0$ y $\forall\ \omega \in \Omega \backslash D, \exists\ N(\omega, \epsilon)$, tal que

$$| X_n(\omega) - X(\omega) | < \epsilon, \quad n > N$$

El tipo de convergencia casi segura se le conoce también como *convergencia fuerte*.

Definición : Se dice que $\{X_n\}$ **converge en probabilidad** a X y se denota por $X_n \xrightarrow{P} X$, si para todo $\epsilon > 0$, $\mathbb{P}\big(| X_n - X | > \epsilon\big) \longrightarrow 0$.

Esto es, si $\forall\ \epsilon > 0$ y $\delta > 0,\ \exists\ N(\epsilon, \delta)$, tal que

$$\mathbb{P}\big(| X_n - X | > \epsilon\big) < \delta, \quad n > N$$

Note que la convergencia en probabilidad implica que $\mathbb{P}\big(| X_n - X | \leq \epsilon\big) \longrightarrow 1$. El tipo de convergencia en probabilidad se le conoce también como *convergencia débil*.

Definición : Se dice que $\{X_n\}$ **converge en media cuadrática** a X y se denota por $X_n \xrightarrow{\text{mc}} X$, si $\mathbb{E}\left[| X_n - X |^2\right] \longrightarrow 0$.

Esto es, si $\forall\ \epsilon > 0\ \exists\ N(\epsilon)$, tal que

$$\mathbb{E}\left[| X_n - X |^2\right] < \epsilon, \quad n > N$$

Definición : Se dice que $\{X_n\}$ **converge en distribución** a X y se denota por $X_n \xrightarrow{D} X$, si $F_n(x) \longrightarrow F(x)$ para todo $x \in \mathcal{C}(F)$, puntos de continuidad de F, la *f. p. a.* de X y las F_n's las correspondientes de X_n.

Esto es, si $\forall\ \epsilon > 0$ y todo $x \in \mathcal{C}(F)$, $\exists\ N(\epsilon, x)$, tal que

$$| F_X(x) - F(x) | < \epsilon, \quad n > N$$

Note que convergencia en distribución no implica la convergencia de las correspondientes *f. d. p.* ó *f. m. p.*. Considere por ejemplo,

$$F_n(x) = \begin{cases} 0 & \text{si} \quad x < 1 - 1/n \\ 1/2 & \text{si} \quad 1 - 1/n \leq x < 1 + 1/n \\ 1 & \text{si} \quad x \geq 1 + 1/n \end{cases}$$

[2] $X_n(\omega) \longrightarrow X(\omega)$ para *"casi todo"* $\omega \in \Omega$, si la convergencia se da en todo $\omega \in \Omega$, excepto quizá para aquellos $\omega \in D$ pero tal que $\mathbb{P}(D) = 0$.

Entonces, $F_n(x) \longrightarrow F(x) = \mathbb{1}_{[1,\infty)}(x)$, que es la función de distribución que asigna toda la probabilidad al punto $x = 1$. Sin embargo, para todo $x \in \mathbb{R}$,

$$f_n(x) = \frac{1}{2}\mathbb{1}_{\{1-\frac{1}{n},1+\frac{1}{n}\}}(x) \longrightarrow f(x) = \frac{1}{2}\mathbb{1}_{\{1\}}(x)$$

que no es *f. d. p.*.

Proposición : Relación entre los modos de convergencia

a). Convergencia en probabilidad implica convergencia en distribución pero no viceversa. $X_n \xrightarrow{P} X \implies X_n \xrightarrow{D} X$.

b). Convergencia casi segura implica convergencia en probabilidad pero no viceversa. $X_n \xrightarrow{\text{cs}} X \implies X_n \xrightarrow{P} X$.

c). Convergencia cuadrática media implica convergencia en probabilidad pero no viceversa. $X_n \xrightarrow{\text{cm}} X \implies X_n \xrightarrow{P} X$.

d). Convergencia casi segura *no* implica convergencia cuadrática media *ni* viceversa.

11.2. Otros resultados límite

Teorema de continuidad de Lévy. Sea $\{F_n\}$ una sucesión de funciones de distribución y F una *f. p. a.*. Sean $\{\varphi_n\}$ y φ las correspondientes funciones características. Entonces,

I. Si $F_n(x) \longrightarrow F(x)$, $x \in \mathcal{C}(F)$, puntos de continuidad de F, entonces, $\varphi_n(t) \longrightarrow \varphi(t)$, para todo $t \in \mathbb{R}$.

II. Si para todo $t \in \mathbb{R}$, $\{\varphi_n(t)\}$ converge a una función $g(t)$, continua en $t = 0$, entonces, g es una función característica, y si F es la correspondiente *f. p. a.* entonces, $F_n(x) \longrightarrow F(x)$, para todo $\mathcal{C}(F)$,

Teorema de mapeo continuo. Sea $\{X_n\}$ y X *v. a.*'s y g una función continua. Entonces,

I. Si $X_n \xrightarrow{P} X$, entonces $g(X_n) \xrightarrow{P} g(X)$.

II. Si $X_n \xrightarrow{D} X$, entonces, $g(X_n) \xrightarrow{D} g(X)$.

Proposición : Sean $\{X_n\}$, $\{Y_n\}$, X y Y, *v. a.*'s. Entonces,

I. Si $X_n \xrightarrow{P} X$ y $Y_n \xrightarrow{P} Y$, entonces $X_n + Y_n \xrightarrow{P} X + Y$.

II. Si $X_n \xrightarrow{\text{cm}} X$ y $Y_n \xrightarrow{\text{cm}} Y$, entonces $X_n + Y_n \xrightarrow{\text{cm}} X + Y$.

III. Si $X_n \xrightarrow{P} X$ y $Y_n \xrightarrow{P} Y$, entonces $X_nY_n \xrightarrow{P} XY$.

Teorema de Slutsky. Sean $\{X_n\}$, $\{Y_n\}$, X y Y, *v. a.*'s y c una constante. Entonces,

I. Si $X_n \xrightarrow{P} X$ y $Y_n \xrightarrow{P} c$, entonces $X_n + Y_n \xrightarrow{P} c + X$

II. Si $X_n \xrightarrow{D} X$ y $Y_n \xrightarrow{P} c$, entonces $X_nY_n \xrightarrow{D} cX$

Nota: no necesariamente se tiene que si $X_n \xrightarrow{D} X$ y $Y_n \xrightarrow{D} Y$, entonces $X_n+Y_n \xrightarrow{D} X+Y$

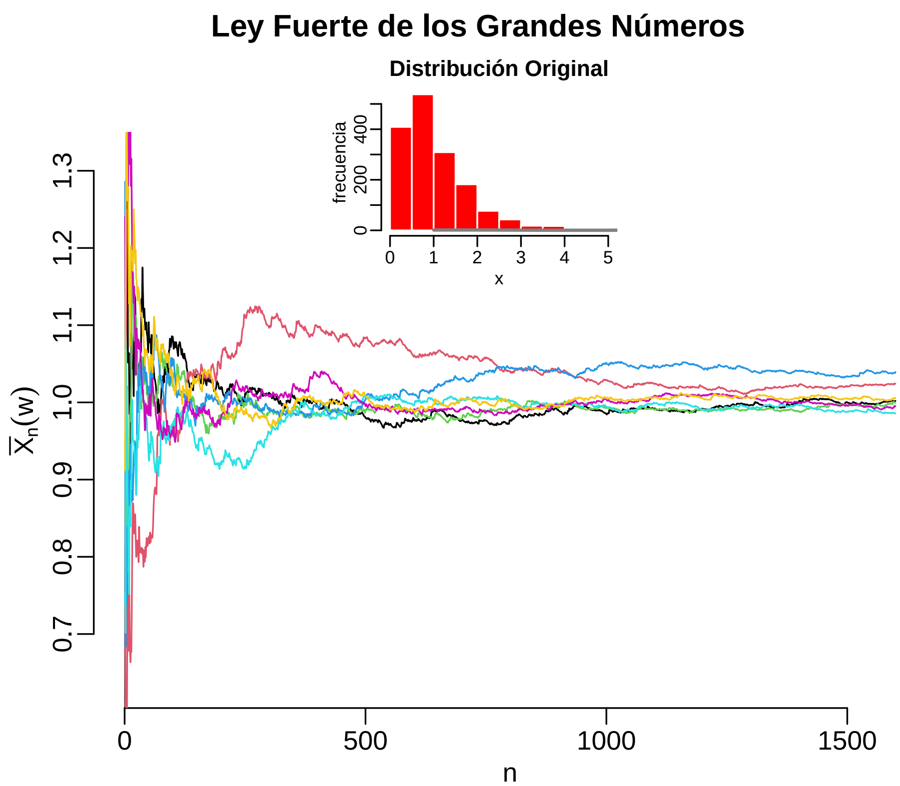


Figura 12: Las distintas trayectorias $\bar{X}_n(\omega)$ correspondientes a distintas $\omega \in \Omega$, van acercándose de valor de μ conforme crece n. En la figura se muestra un histograma de la distribución original Gamma.

11.3. Ley de los grandes números

Sean $X_1, X_2, \ldots$ *v.a.i.i.d.*'s, $S_n = X_1 + \cdots + X_n$ y $\bar{X}_n = \frac{X_1 + \cdots + X_n}{n} = \frac{1}{n} S_n$.

Ley de los grandes números (LGN).

I. **Versión fuerte (LFGN)**. Sean las X_n's con media finita μ. Entonces,

$$\bar{X}_n \xrightarrow{\text{cs}} \mu$$

Y viceversa, si $\bar{X}_n \xrightarrow{\text{cs}} c$, para alguna constante finita c, entonces, $\mathbb{E}[X]$ es finita y $\mathbb{E}[X] = c$.

II. **Versión débil (LDGN)**. Sean las X_n *v. a.*'s con media finita μ. Entonces,

$$\bar{X}_n \xrightarrow{P} \mu$$

Notas:

I. La figura 12 muestra la evolución de distintas **trayectorias** de $\bar{X}_n$, correspondientes a distintas $\omega \in \Omega$, para cuando crece n. El histograma del panel superior de la figura representa la distribución original de las X_i's.

II. La figura 13 muestra la concentración de $\bar{X}_n$, alrededor de la media μ, para cuando crece el tamaño de la muestra n. El histograma del panel superior de la figura representa la distribución original de las X_i's.

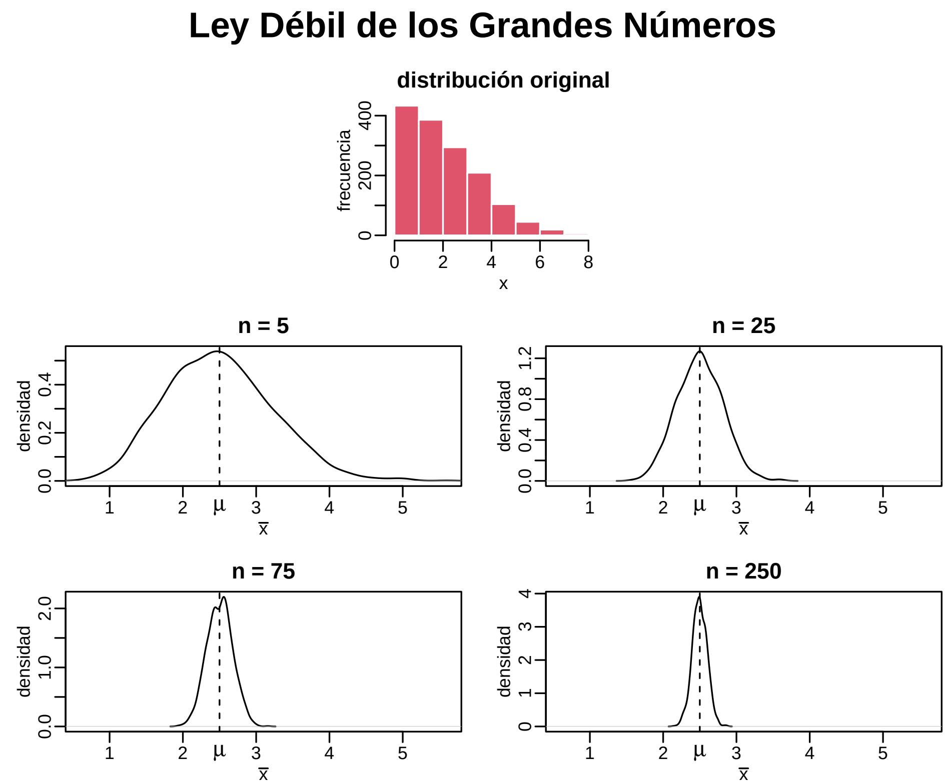


Figura 13: Los valores de $\bar{X}_n$ se va concentrando alrededor de μ conforme crece el tamaño de muestra n. En panel superior se muestra un histograma de la distribución original Poisson con media de 2.5.

Demostración:

I. De un nivel más elevado que el presente curso ...

II. La demostración se sigue del inciso anterior pues convergencia casi segura implica la convergencia en probabilidad. Sin embargo, si además supone que las *v. a.*'s tienen varianza σ^2 finita , su demostración sigue de la desigualdad de Chebyshev. A saber, sea $\epsilon > 0$,

$$\mathbb{P}\left(\mid \bar{X}_n - \mu \mid \geq \epsilon\right) \leq \frac{\sigma^2_{\bar{X}_n}}{\epsilon^2} = \frac{\sigma^2/n}{\epsilon^2} \longrightarrow 0$$

pues $\mathbb{E}[\bar{X}_n] = \mu$ y $\mathrm{var}(\bar{X}_n) = \sigma^2/n$.

Notas:

1. La constante ϵ puede verse como la precisión deseada en la aproximación de $\bar{X}_n$ a μ. La aproximación es buena para n *suficientemente grande*.

2. $\mathbb{P}\left(\mid \bar{X}_n - \mu \mid \geq \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2}$

3. Considere que $X_1, X_2, \ldots$ son *v.a.i.i.d.* con $X \sim \mathrm{Ber}(p)$. Entonces, $\mathbb{E}[X] = p$ y $\mathrm{var}(X) = p(1-p)$. Luego, $\mathbb{E}[\bar{X}_n] = p$ y $\mathrm{var}(\bar{X}_n) = p(1-p)/n$, por lo que para toda $0 < p < 1$,

$$\mathbb{P}\left(\mid \bar{X}_n - p \mid \geq \epsilon\right) \leq \frac{p(1-p)}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2}$$

pues la varianza es máxima cuando $p = 1/2$. Esto es, para toda p, $p(1-p) \leq \frac{1}{2}(1-\frac{1}{2}) = \frac{1}{4}$.

4. Considere dados $\epsilon, \delta > 0$ y suponga que se desea que $\mathbb{P}\left(\mid \bar{X}_n - \mu \mid \geq \epsilon\right) \leq \delta$. Se tiene entonces que

$$\mathbb{P}\left(\mid \bar{X}_n - \mu \mid \geq \epsilon\right) \leq \frac{p(1-p)}{n\epsilon^2} \leq \delta \quad \Longrightarrow \quad n \geq \frac{p(1-p)}{\delta\epsilon^2}$$

   esto es, la condición $\mathbb{P}\left(\mid \bar{X}_n - \mu \mid \geq \epsilon\right) \leq \delta$ se satisface para $n \geq \frac{p(1-p)}{\delta\epsilon^2}$.

   En el caso de *máxima incertidumbre* $p = 1/2$. se tiene $n \geq 1/4\delta\epsilon^2$.

5. Si $\epsilon = 0.05$ y $\delta = 0.10$, entonces $\delta\epsilon^2 = 2.5 \times 10^{-4}$, por lo que si desea que

$$\mathbb{P}(|\bar{X}_n - \mu| \geq 0.05) \leq 0.10, \qquad \text{entonces} \qquad n \geq \frac{1}{4(2.5 \times 10^{-4})} = 1000$$

   Si por alguna razón se considera que $p \approx 0.30$, $p(1-p) = 0.21$ y

$$n \geq \frac{0.2}{2.5 \times 10^{-4}} = 840$$

   y si $p \approx 0.15$, entonces $n \geq 510$.

11.4. Teorema central del límite

Sean $X_1, X_2, \ldots$ *v.a.i.i.d.*'s con media común μ_X y varianza σ_X^2 finita.

Recordar,

$$S_n = \sum_{i=1}^{n} X_i \qquad\qquad \bar{X}_n = S_n/n$$
$$\mathbb{E}[S_n] = n\mu_X \qquad\qquad \mathbb{E}[\bar{X}_n] = \mu_X$$
$$\text{var}(S_n) = n\sigma_X^2 \qquad\qquad \text{var}(\bar{X}_n) = \sigma_X^2/n$$

Luego, cuando $n \longrightarrow \infty$, se sigue que

$$\begin{array}{ccc} \mathbb{E}[S_n] \longrightarrow \pm\infty & ; & \mathbb{E}[\bar{X}_n] \longrightarrow \mu \\ \text{var}(S_n) \longrightarrow +\infty & ; & \text{var}(\bar{X}_n) \longrightarrow 0 \end{array}$$

Note al crecer n, en el caso de la suma S_n, ésta "explota", mientras que la *media muestral* $\bar{X}_n$, se "colapsa" en el sentido que converge a la constante μ.

La versión del teorema central del límite que aquí se presenta es conocida como de Lindberg-Lévy.

Teorema central de límite (TCL) Sean S_n y $\bar{X}_n$ como antes. Entonces, para las variables estandarizadas $Z_n = \frac{S_n - n\mu_X}{\sqrt{n\sigma_X^2}} = \frac{\bar{X}_n - \mu_X}{\sqrt{\sigma_X^2/n}}$, se tiene que

$$Z_n \xrightarrow{D} Z \sim \text{N}(0,1)$$

Demostración: Sin pérdida de generalidad suponga que las propias X_i's ya son estandarizadas, $\mathbb{E}[X_i] = 0$ y $\text{var}(X_i) = 1$ y suponga que X tiene función generadora de momentos m_X finita. Entonces,

I). $m_X(0) = \mathbb{E}[e^{0\cdot X}] = 1$.

II). $m'_X(0) = \frac{d}{dt} m_X(t)|_{t=0} = \mathbb{E}[X] = 0$

III). $m''_X(0) = \frac{d^2}{dt^2} m_X(t)|_{t=0} = \text{var}(X) = 1$

Se tiene también que

I). $m_{S_n}(t) = \left[m_X(t)\right]^n$, por ser *v.a.i.i.d.*.

II). $m_{\bar{X}_n}(t) = \left[m_X(t/n)\right]^n$.

III). $m_{\sqrt{n}\bar{X}_n}(t) = \left[m_X(t/\sqrt{n})\right]^n$.

Sea ahora $L(t) = \log(m_X(t))$. Entonces,

I). $L(0) = 0$.

II). $L'(0) = \dfrac{d}{dt}\log(m_X(0)) = \left.\dfrac{m'_X(t)}{m_X(t)}\right|_{t=0} = 0$

III). $L''(0) = \dfrac{d^2}{dt^2}\log(m_X(0)) = \left.\dfrac{m_X(t)m''_X(t) - \left[m'_X(t)\right]^2}{m^2_X(t)}\right|_{t=0} = \dfrac{1(1)-0^2}{1} = 1$

$$\log(m_{\sqrt{n}\bar{X}_n}(t)) = \log\left[m_X(t/\sqrt{n})\right]^n = n\log(m_X(t/\sqrt{n})) = nL(t/\sqrt{n}) = \frac{L(t/\sqrt{n})}{1/n}$$

$$\begin{aligned}
\lim_{n\to\infty}\frac{L(t/\sqrt{n})}{1/n} &= \lim_{n\to\infty}\frac{L'(t/\sqrt{n})(-\frac{1}{2}n^{-3/2})t}{-n^{-2}}, \qquad \text{Regla de L'Hopital}\\
&= \lim_{n\to\infty}\frac{L'(t/\sqrt{n})t}{2n^{-1/2}}\\
&= \lim_{n\to\infty}\frac{L''(t/\sqrt{n})(-\frac{1}{2}n^{-3/2})t^2}{-n^{-3/2}}, \qquad \text{Regla de L'Hopital}\\
&= \frac{t^2}{2}\lim_{n\to\infty}L''(t/\sqrt{n})
\end{aligned}$$

pues $L''(0) = 1$. Por lo tanto,

$$\log(m_{\sqrt{n}\bar{X}_n}(t)) = \frac{L(t/\sqrt{n})}{1/n} \longrightarrow \frac{t^2}{2}$$

y por continuidad

$$m_{\sqrt{n}\bar{X}_n}(t) \longrightarrow e^{t^2/2}$$

que corresponde a la *f. g. m.* de $Z \sim \mathrm{N}(0,1)$. Se sigue del teorema de continuidad de Lévy que

$$\sqrt{n}\bar{X}_n = \frac{\bar{X}}{1/\sqrt{n}} \xrightarrow{D} Z \sim \mathrm{N}(0,1)$$

El teorema se concluye considerando $W_i = \mu_X + \sigma_X X_i$.

La figura 14 muestra el histograma de promedios al considerar muestras de tamaño $n = 5, 20, 50, 100$. Al aumentar el tamaño de la muestra el histograma se acerca más a una curva "normal". El histograma del panel superior representa la distribución original de las X_i's, en este caso, de una distribución Beta.

Ejemplo : El tiempo de vida de unas lámparas se distribuye exponencial con tiempo medio de vida de 10 días. Tan pronto fallan son reemplazadas por otras idénticas. ¿Cuál es la probabilidad de que se necesiten más de 50 lámparas en un año?

Solución: $X \sim \mathrm{Exp}(\theta = 10)$. $S_n = X_1 + \cdots + X_n$.

$$\mathbb{P}\left(S_{50} < 365\right) = \mathbb{P}\left(Z_n < \frac{365 - 50(10)}{10\sqrt{50}}\right) \overset{\text{TCL}}{\approx} \Phi(-1.91) = 0.028$$

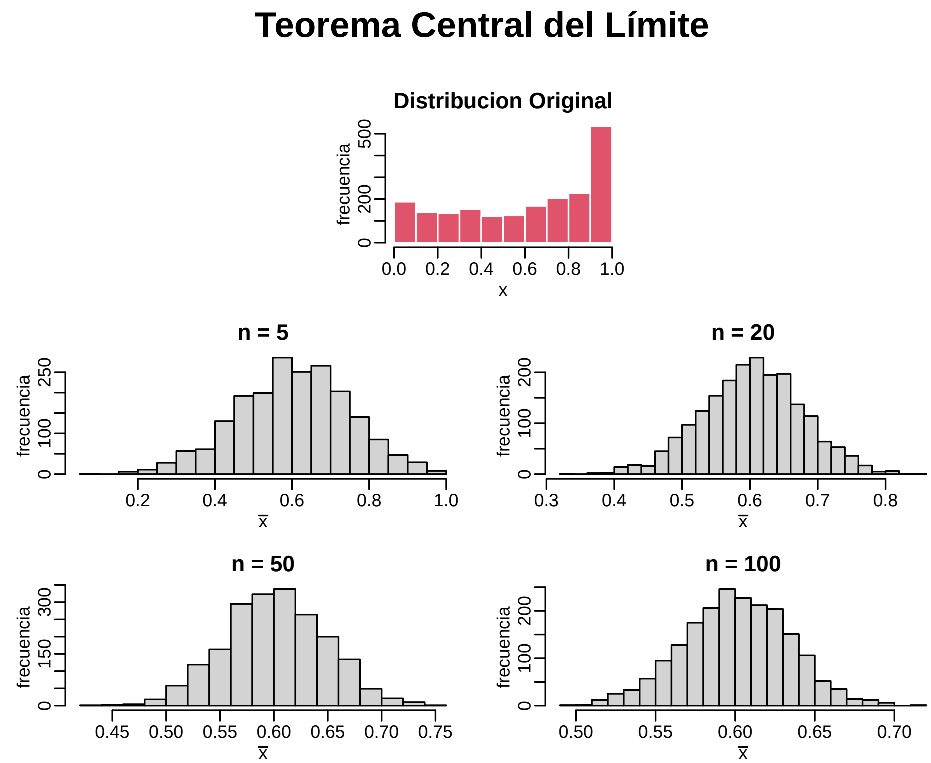


Figura 14: Los valores de $\bar{X}_n$ se va distribuyendo a manera de "campana" al rededor de μ conforme crece el tamaño de muestra n. En el panel central se muestra el histograma de la distribución original Beta.

Ejemplo : Sea $X_1, X_2, \ldots$ *v.a.i.i.d.*'s con media μ y varianza σ^2. Sean $S_n = X_1 + \cdots + X_n$, $\bar{X}_n = \frac{1}{n} S_n$. Aproxime $\mathbb{P}\left(\mid \bar{X}_n - \mu \mid \geq c\right)$.

Solución: Se sigue del teorema central del límite que para n *grande*,

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \dot{\sim} \mathrm{N}(0,1)$$

Entonces,

$$\begin{aligned}
\mathbb{P}\left(\mid \bar{X}_n - \mu \mid \geq c\right) &= 1 - \mathbb{P}\left(-c \leq \bar{X}_n - \mu \leq c\right) \\
&= 1 - \mathbb{P}\left(\frac{-c}{\sigma/\sqrt{n}} \leq \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq \frac{c}{\sigma/\sqrt{n}}\right) \\
&= 1 - \left[\mathbb{P}\left(Z_n \leq \frac{c}{\sigma/\sqrt{n}}\right) - \mathbb{P}\left(Z_n \leq \frac{-c}{\sigma/\sqrt{n}}\right)\right] \\
&\overset{\text{TCL}}{\approx} = 1 - \left[\Phi\left(\frac{c}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{-c}{\sigma/\sqrt{n}}\right)\right] \\
&= 2\Phi\left(\frac{-c}{\sigma/\sqrt{n}}\right)
\end{aligned}$$

Ejemplo : Se toma una muestra aleatoria de tamaño n para determinar el porcentaje que votará por el candidato X.

Sean $X_1, X_2, \ldots$ *v.a.i.i.d.*'s $X_i \sim \text{Ber}(p)$. Luego, $\mathbb{E}[X] = p = \mu_X$ y $\text{var}(X) = p(1-p) = \sigma_X^2 \leq 1/4$.

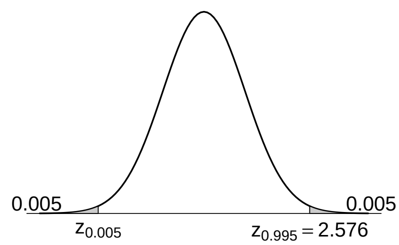
a). Si $n = 900$,

$$\mathbb{P}\left(\mid \bar{X}_n - p \mid \geq 0.025\right) \overset{\text{TCL}}{\approx} 2\Phi\left(\frac{-0.025}{\frac{1}{2}/\sqrt{900}}\right) = 2\Phi(-1.50) = 2(0.067) = 0.1336$$

b). $\mathbb{P}\left(\mid \bar{X}_n - p \mid \geq c\right) = 0.01$. Determine c si $n = 900$.

$$2\Phi\left(\frac{-c}{\sigma/\sqrt{n}}\right) = 0.01$$

$$\begin{aligned} c &= -z_{.005}\sigma/\sqrt{n} \\ &= 2.576 \cdot \frac{1}{2}/\sqrt{900} \\ &= 0.043 \end{aligned}$$



c). Determine n de manera que $\mathbb{P}\left(\mid \bar{X}_n - p \mid \geq 0.025\right) = 0.01$.

$\mathbb{P}\left(\frac{|\bar{X}_n - p|}{\sigma/\sqrt{n}} \geq \frac{0.025}{\sigma/\sqrt{n}}\right) = 0.01$.

Por otro lado, $z_{.995} = 2.576 = \frac{.025}{\frac{1}{2}/\sqrt{n}}$. Así,

$$\sqrt{n} = \frac{2.576}{.025} \cdot \frac{1}{2} = 51.517 \quad \text{y por lo tanto} \quad n \approx 2654$$

11.5. Ejercicios

Refiérase al Cuaderno de Ejercicios sección 11, Barrios and Chambon (2024).

Referencias

Barrios, E. (2024). Apuntes para el curso de Cálculo de Probabilidades I. https://gente.itam.mx/ebarrios/docs/apuntes_CP1.pdf. (6 de enero de 2025).

Barrios, E. and P. Chambon (2024). Cálculo de Probabilidades II – Cuaderno de Ejercicios. https://gente.itam.mx/ebarrios/docs/CuadernoEjercicios_CP2.pdf. (6 de enero de 2025).

Blitzstein, J. K. and J. Hwang (2014). *Intorduction to Probability*. Boca Raton, FL: CRC Press.

Box, G. E. P. and M. E. Muller (1958). A Note on the Generation of Random Normal Deviates. *Annals of the Mathematical Statistics 29*(2), 610–611.

Graybill, F. A. (1961). *Introduction to Linear Statistical Models (Vol I.)*. New York: McGraw-Hill.

Harris, B. (1966). *Theory of Probability*. Reading, MA.: Addison-Wedsley.

Hoel, P. G., S. C. Port, and C. J. Stone (1971). *Introduction to Probability Theory*. Boston: Houghton Miffling Company.

León-García, A. (2008). *Probability, Statistics and Random Processes for Electrical Engineering* (3 ed.). Reading, Massachusetts: Addison-Wesley.

Mood, A. M., F. A. Graybill, and D. C. Boes (1974). *Introduction to the Theory of Statistics* (3rd ed.). Singapore: McGraw-Hill.

Rincón, L. (2014). Introducción a la Probabilidad. https://filedn.com/lwDqCORMczI54WhtiTmXwGp/flip-proba1/mobile/index.html. (6 de enero de 2025).

Ross, S. (2006). *A First Course in Probability* (7th ed.). Upper Saddle River, N.J.: Prentice Hall.

Roussas, G. G. (1997). *A Course in Mathematical Statistics* (2nd ed.). San Diego, CA.: Academic Press.

Schott, J. R. (1997). *Matrix Analysis for Statistics*. New York: Wiley.

Wackerly, D. D., W. Mendenhall III, and R. L. Scheaffer (2008). *Mathematical Statistics with Applications* (7 ed.). Australia: Thomson.